Coursework Declaration and Feedback Form

The Student should complete and sign this part

Student Number: 2615939L	Student Name: Zhengye LI
Programme of Study (e.g. MSc in Electronics and Electrical Engineering):	
Course Code: ENG5059P	Course Name: MSc Project
Name of **First** Supervisor: Umer Ijaz	Name of **Second** Supervisor: Ankush Aggarwa
Title of Project: Community Assembly of Anaerobic Microbiomes During contamination and application to wastewater	

Declaration of Originality and Submission Information

I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements Signed (Student) : Zhengye Li	
Date of Submission : August 18, 2022	

Feedback from Lecturer to Student – to be completed by Lecturer or Demonstrator	
Grade Awarded: Feedback (as appropriate to the coursework which was assessed):	
Lecturer/Demonstrator:	Date returned to the Teaching Office:

Community Assembly of Anaerobic Microbiomes During contamination and application to wastewater

Zhengye Li (Student ID: 2615939L)
Supervisor: Dr.Umer Zeeshan Ijaz
Co-Supervisor: Dr.Ankush Aggarwal

Catalogue

Contents

Abstract

Complex microbial communities are ubiquitous in nature(Aguirre de Carcer, 2019). I investigate the assembly of an anaerobic microbial community for particle flotation driven by a combination of stochastic and deterministic processes. The study was mainly divided into two parts, particle flotation and assembly of anaerobic biomes. Anaerobic particles in these microbial communities present different forms of flotation combination in biological wastewater. During anaerobic digestion of anaerobic organisms, stochastic and deterministic mechanisms dominate. The randomness and determinacy of microbial communities in wastewater treatment are also analyzed in this paper. Particle flotation is an easy problem in the process of anaerobic digestion analysis. There are many reasons for particle flotation, such as composition and location of archaea and bacteria , and filamentous expansion caused by excessive growth of methanogens. In this study, the specific reasons for flotation (the precipitation particles in the flotation particles are rich in anaerobic bacteria, transporters and Arctobacter) were obtained by horizontal comparison of several samples. In addition to this, the assembly of microbial communities was also analyzed by the zero-model approach. Community assembly of anaerobic microbiota during particle flotation is the central idea of this study. I obtained the exact answer of flotation and assembly by analyzing DNA and cDNA of anaerobic biomes in P language. At the same time, several other problems regarding microbial communities in anaerobic digestion were identified during the analysis of the problems.

Keywords: sewage treatment, microbial community, anaerobic digestion, community assembly, particle flotation, randomness, determinacy.

Acknowledgements

Thank you to Dr. Umer Zeeshan Ijaz for his help and guidance on all aspects of bioinformatics and R language in this project. I would like to thank Dr. Uzma for its support in helping us understand the topics covered in this study and for urging and encouraging the progress of my thesis.

1 Background and Aims

1.1 Sewage treatment

The type of sewage treatment process not only determines the sewage purification effect, but also determines the sludge production and sludge treatment method in the sewage purification process.

In the process of sewage treatment, the anaerobic biological treatment of wastewater is also known as anaerobic digestion and anaerobic fermentation in the early stage(Zhang et al., 2020). It refers to the decomposition of organic matter and the production of CH4 and CO2 under anaerobic conditions by the combined action of a variety of (anaerobic or facultative) microorganisms (e.g., methanogens).

Studies in anaerobic microbiology have shown that methanogens are a very special class of archaea (Archea). In addition to taxonomy and its special structure, the main feature of methanogenic bacteria is that they can only use some simple organic matter as substrate. These are mainly some simple one-carbon substances such as formic acid, methanol, methylamines and H2/CO2. Acetic acid is the only two-carbon substance, and other fatty acids containing two carbons or more and alcohols other than methanol cannot be used(Winter et al., n.d.).

1.2 Anaerobic oxidation

Anaerobic digestion in water treatment industry has always been favored by environmentalists, since it has good removal effect, higher reaction rate and to better adapt to the toxic substance, it is more important because of its relative aerobic biological treatment for wastewater is not need to provide a large amount of energy consumption of oxygen transfer, makes the anaerobic biological treatment are widely used in water treatment industry.

The anaerobic digestion process of microbial communities in sewage consists of four parts. The first stage is hydrolysis, followed by fermentation (or acidification) and acetic acid, and finally methanogenesis.

Widely used anaerobic technology(Lettinga et al., 2008): the sludge bed area of ASB reactor is mainly composed of anaerobic sludge with good settling

performance, and the concentration can reach 50-100g/L or higher. Precipitation floating zone mainly depends on the reaction process of the rise of the gas formed by the mixing, sludge concentration is low, usually within 5 to 40 g/L, in the upper portion of the reactor is equipped with gas (methane), solid (sludge), liquid (water) three-phase separator, separator first bending to generate methane bubbles rising process, through the water into the chamber through the catheter. After degassing, the mixed liquid is further separated from the solid and liquid in the settlement area, and the sludge under settlement returns to the reaction area, so that a large number of microorganisms accumulate in the reaction area. The wastewater to be treated is entered by the bottom distribution water system, and the clarified treated water is discharged by overflow from the precipitation area. In UASB reactor can you get a good settlement wins and methane-producing activity of anaerobic sludge particles(Reeburgh, 1980), aspect ratio and relative to other reactor has some advantages: the relative density of granular sludge is smaller than artificial carrier, accomplished by a gas sludge's full contact with the substrate, dispense mixing and reflux sludge equipment and energy consumption; The application of three-phase separator saves the auxiliary degassing device; Granular sludge sedimentation performance is good, avoid the attachment of precipitation separation device and reflux sludge equipment: there is no need to add filler and carrier in the reactor, improve the volume utilization rate.

1.3 Mechanisms of microbial community assembly

Two ecological process theories (based on niche process theory and neutral process theory) jointly describe the mechanism of microbial community formation (assembly).Niche theory postulates that deterministic factors such as species characteristics, interspecies interactions, and environmental conditions control community structure and metabolic functions.(Panosyan et al., 2021) That is, microbial communities are formed by deterministic biological factors (species interactions, such as competition, predation) and abiotic factors (environmental factors, such as pH, and temperature) resulting from different habitat preferences and adaptations of microorganisms(Aguirre de Carcer, 2019). However, the neutral process theory assumes that microorganisms

exhibit stochastic equilibrium in the loss and increase of taxa, that is, stochastic processes (birth, death, migration, speciation, dispersal limitation) shape microbial community structure.
In short, microbial community assembly/community structure is shaped by both deterministic and stochastic ecological processes.

1.4 Deterministic and stochastic processes

Deterministic processes result from the predictable filtering effects of ecological selection on species imposed by biotic and abiotic factors(Wang et al., 2021), which affect the fitness of organisms and thus determine the composition and relative abundance of species. In contrast, stochastic processes involve random births, deaths, probabilistic diffusion, and random changes in the relative abundance of species (ecological drift) and are not the result of environmentally determined fitness. The resulting patterns of species composition are indistinguishable from those generated by random chance alone("MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS", 2022). (It may be that factors affecting species composition are hard to distinguish.)
Deterministic processes include selection of abiotic environmental factors (environmental filtering) and mutual antagonism and synergy between species.
Stochastic processes include unpredictable disturbances, probability diffusion, and random birth-death events(Stegen et al., 2012).
Determinacy and randomness determine the formation of microbial communities.

1.5 Granular sludge & particle flotation

Granular sludge is a special biofilm for the phenomenon of microbial auto condensation found in sewage treatment. According to whether the microbial growth needs oxygen or not, it can be divided into anaerobic granular sludge and aerobic granular sludge(Zhang et al., 2020), which can be converted into each other theoretically. Granular sludge has high mass concentration, and good sedimentation performance (reducing sedimentation tank volume), and there is a large matrix concentration gradient inside (providing an environment conducive to microorganisms). Sewage treatment is used in UASB and SBR.

The formation of granular sludge requires the start-up and operation of the sludge bed for a certain period.

The anaerobic reactor's formation of granular sludge is divided into three stages. The first stage is the initial stage, which mainly carries out sludge domestication to adapt to the capacity of treating the organic matter in wastewater. The second stage is to make the flocculated sludge to granular sludge transformation, so it is necessary to timely improve the loading rate so that the microorganisms can get enough nutrition, so that the gas production and upper-speed increase, causing the sludge bed expansion.(Wang et al., 2021) The third stage is the granular sludge culture period, to realize all sludge granulation and make the reactor reach the highest volume loading rate.

Particle floating refers to the situation in that part of the sinking particles float up in the flow anaerobic sludge bed (UASB) bioreactor(Uemura & Harada, 1993), resulting in the reduction of sewage treatment speed. As a rule, granular sludge should be at the bottom of UASB. Particle flotation occurs for several reasons. For example, the PH value is too low, filamentous bacteria are overgrown, and biogas generation in UASB affects microbial community polymerization(Lettinga et al., 2008), etc. If the cause of particle floating is found and effectively improved, the efficiency and cost of sewage treatment can be greatly improved. In this study, the floating particles and sinking particles were analyzed by R language, and the different DNA and cDNA in floating particles were obtained. When a specific microorganism is found, it can be preliminarily determined that the class of organisms may have an impact on particle floating.

1.6 Aims and Objectives

Composition of anaerobic particles and specific changes during anaerobic digestion.

The structure and assembly patterns of microbial communities.

Cause and form of particle flotation.

2 Methods

2.1 Microbial community sampling

The data for this study were obtained from the database.

2.2 DNA/RNA co-extraction and cDNA synthesis

DNA and RNA were co-extracted from particulate biomass collected from flotsam (n=6DNA and n=6RNA) and the bioreactor layers (n=6DNA and n=6RNA) were precipitated (n= 24 samples in total).

2.3 Statistical analysis

2.3.1Alpha Diversity

Alpha diversity analysis is an important part of microbial diversity analysis. It mainly focuses on the degree of species diversity in local uniform habitat and studies the abundance of species in a sample or multiple samples and the evenness of individual allocation in the community. Alpha diversity index mainly includes Chao1, Ace, Shannon, Simpson, Richness, Goods Coverage, Pielou, Invsimpson and so on("The Use and Types of Alpha-Diversity Metrics in Microbial NGS - CD Genomics", n.d.). Richness- Species abundance; Ace index - assessing the number of OTUs in the community; Chao index - estimated the number of OTUs in the sample by Chao1 algorithm; Shannon& Simpson on behalf of microbial diversity; Pielou index - estimated evenness;

2.3.2 Beta Diversity

For environmental samples, different biomes are often distributed in different samples. Quantifying the differences between these biomes can help researchers understand not only how biologically diverse individual samples are, but also why samples are clustered or dispersed in the way they are. This involves another expression of biome diversity, Beta diversity.

Beta diversity is proposed by Whittaker in 1960 (Li et al., 2017)and defined as the extent to which community composition changes, or the degree to which communities diverged, in relation to complex gradients of the environment or

patterns of the environment. Beta diversity depicts the magnitude in the change of species from one ecosystem to another. β diversity is another name for sample dissimilarity. It quantifies the difference in overall categorical composition between the two samples.

Broadly speaking, Beta diversity analysis consists of two parts: the calculation of distance and the presentation of distance. Beta diversity can not only reflect the diversity distance relationship between samples, but also reflect the degree of differentiation between biological communities.

2.3.3 NRI and NTI

In the NRI and NTI models, the shortest branch length connecting two species on the evolutionary tree represents the evolutionary distance between the two species("NTI.p function - RDocumentation", n.d.). Mean phylogenetic distance (MPD) is the average phylogenetic distance between pairs of all species in a community(Shulakov et al., 2017). Mean nearest phylogenetic taxon distance (MNTD) represents the average evolutionary distance between any species and the closest phylogenetic species in the community. The smaller their value, the closer the species are related to each other. By comparing the MPD and MNTD of the two communities(Hu et al., 2015), we found that the species in community 1 indeed had higher similarities than those in community 2. The number of species varies between communities, which can affect the results of MPD and MNTD. To compare different communities and avoid the influence of different species numbers on MPD and MNTD, Webb et al. (2008) used the Monte-Carlo algorithm to standardize MPD and MNTD, so as to obtain NRI and NTI.

2.3.4 Observation of the top-25 most abundant taxa bar

Microbial communities in floating and settling particles are depicted as stacked bar plots of the relative abundance of the 25 most abundant taxa based on differences in 16S rRNA genes.

2.3.5 Subset Analysis

Subset analysis is based on the selection of the best model according to specific statistical criteria used to determine (evaluate each possible combination of explanatory factors)(*"10.3 - Best Subsets Regression, Adjusted R-Sq, Mallows Cp | STAT 501", n.d.)*.

2.3.6 DeSeq2

DESeq2 is an R package designed for normalization, visualization, and differential expression analysis of high-dimensional quantitative data. It uses Empirical Bayes Techniques to estimate the prior values of log-multiple changes and deviations and computes the posterior values of these statistics(Love et al., 2014).

It was published by Professor Michael Love(michaelisaiahlove@gmail.com) of the University of North Carolina in 2014 and is still being updated and maintained. It is currently the most used R package for differential expression analysis.

2.3.7 Core Heatmap

Core Heatmap is used to describe changes in the core biome. On the one hand,(Yi, n.d.) Core Heatmap marks the parts that need to be analyzed and studied through a grid of colored squares. In the image of the Core Heatmap, the changes in each cell can be clearly seen. Using Core Heatmap, I can analyze which types of microbial communities stand out.

In addition, through the analysis of different Core Heatmap groups, the changes of microbial communities under different parameters can be obtained, and the normalized model can be designed to obtain reasonable solutions (certainty and randomness, diversity, and richness).

2.3.8 Null Modelling (QPE)

The null model is "Compare the result of your partition to the difference of a random network". The Null model can be thought of as the origin of a reference frame, which selects the random network, that is, the network without any structure(hypothesis? et al., 2017). So, the more different you get from the

origin of the reference frame for a network, the better structured the partition is for your network. Therefore, it can be said that Modularity is not performing community discovery, but the evaluation of network structure. Communities are just one manifestation of this structural nature.

QPE algorithm is a quantum phase estimation algorithm("Questions about Quantum Phase Estimation", 2022). This algorithm has been proposed for a long time, but what really brings great impact is the HHL algorithm based on it and various quantum machine learning algorithms based on the HHL algorithm. The most magical effect of phase estimation is to achieve an exponential acceleration compared to algorithms running on traditional computers. A simple example would be to run 2 to the 30th power (about a billion operations) on a traditional computer, but only about 30 times on a quantum computer. Of course, this exponential acceleration effect is conditional on the input and output being qubits. With its advantages, this algorithm can be used in many applications, such as order solving problems, factorization problems and quantum machine learning HHL algorithm.

2.3.9 CODA GLMNET (Multivariate analysis,)

Glmnet is a software package that fits generalized linear and similar models by penalized maximum likelihood. For a grid of values of the regularization parameter lambda, on a logarithmic scale, the regularization path for the net penalty of the lasso or elasticity is calculated.(Tay et al., 2021)In the study I used CODA GLMNET to carry out Multivariate analysis of microbial communities.Multivariate analysis, also known as multivariate statistical analysis, is an important subdiscipline of statistics, which mainly studies the relationship between sets of multiple variables and the relationship between individuals with these variables. Multivariate analysis is widely used in both natural and social sciences, in both theoretical research and applied decision-making(Dempster, 1971). In recent years, with the popularization of computers, the application of multivariate analysis is becoming more and more extensive and in-depth. Multivariate ANALYSIS, SIMPLY PUT, IS a method OF STUDYING THE characteristics of a set of N observations, in which P factors (variables) are observed for each object. The set of observations can be the

whole set or one or more subsets from a larger set; It can be fixed or random: it can be a type variable or a quantity variable: a quantity variable can be continuous or discrete; And the set of variables itself can be one or more subsets taken from a larger set of variables. From the point of view of application, multivariate analysis is to study the relationship between multiple variables, but there is no strict limit on what content is the content of multivariate analysis. It is generally believed that the typical multivariate analysis can be mainly summarized into two types of problems: the first type is to determine the ownership of a certain element, and the second type is to try to reduce the dimension of variables and change variables into independent variables at the same time, so as to better explain the relationship between multiple variables. The second type of method can also be referred to as ranking analysis. In ecological environmental monitoring, ranking analysis can also be divided into two basic categories: environmental ranking, in which the axes are composed of environmental data, and community ranking, in which the axes are composed of community data. The former is ranked by environmental factors, also known as direct ranking; The latter is ranked by the attributes of the community itself (such as the occurrence or not of species, the frequency and coverage of species, etc.), also known as indirect ranking.

2.3.10 Normalized stochasticity Ratio

Normalized stochasticity Ratio (NST), as deterministic, < 50%) and stochastic, & GT; 50%) at the boundary point. NST has limited performance at large spatial scales or under very high ambient noise conditions, but it shows high accuracy (0.90 to 1.00) and accuracy (0.91 to 0.99) in all other simulation scenarios. In addition, zero-model construction algorithms and measures of community similarity have a strong impact on quantifying randomness. The results obtained by NST are higher than the previous ST and NP algorithms in terms of accuracy and precision compared with reality. As the research scale increases (P-S-R-C-G), the accuracy and precision of NST decrease. Different distance algorithms (based on abundance or incidence) can significantly affect the accuracy and precision of NST.

2.3.11Different null model establishment methods

If you don't consider abundance (constrain taxa occurrence frequency), Equiprobable said all species occurrence probability is the same; Proportional means that the occurrence rate is the same as the observed occurrence probability; Fixed means the probability of occurrence is Fixed.

If consider abundance (constrain richness), Equiprobable said species appear probability is the same in all samples; Proportional means that the occurrence rate in a sample is the same as its observed abundance; Fixed means the probability of occurrence is Fixed.

Different null models will also have a large impact on the results.

3 Results

3.1 Taxonomic Identification Comparison

In this study, a variety of samples were extracted to compare DNA and cDNA and to determine whether they were in a Floating or a summary state. From this, I can determine the Floating (top) and summary microbial categories in the microbial community. This allows me to determine the class of microorganisms that occur in particle floating and why when anaerobic organisms in sewage treat sewage. By comparing the four sets of data, I can find out whether microbial community assembly is deterministic or random (or jointly determined), as well as the specific microbial species and assembly process of microbial community assembly. In summary, both particle floating of granular sludge in microbial communities and assembly of microbial communities (diversity and richness, deterministic and stochastic) can be concluded by data comparison.

It is worth mentioning that there are many methods to compare data through R language analysis, which together constitute a relatively comprehensive analysis.

3.2 Diversity Patterns

3.2.1 Alpha Diversity

According to the analysis of Figures I and II, the variation difference between community diversity and community richness of α diversity was observed(Li et al., 2021).

First, by comparing DNA and cDNA in the image, it can be clearly seen that the AVS of DNA is more different, which represents higher diversity of DNA. At the same time, we can see that the community richness of cDNA is higher than that of DNA. DNA is more homogeneous than cDNA.

Second, the Floating and summary genes in the images are compared. It can be clearly observed in the DNA that Floating DNA is slightly more abundant than ~~the~~ fashion DNA in terms of both diversity and richness.

However, for cDNA, Floating cDNA has a similar abundance (cDNA is slightly higher than DNA), and Floating cDNA has a higher diversity and uniformity. This can also be demonstrated through Pielou's Evenness Index (this indicator is strongly influenced by Observed OTU/ASV, which is one of the main shortcomings of this indicator).

Conclusion: The microbial community in the Floating state was higher than that in summary. The genetic material of bacteria and archaea is DNA. Further analysis of the above-described content, bacteria and archaea play a role in sewage treatment, at the same time, they are also very active in anaerobic digestion.

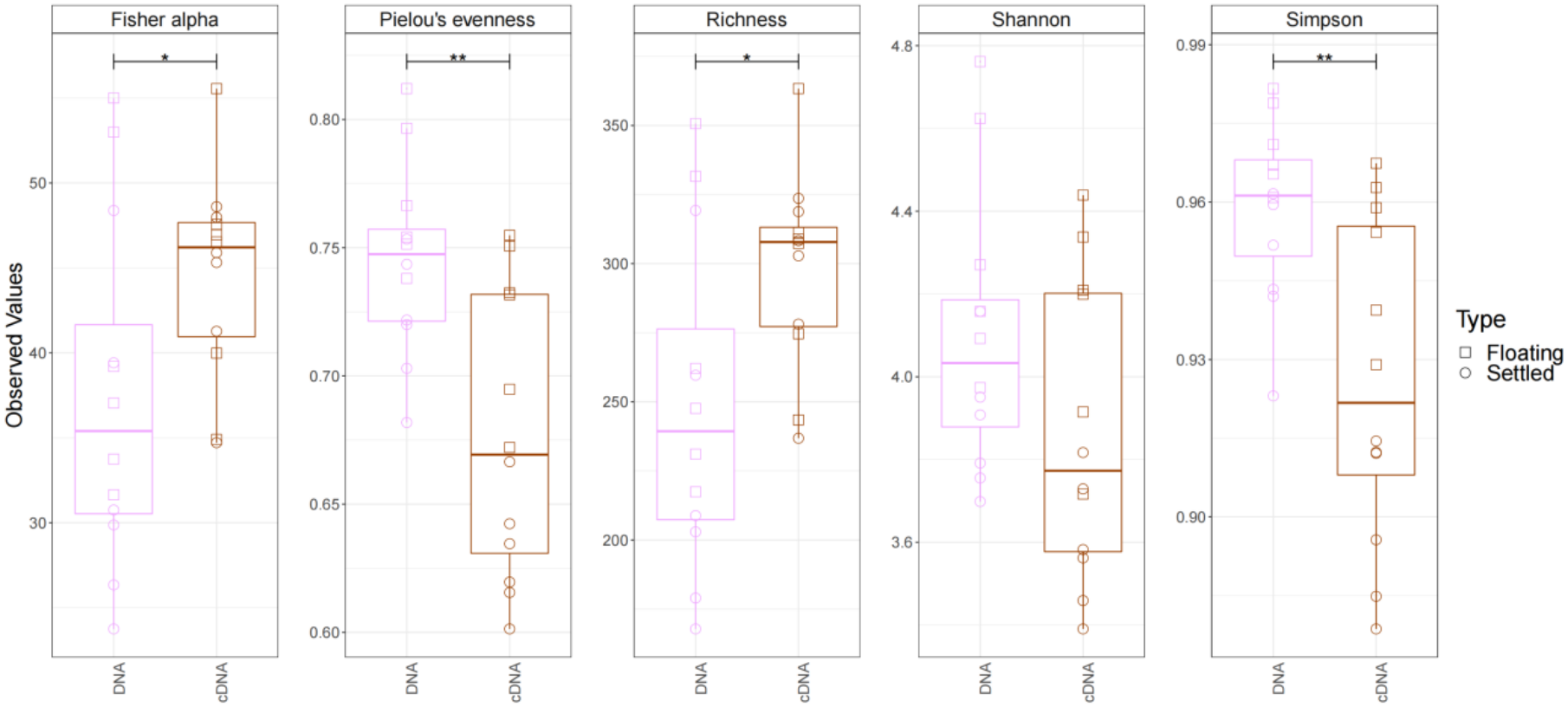


Figure 1. Boxplots of DNA and cDNA in Fisher Alpha, Pielou's Evenness, Richness, Shannon, and Simpson

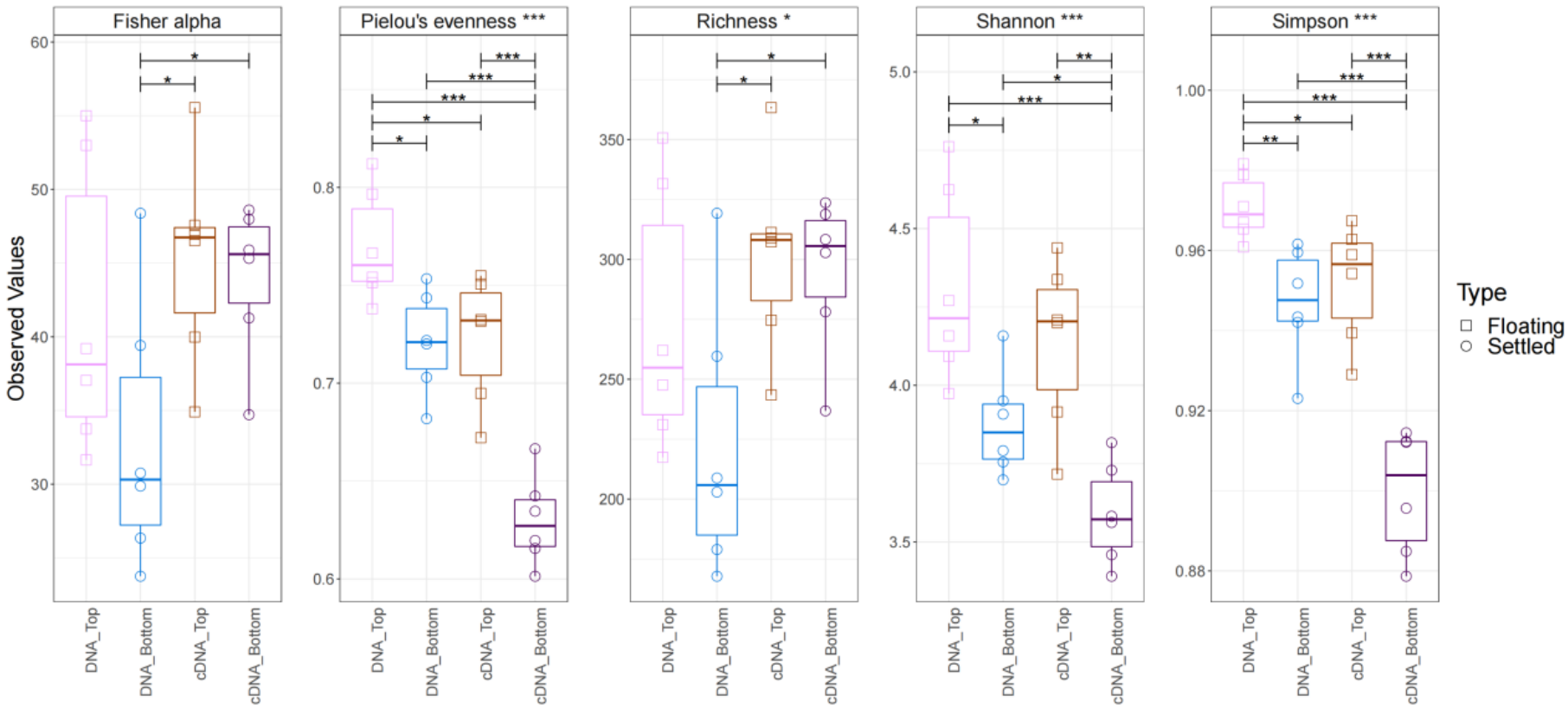


Figure 2. Boxplots of DNA_Top&DNA_Bottom and cDNA_Top&cDNA_Bottom in Fisher Alpha, Pielou's Evenness, Richness, Shannon, and Simpson

3.2.2 Beta Diversity

The Floating and summary states were similar for DNA and for CDNA. This indicates that DNA compositions in Floating and summary states are structurally similar, and cDNA compositions in Floating and summary states are structurally similar. However, DNA and cDNA have different community distribution. Samples with high similarity in community structure tend to cluster together, and samples with very different communities will be far apart.

In settling particles, ASV is significantly reduced. The rarefied richness (ASV) between floating and settling particles is statistically similar. In summary, the distribution of their biome is different, but the number and diversity of microorganisms remain stable. Floating and settling particles are quite similar in terms of both abundance and phylogeny.

The results obtained by β-diversity are slightly different from α-diversity, because diversity is used to measure how diverse the microbiota is within an individual, note that it is a single individual, and does not involve comparisons between individuals. β diversity was calculated to characterize an indicator of The alpha diversity index also affects the beta diversity index. In addition, the graph of β diversity can also represent the similarity of OTU data between samples. As can be seen from the picture, cDNA grouping and DNA grouping

make those with different OTU, but they have a high degree of similarity.

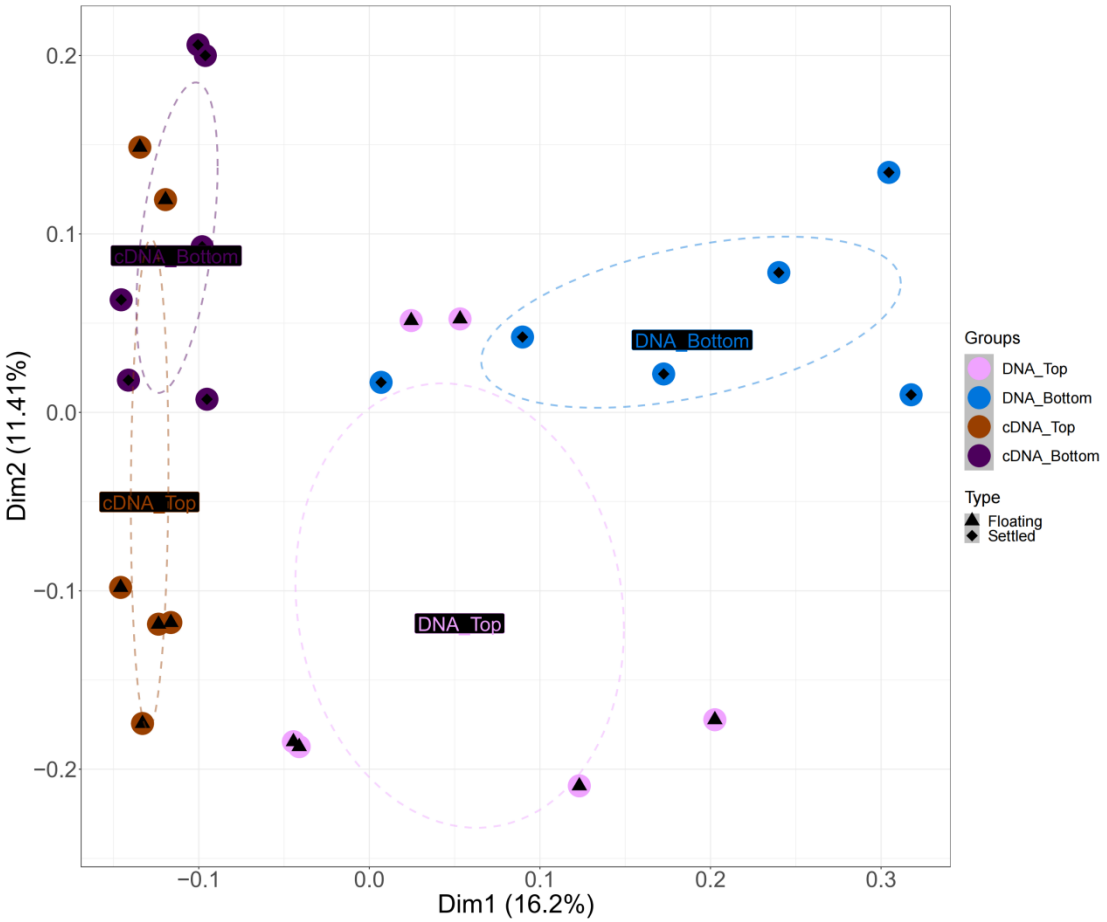


Figure 3. Microbial community distribution

Principle coordinate analysis UniFrac distances.

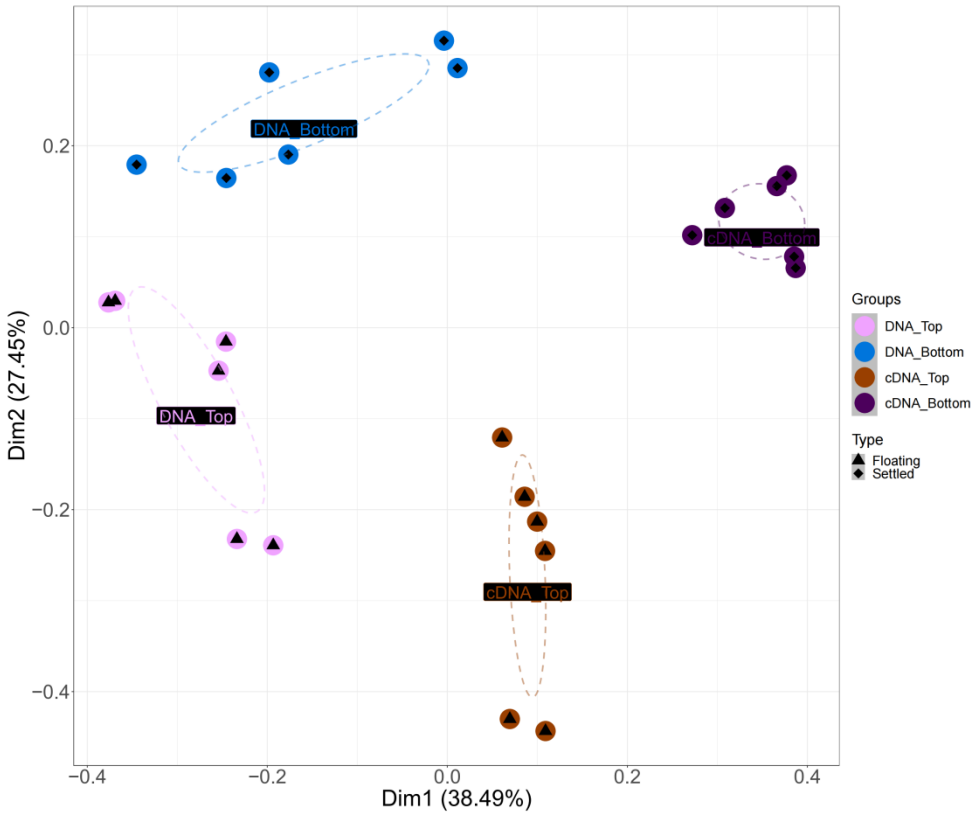


Figure 4. Microbial community distribution

Principle coordinate analysis Bary distances.Distance between groups indicates significant differences in nucleic acids, locations, and reactors.

3.2.3 Taxa bar

The following figure depicts the microbial communities of four data sets (Floating (top) and summary (Bottom) groups for DNA and Floating (top) and summary (Bottom) groups for cDNA. From the data in the figure, Bacteria; *Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae;* Uncultured and Bacteria; *Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidetes vadinHA17;* Uncultured bacterium has the largest proportion in the sample group of DNA. In the two groups of cDNA samples, Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosaetaceae; The proportion of Methanosaeta is much greater than its proportion in DNA. In addition, Bacteria in the cluster of "Floating (top)" and "Bottom" as parameter changes, Bacteria in the cluster of "Bottom"; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Lactococcus has a significantly higher proportion than it does in the Floating (top) grouping. Moreover, in terms of color species, the

microbial species in the DNA group were slightly more than those in the cDNA. This means that their community richness and diversity are relatively higher. Furthermore, in the sewage treatment process of anaerobic digestion by microbial community, some microorganisms are not completely expressed, but most of them play an active role in the process of anaerobic digestion. It is worth mentioning that archaea are prokaryotes, which, like bacteria, do not have a nuclear envelope, and their DNA exists in a circular form. The greater proportion of archaea in the cDNA implies that they have a strong and potentially positive role in the community.

In addition, in the four groups of sample data, the gray component accounts for a large proportion, which means that there are certain errors in the analysis of microbial communities using TAXT method. These errors have greatly disturbed the analysis of community diversity and richness.tentially positive role in the community.

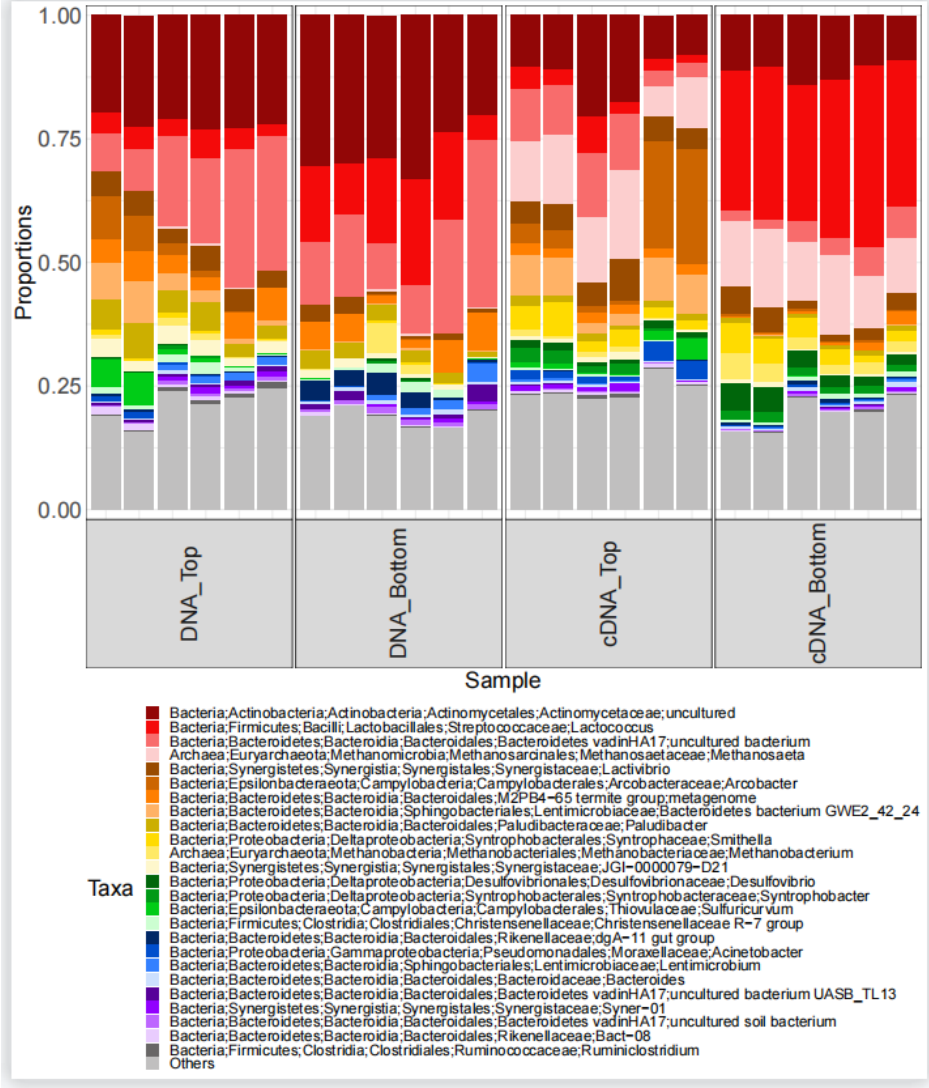


Figure 5. Microbial community in floating and settled granules depicted as a stacked bar chart of the relative abundance of the 25 most abundant taxa , and where “others” represent everything that is not in the top-25.25 the most abundant microbial communities.

3.3 Null Modelling

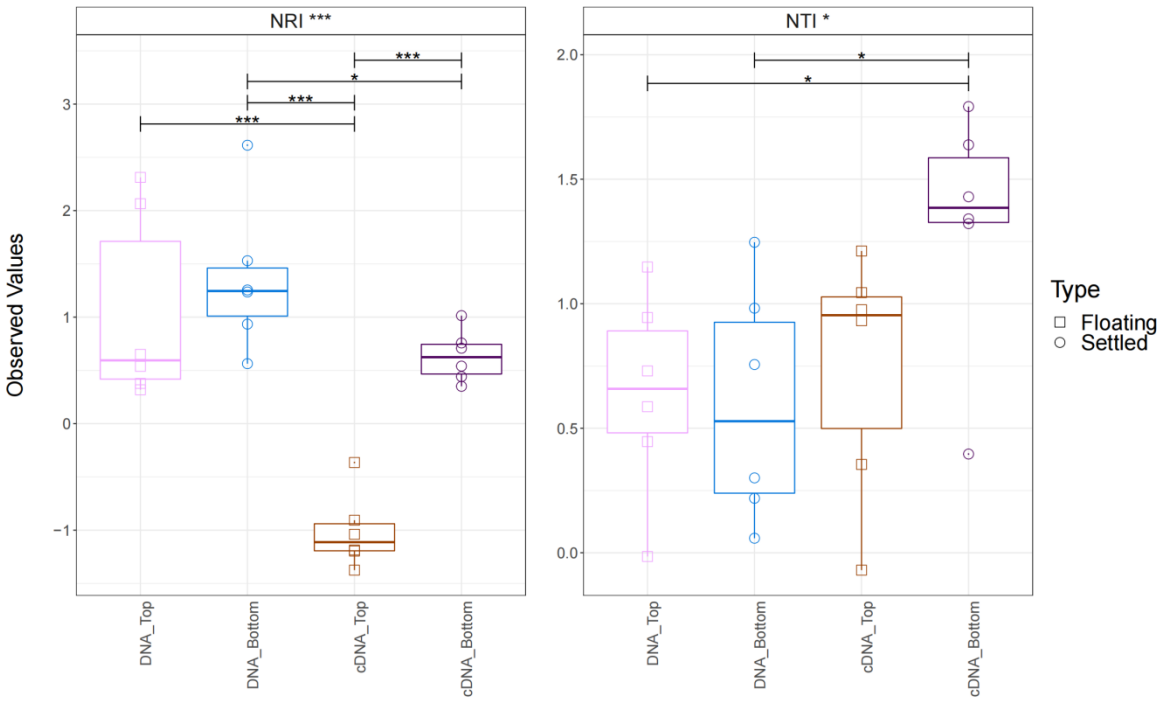
3.3.1 NRI and NTI

The observed values of NRI and NTI of the four groups of parameters in the abscissa are all & GT; 0, which means that all four types have strong environmental pressures. Environmental stress is the embodiment of determinacy in microbial community assembly. This also demonstrates the role of determinacy in microbial community development. At the same time, cDNA

is more stable and similar than DNA. The average NTI of DNA is slightly higher than that of cDNA, which represents a higher upper limit of development as well as diversity of DNA groups.

NRI/NTI can be used to distinguish between "environmental filtering" and "competitive exclusion". Because if NRI/ NTI< 0/-2 implies that microbial communities are subject to stochastic/competitive exclusion. If the NRI/NTI & gt; 0/2 means affected by determinacy in microbial community assembly. As can be seen from the data below, in the case of Weighted, the NRI/ NTI&LT of the microbial community; 0/-2, which confirms that stochastic/competitive exclusion is part of microbial community assembly. In the Unweighted data plot, NTI> 0/2, which indicates that the assembly of deterministic first microorganisms is also interfering. When looking at the DNA groups together, the upper and lower limits of DNA are higher, which means that their diversity is more significant, and in the image in the second column, it can be preliminarily concluded that the DNA group in the microbial community has slightly higher randomness than the cDNA.

(a)



(b)

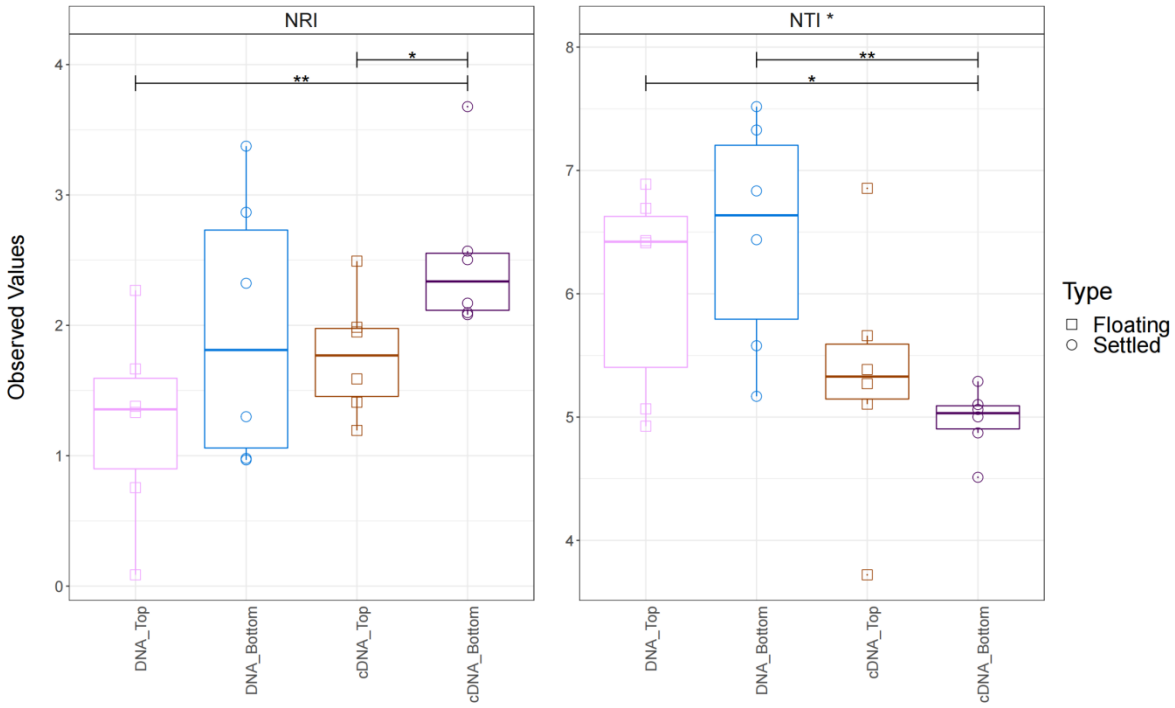


Figure 6

This is a boxplot of NRI/NTI data for DNA and cDNA.Figure6. (a) DNA and cdna data of The NRI/NTI Plot that Represent Unweighted Frequency.
(b) DNA and cdna data of The NRI/NTI Plot that Represent Unweighted IndependentSwap.

3.3.2 NST

NTI, which stands for Nearest-Taxon-index, is a standardized measure of the phylogenetic distance from each taxon in the sample to the nearest taxon, quantifying the degree of terminal clustering. NTI is the inverse of the output of "Ses.MNTD".

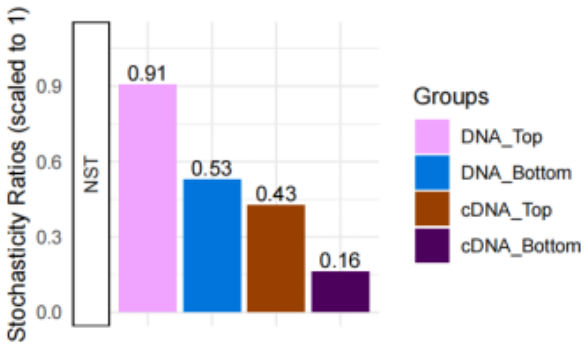
The mean NTI obtained in all communities was significantly different from zero, indicating the existence of average clustering (NTI> 0) or over-dispersion (NTI< 0) In the data analysis of NST, if NST & GT; 0.5 This means increasing value suggests stochasticisty; If NST A< 0.5A This means that the assembly of microbial communities is Deterministic.

Comparing the following sets of data, the NTI data is between plus and minus 2, which shows that the microbial community of the sample is not overly dispersed or overly clustered. Their values are greater than 0.5 and less than

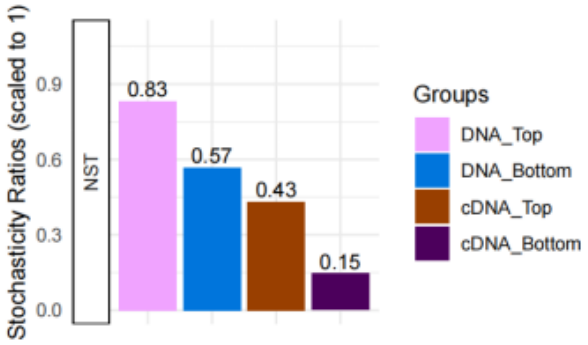
0.5 at the same time, which represents the joint action of certainty and randomness on microbial communities. Compared with the data, the DNA sample group is more deterministic, while the cDNA sample group is more random. This is not exactly the same as NRI and NTI, so more data are needed to prove this. In contrast, the NST analysis of microbial communities is not limited to the following two sets of pictures. When I refer to more data pictures in the appendix, if I normalize the data, the DNA group in microbial communities has a slightly higher randomness than the cDNA group.

(a)

Stochasticity-Ratios_PF_horn_1000_TRUE

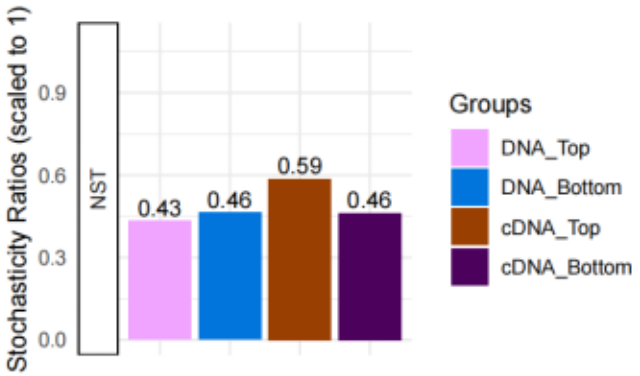


Stochasticity-Ratios_PF_horn_1000_FALSE

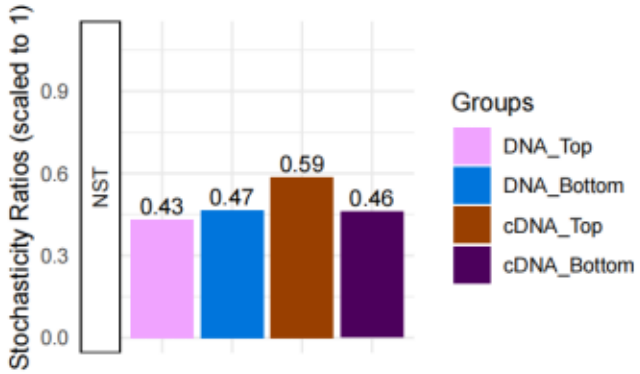


(b)

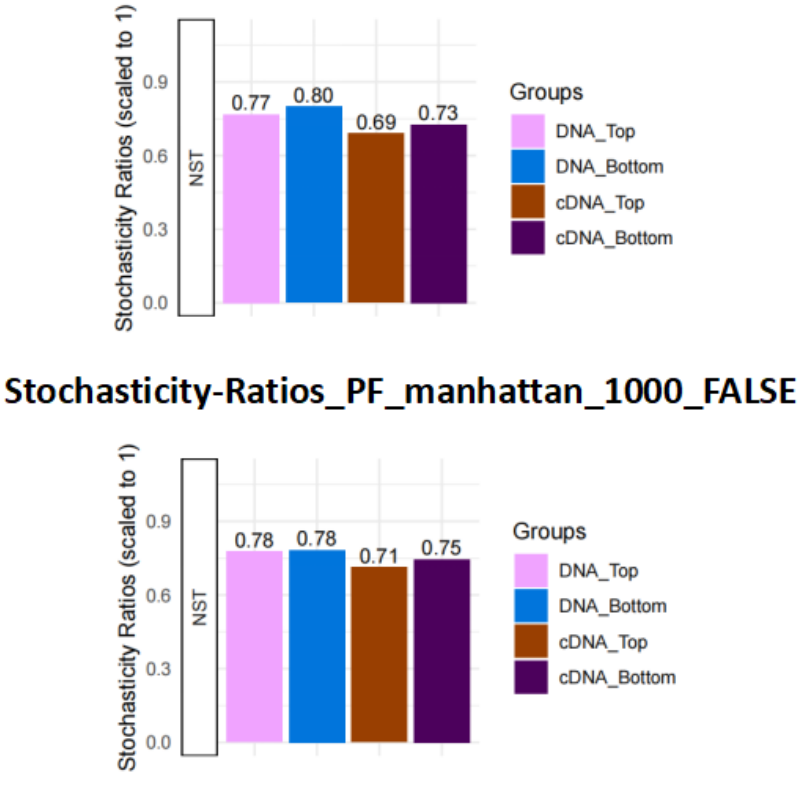
Stochasticity-Ratios_PF_binomial_1000_TRUE



Stochasticity-Ratios_PF_binomial_1000_FALSE



(c)



(d)

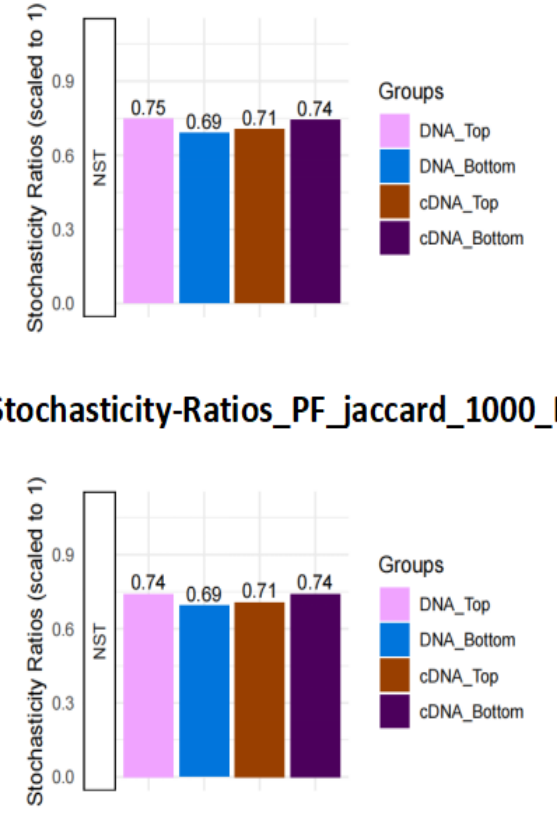


Figure 7. NST data

Histogram of NST data measured by DNA and cDNA under different basal conditions

Figure7. (a) The NST diagram represents The Horn caseStochasticity Ratios of data DNA&cDNA.(b)The NST diagram represents The Binmial caseStochasticity Ratios of data DNA&cDNA. (c) The NST diagram represents The case of ManhattanStochasticity Ratios of data DNA&cDNA. (d) The NST diagram represents The Jaccard case Stochasticity Ratios of data DNA&cDNA.

3.3.3 QPE

Without knowing the effects, I can look at microbial communities based on null Models (QPE) and derive whether randomness is present there (randomness) or whether communities are influenced by "biological factors" or "abiotic factors" (physicochemical parameters). The results of QPE can be divided into four parts. In the absence of any recorded environmental parameters such as PH, temperature, and calprotectin, selection pressure can show whether a microbial sample has potential characteristics of a single environment or multiple environments. Microbial community structure is influenced by the environment, and the results can illustrate what is the percentage of homogeneous selection and what is the percentage of variable selection. In the case of Variable Selection, all four master data are 0. Variable selection does not work, which can be explained by the same environmental conditions of the samples (same temperature, pH, salinity, etc.)(Evans et al., 2016). In the Undominated case, the four sets of data homogenized dispersal around 20 to 30 percent, DNA and cDNA_Top both remained at 6.67 percent, while cDNA_Bottom became 0. Similarly, in the condition of dispersal limitation, except that the values of DNA_Bottom and cDNA_Bottom do not change, the other two groups of data are changed by different magnitudes. This means that the distribution of the microbiota for anaerobic digestion in sewage treatment is limited by some specific factors. It is also proved that floating particles have higher randomness than settling particles. In designing the parameters as DNA and cDNA, the effect of randomness on DNA can be observed, while the effect of randomness on cDNA is not significant. In homogenizing select, the four main data rose by different magnitudes. The value of DNA_Top does not change much, relatively speaking, cDNA_Top and cDNA_Bottom, DNA_Bottom shows a large change, around 60 percent. Because Deterministic species of selection pressure, it means that Deterministic species dominate the development of DNA and cDNA communities.

Focusing on the second image, the beta-diversity of the four sets of data is less than -0.5. This suggests that the populations of both DNA and cDNA are deterministically clustered and similar. In short, although certainty and

randomness jointly dominate the assembly of microbial communities, certainty has a greater impact.

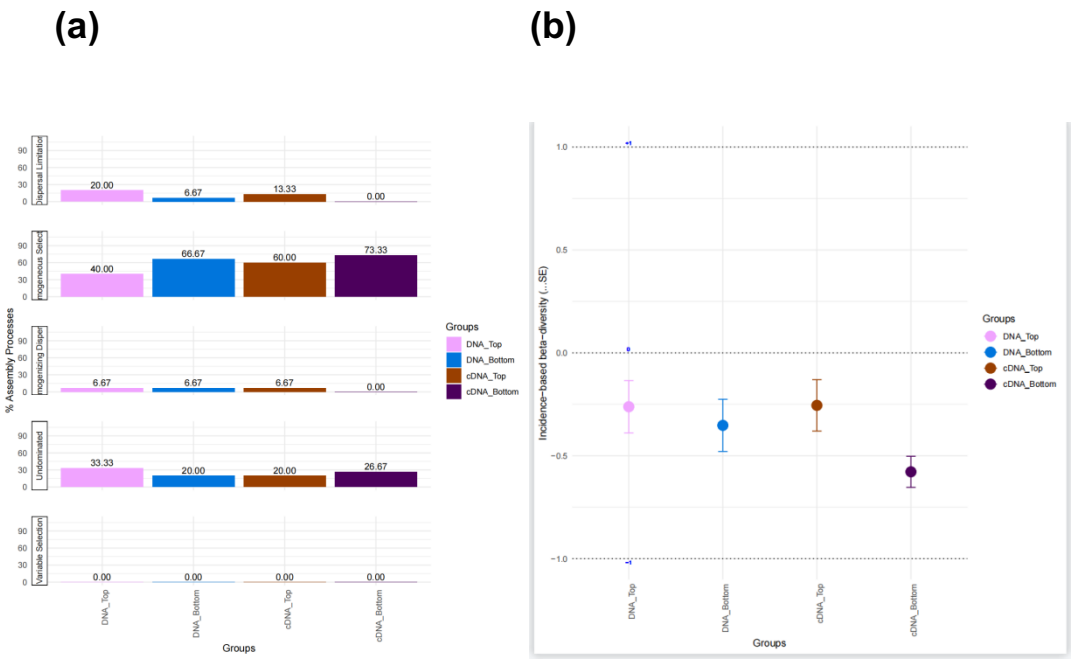


Figure 8. QPE data

Figure8. (a) The results of the QPE randomness (Evans et al., 2016) of microbial communities under different basal conditions. (b) The p results for DNA and cDNA.

3.4 Different analysis

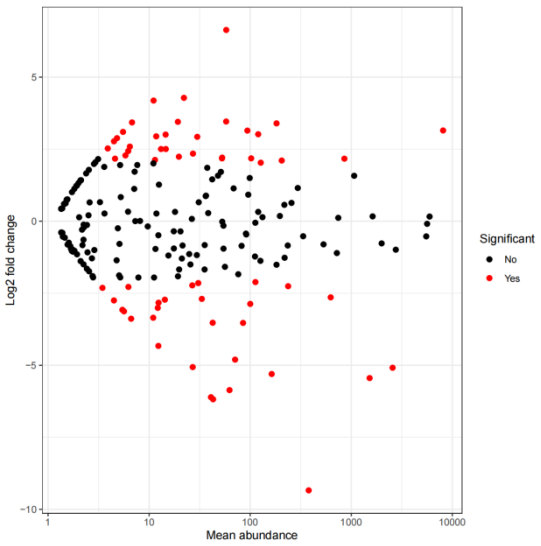
3.4.1 DeSeq2

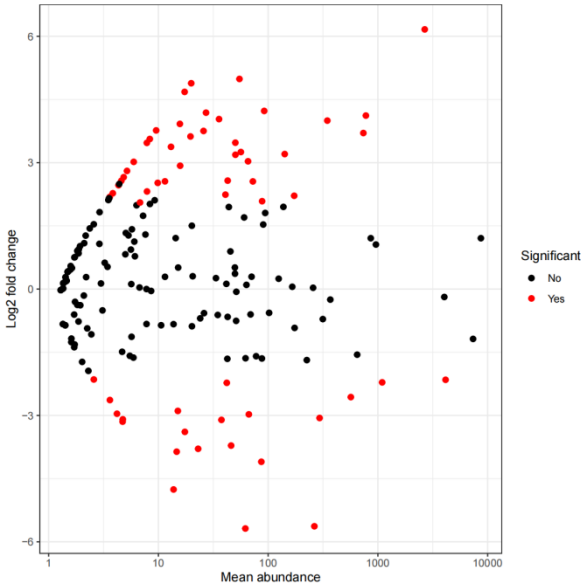
First, Figure 1 is analyzed. In the normalization analysis of DeSeq2, the four groups of data are close, and the degree of normalization is high, and the degree of dispersion is low. Moreover, the mean value of important data of log2 expression level in the ordinate is roughly less than 2 times, which means that the difference of microorganisms between _cDNA_Top_vs_cDNA_Bottom and _DNA_Top_vs_DNA_Bottom is not very significant. On the contrary, the microbial difference of _DNA_Bottom_vs_cDNA_Bottom was significant. The microbes of _DNA_Top_vs_cDNA_Top are slightly different.

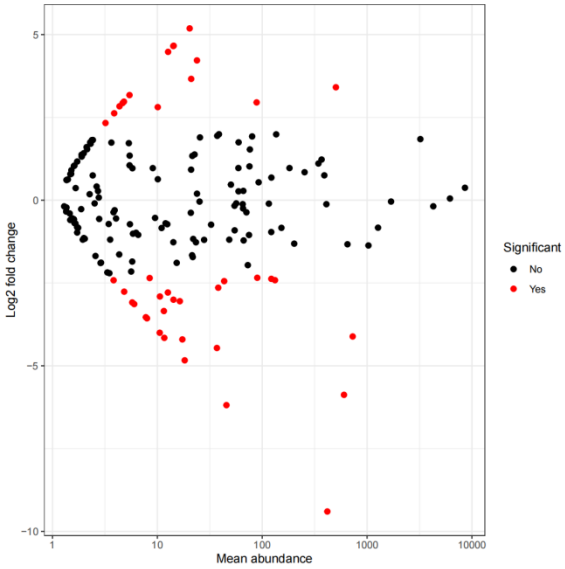
Microorganisms with differential expression characteristics:

By analyzing the latter two data graphs, it can be observed that the microorganisms with differential expression characteristics are Archaea; *Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosaetaceae; Methanosaeta PADj = 7.1078E − 48&archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanospirillaceae; Methanospirillum padj = 5.8918e−11.* This proves that although Archaea does not account for a large proportion of the microbial community, its differences are also expressed, and it is affected by the randomness of microbial community assembly.

In addition, Bacteria series are relatively stable, confirming that the diversity and richness of DNA and cDNA mentioned above are similar.And as can be seen in the data chart, *Bacteria; Firmicutes; Clostridia. Firmicutes;; Proteobacteria; Deltaproteobacteria; Gammaproteobacteria; Betaproteobacteriales; Pseudomonadales;* These types of microorganisms accounted for a higher proportion of the microbial community. They play a dominant role in anaerobic digestion.







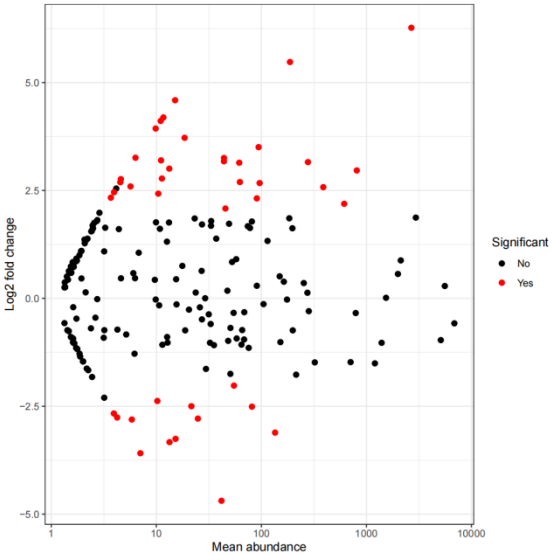


Figure 9. Volcanic maps showing the degree of population dispersion and aggregation of DNA and cDNA.

Figure10. Comparison of community species between DNA and cDNA

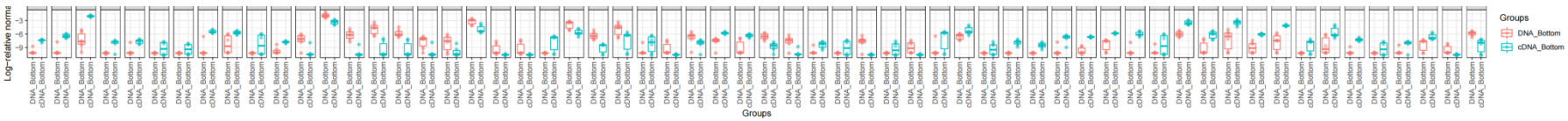


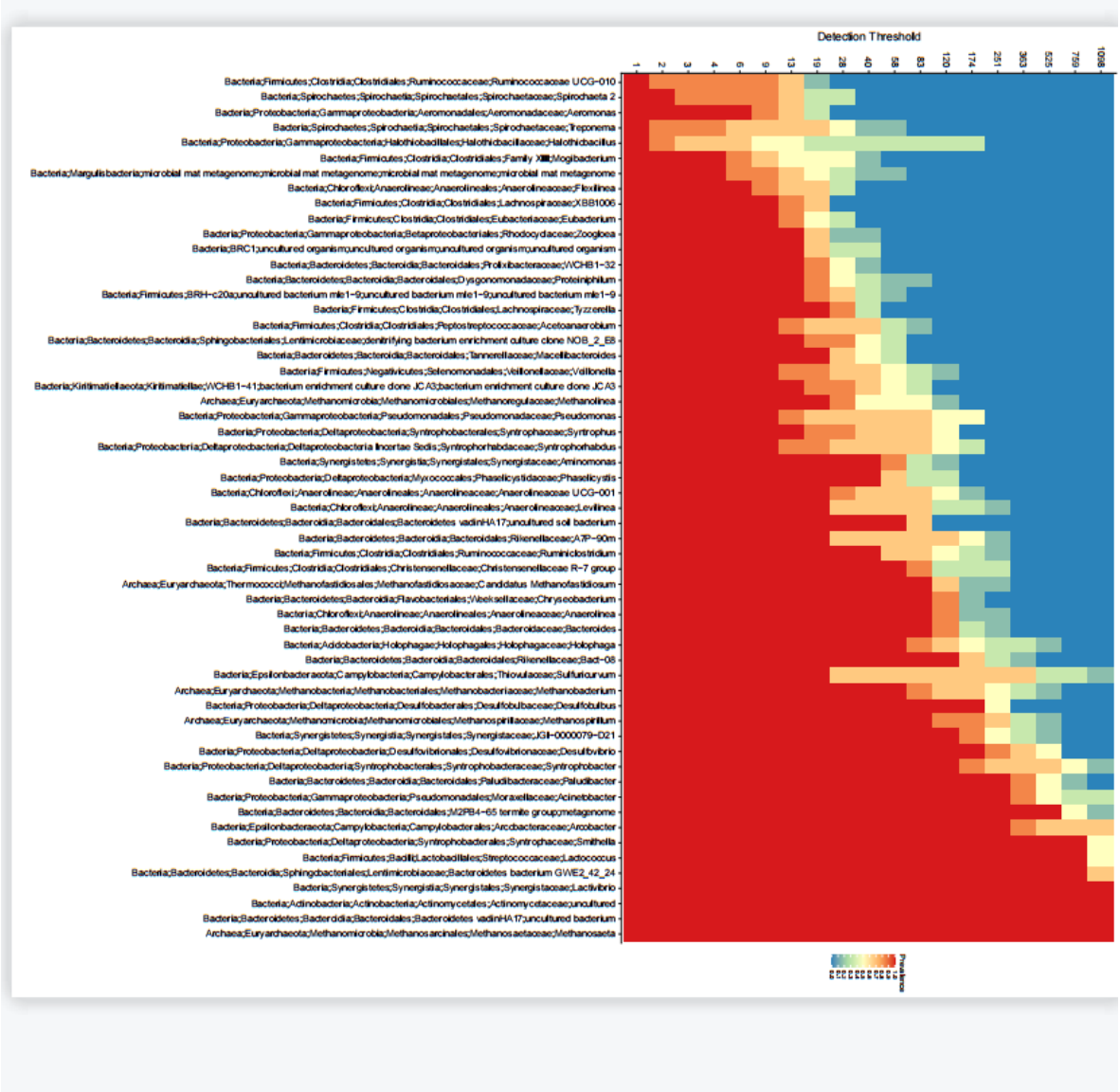
Figure11. Comparison of community species between DNA and cDNA

3.4.2 Core microbiome

The core microbiome is an important part of stability interpretation in complex communities, and the core is usually defined as a group of members shared between microbial communities from similar habitats. Omics approaches to assessing gene expression can enrich ecological insights into the core microbiome.

The core microbiome was analyzed, and the diversity and richness of the microbial community were reflected by different species of microbial populations in the figure. Comparing the two pictures, I can see that most of the biomes have not changed much. *Bacteria; Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminococcaceae UCG - 010 & Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2&Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema&Bacteria; Proteobacteria; Gammaproteobacteria; Halothiobacillales; Halothiobacillaceae; Halothiobacillus,* these groups were relatively variable. *Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales; Methanobacteriaceae; Methanobacterium* showed the most significant change. This is similar to the conclusion reached by DeSeq2. Archaea are microorganisms that express differential characteristics in microbial communities. Because most of the microbial Detection thresholds did not change much, the microbial communities in this study had high stability. This is one of the reasons why microbial community treatment of sewage is widely used. The change of Detection Threshold not only reflects which microorganisms express differential characteristics, but also shows that Archaea is likely to affect particle floating. This kind of requires further experiments to prove.

(a)


Detection Threshold
1
2
3
4
6
9
13
19
28
40
58
83
120
174
251
363
525
759
1098
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcaceae UCG-010
Bacteria;Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae;Spirochaeta 2
Bacteria;Proteobacteria;Gammaproteobacteria;Aeromonadales;Aeromonadaceae;Aeromonas
Bacteria;Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae;Treponema
Bacteria;Proteobacteria;Gammaproteobacteria;Halothiobacillales;Halothiobacillaceae;Halothiobacillus
Bacteria;Firmicutes;Clostridia;Clostridiales;Family XIII;Mogibacterium
Bacteria;Margulisbacteria;microbial mat metagenome;microbial mat metagenome;microbial mat metagenome;microbial mat metagenome
Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Flexilinea
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;XBB1006
Bacteria;Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Eubacterium
Bacteria;Proteobacteria;Gammaproteobacteria;Betaproteobacteriales;Rhodocyclaceae;Zoogloea
Bacteria;BRC1;uncultured organism;uncultured organism;uncultured organism;uncultured organism
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prolixibacteraceae;WCHB1-32
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Dysgonomonadaceae;Proteiniphilum
Bacteria;Firmicutes;BRH-c20a;uncultured bacterium mle1-9;uncultured bacterium mle1-9;uncultured bacterium mle1-9
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Tyzzerella
Bacteria;Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Acetoanaerobium
Bacteria;Bacteroidetes;Bacteroidia;Sphingobacteriales;Lentimicrobiaceae;denitrifying bacterium enrichment culture clone NOB_2_E8
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Tannerellaceae;Macellibacteroides
Bacteria;Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonella
Bacteria;Kiritimatiellaeota;Kiritimatiellae;WCHB1-41;bacterium enrichment culture clone JCA3;bacterium enrichment culture clone JCA3
Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanoregulaceae;Methanolinea
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas
Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae;Syntrophus
Bacteria;Proteobacteria;Deltaproteobacteria;Deltaproteobacteria Incertae Sedis;Syntrophorhabdaceae;Syntrophorhabdus
Bacteria;Synergistetes;Synergistia;Synergistales;Synergistaceae;Aminomonas
Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Phaselicystidaceae;Phaselicystis
Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Anaerolineaceae UCG-001
Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Levilinea
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidetes vadinHA17;uncultured soil bacterium
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;A7P-90m
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminiclostridium
Bacteria;Firmicutes;Clostridia;Clostridiales;Christensenellaceae;Christensenellaceae R-7 group
Archaea;Euryarchaeota;Thermococci;Methanofastidiosales;Methanofastidiosaceae;Candidatus Methanofastidiosum
Bacteria;Bacteroidetes;Bacteroidia;Flavobacteriales;Weeksellaceae;Chryseobacterium
Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Anaerolinea
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides
Bacteria;Acidobacteria;Holophagae;Holophagales;Holophagaceae;Holophaga
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Bact-08
Bacteria;Epsilonbacteraeota;Campylobacteria;Campylobacterales;Thiovulaceae;Sulfuricurvum
Archaea;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobacterium
Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfobulbus
Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanospirillaceae;Methanospirillum
Bacteria;Synergistetes;Synergistia;Synergistales;Synergistaceae;JGI-0000079-D21
Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio
Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae;Syntrophobacter
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Paludibacteraceae;Paludibacter
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Acinetobacter
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;M2PB4-65 termite group;metagenome
Bacteria;Epsilonbacteraeota;Campylobacteria;Campylobacterales;Arcobacteraceae;Arcobacter
Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae;Smithella
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Ladococcus
Bacteria;Bacteroidetes;Bacteroidia;Sphingobacteriales;Lentimicrobiaceae;Bacteroidetes bacterium GWE2_42_24
Bacteria;Synergistetes;Synergistia;Synergistales;Synergistaceae;Lactivibrio
Bacteria;Actinobacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;uncultured
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidetes vadinHA17;uncultured bacterium
Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosaetaceae;Methanosaeta
Prevalence

(b)

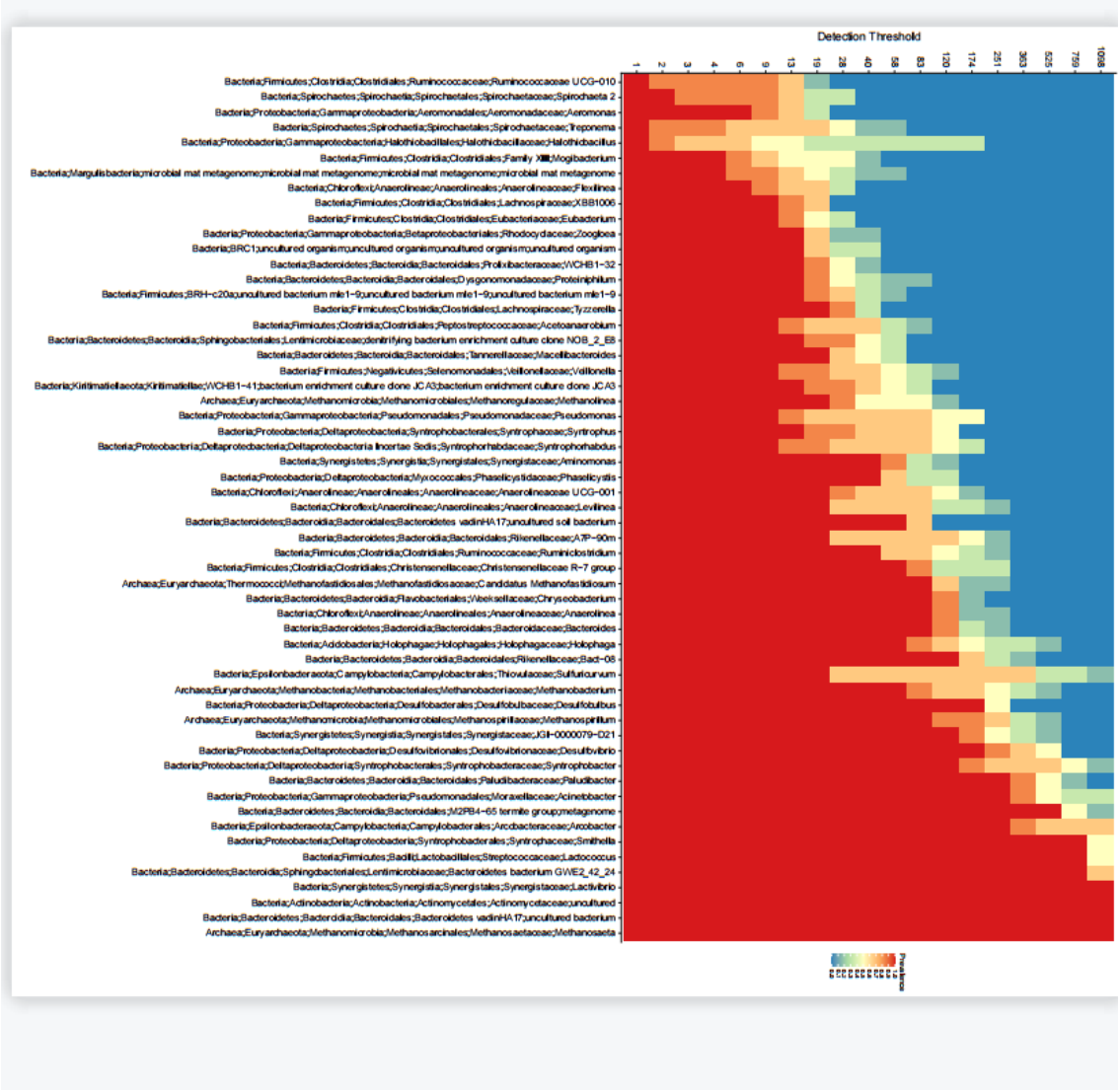


Figure12. Core community abundance data

Figure12. (a) image represents Core community abundance data of cDNA. Figure12. (b) image represents Core community abundance data of DNA. Note:The ASV at the top of the image shows low abundance and the bottom shows high abundance.

3.5 Core microbiome analysis

3.5.1 Subset regression

Richness was the best choice for analyzing the Richness of microbial communities. CDNA was the best model derived from subset analysis. The cross-validation error is 0.18323. This can indicate that the cDNA has a higher

richness. For microbial community diversity, DNA has higher diversity than cDNA. In the second figure, the cross-validation error for DNA is 0.33323. At the same time, it can be seen from the data that DNA has a positive correlation with Shannon. This suggests that DNA contributes to a higher diversity of microbial communities. Conclusions related to this conclusion have also been mentioned in the previous paper.

(a)

	Richness								
Predictors	*Estimates*	*std. Error*	*std. Beta*	*standardized std. Error*	*CI*	*standardized CI*	*Statistic*	*p*	*df*
(Intercept)	-1539.96835	1213.93584	-0.00000	0.17937	-4064.48613 – 984.54942	-0.37303 – 0.37303	-1.26857	2.185e-01	21.00000
Status cDNA	49.95300 *	19.42173	0.47127	0.18323	9.56330 – 90.34270	0.09022 – 0.85232	2.57202	**1.777e-02**	21.00000
Diameter	1430.50090	971.08652	0.26992	0.18323	-588.98408 – 3449.98587	-0.11113 – 0.65097	1.47309	1.556e-01	21.00000
Observations	24								
R^2 / R^2 adjusted	0.295 / 0.228								

** p<0.05 ** p<0.01 *** p<0.001*

(b)

	Shannon								
Predictors	*Estimates*	*std. Error*	*std. Beta*	*standardized std. Error*	*CI*	*standardized CI*	*Statistic*	*p*	*df*
(Intercept)	-6.01738 **	1.58872	0.00000	0.11152	-9.34261 – -2.69216	-0.23341 – 0.23341	-3.78757	**1.245e-03**	19.00000
Status R1	-0.06832	0.09717	-0.09249	0.13154	-0.27170 – 0.13505	-0.36781 – 0.18282	-0.70316	4.905e-01	19.00000
Status R2	0.24087 *	0.09717	0.32608	0.13154	0.03750 – 0.44425	0.05077 – 0.60140	2.47895	**2.273e-02**	19.00000
Status DNA	0.23207 **	0.07934	0.33323	0.11392	0.06602 – 0.39813	0.09480 – 0.57166	2.92518	**8.684e-03**	19.00000
Percent VS	1.22787 ***	0.19834	0.70523	0.11392	0.81273 – 1.64300	0.46679 – 0.94366	6.19065	**6.004e-06**	19.00000
Observations	24								
R^2 / R^2 adjusted	0.753 / 0.702								

** p<0.05 ** p<0.01 *** p<0.001*

Figure13. Subset data

Figure13(a). Subset Regression data table (data on microbial community abundance with Richness as a parameter)

Figure13(b). Subset Regression data table (data on microbial community abundance with Shannon as a parameter)

3.5.2 CODA GLMNET

In the biological monitoring, often need to analyze the relations between the different samples of community structure, the relationship between environmental factors and environmental factors and the causal relationship between community structure changes, a biological community is usually composed of several different species, environmental factors is more complex, multiple attribute of biological and environmental data is the monitoring data of multivariate nature determines the asked for this kind of multivariable In order

to correctly sort out these intricate relationships, it is necessary to use multivariate analysis methods.

As can be seen from the image, Bacteria; Bacteroidetes; Bacteroidia; Sphingobacteriales; Lentimicrobiaceae; Bacteroidetes bacterium GWE2_42_24&Bacteria; Actinobacteria; Actinobacteria; Propionibacteriales; Propionibacteriaceae; The multivariate abundance data for Luteococcus showed the greatest variation. Bacteroidetes had positive effects, while Actinobacteria had negative effects. Both Bacteroidetes and Actinobacteria belong to the bacterial category, while the archaea mentioned above do not belong to the bacterial category. This means that the abundance of the microbial community is more influenced by bacteria. This can also be seen in the images -- Archaea; Euryarchaeota; Thermococci; Methanofastidiosales; Methanofastidiosaceae; Candidatus Methanofastidiosum had a value of just 0.01.

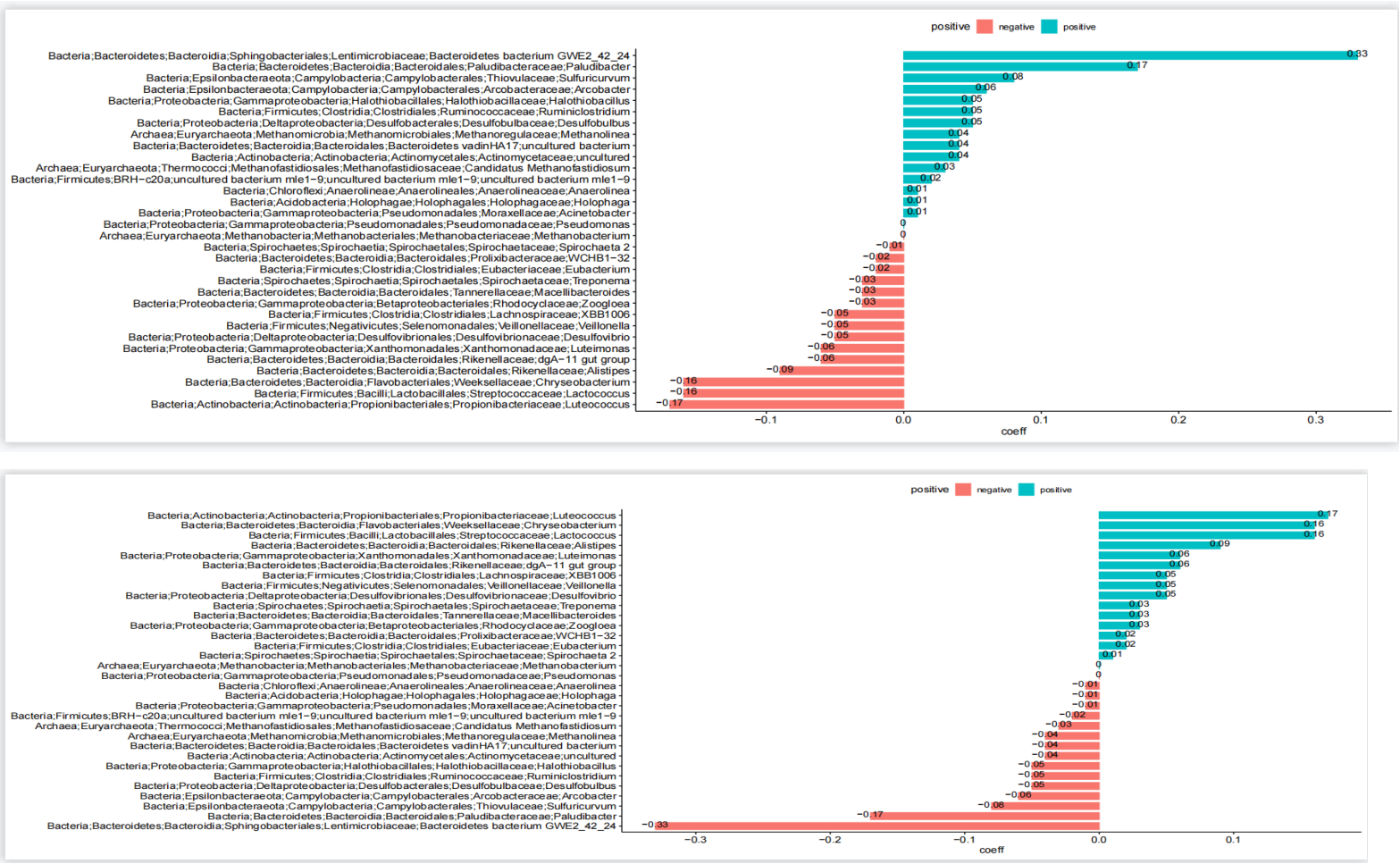


Figure14. Data plot of multivariate abundance data (Each column of the graph represents a biome. The red part is negative, and the green part is positive.)

4 Discussion

4.1 Selection mechanism of particle flotation

Particle flotation is a topic of this paper. When excessive surfactant and grease compounds enter the normal sewage into the aeration tank operation, the partial degradation of the material by specific surfactants forms foam, and makes the foam grow rapidly. These foams are generally white and light, and disappear when the activated sludge reaches maturity. When there is an excess of surfactant in sewage, such substances can affect the stability and permeability of the cytoplasmic membrane, resulting in the loss of some essential components of the cell, resulting in microbial growth arrest and death. During aeration, a large number of bubbles (bubbles) are generated(Theobald, 2014), which are easily attached to the micelle, so that the specific gravity of activated sludge is reduced and floating up. In addition, when the amount of oil in the water is too high, after aeration and mixing, the oil will adhere to the surface of the micelle, resulting in hypoxia death, resulting in the reduction of the specific gravity and floating.

Too high or too low pH impact will affect the catalytic effect of extracellular enzymes and enzymes existing in the cytoplasm and cell wall of activated sludge microorganisms, as well as the absorption of nutrients by microorganisms. When the continuous flow aeration reaction tank pH< 4.0 or pH> At 11.0, in most cases, the microbial activity in the activated sludge was inhibited, or lost activity, or even died, resulting in sludge upfloating("Factors Affecting Performance of Sewage Treatment Plant - UTPedia", 2017).

Influence of water temperature and salt content The suitable temperature range for microorganisms composing activated sludge is generally 15-35° C. When it exceeds 45C, most microorganisms in activated sludge will die and float up (except for long-term acclimated or special microorganisms). The adjustment of pH value of influencer cannot eliminate the influence of alkalinity on activated sludge. Adjusting the pH value of alkaline influent neutralizes alkaline substances, but produces salt. Osmotic pressure varies with salt solution

concentration, which is one of the important factors affecting microbial survival. If the osmotic pressure of the solution in which the microorganism is exposed changes, the cell will die(Laxmi Shrivas, 2014).

Toxic substrates Toxic to aerobic activated sludge microorganisms mainly include: high content of COD, substances (phenols and derivatives, alcohols, aldehydes and some acids, etc.), sulfide, heavy metals and halides. High substrate concentrations can form stable compounds with cellular enzyme activity centers, resulting in matrix inaccessibility, inability to be degraded, and even cell death by poisoning. After heavy metal ions enter human cells, they mainly bind to -SH groups on enzymes or proteins to inactivate or denature them. Trace heavy metal ions can also accumulate in the cell and eventually have toxic effects on microorganisms (micro action). Iodine and chlorine are common halides. Iodine irreversibly binds to the tyrosine of the protein (or enzyme) of the bacteria to generate diiodotyrosine and inactivate the bacteria. Chlorine and water synthesize secondary acid, its decomposition produces strong oxidant. Moreover, the mutation of the material in the waste water makes the microorganism that has been domesticated and can degrade the poison decrease or disappear.

4.2 The advantages and disadvantages

4.2.1 Main advantages

Energy consumption is greatly reduced, and biogas can be recycled;

Sludge production is very low(Molinuevo-Salces et al., 2019);

-- The proliferation rate of anaerobic microorganisms is much lower than that of aerobic microorganisms. The yield Y of acid-producing bacteria is 0.15~0.34kgVSS/kgCOD, that of methanogenic bacteria is about 0.03kgVSS/kgCOD, and that of aerobic microorganisms is about 0.25~0.6kgVSS/kgCOD.

Anaerobic microorganisms may degrade or partially degrade some organic matter which cannot be degraded by aerobic microorganisms.

The reaction process is complex -- anaerobic digestion is a continuous microbial process in which a variety of microorganisms with different properties and functions work together.

4.2.2 Main disadvantages

Sensitive to temperature, pH and other environmental factors;
The water quality of the treated effluent is poor and needs to be further treated by aerobic method.
Large smell;
The removal effect of ammonia nitrogen is not good;

4.3 Community assembly of microbial communities during anaerobic digestion in sewage

The anaerobic digestion of sludge mainly involves hydrolytic acidifying bacteria and methanogenic archaea. Hydrolytic acidifying bacteria play an important role in the process of anaerobic digestion of sludge. Because hydrolytic bacteria can convert the carbohydrates, proteins and lipids in the sludge into simple soluble monomer substances, acidifying bacteria can further convert the hydrolytic products into acidic products (volatile fatty acids), thus providing a carbon source for microbial growth. Obviously, bacteria not only play a key role in the hydrolysis and acidification of organic matter in sludge, but also affect the efficiency of anaerobic digestion. Previous studies have found that archaea account for about 10% of the total microbial population in anaerobic systems . It can be seen that bacteria account for a higher proportion of total microorganisms than archaea, so changes in bacterial community structure will affect archaea community . According to Ahring, anaerobic microorganisms cover at least 20 phyla of bacteria, It includes Proteobacteria, *Firmicutes, Chloroflexi, Spirochaetes, Bacteroidetes, Actinobacteria,* etc. Among the above bacteria,It can be seen that the changes of Chloroflexi, Proteobacteria, Bacteroidetes and Firmicutes communities will have an impact on the anaerobic digestion system, because they are the main phyla in the process of anaerobic digestion.
Bacteroidetes have a degradation effect on complex carbon compounds , such as cellulose and hemicellulose. Previous studies have found that Bacteroidetes can convert cellulose into monosaccharides (mainly glucose) and organic acids , and hemicellulose into D-xylan and glucose(Berman, 2019) . In addition, Bacteroidetes will further decompose glucose after converting hemicellulose

into glucose, because Bacteroidetes are the main bacteria of deglycating sugars . Therefore, the distribution of Bacteroidetes is related to the concentration of volatile fatty acids (VFAs). Obviously, changes in the abundance of Bacteroidetes during anaerobic digestion will lead to changes in the concentration of VFAs in the system, which will have an impact on the pH of the system. During anaerobic digestion, bacteria are involved in both the hydrolysis and acidification stages, while the methanogenesis stage is completely involved by archaea.

Methanogens are mainly divided into three types according to methanogenic pathways, namely acetic acid, hydrogen and methyl methanogens. Most of the methane is mainly produced by acetic acid and hydrogen methanogens . Previous studies have found that *Methanobacterium, Methanosarcina, Methanobrevibacter, Methanosaeta and Methanomicrobium* are the main methanogenic microorganisms . In addition to the above bacteria, *Methanobacteriaceae, Methanospirillaceae, Methanomicrobiaceae, Methanothermobacter, Methanospirllum, Methanoculleus* and Methanomassilliicoccus. Among them, Methanosarcina and Methanosaeta belong to acetic acid *methanogens. Methanosarcina* is a multifunctional methanogens, which has the ability of acetic acid methanogens, hydrogen methanogens and methyl methanogens. Therefore, Methanosarcina can produce methane by using acetic acid, methanol, methylamine, dimethylamine and H2/CO2(De Vrieze et al., 2012) . Methanosarcina is also the only strain that can produce methane through the above three pathways . At the same time, Methanosarcina has a high tolerance to volatile fatty acids (VFAs) and organic carrying rate (OLR) . In the anaerobic environment, the abundance of Methanosarcina and Methanosaeta changed with the change of acetic acid concentration. Previous studies have shown that high concentration acetic acid is suitable for the growth of Methanosarcina, while low concentration ethyl acid is suitable for the growth of Methanosaeta . Recent studies also found that Methanosarcina was the main strain when the acetic acid concentration was 250-500mg COD/L . Except for *Methanosarcina and Methanosaeta,* which belong to the acetic acid type of methanane-producing bacteria, among the detected *Methanosarcina, Methanothermobacter, Methanoculleus*, *Methanospirllum, Methanobacterium, Methanobacteriaceae,*

*Methanospirillaceae, Methanomicro Biaceae and Methanobrevibacter ar*e hydrogen-type methanogens, which mainly use H2/CO2 or formic acid to generate CH4 gas(Liu et al., 2011) . Methanomassiliicoccus is a methyl methanogenic bacterium, which mainly uses methanol (or methylamine) and hydrogen as electron donors to produce CH4 gas bodies . It can be seen that the dominant archaea in anaerobic digestion reactors are mainly acetic acid and hydrogen methanogens, and their common characteristic is that they can use the decomposition products of bacteria (acetic acid, hydrogen, etc.) to produce methane.

4.4 Influencing factors of anaerobic biological treatment

Dissolved oxygen (DO) : about 1~2mg/ L;
Water temperature is one of the important factors. In a certain range, with the increase of temperature, the biochemical reaction rate is accelerated, and the proliferation rate is also accelerated. Cell components, such as proteins and nucleic acids, are sensitive to temperature. When the temperature rises or falls sharply and exceeds a certain limit("envis.org - Factors Affecting Anaerobic Digestion", n.d.), there will be irreversible damage. > 40°C or

nutrients: C, H, O, N nutrients for about 90~97% of the cell composition; The other 3~10% are inorganic elements, mainly P; Domestic sewage generally does not need to add nutrients; Some industrial wastewater needs, generally for aerobic biological treatment process, should be BOD: N: P = 100:5:1 add N and P; Other inorganic nutrient elements: K, Mg, Ca, S, Na, etc. Trace elements: Fe, Cu, Mn, Mo, Si, boron, etc.PH value: The optimal pH of general aerobic microorganisms is between 6.5 and 8.5; pH.Toxic substances (suppressors) : heavy metals; Cyanide; H2S. Halogen elements and their compounds; Phenol, alcohol, aldehyde, etc.; Organic load rate: organic matter in sewage is originally food for microorganisms, but too much, it will not be conducive to microorganisms("What are the influencing factors of anaerobic biological treatment?" , n.d.).

5 Conclusions

The main research content of this report is the assembly of microbial communities in sewage treatment. If the model is ideal, the granular sludge formed by the aggregation of microbial communities will not float, and therefore will not produce floating particles. However, no matter in the experiment or in the actual sewage treatment process, particle floating is a very normal phenomenon. The reported analysis process is divided into four main parts, the first of which is the analysis of the assembly of the microbial community. The diversity and richness of the microbial community. The second part is the zero-model analysis of the microbial community. The analysis of the null model is also focused on the microbial community. As mentioned above, determinism and randomness dominate the main factors of microbial community assembly. At the same time, by determining the influence of certainty and randomness on microbial communities, the diversity and richness of microbial communities in DNA and cDNA groupings and floating and settling groupings can also be deduced in reverse. Two groupings, DNA and cDNA, represent intact and active microbial communities, respectively. This means that after I draw my conclusions, it is clear whether the microbial community doing the anaerobic digestion work is the main factor affecting the microbial community assembly. The third part is the analysis of microbial community differences. For this study, the differences in microbial communities and the expression of special differences suggest that the microbiome is likely to be one of the internal reasons that lead to particle floating and restrict the diversity and richness of microbial communities. The fourth section is the analysis of the core microbial community. The final purpose of this part is similar to that of the first part, which is to explore the influencing factors of microbial community assembly. But this part is more specific to what type of bacteria or archaea, etc. This analysis model can more clearly determine which microbial species affect the diversity and richness of the microbial community. Apart from this, the analysis in Part IV is also linked to certainty and randomness. Determinacy and randomness are themselves part of the development of microbial communities, and one of the aims of this study is to find out which of the two elements has the greater

impact on microbial community assembly. At present, deterministic influences on microbial community assembly are more prominent.

5.1 Floating particles

The extrinsic causes of particle floating have been described in the discussion. In the conclusion, I mainly summarized several internal causes of particle floating. Based on the previous data, Archaea is likely to affect particle floating. For microbial communities, cdnas (active microbiome) may play a larger role in particle floating. At the same time, the diffusion process plays a stronger role in floating particles(Evans et al., 2016).

5.2 Assembly of microbial communities

The assembly of microbial communities will be analyzed in the two aspects mentioned above.

The first part is diversity and richness. From the results in the previous four sections, it can be concluded that the total DNA community) has a higher diversity than the cDNA(active community), which has a higher richness. Specifically for the microbiome, Bacteroidetes had a positive effect on the richness of the microbial community, while Actinobacteria had a negative effect.

The second part is certainty and randomness. From the above, it can be concluded that randomness dominates DNA(total community) grouping in microbial community, and determinism dominates cDNA(active community) grouping in microbial community. When it comes to microorganisms specifically, Archaea is affected by the randomness of microbial community assembly.

Reference

[1] 10.3 - Best Subsets Regression, Adjusted R-Sq, Mallows Cp | STAT 501. PennState: Statistics Online Courses. Retrieved 18 August 2022, from https://online.stat.psu.edu/stat501/lesson/10/10.3.

[2] Aguirre de Cárcer, D. (2019). A conceptual framework for the phylogenetically constrained assembly of microbial communities. Microbiome, 7(1). https://doi.org/10.1186/s40168-019-0754-y

[3] Retrieved 18 August 2022, from https://mindspore.cn/mindquantum/docs/en/r0.6/quantum_phase_estimation.html.

[4] Berman, J. (2019). Bacteria. Taxonomic Guide To Infectious Diseases, 39-119. https://doi.org/10.1016/b978-0-12-817576-7.00003-1

[5] De Vrieze, J., Hennebel, T., Boon, N., & Verstraete, W. (2012). Methanosarcina: The rediscovered methanogen for heavy duty biomethanation. Bioresource Technology, 112, 1-9. https://doi.org/10.1016/j.biortech.2012.02.079

[6] Dempster, A. (1971). An overview of multivariate data analysis. Journal Of Multivariate Analysis, 1(3), 316-346. https://doi.org/10.1016/0047-259x(71)90006-6

[7] envis.org - Factors Affecting Anaerobic Digestion. Envis.org. Retrieved 18 August 2022, from https://envis.org/technology/wastewater-treatment/100-factors-affecting-anaerobic-digestion.

[8] Factors Affecting Performance of Sewage Treatment Plant - UTPedia. Utpedia.utp.edu.my. (2017). Retrieved 18 August 2022, from http://utpedia.utp.edu.my/7250/#:~:text=Although%20there%20are%20many%20other%20factors%20that%20can,national%20company%20is%20identified%20as%20my%20case%20study.

[9] Hu, W., Zhang, Q., Tian, T., Li, D., Cheng, G., & Mu, J. et al. (2015). Relative Roles of Deterministic and Stochastic Processes in Driving the Vertical Distribution of Bacterial Communities in a Permafrost Core from the Qinghai-Tibet Plateau, China. PLOS ONE, 10(12), e0145747. https://doi.org/10.1371/journal.pone.0145747

[10] hypothesis?, W., Du, H., halvorsen, k., Ball, M., & Chernick, M. (2017). What is a null model in regression and how does it relate to the null hypothesis?. Cross Validated. Retrieved 18 August 2022, from https://stats.stackexchange.com/questions/259636/what-is-a-null-model-in-regression-and-how-does-it-relate-to-the-null-hypothesis.

[11] Laxmi Shrivas, V. (2014). Retrieved 18 August 2022, from https://www.slideshare.net/vijalshrivas/microbiology-of-waste-water-treatment.

[12] Lettinga, G., Pol, L., Van Lier, J., & Zeeman, G. (2008). Possibilities and Potential of Anaerobic Wastewater Treatment Using Anaerobic Sludge Bed (ASB) Reactors. Biotechnology, 517-526. https://doi.org/10.1002/9783527620944.ch25

[13] Li, H., Xu, Z., Yan, Q., Yang, S., Van Nostrand, J., & Wang, Z. et al. (2017). Soil microbial beta-diversity is linked with compositional variation in aboveground plant biomass in a semi-arid grassland. Plant And Soil, 423(1-2), 465-480. https://doi.org/10.1007/s11104-017-3524-2

[14]Li, J., Zhu, J., Zhang, H., Peng, Q., Li, X., Zhang, D., & Zhao, W. (2021). Vortex formation mechanism and experimental research of abrasive flow precision machining special inner Curved Surface Based on Large Eddy Simulation. The International Journal Of Advanced Manufacturing Technology, 116(5-6), The 1633-1651. https://doi.org/10.1007/s00170-021-07537-x

[15] Liu, X., Yan, Z., & Yue, Z. (2011). Biogas. Comprehensive Biotechnology, 99-114. https://doi.org/10.1016/b978-0-08-088504-9.00165-3

[16] Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12). https://doi.org/10.1186/s13059-014-0550-8

[17] MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS. (2022), 75(1), 1-13. https://doi.org/10.1128/mmbr.00038-10

[18] Molinuevo-Salces, B., Riaño, B., Hernández, D., & Cruz García-González, M. (2019). Microalgae and Wastewater Treatment: Advantages and Disadvantages. Microalgae Biotechnology For Development Of Biofuel And Wastewater Treatment, 505-533. https://doi.org/10.1007/978-981-13-2264-8_20

[19] NTI.p function - RDocumentation. Rdocumentation.org. Retrieved 18 August 2022, from https://www.rdocumentation.org/packages/iCAMP/versions/1.3.4/topics/NTI.p.

[20] Panosyan, H., Bhattacharjee, K., Manchanda, G., & Pratap Singh, R. (2021). SPRINGER-NATURE (pp. 3-6). Springer.

[21] Questions about Quantum Phase Estimation. Quantum Computing Stack Exchange. (2022). Retrieved 18 August 2022, from https://quantumcomputing.stackexchange.com/questions/27567/questions-about-quantum-phase-estimation.

[22] Reeburgh, W. (1980). Anaerobic methane oxidation: Rate depth distributions in Skan Bay sediments. Earth And Planetary Science Letters, 47(3), 345-352. https://doi.org/10.1016/0012-821x(80)90021-7

[23] Shulakov, A., Egorov, A., & Onipchenko, V. (2017). A phylogenetic analysis of plant communities of the Teberda State Biosphere Reserve. Biology Bulletin Reviews, 7(3), 238-245. https://doi.org/10.1134/s2079086417030057

[24] Stegen, J., Lin, X., Konopka, A., & Fredrickson, J. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. The ISME Journal, 6(9), 1653-1664. https://doi.org/10.1038/ismej.2012.22

[25] Tay, K., Qian, J., & Tay, K. (2021). coda4microbiome for cross-sectional studies. Malucalle.github.io. Retrieved 18 August 2022, from https://malucalle.github.io/statistical-pills/coda4microbiome_cross_sectional.html.

[26] The Use and Types of Alpha-Diversity Metrics in Microbial NGS - CD Genomics. Cd-genomics.com. Retrieved 18 August 2022, from https://www.cd-genomics.com/microbioseq/the-use-and-types-of-alpha-diversity-metrics-in-microbial-ngs.html.

[27] Theobald, D. (2014). Control of activated sludge, including troubleshooting. Processingmagazine.com. Retrieved 18 August 2022, from https://www.processingmagazine.com/home/article/15583654/control-of-activated-sludge-including-troubleshooting.

[28] Uemura, S., & Harada, H. (1993). Microbial characteristics of methanogenic sludge consortia developed in thermophilic U ASB reactors. Applied Microbiology And Biotechnology, 39(4-5), 654-660. https://doi.org/10.1007/bf00205070

[29] Wang, A., Shi, K., Ning, D., Cheng, H., Wang, H., & Liu, W. et al. (2021). Electrical selection for planktonic sludge microbial community function

and assembly. Water Research, 206, 117744. https://doi.org/10.1016/j.watres.2021.117744

[30] Wang, L., Han, M., Li, X., Yu, B., Wang, H., & Ginawi, A. et al. (2021). Mechanisms of niche-neutrality balancing can drive the assembling of microbial community. Molecular Ecology, 30(6), 1492-1504. https://doi.org/10.1111/mec.15825

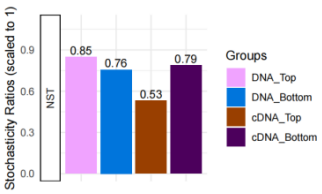
[31] What are the influencing factors of anaerobic biological treatment ? . Kashionbiotech.com. Retrieved 18 August 2022, from https://www.kashionbiotech.com/new-84-408-n.html.

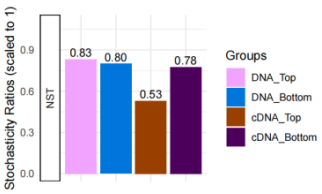
[32]Winter, F., Agarwal, R., Hrdlicka, J., & Varjani, S. CO2 Separation, purification and conversion to chemicals and fuels.

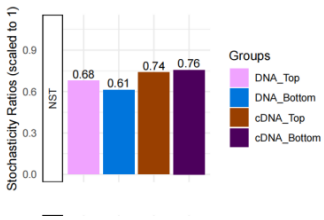
[33] Yi, M. A Complete Guide to Heatmaps. Chartio. Retrieved 18 August 2022, from https://chartio.com/learn/charts/heatmap-complete-guide/.

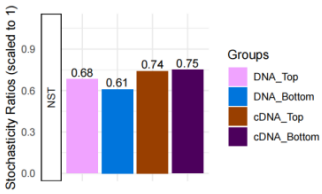
[34] Zhang, B., Ning, D., Van Nostrand, J., Sun, C., Yang, Y., Zhou, J., & Wen, X. (2020). Biogeography and Assembly of Microbial Communities in Wastewater Treatment Plants in China. Environmental Science &Amp; Technology, 54(9), 5884-5892. https://doi.org/10.1021/acs.est.9b07950

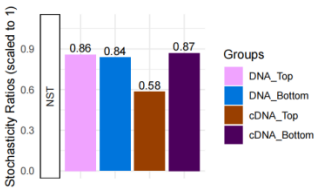
Appendix I

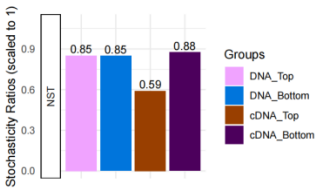

Stochasticity Ratios (scaled to 1)
NST
0.85
0.76
0.53
0.79
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

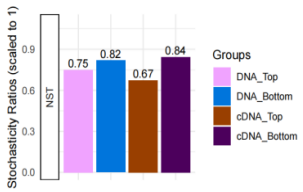

Stochasticity Ratios (scaled to 1)
NST
0.83
0.80
0.53
0.78
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

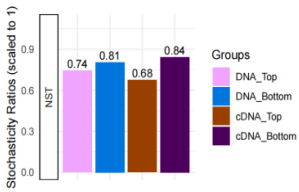

Stochasticity Ratios (scaled to 1)
NST
0.68
0.61
0.74
0.76
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

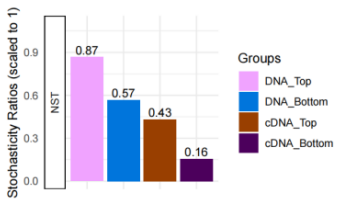

Stochasticity Ratios (scaled to 1)
NST
0.68
0.61
0.74
0.75
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

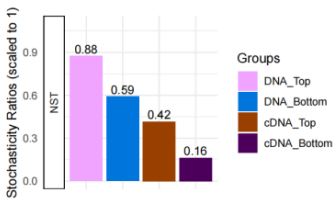

Stochasticity Ratios (scaled to 1)
NST
0.86
0.84
0.58
0.87
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

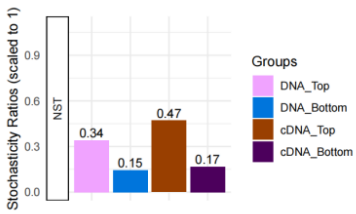

Stochasticity Ratios (scaled to 1)
NST
0.85
0.85
0.59
0.88
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

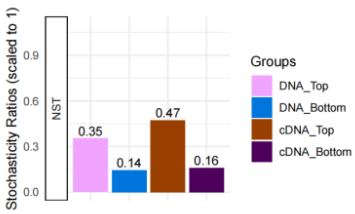

Stochasticity Ratios (scaled to 1)
NST
0.75
0.82
0.67
0.84
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom


Stochasticity Ratios (scaled to 1)
NST
0.74
0.81
0.68
0.84
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom


Stochasticity Ratios (scaled to 1)
NST
0.87
0.57
0.43
0.16
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom


Stochasticity Ratios (scaled to 1)
NST
0.88
0.59
0.42
0.16
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom


Stochasticity Ratios (scaled to 1)
NST
0.34
0.15
0.47
0.17
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom


Stochasticity Ratios (scaled to 1)
NST
0.35
0.14
0.47
0.16
Groups
DNA_Top
DNA_Bottom
cDNA_Top
cDNA_Bottom

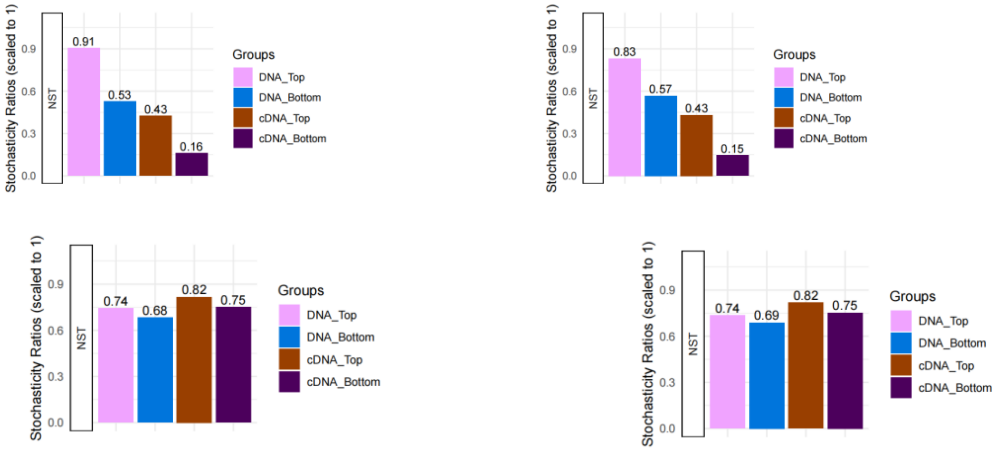


Figure 15 NST-result