

Meta-Analysis of Microbial Communities Reveals The Insights into the Ecological Roles of Microbiota in Wastewater Treatment Plants

P.W. Vidanage

Department of Chemical and Process
Engineering
University of Moratuwa
Moratuwa, Sri Lanka
poornaw@uom.lk

K.P.H.P.P. Kaluthanthri

Department of Chemical and Process
Engineering
University of Moratuwa
Moratuwa, Sri Lanka
hivinkaluthanthri@gmail.com

W.C. Lahiru

Department of Chemical and Process
Engineering
University of Moratuwa
Moratuwa, Sri Lanka
wclahiru100@gmail.com

Abstract—The wastewater treatment process is a complex undertaking with the primary objective of eliminating contaminants and rendering the water suitable for discharge into natural water bodies or for alternative purposes. It holds significant importance in safeguarding public health and preserving the environment. Within this process, microorganisms including bacteria, fungi, and protozoa assume a pivotal role by de-composing organic pollutants and transforming them into harmless by-products. Bioreactors offer an optimal environment for the proliferation and activity of these microorganisms, which are cultivated in large quantities to effectively consume organic substances in the wastewater, thereby reducing pollutant levels. In the secondary treatment stage, microorganisms are also employed to eliminate any remaining pollutants further. Biological wastewater treatment stands as one of the most crucial applications of biotechnology. Wastewater treatment facilities play a vital role in preventing the contamination of natural ecosystems. Traditional mathematical models used in the de-sign of wastewater treatment plants (WWTPs) face limitations in accurately capturing the complex metabolic activities within bacterial cells. However, advancements in sequencing technologies and computation-al capabilities have enabled us to explore the taxonomic and function-al identification of microorganisms present in WWTPs. Furthermore, adopting a systems biology approach offers a comprehensive under-standing of the ecological interactions among microbial communities and their consequential impact on process efficiency. Understanding the ecological dynamics of microbial communities in WWTPs is crucial for unravelling the intricate nutrient transformation processes. Conventional wastewater treatment models often neglect this ecological aspect, but integrating ecological insights into these models, although challenging, is essential for better WWTP design and optimization. Addressing key research questions, such as identifying the members of microbial communities and their roles, is paramount for incorporating ecological insights into WWTP design. By employing meta-analysis, this study uncovers the central microbial components and their interrelationships, providing valuable insights into the factors that influence process dynamics.

Keywords—Wastewater, microbial, community, bioinformatics, bacteria

I. INTRODUCTION

The wastewater treatment process is a complex undertaking with the primary objective of eliminating contaminants and rendering the water suitable for discharge into natural water bodies or for alternative purposes [1]. It holds significant importance in safeguarding public health

and preserving the environment. Within this process, microorganisms including bacteria, fungi, and protozoa assume a pivotal role by decomposing organic pollutants and transforming them into harmless byproducts [2]. Bioreactors offer an optimal environment for the proliferation and activity of these microorganisms, which are cultivated in large quantities to effectively consume organic substances in the wastewater, thereby reducing pollutant levels. In the secondary treatment stage, microorganisms are also employed to further eliminate any remaining pollutants. Biological wastewater treatment stands as one of the most crucial applications of biotechnology. Wastewater treatment facilities play a vital role in preventing the contamination of natural ecosystems and the proliferation of diseases transmitted through sewage, such as cholera and typhoid fever.

A significant number of studies have been done in different wastewater systems, however, most of the studies have been done based on samples taken from individual applications. Only a limited number of studies have been done covering a wide range of applications, analyzing how microbial communities in a wide range can correlate.

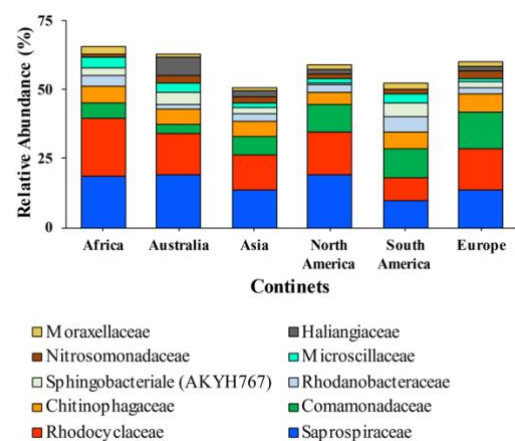


Fig. 1. Relative abundance of most abundant families by continents

The present study endeavors to undertake a meta-analysis of microbial communities encompassing diverse municipal wastewater treatment plants situated in various geographical regions. To achieve this objective, a meticulous examination of publicly accessible data was conducted via comprehensive

metagenomic analysis. Multiple datasets comprising samples from water treatment systems were carefully chosen from the study on municipal WWTPs, and subsequently subjected to analysis utilizing bioinformatic tools. Network analysis, a widely employed data analysis technique, was employed to elucidate taxa that share common ecological roles within the investigated ecosystems [3]. This method has proven to be highly effective in analyzing intricate systems characterized by complex interactions and inferring community structures.

Accordingly, both network analysis and statistical analysis will be executed to discern the identities of microorganisms present and explore their correlations.

II. MATERIALS AND METHODS

A. Bioinformatic Work Flow for the Publicly Available Data

16S rRNA amplicon sequencing datasets from various wastewater sources, obtained from the European Nucleotide Archive [4]. Usegalaxy.org server was used to run the bioinformatic pipe line to analyze the sequencing data retrieved from the selected studies [5].

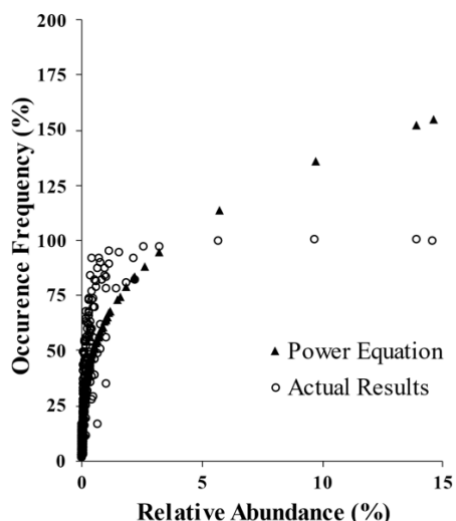


Fig. 2. Occurrence frequency vs Relative abundance actual results and predicted values using power equation

A total of 116 samples from municipal wastewater treatment plants across 6 continents were selected which included 5,536,897 reads (Study: PRJNA509305). The DADA2 genomic data analysis pipeline was used for this analysis [6]. The Filter and Trim step removed low-quality reads, adapter sequences, and contaminants to ensure data accuracy (Truncate read length = 0, minimum read length = 20, and a maximum expected error per read = 2). The plot quality profile function visualized the quality scores across all reads, while the Learn errors function estimated the error model without the need for a reference genome. *Remove Bimera Denovo* step identified and removed chimeric sequences which reduced the screened Amplicon Sequence Variants (ASV) from 80,138 to 39,251. The sequence count provides valuable information about the biological diversity and abundance in the samples. Taxonomic classification of amplicon sequence variants was performed using the *SILVA NR99* database [7].

B. Community Analysis and Mathematical Modelling

ASV data was sorted to obtain relative abundance, average abundance, abundance frequency, and key microorganisms of the samples. Then, the communities were analyzed through clustering using the PAST software [8], which performs a principal component analysis (PCA). Three models (Nachman, Hanski-Gyllenberg, Power,) [9] were used to describe the quantitative relationship between the observed occurrence frequency and relative abundance of ASVs across the selected samples.

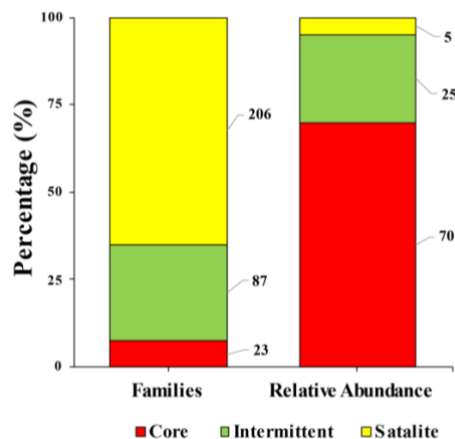


Fig. 3. Average relative abundance of three groups: core, intermittent, and transient groups based on occurrence frequency (%)

C. Network Analysis

The connectedness of microorganisms families were identified by developing a co-occurrence network using Cytoscape® [10] software. R software was used to analyze the relative abundance data of the bacterial families. Statistical significance was tested using *cor.test()* function and screening the correlation with *P* Value < 0.05. Screened nodes (representing the average relative abundance of families) were then represented with edges (representing Pearson's correlation).

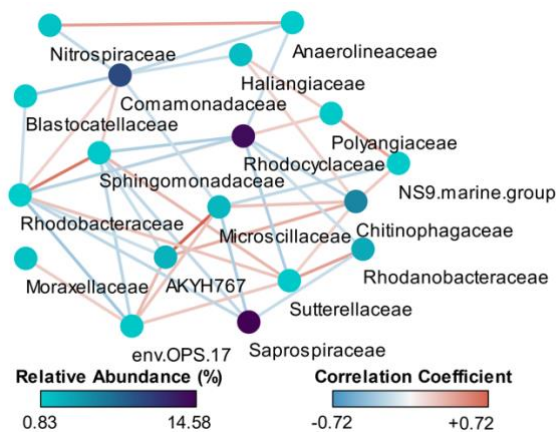


Fig. 4. Network diagram for the bacterial communities representing the association between bacterial families

III. RESULTS

A. Community Analysis and Mathematical Modelling

A total of 316 bacterial families were identified among the 116 samples obtained from municipal wastewater

treatment plants. Figure 1 shows the number of families and their relative abundance. Among these families, a subset of 23 families accounted for a cumulative relative abundance of 70%. This indicates that these 23 families collectively contribute significantly to the overall microbial composition, representing most of the microbial community in the samples. Furthermore, an additional 87 families collectively accounted for 24.99% of the cumulative relative abundance. These families, while less abundant compared to the top 23 families, still contribute substantially to the overall microbial community composition. It is worth noting that the remaining 206 families, which were not included in the top 23 or 87 families, collectively accounted for 4.72% of the cumulative relative abundance.

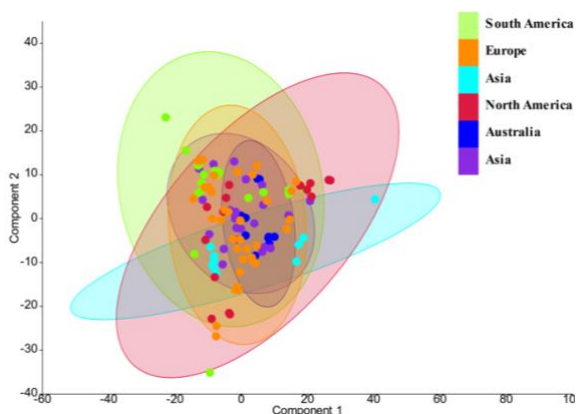


Fig. 5. PCA of all samples. 95% confidence intervals are indicated by shaded area grouping each continent

Although individually less abundant, these families play a role in the overall diversity and composition of the microbial communities in wastewater treatment plants. This distribution of cumulative relative abundance highlights the presence of both dominant and less abundant microbial families in the samples. The dominance of the top 23 families, collectively accounting for 70% of the abundance, (Figure 3) underscores their potential functional importance and influence on the overall microbial community dynamics in wastewater treatment plants. From the one way ANOSIM using PAST[8] it was identified that the communities in South America was significantly different from Asia and Europe samples (Bonfferoni corrected P value < 0.05) [11].

From the mathematical equations used to model the relationship between the prevalence (Occurrence) and abundance, Power equation was fitted with P Value = 0.01 and root sum square (rss) = 59150.11 (Figure 2).

$$P = \frac{\alpha\beta^\mu}{1 + \alpha\beta^\mu}$$

μ and P represent the relative abundance and occurrence frequency, respectively. Fitted values for the model were $\alpha=64.45$, $\beta=0.33$.

B. Network Analysis

Family level association network (Figure 4) consists of 18 nodes and 41 edges. Each node representing a bacterial family was associated with another bacterial family with an average association of 4.56. The network diameter (number

of consecutive edges for longest distance) was 3. No self-loops or multi-edge node pairs are identified in this network. The clustering coefficient for the entire network was 0.17. The highest clustering coefficient was 0.5, and there were 6 nodes with a clustering coefficient ≥ 0.2 .

IV. DISCUSSION

A. Comparative Analysis of the Global Microbial Meta Community of WWTPs

The average relative abundance values within the continents were showing similar pattern with *Saprosiraceae*, *Rhodocyclaceae*, *Comamonadaceae* and *Chitinophagaceae* covering more than 34% of the total abundance of bacterial sequences (Figure 1). This pattern have also been reported by other authors who studied activated sludge system [12], [13]. From the Principal Component Analysis with 95% confidence level, we could not see clear distinguish between the microbial communities from different continents. However with the one-way ANOSIM results we identified South American microbial meta community is different from Asia and Europe which can be a result due to the waste generation patterns, climatic conditions and specific seed cultures used during the startup of the WWTPs used to sequence samples.

The identification of core constituents within the microbial community assumes significance in comprehending the in-situ processes facilitated by bacterial cells in WWTPs. Distinct bacterial taxa possess unique metabolic characteristics that intricately interact with nutrient removal procedures. Empirical models lack the capacity to forecast alterations transpiring within the intricate course of nutrient removal. Through the examination of core families within the community, environmental engineers can gain insights into the intricate material conversion processes. Consequently, this knowledge can aid in effectively maintaining the operational parameters of the WWTPs at desired levels, ensuring the anticipated efficiency in nutrient removal.

The advent of next-generation sequencing and rapid molecular identification techniques has provided WWTP designers with a comprehensive comprehension of the microbial community constituents [14]. This advancement surpasses the reductionist methodology employed in mathematical models. With the enhancement of computational capabilities, it has become feasible to discern various behaviors exhibited by the members of the WWTP community. Consequently, more precise estimations can be attained for any given type of WWT system. In this circumstance the identification of the core microbial members in the community will be crucial. With the increasing numbers of the cells the metabolic functions will have significant influence on the whole system.

B. Network Analysis Revealed the Ecological Association Between the Families

Pearson's correlations indicate the co-occurrence relationship between two identified nodes (bacterial families). The strength of co-occurrence is higher when positive correlations are observed, while it is weaker in the case of negative correlations between the ASVs [15].

Notably, a robust positive correlation was observed between the microbial families *Microscillaceae* and *AKYH767*, suggesting a potential mutualistic or synergistic interaction between these taxonomic groups. It is important to note that both families are commonly found in activated sludge derived from WWTPs. However, the precise ecological nature of this relationship cannot be definitively explained due to the absence of isolates and controlled ecological experiments [16].

The ASV identification method offers improved taxonomic resolution compared to the conventional Operational Taxonomic Units (OTU) approach [17]. This could explain the observed outcome of obtaining 39,251 ASVs through the bioinformatic analysis. Another contributing factor is the inclusion of multiple samples from different continents. Despite the presence of closely related microbial community members, employing a higher taxonomic resolution enables the detection of subtle differences between individual members from two samples at the species level. In this analysis, all ASVs within a specific family level had to be clustered together due to their low prevalence within the selected sample range.

Similarly, a negative correlation ($R=-0.30$) was observed between the microbial families *Sphingomonadaceae* and *Rhodocyclaceae*, indicating a potential ecological rivalry between these groups. Both of these families are known to encompass denitrifying polyphosphate-accumulating organisms (DPAOs) and denitrifying glycogen-accumulating organisms (DGAOs) [18]. In aerobic or anoxic conditions, DPAOs utilize nitrogenous compounds as the electron acceptor to assimilate phosphates from the substrate [19]. Likewise, DGAOs also employ denitrification as the electron acceptor. Given their shared functionalities that are closely associated with electron exchange and competition for substrate phosphate, *Sphingomonadaceae* and *Rhodocyclaceae* can be considered as competitors. Nevertheless, further confirmation of this putative interaction necessitates controlled experiments.

The findings of this study indicate that three families, namely *Microscillaceae*, *Sphingobacteriales (env.OPS 17)*, and *Chitinophagaceae*, have demonstrated notably high betweenness centrality scores (>0.15). These high scores suggest that these taxa may possess significant importance in preserving the structure and functionality of ecological communities. By identifying and separating the modules responsible for maintaining connectivity within the network, the utilization of betweenness centrality has proven valuable in identifying keystone species within the system [20].

Notably, the aforementioned bacterial families exhibit the third highest degree (6) among all the nodes in the network. This observation signifies their substantial connectivity with other bacterial families within the system [16]. Conducting a comprehensive investigation into the characteristics and behaviors of these bacterial families is of paramount importance, as their contributions are likely to be pivotal for achieving the desired functionality within WWTPs.

V. CONCLUSION AND FUTURE WORK

This study provides valuable insights into the structures of microbial communities in activated sludge wastewater

treatment plants (WWTPs) worldwide, specifically focusing on the family level. A methodology has been developed to identify biological interactions using a DNA fingerprinting approach, which serves as a proxy for detecting relative abundance. The findings of this study lay the groundwork for further analysis of functional gene expression and metabolic pathways. Subsequent investigations can be conducted to explore the relationships between metabolic pathways and nutrient removal processes, such as ANMAMOX [21], denitrification [22], EBPR [23], COMAMMOX [24], among others.

The conclusions drawn from this meta-analysis warrant meticulous curation through a comprehensive examination of cellular metabolism. It is imperative to underscore that substantiating the hypotheses formulated in this study demands rigorous controlled experimental studies. Future investigations will involve laboratory-scale examinations wherein the metabolic conversion of the substrate will be conducted with a thorough functional genomic analysis. Notably, there is a pressing need to explore and curate the well-established rationale behind the robust association among microbial community members and the pathways intrinsic to the treatment process. This methodological approach presents a promising avenue for optimizing operations within Wastewater Treatment Plants (WWTPs) and elevating overall efficiency.

VI. REFERENCES

- [1] B. Zhang, X. Xu, and L. Zhu, "Structure and function of the microbial consortia of activated sludge in typical municipal wastewater treatment plants in winter," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017, doi: 10.1038/s41598-017-17743-x.
- [2] B. Yadav *et al.*, *Introduction to wastewater microbiology: special emphasis on hospital wastewater*. BV, 2020. doi: 10.1016/B978-0-12-819722-6.00001-8.
- [3] M. J. Sweet and M. T. Bulling, "On the Importance of the Microbiome and Pathobiome in Coral Health and Disease," *Front. Mar. Sci.*, vol. 4, no. January, pp. 1–11, 2017, doi: 10.3389/fmars.2017.00009.
- [4] European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), "The European Nucleotide Archive," <https://www.embl.org/>, 2023.
- [5] E. Afgan *et al.*, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W537–W544, 2018, doi: 10.1093/nar/gky379.
- [6] B. J. Callahan, P. J. Mcmurdie, M. J. Rosen, A. W. Han, and A. J. A., "DADA2: High resolution sample inference from Illumina amplicon data," *Nat Methods*, vol. 13, no. 7, pp. 581–583, 2016, doi: 10.1038/nmeth.3869.DADA2.
- [7] G. F. Quast C, P. Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools," *Nucleic Acids Res.*, vol. 41, 2013.
- [8] P. . Hammer, Ø., Harper, D.A.T., Ryan, "PAST: Paleontological statistics software package for education and data analysis." 2001. [Online]. Available: http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- [9] P. W. Vidanage, S.-O. O. Ko, and S. Oh, "Metagenomics Uncovers a Core SAR11 Population in Brackish Surface Waters of the Baltic Sea," *Water*, vol. 12, no. 501, 2020, doi: 10.3390/w12020501.
- [10] P. Shannon *et al.*, "Cytoscape: A Software Environment for Integrated Models," *Genome Res.*, vol. 13, no. 22, p. 426, 1971, doi: 10.1101/gr.1239303.metabolite.
- [11] X. Hu *et al.*, "Comparative analysis of the gut microbial communities in forest and alpine musk deer using high-throughput sequencing," *Front. Microbiol.*, vol. 8, no. APR, pp. 1–12, 2017, doi: 10.3389/fmicb.2017.00572.
- [12] S. Xu, J. Yao, M. Ainiwaer, Y. Hong, and Y. Zhang, "Analysis of

Bacterial Community Structure of Activated Sludge from Wastewater Treatment Plants in Winter,” *Biomed Res. Int.*, vol. 2018, p. 8278970, 2018, doi: 10.1155/2018/8278970.

- [13] S. Oh, D. Choi, and C. J. Cha, “Ecological processes underpinning microbial community structure during exposure to subinhibitory level of triclosan,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-40936-5.
- [14] M. de Celis *et al.*, “Tuning up microbiome analysis to monitor WWTPs’ biological reactors functioning,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020, doi: 10.1038/s41598-020-61092-1.
- [15] Z. Liu *et al.*, “Higher sensitivity of microbial network than community structure under acid rain,” *Microorganisms*, vol. 9, no. 1, pp. 1–18, 2021, doi: 10.3390/microorganisms9010118.
- [16] S. Begmatov *et al.*, “The structure of microbial communities of activated sludge of large-scale wastewater treatment plants in the city of Moscow,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi: 10.1038/s41598-022-07132-4.
- [17] M. Chiarello, M. McCauley, S. Villéger, and C. R. Jackson, “Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold,” *PLoS One*, vol. 17, no. 2 February, pp. 1–19, 2022, doi: 10.1371/journal.pone.0264443.
- [18] S. L. de Sousa Rollemberg, A. R. Mendes Barros, P. I. Milen Firmino, and A. Bezerra dos Santos, “Aerobic granular sludge: Cultivation parameters and removal mechanisms,” *Bioresour. Technol.*, vol. 270, no. August, pp. 678–688, 2018, doi: 10.1016/j.biortech.2018.08.130.
- [19] Y. Lin *et al.*, “Balancing denitrifying phosphorus-accumulating organisms and denitrifying glycogen-accumulating organisms for advanced nitrogen and phosphorus removal from municipal wastewater,” *Bioresour. Technol.*, vol. 369, no. 100, p. 128444, 2023, doi: 10.1016/j.biortech.2022.128444.
- [20] S. Banerjee, C. A. Kirkby, D. Schmutter, A. Bissett, J. A. Kirkegaard, and A. E. Richardson, “Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil,” *Soil Biol. Biochem.*, vol. 97, pp. 188–198, 2016, doi: 10.1016/j.soilbio.2016.03.017.
- [21] W. Xiujie, W. Weiqi, Z. Jing, W. Siyu, and L. Jun, “Dominance of *Candidatus saccharibacteria* in SBRs achieving partial denitrification: effects of sludge acclimating methods on microbial communities and nitrite accumulation,” *RSC Adv.*, vol. 9, no. 20, pp. 11263–11271, 2019, doi: 10.1039/C8RA09518C.
- [22] D. A. Hutchins and F. Fu, “Microorganisms and ocean global change,” *Nat. Microbiol.*, vol. 2, no. 6, p. 17058, 2017, doi: 10.1038/nmicrobiol.2017.58.
- [23] D. H. J. Wong and R. G. Beiko, “Transfer of energy pathway genes in microbial enhanced biological phosphorus removal communities,” *BMC Genomics*, vol. 16, no. 1, pp. 1–13, 2015, doi: 10.1186/s12864-015-1752-5.
- [24] S. Rahimi, O. Modin, and I. Mijakovic, “Technologies for biological removal and recovery of nitrogen from wastewater,” *Biotechnol. Adv.*, vol. 43, no. June, p. 107570, 2020, doi: 10.1016/j.biotechadv.2020.107570.