

# Structural differences and specific biomarkers of the microbial communities in wastewater and natural water reservoirs

Inna Kucheroва

[https://github.com/inna-tsymb/metagenomics/tree/main/lecture\\_03](https://github.com/inna-tsymb/metagenomics/tree/main/lecture_03)

## 1. Introduction

Early was hypothesized that natural aquatic ecosystems (rivers/ponds) have fundamental microbiome differences compared to anthropogenic sources (wastewater, sludge). To test this hypothesis and identify specific drivers of this differences, there was conducted a microbiome wide association study. Linear regression on CLR-transformed data combined with logistic regression was conducted, and identified powerful species-specific biomarkers – natural and anthropogenic environments form completely different ecological niches.

**Objective:** To identify specific microbial species responsible for fundamental differences between anthropogenic and natural ecosystems.

**Hypothesis:** Expected that natural environments would have significantly higher biodiversity compared to anthropogenic systems due to lower pollution pressure and more stable ecological conditions.

## 2. Methodology

- Taxon filtering: only species levels was selected, as strains introduce high level of noise and uncertainty.
- Cutoff threshold: inclusive filtering logoc (OR) was applied; species occurred in >10% or having an average relative abundance of >10% were included, which allows to save both stable colonizers and highly specialized species.
- Statistical analysis: main method was linear regression (Type II ANOVA) on CLR-transformed data to access quantitative changes. Logistic regression was used for binary data to validate results.
- Multiple testing correction: FDR (Benjamini-Hochberg) method was applied with significance threshold <0.05, all comparisons were performed in both normalized (Shultz downsampled to n=20) and non-normalized data.

## 3. Results

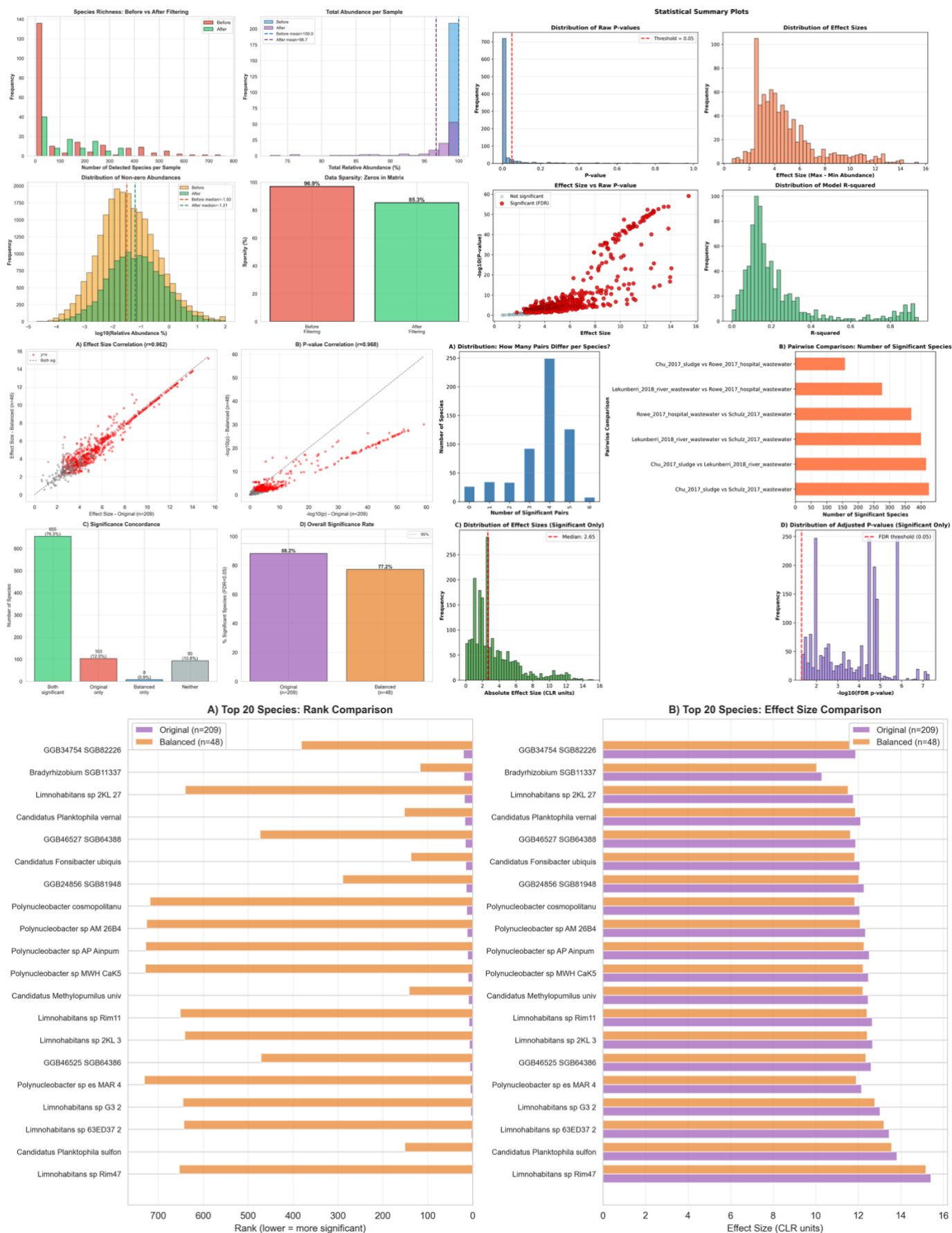
### 3.1. fundamental composition differences

Results confirmed initial hypothesis about different sources form fundamentally different microbiomes. In normalized dataset linear regression identified 758 significant species among 859 tested (88.2%) with ANOVA+FDR. The pairwise method showed that most discriminative comparison was between river water and municipal wastewater, where >91% of species significantly different.

### 3.2. Biomarkers of natural water reservoirs

Showed that all 30 of most statistically significant associated species demonstrate identical pattern: they are extremely abundant in river water and rare in hospital wastewater. Most robust biomarkers of river ecosystem are oligotrophic freshwater specialists:

- *Limnohabitans* sp Rim47: extreme quantitative enrichment ( $p=5.09e-56$ , effect = 15.25 CLR units) demonstrates almost perfect presence: 100% in freshwater vs. 5% in hospital wastewater.



**Fig.1.** Comprehensive assessment of differences in microbiome structure. The panel illustrates quality control metrics after filtering, overall statistics of regression models, pairwise comparisons between wastewater types, and stability of ranks and effect sizes for the top 20 biomarkers. The data confirms the highest discrimination between natural river water and municipal wastewater treatment plants.

- *Candidatus Planktophila sulfonica* and *Polynucleobacter spp.*: also present core of natural microbiome with extremely high effect size. For *C. Planktophila sulfonica*, an extreme level of enrichment in river water was recorded ( $p = 3.45e-92$ ; effect = 13.64 CLR units), with the source of water explaining 87.5% of the variation in the abundance of this species. At the same time, five different species of *Polynucleobacter* were among the top 30 most significant biomarkers of the river environment. Their averaged statistical indicators demonstrate strong association with natural water bodies: the average effect size is about 12 CLR units (p-value range from  $10^{-84}$  to  $10^{-86}$ ).

### 3.3. Biomarkers of anthropogenic ecosystems

In contrast to natural waters, anthropogenic ecosystems have their own specific microbiome. Eg, *Vibrio parahaemolyticus* is positively associated with municipal wastewaters with strong quantitative enrichment ( $p=1.20e-15$ , effect = 8.52 CLR units) and systematic presence: ~90% in wastewater and 0% in rivers.

### 3.4. Validation of methods

There is extremely high concordance between linear (CLR abundance) and logistic (binary presence) regressions. In normalized mode, 708 (82.5%) of species were significant for both models. This indicates that main microbial actors exhibit changes according to the all-or-nothing principle.

## 4. Discussion

Results fully confirm and expand previous hypothesis. Difference between anthropogenic and natural ecosystems lies not only in general biodiversity metrics, but also in complete replacement of ecological niches. Natural waters support the development of obligate freshwater oligotrophs, which evolutionarily adapted to nutrient-poor environments. In contrast, wastewater act like powerful selective factor, replacing natural species with heterotrophs. The selected species with the maximum effect > 10 CLR units are ready candidates for the role of diagnostic biomarkers for environmental monitoring and identification of pollution sources.

## 5. Limitations

Despite the results, this study has a number of limitations:

- Heterogeneity of metadata and lack of physicochemical parameters.
- Sample imbalance and potential batch effect.
- Lack of temporal dynamics.
- Limitation of reference databases.
- Geographic fragmentation.

## 6. Conclusions

Natural and anthropogenic ecosystems have not only different level of diversity, but also fundamentally different and almost unrelated microbial communities. Depending on statistical test, from 87% to 98.2% of the species demonstrated significant difference depending on source of origin.

Natural water reservoirs characterized by totally unique community structure dominated by oligotrophic specialists adapted to conditions with low nutrient content. Wastewaters form their own ecological niches, which suppress natural microbiota and encourage the development of specific heterotrophs and bacteria adapted to engineering conditions.

The identified biomarker species demonstrated extremely large effect sizes 12-15 CLR units and nearly perfect diagnostic presence (100% prevalence in river water versus 0-5% in wastewater). This makes them ideal candidates for the development of source tracking tools and the identification of the impact of wastewater on natural water.

## Literature

- Frederik Schulz et al., Giant viruses with an expanded complement of translation system components. *Science* 356, 82-85 (2017). DOI: 10.1126/science.aal4657
- Chu BTT Petrovich ML, Chaudhary A, Wright D, Murphy B, Wells G, Poretsky R. 2018. Metagenomics Reveals the Impact of Wastewater Treatment Plants on the Dispersal of Microorganisms and Genes in Aquatic Sediments. *Appl Environ Microbiol* 84:e02168-17. <https://doi.org/10.1128/AEM.02168-17>
- Chopyk, J., Nasko, D.J., Allard, S. et al. Seasonal dynamics in taxonomy and function within bacterial and viral metagenomic assemblages recovered from a freshwater agricultural pond. *Environmental Microbiome* 15, 18 (2020). <https://doi.org/10.1186/s40793-020-00365-8>
- Will P. M. Rowe, Craig Baker-Austin, David W. Verner-Jeffreys, Jim J. Ryan, Christianne Micallef, Duncan J. Maskell, Gareth P. Pearce, Overexpression of antibiotic resistance genes in hospital effluents over time, *Journal of Antimicrobial Chemotherapy*, Volume 72, Issue 6, June 2017, Pages 1617–1623, <https://doi.org/10.1093/jac/dkx017>
- Lekunberri I, Balcázar JL, Borrego CM. Metagenomic exploration reveals a marked change in the river resistome and mobilome after treated wastewater discharges. *Environ Pollut.* 2018 Mar;234:538-542. doi: 10.1016/j.envpol.2017.12.001. Epub 2017 Dec 21. PMID: 29220785.