

# Evaluating differences in bacterial association networks from rainforest and converted pasturelands in the Colombian Amazon region using CCLasso and SPRING inference methods

Capstone project report

Network Biology - MSB1014

Academic Year: 2023-2024

 Sebastian Ayala Ruano

Faculty of Science and Engineering

Maastricht University

Student ID: i6314501

`s.ayalaruano@student.maastrichtuniversity.nl`

October 27, 2023

## 1 Introduction

The Earth's microbiome - all microbial communities, their structural components, and metabolic byproducts - constitutes the largest portion of biodiversity on our planet (1). Due to their widespread presence, microbes profoundly impact the biosphere, collectively regulating the global biogeochemical cycles, soil fertility, and other processes. Hence, investigating microbiomes can provide insights into food security, climate change, and land use issues (2; 3). In particular, it is crucial to study the soil microbial diversity of tropical ecosystems facing threats due to extensive deforestation for crop cultivation or pasture expansion (4; 5).

The exponential progress in omics technologies over the last decades has facilitated culture-independent methods for microbiome research (1). Specifically, metagenomics techniques use genetic sequences from environmental samples to characterize the taxonomic profile and quantify the relative abundance of microorganisms in these ecosystems (2; 6). Standard methods to analyze metagenomics data (e.g., diversity indices) focus on specific taxa at the individual level, without considering microbial interactions (7).

Microbial association networks (MANs) provide insights into potential ecological interactions among microbes, including mutualism, competition, and more. Furthermore, these networks can reveal communities that share ecological functions or keystone taxa playing crucial roles in the system (8; 9). In MANs, nodes correspond to Operational Taxonomical Units (OTUs) at a given taxonomic rank, and edges between nodes denote significant co-presence (positive relationships) or mutual exclusion (negative relationships) patterns in OTU abundances across samples. Multiple association metrics are available for inferring MANs, mainly founded on correlation, proportionality, and conditional dependence approaches (10; 11).

In this project, the CCLasso (12) (correlation-based) and SPRING (13) (conditional dependence-based) methods were used to explore differences in microbiomes found in the rainforest and converted pasturelands of the northwest Colombian Amazon region.

## 2 Methods

The NetCoMi R package v1.1 (10) was the central framework to perform the network inference, analysis, visualization, and comparison. All underlying code is accessible through GitHub ([github.com/sayalaruano/img2brain](https://github.com/sayalaruano/img2brain)) under the MIT license.

### 2.1 Dataset and preprocessing

The raw sequencing data used for this project is available at the PRJEB44163 project from the European Nucleotide Archive (ENA) database. This research collected 52 soil samples from the Colombian Amazon region: 36 from rainforest areas and 16 from converted pasturelands. The rainforest samples were the reference with minimal intervention, while the pastureland represented the land use systems.

MGnify (14) is a platform that automates the analysis of metagenomics datasets from ENA and other databases. The abundance table and taxonomic profiles to infer the MANs in this project were obtained from the MGYS00005779 study, which applied the MGnify's pipeline v5 (14). The first step was downloading the data and metadata using the MGnifyR package v0.1 (15), enabling the utilization of the MGnify API in R scripts. Then, the data was preprocessed and manipulated using the Phyloseq v1.44 (16) and Microbiome

v1.22 (17) R packages. The dataset was filtered to retrieve only bacterial OTUs, excluding taxa from other life kingdoms. In addition, the data was aggregated at the family taxonomic level, obtaining 200 OTUs.

## 2.2 Network inference

NetCoMi has a computational workflow that involves calculating associations among OTUs using a specified metric, applying sparsification if necessary, and converting these associations into dissimilarities and subsequently into similarities, resulting in the adjacency matrix for the inferred networks. The inference of the bacterial association networks (BANs) was performed using the CCLasso (12) and SPRING (13) association metrics, which respectively use correlation and conditional dependence approaches.

The parameters required for the CCLasso algorithm included the selection of zero-treatment, sparsification, and dissimilarity methods. In this study, the CCLasso BANs were built applying the "pseudo count" method for zero-treatment (designated as "pseudoZO"), a sparsification threshold of 0.2, and the "signed" method for converting associations into dissimilarities, where the distance is highest for strongly negative associated taxa. The SPRING metric incorporates a rank-based correlation estimator for zero-treatment and employs the Stability Approach to Regularization Selection (StARS) (18) for network sparsification. To construct the SPRING BANs, a 0.02 StARS threshold and the "signed method" for dissimilarity transformation were used, leaving the other parameters as defaults. Thresholds for both networks were chosen from a range of values, 0.1 to 0.9 for CCLasso, and 0.01 to 0.2 for SPRING. The selected thresholds generated a modular structure while minimizing network density.

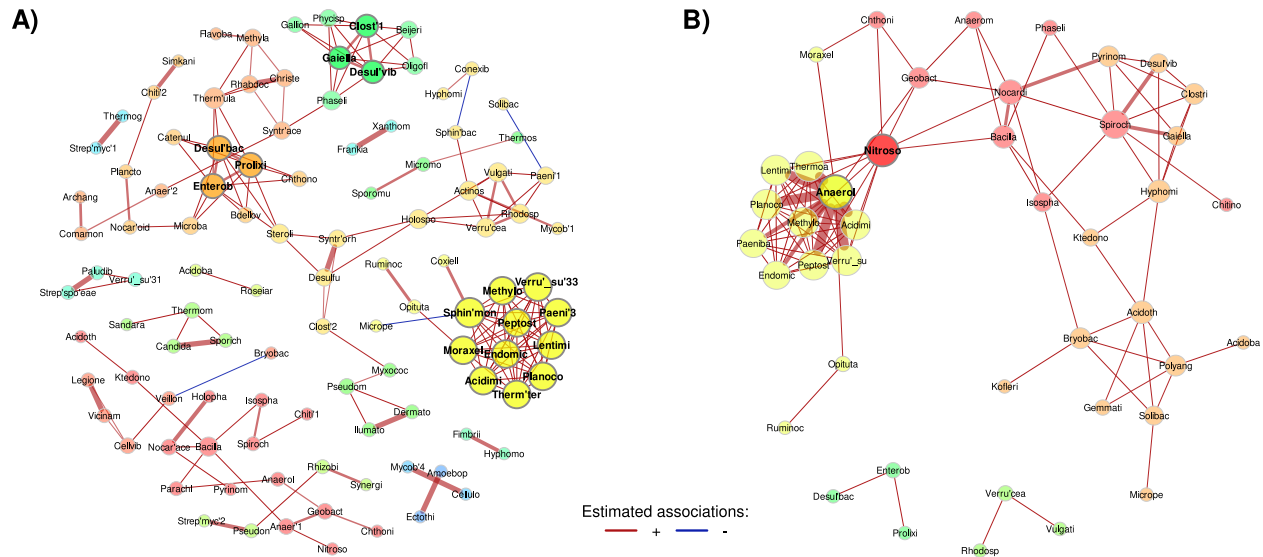
## 2.3 Network analysis, visualization, and comparison

Global topological measures were computed as an initial step to characterize the overall network structures. Following this, central nodes were identified using metrics such as degree, betweenness, and closeness. Furthermore, the fast greedy modularity optimization algorithm (19) was applied to find network communities. The network visualizations were created using the Fruchterman-Reingold force-field layout (20), with nodes' size and color reflecting topological and taxonomic attributes. All singleton nodes were removed from the networks. Ultimately, both visual and quantitative network comparisons were performed, with the quantitative part involving a non-parametric permutation test with 100 repetitions.

# 3 Results and Discussion

## 3.1 Network inference methods comparison

To explore the differences between the two network inference methods, the initial step involved visually comparing the resultant networks and an evaluation of global topological metrics. Figure 1 shows the BANs derived from forest samples. Notably, the CCLasso network has fewer nodes and edges compared to the SPRING-derived network (Figure 1 and Table 1), primarily due to the threshold chosen for the sparsification method. CCLasso performed well with the default threshold of 0.2, while the recommended StARS cutoff of 0.1 for SPRING produced dense networks sensitive to minor parameter adjustments. Hence, extensive parameter tuning was required to establish an appropriate threshold for the SPRING networks, a challenge mentioned in previous reviews (10; 11).



**Figure 1:** Forest bacterial association networks inferred with **A)** SPRING and **B)** CCLasso. Nodes represent family-level OTUs, colored by their community and sized by degree. Red edges correspond to positive associations and blue edges to negative ones. Hubs, defined as nodes with a degree value above the empirical 90% quantile, are highlighted with bold text and borders. OTU labels were shortened to the first seven letters of their names, with the list of full names available in *Supplementary Information*.

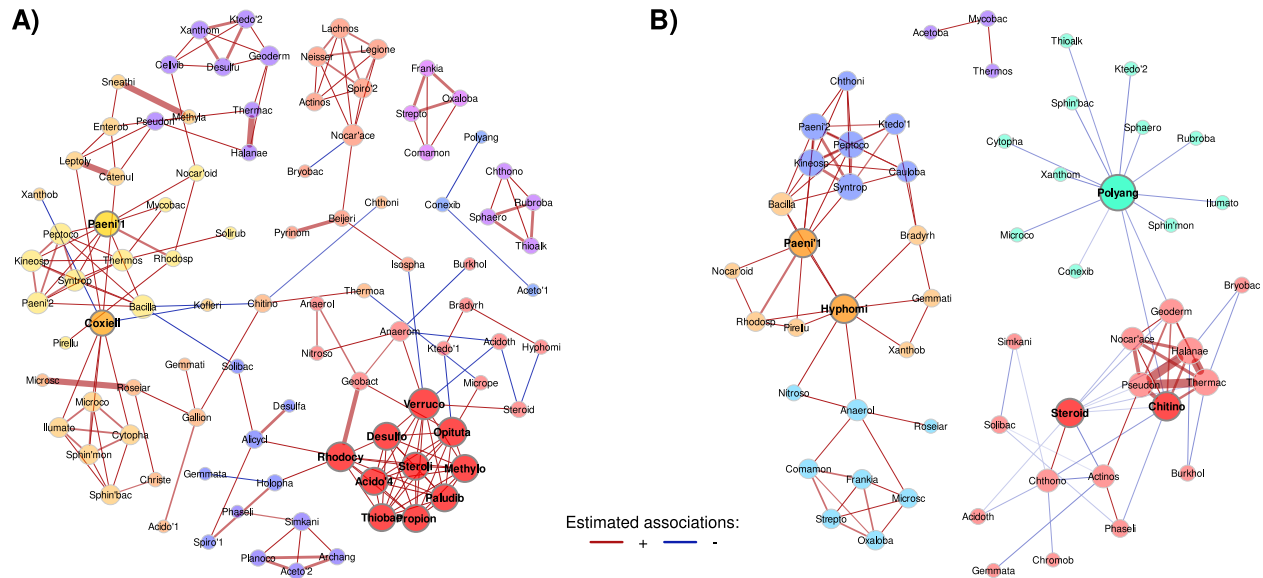
**Table 1:** Global Topological Metrics of the Bacterial Association Networks

Metric	Forest SPRING	Forest CCLasso	Pasture SPRING	Pasture CCLasso
<b>Nodes</b>	108	43	95	56
<b>Edges</b>	183	108	185	113
<b>Number of communities</b>	19	6	10	7
<b>Modularity</b>	0.69	0.46	0.71	0.43
<b>Average path length</b>	6.02	3.20	6.96	2.78
<b>Average degree*</b>	3.39	5.02	3.89	4.04

\* The metric was calculated based on the giant component of the networks.

Moreover, the STRING BAN has more clusters and higher modularity values than the CCLasso network (Table 1). Although elevated values of these measures indicate the presence of a clustering structure, an excessive number of communities may not be desirable, as it increases the likelihood of some groups being artifacts (21). Regarding the giant component average path length, the CCLasso BAN has a lower value than the SPRING network (Table 1), indicating a more efficient information flow in the former. Another difference in forest BANs is the ease of identifying degree-based hub nodes, particularly evident in the CCLasso-derived network (Figure 1), and supported by the lower average degree value in the SPRING network (Table 1). Considering the commonalities between the forest networks, both show comparable clustering coefficients and positive edge percentage values.

Figure 2 presents the pastureland BANs, exhibiting comparable distinctions to those observed in the forest networks. These differences are also reflected in the global topological measures in Table 1. Noteworthy, there is an increase in negative interactions among the OTUs in these networks compared to their forest counterparts, which was statistically confirmed with a significance level of 0.05 using non-parametric permutation testing. This pattern could be possibly linked to antagonistic relationships like competition and predation in the converted land systems (8; 7).



**Figure 2:** Pasture bacterial association networks inferred with **A)** SPRING and **B)** CCLasso. Nodes represent family-level OTUs, colored by their community and sized by degree. Red edges correspond to positive associations and blue edges to negative ones. Hubs, defined as nodes with a degree value above the empirical 90% quantile, are highlighted with bold text and borders. OTU labels were shortened to the first seven letters of their names, with the list of full names available in *Supplementary Information*.

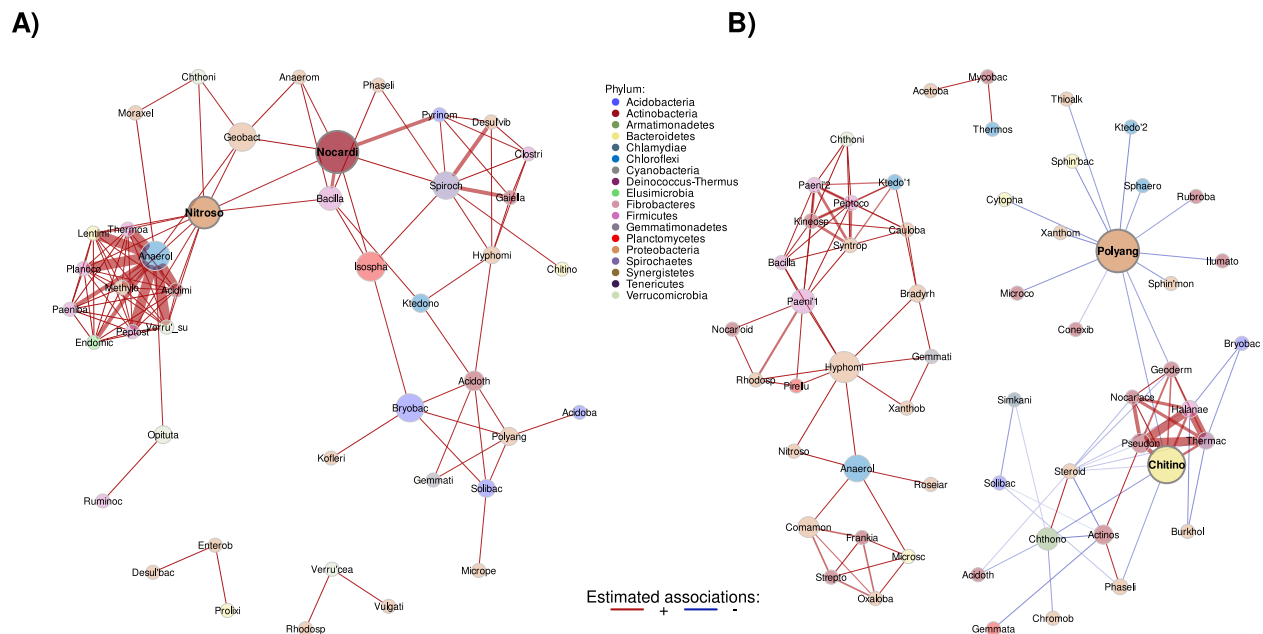
To enhance the comparison of networks from the distinct methods, association coincidences were calculated (Table 2). In both forest and pasture conditions, the low percentage of coincidences indicates substantial contrasts in the predicted interactions from the two association metrics. Considering the information provided, CCLasso inferred graphs with more favorable topological features for capturing microbiome interactions, so they were selected for the biological comparisons in the following section.

**Table 2:** Coincidence Percentages of Associations Among Forest and Pasture Networks Inferred Using SPRING and CCLasso

Association	Forest	Pasture
Positive-Positive	23%	12.98%
Negative-Negative	0	0.76%
Positive-Negative	0	0
No coincidence	77%	86.26%

### 3.2 Networks comparison by land use system

Several studies have reported differences in the bacterial community composition of soil samples taken from tropical rainforests and converted land systems in South America (22; 23) and Southeast Asia (24; 25; 26). Across all locations, a recurring pattern is the decrease of Acidobacteria phylum OTUs in converted systems compared to rainforests (25; 22), along with an opposite trend observed for Actinobacteria (27; 26). These two tendencies are evident in the CCLasso networks (Figure 3), where the pastureland graph has about half the Acidobacteria OTUs and over three times the Actinobacteria OTUs compared to the forest network. The decrease in Acidobacteria OTUs is associated with increased carbon and pH in converted systems (26), while the rise in Actinobacteria may be attributed to their role in negative ecological interactions, potentially related to their antibiotic production. Additionally, the BANs in Figure 3 exhibit differences in specific OTUs, with some reported as keystone species in forests (e.g., Nitrosomonadaceae family (28)) and others as indicators of soil perturbation due to human activity (e.g., Chitinophagaceae family (29)).



**Figure 3: A) Forest and B) Pasture bacterial association networks** inferred with CCLasso. Nodes represent family-level OTUs, colored by their phylum and sized by betweenness. Red edges correspond to positive associations and blue edges to negative ones. Hubs, defined as nodes with a betweenness value above the empirical 90% quantile, are highlighted with bold text and borders. OTU labels were shortened to the first seven letters of their names, with the list of full names available in *Supplementary Information*.

In summary, this study evaluated two methods for inferring BANs from rainforest and converted pasturelands, with CCLasso being the preferred choice due to its topological features and ease of parameter tuning. Moreover, CCLasso networks confirmed prior findings of taxa displaying differences between rainforest and pasturelands at the phylum and family levels, demonstrating the utility of BANs in understanding the impact of land conversion on the rainforest microbiome. However, these methods are simplifications of intricate microbiome systems, so their predictions should be analyzed cautiously.

## References

- [1] S. M. Gibbons and J. A. Gilbert, “Microbial diversity—exploration of natural ecosystems and microbiomes,” *Current Opinion in Genetics & Development*, vol. 35, pp. 66–72, 2015.
- [2] A. Boetius, “Global change microbiology — big questions about small life for our future,” *Nature Reviews Microbiology*, vol. 17, no. 6, pp. 331–332, 2019.
- [3] C. Averill, M. A. Anthony, P. Baldrian, F. Finkbeiner, J. van den Hoogen, T. Kiers, P. Kohout, E. Hirt, G. R. Smith, and T. W. Crowther, “Defending Earth’s terrestrial microbiome,” *Nature Microbiology*, vol. 7, no. 11, pp. 1717–1725, 2022.
- [4] P. Baldrian, R. López-Mondéjar, and P. Kohout, “Forest microbiome and global change,” *Nature Reviews Microbiology*, vol. 21, no. 8, pp. 487–501, 2023.
- [5] S. Pajares, B. J. M. Bohannan, and V. Souza, “Editorial: The Role of Microbial Communities in Tropical Ecosystems,” *Frontiers in Microbiology*, vol. 7, p. 1805, 2016.
- [6] V. de Lorenzo, “15 years of microbial biotechnology: the time has come to think big—and act soon,” *Microbial Biotechnology*, vol. 15, no. 1, pp. 240–246, 2022.
- [7] M. Layeghifard, D. M. Hwang, and D. S. Guttman, “Disentangling Interactions in the Microbiome: A Network Perspective,” *Trends in Microbiology*, vol. 25, no. 3, pp. 217–228, 2017.
- [8] K. Faust and J. Raes, “Microbial interactions: from networks to models,” *Nature Reviews Microbiology*, vol. 10, no. 8, pp. 538–550, 2012.
- [9] S. M. Koder, P. Das, J. A. Gilbert, and H. L. Lutz, “Conceptual strategies for characterizing interactions in microbial communities,” *iScience*, vol. 25, no. 2, 2022.
- [10] S. Peschel, C. L. Müller, E. von Mutius, A.-L. Boulesteix, and M. Depner, “NetCoMi: network construction and comparison for microbiome data in R,” *Briefings in Bioinformatics*, vol. 22, no. 4, p. bbaa290, 2021.
- [11] M. S. Matchado, M. Lauber, S. Reitmeier, T. Kacprowski, J. Baumbach, D. Haller, and M. List, “Network analysis methods for studying microbial communities: A mini review,” *Computational and Structural Biotechnology Journal*, vol. 19, pp. 2687–2698, 2021.
- [12] H. Fang, C. Huang, H. Zhao, and M. Deng, “CCLasso: correlation inference for compositional data through Lasso,” *Bioinformatics*, vol. 31, no. 19, pp. 3172–3180, 2015.
- [13] G. Yoon, I. Gaynanova, and C. L. Müller, “Microbial Networks in SPRING - Semi-parametric Rank-Based Correlation and Partial Correlation Estimation for Quantitative Microbiome Data,” *Frontiers in Genetics*, vol. 10, p. 516, 2019.
- [14] L. Richardson, B. Allen, G. Baldi, M. Beracochea, M. Bileschi, T. Burdett, J. Burgin, J. Caballero-Pérez, G. Cochrane, L. Colwell, T. Curtis, A. Escobar-Zepeda, T. Gurbich, V. Kale, A. Korobeynikov, S. Raj, A. Rogers, E. Sakharova, S. Sanchez, D. Wilkinson, and R. Finn, “MGnify: the microbiome sequence data analysis resource in 2023,” *Nucleic Acids Research*, vol. 51, no. D1, pp. D753–D759, 2023.
- [15] T. Borman, B. Allen, and L. Lahti, “MGnifyR: R interface to EBI MGnify metagenomics resource,” 2023.

- [16] P. J. McMurdie and S. Holmes, “phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data,” *PLOS ONE*, vol. 8, no. 4, p. e61217, 2013.
- [17] L. Lahti and S. Sudarshan, “Microbiome R package,” 2012.
- [18] H. Liu, K. Roeder, and L. Wasserman, “Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models,” in *Advances in Neural Information Processing Systems*, vol. 23, Curran Associates, Inc., 2010.
- [19] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” *Physical Review E*, vol. 70, no. 6, p. 066111, 2004.
- [20] T. M. J. Fruchterman and E. M. Reingold, “Graph drawing by force-directed placement,” *Software: Practice and Experience*, vol. 21, no. 11, pp. 1129–1164, 1991.
- [21] M. E. J. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [22] J. L. M. Rodrigues, V. H. Pellizari, R. Mueller, K. Baek, E. d. C. Jesus, F. S. Paula, B. Mirza, G. S. Hamaoui, S. M. Tsai, B. Feigl, J. M. Tiedje, B. J. M. Bohannan, and K. Nüsslein, “Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 3, pp. 988–993, 2013.
- [23] A. Pedrinho, L. W. Mendes, L. F. Merloti, M. d. C. da Fonseca, F. d. S. Cannavan, and S. M. Tsai, “Forest-to-pasture conversion and recovery based on assessment of microbial communities in Eastern Amazon rainforest,” *FEMS Microbiology Ecology*, vol. 95, no. 3, p. fiy236, 2019.
- [24] S. A. Wood, J. A. Gilbert, J. W. Leff, N. Fierer, H. D’Angelo, C. Bateman, S. M. Gedallovich, C. M. Gillikin, M. R. Gradoville, P. Mansor, A. Massmann, N. Yang, B. L. Turner, F. Q. Brearley, and K. L. McGuire, “Consequences of tropical forest conversion to oil palm on soil bacterial community and network structure,” *Soil Biology and Biochemistry*, vol. 112, pp. 258–268, 2017.
- [25] T. Kaupper, S. Hetz, S. Kolb, S. Yoon, M. A. Horn, and A. Ho, “Deforestation for oil palm: impact on microbially mediated methane and nitrous oxide emissions, and soil bacterial communities,” *Biology and Fertility of Soils*, vol. 56, no. 3, pp. 287–298, 2020.
- [26] L. Lee-Cruz, D. P. Edwards, B. M. Tripathi, and J. M. Adams, “Impact of Logging and Forest Conversion to Oil Palm Plantations on Soil Bacterial Communities in Borneo,” *Applied and Environmental Microbiology*, vol. 79, no. 23, pp. 7290–7297, 2013.
- [27] P. Hill, V. Krištůfek, L. Dijkhuizen, C. Boddy, D. Kroetsch, and J. D. van Elsas, “Land Use Intensity Controls Actinobacterial Community Structure,” *Microbial Ecology*, vol. 61, no. 2, pp. 286–302, 2011.
- [28] K. S. Boyarshin, V. V. Adamova, Z. Wentao, O. Y. Obuhova, M. V. Kolkova, V. A. Nesterenko, O. S. Beshpalova, V. V. Kluyeva, K. A. Degtyareva, Y. N. Kurkina, O. A. Makanina, and I. V. Batlutskaia, “The Effect of Long-Term Agricultural Use on the Bacterial Microbiota of Chernozems of the Forest-Steppe Zone,” *Diversity*, vol. 15, no. 2, p. 191, 2023.



- [29] S. M. Hermans, H. L. Buckley, B. S. Case, F. Curran-Cournane, M. Taylor, and G. Lear, "Bacteria as Emerging Indicators of Soil Condition," *Applied and Environmental Microbiology*, vol. 83, no. 1, pp. e02826–16, 2016.