GMR
The Original

STRUCTURAL DIVERSITY OF THERMOTOLERANT BACTERIA FROM PETROLEUM RESERVOIRS

Sajida Tayyaba[1], Arshia Amin[2]*

[1]Department of Bioinformatics and Biosciences, Faculty of Health and Life Sciences
[2]Capital University of Science and Technology Islamabad
*Corresponding Author: Arshia Amin, Email: arshia.amin@cust.edu.pk, arshiaamin.butt@gmail.com

ABSTRACT
Shotgun metagenomics combined with 16S rRNA amplicon sequencing enables genome-resolved characterization of complex microbial communities without cultivation. These approaches reveal microbial depth, diversity, relative distribution and abundance in the pristine sites where its helping in microbially enhanced oil recovery and bioremediation. This study examines the structure of thermotolerant bacteria in subsurface cores (3,770-4,315 m) of the Nashpa Oil field located in Haripur, Pakistan through the application of 16S amplicon and shotgun metagenomics. Analysis revealed *Proteobacteria* dominance alongside *Firmicutes* and *Bacteroidota*, with genera such as *Alcanivorax*, *Marinobacter*, and *Rhodococcus* which are reported to be involved in hydrocarbon degradation, heat-shock responses and low-oxygen adaptations. These findings highlight the reservoir microbiome as a resource for microbial enhanced oil recovery (MEOR) and hydrocarbon bioremediation.

KEYWORDS: Shotgun metagenomic sequencing; structural profile, MEOR

1. INTRODUCTION

Metagenomics enables the study of soil ecosystems in a way that allows the structure and function of the most diverse microbial communities on Earth to be described (Child et al., 2024). On its own, metagenomics provides potentially role and phylogeny for uncultivated taxa, and evolutionary profiles regarding community function and structure (Zhang et al., 2021). With the rapid development of high-throughput sequencing technologies, metagenomics has gradually evolved from small-scale clone libraries to large-scale, deep sequencing projects, enabling the profiling of thousands of samples in parallel (Pérez-Cobas et al., 2020). Two major experimental strategies that are widely used within metagenomic studies are marker-gene (amplicon) sequencing and shotgun metagenomic sequencing. Marker-gene approaches, such as 16S rRNA gene sequencing for bacteria and archaea, target conserved loci to characterize community composition and relative abundance. They are cost-efficient and appropriate for large cohorts, but they have limited functional information and a lower taxonomic resolution, especially at species or strain level (Durazzi et al., 2021). In contrast, shotgun metagenomics sequences the entire pool of DNA in a sample, enabling simultaneous assessment of taxonomic profiles as well as metabolic pathways, antimicrobial resistance genes and mobile genetic elements. Subsurface petroleum reservoirs constitute extreme deep biosphere environments characterized by high temperature, pressure, salinity, and limited nutrient availability. Despite these harsh conditions, diverse and metabolically active microbial communities persist and play key roles in hydrocarbon degradation, biogeochemical cycling, reservoir souring, and gas production (Salam & Obayori, 2019). Importantly, reservoir-associated microorganisms contribute to processes relevant to microbial enhanced oil recovery (MEOR), including biosurfactant production, biopolymer formation, and biogas generation, which can improve oil mobilization and recovery efficiency (Nam et al., 2023) (Liu et al., 2025). Understanding the structure of microbial communities in petroleum reservoirs is therefore critical for both ecological interpretation of the deep biosphere and the development of biologically driven strategies for reservoir management (Kumar et al., 2024). This approach has been particularly valuable for reconstructing metagenome-assembled genomes (MAGs), which represent uncultivated microbial taxa and have substantially expanded our understanding of microbial diversity in subsurface environments (Mirete et al., 2025). Such genome-resolved analyses are essential for linking microbial community structure to ecological function and MEOR-relevant processes in petroleum reservoirs. IMG or RAST programs can be selected if the study's goal is a reconstructed genome and if the assembly produces large contigs. Metagenome-assembled genomes (MAGs) are a large number of uncultivated genome equivalents from a remarkably diverse range of habitats that have been produced via metagenomics. These MAGs have greatly improved our understanding of microbial diversity and challenged some long-held beliefs derived from studies of isolated or domesticated species by providing insights into previously inaccessible and unknown taxa that were previously unexplored and unknown, often referred to as microbial "dark matter" (Wu et al., 2025). Thus, detailed Shotgun based studies of thermotolerant bacteria from deep subsurface are required in order to establish their phylogenetic placement and signs of thermal adaptation. To date, deep subsurface petroleum core microbiomes from

Pakistan have rarely been investigated using 16S amplicon sequencing, resulting in limited understanding of microbial potential in these reservoir ecosystems. This study integrates 16S rRNA amplicon sequencing to characterize community structure of thermotolerant microbial communities from deep petroleum reservoir cores. These samples capture vertical heterogeneity in subsurface physicochemical conditions, enabling assessment of how microbial community structure is related to hydrocarbon degradation and MEOR relevant processes vary with depth in the reservoir. By describing the results of 16S amplicon analysis we seek to: establish the phylogenetic diversity of thermotolerant bacteria in deep subsurface environments and provide insights into the biotechnological potentials of thermotolerant bacteria from the deep subsurface cores.

2. MATERIALS AND METHODS

2.1 Sample collection

The Nashpa oil field, Haripur, Pakistan (33° 6' 54.9684" N, 71° 5' 43.9260" E), provided the soil samples for 16s amplicon sequencing and MAGs recovery from depths of 3770–4315 metres. Immediately after collection, all samples were stored on dry ice and then transported to the university laboratory at cooled temperatures (Li et al., 2023).

2.2 Alpha diversity analysis

Alpha diversity indices, including Shannon, Chao1, and Simpson, were calculated for each sample. Statistical comparisons among groups were performed using the Kruskal–Wallis rank-sum test. Results are reported as mean ± standard deviation.

2.3 Beta diversity analysis

Beta diversity was calculated using both weighted and unweighted UniFrac distance matrices in QIIME2 , based on a rooted phylogenetic tree generated from aligned 16S rRNA gene sequences. Principal Coordinate Analysis (PCoA) was performed to visualize community differences among samples. Statistical significance between groups was assessed using PERMANOVA with 999 permutations.

2.4 Genome resolved metagenomic recovery

2.4.1 DNA extraction

Extraction of the DNA from the soil samples was performed with a DNeasy PowerSoil kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The amount and purity of the obtained DNA were evaluated via a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States), and the samples were finally stored at -20°C (Pacwa-Płociniczak et al., 2020).

2.4.2 Library construction, preparation, sequencing and MAGs recovery

The Illumina TruSeq DNA Library Prep Kit was used for library preparation. After preparation, shotgun sequencing of the eluted DNA samples was performed by Microsynth (China) on an Illumina NovaSeq system with 2 × 150 bp (Frey et al., 2022). Quality control was performed via FastQC (v0.11.9). The reads were preprocessed by stripping adapter sequences via the Trim Reads tool, with a default parameter-quality score cut-off of 0.05 and up to 2 ambiguous nucleotides. Some samples with poor coverage (too few reads) were excluded from further analysis. Clean reads from the microbiome assembly and MAGs recovery were then assembled via MEGAHIT v1.2.9, following the removal of contigs shorter than 2000 bp.

3. RESULTS AND DISCUSSION

3.1 Physicochemical characteristics of samples for 16s amplicon and MAGs recovery

Physicochemical characteristics (Table 1) summarize pH, EC and phosphorus, nitrate, chromium, cadmium and lead concentration.

Table 1. Physicochemical characteristics of the sampling site

Samples	pH	EC (mS/cm)	Phosphorus (PPM)	Nitrate Nitrogen (PPM)	Chromium (PPM)	Cadmium (PPM)	Lead (PPM)
Control	6.84	16.25	0.36	22.82	33.62	0.11	12.74
NW11-1	8.64	12.27	3.33	11.80	74.38	1.67	31.20
NW11-5	8.89	11.16	3.12	12.96	54.80	2.05	33.25
NW11-10	8.60	11.16	1.25	11.22	82.50	2.05	35.15
NW11-13	8.41	13.92	2.79	19.47	52.87	0.23	22.13
NW11-15	8.58	10.14	4.90	16.10	57.35	0.33	22.23
NW11-20	7.86	0.48	3.12	18.08	84.32	0.23	22.43
NW11-17	8.44	5.58	4.90	17.85	85.50	3.55	34.15

3.2 Community structure (16s based)

Community structure included alpha and beta diversity analysis.

3.2.1 Alpha Diversity

Alpha diversity was evaluated using Shannon, Chao1, and Simpson indices. The mean ± standard deviation for each index was calculated across sample groups. Kruskal–Wallis rank-sum tests revealed no significant differences in alpha diversity among groups for Shannon and Chao1 indices ($p = 0.4289$; Table 2), indicating comparable within-sample diversity across the studied groups.

Table 2. Alpha diversity indices (Shannon, Chao1, and Simpson) summarized as mean ± standard deviation across sample groups. Statistical significance was assessed using the Kruskal–Wallis rank-sum test.

Indices	Mean± SD	P Value
Shannon	6.23 ± 0.46	0.4289
Chao1	1326.17 ± 644.21	0.4289
Simpson	0.997 ± 0.002	0.4289

Note P-values were calculated using the Kruskal–Wallis rank-sum test ($\chi^2 = 7$, df = 7).

3.2.2 Microbial community profiling

Although alpha diversity differences among groups were not statistically significant (Kruskal-Wallis test, p=0.429), descriptive trends in richness and evenness were observed. The control, NW11.1, NW11.5, and NW11.10 samples tended to show comparatively lower richness metrics relative to NW11.15 and NW11.20. However, these differences did not reach statistical significance. Chao1 and ACE richness estimators were consistently higher than the observed species counts across all samples, indicating the potential presence of undetected taxa. Shannon diversity values approached 7 in NW11.15 and NW11.20, suggesting relatively high species richness and evenness in these samples. In contrast, the control and NW11.1 samples exhibited comparatively lower Shannon values, though variability across groups was modest. Simpson indices were close to 1 in all samples, indicating low dominance and generally even community structure across the dataset. Minor variations in inverse Simpson values were observed, with NW11.15 and NW11.20 displaying higher values relative to other groups. However, because statistical comparisons were not significant, these patterns should be interpreted as descriptive rather than indicative of strong ecological differentiation. Good's coverage values ranged between 97.90% and 99.99%, indicating that sequencing depth was sufficient to capture the majority of bacterial diversity present in the reservoir core samples (Kim et al., 2021). Taxonomic profiling of the 16S rRNA amplicon dataset revealed that *Proteobacteria* dominated all samples, with *Firmicutes, Cyanobacteria, and Bacteroidota* present at varying relative abundances. At the class level, *Gammaproteobacteria* were predominant, while *Clostridia* and *Bacilli* were also represented, reflecting the presence of metabolically diverse and potentially stress-tolerant taxa. At finer taxonomic resolution, orders such as *Halobacterales*, *Burkholderiales*, and *Oceanospirillales* were detected across samples, although relative abundances varied. Family-level patterns indicated representation of lineages including *Marinilabiliaceae*, *Izemoplasmataceae*, *Halanaerobiaceae*, and *Aeromonadaceae*. Genus-level analysis identified hydrocarbon-associated taxa such as *Achromobacter*, *Aeromonas, Marinobacter*, *Alcanivorax*, and *Halorubrum*. Several species affiliated with petroleum-associated genera (e.g., *Alcanivorax venustensis*, *Idiomarina aquatica*, *Pseudidiomarina homiensis*) were detected. While these taxa are known from hydrocarbon-impacted environments, functional capabilities inferred from 16S data remain predictive and were further examined using genome-resolved shotgun metagenomic analysis. Overall, taxonomic variation across samples suggests environmental filtering and reservoir-specific microbial adaptation. However, given the absence of statistically significant alpha diversity differences, these compositional shifts likely reflect community restructuring rather than large-scale changes in overall diversity.

3.2.3 Beta Diversity

The beta diversity patterns in relation to genera and species (Figure 1) revealed the patterns of which species co-occur. Some of the environmental factors that are known to cause variations in beta diversity are related to oil pollution. Indeed, studies have shown that bacterial genetic diversity in soils with crude oil pollution is affected, with microbiomics analyses of this microbiome (Vasanthrao et al., 2025). Different PCoA axes (Figure 1) suggest different variances. Clustering clearly shows NW11-17 and NW11-20. NW11-1 and NW11-5 tend to cluster together, implying that these samples are much more alike than different. Similarly, NW11-10 and NW11-15 cluster together, as do NW11-13 and the control sample. In part b, NW11-17 and NW11-20 cluster, whereas NW11-10, NW11-13, and NW11-15 are separated from each other, indicating stark contrasts. Compared with the other plots, the control sample plots exhibit very similar patterns for NW11-1 and NW11-5. The clumped dots on the plot illustrate species turnover, according to how the species composition changes in different locations or under different conditions (Madariaga-Troncoso et al., 2025).

3.3 Genome-Resolved Functional Potential (Shotgun MAGs)

3.3.1 Metagenome binning and Genome Reconstruction Statistics

Metagenome reconstruction statistics (Table 3) contains the information of MAGs and their threshold criteria.

Table 3. Summary of MAG Recovery and Quality Assessment

Metric	Value
Total MAGs recovered	402
High Quality MAGs	216
Medium Quality MAGs	133
Low Quality MAGs	53
High-quality criteria	≥90% completeness, ≤5% contamination
Medium-quality criteria	≥50% completeness, ≤10% contamination

3.3.2 Taxonomic composition of High Quality MAGs

Based on GTDB classification, the 216 high-quality MAGs were distributed across multiple bacterial phyla (Table 4), with clear dominance of *Proteobacteria* and *Actinobacteriota.*

Table 4. Approximate Phylum Level Distribution

Phylum	**Representation (%)**	**Representative Genera**
Proteobacteria	~60-65%	*Pseudomonas, Alcanivorax, Roseovarius, Marinobacter, Paracoccus, Halomonas*
Actinobacteriota	~15-18%	*Streptomyces, Nocardioides*
Bacteroidota	~5-8%	*Wenzhouxiangella*
Planctomycetota	~4-6%	*Gimesia, Phycisphaera*
Verrucomicrobiota	~3-5%	*Mucisphaera*
Other minor phyla	~<5%	Candidate and Uncultured taxa

MAG quality report (Table 5) of the species involved in hydrocarbons degradation, survive in low oxygen conditions and can tolerate high temperature contains the information of completeness, contamination, GC content, Genome size and Contig N50.

Table 5. MAG quality metrics for selected species involved in hydrocarbons degradation, survive in low oxygen conditions and can tolerate high temperature

MAGs ID	GTDB Species	Completeness (%)	Contamination (%)	Contig N50	Genome Size (bp)	GC Content (%)
NW11-2_bins_40	*Azospirillum thermophilum*	98	2.96	26197	3101119	70
NW11-5_bins_54	*Thermohalobaculum* sp.	93.69	3.48	10125	3244858	71
NW11-3_bins_32	*Alcanivorax* sp.	99.94	3.83	123454	4313987	59
NW11-6_bins_52	*Pseudomonas aeruginosa*	99.99	1.24	87742	4305471	60
NW11-4_bins_8	*Streptomyces* sp.	94.84	1.06	33057	3022365	70
NW11-3_bins_20	*Erwinia* sp.	98.89	3.31	270164	3580414	69

These findings suggest that multiple taxa i.e., *Pseudomonas*, *Marinobacter*, *Roseovarius* and *Pseudomonas* are reported to contribute to nitrate reduction with potential role in the reservoir environment, with *Pseudomonas* likely playing a dominant role.

Among the prominent thermotolerant taxa, *Thermohalobaculum* spp., *Azospirillum thermophilum*, *Pseudomonas putida*, *Alcanivorax* spp., *Streptomyces sp., Erwinia sp.* and *Spiribacter* represent strong candidates for MEOR applications. These genomes encode traits consistent with thermotolerance and facultative anaerobic metabolism.

4. DISCUSSION

This study provides one of the first genome-resolved functional insights into microbial communities inhabiting deep subsurface petroleum cores from Pakistan using Shotgun metagenomic sequencing.

Microbial communities of 16s amplicon and shotgun analysis from hydrocarbon-contaminated soil exhibited apparent patterns of taxonomic dominance, stress tolerance ability and metabolic specialization according to structural and

functional profiling, which revealed extreme environmental filtering by crude oil exposure and physicochemical extremes. Alpha diversity analysis, which includes the Shannon, Chao1, and Simpson indices, did not reveal statistically significant differences ($p = 0.4289$), although numerical differences were observed. A comparison of the control samples with the other isolates, such as NW11-10, NW11-13, and NW11-20, revealed that the control samples presented reduced Shannon diversity, highlighting the selective enrichment of hydrocarbon-adapted taxa in contaminated environments, whereas the other sample groups presented high Shannon diversity, indicating that the presence of multiple metabolic functional groups maintained relatively even abundances. A study of Chao1 richness confirmed this trend, which suggests greater species richness in oil-exposed samples because of the low abundance of microorganisms but functionally specialized taxa of microorganisms that contribute to the breakdown of hydrocarbons (Obayori et al., 2024). The results of this study are consistent with global reports indicating that *Proteobacteria* are the most metabolically adaptable responders to hydrocarbon stress, whereas this study revealed that *Proteobacteria* was the dominant phylum across all the samples, followed by *Firmicutes*, *Bacteroidota*, *Actinobacteriota*, and *Halanaerobiota*. *Gammaproteobacteria* and *Alphaproteobacteria* are highly abundant in this phylum, showing drastic shifts towards taxa characterized by alkane degradation, stress tolerance, and biofilm production (Liu et al., 2024; Okafor et al., 2022). The results of the present study revealed the most prevalent genera with several uncultivable bacterial groups, e.g., *Alcanivorax*, *Marinobacter*, *Halomonas*, *Idiomarina*, *Pseudidiomarina*, and *Achromobacter*. These genera can withstand high temperatures and salinities and are essential for the breakdown of hydrocarbons and the synthesis of biosurfactants. *Alcanivorax venustensis*, *Idiomarina aquatica*, *Rhodococcus ruber*, and *Pseudomonas putida*, which are metabolically diverse uncultured organisms, were highlighted by species-level analysis and contribute significantly to the enzymatic functions involved in petroleum compound degradation. The same species distributions were documented in another study of Shotgun metagenomes from oil-contaminated marine and soil environments (Góngora et al., 2024; Peng et al., 2024). Principal coordinate analysis (PCoA) was performed via weighted and unweighted UniFrac distance matrices to separate the bacterial communities according to their physiological role and exposure level. Modest but biologically significant variations in community composition were indicated by the first PCoA, which accounted for 18.5% and 15% of the total variation, whereas the second analysis explained 81.3% of the variation along the first axis and 14.3% along the second axis, which was a greater proportion of the variability and highlighted a significant change in microbial structure (Chen et al., 2024; Mafiana et al., 2021).

Future implications

The taxonomic profile if combined with the functional profiles could describe microbial communities with traits relevant to microbial enhanced oil recovery. MEOR anaerobic and facultative metabolic pathways may allow microbial activity in oxygen-limited environments typical of subsurface oil formations. Additionally, functional redundancy and syntrophic interactions among microbial populations could enhance community stability and maintain biodegradation efficiency, key factors in MEOR applications.

5. CONCLUSION

This work expands current knowledge by providing rare shotgun metagenomic characterization of deep subsurface petroleum cores from Pakistan. Metagenomic studies of thermotolerant microbial communities from deep subsurface environments revealed strong adaptive capacities to hydrocarbon contamination. These results support the possible use of such microbes in environmentally friendly oil recovery and bioremediation approaches. Additionally, the presence of many hypothetical proteins suggests previously unknown enzymatic function reservoirs that require more scientific research. Cultivating representative isolates and developing engineered microbial consortia can be prioritized in the future to facilitate practical and large-scale implementation. This study provides ecological novelty by revealing previously uncharacterized genome-level of thermotolerant deep-reservoir microbial communities to combined thermal, chemical, and oxygen stress. Future work should focus on experimental validation through controlled cultivation, functional expression studies, and reservoir-simulated experiments to confirm in situ activity and recovery efficiency.

Acknowledgements

We acknowledge the “Oil and Gas Development Company Limited” for providing us the raw petroleum reservoir samples” and also for sharing the field issues related to recovery of residual petroleum.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding Statement

Authors have received funding for the analysis of samples but there is no available funding for publishing the data.

Author Information
Authors and Affiliations
Department of Bioinformatics and Biosciences, Faculty of Health and Life Sciences, Capital University of Science and Technology, Islamabad, 44000, Pakistan.

Contributions
Both authors contributed to the conception and design of the study. Conceptualization, hypothesis design and project design was done by A.A. Formal analysis was done by S.T. The first draft of the manuscript was written by S.T. Feedback on previous versions was provided by A.A. Both authors read and approved the final manuscript.

Corresponding author
Correspondence to Arshia amin (arshiaamin.butt@gmail.com)

Availability of data
The sequences used for metagenomic analysis were deposited at NCBI under Bioproject PRJNA1094864- SRP499294 accession numbers SRR285285214–SRR28528521.

Declaration of generative AI in scientific writing
We declare that the generative AI was only used at certain instances as supportive tool. Similarity index with AI is also checked and is within limits.

REFERENCES
1. Chen, Y. J., Altshuler, I., Freyria, N. J., Lirette, A., Góngora, E., Greer, C. W., & Whyte, L. G. (2024). Arctic's hidden hydrocarbon degradation microbes: investigating the effects of hydrocarbon contamination, biostimulation, and a surface washing agent on microbial communities and hydrocarbon biodegradation pathways in high-Arctic beaches. *Environ Microbiome*, *19*(1), 81. https://doi.org/10.1186/s40793-024-00626-w
2. Child, H. T., Wierzbicki, L., Joslin, G. R., & Tennant, R. K. (2024). Comparative evaluation of soil DNA extraction kits for long read metagenomic sequencing. *Access Microbiol*, *6*(9). https://doi.org/10.1099/acmi.0.000868.v3
3. Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep*, *11*(1), 3030. https://doi.org/10.1038/s41598-021-82726-y
4. Frey, B., Varliero, G., Qi, W., Stierli, B., Walthert, L., & Brunner, I. (2022). Shotgun Metagenomics of Deep Forest Soil Layers Show Evidence of Altered Microbial Genetic Potential for Biogeochemical Cycling. *Front Microbiol*, *13*, 828977. https://doi.org/10.3389/fmicb.2022.828977
5. Góngora, E., Lirette, A. O., Freyria, N. J., Greer, C. W., & Whyte, L. G. (2024). Metagenomic survey reveals hydrocarbon biodegradation potential of Canadian high Arctic beaches. *Environ Microbiome*, *19*(1), 72. https://doi.org/10.1186/s40793-024-00616-y
6. Kim, J. W., Hong, Y. K., Kim, H. S., Oh, E. J., Park, Y. H., & Kim, S. C. (2021). Metagenomic Analysis for Evaluating Change in Bacterial Diversity in TPH-Contaminated Soil after Soil Remediation. *Toxics*, *9*(12). https://doi.org/10.3390/toxics9120319
7. Kumar, K. R., Cowley, M. J., & Davis, R. L. (2024). Next-Generation Sequencing and Emerging Technologies. *Semin Thromb Hemost*, *50*(7), 1026-1038. https://doi.org/10.1055/s-0044-1786397
8. Li, Y. Q., Xin, Y., Li, C., Liu, J., & Huang, T. (2023). Metagenomics-metabolomics analysis of microbial function and metabolism in petroleum-contaminated soil. *Braz J Microbiol*, *54*(2), 935-947. https://doi.org/10.1007/s42770-023-01000-7
9. Liu, Q., He, W., Zhang, W., Wang, L., & Tang, J. (2024). Metagenomic analysis reveals the microbial response to petroleum contamination in oilfield soils. *Sci Total Environ*, *912*, 168972. https://doi.org/10.1016/j.scitotenv.2023.168972
10. Liu, S., Rodriguez, J. S., Munteanu, V., Ronkowski, C., Sharma, N. K., Alser, M., Andreace, F., Blekhman, R., Błaszczyk, D., Chikhi, R., Crandall, K. A., Della Libera, K., Francis, D., Frolova, A., Gancz, A. S., Huntley, N. E., Jaiswal, P., Kosciolek, T., Łabaj, P. P., Łabaj, W., Luan, T., Mason, C., Moustafa, A. M., Muralidharan, H. S., Mutlu, O., Mansouri Ghiasi, N., Rahnavard, A., Sun, F., Tian, S., Tierney, B. T., Van Syoc, E., Vicedomini, R., Zackular, J. P., Zelikovsky, A., Zielińska, K., Ganda, E., Davenport, E. R., Pop, M., Koslicki, D., & Mangul, S. (2025). Analysis of metagenomic data. *Nat Rev Methods Primers*, *5*. https://doi.org/10.1038/s43586-024-00376-6
11. Madariaga-Troncoso, D., Vargas, I., Rojas-Villalta, D., Abanto, M., & Núñez-Montero, K. (2025). Metataxonomics Characterization of Soil Microbiome Extraction Method Using Different Dispersant Solutions. *Microorganisms*, *13*(4). https://doi.org/10.3390/microorganisms13040936

12. Mafiana, M. O., Kang, X. H., Leng, Y., He, L. F., & Li, S. W. (2021). Petroleum contamination significantly changes soil microbial communities in three oilfield locations in Delta State, Nigeria. *Environ Sci Pollut Res Int*, *28*(24), 31447-31461. https://doi.org/10.1007/s11356-021-12955-1
13. Mirete, S., Sánchez-Costa, M., Díaz-Rullo, J., González de Figueras, C., Martínez-Rodríguez, P., & González-Pastor, J. E. (2025). Metagenome-Assembled Genomes (MAGs): Advances, Challenges, and Ecological Insights. *Microorganisms*, *13*(5). https://doi.org/10.3390/microorganisms13050985
14. Nam, N. N., Do, H. D. K., Loan Trinh, K. T., & Lee, N. Y. (2023). Metagenomics: An Effective Approach for Exploring Microbial Diversity and Functions. *Foods*, *12*(11), 2140. https://www.mdpi.com/2304-8158/12/11/2140
15. Obayori, O. S., Adesina, O. D., Salam, L. B., Ashade, A. O., & Nwaokorie, F. O. (2024). Depletion of hydrocarbons and concomitant shift in bacterial community structure of a diesel-spiked tropical agricultural soil. *Environ Technol*, *45*(25), 5368-5383. https://doi.org/10.1080/09593330.2023.2291421
16. Okafor, C., Chikere, C., Akaranta, O., & Ntushelo, K. (2022). Crude oil hydrocarbons' effect on soil microbial metagenome from Niger Delta polluted soils. *F1000Research*, *11*, 1108. https://doi.org/10.12688/f1000research.124785.1
17. Pacwa-Płociniczak, M., Biniecka, P., Bondarczuk, K., & Piotrowska-Seget, Z. (2020). Metagenomic Functional Profiling Reveals Differences in Bacterial Composition and Function During Bioaugmentation of Aged Petroleum-Contaminated Soil. *Front Microbiol*, *11*, 2106. https://doi.org/10.3389/fmicb.2020.02106
18. Peng, L., Hou, J., Zhang, Y., Wang, B., Zhang, Y., Zhao, K., Wang, Q., Christie, P., Liu, W., & Luo, Y. (2024). Metagenomic analysis of a thermophilic bacterial consortium and its use in the bioremediation of a petroleum-contaminated soil. *Chemosphere*, *360*, 142379. https://doi.org/10.1016/j.chemosphere.2024.142379
19. Pérez-Cobas, A. E., Gomez-Valero, L., & Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microb Genom*, *6*(8). https://doi.org/10.1099/mgen.0.000409
20. Salam, L., & Obayori, O. (2019). Structural and functional metagenomic analyses of a tropical agricultural soil. *Spanish Journal of Soil Science*, *9*, 1-23. https://doi.org/10.3232/SJSS.2019.V9.N1.01
21. Vasanthrao, R., Nidhin, I. K., Taj, Z., & Chattopadhyay, I. (2025). Comprehensive whole metagenomics analysis uncovers microbial community and resistome variability across anthropogenically contaminated soils in urban and suburban areas of Tamil Nadu, India. *Front Microbiol*, *16*, 1649872. https://doi.org/10.3389/fmicb.2025.1649872
22. Wu, D., Seshadri, R., Kyrpides, N. C., & Ivanova, N. N. (2025). A metagenomic perspective on the microbial prokaryotic genome census. *Sci Adv*, *11*(3), eadq2166. https://doi.org/10.1126/sciadv.adq2166
23. Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front Microbiol*, *12*, 766364. https://doi.org/10.3389/fmicb.2021.766364

Figure Captions

Fig 1. Principal coordinates analysis (PCoA) of microbial community composition based on (a) weighted UniFrac and (b) unweighted UniFrac distance matrices. Axis 1 and Axis 2 explain 18.5% and 15.0% of the variation in the weighted UniFrac analysis and 81.3% and 14.3% of the variation in the unweighted UniFrac analysis, respectively. Samples that cluster closer together share more similar phylogenetic community structures, whereas greater separation indicates increased differences in microbial community composition.

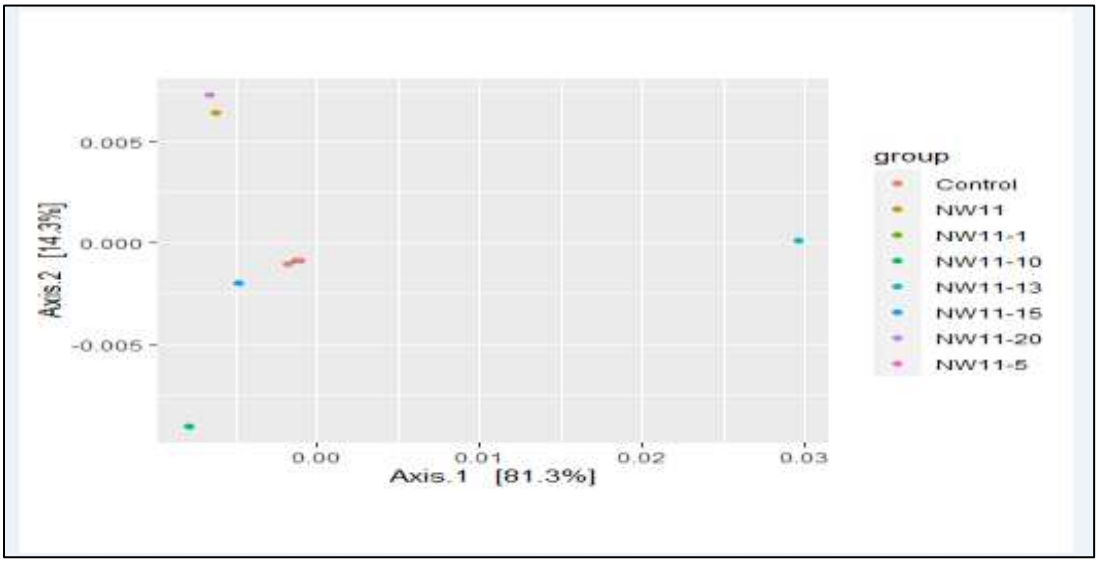
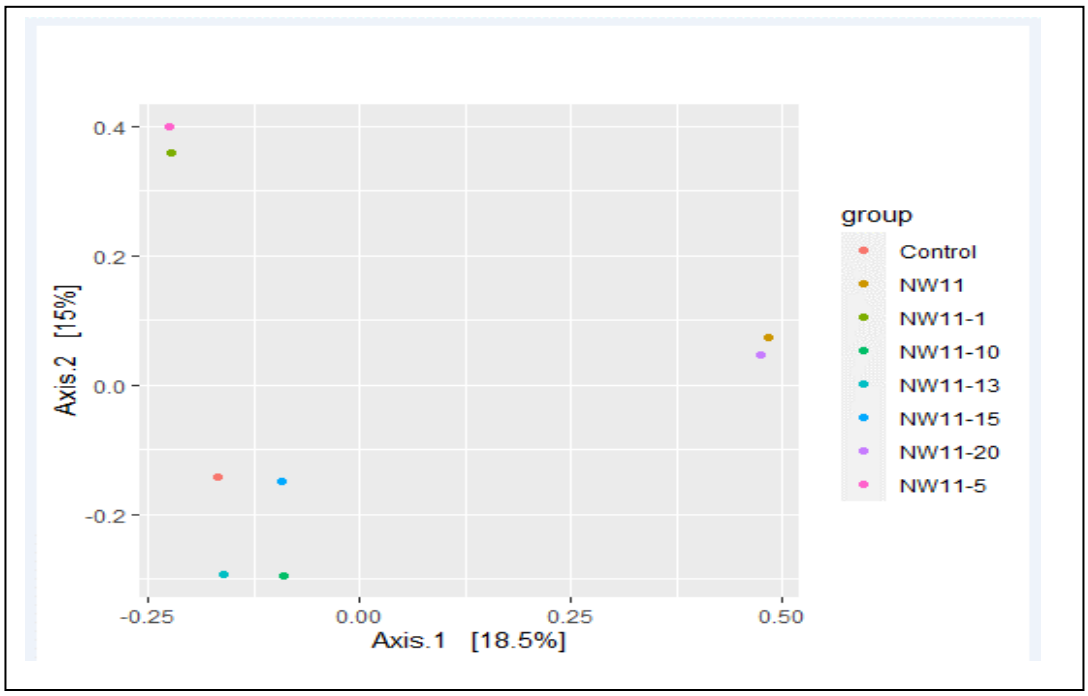


Fig 1. Principal



coordinates analysis (PCoA) of microbial community composition based on (a) weighted UniFrac and (b) unweighted UniFrac distance matrices. Axis 1 and Axis 2 explain 18.5% and 15.0% of the variation in the weighted UniFrac analysis and 81.3% and 14.3% of the variation in the unweighted UniFrac analysis, respectively. Samples that cluster closer together share more similar phylogenetic community structures, whereas greater separation indicates increased differences in microbial community composition.