

Fungal Population Analysis of Hydrocarbons Contaminated Soil: Samples from Taq-taq Oil Field in Koya City, Kurdistan Region, Iraq

Srwa A. Mohammed^{1†}, Taha J. Omar¹ and Ayad H. Hasan^{1,2}

¹Department of Medical Microbiology, Faculty of Science and Health, Koya University, Danielle Mitterrand Boulevard, Koya KOY45, Kurdistan Region – F.R. Iraq

²Department of Biomedical Sciences, College of Health Technology, Cihan University – Erbil, Kurdistan Region – F.R. Iraq

Abstract—Petroleum is often regarded as one of the environmental hazards that pose the greatest threat to human health. After radiation, petroleum and all of its byproducts and wastes are considered to have the second-worst effect on the environment. Demonstrating fungal microbiomes that flourish on soil heavily polluted by petroleum and moderately contaminated soil samples, comparing them with uncontaminated soil samples from Taq-taq (TPOPCO) through metagenomic analysis through sequencing of the 18S-V4 region. Metagenomic analysis is conducted using high-throughput sequencing technology, targeting 18 subunit ribosomal ribonucleic acid amplicons through the Illumina-HiSeq platform. In general, an increase in fungal community richness and diversity is reported in soil contaminated with petroleum. This is determined by counting the number of operational taxonomic units, performing principal coordinate analysis, and calculating α (Chao1 and Shannon indices) and β diversity. The composition of microbial communities is significantly altered by crude oil exposure. At the phylum level, there are considerable transitions between groups B and C for Ascomycota, Basidiomycota, Chytridiomycota, Olpidiomycota, Zoopagomycota, Cryptomycota, and Mucoromycota. In examining Group D relative to Group C, there are significant differences in Ascomycota, Basidiomycota, Chytridiomycota, Olpidiomycota, and Cryptomycota. This study is an important first step in determining and understanding the fungal population of soil extensively contaminated with crude oils of the Taq-taq/Kurdistan Region of Iraq.

Index Terms—Fungal community, Metagenomics, Operational taxonomic unit number, Statistical analysis, V4 18S rRNA gene.

I. INTRODUCTION

Contaminated soil resulting from spilled oil is one of the most serious ecological disasters, adversely affecting both marine

and terrestrial ecosystems and incurring substantial economic losses. Moreover, oil spills are among the most prevalent natural catastrophes. Due to the toxicological properties of the oil's components, oil spills provide possible health risks to both cleanup personnel and coastal people (Laffon, et al., 2016). The Kurdistan Region of Iraq (KRI) is located north and north-east of the Arabian plate. The region is one of the oil-rich areas of Iraq (Shlimon, et al., 2020). One of the most significant fields in this region is the Taq-Taq oil field. As a result of factors such as an increasing worldwide demand, an expanding population, and widespread consumption of petroleum products, the petrochemical sector has been expanding at a steady rate and is continuing to contaminate both marine and terrestrial habitats (Ramadass, et al. 2018; Baoune, et al. 2019). Microorganisms use hydrocarbons as their only carbon source or modify them to reduce their toxicity. Most pathways in microbial hydrocarbon degradation transpire in aerobic circumstances; fungi possess the capability to break down polycyclic aromatic hydrocarbons. Bacteria metabolize hydrocarbons through specialized pathways, including alkane monooxygenase and dioxygenase, whereas fungi employ various hydrocarbons through non-specific enzyme complexes, such as cytochrome P450, lignin peroxidase, manganese peroxidase, and laccase that assist them to decompose lignin as well as cellulose (Asemoloye, et al., 2020). Molecular approaches to studying biodiversity may also be analyzed according to the kind of nucleic acid that was extracted (DNA or RNA) or using an analytical technique that is based on either partial or entire community analysis (Wydro, 2022). Microbial community analysis of contaminated soil with crude oils involves the study of the diverse array of microorganisms present in the soil and their response to oil contamination. Using advanced molecular techniques such as next-generation sequencing, researchers can identify and quantify the different microbial species present in the soil (Bonomo, et al. 2022). The field of metagenomics has revolutionized our understanding of microbial diversity by allowing researchers to investigate the vast array of non-cultivable bacteria and fungi that thrive in diverse environmental settings. Traditional cultivation

ARO-The Scientific Journal of Koya University
Vol. XII, No. 2 (2024), Article ID: ARO.11745. 8 pages
DOI: 10.14500/aro.11745

Received: 03 August 2024; Accepted: 03 December 2024
Regular research paper; Published: 25 December 2024

[†]Corresponding author's e-mail: srwa.ali@koyauniversity.org
Copyright © 2024 Srwa A. Mohammed, Taha J. Omar and Ayad H. Hasan. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-SA 4.0).



methods often fail to capture the full microbial diversity present in the environment. Metagenomics overcomes this limitation by directly sequencing the DNA of uncultured microorganisms, revealing valuable genetic information that can be harnessed for novel biotechnological applications (Chandran, Meena and Sharma, 2020).

In this study, our objective was to characterize and introduce the original surface fungal microbiome aerobically that was associated with crude oils and compare it with clean or control samples of the Taq-TaQ (TTOPCO) oil field soil of the Kurdistan Region/Iraq through metagenomic analysis through sequencing of the 18S-V4 region. This provides valuable contributions to environmental restoration.

II. MATERIALS AND METHODS

A. Sample Collection

Nineteen soil oil samples were collected from the asphalt seep at Koya City Taq Taq Operating Company (TTOPCO) in the KRI. Oil-contaminated soil samples were taken from three places in the seep on October 13, 2020, and a control soil sample was taken from the uncontaminated region. Table I lists the sample sites' coordinates and related attributes. The samples were taken at 5–10 cm deep to obtain representative soil profiles for examination. To ensure proper handling and preservation, all the collected samples were carefully placed into sterilized polyethylene bags.

B. Fungal Detection and Identification from Soil Samples using Molecular Analysis

Nineteen soil samples were extracted for whole DNA using the DNeasy® Power Soil® Pro Kit (LOT 169043419/Qiagen-Germany). Add 250 mg of soil in each PowerBead Pro tube, and add 800 µL of solution CD1 (lysis buffer) in the PowerBead Pro tubes. Then, increasing the vortex time to 20 min and letting the mixture sit at room temperature for an hour, then following the kit protocol, after following all the steps, the eluted DNA was quantified using a Nanodrop spectrophotometer (Nano Drop Spectro 117 432-UK) for evaluating the concentration of DNA. For metagenomic analysis, total DNA that was extracted from all nineteen soil samples was sequenced at BGI Sequencing Centre Hong Kong/China. The outcomes of library construction were displayed. Eighteen samples were passed successfully with 50,000 tag numbers for 18S-V4 with SILVA_18S. 30 ng of qualified DNA template and the 18S rRNA fusion primers are added for polymerase chain reaction (PCR). All PCR

products are purified by Agencourt AMPure XP beads. These beads were dissolved in the elution buffer, and the resulting mixture was used to complete the library construction process. Subsequently, the library size and concentration were determined using the Agilent 2100 Bioanalyzer. Only libraries that met the required criteria were selected for sequencing on the HiSeq 2500 platform (PE300), based on their insert size.

C. Bioinformatics Analysis Workflow

Raw data were filtered to generate high-quality clean reads (Fadrosh, et al., 2014) with truncated reads whose average phred quality values are lower than 20. Then, those whose lengths are 75% of their original lengths after truncation, contaminated by adapter sequences (Martin, et al. 2011), ambiguous bases (N bases), and low-complexity reads should be removed. Paired-end reads were merged using Fast Length Adjustment of Short reads (v1.2.11), with a minimum overlapping length of 15 bp and a mismatching ratio of the overlapped region of ≤ 0.1 (Martin, et al., 2011). After which clean reads that can overlap with each other are merged to tags and further clustered to Operational Taxonomic Units (OTU) with USEARCH (v7.0.1090): OTU are a unified marker for analyzing a taxon unit (seven taxonomy levels) (kingdom, phylum, class, order, family, genus, and species). Software used for statistics: USEARCH (v7.0.1090) (Edgar 2013) and UCHIME (v4.2.40) (Edgar 2011).

III. RESULTS AND DISCUSSIONS

The following fungus genera were discovered by molecular techniques using the same soil samples used for metagenomic analysis: Group A included *Aspergillus lentulus* 28S ribosomal RNA (accession number XR 004500616.1). Group B included *Aspergillus fellis* strain FM324 chromosome 3 (accession number CP066505.1), *Aspergillus luteonubrus* strain MST FP2246 (accession number MT196912.1), and *Aspergillus arizonicus* isolate CCF 5341 (accession number OK321187.1). Group C included *Rhizopus arrhizus* strain SC49B03 (accession number MW113537.1) (Srwa, Taha and Ayad, 2023).

A. Taxonomic Profiling of Fungal Population

The whole DNA extraction was carried out on each sample on an individual basis. It was found through the creation and analysis of 18S rRNA sequencing libraries that the entire complete genome DNA was used for library

TABLE I
SOIL SAMPLES AND SAMPLING LOCATIONS

| Groups of soil samples | Given numbers | Location of the samples | Coordinate (°N, °E) |
|------------------------|----------------|--|--|
| A | 1,2,3,4 | Oil and water mixed samples were collected from an area close to the drilled pool of oil at the Taq-taq oil field. | 35° 40' 33" N 44° 31' 30" E (Sardar and Dler, 2017) |
| B | 5,6,7,8,9 | Underlying and flanking region of the Taq-taq asphalt seep flow. | |
| C | 10,11,12,13,14 | Control: Ten meters away from site Group B, it had not been contaminated by oil spills. | |
| D | 15,16,17,18,19 | Transportation area where the soil was contaminated with spilt oil was 20 m away from site number two. | |

construction, and the findings showed that every single one of the 18 samples was able to pass profitably. Then, raw data are filtered, clean data are created, after which clean data are merged with tags and further clustered to OTU. The preprocessing of the 18S amplicon sequences produced a total of 1296 OTUs. Table II, which were comprised of 1163942 paired reads, respectively, and with the connect tag number ranging from 69129 to 64147 Table III, which included the control samples as well as the results, showed that for Group A soil samples, there were 110 OTU numbers with 137189 tag numbers out of three subsamples, 524 OTU numbers for Group B soil samples with 246320 tag numbers out of five subsamples, the soil samples from Group C have 188 OTU numbers and 249182 tag numbers out of five subsamples, whereas the soil samples from Group D have 447 OTU numbers and 248661 tag numbers out of five subsamples, as shown in Table II. The soil sample from Group B, which had been long-term contaminated with crude oil, had a higher number of OTUs when compared to both the clean soil samples used as a control and other crude oil-contaminated samples. While the soil sample from Group C, which was uncontaminated, had the most diverse tag number when compared to the other groups. And the results showed that out of 475 OTUs, 46 OTUs were related to the kingdom of fungi, while the remaining were other eukaryotes.

TABLE II
OTU STATISTICS

| Sample name | Tag number | OTU number |
|-------------|------------|------------|
| A1 | 49917 | 35 |
| A2 | 39087 | 28 |
| A3 | 48185 | 47 |
| B5 | 49625 | 60 |
| B6 | 49028 | 195 |
| B7 | 49414 | 101 |
| B8 | 48748 | 106 |
| B9 | 49505 | 62 |
| C10 | 49835 | 41 |
| C11 | 49913 | 37 |
| C12 | 49891 | 42 |
| C13 | 49602 | 32 |
| C14 | 49941 | 36 |
| D15 | 49819 | 83 |
| D16 | 49502 | 72 |
| D17 | 49747 | 53 |
| D18 | 49811 | 63 |
| D19 | 49782 | 176 |

TABLE III
ALPHA DIVERSITY RESULTS

| # Alpha | Mean (A) | SD (A) | Mean (B) | SD (B) | Mean (C) | SD (C) | Mean (D) | SD (D) | p-value |
|----------|----------|---------|----------|----------|----------|--------|----------|--------|---------|
| sobs | 36.66 | 9.609 | 104.8 | 54.75126 | 37.6 | 4.037 | 89.4 | 49.661 | 0.005 |
| chao | 43.8 | 15.661 | 127.71 | 69.81464 | 47.45 | 11.023 | 105.835 | 46.765 | 0.011 |
| ace | 51.40 | 24.355 | 128.15 | 66.59791 | 49.400 | 13.044 | 118.056 | 48.220 | 0.021 |
| shannon | 2.01 | 0.367 | 2.92 | 0.42379 | 1.817 | 0.770 | 2.634 | 1.134 | 0.052 |
| simpson | 0.19 | 0.065 | 0.119 | 0.06738 | 0.261 | 0.171 | 0.192 | 0.251 | 0.261 |
| coverage | 0.99 | 0.00009 | 0.999 | 0.0002 | 0.999 | 0.0001 | 0.999 | 0.0001 | 0.170 |

B. OTU Abundance Table/OTU Analysis/OTU Venn Map

The results of the OTU analysis presented on the samples achieved from oil-contaminated soil and clean soil are shown in the following diagrams. Fig. 1 Venn diagram OTU shows specific number of OTU in each group, and samples (A, B, C, and D) were displayed (9, 178, 6, and 108), respectively, while overlapping OTUs for each group sample with other groups (A, B, C, and D) was shown to be 66, 157, 99, and 141, respectively. Thus, the results of the diagram and statistical analysis showed that the region with the largest concentration of OTU 18S fungal communities was the area corresponding to Group B's soil sample, which was a long-term oil-contaminated soil sample. Group D, which consisted of an area that had been contaminated by spilled crude oil, came next after Group B. The findings from the study conducted by Galitskaya, et al. (2021) align closely with the outcomes of the current study. Their investigation demonstrated the utility of Venn diagrams in comparing OTUs across samples or subgroups. These diagrams serve to illustrate the extent of similarity and overlap in OTU compositions among various environmental samples.

As a result, Galitskaya's work indicated that 25% (samples D and C) to 43% (sample S) of fungal OTUs were unique to each sample, with only approximately 45% being consistently present across all samples. This discrepancy could potentially stem from the stronger associations between soil fungi and plant communities compared to the relationships between plants and bacteria within the bulk soil (Mueller, et al. 2014; Sun, et al. 2017).

C. Diversity Analysis across Distinct Locations

Alpha diversity assessment

The evaluation of species diversity within a sample, known as alpha diversity, can be evaluated through various indices, including the observed species index, Chao index, ACE index, Shannon index, Simpson index, and the Good-coverage index. The outcomes of this analysis demonstrated distinct p-values for the 18S fungal community, which were 0.261 for Groups A, B, C, and D combined. Furthermore, a higher value of the Good-coverage index signifies a lower count of undiscovered species in samples among Groups B, C, and D, with p-values of 0.171 as shown in (Fig. 2 and Table IV). In the realm of statistical analysis for alpha diversity, the 18S data indicated peak diversities within soil Group B (14.62) followed by Group D (13.16). The richness value for 18S, soil group B, exhibited greater richness (sobs =

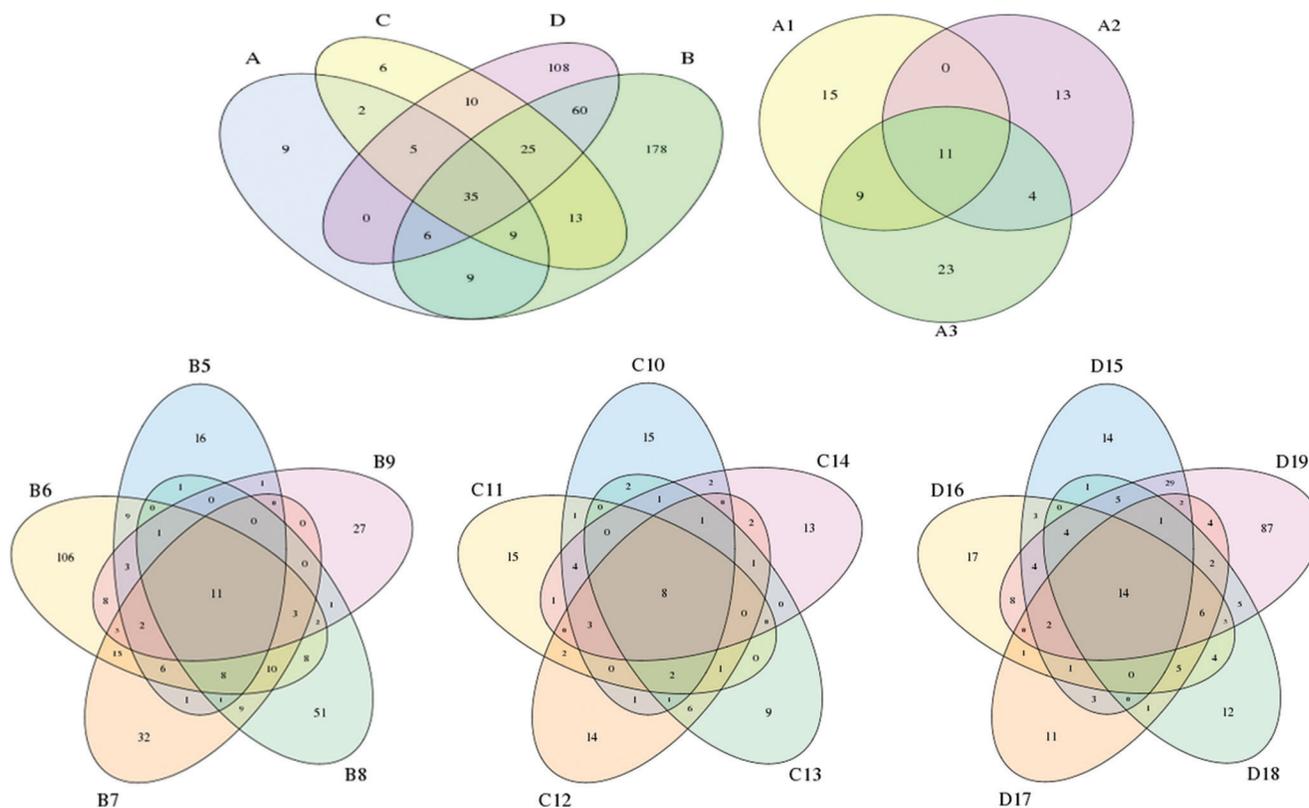


Fig. 1. OTU Venn diagrams.

The Venn diagrams illustrate the distribution of Operational Taxonomic Units (OTUs) among different groups/ samples. Each group/sample is visually distinguished by a unique color fill. The numerical values within the distinct, non-overlapping regions represent the count of OTUs exclusive to each specific group/sample. Conversely, the integers within the intersecting regions indicate the quantity of OTUs that are shared among multiple groups/samples.

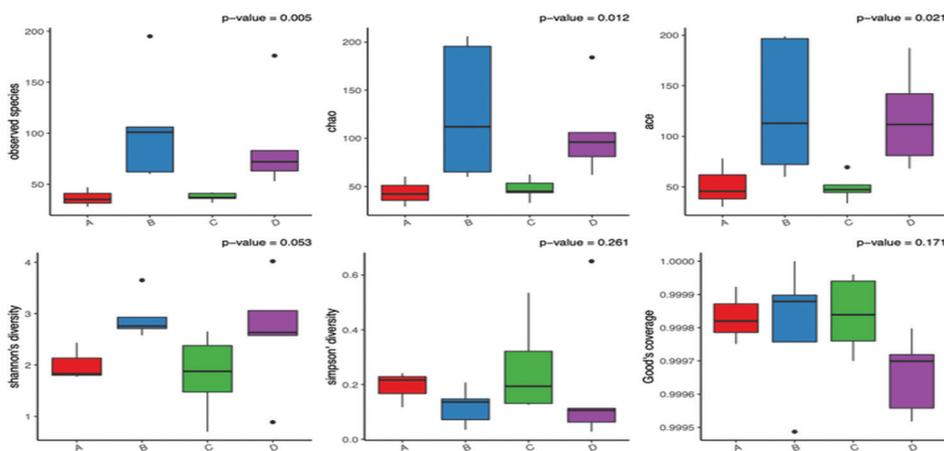


Fig. 2. Boxplot of alpha-diversity.

524.000000). In addition, the Simpson index of diversity showcased its lowest value. The Simpson index's nadir was evident in the subsamples of Group B (B8 = 0.035294) and the subsamples of Group D (D19 = 0.027964), as presented in Table V. The average and standard deviation of alpha diversity were computed for each group. A significance level of $p < 0.05$ indicates noteworthy variations in alpha diversity between at least two groups, underscoring meaningful distinctions in species diversity across these groups. This

outcome potentially suggests that over an extended period of contamination, native microorganisms have acclimatized to thrive within this environment. These microorganisms have seemingly evolved to effectively harness total petroleum hydrocarbons (TPH) as their exclusive carbon and energy source.

Beta diversity analysis

a. Heatmap and Bray-Curtis distance: A heatmap depicting the beta diversity index was generated for 18S datasets.

TABLE IV
TAGS CONNECTION FOR OVERLAPPED PAIRED-END

| Sample name | Total pairs read number | Connect tag number | Connect ratio (%) | Average length and SD |
|-------------|-------------------------|--------------------|-------------------|-----------------------|
| A1 | 68156 | 67440 | 98.95 | 370/21 |
| A2 | 39140 | 39097 | 99.89 | 383/12 |
| A3 | 68278 | 67327 | 98.61 | 383/16 |
| B5 | 50451 | 50393 | 99.89 | 377/13 |
| B6 | 72504 | 72269 | 99.68 | 399/24 |
| B7 | 68407 | 67491 | 98.66 | 380/11 |
| B8 | 62954 | 62279 | 98.93 | 378/9 |
| B9 | 54163 | 54101 | 99.89 | 385/7 |
| C10 | 68233 | 67408 | 98.79 | 377/11 |
| C11 | 68206 | 67309 | 98.68 | 384/6 |
| C12 | 60593 | 60546 | 99.92 | 385/10 |
| C13 | 74302 | 74226 | 99.9 | 381/8 |
| C14 | 68545 | 67677 | 98.73 | 378/4 |
| D15 | 67606 | 66706 | 98.67 | 385/12 |
| D16 | 68171 | 67252 | 98.65 | 386/8 |
| D17 | 68108 | 67193 | 98.66 | 389/4 |
| D18 | 68520 | 67601 | 98.66 | 381/8 |
| D19 | 67605 | 66503 | 98.37 | 391/17 |

TABLE V
STATISTICAL TABLE OF ALPHA DIVERSITY

| Sample name | Sobs | Chao | Ace | Shannon | Simpson | Coverage |
|-------------|------|---------|---------|---------|---------|----------|
| A1 | 35 | 42.2 | 45.678 | 1.770 | 0.216 | 0.999 |
| A2 | 28 | 29 | 30.417 | 1.831 | 0.241 | 0.999 |
| A3 | 47 | 60.2 | 78.109 | 2.436 | 0.117 | 0.999 |
| B5 | 60 | 60 | 60 | 2.757 | 0.147 | 1 |
| B6 | 195 | 195.55 | 196.712 | 2.579 | 0.207 | 0.999 |
| B7 | 101 | 112 | 112.900 | 2.710 | 0.136 | 0.999 |
| B8 | 106 | 206 | 199.028 | 3.649 | 0.035 | 0.999 |
| B9 | 62 | 65 | 72.117 | 2.930 | 0.071 | 0.999 |
| C10 | 41 | 45 | 47.270 | 1.877 | 0.193 | 0.999 |
| C11 | 37 | 53.5 | 51.937 | 1.475 | 0.321 | 0.999 |
| C12 | 42 | 43.5 | 44.410 | 2.652 | 0.131 | 0.999 |
| C13 | 32 | 33 | 33.885 | 2.377 | 0.126 | 0.999 |
| C14 | 36 | 62.25 | 69.500 | 0.703 | 0.534 | 0.999 |
| D15 | 83 | 106 | 111.681 | 2.631 | 0.113 | 0.999 |
| D16 | 72 | 81 | 80.971 | 3.058 | 0.062 | 0.9997 |
| D17 | 53 | 62.1 | 68.121 | 0.889 | 0.650 | 0.999 |
| D18 | 63 | 96 | 142.061 | 2.575 | 0.106 | 0.999 |
| D19 | 176 | 184.076 | 187.447 | 4.019 | 0.027 | 0.999 |

The Bray-Curtis distance, a widely employed index for evaluating dissimilarities between two communities, was employed. This index's range spans from zero to one, with zero indicating an exact similarity in community structure, as illustrated in the accompanying histograms. Among the 18 soil samples analyzed for the 18S dataset, each sample showcased dissimilarity with other subsamples, with values deviating from zero. Notably, soil subsample A1 demonstrated pronounced similarity only with subsamples A3, A2, and B5, while markedly differing from subsamples D16, D17, and D19, as depicted in Fig. 3a.

The research conducted by Galitskaya et al. (2021) employed NMDS analysis of fungal communities in three soil samples contaminated with crude oil, confirming that

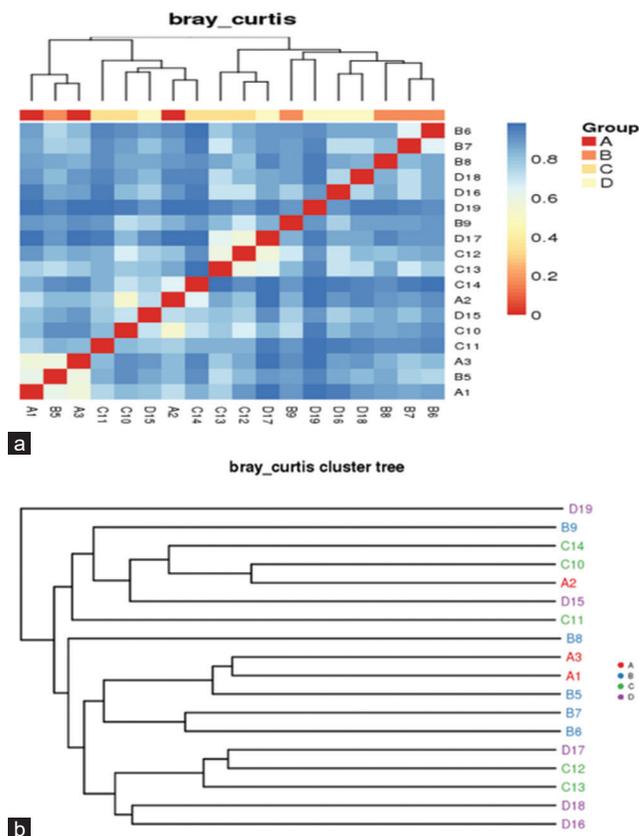


Fig. 3. (a) Heatmap beta-diversity. (b) Sample clustering tree for beta-diversity.

(a) This figure showcases a heatmap depicting the beta diversity indices. A larger index value corresponds to a more pronounced dissimilarity between samples. (b) This figure displays the results of sample clustering analysis. Samples belonging to the same group are depicted with identical colors, and shorter branch lengths or closer distances between samples indicate a higher degree of similarity.

the original fungal communities of the three soils exhibited more significant differences from one another than the bacterial communities and that the fungal communities of the unpolluted soils underwent substantial alterations over time. The primary explanation for the lack of a unified response of the fungal communities in the three distinct soils to petroleum contamination was likely these disparities. Non-metric multidimensional scaling indicated that the successions of fungal communities in contaminated soil varied based on the duration of crude oil contamination.

b. The samples clustering tree displayed the arrangement of the 18 soil samples into four primary clusters representing diverse soil fungal communities. The Bray-Curtis cluster tree indicated the grouping of branches was corresponding to B6, B7, B5, A1, A3, and B8, as well as the clustering of branches corresponding to D17, C12, and C13, and finally, the clustering of branches was corresponding to A2, C10, and C14. Similar fungal populations were observed within these groups, characterized by shorter branch lengths. Conversely, sample D19 stood apart from the fungal communities of the other D samples, as evidenced by its extended branch length, as depicted in Fig. 3b. The study conducted by Gałzka et al.

(2018) demonstrated that soil samples directly obtained from oil wells exhibited the highest levels of biological diversity, as quantified by the Shannon–Weaver index. This discovery aligns with the outcomes of a previous investigation. In a study by Zhang, Ju and Zuo (2018), the metagenomic-based phylogeny of isolated strains predominantly placed them within the Ascomycota phylum, with a few isolates belonging to Basidiomycota.

D. Correlation Analysis/Species-Specific Spearman Coefficients

We conducted a comprehensive investigation of correlations among fungal species (18S), separating connections of significance:

- (1) The *Saccharomyces cerevisiae* group (C) demonstrated the highest abundance of this species. It displayed correlations with *Chaetomium globosum* (1) and *Yarrowia lipolytica* (1), both of which were most abundant in groups C and A, respectively. However, *Saccharomyces cerevisiae* exhibited no correlation with *Pichia kudriavzevii* (0), which were most abundant in groups (C). *Pichia kudriavzevii*, this species showed a notable association with *Yarrowia lipolytica* (1) and *Albugo laibachii* (1), while lacking any connection with *Saccharomyces cerevisiae* (0).
- (2) *Chaetomium globosum*: Significant connections were observed with *Saccharomyces cerevisiae* (0.7) and *Albugo laibachii* (1). It also had a modest correlation with *Pichia kudriavzevii* (0.1) but a negative correlation with *Yarrowia lipolytica* (-0.2), as depicted in Fig. 4. These correlations clarify complicated interplays among species within various groups, providing valuable insights into the ecological dynamics of the studied communities.

E. Species Abundance/Species Composition and Abundance

To brilliantly illustrate the absolute diversity of fungal species within various soil samples, along with their respective composition and distribution, our study employed a multifaceted approach. This included the utilization of species abundance bar plots. This graphical representation provides an insightful visual of the relative abundance of different

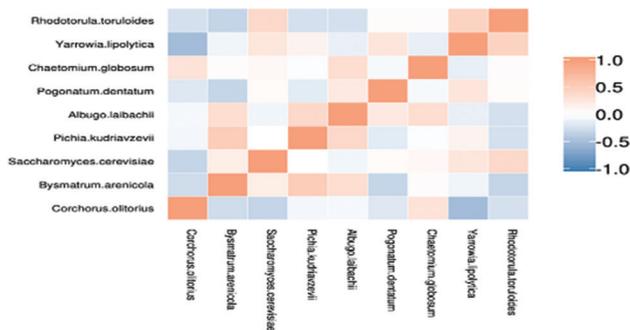


Fig. 4. Species spearman coefficients analysis.

The map shows the correlation between the species level (relative abundance >0.5%). The darker the color, the stronger the correlation between species.

species within the soil samples. Fig. 5 presents an attractive view, showcasing the top 15 most divergent eukaryotic phyla depicted in the annotated phylum barplot. This elucidates the fungal population profiles at the phylum level, highlighting remarkable variations. Our findings reveal noteworthy shifts in fungal phylum abundance between Group B and Group C; we observed significant abundance differences, with shifts in Other (48.47% vs. 38.89%), Ascomycota (2.76% vs. 21.26%), Basidiomycota (3.78% vs. 0.348%), Chytridiomycota (2.193% vs. 0.961%), Olpidiomycota (0% vs. 0.0081%), Zoopagomycota (0.084% vs. 0%), Cryptomycota (0.0154% vs. 0%), and Mucoromycota (0.018% vs. 0%). In exploring Group D in comparison to Group C, we noted substantial differences in Other (58.24% vs. 38.89%), Ascomycota (0.350% vs. 21.26%), Basidiomycota (0.175% vs. 0.348%), Chytridiomycota (0.033% vs. 0.961%), Olpidiomycota (0.357% vs. 0.0081%), and Cryptomycota (0.003% vs. 0%) Table VI.

In support of our fungal microbiome findings, Wang, et al. (2021) analyzed 18S rDNA sequencing data of fungal microbiomes in Ecopile soils. Their analysis revealed that approximately 55% (range 48–64%) of sequence reads could be attributed to the fungal kingdom, while the remaining sequences were assigned to various eukaryotic soil organisms. The majority of fungal reads (an average of 70%) were assigned to the Ascomycota phylum, while other fungal phyla were also identified. At the genus level, they identified 145 different fungal genera, with Ascomycota and Basidiomycota being prominent members of the fungal community.

F. Methods for Differential Species Composition Analysis

In this section, principal component analysis (PCA) was employed for assessing differences in species composition between two sites or two groups. The results revealed a robust correlation between two key variables, which are represented along the X and Y axes as PC1 and PC2, respectively. PC1 and PC2 account for 26.49% and 22.22% of the variance, respectively, Fig. 6. This analysis also delineates discrete taxonomic groups associated with each site, confirming the significant divergence in environmental conditions between Groups A and B, which are horizontally separated from each other. In addition, Group C is notably distant from the other groups, further emphasizing the distinct environmental conditions of Group D when compared to the rest.

These collective studies offer valuable insights into fungal community structures in various environmental contexts, contributing to our understanding of microbiome dynamics and adaptation to contaminants. Numerous studies have consistently reported higher fungal diversity in contaminated soils compared to control soils, as documented in studies by Deshmukh, Khardenavis and Purohit (2016) and Borowik, et al. (2019). These combined findings provide a comprehensive understanding of the fungal community in different soil environments, shedding light on their composition and diversity.

TABLE VI
GROUPS BARPLOT STATISTICAL ANALYSIS

| Taxon | A | A% | B | B% | C | C% | D | D% |
|-----------------|----------|--------|----------|--------|----------|--------|----------|--------|
| Mucoromycota | 0 | 0 | 0.000182 | 0.0182 | 0 | 0 | 0 | 0 |
| Cryptomycota | 0 | 0 | 0.000154 | 0.0154 | 0 | 0 | 0.00003 | 0.003 |
| Zoopagomycota | 0 | 0 | 0.00084 | 0.084 | 0 | 0 | 0 | 0 |
| Olpidiomycota | 0 | 0 | 0 | 0 | 0.000081 | 0.0081 | 0.00357 | 0.357 |
| Chytridiomycota | 0.000262 | 0.0262 | 0.021938 | 2.193 | 0.009619 | 0.961 | 0.000337 | 0.033 |
| Basidiomycota | 0.027334 | 2.733 | 0.037873 | 3.787 | 0.003487 | 0.348 | 0.001753 | 0.175 |
| Ascomycota | 0.204608 | 20.460 | 0.027663 | 2.766 | 0.212647 | 21.26 | 0.003502 | 0.3502 |
| Other | 0.633286 | 63.328 | 0.484723 | 48.47 | 0.388980 | 38.89 | 0.582459 | 58.24 |

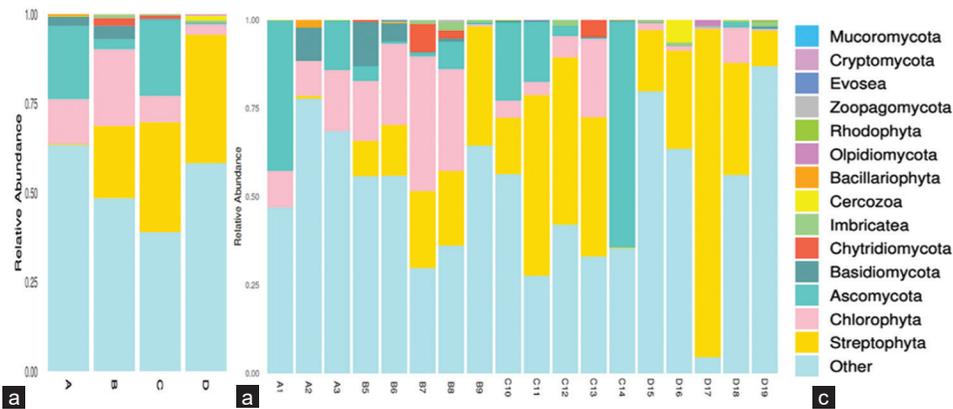


Fig. 5. Phylum abundance barplot (a) groups (b) samples (c) key of a and b.

The GraPhlan map allows us to spatially represent the taxonomic relationships among various species. It visualizes the intricate web of connections between microbial groups, aiding in the understanding of community structure.

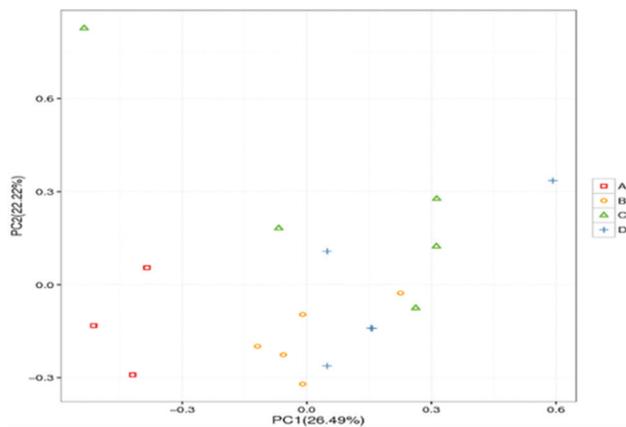


Fig. 6. Species principal component analysis.

Points depicted in various colors and shapes represent samples collected from distinct environments or conditions.

IV. CONCLUSION

Employing high-throughput sequencing technology, we investigated the soil microbial diversity through the analysis of 18S amplicons on the Illumina-HiSeq platform. The results illuminated significant shifts in fungal community structures within long-term crude oil-contaminated soil; this manifested in the emergence of distinct fungal groups. As well as underscores the resilience of soils, even in the face of extensive petroleum contamination. This vital insight holds the potential for exploitation in large-scale oil spill

clean-up initiatives, reaffirming the importance of harnessing the natural capacity of microbial communities to mitigate environmental hazards.

V. ACKNOWLEDGMENT

We would like to thank the staff of the Deanship of Research and Development Center at Koya University for their help and technical support.

REFERENCES

- Asemoloye, M.D., Tosi, S., Daccò, C., Wang, X., Xu, S., Marchisio, M.A., Gao, W., Jonathan, S.G., and Pecoraro, L., 2020. Hydrocarbon degradation and enzyme activities of *Aspergillus oryzae* and *Mucor irregularis* isolated from nigerian crude oil-polluted sites. *Microorganisms*, 8(12), p.1912.
- Baoune, H., Aparicio, J.D., Acuña, A., El Hadj-khelil, A.O., Sanchez, L., Polti, M.A., and Alvarez, A., 2019. Effectiveness of the *Zea mays*-*Streptomyces* association for the phytoremediation of petroleum hydrocarbons impacted soils. *Ecotoxicology and Environmental Safety*, 184, p.109591.
- Bonomo, M.G., Calabrone, L., Scrano, L., Bufo, S.A., Di Tomaso, K., Buongarzone, E., and Salzano, G., 2022. Metagenomic monitoring of soil bacterial community after the construction of a crude oil flowline. *Environmental Monitoring and Assessment*, 194(2), p.48.
- Borowik, A., Wyszowska, J., Kucharski, M., and Kucharski, J., 2019. Implications of soil pollution with diesel oil and bp petroleum with active technology for soil health. *Environmental Research and Public Health*, 16(14), p.2474.

- Chandran, H., Meena, M., and Sharma, K., 2020. Microbial biodiversity and bioremediation assessment through omics approaches, *Frontiers in Environmental Chemistry*, 1, p.570326.
- Deshmukh, R., Khardenavis, A.A., and Purohit, H.J., 2016. Diverse metabolic capacities of fungi for bioremediation. *Indian Journal of Microbiology*, 56(3), pp.247-264.
- Edgar, R.C., 2013. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), pp.996-998.
- Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., and Ravel, J., 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, 2(1), p.6.
- Gałązka, A., Grządziel, J., Gałązka, R., Ukalska-Jaruga, A., Strzelecka, J., and Smreczak, B., 2018. Genetic and functional diversity of bacterial microbiome in soils with long term impacts of petroleum hydrocarbons. *Frontiers in Microbiology*, 9, p.1923.
- Galitskaya, P., Biktasheva, L., Blagodatsky, S., and Selivanovskaya, S., 2021. Response of bacterial and fungal communities to high petroleum pollution in different soils. *Scientific Reports*, 11(1), p.164.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), pp.10-12.
- Mueller, R.C., Paula, F.S., Mirza, B.S., Rodrigues, J.L.M., Nüsslein, K., and Bohannan, B.J.M., 2014. Links between plant and fungal communities across a deforestation chronosequence in the Amazon rainforest. *International Society for Microbial Ecology*, 8(7), pp.1548-1550.
- Ramadass, K., Megharaj, M., Venkateswarlu, K., and Naidu, R., 2018. Bioavailability of weathered hydrocarbons in engine oil-contaminated soil: Impact of bioaugmentation mediated by *Pseudomonas spp.* on bioremediation. *Science of the Total Environment*, 636, pp.968-974.
- Shlimon, A.G., Mansurbeg, H., Othman, R.S., Gittel, A., Aitken, C.M., Head, I.M., Finster, K.W., and Kjeldsen, K.U., 2020. Microbial community composition in crude oils and asphalts from the Kurdistan Region of Iraq. *Geomicrobiology Journal*, 37(7), pp.635-652.
- Srwa, A.M., Taha, J.O., and Ayad, H.H., 2023 Degradation of crude oil and the pure hydrocarbon fractions by indigenous soil microorganisms. *Biologia*, 78(12), pp.3637-3651.
- Sun, S., Li, S., Avera, B.N., Strahm, B.D., and Badgley, B.D., 2017. Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. *Applied and Environmental Microbiology*, 83(14), p.e00966-17.
- Wang, M., Garrido-Sanz, D., Sansegundo-Lobato, P., Redondo-Nieto, M., Conlon, R., Martin, M., Mali, R., Liu, X., Dowling, D.N., Rivilla, R., and Germaine, K.J., 2021. Soil microbiome structure and function in ecopiles used to remediate petroleum-contaminated soil. *Frontiers in Environmental Science*, 9, p.624070.
- Wydro, U., 2022. Soil microbiome study based on DNA extraction: A review. *Water*, 14(24), p.3999.
- Zhang, N., Ju, Z., and Zuo, T., 2018. Time for food: The impact of diet on gut microbiota and human health. *Nutrition*, 51, pp.80-85.