DATA NOTE

Crude oil hydrocarbons' effect on soil microbial metagenome from Niger Delta polluted soils [version 1; peer review: awaiting peer review]

Chidinma Peace Okafor¹,², Chioma Blaise Chikere³,⁴, Onyewuchi Akaranta⁵, Khayalethu Ntushelo⁶

¹Department of Microbiology, Africa Centre of Excellence, Centre for Oilfield Chemicals Research (ACECEFOR), University of Port Harcourt, Choba, Rivers, 500004, Nigeria
²Department of Microbiology, Imo State University, Owerri, Imo State, 460222, Nigeria
³Department of Microbiology, Faculty of Science, University of Port Harcourt, Choba, Rivers, 500004, Nigeria
⁴Department of Environmental Sciences, College of Agriculture and Environmental Sciences (CAES), University of South Africa, Roodepoort, Florida, Gauteng Province, 0003, South Africa
⁵Africa Centre of Excellence Centre for Oilfield Chemical Research (ACE-CEF), University of Port Harcourt, Choba, Rivers State, 500004, Nigeria
⁶Department of Agriculture and Animal Health Science, University of South Africa, Corner Christain De Wet and Pioneer Avenue, Science Campus, Private Bag X6, Florida, 1710, South Africa

Abstract

Crude oil pollution is an endemic environmental problem in the Niger Delta region of Nigeria with background pollution levels greater than the regulatory intervention limit of 5000mg/kg of soil as stipulated in the Environmental Guidelines and Standards for Petroleum Industry in Nigeria (EGASPIN) in most polluted sites. Hence, the essence of this study was to evaluate the extent of damage caused by the pollutant (crude oil hydrocarbons) on the soil physicochemical parameters and microbial communities as well as to determine pollution effects on soil microbial ecosystem services. The soil microbial community composition, diversity, functional genes, and metabolic pathways were studied to evaluate the pollutant effect on soil microbiomes and their ability to utilize petroleum hydrocarbons as carbon and energy sources. Two sites Bodo (N4.620134, 7.282998E) and Ngia-Ama Tombia (N4.9816667; 7.0608333E) were chosen for the study. The America standard testing methods (ASTM International) were used in measuring physicochemical parameters of the soil influencing microbial response and pollutant behavior in the soil environment. Soil pH was measured using a pH meter, total petroleum hydrocarbons (TPH) using gas chromatography with a mass spectrophotometer (GC-MS), polycyclic aromatic hydrocarbons (PAHs) (ASTM D5412-93(2017) e1), Heavy metals analysis (ASTM D8404-21), soil texture (ASTM D6913/D6913M-17). Shotgun metagenomic analysis
was used to determine microbial community composition, following “DNA extraction”, library preparation (Nextera® DNA Flex Library Prep Kit (Illumina, San Diego, CA), and sequencing using Illuminana NovaSeq® 6000. The results were analyzed using bioinformatics pipelines. The sequences generated were deposited in the European nucleotide Archive (ENA) with project accession number PRJEB53529.

**Keywords**
crude oil hydrocarbons, Pollution, Niger Delta, Shotgun metagenomics, Soil microbiomes.

This article is included in the Agriculture, Food and Nutrition gateway.
Introduction
Successful microbial remediation of polluted environments depends largely on the prevailing baseline physicochemical parameters, ecological conditions, available microbial community structure and their numbers and how the pollutant influences the interaction of these factors. For this study crude oil polluted soils were monitored prior to a microbial remediation project. Crude oil is a major pollutant of environmental concern. At certain concentrations it is toxic to microorganisms and other life forms within the receiving environmental compartment. It contains compounds like the polycyclic aromatic hydrocarbons (PAHs) which are priority chemicals of concern according to the United States Environmental Protection Agency (USEPA). PAHs are reported to be toxic, mutagenic, teratogenic to mammals. They bioaccumulate along the food chain and can persist in the environment. Crude oil hydrocarbons also alter the physicochemical parameters like pH, oxygen tension, and nutrient composition of receiving environment thereby affecting microbial activities within the impacted site. The Niger Delta region of Nigeria has a high incidence of crude oil pollution due to the heightened oil and gas industry activities within the region (Nwilo & Badejo, 2005). Most polluted environments within the region exceed both the regulatory intervention value of 40 mg/kg of soil or 70 μg/L for ground waters of PAHs and total petroleum hydrocarbons (TPH) concentration of 5000 mg/kg of soil or 600 μg/L for groundwater as contained in the Environmental Guidelines and Standards for the Petroleum Industry in Nigeria (EGASPIN) (Okafor et al., 2021; UNEP, 2011).

The data collected were to address the specific objectives of the study which were to:

1. Investigate the effect of crude oil hydrocarbons on diversity, distribution, and functional profile of indigenous microbial communities in the polluted soils. It also sought to determine metabolic pathways utilized by the indigenous microbial population in the degradation of the pollutant.

2. Determine the change if any on the physicochemical parameters occasioned by the crude oil spill.

Methods
The data were collected from two polluted sites in Rivers State, Nigeria, namely Bodo community (N4.620134, 7.282998E) and Ngia-Ama Tombia Community (N4.9816667; 7.0608333E) within the Niger Delta region of Nigeria. The Ngia-Ama Tombia site is a moderate low land in close proximity to mangroves and creeks whereas Bodo is an upland environment. Eight (8) samples per 1000 sqm of soil at the depth of 0-30 cm were collected from each site and used for this study. The eight samples were collected, four (4) from points around the pollution point and another four (4) samples about 200m away from the point source. The samples were pooled and homogenized into duplicates per site according to proximity before analysis. The physical properties of the polluted soil were analyzed according to regulatory recommended test methods for physicochemical parameters as contained in (APPENDIX VIII-E1, EGASPIN) (Resources, 2018). Sieve analysis and hydrometer method were used to determine the proportion of sand, salt, clay and sand using the standard test methods for particle-size distribution of both soils and fine-grained soils (ASTM, 2017b; ASTM, 2021a). The soil density, specific gravity and water content by gravimetric method, were also determined to project pollutant movement within the soil. Other parameters within the polluted soil were determined and these physicochemical parameters included pH of the soil, soil conductivity, soil nutrients (nitrates and phosphate concentration) of the polluted soil. pH was analyzed according to the United States Environmental Protection Agency (USEPA, 2004) using the pH meter (Hanna multimeter reader), soil conductivity was by a modification of the APHA 145 method, whereas the nutrients, nitrates and phosphates were determined using the spectrophotometer method (PG, T60) and nutrient pillows from HACH®. The concentration of extractable total petroleum hydrocarbon (eTPH) in the soil was determined using the GC-FID protocol. The residual petroleum hydrocarbons were extracted from the samples by hexane (for slightly contaminated soils) or chloroform (for heavily contaminated soils) and the extract cleaned in activated charcoal column prior to spectrophotometry. Agilent 1760 spectrophotometer with flame ionizing radiation was used for TPH concentration determination. GC-MS protocol was used in the determination of the PAHs concentration in the soil samples after sample clean-up (ASTM, 2017a). Heavy metals analysis was done using the ASTM-D8404-21 method (ASTM, 2021b). The practice covered drying, homogenization, ammonium bifluoride-nitric acid digestion of soil samples and associated quality control (QC) samples for the determination of metals and metalloids using laboratory flame atomic absorption spectrometry (FAAS).

For microbial community analysis, shotgun metagenomic sequencing was used to determine the composition, diversity, spread and functional profile of extant microbial population within the polluted environment. The samples were processed and analyzed with the ZymoBIOMICS® Shotgun Metagenomic Sequencing Service for Microbiome Analysis (Zymo Research, Irvine, CA). Sample processing included DNA extraction, library preparation, and metagenomic DNA sequencing.
DNA extraction
DNA extraction was done using the ZymoBIOMICS®-96 MagBead DNA Kit (Zymo Research, Irvine, CA), Cat. No. D4308 using an automated platform according to the manufacturers’ instructions. Following the extraction, 50 μL of extracted DNA was then used for library preparation.

Shotgun metagenomic library preparation
Genomic DNA samples were profiled with shotgun metagenomic sequencing. Sequencing libraries were prepared with the Nextera® DNA Flex Library Prep Kit (Illumina, San Diego, CA) Cat.No. 20018704 with up to 100 ng DNA input following the manufacturer’s protocol using internal dual-index 8 base pairs (bp) barcodes with Nextera® adapters (Illumina, San Diego, CA). All libraries were quantified with TapeStation® (Agilent Technologies, Santa Clara, CA) and then pooled in equal abundance. The final pool was quantified using quantitative polymerase chain reaction (qPCR). This step amplifies the Bead-linked Transposomes (BLT) tagmented DNA using a limited-cycle PCR program. The PCR step adds the Enhanced PCR mix (EPM), Nextera® DNA CD indexes (24 indexes, 24 samples) Cat.No. 20018707 and sequences required for sequencing cluster generation. Amplification was by five (5) cycles of: 98°C for 45 seconds, 62°C for 30 seconds, 68°C for 2 minutes, 68°C for 1 minutes and held at 10°C on the Applied Biosystems™ SimpliAmp™ Thermal cycler. The metagenomic library had no visualization of Ct value.

Sequencing
Sequencing depth, or coverage, refers to the number of times a reference base is represented within a set of sequencing reads. The higher the sequencing depth, the more sensitive the detection. Sequencing depth can critically affect the profiling of polymicrobial animal and environmental samples when using shotgun metagenomics. The depth needed will be determined by the sensitivity of detection required for an application.

The final library was sequenced on the Illumina NovaSeq® with a sequence depth of >15 M.

Bioinformatics and statistical analysis
Raw sequence reads were trimmed by sliding window with 6bp window size and a quality cutoff of 20 to remove low-quality fractions with size lower than 70 bp and adapters with Trimmomatic-0.33 (Bolger et al., 2014).

Control samples
The ZymoBIOMICS® microbial community standard (Zymo Research, Irvine, CA) was used as a positive control for each DNA extraction, ZymoBIOMICS® microbial community DNA standard (Zymo Research, Irvine, CA) was used as a positive control for each targeted library preparation. Negative controls (i.e. blank extraction control, blank library preparation control) were included to assess the level of bioburden carried by the wet-lab process.

These data set were collected as part of the polluted site characteristic project to determine various effects of hydrocarbon pollution on soil and its indigenous microbial communities. They are also to serve as polluted site reference genomes for comparison on effect of climate and other variables on microbial response to perturbations.

Data availability
The results obtained are deposited with the European Nucleotide Archive (ENA) of the European Molecular Biology Laboratory (EMBL) repository.

European Nucleotide Archive: Crude oil hydrocarbon polluted soil metagenomes. Accession number: PRJEB53529 (Okafot et al., 2022).

This project contains the following underlying data:

- Data file 1. Contains raw sequences of soil metagenomes from Bodo sample (BSP1) with sample accession number ERS12263116
- Data file 2. Contains raw sequences of soil metagenomes from Bodo sample (BSP2) with sample accession number ERS12263117
- Data file 3. Contains raw sequences of soil metagenomes from Tombia sample (TSP1) with samples accession number ERS12257751
- Data file 4. Contains raw sequences of soil metagenomes from Tombia sample (TSP2) with sample accession number ERS12257752

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

References

ASTM: Standard Test method for particle-size distribution (Gradation) of Fine-Grained soils using the sedimentation (Hydrometer) Analysis. Standard No. D7928-21e1. 2021a.

ASTM: Standard Practice for Preparation of Soil Samples by Ammonium Bifluoride-Nitric Acid Digestion for Subsequent Analysis for Metals and Metalloids. Standard No. D8404-21. 2021b.

ASTM: Standard Test Method for Quantification of Complex Polycyclic Aromatic Hydrocarbon Mixtures or Petroleum Oils in Water. Standard No. D5412-93(2017) e1. 2017a.

ASTM: Standard Test Method for Particle size Distribution (Gradation) of soils using sieve analysis. Standard No. D6913/D6913-17. 2017b.

Bolger AM, Lohse M, Usadel B: Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15): 2114-2120. PubMed Abstract | Publisher Full Text

Nwilo PC, Badejo OT: Oil spill problems and management in the Niger Delta. 2005 International Oil Spill Conference, IOSC 2005. 2005;
11400-11403. Publisher Full Text

Okafor CP, Chikere CB, Akaranta O, et al.: Crude oil polluted soil metagenomes. European Nucleotide Archive [Dataset]. 2022. Reference Source

Okafor CP, Udemang NL, Chikere CB, et al. Indigenous microbial strains as bioresource for remediation of chronically polluted Niger Delta soils. Scientific African. 2021; 11: e00682. Publisher Full Text

Resources, D. of P: Environmental Guidelines and Standards for the Petroleum Industry in Nigeria. (3rd ed.). 2018.

UNEP: Environmental Assessment of Ogoniland. 2011; 39: 205.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com