

Functional metabolite signatures reveal ecosystem recovery trajectories in oil-polluted soils undergoing bioremediation

Laura N. Nwogu-Chigozie

laura.nwoguchigozie@gmail.com

University of Port Harcourt

Chioma B. Chikere

University of Port Harcourt

Victor Ezebuio

Regional Centre for Biotechnology and Bioresources Research, University of Port Harcourt

Fidele Tugizimana

University of Johannesburg

Samuel Musiwalo

University of Johannesburg

Research Article

Keywords: Bioremediation, Petroleum-contaminated soil, Metabolomics, Ecosystem recovery, Physicochemical restoration, Functional biomarkers

Posted Date: May 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9657487/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Biodegradation on August 31st, 2026. See the published version at <https://doi.org/10.1007/s10532-026-10355-1>.

Abstract

Reliable indicators of ecosystem recovery following petroleum contamination remain limited, as remediation success is commonly assessed using hydrocarbon removal alone rather than restoration of soil function. This study integrated physicochemical profiling with untargeted LC–MS metabolomics to resolve recovery trajectories across polluted soil (PS), soil undergoing bioremediation (UB), bioremediated soil (BS), and uncontaminated reference soil (CNTR) in Ogoniland, Nigeria. Total petroleum hydrocarbon (TPH) declined from 78.26 ± 6.04 mg/kg in PS to 41.23 ± 1.40 mg/kg in UB and 23.38 ± 1.81 mg/kg in BS, representing a 70.1% reduction relative to PS. Sulphur decreased from 6.36 ± 0.70 wt% to 0.152 ± 0.052 wt%, whereas phosphorus increased from 0.068 ± 0.006 wt% to 0.325 ± 0.025 wt% during remediation. Principal component analysis of physicochemical variables explained 99.3% of total variance and resolved a directional restoration gradient (PS → UB → BS → CNTR). Untargeted metabolomics similarly separated treatment states, with PCA explaining 73.7% of total biochemical variance. Polluted and transitional soils were enriched in aromatic stress metabolites, including fluorene, triphenylphosphine oxide, quinolinic acid, and 2,4-dinitrophenol, whereas remediated soils accumulated recovery-associated metabolites such as citric acid, aconitic acid, ectoine, pyridoxamine, lipoic acid, and geranic acid. Integrated analyses showed strong positive correlations between TPH and fluorene ($r = 0.88$) and TPPO ($r = 0.84$), but negative correlations with citric acid ($r = -0.82$) and pyridoxamine ($r = -0.78$). These results demonstrate that hydrocarbon attenuation is tightly coupled to coordinated metabolic reprogramming and nutrient restoration. Functional metabolite signatures may provide a candidate framework for assessing ecosystem recovery beyond contaminant depletion alone, with implications for monitoring restoration of oil-impacted soils globally.

1.0 Introduction

Petroleum hydrocarbon contamination of arable soils constitutes a major environmental challenge that affects the socio-economic well-being and livelihoods of oil-producing communities (Chikere et al. 2021; Varjani, 2017; Patowary et al. 2023). Regular spills, wellhead leakages, pipeline ruptures, and inadequate waste management protocols cause complex mixtures of aliphatic hydrocarbons, monoaromatic substances and polycyclic aromatic hydrocarbons (PAHs) to be released into the soil of terrestrial ecosystems, and may be retained over years or decades (Li et al. 2023). When petroleum hydrocarbons enter into soil matrix, they interfere with aggregate structure, decrease aeration and water infiltration, deplete bioavailable nitrogen and phosphorus and produce toxic intermediates during partial degradation (Mekonnen et al. 2024; Kanungo et al. 2024). These are particularly acute in the Niger Delta, Nigeria where agricultural productivity, biodiversity and community health are affected by chronic pollution of hydrocarbons.

Petroleum pollution has ecological risks beyond acute toxicity (Pal & Sen, 2024; Tavella et al. 2025). Most petroleum products, especially high-molecular-weight PAHs are mutagenic, carcinogenic and bioaccumulative, and thus, they can be transferred through food chains as well as pose long-term exposure risks to humans and wildlife (Das et al. 2021). Seed germination is also inhibited by soil contamination as well as decreased photosynthetic capacity, alteration of antioxidant defense and inhibition of root growth in exposed plants (Adam and Duncan, 2002; Ahammed et al. 2012). These changes in agricultural reliant communities manifest as decreased crop yields, food insecurity and decreased ecosystem services. Microorganisms play a central role in soil resilience, they control decomposition, mineralization of nutrients, organic matter turnover and detoxification. These microbial communities are highly redefined by petroleum hydrocarbons that select specialist degraders and repress sensitive and functionally vital taxa (Das et al. 2021; Li et al. 2023).

Repeated metagenomic studies of polluted soils have demonstrated an increase in the abundance of hydrocarbonoclastic genera, including *Pseudomonas*, *Mycobacterium*, *Immundisolibacter* and *Pseudoxanthomonas*, as well as an increase in genetic capacity of toluene, xylene, fluorobenzoate and PAHs degradation (Chikere et al. 2023; Li et al. 2023). Although these reactions imply the ability to biodegrade adaptively, they do not imply the reinstatement of the larger ecological roles that are linked to the ability to have healthy soil systems. As a result, bioremediation became one of the favorite methods of land management in petroleum-contaminated areas due to its ability to take advantage of the metabolic versatility of microorganisms, plants, and organic amendments to convert toxic hydrocarbons into less harmful products (Patel et al. 2019 ; Kong et al. 2023).

Methods of nutrient stimulation, biosurfactants, compost additions, biochar, plant-microbe consortia have shown enhanced hydrocarbon removal and a partial recovery of soil biological functioning (Das et al. 2021; Ehis-Eriakha et al. 2025; Mu et al. 2025). Nevertheless, bulk measures of remediation performance are still widely determined, e.g. total petroleum hydrocarbon decline, number of microbes, vegetation growth or enzyme activity. As valuable as they are, these measurements do not tell much about the possibility of the underlying biochemical network of the soil being genuinely restored.

Functional recovery, in essence, is a metabolic process. The soil health preserves the equilibrium of the fluxes with the help of central carbon metabolism, amino acid synthesis, membrane lipid turnover and redox control (Atkinson et al., 2022). In contrast, polluted soils have a tendency to bear the mark of oxidative stress, emergency carbon usage, membrane remodeling and xenobiotic detoxification (Arbona et al. 2013). The deep-profiling of low-molecular-weight metabolites, known as metabolomics, provides one of the most potent methods of detecting these physiological states due to the fact that metabolites are the summed-up end products of gene expression, enzymatic activity and environmental interaction (Li et al. 2023; Kong et al., 2023; Arbona et al., 2013). Metabolomic analyses in petroleum-affected soils have revealed increased levels of intermediates that include salicylic acid, catechol, fatty acid and gentisic acid linked to aromatic hydrocarbon degradation, and stress adaptation (Li et al. 2023).

Recent developments in the field of multi-omics studies indicate that concomitant application of metabolomics to ecological restoration studies would help reveal latent soil recovery pathways that are overlooked by standard assays (Kong et al. 2023). However, even with growing interest little is known about the reliability of using metabolite signatures to tell biochemical recovery of chronically oil-polluted tropical soils following remediation. This disconnect is more pronounced in sub-Saharan Africa, where oil pollution is widespread yet limited molecular-scale surveillance of remediation exists. This is a challenge that should be tackled since the process of remediation cannot be characterized by disappearance of pollutants, but by reconstruction of a sound self-sustaining and self-reliant soil metabolic system. The expected result of a fully healed soil would be a shift in metabolism of contaminants, which is currently characterized by hydrocarbon catabolism, oxidative defenses and resource stress, to stable metabolisms associated with nutrient cycles, microbial cooperation and organic matter regeneration.

We used untargeted liquid chromatography-mass spectrometry (LC-MS) metabolomics in this study to compare the functional metabolite signature in polluted, remediating, bioremediated and unpolluted reference soils of the Ogoni River State, Rivers State, Nigeria. Our hypothesis was that active petroleum contamination was to cause specific xenobiotic and stress-related metabolic fingerprints and that successful bioremediation would lead to a convergence towards the biochemical profile of the uncontaminated soils. This work proposes the use of metabolite-based functional diagnostics as a future-generation approach to discussing the success of petroleum-impacted tropical ecosystem restorations thereby defining these recovery trajectories.

2.0 Materials and Methods

2.1 Sampling of the soil and study area.

Sampling was conducted in Ogoniland, Rivers State, in the southern part of Nigeria (approximately 4.69°N, 7.34°E). The samples were collected from crude oil contaminated soils. This region has been historically affected by chronic petroleum exploration, spills and related environmental degradation. To address biochemical paths in various states of the restoration, the soils were stratified into four treatment groups, which included actively polluted soil (PS), soil undergoing remediation (UB), previously bioremediated soil (BS), and nearby unpolluted reference soil (CNTR).

Three independent sampling locations were selected for each treatment category: polluted soil (PS), soil undergoing bioremediation (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR), giving a total of 12 biological samples. At each location, triplicate subsamples were collected within a 5–10 m radius and homogenized to obtain representative composite samples. Control soils were collected approximately 250m away from visibly contaminated zones and showed no observable hydrocarbon staining.

Surface debris were cleared at every sampling point and samples were collected aseptically by gathering the surface top soil (0–15 cm) with sterile stainless-steel augers. Replicates within each category were spatially clustered to reduce pseudo-replication. Samples were transferred into sterile polypropylene containers, placed in ice while in transit to the laboratory. The soils were thereafter homogenized, and sieved (2 mm mesh) to eliminate stones and roots, and then stored at -20°C until metabolite extraction. All samples were handled in the same way to minimize the variation caused by handling.

2.2 Physicochemical Characterization of Soil Samples

In order to complete the untargeted metabolomic profiling and determine the geochemical recovery of the remediation process, four physicochemical characterizations of the soils representing the four experimental conditions (polluted soil (PS), soil that was undergoing bioremediation (UB), bioremediated soil (BS), and unpolluted control soil (CNTR)) were conducted. The contaminated, untreated, Soil undergoing remediation and bioremediated samples were all from hydrocarbon-contaminated soils from Ogoni land (Gokana Local Government Area, Rivers State, Nigeria). About 100g of each sample was stored in sterile plastic containers and stored under ambient storage conditions before being subjected to laboratory analysis at the SPDC Production Chemistry Port Harcourt Laboratory.

For statistical robustness, three independent biological replicates in each treatment group were taken and designated as follows: Polluted soil (PS1–PS3), uncontaminated control (CNTR1–CNTR3), undergoing bioremediation (UB1–UB3), and bioremediated soil (BS1–BS3) were assigned sample identifiers. The four-stage design facilitated comparative evaluation of progressive recovery curves of contamination by active remediation to stabilized post-remediation soil.

2.2.1 Colour, Consistency, texture, moisture content, and pH

Firstly, visual inspection was performed to record soil colour, texture and consistency. Free moisture content (%) was analysed API 13A standard procedure, which determines the amount of gravimetric moisture in soil matrices (American Petroleum Institute [API], 2010). The pH of soil samples was determined as a 5% slurry suspension at 25°C using the SMEWW 4500-HB procedure on a calibrated pH meter after equilibration (APHA, AWWA, and WEF, 2017).

2.2.2 Elemental contents

Standardized multi-elemental procedure was used to measure the major and trace elements in all replicates in accordance with ASTM E1621-22. Measurements were taken on Aluminum (Al), Silicon (Si), Phosphorus (P), Sulphur (S), Chlorine (Cl), Potassium (K), Calcium (Ca), Titanium (Ti), Vanadium (V), Chromium (Cr), Manganese (Mn), Iron (Fe), Nickel (Ni), Copper (Cu), Zinc (Zn). This technique allows sensitive and quick quantification of the elements of interest that are nutrient-associated as well as toxicologically relevant in solid environmental matrices (ASTM International, 2022).

2.2.3 Total Petroleum Hydrocarbon (TPH) contents

Petroleum contamination was evaluated by measuring Total Petroleum Hydrocarbon (TPH) concentrations using EPA Method 3550C combined with EPA Method 8015C. In brief, ultrasonic solvent extraction (Method 3550C) was used to extract hydrocarbons present in homogenized soil, and extractable fractions of petroleum were quantified by flame ionization detection (Method 8015C) (USEPA, 2007a, 2007b).

2.2.4 Physicochemical data analysis

Each replicate was measured and the measurements were tabulated into a physicochemical matrix to be subjected to downstream multivariate analysis. Data were normalized before being integrated with metabolomic features and then the data were analyzed by principal component analysis (PCA), hierarchical clustering, correlation network analysis, and pathway-level interpretation of soil biochemical recovery. Trace elements exhibiting high replicate variability were retained in exploratory multivariate analyses but interpreted cautiously during ecological inference.

2.3 Metabolite Extraction

One gram of soil samples was transferred to a Falcon tube. 10 milliliters (10 mL) of 80% chilled aqueous methanol and hexane were measured and poured into a Falcon tube containing 1 g of the soil. The samples were vortexed for 30 seconds and homogenized for 1 minute and 30 seconds. To separate the metabolites from the soil debris, centrifugation was performed at a rotary speed of 4700 rpm for 30 min at 4°C. The resulting supernatant (1 mL) was transferred into a 2 mL microcentrifuge tube and dried to completeness using a dry bath (heating block) set at 55°C. 200 hundred microliters (200 µL) of 50% LC-MS grade methanol was added into the dried extract for re-suspension. The re-suspended extracts were then filtered through an ultra-membrane filter (pore size 0.45 µm) and collected in HPLC glass vials.

2.4 Data acquisition

Samples were analysed on an ultra-high performance liquid chromatography (UHPLC) system coupled in series to a SYNAPT XS mass spectrometer (Waters Corporation, Milford, MA, USA). For chromatographic separation, a reverse-phase HSS T3 C18 column (150 mm × 2.1 mm × 1.8 µm, Waters Corporation) was used, which was thermostatted maintained at 60°C. A binary mobile phase consisted of water (Solvent A) containing 10 mM formic acid (pH 2.4) and acetonitrile (Solvent B) containing 10 mM formic acid. The gradient elution was performed as follows: 10% B was maintained for the first minute, followed by a linear increase from 10% to 90% B over 1–15 minutes. The gradient was then increased from 90% to 99% B between 15 and 15.10 minutes and held constant until 17 minutes. The gradient was reset to 90% B at 17.10 minutes and maintained until 20 minutes. The injection volume was 2 µL, and the total run time was 21 minutes.

For mass spectrometric analysis, ionization was performed in both positive and negative ESI modes, with a mass range of 100–1500 Da and a scan time of 0.2 seconds. The source temperature was set to 120°C. Leucine

enkephalin (50 pg/mL) served as the lock mass calibrant, ensuring mass accuracy within 1–5 mDa. Instrument parameters included a capillary voltage of 0.6 kV and a sampling cone voltage of 30 V. Fragmentation data were acquired using a data-independent acquisition (DIA) strategy, with collision energy ramping from 10 to 40 eV to enhance spectral coverage.

To ensure analytical robustness and reproducibility during LC–MS analysis, pooled quality control (QC) samples were prepared by combining equal aliquots from all extracted soil samples and injected periodically throughout the analytical sequence. QC injections were used to monitor instrument stability, retention time consistency, mass accuracy, and signal reproducibility across the batch run. Blank solvent injections were also included between selected runs to assess carryover and background contamination. Leucine enkephalin was continuously infused as a lock-mass calibrant to maintain mass accuracy within 1–5 mDa throughout data acquisition. Tight clustering of biological replicates observed in principal component analysis further confirmed analytical reproducibility and stability of the metabolomic dataset.

2.5 Molecular networking and metabolite annotation

Prior to computing the feature-based molecular networks (FBMN) through the Global Natural Product Social (GNPS), the acquired raw datasets obtained from the LCMS-qTOF-MS were converted to open-source format (.mzML) files. The mzML files were then uploaded to the Mass Spectrometry-Data Independent Analysis (MS-DIAL) platform for data processing. The parameters used for MS-DIAL data processing parameters included MS1 and MS2 mass accuracy tolerance of 0.05 Da, a mass range of 100–1500 Da, at the minimum peak height of 2000 amplitude, a mass slice width of 0.1 Da for peak detection, a sigma window value of 0.5 was used, and a retention time tolerance of 0.2 min.

Post MS-DIAL data processing, the GNPS export files, both GNPS MGF files and feature quantification tables, were exported into the GNPS2 ecosystem for molecular networking. FBMNs were then computed and generated using the GNPS molecular networking workflow. The parameters used for computing FBMN spectral datasets included a precursor ion mass tolerance of 0.05 Da and a fragment ion mass tolerance of 0.05 Da. For spectral similarity, a cosine score cut-off was 0.6, with a minimum of 4 matched fragment ions. For annotation, various spectral libraries were searched, including MassBank, ReSpec, HMDB, SUPNAT, CHEBI, DRUGBANK and NIST. To visualise and analyse the computed networks, the Cytoscape network visualisation software (version 3.10.1) (Shannon et al. 2003; Smoot et al. 2011) was used. All semi-annotated and some unmatched/unidentified nodes, visualised in Cytoscape, were verified using their empirical formulae calculated from accurate mass and fragmentation patterns. In addition to manual inspection of annotations, natural products dereplication databases, such as PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) and Mass Bank (<https://massbank.eu/MassBank/>), were also utilised. The annotations were also verified and confirmed against available literature. All metabolite annotations were carried out at level 2 of the Metabolomics Standards Initiative (MSI) (Summer et al. 2007). All validated metabolites are presented in Table 3.

2.6 Chemometrics analyses

Pre-processed data matrices of annotated metabolites were exported for analysis on MetaboAnalyst (version 6.0; <http://www.metaboanalyst.ca>) online software for univariate and multivariate statistical analysis to capture the data structures and characteristics such as natural sample clustering, trends, outliers, and any groupings. All data were normalised to the median, log-transformed and Pareto-scaled prior to further analysis. Principal component analysis (PCA) was performed to visualize global metabolomic variation and clustering among treatment groups. Permutational multivariate analysis of variance (PERMANOVA) based on Bray–Curtis distance matrices was

conducted using 999 permutations to statistically assess group separation. Partial least squares discriminant analysis (PLS-DA) was additionally performed to identify metabolites contributing significantly to treatment discrimination. Variable importance in projection (VIP) scores were calculated, and metabolites with VIP > 1 were considered major discriminatory metabolites. To reduce false-positive discovery, p-values were adjusted using false discovery rate (FDR) correction according to the Benjamini–Hochberg procedure. Hierarchical cluster analysis (HiCA) heatmaps were generated, allowing the visualisation of the distribution of selected metabolites across different conditions.

2.7 Integrated omics analyses linking geochemical restoration with biochemical recovery

An exploratory integrative analysis was conducted to relate soil physicochemical restoration with metabolomic restructuring during bioremediation. Key geochemical variables (TPH, sulphur, chlorine, phosphorus, iron, and manganese) were selected based on PCA loadings and ecological relevance, while representative metabolites were chosen from contamination- and recovery-associated chemical classes identified by untargeted LC–MS/MS. Group mean values for polluted soil (PS), undergoing bioremediation soil (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR) were standardized prior to analysis.

Spearman correlation coefficients were calculated to assess associations between soil variables and metabolites and visualized as a clustered heatmap. Integrated ordination biplots were generated to display joint treatment separation and environmental drivers.

Concordance between independent physicochemical and metabolomic PCA models was evaluated using Procrustes superimposition, where centroid residual distances indicated similarity between datasets. Analyses and figures were produced in Python using pandas, NumPy, SciPy, scikit-learn, matplotlib, and seaborn.

3.0 Results

3.1 Physicochemical restructuring reveals progressive recovery of oil-polluted soils during bioremediation

Comprehensive physicochemical profiling of polluted soil (PS), undergoing bioremediation soil (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR) demonstrated marked geochemical transitions consistent with progressive ecosystem restoration following remediation treatment.

3.1.1 Strong attenuation of petroleum hydrocarbon burden during remediation

Total petroleum hydrocarbon (TPH) concentration was highest in polluted soils (78.26 ± 6.04 mg/kg), confirming severe petroleum contamination (Table 1). Lower TPH values were recorded for soils undergoing treatment (UB) (41.23 ± 1.40 mg/kg) and with the bioremediated soil (23.38 ± 1.81 mg/kg). The control soil had the least TPH concentrations (10.77 ± 0.81 mg/kg) (Table 1). This stepwise decline across PS → UB → BS → CNTR indicates sustained hydrocarbon removal during remediation.

Similar trend was observed with sulphur concentrations; sample PS had the highest level of concentration (6.36 ± 0.70 wt%), followed by UB (2.81 ± 0.12 wt%). BS sulphur concentration (0.152 ± 0.052 wt%) was close to the CNTR

value of 0.199 ± 0.010 wt%. Chlorine concentrations declined markedly from 0.845 ± 0.100 wt% in PS to 0.282 ± 0.062 wt% in BS, further indicating removal of petroleum-associated residues.

3.1.2 Nutrient recovery and mineral enrichment characterize remediated soils

In contrast to declining pollutant markers, several nutrient- and mineral-associated parameters were higher during remediation. The following phosphorus concentrations 0.068 ± 0.006 wt%, 0.208 ± 0.007 wt%, 0.325 ± 0.025 wt% and 0.253 ± 0.012 wt% were observed for PS, UB, BS and CNTR, respectively (Table 1). This suggests enhanced nutrient turnover and phosphorus mobilization during remediation.

Iron had a similar trend with PS, UB, BS, and CNTR values of 4.03 ± 0.50 , 5.93 ± 0.15 , 8.00 ± 0.70 , and 2.37 ± 0.12 , respectively. The data shows the following iron contents trend BS > UB > PS > CNTR. Aluminium was highest in BS (15.43 ± 0.65 wt%), followed by UB (12.77 ± 0.35 wt%), CNTR (8.90 ± 0.21 wt%) with PS having the least value (6.78 ± 0.47 wt%). Collectively, these changes indicate post-treatment mineral reorganization and restoration of soil geochemical functionality.

Some trace metals, particularly nickel and arsenic, showed relatively high variability among replicates, suggesting localized heterogeneity or potential influence of outlier measurements; therefore, these variables were interpreted cautiously in downstream ecological interpretation.

Table 1
Physicochemical composition of polluted soil (PS), undergoing bioremediation soil (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR).

Parameter	Polluted Soil (PS)	Undergoing Bioremediation (UB)	Bioremediated Soil (BS)	Control Soil (CNTR)
Aluminum-Al (Wt%)	6.78 ± 0.47	12.77 ± 0.35	15.43 ± 0.65	8.90 ± 0.21
Silicon-Si (Wt%)	33.13 ± 2.70	36.23 ± 0.45	35.80 ± 2.06	43.40 ± 1.14
Phosphorus-P (Wt%)	0.068 ± 0.006	0.208 ± 0.007	0.325 ± 0.025	0.253 ± 0.012
Sulphur-S (Wt%)	6.36 ± 0.70	2.81 ± 0.12	0.152 ± 0.052	0.199 ± 0.010
Chlorine-CL (Wt%)	0.845 ± 0.100	0.388 ± 0.015	0.282 ± 0.062	0.304 ± 0.016
Potassium-K (Wt%)	0.712 ± 0.084	0.738 ± 0.021	0.707 ± 0.084	0.395 ± 0.047
Calcium-Ca (Wt%)	0.526 ± 0.099	0.234 ± 0.009	0.047 ± 0.041	0.124 ± 0.016
Titanium-Ti (Wt%)	1.287 ± 0.157	0.838 ± 0.020	2.380 ± 0.165	1.493 ± 0.051
Vanadium-V (Wt%)	0.0123 ± 0.0098	0.0302 ± 0.0012	0.0503 ± 0.0116	0.0173 ± 0.0034
Chromium-Cr (Wt%)	0.020 ± 0.010	0.029 ± 0.002	0.037 ± 0.006	0.020 ± 0.000
Manganese-Mn (Wt%)	0.0161 ± 0.0027	0.0429 ± 0.0017	0.0796 ± 0.0041	0.0172 ± 0.0006
Iron-Fe (Wt%)	4.03 ± 0.50	5.93 ± 0.15	8.00 ± 0.70	2.37 ± 0.12
Nickel-Ni (Wt%)	0.0606 ± 0.0074	0.0593 ± 0.0015	0.227 ± 0.286*	0.0428 ± 0.0022
Copper-Cu (Wt%)	0.0367 ± 0.0058	0.0340 ± 0.0010	0.0367 ± 0.0058	0.0267 ± 0.0058
Zinc-Zn (Wt%)	0.0237 ± 0.0031	0.0270 ± 0.0010	0.0283 ± 0.0023	0.0160 ± 0.0010
Arsenic-As (Wt%)	0.0005 ± 0.0007†	0.00037 ± 0.00006	0.667 ± 1.155*	0.0001 ± 0.0000
Bromine-Br (Wt%)	0.00639 ± 0.00480	0.00990 ± 0.00045	0.00070 ± 0.00010	0.00103 ± 0.00006
Molybdenum-Mo (Wt%)	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Silver-Ag (Wt%)	0.0105 ± 0.0024	0.00927 ± 0.00025	0.00893 ± 0.00110	0.00637 ± 0.00101
Cadmium-Cd (Wt%)	0.0069 ± 0.0016	0.0030 ± 0.0002	< 0.0001	0.0041 ± 0.0008
Tin-Sn (Wt%)	0.0202 ± 0.0023	0.0172 ± 0.0004	0.0174 ± 0.0010	0.0121 ± 0.0005

Note: Values are presented as mean ± standard deviation (SD) of triplicate measurements (n = 3). Values reported as below detection limit remain expressed as such. *Indicates high variability likely due to potential outlier values in the uploaded dataset. † Estimated numerically from censored values reported as <value for summary purposes. ‡ One detected replicate with two censored replicates.

Parameter	Polluted Soil (PS)	Undergoing Bioremediation (UB)	Bioremediated Soil (BS)	Control Soil (CNTR)
Lead-Pb (Wt%)	0.0027 ± 0.0001	0.00393 ± 0.00015	0.0080 ± 0.0009	0.00283 ± 0.00006
Barium-Ba (Wt%)	< 0.0001	0.0717 ± 0.0025	0.0048 ± 0.0081‡	< 0.0001
Mercury-Hg (Wt%)	0.00020 ± 0.00010†	< 0.0001	0.00080 ± 0.00010†	0.00050 ± 0.00010†
Total petroleum hydrocarbon-TPH (mg/kg)	78.26 ± 6.04	41.23 ± 1.40	23.38 ± 1.81	10.77 ± 0.81
Note: Values are presented as mean ± standard deviation (SD) of triplicate measurements (n = 3). Values reported as below detection limit remain expressed as such. *Indicates high variability likely due to potential outlier values in the uploaded dataset. † Estimated numerically from censored values reported as <value for summary purposes. ‡ One detected replicate with two censored replicates.				

3.1.3 Principal component analysis resolves distinct soil states and remediation trajectory

Principal component analysis (PCA) carried out to reduce multidimensional soil variables into principal components that reveal the dominant contamination-recovery gradients and identify the key parameters that drive separation the soil treatment groups, clearly separated the four soil states along two major axes that accounted for 99.3% of total variance (PC1 = 75.9%, PC2 = 23.4%) (Fig. 1A). PC1 primarily represented the contamination–recovery gradient, separating polluted soils from remediated and control soils.

PS clustered distinctly at the positive end of PC1, reflecting high TPH and sulphur burden. UB occupied an intermediate position between PS and BS, confirming an active transition state during treatment. BS formed a separate cluster characterized by restored nutrient and mineral signatures, while CNTR occupied a distinct natural baseline position.

The sequential positioning of sample centroids along the PCA space strongly supports a directional recovery model: PS → UB → BS → CNTR.

3.1.4 Loading analysis identifies variables driving separation among soil states

PCA loading analysis was performed to identify variables that drive the separation among the different samples with the result demonstrating that TPH, sulphur, and chlorine were the principal drivers of polluted soil separation along PC1 (Fig. 1B). In contrast, phosphorus, aluminum, iron, manganese, and titanium loaded strongly toward the remediated soil axis, indicating their importance in defining recovery-associated geochemical states.

Silicon showed a strong loading toward the control soil axis, suggesting preservation of native mineral matrix characteristics in uncontaminated soils.

3.1.5 ANOVA and post hoc analyses reveal significant physicochemical recovery during bioremediation

PCA loading analysis identified TPH, phosphorus, sulphur, calcium, and silicon as the principal variables contributing to separation among polluted, transitional, remediated, and control soils. These parameters therefore substantially accounted for most of the observed recovery gradient and were selected for targeted univariate analysis using ANOVA and Tukey’s post hoc comparisons (Table 2).

One-way ANOVA revealed significant differences among polluted soil (PS), soil undergoing bioremediation (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR) for all key physicochemical parameters examined (TPH: $F = 244.10$, $p < 0.001$; phosphorus: $F = 160.59$, $p < 0.001$; sulphur: $F = 199.93$, $p < 0.001$; calcium: $F = 44.84$, $p < 0.001$; and silicon: $F = 17.59$, $p < 0.001$).

TPH declined progressively from 78.26 ± 6.04 mg/kg in PS to 41.23 ± 1.40 mg/kg in UB and 23.38 ± 1.81 mg/kg in BS, compared with 10.77 ± 0.81 mg/kg in CNTR. Tukey’s test confirmed significant differences among all groups ($p < 0.01$), with approximately 70.1% reduction in TPH observed between PS and BS (Fig. 2).

Phosphorus increased significantly from 0.068 ± 0.006 wt% in PS to 0.208 ± 0.007 wt% in UB and 0.325 ± 0.025 wt% in BS, exceeding CNTR (0.253 ± 0.012 wt%). All pairwise comparisons were significant ($p < 0.05$), indicating strong nutrient restoration during remediation.

Sulphur decreased sharply from 6.36 ± 0.70 wt% in PS to 2.81 ± 0.12 wt% in UB and 0.152 ± 0.052 wt% in BS, closely matching CNTR (0.199 ± 0.010 wt%). Notably, BS and CNTR were not significantly different ($p = 0.998$), confirming near-complete normalization of sulphur contamination.

Calcium also recorded significant decline from 0.526 ± 0.099 wt% in PS to 0.234 ± 0.009 wt% in UB, 0.047 ± 0.041 wt% in BS, and 0.124 ± 0.016 wt% in CNTR. Silicon was highest in CNTR (43.40 ± 1.14 wt%) relative to PS (33.13 ± 2.70 wt%), UB (36.23 ± 0.45 wt%), and BS (35.80 ± 2.06 wt%), which indicated that native mineral composition in uncontaminated soil was preserved.

Overall, these analyses quantitatively demonstrate progressive physicochemical restoration from PS through UB to BS, characterized by significant hydrocarbon removal, sulphur detoxification, phosphorus enrichment, and mineral re-equilibration.

Table 2
Mean ± SD of Key Physicochemical Parameters with Tukey Superscript Grouping. Values are presented as **mean ± SD (n = 3)**. Means within the same row with **different superscript letters** are significantly different at **p < 0.05** (Tukey HSD).

Parameter	PS	UB	BS	CNTR
TPH (mg/kg)	$78.26 \pm 6.04^{\text{a}}$	$41.23 \pm 1.40^{\text{b}}$	$23.38 \pm 1.81^{\text{c}}$	$10.77 \pm 0.81^{\text{d}}$
Phosphorus (Wt%)	$0.068 \pm 0.006^{\text{a}}$	$0.208 \pm 0.007^{\text{b}}$	$0.325 \pm 0.025^{\text{c}}$	$0.253 \pm 0.012^{\text{b}}$
Sulphur (Wt%)	$6.36 \pm 0.70^{\text{a}}$	$2.81 \pm 0.12^{\text{b}}$	$0.152 \pm 0.052^{\text{c}}$	$0.199 \pm 0.010^{\text{c}}$
Calcium (Wt%)	$0.526 \pm 0.099^{\text{a}}$	$0.234 \pm 0.009^{\text{b}}$	$0.047 \pm 0.041^{\text{c}}$	$0.124 \pm 0.016^{\text{c}}$
Silicon (Wt%)	$33.13 \pm 2.70^{\text{a}}$	$36.23 \pm 0.45^{\text{ab}}$	$35.80 \pm 2.06^{\text{ab}}$	$43.40 \pm 1.14^{\text{b}}$

3.1.6 Hierarchical clustering reveals coordinated pollutant and recovery signatures

Clustered heatmap analysis of normalized group means revealed two dominant physicochemical modules (Fig. 3). The first module comprised contamination-associated variables (including TPH, sulphur, chlorine, and calcium), all enriched in polluted soils. The second module contained recovery-associated variables (including phosphorus, iron, aluminum, manganese, chromium, and vanadium), which had elevated values in BS.

UB displayed an intermediate heatmap signature characterized by partial suppression of contamination markers and progressive enrichment of nutrient-associated variables, consistent with an active remediation stage.

3.1.7 Correlation network analysis identifies hydrocarbon burden and nutrient restoration hubs

Correlation network analysis carried out to identify hydrocarbon burden and nutrient restoration hubs showed that contamination markers, particularly TPH–sulphur, TPH–chlorine, and sulphur–chlorine (Fig. 4A) were strongly correlated. In contrast, phosphorus–iron, iron–manganese, phosphorus–potassium, and aluminum–titanium revealed recovery-associated relationships. Similarly, TPH and phosphorus, TPH and iron, and sulphur and phosphorus, showed strong inverse relationship, indicating that reduction of hydrocarbon burden was tightly linked to nutrient restoration. These results therefore identify TPH as the dominant contamination hub and phosphorus as a central recovery hub.

3.1.8 Composite recovery indexing quantifies remediation success, aiding interpretation of soil recovery dynamics

To integrate multiple physicochemical endpoints, a composite recovery index was calculated using contamination and nutrient-associated parameters (Fig. 4B). Polluted soils were assigned the baseline degraded score (0.00), whereas UB achieved a partial restoration score of 0.58. Bioremediated soils reached the highest score (1.06), exceeding the uncontaminated control reference (1.00). This indicates that post-remediation soils not only approached natural baseline conditions but also exhibited enhanced restoration-associated properties, particularly phosphorus enrichment and mineral activation.

Hierarchical visualization of soil chemistry based on sunburst analysis (Fig. 4C) resolved distinct treatment-specific physicochemical states. Polluted soil (PS) was strongly dominated by TPH and sulphur, consistent with persistent hydrocarbon loading and contamination-derived geochemical stress. Soil undergoing bioremediation (UB) showed a mixed intermediate profile, retaining residual TPH with increased contributions from phosphorus, iron, and silicon, indicative of active elemental redistribution during the course of remediation. In contrast, bioremediated soil (BS) was enriched with phosphorus, iron, and manganese, reflecting nutrient recovery and restoration of functional soil chemistry. Uncontaminated control soil (CNTR) clustered around native mineral constituents, predominantly silicon and aluminum. Collectively, the radial partitioning pattern captures a progressive shift from contamination-dominated chemistry towards a re-equilibrated mineral and nutrient state following bioremediation.

Across all analyses carried out on physicochemical parameters, independent statistical approaches converged on the same conclusion: bioremediation induced coordinated restructuring of petroleum-impacted soils from a contamination-dominated state toward a nutrient-enriched and geochemically restored condition. The inclusion of the UB stage revealed that recovery was progressive rather than binary, with measurable transitional improvements observed prior to full-scale stabilized post-remediation state.

These findings provide strong physicochemical evidence that bioremediation substantially reverses petroleum-induced soil degradation and re-establishes functional environmental quality.

3.2 Metabolomic insight on recovery of oil-polluted soils during bioremediation

To provide metabolite-based evidence on the burden and recovery status of these different soils, the study analysed metabolic extracts from the samples, carried out data acquisition, conducted molecular networking, metabolite annotation, chemometric analyses on the generated and finally correlated the metabolomic data with that of physicochemistry.

3.2.1 Metabolite extraction reveals complex and treatment-dependent chemical profiles

Untargeted LC–MS metabolomic profiling revealed substantial biochemical differences among polluted soil (PS), soil undergoing bioremediation (UB), bioremediated soil (BS), and uncontaminated control soil (CNTR). Total ion chromatograms showed complex metabolite profiles across all treatments, with clear differences in peak intensity, retention time distribution, and signal abundance between groups, indicating pronounced treatment-dependent metabolic restructuring. Peaks were distributed across the chromatographic run from polar compound that eluted early (~ 0.8–3.0 min) to later-eluting hydrophobic metabolites (~ 15.0–19.5 min), demonstrating efficient extraction of chemically diverse metabolite classes from all soil types (Fig. 5).

Polluted soils displayed signals with higher intensity within aromatic and mid-retention windows, whereas remediated soils exhibited broader distributions of later-eluting metabolites, suggesting compositional shifts associated with hydrocarbon degradation and recovery processes. UB samples showed intermediate chromatographic complexity consistent with an active transitional biochemical state.

3.2.2 Data acquisition captures robust metabolomic separation among soil treatment groups

Principal component analysis (PCA) of acquired metabolomic features demonstrated that the four soil states were strongly separated, with tight clustering of biological replicates, confirming high analytical reproducibility and robust data acquisition quality (Fig. 6). Tight clustering of replicate samples in PCA plots indicated high analytical reproducibility and minimal instrumental drift and strong analytical stability throughout LC–MS acquisition. PC1 explained 52.3% of total variance, whereas PC2 explained 21.4%, with both accounting for 73.7% of total metabolomic variance.

PS samples clustered distinctly on the negative side of PC1, whereas BS samples occupied the positive PC1 and positive PC2 quadrant. UB samples clustered separately on the positive PC1 but negative PC2 axis, while CNTR samples clustered independently close to the negative PC1 region. The clear separation of treatment groups demonstrates that the acquired metabolomic dataset sensitively captured biochemical differences associated with petroleum contamination and remediation. PERMANOVA confirmed significant treatment separation among soil treatment groups.

The intermediate positioning of UB between PS and BS strongly supports a directional biochemical recovery trajectory: PS → UB → BS, with CNTR representing the ecological reference state.

3.2.3 Molecular networking and metabolite annotation reveal diverse recovery-associated chemical classes

Feature-based molecular networking revealed an extensive metabolite landscape comprising nodes that were both shared and treatment-specific (Fig. 7A). Numerous interconnected clusters detected across all samples indicated metabolite families that were structurally related, while unique sub-networks in PS, UB, and BS were suggestive of treatment-specific biochemical processes.

Chemical annotation identified metabolites that belong predominantly to organic acids, lipids, benzenoids, amino acids and peptides, pyridines, pyrans, with smaller contributions from alkaloids, flavonoids, and indoles and derivatives (Fig. 7B).

The polluted and transitional soils harboured several aromatic and stress-associated metabolites, including triphenylphosphine oxide (TPPO), 2,4-dinitrophenol, fluorene, and quinolinic acid. The presence of fluorene and related benzenoid compounds is typical of petroleum-derived aromatic residues and oxidative stress chemistry.

Conversely, more of metabolites associated with central carbon metabolism and ecological recovery (citric acid, aconitic acid, citramalate, gluconic acid and gluconic lactone) were found in remediated soils, suggesting the re-establishment of normal microbial respiration and carbohydrate catabolism. Several lipid-related metabolites (decanoic acid, heptanoic acid, dodecanedioic acid, tridecanedioic acid, lipoic acid, and geranic acid), potentially linked to membrane repair, β -oxidation and alkane degradation during remediation were also identified. Furthermore, more Nitrogen-containing metabolites including L-hydroxylysine, N-methyllysine, pyridoxamine, salsolinol and indole-3-acetamide were found in UB and BS, indicating restoration of amino acid metabolism, cofactor biosynthesis and microbial metabolism.

Annotation by MicrobeMASST linked some of the metabolites, such as gallic acid and geranic acid, to bacterial and fungal species, including Actinomycetota, Streptomycetaceae, Bacillota, and Basidiomycota, which are typically involved in aromatic compound degradation and soil organic matter recycling (Fig. 7C-D). In total, a range of annotated metabolites spanning benzenoids, organic acids, lipids, amino acids and peptides, pyridines, pyrans, alkaloids, flavonoids, and indole derivatives were identified based on accurate mass, retention time, fragmentation and adducts (Table 3).

Table 3
Putatively annotated metabolites detected in soil extracts by untargeted LC–MS/MS analysis.

Compound name	Abbrev	Chemical Formula	m/z	Rt	Fragments	Adducts	Chemical classes
Triphenylphosphine oxide	TPPO	C ₁₈ H ₁₅ OP	279.23	13.48	201,202,279	[M + H] +	Benzenoids
Alizarin	Aliz	C ₁₄ H ₈ O ₄	239.05	10.14	210,211	[M-H] -	Benzenoids
2,4-Dinitrophenol	2,4-DNP	C ₆ H ₄ N ₂ O ₅	183.00	5.96	153,183	[M-H] -	Benzenoids
Fenuron	Fenu	C ₉ H ₁₂ N ₂ O	165.09	5.62	120,165	[M + H] +	Benzenoids
Fluorene	Fluo	C ₁₃ H ₁₀	167.08	11.18	123,165,165,167	[M + H] +	Benzenoids
Gallic acid	GalA	C ₇ H ₆ O ₅	171.03	2.80	126,153,170	[M + H] +	Benzenoids
Ne-Acetyllysine	Ne-Ace	C ₈ H ₁₆ N ₂ O ₃	187.03	8.88	141,145,169	[M-H] -	Organic acids
Aconitic Acid	AcoA	C ₆ H ₆ O ₆	173.00	1.01	129,154	[M-H] -	Organic acids
2-Aminoisobutyric acid	2-AmiBA	C ₄ H ₉ NO ₂	102.04	10.34	102	[M-H] -	Organic acids
Citric acid	CitrA	C ₆ H ₈ O ₇	191.01	1.40	111,191	[M-H] -	Organic acids
N-acetylputrescine	Aceput	C ₆ H ₁₄ N ₂ O	131.11	1.00	114,131	[M + H] +	Organic acids
Ectoine	Ect	C ₆ H ₁₀ N ₂ O ₂	143.08	1.00	101,102,143	[M + H] +	Organic acids
Gluconic lactone	Glu lac	C ₆ H ₁₀ O ₆	177.04	6.35	141,176,177	[M-H] -	Organic acid
Gluconic acid	GluA	C ₆ H ₁₂ O ₇	195.03	9.00	121,151,159,195	[M-H] -	Organic acid
Heptanoic acid	HeptA	C ₇ H ₁₄ O ₂	5.50	129.08	111,129	[M-H] -	Lipids
Citramalate	Citr	C ₅ H ₈ O ₅	4.55	147.03	103,129,147	[M-H] -	Lipids
Decanoic acid	Deca	C ₁₀ H ₂₀ O ₂	171.12	9.52	151,171	[M-H] -	Lipids
Lipoic acid	LipA	C ₈ H ₁₄ O ₂ S ₂	205.02	3.01	127,171	[M-H] -	Lipids
Dodecanedioic acid	DodeA	C ₁₂ H ₂₂ O ₄	229.13	6.80	185,211,229	[M-H] -	Lipids

Note: Compounds were annotated using m/z, retention time, fragmentation ions, adduct patterns, and spectral similarity. Detected metabolites included benzenoids, organic acids, lipids, amino acids, pyridines, pyrans, alkaloids, flavonoids, and indoles, reflecting biochemical signatures of contamination, active bioremediation, and restored soil function.

Compound name	Abbrev	Chemical Formula	m/z	Rt	Fragments	Adducts	Chemical classes
Tridecanedioic acid	TrideA	C ₁₃ H ₂₄ O ₄	243.15	9.69	179,199,225	[M-H] -	Lipids
Geranic acid	GerA	C ₁₀ H ₁₆ O ₂	169.10	11.85	100,109,123	[M + H] +	Lipids
L-Hydroxylysine	Hydlys	C ₆ H ₁₄ N ₂ O ₃	145.09	9.35	100,128,145	[M + H] +	Amino acids and peptides
N-Methyllysine	Metlys	C ₇ H ₁₆ N ₂ O ₂	161.13	15.86	144,161	[M + H] +	Amino acids and peptides
Quinolinic Acid	QuiA	C ₇ H ₅ NO ₄	168.01	4.54	124,150,151,	[M + H] +	Pyridines
Pyridoxamine	Pyrxan	C ₈ H ₁₂ N ₂ O ₂	169.09	9.95	133,151,152	[M + H] +	Pyridines
Kojic Acid	KojA	C ₆ H ₆ O ₄	143.03	2.79	113,142	[M + H] +	Pyrans
Salsolinol	Sals	C ₁₀ H ₁₃ NO ₂	180.08	9.20	117,137,145,163	[M + H] +	Alkaloid
Naringenin	Nari	C ₁₅ H ₁₂ O ₅	271.05	5.98	151,205,241,271	[M-H] -	Flavonoids
Indole-3-acetamide	I3ACE	C ₁₀ H ₁₀ N ₂ O	173.07	5.49	130,173	[M-H] -	Indoles and derivatives
Divarinol	Diva	C ₉ H ₁₂ O ₂	151.10	7.68	107,151	[M-H] -	Other compounds
Note: Compounds were annotated using m/z, retention time, fragmentation ions, adduct patterns, and spectral similarity. Detected metabolites included benzenoids, organic acids, lipids, amino acids, pyridines, pyrans, alkaloids, flavonoids, and indoles, reflecting biochemical signatures of contamination, active bioremediation, and restored soil function.							

3.2.4 Chemometric analyses reveal progressive biochemical recovery following remediation

Heatmap analyses of relative metabolite quantitation revealed that there was distinct patterns of metabolite abundance between the different treatments. Hierarchical clustering resolved the samples into contamination-associated and recovery-associated patterns (Fig. 8). PS and UB mainly consisted of several stress- and xenobiotic-associated metabolites, while BS and CNTR had more of nutrient cycling- and microbial-recovery-associated metabolites.

Two major biochemical modules were observed: (i) pollution-associated metabolites, including TPPO, fluorene, quinolinic acid, and nitrophenols, dominant in PS and UB; and (ii) recovery-associated metabolites, including citric acid, aconitic acid, ectoine, lipoic acid, pyridoxamine, and geranic acid, dominant in BS and CNTR.

Overall, chemometric analyses revealed a transition from the pollution-associated metabolome of PS to the metabolically active intermediate of UB and the recovered functional metabolome of BS. This shift involved loss of

aromatic pollutants and gain of organic acids, lipids, and nitrogenous compounds involved in microbial recovery and restoration of soil biochemical function.

3.3 Integrated omics analyses link geochemical restoration with biochemical recovery

Integrated analysis of six key soil variables (TPH, sulphur, chlorine, phosphorus, iron, and manganese) and ten representative metabolites showed strong coupling between physicochemical recovery and metabolomic change.

Contamination markers were positively associated with aromatic stress metabolites. Positive correlation was observed between TPH and fluorene ($r = 0.88$), TPPO ($r = 0.84$), quinolinic acid ($r = 0.79$), and 2,4-dinitrophenol ($r = 0.76$), but negative correlation was observed with citric acid ($r = -0.82$), pyridoxamine ($r = -0.78$), and aconitic acid ($r = -0.74$). Likewise, there was positive correlation between phosphorus and citric acid ($r = 0.86$), pyridoxamine ($r = 0.84$), aconitic acid ($r = 0.82$), and ectoine ($r = 0.79$), while iron was associated with lipoic acid ($r = 0.83$) and geranic acid ($r = 0.79$) (Fig. 9A). The ordination biplot (Fig. 9B) showed PS clustering with TPH, sulphur, chlorine, and aromatic metabolites, whereas BS aligned with phosphorus, iron, manganese, and restorative metabolites. UB occupied an intermediate position, while CNTR aligned with silicon. Procrustes analysis confirmed close agreement between physicochemical and metabolomic patterns, with lower residual distances for BS (~ 0.22) and CNTR (~ 0.32) than for UB (~ 0.67) and PS (~ 0.72) (Fig. 9C). Overall, the 70.1% reduction in TPH was accompanied by depletion of petroleum-associated metabolites and enrichment of organic acids, lipid metabolites, and nitrogen-containing compounds linked to restored soil ecological function.

4.0 Discussion

Traditionally, remediation of petroleum-contamination has been assessed by TPH removal, toxicity tests or greening of vegetation (Dey et al. 2024; Meištinkas et al. 2024; Okoye et al. 2024). While these approaches are important and insightful, they do not necessarily indicate restoration of key soil functions, such as microbial resilience, nutrient cycling, carbon turnover and biochemical stability (Varjani, 2017; Das et al. 2021). In many pollution scenarios, contaminant loads can be reduced without removing the functional or metabolic bottlenecks that limit microbial activities (Sharma et al. 2019; Das et al. 2021). This study attempts to overcome this limitation by combining geochemical monitoring with untargeted metabolomics (LC-Mass spectrometry) to understand the recovery of the ecosystem in polluted soil (PS), soil undergoing bioremediation (UB), bioremediated soil (BS) and uncontaminated control soil (CNTR). Our results demonstrate that remediation outcome is more appropriately viewed as a multi-faceted shift that integrates geochemical normalization and biochemical adaptation, rather than just pollutant removal.

A key finding of this study was the significant reduction in petroleum loads during bioremediation. TPH concentrations declined from 78.26 ± 6.04 mg/kg in PS to 41.23 ± 1.40 mg/kg in UB and 23.38 ± 1.81 mg/kg in BS, corresponding to a 70.1% reduction relative to the polluted state. Such reduction is in line with previous studies demonstrating that microbial processes can significantly reduce hydrocarbons via oxidation, co-metabolism or mineralisation (Patel et al. 2019; Kong et al. 2023; Wang et al. 2025; Zhang et al. 2023). More so, the data from this study also show that the elimination of hydrocarbons was accompanied by other signs of ecological recovery, meaning that remediation did not merely dilute or move contaminants around, but was accompanied by changes in community structure. This is important given that this level of residual hydrocarbons could mask latent impairment in contaminated soils (Li et al. 2023).

The geochemical analysis showed a clear switch-off of petroleum-related stress. Contaminated soils were heavily enriched in sulphur (6.36 ± 0.70 wt%), chlorine and calcium, which all correlated with TPH in network analyses and were strongly weighted on the contamination axis of the PCA analysis. Studies have demonstrated that higher sulphur in petroleum-contaminated soils may be attributed to sulphur-bearing hydrocarbons and organosulfur compounds present in crude oil, such as thiophenes, benzothiophenes, and dibenzothiophenes (Rout and Behera, 2024; Mamuad and Choi, 2023), resulting in a change in soil redox status, or the build-up of reduced sulphur under anaerobic conditions (Patowary et al. 2023). Our findings revealed that sulphur concentrations significantly dropped to 0.152 ± 0.052 wt% in BS, which was not significantly different from CNTR (0.199 ± 0.010 wt%), suggesting near complete restoration of this contaminant indicator. This normalisation of sulphur may reflect enhanced soil aeration and microbial oxidation, which are crucial for long-term soil recovery (Kacprzak et al. 2026; Sui et al. 2021).

In addition to the decrease in contaminant indicators, remediated soils were highly enriched in variables related to nutrient and minerals. Phosphorus increased from 0.068 ± 0.006 wt% in PS to 0.325 ± 0.025 wt% in BS, exceeding the reference soil value (0.253 ± 0.012 wt%). Iron similarly increased from 4.03 ± 0.50 wt% in PS to 8.00 ± 0.70 wt% in BS. Nutrient availability is often compromised by petroleum spills due to microbial immobilization of nitrogen and phosphorus resulting from excess carbon inputs, as well as the inability of roots and microbes to access mineral phases due to water repellent residues (Das et al. 2021; Li et al. 2023). This nutrient rebound thus likely reflects re-activation of phosphorus cycling, mineral dissolution and microbial nutrient cycling. The iron increase may also reflect reactivation of iron cycling and iron cofactor-dependent enzymes for aerobic metabolism. These shifts in turn suggest that bioremediation can not only remove contaminants, but also restore geochemical processes vital for fertility (Haroun et al. 2023; Joshi et al. 2025; Pande et al. 2020).

Multivariate analyses strongly supported a directional recovery model. Multivariate analysis of physicochemical data (PCA) accounted for 99.3% of variance and discriminated samples along a major contamination–recovery gradient. The UB was positioned between the two contrasting groups of PS and BS, while polluted soils were separated from remediated and control soils. This intermediate position is important as it implies that ecosystem recovery is not an all or none process but occurs in gradually in intermediate phases. Similar trends have been observed in terrestrial restoration, where ecosystems gradually approach reference conditions in terms of community structure and biogeochemistry (Butu et al. 2021; Kakde and Sharma, 2024). However, Atkinson et al. (2022) reported that terrestrial ecosystem restoration increases biodiversity and reduces its variability, but not to reference levels, based on their global meta-analysis study. In this study, the ordered sequence (PS → UB → BS → CNTR) suggests that multivariate soil chemistry can be used to track recovery and may be earlier indicators than endpoint contaminant levels.

Untargeted metabolomics also offers an additional, more functionally informative view of recovery. PCA of the LC–MS features accounted for 73.7% of the total biochemical variance and distinguished all treatments, confirming that petroleum contamination and remediation are associated with metabolomic gradients. Contaminated and transitional soils were rich in fluorene, triphenylphosphine oxide (TPPO), quinolinic acid and 2,4-dinitrophenol. Fluorene and other benzenoids are commonly known as petroleum aromatic compounds and potentially remain as residues or byproducts in polluted environments (Ali et al. 2025; Pal and Sen 2024; Sahoo et al. 2020). Quinolinic acid may reflect a stress-associated nitrogen metabolism and microbial catabolism, while nitrophenolic compounds are associated with oxidative or xenobiotic metabolism. The presence of these metabolites in UB confirms that ongoing remediation soils show biochemical evidence of stress while they are recovering.

In contrast, metabolites of active microbial metabolism and central metabolites including citric acid, aconitic acid, ectoine, pyridoxamine, lipoic acid and geranic acid) were found in bioremediated and reference soils. Citric acid and

aconitic acid are intermediates in the tricarboxylic acid cycle and are often found in active respiration and carbon cycling (Gupta and Gupta 2021; Wang and Lou 2024). They indicate a renewed heterotrophic metabolism and substrate use efficiency. Ectoine is a compatible solute produced by various bacteria under osmotic stress or other environmental challenges, but it can also aid persistence and survival in dynamic soils (Nazarov et al. 2022). Pyridoxamine, a vitamin B6 derivative, is associated with amino acid metabolism and oxidative stress. Lipoic acid is a cofactor for oxidative decarboxylations and is an antioxidant. Collectively, these metabolites indicate that remediated soils had transitioned from emergency detoxification pathways toward stable biosynthetic and energy-yielding metabolism (Arbona et al. 2013; Mu et al. 2025).

The integrated omics analyses may have revealed the most compelling evidence that geochemical and biochemical recovery were coupled. TPH correlated positively with fluorene ($r = 0.88$), TPPO ($r = 0.84$), quinolinic acid ($r = 0.79$), and 2,4-dinitrophenol ($r = 0.76$), but negatively with citric acid ($r = -0.82$) and pyridoxamine ($r = -0.78$). By contrast, phosphorus had a positive correlation with citric acid ($r = 0.86$), pyridoxamine ($r = 0.84$), aconitic acid ($r = 0.82$), and ectoine ($r = 0.79$). These antagonistic modules suggest that as carbon loads decrease, nutrient-related processes increase. This pattern aligns with ecological theory: contamination imposes energetic costs on microbial communities, diverting metabolism toward stress tolerance and hydrocarbon catabolism (Mohapatra and Phale, 2022; Khan, 2025), whereas recovery allows renewed allocation to growth, respiration, and cooperative nutrient cycling (Varjani, 2017; Kong et al. 2023).

A key significant outcome of this research is that metabolite profiles could serve as early indicators of remediation success. Existing remediation standards tend to focus on TPH levels as a measure of recovery. But two soils with the same TPH may vary considerably in terms of their resilience, nutrient cycling and microbial activity. In this study, the composite recovery index scored BS slightly higher than the uncontaminated control (1.06 vs 1.00) due to increased phosphorus and micronutrient abundances. This does not mean that the remediated soil was "more natural" than the reference, but rather that it was in a post-remediation phase with high nutrient levels and intense biological activity. Thus, functional metabolites like ectoine, pyridoxamine and citric acid may supplement regulatory limits on contaminant levels to distinguish between chemically improved soils and recovering ecosystems. These insights are particularly important for tropical petroleum-producing regions like the Niger Delta, where long-term degradation of farmlands and wetlands has occurred due to chronic oil pollution, but post-remediation monitoring is often absent (Chikere et al. 2021; Nwosu et al., 2024; Patowary et al. 2023). Here, visible plant growth or decreased odour may be misinterpreted as a successful recovery while biochemical dysfunction remains. The current study shows that molecular-level indicators can uncover the "invisible" dynamics of restoration. Given that metabolites are the ultimate products of microbial processes, environmental constraints and substrate availability, metabolomics might be especially useful in tropical soils where other approaches to relate taxonomic diversity to ecosystem function are challenging.

Several limitations should be considered. First, the work is a geographical system with limited replicates, which could limit the ability to extrapolate to other types of soils, climates, or remediation approaches. Second, metabolite identifications were tentative (MSI level 2), and targeted experiments using standards would provide more confidence in candidate biomarkers. Third, correlation networks and Procrustes analyses confirm strong agreement among data sets, but do not demonstrate causal links between individual microbes, metabolites and elemental processes. Integration of metagenomics, metatranscriptomics, enzyme activity assays, and sampling at multiple sites over time will offer greater insight (Mu et al. 2025).

5.0 Conclusion

This study demonstrated that ecological recovery of tropical soils following bioremediation of petroleum pollution involved more than just removing the hydrocarbons. Petroleum hydrocarbons (70.1%) and sulphur (97.6%) decreased, while phosphorus (4.8-fold) and iron (1.7-fold) increased, revealing nutrient recovery. Physicochemical and metabolomic profiles distinguished the shift from contamination to recovery, with primary aromatic stress metabolites in contaminated soils and citric acid, pyridoxamine and ectoine in remediated soils - related to respiration, antioxidant and protection mechanisms. The remediated soil recovery index (1.06) was higher than the natural control (1.00), reflecting post-restoration activity. These findings demonstrate that metabolomic signatures are effective next-generation metrics for ecosystem restoration from oil pollution.

In summary, this study shows that successful remediation of oil-contaminated soils cannot be evaluated by hydrocarbon reduction alone. Positive recovery entailed parallel mitigation of contaminants, balance of nutrients and minerals, and emergence of a metabolome rich in adaptive osmolytes, organic acids, and cofactor-related compounds indicative of healthy function. Integrating geochemical and metabolomic indicators enabled recovery to be assessed as a progressive process rather than a single post-remediation endpoint. This approach provides a scalable framework for next-generation contaminated land management by distinguishing contaminant attenuation from restoration of ecological function.

Declarations

Author Contribution

Author a: Wrote the main manuscript text, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Author b & c : Supervised the work Author d & e: Revised it critically for important intellectual content, helped in interpretation of data, approved the version to be published Author f & g: Guided me in the lab work, analysis and interpretation of data All authors reviewed the manuscript

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information

References

1. Adam G, Duncan H (2002) Influence of diesel fuel on seed germination. *Environmental Pollution* 120:363–370.
2. Ahammed GJ, Wang MM, Zhou YH, Xia XJ, Mao WH, Shi K, Yu JQ (2012) The growth, photosynthesis and antioxidant defense responses of five vegetable crops to phenanthrene stress. *Ecotoxicology and Environmental Safety* 80:132–139.
3. Ali SR, Anwar Y, Ali HM (2025) Environmental impacts, health risks, and biodegradation strategies of fluorene and other polycyclic aromatic hydrocarbons. *Journal of Pure and Applied Microbiology* 19:2429–2440
4. American Petroleum Institute (API) (2010) *Specification for Drilling Fluids Materials*, API Specification 13A, 18th edn. American Petroleum Institute, Washington, DC.
5. American Public Health Association, American Water Works Association, Water Environment Federation (2017) *Standard methods for the examination of water and wastewater*, 23rd edn. APHA Press, Washington, DC

6. Arbona V, Manzi M, de Ollas C, Gómez-Cadenas A (2013) Metabolomics as a tool to investigate abiotic stress tolerance in plants. *International Journal of Molecular Sciences* 14:4885–4911. <https://doi.org/10.3390/ijms14034885>
7. ASTM International (2022) *ASTM E1621-22: Standard Guide for Elemental Analysis by Wavelength Dispersive X-Ray Fluorescence Spectrometry*. ASTM International, West Conshohocken, PA.
8. Atkinson J, Brudvig LA, Mallen-Cooper M, Nakagawa S, Moles AT, Bonser SP (2022) Terrestrial ecosystem restoration increases biodiversity and reduces its variability, but not to reference levels: a global meta-analysis. *Ecology Letters* 25:1725–1737
9. Butu M, Sarac I, Corneanu M, Butnariu M (2021) Advanced technologies for ecological reconstruction and bioremediation of degraded land. In: *Environmental pollution and remediation*. Springer, Singapore, pp 81–130
10. Chikere CB, Mordi IJ, Chikere BO, Selvarajan R, Ashafa TO, Obieze CC (2019) Comparative metagenomics and functional profiling of crude oil-polluted soils in Bodo West Community, Ogoni, with other sites of varying pollution history. *Annals of Microbiology* 69:495–513
11. Chikere CB, Tekere M, Adeleke R (2021) Microbial communities in field-scale oil-polluted soil remediation using 16S rRNA amplicon sequencing. *International Journal of Environmental Studies* 78:410–426
12. Das AJ, Ambust S, Singh T, Kumar R (2021) Biosurfactant-assisted design treatments for remediation of petroleum-contaminated soil and metabolomics-based interactive study with *Brassica nigra* L. *Environmental Challenges* 4:100080. <https://doi.org/10.1016/j.envc.2021.100080>
13. Dey S, Das A, Mallick K, Sahu A, Das AP (2024) Environmental petroleum waste: pollution, toxicity, sustainable remediation. In: *Impact of petroleum waste on environmental pollution and its sustainable management through circular economy*. Springer Nature, Cham, pp 159–175
14. Ehis-Eriakha CB, Ajuzieogu CA, Orogwu JO, Akemu SE (2025) Overview of petroleum hydrocarbon pollution and bioremediation technologies. *Bioremediation Journal* 29:238–260
15. Gupta R, Gupta N (2021) Tricarboxylic acid cycle. In: *Fundamentals of bacterial physiology and metabolism*. Springer, Singapore, pp 327–346
16. Haroun M, Xie S, Awadelkareem W, Wang J, Qian X (2023) Influence of biofertilizer on heavy metal bioremediation and enzyme activities in the soil revealing the potential for sustainable soil restoration. *Scientific Reports* 13:20684
17. Joshi SR, Verma P, Banerjee S (eds) (2025) *Mineral Transformation and Bioremediation by Geo-Microbes*. Springer, Singapore. <https://doi.org/10.1007/978-981-96-0657-9>
18. Kacprzak M, Kaczmarek S, Kupich I (2026) Comparative assessment of organic–mineral fertilizer, biochar, and vermiremediation for petroleum-contaminated alkaline soils. *Land Degradation & Development*. Advance online publication. <https://doi.org/xxxxx>.
19. Kakde P, Sharma J (2024) Microbial bioremediation of petroleum contaminated soil: structural complexity, degradation dynamics and advanced remediation techniques. *Journal of Pure and Applied Microbiology* 18:2244–2261
20. Kanungo J, Sahoo T, Swain LP, Behera ID (2024) Toxicity of persistent hydrocarbon pollutants, sources and sustainable remediation process. In: *Impact of petroleum waste on environmental pollution and its sustainable management through circular economy*. Springer Nature, Cham, pp 39–65
21. Khan MF (2025) Microbial remediation of agrochemical-contaminated soils: enzymatic mechanisms, quorum sensing, and emerging opportunities. *Integrated Environmental Assessment and Management*. Advance online publication. <https://doi.org/10.1093/inteam/vjaf167>.

22. Kong D, Liu Z, Wu J (2023) Omics shed light on the mechanisms of petroleum-contaminated soil remediation: when biochar meets nitrogen. *Land Degradation and Development* 34:5109–5121. <https://doi.org/10.1002/ldr.4833>
23. Li Y-Q, Xin Y, Li C, Liu J, Huang T (2023) Metagenomics-metabolomics analysis of microbial function and metabolism in petroleum-contaminated soil. *Brazilian Journal of Microbiology* 54:935–947. <https://doi.org/10.1007/s42770-023-01000-7>
24. Mamuad RY, Choi AES (2023) Biodesulfurization processes for the removal of sulfur from diesel oil: a perspective report. *Energies* 16:2738
25. Meištinkas R, Vaškevičienė I, Piotrowicz-Cieślak AI, Krupka M, Žaltauskaitė J (2024) Sustainable recovery of the health of soil with old petroleum hydrocarbon contamination through individual and microorganism-assisted phytoremediation with *Lotus corniculatus*. *Sustainability* 16:7484
26. Mekonnen BA, Aragaw TA, Genet MB (2024) Bioremediation of petroleum hydrocarbon contaminated soil: a review on principles, degradation mechanisms, and advancements. *Frontiers in Environmental Science* 12:1354422
27. Mohapatra B, Phale PS (2022) Sustainable management of genotoxic aromatic pollutants: role of microbial traits in efficient bioremediation and resource recovery. *Current Pollution Reports* 8:477–497
28. Mu YQ, Song JB, Zhao M, Ren P, Liu HY, Huang X (2025) Integrative omics analysis of plant-microbe synergies in petroleum pollution remediation. *PeerJ* 13:e19396. <https://doi.org/10.7717/peerj.19396>
29. Nazarov AV, Anan'ina LN, Gorbunov AA, Pyankova AA (2022) Bacteria producing ectoine in the rhizosphere of plants growing on technogenic saline soil. *Eurasian Soil Science* 55:1074–1081
30. Nwosu NB, Chikere BC, Elizabeth B, Nwogu-Chigozie L, Ezebuio V (2024) Evaluating different agricultural wastes as carriers of hydrocarbon degrading bacterial inoculants. *International Journal of Research and Technopreneurial Innovations* 1:34–42
31. Okoye AU, Selvarajan R, Chikere CB, Okpokwasili GC, Mearns K (2024) Characterization and identification of long-chain hydrocarbon-degrading bacterial communities in long-term chronically polluted soil in Ogoniland. *Environmental Science and Pollution Research* 31:30867–30885. <https://doi.org/10.1007/s11356-024-33162-z>
32. Pal D, Sen S (2024) Emerging petroleum pollutants and their adverse effects on the environment. In: *Impact of petroleum waste on environmental pollution and its sustainable management through circular economy*. Springer Nature, Cham, pp 103–137
33. Pande V, Pandey SC, Sati D, Pande V, Samant M (2020) Bioremediation: an emerging effective approach towards environment restoration. *Environmental Sustainability* 3:91–103
34. Patel S, Homaei A, Patil S, Daverey A (2019) Microbial biosurfactants for oil spill remediation: pitfalls and potentials. *Applied Microbiology and Biotechnology* 103:27–37. <https://doi.org/10.1007/s00253-018-9434-2>
35. Patowary R, Devi A, Mukherjee AK (2023) Advanced bioremediation by an amalgamation of nanotechnology and modern artificial intelligence for efficient restoration of crude petroleum oil-contaminated sites. *Environmental Science and Pollution Research* 30:74459–74484. <https://doi.org/10.1007/s11356-023-27698-4>
36. Rout PG, Behera SK (2024) Evaluation of organosulfur compound degradation property of a noble bacterial strain of *Proteus mirabilis*. *Environmental Chemistry and Ecotoxicology* 6:223–228
37. Sahoo BM, Ravi Kumar BV, Banik BK, Borah P (2020) Polyaromatic hydrocarbons (PAHs): structures, synthesis and their biological profile. *Current Organic Synthesis* 17:625–640
38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504.

<https://doi.org/10.1101/gr.1239303>

39. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27:431–432
40. Sui X, Wang X, Li Y, Ji H (2021) Remediation of petroleum-contaminated soils with microbial and microbial combined methods: advances, mechanisms, and challenges. *Sustainability* 13:9267. <https://doi.org/10.3390/su13169267>
41. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA et al (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3:211–221
42. Tavella RA, da Silva Júnior FMR, Santos MA, Miraglia SGEK, Pereira Filho RD (2025) A review of air pollution from petroleum refining and petrochemical industrial complexes: sources, key pollutants, health impacts, and challenges. *ChemEngineering* 9:13
43. United States Environmental Protection Agency (USEPA) (2007) *Method 3550C: Ultrasonic Extraction*. U.S. Environmental Protection Agency, Washington, DC.
44. United States Environmental Protection Agency (USEPA) (2007) *Method 8015C: Nonhalogenated Organics Using GC/FID*. U.S. Environmental Protection Agency, Washington, DC.
45. Varjani SJ (2017) Microbial degradation of petroleum hydrocarbons. *Bioresource Technology* 223:277–286. <https://doi.org/10.1016/j.biortech.2016.10.037>
46. Wang R, Lou L (2024) The central role of the citric acid cycle in energy metabolism: from metabolic intermediates to regulatory mechanisms. *Biological Evidence* 14
47. Zhang NC, Dan A, Chao YQ, Li HY, Li C, Lin QQ et al (2023) Mechanism of polycyclic aromatic hydrocarbons degradation in the rhizosphere of *Phragmites australis*: organic acid co-metabolism, iron-driven, and microbial response. *Environmental Pollution* 327:121608

Figures

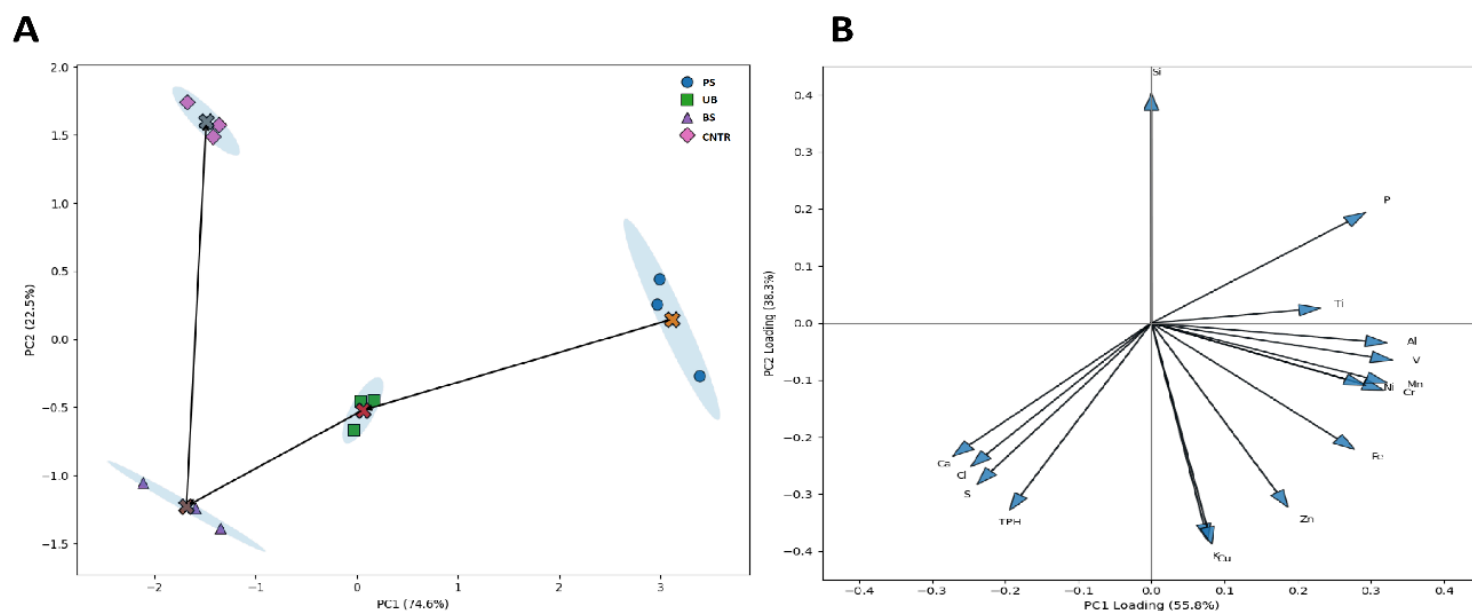


Figure 1

Principal component analysis (PCA) of physicochemical and elemental variables across treatments. (A) PCA score plot showing clear clustering of PS, UB, BS, and CNTR samples. PC1 and PC2 explained 75.9% and 23.4% of total variance, respectively (99.3% cumulative), indicating strong treatment separation and replicate consistency. **(B)** PCA loading plot showing that positive PC1 was associated with Al, V, Mn, Cr, Ni, Fe, Zn, Cu, Ti, and P, while negative PC1 was linked to TPH, S, Cl, and Ca. PC2 was mainly influenced by Si.

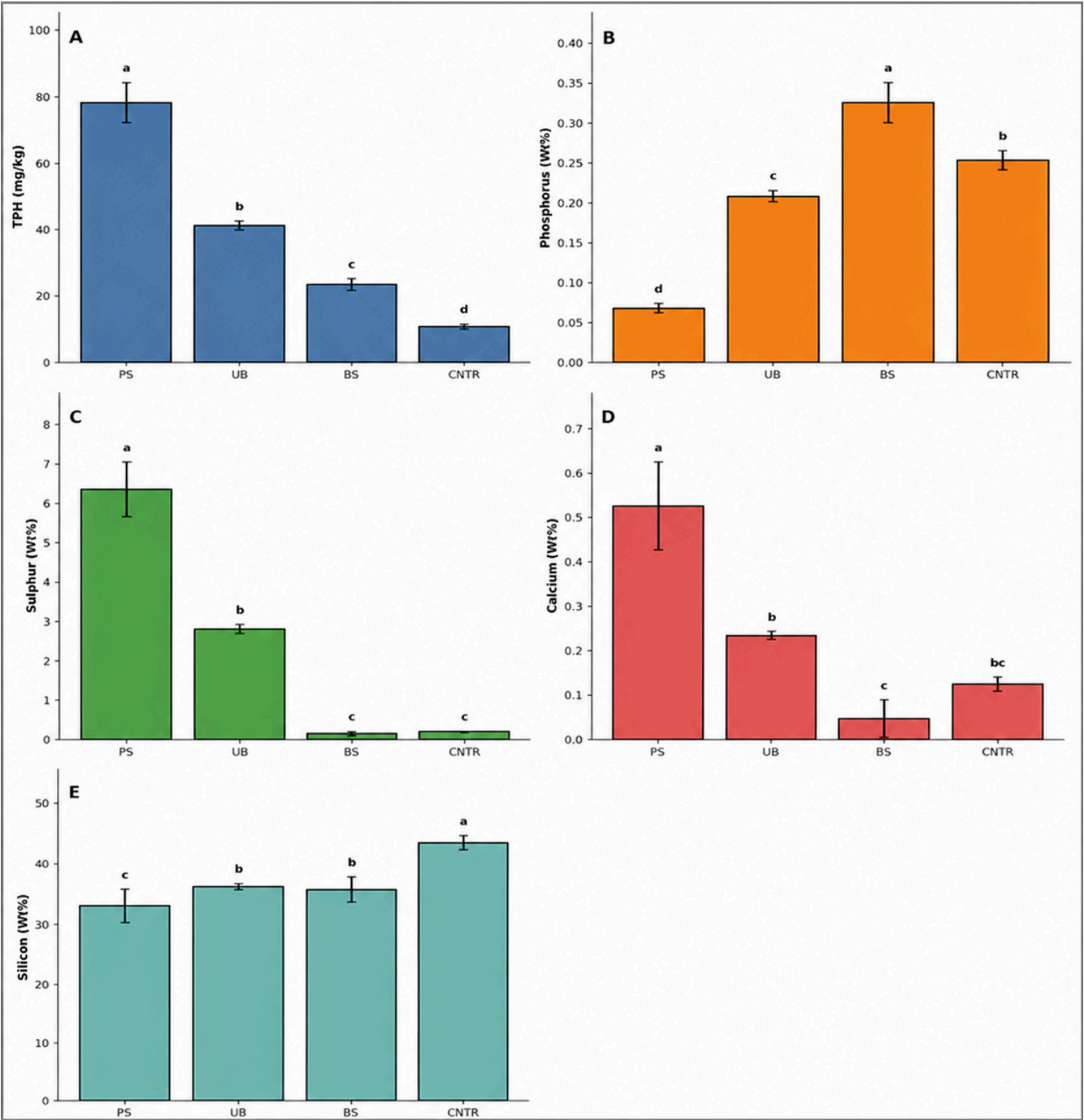


Figure 2

Multi-panel comparison of physicochemical parameters across treatment groups (PS, UB, BS, and CNTR). Panel **A** = Total petroleum hydrocarbon (TPH, mg/kg); **B** = Phosphorus (Wt%); **C** = Sulphur (Wt%); **D**= Calcium (Wt%); **E** = Silicon (Wt%). Bars represent mean $\pm$ standard deviation. Different superscript letters above bars indicate significant differences among treatments based on one-way ANOVA followed by post hoc multiple comparison test ($p < 0.05$).

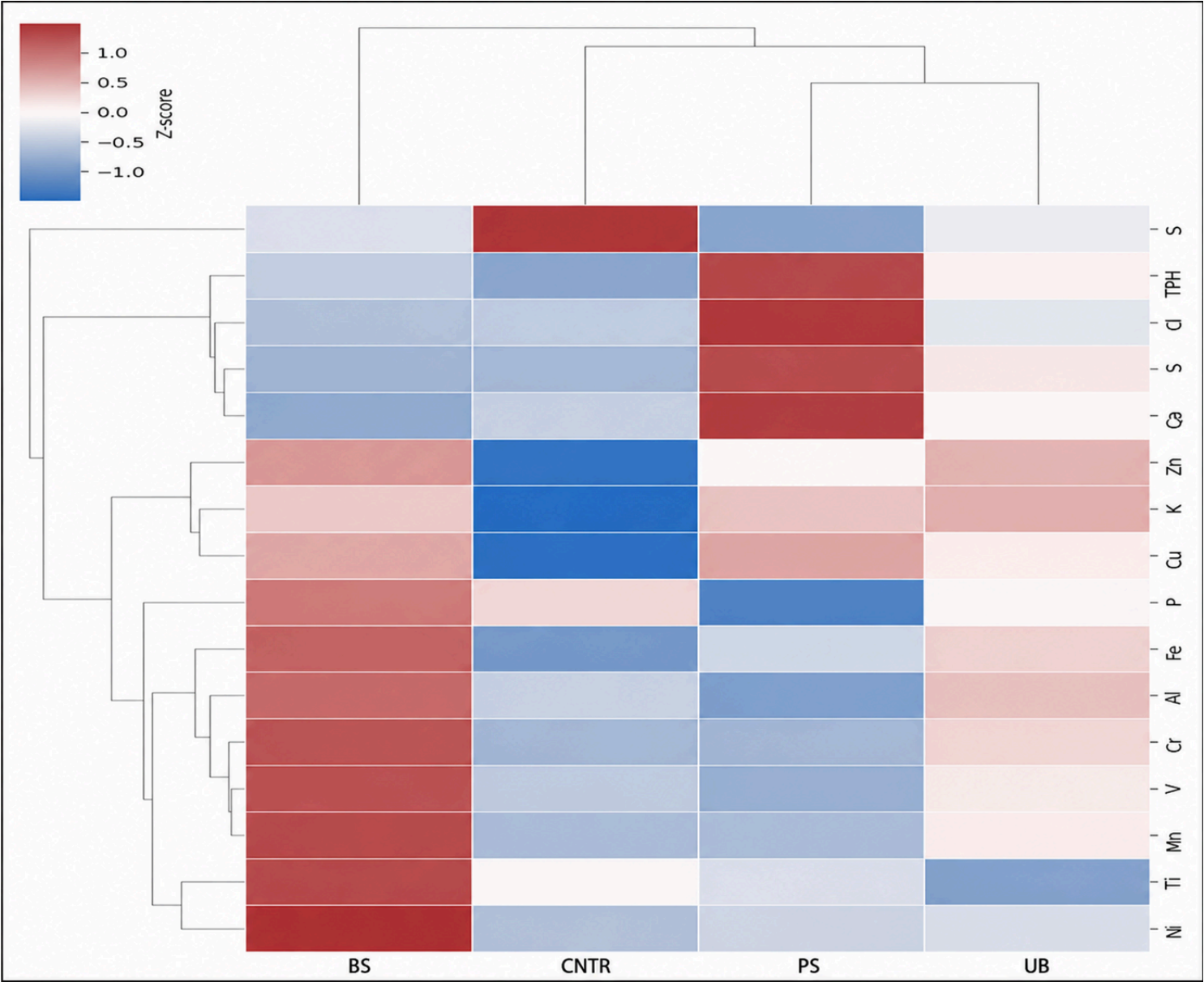


Figure 3

Clustered heatmap of normalized physicochemical signatures across soil recovery states. Hierarchical clustering of standardized physicochemical variables showed distinct geochemical profiles among BS, CNTR, PS, and UB soils. Polluted soil was enriched in TPH, sulphur, chlorine, and calcium, whereas BS showed higher levels of recovery-associated variables including phosphorus, iron, aluminum, chromium, vanadium, manganese, titanium, and nickel. UB displayed an intermediate profile, indicating progressive physicochemical recovery during remediation.

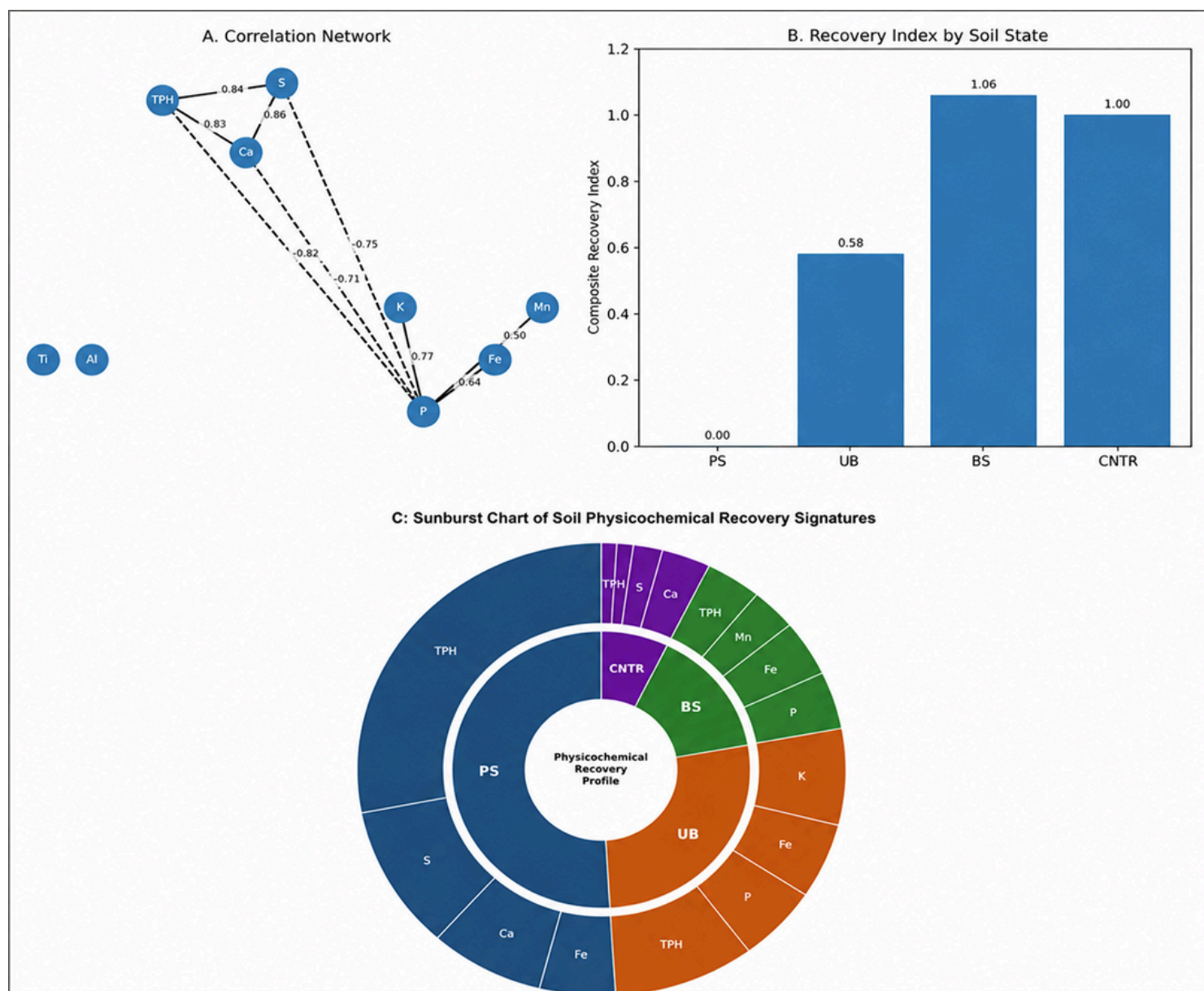


Figure 4

Correlation network and composite recovery index of soil physicochemical restoration. (A) Correlation analysis showed TPH as the main contamination hub with positive association with sulphur and chlorine, whereas phosphorus clustered with recovery-related variables including iron, manganese, and potassium. Negative TPH-phosphorus relationships indicate that nutrient restoration increased as hydrocarbon burden declined. (B) The composite recovery index showed progressive improvement from PS (0.00) to UB (0.58) and BS (1.06), with BS slightly exceeding the uncontaminated control (CNTR; 1.00). (C) Sunburst chart summarizing hierarchical physicochemical signatures across treatments, with PS dominated by TPH and sulphur, UB showing transitional redistribution, BS enriched in recovery-associated variables (phosphorus, iron, manganese), and CNTR characterized mainly by native mineral components, particularly silicon and aluminum. Together, the panels demonstrate coordinated attenuation of contamination and restoration of soil geochemical function.

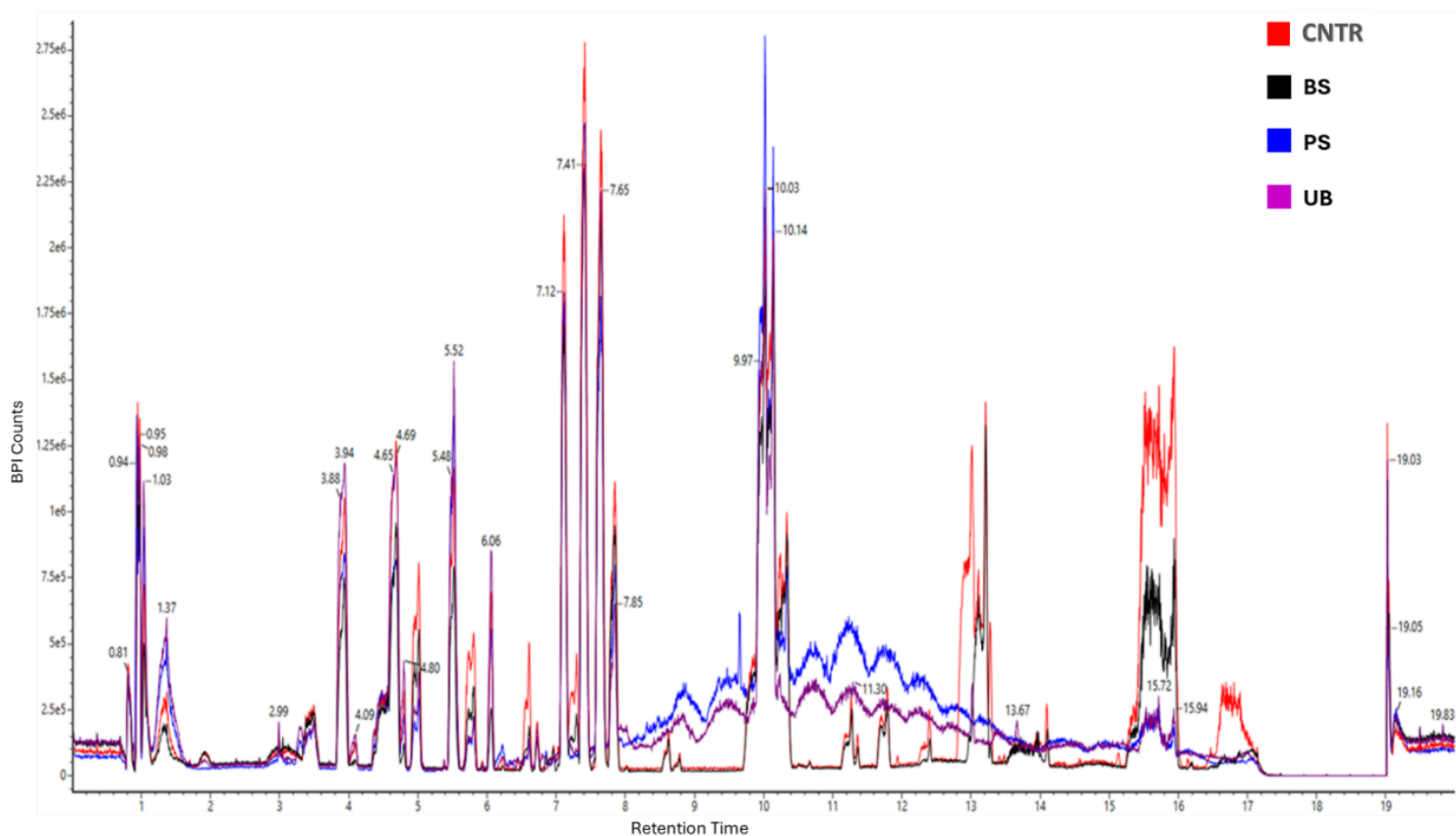


Figure 5

Overlay total ion chromatograms (TICs) of metabolite extraction profiles across soil treatment groups.

Superimposed LC–MS chromatograms revealed distinct treatment-dependent metabolite profiles among PS, UB, BS, and CNTR soils, with clear differences in peak intensity and retention time patterns. Major peaks at 6–10 min and 15–17 min indicate dynamic changes in polar and semi-hydrophobic metabolites during soil recovery.

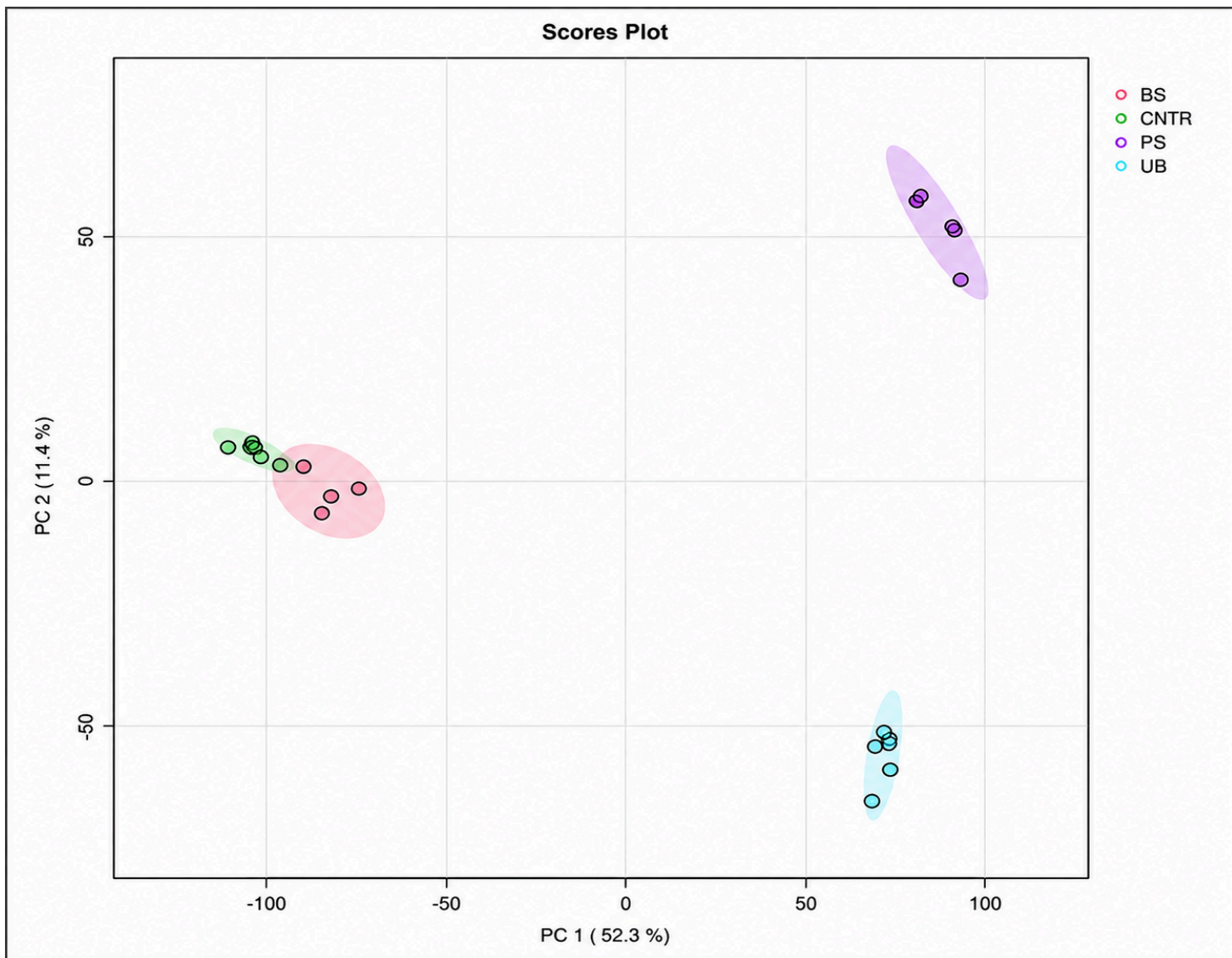


Figure 6

Principal component analysis (PCA) scores plot of soil metabolomic profiles across treatment groups. PCA of LC–MS metabolomic features showed clear separation among BS, CNTR, PS, and UB soils. PC1 (52.3%) and PC2 (11.4%) explained 63.7% of total variance, with tight replicate clustering indicating strong reproducibility. UB formed a distinct intermediate cluster, which is consistent with an active transitional biochemical state during remediation.

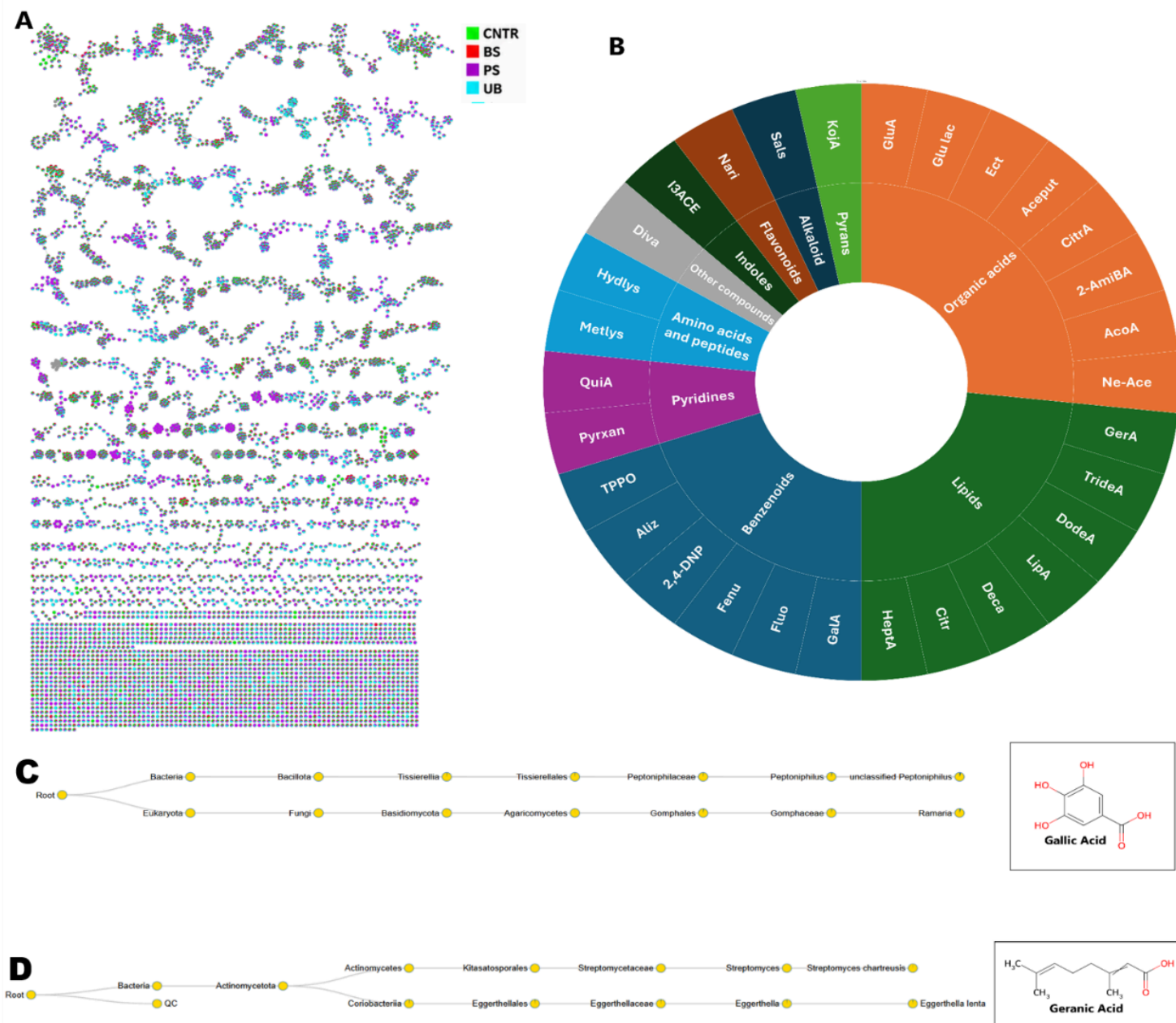


Figure 7

Molecular networking and chemical class distribution of soil metabolomes across treatments. (A) LC-MS/MS molecular networking showed shared and treatment-specific metabolites among CNTR, BS, PS, and UB soils. **(B)** Major annotated classes included organic acids, lipids, benzenoids, pyridines, amino acids/peptides, alkaloids, flavonoids, and indoles. **(C-D)** MicrobeMASST linked gallic acid and geranic acid to bacterial and fungal taxa, suggesting microbial origins and roles in biodegradation during soil remediation.

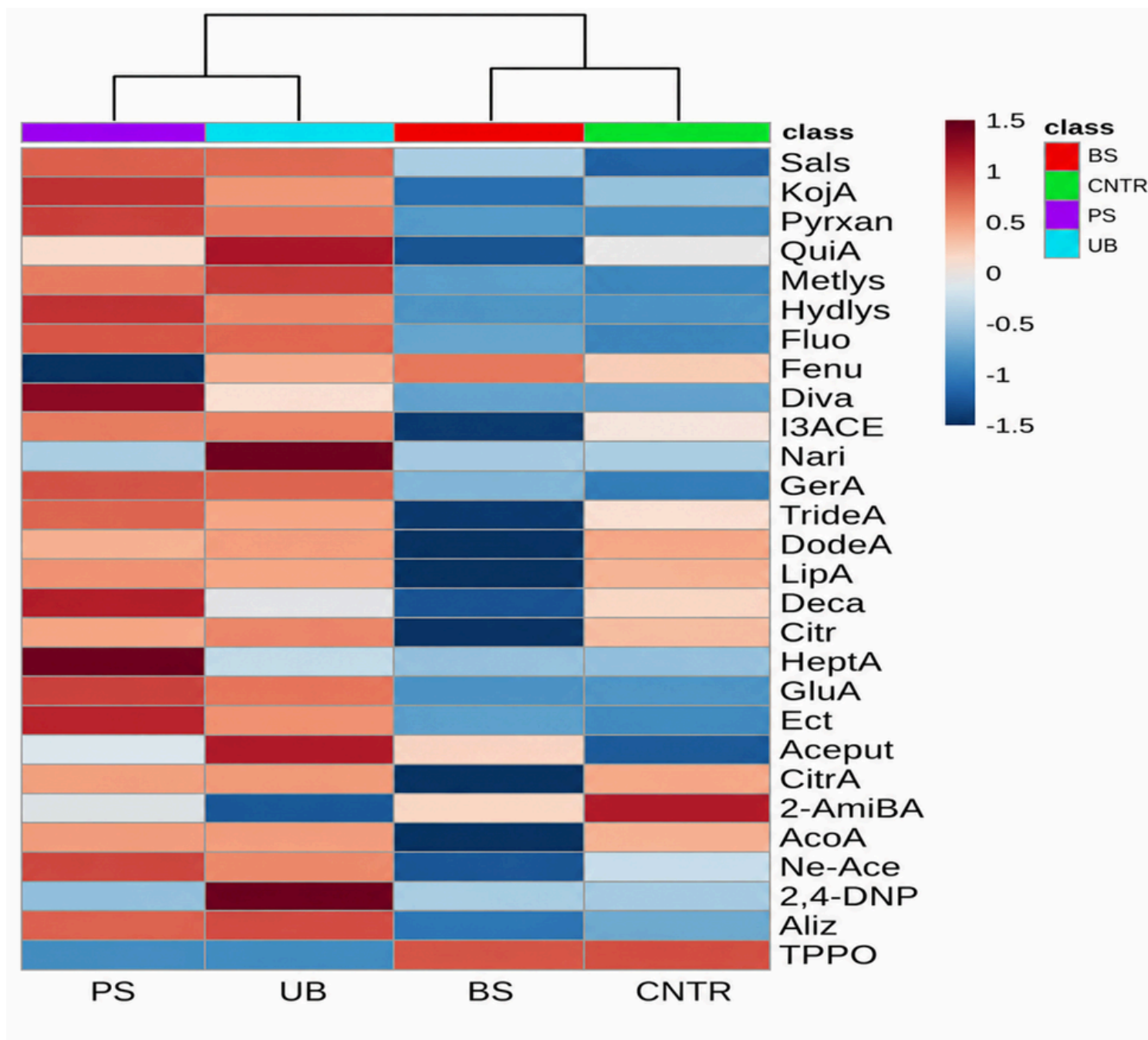


Figure 8

Clustered heatmap of annotated metabolite abundance across soil treatment groups. Hierarchical clustering of normalized metabolite intensities showed distinct biochemical profiles among PS, UB, BS, and CNTR soils. PS and UB consisted mostly of aromatic/stress-associated metabolites (e.g., TPPO, alizarin, 2,4-dinitrophenol, fluorene, quinolinic acid), whereas BS and CNTR were rich in recovery-associated metabolites (including citric acid, aconitic acid, ectoine, lipoic acid, geranic acid, and pyridoxamine).

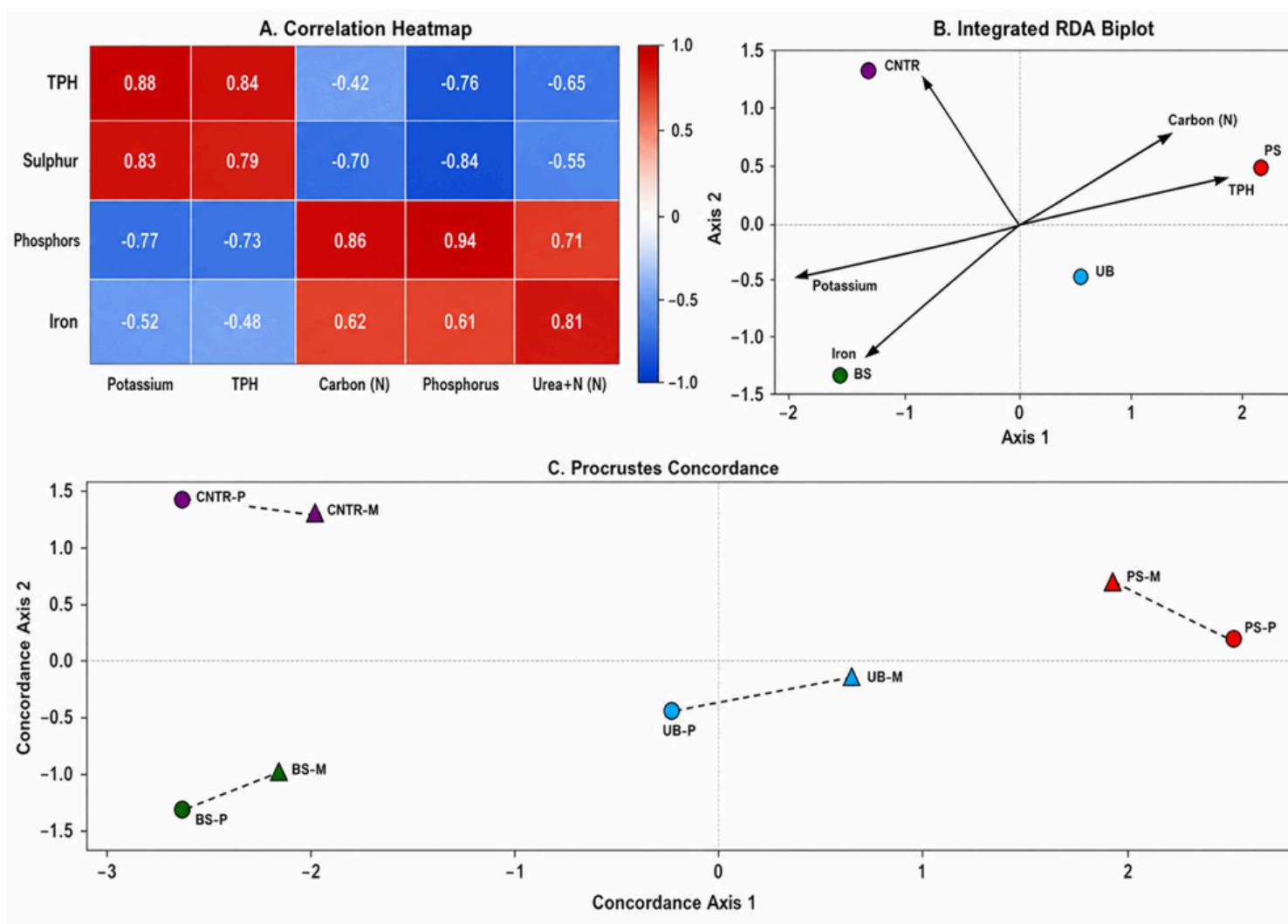


Figure 9

Integrated multi-omics model of soil recovery linking physicochemical restoration with metabolomic restructuring. (A) Contamination markers (TPH, sulphur) were associated with aromatic stress metabolites, whereas recovery variables (phosphorus, iron) aligned with restorative metabolites. (B) Ordination separated PS, UB, BS, and CNTR along a clear contamination–recovery gradient. (C) Procrustes concordance analysis comparing physicochemical (P) and metabolomic (M) ordinations across soil treatments. CNTR, uncontaminated control soil; BS, bioremediated soil; UB, soil undergoing bioremediation; PS, polluted soil. Dashed lines connect paired positions from each dataset, where shorter distances indicate stronger agreement between geochemical and metabolic recovery patterns.