

Metabolomic signatures of functional recovery during petroleum bioremediation in a Niger Delta farmland

Emmanuel Oliver Fenibo

feniboe1478@gmail.com

Hensard University

Matambo Sylvester Tonderayi

University of South Africa

Research Article

Keywords: Bioremediation monitoring, Hydrocarbon-degradation, Niger Delta, Pathway enrichment analysis, Soil metabolomics

Posted Date: July 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10402852/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Functional recovery of petroleum-contaminated soils remains poorly characterized, as remediation success is often judged by contaminant reduction rather than ecosystem-level biochemical responses. This study applied integrated liquid chromatography-mass spectrometry (LC-MS) and ^1H nuclear magnetic resonance (^1H -NMR) metabolomics to track metabolic restructuring during field-scale bioremediation of a polluted farmland in the Niger Delta, Nigeria. Soil samples were collected over 185 days from uncontaminated reference soil, untreated contaminated soil, tillage-watered contaminated soil, and poultry-manure-amended contaminated soil. Integrated analyses generated 266 spectral features, including confidently and putatively annotated compounds associated with osmolytes, organic acids, aromatic signals, and lipid-associated resonances. Principal component analysis and PERMANOVA showed significant treatment- and time-associated metabolomic restructuring ($p < 0.001$). Soils undergoing remediation showed higher relative abundance of branched-chain amino acids, glutamic acid, alanine, serine, and nucleotide-related metabolites, whereas untreated contaminated soils showed stronger representation of aromatic and stress-associated metabolites. Pathway enrichment indicated significant representation of amino acid metabolism, central carbon metabolism, and hydrocarbon degradation pathways. These findings demonstrate that bioremediation is accompanied by substantial metabolic reorganization, highlighting metabolomics as a powerful complement to contaminant-based assessments for evaluating functional recovery in petroleum-impacted ecosystems.

1. INTRODUCTION

Petroleum hydrocarbon contamination remains one of the most pervasive environmental challenges for terrestrial ecosystems worldwide, particularly in oil-producing regions where exploration, transport, refining, and accidental spills are frequent. In soil environments, petroleum hydrocarbons alter physicochemical properties, impair nutrient cycling, reduce agricultural productivity, and impose oxidative and osmotic stress on resident microbial communities. These disturbances disrupt ecosystem functioning and compromise the soil's capacity to sustain essential ecological processes. The Niger Delta region of Nigeria is among the most severely affected petroleum-producing environments globally (Fenibo et al., 2024), where recurrent oil spills and artisanal refining activities have led to widespread and persistent soil contamination.

Bioremediation strategies such as tillage, nutrient amendment, and controlled irrigation are widely used to stimulate indigenous hydrocarbon-degrading microorganisms and accelerate contaminant removal. Traditionally, remediation success has been evaluated using concentration-based endpoints, including total petroleum hydrocarbon (TPH) reduction and changes in soil physicochemical properties. While these metrics provide valuable information on contaminant attenuation, they offer limited insight into the extent of biological recovery and the restoration of ecosystem functionality (Seo et al., 2024). Increasing recognition of this limitation has stimulated interest in functional indicators that reveal the biochemical and physiological responses of soil ecosystems during remediation.

Metabolomics has emerged as a powerful systems-level approach for characterising the small-molecule metabolites that represent the functional endpoint of biological activity. Unlike genomic or proteomic analyses, which indicate biological potential, metabolomics provides a direct snapshot of ongoing biochemical processes and cellular responses to environmental stress. Consequently, metabolomic profiling can reveal alterations in energy metabolism, nutrient assimilation, stress adaptation, and microbial activity that accompany ecosystem disturbance and recovery (Awan et al., 2026). Recent studies have demonstrated the utility of metabolomics for investigating microbial responses to hydrocarbon contamination and for identifying metabolic pathways associated with biodegradation. However, most investigations have been conducted under controlled laboratory conditions, where environmental variability is minimised and ecological complexity is substantially reduced.

Among available analytical platforms, liquid chromatography–mass spectrometry (LC–MS) offers high sensitivity for detecting a broad range of low-abundance metabolites, whereas proton nuclear magnetic resonance (^1H -NMR) spectroscopy provides robust, highly reproducible characterisation of abundant metabolites, osmolytes, and lipid-associated compounds (Choudhary et al., 2026). Integrating these complementary platforms expands metabolome coverage, enhances biological interpretation, and improves confidence in characterising ecosystem responses. Despite these advantages, field-scale applications of integrated LC–MS and ^1H -NMR metabolomics for assessing bioremediation-induced ecosystem recovery remain limited, particularly in chronically petroleum-polluted environments. Furthermore, although hydrocarbon degradation pathways are increasingly recognised as being interconnected with central carbon metabolism, amino acid biosynthesis, nucleotide turnover, and cellular stress-response mechanisms, the extent to which these coordinated metabolic responses occur under realistic field conditions remains poorly understood and emerging. Addressing this knowledge gap is essential for advancing metabolomics from a descriptive research tool towards a functional framework for environmental assessment and remediation monitoring.

Therefore, this study applied integrated LC–MS and ^1H -NMR metabolomics to evaluate metabolic responses during field-scale bioremediation of petroleum-contaminated agricultural soils in Ngia Ama, Niger Delta, Nigeria. We hypothesised that bioremediation would induce measurable shifts in soil metabolomic profiles and that integrated metabolomic analyses would reveal metabolic signatures associated with microbial recovery and the restoration of ecosystem function. Specifically, the study aimed to (i) characterise treatment- and time-dependent changes in soil metabolomes, (ii) identify metabolites associated with contamination and recovery states, and (iii) reconstruct metabolic pathways linked to hydrocarbon degradation and ecosystem restoration. By providing a functional assessment of soil recovery beyond conventional contaminant-based metrics, this work contributes to the development of metabolomics-informed approaches for evaluating environmental remediation outcomes.

2. MATERIALS AND METHODS

2.1 Study site and sampling design

The study was conducted on petroleum hydrocarbon-contaminated agricultural land at Ngia Ama, Rivers State, Niger Delta, Nigeria (4.7947° N, 6.6831° E). The site had a history of chronic contamination due to artisanal crude-oil refining activities. Field-scale bioremediation was implemented, including periodic tillage, watering, and nutrient amendment using poultry droppings. Four soil treatments were evaluated:

1. Uncontaminated reference soil (CS)
2. Contaminated, untreated soil (B30)
3. Contaminated soil subjected to tillage and watering (C)
4. Contaminated soil amended with poultry droppings in addition to tillage and watering (T)

Sampling occurred at baseline (Day 0) and during remediation on Days 28, 57, 91, and 185. Each treatment comprised three independent biological replicates from spatially separated plots to minimise pseudoreplication and capture field-scale heterogeneity. Within each replicate, multiple soil cores (0–15 cm depth) were collected using sterile augers and pooled to form representative composite samples. Samples were homogenised, sieved through a 2-mm mesh to remove debris, transported on ice, and stored at -80°C until metabolomic analysis. All procedures employed sterile equipment to preserve metabolite integrity.

2.2 Extraction and LC–MS analysis of primary metabolites

Primary metabolites were extracted from 1.0 g of homogenised soil using 5 mL of a 75:25 (v/v) methanol: ultrapure water mixture (LC–MS grade) (Mary Celin et al., 2023). Samples were vortexed for 2 min, then centrifuged at $12,000 \times g$ for 15 min at 4°C , and the supernatants were transferred to LC–MS vials. Metabolite profiling was performed on a Shimadzu LCMS-8040 triple quadrupole mass spectrometer with electrospray ionisation (ESI). Chromatography was conducted at 40°C , with a flow rate of 0.4 mL min^{-1} and an injection volume of $1 \mu\text{L}$, using an acetonitrile-water (50:50 v/v) mobile phase. Detection employed multiple reaction monitoring (MRM) for amino acids, nucleobases, nucleosides, organic acids, carnitines, and hydrocarbon-degradation-associated metabolites. Quality assurance included pooled QC samples injected at the start and after every five study samples, procedural blanks, and retention-time monitoring. Features with $> 30\%$ relative standard deviation (RSD) in QC samples were excluded. A randomised injection order minimised batch effects, and instrument calibration followed the manufacturer's guidelines.

2.3 PLS-DA modelling and validity analysis

Partial least squares discriminant analysis was performed on the log-transformed and autoscaled metabolite matrix. Treatment class was encoded as a one-hot response matrix, and PLS-DA models containing one to eight latent components were evaluated. Repeated stratified five-fold cross-validation was used for component assessment, preserving the treatment distribution in each fold. Model selection was based on the highest cross-validated Q^2 value, while avoiding unnecessary addition of latent

components after predictive performance declined. The selected model was then used to calculate fitted explanatory performance, cross-validated predictive performance, and class-assignment error.

Model validity was assessed using R^2Y , Q^2 , cross-validation accuracy, cross-validation error, and permutation testing. R^2Y measured the degree to which the fitted model explained treatment-class variation, while Q^2 measured predictive performance from cross-validated predictions. Cross-validation error was calculated as one minus cross-validation accuracy. To test whether the observed discrimination exceeded random class allocation, treatment labels were randomly reassigned 500 times, and the cross-validated Q^2 value was recalculated after each permutation. The empirical permutation probability was calculated from the proportion of permuted models that equalled or exceeded the observed Q^2 , using a plus-one correction.

2.4 Occurrence frequency, clustering and VIP analysis

Metabolite occurrence frequency was calculated as the percentage of analytical profiles in which the metabolite had a detected, non-zero peak area after importing the report data. Occurrence was summarised across the full sample set and apportioned by treatment class to show the contribution of B30, C, T and CS to total detection frequency. For the hierarchical clustering heatmap, selected primary metabolites were transformed using $\log_{10}(\text{peak area} + 1)$, row-scaled by Z-score, and clustered according to relative abundance patterns. The heatmap visualises relative metabolite enrichment or depletion across treatments and sampling order, not absolute peak intensity. Variable importance in projection scores was calculated from the selected PLS-DA model, and $VIP > 1.0$ was used as the screening threshold for above-average contribution to class discrimination.

2.5 Extraction and ^1H -NMR analysis of secondary metabolites

Secondary metabolites were extracted from 2.0 g of soil using 8 mL of deuterated methanol (CD_3OD , 99.8 atom % D). Samples were vortexed for 2 min, sonicated in an ice-cooled ultrasonic bath for 20 min, and centrifuged at $10,000 \times g$ for 15 min at 4°C . Extraction was repeated twice, and combined supernatants were filtered through $0.22 \mu\text{m}$ PTFE membranes. Extracts were concentrated under nitrogen and reconstituted in $600 \mu\text{L}$ CD_3OD containing 0.05% TSP (internal standard). Spectra were acquired on a Varian 600 MHz ^1H -NMR spectrometer at 298 K using a one-dimensional NOESY-presaturation pulse sequence. 128 scans per sample were collected with a spectral width of 12 ppm, an acquisition time of 2.65 s, and a relaxation delay of 4 s. Fourier transformation, phase and baseline correction, and referencing to TSP at 0.00 ppm were performed. Probabilistic quotient normalisation (PQN) and binning (0.01 ppm) preceded downstream analyses. Residual water and solvent regions were excluded (Ferreira et al., 2024).

2.6 Metabolomic data processing and metabolite annotation

LC–MS chromatograms were processed to generate peak area matrices. Zero values were replaced with one, and metabolites detected in < 80% of replicates were excluded. The remaining missing values were imputed using k-nearest-neighbour (kNN) methods. PQN normalisation, $\log_{10}$ transformation, and autoscaling were applied before analysis. NMR bins were normalised similarly and integrated with LC–MS datasets (Karaman, 2017). Metabolite annotation used KEGG, HMDB, METLIN, MassBank, MetaCyc, and NMRShiftDB. MSI Level 1 metabolites were confirmed with authentic standards; MSI Level 2 metabolites were database-supported annotations. Annotated metabolites from both platforms were merged into a unified dataset for statistical analysis, pathway reconstruction, and biomarker discovery. A complete list of annotated metabolites, identification criteria, database matches, and MSI confidence levels is provided in Table S1. Figure 1 displays the summary of the method section.

2.7 Multivariate and differential metabolite analyses

Preprocessed data were analyzed in MetaboAnalyst 6.0 and R v4.x.x. Global metabolomic variation was assessed by principal component analysis (PCA); treatment- and time-dependent differences were tested via PERMANOVA (999 permutations, Bray-Curtis dissimilarity). Hierarchical clustering and heatmaps used Euclidean distance and Ward's linkage. Linear mixed-effects models evaluated treatment, time, and treatment $\times$ time interactions:

$$Y_{ijk} = \mu + \alpha_i + S_j + (T \times S)_{ij} + b_k + \epsilon_{ijk}$$

where Y_{ijk} is metabolite abundance, μ is overall mean metabolite abundance, α_i is treatment, S_j is sampling time, $(T \times S)_{ij}$ is the treatment $\times$ time interaction, b_k is plot replicate treated as a random effect, and ϵ_{ijk} is the residual error (Cho et al., 2022).

Fixed effects were tested with Type III ANOVA; p-values adjusted via Benjamini–Hochberg FDR. Significant metabolites had FDR < 0.05 and $|\log_2 \text{fold-change}| \geq 1$. Candidate biomarkers were identified via differential abundance, frequency, fold-change, and PCA loadings.

2.8 Pathway enrichment and metabolic network reconstruction

Significant metabolites were mapped to **KEGG and MetaCyc pathways** using MetaboAnalyst 6.0. Over-representation analysis used the hypergeometric test; pathway topology analysis employed relative betweenness centrality (Ma et al., 2019). Pathways with FDR-adjusted $p < 0.05$ and impact > 0.10 were considered enriched. Integrated metabolic networks linked hydrocarbon degradation, central carbon metabolism, amino acid biosynthesis, nucleotide turnover, and lipid metabolism to ecosystem recovery.

2.9 Quality assurance and quality control

Biological triplicates, pooled QC samples, procedural blanks, and retention-time monitoring ensured analytical precision. Features with QC RSD > 30% were excluded. Tight clustering of QC samples in PCA confirmed instrument stability and reproducibility. Metabolite identities were verified across multiple databases and spectral resources before inclusion in downstream analyses.

3. RESULTS

3.1 Treatment- and time-dependent structuring of the soil metabolome

Integrated LC–MS and ^1H -NMR metabolomic profiling showed treatment-specific metabolic variation across the bioremediation gradient. Principal component analysis (PCA) accounted for 45.1% of the total variance, with PC1 and PC2 explaining 28.3% and 16.8%, respectively (Fig. 2). Untreated contaminated soils (B30) separated from the other sample groups, while the amended soils (C and T) and uncontaminated reference soils (CS) occupied distinct ordination regions. The close grouping of quality-control samples near the centre of the PCA plot supported acceptable analytical reproducibility. PERMANOVA showed significant effects of treatment ($R^2 = 0.317$, $p < 0.001$), sampling time ($R^2 = 0.196$, $p < 0.001$), and the treatment $\times$ time interaction ($R^2 = 0.142$, $p < 0.001$), indicating that remediation treatment and temporal progression were associated with measurable metabolomic restructuring.

3.2 LC–MS profiling of primary metabolites

Targeted LC–MS analysis identified 99 primary metabolites, including amino acids, nucleotides, organic acids, and carnitine derivatives. Branched-chain amino acids, particularly leucine and isoleucine, showed consistently high abundance across the treatment groups. Other abundant metabolites included glutamic acid, alanine, serine, adenosine, guanosine, and cytidine. Compared with untreated contaminated soils, amended treatments showed higher relative abundance of metabolites associated with amino acid metabolism, nitrogen assimilation, nucleotide turnover, and energy-related metabolism. Hierarchical clustering showed treatment-related abundance patterns, with glutamic acid and creatine enriched in selected early amended and remediated samples (C0, T91), while 4-hydroxyproline showed stronger signals in T0, C28, C57, and C185 (Fig. 3a). Organic acid intermediates, including malic acid, fumaric acid, citric acid, succinic acid, and 2-ketoglutaric acid, displayed more variable and less frequent abundance patterns. Visual gradients were most evident for adenosine, cytidine, alanine, serine, glutamic acid, 4-hydroxyproline, carnitine, and acetylcarnitine, suggesting that these metabolites contributed to the observed treatment-related metabolomic structure.

Occurrence analysis showed a recurrent set of metabolites across the dataset. Cytidine, leucine, isoleucine, adenosine, and guanosine were detected in all samples (100.0%) (Fig. 3b). Adenine and tyrosine were detected in 97.2% of samples, while glutamic acid and serine were detected in 94.4%. Alanine occurred in 91.7% of samples, and acetylcarnitine occurred in 88.9%. Carnitine showed a lower but still notable occurrence frequency of 69.4%. By contrast, hypoxanthine, 2-ketoglutaric acid, and malic acid were detected less frequently, occurring in 22.2%, 19.4%, and 8.3% of samples, respectively. This distribution indicates that the metabolomic dataset contained a consistently detected biochemical background, together with less frequent organic acid and purine-related signals that may reflect treatment- or time-associated metabolic variation.

Variable importance in projection (VIP) analysis ranked metabolites according to their relative contribution to the selected four-component PLS-DA model. As shown in Fig. 3c, adenosine recorded the highest VIP score (1.949), followed by cytidine (1.686), isocitric acid (1.633), ornithine (1.555), allantoin (1.522), choline (1.487), methionine sulfoxide (1.440), uracil (1.437), nicotinic acid (1.373), cytidine monophosphate (1.369), alanine (1.353), guanosine 3',5'-cyclic monophosphate (1.336), methionine (1.333), 4-hydroxyproline (1.322), and cytosine (1.312). All listed metabolites exceeded the VIP > 1.0 screening threshold, indicating an above-average contribution to class separation in the supervised model. The VIP profile suggests that nucleoside-related metabolites, amino-acid-associated metabolites, and organic-acid-related signals were among the main variables associated with treatment-related separation, and they should therefore be interpreted as candidate discriminant indicators instead of confirmed biomarkers.

3.3 ¹H NMR profiling of secondary metabolites

Untargeted ¹H-NMR profiling resolved 229 secondary metabolites and showed treatment-associated differences across the soil groups (Fig. S1). Choline, betaine, glycine, alanine, acetate, lactate, and lipid-associated resonances contributed more strongly to amended and reference soils. By contrast, citrate, succinate, aromatic signals, and phenolic signals showed stronger representation in untreated contaminated soils. These patterns suggest a remediation-associated shift from contamination-related stress metabolism towards metabolite profiles consistent with anabolic activity, osmotic adjustment, and membrane-related adaptation, although these responses should be interpreted as candidate functional indicators pending direct biological validation.

3.4 Multi-platform integration

Integration of the LC–MS and ¹H-NMR datasets yielded 266 annotated metabolites, comprising 62 metabolites detected by both platforms, 37 LC–MS-specific metabolites, and 167 NMR-specific metabolites, representing 23.3%, 13.9%, and 62.8% of the integrated metabolome, respectively (Fig. S2). Cross-platform correlation analysis ($R = 0.78$, $p < 0.001$) supported concordant treatment-associated metabolic trends across the two analytical platforms. Integrated pathway mapping associated remediated treatments with aminoacyl-tRNA biosynthesis, glycine-serine-threonine metabolism, glycolysis/gluconeogenesis, butanoate metabolism, nitrogen metabolism, and nicotinate-nicotinamide metabolism. By contrast, untreated contaminated soils exhibited a stronger representation of aromatic hydrocarbon degradation, benzoate degradation, xenobiotic biodegradation, fatty acid oxidation, and oxidative stress-related pathways.

3.5 Pathway enrichment and metabolic network reconstruction

Pathway enrichment analysis identified hydrocarbon degradation and central carbon metabolism as prominent pathway groupings associated with the observed treatment-related metabolomic variation (Fig. 4a). Microbial metabolism in diverse environments showed the highest pathway impact, while aromatic hydrocarbon degradation and central carbon pathways contributed substantially to the

reconstructed pathway structure. Network reconstruction associated contamination-related pathways with oxidative stress and xenobiotic metabolism, whereas remediation-associated modules were connected to central carbon metabolism, amino acid biosynthesis, nucleotide turnover, lipid remodelling, and energy-related processes (Fig. 4b). Hydrocarbon degradation modules, including phenol, naphthalene, and ethylbenzene pathways, appeared to converge on central carbon metabolism through catabolic intermediates (Fig. S3). Microbial taxa previously reported in the literature, including *Pseudomonas*, *Ralstonia*, *Acinetobacter*, *Rhodococcus*, *Burkholderia*, *Sphingomonas*, and *Bacillus*, were associated with these pathways, supporting only an indirect functional interpretation of possible microbial community activity.

3.6 Functional interpretation

Primary metabolites, including branched-chain amino acids, glutamic acid, alanine, serine, and nucleotides, were associated with biochemical processes related to protein biosynthesis, nitrogen assimilation, and energy metabolism. Secondary metabolites, including choline, betaine, glycine, acetate, lactate, and lipid-associated resonances, were associated with osmotic adjustment and membrane-related adaptation. Together, the integrated metabolomic profile suggests a remediation-associated transition from contamination-related catabolic stress patterns towards metabolite profiles consistent with anabolic activity and possible functional recovery. However, these changes should be interpreted as candidate functional indicators, not as definitive evidence of restored ecosystem function.

4. DISCUSSIONS

Integrated LC–MS and ¹H-NMR metabolomics provided a sensitive and reproducible approach for assessing biochemical restructuring in petroleum-contaminated soils exposed to different remediation treatments. The combined platforms detected distinct metabolomic patterns associated with untreated contaminated, amended, and reference soils, indicating treatment-associated shifts in soil biochemical profiles and inferred microbial metabolic processes. Instead of confirming complete ecosystem restoration, these data provide a basis for identifying candidate functional indicators of remediation-associated biochemical change. This approach may complement contaminant-based assessment by showing metabolite patterns consistent with improved soil biochemical status, but further validation through microbial activity assays, enzyme measurements, and multi-omics analyses is required before these indicators can be used to confirm restored biological function.

4.1 Metabolic modelling indicates candidate signals of microbial recovery and osmolyte dynamics

Targeted LC–MS profiling detected a diverse set of primary metabolites that differentiated the remediation treatments and provided candidate biochemical indicators of remediation-associated soil change.

Higher relative abundances of branched-chain amino acids, including leucine and isoleucine, together with glutamic acid, alanine, serine, and nucleotide-related metabolites such as adenosine, guanosine, and cytidine, were observed in remediated soils. These metabolites are associated with protein biosynthesis, nitrogen assimilation, nucleotide turnover, and energy-related metabolism, but they should not be interpreted as direct confirmation of restored microbial function. Glutamic acid appeared as an important candidate indicator because of its role in nitrogen–carbon coupling, while the intermittent detection of malic acid and 2-ketoglutaric acid suggested treatment-associated variation in central carbon metabolism (Bayram et al., 2020). These metabolites provide mechanistically informative but still inferential indicators of microbial activity and remediation-associated biochemical recovery (Table 1). Their coordinated abundance patterns suggest a transition from contamination-associated catabolic stress responses towards metabolite profiles consistent with anabolic and growth-supporting biochemical processes. Thus, these findings extend remediation assessment beyond contaminant reduction by identifying candidate functional indicators that require further validation through microbial activity assays, enzyme measurements, and multi-omics analyses.

These observations are consistent with earlier metabolomics studies on petroleum-contaminated soils, although they require cautious interpretation. In Ogoniland, Nigeria, remediation was associated with a shift from stress-associated compounds towards metabolites such as citric acid, ectoine, and lipoic acid, which were interpreted as biochemical signals consistent with microbial recovery (Nwogu-Chigozie et al., 2026). Stable isotope-assisted analysis of PAH-contaminated soils also detected ring-cleavage products entering central metabolic pathways, supporting the interpretation that some of the core metabolites observed in the present study may be associated with hydrocarbon transformation and downstream carbon metabolism (Tian et al., 2018). However, some studies have reported the formation of oxygenated intermediates with higher toxicity, indicating that metabolite changes during remediation do not always translate directly into functional recovery (Spinelli et al., 2024). The lower representation of stress-associated metabolites and the higher relative abundance of remediation-associated compounds observed in this study may therefore reflect differences in contaminant composition, remediation approach, sampling duration, and environmental conditions. These patterns support the value of metabolomics as a sensitive tool for identifying candidate biochemical indicators during bioremediation, but interpretation should remain context-dependent and should be validated through direct microbial activity measurements, enzyme assays, and complementary multi-omics analyses.

Untargeted ^1H -NMR profiling detected secondary metabolites that were associated with physiological adjustment in remediated soils. Higher relative abundances of choline, betaine, glycine, acetate, lactate, and lipid-associated compounds were consistent with osmotic adjustment and membrane-related adaptation under hydrocarbon-associated stress. These metabolites suggest a movement away from contamination-related stress responses towards biochemical profiles compatible with improved cellular regulation and adaptive metabolic activity. By contrast, untreated contaminated soils showed stronger representation of stress-associated metabolites, suggesting continued pressure on catabolic processes. The gradual appearance of anabolic and membrane-related metabolite features in remediated soils points to a possible metabolic reorientation towards biosynthesis, membrane adjustment, and osmotic

regulation, but these patterns should be interpreted as candidate functional indicators instead of definitive evidence of ecological recovery.

Comparable evidence from India showed that biosurfactant-assisted remediation with *Brassica nigra* was associated with the accumulation of osmolytes and lipid derivatives, which may support microbial and plant adaptation under petroleum-related stress (Das et al., 2021). Similarly, Liu et al. (2023) combined metagenomics and metabolomics in China and reported enrichment of aromatic degradation and osmolyte-related pathways, with elevated gentisic acid levels interpreted as being consistent with membrane-related adaptation under hydrocarbon stress. These studies suggest a broadly similar pattern in which polluted soils show stronger representation of stress-associated metabolites, such as fluorene and quinolinic acid, while remediated soils show higher relative abundance of metabolites associated with osmotic adjustment, redox balance, and central metabolism, including ectoine, citric acid, pyridoxamine, and lipoic acid. Functionally, these compounds are linked to osmoprotection, redox regulation, and membrane stabilisation, but their presence should be treated as indirect evidence of microbial adaptation instead of direct confirmation of ecological recovery. This convergence supports the value of untargeted metabolomics, particularly when ^1H -NMR and LC–MS are combined, for identifying candidate functional indicators that complement hydrocarbon depletion data and help describe remediation-associated microbial and soil biochemical responses.

Table 1

Metabolite-based functional biomarkers for monitoring microbial activity and soil ecosystem recovery during bioremediation

Metabolite	Environmental Relevance	Monitoring Application	Reference
Leucine & Isoleucine	Indicators of protein biosynthesis and microbial growth.	Track microbial biomass recovery and soil biological productivity.	Tang et al. (2026)
Alanine & Serine	Central to amino acid metabolism and nitrogen assimilation.	Serve as biomarkers of nutrient cycling and microbial adaptation during remediation.	Kishorekumar et al. (2020)
Glutamic Acid	Hub metabolite in transamination and nitrogen–carbon flux; detected in > 97% of samples.	Stable biomarker of ecosystem recovery and functional nitrogen metabolism.	Li et al. (2025)
Adenosine, Guanosine, Cytidine	Reflect nucleotide turnover and energy metabolism, linked to microbial replication.	Monitor microbial metabolic vigor and biosynthetic capacity in remediated soils.	Soultanas & Janniere (2023)
Malic Acid and 2-Ketoglutaric Acid	Sporadic detection indicates treatment-specific activation of central carbon metabolism.	Reveal pathway-level differences between remediation strategies and microbial responses.	Yang et al. (2025)
Choline & Betaine	Osmoprotectants that stabilise cell membranes and maintain osmotic balance under stress.	Biomarkers of microbial stress tolerance and physiological adaptation in remediated soils.	Goszcz et al. (2025)
Glycine	Functions as an osmolyte and precursor in one-carbon metabolism, linking stress tolerance with biosynthesis.	Indicator of microbial adaptation and metabolic recovery during remediation.	Goszcz et al. (2025)
Acetate & Lactate	Short-chain organic acids reflecting active fermentation and carbon flux.	Track microbial energy metabolism and carbon turnover in recovering soils.	Tang et al. (2026)
Lipid-associated resonances	Evidence of membrane remodelling and stabilisation under contaminant stress.	Functional marker of microbial resilience and long-term ecosystem stabilization.	Erazo-Oliveras et al. (2024)

4.2 Pathway enrichment and metabolic network reconstruction indicate candidate biochemical linkages

Integrated pathway enrichment analysis suggested that hydrocarbon and xenobiotic degradation were associated with organised metabolic linkages rather than isolated biochemical changes. Aliphatic and aromatic hydrocarbons, including PAHs, naphthalene, ethylbenzene, and chlorobenzene, appeared to

connect with shared intermediate pools, including fatty acids, catecholic compounds, and acetyl-CoA. Cytochrome P450-mediated xenobiotic metabolism may contribute to oxidative activation and connect initial hydrocarbon transformation with downstream catabolic pathways (Wen et al., 2025). The reconstructed metabolic networks associated xenobiotic degradation pathways with core cellular metabolism, including carbon flow, nutrient cycling, stress response, and biosynthetic processes. Intermediates such as catechol, protocatechuate, salicylate, benzoate, acetate, and succinate appeared to act as convergence points that may direct carbon flux towards central metabolism. The metabolite network pattern was consistent with previously reported remediation outcomes, including a 93.37% reduction in total petroleum hydrocarbons after 90 days (Fenibo et al., 2025). Poultry-derived nutrient amendment may also have supported greater pathway connectivity and degradation-related activity, although this interpretation remains inferential and should be validated through direct microbial, enzymatic, and multi-omics evidence.

These findings suggest that microbial communities may operate through coordinated enzymatic interactions during hydrocarbon degradation. Activation of one degradation pathway may influence metabolite flow through related pathways, thereby supporting carbon channelling and energy generation under changing environmental conditions. Shared intermediates can connect structurally different hydrocarbon substrates with central carbon metabolism, linking detoxification processes with bioenergetic activity and potential biomass formation. Nutrient availability may further influence pathway activity, indicating that hydrocarbon catabolism depends partly on broader soil nutrient conditions. The distribution and turnover of intermediates such as catechol, protocatechuate, salicylate, benzoate, acetate, and succinate may therefore serve as candidate mechanistic indicators of remediation-associated biochemical change, although they do not independently confirm ecosystem recovery. The observed integration of hydrocarbon degradation pathways is consistent with previous reports of over-represented microbial metabolism across diverse environments (Gigl et al., 2026). Earlier studies have also described sequential and convergent transformations, aromatic ring activation, and cytochrome P450-mediated oxidation as important processes that increase substrate accessibility and support downstream dioxygenase activity (Chikere et al., 2018). The role of intermediate metabolites as possible convergence points is also consistent with mechanistic models of ortho- and meta-cleavage pathways in aromatic hydrocarbon metabolism. Similarly, the potential interpretive value of reconstructed metabolite networks is supported by studies linking coherent metabolite flow with biodegradation efficiency (Sleight et al., 2020), but such network-based interpretations should be treated as inferential and should be validated through direct enzymatic, microbial, and multi-omics evidence.

These results suggest that petroleum hydrocarbon transformation was associated with coordinated metabolic reorganisation, but they do not by themselves confirm complete functional restoration. Metabolite network analysis offers a useful interpretive framework for assessing remediation-associated biochemical change and for identifying candidate indicators of microbial activity. Monitoring the distribution and turnover of intermediate metabolites may help describe the progression of hydrocarbon transformation and possible movement towards improved soil biochemical function, although forecasting mineralisation completeness or confirmed soil recovery requires additional validation.

Nutrient amendments may support stronger pathway connectivity and degradation-related activity, but this interpretation remains inferential. Overall, the study supports the value of integrated pathway and metabolite network analysis for guiding remediation assessment, while recognising that broader application is limited by methodological constraints and uncertainty in functional interpretation. Future work should improve metabolite resolution, combine metabolomics with metagenomics, metatranscriptomics, enzyme assays, and contaminant-degradation measurements, and apply longitudinal validation across independent field conditions to strengthen mechanistic inference and predictive reliability.

Limitations and future directions

Although metabolite identification followed MSI-compliant standards, the resolution of untargeted ^1H -NMR features remains an important limitation because annotation confidence can vary across compounds. Future studies should therefore stratify metabolite annotations by confidence level to improve interpretability and reproducibility. LC-MS provided high analytical sensitivity, but structural ambiguity in untargeted workflows can still affect the assignment of specific biochemical identities, especially for isomeric compounds in complex soil matrices. Functional and taxonomic interpretations based on metabolite patterns also remain indirect. Metabolite profiles can suggest microbial activity and pathway involvement, but they do not directly identify microbial community composition, gene expression, or enzyme-level activity. Thus, future work should combine metabolomics with metagenomic, metatranscriptomic, enzymatic, and culture-based validation to confirm the biological sources and enzymatic drivers of the observed metabolic changes.

The machine-learning models used to identify recovery-associated metabolite predictors also require cautious interpretation. Although they provide useful exploratory insight, their diagnostic and predictive value remains preliminary until tested with rigorous validation procedures. Future studies should include independent datasets, external validation, cross-validation, and performance metrics such as AUC, sensitivity, specificity, and test-set accuracy across different environmental conditions. Without such validation, model outputs should be treated as hypothesis-generating indicators instead of definitive diagnostic tools. Finally, metabolomics provides a high-resolution biochemical snapshot of system status, but it does not establish causality on its own. Future research should use longitudinal sampling to follow temporal changes in degradation pathways and should pair metabolite profiling with enzyme assays, microbial community analysis, and physicochemical monitoring. Such integrated approaches will strengthen mechanistic interpretation and improve the reliability of candidate functional indicators associated with xenobiotic degradation, metabolic network change, and remediation-related soil recovery.

CONCLUSION

Integrated LC-MS and ^1H -NMR metabolomics showed significant biochemical restructuring during field-scale bioremediation of petroleum-contaminated farmland in the Niger Delta. Remediated soils showed

higher relative abundance of metabolites associated with amino acid metabolism, nucleotide turnover, osmotic adjustment, membrane-related adaptation, and central carbon metabolism, while untreated contaminated soils retained stronger representation of stress-associated and aromatic metabolites. These findings suggest that metabolomics can complement conventional contaminant-based assessment by identifying biochemical patterns associated with remediation-related soil change. However, the detected metabolite patterns are interpreted as candidate functional indicators of soil recovery instead of definitive evidence of restored ecosystem function or confirmed microbial functional recovery.

DECLARATIONS

Acknowledgement

We gratefully acknowledge Ntushelo Khayaletu of the University of South Africa for his supervisory guidance during the laboratory analyses.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Declaration of Generative AI Use

ChatGPT (OpenAI) was used to assist with language editing and figure refinement. The authors reviewed, validated, and take full responsibility for all content presented in the manuscript.

CRedit Authorship Contribution Statement

EOF and MST: Conceptualisation, Formal analysis, Writing – original draft, Review

EOF and MST: Methodology, Conceptualisation, Review.

MST: Funding acquisition, Supervision, Review

EOF and MST: Validation, Visualisation.

Funding

This study was not funded by any organisation.

REFERENCES

1. Awan, S., Sharma, S., & Oyege, I. (2026). Metabolomics to Study the Effect of Primary and Secondary Metabolism in Plants Towards Nano Pollution. In R. N. Pudake, U. Jain, & J. M. Al-Khayri (Eds.), *Phyto Nanotoxicology* (Vol. 7, pp. 271–298). Springer Nature Switzerland.
https://doi.org/10.1007/978-3-032-24518-2_11
2. Bayram, S., Fürst, S., Forbes, M., & Kempa, S. (2020). Analysing central metabolism in ultra-high resolution: At the crossroads of carbon and nitrogen. *Molecular Metabolism*, 33, 38–47.
3. Chikere, C. B., Fenibo, E. O., & Akaranta, O. (2018). Comparative effectiveness of activated soil in bioremediation of a farmland polluted soil by polyaromatic hydrocarbon in the Niger Delta. *Journal of Bioremediation and Biodegradation*, 6(9), 456.
4. Cho, S., Preacher, K. J., Yaremych, H. E., Naveiras, M., Fuchs, D., & Fuchs, L. S. (2022). Modelling multilevel nonlinear treatment-by-covariate interactions in cluster randomized controlled trials using a generalized additive mixed model. *British Journal of Mathematical and Statistical Psychology*, 75(3), 493–521. <https://doi.org/10.1111/bmsp.12265>
5. Choudhary, S., Venkatrayalu, G., & Poluri, K. M. (2026). Mechanistic insights into integrated detoxification of hexavalent chromium and polyvinyl chloride derived nanoplastics along with enhanced lipid production by *Chlorella* sp. *Journal of Environmental Chemical Engineering*, 14(4), 123404.
6. Das, A. J., 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.
7. Erazo-Oliveras, A., Muñoz-Vega, M., Salinas, M. L., Wang, X., & Chapkin, R. S. (2024). Dysregulation of cellular membrane homeostasis as a crucial modulator of cancer risk. *The FEBS Journal*, 291(7), 1299–1352. <https://doi.org/10.1111/febs.16665>
8. Fenibo, E. O., Nkuna, R., & Matambo, T. (2024). Impact of artisanal refining activities on bacterial diversity in a Niger Delta fallow land. *Scientific Reports*, 14(1), 3866.
9. Fenibo, E. O., Nkuna, R., & Matambo, T. (2025). Metagenomic Insights Into Biopile Remediation of Petroleum-Contaminated Soil Using Chicken Droppings in Rivers State, Nigeria. *Bioinformatics and Biology Insights*, 19, 11779322251371117. <https://doi.org/10.1177/11779322251371117>
10. Ferreira, O. E., Silva, H. D. M., Alves, A. B., de Aguilar, M. G., Pimenta, L. P. S., Costa, G. H. G., da Cruz, M. C. P., da Silva Bezerra, A. C., & Machado, A. R. T. (2024). Biochar enhances soil interactions and the initial development of sugarcane. *Scientific Reports*, 14(1), 27610.
11. Gigl, F., Abdullahi, M., Benkwitz-Bedford, S., Eastwood, N., Zhou, J., Hollert, H., & Orsini, L. (2026). Challenging Evolutionary Paradigms: *Daphnia* Populations Resurrected From Unpolluted Environments Show Enhanced Detoxification Ability to Aromatic Pollutants. *Molecular Ecology*, 35(4), e70272. <https://doi.org/10.1111/mec.70272>
12. Goszcz, A., Furtak, K., Stasiuk, R., Wójtowicz, J., Musiałowski, M., Schiavon, M., & Dębiec-Andrzejewska, K. (2025). Bacterial osmoprotectants—A way to survive in saline conditions and potential crop allies. *FEMS Microbiology Reviews*, 49, fuaf020.

13. Karaman, I. (2017). Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. In A. Sussulini (Ed.), *Metabolomics: From Fundamentals to Clinical Applications* (Vol. 965, pp. 145–161). Springer International Publishing. https://doi.org/10.1007/978-3-319-47656-8_6
14. Kishorekumar, R., Bulle, M., Wany, A., & Gupta, K. J. (2020). An Overview of Important Enzymes Involved in Nitrogen Assimilation of Plants. In K. J. Gupta (Ed.), *Nitrogen Metabolism in Plants* (Vol. 2057, pp. 1–13). Springer New York. https://doi.org/10.1007/978-1-4939-9790-9_1
15. Li, S., Ahmed, W., Jiang, T., Yang, D., Yang, L., Hu, X., Zhao, M., Peng, X., Yang, Y., Zhang, W., Li, M., & Zhao, Z. (2025). Amino acid metabolism pathways as key regulators of nitrogen distribution in tobacco: Insights from transcriptome and WGCNA analyses. *BMC Plant Biology*, 25(1), 393. <https://doi.org/10.1186/s12870-025-06390-4>
16. Liu, C., Yu, J., Ying, J., Zhang, K., Hu, Z., Liu, Z., & Chen, S. (2023). Integrated metagenomics and metabolomics analysis reveals changes in the microbiome and metabolites in the rhizosphere soil of *Fritillaria unibracteata*. *Frontiers in Plant Science*, 14, 1223720.
17. Ma, J., Shojaie, A., & Michailidis, G. (2019). A comparative study of topology-based pathway enrichment analysis methods. *Bmc Bioinformatics*, 20(1), 546. <https://doi.org/10.1186/s12859-019-3146-1>
18. Mary Celin, S., Sharma, B., Bhanot, P., Kalsi, A., Sahai, S., & Tanwar, R. K. (2023). Trends in environmental monitoring of high explosives present in soil/sediment/groundwater using LC-MS/MS. *Mass Spectrometry Reviews*, 42(5), 1727–1771. <https://doi.org/10.1002/mas.21778>
19. Nwogu-Chigozie, L. N., Chikere, C. B., Ezebuio, V., Tugizimana, F., & Musiwalo, S. (2026). *Functional metabolite signatures reveal ecosystem recovery trajectories in oil-polluted soils undergoing bioremediation*. <https://www.researchsquare.com/article/rs-9657487/latest>
20. Seo, J., Kim, D., Lee, J., Kim, K., Kim, S., & Kim, H. S. (2024). Comparative analysis of assessment models for rehabilitation potential of fish habitat. *Ecological Indicators*, 161, 112003.
21. Sleight, T. W., Khanna, V., Gilbertson, L. M., & Ng, C. A. (2020). Network Analysis for Prioritizing Biodegradation Metabolites of Polycyclic Aromatic Hydrocarbons. *Environmental Science & Technology*, 54(17), 10735–10744. <https://doi.org/10.1021/acs.est.0c02217>
22. Soultanas, P., & Janniere, L. (2023). The metabolic control of DNA replication: Mechanism and function. *Open Biology*, 13(8). <https://royalsocietypublishing.org/rsob/article/13/8/230220/91334>
23. Spinelli, S., Marino, A., Morabito, R., & Remigante, A. (2024). Interplay between metabolic pathways and increased oxidative stress in human red blood cells. *Cells*, 13(23), 2026.
24. Tang, H., Zhang, Q., Yang, J., Ruan, Y., Zhang, X., Pang, S., Qiao, Z., Huang, J., Langford, P. R., Zhou, R., Zhang, H., Zhang, Q., & Li, L. (2026). Metabolic flux analysis revealed key roles of ArcB in ADI pathway and IlvC in BCAA biosynthesis during *Streptococcus suis* anaerobic growth and infection. *Virulence*, 17(1), 2602281. <https://doi.org/10.1080/21505594.2025.2602281>
25. Tian, Z., Vila, J., Yu, M., Bodnar, W., & Aitken, M. D. (2018). Tracing the Biotransformation of Polycyclic Aromatic Hydrocarbons in Contaminated Soil Using Stable Isotope-Assisted

Metabolomics. *Environmental Science & Technology Letters*, 5(2), 103–109.

<https://doi.org/10.1021/acs.estlett.7b00554>

26. Wen, X., Feng, K., Wei, P., Li, J., Li, M., Dou, W., Guo, Z., Zhang, Y., & He, L. (2025). The ROS–FOXO pathway mediates broad-spectrum detoxification of acaricides in *Tetranychus cinnabarinus*. *Communications Biology*, 8(1), 1274.
27. Yang, D., Li, Z., Zhong, K., Tang, X., Ding, Y., Sun, Y., Li, R., & Guan, X. (2025). The Effects of NAA and GA₃ Applications at Preflowering and Preveraison Stages on Sugar–Acid Metabolism and Wine Quality in Marselan Grapes: Integrated Insights from Transcriptomic and Metabolomic Analyses. *Journal of Agricultural and Food Chemistry*, 73(44), 28188–28201.
<https://doi.org/10.1021/acs.jafc.5c08635>

Figures

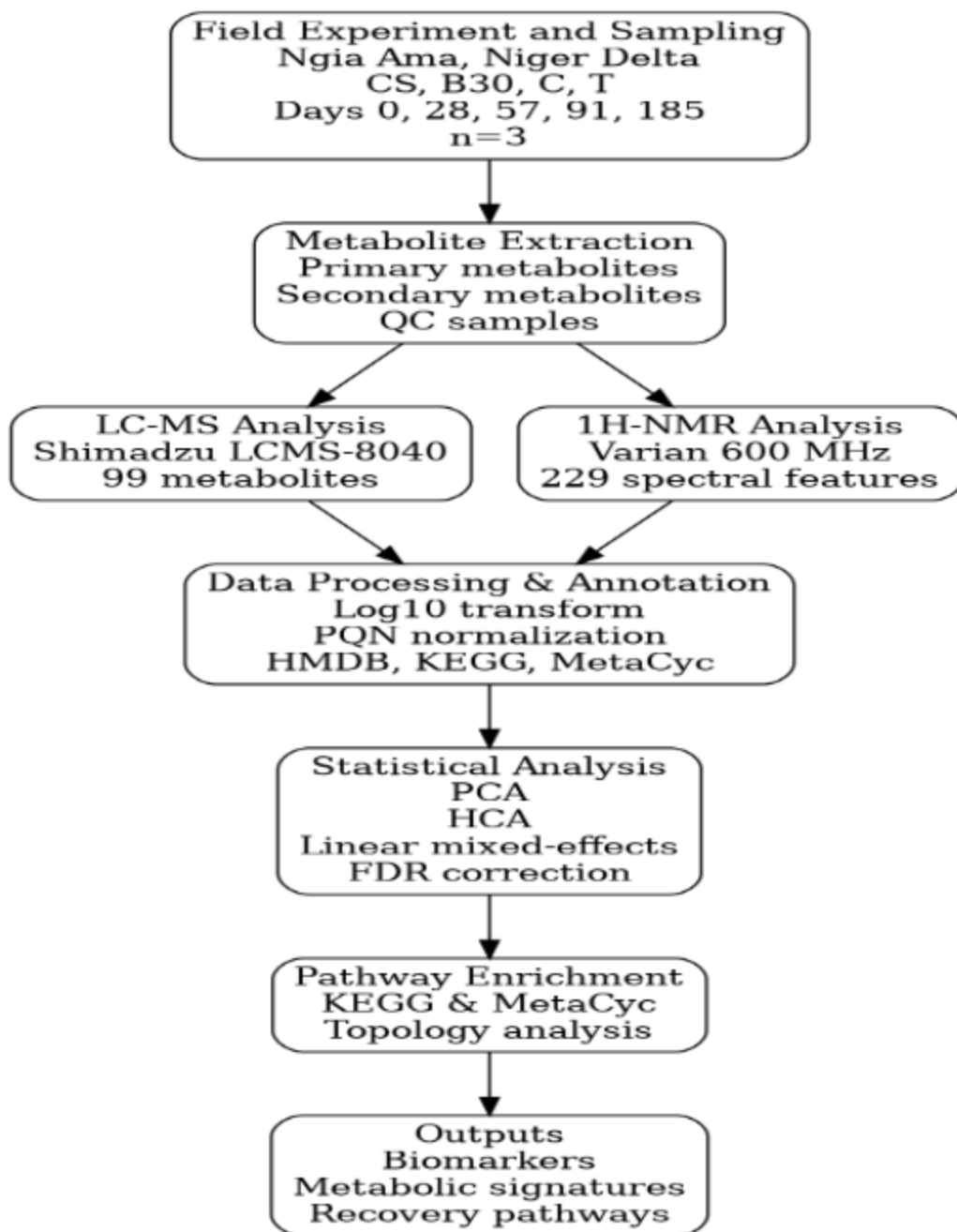


Figure 1

Experimental workflow for LC–MS and ¹H-NMR metabolomic evaluation of petroleum-polluted soils at Ngia Ama, Niger Delta. Soil samples from contaminated, remediated, and uncontaminated plots were subjected to metabolite extraction, LC–MS and ¹H-NMR profiling, data integration, multivariate analysis, biomarker identification, and pathway enrichment to elucidate metabolic mechanisms underpinning hydrocarbon degradation and ecosystem recovery

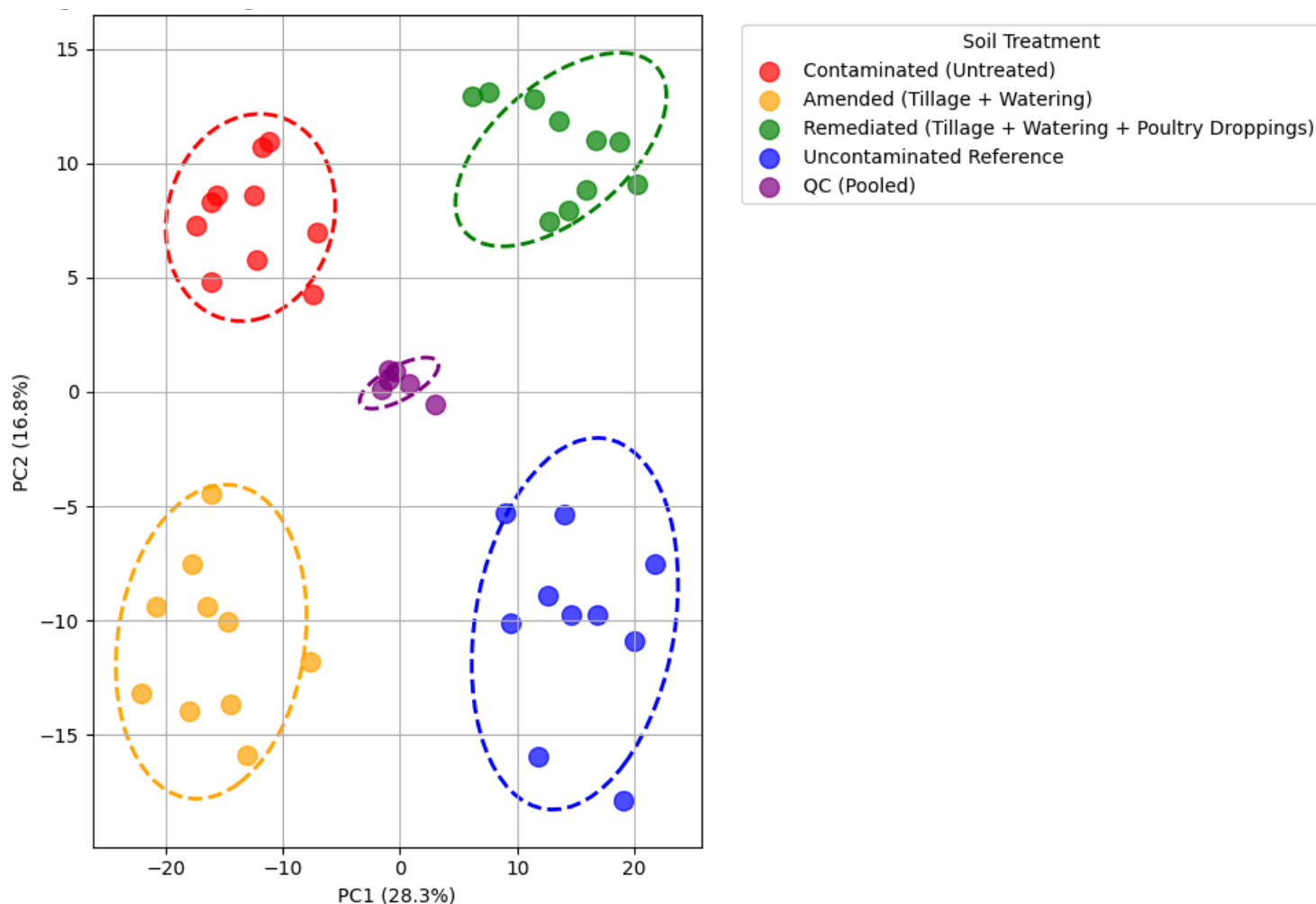


Figure 2

Principal component analysis of integrated LC–MS and ^1H -NMR metabolomic profiles from untreated contaminated, amended, poultry-manure-amended, uncontaminated reference, and pooled QC soil samples. PC1 and PC2 explained 28.3% and 16.8% of the total variance, respectively. Samples showed treatment-associated metabolomic structuring, while pooled QC samples clustered closely near the centre of the ordination, supporting analytical reproducibility. Ellipses indicate group dispersion and should be interpreted as exploratory visual summaries rather than definitive evidence of complete treatment separation.

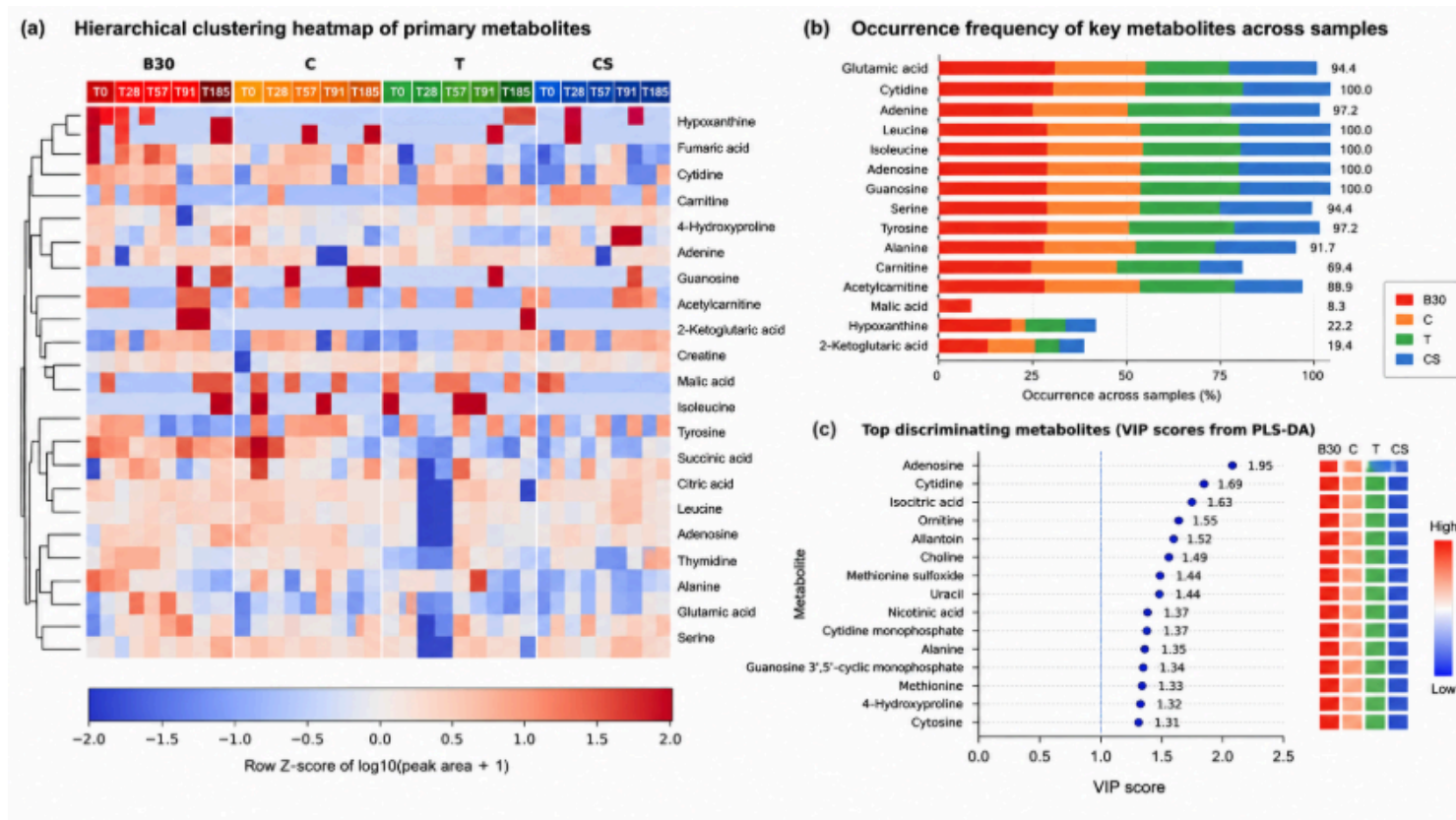


Figure 3

LC-MS-derived metabolic signatures across soil treatments. (a) Hierarchical clustering heatmap showing treatment-specific primary metabolite profiles. (b) Occurrence frequencies of selected metabolites across samples. (c) Candidate discriminant metabolites ranked by VIP scores from the supervised classification model

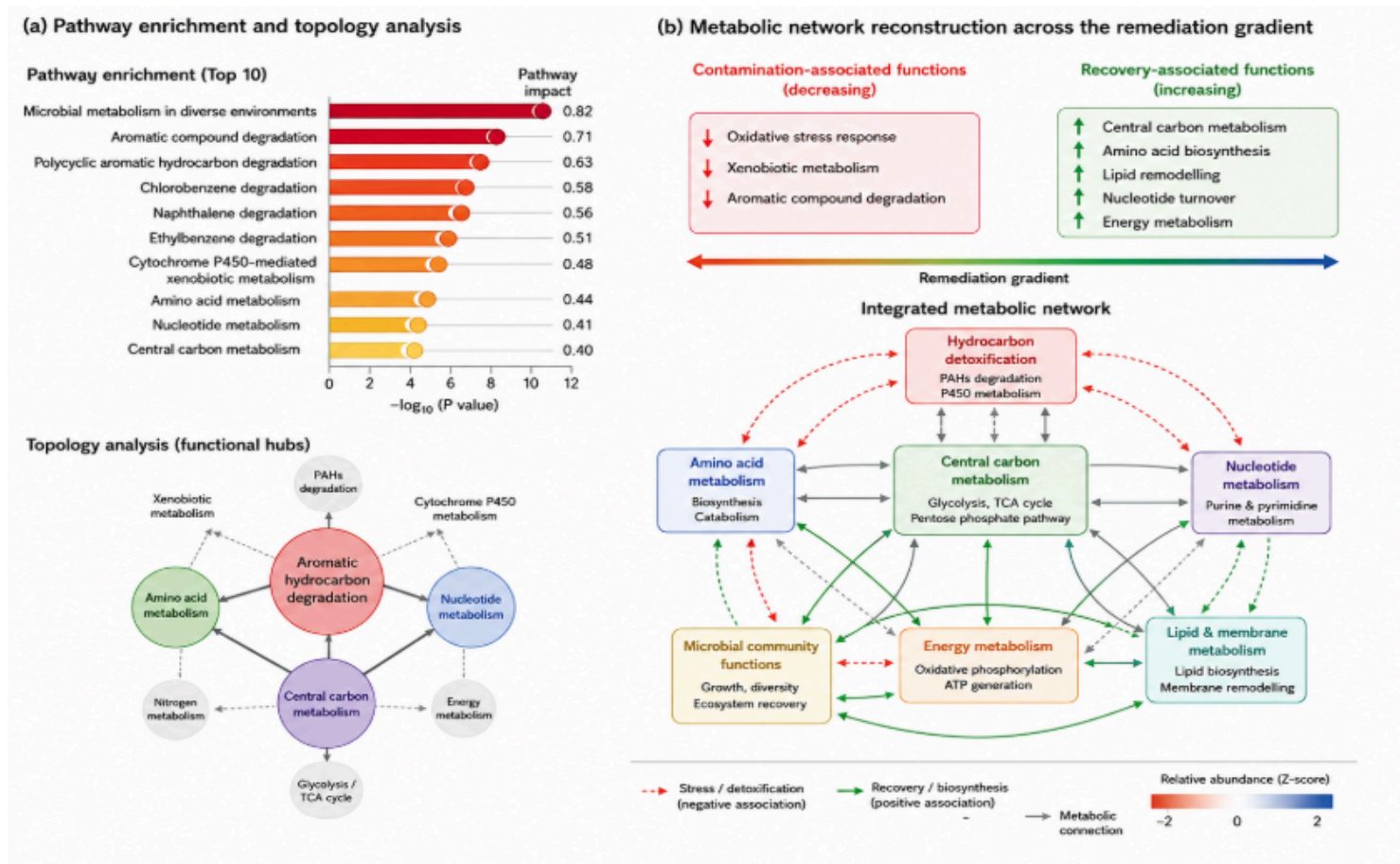


Figure 4

Pathway enrichment and metabolic network reconstruction from integrated LC–MS and ¹H NMR datasets.

(a) Enriched metabolic pathways ranked by significance and pathway impact.

(b) Reconstructed metabolic network showing contamination-associated and recovery-associated modules across the remediation gradient.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixforMetabolomics.docx](#)