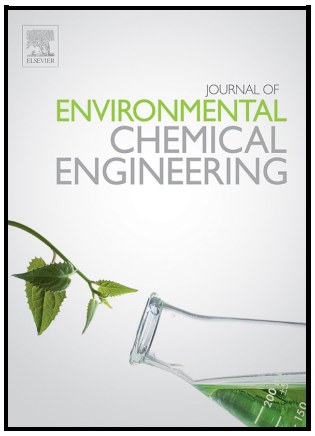


Abundance Correlations of Carbon Fixation and Pollutant Degradation Functions in Oil-Contaminated Soils: Insights Derived from Integrated Metagenomic and Quantitative PCR Analyses

Jiayu Song, Yakui Chen, Yilei Han, Xingchun Li, Zheng Liu, Ziyi Peng, Aimin Fu, Jinman Ren, Diannan Lu, Chunmao Chen

PII: S2213-3437(25)01326-0

DOI: https://doi.org/10.1016/j.jece.2025.116630

Reference: JECE116630

To appear in: *Journal of Environmental Chemical Engineering*

Received date: 3 January 2025
Revised date: 18 March 2025
Accepted date: 11 April 2025

Please cite this article as: Jiayu Song, Yakui Chen, Yilei Han, Xingchun Li, Zheng Liu, Ziyi Peng, Aimin Fu, Jinman Ren, Diannan Lu and Chunmao Chen, Abundance Correlations of Carbon Fixation and Pollutant Degradation Functions in Oil-Contaminated Soils: Insights Derived from Integrated Metagenomic and Quantitative PCR Analyses, *Journal of Environmental Chemical Engineering,* (2025) doi:https://doi.org/10.1016/j.jece.2025.116630

Abundance Correlations of Carbon Fixation and Pollutant Degradation Functions in Oil-Contaminated Soils: Insights Derived from Integrated Metagenomic and Quantitative PCR Analyses

Jiayu Song[1,2‡], Yakui Chen[3‡], Yilei Han[3], Xingchun Li[1,2], Zheng Liu[3], Ziyi Peng[4], Aimin Fu[1,2], Jinman Ren[1,2], Diannan Lu[3*], Chunmao Chen[4]*

[1] State Key Laboratory of Petroleum Pollution Control, Beijing 102206, China

[2] CNPC Research Institute of Safety and Environmental Technology, Beijing 102206, China

[3] Key Lab of Industrial biocatalysis, Ministry of Education, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

[4] College of Chemical Engineering and Environment, China University of Petroleum-Beijing, Beijing 102206, China

‡ These authors contributed equally.

* Corresponding Author

E-mail: hanyl@mail.tsinghua.edu.cn (Yilei Han); ludiannan@tsinghua.edu.cn (Diannan Lu)

Abstract

Microbial communities in petroleum-contaminated soils degrade hydrocarbons while

harboring underexplored CO_2 fixation potential. This study analyzes 44 soil samples from six Chinese petroleum-contaminated sites with varying pollution gradients and physicochemical properties. Metagenomic analysis identified dominant carbon fixation pathways (reverse Tricarboxylic Acid (rTCA) cycle and the Dicarboxylate/4-Hydroxybutyrate (DC/4-HB) cycle) co-occurring with upregulated hydrocarbon degradation genes (alkane 1-monooxygenase (*alkB*) and benzoate 1,2-dioxygenase). Gene co-occurrence analysis demonstrated strong positive correlations (r=0.798) between alkane degradation genes (e.g., *alkB*) and carbon fixation genes (e.g., *cbbL_red*), with polycyclic aromatic hydrocarbon (PAH) degradation genes exhibited moderate associations. Quantitative PCR validation across sites confirmed these relationships. The results uncover a novel metabolic coupling mechanism, where hydrocarbon degradation coincides with enhanced carbon fixation, advancing our understanding of microbial adaptation in contaminated ecosystems. This study provides foundational insights for designing eco-friendly bioremediation strategies leveraging microbial dual metabolic capabilities.

Keywords: petroleum hydrocarbons; biogeochemical cycles; CO_2 assimilation; functional genes; gene quantification；correlation analysis

1. Introduction

The rising global demand for t crude oil- a primary energy source-has intensified oil contamination incidents throughout its lifecycle, including exploration, production, maintenance, transportation, storage, and accidental spills [1]. Such pollution imposes severe ecological burdens [2]. Microbial communities adapt to these conditions through selective pressure, favoring hydrocarbon-degrading taxa [3], [4]. This adaptation drives the proliferation of pollutant-degrading genes [5], such as those encoding alkanes [6] and polycyclic aromatic hydrocarbon

(PAH)-metabolizing enzymes [7]. These microbial activities crucially mitigate petroleum hydrocarbon contamination in soils [8].

Microbial communities in oil-contaminated soils exhibit distinct structural profiles, with growing research interest in there carbon metabolism roles [9]. Notably, photosynthetic and methanotrophic bacteria can utilize petroleum hydrocarbons as energy sources while performing carbon fixation capabilities [10]. Recent advances in genetic and metabolic engineering aim to enhance these oranisms' dual capabilities- pollutiant degradation and carbon sequestration-to boost there environmental and industrial utility [11]. However, studies highlight that petroleum pollutants disrupt interactions between carbon-cycling genes and other crucial soil function networks [12]. A key knowledge gap persists: the regulatory role of carbon fixation genes in modulating carbon cycling during organic pollutant degradation remains poorly characterized [13]. Quantifying the abundance of these functional genes is thus essential to evaluate microbial CO_2 fixation potential in petroleum-contaminated soils and advanceremediation strategies.

Advances in high-through sequencing and metagenomics technologies have revolutionized the exploration of functional gene abundance [14], [15]. These approaches enable researchers to map microbial communities in oil-contaminated soils [16] and identify functional genes linked to carbon cycling and organic pollutant degradation [17]. Quantitative Polymerase Chain Reaction (qPCR) further enhances this capability by quantifying target DNA copies in environmental samples, providing critical data for monitoring bioremediation efficacy [18]. Recognized for its automation, speed, sensitivity, and reliability, qPCR has become a cornerstone tool in bioremediation research [19].

This paper profiles functional genes diversity in petroleum-contaminated soil. Metagenomic data were functional annotation and pathway-mapped using the KEGG database to identify

modules associated with carbon fixation and pollutant degradation. Surface soil samples were collected from six geographically distinct oil fields in China, each exhibiting unique physicochemical properties. We specifically investigate synergistic relationships between carbon fixation genes and pollutant degradation genes. Metagenomic analysis revealed distinct carbon fixation patterns across soil types, with significant abundance correlations to key hydrocarbon degradation genes. Complementary qPCR data validated linkages between critical carbon fixation genes (e.g., *cbbL*) and pollutant degradation markers (e.g., *alkB*). Using these findings, we proposed an integrated pathway linking carbon fixation and pollution degradation via CO_2 metabolic nodes. This work establishes a fundamental framework for understanding gene-level interactions in contaminated ecosystems and offers theoretical foundations for developing multifunctional bioremediation strategies.

2. Materials and methods

2.1 Site description and Soil samples

A total of 44 surface soil samples (0-20 cm depth) were collected from six oil-contaminated fields in China. Site S1 (46°35′N, 125°18′E) exhibites a temperate continental monsoon climate, with an average annual precipitation of 427.5 mm. Site S2 (34°20′N, 107°10′E) shares a similar temperate continental monsoon climate, recording 470 mm annual precipitation. Site S3(31°74′N, 104°46′E), experiences a subtropical monsoon climate and receives 1200 mm of annual rainfall. Sit S4 (38°43′N, 117°30′E) is characterized by a warm temperate semi-humid continental monsoon climate, with 593 mm average annual precipitation. Site S5(45°36′N, 84°42′E) and Site S6(42°55′N, 89°25′E) both feature temperate continental climates, but differ markedly in precipitation: S5 receives 150 nm annually, while S6 receives less than 100nm. All samples were

placed in sterile bags, transported to the laboratory under cold chain conditions, and stored at 4°C. Prior to analysis, each sample underwent a standardized pre-treatment procedure [20], which included homogenization. Each homogenized sample was divided into three aliquots for distinct analyses: (1) physicochemical analysis, (2) functional gene quantification, and (3) metagenomic sequencing.

2.2 Characterization of the physicochemical and nutritional properties

We measured physicochemical properties, total petroleum hydrocarbons (TPH), and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) enzyme activity. Physicochemical properties included pH, water content, electronic conductivity, ammonium nitrogen (NH_4^+-N), nitrate nitrogen (NO_3^--N), total nitrogen (TN), total phosphorus (TP), effective phosphorus (AP), and organic carbon (OC). These analyses followed established protocols [21]. RubisCO activity was quantified using a commercial ELISA kit (Shanghai Rui fan Biotechnology Co., Ltd.) according to the manufacturer's guidelines. TPH concentrations were determined via gas chromatography.

2.3 Quantification of functional genes

The qPCR was utilized to quantify the abundance of functional genes in the soil samples [22]. Target genes included those associated with petroleum hydrocarbon degradation (e.g., *alkB*, *PAH-RHDα GN*, *PAH-RHDα GP*) and carbon fixation (*cbbL*, *cbbM*, *aclB*, and *fhs*), as referenced in prior studies [23], [24], [25], Primer sequences are detailed in **Supplementary Table 1**. All samples were analyzed in triplicate, with the inclusion of a no-template control (NTC) and a negative sample control (NSC).

The qPCR reaction mixture comprised 7.5 μL of 2×SYBR Green Mix, 0.7 μL of each forward

and reverse primer (5 μM), and 10-100 ng of template DNA. Thermal cycling conditions were as follows: initial denaturation at 95°C for 5 minutes, followed by 45 cycles of denaturation (95°C,15s), annealing (55°C, 15s), and extension (72°C, 35s).

For standard curve generation, plasmid DNA carrying target gene sequences was serially diluted (10-fold) across six gradients, ranging from 10^7 to 10^2 copies per reaction. Each reaction well received 2 μL of diluted plasmid. Amplification efficiency ranged from 95% to 110%, with standard curve R^2 value＞ 0.990. Primers specificity was confirmed by single-peak melting curves of amplification products.

Following qPCR, PCR products were analyzed via 1.5% agarose gel electrophoresis. Bands matching the expected size were excised and purified using a gel extraction kit. Purified DNA fragments were ligated into the pMD18-T vector (Takara Bio) via T-A cloning. The ligation products were transformed into XL10-Gold competent cells (Agilent Technologies) through heat shock at 42℃ for 45s. Transformed cells were recovered SOC medium at 37 ℃ and plated onto LB agar supplemented with 100 ug/mL ampicillin. After 16h incubation at 37 ℃, single clones were selected , for plasmid isolation using a Miniprep ki (Qiagen). Plasmid DNA was verified by Sanger sequencing (Genewiz) to confirm target gene insertion.

2.4 Metagenomic analysis

Total microbial DNA was extracted from soil samples using the DNeasy® 96 PowerSoil® Pro QIAcube® HT Kit (Qiagen, Germany). DNA quality was determined by 1.0% agarose gel electrophoresis, while concentrations were quantified with a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific, USA). All samples were subjected to stringent pre-library quality

control via UV-Vis Spectrophotometry (NanoDrop/Thermo Fisher). The A260/A280 and A260/A230 ratios for all samples fell within acceptable ranges (1.85 - 1.92 and 2.05 - 2.22, respectively), meeting stringent quality thresholds for downstream applications [26], [27], [28]. Purified DNA was stored at -80 ℃ until further processing.

DNA was fragmented to ~400 bp using a Covaris M220 ultrasonic processor (Covaris, USA). Libraries were prepared with the NEXTFLEX^{TM} Rapid DNA-Seq Kit (Bioo Scientific, USA), followed by bridge PCR amplification. Sequencing was conducted on the Illumina NovaSeq 6000

p

l

The metagenome assembly was performed using MEGAHIT [29]. A non-redundant gene a
catalog was sequencently constructed from with the assembled sequences. For functional t
annotation, these gene sequences were aligned against the KEGG GENES database using BLASTP f
(Version 2.2.28+). The alignment parameters were set with a stringent e-value cutoff of $1e^{-5}$ to o
ensure reliable matches. Functional annotation was performed through KOBAS 2.0 (Xie et al. r
2011), which implements KO-based classification. Four distinct functional categories were m
analyzed: KO (KEGG Orthology), Pathway, EC, and Module. The abundance of each functional category was determined by summing the normalized abundances of all genes assigned to specific (
annotations within that category.
I

2.5 Data Analysis and Visualization

l

Functional gene analysis and visualization were implemented in Python (version 3.9.16). The u
computational workflow comprised three principal components: 1) statistical analyses including m
principal component analysis (PCA) conducted through the *scikit-learn* package (version 1.5.1), i
2) correlation calculations executed using *pandas'corr* method (version 1.5.3), and 3) visualization

n

a

,

generated by *seaborn* (version 0.12.2) and *matplotlib* (version 3.9.2). Core data processing tasks were managed through *numpy* (version 1.26.4) and *pandas* (version 1.5.3), ensuring efficient handling of large-scale metagenomic datasets.

Soil property analysis employed PCA to characterize multivariate patterns. During pre-processing, data distributions were assessed through statistical diagnostics, with outliers and missing values systematically addressed (**Supplementary Figure 1)**. The final dataset comprised 44 qualified samples meeting quality control criteria. PCA implementation focused on three analytical dimensions: (1) explained variance ratios quantifying component informativeness, (2) statistical significance testing of principal components (PCs), and (3) comparative evaluation of sample characteristics across geographic regions.

An integrated analytical framework was developed to investigate 13 carbon fixation modules and 11 pollutant degradation functional genes across 36 metagenomic samples. The investigation comprised two complementary approaches: (1) Comparative abundance profiling of functional modules/genes through metagenomic read quantification, and (2) Systematic correlation mapping using Pearson's coefficient analysis. Specifically, cross-system interactions were examined by calculating pairwise correlations between carbon fixation pathways and xenobiotic degradation capacities.

The correlation analysis between pollutant degradation genes and carbon fixation genes was conducted against a reference background comprising all functional genes. The analytical workflow proceeded through four sequential stages: (1) initial pairwise correlation quantification between 132 carbon fixation genes and 12905 non-carbon fixation genes using Pearson’s method; (2) establishment of a baseline correlation distribution by aggregating non-carbon fixation genes

as reference variables; (3) Descending prioritization of non-carbon fixation genes based on their mean correlation intensity with carbon fixation genes;(4) stratified analysis focusing on 4599 constitutively expressed genes (detected across all 36 samples) to minimize outlier effects. This refined subset contained 81 high-confidence carbon fixation genes (61.4% retention rate) and 4518 non-carbon fixation genes, including 7/11 target pollutant degradation genes.

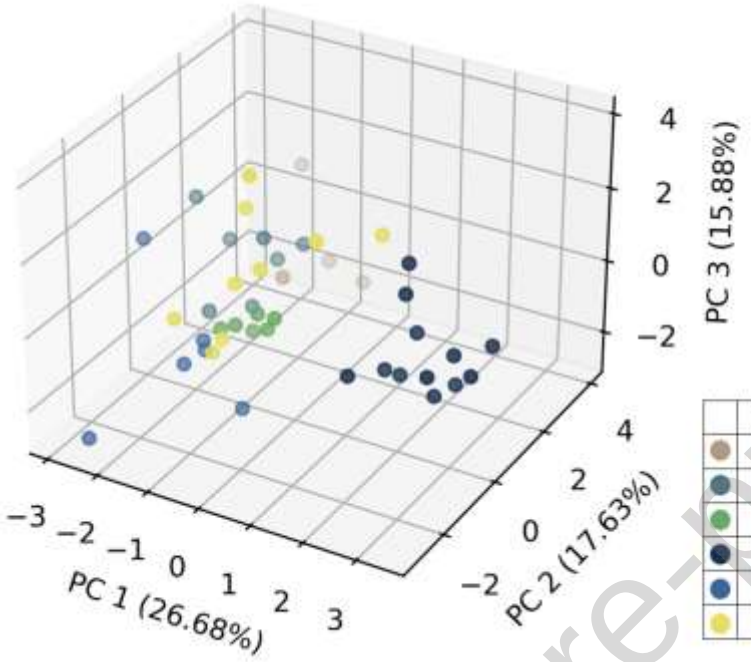
The PCA and correlation analysis were performed on qPCR-derived quantification data following systematic quality control procedures. Prior to PCA implementation, gene abundance distributions were rigorously evaluated, with outliers and missing values addressed according to established protocols (**Supplementary Figure 2)**. Through this preprocessing pipeline, 41 high-quality samples were retained for downstream analysis. The PCA specifically examined three critical aspects: explained variance ratios reflecting component biological relevance, statistical significance of PCs, and spatial differentiation patterns among region sample cluster.

3. Results and Discussion

3.1 Analysis of soil properties

We intially examined distributional diagnostics of physicochemical properties to characterize regional soil heterogeneity, systematically addressing data completeness issues and statistical outliers through standardized protocols (**Supplementary Figure 1).** Following this preprocessing phase, PCA was systematically applied to 11 key physicochemical parameters across quality-controlled samples (**Supplementary Figure 3**). The dimensionality reduction revealed that the first three principal components cumulatively accounted for 60% of the total variance, with subsequent components domonstrating a marked inflection point in explanatory power. This variance partitioning pattern facilitated identification of core parameters driving geographical

differentiation while filtering noise from higher-order components.



	Field	#Samples	TPH Content (mg/kg)
	S1	4	20407.25 ± 13541.08
	S2	7	5657.90 ± 7183.10
	S3	6	2047.67 ± 2382.42
	S4	12	490.71 ± 384.38
	S5	6	200.39 ± 98.16
	S6	9	0.10 ± 0.06

Figure 1 PCA of the physicochemical properties of the samples. Each point in the figure represents a distinct sample, plotted within the multidimensional space determined by the first three PCs. The axes of the figure are annotated with the percentage of variance that each of the PCs explains.

Figure 1 demonstrates distinct geographical differentiation of soil samples across contaminated sites through three-dimensional principal component projections. Regional soils characteristics were delineated based on dominant parameters contributing to these components (**Supplementary Table 2-3)**. Notably, S1- area soils exhibited a unique profile characteristiced by: RubisCO enzyme activity displayed remarkable consistency across all samples (30-80 U/L), explaining its limited contribution to principa component differentiation. This metabolic uniformity aligns with established correlations between RubisCO functionality and stable soil carbon pools [30]. While elevated RubisCO activity is often associate with enhanced carbon assimilation potential in autotrophic microorganisms [31], this enzymatic parameter alone cannot

used in isolation to characterize soil carbon sequestration capacity.

3.2 Analysis of functional modules for carbon fixation

To accurately characterize the carbon sequestration capacity of microorganisms in petroleum-contaminated soil, we employed metagenomic metagenomic sequencing technology. This involved conducting a comprehensive analysis and in-depth characterization of microbial carbon fixation functional genes. The functional genes obtained from metagenomic sequencing were annotated against the KEGG database at three hierarchical levels: KO, module, modules, and pathways. Special emphasis was placed on the carbon fixation functions exhibited by photosynthetic organisms and prokaryotes. These functions are represented in the KEGG database through 2 pathways and 13 modules, as detailed in **Supplementary Table 4**. Key processes included the Calvin cycle, CAM cycle (plant-specific pathway), and rTCA cycle, among others [32].

Figure 2 showed functional gene reads counts at the module level across all samples. It revealed higher gene abundance within the prokaryotic carbon fixation pathway (ko00720, represented in the last seven columns) compared to the photosynthetic biological carbon fixation pathway (ko00710, represented in the first six columns). Within the prokaryotic pathway, the most abundant functional genes were associated with the rTCA cycle (M00173), DC/4-HB cycle (M00374), and 3-HP bicycle (M00376). In the photosynthetic biotic carbon fixation pathway, the Calvin Cycle (M00165) module exhibited the highest gene abundance. This result aligns with recent observations of key carbon fixation genes abundances in samples from petroleum-hydrocarbon-contaminated aquifers [9] and of shale gas field soil [13]. **Supplementary Figure 4** provided a comparative analysis of the different carbon fixation functional modules across these regions. Regarding regional distributions, 3 out of 4 samples demonstrated high gene abundance,

potentially linked to the S1 region's highest pollution levels. For the remaining regions, the gene abundance rank order was: S3 > S2 > S5 > S4 > S6. Notably, S3, S2, and S5 samples exhibited more average physicochemical properties. In contrast, S4 samples were characterized by higher organic carbon content, while the S6 area featured high pH and substantial total nitrogen content.

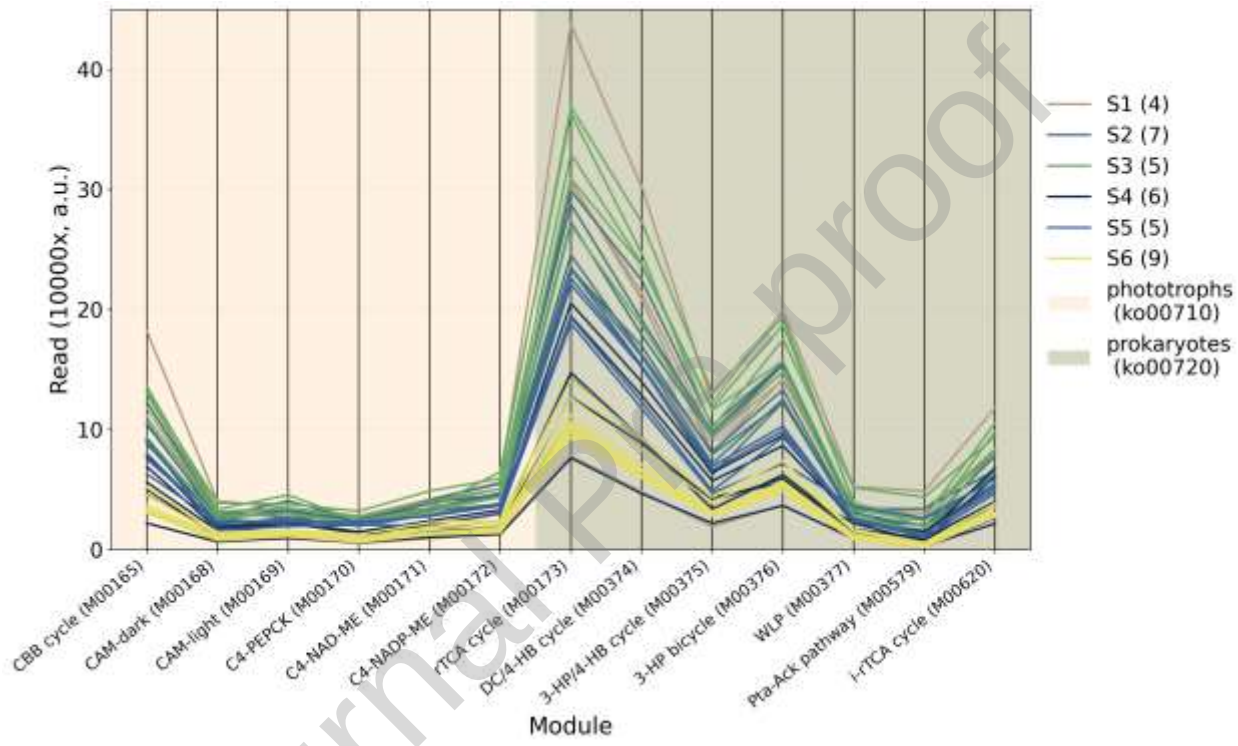


Figure 2. Gene Abundance Distribution in Carbon Fixation Pathways

The *x*-axis shows the carbon fixation functional modules, identified by their KEGG Module numbers. The *y*-axis quantifies the gene abundance in arbitrary units within these modules. Each line in the graph corresponds to a different sample, with varying line colors indicating samples collected from diverse fields. The numbers in parentheses next to the field names denote the sample count. Moreover, different background colors are used to distinguish the metabolic pathways associated with the functional modules.

Although the overall abundance of carbon fixation genes varied among regions, their relative sequence remained consistent. We computed pairwise correlation coefficients for carbon fixation

functional modules. At the module level, diverse carbon fixation functional genes exhibited significant correlations, as shown in **Figure 3**. Data were then aggregated at the pathway level (**Figure 3B)**. Inter-correlation coefficients between the two pathways exceeded 70%, while the intra-correlation coefficients within each pathway were higher than 80%. These results indicated the evolutionary stability of carbon fixation mechanisms in both prokaryotic and eukaryotic organisms [33]. Notably, complementarity and synergistic relationships were observed between prokaryotes and photosynthetic organisms carbon fixation pathways [34].

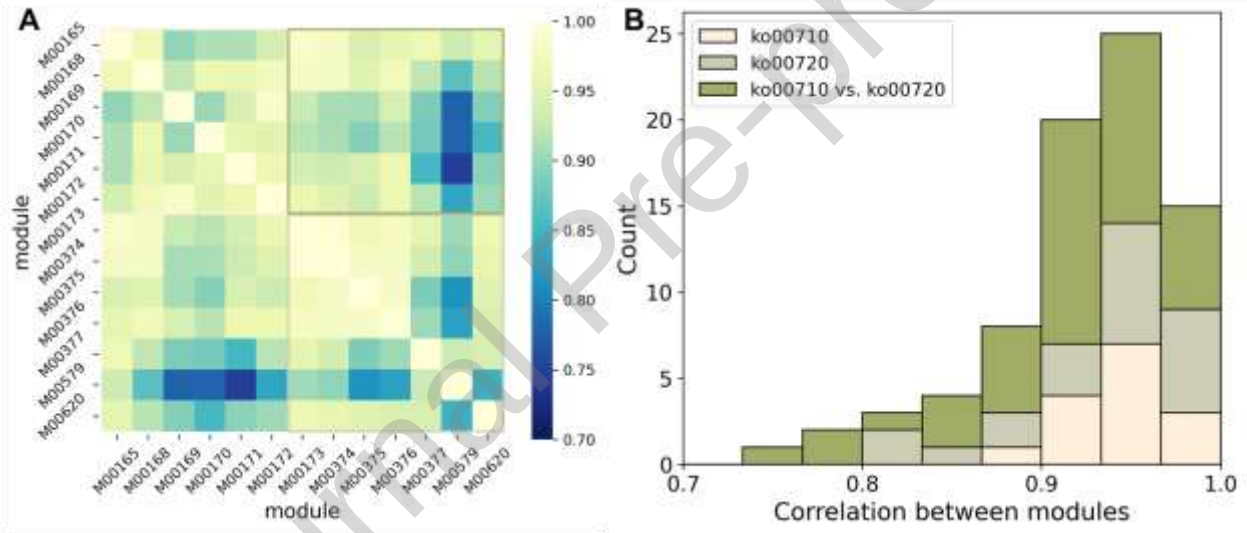


Figure 3 Correlation Analysis of Carbon Fixation Functional Modules' Abundance. (A) Heatmap illustrating the correlation coefficients; (B) Distribution graph of correlation coefficients. The analysis encompasses the autocorrelation within carbon fixation modules in photosynthetic organisms (ko00710), intra-correlation within carbon fixation pathways in prokaryotes (ko00720), and the inter-correlation between these two pathways (ko00710 vs ko00720), presented in stacked format.

3.3 Analysis of functional genes for pollutant degradation

The fields were primarily contaminated with petroleum hydrocarbons, including alkanes and

aromatic compounds. Functional genes associated with alkanes and aromatic hydrocarbons hydrocarbon degradation were identified from the KEGG database [6], [7], as presented in **Supplementary Table 5.** These functional genes were categorized into four groups: (1) aerobic oxidation of alkanes, (2) toluene oxidation via monooxygenase activity, (3) naphthalene degradation through dioxygenase activity, and (4) benzoyl-CoA pathway under anaerobic conditions.

Figure 4 illustrated the reads counts of selected pollutant-degrading functional genes. The data revealed significantly higher expression levels for genes involved in aerobic alkane oxidation and naphthalene dioxygenase activity. Specifically, alkane 1-monooxygenase (K00496) [35], long-chain alkane monooxygenase (K20938) [36], catechol 2,3-dioxygenase (K00446) [37], and benzoate/toluate 1,2-dioxygenase (K05549) [38] showed robust expression. This indicated strong degradation capaacities for both alkanes and aromatic compounds in oil-contaminated sites [39]. Regional comparisons of pollutant-degrading genes expression levels revealed three distinct clusters:(1) high-expression groups: S1 and S3; (2) moderate-expression group: S2 and S4; and (3) low-expression group: S5 and S6. This ranking directly mirrored pollution intensity gradients, suggesting that the microbial communities have evolved adaptive metabolic strategies to optimize survival and functionality under varying contamination pressures [12], [40], [41].

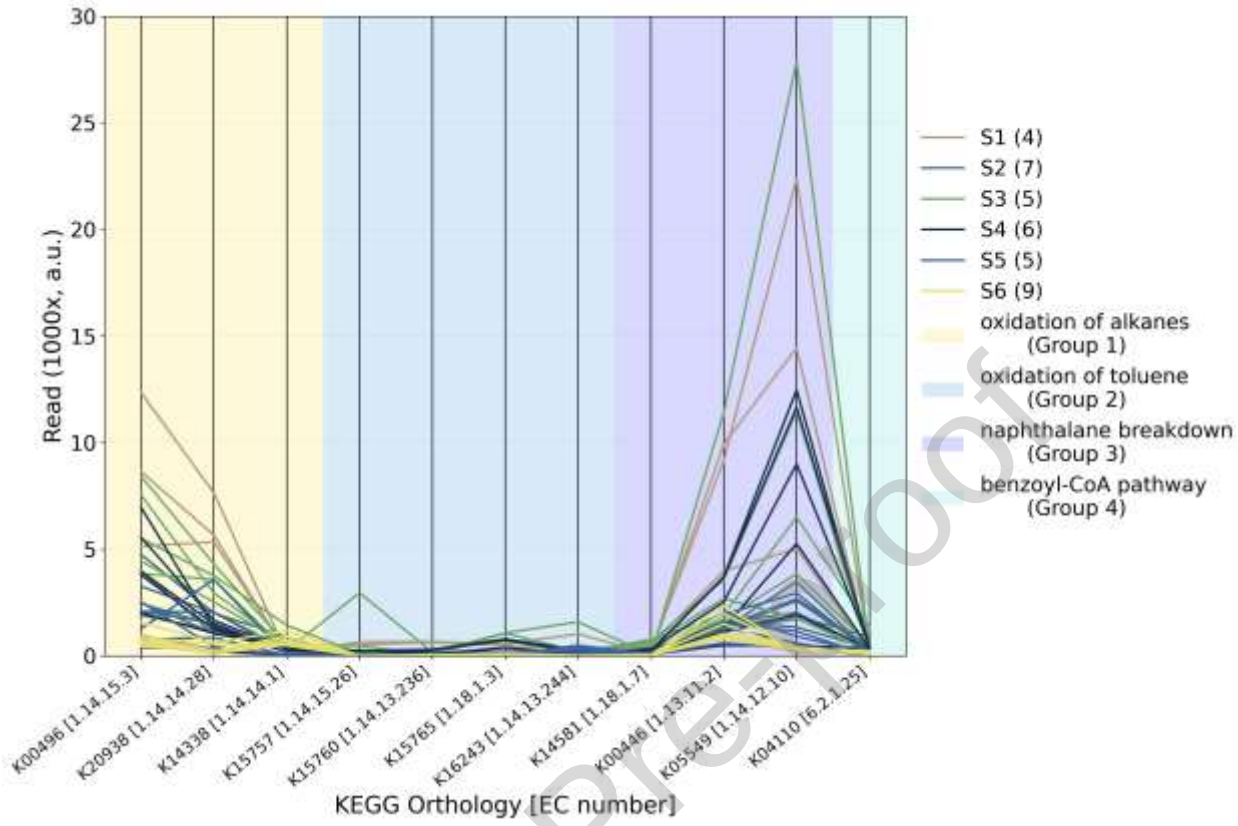


Figure 4 Abundance of Pollutant Degradation Functional Genes. The *x*-axis displays pollution-degrading functional genes, identified by KEGG Ortholog and EC numbers, while the *y*-axis represents the gene abundance in arbitrary units. Each line corresponds to an individual sample, with diverse colors signifying samples from different sources. The background colors are employed to distinguish the functional group.

Pariwise correlations between functional genes were calculated (**Figure 5A)**. **Figure 5B** displayed the overall distribution, revealing a distinct pattern compared to carbon fixation modules correlations (**Figure 3B**). This indicated significant differentiation of pollutant-degrading functions among soil samples from different regions, highlighting the adaptability of degradation functions to diverse environmental conditions. Notably, while carbon fixation functions remained relatively consistent and stable across regions, pollutant-degrading functions demonstrated greater dynamic and specificity to the local contamination levels and types [42].

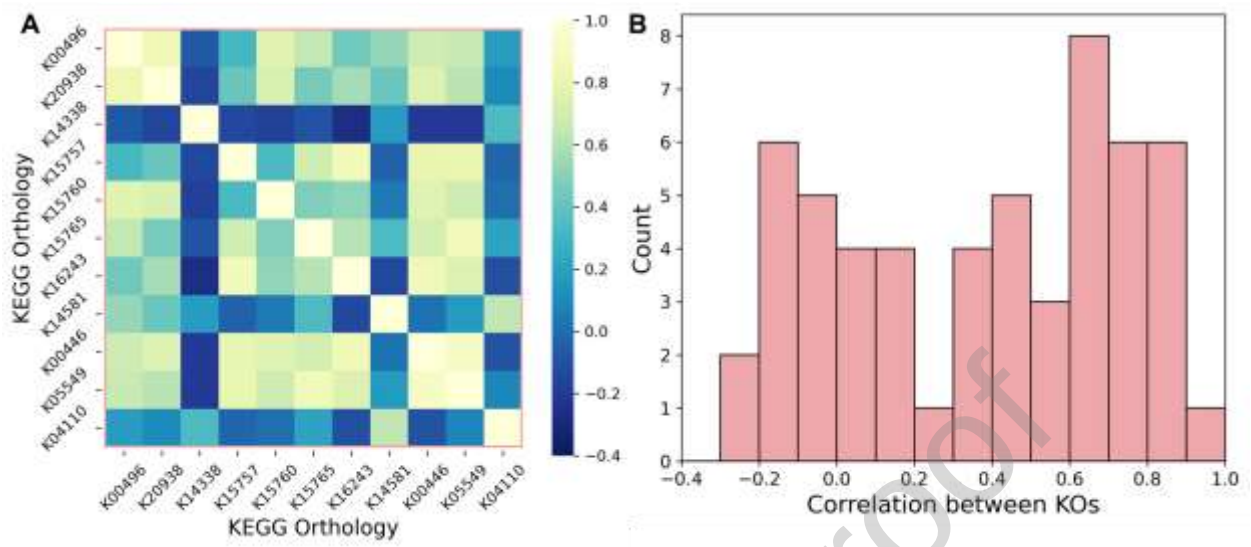


Figure 5 Distribution of Correlations for Pollution-Reducing Functional Genes. (A) Heatmap illustrating the correlation coefficients; (B) The distribution of these correlation coefficients within the upper triangular matrix which excludes the diagonal elements on the heatmap.

3.4 Position of pollutant degradation genes in the overall correlation rankings with carbon fixation genes

Given the significant correlation observed among carbon fixation gene modules, we aggregated all functional genes associated with both pathways. This aggregation aimed to optimize the utilization of the metagenome sequencing data. We the analyzed the correlation distribution between each non-carbon fixation genes and the carbon fixation genes. To ensure unbiased results, we computed pairwise correlations between all non-carbon fixation genes and the carbon fixation functional genes. These correlation coefficients were used to generate a comprehensive ranking.. The relative positions of pollutant-degrading genes within this ranking reflected their functional associations with the carbon fixation.

In the study, 7 functional genes associated with pollutant degradation were identified across

all samples (**Figure 6). Their** distribution patterns revealed varying correlation strength with carbon fixation genes across different functional categories. Notably, two alkyl oxidation genes (K20938 and K00496), which function in alkane degradation [43], [44], occupied the highest positions in the ranking. Four aromatic hydrocarbon oxidase genes (K14581, K00446, K15765 and K05549) demonstrated moderate correlations, ranking between 2600th and 3000th positions. These genes are known to degraded naphthalene and toluene [45], [46], [47]. In contrast, the anaerobic benzoyl-CoA pathway (K04110) [48] exhibited the weakest correlation. **Supplementary Figure 7** presented regional analysis results, showing the frequency of pollutant-degrading genes appearing in the top 2000 rankings. Alkyl oxidation genes occurred four times this threshold, while aromatic compound degradation genes appeared three times. These geographical variations in ranking frequencies suggest potential functional linkages between pollutant degradation proceese and carbon fixation mechanisms.

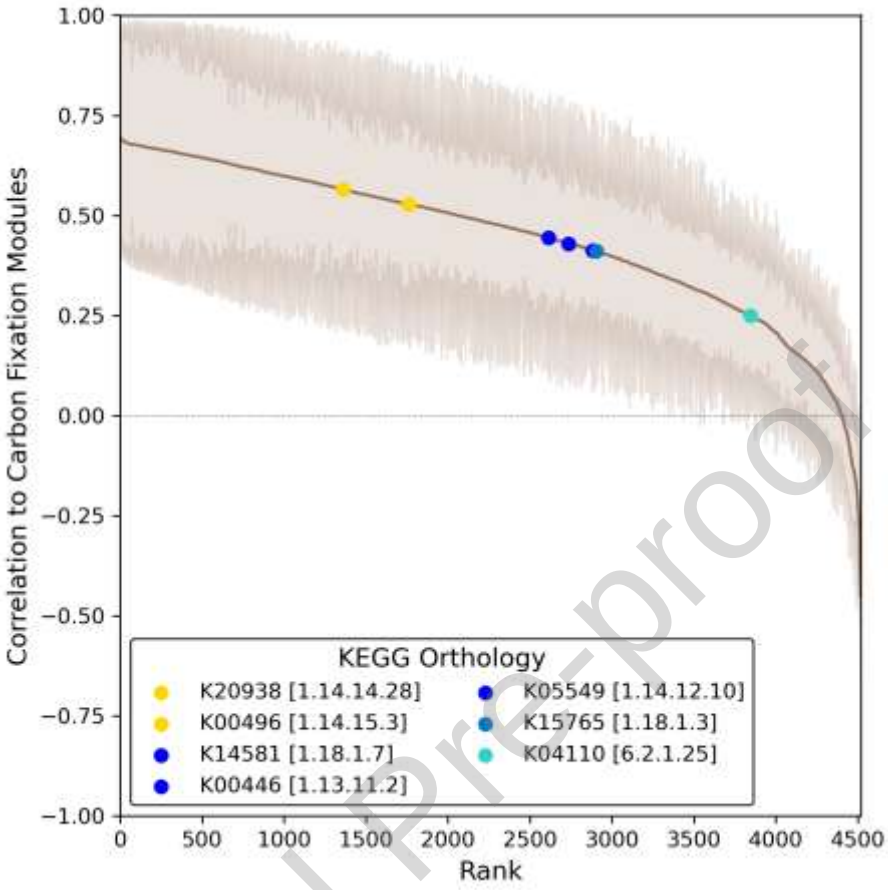


Figure 6 Correlation of Functional Genes with Carbon Fixation Pathways. The data are sorted according to the average magnitude of the correlations. The *x*-axis represents the ranking of the genes, and the *y*-axis shows their correlation values. The mean correlation is illustrated by a solid line, with the shaded area denoting the 95% confidence interval for these values. Colored dots illustrate the positions of the seven pollution-reducing genes, with each color corresponding to a different functional category: yellow for aerobic alkyl oxidation of alkanes (Group 1), dark blue for aerobic monooxygenase activity on toluene (Group 2), light blue for aerobic dioxygenase activity on naphthalene (Group 3), and cyan for anaerobic functions related to the benzoyl-CoA pathway (Group 4).

3.5 Correlation analysis of functional genes determined by qPCR

To validate potential functional linkages between carbon fixation and pollutant degradation

genes, we performed qPCR analysis across all soil samples. The quantified targets included seven carbon fixation genes and three pollutant degradation genes (**Supplementary Table 6**). PCA of 41 qualified samples (**Supplementary Table 7, Supplementary Figure 8**) revealed that the first two principal components explained over 55% of total variance. Notably, the *alkB* gene (alkyl oxidation) and the *cbbL_green2* gene (carbon fixation emerged as major contributors to this variance. Pairwise correlations analysis (**Figure 7**) indentified significant associations: *alkB* showed strong correlation with both *cbbL_red* and *cbbL_green1*, while *PAH_RHDαGN* exhibited marked correlations with *cbbL_green2* and *cbbM*. These findings further support the proposed functional connections between specific pollutant degradation pathways and carbon fixation systems.

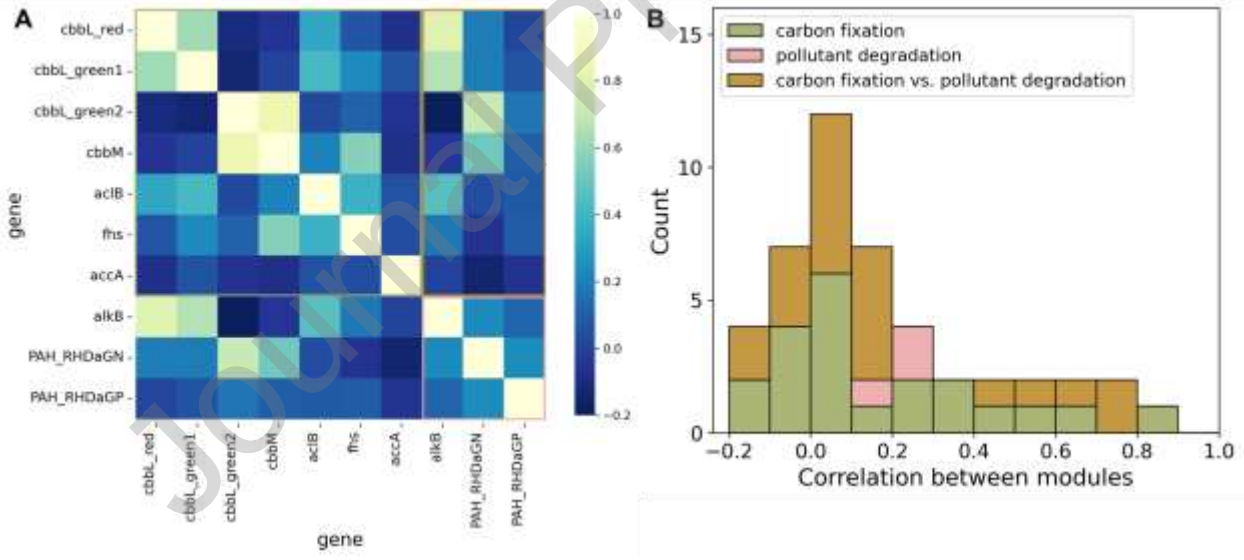


Figure 7 Distribution of correlations for qPCR function genes. (A) Heatmap illustrating the correlation coefficients; (B) The distribution of these correlation coefficients. These includes the correlation distribution between carbon fixation and degradation genes and the distribution of their inter-correlations, presented in a stacked format.

The correlations between four key genes pairs were visualized in **Supplementary Figure 9**.

Results from Supplementary Figure 9A and B revealed strong positive associations between the *alkB* pollutant degradation gene and the carbon fixation genes *cbbL_red* (R=0.798) and *cbbL_green* (R=0.66), with consistent trends observed across diverse field datasets (Supplementary Figure 9A-B). Analogous patterns were identified in **Supplementary Figures 9C and 9D**: the *PAH_RHDαGN* gene exhibits a robust correlation with the *cbbL_green2* gene (R=0.706) and a moderate correlation with the *cbbM* gene (R=0.519). These observations corroborate metagenomic data, further supporting functional interplay between carbon fixation pathways and pollutant degradation mechanisms.

3.6 Correlation analysis of compounds in metabolic pathways

Nine metabolites were identified as common nodes between carbon fixation and pollution degradation pathways: CO_2, pyruvate, acetyl-CoA, acetate, succinate, formate, succinyl-CoA, propanoyl-CoA, and fumarate. These compounds were systematically categorized into up to four reaction groups per metabolite, depending on their roles as substrates or products in pathway reactions. **Figure 8** summarizes the distribution of ko values across samples for these shared biochemical intermediates.

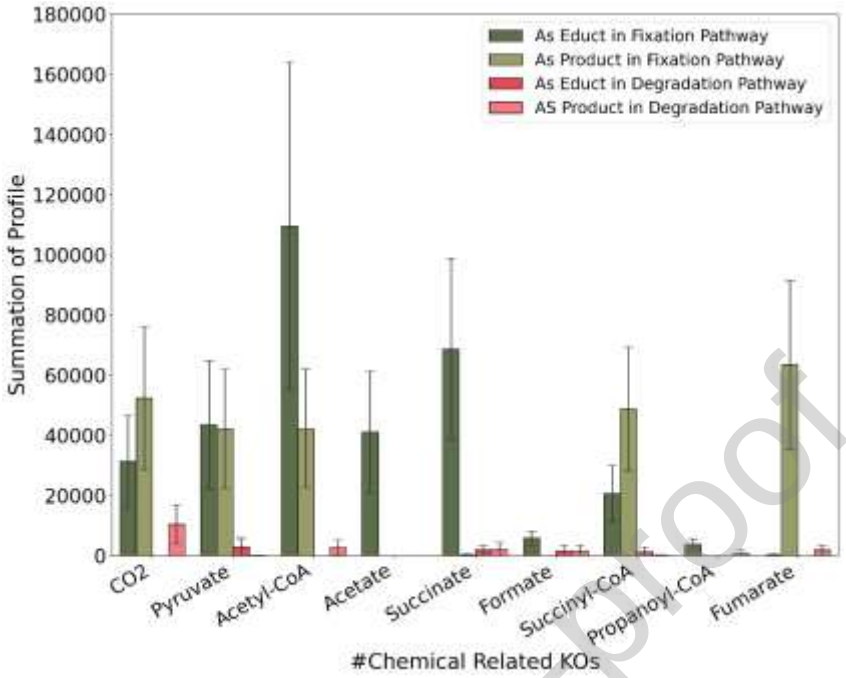


Figure 8 Total expression of compound-related genes in different pathways

Genes linked to CO_2 and Acetyl-CoA demonstrated the highest expression levels among those associated with the nine shared metabolites. We hypothesized that metabolites serving as degradation pathways intermediates might simultaneously act as substrates in carbon fixation reactions. This dual functionality prompted focused analysis on CO_2 and acetyl-CoA-associated processes. Comparative analysis revealed strong correlations (**Supplementary Figure 10**) between expression levels of genes involved in these interconnected metabolic processes, with most exhibiting $R > 0.8$.

Building on correlation analyses, we proposed an integrated microbial mechanism coupling pollutant degradation and carbon fixation in petroleum-contaminated soil (**Figure 9)**. Petorleum hydrocarbons-including alkanes, arenes, and PAHs – are sequentially metabolized through oxygenase-mediated catalysis (*alkB*, *PAH_RHDαGN*, *PAH_RHDαGP*) into hydroxylated intermediates, which oxidoreductases further convert to carboxylates. Subsequent decarboxylation of these compounds releases bioavailable CO_2, serving as the primary inorganic carbon source for

carbon fixation pathways. Metagenomic analyses identified two dominant carbon assimilation toutes: the reductive rTCA cycle (rTCA, M00173) [49] and Calvin-Benson-Bassham Cycle (M00165) [50], both utilizing CO_2, pyruvate, and acetyl-CoA as central metabolites. Notably, EC 1.2.7.- mediated synthesis of pyruvate from acetyl-CoA and CO_2 exemplified this metabolic integration. Quantitative assesements highlight CO_2 generation via three synergistic routes-direct pollutant decarboxylatio, acyl-CoA replenishment through aldehyde oxidation, and lyase-driven pyruvate production from long-chain substrates as the principal coupling mechanism. This framework elucidates how microbial systems concurrently achieve hydrocarbon detoxification and carbon sequestration through shared metabolic nodes, providing critical insights for bioremediation strategies in contaminated ecosystems.

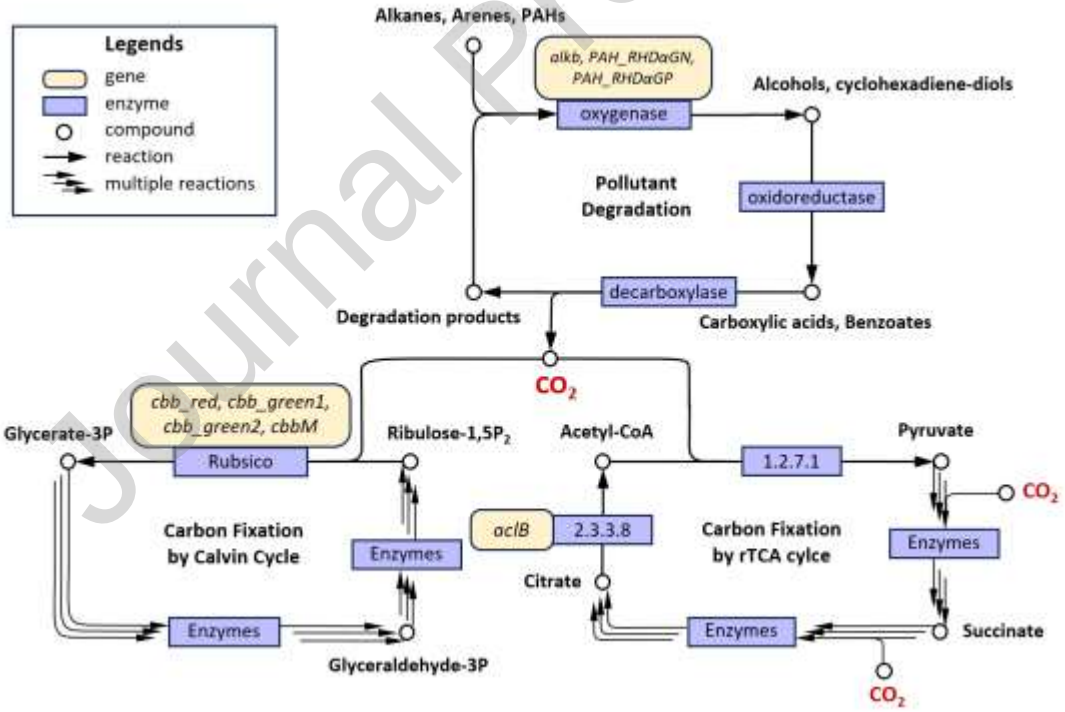


Figure 9 Pathway illustrating the coupling of carbon fixation and pollutant degradation functions.

4. Conclusion

This study established a significant positive correlation between hydrocarbon degradation genes and carbon fixation genes in oil-contaminated soils, demonstrated through integrated metagenomic sequencing and qPCR validation across diverse field samples. Functional analyses highlight particularly strong associations between alkane/polycyclic aromatic hydrocarbon oxidation genes (e.g., *alkB*, *PAH_RHDαGN*) and key carbon fixation markers., Building on these observations, we propose a metabolic coupling mechanism where CO_2 functions as the critical interface—generated via pollutant degradation through sequential oxygenation, carboxylation, and decarboxylation processes, then utilized in carbon fixation pathways dominated by the rTCA and Calvin-Benson-Bassham cycles. Crucially, isotopic tracing of CO_2 flux in targeted carbon cycle nodes could experimentally verify this model. These findings collectively advance mechanistic understanding of microbial self-remediation in oil-polluted ecosystems, informing optimized strategies for concurrent contaminant removal and carbon sequestration.

E-supplementary data for this work can be found in e-version of this paper online.

Abbreviation	
rTCA	reverse Tricarboxylic Acid
DC/4-HB	Dicarboxylate/4-Hydroxybutyrate
alkB	alkane 1-monooxygenase
PAH	Polycyclic aromatic hydrocarbon
qPCR	Quantitative Polymerase Chain Reaction
TPH	total petroleum hydrocarbons
RubisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
NH_4^+-N	ammonium nitrogen
NO_3^--N	nitrate nitrogen

TN	total nitrogen
TP	total phosphorus
AP	effective phosphorus
OC	organic carbon
PCA	principal component analysis
K00496	alkane 1-monooxygenase
K20938	long-chain alkane monooxygenase
K00446	catechol 2,3-dioxygenase
K05549	benzoate/toluate 1,2-dioxygenase

Author Contribution

Jiayu Song: Conceptualization; Writing - original draft; Data curation; Methodology; Writing - review & editing. **Yakui Chen**: Conceptualization; Writing - original draft; Data curation; Formal analysis; Methodology; Visualization; Writing - review & editing**. Yilei Han**: Conceptualization; Writing - original draft; Data curation; Validation; Writing - review & editing. **Xingchun Li**: Supervision; Writing - review & editing. **Zheng Liu**: Supervision; Writing - review & editing. **Ziyi Peng**: Data curation; Experiment; Writing - review & editing. **Aimin Fu**: Experiment. **Jinman Ren**: Experiment. **Diannan Lu**: Funding acquisition; Supervision; Writing - review & editing. **Chunmao Chen**: Writing - review & editing.

Declaration of competing interest

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

Acknowledgements

This research was supported by the National Key Research and Development Program of China under Grant No. 2018YFA0902200 and China National Petroleum Corporation Research Institute of Safety and Environmental Technology (Project No.2021DQ03-A4、No.2023ZZ1303).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to polish the sentences. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Reference

[1]. Ron, Eliora Z, and Eugene Rosenberg. 2014. 'Enhanced bioremediation of oil spills in the sea', *Current Opinion in biotechnology*, 27: 191-94.

[2]. Muthukumar, Balakrishnan, Saravanan Surya, Krithiga Sivakumar, Mohamad S AlSalhi, Tentu Nageswara Rao, Sandhanasamy Devanesan, Paulraj Arunkumar, and Aruliah Rajasekar. 2023. 'Influence of bioaugmentation in crude oil contaminated soil by Pseudomonas species on the removal of total petroleum hydrocarbon', *Chemosphere*, 310: 136826.

[3]. Liang, Yuting, Joy D Van Nostrand, Ye Deng, Zhili He, Liyou Wu, Xu Zhang, Guanghe Li, and Jizhong Zhou. 2011. 'Functional gene diversity of soil microbial communities from five oil-contaminated fields in China', *The ISME journal*, 5: 403-13.

[4]. Varjani, Sunita J. 2017. 'Microbial degradation of petroleum hydrocarbons', *Bioresource technology*, 223: 277-86.

[5]. Abbasian, Firouz, Robin Lockington, Megharaj Mallavarapu, and Ravi Naidu. 2015. 'A comprehensive review of aliphatic hydrocarbon biodegradation by bacteria', *Applied biochemistry and biotechnology*, 176: 670-99.

[6]. Wang, Wanpeng, and Zongze Shao. 2013. 'Enzymes and genes involved in aerobic alkane degradation', *Frontiers in Microbiology*, 4: 116.

[7]. Hendrickx, Barbara, Howard Junca, Jolana Vosahlova, Antje Lindner, Irene Rüegg, Margarete Bucheli-Witschel, Folkert Faber, Thomas Egli, Margit Mau, and Michael Schlömann. 2006. 'Alternative primer sets

for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site', *Journal of microbiological methods*, 64: 250-65.

[8]. Allamin, Ibrahim Alkali, Mohd Izuan Effendi Halmi, Nur Adeela Yasid, Siti Aqlima Ahmad, Siti Rozaimah Sheikh Abdullah, and Yunus Shukor. 2020. 'Rhizodegradation of petroleum oily sludge-contaminated soil using Cajanus cajan increases the diversity of soil microbial community', *Scientific reports*, 10: 4094.

[9]. Ning, Zhuo, Pingping Cai, and Min Zhang. 2024. 'Metagenomic analysis revealed highly diverse carbon fixation microorganisms in a petroleum-hydrocarbon-contaminated aquifer', *Environmental Research*, 247: 118289.

[10]. Erşan, Sevcan, and Junyoung O Park. 2020. 'Light-independent biological conversion of CO2', *Joule*, 4: 2047-51.

[11]. Onyeaka, Helen, and OC Ekwebelem. 2023. 'A review of recent advances in engineering bacteria for enhanced CO2 capture and utilization', *International Journal of Environmental Science and Technology*, 20: 4635-48.

[12]. Gao, Huan, Manli Wu, Heng Liu, Yinrui Xu, and Zeliang Liu. 2022. 'Effect of petroleum hydrocarbon pollution levels on the soil microecosystem and ecological function', *Environmental Pollution*, 293: 118511.

[13]. Chen, Kejin, Rong He, Lingyue Liu, Xin Huang, Juan Ping, Chuan Huang, Xiang Wang, and Yuanyuan Liu. 2022. 'The dominant microbial metabolic pathway of the petroleum hydrocarbons in the soil of shale gas field: carbon fixation instead of CO2 emissions', *Science of The Total Environment*, 807: 151074.

[14]. Di, H.; Zhang, M.; Ning, Z.; Liu, C.; He, Z.; Wang, S.; Kong, S.; Gan, S.; Sun, W.; Xu, Z. 2025. Metagenomic Insights into the Abundance of Iron-Reducing Microorganisms in a Petroleum-Contaminated Iron-Rich Aquifer. *Microorganisms* **,** *13*, 433.

[15]. Gogoi, Bhoirob, Islam Nazim Forid, HSarma emen.2025.Biodegradation of nitro-PAHs by multi-trait PGPR strains isolated directly from rhizosphere soil, The Microbe, 6, 100263.

[16]. Li, Yong-Quan, Ying Xin, Caili Li, Jin Liu, and Tao Huang. 2023. 'Metagenomics-metabolomics analysis of microbial function and metabolism in petroleum-contaminated soil', *Brazilian Journal of Microbiology*, 54: 935-47.

[17]. He, Zhili, Terry J Gentry, Christopher W Schadt, Liyou Wu, Jost Liebich, Song C Chong, Zhijian Huang, Weimin Wu, Baohua Gu, and Phil Jardine. 2007. 'GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes', *The ISME journal*, 1: 67-77.

[18]. Cyplik, Paweł, Marcin Schmidt, Alicja Szulc, Roman Marecik, Piotr Lisiecki, Hermann J Heipieper, Mikołaj Owsianiak, Mikhail Vainshtein, and Łukasz Chrzanowski. 2011. 'Relative quantitative PCR to assess bacterial community dynamics during biodegradation of diesel and biodiesel fuels under various aeration conditions', *Bioresource technology*, 102: 4347-52.

[19]. Laurie, Andrew D, and Gareth Lloyd-Jones. 2000. 'Quantification of phnAc and nahAc in contaminated New Zealand soils by competitive PCR', *Applied and environmental microbiology*, 66: 1814-17.

[20]. Li, Ying, Qi Shen, Xiaochi An, Yuanhuan Xie, Xiuming Liu, and Bin Lian. 2022. 'Organomineral fertilizer application enhances Perilla frutescens nutritional quality and rhizosphere microbial community stability in karst mountain soils', *Frontiers in Microbiology*, 13: 1058067.

[21]. Li, Xiaosen, Yakui Chen, Xianyuan Du, Jin Zheng, Diannan Lu, and Zheng Liu. 2021. 'Ecological response in the integrated process of biostimulation and bioaugmentation of diesel-contaminated soil', *Applied Sciences*, 11: 6305.

[22]. Krause, Hans-Martin, Joe G Ono-Raphel, Edward Karanja, Felix Matheri, Martina Lori, Yina Cifuentes, Stefanie P Glaeser, Andreas Gattinger, Amritbir Riar, and Noah Adamtey. 2023. 'Organic and conventional farming systems shape soil bacterial community composition in tropical arable farming', *Applied Soil Ecology*, 191: 105054.

[23]. Xiao, Ke-Qing, San-An Nie, Peng Bao, Feng-Hua Wang, Qiong-Li Bao, and Yong-Guan Zhu. 2014. 'Rhizosphere effect has no effect on marker genes related to autotrophic CO_2 fixation in paddy soils?', *Journal of Soils and Sediments*, 14: 1082-87.

[24]. Cébron, Aurélie, Marie-Paule Norini, Thierry Beguiristain, and Corinne Leyval. 2008. 'Real-Time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHDα) genes from Gram positive and Gram negative bacteria in soil and sediment samples', *Journal of microbiological methods*, 73: 148-59.

[25]. Rojo, Fernando. 2009. 'Degradation of alkanes by bacteria', *Environmental microbiology*, 11: 2477-90.

[26]. Wei Chenbin, Gu Wei, Tian Rong, Xu Fei, Han Yun, Ji Yuanyuan, Li Tao, Zhu Yu, Lang Peilei, Wu Wenqing. 2022. Comparative analysis of the structure and function of rhizosphere microbiome of the Chinese medicinal herb *Alisma* in different regions. *Arch Microbiol* 204, 448 .

[27]. Tao, Y., Shen, L., Han, S., Li, Z., Cui, Y., Lin, Y., ... & Zhang, Y. (2023). Metagenomic study of carbon metabolism in black soil microbial communities under lead-lanthanum stress. *Journal of Hazardous Materials*, *446*, 130666.

[28]. Zhu, Dong, Ma, Jun, Li, Gang., Rillig, Matthias. C., & Zhu, Yong. Guan. (2022). Soil plastispheres as hotspots of antibiotic resistance genes and potential pathogens. *The ISME journal*, *16*(2), 521-532.

[29]. Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics*, 31: 1674-76.

[30]. Chen, Xiaojuan, Xiaohong Wu, Yan Jian, Hongchao Yuan, Ping Zhou, Tida Ge, Chengli Tong, Dongsheng Zhou, Jinshui Wu. 2014. ‘Carbon dioxide assimilation potential, functional gene amount and RubidCO activity of autotrophic microorganisms in agricultural soils’, *Environmental Science*, 35(3): 1144-1150.

[31]. Yuan, Hongchao, Tida Ge, Xiaohong Wu, Shoulong Liu, Chengli Tong, Hongling Qin, Minna Wu, Wenxue Wei, Jinshui Wu. 2012. ‘Long-term field fertilization alters the diversity of autotrophic bacteria based on the ribulose-1,5-biphosphate carboxylase/oxygenase (RubisCO) large-subunit genes in paddy soil’, *Applied Microbiology and Biotechnology*, 95: 1061-1071.

[32]. Jiang, Q.Y., et al.2022. "Insights into carbon-fixation pathways through metagonomics in the sediments of

deep-sea cold seeps."Marine Pollution Bulletin 176:113458.

[33]. Correa, Sulamita Santos, Junia Schultz, Kyle J Lauersen, and Alexandre Soares Rosado. 2023. 'Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways', *Journal of Advanced Research*, 47: 75-92.

[34]. Gong, Fuyu, Huawei Zhu, Yanping Zhang, and Yin Li. 2018. 'Biological carbon fixation: from natural to synthetic', *Journal of CO2 Utilization*, 28: 221-27.

[35]. Beilen Jan B Van , Funhoff Enrico G. 2007. Alkane hydroxylases involved in microbial alkane degradation.*Applied Microbiology and Biotechnology*. 74(1):13-21.

[36]. Li Liu, Liu Xueqian, Yang Wen, Xu Feng, Wang Wei, Feng Lu, Bartlam Mark, Wang Lei and Rao Zihe.2008. Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. J Mol Biol. 376(2):453-65.

[37]. Mesarch MB, Nakatsu CH, Nies L. 2000 Development of catechol 2,3-dioxygenase-specific primers for monitoring bioremediation by competitive quantitative PCR. *Appl Environ Microbiol*. 66(2):678-83.

[38]. Tian Fang, Guo Guang, Fu Weilian, Li Shiji, Ding Keqiang, Yang Feng, Liang Chengyue.2024. Decolorization and detoxification of Brilliant Crocein GR by a newly enriched thermophilic consortium. *J Environ Manage*. 356:120623.

[39]. Liao Jingqiu, Wang Jie, Jiang Dalin, Wang Michaelcai, Huang Yi. 2015. Long-term oil contamination causes similar changes in microbial communities of two distinct soils. *Applied microbiology and biotechnology*, 2015, 99: 10299-10310.

[40]. Zhang D, Hu Q, Wang B, 2023. Effects of single and combined contamination of total petroleum hydrocarbons and heavy metals on soil microecosystems: Insights into bacterial diversity, assembly, and ecological function. *Chemosphere*, 345: 140288.

[41]. Wani Atif Khurshid, Akhtar Nahid, Sher Farooq, Navarrete Acacio Aparecido and Américo-Pinheiro Juliana Heloisa Pinê 2022. Microbial adaptation to different environmental conditions: molecular perspective of evolved genetic and cellular systems. Archives of Microbiology, 204(2): 144.

[42]. Brzeszcz, Joanna, and Paweł Kaszycki. 2018. 'Aerobic bacteria degrading both n-alkanes and aromatic hydrocarbons: an undervalued strategy for metabolic diversity and flexibility', *Biodegradation*, 29: 359-407.

[43]. Feng, Lu, Wei Wang, Jiansong Cheng, Lei Wang. 2007. 'Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir'. *Biological Sciences*, 104(13): 5602-5607.

[44]. Smits T H M, Witholt B, van Beilen J B. 2003. Functional characterization of genes involved in alkane oxidation by *Pseudomonas aeruginosa*. Antonie Van Leeuwenhoek, 84: 193-200.

[45]. Simon, M J, T D Osslund, R Saunders, B D Ensley, S Suggs, A Harcourt, W C Suen, D L Cruden, D T Gibson, G J Zylstra. 1993. ‘Sequences of genes encoding naphthalene dioxygenase in Pseudomonas putida strains G7 and NCIB 9816-4’. *Gene*, 15, 127(1): 31-37.

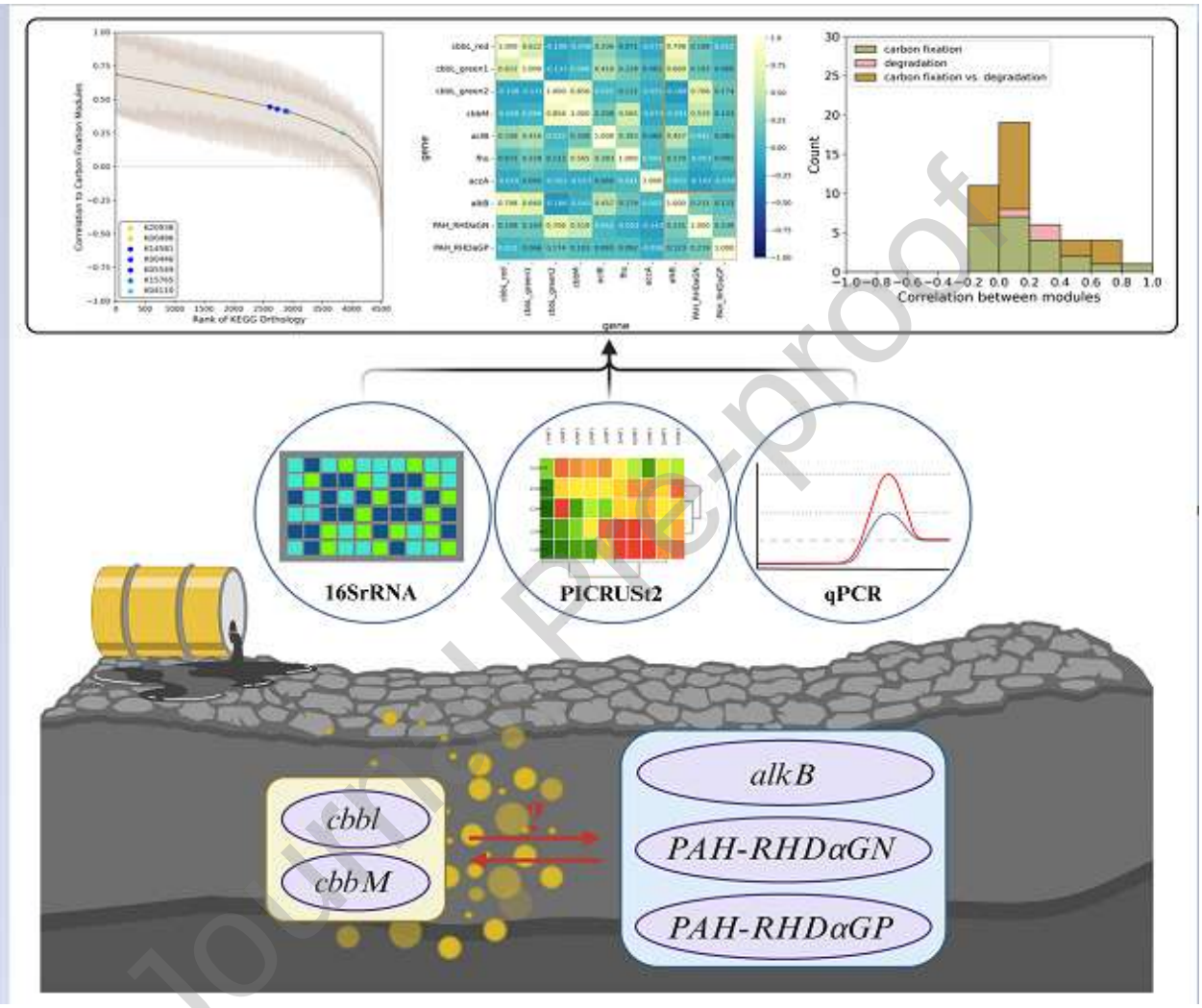
[46]. Benjamin, R C, J A Voss, D A Kunz. 1991. 'Nucleotide sequence of xylE from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7'. Journal of Bacteriology, 173(8): 2724-2728.

[47]. Yen, K M, M R Karl, L M Blatt, M J Simon, R B Winter, P R Fausset, H S Lu, A A Harcourt, K K Chen. 1991. 'Cloning and characterization of a Pseudomonas mendocina KR1 gene cluster encoding toluene-4-monooxygenase'. Journal of Bacteriology, 173(17): 5315-5327.

[48]. Schühle K，Gescher J, Feil U, Paul M, Jahn M, Schagger H, Fuchs G. 2003. 'Benzoate-Coenzyme A Ligase from *Thauera aromatica*: an Enzyme Acting in Anaerobic and Aerobic Pathways'. *Journal of Bacterilolgy*, 185(16): 4920-4929.

[49]. Norio, Kitadai, Masafumi Kameya, Kosuke Fujishima. 2017. 'Origin of the Reductive Tricarboxylic Acid (rTCA) Cycle-Type CO_2 Fixation: A Perspective'. *Life*, 7(4): 39.

[50]. Tamoi, M, Miyasaka H, Shigeoka S, Kanaboshi H. 2001. 'Molecular mechanisms of the resistance to hydrogen peroxide of enzymes involved in the Calvin cycle from halotolerant Chlamydomonas sp W80'. *Archives of Biochemistry and Biophysics*, 390(2): 176-185.



Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Diannan Lu reports was provided by the National Key Research and Development Program of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Analyzed 44 samples from 6 oil-contaminated sites of China, each with distinct physicochemical properties and levels of contamination.
- Employed metagenomic sequencing and qPCR experiments to reveal synergy between carbon fixation and pollution reduction gene abundances.
- Confirmed positive correlations between the alkane-oxidizing gene *alkB* and the aromatic-oxidizing gene *PAH_RHDαGN* with specific carbon fixation genes through qPCR validation.
- The potential coupling mechanism between carbon fixation and pollutant degradation pathways was revealed.