NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Abbasian, Firouz; Palanisami, Thavamani; Megharaj, Mallavarapu; Naidu, Ravi; Lockington, Robin; Ramadass, Kavitha "Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis". Published in Biotechnology Progress Vol. 32, Issue 3, p. 638-648 (2016)

Available from: <http://dx.doi.org/10.1002/btpr.2249>

This is the peer reviewed version of the following article: Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R. and Ramadass, K. (2016), Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis. Biotechnology Progress, 32: 638–648, which has been published in final form at <http://dx.doi.org/10.1002/btpr.2249>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Accessed from: <http://hdl.handle.net/1959.13/1344348>

Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis

Firouz Abbasian^{1,2}, Robin Lockington^{2,3}, Kavitha Ramadass^{2,3}, Thavamani Palanisami^{1,2}, Ravi Naidu^{1,2}, Mallavarapu Megharaj^{1,2}

Firouz Abbasian	(Firouz.abbasian@gmail.com)	+61-8-883-25094
Robin Lockington	(Robin.Lockington@unisa.edu.au)	+61-8-830-26272
Kavitha Ramadass	(Kavitha.Ramadass@unisa.edu.au)	+61-8-830-26288
Thavamani Palanisami	(Thava.Palanisami@newcastle.edu.au)	+61-8-830-26274
Ravi Naidu	(Ravi.Naidu@newcastle.edu.au)	+61-8-830-23057
Mallavarapu Megharaj	(Megh.Mallavarapu@newcastle.edu.au)	+61-8-830-25044

Affiliation:

- 1- Global Centre for Environmental Remediation (GCER), Faculty of Science and Information Technology, the University of Newcastle, Callaghan, NSW2308, Australia.
- 2- Cooperative Research Centre for Environmental Risk Assessment and Remediation of the Environment (CRC CARE), The University of Newcastle, Callaghan, NSW, Australia.
- 3- Future Industry Institute (FFI), University of South Australia, Australia.

Running title: Metagenomic study on Crude oil filed

#Address correspondence to Firouz Abbasian

Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis

Abstract

Soils contaminated with crude oil are rich sources of enzymes suitable for both degradation of hydrocarbons through bioremediation processes and improvement of crude oil during its refining steps. Due to the long term selection, crude oil fields are unique environments for the identification of microorganisms with the ability to produce these enzymes. In this metagenomic study, based on Hiseq Illumina sequencing of samples obtained from a crude oil field and analysis of data on MG-RAST, Actinomycetales (9.8%) were found to be the dominant microorganisms, followed by Rhizobiales (3.3%). Furthermore, several functional genes were found in this study, mostly belong to Actinobacteria (12.35%), which have a role in the metabolism of aliphatic and aromatic hydrocarbons (2.51%), desulfurization (0.03%), element shortage (5.6%) and resistance to heavy metals (1.1%). This information will be useful for assisting in the application of microorganisms in the removal of hydrocarbon contamination and/or for improving the quality of crude oil.

Keywords: Metagenomic study, crude oil well, microbial diversity, alkanes, aromatic hydrocarbons

1. Introduction

Soil contamination with crude oil is a serious concern for human and ecological health ¹. Annually millions of tons of crude oil are extracted to supply energy and primary materials for industries, and this demand has been increasing as a result of growth in population and industries ². In addition to the natural release of hydrocarbons, several industrial activities, including crude oil extraction, transportation, refining and utilisation, especially in automobile service stations, cause environmental pollution ². This situation deteriorates with time when the components are exposed to weathering processes, which lead to increases in the relative levels of higher molecular weight hydrocarbons with higher density and viscosity ³. Although several mechanical and chemical techniques are available for cleaning these contaminants from the environments, *in situ* bioremediation is a very efficient and cost effective approach ⁴. Since the use of native microorganisms for *in situ* removal of these contaminants has shown promise in several studies ⁴, an investigation of the diversity of microorganisms in contaminated areas and of the effects of environmental conditions on the degradation ability of these organisms can improve the management of the bioremediation rate in these sites. Although many studies have been performed to investigate the microbial diversity and the genes involved in degradation of aliphatic and aromatic hydrocarbons ^{5,6}, most of these studies focussed on a special community of microorganisms or on a limited number of genes of interest. Furthermore, these studies mostly focussed on the genes involved in the degradation of hydrocarbons or the production of surfactants, and less consideration has been given to other genes effective in bioremediation, such as heavy metal resistant genes and the genes required for acquisition of essential nutrients. In addition, since the levels of nitrogen, phosphorus, sulfur and iron of the crude oil contaminated sites are very low due to the predominance of hydrocarbons ⁶, investigations of the genes involved in the acquisition of these nutrients can also be useful in the development of optimized physical, chemical and environmental conditions that may improve rate and efficiency of biodegradation process.

Since microbial culture is limited to the isolation and identification of limited numbers of microorganisms, this technique is not suitable for investigation of whole microbial diversity and their genomes of all microorganisms present in a particular environment ⁷. However, recent metagenomic approaches enable researchers to analyse the whole microbial diversity, and the genetic capacity for the active metabolic pathways present in a given environment ⁸. This technique is especially efficient for the study of the biodiversity and genome analysis of complex environmental samples where most of microorganisms cannot be cultured under normal laboratory conditions ⁷. Moreover, this system enable operators to establish a correlation between microbial diversity and

the level of hydrocarbon(s) in a contaminated site ⁷. Up to now, several studies have been performed on the microbial communities and metagenomic capacities of soils contaminated with crude oil and its derivatives ⁹⁻¹¹. These studies have revealed large variations in the biodiversity and abundances of microorganisms in different geographical locations. Furthermore, big differences were observed in the genetic capacities of microbial communities identified at each location. Therefore, studies such as these give excellent opportunities for finding new microbial strains and the genes involved in both bioremediation of hydrocarbon contaminants and improving crude oil quality through refining biological process. In this study, the whole microbial diversity of a crude oil contaminated field soil and the genes involved in the degradation of hydrocarbons and microbial adaptations to this environment was investigated using a metagenomic approach.

2. Material and methods

Overall, 10 kg surface (0-10 cm depth) soil were sampled from different areas at the vicinity of crude oil wells (Perth, Western Australia, Australia). Contaminated soil samples were transported from the field to the laboratory under standard refrigerated (4° C) conditions. To ensure the quality of the work, following sifting the soil using a 2 mm metal sieve, the soil samples were thoroughly mixed and the sampling for DNA extraction was performed from different parts of the soil samples. The soil characteristics were measured according to the protocols described in text books ^{12,13} as 6.8 pH at room temperature, 2.8% moisture and 612 µS conductivity.

2-1. TPH analysis

Following extraction of TPH compounds from soil samples using the sequential ultrasonic solvent extraction method ¹⁴, the level of contamination was quantified by a gas chromatograph fitted with a flame ionization detector (GC-FID, Agilent 6890). The Chromatography process was conducted on a fused-silica capillary column BPX-5 from SGE (15 m × 0.32 mm internal diameter) coated with HP-5 (0.10-µm film thickness). Helium (2.5 mL min⁻¹) was used as the carrier gas, and the FID detector was adjusted at 300°C. Splitless injection with a sample volume of 1 µL was applied and the oven temperature was increased from 50 °C to 300 °C at a gradient of 25 °C min⁻¹ and held at this temperature for 5 min. Overall, the total run time was 19.6 minutes. The quantification of hydrocarbons were performed by Agilent Chemstation Software through integration and calibration of peaks of a standard concentration of an external calibration standard namely Hydrocarbon Window Defining Standard (C₈-C₄₀) from AccuStandard ¹⁵. Overall, five concentrations of external calibration standard in the range expected in the samples were analysed, and a linear curve fit with a R² value of 0.997 was obtained. The CCV (Continuing

Calibration Verification) was analysed at the start and end of every 20 samples and CCV recovery was 95-110% of true value. Hexane was run as blank with every 10 samples to test for cross contamination. The surrogate (*o*-terphenyl) was spiked at a level to produce a recommended extract concentration of 20 µg/mL. Surrogate recoveries ranged 70-120% for all the samples, and 25 mg kg⁻¹ was the minimum concentration of TPH detected (MDL) through this analytical method. To confirm the quality of the TPH analysis, the soil was vortexed properly and four samples were taken for this analysis.

2-2. DNA sequencing using Illumina Hiseq platform and data analysis

To confirm the quality of DNA sequencing, the soil was vortexed properly and two samples were taken for metagenomic analysis. The DNA from the soils was extracted using a power soil DNA kit (MO BIO), according to manufacturer's instructions. First, after a gentle vortex of 0.25 g sediment samples, the contents were mixed with 60 µl solution C1. Following a vigorous vortex step for 10 minutes and a centrifuge at 10,000 x g for 30 S, the supernatants (400-500 µl) were mixed up with 250 µl solution C2 in clean 2-ml tubes. These tubes were shaken gently for 5 S and were passed a 4 °C incubation for 5 minutes. After separation of the supernatant by centrifugation at 10,000 x g for 60 S, up to 600 µl supernatants were poured into a new 2-ml tubes, where were mixed with 200 µl solution C3. Once again, there was an incubation period at 4 °C for 5 minutes and a centrifugation at 10,000 x g for 60 S. A total of up to 750 µl of the resulting supernatants were mixed up with 1200 µl of solution C4 in new clean 2-ml tubes. Following a gentle vortex, a portion of these mixtures (675 µl each time) were loaded onto a spin filter, where centrifugation force at 10,000 x g for 60 S made the mixture pass through the filters and DNA stuck on. This step continued to centrifuge all the mixture. These filters were wash out by adding 500 µl solution C5 and a centrifugation at 10,000 × g for 30 seconds. After transfer of the spin filters into a new 2-ml tube and addition of 100 µl solution C6, a centrifugation step at 10,000 × g for 30 seconds eluted the DNAs into the solution buffer. The levels of DNAs were quantified by a quantifluor dsDNA system (Promega) and were adjusted to a minimum of 100 ng/µl. The DNA samples were then sequenced on the Illumina Hiseq platform by the Australian Genome Research Facility (AGRF).

The raw data obtained from Illumina platform were submitted directly in Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) ¹⁶, where all the data was trimmed by removing uninformative and/or

duplicative reads before any further studies. The microbial diversity of the data was determined using M5RNA annotation source (minimum percentage identity cut off of 97%, an E-value cutoff of $1e-5$ and minimum alignment length cutoff of 50 bp). Furthermore, the M5NR annotation source based on a minimum percentage identity cut off of 70%, an E-value cutoff of $1e-5$ and a minimum alignment length cutoff of 15 bp was used to analyse the best hit classification of genes in this data set. In addition, the hierarchical classification of the genes and their functional abundances were determined through KO (KEGG orthology), NOG (non-supervised orthologous groups) and COG (cluster of orthologous groups) subsystem sources and based on minimum percentage identity cut off of 70%, E-value cutoff of $1e-5$ and cutoff of 15 bp. Further statistical analysis were performed using the excel program.

2-3. Nucleotide sequence accession numbers

This metagenomic analysis of crude oil field soil was registered in NCBI under ID SUB713359, BioSample accession SAMN03106209 and BioProject ID PRJNA263754, and the reads are publicly available in the MG-RAST system as A11 and A12 under project identifiers 4 4582710.3 and 4582711.3.

3. Results

3-1. Gas chromatograph assays

The gas chromatograph assays showed that this soil is highly polluted with many different types of hydrocarbons, including >C16-C34 (33317 ppm), >C10-C16 (18520 ppm) and >C34-C40 (30 ppm), and its TPH (Total Petroleum Hydrocarbon) level is largely high (51867 ppm).

3-2. Statistics of reads produced by HiSeq Illumina

The dataset obtained from the sequencing of DNA from the crude oil well field (A21 and A22) were uploaded to the online metagenomic analysis tool, MG-RAST (metagenomics Rapid Annotation using Subsystem Technology) and were automatically filtered by QC software. These two files contained 30656512 fragments with an average length of 138 ± 25 bp and 128 ± 32 bp (respectively for A21 and A22) and totally 4920764700 base pairs, among them 1259670862 reads (25%) did not pass the QC pipeline. Overall, 23221149 (75%) and 337603 (1.1%) fragments contain predicted proteins and ribosomal RNA genes, respectively, of which 7872294 (25.7%)

and 4743 (0.015%) reads belonged to identified protein and rRNA features. The total GC% value in these dataset was 63 ± 9 %.

3-3. The microbial communities in the crude oil well field

The microbial orders identified in these samples are listed in Figure 1. Regardless of the 7.4% of organisms that were unclassified, the majority of the OTUs (84.2%) identified from the data belonged to bacterial phyla, followed by eukaryota (7.5%), archaea (0.8%) and viruses (0.04%) (Fig. 2). Among the bacterial phyla, the crude oil field soils community was comprised mostly of Proteobacteria (32.1%), Actinobacteria (12.3%), Firmicutes (5.8%), Acidobacteria (2.5%), Bacteroidetes (2.3%), Planctomycetes (2.1%), Gemmatimonadetes (1.6%) and many other bacterial phyla with less than 1% abundance (Fig. 2). At the order level, Actinomycetales (9.8%), Rhizobiales (3.3%), Sphingomonadales (2.9%), Burkholderiales (2.6), Clostridiales (2.5%) and many other orders lower than 2% abundance were the dominant OTUs (Fig. 2). Fungi 2.95% (mostly Glomerellales) and many small animals from different orders (4.32%) were classified as other and were not considered for further studies.

3-4. Classification of microorganisms based on their metabolism and respiration

Following information in the literature ^{17,18} and based on microbial ability to supply their carbon and energy sources, the known microbial orders found in this study were categorized into four different metabolic classes (photoautotrophic microorganisms, photoheterotrophic bacteria, chemolithotrophic bacteria and chemoheterotrophic bacteria). Since some particular species of microorganisms in an order can show a completely different metabolic pathway in comparison to other strains from the same order, the properties of each strain were analysed separately. Based on this analysis, the majority (88%) of these microorganisms belonged to chemoheterotrophic bacteria, followed by photoautotrophic microorganisms (3%), photoheterotrophic bacteria (5%) and chemolithotrophic bacteria (4%) (Fig. 3). Most of the chemoheterotrophic bacteria in this field belonged to aerobic chemoheterotrophic bacteria (62%), followed by facultative aerobic/anaerobic bacteria (23%) while absolute anaerobic bacteria consisted only 15% of this community (Fig. 4).

Based on analysis of the data using the M5NR annotation source, 98.2% of the total genes found in the soil samples collected from crude oil well field belonged to bacteria, followed by Eukaryota (0.7%), Archaea (0.7%) and viruses (0.3%) (Fig. 5). Most of these functional genes (49.14%) belonged to Proteobacteria, followed by

Actinobacteria (12.35%), Planctomycetes (4.07), Acidobacteria (3.4%), Bacteroidetes (3.1%), Firmicutes (2.04%) and many other organisms with less than 2% abundance (Fig. 5). Overall, these genes could be categorized as being involved in several metabolic pathways, including clustering (15.1%), carbohydrate metabolism (10.8%), amino acid metabolism (9.4%), protein metabolism (8.9%), miscellaneous (7.2%), cofactors and vitamins (6.8%) and many others including housekeeping genes at less than 5% (Fig. 6). The genes involved in the metabolism of aromatic hydrocarbons (1.8%) were sub-classified into the genes involved in aerobic (14%) and anaerobic (18%) pathways as well as the genes involved into peripheral (46%) and intermediate (22%) metabolisms.

4. Discussion

Environmental contamination due to petroleum hydrocarbons is of growing public concern as large amounts of these compounds are being released deliberately or inadvertently into the environment ². Crude oil extraction processes lead to the release of massive amounts of these compounds to soil and water, and therefore, crude oil fields are highly contaminated with several types of aliphatic and aromatic hydrocarbons originating from crude oil ¹. Although this phenomenon can lead to tragic environmental problems such as contamination of water sources, these areas also become enriched sources of hydrocarbon degrading microbial populations and the corresponding genes for the degradation of hydrocarbon contaminants². Metagenomics analysis of such environments is a promising approach that allows analysis of these microbial communities and all the functional genes involved in both the adaptation of microorganisms to crude oil contamination and the degradation of aliphatic and aromatic hydrocarbons ⁷. In this study, a HiSeq Illumina platform was employed to investigate the microbial population of soil samples collected from a crude oil well field located in Perth, Australia.

Based on the results, bacteria (mostly Proteobacteria, Actinobacteria and Firmicutes) were the dominant microorganisms in this field, and eukaryotes and archaea comprised only a small portion of this community. Similar results were reported by Yergeau et al ¹⁹ and Liang et al ⁹ through metagenomic analysis on the soil samples obtained from Canadian High Arctic Soils and different areas of China, respectively. Several members of order Actinomycetales were identified in this field, among them genera *Mycobacterium* sp., *Geodermatophilus* sp., *Nocardia* sp. and *Nocardioides* sp., were the most dominant microorganisms. While contamination with crude oil could increase the abundance of *Mycobacterium* sp. in the Canadian High Arctic Soils too, the abundances of

Nocardia sp. and *Nocardioides* sp. were decreased in these samples ¹¹, showing the effects of different environmental conditions, such as soil chemistry and climate, on the microbial community.

Overall, 85% of the microbial strains identified in these samples belonged to obligatory and facultative anaerobic bacteria, which was expected because the soil samples were obtained from near the soil surface, exposed to aerobic conditions. For instance, Dongshan et al ¹⁰ showed that the majority of microorganisms in the subsurface reservoirs contaminated with crude oil belonged to methanogenic microorganisms, which shows the effects of oxygen tension on the microbial communities. In addition to a few gram positive bacteria, such as *Mycobacterium* sp., *Geodermatophilus* sp., *Nocardia* sp. and *Nocardioides* sp. several gram negative bacteria, such as Rhizobiales (mostly *Rhodopseudomonas* sp., *Bradyrhizobium* sp., *Rhizobium*, *Mesorhizobium* sp. and *Methylobacterium* sp.), Burkholderiales (mostly *Burkholderia* sp. and *Leptothrix* sp.), Rhodobacterales (mostly *Rhodospirillum* sp., *Magnetospirillum* sp., *Rhodobacter* sp. and *Azospirillum* sp.), Myxococcales (mostly *Chondromyces* sp., *Stigmatella* sp. and *Haliangium* sp.), Pseudomonadales (*Pseudomonas* sp.), Xanthomonadales (mostly *Xanthomonas* sp., *Stenotrophomonas* sp. and *Xylella* sp.) and Solirubrobacterales (*Conexibacter woesei* DSM 14684) comprised the majority of strictly aerobic chemoheterotrophic bacteria in the crude oil field. Among these bacteria, Pseudomonadales ²⁰, Burkholderiales ²¹, Xanthomonas sp. ²² and Rhodobacterales ²³ are the most frequent microorganisms used for bioremediation and removal of hydrocarbon contamination. However, obligate anaerobic bacteria comprised a small portion of this microbial community and among them, Clostridiales (2.5%) (mostly *Clostridium* sp., *Desulfotomaculum* sp., *Thermincola* sp., *Ruminococcus* sp.), Sphingomonadales and several sulfur reducing bacteria, such as Desulfovibrionales, Desulfuromonadales, and Desulfobacterales were the dominant obligate anaerobic chemoheterotrophic bacteria in this field. The presence and involvement of Clostridiales ²⁴, Sphingomonadales ²⁵ and several sulfate reducing bacteria (SRB) ²⁶ have been reported in several similar studies.

In addition to the chemoheterotrophic bacteria, there were several members of groups of microorganisms belonging to other metabolic categories. Different types of phototrophic microorganisms were found in the field community with the ability to use H₂O (oxygenic photoautotrophic microorganisms), H₂S (non-oxygenic photoautotrophic bacteria) or an organic compound (photoheterotrophic bacteria) as the source of electrons required for CO₂ fixation process. A portion (4%) of these bacteria belonged to several types of chemolithotrophic

bacteria with the ability to use mainly nitrogen, such as Nitrospirales (mainly *Candidatus nitrospira defluvii*) and Nitrosomonadales (*Nitrosomonas* sp.), and iron, especially Acidimicrobiales (*Acidithiomicrobium* sp. P2 and *Acidimicrobium ferrooxidans*). The photoautotrophic bacteria identified in this study fell into two categories; three orders, mostly Chromatiales (1.2%), chlorophyll harbouring-Sphingomonadales (*Erythrobacter* sp.) (0.5%) and a few Chlorobiales (0.1%) belonged to of anoxygenic photoautotrophic bacteria, the other category consisted of different types of oxygenic photoautotrophic bacteria (Cyanobacteria) and eukaryotic green algae. *Erythrobacter* sp. are a group of photoheterotrophic bacteria that obtain their main energy requirements from light, while depending for their growth on organic substrates as carbon sources^{17,18}. This genus of bacteria is able to utilise a variety of aromatic and aliphatic hydrocarbons through an aerobic chemoheterotrophic activity²⁵. The presence of both anoxygenic and oxygenic phototrophic microorganisms in the sites contaminated with hydrocarbons, especially crude oil, have been shown in previous studies as well^{27,28}. The production of molecular oxygen, specifically by the phototrophic microorganisms, as well as nitrogen fixation and the production of simple organics can be critical for the improvement of bioremediation of hydrocarbons by chemoheterotrophic bacteria²⁹. However, as reported by AN et al¹⁰, the phototrophic microorganisms comprise only a small portion of the microbial communities in the crude oil contaminated sites.

Data analysis using The KEGG (Kyoto Encyclopedia of Genes and Genomes) showed that the metagenome capacity of the crude oil contaminated soil was wide enough to allow most known metabolic pathways required for the growth and activity of microorganisms to proceed. The majority of these genes and metabolic pathways are involved in routine microbial activity, such as intracellular clustering processes, metabolism of carbohydrates, amino acids and proteins, cofactors, vitamins, and belonged mainly to bacteria, especially Proteobacteria and Actinobacteria. However, as we will mention below, a small portion of the genes found in this study are involved in the adaptation of the microbial population to high hydrocarbon concentrations and the degradation of these compounds in order for them to be used as a metabolite or to reduce their toxicity.

4-1. Alkane monooxygenases:

Alkane hydroxylating enzymes are a group of monooxygenases responsible for the oxidation of n-alkanes and several other compounds to their corresponding alkan-1-ol³⁰. Depending on the microorganism involved, and the electron carrier and substrates, these enzymes are classified into alkane rubredoxin-dependent monooxygenase, flavin-binding monooxygenase, CYP153 soluble cytochrome P450 (for C5–C12 alkanes), methane

monooxygenase (in methane oxidizing bacteria), propane monooxygenase (in propane-oxidizing bacteria), ammonia monooxygenase (in ammonia-oxidizing bacteria), butane monooxygenases (BMOs) and etc.^{31,32}. While CYP153 soluble cytochrome P450 monooxygenases are responsible for mono-oxidation of the alkanes with C5–C12 length carbon, flavin-binding monooxygenase (AlmA) oxidize long chain (C20–C36) alkanes. Overall, less than 0.01% of the metagenome belonged to alkane hydroxylating enzymes, mainly categorised as alkane rubredoxin-dependent monooxygenases and Flavin-binding monooxygenases (AlmA; for C20–C36 alkanes). These two genes were found in a variety of phylogenetic groups, including Actinomycetales, Burkholderiales, Pseudomonadales, Oceanospirillales, Deinococcales, Rhodobacterales and Sphingobacteriales (Table 2). A limited numbers of microorganisms in this study, including *Rhodococcus* sp. H1, *Mycobacterium marinum* M and *Dietzia cinnamomea* P4 (all belonging to Actinomycetales) and *Salinisphaera* bacterium PC39 (Salinisphaerales) harboured Cyt-P450-dependent monooxygenase genes, as well. Furthermore, the genome of two genera of Actinomycetales, *Mycobacterium* sp. (*M. smegmatis* str. MC2 155, *M. chubuense* NBB4) and *Rhodococcus* sp. (*R. imtechensis*, *Rhodococcus jostii* RHA1), contained propane monooxygenase. In the studies performed by Liang et al.⁹ on soils contaminated with crude oil, the majority of functional genes involved in hydrocarbon degradation belonged to the alkane-1-monooxygenase of *Rhodococcus erythropolis* and *Rhodococcus* sp. Q15. In addition, the soluble methane monooxygenase (sMM), which catalyses the oxidation of methane, was found in a variety of methanotrophic archaea and several bacteria, including *Acidiphilium* sp., *Amycolatopsis* sp., *Beijerinckia* sp., *Crenothrix* sp., *Frankia* sp., *Marinobacter* sp., *Mycobacterium smegmatis* str. MC2 155, *Verminephrobacter eiseniae* EF01-2 and many ammonia oxidising bacteria (Table 2). Indeed, methane monooxygenase in methane oxidizing archaea is related to Ammonia monooxygenases (AMOs) found in ammonia oxidising bacteria, which are able to hydroxylate ammonia as well as methane and many other short length alkanes³³.

4-2. Genes responsible for degradation of aromatic hydrocarbon

Overall 2.5% of the genes found in the soil samples collected from crude oil well field belonged to functional enzymes responsible for the metabolism of aromatic hydrocarbons (Table 3). Among these aromatic hydrocarbon degrading genes 0.3% were responsible for initiation of aerobic catabolism of several types of hydrocarbons. Furthermore, the enzymes responsible for initiation of anaerobic aromatic degrading pathways comprised 0.4% of the total number of the functional genes involved in aromatic hydrocarbon degradation. Other genes in this

category fall into peripheral pathways for catabolism of aromatic compounds (0.6%) and the genes involved in metabolism of central aromatic intermediates (1.2%). A full list of the genes involved in the degradation of aromatic hydrocarbons found in this study is presented in Table 3. These genes are potentially able to catabolize a large variety of aromatic hydrocarbons, such as aromatic amines, benzene, toluene, carbazol, naphthalene and Ethylbenzene, through aerobic metabolism. However, the majority of these genes were responsible for encoding Benzoate (0.13%) and Phenylacetate (0.1%) catabolic pathways. Furthermore, approximately 0.4% of the whole metagenome obtained in this study belonged to the genes encoding the enzymes involved in anaerobic degradation of aromatic hydrocarbons, mostly benzoate. Despite difference in the ratios of these genes and their sequences, a similar list of genes responsible for the degradation of aromatic hydrocarbons was reported in previous studies^{9,10} working on metagenomic studies of crude oil contaminated soils.

4-3. Genes responsible for desulfurization of sulfur-containing hydrocarbon

The desulfurization of sulfur-containing hydrocarbons is one of most important points of interest due to their role in providing sulfur supplementation for hydrocarbon remediating organisms and also in improving the quality of oil-originated fuels³⁴. Several microorganisms have been found with the hydrocarbon desulfurization ability, which employ different enzymatic reactions to remove these elements from their substrates³⁵. These enzymes can be categorized into Alkanesulfonates monooxygenase, which remove sulfur from sulfonated alkanes, and aromaticsulfonate desulfurating enzymes such as Dibenzothiophene (DBT) degrading enzymes, which convert a poly-aromatic sulfur heterocyclic compound to 2-hydroxybiphenyl (2-HBP)³⁵. The later enzyme is ordered usually as an operon consisted of three genes, including dszC (dibenzothiophene monooxygenase), dszA (Dibenzothiophene sulfone monooxygenase) and dszB (2,2-hydroxybiphenyl benzensulfonate desulfinase). The genes encoding both groups of alkane and DBT desulfurating enzymes were found in these metagenomic studies. The genes responsible for desulfurization of Alkanesulfonate, including Alkanesulfonate monooxygenase and Alkanesulfonate transport system, consisted 0.024% of whole microbial genome in these samples, mostly carried by different species of Actinomycetales, Clostridiales, Burkholderiales, Pseudomonadales and Rhizobiales. The genes responsible for desulfurization of dibenzothiophene comprised a minority of the hydrocarbon desulfurizing genes, and were mostly carried on Actinomycetales, Bacillales, Burkholderiales and Rhizobiales. The DNA sequences of many of the Alkanesulfonate and DBT desulfurizing genes found in this study showed significant differences with the existing corresponding genes in public gene databases. Since the DNA sequences of genes

give no idea of their expression and enzymatic activities, further studies are required to isolate and express these genes in laboratories.

4-4. Genes responsible for microbial adaptation and growth in hydrocarbon contaminated soils

Biosurfactants are a group of amphiphilic surface active agents with the ability to reduce surface/interfacial tensions between two immiscible fluids ³⁶. Due to improvement in the solubility of polar compounds in organic solvents, biosurfactants are used in several crude oil dependent technologies, such as transfer of crude oil, enhanced oil recovery and crude oil bioremediation ³⁶. Due to the insolubility of hydrocarbons in water, the production of these biosurfactants looks likely to be necessary for improvements to the bioavailability of these compounds ³⁷. The majority (0.8%) of biosurfactant-encoding genes in this study produce trehalose, which is a disaccharide residue on a glycolipid surfactant substituted by mycolic acids at C-6 and C-6' positions ³⁸. Polyol lipids, including (non)oxyethylenated ethers or esters of a polyol, were the second dominant (0.015%) biosurfactant producing genes. In addition, bacterial exopolysaccharide mono/di-rhamnolipids, which are produced by the action of rhamnosyltransferase 1 and rhamnosyltransferase 1 ³⁹, were the third most dominant (0.001%) group of biosurfactants in this study.

Macro/micro-elements are essential for normal growth of microorganisms, and any shortage of these elements restrict microbial growth. Depending on the elements and the enzymatic capacity of microorganisms, these essential nutrients are supplied from mineral or organic compounds present in the microbial environment. Therefore, microorganisms have to spare a portion of their genome capacity for improving the bioavailability of different elements. Since crude oil is a very deprived environment for many essential elements, especially phosphorus, nitrogen, sulfur, any contamination can lead to changes in the C:N:S:P ratio ⁶. This is the scenario expected in the soil samples obtained from the crude oil contaminated field where the microorganisms spare a significant portion of their genome capacity for improving the bioavailability and acquisition of phosphorus, nitrogen, sulfur, potassium and iron (1.63%, 1.53%, 1.34%, 0.4%, 0.7% respectively).

Furthermore, since crude oils contain different heavy elements, such as arsenic, cadmium, cobalt, zinc, cadmium, copper, mercury, chromium, zinc and many other minerals, these may have toxic effects on microbial growth and activities, ⁴⁰. Therefore, the genomes of the microorganisms present in these environments must contain the genetic capacity to enable microorganisms to be resistant to these toxic elements. Genes resistant to several types

of toxic elements were found in the dataset obtained from our given sample. The genes responsible for microbial resistance to heavy metals found in this study can be classified into two categories. The first group of these genes express several cell membrane integrated channels, such as Co/Zn/Cd efflux system membrane fusion protein, cadmium-transporting ATPase, *copA* ATPase dependent transporter, *cusCFBA* proton-cation antiporter complexes, mercuric transport protein and arsenic efflux pumps, which are involved in excretion of the heavy metals. The second category of these genes, such as multicopper oxidase, organomercurial lyase, mercuric reductase and arsenate reductase, are responsible for reducing the toxicity of the heavy metals through their oxidation, reduction or conjugation with a (in)organic compound.

5. Conclusion

To sum up, based on this metagenomic study, the soils contaminated with crude oil in the oil field area were inhabited by a diversity of microorganisms in which Actinomycetales, Rhizobiales, Burkholderiales and Clostridiales were found as dominant microorganisms. These microorganisms are rich sources of different enzymes useful for the degradation of aliphatic and aromatic hydrocarbons, the production of biosurfactants, and the improvement of the quality of petroleum throughout crude oil refining process. Furthermore, the genetic capacity of these microorganisms provides valuable information regarding the adaptation of hydrocarbon degrading microorganisms to different harsh environmental conditions, such as heavy metal toxicity, hydrocarbon hydrophobicity and the C:N:P imbalances, which are useful for the design of bioremediation systems for hydrocarbon contaminated sites. However, since crude oil wells have been established in different climatic conditions, from hot deserts to frozen tundra, the need for further studies is required to obtain more detailed information regarding the presence and participation of different microbial strains and their corresponding hydrocarbon degrading genes and to assess the activity of each stain in these differing conditions.

Conflicts of interest:

There is no conflicts of interest.

Acknowledgment

The authors would like to appreciate GCER (Global Centre for Environmental Remediation) and CRC-CARE (Cooperative Research Centre for Contamination Assessment and Remediation of the Environment) for all their supports.

References:

1. Jin HM, Kim JM, Lee HJ, Madsen EL, Jeon CO. Alteromonas as a key agent of polycyclic aromatic hydrocarbon biodegradation in crude oil-contaminated coastal sediment. *Environmental science & technology*. 2012;46(14):7731-7740.
2. Abbasian F, Lockington R, Mallavarapu M, Naidu R. A Comprehensive Review of Aliphatic Hydrocarbon Biodegradation by Bacteria. *Appl Biochem Biotech*. 2015;176(3):670-699.
3. Pietroski JP, White JR, DeLaune RD, Wang JJ, Dodla SK. Fresh and weathered crude oil effects on potential denitrification rates of coastal marsh soil. *Chemosphere*. 2015;134:120-126.
4. Yoon S. COMPOSITION, METHOD, AND APPARATUS FOR CRUDE OIL REMEDIATION. US Patent 20,150,014,221; 2015.
5. Suja F, Rahim F, Taha MR, Hambali N, Razali MR, khalid A, Hamzah A. Effects of local microbial bioaugmentation and biostimulation on the bioremediation of total petroleum hydrocarbons (TPH) in crude oil contaminated soil based on laboratory and field observations. *International Biodeterioration & Biodegradation*. 2014;90:115-122.
6. Uhlik O, Strejcek M, Hroudova M, Demnerova K, Macek T. Identification and Characterization of Bacteria with Bioremediation Potential: from Cultivation to Metagenomics. *CHEMICAL LISTY*. 2013;107(8):614-622.
7. Abbasian F, Lockington R, Mallavarapu M, Naidu R. The Integration of Sequencing and Bioinformatics in Metagenomics. *Reviews in Environmental Science and Bio/technology*. 2015;14(3):357-383.
8. Gong J-S, Lu Z-M, Li H, Zhou Z-M, Shi J-S, Xu Z-H. Metagenomic technology and genome mining: emerging areas for exploring novel nitrilases. *Appl Microbiol Biotech*. 2013;97(15):6603-6611.
9. Liang Y, Van Nostrand JD, Deng Y, et al. Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *ISME J*. 2011;5(3):403-413.
10. An D, Caffrey SM, Soh J, et al. Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environmental science & technology*. 2013;47(18):10708-10717.
11. Yergeau E, Sanschagrin S, Beaumier D, Greer CW. Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high arctic soils. *PloS one*. 2012;7(1):1-10.
12. Sparks DL. *Environmental soil chemistry*. Academic press; Amsterdam ; Boston. 2003.
13. Rayment G, Higginson FR. *Australian laboratory handbook of soil and water chemical methods*. Inkata Press Pty Ltd; 1992.
14. Risdon GC, Pollard SJ, Brassington KJ, et al. Development of an analytical procedure for weathered hydrocarbon contaminated soils within a UK risk-based framework. *Analytical chemistry*. 2008;80(18):7090-7096.
15. Miller W, Miller D. A micro-pipette method for soil mechanical analysis. *Communications in Soil Science & Plant Analysis*. Vol 181987:1-15.
16. Glass EM, Meyer F. The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*. 2011:325-331.
17. Whitman WB, Parte AC. *Systematic bacteriology*. Springer; 2009.
18. Schink B, Janssen P, Brune A. *Bergey's Manual of Systematic Bacteriology*. Springer; 2011.

- 454 19. Yergeau E, Sanschagrin S, Beaumier D, Greer CW. Metagenomic analysis of the bioremediation
455 of diesel-contaminated Canadian high Arctic soils. *PLoS one*. 2012;7(1):e30058.
- 456 20. Pacwa-Płociniczak M, Płaza GA, Poliwoda A, Piotrowska-Seget Z. Characterization of
457 hydrocarbon-degrading and biosurfactant-producing *Pseudomonas* sp. P-1 strain as a
458 potential tool for bioremediation of petroleum-contaminated soil. *Environmental Science and*
459 *Pollution Research*. 2014;21(15):9385-9395.
- 460 21. Dobslaw D, Engesser KH. Degradation of toluene by ortho cleavage enzymes in *Burkholderia*
461 *fungorum* FLU100. *Microbial biotechnology*. 2014.
- 462 22. Salam L, Ilori M, Amund O, Numata M, Horisaki T, Nojiri H. Carbazole angular dioxygenation
463 and mineralization by bacteria isolated from hydrocarbon-contaminated tropical African soil.
464 *Environmental Science and Pollution Research*. 2014;21(15):9311-9324.
- 465 23. Auffret MD, Yergeau E, Labbé D, Fayolle-Guichard F, Greer CW. Importance of *Rhodococcus*
466 strains in a bacterial consortium degrading a mixture of hydrocarbons, gasoline, and diesel oil
467 additives revealed by metatranscriptomic analysis. *Appl Microbiol Biotech*. 2014:1-12.
- 468 24. Sherry A, Gray N, Ditchfield A, et al. Anaerobic biodegradation of crude oil under sulphate-
469 reducing conditions leads to only modest enrichment of recognized sulphate-reducing taxa.
470 *International Biodeterioration & Biodegradation*. 2013;81:105-113.
- 471 25. Leys N, Ryngaert A, Bastiaens L, Top E, Verstraete W, Springael D. Culture independent
472 detection of *Sphingomonas* sp. EPA 505 related strains in soils contaminated with polycyclic
473 aromatic hydrocarbons (PAHs). *Microbial ecology*. 2005;49(3):443-450.
- 474 26. Callaghan AV, Morris BEL, Pereira IAC, et al. The genome sequence of *Desulfatibacillum*
475 *alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol*.
476 2012;14(1):101-113.
- 477 27. Jiménez N, Viñas M, Bayona JM, Albaiges J, Solanas AM. The Prestige oil spill: bacterial
478 community dynamics during a field biostimulation assay. *Appl Microbiol Biotech*.
479 2007;77(4):935-945.
- 480 28. Raghukumar C, Vipparthy V, David J, Chandramohan D. Degradation of crude oil by marine
481 cyanobacteria. *Appl Microbiol Biotech*. 2001;57(3):433-436.
- 482 29. Papizadeh M, Ardakani MR, Motamedi H, Rasouli I, Zarei M. C-S Targeted Biodegradation of
483 Dibenzothiophene by *Stenotrophomonas* sp. NISOC-04. *Appl Biochem Biotech*. 2011;165(3-
484 4):938-948.
- 485 30. Li L, Liu X, Yang W, et al. Crystal Structure of Long-Chain Alkane Monooxygenase (LadA) in
486 Complex with Coenzyme FMN: Unveiling the Long-Chain Alkane Hydroxylase. *J Mol Bio*.
487 2008;376(2):453-465.
- 488 31. Redmond MC, Valentine DL, Sessions AL. Identification of novel methane-, ethane-, and
489 propane-oxidizing bacteria at marine hydrocarbon seeps by stable isotope probing. *Appl*
490 *Environ Microbiol*. 2010;76(19):6412-6422.
- 491 32. Sayavedra-Soto L, Hamamura N, Liu C, Kimbrel J, Chang JH AD. The membrane-associated
492 monooxygenase in the butane-oxidizing Gram-positive bacterium *Nocardioides* sp. strain CF8
493 is a novel member of the AMO/PMO family. *Environmental Microbiology Reports*.
494 2011;3:390–396.
- 495 33. De La Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. Cultivation of a thermophilic
496 ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol*. 2008;10(3):810-
497 818.
- 498 34. Milam SN, Reynolds MA, Wellington SL, Buhrman FA. Process for treating a hydrocarbon-
499 containing feed. Google Patents; 2014.
- 500 35. Firouz Abbasian, Robin Lockington, Mallavarapu Megharaj, Naidu R. A Review on the Genetics
501 of Aliphatic and Aromatic Hydrocarbon Degradation. *Appl Biochem Biotech*. 2016; 178(2): 224-
502 250.
- 503 36. Saharan BS, Sahu RK, D S. A Review on Biosurfactants: Fermentation, Current Developments
504 and Perspectives. *Genetic Engineering and Biotechnology Journal*. 2011;2011:1-39.

505 37. Soberón-Chávez G, Maier RM. Biosurfactants: a general overview. *Biosurfactants*: Springer;
506 2011:1-11.

507 38. Lang S, Syldatk C, Rau U. 17 Enzymatic Synthesis and Modification of Glycolipids. *Enzymes in*
508 *Lipid Modification*. 2000;40:363.

509 39. Müller MM, Kügler JH, Henkel M, et al. Rhamnolipids—next generation surfactants? *Journal*
510 *of biotechnology*. 2012;162(4):366-380.

511 40. Osuji LC, Onojake CM. Trace Heavy Metals Associated with Crude Oil: A Case Study of Ebocha-
512 8 Oil-Spill-Polluted Site in Niger Delta, Nigeria. *Chemistry & biodiversity*. 2004;1(11):1708-
513 1715.

514

- 1) A metagenomic study performed on soil obtained from crude oil field
- 2) A full list of microbial diversity was obtained from this study.
- 3) A full list of functional genes involved in degradation of aliphatic hydrocarbon was obtained from this analysis.
- 4) A full list of functional genes involved in degradation of aromatic hydrocarbon was obtained from this analysis.
- 5) The genes required to supply essential elements for the microorganisms in this sample was obtained from this analysis.
- 6) The genes involved in increase resistance of these microorganisms to heavy metals present in the crude oil was listed in this manuscript

Table 1. List of the genes and microorganisms involved in degradation of aliphatic and aromatic hydrocarbons identified in the metagenomic data analysis crude oil filed soil.

Enzymes		Orders	Ratios
	Alkane rubredoxin-dependent monooxygenase	Actinomycetales, Burkholderiales, Pseudomonadales, Flavobacteriales, Legionellales, Oceanospirillales, Rhodobacterales, Rhodospirillales, Myxococcales, Cytophagales, Deinococcales,, Solirubrobacteriales, Sphingobacteriales	0.006%
	Alpha-ketoglutarate-dependent dioxygenase AlkB	Burkholderiales, Enterobacteriales	0.00004%
	Bacterial Flavin-binding monooxygenase (AlmA; for C20–C36 alkanes)	Actinomycetales, Alteromonadales, Burkholderiales, Rhodobacterales, Pseudomonadales, Oceanospirillales, Rhizobiales, Deinococcales, Rhodobacterales, Sphingobacteriales, several Fungi (<i>Ajellomyces dermatitidis</i> SLH14081, <i>Coccidioides posadasii</i> str., <i>Grosmannia clavigera</i> kw1407, <i>Metarhizium anisopliae</i> ARSEF 23)	0.003%
	Propane monooxygenase	Actinomycetale	0.0005%
	Bacterial CYP153: cytochrome P450 (for C5–C12 alkanes)	Actinomycetales, Salinisphaerales	0.0005%
	Methane monooxygenase/amm onia monooxygenase, subunits A, B and C	Actinomycetales, Methylococcales, Nitrosomonadale, Nitrosopumilales, Alteromonadales, Burkholderiales, <i>Chromatiales</i> , Rhizobiales, Rhodospirillales, Sphingobacteriales	0.0008%
	Aerobic pathways	Aromatic Amine Catabolism, Benzoate transport and degradation cluster, Cresol degradation, Gentisate degradation, Phenylacetyl-CoA catabolic pathway (core), Toluene 4-monooxygenase (T4MO), carbazol degradation cluster p-cymene degradation	0.3%
	Anaerobic pathways	Acetophenone carboxylase 1, Anaerobic benzoate metabolism, Anaerobic toluene and ethylbenzene degradation	0.4%
	Central aromatic intermediates	4-Hydroxyphenylacetic acid catabolic pathway, Catechol branch of beta-ketoadipate pathway, Central meta-cleavage pathway of aromatic compound degradation, Homogentisate pathway of aromatic compound degradation, N-heterocyclic aromatic compound degradation, Protocatechuate branch of beta-ketoadipate pathway, Salicylate and gentisate catabolism	0.6%
	Peripheral pathways	Benzoate catabolism, Biphenyl Degradation, Chloroaromatic degradation pathway, Naphtalene and anthracene degradation, Phenol hydroxylase, Phenylpropanoid compound degradation, Quinate degradation, Salicylate ester degradation, Toluene degradation, n-Phenylalkanoic acid degradation, p-Hydroxybenzoate degradation	1.2%

Table 2. List of the genes and microorganisms involved in desulfurization of hydrocarbons identified in crude oil metagenomic data analysis

Pathway	Genes	Microorganisms	Ratio
	Alkanesulfonate monooxygenase (EC 1.14.14.5)	Actinomycetales, <i>Bacillales</i> , Burkholderiales, Enterobacteriales, Pseudomonadales, Rhizobiales, Caulobacteriales, Herpetosiphonales, Rhodospirillales, Sphingomonadales, several Algae and Cyanobacteria such as <i>Anabaena variabilis</i> ATCC 29413, <i>Nostoc punctiforme</i> PCC 73102, <i>Microcystis aeruginosa</i> NIES-843,	0.004%
	Alkanesulfonates transport system permease protein	Actinomycetales, Burkholderiales, Clostridiales, Enterobacteriales, Pseudomonadales, Rhizobiales, Rhodospirillales, Rhodobacteriales, Desulfuromonadales, Bifidobacteriales, Thermoanaerobacteriales, Sphingomonadales, Solirubrobacteriales, Methanosarcinales, Methylophilales, Synergistales, an unclassified bacterium (<i>Thermobaculum terrenum</i> ATCC BAA-798), Caulobacteriales and Myxococcales	0.02%
	dibenzothiophene (DBT) monooxygenase (dszC)	Actinomycetale, <i>Bacillales</i> , Burkholderiales, Rhizobiales and Rhodospirillales	0.002%
	DBT-sulfone monooxygenase (dszA)	Actinomycetales, <i>Bacillales</i> , Burkholderiales, Pseudomonadales and Rhizobiales	0.001%
	dibenzothiophene-5,5-dioxide monooxygenase (dszB)	Actinomycetales, <i>Bacillales</i> , Burkholderiales, Pseudomonadales, Rhizobiales, Solirubrobacteriales and <i>Cyanobacteria</i>	0.001%

Table 3. Ratios of the genes involved in metal resistance and shortage of elements identified in crude oil metagenomic data.

Ability	Elements	Mechanism	Ratios
	Arsenic	Arsenate reductase, Arsenic efflux pump protein, Arsenic resistance protein ArsH, Arsenical pump-driving ATPase, Arsenical-resistance protein ACR3, Arsenical regulatory operons	0.1%
	Cadmium	Cadmium resistance protein, Cadmium-transporting ATPase	
	Cobalt-zinc-cadmium	Cadmium-transporting ATPase, Cation efflux system proteins CusA, Cobalt-zinc-cadmium resistance proteins CzcA and CzcD, Heavy metal RND efflux outer membrane protein (CzcC family), Copper sensory histidine kinase CusS, Heavy metal sensor histidine kinase, Hypothetical protein involved in heavy metal export, Nickel-cobalt-cadmium resistance protein NccB, Putative silver efflux pump, Transcriptional regulatory genes (MerR family, Cd(II)/Pb(II), CusC, CusR, HmrR,	0.007%
	Copper	CopG protein, Copper chaperone, Copper resistance proteins C and D, Copper-translocating P-type ATPase (EC 3.6.3.4), Multicopper oxidase, Magnesium and cobalt efflux protein CorC, CutF precursor, CusS, CusR,	0.84%
	Mercury	Mercuric ion reductase, Mercuric transport proteins (MerC, MerE and MerT), Organomercurial lyase, Periplasmic mercury(+2) binding protein and Mercuric resistance operon regulators	0.09%
	Chromium	Chromate resistance proteins: ChrB, ChrA, ChrC Superoxide dismutase SodM-like protein ChrF	0.03%
	Zinc	Response regulator of zinc sigma-54-dependent two-component system, Zinc resistance-associated protein	0.02%
	Sulfur	Arylsulfatase, Sulfite reductase (HmeA, HmeD, HmeC, HmeB), Sulfur oxidation (Sulfide dehydrogenase, Sulfite oxidase, SoxC, SoxB, SoxD, SoxY, SoxZ, SoxA, SoxR, SoxS, SoxH), Thioredoxin-disulfide reductase, Inorganic Sulfur Assimilation (sulfur ABC-type transporter, Adenylyl-sulfate reductase, Adenylylsulfate kinase, Adenylylsulfate reductase), uptake and metabolism of Alkanesulfonate, DMSP, L-Cystine, Taurine and glutathione	1.34%
	Nitrogen	Allantoin Utilization, Amidase, Ammonia assimilation (Glutamate synthase, Glutamate-ammonia-ligase adenylyltransferase, Glutamine amidotransferase), Cyanate hydrolysis, Denitrification (nitrite reductase, Nitric oxide reductase, Nitrous oxide reductase and their accessory proteins), nitrite reductase, Nitric oxide synthase oxygenase, Nitric-oxide synthase, Nitrilase, Nitrogen fixation (Nitrogenase and its associated proteins)	1.53%
	Phosphorus	Alkylphosphonate utilization, Phosphate ABC transporter, Phosphate-binding DING proteins	1.63%
	Iron	Ferrous and Ferric Transport system, Heme uptake and utilization, Iron(III) dicitrate transport system Fec, Siderophores (Alcaligin, Achromobactin, Aerobactin, Anthrachelin, Bacillibactin, Enterobactin, Petrobactin, Pyoverdine, Petrobactin, Salmochelin, Desferrioxamine E, pyochelin, Staphylobactin, Yersiniabactin, Vibrioferriin)	0.7%
	potassium	efflux system ATP-binding protein and ancillary proteins (KefG, and KefC), K transporter trk, KdpE, KdpD, K/H antiporter ROSB, Kef-type transport system 2, FKBP-type peptidyl-prolyl cis-trans isomerase FkpA precursor, Glutathione-regulated potassium-efflux system ATP-binding protein, Kup system potassium uptake protein, Large-conductance mechanosensitive channel, Potassium channel protein, Potassium voltage-gated channel subfamily KQT, Potassium-transporting ATPase A, B and C chains, Voltage-gated potassium channel, cAMP-dependent Kef-type K ⁺ transport system, pH adaptation potassium efflux system A	0.4%