Title

Microbial community responses to different volatile petroleum hydrocarbon class mixtures in an aerobic sandy soil

Author names and affiliations

George Mangse[a,b,*], David Werner[a], Paola Meynet[a] and Chukwuma C. Ogbaga,[b,c]
[a]School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, England, UK.
[b]Department of Microbiology & Biotechnology, Nile University of Nigeria, Abuja, Nigeria
[c]Department of Biological Sciences, Nile University of Nigeria, Abuja, Nigeria

***Corresponding author:**

George Mangse
Email address: george.mangse@nileuniversity.edu.ng
Present/permanent address: Department of Microbiology & Biotechnology, Nile University of Nigeria, Abuja, Nigeria

Highlights

- Volatile petroleum hydrocarbon (VPH) class mixtures stimulated microbial activity.
- Microbial diversity decreased following addition of different VPH class mixtures.
- VPH classes contributed to shaping the microbial communities in batch treatments.
- Bacterial species demonstrated varying responses to different VPH class mixtures
- NGS platforms did not alter the outcomes of microbial community structures.

Abstract

Volatile Petroleum Hydrocarbon (VPH) class effects on soil microbial composition were investigated using two next-generation sequencing (NGS) techniques – 454 pyrosequencing and ion torrent sequencing. Microbial activity was stimulated by adding different VPH compound classes to the sandy soil in comparison with live controls without VPH addition. Microbial community structure was significantly affected by the various VPH classes. At the genus level, *Rhodococcus, Desulfosporosinus*, *Polaromonas, Mesorhizobium* and *Methylibium* had the highest relative abundances in the straight-chain alkane (str-alk) treated soil as compared to the control ($p < 0.05$, 2 sample t-tests) while *Pseudomonas* was more dominant in the cyclic alkane (cyc-alk) contaminated soil. *Pseudonocardia* was significantly higher in relative abundance in the aromatic hydrocarbon (aro-H) treated batches as compared to the control ($p < 0.05$, 2 sample t-tests). A non-metric multidimensional scaling (NMDS) of the Bray Curtis similarity between microbial communities in the batches revealed at least 60% similarity for each treatment and also showed that VPH class was a statistically significant factor in shaping the bacterial communities in the soil treatments (Global $R = 0.861$, $p < 0.01$). The NGS platforms (454 GS Junior and Ion torrent) compared in this study did not appear to affect the outcomes of the microbial community structure and composition analysis.

Keywords**:** Biodegradation; Next Generation Sequencing; Volatile Petroleum Hydrocarbons; Taxonomy; Community; Bioavailability

Capsule

16S amplicon sequencing revealed how soil microbial communities respond to addition of different volatile petroleum hydrocarbons.

1. Introduction

Petroleum hydrocarbons (PHs) are currently being investigated as an excellent source of carbon and energy for the growth of microbial biomass in contaminated sites (Bushnaf et al., 2017; Igun et al., 2019; Wu et al., 2017). The metabolic pathways for the degradation of PHs have been extensively studied (Van Hamme et al., 2003; Das and Chandran, 2010). PHs are reported to have varying levels of susceptibility to microbial attack and their order of susceptibility to microbial attack is as follows: cyclic alkanes < mono-aromatics < branched alkanes < linear alkanes (Ulrici, 2000). The biodegradation of *n*-alkanes depend on their solubility in soil-pore water which is proportional to the carbon chain length of the molecule (Sikkema et al., 1995). Alkanes of intermediate chain lengths ($C_5 – C_{16}$) are less water soluble than the short-chain length compounds ($C_1 – C_4$) rendering them less bioavailable for degradation and less toxic at high concentrations. Alkanes of longer chain length ($C_{25} – C_{40}$) are not easily degraded due to their hydrophobic nature and poor solubility in water and consequently low bioavailability. Mono-aromatic hydrocarbons on the other hand are comparatively more soluble in water than *n*-alkanes of intermediate chain length (http://chem.sis.nlm.nih.gov/chemidplus/, 2005). Cyclic alkanes were reported to be less preferable substrates for microbial growth in a related laboratory column study by Bushnaf (2014). At the same time, some constituents of petroleum may be persistent and toxic once released into the environment. Toxicity of PHs could be a function of their physico-chemical properties such as molecular structures e.g. complex poly-aromatic hydrocarbons (PAHs) tend to persist in the environment compared to mono-aromatic compounds due to their recalcitrant nature, or the solubility of these compounds in aqueous solutions as solubility enhances the bioavailability of organic compounds to indigenous microorganisms and hence their biodegradability. Above a threshold concentration in the environment, volatile petroleum hydrocarbons (VPHs) may exert some deleterious effects on soil microorganisms. The mechanisms by which these effects occur have been previously reviewed (Heipieper et al., 1991b; Kabelitz et al., 2003; Heipieper and Martinez, 2010).

The use of culture-independent techniques as a means to gaining improved understanding of the microbial community responses to environmental factors has gained wider acceptance and applicability in the 'ecological' community. The advancements in molecular biology techniques such as next-generation sequencing (NGS) has enabled an in-depth exploration of large amounts of sequencing data from environmental samples (Shokalla et al., 2012). Such sequencing data have been used in a variety of environmental applications including studying the effects of soil management types on microbial communities (Nacke et al., 2011) and investigating the health status of an ecosystem by analysing its biodiversity (Hajibabaei et al., 2011).

The 454-pyrosequencing platform has been widely used within the last decade to conduct metagenomics studies of environmental samples particularly due to the long read lengths of sequences and the relatively short run time of the technique (Shokalla et al., 2012) while the ion torrent sequencing platform, which is based upon the real-time detection of hydrogen ion concentration, has been developed to utilise one of three different ion chips 314, 316 or 318 each of which is capable of generating up to 10Mb, 100Mb or 1Gb of sequencing data respectively (Shokalla et al., 2012). Such technologies may serve as an invaluable tool for gaining insights into the effects of petroleum hydrocarbon contamination on soil microbiology. The aim of this study was to investigate the effects of different VPH classes (in separate mixtures) on the response of the indigenous micro-organisms in a gravelly sand using next generation sequencing techniques. Such information can inform the design and monitoring of bioremediation techniques, and environmental risk assessments. Aerobic batch experiments were set up in order to monitor the level of biodegradation occurring within different soil systems and control.

2. Materials and methods

2.1 Soil and chemical pollutant mixtures

The soil used in this study was collected from a construction site within the Newcastle University (UK) campus. Soil had an average gravimetric water content of 10% w/w with about 35% w/w in the particle size range of 0.06 -2.0mm. The chemical preparations are as follows: high purity chemicals (Sigma Aldrich, UK) were mixed into three separate classes of petroleum hydrocarbons. The aromatic hydrocarbon mixture consisted of 1 mL each of toluene, m-xylene and 1,2,4-trimethylbenzene (I,2,4-TMB). Straight chain hydrocarbon mixture consisted of 1mL each of n-octane, pentane, hexane, decane and dodecane while the cycloalkanes/branched alkane was made up of a mixture of iso-octane, cyclohexane, methylcyclohexane and methylcyclopentane (1 mL each). Each PH mixture was made in transparent glass vials and stored in the dark at room temperature (20°C).

2.2 Batch experiments

Batch microcosm experiments were performed in amber vials (65 mL, Jencons, a VWR Division, Leicestershire, UK) closed with Teflon Mininert caps (Supelco, Bellefonte, USA) containing 15 g of gravelly sand (water content: 0.1 g g^{-1} soil d.w.) and inoculated with 30 µL of either aromatic hydrocarbon mix or straight chain hydrocarbon mix or a mixture of cycloalkanes/branched alkane. Soil was thoroughly stirred in each vial using a glass stirrer to break up any lumps and to enhance aeration of soil pore spaces prior to addition of the compound mixtures. Each treatment was prepared in triplicates and an additional set of live soil controls (without petroleum hydrocarbons) was also prepared in triplicates. No VPH gases were detected (data not shown) in the flame ionization detector following injection of headspace gases from batches containing soil. This indicates that the soil used to conduct this study did not contain any background amounts of VPHs which could have introduced some bias in the study. The experiments lasted for 14 days.

2.3 Microbial respiration

Soil respiration was monitored in the batches over a 14-day period by measuring the concentration of headspace CO_2 in each vial containing either soil, or soil inoculated with aromatic hydrocarbon mixture or soil inoculated with straight chain alkane mix or soil and cyclic/branched alkanes. For each soil type, contaminated or non-contaminated, triplicate batches were monitored at room temperature (20 °C) and the analysis of headspace CO_2 was conducted using a Fisons 8060 GC linked to a Fisons MD800 MS with a HP-PLOT-Q capillary column. The instrument was calibrated daily by injecting variable volumes of a CO_2 standard (1%) to generate five-point calibration standard curves. Results for instrument calibration are presented in Figures S1-S3. Data fits of the daily standard measurements revealed that the calibration curves were linear over the duration of the experiment with correlation coefficient (r^2) values ranging from 0.88 - 0.99.

2.4 Sample collection and storage

At the end of the batch experiments, the vials were uncovered and samples were collected in triplicates and stored at -20 °C in filtered-sterile phosphate buffer saline (PBS, Oxoid) 1:1 vol/vol for DNA extraction and PCR amplification. Triplicate samples were also collected from the unamended soil batch and stored for microbial analysis.

2.5 DNA extraction and PCR amplification (454 pyrosequencing library preparation)

Genomic DNA was extracted from 50 mg of soil (wet weight) using the FastDNA Spin kit (MP Biomedicals, UK) according to the manufacturer's instructions. The V4 and most of the V5 regions of 16S rRNA gene was PCR amplified by multiplex PCR reactions (averagely 15 reactions per sample), using primer set 515f (5'-GTGNCAGCMGCCGCGGTAA-3') and 926r (5'-CCGYCAATTYMTTTRAGTTT-3')(Wilhelm et al., 2013), with read length of 400 – 500 base pairs (bp). A unique 8 bp barcode, added to the 5' –end of both the forward and reverse primers through a GA linker, was used to label each sample. The primers were attached to the GS FLX Titanium Adapter A (5' - CGTATCGCCTCCCTCGCGCCATCAG – 3') and adapter B (5' – CTATGCGCCTTGCCAGCCCGCTCAG – 3'). Each PCR reaction was performed in a total volume of 25 µL containing 0.5 µL of DNA template, 0.4 µmol L^{-1} of each universal primer, 0.2mmol L^{-1} dNTPs (PCR grade Nucleotide Mix, Roche), Fast Start High Fidelity Enzyme Blend (2.5U/reaction), and a final concentration of 1.8 mM $MgCl_2$ in the FastStart High Fidelity Reaction Buffer (Roche Diagnostics GmbH, Mannheim, Germany).The following PCR thermal cycling programme was used: an initial denaturation step of 95 °C for 4 minutes followed by 25 cycles of denaturation at 95 °C for 1 minute, annealing of primers at 55 °C for 45 seconds, and elongation at 72 °C for 1 minute. The final elongation step was at 72 °C for 7 minutes. Multiplex PCR amplicons were pooled together and cleaned using the QIAquick PCR purification kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. Prior to 454 sequencing of PCR amplicons, the amount of DNA present in clean PCR products were quantified using a Qubit® 2.0 Fluorometer following the manufacturer's protocol.

2.6 454 pyrosequencing and ion torrent sequencing

Clean PCR amplicons were pooled together (in triplicates) in equimolar concentrations and sequenced on a Roche 454 GS Junior (Macropathology Ltd., Coventry, UK). Sequencing was carried out in a bi-directional manner using unique 8 base-pairs barcoded 515f and 926r primers for both forward and reverse runs. The output data from the sequencing runs in the standard flowgram format (SFF) was filtered for quality and subsequently denoised on a 16 core-computer cluster.

For the ion torrent sequencing, in addition to the sample preparation for PCR, the samples were labelled using a unique 12 base pairs Golay barcode, added to the 5' – end of the forward primers through a GAT spacer, and attached to the Ion Adapter A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') while the reverse primers were attached to the Ion adapter trP1 (5'- CCTCTCTATGGGCAGTCGGTGAT-3'). PCR amplicons were cleaned and size selected using double-sided solid-phase reversible immobilisation (SPRI) beads (AgencourtAMPure XP system, Beckman Coulter). Clean PCR amplicons from samples were also pooled in triplicates in equimolar concentrations and sequenced on a Personal Genome Machine (PGM) using a 316 ion chip available at the School Engineering, Environmental Engineering group, Newcastle University.

2.7 Microbial community structure analysis

Briefly, the reads were filtered for quality (filtering criteria: perfect match to sequence barcode/primer, 200bp minimum sequencing length, minimum quality score = 30). In the pyrosequencing-derived data, the number of filtered reads varied from 772 reads in a soil sample to 8684 in a soil and cyclic/branched alkane sample (Table S1). Whereas for the ion torrent-derived data, the number of sequences obtained per sample after filtering for quality ranged from 22,119 reads in a soil and straight alkane contaminated sample to 31,792 reads in an uncontaminated soil sample and was more evenly distributed compared to the pyrosequencing-derived data (Table S2). The discrepancies in the results particularly in the pyrosequencing dataset could be as a result of variability arising from sequencing efforts of the GS Junior platform. Fewer sequences (<100,000) are generated on this platform containing a relatively moderate level of noise. On the contrary, with the ion torrent sequencing platform, a much higher number of sequences generated (>1,000,000) with the least amount of sequencing noise would imply that a higher number of sequences will pass the filtering step. The purpose of using both platforms in parallel in this study was to shed light on potential implications of such technological differences on the scientific insights obtained.

QIIME Denoiser (Reeder and Knight, 2010) was used to detect and correct sequencing errors for the pyrosequencing-derived data and the data were reintegrated into the QIIME pipeline by inflation. The so obtained sequences were clustered into Operational Taxonomic Unit (OTU) at 97% sequence similarity level by the uclust algorithm, a representative sequence from each OTU was selected and taxonomically identified using the Greengenes database (McDonald et al., 2012; Werner et al., 2012). Representative sequences and correspondent taxonomic assignment were used to build a table of OTU abundances at different levels of taxonomy. The QIIME (v.1.8.0) pipeline (Caporaso et al., 2010) was used to determine the microbial community diversity within each sample and across the 12 samples. The resulting OTU table at the class level (L3) of taxonomy was imported into PRIMER v6 and square root transformed for subsequent beta (β) diversity analysis. The Bray Curtis dissimilarity metric was calculated for L3 OTU table and an average pairwise distance and standard deviation was determined for each pair of samples (Clarke Robert et al., 2006). The resulting Bray Curtis distance matrix was mapped unto a 2 dimensional non-metric multidimensional scaling plot (nMDS) using PRIMER v6.

For alpha diversity analysis in the pyrosequencing-derived dataset, an in silico rarefaction was performed using an OTU table generated in QIIME at a minimum rarefaction depth of 100 in a series of depth and a step-wise increase of 200 sequences, a total number of 10 replicates (multiple rarefactions) at each depth and a maximum rarefaction depth of 8000 in the series of depth. For the ion torrent data analysis, rarefaction was performed at a minimum depth of 4000 sequences in a series of depths, a step-wise increase of 2000 sequences and a maximum rarefaction depth of 21,000 in the series of depth. For the diversity within each sample (alpha diversity), the non-parametric species richness estimator Chao1 and the Faith's phylogenetic diversity (PD) index were determined according to the methods described by Chao (1984) and Faith (1992) respectively in QIIME. The Shannon's diversity index (H') was also determined for each sample as a measure of alpha diversity based on derivations made by Shannon and Weaver (1949). This index of diversity (Shannon and Weaver) is based on the assumption that biological systems contain information that can be analysed in a similar manner to some coded information and that individual species, if sampled randomly from a large community of species are well represented in the samples (Pielou, 1975; Magurran, 2004). Shannon's index of diversity is given by the equation below:

$$H' = -\sum p_i \ln p_i \qquad (1.0)$$

where pi is the proportion of individuals present within the *i*th species and is estimated based on a maximum likelihood estimator *ni/N* according to Pielou (1969).

2.8 Statistical analysis

A 2-Way Analysis of Variance (ANOVA) was performed using Minitab-17 statistical software (Minitab Ltd., Coventry, UK) on the alpha diversity indices in order to compare the interactive effects of the factors under consideration (VPH classes) on species richness and microbial diversity. Analysis of Similarities (ANOSIM) was conducted on the Pearson' product-moment correlation dissimilarity matrix using PRIMER v6. To compare microbial community structure across all samples based on the relative abundance of OTUs, the dominant genera (relative abundance > 1.0% of the total number of sequences) were square root transformed and a Principal Components Analysis (PCA) performed on the transformed data using XLSTAT for Windows (XLSTAT, 2014). A comparison of the relative abundances of identified bacterial groups between soil treatments (VPH classes) was conducted using Microsoft Excel v2010 (Microsoft, Redmond, USA) for significant effects ($p < 0.05$).

3. Results and Discussion

3.1 Biodegradation of volatile petroleum hydrocarbon mixtures in batch systems

CO_2 production profiles for the various petroleum hydrocarbon classes are shown in Figure 1. In the control soil batch, an initial lag phase of approximately 6 days was observed. This was followed by a slow rise in the headspace concentration of CO_2 from day 7 with some variability between individual time points, but a clear overall trend. The maximum cumulative CO_2 concentration for the control batch was 0.003 g/L which, is a factor 4 times higher than the initial CO_2 concentration. A similar pattern was observed for the aromatic hydrocarbon contaminated soil batch in which headspace CO_2 rose slowly to reach a comparable maximum concentration of 0.003 g/L by day 15. This indicates limited respiration on aromatic hydrocarbons in the soil investigated, perhaps due to toxicity limitations. In the cyclic and branched alkane contaminated soil batch, headspace CO_2 concentration rose above those measured in the soil and the aromatic hydrocarbon contaminated batches indicating a more notable contribution of cyclic and branched alkane respiration to the overall soil respiration, as compared to the aromatic hydrocarbon amended soil batches. In the straight-chain alkane contaminated soil batch, a much more significant rise in headspace CO_2 concentration was observed in the batches (Fig. 1) suggesting that soil microbial activity was most stimulated by the addition of the straight chain alkane class of petroleum hydrocarbons in contrast to the other PH classes. CO_2 production in the straight alkane soil batch rose by up to two orders of magnitude to reach a maximum cumulative concentration of 0.112 g/L before entering into a stationary phase from around day 13 (Fig. 1). Straight-chain alkanes of intermediate chain length such as those used in the current batch study may not have been inhibitory to the growth of the microorganisms since they are relatively less soluble in water and consequently less bioavailable to the soil microbes. More so, they are more readily biodegradable compared to other classes of PHs (Ulrici, 2000; Bushnaf et al., 2017), especially cyclic and branched alkanes. For the aromatic hydrocarbons, their higher water solubility at equivalent mass addition, as compared to straight-chain alkanes, is expected to induce a more toxic effect due to the higher bioavailability which may inhibit the aromatic hydrocarbon degradation.

3.2 Microbial diversity and species richness

Microbial diversity and species richness across the different samples as determined by the non-parametric richness estimator (Chao, 1984) and the Faith's Phylogenetic diversity (Faith, 1992) are illustrated in Fig. 2. For the pyrosequencing dataset (Fig. 2a), at an even rarefaction depth of 500 sequences, microbial diversity as determined by the Faith's

phylogenetic diversity (PD) was the lowest in the straight-alkane contaminated soil (31.4±0.6). The control soil had a PD value of 43.3±1.3 followed by the soil & aromatic hydrocarbons (aro-Hs) batch (40.0±1.4) and the soil & cyclic/branched alkane (cyc-alk) batch (35.8±0.6). The species richness estimator (Chao1) was clearly higher in the control soil than in two of the contaminated soils; straight-alkanes (str-alk) and cyc-alk but also slightly higher than the soil contaminated with aromatic hydrocarbons (Fig. 2). Chao1 values for soil, soil & aromatic Hs, soil & str-alkanes, and soil & cyc-alk batches were 797.5±80.1, 733.2±42.0, 580.9±25.3 and 684.5±25.7 respectively. There was a statistically significant difference ($p < 0.05$) between different petroleum hydrocarbon classes for the species richness as indicated by a one-way analysis of variance (ANOVA). The observed number of species at a 97% sequence similarity level were statistically significantly different between the different classes of PH (ANOVA, $p < 0.05$). More specifically, the control soil samples had the highest observed number of species of 320.5±5.5 followed by the aromatic hydrocarbon contaminated soils with an average observed number of species of 289.1±4.7. The straight-chain alkane and cyclic or branched hydrocarbon contaminated soils had an average observed spp. of 229.7±2.0 and 258.6±2.5 respectively. This suggests that contamination of soil with an equivalent dose of different petroleum hydrocarbon classes impacted soil microbial diversity negatively by significantly reducing bacterial species richness with the straight-chain alkane contaminated soil being the most affected (Chao1 and Observe spp.) (Fig. 2). More likely, introducing carbon substrates as VPHs into the soil might have caused a few organisms well adapted to the compounds to increase in abundance thereby 'overshadowing' the rare species in the sequencing results. This could explain the seeming decrease in species diversity of the bacterial community.

Rarefaction plots of the observed number of species in the pyrosequencing data and the ion torrent data are shown in Fig. S4. The results from the current study show that by sampling a higher number of sequences, a higher species richness estimate is obtained in both datasets (454 pyrosequencing and Ion torrent datasets). In the pyrosequencing data, the number of OTUs increased with an increase in the sequencing depth but never attained saturation indicating that the sequences were not exhaustively sampled (Fig. S4a). Similarly, in the ion torrent data (Fig. S4b), the rarefaction curve never reached saturation although they were less steep compared to the rarefaction curve from the pyrosequencing data. Several studies on samples from the environment have highlighted the effects of sequencing efforts on the species richness and microbial diversity of such samples (Roesch et al., 2007; Nacke et al., 2011). Findings from these previous studies correspond with the results from the current study in which the number of sequences generated per sample influences the species richness and diversity indices derived from the data, which can show some variation from one sample to the other, even within the same replicates.

3.3 Taxa distribution across samples

The dominant bacteria phyla representing > 3.0 % of the overall (total) relative abundances in the pyrosequencing-derived data are shown in Fig 3a. They are the *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes*, and the *Proteobacteria* classes *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria* representing 4.6, 14.0, 7.2, 10.4, 3.9, 8.5, 13.4, 6.2, 5.9, and 15.3% respectively across the different samples under investigation.

Further analysis of the phyla level taxa revealed a variation in the relative abundances of the different soil treatments. For instance, the *Acidobacteria* had a higher relative abundance in the uncontaminated soils than in all the treated soil samples ($p < 0.05$, 2 sample t-tests). We are not aware of reports that would indicate that members of the phylum *Acidobacteria* have significant potential to degrade VPHs, which may explain their reduced relative abundance in

the VPH contaminated soil as compared to the control soil. *Actinobacteria* were significantly more abundant in the straight-alkane contaminated soils ($p < 0.05$, 2 sample t-tests) than they were in the uncontaminated soils but were comparable in the other treated soils. Studies by Pidgorskyi and Nogina (2016) demonstrated the potentials of members of *Actinobacteria* to degrade n-alkanes of chain lengths C9-C21 similar to those used in this study. *Firmicutes* were significantly higher in relative abundance in the aromatic hydrocarbon contaminated soils than in the uncontaminated and straight-alkane and cyclic-alkane contaminated soil samples ($p < 0.05$, 2 sample t-tests). Members of this phylum were previously implicated in the degradation of *o*-xylene (Morasch et al., 2004) and benzene (Li et al., 2006) from contaminated sites and aquifers. *Gemmatimonadetes* showed the opposite pattern in which uncontaminated soils had a significantly higher relative abundance compared to the straight-chain alkane and cyclic hydrocarbon contaminated soils but not the aromatic hydrocarbon contaminated soil samples ($p < 0.05$, 2 sample t-tests). Members of this phylum have not been previously implicated in the degradation of straight-chain alkanes and cycloalkanes which may explain the decrease in their relative abundances following exposure of soil to these classes of VPH. The *Alphaproteobacteria* class did not show any significant variation with respect to their relative abundance across the control and treated soil samples. In contrast, the *Gammaproteobacteria* were significantly more dominant in the straight-alkane and cyclic/branched alkane contaminated soils than in the control soils and the aromatic hydrocarbon contaminated soil samples ($p < 0.05$, 2 sample t-tests). Members of the phylum *Alphaproteobacteria* and *Gammaproteobacteria* were reported from previous studies to dominate bacteria communities in petroleum hydrocarbon contaminated beach sands and marine sediments (Head et al., 2006; Yakimov et al., 2007; Kostka et al., 2011).

Similarly, the ion torrent-derived phyla level taxa summary for the dominant phyla are displayed in Figure 3b. The taxa distributions reveal an identical pattern to those obtained from the pyrosequencing-derived data as the dominant phyla were retained and did not show any variation. From a qualitative point of view, the bacteria community composition derived did not change between both next-generation sequencing platforms, and within each treatment at least at the phylum level (Fig. 3) highlighting the reproducibility of results from both next-generation sequencing platforms. From a quantitative viewpoint, however, there were some variations between soil treatments when compared to the taxa distribution from the pyrosequencing-derived data. The *Alphaproteobacteria* were significantly more dominant in the aro-H contaminated soils than the uncontaminated soil but were more dominant in the uncontaminated soil than in the str-alk contaminated soil samples (for both sequencing platforms; $p < 0.05$, 2 sample t-tests). The *Gammaproteobacteria* revealed a similar pattern in the ion-torrent generated data where the phylum significantly dominated in the str-alk and the cyc-alk contaminated soils compared to the aro-H contaminated and the control soil samples.

3.4 OTU functions in petroleum hydrocarbon degradation

At the order level, most of the sequences associated with the dominant phyla *Actinobacteria*, *Firmicutes*, *Alphaproteobacteria* and *Gammaproteobacteria* across the control and treated soil samples are displayed in Table 1. The *Actinomycetales* showed approximately a 5-fold increase in the straight-chain alkane contaminated soils compared to the control while the *Bacillales* were twice as high in the aromatic hydrocarbon treated soils than they were in the control samples (Table 1). *Pseudomonadales* increased in percentage abundance by a factor of approximately 33 and 44 in the straight-chain alkane and cyclic alkane contaminated soils respectively in comparison with the untreated soils while the *Xanthomonadales* increased by a factor of about 2 in the str-alk contaminated soils compared to the control soils. Most of the sequence affiliated to the order *Actinomycetales* at the family level belonged to the

Micrococcaceae, Micromonosporaceae, Mycobacteriaceae, Nocardiaceae, Nocardioidaceae, Pseudonocardiaceae, Streptomycetaceae and *Streptosporangiaceae*.

At the genus level, there was a significant difference with respect to the relative abundances of the OTUs between the controls and the treated soil samples (Table 2). The genus *Rhodococcus* and *Methylibium* had the highest relative abundances in the str-alk treated soils ($p < 0.05$, 2 sample t-Tests) in comparison with the untreated soils. *Rhodococcus* has been reported to possess a remarkable range of diverse catabolic genes plus a resilient physiology which explains its ability to adapt to a wide range of environmental conditions (Larkin et al., 2010). *Pseudonocardia* was more dominant in the aromatic hydrocarbon treated soil compared to the untreated soil ($p < 0.05$, 2 sample t-Tests) but decreased in abundance in the straight alkane treated soil (Table 2). Members of the genus *Pseudonocardia* were associated with the degradation of the aromatic hydrocarbons toluene and benzene in a compost biofilter study by Juteau et al. (1999). The genus *Desulfosporosinus* increased significantly ($p < 0.05$) in relative abundance in the straight-alkane treated soil but showed a decrease in abundance in the cyclic and branched alkane treated soils (Table 2) in comparison with untreated soil batch. *Polaromonas* was more dominant in the straight-alkane treated soil compared to the untreated soil ($p < 0.05$). The genus *Pseudomonas* benefitted significantly from the addition of cyclic/branched alkane and straight-chain alkanes when compared to the uncontaminated soil. The genus comprises of a metabolically versatile category of microorganisms that can live aerobically or anaerobically and have the potential to metabolise a wide range of organic compounds ranging from alkanes (van Beilen et al., 1994; Mukherjee et al., 2010) to aromatic hydrocarbons – toluene (Assinder and Williams, 1990; Mukherjee et al., 2010), and benzene (Mukherjee et al., 2010) which have been utilized as sole carbon sources under aerobic conditions. Members of this genus are also producers of the biosurfactant, rhamnolipids, which enhance the bioavailability of hydrophobic organic compounds in contaminated environments (Perfumo et al., 2006). The genus *Mesorhizobium* was statistically significantly higher in relative abundance in the aro-H treated soil than they were in the control soil ($p < 0.05$). Members of this genus demonstrated capacity to utilize BTEX in one study (Auffret et al., 2015). *Azomonas* was more dominant in the straight alkane contaminated soil while *Lycinibacillus* was more significantly dominant in the cyc-alk treated soils than it was in the untreated soils (controls, $p < 0.05$) (Table 2). Members of the genus *Lycinibacillus* have been isolated from a mixture of petroleum hydrocarbons – n-hexane, toluene, n-decane, and xylene isomers although reported to be less tolerant to PHs compared to other genera (Stancu and Grifoll, 2011). Significant increases in relative abundance relative to the control soil were thus observed for genera known to contain putative petroleum hydrocarbon degrading bacteria (Table 2).

3.5 Inter-relationship between bacterial communities within different soil samples

A mapping of the Bray-Curtis similarity matrix of the operational taxonomic units (OTUs) at the class level (Square root transformed) is shown in the Figure 4. Replicate samples from the same soil treatment type clustered more closely to each other than they did to replicates of samples from other treatments with the exception of a straight-chain alkane treated soil sample which clustered more closely with the cyclic/branched alkane treated soil samples (Fig. 4). The pattern of clustering reveals that petroleum hydrocarbon class was an important factor in shaping the microbial communities within the soil samples. Microbial communities from all samples under investigation clustered at a percentage similarity of 60% while the communities within each replicate clustered at a similarity of 80% indicating a significant level of similarity and reproducibility of results. An analysis of similarities (ANOSIM) was performed between the microbial communities present within each sample based on the Bray-Curtis similarity metric in order to investigate the significance of the factor under

consideration. The results showed that the petroleum hydrocarbon class was a statistically significant factor in shaping the microbial communities within different soil treatments (Global $R = 0.861$, $p < 0.01$).

Similarly, a clustering of samples based on the most dominant OTUs (of the pyrosequencing dataset) at the genus level using a heatmap shows that the samples clustered according to petroleum hydrocarbon classes again highlighting the strong effects of PH classes in shaping microbial community structures in contaminated and uncontaminated soils (Fig. 5). The PH degrading genera *Polaromonas*, *Rhodococcus*, *Desulfosporosinus*, *Lysobacter*, *Pseudomonas* and *Clostridium* were more dominant in the straight-chain alkane contaminated soil (indicated by the green rectangles) but showed the reverse trend in other soil treatments including the control samples. Conversely, the genera *Bacillus*, *Nitrospira*, *Rhodoplanes*, *Nitrosopumilus*, *Planctomyces*, *Hyphomicrobium* and *Candidatus Nitrosospheara* most of which have not been previously associated with petroleum hydrocarbon degradation were more dominant in the aromatic hydrocarbon contaminated soils but less abundant in the straight chain alkane contaminated soil (indicated by red rectangles).

Observations and variables plots of a principal component analysis (PCA) are shown for the dominant bacteria genera (>1% of total relative abundance) for the 454–pyrosequencing (Fig. 6a) and the ion torrent (Fig. 6b) derived 16S rDNA sequence libraries. A close look at the plots for the pyrosequencing derived data shows a clear pattern of demarcation of the soil samples based on the PH class into different quadrants. The 1st principal component (PC) accounted for 67.2% of the variation of the dataset while the 2nd PC accounted for 21.5% variation of the total data (Fig 6a). On the 1st PC, the cyc-alk and Str-alk treated soils were partitioned into the positive axis corresponding to the genera *Pseudomonas*, *Rhodococcus*, *Polaromonas* and *Lysobacter* in the observations plot (Fig. 6a). The untreated (control) soil and aro-H treated soils, on the other hand partitioned into the negative axis of the 1st PC corresponding to the genera *Pedomicrobium*, *Rhodoplanes*, and *Candidatus nitrososphaera*. The percentage contribution of individual genus with respect to variation of the data on the 1st PC is as follows: an unidentified genus had the highest contribution of 31.9% followed by the genus *Pseudomonas* with a contribution of 24.2% (Table S3). The genera *Rhodococcus* and *Polaromonas* contributed 9.3% and 12.9% respectively to the variation of dataset on the 1st PC (Table S3). More specifically, the increase in the relative abundance of the genus *Pseudomonas* was in the direction of cyc-alk contaminated soils while *Rhodococcus*, *Polaromonas* and *Lysobacter* increased in their relative abundances in the direction of the straight-chain alkane contaminated soils as indicated by the vector arrow heads (Fig. 6a). Results of a principal component analysis (PCA) on the ion torrent-derived data (Fig. 6b) compared closely with those of the 454 –pyrosequencing data. The major OTUs (accounting for > 1% of total relative abundance) and contributing to the demarcation patterns of the VPH classes (Fig. 6) were mostly the same for both sequencing platforms suggesting that the choice of next generation sequencing platforms did not significantly alter the outcomes of the microbial community structure analyses.

4. Conclusions

Addition of VPH compound classes stimulated soil microbial activity with the straight-chain alkane contaminated soil producing the most stimulatory effect on microbial degradation followed by the cyclic and branched alkane contaminated soil batch. Aromatic hydrocarbon contaminated batch biodegradation was comparable with the uncontaminated soil batch. The straight-chain alkanes were more readily biodegradable in line with findings from previous studies. Contrary to our proposed hypothesis, bacterial species richness and diversity decreased following the exposure of soil to the contaminants. Thus, addition of the different classes of VPHs resulted in a less even community structure and seemingly less diversity because rare species may be lost from the library due to an increase of genes from PH degrading bacteria. Soil contaminated with a mixture of VPH compounds at low concentrations would therefore benefit from the growth of specialised bacterial species involved in VPH biodegradation. Bacterial communities shifted strongly depending on the VPH class e.g. *Polaromonas* and *Rhodococcus* contributed more to biodegradation in the straight-alkane batch while *Pseudomonas* contributed more to biodegradation in the cyclic and branched alkane contaminated soil batch. The implications of these findings provide insights which may enhance the design of remediation systems and also be of broad interest to researchers studying the microbial ecology of petroleum biodegradation in soil.

Conflicts of interest

The authors declare no conflicts of interest in this article

Acknowledgements

This project was funded by an Overseas Scholarship from the Petroleum Development Technology Fund (PTDF), Nigeria.

References

Assinder, S.J., Williams, P.A.,1990. 'The TOL plasmids: determinants of the catabolism of toluene and the xylenes. ', Advanced Microbial Physiology, 31, pp. 1–69. https://doi.org/10.1016/S0065-2911(08)60119-8

Auffret, M.D., Yergeau, E., Labbe, D., Fayolle-Guichard, F., Greer, C.W., 2015. 'Importance of Rhodococcus strains in a bacterial consortium degrading a mixture of H, gasoline, and diesel oil additives revealed by meta-transcriptomic analysis', Applied Microbiology and Biotechnology, 99, pp. 2419-2430. https://doi.org/10.1007/s00253-014-6159-8

Bushnaf, K.M., Puricelli, S., Saponaro, S., Werner, D., 2011. 'Effect of biochar on the fate of volatile petroleum hydrocarbons in an aerobic sandy soil.', Contaminant Hydrology, 126(3-4), pp. 208-215. https://doi.org/10.1016/j.jconhyd.2011.08.008.

Bushnaf, K.M., Mangse G., Meynet P., Russell J.D., Cirpka O.A., Werner, D., 2017. 'Mechanisms of distinct activated carbon and biochar amendment effects on petroleum vapour biofiltration in soil'. Environmental Science: Processes & Impact, 19, pp. 1260-1269. https://doi.org/ 10.1039/c7em00309a

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Gonzalez, P.A., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. 'QIIME allows analysis of high-throughput community sequencing data', Nature Methods, 7(5), pp. 335-336. https://doi.org/10.1038/nmeth.f.303

Chao, A., 1984. 'Nonparametric Estimation of the Number of Classes in a Population', Scandinavian Journal of Statistics, 11, pp. 265-270. https://www.jstor.org/stable/46159641

Chen, Y.P., Lopez-de-Victoria, G., Lovell, C.R., 1993. 'Utilization of aromatic compounds as carbon and energy sources during growth and N2-fixing bacteria. ', Archives in Microbiology, 159, pp. 207-212. https://doi.org/10.1007/BF00248473.

Clarke Robert, K., Somerfield, P.J., Chapman, M.G., 2006. 'On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages', Experimental Marine Biology and Ecology, 330, pp. 55-79. https://doi.org/10.1016/j.jembe.2005.12.017

Das, N., Chandran, P., 2010. 'Microbial degradation of petroleum hydrocarbon contaminants', Biotechnology Research International, 2011, pp. 1-13. https://doi.org/10.4061/2011/941810

Faith, P.D., 1992. 'Conservation evaluation and phylogenetic diversity', Biological Conservation, 61, pp. 1-10. https://doi.org/10.1016/0006-3207(92)91201-3

Garten Jr, C.T., Cooper, L.W., Post III, W.M., Hanson, P.J., 2000. 'Climate controls on forest soil C isotope ratios in the southern Appalachian Mountains', Ecology, 81, pp. 1108-1119.

Hajibabaei, M., Shokalla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. 'Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos', PLoS One, 6(4), pp. 1-7. https://doi.org/10.1371/journal.pone.0017497

Head, I.M., Jones, D.M., Roling, W.F.M., 2006. 'Marine organisms make a meal of oil', Nature Reviews Microbiology, 4, pp. 173-182. https://doi.org/10.1038/nrmicro1348.

Heipieper, H.J., Keweloh, H., Rehm, H.J., 1991a. 'Influence of phenols on growth and membrane permeability of free and immobilized Escherichia coli', Applied and Environmental Microbiology, 57, p. 5. https://doi.org/10.1128/AEM.57.4.1213-1217.1991

Heipieper, H.J., Keweloh, H., Rehm, H.J., 1991b. 'Influence of phenols on growth and membrane permeability of free and immobilized Escherichia coli', Applied and Environmental Microbiology, 57, pp. 1213-1217. https://doi.org/10.1128/AEM.57.4.1213-1217.1991

Heipieper, H.J., Martinez, P.M., 2010. 'Toxicity of hydrocarbons to microorganisms', in Timmis, K.N. (ed.) Handbook of hydrocarbon and lipid microbiology. Heidelberg, Berlin: Springer-Verlag, pp. 1566-1570. https://doi.org/10.1007/978-3-540-77587-4_108

http://chem.sis.nlm.nih.gov/chemidplus/ 2005. Toxicological Data Network (Accessed: 21/07/2015). https://doi.org/10.1890/0012-9658(2000)081[1108:CCOFSC]2.0.CO_2

Igun O.T., Meynet P., Davenport J.D., Werner D., 2019. 'Impacts of activated carbon amendments, added from the start or after five months, on the microbiology and outcomes of crude oil bioremediation in soil. International Bioterioration & Biodegradation', 142, pp. 1-10 https://doi.org/10.1016/j.ibiod.2019.04.008

Jimenez, N., Vinas, M., Guiu-Aragones, C., M., B., Albaiges, J., Solanas, A.M., 2011. 'Polyphasic approach for assessing changes in an autochthonous marine bacterial community in the presence of Prestige fuel oil and its biodegradation potential', Applied Microbiology and Biotechnology, 91, pp. 823-834. https://doi.org/10.1007/s00253-011-3321-4.

Juteau, P., Rho, D., Larocque, R., and LeDuy, A., 1999. 'Analysis of the relative abundance of different types of bacteria capable of toluene degradation in a compost biofilter. Applied Microbiology & Biotechnology', **52,** 863–868 (1999). https://doi.org/10.1007/s002530051604

Kabelitz, N., Santos, P.M., Heipieper, H.J., 2003. 'Effect of aliphatic alcohols on growth and degree of saturation of membrane lipids in Acinetobacter calcoaceticus', FEMS Microbiology Ecology, 220(2), pp. 223-227. https://doi.org/10.1016/S0378-1097(03)00103-4

Kostka, J.E., Prakash, O., Overholt, W.A., Green, S.J., Freyer, G., Canion, A., Delgardio, J., Norton, N., Hazen, T.C., Huettel, M., 2011. 'Hydrocarbon-degrading bacteria and the bacterail community response in Gulf of Mexico beach sands impacted by the Deepwater Horizon oil spill', Applied Environmental Microbiology, 77(22), pp. 7962-7974. https://doi.org/10.1128/AEM.05402-11.

Larkin, M.J., Kulakov, L.A., Allen, C.C.R., 2010. 'Rhodococcus', in (ed), K.N.T. (ed.) Handbook of hydrocarbon and lipid microbiology. Berlin, Heidelberg: Springer-Verlag, pp. 1839-1852. https://doi.org/10.1007/978-3-540-77587-4_129.

Li, H., Liu, Y.H., Luo, N., Zhang, X.Y., Luan, T.G., Hu, J.M., Wang, Z.Y., Wu, P.C., Chen, M.J., Lu, J.Q., 2006. 'Biodegradation of benzene and its derivatives by a psychrotolerant and moderately haloalkaliphilic Planococcus sp. strain ZD22', Research in Microbiology, 157, pp. 629-636. https://doi.org/10.1016/j.resmic.2006.01.002.

Liu, A., Garcia-Dominguez, E., Rhine, E.D., Young, L.Y., 2004. 'A novel arsenate respiring isolate that can utilize aromatic substrates', FEMS Microbiology Ecology, 48, pp. 323-332. https://doi.org/10.1016/j.femsec.2004.02.008.

Malhautier, L., Quijano b, G., Avezac, M., Rocher, J. and Fanlo, J.L. (2014) 'Kinetic characterization of toluene biodegradation by Rhodococcus erythropolis: Towards a rationale for microflora enhancement in bioreactors devoted to air treatment', Chemical Engineering Journal, 247, pp. 199-204. https://doi.org/10.1016/j.cej.2014.02.099 .

Mattes, T.E., Alexander, A.K., Richardson, P.M., Munk, A.C., Han, C.S., Stothard, P., Coleman, N.V., 2008. 'The Genome of Polaromonas sp. Strain JS666: Insights into the Evolution of a Hydrocarbon- and Xenobiotic-Degrading Bacterium, and Features of Relevance to Biotechnology', Applied and Environmental Microbiology, 74(20), pp. 6405-6416. https://doi.org/10.1128/Aem.00197-08.

McDonald, D., Price, M.N., Goodrich, J.K., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. 'An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea', ISME, 6(3), pp. 610-618. https://doi.org/10.1038/ismej.2011.139.

Morasch, B., Schink, B., Tebbe, C.C., Meckenstock, R.U., 2004. 'Degradation of o-xylene and m-xylene by a novel sulfate-reducer belonging to the genus Desulfotomaculum. ', Archives of Microbiology, 181, pp. 407-417. https://doi.org/10.1007/s00203-004-0672-6

Mukherjee, S., Bardolui, N.K., Karim, S., Patnaik, V.V., Nandy, R.K., Bag, P.K., 2010. 'Isolation and characterization of a mono-aromatic H-degrading bacterium, P. aeruginosa from crude oil', Journal of Environmental Science and Health Part A 45, pp. 1048–1053. https://doi.org/10.1080/10934529.2010.486328

Nacke, H., Thurmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., Schoning, I., Schrumpf, M., Daniel, R., 2011. 'Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils', PLoS One, 6(2), pp. 112. https://doi.org/10.1371/journal.pone.0017000.

Nakatsu, C.H., Hristova, K., Hanada, S., Meng, X., Hanson, J.R., Scow, K.M. and Kamagata, Y. (2006) 'Methylibium petroleiphilum gen. nov., sp. nov., a novel methyl tert-butyl ether-degrading methylotroph of the Betaproteobacteria', International Journal of Systematic and Evolutionary Microbiology, 56, pp. 983-989. https://doi.org/10.1099/ijs.0.63524-0.

Palleroni, N.J., Pieper, D.H., Moore, E.R.B., 2010. 'Microbiology of Hydrocarbon-Degrading Pseudomonas', in Timmis, K.N. (ed.) Handbook of Hydrocarbon and Lipid Microbiology. Berlin, Heidelberg: Springer-Verlag, pp. 1787-1798. https://doi.org/10.1007/978-3-540-77587-4_129

Perfumo, A., Banat, I.M., Canganella, F., Merchant, R., 2006. 'Rhamnolipid production by a novel thermophilic H-degrading P. aeruginosa', Applied Microbial and Cell Physiology, 72, pp. 132-138. https://doi.org/10.1007/s00253-005-0234-0.

Pidgorskyi, V.S., and Nogina, T.M., 2016. Biodegradation of petroleum hydrocarbons by Actinobacteria and Acinetobacter Strains producing biosurfactants. Mikrobiol Z, 78(6): 92 – 103. https:// doi: 10.15407/microbiolj78.06.092

Pielou, E.C., 1969. An introduction to mathematical ecology. New York: Wiley. http://dx.doi.org/10.1126/science.169.3940.43-a

Pielou, E.C., 1975. Ecological Diversity. New York: Wiley InterScience. https://doi.org/10.4319/lo.1977.22.1.0174b.

Prince, R.C., Gramain, A., McGenity, T.J., 2010. 'Prokaryotic hydrocarbon degraders', in Timmis, K.N. (ed.) Handbook of hydrocarbon and lipid microbiology. Berlin, Heidelberg: Springer-Verlag, pp. 1669-1692. https://doi.org/10.1007/978-3-540-77587-4.

Reeder, J., Knight, R., 2010. 'Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions', Nature Methods, 7, pp. 668-669. https://doi.org/10.1038/nmeth0910-668b.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., A.K.M., H., 2007. 'Pyrosequencing enumerates and contrasts soil microbial diversity', ISME, 1, pp. 283-289. https://doi.org/10.1038/ismej.2007.53.

Shannon, C.E., Weaver, W., 1949. The mathematical theory of communication. Urbana, IL: University of Illinois Press. https://doi.org/10.1145/584091.584093 .

Shokalla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. 'Next-generation sequencing technologies for environmental DNA research', Molecular Ecology, 21, pp. 1794-1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x.

Sikkema, J., De Bont, J.A.M., Poolman, B., 1995. 'Mechanisms of membrane toxicity of

hydrocarbons', Microbiological Reviews, 59(2), pp. 201-222. https://doi.org/10.1128/mmbr.59.2.201-222.1995.

Stancu, M.M., Grifoll, M., 2011. 'Multidrug resistance in hydrocarbon-tolerant Gram-positive and Gram-negative bacteria', Journal of General Applied Microbiology., 57, pp. 1-18. https://doi.org/10.2323/jgam.57.1.

Ulrici, W., 2000. 'Contaminant soil areas, different countries and monitoring of contaminants ', in Rehm, H.J. and Reed, G. (eds.) Environmental Process II. Soil decontamination biotechnology. pp. 5-42. https://doi.org/10.1002/9783527620951.ch1

Van Hamme, J.D., Singh, A., Ward, O.P., 2003. 'Recent advances in petroleum microbiology', Microbiological and Molecular Biology Reviews, 67(4), pp. 503-549. https://doi.org/10.1128/mmbr.67.4.503-549.2003

Van Beilen, J.B., Wubbolts, M.G., Witholt, B., 1994. 'Genetics of alkane oxidation by Pseudomonas oleovorans', Biodegredation, 5, pp. 161-174. https://doi.org/10.1007/BF00696457

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., Ley, R.E., 2012. 'Impact of training sets on classification of highthroughput bacterial 16S rRNA gene surveys', ISME, 6, pp. 94-102. https://doi.org/10.1038/ismej.2011.82.

Wilhelm, L., Singer, G.A., Christina, F., Battin, T.J., Besemer, K., 2013. 'Microbial biodiversity in glacier-fed streams', International Society for Microbial Ecology, 7, pp. 16511660. https://doi.org/10.1038/ismej.2013.44

Winderl, C., Penning, H., von Netzer, F., Meckenstock, R.U., Lueders, T., 2010. 'DNA-SIP identifies sulfate-reducing Clostridia as important toluene degraders in tar-oil-contaminated aquifer sediment', International Society for Microbial Ecology, 4, pp. 1314-1325. https://doi.org/10.1038/ismej.2010.54

Wu M., Li W., Dick W.A., Ye, X., Chen K., Kost D., Chen L., 2017. Bioremediation of hydrocarbon degradation in a petroleum-contaminated soil and microbial population and activity determination. Chemosphere, 169, pp. 124-130. https://doi.org/10.1016/j.chemosphere.2016.11.059

Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2007. 'Obligate oil-degrading marine bacteria', Current Opinion in Biotechnology, 18, pp. 257-266. https://doi.org/10.1016/j.copbio.2007.04.006

Figure legends

Fig. 1 Headspace CO_2 concentration (g/L) in different soil batches for Soil only (circle), Soil + aromatic hydrocarbons (triangle), Soil + straight-chain alkanes (square) and Soil + cyclic/branched alkanes (prism).

Fig. 2 A comparison of alpha diversity indices Chao1, Observed number of species, Faith's PD and Shannon's diversity index for **a)** 454 sequencing libraries at an even rarefaction depth of 500 sequences and **b)** Ion torrent sequencing library at an even rarefaction depth of 22,000 sequences.

Fig. 3 Relative abundances of phylum level operational taxonomic units (OTUs) accounting for $\geq 3.0\%$ in **a)** 454 sequencing libraries and **b)** Ion torrent of all classified sequences obtained from control samples and different soil samples contaminated with aromatic hydrocarbons, straight-chain alkanes and cyclic/branched alkanes.

Fig. 4 A Nonmetric Multi-Dimensional Scaling (NMDS) of 16S rRNA 454 -pyrosequencing libraries obtained from control soil samples, aromatic hydrocarbon treated soil, straight-chain alkane treated soils and cyclic/branched alkane treated soils at day 14. Averaged Bray Curtis distance on square root transformed (OTU level 3) data is displayed on the plot.

Fig. 5 A heatmap showing the relationship between relative abundances (> 1.0% of total OTU abundance) and batch samples of 16S rRNA 454 -pyrosequencing libraries using dendograms. Colour codes: Green rectangles = high relative abundance, red rectangles = low relative abundance and black rectangles = intermediate abundance. Samples are clustered in columns and OTUs in rows.

Fig. 6 A principal component analysis (PCA) of the relative abundances of dominant OTUs (>1.0% of total sequences) at the genus level (square root transformed data) of **a)** 454-pyrosequencing-derived dataset and **b)** Ion torrent-derived dataset as affected by petroleum hydrocarbon class. The direction of vectors indicate the direction of change of each variable (OTU relative abundance) so that the observation samples clustered together have similar microbial communities.

Tables

Table 1 Summary of bacterial order detected in the highest relative abundance obtained from DNA-derived 454 pyrosequencing libraries of samples from volatile petroleum hydrocarbon contaminated soils

Table 2 Summary of statistically significant treatment effects (uncontaminated soil batch versus aromatics, straight alkanes and cyclic/branched alkanes soil batches, t-test, two tailed, $p < 0.05$) for OTU identified at the genus level, where members of the genus reportedly degrade petroleum hydrocarbons.

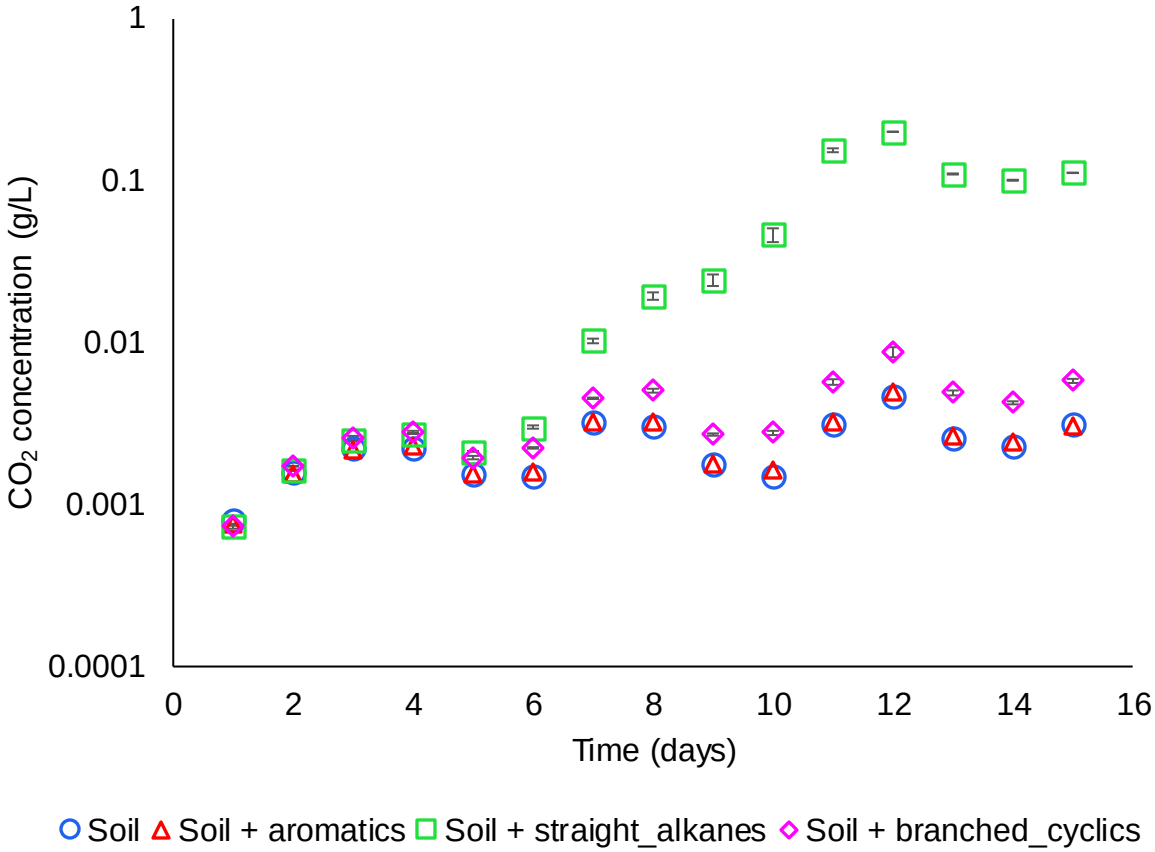

1
0.1
0.01
0.001
0.0001
CO2 concentration (g/L)
0
2
4
6
8
10
12
14
16
Time (days)
Soil
Soil + aromatics
Soil + straight_alkanes
Soil + branched_cyclics

Fig. 1

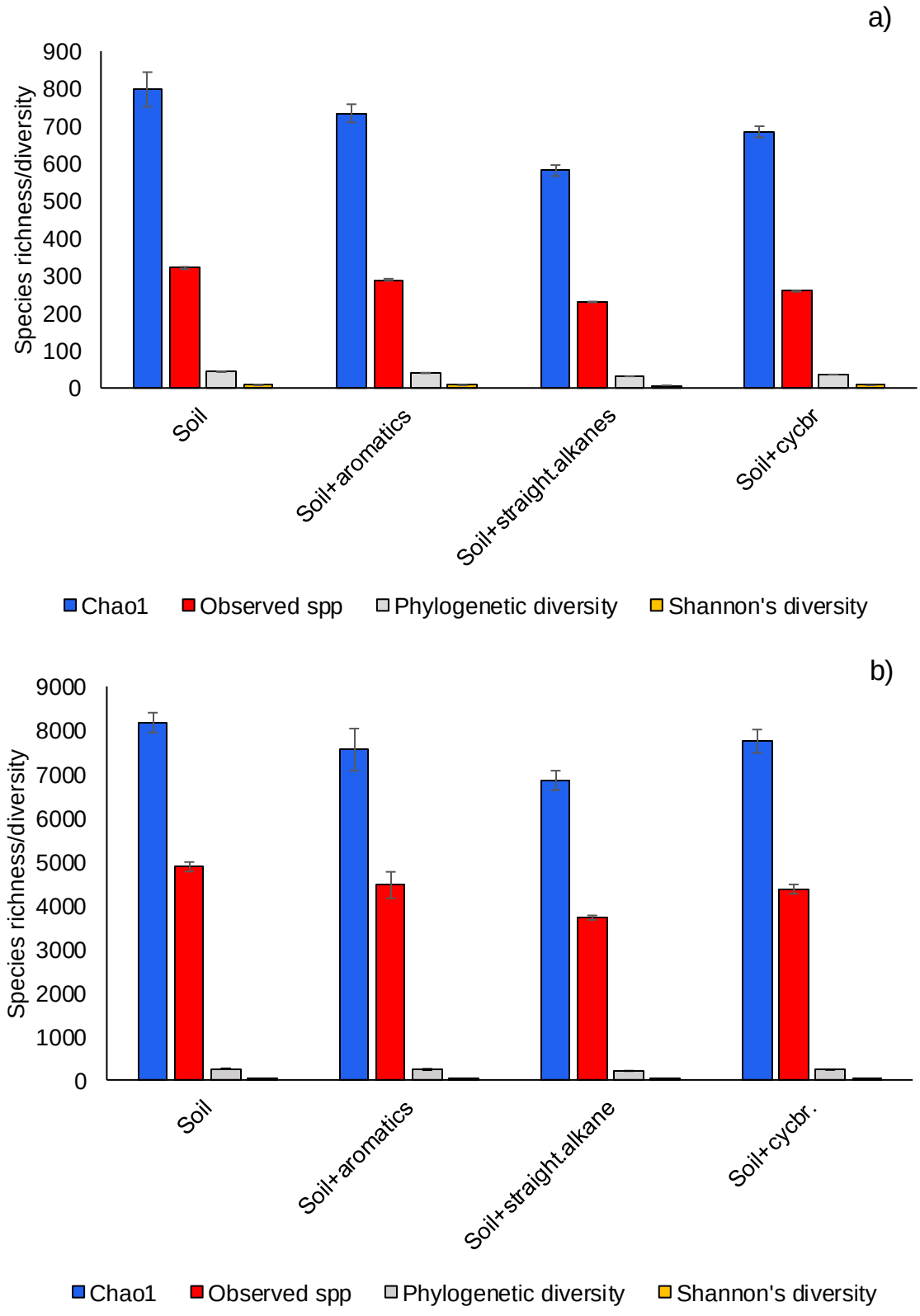

a)
Species richness/diversity
900
800
700
600
500
400
300
200
100
0
Soil
Soil+aromatics
Soil+straight.alkanes
Soil+cycbr
Chao1
Observed spp
Phylogenetic diversity
Shannon's diversity
b)
Species richness/diversity
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Soil
Soil+aromatics
Soil+straight.alkane
Soil+cycbr.
Chao1
Observed spp
Phylogenetic diversity
Shannon's diversity

Fig. 2

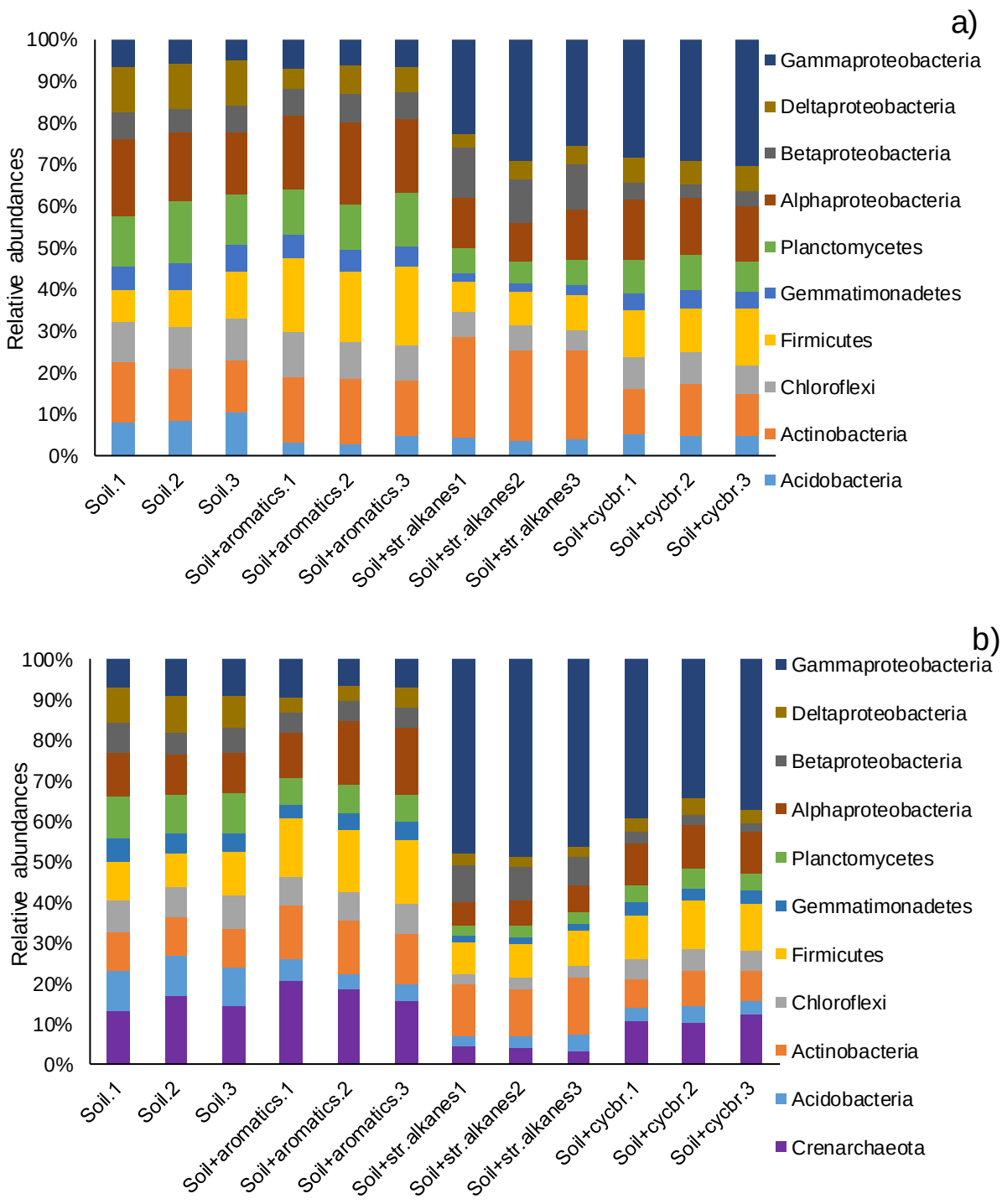

a)
Relative abundances
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Soil.1
Soil.2
Soil.3
Soil+aromatics.1
Soil+aromatics.2
Soil+aromatics.3
Soil+str.alkanes1
Soil+str.alkanes2
Soil+str.alkanes3
Soil+cycbr.1
Soil+cycbr.2
Soil+cycbr.3
Gammaproteobacteria
Deltaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Planctomycetes
Gemmatimonadetes
Firmicutes
Chloroflexi
Actinobacteria
Acidobacteria
b)
Relative abundances
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Soil.1
Soil.2
Soil.3
Soil+aromatics.1
Soil+aromatics.2
Soil+aromatics.3
Soil+str.alkanes1
Soil+str.alkanes2
Soil+str.alkanes3
Soil+cycbr.1
Soil+cycbr.2
Soil+cycbr.3
Gammaproteobacteria
Deltaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Planctomycetes
Gemmatimonadetes
Firmicutes
Chloroflexi
Actinobacteria
Acidobacteria
Crenarchaeota

Fig. 3

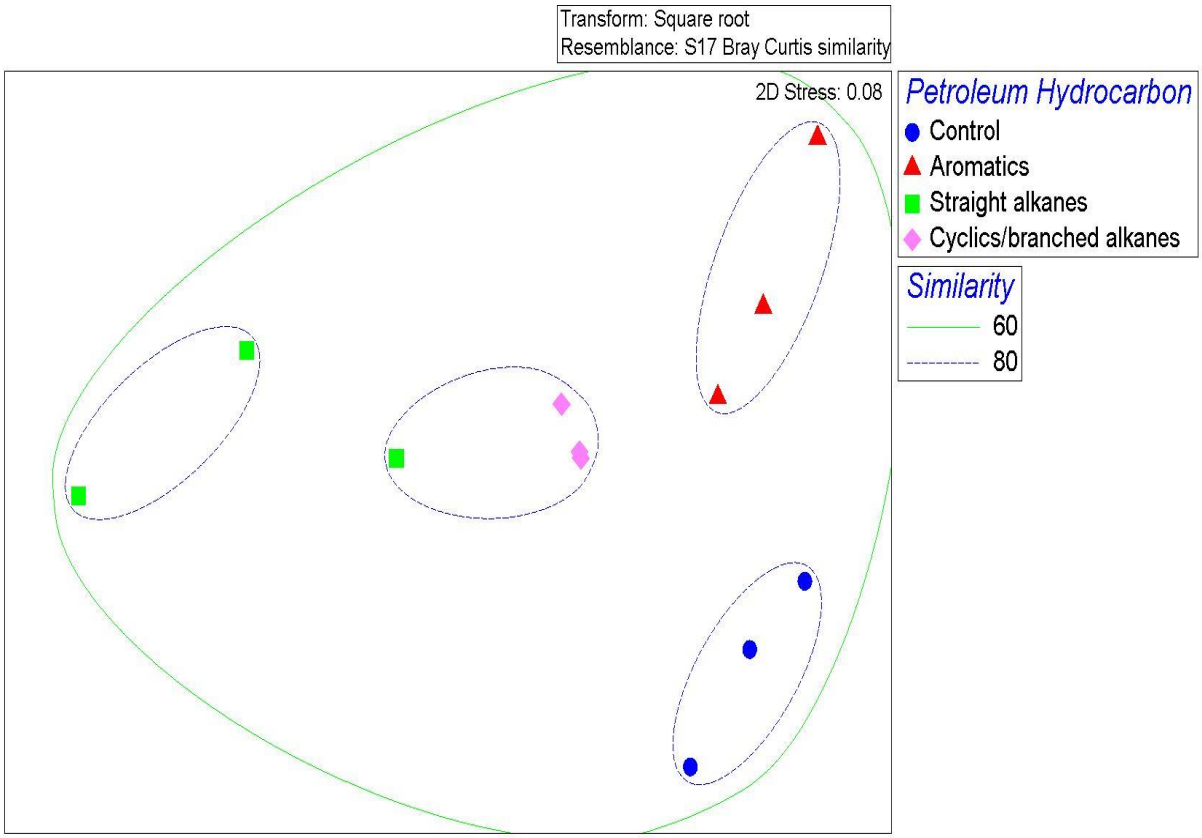

Transform: Square root
Resemblance: S17 Bray Curtis similarity
2D Stress: 0.08
Petroleum Hydrocarbon
Control
Aromatics
Straight alkanes
Cyclics/branched alkanes
Similarity
60
80

Fig. 4

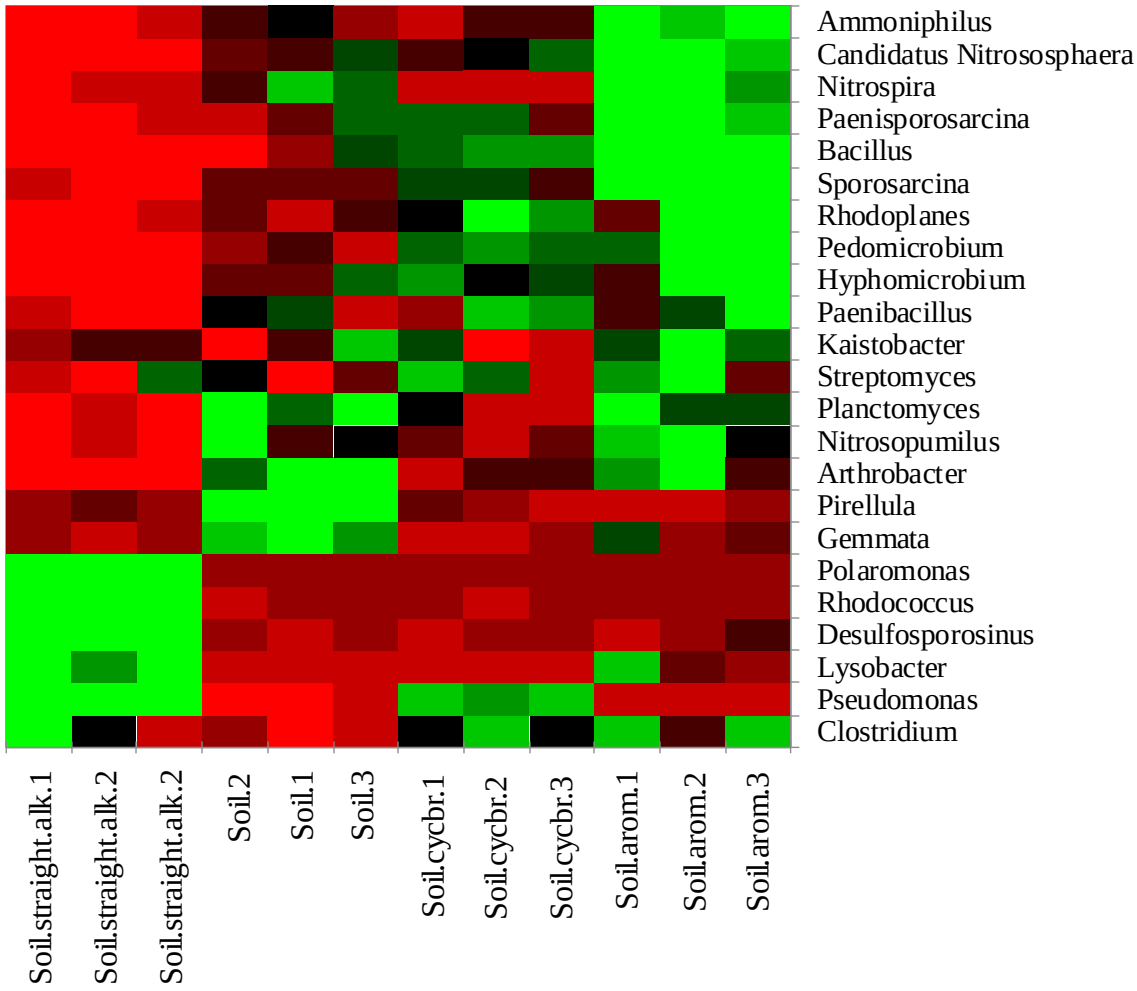

Ammoniphilus
Candidatus Nitrososphaera
Nitrospira
Paenisporosarcina
Bacillus
Sporosarcina
Rhodoplanes
Pedomicrobium
Hyphomicrobium
Paenibacillus
Kaistobacter
Streptomyces
Planctomyces
Nitrosopumilus
Arthrobacter
Pirellula
Gemmata
Polaromonas
Rhodococcus
Desulfosporosinus
Lysobacter
Pseudomonas
Clostridium
Soil.straight.alk.1
Soil.straight.alk.2
Soil.straight.alk.2
Soil.2
Soil.1
Soil.3
Soil.cycbr.1
Soil.cycbr.2
Soil.cycbr.3
Soil.arom.1
Soil.arom.2
Soil.arom.3

Fig. 5

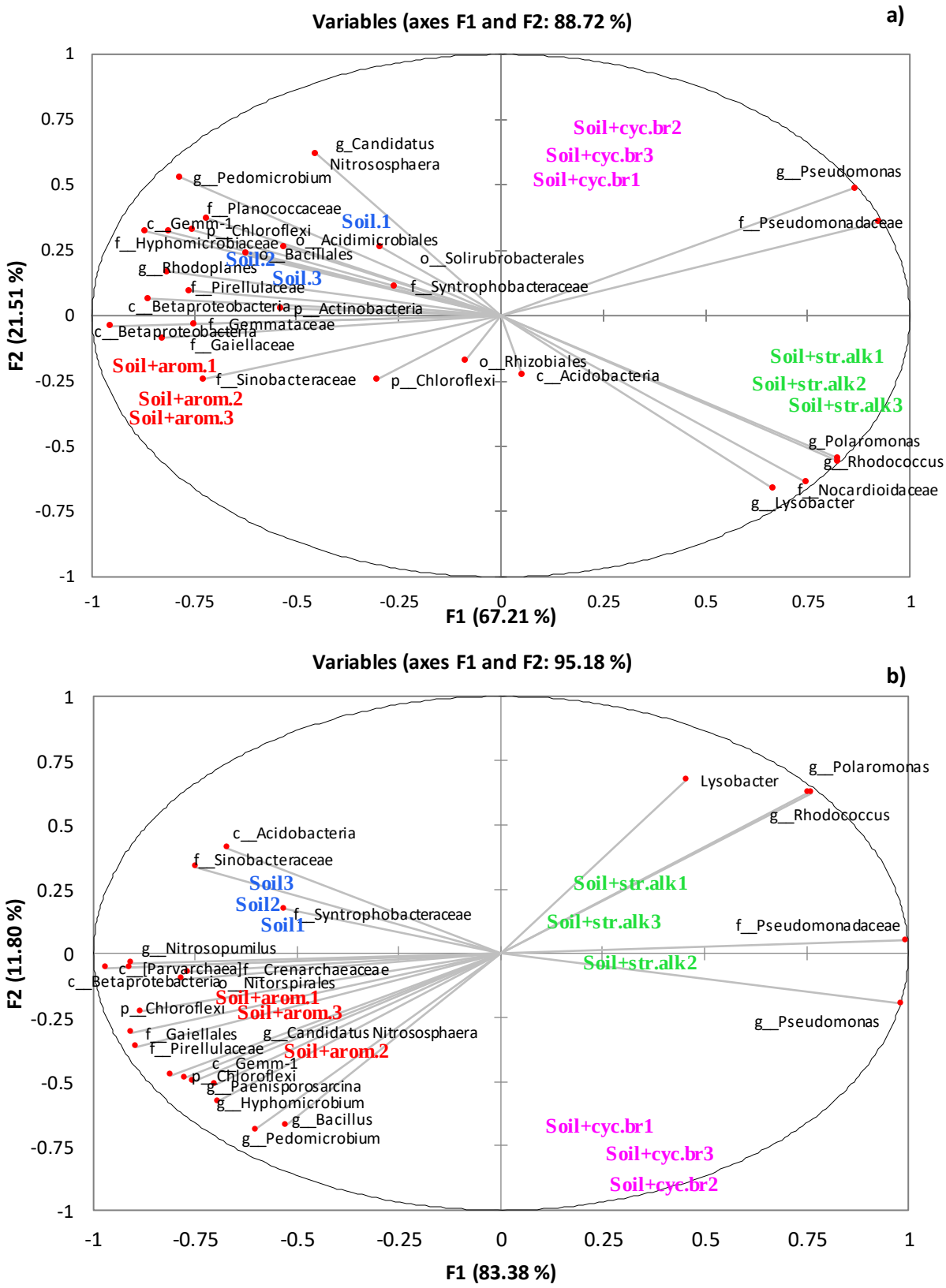

a)
Variables (axes F1 and F2: 88.72 %)
F2 (21.51 %)
F1 (67.21 %)
g_Candidatus Nitrososphaera
g__Pedomicrobium
f__Planococcaceae
c__Gemm-1
p__Chloroflexi
f__Hyphomicrobiaceae
o__Acidimicrobiales
o__Bacillales
o__Solirubrobacterales
g__Rhodoplanes
f__Pirellulaceae
f__Syntrophobacteraceae
c__Betaproteobacteria
p__Actinobacteria
f__Gemmataceae
c__Betaproteobacteria
f__Gaiellaceae
Soil+arom.1
f__Sinobacteraceae
p__Chloroflexi
o__Rhizobiales
c__Acidobacteria
Soil+arom.2
Soil+arom.3
Soil+cyc.br2
Soil+cyc.br3
Soil+cyc.br1
g__Pseudomonas
f__Pseudomonadaceae
Soil.1
Soil.2
Soil.3
Soil+str.alk1
Soil+str.alk2
Soil+str.alk3
g_Polaromonas
g__Rhodococcus
f__Nocardioidaceae
g__Lysobacter
b)
Variables (axes F1 and F2: 95.18 %)
F2 (11.80 %)
F1 (83.38 %)
g__Polaromonas
Lysobacter
g__Rhodococcus
c__Acidobacteria
f__Sinobacteraceae
Soil3
Soil2
Soil1
f__Syntrophobacteraceae
Soil+str.alk1
Soil+str.alk3
f__Pseudomonadaceae
g__Nitrosopumilus
c__[Parvarchaea]f__Crenarchaeaceae
c__Betaproteobacteria
o__Nitorspirales
Soil+str.alk2
Soil+arom.1
p__Chloroflexi
Soil+arom.3
g__Pseudomonas
f__Gaiellales
g__Candidatus Nitrososphaera
f__Pirellulaceae
Soil+arom.2
c__Gemm-1
p__Chloroflexi
g__Paenisporosarcina
g__Hyphomicrobium
g__Bacillus
g__Pedomicrobium
Soil+cyc.br1
Soil+cyc.br3
Soil+cyc.br2

Fig. 6

Tables

Table 1.

OTU classification	%Detected			
	Control	**Aromatics.**	***Straight alk.**	**Cyclics/branched**
Actinomycetales	**3.5±0.4**	**3.4±0.3**	**15.5±0.2**	**2.7±0.1**
Micrococcaceae	0.6±0.1	0.9±0.1	0.6±0.4	0.6±0.1
Micromonosporaceae	0.7±0.6	0.9±0.1	0.5±0.3	0.6±0.1
Mycobacteriaceae	0.1±0.1	0.2±0.1	0.1±0.1	0.1±0.0
Nocardiaceae	0.1±0.1	0.1±0.1	5.5±0.6	0.1±0.1
Nocardioidaceae	1.0±0.1	0.4±0.1	8.1±0.6	0.3±0.1
Pseudonocardiaceae	0.0±0.0	0.0±0.0	0.0±0.0	0.1±0.0
Streptomycetaceae	0.5±0.1	0.6±0.2	0.6±0.2	0.5±0.0
Streptosporangiaceae	0.2±0.1	0.1±0.1	0.1±0.1	0.3±0.0
Bacillales	**5.6±0.4**	**12.1±0.2**	**3.4±0.8**	**7.4±1.3**
Bacillaceae	0.1±0.2	0.3±0.1	0.1±0.0	0.2±0.0
Paenibacillaceae	0.8±0.3	1.6±0.4	0.5±0.3	1.1±0.3
Planococcaceae	1.6±0.4	2.9±0.6	0.6±0.3	1.9±0.4
Thermo-actinomycetaceae	0.2±0.2	0.4±0.2	0.1±0.1	0.3±0.1
Pirellulales	**4.2±0.2**	**5.1±0.8**	**2.3±0.4**	**3.4±0.2**
Pirellulaceae	4.2±0.2	5.1±0.8	2.3±0.4	3.4±0.2
Rhizobiales	**10.6±1.6**	**12.0±1.4**	**5.6±1.4**	**9.3±0.5**
Bradyrhizobiaceae	0.4±0.1	0.3±0.1	0.4±0.2	0.2±0.0
Hyphomicrobiaceae	8.5±0.4	9.7±0.8	3.6±1.0	7.7±0.6
Phyllobacteriaceae	0.3±0.3	0.2±0.1	0.1±0.1	0.1±0.0
Rhizobiaceae	0.2±0.1	0.2±0.2	0.3±0.3	0.1±0.0
Pseudomonadales	**0.5±0.3**	**0.3±0.0**	**16.4±1.8**	**22.2±1.2**
Pseudomonadaceae	0.5±0.3	0.2±0.0	16.3±1.7	22.2±1.2
Xanthomonadales	**3.8±0.6**	**4.2±0.4**	**7.2±1.2**	**2.8±0.3**
Sinobacteraceae	2.4±0.4	2.5±0.4	2.2±0.2	1.8±0.3
Xanthomonadaceae	1.2±0.4	1.6±0.2	5.0±1.1	1.0±0.0

**Straight alk = straight chain alkanes.*

Table 2.

Genus	Soil (Control)	Aromatics	Straight alkanes	Cyclics/branched	Reference
Rhodococcus	5.0E-04	1.2E-03	4.5E-02	9.3E-04	Alkanes, toluene (Larkin et al., 2010; Smits et al., 2001)
Pseudonocardia	5.2E-04	1.8E-03	3.0E-04	9.4E-04	Toluene (Juteau et al., 1999)
Desulfosporosinus	1.1E-03	1.8E-03	1.2E-02	10.0E-04	Toluene (Liu et al., 2004)
Polaromonas	8.4E-04	1.2E-03	5.1E-02	1.2E-03	Toluene (Mattes et al., 2008)
Pseudomonas	1.1E-03	4.2E-03	3.3E-01	2.8E-01	Alkane degradation, Aromatics, hexadecane, benzene and toluene (Mukherjee et al., 2010; Perfumo et al., 2006)
Mesorhizobium	8.5E-04	4.5E-04	1.2E-03	6.5E-04	Benzene, Toluene, Ethylene, m-Xylene. Naphthalene in ONE study and PAHs in another (Auffret et al., 2015)
Methylibium	4.7E-05	1.6E-04	1.3E-03	1.1E-04	Toluene, benzene, ethylbenzene (Nakatsu et al., 2006).
Azomonas	0.0E+00	2.2E-05	5.5E-04	2.5E-04	Some metabolites of aromatic PH such as benzoate, *p*-Toluate and catechol) while simultaneously fixing N2 and may thus play some role in the overall degradation of PH (Chen et al., 1993)
Lycinibacillus	5.0E-04	6.8E-04	3.5E-04	7.1E-04	Mixture of PHs – n-hexane, toluene, n-decane, xylene isomers e.t.c although reported to be less tolerant compared to other genera (Stancu and Grifoll, 2011)

Supplementary Information:

Microbial community responses to different volatile petroleum hydrocarbon class mixtures in an aerobic sandy soil

Author names and affiliations

George Mangse[a,b,*], David Werner[a], Paola Meynet[a] and Chukwuma C. Ogbaga,[b,c]

[a]School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, England, UK.

[b]Department of Microbiology & Biotechnology, Nile University of Nigeria, Abuja, Nigeria

[c]Department of Biological Sciences, Nile University of Nigeria, Abuja, Nigeria

***Corresponding author:**

George Mangse
Email address: george.mangse@nileuniversity.edu.ng
Present/permanent address: Department of Microbiology & Biotechnology, Nile University of Nigeria, Abuja, Nigeria

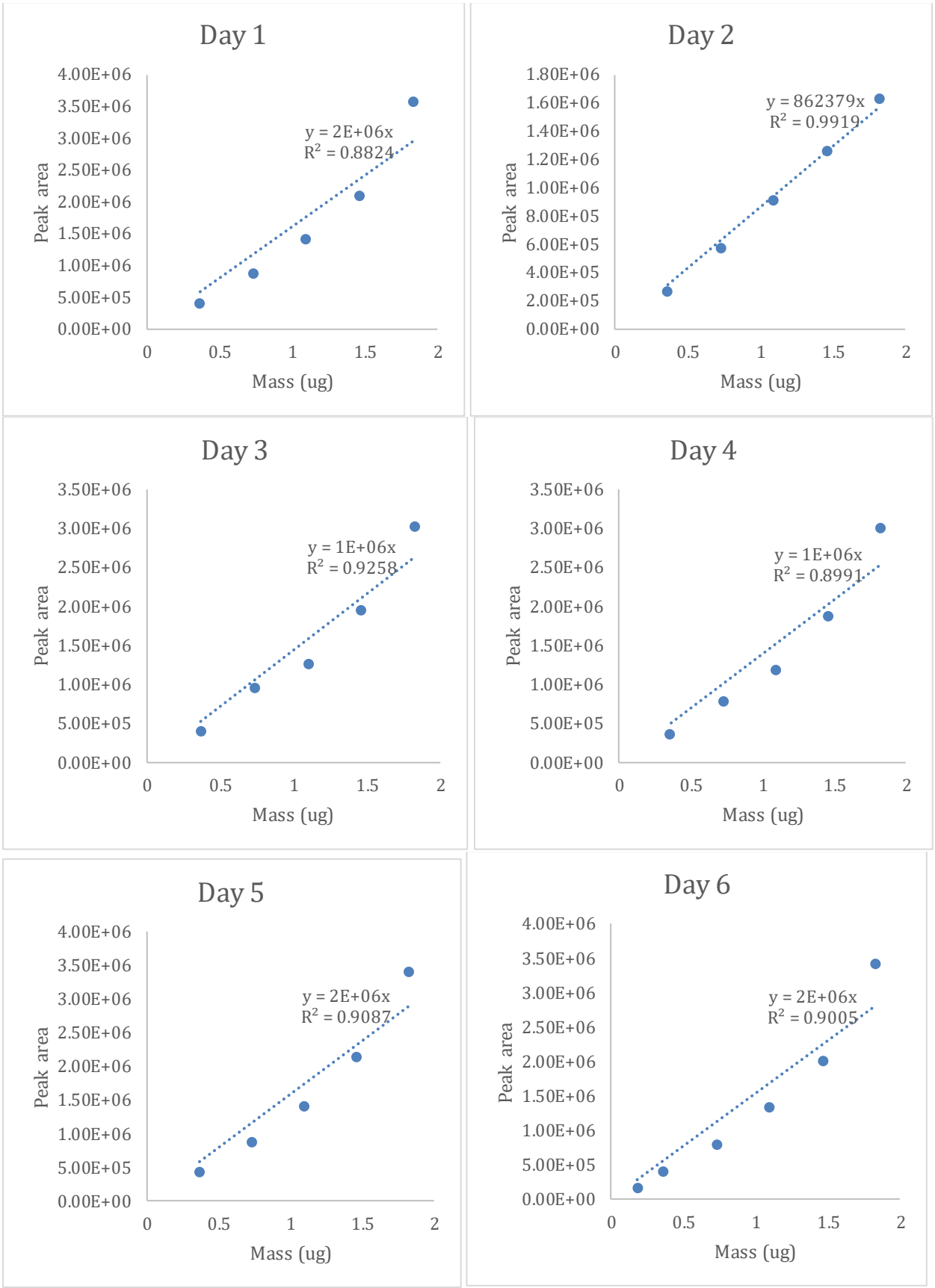


Fig. 1 Callibration standard curves for headspace gas measurements using different volumes of CO_2 generated on a daily basis from day 1 to day 6

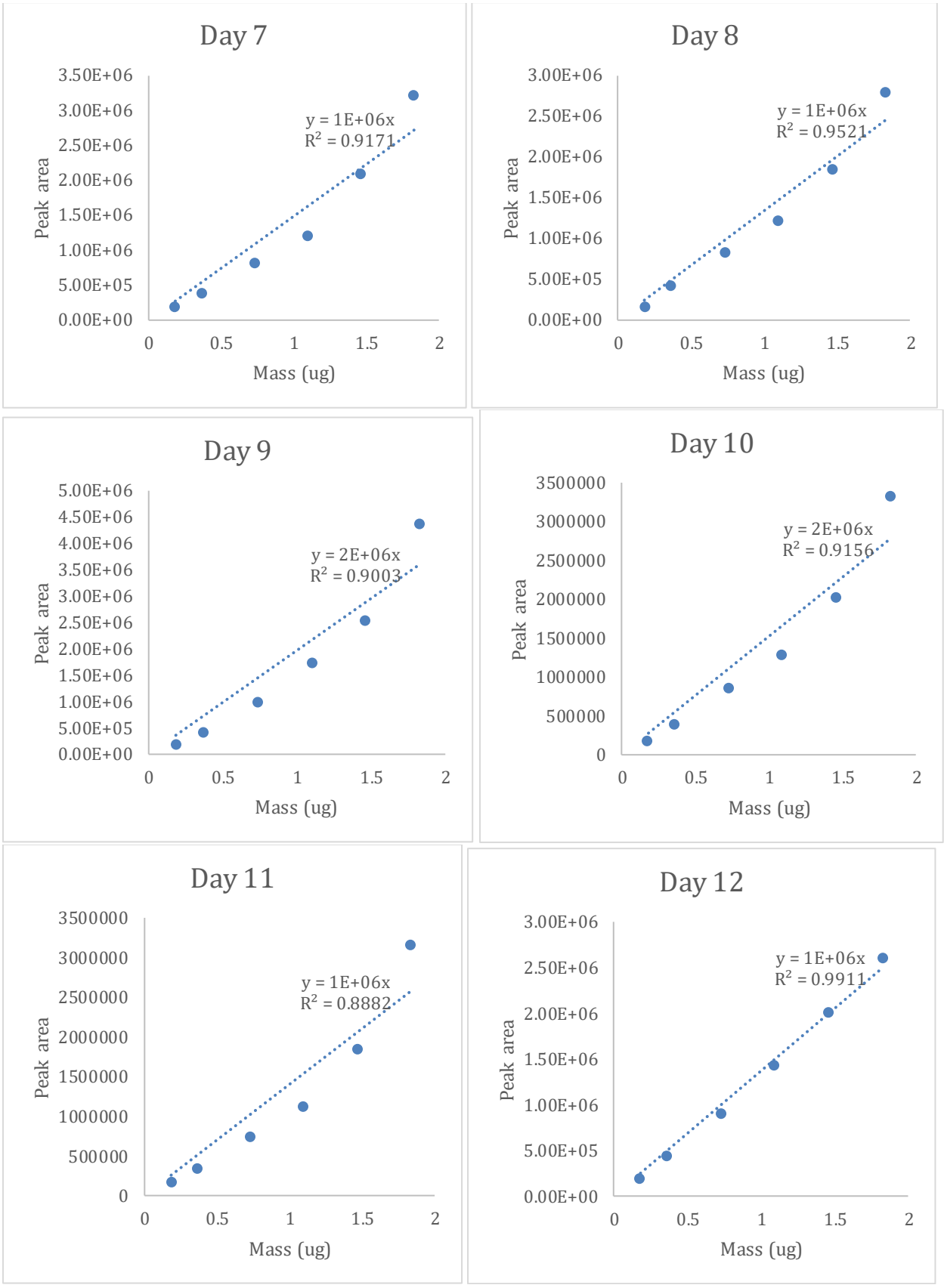


Fig 2. Callibration standard curves for headspace gas measurements using different volumes of CO_2 generated on a daily basis from day 7 to day 12

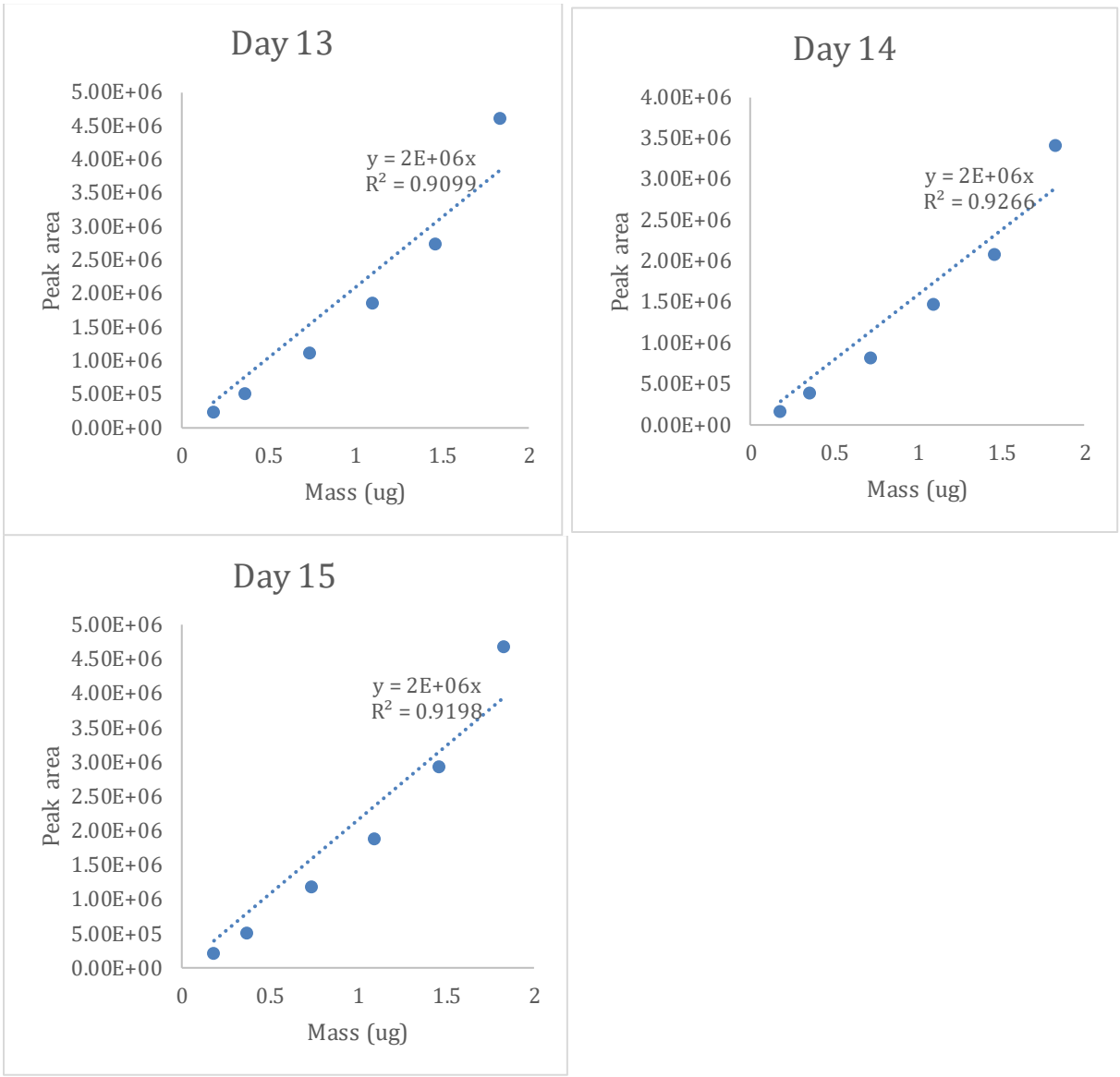
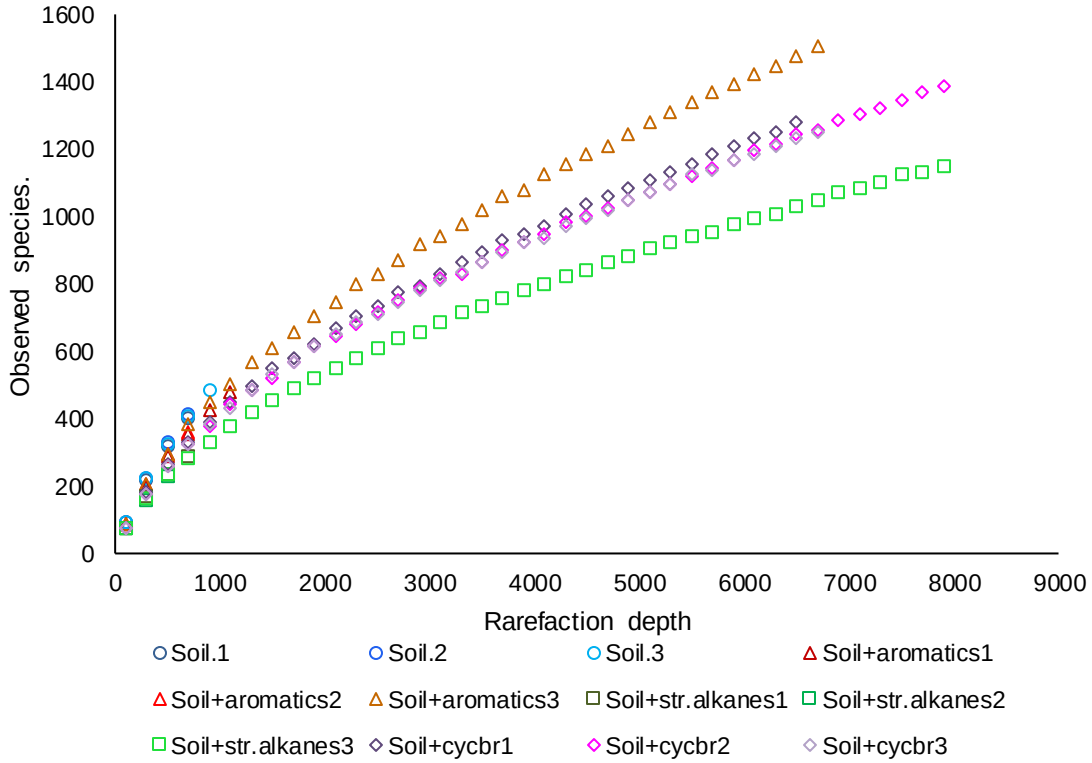


Fig 3. Callibration standard curves for headspace gas measurements using different volumes of CO_2 generated on a daily basis from day 13 to day 15

a)



b)

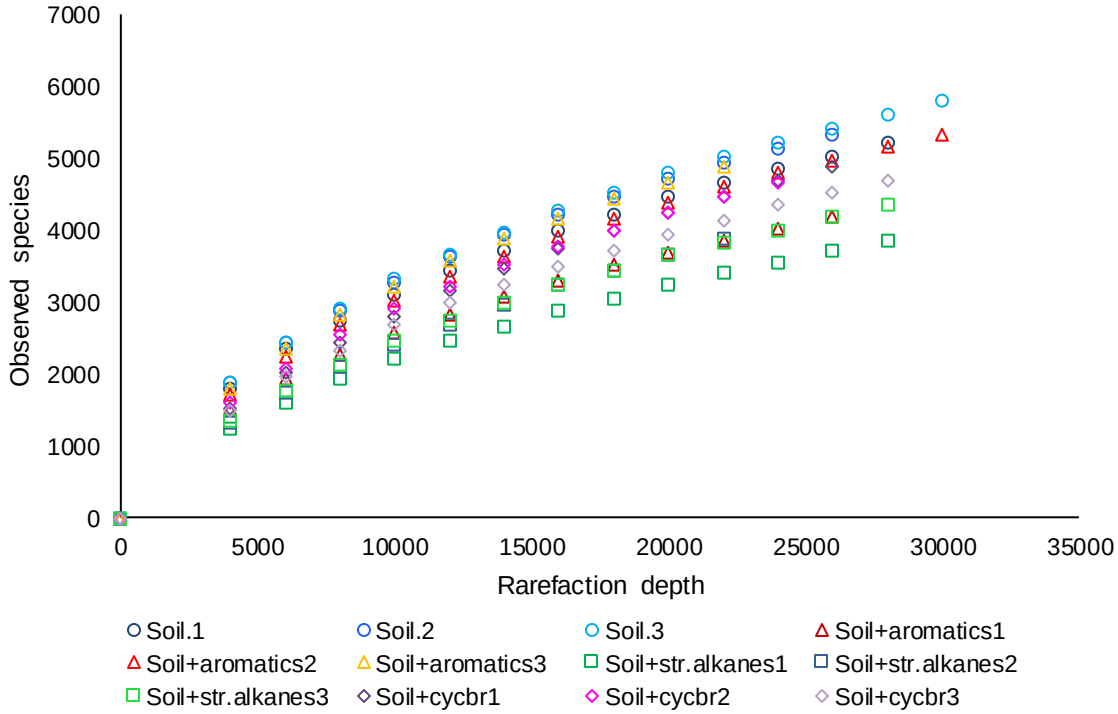


Fig 4. Rarefaction plots showing the effects of sequencing efforts on the observed number of Species for **a)** 454 sequencing and **b)** Ion torrent sequencing platforms.

Table 1. A summary of the number of sequences > 200 base pairs following quality filtering and assigned to the bacterial and archaea domain per sample in the 454-pyrosequencing generated dataset

Sample	Petroleum hydrocarbons	No of sequences > 200 bp	No of sequences assigned to domain bacteria	No of sequences classified below bacteria domain level	No of sequences assigned to domain archaea	No of sequences classified below archaea domain level	Sequences not assigned to any domain
Soil.1	N/A	772	750	749	22	22	0
Soil.2	N/A	856	825	818	30	30	1
Soil.3	N/A	1056	976	972	80	80	0
Soil_arom.1	Aromatics	1111	1069	1061	42	41	0
Soil_arom.2	Aromatics	857	839	829	18	18	0
Soil_arom.3	Aromatics	6867	6775	6732	92	90	0
Soil_str.alk.1	Straight alk.	746	737	736	9	8	0
Soil_str.alk.2	Straight alk.	694	690	689	4	4	0
Soil_str.alk.3	Straight alk.	7976	7839	7810	137	137	0
Soil_cyclbr.1	Cyclics/branched alkanes	6523	6198	6160	324	323	1
Soil_cyclbr.2	Cyclics/branched alkanes	8684	8221	8172	462	459	1
Soil_cyclbr.3	Cyclics/branched alkanes	6782	6610	6577	171	170	1

Table 2. A summary of sequence > 200 base pairs following quality filtering and assigned to the bacterial and archaea domain sample in the Ion torrent (PGM) – generated dataset

Sample	**Petroleum hydrocarbons**	**No of sequences > 200 bp**	**No of sequences assigned to domain bacteria**	**No of sequences classified below domain level**	**No of sequences assigned to domain archaea**	**No of sequences classified below domain level**	**Sequences not assigned to any domain**
Soil.1	N/A	28509	24361	24246	4146	4120	2
Soil.2	N/A	27360	22114	21984	5224	5190	22
Soil.3	N/A	31792	26846	26627	4938	4925	8
Soil_arom.1	aromatics	26042	20359	20266	5669	5647	14
Soil_arom.2	aromatics	30756	24957	24823	5792	5781	7
Soil_arom.3	aromatics	23758	19923	19876	3830	3824	5
Soil_str.alk.1	Straight alk.	29687	28219	28170	1467	1467	1
Soil_str.alk.2	Straight alk.	22119	21114	21078	1000	998	5
Soil_str.alk.3	Straight alk.	29937	28773	28712	1160	1160	4
Soil_cyclbr.1	cyclics/branched alkanes	26116	23168	23101	2947	2935	1
Soil_cyclbr.2	cyclics/branched alkanes	24129	21393	21347	2723	2709	13
Soil_cyclbr.3	cyclics/branched alkanes	28662	24858	24798	3795	3784	9

Table 3. Percentage contribution of variables (dominant OTUs > 1.0% of total relative abundance, square root transformed) to the principal components from the 454-pyrosequencing-generated dataset at genus level of taxonomy (OTU L6). (*) represents unidentified OTUs.

Contribution of the variables (%):		
	Principal Components Factors	
Genus	**F1**	**F2**
Candidatus Nitrososphaera	1.0	5.9
g__Other*	0.0	0.2
g__Other	0.2	0.2
Rhodococcus	9.3	13.3
g__Other	5.3	12.0
g__Other	0.2	0.0
g__Other	1.0	0.0
g__Other	0.1	0.2
g__Other	0.1	0.2
g__Other	0.4	0.3
g__Other	1.6	0.7
g__Other	1.3	1.1
g__Other	0.6	0.3
g__Other	0.5	0.0
g__Other	0.8	0.0
g__Other	0.0	0.1
g__Other	1.4	0.6
Pedomicrobium	1.3	1.9
Rhodoplanes	0.9	0.1
g__Other	0.8	0.0
Polaromonas	12.9	17.7
g__Other	1.8	0.0
g__Other	0.1	0.1
g__Other	31.9	14.9
Pseudomonas	24.2	24.2
g__Other	0.3	0.1
Lysobacter	2.0	6.1

Table 4. Percentage contribution of variables (dominant OTUs > 1.0% of total relative abundance, square root transformed) to the principal components from the Ion torrent generated dataset at genus level of taxonomy (OTU L6).

Contribution of the variables (%)		
	Principal Component Factors	
Genus	**F1**	**F2**
g__Other	0.5	0.0
Nitrosopumilus	2.6	0.0
Candidatus Nitrososphaera	3.4	8.4
g__Other	1.2	0.0
g__Other	0.3	0.8
Rhodococcus	4.4	21.6
g__Other	0.7	0.6
g__Other	0.2	0.6
g__Other	0.4	0.2
Bacillus	0.5	5.3
Paenisporosarcina	0.4	1.5
g__Other	0.2	0.7
g__Other	0.3	0.0
g__Other	0.8	0.9
Hyphomicrobium	0.5	2.6
Pedomicrobium	0.8	7.2
Polaromonas	5.1	23.9
g__Other	1.0	0.0
g__Other	0.2	0.2
g__Other	6.4	0.1
Pseudomonas	69.3	19.4
g__Other	0.4	0.5
Lysobacter	0.4	5.6

NGS Data analysis. Raw pyrosequencing output data from the batch experiments is displayed in Supplementary Table 1. A total of 42,924 sequences passed the quality filtering step out of 71,107 sequences representing approximately 60% of the input sequences. From the total number of quality-filtered sequences, 41,529 sequences were assigned to the domain bacteria out of which 41,305 sequences were classified representing 99%. A total of 1391 sequences were assigned to the archaeal domain and below this domain, 1382 sequences were classified representing 99%. 4 sequences were not assigned to any domain. Average read length for the pyrosequencing dataset was 378 bp and the number of sequences per sample ranged from 772 to 8684 sequences.

With respect to the Ion Torrent sequencing output (Supplementary Table 2), 3,075,520 raw sequences were originally generated from which a total of 328,867 sequences passed the quality filtering step representing approximately 10% of the input sequences. A total of 286,085 sequences were assigned to the bacterial domain representing 86.99% of the quality-filtered sequences while 42,691 sequences were assigned to the archaea domain representing 13% of the total written sequences. Below the bacterial domain, 285,028 sequences were regarded as classified representing 99.6% while below the archaea domain, 42,540 sequences were classified to represent 99.7% of this domain. Average read length for the Ion torrent sequencing dataset was 318 bp and the number of sequences per sample ranged from 22,119 to 31,792 sequences having an average of 27,406± 3021 sequences.

GC-FID analysis, headspace VPH measurements. The headspace concentration of volatile petroleum hydrocarbons in the batches was measured using a HP-7890A Series Gas Chromatograph (Agilent Technologies, Palo Alto, USA). Briefly, 30 µL of samples were injected manually using a Hamilton gas-tight syringe into the machine in split mode, the injector set at 200 °C, flame ionisation detector at 300 °C. Prior to this, a blank sample containing air was run to ensure that there had not been any previous contamination of the syringes and GC columns. Separation of headspace gas was performed on a fused silica capillary column (30 m x 0.25 mm i.d) coated with 0.25µm dimethyl poly-siloxane (HP-5 phase). The column temperature was maintained at 30 °C for 5 minutes and raised to 120 °C at a rate of 10 °C min^{-1} and then held at this temperature for 6 minutes. The hydrogen carrier gas was set at a flow rate of 1 mL min^{-1}, pressure of 50 kPa and split at 10 mls min^{-1}.