Title: Natural attenuation of legacy hydrocarbon spills in pristine soils is feasible despite difficult environmental conditions in the monsoon tropics

Deirdre B. Gleeson[1*], Belinda C. Martin[2], Tim Lardner[1], Andrew S. Ball[3], Kliti Grice[4], Alex I. Holman[4], Alastair Trolove[5], Megan Manix[5], Mark Tibbett[6,7], Gary D. Bending[8], Sally Hilton[8], Megan H. Ryan[1]

[1]UWA School of Agriculture and Environment, University of Western Australia, 35 Stirling Highway, Crawley 6009, AUSTRALIA;

[2]Ooid Scientific, South Fremantle, 6162, AUSTRALIA;

[3]RMIT University, Bundoora VIC 3083, AUSTRALIA;

[4]WA-Organic and Isotope Geochemistry Centre, School of Earth and Planetary Sciences, Curtin University, Kent Street, Bentley 6102, AUSTRALIA;

[5]Horizon Power, 18 Brodie Hall Drive, Bentley 6102, AUSTRALIA;

[6]Department of Sustainable Land Management & Soil Research Centre, School of Agriculture, Policy and Development, University of Reading, Reading, Berkshire RG6 6AH, UK

[7]School of Biological Sciences, The University of Western Australia, Crawley, WA, Australia.

[8]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

Correspondence and requests for materials should be addressed to Deirdre Gleeson (*deirdre.gleeson@uwa.edu.au)

Abstract

The Kimberley region of Western Australia is a National Heritage listed region that is internationally recognised for its environmental and cultural significance. However, petroleum spills have been reported at a number of sites across the region, representing an environmental concern. The region is also characterised as having low soil nutrients, high temperatures and monsoonal rain – all of which may limit the potential for natural biodegradation of petroleum. Therefore, this work evaluated the effect of legacy petroleum hydrocarbons on the indigenous soil microbial community (across the domains Archaea, Bacteria and Fungi) across three sites in the Kimberley region. At each site, soil cores were removed from contaminated and control areas and analysed for total petroleum hydrocarbons, soil nutrients, pH and microbial community profiling (using16S rRNA and ITS sequencing on the Illumina MiSeq Platform). The presence of petroleum hydrocarbons decreased microbial diversity across all kingdoms, altered the structure of microbial communities and increased the abundance of putative hydrocarbon degraders (e.g. *Mycobacterium*, *Acremonium, Penicillium, Bjerkandera* and *Candida*). Microbial community shifts from contaminated soils were also associated with an increase in soil nutrients (notably Colwell P and S). Our study highlights the long-term effect of legacy hydrocarbon spills on soil microbial communities and their diversity in remote, infertile monsoonal soils, but also highlights the potential for natural attenuation to occur in these environments.

Keywords: bioremediation, microbiome, petroleum hydrocarbons, infertile, monsoon

Introduction

Globally, the most widely-distributed environmental pollution is contamination by petroleum hydrocarbons that are found in crude oils, diesels and kerosene, including straight and branched alkanes, cycloalkanes, phenolics, aromatics and polycyclic aromatic hydrocarbon (PAHs) (Grice et al., 2009). Extensive use of petroleum-based products represents a constant threat of spillage, particularly during storage, transport or use via handling incidents. Such spills may result in significant environmental contamination of soils and water, and clean up can be difficult and expensive. This is especially true in pristine and remote environments such as Western Australia's National Heritage listed Kimberley region. This region, bordered by the Pilbara to the south and the Northern Territory to the east, is internationally recognised for its unique and dramatic landscape characterised by its significant biological diversity and cultural history (Commonwealth of Australia, 2011). Over the course of the past 50 years there have been reports of localised petroleum hydrocarbon spills as a result of historical diesel-powered electricity generation at a number of managed sites in this region (Horizon Power, Western Australia) representing an environmental concern.

The soil microbial community plays a major role in hydrocarbon removal in the environment, with biodegradation by indigenous soil microbial communities being one of the primary mechanisms (Bento et al., 2005; Greenwood et al., 2009). The rate of microbially-mediated biodegradation depends on factors such as the nature of the contamination and the suitability of the conditions for microbial activity (Atlas, 1995). Both nutrients and water are known to be limiting factors controlling hydrocarbon biodegradation in soils (Greenwood et al., 2009; Tibbett et al., 2011). Soils in the monsoonal tropics, such as the Kimberley region, represent a particular challenge as conditions for bioremediation are unlikely to be optimal – soils are nutrient-poor and weather conditions, particularly rainfall, are highly variable. The Kimberley region is also one of the hottest regions of Australia, with an average annual mean

temperature of 27 °C and temperatures above 30 °C for most of the year, including during the winter months. Low soil nutrients, sporadic rainfall and high temperatures, combined with the remoteness of the region means that hydrocarbon biodegradation may be limited, and contamination may persist in the soil for a long time.

In our previous work on similar soils in nearby Barrow Island in Western Australia, and in other sites in the Kimberley, we have shown how low concentrations of soil nutrient (particularly nitrogen and phosphorus) restrict hydrocarbon degradation and alter the soil microbial communities (Lin et al., 2014; Tibbett et al., 2011). However, these were short-term, *ex-situ* microcosm studies and, to date, there is a paucity of knowledge on the longer-term effect of terrestrial hydrocarbon spills on microbial communities *in-situ* in environments with naturally low nutrient concentrations that also experience high annual temperatures and monsoonal rainfall.

Here we assessed the impact of legacy hydrocarbon contamination from spills at three sites in the pristine western Australian Kimberley region (Camballin, Fitzroy Crossing and Kununurra) under the control of a regional power generator, Horizon Power. At all sites there were areas with legacy hydrocarbon contamination due to activities associated with power generation. Specifically, we aimed to answer the following questions (1) Has the indigenous microbial community (archaea, bacteria and fungi) been altered by legacy hydrocarbon contamination? (2) Is there an interaction between hydrocarbon contamination, microbial communities and soil nutrients? (3) Is there evidence of natural attenuation in these pristine, low nutrient, monsoonal soils?

Materials and Methods

Field sites and sample collection. Soils were collected from three sites located in the Kimberley region of northern Western Australia (Fitzroy Crossing (18.18°S; 125.56°E),

Camballin (17.99°S; 124.19°E) and Kununurra (15.78°S; 128.71°E) – see Fig. 1A). The region has a tropical monsoonal climate with mean monthly maximum temperatures of >35°C in summer and >30°C in winter (Fig. 1B). The region is also characterised as having a tropical monsoon climate; receiving about 90% of its rainfall during the short wet season (from November to April) when cyclones are common (Fig. 1B). All sites recieve <50 mm of rainfall during the 8-month dry season between the start of April and end of November (Fig. 1B). Two of the sites, Fitzroy Crossing and Camballin, are extremely hot in summer, with average monthly maximum temperatures of ~40°C, and daily maximums reaching up to 47°C (http://www.bom.gov.au/climate/data).

All sites contained native bushland before being cleared between 1964 and 1978, and used for power generation (Horizon Power Technical Report). Hydrocarbon contamination events are recorded as having occurred at each site, but it is possible that other major events were not recorded as historical data related to such events are minimal. Each site may have also been subject to repeated small contamination events over time including diesel, fuel oil and transformer oil.

Field sampling occurred in the dry season between 6 and 10 July 2015 and was conducted by ERM consulting (Perth, Australia). At each site, an area with potential contamination was identified based on knowledge of the site history, previous sampling for contamination, and soil appearance. An adjacent area, likely to be uncontaminated, was then also identified (control soil). Subsequent analysis showed the putatively contaminated and uncontaminated site classifications were not effective predictors of actual hydrocarbon contamination. Therefore, subsequent analysis of microbial communities was based on as-measured contaminated versus uncontaminated sites which used actual total petroleum hydrocarbons measured. Being the dry season, there was little green vegetation present at the sites, but senesced plant material present indicated that all sites, except Kununurra, likely supported the

76 growth of grasses and other herbs in the wet season. Further details of the sites, including
77 contamination history, vegetation and sampling location are presented in Table S1.
78 At each site, three replicate soil cores (n=3) were removed from both the contaminated and
79 control areas. Distance between contaminated and control sampling areas varied among sites
80 but was no less than 5 m and no greater than 20 m. Distance between replicate cores within
81 the contaminated and control areas at each site was no greater than 2 m. Each core was
82 divided into depths from the soil surface of 0–15, 15–30 and 30–45 cm, giving 18 discrete
83 samples at each of the three sites consisting of approximately 1.5 kg per sample which was
84 placed into a cloth bag and then sealed inside a plastic ziplock bag. Soil was stored in iced
85 coolers and shipped to The University of Western Australia (UWA), Perth, within 1–2 days
86 of collection by same-day air freight.
87 The cloth bags of soil were stored in a cool room at 4°C until processing, which occurred
88 within 24 hours. Samples were removed from their bags and passed through a 2-mm sieve
89 into a galvanised steel bucket and thoroughly mixed. Samples were then decanted into a
90 stainless steel tray to form a cone-shaped pile. Subsamples were taken from the pile using a
91 stainless steel spatula or spoon depending on the size of the required subsample. Each
92 subsample comprised soil taken from ten points within the pile: four points at the base of the
93 pile; four points halfway up the pile but between the sampling points at the base; one point at
94 the top of the pile; and one point at the centre of the pile. Aseptic sampling techniques were
95 employed between each sample collection: all utensils being cleaned with a phosphate-free
96 detergent, sprayed with 70% ethanol and wiped dry with clean paper towelling. Subsamples
97 were taken from each sample as follows: (1) 100 g was placed in a vial, snap frozen in liquid
98 nitrogen and stored at –80°C until DNA was later extracted; (2) 100 g was sent for analysis of
99 total petroleum hydrocarbons (TPHs); (3) 50 g was weighed, dried at 105°C for 3 days and

reweighed in order to calculate soil water content; and (4) remaining soil was air-dried (40°C) for measurement of nutrients, pH and particle size.

DNA extraction. DNA extractions were performed on all soils using the PowerSoil™ DNA Isolation Kit (MoBio Laboratories Inc.) following the manufacturer's instructions. Duplicate DNA extractions were pooled and quantified using a Qubit™ and the DNA extract was stored at –40°C before further analysis. Two 20 µl aliquots of each DNA extract were sub-sampled and (1) 20 µl was sent to the Australian Genome Research Facility (AGRF) for bacterial sequencing (16S rRNA) and (2) 20 µl was sent to the University of Warwick, UK for fungal sequencing (ITS). Blank DNA extractions were performed at the same time as sample extractions and amplified alongside DNA samples to ensure no reagent contamination. At the sequencing facilities (AGRF and Warwick University) internal controls were used to ensure that the sequencing data is not compromised (see www.agrf.org.au for full details of their protocols).

Bacterial and Archaeal amplicon sequencing. For each sample, approximately 465 base pairs of the V3/4 region of the bacterial 16S rRNA gene (Mori et al., 2013) were amplified by PCR primer set 341F (5'-CCTAYGGGRBGCASCAG-3') and 806R (5'-GGACTACNNGGGTATCTAAT-3') and sequenced by the Australian Genome Research Facility (AGRF) on the Illumina MiSeq platform using Nextera XT v2 indices and 300 bp paired end technology. Following sequencing, paired-end reads were assembled by aligning the forward and reverse reads using PEAR (version 0.9.5) (Zhang et al., 2014). Ambiguous and chimeric sequences were identified and removed using VSEARCH (version 1.4.0) with the Ribosomal Database Project as reference (Rognes et al., 2016). All downstream analyses were performed in QIIME (version 1.9.1) (Caporaso et al., 2010). Open-reference OTU picking was performed using the SortMeRNA (version 2.0) method with a minimum identity of 97%. Taxonomy was assigned using UCLUST (version 1.2.22) (Edgar, 2010) with the

greengenes database as reference (version 13.8) and sequences were aligned using PyNAST (version 1.2.2).

Fungal amplicon sequencing. For each sample, approximately 260 base pairs of the ITS region of the fungal rRNA gene were amplified by PCR primer set ITS3 (5'-GCATCGATGAAGAACGCAGC-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White et al., 1990) and sequenced on the Illumina MiSeq platform using Nextera XT v2 indices and 300 bp paired end technology. Following sequencing, paired-end reads were assembled by aligning the forward and reverse reads, trimming primers and quality filtering (-fastq_maxee 0.5) using USEARCH and UPARSE software (Edgar, 2010) (version 8.1.1861). Full-length duplicate sequences were removed and sorted by abundance. Singletons or unique reads in the data set were discarded. Sequences were clustered followed by chimera filtering using the ITS Unite database as a reference (uchime_reference_dataset_01.01.2016) (Koljalg et al., 2013). To obtain the number of reads in each OTU, reads were mapped back to OTUs with a minimum identity of 97%. Taxonomy was assigned using QIIME 1.8 (Caporaso et al., 2010) and the ITS Unite database (Koljalg et al., 2013) (sh_qiime_release_s_21.01.2016).

Petroleum hydrocarbons. Total recoverable hydrocarbons (TRHs) and polycyclic aromatic hydrocarbons (PAHs) were assessed in all samples by the National Measurement Institute (NMI) at Kensington, Western Australia (NATA accredited laboratory) with relevant controls and standards incorporated. Total recoverable hydrocarbons were measured using the USEPA Method SW-846-3510C (EPA, 1996) and samples were first extracted with a 50% dichloromethane/acetone solvent under sonication. Prepared extracts were injected into a gas chromatograph (GC) where the separation of individual components was achieved with a non-polar capillary column and detection was by flame ionisation. Note that this method does not include any clean-up procedure; hence, any extracted compound capable of detection by

flame ionisation and eluting within the C_{10}–C_{40} range on the capillary GC column is considered a petroleum hydrocarbon and included in the TPH result. Total recoverable hydrocarbons in the C_6–C_{10} range were analysed using the Purge and Trap technique with the following reported: C_6–C_{10}, >C_{10}–C_{16}, >C_{16}–C_{34}, >C_{34}–C_{40} with the limits of reporting being 25, 50, 100 and 100 mg kg^{-1}, respectively.

Polycyclic aromatic hydrocarbons including naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benz(a)anthracene, chrysene, benzo(b+k)fluoranthene, benzo(a)pyrene, indeno(1,2,3,c,d)pyrene, dibenz(a,h)anthracene and benzo(g,h,i)perylene were measured in all samples using the USEPA Method 8270E-SW-846 (EPA, 2014). These were first extracted with a 50% dichloromethane/acetone solvent under sonication and the extract was analysed using capillary column GC with a mass selective detector or tandem mass selective detector. The ions were generated using electron impact. The limits of reporting were 2 mg kg^{-1} for benzo(b+k)fluoranthene and 1 mg kg^{-1} for the remainder.

Soil nutrients, pH and particle size. The soil was sent to CSBP laboratories (Bibra Lake Western Australia; Australasian Soil and Plant Analysis Council (ASPAC) certified) for analysis. Unless otherwise specified, soil analysis methods followed those of Rayment and Lyons (2010) as follows: bicarbonate-extractable phosphorus (P) and potassium (K) (Colwell, 1963), mineral nitrogen (N) (ammonium-N plus nitrate-N) (Searle, 1984), extractable sulfur (S) (Blair et al., 1991), pH and conductivity in a soil:solution ratio of 1:5 with relevant controls and standards incorporated. Particle size was determined at UWA following the pipette extraction method (Gee and Or, 2002). For some measurements, some samples were below detection limits and these data were converted to 0 for statistical analyses. The limits were as follows: ammonium-N and nitrate-N <1 mg kg^{-1}, extractable P, K and S <1 mg kg^{-1}, and conductivity <0.01 dS m^{-1}.

Data analysis. The *a priori* experimental design consisted of three factors as follows: site (Fitzroy Crossing, Camballin and Kununurra; n=3), contamination (control and contaminated; n=2) and depth (0-15 cm, 15-30 cm and 30-45 cm; n=3) with 3 replications of each resulting in a total of 54 experimental units. Following measurement of TPH levels within each of the 54 samples collected, samples previously designated as 'contaminated' and 'control' were re-labelled as 'TPH detected' or 'TPH not detected' to reflect findings on analysis of TPH levels in the samples. To then determine whether legacy hydrocarbon contamination had a significant effect, samples were first grouped according to whether petroleum hydrocarbons were detected. As not all samples at previously labelled 'contaminated' sites had detectable hydrocarbons, an uneven design resulted among sites, particularly so with depth (Fig. S2). Subsequently, soil depth (0-45 cm) was pooled for both 'TPH dectected' and 'TPH not detected' sites and type III sum of squares was applied for all one-way and two-way ANOVA tests to determine differences in petroleum hydrocarbons and soil chemical properties (e.g. pH, NH_4-N, NO_3-N, Colwell P) among locations. Some parameters were subject to log or square root transformation to achieve normality. All statistical analyses were performed in R (version 4.0.2) (R Core Team, 2017).

Microbial community analysis. Statistical analyses were carried out in R (version 4.0.2) (R Core Team, 2017) using the Phyloseq package (version 1.30.0) (McMurdie and Holmes, 2013), ggplot2 (version 3.2.1) (Wickham, 2009) and vegan packages (version 2.5-6) (Oksanen et al. 2019). For 16S rRNA sequencing data, OTUs identified in less than 5 % of the samples were removed (filter defined by investigating prevalence plots) as well as reads identified as chloroplast, mitochondria or unknown phyla. Archaea were not filtered out and were instead analysed as a separate kingdom. For the fungal ITS data, OTUs identified in less than 2 % of samples were removed, as well as unidentified phyla. Diversity within samples (alpha diversity) was estimated on rarified data (to 31590 reads for 16S and 5690 reads for

ITS) data using Shannon, Chao 1 diversity indices and differences among sites and presence of hydrocarbons was determined by two-way ANOVA (type III sum of squares).

Bray Curtis was used to construct dissimilarity matrices of the communities (beta-diversity) and visualised using Principle Coordinates Analysis (PCoA). Stratified permutational multivariate analysis of variance (PERMANOVA; R vegan function adonis) with 999 permutations was conducted to explore the percentage variance in beta diversity that could be explained by the differences in site and the presence of hydrocarbons. Constrained Analysis of Principal Coordinates (CAP) models for bacterial and fungal communities were built for each site separately and were constrained by soil properties (~ TPH + Colwell P+ S + Colwell K + NH_4-N + NO_3-N + pH + moisture). The final model for each site was optimised using Akaike's Information Criterion (AIC) in a stepwise algorithm and checking the variance inflation factors using the vegan package (version 2.5-6) (Oksanen et al., 2016). The significance of each reduced CAP model was assessed using permutation tests with 999 permutation.

Differential abundance of OTUs between contaminated and non-contaminated soils was performed on variance stabilised data that was agglomerated to genus level using the DeSeq2 package (version 1.24.0) (McMurdie and Holmes, 2014). The significance of differentially abundant taxa was defined by log2 fold change and a Benjamini-Hochberg adjusted p-value of 0.01. All raw sequences have been uploaded to NCBI Sequence Read Archieve (SRA) under the project number PRJNA695356 for 16S rRNA data and PRJNA701380 for ITS

Results

PAHs were below detection limits in all samples (2 mg kg^{-1} for benzo(b+k)fluoranthene and 1 mg kg^{-1} for the remainder of PAHs analysed). For the contaminated soils at each site, Camballin soils had significantly lower TPH, as well as lower C_{16}-C_{34} and C_{34}-C_{40}

hydrocarbon fractions and extractable organic matter (EOM) than Fitzroy Crossing and Kununurra (Table 1). There were no differences in mean recoverable hydrocarbons between Fitzroy Crossing and Kununurra (Table 1).

Kununurra soils contained a greater concentration of inorganic N (both NH_4-N and NO_3-N) and S than Fitzroy Crossing or Camballin, particularly so for the contaminated soils (Table 2). Colwell P was lowest at Fitzroy Crossing, followed by Camballin and highest at Kununarra (Table 2). Colwell P was approximately two times higher in all contaminated soils compared to respective uncontaminated soils (Table 2). Contaminated soils also tended to have lower pH than uncontaminated soils (Table 2).

Archaeal, bacterial and fungal alpha diversity (Chao 1 and Shannon) were lower in petroleum-contaminated soils at all three sites, except Camballin where there was no difference in archaeal diversity between contaminated and uncontaminated soil (Fig. 2). Contaminated soils at Kununarra exhibited the lowest diversity for archaea, bacteria and fungi (Fig. 2). At Camballin archaeal diversity was negatively correlated with organic C only, while bacterial and fungal diversity were negatively correlated with TPH and Colwell P. At Fitzroy Crossing archael and bacterial diversity were negatively correlated with TPH, organic C and soil nutrients, fungal diversity positively correlated with pH only while soil nutrients in general positively correlated with TPH at this site. At Kununurra only bacterial and fungal diversity were negatively correlated with TPH. Archaeal, bacterial and fungal diversity negatively correlated with organic C while archaeal diversity was negatively correlated with soil nutrients (excluding K) and fungal diversity was negatively correlated with soil nutrients (excluding K and NO_3-N).

The composition of archaeal soil communities differed between soils with and without hydrocarbons detected ($F_{1,52} = 3.64$, $p < 0.001$) and site ($F_{2,52} = 2.55$, $p < 0.003$), with 6 % of total variation explained by presence of hydrocarbons, 9 % by site and 7 % by their interaction

(Fig. 3A). The composition of bacterial soil communities also differed according to the presence of hydrocarbons ($F_{1,52} = 7.15$, $p < 0.001$) and location ($F_{2,52} = 4.32$, $p < 0.001$), with 11 % of total variation explained by the presence of hydrocarbons, 13 % explained by site and 7 % explained by the interaction between hydrocarbon presence and site (Fig. 3B). Fungal soil communities differed with the presence of hydrocarbons ($F_{1,52} = 3.24$, $p < 0.001$), and site ($F_{12,52} = 5.0$, $p < 0.001$). However, the presence of hydrocarbons only explained 5 % of the total variation in fungal community structure, whereas site accounted for 16 % of the total variation and 6 % was explained by the interaction between hydrocarbon presence and site (Fig. 3C).

Camballin soils had the smallest differentiation in microbial communities between contaminated and uncontaminated soils of the three sites (Fig. 4A1, B1 and C1). TPH, Colwell P and dissolved inorganic N (NO_3-N and NH_4-N) were also predictors of the microbial community structure in contaminated soils at Camballin, whilst moisture, pH and S were predictors of the communities in uncontaminated Camballin soils (Fig. 4A1, B1 and C1). Archaeal, bacterial and fungal communities in petroleum contaminated soils at Fitzroy Crossing were related to higher concentrations of TPH, Colwell P, S, Colwell K, and NO_3-N, whist uncontaminated soils at Fitzroy Crossing were related to higher soil pH, though this was not the case for archaea (Fig. 4A2, B2, and C2). There was a clear differentiation in archaeal, bacterial and fungal communities between contaminated and uncontaminated soils at the Kununarra site, with contaminated soil microbial communities related to greater NH_4-N, moisture and S for bacteria and archaea and S for fungi (Fig. 4A3, B3 and C3). The two outliers from the Fitzroy Crossing and Kununarra sites were samples collected from the historically contaminated sites that were expected to contain hydrocarbons, but were found to have no recoverable hydrocarbons on analysis.

Only two archaeal genera were differentially abundant among contaminated and non-contaminated soils; Candidate genus *Nitrososphaera,* which was significantly more abundant in uncontaminated soils and *Methanocella,* which was more abundant in contaminated soils (Fig. 5A2, A3). No archaea were significantly different between contaminated and non-contaminated soils at Camballin.

Prevalent (top 10 most abundant) bacterial taxa enriched in contaminated soils included *Bacteriovorax* (*Deltaproteobacteria*), Candidate taxa *Rhodoluna* (*Actinobacteria*), *Thiovirga* (*Gammaproteobacteria*), *LCP 26* (*Caldithrixaceae*) and *Mycobacterium* (*Actinobacteria*) (Fig. 5B1-B3). Prevalent bacterial taxa more enriched in uncontaminated soils across multiple sites included *Roseiflexus* (*Chloroflexi*), *Chloroflexus* (*Chloroflexi*), *Symbiobacterium* (*Firmicutes*), *Nitrospina* (*Deltaproteobacteria*), *Brevibacillus* (*Firmicutes*) (Fig. 5B1-B3).

Prevalent fungal taxa enriched in oil contaminated soils included *Glomus* (*Glomeromycetes*), *Kamienskia* (*Glomeromycetes*), *Gymnoascus* (*Eurotiomycetes*), *Neogaeumannomyces* (*Sordariomycetes*), *Acremonium* (*Sordariomycetes*), *Candida* (*Saccharomycetes*), *Bjerkandera* (*Agaricomycetes*), *Cryptocoryneum* (*Dothideomycetes*), *Chaetomium* (*Sordariomycetes*), *Toxicocladosporium* (*Dothideomycetes*), and *Penicillium* (*Eurotiomycetes*) (Fig. 5C1-C3).

Discussion

Microbial community shifts in legacy oil contaminated soils

Legacy oil spills across the Kimberley region of Western Australia have resulted in significant changes to the structure and diversity of native soil microbial communities. Studies investigating the long term effects (over 20 years) of petroleum hydrocarbons on soil microbial communities are rare, and even more so in the context of the difficult and remote environmental conditions presented at the field sites examined here. Petroleum hydrocarbons

(PH) have previously been associated with decreases in diversity for bacteria and fungi in mostly temperate soils (Borowik et al., 2017; Morais et al., 2016; Sutton et al., 2013) as well significant changes to microbial community structure (Abed et al., 2014; Bacosa and Erdner, 2018; Martin et al., 2016; Morais et al., 2016). This change in community structure and diversity is expected as only those microbes that are able to tolerate changed soil conditions and toxicity of hydrocarbons can survive. In the current study we might have expected that microbial communities would have fully recovered at the point of sampling given the length of time since contamination, in the case of the Camballin site up to 27 years, however this was not the case perhaps indicating that the long-term legacy observed here could potentially be as a result of prevailing climatic and edaphic conditions.

Mycobacterium was among the most prevalently abundant bacteria (top 10 most abundant) that was enriched in contaminated soils at Kununarra and Fitzroy Crossing. *Mycobacterium* is a well-known PH degrader, possessing enzymes also capable of degradation several polycyclic aromatic hydrocarbons (Kweon et al., 2010). Additionally, *Acremonium, Penicillium, Bjerkandera* and *Candida* were prevalent fungi (top 10 most abundant) that were enriched in contaminated soils and have been previously shown to degrade petroleum hydrocarbons (Field et al., 1992; Gargouri et al., 2015; Germida et al., 2002; Govarthanan et al., 2017; Ma et al., 2015). Other microbes that were enriched in contaminated soils from our study have been recovered previously from oil contaminated soils including the hydrogenotrophic methanogenic archaea *Methanocella* (Tischer et al., 2013), the bacteria *Bacteriovorax* (Bacosa and Erdner, 2018) and arbuscular mycorrhiza fungi including *Glomus* (Garcés-ruiz et al., 2017) and *Kamienskia* (Garcés-ruiz and Declerck, 2019). Whilst these microbes may not be directly involved in PH degradation, they appear to benefit from altered competitive dynamics and/or soil conditions associated with PH contamination allowing them to proliferate. Identification of key putative PH degraders in such a harsh environment may

provide evidence for natural attenuation in other similar environments suggesting that key degraders are not universal but relate to the prevailing environment and geography. In the current study the presence of petroleum hydrocarbons reduced the relative abundance of indigenous microbial taxa in these soils, resulting in a reduction in overall diversity. Some notable microbes with lower relative abundance in contaminated soils included the archaeal *Candidate genus Nitrososphaera,* bacteria from the class *Chloroflexia, "*green non-sulfur bacteria" (*Roseiflexus* and *Chloroflexus*) and bacteria from the class *Firmicutes* (*Symbiobacterium* and *Brevibacillus*). *Candidate genus Nitrososphaera* is an autotrophic ammonia-oxidising archaea whose abundance has been previously found to decrease in crude oil contaminated soils (Morais et al., 2016). Urakawa (2012) found that ammonia-oxidising archaea were several times more sensitive to crude oil than ammonia-oxidising bacteria, which may have significant implications for N cycling if these community shifts persist. Further research is required to understand the implications of these community shifts for long-term soil health in these low nutrient environments.

Interaction of petroleum hydrocarbons and soil nutrients

Microbial community structure in contaminated soils was not only impacted by the presence of hydrocarbons but was also related to increased soil nutrients. Generally microbial communities in contaminated soils were associated with increased soil PO_4, NH_4-N, NO_3-N and S, while communities in uncontaminated soils were associated with increased soil pH. This may be related to the nitrogen, sulphur and oxygen (NSO) fraction in hydrocarbons which are sometimes called non-hydrocarbons or resins. NSO compounds are usually present in small quantities but have been found to increase in the residual oil fractions (Bailey et al., 1973; Westlake et al., 1978), similar to our current findings. The major nutrients likely to limit biodegradation of hydrocarbons are P, NH_4-N, and NO_3-N, and these nutrients are all

comparable to concentrations found in other studies in north Western Australia (Tibbett et al., 2011) but with Colwell P found at higher concentration than compared to nitrogenous compounds in the current study. Regardless, these concentrations of nutrients are not uncommon in Australian native ecosystems (Orians and Milewski, 2007; Spain et al., 2015; Westoby, 1988) and are very low in global terms and more comparable to beach sands than post-glacial soils in the northern hemisphere (Martinez and Garcia-Franco, 2008). Contaminated soils at Kununarra were also associated with greater N, S and P than their respective uncontaminated soils. The higher concentrations of S and P in contaminated soils may be due to the use of surfactants (which can contain high levels of sulfonates and phosphates), particularly if the contamination event occurred as a result of washing oil drums as reported (Table S1). Such surfactants, along with other additives in storage containers and flowlines, can have a significantly deleterious effect on soil physical properties (George et al., 2011). Additionally the increased soil Colwell P could be from the oil itself, as phosphate-containing oil-soluble compounds are often added to unrefined and partially refined hydrocarbons to act as anti-fouling agents (Gazulla et al., 2017). The different associations between microbial communities and soil nutrients among contaminated sites reported here emphasise the difficulties in predicting how indigenous soil communities will respond to spills, particularly in these monsoonal, low nutrient soils.

Evidence of natural attenuation in pristine monsoon soils

From the TPH profiles (enriched in heavy fractions), as well as the presence of putative hydrocarbon-degrading microbes, it is evident that all three sites across the Kimberley are undergoing, or have undergone, some degree of natural attenuation. It is possible that microbial degradation of TPH occurred until soil nutrients had run out, and we can speculate that this was then followed by natural attenuation processes. In general, we do report that

375 TPH is positively correlated with soil nutrients, however, which nutrient is positively
376 correlated to TPH is site dependent and potentially also relates to the type of contamination
377 present and any co-contaminants present. While biodegradation is partly responsible,
378 volatilisation and photochemical oxidation are also important natural attenuation pathways,
379 especially in these hot climates, as dry conditions can promote volatilisation of PHs (Dragun,
380 1998). This would partly account for the considerable attenuation of short chain alkanes
381 (<C9) which are most susceptible to volatilisation yet can also be toxic to microorganisms
382 (Atlas, 1995). Typically, C_6 to C_{16} alkanes are biodegraded monoterminally to the
383 corresponding alcohol, aldehyde, and monobasic fatty acid (Singer and Finnerty, 1984). The
384 contaminated soils at Camballin had the lowest TPH and the lowest differentiation in
385 microbial communities between contaminated and uncontaminated soils compared to both
386 Fitzroy Crossing and Kununarra. There is limited information available on historical
387 spills/contamination events at these sites. A spill was recorded (presumed to be diesel) at
388 Camballin in 1994 (27 years before the date of collection), but there is no information
389 regarding the extent and duration of this spill. Leaking transformer oil (at a rate of 20 L day^{-1})
390 was recorded to have occurred at Fitzroy Crossing in 1998, which lasted “some weeks”,
391 whilst Kununarra records note possible leaking waste oil drums as well as potential
392 contamination from washing large (600 kL) oil storage tanks. Because of differences in their
393 contamination history, as well as the poor accuracy of records, we are unable to directly
394 relate the microbial community shifts to total oil exposure within each site. It is possible that
395 the soils at Camballin were either exposed to lower amounts of total hydrocarbons, or that
396 there was indeed only one spill at this site (in 1994), and subsequently there has been more
397 time for natural attenuation to occur and for the microbial community to return to a more
398 ‘natural’ state (i.e. resembling uncontaminated soils). Regardless, the aged TPH profiles
399 along with putative degraders provides evidence of natural attenuation (or at least the

potential for natural attenuation) within all three sites in the Kimberly, despite being in an environment that is typically considered not amenable to natural attenuation (e.g. low nutrients, highly variable rainfall, high temperatures). Further investigation of the stability of the residual hydrocarbons and microbial community shifts over time in these soils would allow a more thorough assessment of the abiotic factors that may be limiting complete degradation of hydrocarbons at these sites and ultimately show the effectiveness of natural attenuation as a sustainable and cost-effective alternative to chemical treatment in these remote locations.

Acknowledgements

This work was funded by an Australian Research Council Linkage Project (LP110201130). GB and SH were funded by the UK Natural Environment Research Council (Grant NE/S010270/1).

References

Abed, R.M.M., Al-Kindi, S., Al-Kharusi, S., 2014. Diversity of Bacterial Communities Along a Petroleum Contamination Gradient in Desert Soils. Microb. Ecol. 69, 95–105. https://doi.org/10.1007/s00248-014-0475-5

Atlas, R.M., 1995. Bioremediation of petroleum pollutants. Int. Biodeterior. Biodegradation 35, 335–336. https://doi.org/10.1016/0964-8305(95)90041-1

Bacosa, H.P., Erdner, D.L., 2018. Hydrocarbon degradation and response of sea floor sediment bacterial community in the northern Gulf of Mexico to light Louisiana sweet crude oil. ISME J. 2532–2543. https://doi.org/10.1038/s41396-018-0190-1

Bailey, N.J.L., Jobson, A.M., Rogers, M.A., 1973. Bacterial degradation of crude oil: Comparison of field and experimental data. Chem. Geol. 11, 203–221.

https://doi.org/10.1016/0009-2541(73)90017-X

Bento, F.M., Camargo, F. a O., Okeke, B.C., Frankenberger, W.T., 2005. Comparative bioremediation of soils contaminated with diesel oil by natural attenuation, biostimulation and bioaugmentation. Bioresour. Technol. 96, 1049–1055. https://doi.org/10.1016/j.biortech.2004.09.008

Blair, G.J., Chinoim, N., Lefroy, R.D.B., Anderson, G.C., Crocker, G.J., 1991. A soil sulfur test for pastures and crops. Soil Res. 29, 619–626.

Borowik, A., Wyszkowska, J., Oszust, K., 2017. Functional diversity of fungal communities in soil contaminated with diesel oil. Front. Microbiol. 8, 1–11. https://doi.org/10.3389/fmicb.2017.01862

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., 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. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth0510-335

Colwell, J.D., 1963. The estimation of the phosphorus fertilizer requirements of wheat in southern New South Wales by soil analysis. Aust. J. Exp. Agric. Anim. Husb. 3, 190–197.

Dragun, J., 1998. The Soil Chemistry of Hazardous Materials, 2nd ed. ed. Amherst Scientific Publisher, Amherst, MA.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

EPA, USA., 2014. Method 8270E (SW-846): Semivolatile Organic Compounds by Gas Chromatography/ Mass Spectrometry (GC/MS). Washington, DC.

EPA, USA., 1996. SW-846 Test Method 3510C: Separatory Funnel Liquid-Liquid Extraction. Washington, DC.

Field, J.I.M.A., Jong, E.D.D.E., Costa, G.F., Bont, J.A.N.A.M.D.E., 1992. Biodegradation of Polycyclic Aromatic Hydrocarbons by New Isolates of White Rot Fungi. Appl. Environ. Microbiol. 58, 2219–2226.

Garcés-ruiz, M., Declerck, C.S.S., 2019. Community composition of arbuscular mycorrhizal fungi associated with native plants growing in a petroleum polluted soil of the Amazon region of Ecuador. Microbiol. Open 1–9. https://doi.org/10.1002/mbo3.703

Garcés-ruiz, M., Senés-guerrero, C., Declerck, S., Pirttilä, A.M., Pagano, M.C., 2017. Arbuscular Mycorrhizal Fungal Community Composition in Carludovica palmata, Costus scaber and Euterpe precatoria from Weathered Oil Ponds in the Ecuadorian Amazon. Front. Microbiol. 8, 1–13. https://doi.org/10.3389/fmicb.2017.02134

Gargouri, B., Mhiri, N., Karray, F., Aloui, F., Sayadi, S., 2015. Isolation and Characterization of Hydrocarbon-Degrading Yeast Strains from Petroleum Contaminated Industrial Wastewater. Biomed Res. Int. 1–11. https://doi.org/10.1155/2015/929424

Gazulla, M.F., Rodrigo, M., Orduña, M., Ventura, M.J., Andreu, C., 2017. Determination of Phosphorus in Crude Oil and Middle Distillate Petroleum Products by Inductively Coupled Plasma–Optical Emission Spectrometry. Anal. Lett. 50, 2465–2474. https://doi.org/10.1080/00032719.2017.1296853

Gee, G.W., Or, D., 2002. Particle size analysis, in: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis, Part 4, Physical Methods. Soils Science Society of America, Madison.

George, S.J., Sherbone, J., Hinz, C., Tibbett, M., 2011. Terrestrial exposure of oilfield flowline additives diminish soil structural stability and remediative microbial function. Environ. Pollut. 159, 2740–2749. https://doi.org/10.1016/j.envpol.2011.05.023

Germida, J.J., Frick, C.M., Farrell, R.E., 2002. Phytoremediation of oil-contaminated soils,

in: Violante, A., Huang, P.., Bollag, J.M., Gianfreda, L. (Eds.), Developments in Soil Science. Elsevier, pp. 169–186.

Govarthanan, M., Fuzisawa, S., Hosogai, T., Chang, Y.-C., 2017. Biodegradation of aliphatic and aromatic hydrocarbons using the filamentous fungus Penicillium sp . CHY-2 and characterization of its manganese peroxidase activity. R. Soc. Chem. 7, 20716–20723. https://doi.org/10.1039/c6ra28687a

Greenwood, P.F., Wibrow, S., George, S.J., Tibbett, M., 2009. Hydrocarbon biodegradation and soil microbial community response to repeated oil exposure. Org. Geochem. 40, 293–300. https://doi.org/10.1016/j.orggeochem.2008.12.009

Grice, K., Lu, H., Atahan, P., Asif, M., Hallmann, C., Greenwood, P., Maslen, E., Tulipani, S., Williford, K., Dodson, J., 2009. New insights into the origin of perylene in geological samples. Geochim. Cosmochim. Acta 73, 6531–6543. https://doi.org/10.1016/j.gca.2009.07.029

Koljalg, U., Nilsson, H., Abarenkov, K., Tedersoo, L., Taylor, A., Bahram, M., Bates, S., Bruns, T., Bengtsson-Palme, J., Callaghan, T., Douglas, B., Drenkhan, T., Eberhardt, U., Duenas, M., Grebenc, T., Griffith, G., Hartmann, M., Kirk, P., Kohout, P., Larsson, E., Lindahl, B., Lucking, R., Martin, M., Matheny, B., Nguyen, N., Niskanen, T., Oja, J., Peay, K., U, P., Peterson, M., Poldmaa, K., Saag, L., Saar, I., Schubler, A., Scott, J., Senes, C., Smith, M., Suija, A., Taylor, L., Teresa, T., Weiss, M., Larsson, K., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271–5277.

Kweon, O., Kim, S.J., Freeman, J.P., Song, J., Baek, S., Cerniglia, C.E., 2010. Substrate specificity and structural characteristics of the novel Rieske nonheme iron aromatic ring-hydroxylating oxygenases NidAB and NidA3B3 from Mycobacterium vanbaalenii PYR-1. MBio 1, 9–14. https://doi.org/10.1128/mBio.00135-10

500 Lin, D.S., Taylor, P., Tibbett, M., 2014. Advanced multivariate analysis to assess remediation
501 of hydrocarbons in soils. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-014-
502 3140-0
503 Ma, X., Ding, N., Peterson, E.C., 2015. Bioaugmentation of soil contaminated with high-
504 level crude oil through inoculation with mixed cultures including Acremonium sp.
505 Biodegradation 26, 259–269. https://doi.org/10.1007/s10532-015-9732-7
506 Martin, B.C., George, S.J., Price, C.A., Shahsavari, E., Ball, A.S., Tibbett, M., Ryan, M.H.,
507 2016. Citrate and malonate increase microbial activity and alter microbial community
508 composition in uncontaminated and diesel-contaminated soil microcosms. Soil 2, 487–
509 498. https://doi.org/10.5194/soil-2-487-2016
510 Martinez, M.L., Garcia-Franco, J.G., 2008. Plant-plant interactions in coastal dunes, in:
511 Coastal Dunes. Springer, Berlin, pp. 205–220.
512 McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is
513 inadmissible. PLoS Comput. Biol. 10, 1–12.
514 https://doi.org/10.1371/journal.pcbi.1003531
515 McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive
516 analysis and graphics of microbiome census data. PLoS One 8, 1–11.
517 https://doi.org/10.1371/journal.pone.0061217
518 Morais, D., Pylro, V., Clark, I.M., Hirsch, P.R., Tótola, M.R., 2016. Responses of microbial
519 community from tropical pristine coastal soil to crude oil contamination. PeerJ 2016, 1–
520 21. https://doi.org/10.7717/peerj.1733
521 Mori, H.I., Maruyama, F.U., Kato, H.I., Toyoda, A.T., Dozono, A.Y., Ohtsubo, Y.O., Nagata,
522 Y.U.J.I., Fujiyama, A.S.A.O., Tsuda, M.A., Kurokawa, K.E.N., 2013. Design and
523 Experimental Application of a Novel Non-Degenerate Universal Primer Set that
524 Amplifies Prokaryotic 16S rRNA Genes with a Low Possibility to Amplify Eukaryotic

525 rRNA Genes. DNA Res. 1–11.
526 Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.,
527 O'Hara, R.B., Simpson, G., Solymos, P., Stevens, H., Szoecs, E., Wagner, H., 2016.
528 Vegan: community ecology package.
529 Orians, G.H., Milewski, A. V., 2007. Ecology of Australia: The effects of nutrient-poor soils
530 and intense fires. Biol. Rev. 82, 393–423. https://doi.org/10.1111/j.1469-
531 185X.2007.00017.x
532 Rayment, G., Lyons, D., 2010. Soil Chemical Methods - Australia. CSIRO Publishing.
533 Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open
534 source tool for metagenomics. PeerJ 4, 1–22.
535 https://doi.org/10.7287/peerj.preprints.2409v1
536 Searle, P.L., 1984. The berthelot or indophenol reaction and its use in the analytical
537 chemistry of nitrogen. A review. Analyst 109, 549–568.
538 Singer, M.E., Finnerty, W.R., 1984. Microbial metaboism of straight chain and branched
539 alkanes, in: Atlas, R.M. (Ed.), Petroleum Microbiology. Macmillan Publishers Limited,
540 pp. 1–61.
541 Spain, A. V., Tibbett, M., Hinz, D.A., Ludwig, J.A., Tongway, D.J., 2015. The mining
542 restoration system and ecosystem development following bauxite mining in a biodiverse
543 environment of the seasonally dry tropics, Northern Territory, Australia, in: Tibbett, M.
544 (Ed.), Mining in Ecologically Sensitive Landscapes. CRC Press, Netherlands, pp. 159–
545 227.
546 Sutton, N.B., Maphosa, F., Morillo, J.A., Al-Soud, W.A., Langenhoff, A.A.M., Grotenhuis,
547 T., Rijnaarts, H.H.M., Smidt, H., 2013. Impact of long-term diesel contamination on soil
548 microbial community structure. Appl. Environ. Microbiol. 79, 619–630.
549 https://doi.org/10.1128/AEM.02747-12

Tibbett, M., George, S.J., Davie, A., Barron, A., Milton, N., Greenwood, P.F., 2011. Just add water and salt: The optimisation of petrogenic hydrocarbon biodegradation in soils from semi-arid Barrow Island, Western Australia. Water. Air. Soil Pollut. 216, 513–525. https://doi.org/10.1007/s11270-010-0549-z

Tischer, K., Kleinsteuber, S., Schleinitz, K.M., Fetzer, I., Spott, O., Stange, F., Lohse, U., Franz, J., Neumann, F., Gerling, S., Schmidt, C., Hasselwander, E., Harms, H., Wendeberg, A., 2013. Microbial communities along biogeochemical gradients in a hydrocarbon-contaminated aquifer 15, 2603–2615. https://doi.org/10.1111/1462-2920.12168

Urakawa, H., Garcia, J.C., Barreto, P.D., Molina, G.A., Barreto, J.C., 2012. Rapid communication A sensitive crude oil bioassay indicates that oil spills potentially induce a change of major nitrifying prokaryotes from the Archaea to the Bacteria. Environ. Pollut. 164, 42–45. https://doi.org/10.1016/j.envpol.2012.01.009

Westlake, D.W.S., Jobson, A.M., Cook, F.D., 1978. In situ degradation of oil in a soil of the boreal region of the Northwest Territories. Can. J. Microbiol. 24, 254–260. https://doi.org/10.1139/m78-044

Westoby, M., 1988. Comparing Australian Ecosystems to Those Elsewhere. Bioscience 38, 549–556. https://doi.org/10.2307/1310763

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. PCR Protocols: a guide to methods and applications. Academic Press, New York.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620. https://doi.org/10.1093/bioinformatics/btt593

Figures and Tables

GRAPHICAL ABSTRACT

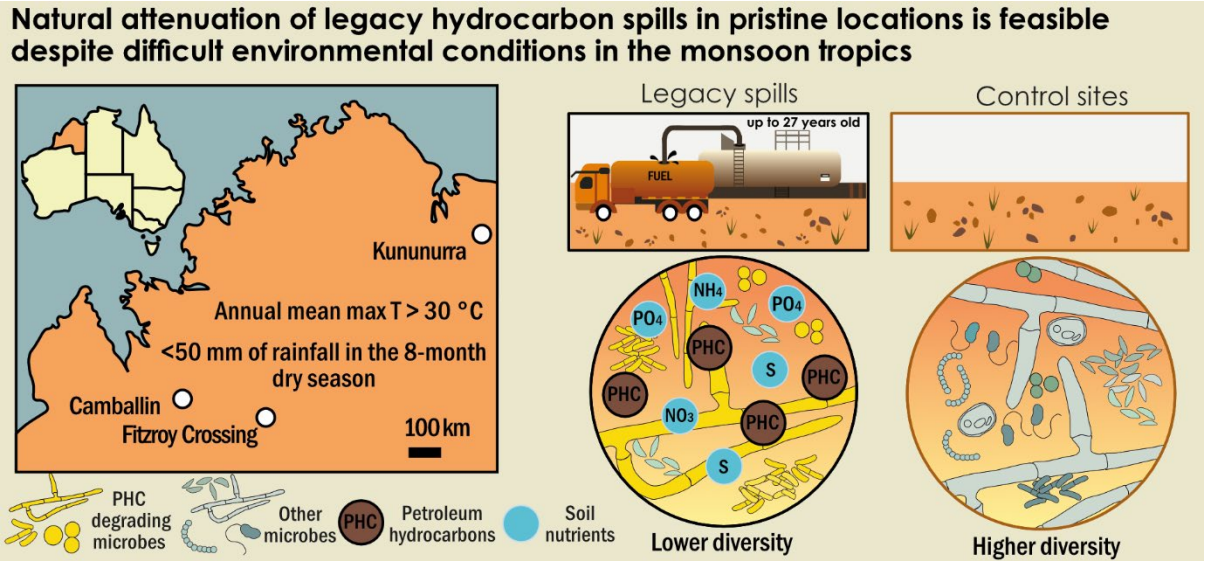


Figure 1. (A) Location of sites within the Kimberley region in Western Australia, Australia. (B) Monthly mean maximum temperature and mean monthly rainfall for the three sites within the Kimberley region. FC = Fitzroy Crossing. Data are means from the years 1991 to 2020 taken from http://www.bom.gov.au/climate/data.

593 Table 1. Mean recoverable hydrocarbons (± 1 SE) from soil cores at contaminated areas within
594 each site. Letters indicate significant (<0.05) post-hoc differences among sites for each
595 hydrocarbon parameter.

Hydrocarbons (mg kg^{-1})	Camballin	Fitzroy Crossing	Kununurra
TPH	280 (39) ^{A}	2725 (608) ^{B}	2229 (542) ^{B}
C_{10}-C_{16}	0 (0)	22 (9)	12 (6)
C_{16}-C_{34}	223 (22) ^{A}	1873 (472) ^{B}	1351 (356) ^{B}
C_{34}-C_{40}	58 (18) ^{A}	830 (131) ^{B}	866 (183) ^{B}
Extractable organic matter (EOM) (mg g^{-1})	0.84 (0.06) ^{A}	6.19 (1.10) ^{B}	9.81 (2.30) ^{B}

596
597
598
599
600
601
602
603
604
605
606
607
608

609 **Table 2.** Mean soil properties (± 1 SE) by site and Total Petroleum Hydrogen (TPH) detection.
610 Uppercase letters indicate significant differences among sites for each TPH detected category,
611 whilst lowercase letters indicate significant differences between TPH detection within each
612 site.

Soil property (mg kg^{-1})	TPH detected	Camballin	Fitzroy Crossing	Kununurra
NH_4-N	No	3.0 (0.2)	1.9 (0.1)	$2.3\ (0.1)^{a}$
	Yes	$2.5\ (0.2)^{A}$	$1.9\ (0.2)^{A}$	$7.3\ (1.0)^{Bb}$
NO_3-N	No	$0.07\ (0.04)^{A}$	$0.30\ (0.12)^{A}$	$1.33\ (0.31)^{Ba}$
	Yes	$0\ (0)^{A}$	$0.63\ (0.18)^{B}$	$3.13\ (0.26)^{Cb}$
Colwell P	No	$5.1\ (0.7)^{Aa}$	$2.5\ (0.2)^{Ba}$	$8.6\ (1.1)^{Ca}$
	Yes	$9.5\ (1.5)^{Ab}$	$6.0\ (0.5)^{Ab}$	$17.5\ (1.3)^{Bb}$
S	No	$4.1\ (0.3)^{A}$	$1.1\ (0.0)^{Ba}$	$4.5\ (1.3)^{Aa}$
	Yes	$2.9\ (0.3)^{A}$	$2.8\ (0.1)^{Ab}$	$23.1\ (1.8)^{Bb}$
Colwell K	No	$90\ (11)^{Aa}$	$68\ (3)^{A/B}$	$56\ (5)^{B}$
	Yes	$42\ (2)^{Ab}$	$85\ (5)^{B}$	$66\ (5)^{B}$
pH ($CaCl_2$)	No	$6.6\ (0.1)^{a}$	$6.8\ (0.1)^{a}$	6.4 (0.1)
	Yes	$6.1\ (0.1)^{b}$	$6.3\ (0.2)^{b}$	6.0 (0.2)
Moisture (%)	No	$1.6\ (0.1)^{A}$	$1.0\ (0.1)^{B}$	$1.6\ (0.1)^{Aa}$
	Yes	$1.5\ (0.1)^{A}$	$0.9\ (0.1)^{A}$	$2.4\ (0.1)^{Bb}$

613
614
615
616
617
618

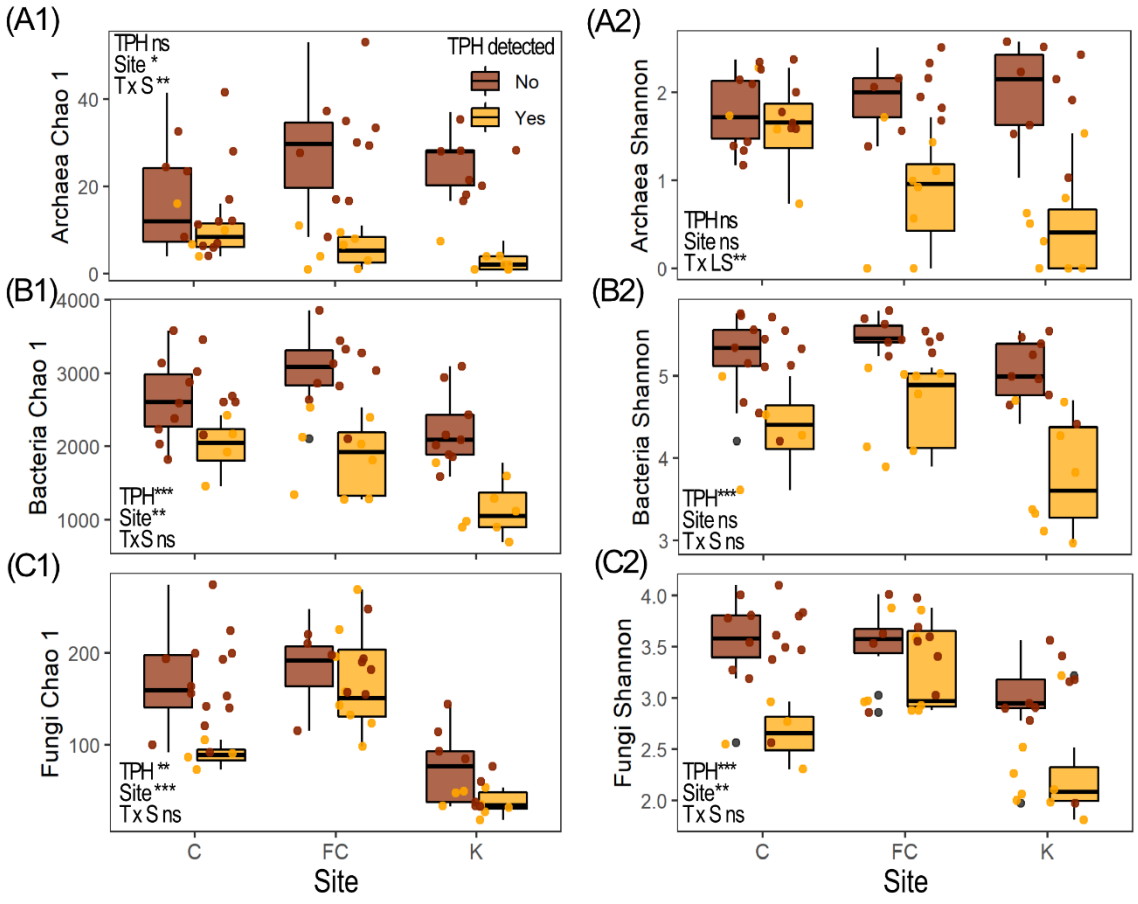
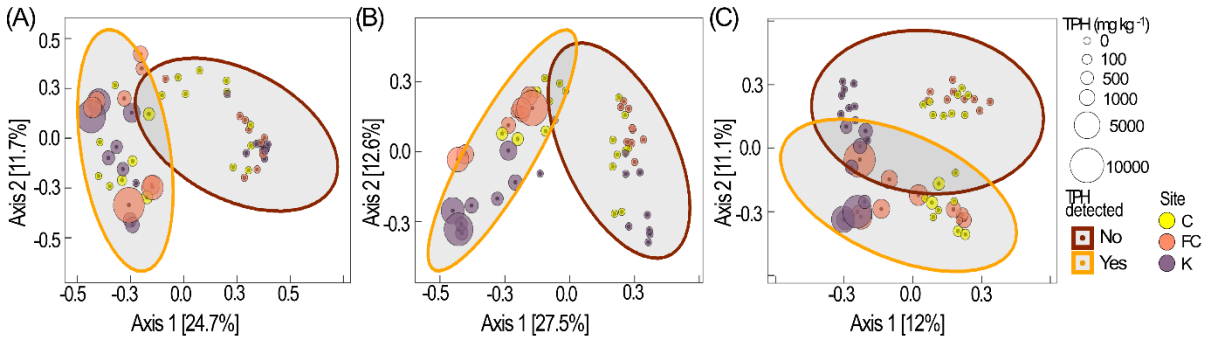


Figure 2. Boxplots showing median and interquartile range (with data points overlain) of estimated alpha diversity (Chao 1 and Shannon) of (A1-A2) archaea (B1-B2) bacteria and (C1-C2) fungi in soils from three sites with (yellow) and without (brown) petroleum hydrocarbons detected. The significance of fixed factors (TPH detected, site, and their interaction (T x S)) are displayed in each plot, where * = 0.05, ** < 0.01, *** < 0.001, and NS= not significant. C = Camballin, FC = Fitzroy Crossing, K = Kununurra. Outliers are shown as grey dots.



632

633

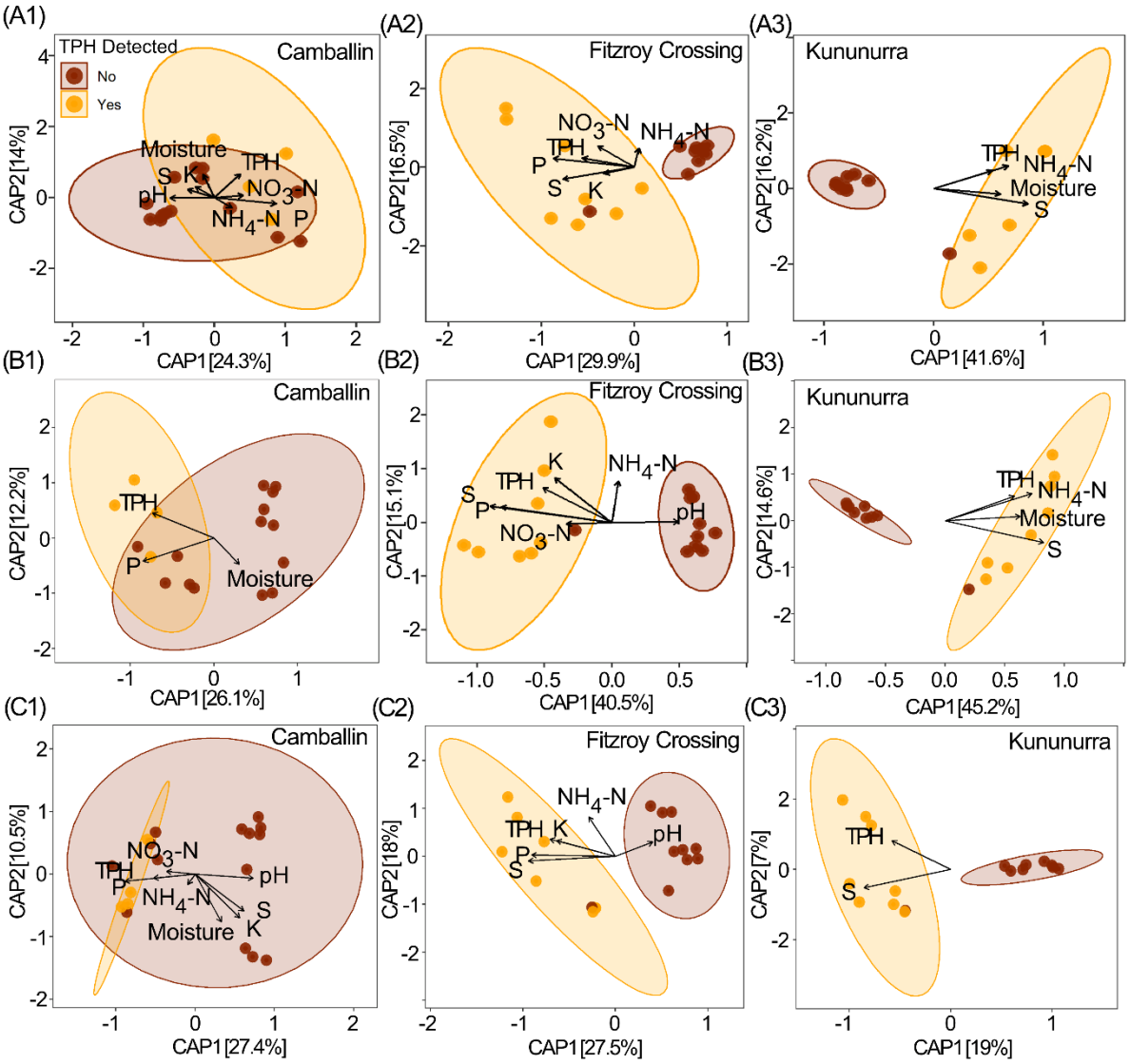
634 **Figure 3.** Principal coordinates analysis (PCoA) of Bray-Curtis distances of (A) archea, (B)
635 bacteria and (C) fungi; 95% confidence ellipses encircle soils with (yellow) and without
636 (brown) petroleum hydrocarbons and are separated by site (C = Camballin, FC = Fitzroy
637 Crossing, K = Kununurra). Size of points corresponds to measured TPH concentration (mg kg-
638 1). For the composition of archaeal soil communities, 6 % of total variation was explained by
639 presence of hydrocarbons, 9 % by site and 7 % by their interaction. For the bacterial soil
640 communities, 11 % of total variation was explained by the presence of hydrocarbons, 13 %
641 explained by site and 7 % explained by the interaction between hydrocarbon presence and site.
642 For the fungal soil communities the presence of hydrocarbons explained 5 % of the total
643 variation, whereas site accounted for 16 % of the total variation and 6 % was explained by the
644 interaction between hydrocarbon presence and site.

645

646

647

648



649

650

651 **Figure 4.** Constrained Analysis of Principal Coordinates (CAP) models of each site constrained
652 by soil properties (~TPH+ Colwell P + S + Colwell K+ NH_4-N + NO_3-N+ pH + Moisture). The
653 final model for each site was optimised using AIC and checking the variance inflation factors
654 and displayed within each panel using vectors for (A1-A3) archaea, (B1-B3) bacteria and
655 (C1—C3) fungi. 95% confidence ellipses encircle soils with (yellow) and without (brown)
656 petroleum hydrocarbons. Outliers present in Fitzroy Crossing and Kununurra were cores
657 collected from the historically contaminated sites that were predicted to contain hydrocarbons,
658 but were found to contain no recoverable hydrocarbons upon analysis.

659

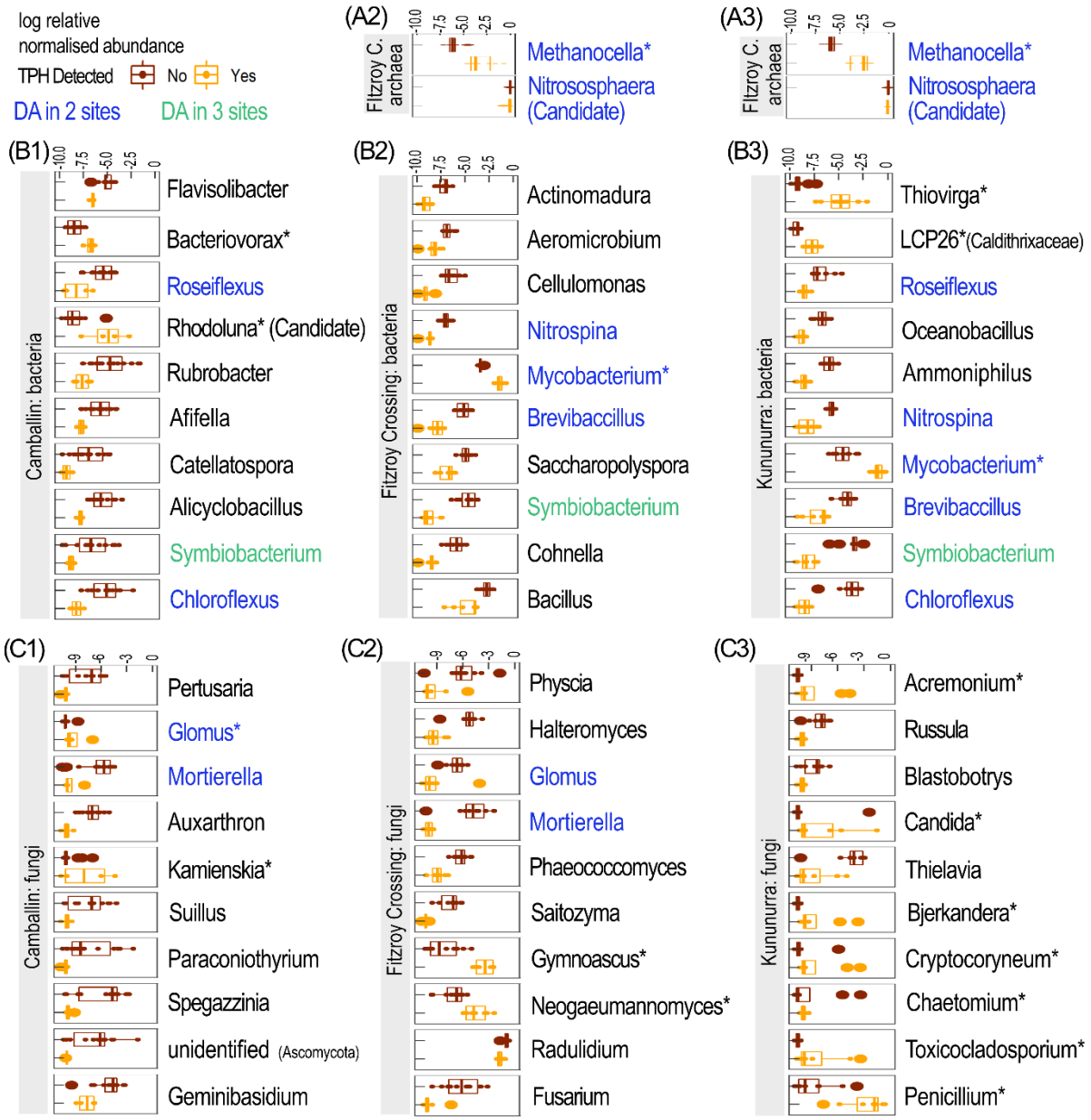


Figure 5. Top ten differentially abundant taxa (at the genus level) between soils with (yellow) and without (brown) petroleum hydrocarbons ($p<0.01$). Analysis was performed for each location and domain separately where A2 to A3 is for archaea, B1 to B3 is for bacteria and C1 to C3 is for fungi. No archaea were significantly different between contaminated and non-contaminated soils at Camballin. An * indicates taxa that were significantly enriched in contaminated soils. The text of OTUs that were differentially abundant (DA; either positive

668 or negative) across two sites are colored blue and those that were shared among all three sites
669 are colored green.