Metagenomic analysis of microbial communities across a transect from low to highly hydrocarbon-contaminated soils in King George Island, Maritime Antarctica

Metagenomic analysis of microbial communities across a transect from low to highly hydrocarbon-contaminated soils in King George Island, Maritime Antarctica

Diogo Jurelevicius*[1], Raphael da Silva Pereira[1], Fabio Faria da Mota[2], Juliano C. Cury[3], Ivan Cardoso de Oliveira[1], Alexandre S. Rosado[1,4], Olivia U. Mason[5], Janet K. Jansson[6], and Lucy Seldin[1]

[1]Instituto de Microbiologia Prof. Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil; ; [2]IOC - Instituto Oswaldo Cruz, Rio de Janeiro, RJ, Brazil; [3]Universidade Federal de São João del-Rei; [4]Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Saudi Arabia; [5]Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, FL, USA; [6]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA.

[*]Corresponding author: Laboratório de Biotecnologia e Ecologia Microbiana, LABEM, Departamento de Microbiologia Geral, Instituto de Microbiologia *Paulo de Góes* (IMPPG), Universidade Federal do Rio de Janeiro, Centro de Ciências da Saúde, Bloco I, Ilha do Fundão, CEP 21941-590, Rio de Janeiro, Brazil.
Phone: 55-21-3938.6741; Fax: 55-21-2560.8344
E-mail: diogoj@micro.ufrj.br

Running title: Microbial communities in hydrocarbon-contaminated Antarctic soils

Declaration of competing interest

The authors declare that they have no competing interests

Acknowledgements

This study was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, financial code 001) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). The authors gratefully acknowledge the Earth Microbiome Program (EMP) for the sequencing data presented in this study.

Abstract

Soil samples from a transect from low to highly hydrocarbon-contaminated soils were collected around the Brazilian Antarctic Station Comandante Ferraz (EACF), located at King George Island, Antarctica. Quantitative PCR (qPCR) analysis of bacterial 16S rRNA genes, 16S rRNA gene (iTag) and shotgun metagenomic sequencing were used to characterize microbial community structure and the potential for petroleum degradation by indigenous microbes. Hydrocarbon contamination did not affect bacterial abundance in EACF soils (bacterial 16S rRNA gene qPCR). However, analysis of 16S rRNA gene sequences revealed a successive change in the microbial community along the pollution gradient. Microbial richness and diversity decreased with the increase of hydrocarbon concentration in EACF soils. The abundance of *Cytophaga*, *Methyloversatilis*, *Polaromonas* and *Williamsia* were positively correlated (p-value = <0.05) with the concentration of total petroleum hydrocarbons (TPH) and/or polycyclic aromatic hydrocarbons (PAH). Annotation of metagenomic data revealed that the most abundant hydrocarbon degradation pathway in EACF soils was related to alkyl derivative-PAH degradation (mainly methylnaphthalenes) via the CYP450 enzyme family. The abundance of genes related to nitrogen fixation increased in EACF soils as the concentration of hydrocarbons increased. The results obtained here are valuable for the future of bioremediation of petroleum hydrocarbon-contaminated soils in polar environments.

Keywords: Antarctic microbiome, hydrocarbon degradation, bioremediation, metagenomic analysis

INTRODUCTION

The use of petroleum-derived fuels by human activity has resulted in the introduction of large quantities of petroleum hydrocarbons into cold regions, including Antarctic soils (Dauner et al., 2015; Martínez Álvarez et al., 2017; Errington et al., 2018). The diesel-fuel (variety "Arctic"; DFA), that is widely used in Antarctic scientific stations as an energy supply is composed of a mixture of aliphatic, aromatic and polyaromatic hydrocarbons (PAH). High concentrations of these compounds are toxic to plant and animals (van Dorst et al., 2014; Errington et al., 2018). However, the clean-up of Antarctic environments often relies on natural attenuation of hydrocarbon-contaminated soils, mainly based on microbial processes (de Jesus et al., 2015a,b) due to the complex logistics of applying physical-chemical and other machine-dependent oil remediation methods (such as biopiles) in Antarctica.

The presence of cold-adapted hydrocarbon-degrading bacteria has previously been demonstrated in Antarctic soils (van Dorst et al., 2014; Vázquez et al., 2017). Hydrocarbon-degrading bacteria related to the genera *Rhodococcus*, *Acinetobacter*, *Pseudomonas*, and *Sphingomonas* were isolated from contaminated and uncontaminated Antarctic soils (Aislabie et al., 2006; de Jesus et al., 2015a). Cury et al. (2014) used 16S rRNA gene cloning and sequencing to demonstrate the difference between the composition of the microbial communities in high- and low-contaminated soils. Functional genes related to hydrocarbon degradation (such as alkane monooxygenase *alkB* and/or genes coding for aromatic ring cleaving dioxygenases) were also previously detected in Antarctic soils (Powell et al., 2010; Guibert et al., 2012; Jurelevicius et al., 2013; Guibert et al., 2016). However, determining the potential of the whole microbial community through a shotgun metagenomic approach has not been carried out so far.

Thus, the hydrocarbon-degradation potential of the microbial communities in Antarctic soils is largely unknown.

Bioremediation (*in situ*) of contaminated Antarctic soils is rarely successful (McWatters et al., 2016). Despite the presence of hydrocarbon degrading bacteria, petroleum hydrocarbons accumulate for long periods in contaminated-Antarctic soils (Cury et al., 2014; de Jesus et al., 2015a; Martínez Álvarez et al., 2017; Vázquez et al., 2017). As shown by Atlas and Hazen (2011) and Aislabie et al. (2012), microorganisms require elements other than carbon for growth and the availability of N, P and K can limit rates of oil biodegradation in polar environments. It is therefore essential to gain a better understanding of the complex interactions between soil physical-chemical proprieties and microbial community distribution and metabolism in order to optimise *in situ* bioremediation of oil-contaminated Antarctic soils.

Here we hypothesized that the composition of the microbial communities and their potential to degrade hydrocarbons vary from low- to high-contaminated Antarctic soils, and are modulated by the distribution of hydrocarbon fractions and nutrient availability. To test these hypotheses, soil samples previously collected and described by Cury et al. (2014) were used here to determine the composition and the potential functions of the microbial communities present across the hydrocarbon contaminated soil transect, and to correlate the microbial data to the hydrocarbon distribution in EACF soils. We combined molecular analyses (qPCR, and 16S rRNA gene and metagenome sequencing) with soil physicochemical data to evaluate: (i) the presence and distribution of compounds (mainly N, P and K) that limit microbial hydrocarbon degradation; (ii) the pattern of hydrocarbon distribution in the contaminated soils; (iii) how contamination modulates the microbial community; and (iv) the presence and distribution of microbial metabolic pathways. Altogether, the data presented here

contribute to the improvement of *in situ* bioremediation of oil-contaminated Antarctic soils.

MATERIALS AND METHODS

Sampling sites

This study was performed with samples collected at the Brazilian Antarctic Station Comandante Ferraz (EACF, 62˚ 05′ S, 058˚ 23.5′ W), King George Island, Antarctic Peninsula, which is part of the South Shetlands archipelago in Maritime Antarctica. All necessary permits were obtained for the field studies described here. In 1989, a spill of 20,000 L of diesel-fuel Arctic (DFA) occurred from a 380-ton DFA storage tank at EACF. It contaminated an area of approximately 100 m in length. After more than 30 years, the soils around EACF showed staining, retained an odor characteristic of DFA, and was shown to have high concentrations of petroleum hydrocarbons (de Jesus et al., 2015a).

Samples were collected in triplicate at five points along a transect starting from low contaminated soils (sE and sD) toward highly hydrocarbon-contaminated soils (sC, sB and sA) near the DFA-storage tanks (500 g of soils per sample, at 0-50 cm depth) as previously described in Cury et al. (2014). At the time of soil collection, the average air temperature in this area was 1.1°C, with an average daily maximum of 3.3°C and an average daily minimum of -1.0°C (means for March 1986–2010, http://antartica.cptec.inpe.br). All collected samples were stored at -80˚C until further analysis.

Determination of abiotic parameters

The physicochemical properties of the soil samples and the quantification of Total

Petroleum Hydrocarbon have previously been characterized (Cury et al., 2014). The soils showed different concentrations of TPH, ranging from 81,941 µg g^{-1} of TPH in sE soil to 426,404 ng g^{-1} in soil sA (Cury et al., 2014). Additionally, we determined the distribution of the aliphatic and aromatic hydrocarbon concentrations in each soil sample. To determine the hydrocarbon concentrations and their fingerprints in the soils sampled, each triplicates were pooled, mixed and oven-dried at 50°C for 4–5 days. Hydrocarbons were extracted from each sample by mixing 25 g of soil with a 50% mixture of residue grade n-hexane and dichloromethane for 8 h. Each hydrocarbon extract was fractionated into F1 (aliphatic) and F2 (aromatic) using a silica gel-alumina chromatography column. Aliphatic hydrocarbons were determined on a Hewlett Packard 5890 II high-resolution gas chromatograph equipped with flame ionization detector (GC-FID), whereas aromatic and polycyclic aromatic hydrocarbons (PAH) were quantitatively analyzed using an Agilent 6890 gas chromatograph coupled to a 5973N mass spectrometer (GC-MS) in the selected ion monitoring (SIM) mode (Bícego et al., 2006).

DNA extraction

Total microbial DNA was extracted and purified from soil using the standard methods described by the Earth Microbiome Program (http://www.earthmicrobiome.org/protocols-and-standards/dna-extraction-protocol/). Briefly, DNA was extracted from 0.25 g of soil using the 96-well format MoBio Powersoil DNA kit on an EpMotion 5075 robot with vacuum (Eppendorf). The quantity and purity of extracted DNA were determined using a NanoDrop 1000 (Thermo Scientific, Suwanee, GA, USA).

Quantitative real-time PCR (q-PCR)

The abundance of the prokaryotic communities (based on the 16S rRNA-encoding genes) was quantified by q-PCR using primers 341F (5'CTACGGGAGGCAGCAG3') and 534R (5'ATTACCGCGGCTGCTGG3'), following previously described qPCR conditions (Lammel et al., 2015). Quantification was carried out in duplicate for each sample on a Rotor Gene 6000 (Corbett Life Science, Australia). The qPCR reaction contained 1 μL DNA, 0.48 μL (0.4 μM) of each primer, 6 μL of GoTaq qPCR Master Mix (2×) and nuclease-free water to the final volume of 12 μL. The following protocol (fast setting) was used: 95 °C for 20 s; 40 cycles of 95 °C for 3 s, 55 °C for 20 s and 72 °C for 45 s; 95 °C for 1 s, 60 °C for 20 s and 95 °C for 1 s (melting curve analysis). Fluorescence was read during the elongation step of each cycle. Specificity of the amplification products was confirmed by melting-curve analysis, and the expected sizes of the amplified fragments were assessed in 1.5% agarose gels stained with ethidium bromide. Standard curves were obtained using serial dilutions (10^7 to 10^1 gene copies μl^{-1}) of specific plasmid-containing cloned fragments. In all cases, significant differences between q-PCR values obtained from soil samples was determined using Tukey's pairwise comparisons (p-value = <0.05).

16S rRNA gene sequencing and statistical analyses

The V4 region of bacterial and archaeal 16S rRNA gene were PCR amplified using the primers 515F and 806R described by Caporaso et al. (2011, 2012) and the conditions standardized by the Earth Microbiome Project (http://www.earthmicrobiome.org/emp-standard-protocols/16s/) as described in the Earth Microbiome Project (http://www.earthmicrobiome.org/emp-standard-protocols/16S/) and in Thompson et al. (2017). Sequencing was carried out using the MiSeq (Illumina, San Diego, CA, USA)

platform. The sequences were demultiplexed and analyzed using the QIIME2 version 2020.2 pipeline (Bolyen et al., 2019). Briefly, Deblur (Amir et al., 2017) was used to control the quality of sequences and to construct a feature table containing the distribution of ASVs (amplicon variants). Multiple sequence alignment of ASVs and phylogenetic trees were obtained using mafft and FastTree programs, respectively. The feature table was rarified to 161,000 sequences for each sample and together with the phylogenetic tree were used for the analyses of alpha and beta diversities.

Alpha diversity were estimated using the observed ASVs counts, the Shannon diversity index (Shannon & Weaver, 1949) and Faith's Phylogenetic Diversity (Faith & Baker, 2007). The values of alpha diversity of each sample were compared using Kruskal-Wallis statistical analysis. Beta diversity was compared using Non Metrical Multidimensional Scaling (NMDS) with Bray-Curtis distance. The difference in microbial composition present in each sample was statistically compared using PERMANOVA (Anderson, 2001). Both NMDS and PERMANOVA analyses were performed in PAST 4.03 software (Hammer et al., 2001).

Taxonomic composition of the samples

The ASVs were assigned to a taxonomy using Greengenes 13_8 database (McDonald et al., 2012) in pre-trained Naive Bayes classifier and the q2-feature-classifier plugin. The absolute abundances of the microbial groups present in each sample were estimated as described in Jian et al. (2020). Briefly, the values obtained from microbial q-PCR were considered to represent the number of microbial cells in each sample. The values of relative abundance of each microbial phyla, classes and genera (based on sequencing data) were considered as the fraction of the represented groups in each sample.

Microbial community functional predictions: Metagenomic library preparation, sequencing and quality assessment

Metagenomic shotgun sequencing libraries were prepared and sequenced at Argonne National Laboratory. For each sample, 1 pg of genomic DNA was used for Illumina's TruSeq library preparation. Libraries were sequenced using the Illumina HiSeq 2 × 100-bp paired-end technology. Sequences that had 5% bp with phred scores less than or equal to 10 were filtered out before bioinformatics analyses. The Ilumina reads were compared to the reference NCBI-nr database (NCBI) using blastx (e-value threshold 10^{-5}). The files of reads, the resulting blastx files and the protein accession to KEGG mapping file were imported into MEGAN5 (Huson & Weber, 2013). The functional classification was automatically calculated by MEGAN5 based on KEGG classification. Further, to analyze the same number of reads per sample, the KEGG orthologies (KOs) counts were rarefied based on the lowest number of metagenomic reads obtained from one sample. The Enzyme Commission (EC) numbers were used to calculate the abundance of reads among the samples related to each enzymatic function.

RESULTS AND DISCUSSION

More than 30 years after the great diesel oil spill at the Brazilian Antarctic Station Comandante Ferraz (EACF), we determined that hydrocarbons have persisted in soil samples from contaminated areas (Cury et al., 2014). The aliphatic and aromatic hydrocarbon fractions differed along the contamination transect. The similarity index of soils based on the distribution of hydrocarbon patterns progressively decreased when highly contaminated soil, sA, was compared to less contaminated soils, sB, sC, sD and sE, respectively (Table S1, Fig. S1). An unresolved complex mixture (UCM) of hydrocarbons comprised 32% (sB) to 58% (sD) of the TPH quantified in Antarctic soils

245 (Table 1, Table S2). The highest concentration of resolved aliphatic hydrocarbons
246 (rAH) was detected in sA (163,000 ng g^{-1}). The concentration of rAH progressively
247 decreased from sA to sB (104,520 ng g^{-1}), sC (38,940 ng g^{-1}), sE (29,040 ng g^{-1}) and sD
248 (28,440 ng g^{-1}) (Table 1, Table S2). These values were highly similar to the previously
249 reported TPH distribution in EACF soils (Cury et al., 2014). Although the highest TPH
250 (Cury et al., 2014) and rAH values were obtained in sA, the concentration of alkanes
251 ranged from n-C12 to n-C35 (including pristane and phytane) was higher in sB
252 (116,301 ng g^{-1}) than in sA (101,292 ng g^{-1}). Moreover, the concentration of PAH was
253 higher in soils sB and sC (57,884 and 43,322 ng g^{-1}, respectively) than in sA (41,321 ng
254 g^{-1}) (Table 1 and Table S2).
255 The chain length of the remaining n-alkanes present in EACF soils ranged from
256 n-C12 to n-C33 in sA, n-C12 to n-C35 in soil sB, n-C12 to n-C31 in sC, n-C12 to n-C24
257 in sD and n-C12 to n-C26 in sE (Table 1). In highly contaminated soils, the more
258 frequent n-alkanes were related to n-C17 followed by n-C16 and n-C18 in sB, and n-
259 C17 followed by n-C16 and n-C15 in sA (Table 1). Among aromatic hydrocarbons,
260 unsubstituted polycyclic aromatic hydrocarbons (PAH) and alkyl derivatives PAH were
261 comprised of compounds with 2–5 rings in almost all samples (Table 1). However,
262 alkyl derivatives of PAH, which are less soluble in water and harder to degrade than
263 unsubstituted PAH (Seo et al., 2009; Dauner et al., 2015; Prus et al., 2015), were the
264 main aromatic compounds detected in Antarctic soils (representing 98 to 99% of PAH
265 found in the soils analyzed here). In high TPH-content soils sA and sB, the main alkyl
266 PAH derivatives were dimethylnaphathalene followed by trimethylnaphathalene and
267 methylnaphthalene (sB) or methylfluorene (sA). In soils sC, sD and sE the main alkyl
268 PAH derivatives were trimethylnaphathalene followed by dimethylnaphathalene and
269 methylnaphthalene (Table 1). There is an order of hydrocarbon removal by microbial

degradation based on the different resistances of hydrocarbons to biodegradation: lighter n-alkanes and unsubstituted PAH are more easily degraded than heavier n-alkanes and alkyl PAH derivatives (Abbasian et al., 2015). This makes it possible to assess the petroleum biodegradation stage in contaminated soils. Interestingly, the n-alkanes found after 30 years in the highly contaminated soils were predominantly short-chain n-alkanes (from n-C16 to n-C18). The predominance of short-chain n-alkanes is typical for recently contaminated soils, not in chronically, long-term contaminated soils (Abbasian et al., 2015). On the other hand, the prevalence of alkyl derivatives (mainly tri- and dimethylnaphthalenes) over methylnaphthalene and unsubstituted PAH suggests the faster modification of aromatic hydrocarbons than of aliphatic hydrocarbons in EACF soils.

The high concentration of TPH detected at the EACF site, discussed above, is not indicative of no biodegradation of hydrocarbons in these soils. As suggested by Prus et al. (2015), microbial degradation of oil contaminants in Antarctica occurs predominantly in the warmer summer season (which comprises only a period of three months/year) and mainly in soils whose microbial communities have previously been enriched by hydrocarbon selection. Eriksson et al. (2001) and Børresen et al. (2007) suggested that freeze–thaw cycles could enhance bioremediation by enhancing specific hydrocarbon degraders in contaminated soils. Therefore, we expected that after long time exposure of the soil microbiome to DFA oil that the microbial community in highly contaminated EACF soils would be primarily comprised of hydrocarbon-degrading specialists.

The structure of microbial community in EACF soils

Bacterial quantification using qPCR analysis showed that the difference in TPH

concentration and in hydrocarbon distribution in EACF did not significantly affect the bacterial abundances present in these soils. Bacterial abundances estimated based on the quantification of 16S rRNA coding gene varied from 7.3 x 10^8 cell g^{-1} in sC to 4.4 x 10^8 cell g^{-1} in sE, Fig. 1A). These results contrast with those of Cury et al. (2014) who used culture-based analyses to show that the total heterotrophic bacteria and hydrocarbon-degrading bacterial populations were unevenly distributed between high- and low-contaminated EACF soils. However, EACF soils containing different hydrocarbon fractions did select for different microbial community compositions and functions when taxonomic analysis of microbial community was evaluated by 16 rRNA gene and metagenomic sequence analyses, which is elaborated on below.

Analysis of 16S rRNA gene sequencing revealed that microbial richness (based on ASVs counts) decreases when comparing soils with low levels of contamination (sD, sE and sC) to highly contaminated soils (sB and sA) (data not shown). The phylogenetic diversity (Faith's-PD index) and diversity (Shannon index) decreased significantly in the highly contaminated soils (p-values = <0.05) (Fig. 1B, 1C). The effect of contamination on microbial structure was confirmed by NMDS ordination of sC, sD and sE soils, that showed a progressive separation from sB and sA soils (Fig. 1D). Similar results were observed analyzing the distance matrix of metagenome annotation (Fig. 1E). Therefore, even with no impact on microbial biomass across the petroleum contamination spectrum in the Antarctic soils analyzed herein, the difference in microbial structure, composition and function, as we observed here, are generally the result of the hydrocarbon contamination in Antarctic soils. Similar results were obtained by Powell et al. (2010), Muangchinda et al. (2015) and Vázquez et al. (2017).

Microbial composition in EACF soils

320 The number of microbial phyla in EACF soils decreased from 14 in sD and sE soils, to
321 9, 10 and 11 in sB, sC and sA, respectively (Fig. 2). The absolute abundance
322 (normalized based on q-PCR values) of the most dominant phyla showed trends
323 according to the degree of contamination (the relative abundances of the most dominant
324 prokaryotic groups present in Antarctic soils are shown in Fig. S2). Proteobacteria
325 increased in abundance as contamination increased (from 2.1 x 10^8 in sE to 4.4 x 10^8 in
326 sA), mainly caused by the increase of betaproteobacterial class (1.6 x 10^8 in sE to 2.7 x
327 10^8 in sA). Specifically, all soils were dominated by members of the Betaproteobacteria
328 class (Fig. 2). In addition, in soils with higher TPH content (sA), Bacteroidetes and
329 Alphaproteobacteria (1.8 x 10^8 and 8.9 x 10^7, respectively) were the most abundant
330 microbial groups (Fig. 2).
331 In the sB and sC soils the absolute abundance of Betaproteobacteria was high
332 (1.9 x 10^8 and 2.0 x 10^8, respectively), followed by Actinobacteria (1.3 x 10^8 and 1.4 x
333 10^8, respectively) and Bacteroidetes (9.2 x 10^7 and 1.1 x 10^8) (Fig. 2). The
334 betaproteobacterial abundance was also high in sD and sE soils (1.2 x 10^8 in sD and sE)
335 but was followed by Actinobacteria (8.9 x 10^7) and Bacteroidetes (5.7 x 10^7) (in soil sD)
336 and by Gammaproteobacteria (7.0 x 10^7) and Bacteroidetes (5.8 x 10^7) in soil sE. The
337 dominance of Proteobacteria, and in particular the increase of Betaproteobacteria in our
338 more contaminated samples is consistent with the findings of Vázquez et al. (2017) in
339 contaminated soils from the southwestern region of King George Island, South Shetland
340 Islands.
341 The difference in microbial community compositions between soils with the
342 lowest levels of hydrocarbons and the most contaminated soils was primarily due to the
343 high relative abundance of Betaproteobacteria from the *Methyloversatilis* genus, which
344 increased from 1.6 x 10^7 in sD to 1.8 x 10^8 in sA paralleling the increase in TPH content

345 in EACF soils (Table 2). In addition, the absolute abundance of ASVs related to
346 *Cytophaga* genus (Bacteroidetes) and to unidentified genus from Opituaceae family
347 (Verrucomicrobia) were high in sA (8.4 x 10^7 and 4.0 x 10^7, respectively) and low to
348 undetected in all other samples (*Cytophaga* abundance ranged from undetected in sB to
349 2.0x10^6 in sD). Both *Methyloversatilis* and unidentified genus from Opituaceae family
350 were positively correlated with TPH concentration in EACF soils (r = 0.7 for both
351 ASVs, p-value = < 0.1) (Table 2). Members of Opituaceae family are not known
352 hydrocarbon degraders, yet *Cytophaga* and *Methyloversatilis* have previously been
353 described as aromatic and PAH degraders (Viñas et al., 2005; Smalley et al., 2015;
354 Rochman et al., 2017). In addition, the genome of *Methyloversatilis discipulorum*,
355 described in Smalley et al. (2015), showed the presence of different genes related to
356 aromatic and PAH degradation, such as genes coding for ring-hydroxylating
357 dioxygenases and catechol 2,3 dioxygenase. Thus, our data is consistent with the
358 metabolism of the known hydrocarbon degrading microorganisms that were dominant
359 in our samples. We also found that previously unrecognized hydrocarbon degraders
360 were abundant in contaminated soils.
361 The concentration of PAH, as well as the toxic effects of total TPH, can
362 modulate the distribution of hydrocarbon degrading microbes in EACF soils. For
363 example, in contrast with the results obtained in soil sA, soils sB and sC showed an
364 increase in abundance *Polaromonas* (7.5 x 10^7 and 8.1 x 10^7, respectively), *Williamsia*
365 (8.6 x 10^7 and 4.1 x 10^7) and unidentified genera from Flavobacteriaceae (4.4 x 10^7 and
366 3.0 x 10^7). While ASVs related to *Polaromonas* were only positively correlated with
367 PAH in EACF soils (r = 0.7, p-value = < 0.5), ASVs related to *Williamsia* and to
368 unidentified genera from Flavobacteriaceae were positively correlated to both PAH and
369 TPH concentration in EACF soil. For example, ASVs related to *Williamsia* were more

correlated with PAH (r = 0.9, p-value = < 0.1), but was also positively correlated to TPH (r = 0.5, p-value = < 0.05) in EACF soils (Table 2). Both *Polaromonas* and *Williamsia* genera have been more described in environments contaminated with PAH than with alkanes. Members of the *Polaromonas* genus are well known for their ability to degrade PAH compounds, including high molecular weight PAH (Yagi et al., 2009, Hanson et al., 2012). Although hydrocarbon-degrading bacteria that were predominant in hydrocarbon-contaminated EACF soils have not yet been isolated from Antarctic soils, clones related to *Polaromonas* were predominant in soils containing up to 1,095 ng g^{-1} PAH from the southwestern region of King George Island (Vázquez et al., 2017). The prevalence of specialized aromatic hydrocarbon-degrading bacteria over alkane-degrading bacteria in EACF soils is consistent with the distribution of the different alkanes and PAH compounds in these soils (as discussed before). Therefore, the results suggest a high potential for microbial degradation of aromatic compounds in EACF soils.

Metagenomic functional analysis of microbial community

Here, we used metagenomic analysis to understand how contamination modulates the distribution of hydrocarbon-degrading pathways in microbes present in Antarctic soils. Metagenomic data were rarefied to a total of 1,011,733 reads per soil (considering the lowest number of reads obtained from EAFC soils). In the raw, unassembled, normalized metagenomic sequence data, genes coding for aerobic and anaerobic pathways of aromatic, PAH and aliphatic hydrocarbon degradation were found in all soils. The most predominant genes coded for (i) phenol hydroxylase (benzene degradation), (ii) enzymes from CYP450 family (phenanthrene 1,2-monooxygenase and 1-methylnaphthalene hydroxylase, PAH and alkyl PAH derivatives degradation), (iii)

benzoate 1,2-dioxygenase (benzoate degradation), (iv) alkane-monooxigenase AlkB (alkanes, including cycloalkanes), (v) biphenyl 2,3-dioxygenase (biphenyl) and (vi) benzoate coenzyme A (anaerobic degradation of aromatic compounds) (Fig. 3, Table S3). Reads coding for CYP450 related to PAH degradation were particularly abundant in sB and sC with the lowest number of these genes observed in soil sA (Fig. 3). Alkane-monooxygenase (AlkB) were most abundant in sA, while the lowest number was observed in sB. Interestingly, genes coding for PAH ring-hydroxylating dioxygenases (such as PAH-RHDs coded by *nidA*, *nahA*, *nirA*) were underrepresented in our metagenomic data (Fig. 3). To confirm these results, we carried out blastx analysis comparing sequences to enzymes coding for PAH-RHDs that were previously described in Antarctic soils (Flocco et al., 2009; Jurelevicius et al., 2012). None of the previously described Antarctic PAH-RHDs was found in the metagenomic data (data not shown). Although Jurelevicius et al. (2012) and Flocco et al. (2009) have shown the presence of PAH-RHDs in contaminated soils, these genes were sparsely detected in both studies and, in some samples, these genes were below the limit of detection for PCR amplification. Therefore, the results obtained here suggest that the remaining PAH alkyl derivatives found in Antarctic soils are mainly metabolized by CYP450 pathways.

The main PAH degradation pathway detected in EACF soils included genes from the methylnaphthalene hydroxylases (from CYP450 family), which convert, for example, 1-methylnaphthalene and 2-methylnaphthalene to 1-hydroxymethyl-naphthalene and 2-hydroxymethyl-naphthalene, respectively (Fig. 4). Further, the majority of the genes coding for the conversion of 1-hydroxymethyl-naphthalene into catechol (central compound of most pathways of aromatic hydrocarbon degradation) were found in all Antarctic soils (Fig. 4). Genes coding for conversion of 2-hydroxymethyl-naphthalene to 4-formylsalicylic acid and/or into 2-naphthoate were

also found (Fig. 4). 2-naphthoate can be further converted to catechol. Furthermore, unsubstituted PAH are likely to be metabolized through the pathway initialized by phenanthrene 1,2-monooxygenase (also from the CYP450 family) which converts phenanthrene to phenanthrene 1,2-oxide and then to 1-phenanthrenol. 1-phenanthrenol can be further converted to 1-methoxyphenanthrene by a methyltransferase (Fig. 4). These pathways for PAH degradation that are based on bacterial CYP450 have been less described and studied than the pathways coded by PAH-RHDs (Flocco et al., 2009; Jurelevicius & Seldin, 2010). Degradation of alkyl derivatives of PAH through methylnaphthalene hydroxylases has been described in bacteria from *Pseudomonas* and *Pseudomonas*-related genera (Seo et al., 2009). On the other hand, the conversion of phenanthrene to 1-methoxyphenanthrene has been described mainly in cyanobacteria (Cerniglia, 1992). Thus, our results showed that the CYP450 pathways for PAH hydrocarbon degradation represent novel, important data regarding the process by which hydrocarbons could potentially be degraded in Antarctic soils by resident microbes.

Despite the presence of hydrocarbon-degrading microorganisms, the limited abundance of nutrients essential to microbial metabolism may have been limiting the complete bioremediation of these contaminated soils (Atlas & Hazen, 2011; Aislabie et al., 2012). For example, Eckford et al. (2002) and Yergeau et al. (2012) reported that essential nutrients needed to support microbial activity (such as N and P) limit the potential for bioremediation in polar environments. As discussed earlier, EACF soils have a high amount of P (Cury et al., 2014), with N likely being the main limiting nutrient in these soils. As the density of microbial biomass is low in most Antarctic soils, the input of nitrogen by the decomposition of organic matter is limited to some specific sites. For example, colonies of penguins, seals and sea lions may contribute to

445 the input of nitrogen (Fritsen et al., 2000; Eckford et al., 2002; Cury et al., 2014).

446 Our results showed that the predominant pathways involved in the nitrogen
447 cycle shifted across the soil samples (Fig. 5A and 5B). The number of reads related to
448 the nitrogenase complex (*nifDKH* operon) increased as the TPH concentration increased
449 in the EACF soils. By contrast, the number of reads of genes coding for nitrification
450 (*nxrAB* and *amoCAB*), assimilatory nitrate reduction (*nasAB*), dissimilatory nitrate
451 reduction (*nirBD*) and for denitrification (*narGHI*, *napAB*, *nirK* and *norBC*) were
452 higher in samples with lower amounts of TPH (Fig. 5A). Therefore, we hypothesize that
453 the microbial community present in soils with higher hydrocarbon contamination levels
454 has adapted over the 30-year period to fix atmospheric N and increase bioavailable N
455 which, in turn, facilitates the growth of aerobic hydrocarbon-degrading bacteria (which
456 outcompete the nitrifiers in these soils) (Fig. 5B). The bioremediation of highly-
457 contaminated EACF soils would then be limited by the activity of diazotrophic bacteria.
458 On the other hand, the soils with lower levels of contamination had the potential for
459 conversion of fixed N to nitrate through nitrification. While the nitrate produced could
460 potentially be converted to ammonium through assimilatory nitrate reduction, we
461 instead found genetic evidence for its use as an electron acceptor for anaerobic
462 respiration through denitrification, which would result in loss of N from the system
463 (Fig. 5B). The potential role of denitrifiers in the degradation of aromatic hydrocarbons
464 in Antarctic soils has also been proposed by Sampaio et al. (2017).

465 As shown before, the increase of TPH and PAH concentrations in Antarctic soils
466 were positively correlated with the abundance of different bacteria, such as bacteria
467 from *Methyloversatilis*, *Cytophaga* and *Polaromonas* genera. Interestingly, genomic
468 analyses have shown the presence of genes coding for the nitrogenase complex in
469 genomes of *M. thermotolerans* and *M. discipulorum* (Smalley et al., 2015), and of

Saccharicrinis fermentans (Inoue et al., 2015) which were reclassified from *Cytophaga fermentans* by Yang et al. (2014). This evidence suggests a possible role for members of the genera *Methyloversatilis* and *Cytophaga* in the biological fixation of nitrogen in the highly contaminated soils of the EACF. Furthermore, *Polaromonas naphthalenivorans*, which was enriched in sB and sC soils, has previously been described as a diazotrophic bacterium (Hanson et al., 2012). Therefore, we predict that the limited availability of N together with hydrocarbon concentrations selected for specialists able to both fix atmospheric nitrogen and to degrade hydrocarbons (Fig. 5B).

In conclusion, our results suggested that bioremediation of low temperature, hydrocarbon contaminated soils may be carried out by N_2-fixing strains that can degrade hydrocarbons *in situ*. Additionally, in contaminated soil, N_2-fixing bacteria can play a role in decreasing the carbon:nitrogen ratio and increasing the rate of bacterial growth and, consequently, the hydrocarbon degradation. N_2-fixing strains have previously been proposed as biofertilizers for stimulating remediation of contaminated soils. Prantera et al. (2002) described two N_2-fixing strains (Alpha and Betaproteobacteria) that were highly effective at degrading benzene, toluene and xylene, showing the potential for application of these strains in the soil environment as biofertilizers. In addition, Hanson et al. (2012) found that nitrogen fixation played an important role in naphthalene biodegradation by *Polaromonas naphthalenivorans.* The prospect for utilization of nitrogen fixing and hydrocarbon degrading bacterial inoculants that we identified from the genera *Methyloversatilis*, *Cytophaga* and *Polaromonas* could be useful for clean-up of contaminated Antarctic soils. Furthermore, all the results obtained here are valuable to the future of bioremediation of petroleum hydrocarbons-contaminated soils in polar environments, and showed the importance of N to microbial degradation of recalcitrant hydrocarbons.

495

496

497

References

Abbasian, F., Lockington, R., Mallavarapu, M. et al. (2015). A comprehensive review of aliphatic hydrocarbon biodegradation by bacteria. Applied Biochemistry and Biotechnology, 176, 670e699.

Aislabie, J., Saul, D.J., Foght, J.M. (2006). Bioremediation of hydrocarbon-contaminated polar soils. Extremophiles, 10, 171-179.

Aislabie, J.M., Ryburn, J., Gutierrez-Zamora, M.L. et al. (2012). Hexadecane mineralization activity in hydrocarbon-contaminated soils of Ross Sea region Antarctica may require nutrients and inoculation. Soil Biology Biochemistry, 45, 49-60.

Amir, A., McDonald, D., Navas-Molina, J.A. et al. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems, 2, e00191-16.

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26 (1), 32-46.

Atlas, R.M., Hazen, T.C. (2011). Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. Environmental Science & Technology, 45, 16, 6709-6715.

Bícego, M.C., Taniguchi, S., Yogui, G.T. et al. (2006). Assessment of contamination by polychlorinated biphenyls and aliphatic and aromatic hydrocarbons in sediments of the Santos and São Vicente Estuary System, São Paulo, Brazil. Marine Pollution Bulletin, 52, 1804-1816.

Bolyen, E., Rideout, J.R., Dillon, M.R. et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37, 852-857.

Børresen, M.H., Barnes, D.L., Rike, A.G. (2007). Repeated freeze–thaw cycles and

523 their effects on mineralization of hexadecane and phenanthrene in cold climate
524 soils. Cold Regions Science and Technology, 49, 215-225.
525 Caporaso, J.G., Lauber, C.L., Walters, W.A. et al. (2012). Ultra-high-throughput
526 microbial community analysis on the Illumina HiSeq and MiSeq platforms. The
527 ISME Journal, 6(8), 1621-1624.
528 Caporaso, J.G., Lauber, C.L., Walters, W.A. et al. (2011). Global patterns of 16S rRNA
529 diversity at a depth of millions of sequences per sample. Proceedings of the
530 National Academy of Sciences, 108, 4516-4522.
531 Cerniglia, C.E. **(**1992). Biodegradation of polycyclic aromatic hydrocarbons.
532 Biodegradation, 3, 351-368.
533 Cury, J.C., Jurelevicius, D.A., Villela, H.D.M. et al. (2014). Microbial diversity and
534 hydrocarbon depletion in low and high diesel-polluted soil samples from Keller
535 Peninsula, South Shetland Islands. Antarctic Science, 27, 1-11.
536 Dauner, A.L., Hernández, E.A., MacCormack, W.P. et al. (2015). Molecular
537 characterisation of anthropogenic sources of sedimentary organic matter from
538 Potter Cove, King George Island, Antarctica. Science of Total Environment,
539 502, 408-416.
540 de Jesus, H.E., Peixoto, R.S., Cury, J.C. et al. (2015a). Evaluation of soil
541 bioremediation techniques in an aged diesel spill at the Antarctic Peninsula.
542 Applied Microbiology and Biotechnology, 99(24), 10815-10827.
543 de Jesus, H.E., Peixoto, R.S., Rosado, A.S. (2015b). Bioremediation in Antarctic Soils.
544 Journal of Petroleum & Environmental Biotechnology, 6, 248.
545 Eckford, R., Cook, F.D., Saul, D. et al. (2002). Free-living heterotrophic nitrogen-fixing
546 bacteria isolated from fuel-contaminated Antarctic soils. Applied and
547 Environmental Microbiology, 68, 5181-5185.

548 Eriksson, M., Ka, J., Mohn, W.W. (2001). Effects of low temperature and freeze-thaw
549 cycles on hydrocarbon biodegradation in Arctic Tundra soil. Applied and
550 Environmental Microbiology, 67(11), 5107-5112.
551 Errington, I., King, C.K., Wilkins, D. et al. (2018). Ecosystem effects and the
552 management of petroleum-contaminated soils on subantarctic islands.
553 Chemosphere, 194, 200-210.
554 Faith, D.P., Baker, A.M. (2007). Phylogenetic diversity (PD) and biodiversity
555 conservation: some bioinformatics challenges. Evolutionary Bioinformatics
556 Online, 2, 121-128.
557 Flocco, C.G., Gomes, N.C., Maccormack, W. et al. (2009). Occurrence and diversity of
558 naphthalene dioxygenase genes in soil microbial communities from the
559 Maritime Antarctic. Environmental Microbiology, 11, 700-714.
560 Fritsen, C.H., Grue, A.M., Priscu, J.C. (2000). Distribution of organic carbon and
561 nitrogen in surface soils in the McMurdo Dry Valleys, Antarctica. Polar
562 Biology, 23, 121-128.
563 Guibert, L.M., Loviso, C.L., Borglin, S. et al. (2016). Diverse bacterial groups
564 contribute to the alkane degradation potential of chronically polluted
565 subantarctic coastal sediments. Microbial Ecology, 71, 100-112.
566 Guibert, L.M., Loviso, C.L., Marcos, M.S. et al. (2012). Alkane biodegradation genes
567 from chronically polluted subantarctic coastal sediments and their shifts in
568 response to oil exposure. Microbial Ecology, 64, 605-616.
569 Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001). PAST: Paleontological statistics
570 software package for education and data analysis. Palaeontologia Electronica,
571 4(1), 9pp.
572 Hanson, B.T., Yagi, J.M., Jeon, C.O. et al. (2012). Role of nitrogen fixation in the

573 autecology of *Polaromonas naphthalenivorans* in contaminated sediments.
574 Environmental Microbiology, 14 (6), 1544-1557.
575 Huson, D.H., Weber, N. (2013). Microbial community analysis using MEGAN.
576 Methods in Enzymology, 531, 465-485.
577 Inoue, J., Oshima, K., Suda, W. et al. (2015). Distribution and evolution of nitrogen
578 fixation genes in the phylum Bacteroidetes. Microbes and Environments, 30(1),
579 44-50.
580 Jian, C., Luukkonen, P., Yki-Järvinen, H. et al. (2020). Quantitative PCR provides a
581 simple and accessible method for quantitative microbiota profiling. PLoS One,
582 15(1), e0227285.
583 Jurelevicius, D.A., Seldin, L. (2010). Bacteria and their genes coding for enzymes
584 involved in bioremediation of petroleum hydrocarbons: Bioremediation:
585 Biotechnology, Engineering and Environmental Management. Mason, A.C.
586 (ed.). New York, USA: Nova Publishers, pp.133-158.
587 Jurelevicius, D., Alvarez, V.M., Peixoto, R. et al. (2013). The use of a combination of
588 *alkB* primers to better characterize the distribution of alkane-degrading bacteria.
589 PLoS One, 8(6), e66565.
590 Jurelevicius, D., Alvarez, V.M., Peixoto, R. et al. (2012). Bacterial polycyclic aromatic
591 hydrocarbon ring-hydroxylating dioxygenases (PAH-RHD) encoding genes in
592 different soils from King George Bay, Antarctic Peninsula. Applied Soil
593 Ecology, 55, 1-9.
594 Lammel, D.R., Nüsslein, K., Tsai, S.M. et al. (2015). Land use, soil and litter chemistry
595 drive bacterial community structures in samples of the rainforest and Cerrado
596 (Brazilian Savannah) biomes in Southern Amazonia. European Journal of Soil
597 Biology, 66, 32-39.

598 Martínez Álvarez, L.M., Ruberto, L., Lo Balbo, A. et al. (2017). Bioremediation of
599 hydrocarbon-contaminated soils in cold regions: Development of a pre-
600 optimized biostimulation biopile-scale field assay in Antarctica. Science of the
601 Total Environment, 590-591, 194-203.
602 McDonald, D., Price, M.N., Goodrich, J. et al. (2012). An improved Greengenes
603 taxonomy with explicit ranks for ecological and evolutionary analyses of
604 bacteria and archaea. The ISME Journal, 6, 610-618.
605 McWatters, R.S., Wilkins, D., Spedding, T. et al. (2016). On site remediation of a fuel
606 spill and soil reuse in Antarctica. Science of the Total Environment, 571, 963-
607 973.
608 Muangchinda, C., Chavanich, S., Viyakarn, V. et al. (2015). Abundance and diversity of
609 functional genes involved in the degradation of aromatic hydrocarbons in
610 Antarctic soils and sediments around Syowa Station. Environmental Science and
611 Pollution Research International, 22(6), 4725-4735.
612 Powell, S., Bowman, J.P., Ferguson, S.H. et al. (2010). The importance of soil
613 characteristics to the structure of alkane-degrading bacterial communities on
614 sub-Antarctic Macquarie Island. Soil Biology and Biochemistry, 42, 2012-2021.
615 Prantera, M., Drozdowicz, A., Leite, S. et al. (2002). Degradation of gasoline aromatic
616 hydrocarbons by two N_2-fixing soil bacteria. Biotechnology Letters, 24, 85-89.
617 Prus, W., Fabiańska, M.J., Łabno, R. (2015). Geochemical markers of soil
618 anthropogenic contaminants in polar scientific stations nearby (Antarctica, King
619 George Island). Science of the Total Environment, 518-519, 266-279.
620 Rochman, F.F., Sheremet, A., Tamas, I. et al. (2017). Benzene and naphthalene
621 degrading bacterial communities in an oil sands tailings pond. Frontiers in
622 Microbiology, 8, 1845.

623 Sampaio, D.S., Almeida, J.R.B., de Jesus, H.E. et al. (2017). Distribution of anaerobic
624 hydrocarbon-degrading bacteria in soils from King George Island, Maritime
625 Antarctica. Microbial Ecology, 74(4), 810-820.
626 Seo, J.S., Keum, Y.S., Li, Q.X. (2009). Bacterial degradation of aromatic compounds.
627 International Journal of Environmental Research and Public Health, 6, 278-309.
628 Shannon, C.E., Weaver, W. (1949). The mathematical theory of communication.
629 Chicago, USA: University of Illinois Press.
630 Smalley, N.E., Taipale, S., De Marco, P. et al. (2015). Functional and genomic diversity
631 of methylotrophic Rhodocyclaceae: description of *Methyloversatilis*
632 *discipulorum* sp. nov. International Journal of Systematic and Evolutionary
633 Microbiology, 65(7), 2227-2233.
634 Thompson, L.R., Sanders, J.G., McDonald, D. et al. (2017). A communal catalogue
635 reveals Earth's multiscale microbial diversity. Nature, 551, 457-463.
636 van Dorst, J., Siciliano, S.D., Winsley, T. et al. (2014). Bacterial targets as potential
637 indicators of diesel fuel toxicity in subantarctic soils. Applied and
638 Environmental Microbiology, 80, 4021-4033.
639 Vázquez, S., Monien, P., Minetti, R.P. et al. (2017). Bacterial communities and
640 chemical parameters in soils and coastal sediments in response to diesel spills at
641 Carlini Station, Antarctica. Science of the Total Environment, 605-606, 26-37.
642 Viñas, M., Sabaté, J., Espuny, M.J. et al. (2005). Bacterial community dynamics and
643 polycyclic aromatic hydrocarbon degradation during bioremediation of heavily
644 creosote-contaminated soil. Applied and Environmental Microbiology, 71(11),
645 7008-7018.
646 Yagi, J.M., Sims, D., Brettin, T. et al. (2009). The genome of *Polaromonas*
647 *naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment,

reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. Environmental Microbiology, 11(9), 2253-2270.

Yang, S.H., Seo, H.S., Woo, J.H. et al. (2014). *Carboxylicivirga* gen. nov. in the family Marinilabiliaceae with two novel species, *Carboxylicivirga mesophila* sp. nov. and *Carboxylicivirga taeanensis* sp. nov., and reclassification of *Cytophaga fermentans* as *Saccharicrinis fermentans* gen. nov., comb. nov. International Journal of Systematic and Evolutionary Microbiology, 64(Pt 4), 1351-1358.

Yergeau, E., Bokhorst, S., Kang, S. et al. (2012). Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. The ISME Journal, 6(3), 692-702.

659 Table 1. Distribution of petroleum hydrocarbons in EACF soils (Table S2 presents the
660 complete list of hydrocarbons found in EACF soils)

	EACF soils				
	sA	**sB**	**sC**	**sD**	**sE**
Total petroleum Hydrocarbons (TPH)*	426,404	325,136	154,850	101,318	81,942
Unresolved complex mixtures (UCM)	162,112	104,315	61,353	58,789	37,717
Resolved aliphatic hydrocarbons (rAH)	163,000	104,520	38,940	28,440	29,040
Total of n-alkanes	101,292	116,301	54,557	14,089	15,185
The main n-alkane detailed (C12-C25)					
C12	737	1,165	1,210	288	212
C13	3,707	3,305	3,358	1,204	1,038
C14	9,788	6,801	8,865	3,538	3,006
C15	13,512	10,721	12,394	3,976	4,023
C16	15,807	16,609	10,313	2,599	3,217
C17	16,448	20,547	6,465	1,133	1,685
Prystane	5,853	5,547	3,344	787	1,001
C18	11,002	15,897	3,096	267	479
Phytane	3,238	4,528	1,267	170	255
C19	6,892	10,464	1,653	70	142
C20	4,285	6,648	910	24	52
C21	3,079	4,528	561	12	25
C22	1,996	2,944	336	7	16
C23	1,573	2,177	249	7	12
C24	1,054	1,410	157	7	13
C25	839	1,036	129	0	0

	EACF soils				
	sA	**sB**	**sC**	**sD**	**sE**
Polycyclic Aromatic Hydrocarbons (PAH)	41,321	57,884	43,322	37,743	25,623
Σ2−3 aromatic rings	41,003	57,522	43,152	37,711	25,398
Σ4−6 aromatic rings	415	423	204	40	274
SUM unmodified PAH	949	1,273	468	201	553
SUM alkyl PAH	40,373	56,611	42,854	37,542	25,070
The main PAH detailed					
Methylnaphthalene	904	2,849	964	927	362
Ethylnaphthalene	627	1,332	689	704	276
Dimethylnaphthalene	17,640	25,183	14,553	12,272	7,070
Trimethylnaphthalene	16,414	18,768	24,676	23,035	16,734
Methylfluorene	1,174	1,806	445	238	269
Methyldibenzothiophene	569	951	342	100	54
Methylphenanthrene	802	1,358	336	65	27
Dimethyldibenzothiophene	973	1,869	413	71	79
Dimethylphenanthrene	677	1,179	186	33	26

* in all cases the units are ng g^{-1}

Table 2. Absolute abundance of the main prokaryotic genera and the linear correlation (Spearman correlation) of bacterial genera distribution and the concentration of TPH and PAH in EACF soils.

	Absolute abundance[a]					Linear correlation[b]	
	sA	**sB**	**sC**	**sD**	**sE**	**TPH**	**PAH**
Archaea	0 (0)	0 (0)	0 (0)	$9.7x10^5$ ($6.8x10^5$)	$3.0x10^5$ ($2.6x10^5$)	-0.64 (**)	-0.64 (**)
o__Acidobacteriales; g__	0 (0)	0 (0)	$4.8x10^6$ ($2.1x10^6$)	$1.1x10^7$ ($3.4x10^6$)	$5.0x10^6$ ($6.7x10^5$)	-0.72 (**)	-0.51 (*)
c__Chloracidobacteria; g__	$1.4x10^6$ (0)	$7.9x10^5$ ($3.4x10^5$)	$4.4x10^7$ ($1.0x10^7$)	$1.5x10^7$ ($6.1x10^6$)	$2.4x10^7$ ($9.9x10^6$)	-0.43	-0.27
g__Arthrobacter	0 (0)	0 (0)	$1.2x10^6$ ($4.2x10^5$)	$1.1x10^7$ ($3.1x10^6$)	$1.7x10^7$ ($8.1x10^6$)	-0.84 (**)	-0.66 (**)
g__Williamsia	$1.2x10^7$ ($3.7x10^6$)	$8.6x10^7$ ($1.3x10^7$)	$4.1x10^7$ ($2.1x10^7$)	$1.4x10^7$ ($2.7x10^6$)	$2.9x10^5$ ($2.5x10^5$)	0.51 (*)	0.90 (**)
f__Patulibacteraceae; g__	$2.6x10^6$ ($8.3x10^5$)	$5.9x10^5$ (0)	$2.6x10^7$ ($6.9x10^6$)	$1.2x10^6$ ($3.4x10^5$)	$2.9x10^5$ ($2.5x10^5$)	0.53 (*)	0.46
f__Flavobacteriaceae; g__	$1.8x10^7$ ($3.1x10^6$)	$4.4x10^7$ ($1.5x10^7$)	$3.0x10^7$ ($1.7x10^7$)	$2.2x10^6$ ($1.0x10^6$)	$8.9x10^6$ ($5.0x10^6$)	0.45	0.71 (**)
g__Flavobacterium	$2.1x10^7$ ($4.7x10^6$)	$2.9x10^7$ ($3.7x10^6$)	$5.1x10^6$ ($2.9x10^6$)	$8.4x10^6$ ($8.5x10^6$)	$5.5x10^6$ ($9.2x10^5$)	0.67 (**)	0.55 (*)
o__Sphingobacteriales; g__	$3.6x10^7$ ($3.8x10^6$)	$2.4x10^6$ (0)	$3.8x10^7$ ($7.7x10^6$)	$3.2x10^7$ ($3.1x10^6$)	$3.3x10^7$ ($9.0x10^6$)	0.00	-0.26
g__Cytophaga	$8.4x10^7$ ($3.3x10^6$)	0 (0)	$1.2x10^6$ ($4.2x10^5$)	$1.7x10^6$ ($1.0x10^6$)	$4.4x10^5$ (0)	0.35	-0.17

	Absolute abundance[a]					Linear correlation[b]	
	sA	**sB**	**sC**	**sD**	**sE**	**TPH**	**PAH**
o__Clostridiales; g__	$3.4x10^6$ ($1.5x10^6$)	$9.5x10^6$ ($1.0x10^6$)	$1.2x10^6$ ($4.2x10^5$)	$2.4x10^5$ ($3.4x10^5$)	0 (0)	0.86 (**)	0.86 (**)
g__Gemmatimonas	0 (0)	0 (0)	$3.4x10^6$ ($1.1x10^6$)	$2.0x10^7$ ($4.8x10^6$)	$1.4x10^7$ ($9.2x10^5$)	-0.79 (**)	-0.62 (**)
g__Brevundimonas	$3.4x10^7$ ($7.9x10^6$)	$1.0x10^7$ ($1.2x10^6$)	$1.2x10^6$ ($8.4x10^5$)	$7.2x10^5$ ($3.4x10^5$)	$1.5x10^5$ ($2.5x10^5$)	0.94 (**)	0.62 (**)
g__Kaistobacter	$2.9x10^6$ ($1.2x10^6$)	$9.9x10^5$ ($3.4x10^5$)	$2.3x10^7$ ($3.6x10^6$)	$8.2x10^6$ ($2.0x10^6$)	$1.5x10^6$ ($2.5x10^5$)	0.03	0.00
c__Betaproteobacteria; g__	0 (0)	$5.9x10^5$ (0)	$7.3x10^6$ ($2.2x10^6$)	$1.6x10^7$ ($5.5x10^6$)	7.8x106 ($2.5x10^5$)	-0.71 (**)	-0.34
g__Aquabacterium	$3.1x10^7$ ($1.7x10^6$)	$1.6x10^6$ ($3.4x10^5$)	$5.1x10^6$ ($1.4x10^6$)	$1.2x10^6$ ($3.4x10^5$)	$1.2x10^6$ ($2.5x10^5$)	0.74 (**)	0.41
g__Polaromonas	$2.1x10^7$ ($4.2x10^5$)	$7.4x10^7$ ($1.2x10^7$)	$8.1x10^7$ ($2.3x10^7$)	$1.7x10^7$ ($6.8x10^5$)	$2.4x10^7$ ($8.8x10^6$)	0.29	0.72 (**)
g__Xenophilus	$1.2x10^7$ ($1.2x10^6$)	$4.1x10^7$ ($1.1x10^7$)	$3.9x10^7$ ($1.4x10^7$)	$7.2x10^6$ (0)	$3.7x10^6$ ($1.3x10^6$)	0.61 (**)	0.91 (**)
g__Methyloversatilis	$1.8x10^8$ ($1.7x10^7$)	$5.6x10^7$ ($7.4x10^6$)	$1.7x10^7$ ($4.8x10^6$)	$1.5x10^7$ ($6.8x10^5$)	$2.6x10^7$ ($9.9x10^6$)	0.70 (**)	0.31
g__Geobacter	$2.2x10^7$ ($5.6x10^6$)	$5.3x10^6$ ($1.2x10^6$)	$6.8x10^6$ ($1.1x10^6$)	$2.4x10^5$ ($3.4x10^5$)	$2.5x10^6$ ($1.1x10^6$)	0.79 (**)	0.50
g__Aquicella	0 (0)	0 (0)	$2.4x10^5$ ($4.2x10^5$)	$2.4x10^5$ ($3.4x10^5$)	$1.7x10^7$ ($6.3x10^6$)	-0.73 (**)	-0.64 (**)

	Absolute abundance[a]					Linear correlation[b]	
	sA	**sB**	**sC**	**sD**	**sE**	**TPH**	**PAH**
f__Opitutaceae; g__	$4.0x10^7$ ($2.0x10^7$)	0 (0)	$2.4x10^5$ ($4.2x10^5$)	0 (0)	0 (0)	0.67 (**)	0.08
g__Chthoniobacter	$7.2x10^5$ (0)	0 (0)	$3.4x10^6$ ($1.1x10^6$)	$1.1x10^7$ ($4.4x10^6$)	$1.7x10^7$ ($2.7x10^6$)	-0.72 (**)	-0.74 (**)
o__Verrucomicrobiales; g__	$9.6x10^5$ ($4.2x10^5$)	0(0)	$2.9x10^6$ ($1.4x10^6$)	$7.7x10^6$ ($4.1x10^6$)	$6.2x10^6$ ($4.4x10^5$)	-0.69 (**)	-0.71 (**)
Bacteria: others	$2.0x10^8$ ($1.7x10^7$)	$2.3x10^8$ ($1.3x10^7$)	$3.4x10^8$ ($2.5x10^7$)	$2.8x10^8$ ($3.2x10^7$)	$2.2x10^8$ ($2.4x10^7$)		

[a] Absolute abundance with standard deviation in parentheses. [b] Positive and negative correlation of the abundance of ASVs with TPH and PAH concentration in EACF soils. (*) and (**) indicate significant differences of Spearman correlation considering $P<0.05$ and $P<0.01$, respectively.

Figure and legends

Fig 1. Quantification of bacterial abundance and diversity in EACF soils. (A) Estimation of bacterial abundance based on quantification of 16S rRNA gene amounts by qPCR; (B) and (C) Diversity analyses (Faith`s PD and Shannon indexes, respectively) of microbial communities present in EACF soils; (D) Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities of 16S rRNA gene (ASVs) distribution. The concentration of TPH, PAH and alkanes in EAFC soils were used as vectors in NMDS analysis. PERMANOVA were used to compare the structure of microbial community present in EAFC soil; (E) Cluster distance based on Bray-Curtis dissimilarities of functions (rarefied KEGG orthologies (KOs) counts) profiles obtained from metagenomic data.

Fig 2. Average of absolute abundance of major bacterial and archaeal groups observed in EACF soils (contributing >1% of the total abundance in at least one sample). Phylum and Proteobacteria class distribution.

Fig 3. Heat map showing the relative abundances of genes involved in degradation of hydrocarbons in the metagenome data of EACF soils (based on rarefied KEGG orthologies (KOs) counts).

Fig 4. Relative abundances (log scale) of the genes involved in the main hydrocarbon degradation pathways observed in EACF soils (metagenome data based on rarefied KEGG orthologies (KOs) counts). (A) Genes of 1-methylnaphthalene degradation to cathecol; (B) genes of 1-methylnaphthalene degradation to 4-formylsalicylic acid and/or 2-naphthoate; (C) genes of phenanthrene degradation to 1-methoxyphenanthrene. The

genes names: 1a - 1-methylnaphthalene hydroxylase [EC:1.14.13.-]; 1b - alcohol dehydrogenase [EC:1.1.1.1]; 1c - 1-naphthaldehyde dehydrogenase [EC:1.2.1.-]; 1d - 1-naphthoic acid dioxygenase [EC:1.14.12.-]; 1e - cis-1,2-dihydroxy-1,2-dihydro-8-carboxynaphthalene dehydrogenase [EC:1.1.-.-]; 1f - 1,2-dihydroxy-8-carboxynaphthalene dioxygenase [EC:1.13.-.-]; 1g - 2-carboxy-2-hydroxy-8-carboxychromene isomerase [EC:5.3.99.-]; 1h - 2-hydroxy-3-carboxy-benzalpyruvate hydratase-aldolase [EC:4.2.1.-]; 1i - 3-formylsalicylate oxidase [EC:1.2.3.-]; 1j - 2-hydroxyisophthalate decarboxylase [EC:4.1.1.-]; ik - salicylate hydroxylase [EC:1.14.13.1]; il - catechol 2,3-dioxygenase [EC:1.13.11.2]. 2m - phenanthrene 1,2-monooxygenase [EC:1.14.13.-]; 2n - [EC:2.1.1.-].

Fig 5. Relative abundances of genes in the nitrogen cycle that were observed in the metagenome data (rarefied KEGG orthologies (KOs) counts). (A) Heat map showing the relative abundances of genes involved in the nitrogen cycle for each of the EACF soils. (B) Nitrogen cycle showing differences between the EACF soils. Red lines depict genes that were enriched in accordance with the decrease of TPH, blue lines depict genes that were enriched in accordance with the increase of TPH and black line: genes not correlated to the TPH concentration. Dashed blue lines represent our hypothesis concerning the fate of fixed nitrogen in hydrocarbon contaminated soils. The genes of (i) nitrogen fixation: *nifDKH* (molybdenum-iron nitrogenase complex) and *anfG* (alternative nitrogenase); (ii) nitrification: *amoCAB* (ammonia monooxygenase), *hao*, (hydroxylamine oxidoreductase) and *nxrAB* (nitrate reductase/nitrite oxidoreductase); (iii) denitrification: *narGHI* (nitrate reductase/nitrite oxidoreductase), *napAB* (periplasmic nitrate reductase), *nirKS* (nitrite reductase (NO-forming)/hydroxylamine reductase), *norBC* (nitric oxide reductase) and *nosZ* (nitrous-oxide reductase); (iv)

Assimilatory nitrate reduction: *nasAB* (assimilatory nitrate reductase), *narB* (ferredoxin-nitrate reductase), and *nirA* (ferredoxin-nitrite reductase); and (v) Dissimilatory nitrate reduction: *nrfAH* (cytochrome c nitrite reductase).