# *Integrating shotgun metagenomics, 16S rRNA gene metabarcoding and culture approaches: a better outlook for functional profiling of a PAH-contaminated soil*

Festa Sabrina[1]*, Granada Marina[1], Irazoqui José M.[2], Cuadros-Orellana Sara[3], Quevedo Claudio[3], Coppotelli Bibiana M.[1], Morelli Irma S.[1,4]

1-Centro de Investigación y Desarrollo en Fermentaciones Industriales, CINDEFI (UNLP, CCT-La Plata, CONICET), La Plata, Argentina

2- Instituto de Investigación de la Cadena Láctea, IdICaL (INTA-CONICET). Rafaela, Santa Fe, Argentina

3- Facultad de Ciencias Agrarias y Forestales, Centro de Biotecnología de los Recursos Naturales, Universidad Católica del Maule, Talca, Chile

4- CIC-PBA, La Plata, Argentina

*Corresponding author: Sabrina Festa
CINDEFI, Street 50 N°227, 1900 La Plata, Argentina
E-mail: sfesta@biotec.quimica.unlp.edu.ar
ORCID: 0000-0002-3683-9305

**Key words**

polycyclic aromatic hydrocarbon, shotgun metagenomics, *Mycolicibacterium*, enrichment culture

**Abstract**

Understanding bacterial diversity and function is critical for designing bioremediation strategies. This research aimed to assess chronically hydrocarbon contaminated soil bacterial diversity and their aromatic compound degradation (ACD) potential by integrating shotgun metagenomic, 16S rRNA gene metabarcoding and culture approaches. While soil metabarcoding showed dominance of Proteobacteria, metagenomics indicated that 99,5% of the sequences were taxonomically assigned to Streptomycetales order and that almost all genes related to ACD were assigned to the latter. To inspect other phyla contribution to ACD, a functional prediction was delved, and two culture approaches were used. PICRUSt2 revealed that ACD pathways were mostly found in Alphaproteobacteria, Actinobacteria and Gammaproteobacteria classes. An enrichment culture (r-EFP) was obtained with pyrene as sole carbon and energy source and a bacterial strain (S19P6), identified as a member of *Mycolicibacterium* genus, was isolated. Both cultures demonstrated the ability to degrade more than 90% of the supplemented pyrene after 21 days of incubation. 16S rRNA gene metabarcoding and shotgun metagenomics approaches in r-EFP indicated predominance of Proteobacteria Phylum and the presence of genes responsible for the degradation of ACD mostly assigned to the predominant phyla. Complementing different methodologies enable the recognition of the metabolic potential of soil Proteobacteria related to ACD.

**Environmental implication**

Polycyclic aromatic hydrocarbons (PAH) are chemical compounds that are hazardous and persistent in the environment. Due to its toxicity, strategies for removing these pollutants are required. Microbial processes are considered one of the most significant and influential in PAH removal; as a consequence, it is crucial for designing bioremediation strategies to understand microbial functional and phylogenetic diversity. This investigation makes a major contribution to research regarding the bacterial processes involved in a chronically contaminated soil by integrating different methodological approaches to try to explain an observed limitation in PAH degradation when applying different bioremediation strategies on a PAH contaminated soil.

## Introduction

Microbial processes are considered one of the most significant and influential in polycyclic aromatic hydrocarbon (PAH) removal (Miller *et al.*, 2004). The understanding of microbial functional and phylogenetic diversity within a contaminated niche has been recommended as critical for risk assessments, microbial monitoring and for designing bioremediation strategies as a consistent approach to clean up environmental pollutants (Gosai *et al.*, 2018). To face anthropogenic contamination, exploring indigenous microbial communities adapted to these environments over long periods of time is expected to provide significant information (Scoma *et al.*, 2017). This implies advanced methodologies like molecular tools able to provide in-depth understanding about the aspects of microbial processes and survival under stressed environment (Mishra *et al.*, 2021).

The diversity and metabolic capacity of soil microbiome can be inferred from 16SrRNA gene metabarcoding and described by shotgun metagenomics. Gene marker approaches coupled to functional prediction software's like PICRUSt rely on the correlation between phylogenetic trees and clusters of genes shared between taxa (Langille *et al.*, 2013). Shotgun metagenomics, on the other hand, provides a direct assessment of the functional attributes of the microbiome (Jovel *et al.*, 2016). However, metagenomics centred approaches alone could not be sufficient to reveal functions belonging to the prokaryotic rare biosphere because most low abundance prokaryotes remain uncultured and consequently understudied (Pascoal, Magalhães and Costa, 2020). Cultivation-based approaches can acquire microorganisms that are undetected by molecular methods and furthermore allow for the verification of metabolic potential determined by the metagenomic data (Prakash *et al.*, 2013). The use of both molecular and culture-based methods gives the opportunity not only to collect information on the composition of microbial communities but also to isolate microorganisms capable of developing a specific activity (Besaury *et al.*, 2013). Isolating environmental microorganisms and studying their physiology under controlled conditions are essential aspects of understanding their ecology (Bartelme *et al.*, 2020). Culture is needed to describe novel species as a function and as the *sine qua non* condition for understanding how the microbial world functions (Alain and Querellou, 2009). That is why an integrative approach should be considered while assessing environmental microbial diversity (Suenaga *et al.*, 2014).

Related isolation techniques are enrichment cultures, in which environmental conditions are tailored to favor certain genotypes (that cross-feed essential metabolites), resulting in self-assembled microbiomes (Souza, Shitut and Kost, 2018; Lawson *et al.*, 2019). While natural consortia are still difficult to scrutinize, enrichment cultures offer a compromise between

natural and synthetic communities (Ponomarova and Patil, 2015) and a good model to study the ecological niche of key taxa and the interactions that can occur between bacterial populations during a bioremediation process (Desai, Pathak and Madamwar, 2010).

This paper attempts to explore the genetic potential of the bacterial community of a chronically PAH contaminated soil. This soil, called IPK, came from a petrochemical area near La Plata city, Argentina. This soil belonged to a landfarming unit where an alkaline petrochemical sludge (with high PAH concentration) was treated; this landfarming unit was closed almost twenty years ago. When sampled, PAH of three and four rings were detected in IPK soil. Different treatments have been studied in order to select a strategy that can achieve the elimination of residuals PAH. Allochthons bioaugmentation and surfactant-enhanced bioremediation were assayed and a limitation in the degradation of contaminants was observed (Festa *et al.*, 2016; Cecotti *et al.*, 2018). To try to understand the observed limitation a metagenomic approach was performed. Furthermore, enrichment cultures could serve as a tool for identification of microbial degraders and metabolic pathways when biological components involved in hydrocarbon degradation in the environment are below detection limits of the techniques used. That is why, considering that the comprehension of the composition and metabolic potential of a soil bacterial community could be also done by exploring bacterial cultures of that soil, and in light of autochthonous bioaugmentation as a possible future strategy for decontamination, a culture analysis was carried out. Therefore, this study integrates approaches to makes a major contribution to research regarding the bacterial processes involved in a bioremediation.

## Material and methods

### 1 Soil characterization

#### 1.1 Physicochemical properties

A chronically PAH contaminated soil, called IPK, was collected from a petrochemical area near La Plata city, Argentina. This soil belonged to a land farming unit where an alkaline petrochemical sludge (with high PAH concentration) was treated. In order to identify the physicochemical properties of IPK an analysis was carried out in the Laboratory of Soil Science at the University of La Plata and showed a clay loam texture, pH of 8, 5.73% (w/w) organic carbon, 9,88% (w/w) soil organic matter, 0,134 (w/w) total nitrogen, C/N ratio, 25 $mg.kg^{-1}$ available phosphorus and $Ca^{+2}$, $Mg^{+2}$, $Na^{+}$ and $K^{+}$ 28,4, 10, 2,14, 1,1 ($cmolc.kg^{-1}$) respectively. All the different analysis performed regarding IPK soil were carried out with sieved soil (decreasing pore size until in a 2-mm mesh.

#### 1.2 Chemical hydrocarbon extraction and quantification and microbial enumeration

Total PAH and aliphatic hydrocarbon in IPK soil were extracted according to Mora *et al.*, (2014) with hexane:acetone (1:1), and the resulting extract was analyzed in a PerkinElmer auto system gas chromatograph with a flame ionization detector (GC-FID) according to Del Panno *et al.*, (2005). Microbial counts were determined as described in previous work (Festa, Coppotelli and Morelli, 2016); in summary viable heterotrophic bacterial counts were determined in R2A agar medium and incubated at 24±2 °C for 10 days. Also, the aromatic hydrocarbon-degrading bacteria was determined microtiter plates, using liquid mineral medium (LMM) and incubated 24±2 °C for 21 days. Most Probable Number (MNP) method was used to enumerate PAH degraders as described by Festa, Coppotelli and Morelli (2016).

### 1.3 DNA extraction for qPCR, 16S rRNA gene metabarcoding and shotgun metagenomic analysis

Total soil DNA extraction was performed in triplicates using E.Z.N.A. ® Soil DNA Kit (Omega Bio-Tek, Inc., Norcross, GA, USA) following manufacturer's instructions. DNA was stored at -20°C until analysis.

Different qPCR were carried out for IPK soil DNA in order to qualitatively and quantitative characterize the genetic potential of the IPK microbial community, using the following *primers*: 1055F (5´-ATGGCTGTCGTCAGCT-3´) and 1392R (5´-ACGGGCGGTGTGTAC–3´) for 16SrRNA gene; PAH-RHDα GN F (5´-GAGATGCATACCACGTKGGTTGGA-3´) and PAH-RHα GN R (5´-AGCTGTTGTGGAAGAYWGTGCMGTT-3´) designed for amplifying the α subunit of the PAH ring hidroxilating dioxygenase gene of gram negative bacteria; and NidAF (5'-TTCCCGAGTACGAGGGATAC-3´) and NidAR (5'-TCACGTTGATGAACGACAAA-3´) designed for pyrene dioxigenase gene (NidA). Amplification program and used pairs of *primers* were designed by Harms *et al.*, (2003),Cébron *et al.*,( 2008) and Peng *et al.*, (2010) respectively.

Quantification of 16SrRNA and PAH-RHDα gene copy number was performed in triplicates in a Rotor-Gene Q® (QIAGEN). For all qPCR assays three technical replicates on samples and negative controls (PCR blanks with only the reaction mix and PCR blanks containing the mix and 1 µl of PCR-grade water) were used. Separate tubes in triplicate were used to measure threshold cycles (Ct) and the identity and purity of the amplified product were checked by analyzing the melting curve at the end of amplification. Reaction was carried out as described by Madueño *et al.*, (2018). Results were expressed as gene copy number.$g^{-1}$ of dry soil.

#### *1.3.1 Metabarcoding of 16S rRNA gene*

16S rRNA metabarcoding analysis for IPK soil was carried out in triplicates using 16S rRNA universal bacterial primers 515F (5' -GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') to amplify V4 hypervariable region of 16S rRNA gene. Sample sequencing was performed using the sequencing platform Illumina NovaSeq 6000 at Molecular Research laboratory (MR DNA; Shallowater, TX) based on established and validated protocols (http://www.mrdnalab.com/) as described in previous work (Festa et al., 2016). Paired end 16S rRNA reads were analyzed using QIIME2 2021.2 pipeline (Bolyen *et al.*, 2019), raw sequence data was demultiplexed and quality filtered followed by denoising with DADA2 to obtain Amplicon Sequence Variants (ASVs). A filtering approach was carried out to remove low-abundance sequences (<0.1% of the average depth of the samples), mitochondrial/chloroplast 16s rRNA sequences and reads that were not classified at the phylum level. After alpha rarefaction and alpha diversity analysis, taxonomy composition was explored Silva 138 trained Naive Bayes classifier 99% OTUs from 515F/806R region of sequences. Metabarcoding analysis was coupled to PICRUSt2 software (v2.4.1) to predict the abundance of gene families KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologs (KOs) and higher-level pathways (MetaCyc) was used (Douglas *et al.*, 2020). PICRUSt2 was executed following the pipeline established by the software developers. Prediction accuracy was estimated by the Nearest Sequenced Taxon Index (NSTI) scores. Bar plots, Bubble plots and Sankey Diagrams were generated using R v. 4.2.0 (R Foundation for Statistical Computing) and ggplot2_3.3.6 (Wickham, 2016).

#### *1.3.2 Metagenomic analysis*

Shotgun metagenomics was performed in Corporación Corpogen (https://www.corpogen.org/), using an Illumina Miseq sequencing platform. Raw reads were quality trimmed using

Trimmomatic (version 0.39; Bolger, Lohse and Usadel, 2014), to remove adapters and low quality regions, and an estimation of the total coverage obtained was calculated using Nonpareil (version 3.2; Rodriguez-R and Konstantinidis, 2014). An assembly was obtained using SPAdes (version 3.11.1; Bankevich *et al.*, 2012) and contigs were binned using MaxBin2 (version 2.2.5; Wu, Simmons and Singer, 2016). The quality of the bins obtained was checked using CheckM (version 1.0.11; Parks *et al.*, 2015). Then, genes were predicted on the assembled contigs using prodigal (version 2.6.3). Finally, these sequences were annotated using Interproscan (version 5.50.84; Blum *et al.*, 2021)and kofamscan (version 1.3.0; Aramaki *et al.*, 2020). All metabolic pathways regarding the degradation of aromatic compounds, and enzymes described for each step were obtained from MetaCyc (Caspi *et al.*, 2018). For those enzymes described in the InterPro or KEGG database, we extracted their annotation. In order to reconstruct the metabolic pathways, we compared the annotation from the MetaCyc enzymes with that of our predicted proteins and used CytoScape (version 3.9.1) to construct the metabolic network. Also, GhostKOALA annotation server (Kanehisa, Sato and Morishima, 2016) was used to find potential functions related to the metabolic pathways of interest.

## 2. Evaluating pyrene degrading communities

### 2.1 Obtaining a pyrene degrading enrichment culture

A pyrene-degrading bacterial enrichment culture was obtained from IPK soil by five sequential transfers in LMM with 1000 mg.l$^{-1}$ of pyrene as sole carbon and energy source and 0,05 g.l$^{-1}$ yeast extract. Biomass transfer:culture medium was 1:10 ratio and each transfer was incubated 7 days at 28°C and 150 rpm. Last bacterial suspension was finally conserved in 40% of glycerol at -80 °C. When need it for an assay the enrichment culture was reactivated in LMM supplemented with 1000 mg.l$^{-1}$ of pyrene and 0,05 g.l$^{-1}$ yeast extract for 7 days at 28 °C and 150 rpm. The reactivated enrichment culture was named r-EFP

#### *2.1.1 Pyrene degradation kinetics*

The pyrene degradation by the enrichment culture was determined in LMM with 100 mg. l $^{-1}$ of pyrene as a sole carbon and energy source and 0,05 g.l$^{-1}$ yeast extract. The cultures were inoculated with the reactivated r-EFP and incubated at 28°C and 180 rpm (in triplicates) during 21 days of incubation. An abiotic control was incubated in the same conditions.

At different incubation times remaining pyrene was measured in the r-EFP cultures. A chemical extraction was carried out three times with ethyl acetate as previously described (Festa, Coppotelli and Morelli, 2013). The ethyl acetate extracts were analyzed by reversed-phase high-pressure liquid chromatography (HPLC) using a Waters chromatograph with a Symmetry Waters C18 column (15 cm 4.6 mm i.d., bead size 5 mm, pore size 100 A) and a diode-array detector. A linear gradient of 15 mM phosphoric acid in nanopure water solution and methanol (20:80 to 5:95 vol.vol$^{-1}$) over 15 min and a flow rate of 1 ml.min$^{-1}$ was used.

#### *2.1.2 DNA extraction for 16S rRNA gene metabarcoding and shotgun metagenomic analysis*

DNA extraction was carried out using E.Z.N.A. ® Soil DNA Kit (Omega Bio-Tek, Inc., Norcross, GA, USA) following manufacturer's instructions. DNA was stored at -20°C until analysis.

16S rRNA gene metabarcoding and shotgun metagenomics were performed in an Illumina NovaSeq 6000 sequencing platform at Novogene (https://en.novogene.com/). Metabarcoring analysis was carried out using 16S rRNA universal bacterial primers 515F (5' -GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') to amplify V4

hypervariable region of 16S rRNA gene. Paired end sequence data analysis was performed as described above in this manuscript for DNA IPK soil sample, using QIIME2 Pipeline (version 2021.2).

Shotgun metagenomic raw reads were quality trimmed using Trimmomatic (version 0.39). An assembly was obtained using MegaHit (version 1.2.9; Li *et al.*, 2015) and the quality was checked using QUAST (Gurevich *et al.*, 2013). Then, genes were predicted and annotated on the assebmled contigs using Prokka (Seemann, 2014). Contigs were binned using Binnacle (version 2.2.5; Muralidharan *et al.*, 2021) and the quality of the bins obtained was checked using CheckM (version 1.0.11; Parks *et al.*, 2015). Taxonomical assignments of the metagenome assembled genomes was carried out using PhyloPhlAn (version 3.0.36, Segata *et al.*, 2013) and CAT/BAT (Bastiaan von Meijenfeldt *et al.*, 2019). All metabolic pathways regarding the degradation of aromatic compounds, and enzymes described for each step were obtained from MetaCyc (Caspi *et al.*, 2018), KEGG AND AromaDeg database (Duarte *et al.*, 2014). GhostKOALA annotation server (Kanehisa, Sato and Morishima, 2016) was used to find potential functions related to the metabolic pathways of interest.

**2.2 Isolation of a pyrene-degrading bacterial strain**

A portion of IPK soil was suspended it in an Erlenmeyer with physiological solution 0,85% $w.v^{-1}$. Dilutions were carried out from the suspension and an aliquot was spread on agar plates containing R3A (Vecchioli, Del Panno and Painceira, 1990) supplemented an 1% agarose layer with 0, 46 mg/ml of pyrene. After 28 days of incubation, bacterial colonies demonstrating a transparent halo around the colony were isolated.

**2.1 Biodegradation studies and molecular characterization of S19P6 isolated strain**

Cultures with 100 ppm of pyrene as carbon and energy source were inoculated with S19P6 strain and incubated at 28°C and 180 rpm (in triplicates). These cultures were used to quantify pyrene degradation and to evaluate the growth curve of S19P6 using pyrene as sole carbon and energy source at different incubation times. An abiotic control was incubated in the same conditions.

For pyrene degradation kinetics cultures were extracted three times with ethyl acetate and analyzed as explained above by HPLC. Growth curve of S19P6 was monitored by counting colony-forming units (CFU) after spreading 0.1 ml of an appropriate dilution on R2A medium and incubating the plates in the dark for 7 days at 28°C (Festa, Coppotelli and Morelli, 2013).

In order to stablish the identity of the isolated strain, DNA was extracted with boiling method (Festa, Coppotelli and Morelli, 2013) and a PCR amplification using 16S rRNA gene targeted primers was carried out. Resulting amplicon was sequenced by Sanger sequencing service in Macrogen. Nucleotide sequences were compared to those in the National Center for Biotechnology Information GenBank database by using BLAST program. To study the phylogenetic relationship between the 16S rRNA gene sequences from the isolated strain and selected sequences available in GenBank, a distance-based evolutionary tree was constructed using Neighbor-Joining method and p-distance algorithm with the Molecular Evolutionary Genetics Analysis package (MEGA version 4.0). The robustness of the phylogeny was tested by bootstrap analysis with 500 iterations

Presence of genes related to key enzymes involved in the degradation pathway of pyrene in S19P6 was evaluated by PCR using two sets of primers targeting *nidA* (as described above) and *pahE* genes. For the latter, the used primers (pahE4F and pahE4R) were designed by Liang, Huang and Wang, (2019). PCR conditions were carried out as described by those authors.

## Data deposition

Metagenomic and amplicon sequence data are available at the NCBI Short-Read Archive under project accession number PRJNA889761 for IPK soil and under project accession number PRJNA906488 for r-EFP enrichment culture datasets. Full length 16S rRNA gene sequence of the isolated strain is available in NCBI GenBank under accession number OP899840.

## Results

## 1. Soil characterization

### 1.1 Hydrocarbon quantification and microbial enumeration

Chemical extraction was performed in order to describe PAH content of IPK soil. As it can be seen from the data shown in table 1, IPK soil presented a total PAH hydrocarbon concentration of 412.3±24.8 $mg.kg^{-1}$ of dry soil. Notably, a predominance of three and four rings PAH was observed with pyrene as major concentration hydrocarbon (Table 1), 119.3±4.3 $mg.kg^{-1}$ of dry soil, followed by acenaphtylene 79.1± 3.3 $mg.kg^{-1}$ of dry soil.

Regarding bacterial enumeration, an estimation of total heterotrophic cultivable bacteria, PAH-degraders and specifically pyrene degrading bacteria, was assessed by culture dependent and culture independent techniques.

Heterotrophic cultivable bacteria revealed a concentration of $4.15.10^{+6} \pm 3.08.10^{+5}$ CFU. g dry $soil^{-1}$ while pyrene and PAH degrading bacteria counts were four orders of magnitude lower, $3.26.10^{+2} \pm 1.58.10^{+2}$ and $4.09.10^{+2} \pm 4.16.10^{+1}$ respectively. Quantification of 16S rRNA gene and PAH-RDHα GN functional gene was performed by a qPCR assay. The obtained value for total bacteria was $4.27.10^{+10} \pm 2.52.10^{+10}$ 16S rRNA copies. g dry $soil^{-1}$ and for the functional gene was $2.16.10^{+8} \pm 1.72.10^{+8}$ PAH-RHDα copies. g dry $soil^{-1}$; the difference between those values was not as high as for the culture dependent approach.

### 1.2 Diversity analysis and prediction of functional potential

Taxonomic and diversity analysis was accomplished by 16Sr RNA gene metabarcoding. After proper analysis, a total of 382,461 sequences were obtained for IPK triplicates samples and were grouped in 369 amplicon sequence variants (ASVs). The rarefaction curves (Figure S1) and Good´s coverage (value of 1) illustrated that the effort was enough to cover most of the diversity of the samples. As diversity measures, Hill-numbers, ${}^{0}D$, ${}^{1}D$ and ${}^{2}D$ were 365, 110.77 and 31.21 respectively. The decrease observed between ${}^{0}D$ and ${}^{1}D$ values suggested that IPK soil presents a highly impacted and uneven bacterial community, with low richness and diversity (Table S1).

At phylum level a clear dominance of Proteobacteria (53.7% of relative abundance) was observed, followed by Actinobacteria (26.3% of relative abundance) (Figure 1). Within Proteobacteria, Immundisolibacterales was the most represented order in the samples (14.8% of relative abundance) followed by Burkholderiales order (5.0% of relative abundance); Actinobacteria was mainly represented by Microtrichales order (15.6% of relative abundance).

In order to assess the functional potential of IPK soil, a predictive functional profile of the community was assessed using PICRUSt2. Mean Weighted Nearest Sequenced Taxon Index (NSTI) value was 0.22 ±0.01. When looking to the predicted pathways associated to the degradation of aromatic compounds (Figure 2a), it can be seen that the ortho-cleavage pathway

of protocatechuate (P3) was the most significantly (p<0,01 ANOVA-Tukey) represented, followed by the meta- and orto- cleavage of catechol pathways (P4 and P5 respectively) and the superpathway of salicylate degradation (P8). Additionally, a prediction of the contribution of each taxa to the mentioned pathways was carried out. A clear predominance of Alphaproteobacteria, Actinobacteria, Gammaproteobacteria and Acidimicrobiia classes was observed, representing more than the 80% of the contribution to each pathway (Figure.2a.). In particular, Alphaproteobacteria class was the major contributor to ortho-cleavage of protocatechuate (P3). A selection of the most represented KO for each of those pathways were carried out and taxa contribution to those KO was predicted and showed in Figure 2b. Regarding functions related to the ortho-cleavage of Protocatechuate (green bars), a higher proportion of K01607 (EC:4.1.1.44) and K01055 (3.1.1.24) was found in Microtrichales and Hyphomicrobiales orders. Concerning meta-cleavage of catechol (yellow bars): K07104 (1.13.11.2) was mostly found to be related to Hyphomicrobiales order, while K01821 (5.3.2.6) was equally found in Hyphomicrobiales and Burkholderiales. However, Xanthomonadales and Dongiales were major contributors to K02554 (4.2.1.80), indicating that Proteobacteria Phylum (blue bars) was the major contributor to this pathway.

## 1.3 Metagenomic analysis

### *1.3.1 Soil taxonomic profile*

From the sequencing of total DNA IPK soil, approximately 11 millon of high quality raw paired end sequences were obtained. Sample coverage in Nonpareil software was analyzed, the result indicated that IPK sample was oversample, meaning full dataset coverage was achieved, inferring the presence of low diversity community in IPK soil (Figure S2).

The taxonomic description of the soil is shown in figure 3a. About 97% of the cleaned reads were taxonomically classified and, almost all were assigned to Actinobacteria phylum (99,81%), specifically Streptomycetales order (99,54%). A vast minority of all sequences were assigned to two other phyla (Figure 3a) Proteobacteria (0,11%), represented by Pseudomonadales (0,02%) and Immundisolibacterales order (0.01%), and Firmicutes (0,05%) represented by Bacilliales order (0,03%). After the assembly of reads a total of 5357 contigs were obtained, being 242 contigs larger than 10 kbp. Around 6300 genes were predicted in those contigs (Figure 3b), where almost 99% of genes were assigned to *Streptomyces* genus. After a binning procedure a bin of a size of almost 5 Mb (81,6% of completeness and 0,86% of contamination) composed of 349 contigs was recovered, it possesses 73% of GC content and was taxonomically assigned to Streptomycetaceae family.

### *1.3.2 Functional characterization of IPK soil community*

*Functional categories related to hydrocarbon degradation*

Different approaches were used to search for genes related to aromatic compound degradation in IPK soil metagenome.

Based on GhostKOALA annotation, a total of 12 degradation-related genes were found, 10 assigned to Actinobacteria phylum, mainly to Streptomycetales order, and 2 to Proteobacteria, in particular to Burkholderiales and Hyphomicrobiales orders (Table 2). Meta-cleavage of protocatechuate was the most represented pathway (with four sequences assigned to this

pathway), followed by meta-cleavage of catechol (with three KO sequences assigned to this pathway), all represented by Actinobacteria related sequences. K14584 (related to naphthalene degradation) was assigned to Burkholderiales order and K01055 (related to ortho-cleavage of catechol and protocatechuate degradation) to Hyphomicrobiales.

As only a few aromatic compound degradation related genes were found using GhostKOALA approach, a comprehensive analysis of the enzymatic domains present in IPK soil metagenome was carried out using MetaCyc database. As a result, a metabolic network relating the aromatic compound degradation pathways found is shown in Figure 4. All the enzymatic domains associated to those pathways were found to be assigned to Streptomycetales order. In the constructed network, the width of the lines connecting two compounds refers to the percentage of completeness of the pathway involved (referring to what percentage of the enzymatic domains were found in relation to the total), and the color refers to what type of dioxygenase was found in the metagenome to be participating in that pathway.

With this approach, almost all genes codifying for enzymes of naphthalene degradation upper pathway were found (78% of completeness, from naphthalene to salicylate). Concerning the initial step of the pathway, the presence of dioxygenase enzyme belonging to Rieske non-heme iron oxygenases family, capable of participating in this and in other several pathways (benzene, p-cumate, biphenyl, phthalate and salicylate degradation) was found (green lines). Also, the domain of a key enzyme responsible for the fifth step of the upper pathway, a trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (EC: 4.1.2.45, K14585) was recovered from the metagenome.

Regarding central intermediates, almost the complete set of enzymes (between 80% and 100%) was found for phthalate degradation, salicylate and catechol degradation (meta-cleavage pathway). However, for the latter, the gene coding for the enzyme responsible for the meta cleavage was not found in IPK metagenome (catechol 2,3-dioxygenase). Despite 57% of the enzymes responsible for catechol ortho-cleavage pathway were found, a key enzyme, the intradiol dioxygenase responsible for the cleavage of catechol was present in the metagenome (catechol 1,2-dioxygenase, pink line). A high percentage (between 60 and 79%) of completeness was observed for protocatechuate degradation (meta-cleavage pathway), protocatechuate degradation (para-cleavage pathway) plus 2-hydroxypenta-2,4-dienoate degradation, gentisate degradation and salicylate degradation plus gentisate degradation II.

Within protocatechuate metabolic pathways, protocatechuate 2,3-dioxygenase (responsible for the initial step in the para-cleavage pathway) was found. Nor protocatechuate 4,5-dioxygenase neither protocatechuate 3,4-dioxygenase genes were observed in the metagenome.

As dioxygenase responsible for the initial attack of pyrene did not have functional gene number and therefore it was no possible to predict its presence in IPK soil metagenome, a PCR was carried out with primers directed to a gene codifying for the PAH dioxygenase involved in pyrene initial attack (*nidA*). A positive result was obtained, indicating that some members of IPK bacterial community could be able to initiate PAH/pyrene degradation (Figure S3).

## 2. Evaluating pyrene degrading communities

### 2.1 Pyrene degrading enrichment culture r-EFP

In order to use culture approach as another tool for identification of microbial degraders, a pyrene degrading bacterial enrichment culture, called r-EFP was obtained from IPK soil using

pyrene as sole carbon and energy source. The pyrene degradation kinetic determined during 21 days of incubation in r-EFP cultures is shown in figure 5. Although pyrene elimination was not observed during the first 7 days of incubation, a significant decrease of the supplemented pyrene was detected after 15 days of incubation (50%). Furthermore, at day 21 of incubation a significantly higher degradation percentage (94.9%) was measured in r-EFP cultures.

### *2.1.1 r-EFP diversity and functional potential*

The bacterial composition of the enriched culture r-EFP was assesed by 16S rRNA metabarcoding. A total of 272080 paired end sequences were obtained after a denoising and filtering procedure of the raw data. Rarefaction curves (Figure S4) demonstrated that the sequencing depth was enough to cover r-EFP culture diversity. A total of 67 amplicon ASV were obtained. Further taxonomic classiffication of these ASV showed preponderance of two bacterial phyla, Proteobacteria (87.6% of relative abundance) and Bacteroidota (11.9% of relative abundance). As can be seen in Figure 6, at order level Pseudomonadales (49.1%) were predominant, followed by Enterobacterales (23.1%), Sphingobacteriales (10.9%), Hyphomicrobiales (8.8%), Xanthomonadales (3.8%) and Burkholderiales (1.8%). Orders between 0.1% and 0.05% of relative abundance were Corynebacteriales, Micrococcales, Chitinophagales and Sphingomonadales.

Insight into bacterial diversity and catabolic potential analysis was accomplished by a shotgun metagenome approach of r-EFP culture. A total of 75250806 paired end raw sequences were obtained and after a quality filtering, trimming and assembly process a total of 90707 contigs were generated, 395 larger than 50000 bp. After annotation a total of 203798 coding DNA sequences (CDSs) were obtained. Taxonomical assigments of the reads (Figure 6) revealed a clear predominance of Pseudomonadales (48.8%), Hyphomicrobiales (5.1%), Sphingobacteriales (4,7%), Xanthomonadales (4.7%) Enterobacterales (3.7%), Sphingomonadales (2.6%), Micrococcales (2.4%) and Burkholderiales (1.8%) orders were found in more than 1% of relative abundance. Instead, for Chitinophagales and Corynebacteriales orders a minor number of sequences were found, representing 0.4% and 0.3% of relative abundance respectively.

After a binning process, eight high-quality metagenome assembled genomes were recovered (completeness ≥ 90% and contamination < 5%). Table 3 provides an overview of each MAG features such as completeness (%), contamination (%), size (bp), number of contigs, CDS and GC content (%). GC content ranged from 40,37% to 70,06%. The larger MAG size was 6531845 bp with 100% completeness, whereas the shortest one was 3174027 bp with 95,77% completeness. Taxonomic classification (Table 3) was carried out using a variety of approaches and softwares. Four MAGs were assigned to Proteobacteria (MAG3, MAG5, MAG9, MAG17), three to Bacteroidetes (MAG7, MAG8, MAG14) and one to Actinobacteria (MAG12). Regarding Proteobacteria phylum, MAG3 was classified as Gammaproteobacteria class, MAG5, MAG17 to Xanthomonadaceae familiy (MAG5, MAG17) and MAG9 to Hyphomicrobiales order. Regarding Bacteroidetes phylum, MAG7 was classified as belonging to Chitinophagia class and MAG8 and MAG14 to Sphingobacteria class, although MAG14 was further classified to Sphingobacteriaceae family. MAG12 was classified to Actinomycetia class and 16S rRNA gene sequence showed a 99% similarity with Micrococcales order.

Degradation potential of aromatic compounds in the bacterial culture enrichment was explored by examining the presence of sequences coding for proteins responsible for each step in peripheral and central intermediate metabolic pathways and evaluating which member of the

enrichment could contribute to these steps. Overall, approximately 502 annotated genes sequences related to these pathways were found. Integrating GhostKOALA annotation results, KEGG and MetaCyc database a reconstruction of nine aromatic compounds aerobic degradation described pathways was carried out (Figure 7 and 8 and table S2). Counting total sequences annotated for those pathways demonstrated that orto-cleavage of protocatechuate was the most represented pathway, followed by catechol degradation pathway via meta-cleavage and orto-cleavage (Figure 7a). Three main bacterial classes, Alpha-, Beta- and Gammaproteobacteria were found to be the mayor contributors to those pathways (Figure 7b). Among those classes, four bacterial orders contributed to more than the 73% of all sequences, these are Burkholderiales, Hyphomicrobiales, Pseudomonadales and Sphingomonadales (Figure 7c). In figure 8, it is shown that r-EFP could be able to perform three steps of naphthalene degradation to salicylic acid, mainly associated to Sphingomonadales and Burkholderiales orders. 1-Hydroxy-2-naphthoic acid central intermediate of phenanthrene and pyrene degradation could follow two degradation pathways, (1) one leading to salicylic acid, with further conversion to catechol or gentisate, and the (2) other to phthalic acid, leading to the formation of protocatechuate. Regarding the pathway leading to salicylic acid, gene sequences implicated in the conversion of 1-hydroxy-2-naphthoic to naphthalene-1,2-diol (K00480, EC:1.14.13.1, coding for a salicylate hydroxylase) were mainly assigned to Hyphomicrobiales and Burkholderiales orders. Salicylic acid conversion to catechol or gentisate and further gentisate degradation functions were mostly found in Burkholderiales and Hyphomicrobiales affilated gene sequences.

For catechol degradation through meta- and orto- cleavage annotated sequences included more bacterial orders than all the other analyzed pathways, being Burkholderiales, Hyphomicrobiales, and Pseudomonadales orders the predominant ones for both cleavage mechanisms and additionally Sphingomonadales order for meta-cleavage. Most represented functions involved in catechol degradation though meta-cleavage were K01821 and K02554, coding for a 2-hydroxymuconate tautomerase (EC:5.3.2.6) and 2-oxopent-4-enoate hydratase (EC:4.2.1.80) respectively, and though orto-cleavage was K01055 a beta-ketoadipate enol-lactone hydrolase (EC:3.1.1.24). Concerning the metabolism of 1-hydroxy-2-naphthoic acid to phthalic acid, no genes were found in r-EFP metagenome. However, genes involved in further degradation of phthalic acid to protocatechuate and the latter to TCA intermediates were annotated. Conversion to protocatechuate was mainly detected from Burkholderiales, Sphingomonadales and Hyphomicrobiales orders. Genes involved in downstream conversion for meta- and orto-cleavage protocatechuate were annotated and mainly found to be affiliated to Hyphomicrobiales and Burkholderiales for the first pathway and also Pseudomonadales and Bacillales for the latter. Most represented genes in these pathways were K10218 (coding for a 4-carboxy-4-hydroxy-2-oxoadipate aldolase, E.C: 4.1.3.17) and K01607 (coding for 4-carboxymuconolactone decarboxylase, E.C: 4.1.1.44).

In order to assess the presence of dioxygenases coding genes, AromaDeg database was used and as a result, different cluster of oxygenase families, correlating with the native substrates oxidized by their members, were obtained. Rieske non heme iron oxygenases and extradiol dioxygenases were found in r-EFP metagenome (Table 4), being the two most represented the viccinal oxygen chelate superfamily and the LigB extradiol. Concerning Rieske non heme iron oxygenases, (sequences of the biphenyl/naphthalene, phthalate, salicylate and benzoate family were found to be assigned predominately to Alpha- and Beta-proteobacteria. Referring to extradiol dioxygenases, sequences were mostly assigned to Alpha- and Gammaproteobacteria.

Functional annotation was carried out on each MAGs and the presence of genes related to aromatic compound degradation was evaluated. Table S3 shows that all MAGs possess genes involved in at least one of the analyzed pathways, however no genes coding for enzymes related to gentisate degradation were found. Additionally, PCR gene amplification was carried out with two set of primers, *nidA* and PAH-RHDα. For both sets, positive product amplification of the expected size was observed (Figure S5).

## 2.2 Isolation and characterization of pyrene-degrading bacterial strain

In order to isolate PAH degrading strains from IPK soil, a bacterial count on a pyrene-R3A plate was carried out and screened after 28 days to find a solubilizing/degradation halo zone around colonies. This halo was found on a small, bright and yellow colony. S19P6 strain was isolated.

Pyrene elimination by S19P6 strain was measured in liquid mineral medium culture supplemented with 0,05 $g.l^{-1}$ yeast extract and 100 ppm of pyrene. Figure 9a shows that after 15 days of incubation S19P6 strain was able to eliminate more than 85% of the supplemented pyrene as sole carbon and energy source and complete pyrene elimination was observed after 28 days of incubation. No pyrene elimination was observed in the abiotic control. In addition, growth curve of strain S19P6 was performed during pyrene degradation (Figure 9b). An exponential phase was observed during the first 7 days of incubation reaching three orders of magnitude higher than the beginning of the assay, followed by a stationary phase at the end of the incubation period.

### *2.2.1 Taxonomic and functional potential characterization of the isolated strain S19P6*

16S rRNA gene sequence analysis of strain S19P6 was carried out, indicating that strain had 99% similarity to sequence assigned to *Mycolicibacterium* genus and was phylogenetically related to members of that genus (Figure S6). Furthermore, 16S rRNA gene sequence was compared with ASV sequences from metabarcoding analysis of IPK soil and r-EFP culture enrichment and with 16S rRNA sequences annotated in r-EFP shotgun metagenome. As a result, 100% similarity with an ASV sequence from IPK soil and culture enrichment r-EFP (Table 5) was found.

To assess pyrene degradation potential of S19P6 strain a PCR was carried out with two sets of primers, one targeting pyrene dioxygenase gene *nidA* and the second one targeting the *pahE* codifying gene. Positive amplification (product of the expected size) was found in S19P6 for both primers sets (Table 5).

## Discussion

The impact of long-term PAH contamination on IPK soil bacterial community was assessed through culture-dependent and independent complementary approaches, ranging from enrichment culture and pure culture studies focusing on its metabolic potential, to community-scale research trying to determine the link between the diversity and functional capabilities of PAH-degraders. To identify key functions and microorganisms involved in the degradation of environmental pollutants is essential to understand the ecological mechanisms of environmental recovery (Gomes *et al.*, 2010; Gannes and Hickey, 2017).

The introduction of xenobiotic compounds generates selective loads to the microbial communities in develop mechanism for degradation of such compounds (Gosai *et al.*, 2018). In

biodegradation studies, the term adaptation refers to a phenomenon that can occur both in mixed microbial communities and in individual microbial lineages. The mechanisms that would allow adaptation, proposed by van der Meer *et al.*, (1992), are: the induction and / or repression of specific enzymes and / or the enrichment of the microbial population capable of transforming the compound of interest. It was demonstrated by molecular and culture approaches that IPK bacterial community is a highly impacted community. By 16S rRNA metabarcoding analysis a low diversity community was observed with two predominant phyla Proteobacteria and Actinobacteria (Figure 1 and Table S1). Previous studies evaluating diversity of PAH contaminated soils also reported the predominance of the same phylum found in this study (Haleyur *et al.*, 2019; Lu *et al.*, 2019); hydrocarbons tend to suppress certain sensitive groups to select primarily for subgroups of the Actinobacteria and Proteobacteria (Stefani *et al.*, 2015). Regarding predominant orders, Immundisolibacterales (14,80% of relative abundance) and Microtrichales (15.30% of relative abundance) were found (Figure 1). Corteselli, Aitken and Singleton, (2017) reported that a member of Immundisolibacterales order was able to grow on pyrene, phenanthrene, anthracene, benz[a]anthracene and fluorene as a sole carbon and energy source under mesophilic temperature, neutral pH and low salinity conditions. In contrast, no link to PAH degradation was found in literature for Microtrichales order. Shotgun metagenomics results showed that IPK community present a clear predominance of Streptomycetales order (Figure 3a) with almost all annotated genes assigned to this order (Figure 3b). In accordance with the present results, previous studies have reported that Streptomycetales order is widely found in hydrocarbon polluted, with a substantial role in soil ecology and being encountered to be the most abundant isolated genus (Anandan, Dharumadurai and Manogaran, 2016; Baoune *et al.*, 2019).

It can therefore be notice that taxonomic classification carried out with both molecular approaches was divergent (Figure 1 and 3a). A possible explanation could be the difference in 16S rRNA gene copy number and genome size of the members of the community members. Members of Proteobacteria are usually considered to be fast-growing bacteria (r-strategists) with high 16S rRNA gene copy number, whereas Actinobacteria are mostly slow-growing organisms (K-strategists) with low 16S rRNA gene copy number (Klappenbach, Dunbar and Schmidt, 2000; Brzeszcz *et al.*, 2016; Pereira-Flores, Glöckner and Fernandez-Guerra, 2019). Abundance estimation based on the 16S rRNA sequence counts tend to underestimate the abundance of taxa with low 16S rRNA copy numbers and to overestimate taxa with high 16S rRNA copy numbers (Větrovský and Baldrian, 2013). Microorganisms related to Streptomycetales order have linearized chromosomes commonly over 8 Mbp with a high G+C content (Law *et al.*, 2019), representing a challenge for PCR amplification(Baldrian, 2019). Besides, a larger genome has greater possibilities of being sampled than a smaller genome, even if they occur in equal abundances, as it will produce more fragments per genome (Angly *et al.*, 2009). A mayor influence in community analysis is the abundance of the members of the community but also their genome sizes and 16S rRNA gene copy numbers of the abundant microorganisms are important driving factors of the analysis. Evaluation using only 16S rRNA gene fragment abundances should be done if community members share characteristics close to the average genome size and 16S rRNA gene copy number (Frank and Sørensen, 2011). Other potential source of variation in microbial community profiling could be the DNA extraction approaches and the 16S rRNA gene variable region primers used for sequencing (Han *et al.*, 2020). In this study not only DNA extractions were carried out using the same commercial kit,

but also other authors could detect Streptomycetaceae family using the same pair of primers employed for 16S rRNA gene sequencing (Ho, Di Lonardo and Bodelier, 2017; Kuang *et al.*, 2018; BenIsrael *et al.*, 2020; Shimoi *et al.*, 2020).

Considering the above-mentioned limitation of both techniques, to further characterize IPK soil community two culture approaches were performed, a culture enrichment and a strain isolation. Microbial enrichment are natural microbiomes cultivated *ex situ* in a defined growth medium under well-controlled conditions that, by means of sequential transfers with a selective pressure, yield a specialized microbial consortium, where most niches are constructed by the microorganisms in the original community (Jiménez, Mares and Salles, 2018; Estrela, Sánchez and Rebolleda-Gómez, 2021). Because functional genes belonging to members of IPK bacterial community other than Streptomycetales order were disguised by the latter and taking advantage of the increased accessibility and tractability of these cultures compared to natural communities (Blasche *et al.*, 2017), the potential functions related to aromatic compound degradation of IPK soil community were also assessed by an enrichment culture and by an isolated strain S19P6. Previously, Gomes *et al.*, (2010) took advantage of this kind of cultures to study, by fingerprint analyses, potential functions present in an enriched community and undetected in the source environment.

A pyrene degrading enrichment culture r-EFP was obtained from soil. After confirming pyrene elimination (Figure 5), a 16SrRNA gene metabarcoding analysis, and a metagenomic approach was carried out to assess diversity and functional potential of r-EFP. Regarding r-EFP microbial taxa, contrary to IPK soil results obtained in this research, bacterial phylum and orders agreed across sequencing methodologies (Figure 6), as reported by other authors (Jovel *et al.*, 2016; Brumfield *et al.*, 2020). Both, 16SrRNA gene metabarcoding and shotgun metagenomic approaches, indicated a clear predominance of Proteobacteria Phyla, being Pseudomonadales order the one present in almost 50% of relative abundance, a vastly known PAH degrading related order (Ma, Xu and Jia, 2013; Ghosh, Jasmine and Mukherji, 2014; Swati, Ghosh and Thakur, 2019). Although bacterial orders were common in both approaches differences in relative abundances were observed. Members of the other bacterial orders found in r-EFP previously related to pyrene degradation by other authors were Sphingomonadales (Hesham et al., 2014; Vila, Tauler and Grifoll, 2015), Burkholderiales (Al-Thukair and Malik, 2016), Corynebacterales (Wu et al., 2019), Enterobacterales (Gupta et al., 2020), Micrococcales (Gauthier et al., 2003), Sphingobacteriales (Zhao et al., 2013), Xanthomonadales (Mangwani et al., 2014) and Hyphomicrobiales (Ortega-González *et al.*, 2015). With regards to *Chitinophagales* orders, no isolated pyrene degrading microorganism related to this order was found. As reported by Kwon, Kwon and Kim, (2019) only thirteen isolates from the phylum Bacteroidetes were confirmed to degrade hydrocarbon, but none from the order mentioned above. When focusing on potential functions related to aromatic compound degradation, members of r-EFP enrichment culture held genes codifying for enzymes responsible for the degradation of naphthalene, 1-hydroxy-2-naphthoic acid, salicylic acid, gentisate, phthalic acid, catechol and protocatechuate (Figure 7 and 8). Predominant pathway in IPK soil predicted by PICRUSt2 was concordant with the results obtained with r-EFP metagenome. Functional profiling with both approaches demonstrated that protocatechuate degradation via ortho-cleavage was the most represented pathway not only in r-EFP metagenome (Figure 7) but also in IPK soil PICRUSt2 prediction (Figure 2a). By both approaches K01607 (Figure 2b and 8) (4-carboxymuconolactone decarboxylase) was found to be represented by members of Proteobacteria and Actinobacteria,

being Hyphomicrobiales order one of the mayor contributors. No genes related to this function was found in IPK soil shotgun metagenome. Regardless not having found genes responsible for the conversion of 1-hydroxy-2-naphthoic acid to phtalic acid, sequences related to the degradation of the latter compound were found (Figure 8) so, degradation could occur also via phthalic acid. Sequences related to dioxygenase coding genes belong primarily to Alpha-, Beta- and Gammaproteobacteria classes (Figure 8). The presence of dioxygenases coding sequences was demonstrated by shotgun metagenome analysis (Table 4) but also by PCR using specific primer sets designed to amplify *nidA* gene for *Mycobacterium* genus (Debruyn, Chewning and Sayler, 2007) and genes codifying for the alpha subunit of the PAH-ring hydroxylating dioxygenases (Cébron *et al.*, 2008) (Figure S5 and S6). The positive amplification in *nidA* PCR could be due to, despite being in very low relative abundance, Corynebacterales order is present in r-EFP (Figure 6) and so is a genus of this order, *Mycobacterium*, as demonstrated by the similarity found in S19P6 isolated strain 16S rRNA gene with an ASV found in r-EFP (Table 5). Furthermore, *nidA* gene could be in the genome of a gram-negative bacteria in r-EFP as demonstrated by Klankeo et al., (2007) who found the presence of *nidA* gene in a pyrene degrading Pseudoxanthomonas sp. isolated from soil. As for PAH-RHDα PCR positive result, used primer set was designed to amplify genes common to Gram negative PAH degrader (Cébron *et al.*, 2008) including sequences belonging to Pseudomonadales, Sphingomonadales and Burkholderiales order, all of them found in r-EFP (Figure 6). Binning of the r-EFP metagenome reveled eight high quality metagenome assembled genomes (completeness ≥90% and contamination ≤ 5%) (Table 3). Interestingly, none of them were taxonomically assigned to Pseudomonadales predominant order but to orders present in minor abundance in the enrichment culture; this result is contrary to a previous study which reconstructed high quality MAGs from the most abundant OTUs identified in a biphenyl-degrading soil bacterial consortium (Daniel Garrido-Sanz, Javier Manzano, Marta Martín, 2018). Sequences related to the analyzed aromatic compound degradation annotated pathways were found in the recovered MAGs. Taxa contribution to the initial pyrene degradation steps was not possible to determine as enzymes involved have no functional gene numbers (Wang *et al.*, 2016). Even more, the observed functional redundancy did not allow to assert which microorganisms present in r-EFP could perform and ecological function or not (Rodríguez *et al.*, 2016), so isolating members of r-EFP could be of great help to understand the ecological niche of each in the degradation of pyrene.

Furthermore, a pyrene degrading strain S19P6, identified as a member of *Mycolicibacterium* genus (Table 5), was isolated from IPK soil demonstrating ability to grow and degrade pyrene as sole carbon and energy source after 28 days of incubation (Figure 9a and b). Also, the presence of *nidA* and *pahE* gene found by PCR in S19P6 strain further confirm the pyrene degradation capacity (Table 5). pahE gene product catalyzes a significantly important step (the fifth) of the aerobic PAH degradation pathway; its presence could be indicative that bacteria could grow solely on PAH, not only add oxygen to this compound (Liang, Huang and Wang, 2019). pahE gene has recently been postulated as the best functional marker to indicate PAH degradation (Liang *et al.*, 2022). Among many bacterial isolates capable of degrading pyrene, the genus *Mycobacterium*, which can utilize pyrene as a sole carbon and energy source, has been most widely studied (Kim *et al.*, 2018; Yang *et al.*, 2021). Interestingly, neither the S19P6 isolated strain nor the members of r-EFP were present between the most abundant taxa found by 16S rRNA metabarcoding (Figure 1) nor shotgun metagenomics (Figure 3) in IPK soil. This result further supports the existing gap between culture dependent and independent approaches

previously described by other researchers. Several reports have shown that not only the most abundant microorganisms detected by culture independent techniques were not isolated by culture approaches (Stefani *et al.*, 2015) but also that microorganisms captured by culturing method are present in low abundance or absent in the sequencing analysis (Lee *et al.*, 2016; Wang *et al.*, 2017). The enrichment and pure culture isolation strategies tend to favor the growth of members with r-strategies (Alain and Querellou, 2009) and also the "conditional rare taxa". This kind of rare taxa could respond to natural or laboratory selective pressures during culture enrichments becoming abundant by the changing conditions (Pascoal, Magalhães and Costa, 2020). It was demonstrated that culture-based methodologies could complement and increase knowledge obtained from independent-culture ones (Oita *et al.*, 2021) when characterizing soil microbiomes.

## Conclusion

Previous bioremediation strategies on IPK soil as allochthons bioaugmentation and surfactant-enhanced bioremediation were assayed and a limitation in the degradation of contaminants was observed (Festa *et al.*, 2016; Cecotti *et al.*, 2018). However, research carried out in this work demonstrated that members of the bacterial community of the analyzed chronically PAH contaminated soil have the genetic potential related to aromatic compound degradation. For the abovementioned reason, other bioremediation strategies to trigger this non-active potential need to be tested, such as the addition of easily accessible carbon source (biostimulation) or/and autochthonous bioaugmentation with the obtained cultures.

## Acknowledgments

This work was financially supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Agencia Nacional de Promoción Científica y Tecnológica (PICT 2019-01805, PICT 2018-1889) and United Nations University for Biotechnology of Latin America and Caribe (UNU-BIOLAC).

## Declaration of Competing Interest

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

## Figure legends

**Figure 1.** Bacterial composition (relative abundance) at phylum and order level (>3% of relative abundance) in IPK soil.

**Figure 2a.** Bacterial classes contribution to metabolic pathways related to the degradation of aromatic compounds according to PICRUSt2 prediction. Results are shown as a mean of triplicate independent samples. P1:Catechol degradation to β-ketoadipate; P2:Protocatechuate degradation (meta-cleavage pathway); P3:Protocatechuate degradation (ortho-cleavage pathway); P4:Catechol degradation (meta-cleavage pathway); P5:Catechol degradation (ortho-cleavage pathway); P6:Catechol degradation to 2-oxopent-4-enoate II; P7:Catechol degradation (meta-cleavage pathway); P9:Superpathway of salicylate degradation.

**Figure 2b**. Sankey diagram showing the absolute contribution of chosen PICRUSt2 predicted KO by bacterial orders (orders with less contribution than 10% were grouped as "Other orders"). Vertical nodes are proportional to size. Green node: functions related to Protocatechuate degradation (ortho-cleavage pathway); yellow nodes: functions related to Catechol degradation (meta-cleavage pathway); red nodes: orders belonging to Actinobacteria Phylum; blue nodes: orders belonging to Proteobacteria Phylum; grey node: other orders.

**Figure 3a**. Relative abundance (%) of bacterial orders in IPK soil assessed by shotgun metagenomic (orders with less than 0.01% of relative abundance were grouped as "Other orders").

**Figure 3b**. Barplot showing the number of genes annotated in IPK soil metagenome and its taxonomic assignment.

**Figure 4.** Metabolic network plot showing the aromatic compounds degradation related functions found in IPK shotgun metagenome. Line width: completeness of the pathway involved; The line color indicates type of dioxygenase found in the metagenome to be participating in that pathway (pink line: catechol 1.2-dioxygenase; light blue line: protocatechuate 2.3-dioxygenase; green line: Rieske non-heme iron oxygenases family dioxygenase). Key pathway intermediates are shown in red.

**Figure 5.** Pyrene concentration ($mg.l^{-1}$) in the r-EFP culture growing in LMM with pyrene as a sole carbon and energy source during 21 days incubation. Results are means of triplicate independent experiments. Bars represent standard deviations.

**Figure 6.** Bacterial composition (relative abundance) at order level revealed by 16S rRNA gene sequencing (>0.05% of relative abundance) and shotgun metagenome (>0.2% of relative abundance) analysis in r-EFP.

**Figure 7.** Heatmaps showing a) the total count of sequences related to aromatic compound degradation pathways in the r-EFP shotgun metagenome and b) bacterial classes and c) orders contributing to those pathways

**Figure 8.** Heatmap showing bacterial orders contribution to KEGG orthologs involved in aromatic compound degradation found in r-EFP shotgun metagenome using GhostKOALA and AromaDeg database.

**Figure 9**. a) Pyrene concentration ($mg.l^{-1}$) in the S19P6 cultures growing in LMM with pyrene as a sole carbon and energy source during 28 days of incubation. b) Growth of S19P6 strain on pyrene as an only carbon and energy source in LMM during 28 days of incubation. Results are means of triplicate independent experiments. Bars represent standard deviations.

**Figure S1.** Rarefaction curves of IPK sequencing data indicating the observed features (y-axis) as a function of sequencing depth (x-axis).

**Figure S2.** Estimated average coverage of the IPK metagenomic dataset using Nonpareil Software.

**Figure S3.** Agarose gel electrophoresis of PCR products of *nidA* gene of IPK soil sample.

**Figure S4.** Rarefaction curves of r-EFP sequencing data indicating the observed features (y-axis) as a function of sequencing depth (x-axis).

**Figure S5.** Agarose gel electrophoresis of PCR products of *nidA* gene in r-EFP (r-E).

**Figure S6.** Neighbor-joining tree based on a distance matrix analysis of 16S rDNA partial sequences of S19P6 isolated strain and selected sequences available in GenBank was constructed using the MEGA software version 4. The numbers at each node correspond to the bootstrap per cent values. The space bar indicates 0.01 sequence variation.

## Supplementary material

## References


Al-Thukair, A.A. and Malik, K. (2016) 'Pyrene metabolism by the novel bacterial strains Burkholderia fungorum (T3A13001) and Caulobacter sp (T2A12002) isolated from an oil-polluted site in the Arabian Gulf', *International Biodeterioration and Biodegradation*, 110, pp. 32–37. doi:10.1016/j.ibiod.2016.02.005.

Alain, K. and Querellou, J. (2009) 'Cultivating the uncultured: Limits, advances and future challenges', *Extremophiles*, 13(4), pp. 583–594. doi:10.1007/S00792-009-0261-3/FIGURES/3.

Anandan, R., Dharumadurai, D. and Manogaran, G.P. (2016) 'An Introduction to Actinobacteria', *Actinobacteria: Basic and Biotechnology Applications*, pp. 1–36. doi:http://dx.doi.org/10.5772/57353.

Angly, F.E. *et al.* (2009) 'The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes', *PLoS Computational Biology*, 5(12). doi:10.1371/journal.pcbi.1000593.

Aramaki, T. *et al.* (2020) 'KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold', *Bioinformatics*, 36(7), pp. 2251–2252. doi:10.1093/bioinformatics/btz859.

Baldrian, P. (2019) 'The known and the unknown in soil microbial ecology', *FEMS Microbiology Ecology*, 95(2), pp. 1–9. doi:10.1093/femsec/fiz005.

Bankevich, A. *et al.* (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology*, 19(5), pp. 455–477. doi:10.1089/cmb.2012.0021.

Baoune, H. *et al.* (2019) 'Bioremediation of petroleum-contaminated soils using Streptomyces sp. Hlh1', *Journal of Soils and Sediments*, 19(5), pp. 2222–2230. doi:10.1007/s11368-019-02259-w.

Bartelme, R.P. *et al.* (2020) 'Influence of Substrate Concentration on the Culturability of Heterotrophic Soil Microbes Isolated by High-Throughput Dilution-to-Extinction Cultivation', 5(1), pp. 1–15.

Bastiaan von Meijenfeldt, F.A. *et al.* (2019) 'Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT', *Genome biology*, 20, p. 217. doi:10.1186/s13059-019-1817-x.

BenIsrael, M. *et al.* (2020) 'Quantification of toluene phytoextraction rates and microbial biodegradation functional profiles at a fractured bedrock phytoremediation site', *Science of the Total Environment*, 707. doi:10.1016/j.scitotenv.2019.135890.

Besaury, L. *et al.* (2013) 'Culture-Dependent and Independent Studies of Microbial Diversity in

Highly Copper-Contaminated Chilean Marine Sediments', *Microbial Ecology* [Preprint]. doi:10.1007/s00248-012-0120-0.

Blasche, S. *et al.* (2017) 'Model microbial communities for ecosystems biology', *Current Opinion in Systems Biology*, 6, pp. 51–57. doi:10.1016/J.COISB.2017.09.002.

Blum, M. *et al.* (2021) 'The InterPro protein families and domains database: 20 years on', *Nucleic Acids Research*, 49(D1), pp. D344–D354. doi:10.1093/nar/gkaa977.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics* [Preprint]. doi:10.1093/bioinformatics/btu170.

Bolyen, E. *et al.* (2019) 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*, 37, pp. 852–857. doi:10.1038/s41587-019-0190-3.

Brumfield, K.D. *et al.* (2020) 'Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data', (July 2014), pp. 1–21.

Brzeszcz, J. *et al.* (2016) 'r-strategist versus K-strategist for the application in bioremediation of hydrocarbon-contaminated soils', *International Biodeterioration and Biodegradation* [Preprint]. doi:10.1016/j.ibiod.2015.10.001.

Caspi, R. *et al.* (2018) 'The MetaCyc database of metabolic pathways and enzymes', *Nucleic Acids Research*, 46(D1), pp. D633–D639. doi:10.1093/nar/gkx935.

Cébron, A. *et al.* (2008) 'Real-Time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHD α ) genes from Gram positive and Gram negative bacteria in soil and sediment samples', *Journal of Microbiological Methods* [Preprint]. doi:10.1016/j.mimet.2008.01.009.

Cecotti, M. *et al.* (2018) 'Efficiency of surfactant-enhanced bioremediation of aged polycyclic aromatic hydrocarbon-contaminated soil: Link with bioavailability and the dynamics of the bacterial community', *Science of the Total Environment*, 634, pp. 224–234. doi:10.1016/j.scitotenv.2018.03.303.

Corteselli, E.M., Aitken, M.D. and Singleton, D.R. (2017) 'Description of Immundisolibacter cernigliae gen. Nov., sp. nov., a high-molecular-weight polycyclic aromatic hydrocarbondegrading bacterium within the class Gammaproteobacteria, and proposal of Immundisolibacterales ord. nov. and Immundisolibacteraceae fa', *International Journal of Systematic and Evolutionary Microbiology*, 67(4), pp. 925–931. doi:10.1099/ijsem.0.001714.

Daniel Garrido-Sanz, Javier Manzano, Marta Martín, M.R.-N. and R.R. (2018) 'Metagenomic Analysis of a Biphenyl-Degrading Soil Bacterial Consortium Reveals the Metabolic Roles of Specific Populations', 9. doi:10.3389/fmicb.2018.00232.

Debruyn, J.M., Chewning, C.S. and Sayler, G.S. (2007) 'Comparative quantitative prevalence of Mycobacteria and functionally abundant nidA, nahAc, and nagAc dioxygenase genes in coal tar contaminated sediments', *Environmental Science and Technology*, 41(15), pp. 5426–5432. doi:10.1021/es070406c.

Desai, C., Pathak, H. and Madamwar, D. (2010) 'Advances in molecular and "-omics" technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites', *Bioresource Technology* [Preprint]. doi:10.1016/j.biortech.2009.10.080.

Douglas, G.M. *et al.* (2020) 'PICRUSt2 for prediction of metagenome functions', *Nature Biotechnology*, 38(6), pp. 669–673. doi:10.1038/s41587-020-0550-z.

Duarte, M. *et al.* (2014) 'AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics', *Database* [Preprint]. doi:10.1093/database/bau118.

Estrela, S., Sánchez, Á. and Rebolleda-Gómez, M. (2021) 'Multi-Replicated Enrichment Communities as a Model System in Microbial Ecology', *Frontiers in Microbiology*, 12(April).

doi:10.3389/fmicb.2021.657467.

Festa, S. *et al.* (2016) 'Monitoring the impact of bioaugmentation with a PAH-degrading strain on different soil microbiomes using pyrosequencing', *FEMS Microbiology Ecology* [Preprint]. doi:10.1093/femsec/fiw125.

Festa, S., Coppotelli, B.M. and Morelli, I.S. (2013) 'Bacterial diversity and functional interactions between bacterial strains from a phenanthrene-degrading consortium obtained from a chronically contaminated-soil', *International Biodeterioration and Biodegradation* [Preprint]. doi:10.1016/j.ibiod.2013.06.006.

Festa, S., Coppotelli, B.M. and Morelli, I.S. (2016) 'Comparative bioaugmentation with a consortium and a single strain in a phenanthrene-contaminated soil: Impact on the bacterial community and biodegradation', *Applied Soil Ecology* [Preprint]. doi:10.1016/j.apsoil.2015.08.025.

Frank, J.A. and Sørensen, S.J. (2011) 'Quantitative metagenomic analyses based on average genome size normalization', *Applied and Environmental Microbiology*, 77(7), pp. 2513–2521. doi:10.1128/AEM.02167-10.

Gannes, V. de and Hickey, W.J. (2017) *Genetic Adaptations of Bacteria for Metabolism of Polycyclic Aromatic Hydrocarbons*, *Microbial Ecotoxicology*. doi:10.1007/978-3-319-61795-4.

Gauthier, E. *et al.* (2003) 'Initial characterization of new bacteria degrading high-molecular weight polycyclic aromatic hydrocarbons isolated from a 2-year enrichment in a two-liquid-phase culture system', *Journal of Applied Microbiology*, 94(2), pp. 301–311.

Ghosh, I., Jasmine, J. and Mukherji, S. (2014) 'Biodegradation of pyrene by a Pseudomonas aeruginosa strain RS1 isolated from refinery sludge', *Bioresource Technology*, 166, pp. 548–558. doi:10.1016/j.biortech.2014.05.074.

Gomes, N.C.M. *et al.* (2010) 'Mangrove microniches determine the structural and functional diversity of enriched petroleum hydrocarbon-degrading consortia', *FEMS Microbiology Ecology*, 74(2), pp. 276–290. doi:10.1111/j.1574-6941.2010.00962.x.

Gosai, H.B. *et al.* (2018) 'Concentrations, input prediction and probabilistic biological risk assessment of polycyclic aromatic hydrocarbons (PAHs) along Gujarat coastline', *Environ Geochem Health*, 40, pp. 653–665. doi:10.1007/s10653-017-0011-x.

Gupta, B. *et al.* (2020) 'Comparative evaluation of growth kinetics for pyrene degradation by Acinetobacter pittii NFL and Enterobacter cloacae BT in the presence of biosurfactant', *Bioresource Technology Reports*, 9(November 2019), p. 100369. doi:10.1016/j.biteb.2019.100369.

Gurevich, A. *et al.* (2013) 'QUAST: quality assessment tool for genome assemblies.', *Bioinformatics (Oxford, England)*, 29(8), pp. 1072–1075. doi:10.1093/bioinformatics/btt086.

Haleyur, N. *et al.* (2019) 'Influence of bioaugmentation and biostimulation on PAH degradation in aged contaminated soils: Response and dynamics of the bacterial community', *Journal of Environmental Management*, 238(March), pp. 49–58. doi:10.1016/j.jenvman.2019.02.115.

Han, D. *et al.* (2020) 'Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing', *Journal of Advanced Research*, 26(1), pp. 111–121. doi:10.1016/j.jare.2020.07.010.

Harms, G. *et al.* (2003) 'Real-time PCR quantification of nitrifying bacteria in a municipal wastewater treatment plant', *Environmental Science and Technology* [Preprint]. doi:10.1021/es0257164.

Hesham, A.E.L. *et al.* (2014) 'Biodegradation ability and catabolic genes of petroleum-degrading Sphingomonas koreensis strain ASU-06 isolated from Egyptian oily soil', *BioMed*

*Research International*, 2014. doi:10.1155/2014/127674.

Ho, A., Di Lonardo, D.P. and Bodelier, P.L.E. (2017) 'Revisiting life strategy concepts in environmental microbial ecology', *FEMS Microbiology Ecology*, 93(3), pp. 1–14. doi:10.1093/femsec/fix006.

Jiménez, D.J., Mares, M.C. De and Salles, J.F. (2018) 'Temporal expression dynamics of plant biomass-degrading enzymes by a synthetic bacterial consortium growing on sugarcane bagasse', *Frontiers in Microbiology*, 9(FEB), pp. 1–13. doi:10.3389/fmicb.2018.00299.

Jovel, J. *et al.* (2016) 'Characterization of the gut microbiome using 16S or shotgun metagenomics', *Frontiers in Microbiology*, 7(APR), pp. 1–17. doi:10.3389/fmicb.2016.00459.

Kanehisa, M., Sato, Y. and Morishima, K. (2016) 'BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences', *Journal of Molecular Biology*, 428(4), pp. 726–731. doi:https://doi.org/10.1016/j.jmb.2015.11.006.

Kim, D.W. *et al.* (2018) 'Comparative genomic analysis of pyrene-degrading Mycobacterium species: Genomic islands and ring-hydroxylating dioxygenases involved in pyrene degradation', *Journal of Microbiology*, 56(11), pp. 798–804. doi:10.1007/s12275-018-8372-0.

Klappenbach, J.A., Dunbar, J.M. and Schmidt, T.M. (2000) 'rRNA operon copy number reflects ecological strategies of bacteria', *Applied and Environmental Microbiology*, 66(4), pp. 1328–1333. doi:10.1128/AEM.66.4.1328-1333.2000.

Kuang, S. *et al.* (2018) 'Soil microbial community structure and diversity around the aging oil sludge in yellow river delta as determined by high-throughput sequencing', *Archaea*, 2018. doi:10.1155/2018/7861805.

Kwon, K., Kwon, Y.M. and Kim, S.J. (2019) 'Aerobic Hydrocarbon-Degrading Bacteroidetes', in *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*, pp. 73–91.

Langille, M.G.I. *et al.* (2013) 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences', *Nature Biotechnology*, 31(9), pp. 814–821. doi:10.1038/nbt.2676.

Law, J.W.-F. *et al.* (2019) 'A Review on Mangrove Actinobacterial Diversity: The Roles of Streptomyces and Novel Species Discovery', *Progress In Microbes & Molecular Biology*, 2(1), pp. 1–10. doi:10.36877/pmmb.a0000024.

Lawson, C.E. *et al.* (2019) 'Common principles and best practices for engineering microbiomes', *Nature Reviews Microbiology*, 17(12), pp. 725–741. doi:10.1038/s41579-019-0255-9.

Lee, S.A. *et al.* (2016) 'Comparative analysis of bacterial diversity in the rhizosphere of tomato by culture-dependent and -independent approaches', *Journal of Microbiology*, 54(12), pp. 823–831. doi:10.1007/s12275-016-6410-3.

Li, D. *et al.* (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics*, 31(10), pp. 1674–1676. doi:10.1093/bioinformatics/btv033.

Liang, C. *et al.* (2022) 'Shifts of the new functional marker gene (pahE) of polycyclic aromatic hydrocarbons (PAHs) degrading bacterial population and its relationship with PAHs biodegradation', *Journal of Hazardous Materials*, 437, p. 129305. doi:10.1016/J.JHAZMAT.2022.129305.

Liang, C., Huang, Y. and Wang, H. (2019) 'pahE, a Functional Marker Gene for Polycyclic Aromatic Hydrocarbon-Degrading Bacteria', 85(3), pp. 1–17.

Lu, C. *et al.* (2019) 'A PAH-degrading bacterial community enriched with contaminated agricultural soil and its utility for microbial bioremediation', *Environmental Pollution*, 251, pp. 773–782. doi:10.1016/j.envpol.2019.05.044.

Ma, J., Xu, L. and Jia, L. (2013) 'Characterization of pyrene degradation by Pseudomonas sp. strain Jpyr-1 isolated from active sewage sludge', *Bioresource Technology*, 140, pp. 15–21. doi:10.1016/j.biortech.2013.03.184.

Madueño, L. *et al.* (2018) 'Insights into the mechanisms of desiccation resistance of the Patagonian PAH-degrading strain Sphingobium sp. 22B', *Journal of Applied Microbiology*, 124(6), pp. 1532–1543. doi:10.1111/jam.13742.

Mangwani, N. *et al.* (2014) 'Characterization of Stenotrophomonas acidaminiphila NCW-702 biofilm for implication in the degradation of polycyclic aromatic hydrocarbons', *Journal of Applied Microbiology*, 117(4), pp. 1012–1024. doi:https://doi.org/10.1111/jam.12602.

van der Meer, J.R. *et al.* (1992) 'Molecular mechanisms of genetic adaptation to xenobiotic compounds', *Microbiological reviews*, 56(4), pp. 677–694. doi:10.1128/MR.56.4.677-694.1992.

Miller, C.D. *et al.* (2004) 'Isolation and characterization of polycyclic aromatic hydrocarbon-degrading mycobacterium isolates from soil', *Microbial Ecology*, 48(2), pp. 230–238. doi:10.1007/s00248-003-1044-5.

Mishra, S. *et al.* (2021) 'Recent Advanced Technologies for the Characterization of Xenobiotic-Degrading Microorganisms and Microbial Communities', *Frontiers in Bioengineering and Biotechnology*, 9(February). doi:10.3389/fbioe.2021.632059.

Mora, V.C. *et al.* (2014) 'Remediation of phenanthrene-contaminated soil by simultaneous persulfate chemical oxidation and biodegradation processes', *Environmental Science and Pollution Research*, 21(12), pp. 7548–7556.

Muralidharan, H.S. *et al.* (2021) 'Binnacle: Using Scaffolds to Improve the Contiguity and Quality of Metagenomic Bins', *Frontiers in Microbiology*, 12(February), pp. 1–15. doi:10.3389/fmicb.2021.638561.

Oita, S. *et al.* (2021) 'Methodological Approaches Frame Insights into Endophyte Richness and Community Composition', *Microbial Ecology*, 82(1), pp. 21–34. doi:10.1007/s00248-020-01654-y.

Ortega-González, D.K. *et al.* (2015) 'Evaluation of the Removal of Pyrene and Fluoranthene by Ochrobactrum anthropi, Fusarium sp. and Their Coculture', *Applied Biochemistry and Biotechnology*, 175(2), pp. 1123–1138. doi:10.1007/s12010-014-1336-x.

Del Panno, M.T. *et al.* (2005) 'Effect of petrochemical sludge concentrations on microbial communities during soil bioremediation', *FEMS Microbiology Ecology* [Preprint]. doi:10.1016/j.femsec.2005.01.014.

Parks, D.H. *et al.* (2015) 'CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes', *Genome Research*, 25(7), pp. 1043–1055. doi:10.1101/GR.186072.114.

Pascoal, F., Magalhães, C. and Costa, R. (2020) 'The Link Between the Ecology of the Prokaryotic Rare Biosphere and Its Biotechnological Potential', *Frontiers in Microbiology*, 11(February). doi:10.3389/fmicb.2020.00231.

Peng, J.J. *et al.* (2010) 'Dynamic changes in functional gene copy numbers and microbial communities during degradation of pyrene in soils', *Environmental Pollution*, 158(9), pp. 2872–2879. doi:10.1016/j.envpol.2010.06.020.

Pereira-Flores, E., Glöckner, F.O. and Fernandez-Guerra, A. (2019) 'Fast and accurate average genome size and 16S rRNA gene average copy number computation in metagenomic data', *BMC bioinformatics*, 20(1), p. 453. doi:10.1186/s12859-019-3031-y.

Ponomarova, O. and Patil, K.R. (2015) 'Metabolic interactions in microbial communities: Untangling the Gordian knot', *Current Opinion in Microbiology* [Preprint].

doi:10.1016/j.mib.2015.06.014.

Prakash, O. *et al.* (2013) 'Microbial cultivation and the role of microbial resource centers in the omics era', *Applied Microbiology and Biotechnology* [Preprint]. doi:10.1007/s00253-012-4533-y.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2014) 'Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets', *Bioinformatics*, 30(5), pp. 629–635. doi:10.1093/BIOINFORMATICS/BTT584.

Rodríguez, R.A. *et al.* (2016) 'Response to comments on "Uncertainty principle in niche assessment: A solution to the dilemma redundancy vs. competitive exclusion, and some analytical consequences"', *Ecological Modelling* [Preprint]. doi:10.1016/j.ecolmodel.2016.09.014.

Scoma, A. *et al.* (2017) 'Self-healing capacity of deep-sea ecosystems affected by petroleum hydrocarbons', *EMBO reports*, 18(6), pp. 868–872. doi:10.15252/EMBR.201744090.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. doi:10.1093/bioinformatics/btu153.

Segata, N. *et al.* (2013) 'PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes.', *Nature communications*, 4, p. 2304. doi:10.1038/ncomms3304.

Shimoi, Y. *et al.* (2020) 'Effects of chitin degradation products N-acetylglucosamine and N,N'-diacetylchitobiose on chitinase activity and bacterial community structure in an incubated upland soil', *Soil Science and Plant Nutrition*, 66(3), pp. 429–437. doi:10.1080/00380768.2020.1767488.

Souza, G.D., Shitut, S. and Kost, C. (2018) 'Natural Product Reports interactions in bacteria †', *Natural Product Reports*, 35, pp. 455–488. doi:10.1039/C8NP00009C.

Stefani, F.O.P. *et al.* (2015) 'Culture-dependent and -independent methods capture different microbial community fractions in hydrocarbon-contaminated soils', *PLoS ONE* [Preprint]. doi:10.1371/journal.pone.0128272.

Suenaga, H. *et al.* (2014) 'Diversity of extradiol dioxygenases in aromatic-degrading microbial community explored using both culture-dependent and culture-independent approaches', *FEMS Microbiology Ecology* [Preprint]. doi:10.1111/1574-6941.12390.

Swati, Ghosh, P. and Thakur, I.S. (2019) 'Biodegradation of pyrene by Pseudomonas sp. ISTPY2 isolated from landfill soil: Process optimisation using Box-Behnken design model', *Bioresource Technology Reports*, 8(October), p. 100329. doi:10.1016/j.biteb.2019.100329.

Vecchioli, G.I., Del Panno, M.T. and Painceira, M.T. (1990) 'Use of selected autochthonous soil bacteria to enhanced degradation of hydrocarbons in soil.', *Environ. Pollut.*, 67, pp. 249–258.

Větrovský, T. and Baldrian, P. (2013) 'The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses', *PLoS ONE* [Preprint]. doi:10.1371/journal.pone.0057923.

Vila, J., Tauler, M. and Grifoll, M. (2015) 'Bacterial PAH degradation in marine and terrestrial habitats', *Current Opinion in Biotechnology*, 33, pp. 95–102. doi:10.1016/j.copbio.2015.01.006.

Wang, B. *et al.* (2017) 'Environment Comparative analysis of microbial communities during enrichment and isolation of DDT-degrading bacteria by culture-dependent and -independent methods', *Science of the Total Environment*, 590–591, pp. 297–303. doi:10.1016/j.scitotenv.2017.03.004.

Wang, H. *et al.* (2016) 'Co-acclimation of bacterial communities under stresses of hydrocarbons with different structures', *Scientific Reports*, 6. doi:10.1038/srep34588.

Wickham, H. (2016) 'ggplot2: Elegant Graphics for Data Analysis.', *Springer-Verlag* [Preprint].

Wu, F. *et al.* (2019) 'Pyrene Degradation by Mycobacterium gilvum: Metabolites and Proteins Involved', *Water, Air, and Soil Pollution*, 230(3). doi:10.1007/s11270-019-4115-z.

Wu, Y.W., Simmons, B.A. and Singer, S.W. (2016) 'MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets', *Bioinformatics*, 32(4), pp. 605–607. doi:10.1093/BIOINFORMATICS/BTV638.

Yang, J. *et al.* (2021) 'Colonization and performance of a pyrene-degrading bacterium Mycolicibacterium sp. Pyr9 on root surfaces of white clover', *Chemosphere*, 263, p. 127918. doi:10.1016/j.chemosphere.2020.127918.

Zhao, J.-K. *et al.* (2013) 'Parapedobacter pyrenivorans sp. nov., isolated from a pyrene-degrading microbial enrichment, and emended description of the genus Parapedobacter', *International Journal of Systematic and Evolutionary Microbiology*, 63(Pt_11), pp. 3994–3999. doi:10.1099/ijs.0.051938-0.