Article Type: Original Article

Metagenomics reveals the high PAH-degradation potential of abundant uncultured bacteria from chronically-polluted subantarctic and temperate coastal marine environments

Claudia L. Loviso[#], Mariana Lozada[#], Lilian M. Guibert, Matías A. Musumeci, Sandra Sarango Cardenas, Ruud V. Kuin, Magalí S. Marcos and Hebe M. Dionisi*

[#]These authors contributed equally to this work.

Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CENPAT-CONICET), Blvd. Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina.

Abbreviated running headline: uncultured PAH-degrading bacteria

E-mail addresses: C.L.L.: loviso@cenpat-conicet.gob.ar, M.L.: lozada@cenpat-conicet.gob.ar, L.M.G.: guibert@cenpat-conicet.gob.ar, M.A.M.: musumeci@cenpat-conicet.gob.ar, S.S.C.: sandra.sarango.c@gmail.com, R.V.K.: rvjkuin@gmail.com, M.S.M. magali@cenpat-conicet.gob.ar, H.M.D. hdionisi@cenpat-conicet.gob.ar.

***Correspondenc**e: Hebe M. Dionisi, Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CENPAT-CONICET), Blvd. Brown 2915,

U9120ACD, Puerto Madryn, Chubut, Argentina. Tel. +54-280-4883184 extension 1251; Fax +54-280-4883543. e-mail: hdionisi@cenpat-conicet.gob.ar.

Current addresses: Sandra Sarango Cardenas, Teconec SAC, Av. La Molina 3365, La Molina, Perú; Magalí S. Marcos, Laboratorio de Microbiología y Biotecnología, Instituto Patagónico para el Estudio de los Ecosistemas Continentales (IPEEC, CENPAT-CONICET).

Abstract

Aims: To investigate the potential to degrade polycyclic aromatic hydrocarbons (PAHs) of yet-to-be cultured bacterial populations from chronically-polluted intertidal sediments.

Methods and Results: A gene variant encoding the alpha subunit of the catalytic component of an aromatic ring-hydroxylating oxygenase (RHO) was abundant in intertidal sediments from chronically-polluted subantarctic and temperate coastal environments, and its abundance increased after PAH amendment. Conversely, this marker gene was not detected in sediments from a non-impacted site, even after a short-term PAH exposure. A metagenomic fragment carrying this gene variant was identified in a fosmid library of subantarctic sediments. This fragment contained five pairs of alpha and beta subunit genes and a lone alpha subunit gene of oxygenases, classified as belonging to three different RHO functional classes. *In silico* structural analysis suggested that two of these oxygenases contain large substrate-binding pockets, capable of accepting high molecular weight PAHs.

Conclusions: The identified uncultured microorganism presents the potential to degrade aromatic hydrocarbons with various chemical structures, and could represent an important member of the PAH-degrading community in these polluted coastal environments.

Significance and Impact of Study: This work provides valuable information for the design of environmental molecular diagnostic tools and for the biotechnological application of RHO enzymes.

Keywords: intertidal sediments; polycyclic aromatic hydrocarbons; ring-hydroxylating oxygenases; qPCR; metagenomic library; protein modelling.

INTRODUCTION

Urbanized coastal marine environments receive a constant input of polycyclic aromatic hydrocarbons (PAHs). Due to their hydrophobic nature and slow biodegradability, these compounds accumulate in sediments and in some marine organisms, where they can elicit adverse effects on both human and environmental health (Nikolaou *et al.* 2009). The management of contaminated sediments greatly depends on natural or enhanced biodegradation processes (Himmelheber and Hughes 2014). The molecular analysis of key genes related to pollutant biodegradation can provide valuable information concerning the presence or activity of the microbial populations responsible for the removal of environmental contaminants. Furthermore, this information sets the basis for the design of molecular biological tools for environmental bioremediation applications (Mahendra *et al.* 2012). The genes encoding the large (α) subunit of PAH dioxygenases have been widely utilized as biomarkers to analyze the potential of a microbial community to degrade PAHs under aerobic conditions (Iwai *et al.* 2011). These enzymes catalyze the first and limiting step in the activation of these compounds, the dihydroxylation of one of the aromatic rings of the molecule to generate a cis-dihydrodiol (Parales and Resnick 2006). PAH dioxygenases are members of a large enzyme family named ring-hydroxylating oxygenases (RHOs). These multicomponent enzyme systems are composed of a terminal oxygenase presenting an homo-

(α_n) or hetero-multimeric ($\alpha_n\beta_n$) structure, and one or two soluble proteins that transfer electrons from NAD(P)H to the oxygenase (Chakraborty *et al.* 2012).

Many vulnerable coastal environments of Patagonia (Argentina), such as breeding and feeding grounds of seabirds and marine mammals, are close to oil producing areas or urban settlements, raising concerns about the possible effects of hydrocarbon pollution (Barragán Muñoz *et al.* 2003; Yorio 2009). Moderate to high levels of PAHs have been detected in both sediments and benthic fauna at a number of coastal areas (Commendatore and Esteves 2007; Lozada *et al.* 2008; Amin *et al.* 2011; Commendatore *et al.* 2012; Marcos *et al.* 2012). The diversity of the PAH-degrading community from these chronically-polluted environments is just starting to emerge. The majority of the sequences identified in studies targeting RHO α-subunit genes shared low to moderate identity values with previously reported oxygenases (Lozada *et al.* 2008; Marcos *et al.* 2009; Dionisi *et al.* 2011). The bacterial populations carrying some of these gene variants were found to be abundant and stable, although mostly restricted to cold coastal environments (Marcos *et al.* 2012). In contrast, the *phnA1* gene, previously identified in PAH-degrading strains belonging to the genus *Cycloclasticus* (Kasai *et al*. 2003), was found to be abundant in only half of the analyzed sediment samples from both temperate and subantarctic environments of Patagonia (Marcos *et al*. 2012). However, a strong positive correlation between *phnA1* gene abundances and low molecular weight PAH concentrations suggested that members of this cosmopolitan genus could play a role in the degradation of these compounds in coastal environments of Patagonia. More information is still needed to be able to select marker genes suitable for molecular environmental diagnostics of hydrocarbon pollution in the more than 2,000 km of the Patagonian coast.

In this work, we used multiple molecular approaches to identify environmentally relevant members of the PAH-degrading community from coastal environments of Patagonia exposed to different climates. First, we identified a RHO α-subunit gene variant that was

abundant in PCR clone libraries from both temperate and subantarctic sediments. Secondly, we designed a quantitative PCR (qPCR) assay to estimate the abundance of microorganisms carrying this gene in intertidal sediments from both types of environments, as well as to evaluate their response after PAHs and crude oil exposure. Lastly, in order to recover the full-length sequence of this gene variant, as well as to obtain information about the capabilities of these yet-uncultured microorganisms, we constructed a large-insert metagenomic library from polluted intertidal sediments. We identified a metagenomic fragment carrying this marker gene, which encoded six oxygenases with various potential substrates, including high molecular weight PAHs.

MATERIALS AND METHODS

Sample collection. Figure 1 shows the three selected sampling locations along the Patagonian coast. An intertidal sediment sample was obtained in Fracasso Beach (PF, 42°25' S, 64°7' W), a non-impacted environment located in the San José Gulf, Valdés Peninsula (sample PF08). Chronically polluted sediment samples were retrieved from: (a) Cordova Cove (CC, 45°45' S, 67°22' W), a temperate environment situated in the San Jorge Gulf (samples CC08-1, CC08-2, CC10-1 and CC10-2), and (b) Ushuaia Bay (UB, 54°48' S, 68°17' W), a subantarctic environment located within the Beagle Channel (samples EM06, OR05, OR06, OR07 and OR08). For each of these samples, intertidal sediments (top 3 cm) were retrieved along the low-tide line at ten random points using acrylic cores with an inner diameter of 4.4 cm and stored at 4°C during transport to the laboratory. The ten sub-samples were mixed thoroughly to produce a composite sample, which was used immediately to construct the experimental systems, or stored at –80°C for molecular analyses and -20°C for chemical analysis. Samples were named according to their sampling location (PF, CC), or site within UB (EM, OR) followed by the last two digits of the sampling year, with an additional number for the two CC samples obtained in the same date. Table S1 shows a summary of the

samples used in this study. Detailed information of the sampling locations, dates and PAH concentrations can be found in previous studies (Lozada *et al.* 2008; Marcos *et al.* 2012).

Experimental systems. Sediment-in-seawater slurries were built in 500-ml flasks and consisted of approximately 40 g of sediment [samples PF08, CC08-2 or OR08, (Marcos *et al.* 2012)] and 80 ml of 0.45-µm filtered natural seawater, as previously described (Guibert *et al.* 2012). Slurries were left without further hydrocarbon addition, or were amended with the following hydrocarbons: 0.34 g of phenanthrene, 0.34 g of pyrene or 0.46 ml of light crude oil (from the San Jorge Basin, Comodoro Rivadavia, Argentina). Slurries were incubated in the dark at 15°C for 20 days, with constant agitation at 150 rev min^{-1}. Samples were obtained from the slurries without decanting, centrifuged 5 min at 6000 *g*, the supernatant was discarded and the sediment was stored at -80ºC for qPCR analysis.

Metagenomic DNA extraction. The DNA used for the construction of PCR clone libraries and for qPCR analysis was purified in duplicate from 0.5 to 0.8 g wet sediment using the FastDNA®SPIN kit for soil (MP Biomedicals, Santa Ana, CA, USA), as previously described (Lozada *et al.* 2008). Two independent DNA extractions per sample were pooled in equal mass amounts and used as template in PCR-based assays. The metagenomic DNA used for the construction of the fosmid library was purified from sediment sample OR07 following the protocol described by Zhou *et al.* (1996). DNA concentrations were determined using the DNA binding-fluorophore EvaGreen® (Biotium, Inc., Hayward, CA, USA) in a Chromo4 thermal cycler (Bio-Rad, Hercules, CA, USA) (Wang *et al.* 2006) or Hoechst 33258 dye (Amersham Biosciences, Piscataway, NJ) in a Hoefer DyNA Quant 200 fluorometer (Hoefer Scientific Instruments, San Francisco, CA).

Construction and screening of PCR clone libraries. The PCR clone libraries were constructed using primers Nah-for and Ac596r [(Wilson *et al.* 1999; Zhou *et al.* 2006), Table S2]. PCR amplifications were performed in a Chromo4 thermal cycler in 25-µl reactions

containing 50 mmol l^{-1} KCl, 10 mmol l^{-1} Tris-HCl pH 9.0, 0.1% Triton X-100, 1.5 mmol l^{-1} $MgCl_2$, 0.2 μmol l^{-1} dNTPs, 2 μmol l^{-1} primer Nah-for, 0.5 μmol l^{-1} primer Ac596R, 1 U of T-PLUS DNA polymerase (Inbio-Highway, Tandil, Argentina) and 1 to 10 ng of metagenomic DNA. The PCR program is indicated in Table S2. PCR products from UB sediment samples (OR06, OR07, OR08 and EM06) were combined, purified with PCR WIZARD® SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA), and cloned into the pCR®4.0 vector using the TA Cloning kit for sequencing (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. PCR products from CC sediment samples (CC08-1, CC08-2, CC10-1 and CC10-2) were processed following the same protocol. For the analysis of the experimental system of UB sediments exposed to crude oil [OR08-oil, (Guibert *et al.* 2012)], six independent PCR reactions were combined and run in a 1.5% agarose gel. The band with the expected amplicon size was excised and purified using the PCR WIZARD® SV Gel and PCR Clean-Up System before cloning. Plasmids were purified from randomly chosen clones using QIAprep® Spin Miniprep Kit (QIAGEN Inc., Valencia, CA, USA), and 20 inserts (8 clones from both UB and OR08-oil libraries, and 4 clones from CC library) were sequenced using the vector primer M13F at the sequencing service of INTA-CASTELAR (Hurlingham, Buenos Aires, Argentina).

qPCR assays. The quantification of the target genes (RHO α subunit gene variants T and *phnA1*, as well as bacterial 16S rRNA gene) was carried out in DNA purified from intertidal sediments and experimental systems using SybrGreen I-based qPCR assays. Primer3 software (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used for the design of the primer set targeting gene variant T, based on the consensus sequence of the gene fragments identified in the PCR clone libraries. The specificity of the potential primers was evaluated by comparing their sequences with the NCBI database using blastn tool. Assay optimization and sample analysis were performed as previously described (Marcos *et al.*

2012). The qPCR reactions were carried out in a Chromo4 thermal cycler in a 20-µl volume containing 1× PerfeCTa® SYBR® Green Supermix (Quanta BioSciences, Inc., Gaithersburg, MD) or Mezcla Real (Biodynamics, Buenos Aires, Argentina), the primers at the concentrations indicated in Table S2, and 0.2 to 2.5 ng of template DNA. The amplification program for each qPCR assay is indicated in Table S2. Non-template controls were included in all runs. Plasmids carrying a larger fragment of the target genes were used as standard for each assay. Each plasmid was purified, linearized and quantified as previously described (Marcos *et al.* 2012). A standard curve with target concentrations between 5 and 10^8 copies reaction^{-1} was calculated for each run. Reactions were performed in triplicate or quadruplicate, and an additional spiking reaction was added (10^6 copies of standard DNA reaction^{-1}) in each sample and target, in order to correct for PCR inhibition (Marcos *et al.* 2012). The specificity of the reactions was confirmed by comparing melting curves from the samples and the standards and by agarose gel electrophoresis. The qPCR data are reported as copies of the target gene µg DNA^{-1}. Amplification efficiencies were calculated as Eff = $(10^{(-1/\text{slope})})^{-1}$ (Pfaffl 2004).

Construction and screening of the sediment metagenomic library. A fosmid library was constructed using the metagenomic DNA isolated from sample OR07 (Marcos *et al.* 2012) using the CopyControl™ HTP Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI, USA) following the protocol suggested by the manufacturer. The library was plated on LB agar with 12.5 µg ml^{-1} chloramphenicol (approximately 100 clones per plate). After an overnight incubation at 37ºC, the bacterial cells from each plate were resuspended with 1 ml of LB medium containing 20% glycerol and stored at -80°C. Primers Nah-for and Ac596r were chosen for the molecular screening of the metagenomic library, using the same conditions as described above with the exception that whole-cell PCR amplifications were used to simplify the screening. Cells from each group of clones were washed and

resuspended in sterile distilled water and used as template in PCR reactions. The PCR products were visualized in agarose gels. PCR products of the expected size were sequenced to confirm their specificity before isolation and screening of individual clones.

Fosmid DNA extraction and sequencing. Fosmid DNA was purified from 10 ml of cells previously induced for 15 h [LB medium with 12.5 μg ml^{-1} chloramphenicol and 1X CopyControl Fosmid Autoinduction Solution (Epicentre Biotechnologies, Madison, WI, USA)], using Qiagen Miniprep Kit (Qiagen, Valencia, CA, USA). The purified fosmid DNA was sequenced at INDEAR (Rosario, Argentina) using 454/Roche sequencing platform.

Functional annotation and taxonomic assignment of the metagenomic fragment. After read assembly (peak depth of 39), gene prediction and annotation were performed using the Integrative Services for Genomic Analysis (ISGA, http://isga.cgb.indiana.edu) and the Rapid Annotation using Subsystem Technology (RAST, http://rast.nmpdr.org), and further manually curated. Conserved amino acid sequences and potential biochemical functions of predicted genes were confirmed using InterProScan program (Goujon *et al.* 2010), Kyoto Encyclopedia of Genes and Genomes database (KEGG) (Kanehisa and Goto 2000), pfam (Finn *et al.* 2014), COG (Tatusov *et al.* 2001) and blast analyses (McGinnis and Madden 2004). G+C % content was calculated using GC Calculator (www.genomicsplace.com/gc_calc.html). The taxonomic assignment of the metagenomic fragment was performed using PhyloPythiaS (Patil *et al.* 2012) and Megan5 (Huson *et al.* 2007). In the case of PhyloPythyiaS, the sequence was split into ~3-kb fragments. The model type used was generic 2013, which includes 4,522 genomes from 800 different genera. The parameters used in Megan 5 were: blastp of all coding sequences with min-score = 100, min-support = 3, top-percent = 10, win-score = 0, min-complexity = 0.44.

Phylogenetic and *in silico* structural analyses of RHO sequences. Multiple sequence alignments of RHO α-subunit sequences were performed using ClustalX with default settings

(Larkin *et al.* 2007), accessed as a web service from Jalview multiple sequence alignment editing, visualisation and analysis program (Waterhouse *et al.* 2009). Phylogenetic trees were constructed using the Molecular Evolutionary Genetics Analysis software [MEGA 6.06, (Tamura *et al.* 2013)], with the Neighbour-Joining method and Jones Taylor Thornton matrix as amino acid replacement model. Robustness of the inferred tree topology was verified by bootstrapping with 1,000 replications. The RHObase program was also used for RHO classification into functional classes, as well as for the prediction of their potential substrates (Chakraborty *et al.* 2014). The hypothetical structure models of the three class A RHO sequences identified in the metagenomic fragment were generated using the Swiss-Model server (Biasini *et al.* 2014), using the crystal structures of oxygenases from *Sphingomonas sp* CHY-1 (PhA1, PDB 2CKF; Jakoncic *et al.* 2007) and *Sphingomonas yanoikuyae* B1 (BphA1, PDB 2GBW; Ferraro *et al.* 2007) as templates, as indicated. The models were checked using the servers QMEAN (Benkert *et al.* 2008), Verify3D (Liithy *et al.* 1992) and VADAR (Willard *et al.* 2003). The CASTp server (Dundas *et al.* 2006) was used for identifying the catalytic cavities and calculating their respective volume, while the height, width and length were measured using UCSF Chimera software (Pettersen *et al.* 2004). For docking analysis, enzyme-substrate complexes were generated based on the information on the substrate binding mode in the active site of previously determined RHO crystal structures. Phenanthrene and pyrene were fit in the catalytic cavities in accordance with the geometry of phenanthrene displayed in the reported crystal structure (PDB 2HMK; Ferraro *et al.* 2006), following a reported procedure (Kweon *et al.* 2010). Briefly, the modelled three-dimensional structures of the M117 dioxygenases and their templates were superimposed to the crystal structure of naphthalene 1,2-dioxygenase bound to phenanthrene (PDB 2HMK) by using UCSF Chimera. After superimposition, phenanthrene and pyrene molecules were overlapped to the bound phenanthrene, preserving the distances of the to-be-hydroxylated target atoms of

the ligands to the catalytic iron. Docking analyses were also performed using SwissDock server (Grosdidier *et al.* 2011a; Grosdidier *et al.* 2011b). The Pose and Rank server was used to score the protein-ligand complexes (Fan *et al.* 2011).

Sequence accession numbers. Sequences obtained in this study have been deposited in the NCBI database under accession numbers KM102501 to KM102507 (PCR clone libraries) and KP330468 (M117 fosmid clone).

RESULTS

Identification of RHO α-subunit gene fragments

In order to identify functional marker genes related to PAH biodegradation in coastal environments of Patagonia, we selected a high-coverage primer set targeting RHO α-subunit genes based on the analysis of primer specificity and coverage recently reported by Iwai and collaborators (2011). To obtain a better representation of the diversity of PAH dioxygenase genes in each selected chronically-polluted environment, we combined the PCR products obtained from 4 different sediment samples from each sampling location before cloning (Table S1). Total PAH concentrations in these samples varied between 378 and 4,127 μg kg^{-1} dry weight sediment (Table S1; Lozada *et al.* 2008; Marcos et al. 2012). Primers Nah-for and Ac596r (Table S2) yielded strong and reproducible PCR amplifications from samples of both temperate and subantarctic environments of Patagonia (CC and UB, respectively; Figure 1), as well as from an experimental system constructed with a UB sediment sample and amended with crude oil (OR08-oil; Guibert *et al.* 2012). The cloning and sequencing of these amplicons revealed four distinct variants of RHO α-subunit genes, sharing only 19.5 to 51.2 % identity at the amino acid level. One of the gene variants was closely related to gene fragments previously amplified from UB sediments (gene variant E, Table 1). The other three variants presented only moderate identities at the amino acid level with PAH dioxygenases

identified in cultured bacteria, and were named T, U and V following the nomenclature used in previous works (Lozada *et al.* 2008; Marcos *et al.* 2009). The most abundant clone type corresponded to gene variant T, which was detected in sediments from both subantarctic and temperate coastal environments and overall represented 60 % of the analyzed clones (Table 1).

Abundance of gene variant T in sediment samples and experimental systems

To be able to estimate the abundance of gene variant T in intertidal sediment samples, as well as to evaluate changes in the abundance of microorganisms carrying this gene after exposure to phenanthrene, pyrene or crude oil, we designed and optimized a qPCR assay targeting this gene (Table S2). We compared the abundances of this gene with those of *phnA1*, identified in PAH-degrading strains belonging to the genus *Cycloclasticus* (Kasai *et al.* 2003). We analyzed the relative abundance of the bacterial 16S rRNA gene as reference. Gene variant T was not detected in sample PF08, retrieved from a temperate coastal environment not exposed to anthropogenic pollution, nor was observed in slurries constructed using this sediment sample after hydrocarbon amendment (Figure 2A). In contrast, gene variant T was abundant in polluted sediment samples CC08-2 and OR08, and further increase its abundance in most experimental systems constructed with these sediments and exposed to crude oil, phenanthrene or pyrene (Figure 2B and C). These results indicate that microorganisms carrying this gene were not only abundant in sediments from both temperate and subantarctic environments, but these populations were able to grow after PAH exposure in the experimental systems. In comparison, the *phnA1* gene from *Cycloclasticus* spp. was undetectable or at abundances below the quantification limit of the assay in the three analyzed sediment samples, but showed a dramatic increase in abundance when chronically-polluted sediments were amended with crude oil or PAHs. In non-impacted sediments, *phnA1* gene abundances increased after pyrene and crude oil exposure, but not after phenanthrene

amendment (Figure 2A). This result suggests the presence of pyrene degrading *Cycloclasticus* strains in pristine environments of Patagonia, which would be able to respond rapidly in the event of an oil spill.

We analyzed the prevalence and dynamics of microbial populations carrying gene variant T in sediments of UB, by assessing the abundance of this gene in a series of intertidal sediment samples obtained during three consecutive years at the same site. Gene variant T was found at high relative abundances in all the analyzed samples (Figure 2D). These results indicate that, at least in the sampled period, the microorganisms carrying this gene variant represented stable population/s within the microbial community indigenous of this site.

Genomic context of gene variant T

In order to recover a full-length sequence of gene variant T as well as to obtain information about its genomic context, we constructed a large-insert metagenomic library from intertidal sediment sample OR07. This sample was chosen based on the overall abundance of novel RHO α subunit gene variants (Marcos *et al.* 2012 and this study). The library consisted of approximately 46,000 clones covering 1.6 Gb of metagenomic information. The same primer set and conditions used for the construction of the PCR clone libraries were applied for the sequence-based screening of the metagenomic library. One fosmid clone (M117) carried a gene that shared 97.7 to 98.8 % identity at the amino acid level with the gene variant T identified in the PCR clone libraries. None of the other gene variants identified in the PCR clone libraries were detected in the metagenomic library. Clone M117 had an insert size of 36.7 kb, a G+C content of 54.1 %, and contained 38 coding sequences (Figure 3 and Table S3). The potential source organism of this metagenomic fragment was evaluated using both PhyloPythiaS, a composition-based taxonomic classifier (Patil *et al.* 2012), and Megan, a sequence similarity-based method (Huson *et al.* 2007). PhyloPythiaS binned this metagenomic fragment within the Gammaproteobacteria class.

Megan assigned 21 % of the coding sequences to this class and 45 % of the sequences to Proteobacteria, while the rest of the coding sequences binned to Bacteria or received no assignment. Discontiguous megablast algorithm within blastn (McGinnis and Madden 2004) indicated that the identified metagenomic fragment was most closely related to the betaproteobacterium *Rhodocyclaceae bacterium* PG1-Ca6 (NCBI Accession number CP010554), with a 68 % identity and 32 % coverage. However, the alignment was highly fragmented and over 1.5 Mb of the chromosome of this microorganism. The second most related genome was the gammaproteobacterium *Cyclocasticus* sp. P1 (CP003230, 67 % identity and 27 % coverage). These results suggest that this fragment belongs to a member of the phylum Proteobacteria, although divergent from currently described PAH-degrading microorganisms.

Fosmid M117 contained a lone gene encoding the α-subunit of an oxygenase (M117-16) as well as five pairs of adjacent and codirectional α and β subunits genes (M117-22/21, 23/24, 33/32, 36/35 and 38/37; Figure 3 and Table S3). M117-33 was the gene detected in the molecular screening of the metagenomic library, corresponding to gene variant T. The majority of the oxygenase sequences identified in this fragment shared moderate identity values with their closest matches of the NCBI database in Blastp analysis, which were dioxygenase sequences identified in isolates belonging to the classes Alpha-, Beta- and Gammaproteobacteria, as well as Bacilli (Table S4). Two methods were used to classify these terminal oxygenases into functional classes, a phylogenetic analysis of the α subunit sequences (Figure S1) and the prediction tool of the Ring-Hydroxylating Oxygenase database (RHObase; Chakraborty *et al.* 2014). The results of the two analyses were in agreement (Table S4). This server was also used to evaluate the potential substrates of these oxygenases (Table S4). The lone α subunit gene (M117-16) was classified as class C RHO. Despite the absence of an adjacent gene encoding the β subunit in the identified fragment, this sequence

clustered with oxygenases with a hetero-multimeric structure, classified as type IV based on their electron transport protein components (Figure S1). The putative substrates of this enzyme were various carboxylated aromatics, such as salicylate, substituted salicylates, dihydroxybenzoates and anthranilate. Sequences M117-21 to 24 corresponded to two contiguous sets of α and β subunits with opposite orientations (Figure 3), classified within the functional class B. The α-subunit sequences did not cluster with reference sequences classified as type I, II and IV (Figure S1), and their putative substrates corresponded to carboxylated aromatics, like benzoate and toluate.

The last three pairs of oxygenase sequences identified in the metagenomic fragment coded for the terminal component of class A RHOs (Figure 3 and Table S4). All three α-subunit sequences clustered with type IIIαβ RHO enzymes (Figure S1), and their potential substrates included various aromatic hydrocarbons, such as PAHs, arylbenzenes, and/or alkylbenzenes (Table S4). Three-dimensional models of these oxygenases were constructed to evaluate the potential shape and size of their active sites. These models were based on the crystal structure of PhnA1 from *Sphingomonas sp* CHY-1 (PDB 2CKF) for sequences M117-33/32 or BphA1 from *S. yanoikuyae* B1 (PDB 2GBW) for sequences M117-36/35 and M117-38/37. M117-38 was truncated, but contained a complete catalytic domain. A hybrid sequence was then generated using 166 amino acids of the N-terminus end of its closest relative, sequence NP_049184 from *Novosphingobium aromaticivorans* F199. The sequence identities of the modelled RHO α-subunit sequences with their template ranged from 42 to 51%. Differences in both composition and conformation of the active sites were observed in the three modelled enzymes, when compared with their template (Figure S2). In agreement, the catalytic cavities were also quite different (Figure S3). The model for sequences M117-38/37 showed the largest catalytic cavity dimensions, which was 50 % larger than the one from its template structure and 30% larger than the active site dimension calculated for

sequence pair M117-33/32 (Table S5). Docking analyses were performed to assess if phenanthrene and pyrene (PAHs evaluated in the experimental systems) could potentially be substrates of these oxygenases (Figure S3). These studies were based on the binding mode of phenanthrene in the crystal structure of naphthalene 1,2-dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4 (Ferraro *et al.* 2006). The score values for phenanthrene binding in the three sequences were similar to the one calculated for PhnA1 from *Sphingomonas* CHY-1 (Table S6), which is able to bind and oxidize this ligand (Demaneche *et al.* 2004). In the case of pyrene, however, better score values were obtained for M117-38/37 and M117-36/35 than for M117-33/32. These results suggest that while phenanthrene could fit in the active sites of the three modelled oxygenases, pyrene could probably only be accommodated in the catalytic pocket of enzymes encoded by sequences M117-36/35 and M117-38/37.

The other coding sequences identified in the metagenomic fragment were mostly related to general cellular processes, such as the biosynthesis of tryptophan, lysine, purine and folate, cell division, lipid metabolism, stress response, signal transduction and trans editing activity of D-tyrosyl-tRNA, among others (Figure 3 and Table S3). No genes encoding electron transport protein components were found in the metagenomic fragment, nor genes related to horizontal gene transfer events.

DISCUSSION

Intertidal environments are highly dynamic, and sediment microbial communities from this zone are heavily exposed to abiotic factors (Van Colen *et al*. 2014). The extensive coast of eastern Patagonia features important differences in near-surface air temperatures and precipitations (Garreaud *et al.* 2013). Arid conditions, strong winds and moderate temperatures are the predominant climatic conditions in northern and central Patagonia, while a humid and colder climate is prevalent in subantarctic environments such as the Beagle Channel (Marcos *et al.* 2012). These differences in the environmental conditions might

represent an important factor influencing the structure of the PAH-degrading microbial community. Previous works were successful in identifying various novel RHO α-subunit gene fragments in Ushuaia Bay (UB), a subantarctic coastal environment (Lozada *et al.* 2008; Marcos *et al.* 2009; Dionisi *et al.* 2011), but a similar strategy applied to sediments from polluted coastal environments of Northern or Central Patagonia resulted in either unspecific amplification or the retrieval of previously characterized sequences that were found often present al low abundances (Lozada *et al.* 2008; Marcos *et al.* 2012). A more complete knowledge of the bacterial populations with PAH-degradation potential in different coastal environments of Patagonia is still needed. This information is not only necessary for a better understanding of PAH-degradation processes, but also for the selection of suitable marker genes for the design of environmental diagnostic tools for this region (Mahendra et al. 2012).

In this work, we identified a gene variant encoding a RHO α subunit that was present in intertidal sediment samples from two chronically-polluted sites distanced approximately 1,000 km and belonging to different biogeographic regions (Marcos *et al.* 2012). The closest relative of this gene was an aromatic-ring-hydroxylating dioxygenase from the pyrene degrading strain *Cycloclasticus* sp. P1 (Lai *et al.* 2012), although these sequences only shared a moderate identity. Other close relatives of gene variant T were dioxygenase sequences identified in *Novosphingobium* and *Sphingomonas* strains (data not shown). We designed a qPCR assay targeting this gene with the sequence information obtained in the clone libraries, as PCR clone libraries are not suitable to accurately assess the abundance of target genes in environmental samples (Sipos *et al.* 2010). We used this assay to estimate the abundance of uncultured microorganisms containing this gene in the sediments and to assess the short-term response of these populations after PAH or crude oil amendment. This analysis showed that these microorganisms were stable members of the PAH-degrading community in chronically polluted environments, and in some experimental conditions further increased their

abundance. In contrast, we were not able to detect this gene in sediments with low anthropogenic impact, even after exposure to high concentration of PAHs in experimental systems. These results suggest that this gene could constitute a valuable biomarker of chronic pollution in coastal environments of Patagonia, although further studies are needed to evaluate the role of these populations in PAH biodegradation processes.

Obligate hydrocarbonoclastic bacteria belonging to the genus *Cycloclasticus* have long been recognized as predominant PAH degrading microorganisms in marine sediments (Staley 2010). In the polluted sediments analyzed in this study, the functional marker gene for these microorganisms (*phnA1*), was present at abundances below the quantification or detection limits of the technique, in agreement with a previous study (Marcos *et al.* 2012). However, bacteria carrying the *phnA1* gene were able to grow rapidly at the expense of the added PAHs in the experimental systems built with polluted sediments. *Cycloclasticus* spp. have been shown to thrive in the presence of PAHs, in particular in seawater, reaching high abundances (Kasai *et al.* 2002; Teira *et al.* 2007; Yang *et al.* in press). This shift has been associated with the biodegradation of low molecular weight compounds, and it is typically transient (Teira *et al.* 2007). It is possible that, while the microorganisms carrying gene variant T are already established members of the benthic community, the conditions used in the experimental systems (seawater saturation, and high oxygen levels due to agitation) were more adequate than those from the original sediments for the growth of bacteria belonging to the genus *Cycloclasticus*.

A rapid response to hydrocarbon exposure of the bacterial populations targeted by the qPCR assays in chronically-polluted sediments was consistent with the long-standing notion that pre-exposure results in a pre-adaptation to pollutants and therefore a faster response in a further pollution event (Païssé *et al.* 2010). Possibly, the time frame chosen for this experiment was too short to be able to observe a similar response in PF sediments, which did

not have a history of exposure to anthropogenic hydrocarbons and contained no detectable PAHs (Marcos *et al.* 2012). Interestingly, a moderate increase in *phnA1* gene abundance was observed after pyrene amendment in these sediments. Several *Cycloclasticus* strains have shown the ability to degrade pyrene (Wang *et al.* 2008; Lai *et al.* 2012; Cui *et al.* 2014). The increase in *phnA1* gene abundance observed in all the experimental systems exposed to pyrene suggests that *Cycloclasticus* strains with similar capabilities could be present in intertidal sediments of all three sites.

In this work, we used a metagenomic approach to increase our knowledge of the yet-to-be-cultured microorganisms carrying gene variant T. The fosmid metagenomic library was constructed using an intertidal sediment sample obtained near an oil jetty that is used to discharge refined petroleum products to shore storage tanks. The operation of this oil terminal represents a source of chronic hydrocarbon pollution to the coastal environment located near this jetty (Commendatore *et al.* 2012). Aliphatic hydrocarbon diagnostic indices in this sample provided evidence of ongoing biodegradation processes (Guibert *et al.* 2012). Moderate levels of PAHs were detected in this sample, which included naphthalene, acenaphthene, anthracene, fluoranthene as well as pyrene (Marcos *et al.* 2012). Assuming an average prokaryotic genome size of 3.8 Mb (Tomazetto *et al.* 2015), this metagenomic library contains a genomic information equivalent to approximately 470 prokaryotic genomes. Due to the high diversity of the sediment microbial community at this site, with an estimated richness of more than 3,000 OTUs (defined at 0.03 distance threshold) only for the bacterial fraction (Guibert *et al.* 2012), the metagenomic library presents a low coverage of the microbial community indigenous of this site. However, due to the high relative abundance of the microorganisms carrying this gene variant, the size of the metagenomic library was sufficient for the identification of one clone carrying the biomarker gene.

The identified metagenomic fragment contained sequences encoding six oxygenase component of RHOs, which were classified as belonging to three different functional classes (Chakraborty *et al.* 2012) and presented potential substrates with various chemical structures (Chakraborty *et al.* 2014). The recent sequencing of the genome of PAH-degrading bacterial strains showed that these microorganisms often carry multiple genes encoding the oxygenase component of RHOs, distributed throughout their genome (Lai *et al.* 2012; Zhang and Anderson 2012; Cui *et al.* 2013, Khara *et al*. 2014; Singleton *et al.* 2015). In fact, the organization of the genes involved in PAH biodegradation pathways originally described in *Pseudomonas* strains (two gene clusters encoding the enzymes involved in the upper and lower pathways), is now known to be extremely rare (Suenaga *et al.* 2009). For instance, *Cycloclasticus* strains, which are highly specialized in the utilization of aromatic compounds, were found to carry at least twelve RHO α-subunit genes (Lai *et al.* 2012; Cui *et al.* 2013). Metabolically versatile Sphingomonad strains, capable of degrading various aromatic compounds, contained seven pairs of genes coding for α and β subunits of RHOs organized in several clusters, and only one set of genes encoding the electron transport system (Khara *et al*. 2014). Similarly, a novel microorganism belonging to the *Rhodocyclaceae* family (strain PG1) had eight sets of genes coding for RHO enzymes (Singleton *et al.* 2015). The majority of α and β subunit sequences identified in the metagenomic fragment were related to oxygenases identified in these three groups of PAH-degrading bacteria, although the moderate identity values suggest that the identified microorganism is not affiliated with either of these groups. Interestingly, the genome of strain PG1 (isolated from contaminated soil) also contains a coding sequence with 85 % identity at the amino acid level with gene variant A, identified in UB sediments using a PCR based approach (Lozada *et al.* 2008, Marcos *et al.* 2012).

Three pairs of RHO α and β subunits, located in less than 7 kb within the metagenomic fragment, encoded class A oxygenases. Class A RHOs preferentially hydroxylate α,β-positions of the aromatic ring with respect to an adjacent fused aromatic ring or a phenyl/alkyl/chloro-substitution, and include PAH dioxygenases (Chakraborty *et al.* 2012). Three-dimensional models of these oxygenases predicted catalytic pockets with different volumes and shapes, and possibly different substrate preferences. When compared with the enzymes used as template, PhnA1 from *Sphingomonas* sp. CHY-1 and BphA1 from *S. yanoikuyae* B1, the identified oxygenases contained similar or larger catalytic pockets, and two of these oxygenases seemed to be able to accommodate pyrene. Naphthalene was the preferred substrate of PhnA1 (Demaneche *et al.* 2004), while BphA1 recognized preferentially biphenyl (Yu *et al.* 2007). The oxygenase encoded by sequences M117-38/37 presented a much larger catalytic pocket, which was comparable in size with NidAB from *Mycobacterium vanbaalenii* PYR-1, whose preferential substrate is pyrene (Kweon *et al.* 2010). Overall, our results suggest that the three class A oxygenases encoded in fragment M117 might be able to hydroxylate phenanthrene and two of them pyrene, which were the PAHs evaluated in the experimental systems. Further studies, including the heterologous expression of these enzymes and their enzymatic characterization, will be needed to validate this hypothesis. Although the genes encoding the electron transport components of these enzymes were not present within the metagenomic fragment, it is still possible to express these enzymes using compatible components from related bacteria. This characterization will not only allow to gain a better understanding of the biodegradation capabilities of this uncultured microorganism, but also represents an essential step to explore the potential biotechnological applications of the identified enzymes (Parales and Resnick 2007; Allen 2012).

In this work, we used three different molecular approaches to identify and characterise a novel proteobacterium that appears to be highly specialized in the degradation of PAHs. The analysis of marine metagenome datasets available in public databases could not identify sequences with high identities with variant T (data not shown). Since only a small fraction of the marine microbial diversity is currently covered by these metagenomic datasets, this result cannot exclude the possibility that this biomarker gene could still be present in marine environments. The evidence obtained so far indicates that this microorganism has a biogeographic distribution at least extending polluted coastal environments of Patagonia. This RHO α-subunit gene, distinctive of chronically-polluted sediments from both temperate and subantarctic environments, represents a first step for its use as a biomarker in the analysis of sediment samples during natural or enhanced attenuation of polluted coastal environments of Patagonia. PAHs are highly persistent in coastal sediments, and molecular biological tools targeting PAH-degrading populations could provide valuable information for the management of polluted coastal environments, potentially reducing both time and total costs for site restoration (Interstate Technology Regulatory Council 2011).

ACKNOWLEDGMENTS

ML and HMD are staff members from The Argentinean National Research Council (CONICET). At the time of this study, CLL, LMG and MSM were recipients of graduate student fellowships and MAM was a recipient of a postdoctoral fellowship from CONICET. RVK (an undergraduate student from Leiden University of Applied Sciences, The Netherlands) and SVS (a master student at the University of Rouen, France) participated of this work while carrying out training internships at the Environmental Microbiology Laboratory (CENPAT-CONICET). This work was funded by grants from CONICET, the National Agency for the Promotion of Science and Technology from the Argentinean Ministry of Science and Technology (ANPCyT, MINCyT) and the Blacksmith Institute. We

thank M. Gil, J. L. Esteves, H. Ocariz, A. Torres, L. Bala and L. Musmeci for their invaluable contribution during sample collection.

Conflict of interest: No conflict of interest declared.

REFERENCES

Allen, C.C.R. (2012) Ring-hydroxylating dioxygenases: biocatalysts with remarkable Substrate specificity and exquisite product stereochemistry. *J Biocatal Biotransformation* **1**, e104.

Amin, O.A., Comoglio, L.I. and Sericano, J.L. (2011) Polynuclear aromatic and chlorinated hydrocarbons in mussels from the coastal zone of Ushuaia, Tierra del Fuego, Argentina. *Environ Toxicol Chem* **30**, 521-529.

Barragán Muñoz, J.M., Dadón, J.R., Matteucci, S.D., Morello, J.H., Baxendale, C. and Rodríguez, A. (2003) Preliminary basis for an integrated management program for the coastal zone of Argentina. *Coast Manage* **31**, 55-77.

Benkert, P., Tosatto, S.C. and Schomburg, D. (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct, Funct, Bioinf* **71**, 261-277.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M. and Bordoli, L. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nuc Acids Res* **42(W1)**, W252-W258.

Commendatore, M. and Esteves, J. (2007) An assessment of oil pollution in the coastal zone of Patagonia, Argentina. *Environ Manage* **40**, 814-821.

Commendatore, M.G., Nievas, M.L., Amin, O. and Esteves, J.L. (2012) Sources and distribution of aliphatic and polyaromatic hydrocarbons in coastal sediments from the Ushuaia Bay (Tierra del Fuego, Patagonia, Argentina). *Mar Environ Res* **74**, 20-31.

Cui, Z., Xu, G., Gao, W., Li, Q., Yang, B., Yang, G. and Zheng, L. (2014) Isolation and characterization of *Cycloclasticus* strains from Yellow Sea sediments and biodegradation of pyrene and fluoranthene by their syntrophic association with *Marinobacter* strains. *Int Biodeter Biodeg* **91**, 45-51.

Cui, Z., Xu, G., Li, Q., Gao, W. and Zheng, L. (2013) Genome sequence of the pyrene-and fluoranthene-degrading bacterium *Cycloclasticus* sp. strain PY97M. *Genome Announc* **1**, e00536-00513.

Chakraborty, J., Ghosal, D., Dutta, A. and Dutta, T.K. (2012) An insight into the origin and functional evolution of bacterial aromatic ring-hydroxylating oxygenases. *J Biomol Struct Dyn* **30**, 419-436.

Chakraborty, J., Jana, T., Saha, S. and Dutta, T.K. (2014) RHObase: A database of bacterial aromatic ring-hydroxylating oxygenases in the management of bioremediation and biocatalysis of aromatic compounds. *Environ Microbiol Rep* **6**, 519–523.

Demaneche, S., Meyer, C., Micoud, J., Louwagie, M., Willison, J.C. and Jouanneau, Y. (2004) Identification and functional analysis of two aromatic-ring-hydroxylating dioxygenases from a *Sphingomonas* strain that degrades various polycyclic aromatic hydrocarbons. *Appl Environ Microbiol* **70**, 6714-6725.

Dionisi, H.M., Lozada, M., Marcos, M.S., Di Marzio, W.D. and Loviso, C.L. (2011) Aromatic Hydrocarbon Degradation Genes from Chronically Polluted Subantarctic Marine Sediments. In *Handbook of Molecular Microbial Ecology II: Metagenomics in*

Different Habitats ed. De Bruijn, F.J. pp: 461-473. Hoboken, New Jersey, Wiley/Blackwell.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nuc Acids Res* **34**, W116-W118.

Fan, H., Schneidman-Duhovny, D., Irwin, J.J., Dong, G., Shoichet, B.K. and Sali, A. (2011) Statistical potential for modeling and ranking of protein-ligand interactions. *J Chem Inf Model* **51**, 3078-3092.

Ferraro, D., Brown, E., Yu, C.-L., Parales, R., Gibson, D. and Ramaswamy, S. (2007) Structural investigations of the ferredoxin and terminal oxygenase components of the biphenyl 2,3-dioxygenase from *Sphingobium yanoikuyae* B1. *BMC Struct Biol* **7**, 10.

Ferraro, D.J., Okerlund, A.L., Mowers, J.C. and Ramaswamy, S. (2006) Structural basis for regioselectivity and stereoselectivity of product formation by naphthalene 1, 2-dioxygenase. *J Bacteriol* **188**, 6986-6994.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L. and Mistry, J. (2014) Pfam: the protein families database. *Nuc Acids Res* **42 (D1),** D222-D230.

Garreaud, R., Lopez, P., Minvielle, M. and Rojas, M. (2013) Large-scale control on the Patagonian climate. *J Climate* **26**, 215-230.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nuc Acids Res* **38**, W695-W699.

Grosdidier, A., Zoete, V. and Michielin, O. (2011a) Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem* **32**, 2149-2159.

Grosdidier, A., Zoete, V. and Michielin, O. (2011b) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nuc Acids Res* **39**, W270-W277.

Guibert, L.M., Loviso, C.L., Marcos, M.S., Commendatore, M.G., Dionisi, H.M. and Lozada, M. (2012) Alkane biodegradation genes from chronically polluted Subantarctic coastal sediments and their shifts in response to oil exposure. *Microb Ecol* **64**, 605-616.

Himmelheber, D.W. and Hughes, J.B. (2014) In Situ Biotransformation of Contaminants in Sediments. In *Processes, Assessment and Remediation of Contaminated Sediments* ed. Reible, D.D. pp.263-304. New York: Springer.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17**, 377-386.

Interstate Technology Regulatory Council, E.M.D. (2011) *Environmental Molecular Diagnostics Fact Sheets* No. EMD-1 (http://www.itrcweb.org/GuidanceDocuments/EMD1.pdf).

Iwai, S., Johnson, T.A., Chai, B., Hashsham, S.A. and Tiedje, J.M. (2011) Comparison of the specificities and efficacies of primers for aromatic dioxygenase gene analysis of environmental samples. *Appl Environ Microbiol* **77**, 3551-3557.

Jakoncic, J., Jouanneau, Y., Meyer, C. and Stojanoff, V. (2007) The crystal structure of the ring-hydroxylating dioxygenase from *Sphingomonas* CHY-1. *FEBS J* **274**, 2470-2481.

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nuc Acids Res* **28**, 27-30.

Kasai, Y., Kishira, H. and Harayama, S. (2002) Bacteria belonging to the genus *Cycloclasticus* play a primary role in the degradation of aromatic hydrocarbons released in a marine environment. *Appl Environ Microbiol* **68**, 5625-5633.

Kasai, Y., Shindo, K., Harayama, S. and Misawa, N. (2003) Molecular characterization and substrate preference of a polycyclic aromatic hydrocarbon dioxygenase from *Cycloclasticus* sp. strain A5. *Appl Environ Microbiol* **69**, 6688-6697.

Khara, P., Roy, M., Chakraborty, J., Ghosal, D., and Dutta, T.K. (2014) Functional characterization of diverse ring-hydroxylating oxygenases and induction of complex aromatic catabolic gene clusters in *Sphingobium* sp. PNB. *FEBS Open Bio* **4**, 290-300.

Kweon, O., Kim, S.-J., Freeman, J.P., Song, J., Baek, S. and Cerniglia, C.E. (2010) Substrate specificity and structural characteristics of the novel rieske nonheme iron aromatic ring-hydroxylating oxygenases NidAB and NidA3B3 from *Mycobacterium vanbaalenii* PYR-1. *MBio* **1**, e00135-00110.

Lai, Q., Li, W., Wang, B., Yu, Z. and Shao, Z. (2012) Complete genome sequence of the pyrene-degrading bacterium *Cycloclasticus* sp. strain P1. *J Bacteriol* **194**, 6677-6677.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Liithy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85.

Lozada, M., Riva Mercadal, J.P., Guerrero, L.D., Di Marzio, W.D., Ferrero, M.A. and Dionisi, H.M. (2008) Novel aromatic ring-hydroxylating dioxygenase genes from coastal marine sediments of Patagonia. *BMC Microbiol* **8**, 50.

McGinnis, S. and Madden, T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nuc Acids Res* **32**, W20-W25.

Mahendra, S., Gedalanga, P., Kotay, S.M., Torres, C.I., Butler, C.S. and Goel, R. (2012) Advancements in molecular techniques and applications in environmental engineering. *Water Environ Res* **84**, 814-844.

Marcos, M.S., Lozada, M., Di Marzio, W.D. and Dionisi, H.M. (2012) Abundance, dynamics, and biogeographic distribution of seven polycyclic aromatic hydrocarbon dioxygenase gene variants in coastal sediments of Patagonia. *Appl Environ Microbiol* **78**, 1589-1592.

Marcos, M.S., Lozada, M. and Dionisi, H.M. (2009) Aromatic hydrocarbon degradation genes from polluted Subantarctic marine sediments. *Lett Appl Microbiol* **49**, 602-608.

Nikolaou, A., Kostopoulou, M., Lofrano, G., Meric, S., Petsas, A. and Vagi, M. (2009) Levels and toxicity of polycyclic aromatic hydrocarbons in marine sediments. *Trends Analyt Chem* **28**, 653-664.

Païssé, S., Goñi-Urriza, M., Coulon, F. and Duran, R. (2010) How a bacterial community originating from a contaminated coastal sediment responds to an oil input. *Microb Ecol* **60**, 394-405.

Parales, R.E. and Resnick, S.M. (2006) Aromatic Ring Hydroxylating Dioxygenases. In *Pseudomonas*. pp. 287-340. The Netherlands: Springer.

Parales, R.E. and Resnick, S.M. (2007) Applications of Aromatic Hydrocarbon Dioxygenases. In *Biocatalysis in the Pharmaceutical and Biotechnology Industries* ed. Patel, R.N. pp. 299-331. Boca Raton, FL: CRC Press.

Patil, K.R., Roune, L. and McHardy, A.C. (2012) The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS One* **7**, e38581.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

Pfaffl, M.W. (2004) Quantification Strategies in Real-Time PCR. In *AZ of Quantitative PCR* ed. Bustin, S.A. pp.89-113. La Jolla, CA: International University Line (IUL).

Singleton, D.R., Dickey, A.N., Scholl, E.H., Wright, F.A. and Aitken, M.D. (2015) Complete genome sequence of a novel bacterium within the family *Rhodocyclaceae* that degrades polycyclic aromatic hydrocarbons. *Genome Announc* **3**, e00251-15.

Sipos, R., Székely, A., Révész, S. and Márialigeti, K. (2010) Addressing PCR Biases in Environmental Microbiology Studies. In *Bioremediation: Methods and Protocols (Methods in Molecular Biology nº 599)* ed. Cummings, S.P. pp. 37-58. New York: Springer-Verlag.

Staley, J.T. (2010) *Cycloclasticus*: A Genus of Marine Polycyclic Aromatic Hydrocarbon Degrading Bacteria. In *Handbook of Hydrocarbon and Lipid Microbiology, part 19* ed. McGenity, T. pp. 1781-1786. Berlin Heidelberg: Springer.

Suenaga, H., Koyama, Y., Miyakoshi, M., Miyazaki, R., Yano, H., Sota, M., Ohtsubo, Y., Tsuda, M. and Miyazaki, K. (2009) Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. *ISME J* 3, 1335-1348.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molec Biol Evol* **30**, 2725-2729.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nuc Acids Res* **29**, 22-28.

Teira, E., Lekunberri, I., Gasol, J.M., Nieto-Cid, M., Álvarez-Salgado, X.A. and Figueiras, F.G. (2007) Dynamics of the hydrocarbon-degrading *Cycloclasticus* bacteria during mesocosm-simulated oil spills. *Environ Microbiol* **9**, 2551-2562.

Tomazetto, G., Wibberg, D., Schlüter, A. and Oliveira, V.M. (2015) New FeFe-hydrogenase genes identified in a metagenomic fosmid library from a municipal wastewater treatment plant, as revealed by high-throughput sequencing. *Res Microbiol* 166, 9-19.

Van Colen, C., Underwood, G.J., Serôdio, J. and Paterson, D.M. (2014) Ecology of intertidal microbial biofilms: Mechanisms, patterns and future research needs. *J Sea Res* **92**, 2-5.

Wang, B., Lai, Q., Cui, Z., Tan, T. and Shao, Z. (2008) A pyrene-degrading consortium from deep-sea sediment of the West Pacific and its key member *Cycloclasticus* sp. P1. *Environ Microbiol* **10**, 1948-1963.

Wang, W., Chen, K. and Xu, C. (2006) DNA quantification using EvaGreen and a real-time PCR instrument. *Anal Biochem* **356**, 303-305.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

Wilson, M.S., Bakermans, C. and Madsen, E.L. (1999) In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. *Appl Environ Microbiol* **65**, 80-87.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D. and Wishart, D.S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nuc Acids Res* **31**, 3316-3319.

Yang, T., Nigro, L.M., Gutierrez, T., Joye, S.B., Highsmith, R. and Teske, A. Pulsed blooms and persistent oil-degrading bacterial populations in the water column during and after the Deepwater Horizon blowout. *Deep Sea Research Part II: Topical Studies in Oceanography*. In press.

Yorio, P. (2009) Marine protected areas, spatial scales, and governance: implications for the conservation of breeding seabirds. *Conserv Lett* **2**, 171-178.

Yu, C.L., Liu, W., Ferraro, D.J., Brown, E.N., Parales, J.V., Ramaswamy, S., Zylstra, G.J., Gibson, D.T. and Parales, R.E. (2007) Purification, characterization, and crystallization of the components of a biphenyl dioxygenase system from *Sphingobium yanoikuyae* B1. *J Ind Microbiol Biotechnol* **34**, 311-324.

Zhang, C. and Anderson, A.J. (2012) Multiplicity of genes for aromatic ring-hydroxylating dioxygenases in *Mycobacterium* isolate KMS and their regulation. *Biodegradation* **23**, 585–596.

Zhou, H.W., Guo, C.L., Wong, Y.S. and Tam, N.F.Y. (2006) Genetic diversity of dioxygenase genes in polycyclic aromatic hydrocarbon-degrading bacteria isolated from mangrove sediments. *FEMS Microbiol Lett* **262**, 148-157.

Zhou, J., Bruns, M. and Tiedje, J. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**, 316-322.

FIGURE LEGENDS

Figure 1. Sampling locations. PF: Fracasso Beach, San José Gulf, Valdés Peninsula (non-impacted site, 42°25' S, 64°7' W), CC: Cordova Cove, San Jorge Gulf (chronically-polluted site, temperate climate, 45°45' S, 67°22' W), UB: Ushuaia Bay, Tierra del Fuego Island (chronically-polluted site, subantarctic climate, 54°48' S, 68°17' W).

Figure 2. Abundance of gene variants T and *phnA1* in sediment samples and experimental systems. qPCR assays targeting gene variants T and *phnA1*from *Cycloclasticus* spp., as well as bacterial 16S rRNA genes were applied to sediment samples and experimental systems from PF (A), CC (B), and UB (C) sampling locations. Relative abundance of the analyzed genes in sediment samples (PF08, CC08, OR08), slurries without further hydrocarbon addition (PF08-exp, CC08-exp, OR08-exp), slurries with phenanthrene addition (PF08-phe, CC08-phe, OR08-phe), slurries with pyrene addition (PF08-py, CC08-py, OR08-py), and slurries with crude oil addition (PF08-oil, CC08-oil, OR08-oil) are shown. Circle: detected at a concentration below the quantification limit of the assay (8×10^2 copies μg^{-1} DNA); *: not determined. D- Quantification of gene variant T in sediment samples retrieved from UB site in three consecutive years. Black bars, gene variant T; gray bars, *phnA1* gene; white bars, bacterial 16S rRNA genes.

Figure 3. Gene organization of metagenomic fragment M117. The predicted protein-coding sequences are shown with arrows, coloured by COG categories: dark blue, amino acid metabolism and transport [E]; orange, lipid metabolism [I]; yellow, coenzyme metabolism [H]; purple, cell cycle control and mitosis [D]; light green, nucleotide metabolism and transport [F]; pink, cell wall/membrane/envelope biogenesis [M]; brown, signal transduction [T]; light blue, translation [J]; red, inorganic ion transport and metabolism [P]; dark green, replication and repair [L]; fuchsia, energy production and conversion [C]; gray, general functional prediction only [R], function unknown [S] or not in COG. Detailed information of

the functional annotation of these sequences can be found in Table S3. Genes encoding α and β subunits of the terminal oxygenases are indicated with gray rectangles. The numbers of the coding sequences are indicated on the top, with the sequence corresponding to gene variant T shown with gray background.

Supporting Information

Table S1. Sediment samples, total PAH concentrations and molecular analyses performed in this study

Table S2. Primers and conditions used in this study for PCR-based analyses

Table S3. Putative function of the coding sequences identified in fosmid M117

Table S4. Best Blastp matches in the NCBI database to the oxygenase sequences identified in fosmid M117

Figure S1. Phylogenetic trees of RHO α-subunit sequences identified in fosmid M117. Neighbor-joining trees including metagenomic sequences (in red), previously classified sequences (Chakraborty *et al.* 2012) and the closest homologs of the metagenomic sequences from the NCBI database (bold fonts). Sequence name and strain name (in brackets) is indicated in each case. RHO classification according to the scheme proposed for Chakraborty *et al.* (2012) is indicated on the right. Figure S2a, class A RHOs; Figure S2b, class B RHOs; Figure S2c, class C RHOs. Bootstrap values were calculated as percentage of 1,000 replicates, with only values ≥50% shown in the figure. The scale bar represents the inferred amino acid changes per position.

Figure S2. Active sites of the modelled oxygenases. Stick representations identify the amino acids interacting with the catalytic iron (sphere), whereas line representations show the amino acids involved in substrate-binding. A) M117-33; B) M117-36; C) M117-38. The amino acids composing the active site of the templates, identified from previous works [(Jakoncic *et al.* 2007) for A, (Ferraro *et al.* 2007) for B and C], are shown in green. The amino acids belonging to the modelled enzyme are shown in red, and only those residues differing from the template sequence are identified in red. The figure was constructed with PyMOL (0.99RC6).

Figure S3. Topology of the catalytic cavities of the modelled oxygenases. A) oxygenase from *Sphingomonas sp* CHY-1 (PDB 2CKF) (Jakoncic *et al.* 2007); B) oxygenase from *Sphingomonas yanoikuyae* (PDB 2GBW) (Ferraro *et al.* 2007); C) M117-33 (template 2CKF). D) M117-36 (template 2GBW). E) M117-38 (template 2GBW). The spatial disposition of pyrene was obtained from docking analysis. The sphere represents the catalytic iron.

Table S5. Active site dimensions of the modelled class A oxygenases and their respective templates

Table S6. Pose and Rank scores for docking analysis of different complexes between the oxygenases and phenanthrene or pyrene

Table 1. Analysis of the RHO α-subunit gene variants identified in the PCR clone libraries.

Gene variant	Sequence length (bp)	Library[a] (% clones)[b]	Closest homolog[c]	Accession number	Identity[d] (%)	Query coverage (%)
E	257	UB (38) OR08-oil (38)	Aromatic ring-hydroxylating dioxygenase, partial [uncultured bacterium], intertidal sediments, Ushuaia Bay, Argentina	CAP60623	85	98
			Naphthalene 1,2-dioxygenase [*Polycyclovorans algicola* TG408]	WP_029889175	84	98

T	260	UB (50) OR08-oil (62) CC (75)	Aromatic-ring-hydroxylating dioxygenase subunit alpha-like protein [*Cycloclasticus* sp. P1]	YP_006837517	67	96
U	245	UB (12)	(2Fe-2S)-binding protein [*Novosphingobium malaysiense*]	WP_039290189	60	98
V	275	CC (25)	Naphthalene dioxygenase iron sulfur protein, partial [*Pseudomonas* sp. 5N1-1]	CAD42906	47	96

[a]Libraries: UB, library constructed from Ushuaia Bay sediment samples OR06, OR07, OR08 and EM06 (Lozada *et al.* 2008; Marcos *et al.* 2012); OR08-oil, library constructed from OR08-oil experimental system (Guibert *et al.* 2012); CC, library constructed from Cordova Cove sediment samples CC08-1, CC08-2, CC10-1 and CC10-2 (Marcos *et al.* 2012). Further information can be found in Material and Method and Table S1.

[b]Percentage of clones of the PCR-clone library corresponding to each gene variant is indicated in parentheses.

[c]Closest homolog of the NCBI database, showing the highest Blastp maximum score.

[d]Percent identity at the amino acid level shared with the closest homolog.