

ORIGINAL ARTICLE

# Aromatic hydrocarbon degradation genes from chronically polluted Subantarctic marine sediments

M.S. Marcos, M. Lozada and H.M. Dionisi

Environmental Microbiology Laboratory, Patagonian National Research Center (CENPAT – CONICET), Puerto Madryn, Chubut, Argentina

## Keywords

dioxygenases, intertidal sediments, polycyclic aromatic hydrocarbons, ring-hydroxylating oxygenases, Subantarctic marine environments.

## Correspondence

Hebe M. Dionisi, Environmental Microbiology Laboratory, Patagonian National Research Center, (CENPAT – CONICET), Boulevard Brown 2915, Puerto Madryn (U9120ACF), Chubut, Argentina.  
E-mail: hdionisi@cenpat.edu.ar

2009/0219: received 4 February 2009,  
revised 2 June 2009 and accepted  
3 August 2009

doi:10.1111/j.1472-765X.2009.02711.x

## Abstract

**Aim:** The goal of this study was to identify functional targets to detect polycyclic aromatic hydrocarbon (PAH)-degrading bacterial populations in cold marine ecosystems.

**Methods and Results:** We designed a degenerate primer set targeting genes encoding the  $\alpha$  subunit of PAH-dioxygenases from Gram-positive bacteria. This primer set was used to amplify gene fragments from metagenomic DNA isolated from Subantarctic marine sediments (Ushuaia Bay, Argentina). These gene fragments were cloned and sequenced. We identified 14 distinct groups of genes, most of them showing significant relatedness with dioxygenases from Gram-positive bacteria of the genera *Rhodococcus*, *Mycobacterium*, *Nocardioides*, *Terrabacter* and *Bacillus*. The level of identity with these genes, however, was low to moderate (33–62% at the amino acid level).

**Conclusion:** These results indicate the presence of a high diversity of hitherto unidentified dioxygenase genes in this cold polluted environment.

**Significance and Impact of the Study:** Subantarctic marine ecosystems are particularly vulnerable to hydrocarbon pollution, and the development of environmental restoration strategies for these environments is pressing. The information obtained in this work will be the starting point for the design of quantitative molecular tools to analyse the abundance and dynamics of these aromatic hydrocarbon-degrading bacterial populations in the marine environment.

## Introduction

Cold coastal regions are particularly vulnerable to hydrocarbon pollution resulting from both continuous low-level inputs and major tanker accidents (Margesin and Schinner 1999). Although natural attenuation processes are known to reduce contamination levels even in cold climates, low temperatures hamper hydrocarbon biodegradation rates and affect the physicochemical behaviour of hydrocarbons (Coulon *et al.* 2007). Subantarctic marine ecosystems may be even more susceptible to oil pollution because of their increased UV radiation levels, as a synergistic effect of UV and hydrocarbon exposure on marine organisms has been observed (Sargian *et al.* 2007). Therefore, the development of adequate remedial options for this region is pressing. There is tremendous potential for the use of high-throughput molecular methods for

analysing the presence, relative abundance and dynamics of key hydrocarbon-degrading bacterial populations. However, insufficient knowledge regarding key biomarkers is a major barrier for their implementation in field diagnostics and for monitoring bioremediation performance (SERDP 2005). Although functional biomarkers do not provide phylogenetic information about the host organisms because of the possibility of horizontal gene transfer events (Springael and Top 2004), they are more appropriate for this use, as are able to provide information about the degradative potential of the detected populations.

Rieske nonheme iron aromatic ring-hydroxylating oxygenases (RHOs) catalyse the insertion of molecular oxygen into benzene rings, a common first step in the bacterial degradation of aromatic compounds (Ferraro *et al.* 2005). RHOs are multicomponent enzymes,

comprising a reductase that obtains electrons from NAD(P)H, often a ferredoxin that shuttles the electrons and a catalytic oxygenase component (Ferraro *et al.* 2005). The oxygenase, frequently incorporating both atoms of dioxygen into the substrate, can have either a homo- ( $\alpha_n$ ) or a hetero-oligomeric ( $\alpha_n\beta_n$ ) structure (Kweon *et al.* 2008). The  $\alpha$  subunits contain the catalytic domain, and genes encoding for this subunit are often used as markers to study the biodegradation of aromatic compounds (Gibson and Parales 2000). Our goal was to identify functional targets for polycyclic aromatic hydrocarbon (PAH) degradation in chronically polluted Subantarctic marine sediments. Using a primer set designed based on conserved regions of dioxygenase  $\alpha$  subunits from PAH biodegradation pathways previously found in Gram-positive bacteria, we identified 14 distinct groups of dioxygenase gene fragments in coastal sediments, most of them distantly related to previously described sequences. This work significantly expands our knowledge of aromatic hydrocarbon biodegradation genes in cold marine ecosystems.

## Materials and methods

### Sampling and DNA extraction

Sampling sites are located in Ushuaia Bay, Argentina. Biogeographically, these sites are situated within the Subantarctic subregion according to a delimitation scheme proposed for Latin America and the Caribbean (Morrone 2000). The sites are located in close proximity to a pier (Orion Plant) used for loading and unloading of petroleum-derived products. As a consequence, the coastal sediments at these sites are frequently exposed to small accidental hydrocarbon spillages. In addition, this area is subjected to high UV-B radiation levels, and the coastal environment receives the input of pluvial discharges, untreated sewage and small rivers, causing spatial and temporal fluctuations in salinity and nutrient concentrations (Gil *et al.* 2006). More detailed information about sampling sites can be found at Lozada *et al.* (2008).

Composite surficial sediments were sampled manually along the low tide line at seven to ten random points per site, using acrylic cores with an inner diameter of 4.4 cm. The sediment surficial layer (0–3 cm) was obtained by inserting the core 3 cm depth and releasing the extracted column. Samples were processed as previously described (Lozada *et al.* 2008). High molecular mass DNA was purified using a FastDNA<sup>®</sup> SPIN Kit for Soil (Q-BIOgene, Carlsbad, CA, USA), according to the manufacturer's instructions with minor modifications, as reported in Lozada *et al.* (2008).

### PCR, cloning and sequence analysis

Amplification of dioxygenase gene fragments were carried out as previously described (Lozada *et al.* 2008), with the following modifications: (i) primers NMR331f (5'-TGCC CKTACCACGGYTGG-3') and NMR1134r (5'-CTCGGCG TCGTCCTGYTC-3') were used for amplification and (ii) the amplification programme included 5 min at 94°C, 40 cycles of 30 s at 94°C, 30 s at 53°C and 60 s at 72°C, with an elongation step of 30 min at 72°C. Catabolic gene libraries were constructed using pCR<sup>®</sup>4.0 vector (TA Cloning kit for sequencing; Invitrogen, Carlsbad, CA, USA). Clones were sequenced commercially at Macrogen (Seoul, Korea), and the obtained sequences were compared to the GenBank database, using the Basic Local Alignment Search Tool TBLASTX (McGinnis and Madden 2004). Phylogenetic trees were constructed using the neighbour-joining algorithm in the Molecular Evolutionary Genetics Analysis software MEGA4 (Tamura *et al.* 2007). To test the inferred phylogeny, a bootstrap test with 1000 replications was used. Deduced amino acid sequences from representative clones were also compared to three additional databases: Clusters of Orthologous Groups (COG, <http://www.ncbi.nlm.nih.gov/COG/>, 27 May 2009), Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>, 27 May 2009) and The Gene Ontology (GO, <http://www.geneontology.org/>, 27 May 2009).

### Diversity analysis

The DOTUR software (Schloss and Handelsman 2005) was used to group sequences based on distance at the amino acid level. The Shannon's diversity and Simpson's dominance indexes (Shannon and Weaver 1949; Simpson 1949) were also computed using DOTUR, while the Chao's Jaccard abundance-based similarity index (Chao *et al.* 2005) was calculated using ESTIMATES (Colwell 2005). Library coverage indexes were calculated as described by Good (1953).

### Nucleotide sequence accession numbers

The sequences determined in this study have been deposited at the GenBank nucleotide database under accession numbers FJ493400–FJ493456.

## Results

### Library construction

A degenerate primer set was designed targeting the large subunit of PAH dioxygenase genes from Gram-positive

bacteria available at the GenBank database by June 2004. The alignment included genes identified in the marine bacterium *Nocardioides* sp. KP7 (AB017794), in *Mycobacterium* spp. (AF546905, AF548347, AF548345, AJ494745, AF546904, AF249301, AF548343, AB179737, AY365117, AY330100, AY330098, AY330102) and in *Rhodococcus* spp. (AY392423, AY392424, AF121905, AB110633, AF082663, AJ401612, AB024936). When this primer set was recently compared to the GenBank database using the BLAST program, 12 more PAH dioxygenase genes from Gram-positive bacteria belonging to the genera *Rhodococcus* and *Mycobacterium* showed perfect matches with both primers. No perfect matches with dioxygenase genes from Gram-negative bacteria were found.

Six sediment samples retrieved from Ushuaia Bay, Tierra del Fuego Island, Argentina [SC04, EM06, OR04, OR05, OR06 and OR07, (Lozada *et al.* 2008)] were analysed to detect the presence of PAH dioxygenase genes related to Gram-positive bacteria using primers NMR331f and NMR1134r. Only two samples from the same site, OR05 and OR07, produced amplification products of the expected size (data not shown), which were used to generate libraries NMR-OR05 and NMR-OR07. However, these genes could still be present in the other four samples (SC04, EM06, OR04 and OR06), although under the detection limit of the technique. Seventy-four clones from library NMR-OR05 and 96 clones from library NMR-OR07 were sequenced and compared to the GenBank database. Clones showing no matches with dioxygenases were removed from further analyses, as these fragments were assumed to be produced as the result of mispriming events.

### Library analyses

Sixty-two sequences from both libraries produced significant alignments with dioxygenase sequences deposited at the GenBank database, and most of these matches were associated with aromatic hydrocarbon degradation genes from Gram-positive bacteria of the genera *Rhodococcus*, *Mycobacterium*, *Nocardioides*, *Terrabacter* and *Bacillus*. Protein sequence identities with these genes, however, were rather low, ranging from 33% to 62% in the analysed fragment (248–271 amino acids).

We used the DOTUR software (Schloss and Handelsman 2005) to group these sequences based on their distance. Considering identities at the protein level  $\geq 86.7\%$  among members of the same group, this software identified 14 distinct groups. At this level of resolution, sequence identities between members of different groups were  $\leq 67\%$ . To continue with a previous nomenclature (Lozada *et al.* 2008), these groups were named F–S. We compared deduced amino acid sequences from clones assigned to each of these groups with COG, KEGG and GO databases

(Supporting information Table S1) to infer the possible function of these fragments. These curated databases are currently being used to reliably predict gene function from the short sequences typically generated in metagenomic studies, such as pyrosequencing reads (Mou *et al.* 2008). All the analysed sequences matched proteins described as dioxygenases or putative dioxygenases, with *E*-values lower than  $10^{-15}$ ,  $10^{-22}$  and  $10^{-19}$ , for COG, KEGG and GO databases, respectively. These results indicate that the 14 groups of sequences identified in this study encode the  $\alpha$  subunit of dioxygenases present in these Subantarctic, chronically polluted marine sediments.

Both libraries were equally diverse and showed low dominance values (Table 1). However, they were only moderately similar (Chao's Jaccard abundance-based similarity index 0.59), as only 4 of the 14 groups (F, G, I and M) were shared by both libraries. It must be noted, however, that these analyses were performed to compare both gene libraries, and these diversity values cannot be extrapolated to the microbial community (Lozada *et al.* 2008).

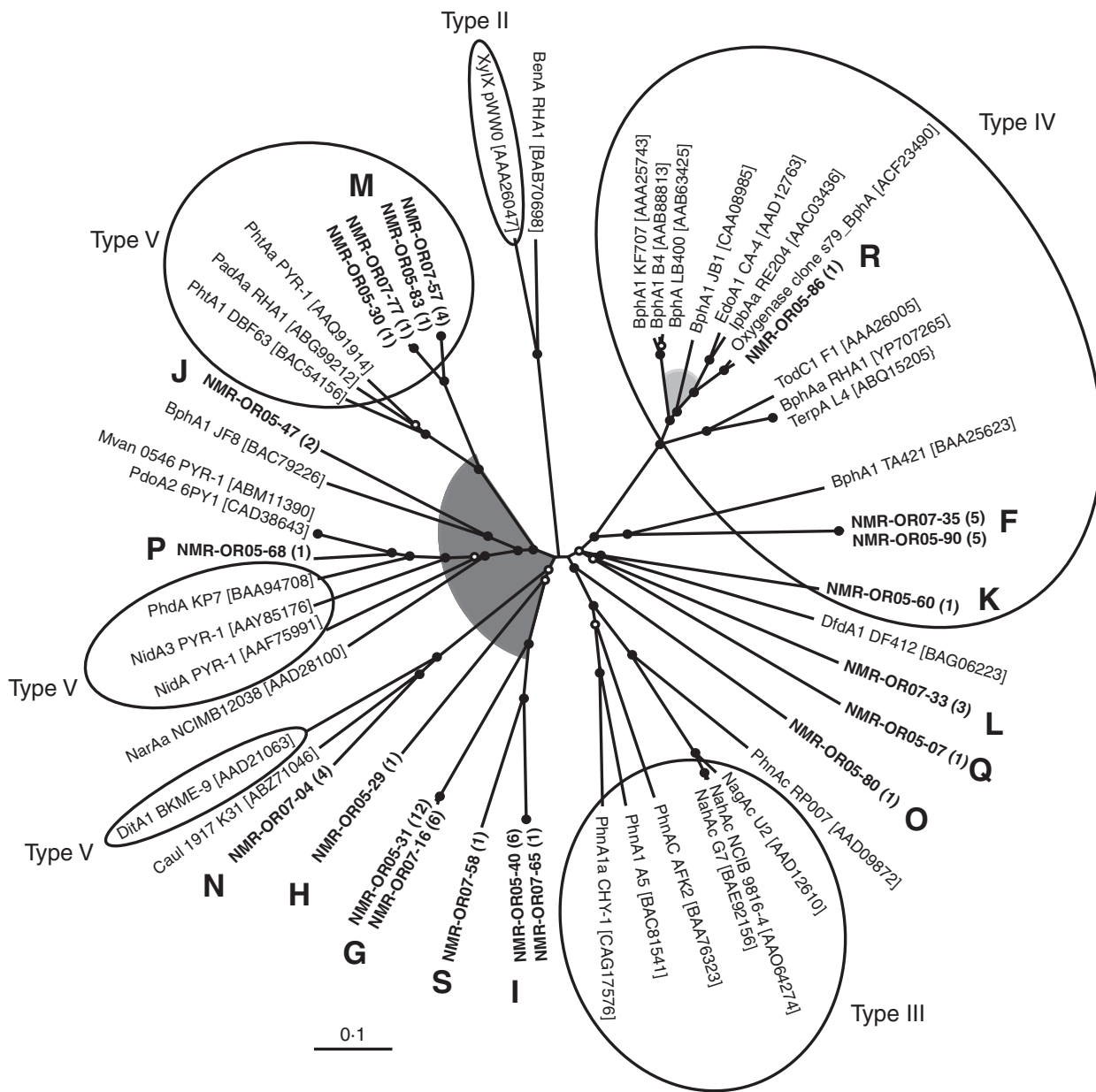
### Phylogenetic analyses

Figure 1 shows the phylogenetic relationships among the dioxygenase sequences identified in this study and with selected sequences from the GenBank database. Most sequences formed deeply rooted branches with previously described dioxygenases. In an effort to classify these highly divergent sequences, we used a new classification system recently proposed by Kweon *et al.* (2008). This classification recognizes five distinct RHO types taking into consideration not only the sequence information, but also the interactions between the enzyme components. This is relevant because RHOs are multicomponent enzymes, as both the oxygenase and the electron transport chain components need to be considered in a functional approach (Kweon *et al.* 2008). Moreover, these authors demonstrated that distance analysis of protein sequences of the oxygenase component can be used for this classification when the information for the electron transport chain components is not available (Kweon *et al.* 2008). Using this strategy, they could establish a threshold value of 0.61 for the reliable assignment of sequences to a

**Table 1** Diversity analysis of libraries NMR-OR05 and NMR-OR07

Library	<i>S</i>	<i>H</i>	<i>L</i>	<i>S</i> <sub>n</sub>	<i>C</i> (%)
NMR-OR05	11	1.94	0.18	6	81.8
NMR-OR07	7	1.79	0.15	2	92.0

*S*, number of dioxygenase groups; *H*, Shannon's diversity index; *L*, Simpson's dominance index; *S*<sub>n</sub>, singletons (number of species with only one clone); *C* (%), library coverage.



**Figure 1** Phylogenetic relationships among clone sequences and selected dioxygenases from the GenBank database. Bootstrap percentages higher than 50% are shown in the nodes (51–75%: white circles; 76–100%: black circles). Scale bar: substitutions per amino acid position. Shaded areas: sequences containing indels (dark grey: indel at position 321; light grey: indel at position 146). Oxygenase types II–V are marked in circles. Clone sequences from this study are boldfaced (groups F–S). Sequences 98% identical or more at the amino acid level were grouped; the number of clones represented by these sequences is shown in parenthesis. Enzyme names, strains and accession numbers are indicated in the tree.

particular enzyme type. We established new threshold values for each oxygenase type, from distance values recalculated based on the positions included in our dioxygenase fragments.

According to this analysis, group M clearly falls within type V enzyme (mean distance 0.58, threshold value 0.62), while distance values for group P are borderline for

this enzyme type (mean distance 0.62). Type V enzymes are very diverse in terms of substrate, and are characterized by the presence of [3Fe-4S]-type ferredoxins. This type includes enzymes typically found in Gram-positive micro-organisms (mostly *Actinobacteria*): phenanthrene/naphthalene/anthracene dioxygenase from the marine bacterium *Nocardioides* sp. KP7, phenanthrene and pyrene

dioxygenases from *Mycobacterium*, as well as phthalate dioxygenases from *Rhodococcus*, *Mycobacterium*, *Terrabacter* and *Arthrobacter* (Kweon *et al.* 2008). Dioxygenase groups G, H, I, J, N and S present distance values too high to allow their inclusion in this enzyme type. However, analysing amino acid sequence alignments, we detected the presence of an insertion that was shared by the above-mentioned groups, together with all sequences belonging to type V (Fig. 1), supporting a common origin. Conserved insertions and deletions (i.e. indels) are signature sequences located at the same position in homologous sequences and flanked by conserved sequence blocks. They have been used to infer evolutionary relationships among genes, as they rarely occur because of independent mutational events (Gupta 1998). This indel starts at position 321 (*Mycobacterium vanbaaleonii* PYR-1 NidA numbering, Supporting information Fig. S1a) and is not fully conserved. The dendrogram produced from the alignment of the positions corresponding to this indel largely reflects the phylogenetic relationships of the complete fragment (data not shown), suggesting that this insertion was acquired at an early stage of the evolutionary divergence of type V RHOs.

On the other hand, groups F, K and R belong to the type IV enzyme system (0.68, 0.68 and 0.50, respectively, threshold 0.70). This enzyme type includes mainly biphenyl and polychlorinated biphenyl dioxygenases from diverse genera and is characterized by the presence of [2Fe-2S]-type ferredoxins in the electron transport chain (Kweon *et al.* 2008). A second indel was found in all the sequences from a monophyletic group that includes group R from this study and part of type IV enzymes (Fig. 1). This indel starts at position 146 (*Burkholderia xenovorans* LB400 BphA numbering, Supporting information Fig. S1b) and is completely conserved, suggesting that it has been recently acquired. The three remaining groups of sequences found in this work (L, Q and O) could not be assigned to a particular enzyme type.

## Discussion

In a recent study, we characterized PAH-degrading bacteria in intertidal sediments from the Argentinean coast of Patagonia using a functional, culture-independent approach, targeting genes encoding the  $\alpha$  subunit of PAH dioxygenases previously described in Gram-negative bacteria (Lozada *et al.* 2008). In that study, we identified five novel groups of dioxygenase genes in sediments from a chronically polluted environment of Ushuaia Bay. In this study, we continued with the characterization of PAH-degrading populations from coastal marine sediments of Ushuaia Bay, targeting dioxygenase genes related to those previously found in Gram-positive bacteria. In the marine

environment, the most widely studied Gram-positive PAH-degrading strain has been *Nocardioides* sp. KP7 (Saito *et al.* 2000), collected from a polluted beach in Kuwait and capable of growing at 42°C on phenanthrene as a sole carbon source. Naphthalene degraders of various genera have also been isolated from tropical intertidal sediments, although no information about their catabolic genes is available (Zhuang *et al.* 2003). Environmental conditions at these places are very different from cold environments such as Subantarctic sediments. In this work, we identified 14 distinct groups of dioxygenase sequences in two libraries constructed from intertidal sediment samples obtained from the same site at different seasons 18 months apart. Four of these groups were found in both libraries, suggesting that populations containing these genes are fairly stable and relatively abundant within the microbial community indigenous at this site. None of the sequences used to design the primers was found in the libraries; moreover, the level of relatedness with previously described sequences was rather low.

There are important knowledge gaps on the structure and diversity of RHO enzymes, making them difficult to classify. In addition, there seems to be a sampling bias toward the Northern hemisphere for the characterization of hydrocarbon degrading bacterial populations (Yakimov *et al.* 2007). Perhaps because of these facts, only four out of the 14 dioxygenase groups identified in this study could be reliably assigned to enzyme types described by Kweon by distance analysis (Kweon *et al.* 2008). Nevertheless, we identified an insertion–deletion which seems to be distinctive of enzyme type V, as it is present in all sequences belonging to this type but absent in sequences belonging to the other described enzyme types. This indel allowed us to relate seven more groups with enzyme type V, mainly found in Gram-positive bacteria. Further studies of complete genome regions will be needed to identify the other genes of these multicomponent enzymes, allowing the conclusive assignment of these sequences to a particular enzyme type.

Using this new classification system (Kweon *et al.* 2008), we also analysed the novel dioxygenase sequences identified in sediments of Ushuaia Bay in a previous study (Lozada *et al.* 2008). In contrast with the sequences found in this study, all five groups of dioxygenases clearly fell within the type III enzyme, which includes naphthalene and phenanthrene dioxygenases from micro-organisms of the Phylum Proteobacteria, such as *Pseudomonas*, *Cycloclasticus*, *Ralstonia*, *Sphingomonas*, *Alcaligenes* and *Burkholderia*. Considering both studies, we detected three out of the five described groups of RHO enzymes in this Subantarctic marine environment. Using quantitative molecular assays such as real-time PCR or microarrays, it will be possible to analyse the ecological significance of the bacterial

populations carrying these genes and ultimately to identify the key players for aromatic hydrocarbon biodegradation in this highly vulnerable environment.

## Acknowledgements

This work was supported by grants from the National Research Council of Argentina (CONICET) and the Global Environmental Facility (Patagonian GEF, administered by the Environment and Sustainable Development Secretariat, Argentina). M.S.M. is a fellow, and M.L. and H.M.D. are staff members from CONICET. We thank M. Gil, J.L. Esteves, H. Ocariz and A. Torres for their help in sample collection.

## References

- Chao, A., Chazdon, R.L., Colwell, R.K. and Shen, T.-J. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* **8**, 148–159.
- Colwell, R.K. (2005) *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples*, ver 7.5. User's Guide and Application. <http://purl.oclc.org/estimates>.
- Coulon, F., McKew, B.A., Osborn, A.M., McGenity, T.J. and Timmis, K.N. (2007) Effects of temperature and biostimulation on oil-degrading microbial communities in temperate estuarine waters. *Environ Microbiol* **9**, 177–186.
- Ferraro, D.J., Gakhar, L. and Ramaswamy, S. (2005) Rieske business: structure–function of Rieske non-heme oxygenases. *Biochem Biophys Res Commun* **338**, 175–190.
- Gibson, D.T. and Parales, R.E. (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr Opin Biotechnol* **11**, 236–243.
- Gil, M., Amín, O., Torres, A.I. and Esteves, J.L. (2006) Variabilidad de la calidad de agua del sistema Bahía Ushuaia-Bahía Golondrina (BU-BG), Tierra del Fuego, Argentina. VI Jornadas Nacionales de Ciencias del Mar. Puerto Madryn, Chubut, Argentina.
- Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264.
- Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among Archaeobacteria, Eubacteria, and Eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.
- Kweon, O., Kim, S.-J., Baek, S., Chae, J.-C., Adjei, M.D., Baek, D.-H., Kim, Y.-C. and Cerniglia, C.E. (2008) A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem* **9**, 11.
- 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.
- Margesin, R. and Schinner, F. (1999) Biological decontamination of oil spills in cold environments. *J Chem Technol Biotechnol* **74**, 381–389.
- McGinnis, S. and Madden, T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20–W25.
- Morrone, J.J. (2000) Biogeographic delimitation of the Subantarctic subregion and its provinces. *Revista del Museo Argentino de Ciencias Naturales* **2**, 1–15.
- Mou, X., Sun, S., Edwards, R.A., Hodson, R.E. and Moran, M.A. (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**, 708–711.
- Parales, R.E. and Resnick, S.M. (2004) Aromatic hydrocarbon dioxygenases. In *Biodegradation and Bioremediation* ed. Sing, A. and Ward, O.P. pp. 175–195. Berlin: Springer.
- Saito, A., Iwabuchi, T. and Harayama, S. (2000) A novel phenanthrene dioxygenase from *Nocardioideis* sp. strain KP7: expression in *Escherichia coli*. *J Bacteriol* **182**, 2134–2141.
- Sargian, P., Mas, S., Pelletier, É. and Demers, S. (2007) Multiple stressors on an Antarctic microplankton assemblage: water soluble crude oil and enhanced UVR level at Ushuaia (Argentina). *Polar Biol* **30**, 829–841.
- Schloss, P.D. and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**, 1501–1506.
- Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Simpson, E.H. (1949) Measurement of diversity. *Nature* **163**, 688.
- Springael, D. and Top, E.M. (2004) Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol* **12**, 53–58.
- Strategic Environmental Research and Development Program (SERDP) (2005) *SERDP and ESTP Expert Panel Workshop on Research and Development Needs for the Environmental Remediation Application of Molecular Biological Tools*. Charlottesville, Virginia: Strategic Environmental Research and Development program and Environmental Security Technology Certification Program.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.
- Yakimov, M.M., Timmis, K.N. and Golyshin, P.N. (2007) Obligate oil-degrading bacteria. *Curr Opin Biotechnol* **18**, 257–266.
- Zhuang, W.Q., Tay, J.H., Maszenan, A.M., Krumholz, L.R. and Tay, S.T.L. (2003) Importance of Gram-positive naphthalene-degrading bacteria in oil-contaminated tropical marine sediments. *Lett Appl Microbiol* **36**, 251–257.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** (a) Partial sequence alignment of dioxygenases showing an indel at position 321. (b) Partial sequence alignment of dioxygenases showing an indel at position 146.

**Table S1.** Match results of clone sequences from libraries NMR-OR05 and NMR-OR07 in four different databases: NCBI, COG, KEGG and GO.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.