AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

Functional Metagenomics of a Biostimulated Petroleum-Contaminated Soil Reveals an Extraordinary Diversity of Extradiol Dioxygenases

Laura Terrón-González,[a,b]* Guadalupe Martín-Cabello,[a,b] Manuel Ferrer,[c] Eduardo Santero[a,b]

Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide/Consejo Superior de Investigaciones Científicas/Junta de Andalucía,[a] and Departamento de Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide,[b] Seville, Spain; Instituto de Catálisis y Petroleoquímica, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain[c]

A metagenomic library of a petroleum-contaminated soil was constructed in a fosmid vector that allowed heterologous expression of metagenomic DNA. The library, consisting of 6.5 Gb of metagenomic DNA, was screened for extradiol dioxygenase (Edo) activity using catechol and 2,3-dihydroxybiphenyl as the substrates. Fifty-eight independent clones encoding extradiol dioxygenase activity were identified. Forty-one different Edo-encoding genes were identified. The population of Edo genes was not dominated by a particular gene or by highly similar genes; rather, the genes had an even distribution and high diversity. Phylogenetic analyses revealed that most of the genes could not be ascribed to previously defined subfamilies of Edos. Rather, the Edo genes led to the definition of 10 new subfamilies of type I Edos. Phylogenetic analysis of type II enzymes defined 7 families, 2 of which harbored the type II Edos that were found in this work. Particularly striking was the diversity found in family I.3 Edos; 15 out of the 17 sequences assigned to this family belonged to 7 newly defined subfamilies. A strong bias was found that depended on the substrate used for the screening: catechol mainly led to the detection of Edos belonging to the I.2 family, while 2,3-dihydroxybiphenyl led to the detection of most other Edos. Members of the I.2 family showed a clear substrate preference for monocyclic substrates, while those from the I.3 family showed a broader substrate range and high activity toward 2,3-dihydroxybiphenyl. This metagenomic analysis has substantially increased our knowledge of the existing biodiversity of Edos.

Large amounts of several types of contaminants are released into the environment due to industrial activities and accidental spills. Many of these contaminants, such as aromatic hydrocarbons, have complex and stable chemical structures, which make them prevail in the environment for a long time, thus resulting in the contamination of ecosystems. Biological treatment of industrial wastewater and bioremediation of contaminated soils and water are therefore critical to prevent or combat this contamination.

Standardized culturing techniques have been successfully applied for many years to isolate many different types of bacteria that are capable of utilizing a variety of aromatic hydrocarbons and to characterize their biodegradation pathways and the coding genes responsible for these capabilities, which have shown that bacteria can adapt to utilize a plethora of organic contaminants, including xenobiotics. The aerobic biodegradation of aromatics has been well documented and has been shown to follow two pathways, involving either intradiol or extradiol cleavage of the aromatic rings of di- or trihydroxylated intermediates. The intradiol dioxygenases that have been described to date appear to belong to the same superfamily, but extradiol dioxygenases (Edos) of three different phylogenetic origins have been reported (1, 2). Type I extradiol dioxygenases are more numerous and belong to the vicinal oxygen chelate superfamily (3). They can be small one-domain (also designated class I extradiol dioxygenases) (4) or large two-domain (class II) enzymes. Type II or LigB superfamily (also called class III) Edos comprise unrelated dioxygenases with homomeric or heteromeric subunit compositions (5, 6, 7). Type III Edos, such as gentisate dioxygenases, belong to the cupin superfamily (8). In addition, other unrelated dioxygenases can cleave the aromatic rings of intermediates such as hydroquinone (9, 10).

However, in spite of their excellent performance under laboratory conditions, the application of promising bacteria in the field has been disappointing (11). The extraordinary potential of contaminant-degrading bacteria has not yet been sufficiently exploited, probably because of our limited knowledge of the catabolic capabilities and the performance levels of the microbial communities in the environment (2). A better knowledge of the biodegradation potential of the natural environment and the processes that are involved as well as the diversity and ecology of biodegrading bacterial strains would improve our strategies for decontaminating polluted sites.

Cultivated microorganisms represent only a very small fraction of the actual diversity of microorganisms that are present in nature (12, 13) and therefore provide a very limited picture of the actual capabilities of natural niches in biodegrading contaminants (2). Metagenomic analysis potentially allows access to all of the genetic resources that are present in an environment, regardless of whether they belong to microorganisms that can be cultured in the laboratory, and therefore represents an alternative approach to unraveling the biodiversity of microorganisms and their activities

Received 24 November 2015 **Accepted** 19 January 2016

Accepted manuscript posted online 19 February 2016

Citation Terrón-González L, Martín-Cabello G, Ferrer M, Santero E. 2016. Functional metagenomics of a biostimulated petroleum-contaminated soil reveals an extraordinary diversity of extradiol dioxygenases. Appl Environ Microbiol 82:2467–2478. doi:10.1128/AEM.03811-15.

Editor: R. E. Parales, University of California—Davis

Address correspondence to Eduardo Santero, esansan@upo.es.

* Present address: Laura Terrón-González, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03811-15.

(14). Metagenomic functional analyses, which are based on the detection of phenotypic changes in a bacterial host due to its acquisition of a metagenomic library clone encoding a particular activity, have the advantage of allowing straightforward identification of activities even if the coding genes are so unrelated to those in databases that they cannot be detected by sequence analysis.

Edos, the key enzymes defining biodegradation pathway types, have been used as functional markers for assessing the biodegradation potentials of bacterial communities (2). A functional screening based on the development of a yellow color resulting from extradiol ring cleavage in catechol has been successfully used to detect genes coding for Edo activity in a number of previous metagenomic functional analyses (15, 16, 17). In spite of the large diversity of already known Edo sequences, some sequences have led to the definition of new subfamilies of type I Edos and have contributed to our understanding of catabolic pathways *in situ*. However, even in the most extensive analyses, intrinsic biodiversity has not been very high because a large fraction of the identified Edos were very similar, differing by only a few amino acids (15, 16).

In general, metagenomic functional analyses have two main limitations. On the one hand, a lack of expression of metagenomic DNA in a surrogate host where a DNA library is maintained may prevent the detection of clones from highly unrelated bacteria, thus favoring the identification of genes in bacteria that are more closely related to the host. On the other hand, the indicator substrate that is used for functional screening may bias the identification of enzymes in favor of those more active toward substrates that are most similar to the indicator. In an attempt to circumvent these limitations, we used a fosmid expression vector that promotes the transcription of metagenomic DNA to construct a metagenomic library and two different indicators during functional screening to perform a new metagenomic functional analysis of Edos from a soil contaminated with crude oil. The identified Edos led to the definition of 10 new subfamilies of type I Edos and to the ascription of metagenomic type II Edos to 2 newly defined families of type II enzymes, thus representing the highest biodiversity of Edos reported for a particular site.

MATERIALS AND METHODS

Plasmids, bacterial strains, and growth conditions. The fosmid expression vector pMPO579 (18) was used to construct the metagenomic library. The helper plasmid pRK2013 (19) in DH5α [F$^-$ $^{\phi 80d}$*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17 supE44 thi-1 gypA relA1*] (20) was used to transfer the metagenomic library by conjugation to other strains. The *Escherichia coli* strain EPI300-T1^R [F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) (Str^r) ϕ80d*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(*ara, leu*)*7697 galU galK* λ$^-$ *rpsL nupG trfA tonA dhfr*] was used to transfect and maintain the metagenomic library. Its derivative MPO554 Nal^r, bearing an N antitermination system (18), was used for activity screening. The *E. coli* strains were routinely grown in Luria-Bertani (LB) medium at 37°C.

Site description and sample collection. Soil samples (approximately 40 g, wet weight) were obtained from a parcel on the southern Iberian Peninsula contiguous to a refinery plant (Huelva, Andalucía, Spain; lat 37.1890, long −6.9079; altitude of 126 m). The site has been anthropogenically influenced and is a crude oil-contaminated site; the contaminated soil was subjected to biostimulation for 6 months. The top biostimulated soil was sampled at a depth of 5 to 10 cm on 22 June 2010, and the samples were kept in open plastic bags in the dark at room temperature and processed immediately.

Construction of the metagenomic library. First, bacteria were extracted by the direct addition of 100 ml of disruption buffer (0.2 M NaCl, 50 mM Tris-HCl, pH 8.0) to 40 g of the soil sample, which was mixed overnight with shaking at 4°C. The suspension was centrifuged at low speed (400 × *g* for 3 min), and the supernatant was poured over 10 ml of a Nycodenz resin (Axis-Shield) solution (1.3 g ml^{-1} in disruption buffer) and centrifuged again at 10,000 × *g* for 40 min at 4°C. The bacterium-containing band at the interface between the Nycodenz and the aqueous layer was recovered and mixed with phosphate-buffered saline (PBS) (136.89 mM NaCl, 2.70 mM KCl, 10 mM Na_2HPO_4, 1.98 mM KH_2PO_4, pH 7.4). The cells were pelleted by centrifugation at 10,000 × *g* for 20 min and resuspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA was then extracted with a Gnome DNA isolation kit (MP Biomedicals). Using this method, we obtained 4 μg of DNA from 40 g of soil, with an average size of approximately 35 kb.

The metagenomic library was constructed in the fosmid expression vector pMPO579 according to the CopyControl fosmid library production kit protocol (Epicentre Biotechnologies, WI, USA) with some modifications as described in the work of Terrón-González et al. (18).

Identification of clones conferring extradiol dioxygenase activity. The metagenomic library was transferred by triparental matings (19) to the MPO554 Nal^r strain, with DH5α/pRK2013 as the helper strain. Conjugative matings were performed on LB agar without antibiotic selection overnight at 37°C. The mating mixtures were then plated on LB agar with the necessary antibiotics for transconjugant selection (12.5 mg liter^{-1} chloramphenicol and 15 mg liter^{-1} nalidixic acid), 1 mM arabinose for increasing the copy number of the fosmid, and 1 mM salicylate for increasing the transcription levels. The plates were incubated at 30°C for 48 h. Then, the resultant colonies were screened for extradiol dioxygenase activity by spraying them with 1% (wt/vol) catechol in water or 0.5% 2,3-dihydroxybiphenyl (DHBP) in ethanol. Yellow coloration due to the formation of 2-hydroxymuconic semialdehyde or 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid, respectively, was observed in the positive clones. All of the positive clones were analyzed by restriction enzyme digestions.

Activity assays. Extradiol dioxygenase activity in resting cell assays using 2,3-dihydroxybiphenyl, catechol, 3-methylcatechol, 4-methylcatechol, and 4-chlorocatechol as the substrates was assayed by measuring the formation of the corresponding ring fission products in 50 mM NaK phosphate buffer (pH 6.8) (21). The extinction coefficients used for the ring fission products of the substrates were as follows: 2,3-dihydroxybiphenyl, λ_{max} = 434 nm, ε = 13.2 mM^{-1} cm^{-1}; catechol, λ_{max} = 375 nm, ε = 36 mM^{-1} cm^{-1}; 3-methylcatechol, λ_{max} = 388 nm, ε = 13.8 mM^{-1} cm^{-1}; 4-methylcatechol, λ_{max} = 382 nm, ε = 28.1 mM^{-1} cm^{-1}; 4-chlorocatechol, λ_{max} = 379 nm, ε = 39.6 mM^{-1} cm^{-1} (21, 22). The assays were performed using the selected positive single fosmid clones in the MPO554 strain. The bacteria were grown in the presence of 1 mM salicylate and 1 mM arabinose (to increase heterologous expression and fosmid copy number, respectively) at 30°C overnight The cells were harvested by centrifugation at 16,000 × *g* for 1 min at 4°C, washed and resuspended in 50 mM NaK phosphate buffer (pH 6.8) to an approximate optical density at 600 nm (OD_{600}) of 5, and used immediately. The reaction was performed using 100 μl of the prepared cell suspensions and initiated by adding the corresponding substrate: 5 μl of a 2,3-dihydroxybiphenyl solution in H_2O (10 mM), 5 μl of a catechol solution in ethanol (10 mM), 4 μl of a 3-methylcatechol solution in H_2O (10 mM), 8 μl of a 4-methylcatechol solution in H_2O (10 mM) and 10 μl of a 4-chlorocatechol solution in ethanol (10 mM). The final substrate concentrations in the reaction mixtures were 0.5 mM, 0.5 mM, 0.4 mM, 0.8 mM, and 1 mM, respectively. Reactions were monitored at room temperature in a POLARstar Omega multifunctional microplate reader (BMG Labtech GmbH, Germany) with a Costar 96 microplate at the indicated wavelength. Activity was calculated as the amount of product generated per unit of time referred to one unit of OD_{600}. All values were corrected against a blank nonenzymatic transformation. Finally, the activity of each

clone relative to the activity toward its most preferred substrate was calculated as a percentage. Three independent replicates of each clone were performed, and average values were calculated.

DNA sequencing and data analysis. The fosmids were sequenced with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Lifesequencing S.L. (Valencia, Spain), and the resultant sequences were assembled using a Newbler GS v.2.3 *de novo* assembler (Roche) (for details, see reference 23). Three independent pools were done and sequenced. Pool 1 contains 16 fosmids (UPO33 to UPO38 and UPO40 to UPO49), pool 2 contains 17 fosmids (UPO50, UPO51, UPO52, UPO54, UPO55, UPO57, UPO58, UPO60, UPO61, UPO64, UPO65, UPO66, UPO67, UPO68, UPO69, UPO71, and UPO72), and pool 3 contains 15 fosmids (UPO53, UPO74, UPO75, UPO76, UPO77, UPO78, UPO79, UPO80, UPO85, UPO86, UPO87, UPO88, UPO89, UPO90, and UPO91). For the preparation of libraries for next-generation sequencing, we pooled fosmids for each pool in a minicentrifuge tube in equimolar ratios. DNA fragmentation, library preparation, and sequencing were performed according to the manufacturer's instructions. To identify fosmids in the pool, we initially performed Sanger sequencing of the termini of each fosmid using the T7 (TAATACGACTCACTATAGGG) and pMPO579Rev (TCTCCTTTACTCATATGTATATCTCC) primers and aligned resulting reads with all contigs obtained for a pool using a local BLAST algorithm. In parallel, identity was also performed by PCR amplification using internal primers designed for each of the largest contigs. A total of 84,200 reads with a mean read length of 360 nucleotides that provides a total number of 47.46 Mb were obtained. After assembly, the contigs (assigned to each fosmid) containing genes encoding Edo proteins were identified and submitted to public databases. The length of the contigs per fosmid (as submitted to database) is shown in Table S1 in the supplemental material.

The assembled sequences were compared to those in the NCBI database using the BLASTx and BLASTn programs (24) and annotated according to their similarities. Open reading frames (ORFs) potentially coding for extradiol dioxygenase enzymes were assigned to clones by PCR and subsequently confirmed by standard Sanger sequencing default parameters. Multiple-sequence alignment was carried out using CLUSTALX software, and a distance matrix and a phylogenetic tree were constructed using the neighbor-joining method (25) and visualized with TreeView software.

Chemicals. All chemicals used for enzymatic tests were of analytical grade and were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO, USA). The kits that were applied in this study and their individual manufacturers have been noted in the text.

Nucleotide sequence accession numbers. The metagenomic sequences obtained from the fosmids are available under the EMBL/GenBank/DDBJ accession numbers KU144963 to KU145003.

RESULTS

Construction of a metagenomic fosmid library. The sample that was used for library construction was 40 g of soil from a refinery that had been contaminated with crude oil and that had been subjected to bioremediation through biostimulation for 6 months. Biomass was extracted by soil dispersion and low-speed centrifugation, and a fraction that was enriched in bacteria was isolated after high-speed Nycodenz density gradient centrifugation (26). A total of 4 μg of 30- to 40-kb-long DNA fragments was isolated from the bacterial fraction and was used to construct the library in a pMPO579 vector, a fosmid vector constructed to facilitate gene expression of metagenomic DNA in *E. coli* by transcription initiation from a salicylate-inducible heterologous promoter, followed by a *nut* (N utilization) site that allows transcript antitermination by the N antitermination protein from the lambda phage (18).

The constructed library, which consisted of 185,000 independent clones, was stored *en masse* in a single EPI300-T1 culture that was divided in small samples and kept frozen until use. Restriction analyses of randomly selected clones indicated that each of the clones contained different DNA inserts with an average size of 35 kb; therefore, the library harbored approximately 6.5 Gb of metagenomic DNA.

Functional screening of the metagenomic library. The metagenomic library was screened for Edo activity in the specialized strain MPO554 Nal^r, which produces the N antitermination protein, thus allowing heterologous gene transcription of the metagenomic DNA. To accomplish this, the library was transferred from EPI300-T1 into the specialized strain by conjugation in triparental matings. The conjugal transfer frequency of the fosmid clones was estimated to be 7 to 10% of the recipient cells.

For functional screening, the mating mixture was diluted to obtain 1×10^3 to 4×10^3 CFU of transconjugant cells per plate (15-cm diameter) in LB selective medium that also contained salicylate to induce heterologous transcription and arabinose to increase the fosmid copy number. The grown colonies were sprayed with either catechol or 2,3-dihydroxybiphenyl (2,3-DHBP) in an attempt to maximize the identification of Edos with different preferences for mono- or diaromatic substrates. Approximately 260,000 colonies distributed across 185 plates were sprayed with catechol, while a similar number of colonies distributed across 150 plates were sprayed with 2,3-DHBP. When using catechol as a substrate, 28 yellow colonies were identified and confirmed, which resulted in 17 different clones (designated UPO33 to UPO49) showing a distinct restriction fragment pattern after digestion with two restriction enzymes. On the other hand, screening using 2,3-DHBP detected 67 positive hits that resulted in up to 44 different fragment patterns after restriction enzyme digestion. Only 3 clones identified using 2,3-DHBP had a restriction pattern previously identified in the screening using catechol (UPO35, UPO42, and UPO48); therefore, the screening resulted in a total of 58 independent clones bearing Edo activity.

Testing the effect of salicylate-induced heterologous expression on the development of a yellow color in the already-identified clones showed that one-third of the clones still developed sufficient yellow color in the absence of salicylate and could therefore possibly be detected in a functional screening in the absence of heterologous expression. However, the remaining clones either barely developed yellow color or did not show color at all in the absence of salicylate. Therefore, the use of the heterologous expression system herein apparently increased the frequency of positive hits by at least 3-fold.

The low level of clone repetition among the identified fosmids (58 different restriction patterns among 95 positive hits) suggests that the functional screening was not saturating; presumably, additional independent positive clones would have appeared if more colonies had been screened.

Identification of Edo genes and assignment to clones. DNA from 48 clones (out of 58 initially selected) showing different restriction patterns was sent for next-generation sequencing. Automatic functional annotation of the contigs initially revealed 24 different genes encoding potential extradiol dioxygenases. However, 17 genes that were initially annotated as glyoxalase/bleomycin resistance protein/dioxygenase actually belonged to the extradiol dioxygenase family; therefore, the total number of Edo-encoding genes found in this functional screening was 41.

The Edo-encoding genes were ascribed to clones by PCR am-

plification of each clone in the pool, with one internal and one external primer designed for each *edo* gene identified in the pool and resequencing of the positive clones with the primers used for the PCR. All 41 genes were unambiguously ascribed to at least one clone, as shown in Table S2 in the supplemental material.

The most abundant Edo-encoding gene in the metagenomic library appeared to be $edoD_1$, which appeared in 4 clones, followed by $edoQ_1$ and $edoS_1$, each of which was ascribed to 3 clones, and $edoC_1$, $edoB_1$, $edoS_2$ and *edoW*, each of which was found in 2 clones. The UPO60 clone that codes for $edoQ_1$ has a genomic context very similar to but different from those of UPO37 and UPO89 because a 1.4-kb deletion and up to 22 point substitutions were found compared to the other clones; therefore, it likely arises from a different genome. The remaining genes were assigned to just one clone. Four of the clones contained two *edo* genes: $edoY_3$ and *edoP* in UPO45, $edoX_2$ and $edoB_3$ in UPO51, $edoY_1$ and $edoY_2$ in UPO64, and $edoX_1$ and $edoB_2$ in UPO67.

Pairwise comparisons of positive clones indicated that although $edoQ_2$, $edoQ_3$, $edoQ_4$, and $edoQ_5$ were different, there were actually very few differences among them, all of which resulted in synonymous codons; therefore, they all encoded the same protein, which was designated $EdoQ_2$. The protein groups $EdoB_2$-$EdoB_3$, $EdoC_1$-$EdoC_2$-$EdoC_5$, $EdoD_1$-$EdoD_2$, $EdoS_4$-$EdoS_5$, and $EdoX_1$-$EdoX_2$ possessed at least 98% identity with each other. The remaining Edos possessed 95% identity or less with any of the others.

Many Edos are involved in the biodegradation of aromatics and were at one point closely linked to other genes encoding biodegradation enzymes. This is the case for most of the *edo* genes identified here because, out of the 41 identified *edo* genes, only *edoG* and *edoV* appeared to be clearly unlinked to any other biodegradation genes. The *edoG* gene was linked to several transposases; however, intriguingly, *edoV* was flanked by amino acid metabolism genes that are common in bacterial genomes.

Phylogenetic analysis of metagenomic Edos. An initial BLAST analysis of each Edo enzyme against the UniProtKB/Swiss-Prot database revealed that the sequences of all but EdoW showed less than 60% identity with enzymes that have been identified in cultured bacteria, and 16 of them showed 40% or less identity to any characterized Edo (Table 1). The most divergent Edo was $EdoY_3$, which showed 31% identity with the most similar Edo sequence in the nonredundant database.

Extradiol dioxygenases are grouped as types I, II, and III, which have completely different evolutionary origins. Most of the metagenomic Edo sequences were clearly ascribed to type I, which comprises the oxygen vicinal chelate superfamily of extradiol dioxygenases. However, 5 of the sequences were unambiguously assigned to type II. No metagenomic Edo was ascribed as a type III ring cleavage dioxygenase belonging to the cupin superfamily.

Type I Edos can be highly divergent in sequence and are divided into different families (I.1 to I.6) that share less than 22% identity (1, 27). The I.1 family is composed of short, single-domain enzymes, while the other families are composed of large two-domain dioxygenases. All of the identified type I Edos were two-domain enzymes and were ascribed to families I.2 to I.5.

Edos belonging to families I.4, I.5, and I.6 are less-abundant two-domain enzymes and have not been divided into subfamilies in previous studies. We identified a total of 6 sequences that can be ascribed to families I.4 and I.5 (Fig. 1). Four of these sequences were divergent enough to enable the definition of I.4.B and I.5.B subfamilies because they shared less than 54% identity with the most similar sequence in the families (1). $EdoB_1$, $EdoB_2$, and $EdoB_3$, together with other uncharacterized enzymes in the database, defined subfamily I.4.B, while EdoF, together with other uncharacterized proteins in the database, defined subfamily I.5.B.

Ten metagenomic Edos were ascribed to the I.2 family of Edos, which are mainly involved in the cleavage of single aromatic ring molecules (Fig. 2). For simplicity, subfamilies to which no metagenomic Edo was ascribed are represented by just one sequence. Most of the sequences were sufficiently similar to be ascribed to any of the 7 previously defined subfamilies, but EdoH was different enough to define, together with other sequences in the nonredundant database, an eighth subfamily, which was designated I.2.H.

A total of 17 Edos were ascribed to the I.3 family, which is composed of enzymes with high activities toward molecules with more than one aromatic ring (Fig. 3). This family was previously divided into 14 subfamilies, but only 2 of them, I.3A and I.3.E, harbored metagenomic Edos (EdoW and EdoV, respectively). The remaining subfamilies are therefore represented for simplicity by a single sequence from the database in Fig. 3. Fifteen out of the 17 Edo sequences belonging to this family were so different that they could not be ascribed to previously described subfamilies and led to the definition of a total of 7 new subfamilies, designated I.3.O to I.3.U.

Type II Edos are less abundant, but they can be more diverse than type I Edos in relation to subunit composition, which can be homo- or heteromultimeric. Based on their composition and phylogenetic relationships, type II Edos can be separated into at least 7 families (Fig. 4); our sequences were related to two of these families.

The II.1 family comprises a group of heteromultimeric enzymes that are composed of a large β subunit of approximately 270 to 280 residues and a small, unrelated, α subunit, these subunits being encoded by two contiguous genes. Sequences in Fig. 4 correspond to the large β subunit. $EdoY_1$ and $EdoY_2$ are also apparently encoded by a pair of contiguous genes for α and β subunits, and sequences of their β subunit are similar enough to ascribe them to this family (≥49% identity). Similar genomic contexts coding for heteromultimeric Edos were found for CarB and $EdoY_3$ enzymes. In these cases, the sequences of their respective β subunits are much more divergent, although they could also be ascribed to the same family since they show 29 to 33% identity with the II.1 sequences.

The II.2 family is also composed of heteromultimeric Edos with similar composition but with slightly larger β subunits (274 to 302 residues) that are substantially divergent from those of the II.1 family. To this family belongs the well-characterized enzyme LigB (28). Interestingly, PcmA, GalA, and DesB are even larger homomeric proteins that are actually composed of the α and β domains fused in the same polypeptide (6, 29, 30).

A third group of heteromultimeric proteins defining the II.3 family comprises Amn and CnbC Edos that are composed of a large, active β subunit (305 to 314 residues) and an α subunit much larger than those shown by the enzymes in the II.1 and II.2 families, which actually shows similarity to its β subunit counterpart. Both α and β subunits are shown in Fig. 5.

HpaD is an Edo sequence of 283 residues encoded by a single gene in a very compact *hpaGEDFHI* operon that lacks sequences resembling an α subunit (31). Thus, there is no indication that

TABLE 1 List of identified metagenomic Edos[a]

Edo name	Edo classification	Length (aa)	Most similar protein	Host organism	Bacterial division (class) or other	Amino acid identity (%)
$EdoA_1$	I.2.A	307	Catechol 2,3-dioxygenase	*Pseudomonas* sp. strain CF600	*Gammaproteobacteria*	57
$EdoA_2$	I.2.A	307	Catechol 2,3-dioxygenase	*Pseudomonas* sp. strain CF600	*Gammaproteobacteria*	57
$EdoA_3$	I.2.A	307	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	57
$EdoB_1$	I.4.B	306	Catechol 2,3-dioxygenase	*Cupriavidus necator*	*Betaproteobacteria*	28
$EdoB_2$	I.4.B	306	Catechol 2,3-dioxygenase	*Cupriavidus necator*	*Betaproteobacteria*	28
$EdoB_3$	I.4.B	306	Catechol 2,3-dioxygenase	*Cupriavidus necator*	*Betaproteobacteria*	28
$EdoC_1$	I.2.C	316	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	39
$EdoC_2$	I.2.C	313	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	38
$EdoC_3$	I.2.C	314	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	40
$EdoC_4$	I.2.C	314	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	45
$EdoC_5$	I.2.C	316	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	39
$EdoD_1$	I.4.A	306	Catechol 2,3-dioxygenase	*Cupriavidus necator*	*Betaproteobacteria*	58
$EdoD_2$	I.4.A	306	Catechol 2,3-dioxygenase	*Cupriavidus necator*	*Betaproteobacteria*	59
EdoF	I.5.B	303	Catechol 2,3-dioxygenase	*Rhodococcus rhodochrous*	*Actinobacteria*	30
EdoG	I.2.B	290	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	48
EdoH	I.2.H	310	Catechol 2,3-dioxygenase	*Pseudomonas putida*	*Gammaproteobacteria*	46
$EdoO_1$	I.3.O	289	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	39
$EdoO_2$	I.3.O	290	2,3-Dihydroxybiphenyl dioxygenase	*Burkholderia xenovorans* LB400	*Betaproteobacteria*	40
$EdoO_3$	I.3.O	289	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida* F1	*Gammaproteobacteria*	39
$EdoO_4$	I.3.O	289	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida* F1	*Gammaproteobacteria*	40
EdoP	I.3.P	294	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	51
$EdoQ_1$	I.3.Q	291	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	51
$EdoQ_2$	I.3.Q	291	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	48
EdoR	I.3.R	293	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	52
$EdoS_1$	I.3.S	300	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	53
$EdoS_2$	I.3.S	300	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	54
$EdoS_3$	I.3.S	300	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	54
$EdoS_4$	I.3.S	300	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	52
$EdoS_5$	I.3.S	300	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	52
EdoT	I.3.T	303	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	46
EdoU	I.3.U	329	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas pseudoalcaligenes*	*Gammaproteobacteria*	33
EdoV	I.3.E	310	2,3-Dihydroxybiphenyl dioxygenase	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	58
EdoW	I.3.A	293	2,3-Dihydroxybiphenyl dioxygenase	*Pseudomonas* sp. strain KKS102	*Gammaproteobacteria*	87
$EdoX_1$	II	274	Extradiol ring cleavage dioxygenase	*Arabidopsis thaliana*	*Eukaryota*, plant	45
$EdoX_2$	II	265[b]	Extradiol ring cleavage dioxygenase	*Arabidopsis thaliana*	*Eukaryota*, plant	45
$EdoY_1$	II	276	2,3-Dihydroxyphenyl-propionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase	*Photorhabdus luminescens* subsp. *laumondii* TTO1	*Gammaproteobacteria*	26
$EdoY_2$	II	276	2,3-Dihydroxyphenyl propionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase	*Photorhabdus luminescens* subsp. *laumondii* TTO1	*Gammaproteobacteria*	25
$EdoY_3$	II	278	Protocatechuate 4,5-dioxygenase beta chain	*Sphingomonas paucimobilis*	*Alphaproteobacteria*	31

[a] Under Edo classification, the type and, where appropriate, the family and subfamily of each Edo are indicated. The most similar sequences in the UniProt/Swiss-Prot databases and their percent identity are also shown. The $edoQ_2$, $edoQ_3$, $edoQ_4$, and $edoQ_5$ genes code for the same protein, designated $EdoQ_2$.
[b] Partial sequence; the $edoX_2$ gene is not complete in the metagenomic DNA insert.

HpaD could be a heteromultimer. This is in contrast to HpcB (32), a homomultimeric protein whose β subunit, of 276 residues, is 98% identical to HpaD. Although HpaD shows 15 to 26% identity to Edos of the II.3 family, given their subunit composition and shorter length, we propose to ascribe them to a new II.4 family.

The II.5 family comprises well-characterized homomultimeric enzymes of approximately 313 to 321 residues, such as the MhpB proteins of different bacteria.

The II.6 family is constituted by a group of smaller (255 to 274 residues), apparently homomeric proteins because contiguous genes coding for α subunits are not found. This group, to which $EdoX_1$ and $EdoX_2$ belong, contains the Edo enzymes Diox PORGR and Diox ARATH, which are of plant origin and act as extradiol dioxygenases in biosynthetic reactions (33, 34). The most divergent enzyme, YgiD, is a monomeric 4,5-dihydroxyphenylalanine (DOPA)-extradiol dioxygenase from *E. coli* (35).

Finally, the II.7 family comprises the large homomeric Edos LigZ (334 amino acids [aa]) and DesZ (330 aa) involved in ring opening of different lignin-derived products (36, 37). Unlike the largest members of the II.2 family, these sequences do not contain α subunit domains.

Extradiol dioxygenase activities of clones using different substrates. The activities of the 58 selected clones encoding Edo enzymes toward 2,3-DHBP, catechol, and methyl- or chloro-substituted catechols were tested to establish their substrate ranges and preferences. To accomplish this, the selected clones were grown in the presence of salicylate and arabinose to ensure the expression of the corresponding genes, and their activities were measured in resting cell assays.

Because absolute levels of Edo activity may substantially vary from one clone to another, Fig. 5 represents the activity of each clone relative to the activity toward its most preferred substrate, to best represent both the substrate range and preference of each clone. 3-Methylcatechol was the most preferred substrate because 39% of the clones (17 of 44) showed their highest activity toward it, followed by 4-methylcatechol, for which 10

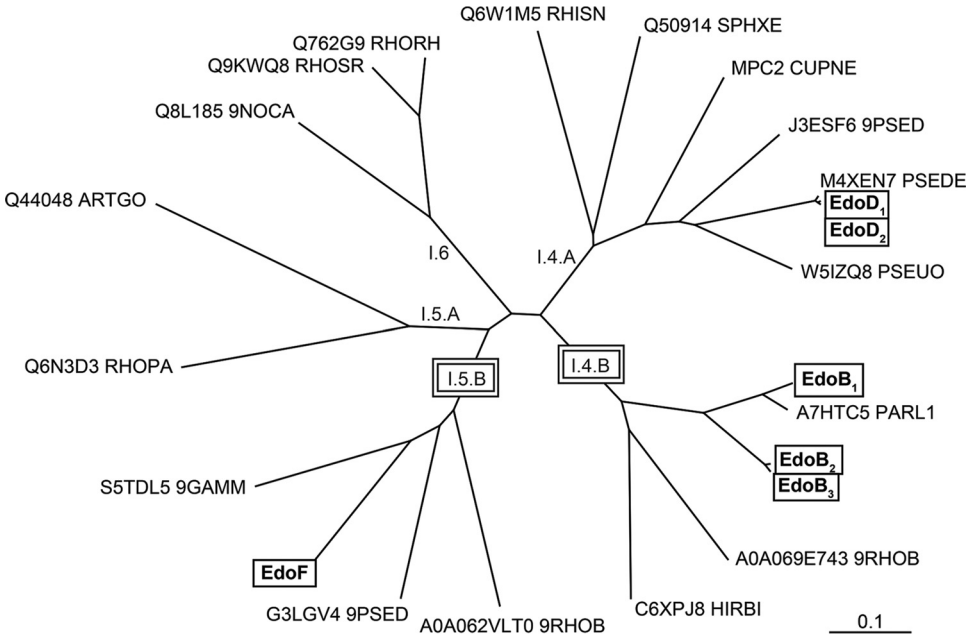

FIG 1 Phylogenetic tree of metagenomic Edos and previously sequenced type I subfamily I.4, I.5, and I.6 Edos. The scale indicates the number of amino acid substitutions per position. The metagenomic Edos identified in this study are in boxes, and the new subfamilies are in double boxes. The previously identified Edos included in the phylogenetic tree and their UniProt accession numbers are as follows: MPC2 CUPNE, metapyrocatechase 2, *Cupriavidus necator* (P17296); Q6W1M5 RHISN, metapyrocatechase, *Rhizobium* sp. (Q6W1M5); Q50914 SPHXE, 2,3-dihydroxybiphenyl dioxygenase, *Sphingobium xenophagum* (Q50914); Q9KWQ8 RHOSR, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus* sp. (Q9KWQ8); Q8L185 9NOCA, extradiol dioxygenase, *Rhodococcus* sp. (Q8L185); Q6N3D3 RHOPA, putative catechol 2,3-dioxygenase, *Rhodopseudomonas palustris* (Q6N3D3); Q44048 ARTGO, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, *Arthrobacter globiformis* (Q44048); Q762G9 RHORH, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus rhodochrous* (Q762G9); M4XEN7 PSEDE, MhqB protein, *Pseudomonas denitrificans* (M4XEN7); W5IZQ8 PSEUO, metapyrocatechase 2, *Pseudomonas* sp. (W5IZQ8); J3ESF6 9PSED, glyoxalase/bleomycin resistance protein/dioxygenase superfamily, *Pseudomonas* sp. (J3ESF6); A7HTC5 PARL1, glyoxalase/bleomycin resistance protein/dioxygenase, *Parvibaculum lavamentivorans* (A7HTC5); A0A069E743 9RHOB, glyoxalase/bleomycin resistance protein/dioxygenase, *Hyphomonas adhaerens* (A0A069E743); C6XPJ8 HIRBI, glyoxalase/bleomycin resistance protein/dioxygenase, *Hirschia baltica* (C6XPJ8); S5TDL5 9GAMM, extradiol dioxygenase of the vicinal chelate superfamily (2,3-dihydroxy-*p*-cumate dioxygenase), *Cycloclasticus zancles* (S5TDL5); G3LGV4 9PSED, CmtC, *Pseudomonas* sp. (G3LGV4); A0A062VLT0 9RHOB, glyoxalase/bleomycin resistance protein/dioxygenase, *Hyphomonas polymorpha* (A0A062VLT0).

clones showed the highest preference. Finally, 2,3-DHBP and catechol, the substrates used in the functional screening, were the most preferred substrates in only 9 and 7 clones, respectively. On the other hand, none of the clones showed their highest activities toward 4-chlorocatechol. In fact, this was one of the least preferred substrates for 23 of the clones, most of which showed less than 5% of the activity that was shown toward their most preferred substrates, and 5 clones showed a relative activity between 5 and 10%. Therefore, 4-chlorocatechol was the least favorable substrate. This is in full agreement with previously described Edo activities and with the assertion that chloroaromatics are most commonly degraded through a modified ortho-cleavage pathway (38).

A number of clones, such as UPO51 (containing $EdoX_2$/$EdoB_3$), showed quite narrow substrate preference profiles, showing strong preferences for only one of the 5 tested substrates and 30% activity or less with any other. In contrast, other clones showed a much broader preference spectrum; for example, UPO80 (containing $EdoS_1$) exhibited on its least preferred substrate more than 68% of its activity on its most preferred substrate. The best substrates for this enzyme appeared to be the 4-substituted monoaromatics. Of all the clones, UPO80 was the best at using 4-chlorocatechol, showing a relative activity of 96% when using this substrate.

DISCUSSION

In this study, a metagenomic functional analysis was performed to estimate the relative abundance and diversity of extradiol dioxygenases, which are key enzymes in the biodegradation of aromatic contaminants. We further compared them to those obtained from other contaminated sites.

To maximize the identification of positive hits, a vector enabling the heterologous expression of metagenomic DNA in an *E. coli* surrogate host was used. Testing the color development of the already-identified clones under both conditions indicated that most of them would have been missed if the heterologous expression system had not been induced during screening. Additionally, 64% of the sequenced clones possessed an *edo* gene that was read in the same direction as the salicylate-inducible promoter. Therefore, there is a clear bias toward the orientation promoting heterologous transcription of the *edo* genes. These two results empirically support the idea that the expression of metagenomic genes in a surrogate host harboring a library is a major limiting factor in identifying positive hits during functional screening (14, 18, 39).

A second factor preventing the identification of additional positive hits is the sensitivity and efficiency of a screening (18). The screening using 2,3-DHBP as an indicator resulted in 2.5-fold-more positive clones than the screening on the same number of colonies using catechol (44 versus 17). This resulted in the iden-

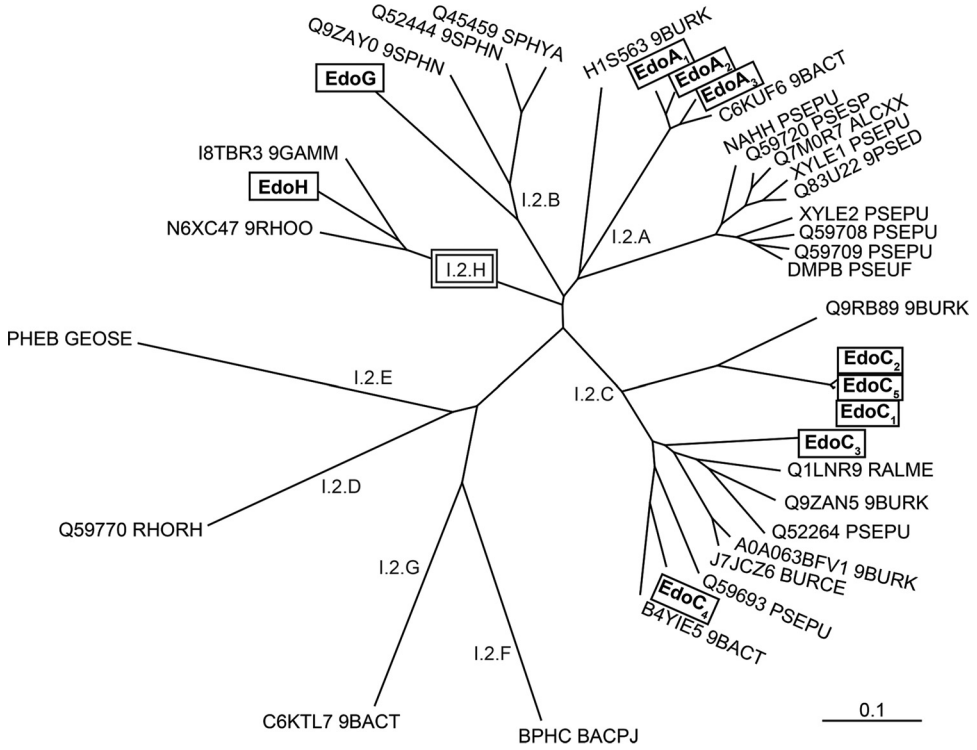

FIG 2 Phylogenetic tree of metagenomic Edos and previously sequenced type I subfamily I.2 Edos. The scale indicates the number of amino acid substitutions per position. The metagenomic Edos identified in this study are in boxes, and the new subfamilies are in double boxes. The previously identified Edos included in the phylogenetic tree and their UniProt accession numbers are as follows: BPHC BACPJ, manganese-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Bacillus* sp. (Q8GR45); Q59770 RHORH, catechol 2,3-dioxygenase, *Rhodococcus rhodochrous* (Q59770); PHEB GEOSE, metapyrocatechase, *Geobacillus stearothermophilus* (P31003); Q59693 PSEPU, catechol 2,3-dioxygenase II, *Pseudomonas putida* (Q59693); Q9ZAN5 9BURK, catechol 2,3-dioxygenase, *Comamonas* sp. (Q9ZAN5); Q52264 PSEPU, 3-methylcatechol 2,3-dioxygenase, *Pseudomonas putida* (Q52264); Q52444 9SPHN, catechol 2,3-dioxygenase, *Sphingomonas agrestis* (Q52444); Q45459 SPHYA, catechol 2,3-dioxygenase, *Sphingobium yanoikuyae* (Q45459); XYLE2 PSEPU, metapyrocatechase, *Pseudomonas putida* (Q04285); Q59708 PSEPU, catechol 2,3-dioxygenase, *Pseudomonas putida* (Q59708); DMPB PSEUF, metapyrocatechase, *Pseudomonas* sp. (P17262); Q59709 PSEPU, catechol 2,3-dioxygenase, *Pseudomonas putida* (Q59709); NAHH PSEPU, metapyrocatechase, *Pseudomonas putida* (P08127); Q59720 PSESP, catechol 2,3-dioxygenase, *Pseudomonas* sp. (Q59720); Q7M0R7 ALCXX, catechol 2,3-dioxygenase, *Alcaligenes xylosoxidans* (Q7M0R7); XYLE1 PSEPU, metapyrocatechase, *Pseudomonas putida* (P06622); Q83U22 9PSED, catechol 2,3-dioxygenase, *Pseudomonas* sp. (Q83U22); C6KUF6 9BACT, catechol 2,3-dioxygenase, uncultured bacterium CLON2H2 (C6KUF6); C6KTL7 9BACT, catechol 2,3-dioxygenase, uncultured bacterium CLON1H11 (C6KTL7); Q9ZAY0 9SPHN, catechol 2,3-dioxygenase, *Sphingomonas* sp. (Q9ZAY0); H1S563 9BURK, catechol 2,3-dioxygenase, *Cupriavidus basilensis* (H1S563); I8TBR3 9GAMM, catechol 2,3-dioxygenase, *Hydrocarboniphaga effusa* (I8TBR3); N6XC47 9RHOO, catechol 2,3-dioxygenase, *Thauera* sp. (N6XC47); Q9RB89 9BURK, catechol 2,3-dioxygenase, *Burkholderia sartisoli* (Q9RB89); A0A063BFV1 9BURK, catechol 2,3-dioxygenase, *Burkholderia* sp. (A0A063BFV1); J7JCZ6 BURCE, catechol 2,3-dioxygenase, *Burkholderia cepacia* (J7JCZ6); Q1LNR9 RALME, metapyrocatechase (catechol 2,3-dioxygenase), *Ralstonia metallidurans* (Q1LNR9); B4YIE5 9BACT, catechol 2,3-dioxygenase, uncultured bacterium (B4YIE5).

tification of only 16 *edo* genes using catechol, compared to the 27 genes that were identified using 2,3-DHBP (6 genes were identified with both substrates [Table 1]), which suggests that many of the enzymes in the current sample are not sufficiently active toward catechol to enable detection; therefore, although catechol has been the most common substrate used to detect Edo activity (15, 16, 40, 41), it should not be the indicator of choice for its functional detection.

The distribution of *edo* genes among the clones was quite even; the most abundant gene was $edoD_1$, which was found in just 4 of the sequenced clones. This clearly indicates that the population of *edo*-containing bacterial genomes that was identified by this functional screening was not dominated by one or a few predominant bacteria that were particularly abundant in the soil sample but rather had an even distribution. Intriguingly, this is in contrast to what was found in previous, more extensive metagenomic analyses that identified *edo* genes from a coke plant wastewater treatment sample (15) and from soil contaminated with jet fuel (16), both of which showed more extensive repetitions of a number of highly abundant Edo sequences among the identified clones. In the case of the coke plant sample, 20 of the 38 identified sequences were actually repetitions of a few highly similar variant sequences of an Mn^{2+}-dependent Edo (42) encoded by a plasmid that was designated pSKYE1 (43). Although the coke plant wastewater had a mixture of several aromatic compounds that could be used as carbon sources (15), it was highly enriched in phenol, and the high prevalence of Edo enzymes may be explained by their proposed roles in phenol detoxification (41). In the case of the contaminated soil, the *edo* sequences were initially amplified from positive clones with selected primers, which may impose a bias toward selecting highly similar sequences.

Previous studies have suggested that substrate preferences may be different for Edos belonging to different phylogenetic ascriptions. Thus, in general, Edos of the I.2 family prefer monoaromatic substrates, with the possible exception of members of the subfamily I.2.G, which show almost equal activities for catechol

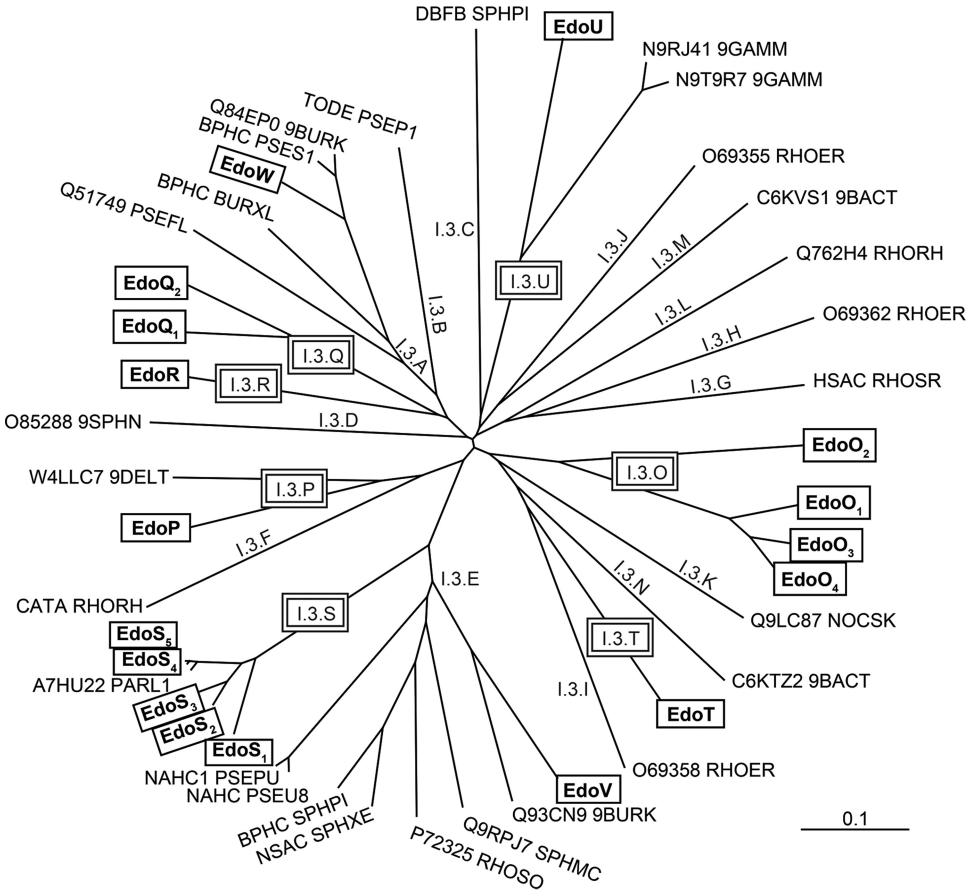

FIG 3 Phylogenetic tree of metagenomic Edos and previously sequenced type I subfamily I.3 Edos. The scale indicates the number of amino acid substitutions per position. The metagenomic Edos identified in this study are in boxes, and the new subfamilies are in double boxes. The previously identified Edos included in the phylogenetic tree and their UniProt accession numbers are as follows: Q9RPJ7 SPHMC, 1,2-dihydroxynaphthalene dioxygenase ThnC, *Sphingopyxis macrogoltabida* (Q9RPJ7); O85288 9SPHN, extradiol dioxygenase, *Sphingomonas* sp. (O85288); P72325 RHOSO, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus* sp. (P72325); NSAC SPHXE, 1,2-dihydroxynaphthalene dioxygenase, *Sphingobium xenophagum* (P74836); NAHC PSEU8, 1,2-dihydroxynaphthalene dioxygenase, *Pseudomonas* sp. (P0A108); NAHC1 PSEPU, 1,2-dihydroxynaphthalene dioxygenase, *Pseudomonas putida* (P11861); BPHC BURXL, biphenyl-2,3-diol 1,2-dioxygenase, *Burkholderia xenovorans* (P47228); TODE PSEP1, 3-methylcatechol 2,3-dioxygenase, *Pseudomonas putida* (P13453); BPHC PSES1, biphenyl-2,3-diol 1,2-dioxygenase, *Pseudomonas* sp. (P17297); Q51749 PSEFL, 3-isopropylcatechol dioxygenase, *Pseudomonas fluorescens* (Q51749); CATA RHORH, metapyrocatechase, *Rhodococcus rhodochrous* (Q53034); O69355 RHOER, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69355); Q762H4 RHORH, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus rhodochrous* (Q762H4); O69362 RHOER, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69362); HSAC RHOSR, iron-dependent extradiol dioxygenase, *Rhodococcus* sp. (Q9KWQ5); Q9LC87 NOCSK, extradiol dioxygenase, *Nocardioides* sp. (Q9LC87); O69358 RHOER, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Rhodococcus erythropolis* (O69358); Q84EP0 9BURK, 2,3-dihydroxybiphenyl dioxygenase, *Cupriavidus oxalaticus* (Q84EP0); Q93CN9 9BURK, extradiol dioxygenase DbtC, *Burkholderia* sp. (Q93CN9); DBFB SPHPI, 2,2′,3-trihydroxybiphenyl dioxygenase, *Sphingomonas paucimobilis* (P47243); C6KVS1 9BACT, 2,3-dihydroxybiphenyl 1,2-dioxygenase, uncultured bacterium CLON7E11 (C6KVS1); C6KTZ2 9BACT, 2,3-dihydroxybiphenyl 1,2-dioxygenase, uncultured bacterium CLON1F2 (C6KTZ2); BPHC SPHPI, biphenyl-2,3-diol 1,2-dioxygenase, *Sphingomonas paucimobilis* (P11122); A7HU22 PARL1, glyoxalase/bleomycin resistance protein/dioxygenase, *Parvibaculum lavamentivorans* (A7HU22); W4LLC7 9DELT, uncharacterized protein, "*Candidatus* Entotheonella" sp. (W4LLC7); N9RJ41 9GAMM, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Acinetobacter* sp. (N9RJ41); N9T9R7 9GAMM, 2,3-dihydroxybiphenyl 1,2-dioxygenase, *Acinetobacter* sp. (N9T9R7).

and 2,3-DHBP (15). Conversely, Edos of family I.3 prefer polycyclic molecules (1, 2). All of the I.2 family Edos that were characterized here showed preferences for monocyclic substrates (catechol or substituted catechols) and, more strikingly, showed very low activities against 2,3-DHBP, with the exception of $EdoC_3$. Therefore, these results support the assertion that Edos of the I.2 family cannot efficiently transform large polycyclic substrates. However, although members of the I.3 family showed high relative activities toward 2,3-DHBP (at least 29% of its maximal activity), this was not the preferred substrate for most of them. Actually, 13 out of the 16 Edos of the I.3 family showed maximal activities with monocyclic substrates; 11 of them preferred methyl-substituted catechols. Therefore, family I.3 should not simply be considered a family of Edos for polycyclic substrates, even though some characterized family I.3 enzymes, such as those from *Burkholderia xenovorans* LB400 (subfamily I.3.A) or *Rhodococcus globerulus* P6 (subfamily I.3.B), have shown a very strong preference for 2,3-DHBP (44, 45); rather, many of them are Edos with broad substrate specificities that allow for the efficient transformation of 2,3-DHBP.

The high abundance and diversity of Edos belonging to the I.3 family that were found in this study are in contrast to what has been found in previous metagenomic functional screenings (15, 16), in which family I.3 was less well represented. This difference

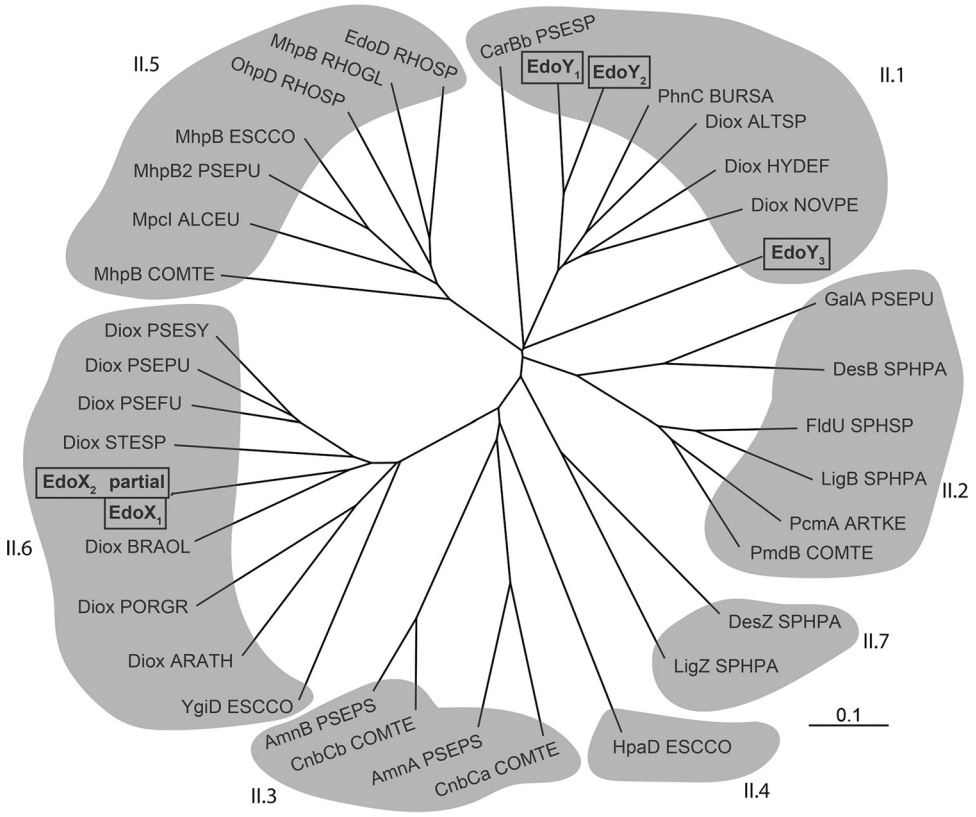

FIG 4 Phylogenetic tree of metagenomic Edos and previously sequenced type II Edos. The scale indicates the number of amino acid substitutions per position. The metagenomic Edos identified in this study are in boxes. The $EdoX_2$ sequence was partial and has 9 amino acids missing from its C terminus compared to that of $EdoX_1$. The previously identified Edos included in the phylogenetic tree and their UniProt accession numbers are as follows: PhnC BURSA, extradiol dioxygenase, *Burkholderia sartisoli* (Q9ZHH5); Diox ALTSP, protocatechuate 4,5-dioxygenase subunit beta, *Alteromonas* sp. strain SN2 (F5Z5S4); Diox HYDEF, uncharacterized protein, *Hydrocarboniphaga effusa* AP103 (I7ZHF5); Diox NOVPE, protocatechuate 3,4-dioxygenase, *Novosphingobium pentaromativorans* US6-1 (G6EGK0); MhpB2 PSEPU, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase 2, *Pseudomonas putida* (Q9F9U5); OhpD RHOSP, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Rhodococcus* sp. (Q9KH19); MhpB RHOGL, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Rhodococcus globerulus* (O05146); MhpB COMTE, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Comamonas testosteroni* (Q9S157); Diox PSEFU, extradiol ring cleavage dioxygenase protein subunit B, *Pseudomonas fulva* 12-X (F6AH36); Diox PSEPU, extradiol ring cleavage dioxygenase subunit B, *Pseudomonas putida* HB3267 (L0FGE4); Diox PSESY, subunit of aromatic ring-opening dioxygenase, *Pseudomonas syringae* BRIP34881 (L7G5A4); YgiD ESCCO, uncharacterized protein/DODA-type extradiol aromatic ring-opening dioxygenase family, *Escherichia coli* K-12 (P24197); Diox ARATH, extradiol ring cleavage dioxygenase, *Arabidopsis thaliana* (Q949R4); Diox PORGR, 4,5-DOPA dioxygenase extradiol, DODA, *Portulaca grandiflora* (Q7XA48); CnbCa COMTE, 2-aminophenol 1,6-dioxygenase subunit alpha, *Comamonas testosteroni* (Q6J1Z5); CnbCb COMTE, 2-aminophenol 1,6-dioxygenase subunit beta, *Comamonas testosteroni* (Q6J1Z6); AmnA PSEPS, 2-aminophenol-1,6-dioxygenase alpha subunit, AmnA, *Pseudomonas pseudoalcaligenes* (O34137); AmnB PSEPS, 2-aminophenol-1,6-dioxygenase beta subunit, *Pseudomonas pseudoalcaligenes* (O24680); CarBb PSESP, catalytic subunit of meta-cleavage enzyme, *Pseudomonas* sp. (O32474); DesZ SPHPA, 3-*O*-methylgallate 3,4-dioxygenase, *Sphingomonas paucimobilis* (Q7WYU8); EdoD RHOSP, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Rhodococcus* sp. strain I1 (O69791); FldU SPHSP, FldU protein, *Sphingomonas* sp. strain LB126 (Q9L3A6); GalA PSEPU, gallate dioxygenase, *Pseudomonas putida* KT2440 (Q88JX5 plus 80 aa at N terminus: the ORF for GalA apparently starts 240 nucleotides upstream of the annotated start codon and thus codes for a protein with 80 additional residues in its N terminus compared to that of Q88JX5); HpaD ESCCO, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, *Escherichia coli* (Q46980); MhpB ESCCO, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Escherichia coli* K-12 (P0ABR9); LigB SPHPA, beta subunit of protocatechuate 4,5-dioxygenase, *Sphingomonas paucimobilis* (Q7DH56); LigZ SPHPA, biphenyl ring cleavage enzyme, *Sphingomonas paucimobilis* (O82834); MpcI ALCEU, 2,3-dihydroxyphenylpropionate/2,3-dihydroxycinnamic acid 1,2-dioxygenase, *Alcaligenes eutrophus* (P17295); PcmA ARTKE, protocatechuate 4,5-dioxygenase, *Arthrobacter keyseri* (Q9AGL8); PmdB COMTE, protocatechuate 4,5-dioxygenase beta subunit, *Comamonas testosteroni* (Q8RNX9); Diox BRAOL, extradiol ring cleavage dioxygenase, *Bradyrhizobium oligotrophicum* S58 (M4Z500); Diox STESP, extradiol dioxygenase subunit, *Stenotrophomonas* sp. strain G205 (A0A0A1GK74); DesB SPHPA, gallate dioxygenase, *Sphingomonas paucimobilis* (Q5NTE5).

might be due to real differences between the populations of Edos from different sites. However, when analyzing the substrate preferences of the clones for catechol versus 2,3-DHBP, a strong preference for catechol or methyl-substituted derivatives and very low activity toward 2,3-DHB was found among those initially identified by screening with catechol (Fig. 5, clones UPO33 to UPO49). As a consequence, screening with catechol, the substrate indicator used in previous analyses, resulted in a strong biased identification of Edos in favor of those belonging to family I.2. On the other hand, the clones that were detected using 2,3-DHBP (UPO50 to UPO91) led to the identification of most of the Edos belonging to families I.3, I.4, and I.5 but only 2 of the 10 Edos of family I.2. These results clearly show the great bias that the substrate indicator used during screening may impose on these functional screenings. In any case and as a general rule, the best option is probably to use a mixture of as many substrates as possible to maximize the

Clone	EDO	Family / Subfamily	Screening	Relative activity				
				DHBP	CAT	3MCAT	4MCAT	4ClCAT
UPO33	$EdoC_5$	I.2.C	CAT		•			
UPO35[1]	$EdoC_2$	I.2.C	CAT, DHBP		•			
UPO36	$EdoA_1$	I.2.A	CAT				•	
UPO38	$EdoC_4$	I.2.C	CAT		•			
UPO39	N.S.	N.S.	CAT			•		
UPO41	EdoH	I.2.H	CAT				•	
UPO42[1]	EdoV	I.3.E	CAT, DHBP			•		
UPO43	$EdoA_3$	I.2.A	CAT				•	
UPO45	$EdoY_3$/P	II.1/I.3.P	CAT	•				
UPO46	$EdoC_3$	I.2.C	CAT				•	
UPO47	$EdoA_2$	I.2.A	CAT				•	
UPO48[1]	$EdoO_4$	I.3.O	CAT, DHBP			•		
UPO49	$EdoC_1$[2]	I.2.C	CAT		•			
UPO51	$EdoX_2$/B_3	II.6/I.4.B	DHBP	•				
UPO53	EdoU	I.3.U	DHBP	•				
UPO54	$EdoO_2$	I.3.O	DHBP			•		
UPO55	$EdoD_1$	I.4.A	DHBP			•		
UPO56	N.S.	N.S.	DHBP	•				
UPO57	EdoG	I.2.B	DHBP		•			
UPO59	N.S.	N.S.	DHBP		•			
UPO61	EdoR	I.3.R	DHBP		•			
UPO62	N.S.	N.S.	DHBP			•		
UPO63	N.S.	N.S.	DHBP			•		
UPO64	$EdoY_1$/Y_2	II.1	DHBP			•		
UPO67	$EdoX_1$/B_2	II.6/I.4.B	DHBP			•		
UPO68	$EdoB_1$	I.4.B	DHBP			•		
UPO70	N.S.	N.S.	DHBP			•		
UPO71	EdoF	I.5.B	DHBP	•				
UPO74	$EdoQ_2$	I.3.Q	DHBP			•		
UPO75	$EdoO_3$	I.3.O	DHBP			•		
UPO76	EdoT	I.3.T	DHBP		•			
UPO77	$EdoS_4$	I.3.S	DHBP				•	
UPO78	$EdoD_2$	I.4.A	DHBP			•		
UPO80	$EdoS_1$	I.3.S	DHBP				•	
UPO81	N.S.	N.S.	DHBP			•		
UPO82	N.S.	N.S.	DHBP			•		
UPO83	N.S.	N.S.	DHBP	•				
UPO84	N.S.	N.S.	DHBP	•				
UPO85	$EdoS_2$[2]	I.3.S	DHBP				•	
UPO86	$EdoS_5$	I.3.S	DHBP				•	
UPO87	$EdoB_1$	I.4.B	DHBP	•				
UPO89	$EdoQ_1$[2]	I.3.Q	DHBP	•				
UPO90	$EdoO_1$	I.3.O	DHBP			•		
UPO91	$EdoS_3$	I.3.S	DHBP				•	

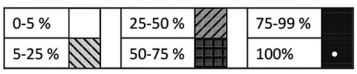

FIG 5 Relative activities of clones bearing *edo* genes. The clone names, the enzyme names, their phylogenetic ascriptions, and the indicator substrates used in the screenings where the positive clones were detected are shown along with their relative activities toward 2,3-dihydroxybiphenyl (DHBP), catechol (CAT), 3-methylcatechol (3MCAT), 4-methylcatechol (4MCAT), and 4-chlorocatechol (4ClCAT). A total of 58 clones were assayed, but those encoding the same Edos had similar profiles, and only one representative is shown, with the exception of UPO68 and UPO87, which bear $EdoB_1$ but showed a very different pattern of preference. Footnote 1, clones UPO35, UPO42, and UPO48 were identified in both screenings. Footnote 2, $EdoC_1$ was also identified with DHBP in UPO50. $EdoQ_1$ and $EdoS_2$ were also identified with catechol in UPO37 and UPO44, respectively. N.S., not sequenced.

detection of enzymatic activities with different substrate preferences.

This analysis also identified 5 type II Edos belonging to 2 different families. Compared to type I Edos, type II Edos have been less well characterized, have fewer available sequences, and have not been previously identified in metagenomic functional analysis. This has been attributed to their presumed scarcity in the environment (41). However, the present analysis revealed that most of the clones encoding type II Edos (UPO51, UPO64, and UPO67) were detected using 2,3-DHB and had very low relative activities with catechol. UPO45, the only positive clone detected using catechol, actually codes for an additional Edo of the I.3.P subfamily; therefore, its high activity against catechol might be due to the additional encoded enzyme. These results combined with the substrates utilized by the type II Edos that have been characterized so far suggest that at least some families (II.1 and II.6) of type II Edos may not be particularly active against catechol. On the other hand, the enzymes of the II.2 and II.5 families that have been characterized to date (not found in this screening), such as MhpB and McpI (4), GalA (6), DesB and LigAB (27), and PmdAB (46), are clearly involved in cleaving the aromatic rings of monocyclic substrates. This reinforces the view that the substrate preferences of Edos may be related to their phylogenetic origins (1, 2).

In spite of the wide variety of Edo enzymes that were identified, no members of the cupin superfamily (type III extradiol dioxygenases) were detected in the current or previous metagenomic functional analyses for extradiol dioxygenases. It was presumed that these enzymes would be present in the soil sample because they are involved in a number of aromatic degradation pathways. Although they are ring cleavage dioxygenases, the failure to identify them was not surprising because their substrates do not actually maintain a diol character (47); therefore, most of these enzymes may not be active toward the catecholic substrate indicators that were used in these analyses.

This study identified 41 new *edo* coding genes, 36 of which are for type I enzymes and are distributed in families I.2 to I.5. There were a huge number of known Edo sequences belonging to the I.2 family, and only one of the 10 new Edos belonging to this family was divergent enough to be defined as a new subfamily. This suggests that most of the diversity of the family I.2 Edos in the crude oil-contaminated soil was already identified. However, this is not the case for the 26 remaining sequences, which were highly divergent from what was previously known, as only 4 were similar enough to be ascribed to previously defined subfamilies. Particularly striking is the biodiversity of the I.3 family sequences found in this study, because only 2 of the 17 sequences were ascribed to previously defined subfamilies, and the remaining 15 sequences defined up to 7 new subfamilies, in spite of the large number of I.3 sequences and subfamilies that have already been defined.

The phylogenetic analysis of the sequences identified in this study revealed the highest diversity of Edo sequences found in a particular site thus far and show that, in spite of the quite large

number of Edo sequences that are already known, we may still be far from understanding the entirety of the biodiversity of these enzymes in nature.

ACKNOWLEDGMENTS

We thank Encarnación Carrillo for her comments/criticisms, Nuria Pérez for technical help, and all of the members of the Santero laboratory for their insights and helpful suggestions.

L.T.-G. designed and performed most of the experiments and analyzed the results, G.M.-C. also contributed to the functional screening and activity assays, M.F. participated in sequence analyses of the fosmids and discussion of the results, and E.S. designed the general strategy, supervised the work, analyzed data, and wrote the paper. All authors reviewed the manuscript.

FUNDING INFORMATION

This work was supported by grants CSD2007-0005 and BIO2011-24003, cofunded by the Spanish Ministerio de Educación y Ciencia and the European Regional Development Fund, to Eduardo Santero; by the Spanish Ministry of Economy and Competitiveness within the ERA NET-IB2 program, grant number ERA-IB-14-030 to Manuel Ferrer; and by a fellowship from the FPU program (Ministerio de Educación y Ciencia, Spain) that was awarded to Laura Terrón-González. Manuel Ferrer gratefully acknowledges the financial support provided by the European Community projects MAGIC-PAH (FP7-KBBE-2009-245226) and KILL-SPILL (FP7-KBBE-2012-312139) and the Spanish Ministry of Economy and Competitiveness (grants PCIN-2014-107 and BIO2014-54494-R). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. **Eltis LD, Bolin JT.** 1996. Evolutionary relationships among extradiol dioxygenases. J Bacteriol **178:**5930–5937.
2. **Vilchez-Vargas R, Junca H, Pieper DH.** 2010. Metabolic networks, microbial ecology and 'omics' technologies: towards understanding *in situ* biodegradation processes. Environ Microbiol **12:**3089–3104. http://dx.doi.org/10.1111/j.1462-2920.2010.02340.x.
3. **Gerlt JA, Babbitt PC.** 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem **70:**209–246. http://dx.doi.org/10.1146/annurev.biochem.70.1.209.
4. **Spence EL, Kawamukai M, Sanvoisin J, Braven H, Bugg TD.** 1996. Catechol dioxygenases from *Escherichia coli* (MhpB) and *Alcaligenes eutrophus* (MpcI): sequence analysis and biochemical properties of a third family of extradiol dioxygenases. J Bacteriol **178:**5249–5256.
5. **Sugimoto K, Senda T, Aoshima H, Masai E, Fukuda M, Mitsui Y.** 1999. Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. Structure **7:**953–965. http://dx.doi.org/10.1016/S0969-2126(99)80122-1.
6. **Nogales J, Canales A, Jimenez-Barbero J, Garcia JL, Diaz E.** 2005. Molecular characterization of the gallate dioxygenase from *Pseudomonas putida* KT2440. The prototype of a new subgroup of extradiol dioxygenases. J Biol Chem **280:**35382–35390.
7. **Zhen D, Liu H, Wang SJ, Zhang JJ, Zhao F, Zhou NY.** 2006. Plasmid-mediated degradation of 4-chloronitrobenzene by newly isolated *Pseudomonas putida* strain ZWL73. Appl Microbiol Biotechnol **72:**797–803. http://dx.doi.org/10.1007/s00253-006-0345-2.
8. **Dunwell JM, Khuri S, Gane PJ.** 2000. Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. Microbiol Mol Biol Rev **64:**153–179. http://dx.doi.org/10.1128/MMBR.64.1.153-179.2000.
9. **Moonen MJ, Kamerbeek N, Westphal AH, Boeren S, Janssen D, Fraaije M, van Berkel W.** 2008. Elucidation of the 4-hydroxyacetophenone catabolic pathway in *Pseudomonas fluorescens* ACB. J Bacteriol **190:**5190–5198. http://dx.doi.org/10.1128/JB.01944-07.
10. **Moonen MJ, Synowsky SA, Berg WA, Westphal AH, Heck AJ, Heuvel RH, Fraaije MW, van Berkel WJ.** 2008. Hydroquinone dioxygenase from *Pseudomonas fluorescens* ACB: a novel member of the family of nonheme-iron(II)-dependent dioxygenases. J Bacteriol **190:**5199–5209. http://dx.doi.org/10.1128/JB.01945-07.
11. **Thompson IP, van der Gast CJ, Ciric L, Singer AC.** 2005. Bioaugmentation for bioremediation: the challenge of strain selection. Environ Microbiol **7:**909–915. http://dx.doi.org/10.1111/j.1462-2920.2005.00804.x.
12. **Rappe MS, Giovannoni SJ.** 2003. The uncultured microbial majority. Annu Rev Microbiol **57:**369–394. http://dx.doi.org/10.1146/annurev.micro.57.030502.090759.
13. **Leigh MB, Pellizari VH, Uhlik O, Sutka R, Rodrigues J, Ostrom NE, Zhou J, Tiedje JM.** 2007. Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J **1:**134–148. http://dx.doi.org/10.1038/ismej.2007.26.
14. **Terrón-González L, Genilloud O, Santero E.** 2014. Potential and limitations of metagenomic functional analyses, p 1–43. *In* Benedetti C (ed), Metagenomics, methods, applications and perspectives. Nova Publishers, New York, NY.
15. **Suenaga H, Ohnuki T, Miyazaki K.** 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ Microbiol **9:**2289–2297. http://dx.doi.org/10.1111/j.1462-2920.2007.01342.x.
16. **Brennerova MV, Josefiova J, Brenner V, Pieper DH, Junca H.** 2009. Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. Environ Microbiol **11:**2216–2227. http://dx.doi.org/10.1111/j.1462-2920.2009.01943.x.
17. **Pessoa TBA, de Souza SS, Cerqueira AF, Rezende RP, Pirovani CP, Dias JCT.** 2013. Construction and validation of metagenomic DNA libraries from landfarm soil microorganisms. Genet Mol Res **12:**2148–2155. http://dx.doi.org/10.4238/2013.June.28.2.
18. **Terrón-González L, Medina C, Limón-Mortés MC, Santero E.** 2013. Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. Sci Rep **3:**1107. http://dx.doi.org/10.1038/srep01107.
19. **Figurski DH, Helinski DR.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci U S A **76:**1648–1652. http://dx.doi.org/10.1073/pnas.76.4.1648.
20. **Hanahan D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol **166:**557–580. http://dx.doi.org/10.1016/S0022-2836(83)80284-8.
21. **Andújar E, Hernáez MJ, Kaschabek SR, Reineke W, Santero E.** 2000. Identification of an extradiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product. J Bacteriol **182:**789–795. http://dx.doi.org/10.1128/JB.182.3.789-795.2000.
22. **Heiss G, Stolz A, Kuhm AE, Müller C, Klein J, Altenbuchner J, Knackmuss H-J.** 1995. Characterization of a 2,3-dihydroxybiphenyl dioxygenase from the naphthalenesulfonate-degrading bacterium strain BN6. J Bacteriol **177:**5865–5871.
23. **Alcaide M, Stogios PJ, Lafraya Á, Tchigvintsev A, Flick R, Bargiela R, Chernikova TN, Reva ON, Hai T, Leggewie CC, Katzke N, La Cono V, Matesanz R, Jebbar M, Jaeger KE, Yakimov MM, Yakunin AF, Golyshin PN, Golyshina OV, Savchenko A, Ferrer M.** 2015. Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. Environ Microbiol **17:**332–345. http://dx.doi.org/10.1111/1462-2920.12660.
24. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
25. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406–425.
26. **Bakken LR, Lindahl V.** 1995. Recovery of bacterial cells from soil, p 9–27. *In* van Elsas JD, Trevors JT (ed), Nucleic acids in the environment: methods and applications. Springer-Verlag, Berlin, Germany.
27. **Sakai M, Masai E, Asami H, Sugiyama K, Kimbara K, Fukuda M.** 2002. Diversity of 2,3-dihydroxybiphenyl dioxygenase genes in a strong PCB degrader, Rhodococcus sp. strain RHA1. J Biosci Bioeng **93:**421–427.
28. **Noda Y, Nishikawa S, Shiozuka K, Kadokura H, Nakajima H, Yoda K, Katayama Y, Morohoshi N, Haraguchi T, Yamasaki M.** 1990. Molecular cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis*. J Bacteriol **172:**2704–2709.
29. **Kasai D, Masai E, Miyauchi K, Katayama Y, Fukuda M.** 2005. Characterization of the gallate dioxygenase gene: three distinct ring cleavage di-

oxygenases are involved in syringate degradation by *Sphingomonas paucimobilis* SYK-6. J Bacteriol **187:**5067–5074. http://dx.doi.org/10.1128/JB.187.15.5067-5074.2005.

30. **Eaton RW.** 2001. Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B. J Bacteriol **183:**3689–3703. http://dx.doi.org/10.1128/JB.183.12.3689-3703.2001.
31. **Prieto MA, Diaz E, Garcia JL.** 1996. Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: engineering a mobile aromatic degradative cluster. J Bacteriol **178:**111–120.
32. **Roper DI, Cooper RA.** 1990. Subcloning and nucleotide sequence of the 3,4-dihydroxyphenylacetate (homoprotocatechuate) 2,3-dioxygenase gene from *Escherichia coli* C. FEBS Lett **275:**53–57. http://dx.doi.org/10.1016/0014-5793(90)81437-S.
33. **Christinet L, Burdet FX, Zaiko M, Hinz UG, Zryd J-P.** 2004. Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in *Portulaca grandiflora*. Plant Physiol **134:**265–274. http://dx.doi.org/10.1104/pp.103.031914.
34. **Weng JK, Li Y, Mo H, Chapple C.** 2012. Assembly of an evolutionarily new pathway for alpha-pyrone biosynthesis in *Arabidopsis*. Science **337:**960–964. http://dx.doi.org/10.1126/science.1221614.
35. **Gandía-Herrero F, García-Carmona F.** 2014. *Escherichia coli* protein YgiD produces the structural unit of plant pigments betalains: characterization of a prokaryotic enzyme with DOPA-extradiol-dioxygenase activity. Appl Microbiol Biotechnol **98:**1165–1174. http://dx.doi.org/10.1007/s00253-013-4961-3.
36. **Peng X, Egashira T, Hanashiro K, Masai E, Nishikawa S, Katayama Y, Kimbara K, Fukuda M.** 1998. Cloning of a *Sphingomonas paucimobilis* SYK-6 gene encoding a novel oxygenase that cleaves lignin-related biphenyl and characterization of the enzyme. Appl Environ Microbiol **64:**2520–2527.
37. **Kasai D, Masai E, Miyauchi K, Katayama Y, Fukuda M.** 2004. Characterization of the 3-O-methylgallate dioxygenase gene and evidence of multiple 3-O-methylgallate catabolic pathways in *Sphingomonas paucimobilis* SYK-6. J Bacteriol **186:**4951–4959. http://dx.doi.org/10.1128/JB.186.15.4951-4959.2004.
38. **Reineke W.** 1998. Development of hybrid strains for the mineralization of chloroaromatics by patchwork assembly. Annu Rev Microbiol **52:**287–331. http://dx.doi.org/10.1146/annurev.micro.52.1.287.
39. **Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF.** 2010. Expanding small-molecule functional metagenomics through parallel screening of broad host-range cosmid environmental DNA libraries in diverse *Proteobacteria*. Appl Environ Microbiol **76:**1633–1641. http://dx.doi.org/10.1128/AEM.02169-09.
40. **Chatfield LK, Williams PA.** 1986. Naturally occurring TOL plasmids in *Pseudomonas* strains carry either two homologous or two nonhomologous catechol 2,3 oxygenase genes. J Bacteriol **168:**878–885.
41. **Suenaga H, Mizuta S, Miyazaki K, Yaoi K.** 2014. Diversity of extradiol dioxygenases in aromatic-degrading microbial community explored using both culture-dependent and culture-independent approaches. FEMS Microbiol Ecol **90:**367–379. http://dx.doi.org/10.1111/1574-6941.12390.
42. **Suenaga H, Mizuta S, Miyazaki K.** 2009. The molecular basis for adaptive evolution in novel extradiol dioxygenases retrieved from the metagenome. FEMS Microbiol Ecol **69:**472–480. http://dx.doi.org/10.1111/j.1574-6941.2009.00719.x.
43. **Suenaga H, Koyama Y, Miyakoshi M, Miyazaki R, Yano H, Sota M, Ohtsubo Y, Tsuda M, Miyazaki K.** 2009. Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. ISME J **3:**1335–1348. http://dx.doi.org/10.1038/ismej.2009.76.
44. **Vaillancourt FH, Labbe G, Drouin NM, Fortin PD, Eltis LD.** 2002. The mechanism-based inactivation of 2,3-dihydroxybiphenyl 1,2-dioxygenase by catecholic substrates. J Biol Chem **277:**2019–2027. http://dx.doi.org/10.1074/jbc.M106890200.
45. **McKay DB, Prucha M, Reineke W, Timmis KN, Pieper DH.** 2003. Substrate specificity and expression of three 2,3-dihydroxybiphenyl 1,2-dioxygenases from *Rhodococcus globerulus* strain P6. J Bacteriol **185:**2944–2951. http://dx.doi.org/10.1128/JB.185.9.2944-2951.2003.
46. **Mampel J, Providenti MA, Cook AM.** 2005. Protocatechuate 4,5-dioxygenase from *Comamonas testosteroni* T-2: biochemical and molecular properties of a new subgroup within class III of extradiol dioxygenases. Arch Microbiol **183:**130–139. http://dx.doi.org/10.1007/s00203-004-0755-4.
47. **Fetzner S.** 2012. Ring-cleaving dioxygenases with a cupin fold. Appl Environ Microbiol **78:**2505–2514. http://dx.doi.org/10.1128/AEM.07651-11.