39 Phylogenomics of Aerobic Bacterial Degradation of Aromatics

D. Pérez-Pantoja[1] · *R. Donoso*[1] · *H. Junca*[2,3] · *B. González*[1] · *D. H. Pieper*[3]
[1]Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, NM-EMBA, CASEB, P. Universidad Católica de Chile, Santiago, Chile
[2]Centro Colombiano de Metagenómica y Bioinformática de Ambientes Extremos (GeBiX) – Grupo de Genética Molecular, Corporación CorpoGen, Bogotá, Colombia
[3]Biodegradation Research Group, Department of Microbial Pathogenesis, HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
dpi@helmholtz-hzi.de

K. N. Timmis (ed.), *Handbook of Hydrocarbon and Lipid Microbiology*, DOI 10.1007/978-3-540-77587-4_95,

Abstract: Aromatic compounds are widely distributed in nature. They are found as lignin components, aromatic amino acids, and xenobiotic compounds, among others. Microorganisms, mostly bacteria, degrade an impressive variety of such chemical structures. Various aerobic aromatic catabolic pathways have been reported in bacteria, which typically consist of activation of the aromatic ring through oxygenases or CoA ligases and ring cleavage of di- or trihydroxylated intermediates or dearomatized CoA derivatives. We survey almost 900 sequenced bacterial genomes for the presence of genes encoding key enzymes of aromatic metabolic pathways, including ring-cleavage enzymes as well as enzymes activating aromatics or dearomatizing CoA derivatives. The metabolic diversity is discussed from two angles: the spread of such key activities among different bacterial phyla and the overall metabolic potential of members of bacterial genera.

1 Introduction

A few non-mutually exclusive choices are possible to address the analysis of the genetic basis of bacterial degradation of aromatic compounds. One is to select a few well-studied bacterial catabolic models and go in depth into their genetic organization of aromatic catabolism genes (Jimenez et al., 2002; Pérez-Pantoja et al., 2008). Another approach is to select a few central catabolic pathways and to assess the similarities and differences in gene organization, substrate range, and regulatory elements, among the bacteria where such pathways have been described. A third possibility is to look for all the aromatic catabolism pathways present in bacteria, searching in the growing database of sequenced bacterial genomes. The latter, by definition, is a less in-depth analysis but has the broader coverage possible today. We selected the latter approach, because we think it provides clues on the distribution of catabolic properties among bacterial phyla, gives some hints on the ecological functions of specific bacterial groups, defines underscored research objectives, and gives a better overview of the genetic basis of bacterial catabolism of aromatics. The phylogenomic approach to study the organization of aromatic degradation is based on the selection of sequences of key catabolic functions to fish into the sequenced genome database, followed by refinement of the positive scores. With this information, the genomes can be analyzed in terms of presence/absence of catabolic abilities among bacterial groups, new enzyme families based on the sequence similarity be defined, new putative functions be suggested, and evolutionary links among different groups of sequences be addressed. Of course such approach has some limitations, as most of the new data are not supported by biochemical or genetic studies. To minimize such limitations, the selected sequence probes were derived from both biochemical and genetic well-studied systems. One of the main purposes of the following material is to provide to the reader new research venues to get a deeper knowledge on bacterial catabolism of aromatics.

2 Aerobic Aromatic Catabolic Routes

Bacterial degradation of aromatic compounds and their haloaromatic derivatives has been well studied (See Chapter 4, Vol. 2, Part 2; Chapter 5, Vol. 2, Part 2). Various pathways for degradation of these compounds by bacteria have been reported. The activation of the

aromatic ring commonly proceeds by members of one of three superfamilies: the Rieske non-heme iron oxygenases usually catalyzing the incorporation of two oxygen atoms (although some members of this superfamily also catalyze monooxygenations) (Gibson and Parales, 2000), the flavoprotein monooxygenases (van Berkel et al., 2006), and the soluble diiron multicomponent oxygenases (Leahy et al., 2003). Further metabolism is achieved through di- or trihydroxylated aromatic intermediates. Alternatively, activation is mediated by CoA ligases and the formed CoA derivatives are subjected to oxygenations. This can proceed through 2-aminobenzoyl-CoA monooxygenase/reductase, an enzyme that catalyzes both monooxygenation and hydrogenation, and where the N-terminal part of the protein shows similarities to single-component flavin monooxygenases (Buder and Fuchs, 1989). Alternatively, the aromatic CoA derivative is attacked by multicomponent enzymes, where the oxygenase subunits belong to the diiron oxygenases, like in phenylacetyl-CoA (Ismail et al., 2003) or benzoyl-CoA oxygenase (Zaar et al., 2004). Various further key reactions channeling aromatics to central di- or trihydroxylated intermediates, such as the processing of side chains or demethylations, will not be discussed here (See ❯ Chapter 4, Vol. 2, Part 2).

The further aerobic degradation of di- or trihydroxylated intermediates can be catalyzed by either intradiol or extradiol dioxygenases. While all intradiol dioxygenases described thus far belong to the same superfamily, members of at least three different families are reported to be involved in the extradiol ring cleavage of hydroxylated aromatics. Type I extradiol dioxygenases (e.g., catechol 2,3-dioxygenases) belong to the vicinal oxygen chelate superfamily enzymes (Gerlt and Babbitt, 2001), the type II or LigB superfamily of extradiol dioxygenases which comprise among other protocatechuate 4,5-dioxygenases (Sugimoto et al., 1999) and the type III enzymes such as gentisate dioxygenases which comprise enzymes belonging to the cupin superfamily (Dunwell et al., 2000). However, even though belonging to different families, all three types of extradiol dioxygenases share similar active sites and all type I, type II, and various type III enzymes have the same iron ligands, two histidine and one glutamate, that constitute the 2-His 1-carboxylate structural motif. The recently identified benzoquinol 1,2-dioxygenase from the 4-hydroxyacetophenone-degrading *Pseudomonas fluorescens* ACB that displays no significant sequence identity with known dioxygenases may constitute the prototype of a novel fourth class of Fe^{2+}-dependent dioxygenases (Moonen et al., 2008).

3 Sequenced Bacterial Genomes

Currently (as of September 2008) approximately 1,000 genomes have been sequenced and three quarters of them finished. For the purpose of this review, we concentrated on genomes that were simultaneously represented in both the Integrated Microbial Genomes (IMG) database at DOE Joint Genome Institute (JGI) (img.jgi.doe.gov/cgi-bin/pub/main.cgi?page=home) and the National Center for Biotechnology Information (NCBI) database at National Institute of Health (NIH) (www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), summing up to 822 genomes. The number of representatives of the bacterial phyla in these public databases is highly variable: from a very few members from the phyla Aquificae (2) Acidobacteria (2), Chlamydiae (11), Chlorobi (10), Chloroflexi (8), Deinococcus/Thermus (4), Fusobacteria (2), Lentisphaerae (2), Planctomycetes (3), Spirochaetes (9), Thermotogae (6), and Verrucomicrobia (1); the medium represented phyla: Actinobacteria (53), Bacteroidetes (28), Cyanobacteria (40), and the Proteobacteriales δ- (23) and ε- classes (28); and the highly represented phylum Firmicutes (182) and the α- (112) β- (71) and γ- (223) classes of Proteobacteria (besides two

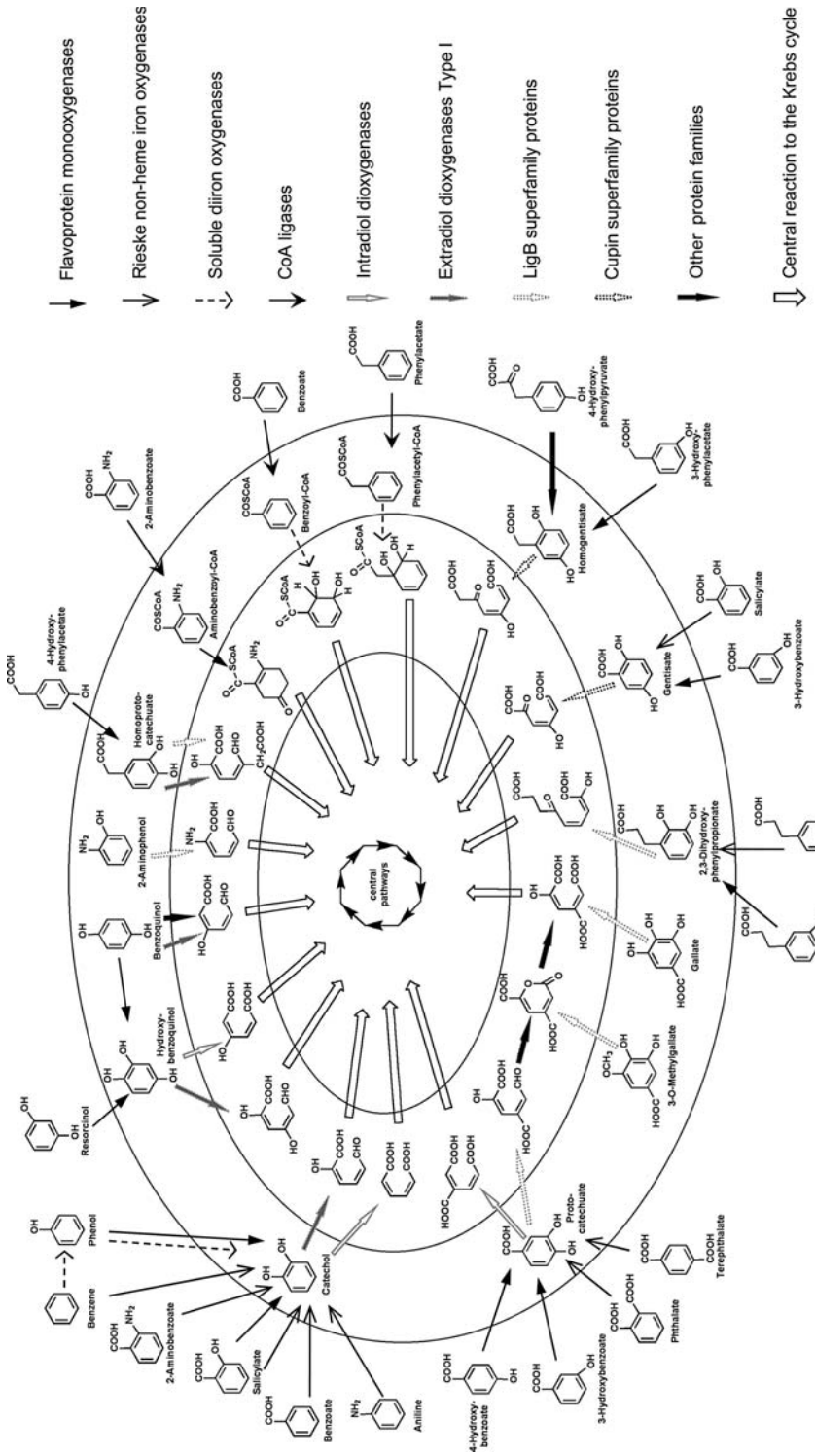

Figure 1 (Continued)

unclassified Proteobacteria). Despite of that, the number of bacterial genomes is now significant to search for the presence/absence of the main catabolic pathways for aromatic compounds to provide a reasonable idea about the spread of these catabolic abilities among the main phylogenetic groups.

4 Spread of Members of Gene Families

4.1 Intradiol Dioxygenases

The intradiol cleavage of catechol to muconate and of protocatechuate to 3-carboxymuconate by catechol 1,2-dioxygenases and protocatechuate 3,4-dioxygenases, respectively, is a central reaction in the metabolism of various aromatic compounds (Fig. 1). Hydroxybenzoquinol (1,2,4-trihydroxybenzene) is also a central intermediate in the degradation of a variety of aromatic compounds such as resorcinol (Fig. 1), with hydroxybenzoquinol 1,2-dioxygenase as key enzyme, catalyzing intradiol cleavage to form 3-hydroxy-*cis,cis*-muconate and its tautomer, maleylacetate. Among the different groups of enzymes significant metabolic cross-reactivity is usually not observed. Phylogenetic analysis of the deduced protein sequences of intradiol dioxygenases encoded in the genomes of bacteria sequenced so far showed the presence of seven clusters as indicated in Fig. 2.

Based on biochemical or genetically validated representatives, cluster 1 comprises hydroxybenzoquinol dioxygenases, cluster 2 proteobacterial catechol 1,2-dioxygenases, cluster 3 actinobacterial catechol 1,2-dioxygenases, and clusters 5 and 7 the α- and β-subunits of protocatechuate 3,4-dioxygenases, respectively. Enzymes of cluster 6 are obviously related to the β-subunits of protocatechuate dioxygenases, however, in no case genes encoding these enzymes are clustered with genes encoding putative α- subunits, and the function of these enzymes remains to be elucidated. Similarly, the function of enzymes of cluster 4 wait for clarification.

Intradiol dioxygenases are nearly exclusively found in two phyla, the Actinobacteria and the Proteobacteria. However, protocatechuate 3,4-dioxygenases were observed in one of the two sequenced Deinococci, i.e., *Deinococcus geothermalis* DSM 11300 and one of the two sequenced Acidobacteria, i.e., *Solibacter usitatus* Ellin6076. Considering the wide spread of Acidobacteria in the environment, their involvement in aromatic degradation under natural

Figure 1

Aerobic metabolism of aromatics via di- or trihydroxylated intermediates, or via CoA derivatives. Peripheral hydroxylation reactions can be catalyzed by flavoprotein monooxygenases, Rieske non-heme iron oxygenases or soluble diiron oxygenases. Alternatively, aromatics can be activated through CoA ligases followed by dearomatization catalyzed by members of the flavoprotein monooxygenases or soluble diiron oxygenases. 4-Hydroxyphenylpyruvate dioxygenase is indicated by a Central di- or trihydroxylated intermediates are subjected to ring cleavage by intradiol dioxygenases or extradiol dioxygenases of the vicinal chelate superfamily, the LigB superfamily or the cupin superfamily. Ring-cleavage products are channeled to the Krebs cycle via central reactions.

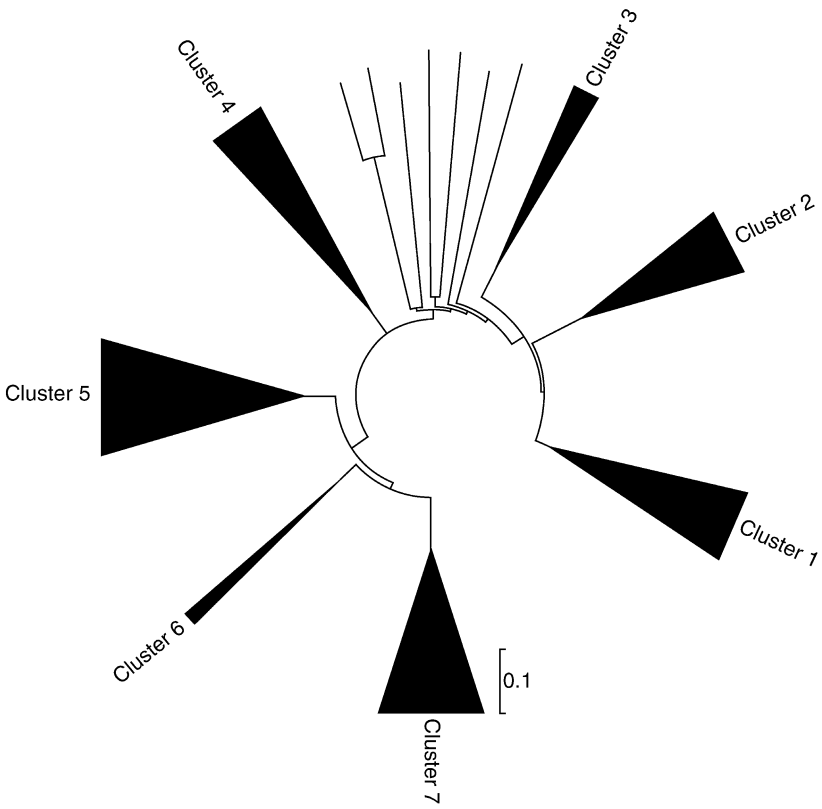

Figure 2

Evolutionary relationships among intradiol dioxygenases. The evolutionary history was inferred using the neighbor joining method after alignment of sequences using MUSCLE (Edgar, 2004). All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Wedges represent enzyme clusters as described in the text. Deduced protein sequences not falling inside the defined clusters are also indicated. Wedge length is a measure of evolutionary distance from the common ancestor. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007). Cluster 1 comprises hydroxybenzoquinol dioxygenases, cluster 2 proteobacterial catechol 1,2-dioxygenases, cluster 3 actinobacterial catechol 1,2-dioxygenases, and clusters 5 and 7 the α- and β-subunits of protocatechuate 3,4-dioxygenases, respectively. The functions of enzymes of clusters 4 and 6 remain to be elucidated.

conditions has to be considered. Actually, Acidobacteria have been implied to be involved in the biogeochemical cycles of rhizosphere soil (Lee et al., 2008).

Regarding catechol 1,2-dioxygenases, where two lineages have previously been described (Eulberg et al., 1997), phylogenetic analysis confirmed that cluster 3 enzymes are restricted to members of the order *Actinomycetales* of the Actinobacteria, and catechol intradiol cleavage pathways were observed in the majority of *Corynebacteria, Arthrobacter, Mycobacteria, and Nocardiaceae*. Usually, Actinobacteria possessing a catechol intradiol cleavage pathway also harbor a protocatechuate intradiol cleavage. However, *Streptomyces* strains seem to be

endowed only with the protocatechuate branch. A hydroxybenzoquinol pathway seems to be spread only in *Corynebacteria* and out of the *Mycobacteria*, only *Mycobacterium smegmatis* and *M. vanbaalenii* are endowed with such a pathway.

As shown in Table 1, intradiol dioxygenases can be identified in 11 out of 19 α-proteobacterial, 2 out of 10 β-proteobacterial, and 4 out of 29 γ-proteobacterial families and are absent in δ- or ε- proteobacteria. Significant differences in gene spread were observed among families. Catechol intradiol pathways are observed in nearly all *Pseudomonas* strains and are absent only from the genomes of *P. syringae* and *P. mendocina*. The last one is also the only *Pseudomonas* strain devoid of a protocatechuate intradiol pathway. Similarly, both protocatechuate and catechol pathways are observed in all *Burkholderia* genomes. Interestingly, catechol intradiol cleavage pathways were only exceptionally observed in α-Proteobacteria. In contrast, a catechol pathway is absent in *Rhizobiaceae*, which, however, often bear a hydroxybenzoquinol pathway. Also *Bradyrhizobiaceae*, none of which has a catechol pathway, are usually endowed with a hydroxybenzoquinol pathway except for *Nitrobacter* strains.

4.2 EXDO I Family

The extradiol ring cleavage of catechol is typically catalyzed by type I extradiol dioxygenases (EXDO I), which belong to the vicinal oxygen chelate superfamily (Gerlt and Babbitt, 2001).

Table 1

Intradiol dioxygenases observed in genomes of Proteobacteria

Class	Family	Protocatechuate 3,4-dioxygenase (Pca34)	Catechol 1,2-dioxygenase (Cat12)	Hydroxybenzo-quinol dioxygenase (Hqu)
α	*Caulobacteraceae* (2)	++	–	–
α	*Aurantimonadaceae* (2)	+	–	–
α	*Bradyrhizobiaceae* (11)	(–)	–	++
α	*Brucellaceae* (6)	++	–	–
α	*Methylobacteriaceae* (3)	+	+	–
α	*Phyllobactriaceae* (3)	++	–	–
α	*Rhizobiaceae* (6)	++	–	++
α	*Rhodobacteraceae* (24)	++	–	+
α	*Xanthobacteriaceae* (2)	++	+	++
α	*Acetobacteraceae* (3)	+	–	–
α	*Sphingomonadaceae* (5)	(–)	+	(–)
β	*Burkholderiaceae* (43)	++	++	+
β	*Comamonadaceae* (8)	(–)	+	+
γ	*Oceanosprillaceae* (3)	–	+	+
γ	*Moraxellaceae* (5)	+	+	–
γ	*Pseudomonadaceae* (19)	++	++	(–)
γ	*Xanthomonadaceae* (11)	+	–	–

++; More than 60% of the sequenced genomes of these bacterial taxa comprise a gene encoding the mentioned activity (number of sequenced representatives is given in parentheses); +, between 20 and 60%; (–), less than 20%; –, not observed

The EXDO I family comprises enzymes that catalyze the dioxygenolytic ring fission of the catecholic derivatives in several bacterial mono- and polyaromatics biodegradation pathways (Eltis and Bolin, 1996) (*Fig. 1*) like those involved in degradation of benzene, toluene, phenol, biphenyl, naphthalene, dibenzofuran, 4-hydroxyphenylacetate, *p*-cymene, or diterpenoid compounds such as abietate. They catalyze the *meta*-cleavage of catechol to 2-hydroxymuconic semialdehyde (catechol 2,3-dioxygenases, C23O), of 2,3-dihydroxybiphenyl (2,3-dihydroxybiphenyl 1,2-dioxygenases, BphC), 1,2-dihydroxynaphthalene (NahC), homoprotocatechuate (homoprotocatechuate 2,3-dioxygenases, HpaD), 2,3-dihydroxy-*p*-cumate (2,3-dihydroxy-*p*-cumate-3,4-dioxygenases CmtC), and 7-oxo-11,12-dihydroxydehydroabietate (DitC), among others (see also *Fig. 1*).

In many cases, the respective genes are localized in catabolic pathway gene clusters such that their actual function can easily be deduced. However, in various cases multiple EXDO I activities are observed in a single strain and often their function remains unproven (Maeda et al., 1995). Here, the names are given to the enzymes according to the preferential activity observed, but in many cases there is a range of structurally similar substrates that can be metabolized by the same enzyme with varying catalytic efficacies and the "natural" substrate has not yet been identified.

Because genome annotations pipelines are in many cases using the NCBI Conserved Domains Database (CDD), which is in turn, interconnected with the Wellcome Trust Sanger Institute Pfam database descriptions, all EXDO I genes found in the genome sequences are recognized and annotated with the superfamily name as Glyoxalase/bleomycin resistance protein/dioxygenase (InterPro: IPR004360, pfam00903: Glyoxalase). However, in the majority of cases, a more precise annotation of several genomic sequences as EXDO I would be possible, as they show conservation of the Prosite PS00082 extradiol ring-cleavage dioxygenases signature [GNTIV]-x-H-x(5,7)-[LIVMF]-Y-x(2)-[DENTA]-P-x-[GP]-x(2,3)-E.

Phylogenetic analysis of the deduced protein sequences of EXDO I encoded in the genomes of bacteria sequenced so far, and retrieved after iterative PSI Blast searches using representative proteins of major clusters where a function has been described as seeds show the presence of three major evolutionary lineages (*Fig. 3*).

One of these lineages (cluster 1) comprises nearly all EXDO I proteins of validated function. Ten subclusters (A–J) grouping proteins associated with different substrate specificities can be differentiated. Subcluster 1A comprises enzymes experimentally validated as C23O. Interestingly, there is a high redundancy in genomes, as the 28 identified genes are observed in only 18 strains. Out of these, 13 strains belong to the β-proteobacteria and C230 is mainly observed in *Burkholderia*, *Cupriavidus*, and *Ralstonia* genomes. This contrasts previous reports on C23Os, which were predominantly characterized from *Pseudomonas* strains (Eltis and Bolin, 1996). However, in none of the sequenced *Pseudomonas* a homologous gene is observed. It has, however, to be noted that most of such genes have previously been reported on plasmids rather than in the chromosome of the strains, such as the case for *P. putida* KT2440 where the IncP-9 TOL plasmid pWW0 is present (Williams and Murray, 1974), but not included in the same genome project. It is also interesting to note that the Actinobacterium *R. jostii* RHA1 has a predicted C23O of this kind.

Subcluster 1B groups putative homoprotocatechuate 2,3-dioxygenases of the actinobacterial lineage. As expected from literature, the respective encoding genes are present in Actinobacteria (Vetting et al., 2004), and observed in 5 out of 53 genomes. They are absent from any β- and γ-proteobacterial genomes, but surprisingly most abundant in α-proteobacterial genomes

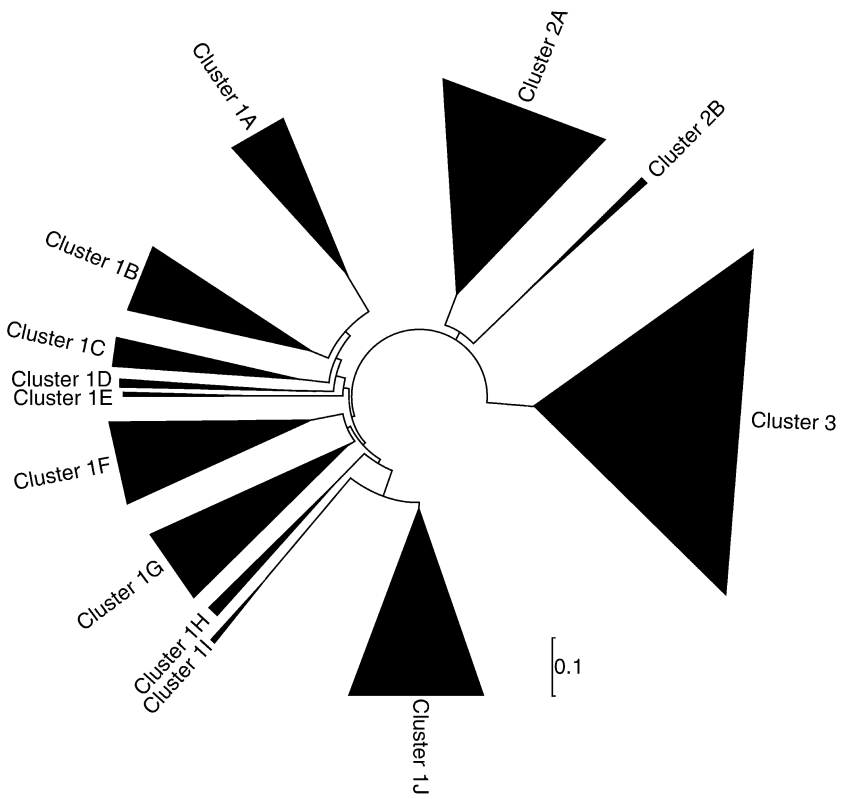

Figure 3

Evolutionary relationships among type I extradiol dioxygenases (EXDO I). Subcluster 1A comprises catechol 2,3-dioxygenases, subcluster 1B putative homoprotocatechuate 2,3-dioxygenases, subcluster 1C proteins related to BphC of *Bacillus* sp. JF8, subcluster 1D proteins related to NahC of *Bacillus* sp. JF8, subcluster 1G proteins related to DntD of *Burkholderia* sp. DNT or BphC3 and BphC4 of *R. jostii* RHA1, subcluster 1H proteins similar to those capable to cleave 2,3-dihydroxy-*p*-cumate, subcluster 1I proteins related to those involved in diterpenoid degradation, and subcluster 1J enzymes with similarities to those being active mainly against bicyclic and higher condensed dihydroxylated aromatics. Subcluster 2B comprises so-called one-domain extradiol dioxygenases and cluster 3 proteins related to LinE chlorobenzoquinol 1,2-dioxygenases and PcpA 2,6-dichlorobenzoquinol 1,2-dioxygenases. However, the function of the majority of enzymes of cluster 3 as well as of enzymes of subclusters 1E, 1F, and 2A remains to be elucidated.

(16 genomes), specifically in *Bradyrhizobiaceae* and *Rhodobacteraceae*, even though proteobacterial homoprotocatechuate 2,3-dioxygenases are generally assumed to be members of the LigB family (see below) (Roper and Cooper, 1990). It is also interesting to note that such genes were found also outside the Actinobacteria and Proteobacteria, and are present in both sequenced *Deinococcus* and in both sequenced *Thermus* strains as well as in three *Bacillaceae*.

Subcluster 1C groups proteins related to BphC of *Bacillus* sp. JF8 involved in biphenyl degradation by this strain (Hatta et al., 2003). Related proteins are not encoded in any of the sequenced Bacilli, but astonishingly in all four genomes available of *Chloroflexaceae* strains and in a few actinobacterial species, including one protein of *R. jostii* RHA1, however, not having a taxonomically linked distribution in lower levels. Similarly, proteins related to NahC 1,2-dihydroxynaphthalene dioxygenase of *Bacillus* sp. JF8 (Miyazawa et al., 2004) (subcluster 1D) are not observed in any *Bacillus* species, but encoded in four α-proteobacterial genomes. Also the three subcluster 1E proteins, where no closely related proteins have been characterized so far, are encoded in two α-proteobacterial genomes.

Subcluster 1F proteins are encoded by all 34 genomes available of *Burkholderia* and various other proteobacterial genomes, however, their actual function still remains to be elucidated.

Subcluster 1G comprises proteins such as DntD of *Burkholderia* sp. DNT responsible for *meta*-cleavage of trihydroxytoluene, which is also active on catechol (Haigler et al., 1999) but includes as well various proteins of proven activity against 2,3-dihydroxybiphenyl such as BphC3 and BphC4 of *R. jostii* RHA1, both being reported as being practically inactive with catechol (Sakai et al., 2002). Similar proteins are mainly observed in genomes of Actinobacteria, with *R. jostii* RHA1 harboring three of such genes, and α- and β-Proteobacteria. Proteins similar to those capable to cleave 2,3-dihydroxy-*p*-cumate (subcluster 1H) are only found in four genomes including *P. putida* F1 reported to exhibit such activity (Eaton, 1996) and *B. xenovorans* LB400, indicating that it is not a widespread activity. Similarly, proteins related to those involved in diterpenoid degradation (subcluster 1I) (Martin and Mohn, 2000), are not common in the genomes analyzed, showing only hits in *Caulobacter* sp. K31 and the already described activity of *B. xenovorans* LB400 (Smith et al., 2007). Subcluster 1J comprises a variety of enzymes with similarities to members of subfamilies I.4, I.5, and I.3.E being active mainly against bicyclic and higher condensed dihydroxylated aromatics (Eltis and Bolin, 1996). An overall of 68 such proteins could be observed to be encoded in thus far sequenced genomes. Respective genes are observed in 11 of 17 Mycobacterial genomes, which is not astonishing, as various sequenced *Mycobacteria* were selected for their capability to mineralize polycyclic aromatics. They are also observed in all three Nocardiaceae genomes, with *R. jostii* RHA1 harboring six such genes. In addition, eight α-, eight β-, and five γ-proteobacterial strains harbor such enzyme. Out of the *Pseudomonas*, it was observed only in the *P. putida* F1 genome (Zylstra et al., 1988).

The majority of the approximately 100 protein sequences conforming cluster 2 contain the Prosite PS00082 extradiol ring-cleavage dioxygenase signature described above. Subcluster 2B comprises BphC6 of *R. jostii* RHA1 (ABO34703) and other previously characterized so-called one-domain extradiol dioxygenases such as BphC2 and BphC3 from *R. globerulus* P6 with reported activity against 2,3-dihydroxybiphenyl (Asturias and Timmis, 1993) (subfamily I.1 as defined by Eltis and Bolin (Eltis and Bolin, 1996)). However, besides BphC6 of strain RHA1, no further enzyme of this type was found to be encoded in the genomes analyzed, and proteins with similarity to subcluster 2A proteins have not yet been functionally characterized.

Ring-cleavage dioxygenases involved in the turnover of (chloro)benzoquinols and (chloro)hydroxybenzoquinols have been identified from various microorganisms degrading γ-hexachlorocyclohexane or chlorophenols, and comprise LinE chlorobenzoquinol/benzoquinol 1,2-dioxygenases, which preferentially cleaves aromatic rings with two hydroxyl groups at *para* positions (Miyauchi et al., 1999) and PcpA 2,6-dichlorobenzoquinol 1,2-dioxygenases (Xu et al., 1999). These proteins are comprised in cluster 3, and are the only validated extradiol dioxygenases observed in this cluster. Compared to cluster 1, cluster 3 is so divergent that even

the Superfam HMM system recognizes the validated LinE/PcpA sequences as part of the Glyoxalase/bleomycin resistance protein/dioxygenase superfamily but belonging to the family of Glyoxalase I (lactoylglutathione lyase). Only the genomes of *Cupriavidus necator* H16 and JMP134 contain sequences that may have encode chlorobenzoquinol dioxygenases. It should be noted that one of the sequences of *C. necator* JMP134 is clustered with a gene similar to the one described from *P. putida* HS12 encoding nitrobenzene nitroreductase, which is also clustered with a putative benzoquinol extradiol dioxygenase (Park and Kim, 2000).

4.3 Lig B Superfamily

A second family of extradiol dioxygenases is the so-called LigB family (Sugimoto et al., 1999). LigB type extradiol dioxygenases are well established as being responsible for the degradation of protocatechuate via the protocatechuate 4,5-dioxygenase pathway. Protocatechuate dioxygenases are composed of two distinct subunits, with the active site being located in the β-subunit. Also, proteobacterial homoprotocatechuate 2,3-dioxygenases as the one described in *Escherichia coli* (Roper and Cooper, 1990) belong to the type II or LigB superfamily of extradiol dioxygenases whereas actinobacterial homoprotocatechuate 2,3-dioxygenases are supposed to belong to the EXDO I (Vetting et al., 2004). A further well-documented group of LigB-type extradiol dioxygenases are the 2,3-dihydroxyphenylpropionate 1,2-dioxygenases which, like LigB-type homoprotocatechuate dioxygenases, consist only of one type of subunit (Diaz et al., 2001). Recent analyses have revealed various other substrates that are cleaved by LigB-type extradiol dioxygenases. Aminophenol 1,6-dioxygenases (*Fig. 1*) are, like protocatechuate 4,5-dioxygenases, composed of two distinct subunits, with the β-subunits containing the active site (Takenaka et al., 1997). Gallate dioxygenases have so far been described in *S. paucimobilis* SYK-6 (Kasai et al., 2005) and *P. putida* KT2440 (Nogales et al., 2005), and are specific for this substrate and do not transform protocatechuate, whereas gallate transformation by protocatechuate 4,5-dioxygenases has been reported. Both gallate dioxygenases have sizes significantly larger than those of the β-subunits of protocatechuate dioxygenases. Analysis of the primary structure revealed that the N-terminal regions showed a significant amino acid sequence identity with the β-subunit of protocatechuate 4,5-dioxygenases, whereas the C-terminal region has similarity to the corresponding small α-subunit (Nogales et al., 2005). It was therefore suggested that gallate dioxygenases are two-domain proteins that have evolved from the fusion of large and small subunits. Additional LigB-type enzymes have been described to be involved in the degradation of methylgallate (Kasai et al., 2004) or of bi- and polycyclic aromatics (Laurie and Lloyd-Jones, 1999).

Phylogenetic analysis of the deduced protein sequences of LigB-type proteins encoded in the genomes of bacteria sequenced so far allowed the identification of six clusters (*Fig. 4*).

Cluster 1 comprises three subclusters, which contain protocatechuate 4,5-dioxygenase β-subunits (*Fig. 4*, cluster 1A), gallate dioxygenases (cluster 1B), and a group of related proteins where no member has been characterized thus far (cluster 1C). Respective genes were nearly exclusively observed in α-, β-, and γ-Proteobacteria and only 1 of the 53 analyzed actinobacterial genomes (*Arthrobacter* sp. FB24) has a protocatechuate 4,5-dioxygenase encoding gene. Protocatechuate 4,5-dioxygenases are predominantly observed in *Comamonadaceae* and *Bradyrhizobiaceae*, specifically *Bradyrhizobium* and *Rhodopseudomonas* strains and are mainly composed of two distinct subunits as evidenced by two subsequent genes encoding the respective subunits. However, putative gene fusions are observed in *Arthrobacter* and

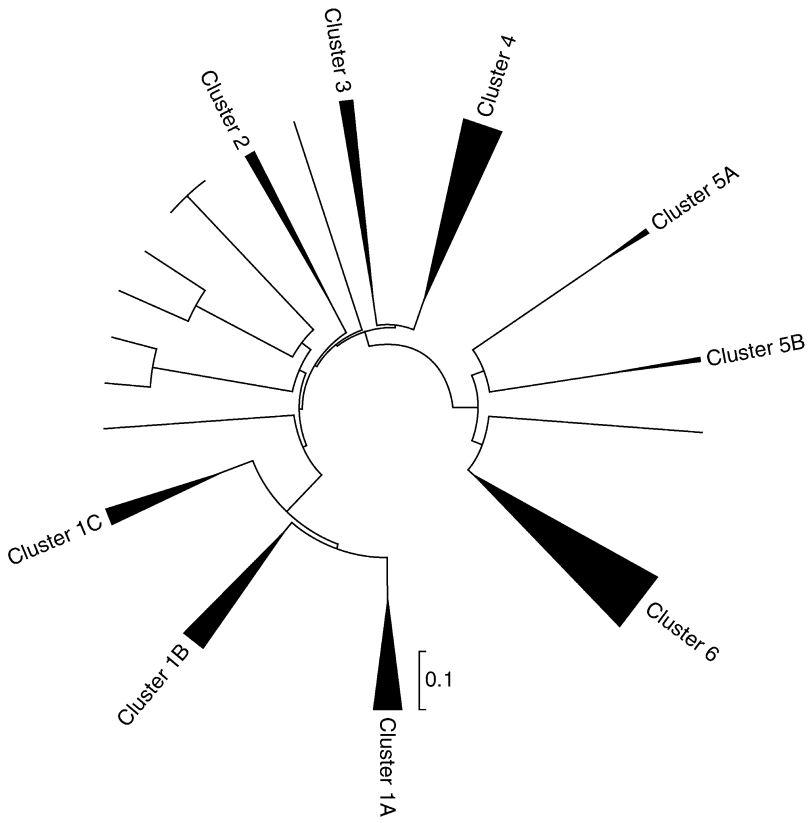

Figure 4

Evolutionary relationships among LigB-type dioxygenases. Subcluster 1A comprises protocatechuate 4,5-dioxygenase β-subunits, subcluster 1B gallate dioxygenases, cluster 2, enzymes most closely related to PhnC of *Burkholderia* sp. strain RP007 or CarBb of *P. resinovorans* CA10, cluster 3 enzymes related to DesZ of *Sphingomonas paucimobilis* SYK-6, cluster 4 2,3-dihydroxyphenylpropionates 1,2-dioxygenases, cluster 5 the β- and α-subunits (clusters 5A and B, respectively) of 2-aminophenol 1,6-dioxygenases, and cluster 6 homoprotocatechuate 2,3-dioxygenases. The function of enzymes of subcluster 1C remains to be elucidated.

Verminephrobacter. Even though one of the two gallate dioxygenases characterized so far was reported in a *Sphingomonas* strain (Kasai et al., 2005), gallate dioxygenase encoding genes are not observed in any of the 112 sequenced α-Proteobacteria and are thus not a dominant trait in this group. In contrast, gallate dioxygenases are obviously encoded in the genomes of three of four sequenced *P. putida* strains. The supposed gallate dioxygenases are mainly fusions of α- and β-subunits, like in *P. putida* KT2440 (Nogales et al., 2005), however, seem to consist of separate subunits in *Xanthomonas* and *Chromohalobacter*. Dioxygenases belonging to the third subcluster are usually composed of α- and β-subunits, and are in 10 out of 12 cases encoded in genomes, which also encode a protocatechuate 4,5-dioxygenase pathway.

A second cluster (cluster 2, *Fig. 4*) comprises enzymes most closely related to those involved in bi- and polycyclic aromatic degradation such as PhnC involved in the degradation

of polycyclic aromatics by *Burkholderia* sp. strain RP007 (Laurie and Lloyd-Jones, 1999), CarBb involved in the degradation of carbazol by *P. resinovorans* CA10 (Sato et al., 1997a), or BphC6 involved in the degradation of fluorene by *Rhodococcus rhodochrous* K37 (Taguchi et al., 2004). However, no clear association with a capability to degrade such compounds was evident, and the respective enzymes are spread among very different groups of Actinobacteria and Proteobacteria. The corresponding genes are absent from strains selected for genome sequencing due to their exceptional capability to degrade aromatics such as *M. vanbaalenii* Pyr, *M. gilvium* PYR-GCK, *R. jostii* RHA1, or *B. xenovorans* LB400.

Cluster 3 comprises enzymes related to DesZ methylgallate dioxygenase of *Sphingomonas paucimobilis* SYK-6, where 7 out of 11 proteins are observed in *Mycobacterium* strains, however, their function remains to be elucidated.

A fourth cluster obviously comprises 2,3-dihydroxyphenylpropionate 1,2-dioxygenases. The respective enzymes are most dominantly observed to be encoded in the genomes of *Enterobacteriaceae*, and specifically observed in 13 out of 18 *E. coli* strains sequenced and in *Shigella sonnei*. Interestingly, related enzymes are also observed to be encoded by 9 out 17 Mycobacterial genomes. Their function, however, remains to be proven.

A fifth cluster comprises 2-aminophenol 1,6-dioxygenases (*Fig. 4*, clusters 5A and B comprising the β- and α-subunits, respectively). Only two of these enzymes are observed to be encoded by previously sequenced genomes, i.e., *B. xenovorans* LB400 and *P. putida* W619, indicating such pathways to be present only in very few specialized bacteria. In contrast, homoprotocatechuate 2,3-dioxygenases (cluster 6) are observed to be widespread, and in contrast to previous assumptions that LigB-type homoprotocatechuate 2,3-dioxygenases were restricted to proteobacteria, homologues are also observed in two Actinobacteria, and the genomic context suggest that those enzymes actually are part of a functional homoprotocatechuate pathway. A homologue is also observed in *Bacillus licheniformis*.

4.4 Cupin Dioxygenases

Several extradiol dioxygenases of aromatic degradation pathways have been described to belong to the cupin superfamily (Dunwell et al., 2000) sharing a common architecture and including key enzymes such as gentisate 1,2-dioxygenase (involved in the degradation of salicylate or 3-hydroxybenzoate, *Fig. 1*), homogentisate 1,2-dioxygenase (involved in the degradation of phenylalanine and tyrosine) (Arias-Barrau et al., 2004) and 3-hydroxyanthranilate 3,4-dioxygenase (involved in tryptophan degradation) (Kurnasov et al., 2003; Muraki et al., 2003). The phylogenomic analysis of this type of dioxygenases in the genomes of bacteria sequenced so far shows that homogentisate dioxygenase is the enzyme with the broadest distribution in bacterial families. This may be explained by the key role in the degradation of the aromatic amino acids phenylalanine and tyrosine in several organisms, including eukaryotes. Putative genes encoding this enzyme are strongly represented in Proteobacteria, being identified in 10 out of 19 α-, 5 out of 10 β-, 16 out of 29 γ-, and 4 out of 11 δ-proteobacterial families, although they were absent in ε-proteobacteria. In the families *Bradyrhizobiaceae, Rhizobiaceae, Alcaligenaceae, Burkholderiaceae, Shewanellaceae, Legionellaceae, Pseudomonadaceae*, and *Vibrionaceae*, a respective gene can be observed in nearly all genomes sequenced. Homogentisate 1,2-dioxygenase was the unique aromatic ring-cleavage enzyme found in sequenced representatives of the families *Hyphomonadaceae, Neisseriaceae, Aeromonadaceae, Idiomarinaceae, Moritellaceae, Chromatiaceae, Legionellaceae, Hahellaceae, Bdellovibrionaceae,*

Cystobacteraceae, and *Nannocystaceae*. In addition, genes putatively encoding homogentisate 1,2-dioxygenase are also found in members of the non-proteobacterial orders *Actinomycetales, Flavobacteriales, Sphingobacteriales*, and *Bacillales*.

Gentisate 1,2-dioxygenase is the ring-cleavage enzyme involved in catabolism of salicylate and 3-hydroxybenzoate, among other aromatics (Fig. 1). In comparison to homogentisate 1,2-dioxygenases, gentisate 1,2-dioxygenases show a narrow distribution in bacterial families of proteobacteria being identified only in six α-, three β-, and three γ-proteobacterial families and being absent from δ- and ε-proteobacteria. The number of members with putative gentisate 1,2-dioxygenase genes inside the 12 proteobacterial families owing this enzyme is also significantly lower than the percentage of homogentisate 1,2-dioxygenase carrying members. Inside the *Comamonadaceae* however, six out of eight members harbor a gentisate 1,2-dioxygenase, but only one a homogentisate dioxygenase. Similarly, homogentisate dioxygenases are absent from the genomes of *Enterobacteriaceae*, although *Salmonella, Serratia*, and some *E. coli* strains are endowed with a gentisate dioxygenase. In addition to Proteobacteria, gentisate 1,2-dioxygenase genes can be found in *Corynebacteriaceae, Micrococcaceae, Mycobacteriaceae, Nocardiaceae*, and *Bacillaceae*.

3-Hydroxyanthranilate 3,4-dioxygenase catalyzes the conversion of 3-hydroxyanthranilate to 2-amino-3-carboxymuconic semialdehyde during tryptophan degradation via the kynurenine pathway. This extradiol dioxygenase is the cupin-type dioxygenase with the narrowest distribution since it is only found and with a low representativity in *Brucellaceae, Rhodobacteraceae, Sphingomonadaceae, Burkholderiaceae, Shewanellaceae, Xanthomonadaceae*, and *Myxococcaceae* in Proteobacteria and in *Flavobacteriaceae, Flexibacteraceae*, and *Bacillaceae* in non-proteobacterial families.

4.5 Other Extradiol Dioxygenases

Recently, a novel Fe^{2+}-dependent dioxygenase, benzoquinol 1,2-dioxygenase, which is a $\alpha_2\beta_2$ heterotetramer where the α- and β-subunits displayed no significant sequence identity with other dioxygenases and which catalyzes the ring fission of a wide range of benzoquinols to the corresponding 4-hydroxymuconic semialdehydes, has been described in *P. fluorescens* ACB (Moonen et al., 2008). Putative genes encoding both subunits of benzoquinol 1,2-dioxygenase show a highly narrow distribution since they are almost exclusively found in *Burkholderia* with the exceptions of *P. luminescens* subsp. *laumondii* TTO1 and *P. aeruginosa* PA7 strains, in spite to be originally identified in a 4-hydroxyacetophenone-degrading *P. fluorescens* strain (Moonen et al., 2008). The origin of this type of dioxygenase remains to be clarified.

4.6 Diiron Oxygenases

Soluble diiron oxygenases comprise an evolutionary-related family of enzymes capable to monooxygenate benzene/toluene to phenol/methylphenol and phenols to catechols (Leahy et al., 2003). Sequence comparisons of the respective α-subunits with the PaaA oxygenase subunit of phenylacetyl-CoA oxygenase and the BoxB oxygenase of benzoyl-CoA oxygenase strongly suggest that also these enzymes belong to the family of soluble diiron oxygenases.

Benzene/toluene monooxygenases and phenol monooxygenases of the soluble diiron oxygenase family are enzyme complexes including an electron transport system comprising

a reductase (and, in some cases, a ferredoxin), a catalytic effector and a terminal heteromultimeric oxygenase composed by α, β, and γ subunits whose α-subunits are assumed to be the site of substrate hydroxylation (Leahy et al., 2003). According to the presence of genes putatively coding for α subunit, benzene/toluene multicomponent monooxygenase are found almost exclusively in β-Proteobacteria, including *Burkholderia, Cupriavidus, Ralstonia, Methylibium*, and *Dechloromonas* strains with the only exceptions of *Bradyrhizobium* sp. BTAi1 and *Frankia* sp. CcI3. In the β-proteobacterial strains, the benzene/toluene multicomponent monooxygenase are associated with a phenol/methylphenol multicomponent monooxygenase. On the other hand, the phenol/methylphenol multicomponent monooxygenases showed a slightly broader distribution since in addition to the above mentioned strains, such genes are also identified in *Acidovorax* and *Verminephrobacter* strains and even in γ-proteobacterial families such as *Alteromonadaceae* and *Pseudomonadaceae*.

In contrast to the limited distribution of the above described multicomponent monooxygenases, multicomponent phenylacetyl-CoA oxygenases are broadly distributed in Proteobacteria being identified in 6 out of 19 α-, 5 out of 10 β-, and 8 out of 29 γ-proteobacterial families. They are, however, absent from δ- and ε-proteobacteria. The families *Rhodobacteraceae, Bradyrhizobiaceae, Alcaligenaceae, Burkholderiaceae, Rhodocyclaceae, Enterobacteriaceae*, and *Pseudomonadaceae* include a significant number of strains with such genes. Several representatives are also found in non-proteobacterial families, predominantly Actinobacteria such as *Streptomycetaceae, Pseudonocardiaceae, Nocardiaceae, Micrococcaceae, Corynebacteriaceae, Brevibacteriaceae*, and *Acidothermaceae*, and also in *Flavobacteriaceae* and *Bacillaceae* families.

Benzoyl-CoA oxygenase encoding genes are exclusively found in some families of the α- and β-proteobacteria: *Bradyrhizobiaceae, Rhodospirillaceae, Comamonadaceae, Burkholderiaceae*, and *Rhodocyclaceae*, and predominantly in the last two families in which the pathway was also originally described (Denef et al., 2004; Zaar et al., 2004).

4.7 Flavoprotein Monooxygenases

Flavoprotein monooxygenases are involved in a wide variety of biological processes including biosynthesis of antibiotics and siderophores or biodegradation of aromatics. They have been classified according to sequence and structural data in six classes (van Berkel et al., 2006), with classes A, D, and F being of special importance for aromatic degradation. Class A enzymes are considered to be widely distributed in different bacterial taxa and typically *ortho-* or *para*-hydroxylate aromatic compounds that contain an activating hydroxyl- or amino-group (van Berkel et al., 2006). In fact, it is interesting to note that according to genome annotations, a huge set of bacteria contain enzymes capable of 4-hydroxybenzoate 3-hydroxylation, salicylate 1-hydroxylation or 2,4-dichlorophenol 6-hydroxylation. Regarding the fact that the capability to mineralize chloroaromatics is not widespread in bacteria and chlorocatechol genes, usually necessary to achieve mineralization of chloroaromatics are, among the sequenced genomes only observed in the two bacteria well studied for such capability, i.e., *B. xenovorans* LB400 (Chain et al., 2006) and *C. necator* JMP134 (Pérez-Pantoja et al., 2008), the annotated widespread of enzymes involved in dichlorophenol degradation is astonishing. A phylogenetic analysis of proteins related to enzymes of class A flavoproteins using proteins of documented function (salicylate 1-hydroxylases, 3-hydroxybenzoate 4-hydroxylases, 2-aminobenzoyl-CoA monooxygenases/reductases, 4-hydroxybenzoate 3-hydroxylases, among others) as seeds show that these oxygenases can be grouped into six distinct protein clusters (enzymes related to

UbiH involved in ubiquinone biosynthesis will not be discussed here). Only one of these clusters comprises enzymes, which, based on characterized representatives, can be assumed to catalyze a single defined activity, i.e., the 3-hydroxylation of 4-hydroxybenzoate. As with the majority of aromatic degradative properties, the respective enzymes are predominantly observed in Actinobacteria and Proteobacteria. However, they are also observed in one of two Acidobacteria, in *Pedobacter* of the Bacteroidetes, in one *Deinococcus* and in 1 of 28 *Bacillaceae*. No other monocomponent flavoprotein monooxygenases discussed in this section are observed in these orders. Among the Actinobacteria, 4-hydroxybenzoate 3-hydroxylases are observed in roughly one third of the families, including *Arthrobacter* and *Streptomyces*, but interestingly were absent from any of the 17 *Mycobacterium* analyzed. It is a dominant trait in α-Proteobacteria, specifically in *Bradyrhizobiaceae* and *Rhodobacteraceae*. Also among β-Proteobacteria, all 34 *Burkholderia*, three *Cupriavidus*, four *Ralstonia*, and six out of eight *Comamonadaceae* are endowed with such capability. In contrast, such activity is rare in γ-Proteobacteria with the exception of *Pseudomonadaceae*, where 17 out of 18 strains (exception again *P. mendocina*) have a 4-hydroxybenzoate 3-hydroxylase. Similarly, such activity is spread among *Acinetobacter* and *Xanthomonas* strains. Among the *Enterobacteriaceae*, only *Klebsiella pneumoniae* and *Serratia proteomaculans* have a 4-hydroxybenzoate 3-hydroxylase.

Also the aminobenzoyl-CoA pathway (Altenschmidt and Fuchs, 1992) seems to be strongly represented among the thus far sequenced bacteria. In a phylogenetic analysis, the aminobenzoyl-CoA oxygenases seem to be related to salicylyl-CoA 5-hydroxylase from *Streptomyces* sp. WA46 (Ishiyama et al., 2004) channeling salicylate to gentisate. However, in contrast to the organization in strain WA46 where the oxygenase encoding gene is clustered with a gentisate dioxygenase, function as a salicylyl-CoA 5-hydroxylase can be suggested only in a few cases, such as in *S. wittichii* RW1, since a gentisate pathway is absent from the genomes of various strains including the two *Streptomyces* strains sequenced. Overall, homologues to aminobenzoyl-CoA oxygenases are observed in 44 genomes comprising Actinobacteria (five genomes) such as *Streptomyces* or *Saccharopolyspora erythraea* NRRL 2338. In Proteobacteria this pathway is absent in γ-Proteobacteria, but it is observed in *Plesiocystis pacifica* SIR-1 (a δ-proteobacterium). The pathway is abundant in β-Proteobacteria such as *Azoarcus* strains, where this metabolic route was initially established (Altenschmidt and Fuchs, 1992), but also in *Comamonadaceae* (six of eight genomes), *Ralstonia* (all four genomes), *Cupriavidus* (all three genomes), and α-proteobacteria such as *Bradyrhizobium* strains (all three genomes) or *Rhodobacteraceae* (11 of 24 genomes).

A large number of genes in bacterial genomes (nearly 100) are annotated as encoding salicylate 1-hydroxylases. However, a phylogenetic analysis taking into account validated salicylate 1-hydroxylases, identified only two of such proteins (amino acid sequence identity >40% to validated NahG proteins [Yen and Gunsalus, 1982]) encoded in the genome of *A. baylyi* ADP1 (as previously described [Jones et al., 2000]) and *P. putida* GB-1 (see *Fig. 5*, cluster 1). Also enzymes related to NahW, a second evolutionary lineage of salicylate 1-hydroxylases (Bosch et al., 1999b) are scarce and only seven homologues (four of them encoded by *Burkholderia* genomes) are identified (sequence identity >35%) (see *Fig. 5*, cluster 5). In contrast, various enzymes (observed in 22 genomes) clustered with enzymes of proven function as 3-hydroxybenzoate 6-hydroxylases (*Fig. 5*, cluster 10) and were observed, among others, in three *Corynebacteria*, two *Arthrobacter,* seven *Burkholderiaceae*, and three *Comamonadaceae* strains. Other enzymes annotated as salicylate hydroxylases (16) show high similarity (>60% identity) and cluster together with 6-hydroxynicotinate 3-monooxygenase of *P. fluorescens* TN5 (Nakano et al., 1999) such that their function as salicylate hydroxylases is

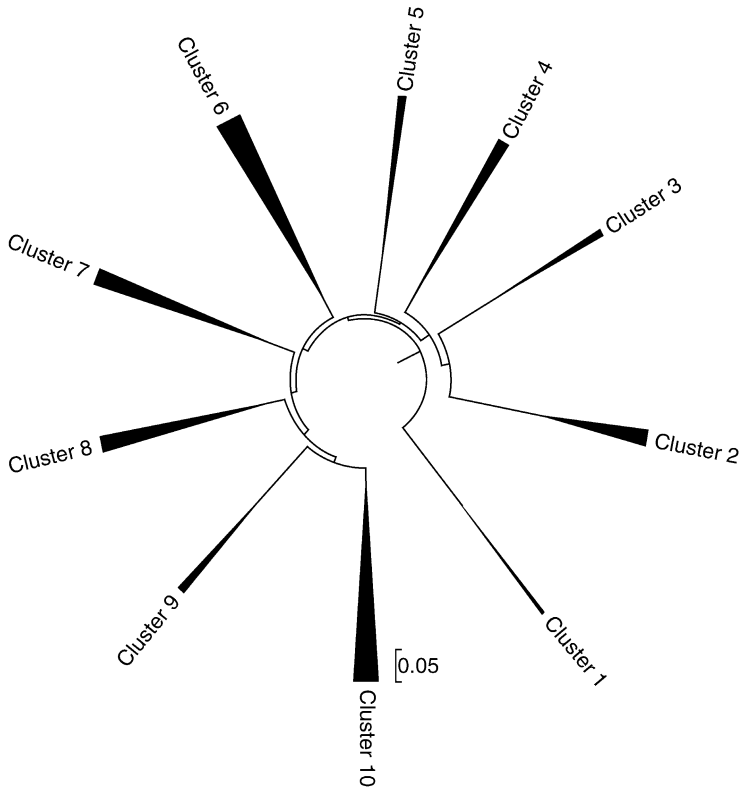

Figure 5

Evolutionary relationships among proteins related to NahG, or NahW-type salicylate 1-hydroxylases and 3-hydroxybenzoate 6-hydroxylases. Clusters 1 and 5 comprise salicylate 1-hydroxylases related to NahG or NahW salicylate 1-hydroxylases, cluster 10 3-hydroxybenzoate 6-hydroxylases, and cluster 2 enzymes related to 6-hydroxynicotinate 3-monooxygenase of *Pseudomonas fluorescens* TN5. The function of enzymes of other clusters remains to be elucidated.

questionable (*Fig. 5*, cluster 2). The same holds true for a further more than 100 additional sequences, out of which 69 (*Fig. 5*, cluster 6–9) are, among enzymes with validated function, phylogenetically most closely related to 3-hydroxybenzoate 6-hydroxylases. However their genomic contexts indicate different functions.

A similar situation holds for enzymes annotated as 3-hydroxyphenylpropionate monooxygenases. An overall of 24 proteins showed significant similarity (>40% identity) with respective validated enzymes and, in phylogenetic analysis, clustered together in one evolutionary branch. These enzymes are predominantly observed in *Mycobacterium* (seven genomes) and *Enterobacteriaceae* (mainly *E. coli*, 11 genomes, but also in *K. pneumoniae* and *S. sonnei*), as well as in *B. vietnamiensis, B. xenovorans, C. necator* JMP134, and *P. putida* W619. Other enzymes annotated as 3-hydroxyphenylpropionate monooxygenases show significant similarity to either resorcinol monooxygenase of *C. glutamicum* (Huang et al., 2006) or to GdmM involved in formation of the geldanamycin benzoquinoid system by *S. hygroscopicus*

AM 3672 (Rascher et al., 2005) and are thus highly improbable to function as 3-hydroxyphenylpropionate monooxygenase.

A 3-hydroxyphenylacetate 6-hydroxylase forming homogentisate has been recently described in *P. putida* U being composed of the hydroxylase and a small coupling protein, constituting a novel type of two-component hydroxylase, distinct from the classical two-component flavoprotein monooxygenases (Arias-Barrau et al., 2005). Seventeen homologues (>40% sequence identity, clustering on the same phylogenetic branch) are observed in 16 of the so far sequenced genomes and usually two subsequent genes encoding for the coupling protein and the monooxygenase can be identified. Interestingly, in contrast to the first and thus far only observation in *Pseudomonas*, such genes are absent from all 17 sequenced *Pseudomonas* strains and all other γ-proteobacterial genomes but frequently found in *Burkholderia* (5 of 34 genomes), *Cupriavidus* (two of three genomes), and *Comamonadaceae* (four out of eight genomes).

Also, various flavoprotein monooxygenases are annotated as 2,4-dichlorophenol hydroxylases. However, enzymes related to valid 2,4-dichlorophenol hydroxylases (>40% sequence identity) also comprise phenol hydroxylases such as PheA from *Pseudomonas* sp. strain EST1001, which transforms phenol and 3-methylphenol, but not 2,4-dichlorophenol (Nurk et al., 1991), ChqA chlorobenzoquinol monooxygenase of *Pimelobacter simplex* (AY822041), HpbA 2-hydroxybiphenyl-3-monooxygenase from *P. azelaica* HBP1, which is capable of oxidizing various 2-substituted phenols, but not phenol (Suske et al., 1997), OhpB 3-(2-hydroxyphenyl)propionic acid monooxygenase from *R. aetherivorans* I24 (DQ677338) and MhqA methylbenzoquinol monooxygenase from *Burkholderia* NF100 (Tago et al., 2005). Thus, enzymes of this group typically share the capability to transform 2-substituted phenols, but are obviously recruited for different metabolic routes and involve pathways where the ring-cleavage substrate is a dihydroxylated compound, but also routes where the ring-cleavage substrate is trihydroxylated. The function of these proteins, therefore, cannot be deduced from similarity measures or from phylogenetic analysis. An overall of 18 proteins can be identified as belonging to this cluster, and beside the two characterized 2,4-dichlorophenol hydroxylases from *C. necator* JMP134 only two genomes (*Rhizobium leguminosarum* and *Bradyrhizobium* sp. ORS278) comprise proteins clustering with 2,4-dichlorophenol hydroxylases. However, the genetic environment of the encoding genes does not give a direct support for such a function. Further proteins of this cluster are observed to be scattered among Actinobacteria and Proteobacteria with *R. jostii* RHA1 encoding for three of such proteins.

Interestingly, a distinct group of flavoprotein monooxygenases exhibiting approximately 30% of sequence identity to the above described monooxygenases is also typically annotated as phenol hydroxylases. This annotation seems to be due to some similarity to the phenol hydroxylase (30–35% identity) of *Trichosporon cutaneum* (Enroth et al., 1994), however, phylogenetic analysis shows that a set of 29 proteins (typically with identities >50%) is most closely related to proteins of validated function as 3-hydroxybenzoate 4-hydroxylases, previously assumed to be restricted to *Comamonas* strains (Hiromoto et al., 2006). In fact, inside the β-proteobacteria such genes are only observed in *C. testosteroni* and *B. phymatum*, however, also three γ-Proteobacteria harbor such gene, and 3-hydroxybenzoate-4-hydroxylases seem to be frequently encoded in the genome of α-Proteobacteria (12 genomes), specifically in *Bradyrhizobium* strains (all three genomes) and *Rhodobacteraceae* (6 out of 24 genomes). Also seven Actinobacteria seem to harbor such activity (among them two *Corynebacterium* species and both sequenced *Arthrobacter* strains), indicating this activity to be more widespread than previously thought.

Nearly 20 enzymes were annotated as pentachlorophenol monooxygenases, an activity previously reported, for example, in *Sphingobium chlorophenolicum* (Cai and Xun, 2002). However, none of these proteins showed sequence identities >35% to validated PcpB proteins, and only a group of enzymes typically encoded in *Burkholderia* genomes could be shown to be evolutionary related, however, their function as PCP monooxygenases seems highly improbable.

Styrene monooxygenases (StyA) have been identified in various *Pseudomonas* strains (Beltrametti et al., 1997), and were classified as Class E flavoprotein monooxygenases, however, they are evolutionary related to the Class A flavoprotein monooxygenases (van Berkel et al., 2006). Interestingly, none of the sequenced *Pseudomonas* strains harbor such a gene. Eight phylogenetically related proteins are observed in genome sequencing projects, however their function as such monooxygenases remains speculative.

Two-component aromatic hydroxylases such as 4-hydroxyphenylacetate 3-hydroxylases from *E. coli* (Diaz et al., 2001) consisting of an oxidoreductase and an oxygenase were classified as type D flavoprotein monooxygenases (Ballou et al., 2005) and have no structural or sequence similarities to the single-component enzymes described above. Iterative Psi-blast searches identified nearly 100 of such enzymes putatively involved in aromatic metabolism to be encoded in sequenced genomes and phylogenetic analysis indicated the presence of eight evolutionary lines (see *Fig. 6*).

Two of the branches contain the proteobacterial (*Fig. 6*, cluster 1) and non-proteobacterial (*Fig. 6*, cluster 7) 4-hydroxyphenylacetate 3-hydroxylases with an identity of members of the different cluster of approximately 30%. Proteins located on the same phylogenetic branch as validated 4-hydroxyphenylacetate 3-hydroxylases from *Thermus* or *Geobacillus* (Hawumba et al., 2007; Kim et al., 2007) are observed in only three Actinobacteria, but in both sequenced *Deinococci* and in both *Thermus* strains. It is also a dominant trait in *Bacillaceae* (13 out of 28 genomes).

Among the Proteobacteria, 4-hydroxyphenylacetate 3-hydroxylation by enzymes of this cluster is a trait nearly exclusively observed in γ-proteobacteria, predominantly in *Enterobacteriaceae* (19 out of 61 genomes) and *Pseudomonas* (5 out of 18 genomes), and outside of this group only in two α-proteobacteria.

The cluster of proteins most closely related to these proteobacterial 4-hydroxyphenylacetate 3-hydroxylases (50–60% identity) comprises those with high similarity to phenol hydroxylase PheA of *Geobacillus thermoleovorans* (Duffner and Muller, 1998), *R. erythropolis* (CAJ01325), 4-nitrophenol hydroxylase of *Rhodococcus* sp. PN1 (Takeo et al., 2003), an enzyme which also acts as a phenol hydroxylase, and 4-coumarate 3-hydroxylase of *Saccharothrix espanaensis* involved in the formation of caffeic acid (Takeo et al., 2003) (see *Fig. 6*, cluster 2). Interestingly, respective genes are practically absent from proteobacteria and only observed in *Photorhabdus* and *Saggitula*, but observed in one of the two *Thermus* strains sequenced, in all *Chloroflexaceae* and in some Actinobacteria such as *R. jostii* RHA1, which harbors four homologues.

A further group of proteins show similarity to PvcC, previously assumed to be involved in pyoverdin synthesis, but recently shown to be involved in the formation of pseudoverdine and paerucumarin by *P. aeruginosa* (Takeo et al., 2003) (*Fig. 6*, cluster 3). Interestingly, respective genes and gene clusters are exclusively observed in *P. aeruginosa, B. mallei, B. pseudomallei*, and *B. thailandensis*.

A further cluster of six proteins, also typically annotated as 4-hydroxyphenylacetate 3-hydroxylases is related to TcpA 2,4,6-trichlorophenol monooxygenases of *C. necator*

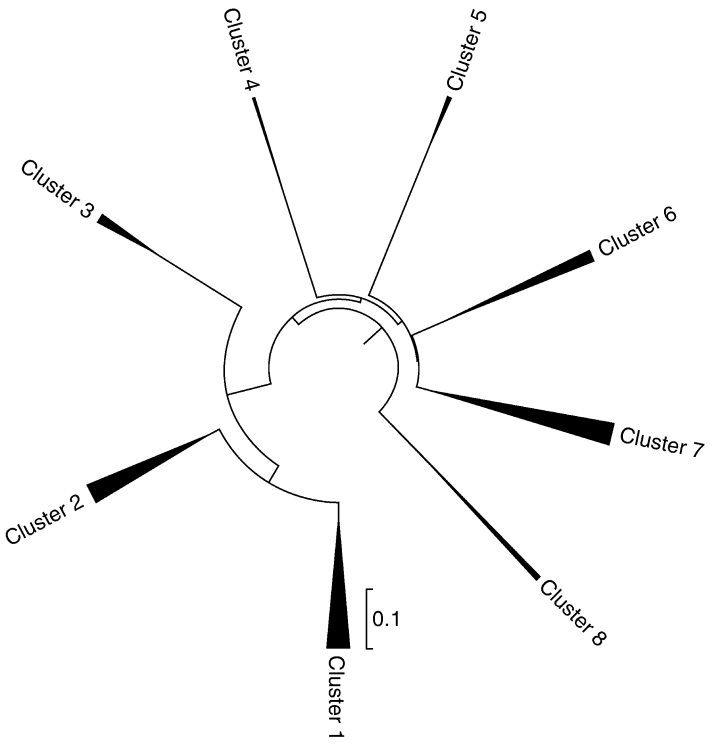

Figure 6

Evolutionary relationships among the large subunits of two-component flavoprotein monooxygenases related to 4-hydroxyphenylacetate 3-hydroxylase from *Escherichia coli*. Clusters 1 and 7 comprise 4-hydroxyphenylacetate 3-hydroxylases of proteobacteria and non-proteobacteria, cluster 2 proteins related with PheA phenol hydroxylase of *Geobacillus thermoleovorans*, and cluster 3 proteins with similarity to PvcC of *P. aeruginosa* (Takeo et al., 2003) (*Fig. 6*, cluster 3). The function of enzymes of other clusters remains to be elucidated.

JMP134 (Sanchez and Gonzalez, 2007), however, the function of these proteins also remains to be elucidated (*Fig. 6*, cluster 8).

A different type of two-component aromatic hydroxylases consisting also of a reductase and an oxygenase has been described recently (Thotsaporn et al., 2004). This type has been also classified as type D flavoprotein monooxygenases (Ballou et al., 2005) but it is able to use FMN, FAD, and riboflavin for hydroxylation in contrast to HpaB, PheA, and TcpA, which specifically uses only reduced FAD (Thotsaporn et al., 2004). The best studied representative of this group is 4-hydroxyphenylacetate 3-hydroxylase from *A. baumannii* but it shows very low identity with the 4-hydroxyphenylacetate 3-hydroxylases described previously in *E. coli*, *P. aeruginosa*, or *T. thermophilum* (Thotsaporn et al., 2004). Although the different types of 4-hydroxyphenylacetate 3-hydroxylase catalyze the same reaction, they have significant differences in the details of the mechanisms involved (Ballou et al., 2005). Genes putatively coding for enzymes similar to the *A. baumannii*-type of 4-hydroxyphenylacetate 3-hydroxylase are found in some strains of α- and γ-proteobacteria: *S. stellata*, *R. sphaeroides*, *Marinomonas* sp., *V. shilonii*, *V. vulnificus*, *A.vinelandii*, *P. entomophila*, and one *P. putida* strain. Additional

enzymes of this kind of two-component aromatic hydroxylases includes naphthoate 2-hydroxylase (NmoAB) described in *Burkholderia* sp. JT1500 (Deng et al., 2007) with homologous genes in some *Bradyrhizobium* and *Cupriavidus* strains and resorcinol hydroxylase from *Rhizobium* sp. MTP-10005 (GraAD) (Yoshida et al., 2007) with homologous genes in the related strains *A. tumefaciens* and *R. leguminosarum* and in the β-proteobacterium *Polaromonas* sp. JS666.

4.8 Rieske Non-Heme Iron Oxygenases

The so-called Rieske non-heme iron oxygenases are one of the key families of enzymes important for aerobic activation and thus degradation of aromatics such as benzoate, benzene, toluene, phthalate, naphthalene, or biphenyl (*Fig. 1*) (Gibson and Parales, 2000). Members of this family also catalyze monooxygenations, such as salicylate 1- or salicylate 5-hydroxylases or demethylations, such as vanillate O-demethylases. They are multicomponent enzyme complexes consisting of a terminal oxygenase component (iron–sulfur protein [ISP]) and electron transport proteins (a ferredoxin and a reductase or a combined ferredoxin-NADH-reductase). The catalytic ISPs are usually heteromultimers composed of a large α-subunit containing a Rieske-type [2Fe-2S] cluster, with a mononuclear nonheme iron oxygen activation center, and a substrate-binding site modulating substrate specificity and a small β-subunit, however, some enzymes, such as phthalate 4,5-dioxygenases contain an oxygenase composed only of α-subunits.

Phylogenetic analyses of Rieske non-heme iron oxygenases show that sequences obtained in our searches can be grouped into three main divergent clusters or divisions, where only two of them comprise proteins of validated function and are thus discussed here. One of these two divisions comprises the so-called phthalate family including vanillate demethylases (Gibson and Parales, 2000). Four clusters of this division contain oxygenases of proven function to dioxygenate aromatics, i.e., phthalate 4,5-dioxygenases (Nomura et al., 1992), isophthalate dioxygenase (Wang et al., 1995), phenoxybenzoate dioxygenase (Dehmel et al., 1995), and carbazol dioxygenase (Sato et al., 1997b). Genes putatively encoding phthalate 4,5-dioxygenases are nearly exclusively observed in β-proteobacteria (seven genomes) except for an amazing five homologues possibly encoded in the genome of *Rhodobacterales bacterium* HTCC2654. Similarly, genes putatively encoding isophthalate dioxygenases are predominantly observed in β-proteobacterial genomes (overall in five), but also in one γ-proteobacterium and in two α-proteobacteria, among them strain HTCC2654. A similar spread is observed for enzymes related to phenoxybenzoate dioxygenase (observed in seven β-, four α-, and one γ-Proteobacterium). Genes putatively encoding carbazol dioxygenases are not observed in any sequenced genome.

Most of the currently characterized Rieske non-heme iron oxygenases are concentrated in a well-defined division (see *Fig. 7*). The significant amount of validly described enzymes allows assignment of putative functions to most of the respective enzymes encoded in sequenced genomes.

Benzoate dioxygenases (cluster A1) are most widely distributed and can be observed in the genomes of Actinobacteria as well as α-, β-, and γ-proteobacteria. Most importantly, such enzymes are observed in 32 out of 34 *Burkholderia* strains, 14 out of 18 *Pseudomonas* strains, and 4 out of 17 Mycobacteria. Anthranilate can be transformed either by two-component anthranilate dioxygenases such as the one described from *Acinetobacter baylyi* ADP1

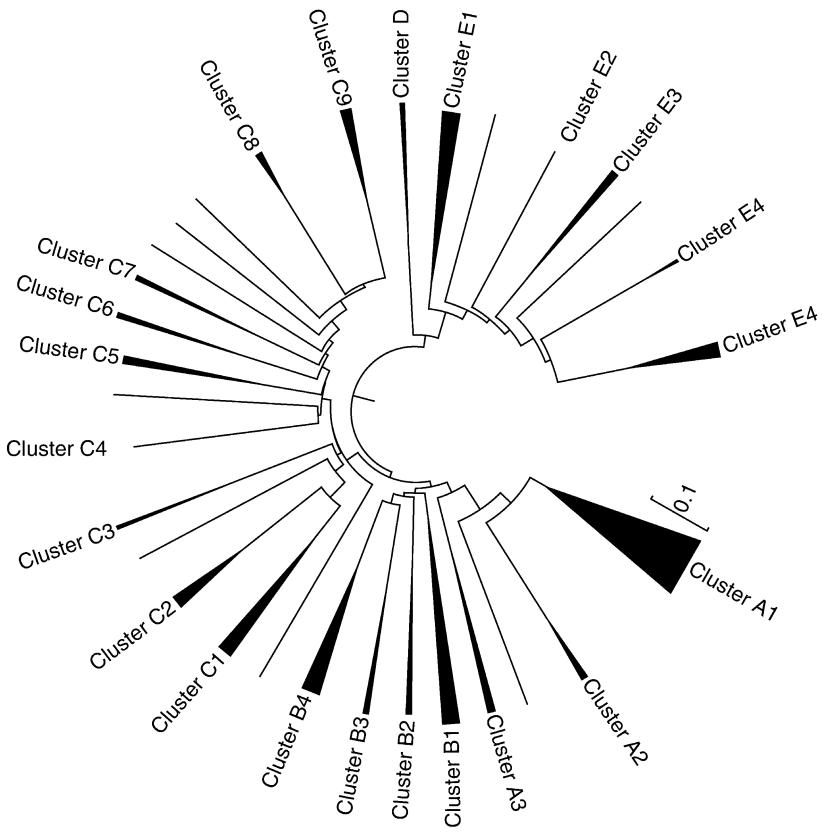

Figure 7

Evolutionary relationships among the α-subunits of Rieske non-heme iron oxygenases excluding phthalate family enzymes. A function can be assigned to proteins of some of the clusters shown as follows: cluster A1, benzoate dioxygenases; cluster A2, two component anthranilate dioxygenases; cluster A3, proteins related with *p*-cumate dioxygenases; cluster B3, aniline dioxygenases; cluster C1, NidA-type dioxygenases; cluster C2, phthalate 3,4-dioxygenases; cluster C3, proteins related with diterpenoid dioxygenases; cluster C5, NahA-type naphthalene dioxygenases; cluster 6, proteins related with ethylbenzene dioxygenase from *R. jostii* RHA1; cluster C8, 3-phenylpropionate dioxygenases; cluster C9, benzene/toluene/isopropylbenzene/biphenyl dioxygenases; cluster E1, salicylate 5-hydroxylases; cluster E2, 2-chlorobenzoate dioxygenases; cluster E3, terephthalate dioxygenases; cluster E4, salicylate 1-hydroxylases; and cluster E5, three component anthranilate dioxygenases. The function of enzymes of other clusters remains to be elucidated.

(Eby et al., 2001) (cluster A2) or by three-component anthranilate dioxygenase as the one from *Burkholderia cepacia* DBO1 (Chang et al., 2003) (cluster E5). Genome analysis clearly showed that two-component dioxygenases are obviously restricted to γ-proteobacteria and are only observed in seven *Pseudomonas* genomes and, as described, in *A. baylyi*. In contrast, three-component anthranilate dioxygenases are exclusively observed in *Burkholderia* genomes and present in 31 out of 34 sequenced strains. Cluster A3 comprises proteins phylogenetically

related with known *p*-cumate dioxygenases. These sequence relatives are found in five of the sequenced *Pseudomonas* genomes but also in *S. wittichi* RW1 and *B. xenovorans* LB400.

Cluster B3 comprises proteins similar to aniline dioxygenases, and similar sequences are found only in *Nocardioides* sp. JS614 and *Bradyrhizobium* sp. BTAi1, indicating a very restricted distribution of such activity. Further related sequences, where no specific function can be postulated (clusters B1, B2, and B4) were predominantly observed in *Burkholderiaceae*.

Proteins of cluster C1 exhibit similarity to proteins involved in the degradation of polycylic aromatics by Actinobacteria, exemplified by NidA of *M. vanbaalenii* PYR-1 (Stingley et al., 2004a) and thus putatively have a function in degradation of polycyclic aromatics. In accordance with this assumption, respective proteins are found to be encoded in the genomes of five environmental Mycobacteria and up to four different such proteins are observed per genome. As NidA-like proteins, also sequences putatively encoding phthalate 3,4-dioxygenases (Stingley et al., 2004b) (cluster C2) are exclusively observed in Actinobacteria, differentiating them from β-proteobacteria which obviously degrade phthalate by phthalate 4,5-dioxygenases. Phthalate 3,4-dioxygenases were observed to be encoded in genomes of Mycobacteria comprising a NidA sequence, but also in *M. avium* strains, *R. jostii* RHA1, and *Arthrobacter* sp. FB24.

Group C3 proteins, comprising diterpenoid dioxygenases-like proteins (Martin and Mohn, 1999) are having a very restricted distribution in the genomes available so far, being found only in *Caulobacter* sp. K31, *Sphingomonas* sp. SKA58, *S. wittichii* RW1, and *B. xenovorans* LB400 genomes (Smith et al., 2007).

Naphthalene and phenanthrene dioxygenases related to NahA of *P. stutzeri* AN10 (Bosch et al., 1999a) have previously been observed in various *Pseudomonas, Sphingomonas, Burkholderia, Cycloclasticus, Acidovorax*, and *Ralstonia* isolates. The genomic survey indicates such activities (see cluster C5) not to be widespread and similar sequences are only observed in genomes of *N. aromaticivorans* DSM 12444, *Acidovorax* sp. JS42, and *P. naphthalenivorans* CJ2. Also sequences related to ethylbenzene dioxygenase from strain RHA1 (Iwasaki et al., 2006) (cluster C6) are additionally observed only in of *Azotobacter vinelandii* AvOP and *N. aromaticivorans* DSM 12444.

Sequences indicating to encode 3-phenylpropionate dioxygenases (cluster C8) are exclusively observed in *Enterobacteriaceae,* and interestingly observed in all *Shigella* spp. strains (seven genomes) and 11 of 17 *E. coli*.

Cluster C9 is composed of benzene/toluene/isopropylbenzene/biphenyl dioxygenases (Witzig et al., 2006), enzymes typically involved in the degradation of the respective compounds, where a broad set of both proteobacterial and actinobacterial isolates is available. Respective sequences are only observed in the four genomes of strains previously reported to harbor such activity (*P. putida* F1, *B. xenovorans* LB400, *P. napthalenivorans* CJ2, and *R. jostii* RHA1).

Cluster E comprises enzymes acting on *ortho*- or *para*-substituted benzoates and include salicylate 5-hydroxylases (Fuenmayor et al., 1998) (cluster E1), salicylate 1-hydroxylases (Pinyakong et al., 2003) (cluster E4), 2-chlorobenzoate dioxygenases (cluster E2), three-component anthranilate dioxygenases (cluster E5, see above), and terephthalate dioxygenases (Sasoh et al., 2006) (cluster E3). Respective sequences are nearly exclusively observed in β-proteobacteria and in Sphingomonads out of the α-proteobacteria and only terephthalate dioxygenases are also observed in Actinobacteria, i.e., *R. jostii* RHA1 and *Arthrobacter aurescens* T1, which corresponds with various reports of *Rhodococci* being capable of degrading terephthalate. Terephthalate dioxygenases are also observed in *B. xenovorans* LB400 and

C. testosteroni with members of last mentioned genus also often being implicated in terephthalate degradation (Sasoh et al., 2006).

Salicylate 5-hydroxylases were observed in two *Cupriavidus* strains, both *Polaromonas* strains, and both *R. solanacearum* isolates in accordance with such activity being first described from a *Ralstonia* strain (Fuenmayor et al., 1998). Also *S. wittichii* RW1 seems to harbor such activity. In contrast, a Rieske-type salicylate 1-hydroxylase was only observed in *N. aromaticivorans* DSM 12444, also in accordance with the fact that such activities so far have only been described in Sphingomonads. Also putative 2-chlorobenzoate 1,2-dioxygenases are rare and a putative homologue is only observed in the genome of *B. xenovorans* LB400.

5 Metabolism Diversity

A very exciting question can be addressed based on the phylogenomic analyses carried out here: What is the diversity of catabolic properties within phylogenetic groups? However, before answering such question, a definition about the "unit of catabolic diversity" must first be addressed. The first unit level is pathway diversity. It refers to the presence in one bacterium or bacterial group of different ways to degrade one compound (i.e., intradiol versus extradiol ring cleavage; classical aromatic ring oxidation versus a CoA-dependent pathway, etc.). This level of diversity is the thickest and provides the most powerful versatility because it allows the microorganism to choose among very different ways to metabolize the compound. The second level of "unit of diversity" is the enzymatic diversity. It refers to the same biochemical reaction or catabolic step carried out by completely different enzymes. For example, enzymes belonging to three different families can perform phenol conversion to catechol: single-component flavoprotein monoxygenases, diiron oxygenases, or two-component monooxygenases. This level of catabolic diversity is finer than the previous one, but still significant because it allows for versatility at the biochemical level, i.e., different substrate affinities, different cofactor requirements, inhibitor effects, among others. The third level of catabolic diversity is the genetic diversity, or classical gene redundancy: the same biochemical step may be performed by very similar enzymes encoded by different genes. It is assumed that the main point of diversity here is at the regulatory level.

Although a gross measure of catabolic versatility, in the following three sections the pathway diversity will be used as a diversity unit for aromatic catabolism properties of a taxonomic group. This is especially relevant to account for the diversity of central pathways as defined in *Table 2*.

5.1 Metabolism by Bacteria Outside the Actinobacterial and Proteobacterial Phyla

When the genome database is searched for the aromatic catabolic pathways listed in *Table 2*, using the corresponding representative gene sequences, an unequal distribution of these markers among phyla and genera is easily noticed. Only members of 8 out of 17 phyla where representatives have been sequenced show the presence of the catabolic gene markers described above. However, it should be also noted that among the phyla showing absence of aromatic catabolic pathway markers, often only a few representatives have been sequenced, such as one Verrucomicrobia, two Aquificae, Fusobacteria, or Lentispharea strains, three

Table 2

Key groups of catabolic enzymes discussed in the metabolic diversity section

Enzyme group	Family	Pathway marker	Enzyme function	Abbreviation
Protocatechuate 3,4-dioxygenase	Intradiol dioxygenase	++	Intradiol cleavage	Pca34
Catechol 1,2-dioxygenase	Intradiol dioxygenase	++	Intradiol cleavage	Cat12
Hydroxybenzoquinol 1,2-dioxygenase	Intradiol dioxygenase	++	Intradiol cleavage	Hqu
Chlorocatechol 1,2-dioxygenase	Intradiol dioxygenase	+	Intradiol cleavage	Cca
Catechol 2,3-dioxygenase	Type I extradiol dioxygenase	++	Extradiol cleavage	Cat23
2,3-Dihydroxybiphenyl 1,2-dioxygenase	Type I extradiol dioxygenase	#	Extradiol cleavage	Dhb
Homoprotocatechuate 2,3-dioxygenase	Type I extradiol dioxygenase	++	Extradiol cleavage	Hpc_{EXDOI}
Protocatechuate 4,5-dioxygenase	LigB-type dioxygenase	++	Extradiol cleavage	Pca45
Gallate 4,5-dioxygenase	LigB-type dioxygenase	++	Extradiol cleavage	Gal
Homoprotocatechuate 2,3-dioxygenase	LigB-type dioxygenase	++	Extradiol cleavage	Hpc_{LigB}
2,3-Dihydroxyphenylpropionate 1,2-dioxygenase	LigB-type dioxygenase	++	Extradiol cleavage	Dhp
2-Aminophenol 1,6-dioxygenase	LigB-type dioxygenase	+	Extradiol cleavage	Amn
Gentisate 1,2-dioxygenase	Cupin superfamily dioxygenase	++	Extradiol cleavage	Gen
Homogentisate 1,2-dioxygenase	Cupin superfamily dioxygenase	++	Extradiol cleavage	Hge
3-Hydroxyanthranilate 3,4-dioxygenase	Cupin superfamily dioxygenase	++	Extradiol cleavage	Han
Benzoquinol 1,2-dioxygenase	Type IV extradiol dioxygenase	+	Extradiol cleavage	Bqu
Benzoyl-CoA oxygenase	Soluble diiron oxygenase	++	Dearomatization	Box
Phenylacetyl-CoA oxygenase	Soluble diiron oxygenase	++	Dearomatization	Paa
2-Aminobenzoyl-CoA monooxygenase/reductase	Class A flavoprotein monooxygenase	++	Dearomatization	Abc
4-Hydroxybenzoate 3-hydroxylase	Class A flavoprotein monooxygenase		Forming protocatechuate	Phb3H
3-Hydroxybenzoate 4-hydroxylase	Class A flavoprotein monooxygenase		Forming protocatechuate	Mhb4H

Table 2 (Continued)

Enzyme group	Family	Pathway marker	Enzyme function	Abbreviation
Salicylate 1-hydroxylase	Class A flavoprotein monooxygenase		Forming catechol	Ohb1H
3-Hydroxybenzoate 6-hydroxylase	Class A flavoprotein monooxygenase		Forming gentisate	Mhb6H
3-Hydroxyphenylpropionate 2-hydroxylase	Class A flavoprotein monooxygenase		Forming 2,3-dihydroxyphenylpropionate	Mhp2H
3-Hydroxyphenylacetate 6-hydroxylase	Class A flavoprotein monooxygenase		Forming homogentisate	Mha6H
Phenol/benzoquinol hydroxylase	Class A flavoprotein monooxygenase		Forming catechol/ hydroxybenzoquinol	Pbq2H
4-Hydroxyphenylacetate 3-hydroxylases	Class D flavoprotein monooxygenase		Forming homoprotocatechuate	Pha3H
Phenol 2-hydroxylase	Class D flavoprotein monooxygenase		Forming catechol	Ph2H
Chlorophenol 4-hydroxylase	Class D flavoprotein monooxygenase		Forming chlorobenzoquinol	Ph4H
4-Hydroxyphenylacetate 3-hydroxylases	Class D* flavoprotein monooxygenase		Forming homoprotocatechuate	Pha3H
Resorcinol 4-hydroxylase	Class D* flavoprotein monooxygenase		Forming hydroxybenzoquinol	Res4H
Terephthalate 1,2-dioxygenase	Rieske nonheme iron oxygenase		Channeling to protocatechuate	TphDO
Phthalate 3,4-dioxygenase	Rieske nonheme iron oxygenase		Channeling to protocatechuate	Pht34DO
Anthranilate 1,2-dioxygenase (2 component)	Rieske nonheme iron oxygenase		Channeling to catechol	AntDO
Anthranilate 1,2-dioxygenase (3 component)	Rieske nonheme iron oxygenase		Channeling to catechol	AntDO
Benzoate 1,2-dioxygenase	Rieske nonheme iron oxygenase		Channeling to catechol	BenDO
Salicylate 5-hydroxylase	Rieske nonheme iron oxygenase		Channeling to gentisate	Sal5H
Phenylpropionate 2,3-dioxygenase	Rieske nonheme iron oxygenase		Channeling to 2,3-dihydroxyphenylpropionate	PhpDO
Biphenyl 2,3-dioxygenase type	Rieske nonheme iron oxygenase		Activation of hydrophobic aromatics	BphDO
Naphthalene inducible dioxygenase (NidA) type	Rieske nonheme iron oxygenase		Polycyclic aromatic degradation	NidDO

Table 2 (Continued)

Enzyme group	Family	Pathway marker	Enzyme function	Abbreviation
Phthalate dioxygenase 4,5-dioxygenase	Rieske nonheme iron oxygenase		Channeling to protocatechuate	PhtDO
Isophthalate dioxygenase	Rieske nonheme iron oxygenase		Channeling to protocatechuate	IphDO
Toluene/benzene monooxygenase	Soluble diiron oxygenase		Forming phenol	Tmo
Phenol monooxygenase	Soluble diiron oxygenase		Forming catechol	Pmo

Only enzymes where a function could be assigned with high probability are included in the list
Eleven groups of aromatic ring-cleavage activities (homoprotocatechuate 2,3-dioxygenases, even though belonging to different enzyme families were defined as one activity) and all groups of enzymes catalyzing dearomatization of aromatic CoA derivatives were defined as abundant, as they are observed in more than ten sequenced genomes and are marked as ++
Three groups of aromatic ring-cleavage activities were defined as less abundant, as they were observed in ten or less sequenced genomes and are marked as +
2,3-Dihydroxybiphenyl 1,2-dioxygenases (marked #) are not included in the list of aromatic catabolic pathway markers discussed in the proteobacterial section, as they are assumed to have their function in the metabolism of bi- and polycyclic aromatics rather than monocyclic aromatics
Class D* flavoprotein monooxygenases refers to enzymes capable of using FMN, FAD, and riboflavin for hydroxylation

Planctomycetes, six Thermotogae, or nine Spirochaetes. Specifically in case the phylum contains aerobic species, only further genome analysis will reveal if such capabilities are in fact absent. Aromatic metabolic pathways were also absent from Chlamydiaea (11 genomes) where cultured representatives are obligate intracellular parasites of eukaryotic cells, the typically strict anaerobic Chlorobi (10 genomes), but also from Cyanobacteria (40 genomes), even though, for example, phenol degradation by the cyanobacterium *Phormidium valderianum* has been reported (Shashirekha et al., 1997).

Most of the catabolic markers analyzed here are exclusively observed in Proteobacteria and Actinobacteria. This may be due to the fact that an immense amount of work has been invested specifically on elucidation of aromatic degradation in easy to culture members of these phyla. It thus cannot be excluded that novel groups of catabolic enzymes will be identified from other phyla. However, members of certain catabolic gene families can be observed in some representatives of other genera, such that the genome survey performed here is valid to get a reasonable overview of metabolic properties also from other phyla. For example, members of the cupin family, i.e., gentisate 1,2-dioxygenase, homogentisate 1,2-dioxygenase, and 3-hydroxyanthranilate 3,4-dioxygenase are all observed in other phyla, with homogentisate 1,2-dioxygenase being observed in Bacteroidetes, Chloroflexi, and Firmicutes (Bacilli). Bacilli and Bacteroidetes were also indicated not only to encode gentisate 1,2-dioxygenase and 3-hydroxyanthranilate 3,4-dioxygenase, but also a phenylacetate degradative pathway. In contrast to ring-cleavage pathways mediated by members of the cupin family, pathways mediated by other extradiol dioxygenases or intradiol dioxygenases are scarce outside of the Actinobacterial and Proteobacterial phyla. Intradiol cleavage dioxygenases are observed in Acidobacteria and the Thermus/Deinococcus phylum, among the LigB-type extradiol dioxygenases only

homoprotocatechuate 2,3-dioxygenases is observed in *Bacilli* and out of EXDO I proteins only homoprotocatechuate 2,3-dioxygenase is observed in *Bacilli* and Thermus/Deinococcus. Exceptional is the detection of distinct EXDO I proteins in Chloroflexi.

Even though only two Acidobacteria and four Deinococcus/Thermus strains have been sequenced, the genomic survey indicates aromatic metabolic properties to be spread among those phyla. It can be suggested that *S. usitatus* Ellin6076 is capable to degrade 4-hydroxybenzoate via protocatechuate followed by intradiol cleavage and 4-hydroxyphenylpyruvate via the homogentisate pathway. Further capabilities of Acidobacteria thus remain to be discovered. All four members of the phylum Deinococcus/Thermus obviously share the capability to degrade 4-hydroxyphenylacetate via homoprotocatechuate and *D. geothermalis* DSM 11300 seems to harbor the capability to degrade 4-hydroxybenoate via protocatechuate and intradiol cleavage. Intradiol cleavage seems to be absent from Chloroflexi, Bacteroidetes, and Firmicutes. Interestingly, Chloroflexi can be proposed to be phenol degraders catabolizing it via catechol and *meta*-cleavage. Among Bacteroidetes, the homogentisate pathway and astonishingly the 3-hydroxyanthranilate pathway, in addition to the phenylacetate degradative pathway, seem to be spread among members of the orders *Flavobacteriales* and *Sphingobacteriales*. Out of the Firmicutes, only Bacillaceae (members of the genera *Bacillus, Exiguobacterium, Geobacillus*, and *Oceanobacillus* have been sequenced) seem to harbor aromatic metabolic properties. Unfortunately, no *Paenibacillus* genome sequence is available so far. *Bacillus* strains such as *Bacillus* sp. JF8 (Shimura et al., 1999), *B. subtilis* IS13 (Shimura et al., 1999), and others have been shown to be capable of degrading aromatics such as biphenyl, guaiacol, cinnamate, coumarate, or ferulate (Peng et al., 2003), and *Paenibacilli* such as *P. naphthalenovorans*, *Paenibacillus* sp. strain YK5, or *Paenibacillus* sp. KBC101 (Daane et al., 2002; Iida et al., 2006; Sakai et al., 2005) are shown to be capable of degrading naphthalene, dibenzofuran, or biphenyl. Thus, the metabolic diversity of *Bacillaceae* is clearly underrepresented by the currently sequenced 28 genomes, which indicate metabolic properties similar to those of Bacteroidetes, such as a spread of the homogentisate pathway in *Bacillus* and the presence of the 3-hydroxyanthranilate and the gentisate pathway in addition to the phenylacetate degradative pathway in members of different genera. In addition, 4-hydroxyphenylacetate degradation via homoprotocatechuate seems to be also a capability spread among *Bacillaceae*.

5.2 Actinobacteria

Aromatic metabolic routes can be observed in 12 out of 20 families from the phylum Actinobacteria and pathways analyzed here are absent in *Actinomycetaceae, Cellulomonadaceae, Kineosporiaceae, Microbacteriaceae, Nocardiopsaceae, Propionibacteriaceae, Bifidobacteriaceae*, and *Coriobacteriaceae*. Within the *Corynebacterium* genus, *C. diphteriae* and *C. jeijekum*, a nocosomial pathogen have no aromatic catabolic pathways. Interestingly, they have the smaller genomes of this group. A similar situation is observed within the *Mycobacteria*, as *M. leprae*, *M. bovis,* and *M. tuberculosum* also have no aromatic catabolic pathways and the smaller genomes of this group. In contrast, environmental Mycobacteria are characterized by an enormous metabolic potential, however, it should be noted that *M. vanbaalenii* Pyr1, *M. gilvum* PYR-GCK, as well as strains JLS, KMS, and MCS have been sequenced due to their capability to degrade various polycyclic aromatics reflected in the presence of up to four NidA-type Rieske non-heme iron oxygenases for initiating metabolism of PAHs and up to six BphC type I extradiol dioxygenases per genome.

However, not only *Mycobacteria* are endowed with a high metabolic potential. In contrast to members of all phyla described above, Actinobacteria not only often comprise a homogentisate pathway, which is observed in seven families, but also a protocatechuate intradiol cleavage pathway observed in eight families and more than one third of sequenced strains. Typically, actinobacterial strains endowed with a protocatechuate pathway also harbor a protocatechuate forming 4-hydroxybenzoate 3-hydroxylase such as both *Micrococcaceae* or *Streptomycetaceae* (see *Table 3*) and often a 3-hydroxybenzoate 4-hydroxylase (such as both *Micrococcaceae*), indicating protocatechuate to be a central intermediate of various metabolic routes. Interestingly, *Mycobacteria* harboring a protocatechuate intradiol cleavage do not contain any of the aforementioned genes, but typically a phthalate dioxygenase.

Table 3 shows an overview of catabolic markers observed at least twice in genomes of actinobacterial families, from which at least two genomes have been sequenced. Two observations are evident from the table. First, *Corynebacteriaceae, Nocardiaceae*, and specifically *Micrococcaceae* are endowed with a broad metabolic potential. However, it should be noted that among the three *Nocardiaceae*, *R. jostii* RHA1 has a metabolic potential much broader than *Nocardia farcinica* IFM 10152 or *Nocardioides* sp. JS614. Unfortunately, no more sequences of the reported highly versatile *Rhodococcus* genus (van der Geize and Dijkhuizen, 2004) are available thus far. Also various reports on the metabolic versatility of *Arthrobacter* strains are known (Nordin et al., 2005). In contrast, *Corynebacteria* just recently have become the focus of more intense metabolic investigations (Huang et al., 2006). Second, the table shows a clear cooccurrence of ring-cleavage activity markers as well as of markers for peripheral activities, supporting that our annotation efforts are appropriate to deduce metabolic potential.

5.3 Proteobacteria

Three of the five classes of Proteobacteria (α, β, and γ) concentrate the vast majority of the reported catabolic pathways towards aromatic compounds that can be traced in the current genome databases (*Table 4*). Only a couple of aromatic catabolic pathways (Pca34, Hge, and Han) are found in some strains of the *Myxococcales* order of δ proteobacteria and none in the ε proteobacterial class.

The **α class of Proteobacteria** has an uneven distribution of aromatic catabolic gene markers. None of the members of the three families of the order *Rickettsiales* have such catabolic properties. The small genome size of these members may be related to this trait. Aromatic ring-cleavage pathways are also absent from all members of the *Parvularculaceae, Bartonellacea*e, and *Erythrobacteraceae* families and some members of the *Aurantimonadaceae, Bradyrhizobiaceae, Methylobacteriaceae, Phyllobacteraceae, Rhodobacteraceae, Acetobacteracea, Rhodospirillaceae,* and *Sphingomonadaceae* families. In contrast, four α- proteobacterial strains (*Bradyrhizobium* sp. BTAi1, *S. wittichii* RW1, *Sagittula stellata* E-37, and *Silicibacter pomeroyii* DSS-3) have 8–9 out of the 14 main pathways and another three strains (*Bradyrhizobium japonicum* USDA110, *Bradyrhizobium* sp. ORS278, and *Jannaschia sp.* CCS1) have seven main aromatic catabolic pathways suggesting *Bradyrhizobium* strains to be metabolically highly versatile.

The most broadly distributed pathways in the α class of proteobacteria are Pca34 and Hge being observed in 30–40% of the sequenced genomes and in 11 and 9 families, respectively. Some catabolic pathways are only seldomly found in members of this proteobacterial class, and only *N. aromaticivorans* DSM 12444 has the Cat23 pathway and only *X. autotrophicus* Py2

Table 3

Catabolic gene markers of Proteobacteria

Actinobacterial Families	Pca34	Phb3H	Mhb4H	Pht34DO	TphDO	Cat12	BenDO	Gen	Mhb6H	Hge	HppDO	Mha6H	Dhp	Mhp2H	Hpc_{EXDOI}	Pha3H	NidDO	Dhb	Paa	Abc
Cellulomonadaceae (2)																				
Corynebacteriaceae (5)	++	++	+			++	++	++	++											+
Frankiaceae (3)										+	+		+							
Microbacteriaceae (2)																				
Micrococcaceae (2)	++	++	++	+	+	+		++	++						+	+				++
Micromonosporaceae (2)										++	++									+
Mycobacteriaceae (17)	+			+		+	+	+		+			+		(+)	(+)	+	++		
Nocardioidaceae (3)	+	+	+	+	+	++	+	+	+	+	++	+	+	+	+	+		++	+	++
Streptomycetaceae (2)	++	++									++								++	++
Bifidobacteriaceae (4)																				

++; More than 60% of the sequenced genomes of these proteobacterial families comprise a gene encoding the mentioned activity (number of sequenced representatives is given in parentheses); +, between 20 and 60%; (+), less than 20%. For abbreviations, see *Table 2*. Only families where at least two members have been sequenced are included in the analysis

Table 4

Catabolic gene markers of Actinobacteria

α Proteobacterial Families	Pca34	Pca45	Phb3H	Mhb4H	Pht34DO	IphDO	TphDO	Cat12	Cat23	Ohb1H	AntDO	BenDO	TMO	PMO	Ph2H	Pbq2H	Hqu	Res4H	Bqu	Gen	Mhb6H	Ohb5H	Hge	HppDO	Mha6H	Han	Dhp	Mhp2H	PhpDO	Gal	Hpc_{Lig8}	Hpc_{EXDOI}	Pha3H	Paa	Box	Abc
Caulobacteraceae (2)	++		++																	+			++	++												
Aurantimonadaceae (2)	+		+	+																																+
Bartonellaceae (3)																																				
Bradyrhizobiaceae (11)	(+)	+	++	+		(+)							(+)			(+)	++			+			++	+							+	+		++	(+)	+
Brucellaceae (6)	++		++																							(+)						+				
Methylobacteriaceae (3)	+		+	+				+				+											+													
Phyllobacteriaceae (3)	++		++																	+	+		++	++								+	+			+
Rhizobiaceae (5)	++		++													(+)	+	+					++	++										+		
Xanthobacteraceae (2)	++		++					+				+					++						+				+							+		+
Hyphomonadaceae (3)																							+	++												+
Rhodobacteraceae (24)	++		++	+				(+)				(+)			(+)		+			(+)	(+)		+	+	(+)	(+)					(+)	+	(+)	+	(+)	+
Acetobacteraceae (3)	+																																	+		
Rhodospirillaceae (3)			++																												++				++	
Anaplasmataceae (10)																																				
Rickettsiaceae (16)																																				
SAR11 (2)																																				
Erythrobacteraceae (3)																								++												
Sphingomonadaceae (5)	(+)	(+)	(+)					+	(+)			(+)				(+)	(+)			+			+	++	(+)	(+)					(+)			(+)		(+)
β Proteobacterial Families																																				
Alcaligenaceae (3)						+																	++	++							+			++		+
Burkholderiaceae (43)	++	(+)	++	(+)	(+)	(+)	(+)	++	(+)	(+)	++	++	(+)	(+)		(+)	+		(+)	+	(+)	(+)	++	++	(+)	(+)	(+)	(+)		(+)	++			++	(+)	+
Comamonadaceae (8)	(+)	++	++	(+)	+	(+)	(+)	+	+			(+)		(+)			+	(+)		++	+	+	(+)	++	+	(+)								+	++	++

Table 4 (Continued)

α Proteobacterial Families	Pca34	Pca45	Phb3H	Mhb4H	Pht34DO	IphDO	TphDO	Cat12	Cat23	Ohb1H	AntDO	BenDO	TMO	PMO	Ph2H	Pbq2H	Hqu	Res4H	Bqu	Gen	Mhb6H	Ohb5H	Hge	HppDO	Mha6H	Han	Dhp	Mhp2H	PhpDO	Gal	Hpc_{LigB}	Hpc_{EXDOI}	Pha3H	Paa	Box	Abc
Oxalobacteraceae (2)																																				
Neisseriaceae (5)																							(+)													
Nitrosomonadaceae (3)																																				
Rhodocyclaceae (3)		+	+						+			+	+	+		+				+	+	+		+			++							++	++	++
γ Proteobacterial Families																																				
Aeromonadaceae (7)	(+)					(+)		+	(+)			(+)		(+)									+	+										+		
Idiomarinaceae (2)																							++	++												
Pseudoalteromonadaceae (3)		+	+																				++	+												
Psychromonadaceae (2)																																		+		
Shewanellaceae (18)																							++			(+)								(+)		
Ectothiorhodospiraceae (3)																							+											+		
Enterobacteriaceae (61)	(+)		(+)					(+)		(+)		(+)			(+)				(+)	(+)							+	+	+	(+)	+		+	(+)		
Coxiellaceae (5)																																				
Legionellaceae (4)																							++	++												
Oceanospirillaceae (3)		+	+	+				+									+			+		+	++	++							+		+	++		
Pasteurellaceae (21)																															(+)		(+)			
Moraxellaceae (5)	+		+					+		(+)	(+)	+				(+)				(+)														(+)		
Pseudomonadaceae (19)	++		++					++	(+)		+	+		(+)			(+)		(+)	+	(+)		++	++			(+)	(+)		+	+		+	+		
Francisellaceae (7)																																				
Thiotrichaceae (2)																																				
Vibrionaceae (30)																							++	++							(+)		(+)			
Xanthomonadaceae (11)	+		+																				++	++		++				+						

++; More than 60% of the sequenced genomes of these Actinobacterial families comprise a gene encoding the mentioned activity (number of sequenced representatives is given in parentheses); +, between 20 and 60%; (+), less than 20%. For abbreviations, see *Table 2*. Only families where at least two members have been sequenced are included in the analysis

has the Dhp pathway. The Cca, Amn, and Bqu pathways are not found in any α proteobacterial genome.

Regarding peripheral pathways, α proteobacterial strains endowed with a protocatechuate pathway also harbor a Phb3H and with lower frecuency a Mhb4H. Isomers of phthalate seems not to be typical substrates for α proteobacteria, since with the exception of IphDO in *B. japonicum* USDA110, phthalate, isophthalate, or terephthalate dioxygenases are not found. BenDO are usually observed in strains endowed with a Cat12 pathway and strains endowed with Hge usually also harbor HppDO encoding genes.

The **β class of proteobacteria** harbors all major central aromatic catabolic pathways listed in *Table 2*. The distribution of these catabolic pathways among β-proteobacterial strains has some points to be noted. Except for the presence of the Hge catabolic pathway in *C. violaceum*, the families *Oxalobacteraceae, Neisseriaceae*, and *Nitrosomonadaceae* are devoid of the investigated aromatic catabolic properties. Specifically, members of the *Burkholderiaceae* and *Comamonadaceae* show a high metabolic potential and usually harbor a broad set of aromatic pathways and members of the *Burkholderia*, *Cupriavidus*, *Ralstonia*, *Delftia*, and *Polaromonas* genera comprise up to 11 out of the 14 major central aromatic pathways (Pérez-Pantoja et al., 2008). *Polynucleobacter* sp. QLW-P1DMWA-1 is the only member of the *Burkholderiaceae* family that has no such catabolic pathway (the smallest genome among them); and *Limnobacter* sp. MED105 (the second smallest genome) has only Cat23. Members of the *Alcaligenaceae* and *Rhodocyclaceae* are obviously relatively limited in their aromatic catabolic potential. It should be noted that *Rhodocyclaceae* comprise genera such as *Azoarcus, Thauera*, or "*Aromatoleum*," nitrate-reducing bacteria that contribute significantly to the biodegradation of aromatic compounds in anoxic waters and soils and that are endowed with several pathways for anaerobic catabolism of aromatics. It has, however, also been shown that aerobic aromatic pathway are functional in these bacteria (Rabus, 2005).

The most abundant pathways in the β class are Paa, Hge, Cat12, and Pca34, which are found in 60% or more of the sequenced genomes available. In contrast, Gal, Dhb, and Han are only observed in <10% of the sequenced genomes. Out of the rare pathways, the benzoquinol pathway is observed in an astonishingly eight *Burkholderia* genomes, being, with one exception, absent from any other genome sequenced thus far.

In coherence with the observation of a broad set of central aromatic pathways, a large diversity of peripheral pathways is found in β-proteobacteria, predominantly in *Burkholderiaceae* and *Comamonadaceae*. Phb3H, Mhb4H, Pht34DO, IphDO, and TphDO, among others, are funneling to protocatechuate, whereas Ohb1H, AntDO, PMO, and BenDO are channeling to catechol, with BenDO being the most abundantly observed enzyme. β-Proteobacterial strains endowed with Hge usually harbor also HppDO as funneling activity and less frequently also Mha6H, whereas Mhb6H and Ohb5H channel to Gen. Astonishingly β-proteobacterial genomes encoding Hpc_{LigB} (frequently observed in *Burkholderiaceae*) are devoid of genes encoding Pha3H.

The **γ class of Proteobacteria** has a profile that is different from the α and β classes. Three families (*Coxiellaceae, Francisellaceae*, and *Thiotrichaceae*) are completely devoid of catabolic pathways for aromatics, and four other families (*Pasteurellaceae, Legionellaceae, Psychromonadaceae*, and *Idiomarinaceae*) only show one main pathway. The catabolically most versatile family are the *Pseudomonadaceae*, the genomes of which encode 10 out of 14 main pathways. Also *Oceanospirillaceae* and *Enterobacteriaceae* show a broad catabolic potential manifested by the presence of seven main pathways, whereas five main pathways are observed in *Alteromonadaceae* and four main pathways in *Moraxellaceae* and *Xanthomonadaceae*. The most

abundant pathway in γ-proteobacteria is Hge, found in nearly 90% of the genomes. Also Hpc_{LigB}, Pca34, Paa, and Cat12 pathways are frequently observed (in 20–40% of the sequenced genomes), whereas Hqu, Cat23, and Pca45 are found only in <10% of the genomes. Of the main pathways only Box and Abc are absent in γ-proteobacteria. Within the environmental relevant *Pseudomonadaceae*, specifically Hge and both branches of the 3-oxoadipate pathway, Cat12 and Pca34, are observed. On the contrary, Hge was absent from genomes of *Enterobacteriaceae* and Cat12 and Pca34 pathways were only seldomly observed. In this important family the most relevant pathways are Hpc_{LigB}, Dhp, and Paa.

The analyzed peripheral pathways are less widespread in γ-proteobacteria compared to β-proteobacteria. Out of the four pathways channeling to protocatechuate (*Table 2*) only Phb3H is found at a significant abundance. As reactions funneling to catechol, only AntDO and BenDO are observed frequently. Also Mhb6H and Ohb5H funneling to gentisate are seldomly observed. On the contrary, genomes encoding Hge typically also encode HppDO across most γ proteobacterial families and Pha3H is frequently found in strains endowed with Hpc_{LigB}. In the *Enterobacteriaceae*, strains endowed with a Dhp pathway usually also harbor Mhp2H and PhpDO.

A general view of α, β, and γ class of proteobacteria shows that the most widespread aromatic catabolic pathways are Phb3H-Pca34, BenDO-Cat12, HppDO-Hge, and Paa.

5.4 Pathway Redundancy

The existence of alternative or redundant routes for the catabolism of some aromatic compounds has been well documented. The preference for utilization of one route not taking into consideration the presence of a second alternative route for the same compound is usually determined by environmental conditions such as carbon source availability, aromatic compounds concentration, and oxygen availability. The so far more common example of redundant routes is the ring cleavage of catechol by Cat12 or Cat23 pathways. In strains such as *C. necator 335* or *P. putida* P8 it was shown that both pathways are simultaneously induced in the presence of a high concentration of benzoate (Ampe and Lindley, 1996; Cao et al., 2008). The simultaneous presence of putative genes coding for Cat12 and Cat23 in the same bacterium is found in 11 genomes sequences and is the most frequent case of redundant routes found in the phylogenomic analysis used here. Most of the strains showing both catechol pathways are β-proteobacteria (e.g., *Burkholderia* sp. 383 and *C. necator* JMP134) but also examples in α-proteobacteria (*Novosphingobium aromaticivorans* DSM 12444), γ-proteobacteria (*Marinobacter algicola* DG893), and actinobacteria (*R. jostii* RHA1) are found. Several of these metabolically redundant bacteria were isolated by their ability to grow on BTEX (benzene/toluene/ethylbenzene/xylene), PCBs (polychlorinated biphenyls), or chloro- and nitroaromatics.

The simultaneous presence of *ortho*- and *meta*-cleavage for protocatechuate has been previously described in *A. keyseri* 12B (Eaton and Ribbons, 1982). In this strain the *meta*-cleavage of protocatechuate is induced during catabolism of phthalate and the *ortho*-cleavage is induced during the catabolism of 4-hydroxybenzoate. Two strains with complete genome sequence, *Arthrobacter* sp. FB24 and *B. phymatum* STM815, show the simultaneous presence of genes putatively coding for Pca34 and Pca45 and it would be interesting to determine if the conditions for the expression of both pathways are similar to the conditions described in *A. keyseri* 12B.

Aerobic catabolism of benzoate is another well-studied example of alternative pathways for aromatics. Benzoate can be channeled to catechol by BenDO or transformed to benzoyl-CoA by benzoate CoA ligase and then dihydroxylated by Box to be subject of non-oxygenolytic ring cleavage. The presence of genes coding for both routes is verified in seven strains belonging to the *Burkholderiales* including *B. xenovorans* LB400, whose benzoate/catechol and Box pathways are differentially expressed under diverse physiological conditions such as growth phase (Denef et al., 2006). It has been suggested that the Box pathway is used to catabolize benzoate under conditions of reduced oxygen tension (Denef et al., 2006), and it would be interesting if that possibility is valid for other strains showing this kind of pathway redundancy. A similar example of pathway redundancy is the potential for anthranilate catabolism in *B. xenovorans* LB400 endowed with an AntDO function channeling this compound to catechol and an Abc function to degrade it by a CoA-dependent pathway.

A further example of pathway redundancy is the potential to degrade 3-hydroxybenzoate, which could be performed via gentisate mediated by a Mhb6H, or via protocatechuate mediated by a Mhb4H. No examples of such pathway redundancy have been described in the literature, however, four Actinomycetales (*A. aurescens* TC1, *Arthrobacter* sp. FB24, *C. glutamicum* ATCC 13032, and *C. glutamicum* R) show the presence of genes putatively coding for both hydroxylases. If such pathway redundancy is effective and the specific conditions for the expression of each pathway are an interesting issues to be elucidated. Other theoretical possibilities of pathway redundancy such as salicylate being channeled to catechol by Ohb1H or to gentisate by Ohb5H, are not found in genomes sequenced thus far.

An additional level of redundancy revealed by this phylogenomic analysis is the presence of enzymes belonging to different protein families capable of catalyzing similar reactions. At least two examples of such redundancy are found. Genes coding for homoprotocatechuate 2,3-dioxygenases of the EXDO I as well as of the LigB family are observed in strains of *R. palustris*, *Roseobacter* sp., *S. pomeroyi*, and *S. aggregata*. Usually, however, only one of these genes is clustered with genes encoding enzymes to funnel the ring-cleavage product to Krebs cycle intermediates. Intriguing is the situation in *S. aggregata* IAM 12614 and *Roseobacter* sp. *SK209-2-6* where both genes are clustered.

Various organisms harbor two of three different types of phenol hydroxylases: Ph2H, Pbq2H, and Pmo. However, in this case it can be speculated that these enzymes differ in their substrate range and thus extend the spectrum of phenolic compounds that can be metabolized by the respective host.

5.5 Gene Redundancy

Redundancy of catabolic functions is an intriguing feature found in several bacterial genomes. Functionality of redundant pathway modules has been proven in some cases (Aoki et al., 1984; Perez-Pantoja et al., 2003; Seto et al., 1995), and it can be speculated that they play a role in fine-tuning of the expression of these catabolic properties under different environmental conditions including carbon source availability. The analysis of gene marker redundancy in the sequenced genomes shows that out of main aromatic catabolism pathways only Hpc and Han were never found repeated, it should however be noted (see above) that the simultaneous presence of homoprotocatechuate 2,3-dioxygenases of different families is in fact observed. Pca45, Gal, and Abc pathways are found repeated only in one genome each. On the other hand, pathways showing a broader distribution of redundancy are Gen found in 14 strains

from 14 different genera, and Cat12 redundant in 17 genomes of strains of six different genera. In both cases approximately 20% of genomes harboring the respective pathway comprise at least two copies of the gene encoding the ring-cleavage activity. An even higher abundance of redundancy is observed for Cat23 and Hqu, being observed as redundant in approximately 40% of genomes, which comprise this pathway. Some gene markers are even observed in three or four copies per genome.

Different advantages of having multiple pathway copies have been reported. On one hand, even though belonging to the same subfamily, encoded enzymes may differ significantly in substrate specificity, allowing a broader range of substrates to be dissimilated via a given pathway. Even single amino acid differences may significantly influence such specificity as reported by catechol 2,3-dioxygenases, Rieske non-heme iron oxygenases or soluble diiron monooxygenases, respectively (Beil et al., 1998; Junca et al., 2004; Tao et al., 2004). In case of catechol 1,2-dioxygenases, also significant differences in substrate specificity can be observed for related enzymes, and it was recently shown for *P. reinekei* MT1 that one catabolic gene cluster is suited for the metabolism of aromatics dissimilated via catechol, whereas the function of the second gene cluster is to channel methyl- but also chloroaromatics into an *ortho*-cleavage route (Cámara et al., 2007).

Inspection of the genetic environment of catechol 1,2-dioxygenase encoding genes reveals that one copy is usually located in a gene cluster comprising genes encoding enzymes for channeling the ring-cleavage product to Krebs cycle intermediates, whereas the second copy is associated with genes encoding peripheral enzymes, such as benzoate or phenol. One advantage of this strategy and generally of the presence of multiple ring-cleavage enzymes may be the avoidance of accumulation of toxic catecholic intermediates (Schweigert et al., 2001) from abundant substrates (Perez-Pantoja et al., 2003). Interestingly, even though there is a high level of redundancy with regard to the respective ring-cleavage enzymes, complete downstream pathways channeling to the Krebs cycle are only seldomly redundant. Similar to the situation for Cat12 being often connected with BenDO, out of multiple gene copies, one Pca34 is typically connected with Phb3h, and one Gen is typically connected with Mhb6h. An interesting network of redundant functions is observed in *C. necator* JMP134, where one Cat12 is associated with BenDO, and the other with one of the two copies of Pmo. The second copy of Pmo is in turn associated with a Cat23.

Strains with a high level of gene redundancy for aromatic catabolism pathways include *B. japonicum* USDA110, *S. wittichii* RW1, *B. xenovorans* LB400, and *C. necator* H16 with three redundant central ring-cleavage pathways, and the remarkably *C. necator* JMP134 that has five such redundant pathways. It should be noted that gene redundancy in these bacteria include also several cases of redundancy of peripheral pathways, showing the broad diversity of catabolic functions affected by this feature and suggesting a significant contribution to the catabolic potential for these bacteria. The existence of a high level of gene redundancy in metabolically versatile bacteria would be in line with the recently described concept of "ecoparalogs": under changing environmental circumstances (e.g., salinity, temperature, oxygen tension) bacteria could adapt to these changes by two or more copies of the genes affected by environmental fluctuations and using specialized paralogs, each one performing the same function under different conditions (Sanchez-Perez et al., 2008).

5.6 Superbugs

Several bacterial genomes are interesting because of the number of aromatic catabolism pathways and functions that they encode. Of course, several of these bacteria such as *P. putida*

KT2440 (Jimenez et al., 2002), *B. xenovorans* LB400 (Chain et al., 2006) *R. jostii* RHA1 (McLeod et al., 2006), or *C. necator* JMP134 (Pérez-Pantoja et al., 2008) were selected for genome sequencing projects as they have been used as bacterial models for aromatic metabolism studies. Other bacteria have been also included in genome sequencing projects because they were reported to be versatile aromatic degraders or harbor interesting metabolic routes (*M. vanbaalenii* PYR-1, *M. gilvum* PYR-GCK, *S. wittichii* RW1, *N. aromaticivorans* DSM 12444, *B. vietnamiensis* G4, and P. *naphthalenivorans* CJ2 among others). Although with expected differences, these bacteria all possess more than six main or rare catabolic pathways. As indicated in the corresponding sections above, some of these bacteria (*S. wittichii* RW1, *R. jostii* RHA1, *B. xenovorans* LB400, and especially *C. necator* JMP134) also possess significant levels of gene, enzyme, and pathway redundancy, which add to the catabolic potential. At least in three cases (*P. putida* KT2440, *B xenovorans* LB400, and *C. necator* JMP134) the predicted catabolic properties have been analyzed by metabolic reconstruction studies linking *in vivo* studies with *in silico* analysis (Chain et al., 2006; Jimenez et al., 2002; Pérez-Pantoja et al., 2008), and with transcriptional profiling for some pathways (Denef et al., 2004). It is worth mentioning that *R. jostii* RHA1, *B. xenovorans* LB400, and *C. necator* JM134 have 8, 9, and 11 out of the 14 main pathways defined in *Table 2*, respectively, being among the catabolically most versatile bacteria reported so far.

The inspection of the genome database allows the finding of other, sometimes unexpected, bacteria with broad aromatic metabolic potential. All three *Cupriavius* strains contain 9–11 of the main pathways and among *Burkholderia* strains, *B. phymatum* STM815, and *Burkholderia* sp. 383 deserve special attention, as they both contain ten of the main and one of the rare pathways. Also *Bradyrhizobia*, for which three genomes are available can be regarded as exceptionally versatile comprising seven to eight major pathways, as is also the case for some marine *Rhodobacteraceae* (*Silicibacter pomeroyi* DSS-3 [Moran et al., 2004], *Sagittula stellata* E-37 [Gonzalez et al., 1997], *Jannaschia* sp. CCS1), *Azoarcus* sp. BH72 and various *Comamonadaceae* or *Burkholderiaceae*. These bacteria are, therefore, choices to perform metabolic reconstruction studies as indicated above, in order to demonstrate their catabolic potential. Interestingly, *Pseudomonas* strains only comprise up to six main metabolic pathways, with *P. putida* W619 isolated from the Black Cottonwood tree having the broadest metabolic potential.

It can be assumed that these highly versatile catabolic bacteria live in environments where a variety of aromatic carbon sources are present. One kind of such habitats is the rhizosphere of plants, since it is expected that their exudates contain a myriad of organic carbon sources, most of them in tiny amounts. Interestingly, several of the bacteria listed above have been isolated or proposed to thrive in rhizospheric habitats. Even more, some of them have been described to produce beneficial effects on plants (*Bradyrhizobium*, *Burkholderia*) suggesting a mutually positive interaction between plants and versatile aromatic degraders.

6 Research Needs

Hundreds of bacterial genomes have been completely sequenced, several of which are important paradigms for pollutant transformation pathways. Such complete information of bacterial cells will allow in concert with transcriptomic and proteomic studies the analysis of the detailed behavior and physiology of these organisms, the development of bioinformatic models, and also a predictive modeling. However, various other aspects have to be considered

to reach such goals. First, misannotations in bacterial genome projects are too frequent. However, the identification in databases of proteins of which a function has been proven, to allow a comparison with the protein of interest becomes more and more complicated with the overwhelming data arising from the sequencing projects. Cured databases with valid information are necessary, such as the ribosomal database project (RDP) at rdp.cme.msu.edu or the TCDB transport classification database at www.tcdb.org, which have been developed to facilitate analysis of 16S rDNA or membrane transport proteins. Second, the amount of genes coding for proteins with unknown function is immense, and even broader than annotations may suggest. Metabolic reconstruction work may help to elucidate metabolic routes for which the genetic basis has not yet been explored (Nogales et al., 2008; Pérez-Pantoja et al., 2008). In this context, it is also important to note that an immense amount of valuable information is available on the biochemistry of metabolic pathways, and even though mainly dating back 30 or more years ago is a source that should be recommended for reading.

Even though the list of bacterial genomes sequenced or in the process of sequencing is enormous, the current database is still highly biased for easy to culture microorganisms, as expected, and even some environmentally important groups such as *Rhodococcus* sp. are represented just by one genome. Efforts should be directed towards a better understanding of the diversity inside such genera. However, it should be also noted that significant efforts are needed to really harvest the information available from already sequenced genomes.

References

Altenschmidt U, Fuchs G (1992) Novel aerobic 2-aminobenzoate metabolism. Purification and characterization of 2-aminobenzoate-CoA ligase, localisation of the genes on a 8-kbp plasmid, and cloning and sequencing of the genes from a denitrifying *Pseudomonas* sp. Eur J Biochem 205: 721–727.

Ampe F, Lindley ND (1996) Flux limitations in the *ortho* pathway of benzoate degradation of *Alcaligenes eutrophus*: metabolite overflow and induction of the *meta* pathway at high substrate concentrations. Microbiology 142: 1807–1817.

Aoki K, Konohana T, Shinke R, Nishira H (1984) Two catechol 1,2-dioxygenases from aniline-assimilating bacterium, *Frateuria* species ANA-18. Agric Biol Chem 48: 2097–2104.

Arias-Barrau E, Olivera ER, Luengo JM, Fernandez C, Galan B, Garcia JL, Diaz E, Minambres B (2004) The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. J Bacteriol 186: 5062–5077.

Arias-Barrau E, Sandoval A, Naharro G, Olivera ER, Luengo JM (2005) A two-component hydroxylase involved in the assimilation of 3-hydroxyphenyl acetate in *Pseudomonas putida*. J Biol Chem 280: 26435–26447.

Asturias JA, Timmis KN (1993) Three different 2,3-dihydroxybiphenyl-1,2-dioxygenase genes in the gram-positive polychlorobiphenyl-degrading bacterium *Rhodococcus globerulus* P6. J Bacteriol 175: 4631–4640.

Ballou DP, Entsch B, Cole LJ (2005) Dynamics involved in catalysis by single-component and two-component flavin-dependent aromatic hydroxylases. Biochem Biophys Res Commun 338: 590–598.

Beil S, Mason JR, Timmis KN, Pieper DH (1998) Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene. J Bacteriol 180: 5520–5528.

Beltrametti F, Marconi AM, Bestetti G, Colombo C, Galli E, Ruzzi M, Zennaro E (1997) Sequencing and functional analysis of styrene catabolism genes from *Pseudomonas fluorescens* ST. Appl Environ Microbiol 63: 2232–2239.

Bosch R, GarciaValdes E, Moore ERB (1999a) Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. Gene 236: 149–157.

Bosch R, Moore ERB, GarciaValdes E, Pieper DH (1999b) Nah W, a novel, inducible salicylate hydroxylase involved in mineralization of naphthalene by *Pseudomonas stutzeri* AN10. J Bacteriol 181: 2315–2322.

Buder R, Fuchs G (1989) 2-Aminobenzoyl-CoA monooxygenase/reductase, a novel type of flavoenzyme. Purification and some properties of the enzyme. Eur J Biochem 185: 629–635.

Cai M, Xun LY (2002) Organization and regulation of pentachlorophenol-degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. J Bacteriol 184: 4672–4680.

Cao B, Geng A, Loh K (2008) Induction of ortho- and meta-cleavage pathways in *Pseudomonas* in biodegradation of high benzoate concentration: MS identification of catabolic enzymes. Appl Microbiol Biotechnol 81: 99–107.

Chain PS, Denef VJ, Konstantinidis KT, Vergez LM, Agullo L, Reyes VL, Hauser L, Cordova M, Gomez L, Gonzalez M, Land M, Lao V, Larimer F, LiPuma JJ, Mahenthiralingam E, Malfatti SA, Marx CJ, Parnell JJ, Ramette A, Richardson P, Seeger M, Smith D, Spilker T, Sul WJ, Tsoi TV, Ulrich LE, Zhulin IB, Tiedje JM (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc Natl Acad Sci USA 103: 15280–15287.

Chang HK, Mohseni P, Zylstra GJ (2003) Characterization and regulation of the genes for a novel anthranilate 1,2-dioxygenase from *Burkholderia cepacia* DBO1. J Bacteriol 185: 5871–5881.

Cámara B, Bielecki P, Kaminski F, dos Santos VM, Plumeier I, Nikodem P, Pieper DH (2007) A gene cluster involved in degradation of substituted salicylates via *ortho* cleavage in *Pseudomonas* sp. strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate. J Bacteriol 189: 1664–1674.

Daane LL, Harjono I, Barns SM, Launen LA, Palleroni NJ, Häggblom MM (2002) PAH-degradation by *Paenibacillus* spp. and description of *Paenibacillus naphthalenovorans* sp. nov., a naphthalene-degrading bacterium from the rhizosphere of salt marsh plants. Int J Syst Evol Microbiol 52: 131–139.

Dehmel U, Engesser K-H, Timmis KN, Dwyer DF (1995) Cloning, nucleotide sequence, and expression of the gene encoding a novel dioxygenase involved in metabolism of carboxydiphenyl ehters in *Pseudomonas pseudoalcaligenes* POB310. Arch Microbiol 163: 35–41.

Denef VJ, Park J, Tsoi TV, Rouillard JM, Zhang H, Wibbenmeyer JA, Verstraete W, Gulari E, Hashsham SA, Tiedje JM (2004) Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. Appl Environ Microbiol 70: 4961–4970.

Denef VJ, Klappenbach JA, Patrauchan MA, Florizone C, Rodrigues JL, Tsoi TV, Verstraete W, Eltis LD, Tiedje JM (2006) Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400. Appl Environ Microbiol 72: 585–595.

Deng D, Li X, Fang X, Sun G (2007) Characterization of two components of the 2-naphthoate monooxygenase system from *Burkholderia* sp. strain JT1500. FEMS Microbiol Lett 273: 22–27.

Diaz E, Ferrandez A, Prieto MA, Garcia J (2001) Biodegradation of aromatic compounds by *Escherichia coli.* Microbiol Mol Biol Rev 65: 523–569.

Duffner FM, Muller R (1998) A novel phenol hydroxylase and catechol 2,3-dioxygenase from the thermophilic *Bacillus thermoleovorans* strain A2: nucleotide sequence and analysis of the genes. FEMS Microbiol Lett 161: 37–45.

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.

Eaton RW (1996) *p*-Cumate catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA carrying the *cmt* operon. J Bacteriol 178: 1351–1362.

Eaton RW, Ribbons DW (1982) Metabolism of dibutylphthalate and phthalate by *Micrococcus* sp. strain 12B. J Bacteriol 151: 48–57.

Eby DM, Beharry ZM, Coulter ED, Kurtz DM, Neidle EL (2001) Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. J Bacteriol 183: 109–118.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

Eltis LD, Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. J Bacteriol 178: 5930–5937.

Enroth C, Huang W, Waters S, Neujahr H, Lindqvist Y, Schneider G (1994) Crystallization and preliminary X-ray analysis of phenol hydroxlase from *Trichosporon cutaneum.* J Mol Biol 238: 128–130.

Eulberg D, Golovleva LA, Schlomann M (1997) Characterization of catechol catabolic genes from *Rhodococcus erythropolis* 1CP. J Bacteriol 179: 370–381.

Fuenmayor SL, Wild M, Boyes AL, Williams PA (1998) A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J Bacteriol 180: 2522–2530.

Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem 70: 209–246.

Gibson DT, Parales RE (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. Curr Opin Biotechnol 11: 236–243.

Gonzalez JM, Mayer F, Moran MA, Hodson RE, Whitman WB (1997) *Sagittula stellata* gen. nov.,

sp. nov., a lignin-transforming bacterium from a coastal environment. Int J Syst Bacteriol 47: 773–780.

Haigler BE, Johnson GR, Suen WC, Spain JC (1999) Biochemical and genetic evidence for meta-ring cleavage of 2,4,5-trihydroxytoluene in *Burkholderia* sp. strain DNT. J Bacteriol 181: 965–972.

Hatta T, Mukerjee-Dhar G, Damborsky J, Kiyohara H, Kimbara K (2003) Characterization of a novel thermostable Mn(II)-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase from a polychlorinated biphenyl- and naphthalene-degrading *Bacillus* sp. JF8. J Biol Chem 278: 21483–21492.

Hawumba JF, Brözel VS, Theron J (2007) Cloning and characterization of a 4-hydroxyphenylacetate 3-hydroxylase from the thermophile *Geobacillus* sp. PA-9. Curr Microbiol 55: 480–484.

Hiromoto T, Fujiwara S, Hosokawa K, Yamaguchi H (2006) Crystal structure of 3-hydroxybenzoate hydroxylase from Comamonas testosteroni has a large tunnel for substrate and oxygen access to the active site. J Mol Biol 364: 878–896.

Huang Y, Zhao KX, Shen XH, Chaudhry MT, Jiang CY, Liu SJ (2006) Genetic characterization of the resorcinol catabolic pathway in *Corynebacterium glutamicum*. Appl Environ Microbiol 72: 7238–7245.

Iida T, Nakamura K, Izumi A, Mukouzaka Y, Kudo T (2006) Isolation and characterization of a gene cluster for dibenzofuran degradation in a new dibenzofuran-utilizing bacterium, *Paenibacillus* sp. strain YK5. Arch Microbiol 184: 305–315.

Ishiyama D, Vujaklija D, Davies J (2004) Novel pathway of salicylate degradation by *Streptomyces* sp. strain WA46. Appl Environ Microbiol 70: 1297–1306.

Ismail W, El-Said Mohamed M, Wanner BL, Datsenko KA, Eisenreich W, Rohdich F, Bacher A, Fuchs G (2003) Functional genomics by NMR spectroscopy. Phenylacetate catabolism in *Escherichia coli*. Eur J Biochem 270: 3047–3054.

Iwasaki T, Miyauchi K, Masai E, Fukuda M (2006) Multiple-subunit genes of the aromatic-ring-hydroxylating dioxygenase play an active role in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 72: 5396–5402.

Jimenez JI, Minambres B, Garcia JL, Diaz E (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environ Microbiol 4: 824–841.

Jones RM, Pagmantidis V, Williams PA (2000) sal genes determining the catabolism of salicylate esters are part of a supraoperonic cluster of catabolic genes in *Acinetobacter* sp. strain ADP1. J Bacteriol 182: 2018–2025.

Junca H, Plumeier I, Hecht HJ, Pieper DH (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. Microbiology 150: 4181–4187.

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.

Kasai D, Masai E, Miyauchi K, Katayama Y, Fukuda M (2005) Characterization of the gallate dioxygenase gene: three distinct ring cleavage dioxygenases are involved in syringate degradation by *Sphingomonas paucimobilis* SYK-6. J Bacteriol 187: 5067–5074.

Kim SH, Miyatake H, Hisano T, Iwasaki W, Ebihara AKM (2007) Crystallization and preliminary X-ray analysis of the oxygenase component (HpaB) of 4-hydroxyphenylacetate 3-monooxygenase from *Thermus thermophilus* HB8. Acta Crystallogr Sect F Struct Biol Cryst Commun 63: 556–559.

Kurnasov O, Jablonski L, Polanuyer B, Dorrestein P, Begley T, Osterman A (2003) Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. FEMS Microbiol Lett 227: 219–227.

Laurie AD, Lloyd-Jones G (1999) Conserved and hybrid meta-cleavage operons from PAH-degrading *Burkholderia* RP007. Biochem Biophys Res Commun 262: 308–314.

Leahy JG, Batchelor PJ, Morcomb SM (2003) Evolution of the soluble diiron monooxygenases. FEMS Microbiol Rev 27: 449–479.

Lee SH, Ka JO, Cho JC (2008) Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. FEMS Microbiol Lett 285: 263–269.

Maeda M, Chung S-Y, Song E, Kudo T (1995) Multiple genes encoding 2,3-dihydroxybiphenyl 1,2-dioxygenase in the gram-positive polychlorinated biphenyl-degrading bacterium *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem. Appl Environ Microbiol 61: 549–555.

Martin VJ, Mohn WW (2000) Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. J Bacteriol 182: 3784–3793.

Martin VJJ, Mohn WW (1999) A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-degrading bacterium *Pseudomonas abietaniphila* BKME-9. J Bacteriol 181: 2675–2682.

McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJ, Holt R, Brinkman FS, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a

catabolic powerhouse. Proc Natl Acad Sci USA 103: 15582–15587.

Miyauchi K, Adachi Y, Nagata Y, Takagi M (1999) Cloning and sequencing of a novel meta-cleavage dioxygenase gene whose product is involved in degradation of gamma-hexachlorocyclohexane in *Sphingomonas paucimobilis*. J Bacteriol 181: 6712–6719.

Miyazawa D, Mukerjee-Dhar G, Shimura M, Hatta T, Kimbara K (2004) Genes for Mn(II)-dependent NahC and Fe(II)-dependent NahH located in close proximity in the thermophilic naphthalene and PCB degrader, *Bacillus* sp. JF8: cloning and characterization. Microbiology 150: 993–1004.

Moonen MJ, Synowsky SA, van den Berg WA, Westphal AH, Heck AJ, van den 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.

Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J, Ward N (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature 432: 910–913.

Muraki T, Taki M, Hasegawa Y, Iwaki H, Lau PCK (2003) Prokaryotic homologs of the eukaryotic 3-hydroxyanthranilate 3,4-dioxygenase and 2-amino-3-carboxymuconate-6-semialdehyde decarboxylase in the 2-nitrobenzoate degradation pathway of *Pseudomonas fluorescens* strain KU-7. Appl Environ Microbiol 69: 1564–1572.

Nakano H, Wieser M, Hurh B, Kawai T, Yoshida T, Yamane T, Nagasawa T (1999) Purification, characterization and gene cloning of 6-hydroxynicotinate 3-monooxygenase from *Pseudomonas fluorescens* TN5. Eur J Biochem 260: 120–126.

Nogales J, Palsson B, Thiele I (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst Biol 2: 79.

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.

Nomura Y, Nakagawa M, Ogawa N, Harashima S, Oshima Y (1992) Genes in PHT plasmid encoding the initial degradation pathway of phthalate in *Pseudomonas putida*. J Ferm Bioeng 74: 333–344.

Nordin K, Unell M, Jansson JK (2005) Novel 4-chlorophenol degradation gene cluster and degradation route via hydroxyquinol in *Arthrobacter chlorophenolicus* A6. Appl Environ Microbiol 71: 6538–6544.

Nurk A, Kasak L, Kivisaar M (1991) Sequence of the gene (*pheA*) encoding phenol monooxygenae from *Pseudomonas* sp. EST1001: expression in *Escherichia coli* and *Pseudomonas putida*. Gene 102: 13–18.

Park HS, Kim HS (2000) Identification and characterization of the nitrobenzene catabolic plasmids pNB1 and pNB2 in *Pseudomonas putida* HS12. J Bacteriol 182: 573–580.

Peng X, Misawa N, Harayama S (2003) Isolation and characterization of thermophilic *Bacilli* degrading cinnamic, 4-coumaric, and ferulic acids. Appl Environ Microbiol 69: 1417–1427.

Perez-Pantoja D, Ledger T, Pieper DH, Gonzalez B (2003) Efficient turnover of chlorocatechols is essential for growth of *Ralstonia eutropha* JMP134 (pJP4) in 3-chlorobenzoic acid. J Bacteriol 185: 1534–1542.

Pérez-Pantoja D, De la Iglesia R, Pieper DH, Gonzalez B (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. FEMS Microbiol Rev 32: 736–794.

Pinyakong O, Habe H, Yoshida T, Nojiri H, Omori T (2003) Identification of three novel salicylate 1-hydroxylases involved in the phenanthrene degradation of *Sphingobium* sp. strain P2. Biochem Biophys Res Commun 301: 350–357.

Rabus R (2005) Functional genomics of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. Appl Microbiol Biotechnol 68: 580–587.

Rascher A, Hu Z, Buchanan GO, Reid R, Hutchinson CR (2005) Insights into the biosynthesis of the benzoquinone ansamycins geldanamycin and herbimycin, obtained by gene sequencing and disruption. Appl Environ Microbiol 71: 4862–4871.

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.

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.

Sakai M, Ezaki S, Suzuki N, Kurane R (2005) Isolation and characterization of a novel polychlorinated biphenyl-degrading bacterium, *Paenibacillus* sp. KBC101. Appl Microbiol Biotechnol 68: 111–116.

Sanchez MA, Gonzalez B (2007) Genetic characterization of 2,4,6-trichlorophenol degradation in *Cupriavidus necator* JMP134. Appl Environ Microbiol 73: 2769–2776.

Sanchez-Perez G, Mira A, Nyiro G, Pasić L, Rodriguez-Valera F (2008) Adapting to environmental changes using specialized paralogs. Trends Genet 24: 154–158.

Sasoh M, Masai E, Ishibashi S, Hara H, Kamimura N, Miyauchi K, Fukuda M (2006) Characterization of the terephthalate degradation genes of *Comamonas* sp. strain E6. Appl Environ Microbiol 72: 1825–1832.

Sato S, Ouchiyama N, Kimura T, Nojiri H, Yamane H, Omori T (1997a) Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: Nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase. J Bacteriol 179: 4841–4849.

Sato SI, Nam JW, Kasuga K, Nojiri H, Yamane H, Omori T (1997b) Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10. J Bacteriol 179: 4850–4858.

Schweigert N, Zehnder AJB, Eggen RIL (2001) Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. Environ Microbiol 3: 81–91.

Seto M, Masai E, Ida M, Hatta T, Kimbara K, Fukuda M, Yano K (1995) Multiple polychlorinated biphenyl transformation systems in the gram-positive bacterium *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 61: 4510–4513.

Shashirekha S, Uma L, Subramanian G (1997) Phenol degradation by the marine cyanobacterium *Phormidium valderianum* BDU 30501. J Ind Microbiol Biotechnol 19: 130–133.

Shimura M, MukerjeeDhar G, Kimbara K, Nagato H, Kiyohara H, Hatta T (1999) Isolation and characterization of a thermophilic *Bacillus* sp. JF8 capable of degrading polychlorinated biphenyls and naphthalene. FEMS Microbiol Lett 178: 87–93.

Smith DJ, Park J, Tiedje JM, Mohn WW (2007) A large gene cluster in *Burkholderia xenovorans* encoding abietane diterpenoid catabolism. J Bacteriol 189: 6195–6204.

Stingley RL, Khan AA, Cerniglia CE (2004a) Molecular characterization of a phenanthrene degradation pathway in *Mycobacterium vanbaalenii* PYR-1. Biochem Biophys Res Commun 322: 133–146.

Stingley RL, Brezna B, Khan AA, Cerniglia CE (2004b) Novel organization of genes in a phthalate degradation operon of *Mycobacterium vanbaalenii* PYR-1. Microbiology 150: 3749–3761.

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.

Suske WA, Held M, Schmid A, Fleischmann T, Wubbolts MG, Kohler HPE (1997) Purification and characterization of 2-hydroxybiphenyl 3-monooxygenase, a novel NADH-dependent, FAD-containing aromatic hydroxylase from *Pseudomonas azelaica* HBP1. J Biol Chem 272: 24257–24265.

Tago K, Sato J, Takesa H, Kawagishi H, Hayatsu M (2005) Characterization of methylhydroquinone-metabolizing oxygenase genes encoded on plasmid in *Burkholderia* sp. NF100. J Biosci Bioeng 100: 517–523.

Taguchi K, Motoyama M, Kudo T (2004) Multiplicity of 2,3-dihydroxybiphenyl dioxygenase genes in the Gram-positive polychlorinated biphenyl degrading bacterium *Rhodococcus rhodochrous* K37. Biosci Biotechnol Biochem 68: 787–795.

Takenaka S, Murakami S, Shinke R, Hatakeyama K, Yukawa H, Aoki K (1997) Novel genes encoding 2-aminophenol 1,6-dioxygenase from *Pseudomonas species* AP-3 growing on 2-aminophenol and catalytic properties of the purified enzyme. J Biol Chem 272: 14727–14732.

Takeo M, Yasukawa T, Abe Y, Niihara S, Maeda Y, Negoro S (2003) Cloning and characterization of a 4-nitrophenol hydroxylase gene cluster from *Rhodococcus* sp. PN1. J Biosci Bioeng 95: 139–145.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetic analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

Tao Y, Fishman A, Bentley WE, Wood TK (2004) Altering toluene 4-monooxygenase by active-site engineering for the synthesis of 3-methoxycatechol, methoxyhydroquinone, and methylhydroquinone. J Bacteriol 186: 4705–4713.

Thotsaporn K, Sucharitakul J, Wongratana J, Suadee C, Chaiyen P (2004) Cloning and expression of *p*-hydroxyphenylacetate 3-hydroxylase from *Acinetobacter baumannii*: evidence of the divergence of enzymes in the class of two-protein component aromatic hydroxylases. Biochim Biophys Acta 1680: 60–66.

van Berkel WJ, Kamerbeek NM, Fraaije MW (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. J Biotechnol 124: 670–689.

van der Geize R, Dijkhuizen L (2004) Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. Curr Opin Microbiol 7: 255–261.

Vetting MW, Wackett LP, Que L, Lipscomb JD, Ohlendorf DH (2004) Crystallographic comparison of manganese- and iron-dependent homoprotocatechuate 2,3-dioxygenases. J Bacteriol 186: 1945–1958.

Wang YZ, Zhou Y, Zylstra GJ (1995) Molecular analysis of isophthalate and terephthalate degradation by *Comamonas testosteroni* YZW-D. Environ Health Perspect 103: 9–12.

Williams PA, Murray K (1974) Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. J Bacteriol 120: 416–423.

Witzig R, Junca H, Hecht HJ, Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. Appl Environ Microbiol 72: 3504–3514.

Xu L, Resing K, Lawson SL, Babbitt PC, Copley SD (1999) Evidence that pcpA encodes 2,6-dichlorohydroquinone dioxygenase, the ring cleavage enzyme required for pentachlorophenol degradation in *Sphingomonas chlorophenolica* strain ATCC 39723. Biochemistry 38: 7659–7669.

Yen KM, Gunsalus IC (1982) Plasmid gene organization: naphthalene/salicylate oxidation. Proc Natl Acad Sci USA 79: 874–878.

Yoshida M, Oikawa T, Obata H, Abe K, Mihara H, Esaki N (2007) Biochemical and genetic analysis of the gamma-resorcylate (2,6-dihydroxybenzoate) catabolic pathway in *Rhizobium* sp. strain MTP-10005: identification and functional analysis of its gene cluster. J Bacteriol 189: 1573–1581.

Zaar A, Gescher J, Eisenreich W, Bacher A, Fuchs G (2004) New enzymes involved in aerobic benzoate metabolism in *Azoarcus evansii*. Mol Microbiol 54: 223–238.

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the tod operon. Appl Environ Microbiol 54: 1498–1503.