This is a pre- or post-print of an article published in
Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M.,
Seeger, M., Pieper, D.H., González, B.
Genomic analysis of the potential for aromatic compounds
biodegradation in Burkholderiales
(2012) Environmental Microbiology, 14 (5), pp. 1091-1117.

1

1 Environmental Microbiology and Environmental Microbiology Reports

2 Mini-review

3 Genomic analysis of the potential for aromatic compounds biodegradation in

4 *Burkholderiales*

5 Danilo Pérez-Pantoja¹, Raúl Donoso^{1,4}, Loreine Agulló², Macarena Córdova², Michael
6 Seeger², Dietmar H. Pieper³ and Bernardo González^{1,4*}

7¹ Center for Advanced Studies in Ecology and Biodiversity. Millennium Nucleus in Plant
8 Functional Genomics. Facultad de Ciencias Biológicas, P. Universidad Católica de Chile.
9 Santiago, Chile.

10² Laboratorio de Microbiología Molecular y Biotecnología Ambiental, Departamento de
11 Química, Center for Nanotechnology and Systems Biology, Universidad Técnica
12 Federico Santa María, Valparaíso, Chile.

13³ Microbial Interactions and Processes Research Group, Department of Medical
14 Microbiology, HZI - Helmholtz Centre for Infection Research. Braunschweig, Germany

15⁴ Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez. Santiago, Chile.

16 Running title: Degradation of aromatic compounds in *Burkholderiales*

17 Keywords: *Alcaligenaceae*, aromatic compounds, *Burkholderiaceae*, catabolic pathways,
18 *Comamonadaceae*

19 *Corresponding author. Facultad de Ingeniería y Ciencias. Universidad Adolfo Ibáñez.

20 Postal Code: 7941169. Santiago, Chile. Phone: +56 (2) 3311619. E-mail:

21 bernardo.gonzalez@uai.cl

Abstract

23The relevance of the β -proteobacterial *Burkholderiales* order in the degradation of
24a vast array of aromatic compounds, including several priority pollutants, has been
25largely assumed. In this review, the presence and organization of genes encoding
26oxygenases involved in aromatics biodegradation in 80 *Burkholderiales* genomes
27is analyzed. This genomic analysis underscores the impressive catabolic potential
28of this bacterial lineage, comprising nearly all of the central ring-cleavage pathways
29reported so far in bacteria and most of the peripheral pathways involved in
30channelling of a broad diversity of aromatic compounds. The more widespread
31pathways in *Burkholderiales* include protocatechuate *ortho* ring-cleavage, catechol
32*ortho* ring-cleavage, homogentisate ring-cleavage and phenylacetyl-CoA ring
33cleavage pathways found in at least 60% of genomes analyzed. In general, a
34genus-specific pattern of positional ordering of biodegradative genes is observed in
35the catabolic clusters of these pathways indicating recent events in its evolutionary
36history. In addition, a significant bias towards secondary chromosomes, now
37termed chromids, is observed in the distribution of catabolic genes across
38multipartite genomes, which is consistent with a genus-specific character. Strains
39isolated from environmental sources such as soil, rhizosphere, sediment or sludge
40show a higher content of catabolic genes in their genomes compared to strains
41isolated from human, animal or plant hosts, but no significant difference is found
42among *Alcaligenaceae*, *Burkholderiaceae* and *Comamonadaceae* families,
43indicating that habitat is more of a determinant than phylogenetic origin in shaping
44aromatic catabolic versatility.

45Introduction

46

47Aromatic compounds are widespread in nature, being found as lignin and
48petroleum components, xenobiotic chemicals, aromatic amino acids, and
49constituents of plant exudates, among other sources. The aerobic degradation of
50aromatic compounds and their halogenated derivatives by bacteria has been well
51studied (Pérez-Pantoja *et al.*, 2010a; Pieper *et al.*, 2010). The general principle of
52aromatics degradation indicates that a broad range of peripheral reactions
53transforms a huge variety of compounds to a restricted set of central intermediates,
54which are subject to ring-cleavage and subsequent funnelling into the Krebs cycle.
55Typically, peripheral reactions consist in activation of the aromatic ring through
56oxygenases and/or CoA ligases generating di- or trihydroxylated intermediates
57and/or dearomatized CoA derivatives (Fig. 1) (Pérez-Pantoja *et al.*, 2010a; Pieper
58*et al.*, 2010). The activation of the aromatic ring through hydroxylation (Fig. 1, outer
59circle) is commonly catalyzed by members of one of three oxygenase families: the
60Rieske non-heme iron oxygenases which frequently catalyze the incorporation of
61two oxygen atoms (Gibson and Parales, 2000), the flavoprotein monooxygenases
62(van Berkel *et al.*, 2006), and the soluble diiron multicomponent oxygenases
63(Leahy *et al.*, 2003). The subsequent ring cleavage of di- or trihydroxylated
64intermediates (Fig. 1, inner circle) can be catalyzed by either intradiol or extradiol
65dioxygenases. While all intradiol dioxygenases described thus far belong to the
66same superfamily, members of at least three different families are reported to be
67involved in the extradiol ring cleavage of hydroxylated aromatics. Altogether, the
68inventory of oxygenases involved in activation and cleavage of the aromatic ring is

69extensive and phylogenetically diverse, including several different families that
70have been recently compiled for phylogenomic studies (Pérez-Pantoja *et al.*,
712010b). Analysis of the distribution of these aromatic oxygenases-encoding genes
72among available bacterial genome sequences has shown that members of the
73*Burkholderiales* order of the β -proteobacteria harbor the most impressive potential
74for aromatic compounds catabolism. These biodegradative abilities become
75outstanding in highly specialized degraders belonging to this bacterial lineage
76whose genomic sequences have been completed (Chain *et al.*, 2006; Gross *et al.*,
772008; Mattes *et al.*, 2008; Pérez-Pantoja *et al.*, 2008). Based on the available
78literature in the biodegradation field, the exceptional catabolic potential of
79*Burkholderiales* revealed by genomic analysis is not surprising because of the
80large number of strains belonging to this order that have been isolated by its
81biodegradative abilities (a non-comprehensive list restricted to genus *Burkholderia*
82is found in Denef, 2007). The relevance of *Burkholderiales* for environmental
83biotechnology has been largely proposed based on these culture-dependent
84methods (O'Sullivan and Mahenthiralingam, 2005; Chiarini *et al.*, 2006; Denef,
852007). However, the unequivocal confirmation of *Burkholderiales* preponderance in
86bioremediation processes occurring in multiple environmental systems has been
87raised with the emergence of stable isotope probing techniques that links functional
88activity to specific members of microbial communities (Madsen, 2006). Using this
89*in situ* approach, a key biodegradative role for *Burkholderiales* has been reported
90in bioremediation of multiple aromatic compounds including: polychlorobiphenyls
91(Tillmann *et al.*, 2005; Uhlik *et al.*, 2009), toluene (Sun *et al.*, 2010), benzene (Liou

92 *et al.*, 2008; Xie *et al.*, 2011), benzoate (Pumphrey *et al.*, 2008), salicylate
 93 (Singleton *et al.*, 2005), phenol (Manefield *et al.*, 2005), pentachlorophenol
 94 (Mahmood *et al.*, 2005), 2,4-dichlorophenoxyacetate (Cupples and Sims, 2007),
 95 naphthalene (Jeon *et al.*, 2003; Singleton *et al.*, 2005) and phenanthrene
 96 (Singleton *et al.*, 2005). Such multiple evidences consolidate *Burkholderiales* as
 97 the main player in microbial ecology of bioremediation treatments for aromatic
 98 decontamination.

99 The order *Burkholderiales* belongs to the β subclass of proteobacteria and
 100 comprises the families *Burkholderiaceae*, *Oxalobacteraceae*, *Alcaligenaceae* and
 101 *Comamonadaceae* including strictly aerobic and facultative anaerobic
 102 chemoorganotrophs, obligate and facultative chemolithotrophs, nitrogen-fixing
 103 organisms, as well as plant, animal, and human pathogens (Garrrity *et al.*, 2005);
 104 therefore constituting a phenotypic-, metabolic- and ecologically diverse bacterial
 105 lineage. The molecular basis for this multi level, remarkable versatility has started
 106 to be unveiled by the large quantity of sequenced genomes now publicly available,
 107 allowing comparative genomic analyses of metabolic traits. This prompted us to
 108 carry out a detailed genomic analysis on the presence, organization and
 109 phylogenetic relationships among oxygenase encoding genes involved in aromatic
 110 catabolism in the *Burkholderiales* order. This study improves our understanding of
 111 the biodegradative potential of this bacterial group, revealing the importance of
 112 aromatic catabolism for the habitat adaptability of bacteria.

113

115 *Distribution of catabolic functions in Burkholderiales*

116 Genomes belonging to proteobacterial strains (nearly 1200) make up the more
117 abundant (almost half) bacterial phylum in the database of genomic sequences.
118 The sequenced proteobacterial genomes comprise 167 genomes of β -
119 proteobacteria that in turn contain 110 genomes belonging to strains of the
120 *Burkholderiales* order. Thus, *Burkholderiales* is the second most represented
121 proteobacterial order in the genome database following *Enterobacteriales*. The 80
122 *Burkholderiales* genomes analyzed (Fig. 2, Table S1 of Supplementary Material)
123 comprise representatives of all four *Burkholderiales* families and include at least 48
124 species from 23 genera. However, the distribution of members belonging to the
125 different *Burkholderiales* families in these 80 sequenced genomes is not
126 homogeneous since 48 strains belong to the *Burkholderiaceae* family. The other
127 strains are distributed among the *Comamonadaceae* (16), *Alcaligenaceae* (8) and
128 *Oxalobacteraceae* (5) families (Fig. 2). Three unclassified *Burkholderiales* species,
129 *Thiomonas intermedia* K12, *Methylibium petroleiphilum* PM1 and *Leptothrix*
130 *cholodnii* SP-6 were also analyzed.

131 The genome database for *Burkholderiales* includes bacteria that can be
132 classified, somewhat artificially, based on their source of isolation: (1) Human host,
133 mainly opportunistic pathogens (18 strains, designated **H** in Table S1); (2) Animal
134 host, mainly pathogenic (4 strains, **Z**); (3) Plant pathogens (6 strains, **P**); (4)
135 Rhizosphere and root nodules (12 strains, **R**); (5) Soil (11 strains, **S**); (6)

136 Wastewater and sludge (12 strains, **W**); (7) Sediment (7 strains, **D**); (8)
 137 Endosymbionts (3 strains, **E**), and (9) Miscellaneous sources (7 strains, **U**). It
 138 should be kept in mind that the isolation source does not necessarily reflect the
 139 actual habitat/ecological function of a strain in nature. For instance, different strains
 140 of a given species have been isolated from different environments (e.g.
 141 *Burkholderia thailandensis*) and may thus be considered to thrive in different
 142 environmental niches. Then, correlations between the habitat and the
 143 presence/absence of catabolic properties (see the end of this section), as well as
 144 with other phenotypic traits, should be taken with caution.

145 To address the distribution of aromatic degradation capabilities throughout
 146 the genomes of sequenced members of the *Burkholderiales* order, 48 oxygenase
 147 encoding gene markers covering aromatic ring activating reactions and central
 148 ring-cleavage pathways were surveyed (Table 1). Well-characterized key
 149 oxygenase sequences were used as initial seeds to search for orthologous
 150 sequences. The distribution of aromatic catabolic functions among the
 151 *Burkholderiales* members is highly variable (Fig. 2). The number of detected
 152 catabolic sequence markers ranges from zero in seven strains (*Herminimonas*
 153 *arsenicoxydans*, *Janthinobacterium* sp. Marseille, *Oxalobacter formigenes*
 154 HOxBLS, *Oxalobacter formigenes* OxCC13, *Taylorella equigenitalis* MCE9,
 155 *Polynucleobacter* sp. QLW-P1DMWA-1 and *Polynucleobacter necessarius* subsp.
 156 *necessarius* STIR 1) to 25 in *B. lata* 383, 30 in *B. xenovorans* LB400), and 33 in
 157 *Cupriavidus pinatubonensis* JMP134, formerly *Ralstonia eutropha* and *C. necator*
 158 (Sato *et al.*, 2006). The number of genomes scoring positive for the presence of a

particular marker is also highly variable. While gene markers tracking the phenylacetyl-CoA oxygenase/reductase (Paa), homogentisate 1,2-dioxygenase (Hge), 4-hydroxyphenylpyruvate dioxygenase (HppDO), homoprotocatechuate 2,3-dioxygenase (Hpc), benzoate 1,2-dioxygenase (BenDO), catechol 1,2-dioxygenase (Cat12), 4-hydroxybenzoate 3-hydroxylase (Phb3H) and protocatechuate 3,4-dioxygenase (Pca34) key catabolic activities are highly represented in *Burkholderiales* genomes (the number of genomes possessing such markers ranged between 39 – 72), four oxygenase encoding genes are found only once and five only twice in these genomes (Fig. 2). The overall distribution of catabolic genes in the *Burkholderiales* order is more heterogeneous than in other bacterial orders (Pérez-Pantoja *et al.*, 2010b), which might reflect the ecological diversity of this group.

Significantly different presence/absence patterns are observed within members of *Burkholderiales*. Genomes of members of the three main genera of the *Burkholderiaceae* family (*Burkholderia*, *Cupriavidus* and *Ralstonia*) encode a significant number of catabolic functions (usually more than nine). An average of 14 catabolic oxygenases are encoded in the 35 *Burkholderia* strains analyzed. As an exception, in the genome of *B. rhizoxinica* HKI 0454 only four of these proteins are detected. This may reflect the fact that this strain is a fungal endosymbiont whose genome is significantly smaller than that of the other *Burkholderia* species (Table S1), and consequently shows a reduced metabolic versatility (Lackner *et al.*, 2011). The Pca34, Phb3H, Cat12, Hge, HppDO and Paa encoding genes are present in all *Burkholderia* species, with the exception of strain HKI 0454 that only

182has the *paa* genes. Members of the phylogenetic lineage formed by *B. phymatum*,
 183*B. graminis*, *B. phytofirmans*, *B. xenovorans*, *Burkholderia* sp. CCGE 1001,
 184CCGE1002, CCGE1003, H160, and Ch1-1 (Fig. 2), have catabolic properties
 185different from those of the so called “*B. cepacia* complex” and the so called
 186“*Pseudomallei* group”, and are the only *Burkholderia* species that harbor genes
 187encoding the Abc and Box proteins (Fig. 2), both comprising CoA- dependent
 188oxidative pathways. The overall metabolic potential of these nine *Burkholderia*
 189strains is, in general, more similar to that of strains of the *Cupriavidus* and
 190*Ralstonia* genera than to strains of the “*B. cepacia* complex”. The aromatic
 191metabolic potential of *Burkholderiales* strains not belonging to the genera
 192*Burkholderia*, *Cupriavidus* and *Ralstonia* is highly heterogeneous. For instance, the
 193Pca34, BenDO and Hpc markers, frequently found in *Burkholderia*, *Cupriavidus*
 194and *Ralstonia*, are encoded only in 3-5 genomes of other *Burkholderiales* strains;
 195and AntDO and laaDO proteins, commonly encoded in *Burkholderia* genomes, are
 196absent in all of the rest, except in *Herbaspirillum seropedicae* SmR1. In contrast,
 197several representatives with unique metabolic properties were observed in this
 198heterogeneous group. The *p*-cresol methylhydroxylase (PcmH) encoding genes
 199are exclusively observed in two of the *Comamonas testosteronii* genomes and in
 200*Polaromonas naphthalenivorans* CJ2; and *C. testosteronii* KF-1 is the only strain
 201whose genome encodes a 2,4,5-trihydroxytoluene dioxygenase (Tht).

202 Multiple replicons (multipartite) genomes are found in all *Burkholderia*,
 203*Cupriavidus* and *Ralstonia* strains, including the small genome of *B. rhizoxinica*,
 204and in nine other *Burkholderiales* strains. About 70% of completed genomes

205 contain more than one replicon (Table S1). The term “chromid” has been recently
 206 proposed for secondary chromosomes to distinguish them from both chromosomes
 207 and plasmids, according to the fact that these replicons have a distinct and
 208 consistent set of properties (Harrison *et al.*, 2010). By its utility we adopted this
 209 nomenclature to analyze several interesting facts concerning the distribution of
 210 catabolic gene markers among multipartite genomes (numbered as “2” and “3” in
 211 Fig. 2). The distribution of the relative content of catabolic genes in the main
 212 *Burkholderiales* genera harboring multipartite genomic structure, i.e. the number of
 213 these oxygenases encoding genes relative to the total number of genes in each
 214 replicon type, show a strong bias towards chromids against the chromosome (Fig.
 215 3A). Statistical analysis (t-test) shows that these differences in catabolic content
 216 among replicon types are all significant at a confidence level of 95%. Examples are
 217 Pca34 and Hpc proteins, which are encoded on chromids in all *Burkholderia* and
 218 *Cupriavidus* strains (except *B. rhizoxinica*). Chromids also predominantly encode
 219 Cat12 proteins. Noteworthy, only two of the ring-cleavage oxygenases analyzed
 220 are preferentially encoded in the chromosome: homogentisate 1,2-dioxygenase
 221 (Hge) and phenylacetyl-CoA oxygenase/reductase (Paa). Hge is typically encoded
 222 in the chromosome, except in *Burkholderia glumae* BGR1 and members of
 223 *Cupriavidus* and *Ralstonia* genera. In strains containing multiple gene copies of
 224 Hge (e.g. *B. xenovorans* LB400 and *Burkholderia* sp. CCGE1002), at least one
 225 gene is located on the chromosome. The Paa protein is also commonly encoded in
 226 the chromosome, except in *Ralstonia* genomes. Again, additional copies are
 227 localized elsewhere in some genomes such as those of *C. necator* H16, *C.*
 228 *pinatubonensis* JMP134, *C. taiwanensis* LMG19424 and *Burkholderia* sp.

229CCGE1002. Coincident with its preferential chromosomal location, Hge and Paa
 230are the ring-cleavage oxygenases that show the broadest distribution in
 231*Burkholderiales* genomes. Both features suggest that these catabolic routes are
 232“core” pathways in *Burkholderiales*, which is additionally supported by their
 233involvement in metabolization of aromatic amino acids. On the other hand, the only
 234central catabolic markers encoded in plasmids are Cat23 in *B. vietnamiensis* G4,
 235Han in *Polaromonas* sp. JS666, and Cca in *Achromobacter xylosooxidans* A8 and
 236*C. pinatubonensis* JMP134 (designated as “P” in Fig. 2), in accordance with the
 237fact that these pathways have often been localized on catabolic megaplasms
 238(Vedler, 2009). A few trends can also be observed when the distribution of
 239peripheral reactions markers among multipartite genomes is analyzed (Fig. 2). The
 240Phb3H, VanOD, IvaOD, AntDO, IaaDO, DapDO, Sal1H, Mhp2H and Mha6H
 241catabolic markers are predominantly encoded in chromids. The HppDO marker
 242encoding 4-hydroxyphenylpyruvate dioxygenase that forms homogentisate, the
 243most widely distributed marker found in this survey, is always located in the
 244chromosome, except in the four *Cupriavidus* strains where both this peripheral
 245marker and the corresponding Hge central marker are encoded in chromids.
 246Interestingly, the configuration of these catabolically linked oxygenases being
 247encoded in different replicons is only observed in the *Ralstonia* strains and in *B.*
 248*glumae* BGR1. This genetic arrangement would correspond to an intermediate
 249stage in the “migration” of this pathway from chromosome to chromid that has
 250occurred in the *Cupriavidus* genus (Fig. 2).

251 The relation between genome size and content of catabolic oxygenases was
 252 compared in *Burkholderiales* (Fig. 3B). Interestingly, the percentage of aromatic
 253 oxygenases encoding genes increases with genome size, but the relative content
 254 of enzymes remains fixed, suggesting that the catabolism of aromatic compounds
 255 has increased relevance in *Burkholderiales* metabolisms, as their genomes get
 256 larger. Additionally, it should be noted that some strains as *L. cholodnii* SP-6, *B.*
 257 *petrii* DSMZ 12804, *C. testosteroni* CNB-2 and *P. naphthalenivorans* CJ2, show a
 258 disproportionately high content of aromatic oxygenases; and especially remarkable
 259 is the catabolic specialization of *C. pinatubonensis* JMP134 (Lykidis *et al.* 2010).

260 Can some conclusions about the presence of these aromatic catabolism
 261 properties on the ecological or “life-style” groups defined above be drawn? This
 262 would be, of course, one of the relevant outcomes of these analyses. In general,
 263 within the *Burkholderiaceae* family, human (Table S1, **H**) and/or zoonotic
 264 pathogens (Table S1, **Z**) have more aromatic catabolism potential than expected
 265 for animal host-associated species. This may be explained by the fact that these
 266 strains are predominantly opportunistic pathogens, and, therefore thrive as free-
 267 living organisms. Alternatively, a role for aromatic catabolism in pathogenesis has
 268 been proposed in *Burkholderia* (Law *et al.*, 2008) as expanded upon in the last
 269 section. Interestingly, the lineage comprising strains *B. phymatum*, *B. graminis*, *B.*
 270 *xenovorans*, *B. phytofirmans*, *Burkholderia* sp. CCGE1001, CCGE1002,
 271 CCGE1003, Ch1-1, and H160, which seems to be composed of bacteria that have
 272 in common the ability to interact beneficially with plants (Table S1) (Balandreau
 273 and Mavingui, 2006), possesses a different and greater catabolic potential towards

aromatic compounds (Fig. 2). The *Cupriavidus* and *Ralstonia* genera, also belonging to the *Burkholderiaceae* family have catabolic properties closely related to that divergent *Burkholderia* group, and are also able to establish beneficial interactions with plants, except for *R. solanacearum* strains (Amadou *et al.*, 2008; Kataoka and Futai, 2009). The broader catabolic abilities of rhizobacterial strains compared to pathogenic members of *Burkholderiaceae* family is reliably due to the nutrient-rich niche generated as the result of exudation of a large diversity of compounds, a lot of these having aromatic structures (Bais *et al.*, 2006).

Although a high heterogeneity is observed among members of other *Burkholderiales* families, some gross trends can also be detected (Fig. 2). In contrast to members of the *Burkholderiaceae* family that seem to be relatively similar in their metabolic properties, *Comamonas*, *Variovorax*, *Deftia*, and *Polaromonas* strains, members of the *Comamonadaceae* family, have a quite distinguishable catabolic pattern, whereas the other genera of this family, such as *Acidovorax*, *Alicyclophilus* or *Methylobium* have a restricted aromatic catabolic potential. The *Alcaligenaceae* family harbors two *Achromobacter* species with interesting catabolic properties, whereas among *Bordetella* species, mainly human and animal pathogens, only the sediment isolate *B. petrii* has remarkable catabolic abilities, which however are typically encoded on genomic islands (Gross *et al.*, 2008). Members of the *Oxalobacteraceae* family encode very few aromatic catabolism markers (Fig. 2), with the sole exception of *H. seropedicae*, which is the only representative associated with plants (Table S1).

296 The relative content of aromatic oxygenases encoding genes belonging to
 297 peripheral and central catabolic pathways was compared in *Burkholderiales*
 298 genomes classified by isolation source in three gross groups: H/Z (animal hosts),
 299 R/S (bulk and rhizospheric soils) or W/D (wastewater or sludge and sediment); or
 300 by its phylogenetic affiliation considering the three main families showing catabolic
 301 potential: *Burkholderiaceae*, *Comamonadaceae* and *Alcaligenaceae* (Fig. 4). It can
 302 be observed that the mean of the relative content of catabolic oxygenases is nearly
 303 similar among the three families but a bias towards non host-associated strains is
 304 found. The patterns of catabolic gene content among groups are similar for
 305 peripheral or central pathways. Statistical analysis (t-test) confirms that these
 306 differences in biodegradative potential regarding host- or non-host association are
 307 all significant at a confidence level of 95%. The fact that oxygenase content shows
 308 disparity when grouped by isolation source and not when clustered by taxonomical
 309 family, suggests that shaping of the catabolic potential is more influenced by
 310 habitat than by phylogenetic origin.

311 The overall genomic distribution of catabolic oxygenases encoding genes of
 312 four main ring-cleavage pathways (catechol, protocatechuate, gentisate and
 313 homogentisate) and their peripheral reactions was more thoroughly analyzed in six
 314 selected *Burkholderiales* strains (Fig. S1 of Supplementary Material). The strains
 315 *B. lata* 383, *B. xenovorans* LB400, *C. pinatubonensis* JMP134, *R. pickettii* 12J, *D.*
 316 *acidovorans* SPH-1 and *B. petrii* DSM 12804 were selected based on their superior
 317 catabolic properties, representing six relevant taxonomic groups within
 318 *Burkholderiales*. In general, a supraoperonic clustering of catabolic genes that

319 encode enzymes channelling different aromatic compounds into a common central
 320 pathway (catabolic islands), as has been reported in the γ -proteobacterium
 321 *Acinetobacter baylyi* ADP1, a well-known catabolic strain belonging to
 322 *Pseudomonadales* (Elsemore and Ornston, 1994), is not observed on the genomes
 323 of the six strains (Fig. S1). On the other hand, the genomes of *B. petrii* DSM 12804
 324 and *B. xenovorans* LB400 possess several catabolic genes included in genomic
 325 islands, suggesting that horizontal gene transfer played a significant role in
 326 establishing the catabolic abilities of these strains (Gross *et al.*, 2008; Chain *et al.*,
 327 2006). In order to roughly evaluate the role of horizontal gene transfer in the
 328 establishment of the catabolic markers in these six selected *Burkholderiales*
 329 strains, the G+C content of the biodegradative genes and their genomes were
 330 compared. The G+C content of the catabolic gene markers shows values not
 331 significantly different from the mean G+C content of all genes in the respective
 332 replicons, suggesting that most catabolic genes were ‘trapped’ within their
 333 genomes over a long period of evolution or, alternatively, being acquired from
 334 genomes having a similar G+C content (Fig. S1). Additional sequence analyses,
 335 using IslandViewer (Langille and Brinkman 2009), clarify a putative allochthonous
 336 origin for these catabolic markers. Except in a few cases, these analyses confirm
 337 that most of catabolic genes considered are not included in genomic islands
 338 putatively acquired by horizontal transfer (Fig. S1). It should be noted that the
 339 catabolic pathways considered in the preceding analyses are the most widely
 340 distributed and consequently does not constitute an “exotic” feature, as the routes
 341 previously reported to be acquired by horizontal gene transfer in some of these

bacteria (Gross *et al.*, 2008; Chain *et al.*, 2006). Finally, regarding the overall genomic distribution of catabolic markers, no evident similarities or patterns in gene distribution through the selected genomes are observed among these six strains. Noteworthy, even the catabolic gene scattering in the *Burkholderia* strains considered (both possessing similar tripartite genomic structure) differs drastically (Fig. S1).

Due to its widespread distribution and essential role in catabolism of a wide array of compounds, an in-depth analysis on the phylogenetic relationship and gene clustering of the central pathways channelling four key ring-cleavage intermediates, i.e. protocatechuate, catechol, homogentisate and gentisate was performed. The results of this analysis are presented in the following sections.

353

Protocatechuate ring cleavage pathways and peripheral routes

Protocatechuate (Pca) is a key intermediate in the degradation of the plant biopolymer lignin and several other aromatic compounds (Fig. 1). In *Burkholderiales*, it has been reported that Pca is the ring-cleavage substrate in the degradation of 3-hydroxybenzoate (Chang and Zylstra, 2008), 4-hydroxybenzoate (Zylstra *et al.*, 1989), isovanillate, vanillate, veratrate, phthalate, isophthalate (Chang and Zylstra, 1998; Providenti *et al.*, 2001; 2006a), terephthalate (Sasoh *et al.*, 2006), 3-chlorobenzoate (Nakatsu and Wyndham, 1993) and 3-nitrobenzoate (Providenti *et al.*, 2006b), among others. The ring cleavage of Pca can be performed by two distinct mechanisms, either through intradiol cleavage by

364Pca34DO, a member of the intradiol dioxygenase superfamily or through extradiol
 365cleavage by Pca45DO, a member of the LigB superfamily. Pca34DO is an enzyme
 366composed of two different subunits which, however, share substantial amino acid
 367identity (Yoshida *et al.*, 1976). Typically adjacently located *pcaGH* genes, broadly
 368distributed in *Cupriavidus*, *Ralstonia* and *Burkholderia* strains, encode these
 369subunits (Fig. 5). Analysis of the phylogenetic relationship among the α subunits of
 370Pca34DO (*pcaG* gene) indicates that the Pca34DO gene products from *Ralstonia*
 371strains are more closely related to those of an α - and γ -proteobacteria than to the
 372Pca34DO gene products from *Burkholderia* spp. (Fig. 5).

373 In *Burkholderia* and *Ralstonia* strains, *pcaGH* genes are not clustered with
 374any other genes encoding enzymes performing peripheral reactions generating
 375Pca, or with genes encoding enzymes performing subsequent reactions of the Pca
 376*ortho* cleavage pathway (Fig. 5). In contrast, in all three *Cupriavidus* strains, which
 377harbor a protocatechuate intradiol cleavage pathway, *pcaGH* genes are clustered
 378with genes coding for Phb3H (*pobA*) transforming 4-hydroxybenzoate to
 379protocatechuate, as well as with genes encoding 3-carboxy-*cis,cis*-muconate
 380cycloisomerase (*pcaB*), a fusion enzyme composed of 4-carboxymuconolactone
 381decarboxylase and 3-oxoadipate enol-lactone hydrolase domains (*pcaL*) and a
 382putative transporter for Pca (Fig. 5). It should be noted that the clustering of
 383Pca34DO encoding genes with subsequent Pca *ortho* cleavage pathway genes
 384and/or genes encoding enzymes of peripheral pathways channelling to Pca is
 385commonly observed in α - and γ -proteobacterial strains, although a different
 386organization is found in *Cupriavidus* strains (Fig. 5).

387 It is remarkable that only two *Burkholderiaceae* strains *C. taiwanensis*
 388 LMG19424 and *B. rhizoxinica* HKI 0454- lack Pca34DO. Since even pathogenic
 389 strains maintain such genes in their genomes, this pathway can be assumed to be
 390 highly relevant for the environmental fitness of free-living strains. *C. taiwanensis*
 391 LMG19424 is a nitrogen-fixing symbiont of legumes and *B. rhizoxinica* HKI 0454 is
 392 an intracellular symbiont of the phytopathogenic fungus *Rhizopus microsporus*
 393 (Table S1). Both strains are noticeably less metabolically versatile than other
 394 *Cupriavidus* or *Burkholderia* strains (Fig. 2), (Amadou *et al.*, 2008; Lackner *et al.*,
 395 2011), indicating that during the evolution of their genomes to the symbiotic
 396 lifestyle, many catabolic traits were lost.

397 Only five *Burkholderiales* strains outside of the *Burkholderiaceae* family
 398 contain Pca34DO encoding genes (Fig. 2): strains *Polaromonas* sp. JS666, *V.*
 399 *paradoxus* S110 and *V. paradoxus* EPS of the *Comamonadaceae* family (these
 400 strains are exceptional since members of this family usually perform a *meta*
 401 cleavage of Pca, as described below), *A. piechaudii* ATCC43553 and *B. petrii* DSM
 402 12804. In the latter strain, Pca34DO genes are clustered with genes encoding
 403 subsequent Pca *ortho* cleavage pathway steps (Fig. 5), and are located in the
 404 proximity of genes encoding phthalate 3,4-dioxygenase (PhtDO). The acquisition of
 405 the Pca *ortho* cleavage pathway through horizontal transfer in *B. petrii* DSM 12804
 406 is suggested by the location of the respective gene cluster on a genomic island
 407 (Gross *et al.*, 2008).

408 With the sole exception of *B. petrii* DSM 12804 (Fig. 2), all *Burkholderiales*
 409 strains harboring Pca34DO encoding genes also have a Phb3H encoding gene,

410 indicating that 4-hydroxybenzoate is probably the major substrate degraded
 411 through the Pca *ortho* cleavage pathway. This is supported by the fact that
 412 downstream enzymes of the Pca *ortho* cleavage pathway (3-carboxymuconate
 413 *cis,cis* cycloisomerase, 4-carboxymuconolactone decarboxylase and enol-lactone
 414 hydrolase) are encoded in the same gene cluster as Phb3H in all strains of the
 415 "*Burkholderia cepacia* complex" as well as in *B. glumae* BGR1 and *B. phyatum*
 416 STM815. The gene order is conserved with genes encoding Phb3H being followed
 417 by genes encoding the α - and β -subunits of succinyl-CoA:3-oxoadipate CoA
 418 transferase, 3-carboxy-*cis,cis*-muconate cycloisomerase, 3-oxoadipate enol-
 419 lactone hydrolase and 4-carboxymuconolactone decarboxylase and preceded by a
 420 gene encoding a LysR type transcriptional regulator. A gene encoding a Major
 421 Facilitator Superfamily protein is additionally encoded in all such gene clusters of
 422 "*Burkholderia cepacia* complex" strains. This contrasts the situation in *Ralstonia*
 423 strains, where the Phb3H encoding gene is located separately from Pca *ortho*
 424 cleavage pathway encoding genes. Here, the genes encoding enzymes for
 425 catabolizing the Pca *ortho* cleavage product are organized in a cluster that also
 426 contains 3-oxoadipyl-CoA thiolase and thus encode enzymes for the
 427 transformation of 3-carboxy-*cis,cis*-muconate to succinyl-CoA and acetyl-CoA, in
 428 the order: α - and β -subunits of succinyl-CoA:3-oxoadipate CoA transferase, 3-
 429 oxoadipyl-CoA thiolase, 3-carboxy-*cis,cis*- muconate cycloisomerase, 3-oxoadipate
 430 enol-lactone hydrolase and 4-carboxymuconolactone decarboxylase.

431 Two different oxygenases channelling substrates to 4-hydroxybenzoate (Fig.
 432 1) are observed in *Burkholderiales*: DapDO found in ten *Burkholderia* strains and in

433 *C. testosteroni* KF-1, and PcmH found in *P. naphthalenivorans* CJ2 and *C.*
 434 *testosteroni* strains CNB-2 and S44 (Fig. 2). Other peripheral enzymes channelling
 435 substrates directly to Pca (Fig. 1), where encoding genes are observed in three,
 436 four, eight, thirteen and twenty seven *Burkholderiales* strains are lphDO, ThpDO,
 437 lvaOD, PhtDO, and VanOD, respectively, predominantly in *Burkholderia* and
 438 *Comamonadaceae* (Fig. 2). The broad diversity of peripheral oxygenases
 439 channelling lignin-derived compounds to protocatechuate in *Burkholderia* strains,
 440 which only comprise a Pca *ortho* cleavage pathway, strongly suggests a key role of
 441 *Burkholderia* for the recycling of plant derived carbon of aromatic compounds.

442 The Pca *meta* cleavage pathway has been mainly described in strains of the
 443 *Comamonadaceae* family (Providenti *et al.*, 2001; Mampel *et al.*, 2005). This
 444 cleavage reaction is catalyzed by a heteromultimeric Pca45DO composed by two
 445 unrelated subunits with the encoding genes typically located adjacent to each other
 446 (Fig. 6). Genome surveys show such genes to be predominantly localized in
 447 genomes of members of the *Comamonadaceae* family with seven out of sixteen
 448 strains where the complete genome sequence is available harboring such genes
 449 (Fig. 2). In contrast to the typical organization as a heterodimer, a fusion protein is
 450 encoded in the genome of *V. eiseniae* EF01-2 (Fig. 6). Only two *Burkholderiales*
 451 strains outside the *Comamonadaceae* family are found to harbor a Pca45DO gene,
 452 i.e. *L. cholodnii* SP-6, and *B. phymatum* STM815. The last case is remarkable
 453 since *B. phymatum* STM815 also possesses a Pca *ortho* cleavage pathway (Fig. 2,
 454 previous section, Fig. 5).

455 Analysis of the phylogenetic relationship among Pca45DO gene products
 456 shows that enzymes encoded within *Burkholderiales* genomes are closely related
 457 to each other and different from those of the α -proteobacterial
 458 *Sphingomonadaceae* family (represented by the enzyme from *S. paucimobilis*
 459 SYK-6, Fig. 6), the other proteobacterial family where Pca *meta* cleavage has been
 460 well described (Noda et al. 1990). Exceptional are the enzymes of *B. phymatum*
 461 STM815 and one of Pca45DO proteins encoded in the genomes of *C. testosteroni*
 462 strains CNB-2 and S44, which are only distantly related to both Pca45DO of other
 463 *Burkholderiales* as well as those of *Sphingomonadaceae* and thus can be
 464 assumed to be of a different origin (Fig. 6).

465 Some strains harbor multiple Pca *meta* cleavage gene clusters. The full set
 466 of genes encoding enzymes for the transformation of the Pca 4,5-dioxygenolytic
 467 ring-cleavage product, 4-carboxy-2-hydroxymuconate-6-semialdehyde to pyruvate
 468 and oxaloacetate (4-carboxy-2-hydroxymuconate semialdehyde dehydrogenase, 2-
 469 pyrone-4,6-dicarboxylate hydrolase, 4-oxalomesaconate hydratase and 4-carboxy-
 470 4-hydroxy-2-oxoadipate aldolase) are typically clustered with the Pca45DO-
 471 encoding gene in one gene cluster (Fig. 6), except in *V. eiseniae* EF01-2, where
 472 the cluster lacks an aldolase-encoding gene. Typically, the organization of these
 473 gene clusters in *Burkholderiales* strains follows a hydratase-aldolase-hydrolase-
 474 Pca45DO-dehydrogenase gene order, with the insertion of genes of unknown
 475 function in a few cases. This gene organization clearly differs from the organization
 476 of Pca *meta* cleavage gene clusters in *Sphingomonas* strains (Fig. 6). Recently, a
 477 putative 4-oxalomesaconate tautomerase required for Pca degradation in

478 *Comamonas* sp. strain E6 has been reported (Kamimura *et al.*, 2010). Respective
 479 genes are also observed in most Pca45 *meta* cleavage gene clusters (Fig. 6).

480 Several oxygenases have been described to transform lignin-derived
 481 compounds to Pca followed by Pca *meta* cleavage (Fig. 1), especially in
 482 *Comamonadaceae* strains (Fig. 2). All *Burkholderiales* strains harboring a Pca
 483 *meta* cleavage pathway also own a gene encoding the 4-hydroxybenzoate
 484 hydroxylase (Phb3H) (Fig. 2), although Phb3H-encoding genes are not included in
 485 Pca *meta* cleavage gene clusters. An exception to this are the gene clusters of *R.*
 486 *ferrireducens* T118 and one of the gene clusters of the *C. testosteroni* strains CNB-
 487 2 and S44 (Fig. 6), where one cluster comprises genes encoding for Phb3H and
 488 Pca45DO, and a second cluster comprising genes for the full downstream
 489 catabolic pathway. In contrast to Phb3H-encoding genes, 3-hydroxybenzoate
 490 hydrolase (Mhb4H)-encoding genes are seldom found in *Burkholderiales* and
 491 among strains harboring a Pca *meta* cleavage, are found only in the three
 492 sequenced *C. testosteroni* strains, two *Burkholderia* spp. strains and *B. phymatum*
 493 STM815 (Fig. 2). Vanillate 3-O-demethylase (VanOD) and phthalate 3,4-
 494 dioxygenase (PhtDO) are observed in eight and five of the *Burkholderiales* strains
 495 owning a Pca *meta* cleavage pathway, respectively. It is noteworthy that the full set
 496 of five Rieske-type oxygenases acting on the three phthalate isomers, vanillate and
 497 isovanillate are found in *C. testosteroni* KF-1 (Fig. 2) and four or two of these in *C.*
 498 *testosteroni* CNB-2 and SA44, respectively, in agreement with previous reports on
 499 other *Comamonas* strains (Providenti *et al.*, 2006a; Sasoh *et al.*, 2006; Fukuhara
 500 *et al.*, 2010).

501 Gallate dioxygenases, like Pca45DO belonging to the LigB superfamily of
 502extradiol dioxygenases, have so far been described in *S. paucimobilis* SYK-6
 503(Kasai *et al.*, 2005) and *P. putida* KT2440 (Nogales *et al.*, 2005), and are specific
 504for this substrate and do not transform Pca. Both gallate dioxygenases have sizes
 505significantly larger than those of the β -subunits of Pca45DO and were suggested to
 506have evolved from the fusion of large and small subunits. Respective proteins are
 507encoded in thirteen *Burkholderiales* genomes and based on their similarities could
 508be grouped into two types (GI and GII in Fig. 6). The ring-cleavage of gallate gives
 5094-oxalomesaconate, and thus, the transformation of gallate to pyruvate and
 510oxaloacetate requires only a 4-oxalomesaconate hydratase and a 4-carboxy-4-
 511hydroxy-2-oxoadipate aldolase as downstream enzymes. Accordingly, all gene
 512clusters encoding a gallate dioxygenase of type II are devoid of genes encoding 4-
 513carboxy-2-hydroxymuconic semialdehyde dehydrogenase and 2-pyrone-4,6-
 514dicarboxylate hydrolase (Fig. 6), enzymes necessary only for Pca degradation.
 515This supports the notion that these gene clusters have evolved for gallate
 516degradation (Nogales *et al.* 2011). Gene clusters encoding a type I gallate
 517dioxygenase encode only the ring-cleavage activity. In this case, a second gene
 518cluster comprising genes encoding, among functional genes, a 4-oxalomesaconate
 519hydratase and a 4-carboxy-4-hydroxy-2-oxoadipate aldolase, are located
 520elsewhere in the chromosome. It should be noted that genes encoding putative 4-
 521oxalomesaconate tautomerase are also usually comprised in gallate catabolic
 522gene clusters, however, their function remains to be elucidated.

523

524 Catechol ring cleavage pathways and peripheral routes

525 Major work on elucidation of catechol degradation has so far been
 526 performed in *Pseudomonas* strains (Harwood and Parales, 1996; Jimenez *et al.*,
 527 2002). These organisms typically harbor a *catRBCA* gene cluster, encoding a LysR
 528 type regulator (*catR*), a catechol 1,2-dioxygenase (Cat12DO) forming muconate
 529 (*catA*), a muconate cycloisomerase forming muconolactone (*catB*) and a
 530 muconolactone isomerase forming enol-lactone (*catC*), which is recognized as a
 531 common intermediate in the degradation of catechol and protocatechuate, and
 532 then hydrolyzed by a enol-lactone hydrolase typically not encoded in the *cat* gene
 533 cluster. The *catBCA* cluster (*cat* operon) thus comprises all genes exclusively
 534 responsible for catechol degradation forming the catechol branch of the 3-
 535 oxoadipate pathway (Fig. 1). A complete catechol branch is frequent only in
 536 *Burkholderiaceae*, as nearly all *Burkholderia*, *Cupriavidus* and *Ralstonia* genomes
 537 analyzed so far possess such genes (Fig. 2). All strains of the “*B. cepacea*
 538 complex” harbor a *catRBAC* gene cluster, contrasting the typical *catRBCA* order
 539 observed in *Pseudomonads* (Fig. 7). These closely related gene clusters (type Ia)
 540 are preceded by genes encoding an AntDO anthranilate dioxygenase (Fig. 7),
 541 indicating the major role of the catechol branch to channel anthranilate formed
 542 during the degradation of tryptophan into the 3-oxoadipate pathway. Even though
 543 the same gene order is conserved in *cat* operons of *B. phymatum*, *B. phytofirmans*,
 544 *B. xenovorans* and *Burkholderia* sp. Ch1-1, the respective genes are not located in
 545 a cluster with AntDO (Fig. 7, type Ib). However, both *Burkholderia* sp. Ch1-1 and
 546 *B. xenovorans* harbor an AntDO localized elsewhere (Fig. 2). Despite the absence

547of AntDO encoding genes in some strains outside the *B. cepacia* complex, they are
 548also capable of degrading anthranilate, due to the presence of a gene cluster
 549allowing degradation via aminobenzoyl-CoA (Abc). This pathway is typically absent
 550from *Burkholderia* strains of the “*B. cepacia* complex” (Fig. 2). Like *Burkholderia*
 551strains outside of the *B. cepacia* complex, all *Cupriavidus*, *Ralstonia*, four out of
 552five *Bordetella* spp., as well as the majority of members of the *Comamonadaceae*
 553family harbor the anthranilate degradative pathway via aminobenzoyl-CoA,
 554whereas AntDO is restricted to members of the *Burkholderia* genus. The only
 555exception to this is *H. seropedicae* SmR1, where an AntDO encoding gene
 556precedes the muconate cycloisomerase encoding gene (Fig. 7). Interestingly, in *B.*
 557*phytofirmans*, *B. xenovorans* and *Burkholderia* sp. Ch1-1 the *catBAC* gene cluster
 558is framed by genes encoding a Rieske non-heme iron oxygenase, where the
 559deduced sequences of the α -subunits share >60% of identity with that of *lacC*
 560gene supposed to be involved in indole 3-acetate degradation by *P. putida* 1290
 561(Leveau and Gerards, 2008).

562 Besides the above mentioned type Ia gene cluster, a subset of strains of the
 563“*B. cepacia* complex” (Fig. 2) comprises a second *cat* gene cluster (type IV, Fig. 7),
 564which is typically connected with genes coding for a benzoate dioxygenase
 565(BenDO). As an example, in *B. multivorans* and *B. ambifaria*, a complete *catBCA*
 566cluster is connected to genes encoding for subunits of BenDO. This organization
 567resembles that previously observed in the γ -proteobacterium *K. pneumoniae* 342
 568(Fouts *et al.*, 2008), and the *cat* gene organization is also similar to that typically
 569observed in *Pseudomonads* (Jimenez *et al.*, 2002). Accordingly, the *catA* gene

570products of these type IV clusters are more closely related to γ -proteobacterial
 571*catA* gene products than to type I *catA* gene products (Fig. 7), indicating a common
 572origin.

573 All "*B. cepacia* complex" strains encoding type IV *cat* gene cluster also
 574harbor a complete type I *cat* gene cluster. For this reason, it can be argued that a
 575second complete gene cluster may be an environmental burden. Accordingly,
 576different variants of the type IV cluster harboring *catBA* or even only *catA* genes
 577are observed (Fig. 7). This indicates the presence of a second gene copy of *catA*
 578to be of a selective advantage. A reason for this could be the necessity to avoid the
 579accumulation of highly toxic catechol intermediates (Schweigert *et al.*, 2001).
 580Incorporation of *catA* in a benzoate gene cluster would avoid such accumulation,
 581but directly produce the non toxic *cis,cis*-muconate, which is reported as an inducer
 582of the type I *cat* gene cluster of *Burkholderia* sp. TH2 (Suzuki *et al.*, 2002) and of
 583*Pseudomonas cat* gene clusters (Parsek *et al.*, 1992). Besides strains of the "*B.*
 584*cepacia* complex", *B. graminis*, *B. glumae*, *B. lata*, *B. phytofirmans*, as well as
 585various *Burkholderia* strains (uncharacterized to the species level) and *R. pickettii*
 586strains harbor gene clusters comprising a *catA* gene phylogenetically related to
 587those of the type IV cluster. Such genes are absent from *B. xenovorans* and *B.*
 588*phydatum* (Fig. 7).

589 The localization of *cat* genes in a cluster with *ben* genes is also observed in
 590three of the four sequenced *Cupriavidus* strains and in two of the three *R.*
 591*solanacearum* strains (Fig. 7, type III). They are absent from the genomes of *C.*
 592*taiwanensis*, which is devoid of most pathways involving dihydroxylated
 593intermediates, and also of *R. solanacearum* GMI1000. Even though the structure of

these gene clusters is similar to type IV gene clusters defined above, it should be noted that these *catA* gene products are phylogenetically more closely related to type I *catA* gene products (Fig. 7), indicating the association of *cat* genes into *ben* gene clusters to have occurred at least twice in evolution.

Both *C. pinatubonensis* JMP134 and *C. metallidurans* CH34 carry a second *catA* gene copy (Fig. 7, type II). In both strains, genes encoding a multicomponent phenol monooxygenase (PMO) precede the second copy. Noteworthy, in strains so far described as capable of degrading phenols via PMO, degradation of the catechol intermediate occurs by extradiol cleavage through catechol 2,3-dioxygenases (Cat23DO) (Nordlund and Shingler, 1990; Nordlund *et al.*, 1993; Merimaa *et al.*, 2006). In fact, both strains harbor two copies of PMO encoding genes, with the second copy of the genes localized in a gene cluster with a Cat23DO. Multicomponent PMO are typically of broad substrate specificity and transform methylphenols with high activity (Powlowski and Shingler, 1994; Watanabe *et al.*, 1998). Such activity would channel methylphenols in the *ortho* cleavage pathway, which is usually not suited to mineralize them (Catelani *et al.*, 1971; Knackmuss *et al.*, 1976). However, early studies on *C. pinatubonensis* JMP134 have indicated a high Cat12DO activity only after growth with phenol and not with methylphenols (Pieper *et al.*, 1989), indicating a specific regulation that restricts the flux of methylaromatics into the *ortho*-cleavage pathway. In addition, strain JMP134 is one of the few strains reported thus far which can mineralize methylaromatics via a catechol intradiol cleavage pathway due to the presence of a catabolic gene module encoding enzymes for 4-methylmuconolactone transformation (Erb *et al.*, 1998), which is otherwise a dead-end product for various

618strains (Catelani *et al.*, 1971; Knackmuss *et al.*, 1976). Interestingly, such genes,
 619which could have been recruited in strain JMP134 for 4-methylphenol metabolism
 620via the *ortho* cleavage pathway, are absent from the genome of *C. metallidurans*
 621CH34, but observed on the pHG1 megaplasmid of *C. necator* H16 that harbors
 622only a single phenol hydroxylase gene cluster containing a Cat23DO gene.

623 Out of the sixteen *Comamonadaceae* strains sequenced, only *D.*
 624*acidovorans* SPH-1, *Acidovorax* sp. JS42, *Acidovorax ebreus* TPSY, *Alicyclophilus*
 625*denitrificans* BC and *P. naphthalenivorans* CJ2 harbor a *cat* gene cluster (Figs. 2
 626and 7), indicating such activity is not a common trait in this family. Similarly, in the
 627*Alcaligenaceae*, only both sequenced *Achromobacter* species and *B. petrii* harbor
 628a catechol *ortho* cleavage pathway. The genome sequence of *B. petrii*, however,
 629indicates this gene cluster to be located on a genomic island and thus to have
 630been recruited from foreign DNA (Gross *et al.*, 2008).

631 The extradiol ring-cleavage of catechol and methylsubstituted catechols is
 632typically catalyzed by family I.2 type I extradiol dioxygenases (Eltis and Bolin,
 6331996). Contrasting genes encoding Cat12DO, which are observed in nearly all
 634*Burkholderiaceae*, Cat23DO genes are found only in eight *Burkholderiaceae*,
 635among them three of the four sequenced *Cupriavidus* strains (Figs. 2 and 8).
 636Outside the *Burkholderiaceae*, Cat23DO genes are observed in five
 637*Comamonadaceae*, including *A. denitrificans* BC proposed to grow on benzene via
 638extradiol cleavage of catechol using oxygen generated during chlorate respiration
 639(Weelink *et al.*, 2008), as well as in *Limnobacter* sp. MED105, *L. cholodnii* SP-6
 640and *M. petroleiphilium* PM1.

641 Interestingly, out of the nineteen Cat23DO encoding genes, which are
 642 observed in eighteen *Burkholderiales* gene clusters (the gene cluster of
 643 *Acidovorax* sp. JS42 comprises only two distantly related genes encoding
 644 Cat23DO), all but one are observed in gene clusters encoding a multicomponent
 645 PMO (Fig. 8). Such a high abundance of respective gene clusters in
 646 *Burkholderiales* is astonishing, especially taking into account that these
 647 microorganisms are not typically isolated based on their capability to degrade
 648 phenol. Noteworthy, such a gene cluster is the only aromatic catabolic trait
 649 observed in *Limnobacter* sp. MED105 (Fig. 2).

650 Cat23DO of *Burkholderiales* can be phylogenetically grouped into three
 651 distinct clusters (Fig. 8), with the majority of them being related to Cat23DO CbzE
 652 of *P. putida* GJ31 (Kaschabek *et al.*, 1998) or TbuE of *R. pickettii* PK01 (Kukor and
 653 Olsen, 1996) involved in chlorobenzene or toluene degradation (subfamily I.2.C
 654 according to the classification by Eltis and Bolin, 1996). However, the proteins of
 655 *Limnobacter* sp. MED105, and *C. necator* H16 as well as one of the proteins of *M.*
 656 *petroleiphilum* PM1 belong to a distinct protein cluster more similar to subfamily
 657 I.2.A proteins, such as XylE of *P. putida* mt-2 (Fig. 8), indicating that Cat23DO
 658 phylogeny does not follow strain phylogeny. Three further Cat23DO proteins (Fig.
 659 8) are related to LapB involved in alkylphenol degradation by *Pseudomonas* sp.
 660 strain KL28 (Jeong *et al.*, 2003).

661 Interestingly, gene clusters comprising subfamily I.2.C Cat23DO highly
 662 similar to the ones found in *Burkholderiales* genomes, have also been reported in
 663 previous studies to be frequently clustered with genes encoding aniline

664dioxygenase. One such case is *Delftia* sp. AN3 (Liu *et al.*, 2002). However, such
 665an organization is not observed in any of the sequenced genomes. The clustering
 666of highly similar *meta* cleavage pathways with distinct peripheral reactions may
 667indicate that these pathways have been formed recently by the combination of
 668appropriate pathway gene modules.

669 Four of the gene clusters also contain genes encoding for a
 670benzene/toluene monooxygenase (Leahy *et al.*, 2003), which should allow the
 671hosts, in concert with phenol monooxygenase, to grow on benzene/toluene via two
 672successive monooxygenations. However, some phenol monooxygenases, among
 673them toluene 2-monooxygenase of strain G4, sequentially oxidize toluene to 2-
 674methylphenol and further to 3-methylcatechol (Newman and Wackett, 1995), even
 675though, as evident for phenol monooxygenase of *P. stutzeri* OX1, phenol seems to
 676be the highly preferred substrate (Cafaro *et al.*, 2004). Thus, the fact that *A.*
 677*denitrificans* BC is capable of growth on benzene in the absence of a
 678benzene/toluene monooxygenase (and a Rieske non heme iron oxygenase
 679capable to dioxygenate benzene) may be due to the broad substrate specificity of
 680the respective phenol monooxygenase.

681 The degradation of catechol and 4-methylcatechol (and thus of phenol and
 6824-methylphenol) requires the presence of the dehydrogenase branch of the *meta*-
 683cleavage pathway suited for metabolism and specifically a 2-hydroxymuconic
 684semialdehyde dehydrogenase for rapid dehydrogenation of the ring-cleavage
 685products 2-hydroxymuconic or 2-hydroxy-4-methylmuconic semialdehyde
 686(Harayama *et al.*, 1987), followed by isomerization and decarboxylation to yield 2-

687hydroxypent-2,4-dienoate. In contrast, the degradation of 3-methylcatechol (formed
688from 2-methylphenol) requires the presence of a 2-hydroxymuconic semialdehyde
689hydrolase for rapid hydrolysis of the ketonic ring-cleavage product 2-hydroxy-6-
690oxo-2,4-heptadienoate to yield 2-hydroxypent-2,4-dienoate (Harayama *et al.*,
6911987). This central intermediate of the hydrolytic and dehydrogenase branch is
692further dissembled by a hydratase giving 4-hydroxy-2-oxovalerate, and an aldolase
693giving pyruvate and acetaldehyde. All respective genes are, for example, present
694in the archetype *meta* cleavage pathway gene cluster localized on the pWW0 TOL
695plasmid of *P. putida* mt-2 (Harayama and Rekik, 1990), together with an
696acetaldehyde dehydrogenase encoding gene. Moreover, typically a ferredoxin
697encoding gene is observed upstream of the gene encoding Cat23DO, and the
698respective proteins have been reported to have a reactivating function through
699reduction of the iron atom in the active site of the enzyme (Hugo *et al.*, 1998).
700Similar ferredoxins are observed to be encoded upstream of the Cat23DO genes in
701*Burkholderiales* (Fig. 8), indicating the gene product to be of a significant selective
702advantage. Also typically, both a 2-hydroxymuconic semialdehyde hydrolase and
703enzymes of the dehydrogenase branch are encoded in *meta* cleavage pathway
704gene clusters of *Burkholderiales*, together with genes encoding downstream
705enzymes (Fig. 8). Exceptions to this are the gene clusters of *C. necator* H16, which
706should thus not be capable of growth on 2-methylphenol via 3-methylcatechol and
707one of the two gene clusters of *M. petroleiphilum* PM1. As is the case for *B. lata*
708383, where one of the gene clusters is devoid of genes encoding proteins of the
709dehydrogenase branch, also *M. petroleiphilum* PM1 harbors one complete gene
710cluster, thus allowing degradation of differently substituted catechols.

711

712

713 *Homogentisate pathway and peripheral reactions*

714 Homogentisate is a central metabolite formed during the aerobic catabolism
715 of aromatic amino acids such as phenylalanine and tyrosine in many bacteria
716 (Arias-Barrau *et al.*, 2004). These amino acids are converted to 4-
717 hydroxyphenylpyruvate and subsequently transformed into homogentisate by a 4-
718 hydroxyphenylpyruvate dioxygenase (HppDO) (Fitzpatrick, 2003). This enzyme is a
719 Fe^{2+} dependent, non-heme oxygenase that catalyzes the conversion of 4-
720 hydroxyphenylpyruvate to homogentisate in a reaction involving decarboxylation,
721 substituent migration and aromatic oxygenation (Morán, 2005) (Fig. 1). 3-
722 hydroxyphenylacetate is also channelled into the homogentisate pathway, as
723 described in *Pseudomonas* and *Burkholderia* species (Arias-Barrau *et al.*, 2005;
724 Méndez *et al.*, 2011). 3-hydroxyphenylacetate monooxygenase has been
725 described in *P. putida* strain U (Arias-Barrau *et al.*, 2005) as being composed of
726 the hydroxylase and a small coupling protein, constituting a novel type of two-
727 component hydroxylase, distinct from the classical two-component flavoprotein
728 monooxygenases.

729 The homogentisate pathway converts homogentisate into fumarate and
730 acetoacetate (Arias-Barrau *et al.*, 2004). The homogentisate dioxygenase (Hge),
731 an enzyme of the cupin superfamily of extradiol dioxygenases (Dunwell *et al.*,
732 2000) cleaves the aromatic ring between the acetyl substituent and the proximal

733hydroxyl group, yielding maleylacetoacetate. This product is isomerized by a GSH-
 734dependent maleylacetoacetate isomerase (Suemori *et al.*, 1995) into
 735fumarylacetoacetate, which in turn is hydrolyzed by a fumarylacetoacetate
 736hydrolase to give fumarate and acetoacetate. However, in some bacteria
 737maleylacetoacetate can be directly hydrolyzed into maleate and acetoacetate
 738(Crawford, 1976).

739 In *Burkholderiales*, the homogentisate pathway is highly represented (Figs.
 7402 and 9) and can be observed in most sequenced members of the families
 741*Alcaligenaceae* (except for *Taylorella equigenitalis*) and *Burkholderiaceae* (except
 742for *B. rhizoxinica*, *Poynucleobacter* spp, and *Limnobacter* sp.), but is absent from
 743most *Comamonadaceae* (with the exception of *Polaromonas* sp. JS666, where
 744such a pathway is plasmid encoded, and *Variovorax* spp.). The catabolic operons
 745(*hmg* gene clusters) typically comprise genes encoding Hge (*hmgA*) and
 746fumarylacetoacetate hydrolase (*hmgB*). Fumarylacetoacetate hydrolases form a
 747monophyletic group and share >45% of sequence identity and >50% of sequence
 748identity with fumarylacetoacetate hydrolase from *P. putida* U. Genes encoding
 749maleylacetoacetate isomerase are typically absent from the *hmg* gene clusters.
 750This is in contrast to the organization in other proteobacteria such as *P. putida*, *P.*
 751*fluorescens*, *S. meliloti*, and *S. pomeroyi* (Fig. 9), where *hmg* gene clusters
 752typically comprise a maleylacetoacetate isomerase-encoding gene.

753 Different types of *hmg* gene clusters can be easily defined. Members of the
 754*Burkholderiaceae* family usually comprise *hmgAB* genes framed by genes
 755encoding up to three transporters of the major facilitator superfamily (MFS) where

the proteins encoded by genes preceding *hmgA* share significant similarity to the PcaK 4-hydroxybenzoate transporter of *P. putida* PRS2000 (>30% sequence identity) (Harwood *et al.*, 1994). A gene encoding a flavoprotein monooxygenase typically precedes the *hmg* gene clusters of *Burkholderia* strains, and also in various other *Burkholderiales* such as *Bordetella* spp, *R. pickettii* or *Polaromonas* sp. JS666. In a phylogenetic analysis, the flavoprotein monooxygenases encoded by *Polaromonas* sp JS666 and *B. xenovorans* LB400 (BxeA2725) cluster with those encoding 3-hydroxyphenylacetate 6-hydroxylase of *P. putida* U (44 and 47% of identity, respectively) and can be assumed to encode 3-hydroxyphenylacetate 6-hydroxylases. The other flavoprotein monooxygenases encoded in *hmg* gene clusters share >63% of sequence identity but are distantly related to 3-hydroxyphenylacetate 6-hydroxylases (34-44% of sequence identity). Their conserved inclusion in the *hmg* gene cluster indicates that they channel substrates to homogentisate, however, their actual function remains to be elucidated.

At first glance, the organization of *hmg* gene clusters outside of members of the *Burkholderia* genus seems to be relatively conserved as far as a transcriptional regulator precedes the *hmgAB* genes. In *P. putida* U, a divergently transcribed IclR family regulator regulates *hmgAB* catabolic genes (Arias-Barrau *et al.*, 2004). Such an organization seems the exception in *Burkholderiales*, since in sequenced members of the *Alcaligenaceae* family and *Ralstonia* sp. the cluster is preceded by a LysR type regulator-encoding gene, and in *Cupriavidus* sp. by a TetR regulator-encoding gene. The gene cluster of *Polaromonas* JS666 comprises a MarR type regulator encoding gene and one of the additional gene clusters of *B. xenovorans*

779LB400 (BxeC0994) a LuxR type regulator encoding gene. Typically, only one *hmg*
 780gene cluster is present, with the exception of *Burkholderia* sp. CCGE1002 (two
 781clusters), *Burkholderia* sp. Ch1-1 (two clusters and one containing a truncated Hge
 782encoding gene) and *B. xenovorans* LB400 containing an impressive four such
 783gene clusters. Recently it has been reported that *B. xenovorans* LB400 degrades
 7843-hydroxyphenylacetate and 4-hydroxyphenylacetate through both homogentisate
 785and homoprotocatechuate pathways, suggesting a very complex regulation of
 786these four Hge genes (Méndez *et al.*, 2011).

787 Among pathways channelling to homogentisate, a homogentisate forming 4-
 788hydroxyphenylpyruvate dioxygenase is present in all strains harboring an *hmg*
 789gene cluster but it is always located elsewhere on the genome. Also, a tyrosine
 790forming phenylalanine hydroxylase is present in nearly all such strains except
 791*Achromobacter* and *Bordetella* strains. Interestingly, such genes are also observed
 792in strains where the *hmg* pathway is absent, such as *Acidovorax* or *Comamonas*
 793strains.

794

795 *Gentisate pathway and peripheral reactions*

796 Gentisate (2,5-dihydroxybenzoate) is a central metabolite formed during the
 797catabolism of diverse aromatic compounds such as salicylate (2-hydroxybenzoate),
 7983-hydroxybenzoate and some phenolic compounds including *m*-cresol, 2,5-xyleneol
 799and 3,5-xyleneol (Poh and Bayly, 1980), and even has a key role in the degradation
 800of several polycyclic aromatic compounds such as naphthalene and phenanthrene

801(Lee *et al.*, 2011; Tittabutr *et al.* 2011). In the gentisate central pathway, gentisate
 8021,2-dioxygenase (Gen), a member of the cupin superfamily, cleaves the aromatic
 803ring between the carboxyl substituent and the proximal hydroxyl group to yield
 804maleylpyruvate. Maleylpyruvate is isomerized by a glutathione-dependent
 805maleylpyruvate isomerase to fumarylpyruvate. A fumarylpyruvate hydrolase
 806hydrolyzes fumarylpyruvate to fumarate and pyruvate. Maleylpyruvate can also be
 807converted into pyruvate and maleate by a maleylpyruvate hydrolase in a
 808glutathione-independent way (Crawford and Frick, 1977; Bayly *et al.*, 1980).

809 In contrast to the homogentisate pathway, the gentisate pathway is present
 810only in a few *Burkholderia* strains (Fig. 2), namely *B. cenocepacia*, *B. multivorans*
 811and *B. lata* 383 (but interestingly absent from closely related strains such as *B.*
 812*ambifaria*). It is also found in *Burkholderia* sp. CGE1001 and 1003 as well as in *B.*
 813*xenovorans* LB400. The gentisate pathway is widespread in *Comamonadaceae*
 814and *Alcaligenaceae*, and is present in *Achromobacter* and some *Bordetella* strains
 815(Fig. 2). The gentisate catabolic gene clusters in *Burkholderiales* (Fig. 10) typically
 816contain all three catabolic genes, an organization also observed in other bacteria.
 817Gentisate dioxygenases of *Burkholderiales* are quite diverse and share only down
 818to >30% of sequence identity. As it is shown in Fig. 10, they can be grouped into at
 819least five distinct clusters, where the identity among members of each cluster is
 820>50%. A similar diversity is seen in fumarylpyruvate hydrolases, which may also be
 821sorted into five distinct groups. Most putative fumarylpyruvate hydrolases encoded
 822in group I gentisate gene clusters are, among validly described fumarylpyruvate
 823hydrolases, most closely related to that of *K. pneumoniae* M5a1 (Jones and

824Cooper, 1990), whereas fumarylpyruvate hydrolases Bpro_0982 and Lcho_0798 of
 825*Polaromonas* sp. JS666 and *Leptothrix cholodnii* SP-6 are most closely related to
 826that of *P. putida* AK5 (ACO92376). Interestingly, fumarylpyruvate hydrolases
 827related to that of *Ralstonia* sp. U2 are not only encoded by most gentisate gene
 828clusters of group III but also by gentisate gene clusters of *Comamonas* strains
 829(group V). While all of the above mentioned fumarylpyruvate hydrolases share
 830>50% of sequence identity, hydrolase enzymes encoded by group II and IV gene
 831clusters are only distantly related to archetype fumarylpyruvate hydrolases, sharing
 832with them <30% of sequence identity. Interestingly, such a novel fumarylpyruvate
 833hydrolase is also encoded in one of the gentisate gene clusters of *P.*
 834*naphthalenivorans* CJ2 (Pnap2454), even though the encoded gentisate
 835dioxygenase shares significant sequence identity with that of *Ralstonia* sp. U2.

836 Enzyme diversity also extends to maleylpyruvate isomerases encoded by
 837gentisate gene clusters. In contrast to homogentisate gene clusters where an
 838isomerase-encoding gene is usually absent, most gentisate gene clusters comprise
 839such a gene. Gentisate gene clusters I, III and V typically comprise genes
 840encoding a glutathione dependent maleylpyruvate isomerase of the class zeta
 841subfamily of glutathione S-transferases with high similarity to maleylpyruvate
 842isomerases of validated function (Jones and Cooper, 1990; Marsh *et al.*, 2008).
 843The phylogeny follows that described above for fumarylpyruvate hydrolases.
 844Cognate genes are absent from type II gentisate gene clusters, while type IV
 845clusters typically comprise a gene encoding a putative maleate *cis-trans* isomerase
 846with high similarity (>60% sequence identity) to maleate isomerase from

847 *Alcaligenes faecalis* (Hatakeyama *et al.*, 1997). It could be reasoned that a
 848 glutathione independent maleate isomerase is recruited for gentisate degradation
 849 in these gene clusters.

850 The function of the majority of gentisate gene clusters is obviously the
 851 degradation of 3-hydroxybenzoate, as respective genes encoding proteins with
 852 >45% of sequence identity with 3-hydroxybenzoate 6-hydroxylase Pnap3144 from
 853 *P. naphthalenivorans* CJ2 (Park *et al.*, 2007) are also found here. However, genes
 854 of type IV and V clusters usually do not comprise 3-hydroxybenzoate 6-
 855 hydroxylase encoding genes. Whether the genes observed in *Comamonadaceae*,
 856 as encoding proteins with 33-38% identity with 3-hydroxybenzoate 6-hydroxylase
 857 Pnap3144 are functional remains to be elucidated. A second function of the
 858 gentisate gene cluster is in the degradation of salicylate after 5-hydroxylation
 859 through a multicomponent Rieske non-heme iron oxygenase. Six of the gentisate
 860 catabolic gene clusters identified here comprise genes encoding a salicylate 5-
 861 hydroxylase (Fig. 10). Additionally, genes encoding salicylate 5-hydroxylases,
 862 indicating the capability of the host to grow on salicylate via gentisate as an
 863 intermediate, are found elsewhere in the genomes of both *V. paradoxus* strains,
 864 both *Achromobacter* strains, three of four *Acidovorax* strains, two *Cupriavidus*
 865 strains, as well as in *B. petrii*, indicating such capability to be widespread in
 866 *Burkholderiales*. The turnover of salicylate through gentisate has been usually
 867 related to the degradation of polycyclic compounds in *Burkholderiales* (Lee *et al.*,
 868 2011; Tittabutr *et al.*, 2011); however, no Phn and Nag-like dioxygenases involved
 869 in the degradation of phenanthrene and naphthalene, respectively, seem to be

870 encoded in the *Burkholderiales* genomes analyzed with the exception of *P.*
 871 *naphthalenivorans* CJ2 isolated by its ability to degrade naphthalene (Jeon et al.,
 872 2006). In this bacterium, one (Pnap_3145) of three Gen encoding genes included
 873 in type III catabolic clusters (Fig. 10), has been shown to have a key role in
 874 naphthalene degradation and it has been proposed that this may prevent the
 875 accumulation of toxic intracellular levels of gentisate due its higher affinity for this
 876 substrate (Lee et al., 2011).

877 In three of the group IV catabolic gene clusters (Fig. 10), the Gen encoding
 878 gene is preceded by genes encoding flavoprotein monooxygenases with significant
 879 similarity to 2-aminobenzoyl-CoA monooxygenase/reductase from *Azoarcus*
 880 *evansii* which transforms aminobenzoyl-CoA into 2-amino-5-oxo-cyclohex-1-ene-1-
 881 carbonyl-CoA (60-62% of sequence identity) but also with similarity to salicylyl-CoA
 882 monooxygenase from *Streptomyces* WA46 (40-42% of sequence identity)
 883 (Ishiyama et al., 2004). As the gene products can be supposed to channel to
 884 gentisate, it may be reasoned that they encode salicylyl-CoA monooxygenases.
 885 Nevertheless it remains to be determined if genes encoding enzymes with
 886 similarity to indolepyruvate ferredoxin oxidoreductase (Fig. 10) actually encode
 887 enzymes that can catalyze the oxidative decarboxylation of aromatic compounds
 888 such as 2-hydroxyphenylglyoxylate. Overall, the analysis of gentisate catabolic
 889 gene clusters indicates a high diversity of gene organization and a high rate of
 890 gene rearrangement.

891

892 **Research needs**

893

894 In general terms, this analysis highlights the amazing catabolic potential of
895 *Burkholderiales* towards aromatic compounds. One evident aspect of these traits is
896 the potential utilization in environmental biotechnology for the biodegradation of
897 aromatic pollutants. This topic has been addressed in recent reviews (O'Sullivan
898 and Mahenthiralingam, 2005; Denef, 2007) and is not discussed here. However,
899 some emerging topics could attract additional interest on the biodegradation of
900 aromatic compounds by *Burkholderiales*.

901 One of these aspects is the possible role of aromatic catabolic genes for
902 pathogenicity of some members of the “*B. cepacia* complex”. In *B. cenocepacia*,
903 the requirement of a functional phenylacetic acid catabolic pathway for full
904 pathogenicity in the *Caenorhabditis elegans* host model has recently been shown
905 (Law *et al.*, 2008) as well as the induction of this route during growth in synthetic
906 cystic fibrosis medium (Hamlin *et al.*, 2009; Yudistira *et al.*, 2011). It is pertinent to
907 remember that the phenylacetate catabolic route is one of the most widespread
908 catabolic pathways in *Burkholderiales* (Fig. 2) and it has been suggested that this
909 compound or a derived catabolite may act as a signal that triggers multiple
910 physiological processes (Patrauchan *et al.* 2011), probably indicating a key role in
911 the ecological fitness of this group in multiple environments. Further research is
912 needed to understand the specific role of this pathway in host-pathogen
913 interactions.

914 An additional interesting aspect of aromatic catabolism in *Burkholderiales* is
 915 its potential role in the degradation of antibiotics. In a recent study, hundreds of soil
 916 strains were isolated for their ability to grow on antibiotics as a sole carbon source.
 917 This study included 18 different antibiotics representing eight major classes of
 918 natural and synthetic origin, the structures of various of which comprised an
 919 aromatic moiety (Dantas *et al.*, 2008). Phylogenetic profiling of the isolates
 920 revealed a diverse set of species belonging to eleven different orders, however,
 921 41% of the species isolated belonged to the *Burkholderiales*, followed by
 922 *Pseudomonadales* (24%) (Dantas *et al.*, 2008). It seems probable that the
 923 importance of *Burkholderiales* in antibiotics degradation is related to the broad
 924 catabolic potential for aromatic degradation. In this sense, it should be mentioned
 925 that during the mining of catabolic oxygenases performed for the present analysis,
 926 a huge number of genes putatively encoding aromatic oxygenases remain with
 927 unassigned substrates/functions, constituting candidates to be tested in search of
 928 the genes responsible for antibiotic degradation, among other possibilities.

929

930 **Acknowledgements**

931 Support provided by the research program FONDAP 1501 0001, program 7, the
 932 Millennium Scientific Initiative through grants P/04-007-F and P06-009-F;
 933 FONDECYT grants 1070343, 1070507, 7080148, 7090079, USM grants 13110
 934 and 130948; BMBF-CONICYT 2007-2009 and PCBT-Network (R-12). DPP, RAD
 935 and MC acknowledge CONICYT fellowships. DHP acknowledges support by the

936projects BACSIN and MAGICPAH from the European Commission. Authors
937acknowledge Verónica Morgante for valuable contributions to composition of
938Figure 2 and Max Chavarría for helpful support in statistical analysis.

939 Legends to figures

940 **Figure 1.** Overview of peripheral and ring-cleavage pathways for bacterial
 941 catabolism of aromatic compounds. The inner circle includes the structures of
 942 dearomatized and ring-cleavage products. The outer circle includes the structures
 943 of aryl-CoA and dihydroxylated ring-cleavage intermediates. Dotted lines indicate
 944 multiple steps. Gene markers listed in Table 1 are bolded.

945 **Figure 2.** Distribution of catabolic gene markers through genomes of
 946 *Burkholderiales* strains taxonomically ordered on a 16S rRNA phylogenetic tree.
 947 Maximum likelihood phylogenetic tree (left), for 16S rRNA sequences extracted
 948 from the 80 *Burkholderiales* genomes constructed using MEGA 5 software
 949 (Tamura *et al.*, 2011). The tree constructed based on the general time reversible
 950 model of nucleotide substitution including a gamma function with five categories as
 951 proposed by Akaike information criterion for selection of appropriate evolutionary
 952 model using jModeltest software (Posada and Buckley, 2004). Peripheral and
 953 central pathways markers (right), linked by ring-cleavage intermediates are
 954 grouped in shaded or non-shaded interleaved squares. Numbers denote presence
 955 of the gene marker sequence in the chromosome (1) or chromids (2, 3) when
 956 genomic structure is available. (P) Plasmid location. (D) Draft genome sequence,
 957 avoiding the precise location of gene markers. The total count of gene markers for
 958 each strain is indicated at the external right column. The total count per gene
 959 marker in all strains is indicated on the bottom row. Protein similarity searches
 960 performed with BLASTP program from the NCBI website using the default
 961 parameters (Altschul *et al.*, 1997).

Figure 3. Relative content of catabolic gene markers in *Burkholderiales* strains according to replicon composition and genome size. (A) Percent of aromatic oxygenase-encoding genes per total genes encoded in different replicon types of *Burkholderiales* genera harboring multipartite genomes. (B) Relation between genome size and percentage of aromatic oxygenase-encoding genes and total enzyme-encoding genes in *Burkholderiales* strains harboring at least one catabolic gene marker. The numbers of enzyme-encoding genes obtained from Integrated Microbial Genomes (IMG) system.

Figure 4. Relative content of peripheral and central catabolic markers per total genes encoded in genomes of *Burkholderiales* strains grouped by (A) isolation sources including human and animal hosts [H/Z], soil and rhizosphere [S/R] and wastewater or sludge and sediment [W/D]; and (B) taxonomic families harboring sizeable biodegradative properties.

Figure 5. Maximum likelihood phylogenetic tree for alpha subunit of protocatechuate 3,4-dioxygenase (Pca34) and gene clusters in which the respective dioxygenase-encoding genes are found. After initial identification of homologous sequences considering similarity by the chance expectation values and scores of BLAST probing, the whole sets of homologous proteins were subject to phylogenetic analysis including proteins with well-known function to refine genuine orthologous identification. The deduced protein sequences were aligned with MUSCLE using default values (Edgar, 2004), and alignments were manually checked and further edited to remove misaligned regions. The alignments of each data set of homologous proteins were used to evaluate models of sequence

evolution using Prottest v.2.4 (Abascal *et al.*, 2005). The Akaike information criterion was used to select the most appropriate evolutionary model for each data set. Maximum likelihood phylogenetic trees were constructed based on the selected models using PhyML 3.0 (Guindon *et al.*, 2010) and MEGA 5 (Tamura *et al.*, 2011) software. Confidence in phylogenetic inference was assessed using non-parametric bootstrap resampling. Branches tagged with dots indicate sequences with confirmed functionality not belonging to *Burkholderiales* strains that have been incorporated for additional phylogenetic comparisons. The genes encoding both subunits of Pca34 are aligned to one another. The functions of the different types of genes included in the clusters are indicated on the right. Dotted arrows indicate the presence of additional genes related to peripheral pathways channelling to protocatechuate. The sizes of genes and intergenic regions are to scale.

Figure 6. Maximum likelihood phylogenetic tree for gallate dioxygenase (Gal) and protocatechuate 4,5-dioxygenase (Pca45) and gene clusters in which the respective dioxygenase-encoding genes are found. Branches tagged with dots indicate sequences with confirmed functionality not belonging to *Burkholderiales*, or belonging to *Burkholderiales* strains not included in this study, that have been incorporated for additional phylogenetic comparisons. The genes encoding Gal and Pca45 are aligned to one another. The functions of the different types of genes included in the clusters are indicated at the bottom. Other details as indicated in legend of Fig. 5.

Figure 7. Maximum likelihood phylogenetic tree for catechol 1,2-dioxygenase (Cat12) and gene clusters in which the respective dioxygenase-encoding genes

1008are found. Branches tagged with dots indicate sequences with confirmed
 1009functionality not belonging to *Burkholderiales*, or belonging to *Burkholderiales*
 1010strains not included in this study, that have been incorporated for additional
 1011phylogenetic comparisons. The genes encoding Cat12 are aligned to one another.
 1012Dotted arrows indicate the presence of additional genes related to peripheral
 1013pathways channelling to catechol. A bidirectional arrow indicates truncated
 1014sequences. Other details as indicated in legend of Fig. 5.

1015**Figure 8.** Maximum likelihood phylogenetic tree for catechol 2,3-dioxygenase
 1016(Cat23) and gene clusters in which the respective dioxygenase-encoding genes
 1017are found. Branches tagged with dots indicate sequences with confirmed
 1018functionality not belonging to *Burkholderiales*, or belonging to *Burkholderiales*
 1019strains not included in this study, that have been incorporated for additional
 1020phylogenetic comparisons. The genes encoding Cat23 are aligned to one another.
 1021The functions of the different types of genes included in the clusters are indicated
 1022at the bottom. A bidirectional arrow indicates truncated sequences. Other details as
 1023indicated in legend of Fig. 5.

1024**Figure 9.** Maximum likelihood phylogenetic tree for homogentisate dioxygenase
 1025(Hge) and gene clusters in which the respective dioxygenase-encoding genes are
 1026found. Branches tagged with dots indicate sequences with confirmed functionality
 1027not belonging to *Burkholderiales* that have been incorporated for additional
 1028phylogenetic comparisons. The genes encoding Hge are aligned to one another. A
 1029bidirectional arrow indicates truncated sequences. Other details as indicated in
 1030legend of Fig. 5.

1031 **Figure 10.** Maximum likelihood phylogenetic tree for gentisate dioxygenase (Gen)
1032 and gene clusters in which the respective dioxygenase-encoding genes are found.
1033 Branches tagged with dots indicate sequences with confirmed functionality not
1034 belonging to *Burkholderiales*, or belonging to *Burkholderiales* strains not included
1035 in this study, that have been incorporated for additional phylogenetic comparisons.
1036 The genes encoding Gen are aligned to one another. Dotted arrows indicate the
1037 presence of additional genes related to peripheral pathways channelling to
1038 gentisate. Other details as indicated in legend of Fig. 5.

1039Supplementary Material (Pérez-Pantoja *et al.*)

1040**Table S1:** *Burkholderiales* strains information and genome statistics.

1041

1042**Figure S1:** Chromosomal location of catabolic gene markers belonging to
1043peripheral and central routes linked by catechol, protocatechuate, gentisate and
1044homogentisate as ring-cleavage intermediates, in the genomes of selected
1045*Burkholderiales* strains. Replicons are represented linearly with endpoints
1046corresponding to those in the Entrez Genome database. The genomic size and
1047mean G+C content of total genes including the standard deviation are indicated at
1048the right and left endpoints of each replicon, respectively. The G+C content of each
1049gene marker is indicated next to the symbol representing it. Catabolic gene
1050markers included in genomic islands as predicted by IslandViewer (Langille and
1051Brinkman 2009) are indicated with an asterisk next to the symbol representing it.

Table 1. Catabolic enzyme markers

Enzyme	Family	Enzyme function	Abbreviation
Protocatechuate 3,4-dioxygenase	Intradiol dioxygenase	Intradiol cleavage	Pca34
Catechol 1,2-dioxygenase	Intradiol dioxygenase	Intradiol cleavage	Cat12
Hydroxyquinol 1,2-dioxygenase	Intradiol dioxygenase	Intradiol cleavage	Hxq
Chlorocatechol 1,2-dioxygenase	Intradiol dioxygenase	Intradiol cleavage	Cca
Catechol 2,3-dioxygenase	Type I extradiol dioxygenase	Extradiol cleavage	Cat23
2,3-dihydroxy- <i>p</i> -cumate-3,4-dioxygenase	Type I extradiol dioxygenase	Extradiol cleavage	Dhc
Chlorohydroquinone 1,2-dioxygenase	Type I extradiol dioxygenase	Extradiol cleavage	Chq
2,4,5-Trihydroxytoluene dioxygenase	Type I extradiol dioxygenase	Extradiol cleavage	Tht
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
2,3-Dihydroxyphenylpropionate 1,2-dioxygenase	LigB type dioxygenase	Extradiol cleavage	Dhp
2-Aminophenol 1,6-dioxygenase	LigB type dioxygenase	Extradiol cleavage	Amn
Hydroquinone 1,2-dioxygenase	Type IV extradiol dioxygenase	Extradiol cleavage	Hqo
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
Benzoyl-CoA oxygenase/reductase	Soluble diiron oxygenase	Dearomatization	Box
Phenylacetyl-CoA oxygenase/reductase	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
Salicylate 1-hydroxylase	Class A flavoprotein monooxygenase	Forming catechol	Sal1H
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
(di)Chlorophenol hydroxylase	Class A flavoprotein monooxygenase	Forming (di)chlorocatechol	CphH
4-hydroxyphenylacetate 3-hydroxylase	Class D flavoprotein monooxygenase	Forming homoprotocatechuate	Pha3H
Trichlorophenol 4-hydroxylase	Class D flavoprotein monooxygenase	Forming chlorohydroxyquinol	Tcp4H

Resorcinol 4-hydroxylase	Class D flavoprotein monooxygenase	Forming hydroxyquinol	Res4H
Vanillate 3-O demethylase	Rieske non-heme iron oxygenase	Channelling to protococatechuate	VanOD
Isovanillate 4-O demethylase	Rieske non-heme iron oxygenase	Channelling to protococatechuate	IvaOD
Terephthalate 1,2-dioxygenase	Rieske non-heme iron oxygenase	Channelling to protococatechuate	TphDO
Phthalate 3,4-dioxygenase	Rieske non-heme iron oxygenase	Channelling to protococatechuate	PhtDO
Isophthalate 3,4-dioxygenase	Rieske non-heme iron oxygenase	Channelling to protococatechuate	IphDO
Anthranilate 1,2-dioxygenase	Rieske non-heme iron oxygenase	Channelling to catechol	AntDO
Benzoate 1,2-dioxygenase	Rieske non-heme iron oxygenase	Channelling to catechol	BenDO
2-Halobenzoate 1,2-dioxygenase	Rieske non-heme iron oxygenase	Channelling to catechol	OhbDO
Biphenyl 2,3-dioxygenase	Rieske non-heme iron oxygenase	Channelling to catechol	BphDO
Indole 3-acetate dioxygenase	Rieske non-heme iron oxygenase	Channelling to catechol	IaaDO
Chlorobenzene dioxygenase	Rieske non-heme iron oxygenase	Channelling to chlorocatechol	CbzDO
<i>p</i> -Cumate 2,3-dioxygenase	Rieske non-heme iron oxygenase	Channelling to 2,3-dihydroxy- <i>p</i> -cumate	CmtDO
Salicylate 5-hydroxylase	Rieske non-heme iron oxygenase	Channelling to gentisate	Sal5H
Toluene/benzene monooxygenase	Soluble diiron oxygenase	Forming phenol	Tmo
Phenol monooxygenase	Soluble diiron oxygenase	Forming catechol	Pmo
4-Hydroxyphenylpyruvate dioxygenase	α -keto acid dependent dioxygenase	Forming homogentisate	HppDO
2,4'-Dihydroxyacetophenone dioxygenase	Cupin superfamily dioxygenase	Forming 4-hydroxybenzoate	DapDO
<i>p</i> -Cresol methylhydroxylase	Flavocytochrome C	Forming 4-hydroxybenzoate	PcmH

1053

1054**References**

1055

1056Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: Selection of best-fit
1057models of protein evolution. *Bioinformatics* **21**:2104-2105.

1058Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and
1059Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein
1060database search programs. *Nucleic Acids Res* **25**:3389-3402.

1061Amadou, C., Pascal, G., Mangenot, S., Glew, M., Bontemps, C., Capela, D., *et al.*
1062(2008). Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and
1063comparative genomics of rhizobia. *Genome Res* **18**:1472-1483.

1064Arias-Barrau, E., Olivera, E.R., Luengo, J.M., Fernández, C., Galán, B., García,
1065J.L., *et al.* (2004). The homogentisate pathway: a central catabolic pathway
1066involved in the degradation of L-phenylalanine, L-tyrosine, and 3-
1067hydroxyphenylacetate in *Pseudomonas putida*. *J Bacteriol* **186**:5062-5077.

1068Arias-Barrau, E., Sandoval, A., Naharro, G., Olivera, E.R., and Luengo, J.M.
1069(2005). A two-component hydroxylase involved in the assimilation of 3-
1070hydroxyphenyl acetate in *Pseudomonas putida*. *J Biol Chem* **280**: 26435-26447.

1071Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S., and Vivanco, J.M. (2006). The role of
1072root exudates in rhizosphere interactions with plants and other organisms. [*Annu.*](#)
1073[*Rev Plant Biol* **57**:233-266.](#)

- 1074Balandreau, J., and Mavingui, P. (2006). Beneficial interactions of *Burkholderia*
 1075spp. with plants. In *Burkholderia: Molecular Microbiology and Genomics*. Coeyne,
 1076T., and Vandamme, P. (eds). Wymondham, Norfolk: Horizon Bioscience, pp. 129-
 1077151.
- 1078Bayly, R.C., Chapman, P.J., Dagley, S., and Di Berardino, D. (1980). Purification
 1079and some properties of maleylpyruvate hydrolase and fumarylpyruvate hydrolase
 1080from *Pseudomonas alcaligenes*. *J Bacteriol* **143**:70-77.
- 1081Cafaro, V., Izzo, V., Scognamiglio, R., Notomista, E., Capasso, P., Casbarra, A., *et*
 1082*al.* (2004). [Phenol hydroxylase and toluene/o-xylene monooxygenase from](#)
 1083[Pseudomonas stutzeri OX1: interplay between two enzymes.](#) *Appl Environ*
 1084*Microbiol* **70**:2211-2219.
- 1085Catelani, D., Fiecchi, A., and Galli, E. (1971). Dextro-gamma-carboxymethyl-
 1086gamma-methyl-delta-alpha-butenolide. A 1,2-ring-fission product of 4-
 1087methylcatechol by *Pseudomonas desmolyticum*. *Biochem J* **121**:89-92.
- 1088Chain, P.S., Deneff, V.J., Konstantinidis, K.T., Vergez, L.M., Agullo, L., Reyes, V.L.,
 1089*et al.* (2006). *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp
 1090genome shaped for versatility. *Proc Natl Acad Sci USA* **103**:15280-15287.
- 1091Chang, H.K., and Zylstra, G.J. (1998). Novel organization of the genes for
 1092phthalate degradation from *Burkholderia cepacia* DBO1. *J Bacteriol* **180**:6529-
 10936537.

1094Chang, H.K., and Zylstra, G.J. (2008). Examination and expansion of the substrate
 1095range of *m*-hydroxybenzoate hydroxylase. *Biochem Biophys Res Commun*
 1096**371**:149-153.

1097Chiarini, L., Bevivino, A., Dalmastri, C., Tabacchioni, S., and Visca, P. (2006).
 1098*Burkholderia cepacia* complex species: health hazards and biotechnological
 1099potential. *Trends Microbiol* **14**:277-286.

1100Crawford, R.L. (1976). Degradation of homogentisate by strains of *Bacillus* and
 1101*Moraxella*. *Can J Microbiol* **22**:276-280.

1102Crawford, R.L., and Frick, T.D. (1977). Rapid spectrophotometric differentiation
 1103between glutathione-dependent and glutathione-independent gentisate and
 1104homogentisate pathways. *Appl Environ Microbiol* **34**:170-174.

1105Cupples, A.M., and Sims, G.K. (2007). Identification of *in situ* 2,4-
 1106dichlorophenoxyacetic acid-degrading soil microorganisms using DNA-stable
 1107isotope probing. *Soil Biol Biochem* **39**:232–238.

1108Dantas, G., Sommer, M.O., Oluwasegun, R.D., and Church, G.M. (2008). Bacteria
 1109subsisting on antibiotics. *Science* **320**:100-103.

1110Denef, V.J. (2007). Biodegradation of organic anthropogenic pollutants by
 1111*Burkholderia* species. In *Burkholderia: Molecular Microbiology and Genomics*.
 1112Coenye, T., and Vandamme, P. (eds). Wymondham, Norfolk: Horizon Bioscience,
 1113pp. 177-201.

- 1114Dunwell, J.M., Khuri, S., and Gane, P.J. (2000). Microbial relatives of the seed
 1115storage proteins of higher plants: conservation of structure and diversification of
 1116function during evolution of the cupin superfamily. *Microbiol Mol Biol Rev* **64**:153-
 1117179.
- 1118Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced
 1119time and space complexity. *BMC Bioinformatics* **5**:113.
- 1120Elsemore, D.A., and Ornston, L.N. (1994). The *pca-pob* supraoperonic cluster of
 1121*Acinetobacter calcoaceticus* contains *quiA*, the structural gene for quinate-
 1122shikimate dehydrogenase. *J Bacteriol* **176**:7659-7666.
- 1123Eltis, L.D., and Bolin, J.T. (1996). Evolutionary relationships among extradiol
 1124dioxygenases. *J Bacteriol* **178**:5930-5937.
- 1125Erb, R.W., Timmis, K.N., and Pieper, D.H. (1998). Characterization of a gene
 1126cluster from *Ralstonia eutropha* JMP134 encoding metabolism of 4-
 1127methylmuconolactone. *Gene* **206**:53-62.
- 1128Fitzpatrick, P.F. (2003). [Mechanism of aromatic amino acid hydroxylation.](#)
 1129*Biochemistry* **42**:14083-14091
- 1130Fouts, D.E., Tyler, H.L., DeBoy, R.T., Daugherty, S., Ren, Q., Badger, J.H., *et al.*
 1131(2008). Complete genome sequence of the N₂-fixing broad host range endophyte
 1132*Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet*
 1133**4**:e1000141.

- 1134Fukuhara, Y., Inakazu, K., Kodama, N., Kamimura, N., Kasai, D., Katayama, Y., *et*
 1135*al.* (2010). Characterization of the isophthalate degradation genes of *Comamonas*
 1136*sp.* strain E6. *Appl Environ Microbiol* **76**:519-527.
- 1137Garrrity, G.M., Bell, J.A., and Lilburn, T. (2005). Order I. *Burkholderiales* ord. nov. In
 1138Bergey's Manual of Systematic Bacteriology. Brenner, D.J., Krieg, N.R., Staley,
 1139J.T., and Garrrity, G.M. (eds). New York: Springer, p. 575.
- 1140Gibson, D.T., and Parales, R.E. (2000). Aromatic hydrocarbon dioxygenases in
 1141environmental biotechnology. *Curr Opin Biotechnol* **11**:236-243.
- 1142Gross, R., Guzmán, C.A., Sebahia, M., dos Santos, V.A., Pieper, D.H., Koebnik,
 1143R., *et al.* (2008). The missing link: *Bordetella petrii* is endowed with both the
 1144metabolic versatility of environmental bacteria and virulence traits of pathogenic
 1145*Bordetellae*. *BMC Genomics* **9**:449.
- 1146Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O.
 1147(2010). New algorithms and methods to estimate maximum-likelihood phylogenies:
 1148Assessing the performance of PhyML 3.0. *Syst Biol* **59**:307-321.
- 1149Hamlin, J.N., Bloodworth, R.A., and Cardona, S.T. (2009). Regulation of
 1150phenylacetic acid degradation genes of *Burkholderia cenocepacia* K56-2. *BMC*
 1151*Microbiol* **9**:222.
- 1152Harayama, S., Mermod, N., Rekik, M., Lehrbach, P.R., and Timmis, K.N. (1987).
 1153Roles of the divergent branches of the meta-cleavage pathway in the degradation
 1154of benzoate and substituted benzoates. *J Bacteriol* **169**:558-564.

1155Harayama, S., and Rekik, M. (1990). The *meta* cleavage operon of TOL
1156degradative plasmid pWW0 comprises 13 genes. *Mol Gen Genet* **221**:113-120.

1157Harrison, P.W., Lower, R.P., Kim, N.K., and Young, J.P. (2010). Introducing the
1158bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol* **18**:141-
1159148.

1160Harwood, C.S., Nichols, N.N., Kim, M.K., Ditty, J.L., and Parales, R.E. (1994).
1161Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement
1162in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. *J Bacteriol*
1163**176**:6479-6488.

1164Harwood, C.S., and Parales, R.E. (1996). The beta-ketoadipate pathway and the
1165biology of self-identity. *Annu Rev Microbiol* **50**:553–590.

1166Hatakeyama, K., Asai, Y., Uchida, Y., Kobayashi, M., Terasawa, M., and Yukawa,
1167H. (1997). Gene cloning and characterization of maleate *cis-trans* isomerase from
1168*Alcaligenes faecalis*. *Biochem Biophys Res Commun* **239**:74-79.

1169Hugo, N., Armengaud, J., Gaillard, J., Timmis, K.N., and Jouanneau, Y. (1998). A
1170novel [2Fe-2S] ferredoxin from *Pseudomonas putida* mt-2 promotes the reductive
1171reactivation of catechol 2,3-dioxygenase. *J Biol Chem* **273**:9622-9629.

1172Ishiyama, D., Vujaklija, D., and Davies, J. (2004). Novel pathway of salicylate
1173degradation by *Streptomyces* sp. strain WA46. *Appl Environ Microbiol* **70**:1297-
11741306.

1175 Jeon, C.O., Park, W., Padmanabhan, P., DeRito, C., Snape, J.R., and Madsen,
 1176 E.L. (2003). Discovery of a bacterium, with distinctive dioxygenase, that is
 1177 responsible for *in situ* biodegradation in contaminated sediment. *Proc Natl Acad*
 1178 *Sci USA* **100**:13591-13596.

1179 Jeon, C.O., Park, M., Ro, H.S., Park, W., and Madsen, E.L. (2006). The
 1180 naphthalene catabolic (*nag*) genes of *Polaromonas naphthalenivorans* CJ2:
 1181 evolutionary implications for two gene clusters and novel regulatory control. *Appl*
 1182 *Environ Microbiol* **72**:1086-1095.

1183 Jeong, J.J., Kim, J.H., Kim, C.K., Hwang, I., and Lee, K. (2003). 3- and 4-
 1184 alkylphenol degradation pathway in *Pseudomonas* sp. strain KL28: genetic
 1185 organization of the *lap* gene cluster and substrate specificities of phenol
 1186 hydroxylase and catechol 2,3-dioxygenase. *Microbiology* **149**:3265-3277.

1187 Jiménez, J.I., Miñambres, B., Garcia, J.L., and Díaz, E. (2002). Genomic analysis
 1188 of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ*
 1189 *Microbiol* **4**:824-841.

1190 Jones, D.C., and Cooper, R.A. (1990). Catabolism of 3-hydroxybenzoate by the
 1191 gentisate pathway in *Klebsiella pneumoniae* M5a1. *Arch Microbiol* **154**:489-495.

1192 Kamimura, N., Aoyama, T., Yoshida, R., Takahashi, K., Kasai, D., Abe, T., *et al.*
 1193 (2010). Characterization of the protocatechuate 4,5-cleavage pathway operon in
 1194 *Comamonas* sp. strain E6 and discovery of a novel pathway gene. *Appl Environ*
 1195 *Microbiol* **76**:8093-8101.

- 1196 [Kasai, D.](#), [Masai, E.](#), [Miyauchi, K.](#), [Katayama, Y.](#), and [Fukuda, M.](#) (2005).
 1197 Characterization of the gallate dioxygenase gene: three distinct ring cleavage
 1198 dioxygenases are involved in syringate degradation by *Sphingomonas*
 1199 *paucimobilis* SYK-6. [J Bacteriol](#) **187**:5067-5074.
- 1200 Kaschabek, S.R., Kasberg, T., Muller, D., Mars, A.E., Janssen, D.B., and Reineke,
 1201 W. (1998). Degradation of chloroaromatics: purification and characterization of a
 1202 novel type of chlorocatechol 2,3-dioxygenase of *Pseudomonas putida* GJ31. *J*
 1203 *Bacteriol* **180**:296-302.
- 1204 Kataoka, R., and Futai, K. (2009). A new mycorrhizal helper bacterium, *Ralstonia*
 1205 species, in the ectomycorrhizal symbiosis between *Pinus thunbergii* and *Suillus*
 1206 *granulatus*. *Biol Fertil Soils* **45**:315–320.
- 1207 Knackmuss, H.-J., Hellwig, M., Lackner, H., and Otting, W. (1976). Cometabolism
 1208 of 3-methylbenzoate and methylcatechols by a 3-chlorobenzoate utilizing
 1209 *Pseudomonas*: accumulation of (+)-2,5-dihydro-4-methyl- and (+)-2,5-dihydro-2-
 1210 methyl-5-oxo-furan-2-acetic acid. *Eur J Appl Microbiol* **2**:267-276.
- 1211 Kukor, J.J., and Olsen, R.H. (1996). Catechol 2,3-dioxygenases functional in
 1212 oxygen-limited (hypoxic) environments. *Appl Environ Microbiol* **62**:1728-1740.
- 1213 Lackner, G., Moebius, N., Partida-Martínez, L., and Hertweck, C. (2011). Complete
 1214 genome sequence of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus*
 1215 *microsporus*. *J Bacteriol* **193**:783-784.

1216Langille, M.G., and Brinkman, F.S. (2009). IslandViewer: an integrated interface for
 1217computational identification and visualization of genomic islands. *Bioinformatics*
 1218**25**:664-665.

1219Law, R.J., Hamlin, J.N., Sivro, A., McCorrister, S.J., Cardama, G.A., and Cardona,
 1220S.T. (2008). A functional phenylacetic acid catabolic pathway is required for full
 1221pathogenicity of *Burkholderia cenocepacia* in the *Caenorhabditis elegans* host
 1222model. *J Bacteriol* **190**:7209-7218.

1223Leahy, J.G., Batchelor, P.J., and Morcomb, S.M. (2003). Evolution of the soluble
 1224diiron monooxygenases. *FEMS Microbiol Rev* **27**:449-479.

1225Lee, H.J., Kim, J.M., Lee, S.H., Park, M., Madsen, E.L., and Jeon, C.O. (2011).
 1226Characterization of gentisate-1,2-Dioxygenase in the third naphthalene catabolic
 1227gene cluster of *Polaromonas naphthalenivorans* CJ2: Role in naphthalene
 1228degradation. *Microbiology* doi:10.1099/mic.0.049387-0

1229Leveau, J.H., and Gerards, S. (2008). Discovery of a bacterial gene cluster for
 1230catabolism of the plant hormone indole 3-acetic acid. *FEMS Microbiol Ecol* **65**:238-
 1231250.

1232Liou, J.S.C., DeRito, C.M., and Madsen, E.L. (2008). Field-based and laboratory
 1233stable isotope probing surveys of the identities of both aerobic and anaerobic
 1234benzene-metabolizing microorganisms in freshwater sediment. *Environ Microbiol*
 1235**10**:1964–1977.

1236Liu, Z., Yang, H., Huang, Z., Zhou, P., and Liu, S.J. (2002). Degradation of aniline
 1237by newly isolated, extremely aniline-tolerant *Delftia* sp. AN3. *Appl Microbiol*
 1238*Biotechnol* **58**:679-682.

1239Lykidis, A., Pérez-Pantoja, D., Ledger, T., Mavromatis, K., Anderson, I.J., Ivanova,
 1240N.N., *et al.* (2010). The complete multipartite genome sequence of *Cupriavidus*
 1241*necator* JMP134, a versatile pollutant degrader. *PLoS One* **5**:e9729.

1242Madsen, E.L. (2006). The use of stable isotope probing techniques in bioreactor
 1243and field studies on bioremediation. *Curr Opin Biotechnol* **17**:92-97.

1244Mahmood, S., Paton, G.I., and Prosser, J.I. (2005). Cultivation-independent *in situ*
 1245molecular analysis of bacteria involved in degradation of pentachlorophenol in soil.
 1246*Environ Microbiol* **7**:1349–1360.

1247Mampel, J., Providenti, M.A., and Cook, A.M. (2005). Protocatechuate 4,5-
 1248dioxygenase from *Comamonas testosteroni* T-2: biochemical and molecular
 1249properties of a new subgroup within class III of extradiol dioxygenases. *Arch*
 1250*Microbiol* **183**:130-139.

1251Manefield, M., Griffiths, R.I., Leigh, M.B., Fisher, R., and Whiteley, A.S. (2005).
 1252Functional and compositional comparison of two activated sludge communities
 1253remediating coking effluent. *Environ Microbiol* **7**:715–722.

1254Marsh, M., Shoemark, D.K., Jacob, A., Robinson, C., Cahill, B., Zhou, N.Y., *et al.*
 1255(2008). Structure of bacterial glutathione-S-transferase maleyl pyruvate isomerase
 1256and implications for mechanism of isomerisation. *J Mol Biol* **384**:165-177.

1257Mattes, T.E., Alexander, A.K., Richardson, P.M., Munk, A.C., Han, C.S., Stothard,
1258P., and Coleman, N.V. (2008). The genome of *Polaromonas* sp. strain JS666:
1259insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium,
1260and features of relevance to biotechnology. *Appl Environ Microbiol* **74**:6405-6416.

1261Méndez, V., Agulló, L., González, M., and Seeger, M. (2011). The homogentisate
1262and homoprotocatechuate central pathways are involved in 3- and 4-
1263hydroxyphenylacetate degradation by *Burkholderia xenovorans* LB400. *PLoS ONE*
1264**6**:e17583.

1265Merimaa, M., Heinaru, E., Liivak, M., Vedler, E., and Heinaru, A. (2006). Grouping
1266of phenol hydroxylase and catechol 2,3-dioxygenase genes among phenol- and *p*-
1267cresol-degrading *Pseudomonas* species and biotypes. *Arch Microbiol* **186**:287-296.

1268Moran, G.R. (2005). 4-Hydroxyphenylpyruvate dioxygenase. *Arch Biochem*
1269*Biophys* **433**:117-128.

1270Nakatsu, C.H., and Wyndham, R.C. (1993). Cloning and expression of the
1271transposable chlorobenzoate-3,4-dioxygenase genes of *Alcaligenes* sp. strain
1272BR60. *Appl Environ Microbiol* **59**:3625-3633.

1273Newman, L.M., and Wackett, L.P. (1995). Purification and characterization of
1274toluene 2-monooxygenase from *Burkholderia cepacia* G4. *Biochemistry* **34**:14066-
127514076.

- 1276Noda, Y., Nishikawa, S., Shiozuka, K., Kadokura, H., Nakajima, H., Yoda, K., *et al.*
 1277(1990). Molecular cloning of the protocatechuate 4,5-dioxygenase genes of
 1278*Pseudomonas paucimobilis*. *J Bacteriol* **172**:2704-2709.
- 1279Nogales, J., Canales, A., Jiménez-Barbero, J., García, J.L., and Díaz, E. (2005).
 1280[Molecular characterization of the gallate dioxygenase from *Pseudomonas putida*](#)
 1281[KT2440. The prototype of a new subgroup of extradiol dioxygenases.](#) *J Biol Chem*
 1282**280**:35382-35390.
- 1283Nogales, J., Canales, A., Jiménez-Barbero, J., Serra, B., Pingarrón, J.M., García,
 1284J.L., and Díaz, E. (2011). Unravelling the gallic acid degradation pathway in
 1285bacteria: the *gal* cluster from *Pseudomonas putida*. *Mol Microbiol* **79**:359-374.
- 1286Nordlund, I., and Shingler, V. (1990). Nucleotide sequences of the *meta*-cleavage
 1287pathway enzymes 2-hydroxymuconic semialdehyde dehydrogenase and 2-
 1288hydroxymuconic semialdehyde hydrolase from *Pseudomonas* CF600. *Biochim*
 1289*Biophys Acta* **1049**:227-330.
- 1290Nordlund, I., Powlowski, J., Hagström, A., and Shingler, V. (1993). Conservation of
 1291regulatory and structural genes for a multi-component phenol hydroxylase within
 1292phenol-catabolizing bacteria that utilize a *meta*-cleavage pathway. *J Gen Microbiol*
 1293**139**:2695-2703.
- 1294O'Sullivan, L.A., and Mahenthiralingam, E. (2005). Biotechnological potential within
 1295the genus *Burkholderia*. *Lett Appl Microbiol* **41**:8-11.

1296Park, M., Jeon, Y., Jang, H.H., Ro, H.S., Park, W., Madsen, E.L., and Jeon, C.O.
 1297(2007). Molecular and biochemical characterization of 3-hydroxybenzoate 6-
 1298hydroxylase from *Polaromonas naphthalenivorans* CJ2. *Appl Environ Microbiol*
 1299**73**:5146-5152.

1300Parsek, M.R., Shinabarger, D.L., Rothmel, R.K., and Chakrabarty, A.M. (1992).
 1301[Roles of CatR and cis,cis-muconate in activation of the catBC operon, which is](#)
 1302[involved in benzoate degradation in *Pseudomonas putida*. *J Bacteriol* **174**:7798-](#)
 13037806.

1304Patrauchan, M.A., Parnell, J.J., McLeod, M.P., Florizone, C., Tiedje, J.M., and
 1305Eltis, L.D. (2011). Genomic analysis of the phenylacetyl-CoA pathway in
 1306*Burkholderia xenovorans* LB400. *Arch Microbiol* doi:10.1007/s00203-011-0705-x.

1307Pérez-Pantoja, D., De la Iglesia, R., Pieper, D.H., and González, B. (2008).
 1308Metabolic reconstruction of aromatic compounds degradation from the genome of
 1309the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS*
 1310*Microbiol Rev* **32**:736-794.

1311Pérez-Pantoja, D., González, B., and Pieper, D.H. (2010a). Aerobic degradation of
 1312aromatic hydrocarbons. In *Handbook of Hydrocarbon and Lipid Microbiology*.
 1313Timmis, K.N. (ed). Berlin: Springer-Verlag Berlin Heidelberg, pp. 799-837.

1314Pérez-Pantoja, D., Donoso, R., Junca, H., González, B., and Pieper, D.H. (2010b).
 1315Phylogenomics of aerobic bacterial degradation of aromatics. In *Handbook of*
 1316*Hydrocarbon and Lipid Microbiology*. Timmis, K.N. (ed). Berlin: Springer-Verlag
 1317Berlin Heidelberg, pp. 1355-1397.

1318Pieper, D.H., Engesser, K.H., and Knackmuss, H.-J. (1989). Regulation of
 1319catabolic pathways of phenoxyacetic acids and phenols in *Alcaligenes eutrophus*
 1320JMP134. *Arch Microbiol* **151**:365-371.

1321Pieper, D.H., González, B., Cámara, B., Pérez-Pantoja, D., and Reineke, W.
 1322(2010). Aerobic degradation of chloroaromatics. In Handbook of Hydrocarbon and
 1323Lipid Microbiology. Timmis, K.N. (ed). Berlin: Springer-Verlag Berlin Heidelberg,
 1324pp. 839-864.

1325Poh, C.L., and Bayly, R.C. (1980). Evidence for isofunctional enzymes used in *m*-
 1326cresol and 2,5-xylene degradation via the gentisate pathway in *Pseudomonas*
 1327*alcaligenes*. *J Bacteriol* **143**:59-69.

1328Posada, D., and Buckley, T.R. (2004). Model selection and model averaging in
 1329phylogenetics: advantages of the AIC and Bayesian approaches over likelihood
 1330ratio tests. *Syst Biol* **53**:793-808.

1331Powlowski, J., and Shingler, V. (1994). Genetics and biochemistry of phenol
 1332degradation by *Pseudomonas* sp. CF600. *Biodegradation* **5**:219-236.

1333Providenti, M.A., Mampel, J., MacSween, S., Cook, A.M., and Wyndham, R.C.
 1334(2001). *Comamonas testosteroni* BR6020 possesses a single genetic locus for
 1335extradiol cleavage of protocatechuate. *Microbiology* **147**:2157-2167.

1336Providenti, M.A., O'Brien, J.M., Ruff, J., Cook, A.M., and Lambert, I.B. (2006a).
 1337Metabolism of isovanillate, vanillate, and veratrate by *Comamonas testosteroni*
 1338strain BR6020. *J Bacteriol* **188**:3862-3869.

1339Providenti, M.A., Shaye, R.E., Lynes, K.D., McKenna, N.T., O'Brien, J.M., Rosolen,
 1340S., *et al.* (2006b). The locus coding for the 3-nitrobenzoate dioxygenase of
 1341*Comamonas* sp. strain JS46 is flanked by IS1071 elements and is subject to
 1342deletion and inversion events. *Appl Environ Microbiol* **72**:2651-2660.

1343Pumphrey, G.M., and Madsen, E.L. 2008. Field-based stable isotope probing
 1344reveals the identities of benzoic acid-metabolizing microorganisms and their *in situ*
 1345growth in agricultural soil. *Appl Environ Microbiol* **74**:4111-4118.

1346Sasoh, M., Masai, E., Ishibashi, S., Hara, H., Kamimura, N., Miyauchi, K., and
 1347Fukuda, M. (2006). Characterization of the terephthalate degradation genes of
 1348*Comamonas* sp. strain E6. *Appl Environ Microbiol* **72**:1825-1832.

1349Sato, Y., Nishihara, H., Yoshida, M., Watanabe, M., Rondal, J.D., Concepcion,
 1350R.N., and Ohta, H. (2006). *Cupriavidus pinatubonensis* sp. nov. and *Cupriavidus*
 1351*laharis* sp. nov., novel hydrogen-oxidizing, facultatively chemolithotrophic bacteria
 1352isolated from volcanic mudflow deposits from Mt. Pinatubo in the Philippines. *Int J*
 1353*Syst Evol Microbiol* **56**:973-978.

1354Schweigert, N., Zehnder, A.J., and Eggen, R.I. (2001). Chemical properties of
 1355catechols and their molecular modes of toxic action in cells, from microorganisms
 1356to mammals. *Environ Microbiol* **3**:81-91.

1357Singleton, D.R., Powell, S.N., Sangaiah, R., Gold, A., Ball, L.M., and Aitken, M.D.
 1358(2005). Stable-isotope probing of bacteria capable of degrading salicylate,
 1359naphthalene, or phenanthrene in a bioreactor treating contaminated soil. *Appl*
 1360*Environ Microbiol* **71**:1202–1209.

1361 Suemori, A., Nakajima, K., Kurane, R., and Nakamura, Y. (1995). *o*-, *m*- and *p*-
 1362 hydroxybenzoate degradative pathways in *Rhodococcus erythropolis*. *FEMS*
 1363 *Microbiol Lett* **125**:31-35.

1364 Sun, W., Xie, S., Luo, C., and Cupples, A.M. (2010). Direct link between toluene
 1365 degradation in contaminated-site microcosms and a *Polaromonas* strain. *Appl*
 1366 *Environ Microbiol* **76**:956-959.

1367 Suzuki, K., Ichimura, A., Ogawa, N., Hasebe, A., and Miyashita, K. (2002).
 1368 Differential expression of two catechol 1,2-dioxygenases in *Burkholderia* sp. strain
 1369 TH2. *J Bacteriol* **184**:5714-5722.

1370 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.
 1371 (2011). MEGA5: Molecular evolutionary genetics analysis using maximum
 1372 likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*
 1373 doi:10.1093/molbev/msr121.

1374 Tillmann, S., Strompl, C., Timmis, K.N., and Abraham, W.R. (2005). Stable isotope
 1375 probing reveals the dominant role of *Burkholderia* species in aerobic degradation of
 1376 PCBs. *FEMS Microbiol Ecol* **52**:207–217.

1377 Tittabutr, P., Cho, I.K., and Li, Q.X. (2011). Phn and Nag-like dioxygenases
 1378 metabolize polycyclic aromatic hydrocarbons in *Burkholderia* sp. C3.
 1379 *Biodegradation* doi:10.1007/s10532-011-9468-y

1380 Uhlik, O., Jecna, K., Mackova, M., Vlcek, C., Hroudova, M., Demnerova, K., Paces,
 1381 V., and Macek, T. (2009). Biphenyl-metabolizing bacteria in the rhizosphere of

1382horseradish and bulk soil contaminated by polychlorinated biphenyls as revealed
1383by stable isotope probing. *Appl Environ Microbiol* **75**:6471–6477.

1384van Berkel, W.J., Kamerbeek, N.M., and Fraaije, M.W. (2006). Flavoprotein
1385monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* **124**:670-
1386689.

1387Vedler, E. (2009). Megaplasmiids and the degradation of aromatic compounds by
1388soil bacteria. In *Microbial Megaplasmiids*. Schwartz, E. (ed). Berlin:Springer-Verlag
1389Berlin Heidelberg, pp. 33-53.

1390Watanabe, K., Teramoto, M., Futamata, H., and Harayama, S. (1998). Molecular
1391detection, isolation, and physiological characterization of functionally dominant
1392phenol-degrading bacteria in activated sludge. *Appl Environ Microbiol* **64**:4396-
13934402.

1394Weelink, S.A.B., Tan, N.C.G., ten Broeke, H., van den Kieboom, C., van Doesburg,
1395W., Langenhoff, A.A.M., *et al.* (2008). Isolation and characterization of
1396*Alicyclophilus denitrificans* strain BC, which grows on benzene with chlorate as the
1397electron acceptor. *Appl Environ Microbiol* **74**:6672-6681.

1398Xie, S., Sun, W., Luo, C., and Cupples, A.M. (2011). Novel aerobic benzene
1399degrading microorganisms identified in three soils by stable isotope probing.
1400*Biodegradation* **22**:71-81.

1401Yoshida, R., Hori, K., Fujiwara, M., Saeki, Y., and Kagamiyama, H. (1976).
1402Nonidentical subunits of protocatechuate 3,4-dioxygenase. *Biochemistry* **15**:4048-

14034053.

1404Yudistira, H., McClarty, L., Bloodworth, R.A., Hammond, S.A., Butcher, H., Mark,
1405B.L., and Cardona, S.T. (2011). Phenylalanine induces *Burkholderia cenocepacia*
1406phenylacetic acid catabolism through degradation to phenylacetyl-CoA in synthetic
1407cystic fibrosis sputum medium. *Microb Pathog* **51**:186-193.

1408Zylstra, G.J., Olsen, R.H., and Ballou, D.P. (1989). Cloning, expression, and
1409regulation of the *Pseudomonas cepacia* protocatechuate 3,4-dioxygenase genes. *J*
1410*Bacteriol* **171**:5907-5914.