Environmental Microbiology (2002) **4**(12), 799–808

Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440

K. E. Nelson,[1] C. Weinel,[2] I. T. Paulsen,[1] R. J. Dodson,[1] H. Hilbert,[3] V. A. P. Martins dos Santos,[4] D. E. Fouts,[1] S. R. Gill,[1] M. Pop,[1] M. Holmes,[1] L. Brinkac,[1] M. Beanan,[1] R. T. DeBoy,[1] S. Daugherty,[1] J. Kolonay,[1] R. Madupu,[1] W. Nelson,[1] O. White,[1] J. Peterson,[1] H. Khouri,[1] I. Hance,[1] P. Chris Lee,[1] E. Holtzapple,[1] D. Scanlan,[1] K. Tran,[1] A. Moazzez,[1] T. Utterback,[1] M. Rizzo,[1] K. Lee,[1] D. Kosack,[1] D. Moestl,[3] H. Wedler,[3] J. Lauber,[3] D. Stjepandic,[5] J. Hoheisel,[5] M. Straetz,[4] S. Heim,[4] C. Kiewitz,[2] J. Eisen,[1] K. N. Timmis,[4] A. Düsterhöft,[3] B. Tümmler[2] and C. M. Fraser[1]*

[1]*The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA.*
[2]*Klinische Forschergruppe, OE 6711, Medizinische Hochschule Hannover (MHH), Hannover, Germany.*
[3]*Qiagen GmbH, Hilden, Germany.*
[4]*Gesellschaft für Biotechnologische Forschung (GBF), Braunschweig, Germany.*
[5]*Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany.*

Summary

***Pseudomonas putida* is a metabolically versatile saprophytic soil bacterium that has been certified as a biosafety host for the cloning of foreign genes. The bacterium also has considerable potential for biotechnological applications. Sequence analysis of the 6.18 Mb genome of strain KT2440 reveals diverse transport and metabolic systems. Although there is a high level of genome conservation with the pathogenic Pseudomonad *Pseudomonas aeruginosa* (85% of the predicted coding regions are shared), key virulence factors including exotoxin A and type III secretion systems are absent. Analysis of the genome gives insight into the non-pathogenic nature of *P. putida* and points to potential new applications in agriculture, biocatalysis, bioremediation and bioplastic production.**

Received 10 October, 2002; accepted 24 October, 2002. *For correspondence. E-mail cmfraser@tigr.org; Tel. (+1) 301 838 0200; Fax (+1) 301 838 0208.

© 2002 Blackwell Science Ltd

Introduction

Pseudomonads are ubiquitous bacteria that belong to the gamma subclass of the Proteobacteria (Palleroni, 1984). These bacteria engage in important metabolic activities in the environment, including element cycling and the degradation of biogenic and xenobiotic pollutants (Timmis, 2002). Pseudomonads have considerable potential for biotechnological applications, particularly in the areas of bioremediation (Dejonghe *et al*., 2001), biocatalysis (Schmid *et al*., 2001), as biocontrol agents in plant protection (Walsh *et al*., 2001) and for the production of novel bioplastics (Olivera *et al*., 2001).

Pseudomonas putida strain KT2440 (Bagdasarian *et al*., 1981; Regenhardt *et al*., 2002) is the best characterized saprophytic Pseudomonad that has retained its ability to survive and function in the environment. The bacterium is the plasmid-free derivative of a toluene-degrading bacterium, originally designated *Pseudomonas arvilla* strain mt-2 (Kojima *et al*., 1967) and subsequently reclassified as *P. putida* mt-2 (Williams and Murray, 1974; Nakazawa, 2002). Strain KT2440 is the first Gram-negative soil bacterium to be certified as a safety strain by the Recombinant DNA Advisory Committee (Federal Register, 1982) and is the preferred host for cloning and gene expression for Gram-negative soil bacteria (Ramos *et al*., 1987). In addition to the metabolic potential of this bacterium, the ability of *P. putida* KT2440 to colonize the rhizosphere of crop plants (Espinosa-Urgel *et al*., 2002) may facilitate the development of biopesticides and plant growth promoters.

Results

The genome of strain KT2440 is a single circular chromosome, 6 181 863 bp in length with an average G+C content of 61.6% (http://www.tigr.org/; GenBank accession number AE015451). A total of 5420 open reading frames (ORFs) with an average length of 998 bp were identified (Fig. 1). These ORFs were searched against TIGRFAM (Haft *et al*., 2001) and PFAM hidden Markov models, and also against a non-redundant protein database. Preliminary name and role assignments were generated. After manual curation of these ORFs, putative role assignments could be made for 3571 ORFs, with another 600 (11.1%)

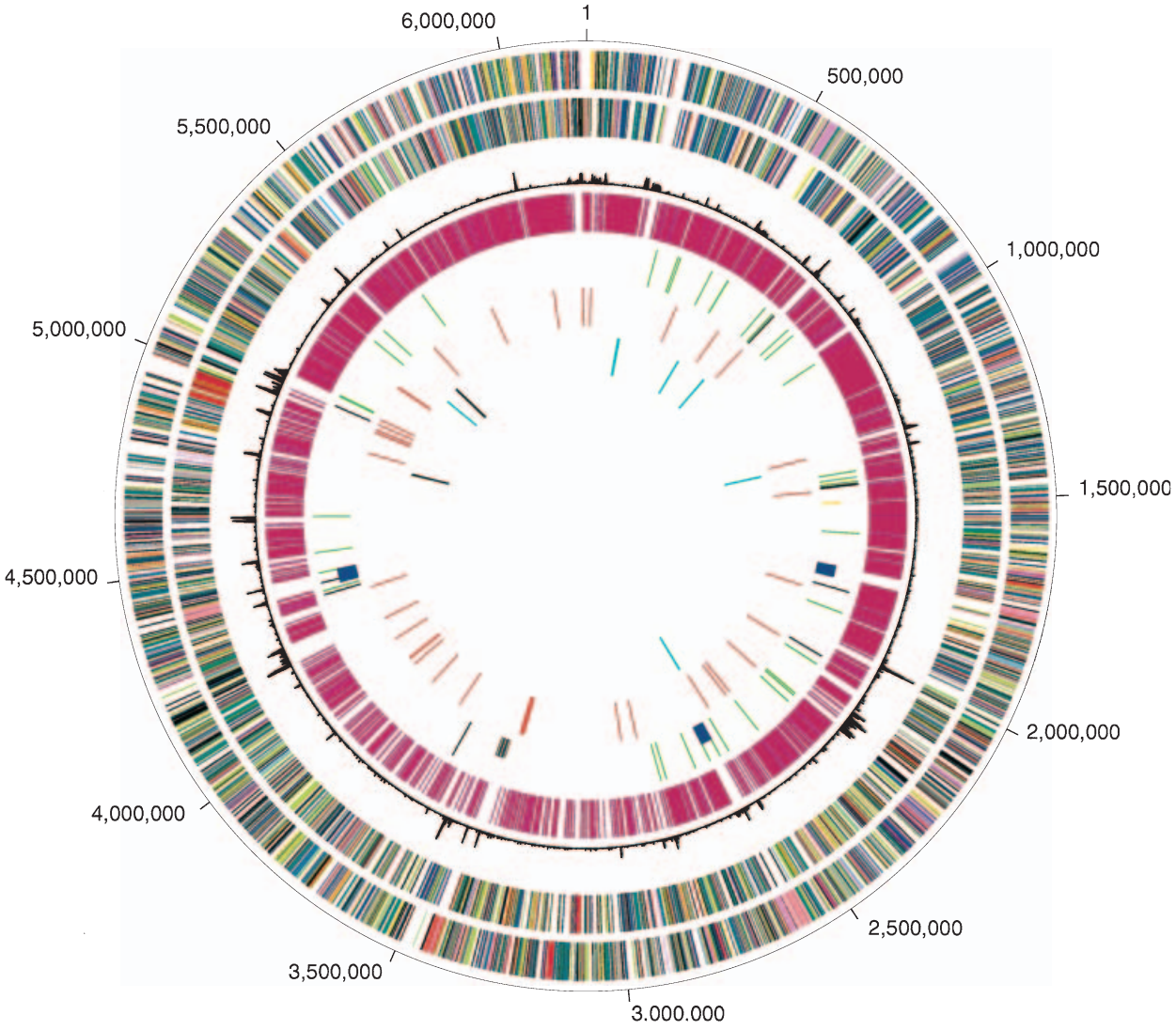

Fig. 1. Circular representation of the *P. putida* KT2440 genome. Outer circle, predicted coding regions on the plus strand colour coded by role categories: salmon, amino acid biosynthesis; light blue, biosynthesis of cofactors, prosthetic groups and carriers; light green, cell envelope; red, cellular processes; brown, central intermediary metabolism; yellow, DNA metabolism; green, energy metabolism; purple, fatty acid and phospholipid metabolism; pink, protein fate/synthesis; orange, purines, pyrimidines, nucleosides, nucleotides; blue, regulatory functions; grey, transcription; teal, transport and binding proteins; black, hypothetical and conserved hypothetical proteins. Second circle, predicted coding regions on the minus strand colour coded by role categories. Third circle, atypical trinucleotide composition of the genome. Fourth circle, top hits to the *P. aeruginosa* genome ($P < 10^{-60}$). Fifth circle, transposable elements (green), phage regions (blue), pyocins (yellow). Sixth circle, tRNAs in red. Seventh circle, rRNAs in blue, and structural RNAs in black.

being unique to *P. putida*. A total of 1037 (19.1%) of the ORFs encode conserved hypothetical proteins. Comparative genome analysis with the currently completed microbial genomes (http://www.tigr.org) revealed that 4610 ORFs (85%) of the KT2440 genome have homologues in the *Pseudomonas aeruginosa* PAO1 genome (Stover *et al*., 2000) ($P < 10^{-5}$), with 3325 (61%) having their best hit to predicted proteins in PAO1 ($P < 10^{-30}$) (Fig. 1). Five hundred and eight *P. putida* ORFs (9.4%) were identified as putative duplications with the largest family composed of 41 transposase genes. Within this group are seven novel multicopy insertion sequence (IS) elements (ISPpu8–11 and ISPpu13–15; see http://www-is.biotoul.fr/) that resemble members of the IS*66* and IS*110* gene families (Mahillon and Chandler, 1998). The target site of the ISPpu9 and ISPpu10 insertions is a unique position in a conserved 23 bp sequence (TTCGCGGGT(A/G)AACC CGCTCCTAC). These insertion sites conform to the consensus sequence of a species-specific repetitive extragenic palindromic (REP) sequence identified previously in *P. putida* KT2440 (Aranda-Olmedo *et al*., 2002). Thus, ISPpu9 and ISPpu10 are examples of insertion

sequences that selectively target a REP sequence (Wilde *et al*., 2001). This might represent a strategy for propagation that avoids harm to the host cell, as the insertion events occur outside essential genes. Seven complete ribosomal operons (5S-16S-23S), two of which are arranged in tandem (Weinel *et al*., 2001), nine full-length group II introns, four regions that encode bacteriophage and two structural RNAs (PpSRP1 and PptmRNA1) were identified in the KT2440 genome. The reader is referred to the article by Weinel *et al*. (2002) in this issue for a more comprehensive description of gene islands and genome signature in KT2440.

Metabolic analysis

Genome analysis of KT2440 reveals metabolic pathways for the transformation of a variety of aromatic compounds (Table 1). Several of these aromatics (ferulate, coniferyl and coumaryl alcohols, aldehydes and acids, vanillate, *p*-hydroxybenzoate, protocatechuate) are derivatives of lignin and may arise during decomposition of plant materials. *P. putida* appears to modify the diverse structures of these aromatics to common intermediates that can be fed into central pathways (Dagley, 1971). Initial steps in the metabolism of ferulic acid, 4-hydroxybenzoate or benzoate, for example, would be mediated by different enzymes with all routes ultimately converging via protocatechuate or catechol to the 3-oxoadipate pathway (Harwood and Parales, 1996). This convergent strategy is also seen with substrates that can be metabolized by the phenylacetyl-CoA pathway (Luengo *et al*., 2001).

Oxygenases and oxidoreductases play key roles in the chemical transformation of recalcitrant compounds (Resnick and Gibson, 1996; http://umbbd.ahc.umn.edu/). Analysis of the KT2440 genome reveals 18 dioxygenases, including the structurally and functionally thoroughly char-

Table 1. Putative substrates identified from *in silico* whole-genome analysis of *P. putida* KT2440.

Putative substrate	Potential end-products	Most similar to	ORF number
Aromatic/aliphatic sulphonates	Phenol + sulphite	*Agrobacterium tumefaciens* *Pseudomonas putida*	PP3219; PP2765 PP0241–PP0236
Benzoate, toluate	Catechol (2-hydro-1,2-dihydroxybenzoate)	*Pseudomonas putida*	PP3161–PP3164
Catechol (2-hydro-1,2-dihydroxybenzoate)	*cis*,*cis*-muconic acid	*Pseudomonas putida*	PP3713; PP3166
Coniferyl alcohol	Ferulic acid (4-hydroxy-3-methoxycinnamate)	*Pseudomonas* sp.	PP5120
2-Cyclohexen-1-one	Cyclohexanone	*Pseudomonas syringae*	PP1478; PP1254; PP2489
4-Hydroxybenzoate	Protocatechuate (3,4-dihydroxybenzoate)	*Pseudomonas putida*	PP3537
Hippurate (benzoylglycine)	Benzoate + glycine	*S. meliloti*	PP2704
Isoquinoline	Isoquinolin-1(^{2}H)-one	*Pseudomonas dimuta* *S. meliloti*	PP3622–PP3621; PP2478–PP2477
Ferulic acid (4-hydroxy-3-methoxycinnamate)	Vanillin (4-hydroxy-3-methoxybenzaldehyde)	*Pseudomonas* sp.	PP3354–PP3358
Maleate	Fumarate	*Serratia marcescens*	PP3942
Phenolsulphates	Phenol + sulphate	*Pseudomonas aeruginosa*	PP3352
Phenolacetate	Phenol + acetate	*Pseudomonas fluorescens*	PP5253
Phenylalanine	Tyrosine	*Pseudomonas aeruginosa*	PP4490
Phenylacetic acid 2-phenylethylamine phenylacetaldehyde phenylalkanoic acids	TCA intermediates	*Pseudomonas putida*	PP3284–PP3270
3-Polyhydroxybutyric acid phenylalkanoic acids	Polyhydroxyalkanoates (polyesters)	*P. putida* U	PP5006–PP5003
Propanediol	3-Hydroxypropanol	*Klebsiella pneumoniae*	PP2803
Protocatechuate (3,4-dihydroxybenzoate)	Succinate, acetate	*Pseudomonas putida*	PP4656–PP4655; PP3952–PP3951; PP1381–PP1379
Quinate	5-Dehydroquinate to protocatechuate	*Xanthomonas campestris*	PP3569
Taurine	Aminoacetaldehyde + succinate		PP0230; PP0169; PP4466
TNT (2,4,6 trinitrotoluene)	2-Hydroxylamino-4,6-dinitrotoluene and 4-hydroxylamino-2, 6-dinitrotoluene	*Pseudomonas putida*	PP0920
Vanillin (4-hydroxy-3-methoxybenzaldehyde)	Protocatechuate (3,4-dihydroxybenzoate)	*Pseudomonas* sp.	PP3357; PP3737–PP3736
Stachydrine	Proline	*S. meliloti*	PP4753–PP4752

The ORFs indicated by 'ORF number' represent the subset in the respective pathway of which homologues with experimentally demonstrated function exist in the organism listed in the adjacent column.

acterized heterodimeric $(\alpha\beta)_3$ benzoate dioxygenase (PP3161–PP3163) (Wolfe *et al*., 2002) $(\alpha\beta)_4$ protocatechuate 3,4-dioxygenase (PP4656–PP4655) (Bull and Ballou, 1981) and two homotetrameric α_4 catechol 1,2-dioxgenases (PP3713 and PP3166) (Kita *et al*., 1999). Three homologues of genes coding for taurine family dioxygenases (PP0230, PP0169, PP4466) and genes encoding 15 monooxygenases, mostly of unknown specificity, are also present. The KT2440 genome encodes 80 oxidoreductases of unknown substrate specificity, at least half of which are located within gene clusters encoding oxidoreductases, transporters and ferredoxins/cytochromes. Two paralogues (PP2489 and PP1478) and an orthologue (PP1254) of the previously described *P. putida xen*A exist that encode an enzyme shown to reduce 2-cyclohexen-1-one. The single *xen*B (PP0920) is homologous to the *Pseudomonas fluorescens xen*B that transforms 2,4,6-trinitrotoluene (TNT) into more than a dozen metabolites by aromatic ring reduction and nitro group reduction (Blehert *et al*., 1999; Pak *et al*., 2000). In addition, the genome encodes 51 putative hydrolases, more than 62 transferases and 40 dehydrogenases, all of unknown substrate specificity and belonging to a number of different protein families. The *ssuFBCDAE* operon (PP0241–PP0235) may enable growth on aromatic and aliphatic sulphonates as the sulphur source, and the 12 glutathione-*S*-transferases that are encoded by the genome may be involved in detoxification processes (Vuilleumier and Pagni, 2002). The presence of three putative chlorohydrolases (PP3209, PP2584 and PP5036) implies a capacity for hydrolysing chloride substituents from chloroorganics. A putative chloride channel (PP3959) was identified that may allow for the extrusion of chloride ions. The catabolic potential of this strain is highlighted in Fig. 2.

Genome analysis of KT2440 reveals that extensive metabolic abilities are chromosomally encoded and implicates mobile genetic elements in their acquisition. For instance, the 82 transposase genes, nine group II introns, a newly identified Tn*7*-like element (PP5407–PP5404) and the previously characterized Tn*4652* (PP2984–PP2964) (Bayley *et al*., 1977) are clustered in 47 regions of the KT2440 chromosome, 24 of which (51%) are flanked directly by genes involved in energy metabolism and transport. These regions may be involved in restructuring the chromosome or modifying host gene expression. For example, the 64.8 kb insertion in a thymidylate kinase gene (N-terminus, PP1919; C-terminus, PP1965) encodes many genes involved in energy metabolism including four putative oxidoreductases, two cytochrome P450 family proteins and a vanillate demethylase, and IS elements disrupt at least five genes including homologues of a malate/L-lactate dehydrogenase, a peptidase and a glutaminase.

Consistent with the extensive metabolic versatility for the degradation of aromatics, KT2440 encodes more putative transporters for aromatic substrates than any currently sequenced microbial genome, including multiple homologues of the *Acinetobacter calcoaceticus* benzoate transporter BenK and of the *P. putida* 4-hydroxybenzoate transporter PcaK (Nichols and Harwood, 1997). In addition, KT2440 has 23 members of the BenF/PhaK/OprD family of porins that includes outer membrane channels implicated in the uptake of aromatic substrates (Olivera *et al*., 1998; Cowles *et al*., 2000).

KT2440 also possesses ≈ 350 cytoplasmic membrane transport systems, 15% more than *P. aeruginosa*, including twice as many predicted ATP-binding cassette (ABC) amino acid uptake transporters. In addition, the presence of 11 LysE family amino acid efflux transporters (*P. aeruginosa* only has one) suggests a key role in preventing the accumulation of inhibitory levels of amino acids and their analogues in the cell. The five predicted APC (amino acid/polyamine/organocation) family GABA (gamma-aminobutyric acid) transporters may be involved in the uptake of butyric acid, which can be converted to polyhydroxyalkanoic acids (bioplastics) (Garcia *et al*., 1999). *P. putida* has a paucity of carbohydrate uptake systems with, for instance, only a single PTS transporter with probable specificity for fructose. Dicarboxylates that appeared to be a significant nutrient source for *P. aeruginosa* may be of less significance for KT2440, which has only two incomplete TRAP family dicarboxylate transporters compared with at least four in *P. aeruginosa*. *P. putida* KT2440 has a large repertoire of ABC amino acid transporters, and this could reflect a physiological emphasis on the metabolism of amino acids and their derivatives for successful competition in the rhizosphere (see below). Also consistent with its ability to colonize plant roots, KT2440 possesses a predicted ABC family opine transporter (PP4458–PP4453) that has been described previously for other rhizosphere microorganisms (Lyi *et al*., 1999), and enzymes for the metabolism of opines (Fig. 2), suggesting an ability to exploit plant-produced opines induced in the rhizosphere by other bacterial species. Moreover, an efflux system for the export of fusaric acid (PP1266–PP1263) was identified, a common fungal toxin produced by phytopathogens, such as *Fusarium oxysporum* (Schnider-Keel *et al*., 2000), which reinforces the potential of *P. putida* for biocontrol of fungal pathogens of plants.

Pseudomonas putida exhibits a complex repertoire of chemosensory systems, including one for flagella-mediated swimming (PP4332–PP4340, PP4392–PP4393), one for twitching motility towards chemoattractants (PP4992–PP4987) and one as yet uncharacterized but also involved in chemotaxis and motility (PP1494–PP1488) that are shared with *P. aeruginosa* PA01. Another putative chemosensory system that is unique to KT2440 (PP3762–PP3757) is composed of a CheY-like protein, response regulator/sensor kinase fusion protein,

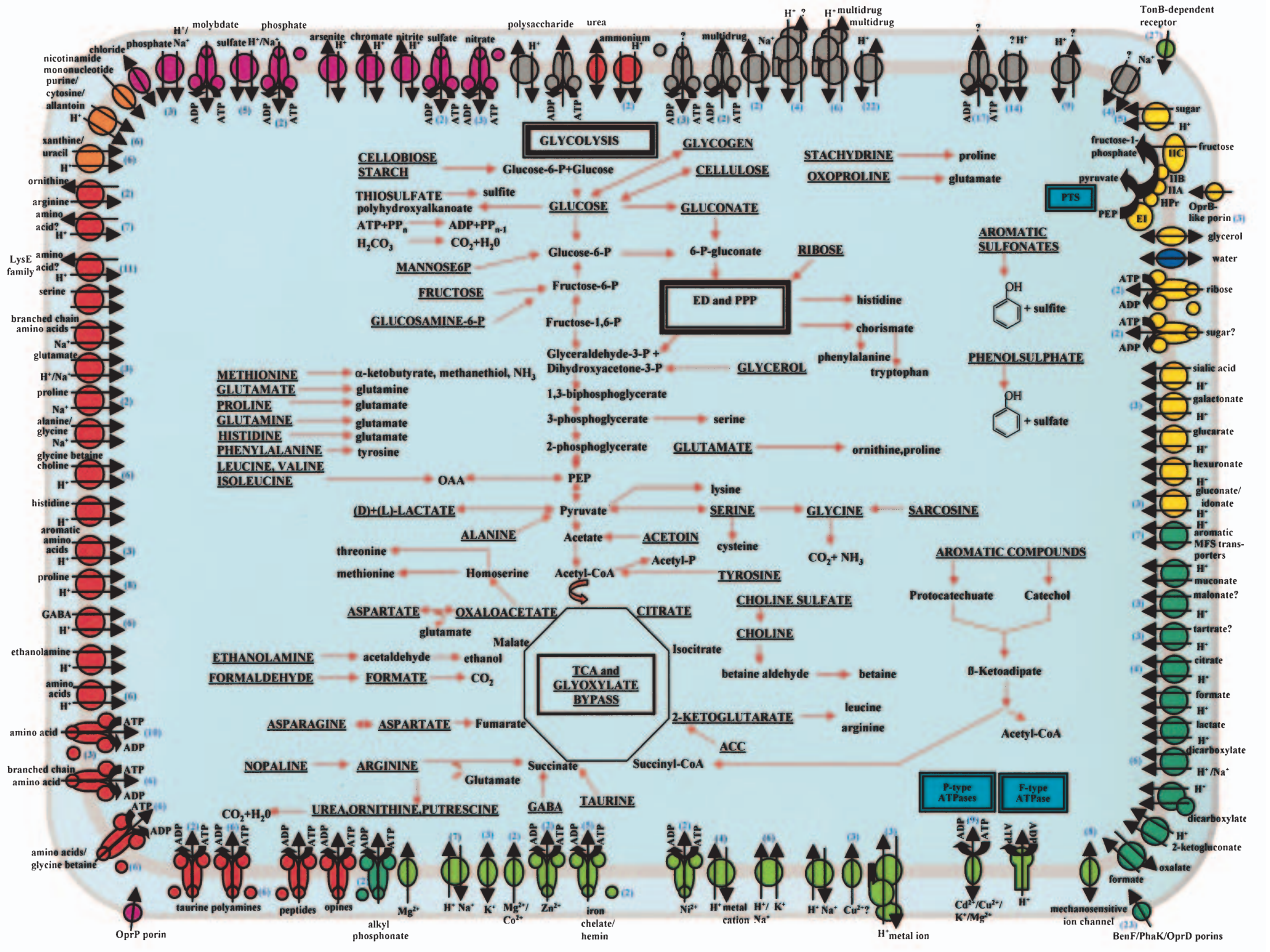

Fig. 2. Overview of metabolism and transport in *P. putida*. Predicted pathways for energy production and metabolism of organic compounds are shown. Predicted transporters are grouped by substrate specificity: inorganic cations (light green), inorganic anions (pink), carbohydrates (yellow), amino acids/peptides/amines/purines/pyrimidines and other nitrogenous compounds (red), carboxylates, aromatic compounds and other carbon sources (dark green), water (blue), drug efflux and other (dark grey). Question marks indicate uncertainty about the substrate transported. Export or import of solutes is designated by the direction of the arrow through the transporter. The energy-coupling mechanisms of the transporters are also shown: solutes transported by channel proteins are shown with a double-headed arrow; secondary transporters are shown with two arrowed lines indicating both the solute and the coupling ion; ATP-driven transporters are indicated by the ATP hydrolysis reaction; transporters with an unknown energy-coupling mechanism are shown with only a single arrow. The P-type ATPases are shown with a double-headed arrow to indicate that they include both uptake and efflux systems. Where multiple homologous transporters with similar substrate predictions exist, the number of that type of transporter is indicated in parentheses. The outer and inner membrane are sketched in grey, the periplasmic space is indicated in light turquoise and the cytosol in turquoise.

CheB-like protein, CheR-like protein and a series of histidine sensor kinase fusion proteins. Of the 27 methyl-accepting protein (MCP) genes that are scattered throughout the chromosome, three encode receptors that probably sense amino acids (PP0320, PP2249, PP1371), and three appear to encode aerotaxis receptors (PP2257, PP2111, PP4521), which may be important for migration to preferred environmental conditions.

Comparisons of P. putida *with other Pseudomonads*

Although *P. putida* KT2440 has been approved as a biological safety strain, the bacterium shares a close evolutionary relationship with the opportunistic human pathogen *P. aeruginosa* (Bodey *et al.*, 1983) and other pseudomonads that are plant pathogens. The genome of *P. putida* contains 1231 ORFs (22.7% of total) that are absent from that of *P. aeruginosa*. These include 226 conserved hypothetical proteins, 575 hypothetical proteins, a cellulose biosynthesis operon, a novel polysaccharide biosynthesis operon, 36 putative transposable elements, three phage, nine group II introns, two toluene resistance proteins, 75 proteins that are involved in energy metabolism and type IV pilus biosynthesis genes. The presence of an operon for the biosynthesis of cellulose (PP2638–PP2634), as seen in *Agrobacterium tumefa-*

ciens and *Rhizobium* sp., as well as a previously uncharacterized operon for the synthesis of polysaccharides (PP3142–PP3132) imply that the polymers synthesized by these two gene clusters may be important for the attachment of KT2440 to plant roots. Among the 1281 ORFs (22.7% of total) that are present in *P. aeruginosa* but absent from the genome of *P. putida* ($P < 10^{-05}$) are 852 hypothetical proteins and 137 conserved hypothetical proteins. Key determinants of virulence and virulence-associated traits present in the *P. aeruginosa* genome, but apparently absent from that of *P. putida*, include genes for exotoxin A, elastase, exolipase, phospholipase C, alkaline protease, a type III secretion pathway, two Nramp manganese/iron transporters and an operon for the synthesis of rhamnolipids.

The only genes in *P. putida* KT2440 specifying products related to virulence-associated traits are adhesion proteins, specified by PP0168, PP1449 and PP0806, with very high molecular weights (PP0168 is the largest gene in the genome and specifies a protein of 8682 amino acids) that are homologous to adhesin genes of enteropathogens. However, Espinosa-Urgel *et al.* (2000) originally identified these proteins as being essential for seed colonization by *P. putida*. Moreover, homologues of the proteins encoded by the KT2440 gene cluster PP5231–PP5226 have been shown to promote plant rhizosphere interactions in *P. fluorescens* (Dekkers *et al.*, 1998). Thus, the PP0168, PP1449 and PP0806 proteins are probably not virulence-mediating traits in KT2440 but, rather, are probably general adhesion molecules that, in conjunction with the gene cluster PP5231–PP5226, play roles in the rhizosphere lifestyle that have been reported for *P. putida*.

Additional systems that resemble virulence-associated traits were identified by whole-genome analyses. Quorum sensing, for example, is a virulence-associated trait in *P. aeruginosa*. Although KT2440 has genes for homoserine lactone production, it apparently does not produce detectable amounts of this signalling molecule (L. Eberl, personal communication).

The mucoid morphotype caused by overproduction of alginate is one hallmark of chronic *P. aeruginosa* infection in the lungs of cystic fibrosis patients (Henry *et al.*, 1992). Twenty-three of the 24 genes of the alginate regulatory and biosynthesis system (Ohman *et al.*, 1996) are present in the KT2440 genome, although the transcriptional regulatory gene *algM*/*mucC* is absent. Loss of AlgM leads to a non-mucoid phenotype in *P. aeruginosa*, suggesting that the lack of AlgM in KT2440 accounts for its non-mucoid morphotype under standard culturing conditions. It remains to be seen whether, under stress conditions such as antibiotic stress (Govan *et al.*, 1981), alginate biosynthesis in *P. putida* may be induced. *P. aeruginosa* is also known for its intrinsic resistance to antimicrobial compounds probably in large part because of its complement of RND/MFP/OMF efflux systems and, in particular, MexAB-OprM (Poole, 2001). The KT2440 genome specifies a similar complement of RND efflux systems, and phylogenetic analysis (Fig. 3) strongly suggests that most of the *P. aeruginosa* systems, including MexB, MexD and MexF, have orthologues in *P. putida*. The location of some of these genes in KT2440 adjacent to those of the BenF/PhaK family porins suggests that their actual physiological role may be the efflux of toxic substrates or metabolites that may accumulate in the cell. One RND efflux pump of KT2440, encoded by PP1386–PP1385, has been shown experimentally to be involved in toluene resistance (Rojas *et al.*, 2001), and at least one ABC transporter exhibits similarity to efflux pumps for organic solvents (PP0960–PP0958; Godoy *et al.*, 2001).

Finally, a 33.5 kb region that spans PP2815–PP2790 in KT2440 shares gene order and 90% similarity with part of the *P. aeruginosa* genomic island PAGI-1, a 48.9 kb chromosomal region present in 85% of *P. aeruginosa* clinical isolates from sepsis and urinary tract infections (Liang *et al.*, 2001). The first four ORFs of PAGI-1, which are predicted to be involved in oxidative stress response, and

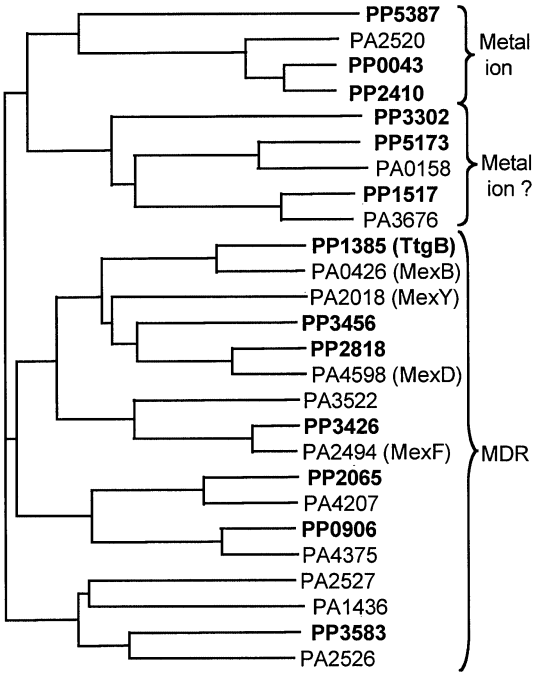

Fig. 3. Phylogenetic representation of the RND family efflux systems in *P. putida* KT2440 and *P. aeruginosa* PA01. Amino acid sequences of the RND proteins were aligned using CLUSTALW, and a neighbour-joining tree was generated from the alignment using the PAUP program. All nodes had bootstrap values in excess of 85%. The probable specificity of each cluster of proteins is indicated on the right, and proteins for which experimental evidence is available are named (TtgB, MexB/Y/D/F); for details, see Poole (2001). One *P. putida* RND transporter with a predicted frameshift (PP3584) was excluded from the alignment and tree construction.

the final 21 ORFs, which encode transposases or hypothetical proteins, are missing from *P. putida*. PP2799 (encoding a putative class III aminotransferase) is specific to KT2440. The last 21 ORFs of PAGI-1 have an atypically low GC content, suggesting that the *P. aeruginosa* strains acquired the island by horizontal gene transfer. In KT2440, the PAGI-1 homologues appear to be part of the conserved core genome, implying that the shared ORFs may not be key virulence traits (Liang *et al*., 2001).

Comparison of the KT2440 genome with those of the phytopathogens *Pseudomonas syringae*, *Ralstonia solanacearum*, *Xanthomonas campestris* and *Xylella fastidiosa* reveals that KT2440 lacks the genes of practically all known plant-related virulence traits, such as type III secretion systems and corresponding secreted substrates (Fouts *et al*., 2002; Guttman *et al*., 2002; Salanoubat *et al*., 2002), as well as plant cell wall-degrading enzymes (Cao *et al*., 2001). The KT2440 genome does contain two operons (PP3790–PP3781 and PP2788–PP2777) related to those for the biosynthesis of secondary metabolites, such as phytotoxin peptides and antibiotics respectively (Bender *et al*., 1999).

Discussion

Pseudomonas putida is a paradigm of a class of cosmopolitan opportunistic bacteria found in terrestrial and aquatic environments everywhere that can use a vast range of substrates for their growth. This study has revealed an unusual wealth of putative determinants for transporters, mono- and dioxygenases, oxidoreductases, ferredoxins and cytochromes, dehydrogenases, sulphur metabolism proteins, etc., and of efflux pumps and glutathione-*S*-transferases ordinarily associated with protection against toxic substrates and metabolites that provides the genetic basis for the exceptional metabolic versatility of *P. putida*.

The analyses presented here provide an opportunity for the development of biotechnological applications. The extremely high number and diversity of enzymes and transporters for secondary metabolites support activities that considerably exceed the previously known metabolic spectrum of KT2440. Whole-genome analysis further underscores the utility of KT2440 for the experimental design of novel pathways for the catabolism of organic pollutants and the potential for bioremediation of soils contaminated with such compounds. As the limited number of enzymes that have been characterized in KT2440 so far typically mediate difficult chemical reactions, some of the newly identified systems revealed by this genome analysis may equally have promise for biotransformations, such as those for the production of epoxides, substituted catechols, enantiopure alcohols and amides and heterocyclic compounds (Faber, 2000). In addition, identification of genes for the synthesis of polyhydroxyalkanoic acids may facilitate the development of novel bioplastics (Garcia *et al*., 1999; Gorenflo *et al*., 2001; Luengo *et al*., 2001; Olivera *et al*., 2001).

Large-scale production of recombinant products and release of recombinant strains into the environment requires the use of a well-characterized host strain with impeccable biosafety credentials. Genomic comparisons between the saprophytic KT2440 strain and plant- and animal-pathogenic Pseudomonads have highlighted differences that account for the different lifestyles and biological properties of these organisms. KT2440 lacks a spectrum of key virulence determinants of pathogenic Pseudomonads that mediate host damage, including exotoxins, specific hydrolytic enzymes, type III secretion systems, factors mediating hypersensitive responses, etc., which clearly account for its avirulence. Genetic determinants that are shared between KT2440 and pathogenic species suggest that certain properties, such as adhesion and polymer biosynthesis, type IV pili, adhesins, stress-related proteins and master regulators such as *gac*A, which are commonly considered to be important for pathogenesis in both plants and animals (Cao *et al*., 2001), may in fact only be important for effective colonization and survival on surfaces, i.e. may be general survival functions and not obligatorily related to pathogenesis, which requires damage of a host. This genomic analysis has thus confirmed the avirulence of *P. putida* KT2440, provided a definitive genetic basis for the biosafety characteristics of this bacterium and generated a database upon which the environmental and biotechnological behaviour of this strain can be interpreted.

Experimental procedures

Sequencing, assembly and gap closure

Cloning, sequencing and assembly were performed as described previously for genomes sequenced by TIGR. Basically, two small-insert plasmid libraries (1.5–2.5 kb) were generated by random mechanical shearing of genomic DNA. Two large-insert libraries were generated in BAC and cosmid clones respectively. In the initial random sequencing phase, approximately eightfold sequence coverage was achieved. The sequences from all libraries were assembled jointly using TIGR ASSEMBLER. Sequence gaps were closed by editing the ends of sequence traces and/or primer walking on plasmid clones. Physical gaps were closed by combinatorial polymerase chain reaction (PCR) followed by sequencing of the PCR product.

ORF prediction and gene family identification

An initial set of ORFs likely to encode proteins was identified by GLIMMER (Salzberg *et al*., 1998), and those shorter than 30 codons were eliminated. ORFs that overlapped were

inspected visually and, in some cases, removed. ORFs were searched against a non-redundant protein database as described previously. Frameshifts and point mutations were detected and corrected where appropriate as described previously. Remaining frameshifts and point mutations are considered authentic, and corresponding regions were annotated as 'authentic frameshift' or 'authentic point mutation' respectively. Two sets of hidden Markov models (HMMs) were used to determine ORF membership in families and superfamilies. These included 721 HMMs from PFAM version 2.0 and 631 HMMs from the TIGR orthologue resource. TOPPRED was used to identify membrane-spanning domains (MSDs) in proteins.

Comparative genomics

All genes and predicted proteins from the *P. putida* genome, as well as from all other published completed genomes (see http://www.tigr.org), were compared using BLAST. For the identification of recent gene duplications, all genes from the *P. putida* KT2440 genome were compared with each other. A gene was considered to be recently duplicated if the most similar gene (as measured by *P*-value) was another gene within the same genome (relative to genes from other genomes).

Atypical nucleotide composition

χ^2 analysis is designed to highlight regions of the genome with atypical nucleotide composition (Fig. 1). A χ^2 statistic was computed comparing the trinucleotide composition on DNA subsequences in 2000 bp windows with the rest of the genome. The distribution of all 64 trinucleotides (3-mers) in the complete genome sequence was computed, followed by the 3-mer distribution in 2000 bp windows across the genome. We used windows that overlapped by half their length, 1000 bp. A large value indicates that the composition in the DNA subsequence is atypical of the genome.

Database submission

The nucleotide sequence of the whole genome of *P. putida* was submitted to GenBank under accession number AE015451.

Acknowledgements

The authors would like to thank Michael Holmes, Michael Heaney, Herve Tettelin, John Heidelberg, Michael Kertesz, Prabhakar Salunkhe, Michael Brown and Ed Moore for assistance at various stages of the project. Financial support by the US Department of Energy and the Bundesministerium für Bildung und Forschung is gratefully acknowledged.

References

Aranda-Olmedo, I., Tobes, R., Manzanera, M., Ramos, J.L., and Marques, S. (2002) Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*. *Nucleic Acids Res* **15:** 1826–1833.

Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F.C., Bagdasarian, M.M., Frey, J., and Timmis, K.N. (1981) Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. *Gene* **16:** 237–247.

Bayley, S.A., Duggleby, C.J., Worsey, M.J., Williams, P.A., Hardy, K.G., and Broda, P. (1977) Two modes of loss of the Tol function from *Pseudomonas putida* mt-2. *Mol Gen Genet* **154:** 203–204.

Bender, C.L., Alarcon-Chaidez, F., and Gross, D.C. (1999) *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **63:** 266–292.

Blehert, D.S., Fox, B.G., and Chambliss, G.H. (1999) Cloning and sequence analysis of two *Pseudomonas* flavoprotein xenobiotic reductases. *J Bacteriol* **181:** 6254–6263.

Bodey, G.P., Bolivar, R., Fainstein, V., and Jadeja, L. (1983) Infections caused by *Pseudomonas aeruginosa*. *Rev Infect Dis* **5:** 279–313.

Bull, C., and Ballou, D.P. (1981) Purification and properties of protocatechuate 3,4-dioxygenase from *Pseudomonas putida*. A new iron to subunit stoichiometry. *J Biol Chem* **256:** 12673–12680.

Cao, H., Krishnan, G., Goumnerov, B., Tsongalis, J., Tompkins, R., and Rahme, L.G. (2001) A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. *Proc Natl Acad Sci USA* **98:** 14613–14618.

Cowles, C.E., Nichols, N.N., and Harwood, C.S. (2000) BenR, a XylS homologue, regulates three different pathways of aromatic acid degradation in *Pseudomonas putida*. *J Bacteriol* **182:** 6339–6346.

Dagley, S. (1971) Catabolism of aromatic compounds by micro-organisms. *Adv Microb Physiol* **6:** 1–46.

Dejonghe, W., Boon, N., Seghers, D., Top, E.M., and Verstraete, W. (2001) Bioaugmentation of soils by increasing microbial richness: missing links. *Environ Microbiol* **3:** 649–657.

Dekkers, L.C., Phoelich, C.C., van der Fits, L., and Lugtenberg, B.J. (1998) A site-specific recombinase is required for competitive root colonization by *Pseudomonas fluorescens* WCS365. *Proc Natl Acad Sci USA* **95:** 7051–7056.

Espinosa-Urgel, M., Salido, A., and Ramos, J.L. (2000) Genetic analysis of functions involved in adhesion of *Pseudomonas putida* to seeds. *J Bacteriol* **182:** 2363–2369.

Espinosa-Urgel, M., Kolter, R., and Ramos, J.L. (2002) Root colonization by *Pseudomonas putida*: love at first sight. *Microbiology* **148:** 341–343.

Faber, K. (2000) *Biotransformations in Organic Chemistry*. Springer Verlag, Berlin.

Federal Register (1982) Certified host–vector systems, p. 17197.

Fouts, D.E., Abramovitch, R.B., Alfano, J.R., Baldo, A.M., Buell, C.R., Cartinhour, S., *et al.* (2002) Genomewide identification of *Pseudomonas syringae* pv. tomato DC3000

promoters controlled by the HrpL alternative sigma factor. *Proc Natl Acad Sci USA* **99:** 2275–2280.

Garcia, B., Olivera, E.R., Minambres, B., Fernandez-Valverde, M., Canedo, L.M., Prieto, M.A., *et al.* (1999) Novel biodegradable aromatic plastics from a bacterial source. Genetic and biochemical studies on a route of the phenylacetyl-coa catabolon. *J Biol Chem* **274:** 29228–29241.

Godoy, P., Ramos-Gonzalez, M.I., and Ramos, J.L. (2001) Involvement of the TonB system in tolerance to solvents and drugs in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **183:** 5285–5292.

Gorenflo, V., Schmack, G., Vogel, R., and Steinbüchel, A. (2001) Development of a process for the biotechnological large-scale production of 4-Hydroxyvalerate-containing polyesters and characterization of their physical and mechanical properties. *Biomacromolecules* **2:** 45–57.

Govan, J.R., Fyfe, J.A., and Jarman, T.R. (1981) Isolation of alginate-producing mutants of *Pseudomonas fluorescens*, *Pseudomonas putida* and *Pseudomonas mendocina*. *J Gen Microbiol* **125:** 217–220.

Guttman, D.S., Vinatzer, B.A., Sarkar, S.F., Ranall, M.V., Kettler, G., and Greenberg, J.T. (2002) A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science* **295:** 1722–1726.

Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., and White, O. (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29:** 41–43.

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

Henry, R.L., Mellis, C.M., and Petrovic, L. (1992) Mucoid *Pseudomonas aeruginosa* is a marker of poor survival in cystic fibrosis. *Pediatr Pulmonol* **12:** 158–161.

Kita, A., Kita, S., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K., *et al.* (1999) An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure Fold Des* **15:** 25–34.

Kojima, Y., Fujisawa, H., Nakazawa, A., Nakazawa, T., Kanetsuna, F., Taniuchi, H., *et al.* (1967) Studies on pyrocatechase. I. Purification and spectral properties. *J Biol Chem* **242:** 3270–3278.

Liang, X., Pham, X.Q., Olson, M.V., and Lory, S. (2001) Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **183:** 843–853.

Luengo, J.M., Garcia, J.L., and Olivera, E.R. (2001) The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications. *Mol Microbiol* **39:** 1434–1442.

Lyi, S.M., Jafri, S., and Winans, S.C. (1999) Mannopinic acid and agropinic acid catabolism region of the octopine-type Ti plasmid pTi15955. *Mol Microbiol* **31:** 339–347.

Mahillon, J., and Chandler, M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62:** 725–774.

Nakazawa, T. (2002) Travels of a *Pseudomonas*, from Japan around the world. *Environ Microbiol* **4:** 782–786.

Nichols, N.N., and Harwood, C.S. (1997) PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*. *J Bacteriol* **179:** 5056–5061.

Ohman, D.E., Mathee, K., McPherson, C.J., DeVries, C.A., Ma, S., Wozniak, D.J., and Franklin, M.J. (1996) Regulation of the alginate (algD) operon in *Pseudomonas aeruginosa*. In *Molecular Biology of Pseudomonads.* Nakazawa, T., Haas, D., and Silver, S. (eds). Washington, DC: American Society for Microbiology Press, pp. 472–483.

Olivera, E.R., Minambres, B., Garcia, B., Muniz, C., Moreno, M.A., Ferrandez, A., *et al.* (1998) Molecular characterization of the phenylacetic acid catabolic pathway in *Pseudomonas putida* U: the phenylacetyl-CoA catabolon. *Proc Natl Acad Sci USA* **95:** 6419–6424.

Olivera, E.R., Carnicero, D., Jodra, R., Minambres, B., Garcia, B., Abraham, G.A., *et al.* (2001) Genetically engineered *Pseudomonas*: a factory of new bioplastics with broad applications. *Environ Microbiol* **3:** 612–618.

Pak, J.W., Knoke, K.L., Noguera, D.R., Fox, B.G., and Chambliss, G.H. (2000) Transformation of 2,4,6-trinitrotoluene by purified xenobiotic reductase B from *Pseudomonas fluorescens* I-C. *Appl Environ Microbiol* **66:** 4742–4750.

Palleroni, N. (1984) *Family I. Pseudomonaceae.* Baltimore, MD: Williams & Wilkins.

Poole, K. (2001) Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *J Mol Microbiol Biotechnol* **3:** 255–264.

Ramos, J.L., Wasserfallen, A., Rose, K., and Timmis, K.N. (1987) Redesigning metabolic routes: manipulation of TOL plasmid pathway for catabolism of alkylbenzoates. *Science* **235:** 593–596.

Regenhardt, D., Heuer, H., Heim, S., Fernandez, D.U., Strömpl, C., Moore, E.R.B., and Timmis, K.N. (2002) Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environ Microbiol* **4:** 912–915.

Resnick, S.M., and Gibson, D.T. (1996) Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. *J Ind Microbiol* **17:** 438–457.

Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J.L., and Segura, A. (2001) Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J Bacteriol* **183:** 3967–3973.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., *et al.* (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415:** 497–502.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26:** 544–548.

Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature* **409:** 258–268.

Schnider-Keel, U., Seematter, A., Maurhofer, M., Blumer, C., Duffy, B., Gigot-Bonnefoy, C., *et al.* (2000) Autoinduction of 2,4-diacetylphloroglucinol biosynthesis in the biocontrol agent *Pseudomonas fluorescens* CHA0 and repression by the bacterial metabolites salicylate and pyoluteorin. *J Bacteriol* **182:** 1215–1225.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., *et al.* (2000) Complete genome

sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406:** 959–964.

Timmis, K.N. (2002) *Pseudomonas putida*: a cosmopolitan opportunist *par excellence*. *Environ Microbiol* **4:** 779–781.

Vuilleumier, S., and Pagni, M. (2002) The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. *Appl Microbiol Biotechnol* **58:** 138–146.

Walsh, U.F., Morrissey, J.P., and O'Gara, F. (2001) Pseudomonas for biocontrol of phytopathogens: from functional genomics to commercial exploitation. *Curr Opin Biotechnol* **12:** 289–295.

Weinel, C., Tümmler, B., Hilbert, H., Nelson, K.E., and Kiewitz, C. (2001) General method of rapid Smith/Birnstiel mapping adds for gap closure in shotgun microbial genome sequencing projects: application to *Pseudomonas putida* KT2440. *Nucleic Acids Res* **15:** E110.

Weinel, C., Nelson, K.E., and Tümmler, B. (2002) Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ Microbiol* **4:** 809–818.

Wilde, C., Bachellier, S., Hofnung, M., and Clement, J.M. (2001) Transposition of IS1397 in the family Enterobacteriaceae and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units. *J Bacteriol* **183:** 4395–4404.

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

Wolfe, M.D., Altier, D.J., Stubna, A., Popescu, C.V., Munck, E., and Lipscomb, J.D. (2002) Benzoate 1,2-dioxygenase from *Pseudomonas putida*: single turnover kinetics and regulation of a two-component Rieske dioxygenase. *Biochemistry* **30:** 9611–9626.