The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis

Eugeni Belda,1,2‡ Ruben G. A. van Heck,3‡ Maria José Lopez-Sanchez,1,4 Stéphane Cruveiller,1 Valérie Barbe,5 Claire Fraser,6 Hans-Peter Klenk,7,8 Jörn Petersen,7 Anne Morgat,9 Pablo I. Nikel,10 David Vallenet,1 Zoé Rouy,1 Agnieszka Sekowska,4 Vitor A. P. Martins dos Santos,3 Víctor de Lorenzo,10 Antoine Danchin4 and Claudine Medigue1

1Alternative Energies and Atomic Energy Commission (CEA), Genomic Institute & CNRS-UMR8030 & Evry University, Laboratory of Bioinformatics Analysis in Genomics and Metabolism, 2 rue Gaston Crémieux, 91057 Evry, France.
2Institut Pasteur, Unit of Insect Vector Genetics and Genomics, Department of Parasitology and Mycology, 28, rue du Dr. Roux, Paris, Cedex 15, 75724, France.
3Laboratory of Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, Building number 316, 6703 HB Wageningen, The Netherlands.
4AMAbiotics SAS, Institut du Cerveau et de la Moelle Épinière, Hôpital de la Pitié-Salpétrière, Paris, France.
5Alternative Energies and Atomic Energy Commission (CEA), Genomic Institute, National Sequencing Center, 2 rue Gaston Crémieux, 91057 Evry, France.
6Institute for Genome Sciences, Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA.
7Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
8School of Biology, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.
9Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Geneva, CH-1206, Switzerland.
10Systems and Synthetic Biology Program, Centro Nacional de Biotecnologia (CNB-CSIC), C/Darwin 3, 28049, Madrid, Spain.

Summary

By the time the complete genome sequence of the soil bacterium *Pseudomonas putida* KT2440 was published in 2002 (Nelson et al., 2002) this bacterium was considered a potential agent for environmental bioremediation of industrial waste and a good colonizer of the rhizosphere. However, neither the annotation tools available at that time nor the scarcely available omics data—let alone metabolic modeling and other nowadays common systems biology approaches—allowed them to anticipate the astonishing capacities that are encoded in the genetic complement of this unique microorganism. In this work we have adopted a suite of state-of-the-art genomic analysis tools to revisit the functional and metabolic information encoded in the chromosomal sequence of strain KT2440. We identified 242 new protein-coding genes and re-annotated the functions of 1548 genes, which are linked to almost 4900 PubMed references. Catabolic pathways for 92 compounds (carbon, nitrogen and phosphorus sources) that could not be accommodated by the previously constructed metabolic models were also predicted. The resulting examination not only accounts for some of the known stress tolerance traits known in *P. putida* but also recognizes the capacity of this bacterium to perform difficult redox reactions, thereby multiplying its value as a platform microorganism for industrial biotechnology.

Introduction

*Pseudomonas putida* is a soil bacterium generally recognized as safe (GRAS). Belonging to a somewhat fuzzy clade of the Pseudomonadales (Palleroni, 1984), it has been used for decades as a model environmental organism with activity against aromatic pollutants. In 2002, in a successful transatlantic collaboration, scientists at The Institute for Genomic Research (The United States) and at four research centers in Germany deciphered and analyzed the genome of strain KT2440 (Nelson et al., 2002). This strain, which can be used to dispose of organic pollutants in the soil, promotes plant growth and fights plant
diseases (Regenhardt et al., 2002). Regenhardt et al. highlighted the complex and versatile metabolism that gives *P. putida* an important role not only in academic research on soil bacteria but also as an agent for environmental cleanup and other biotechnological uses. Yet, the genome analysis tools available at the time were able to extract only a small portion of the wealth of biological activities encoded in the chromosome of this bacterium.

In this work we set out to revisit the metabolic and physiological setup of this organism by re-analyzing the content of its genome using several approaches. We first re-sequenced the *P. putida* KT2440 wild-type strain, in parallel with that of a streamlined derivative as a control for possible evolution in laboratory settings (Leprince et al., 2012) and compared it to the original published sequence (Nelson et al., 2002). Combined with transcriptomic data analysis (Frank et al., 2011; Kim et al., 2013), a complete structural re-annotation of the KT2440 genome sequence led us to eliminate original erroneously predicted protein-coding genes, to correct disrupted genes and to identify potential new genes, some of which encode enzymatic activities. In a second step, we functionally re-annotated these genes based on recent progress in our knowledge of metabolic pathways (Silby et al., 2011; Wu et al., 2011). Thirdly, we used this re-annotation to reconcile in silico predictions from Genome-Scale constraint-based Metabolic Models (GSMMs) (Nogales et al., 2008; Puchalka et al., 2008; Sohn et al., 2010; Oberhardt et al., 2011) with metabolic phenotype data obtained with BIOLOG plates and transcriptomics experiments (Bochner et al., 2001; van Duuren et al., 2013). The updated annotation was then used to extend the GSMM iJP962 (Oberhardt et al., 2011) with newly curated Gene-Protein-Reaction (GPR) associations. Finally, this extended GSMM was evaluated for its ability to correctly predict positive/negative phenotypes of wild-type and mutant strains.

During the curation process, we surveyed metabolic pathways involved in coping with stressful environments and explored in some details the general context of aromatic compounds degradation. In biochemical terms, the synthesis of aromatic molecules is costly as it requires much energy and reducing power (Akashi and Gojobori, 2002). In genetic terms, the synthesis and degradation of aromatic compounds are costly too, because of the fairly large number of genes involved in these processes. In physico-chemical terms, the degradation of aromatics is problematic due to the fact that, because of their electronic set up, they tend to cross membranes when uncharged, often disrupting the lipid bilayer of the membrane and leaking in and out of the compartments where they should be confined (Saparov et al., 2006). Furthermore, because of this property, they frequently behave as proton-carriers that shunt the vectorial proton transport that would be used to build up ATP otherwise (i.e., chemical uncoupling). As a consequence, catabolic processes must be compartmentalized in a way that matches proton availability with the propensity of a protonated molecule to pass through the membrane (Kell and Oliver, 2014; de Lorenzo, 2015). This requires an efficient management of transport processes and control of the electrochemical potential of the cell as well as osmolarity. For this reason, we explored the metabolic capacity of *P. putida*, as indicated by the presence of relevant genes, in the context of control of osmolarity, control of proton availability and aromatic compounds degradation. Taken together, the novelties and metabolic updates presented in this work should contribute to the implementation of biocatalysis strategies using *P. putida* as a chassis for Synthetic Biology constructs.

The updated *P. putida* KT2440 genome sequence is deposited at the International Nucleotide Sequence Data Collaboration (identical accession number: AE015451, version 2). The re-annotated data can also be explored and downloaded using the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope). The curated genome-scale metabolic network is available at the MicroCyc repository (http://www.genoscope.cns.fr/agc/microcyc) and can be downloaded using the “Download Data” functionality of the “Search/Export” menu of the MicroScope platform. Finally, the updated metabolic model is available in the Supporting Information (SBML file format).

**Results and discussion**

**New features of the genome of strain KT2440**

*Pseudomonas putida* genome sequence and its structural re-annotation. The revised *P. putida* genome has 10 additional nucleotides compared with the earlier version (6 181 873 bp instead of 6 181 863 bp). We left out unmodified 140 regions in the re-sequenced genome (the largest being 5-kb long), encompassing regions annotated as rRNAs, tRNAs, transposons and group II intron-encoding sequences. The *P. putida* genome displays a GC content of 61.5%. The consensus sequence correction (see “Experimental Procedures” section) shows that the original sequence was of outstanding quality (Nelson et al., 2002). Indeed, among the 83 detected variations, 46 accounted for Single Nucleotide Polymorphisms (SNPs), 23 for short insertions and 14 for small deletions. It is known that strains kept in laboratories tend to evolve (Barrick et al., 2009). In order to substantiate the validity of our re-sequencing of the genome, we compared the regions of variation with the sequence of a streamlined mutant (Leprince et al., 2012): 95% of the variations were present in both sequences, showing that they were present at a very early stage and did not arise while handled in our laboratory. A significant part of the events (54) were found to affect 20 CoDing Sequences (CDSs; see Supporting Information Table S1). In most cases, insertion/deletion (InDel)
events restored the reading frame (i.e., either the new CDS is longer than the published one or two CDSs are fused in one gene, see Supporting Information Table S1). Only PP_0253 (encoding phosphoenolpyruvate carboxy kinase) and PP_5662 (encoding two fragments of a conserved gene of unknown function) remain pseudogenes (see below). Curiously, PP_5662 (21 SNPs) and PP_4302 (12 SNPs; encoding urea transporter) gather most of the detected SNPs (72%, either transitions or transversions). The re-annotated genome sequence (see “Experimental Procedures” section) comprises 5,592 CDSs plus 56 fragments of CDS (vs. 5,350 CDS stored in the last release of the NCBI GenBank file, NC_002947, or in the Pseudomonas.com database (Winsor et al., 2011)), 22 rRNA genes, and 75 tRNA genes. The non-coding regions account for 11.5% of the P. putida genome and contain 7.5% of repeated sequences. Only nine non-coding regions of more than 1 kb have been identified (Supporting Information Table S2). Among the annotated CDSs (complete genes and pseudogenes), we identified (i) common gene annotations between the original data and the AMIGene predictions: 5301 genes (94.8%), the original start codon positions of which were automatically kept; (ii) gene annotation unique to the original GenBank file: 116 genes, and (iii) gene annotation unique to the AMIGene prediction: 607 potential new CDSs. Following the manual curation process described in the “Experimental Procedures” section, 311 CDSs unique to the present version and 36 CDSs unique to the original annotation were kept. Moreover, 102 original CDSs were considered false positive predictions and removed from the final set of genes (Supporting Information Table S3). All of them would encode proteins of unknown function and 38 (37.2%) were found at a position where a new gene has been annotated, generally on the complementary strand [see Médiège et al. (2002) for a similar rationale used for annotation of Helicobacter pylori genes] (Supporting Information Tables S3 and S4). As shown in Supporting Information Table S4, the validity of most of the 311 newly annotated genes is supported by transcription expression profiling (Frank et al., 2011; Kim et al., 2013) and/or by sequence similarity with authentic genes: for example, PP_5706 encodes a protein involved in the Sec translocation complex (SecG subunit), and PP_5602 encodes the α subunit of the quinohemoprotein amine dehydrogenase (the peaA gene within the peaACB operon) which is known to be involved in the conversion of 2-phenylethylamine and 2-phenylethanol into phenylacetic acid in P. putida U (Arias et al., 2008). Indeed, 118 newly annotated genes among the 143 novel CDSs listed in Supporting Information Table S2 of the publication by (Frank et al., 2011), show up in Supporting Information Table S4 (the 25 missing ones correspond to predicted genes that were considered as false positives by our curation process). Forty-five new genes (14.5%) were assigned a gene product type and a biological process (Supporting Information Table S4) whereas the remaining genes (266) correspond to functions that remain to be identified.

Remaining pseudogenes. The re-sequencing process followed by expert curation of gene fragments and fusion/fission events using the MicroScope platform (Vallenet et al., 2013), identified a total of 71 CDSs as partial genes (14), pseudogenes (54 fragments of CDSs corresponding to 27 pseudogenes, and 1 CDS, PP_3752, which contains an internal stop codon) and one programmed frameshift (2 CDSs corresponding to the peptide chain release factor 2 gene, prfB). Partial genes were essentially grouped into classes of genes either encoding proteins containing Rh domains, transcriptional regulators (LysR family) or transposases (Supporting Information Table S5). Two of the 27 pseudogenes are of particular interest:

- The gene PP_0253 is split into two fragments that have 100% amino acid identity with fragments of the pckA gene encoding phosphoenolpyruvate carboxykinase (ATP dependent) in P. putida F1 (UniProt entry A5VX32). This enzyme is involved in gluconeogenesis, where it catalyzes the conversion of oxaloacetate (OAA) to phosphoenolpyruvate (PEP). The present UniProt functional annotation is supported by sequence similarity using the UniRule annotation procedure (The UniProt Consortium, 2014). Indeed, similarity with an experimentally validated phosphoenolpyruvate carboxykinase is found with the Staphylococcus aureus PckA protein (Q2G1W2, 45.4% amino-acid identity) (Scovill et al., 1996). The underlying reason for this loss of function in strain KT2440 is unknown, but we note that this enzyme is a key enzyme required for gluconeogenesis, under conditions where P. putida strains display a tight regulation of the balance between fluxes going from glucose to pyruvate and from succinate to pyruvate (La Rosa et al., 2015). In Escherichia coli O157:H7, PckA is important for maintaining the pathogenic bacteria in competition with the bulk of the microbiota (Bertin et al., 2014); inactivation of the gene may contribute to the GRAS phenotype of strain KT2440. Additionally, the enzyme is allosterically regulated by Ca²⁺ in other γ-proteobacteria (Sudom et al., 2003), and this feature might point at a particular role of the inactivation of this gene in the P. putida KT2440 niche.

- The two fragments of gene PP_1919 encode a protein similar to E. coli K-12 thymidylate kinase (Tmk protein; > 50% identity), a key enzyme for DNA synthesis. This protein catalyzes the phosphorylation of deoxythymidine monophosphate (dTMP) to deoxythymidine diphosphate (dTDP) in the presence of ATP and
Mg^{2+}. Tmk is essential for DNA synthesis and cell growth in E. coli (Reynes et al., 1996) and it would be expected to be essential in P. putida as well. Interestingly, in strain KT2440, but not in other sequenced P. putida strains, the tmk gene has been disrupted by the integration of a large genomic island of about 65 kb (the 3’-end of the first part of tmk is found at position 2 162 696 bp, while the 5’-end of the second part is found at position 2 227 487 bp). This region is obviously of phage origin (it contains genes for phage integrases, a transcriptional regulator of the Cro/cI family as well as site-specific recombinases), and harbors several clusters of metabolic genes (monooxygenases, dehydrogenases, etc.) together with a cluster of genes involved in arsenic resistance (PP_1927-PP_1930). Remarkably, PP_1964, the prophage gene located next to the truncated tmk gene, is likely to encode a deoxyribonucleotide monophosphate kinase (Mikoulinskaia et al., 2004), that could substitute for the missing essential tmk gene. Alternatively, the two halves of the tmk gene could be expressed separately and the resulting polypeptides reconstruct the enzyme activity through protein trans-complementation, a possibility currently under investigation.

**Functional re-annotation of protein-coding genes.** The outcome of the automatic functional annotation procedure was followed by manual curation of P. putida genes previously recorded as encoding unknown functions, while showing significant similarity with one of the protein and domain resources used in the platform (see “Experimental Procedures” section). Among those, 197 CDSs were reviewed (Supporting Information Table S6). Most of these proteins were labeled as (putative) enzymes (56%), (putative) transporters (20%) or (putative) regulators (9%). We further annotated 61 genes encoding proteins highly similar to proteins with functions experimentally demonstrated either in *Pseudomonas* species/genus or in other organisms. This is the case for genes involved in the catabolism of carnitine [PP_0301 to PP_0305; (Wargo and Hogan, 2009; Bastard et al., 2014)], in phenylethylamine degradation [PP_3459 and PP_3460; (Arias et al., 2008)], in gallate degradation [PP_2513, PP_2514 and PP_2515; (Nogales et al., 2011)], and in urate degradation [PP_4287; (Ramazzina et al., 2006)]. In order to provide accurate annotations, the global curation process was directed by the results of the growth phenotype data obtained in this work as well as extracted from experimentally based literature (see next section).

Overall, the function of 1548 genes has been manually re-annotated and linked to updated literature references (4837 PubMed references in the current annotation release). To provide a comprehensive reconstruction of the global metabolic map of *P. putida*, the utmost care was taken in the curation of associations between genes encoding enzymes and the biochemical reactions they catalyze. A total of 1485 CDSs has been associated to 1898 chemical reactions [1406 reactions from MetaCyc (Caspi et al., 2014) and 492 from Rhea (Morgat et al., 2015)] comprising a total of 3185 gene-reaction associations. In these associations, the role of 229 genes, displaying a high degree of similarity with their counterparts, was automatically annotated via transfer of the related *E. coli* K-12 reactions (see “Experimental Procedures” section). In the current update of the *P. putida* KT2440 genome annotation, about 21% of the protein-coding genes still remain of unknown function. A summary of the main *P. putida* KT2440 genome annotation updates in comparison with the original annotation can be found in Table 1.

An updated view of strain KT2440 metabolic capabilities through genome-scale modeling and phenotyping data. The updated genome annotation and corresponding functions were subsequently reviewed by computer simulations, assessing their contribution to the GSMM iJP962 (Oberhardt et al., 2011), which progresses toward a comprehensive model of the current knowledge of *P. putida* metabolism. First, we pinpointed knowledge gaps in the original GSMM by comparing its *in silico* growth predictions to the output of BIOLOG experiments on carbon, nitrogen and phosphorus sources. This comparison identified 108 compounds, the *in silico* growth prediction of which did not match the BIOLOG outcome. Furthermore, we added an extra set of 12 aromatic compounds that were not included in the BIOLOG assay but were known to serve as carbon source to *P. putida* (Jiménez et al., 2002; Kim et al., 2006). Eventually, the knowledge gap set comprised a total of 120 compounds, among which 43 carbon sources, 43 nitrogen sources, 31 phosphorus sources and 3 compounds that are both carbon and nitrogen sources (uridine, glycyl-glutamate and alanine-glycine) (Table 2).

Initial expansion of the iJP962 model with the automatically reconstructed metabolic network yielded a disappointing total of only 3 (all nitrogen sources) out of 120 compounds, the knowledge gap of which could be closed (i.e., a complete degradation route with reactions connecting the query compound to the central metabolism was present). This observation suggested that the metabolic model and the automatic genome re-annotation were missing catabolic pathways for the remaining 117 compounds. This prompted us to include the full set of 120 compounds as a starting point for a manual metabolic pathway curation process (further described in “Experimental Procedures” section). The outcome of this effort allowed us to identify catabolic pathways for 92 of these compounds (32/43 carbon, 28/43 nitrogen, 29/31 phosphorus and 3/3 carbon and nitrogen sources; see Table 2). Some of those metabolic routes, absent from public metabolic pathways...
Mechanisms of control of osmolarity

Living in polluted environments, *P. putida* needs to cope with highly variable concentrations of osmolytes. It must, therefore, build up a matching opportunity to control osmolarity by shuttling between synthesis, degradation and transport of osmolytes. This is reflected in its genome sequence by the concerted presence of genes involved in these biological processes.

Osmoregulation metabolism and transport of osmolytes.

Potassium glutamate is a major regulator of osmolarity in a large panel of organisms (Gralla and Vargas, 2006). The Kdp and Trk transport systems mediate osmoregulatory K⁺ uptake in a wide range of Bacteria and Archaea. In contrast to what was initially published with the sequence of the genome of *P. putida* KT2440 (Nelson et al., 2002), a complete Kdp system is present in this strain (the *kdpCBAF* operon; Supporting Information Table S8). It contains a functional high affinity P-type ATPase-K⁺ transporter encoded by the now functional *kdpB* gene, previously annotated as a pseudogene. Furthermore, we have identified and annotated a novel gene which encodes the small non-essential KdpF subunit (29 amino acids) that binds and stabilizes the whole protein complex (Gassel et al., 1999). Expression of this gene is dependent on a two-component regulatory system, encoded by *kdpD* (the sensor kinase component) and *kdpE* (the response regulator component), that activates the expression of the *kdpCBAF operon under conditions of severe K⁺ limitation or osmotic upshift (Ballal et al., 2007).

In terms of compatible solutes transport, strain KT2440 has a functional counterpart of the proline/betaine symporter (ProP), a multidrug efflux protein of the major facilitator superfamily (MFS) that mediates the uptake and accumulation of either one of these two osmoprotectants in *E. coli* K-12. ProP allows for adaptation to increasing osmotic pressure by acting as transporter and osmosensor (MacMillan et al., 1999). Exploration of the synteny conservation between *P. putida*, *P. aeruginosa* and *P. syringae* allowed us to

Table 1. Summary of the main *P. putida* KT2440 features annotation update in comparison with the original one.

|                              | New annotations | Original annotations |
|------------------------------|-----------------|---------------------|
| CDS                          | 5592            | 5350                |
| Unknown functions/hypothetical proteins | 1151 (*1)       | 1505                |
| Pseudogenes                  | 28 (*2)         | 9 (*3)              |
| Partial genes                | 14              | 61 (*4)             |
| Additional genes             | 311             |                     |
| False positive genes in original annotations | 102          |                     |
| rRNA genes                   | Total number    | 22                  |
| tRNA genes                   | Total number    | 75                  |
| EC number annotation         | CDS associated with an EC number | 1250 | 463          |
|                              | Total unique EC numbers | 902  | 360          |
|                              | Complete EC numbers | 811  | 360          |
|                              | Partial EC numbers | 91              | 0            |
| GPR associations             | Number of CDSs associated to reactions | 1485 | 0            |
|                              | Number of reactions | 1898 (*5) | 0         |
|                              | Total number of GPR associations | 3185 | 0           |
| PMID annotations             | Genes with associated PMID references | 1371 | 18          |
|                              | Number of different PMID references | 4837 | 1           |

(*1) 1040 conserved proteins of unknown function + 111 proteins of unknown function.

(*2) 28 pseudogenes made of 54 fragments of CDSs corresponding to 27 pseudogenes, and 1 CDS, *PP_3752*, which contains an internal stop codon.

(*3) 9 genes without/product annotation; in/note = "This region contains a pseudogene, one or more premature stops, and is not the result of a sequencing artifact.” following the sequencing and the manual curation processes these 9 pseudogenes have been re-annotated as functional.

(*4) 61 genes without/product annotation; in/note = "This region contains an authentic frame shift and is not the result of a sequencing artifact.” The sequence of 10 of these partial genes has been corrected after the re-sequencing process.

(*5) 1406 MetaCyc reactions (Caspì et al., 2014) + 492 Rhea reactions (Morgat et al., 2015).
identify additional transporters that may operate together to span the whole physiological range of osmolality and provide optimal uptake of glycine-betaine and choline osmoprotectant molecules from the environment (Fig. 1 and Supporting Information Table S8). As reported in experiments performed with \textit{P. aeruginosa} (Wargo, 2013), three transporters of the BBCT family (BetT-I, BetT-II and BetT-III) could transport glycine-betaine (BetT-II) and choline (BetT-I and BetT-III), and thus confer osmoprotection [as shown in \textit{P. syringae}, when they are expressed in a hyperosmotic environment (Chen and Beattie, 2008)]. Moreover, a complete choline-betaine-carnitine (CBC) ABC transport system is encoded in the \textit{cbcXWV} operon. The expression of the operon is induced by an AraC-family transcriptional activator (encoded by \textit{gbdR}) in response to glycine-betaine and dimethylglycine (Chen \textit{et al}., 2010; Wargo, 2013). In fact, three different periplasmic substrate-binding proteins in \textit{P. putida}, encoded by the \textit{cbcX}, \textit{caiX} and \textit{betX} genes (Fig. 1 and Supporting Information Table S8), show high specificity for choline, carnitine and betaine, respectively (Chen \textit{et al}., 2010). Finally, a small multidrug resistance (SMR) protein, homolog of the \textit{E. coli} K-12 EmrE protein, is also present in \textit{P. putida} \textit{KT2440}. It could be associated to choline and glycine-betaine export in response to intracellular levels of both osmoprotectants (Bay and Turner, 2012).

\textbf{Glycine-betaine degradation.} In addition to the annotation of choline and glycine-betaine transporter genes, the Table 2. Results of the integration of the updated catabolic pathways into the metabolic model \textit{iJP962}.

\begin{tabular}{|l|c|c|c|c|}
\hline
\textbf{Carbon Sources} & \textbf{\textit{iJP962}} & \textbf{\textit{iJP962}} & \textbf{\textit{iJP962}} & \textbf{\textit{iJP962}} \\
 & \textbf{+ Pre} & \textbf{+ Cur} & \textbf{+ Pre} & \textbf{+ Cur} \\
\hline
L-Alanyl-Glycine & L-Histidine & & & \\
Glycyl-L-Proline & Uracil & & & \\
Glycyl-L-Glutamic Acid & Xanthine & & & \\
\(\gamma\)-Hydroxy-Butyric Acid & Allantoin & & & \\
\(\alpha\)-Hydroxy Glutaric Acid-Y-Lactone & Gly-Met & & & \\
\(\alpha\)-Glucose & Met-Ala & & & \\
Butyric Acid & L-Cysteine & & & \\
Dihydroxyacetone & Al-Asp & & & \\
L-Pyrogulatamic Acid & Ala-Gln & & & \\
Uridine & Ala-Glu & & & \\
4-Hydroxy-L-Proline (\textit{trans}) & L-Alanyl-Glycine & & & \\
\(\alpha\)-Hydroxy-Butyric Acid & Ala-His & & & \\
D-Galacturonic Acid & Ala-Leu & & & \\
D-Gluronic Acid & Ala-Thr & & & \\
Quinic Acid & Gly-Asn & & & \\
\(\beta\)-Phenyethylamine & Gly-Gln & & & \\
Bromo-Succinic Acid & Glycyl-L-Glutamic Acid & & & \\
D,L-Carnitine & L-Pyrogulatamic acid & & & \\
D-Ribose & Cytidine & & & \\
D-Ribono-1,4-Lactone & Uridine & & & \\
L-Alaminamide & Inosine & & & \\
Methyl Pyruvate & Xanthosine & & & \\
*Gallate & Uric acid & & & \\
*Glycine Betaine & D-Serine & & & \\
*Choline & D-Valine & & & \\
*Sulfate choline & D,L-\(\alpha\)-Amino-N-Butyric acid & & & \\
*Ferulate & L-Amino-N-Valeric acid & & & \\
*Phenylacetate & L-Methionine & & & \\
*Vanil late & \(\beta\)-Phenyethylamine & & & \\
*Vanilline & D-Asparagine & & & \\
*Coniferyl alcohol & & & & \\
*p-Coumarate & & & & \\
*Caffeate & & & & \\
*Nicotinate & & & & \\
\hline
\end{tabular}

All 96 compounds that were part of the initial 120 knowledge gaps and for which a degradation pathway was ultimately identified are included. These include: 23 BIOLOG Carbon sources, 12 Literature-based Carbon sources (indicated by *), 31 BIOLOG Nitrogen sources and 29 BIOLOG Phosphorus sources. \textit{iJP962} was either expanded with the predicted reaction set (+ Pre), or with the curated reaction set (+ Cur). The colors represent no-growth (red), growth (green) or growth with the addition of an artificial transporter (orange).
Annotation of genes involved in the aerobic degradation of these compounds has also been updated. The betIBA operon (Fig. 1 and Supporting Information Table S8) encodes a choline-responsive transcriptional repressor (BetI), and two enzymes, a choline oxidase (BetA) and a betaine aldehyde dehydrogenase (BetB), responsible for the two-step conversion of choline to glycine-betaine (Rkenes et al., 1996; Velasco-García et al., 2006; Ziegler et al., 2010). As in P. aeruginosa, the genes encoding the choline transporter BetT1 and the betIBA operon are divergently transcribed in P. putida KT2440, allowing rapid transcriptional response to choline (Rkenes et al., 1996). Finally, comparative genomics allowed us to identify orthologs of the P. aeruginosa PAO1 genes involved in the three-steps demethylation of glycine-betaine to glycine, a metabolic pathway essential for growth with glycine-betaine as the sole carbon source (Wargo et al., 2008). This pathway includes a novel demethylase activity associated to the GbcAB enzyme complex that catalyzes the initial demethylation of glycine-betaine to dimethylglycine and formaldehyde. This operates via a process involving a dioxygenase and differs from the process mediated by the betaine-homocysteine S-methyltransferase present in other choline degraders like Sinorhizobium meliloti (Smith et al., 1988; Wargo et al., 2008). In P. putida KT2440, an heterodimeric flavin-linked oxidoreductase, encoded by the dgcA and dgcB genes (Supporting Information Table S8), catalyzes the second demethylation reaction of dimethylglycine to sarcosine, which is further demethylated to glycine in a reaction catalyzed by a heterotetrameric sarcosine oxidase complex encoded by the gene cluster soxDAG (Fig. 1).

**Trehalose-glycerol metabolism.** Due to its electroneutral nature and its role as a protein stabilizer, the disaccharide trehalose is a major osmoprotectant in bacterial cells (Kausik and Bhat, 2003; Ruhal et al., 2013). The P. putida genome re-annotation process revealed a complex metabolic scenario where trehalose could play a central role both in osmoregulation and in the metabolism of glycogen (Fig. 2). This differs from the metabolic profile present in most γ-Proteobacteria where this role is fulfilled by monosaccharide nucleoside diphosphates (Chandra et al., 2011). P. putida KT2440 lacks the ostAB genes encoding enzymes involved in the two-step trehalose biosynthesis.
pathway from UDP-glucose via a trehalose-6-phosphate intermediate (Kaasen et al., 1992). Rather, it displays two alternative pathways for trehalose biosynthesis (Fig. 2A and Supporting Information Table S8). The first one involves the PP_4053 protein (previously annotated as a generic glycosyl hydrolase) that is highly similar to the malto-oligosyl trehalose synthase (TreY) from other P.putida strains. Together with the malto-oligosyl trehalose hydrolase (encoded by the treZ gene), TreY catalyzes the biosynthesis of trehalose from glycogen (Kobayashi et al., 1996; De Smet et al., 2000). The second pathway is associated to two different trehalose synthases (coded by treSA and treSB genes) that catalyze the reversible single-step conversion of maltose to trehalose (Lee et al., 2005; Chandra et al., 2011; Ruhal et al., 2013). These enzymes belong to two evolutionary distinct lineages. The corresponding genes do not display sequence similarity and are involved in different genomic and metabolic contexts. The treSB gene (PP_4059, Supporting Information Table S8) encodes a fused protein (a trehalose synthase belonging to a family widely distributed across different bacterial lineages and a maltokinase) and is clustered with genes encoding the glycogen branching enzyme GlgB, and the α-1,4-glucan-maltose-1-phosphate maltotransferase GlgE (Fig. 2B). These genes form an operon for a novel glycogen biosynthesis pathway similar to the variant recently discovered in Mycobacteria (Fig. 2C), that uses α-maltose-1-phosphate instead of UDP-glucose-6-phosphate as the building block to extend glucan chains (Elbein et al., 2010; Chandra et al., 2011; Ruhal et al., 2013). By contrast, the trehalose synthase encoded by the treSA gene (PP_2918; Supporting Information Table S8) belongs to a small family of highly active trehalose synthases. It has been biochemically characterized in P.stutzeri CJ38 as a biocatalyst of biotechnological interest for the production of trehalose (Lee et al., 2005). We propose that this second P.putida trehalose synthase may have a role in the control of osmolarity.

Control of the proton gradient

Pseudomonas putida KT2440 is an obligate aerobe that uses the EDEMP cycle (composed by activities from the Entner–Doudoroff, the incomplete Embden–Meyerhof–
Parnas, and the pentose phosphate pathway) to process glucose (Nikel et al., 2015). Furthermore, it lacks the glucose-specific phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) that usually fuels in the Embden-Meyerhof-Parnas pathway in other bacteria, such as E. coli. Yet, apart from sugars its growth environment provides a considerable number of compounds that may enter its metabolism at various points. This in turn requires the presence of a large number of transport systems, as illustrated by the coding capacity of its genome. The processes encompassing oxygen availability and utilization, carbon catabolism and transport suggest that a considerable amount of protons are involved: they could be channeled during respiration to form ATP and in proton/metabolite co-transport activities. P. putida KT2440 possesses counterparts of the cytochrome bo oxidase and the cytochrome bc1 oxidase found in many bacteria. However, it does not have a counterpart of E. coli cytochrome bd-II oxidase (AppCD). The activity of cytochrome oxidases contributes to build up a proton motive force (Bettenbrock et al., 2014). The proton gradient thereby generated is challenged when the pH of the environment varies. We thus wanted to explore the way the bacterium maintains proton homeostasis through critical examination of its genome sequence. P. putida is a neutrophilic organism and harbors a standard version of most of the general processes involving protons (ATP synthase, assembly of flagellar motor, NADH/NADPH balance, etc.). It differs however from other classes of γ-Proteobacteria such as Enterobacteria in the way it manages the acid resistance response and the transport of protons.

**Acid resistance response.** The acid stress response involves many different processes in species having a periplasm (Lund et al., 2014), where some enzymes may have an acidic optimum pH for activity [e.g., AppA in E. coli (Golovan et al., 2000), a gene not found in P. putida]. The results of the functional re-annotation shows that P. putida KT2440 has orthologs of the E. coli K-12 genes encoding the alternative sigma factor RpoS and the cAMP receptor protein Crp, that constitute the glucose-repressed Acid Resistance system AR1 allowing cell survival at pH \( \approx 2.5 \) (Foster, 2004; Milanesio et al., 2011) (Supporting Information Table S9). In P. putida RpoS restores the acid resistance phenotype missing in rpoS-deficient E. coli mutants. Yet, it seems that in P. putida the role of RpoS is mainly associated to adaptation to carbon starvation conditions (Ramos-Gonzalez and Molin, 1998). RpoS and Crp are global regulators which control expression of multiple genes (regulons) under conditions when global resource allocation needs to be modified as the environment changes (Hui et al., 2015). However, the regulatory network of both Crp and RpoS activities is noticeably different in P. putida when compared with that of E. coli, in line with the widely different niches of the organisms (Venturi, 2003; Milanesio et al., 2011).

Expression profiling studies in the KT2440 strain using different carbon sources revealed a strong expression of rpoS in cells growing with glycine and fructose as carbon sources (Frank et al., 2011; Kim et al., 2013). A further difference can be pointed out: P. putida KT2440 has neither orthologs of the E. coli decarboxylase-antiporter systems AR2 [glutamate-decarboxylase isozymes GadA, GadB and 4-aminobutanoate (GABA)-glutamate antipporter GadC], nor of AR3 (degradative arginine-decarboxylase AdiA and agmatine-arginine antipporter AdiC). As far as the Acid Resistance system 4 (AR4) is concerned, the PP_4140 gene, previously annotated as a pseudogene, was now found to be complete. It is similar to the E. coli lysine decarboxylase (Idcc gene, encoding a constitutive form of the lysine decarboxylase). However, there is no signal of neighbor cadaverine-lysine antipporter CadB characteristic of the AR4 system (Foster, 2004). Overall P. putida lacks most of the acid stress response present in Enterobacteria. This may contribute to its recognized lack of pathogenicity, but it needs to be taken into account when P. putida is used for biocatalysis in a reactor as well when the organism is used for in situ or ex situ bioremediation of polluted environments.

Otherwise, P. putida KT2440 has functional alternative pathways for the degradation of both L-arginine and GABA, which involve enzymatic activities induced in high pH conditions in E. coli. The P. putida annotated homologous genes are listed in Supporting Information Table S9.

**Transport of protons**

Protons are involved in many transport systems, including vectorial transport for ATP synthesis, as well as in the mechanical rotation of flagella. P. putida KT2440 has two counterparts of the Na\(^{+}\)/H\(^+\) antipporter NhaA (Supporting Information Table S9), the best-understood antipporter which helps maintain the internal pH, protecting cells from excess sodium at high pH (Dover and Padan, 2001; Stanick et al., 2002). This species also harbors a putative multidrug efflux protein MdfA that extends the pH tolerance range up to pH = 10 in E. coli, taking over when NhaA is deleted (Lewinson et al., 2004). However, we did not identify a homolog to the positive regulator NhaR, which controls NhaA activity during exponential growth (Rahman-Manor et al., 1992; Carmel et al., 1997). This may be compensated for by a possible activity under the control of the functional RpoS sigma factor together with genes of the RpoS regulon also involved in pH homeostasis in the stationary growth phase (Dover and Padan, 2001). The P. putida KT2440 genome also harbors a second pH-independent Na\(^{+}\)/H\(^+\) antipporter, NhaB (Pinner et al., 1993; Padan et al., 2005), as well as five proton-sodium antiporters of the monovalent cation:proton antipporter (CPA) families CPA1 and CPA2 (nhaB and nhaP genes;
Supporting Information Table S9). Three additional glutathione-gated K\(^+\) efflux systems (kef genes) of the CPA2 family have also been found. They are likely to be important for coping with mechanical stress induced by the considerable variations of metabolites charges and concentrations associated with \(P\). \textit{putida} metabolism in a chemically polluted environment. In \textit{B. subtilis} the essential operon \textit{mrpABCDEFG} encodes a transport system of the CPA3 family, which provides Na\(^+\)/H\(^+\) antiporter activity and functions in resistance toward several different compounds and pH homeostasis (Brett \textit{et al.}, 2005; Kajiyama \textit{et al.}, 2007). As in bacteria from a great many other clades (\textit{e.g.}, in \textit{Bdellovibrio bacteriovorus}, \textit{Bordetella pertussis}, \textit{Deinococcus radiodurans} and \textit{Mycobacterium smegmatis}), but not in \textit{E. coli}, \textit{P. putida} KT2440 has a complete operon counterpart of \textit{phaABCDFEG} (Supporting Information Table S9). Given the role of this system in pH adaptation and cholate resistance, we propose to rename this operon \textit{mrp} for “multiple resistance and pH adaptation locus.” This also allows to distinguish the \textit{P. putida} Na\(^+\)/H\(^+\) pumping function from the biosynthesis of polyhydroxalkanoates (pha genes). This system is widely present in Pseudomonadales, where its organization differs slightly from that of Firmicutes: MrpA counterpart is fused to MrpB (MrpAB protein). Moreover, 22 additional Major Facilitator Superfamily transporters (MFS) were identified in the \textit{P. putida} KT2440 proteome; they could contribute to pH homeostasis through additional Na\(^+\)/H\(^+\) or K\(^+\)/H\(^+\) antiporter activities, as it is the case for the multidrug efflux protein MdfA in \textit{E. coli} or the tetracycline resistance protein TetL in \textit{B. subtilis} (Padan \textit{et al.}, 2005).

Finally, the annotation of five genes encoding periplasmic and outer-membrane proteins associated to pH homeostasis has been updated in \textit{P. putida} KT2440 (Supporting Information Table S9). These genes include the extreme base-induced membrane-bound redox modulator Aix (Stancik \textit{et al.}, 2002), as well as the peptidyl-prolyl cis-trans isomerase SurA, which is necessary for proper folding of outer membrane proteins and whose inactivation is lethal in stationary phase under elevated pH conditions (Tormo \textit{et al.}, 1990; Foster, 1999).

Degradation of carbon-skeleton aromatics

In addition to the outcome of BIOLOG experiments, much experimental evidence has been reported since the first publication of the genome sequence of \textit{P. putida} KT2440. This had impacted the annotation of genes involved in the degradation of aromatic compounds (Dos Santos \textit{et al.}, 2004; Nogales \textit{et al.}, 2005; Kim \textit{et al.}, 2006; Wu \textit{et al.}, 2011). The present upgrade includes genes involved in the central aromatic compounds degradation pathways as well as a variety of connected pathways. A summary of the new vision of the aromatic catabolism of strain KT2440 is represented in Fig. 3, and detailed in the Supporting Information Results file (see Supporting Information Table S10 for a complete description).

We propose a candidate gene for the orphan enzyme \textit{i.e.}, a defined enzyme without assigned sequence) responsible for the first redox step of the two-step degradation of coniferyl alcohol to ferulate (Jiménez \textit{et al.}, 2002; Nogales \textit{et al.}, 2008). The \textit{PP_2426} gene, corresponding to the alcohol dehydrogenase activity CalA (EC 1.1.1.194; Supporting Information Table S10), is likely to encode a coniferyl dehydrogenase, with somewhat promiscuous activity. This candidate gene shows significant similarity with cinnamyl-alcohol dehydrogenases of plant origin (about 50% amino acid identity over the whole protein length), which are also able to act on coniferyl alcohol (see IUBMB annotation, EC 1.1.1.195). The proposed coniferyl dehydrogenase CalA (\textit{PP_2426}) would work together with coniferyl aldehyde dehydrogenase CalB (\textit{PP_5120}, EC 1.2.1.68) (Overhage \textit{et al.}, 1999; Jiménez \textit{et al.}, 2002). However, an experimental validation is necessary to substantiate this prediction.

Degradation of nucleotides and other heterocyclic aromatics

The positive redox phenotypes observed in BIOLOG experiments using uracil and thymine as nitrogen sources led us to re-annotate a gene cluster which contains all the genes involved in the reductive pathway of pyrimidine nucleotides (West, 2001; Schnackerz and Dobritzsch, 2008) (Supporting Information Table S10). This pathway starts with the reduction of uracil and thymine to the corresponding 5,6-dehydro-derivatives by a type II NADPH-dependent dihydroxypyrimidine dehydrogenase (DPD) enzyme complex PydXA (Osterman, 2006; Hidese \textit{et al.}, 2011). The dehydroxypyrimidines are subsequently hydrolyzed by a bifunctional D-hydantoinase/dihydropyrimidinase \textit{pydB} gene and a β-ureidopropionase \textit{hyuC} gene) into β-alanine and 3-amino-isobutyrate respectively \textit{(West, 2001; Schnackerz and Dobritzsch, 2008). The \textit{PP_4036} gene \textit{pydB}, was originally annotated as a pseudogene (sequencing error), but it is likely to be fully functional as...
it encodes a protein highly similar to the experimentally characterized d-hydrantoinase/dihydropyrimidinase from P. putida (Arthrobacter capsulatus) (Chien et al., 1998). The gene cluster also encodes a permease commonly present in β- and γ-Proteobacteria (pydP gene), as well as a transcriptional regulator (the PP_4039 gene is similar to the E. coli rutR gene) (Supporting Information Table S10).

In the same way, the positive redox phenotype observed in BIOLOG experiments when using xanthine, urate or allantoin as nitrogen sources, allowed us to upgrade the annotation of a gene cluster involved in the transport and degradation of purine nucleotides (Supporting Information Table S10). It includes the xanthine dehydrogenase enzyme complex XdhABC that catalyzes the NAD\(^+\)-dependent oxidation of hypoxanthine and xanthine to urate (Parschat et al., 2001), and two of the three enzymes involved in the degradation of urate to S-allantoin (Ramazzina et al., 2006): the hydroxysisourate hydroxylase (PucM) and the 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) decarboxylase (PucL). These proteins belong to two chromosomal clusters and share homologies with eukaryotic and prokaryotic proteins (COG2351 and COG3195, respectively). They also display similar co-evolution phylogenetic profiles (Engelen et al., 2012; Valleynet et al., 2013). This suggests a common evolutionary gain and loss history, as illustrated in other organisms harboring this pathway (Ramazzina et al., 2006). Furthermore, S-allantoin can be degraded in four steps to glyoxylate via S-ureidoglycine as an intermediate, releasing ammonia and urea. The first step involves a novel metal-independent allantoinase encoded by the puuE gene that differs from the E. coli K-12 allantoinase (allB gene) (Ramazzina et al., 2008).
Finally, the annotation of the nic gene cluster (nicPTFEDCXRBAS), responsible for the aerobic degradation of nicotinate to fumarate, has also been updated. It allows *P. putida* KT2440 to grow with nicotinate as both nitrogen and carbon source (Jiménez et al., 2008).

**Toward an extended view of the KT2440 metabolic model**

The updated genome annotation provided us with a list of functions, for example, chemical conversions, that were not previously identified in *P. putida*. However, the effect of an individual function on systems-wide behavior is not straightforward. For example, a candidate degradation pathway can eventually be deemed nonfunctional if its byproducts cannot be further processed. We decided to assess the full impact of the updated annotation by complementing an existing genome-scale metabolic model with the new reactions. This allowed us to check whether the identified enzymatic conversions could truly function in the context of the former knowledge of *P. putida* metabolism, and to pinpoint additional knowledge gaps to be addressed in future studies. Specifically, for 96 out of the 120 defined knowledge gaps, we identified a probable degradation pathway during the targeted manual annotation process. Together, these pathways comprised a total of 253 reactions, 234 of which have been assigned to one or more genes and integrated into MicroScope. Moreover, 43 new ChEBI compounds and 73 new RHEA reactions were created during this curation process.

To assess whether these reactions indeed coped with the knowledge gaps, we expanded the iJP962 metabolic model with the degradation pathways and mimicked *in silico* the BIOLOG experiment (see “Experimental Procedure” section). Surprisingly, this expansion led to an *in silico* positive phenotype for only 20 compounds (out of 96). However, it is important to recall that the BIOLOG setup does not measure growth per se but the integrated activity of redox networks (Bochner, 1989). This relatively small improvement prompted us to inspect the remaining cases in more detail (Supporting Information Table S12). A major issue turned out to be the difficulty in identifying transport proteins for specific compounds; our list of curated reactions only contained 23 transporters. To further test the existence of degradation pathways, we complemented the GSSM with ad hoc transport reactions that behaved as passive diffusion reactions. This improved the outcome of the model, as 72 out of 96 degradation pathways were now functional. Interestingly, even in the original model the addition of ad hoc transporters resolved 10 of the knowledge gaps, indicating that for some compounds the lack of a transport reaction was the only functional step preventing *in silico* growth. This procedure also led to *in silico* positive growth phenotypes for 14 compounds with a negative BIOLOG phenotype (Supporting Information Table S12), demonstrating that it is essential to get experimental evidence for transport systems. Although such results require future *in vitro* confirmation, they suggest that the range of suitable substrates for *P. putida* may be increased with the sole identification of the corresponding transporter proteins. This observation highlights an essential area for future research that will lead to improve GSSMs.

Still, successful *in silico* metabolite degradation was yet to be achieved for 24 out of the 96 compounds with identified degradation pathways. The underlying causes of these remaining knowledge gaps may be roughly divided into four categories (Supporting Information Table S13):

i. **Level of detail.** The degradation pathways for seven compounds involved ill-defined metabolite classes, such as “NADPORNOP,” and “Oxidized-cytochromes.” Where possible, we replaced these with specific instances of these classes, such as NAD and ferricytochrome.

ii. **Byproduct accumulation.** The degradation pathways for six compounds resulted in by-products that the *in silico* cell was unable to dispose of. In particular, five degradation pathways led to an accumulation of sulfur-containing compounds. We complemented the model with sulfate, hydrogen sulfide and sulfate exporters, which allowed successful degradation of 4/5, 1/5 and 5/5 compounds. We show below that *P. putida* KT2440 has 11 candidate tauE genes, which may encode a sulfite exporter. The sulfite export reaction and the 11 corresponding genes were thus added to the curated reaction list.

iii. **Reaction reversibility.** The degradation of one compound, D-glucosamine-6-phosphate, was hampered by a reaction that was irreversible in the model, but reversible according to external sources such as MetaCyc (Caspi et al., 2014) and Brenda (Chang et al., 2015). We adjusted the reaction accordingly.

iv. **Open issues.** Ten out of the degradation pathways led to the production of dead-end metabolites in the model. Dead-end metabolites are metabolites that can either only be produced, or only consumed in the model. We were unable to link possible degradation pathways for these compounds to *P. putida* genes. These non-functioning degradation pathways and the corresponding metabolites highlight a remaining knowledge gap in *P. putida* metabolism to be addressed in future studies.

In addition, we assessed how the expanded model performs in a broader *in silico* growth analysis including both wild-type and mutant growth predictions. We distinguished between predictions for wild-type growth and for mutant growth because these reflect different qualities of a
Comparison to other Pseudomonas putida strains

Udaondo and co-workers have recently reported a comparative analysis of the genomes of nine *P. putida* strains aimed at determining the core collection of genes that give identity to this species (Udaondo et al., 2015). Although the number of strains examined is somewhat limited, the results revealed the lack of pathogenic traits (e.g., exotoxins and type III secretion systems are absent in all cases) and the centrality of the Entner–Doudoroff pathway as the key route for consumption of carbohydrates. Such a core genome [paleome (Acevedo-Rocha et al., 2013; Yang et al., 2015)] of *P. putida* consisted of approximately 3380 genes, a good share of which encoded transporters, both for nutrients and for electrons, which seemingly enable aerobic metabolism under different oxygen regimes. Other genes of the core set determined the pentoses phosphate cycle, arginine and proline metabolism, and different routes for degradation of aromatic chemicals. Amino acid metabolism (synthesis and degradation) was very conserved as well and encoded in each case complete set of transporters, enzymes and regulators. Flagellar biosynthesis and genes for biofilm formation belong to the *P. putida* core genome as well.

Despite a large number of differences between strains, the wealth of information on strain KT2440 discussed above makes this specimen the reference for the whole group. Many of the general traits discussed above that make special strain KT2440 can be properly extended to other members of the *P. putida* group (Nikel et al., 2014), with the caveat that the *P. putida* group is somewhat fuzzy, strain 2440 lying slightly distant from the reference type strain DSM291 (Ye et al., 2014).

### Table 3. Model evaluation. *JP962* as well as the extensions based on predicted (+Pre) and curated (+Cur) degradation pathways were tested in terms of phenotype predictions (growth/no-growth).

|                | *JP962* | *JP962* + Pre | *JP962* + Cur |
|----------------|---------|---------------|---------------|
| Metabolites    | 980     | 1375          | 1122          |
| Reactions      | 1066    | 1533          | 1256          |
| Genes          | 949     | 1203          | 1053          |
| **Wild-type predictions** | | | |
| Specificity    | 0.90    | 0.86          | 0.88          |
| Sensitivity    | 0.42    | 0.55          | 0.75          |
| Accuracy       | 0.59    | 0.66          | 0.79          |
| **Mutant predictions** | | | |
| Coverage       | 0.70    | 0.68          | 0.70          |
| Specificity    | 0.74    | 0.56          | 0.72          |
| Sensitivity    | 0.72    | 0.71          | 0.80          |
| Accuracy       | 0.73    | 0.60          | 0.75          |

We used both wild-type and mutant growth data (Puchalka et al., 2008; Molina-Henares et al., 2010). The experimental mutant data comprised gene knockout data in defined media as well as experimentally verified auxotrophies.
annotations as means to provide a standard for use of this organism as a versatile chassis for both fundamental and biotechnological endeavors. Over the last few years, *P. putida* strains have been increasingly recognized for their potential to host bioreactions that other model bacteria fail to execute (e.g., strongly oxidative biotransformations). An attractive trait of strain KT2440 that makes it adequate for such applications is the fact that this bacterium harbors a large number of metabolic and stress-endurance properties optimal for biotechnological needs. In our present study, we further highlight the potential of *P. putida* for biotransformations and biodegradation by disclosing mechanisms controlling osmolarity and pH homeostasis. While resequencing per se only provided marginal improvement in the sequence, the update of the chassis allowed us to propose a consistent picture of *P. putida* metabolism. Coupled with experimental data using the BIOLOG setup this allowed us to improve considerably the outcome of a systems biology approach where a model metabolism of the organism could be matched with experimental data. The present state of affairs demonstrates that while there remain some knowledge gaps in the *P. putida* metabolism, we now have a clear picture of its overall functioning. Our approach pinpointed a specific deficiency in our knowledge: we need to considerably improve explicit identification of transport systems. This should be a major task for the immediate future of studies with the *P. putida* chassis, but also for other chassis as well.

In this respect, the present update of the genome sequence of *P. putida* and its annotation emphasized considerable differences with the ubiquitous model used as a chassis in many studies, *E. coli* K-12. Indeed, Enterobacteria and related bacteria differ considerably from Pseudomonadales, and *P. putida* may be an excellent reference model of this clade. Beside metabolic differences that have been outlined in the present article, the way DNA is handled is quite different in these clades, and this may be of importance for studies involving DNA constructs meant to provide novel metabolic engineering approaches. An example of this are the different contingents of DNA polymerase III proteins in different species. In *P. putida* one finds four different DNA polymerase III proteins, three variants of DnaE (DnaE1, DnaE2 and DnaE3) and a second type, PolC (Timinskas et al., 2014). Organisms such as *B. subtilis* combine DnaE1 and PolC (Engelen et al., 2012). By contrast, *E. coli* has only DnaE1. A second DnaE variant appears as a heterologous subunit of the enzyme when the length of the genome sequence increases. Furthermore, the presence of DnaE2 together with DnaE1 is linked to bacteria featuring large GC-rich genomes and living in aerobic environments (Timinskas et al., 2014), as in the case of *P. putida* (*dnaEA: PP_1606* and *dnaEB: PP_3119*). Analysis of the co-evolution of the genes that are present in parallel with DnaE2 will certainly help identification of functions that are highly relevant both to the ecological niche of the organism and to its use as a cell factory.

**Experimental procedures**

*Pseudomonas putida* sequencing

The genome sequences of *P. putida* KT2440 DSM 615 and of a mutant strain TEC1 401-1, were obtained using Illumina sequencing technology. The wild-type strain is the one sequenced in 2002, coming from the same original glycerol stock deposited in the DSMZ collection, and the mutant strain was generated by experimental genome reduction over strain KT2440 by the group of Vitor Martins dos Santos in Wageningen University [Microme WP3, (Leprince et al., 2012)]. Paired-ends libraries were prepared with fragment size comprised between 300 and 600 bp and sequenced on HiSeq2000 (100 nt length). A total of 8 786 896 reads were produced for the *P. putida* KT2440 wild-type strain and 11 021 169 reads for the mutant, leading to 1.6 and 2.1 Gb, respectively. Sequence reads were processed to remove low-quality reads and mapped over the *P. putida* KT2440 reference genome sequence.

SNPs/InDels detection strategy

High Throughput Sequencing (HTS) data were analyzed using the PALOMA pipeline (Crueviller S., unpublished) implemented in the Microscope platform (Vallenet et al., 2013). The current pipeline is a "Master" shell script that launches the various modules of the analysis (i.e., a collection of in-house software written in C) and controls for all tasks having been completed without errors. In a first step, the HTS data quality was assessed by including options like reads trimming or merging/split paired-end/mate-paired reads. In a second step, reads were mapped onto the original sequence of *P. putida* str. KT2440 (Accession Number NC_002947; AE015451.1) using the SSAHA2 package (Ning et al., 2001). Unique matches having an alignment score equal to at least half of their length were retained as seeds for full Smith–Waterman realignment (Smith and Waterman, 1981) keeping at both sides a region of the reference genome extended by five nucleotides. All computed alignments were then screened for discrepancies between read and reference sequences and in fine, a score based on coverage, allele frequency, quality of bases and strand bias was computed for each detected event to assess its relevance. The results generated are available at the MicroScope platform (http://www.genoscope.cns.fr/agc/microscope).

Consensus sequence correction

To correct the original sequence of *P. putida* KT2440, the PALOMA pipeline was run with stringent parameters for the "SNP calling" step (allelic frequency set to 0.8 with at least 10 reads mapping the position, a balance of forward reads to reverse reads set to 0.33). This analysis led to a relatively small amount of variations compared with the original one, showing

© 2016 The Authors. *Environmental Microbiology* published by Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology* 18, 3403–3424
that the 2002 sequence was of excellent quality (Nelson et al., 2002). An automated process was subsequently implemented to generate a new version of the sequence of the *P. putida* strain KT2440 genome using both the original sequence and the list of detected variations as inputs. During the process, uncovered areas of the reference genome were reported as well, corresponding either to repeats (discarded by default during the reads mapping step) or potentially large deletions in the re-sequenced genome.

**RNA-Seq Analysis**

The complete transcriptome high-throughput sequencing data published in Kim et al. (2013) was retrieved from the GEO database ([Barrett et al., 2013]; accession no. GSE42491). Data were then analyzed in the MicroScope platform with the workflow TAMARA (Vallenet et al., 2013). The current pipeline is a “Master” shell script that launches the various parts of the analysis (i.e., a collection of Shell/Perl/R scripts) and checks that all tasks are completed without error. Reads preprocessing and mapping steps are performed in the same way as the PALOMA pipeline (see “SNPs/InDels detection strategy” section for details). After reads were mapped on the newly annotated *P. putida* strain KT2440 genome, we minimized the false positive discovery rate using SAMtools [v0.1.8: (Li et al., 2009)] to extract reliable alignments from SAM-formatted files. The number of reads matching each genomic object of the reference genome was then calculated with the Bioconductor-GenomicFeatures package (Lawrence et al., 2013). When reads matched several genomic objects, the count number was weighted so as to keep the total number of reads constant. Finally, the Bioconductor-DESeq package (Anders and Huber, 2010) was used with default parameters to normalize raw count data based on negative binomial distribution and to determine whether expression levels differed between conditions.

**Structural re-annotation of the Pseudomonas putida genome**

The corrected genome sequence was subsequently processed by the MicroScope pipeline for complete structural and functional annotation (Vallenet et al., 2013). Gene prediction was performed using the AMIGene software (Bocs et al., 2003) and the microbial gene finding program Prodigal (Hyatt et al., 2010) known for its capability to locate the translation initiation site with great accuracy. The predicted genes were compared with those listed in the original annotation (AE015451, version: 05-MAR-2010). Manual curation was performed on the two sets of unique genes (see “Results” section) by taking into account transcriptomic information from (Frank et al., 2011) and (Kim et al., 2013) experiments, as well as conservation of sequence similarity and genomic context with homologs in other genomes. Predicted small CDSs having a coding prediction value inferior to 0.3 and which are orphans in terms of sequence similarity and not involved in a synteny group were discarded, unless they showed a signal with one of the transcriptomic experiments. A total of 80 unique GenBank genes, and of 296 unique AMIGene CDSs were considered false positive predictions and discarded from the final annotations (artifact status). These genes are kept in our database as “obsolete” genomic objects, but they are removed from the *P. putida* KT2440 genome annotation deposited at the International Nucleotide Sequence Data Collaboration (identical accession number: AE015451, version 2). The 309 unique AMIGene predictions considered as newly predicted *P. putida* genes are numbered starting from the last original annotation (**PP_5420**) (i.e., **PP_5521**, Supporting Information Table S4).

The RNAmmer (Lagesen et al., 2007) and tRNAscan-SE (Lowe and Eddy, 1997) programs were used to predict rRNA and tRNA-encoding genes, respectively, whereas other RNA structures like small RNAs and riboswitches were identified using the RFAM database (Burge et al., 2013) (n = 65) and from publications (n = 3) (Frank et al., 2011). Finally, intrachromosomal repeats were detected using the method described by (Achaz et al., 2000).

**Functional automatic annotation**

The predicted/annotated genes were subjected to sequence similarity searches using the gapped blastP algorithm against the UniProtKB protein sequence knowledgebase (The UniProt Consortium, 2014) and several protein family resources: COG (Galperin et al., 2015), HAMAP (Pedruzzi et al., 2015) and FIGfam (Meyer et al., 2009). They were also processed using the InterProScan software to predict potential sequence motifs, patterns and protein family assignments compiled in InterPro (Mitchell et al., 2015). In addition, genes encoding enzymes were also classified using the PRIAM profiles (Claudel-Renard et al., 2003). In terms of predicted structural features, α-helical transmembrane regions were searched with the TMHMM program (Krogh et al., 2001) and signal peptides with SignalP (Petersen et al., 2011). To predict probable subcellular localization of the annotated protein in the cell, PSORTb predictions were also carried out (Yu et al., 2011).

Using the MicroScope platform, *E. coli* K-12 expert annotation is already an ongoing process since the work described in (Touchon et al., 2009), with a main focus in the curation of GPR associations coming from EcoCyc (Keseler et al., 2013) and literature data. Then, in order to (re)assign functions to each *P. putida* KT2440 annotated genes, bi-directional best-hit (BBH) between *P. putida* KT2440 and *E. coli* K-12 genes were first identified by BLASTP, and annotation transfer from *E. coli* K-12 to *P. putida* KT2440 genes was carried out based on this BBH relationships and the following similarity thresholds: 50% identity on 80% of the length of the longest protein, or 40% identity on 80% of the length of the longest protein in case of shared genomic context or FIGfam protein families assignments (Meyer et al., 2009). *P. putida* KT2440 annotation transfer includes the transfer of these GPR associations from *E. coli* K-12 counterpart, a feature that improves the subsequent genome-scale metabolic network reconstruction (see below). A total of 706 genes were re-annotated using this process. *P. putida* genes escaping the *E. coli* K-12 functional annotation transfer were annotated following the standard MicroScope procedure (Vallenet et al., 2013). Finally, during the curation process of gene function, chosen gene names conform to the nomenclature conventions derived from (Demerec et al., 1966).
Automatic genome-scale metabolic network reconstruction

The metabolic network of *P. putida* KT2440 was reconstructed from the re-annotated genome sequence stored in PKGD B using the MicroScope automatic reconstruction pipeline, which is based on the BioCyc pathway reconstruction software (Karp et al., 2011). Pathway Tools uses the set of genome annotations as input data to automatically project the set of reference metabolic pathways stored in MetaCyc database (Caspi et al., 2014), generating a specific Pathway Genome Database (PGDB) in a two-step process: first, the Reactome projection step, where associations between genes and metabolic reactions are inferred from gene annotations, and the Pathway projection step where reference pathways are projected based on these gene-reaction associations (see Karp et al., 2011 for further details on the algorithm). The Reactome projection step in MicroScope is enhanced by implementing an export procedure from MicroScope PKGD B to Pathway Tools input format that directly associates the genes to MetaCyc reaction identifiers coming from manual validation by MicroScope curators or from automatic reaction transfer from reference organisms (Vallenet et al., 2013). This allowed minimization of the over-prediction or missing of relevant enzymatic reactions resulting from inaccurate or unclear textual annotations. It should be noticed that, GPR associations rules of the PathoLogic algorithm does not distinguish between AND and OR statements, that is, isozymes and protein complexes are equally considered (using a OR relationship). The reconstructed genome scale metabolic network of *P. putida* KT2440 is included in the MicroCyc repository available at http://www.genoscope.cns.fr/agc/microcyc.

Namespace conversion

To integrate the novel GPRs into the GSMM JP962, the GSMM was first converted into the standardized MnXRef namespace (Bernard et al., 2014) at MetaNetX.org (Ganter et al., 2013) to facilitate integration of reaction sets from external sources. Subsequently, the novel reaction sets were also converted into the MnXRef namespace using a custom script that takes into account the metabolites that are already included in the converted GSMM. This was done in order to account for possible differences in level of detail between the different sources. For example, the compound glucose can correspond to d-glucose or l-glucose, which in turn can correspond to α-d-glucose, β-d-glucose, α-l-glucose or β-l-glucose. In order to correctly connect new reactions to an existing GSMM, their metabolites thus need to not only be converted into the same namespace, but also at the correct level of detail. Metabolite names that had multiple plausible alternatives in the GSMM were manually checked following the logic of metabolic reactions (Danchin and Sekowska, 2014).

Model extension

For each predicted or curated reaction from the functional re-annotation process (hereafter: new reactions), the GSMM JP962 was first scanned to search for model reactions involving the same set of metabolites. Only if no such reaction existed was the new reaction added to the model. Otherwise, the existing and new reactions were compared in terms of reaction directionality and associated genes. If the new reaction had been manually curated, the GSMM reaction was updated in terms of both reaction directionality and gene associations. However, if the new reaction had not been manually curated the GSMM reaction directionality was left unchanged. In addition, the gene associations of the GSMM reaction were only updated if it was an orphan reaction.

Growth phenotype data using BIOLOG experiments

*Pseudomonas putida* KT2440 DSM 6125 was tested for its ability to utilize different carbon (C), nitrogen (N) and phosphorus (P) sources, using BIOLOG PM01, PM02A, PM03B and PM04A MicroPlates (Bochner et al., 2001). Bacteria were grown overnight on nutrient agar plates (DSMZ medium 1) at 28°C. Biolog experiments were performed according to the modified protocol “PM Procedures for *E. coli* and other GN Bacteria” (Biolog, Inc. 16 Jan 2006; see Supporting Information). Subsequently, for PM1 and PM2A experiments cells were transferred and suspended into 20 ml of Inoculating Fluid IF-0 to achieve 85% T (transmittance) in the BIOLOG Turbidimeter. About 240 μl Dye Mix A and 3760 μl H2O were added to a final volume of 24 ml. Each well of PM01 and PM02A MicroPlates (carbon sources) were inoculated with 100 μl of the 85% T cell suspension. PM3B and PM4A experiments require an appropriate carbon source, and a stock solution of 2 M sodium succinate and 200 μM ferric citrate was used as an additive as recommended in the PM procedures for gram-negative bacteria. Initial experiments with 85% T resulted in a strong metabolic response for both, the different substrates but also the negative control (PM3B—A1 [nitrogen]; PM4A—A1 [phosphorus], F1 [sulfur]). Accordingly, the amount of cells was successively reduced in a series of test experiments to a turbidity of 98%, which resulted in sufficient signal strength of the tested substrates combined with a comparably low conversion of the dye in the negative control. For the nitrogen plate PM3B, the optimized inoculation fluid contained 10 ml IF-0, 120 μl Dye Mix A, 60 μl additives and 1820 μl H2O, whereas the inoculation fluid of the phosphorus and sulfur plates PM4A contained 10 ml IF-0, 120 μl Dye Mix A, 120 μl additives and 1760 μl H2O. All PM plates were sealed with parafilm and inoculated in the OmniLog plate reader at 28°C. The conversion of the tetrazolium dye was measured and monitored all 15 min at OD590 for 4 days (96 h). The read-outs were analyzed with MicroLog software applying the automatic threshold option. BIOLOG measures above/below the threshold were considered as positive/negative phenotypes, respectively. The reading of plates involving sulfur compounds did not provide reliable results, presumably because the BIOLOG set up does not measure growth per se, but, rather, reflects an integrated view of the redox network of the cells in a particular environment.

In silico growth simulations

FBA was performed using the Cobra Toolbox (Schellenberger et al., 2011) with MatLab (MathworksInc., Natick, MA) and the gurobi solver (Gurobi Optimization, Houston, TX).
simulations were performed based on the GSMM JIP962 of *P. putida* KT2440 (Oberhardt *et al.*, 2011). Each well of the BIOLOG Microplates was simulated by adjusting the *in silico* medium to the available C-, N-, P- and S-sources. Specifically, the *in silico* media contained: bicarbonate, CO$_2$, cobalt, dihydrogen, iron, magnesium, nickel, oxygen, potassium, H$^+$, sodium, water, succinate (not in C-source tests), phosphate (not in P-source tests), sulfate (not in S-source tests), and the compound specific to the BIOLOG well (see Supporting Information Tables S12 and S14). We discriminated between growth and non-growth phenotypes based on a threshold value of 10$^{-6}$ gdw/gdw/h.

**In silico gene essentiality analysis**

The Cobra Toolbox (Schellenberger *et al.*, 2011) and the Gurobi solver (Gurobi Optimization., Houston, TX) were used for FBA simulations with MatLab (Mathworks Inc., Natnick, MA) based on the GSMM JIP962 of *P. putida* KT2440 (Oberhardt *et al.*, 2011). We expanded the original gene essentiality test-set of JIP815 (Puchalaka *et al.*, 2008) by adding new experimental data including auxotrophies (Molina-Henares *et al.*, 2010). Each knockout gene was simulated by blocking its associated reactions using the “deleteModelGenes” function of the COBRA toolbox. The *in silico* media were adjusted to the minimal media described for the experiments in Puchalaka *et al.* (2008) and Molina-Henares *et al.* (2010) (see Supporting Information Tables S12 and S14). We discriminated between growth and non-growth phenotypes of the mutants based on a threshold value equal to 50% of the wild-type growth rate in the same conditions.

**Metabolic network curation process**

Positive phenotypes in BIOLOG experiments not supported by metabolic model simulations were manually curated using tools and curation interfaces available in MicroScope (Vallenet *et al.*, 2013), in order to find potential catabolic pathways for the corresponding compounds. This includes the analysis of pre-computed results of several computational methods used in the functional annotation process (see above). In addition, genome-context methods available in MicroScope were also used in order to guide functional annotation curation and pathway hole filling: they are based on co-evolution of phylogenetic profiles with functionally related genes (Engelen *et al.*, 2012) and the conservation of genomic and metabolic context through the CANOE strategy used to find candidate genes for orphan enzymatic activities (Smith *et al.*, 2012). The outcome of these methods was further improved by extensive manual literature searches to add additional support to functional assignments. Finally, before the integration into the GSMM JIP962, GPR associations which include more than one gene, were manually curated to distinguish isozymes (OR relationships) from protein complexes (AND relationships).

Biochemical reactions and Gene-Reaction associations resulting from the curation process were manually validated in MicroScope using the MetaCyc (Caspi *et al.*, 2014) and the Rhea (Morgat *et al.*, 2015) reaction databases. Rhea was mainly used to manage biochemical reactions that are absent from the current MetaCyc repository. This implies the creation of new reactions directly in the Rhea database, starting with chemical compounds defined in the Chemical Entities of Biological Interest ontology (ChEBI) (Hastings *et al.*, 2013); reactions are stoichiometrically balanced for mass and charge at pH 7.3 (Morgat *et al.*, 2015). Similarly, in case of missing compounds in ChEBI with correct 2D structure at pH 7.3, the corresponding compounds were created *de novo* in ChEBI using the Marvin suite of tools from ChemAxon (http://www.chemaxon.com).

**Acknowledgments**

This work was supported by European Union’s 7th Framework Programme Microme FP7-KBBE-2007-3-2-08-222886. We would like to thank Brendan Ryback for his help with adding reactions to the existing GSMM, Victoria Michael for her support with BIOLOG experiments, and Kristian Axelsen for reading this manuscript.

**References**

Acevedo-Rocha, C.G., Fang, G., Schmidt, M., Ussery, D.W., and Danchin, A. (2013) From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet* 29: 273–279.

Achaz, G., Coissac, E., Viani, A., and Netter, P. (2000) Analysis of intrachromosomal duplications in yeast *Saccharomyces cerevisiae*: a possible model for their origin. *Mol Biol Evol* 17: 1268–1275.

Akashi, H., and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proc Natl Acad Sci U S A* 99: 3695–3700.

Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106.

Arias, S., Olivia, E.R., Arcos, M., Naharro, G., and Luengo, J.M. (2008) Genetic analyses and molecular characterization of the pathways involved in the conversion of 2-phenylethylamine and 2-phenylethanol into phenylacetic acid in *Pseudomonas putida* U. *Environ Microbiol* 10: 413–432.

Ballal, A., Basu, B., and Apte, S.K. (2007) The Kdp-ATPase system and its regulation. *J Bacteriol* 189: 559–568.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 41: D991–D995.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1247.

Bastard, K., Smith, A.A.T., Vergne-Vaxelaire, C., Perret, A., Zamarucha, A., De Melo-Minardi, R., *et al.* (2014) Revealing the hidden functional diversity of an enzyme family. *Nat Chem Biol* 10: 42–49.

Bay, D.C., and Turner, R.J. (2012) Small multidrug resistance protein EmrE reduces host pH and osmotic tolerance to metabolic quaternary cation osmoprotectants. *J Bacteriol* 194: 5941–5948.

Bernard, T., Bridge, A., Morgat, A., Moretti, S., Xenarios, I., and Pagni, M. (2014) Reconciliation of metabolites and
biochemical reactions for metabolic networks. *Brief Bioinform* **15**: 123–135.

Bertin, Y., Deval, C., de la Foye, A., Masson, L., Gannon, V., Harel, J., *et al.* (2014) The gluconeogenesis pathway is involved in maintenance of enterohemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. *PLoS One* **9**: e98367.

Bettenbrock, K., Bai, H., Ederer, M., Green, J., Hellinger, K.J., Holcombe, M., *et al.* (2014) Towards a systems level understanding of the oxygen response of *Escherichia coli*. *Adv Microb Physiol* **64**: 65–114.

Bochner, B.R. (1989) Sleuthing out bacterial identities. *Nature* **339**: 157–158.

Bochner, B.R., Gadzinski, P., and Panomitros, E. (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* **11**: 1246–1255.

Bocs, S., Cruveiller, S., Vallenet, D., Nuel, G., and Medigue, C. (2003) AMIGene: annotation of Microbial Genes. *Nucleic Acids Res* **31**: 3723–3726.

Brett, C.L., Donowitz, M., and Rao, R. (2005) Evolutionary origins of eukaryotic sodium/proton exchangers. *Am J Physiol Cell Physiol* **288**: C223–C239.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., *et al.* (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41**: D226–D232.

Carman, O., Rahav-Manor, O., Dover, N., Shaanan, B., and Padan, E. (1997) The Na⁺⁻specific interaction between the LysR-type regulator, NhaR, and the nhaA gene encoding the Na⁺/H⁺ antiporter of *Escherichia coli*. *EMBO J* **16**: 5922–5929.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42**: D459–D471.

Chandra, G., Chater, K.F., and Bornemann, S. (2011) Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology* **157**: 1565–1572.

Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., *et al.* (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res* **43**: D439–D446.

Chen, C., and Beattie, G.A. (2008) *Pseudomonas syringae* BetT is a low-affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. *J Bacteriol* **190**: 2717–2725.

Chen, C., Malek, A.A., Wargo, M.J., Hogan, D.A., and Beattie, G.A. (2010) The ATP-binding cassette transporter Cbc (choline/betaine/carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds. *Mol Microbiol* **75**: 29–45.

Chien, H.R., Jih, Y.L., Yang, W.Y., and Hsu, W.H. (1998) Identification of the open reading frame for the *Pseudomonas putida* D-hydantoinase gene and expression of the gene in *Escherichia coli*. *Biochim Biophys Acta* **1395**: 68–77.

Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003) Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* **31**: 6633–6639.

Danchin, A., and Sekowska, A. (2014) The logic of metabolism and its fuzzy consequences. *Environ Microbiol* **16**: 19–28.

De Lorenzo, V. (2015) It’s the metabolism, stupid! *Environ Microbiol Rep* **7**: 18–19.

De Smet, K.A., Weston, A., Brown, I.N., Young, D.B., and Robertson, B.D. (2000) Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology* **146**: 199–208.

Demerec, M., Adelberg, E.A., Clark, A.J., and Hartman, P.E. (1966) A proposal for a uniform nomenclature in bacterial genetics. *Genetics* **54**: 61–76.

Dos Santos, V.A.P.M., Heim, S., Moore, E.R.B., Strätz, M., and Timmis, K.N. (2004) Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environ Microbiol* **6**: 1264–1286.

Dover, N., and Padan, E. (2001) Transcription of nhaA, the main Na⁺/H⁺ antiporter of *Escherichia coli*, is regulated by Na⁺ and growth phase. *J Bacteriol* **183**: 644–653.

Elbein, A.D., Pastuszak, I., Tackett, A.J., Wilson, T., and Pan, Y.T. (2010) Last step in the conversion of trehalose to glycen: a mycobiological enzyme that transfers maltose from maltose 1-phosphate to glycen. *J Biol Chem* **285**: 9803–9812.

Engelen, S., Vallenet, D., Médigue, C., and Danchin, A. (2012) Distinct co-evolution patterns of genes associated to DNA polymerase III DnaE and PolC. * BMC Genom* **13**: 69.

Foster, J.W. (1999) When protons attack: microbial strategies of acid adaptation. *Curr Opin Microbiol* **2**: 170–174.

Foster, J.W. (2004) *Escherichia coli* acid resistance: tales of an amateur acidiophile. *Nat Rev Microbiol* **2**: 898–907.

Frank, S., Klockgether, J., Hagendorf, P., Geffers, R., Schöck, U., Pohl, T., *et al.* (2011) *Pseudomonas putida* KT2440 genome update by cDNA sequencing and microarray transcriptomics. *Environ Microbiol* **13**: 1309–1326.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**: D261–D269.

Ganter, M., Clowes, V., Moretti, S., Stelling, J., and Pagni, M. (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* **29**: 815–816.

Gassner, M., Mollenkamp, T., Puppe, W., and Altendorf, K. (1999) The KdpF subunit is part of the K⁺⁻translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex in vitro. *J Biol Chem* **274**: 37901–37907.

Golovan, S., Wang, G., Zhang, J., and Forsberg, C.W. (2000) Characterization and overproduction of the *Escherichia coli* appA encoded bifunctional enzyme that exhibits both phytase and acid phosphatase activities. *Can J Microbiol* **46**: 59–71.

Gralla, J.D., and Vargas, D.R. (2006) Potassium glutamate as a transcriptional inhibitor during bacterial osmoregulation. *EMBO J* **25**: 1515–1521.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., *et al.* (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res* **41**: D456–D463.

Heavner, B.D., Smallbone, K., Price, N.D., and Walker, L.P. (2013) Version 6 of the consensus yeast metabolic network
refines biochemical coverage and improves model performance. *Database (Oxford)* 2013: bat059.

Hidese, R., Mihara, H., Kurihara, T., and Esaki, N. (2011) *Escherichia coli* dihydropyrimidine dehydrogenase is a novel NAD-dependent heterotetramer essential for the production of 5,6-dihydrouracil. *J Bacteriol* 193: 989–993.

Hui, S., Silverman, J.M., Chen, S.S., Erickson, D.W., Basan, M., Wang, J., et al. (2015) Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Syst Biol* 11: e784.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11: 119.

Jiménez, J.J., Minambres, B., García, J.L., and Díaz, E. (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ Microbiol* 4: 824–841.

Jiménez, J.J., Canales, A., Jiménez-Barbero, J., Ginalski, K., Rychlewski, L., García, J.L., and Díaz, E. (2008) Deciphering the genetic determinants for aerobic nicotinic acid degradation: the nic cluster from *Pseudomonas putida* KT2440. *Proc Natl Acad Sci U S A* 105: 11329–11334.

Kaasen, I., Falkenberg, P., Styrvold, O.B., and Strøm, A.R. (1992) Molecular cloning and physical mapping of the otsBA genes, which encode the osmoregulatory trehalose pathway of *Escherichia coli* evidence that transcription is activated by katf (AppR). *J Bacteriol* 174: 889–898.

Kajiyama, Y., Otagiri, M., Sekiguchi, J., Kosono, S., and Kudo, T. (2007) Complex formation by the mrpABCDEFG genes, which encode the osmoregulatory trehalose pathway. *Environ Microbiol* 9: 1549–1558.

Karp, P.D., Latendresse, M., and Caspi, R. (2011) The pathway tools pathway prediction algorithm. *Stand Genom Sci* 5: 424–429.

Kaushik, J.K., and Bhat, R. (2003) Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *J Biol Chem* 278: 26458–26465.

Kell, D.B., and Oliver, S.G. (2014) How drugs get into cells: molecular strategies in neutralophilic bacteria. *FEMS Microbiol Rev* 38: 1091–1125.

MacMillan, S.V., Alexander, D.A., Culham, D.E., Kunte, H.J., Marshall, E.V., Rochon, D., and Wood, J.M. (1999) The ion coupling and organic substrate specificities of osmoregulatory transporter ProP in *Escherichia coli*. *Biochim Biophys Acta* 1420: 30–44.

Médigue, C., Wong, B.C.-Y., Lin, M.C.-M., Bocs, S., and Danchin, A. (2002) The secE gene of *Helicobacter pylori*. *J Bacteriol* 184: 2837–2840.

Meyer, F., Overbeek, R., and Rodriguez, A. (2009) FIGfams: yet another set of protein families. *Nucleic Acids Res* 37: 6643–6654.

Milanesio, P., Arce-Rodriguez, A., Muñoz, A., Calles, B., and de Lorenzo, V. (2011) Regulatory exaptation of the catabolite repression protein (Crp)-CAMP system in *Pseudomonas putida*. *Environ Microbiol* 13: 324–339.
Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**: D213–D221.

Molina-Henares, M.A., de la Torre, J., García-Salamanca, A., Molina-Henares, A.J., Herrera, M.C., Ramos, J.L., and Duque, E. (2010) Identification of conditionally essential genes for growth of *Pseudomonas putida* KT2440 on minimal medium through the screening of a genome-wide mutant library. *Environ Microbiol* **12**: 1468–1485.

Morgat, A., Axelsen, K.B., Lombardot, T., Alcántara, R., Aimo, L., Zerara, M., et al. (2015) Updates in Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res* **43**: D459–D464.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., et al. (2001) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799–808.

Nikel, P.I., Martínez-García, E., and de Lorenzo, V. (2014) Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **12**: 368–379.

Nikel, P.I., Chavarria, M., Fuhrer, T., Sauer, U., and de Lorenzo, V. (2015) *Pseudomonas putida* KT2440 strain metabolizes glucose through a cycle formed by enzymes of the Entner-Doudoroff, Embden-Meyerhof-Parnas, and pentose phosphate pathways. *J Biol Chem* **290**: 25920–25932.

Ning, Z., Cox, A.J., and Mullikin, J.C. (2001) SSAHA: a fast search method for large DNA databases. *Genome Res* **11**: 1725–1729.

Nogales, J., Canales, A., Jiménez-Barbero, J., García, J.L., and Díaz, E. (2005) Molecular characterization of the galactose dioxygenase from *Pseudomonas putida* KT2440. The prototype of a new subgroup of extradiol dioxygenases. *J Biol Chem* **280**: 35382–35390.

Nogales, J., Palsson, B.O., and Thiele, I. (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iNJ476 as a cell factory. *BMC Syst Biol* **2**: 79.

Nogales, J., Canales, A., Jiménez-Barbero, J., García, J.L., and Díaz, E. (2011) Unraveling the gallic acid degradation pathway in bacteria: the gal cluster from *Pseudomonas putida*. *Mol Microbiol* **79**: 359–374.

Oberhardt, M.A., Puchalka, J., Martíns dos Santos, V.A.P., and Papin, J.A. (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Comput Biol* **7**: e1001116.

Osterman, A. (2006) A hidden metabolic pathway exposed. *Proc Natl Acad Sci U S A* **103**: 5637–5638.

Overhage, J., Priefert, H., and Steinbüchel, A. (1999) Biochemical and genetic analyses of furulic acid catabolism in *Pseudomonas* sp. strain HR199. *Appl Environ Microbiol* **65**: 4837–4847.

Padan, E., Bibi, E., Ito, M., and Kruwich, T.A. (2005) Alkaline pH homeostasis in bacteria: new insights. *Biochim Biophys Acta* **1717**: 67–88.

Palleroni, N.J. (1984) Genus *Pseudomonas*. In *Berger's Manual of Systematic Bacteriology*, Vol. 1, Krieg, N.R., and Stanley, J.T. (eds). Baltimore, MD: Williams & Wilkins, pp. 141–199.

Parschat, K., Canne, C., Hüttermann, J., Kappl, R., and Fetzner, S. (2001) Xanthine dehydrogenase from *Pseudomonas putida* 86: specificity, oxidation-reduction potentials of its redox-active centers, and first EPR characterization. *Biochim Biophys Acta* **1544**: 151–165.

Pedrazzi, U., Rivoire, C., Auchincloss, A.H., Coudert, E., Keller, G., de Castro, E., et al. (2015) HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res* **43**: D1064–D1070.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0.2 discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786.

Pinner, E., Kotler, Y., Padan, E., and Schuldiner, S. (1993) Physiological role of nhaB, a specific Na+/H+ antiporter in *Escherichia coli*. *J Biol Chem* **268**: 1729–1734.

Puchalka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., et al. (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol* **4**: e1000210.

Rahav-Manor, O., Carmel, O., Karpel, R., Taglicht, D., Glaser, G., Schuldiner, S., and Padan, E. (1992) NhaA, a protein homologous to a family of bacterial regulatory proteins (LysR), regulates nhaA, the sodium proton antiporter gene in *Escherichia coli*. *J Biol Chem* **267**: 10433–10438.

Ramazzina, I., Folli, C., Secchi, A., Berni, R., and Percudani, R. (2006) Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. *Nat Chem Biol* **2**: 144–148.

Ramazzina, I., Cendon, L., Folli, C., Berni, R., Monteverdi, D., Zanotti, G., and Percudani, R. (2008) Logical identification of an allantoinase analog (puuB) recruited from poly saccharide deacetylases. *J Biol Chem* **283**: 23295–23304.

Ramos-González, M.I., and Molin, S. (1998) Cloning, sequencing, and phenotypic characterization of the rpoS gene from *Pseudomonas putida* KT2440. *J Bacteriol* **180**: 3421–3431.

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.

Reynes, J.P., Tiraby, M., Baron, M., Drocourt, D., and Tiraby, G. (1996) *Escherichia coli* thymidylate kinase: molecular cloning, nucleotide sequence, and genetic organization of the corresponding *tmk* locus. *J Bacteriol* **178**: 2804–2812.

Rkenes, T.P., Lamark, T., and Strom, A.R. (1996) DNA-binding properties of the BetI repressor protein of *Escherichia coli*: the inducer choline stimulates BetI-DNA complex formation. *J Bacteriol* **178**: 1663–1670.

Ruhl, R., Kataria, R., and Choudhury, B. (2013) Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation. *Microbiol Biotechnol* **6**: 493–502.

Saparov, S.M., Antonenko, Y.N., and Pohl, P. (2006) A new model of weak acid permeation through membranes revisited: does Overton still rule? *Biophys J* **90**: L86–L88.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* **6**: 1290–1307.
Schnackerz, K.D., and Dobritzsch, D. (2008) Amidohydrolases of the reductive pyrimidine catabolic pathway purification, characterization, structure, reaction mechanisms and enzyme deficiency. *Biochim Biophys Acta* 1784: 431–444.

Scovill, W.H., Schreier, H.J., and Bayles, K.W. (1996) Identification and characterization of the pckA gene from *Staphylococcus aureus*. *J Bacteriol* 178: 3362–3364.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., Levy, S.B., and Jackson, R.W. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 35: 652–680.

Smith, T.F., and Waterman, M.S. (1981) Identification of common molecular subsequences. *J Mol Biol* 147: 195–197.

Smith, L.T., Pocard, J.A., Bernard, T., and Le Rudulier, D. (1988) Osmotic control of glycine betaine biosynthesis and degradation in *Rhizobium mellioti*. *J Bacteriol* 170: 3142–3149.

Smith, A.A.T., Belda, E., Viana, A., Medigue, C., and Vallenet, D. (2012) The CanOE strategy: integrating genomic and metabolic contexts across multiple prokaryote genomes to find candidate genes for orphan enzymes. *PLoS Comput Biol* 8: e1002540.

Sohn, S.B., Kim, T.Y., Park, J.M., and Lee, S.Y. (2010) In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnol J* 5: 739–750.

Stancik, L.M., Stancik, D.M., Schmidt, B., Barnhart, D.M., Yoncheva, Y.N., and Slonczewski, J.L. (2002) pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *J Bacteriol* 184: 4246–4258.

Sudom, A., Walters, R., Pastushok, L., Goldie, D., Prasad, L., Delbaere, L.T.J., and Goldie, H. (2003) Mechanisms of activation of phosphoenolpyruvate carboxykinase from *Escherichia coli* by Ca\(^{2+}\) and of desensitization by trypsin. *J Bacteriol* 185: 4233–4242.

The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res* 43: D204–D212.

Timinskas, K., Balvočiutė, M., Timinskas, A., and Venclovas, C. (2014) Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic Acids Res* 42: 1393–1413.

Tormo, A., Almirón, M., and Kolter, R. (1990) *surA*, an *Escherichia coli* gene essential for survival in stationary phase. *J Bacteriol* 172: 4339–4347.

Touchon, M., Hoece, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., *et al.* (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5: e1000344.

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J.L. (2015) Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environ Microbiol*, In press, DOI: 10.1111/1462-2920.13015.

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., *et al.* (2013) MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res* 41: D636–D647.

Van Duuren, J.B.J.H., Puchalka, J., Mars, A.E., Bucker, R., Eggink, G., Wittmann, C., and dos Santos, V.A.P.M. (2013) Reconciling in vivo and in silico key biological parameters of *Pseudomonas putida* KT2440 during growth on glucose under carbon-limited condition. *BMC Biotechnol* 13: 93.

Velasco-García, R., Villalobos, M.A., Ramirez-Romero, M.A., Mújica-Jiménez, C., Iturriaga, G., and Muñoz-Clare, R.A. (2006) Betaine aldehyde dehydrogenase from *Pseudomonas aeruginosa* cloning, over-expression in *Escherichia coli*, and regulation by choline and salt. *Arch Microbiol* 185: 14–22.

Venturi, V. (2003) Control of rpoS transcription in *Escherichia coli* and *Pseudomonas*: why so different? *Mol Microbiol* 49: 1–9.

Wargo, M.J. (2013) Homeostasis and catabolism of choline and glycine betaine: lessons from *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 79: 2112–2120.

Wargo, M.J., and Hogan, D.A. (2009) Identification of genes required for *Pseudomonas aeruginosa* carnitine catabolism. *Microbiology* 155: 2411–2419.

Wargo, M.J., Szewrgold, B.S., and Hogan, D.A. (2008) Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J Bacteriol* 190: 2690–2699.

West, T.P. (2001) Pyrimidine base catabolism in *Pseudomonas putida* biotype B. *Antonie van Leeuwenhoek* 80: 163–167.

Winsor, G.L., Lam, D.K.W., Fleming, L., Lo, R., Whiteside, M.D., Yu, N.Y., *et al.* (2011) *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res* 39: D596–D600.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. (2011) Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiol Rev* 35: 299–323.

Yang, L., Tan, J., O’Brien, E.J., Monk, J.M., Kim, D., Li, H.J., *et al.* (2015) Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. *Proc Natl Acad Sci* 112: 201501384.

Ye, L., Hildebrand, F., Dingemans, J., Ballet, S., Laus, G., Matthijs, S., *et al.* (2014) Draft genome sequence analysis of a *Pseudomonas putida* W15Oct28 strain with antagonistic activity to Gram-positive and *Pseudomonas* sp. pathogens. *PLoS One* 9: e110038.

Yu, N.Y., Laird, M.R., Spencer, C., and Brinkman, F.S.L. (2011) PSORTdb—an expanded, auto-updated, user-friendly protein subcellular localization database for Bacteria and Archaea. *Nucleic Acids Res* 39: D241–D244.

Ziegler, C., Bremer, E., and Kramer, R. (2010) The BCCT family of carriers: from physiology to crystal structure. *Mol Microbiol* 78: 13–34.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Fig. S1** (A) Genomic organization of the *ddp* operon in *P. putida* KT2440 and *P. aeruginosa* PAO1, responsible for the dipeptide degradation. Only *P. aeruginosa* *mdpA* and *psdR* genes do not have orthologs in *P. putida* KT2440. (B) Protein domain organization of the *mdpA* gene in *P. aeruginosa* PAO1 (PA4498) coding for a metallopeptidase involved in dipeptide degradation. In *P. putida* KT2440, three genes coding for putative peptidases have similar protein domain architecture to that of *mdpA*.
Supplementary Results: curation of *P. putida* KT2440 metabolic pathways

**Table S1.** SNPs/InDels identified in coding regions from the re-sequenced wild-type *P. putida* KT2440.

**Table S2.** Non-coding regions larger than 1kb in the re-sequenced genome of *P. putida* KT2440.

**Table S3.** CDS in original annotation removed in the re-annotated genome sequence of *P. putida* KT2440.

**Table S4.** Novel CDSs in the re-annotated genome sequence of *P. putida* KT2440.

**Table S5.** Pseudogenes in the re-annotated genome sequence of *P. putida* KT2440.

**Table S6.** Original genes of unknown function with updated functional annotation in the re-annotated genome sequence of *P. putida* KT2440.

**Table S7.** Functional annotation of curated genes with extended substrate specificity in *P. putida* KT2440.

**Table S8.** Functional annotation of curated genes involved in control of osmolarity in *P. putida* KT2440.

**Table S9.** Functional annotation of curated genes involved in control of proton availability in *P. putida* KT2440.

**Table S10.** Functional annotation of curated genes involved in degradation of aromatic compounds in *P. putida* KT2440.

**Table S11.** Functional annotation of curated genes involved in carnitine degradation and pyoverdine biosynthesis in *P. putida* KT2440.

**Table S12.** Growth phenotype predictions of the *P. putida* KT2440 metabolic models for BIOLOG experiments and additional literature data.

**Table S13.** Overview of the knowledge gap compounds and the required steps for degradation in the *P. putida* KT2440 metabolic models.

**Table S14.** *In-silico* growth media used for flux balance analysis simulations over *P. putida* KT2440 metabolic models.

**Table S15.** Automatically predicted gene-reaction associations from the re-annotated *P. putida* KT2440 genome sequence.