

The Complete Genome of *Comamonas testosteroni* Reveals Its Genetic Adaptations to Changing Environments^{▽†}

Ying-Fei Ma,^{1,2} Yun Zhang,¹ Jia-Yue Zhang,¹ Dong-Wei Chen,¹ Yongqian Zhu,³ Huajun Zheng,³
Sheng-Yue Wang,³ Cheng-Ying Jiang,¹ Guo-Ping Zhao,^{3,4} and Shuang-Jiang Liu^{1,2*}

State Key Laboratory of Microbial Resources¹ and Environmental Microbiology and Biotechnology Research Center at Institute of Microbiology,² Beijing 100101, People's Republic of China, and Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai,³ and Shanghai Institutes for Biological Sciences,⁴ Shanghai 201203, People's Republic of China

Received 25 April 2009/Accepted 27 August 2009

Members of the gram-negative, strictly aerobic genus *Comamonas* occur in various environments. Here we report the complete genome of *Comamonas testosteroni* strain CNB-2. Strain CNB-2 has a circular chromosome that is 5,373,643 bp long and has a G+C content of 61.4%. A total of 4,803 open reading frames (ORFs) were identified; 3,514 of these ORFs are functionally assigned to energy production, cell growth, signal transduction, or transportation, while 866 ORFs encode hypothetical proteins and 423 ORFs encode purely hypothetical proteins. The CNB-2 genome has many genes for transportation (22%) and signal transduction (6%), which allows the cells to respond and adapt to changing environments. Strain CNB-2 does not assimilate carbohydrates due to the lack of genes encoding proteins involved in glycolysis and pentose phosphate pathways, and it contains many genes encoding proteins involved in degradation of aromatic compounds. We identified 66 Tct and nine TRAP-T systems and a complete tricarboxylic acid cycle, which may allow CNB-2 to take up and metabolize a range of carboxylic acids. This nutritional bias for carboxylic acids and aromatic compounds enables strain CNB-2 to occupy unique niches in environments. Four different sets of terminal oxidases for the respiratory system were identified, and they putatively functioned at different oxygen concentrations. This study conclusively revealed at the genomic level that the genetic versatility of *C. testosteroni* is vital for competition with other bacteria in its special niches.

The members of the genus *Comamonas* are gram-negative, strict aerobes and frequently occur in diverse habitats, including activated sludge, marshes, marine habitats, and plant and animal tissues (4, 12, 13). They grow on organic acids, amino acids, and peptone, but they rarely attack carbohydrates. Some species, such as *Comamonas testosteroni*, can also mineralize complex and xenobiotic compounds, such as testosterone (17) and 4-chloronitrobenzene (CNB) (54). Their diversified niches make *Comamonas* species environmentally important and also suggest that the genus *Comamonas* represents a group of bacteria that can adapt very well, both ecologically and physiologically, to environments.

To understand better how environmental microbes adapt to their environments, many well-known environmental microbes, such as *Pseudomonas putida* (53) and *Rhodococcus* sp. strain RAH1 (31), have been sequenced. The genome data for these organisms, as well as other environmental microbes, provide not only an understanding of physiological and environmental functions at the genetic level but also a starting point for systems biology analyses of these microbes. Until now, none of the *Comamonas* species has been sequenced, although

these organisms represent an important group of environmental microbes.

C. testosteroni strain CNB-1 was isolated from CNB-contaminated activated sludge and grows with CNB as a sole source of carbon and nitrogen, and it has been used successfully for rhizoremediation of CNB-polluted soil (25). Strain CNB-1 has a circular chromosome and a large plasmid, and the genes involved in the degradation of CNB on plasmid pCNB1 were identified previously (28). In the present study, the genome of strain CNB-2, which was derived from strain CNB-1, was sequenced, and a genome analysis was performed parallel to physiological experiments. The aim of this work was to obtain genetic insight into how *C. testosteroni* adapts to changing and diverse environments.

MATERIALS AND METHODS

Bacterial strain and physiological and biochemical tests. Strain CNB-2 is a mutant of *C. testosteroni* CNB-1 that lost the degrading plasmid pCNB1 (28). The physiological and biochemical properties of both CNB-1 and CNB-2 were determined by using API 20NE and API 50CH identification systems (bioMérieux, Hazelwood, MO) at 30°C.

Genome sequencing. The genome of CNB-2 was sequenced by using a massive parallel pyrosequencing technology (454 GS 20) (30). All useful reads were assembled into 545 contigs (>500 bp) covering the genome 25-fold. The gaps were closed by using a multiplex PCR technique (10, 48). Finally, the sequences were assembled using the PHRED_PHRAP_CONSED software package with the Unix system (10, 48). The final consensus quality score of each base was more than 30. Genome assembly was assessed by performing long-range PCR and GC skew analysis (see Fig. S1 in the supplemental material).

Sequence analysis. Coding sequences were predicted by combining results obtained with Glimmer packages and GeneMark online (6, 27). Translational products of the coding sequences were annotated by searching the GenBank

* Corresponding author. Mailing address: Institute of Microbiology, Chinese Academy of Sciences, Bei-Chen-Xi-Lu No. 1, Chao-Yang District, Beijing 100101, People's Republic of China. Phone: 86-10-64807423. Fax: 86-10-64807421. E-mail: liusj@sun.im.ac.cn.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

[▽] Published ahead of print on 4 September 2009.

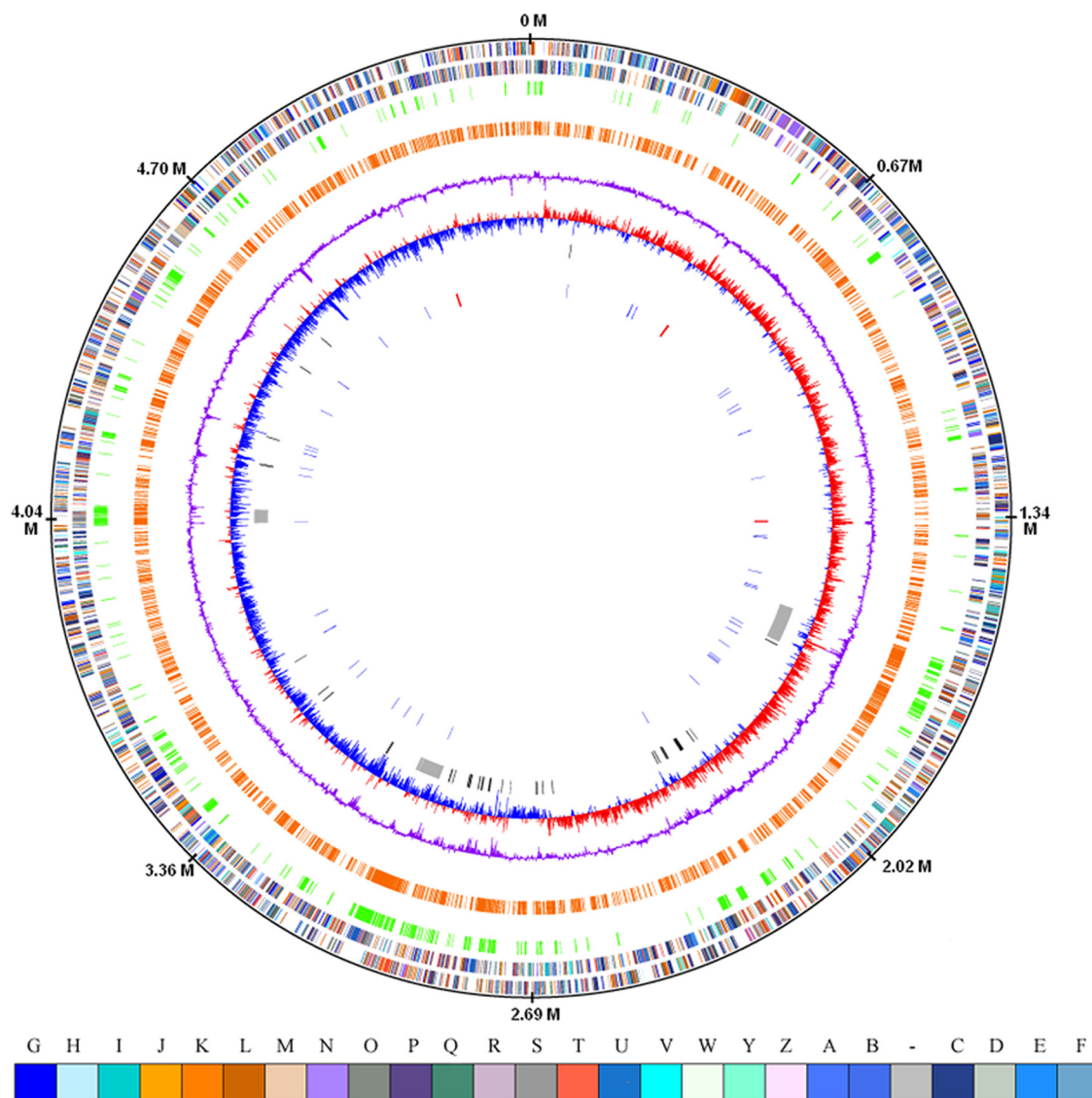


FIG. 1. Circular chromosome of *C. testosteroni* CNB-2. The scale on the outside indicates the size (in Mbp). The first two rings (from the outside in) indicate the genes in the forward and reverse strands, respectively. Rings 3 and 4 indicate the genes of CNB-2 that have no significant similarity to the genes of *C. testosteroni* KF-1 (green) and *P. putida* KT2440 (red). Rings 5 and 6 indicate the G+C content and GC skew $[(C - G)/(C + G)]$, respectively. Ring 7 indicates prophages (gray) and genes encoding transposases, recombinases, and integrases (black). Ring 8 indicates tRNA genes (blue) and *rRNA* operons (tomato). Position 1 of the chromosome was assigned to the first nucleotide of the DnaA gene. In rings 1 and 2, the colors of the genes indicate the COG categories, as indicated at the bottom.

database using the basic local alignment search tool (BLAST) for proteins. Putative functions of translation products were confirmed by using KEGG, TIGRFams, and Clusters of Orthologous Groups (COGs) (14, 33, 46). The tRNA genes were identified by using the tRNAscan-SE tool (26). The rRNA genes were identified by BLASTN using the 16S rRNA gene of strain CNB-1. Transporter genes were predicted by analyzing all annotated genes (putative proteins) with the Transport Classification Database (<http://www.tcd.org>) (39). The metabolic pathway of CNB-2 was constructed in silico by using KEGG tools (33). The protein domains were predicted using the NCBI Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) (29). Transmembrane domains and signal peptide regions were identified with TMHMM2.0 and SignalP3.0, respectively (2, 23).

Genome comparison. The genome data for *C. testosteroni* KF-1 (accession no. NZ_AAUJ000000000), *Escherichia coli* ATCC 8739 (CP000946), and *P. putida* KT2440 (AE015451) were retrieved from a website (<http://www.ncbi.nlm.nih.gov/genomes>). BLASTP was employed for protein sequence analysis. Significant

similarity of proteins was defined using an *e* value of $<1e-05$ and $>30\%$ identity.

Nucleotide sequence accession number. The complete genome sequence and annotation of *C. testosteroni* CNB-2 have been deposited in the GenBank database under accession number CP001220.

RESULTS AND DISCUSSION

General genome features. Strain CNB-2 has a circular chromosome that is 5,373,643 bp long and has a G+C content of 61.4% (Fig. 1). It contains 4,803 predicted open reading frames (ORFs), three rRNA operons, and 79 tRNA genes for all 20 amino acids (Table 1). Based on BLASTP searches (*e* value, $<1e-5$), 3,514 ORFs encode proteins with assigned functions,

TABLE 1. General features of the genomes of *C. testosteroni* CNB-2 and KF-1

Strain	Total size (bp)	G+C content (%)	Coding density (%)	No. of ORFs				Avg ORF length (bp)	No. of rRNA operons	No. of tRNA genes
				Total	With assigned function	Encoding conserved proteins	Encoding hypothetical proteins			
CNB-2	5,373,643	61	86	4,803	3,514	866	423	963	3	79
KF-1 ^a	6,026,527	61	86	5,492	4,266	952	NA ^b	952	6	96

^a Data for the unfinished genome of strain KF-1 were obtained from the GenBank database.

^b NA, not analyzed.

866 conserved proteins, and 423 hypothetical proteins. Using COG functional assignment, 4,649 ORFs were predicted. The gene coding sequences cover 86.4% of the genome, and the average length of the coding sequences is 963 bp. The likely origin of replication (*oriC*) was located between the ribosomal protein L34 and DnaA genes based on GC skew analysis (see Fig. S1 in the supplemental material). Four putative DnaA boxes (DnaA binding sites) were found between the L34 and DnaA genes.

MGEs, HGT, and adaptation to environments. The genome of CNB-2 contains large numbers of mobile genetic elements (MGEs) (Table 2). Three putative prophages (prophages I, II, and III) were identified. Remarkably, these prophages harbor accessory genes possibly acquired by horizontal gene transfer (HGT) from other bacterial species. For example, a gene cluster encoding putative terephthalate dioxygenase (CtCNB1_1493 to CtCNB1_1496) was identified in prophage I. A total of 39 ORFs encoding various transposases and integrases of the IS21, IS4, IS3, and IS110 families were discovered (Table 2). The presence of highly homologous or even identical ORFs among these 39 ORFs suggests that recent gene transposition or gene duplication events occurred in the genome. The MGEs in the CNB-2 genome are located mainly in the region near the terminal DNA replication site.

HGT might result in significant changes in the G+C content of the genome and provide additional physiological capacity to the cells. In the region from 2.75 to 2.97 Mbp, the G+C content is 57.0%, which is lower than the G+C content of the whole genome (61.4%). Based on the observation that there are genes encoding many putative transposases in this region (Fig. 1), it was deduced that strain CNB-2 acquired this region via HTG. Functionally, this region of the CNB-2 genome contains gene clusters putatively involved in xenobiotic compound degradation (e.g., protocatechuate degradation pathway) and heavy metal resistance (copper and arsenate resistance).

Comparative genome. A different strain of *C. testosteroni*, strain KF-1, is being sequenced (<http://www.ncbi.nlm.nih.gov>). Based on 16S rRNA gene analysis, strains CNB-2 and KF-1 are close phylogenetic relatives (99% identity). In addition, they occur in similar niches (activated sludge) and are able to degrade xenobiotics. A comparison of the two genomes showed that they are highly similar to each other; the genetic evidence that these strains are ecologically and physiologically similar includes the finding that 4,272 of the 4,803 ORFs of strain CNB-2 are significantly similar to their counterparts in strain KF-1, whose genome contains 5,492 ORFs. However, the results also revealed differences between CNB-2 and KF-1 (Table 1). The difference in the sizes of the two genomes is ca. 0.7

Mbp. The three prophages of CNB-2 were not found in the KF-1 genome. Corresponding ORFs were not found in KF-1 for a total of 531 ORFs of CNB-2 that are frequently associated with MGEs. The genomic differences between CNB-2 and KF-1 are at least partially due to the abundance of MGEs in the CNB-2 genome. The results of a further comparative analysis of the genomes of strains CNB-2, KF-2, *P. putida* KT2440, and *E. coli* ATCC 8739 are shown in Fig. S2 in the supplemental material.

TABLE 2. Mobile genetic elements in the *C. testosteroni* CNB-2 genome

Element	Gene(s)
Prophages	
I.....	CtCNB1_1435 to CtCNB1_1533
II.....	CtCNB1_2607 to CtCNB1_2676
III.....	CtCNB1_3548 to CtCNB1_3599
Transposases	
IS21.....	CtCNB1_1935, CtCNB1_1936, CtCNB1_1978, CtCNB1_1979, CtCNB1_2459, CtCNB1_2458, CtCNB1_2521, CtCNB1_2522, CtCNB1_2582, CtCNB1_2583, CtCNB1_3734, CtCNB1_3735, CtCNB1_2520, CtCNB1_2522, CtCNB1_1546, CtCNB1_1547
Integrases.....	CtCNB1_1933, CtCNB1_1982, CtCNB1_2289, CtCNB1_2318, CtCNB1_3007, CtCNB1_4129, CtCNB1_1938, CtCNB1_1940, CtCNB1_2411, CtCNB1_2336, CtCNB1_2006, CtCNB1_2333, CtCNB1_1888, CtCNB1_2002, CtCNB1_2003, CtCNB1_2475, CtCNB1_2476, CtCNB1_3143, CtCNB1_1900, CtCNB1_1901, CtCNB1_1888, CtCNB1_2002, CtCNB1_2003, CtCNB1_2475, CtCNB1_2476, CtCNB1_3143, CtCNB1_1900, CtCNB1_1901, CtCNB1_2333
IS116/IS110/IS902.....	CtCNB1_2462, CtCNB1_2490, CtCNB1_2491
IS3/IS911.....	CtCNB1_1899, CtCNB1_2479, CtCNB1_4038, CtCNB1_3809
IS4.....	CtCNB1_0110, CtCNB1_2435, CtCNB1_2460, CtCNB1_2461, CtCNB1_2571, CtCNB1_2765, CtCNB1_3023, CtCNB1_2488, CtCNB1_2763
Tn3.....	CtCNB1_2516
Others.....	CtCNB1_3808, CtCNB1_4037, CtCNB1_3731, CtCNB1_2334, CtCNB1_2478

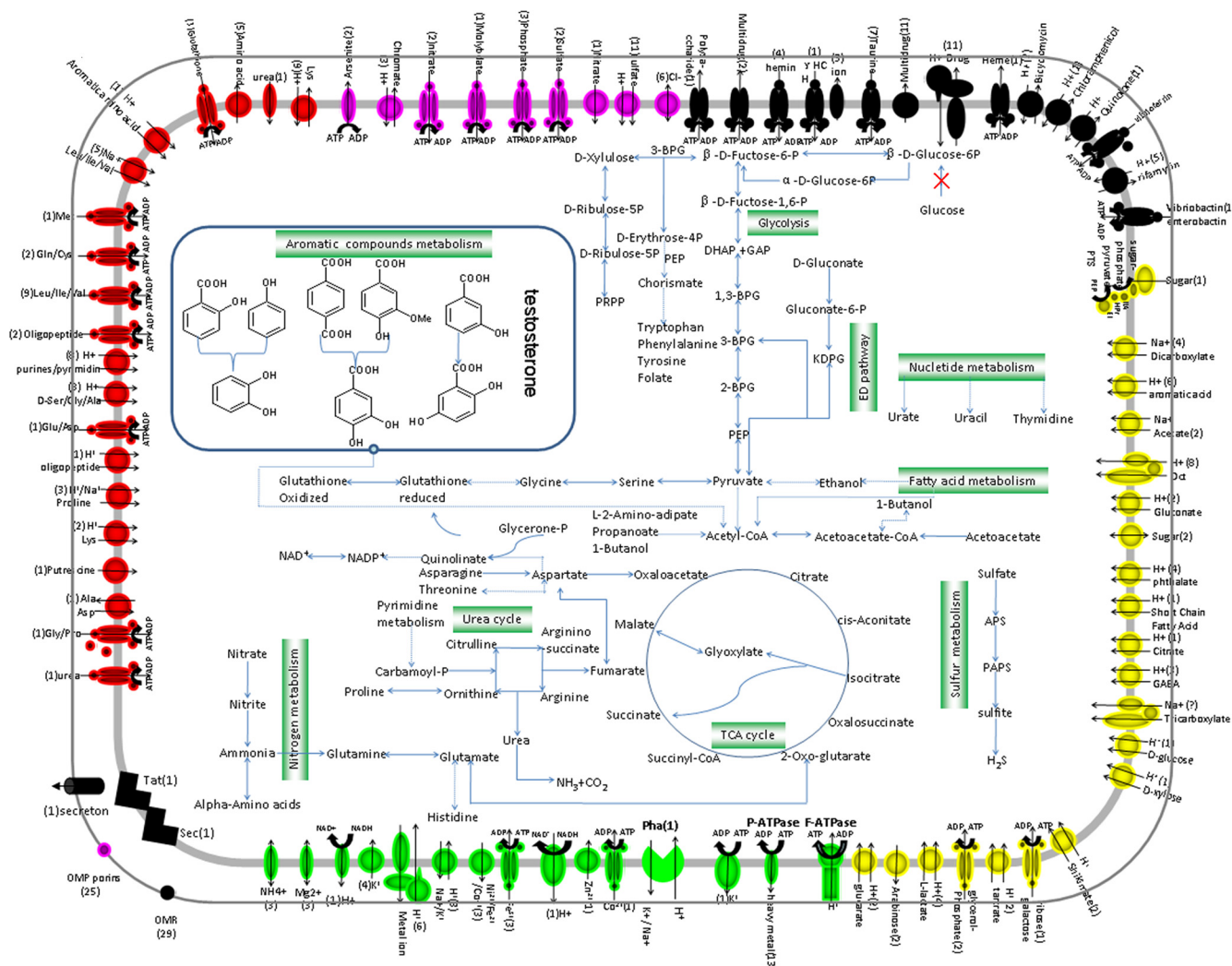


FIG. 2. Overview of metabolism and transport in *C. testosteroni* CNB-2. Transporters are grouped by substrate specificity, as follows: inorganic cations (green); inorganic anions (pink); carbohydrates (yellow); amino acids, peptides, amines, purines, pyrimidines, and other nitrogenous compounds (red); and drug efflux and other (black). Arrows indicate the direction of substrate transport. DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde phosphate; PEP, phosphoenolpyruvate; BPG, bisphosphoglycerate; PRPP, phosphoribosylpyrophosphate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; ED, Entner-Doudoroff; APS, adenosine phosphosulfate; PAPS, phosphoadenosine phosphosulfate; TCA, tricarboxylic acid; Me, methyl; OMP, outer membrane protein; OMR, outer membrane receptor; GABA, γ -aminobutyrate.

General metabolism of strain CNB-2 and core physiological features of the genus *Comamonas*. Strain CNB-2 grows on mineral medium, indicating that it is able to synthesize all its cellular components, such as fatty acids, purine and pyrimidine nucleotides, and the 20 amino acids. The results of the in silico genomic analysis supported this hypothesis. More information concerning general metabolism is shown in Fig. 2.

The genus *Comamonas* was defined as a genus that has a poor ability to assimilate carbohydrates. A total of 49 carbon sources were tested in this study, and CNB-2 was not able to assimilate any sugar except glycerol and gluconate. The glycolysis in glucose catabolism was incomplete due to the lack of hexokinase and glucokinase genes. Genes encoding glucose-6-phosphate 1-dehydrogenase and 6-phosphogluconolactonase of the pentose phosphate pathway were not found in the CNB-2 genome. Thus, oxidation of glucose via this pentose phosphate pathway was not possible (Fig. 2). However, the

nonoxidative part of the pentose phosphate pathway was complete, and five-carbon sugars were generated for biosynthesis. An almost complete Entner-Doudoroff pathway was constructed in silico for gluconate assimilation (Fig. 2), and physiological tests showed that strain CNB-2 grew on gluconate. In contrast to the poor ability of *Comamonas* species to metabolize carbohydrates, members of the genus *Pseudomonas*, such as *P. putida* KT2440, are able to metabolize many sugars (19). This genome project examining a representative of *Comamonas* species revealed that these species do not have a complete sugar-metabolizing pathway. Based on these findings, it was proposed that *Comamonas* and *Pseudomonas* species, which are two groups of environmental microbes, might play different roles in the geobiochemical cycles of elements.

C. testosteroni was named based on its ability to metabolize testosterone. A cluster of genes (CtCNB1_1351 to CtCNB1_1362) similar to the previously identified genes responsible for

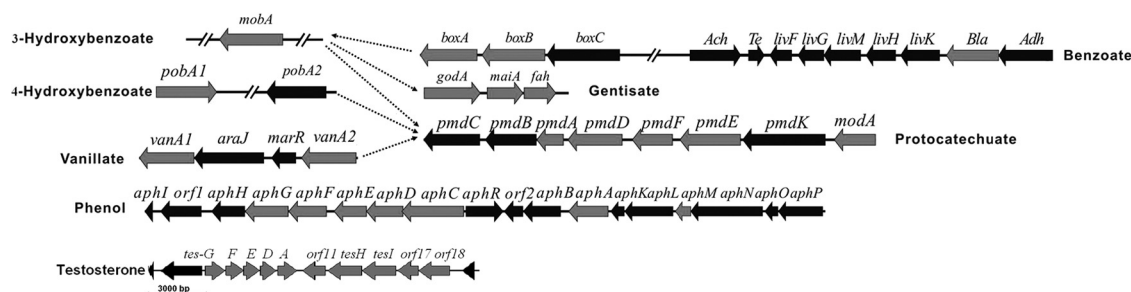


FIG. 3. Gene clusters involved in aromatic compound metabolism in *C. testosteroni* strain CNB-2. *aphA*, catechol 2,3-dioxygenase; *aphB*, protein involved in meta pathway of phenol degradation; *aphR*, gene encoding transcriptional regulator; *aphC*, gene encoding 2-hydroxymuconic semialdehyde dehydrogenase; *aphD*, gene encoding 2-hydroxymuconic semialdehyde hydrolase; *aphE*, gene encoding 2-hydroxypent-2,4-dienoate hydratase; *aphF*, gene encoding acetaldehyde dehydrogenase (acetylating); *aphG*, gene encoding 4-hydroxy-2-oxovalerate aldolase; *aphH*, gene encoding 4-oxalocrotonate decarboxylase; *aphI*, gene encoding 4-oxalocrotonate tautomerase; *Bla*, gene encoding benzoate-CoA ligase family; *boxA*, gene encoding benzoyl-CoA oxygenase/reductase component A; *boxB*, gene encoding benzoyl-CoA oxygenase component B; *boxC*, gene encoding enoyl-CoA hydratase/isomerase; *modA*, gene encoding ABC-type molybdate transport system periplasmic component; *pmdA*, gene encoding protocatechuate 4,5-dioxygenase alpha subunit; *pmdB*, gene encoding protocatechuate 4,5-dioxygenase beta subunit; *pmdC*, gene encoding 2-hydroxy-4-carboxymuconic semialdehyde dehydrogenase; *pmdD*, gene encoding 2-pyrone-4,6-dicarboxylic acid hydrolase; *pmdE*, gene encoding 4-oxalomesaconate hydratase; *pmdF*, gene encoding 4-hydroxy-4-methyl-2-oxoglutarate aldolase; *pmdK*, gene encoding transporter; *vanA1*, gene encoding vanillate monooxygenase; *godA*, gene encoding gentisate 1,2-dioxygenase; *maiA*, gene encoding maleylacetoacetate isomerase; *fah*, gene encoding fumarylacetoacetate hydrolase. The arrows indicate putative pathways of the compounds. The gray arrows indicate the genes that may function in a degradation process, while the black arrows indicate the genes with unknown functions.

testosterone degradation in *C. testosteroni* TA441 was found in the strain CNB-2 genome (1, 16–18) (Fig. 3).

As an important environmental microbe, strain CNB-2 is able to utilize many other kinds of compounds, such as aromatics and short-chain fatty acids, as carbon sources. In the CNB-2 genome, 37, 18, and 47 genes encoding putative dioxygenases, hydroxylases, and oxidoreductases, respectively, were annotated for aromatic and cyclic hydrocarbon degradation. Strain CNB-2 was able to grow on benzoate, gentisate, phenol, 3-hydroxybenzoate, 4-hydroxybenzoate, protocatechuate, and vanillate, and the corresponding gene clusters in the genome were identified (Fig. 3). Key genes involved in degradation of the following compounds were identified in the CNB-2 genome: catechol (CtCNB1_3144 to CtCNB1_3155), protocatechuate (CtCNB1_2740 to CtCNB1_2747), and gentisate (CtCNB1_2777 to CtCNB1_2780). A range of peripheral enzymes that direct aromatic compounds into central pathways were also discovered, such as benzoyl-coenzyme A (benzoyl-CoA) ligase (CtCNB1_0097), benzoyl-CoA oxygenase (CtCNB1_0065 to CtCNB1_0067), phenol hydroxylase (CtCNB1_3158 to CtCNB1_3162), 3-hydroxybenzoate hydroxylase (CtCNB1_3410), 4-hydroxybenzoate hydroxylase (CtCNB1_2156), and vanillate monooxygenase (CtCNB1_4192). Also, 66 Tct systems and nine TRAP-T systems (see Table S1 in the supplemental material) for transportation of carboxylic acids were identified. The carboxylic acids were metabolized through the tricarboxylic acid cycle to provide energy and key intermediates for cell growth.

Strain CNB-2 uses nitrate and ammonium as nitrogen sources, and three genes (CtCNB1_0130, CtCNB1_0202, and CtCNB1_0616) encoding ammonium channels and two gene clusters (CtCNB1_0586 to CtCNB1_0588 and CtCNB1_1234 to CtCNB1_1236) for ABC-type nitrate transporters were found in the genome. Inorganic nitrogen is incorporated into glutamine by glutamine synthetase (CtCNB1_3271). Additionally, genes encoding putative transporters for import of exogenous amino acids, oligopeptides, and branched-chain amino acids were identified in the CNB-2 genome. These compounds

could be metabolized through the urea cycle and tricarboxylic acid cycle, which could supply not only an adequate nitrogen source but also a carbon source for cell growth.

Transport systems. Strain CNB-2 contains genes for a broad range of transport proteins (Fig. 2; see Table S1 in the supplemental material). A total of 22% of all genes were predicted to be involved in substrate transport. This value is higher than the values for *E. coli*, *Haemophilus influenzae*, and *Helicobacter pylori*, which live in relatively closed environments, but it is similar to the values for microbes such as *P. putida* that live in open environments. Ten genes were predicted to encode a channel transporter (TC #1.A) responsible for ammonia, magnesium-cobalt, and urea-amide. A total of 378 genes probably encode carriers (TC #2) for sugars, polyols, drugs, neurotransmitters, Krebs cycle metabolites, phosphorylated glycolytic intermediates, amino acids, peptides, osmolytes, siderophores, iron siderophores, nucleosides, organic anions, and inorganic anions. Thirteen predicted APC family transporters (TC #2.A.3) may be involved in amino acid transport in the cells. Thirty-six genes encode 12 protein complexes of the RND family, which catalyze drug or heavy metal efflux via an H⁺ antiport mechanism. In the genome of CNB-2, the group of transporters (TC #3.A.) driven by P~P bond hydrolysis is the largest group, and such transporters export or import a large number of compounds, including amino acids, drugs, metal ions, inorganic anions, proteins, and aromatic compounds.

Effective transport is important for microbial evolution. Import of compounds into cells is vital for metabolic pathways, and export of toxic compounds out of cells is also vital for bacterial survival in harsh environments. As mentioned above, 22% of the genes in the CNB-2 genome were predicted to encode channels, electrochemical potential-driven transporters, ABC-type transporters, RND efflux systems, protein secretion systems, and other transport-related proteins based on the results obtained using BLASTP searches in the Transport Classification Database (e value, <10e−5). This value is similar to the values for bacteria living in soil or plants, including

Streptomyces coelicolor (35), *P. putida* KT2440 (53), and *Pseudomonas aeruginosa* (24), but it is higher than the values for symbiotic and pathogenic bacteria, such as *Mycobacterium leprae* (51). Further analysis revealed that *P. aeruginosa* (24) and *Agrobacterium tumefaciens* (5) possess high numbers of transporters for sugars, amino acids, and peptides and that *C. testosteroni* CNB-2 and *P. putida* KT2440 (53) have many fewer transporters for sugars. Interestingly, *P. putida* utilizes many sugars for growth, but strains of *Comamonas* species, including strain CNB-2, assimilate few sugars.

Secretion systems. Protein export that controls secretion of toxins, hydrolytic enzymes, and adhesins has an important role in *Comamonas* species living in various niches and colonizing different surfaces. Four types of protein secretion have been identified in gram-negative bacteria (40), but only two types were found when the CNB-2 genome was examined. The putative type I secretion system of strain CNB-2 consists of proteins encoded by five gene clusters (CtCNB1_2177 to CtCNB1_2179, CtCNB1_2281 to CtCNB1_2285, CtCNB1_2442 to CtCNB1_2446, CtCNB1_3019 to CtCNB1_3022, and CtCNB1_3183 to CtCNB1_3186), including a pore-forming outer membrane protein, a membrane fusion protein, and an inner membrane ATP-binding cassette protein. The type I secretion system is also involved in exporting arabinose, hemolysin, and drugs from the cell. The type II secretion system functions via Sec, Tat, and Pul export pathways (3, 52), and the type II secretion system genes are usually conserved in gram-negative bacteria (7). The Sec genes of CNB-2 were identified and similar to those of *E. coli*, but they were not clustered, except for the SecD, SecE, and YajC genes (CtCNB1_4737 to CtCNB1_4739), which were organized consecutively and encoded a putative protein complex. The Tat genes, *tatABC* (CtCNB1_0735 to CtCNB1_0737), were located in a putative operon, but a *tatE* gene was not found. As suggested by previous studies (41), the function of TatE might be replaced by TatA. A gene cluster (CtCNB1_0617 to CtCNB1_0634) that encoded proteins similar to proteins of *Klebsiella oxytoca* (38) was identified in the CNB-2 genome.

Signal transduction, regulation, and responses to environmental changes. In order to respond to environmental changes in a quick and efficient way, environmental bacteria have evolved various signal-sensing and transduction systems (36, 44). According to COG and Conserved Domain Database analyses (47), 300 ORFs in the CNB-2 genome were assigned to functions for signal transduction and regulation. Furthermore, 58 ORFs were predicted to encode putative response regulators, 46 ORFs were predicted to encode putative histidine kinases, and 10 ORFs were predicted to encode putative hybrid sensory kinases. These putative regulators and kinases were further grouped into 43 complete two-component signal transduction systems (see Table S2 in the supplemental material). Moreover, 64 ORFs that contained one output domain (50) and encoded proteins that putatively functioned as one-component signal transduction systems were identified in the CNB-2 genome (see Table S3 in the supplemental material). The high number of signal transduction and regulation genes in the genome of CNB-2 is comparable to the number in the *P. aeruginosa* PAO1 genome (45) and is higher than the number in the *E. coli* genome (32).

The comprehensive and diverse signal transduction and reg-

ulation systems enable strain CNB-2 respond to various environmental changes. This conclusion is supported by analysis of functional domains of the signal transduction systems. Various input (signal-sensing) domains, such as HAMP, PAS, KdpD, and CHASE, of one- and two-component systems were discovered (see Tables S3 and S4 in the supplemental material) (8). In particular, the PAS domain (55) was frequently identified (see Tables S2 and S3 in the supplemental material), and three PAS domains were found in the putative CtCNB1_4372 protein (see Table S3 in the supplemental material). The CtCNB1_4372 protein, a putative one-component signal transduction system with a putative ammonium monooxygenase neighbor, might be involved in the response to changing oxygen concentrations and in regulation of ammonium assimilation by strain CNB-2.

Chemotaxis is an important microbial response to environmental signals, and chemotactic responses to aromatic compounds have been characterized for *Pseudomonas* and *Ralstonia* species (11, 34). Our results showed that strain CNB-2 exhibited chemotaxis toward succinate, while strain CNB-1 exhibited chemotaxis toward both succinate and CNB. Although the putative genes linking the sensing of chemical signals and cell movement could not be identified at this stage, some related genes were identified in the CNB-2 genome. One locus (CtCNB1_0471 to CtCNB1_0475) encoding three putative regulators and one putative kinase localized between the genes for flagellar biosynthesis and chemotaxis, and ORFs encoding putative proteins (CtCNB1_0475, CtCNB1_0476, and CtCNB1_0479) that were homologous to the CheA, CheY, and CheB proteins were found in the CNB-2 genome.

Drug and heavy metal resistance. Several loci of the CNB-2 genome were predicted to be responsible for drug and heavy metal export (see Table S1 in the supplemental material). Three gene clusters (CtCNB1_2506 to CtCNB1_2509, CtCNB1_3138 to CtCNB1_3141, and CtCNB1_3971 to CtCNB1_3974) encode the system for exporting arsenate. Analysis of the CNB-2 genome revealed that this strain has six CzcA-type heavy metal efflux pumps and 11 RND-type efflux transporters for various drugs (Fig. 2; see Table S2 in the supplemental material). The genes encoding these pumps and transporters possibly conferred additional resistance to heavy metals and drugs to strain CNB-2. Moreover, strain CNB-2 has many chaperones that help the cell overcome toxicity. Toxic drugs and heavy metals could induce heat shock proteins (42, 49), as genes encoding heat shock proteins (HslJ [CtCNB1_3317], IbpA [CtCNB1_3446], HslU [CtCNB1_4124], and HslJ [CtCNB1_4571]) were identified in the genome of strain CNB-2 (see Table S4 in the supplemental material). Interestingly, the drug and heavy metal resistance genes are associated with MGEs. For example, four gene clusters (CtCNB1_2496 to CtCNB1_2499, CtCNB1_2507 to CtCNB1_2510, CtCNB1_2513 to CtCNB1_2520, and CtCNB1_2521 to CtCNB1_2524) encoding P-type ATPase, the CzcA family efflux pump, and arsenate resistance are flanked by genes encoding Tn3-type transposases, IS116-IS110-IS902-type transposases, and IS4-type transposases, respectively.

Responses to the availability of an electron acceptor and rhizosphere association with plants. Strain CNB-2 is strictly aerobic and requires oxygen molecules as the final electron acceptor. Genes encoding all the aerobic respiratory chain

members were tentatively identified in the CNB-2 genome. Four putative terminal oxidase complexes, encoded by CtCNB1_1811 to CtCNB1_1814, CtCNB1_2157 to CtCNB1_2161, CtCNB1_3937 to CtCNB1_3941, and CtCNB1_3431 to CtCNB1_3432, that were highly similar to *cbb3*-type, *bo3*-type, *aa3*-type, and *bd*-type complexes (20) were found. It was reported previously that *cbb3*-type and *bd*-type terminal oxidases functioned at lower oxygen concentrations in endosymbiotic rhizobia (37) and under microaerobic conditions in nitrogen-fixing bacteria (15, 22) and in association with animals (21, 43). Thus, the robust competitiveness of strain CNB-1 in the rhizosphere and wastewater treatment plants may be attributed to the presence of multiple terminal oxidases (25). Genome analysis suggested a putative transhydrogenase (CtCNB1_0257 to CtCNB1_0258) that might regulate energy metabolism in strain CNB-2 at different oxygen concentrations.

Our physiological experiment showed that CNB-2 could respire with nitrate aerobically to produce energy. Accordingly, a gene cluster consisting of five genes (*napABCDE*) that is responsible for the first step of aerobically denitrification (9) and encodes putative periplasmic nitrate reductases was identified. Thus, strain CNB-2 has multiple terminal oxidases depending on the oxygen concentration and is able to use nitrate as an electron acceptor.

Conclusion. *C. testosteroni* CNB-2 is the first member of the genus *Comamonas* that has been completely sequenced. Its genome size is close to the genome sizes previously reported for *Pseudomonas* strains, including *P. putida* GB-1 (6.1 Mb) (accession no. CP000926), *P. putida* KT2440 (6.2 Mb) (AE015451), and *Pseudomonas stutzeri* A1501 (4.6 Mb) (CP000304). *Comamonas* species have been found to be ubiquitous in terrestrial and aquatic environments and to be associated with plants and animals. They have also been found in severely polluted environments. The diverse niches occupied by *Comamonas* species reflect their effective adaptation to various physiochemical conditions. In this study, the high genetic enrichment and capacity of *C. testosteroni* CNB-2 revealed its evolutionary adaptations and genetic foundation for living in diverse and complex ecological niches. The first description of a *C. testosteroni* genome provided in this paper should significantly facilitate further basic and applied investigations of *Comamonas* species.

ACKNOWLEDGMENTS

This work was supported by grants 30725001 and 30621005 from the National Natural Science Foundation of China.

REFERENCES

1. Arai, H., T. Ohishi, M. Y. Chang, and T. Kudo. 2000. Arrangement and regulation of the genes for meta-pathway enzymes required for degradation of phenol in *Comamonas testosteroni* TA441. *Microbiology* **146**:1707–1715.
2. Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783–795.
3. Berks, B. C., T. Palmer, and F. Sargent. 2003. The Tat protein translocation pathway and its role in microbial physiology. *Adv. Microb. Physiol.* **47**:187–254.
4. Chou, J. H., S. Y. Sheu, K. Y. Lin, W. M. Chen, A. B. Arun, and C. C. Young. 2007. *Comamonas odontotermitis* sp. nov., isolated from the gut of the termite *Odontotermes formosanus*. *Int. J. Syst. Evol. Microbiol.* **57**:887–891.
5. De Costa, D. M., K. Suzuki, M. Satou, and K. Yoshida. 2001. Genome analysis of *Agrobacterium tumefaciens*: linkage map and genetic features of the left region of the linear chromosome. *Genes Genet. Syst.* **76**:363–371.
6. Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
7. Filloux, A., G. Michel, and M. Bally. 1998. GSP-dependent protein secretion in Gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS Microbiol. Rev.* **22**:177–198.
8. Galperin, M. Y., A. N. Nikolskaya, and E. V. Koonin. 2001. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* **203**:11–21.
9. Gates, A. J., R. O. Hughes, S. R. Sharp, P. D. Millington, A. Nilavongse, J. A. Cole, E. R. Leach, B. Jepson, D. J. Richardson, and C. S. Butler. 2003. Properties of the periplasmic nitrate reductases from *Paracoccus pantotrophus* and *Escherichia coli* after growth in tungsten-supplemented media. *FEMS Microbiol. Lett.* **220**:261–269.
10. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
11. Grimm, A. C., and C. S. Harwood. 1997. Chemotaxis of *Pseudomonas* spp. to the polyaromatic hydrocarbon naphthalene. *Appl. Environ. Microbiol.* **63**:4111–4115.
12. Gul, M., P. Ciragil, E. Bulbuloglu, M. Aral, S. Alkis, and F. Ezberci. 2007. *Comamonas testosteroni* bacteremia in a patient with perforated acute appendicitis. Short communication. *Acta Microbiol. Immunol. Hung.* **54**:317–321.
13. Gumaelius, L., G. Magnusson, B. Pettersson, and G. Dalhammar. 2001. *Comamonas denitrificans* sp. nov., an efficient denitrifying bacterium isolated from activated sludge. *Int. J. Syst. Evol. Microbiol.* **51**:999–1006.
14. Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen, and O. White. 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**:41–43.
15. Hill, S., S. Violette, A. T. Smith, and C. Anthony. 1990. Roles for enteric d-type cytochrome oxidase in N_2 fixation and microaerobiosis. *J. Bacteriol.* **172**:2071–2078.
16. Horinouchi, M., T. Hayashi, H. Koshino, T. Kurita, and T. Kudo. 2005. Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostane-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in *Comamonas testosteroni* TA441. *Appl. Environ. Microbiol.* **71**:5275–5281.
17. Horinouchi, M., T. Hayashi, T. Yamamoto, and T. Kudo. 2003. A new bacterial steroid degradation gene cluster in *Comamonas testosteroni* TA441 which consists of aromatic-compound degradation genes for seco-steroids and 3-ketosteroid dehydrogenase genes. *Appl. Environ. Microbiol.* **69**:4421–4430.
18. Horinouchi, M., T. Kurita, T. Yamamoto, E. Hatori, T. Hayashi, and T. Kudo. 2004. Steroid degradation gene cluster of *Comamonas testosteroni* consisting of 18 putative genes from meta-cleavage enzyme gene *tesB* to regulator gene *tesR*. *Biochem. Biophys. Res. Commun.* **324**:597–604.
19. Jimenez, J. I., B. Minambres, J. L. Garcia, and E. Diaz. 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**:824–841.
20. Jimenez de Bagues, M. P., S. Loisel-Meyer, J. P. Liautaud, and V. Jubier-Maurin. 2007. Different roles of the two high-oxygen-affinity terminal oxidases of *Brucella suis*: cytochrome *c* oxidase, but not ubiquinol oxidase, is required for persistence in mice. *Infect. Immun.* **75**:531–535.
21. Kaminski, P. A., C. L. Kitts, Z. Zimmermann, and R. A. Ludwig. 1996. *Azorhizobium caulinodans* uses both cytochrome *bd* (quinol) and cytochrome *cbb₃* (cytochrome *c*) terminal oxidases for symbiotic N_2 fixation. *J. Bacteriol.* **178**:5989–5994.
22. Kelly, M. J., R. K. Poole, M. G. Yates, and C. Kennedy. 1990. Cloning and mutagenesis of genes encoding the cytochrome *bd* terminal oxidase complex in *Azotobacter vinelandii*: mutants deficient in the cytochrome *d* complex are unable to fix nitrogen in air. *J. Bacteriol.* **172**:6010–6019.
23. Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
24. Lewenza, S., J. L. Gardy, F. S. Brinkman, and R. E. Hancock. 2005. Genome-wide identification of *Pseudomonas aeruginosa* exported proteins using a consensus computational strategy combined with a laboratory-based PhoA fusion screen. *Genome Res.* **15**:321–329.
25. Liu, L., C. Y. Jiang, X. Y. Liu, J. F. Wu, J. G. Han, and S. J. Liu. 2007. Plant-microbe association for rhizoremediation of chloronitroaromatic pollutants with *Comamonas* sp. strain CNB-1. *Environ. Microbiol.* **9**:465–473.
26. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
27. Lukashin, A. V., and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
28. Ma, Y. F., J. F. Wu, S. Y. Wang, C. Y. Jiang, Y. Zhang, S. W. Qi, L. Liu, G. P. Zhao, and S. J. Liu. 2007. Nucleotide sequence of plasmid pCNB1 from *Comamonas* strain CNB-1 reveals novel genetic organization and evolution for 4-chloronitrobenzene degradation. *Appl. Environ. Microbiol.* **73**:4477–4483.
29. Marchler-Bauer, A., J. B. Anderson, M. K. Derbyshire, C. DeWeese-Scott, N. R. Gonzales, M. Gwadz, L. Hao, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, D. Krylov, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, N. Thanki, R. A. Yamashita, J. J.

- Yin, D. Zhang, and S. H. Bryant. 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **35**:D237–D240.
30. Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
 31. McLeod, M. P., R. L. Warren, W. W. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. Jones, R. Holt, F. S. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn, and L. D. Eltis. 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. USA* **103**:15582–15587.
 32. Mizuno, T. 1997. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.* **4**:161–168.
 33. Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**:29–34.
 34. Pandey, G., A. Chauhan, S. K. Samanta, and R. K. Jain. 2002. Chemotaxis of a *Ralstonia* sp. SJ98 toward co-metabolizable nitroaromatic compounds. *Biochem. Biophys. Res. Commun.* **299**:404–409.
 35. Parche, S., H. Nothhaft, A. Kamionka, and F. Titgemeyer. 2000. Sugar uptake and utilisation in *Streptomyces coelicolor*: a PTS view to the genome. *Antonie van Leeuwenhoek* **78**:243–251.
 36. Parkinson, J. S. 1993. Signal transduction schemes of bacteria. *Cell* **73**:857–871.
 37. Preisig, O., R. Zufferey, L. Thony-Meyer, C. A. Appleby, and H. Hennecke. 1996. A high-affinity *cbb₃*-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J. Bacteriol.* **178**:1532–1538.
 38. Pugsley, A. P., and O. Possot. 1993. The general secretory pathway of *Klebsiella oxytoca*: no evidence for relocation or assembly of pilin-like PulG protein into a multiprotein complex. *Mol. Microbiol.* **10**:665–674.
 39. Saier, M. H., Jr. 1998. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. Academic Press, San Diego, CA.
 40. Saier, M. H., Jr. 2006. Protein secretion and membrane insertion systems in Gram-negative bacteria. *J. Membr. Biol.* **214**:75–90.
 41. Sargent, F., N. R. Stanley, B. C. Berks, and T. Palmer. 1999. Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. *J. Biol. Chem.* **274**:36073–36082.
 42. Segura, A., P. Godoy, P. van Dillewijn, A. Hurtado, N. Arroyo, S. Santacruz, and J. L. Ramos. 2005. Proteomic analysis reveals the participation of energy- and stress-related proteins in the response of *Pseudomonas putida* DOT-T1E to toluene. *J. Bacteriol.* **187**:5937–5945.
 43. Smith, M. A., M. Finel, V. Korolik, and G. L. Mendz. 2000. Characteristics of the aerobic respiratory chains of the microaerophiles *Campylobacter jejuni* and *Helicobacter pylori*. *Arch. Microbiol.* **174**:1–10.
 44. Stock, J. B., A. M. Stock, and J. M. Mottonen. 1990. Signal transduction in bacteria. *Nature* **344**:395–400.
 45. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warren, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**:959–964.
 46. Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**:41.
 47. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
 48. Tettelin, H., D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**:500–507.
 49. Toes, A. C., M. H. Daleke, J. G. Kuenen, and G. Muyzer. 2008. Expression of *copA* and *cusA* in *Shewanella* during copper stress. *Microbiology* **154**:2709–2718.
 50. Ulrich, L. E., E. V. Koonin, and I. B. Zhulin. 2005. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* **13**:52–56.
 51. Vissa, V. D., and P. J. Brennan. 2001. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol.* **2**:REVIEWS102.
 52. Voulhoux, R., G. Ball, B. Ize, M. L. Vasil, A. Lazdunski, L. F. Wu, and A. Filloux. 2001. Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J.* **20**:6735–6741.
 53. Weinle, C., K. E. Nelson, and B. Tummeler. 2002. Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ. Microbiol.* **4**:809–818.
 54. Wu, J. F., C. Y. Jiang, B. J. Wang, Y. F. Ma, Z. P. Liu, and S. J. Liu. 2006. Novel partial reductive pathway for 4-chloronitrobenzene and nitrobenzene degradation in *Comamonas* sp. strain CNB-1. *Appl. Environ. Microbiol.* **72**:1759–1765.
 55. Zhulin, I. B., B. L. Taylor, and R. Dixon. 1997. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**:331–333.