CrossMark
click for updates

# Complete Genome Sequence of *Rhodococcus* sp. Strain IcdP1 Shows Diverse Catabolic Potential

Jie Qu,[a,b] Li-Li Miao,[a] Ying Liu,[a] Zhi-Pei Liu[a]

State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China[a]; University of Chinese Academy of Sciences, Beijing, China[b]

The complete genome sequence of *Rhodococcus* sp. strain IcdP1 is presented here. This organism was shown to degrade a broad range of high-molecular-weight polycyclic aromatic hydrocarbons and organochlorine pesticides. The sequence data can be used to predict genes for xenobiotic biodegradation and metabolism.

Address correspondence to Zhi-Pei Liu, liuzhp@sun.im.ac.cn.

*Rhodococcus* is a metabolically versatile genus with remarkable ability to catabolize a wide range of organic compounds (1). Several *Rhodococcus* strains have been isolated from various contaminated environments (2–6). *Rhodococcus* sp. strain IcdP1 was isolated from contaminated soil of the abandoned Beijing Coking Plant in China. It can degrade numerous high-molecular-weight polycyclic aromatic hydrocarbons (HMW-PAHs), such as benzo-[*a*]pyrene and indeno[1,2,3-*cd*]pyrene, and organochlorine pesticides (OCPs), including hexachlorocyclohexanes and 1,1,1-trichloro-2,2-bis(4-chlorophenyl) ethane. It is useful in the bioremediation of contaminated environments.

Whole-genome shotgun (WGS) DNA sequencing was performed by Majorbio (Shanghai, China) using paired-end sequencing with HiSeq 2000 (Illumina) and long sequencing with PacBio RS (Pacific Biosciences). Three Illumina libraries, two paired-end libraries (300 bp and 800 bp) and a 2-kb mate pair library were constructed and sequenced, giving approximately 219-fold genome coverage (BioProject numbers SRX997125, SRX1009942, and SRX1009943, respectively). An approximately 8- to 10-kb-insert PacBio library was constructed and sequenced on an 8 single-molecule real-time (SMRT) cell, yielding 119,255 reads (average length, 7,446 bp) and >150× genome coverage (BioProject number SRX1009944). *De novo* assembly of the Illumina and PacBio reads was carried out using a novel correction algorithm and assembly strategy (7). The Illumina sequences were assembled into 235 contigs, with an $N_{50}$ length of 74,570 bp by SOAP*denovo* version 2.04 (8), and the contigs were joined into 18 scaffolds. The error inherent in long single-molecule sequences was corrected using BLASR, and then the PacBio-corrected reads (PBcR) were *de novo* assembled into scaffolds alone by overlapping sequencing reads and subsequently assembled with Celera Assembler 8.0 (7).

The genome was analyzed by Rapid Annotations using Subsystems Technology (RAST) (9). The genome size was 5,922,928 bp, comprising 5,480 open reading frames and 430 subsystems. The G+C content was 70.62%. Fifty-four tRNAs were identified by tRNAscan-SE version 1.3.1 (10), and 12 rRNAs were predicted by Barrnap 0.4.2. RAST annotation indicates that *Rhodococcus opacus*

B4 (score, 515) and *Rhodococcus jostii* RHA1 (score, 485) are its closest neighbors. The subsystem categories representing the metabolisms of carbohydrates; amino acids and derivatives; and cofactors, vitamins, prosthetic groups, and pigments are the most abundant, as they account for 631, 622, and 451 proteins, respectively. Three hundred eighty-eight open reading frames (ORFs) are involved in the metabolism of fatty acids, lipids, and isoprenoids, while 145 ORFs are involved in the metabolism of aromatic compounds. In addition, the genome contains at least 50 dioxygenase genes and 20 putative cytochrome P450 genes, which enables it to initiate the oxidation of persistent organic pollutants. Analysis of the genome of IcdP1 will help us to identify diverse biodegradation genes and, further, to elucidate the degradation pathways for HMW-PAHs and OCPs.

**Nucleotide sequence accession number.** The complete genome sequence of *Rhodococcus* sp. strain IcdP1 has been assigned GenBank accession no. CP011341.

## ACKNOWLEDGMENTS

## REFERENCES

1. **Van der Geize R, Dijkhuizen L**. 2004. Harnessing the catabolic diversity of *rhodococci* for environmental and biotechnological applications. Curr Opin Microbiol **7**:255–261. http://dx.doi.org/10.1016/j.mib.2004.04.001.
2. **Di Gennaro P, Rescalli E, Galli E, Sello G, Bestetti G**. 2001. Characterization of *Rhodococcus opacus* R7, a strain able to degrade naphthalene and *o*-xylene isolated from a polycyclic aromatic hydrocarbon-contaminated soil. Res Microbiol **152**:641–651. http://dx.doi.org/10.1016/S0923-2508(01)01243-8.
3. **Grund E, Denecke B, Eichenlaub R**. 1992. Naphthalene degradation via salicylate and gentisate by *Rhodococcus* sp. strain B4. Appl Environ Microbiol **58**:1874–1877.
4. **Kimura N, Kitagawa W, Mori T, Nakashima N, Tamura T, Kamagata Y**. 2006. Genetic and biochemical characterization of the dioxygenase involved in lateral dioxygenation of dibenzofuran from *Rhodococcus opacus*

strain SAO101. Appl Microbiol Biotechnol **73**:474–484. http://dx.doi.org/10.1007/s00253-006-0481-8.

5. **Liu L, Schmid RD, Urlacher VB**. 2006. Cloning, expression, and characterization of a self-sufficient cytochrome P450 monooxygenase from *Rhodococcus ruber* DSM 44319. Appl Microbiol Biotechnol **72**:876–882. http://dx.doi.org/10.1007/s00253-006-0355-0.

6. **Song X, Xu Y, Li G, Zhang Y, Huang T, Hu Z**. 2011. Isolation, characterization of *Rhodococcus* sp. P14 capable of degrading high-molecular-weight polycyclic aromatic hydrocarbons and aliphatic hydrocarbons. Mar Pollut Bull **62**:2122–2128. http://dx.doi.org/10.1016/j.marpolbul.2011.07.013.

7. **Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM**. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nat Biotechnol **30**:693–700. http://dx.doi.org/10.1038/nbt.2280.

8. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J**. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res **20**:265–272. http://dx.doi.org/10.1101/gr.097261.109.

9. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.

10. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**:955–964. http://dx.doi.org/10.1093/nar/25.5.0955.