

# Draft Genome Sequence of *Rhodococcus opacus* Strain M213 Shows a Diverse Catabolic Potential

Ashish Pathak,<sup>a</sup> Stefan J. Green,<sup>b,c</sup> Andrew Ogram,<sup>d</sup> Ashvini Chauhan<sup>a</sup>

School of the Environment, Florida A&M University, Tallahassee, Florida, USA<sup>a</sup>; DNA Services Facility, University of Illinois at Chicago, Chicago, Illinois, USA<sup>b</sup>; Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois, USA<sup>c</sup>; Soil and Water Science Department, University of Florida, Gainesville, Florida, USA<sup>d</sup>

A.P. and S.J.G. contributed equally to this article.

**Soil-borne Gram-positive bacteria from the genus *Rhodococcus* metabolize a range of aromatic hydrocarbons and also produce a variety of value-added products, such as triacylglycerols and steroids. We report the draft genome sequence of *Rhodococcus opacus* strain M213 (9,193,504 bp with a G+C content of 66.99%), providing a comprehensive understanding of the repertoire of metabolic genes of this strain.**

Received 10 December 2012 Accepted 21 December 2012 Published 14 February 2013

Citation Pathak A, Green SJ, Ogram A, Chauhan A. 2013. Draft genome sequence of *Rhodococcus opacus* strain M213 shows a diverse catabolic potential. *Genome Announc.* 1(1):e00144-12. doi:10.1128/genomeA.00144-12.

Copyright © 2013 Pathak et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Ashvini Chauhan, ashvini.chauhan@famou.edu.

Much of what is known about the biochemistry and genetics of naphthalene (NAP) metabolism is based on *Pseudomonas putida* G7 and related gammaproteobacteria (1, 2). The catabolic pathways of soil *Actinobacteria*, such as the rhodococci, are not homologous to those of the pseudomonads and hence remain poorly understood (3–5). We are interested in NAP degradation by *Rhodococcus opacus* strain M213, which was isolated from a fuel oil-contaminated soil sample (6). Previously described NAP degradative pathways generate salicylate (SAL) as a metabolic intermediate (1, 2, 7). Several lines of evidence suggest that *R. opacus* M213 encodes an alternate pathway, in which *o*-phthalate is generated as a key metabolic intermediate during growth on NAP (6, 8).

To understand fully the metabolic potential of strain M213, genomic DNA was prepared for shotgun sequencing using the Nextera kit (Epicenter, Madison, WI), with size selection (400- to 800-bp fragments) performed using a Pippin Prep automated electrophoresis instrument (Sage Scientific, Beverly, MA) and sequenced using 100-base paired-end sequencing on an Illumina HiSeq 2000 system. Approximately 87 M reads were generated in pairs and assembled by the *de novo* assembler within the software package CLC Genomics Workbench v5.0 (CLCbio, Cambridge, MA). A total of 483 contigs of length  $\geq 200$  bases were generated, with a sum of  $\sim 9.2$  Mb, an  $N_{50}$  of 79,111 bases, and an average coverage of  $>800\times$ . Note that 95% of the sequence data assembled were present in the 158 largest contigs ( $N_{95}$ , 9.4 kb), and a total of 350 contigs of  $\geq 500$  bases were assembled (99.55% of total assembly). Contigs were successfully used for annotation and gene prediction by Integrated Microbial Genomes (IMG) Expert Review ER (9) using Prodigal (10), which compares the translated proteins with the nonredundant proteins database (NR) at GenBank, Pfam (11), TIGRFam (12), InterPro (13), Kyoto Encyclopedia of Genes and Genomes (KEGG) (12) and Clusters of Orthologous Groups (COG) (14) databases using BLASTp and

HMMER. The genome was also analyzed by Rapid Annotations using Subsystems Technology (RAST) (15).

Heat map analysis and principal components analysis (PCA) using IMG ER showed that M213 significantly differs from the other eight rhodococci for which genome sequences are available. Strain M213 contains a total of 8,942 putative genes, with 75.14% of genes associated with protein-coding functions, whereas 24.06% genes have unknown functions. Approximately 22% of the protein-coding genes were connected to KEGG pathways, with 401 genes involved in the metabolism of polycyclic aromatic hydrocarbons (PAHs) (naphthalene, phenanthrene, anthracene, and benzo[a]pyrene) and an array of halogenated aromatics and aromatic hydrocarbons, including pesticides (dichlorodiphenyl-trichloroethane [DDT] and atrazine). With respect to NAP, genes similar to those involved in the oxidation of both salicylate and *o*-phthalate were detected, suggesting the possibility of dual pathways for NAP degradation in strain M213. Also present were 284 genes for the biosynthesis of terpenoids, polyketides, and other secondary metabolites, such as caffeine, flavonoids, indole, isoquinoline, and alkaloids, along with 300 genes associated with unsaturated and saturated fatty acid biosynthesis and metabolism; this makes strain M213 a lucrative candidate for the industrial production of biofuel precursors and steroids.

**Nucleotide sequence accession numbers.** This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AJYC00000000](https://www.ncbi.nlm.nih.gov/nuccore/AJYC00000000). The version described in this article is the second version, [AJYC02000000](https://www.ncbi.nlm.nih.gov/nuccore/AJYC02000000).

## ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Defense (DoD) grants W911NF-10-1-0146 and W911NF-10-R-0006.

## REFERENCES

1. Goyal AK, Zylstra GJ. 1997. Genetics of naphthalene and phenanthrene degradation by *Comamonas testosteroni*. *J. Ind. Microbiol. Biotechnol.* 19:401–407.

2. Habe H, Omori T. 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Biosci. Biotechnol. Biochem.* 67: 225–243.
3. Kulakov LA, Allen CC, Lipscomb DA, Larkin MJ. 2000. Cloning and characterization of a novel cis-naphthalene dihydrodiol dehydrogenase gene (*narB*) from *Rhodococcus* sp. NCIMB12038. *FEMS Microbiol. Lett.* 182:327–331.
4. Kulakov LA, Delcroix VA, Larkin MJ, Ksenzenko VN, Kulakova AN. 1998. Cloning of new *Rhodococcus* extradiol dioxygenase genes and study of their distribution in different *Rhodococcus* strains. *Microbiology* 144: 955–963.
5. Larkin MJ, Allen CC, Kulakov LA, Lipscomb DA. 1999. Purification and characterization of a novel naphthalene dioxygenase from *Rhodococcus* sp. strain NCIMB12038. *J. Bacteriol.* 181:6200–6204.
6. Uz I, Duan YP, Ogram A. 2000. Characterization of the naphthalene-degrading bacterium, *Rhodococcus opacus* M213. *FEMS Microbiol. Lett.* 185:231–238.
7. Grund E, Denecke B, Eichenlaub R. 1992. Naphthalene degradation via salicylate and gentisate by *Rhodococcus* sp. Strain B4. *Appl. Environ. Microbiol.* 58:1874–1877.
8. Pathak A, Cooper B, Wafula D, Ogram A, Chauhan A. 2011. Evidence that *Rhodococcus opacus* strain M213 degrades naphthalene via a previously undescribed metabolic pathway. Strategic Environmental Research and Development Program (SERDP) Symposium, Washington, DC.
9. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG er: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271–2278.
10. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
11. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. 2008. The Pfam protein families database. *Nucleic Acids Res.* 36:D281–D288.
12. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O. 2007. TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.* 35:D260–D264.
13. Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.
14. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
15. 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.