

Address for correspondence: Susanna Esposito, Pediatric Clinic 1, Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via Commenda 9, 20122 Milan, Italy; email: susanna.esposito@unimi.it

## Genome Sequencing of Pathogenic *Rhodococcus* spp.

**To the Editor:** Infections caused by non-*equi Rhodococcus* spp. are uncommon but can cause severe pneumonia and bloodstream infections with sepsis (1–4). Increasing prevalence of immune-compromising illnesses and use of immunosuppressive agents might contribute to re-emergence of these pathogens (5,6). *Rhodococcus* infections may go undiagnosed or misclassified because of difficulties in laboratory identification, nomenclatural instability, and similarity of signs and symptoms to *Mycobacterium tuberculosis* infection (e.g., increasing cough, dyspnea, hemoptysis, and weight loss; failure to respond to broad-spectrum antimicrobial drugs; progressive cavitory lesion on repeated chest imaging).

A 73-year-old immunocompetent man with chronic obstructive pulmonary disease who lived on a ranch in the southwestern United States was evaluated for possible malignancy of the left lung. Computed tomography showed expansion of an upper lobe cavitory lesion, and he underwent a partial lobectomy. *Rhodococcus* spp. infection was suspected on the basis of the finding of the progressive lesion, salmon-pink colony growth on chocolate agar, and gram-positive

coccobacilli on Gram stain. The isolate, R1101, could not be further identified by using commercial automated systems in the laboratory and so was subjected to genome study.

Alignment of 16S rRNA gene sequences showed that R1101 was most closely related to *R. rhodochrous*, with identities of 98.78%, 99.32%, and 99.86% to the 16S rRNA gene for type strains *R. rhodochrous* DSM43271, *R. pyridinivorans* PDB9, and *R. gordonae* W4937 (7), respectively. In contrast, sequence identity with *R. equi* type strain DSM20307 was 95.57%. Phylogenetic analysis with *Rhodococcus* 16S rRNA genes suggested that R1101 was a member of *R. rhodochrous* and evolutionarily separate from *R. equi* (Figure, panel A). A query against the National Center for Biotechnology Information nonredundant DNA database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) found 60 rhodococcal 16S rRNA sequences that are 99%–100% identical to R1101, which demonstrates worldwide distribution of these closely related strains in diverse environments.

Optical genome mapping (8) of R1101 generated a consensus full-length map showing a circular chromosome map of »4.32 Mb ( $\pm$  5%) in length. The whole-genome restriction map for R1101 was compared with in silico *NcoI* restriction maps of 33 complete genome sequences from the family *Nocardiaceae*, retrieved from GenBank; these sequences included *R. equi*. No substantial similarity (alignment scores >15%) was detected between R1101 and these genomes. Thus, we determined that R1101 was not an *R. equi* isolate.

Pyrosequencing of the R1101 genome yielded an average coverage depth of 14-fold. The total length of 1,020 assembled de novo contigs was 4.65 Mb, with a G + C content of 68%. A total of 3,969 putative protein-encoding genes were identified. Whole-genome phylogeny was used to show evolutionary distances between R1101 and the sequenced *Rhodococcus* species (Figure, panel B). We sampled  $\approx$ 50.4 kb of sequences, encoding 67 putative proteins shared among all selected rhodococci, to generate the

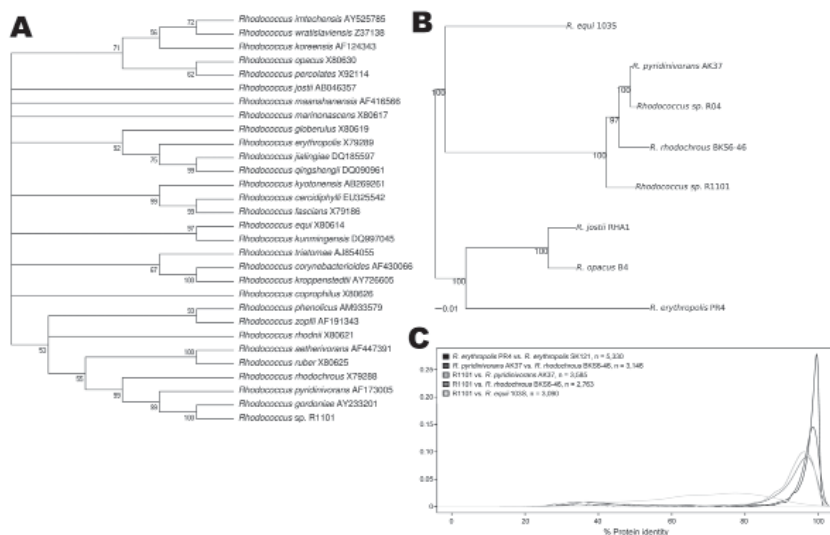


Figure. Phylogenetic and proteomic comparison of isolate R1101 and *Rhodococcus* spp. strains. A) 16S rRNA-based phylogenetic analysis. Complete or partial 16S rRNA gene sequences for 29 *Rhodococcus* spp. type strains were aligned with complete sequence of 16S rRNA gene for isolate R1101. GenBank accession numbers are provided. B) Genome-wide phylogenetic analysis. A sample of 67 protein-coding genes shared among genomes of isolate R1101 and 6 *Rhodococcus* spp. strains were analyzed. Scale bar indicates 0.01 nt substitutions per site. An expanded version of this figure including panel C is available online ([wwwnc.cdc.gov/EID/article/18/11/12-0818-F1.htm](http://wwwnc.cdc.gov/EID/article/18/11/12-0818-F1.htm)).

phylogenetic tree. Results confirmed that R1101 is a member of *R. rhodochrous* and is phylogenetically distant from *R. equi*.

We used a BLASTP search to determine the similarity of predicted R1101 proteins to homologous proteins from *R. equi* 103S, *R. erythropolis* (PR4 and SK121), and *R. pyridinivorans* (AK37 and BKS6-46) (Figure, panel C, Appendix, [wwwnc.cdc.gov/EID/article/18/11/12-0818-F1.htm](http://wwwnc.cdc.gov/EID/article/18/11/12-0818-F1.htm)). These data indicate that R1101 is highly similar to but distinct from *R. pyridinivorans* and *R. rhodochrous* at the protein sequence level and is more distantly related to *R. equi*.

Each R1101 protein  $\geq 1$  database blast hit with an E-value  $\leq 1 \times 10^{-10}$  was assigned a best match to another species on the basis of its bit score. Consistent with the 16S rRNA-based phylogenetic analysis, >92% of the R1101 proteins showed greatest similarity to a protein from *R. pyridinivorans*, *R. rhodochrous*, or *Rhodococcus* sp. R04. However, 120 proteins (3%) have highest homology to their counterparts in the pathogen *R. equi*; 10 of these are unique to R1101 and *R. equi*. Furthermore, genes with the greatest similarity to *R. equi* are not randomly distributed throughout the R1101 genome.

We calculated the probability of observing groups of adjacent R1101 genes that are most similar to *R. equi* and observed 1 cluster of 5 adjacent genes ( $p = 2.33 \times 10^{-8}$ ), 2 groups of 6 adjacent genes ( $p = 6.75 \times 10^{-10}$ ), and 2 blocks of 7 contiguous genes ( $p = 1.94 \times 10^{-11}$ ). Nucleotide BLAST analysis showed that R1101 sequence contigs 604, 456, 139, and 610 align to nt 4454241–4469589 of the *R. equi* 103S chromosome, with >97% identity at the DNA level. The 12 proteins wholly or partially encoded within this 15.3-kb region show >99% identity at the amino acid level with their R1101 orthologs. Among them, contig 139 encodes a cluster of 7 consecutive proteins with high homology to *R. equi*; this portion of the R1101 ge-

nome is likely to have been acquired from *R. equi* through horizontal gene transfer. R1101 may have acquired clusters of virulence-related genes, such as those crucial iron-uptake proteins (9), from the phylogenetically distant species *R. equi*.

In conclusion, infections caused by non-*equi* *Rhodococcus* spp. are rare, especially in immunocompetent patients, but may represent an emerging threat. Specialized diagnostics such as genome sequencing may be needed to accurately identify these pathogens.

#### Acknowledgments

We thank Stephanie Ng, Roscea J Stelsel, and Christopher King for their clinical case study and Fatma Onmus-Leone and Luis J. Martinez for discussions.

**Jun Hang,<sup>1</sup> Robert J. Clifford,<sup>1</sup>  
Yu Yang, Matthew C. Riley,  
Rupal M. Mody,  
Robert A. Kuschner,  
and Emil P. Lesho**

Author affiliations: Walter Reed Army Institute of Research, Silver Spring, Maryland, USA (J. Hang, R.J. Clifford, Y. Yang, M.C. Riley, R.A. Kuschner, E.P. Lesho); and William Beaumont Army Medical Center, El Paso, Texas, USA (R.M. Mody)

DOI: <http://dx.doi.org/10.3201/eid1811.120818>

#### References

1. Spark RP, McNeil MM, Brown JM, Lasker BA, Montano MA, Garfield MD. *Rhodococcus* species fatal infection in an immunocompetent host. *Arch Pathol Lab Med*. 1993;117:515–20.
2. Park SD, Uh Y, Jang IH, Yoon KJ, Kim HM, Bae YJ. *Rhodococcus erythropolis* septicaemia in a patient with acute lymphocytic leukaemia. *J Med Microbiol*. 2011;60:252–5. <http://dx.doi.org/10.1099/jmm.0.015412-0>
3. Baba H, Nada T, Ohkusu K, Ezaki T, Hasegawa Y, Paterson DL. First case of bloodstream infection caused by *Rhodococcus erythropolis*. *J Clin Microbiol*. 2009;47:2667–9. <http://dx.doi.org/10.1128/JCM.00294-09>

<sup>1</sup>These authors contributed equally to this article.

4. Hart DH, Peel MM, Andrew JH, Burdon JG. Lung infection caused by *Rhodococcus*. *Aust N Z J Med*. 1988;18:790–1. <http://dx.doi.org/10.1111/j.1445-5994.1988.tb00182.x>
5. Letek M, González P, Macarthur I, Rodríguez H, Freeman TC, Valero-Rello A, et al. The genome of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet*. 2010;6:e1001145. <http://dx.doi.org/10.1371/journal.pgen.1001145>
6. van der Geize R, Grommen AW, Hessels GI, Jacobs AA, Dijkhuizen L. The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development. *PLoS Pathog*. 2011;7:e1002181. <http://dx.doi.org/10.1371/journal.ppat.1002181>
7. Jones AL, Brown JM, Mishra V, Perry JD, Steigerwalt AG, Goodfellow M. *Rhodococcus gordoniae* sp. nov., an actinomycete isolated from clinical material and phenol-contaminated soil. *Int J Syst Evol Microbiol*. 2004;54:407–11. <http://dx.doi.org/10.1099/ijs.0.02756-0>
8. Riley MC, Kirkup BC Jr, Johnson JD, Lesho EP, Ockenhouse CF. Rapid whole genome optical mapping of *Plasmodium falciparum*. *Malar J*. 2011;10:252. <http://dx.doi.org/10.1186/1475-2875-10-252>
9. De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y, Barry CE III. The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc Natl Acad Sci U S A*. 2000;97:1252–7. <http://dx.doi.org/10.1073/pnas.97.3.1252>

Address for correspondence: Emil P. Lesho, Walter Reed Army Institute of Research, 503 Robert Grant Ave, Silver Spring, MD, 20902, USA; email: [carolinelesho@yahoo.com](mailto:carolinelesho@yahoo.com)

#### Letters

Letters commenting on recent articles as well as letters reporting cases, outbreaks, or original research are welcome. Letters commenting on articles should contain no more than 300 words and 5 references; they are more likely to be published if submitted within 4 weeks of the original article's publication. Letters reporting cases, outbreaks, or original research should contain no more than 800 words and 10 references. They may have 1 Figure or Table and should not be divided into sections. All letters should contain material not previously published and include a word count.