Draft Genome Sequence of *Raoultella terrigena* R1Gly, a Diazotrophic Endophyte

M. Schicklberger,[a] N. Shapiro,[b] D. Loqué,[c] T. Woyke,[b] R. Chakraborty[a]

Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, USA[a]; DOE Joint Genome Institute (JGI), Walnut Creek, California, USA[b]; LBNL DOE Joint BioEnergy Institute (JBEI), Emeryville, California, USA[c]

***Raoultella terrigena* R1Gly is a diazotrophic endophyte isolated from surface-sterilized roots of *Nicotiana tabacum*. The whole-genome sequence was obtained to investigate the endophytic characteristics of this organism at the genetic level, as well as to compare this strain with its close relatives. To our knowledge, this is the first genome obtained from the *Raoultella terrigena* species and only the third genome from the *Raoultella* genus, after *Raoultella ornitholytic* and *Raoultella planticola*. This genome will provide a foundation for further comparative genomic, metagenomic, and functional studies of this genus.**

Address correspondence to R. Chakraborty, rchakraborty@lbl.gov.

*R*aoultella, a genus named after the French bacteriologist Didier Raoult, comprises Gram-negative, oxidase-negative, facultative anaerobic organisms, with both respiratory and fermentative types of metabolism. *Raoultella* belongs to the family *Enterobacteriaceae*, and based on phylogenetic analysis of the 16S rRNA and *rpoB* genes, *Raoultella* was divided from the genus *Klebsiella* (1). We isolated *Raoultella* strain R1Gly after enriching surface-sterilized root samples of *Nicotiana tabacum* in nitrogen-free HGB medium (2), using glycerol as the sole carbon source. This strain grew optimally at 30°C, and its ability to fix atmospheric nitrogen (N) was confirmed by the presence of the *nifH* gene and by the acetylene reduction assay (3). The draft genome sequence revealed a 5.7-Mb genome with 57.84 mol% G+C content, which is comparable to those of the diazotrophs *Raoultella planticola* (5.8 Mb; 55.4 mol%) and *Raoultella ornithinolytica* S12 (5.5 Mb; 57.47 mol%) as well as the nondiazotrophs *Raoultella ornithinolytica* B6 (5.3 Mb; 55.75 mol%) and *Raoultella ornithinolytica* TNT (5.6 Mb; 55.5 mol%) (4–7). The COG predictions categorize 2,062 of the 5,239 protein-encoding genes as pertaining to metabolism and transport, 1,752 to intracellular processes, and 475 to extracellular processes and 229 as having unknown functions.

Interestingly, apart from genes for biological nitrogen fixation, strain R1Gly contains all necessary genes for the tryptophan-dependent production of the plant hormone indole acetic acid (*ipdC*) and the production of 2,3-butanediol and acetoin volatiles previously shown to promote growth in *Arabidopsis* (8).

The genomic DNA was isolated using a cetyltrimethylammonium bromide (CTAB)-based extraction from cells grown overnight at 30°C on LB medium. The draft genome of *Raoultella terrigena* R1Gly was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology (9). A Pacbio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 322,267 filtered subreads totaling 639.3 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads were assembled using HGAP (version: 2.2.0.p1) (10). The final draft assembly contained 10 contigs in 10 scaffolds, totaling 5.7 Mbp in size. The input read coverage was 151.9×.

Genes were identified using Prodigal (11), followed by a round of manual curation using GenePRIMP (12). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database and the UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScan-SE tool (13) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (14). Other noncoding RNAs such as the RNA components of the protein secretion complex and RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERence of RNA ALignment (INFERNAL). Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes (IMG) platform (15) developed by the JGI.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under project accession no. LANE00000000. The version described in this paper is the version LANE00000000.1.

ACKNOWLEDGMENTS

REFERENCES

1. **Drancourt M, Bollet C, Carta A, Rousselier P**. 2001. Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int J Syst Evol Microbiol **51:**925–932. http://dx.doi.org/10.1099/00207713-51-3-925.

2. **Holguin GG, Bashan Y**. 1992. Two new nitrogen-fixing bacteria from the rhizosphere of mangrove trees: their isolation, identification and *in vitro* interaction with rhizosphere *Staphylococcus* sp. FEMS Microbiol Ecol **101:** 207–216.

3. **Elbeltagy A, Nishioka K, Sato T, Suzuki H, Ye B, Hamada T, Isawa T, Mitsui H, Minamisawa K**. 2001. Endophytic colonization and *in planta* nitrogen fixation by a *Herbaspirillum* sp. isolated from wild rice species. Appl Environ Microbiol **67:**5285–5293. http://dx.doi.org/10.1128/AEM.67.11.5285-5293.2001.

4. **Jothikumar N, Kahler A, Strockbine N, Gladney L, Hill VR**. 2014. Draft genome sequence of *Raoultella planticola*, isolated from river water. Genome Announc **2**(5):e01061-14. http://dx.doi.org/10.1128/genomeA.01061-14.

5. **Bao W, Zhou Y, Jiang J, Xu Z, Hou L, Leung FC**. 2015. Complete genome sequence of *Raoultella ornithinolytica* strain S12, a lignin-degrading bacterium isolated from forest soil. Genome Announc **3**(2):e00104-15. http://dx.doi.org/10.1128/genomeA.00104-15.

6. **Shin SH, Um Y, Beak JH, Kim S, Lee S, Oh MK, Kim YR, Lee J, Yang KS**. 2013. Complete genome sequence of *Raoultella ornithinolytica* strain B6, a 2,3-butanediol-producing bacterium isolated from oil-contaminated soil. Genome Announc **1**(3):e00395-13. http://dx.doi.org/10.1128/genomeA.00395-13.

7. **Thijs S, Van Hamme J, Gkorezis P, Rineau F, Weyens N, Vangronsveld J**. 2014. Draft genome sequence of *Raoultella ornithinolytica* TNT, a trinitrotoluene-denitrating and plant growth-promoting strain isolated from explosive-contaminated soil. Genome Announc **2**(3):e00491-14. http://dx.doi.org/10.1128/genomeA.00491-14.

8. **Ryu CM, Farag MA, Hu CH, Reddy MS, Wei HX, Paré PW, Kloepper JW**. 2003. Bacterial volatiles promote growth in *Arabidopsis*. Proc Natl Acad Sci U S A **100:**4927–4932. http://dx.doi.org/10.1073/pnas.0730845100.

9. **Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S**. 2009. Real-time DNA sequencing from single polymerase molecules. Science **323:**133–138. http://dx.doi.org/10.1126/science.1162986.

10. **Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.

11. **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. http://dx.doi.org/10.1186/1471-2105-11-119.

12. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC**. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods **7:**455–457. http://dx.doi.org/10.1038/nmeth.1457.

13. **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.

14. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO**. 2007. Silva: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:**7188–7196. http://dx.doi.org/10.1093/nar/gkm864.

15. **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. http://dx.doi.org/10.1093/bioinformatics/btp393.