genomeA
Journals.ASM.org

Complete Genome Sequence of *Streptomyces* sp. Strain CCM_MD2014, Isolated from Topsoil in Woods Hole, Massachusetts

Richard M. Mariita,[a] Srijak Bhatnagar,[b] Kurt Hanselmann,[c] Mohammad J. Hossain,[a] Jonas Korlach,[d] Matthew Boitano,[d] Richard J. Roberts,[e] Mark R. Liles,[a] Anthony G. Moss,[a] Jared R. Leadbetter,[f] Dianne K. Newman,[g,h,i] Scott C. Dawson[j]

Department of Biological Sciences, Auburn University, Alabama, USA[a]; Microbiology Graduate Group, University of California Davis, Davis, California, USA[b]; Department of Earth Sciences, ETH Zurich, Zürich, Switzerland[c]; Pacific Biosciences, Menlo Park, California, USA[d]; New England BioLabs, Ipswich, Massachusetts, USA[e]; Linde Center for Global Environmental Science, California Institute of Technology, Pasadena, California, USA[f]; Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA[g]; Howard Hughes Medical Institute, Pasadena, California, USA[h]; Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA[i]; Department of Microbiology and Molecular Genetics, University of California Davis, Davis, California, USA[j]

R.M.M. and S.B. are co-first authors.

Here, we present the complete genome sequence of *Streptomyces* sp. strain CCM_MD2014 (phylum *Actinobacteria*), isolated from surface soil in Woods Hole, MA. Its single linear chromosome of 8,274,043 bp in length has a 72.13% G+C content and contains 6,948 coding sequences.

Received 30 October 2015 **Accepted** 4 November 2015 **Published** 31 December 2015

Citation Mariita RM, Bhatnagar S, Hanselmann K, Hossain MJ, Korlach J, Boitano M, Roberts RJ, Liles MR, Moss AG, Leadbetter JR, Newman DK, Dawson SC. 2015. Complete genome sequence of *Streptomyces* sp. strain CCM_MD2014, isolated from topsoil in Woods Hole, Massachusetts. Genome Announc 3(6):e01506-15. doi:10.1128/genomeA.01506-15.

Address correspondence to Scott C. Dawson, scdawson@ucdavis.edu.

The genus *Streptomyces* belongs to the phylum *Actinobacteria*, the members of which are Gram-positive bacteria that are ubiquitous in soils and typically have a genome with a high G+C content. They are known for their capacity for secondary metabolite synthesis and expression of novel enzymes (1). The strain in this study, *Streptomyces* sp. CCM_MD2014, was cocultured with *Curtobacterium* sp. strain MR_MD2014 from topsoil around a rusty fire hydrant in Woods Hole, MA (41°31′44.65″N 70°40′21.5″W) on 7 July 2014, using isolation protocols modified from those of El-Nakeeb and Lechevalier (2). This organism was cultivated as part of the 2014 Microbial Diversity Summer Program at the Marine Biological Laboratory in Woods Hole, MA.

DNA was extracted from the coculture using the Promega Wizard genomic DNA purification kit with 1 h of lysozyme digestion. The DNA was quantified using the Promega QuantiFluor double-stranded DNA (dsDNA) system and then size selected for a minimum length of 4 kb. Size-selected DNA was sequenced on a Pacific Biosciences RSII sequencing platform with P5C3 chemistry. The sequenced fragments were assembled using HGAP3 on the SMRT Portal (3). The final assembled genome consisted of a single linear chromosome that was 8,274,043 bp long, with a 72.13% G+C content and sequencing coverage of 89×.

The genome was annotated using NCBI's Prokaryotic Genome Annotation Pipeline version 2.8 (rev. 449627) (4, 5). There were 6,948 coding sequences (CDSs), 6 rRNA operons, and 68 tRNA genes. Neither Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) nor prophages were detected in the genome by CRISPRFinder (6) and PHAST (7), respectively. The antiSMASH pipeline (8) predicted 52 secondary metabolite biosynthetic genetic clusters in the genome, including genes for lantipeptides, terpenes, siderophores, polyketide synthases type I and II, bacteriocins, and nonribosomal peptide synthase genes. REBASE (9) identified 13 candidate methylase genes, and the following methylated motifs were found using Pacific Biosciences SMRT Portal analysis: A^{m6}AGNNNNNNNTCCG, CATCC^{m6}AG, CC^{m6}AGNNNGTCG, GACGA^{m6}AC, and GCCGGC (see organism number 13448 on REBASE website for more details). A multilocus phylogenetic analysis using *rpoB*, *tryB*, *recA*, *gyrB*, and *atpD* and average nucleotide identity (ANI) (10) of 93.8% indicated that Streptomyces coelicoflavus ZGO656 is the closest sequenced genome. A concatenated alignment of these genes was used for phylogenetic reconstruction.

Nucleotide sequence accession number. The complete genome sequence of *Streptomyces* sp. strain CCM_MD2014 is available through GenBank under the accession number CP009754.

ACKNOWLEDGMENTS

The isolation and culturing of this organism were carried out as part of the 2014 Microbial Diversity Summer Program at the Marine Biological Laboratory in Woods Hole, MA. The program was supported by generous grants from the Simons Foundation (grant 309981 to the Marine Biological Laboratory), the Beckman Foundation, and the Howard Hughes Medical Institute. This material is based upon work supported by the National Aeronautics and Space Administration (grant NNA13AA92A), the National Science Foundation (grant MCB-1417876), and the U.S. Department of Energy (grant DE-FG02-85ER13361). R.M.M. was supported by a fellowship from Alabama EPSCoR GRSP, Auburn University's Cell and Molecular Biosciences (CMB) Summer Fellowship, Selman A. Waksman Endowed Scholarship in Microbial Diversity, Bernard Davis Endowed Scholarship (no. 47802012050), an Auburn Graduate School Travel Award, and the Joseph Kirby Farrington Endowed Fund for Excellence. These funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

Promega Corporation donated the molecular reagents used in this project. The sequencing for this organism was supported and carried out

by Pacific Biosciences. We thank George O'Toole, Alison Butler, Emil Ruff, Arpita Bose, Suzanne Kern, Louis Kerr, Kurt Dahlstrom, Alice Michel, and Cristian Salgado for their assistance and advice.

R.J.R. works for New England BioLabs, a company that sells research reagents, including restriction enzymes and DNA methylases, to the scientific community. J.K. and M.B. are full-time employees at Pacific Biosciences, a company that develops single-molecule real-time sequencing technologies, including the sequencing platform used in this project.

REFERENCES

1. **Hodgson EK, Fridovich I**. 1975. The interaction of bovine erythrocyte superoxide dismutase with hydrogen peroxide. Inactivation of the enzyme. Biochemistry **14:**5294–5299. http://dx.doi.org/10.1021/bi00695a010.
2. **El-Nakeeb MA, Lechevalier HA**. 1963. Selective isolation of aerobic actinomycetes. Appl Microbiol **11:**75–77.
3. **Chin C, 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.
4. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O**. 2008. Toward an Online Repository of Standard Operating Procedures (SOPs) for (Meta) genomic annotation. Omics J Integr Biol **12:**137–141. http://dx.doi.org/10.1089/omi.2008.0017.
5. **Tatusova T, DiCuccio MD, Badretdin A, Chetvernin V, Ciufo S, Li W**. 2013. Prokaryotic genome annotation pipeline, the NCBI handbook, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.
6. **Grissa I, Vergnaud G, Pourcel C**. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35:**W52–W57. http://dx.doi.org/10.1093/nar/gkm360.
7. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS**. 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347–W352. http://dx.doi.org/10.1093/nar/gkr485.
8. **Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T**. 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res **41:**W204–W212. http://dx.doi.org/10.1093/nar/gkt449.
9. **Roberts RJ, Vincze T, Posfai J, Macelis D**. 2014. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res **43:**D298–D299.
10. **Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A**. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res **43:**6761–6771. http://dx.doi.org/10.1093/nar/gkv657.