Complete Genome Assembly of *Corynebacterium* sp. Strain ATCC 6931

H. E. Daligault,[a] K. W. Davenport,[a] T. D. Minogue,[b] K. A. Bishop-Lilly,[c,d] D. C. Bruce,[a] P. S. Chain,[a] S. R. Coyne,[b] K. G. Frey,[c,d] J. Jaissle,[b] G. I. Koroleva,[e] J. T. Ladner,[e] P-E. Li,[a] L. Meincke,[a] A. C. Munk,[a] G. F. Palacios,[e] C. L. Redden,[c,d] ⓘD S. L. Johnson[a]

Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USA[a]; Diagnostic Systems Division (DSD), United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland, USA[b]; Naval Medical Research Center (NMRC-Frederick), Fort Detrick, Maryland, USA[c]; Henry M. Jackson Foundation, Bethesda, Maryland, USA[d]; Center for Genome Sciences (CGS), USAMRIID, Fort Detrick, Maryland, USA[e]

The genus *Corynebacterium* is best known for the pathogen *C. diphtheriae*; however, it contains mostly commensal and non-pathogenic, as well as several opportunistic, pathogens. Here, we present the 2.47-Mb scaffolded assembly of the type strain, *Corynebacterium* sp. ATCC 6931 (NCTC 1914), as deposited into GenBank under accession number CP008913.

The genus *Corynebacterium* includes slow-growing Gram-positive rod-shaped chemo-organotrophs. The genus is named for their "knotted rod" (Greek *corönë*) cellular shape and was first described in the 1890s (*Corynebacterium diphtheriae*, the causative agent of diphtheria). Most described species are commensals or soil/freshwater dwelling; however, there are several species known to be opportunistic pathogens, especially of the immunocompromised (1). Here, we present a scaffolded genome assembly of the type strain *Corynebacterium* sp. ATCC 6931 (NCTC 1914).

High-quality genomic DNA was extracted from a purified isolate using QIAgen Genome Tip-500 at USAMRIID-DSD. Specifically, a 100-mL bacterial culture was grown to the stationary phase and nucleic acid was extracted per the manufacturer's recommendations. Sequence data were generated using two separate Illumina technologies (2, 3). We constructed and sequenced two Illumina libraries: standard (unpaired) 100-bp reads (300-fold genome coverage) and a separate long-insert paired-end library (9-fold genome coverage, 7,954 ± 2,596 bp insert). Data from the two libraries were assembled together in Newbler (Roche) and the consensus sequences computationally shredded into 2-kbp overlapping fake reads (shreds). The raw reads were also assembled in Velvet and those consensus sequences computationally shredded into 1.5-kbp overlapping shreds (4). All draft data were then assembled together in ALLPATHS and the consensus sequences were computationally shredded into 10-kbp overlapping shreds (5). We then integrated the Newbler consensus shreds, Velvet consensus shreds, ALLPATHS consensus shreds, and a subset of the long-insert read-pairs using parallel Phrap (High Performance Software). Possible misassemblies were corrected and some gap closure was accomplished with manual editing in Consed (6–8).

Automatic annotation of the *Corynebacterium* sp. ATCC 6931 genome assembly utilized an Ergatis-based workflow at LANL with minor manual curation. The closed, quality-checked, annotated assembly is available in NCBI (CP008913), and the raw data can be provided upon request. Preliminary review of the annotation indicates 26 genes associated with antibiotic and toxic-metal resistance, in addition to several genes associated with a *Mycobacterium* virulence operon (9). Based on a BWA (10) contig classifier and Mega-BLAST (11) to the NCBI nucleotide database, this strain is most closely related to *C. urealyticum* (known to cause antibiotic resistant urinary tract infections) (12); however, further placement efforts are warranted.

Nucleotide sequence accession number. The finished genome assembly for *Corynebacterium* sp. ATCC 6931 has been deposited in GenBank under accession number CP008913.

ACKNOWLEDGMENTS

REFERENCES

1. **Bernard K.** 2012. The genus *Corynebacterium* and other medically relevant coryneform-like bacteria. J. Clin. Microbiol. **50:**3152–3158. http://dx.doi.org/10.1128/JCM.00796-12.
2. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433–438. http://dx.doi.org/10.1517/14622416.5.4.433.
3. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380. http://dx.doi.org/10.1038/nature03959.

4. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.

5. **Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB.** 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. Genome Res. **18**:810–820. http://dx.doi.org/10.1101/gr.7337908.

6. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated sequencer traces using Phred. I: accuracy assessment. Genome Res. **8**:175–185. http://dx.doi.org/10.1101/gr.8.3.175.

7. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred. II: error probabilities. Genome Res. **8**:186–194.

8. **Gordon D, Abajian C, Green P.** 1998. *Consed*: a graphical tool for sequence finishing. Genome Res. **8**:195–202. http://dx.doi.org/10.1101/gr.8.3.195.

9. **Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. **33**:5691–5702. http://dx.doi.org/10.1093/nar/gki866.

10. **Li H.** 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. http://arxiv.org/abs/1303.3997.

11. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

12. **Soriano F, Tauch A.** 2008. Microbiological and clinical features of *Corynebacterium urealyticum*: urinary tract stones and genomics as the Rosetta Stone. Clin. Microbiol. Infect. **14**:632–643. http://dx.doi.org/10.1111/j.1469-0691.2008.02023.x.