Complete Genome Sequence of *Coprothermobacter proteolyticus* DSM 5265

Alexandra Alexiev, David A. Coil, Jonathan H. Badger, Julie Enticknap, Naomi Ward, Frank T. Robb, Jonathan A. Eisen

University of California Davis Genome Center, Davis, California, USA; J. Craig Venter Institute, La Jolla, California, USA; Cumnor House School, Danehill, West Sussex, United Kingdom; University of Wyoming, Laramie, Wyoming, USA; University of Maryland School of Medicine, Baltimore, Maryland, USA; University of California Davis, Department of Evolution and Ecology, Department of Medical Microbiology and Immunology, Davis, California, USA

Here we present the complete 1,424,912-bp genome sequence of *Coprothermobacter proteolyticus* DSM 5265, isolated from a thermophilic digester fermenting tannery wastes and cattle manure.

**Coprothermobacter proteolyticus** is a nonmotile, non-spore-forming, rod-shaped, Gram-negative anaerobic bacterium isolated from a thermophilic consortium fermenting tannery wastes and cattle manure (1). *C. proteolyticus* has increased utilization of fructose, mannose, glucose, maltose, and sucrose with the addition of yeast extract with either rumen fluid or Trypticase. Fractionation of fructose, mannose, glucose, maltose, and sucrose with *C. proteolyticus* was selected in 2002 as part of a National Science Foundation-funded “Assembling the Tree of Life” project at the Institute for Genomic Research (TIGR) to sequence the genomes of representatives of the seven phyla of bacteria that at the time had cultured representatives but no available genome sequence. *C. proteolyticus* DSM 5265 was grown in DSM medium 481, and DNA was extracted using standard techniques. Sanger sequencing and genome assembly were performed as previously described (7). All predicted proteins larger than 30 amino acids were predicted as previously described (11). The 5’ regions of the CDSs were inspected for ribosomal binding sites and transcriptional terminators. Two sets of hidden Markov models were used to determine CDS membership in families and superfamilies: Pfam v11.0 and TIGRFAMs 3.0 (13). Pfam v11.0 hidden Markov models were also used with a constraint of a minimum of two hits to find repeated domains within proteins and mask them. This annotation was submitted with the genome in 2008, but in 2014 we requested an in-place update of the annotation from NCBI, using their integrated PGAP pipeline (14).

**Nucleotide sequence accession numbers.** This genome sequence has been deposited at DDBJ/EMBL/GenBank under the accession no. CP001145. The version described in this paper is version CP001145.1.

**ACKNOWLEDGMENTS**

This work was funded by the National Science Foundation “Assembling the Tree of Life” grant 0228651, overseen by Jonathan A. Eisen and Naomi Ward. Sanger sequencing was performed at the Institute for Genomic Research (TIGR), in Rockville, MD.

We thank the many others who contributed to this project, including Tara Holley, Martin Wu, Liz O’Connor, Hoda Khouri, Kisha Watkins, William Nelson, Claire Fraser, James Saskwa, Jeremy Selengut, Daniel Haft, Jan Weidman, Yasmin Mohamed, Grace Pai, Shannon Smith, Tamara Feldblyum, Terry Utterback, and Mihai Pop.


