Draft Genome Sequence of *Comamonas thiooxydans* Strain S23[T] (DSM 17888[T]), a Thiosulfate-Oxidizing Bacterium Isolated from a Sulfur Spring in India

Kunwar Digvijay Narayan,[a] Jhasketan Badhai,[a] William B. Whitman,[b] Subrata K. Das[a]

Department of Biotechnology, Institute of Life Sciences, Nalco Square, Bhubaneswar, India[a]; Department of Microbiology, University of Georgia, Athens, Georgia, USA[b]

The genus *Comamonas* contains species isolated from various environments, such as termite guts, wetlands, activated sludge, soil, humans, and fresh water. Here, we report the draft genome sequence of *Comamonas thiooxydans* strain S23[T] capable of oxidizing thiosulfate under mixotrophic growth conditions. Based upon draft genome sequencing, the genome is 5.3 Mb and encodes 4,767 proteins. The *Comamonas thiooxydans* whole-genome sequence will help understand the metabolic diversity in sulfur oxidation pathways.

Address correspondence to Subrata K. Das, subratkdas@hotmail.com.

Comamonas thiooxydans strain S23[T] (DSM 17888[T]) is a Gram-negative aerobic rod. It belongs to the family *Comamonadaceae* of the class *Betaproteobacteria*. The bacterium was isolated from a sulfur spring located at Athmallik, Orissa, India (1). It is the first reported member of the genus *Comamonas* that is capable of oxidizing thiosulfate under mixotrophic growth conditions (2, 3).

The draft genome of *Comamonas thiooxydans* strain S23[T] was generated at the Department of Energy (DOE) Joint Genome Institute (JGI) using the Illumina HiSeq 2000 platform (4). An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform, which generated 8,941,414 reads totaling 1,350.2 Mb. The filtered Illumina reads were assembled using the Velvet (5), wgsim (https://github.com/lh3/wgsim), and Allpaths-LG (6) tools. The final draft assembly contains 45 contigs in 40 scaffolds, totaling 5.3 Mb, with an input read coverage of 253.1-fold. The largest and N_{50} contigs are 832.8 kb and 312.4 kb, respectively, with a G+C content of 62.0%.

The genome was annotated using the JGI Microbial Genome Annotation Pipeline (7). Genes were identified using the Prodigal (8) and GenePRIMP (9) programs. The tRNA, rRNA, and other noncoding RNA genes were identified by searching the genome using the tRNAscan-SE tool (10), rRNA gene models built from SILVA (11) and INFERNAL (http://infernal.janelia.org). The draft genome contains 4,767 coding sequences (CDSs), 68 tRNAs, 11 rRNAs, and 10 other RNAs.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LIOM00000000. The version described in this paper is the first version, LIOM01000000.

ACKNOWLEDGMENTS

REFERENCES

1. **Narayan KD, Pandey SK, Das SK**. 2010. Characterization of *Comamonas thiooxidans* sp. nov., and comparison of thiosulfate oxidation with *Comamonas testosteroni* and *Comamonas composti*. Curr Microbiol **61**:248–253. http://dx.doi.org/10.1007/s00284-010-9602-9.
2. **Pandey SK, Narayan KD, Bandyopadhyay S, Nayak KC, Das SK**. 2009. Thiosulfate oxidation by *Comamonas* sp. S23 isolated from a sulfur spring. Curr Microbiol **58**:516–521. http://dx.doi.org/10.1007/s00284-009-9357-3.
3. **Narayan KD, Pandey SK, Das SK**. 2011. List of new names and new combinations previously effectively, but not validly, published—validation list no. 141. Int J Syst Evol Microbiol **61**:2025–2026. http://dx.doi.org/10.1099/ijs.0.036640-0.
4. **Bennett S**. 2004. Solexa Ltd. Pharmacogenomics **5**:433–438. http://dx.doi.org/10.1517/14622416.5.4.433.
5. **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.
6. **Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB**. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A **108**:1513–1518. http://dx.doi.org/10.1073/pnas.1017351108.
7. **Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC**. 2009. The DOE-JGI standard operating procedure for the annotations of microbial genomes. Stand Genomic Sci **1**:63–67. http://dx.doi.org/10.4056/sigs.632.
8. **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.

9. **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.

10. **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.

11. **Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glckner FO**. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:**2188–7196.