

Draft Genome Sequence of *Thauera* sp. Strain SWB20, Isolated from a Singapore Wastewater Treatment Facility Using Gel Microdroplets

Armand E. K. Dichosa,^a Karen W. Davenport,^a Po-E Li,^a Sanaa A. Ahmed,^a Hajnalka Daligault,^a Cheryl D. Gleasner,^a Yuliya Kunde,^a Kim McMurry,^a Chien-Chi Lo,^a Krista G. Reitenga,^{a*} Ashlynn R. Daughton,^{a*} Xiaohong Shen,^{a*} Seth Frietze,^{a*} Dongping Wang,^a Shannon L. Johnson,^a Daniela I. Drautz-Moses,^b Stephan Schuster,^b Patrick S. Chain,^a Cliff Han^a

Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA^a; Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore^b

* Present address: Krista G. Reitenga, University of Maryland School of Medicine, Baltimore, Maryland, USA; Ashlynn R. Daughton, Defense Systems and Analysis, Los Alamos National Laboratory, Los Alamos, New Mexico, USA; Xiaohong Shen, Beacon Analytical System, Inc., Saco, Maine, USA; Seth Frietze, College of Natural and Health Sciences, University of Northern Colorado, Greeley, Colorado, USA.

We report here the genome sequence of *Thauera* sp. strain SWB20, isolated from a Singaporean wastewater treatment facility using gel microdroplets (GMDs) and single-cell genomics (SCG). This approach provided a single clonal microcolony that was sufficient to obtain a 4.9-Mbp genome assembly of an ecologically relevant *Thauera* species.

Received 4 February 2015 Accepted 5 February 2015 Published 19 March 2015

Citation Dichosa AEK, Davenport KW, Li P-E, Ahmed SA, Daligault H, Gleasner CD, Kunde Y, McMurry K, Lo C-C, Reitenga KG, Daughton AR, Shen X, Frietze S, Wang D, Johnson SL, Drautz-Moses DI, Schuster S, Chain PS, Han C. 2015. Draft genome sequence of *Thauera* sp. strain SWB20, isolated from a Singapore wastewater treatment facility using gel microdroplets. *Genome Announc* 3(2):e00132-15. doi:10.1128/genomeA.00132-15.

Copyright © 2015 Dichosa et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Armand E. K. Dichosa, armand@lanl.gov, or Stephan Schuster, scs@bx.psu.edu.

Species of the betaproteobacterial *Thauera* genus have been characterized as being able to aerobically (1–3) or anaerobically (4–7) degrade aromatic compounds under denitrifying conditions, as well as to oxidize organic acids and alcohols (8). Due to the potential of *Thauera* species in bioremediation applications and for their ability to produce exopolysaccharides (9), much interest has been focused on assembling and annotating the genomes of these ecologically relevant bacterial species. Interestingly, all the currently available *Thauera* sp. genomes were obtained from isolates from sludge/wastewater treatment facilities (9, 10).

Recently, our team utilized the combined technologies of gel microdroplets (GMDs) and single-cell genomics (SCG) to obtain several near-complete genomes of novel bacterial species inhabiting the human oral and gut microbiomes (11, 12). We applied this approach to coculture a complex sewage wastewater microbial community to attempt the recovery of as many diverse bacterial representatives as possible. We flow-sorted 176 GMDs and then subjected each to whole-genome amplification and bacterial 16S rRNA phylotyping (11–14). Of the 33 recovered bacterial genera (data not presented), one GMD was phylotyped as *Thauera* (named SWB20), having 99.9% 16S rRNA gene sequence identity (1,409 bp) to *Thauera* sp. MZ1T strain MZ1 (GenBank accession no. NR_074711.1). Due to its ecological significance, the amplified genome products of SWB20 were used to create libraries for the Illumina MiSeq (Illumina, Inc., San Diego, CA) and PacBio (Pacific Biosciences, Menlo Park, CA) sequencing platforms. The Illumina genome assemblies included both IDBA-UD version 1.1.0 (15) and Velvet version 1.2.08 (16) for 300× coverage. The PacBio genome assemblies utilized HGAP version 2.1.1 (17) for 159× coverage. The hybrid assemblies yielded 21 contigs (N_{50} , 343,694; N_{90} , 154,439; maximum, 1,000,345; minimum, 14,734)

for a near-complete 4,927,396-bp genome with 66.2% G+C content. Only contigs >14,000 bp from the final assembly were used for the subsequent annotations (18), which showed 4,421 protein-encoding genes (PEGs), 11 rRNAs (three 16S rRNAs, four 23S rRNAs, and four 5S rRNAs), and 61 tRNAs.

We performed genomic phylogeny and single-nucleotide polymorphism (SNP) comparison analyses (our unpublished data) of SWB20 against the previously published *Thauera* genomes: *T. aminoaromatica* strain MZ1T (9, 19) (also known as MZ1T; GenBank accession no. NC_011662.2), *Thauera* sp. strain 27 (accession no. AMXB01000000), *Thauera* sp. strain 28 (accession no. AMXA01000000), *T. linaloolentis* 47Lol (accession no. AMXE01000000), *Thauera* sp. strain 63 (accession no. AMXC01000000), and *T. aminoaromatica* S2 (also known as S2; accession no. AMXD01000000) (10). We found that SWB20 clusters with MZ1T and S2, while MZ1T and S2 are more closely related (data not presented). Our SNP analysis yielded a core genome size of 1,064,233 bp across all seven *Thauera* genomes. SWB20 had 0.66% and 0.58% SNP composition with MZ1T and S2, respectively, while S2 had 0.41% SNP composition compared to that of MZ1T, thereby supporting the close phylogenetic topography of these three *Thauera* strains. Both the close phylogeny and very low SNP differences compared to MZ1T and S2 suggest that SWB20 may be a unique *T. aminoaromatica* strain. Compared to the remaining *Thauera* sp. genomes, SWB20 averaged ~10% SNP composition (data not shown).

Nucleotide sequence accession numbers. The draft genome sequence of *Thauera* sp. strain SWB20 has been deposited as a whole-genome shotgun project at DDBJ/EMBL/GenBank under the accession no. [JTDM000000000](https://www.ncbi.nlm.nih.gov/nuccore/JTDM000000000) (BioProject ID PRJNA267225). The version described in this paper is version JTDM01000000.

ACKNOWLEDGMENTS

We thank Steve Turner of Pacific Biosciences for supplying the PacBio sequencing materials.

This genome-sequencing project was partially supported by the Los Alamos National Laboratory through the Laboratory Directed Research and Development Fund, grant 20100034DR. It was released under grant LA-UR-14-23425.

REFERENCES

- Song B, Young LY, Palleroni NJ. 1998. Identification of denitrifier strain T1 as *Thauera aromatica* and proposal for emendation of the genus *Thauera* definition. *Int J Syst Bacteriol* 48:889–894. <http://dx.doi.org/10.1099/00207713-48-3-889>.
- Scholten E, Lukow T, Auling G, Kroppenstedt RM, Rainey FA, Diekmann H. 1999. *Thauera mechernichensis* sp. nov., an aerobic denitrifier from a leachate treatment plant. *Int J Syst Bacteriol* 49:1045–1051. <http://dx.doi.org/10.1099/00207713-49-3-1045>.
- Song B, Palleroni NJ, Kerkhof LJ, Häggblom MM. 2001. Characterization of halobenzoate-degrading, denitrifying *Azoarcus* and *Thauera* isolates and description of *Thauera chlorobenzoica* sp. nov. *Int J Syst Evol Microbiol* 51:589–602. <http://ijs.sgmjournals.org/content/51/2/589>.
- Tschech A, Fuchs G. 1987. Anaerobic degradation of phenol by pure cultures of newly isolated denitrifying pseudomonads. *Arch Microbiol* 148:213–217. <http://dx.doi.org/10.1007/BF00414814>.
- Evans PJ, Mang DT, Kim KS, Young LY. 1991. Anaerobic degradation of toluene by a denitrifying bacterium. *Appl Environ Microbiol* 57:1139–1145.
- Macy JM, Rech S, Auling G, Dorsch M, Stackebrandt E, Sly LI. 1993. *Thauera selenatis* gen. nov., sp. nov., a member of the beta subclass of *Proteobacteria* with a novel type of anaerobic respiration. *Int J Syst Bacteriol* 43:135–142. <http://dx.doi.org/10.1099/00207713-43-1-135>.
- Anders H-J, Kaetzke A, Kämpfer P, Ludwig W, Fuchs G. 1995. Taxonomic position of aromatic-degrading denitrifying pseudomonad strains K 172 and KB 740 and their description as new members of the genera *Thauera*, as *Thauera aromatica* sp. nov., and *Azoarcus*, as *Azoarcus evansii* sp. nov., respectively, members of the beta subclass of the *Proteobacteria*. *Int J Syst Bacteriol* 45:327–333. <http://dx.doi.org/10.1099/00207713-45-2-327>.
- Dubbels BL, Sayavedra-Soto LA, Bottomley PJ, Arp DJ. 2009. *Thauera butanivorans* sp. nov., a C2-C9 alkane-oxidizing bacterium previously referred to as “*Pseudomonas butanovora*.” *Int J Syst Evol Microbiol* 59:1576–1578. <http://dx.doi.org/10.1099/ijs.0.000638-0>.
- Jiang K, Sanseverino J, Chauhan A, Lucas S, Copeland A, Lapidus A, Del Rio TG, Dalin E, Tice H, Bruce D, Goodwin L, Pitluck S, Sims D, Brettin T, Detter JC, Han C, Chang YJ, Larimer F, Land M, Hauser L, Kyrpides NC, Mikhailova N, Moser S, Jegier P, Close D, DeBruyn JM, Wang Y, Layton AC, Allen MS, Sayler GS. 2012. Complete genome sequence of *Thauera aminoaromatica* strain mz1t. *Stand Genomic Sci* 6:325–335. <http://dx.doi.org/10.4056/signs.2696029>.
- Liu B, Frostegård A, Shapleigh JP. 2013. Draft genome sequences of five strains in the genus *Thauera*. *Genome Announc* 1(1):e00052-12. <http://dx.doi.org/10.1128/genomeA.00052-12>.
- Dichosa AE, Daughton AR, Reitenga KG, Fitzsimons MS, Han CS. 2014. Capturing and cultivating single bacterial cells in gel microdroplets to obtain near-complete genomes. *Nat Protoc* 9:608–621. <http://dx.doi.org/10.1038/nprot.2014.034>.
- Fitzsimons MS, Novotny M, Lo C-C, Dichosa AE, Yee-Greenbaum JL, Snook JP, Gu W, Chertkov O, Davenport KW, McMurry K, Reitenga KG, Daughton AR, He J, Johnson SL, Gleasner CD, Wills PL, Parson-Quintana B, Chain PS, Detter JC, Lasken RS. 2013. Nearly finished genomes produced using gel microdroplet culturing reveal substantial intraspecies genomic diversity within the human microbiome. *Genome Res* 23:878–888. <http://dx.doi.org/10.1101/gr.142208.112>.
- Close DW, Ferrara F, Dichosa AE, Kumar S, Daughton AR, Daligault HE, Reitenga KG, Velappan N, Sanchez TC, Iyer S, Kiss C, Han CS, Bradbury AR. 2013. Using phage display selected antibodies to dissect microbiomes for complete *de novo* genome sequencing of low abundance microbes. *BMC Microbiol* 13:270. <http://dx.doi.org/10.1186/1471-2180-13-270>.
- Dichosa AE, Fitzsimons MS, Lo C-C, Weston LL, Preteska LG, Snook JP, Zhang X, Gu W, McMurry K, Green LD, Chain PS, Detter JC, Han CS. 2012. Artificial polyploidy improves bacterial single cell genome recovery. *PLoS One* 7:e37387. <http://dx.doi.org/10.1371/journal.pone.0037387>.
- Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <http://dx.doi.org/10.1093/bioinformatics/bts174>.
- 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>.
- Chin C-S, 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>.
- Orvis J, Crabtree J, Galens K, Gussman A, Inman JM, Lee E, Nampally S, Riley D, Sundaram JP, Felix V, Whitty B, Mahurkar A, Wortman J, White O, Angiuoli SV. 2010. Ergatis: a Web interface and scalable software system for bioinformatics workflows. *Bioinformatics* 26:1488–1492. <http://dx.doi.org/10.1093/bioinformatics/btq167>.
- Lajoie CA, Layton AC, Gregory IR, Sayler GS, Taylor DE, Meyers AJ. 2000. Zooglyph clusters and sludge dewatering potential in an industrial activated-sludge wastewater treatment plant. *Water Environ Res* 72:56–64. <http://dx.doi.org/10.2175/106143000X137112>.