

Draft Genome Sequence of *Thermotoga maritima* A7A Reconstructed from Metagenomic Sequencing Analysis of a Hydrocarbon Reservoir in the Bass Strait, Australia

Brodie Sutcliffe,^a David J. Midgley,^a Carly P. Rosewarne,^a Paul Greenfield,^b Dongmei Li^a

CSIRO Animal, Food and Health Sciences, North Ryde, NSW, Australia^a; CSIRO Mathematics, Informatics and Statistics, North Ryde, NSW, Australia^b

The draft genome sequence of *Thermotoga maritima* A7A was obtained from a metagenomic assembly obtained from a high-temperature hydrocarbon reservoir in the Gippsland Basin, Australia. The organism is predicted to be a motile anaerobe with an array of catabolic enzymes for the degradation of numerous carbohydrates.

Received 5 August 2013 Accepted 7 August 2013 Published 5 September 2013

Citation Sutcliffe B, Midgley DJ, Rosewarne CP, Greenfield P, Li D. 2013. Draft genome sequence of *Thermotoga maritima* A7A reconstructed from metagenomic sequencing analysis of a hydrocarbon reservoir in the Bass Strait, Australia. *Genome Announc.* 1(5):e00688-13. doi:10.1128/genomeA.00688-13.

Copyright © 2013 Sutcliffe et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to David J. Midgley, david.midgley@csiro.au.

The quality and efficiency of oil recovery from high-temperature oil reservoirs are affected by the activity of endogenous microbes (1). In a step toward unraveling the microbial communities in oil-associated subsurface environments, we examined the metagenome of a hydrocarbon reservoir (named “Tuna”; 38°10’S, 148°25’E) obtained from the A7A well in Gippsland Basin, Australia. The A7A well is ~1,960 m deep and has a temperature of 102°C, a pH of 7.2, and a salinity of 2.68%.

The metagenome was sequenced on the Illumina HiSeq 2000 system, which generated 2 × 100-bp paired-end reads. The resulting ~50 million metagenomic reads were corrected using Blue (<http://www.bioinformatics.csiro.au/blue>), prior to being assembled using Velvet v1.2.07 ($k = 41$). The first step in selecting the contigs belonging to the genome of *Thermotoga maritima* strain A7A was to assign each of the contigs in the metagenomic assembly to a family by comparing the 25-bp k -mers from each contig to all the unique k -mers found in ~4,000 microbial genome sequences and draft genome sequences from GenBank. Contigs were also binned using characteristic trinucleotide frequency signatures (2) and covariance principal components analysis (PCA) plots. Where these two methods disagreed, BLAST-based (3) binning was undertaken to confirm the identities of the remaining contigs.

In total, 116 contigs ($\geq 1,034$ bp) were assigned to *T. maritima* A7A; the contig lengths ranged from 1,034 to 116,114 bp, with mean, median, and N_{50} lengths of 15,194, 8,670, and 27,335 bp, respectively. In total, the contigs comprise 1,762,576 bp, with an overall G+C content of 45.7%. The closest sequenced relative of *T. maritima* A7A is *T. maritima* strain RQ2 (4), and the two strains share a 16S rRNA gene identity of 100% (over 980 bp). The contigs for the draft genome sequence were annotated using the Integrated Microbial Genomes Expert Review (IMG ER) pipeline (5), which identified 1,987 coding genes. Although *T. maritima* A7A is an anaerobe, like other *T. maritima* strains, it possesses an NAD(P)H oxidoreductase, rubredoxin, the flavodiiron protein, and neelaredoxin, encoded by genes laterally acquired from the

archaeal order *Thermococcales* (6), which confer tolerance to low levels of oxygen.

Previous work has found that *Thermotoga* species boast a large number of carbohydrate-active enzymes relative to their genome size (7) and that species within the genus differ in their possession of these enzymes (8). The genome appears to contain genes for the utilization of various disaccharides: sucrose, cellobiose, and maltose, along with the monosaccharides fructose, glucose, and xylose. Further analysis of carbohydrate-active enzymes using the dbCAN web service (9) indicates that in terms of carbohydrate usage, *T. maritima* A7A is most similar to *T. maritima* RQ2, although *T. maritima* A7A appears to lack the fructose phosphotransferase system (PTS) that is present in RQ2. Compared with the other *Thermotoga* species, *T. maritima* A7A appears to possess two carbohydrate-active enzymes not previously described in the genus, an esterase from family CE12 and a glycosyltransferase from family GT27.

Along with *T. maritima* A7A, the microbial community at the A7A well is dominated by species from the genera *Thermoanaerobacter* (10) and *Desulfonauticus*. Examinations of the putative interactions among these organisms in this environment are ongoing.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AUNF00000000](https://www.ncbi.nlm.nih.gov/nuccore/AUNF00000000). The version described in this paper is version AUNF01000000.

ACKNOWLEDGMENTS

This research was supported by CSIRO and the Victorian Department of Primary Industries.

We thank James Buckingham from Exxon Mobil for organizing the formation water sample collection. We also thank NMI for the chemistry analysis of the formation water samples and the University of Western Sydney for the Illumina sequencing.

REFERENCES

- Magot M, Ollivier B, Patel BK. 2000. Microbiology of petroleum reservoirs. *Antonie Van Leeuwenhoek* 77:103–116.

2. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10:R85. doi:[10.1186/gb-2009-10-8-r85](https://doi.org/10.1186/gb-2009-10-8-r85).
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
4. Swithers KS, Dipippo JL, Bruce DC, Detter C, Tapia R, Han S, Saunders E, Goodwin LA, Han J, Woyke T, Pitluck S, Pennacchio L, Nolan M, Mikhailova N, Lykidis A, Land ML, Brettin T, Stetter KO, Nelson KE, Gogarten JP, Noll KM. 2011. Genome sequence of *Thermotoga* sp. strain RQ2, a hyperthermophilic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores. *J. Bacteriol.* 193: 5869–5870.
5. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40(Database issue):D115–D122. doi:[10.1093/nar/gkr1044](https://doi.org/10.1093/nar/gkr1044).
6. Le Fourn C, Fardeau ML, Ollivier B, Lojou E, Dolla A. 2008. The hyperthermophilic anaerobe *Thermotoga maritima* is able to cope with limited amount of oxygen: insights into its defence strategies. *Environ. Microbiol.* 10:1877–1887.
7. Chhabra SR, Shockley KR, Connors SB, Scott KL, Wolfinger RD, Kelly RM. 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Mol. Biochem.* 278:7540–7552.
8. Frock AD, Gray SR, Kelly RM. 2012. Hyperthermophilic *Thermotoga* species differ with respect to specific carbohydrate transporters and glycoside hydrolases. *Appl. Environ. Microbiol.* 78:1978–1986.
9. Yin Y, Mao X, Yang JC, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40:W445–W451. doi:[10.1093/nar/gks479](https://doi.org/10.1093/nar/gks479).
10. Li D, Greenfield P, Rosewarne CP, Midgley DJ. 2013. Draft genome sequence of *Thermoanaerobacter* sp. strain A7A reconstructed from a metagenome obtained from a high-temperature hydrocarbon reservoir in the Bass Strait, Australia. *Genome Announc.* 1(5):e00701-13. doi:[10.1128/genomeA.00701-13](https://doi.org/10.1128/genomeA.00701-13).