Genome Sequence of *Thermotoga* sp. Strain RQ2, a Hyperthermophilic Bacterium Isolated from a Geothermally Heated Region of the Seafloor near Ribeira Quente, the Azores

Kristen S. Swithers,[1] Jonathan L. DiPippo,[1] David C. Bruce,[2] Christopher Detter,[2] Roxanne Tapia,[2] Shunsheng Han,[2] Elizabeth Saunders,[2]‡ Lynne A. Goodwin,[2] James Han,[3] Tanja Woyke,[3] Sam Pitluck,[3] Len Pennacchio,[3] Matthew Nolan,[3] Natalia Mikhailova,[3] Athanasios Lykidis,[3] Miriam L. Land,[4] Thomas Brettin,[4] Karl O. Stetter,[6] Karen E. Nelson,[5] J. Peter Gogarten,[1] and Kenneth M. Noll[1]*

*Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269-3125[1];
Los Alamos National Laboratory, Los Alamos, New Mexico 87545[2]; DOE Joint Genome Institute,
Walnut Creek, California 94598[3]; Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831[4];
The J. Craig Venter Institute, Rockville, Maryland[5]; and Lehrstuhl für Mikrobiologie,
Universität Regensburg, D-93053, Regensburg, Germany[6]*

Thermotoga sp. strain RQ2 is probably a strain of *Thermotoga maritima*. Its complete genome sequence allows for an examination of the extent and consequences of gene flow within *Thermotoga* species and strains. *Thermotoga* sp. RQ2 differs from *T. maritima* in its genes involved in *myo*-inositol metabolism. Its genome also encodes an apparent fructose phosphotransferase system (PTS) sugar transporter. This operon is also found in *Thermotoga naphthophila* strain RKU-10 but no other *Thermotogales*. These are the first reported PTS transporters in the *Thermotogales*.

Thermotoga sp. strain RQ2 is a hyperthermophilic anaerobic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores (7). Although this strain has not been characterized, its 16S rRNA gene sequence and the amino acid composition of some of its proteins show that it belongs to the genus *Thermotoga* and is probably a strain of *Thermotoga maritima* (10, 12). Examinations of its gene content using cosmid clones demonstrated a highly conserved pattern of gene organization between strains RQ2 and MSB8 (10–12). However, significant differences were also observed, and numerous genes were apparently acquired by this strain via horizontal gene transfer. Consequently, a more thorough sequencing of its genome was warranted to show the extent of subspecies genomic variation in *T. maritima*.

Genomic DNA was prepared from a culture of *Thermotoga* sp. RQ2 provided by Karl Stetter. The genome of *Thermotoga* sp. RQ2 was Sanger sequenced using a combination of 2.0-kbp, 2.6-kbp, and 6-kbp DNA libraries. All general aspects of library construction and sequencing can be found at the JGI website (http://www.jgi.doe.gov/). Reads from all libraries were assembled using the Phred/Phrap/Consed assembler (4–6). Possible misassemblies were corrected and gaps between contigs were closed by editing in Consed or by custom primer walks from subclones or PCR products.

Genes were identified using Glimmer (3) and Critica (1) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation to eliminate obvious overlaps. In order to determine a product description for each of the predicted coding sequences (CDSs), they were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Noncoding genes and miscellaneous features were predicted using tRNAscan-SE (9), TMHMM (8), and signalP (2).

The complete *Thermotoga* sp. RQ2 genome consists of 1,877,693 bp in a single circular chromosome. The genome has an average G+C content of 46%, and there are a total of 1,819 predicted protein-coding genes, one ribosomal operon, 46 tRNAs, and 34 pseudogenes.

Compared to the other publicly available *Thermotoga* genomes (*T. maritima*, *T. petrophila*, *T. neapolitana*, and *T. naphthophila*), several contiguous genes involved in *myo*-inositol metabolism were apparently lost. RQ2 has other genes for *myo*-inositol metabolism that are present in those *Thermotoga* species in a region containing *ino-1* that was previously found to have undergone interspecies recombination in *T. maritima* (11). The *Thermotoga* ribonucleotide reductase transcriptional repressor NrdR is also absent. NrdR represses the transcription of the ribonucleotide reductase genes (13), so *Thermotoga* sp. RQ2 probably uses a different regulatory mechanism. The strain RQ2 genome encodes a fructose phosphotransferase system (PTS) transporter (TRQ2_0637-0640). *T. naphthophila* is the only other *Thermotogales* species that has PTS-encoding genes. This is the first report of PTS systems in the *Thermotogales*. This genome sequence and those of the other *Thermotoga* species will allow in-depth comparative analyses of gene flow within this genus.

* Corresponding author. Mailing address: Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269-3125. Phone: (860) 486-4688. Fax: (860) 486-4331. E-mail: kenneth.noll @uconn.edu.
‡ Deceased.

Nucleotide sequence accession number. The complete genome sequence of *Thermotoga* sp. strain RQ2 is available in GenBank under accession number CP000969.

REFERENCES

1. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
2. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
3. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
4. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
5. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
6. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
7. **Huber, R., et al.** 1986. Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch. Microbiol. **144:**324–333.
8. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
9. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
10. **Nesbo, C. L., and W. F. Doolittle.** 2003. Targeting clusters of transferred genes in Thermotoga maritima. Environ. Microbiol. **5:**1144–1154.
11. **Nesbo, C. L., S. L'Haridon, K. O. Stetter, and W. F. Doolittle.** 2001. Phylogenetic analyses of two "archaeal" genes in thermotoga maritima reveal multiple transfers between archaea and bacteria. Mol. Biol. Evol. **18:**362–375.
12. **Nesbo, C. L., K. E. Nelson, and W. F. Doolittle.** 2002. Suppressive subtractive hybridization detects extensive genomic diversity in Thermotoga maritima. J. Bacteriol. **184:**4475–4488.
13. **Panosa, A., I. Roca, and I. Gibert.** 2010. Ribonucleotide reductases of Salmonella typhimurium: transcriptional regulation and differential role in pathogenesis. PLoS One **5:**e11328.