Draft Genome Sequence of *Thermus* sp. Isolate 2.9, Obtained from a Hot Water Spring Located in Salta, Argentina

Laura E. Navas,[a,c] Marcelo F. Berretta,[a,c] Elio M. Ortiz,[a] Graciela B. Benintende,[a] Ariel F. Amadio,[b,c] Rubén O. Zandomeni[a,c]

Instituto de Microbiología y Zoología Agrícola, Instituto Nacional de Tecnología Agropecuaria (INTA), Castelar, Buenos Aires, Argentina[a]; Estación Experimental Agropecuaria Rafaela, Instituto Nacional de Tecnología Agropecuaria (INTA), Rafaela, Santa Fe, Argentina[b]; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina[c]

A.F.A. and R.O.Z. contributed equally to and directed different aspects of the work.

Thermus sp. isolate 2.9 was obtained from a hot water spring in Salta, Argentina. Here, we report the draft genome sequence (2,485,434 bp) of this isolate, which consists of 11 scaffolds of >10 kbp and 2,719 protein-coding sequences.

Address correspondence to Laura E. Navas, navas.laura@inta.gob.ar.

Bacteria belonging to *Thermus* genus have been isolated mostly from hot spring environments. Species from this genus contain relevant genes with potential biotechnological applications as sources of thermostable enzymes. There is also interest in studying the mechanism involved in bacterial adaptation to extreme natural environments. *Thermus* sp. isolate 2.9 is a Gram-negative aerobic, rod-shaped, thermophilic bacterium isolated from a hot water spring in Rosario de la Frontera, Salta, in the northwest of Argentina.

The draft genome sequence data were generated using a combination of the Roche 454 GS and Illumina MiSeq platforms, producing unpaired and paired-end reads, respectively, together with an optical map of the genome. The 454 FLX shotgun data (Macrogen, Seoul, South Korea) consisted of 215,557 reads achieving 32-fold coverage of the genome. First, Newbler was used to assemble those reads in 119 contigs of >1 kbp. The contigs were aligned against the optical map using MapSolver version 3.2 and also against the genomes of *Thermus thermophilus* HB8, *T. thermophilus* HB27, and *Thermus aquaticus* Y51MC23 using Mauve 2.3.1 (1). The analysis of the alignments allowed the ordering of contigs, and 32 gaps were closed by primer walking and sequencing of the PCR products over the gap, reducing the number of contigs to 87. To improve the genome assembly, an Illumina paired-end library was constructed using long jump distance technology with an insert size of 8 kbp (MWG Eurofins), generating 953,708 paired reads with an average length of 124 bp, which achieved 110-fold coverage. Finally, scaffolding and *in silico* gap filling were performed with SSCAPE 2.0 (2) and GapFiller 1.10 (3), respectively, resulting in 11 scaffolds of >10 kbp, containing 62 contigs. It represents an efficient strategy for sequencing genomes with a high G+C content and high proportions of repeated sequences, in which obtaining a single contig is unlikely (4).

The final assembly of *Thermus* sp. isolate 2.9 presented a total size of 2,485,434 bp, with an average G+C content of 67.29%. It was subjected to automated annotation using the RAST server 2.0 (5) and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). The tRNA and rRNA genes were predicted using tRNASCan-SE 1.21 (6) and RNAmmer 1.2 (7), respectively.

Annotation by RAST predicted 2,719 protein-coding genes, 45% of which were assigned to 360 subsystems. *Thermus* sp. isolate 2.9 has one copy of 23S/5S and 16S rRNA genes and 46 tRNA genes. *T. thermophilus* HB8 (score, 529), HB27 (score, 526), and *T. aquaticus* (score, 424) were reported by RAST to be the closest neighbors of *Thermus* sp. isolate 2.9. Also, analysis using BLAST showed that three scaffolds, corresponding to 394,087 bp, have similarities with known plasmids of *T. thermophilus* strains.

From all enzymes identified by the annotation, 435 were selected for their potential applications in agriculture, biosensors, biotechnology, environment, energy, food, medicine, and other industries. In order to assess their functionality, the cloning and expression of some of these enzymes in *Escherichia coli* are under way. Genome-based knowledge of thermophiles is of great importance and interest for not only assessing the diversity of their enzymes but also understanding their metabolic adaptations to extreme living conditions.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no. JTJB00000000. The version described in this paper is version JTJB01000000.

REFERENCES

1. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394–1403. http://dx.doi.org/10.1101/gr.2289704.

2. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. BioInformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.

3. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with Gap-Filler. Genome Biol **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.

4. **Metzker ML.** 2010. Sequencing technologies—the next generation. Nat Rev Genet **11:**31–46. http://dx.doi.org/10.1038/nrg2626.

5. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.

6. **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.

7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.