Complete Genome Sequence of the Thermophilic Bacterium *Thermus* sp. Strain CCB_US3_UF1

Beng Soon Teh,[a] Ahmad Yamin Abdul Rahman,[a] Jennifer A. Saito,[a] Shaobin Hou,[b] and Maqsudul Alam[a]

Centre for Chemical Biology, Universiti Sains Malaysia, Penang, Malaysia,[a] and Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA[b]

Thermus sp. strain CCB_US3_UF1, a thermophilic bacterium, has been isolated from a hot spring in Malaysia. Here, we present the complete genome sequence of *Thermus* sp. CCB_US3_UF1.

*T*hermus spp. are Gram-negative, aerobic, nonsporulating, and rod-shaped thermophilic bacteria (3). Members of the genus *Thermus* have potential in biotechnological applications as sources of thermostable DNA polymerase used in PCR techniques (10, 11). Here, we report the complete genome sequence of *Thermus* sp. strain CCB_US3_UF1, isolated from a hot spring in Ulu Slim, Perak, Malaysia, at 92.4°C, pH 7.

The genomic DNA of *Thermus* sp. CCB_US3_UF1 was extracted using a modified phenol-chloroform protocol (5). The whole-genome sequencing of *Thermus* sp. CCB_US3_UF1 was performed using Roche 454 and Solexa paired-end sequencing technology. A 3-kb genomic library was constructed, and 97,991 paired-end reads and 54,397 single-end reads were generated using the GS FLX system, providing 21.14-fold genome coverage. A total of 3,469,788 reads from the 3-kb library were produced to reach a depth of 115-fold coverage with an Illumina Solexa GA IIx (Illumina, San Diego, CA). These reads were mapped to the scaffolds using the Burrows-Wheeler alignment (BWA) tool (7).

The complete genome of *Thermus* sp. CCB_US3_UF1 is composed of a single circular chromosome of 2,243,772 bp and a plasmid of 19,716 bp, with G+C contents of 68.6% and 65.6%, respectively. There are 2,247 predicted coding sequences (CDS), 2 rRNA operons, and 48 tRNA genes. There are 32 predicted CDS in the plasmid. The automated annotation of the genome was done using the DIYA (Do-It-Yourself Annotator) pipeline (12). Open reading frames (ORFs) were identified using Glimmer3 (4), followed by a protein similarity search using BLAST (1) against UNIREF (13), RPS-BLAST against CDD (9), and Asgard (2). Transfer RNAs were predicted by using tRNAscan-SE (8), while ribosomal RNAs were identified by using RNAmmer (6).

The genome reveals that *Thermus* sp. CCB_US3_UF1 possesses numerous transporters for efficient substrate and nutrient uptake and for utilization of various energy sources.

Nucleotide sequence accession numbers. The genome sequences of *Thermus* sp. CCB_US3_UF1 have been deposited in GenBank under accession numbers CP003126 (chromosome) and CP003127 (plasmid).

ACKNOWLEDGMENT

REFERENCES

1. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
2. **Alves JM, Buck GA.** 2007. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. Chem. Biodivers. **4**:2593–2602.
3. **Brock TD, Freeze H.** 1969. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. J. Bacteriol. **98**:289–297.
4. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
5. **Fulton J, Douglas T, Young M.** 2009. Isolation of viruses from high temperature environments, p 43–54. *In* Clokie M and Kropinski AM (ed), Bacteriophages: methods and protocols. Volume 1: isolation, characterization, and interactions. Humana Press, New York, NY.
6. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
7. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760.
8. **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.
9. **Marchler-Bauer A, et al.** 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. **39**:D225–D229.
10. **Niehaus F, Bertoldo C, Kahler M, Antranikian G.** 1999. Extremophiles as a source of novel enzymes for industrial application. Appl. Microbiol. Biotechnol. **51**:711–729.
11. **Pantazaki AA, Pritsa AA, Kyriakidis DA.** 2002. Biotechnologically relevant enzymes from *Thermus thermophilus*. Appl. Microbiol. Biotechnol. **58**:1–12.
12. **Stewart AC, Osborne B, Read TD.** 2009. DIYA: a bacterial annotation pipeline for any genomics lab. Bioinformatics **25**:962–963.
13. **Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH.** 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics **23**:1282–1288.