

# Draft Genome Sequence of the Thermophile *Thermus filiformis* ATCC 43280, Producer of Carotenoid-(Di)glucoside-Branched Fatty Acid (Di)esters and Source of Hyperthermostable Enzymes of Biotechnological Interest

Fernanda Mandelli,<sup>a,b</sup> Brenda Oliveira Ramires,<sup>a</sup> Matthew Brian Couger,<sup>c</sup> Douglas A. A. Paixão,<sup>a</sup> Cesar M. Camilo,<sup>d</sup> Igor Polikarpov,<sup>d</sup> Rolf Prade,<sup>c</sup> Diego M. Riaño-Pachón,<sup>a</sup> Fabio M. Squina<sup>a</sup>

Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, SP, Brazil<sup>a</sup>; Departamento de Ciência de Alimentos da Faculdade de Engenharia de Alimentos, UNICAMP, Campinas, SP, Brazil<sup>b</sup>; Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA<sup>c</sup>; Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil<sup>d</sup>

F.M. and B.O.R. contributed equally to this work.

**Here, we present the draft genome sequence of *Thermus filiformis* strain ATCC 43280, a thermophile bacterium capable of producing glycosylated carotenoids acylated with branched fatty acids and enzymes of biotechnological potential.**

Received 6 April 2015 Accepted 7 April 2015 Published 14 May 2015

**Citation** Mandelli F, Oliveira Ramires B, Couger MB, Paixão DAA, Camilo CM, Polikarpov I, Prade R, Riaño-Pachón DM, Squina FM. 2015. Draft genome sequence of the thermophile *Thermus filiformis* ATCC 43280, producer of carotenoid-(di)glucoside-branched fatty acid (di)esters and source of hyperthermostable enzymes of biotechnological interest. *Genome Announc* 3(3):e00466-15. doi:10.1128/genome.A00466-15.

**Copyright** © 2015 Mandelli 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 Diego M. Riaño-Pachón, [diego.riano@bioetanol.org.br](mailto:diego.riano@bioetanol.org.br), or Fabio M. Squina, [fabio.squina@bioetanol.org.br](mailto:fabio.squina@bioetanol.org.br).

*Thermus filiformis* strain (ATCC 43280—obtained from the National Institute for Quality Control in Health, FIOCRUZ—Rio de Janeiro, Brazil) was first isolated from a hot spring in New Zealand (1). *T. filiformis* forms long filaments consisting of chains of cells and can thus be morphologically distinguished from other *Thermus* strains (1). This strain is particularly interesting, as it was reported to produce thermozeaxanthins and thermobiszeaxanthins, which are carotenoid-(di)glucoside-branched fatty acid (di)esters (2). Carotenoids are pigments known for their ability to act as potent antioxidants, protecting cells and tissues from the damaging effects of reactive oxygen and nitrogen species. Besides carotenoid production, *T. filiformis* can be considered a source of hyperthermostable enzymes of biotechnological interest.

Here we present the genome sequence of *T. filiformis* ATCC 43280. This genome was sequenced on the Illumina MiSeq sequencing system, generating 7,099,814 paired-end reads of 250 bp (insert size, 250 bp) and 2,124,227 mate-pair reads of 150 bp (insert size between 2 Kbp and 15 Kbp). Paired-end reads were pre-processed with Trimmomatic (3), and mate pairs with NextClip (4), resulting in 1,857,428 and 621,279 cleaned reads, respectively. The genome size was estimated to be 3.1 Mbp based on k-mer count statistics (5), with an estimated coverage of 580×. Chromosomal assembly was carried out with SPAdes v3.5.0 using an ensemble of k-mer values (K = [21, 33, 55, 77, 99, 127]). Presence of typical bacterial marker genes was assessed using Amphora2 (6). The resulting assembly has 40 scaffolds, with a total length of 2,386,081 bp and an  $N_{50}$  of 551,922 bp. The average G+C content of the genome is 68.9%, which is similar to related species, i.e., *Thermus thermophilus* HB8, GC of 69.5% (7); *T. oshimai* JL-2, GC of 68.6% (7); *T. scotoductus* SA-01, GC of 64.9% (8); *T. islandicus* DSM251543, GC of 68.3% (assembly, ATXJ00000000.1); *Ther-*

*mus antranikianii* DSM12462, GC of 64.8% (assembly, GCA\_000423905.1); and *Thermus aquaticus* Y51MC23, GC of 68.0% (assembly, ABVK00000000.2). Gene prediction was carried out with the NCBI Prokaryotic Genome Annotation Pipeline (9). A total of 2,405 genes were identified, of these, there are 2,211 protein-encoding genes, 6 rRNA genes, 47 tRNAs, and 9 pseudo-genes. Gene content is similar to that of related species, i.e., *T. thermophilus* (2,035 genes), *T. oshimai* (2,205 genes), *T. scotoductus* (2,503 genes), *T. islandicus* (2,289 genes), *T. antranikianii* (2,225 genes), and *T. aquaticus* (2,593 genes). The current genome assembly provides a preliminary landscape of the genomic and metabolic capabilities of *T. filiformis*.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [JPSL00000000](https://ncbi.nlm.nih.gov/assembly/JPSL00000000/). The version described in this paper is version JPSL02000000.

## ACKNOWLEDGMENTS

This work was financially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP: 2014/50371-8) and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: 442333/2014-5 and 310186/2014-5). F.M. received a fellowship from CNPq (142685/2010-0).

## REFERENCES

- Hudson JA, Morgan HW, Daniel RM. 1987. *Thermus filiformis* sp. nov., a filamentous Caldoactive bacterium. *Int J Syst Bacteriol* 37:431–436. [http://dx.doi.org/10.1099/00207713-37-4-431](https://doi.org/10.1099/00207713-37-4-431).
- Mandelli F, Miranda VS, Rodrigues E, Mercadante AZ. 2012. Identification of carotenoids with high antioxidant capacity produced by extremophile microorganisms. *World J Microbiol Biotechnol* 28:1781–1790. [http://dx.doi.org/10.1007/s11274-011-0993-y](https://doi.org/10.1007/s11274-011-0993-y).

3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
4. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 30:566–568. <http://dx.doi.org/10.1093/bioinformatics/btt702>.
5. Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764–770. <http://dx.doi.org/10.1093/bioinformatics/btr011>.
6. Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28:1033–1034. <http://dx.doi.org/10.1093/bioinformatics/bts079>.
7. Murugapiran SK, Huntemann M, Wei CL, Han J, Dettler JC, Han CS, Erkkila TH, Teshima H, Chen A, Kyrpides N, Mavrommatis K, Markowitz V, Szeto E, Ivanova N, Pagani I, Lam J, McDonald AI, Dodsworth JA, Pati A, Goodwin L, Peters L, Pitluck S, Woyke T, Hedlund BP. 2013. Whole-genome sequencing of *Thermus oshimai* JL-2 and *Thermus thermophilus* JL-18, incomplete denitrifiers from the United States Great Basin. *Genome Announc* 1(1):e00106-12. <http://dx.doi.org/10.1128/genomeA.00106-12>.
8. Gounder K, Brzuszkiewicz E, Liesegang H, Wollherr A, Daniel R, Gottschalk G, Reva O, Kumwenda B, Srivastava M, Bricio C, Berenguer J, van Heerden E, Litthauer D. 2011. Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01. *BMC Genomics* 12:577. <http://dx.doi.org/10.1186/1471-2164-12-577>.
9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufu S, Li W. 2013. Prokaryotic genome annotation pipeline. In Beck J, Benson D, Coleman J, Hoepfner M, Johnson M, Maglott D, Mizrahi I, Morris R, Ostell J, Pruitt K, Rubinstein W, Sayers E, Sirotkin K, Tatusova T (ed), *The NCBI handbook*, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.