

Genome Sequences of Four Italian *Streptococcus thermophilus* Strains of Dairy Origin

Laura Treu,^a Veronica Vendramin,^b Barbara Bovo,^b Stefano Campanaro,^c Viviana Corich,^b Alessio Giacomin^b

Department of Environmental Engineering, Technical University of Denmark, Kongens Lyngby, Denmark^a; Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Padua, Italy^b; Department of Biology, University of Padova, Padua, Italy^c

This report describes the genome sequences of four *Streptococcus thermophilus* strains, namely, TH982, TH985, TH1477, and 1F8CT, isolated from different dairy environments from the Campania and the Veneto regions in Italy. These data are aimed at increasing the genomic information available on this species, which is of paramount importance for the dairy industry.

Received 18 February 2014 Accepted 21 February 2014 Published 13 March 2014

Citation Treu L, Vendramin V, Bovo B, Campanaro S, Corich V, Giacomin A. 2014. Genome sequences of four Italian *Streptococcus thermophilus* strains of dairy origin. *Genome Announc.* 2(2):e00126-14. doi:10.1128/genomeA.00126-14.

Copyright © 2014 Treu 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 Viviana Corich, viviana.corich@unipd.it.

Streptococcus thermophilus is a species of great importance for the preparation of numerous dairy products (1). In fact, besides lactic acid production, several other technological characteristics are possessed by strains of this species.

The first two strains of *S. thermophilus*, TH982 and TH985, were isolated in the Campania region from whey and curd, respectively, obtained from mozzarella di bufala campana DOP cheese (pasta filata cheese) production. *S. thermophilus* strains TH1477 and 1F8CT were isolated in the Veneto region from cow milk and curd, respectively, from Grana Padano DOP cheese (long-ripening cheese) production. These strains possess a variety of functional properties of technological interest for the dairy industry, such as slow (1F8CT) or rapid (TH1477, TH982, and TH985) acidifying capabilities, low fermentation and growth rates (1F8CT), and pigment biosynthesis (TH1477).

Here, we present the genome sequences of *S. thermophilus* strains TH982, TH985, TH1477, and 1F8CT, generated using an Illumina MiSeq platform (with 1-kb mate-pair libraries) at the Ramaciotti Centre, Sydney, Australia. Sequencing coverages of 218×, 183×, 142×, and 261× were obtained corresponding to 1,635,998, 1,410,685, 1,258,918, and 1,544,110 paired-end reads, respectively (2 × 250 bp). The files generated were assembled with Velvet version 1.2.10 (2) and ABySS software version 1.3.5 (3) (optimal *k*-mer, 131). The consensus sequences of the two assemblies were manually compared. Between 52 and 84 scaffolds were obtained, with total sizes of 1,733,024, 1,838,250, 1,879,471, and 1,742,121 bp for strains TH982, TH985, TH1477, and 1F8CT, respectively, with a G+C content of 39%. All strains had the scaffolds assembled into a single circular chromosome by aligning them against the reference genome of *S. thermophilus* CNRZ1066 (assembly ASM1184v1). Several plasmid sequences were detected by BLAST analysis in two scaffolds of strain TH985, in three of TH1477, and in five of 1F8CT.

Protein-coding open reading frames (ORFs) were predicted and annotated using the RAST server (4). The numbers of predicted protein-coding genes (CDSs) in TH982, TH985, TH1477, and 1F8CT strains were 1,924, 1,952, 1,986 and 1,864, respec-

tively. Furthermore, 47, 69, 55, and 51 RNA genes were found, respectively. These genomes contain few phage sequences and no transposase-coding genes, while 20 to 26 clusters of regularly interspaced short palindromic repeats (CRISPRs) were found.

A comparison with the reference genome of *S. thermophilus* CNRZ1066 highlighted numerous CDSs that were exclusive to each strain. In particular, strains from the Veneto region include numerous sequences belonging to the oxidative stress responding genes, while those from Campania contain unique genes for choline and betaine uptake and betaine biosynthesis, which are important in the osmotic stress response.

These data are intended to increase the availability of genomes of *S. thermophilus* strains of dairy origin (5, 6) in order to better understand their biodiversity and their known and potential technological properties.

Nucleotide sequence accession numbers. The sequences of this whole-genome shotgun project have been deposited at DDBJ/EMBL/GenBank under the accession no. [AZTL000000000](https://www.ncbi.nlm.nih.gov/nuclink/AZTL000000000), [AZTM000000000](https://www.ncbi.nlm.nih.gov/nuclink/AZTM000000000), [AZTJ000000000](https://www.ncbi.nlm.nih.gov/nuclink/AZTJ000000000), and [AZTK000000000](https://www.ncbi.nlm.nih.gov/nuclink/AZTK000000000) for *S. thermophilus* strains TH982, TH985, TH1477, and 1F8CT, respectively. The versions described in this paper are AZTL01000000, AZTM01000000, AZTJ01000000 and AZTK01000000, respectively.

ACKNOWLEDGMENTS

This study was supported in part by POR Veneto “Fondo Europeo di Sviluppo Regionale” 2007–2013—Asse 1, Azione 1.1.2

We thank Angiolella Lombardi for providing *S. thermophilus* strains.

REFERENCES

- Mora D, Fortina MG, Parini C, Ricci G, Gatti M, Giraffa G, Manachini PL. 2002. Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J. Appl. Microbiol.* 93: 278–287. <http://dx.doi.org/10.1046/j.1365-2672.2002.01696.x>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.

3. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
4. 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>.
5. Treu L, Vendramin V, Bovo B, Campanaro S, Corich V, Giacomini A. 2014. Whole-genome sequences of *Streptococcus thermophilus* strains TH1435 and TH1436 isolated from raw goat milk. *Genome Announc.* 2(1): e01129-13. <http://dx.doi.org/10.1128/genomeA.01129-13>.
6. Treu L, Vendramin V, Bovo B, Campanaro S, Corich V, Giacomini A. 2014. Genome sequences of *Streptococcus thermophilus* strains MTH17CL396 and M17PTZA496 from fontina, an Italian PDO cheese. *Genome Announc.* 2(1):e00067-14. <http://dx.doi.org/10.1128/genomeA.00067-14>.