



# Draft Genome Sequences of 18 *Streptococcus* Strains Isolated from Live Dietary Supplements and Cultured Food Products

Tammy J. Barnaba,<sup>a</sup> Jayanthi Gangiredla,<sup>a</sup> Mark K. Mammel,<sup>a</sup>  David W. Lacher,<sup>a</sup>  Carmen Tartera<sup>a</sup>

<sup>a</sup>Division of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland, USA

**ABSTRACT** We present the genome sequences of 18 *Streptococcus* isolates from 8 different dietary supplements and 9 cultured food products. Strains from this species naturally colonize the human mouth and upper respiratory tract. Studies have shown that *S. thermophilus* and *S. salivarius* strains confer oral health benefits to their host with little to no risk of pathogenic infection.

*Streptococcus* is a Gram-positive, lactic acid bacterium found in dairy products and dietary supplements. *S. thermophilus* use in fermentation of milk products has been linked to the facilitation of dairy product digestion (1, 2). *S. salivarius* naturally resides in the mouth and upper respiratory tract and is one of the first beneficial microbes to colonize infants (3). This species is thought to be beneficial in the reduction of dental plaque and periodontal health when administered orally (4, 5). It has proven difficult to distinguish *S. thermophilus* from *S. salivarius* using medium plating methods, but genomic sequencing accurately identified them within fecal samples from individuals who had consumed dairy products (6). Dietary supplements containing these strains are becoming more prevalent; thus, the need for high-quality genomic sequences representing commonly used strains is crucial for proper identification. Here, we present the draft genome sequences of 2 *S. salivarius* and 16 *S. thermophilus* strains isolated from dietary supplements and cultured food products for classification (Table 1).

A serving from each product was suspended in 5 mL of saline. Then, 100  $\mu$ L from serial dilutions was plated on MRS agar plates and incubated at 30°C aerobically and 37°C aerobically, CO<sub>2</sub>-enriched, microaerophilic and anaerobic conditions. Colonies were selected based on morphologic differences. DNA extraction from overnight cultures of purified colony isolates was performed with the QIAcube using the DNeasy blood and tissue/Gram-positive bacterium or yeast protocol (Qiagen, Hilden, Germany). A DNA clean and concentrator kit (Zymo Research, Irvine, CA) was used to increase the quality of the DNA by removing potential inhibitors that could affect the sequencing process, as well as to concentrate the sample from 150  $\mu$ L to a final volume of 6  $\mu$ L. DNA quality and concentration were determined using a Qubit 3.0 fluorometer (Life Technologies, Burlington, CA). Sequencing libraries were prepared with 1 ng of DNA using the Nextera XT DNA library preparation kit and the Nextera XT index kit v2 set A (Illumina, San Diego, CA) and sequenced on either the Illumina MiSeq system using the MiSeq reagent kit v2 (500 cycles) or the NextSeq system using the NextSeq 500/550 midoutput kit v2.5 (300 cycles). The resulting 250-bp (MiSeq) and 150-bp (NextSeq) paired-end reads were quality controlled using FastQC v0.11.9 (Q > 30) (7), and then the reads were trimmed using Trimmomatic v0.38.1 (8) and *de novo* assembled using SPAdes v3.8.2 (9). The completeness of assemblies was determined using BUSCO v5.0.0 and Lactobacillales\_odb10, with genome assemblies completed from 93.7% to 99.5% (10). Species-level taxonomic identification and abundances were inferred for all samples using our in-house kmer database (k = 30) (11). Default parameters were used for all the analytical tools.

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Carmen Tartera, Carmen.Tartera@fda.hhs.gov.

The authors declare no conflict of interest.

**Received** 15 March 2022

**Accepted** 30 April 2022

**Published** 26 May 2022

**TABLE 1** *Streptococcus* isolates sequenced in this study

Species	Strain	Source	Sequence Platform	SRA no.	BioSample no.	GenBank WGS accession no.	No. of reads	Genome coverage (x)	No. of contigs	No. of predicted genes	Largest contig (bp)	Total length (bp)	$N_{50}$ (bp)	GC (%)	Lineage dataset	Complete	Missing	Total	%
<i>S. salivarius</i>	DS85_408 <sup>a</sup>	Dietary supplement <sup>b</sup>	NextSeq 500/550	SRR11910125	SAMN15078926	JAIQVW0000000000	3,817,804	119	58	2,181	307,204	2,320,947	124,435	39.4	lactobacillales_odb10	400	2	402	99.5
<i>S. salivarius</i>	DS85_40A <sup>a</sup>	Dietary supplement <sup>b</sup>	NextSeq 500/550	SRR11910126	SAMN15078925	JAIQVW5000000000	3,984,844	69	84	2,154	176,391	2,290,943	65,821	39.4	lactobacillales_odb10	400	2	402	99.5
<i>S. thermophilus</i>	DS14_14	Dietary supplement <sup>b</sup>	MISeq	SRR5310871	SAMN06464115	JAIQVW0000000000	3,061,692	392	62	1,836	207,436	1,776,323	73,135	38.9	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF2_14	Yogurt product	MISeq	SRR5310877	SAMN06464117	JAIQVW0000000000	814,748	92	76	1,824	194,267	1,742,651	53,094	39.0	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF1_14	Kefir product	MISeq	SRR5310877	SAMN06464116	JAIQVW0000000000	4,348,482	238	44	1,814	203,892	1,742,991	124,918	39.0	lactobacillales_odb10	388	14	402	95.5
<i>S. thermophilus</i>	DS76_14	Dietary supplement <sup>b</sup>	NextSeq 500/550	SRR11910219	SAMN15078890	JAIQVW0000000000	6,496,696	202	62	1,822	122,821	1,736,106	63,115	38.9	lactobacillales_odb10	392	10	402	97.5
<i>S. thermophilus</i>	DS51_14	Dietary supplement <sup>b</sup>	NextSeq 500/550	SRR11910208	SAMN15078890	JAIQVW0000000000	9,984,694	20	97	1,759	104,555	1,646,584	41,717	39.2	lactobacillales_odb10	398	4	402	99.0
<i>S. thermophilus</i>	CF5_14	Nondairy yogurt product	NextSeq 500/550	SRR11910326	SAMN15078648	JAIQVW0000000000	6,256,254	565	56	1,849	137,000	1,777,217	65,705	38.9	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	DS03_14	Dietary supplement <sup>b</sup>	NextSeq 500/550	SRR11910242	SAMN15078736	JAIQVW0000000000	2,494,996	89	64	1,867	122,822	1,774,216	61,113	38.9	lactobacillales_odb10	398	4	402	99.0
<i>S. thermophilus</i>	DS30_14	Dietary supplement <sup>b</sup>	MISeq	SRR11910360	SAMN15078662	JAIQVW0000000000	938,788	68	61	1,922	175,129	1,848,774	70,646	38.8	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	DS30_14	Yogurt product	NextSeq 500/550	SRR11910418	SAMN15078680	JAIQVW0000000000	1,259,266	96	81	1,853	117,608	1,766,804	52,466	39.0	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF13_14	Yogurt product	NextSeq 500/550	SRR11910376	SAMN15078605	JAIQVW0000000000	1,669,016	105	46	1,892	174,923	1,808,551	93,887	38.9	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF12_14	Yogurt product	NextSeq 500/550	SRR11910216	SAMN15078601	JAIQVW0000000000	9,134,166	484	55	1,925	313,755	1,848,466	100,076	38.8	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF3_14	Nondairy yogurt product	NextSeq 500/550	SRR11910321	SAMN15078635	JAIQVW0000000000	8,713,470	173	52	1,778	204,361	1,704,260	97,302	38.9	lactobacillales_odb10	377	25	402	93.8
<i>S. thermophilus</i>	CF10_14	Cheese product	NextSeq 500/550	SRR11910350	SAMN15078639	JAIQVW0000000000	1,720,144	123	59	1,847	145,424	1,759,255	70,976	39.0	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF4_14	Kefir product	NextSeq 500/550	SRR11910241	SAMN15078639	JAIQVQ0000000000	5,057,426	88	69	1,866	158,655	1,771,462	62,823	38.9	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	DS61_14	Dietary supplement <sup>b</sup>	MISeq	SRR11910392	SAMN15078830	JAIQVW0000000000	4,364,456	550	49	1,847	162,842	1,778,590	71,039	38.9	lactobacillales_odb10	399	3	402	99.3
<i>S. thermophilus</i>	CF8_14	Yogurt product	NextSeq 500/550	SRR11910258	SAMN15078656	JAIQVW0000000000	1,755,666	75	68	1,851	130,762	1,773,459	58,946	38.9	lactobacillales_odb10	399	3	402	99.3

<sup>a</sup> Two different strains from the same dietary supplement (DS85).

<sup>b</sup> Dietary supplement for general population.

<sup>c</sup> Dietary supplement for women.

Depth of coverage for the draft genomes ranged from 20 to 565 $\times$ , with genome sizes ranging from 1,646,854 to 2,320,947 bp. The number of contigs ranged from 44 to 97, while the  $N_{50}$  values ranged from 41,717 to 124,918 bp. The GC content varied from 38.8 to 39.4%. The sequences of these 18 *Streptococcus* isolates were submitted to GenBank and included two *S. salivarius* strains and 16 *S. thermophilus* strains.

**Data availability.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under BioProject number [PRJNA336518](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA336518) and under accession numbers [JAIQWE000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JAIQWE000000000) to [JAIQXB000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JAIQXB000000000). The assemblies described in this paper are the first versions. All annotations were done using the NCBI PGAP pipeline v5.2 (12).

## ACKNOWLEDGMENTS

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration (FDA), or the U.S. Government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the FDA.

## REFERENCES

1. Iyer R, Tomar SK, Maheswari TU, Rameshwar S. 2010. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. *Int Dairy J* 20:133–141. <https://doi.org/10.1016/j.idairyj.2009.10.005>.
2. Rul F, Ben-Yahia L, Chegdani F, Wrzosek L, Thomas S, Noordine ML, Gitton C, Cherbuy C, Langella P, Thomas M. 2011. Impact of the metabolic activity of *Streptococcus thermophilus* on the colon epithelium of gnotobiotic rats. *J Biol Chem* 286:10288–10296. <https://doi.org/10.1074/jbc.M110.168666>.
3. Kaci G, Goudercourt D, Dennin V, Pot B, Doré J, Ehrlich SD, Renault P, Blottière HM, Daniel C, Delorme C. 2014. Anti-inflammatory properties of *Streptococcus salivarius*, a commensal bacterium of the oral cavity and digestive tract. *Appl Environ Microbiol* 80:928–934. <https://doi.org/10.1128/AEM.03133-13>.
4. Burton JP, Drummond BK, Chilcott CN, Tagg JR, Thomson WM, Hale J, Wescombe PA. 2013. Influence of the probiotic *Streptococcus salivarius* strain M18 on indices of dental health in children: a randomized double-blind, placebo-controlled trial. *J Med Microbiol* 62:875–884. <https://doi.org/10.1099/jmm.0.056663-0>.
5. Burton JP, Chilcott CN, Moore CJ, Speiser G, Tagg JR. 2006. A preliminary study of the effect of probiotic *Streptococcus salivarius* K12 on oral malodour parameters. *J Appl Microbiol* 100:754–764. <https://doi.org/10.1111/j.1365-2672.2006.02837.x>.
6. Martinović A, Cocuzzi R, Arioli S, Mora D. 2020. *Streptococcus thermophilus*: to survive, or not to survive the gastrointestinal tract, that is the question! *Nutrients* 12:2175. <https://doi.org/10.3390/nu12082175>.
7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
9. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
10. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
11. Leonard SR, Mammel MK, Lacher DW, Elkins CA. 2015. Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach. *Appl Environ Microbiol* 81:8183–8191. <https://doi.org/10.1128/AEM.02601-15>.
12. Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.