AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequences of Bacteriocin-Producing *Streptococcus thermophilus* Strains ST106 and ST109

John A. Renye, Jr.,[a] David S. Needleman,[b] Dennis H. Steinberg[a]

[a]Dairy and Functional Foods Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, Pennsylvania, USA
[b]Office of the Center Director, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, Pennsylvania, USA

ABSTRACT *Streptococcus thermophilus* strains ST106 and ST109 produce broad-spectrum bacteriocins encoded within a bacteriocin-like peptide (*blp*) gene cluster. This study reports the complete genome sequences for both strains, with the ST109 chromosome containing 1,788,866 nucleotides (nt) and 1,572 predicted genes, and ST106 having 1,856,083 nt and 1,601 predicted genes.

Streptococcus thermophilus is routinely isolated from raw milk and has been designated a generally recognized as safe (GRAS) bacterium due to its extensive use in the production of cheese and yogurt. Some *S. thermophilus* strains have been investigated for food safety and probiotic applications due to their production of broad-spectrum bacteriocins (1). In several strains, bacteriocins are encoded at a chromosomal locus which resembles the bacteriocin-like peptide (*blp*) gene cluster in *Streptococcus pneumoniae* (2). The *blp* gene cluster has been characterized for *S. thermophilus* strains LMD-9 (3), ST106, and ST118 (4), where bacteriocin production was induced with a recombinant or synthetic quorum sensing peptide, BlpC-30, and for strain B59671 (5, 6), where production occurs naturally.

Here, we report the complete genome sequences for *S. thermophilus* strains ST106 and ST109, raw milk isolates from our in-house culture collection. The strains have previously been reported to produce the broad-spectrum bacteriocins thermophilin 106 and thermophilin 109, respectively (4, 7). *S. thermophilus* cultures were grown overnight in tryptone-yeast extract-lactose (TYL) medium, and genomic DNA was isolated from mutanolysin-pretreated cells using the DNeasy blood and tissue kit (Qiagen). Genomes were sequenced using a Pacific Biosciences RS II system (PacBio, Menlo Park, CA) by the Genomics Core Facility, Clinical and Translational Research Institute at Drexel College of Medicine (Philadelphia, PA). P6-C4 PacBio chemistry was performed using one single-molecule real-time (SMRT) cell per genome (8), producing 260,568 reads (mean subread length, 3,368 nucleotides [nt]; N_{50} value, 4,826 nt) and 284,854 reads (mean subread length, 3,549 nt; N_{50} value, 4,973 nt) for ST106 and ST109, respectively. *De novo* assembly was achieved using the HGAP assembler (version 2.3), estimating a genome size of 2 Mb, which yielded a single contig for strain ST109, with a mean coverage of 565-fold, and a single contig for ST106, with a mean coverage of 469-fold. Both chromosomes were circularized using Circlator (version 1.0.2) and resequenced with RS_Modification_and _Motif_Anlysis (version 2.3) using default settings (9). The genomes were analyzed with Taxator-*tk* (version 1.2) (10) using the nonredundant-microbial_20140513 database (http://research.bifo.helmholtz-hzi.de/software), the megan-lca algorithm, and custom Python scripts for postprocessing. The results confirmed the strain identities and showed that the fully assembled ST109 genome consisted of 1,788,866 nt with 39.2% G+C content, and ST106 contained 1,856,083 nt with 39.3% G+C content.

Genome annotation using the NCBI Prokaryotic Genome Annotation Pipeline ver-

Citation Renye JA, Jr, Needleman DS, Steinberg DH. 2019. Complete genome sequences of bacteriocin-producing *Streptococcus thermophilus* strains ST106 and ST109. Microbiol Resour Announc 8:e01336-18. https://doi.org/10.1128/MRA.01336-18.

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to John A. Renye, Jr., john.renye@ars.usda.gov.

Received 30 September 2018
Accepted 4 January 2019
Published 14 February 2019

sion 4.6 (11, 12) showed a total of 1,572 and 1,601 protein-encoding genes in ST109 and ST106, respectively, with both strains possessing 67 tRNAs and 6 rRNAs (5S, 16S, and 23S rRNA). Similar to other *S. thermophilus* strains (6, 13), the ST109 genome contained 13.5% nonfunctional genes; however, ST106 possessed 323 pseudogenes, corresponding to 16.8% of the predicted coding sequences. Previous studies showed that thermophilin 109 was naturally produced and may be encoded within the *blp* gene cluster (4, 7). The Clone Manager software version 9 (Sci-Ed Software, NC) was used to compare the chromosomal locus flanked by *blpA* and *blpX,* which consisted of 13,350 nt in ST106 and 13,332 nt in ST109. Alignment of the two loci revealed 18 mismatched nucleotides. The completed ST106 and ST109 genomes are expected to provide additional information needed to understand the molecular mechanisms regulating the production of thermophilins 106 and 109.

Data availability. The complete genome sequences for *S. thermophilus* ST106 and ST109 have been deposited in GenBank under the accession numbers CP031881 and CP031545 and in the NCBI SRA database under the accession numbers SRP162362 and SRP155251, respectively.

ACKNOWLEDGMENTS

We thank Garth Ehrlich and staff from Drexel College of Medicine for their assistance in sequencing and analysis of the genomes.

The mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

REFERENCES

1. Rossi F, Marzotto M, Cremonese S, Rizzotti L, Torriani S. 2013. Diversity of *Streptococcus thermophilus* in bacteriocin production; inhibitory spectrum and occurrence of thermophilin genes. Food Microbiol 35:27–33. https://doi.org/10.1016/j.fm.2013.02.006.
2. de Saizieu A, Gardes C, Flint N, Wagner C, Kamber M, Mitchell TJ, Keck W, Amrein KE, Lange R. 2000. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol 182:4696–4703.
3. Fontaine L, Boutry C, Guedon E, Guillot A, Ibrahim M, Grossiord B, Hols P. 2007. Quorum-sensing regulation of the production of Blp bacteriocins in *Streptococcus thermophilus*. J Bacteriol 189:7195–7205. https://doi.org/10.1128/JB.00966-07.
4. Somkuti GA, Renye JA, Jr. 2014. Effect of BlpC-based quorum-sensing induction peptide on bacteriocin production in *Streptococcus thermophilus*. J Food Res 4:88–96. https://doi.org/10.5539/jfr.v4n1p88.
5. Renye JA, Jr, Somkuti GA. 2013. BlpC-regulated bacteriocin production in *Streptococcus thermophilus*. Biotechnol Lett 35:407–412. https://doi.org/10.1007/s10529-012-1095-0.
6. Renye JA, Jr, Needleman DS, Somkuti GS, Steinberg DH. 2018. Complete genome sequence of Streptococcus thermophilus strain B59671, which naturally produces the broad-spectrum bacteriocin thermophilin 110. Genome Announc 5:e01213-17. https://doi.org/10.1128/genomeA.01213-17.
7. Renye JA, Jr, Somkuti GA, Garabal JI, Steinberg DH. 2016. Bacteriocin production by *Streptococcus thermophilus* in complex growth media. Biotechnol Lett 38:1947–1954. https://doi.org/10.1007/s10529-016-2184-2.
8. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
9. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circulator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
10. Dröge J, Gregor I, McHardy AC. 2015. Taxator-tk: precise taxonomic assignment of metagenomes by fast approximation of evolutionary neighborhoods. Bioinformatics 31:817–824. https://doi.org/10.1093/bioinformatics/btu745.
11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin 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.
12. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: and update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.
13. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, Fonstein M, Overbeek R, Kyprides N, Purnelle B, Prozzi D, Ngui K, Masuy D, Hancy F, Burteau S, Boutry M, Delcour J, Goffeau A, Hols P. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat Biotechnol 22:1554–1558. https://doi.org/10.1038/nbt1034.