Complete Genome Sequence of *Streptococcus ruminantium* sp. nov. GUT-187[T] (=DSM 104980[T] =JCM 31869[T]), the Type Strain of *S. ruminantium*, and Comparison with Genome Sequences of *Streptococcus suis* Strains

Mari Tohya[1], Tsutomu Sekizaki[2], and Tohru Miyoshi-Akiyama[1,]*

[1]Pathogenic Microbe Laboratory, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan

[2]Research Center for Food Safety, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

*Corresponding author: E-mail: takiyam@ri.imcj.go.jp.

Abstract

Streptococcus ruminantium sp. nov. of type strain GUT-187[T], previously classified as *Streptococcus suis* serotype 33, is a recently described novel streptococcal species. This study was designed to determine the complete genome sequence of *S. ruminantium* GUT-187[T] using a combination of Oxford Nanopore and the Illumina platform, and to compare this sequence with the genomes of 27 *S. suis* representative strains. The genome of GUT-187[T] was 2,090,539 bp in size, with a GC content of 40.01%. This genome contained 1,961 predicted protein coding DNA sequences (CDSs); of these, 1,685 (85.9%) showed similarity with *S. suis* CDSs. Of the remaining 276 CDSs, 81 (29.3%) showed some degree of similarity with CDSs of other streptococcal species. The genome of GUT-187[T] contained no intact prophage. The numbers of prophages and CRISPR spacers, as well as the presence or absence of genes encoding CRISPR-associated proteins, differed in *S. ruminantium* and *S. suis*. A phylogenetic analysis indicates that GUT-187[T] may be outgroup to the *S. suis* strains in our sample, thereby justifying its classification as distinct species. Gene mapping indicated 10.2 times of massive genome rearrangements in average occurred between *S. ruminantium* and *S. suis*. There was no significant statistical difference in clusters of orthologous group distribution between *S. ruminantium* and *S. suis*.

Key words: *Streptococcus ruminantium*, *Streptococcus suis*, complete genome sequence, comparative genomics, novel species.

Introduction

Streptococcus ruminantium is a recently described novel streptococcal species (Tohya et al., 2017). Although *S. ruminantium* was previously recognized to be *Streptococcus suis* serotype 33 (Tohya et al., 2017), several studies analyzing the taxonomic status of *S. suis* suggested that several serotype reference strains, including those for *S. suis* serotype 33, differed from the *S. suis* taxon, suggesting that these strains were not authentic *S. suis* (Tien le et al., 2013; Ishida et al., 2014; Arai et al., 2015). To further clarify that *S. ruminantium* and *S. suis* are distinct species, we sequenced the complete genome of the type strain *S. ruminantium* GUT-187[T] (=DSM 10498[OT]=JCM 31869[T]) and compared its sequence with those of 27 *S. suis* representative strains.

Materials and Methods

Determination of the Whole Genome Sequence of *S. ruminantium* GUT-187[T]

The GUT-187[T] genome was prepared using enzymatic lysis methods as described previously (Nishijima et al., 2016) and subjected to MinION genome sequencing using a flow cell (R9.5) (Oxford NANOPORE). Libraries were prepared using Rapid sequencing kit R9 version (Oxford NANOPORE).

Complete Genome Sequence of *S. ruminantium* sp. nov. GUT-187^T

GBE
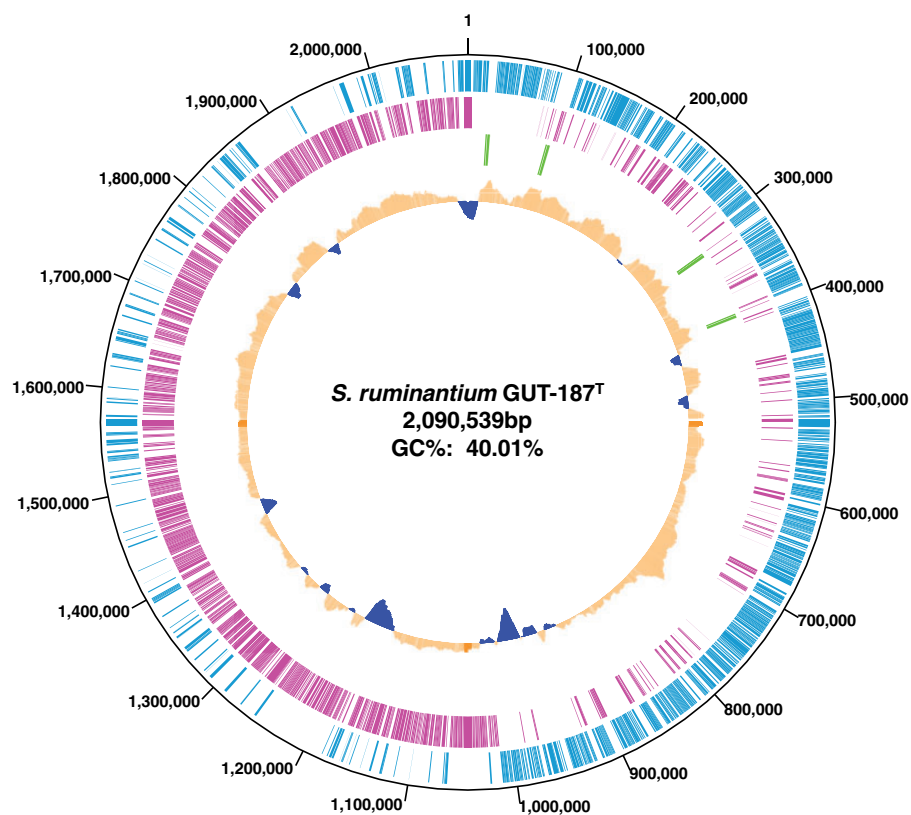
Fig. 1.—Circular representation of the genome of *Streptococcus ruminantium* sp. nov. GUT-187^T. Circle 1 (outermost circle) indicates the distance from the putative origin of replication. Circle 2 shows annotated CDSs encoded on the forward (light blue) and reverse (yellow) chromosomal strands. The rRNA genes (green) are shown in circle 3. Circle 4 (innermost circle) shows the G + C content with greater and less than average (0.40) in blue and orange, respectively.

Illumina sequencing was performed with Nextera XT to prepare a sequencing library, followed by MiSeq sequencing (Illumina), which yielded 301 bp pair end reads. Both procedures were performed according to the manufacturers' instructions. Approximately 300 Mbp of Nanopore data and 598,228 pair end reads of Illumina were used for genome assembly.

Bioinformatics Analyses

Genome de novo assembly was performed using SPAdes (Esgleas et al., 2005) in hybrid and careful mode, resulting in two contigs. Gaps between contigs were filled by a standard PCR method using LATaq (Takara). Amplicons were sequenced by MiSeq, as described above. CDSs were identified and annotated by glimmer (Delcher et al., 2007), using commercial software (in silico Molecular Cloning; in silico biology, Japan). The genome sequence was registered at the DNA Data Bank of Japan (DDBJ) under accession number AP018400. Multi-locus sequence typing (MLST) was determined using *S. suis* MLST Databases (https://pubmlst.org/ssuis/; King et al., 2002). Prophages in the genomes were analyzed using PHAST (http://phast.wishartlab.com; Zhou et al., 2011). CRISPRs were detected using CRISPR finder

(http://crispr.i2bc.paris-saclay.fr/Server; Grissa et al., 2007). Genome rearrangement maps were created using in silico Molecular Cloning software (in silico biology). The presence or absence of genes encoding CAS proteins in the genomes was analyzed using the TBlastX program and a custom streptococcal CAS protein database. Concatenated SNP sequences were aligned with MAFFT (Katoh et al. 2017). A Neighbor-Joining phylogenetic tree (Saitou and Nei 1987) was estimated using CLC genomics workbench (QIAGEN), a commercial software. The tree was midpoint rooted (Graur 2016). Gene mapping to analyze genome rearrangements and clusters of orthologous groups (COG) analysis was performed using in silico Molecular Cloning software (in silico biology). The proportions of COG categories were analyzed using CLC genomics workbench (QIAGEN).

Results

We utilized Oxford Nanopore and the Illumina platform to determine the complete genome sequence of GUT-187^T. The hybrid assembly approach using data from both platforms was highly effective, as well as being less costly than other platforms.

Fig. 2.—Rooted phylogenetic tree of *Streptococcus ruminantium* and *Streptococcus suis*. The Neighbor-Joining phylogenetic tree was estimated using CLC genomics workbench (QIAGEN), a commercial software. Tip labels are aligned. The genetic distances between the major nodes and bootstrap values are shown. Strains harboring CRISPR-associated proteins (CAS) are indicated in boxes.

The GUT-187[T] genome was 2,090,539 bp in size, with a GC content of 40.01%, comparable to those of the 27 *S. suis* representative strains in our study. The genome of GUT-187[T] contained 1,961 predicted protein coding DNA sequences (CDSs) (fig. 1, supplementary table S1, Supplementary Material online), comparable to the mean ± SD number of CDs in the 27 *S. suis* representative strains (1972.0 ± 88.4). Of the 1,961 CDSs in GUT-187[T], 1,685 (85.9%) were homologous with CDSs of *S. suis*. Among the 276 remaining CDSs, 81 (29.3%) showed some degree of similarity with CDSs of other streptococcal species.

MLST of GUT-187[T] was unique when compared with those of the *S. suis* representative strains (supplementary table S1, Supplementary Material online). The GUT-187[T] genome contained no intact prophages and only one remnant of prophage, located in the 97,740–108,427 bp region. The numbers of prophages and CRISPR spacers, as well as the presence or absence of genes encoding CRISPR-associated proteins (CAS) proteins, differed between *S. ruminantium*

and *S. suis*, with no associations between the presence or absence of CRISPR and CAS and the number of prophages although it is well known that CRISPR-CAS system counteracts invasion of foreign genetic materials (supplementary table S1, Supplementary Material online; Marraffini and Sontheimer 2010).

The phylogenetic analysis in conjunction with midpoint rooting supports the hypothesis that GUT-187[T] belong to a spices other than *S. suis* (fig. 2). Notably, the *S. suis* strains harboring CAS clustered together in the phylogenetic tree and GUT-187[T] also harbored CAS.

BLASTN analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) using the complete genome sequence of GUT-187[T] as a query showed that the top 37 hits were *S. suis*, while the query coverage was around 50% (supplementary table S2, Supplementary Material online). When the genome sequence of *S. suis* 6407 (Accession # CP008921.1) was used as a query, the top 35 hits were *S. suis* strains and

Complete Genome Sequence of *S. ruminantium* sp. nov. GUT-187[T]

GBE
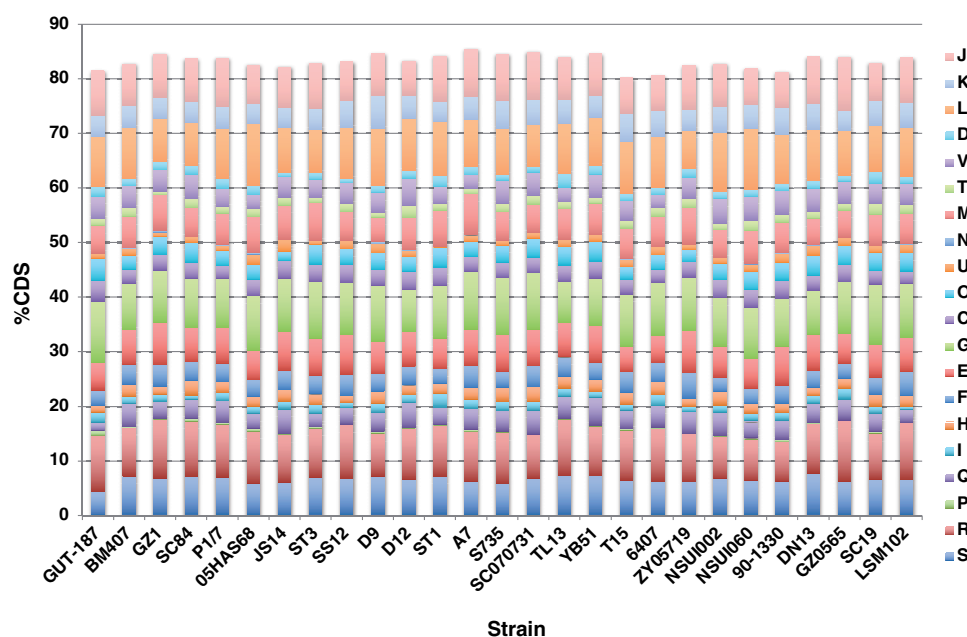
FIG. 3.—Comparison of GUT-187[T] with the *Streptococcus suis* representative strains using clusters of orthologous groups (COG). Category abbreviations: J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; D, cell cycle control, cell division, and chromosomal partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, and chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, biosynthesis, transport, and catabolism of secondary metabolites; R, general function predicted only; S, function unknown. *Streptococcus suis SS2-1* was omitted from this analysis because its genome information had not been annotated at the time of analysis.

the query coverage ranged from 72% to 83% (data not shown), indicating a taxonomic difference between *S. ruminantium* and *S. suis*.

Mapping of genes in the GUT-187[T] genome against *S. suis* to analyze genome rearrangements showed that massive inversions had occurred at around 450 kbp and 1,550 kbp in all *S. suis* strains except strains NSUI060 and 90-1330 (supplementary table S3, Supplementary Material online). Detailed investigation of the inversion points indicated that number of inversions in the *S. suis* strains in comparison with GUT-187[T] are 10.2 ± 2.3 (mean ± SD, ranged from 6 to 15, supplementary table S3, Supplementary Material online). The inversion regions of *S. ruminantium* is different from the insertion point of prophage (97,740–108,427 bp region, locus_tag ranged SR187_0490 from to SR187_0525), and there are no obvious mobile genetic elements, such as insertion sequences or transposons in the inversion regions (data not shown). Thus, we are not able to show direct evidence how the rearrangement occurred.

The ratio of each category of COG was found to be relatively conserved in the *S. ruminantium* and *S. suis* strains (fig. 3). There were no significant differences in the proportions of COG categories between GUT-187 and *S. suis* strains or between CAS positive and negative strains (data not shown). COG categories J, L, G, and R and, to a lesser extent,

COG categories M, E, and S, were relatively abundant among the strains.

In conclusion, the complete genome sequence of *S. ruminantium* further supports its classification as a distinct species. The sequence data may also enable the development of methods to analyze its epidemiology, as well as rapid diagnostic assays.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Arai S, et al. 2015. Development of loop-mediated isothermal amplification to detect *Streptococcus suis* and its application to retail pork meat in Japan. Int J Food Microbiol. 208:35–42.

Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23(6):673–679.

Esgleas M, Lacouture S, Gottschalk M. 2005. *Streptococcus suis* serotype 2 binding to extracellular matrix proteins. FEMS Microbiol Lett. 244(1):33–40.

Graur D. 2016. Molecular and Genome Evolution EBook. Basingstoke, UK: Palgrave Macmillan.

Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35(Web Server):W52–W57.

Ishida S, et al. 2014. Development of an appropriate PCR system for the reclassification of *Streptococcus suis*. J Microbiol Methods. 107:66–70.

Katoh K, Rozewicki J, Yamada KD. 2017 Sep 6. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinformatics. doi: 10.1093/bib/bbx108.

King SJ, et al. 2002. Development of a multilocus sequence typing scheme for the pig pathogen *Streptococcus suis*: identification of virulent clones and potential capsular serotype exchange. J Clin Microbiol. 40(10):3671–3680.

Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: rNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 11(3):181–190.

Nishijima S, et al. 2016. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. DNA Res. 23(2): 125–133.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4, 406–425

Tien le HT, Nishibori T, Nishitani Y, Nomoto R, Osawa R. 2013. Reappraisal of the taxonomy of *Streptococcus suis* serotypes 20, 22, 26, and 33 based on DNA–DNA homology and sodA and recN phylogenies. Vet Microbiol. 162(2–4):842–849.

Tohya M, et al. 2017. Defining the taxonomic status of *Streptococcus suis* serotype 33: the proposal for *Streptococcus ruminantium* sp. nov. Int J Syst Evol Microbiol. 67(9):3660–3665.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39(Suppl):W347–W352.

Associate editor: Dan Graur