Complete Genome Sequence of *Streptococcus dysgalactiae* subsp. *equisimilis* 167 Carrying Lancefield Group C Antigen and Comparative Genomics of *S. dysgalactiae* subsp. *equisimilis* Strains

Shinya Watanabe, Teruo Kirikae, and Tohru Miyoshi-Akiyama*

Department of Infectious Diseases, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan

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

Accepted: July 31, 2013

Data deposition: The nucleotide sequence of the chromosome of SDSE 167 has been deposited in the DNA Database of Japan under accession AP012976.

Abstract

*Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) is an emerging human pathogen that causes life-threatening invasive infections such as streptococcal toxic shock syndrome. Recent epidemiological studies reveal that invasive SDSE infections have been increasing in Asia, Europe, and the United States. Almost all SDSE carry Lancefield group G or C antigen. We have determined the complete genome sequence of a human group C SDSE 167 strain. A comparison of its sequence with that of four SDSE strains, three in Lancefield group G and one in Lancefield group A, showed approximately 90% coverage. Most regions showing little or no homology were located in the prophages. There was no evidence of massive rearrangement in the genome of SDSE 167. Bayesian phylogeny using entire genome sequences showed that the most recent common ancestor of the five SDSE strains appeared 446 years ago. Interestingly, we found that SDSE 167 harbors sugar metabolizing enzymes in a unique region and streptodornase in the phage region, which presumably contribute to the degradation of host tissues and the prompted covRS mutation, respectively. A comparison of these five SDSE strains, which differ in Lancefield group antigens, revealed a gene cluster presumably responsible for the synthesis of the antigenic determinant. These results may provide the basis for molecular epidemiological research of SDSE.

Key words: *Streptococcus dysgalactiae* subsp. *equisimilis*, Lancefield group C, complete genome sequence, Bayesian phylogeny.

Introduction

*Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) belongs to Lancefield group G or C (or, more rarely, A) streptococci (Vandamme et al. 1996; Takahashi et al. 2011). Although previously considered much less pathogenic in humans than group A (GAS: *Streptococcus pyogenes*) and group B streptococci, SDSE has been increasingly reported to cause invasive infections, such as sepsis and streptococcal toxic shock-like syndrome (Sylvestky et al. 2002; Cohen-Poradosu et al. 2004; Liao et al. 2008). Recent epidemiological studies have shown that SDSE contributes significantly to the burden of invasive infections caused by β-hemolytic streptococci (Takahashi et al. 2011).

Genome analyses of SDSE and targeted microarray analyses of GAS virulence genes in 58 SDSE strains isolated from infected humans have shown that clinically isolated SDSE strains have many of the important virulence factors present in GAS, including streptolysin S (SLS), streptolysin O (SLO), streptokinase, and antiphagocytic surface M proteins (Davies et al. 2007). Several important GAS virulence factors, however, are missing from the SDSE genome, including SpeB, a chromosomally encoded cysteine protease, and a hyaluronic acid capsule. Moreover, SDSE lacks superantigenic activity because its speG gene, encoding a superantigen homolog, does not show superantigenic activity against human peripheral mononuclear cells (Zhao et al. 2007).

In this study, we determined, for the first time, the complete genome sequence of a group C SDSE 167 strain isolated from a human patient and shown, using a mouse model, to be the most virulent strain. The 167 genome was compared...
with the complete genome sequences of four SDSE strains, three in Lancefield group G and one in Lancefield group A.

Materials and Methods

Bacterial Strains and Virulence in Mice (or Pathogenicity against Mice)

SDSE 167 strain was isolated from a patient with an invasive infection in Japan; the other completely sequenced strains described are listed in table 1. SDSE was cultured in 5% sheep blood agar or brain–heart infusion medium at 37°C under 5% CO₂ as described (Miyoshi-Akiyama et al. 2003a). The virulence of these SDSE strains listed in supplementary table S1, Supplementary Material online, was compared using a mouse i.p. infection model (Miyoshi-Akiyama et al. 2003b). Protocols of all animal experiments were approved by the ethical committee of the National Center for Global Health and Medicine based on the “Basic Guidelines for Proper Conduct of Animal Testing and Related Activities in the Research Institutions under the Jurisdiction of the Ministry of Health, Labour and Welfare (MHLW) of Japan.”

Preparation of Genomic DNA and Genome Sequencing

Streptococci were lysed as described (Miyoshi-Akiyama et al. 2003a), and genomic DNA was purified using DNeasy Blood & Tissue kits (QIAGEN). An 8-kb pair-end library of the SDSE 167 genome was prepared and sequenced using GS junior according to the manufacturer’s instruction (Roche). This generated 230,950 reads, covering 41,119,010 bp (19.8-fold coverage), which were assembled into scaffolds and contigs. Gap filling was performed by conventional Sanger sequencing of the polymerase chain reaction (PCR) fragments based on brute force PCR among the contigs and scaffolds. The nucleotide sequence of the chromosome of SDSE 167 has been deposited in the DNA Database of Japan under accession no. AP012976.

In Silico Analyses

MetaGeneAnnotator was used for primary CDS extraction (Noguchi et al. 2008), with initial functional assignment and manual correction performed by in silico molecular cloning (in silico biology, inc.). Prophage regions and clustered regularly interspaced short palindromic repeats (CRISPRs) were identified by Prophage Finder (Bose and Barber 2006) and CRISPRFinder (Grissa et al. 2007), respectively.

Phylogenetic Analyses

Whole-genome sequences were aligned with MAFFT (Katoh and Standley 2013). The evolutionary model (simple HYK) was chosen based on the results obtained with jModelTest 2.1.2 (Darriba et al. 2012) and convergence of the tree during preliminary phylogenetic analyses. A post-probable phylogenetic tree was constructed from genome sequence alignment with BEAST (Drummond and Rambaut 2007). BEAST was also used to estimate time from the most recent appearance of a common ancestor. The sequence alignments used are available from the corresponding author upon request.

Table 1

| Strain       | 167   | AC-2713 | ATCC 12394 | GGS_124 | RE378 |
|--------------|-------|---------|------------|---------|-------|
| Length (nt)  | 2,076,397 | 2,179,445 | 2,159,491  | 2,106,340 | 2,151,145 |
| G + C%       | 39.57 | 39.52   | 39.5   | 39.58  | 39.49 |
| rRNA operon  | 5     | 5       | 5      | 5      | 5     |
| tRNA         | 56    | 57      | 57     | 57     | 56    |
| emm          | stL839| stG485.0| stG166b.0| stg480.0| stg6792|
| Lancefield C | C     | A       | G      | G      |       |
| Acc. no.     | AP012976 | NC_019042 | NC_017567 | NC_012891 | NC_018712 |
| Isolation year | 2003  | 1999    | 1939   | 2006   | 2007 |
| Locus tag of putative enzyme | SDSE167_0822 to SDSE167_0826 | SDSE12394_04095 to SDSE12394_04110 | SDEG_0759 to SDEG_0762 | GGS_0731 to GGS_0734 |
| The Blast hits | Group C SDSE, Streptococcus equi subsp. equi, Streptococcus equi subsp. zooepidemicus | Group A SDSE, S. pyogenes | Group G SDSE | Group G SDSE |
| Reference    | This study | Suzuki et al. 2011 | Brandt et al. 1999 | Shimomura et al. 2011 | Yoshida et al. 2011 |

**Note:** Overview of SDSE strains used for comparisons with the 167 genome.
PCR Analysis

Conventional PCR to analyze the distribution of genes identified in this study was performed using TAKARA LA Taq according to the manufacturer’s instruction (TAKARA BIO Inc.). Primers used to amplify the corresponding genes are listed in supplementary table S2, Supplementary Material online.

Results and Discussion

SDSE 167, carrying Lancefield group C antigen, was isolated from an invasively infected human patient in 2003. We found that it was the most virulent SDSE strain isolated with an LD$_{50}$ of $9.6 \times 10^5$ CFU/mouse in our SDSE collection having LD$_{50}$ values ranging from $9.6 \times 10^5$ to $4.5 \times 10^7$ CFU/mouse (supplementary table S1, Supplementary Material online).

The SDSE 167 genome consists of a single circular chromosome of 2,076,397 bp with an average GC content of 39.57% (fig. 1 and table 1). The chromosome was shown to contain a total of 2223 protein-encoding genes, and 56 tRNA genes for all amino acids. In addition, the chromosome harbors two prophage-like elements (fig. 1 and table 2). To analyze evolutionary relationship of SDSE 167 with other SDSE strains, whole genome data of the five SDSE strains listed in table 1, including 167, were compared. Genome coverage analysis using the Blast algorithm indicated that approximately 90% of genome is shared among the five SDSE strains (supplementary table S3, Supplementary Material online).

![Circular representation of the genome of Streptococcus dysgalactiae subsp. equisimilis strain 167.](image-url)

**Fig. 1.**—Circular representation of the genome of Streptococcus dysgalactiae subsp. equisimilis strain 167. Circle 1 (outermost circle) indicates the distances from the putative origin of replication. Circles 2 and 3 show annotated CDS encoded by the forward (light blue) and reverse (pink) chromosomal strands, respectively. Circle 4 shows the rrs operons. Circle 5 shows prophages (green). Circle 6 shows unique regions found in 167 other than in phages, including regions encoding enzymes involved in sugar transfer and sugar metabolism and the FCT region. Circle 7 (innermost circle) shows the G+C content with more and less than average (0.40) in purple and brown, respectively.
| Strain       | Phage No. | Length | Best Hit of Blast                                                                 |
|-------------|-----------|--------|----------------------------------------------------------------------------------|
| 167         | φ1        | 37,972 | Streptodornase (Sdn)                                                              |
|             |           |        | Putative cell wall hydrolase, lysin                                              |
|             |           |        | Putative holin                                                                    |
|             |           |        | Putative hyaluronidase                                                             |
|             |           |        | Head maturation protease                                                           |
|             |           |        | Site-specific recombinase                                                          |
|             |           |        | Putative transcriptional activator                                                |
|             |           |        | Putative C5 methylase MarMP1                                                        |
|             |           |        | Single-strand binding protein                                                      |
|             |           |        | Putative replisome organizer                                                       |
|             | φ2        | 18,822 | Defective                                                                        |
|             | φ1        | 10,880 | gp44 clamp loader subunit                                                          |
|             |           |        | Site-specific recombinase                                                          |
|             |           |        | Putative cell wall hydrolase, lysin                                              |
|             |           |        | Holin                                                                            |
|             |           |        | Putative platelet-binding protein                                                 |
|             |           |        | ClpP-like protease                                                                |
|             |           |        | Putative portal protein                                                           |
|             |           |        | Putative DNA methylase                                                            |
|             |           |        | Transferase                                                                       |
|             |           |        | Putative helicase                                                                 |
|             |           |        | Putative DNA polymerase A domain                                                   |
| AC-2713     | φ4        | 38,710 | DNA cytosine methylase                                                            |
|             |           |        | Transcriptional regulator                                                          |
|             |           |        | Putative repressor protein                                                         |
|             | φ3        | 5,854  | Defective                                                                        |
| ATCC12394   | φ1        | 10,872 | gp44 clamp loader subunit                                                          |
|             | φ2        | 11,328 | DNA polymerase accessory protein                                                   |
|             | φ3        | 11,990 | IMPB                                                                             |
|             | φ4        | 28,611 | Putative DNA polymerase III delta prime subunit                                   |

(continued)
Genome rearrangement analysis showed no evidence of massive recombination among these SDSE strains (fig. 2a). Some regions showing diversity are located in the prophage regions, as omitting the prophage regions from analysis resulted in decreased rearrangements (fig. 2b). The relative stability of the whole genomes of the SDSE strains allowed alignment and analysis of the phylogenetic evolution of SDSE using whole-genome sequences. This phylogenetic analysis using BEAST, which is designed to reconstruct evolutionary history over time from sampled DNA sequences using a post-probabilistic approach (Drummond and Rambaut 2007), indicated that the genetic distance of 167 is relatively far from the others, suggesting that the most recent common ancestor of all five SDSE strains appeared about 446 years ago (fig. 3a). Essentially the same results were obtained from analysis of core genome sequences, after omitting the phage regions (fig. 3b).

We compared prophage regions among the SDSE strains (table 2). SDSE 167 harbors two prophages, with a gene encoding streptodornase (sdzD), which presumably contribute to the prompted covRS mutation in vivo responsible for conversion of GAS into more virulent phenotype (Walker et al. 2007) found in prophage 1. The sdzD gene was shared among the SDSE isolates (supplementary table S2, Supplementary Material online). A comprehensive homology search of prophages found in SDSE showed that some of the prophages are shared, whereas the two prophages in 167 showed relatively low identity values, indicating that these prophages are essentially unique (supplementary table S4, Supplementary Material online). We also analyzed the genomes for the presence of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated sequence proteins (CASs) (Bhaya et al. 2011; Wiedenheft et al. 2012). CRISPR relies on small RNAs for sequence-specific detection and cleaving of foreign nucleic acids, including bacteriophages and plasmids. All SDSE strains sequenced previously harbor at least one probable CRISPR with more than 20 spacers. In contrast, newly sequenced 167 does not harbor any probable CRISPR or CAS. Thus, presumably there is no interference by the CRISPR system in 167, and SDSE 167 may be prone to infection by phages. Analysis of a larger number of group C SDSE isolates is necessary to determine whether the absence of CRISPR interference is common to these isolates (table 3).

Analysis of unique regions other than prophage regions in SDSE 167 showed that SDSE 167 harbors two unique gene clusters, which are not found in the other four SDSE strains. One cluster encodes glycosyl transferase and membrane proteins (table 2, locus_tag: SDSE167_0822 to SDSE167_0826). The finding, that this cluster is surrounded by other carbohydrate modifying enzymes, such as α-(1,2)-rhamnooligosaccharide transferase, and α-L-Rhaalpha-1,3-L-rhamnooligosaccharide transferase...
This region shows 65% identity with six genes of *Streptococcus mutans* (rgpA through rgpF), whose disruption results in a loss of serotype-specific antigenicity, specified by the glucose side chains of rhamnose–glucose polysaccharide from the cell wall (Yamashita et al. 1998), suggesting that this region may be involved in the synthesis of Lancefield group C antigen. Blast analyses of the cluster as well as the corresponding region of the other SDSE strains, three in Lancefield group G and one in Lancefield group A, indicated that these clusters showed identity with those of bacteria carrying group C, G, and A antigens, respectively (table 1). PCR analysis using the regions-specific primers showed that corresponding regions of each putative group antigen were carried by the SDSE isolates (supplementary table S1, Supplementary Material online). Although phenotypic analysis is necessary to elucidate the functional roles of these clusters, their sequences may be used in place of serotyping to identify Lancefield group antigens.

The other unique region found in SDSE 167 is located at SDSE167_0913 to 0915. This region did not show significant homology with genome of *Streptococcus pyogenes* except for first 1 kb region, which encodes two hypothetical proteins only at nucleic acid level. The remaining region shows weak identity with the PTS system, galactitol-specific IIIC component of *Enterococcus faecium* NRRL B-2354 and ribulose-phosphate 3-epimerase of *Streptococcus agalactiae* 09mas018883. This region contains sugar metabolizing enzymes, including tagatose-6-phosphate kinase, phosphoenolpyruvate-dependent sugar, the phosphotransferase system, the phosphotransferase system protein, the PTS system galactitol-specific IIIC component, PTS system galactitol-specific IIIC component, predicted protein, class II aldolase/adducin, and allulose-6-phosphate 3-epimerase. Our recent microarray results suggested that, upon injection into mouse peritoneal...
cavities, SDSE degrades host tissue polysaccharides by secreting poly/oligosaccharide lyases, while simultaneously using the Entner–Doudoroff pathway to metabolize acquired carbohydrates (Watanabe et al. 2013), and this region was found in 167 among the SDSE isolates (supplementary table S1, Supplementary Material online). Thus, this unique region containing sugar metabolizing enzymes may contribute to the higher virulence of SDSE 167.

In conclusion, we determined, for the first time, the complete genome sequence of a group C SDSE strain 167 and compared it with the genome sequences of other SDSE strains. Our results may provide insight into the pathogenic mechanism of SDSE and may form the basis of molecular epidemiological research on these highly virulent bacteria.

Supplementary Material
Supplementary tables S1–S4 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
The authors thank Mrs. Komiya and Mrs. Sakurai for excellent work in the genome analysis. This work was partly supported...
by JSPS KAKENHI Grant Number 24390109 (to T.M.A.) and 25860330 (to S.W.) and by a Grant for International Health Research (23A-301) from the Ministry of Health, Labour and Welfare of Japan (to T.K.).

**Literature Cited**

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 45:273–297.

Bose M, Barber RD. 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. In Silico Biol. 6:223–227.

Brandt CM, Haase G, Schnitzler N, Zbiden R, Lütticken R. 1999. Characterization of blood culture isolates of Streptococcus dysgalactiae subsp. equisimilis possessing Lancefield’s group A antigen. J Clin Microbiol. 37:4194–4197.

Cohen-Poradosu R, et al. 2004. Group G streptococcal bacteremia in Jerusalem. Emerg Infect Dis. 10:1455–1460.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

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

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Liao CH, Liu LC, Huang YT, Teng LI, Hsueh PR. 2008. Bacteremia caused by group G streptococci, Taiwan. Emerg Infect Dis. 14:837–840.

Miyoshi-Akiyama T, et al. 2003a. Streptococcus dysgalactiae-derived mitogen (SDM), a novel bacterial superantigen: characterization of its biological activity and predicted tertiary structure. Mol Microbiol. 47:1589–1599.

Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. 15:387–396.

Shimomura Y, et al. 2011. Complete genome sequencing and analysis of a Lancefield group G Streptococcus dysgalactiae subsp. equisimilis strain causing streptococcal toxic shock syndrome (STSS). BMC Genomics. 12:17.

Suzuki H, et al. 2011. Comparative genomic analysis of the Streptococcus dysgalactiae species group: gene content, molecular adaptation, and promoter evolution. Genome Biol Evol. 3:168–185.

Vandamme P, Pot B, Falsen E, Kersters K, Devriese LA. 1996. Taxonomic study of lancefield streptococcal groups C, G, and L (Streptococcus dysgalactiae) and proposal of S. dysgalactiae subsp. equisimilis subsp. nov. Int J Syst Bacteriol. 46:774–781.

Walker MJ, et al. 2007. DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. Nat Med. 13:981–985.

Watanabe S, Shimomura Y, Ubukata K, Kiriae T, Miyoshi-Akiyama T. 2013. Concomitant regulation of host tissue-destroying virulence factors and carbohydrate metabolism during invasive diseases induced by group G streptococci. J Infect Dis. Advance Access published July 30, 2013, doi: 10.1093/infdis/jit353.

Zhao J, et al. 2007. Cloning, expression, and characterization of the superantigen streptococcal pyrogenic exotoxin G from Streptococcus dysgalactiae. Infect Immun. 75:1721–1729.