JB Journals.ASM.org

Complete Genome Sequence of *Streptococcus mutans* GS-5, a Serotype *c* Strain

Saswati Biswas and Indranil Biswas

Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, Kansas, USA

***Streptococcus mutans*, a principal causative agent of dental caries, is considered to be the most cariogenic among all oral streptococci. Of the four *S. mutans* serotypes (*c*, *e*, *f*, and *k*), serotype *c* strains predominate in the oral cavity. Here, we present the complete genome sequence of *S. mutans* GS-5, a serotype *c* strain originally isolated from human carious lesions, which is extensively used as a laboratory strain worldwide.**

Streptococcus mutans, a colonizer of the supragingival tooth surface, is a part of a complex microflora comprising ~700 species. To maintain its dominant presence and to cause dental caries, this acidogenic and aciduric organism can drastically and quickly reduce the pH of its surroundings, leading to demineralization of tooth enamel (10). *S. mutans* also secretes antimicrobial peptides (mutacins) to suppress the growth of other competing species. Of the four serotypes of *S. mutans*, the predominant oral isolates are of serotype *c* (14). GS-5 is a representative serotype *c* strain originally isolated from carious lesions 45 years ago (7) and has been extensively used in genetic and biochemical studies of virulence (10). This strain also produces a unique two-peptide lantibiotic mutacin, known as SmbAB (17), which is encoded by only ~8% of the isolates (16). The recent completion of three *S. mutans* genome sequences (UA159, NN2025, and LJ23 [1, 2, 13]) indicates a large degree of diversity and genome rearrangement within the species. In the work presented here, we determined the complete genome sequence of GS-5, which will allow us to gain further insight into overall genetic variation in *S. mutans*.

Genome sequencing was performed using a combination of strategies that included Illumina GA-IIx and Roche GS junior technology at Cofactor Genomics (St. Louis, MO). The Illumina sequencing used two genomic libraries, SIPE (~300 bp) and LIMP (~3 kb), which generated ~21 million and ~15 million paired reads, respectively. Roche 454 GS junior generated 142,832 reads (average length, ~450 nucleotides) that provided ~26× coverage of the entire genome. Several assembly procedures were applied, and manual editing was performed with the 454 GS junior data. The remaining gaps and unassertive assembled regions were verified by PCR/ABI sequencing to obtain a single contig onto which Illumina data were subsequently mapped for further refinement. The genome was annotated using IGS Annotation Engine (5), and The NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) was employed for submission (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

The GS-5 genome encodes a 2,027,088-bp circular molecule with a G+C content of 36.8%. The sequence information is consistent with a previously generated physical map of the GS-5 genome (3). The putative origin of replication was mapped near nucleotide position 1,553 by OriFinder (6). GS-5 harbors 65 tRNA genes, 15 rRNA genes, and 1,985 coding sequences (CDS). Among the CDS, 91 exported proteins, 23 lipoproteins, 11 wall proteins, and 462 membrane proteins were identified by SLEP (8). No putative phages or prophages were predicted by Prophinder (12). However, two complete and one incomplete CRISPRs (clustered regularly interspaced short palindromic repeats) and numerous insertion sequences were identified with CRISPRFinder (9) and ISfinder (15), respectively. Furthermore, in addition to SmbAB, at least 10 other putative mutacins were identified by BAGEL2 (4). Several large (>1.0-kb) duplicated regions were identified in the genome by REPuter (11). Sequence alignment indicated that genome rearrangement occurred between GS-5 and NN2025 across the replication axis but not between GS-5 and UA159. The GS-5 complete genome will allow for in-depth comparative genomics to unravel the extent of genome rearrangements and dynamics in *S. mutans* and to better understand the adaptive life style of this pathogen.

Nucleotide sequence accession number. The complete genome sequence of *S. mutans* GS-5 was deposited in GenBank under accession number CP003686.

ACKNOWLEDGMENTS

We thank Michelle Giglio for helping us with the IGS Annotation Engine (http://ae.igs.umaryland.edu/cgi/index.cgi) and Jeff Banas (University of Iowa) for helpful discussions.

This work was supported in part by NIDCR grants DE016686 and DE021664 to I.B.

REFERENCES

1. **Aikawa C, et al.** 2012. Complete genome sequence of the serotype k *Streptococcus mutans* strain LJ23. J. Bacteriol. **194**:2754–2755.
2. **Ajdic D, et al.** 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. U. S. A. **99**:14434–14439.
3. **Cappiello MG, et al.** 1999. Physical and genetic map of *Streptococcus mutans* GS-5 and localization of five rRNA operons. Oral Microbiol. Immunol. **14**:225–232.
4. **de Jong A, van Heel AJ, Kok J, Kuipers OP.** 2010. BAGEL2: mining for bacteriocins in genomic data. Nucleic Acids Res. **38**:W647–W651.
5. **Galens K, et al.** 2011. The IGS standard operating procedure for automated prokaryotic annotation. Stand. Genomic Sci. **4**:244–251.
6. **Gao F, Zhang CT.** 2008. Ori-Finder: a web-based system for finding oriCs

Received 21 June 2012 Accepted 26 June 2012

Address correspondence to Indranil Biswas, ibiswas@kumc.edu.

doi:10.1128/JB.01106-12

in unannotated bacterial genomes. BMC Bioinformatics **9**:79. doi:10.1186/1471-2105-9-79.

7. **Gibbons RJ, Berman KS, Knoettner P, Kapsimalis B.** 1966. Dental caries and alveolar bone loss in gnotobiotic rats infected with capsule forming streptococci of human origin. Arch. Oral Biol. **11**:549–560.
8. **Giombini E, Orsini M, Carrabino D, Tramontano A.** An automatic method for identifying surface proteins in bacteria: SLEP. BMC Bioinformatics **11**:39. doi:10.1186/1471-2105-11-39.
9. **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.
10. **Kuramitsu H.** 1993. Virulence factors of mutans streptococci: role of molecular genetics. Crit. Rev. Oral Biol. Med. **4**:159–176.
11. **Kurtz S, et al.** 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29**:4633–4642.
12. **Lima-Mendez G, Van Helden J, Toussaint A, Leplae R.** 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. Bioinformatics **24**:863–865.
13. **Maruyama F, et al.** 2009. Comparative genomic analyses of *Streptococcus mutans* provide insights into chromosomal shuffling and species-specific content. BMC Genomics **10**:358. doi:10.1186/1471-2164-10-358.
14. **Nakano K, et al.** 2007. Serotype distribution of *Streptococcus mutans* a pathogen of dental caries in cardiovascular specimens from Japanese patients. J. Med. Microbiol. **56**:551–556.
15. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. **34**:D32–D36.
16. **Waterhouse JC, Russell RR.** 2006. Dispensable genes and foreign DNA in *Streptococcus mutans*. Microbiology **152**:1777–1788.
17. **Yonezawa H, Kuramitsu HK.** 2005. Genetic analysis of a unique bacteriocin, Smb, produced by *Streptococcus mutans* GS5. Antimicrob. Agents Chemother. **49**:541–548.