JB Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Genome Sequence of the Serotype *k Streptococcus mutans* Strain LJ23

Chihiro Aikawa,[a] Nayuta Furukawa,[a] Takayasu Watanabe,[a] Kana Minegishi,[b] Asuka Furukawa,[b] Yoshinobu Eishi,[b] Kenshiro Oshima,[c] Ken Kurokawa,[d] Masahira Hattori,[c] Kazuhiko Nakano,[e] Fumito Maruyama,[a] Ichiro Nakagawa,[a] and Takashi Ooshima[e]

Section of Bacterial Pathogenesis, Tokyo Medical and Dental University Graduate School of Medical and Dental Sciences, Tokyo, Japan[a]; Department of Human Pathology, Tokyo Medical and Dental University Graduate School of Medical and Dental Sciences, Tokyo, Japan[b]; Department of Computational Biology, The University of Tokyo Graduate School of Frontier Sciences, Kashiwa, Chiba, Japan[c]; Division of Information Biotechnology, Department of Bioinformation Engineering, Tokyo Institute of Technology School and Graduate School of Bioscience and Biotechnology, Midori-ku, Yokohama, Kanagawa, Japan[d]; and Department of Pediatric Dentistry, Osaka University Graduate School of Dentistry, Suita, Osaka, Japan[e]

Streptococcus mutans **is the major pathogen of dental caries and occasionally causes infective endocarditis. Here we report the complete genome sequence of serotype *k S. mutans* strain LJ23, which was recently isolated from the oral cavity of a Japanese patient.**

Streptococcus mutans, a major pathogen of dental caries, is classified into serotypes *c*, *e*, *f*, and *k* (17). Approximately 70 to 80% of oral isolates are serotype *c*, while the distribution frequency of serotype *k* is ~5% (12). Serotype *k* strains are not only dental caries pathogens but have been increasingly implicated in infective endocarditis (6, 11). Furthermore, it was recently reported that infection with serotype *k* strains expressing the collagen-binding protein (CBP) is a potential risk factor for hemorrhagic stroke in mice (10). However, the relationship between serotype and clinical conditions remains unclear. The complete genome sequences of bacteria might help to reveal such relationships, but so far, only the genomes of two *S. mutans* serotype *c* strains, NN2025 and UA159, have been fully sequenced (1, 9). Here we describe the first complete genome sequence of a serotype *k* strain, LJ23, isolated in Japan.

The complete genome sequence of strain LJ23 was determined by a combination of pyrosequencing (143,616,336-bp sequence, 71-fold coverage) and the Sanger method (35,189,307-bp sequence, 17-fold coverage). The pyrosequencing and Sanger reads were assembled using Newbler and Phred/Phrap/Consed, respectively. Gaps between adjacent contigs were closed by sequencing PCR amplicons from genomic DNA. Protein-coding sequences (CDSs) were predicted using a combination of MetaGeneAnnotator (14), GLIMMER (3), and the IMCGE software (In Silico Biology Co., Ltd., Japan). Annotation of CDSs was based on the results of BLASTP searches against the NCBI nonredundant protein database. Insertion sequences (ISs), conjugative transposons, and clustered regularly interspaced short palindromic repeats (CRISPRs) were identified using ISfinder (16), a combination of Mauve (2) and GenomeMatcher (15), and CRISPRFinder (5), respectively. Nontranslated genes were predicted using tRNAscan-SE (8), RNAmmer (7), and Rfam (4).

The genome of *S. mutans* LJ23 contains a single circular chromosome (2,015,626 bp, 37.05% GC content). The chromosome contains 1,921 CDSs, five rRNA operons, 65 tRNA sequences, 25 noncoding RNAs, 13 ISs, and two CRISPRs.

All-to-all BLASTP analysis with NN2025 and UA159 protein sequences showed that LJ23 possesses 80 strain-specific CDSs. It was previously shown that the distribution of the CBP-encoding *cnm* gene in clinical isolates was ~10% and predominant in serotype *k* or *f* strains (11, 13), and *cnm* was found to be an LJ23-specific gene. Approximately half of the strain-specific CDSs were annotated as hypothetical proteins, and 33 CDSs were on mobile elements. Several genes on one mobile element showed high homology with proteins produced by *Streptococcus pneumoniae*, suggesting that LJ23 may have acquired these CDSs via horizontal gene transfer from this species. Dot plot analysis comparison with the NN2025 and UA159 genome sequences indicated that genome rearrangements occurred between LJ23 and UA159 along the replication axis but not between LJ23 and NN2025. This may be due to the fact that UA159 was isolated in the United States, whereas both LJ23 and NN2025 were isolated in Japan. There remain many questions regarding how the oral environment affects bacterial pathogenicity under different clinical conditions. We believe that genomic analysis of LJ23 could lead to new insights into the mechanisms of pathogenicity of *S. mutans*.

Nucleotide sequence accession number. The complete genome sequence of *S. mutans* LJ23 was deposited in the DDBJ/EMBL/GenBank databases under accession no. AP012336. In addition, genome information for *S. mutans* LJ23 can be downloaded from our laboratory's website (http://www.tmd.ac.jp/grad/bac/).

ACKNOWLEDGMENTS

This work was supported by grants-in-aid for scientific research 21390497, 21390487, 22592032, and 22592078; by the Japanese Ministry of Education, Global Center of Excellence (GCOE) Program International Research Center for Molecular Science in Tooth and Bone Diseases, MEXT, Japan; and by a funding program for Next Generation World-Leading Researchers (LS041), Japan Society for the Promotion of Science.

REFERENCES

1. **Ajdić D, et al.** 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. U. S. A. **99**:14434–14439.
2. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple

Received 3 March 2012 **Accepted** 13 March 2012

Address correspondence to Fumito Maruyama, fumito-m.bac@tmd.ac.jp.

doi:10.1128/JB.00350-12

alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
3. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27**: 4636–4641.
4. **Griffiths-Jones S, et al.** 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. **33**:D121–D124.
5. **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.
6. **Hamada S, Slade HD.** 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. Microbiol. Rev. **44**:331–384.
7. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
8. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
9. **Maruyama F, et al.** 2009. Comparative genomic analyses of *Streptococcus mutans* provide insights into chromosomal shuffling and species-specific content. BMC Genomics **10**:358.
10. **Nakano K, et al.** 2011. The collagen-binding protein of S*treptococcus mutans* is involved in haemorrhagic stroke. Nat. Commun. **2**:485.
11. **Nakano K, et al.** 2007. *Streptococcus mutans* clonal variation revealed by multilocus sequence typing. J. Clin. Microbiol. **45**:2616–2625.
12. **Nakano K, Nomura R, Nakagawa I, Hamada S, Ooshima T.** 2004. Demonstration of *Streptococcus mutans* with a cell wall polysaccharide specific to a new serotype, *k*, in the human oral cavity. J. Clin. Microbiol. **42**:198–202.
13. **Nakano K, et al.** 2010. Molecular characterization of *Streptococcus mutans* strains containing the *cnm* gene encoding a collagen-binding adhesin. Arch. Oral Biol. **55**:34–39.
14. **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.
15. **Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M.** 2008. GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics **9**:376.
16. **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.
17. **Tsukioka Y, Yamashita Y, Oho T, Nakano Y, Koga T.** 1997. Biological function of the dTDP-rhamnose synthesis pathway in *Streptococcus mutans*. J. Bacteriol. **179**:1126–1134.