

Draft Genome Sequences of Two Clinical Isolates of *Streptococcus mutans*

Hui Zheng,^a Lihong Guo,^b Ning Du,^b Jiuxiang Lin,^a Lai Song,^c Guiming Liu,^c Feng Chen^b

Department of Orthodontics, Peking University School and Hospital of Stomatology, Beijing, China^a; Central of Laboratory, Peking University School and Hospital of Stomatology, Beijing, China^b; CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China^c

We report the draft genome sequences of PKUSS-HG01 and PKUSS-LG01, two clinical isolates of *Streptococcus mutans* from human dental plaque. The genomics information will facilitate the study of the mechanisms of pathogenicity and evolution of *S. mutans*.

Received 21 April 2014 Accepted 5 May 2014 Published 12 June 2014

Citation Zheng H, Guo L, Du N, Lin J, Song L, Liu G, Chen F. 2014. Draft genome sequences of two clinical isolates of *Streptococcus mutans*. *Genome Announc.* 2(3):e00441-14. doi:10.1128/genomeA.00441-14.

Copyright © 2014 Zheng et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Guiming Liu, liugm@big.ac.cn, or Feng Chen, moleculecf@gmail.com.

Streptococcus mutans is a Gram-positive organism that is the primary causative agent in the formation of dental caries in humans. *S. mutans* metabolizes a variety of carbohydrates, generating organic acid in dental plaque. The acidic environment results in mineral loss on the surface of teeth and eventually causes tooth decay (1). Furthermore, it has been reported that *S. mutans* is implicated not only in infective endocarditis (2) but also in hemorrhagic stroke (3). To better understand the molecular basis of mechanisms of pathogenicity, degree of diversity, and genome rearrangement within the species, we sequenced the genomes of two strains from human dental plaque.

Whole-genome sequencing was performed using an Illumina Hi-Seq 2000 platform with an insert size of 500 bp, according to the manufacturer's instructions. Paired sequences were generated, with total coverage of 315 and 310 for PKUSS-HG01 and PKUSS-LG01, respectively. For each genome, *de novo* assembly was performed using SOAPdenovo (4), and then the gaps within scaffolds were closed with GapFiller (5). Open reading frames (ORFs) were predicted with Glimmer 3.02 (6). Putative ribosomal binding sites were identified with RBSfinder (7). tRNAs and rRNAs were predicted using tRNAscan-SE (8) and RNAmmer (9), respectively. Functional annotations were performed by search against the NCBI nr, Swiss-Prot (10), Clusters of Orthologous Groups (COG) (11), KEGG (12), and InterProScan (13) databases.

We obtained 16 and 17 contigs larger than 500 bp for PKUSS-HG01 and PKUSS-LG01, respectively. The genomic sequence of PKUSS-HG01 comprises 1,951 coding sequences (CDSs), 3 rRNA loci, and 43 tRNAs, with a G+C content of 36.6%. In comparison, the genome sequence of PKUSS-LG01 contains 1,935 CDSs, 3 rRNA loci, and 39 tRNAs, with a G+C content of 36.6%.

Totals of 1,606 and 1,598 CDSs were assigned to COGs, and 1,064 and 1,057 CDSs can be annotated into 146 and 147 pathways by using KAAS (14) for PKUSS-HG01 and PKUSS-LG01, respectively. All-to-all BLASTP analysis with PKUSS-HG01 and PKUSS-LG01 protein sequences showed that they possess 1,865 orthologous proteins. These genome sequences will enable further study

of the mechanisms of pathogenicity and evolution and facilitate future drug development.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AXSW000000000](https://www.ncbi.nlm.nih.gov/nuclink/AXSW000000000) for strain PKUSS-HG01 and [AXSX000000000](https://www.ncbi.nlm.nih.gov/nuclink/AXSX000000000) for strain PKUSS-LG01. The versions described in this paper are versions AXSW01000000 and AXSX01000000.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (grants 81200762 and 40906097).

REFERENCES

1. Banas JA. 2004. Virulence properties of *Streptococcus mutans*. *Front. Biosci.* 9:1267–1277. <http://dx.doi.org/10.2741/1305>.
2. Hamada S, Slade HD. 1980. Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol. Rev.* 44:331–384.
3. Nakano K, Hokamura K, Taniguchi N, Wada K, Kudo C, Nomura R, Kojima A, Naka S, Muranaka Y, Thura M, Nakajima A, Masuda K, Nakagawa I, Speziale P, Shimada N, Amano A, Kamisaki Y, Tanaka T, Umemura K, Ooshima T. 2011. The collagen-binding protein of *Streptococcus* is involved in haemorrhagic stroke. *Nat. Commun.* 2:485. <http://dx.doi.org/10.1038/ncomms1491>.
4. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience.* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
5. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
6. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
7. Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. 2001. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 17:1123–1130. <http://dx.doi.org/10.1093/bioinformatics/17.12.1123>.
8. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detec-

- tion of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
9. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
 10. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31:365–370. <http://dx.doi.org/10.1093/nar/gkg095>.
 11. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29:22–28. <http://dx.doi.org/10.1093/nar/29.4.e22>.
 12. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36:D480–D484.
 13. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120. <http://dx.doi.org/10.1093/nar/gni118>.
 14. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185. <http://dx.doi.org/10.1093/nar/gkm321>.