

DATA NOTE

Open Access



Complete genome sequence of *Streptococcus hominis* isolated from subgingival biofilm

Seok Bin Yang^{1,2†}, Doyun Ku^{1,2†}, Ji-Hoi Moon¹, Jae-Hyung Lee¹, Sang Wook Kang³, Hak Kyun Kim^{4*} and Kyu Hwan Kwack^{1*}

Abstract

Objective *Streptococcus hominis* is a recently described species within the genus *Streptococcus*, yet its genomic characteristics remain poorly understood, particularly in the context of the oral microbiome. Previously, only two complete genomes from non-oral sources were available. To address this gap, we sequenced and analyzed *S. hominis* strain KHUD_010, isolated from the subgingival biofilm of a healthy Korean adult.

Data description Genomic DNA from KHUD_010 was extracted and confirmed as *S. hominis* by 16 S rRNA gene sequencing. Whole-genome sequencing using the PacBio Sequel II platform generated 135,974 HiFi reads (N50: 10,345 bp). *De novo* assembly with SMRT Link v11.0 produced a single circular chromosome of 1,883,665 bp with 39.04% GC content. Annotation via the NCBI Prokaryotic Genome Annotation Pipeline predicted 1,793 protein-coding genes, four rRNA operons (5 S, 16 S, 23 S), and 120 tRNAs. BUSCO analysis showed 99.1% completeness. Comparative genomics with NSJ-17 and UMB6992B revealed 1,416 core, 223 dispensable, and 398 strain-specific gene clusters. KHUD_010 harbored 18 unique gene clusters comprising 20 genes, mostly assigned to COG category L (replication, recombination, repair). This high-quality genome expands the genomic landscape of *S. hominis* and provides a valuable reference for future studies on oral microbiome diversity and host adaptation.

Keywords *Streptococcus hominis*, Genome, Subgingival biofilm

Objective

The human oral cavity harbors a complex microbial ecosystem in which *Streptococcus* species play key roles in both maintaining oral health and contributing to disease [1]. These bacteria are among the earliest tooth colonizers and are involved in biofilm formation, nutrient metabolism, and immune modulation [1, 2]. While species such as *S. mitis*, *S. oralis*, *S. sanguinis*, and *S. mutans* have been well characterized [3], large-scale metagenomic studies employing NGS have revealed a broader diversity of oral streptococci, including previously uncultured and uncharacterized taxa [4–7]. Although *Streptococcus* remains one of the most dominant genera in the oral microbiota, many of its constituent species remain

[†]Seok Bin Yang and Doyun Ku these authors contributed equally to this work.

*Correspondence:

Hak Kyun Kim
hakyun@cau.ac.kr
Kyu Hwan Kwack
dmdkwack@khu.ac.kr

¹Department of Oral Microbiology, College of Dentistry, Kyung Hee University, Seoul 02447, Republic of Korea

²Department of Dentistry, Graduate School, Kyung Hee University, Seoul 02447, Republic of Korea

³Department of Oral and Maxillofacial Pathology, College of Dentistry, Kyung Hee University, Seoul 02447, Republic of Korea

⁴Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea



genomically undefined, limiting our understanding of their ecological functions and clinical relevance [8].

Streptococcus hominis is a recently described species with only two complete genomes currently available, both from extra-oral sources (fecal and urinary isolates). To date, no oral isolates have been sequenced. In this study we report the complete genome sequence of *S. hominis* strain KHUD_010, isolated from the subgingival biofilm of a healthy Korean adult. Comparative analysis with the two existing strains, NSJ-17 (feces, China) [9] and UMB6992B (urine, USA) [10], provides insight into species-level diversity and strain-specific genomic traits. This work expands the genomic repertoire of *S. hominis* and lays the groundwork for future studies of its role in the oral microbiome.

Data description

The *S. hominis* strain KHUD_010 was isolated from the subgingival biofilm of a healthy Korean female adult using Mitis Salivarius agar under anaerobic conditions at 37°C. Genomic DNA was extracted using the Wizard HMW DNA Extraction Kit (Promega, USA). Species identification was confirmed by sequencing the full-length 16S rRNA gene (~1.5 kb) amplified with universal bacterial primers (forward: 5'-AGAGTTTGATCCTGGC TCAG-3'; reverse: 5'-GGTTACCTTGTACGACTT-3') [11] and performing a BLAST search against the GenBank database.

Whole-genome sequencing was conducted using the PacBio Sequel II platform. A Single-Molecule Real-Time (SMRT) bell library was prepared according to the manufacturer's protocol (Pacific Biosciences). One SMRT cell generated 135,974 HiFi reads, with an average read length of 9,396 bp and an N50 of 10,345 bp (Data file 1). *De novo* assembly was performed using the Microbial Genome Analysis pipeline in SMRT Link v11.0 (<https://www.pacb.com/support/software-downloads/>) with default parameters. Annotation was carried out using the NCBI Prokaryotic Genome Annotation Pipeline.

The complete genome of *S. hominis* KHUD_010 consists of a single circular chromosome of 1,883,665 bp with a G+C content of 39.04% (Data file 2 and 3). It encodes 1,793 protein-coding genes, four rRNA operons (5 S, 16 S, 23 S) and 120 tRNAs. Genome completeness was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.8.2 [12] with the "bacteria_odb12" dataset (3,130 genomes, 116 BUSCOs), yielding a completeness score of 99.1% (Data file 4). Based on the NCBI taxonomy check, strain KHUD_010 showed 99.94% ANI and 99.86% coverage with *S. hominis* (GCA_014287335.1), supporting its species-level identification. Comparative genomic analysis was conducted using two publicly available *S. hominis* genomes: NSJ-17 (feces, China) and UMB6992B (urine, USA).

Orthologous clustering identified 1,416 core gene clusters shared among the three strains, 223 dispensable clusters, and 398 strain-specific clusters (Data file 5). Strain KHUD_010 harbored 18 unique gene clusters containing 20 genes (Data file 6). Functional annotation of these genes revealed enrichment in COG category L (replication, recombination, and repair), with additional genes mapped to categories S (function unknown), V (defense mechanisms), and D (cell cycle control). These features suggest the presence of genomic adaptations relevant to survival in the subgingival environment.

The genome of KHUD_010 represents the first complete genome of an oral *S. hominis* isolate. It significantly expands the limited genomic repertoire of this species and provides a valuable resource for investigating strain-level diversity, host-specific adaptation, and the ecological roles of streptococci in the oral microbiome. As the third complete genome available for *S. hominis*, it enables broader comparative genomic analyses and functional studies of this understudied species. Table 1 summarizes Data files 1–6 and Datasets 1–2.

Limitation

None.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	HiFi read length distribution of <i>Streptococcus hominis</i> KHUD_010	Portable Data Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29527361.v1) [13]
Data file 2	Genome features of <i>Streptococcus hominis</i> KHUD_010	Office Open XML Spreadsheet file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.29527409.v2) [14]
Data file 3	Circular genome map of <i>Streptococcus hominis</i> KHUD_010 with annotated features	Portable Data Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29527520.v1) [15]
Data file 4	Short BUSCO summary	Portable Data Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29527457.v3) [16]
Data file 5	Upset plot of orthologous gene clusters among three <i>Streptococcus hominis</i> strains (KHUD_010, NSJ_17, and UMB6992B)	Portable Data Format file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.29527553.v1) [17]
Data file 6	KHUD_010 strain: 18 unique gene clusters containing 20 genes	Office Open XML Spreadsheet file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.29527433.v5) [18]
Data set 1	Sequencing read dataset of <i>Streptococcus hominis</i> KHUD_010	Fastq file (.fastq.gz)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRX28051166)
Data set 2	Genome assembly of <i>Streptococcus hominis</i> KHUD_010	FASTA/GenBank/ASN.1	NCBI Genome assembly (https://identifiers.org/ncbi/insdc.gca:GCA_040207525.1) [20]

Abbreviations

NGS	Next Generation Sequencing
BLAST	Basic Local Alignment Search Tool
SMRT	Single-Molecule Real-Time
NCBI	National Center for Biotechnology Information
BUSCO	Benchmarking Universal Single-Copy Orthologs
COG	Clusters of Orthologous Groups of proteins

Authors' contributions

JHM, KHK and HKK conceived and designed the experiments. SBY and SWK performed strain isolation, cultivation and DNA extraction. DK and JHL performed the genome analysis. The manuscript was written by SBY and DK and revised by KHK and HKK. The author(s) read and approved the final manuscript.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT, RS-2023-00280791, RS-2024-00358942), Korea Basic Science Institute (National research Facilities and Equipment Center) grant funded by the Korea government (MSIT, RS-2024-00404660).

Data availability

The raw sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRX28051166 (Data Set 1) [19]. The genome assembly of the *S. hominis* KHUD_010 was submitted to NCBI GenBank and are available under the accession number GCF_040207525.1 (Data set 2) [20].

Declarations**Ethics approval and consent to participate**

This study was approved by the Ethical Committee of the Kyung Hee University Dental Hospital (KHD IRB 1606-5) and was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013. All subjects provided informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 July 2025 / Accepted: 28 August 2025

Published online: 29 September 2025

References

- Lamont RJ, Koo H, Hajishengallis G. The oral microbiota: dynamic communities and host interactions. *Nat Rev Microbiol*. 2018;16(12):745–59. <https://doi.org/10.1038/s41579-018-0089-x>.
- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*. 2005;43(11):5721–32. <https://doi.org/10.1128/JCM.43.11.5721-5732.2005>.
- Abranches J, Zeng L, Kajfasz JK, Palmer SR, Chakraborty B, Wen ZT, et al. Biology of oral streptococci. *Microbiol Spectr*. 2018. <https://doi.org/10.1128/microbiolspec.gpp3-0042-2018>.
- Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simón-Soro A, Pignatelli M, et al. The oral metagenome in health and disease. *ISME J*. 2012;6(1):46–56. <https://doi.org/10.1038/ismej.2011.85>.
- Yamagishi J, Sato Y, Shinozaki N, Ye B, Tsuboi A, Nagasaki M, et al. Comparison of boiling and robotics automation method in DNA extraction for metagenomic sequencing of human oral microbes. *PLoS ONE*. 2016;11(4):e0154389. <https://doi.org/10.1371/journal.pone.0154389>.
- Solbiati J, Frias-Lopez J. Metatranscriptome of the oral microbiome in health and disease. *J Dent Res*. 2018;97(5):492–500. <https://doi.org/10.1177/0022034518761644>.
- Pang L, Wang Y, Ye Y, Zhou Y, Zhi Q, Lin H. Metagenomic analysis of dental plaque on pit and fissure sites with and without caries among adolescents. *Front Cell Infect Microbiol*. 2021;11:740981. <https://doi.org/10.3389/fcimb.2021.740981>.
- Okahashi N, Nakata M, Kuwata H, Kawabata S. Oral mitis group streptococci: a silent majority in our oral cavity. *Microbiol Immunol*. 2022;66(12):539–51. <https://doi.org/10.1111/1348-0421.13028>.
- National Center for Biotechnology Information. Genome assembly https://identifiers.org/ncbi/insdc.gca:GCA_014287335.1.
- National Center for Biotechnology Information. Genome assembly https://identifiers.org/ncbi/insdc.gca:GCA_040717855.1.

11. Srinivasan R, Karaoz U, Volegova M, et al. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS ONE*. 2015;10(2):e0117617. <https://doi.org/10.1371/journal.pone.0117617>.
12. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–54. <https://doi.org/10.1093/molbev/msab199>.
13. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. HiFi read length distribution of *Streptococcus hominis* KHUD_010. figshare <https://doi.org/10.6084/m9.figshare.29527361.v1>.
14. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. Genome features of *Streptococcus hominis* KHUD_010. figshare <https://doi.org/10.6084/m9.figshare.29527409.v2>.
15. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. Circular genome map of *Streptococcus hominis* KHUD_010 with annotated features. figshare <https://doi.org/10.6084/m9.figshare.29527520.v1>.
16. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. Short. BUSCO summary. figshare <https://doi.org/10.6084/m9.figshare.29527457.v3>.
17. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. Upset plot of orthologous gene clusters among three *Streptococcus hominis* strains (KHUD_010, NSJ_17, and UMB6992B). figshare <https://doi.org/10.6084/m9.figshare.29527553.v1>.
18. Yang S, Ku D, Moon J-H, Lee J-H, Kang S-W, Kim H, Kwack K. KHUD_010 strain: 18 unique gene clusters containing 20 genes. figshare <https://doi.org/10.6084/m9.figshare.29527433.v5>.
19. NCBI Sequence Read Archive. 2025. <https://identifiers.org/ncbi/insdc.sra:SRX28051166>. Accessed 19 Mar 2025.
20. NCBI. Genome. assembly. 2024. https://identifiers.org/ncbi/insdc.gca:GCA_040207525.1. Accessed 20 Jun 2024.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.