

DATA NOTE

Open Access



# Complete genome sequence of *Schaalia odontolytica* isolated from subgingival biofilm

Eun-Young Jang<sup>1,2</sup>, Jeewan Chun<sup>1,2</sup>, Kyu Hwan Kwack<sup>1</sup>, Ji-Hoi Moon<sup>1\*</sup> and Jae-Hyung Lee<sup>1\*</sup>

## Abstract

**Objective** Recent advancements in genome-based taxonomic classification propose the reclassification of certain *Actinomyces* species into new genera, including *Schaalia*. *Schaalia odontolytica*, the type species within this genus, is frequently found in the human oral cavity and has been associated with actinomycotic lesions. Currently, only two complete genomes of *S. odontolytica* strains have been reported. Recognizing the limited research on subspecies-level variation of *S. odontolytica*, we conducted genome sequencing of strain KHUD\_008, isolated from a Korean periodontitis patient's subgingival biofilm. Additionally, we performed a comparative genome analysis using previously sequenced genomes of strain XH001 and strain FDAARGOS\_732, both derived from the human oral cavity.

**Data description** Pacific Biosciences Sequel II sequencing generated 15,904 and 76,557 raw sequencing sub-reads, which were integrated to assemble the *de novo* genome using the Microbial Genome Analysis pipeline in the Single-Molecule Real-Time Analysis. The genome assembly completeness, assessed by Benchmarking Universal Single-Copy Orthologs, reached 99.2%. The genome is 2,389,595 bp with a GC content of 66.37%, and contains 2,002 protein-coding genes, 9 rRNAs, and 48 tRNA. Comparative analysis with two previously sequenced strains revealed many strain-specific genes in KHUD\_008, primarily related to envelope biogenesis and replication/recombination/repair processes.

**Keywords** *Schaalia odontolytica*, Genome, Subgingival biofilm

## Objective

In the human microbiome, the phylum *Actinobacteria* encompasses more than 30 genera, some of which have the potential to cause human actinomycosis, a chronic granulomatous infectious disease [1–3]. *Actinomyces israelii* (originally named *Streptothrix israeli*) was

identified as the etiological agent responsible for actinomycosis by Kruse in 1896 [2, 4]. Other *Actinomyces* species, including *A. naeslundii*, *A. odontolyticus*, and *A. viscosus*, have also been implicated in actinomycotic lesions in humans [2]. However, recent advancements in genome-based taxonomic classification of the phylum *Actinobacteria* have indicated the need for revising several orders, families, and genera, as well as numerous species and a few subspecies [5]. As part of this reclassification, certain *Actinomyces* species are proposed to be moved into new genera, such as *Schaalia*. For example, two *Actinomyces odontolyticus* species strains, XH001 [6] and F0309 [7], most studied as references, were moved into *Schaalia odontolytica*.

\*Correspondence:

Ji-Hoi Moon  
prudence75@khu.ac.kr  
Jae-Hyung Lee  
jaehlee@khu.ac.kr

<sup>1</sup>Department of Oral Microbiology, College of Dentistry, Kyung Hee University, Seoul 02447, Republic of Korea

<sup>2</sup>Department of Dentistry, Graduate School, Kyung Hee University, Seoul 02447, Republic of Korea



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The genus *Schaalia* comprises aerobic Gram-positive organisms that typically exhibit straight to slightly curved rod-shaped cells, although some species may produce coccoid or coccobacillary cells [5]. Their peptidoglycan composition includes L-lysine and N-acetylated muramic acid [5]. The genomic G+C content of *Schaalia* species ranges from 56 to 70%. The type species within this genus is *Schaalia odontolytica*, which commonly inhabits the oral cavity [3, 5]. Currently, only two complete genomes of *S. odontolytica* strains have been reported. Recognizing the limited research on subspecies-level variation of *S. odontolytica*, we conducted genome sequencing of strain KHUD\_008, which was isolated from the subgingival biofilm of a Korean patient with periodontitis.

### Data description

The bacterial strain KHUD\_008 was obtained by culturing subgingival bacterial biofilms of a Korean patient in Trypticase soy agar with defibrinated sheep blood under anaerobic conditions at 37 °C. Genomic DNA was extracted using Wizard HMW DNA Extraction Kit (promega). The complete 16 S rRNA gene sequence (~1.5 kb) was amplified using universal eubacteria primers (forward primer: 5'-AGAGTTTGATCCTGGCTCAG-3'; reverse primer: 5'-GGTTACCTTGTACGACTT-3') [8].

**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	Subread length distribution of sequencing reads	Portable Data Format file (.pdf)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.23591493">https://doi.org/10.6084/m9.figshare.23591493</a> ) [10]
Data file 2	Genome features of <i>Schaalia odontolytica</i> KHUD_008	Portable Data Format file (.pdf)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.23591523">https://doi.org/10.6084/m9.figshare.23591523</a> ) [11]
Data file 3	Short BUSCO summary	Portable Data Format file (.pdf)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.23591538">https://doi.org/10.6084/m9.figshare.23591538</a> ) [12]
Data file 4	Strain KHUD_008 specific genes belonging to COG L and COG M	Portable Data Format file (.pdf)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.23591544">https://doi.org/10.6084/m9.figshare.23591544</a> ) [13]
Data set 1	Sequencing read dataset 1 of <i>Schaalia odontolytica</i> KHUD_008	Fastq file (.fastq.gz)	NCBI Sequence Read Archive ( <a href="https://identifiers.org/ncbi/insdc.sra:SRX19887473">https://identifiers.org/ncbi/insdc.sra:SRX19887473</a> ) [14]
Data set 2	Sequencing read dataset 2 of <i>Schaalia odontolytica</i> KHUD_008	Fastq file (.fastq.gz)	NCBI Sequence Read Archive ( <a href="https://identifiers.org/ncbi/insdc.sra:SRX19887474">https://identifiers.org/ncbi/insdc.sra:SRX19887474</a> ) [15]
Data set 3	Genome assembly of <i>Schaalia odontolytica</i> KHUD_008	FASTA / GenBank / ASN.1	NCBI Genome assembly ( <a href="https://identifiers.org/ncbi/insdc.gca:GCA_024584435.1">https://identifiers.org/ncbi/insdc.gca:GCA_024584435.1</a> ) [16]

After purifying and sequencing the PCR product, species identification was performed *via* a BLAST search against the GenBank database.

For whole genome sequencing of *S. odontolytica* KHUD\_008, a Single-Molecule Real-Time (SMRT) bell library was prepared following the manufacturer's instructions (Pacific Biosciences). Two sequencing runs were performed, yielding 15,904 and 76,557 raw sequencing subreads with average read lengths of 8,559 and 8,610 bp using Pacific Biosciences Sequel II with 2.0 sequencing chemistry (Data file 1). The *de novo* genome of strain KHUD\_008 was assembled using the Microbial Genome Analysis pipeline in the SMRT Analysis version 11.0 (<https://www.pacb.com/support/software-downloads/>) with default parameters, incorporating two sequencing run datasets. The NCBI Prokaryotic Genome Annotation Pipeline was employed for gene annotation.

The genome of *S. odontolytica* KHUD\_008 is 2,389,595 bp long with a G+C content of 66.37% (Data file 2). It contains 2,002 protein-coding genes, 9 rRNAs (5 S, 16 S, 23 S), and 48 tRNAs. To assess genome completeness, Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.7 [9] was used with “bacteria\_odb10 dataset” (number of genomes: 4085, number of BUSCOs: 124), resulting in a genome assembly completeness of 99.2% (Data file 3). Additionally, we conducted a comparative genome analysis with previously sequenced genomes of strain XH001 and strain FDAARGOS\_732, both isolated from the human oral cavity. This analysis revealed 103 strain-specific genes in KHUD\_008, excluding unclassified and “unknown” genes. Strain-specific genes were found to be enriched in COG L (replication, recombination and repair) and COG M (cell wall, membrane, envelope biogenesis) categories (Data file 4).

We believe that the dataset, presented alongside the complete genome of *S. odontolytica* KHUD\_008, forms a solid foundation for future studies on the genomic, transcriptomic and phenotypic characterization of this species, as well as other closely related species. As the third publicly available complete genome sequence of *S. odontolytica*, this data will be a valuable resource for comparative genomics, enabling more comprehensive analyses and discoveries regarding strain-level variation. Table 1 provides the links to Data files 1–4 and Data sets 1–3.

### Limitation

Initially, we obtained the sequencing subread data using Pacific Biosciences Sequel II, and the number of the subread was less than 20,000. In order to obtain a more comprehensive dataset, we performed an additional sequencing run using the same sequencing library, resulting in over 70,000 subreads. By combining the datasets from both sequencing runs, we successfully generated the complete and circular genome for *S. odontolytica*

**KHUD\_008.** Therefore, we believe that there are no limitations in terms of data acquisition and the complete genome assembly process in this study.

#### Abbreviations

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

#### Author contributions

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

#### Funding

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT (NRF-2021R1A2C2008180, NRF-2022R1F1A1071248).

#### Data availability

Data files 1–4 described in this Data note can be freely and openly accessible on Figshare (<https://figshare.com/>) [10–13]. Data sets 1–3 are available on the NCBI database. The raw reads have been submitted to the NCBI Sequence Read Archive under the accession number SRX19887473 [14] and SRX19887474 [15] (Data set 1 and 2). The genome assembly of the *S. odontolytica* KHUD\_008 was submitted to NCBI GenBank and are available under the accession number GCA\_024584435.1 (Data set 3) [16].

#### 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: 29 June 2023 / Accepted: 8 December 2023

Published online: 09 February 2024

#### References

1. Hall V. *Actinomyces*-gathering evidence of human colonization and Infection. *Anaerobe*. 2008;14:1–7.
2. Könönen E, Wade WG. *Actinomyces* and related organisms in human Infections. *Clin Microbiol Rev*. 2015;28:419–42.
3. Podar NA, Klingeman D, Miranda-Sanchez F, Dewhirst FE, Podar M. Draft genome sequence of *Schaalia odontolytica* strain ORNL0103, a basibiont of *Candidatus Saccharibacteria* HMT352. *Microbiol Resour Announc*. 2021;10:e0079321.
4. Kruse W. Systematik der Streptothricheen und Bakterien, p 48–96. In Flugge C, Die Mikroorganismen, 3rdvol 2, Vogel. Leipzig, Germany. 1896.
5. Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk HP, Goodfellow M, Göker M. Genome-based taxonomic classification of the phylum *Actinobacteria*. *Front Microbiol*. 2018;9:2007.
6. National Center for Biotechnology Information. Genome assembly [https://identifiers.org/ncbi/insdc.gca:GCA\\_001462375.1](https://identifiers.org/ncbi/insdc.gca:GCA_001462375.1).
7. National Center for Biotechnology Information. Genome assembly [https://identifiers.org/ncbi/insdc.gca:GCA\\_000163415.1](https://identifiers.org/ncbi/insdc.gca:GCA_000163415.1).
8. Srinivasan R, Karaoz U, Volegova M, MacKichan J, Kato-Maeda M, Miller S, Nadarajan R, Brodie EL, Lynch SV. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS ONE*. 2015;10:e0117617.
9. Manni M, Berkeley MR, Seppely 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;38:4647–54.
10. Jang E-Y, Chun J, Kwack KK, Moon J-H, Lee J-H. Subread length distribution of sequencing reads. [figshare https://doi.org/10.6084/m9.figshare.23591493](https://doi.org/10.6084/m9.figshare.23591493) (2023).
11. Jang E-Y, Chun J, Kwack KK, Moon J-H, Lee J-H. Genome features of *Schaalia odontolytica* KHUD\_008. [figshare https://doi.org/10.6084/m9.figshare.23591523](https://doi.org/10.6084/m9.figshare.23591523) (2023).
12. Jang E-Y, Chun J, Kwack KK, Moon J-H, Lee J-H, Short. BUSCO summary. [figshare https://doi.org/10.6084/m9.figshare.23591538](https://doi.org/10.6084/m9.figshare.23591538) (2023).
13. Jang E-Y, Chun J, Kwack KK, Moon J-H, Lee J-H. Strain KHUD\_008 specific genes belonging to COG L and COG M. [figshare https://doi.org/10.6084/m9.figshare.23591544](https://doi.org/10.6084/m9.figshare.23591544) (2023).
14. NCBI Sequence Read Archive. <https://identifiers.org/ncbi/insdc.sra:SRX19887473> (2023).
15. NCBI Sequence Read Archive. <https://identifiers.org/ncbi/insdc.sra:SRX19887474> (2023).
16. NCBI. Genome. assembly. [https://identifiers.org/ncbi/insdc.gca:GCA\\_024584435.1](https://identifiers.org/ncbi/insdc.gca:GCA_024584435.1) (2022).

#### Publisher's Note

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