

# Draft genome sequence of endometrial *Schaalia turicensis* strain R31 from a patient with adenomyosis

Nicole Jimenez,<sup>1</sup> Matthew Hawkins,<sup>2</sup> Bonnie Hurwitz,<sup>3</sup> Melissa M. Herbst-Kralovetz<sup>1,2,4</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 4.

**ABSTRACT** The endometrial *Schaalia turicensis* strain R31 genome was isolated from a post-hysterectomy endometrial glycerol swab obtained from a woman diagnosed with adenomyosis. Here, we report a 1.97 Mb complete draft genome sequence of this strain.

**KEYWORDS** Actinomyces, *Schaalia turicensis*, female reproductive tract, uterus, endometrium, adenomyosis

*Actinomyces turicensis* was reclassified (1) as *Schaalia turicensis*. These Gram-positive facultative anaerobes exist in the human oral (2, 3), gut (4), and urogenital tracts (5–7). *S. turicensis* is an opportunistic pathogen with cases involving urethritis (5), bacteremia (8), meningitis (9), soft tissue infections (10), and wound infections (11).

In this announcement, we describe the draft genome of *S. turicensis* R31, which was isolated from the endometrium, thereby expanding our understanding of *S. turicensis* in gynecologic health.

This isolate was obtained from an endometrial swab of a patient diagnosed with adenomyosis, in which endometrial-like tissue grows into the myometrium. This was conducted as part of a multi-omic study investigating patients undergoing hysterectomy (12–14). The Institutional Review Board of the University of Arizona approved this study (reference no. 1708726047), and written informed consent was obtained from the study participant. Swabs were frozen in Amies transport media with 10% glycerol. Serial dilutions were performed to isolate the bacteria under anaerobic conditions at 37°C on Tryptic Soy Agar supplemented with 5% sheep's blood for 48 h. Bacterial DNA was extracted using the Qiagen DNeasy PowerSoil Pro Kit (MO BIO Laboratories, Carlsbad) and sequenced at the University of Arizona PANDA Core for Genomics and Microbiome Research. Paired-end sequencing was performed using Illumina's PCR-Free Library Prep Kit and the NextSeq 1000 Platform (300-cycle) with read length of 35–151 bp. Trimmomatic (v0.39, ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, SLIDINGWINDOW:4:20, MINLEN:100, HEADCROP:15) (15) improved read quality and was assessed using FastQC (v0.11.9) (16). Kraken2 (v2.1.3) (17) and Bracken (v2.8) (17) were used for species-level read classification based on the k2\_plus database (downloaded on 2023-06-05) (18). KrakenTools (v1.2) (17) (extract\_kraken\_reads.py, --taxid 9606, --include-children) was used to separate human from microbial reads. Assembly was performed using Unicycler (v16.0) (19), followed by quality checks with Checkm2 (v1.0.1, -m 500) (20) and Quast (v5.2.0) (21). Annotation was performed using PGAP (v 6.1) (22). Default parameters were used for all tools unless otherwise specified. All code is available on GitHub ([https://github.com/hurwitzlab/vaginal\\_genome\\_assembly](https://github.com/hurwitzlab/vaginal_genome_assembly)). Genomic analyses were performed on the Bacterial and Viral Bioinformatics Resources Center website (23).

The draft genome of *S. turicensis* R31 comprises 38 contigs, totaling 1,973,689 base pairs, with a GC content of 56.83%. The assembly exhibits moderate contiguity with an N50 of 102,621 bp. Annotation revealed 1,750 coding sequences, 47 tRNA genes, and

**Editor** Catherine Putonti, Loyola University Chicago, Chicago, Illinois, USA

Address correspondence to Melissa M. Herbst-Kralovetz, mherbst1@email.arizona.edu.

M.M.H.-K is a paid consultant for Freya Biosciences; however, none of this work was shared with or licensed to this company or any other commercial entity. All other authors declare no competing interests.

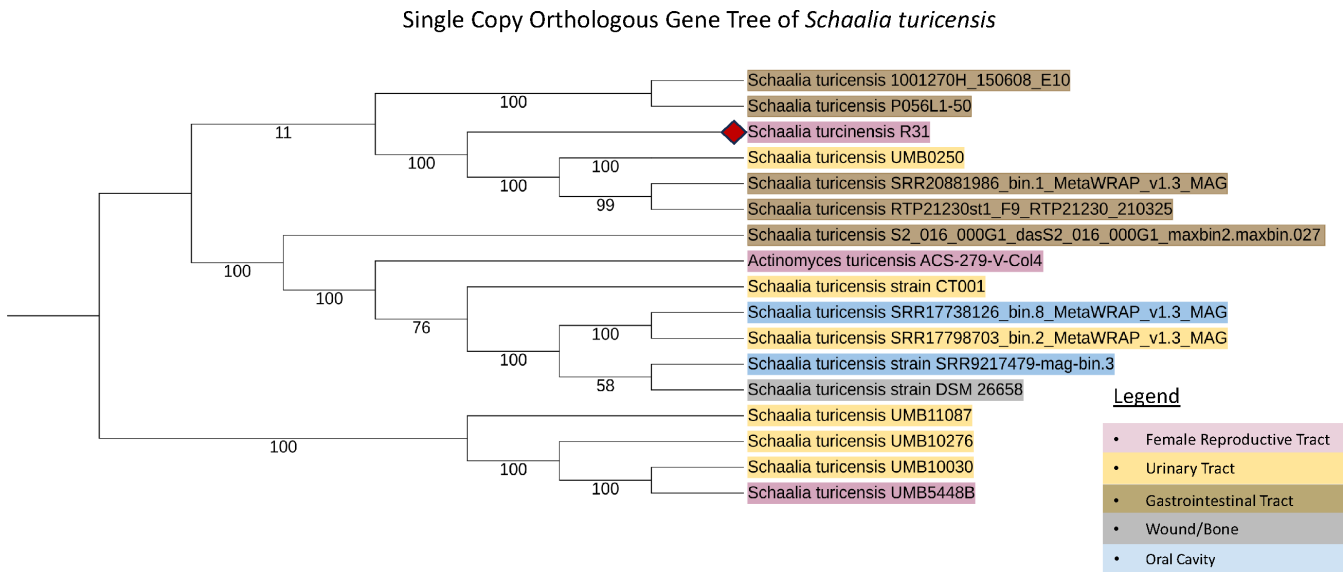
See the funding table on p. 4.

**Received** 25 August 2025

**Accepted** 26 October 2025

**Published** 24 November 2025

Copyright © 2025 Jimenez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



**FIG 1** Single-copy orthologous phylogenetic tree among 16 publicly available *S. turicensis* strains and R31. Single-copy orthologous genes were used to create the bacterial genome tree ( $n = 278$ ) and were aligned by codon using RaXML as part of the BV-BRC bacterial genome tree pipeline using default parameters. The tree comprises 11 whole-genome assemblies and five metagenome-assembled genomes of *S. turicensis*. *S. turicensis* R31 is the strain discussed in this microbial resource announcement and indicated by a red diamond. The colors of *S. turicensis* strain names are based on the environmental source from which the genomes originated.

three rRNA genes. Of these, 229 were annotated as hypothetical proteins, while 1,521 had functional assignments.

Subsystem classification identified metabolism (196 genes) as the largest functional group (Table 1). Amino acid metabolism made up 34.7% of the genome's capability. Protein comparisons with public *S. turicensis* strains indicated 438 shared protein families and 24 protein families unique to R31 (Table 1). Two of these were related to cell wall/invasion-associated proteins and histone acetyltransferase, while 22 families were of hypothetical origin (Table 1). Antibiotic resistance genes (24 genes) were identified, which may be crucial for treating actinomycosis (Table 1).

Kraken2 classification confirms the R31 identity within the *Schaalia* genus. Additional strain-specific phylogenetic analysis revealed a non-specific placement based on isolation source (Fig. 1), with an average nucleotide identity of 96.8% to other *S. turicensis* in Clade 1 and 68% in Clade 2 (Table 1). Together, these findings highlight the fundamental genomic capabilities of the endometrial *S. turicensis* R31 genome, laying a foundation for investigating its role in gynecologic health.

**TABLE 1** Genome characteristics table for *S. turicensis* R31<sup>a</sup>

| Genome name                                              | <i>S. turicensis</i> R31                                                                                                                                                                                                                                                                             |
|----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Isolate information                                      |                                                                                                                                                                                                                                                                                                      |
| Isolation source                                         | Endometrium                                                                                                                                                                                                                                                                                          |
| Health status                                            | Adenomyosis                                                                                                                                                                                                                                                                                          |
| Strain identity                                          |                                                                                                                                                                                                                                                                                                      |
| KRAKEN2 taxonomy                                         | <i>Schaalia turicensis</i>                                                                                                                                                                                                                                                                           |
| Average nucleotide identity to representative strain (%) | 95.95                                                                                                                                                                                                                                                                                                |
| Average nucleotide identity to Clade 1 (mean) (%)        | 96.76                                                                                                                                                                                                                                                                                                |
| Average nucleotide identity to Clade 2 (mean) (%)        | 67.97                                                                                                                                                                                                                                                                                                |
| Genome characteristics                                   |                                                                                                                                                                                                                                                                                                      |
| Raw reads                                                | 4058498                                                                                                                                                                                                                                                                                              |
| Genome size (bp)                                         | 1,973,689                                                                                                                                                                                                                                                                                            |
| Number of contigs                                        | 38                                                                                                                                                                                                                                                                                                   |
| Contig N50 (bp)                                          | 102,621                                                                                                                                                                                                                                                                                              |
| Contig L50                                               | 6                                                                                                                                                                                                                                                                                                    |
| GC (%)                                                   | 56.83%                                                                                                                                                                                                                                                                                               |
| Genome coverage                                          | 200×                                                                                                                                                                                                                                                                                                 |
| Number of 5S rRNA                                        | 1                                                                                                                                                                                                                                                                                                    |
| Number of 16S rRNA                                       | 1                                                                                                                                                                                                                                                                                                    |
| Number of 23S rRNA                                       | 1                                                                                                                                                                                                                                                                                                    |
| Number of tRNAs                                          | 47                                                                                                                                                                                                                                                                                                   |
| Number of CDS                                            | 1,750                                                                                                                                                                                                                                                                                                |
| Number of CDS with functional assignments                | 1,521                                                                                                                                                                                                                                                                                                |
| Number of unique protein families                        | 24                                                                                                                                                                                                                                                                                                   |
| Unique protein family IDs (PGFAMs) – hypothetical        | PGF_12915410, PGF_06376995, PGF_12422020, PGF_11667696, PGF_00274664, PGF_01749377, PGF_00426553, PGF_00273925, PGF_12868907, PGF_08271889, PGF_02376408, PGF_00978629, PGF_06147024, PGF_02629833, PGF_04698765, PGF_00116012, PGF_05143424, PGF_00208348, PGF_05574635, PGF_00263229, PGF_02376410 |
| Unique protein family IDs (PGFAMs) – non-hypothetical    | PGF_00041234, PGF_10350959, PGF_10402854                                                                                                                                                                                                                                                             |
| Number of antibiotic resistance genes (PATRIC)           | 24                                                                                                                                                                                                                                                                                                   |
| Antibiotic resistance gene IDs                           | Alr, Ddl, Dfr, dxr, EF-G, EF-Tu, folA, folP, GdpD, gidB, gyrA, gyrB, iso-tRNA, LpqB, MtrA, MtrB, MurA, PgsA, rpoB, rpoC, rho, S10p, S12p                                                                                                                                                             |

<sup>a</sup>The table includes isolate information, including the source and health status of the patient from which the isolate was obtained. Further strain identity information, including taxonomy classification based on two databases (Kraken2) and average nucleotide identity compared to all *S. turicensis* genomes, genomes within Clade 1, and genomes within Clade 2, is highlighted in the phylogenetic tree. In addition, genomic characteristics were obtained from annotated and evaluated with BV-BRC, including genome size, number of contigs, contig N50, contig L50, GC content, genome coverage, number of 5S rRNA, 16S rRNA, 23S rRNA, tRNAs, CDS, and CDS with functional assignments. The data also include the number of unique protein families specific to the R31 genome and their corresponding PATRIC cross-genus family IDs, as well as the number of antibiotic resistance genes identified by PATRIC, along with their gene IDs. Methods and version numbers for genome assembly can be found at [https://github.com/hurwitzlab/vaginal\\_genome\\_assembly](https://github.com/hurwitzlab/vaginal_genome_assembly).

## ACKNOWLEDGMENTS

This study was primarily supported by the Mary Kay Foundation Translational Research Grant to M.M.H-K. The authors would like to thank the patient who participated in the parent study, as well as the clinical team and clinical research coordinators for their assistance with patient recruitment, sample collection, and clinical data collection. The authors would also like to acknowledge Dr. Pawel Laniewski for additional bacterial cultivation support, including funding from Arizona Biomedical Research Center New Investigator Award to P.L.(RFGA2022-010-05). N.R.J. was supported by the American Cancer Society Postdoctoral Fellowship, PF-24-1316926-01-MM, for work on the endometrial microbiome. N.R.J. is part of the Guiding U54 Investigator Development to Sustainability (GUIDeS) shared resource, a part of the Partnership for Native American Cancer Prevention, which is funded under parallel grants, U54CA143924 and U54CA143925, and supports early-stage investigators. The authors would also like to

acknowledge the University of Arizona BIO5 Institute and Keep Engaging Youth in Science (KEYS) Research internship in which M.H. participated and conducted this work.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Obstetrics and Gynecology, College of Medicine-Phoenix, University of Arizona, Phoenix, Arizona, USA

<sup>2</sup>Department of Basic Medical Sciences, College of Medicine-Phoenix, University of Arizona, Phoenix, Arizona, USA

<sup>3</sup>Department of Biosystems Engineering, University of Arizona, Tucson, Arizona, USA

<sup>4</sup>University of Arizona Cancer Center, Tucson, Arizona, USA

## AUTHOR ORCIDS

Nicole Jimenez  <http://orcid.org/0000-0001-9755-0022>

Bonnie Hurwitz  <http://orcid.org/0000-0001-8699-957X>

Melissa M. Herbst-Kralovetz  <http://orcid.org/0000-0002-8540-5917>

## FUNDING

| Funder                                                 | Grant(s)            | Author(s)                   |
|--------------------------------------------------------|---------------------|-----------------------------|
| <a href="#">American Cancer Society</a>                | PF-24-1316926-01-MM | Nicole Jimenez              |
| <a href="#">Mary Kay Ash Foundation</a>                |                     | Melissa M. Herbst-Kralovetz |
| <a href="#">Arizona Biomedical Research Commission</a> | RFGA2022-010-05     | Melissa M. Herbst-Kralovetz |
| <a href="#">National Cancer Institute</a>              | U54CA143924         | Melissa M. Herbst-Kralovetz |

## AUTHOR CONTRIBUTIONS

Nicole Jimenez, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing | Matthew Hawkins, Investigation, Writing – original draft, Writing – review and editing | Bonnie Hurwitz, Formal analysis, Writing – review and editing | Melissa M. Herbst-Kralovetz, Conceptualization, Supervision, Validation, Writing – review and editing

## DATA AVAILABILITY

The draft genome sequence of *Schaalia turicensis* strain R31 has been deposited in the Sequence Read Archive (SRA) under accession number [SRR35785846](#), after filtering out sequences below 200 bp. This genome is part of a larger BioProject ([PRJNA1036657](#)) associated with 16S rRNA level data of vaginal ([SAMN38145775](#)) and rectal ([SAMN38145877](#)) microbiomes. The genome BioSample identification is [SAMN50643546](#). The genome annotation can also be found at GenBank [GCA\\_052216585.1](#) and additional genome assembly, annotation, protein family, and phylogenetic analysis information at BV-BRC (<https://www.bv-brc.org/workspace/jimeneznr@patricbrc.org/Schaalia>). Additional assembly information and parameters utilized can be found at [https://github.com/hurwitzlab/vaginal\\_genome\\_assembly](https://github.com/hurwitzlab/vaginal_genome_assembly).

## REFERENCES

- Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk H-P, Goodfellow M, Göker M. 2007. Genome-based taxonomic classification of the phylum Actinobacteria. *Front Microbiol* 9. <https://doi.org/10.3389/fmicb.2018.02007>
- Vassa N, Mubarik A, Patel D, Muddassir S. 2019. *Actinomyces turicensis*: an unusual cause of cervicofacial actinomycosis presenting as ludwig angina in an immunocompromised host - case report and literature review. *IDCases* 18:e00636. <https://doi.org/10.1016/j.idcr.2019.e00636>
- Tabaksert A, Kumar R, Raviprakash V, Srinivasan R. 2021. *Actinomyces turicensis* parapharyngeal space infection in an immunocompetent host: first case report and review of literature. *Access Microbiol* 3:000241. <https://doi.org/10.1099/acmi.0.000241>
- Patel A, Bhushan S. 2025. Recurrent diverticulitis complicated by *Actinomyces turicensis* abscess. *ACG Case Rep J* 12:e01658. <https://doi.org/10.14309/crj.0000000000001658>
- Nokchan N, Wongsurawat T, Jenjaroenpun P, Nitayanon P, Tribuddharat C. 2021. Complete genome sequence of *Schaalia turicensis* strain CT001,

- isolated from a patient with gonococcal urethritis in Thailand. Microbiol Resour Announc 10:e0083621. <https://doi.org/10.1128/MRA.00836-21>
6. Sarumathi D, Gopichand P, Sejjal K, Priyamvada PS, Mandal J. 2020. *Actinomyces turicensis* causing urinary tract infection in nephrotic syndrome patient- a case report. JCDR. <https://doi.org/10.7860/JCDR/2020/44091.13734>
  7. Liu Y, Li M, Ding Y, Gao Y, Li T, Wang X. 2025. *Schaalia turicensis* endometritis in a diabetic patient: MALDI-TOF MS overcomes diagnostic barriers in a non-IUD-associated case. Front Med 12:1593385. <https://doi.org/10.3389/fmed.2025.1593385>
  8. Kansara T, Majmundar M, Doshi R, Ghosh K, Saeed M. 2020. A case of life-threatening *Actinomyces turicensis* bacteremia. Cureus 12:e6761. <https://doi.org/10.7759/cureus.6761>
  9. Kocsis B, Tiszlavicz Z, Jakab G, Brassay R, Orbán M, Sárkány Á, Szabó D. 2018. Case report of *Actinomyces turicensis* meningitis as a complication of purulent mastoiditis. BMC Infect Dis 18:686. <https://doi.org/10.1186/s12879-018-3610-y>
  10. Liu P, Sun K, Li R, Chen X. 2024. Case report: a rare case of skin abscess caused by coinfection of *Actinobaculum schaalii* and *Actinomyces turicensis* Front Cell Infect Microbiol 14:1378197. <https://doi.org/10.3389/fcimb.2024.1378197>
  11. Cronin JT, Richards BW, Skedros JG. 2023. *Schaalia* (formerly *actinomyces*) *turicensis* infection following open rotator cuff repair. Cureus 15:e34242. <https://doi.org/10.7759/cureus.34242>
  12. Lorentzen GM, Łaniewski P, Cui H, Mahnert ND, Mourad J, Borst MP, Willmott L, Chase DM, Roe DJ, Herbst-Kralovetz MM. 2024. Cervicovaginal metabolome and tumor characteristics for endometrial cancer detection and risk stratification. Clin Cancer Res 30:3073–3087. <https://doi.org/10.1158/1078-0432.CCR-23-2934>
  13. Jimenez NR, Herman CR, Łaniewski P, Cope E, Lee K, Mahnert ND, Chase DM, Caporaso JG, Herbst-Kralovetz MM. 2025. Navigating complexities of polymorphic microbiomes in endometrial cancer. NPJ Biofilms Microbiomes 11:85. <https://doi.org/10.1038/s41522-025-00690-1>
  14. Łaniewski P, Cui H, Mahnert ND, Mourad J, Borst MP, Willmott L, Chase DM, Roe DJ, Herbst-Kralovetz MM. 2022. Protein biomarkers in cervicovaginal lavages for detection of endometrial cancer. Biomark Res 10:88. <https://doi.org/10.1186/s40364-022-00438-5>
  15. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
  16. Tully B. 2016. Quality assessment: FastQC v1. <https://doi.org/10.17504/protocols.io/fa3bign>
  17. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, Salzberg SL, Steinegger M. 2022. Metagenome analysis using the Kraken software suite. Nat Protoc 17:2815–2839. <https://doi.org/10.1038/s41596-022-00738-y>
  18. Langmead B. 2025. Index Zone by BenLangmead. Available from: <https://benlangmead.github.io/aws-indexes/k2>
  19. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
  20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
  21. Mikheenko A, Saveliev V, Hirsch P, Gurevich A. 2023. WebQUAST: online evaluation of genome assemblies. Nucleic Acids Res 51:W601–W606. <https://doi.org/10.1093/nar/gkad406>
  22. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. Nucleic Acids Res 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
  23. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, et al. 2023. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res 51:D678–D689. <https://doi.org/10.1093/nar/gkac1003>