

Draft genome sequence of endometrial *Alloscardovia omnicolens* strain Y28 identified in endometrial cancer

Gwenllian A. Anwyl,^{1,2} Nicole R. Jimenez,³ Bonnie Hurwitz,⁴ Melissa M. Herbst-Kralovetz^{1,3,5}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We report a 1.8 Mb draft genome of *Alloscardovia omnicolens* Y28 isolated from a patient with endometrial cancer.

KEYWORDS *Alloscardovia omnicolens*, endometrial cancer, endometrial microbiota, human microbiome, gynecologic health

Alloscardovia omnicolens was first identified in Europe in 2007, following isolation from diverse human samples, including the urogenital tract (1), bloodstream (2), oral cavity, and abscesses of the lung and aortic valve (2). It is a gram-positive, non-spore-forming, non-motile rod that grows facultatively anaerobically, occurring singly, as pairs, or chains (2). In clinical samples, *A. omnicolens* is frequently co-isolated with *Bifidobacterium* spp. (3). Although usually considered commensal (4), *A. omnicolens* has been associated with urinary tract infections and bacteremia (1), suggesting an opportunistic role.

This announcement documents the draft genome of the endometrial *A. omnicolens* strain Y28, clarifying its genomic characteristics and evolutionary relationship, to advance understanding of endometrial microbiota.

As part of a larger multi-omic study on endometrial cancer and benign condition (5–7), this research project was conducted at the University of Arizona. The Institutional Review Board approved the study (reference no. 1708726047), and the participant diagnosed with endometrial cancer provided written informed consent. Endometrial samples were collected from the bifurcated uterus using sterile swabs. Swabs were stored in Amies transport media with 10% glycerol, after which microbial isolates were obtained and cultured 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 & Microbiome Research. Paired-end sequencing was performed using Illumina's PCR-Free Library Prep and the Next-Seq 1000 Platform (300-cycle) with a 2×150 bp read length. Trimmomatic (v0.39) (8) improved read quality and was assessed with FastQC (v0.11.9) (9). Kraken2 (v2.1.3) (10) and Bracken (v2.8) (10) were used for species-level read classification based on the k2_pluspf database (downloaded on 2023-06-05) (11). Krakentools (v1.2) (10) (extract_kraken_reads.py, --taxid 9606, --include-children) was used to separate human from microbial reads. Assembly utilized Unicycler (v16.0) (12) using SPAdes (13), followed by quality checks with Checkm2 (v1.0.1, -m 500) (14) and Quast (v5.2.0) (15). Annotation was performed using PGAP (v 6.1) (16). Default parameters were used for all tools unless otherwise specified. All code is available on GitHub (https://github.com/hurwitzlab/vaginal_genome_assembly). Genomic analyzes were performed using the Bacterial and Viral Bioinformatics Resources Center (17, 18).

The draft of *A. omnicolens* Y28 comprises 13 contigs, with 1,844,906 bp, an N50 of 1,026,604, and a GC content of 46.84% (Table 1). Genome annotation revealed

Editor Daria Van Tyne, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

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

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 3 December 2025

Accepted 16 January 2026

Published 10 February 2026

Copyright © 2026 Anwyl 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/).

TABLE 1 Genomic assembly and characteristics of *A. omnicolens* strain Y28^a

Genome name	<i>A. omnicolens</i> Y28
Isolate Information	
Isolation Source	Endometrium
Health Status	Endometrium Cancer
Strain Identity	
BV-BRC Genome Similarity	<i>A. omnicolens</i> DSM 21503
Genome Similarity Assembly Accession	GCA_002259705.1
K-mer Count	4/1000
K-mer Distance	0.23011
Average nucleotide identity to the Representative strain (%)	98.80
Genome characteristics	
Genome size (bp)	1,844,906
Number of Contigs	13
Total Raw reads	1,195,459
Contig N50 (bp)	1,026,604
Contig L50	1
GC (%)	46.84
Genome coverage	200x
Number of 5S rRNA	1
Number of 16S rRNA	2
Number of 23S rRNA	2
Number of tRNA	46
Number of CDS	1,568
Number of CDS with functional assignments	1,536
Number of Unique protein families	8
Unique protein Families IDs (PGFAMs-Hypothetical)	PGF_05514654; PGF_04425848; PGF_01749992; PGF_05057205; PGF_05115053; PGF_10607971; PGF_01750103; PGF_05158714
Unique protein Families IDs (PGFAMs- Non-Hypothetical)	0
Number of Antibiotic Resistance Gene IDs (PATRIC)	20
Antibiotic Resistance Gene IDs	MtrA; gyrB; gyrA; S10p; Alr; Ddl; MurA; gidB; rpoB; MtrB; S12p; rho; Ddl; Iso-tRNA; folP; EF-G; EF-Tu; PgsA; folA; Dfr; rpoC

^aThe genome of *A. omnicolens* was isolated from the endometrium of a patient with endometrial cancer. Genus and species identification was performed by using genome similarity and validated by Average Nucleotide Identity. The table summarizes genome assembly statistics, including the number of contigs, total raw reads, N50, genome size, and GC content, as well as genome annotation results. Key genomic features reported include the number of coding sequences (CDS), rRNA and tRNA genes, as well as the proportion of genes with functional assignments. Comparative analysis with reference strain *A. omnicolens* DSM 21503 is provided, along with average nucleotide identity (ANI) and k-mer metrics. Additionally, utilizing BV-BRC we identified the number of protein families and unique protein families identified by PATRIC, and predicted antibiotic resistance genes and their corresponding PATRIC gene identifiers.

1,568 coding sequences (CDS), including 1,536 functionally assigned proteins and 32 hypothetical proteins, alongside 46 tRNA and 5 rRNA genes (Table 1).

Strain Y28 belongs to the *Alloscardovia* genus, sharing 98.8% nucleotide identity with the representative strain (Table 1) also observed in the phylogenetic tree (Fig. 1). RASTtk (19) identified that Y28 shared 1,048 protein families with publicly available *A. omnicolens* whole-genome strains but also possessed eight unique hypothetical proteins (Table 1). Amino acid metabolism represented 35.5% of the metabolism-associated genes, with high levels of lysine, threonine, and cysteine pathways (28 genes). Such metabolic alterations are increasingly recognized in gynecologic health (20, 21), where disruptions in these pathways may contribute to endometrial cancer (22). The presence of

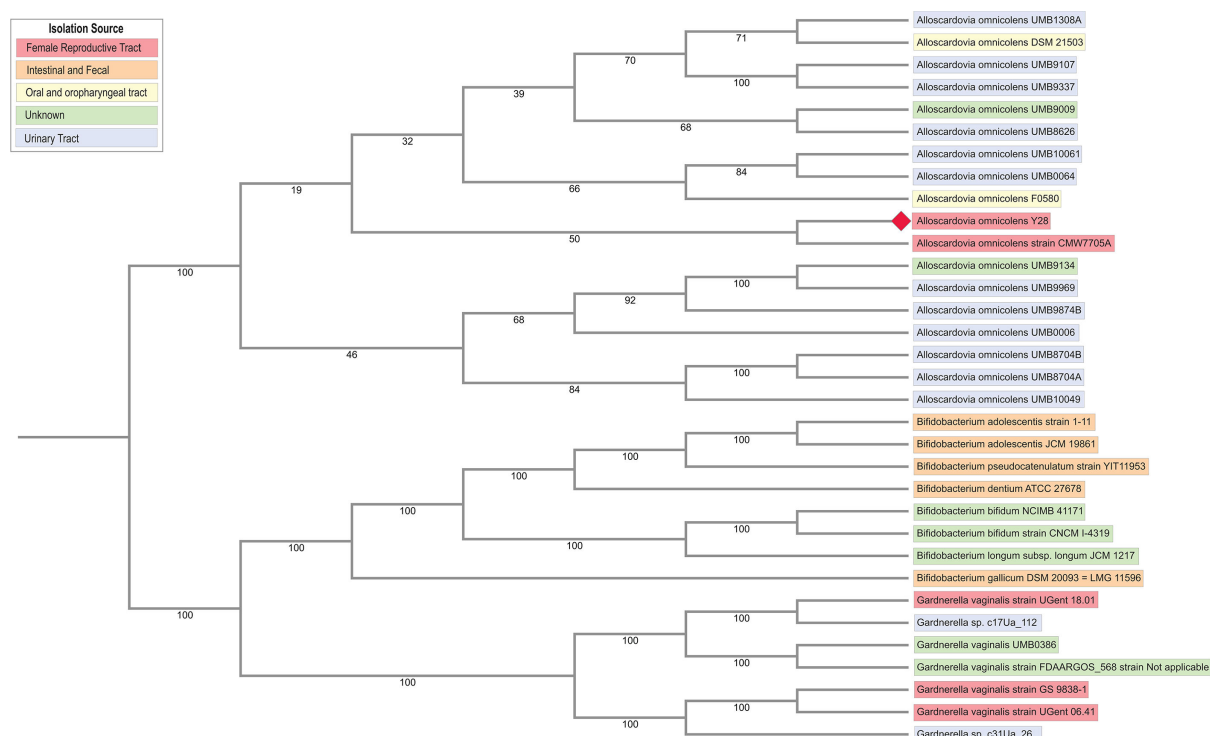


FIG 1 Single Copy Orthologous Phylogenetic Tree amongst publicly available *A. omnicolens* strains, Y28 and related *Bifidobacteriaceae* Species. Single-copy orthologous genes and RaxML were used to create the phylogenetic tree ($n = 415$) and were aligned by codon using MUSCLE. The tree includes 34 whole genomes, all from human hosts. *A. omnicolens* strain Y28 is discussed in this microbial resource announcement and indicated by a red diamond. Colors of strain names are based on their isolation source. Bootstrap values (shown at nodes) reflect the robustness of each branching point.

these pathways may reflect adaptations to the endometrial niche, suggesting potential interactions with the host.

ACKNOWLEDGMENTS

This study received primary support from the Mary Kay Foundation Translational Research Grant awarded to M.M.H-K. We extend our gratitude to the patient who participated in the parent study, as well as the clinical team and research coordinators for their invaluable assistance with patient recruitment, sample collection, and clinical data gathering. We also acknowledge Dr. Pawel Laniewski for providing additional support for bacterial cultivation, including funding from the Arizona Biomedical Research Center New Investigator Award (RFGA2022-010-05).

Additional funding was provided by the University of Arizona Health Sciences Strategic Plan Initiative 4.1. Personalized Defense. The American Cancer Society Postdoctoral Fellowship, PF-24-1316926-01-MM, supported N.R.J.'s 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.

AUTHOR AFFILIATIONS

¹Department of Basic Medical Sciences, College of Medicine-Phoenix, University of Arizona, Phoenix, Arizona, USA

²Department of Life Sciences, University of Bath, Bath, United Kingdom

³Department of Obstetrics and Gynecology, College of Medicine-Phoenix, University of Arizona, Phoenix, Arizona, USA

⁴Department of Biosystems Engineering, University of Arizona, Tucson, Arizona, USA

⁵University of Arizona Cancer Center, Tucson, Arizona, USA

AUTHOR ORCIDs

Gwenllian A. Anwyl  <http://orcid.org/0009-0004-6173-5331>

Nicole R. 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)
Mary Kay Foundation		Melissa M. Herbst-Kralovetz
The American Cancer Society Postdoctoral Fellowship	PF-24-1316926-01-MM	Nicole R. Jimenez
Arizona Biomedical Research Commission	RFGA2022-010-05	Melissa M. Herbst-Kralovetz
National Cancer Institute - Partnership for Native American Cancer Prevention	U54CA143924	Melissa M. Herbst-Kralovetz

AUTHOR CONTRIBUTIONS

Gwenllian A. Anwyl, Formal analysis, Visualization, Writing – original draft, Writing – review and editing | Nicole R. Jimenez, Conceptualization, Investigation, Validation, Writing – review and editing | Bonnie Hurwitz, Formal analysis, Methodology, Validation, Writing – review and editing | Melissa M. Herbst-Kralovetz, Conceptualization, Funding acquisition, Supervision, Validation, Writing – review and editing

DATA AVAILABILITY

The draft genome sequence of *A. omnicolens* Y28 has been deposited at the Sequence Read Archive (SRA) under accession number [NZ_JBRAVA0000000000](https://www.ncbi.nlm.nih.gov/sra/NZ_JBRAVA0000000000). This genome is part of a larger BioProject [PRJNA1354867](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1354867). The genome BioSample identification is [SAMN51299389](https://www.ncbi.nlm.nih.gov/biosample/SAMN51299389). The genome annotation can also be found at GenBank [GCF_052772085.1](https://www.ncbi.nlm.nih.gov/genbank/GCF_052772085.1), and additional genome assembly, annotation, protein family, and phylogenetic analysis information are available at BV-BRC (<https://www.bv-brc.org/workspace/jimeneznr@patricbrc.org/Alloscardovia>) with access via a free account. Additional assembly information and parameters utilized can be found at https://github.com/hurwitzlab/vaginal_genome_assembly.

REFERENCES

- Ogawa Y, Koizumi A, Kasahara K, Lee S-T, Yamada Y, Nakano R, Yano H, Mikasa K. 2016. Bacteremia secondary to *Alloscardovia omnicolens* urinary tract infection. *J Infect Chemother* 22:424–425. <https://doi.org/10.1016/j.jiac.2015.12.013>
- Huys G, Vancanneyt M, D'Haene K, Falsen E, Wauters G, Vandamme P. 2007. *Alloscardovia omnicolens* gen. nov., sp. nov., from human clinical samples. *Int J Syst Evol Microbiol* 57:1442–1446. <https://doi.org/10.1099/ijs.0.64812-0>
- Mahlen SD, Clarridge JE. 2009. Site and clinical significance of *Alloscardovia omnicolens* and *Bifidobacterium* species isolated in the clinical laboratory. *J Clin Microbiol* 47:3289–3293. <https://doi.org/10.1128/JCM.00555-09>
- Brown MK, Forbes BA, Stitley K, Doern CD. 2016. Defining the clinical significance of *Alloscardovia omnicolens* in the urinary tract. *J Clin Microbiol* 54:1552–1556. <https://doi.org/10.1128/JCM.03084-15>
- 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>
- 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>
- 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>
- 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>

9. Tully B. 2016. Quality assessment: FastQC v1. <https://doi.org/10.17504/protocols.io/fa3bign>
10. 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>
11. Langmead B. 2025. Kraken 2, KrakenUniq and Bracken indexes. Available from: <https://benlangmead.github.io/aws-indexes/k2>
12. 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>
13. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
14. 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>
15. 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>
16. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretidin 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>
17. 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>
18. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, et al. 2020. The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>
19. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA 3rd, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>
20. Razghonova Y, Mika A, Czapiewska M, Stanczak A, Zygowska P, Wydra DG, Sledzinski T, Abacjew-Chmylko A. 2025. Endometrial cancer is associated with altered metabolism and composition of fatty acids. *Int J Mol Sci* 26:3322. <https://doi.org/10.3390/ijms26073322>
21. Li J, Yang H, Zhang L, Zhang S, Dai Y. 2023. Metabolic reprogramming and interventions in endometrial carcinoma. *Biomedicine Pharmacotherapy* 161:114526. <https://doi.org/10.1016/j.biopha.2023.114526>
22. Skorupa A, Poński M, Ciszek M, Cichoń B, Klimek M, Witek A, Pakuła S, Boguszewicz Ł, Sokół M. 2021. Grading of endometrial cancer using ¹H HR-MAS NMR-based metabolomics. *Sci Rep* 11:18160. <https://doi.org/10.1038/s41598-021-97505-y>