![AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements]

Complete Genome Sequences of *Ezakiella coagulans* C0061C1 and *Fenollaria massiliensis* C0061C2

Michael T. France,[a,b] Jack Clifford,[a,b] Shilpa Narina,[a,b*] Lindsay Rutt,[a,b] Jacques Ravel[a,b]

aInstitute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA
bDepartment of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

ABSTRACT *Ezakiella coagulans* and *Fenollaria massiliensis* are two obligate anaerobic bacteria in the family *Peptoniphilaceae* and are both uncommon members of the human vaginal microbiota. We isolated a strain of each bacterium from the same vaginal swab specimen and here report the first complete genome sequences of the two species.

*E*zakiella coagulans* and *Fenollaria massiliensis* are obligate anaerobic bacteria in the family *Peptoniphilaceae* (1, 2). Both species are infrequent members of the human vaginal microbiota (VMB), and, when present, typically represent only a small fraction of the community (3). The VMB is often dominated by single species of *Lactobacillus*, and such communities have been associated with decreased risk for several adverse health outcomes (4). However, both *E. coagulans* and *F. massiliensis* are typically not found to be coresident with *Lactobacillus* species and instead cohabit communities that are characterized by a more even collection of other obligate and facultative anaerobes (3), e.g., *Gardnerella*, *Prevotella*, and *Atopobium*. The relationship between vaginal health and *E. coagulans* and *F. massiliensis* has yet to be determined. Here, we report the complete genome sequences of these two species.

One strain each of *E. coagulans* (C0061C1) and *F. massiliensis* (C0061C2) were isolated from a midvaginal swab specimen collected from a woman who identified as African American (Fig. 1, right). Swab material was first resuspended in 1 mL of *Brucella* broth (with hemin and vitamin K) and then plated on human blood Tween bilayer agar (5). The plate was incubated for 72 h under anaerobic conditions (gas mixture: 5% H_2, 10% CO_2, and 85% N_2), and then the strains were isolated. Genomic DNA was extracted, as described previously (5), with the MasterPure complete DNA purification kit (Lucigen, Middleton, WI, USA), sequencing libraries were prepared using the SMRTBell Express template preparation kit 2.0 (Pacific Biosciences [PacBio], Menlo Park, CA, USA), and libraries were size selected targeting 17 kbp using a BluePippin system (Sage Science, Beverly, MA, USA). Sequencing was performed using a Sequel II instrument and single-molecule real-time (SMRT) Cell 8M (PacBio) and produced 2,431,328 reads for C0061C1 ($15,387\times$ coverage; read N_{50}, 12,999 bp) and 2,951,446 reads for C0061C2 ($23,111\times$ coverage; read N_{50}, 12,997 bp). Read correction, trimming, and assembly were performed using Canu (6) (v2.1.1; coverage limit, $1,500\times$; target genome size, 2 Mbp; minimum read length, 1 kbp). Each assembly produced a single large contig, which was circularized using Simple-Circularise (https://github.com/Kzra/Simple-Circularise) (v1; default settings), rotated using Circlator (7) (v1.5.5; default settings), and taxonomically identified using GTDB-Tk (8) (v1.5.1; default settings).

E. coagulans C0061C1 had a genome that was 1.99 Mbp in length (GC content, 35.2%) and encoded 3 rRNA operons, 49 tRNAs, and 1,904 additional coding sequences. The genome of *F. massiliensis* C0061C2 had a smaller length of 1.61 Mbp (GC content, 34.7%) and encoded 4 rRNA operons, 48 tRNAs, and 1,514 additional coding sequences. Prediction and annotation of coding sequences were performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (9–11) (v6.0; default settings). Available whole-genome sequences of other members of *Peptoniphilaceae* and the outgroup *Peptostreptococcus anaerobius* were acquired
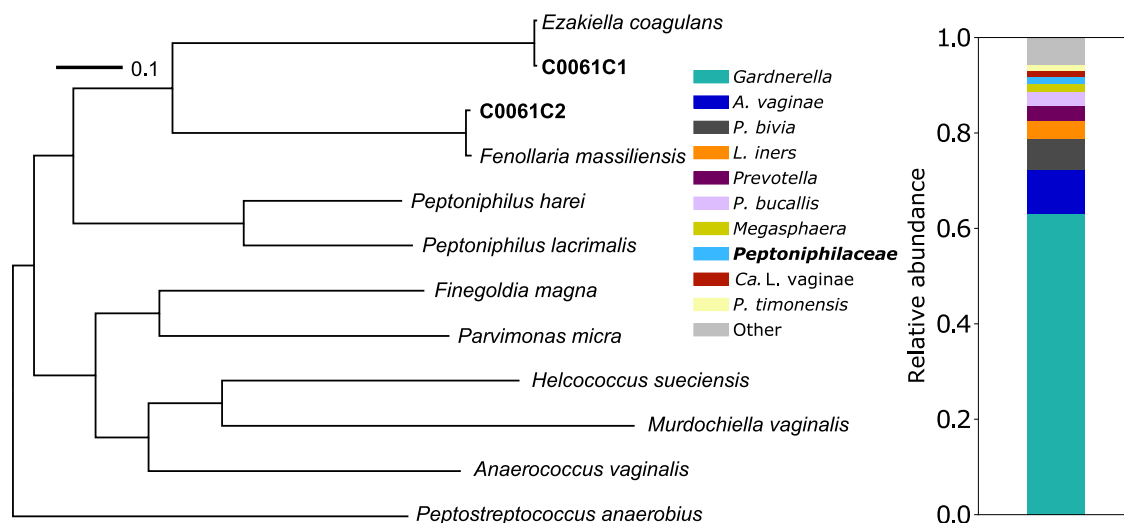
FIG 1 (Right) Taxonomic composition of the VMB at the time of isolation, estimated using previously published shotgun metagenomic data (15) and the VIRGO nonredundant gene catalog (16). (Left) Phylogenetic tree constructed from a concatenated alignment of the amino acid sequences of 514 single-copy core genes. Novel whole-genome sequences of strains C0061C1 and C0061C2 (bold) are most similar to conspecific draft genome sequences. Available genome sequences of *E. coagulans* (GenBank assembly accession number GCF_003096635.1), *F. massiliensis* (GenBank assembly accession number GCF_000312505.2), *Peptoniphilus harei* (GenBank assembly accession number GCF_900638565.1), *Peptoniphilus lacrimalis* (GenBank assembly accession number GCF_000378725.1), *Finegoldia magna* (GenBank assembly accession number GCF_013267535.1), *Parvimonas micra* (GenBank assembly accession number GCF_003454775.1), *Helcococcus sueciensis* (GenBank assembly accession number GCF_000423145.1), *Murdochiella vaginalis* (GenBank assembly accession number GCF_900119705.1), *Anaerococcus vaginalis* (GenBank assembly accession number GCF_016127475.1), and *Peptostreptococcus anaerobius* (GenBank assembly accession number GCF_900454605.1) were obtained from the NCBI RefSeq database.

from the RefSeq database, and genes present in all strains, including C0061C1 and C0061C2, were identified using orthoMCL (12) (v2.0.9; sequence identity, 70%; E value, $\leq 10^{-5}$), aligned using ClustalW2 (13) (v2.1; default settings), and concatenated. A phylogenetic tree (Fig. 1, left) was then constructed from the alignment using RAxML-NG (14) (v1.0.1; molecular evolutionary model, LG; starting trees, 10 parsimony and 10 random [bootstrapping converged after 50 replicates]). *E. coagulans* C0061C1 and *F. massiliensis* C0061C2 were most closely related to draft genome sequences from conspecific strains, and the two species were sister to one another.

The swab specimen was collected after informed consent was obtained from the participant, who also provided consent for storage of the material and its use in future research studies related to women's health. The original study was approved by the University of Maryland School of Medicine institutional review board.

Data availability. The *E. coagulans* and *F. massiliensis* genome sequences were deposited in NCBI GenBank with the accession numbers CP096650 and CP096649, respectively. The raw PacBio sequence reads were deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA832302, with the SRA accession numbers SRX15015913 and SRX15015914. All scripts used in the assembly and analysis of the genome sequences are available at https://github.com/ravel-lab/Pepton_WGS.

REFERENCES

1. Garcia-Lopez M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, Hahnke RL, Goker M. 2019. Analysis of 1,000 type-strain genomes improves taxonomic classification of *Bacteroidetes*. Front Microbiol 10:2083. https://doi.org/10.3389/fmicb.2019.02083.

2. Pagnier I, Croce O, Robert C, Raoult D, La Scola B. 2014. Non-contiguous finished genome sequence and description of *Fenollaria massiliensis* gen. nov. sp. nov., a new genus of anaerobic bacterium. Stand Genomic Sci 9:704–717. https://doi.org/10.4056/sigs.3957647.

3. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, Waetjen LE, Brotman RM, Ravel J. 2020. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. Microbiome 8:166. https://doi.org/10.1186/s40168-020-00934-6.

4. France M, Alizadeh M, Brown S, Ma B, Ravel J. 2022. Towards a deeper understanding of the vaginal microbiota. Nat Microbiol 7:367–378. https://doi.org/10.1038/s41564-022-01083-2.

5. France MT, Rutt L, Narina S, Arbaugh S, McComb E, Humphrys MS, Ma B, Hayward MR, Costello EK, Relman DA, Kwon DS, Ravel J. 2020. Complete genome sequences of six *Lactobacillus iners* strains isolated from the human vagina. Microbiol Resour Announc 9:e000234-20. https://doi.org/10.1128/MRA.00234-20.

6. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

7. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.

8. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

9. 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.

10. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.

11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

12. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503.

13. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

14. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

15. France MT, Fu L, Rutt L, Yang H, Humphrys MS, Narina S, Gajer PM, Ma B, Forney LJ, Ravel J. 2022. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. Genome Biol 23:66. https://doi.org/10.1186/s13059-022-02635-9.

16. Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, Brotman RM, Ravel J. 2020. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. Nat Commun 11:940. https://doi.org/10.1038/s41467-020-14677-3.