

2016

Genome sequences of 14 Firmicutes strains isolated from the human vagina

Grace E. Deitzler
Washington University School of Medicine in St. Louis

Maria J. Ruiz
Washington University School of Medicine in St. Louis

Cory Weimer
Washington University School of Medicine in St. Louis

SoEun Park
Washington University School of Medicine in St. Louis

Lloyd Robinson
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

Recommended Citation

Deitzler, Grace E.; Ruiz, Maria J.; Weimer, Cory; Park, SoEun; Robinson, Lloyd; Hallsworth-Pepin, Kymberlie; Wollam, Aye; Mitreva, Makedonka; Lewis, Amanda L.; and Lewis, Warren G., "Genome sequences of 14 Firmicutes strains isolated from the human vagina." *Genome Announcements*. 4, 5. 2016. (2016).
https://digitalcommons.wustl.edu/open_access_pubs/5629

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

Authors

Grace E. Deitzler, Maria J. Ruiz, Cory Weimer, SoEun Park, Lloyd Robinson, Kymberlie Hallsworth-Pepin, Aye Wollam, Makedonka Mitreva, Amanda L. Lewis, and Warren G. Lewis

Genome Sequences of 14 *Firmicutes* Strains Isolated from the Human Vagina

Grace E. Deitzler,^{a,d} Maria J. Ruiz,^{a,d} Cory Weimer,^{a,d} SoEun Park,^{a,d} Lloyd Robinson,^{a,d} Kymberlie Hallsworth-Pepin,^e Aye Wollam,^e Makedonka Mitreva,^{c,e} Amanda L. Lewis,^{a,b,d} Warren G. Lewis^{c,d}

Department of Molecular Microbiology,^a Department of Obstetrics and Gynecology,^b Department of Medicine,^c Center for Women's Infectious Disease Research,^d and McDonnell Genome Institute,^e Washington University School of Medicine, St. Louis, Missouri, USA

For a commentary on this article, see [doi:10.1128/genomeA.00890-16](https://doi.org/10.1128/genomeA.00890-16).

Research on vaginal infections is currently limited by a lack of available fully sequenced bacterial reference strains. Here, we present strains (now available through BEI Resources) and genome sequences for a set of 14 vaginal isolates from the phylum *Firmicutes*. These genome sequences provide a valuable resource for future research in understanding the role of Gram-positive bacteria in vaginal health and disease.

Received 3 July 2016 Accepted 4 August 2016 Published 29 September 2016

Citation Deitzler GE, Ruiz MJ, Weimer C, Park S, Robinson L, Hallsworth-Pepin K, Wollam A, Mitreva M, Lewis AL, Lewis WG. 2016. Genome sequences of 14 *Firmicutes* strains isolated from the human vagina. *Genome Announc* 4(5):e00888-16. doi:10.1128/genomeA.00888-16.

Copyright © 2016 Deitzler 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/).

Address correspondence to Amanda L. Lewis, allewis@wustl.edu, or Warren G. Lewis, WarrenGLewisPhD@gmail.com.

A variety of health outcomes have been associated with specific bacteria or bacterial patterns colonizing the human vagina. For example, bacterial vaginosis (BV) is a common dysbiosis of the vaginal microbiota that is associated with increased risks of sexually transmitted infections and serious complications during pregnancy, such as intrauterine infections and preterm labor (1–6). BV is generally characterized by the absence of vaginal *Lactobacillus* and overgrowth of diverse microbes, including *Firmicutes*, such as *Anaerococcus*, *Finegoldia*, *Megasphaera*, *Peptoniphilus*, and *Veillonella* species. However, little is known about the strategies these bacteria use to colonize the vagina or whether they are involved in the causes and complications associated with BV. We isolated 15 *Firmicutes* from vaginal swabs of healthy and BV-affected pregnant and nonpregnant women. While some of these bacterial species may have sequenced reference strains, there are very few strains from the human vagina whose genomes have been previously sequenced.

Vaginal swabs were collected from nonpregnant and pregnant women according to Washington University institutional review board (IRB)-approved protocols 201108155 and 20110382, respectively. Organisms were isolated from these swabs anaerobically on solid medium. Isolation procedures and clinical information will be described in a future publication. The culture conditions for each strain have been provided to BEI Resources. Sequencing of 16S rRNA genes >1,400 bp were used to initially assign bacteria genus and species identifications by comparison with the NCBI ribosomal database. Genomic DNA was obtained using the Wizard genomic DNA purification kit (Promega).

De novo assembly of genomes were conducted using the One Button Velvet assembly pipeline (version 1.1.06) (7), with hash sizes of 31, 33, and 35 after downsizing the sample input data to 100× coverage. A minimum length for contigs was set (post-assembly) at 200 bp. We performed a screen for core genes (as defined by the HMP [8]) on many of the assemblies to test for com-

TABLE 1 Strain names and accession numbers

Species	Strain	BEI catalog no.	Nucleotide sequence accession no.
<i>Anaerococcus tetradius</i>	MJR8151	HMS-1268	LRPM00000000
<i>Bacillus coagulans</i>	GED7749B	HMS-1281	LRPN00000000
<i>Clostridium perfringens</i>	MJR7757A	HMS-1290	LRPU00000000
<i>Enterococcus faecium</i>	MJR8396B	HMS-1267	LRPV00000000
<i>Finegoldia magna</i>	GED7760A	HMS-1285	LRPW00000000
<i>Megasphaera</i> sp.	MJR9396C	HMS-1269	LRVC00000000
<i>Peptoniphilus harei</i>	CMW7756A	HMS-1297	LRQE00000000
<i>Peptostreptococcus anaerobius</i>	MJR8628A	HMS-1263	LSQZ00000000
<i>Staphylococcus lugdunensis</i>	MJR7738	HMS-1293	LRQI00000000
<i>Staphylococcus simulans</i>	MJR7712	HMS-1283	LRQJ00000000
<i>Streptococcus mitis</i>	CMW7705B	HMS-1296	LRQR00000000
<i>Streptococcus pasteurianus</i>	GED7275A	HMS-1273	LSRA00000000
<i>Streptococcus salivarius</i>	GED7778A	HMS-1287	LRQS00000000
<i>Veillonella atypica</i>	CMW7756B	HMS-1301	LRQT00000000

pleteness of the genome. Gene annotation was performed using both *ab initio* and evidence-based (BLAST) predictions. Coding sequences were predicted using GeneMark and Glimmer3 (9, 10). Intergenic regions not identified by GeneMark and Glimmer3 were searched by BLAST in NCBI's nonredundant bacterial (NR) database. The best prediction for each open reading frame was selected by evaluating all predictions against the best evidence (nonredundant bacterial, NR, and Pfam [11]) and resolving overlaps between adjacent coding genes. tRNA genes were determined using tRNAscan-SE (12) and noncoding RNA genes by RNAmmer (13) and Rfam (14). Metabolic pathways and subcellular localization were predicted using KEGG (15) and PSORTb (16), respectively, and functional domains were evaluated using InterProScan (17).

Accession number(s). These whole-genome shotgun projects have been deposited in GenBank under the accession numbers listed in Table 1. The sequences described in this paper are the first versions. We have also made the strains available to the research community by depositing them with the Biodefense and Emerging Infections (BEI) Research Resource Repository (see BEI numbers in Table 1).

ACKNOWLEDGMENTS

We thank Justin Perry for technical assistance, Deborah Frank for editorial assistance, and the Women's and Infant's Health Specimen Consortium (especially Michele Landeau) for assistance with the collection of vaginal swabs from pregnant women. We also thank Jenifer Allsworth, Jeff Peipert, and Jennifer Bick of the Contraceptive CHOICE project for providing specimens from nonpregnant women. We also appreciate the generosity of the women who participated in this research.

This work was funded by the Burroughs-Wellcome Fund preterm Birth Initiative (A.L.L.). Summer stipends were provided by the Planning Center for Research in Benign Urology Summer Research Program P20DK097798 (G.E.D.) and the Washington University in St. Louis Global Health Center of the Institute for Public Health (M.J.R.).

FUNDING INFORMATION

This work, including the efforts of Makedonka Mitreva, was funded by HHS | National Institutes of Health (NIH) (5U54HG004968). This work, including the efforts of Grace E. Deitzler, was funded by HHS | National Institutes of Health (NIH) (P20DK097798). This work, including the efforts of Amanda L. Lewis and Warren G. Lewis, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (NIH R01 AI114635). This work, including the efforts of Amanda L. Lewis, was funded by Burroughs Wellcome Fund (BWF).

REFERENCES

- Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. 2008. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS* 22:1493–1501. <http://dx.doi.org/10.1097/QAD.0b013e3283021a37>.
- Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, Cotch MF, Edelman R, Pastorek JG, Jr, Rao AV. 1995. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *N Engl J Med* 333:1737–1742. <http://dx.doi.org/10.1056/NEJM199512283332604>.
- Hitti J, Hillier SL, Agnew KJ, Krohn MA, Reisner DP, Eschenbach DA. 2001. Vaginal indicators of amniotic fluid infection in preterm labor. *Obstet Gynecol* 97:211–219.
- Holst E, Goffeng AR, Andersch B. 1994. Bacterial vaginosis and vaginal microorganisms in idiopathic premature labor and association with pregnancy outcome. *J Clin Microbiol* 32:176–186.
- Wiesenfeld HC, Hillier SL, Krohn MA, Landers DV, Sweet RL. 2003. Bacterial vaginosis is a strong predictor of *Neisseria gonorrhoeae* and *Chlamydia trachomatis* infection. *Clin Infect Dis* 36:663–668. <http://dx.doi.org/10.1086/367658>.
- Zhang X, Xu X, Li J, Li N, Yan T, Ju X. 2002. Relationship between vaginal sialidase bacteria vaginosis and chorioamnionitis. *Zhonghua Fu Chan Ke Za Zhi* 37:588–590.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
- Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486:215–221. <http://dx.doi.org/10.1038/nature11209>.
- Borodovsky M, Mills R, Besemer J, Lomsadze A. 2003. Prokaryotic gene prediction using GeneMark and GeneMark.Hmm. *Curr Protoc Bioinformatics* Chapter 4:Unit 4.5. <http://dx.doi.org/10.1002/0471250953.bi0405s01>.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641. <http://dx.doi.org/10.1093/nar/27.23.4636>.
- Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. 2008. The Pfam protein families database. *Nucleic Acids Res* 36:D281–D288. <http://dx.doi.org/10.1093/nar/gkm960>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 33:D121–D124. <http://dx.doi.org/10.1093/nar/gki081>.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280. <http://dx.doi.org/10.1093/nar/gkh063>.
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization sub-categories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615. <http://dx.doi.org/10.1093/bioinformatics/btq249>.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* 33:W116–W120. <http://dx.doi.org/10.1093/nar/gki442>.