JAMIA Open, 2024, **7(3)**, ooae099
https://doi.org/10.1093/jamiaopen/ooae099
Application Notes

Application Notes

VMAP: Vaginal Microbiome Atlas during Pregnancy

Antonio Parraga-Leo, MS[1,2,3,4,5], Tomiko T. Oskotsky, MD[1,2,3], Boris Oskotsky, PhD[1,2], Camilla Wibrand, MBBS[1,2,3,6], Alennie Roldan, BS[1,2,3], Alice S. Tang, PhD[1,2,3], Connie W.Y. Ha, PhD[7,8], Ronald J. Wong, PhD[9], Samuel S. Minot, PhD[10], Gaia Andreoletti, PhD[1,2,3], Idit Kosti, PhD[1,2,3], Kevin R. Theis, PhD[11], Sherrianne Ng, PhD[12,13], Yun S. Lee, PhD[12,13], Patricia Diaz-Gimeno, PhD[5], Phillip R. Bennett, MD, PhD[12,13], David A. MacIntyre, PhD[12,13], Susan V. Lynch, PhD[7,8], Roberto Romero, MD, DMedSci[14,15,16], Adi L. Tarca, PhD[14,17,18,19], David K. Stevenson, MD[9,20], Nima Aghaeepour, PhD[9,21,22], Jonathan L. Golob, MD, PhD[3,23], Marina Sirota, PhD[1,2,3,*]

[1]Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA 94143, United States, [2]Department of Pediatrics, University of California San Francisco, San Francisco, CA 94158, United States, [3]March of Dimes Prematurity Research Center at the University of California San Francisco, San Francisco, CA 94143, United States, [4]Department of Pediatrics, Obstetrics and Gynaecology, Universidad de Valencia, Valencia 46010, Spain, [5]IVIRMA Global Research Alliance, IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia 28023, Spain, [6]Department of Pediatrics and Adolescent Medicine, Aarhus University Hospital, Aarhus 8200, Denmark, [7]Benioff Center for Microbiome Medicine, Department of Medicine, University of California, San Francisco, CA 94143, United States, [8]Division of Gastroenterology, Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, United States, [9]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, United States, [10]Data Core, Fred Hutchinson Cancer Center, Seattle, WA 98109, United States, [11]Department of Biochemistry, Microbiology and Immunology, Wayne State University, Detroit, MI 48201, United States, [12]Imperial College Parturition Research Group, Institute of Reproductive and Developmental Biology, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom, [13]March of Dimes Prematurity Research Centre at Imperial College London, London W12 0NN, United Kingdom, [14]Pregnancy Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, US Department of Health and Human Services, Bethesda, MD 20892, United States, [15]Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI 48109, United States, [16]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, United States, [17]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, United States, [18]Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201, United States, [19]Department of Computer Science, Wayne State University College of Engineering, Detroit, MI 48202, United States, [20]Center for Academic Medicine, Stanford University School of Medicine, Stanford, CA 94304, United States, [21]Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Stanford, CA 94305, United States, [22]Department of Biomedical Data Sciences, Stanford University School of Medicine, Stanford, CA 94305, United States, [23]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, United States

*Corresponding author: Marina Sirota, PhD, Bakar Computational Health Sciences Institute, Department of Pediatrics, University of California San Francisco, 490 Illinois Street, Floor 2, San Francisco, CA 94143, United States (marina.sirota@ucsf.edu)

Antonio Parraga-Leo and Tomiko T. Oskotsky contributed equally and are considered co-first authors of this work.

Jonathan L. Golob and Marina Sirota contributed equally and are considered co-last authors of this work.

Abstract

Objectives: To enable interactive visualization of the vaginal microbiome across the pregnancy and facilitate discovery of novel insights and generation of new hypotheses.

Material and Methods: Vaginal Microbiome Atlas during Pregnancy (VMAP) was created with R shiny to generate visualizations of structured vaginal microbiome data from multiple studies.

Results: VMAP (http://vmapapp.org) visualizes 3880 vaginal microbiome samples of 1402 pregnant individuals from 11 studies, aggregated via open-source tool MaLiAmPi. Visualized features include diversity measures, VALENCIA community state types, and composition (phylotypes, taxonomy) that can be filtered by various categories.

Discussion: This work represents one of the largest and most geographically diverse aggregations of the vaginal microbiome in pregnancy to date and serves as a user-friendly resource to further analyze vaginal microbiome data and better understand pregnancies and associated outcomes.

Conclusion: VMAP can be obtained from https://github.com/msirota/vmap.git and is currently deployed as an online app for non-R users.

Received: February 29, 2024; **Revised:** August 13, 2024; **Editorial Decision:** September 5, 2024; **Accepted:** September 9, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of the American Medical Informatics Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Lay Summary

The vaginal microbiome plays a significant role in birth outcomes, including preterm and spontaneous labor. Despite the increasing number of microbiome studies being performed, the information remains separated and difficult to compare due to technical challenges in its aggregation. This fact hinders analyses of large and widely-representative data. To address this drawback, we present the Vaginal Microbiome Atlas during Pregnancy (VMAP), a web application that enables the visualization of 11 different microbiome studies, including a total of 3880 vaginal microbiome samples aggregated leveraging an open-source tool, MaLiAmPi. VMAP offers an online platform for exploring robust vaginal microbiome data in the context of pregnancy from multiple studies, providing interactive visualization of diverse microbiome features, allowing a better understanding of the role of the vaginal microbiome in preterm birth, and facilitating the generation of new research hypotheses.

Key words: bioinformatics; 16S rRNA sequencing; microbiome; visualization; data integration.

Introduction

The vaginal microbiome plays a significant role in birth outcomes, including spontaneous and preterm labor.[1–7] An increasing number of microbiome studies in the past decade, thanks to advances in high-throughput sequencing, have permitted greater insights into the role of the microbiome in human health and disease[8]; however, efforts to aggregate microbiome datasets to better understand the microbiome through analysis of large, widely-representative data has been precluded by technical challenges.[9] In particular, the dominant approach to harmonize microbiome data include closed reference operational taxonomic units (OTUs) and projection to taxonomy. The main drawback is this technology is highly dependent on the matching of microbiome communities to the reference which hinders the process of harmonization.[10] In this work, we present an atlas of the vaginal microbiota during pregnancy leveraging data across 11 studies harmonized by the open-source tool, MaLiAmPi.[10] This algorithm places 16S rRNA gene variable region amplicon sequence variants onto a common phylogenetic tree, successfully correcting for the technical differences and retaining more true entropy. Although other microbiome atlases exist, such as microbioTA,[11] here we present a specific atlas of vaginal microbiome across pregnancy that encompasses a total of 11 projects. This dataset represents one of the largest and most geographically diverse aggregations of the vaginal microbiome in pregnancy to date—3880 samples from 1402 pregnant individuals—harmonized into a set of generalizable features suitable for integration of new microbiota data post-hoc. A rich set of paired metadata is included, including collection week during gestation (by specimen), week of delivery, and maternal NIH racial category. Approximately one-third of the pregnancies in this data set were of women delivering preterm, including just over 1/8th who delivered early preterm. This dataset and features derived from these data can be interactively visualized via our Vaginal Microbiome Atlas during Pregnancy (VMAP) application, a resource for understanding the vaginal microbiome in the context of pregnancy (http://vmapapp.org).

Materials and methods

The dataset was constructed by aggregating and processing vaginal microbiome data from the public domain[12–16] in addition to including newly generated data not yet released to the public. The publicly available data were from 9 studies, representing 3578 samples across 1268 individuals of whom 851 delivered term and 247 preterm (32-37 weeks of gestation) and 170 early preterm (before 32 weeks of gestation). Two additional unpublished datasets are included. One is from Wayne State University consisting of 159 samples across 60 individuals among whom 40 (66.7%) had term deliveries and 15 (25.0%) had preterm deliveries and 5 (8.3%) who had early preterm deliveries. A newly generated dataset which comprises 143 vaginal microbiome samples from 74 individuals, up to 3 samples (one sample per trimester) for each individual, with 44 individuals (59.5%) having term deliveries, 25 individuals (33.8%) having preterm deliveries, and 5 individuals (6.8%) having early preterm deliveries.[17]

To harmonize the raw amplicon sequence reads from such a diverse set of underlying approaches, we developed a novel phylogenetic-based approach implemented in an open source workflow called MaLiAmPi.[10] Briefly, MaLiAmPi is a Nextflow pipeline based on 4 important steps: (1) The generation of ASVs from FASTQ files, (2) The selection of a repository of 16S rRNA alleles, (3) The generation of a reference package containing a phylogenetic tree, and (4) The placement of these ASVs onto the reference phylogenetic tree. This pipeline is a promising approach for harmonizing microbiome data improving the current techniques such as those based on closed-OTUs.

We further compute the alpha-diversity of communities (diversity measures include Shannon, Inverse Simpson, Balance weighted phylogenetic diversity [bwpd], phylogenetic entropy, quadratic, unrooted phylogenetic diversity, and rooted phylogenetic diversity), weighted phylogenetic (KR) distance between communities, provide taxonomic assignments to each ASV, and cluster ASVs into phylotypes which serve as a taxonomy-independent representative of (phylogenetically) closely related organisms. In addition, VALENCIA[18] was used to provide the microbiome community state type (CST) of each sample.

VMAP is an interactive web application implemented in R shiny (version = 1.7.5.1) and HTML 5 to visualize, filter, and explore microbiome data.

Results

VMAP (http://vmapapp.org) interactively visualizes vaginal microbiome features during healthy term or adverse-outcome associated pregnancies across 11 studies, representing 3880 samples from 1402 individuals, of whom 935 delivered at term, 287 delivered at preterm, and 180 whose deliveries were early preterm. The demographics and information about the source studies are presented in Table 1. The vaginal microbiome data are from individuals of different racial backgrounds, including Black or African American, White, American Indian, Alaska native, Asian, and unknown, and of varying age. Additionally, samples were collected during the 3 pregnancy trimesters allowing insight into microbiome changes across time (Table S1). Figure 1 depicts a schematic of all the features that are captured by the atlas.

Table 1. Summary of studies, individuals, samples, and 16S rRNA gene sequencing information.

Studies	Accession IDs	SDY465, PRJEB11895, PRJEB21325, PRJEB30642, PRJNA242473, PRJNA294119, PRJNA393472, PRJNA430482, PRJNA504518, PRJEB12577, SDY2187
	Centers	Imperial College London, Stanford University, University of Maryland, University of Pennsylvania, Virginia Commonwealth, Washington University, Wayne State University
Individuals	*n*	1402
Age range, *n* (%)	Unknown	578 (41.2)
	Below 28	383 (27.3)
	Above 29	441 (31.5)
Race, *n* (%)	Race: American Indian or Alaska Native	8 (0.6)
	Race: Asian	84 (6.0)
	Race: Black or African American	824 (58.9)
	Race: Native Hawaiian or Other Pacific Islander	3 (0.5)
	Race: White	414 (29.6)
	Race: unknown	69 (5.0)
Ethnicity, *n* (%)	Ethnicity: Hispanic or Latino	46 (3.3)
	Ethnicity: others	95 (6.8)
	Ethnicity: unknown	1261 (89.9)
Delivery, *n* (%)	Term	935 (66.7)
	Preterm	287 (20.5)
	Early preterm	180 (12.8)
Samples	*n*	3880
Delivery, *n* (%)	Term	2779 (71.6)
	Preterm	745 (19.2)
	Early preterm	356 (9.2))
Trimester, *n* (%)	First	283 (7.3)
	Second	2635 (67.9)
	Third	962 (24.8)
Study, *n* (%)	SDY465 (A)	231 (6.0)
	PRJEB11895 (B)	33 (0.9)
	PRJEB21325 (C)	144 (3.7)
	PRJEB30642 (D)	134 (3.5)
	PRJNA242473 (E)	168 (4.3)
	PRJNA294119 (F)	145 (3.7)
	PRJNA393472 (G)	957 (24.7)
	PRJNA430482 (H)	216 (5.6)
	PRJNA504518 (I)	1467 (37.8)
	PRJEB12577 (J)	83 (2.1)
	SDY2187 (S)	143 (3.6)
	SDY2187 (W)	159 (4.1)
16S rRNA sequencing	V Region sequences	V1—V2, V1—V3, V3—V4, V3—V5, or V4
	Instruments	454 GS FLX Titanium, Illumina HiSeq 2500, Illumina HiSeq 4000, Illumina MiSeq, or Illumina NextSeq 550

This tool allows customizable visualization of the data at several levels including demographics of the cohort at the sample or individual level organized by outcome (term, preterm, early preterm), race and ethnicity distributions, as well as breakdown per project (see web tutorial). Demographics are shown as stacked barplots where the relative frequency of each type of category selected can be seen. Furthermore, stacked barplots are updated according to the demographic feature selected (Figure S1). Alpha-diversity measures are also shown in several ways including overall correlation plots of the individual measures and their distribution as well as violin plots across various stratifications (Figure S2). Different diversity measures are customizable allowing users to select their measures of interest with a legend below the chart explaining each of them. CSTs are visualized via alluvial plots longitudinally across trimesters allowing for various stratifications of interest. The alluvial plots allow the observation of changes in the relative composition of CSTs across trimesters according to the feature selected, representing each stream as an individual CST (Figure S3A). The prevalence and relative abundance of closely related microbes (as represented by phylotypes) or taxa are displayed as heatmaps (Figure S3B). Dimensionality reduction plots, namely Uniform Manifold Approximation and Projections (UMAPs), are used to visualize phylotype data per sample or individual comparing different features (Figure S4). Finally, it is worth noting that VMAP contains a tutorial on what each chart represents and how to use the filters to facilitate the use of this application.

Discussion

VMAP provides a means to view a rich set of features derived from vaginal microbiota data (including alpha-diversity, CSTs, and the composition of specific microbes and taxa) associated with core outcomes (delivery group) and metadata (eg, NIH racial categories, maternal age) that can be sliced and viewed interactively. VMAP can be a resource for those wishing to relate and understand vaginal microbiota data within the context of pregnancy. The generalizability of these features across studies using distinct underlying techniques

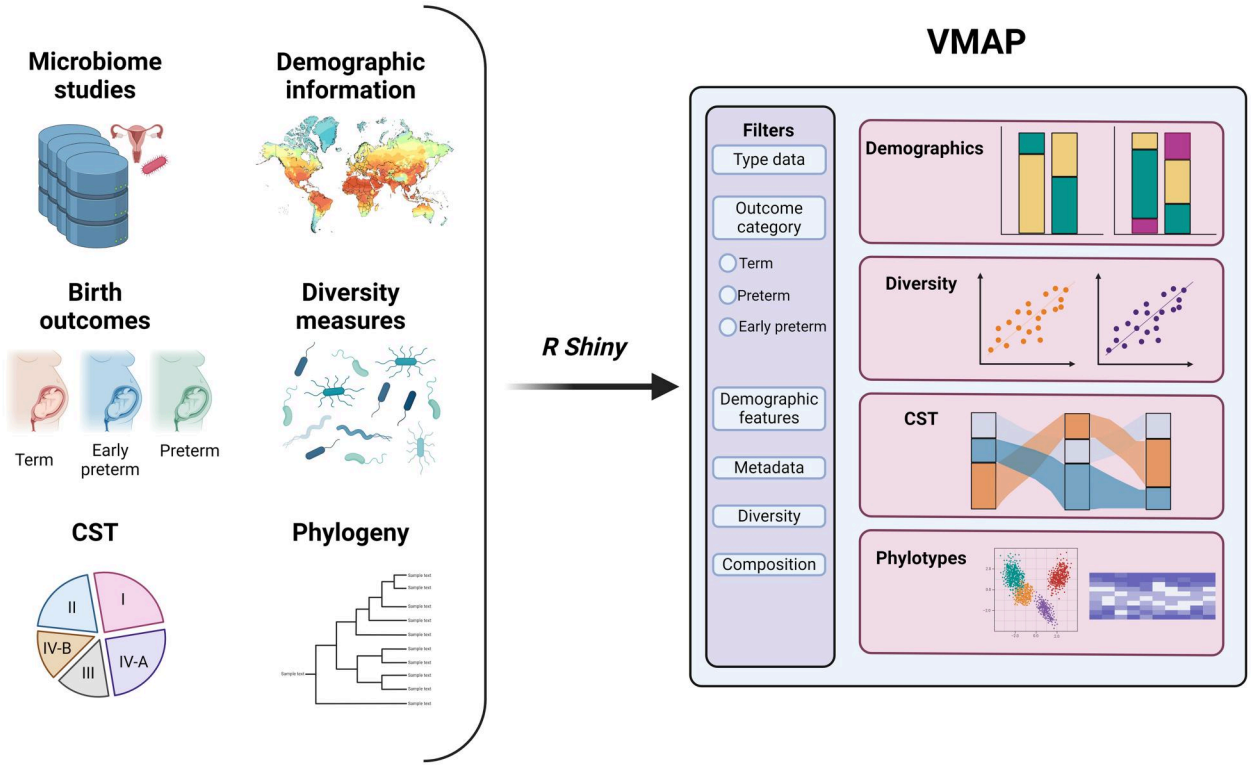


Figure 1. VMAP RShiny App. VMAP encompasses information from different microbiome studies gathering demographic information, birth outcomes, diversity measures, CST, and phylogeny (left). A user can visualize demographic features of the cohort, diversity measures (including correlations between them) by outcome of interest, alluvial plots of CSTs, dimensionality reduction plots of phylogeny, and heatmaps of taxonomical features (right). The visualizations are customizable by the filters in the bar on the left, allowing one to choose which projects and samples to visualize. Abbreviation: CST = community state types. Created with BioRender.com.

(eg, targeting different 16S rRNA gene variable regions or novel sequencing platforms) and the ability to integrate new data into the existing set of features are significant advances for microbiome research. For example, this data set and approach were successfully employed as the basis for a machine learning ("DREAM") challenge predicting preterm and early-preterm birth where over 300 teams participated with the goal of building machine learning models to predict which women would deliver preterm.[17] This challenge used this novel approach to integrate data *post-hoc* to validate machine learning models against datasets unavailable to participants. VMAP will enable the scientific community to visualize and study all of this information, both to better understand the role of the microbiome in preterm birth and to generate new research hypotheses. In addition, filtering by metadata information can reveal trends in microbiome data in specific groups of interest, such as age groups or trimesters.

The resource has limitations that should be considered. It is based on publicly available data which might not have full clinical or demographic annotations of the samples in the metadata. While the sample size of the study is considerable with representation of individuals of diverse backgrounds, it may not be representative of the entire population of pregnant women from around the world.

This work serves as the basis for several potential follow-up opportunities. Our VMAP visualization resource can further be extended to include non-pregnancy datasets as well as microbiome data across body sites by leveraging the MaLiAmPi harmonization technique that can incorporate additional datasets post hoc.[10] In addition, other adverse pregnancy outcomes such as recurrent pregnancy loss can be investigated and included in the expansion of this resource with the aim of observing differences and similarities as well as improving the understanding of these reproductive conditions. Finally, the microbiome data can further be integrated with other omics measures to better understand healthy human pregnancy and those associated with adverse outcomes. VMAP will be valuable for more robust interpretation of novel datasets seeking to relate the vaginal microbiome to pregnancy outcomes—serving as a large and regularized set of data and metadata for comparison.

Conclusion

VMAP offers a platform for exploring robust vaginal microbiome data in the context of pregnancy from multiple studies, providing interactive visualization and assessment of diverse features. Despite current limitations such as clinical annotation accuracy, VMAP opens avenues into further work—extending to other outcomes and incorporating additional microbiome data including non-pregnancy datasets, data from other body sites, and other omics measures—for a deeper understanding of human health.

Acknowledgments

We thank members of the Sirota Lab, University of California, San Francisco, for useful discussion.

Author contributions

Jonathan L. Golob and Marina Sirota conceived the study. Connie W.Y. Ha, Ronald J. Wong, and Adi L. Tarca generated and shared data for the validation dataset. Jonathan L. Golob, Tomiko T. Oskotsky, Alice S. Tang, Alennie Roldan, and Samuel S. Minot aggregated the training datasets. Jonathan L. Golob and Alennie Roldan normalized the training and validation datasets. Antonio Parraga-Leo and Jonathan L. Golob developed the application. Tomiko T. Oskotsky, Jonathan L. Golob, and Marina Sirota were major contributors to writing the manuscript. All authors read and approved the final manuscript.

Supplementary material

Supplementary material is available at *JAMIA Open* online.

Funding

This study was supported by the March of Dimes (A.P.-L., T.T. O., A.R., A.S.T., C.W.Y.H., R.J.W., G.A., I.K., S.N., Y.S.L.L., P.R.B., D.A.M., S.V.L., D.K.S., N.A., J.L.G., M.S.); Spanish Ministry of Science, Innovation and Universities through FPU program [FPU18/01777, EST22/00170] (A.P.-L.); Instituto de Salud Carlos III (ISCIII) through Miguel Servet program [CP20/00118] and co-funded by European Union (P.D.-G); and the National Institutes of Health [R35GM138353, 1R01HL139844, 3P30AG066515, 1R61NS114926, 1R01AG058417] (N.A.), [R01HD105256] (N.A., M.S.), and P01HD106414 (N.A.); the Burroughs Welcome Fund (N.A.); the Alfred E. Mann Foundation (N.A.); and the Robertson foundation (N.A.).

This work was supported, in part, by the Pregnancy Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, United States Department of Health and Human Services (NICHD/NIH/DHHS); and, in part, by federal funds from NICHD/NIH/DHHS (Contract No. HHSN275201300006C). Dr. Roberto Romero has contributed to this work as part of his official duties as an employee of the United States Federal Government. Dr. Tarca is also supported by the Wayne State University Perinatal Initiative for Maternal, Perinatal and Child Health.

Conflicts of interest

Antonio Parraga-Leo and Patricia Diaz-Gimeno are receiving hononaria from the IVI Foundation. All other authors declare no financial or non-financial competing interests.

Data availability

The RShiny app is available: http://vmapapp.org.

Datasets were downloaded from ImmPort[13] via the March of Dimes Preterm Birth database[12] (Study SDY465), from the NCBI Sequence Read Archive[15] (BioProjects PRJNA242473, PRJNA294119, PRJNA393472, and PRJNA430482), the Sequence Read Archive of the European Nucleotide Archive[16] (Projects PRJEB11895, PRJEB12577, PRJEB21325, and PRJEB30642), and the database of Genotypes and Phenotypes (dbGaP)[14] (accession number phs001739.v1.p1).

Additional associated metadata were requested and obtained from the RAMS Registry (https://ramsregistry.vcu.edu) (PRJNA430482) or from the senior author (Projects PRJEB11895, PRJEB12577, PRJEB21325, and PRJEB30642), or downloaded from the publication by the Kindinger et al[4,5] (PRJEB11895 and PRJEB12577).

Data for accession number phs001739.v1.p1 are exclusively available via dbGaP after following the application procedures there.

The aggregated vaginal microbiome dataset can be downloaded from the March of Dimes Prematurity Research Database excluding the datasets that we are unable to share due to legal restrictions (**SDY2187**).

Ethics statement

This work was approved by the National Heart, Lung, and Blood Institute (NHLBI) Clinical Data Science Institutional Review Board (CDS-IRB) in study number 2021-040, and reliance was granted to the NHLBI CDS-IRB by the University of California, San Francisco Institutional Review Board in study number 21-35274.

References

1. Hyman RW, Fukushima M, Jiang H, et al. Diversity of the vaginal microbiome correlates with preterm birth. *Reprod Sci*. 2014;21:32-40. https://doi.org/10.1177/1933719113488838
2. DiGiulio DB, Callahan BJ, McMurdie PJ, et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci USA*. 2015;112:11060-11065. https://doi.org/10.1073/pnas.1502875112
3. Callahan BJ, DiGiulio DB, Goltsman DSA, et al. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. *Proc Natl Acad Sci USA*. 2017;114:9966-9971. https://doi.org/10.1073/pnas.1705899114
4. Kindinger LM, Bennett PR, Lee YS, et al. The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. *Microbiome*. 2017;5:6. https://doi.org/10.1186/s40168-016-0223-9
5. Kindinger LM, MacIntyre DA, Lee YS, et al. Relationship between vaginal microbial dysbiosis, inflammation, and pregnancy outcomes in cervical cerclage. *Sci Transl Med*. 2016;8:350ra102. https://doi.org/10.1126/scitranslmed.aag1026
6. Brown RG, Marchesi JR, Lee YS, et al. Vaginal dysbiosis increases risk of preterm fetal membrane rupture, neonatal sepsis and is exacerbated by erythromycin. *BMC Med*. 2018;16:9. https://doi.org/10.1186/s12916-017-0999-x
7. Brown RG, Al-Memar M, Marchesi JR, et al. Establishment of vaginal microbiota composition in early pregnancy and its association with subsequent preterm prelabor rupture of the fetal membranes. *Transl Res*. 2019;207:30-43. https://doi.org/10.1016/j.trsl.2018.12.005
8. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4516-4522. https://doi.org/10.1073/pnas.1000080107
9. Schmidt TSB, Matias Rodrigues JF, von Mering C. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ Microbiol*. 2015;17:1689-1706. https://doi.org/10.1111/1462-2920.12610
10. Minot SS, Garb B, Roldan A, et al. MaLiAmPi enables generalizable and taxonomy-independent microbiome features from technically diverse 16S-based microbiome studies. *Cell Rep Methods*. 2023;3:100639. https://doi.org/10.1016/j.crmeth.2023.100639

11. Wang P, Zhang S, He G, et al. *microbioTA*: an atlas of the microbiome in multiple disease tissues of *Homo sapiens* and *Mus musculus*. *Nucleic Acids Res*. 2023;51:D1345-D1352. https://doi.org/10.1093/nar/gkac851
12. Sirota M, Thomas CG, Liu R, et al.; March of Dimes Prematurity Research Centers. Enabling precision medicine in neonatology, an integrated repository for preterm birth research. *Sci Data*. 2018;5:180219. https://doi.org/10.1038/sdata.2018.219
13. Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res*. 2014;58:234-239. https://doi.org/10.1007/s12026-014-8516-1
14. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181-1186. https://doi.org/10.1038/ng1007-1181
15. Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. 2011;39:D19-D21. https://doi.org/10.1093/nar/gkq1019
16. Leinonen R, Akhtar R, Birney E, et al. The European nucleotide archive. *Nucleic Acids Res*. 2011;39:D28-31. https://doi.org/10.1093/nar/gkq967
17. Golob JL, Oskotsky TT, Tang AS, et al.; Preterm Birth DREAM Community. Microbiome preterm birth DREAM challenge: crowdsourcing machine learning approaches to advance preterm birth research. *MedRxiv Prepr Serv Health Sci*. 2023:2023.03.07.23286920. https://doi.org/10.1101/2023.03.07.23286920
18. France MT, Ma B, Gajer P, et al. Valencia: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome*. 2020;8:166. https://doi.org/10.1186/s40168-020-00934-6

© The Author(s) 2024. Published by Oxford University Press on behalf of the American Medical Informatics Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
JAMIA Open, 2024, 7, 1–6
https://doi.org/10.1093/jamiaopen/ooae099
Application Notes