Identification of shared bacterial strains in the vaginal microbiota of reproductive-age mothers and daughters using genome-resolved metagenomics

Running title: Strain sharing in the VM of mothers and daughters

M. France[1,2], S. Brown[3], A. Rompalo[4], R. M. Brotman[1,3], J. Ravel[*1,2]

[1] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD USA.

[2] Department of Microbiology and Immunology, University of Maryland School of Medicine

[3] Department of Epidemiology and Public Health, University of Maryland School of Medicine

[4]Division of Infectious Diseases, John Hopkins School of Medicine

*Corresponding author, available at jravel@som.umaryland.edu

Competing Interests

Dr. Ravel is a co-founder of LUCA Biologics, a biotechnology company focused on translating microbiome research into live biotherapeutics for women's health. All other authors declare that they have no competing interests.

Abstract

It has been suggested that the human microbiome might be vertically transmitted from mother to offspring and that early colonizers may play a critical role in development of the immune system. Studies have shown limited support for the vertical transmission of the intestinal microbiota but the derivation of the vaginal microbiota remains largely unknown. Although the vaginal microbiota of children and reproductive age cis women differ in composition, the vaginal microbiota could be vertically transmitted. To determine whether there was any support for this hypothesis, we examined the vaginal microbiota of daughter-mother pairs from the Baltimore metropolitan area (ages 14-27, 32-51; n=39). We assessed whether the daughter's microbiota was similar in composition to their mother's using metataxonomics. Permutation tests revealed that while some pairs did have similar vaginal microbiota, the degree of similarity did not exceed that expected by chance. Genome-resolved metagenomics was used to identify shared bacterial strains in a subset of the families (n=22). We found a small number of bacterial strains that were shared between mother-daughter pairs but identified more shared strains between individuals from different families, indicating that vaginal bacteria may display biogeographic patterns. Earlier-in-life studies are needed to demonstrate vertical transmission of the vaginal microbiota.

Importance

Early colonizers of our microbiota are theorized to play an important role in the development of our immune system, yet we know little about how these communities are established. Vertical transmission from mother to offspring at the time of birth is theorized to be a major source of early colonizers but limited evidence supporting this process has only been shown for the intestinal tract microbiota. The provenance of the vaginal microbiota is largely unknown, although some have posited it is similarly vertically transmitted. We examined the vaginal microbiota of mother-daughter

pairs and found limited evidence in support of this hypothesis. However, our analysis also revealed putative biogeographic patterns in the distribution of the strains which comprise the vaginal microbiota. Our results give insight into the role of vertical transmission for the vaginal microbiota and motivate future studies on the biogeography of these bacteria.

Introduction

The human body is colonized by microbial populations which together comprise our microbiota (1). These populations have been shown to be critical determinants of our health and well-being (2, 3) and are founded early in life (4). Initial colonization of newborn infants primarily occurs during and immediately following birth (5, 6), although there is an active debate on whether *in utero* seeding plays a role in this early colonization (7-13). Establishment of the microbiota is theorized to be critical to the programming of the neonatal immune system (14-17). The maternal microbiota has long been hypothesized to be a major contributor of microbial strains to their newborn offspring through a process of vertical transmission (18). As the neonate moves through the vaginal canal, it is expected to be exposed to the mother's vaginal microbiota and perhaps their fecal and skin microbiota. It follows then that the microbiota of neonates born via C-section has been observed to transiently differ in composition from those born via vaginal delivery (14, 19, 20). Studies seeking to demonstrate this process of vertical transmission have identified shared bacterial phylotypes in the microbiota of mothers and their neonates using 16S rRNA amplicon sequencing (21, 22). However, these data lack the resolution necessary to identify shared strains (23). The most convincing evidence for vertical transmission comes from studies which either used cultivation or shotgun metagenomic based techniques to identify bacterial strains in the microbiota of mothers and their infants (24-28). Yet, these probable vertically transmitted strains have been shown to not comprise the majority of the neonate's microbiota and to be short-lived in the neonatal microbiota (24). More study is needed to define the provenance of a neonate's microbiota.

Much of the work on maternal microbiota transmission has focused on the neonate's intestinal, skin, or oral microbiota (24-28). The source of the bacterial species and strains that inhabit the

vagina is not known. Reproductive-age cis women routinely have communities which are dominated by *Lactobacillus* with *L. crispatus*, *L. iners*, *L. jensenii*, and *L. gasseri* being the most prevalent species (29, 30). A significant proportion of these women, however, have communities which do not contain a high relative abundance of lactobacilli and instead are characterized by a more even distribution of several obligate or facultative anaerobes including species in the *Gardnerella*, *Atopobium*, and *Prevotella* genera (29, 30). These *Lactobacillus* deficient communities are more common among women of Hispanic and African descent (29-31) and have been associated with increased risk for adverse health events, including reproductive tract infections (32-36). Less is known about the microbial communities which comprise the vaginal microbiota of pre-pubertal children, but they have been shown to differ in composition and in bacterial load from those found in reproductive-age cis women (37-39). Any bacteria which are transferred at the time of birth may not be capable of surviving in a child's vagina, which has been shown to have neutral or alkaline pH and to have a paucity of *Lactobacillus* (38). It is not until early puberty that the species which are common in a reproductive age cis women's vaginal microbiota (e.g. *L. crispatus*, *L. iners*, *G. vaginalis*) gain dominance in an adolescent's vaginal microbiota (40). While it is thought that pubertal hormonal and physiological changes which occur are responsible for this shift in the composition of the vaginal microbiota (41-43), it is not clear where the strains come from. It could be that they are of maternal origin and have persisted at low abundance throughout early-life or that they are acquired later in life through some other mechanism.

To determine whether there was evidence for the vertical transmission of the vaginal microbiota which persisted into adolescence and adulthood, we characterized the vaginal microbiota of pre-menopausal mothers and their post-menarcheal daughters. Metataxonomics was used to

investigate similarities in community composition between the mother-daughter pairs and genome-resolved metagenomics was used to identify bacterial strains which the two had in common.

Results

Similarity in the taxonomic composition of mothers and their daughters

We first asked whether mothers and their daughters had vaginal microbiota which were similar in taxonomic composition. Metataxonomics was used to assess the composition of the vaginal microbiota for 42 mothers and their 45 daughters (Figure 1). One family had two daughters, and another had three (26 & 23, respectively). Similarity between communities was assessed using Yue & Clayton's θ, which is function of the relative abundances of shared and non-shared species in the communities. While there were several examples of mother-daughter pairs which had very similar in taxonomic composition (e.g. families 1, 8, 23), there were also several which did not (e.g. families 6, 15, 24). The average similarity between mother-daughter pairs was 0.3, indicating that most were found to have communities with different compositions. Similarity between mother-daughter pairs was higher when the daughter had a community state type (CST) that was not dominated by *Lactobacillus* (CST IV, Figure 2A).

Because the taxonomic composition of the human vaginal microbiota routinely resembles one of a limited number of configurations, it is expected that there can be a degree of similarity between entirely unrelated individuals. To determine whether the observed similarity between mother-daughter pairs exceeded that expected by chance alone we used a permutation test. Taxonomic profiles were shuffled and the similarity between these randomized mother-daughter pairs was assessed in the same manner. As can be seen in Figure 2B, the distribution of similarity scores for the permuted data did not differ substantially from the observed distribution. The observed data was found to have slightly fewer pairs with θ value between 0-0.1 ($p=0.02$); and slightly more pairs

with a θ value between 0.1-0.2 (p=0.02). This result indicates that the observed similarity in composition between mother-daughter pairs was not different than that for randomly selected mother-daughter pairs.

Genome resolved metagenomics

In the above analysis we demonstrated that most mother-daughter pairs did not have vaginal microbiota with similar taxonomic profiles. Yet, many pairs were found to have species in common, just at different relative abundances (e.g. *L. crispatus* in family 4). To determine whether the populations of these shared species were comprised of the same strain(s), we selected 22 families (47 participants) to conduct shotgun metagenomic sequencing (denoted by black diamonds in Figure 1). The resulting metagenomes were assembled and binned allowing us to recover 225 near-complete MAGs (Figure 3). We recovered about 5 MAGs per metagenome with a minimum of 1 and a maximum of 13. The recovered MAGs included: 14 *Atopobium*, 12 "*Ca.* Lachnocurva vaginae", 53 *Gardnerella*, 16 *L. crispatus*, 27 *L. iners*, 6 *L. jensenii*, 29 *Prevotella*, 10 *S. amnii*, and 10 *S. sanguinegens*. The remaining 48 MAGs comprised 24 species. We were able to recover four *Gardnerella* MAGs from a single metagenome which contained *Gardnerella.*

Identification of shared strains

To identify pairs of MAGs which originate from the same bacterial strain we used the inStrain tool and associated workflow. If the vaginal microbiota is vertically transmitted, we expected that mothers and daughters which had species in common, might also have strains in common. Furthermore, if this transmission had happened at the time of birth, the mother and daughter strains should also not be identical but instead show some degree of sequence divergence consistent with the amount of time past. For this reason, we used a stringent threshold for defining shared strains of at least 99.9% sequence identity and at least 70% overlap. Among our set of 225 MAGs, we identified 49 pairs which met this threshold. Among these, ten were between a mother and

daughter from the same family, representing six mother-daughter pairs. These ten pairs of strains were found at both high and low relative abundances and their abundance was generally similar between the mother's and daughter's communities (Table 1). Daughters in pairings which were found to share strains were younger at the time of sampling than those in pairings which were not found to share strains (15.17 versus 19.62; t=-4.31; p<0.001). No trend was observed with the mother's age (39.6 versus 41.9; t=-0.81; p=0.45). Of these six mother-daughter pairs, three were found to share *L. crispatus* strains (families 1, 3, and 5) with family 5 also sharing MAGs classified as Clostridiales and *P. timonensis*. Family 30 shared MAGs classified *Gardnerella*, *L. iners*, and *A. rimae* while family 21 shared only "*Ca*. L. vaginae" and family 41, only *P. timonensis*. The remaining 39 instances of shared strains were between daughters and mothers from different families (n=18), mothers from different families (n=13), and daughters from different families (n=8). These pairings were parsed into a diagram representing the network of shared strains among the mothers and daughters in this study (Figure 4). A large part of this network was comprised of a *L. crispatus* and a *L. jensenii* strain which were identified in five and four metagenomes, respectively.

Comparison of sequence identity among strain shared within versus between families

We next asked whether the strains identified as shared within mother-daughter pairs were more or less similar than those shared between families. Sequence similarity was measured as the number of singe nucleotide polymorphisms (SNPs) per megabase pair of aligned sequence. We found strains shared within families trended to being more similar to one another than those shared between families, but this difference was not significant ($t_{15.1}$=-1.52, p=0.149). For the strains shared within families these values were used in combination with the daughter's age to calculate the per year substitution rate under the hypothesis of vertical transmission at the time of birth. These values ranged from $1.05*10^{-6}$ to $3.24*10^{-5}$ per base pair per year and are listed in Table 1.

Discussion

The origin of the bacterial strains which constitute the vaginal microbiota is not currently known. We found limited evidence for vertical transmission of these bacteria from a mother to her daughter which had persisted through the daughter's adolescence. While some mother-daughter pairs were found to have communities of similar taxonomic composition, the observed similarities could be explained by chance. A small subset of the mother-daughter pairs were also found to have bacterial strains in common, consistent with vertical transmission, but shared strains were more frequently identified in unrelated individuals. These results do not eliminate vertical transmission as a possible mechanism by which the vaginal microbiota is founded but rather suggest that mothers and daughters do not necessarily have similar vaginal microbiota, later in life. Because we examined the communities years after the birth of the daughter, there was plenty of time for either the mother's or the daughter's vaginal microbiota to experience strain turnover. Longitudinal studies have indicated that the vaginal microbiota of reproductive-age cis women does experience changes in composition over time, although the studies followed women for only a few months (44, 45). Daughters that were found to share strains with their mother were also younger than those which were not, further suggesting that time may play a role. It is not difficult to imagine that the populations of some bacterial strains might go extinct over the course of a person's life. The mechanisms by which new bacterial strains might be introduced into the vaginal microbiota are not well understood. Unprotected vaginal sex and other sexual practices could result in the introduction of new bacterial strains (46-48), but there are likely other mechanisms as well. Strain turnover in either the mother or the daughter's vaginal microbiota would erode any signal of vertical transmission.

In our analysis, six mother-daughter pairs were found to have matching bacterial strains in their vaginal microbiota. Under the vertical transmission at the time of birth, substitutions are expected to accumulate in both the mother's and the daughter's populations, as they evolve independently post-transmission. We used the daughter's age to calculate the substitution rate under this hypothesis and arrived at values between $3*10^{-5}$ and $1*10^{-6}$ substitutions per site-year for each shared strain. It is difficult to say how if our observed values fit the vertical transmission narrative, as the expected substitution rate is not well understood. A study on the population genomics of *Neisseria gonorrhoeae*, a sexually transmitted pathogen, estimated a rate of $3*10^{-5}$ substitutions per site-year (49). Another study surveyed of the substitution rate experienced by a number of human pathogens estimated rates between 10^{-5} and 10^{-8} substitutions per site-year (50). They also found a strong negative relationship between the estimated substitution relationship and the timescale over it was measured. The authors suggest that this relationship results from the accumulation of deleterious mutations which have yet to be purged by purifying selection (50). The timescale separating our hypothesized vertically transmitted strains is rather short, consistent with our relatively high estimated substitution rates. Many of the bacteria common to the human vagina have reduced genome sizes and have lost components of DNA repair machinery (51, 52). These bacteria may experience higher than average mutation rates (53-56) which could lend itself to higher estimates of their substitution rate (57). We cannot say for certain that our observation of shared strains among these six mother-daughter pairs is the result of the vertical transmission at time of birth, but we find this explanation is reasonable.

The majority of shared strains were identified in women from unrelated families. We hypothesize that this observation may reflect the biogeography of vaginal bacteria. The mothers and daughters included in this study were all living in the Baltimore metropolitan area at the time of sampling.

These individuals may be more likely to share bacterial strains with one another simply due to their geographic proximity. Lourens Baas Becking put forth the hypothesis that bacteria did not display biogeographic patterns, suggesting that in the microbial world "everything is everywhere, but the environment selects" (58). In the years since Bass-Becking put forth his hypothesis, there have been several demonstrations to the contrary (59-63). The dispersal of some bacterial species appears to be constrained leading them to exhibit biogeographic patterns. We suggest that many of the species common to the vaginal microbiota (e.g. *L. crispatus*, *L. iners*, *L. jensenii*, *G. vaginalis*, "Ca. L. vaginae", *A. vaginae*) are among those likely to have their dispersal constrained. With the except of *L. crispatus*, which is sometimes found in the intestinal tracts of chickens (64, 65), these species are not routinely found anywhere other than the vagina. Many of these bacteria are also fastidious and require anaerobic conditions for their robust growth. It is therefore not clear how they might disperse over the great distances necessary to erode biogeographic patterns, except by means of their host. If dispersal is primarily achieved via sexual activity, sexual networks could underpin biogeographic patterns observed for vaginal bacteria (66). Additional studies of participants from around the world might further illuminate these biogeographic patterns and the factors which govern them. Results from such studies could have translational impact and would be informative on the necessity of developing geo-adapted probiotic formulations to modulate the vaginal microbiota.

We implemented a stringent sequence similarity cutoff (≥99.9% similarity) to identify strains which may have been vertically transferred from mother to daughter. This is because it is not enough to identify instances where two bacterial assemblies belong to the same lineage. Their genome sequences must also be similar enough that any observed sequence differences can be explained by the post-transmission evolution of the two populations. In our case, the transmission was

hypothesized to have occurred at birth, meaning that the two populations had evolved independently, in most cases, for about two decades. Other studies which have examined the maternal transmission of microbes from mother to offspring have done so shortly after birth (24-26). In this case, there should be minimal sequence differences between the two populations of vertically transferred strains. It makes sense then, to utilize even more stringent sequence similarity cutoffs than that used here (e.g. ≥99.99% similarity). Marker gene based tools like StrainPhlan (23), do not have the capability to implement such genome-wide sequence similarity cutoffs and therefore may not be the best tool for this analysis. We, as have others (67), advocate for the use of appropriate sequence similarity cutoffs to identify recent microbial transfer events. Identifying the same species or even the same bacterial strain in two samples is not enough.

This study has a number of important limitations which should be considered. First, we examined the mother's and daughter's vaginal microbiota not at birth, but instead sometime after the daughter had experienced menarche. This means that there was plenty of time for the daughter or the mother to gain or lose bacterial strains in their communities. Second, we are missing a great deal of metadata which could help explain why some mother-daughter pairs were found to share bacterial strains, and some were not. For example, we do not know which daughters were born by cesarean section and which were vaginally delivered. Nor do we have sexual behavior or partner history data for the participants in this study. Third, the participants had originally been enrolled in a douching intervention study: douching may influence the composition of the vaginal microbiota (68). Finally, the cohort examined was relatively small and included only women in the greater Baltimore metropolitan area who identified as Black or African American.

Yet even with these limitations, we did identify several mother-daughter pairs which did share strains, which were similar in sequence enough to have been vertically transmitted at the time of birth. These results motivate future studies which investigate the extent to which the vaginal microbiota is vertically transmitted and the importance of this event to the daughter's future reproductive health. We also identified shared strains in unrelated individuals, suggesting that vaginal bacteria might display biogeographic patterns. These patterns could be confirmed by large scale, multi-regional studies which examine the extent to which strains of vaginal bacteria show geographic specificity.

Methods

Cohort description/sample collection

We characterized the vaginal microbiota of 87 reproductive-age cis women including 45 daughters and their 42 mothers. The average age of the daughters was 19 (14-35) and the average age of the mothers was 41 (32-51). Participants included in this study self-identified as Black or African American and took part in a douching intervention study (69). Vaginal swab specimens were collected from participants and stored at -80°C. The study was approved by the internal review boards at the University of Maryland and the Johns Hopkins University School of Medicine.

DNA extraction

DNA was extracted from 200 µL of vaginal swab specimen resuspended in 1ml of phosphate buffered saline transport medium. DNA extractions were performed using the MagAttract PowerMicrobiome DNA/RNA Kit (Qiagen; Hilden, Germany) and bead-beating on a TissueLyser II (Qiagen) according to the manufacturer's instructions and automated onto a Hamilton STAR robotic platform (Hamilton Robotics; Reno, NV, USA).

16S rRNA gene sequencing

The V3V4 region of the 16S rRNA gene was amplified and sequenced as described previously (70). The protocol utilizes two amplification steps: one which targets the V3V4 region and one which adds barcoded sequencing (primer sequences in supplemental table 1). Pooled amplicons were then deep sequenced an Illumina HiSeq 2500 (Illumina; San Diego, CA, USA) and the resulting paired end sequence reads were processed using DADA2 (71) to identify amplicon sequence variants (ASVs) and remove chimeric sequences as described previously (29). The median number of sequences per sample following processing was 16 990 (range: 182-36 401). Each ASV was assigned to a taxonomic group using the RDP Naïve Bayesian Classifier (72) trained with the SILVA 16S rRNA gene database (73). Genera common in the vaginal environment (e.g. *Lactobacillus*, *Gardnerella*, *Prevotella*, *Sneathia*, and *Mobiluncus*) were further classified at the species level using speciateIT (version 1.0, http://ravel-lab.org/speciateIT). The sequence counts attributed to ASVs assigned to the same phylotype were added together. Samples for which less than 500 sequences were generated were dropped from the analysis. Phylotypes with a study-wide average relative abundance of $< 10^{-4}$ were removed. The final dataset contained 81 samples and 100 phylotypes and can be found in Supplemental Table 1. Taxonomic profiles were assigned to community state types (CSTs) using VALENCIA (29). Similarity in taxonomic composition between mother-daughter pairs was assessed using the Yue-Clayton θ (74). A permutation test was performed to determine

whether these observed values differed with that expected by chance alone. Taxonomic profiles for mothers and daughters were each shuffled randomly and then similarity was assessed in the same manner. This process was repeated 100 times.

Shotgun metagenomics

Shotgun metagenomic data was generated for 22 families and included 47 total samples (22 mothers and 25 daughters). The samples selected for shotgun metagenomics are indicated in Figure 1. Families which either had similar taxonomic profiles or had species in common were selected for this analysis. Shotgun metagenomic sequence libraries were prepared from the extracted DNA using Illumina Nextera XT Flex kits according to manufacturer recommendations. The resulting libraries were sequenced on an Illumina HiSeq 4000 (10 per lane,150 bp paired-end mode) at the Genomic Resource Center at the University of Maryland School of Medicine. The average number of read pairs generated for each library was 37,000,000 (range: 24,500,000 to 81,800,000). Human reads were identified in the resulting sequence datasets using BMtagger and removed (ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/). Sequence datasets were further processed using sortmeRNA (75) to identify and remove ribosomal RNA reads and the remaining reads were trimmed for quality (4 bp sliding window, average quality score threshold Q15) using Trimmomatic v0.3653 (76). Reads trimmed to less than 75bp were removed from the dataset.

Genome resolved metagenomics

The taxonomic composition of each metagenome was established by mapping to the VIRGO non-redundant gene catalog (77). *De novo* assembly was performed on each metagenome using metaspades (78, 79) with k-mer sizes: 21, 33, 55, 77, 99, 101, and 127. The resulting assemblies were separated into single genome bins using a reference guided approach. For each metagenome, the sequence reads were mapped back to the corresponding assembly, to establish the contig coverage, and to the VIRGO gene catalog, to establish taxonomy of the contig. Contigs

317 demonstrating at least 5X coverage and which were found to have at least 90% of the reads
318 mapping to VIRGO genes with the same taxonomic annotation were separated into species bins.
319 The species bins were further split into metagenome assembled genomes (MAGs) based on
320 differences in contig coverage. Quality of the resulting MAGs were examined using checkM (80)
321 and those demonstrating at least 80% completion and less than 5% contamination were used in the
322 subsequent analyses (Supplemental Table 2). The average completeness of the MAGs was
323 97.04% (80.9%-100%) and the average contamination was 1.05% (0.0%-4.94%). Genes were
324 identified in each MAG using prodigal (81) and OrthoMCL was used to identify those which were
325 common to at least 95% of the MAGs (82). Thirteen such genes were identified and their amino
326 acid sequences were individually aligned using Muscle (83) and then concatenated into a single
327 alignment using phyutility (84). PartitionFinder was used to select an appropriate partitioning
328 scheme and model of molecular evolution (85). The Phylogeny of the 225 MAGs was established
329 using RaxML-ng with 10 parsimony and ten random starting trees (86). Bootstrap convergence was
330 detected using the autoMRE setting and occurred after 750 replicates. Relative abundance of
331 MAGs in their resident communities was approximated as the percent of reads from the
332 metagenome mapping to the MAG.

333 ## Identification of shared strains

334 Similarity between the MAGs was assessed using inStrain (67). An all-versus-all strategy was used
335 wherein a separate inStrain profile was built by mapping the sequence reads from each
336 metagenome against the MAGs recovered from each participant using Bowtie2 (87). This resulted
337 in 2209 inStrain profiles. For each participant, the set of 47 inStrain profiles were then summarized
338 using the inStrain compare function with Ward linkage, which afforded the determination of
339 coverage overlap and sequence similarity between the sequences reads of each metagenome and
340 the MAGs of each participant. We then applied a stringent sequence similarity threshold (70%

coverage and at least 99.9% sequence similarity) to identify participants with shared bacterial strains. A network diagram representing strain sharing was built from the using the NetworkX python package (https://networkx.org/). Sequence similarity between shared strains identified in the same family versus different families were compared using a student's t-test.

Data and code availability

Raw 16S rRNA amplicon and human removed shotgun metagenomic sequence data has been deposited in the NCBI SRA under the Bioproject PRJNA779415. Python and R scripts used in the analysis of these data and in the generation of figures are available at: https://github.com/ravel-lab/MotherDaughter.

Acknowledgements

This work was supported in part by funding from the National Institute of Child Health and Development, the National Institute of Allergy and Infectious Disease, and the National Institute of Nursing Research under the grants R01AI060892, R01NR015495, and R01AI116799.

Competing Interests

Dr. Ravel is a co-founder of LUCA Biologics, a biotechnology company focused on translating microbiome research into live biotherapeutics for women's health. All other authors declare that they have no competing interests.

References

1. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.

2. Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. Nat Rev Genet 13:260-70.

3. Pflughoeft KJ, Versalovic J. 2012. Human Microbiome in Health and Disease, vol 7, p 99-122. Annual Reviews

4. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. 2019. The Human Microbiome and Child Growth - First 1000 Days and Beyond. Trends Microbiol 27:131-147.

5. Gritz EC, Bhandari V. 2015. The human neonatal gut microbiome: a brief review. Front Pediatr 3:17.

6. Moore RE, Townsend SD. 2019. Temporal development of the infant gut microbiome. Open Biol 9:190128.

7. Leiby JS, McCormick K, Sherrill-Mix S, Clarke EL, Kessler LR, Taylor LJ, Hofstaedter CE, Roche AM, Mattei LM, Bittinger K, Elovitz MA, Leite R, Parry S, Bushman FD. 2018. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. Microbiome 6:196.

8. Perez-Munoz ME, Arrieta MC, Ramer-Tait AE, Walter J. 2017. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. Microbiome 5:48.

9. Rackaityte E, Halkias J, Fukui EM, Mendoza VF, Hayzelden C, Crawford ED, Fujimura KE, Burt TD, Lynch SV. 2020. Viable bacterial colonization is highly limited in the human intestine in utero. Nat Med 26:599-607.

10. Romero R, Miranda J, Chaemsaithong P, Chaiworapongsa T, Kusanovic JP, Dong Z, Ahmed AI, Shaman M, Lannaman K, Yoon BH, Hassan SS, Kim CJ, Korzeniewski SJ, Yeo L, Kim YM. 2015. Sterile and microbial-associated intra-amniotic inflammation in preterm prelabor rupture of membranes. J Matern Fetal Neonatal Med 28:1394-409.

11. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. 2014. The placenta harbors a unique microbiome. Science Translation Medicine 6:ra65237.

12. Silverstein RB, Mysorekar IU. 2021. Group therapy on in utero colonization: seeking common truths and a way forward. Microbiome 9:7.

13. Walter J, Hornef MW. 2021. A philosophical perspective on the prenatal in utero microbiome debate. Microbiome 9:5.

14. Wampach L, Heintz-Buschart A, Fritz JV, Ramiro-Garcia J, Habier J, Herold M, Narayanasamy S, Kaysen A, Hogan AH, Bindl L, Bottu J, Halder R, Sjoqvist C, May P, Andersson AF, de Beaufort C, Wilmes P. 2018. Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. Nat Commun 9:5091.

15. de Aguero MG, Ganal-Vanarburg SC, Fuhrer T, Rupp S, Uchimura Y, Li H, Steinert A, Heikenwalder M, Hapfelmeier S, Sauer U, McCoy KD, Macpherson AJ. 2016. The maternal microbiota drives early postnatal innate immune development. Science 351:1296-1302.

16. Dzidic M, Boix-Amoros A, Selma-Royo M, Mira A, Collado MC. 2018. Gut Microbiota and Mucosal Immunity in the Neonate. Med Sci (Basel) 6.

17. Sanidad KZ, Zeng MY. 2020. Neonatal gut microbiome and immunity. Curr Opin Microbiol 56:30-37.

18. Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG. 2015. The infant microbiome development: Mom matters, vol 21, p 109-117.

19. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A 107:11971-5.

20. Stinson LF, Payne MS, Keelan JA. 2018. A Critical Review of the Bacterial Baptism Hypothesis and the Impact of Cesarean Delivery on the Infant Microbiome, vol 5, p 135. Frontiers.

21. Mortensen MS, Rasmussen MA, Stokholm J, Brejnrod AD, Balle C, Thorsen J, Krogfelt KA, Bisgaard H, Sorensen SJ. 2021. Modeling transfer of vaginal microbiota from mother to infant in early life. Elife 10.

22. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J. 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe 17:690-703.

23. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 27:626-638.

24. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo M, Beghini F, Bertorelli R, De Sanctis V, Bariletti I, Canto R, Clementi R, Cologna M, Crifo T, Cusumano G, Gottardi S, Innamorati C, Mase C, Postai D, Savoi D, Duranti S, Lugli GA, Mancabelli L, Turroni F, Ferrario C, Milani C, Mangifesta M, Anzalone R, Viappiani A, Yassour M, Vlamakis H, Xavier R, Collado CM, Koren O, Tateo S, Soffiati M, Pedrotti A, Ventura M, Huttenhower C, Bork P, Segata N. 2018. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe 24:133-145 e5.

25. Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, Selvenius J, Oikarinen S, Hyoty H, Virtanen SM, Ilonen J, Ferretti P, Pasolli E, Tett A, Asnicar F, Segata N, Vlamakis H, Lander ES, Huttenhower C, Knip M, Xavier RJ. 2018. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. Cell Host Microbe 24:146-154 e4.

26. Wang S, Zeng S, Egan M, Cherry P, Strain C, Morais E, Boyaval P, Ryan CA, E MD, Ross RP, Stanton C. 2021. Metagenomic analysis of mother-infant gut microbiome reveals global distinct and shared microbial signatures. Gut Microbes 13:1-24.

27. Milani C, Mancabelli L, Lugli GA, Duranti S, Turroni F, Ferrario C, Mangifesta M, Viappiani A, Ferretti P, Gorfer V, Tett A, Segata N, van Sinderen D, Ventura M. 2015. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. Appl Environ Microbiol 81:7078-87.

28. Makino H, Kushiro A, Ishikawa E, Muylaert D, Kubota H, Sakai T, Oishi K, Martin R, Ben Amor K, Oozeer R, Knol J, Tanaka R. 2011. Transmission of intestinal Bifidobacterium longum subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. Appl Environ Microbiol 77:6788-93.

29. 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.

30. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women., vol 108, p 4680-4687.

31. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, Foster JA, Forney LJ. 2007. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. ISME J 1:121-33.

32. Ma B, Forney LJ, Ravel J. 2012. Vaginal Microbiome: Rethinking Health and Disease, vol 66, p 371-389.

33. Gosmann C, Anahtar MN, Handley SA, Huttenhower C, Farcasanu M, Abu-Ali G, Bowen BP, Padavattan N, Desai C, Droit L, Moodley A, Dong M, Chen Y, Ismail N, Ndung'u T, Ghebremichael MS, Wesemann DR, Mitchell CM, Dong KL, Huttenhower C, Walker BD, Virgin HW, Kwon DS. 2017. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women, vol 46, p 29-37.

34. Tamarelle J, de Barbeyrac B, Le Hen I, Thiebaut A, Bebear C, Ravel J, Delarocque-Astagneau E. 2018. Vaginal microbiota composition and association with prevalent Chlamydia trachomatis infection: a cross-sectional study of young women attending a STI clinic in France. Sex Transm Infect 94:616-618.

35. Elovitz MA, Gajer P, Riis V, Brown AG, Humphrys MS, Holm JB, Ravel J. 2019. Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. Nat Commun 10:1305.

36. Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. 2015. Lactobacillus species as biomarkers and agents that can promote various aspects of vaginal health, vol 6, p 1-18.

37. Gerstner GJ, Grunberger W, Boschitsch E, Rotter M. 1982. Vaginal organisms in prepubertal children with and without vulvovaginitis. Archives in gynecology 231:247-252.

38. Hammerschlag MR, Alpert S, Onderdonk AB, Thurston P, Ellen D, McCormack WM, Bartlett JG. 1978. Anaerobic microflora of the vagina in children. AJOG 131:853-856.

39. Hill GB, St. Claire KK, Gutman LT. 1995. Anaerobes predominate among the vaginal microflora of prepubertal girls. Clinical Infectious Diseases 20:S269-S270.

40. Hickey RJ, Zhou X, Settles ML, Erb J, Malone K, Hansmann MA, Shew ML, Pol BVD, Fortenberry JD, Forney LJ, Forney J. 2015. Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women, vol 6, p e00097-15.

41. Hickey RJ, Zhou X, Pierson JD, Ravel J, Forney LJ. 2012. Understanding vaginal microbiome complexity from an ecological perspective, vol 160, p 267-282.

42. Kaur H, Merchant M, Haque MM, Mande SS. 2020. Crosstalk Between Female Gonadal Hormones and Vaginal Microbiota Across Various Phases of Women's Gynecological Lifecycle. Front Microbiol 11:551.

43. Nunn KL, Ridenhour BJ, Chester EM, Vitzthum VJ, Fortenberry JD, Forney LJ. 2019. Vaginal Glycogen, Not Estradiol, Is Associated With Vaginal Bacterial Community Composition in Black Adolescent Women. J Adolesc Health 65:130-138.

44. Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, Koenig SSK, Fu L, Ma Z, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota, vol 4, p 132ra52.

45. Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, Agnew KJ, Marrazzo JM, Fredricks DN. 2010. Temporal variability of human vaginal bacteria and relationship with Bacterial vaginosis. PLoS ONE 5:e10197.

46. Mandar R, Punab M, Borovkova N, Lapp E, Kiiker R, Korrovits P, Metspalu A, Krjutskov K, Nolvak H, Preem JK, Oopkaup K, Salumets A, Truu J. 2015. Complementary seminovaginal microbiome in couples. Res Microbiol 166:440-447.

47. Vodstrcil LA, Walker SM, Hocking JS, Law M, Forcey DS, Fehler G, Bilardi JE, Chen MY, Fethers KA, Fairley CK, Bradshaw CS. 2015. Incident bacterial vaginosis (BV) in women who have sex with women is associated with behaviors that suggest sexual transmission of BV. Clin Infect Dis 60:1042-53.

48. Mehta SD, Zhao D, Green SJ, Agingu W, Otieno F, Bhaumik R, Bhaumik D, Bailey RC. 2020. The Microbiome Composition of a Man's Penis Predicts Incident Bacterial Vaginosis in His Female Sex Partner With High Accuracy. Front Cell Infect Microbiol 10:433.

49. Perez-Losada M, Crandall KA, Zenilman J, Viscidi RP. 2007. Temporal trends in gonococcal population genetics in a high prevalence urban community. Infect Genet Evol 7:271-8.

50. Duchene S, Holt KE, Weill FX, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016. Genome-scale rates of evolutionary change in bacteria. Microb Genom 2:e000094.

51. France MT, Mendes-Soares H, Forney LJ. 2016. Genomic comparisons of Lactobacillus crispatus and Lactobacillus iners reveal potential ecological drivers of community composition in the vagina, vol 82, p 7063-7073.

52. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G. 2011. At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1., vol 108, p 4688-4695.

53. Acosta S, Carela M, Garcia-Gonzalez A, Gines M, Vicens L, Cruet R, Massey SE. 2015. DNA Repair Is Associated with Information Content in Bacteria, Archaea, and DNA Viruses. J Hered 106:644-59.

54. Batut B, Knibbe C, Marais G, Daubin V. 2014. Reductive genome evolution at both ends of the bacterial population size spectrum. Nat Rev Microbiol 12:841-50.

55. Bourguignon T, Kinjo Y, Villa-Martin P, Coleman NV, Tang Q, Arab DA, Wang Z, Tokuda G, Hongoh Y, Ohkuma M, Ho SYW, Pigolotti S, Lo N. 2020. Increased Mutation Rate Is Linked to Genome Reduction in Prokaryotes. Curr Biol 30:3848-3855 e4.

56. Marais GA, Calteau A, Tenaillon O. 2008. Mutation rate and genome reduction in endosymbiotic and free-living bacteria. Genetica 134:205-10.

57. Bromham L. 2009. Why do species vary in their rate of molecular evolution? Biol Lett 5:401-4.

58. Baas-Becking LGM. 1934. Geobiologie of Inleiding Tot de Milieukunde. W. P. van Stockum & Zoon N. V., The Hague, The Netherlands.

59. Cho JC, Tiedje JM. 2000. Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil, vol 66, p 5448-5456.

60. Remold SKSK, Purdy-Gibson MEME, France MTMT, Hundley TCTC. 2015. Pseudomonas putida and Pseudomonas fluorescens species group recovery from human homes varies seasonally and by environment, vol 10, p e0127704.

61. Amenyogbe N, Dimitriu P, Smolen KK, Brown EM, Shannon CP, Tebbutt SJ, Cooper PJ, Marchant A, Goetghebuer T, Esser M, Finlay BB, Kollman TR, Mohn WW. 2020. Biogeography of the relationship between the child gut microbiome and innate immune system. mBio 12:e03079-20.

62. Li SP, Wang P, Chen Y, Wilson MC, Yang X, Ma C, Lu J, Chen XY, Wu J, Shu WS, Jiang L. 2020. Island biogeography of soil bacteria and fungi: similar patterns, but different mechanisms. ISME J 14:1886-1896.

63. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach AL, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map, vol 4, p 102-112.

64. Pan M, Hidalgo-Cantabrana C, Barrangou R. 2020. Host and body site-specific adaptation of Lactobacillus crispatus genomes. NAR Genomics and Bioinformatics 2.

65. Adhikari B, Kwon YM. 2017. Characterization of the Culturable Subpopulations of Lactobacillus in the Chicken Intestinal Tract as a Resource for Probiotic Development. Front Microbiol 8:1389.

66. Kenyon C, Buyze J, Klebanoff M, Brotman RM. 2018. The role of sexual networks in studies of how BV and STIs increase the risk of subsequent reinfection. Epidemiol Infect 146:2003-2009.

67. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. 2021. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol doi:10.1038/s41587-020-00797-0.

68. Brotman RM, Klebanoff MA, Nansel TR, Andrews WW, Schwebke JR, Zhang J, Yu KF, Zenilman JM, Scharfstein DO. 2008. A longitudinal study of vaginal douching and bacterial vaginosis--a marginal structural modeling analysis. Am J Epidemiol 168:188-96.

69. Mark H, Sherman SG, Nanda J, Chambers-thomas T, Barnes M, Rompalo A. 2010. What has Changed about Vaginal Douching among African American Mothers and Daughters ? 27:418-424.

70. Holm JB, Humphrys M, Robinson CK, Settles ML, Ott S, Fu L, Yang H, Gajer P, He X, McComb EJ, Gravitt P, Ghanem KG, Brotman RM, Ravel J. 2019. Ultra-high throughput multiplexing and sequencing of >500 bp amplicon regions on the Illumina HiSeq 2500 platform, vol 4, p e00029-19.

71. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data, vol 13, p 581-583.

72. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-7.

73. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Glo FO, Yarza P. 2013. The SILVA ribosomal RNA gene database project : improved data processing and web-based tools. 41:590-596.

74. Yue JC, Clayton MK. 2005. A Similarity Measure Based on Species Proportions. Communications in Statistics - Theory and Methods 34:2123-2131.

75. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211-3217.

76. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data., vol 30, p 2114-20. Oxford University Press.

77. 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.

78. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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 19:455-77.

79. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824-834.

80. 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 25:1043-55.

81. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokayotic gene recognition and translation site identification. BMC Bioinformatics 11:119.

82. Li L. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13:2178-2189.

83. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

84. Smith SA, Dunn CW. 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data, vol 24, p 715-716.

85. Lanfear R, Calcott B, Ho SY, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695-701.

86. 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.

87. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2, vol 9, p 357-359. Nature Publishing Group.

Figure legends

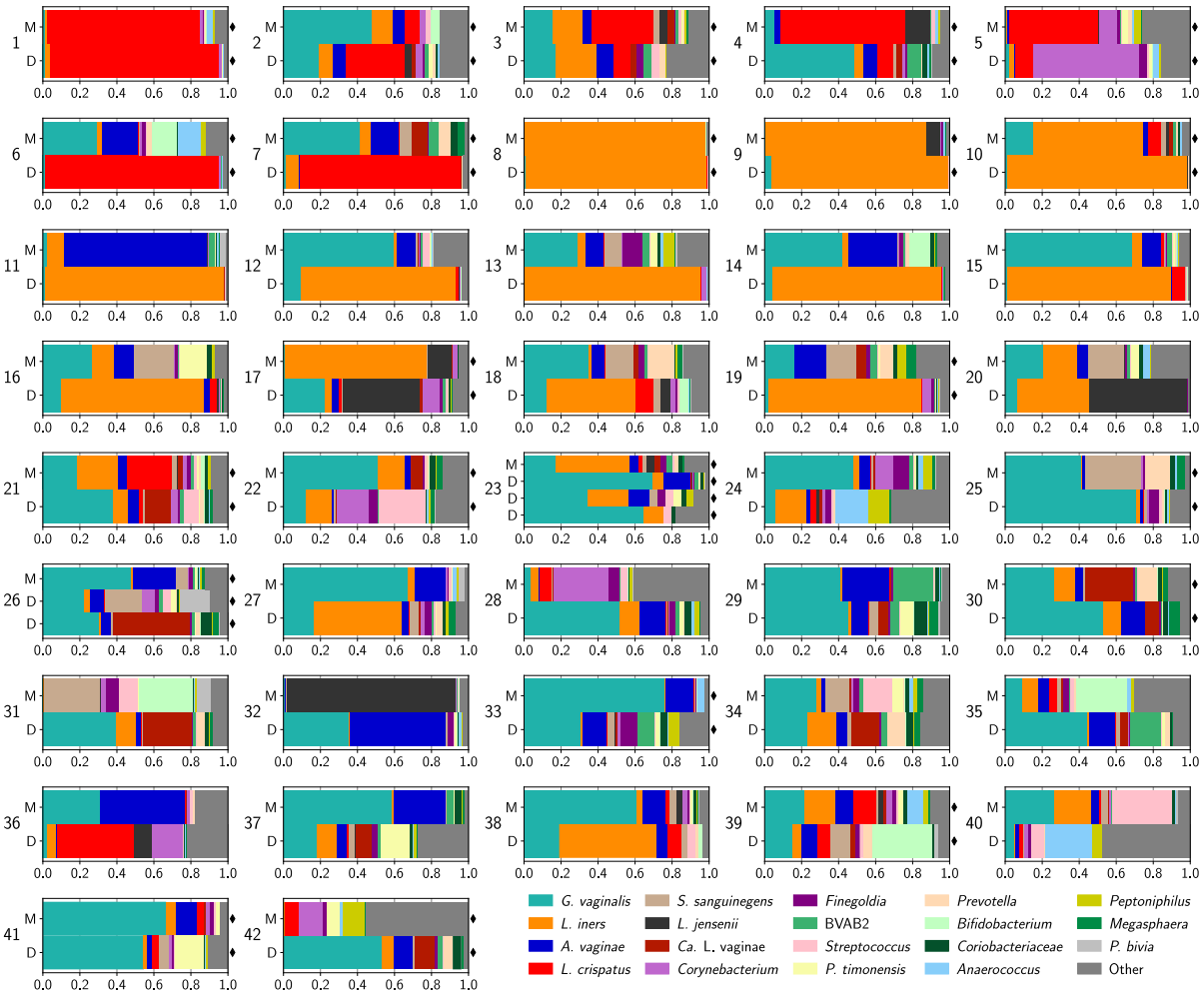


Figure 1: Taxonomic composition of the vaginal microbiota of reproductive age mothers and daughters. Stacked bars represent the relative abundances of individual bacterial phylotypes. Each plot displays the profiles for members belonging to the same family (M-Mother, D-Daughter). Two families (23,27) had 2 and 3 daughters, respectively. Samples denoted with black diamonds were selected for shotgun metagenomic analysis.

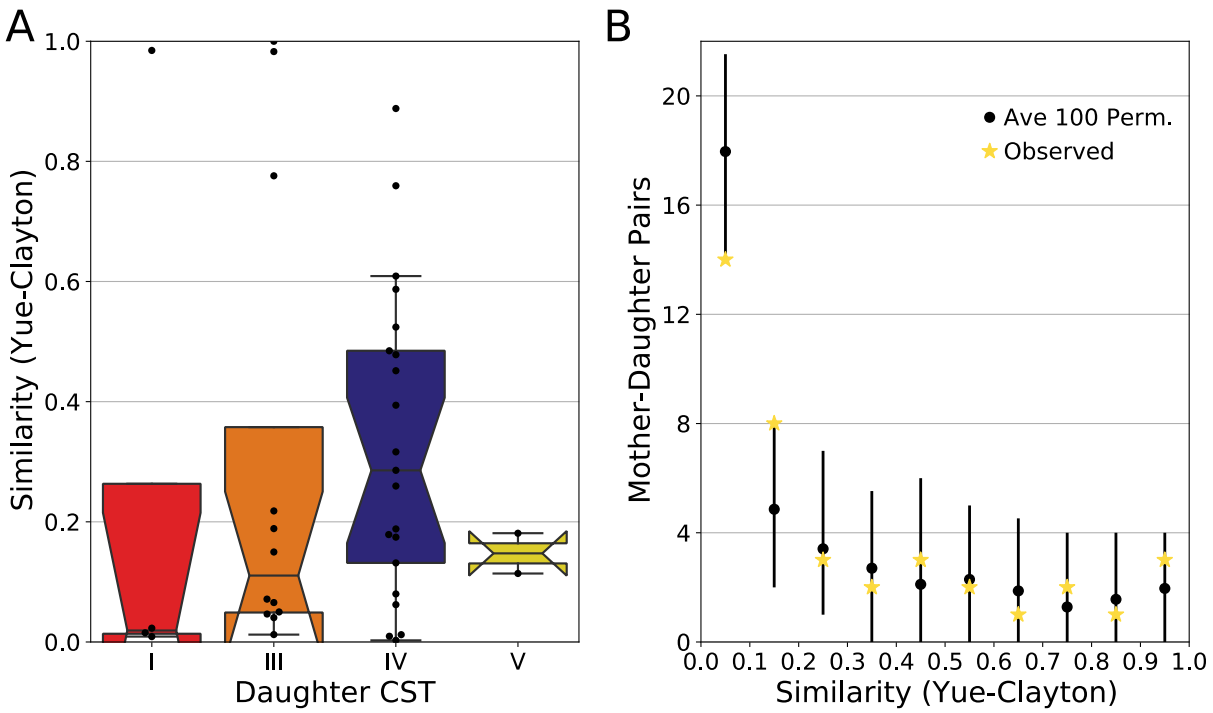
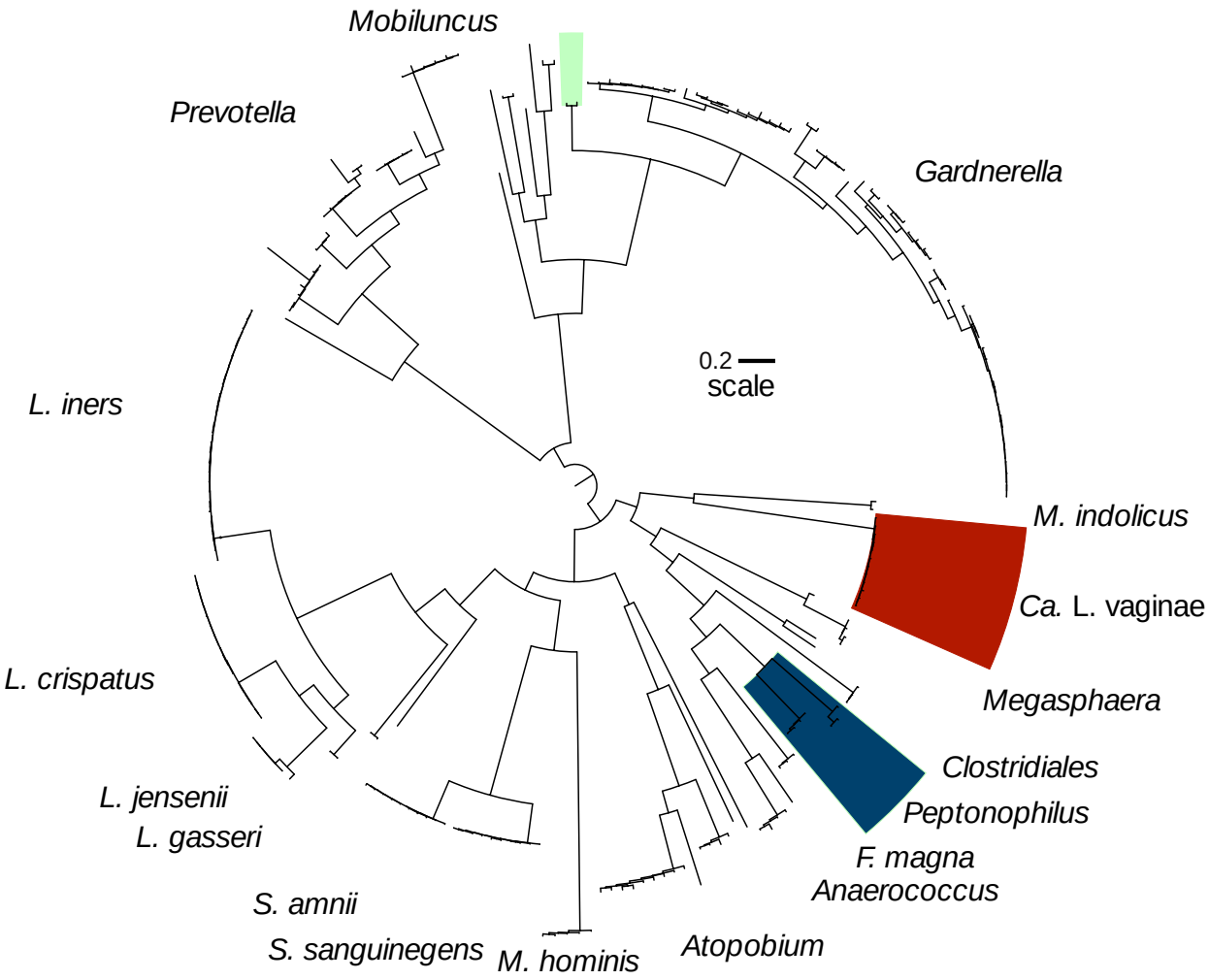


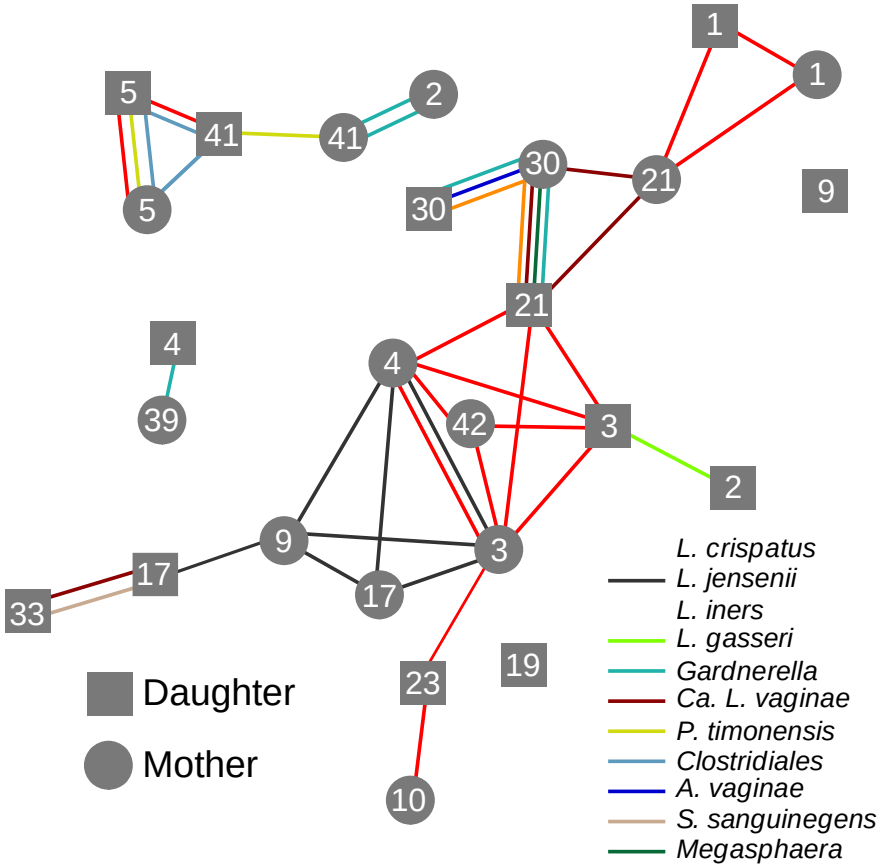
Figure 2: Similarity in the taxonomic composition between mother-daughter pairs, delineated by the daughter's CST assignment (A). Higher values of the Yue-Clayton index signify communities that bear greater taxonomic compositional similarity. Permutation tests were used to establish whether the observed similarities between mothers and their daughters were different than that expected by chance alone (B). Black points represent the average number of permuted mother-daughter pairs whose similarity fell within 0.1 increments of the Yue-Clayton similarity index, while yellow stars represent the observed number of pairs. Error bars span the range between the 2.5% and 97.5% quantiles of the 100 random permutations.



626

627 **Figure 3:** Phylogenetic tree displaying the taxonomic diversity of the 225 metagenome assembled
628 genomes (MAGs) derived from the shotgun metagenomic data generated for 22 families. The
629 maximum likelihood phylogeny was established using a concatenated alignment of the amino acid
630 sequences for 13 orthologous genes which were found to be present in at least 99% of the MAGs:
631 *gapD*, *gltX, ileS*, *pheS*, *pheT*, *cysS*, *hisS, uvrD*, *ruvX*, *rpsO*, *Ffh*, *obgE*, and *lepA*.

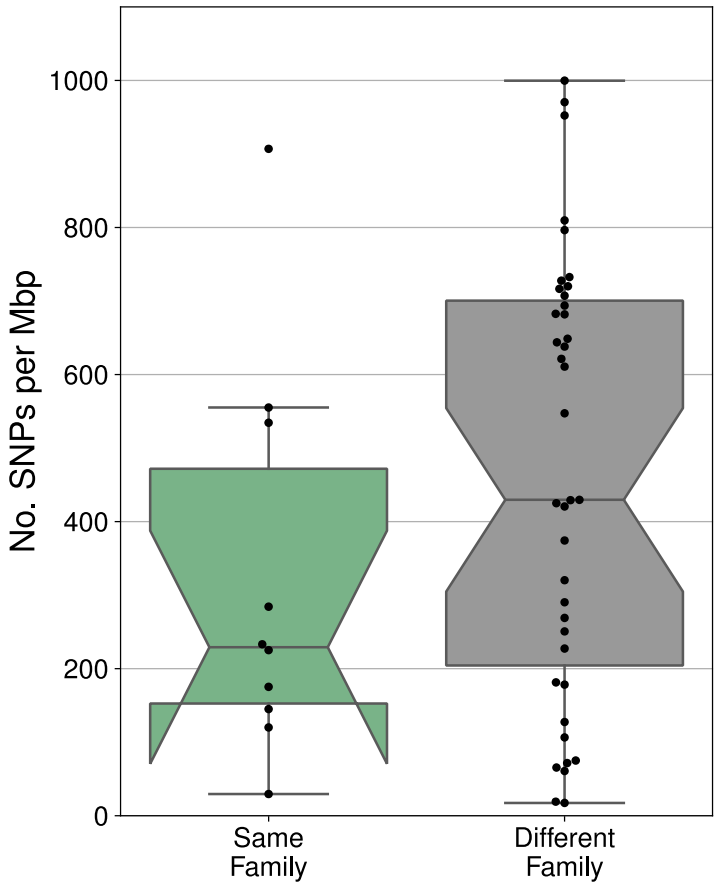
632



633

634 **Figure 4**: Network diagram of shared bacterial strains identified in this cohort. A stringent threshold,
635 99.9% sequence identity, 70% coverage, was used to identify shared strains. Lines represent the
636 shared bacterial strains and connect the participants in which the strain was found. Mothers are
637 represented by circles and daughters by squares. Numbers on the nodes signify the family and can
638 be linked back to the taxonomic profile of the participant using Figure 1.

639



640

641 **Figure 5:** Number of single nucleotide polymorphisms identified in strains found to be shared

642 between members the same family or between members of different families. Values are scaled per

643 mega base pair of compared sequence.

Table 1: MAGs identified in mother-daughter pairs

Taxonomy	Family	Daughter's age	Rel. Abund Daughter†	Mother's age	Rel. Abund. Mother†	Substitution Rate*
"*Ca* . L. vaginae"	21	18	9.21%	41	2.80%	1.48E-05
A. vaginae	30	14	2.22%	42	0.46%	1.05E-06
Clostridiales Family	5	14	2.15%	36	2.31%	5.18E-06
Gardnerella	30	14	47.42%	42	16.56%	1.98E-05
L. crispatus	3	15	44.54%	32	72.13%	4.00E-06
L. crispatus	1	15	80.91%		75.57%	9.48E-06
L. crispatus	5	14	31.88%	36	39.26%	6.26E-06
L. iners	30	14	12.34%	42	5.36%	8.32E-06
P. timonensis	41	15	16.16%	47	1.63%	7.50E-06
P. timonensis	5	14	2.18%	36	2.83%	3.24E-05
**per base pair per year*						
†Relative abundance in community						

644