



## Comparative metagenomic analysis of the vaginal microbiome in healthy women

Fei Liu<sup>a,1</sup>, Yingshun Zhou<sup>b,1</sup>, Lisi Zhu<sup>c,1</sup>, Zhiyi Wang<sup>d</sup>, Liangkun Ma<sup>e,\*</sup>, Yuan He<sup>f,\*\*</sup>, Pengcheng Fu<sup>g,\*\*\*</sup>

<sup>a</sup> CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

<sup>b</sup> Department of Pathogenic Biology, Pathogenic Biology platform, School of Basic Medical Sciences, Southwest Medical University, Luzhou, 646000, Sichuan Province, China

<sup>c</sup> Laboratory of Clinical Genetics, Medical Research Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China

<sup>d</sup> CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

<sup>e</sup> Peking Union Medical College Hospital (Dongdan Campus), No. 1 Shuai Fu Yuan, Beijing, 100730, China

<sup>f</sup> National Human Genetic Resources Center, National Research Institute for Health and Family Planning, Beijing, China

<sup>g</sup> Mybiome (Beijing) Medical Consulting Co. LTD, China

### ARTICLE INFO

#### Keywords:

Vaginal microbiome  
Metagenome  
Metabolic pathways  
Genome  
Strains

### ABSTRACT

The composition of these vaginal microbiome has a significant impact on women's health. However, few studies have characterized the vaginal microbiome of healthy Chinese women using metagenomic sequencing. Here, we carried out a comparative metagenomic analysis to survey taxonomic, functional levels, and microbial communities' genome content in healthy women's vaginal microbiome. Overall, we observed a total of 111 species, including all dominant vaginal *Lactobacillus* species, such as *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*. Unlike microbial taxa, several pathways were ubiquitous and prevalent across individuals, including adenosine ribonucleotides de novo biosynthesis and pyruvate fermentation to acetate and lactate II. Notably, our diversity analysis confirmed a significant difference in healthy women from different ethnic groups. Moreover, we binned vaginal assemblies into 62 high-quality genomes, including 9 *L. iners*, 7 *A. vaginae*, 5 *L. jensenii*, and 5 *L. crispatus*. We identified the pan and core genomes of *L. iners* and *A. vaginae* and revealed the genetic diversity. Primary differences between strains were the hypothetical genes and mobile element-like genes. Our results provide a framework for understanding the implications of the female reproductive tract's composition and functional potential and highlight the importance of genome-resolved metagenomic analysis to further understand the human vaginal microbiome.

### 1. Introduction

The female reproductive tract's microbiome has implications for women's reproductive health [1,2]. Recent reports have suggested that the vaginal microbiome composition has a significant

population-specific impact on preterm birth risk [3–8]. A vaginal microbiome dominated by *Lactobacillus* species is more related to the health of the female reproductive tract. They can keep the stability of the vaginal environment and maintain low pH by producing lactic acid. In contrast, a microbiome dominated by complex microbial communities

**Abbreviations:** CSTs, community state types; TNF, tetranucleotide frequency; PERMANOVA, Permutational multivariate analysis of variance; NMDS, nonmetric multidimensional scaling; PCoA, Principal Coordinate Analysis.

Peer review under responsibility of KeAi Communications Co., Ltd.

\* Corresponding author.

\*\* Corresponding author.

\*\*\* Corresponding author.

**E-mail addresses:** [liuf@im.ac.cn](mailto:liuf@im.ac.cn) (F. Liu), [yingshunzhou@swmu.edu.cn](mailto:yingshunzhou@swmu.edu.cn) (Y. Zhou), [julith-1@163.com](mailto:julith-1@163.com) (L. Zhu), [wangzhiyi@im.ac.cn](mailto:wangzhiyi@im.ac.cn) (Z. Wang), [maliangkun@pumch.cn](mailto:maliangkun@pumch.cn) (L. Ma), [heyuan8506@126.com](mailto:heyuan8506@126.com) (Y. He), [phillis1990@163.com](mailto:phillis1990@163.com) (P. Fu).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.synbio.2021.04.002>

Received 28 September 2020; Received in revised form 3 April 2021; Accepted 6 April 2021

2405-805X/© 2021 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of *Gardnerella*, *Atopobium*, *Dialister*, and other anaerobes has been associated with a higher risk for acquisition of sexually transmitted infections and pelvic inflammatory disease [9–11]. The vaginal bacterial community composition has been classified into community state types (CSTs) based on the presence and relative abundance of species characterized by the 16S rRNA gene sequencing [10,12].

Studies of the vaginal-associated microbiome have mainly relied on 16S rRNA gene amplicon sequencing [12,15], which has low taxonomic resolution and cannot perform species-specific functional analysis. In addition, bacterial species strains often exhibit substantial diversity in gene content [13,14]. A substantial portion of reproductive-age women has a vaginal microbiota dominated by *L. iners* [16]. However, its relationship to vaginal health is complicated [17]. High-quality genomes of *L. iners* are still limited. The persistence of an adherent bacterial biofilm, including *A. vaginae*, seems to be the primary reason for the failure of bacterial vaginosis (BV) treatment [18]. *A. vaginae* is often highly resistant to metronidazole [19]. Whole metagenomic shotgun sequencing can address these limitations [20,21]. However, only a few studies have applied it to the vaginal microbiome [22]. Little reviews have characterized the vaginal microbiome of healthy Chinese women using metagenomic sequencing.

Here, we performed a comparative metagenomic analysis to survey taxonomy, functional potential, and microbial communities' genome content in healthy women's vaginal microbiome using metagenome sequencing.

## 2. Materials and methods

### 2.1. Dataset collection

Metagenomes used in this study included 82 newly in-house sequenced datasets and 133 randomly selected human vaginal metagenome datasets in the HMP (HMP1-II) cohort [22] downloaded from the National Center for Biotechnology Information Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra>). The SRA datasets were converted to fastq using the fastq-dump module in the NCBI SRA Toolkit. The sample ID, sequencing platform, reads length, reads number, data size, and accession numbers for each dataset were shown in Supporting Information Table S1. A total of 210 women were selected in the Peking Union Medical College Hospital from July 2016 to October 2017, and the study was performed with the Ethical Committee's approval. The following subjects were excluded from the study: those who used antibiotics within the previous 60 days; those who had any of the symptoms or similar to pruritus, leukorrhea abnormalities, urinary tract infection, and diarrhea within four weeks; the presence of a moderate or severe illness at the time of enrollment; pregnant or lactating; vaginal discharge with blood. After selection, only 82 of these were enrolled. For each participant, a cotton swab was used to collect vaginal discharge from the posterior fornix. All in-house metagenomic data were derived from vaginal swab samples and after obtaining informed consent from all the participants. Samples were stored at  $-20^{\circ}\text{C}$  for up to 2 weeks before being transferred to  $-80^{\circ}\text{C}$ . All women were not pregnant, of reproductive age, ranging from 18 to 45 years.

### 2.2. DNA isolation, library construction, and whole metagenomic sequencing

The samples underwent DNA extraction using standard methods. Briefly, cells were lysed with the addition of an enzymatic cocktail: 50  $\mu\text{l}$  of lysozyme (10 mg/ml), 6  $\mu\text{l}$  of mutanolysin (25,000 U/ml Sigma-Aldrich), 3  $\mu\text{l}$  of lysostaphin (4000 U/ml in sodium acetate; Sigma-Aldrich), and 41  $\mu\text{l}$  of TE50 buffer (10 mM Tris-HCL and 50 mM EDTA, pH 8.0). The mixture was incubated for 1h at  $37^{\circ}\text{C}$  and then disrupted by bead beating for 2 min at room temperature in a Mini-Beadbeater-96. DNA was isolated using the QIAamp DNA Mini Kit (Qiagen). Samples were eluted with  $2 \times 200 \mu\text{l}$  of TE buffer (10 mM Tris-

Cl, 0.5 mM EDTA; pH 9.0). A NanoDrop was used to measure DNA concentration and purity. DNA was sheared to an average length of 300–500 bp. DNA libraries were constructed using NEB Next Ultra DNA Library Prep Kit for Illumina following the manufacturer's recommendations. The library quality was assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). Barcoded TruSeq libraries were sequenced with an Illumina HiSeq 2500 platform.

### 2.3. Metagenomic analysis

Sequencing data were processed using KneadData v0.7.2 (<https://bitbucket.org/biobakery/kneaddata/wiki/Home>) for quality trimming and human DNA depletion with default settings. Briefly, reads were mapped to the GRCh38 reference library with standard masking using Bowtie2, and human reads were filtered out. Then, Trimmomatic was used to filter low-quality reads. MetaPhlan2 [23] was used to perform taxonomic classification and profiling by mapping the filtered reads against a library of clade-specific markers. Functional profiling was performed using HUMAnN2 in UniRef90 mode [24]. The filtered reads were mapped to the pangenomes species identified during the taxonomic profiling by MetaPhlan2 using Bowtie2. We defined a 'core' pathway as one that was detected with relative abundance  $>10^{-4}$  in at least 75% of samples. The pangenomes have been pre-annotated to their respective UniRef90 families. The human vaginal non-redundant gene catalog (VIRGO) [25], an integrated and comprehensive resource to establish taxonomic and functional profiling of vaginal microbiomes, was also used to characterize the functional composition of our sequenced vaginal metagenomic datasets. LEfSe [26] analysis was used to identify discriminative taxa between different subjects. Statistically significant groups were reported with linear discriminant analysis (LDA) scores  $>3$ . Illumina sequencing reads were assembled using IDBA-UD [27], a de novo assembler for metagenomic sequencing data with highly uneven depth. Genome bins were performed based on the calculated sequence depth and tetranucleotide frequency (TNF) of each scaffold using MetaBAT2 [28]. Completeness and contamination of all bins created were assessed using CheckM version v1.0.18 [29]. To assign a high-quality bin, we used a threshold of marker gene completeness  $>90\%$  and contamination  $<5\%$ . GTDB-Tk [30] was used to assign taxonomy to assembled bins.

Nonmetric multidimensional scaling (NMDS) and Principal coordinate analysis (PCoA) were performed to evaluate differences in taxonomy profiles among samples based on the Bray–Curtis distance of species relative abundance. Permutational multivariate analysis of variance (PERMANOVA) between different groups was done with adonis in vegan with a similarity index using 9999 permutations. Shannon diversity, redundancy analysis, and heatmaps were prepared in R with vegan and ggplot2 packages.

### 2.4. *L. iners* and *A. vaginae* phylogenetic trees and genome comparisons

Thirty-one publicly available *L. iners* genomes plus the nine *L. iners* genomes recovered in this study were analyzed. *L. iners* genomes used in this study are: GCF\_000149065.1, GCF\_000149085.1, GCF\_000149105.1, GCF\_000149125.1, GCF\_000149145.1, GCF\_000160875.1, GCF\_000177755.1, GCF\_000179935.1, GCF\_000179955.1, GCF\_000179975.1, GCF\_000179995.1, GCF\_000185405.1, GCF\_000191685.1, GCF\_000191705.1, GCF\_000204435.1, GCF\_000227195.1, GCF\_001435015.1, GCF\_002871595.1, GCF\_002884695.1, GCF\_002884705.1, GCF\_002892385.1, GCF\_009556455.1, GCF\_009857205.1, GCF\_010748955.1, GCF\_011058695.1, GCF\_011058715.1, GCF\_011058735.1, GCF\_011058755.1, GCF\_011058775.1, GCF\_011058795.1, and GCF\_902374445.1. Similarly, seven *A. vaginae* genomes recovered in this study and six publicly available *A. vaginae* genomes (*A. vaginae* PB189-T1-4 (GCF\_000179715) with an abnormal high GC% was removed in the study) were used. The NCBI assembly

numbers for the publicly available *A. vaginae* genomes used in this study are: GCF\_000159235.2, GCF\_000178335.1, GCF\_001049775.1, GCF\_001562845.1, and GCF\_900445305.1. kSNP3.0 [31] was used to build phylogenetic trees from whole-genome SNP patterns. The core and pan-genome were estimated using Roary [32] at standard settings with GFF3 files generated by PROKKA [33] used as input files. The Comprehensive Antibiotic Resistance Database (CARD) was used to predict antibiotic-resistant genes [34].

### 3. Result

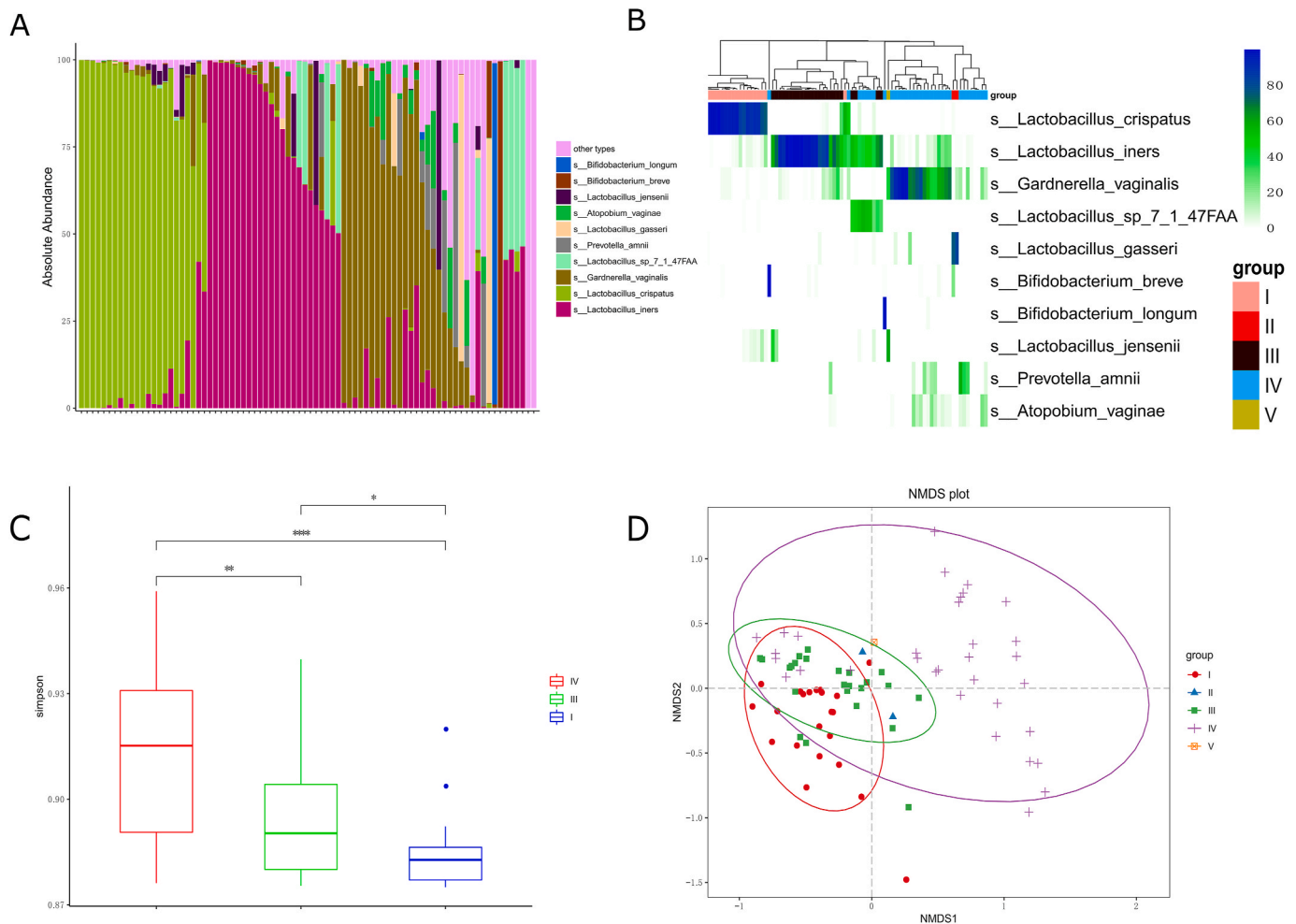
We recruited 210 women in the Peking Union Medical College Hospital from July 2016 to October 2017. After selection, only 82 of these were enrolled. We performed shotgun metagenomic sequencing on 82 vaginal samples (Supplementary Table S2). These metagenomes have about 5.8 billion reads, 92% of which were identified as human sequences and removed for the following analysis. We found that vaginal metagenomes dominated by *Lactobacillus* species had significantly higher proportions of human sequence reads than those from *Lactobacillus* deficient datasets (94.2% vs. 89.3%;  $P < 0.05$ ), which was consistent with the previous report [25]. There are some reasons for this,

such as low-diversity bacterial communities, and the problem in isolation of pure DNA from Gram-positive bacteria arises from the difficulty in lysis of bacterial cells and associated metabolites.

#### 3.1. Characterization of vaginal community composition and structure

Unlike gut or oral communities, vaginal communities were generally dominated by one species, based on whole-genome abundance measures (Fig. 1A). Overall, we observed a total of 111 species in the vaginal microbiota of these women. We identified all dominant vaginal *Lactobacillus* species, including *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*, as well as common facultative and strict anaerobic vaginal species such as *Gardnerella vaginalis*, *Atopobium vaginae*, and *Prevotella amnii*.

Fig. 1B shows the taxonomic profiles encompassing the five previously reported vaginal community state types (CSTs), including CST I (*L. crispatus*-dominated), II (*L. gasseri*-dominated), III (*L. iners*-dominated), IV (array of strict and facultative anaerobes), and V (*L. jensenii*-dominated) with frequencies in this set of metagenomes of 23.9%, 2.2%, 26.1%, 35.9%, and 1.1%, respectively (Supplementary Table S2). CST I is always associated with a healthy vagina, while CST III is more prone to vaginal dysbiosis [35,36]. Studies have shown that the protective effect



**Fig. 1.** Community structure of the vaginal microbiome in different subjects. (A) Relative abundance of the top 10 most abundant vaginal taxa are displayed. (B) Heatmap of relative abundance of microbial taxa based on the ten most abundant taxa. (C) Shannon diversity indices calculated for the vaginal communities in this study (groups II and V with few samples were excluded). (D) NMDS plot showing the similarities among all the samples according to the community structure. The confidence ellipses are drawn with a confidence level of 0.90. The samples with different CSTs are plotted with different colors. Unlike gut or oral communities, vaginal communities were generally dominated by one species, based on whole-genome abundance measures. Overall, a total of 111 species were observed in the vaginal microbiota of these women. All dominant vaginal *Lactobacillus* species, including *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*, as well as common facultative and strict anaerobic vaginal species such as *Gardnerella vaginalis*, *Atopobium vaginae*, and *Prevotella amnii* were identified in this study. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

of *L. crispatus* against BV is intrinsically associated with the ability to produce lactic acid and bacteriocin that maintain the healthy state of the vagina [37].

The Shannon diversity indices and beta-diversity analysis of the communities indicated that the most diverse communities were CST IV. Fig. 1C shows communities in CST I have the lowest Shannon diversity values, whereas CST IV communities have the highest Shannon diversity values. Fig. 1D shows the relationships among communities by nonmetric multidimensional scaling (NMDS). The samples in CST I and CST III grouped tightly, whereas the CST IV samples spanned the entire space and showed the highest variation compared with the other samples. Although CST IV communities were not dominated by *Lactobacillus* sp., *L. iners* and *L. crispatus* were detected in 87.8% and 48.5% of CST IV communities, respectively (Supplementary Table S2).

### 3.2. Functional annotations of the human vagina

According to the previous studies, our results suggest that the human vagina has no single core microbiome [12,22]. Several core microbiomes can be defined by the community groups depicted. Unlike microbial taxa, several pathways were ubiquitous and prevalent across individuals (Fig. 2A), such as adenosine ribonucleotides de novo biosynthesis and pyruvate fermentation to acetate and lactate II. The most abundant of these core pathways include adenosine ribonucleotides de novo biosynthesis, which was contributed principally by *L. iners* in CST III (Fig. 2B). In contrast, adenosine ribonucleotides de novo biosynthesis was contributed principally by unclassified species in CST I. As the database for different types of *Lactobacilli* is not comprehensive, these unclassified species could be *L. crispatus*. We found that some samples were blank even with the same bacteria species (Fig. 2C). For example, in the *L. iners* dominated samples, some do not have the listed pathway detected, indicating that the *L. iners* strains may be unique.

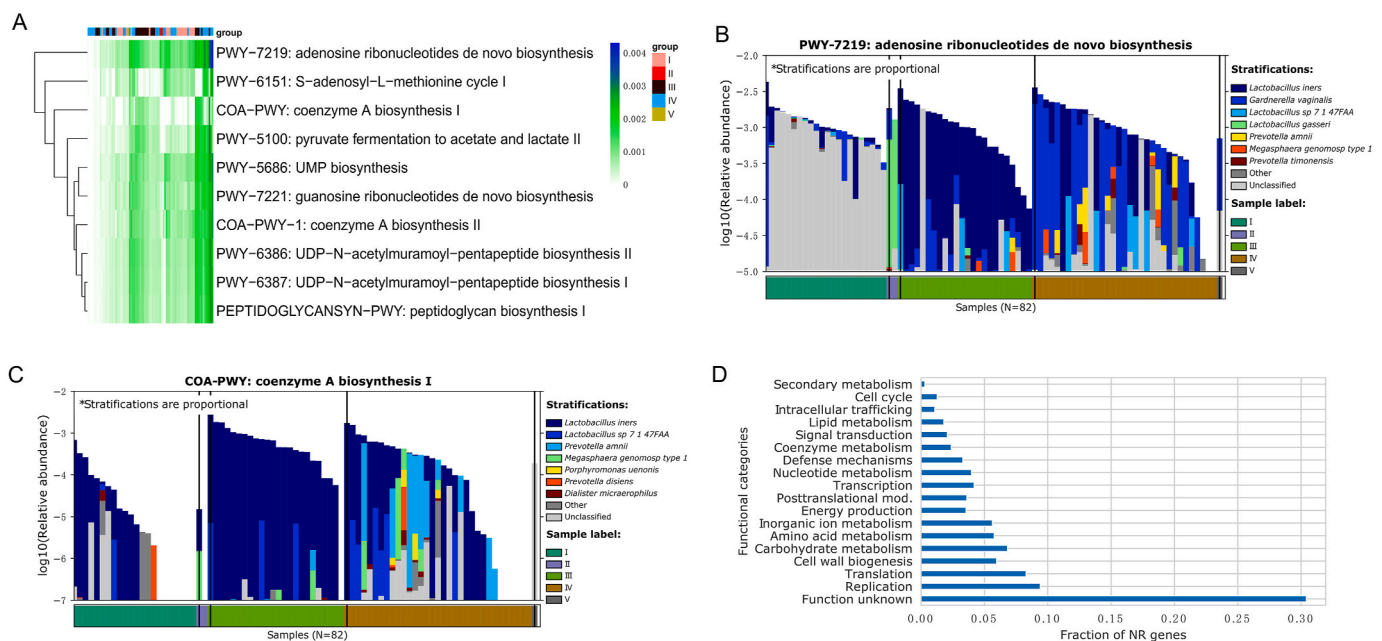
We also used the human vaginal non-redundant gene catalog (VIRGO) to characterize the functional composition of our sequenced vaginal metagenomic datasets. The number of reads per non-redundant

gene per sample with functional annotations, including KEGG, eggNOG, and EC number, can be found in Supplementary Table S3. Overall, 199,706 genes were assigned a functional annotation from at least one source. Specifically, a total of 73,842 unigenes were assigned to 198 KEGG pathways, and 198,756 unigenes were assigned to the eggNOG functions in this study. Fig. 3D shows an overview of the eggNOG functions encoded in VIRGO. Carbohydrate metabolism, amino acid metabolism, and inorganic ion metabolism were the top three in the eggNOG category “metabolism”.

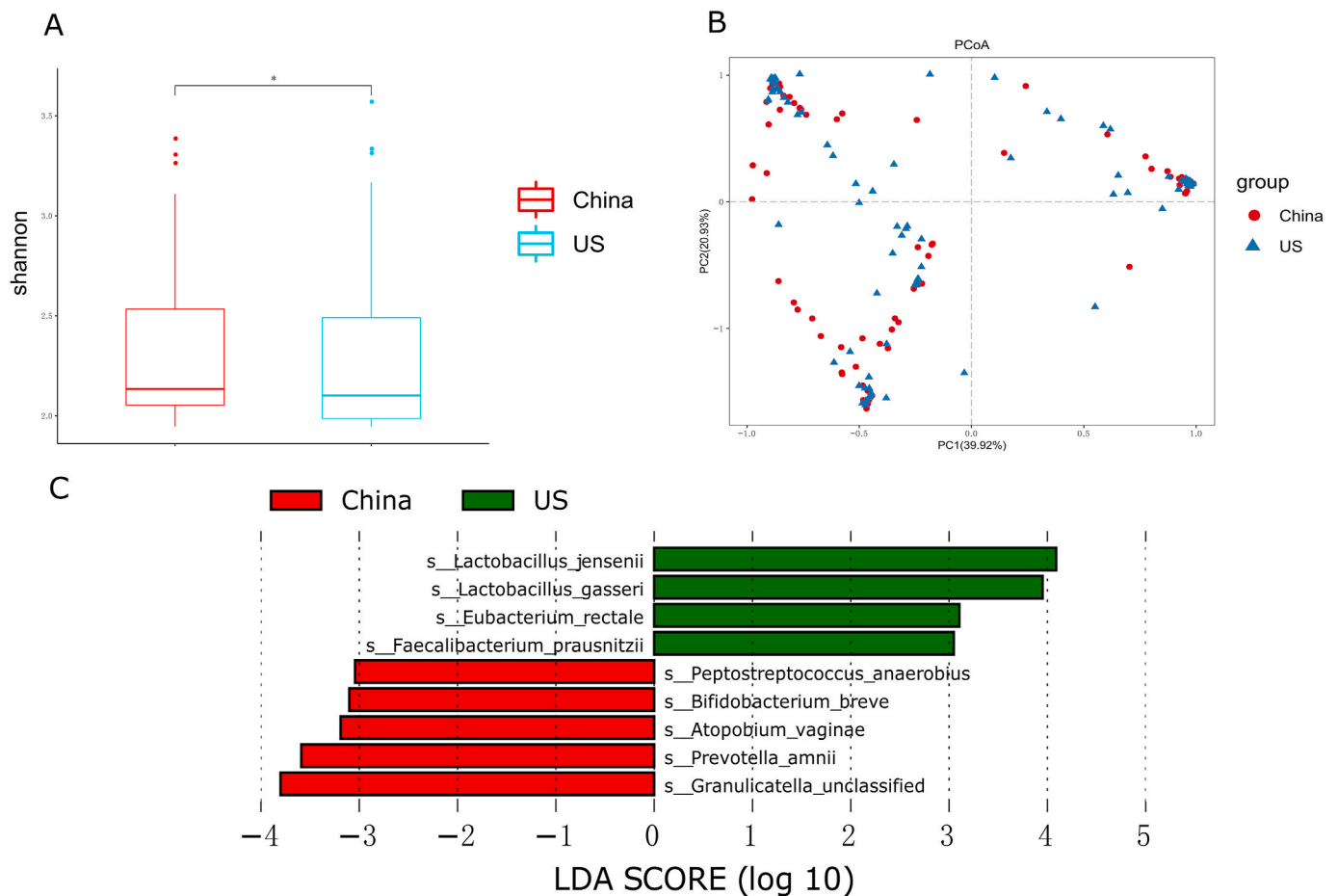
### 3.3. A comparison of the vaginal microbiome in healthy women from different ethnic groups

We downloaded 133 metagenome datasets of the human vagina in the HMP cohort [22] from the National Center for Biotechnology Information (NCBI) Sequence Read Archive. We performed a comparison of the vaginal microbiome in 82 newly in-house sequenced Chinese datasets and 133 HMP American datasets. Although the types of vaginal microbial communities found in healthy Chinese women were similar to American women, alpha and beta diversity confirmed a significant difference between these two groups. As shown in Fig. 3A, Chinese communities have higher Shannon diversity values. Similarly, we employed Principal coordinate analysis (PCoA) to assess the variation of vaginal microbial communities between Chinese and American women and found significant differences in diversities based on the Bray–Curtis distance (Fig. 3B; adonis,  $P < 0.001$ ;  $R^2 = 0.033$ ). Each point represents the vaginal community of an individual in Fig. 3B. Some subjects were very similar and clustered together in the figure, while others had little similarity to any other subjects.

LEfSe analysis revealed that many differences in the abundance of communities existed between these two groups (Fig. 3C). The LDA scores indicated that the relative abundances of *Granulicatella unclassified*, *Prevotella amnii*, and *Atopobium vaginae* were much more enriched in Chinese than those in American (Fig. 3C). The most differentially abundant species in American was *Lactobacillus jensenii*. To quantify the



**Fig. 2.** Functional analysis of the vaginal microbiome in different subjects. (A) Heatmap of relative abundance of the 10 most abundant metabolic pathways in the vaginal metagenomes used in this study. The legend indicates microbiome samples by the five CSTs. Functional profiling was performed using HUMAnN2 in UniRef90 mode. (B–C) Contributive diversity of core vaginal microbiome pathways. The top of each set of stacked bars indicates the total stratified abundance of the pathway within a single sample (log-scaled). Species and ‘unclassified’ stratifications are proportionally scaled within the total bar height. Community groups can define several core microbiomes. Unlike microbial taxa, several pathways were ubiquitous and prevalent across individuals, such as adenosine ribonucleotides de novo biosynthesis and pyruvate fermentation to acetate and lactate II. (D) Functional distribution of non-redundant genes in VIRGO using EggNOG (v4.5).



**Fig. 3.** Comparing the vaginal microbiome diversity in healthy Chinese and American women. (A) Shannon diversity indices calculated for the vaginal communities. (B) Two-dimensional scatter plot constructed by PCoA analysis among vaginal microbial communities. (C) Histogram of the linear discriminant analysis (LDA) scores (log10) computed for features with differential abundance in the healthy Chinese and American subjects. Statistically significant groups were reported with LDA scores >3. The samples from Chinese and American women are designated with different colors. Chinese communities have higher Shannon diversity values. Although the types of vaginal microbial communities found in healthy Chinese women were similar to American women, adonis analysis confirmed a significant difference between these two groups. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

relative influence of dominant taxon (*L. crispatus* and *L. iners*) on community composition, we performed redundancy analysis (RDA) on microbial communities' relative abundances. Our results indicated that *L. crispatus* and *L. iners* were highly significant sources of variation in community abundances ( $P < 0.001$ ;  $R^2 > 0.98$ ).

### 3.4. Assembly of 62 draft microbial genomes from the human vagina

We carried out a metagenomic assembly of individual samples from the above analyzed 215 vaginal datasets and created 834 metagenomic bins from the single-sample binning. After dereplication to 95% ANI, 62 high-quality genomes (>90% completeness and <5% contamination) remained, including 9 *L. iners*, 7 *A. vaginae*, 5 *L. jensenii*, and 5 *L. crispatus*.

We compared the high-quality genomes of 9 strains of *L. iners* from the vaginal samples with 31 publicly available *L. iners* genomes (Fig. 4A). *L. iners* strains recovered in this study were distributed evenly across the phylogenetic tree. We annotated an average of 1226 genes per strain. This set of 40 *L. iners* genomes comprised 3495 different gene orthologous groups (Supplementary Table S4). The core genome consisted of 735 genes (which correspond to ~60% of a given genome), which displayed high gene conservation and low genetic diversity in *L. iners* strains.

We compared *A. vaginae* strains recovered in this study with other published *A. vaginae* strains to determine genetic heterogeneity.

Phylogenetic analysis revealed that *A. vaginae* strains recovered here were divergent from previously sequenced strains (Fig. 4 4B). Pan-genome analysis in *A. vaginae* species contained 2218 protein-coding sequences with 28.2% (625) core genes and 71.8% (1593) shell genes (Supplementary Table S5).

Fig. 5 shows whole-genome alignments from the *L. iners* clinical isolates described in this study. Primary differences between strains were the hypothetical genes and mobile element-like genes, including genes involved in CRISPR-associated, restriction-modification systems, virulence-associated, and antibiotic resistance.

Among seven strains of *A. vaginae* from this study, we detected the tetracycline-resistant ribosomal protection protein *tetM* in one of these strains. Moreover, among nine strains of *L. iners*, Erm 23S ribosomal RNA methyltransferase *ErmB* was found.

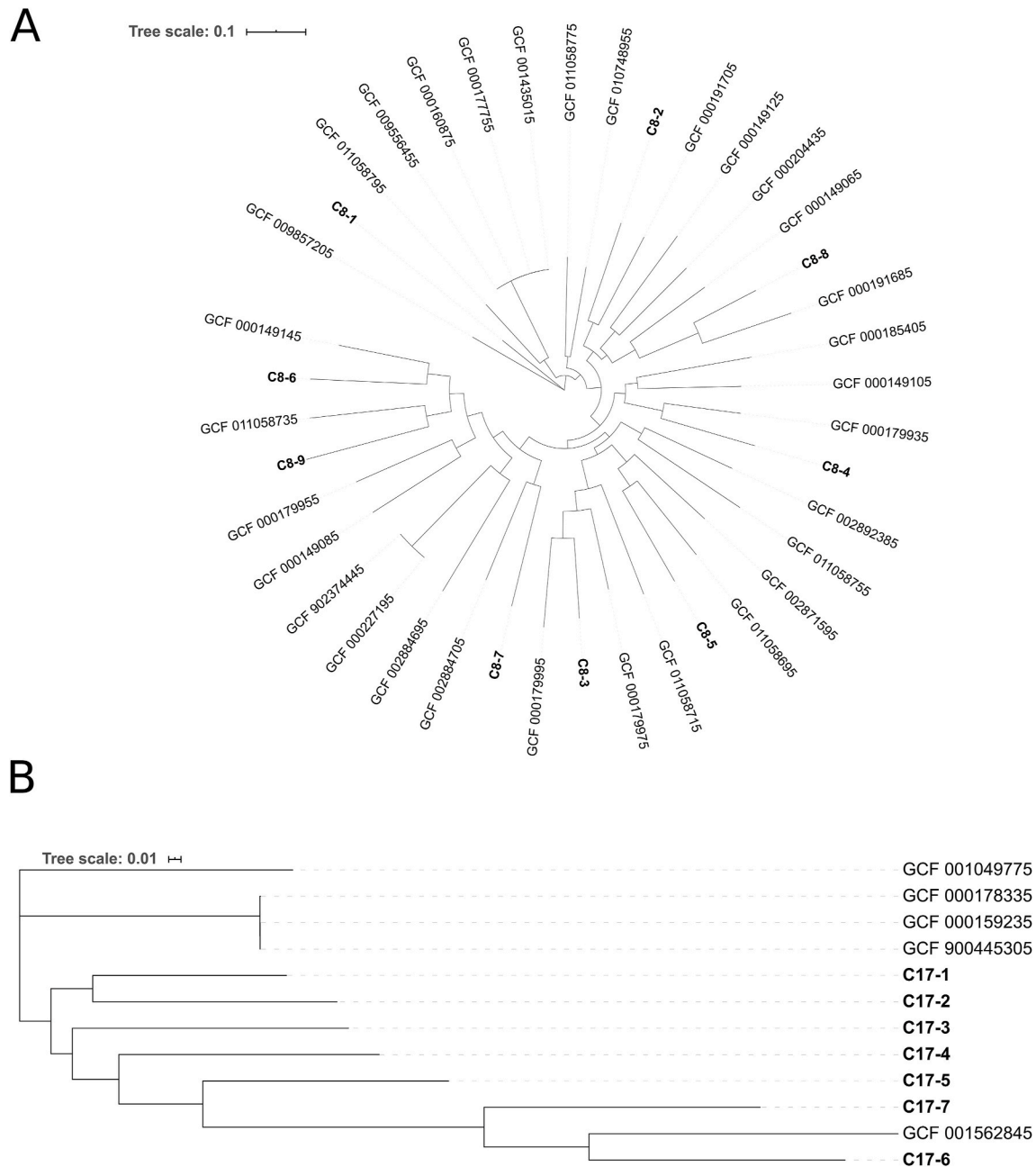
## 4. Discussion

This study carried out a comparative metagenomic analysis to survey taxonomic, functional levels, and microbial communities' genome content in healthy women's vaginal microbiome. Studies of the vaginal-associated microbiome have mainly relied on 16S rRNA gene amplicon sequencing [12,15], which has limitations in taxonomic resolution and the lack of ability to perform species-specific functional analysis. Strains of a bacterial species often exhibit substantial diversity in gene content. The five previously reported vaginal CSTs identified in Chinese women,

and the most diverse communities were those of group IV, consistent with the previous study [12]. Compared to the recently reported vaginal microbiome of Chinese pregnant women [38], this study found more *L. iners*-dominated and less *L. crispatus*-dominated women. Notably, the diversity of community types in Chinese was higher than in American (Fig. 3). Consistent with previous reports [34,39], our findings suggest some differences in different racial groups in terms of the bacterial species found in the vagina. However, the reasons for such differences are still unclear but could be based on differences in innate and adaptive immune systems, contraceptive usage, sex activities, topical antibiotics

exposure, and other genetically determined differences between hosts [17]. Besides, the sample collection method, DNA extraction, and sequencing could all bring biases to comparing microbiome composition across datasets.

Compared to other 16S rRNA gene sequencing studies, this study can perform gene-level analysis. We identified conserved core functions among individuals despite differences in their species composition (Fig. 2). Interestingly, all communities contained members assigned to genera known to produce lactic acid, including *Lactobacillus*, *Megasphaera*, and *Atopobium*. An important catabolic function, namely the



**Fig. 4.** Phylogenetic tree reconstructed from core SNPs (SNPs present in all genomes analyzed) with kSNP3. (A) Phylogenetic tree of 40 *L. iners* genomes, including 31 available in GenBank. The NCBI assembly numbers for the publicly available *L. iners* genomes used in this study are: GCF\_000149065.1, GCF\_000149085.1, GCF\_000149105.1, GCF\_000149125.1, GCF\_000149145.1, GCF\_000160875.1, GCF\_000177755.1, GCF\_000179935.1, GCF\_000179955.1, GCF\_000179975.1, GCF\_000179995.1, GCF\_000185405.1, GCF\_000191685.1, GCF\_000191705.1, GCF\_000204435.1, GCF\_000227195.1, GCF\_001435015.1, GCF\_002871595.1, GCF\_002884695.1, GCF\_002884705.1, GCF\_002892385.1, GCF\_009556455.1, GCF\_009857205.1, GCF\_010748955.1, GCF\_011058695.1, GCF\_011058715.1, GCF\_011058735.1, GCF\_011058755.1, GCF\_011058775.1, GCF\_011058795.1, and GCF\_902374445.1. (B) Phylogenetic tree of 12 *A. vaginae* genomes, including five available in GenBank. The NCBI assembly numbers for the publicly available *A. vaginae* genomes used in this study are: GCF\_000159235.2, GCF\_000178335.1, GCF\_001049775.1, GCF\_001562845.1, and GCF\_900445305.1. The genomes recovered in the current study are shown in bold.

production of lactic acids, such as pyruvate fermentation to acetate and lactate II, was conserved among communities despite differences in the species composition. In addition, vaginal communities had many genes for nutrient (carbohydrates and amino acids) utilization, consistent with the previous report [25].

*Lactobacillus iners* dominated vaginal bacterial communities of many women [10,12]. *Atopobium vaginae* is associated with bacterial vaginosis in around 80% of the cases and might involve therapeutic failures [40]. Among *L. iners* and *A. vaginae* genomes studied here, *A. vaginae* exhibited a greater diversity, which is consistent with extensive and long-term populations of *A. vaginae* in humans as part of an ancestral state. *L. iners* showed less diversity, reflecting more recent colonization or a smaller historical population size during human history. Comparative analysis revealed that strains recovered in this study were divergent from previously sequenced strains (Figs. 4 and 5), indicating the novelty of this study and the importance of strain-level analysis of metagenomes. Mobile element-like genes, including virulence and antibiotic resistance genes, existed between strains, demonstrating their adaptation in different environments related to vaginal health.

This study has several limitations. First, the relatively small number of samples included in the analysis limited this study. Likely, a larger

sample with diverse countries of origin would probably reveal more information. Second, our research subjects were recruited only in Beijing. The generalizability of these results will be augmented when applied to other populations of women.

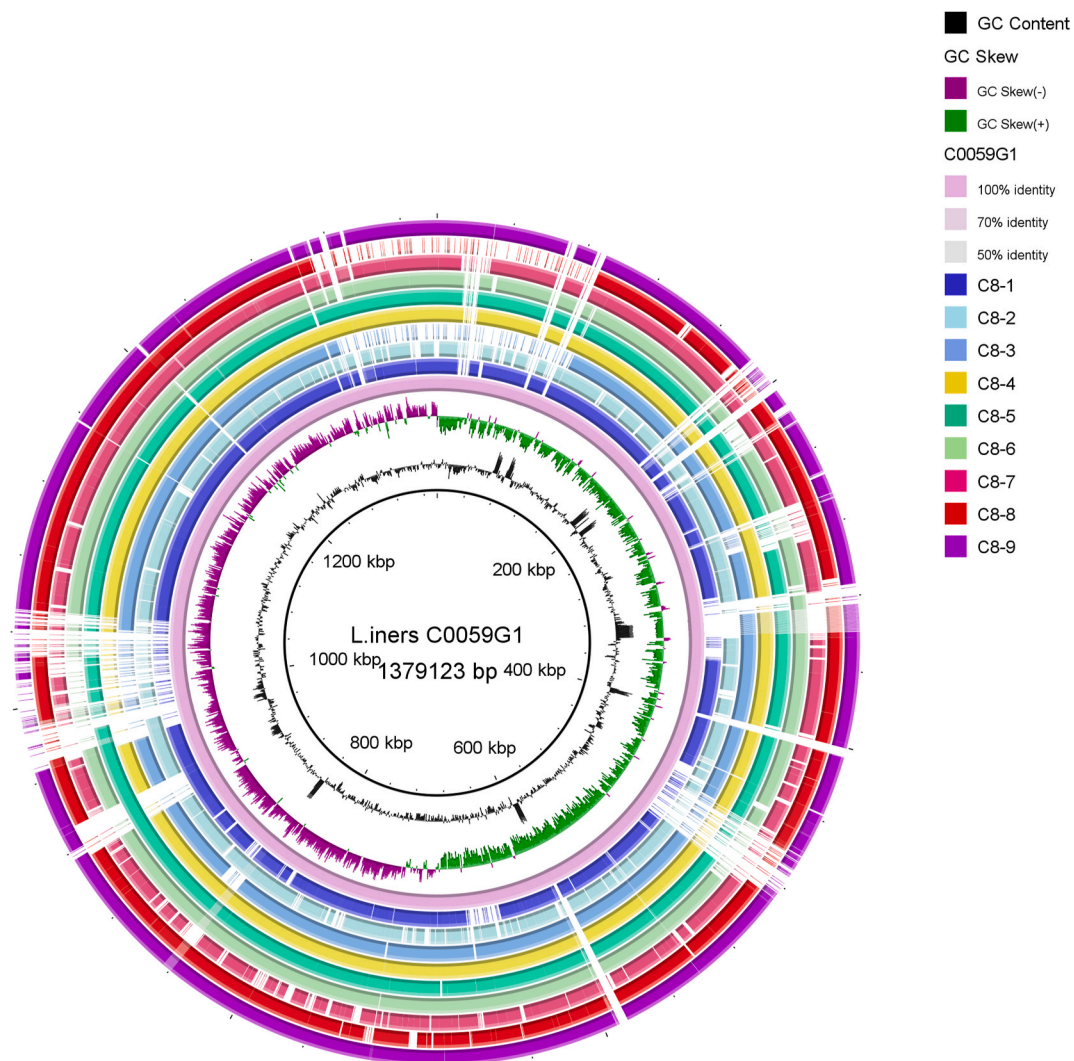
Our results suggest the species-specific composition and functional potential in the vaginal microbiome of healthy women. We highlight the importance of strain-level analysis further to understand the human vaginal microbiome and disease risks.

#### Nucleotide sequence accession numbers

All genomic sequences analyzed in the study have been deposited in GenBank using the accession number PRJNA630342.

#### Author contributions

Fei Liu: Investigation, Visualization, Methodology, Software, Writing - original draft. Yingshun Zhou: Investigation, Formal analysis. Lisi Zhu: Methodology, Investigation. Zhiyi Wang: Validation. LiangKun Ma: Supervision, Methodology. Yuan He: Conceptualization, Methodology. Pengcheng Fu: Supervision, Methodology.



**Fig. 5.** Whole-genome alignments from the *L. iners* clinical isolates described in this study. The two innermost rings represent the GC content of that area and the GC-skew, respectively. The following ten rings each represent one genome of the *L. iners* strains. The saturation of the color in these rings indicates a BLAST hit of identity. Primary differences between strains were the hypothetical genes and mobile element-like genes, including genes involved in CRISPR-associated, restriction-modification systems, virulence-associated, and antibiotic resistance. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

## Declaration of competing interest

The authors declare that they have no competing interests.

## Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (2017YFD0400602 and 2016YFC1000102-2), the National Natural Science Foundation of China (31500114), and the Beijing Municipal Natural Science Foundation (7192169).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.synbio.2021.04.002>.

## References

- Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019;25(4):679–89.
- Chen C, Song X, Wei W, Zhong H, Dai J, Lan Z, Li F, Yu X, Feng Q, Wang Z, et al. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat Commun* 2017;8(1):875.
- Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, et al. The vaginal microbiome and preterm birth. *Nat Med* 2019;25(6):1012–21.
- Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, Agnew KJ, Marrazzo JM, Fredricks DN. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS One* 2010;5(4):e10197.
- Haque MM, Merchant M, Kumar PN, Dutta A, Mande SS. First-trimester vaginal microbiome diversity: a potential indicator of preterm delivery risk. *Sci Rep* 2017;7(1):16145.
- Son K-A, Kim M, Kim YM, Kim SH, Choi S-J, Oh S-Y, Roh C-R, Kim J-H. Prevalence of vaginal microorganisms among pregnant women according to trimester and association with preterm birth. *Obstet Gynecol Sci* 2018;61(1):38–47.
- Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Bieda J, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* 2014;2:18.
- Subramaniam A, Kumar R, Cliver SP, Zhi D, Szychowski JM, Abramovici A, Biggio JR, Lefkowitz EJ, Morrow C, Edwards RK. Vaginal microbiota in pregnancy: evaluation based on vaginal flora, birth outcome, and race. *Am J Perinatol* 2016;33(4):401–8.
- Younes JA, Lievens E, Hummelen R, van der Westen R, Reid G, Petrova MI. Women and their microbes: the unexpected friendship. *Trends Microbiol* 2018;26(1):16–32.
- DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS, Wong RJ, Shaw G, et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* 2015;112(35):11060–5.
- Bradshaw CS, Sobel JD. Current treatment of bacterial vaginosis—limitations and need for innovation. *J Infect Dis* 2016;214(Suppl 1):S14–20.
- Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 2011;108(Suppl 1):4680–7.
- Liu F, Zhu Y, Yi Y, Lu N, Zhu B, Hu Y. Comparative genomic analysis of *Acinetobacter baumannii* clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. *BMC Genom* 2014;15:1163.
- Xu Y, Liu F, Chen S, Wu J, Hu Y, Zhu B, Sun Z. In vivo evolution of drug-resistant *Mycobacterium tuberculosis* in patients during long-term treatment. *BMC Genom* 2018;19(1):640.
- Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UM, Zhong X, Koenig SS, Fu L, Ma ZS, Zhou X, et al. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 2012;4(132):132ra152.
- Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* 2011;108(Suppl 1):4680–7.
- Petrova MI, Reid G, Vanechoutte M, Lebeer S. *Lactobacillus iners*: friend or foe? *Trends Microbiol* 2017;25(3):182–91.
- Togni G, Battini V, Bulgheroni A, Mailland F, Caserini M, Mendling W. In vitro activity of nifuratel on vaginal bacteria: could it be a good candidate for the treatment of bacterial vaginosis? *Antimicrob Agents Chemother* 2011;55(5):2490–2.
- De Backer E, Verhelst R, Verstraelen H, Claeys G, Verschraegen G, Temmerman M, Vanechoutte M. Antibiotic susceptibility of *Atopobium vaginae*. *BMC Infect Dis* 2006;6:51.
- Cao J, Hu Y, Liu F, Wang Y, Bi Y, Lv N, Li J, Zhu B, Gao GF. Metagenomic analysis reveals the microbiome and resistome in migratory birds. *Microbiome* 2020;8(1):26.
- Liu F, Wang Y, Gao GF, Zhu B. Metagenomic analysis reveals the abundance and diversity of ARGs in children's respiratory tract microbiomes. *J Infect* 2020;80(2):232–54.
- Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, et al. Strains, functions and dynamics in the expanded human microbiome project. *Nature* 2017;550(7674):61–6.
- Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;12(10):902–3.
- Franzosa EA, McIver LJ, Raunavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018;15(11):962–8.
- Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, Brotman RM, Ravel J. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat Commun* 2020;11(1):940.
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;12(6):R60.
- Peng Y, Leung HC, Yiu SM. Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28(11):1420–8.
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;7:e7359.
- Parks DH, Imelfort M, Skennerton CT, Hugenholz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25(7):1043–55.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36(10):996–1004.
- Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 2015;31(17):2877–8.
- Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31(22):3691–3.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068–9.
- Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Boucard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;48(D1):D517–25.
- Jakobsson T, Forstum U. *Lactobacillus iners*: a marker of changes in the vaginal flora? *J Clin Microbiol* 2007;45(9):3145.
- Verstraelen H, Verhelst R, Claeys G, De Backer E, Temmerman M, Vanechoutte M. Longitudinal analysis of the vaginal microflora in pregnancy suggests that *L. crispatus* promotes the stability of the normal vaginal microflora and that *L. gasseri* and/or *L. iners* are more conducive to the occurrence of abnormal vaginal microflora. *BMC Microbiol* 2009;9:116.
- Breshears LM, Edwards VL, Ravel J, Peterson ML. *Lactobacillus crispatus* inhibits growth of *Gardnerella vaginalis* and *Neisseria gonorrhoeae* on a porcine vaginal mucosa model. *BMC Microbiol* 2015;15:276.
- He Y, Huang Y, Zhang Z, Yu F, Zheng Y. Exploring profile and potential influencers of vaginal microbiome among asymptomatic pregnant Chinese women. *PeerJ* 2019;7:e8172.
- Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, Foster JA, Forney LJ. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J* 2007;1(2):121–33.
- Polatti F. Bacterial vaginosis, *Atopobium vaginae* and nifuratel. *Curr Clin Pharmacol* 2012;7(1):36–40.