scientific reports

Check for updates

OPEN

Urogenital colonization and pathogenicity of *E. Coli* in the vaginal microbiota during pregnancy

Nassim Boutouchent[1,2], Thi Ngoc Anh Vu[3], Luce Landraud[4] & Sean P. Kennedy[1]✉

This study explores the role of the vaginal microbiota (VM) in the pathophysiology of asymptomatic bacteriuria (ASB) in a cohort of 1,553 pregnant women. Worldwide, *E. coli* remains the most common etiological agent of bacteriuria during pregnancy and also a major causative agent of newborn infections. A healthy VM is typically characterized by low diversity and is dominated by lactic acid-producing species, notably those from the *Lactobacillus* genus. Our results point to decreases in *Lactobacillus spp* associated with an increase of gut-microbiota-associated species from the Enterobacterales order. *Escherichia coli* exhibited the most pronounced increase in abundance within the VM during bacteriuria and was notably associated with ASB. Molecular typing and antimicrobial resistance characterization of 72 metagenome assembled *E. coli* genomes (MAGs) from these pregnant women revealed a genomic signature of extraintestinal pathogenic *E. coli* ("ExPEC") strains, which are involved in various extraintestinal infections such as urinary tract infections, newborn infections and bacteremia. Microbial diversity within the vaginal samples from which an *E. coli* MAG was obtained showed a substantial variation, primarily marked by a decrease in abundance of *Lactobacillus* species. Overall, our study shows how disruption in key bacterial group within the VM can disrupt its stability, potentially leading to the colonization by opportunistic pathogens.

Keywords Vaginal microbiota, Microbial diversity, Genome Assembly, Extra-intestinal pathogenic *E. Coli* (ExPEC), Bacteriuria, Urinary tract infection (UTI)

Urinary tract infections (UTIs) are among the most common infectious diseases, disproportionally affecting women and have a high recurrence rate of 20–30%[1]. These infections are identified through both clinical assessments and microbiological testing. While UTIs are systematically treated in the general population, asymptomatic bacteriuria (ASB) – defined by urine culture $\geq 10^5$CFU/mL, without symptoms– is screened for and treated only in specific populations, including patients undergoing endourological procedures and pregnant women[2,3]. Current knowledge suggests a link between untreated ASB and the onset of severe UTIs, such as acute pyelonephritis (APN) during pregnancy[2,4,5]. The progression from ASB to APN could be associated with serious maternal and fetal complications, such as anemia, sepsis, preterm birth, and low birth weight. Given these risks, the prevailing medical consensus recommends routine screening and treatment of ASB during pregnancy to mitigate risks of adverse outcomes[6]. Increasing resistance to first line antibiotic treatment, such as beta-lactam among Enterobacterales, however, complicates clinical management of these conditions. Furthermore, the toxicity of second line treatments such as Nitrofurantoin and Trimethoprim renders the management of these conditions increasingly challenging[6–8].

Worldwide, *Escherichia coli* remains the predominant bacterium isolated in both UTIs and ASB during pregnancy, accounting for more than half of reported cases[1,9]. Following *E. coli*, species such as *Klebsiella pneumoniae*, *Staphylococcus aureus*, and other members of the Enterobacterales order are commonly identified. Although generally considered a commensal of the gut microbiota, *E. coli* exhibits extensive genomic plasticity, which facilitates a wide array of pathogenic capabilities. These strains are classified into various pathovars, including seven intestinal pathogenic *E. coli* "InPEC" pathovars, and three extraintestinal pathogenic "ExPEC" pathovars. ExPEC pathovars notably include uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli*(NMEC),

[1]Department of Computational Biology, Institut Pasteur, Université Paris Cité, 75015 Paris, France. [2]Département de Microbiologie, CHU de Rouen, 76000 Rouen, France. [3]VNU-Institute of Microbiology and Biotechnology, Vietnam National University, Hanoi, Vietnam. [4]Université Paris Cité and Université Sorbonne Paris Nord, INSERM, IAME, F-75018 Paris, France. ✉email: sean.kennedy@pasteur.fr

and those associated with pneumonia[10,11]. Molecular typing methods have provided essential information in the characterization and epidemiological study of *E. coli* strains. These methods have shown the species to be classified into eight phylogenetic groups (A to G), forming two clusters, based on their phylogenetic relationships: phylogroups B2, G, F, D, and phylogroups A, B1, C, E[12,13]. Such classification is interesting as it strongly correlates with the pathogenic potential of strains. Notably, it has been demonstrated that B2 and D phylogroup strains harbor a rich repertoire of virulence genes that enable adhesion, iron acquisition, immune evasion, and toxin production, leading to extraintestinal infections such as UTIs, neonatal meningitis and bacteriemia[11,14,15]. However, there is no specific "genetic signature" that reliably distinguishes UPEC from other specific pathovars within ExPEC strains[16–18].

Our understanding of the UTI pathophysiology has significantly advanced over the past two decades. It is now evident that the pathogenesis of these infections is complex, relying on a dynamic process of bacterial colonization[18,19]. Certain bacteria, such as UPEC, colonize the peri-urethral flora and ascend through the UT, triggering inflammation and causing disease. A study conducted by Magruder et al. (2019) demonstrated that the increased abundance of *E. coli* in the gut microbiota was associated with future development of *E. coli* UTIs[20]. Additionally, the strains isolated from the gut showed a higher degree of genomic similarity with those present in the urine of the same individuals, providing further evidence for the hypothesis that the gut microbiota serves as a reservoir for UPEC strains. The vaginal microbiome necessarily plays in important role in any model postulating that intestinally derived UPEC strains are responsible for ascending colonization. Potential pathogens must interact with this distinct host ecosystem, differing from that of the intestinal milieu and characterized by a unique physicochemical environment[21].

The vaginal microbiota (VM) has emerged as a pivotal component in understanding women's health and perinatal health of newborns[22,23]. Unlike the diverse gut microbiota, the VM exhibits a more restricted microbial diversity[22,24,25]. Within this community, Lactobacilli stand out as key players, shaping a vaginal environment that is generally considered healthy[23]. Four *Lactobacillus* species are dominant and anchor distinct community state types (CSTs) – *Lactobacillus crispatus* (CST I), *Lactobacillus gasseri* (CST II), *Lactobacillus iners* (CST III), and *Lactobacillus jensenii*(CST V)[26,27]. These lactobacilli produce lactic acid from glycogen degradation products found in epithelial cells, resulting in an acidic vaginal pH (pH < 4.5), creating an environment unfavorable for pathogenic bacteria colonization and growth[28,29]. While *L. crispatus* is generally associated with optimal health, the role of *L. iners* is less certain and is currently under investigation for its potential association with dysbiotic states or infections[30,31]. Further, not all communities can be classified into standard CSTs. Some communities demonstrate reduced lactic acid-producing species and increased anaerobic bacteria. The CST IV community is characterized by lower *Lactobacillus* presence accompanied with greater overall diversity including an increase in anaerobic organisms such as *Gardnerella vaginalis*[26]. This community is not always associated with pathology, but in some women, it may contribute to symptomatic dysbiosis, known as bacterial vaginosis (BV)[32]. The composition of the VM undergoes changes throughout a woman's life. For example, during pregnancy, the VM typically shows reduced richness and diversity compared to non-pregnant women, with increasing dominance of *Lactobacillus* species[33]. Hormonal fluctuations during pregnancy play an important role in shaping vaginal microbiota community (VMC) dynamics, with higher estrogen levels inducing greater glycogen production, promoting a shift towards a stable microbiota state characterized by the dominance of acidophilic lactobacilli[22,34].

Based on numerous observational studies, it has been hypothesized that the VM plays an important role in the pathogenesis of UTIs[35–37]. One hypothesis is founded on a reported increased risk of UTIs associated with decreased *Lactobacillus*-derived H_2O_2levels and the presence of BV, suggesting that a disruption in the composition of this microbiota could favor the onset of UTIs[38]. The vaginal compartment itself has been proposed as a reservoir for *E. coli*and other Enterobacterales capable of causing UTIs[21,39,40]. A compelling argument could be made that an increase in *E. coli* associated with a decrease in lactobacilli potentially underlines the interconnection between intestinal and urogenital microbiota in the development of UTIs. Critically, few studies have addressed the impact of the VM diversity and its role during bacteriuria in pregnancy. Although *E. coli* is a principal etiological agent of UTIs and also in newborn infections, its characterization within the vaginal compartment remains poorly documented. In this study, we describe the VM diversity associated with ASB during pregnancy, and we characterize *E. coli* strains present in the vaginal compartment, thereby enhancing our understanding in bacterial dynamic that could influence maternal and neonatal health outcomes.

Materials and methods

Ethics declaration

The reported study (RCB # 2017-A02755-48) was reviewed by the French ethics committee, "Comité de Protection des Personnes Nord-Ouest III" (Caen, France), who granted approval on November 4th, 2017 (CPP dossier: 2017-74). All individuals enrolled in the InSPIRe cohort provided signed informed consent that included authorization for the collection of medical information and biological samples. The present work includes only the non-human fraction of DNA extracted from vaginal swabs from expectant mothers. Clinical metadata of patients was collected in a secured electronic database. Exports from the database used for analysis were anonymized. Processed vaginal swabs were labeled with corresponding anonymous inclusion numbers. All methods and protocols used in this study were carried out in accordance with the principles outlined in the Declaration of Helsinki.

Dataset and clinical design

We preformed shotgun metagenomic sequencing data analysis of vaginal swab samples collected from an observational clinical study of pregnant women during the third trimester, conducted over the period from 2019 to 2023, and involving three French hospitals in the Ile de France region: Hôpital Cochin, Bichat and Louis Mourier. DNA extraction, shotgun metagenomic sequencing and taxonomic assignment was performed as

reported in Baud et al. (2023)[41]. Low-quality reads < Q30 mean Phred scores were removed and Trimmomatic (v 0.39) was used to remove any adapter or primer sequences. Samples groups, described below, were drawn from a cohort of 1950 samples. Human reads were removed by mapping to the human genome (hg38) using kraken2[42]. Taxonomic classification was performed with moonbase tool (ver. 1.0.0), employing both MetaPhlan3 and Kraken2[43,44]. Species-level abundance counts were adjusted using Bracken[45]. Analyses using relative abundances were computed by dividing individual taxa reads by total reads for each sample. Analyses using absolute read quantifications were normalized using DESeq2-type geometric mean normalization, implemented through the moonstone library (ver.1.0.0)[46].

Three groups were defined to analyze VM biodiversity during urinary tract colonization: (1) Asymptomatic bacteriuria (ASB) group, characterized as women with urine cultures $\geq 10^5$ CFU/mL at the time of the vaginal sampling. (2) Without ASB during pregnancy (Wo-ASB), consisting of women with negative urine culture at the moment of the vaginal sampling, matched with the ASB group to limit confounding factors such as antibiotic use during pregnancy. (3) A control group (HG) of healthy women was constituted from samples collected during routine pregnancy follow-up consultations between 34 and 36 weeks of gestation. Criteria for the control group included a body mass index (BMI) below 30 Kg/m^2, no significant medical antecedents, absence of alcohol or tobacco use, no history of preterm delivery or UTIs, and no antibiotic use during pregnancy. The HG group provides a baseline vision for a healthy state of the VM.

Analysis environment and statistical testing

All the analyses were performed in a Python (ver. 3.8) environment. Statistical tests were conducted using the SciPy library (ver.1.13) with an alpha risk of 0.05. P -values were adjusted using the Benjamini-Hochberg (BH) correction to control the false discovery rate (FDR) with a cutoff of 5%. All reported p-values are those adjusted for multiple testing. When appropriate, analyses were adjusted for gestational age using ANCOVA with the pg.ancova() function from the pingouin library (ver. 0.5.5). Figures were generated using Plotly library (ver. 5.18).

Co-abundance correlation analysis

SparCC (Sparse Correlations for Compositional data) algorithm implemented through FastSpar (ver. 1.0.0) was used to estimate pairwise correlation values between species[47,48]. To explore competitive relationships between taxa within the vaginal ecosystem, an abundance correlation matrix was computed for the 50 most abundant species. Hierarchical clustering was subsequently applied using the average linkage method, based on the correlation distances between species profiles, to identify communities that are co-abundant. The clustermap() function in the Seaborn library (ver. 0.13.1) was used for analysis and visualization.

Diversities computation and principal coordinates analysis

Intra-sample diversity, α-diversity, of samples was assessed by Shannon index, and Faith's phylogenetic diversity (PD) using the scikit-bio.diversity.alpha.diversity module (ver. 0.5.9). A Mann-Whitney *U* statistical test was subsequently applied to test for statistical significance. For inter-sample, β-diversity, Bray-Curtis (BC) and Weighted UniFrac (w-UniFrac) dissimilarity matrices were computed using beta.diversity() functions[41]. The significance of variations among groups was evaluated using the Permutational Multivariate Analysis of Variance (PERMANOVA) diversity module on the scikit-bio library (ver. 0.5.9). Faith's PD and w-UniFrac indices used a 16 S rDNA phylogenetic tree to incorporate phylogenetic diversity information.

Principal Coordinates Analysis was computed from the generated BC and w-UniFrac matrices using the pcoa() function from the scikit-bio library's stats.ordination module. Visualization of the first two components, (PC1) and (PC2), was generated to observe the spatial distribution of groups.

Linear discriminant analysis with size effect (LEfSe)

Biomarkers identification was performed using Linear Discriminant Analysis (LDA). An in-house Lefse_analysis Python class pipeline, based on Segata et al.'s algorithm and optimized for two-group comparisons, was built using the sklearn.discriminant_analysis LinearDiscriminantAnalysis module of the scikit-learn library (ver. 1.3.1)[49]. Lefse_analysis takes relative species abundances and a binary group label column, corresponding to the two groups being compared. Briefly, a Mann-Whitney *U* test for pairwise comparisons of relative abundances across taxa and BH correction (FDR $\leq$ 5%) was applied to p-values. Directionality of abundance was calculated for the two-group analysis by comparing mean quantification values for each significant taxa. Pandas (ver. 2.0.3) was used to assemble result: impacted group per species, log-transformed LDA scores, and the direction of abundance. The interpretability of results is enhanced by phylogenetic mapping of significant taxa to the 16 S rRNA gene phylogenetic tree using the dendropy module (ver. 4.6.1). We focused on biologically relevant species with relative abundances $> 10^{-6}$. We then aligned the sample counts with the smallest group through bootstrap resampling ($n = 1000$) to minimize sample size discrepancies. Finally, iTOL (ver. 6.0) was used to annotate the output phylogenetic tree[50].

Linear regression analyses

Using the Bracken-corrected and normalized counts, we constructed a linear regression model to assess the association of bacteriuria with taxa at different phylogenetic levels, starting from the order level. Significant associations with p-values < 0.01 were retained and subjected to a second model to evaluate their variation while accounting for covariates. Adjusted p-values < 0.05 were considered significant. All linear regressions were performed using the OLS() function from the statsmodels library (version 0.13.2).

Snakemake pipeline for *E. coli* Genome Assembly and Annotation

Samples containing at least *E. coli* 30,000 reads, or ~2X coverage of the expected genome size, were selected for genomic assembly. Sequence data were processed through an in-house Snakemake (ver. 8.11.6) pipeline consisting of three main stages: mono-species filtering, assembly, and annotation. First, metagenomic reads were filtered by mapping to a database (DB) of representative *E. coli* genomes. Representative genomes were selected for completeness and diversity representation, and included 2 RefSeq reference genomes and 30 clinical isolate genomes (Supplemental Table S1). Individual sample reads were mapped against this DB using Bowtie2 (ver. 2.5.1), retaining read pairs where at least one read aligned end-to-end with the reference sequence for the subsequent assembly stage[51]. Second, this subset of exclusively E. coli read pairs were assembled with SPAdes (ver. 3.15.5)[52]. Finally, metagenomic assembled genomes (MAGs) were annotated using Prokka (ver. 1.14.5)[53]. Additional analyses involved the utilization of Panaroo (ver. 1.5.0) for pan-genome analysis, and IQ-TREE (ver. 1.6.12) for generating a phylogenetic comparisons[54,55]. Quality metrics for assemblies were evaluated using CheckM (ver 1.2.3)[56].

Molecular typing and antimicrobial resistance (AMR)

E. coli phylogroup determination was conducted using ClermonTyping method which classifies phylogroups based on specific gene markers[13]. Sequence type (ST) identification was performed using MLST (ver. 2.23.0)[57]. Additionally, *fimH* allele and O: H typing was carried out using FimTyper (ver. 1.0.1) and ectyper (ver. 0.9.0), respectively. As a control for genome completeness, we also confirm the presence of the core gene *rpoB* in all phylotyped assemblies. ResFinder tool (ver. 4.4.3) was used to assess the antimicrobial resistance (AMR) profiles of assembled genomes[58].

Results

Clinical characteristics of pregnant women among groups

In our study, vaginal swabs from 1,553 pregnant women were analyzed, focusing on three groups to explore the microbial diversity during bacteriuria. The ASB group comprised 42 women identified as having positive urine cultures obtained at the time of vaginal sampling. The prevalence of ASB in our cohort was low but consistent with the incidence rates of ASB in western countries, which range from 2 to 7% among pregnant women[2]. The Wo-ASB group included 324 women who did not experienced bacteriuria during their pregnancy and were matched with ASB individual across major clinical variables. Lastly, the HG group encompassed 138 women who met stringent health criteria, including absence of significant medical antecedents and antibiotic use. Regarding anthropometric variables, the age was consistent across all groups, averaging 33 years. The average BMI for all groups remained within a normal range (18.5–24.9 Kg/m^2), between 22.3 and 24.7 Kg/m^2. No significant covariates were identified in the medical histories collected for this study when comparing individuals according to the occurrence of bacteriuria (Table 1). Likewise, antibiotic consumption during pregnancy was comparable between these two groups (Wo-ASB and ASB), with an average interval of three months between the last antibiotic intake and vaginal sampling. The prescribed class of antibiotics were also consistent across these groups, with amoxicillin being the most frequently used antibiotic during pregnancy (Table 1).

Microbial abundance profile during bacteriuria

In both the HG and Wo-ASB groups, *Lactobacillus* species—specifically *L. crispatus*, *L. iners*, *L. gasseri*, and *L. jensenii*—predominated the VM, with *L. crispatus* being the most prevalent (Fig. 1a). The ASB group, however, displayed significant differences in microbial abundances and community composition when we compared these individuals to the other groups (Fig. 1b). At the phylum level, Proteobacteria exhibited a twofold increase, while there was a corresponding twofold decrease in Firmicutes *(p-value < 0.0001)*. A further significant increase in the Ascomycota was observed when compared to the Wo-ASB *(p-value = 0.002)*, and the HG groups (*p-value = 0.01)*. At the family-level, significant differences were also evident within the ASB group, most notably we detected an increase in Enterobacteriaceae and a decrease in Lactobacillaceae, suggesting alterations in microbial community structure within the VM during ASB (Fig. 1b). Furthermore, *E. coli* within the Enterobacterales order exhibited a substantial increase in abundance compared to the Wo-ASB *(p-value < 0.0001)* and HG groups *(p-value < 0.0001)*, placing it among the top five predominant species in the ASB group (Fig. 1a). The prevalence of other species commonly isolated during bacteriuria in pregnant women, such as those from the *Klebsiella*, *Enterobacter*, and *Morganella* genera, was also significantly higher in the ASB group than in the Wo-ASB *(p-values = 0.002; 0.002; 0.004*, *respectively)* and HG group *(p-values = 0.008*, *0.006*, *0.001*, *respectively)*.

Microbial co-abundance clusters among the VMC

To investigate competitive and cooperative interactions within the VMC, a correlation matrix for the 50 most abundant species was calculated. Hierarchical clustering of the matrix, based on the correlation distances between species profiles, was used to investigate VM community structure. This analysis revealed three distinct co-abundance clusters (CACs) (Fig. 1c). The first cluster, CAC I, is characterized by facultative or strict anaerobes, including genera such as *Gardnerella and Prevotella*, along with species like *Aerococcus christensenii* and *Atopobium vaginae* commonly associated with vaginal dysbiosis. Notably, *L. iners*, a species whose impact on vaginal health is still debated, is observed to co-occur with these dysbiosis-associated species.

The second cluster, CAC II, comprises phylogenetically related species primarily from the *Lactobacillus* genus, including *L. crispatus*, *L. jensenii*, *L. johnsonii* and *L. gasseri*, all known for their hydrogen peroxide (H_2O_2) production and their role in promoting a healthy vaginal environment. In contrast, the third cluster, CAC III, comprises organisms often identified as opportunistic pathogens, such as various Enterobacterales (e.g. *E. coli*, *K. pneumoniae*, and *Proteus mirabilis*), alongside *S. aureus*, and *E. faecium*. Notably, CAC II displayed a wide range of paired correlations, both positive and negative, with a notable prevalence of antagonistic interactions

	Wo-ASB ($n=324$)	ASB ($n=42$)	p-value	adjusted p-value
Average age (years)	33.0	32.9	*ns*	*ns*
Average BMI (Kg/m^2)	24.7	24.5	*ns*	*ns*
Average Gestational Age (weeks)	33.5	27.3	**< 0.0001**	**< 0.0001**
Time of vaginal sampling			**< 0.0001**	**< 0.0001**
Before delivery	150	37		
Delivery room	174	5		
Average Pregnancy Term (weeks)	39.1	33.6	**< 0.0001**	**< 0.0001**
Smoking during pregnancy (%)	6.27 *(19/303)*	5 *(2/40)*	*ns*	*ns*
History of premature delivery (%)	18.2 *(50/274)*	16.7 *(6/36)*	*ns*	*ns*
Type of pregnancy			*ns*	*ns*
Simple pregnancy	312	38		
Twin pregnancy	12	4		
Gestational diabetes (%)	26.5 *(68/256)*	27.2 *(9/33)*	*ns*	*ns*
Pre-eclampsia (%)	3.5 *(11/313)*	2.4 *(1/41)*	*ns*	*ns*
FGR (%)	4.5 *(14/310)*	7.7 *(3/39)*	*ns*	*ns*
HIV (%)	0.6 *(2/322)*	5 *(2/40)*	*ns*	*ns*
Cerclage (%)	3.8 *(12/312)*	10.5 *(4/38)*	*ns*	*ns*
Delivery mode			*ns*	*ns*
Cesarean	29	3		
Labor induction	119	16		
Spontaneous labor	176	23		
Antibiotic therapy during pregnancy (%)	28.6 (79/276)	38.1 (16/42)	*ns*	*ns*
Prescribed antibiotic				
Amoxicillin	51.9 (41/79)	62.5 (10/16)		
Cefixime	2.5 (2/79)	0 (0/16)		
3rd generation cephalosporin	10.1 (8/79)	6.2 (1/16)		
Metronidazole	16.5 (13/79)	12.5 (2/16)		
Clindamycin	5 (4/79)	0 (0/16)		
Others	13.9 (11/79)	18.7 (3/16)		
Time between last antibiotic intake and vaginal sampling (days)	82.7	82.2	*ns*	*ns*

Table 1. Overview of clinical characteristics among the Wo-ASB and ASB groups. Quantitative variables analyzed by t-test, and categorical variables evaluated with Fisher's exact test. BH corrected p-values are shown. ns: not significant ($p > 0.05$); FGR: Fetal growth restriction; HIV: Human immunodeficiency virus.

(Fig. 1c). These findings underscore the dynamic interactions among these microbial communities, which may influence the overall stability and health of the VM. The antagonistic interactions, particularly, suggest a competitive environment where *Lactobacillus* species play a pivotal role in maintaining microbial balance against opportunistic pathogens.

Comparative analysis of vaginal microbiome diversity across groups

The Shannon index is a common α-diversity metric, which is sensitive to species richness—the array of species present—and also measures evenness, indicating how uniformly individuals are distributed across these species[59]. In our study, the Shannon index revealed no significant differences among the ASB, HG, and Wo-ASB groups (Fig. 2a). However, the community structure of the VM is unique and *Lactobacillus* species are often interchangeable. We therefore complemented our analysis by using Faith's Phylogenetic Diversity (PD), which accounts for the phylogenetic dimension in the assessment of intra-sample diversity. Defined as the cumulative length of all branches on a phylogenetic tree that are unique to each species present in the sample, the Faith's PD index considers the evolutionary relationships among species[60]. This approach revealed significant variations among the groups. There were substantial differences in index between the ASB group and both the Wo-ASB *(p-value = 0.0009)* and the HG groups *(p-value = 0.0002)*. These results, taken together, suggest a difference in the phylogenetic diversity of the VM as opposed to a significant increase in the number of species within the ASB group. However, when considering gestational age in the statistical model we cannot say that Lactobacilli species, dominate members of a health VM, are specifically displaced ($p\text{-value} = 0.6$ for ASB vs. Wo-ASB and *0.2* for ASB vs. HG).

To assess inter-group differences in VM across HG, Wo-ASB, and ASB groups, we compared Bray-Curtis (BC) and w-UniFrac dissimilarity matrices. The key distinction between these matrices lies again in the incorporation of phylogenetic relationships by the w-UniFrac matrix[61,62]. The BC dissimilarity matrix suggested weak differences between the groups ($p\text{-value} = 0.06$), while the UniFrac dissimilarity matrix revealed significant

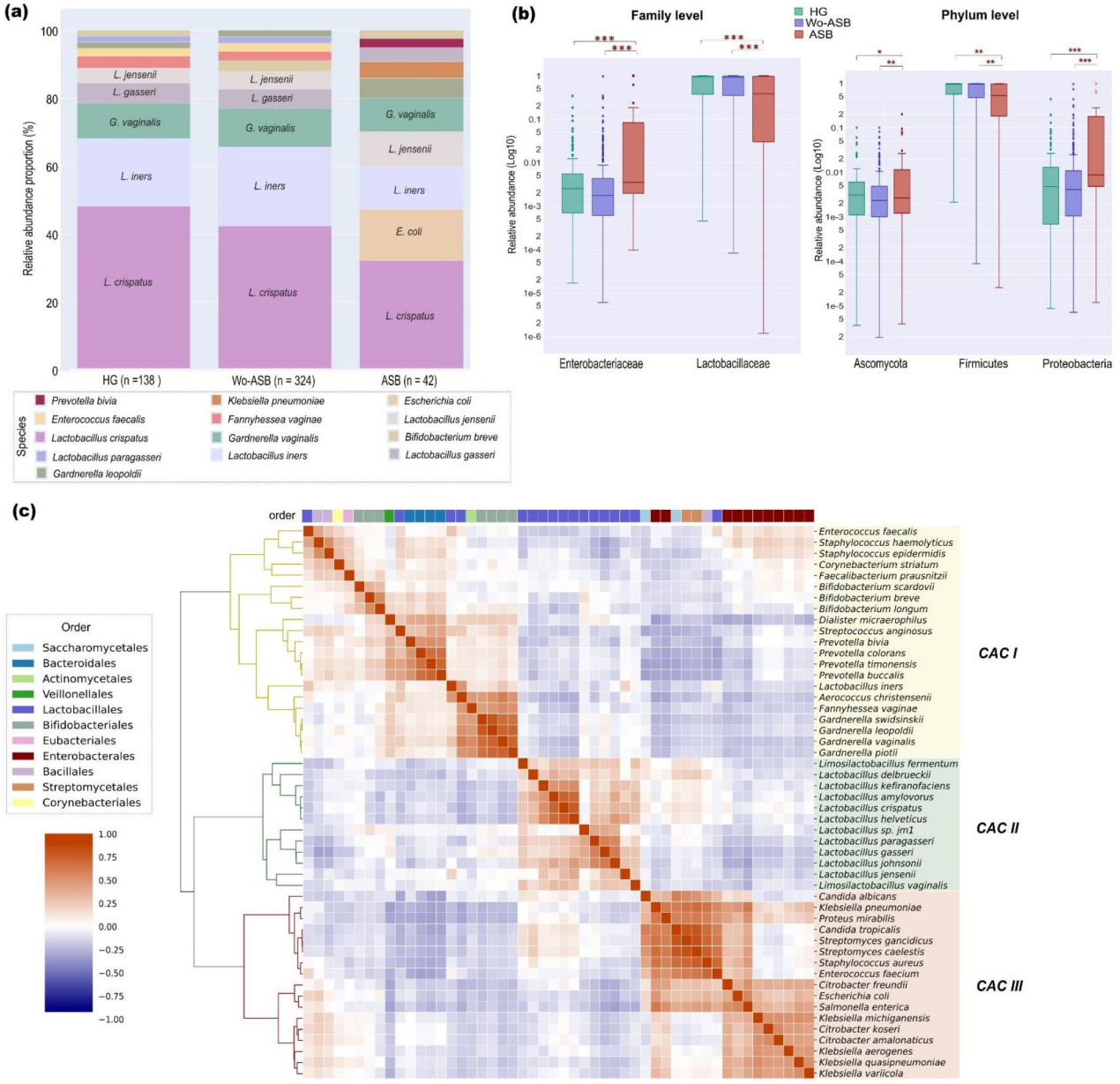

Fig. 1. Microbial Abundance and Correlation Profiles Within the VM of Pregnant Women. (**a**) Stacked bar chart illustrating the percentage abundance of the 10 predominant species in the HG, Wo-ASB and ASB groups. (**b**) Box plot displaying the relative abundances of various taxa (at the family and phylum levels) among the groups. The bounding box represent first and third quartiles, center line indicates the median, and the whiskers extend to 1.5 times the interquartile range from the quartiles. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. (**c**) Hierarchical clustering heatmap of the correlation matrix, computed using SparCC and plotted with the Seaborn (ver.0.13.1) clustermap() function, illustrating the abundance correlations among the 50 most predominant species within the VM of 180 pregnant women from the HG and ASB groups. CAC: Co-Abundance Cluster.

differences (*p-value = 0.001*) (Fig. 2b). Pairwise comparisons between the ASB group and the other two groups (Wo-ASB and HG) showed significant differences in the BC matrix *(p-values = 0.034 and 0.02, respectively)* and the w-UniFrac matrix *(p-values = 0.002 and 0.001, respectively)*. To ensure the robustness of these findings and mitigate biases due to sample size, bootstrap resampling ($n = 1000$) was employed, standardizing the sample sizes across groups to 42. This adjustment abrogated previously significant differences in the BC matrix among these groups (*p-value* of 0.29 and 0.1, respectively), but significant compositional differences persisted in the w-UniFrac matrix (*p-values of 0.007 and 0.002*, respectively). When adjusting for gestational age at the time of sampling and final pregnancy length, the differences were not significant. However, when gestational age at sampling and pregnancy length were considered separately, PERMANOVA tests showed that both explained a significant portion of the variance (*p-value 0.02 and 0.01, respectively*).

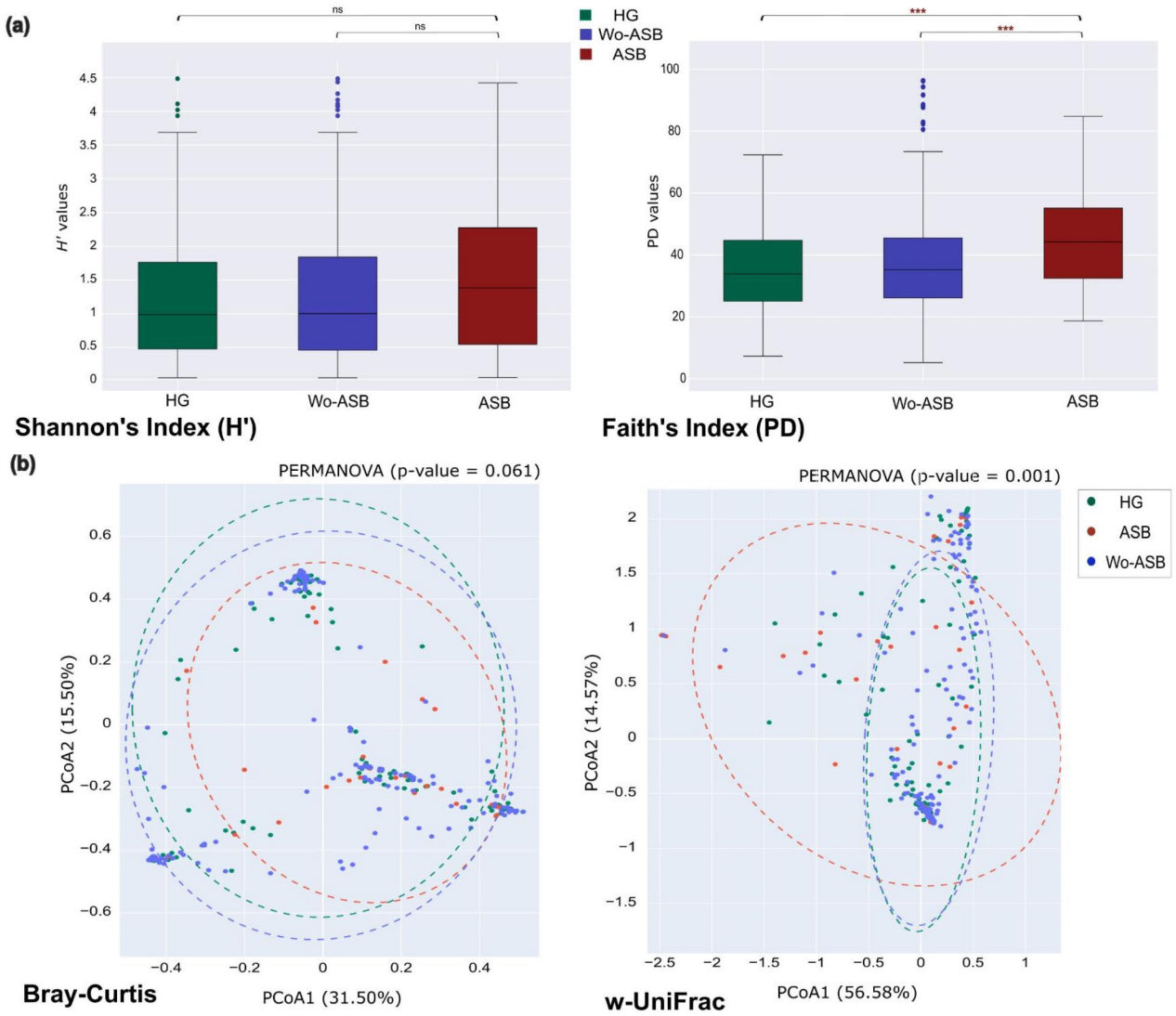

Fig. 2. Comparative Analysis of α-Diversity Metrics and Principal Coordinates Analysis Visualization of β-Diversity Matrices. (**a**) Box plots illustrating the distribution of Shannon (*H'*) and Faith's phylogenetic diversity (PD) indices across HG, Wo-ASB, and ASB groups. The bounds of the box represent the first and third quartiles, center line indicates the median, and the whiskers extend to 1.5 times the interquartile range from the quartiles. ***$p < 0.001$; **ns**: not significant ($p > 0.05$). (**b**) Principal Coordinates Analysis 2D projection of Bray-Curtis and weighted-UniFrac matrices, showing the first two principal components (PC1 and PC2). Points show individual samples, colored by group (HG in green, ASB in red and Wo-ASB in blue. Colored ellipses indicate the 95% confidence intervals for each group.

LEfSe analysis reveals alterations in VM during bacteriuria

To further explore specific taxa differences observed between these ecosystems, a LEfSe workflow adapted for pairwise analysis was computed. This workflow yielded three main results: it identified bacterial species and their phylogenetic relationships which differentiate groups, it calculated species relative abundance between groups, and assessed their strength in contributing to group differentiation (Supplemental Figure S1).

LEfSe analysis of HG and ASB groups identified 52 bacterial species, 43 of which had a significant impact on group separation, indicated by an LDA (log10) score exceeding 2 (Fig. 3). Notably, species more abundant in the HG were phylogenetically related and all belonged to the *Lactobacillus* genus, thus defining a core community for the healthy group. Conversely, the ASB group exhibited a distinct and complex metagenomic signature, characterized by a reduction in *Lactobacillus* as well as an increase in various phylogenetically distant species. One of the key findings from this analysis was the identification of numerous species belonging to traditionally gut-associated orders—such as Enterobacterales and Bacteroidales—which significantly influenced the distinction between the ASB and HG groups (Fig. 3). Although *E. coli* did not significantly influence the differentiation between these groups, the elevated abundance of other species frequently isolated during UTIs in pregnancy, such as *P. mirabilis*, *Enterobacter cloacae*, *Citrobacter freundii*, and *S. aureus*, significantly

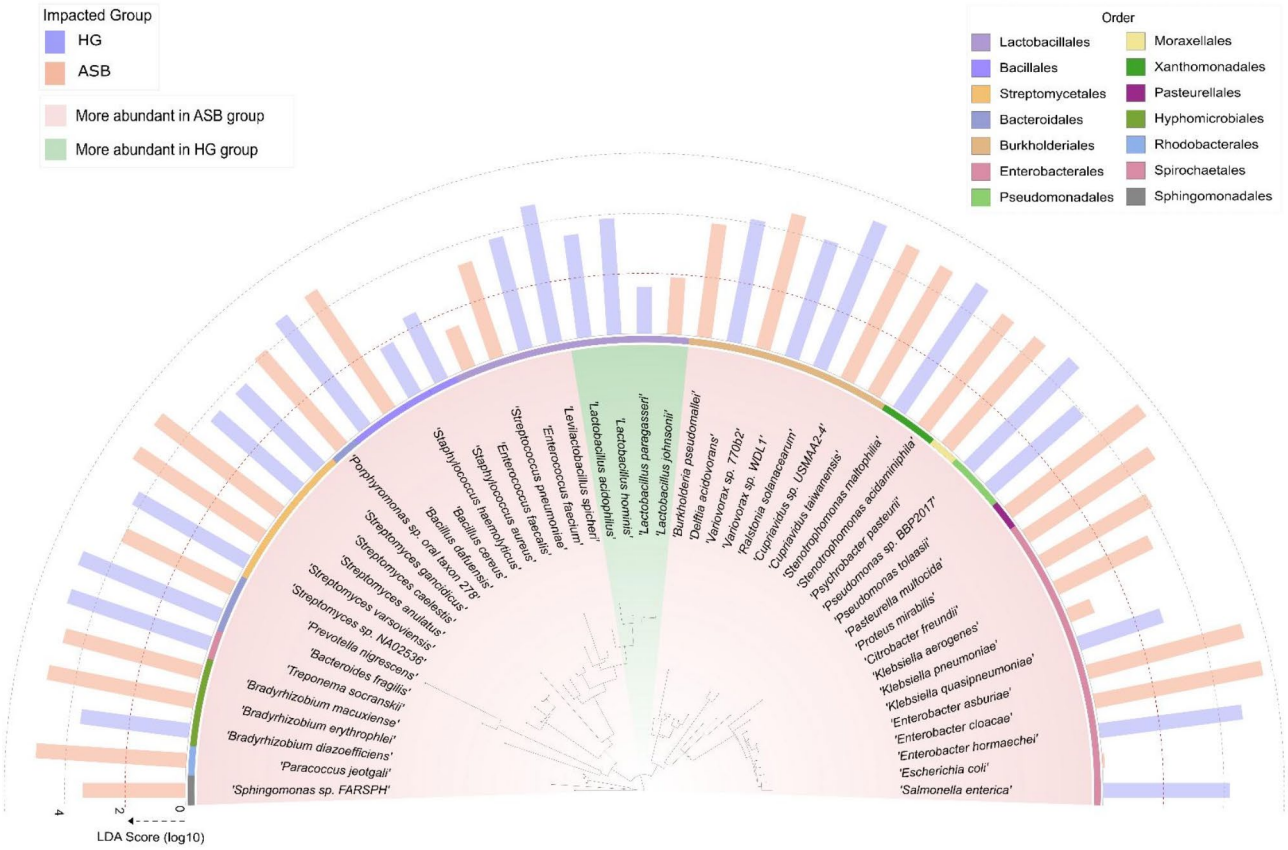

Fig. 3. 16 S rRNA gene Phylogenetic Tree Displaying Species Identified by LEfSe Workflow. Phylogenetic tree (inner arc) displaying bacterial species identified through LEfSe workflow, with their Linear Discriminant Analysis (LDA) scores. Species are color-coded: green indicates higher abundance in the HG group, and red denotes greater abundance in the ASB group. Center arc shows the order of identified markers, with exterior bars showing effect sizes of species on group separation. A red line at a log10 LDA score of 2 indicates the threshold for a significant impact on group differentiation. Lactobacillales are shown in lavender, Bacillales in deep blue, Streptomycetales in bright orange, Bacteroidales in blue gray, Burkholderiales in pale orange, Enterobacterales in muted pink, Pseudomonadales in light green, Moraxellales in pale yellow, Xanthomonadales in dark green, Hyphomicrobiales in olive green, and Pasteurellales in dark purple.

impacted the differentiation among these groups. Other species have also been observed, including anaerobic species such as *Prevotella nigrescens* and *Porphyromonas sp.*commonly noted during BV[32,63]. Concurrently, the influence on group differentiation was also noted to be exerted by species from the orders Pseudomonadales, Xanthomonadales, and Burkholderiales. The detection of these species, found in both environmental ecosystems and human-associated microbiomes, is further evidence of alterations in the physicochemical environment of the VM, reducing barriers against diverse colonization.

Applying the LEfSe workflow between the Wo-ASB and ASB groups identified more specific markers. At the order level, Enterobacterales contributes modestly to the differentiation between these groups, with a log10 LDA score of 0.44. More significant markers were identified at the genus level. *Morganella* and *Salmonella*, both from the *Enterobacteriaceae* family, were identified as having a significant impact on distinguishing the ASB group, with log10 LDA scores of 3.28 and 3.15, respectively, and both were more abundant in the ASB group. At the species level, *E. coli* was identified as having a minor influence on the distinction between the Wo-ASB and ASB groups, with a log10 LDA score of 0.4.

Multivariate linear regression analysis reveals association of enterobacteria with bacteriuria across different taxonomic levels

The VM is associated with gestational age. *Lactobacillus* abundance generally increases in late-term pregnancies and a dysbiotic VM is associated with higher rates of pre-term delivery[33,41]. In our study, women were sampled at their initial screen and at delivery, with bacteriuria samples evidencing significantly earlier gestational age compared to those in the Wo-ASB group (Table 1). We investigated both the effect of earlier sampling, to account for this covariate, as well as the association of ASB samples with earlier delivery, by using a multivariate linear regression model.

This analysis revealed that gestational age influenced the association between the abundance of different taxa within the VM and the occurrence of bacteriuria (Table 2 and Supplemental Table S2). In a multivariate model adjusted for gestational age and time of sampling (before or during delivery), several taxa were significantly associated with bacteriuria. Enterobacterales, the family Enterobacteriaceae, along with *E. coli*, displayed significant associations with bacteriuria (Table 2). Burkholderiales, Hyphomicrobiales, and Microbacteriaceae showed significant correlations at higher taxonomic levels, without specific genera or species being consistently associated with bacteriuria. Models adjusting for both group and gestational age indicated that *Lactobacillales* was not the primary factor in bacteriuria. These findings are consistent with the diversity analyses, revealing that VM composition, affected by gestational age, is associated with the elevated presence of potential pathogens. A distinct microbial signature involving Enterobacterales and particularly *E. coli* is evident.

E. coli Genome Assembly and Annotation

Based on diversity analyses results and given that *E. coli* is a predominant etiology of UTIs and newborn infections, we specifically targeted this species for further investigation. Among the 1,950 samples analyzed, 102 were selected which had ≥ 30,000 *E. coli*-mapping reads, representing ~ 2X genome coverage. Assembly of prefiltered *E. coli* reads for each sample, by SPAdes, resulted in 72 *E. coli* MAGs from an average of 1.43×10^6 reads (IQR 3.4×10^6 – 7.4×10^6). Samples that failed assembly were not analyzed further. We analyzed the genomic annotation CSV table produced by Prokka, along with an IQ-TREE file representing the phylogenetic relationships among the strains.

The median genome size for the 72 *E. coli* MAGs is 5.1 Mb (IQR 4.9–5.3 Mb), which aligns with the type strain *E. coli* genome size (4.6 Mb for *E. coli*K12)[64]. Seven MAGs had a genome size < 2 Mb. The N50 values had a mean of 80 kb (SD +/- 69 kb), indicating that the assemblies vary in quality, with many genomes well-represented by relatively long contigs. N90 values show a mean length of 14.8 kb (SD +/- 16.7 kb), suggesting that smaller contigs also contribute significantly to the assemblies. CheckM analysis reported median completeness of assemblies of 99.6% (IQR 98.6–99.6) and contamination levels of median 0.93% (IQR 0.25–12.25), with 74% (53/72) exhibiting contamination levels below 10% (Supplemental Table S3). The median number of unique genes per genome was 3,500 (IQR 3,400-3,500). Annotations were found for a mean of 1,708 ORFs per genome. For virulence associated genes (VAGs) involved in *E. coli* pathogenicity we confirmed their species assignment using BLAST against the NCBI Core Nucleotide DB for MAGs with > 10% contamination. All analyzed virulence-associated genes showed *E. coli* as the top hit, with > 95% gene identity and e-values of $< 10^{-50}$. Additional descriptive statistics and quality metrics for the assembled genomes are provided in the supplemental tables (Supplemental Table S3 and S4) and figures (Supplemental Figure S2).

Molecular typing of *E. coli* revealed a genomic signature of "ExPEC" strains within the vaginal compartment

Among the 72 *E. coli* MAGs, the predominant *E. coli* phylogroups identified were A, B1, B2, D, and F, covering five of the seven main groups, with C and E absent. Detailed analysis revealed that 62% (39/63) of the phylogrouped strains belonged to B2, the most represented group, followed by equal prevalence of 17.5% (11/63) for both A and D phylogroups. In contrast, B1 and F phylogroups were minor, each represented by only one strain (1.7%, 1/63). The sequence types (ST) 1193 and 131 from B2 were the most common, representing 17% (8/46) and 15% (7/46) of the strains with assigned STs, respectively. Additional B2-associated STs, such as 95, 127, and 73, along with ST 69 from phylogroup D, were also identified. Collectively, strains within these lineages have already been described as "ExPEC"[65]. In contrast, phylogroup A, typically consisting of strains commensal to the digestive

Features	Group + GA + Time of sampling			
	Beta	SE	SZ	q-value (group)
o_Burkholderiales	-0.002	0.001	-1.644	0.144
o_Enterobacterales	-0.077	0.029	-2.644	**0.021**
o_Hyphomicrobiales	-0.026	0.007	-3.788	**0.001**
o_Lactobacillales	0.089	0.080	1.119	0.330
o_Micrococcales	-0.001	0.000	-2.207	0.056
f_Burkholderiaceae	-0.001	0.001	-1.422	0.234
f_Enterobacteriaceae	-0.077	0.029	-2.666	**0.024**
f_Microbacteriaceae	-0.001	0.000	-2.207	0.056
g_Escherichia	-0.057	0.022	-2.556	**0.03**
g_Microbacterium	-0.001	0.000	-2.207	0.056
g_Variovorax	0.000	0.000	-1.747	0.122
s_Escherichia_coli	-0.057	0.022	-2.556	**0.01**

Table 2. Multivariate Linear Regression Analysis of Vaginal Microbiota Taxa Associated with Bacteriuria. The model adjusts for group (ASB or Wo-ASB), gestational age (GA), and time of vaginal sampling (before or during delivery). For each microbiota taxon, the table reports the regression coefficient (Beta), standard error (SE), standardized coefficient (SZ), and the adjusted p-value for multiple comparison (q-value).

tract, was represented by various STs (Table 3). These findings underscore the significant genetic diversity of *E. coli* within this ecosystem, predominantly comprised of strains implicated in various extra-intestinal pathologies. Among the sequenced strains, nine were found in the VM of pregnant women during ASB condition, seven of which were successfully typed and linked to phylogroup B2. These included STs 1193 (3/7), 95 (2/7), 73 (1/7), and 131 (1/7) (Fig. 4). Most of these strains, however, were found in the VM of subjects labeled as "other," indicating those not categorized into the predefined groups ASB, Wo-ASB, or HG.

We investigated the prevalence of genes associated with the pathogenicity of certain *E. coli* strains, focusing on those associated with adhesion to the urothelium (*fimH* and *papG*), iron uptake (*fyuA*, *iutA*, *fepA*), toxin production (*hlyA*, *hlyB*, *hlyD*, *sat*), and the synthesis of the capsular structure (*kpsM*, *kpsT*, *ompT*) (Fig. 4). Detailed information on these virulence associated genes (VAGs), including their functions and frequencies, is provided in supplemental table S5. A high frequency of VAGs involved in iron uptake, with *fyuA* present in 91% of genomes and *fepA* in 88% was noted. The *fimH* gene, associated with adhesion, was also highly prevalent, detected in 90% of the analyzed MAGs. Additionally, we observed a substantial prevalence of the genes *kpsM* (72%) and *kpsT* (40%). These VAGs encode proteins from the ABC transporter superfamily, essential for synthesizing the *E. coli* capsule and critical virulence factors in K1 strains implicated in neonatal meningitis[66].

AMR profiles of *E. coli* in the VM of pregnant women

Analysis of the resistance profile of the 72 *E. coli* MAGs revealed that 49 strains (68%) harbor at least one gene conferring phenotypic resistance to β-lactams, predominantly mediated by bla_{TEM} genes (64%, 46/72), which encode Ambler class A β-lactamases (Table 3). The $bla_{\text{TEM-1b}}$gene, a plasmid-encoded penicillinase, was particularly prevalent, detected in 56% of the strains. This gene imparts phenotypic resistance to penicillin, such as amoxicillin, which is commonly utilized as first-line treatments during pregnancy[6,67]. Predominantly, bla_{TEM-1b} was found in MAGs from phylogroup B2, specifically ST1193 *fimH64* and ST131 *fimH41* (Fig. 4). These MAGs also harbored resistances genes associated with phenotypic resistant to multiple antibiotics, including fluoroquinolones, with all ST1193 harboring mutations conferring resistance to ciprofloxacin. Furthermore, variable detection of genes for resistance to macrolides, tetracyclines, aminoglycosides and sulfonamides was also observed (Supplemental Table S7). The incidence of extended-spectrum β-lactamase (ESBL) was relatively low, identified in only four MAGs from phylogroups A and B2, specifically ST1312 *fimH198* and ST131 *fimH30*,

Phylogroup	ST (*n*/*N*)	ARGs
A (11/63) 17.5%	ST540 (1/11)	*ND*
	ST10 (1/11)	$bla_{TEM}(3/3)$
	ST34 (1/11)	
	ST227 (1/11)	
	ST1312 (1/11)	$bla_{CTX-M}(1/1)$
	Unknown (6/11)	$bla_{CTX-M}(2/6)$ $bla_{CMY-48}(1/6)$ $bla_{TEM}(6/6)$
B1 (1/63) 1.7%	ST 56 (1/1)	*ND*
B2 (39/63) 62%	ST1193 (8/39)	$bla_{TEM}(6/8)$
	ST131 (7/39)	$bla_{CTX-M}(1/7)$ $bla_{OXA-1}(1/7)$ $bla_{TEM}(6/7)$
	ST127 (4/39)	$bla_{TEM}(4/4)$
	ST73 (4/39)	$bla_{TEM}(3/4)$
	ST95 (3/39)	$bla_{TEM}(1/3)$
	ST12 (2/39)	$bla_{TEM}(2/2)$
	ST404 (1/39)	$bla_{TEM}(1/1)$
	ST420 (1/39)	*ND*
	ST681 (1/39)	
	ST91 (1/39)	
	ST2015 (1/39)	
	Unknown (6/39)	$bla_{TEM}(4/6)$ $bla_{OXA-1}(1/6)$ $bla_{CTX-M}(1/6)$
C (1/63) 1.7%	ST88 (1/1)	$bla_{TEM}(1/1)$
D (11/63) 17.5%	ST69 (5/9)	$bla_{TEM}(8/11)$
	ST3268 1/9)	
	Unknown (5/11)	
Unknown (9/72) 12.5%	-	$bla_{TEM}(1/9)$
		$bla_{CFE}(1/9)$

Table 3. Overview of *E. coli* Phylogroups, sequence type (ST), and prevalence of Antimicrobial Resistance genes (ARGs). ND: No AMR genes Detected. ARGs: Antimicrobial Resistance Genes.

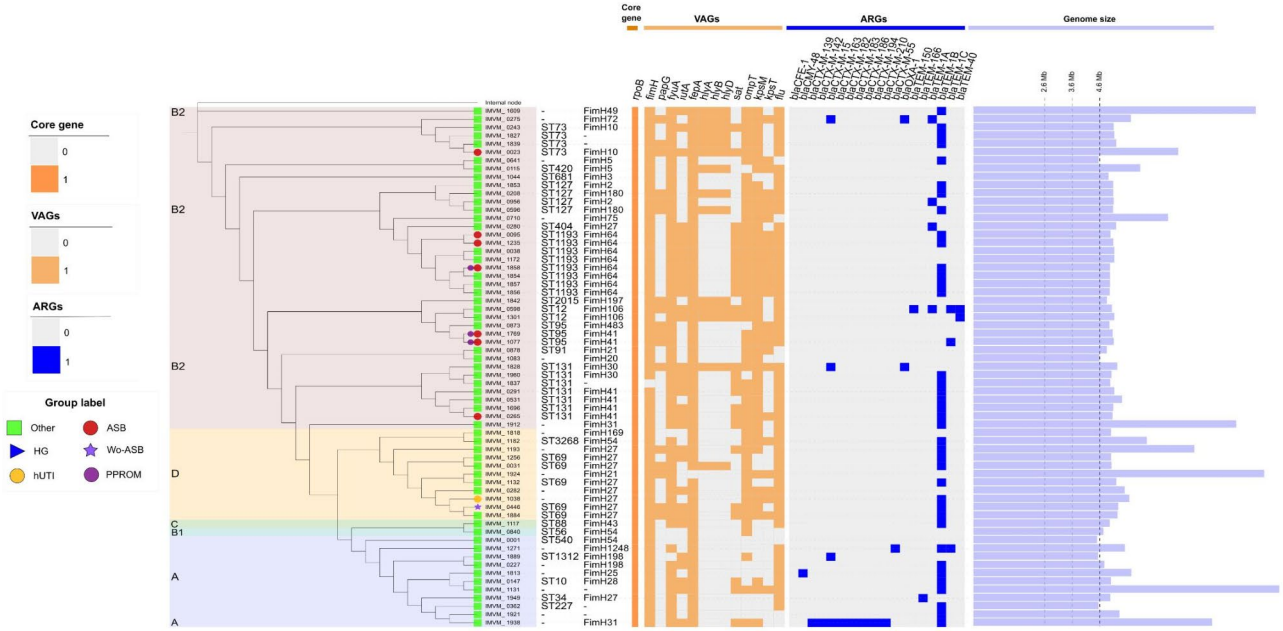

Fig. 4. Phylogenetic Tree and Characteristics of *E. coli* MAGs, within the VM of Pegnant Women. Phylogenetic tree, generated from IQ-TREE analysis based on core genome alignments of 63 of 72 *E. coli* MAGs. MAGs lacking typing information (9/72) were excluded to improve the clarity of the graph. From left to right, annotations show typing information, presence or absence of virulence-associated genes (VAGs), antimicrobial resistance genes (ARGs) and genome size. Group affiliations and preterm premature rupture of membranes (PPROM) status are indicated on the left. ASB: Asymptomatic bacteriuria; Wo-ASB: Women without asymptomatic bacteriuria during pregnancy; HG: Healthy women without any medical history or antibiotic use during pregnancy; hUTI: Pregnant women with a history of UTIs; Other: Samples from women not included in HG, ASB orWo-ASB groups. A detailed figure that includes all the 72 *E. coli* MAGs, with an additional quality marker (N50), is available as supplemental data (Supplemental Figure S3). All the molecular typing results for each MAG are detailed in supplementary table S6.

respectively. In all cases, this high-level resistance to β-lactams was attributed to the $bla_{\mathrm{CTX-M}}$ genes (Table 3 and Supplemental Table S8).

Microbial diversity associated with the presence of *E. coli* within the vaginal compartment

Microbial diversity of the vaginal samples for which an *E. coli* genome had been assembled were analyzed (group: EC) and compared with those from the group defined as healthy (HG) (Fig. 5). The presence of *E. coli* within the vaginal environment was associated with a different microbial community characterized not only with a marked decrease in *Lactobacillus* species, particularly *L. crispatus*, but also with a significant increase abundance in other Enterobacterales genera, such as *Citrobacter and Enterobacter* ($p\text{-value} = 1.42 \times 10^{-5} and\ 2.1 \times 10^{-5} respectively$) (Fig. 5a and b). Furthermore, the comparison of diversity indices also revealed considerable differences (Fig. 5c and d). Faith's PD index showed significant increase in intra-sample PD within the EC group ($p\text{-value} < 0.0001$), which was, notably, not captured when comparing the Shannon index ($p\text{-value} = 0.40$) (Fig. 5c). Analyses and comparison of the BC and w-UniFrac dissimilarity matrices, using PERMANOVA, across these two groups (EC and HG) also revealed significant differences with p-value of 0.001 for both matrices.

Discussion

Vaginal microbiota composition and changes during ASB

Although asymptomatic, ASB during pregnancy poses significant risks due to its potential to progress to severe infections, thus necessitating systematic screening and antibiotic treatment to prevent unfavorable pregnancy outcomes[2,36]. Recent insights into a gut reservoir for UPEC, along with the recognition that these uropathogens can ascend from the gut to colonize the UT, have highlighted the importance of exploring the vaginal microbiota's role in such conditions[19,20]. A primary objective of this study was to investigate the differences in the VM of pregnant women experiencing ASB compared to women without bacteriuria (Wo-ASB) and a group of healthy women with no medical history who had not received antibiotics during pregnancy (HG).

Our analysis reveals compositional variations in the VM during bacteriuria across different taxonomical levels compared to those included in the Wo-ASB and HG groups. These differences were effectively captured using Faith's PD and w-UniFrac for α-diversity and β-diversity analyses, respectively, which both factor phylogenetic differences into diversity scores[60,61]. This is especially valuable in the VM ecosystem, which is dominated by closely related, and often interchangeable, species within the *Lactobacillus* genus. Differences in microbial diversity observed in ASB samples appears to be partly influenced by variations in gestational age at the time of

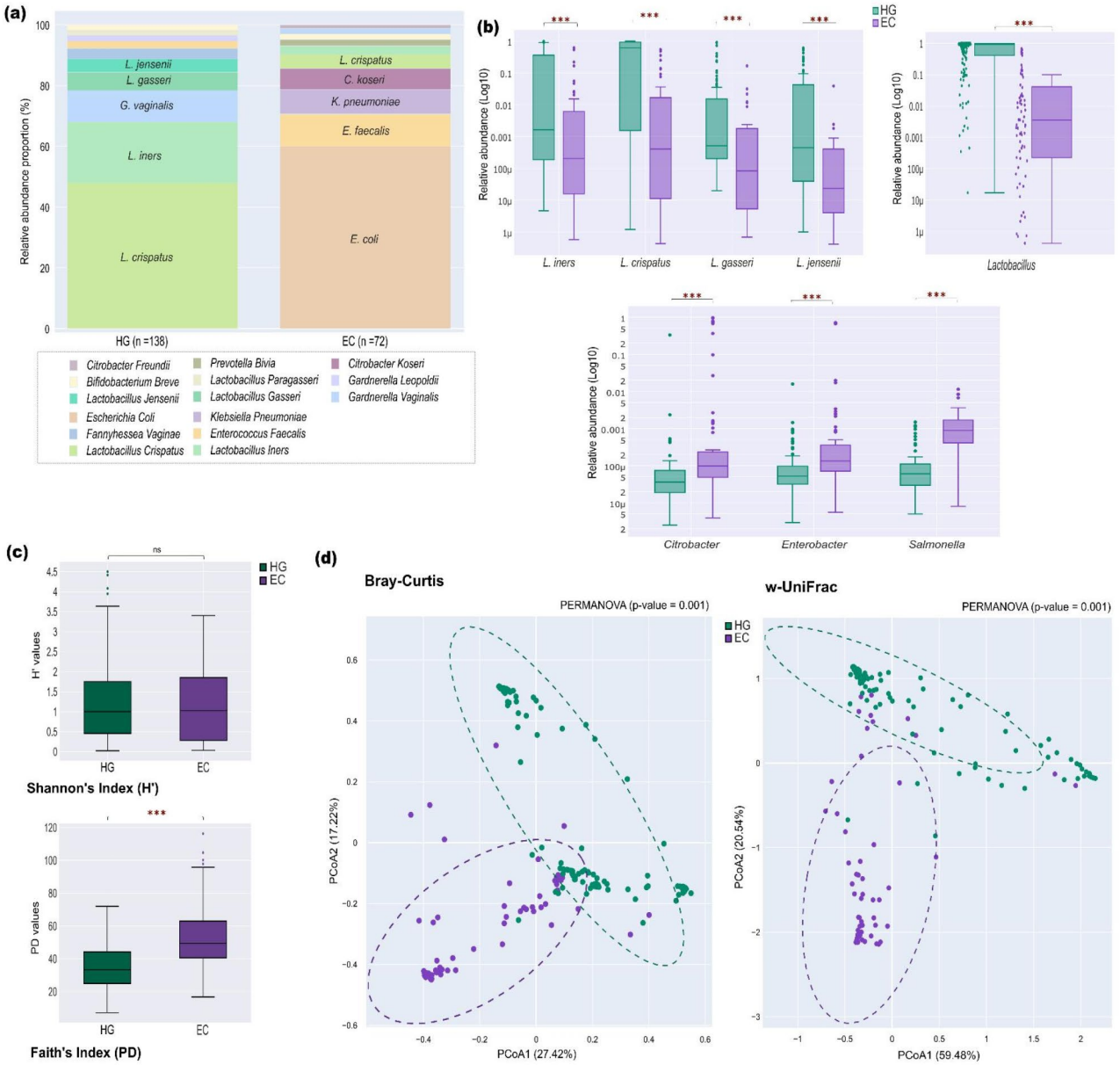

Fig. 5. Comparative Analysis of Microbial Diversity Between EC and HG Groups. The EC group includes 72 samples with assembled *E. coli* genomes, compared against 132 samples from the HG group, identified as healthy women. (**a**) Stacked bar chart illustrating the percentage abundance of the 10 predominant species among the EC and the HG group. (**b**) Box plot illustrating the relative abundances of the four most prevalent *Lactobacillus* species and the overall *Lactobacillus* genus, and key Enterobacteriaceae genera with significant differences. (**c**) Box plots illustrating the distribution of Shannon (H') and Faith's phylogenetic diversity (PD) indices across EC and HG group. The bounds of the box represent the first and third quartiles, center line indicates the median, and the whiskers extend to 1.5 times the interquartile range from the quartiles. ***$p < 0.001$; **ns**: not significant ($p > 0.05$). (**d**) Principal Coordinates Analysis from the Bray-Curtis and weighted-UniFrac matrices, showing the first two principal components (PC1 and PC2) for a 2D projection. Each point represents an individual sample, colored by group (EC in purple, HG in green). The colored ellipses depict the 95% confidence intervals for each group.

sampling and pregnancy length. A well-established correlation exists between increasing *Lactobacillus* species abundance and advancing gestational age, particularly during the third trimester[33,68]. Additionally, a disruption in the VM has been largely reported being associated with preterm birth. Women in our study with bacteriuria had a shorter pregnancy durations compared to the Wo-ASB group and were subsequently sampled at earlier timepoints. These confounding factors, while not unique to our study, complicates the interpretation of observed microbial diversity differences. Indeed, the multivariate model accounting with gestational age did not explain *Lactobacillus* variation in the genus, whereas a model with pregnancy length yielded a significant association

(*p-value = 0.02*). Furthermore, an increase abundance of Enterobacterales, and particularly *E. coli* persisted despite controlling for these covariates. Thus, while variations in biodiversity cannot be directly attributed to bacteriuria, our findings suggest the presence of opportunistic pathogens, including *E. coli*, within an altered microbial ecosystem and associated with gestational age and pregnancy length.

Over the last two decades, culture-based studies have established a significant link between vaginal health and UTIs in both pregnant and non-pregnant women, demonstrating that a dysbiotic VM increases the risk of UTIs[35,38–40]. Magruder et al. (2019) highlighted a gut-UT axis by identifying a digestive reservoir of uropathogens that colonize the UT. In our study, we noted a marked increase of gut-associated species, particularly Enterobacterales, with several species forming part of a cluster (CAC III), and *E. coli*, a predominant etiology of UTIs. These findings support the VM as an important element in the pathophysiology of bacteriuria, pointing to a broader gut-urogenital axis.

Lactobacillus species serve as a critical factor in maintaining vaginal health[24,35]. Our correlation analysis supports these findings (Fig. 1c), indicating that several H_2O_2-producing *Lactobacillus* species, co-abundant within a correlation cluster (CAC II), are pivotal in maintaining vaginal health. These lactobacilli, particularly *L. crispatus*, exert a suppressive impact on anaerobes and opportunistic pathogens. These results align with the previous modules described by Lebeer et al. (2023), showing negative correlations between lactobacilli and anaerobic genera such as *Gardnerella* and *Prevotella*[69]. These CACs, although distinct from the Community State Types (CST) described by Ravel et al. (2011), provide additional information on the interactions between various microbial species or groups. Building on the established benefits of lactobacilli, these results approach the ecological concept of "keystone species" or "keystone module", defined by the essential role that a bacterium or group can play by its presence ("presence-to-trait") or abundance ("abundance-to-trait")[70,71]. These observations will help further our understanding of the underlying mechanisms and their impact on maintaining the stability and health of this ecosystem. A case-in-point is the correlation of *L. iners* with anaerobic species, supporting the hypothesis that *L. iners*is associated with a more diverse vaginal microbiota which could favor infiltration of anaerobic species[30,31].

E. coli characteristics in VM during pregnancy and their clinical implications

Earlier studies have reported an increase in *E. coli* within the VM as associated with BV and a decrease in *Lactobacillus* species[39,72]. This pattern suggests that the vagina may also serve as a reservoir for uropathogenic strains[36]. Genomic assembly of *E. coli* species within the VM in this study revealed a significant overrepresentation of phylogroups B2 and D, collectively accounting for 79% of the 63 classified strains, with phylogroup B2 alone comprising 62%. The high prevalence of these phylogroups, is concerning due to their association with a wide spectrum of extra-intestinal infections, including UTIs, sepsis, and invasive infections in neonates[11]. We observed that ST1193 *fimH64* is the predominant strain within phylogroup B2, followed by ST131 *fimH30* and *fimH*41, both of which are recognized as part of the pandemic "ExPEC" lineages[73]. The ST1193 lineage, in particular, is under increased surveillance due to its multidrug resistance, high virulence, and rapid global spread. Recent findings by Malaure et al. (2024) have further highlighted ST1193's role in early-onset neonatal infections, emphasizing the strain's high virulence and its impact on newborn outcomes[15]. Currently reported strains within this lineage display a broad spectrum of antibiotic resistance[73,74]. This includes variable resistance to beta-lactams and macrolides, as well as consistent resistance to ciprofloxacin, which were confirmed in our study (Supplemental Table S7). These findings underscore the need for heightened vigilance, due to the potential for clinical infection as well as the risk of these strains spreading within neonatal care units. The identification of these "ExPEC" genomic signatures within the VM of pregnant women, who may not have experienced infectious events during pregnancy, highlights their robust colonizing capabilities across diverse environments. The high prevalence of the *fimH* gene in 90% of *E. coli* MAGs, coding for type 1 fimbriae, known to play a critical role in adhesion to vaginal epithelial cells and also recognized as a key factor in urothelial adhesion and gut colonization, supports this hypothesis[21,75].

Investigating the microbial diversity within the samples from which an *E. coli* genome has been assembled has shown common signatures to those reported by culture-based studies[39,72]. Although our analysis were specifically targeted to a narrow selection of samples dominated by a species not typically prevalent in the general population, they reveal that the presence of *E. coli* correlates with a substantial disruption in microbial diversity. This includes the depletion of H_2O_2-producing *Lactobacillus* species, with an associated increase in phylogenetical diversity (Fig. 5), suggesting a substantial differences in the physicochemical vaginal environment. Combined these observations suggest that H_2O_2-producing *Lactobacillus* species could have a protective effect against *E. coli* colonization, highlighting the potential benefits of probiotics in preventing of such condition. Further studies are needed to explore the interactions between these species and the underlying mechanisms involved. In summary, we suggest that the genomic ability of these strains to colonize diverse environments, coupled with the disruption of a protective vaginal environment, may contributed to the presence of pathogenic *E. coli* strains in the vaginal compartment. As such, future work should take into account the high probability of a link between the VM, ASB and UTIs.

Challenges and future directions

While this study highlights important modifications in the VM associated with ASB, some limitations will require further investigation. Although statistical methods reveal a number of significant differences among groups, the small sample size merits further validation with a larger cohort, centered on ASB, to further confirm results. The unique microbial signatures at the moment of bacteriuria could be refined with a targeted cohort that includes control samples from the same individuals in their non-affected state. Genomic assembly and phylotyping from metagenomic data are a promising means of further exploiting metagenomic data yet remains challenging and often impossible for less abundant community members. Addition complications include the

potential for assembling multiple strains of the same species in the same sample, along with the complexity of resolving hypervariable and repetitive regions[76]. Our approach effectively determined whether the strains identified in the VM were described as pathogenic but could be enhanced in the future by the incorporation of both clinical isolation and long-read sequencing. Given the increasing consensus in the literature pointing to the protective nature of lactobacilli, there is a need to develop experimental models to better mimic the vaginal environment in order to validate and explore observed interactions between the observed clusters of H_2O_2-producing *Lactobacillus* species and various pathogenic strains. Such studies could lead to targeted, non-antibiotic strategies to maintain or restore a healthy microbial balance, essential for preventing infections and enhancing maternal and neonatal health.

In conclusion, our study broadens the understanding of microbial community resilience by demonstrating how decrease in a core group of *Lactobacillus* species is linked to colonization by diverse species, including opportunistic pathogens that could impact pregnancy outcomes. Our results highlight a genomic signature that supports the vaginal environment as a potential reservoir for "ExPEC" strains, which may cause complications for both mother and child during pregnancy.

Data availability

The human-filtered metagenomic sequencing data used for analysis is available in the European Nucleotide Archive (ENA) repository: Project accession PRJEB77434, and files accessions ERS20601047 - ERS20603005.

Received: 24 July 2024; Accepted: 14 October 2024
Published online: 26 October 2024

References

1. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
2. Ansaldi, Y., Martinez, T. & Weber, B. Urinary tract infections in pregnancy. *Clin. Microbiol. Infect.* **S1198743X22004311** https://doi.org/10.1016/j.cmi.2022.08.015 (2022).
3. Nicolle, L. E. et al. Clinical practice Guideline for the management of Asymptomatic Bacteriuria: 2019 update by the infectious diseases Society of Americaa. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciy1121 (2019).
4. Kalinderi, K., Delkos, D., Kalinderis, M., Athanasiadis, A. & Kalogiannidis, I. Urinary tract infection during pregnancy: Current concepts on a common multifaceted problem. *J. Obstet. Gynaecol.* **38**, 448–453 (2018).
5. Wingert, A. et al. Asymptomatic bacteriuria in pregnancy: Systematic reviews of screening and treatment effectiveness and patient preferences. *BMJ Open.* **9**, e021347 (2019).
6. Corrales, M., Corrales-Acosta, E. & Corrales-Riveros, J. G. Which antibiotic for urinary tract infections in pregnancy? A literature review of International guidelines. *JCM*. **11**, 7226 (2022).
7. Farfour, E. et al. Antimicrobial Resistance in Enterobacterales recovered from urinary tract infections in France. *Pathogens*. **11**, 356 (2022).
8. Dunne, M. W., Aronin, S. I., Yu, K. C., Watts, J. A. & Gupta, V. A multicenter analysis of trends in resistance in urinary Enterobacterales isolates from ambulatory patients in the United States: 2011–2020. *BMC Infect. Dis.* **22**, 194 (2022).
9. Belete, M. A. & Saravanan, M. A. Systematic review on drug resistant urinary tract infection among pregnant women in developing countries in Africa and Asia; 2005–2016. *IDR*. **13**, 1465–1477 (2020).
10. Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. Pathogenic Escherichia coli. *Nat. Rev. Microbiol.* **2**, 123–140 (2004).
11. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic Escherichia coli. *Nat. Rev. Microbiol.* **19**, 37–54 (2021).
12. Clermont, O., Christenson, J. K., Denamur, E. & Gordon, D. M. The C lermont *E scherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **5**, 58–65 (2013).
13. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. & Clermont, O. ClermonTyping: An easy-to-use and accurate in silico method for Escherichia Genus strain phylotyping. *Microb. Genomics* **4**, (2018).
14. Sannes, M. R., Kuskowski, M. A., Owens, K., Gajewski, A. & Johnson, J. R. Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia and uninfected control subjects. *J. INFECT. DIS.* **190**, 2121–2128 (2004).
15. Malaure, C. et al. Early-onset infection caused by *Escherichia coli* sequence type 1193 in late Preterm and full-term neonates. *Emerg. Infect. Dis.* **30**, 20–28 (2024).
16. Brzuszkiewicz, E. et al. How to become a uropathogen: Comparative genomic analysis of extraintestinal pathogenic Escherichia coli strains. *Proc. Natl. Acad. Sci. U S A*. **103**, 12879–12884 (2006).
17. Schreiber, H. L. et al. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. *Sci. Transl. Med.* **9**, eaaf1283 (2017).
18. Tamadonfar, K. O., Omattage, N. S., Spaulding, C. N. & Hultgren, S. J. Reaching the end of the line: Urinary tract infections. *Microbiol. Spectr.* **7**, 7.3.17. (2019).
19. Meštrović, T. et al. The role of gut, Vaginal, and urinary microbiome in urinary tract infections: From bench to Bedside. *Diagnostics*. **11**, 7 (2020).
20. Magruder, M. et al. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. *Nat. Commun.* **10**, 5521 (2019).
21. Brannon, J. R. et al. Invasion of vaginal epithelial cells by uropathogenic Escherichia coli. *Nat. Commun.* **11**, 2803 (2020).
22. Zhu, B., Tao, Z., Edupuganti, L., Serrano, M. G. & Buck, G. A. Roles of the microbiota of the female reproductive tract in gynecological and reproductive health. *Microbiol. Mol. Biol. Rev.* **86**, e00181–e00121 (2022).
23. France, M., Alizadeh, M., Brown, S., Ma, B. & Ravel, J. Towards a deeper understanding of the vaginal microbiota. *Nat. Microbiol.* **7**, 367–378 (2022).
24. Ma, B., Forney, L. J. & Ravel, J. Vaginal microbiome: Rethinking health and disease. *Annu. Rev. Microbiol.* **66**, 371–389 (2012).
25. Ng, S. et al. Large-scale characterisation of the pregnancy vaginal microbiome and sialidase activity in a low-risk Chinese population. *Npj Biofilms Microbiomes*. **7**, 89 (2021).
26. Ravel, J. et al. Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. U S A*. **108**, 4680–4687 (2011).
27. France, M. T. et al. VALENCIA: A nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome*. **8**, 166 (2020).
28. Spear, G. T. et al. Human α-amylase Present in Lower-Genital-Tract Mucosal fluid processes glycogen to support vaginal colonization by Lactobacillus. *J. Infect. Dis.* **210**, 1019–1028 (2014).
29. Chee, W. J. Y., Chew, S. Y. & Than, L. T. L. Vaginal microbiota and the potential of Lactobacillus derivatives in maintaining vaginal health. *Microb. Cell. Fact.* **19**, 203 (2020).

30. Zheng, N., Guo, R., Wang, J., Zhou, W. & Ling, Z. Contribution of Lactobacillus iners to Vaginal Health and diseases: A systematic review. *Front. Cell. Infect. Microbiol.* **11**, 792787 (2021).
31. Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. Lactobacillus iners: Friend or foe? *Trends Microbiol.* **25**, 182–191 (2017).
32. Chen, X., Lu, Y., Chen, T. & Li, R. The female vaginal microbiome in Health and bacterial vaginosis. *Front. Cell. Infect. Microbiol.* **11**, 631972 (2021).
33. Romero, R. et al. The vaginal microbiota of pregnant women varies with gestational age, maternal age, and parity. *Microbiol. Spectr.* **11**, e03429–e03422 (2023).
34. Serrano, M. G. et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. *Nat. Med.* **25**, 1001–1011 (2019).
35. Stapleton, A. E. The vaginal microbiota and urinary tract infection. *Microbiol. Spectr.* **4**, 4.6.37. (2016).
36. Lewis, A. L. & Gilbert, N. M. Roles of the vagina and the vaginal microbiota in urinary tract infection: Evidence from clinical correlations and experimental models. *GMS Infect. Dis.*, **8:Doc02** (2020). https://doi.org/10.3205/ID000046
37. Dominoni, M. et al. Microbiota ecosystem in recurrent cystitis and the immunological microenvironment of urothelium. *Healthcare*. **11**, 525 (2023).
38. Hillebrand, L., Harmanli, O. H., Whiteman, V. & Khandelwal, M. Urinary tract infections in pregnant women with bacterial vaginosis. *Am. J. Obstet. Gynecol.* **186**, 916–917 (2002).
39. Gupta, K. et al. Inverse Association of H2O2-Producing Lactobacilli and Vaginal Escherichia coli colonization in women with recurrent urinary tract infections. *J. Infect. Dis.* **178**, 446–450 (1998).
40. Stamey, T. A. & Sexton, C. C. The role of Vaginal colonization with Enterobacteriaceae in recurrent urinary infections. *J. Urol.* **113**, 214–217 (1975).
41. Baud, A. et al. Microbial diversity in the vaginal microbiota and its link to pregnancy outcomes. *Sci. Rep.* **13**, 9061 (2023).
42. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
43. Baud, A. & Kennedy, S. P. Targeted metagenomic databases provide improved analysis of Microbiota samples. *Microorganisms*. **12**, 135 (2024).
44. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*. **10**, e65088 (2021).
45. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic Survey Data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
48. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: Rapid and scalable correlation estimation for compositional data. *Bioinformatics*. **35**, 1064–1066 (2019).
49. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
50. Letunic, I. & Bork Interactive tree of life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **gkae268**https://doi.org/10.1093/nar/gkae268 (2024).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
52. Bankevich, A. et al. SPAdes: A New Genome Assembly Algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
53. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. **30**, 2068–2069 (2014).
54. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
55. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
56. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
57. Larsen, M. V. et al. Multilocus sequence typing of total-genome-sequenced Bacteria. *J. Clin. Microbiol.* **50**, 1355–1361 (2012).
58. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genomics* **8**, (2022).
59. Shannon, C. & Weaver, W. The Mathematical Theory of Communication.
60. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
61. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. UniFrac: An effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172 (2011).
62. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
63. Lithgow, K. V. et al. Protease activities of vaginal Porphyromonas species disrupt coagulation and extracellular matrix in the cervicovaginal niche. *Npj Biofilms Microbiomes*. **8**, 8 (2022).
64. Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, 1453–1462 (1997).
65. Manges, A. R. et al. Global extraintestinal pathogenic Escherichia coli (ExPEC) lineages. *Clin. Microbiol. Rev.* **32**, e00135–e00118 (2019).
66. Pigeon, R. P. & Silver, R. P. Topological and mutational analysis of KpsM, the hydrophobic component of the ABC-transporter involved in the export of polysialic acid in *Escherichia coli* K1. *Mol. Microbiol.* **14**, 871–881 (1994).
67. Salverda, M. L. M., De Visser, J. A. G. M. & Barlow, M. Natural evolution of TEM-1 β-lactamase: Experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **34**, 1015–1036 (2010).
68. Kervinen, K. et al. Parity and gestational age are associated with vaginal microbiota composition in term and late term pregnancies. *eBioMedicine*. **81**, 104107 (2022).
69. Lebeer, S. et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nat. Microbiol.* **8**, 2183–2195 (2023).
70. Power, M. E. et al. Challenges in the Quest for keystones. *BioScience*. **46**, 609–620 (1996).
71. Amit, G. & Bashan, A. Top-down identification of keystone taxa in the microbiome. *Nat. Commun.* **14**, 3951 (2023).
72. Navas-Nacher, E. L. et al. Relatedness of *Escherichia coli* Colonizing women longitudinally. *Mol. Urol.* **5**, 31–36 (2001).
73. Pitout, J. D. D., Peirano, G., Chen, L., DeVinney, R. & Matsumura, Y. Escherichia coli ST1193: Following in the footsteps of E. Coli ST131. *Antimicrob. Agents Chemother.* **66**, e00511–e00522 (2022).
74. Wyrsch, E. R., Bushell, R. N., Marenda, M. S., Browning, G. F. & Djordjevic, S. P. Global phylogeny and F virulence plasmid carriage in Pandemic Escherichia coli ST1193. *Microbiol. Spectr.* **10**, e02554–e02522 (2022).
75. Spaulding, C. N. et al. Selective depletion of uropathogenic E. Coli from the gut by a FimH antagonist. *Nature*. **546**, 528–532 (2017).
76. Lapidus, A. L. & Korobeynikov, A. I. Metagenomic data assembly – The way of decoding unknown microorganisms. *Front. Microbiol.* **12**, 613791 (2021).

Acknowledgements
This work was supported by the Programme d'Investissements d'avenir and bpifrance (Structuring R&D Project for Competitiveness – PSPC): # DOS0053477 SUB et DOS0053473 ARWe also acknowledge the financial support provided by the University Hospital of Rouen Normandy through the "Bourse Année Recherche (BAR)" grant for the year 2023-2024. A full list of contributors to this study, collectively called the InSPIRE Consortium, can be found in the accompanying Supplementary Information files.

Author contributions
N.B. contributed to conception, performed computational analysis, prepared all figures and drafted the manuscript. T.G.A.V. developed the genome assembly Snakemake pipeline. L.L. provided expert review and editing. S.K. conceived the project, provided the dataset, supervised the work and edited the manuscript. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-76438-2.

Correspondence and requests for materials should be addressed to S.P.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.