

ARTICLE OPEN



Shotgun sequencing of the vaginal microbiome reveals both a species and functional potential signature of preterm birth

Conor Feehily^{1,2}, David Crosby³, Calum J. Walsh^{1,2}, Elaine M. Lawton^{1,2}, Shane Higgins³, Fionnuala M. McAuliffe^{3,4}✉ and Paul D. Cotter^{1,2,4}✉

An association between the vaginal microbiota and preterm birth (PTB) has been reported in several research studies. Population shifts from high proportions of lactobacilli to mixed species communities, as seen with bacterial vaginosis, have been linked to a twofold increased risk of PTB. Despite the increasing number of studies using next-generation sequencing technologies, primarily involving 16S rRNA-based approaches, to investigate the vaginal microbiota during pregnancy, no distinct microbial signature has been associated with PTB. Shotgun metagenomic sequencing offers a powerful tool to reveal community structures and their gene functions at a far greater resolution than amplicon sequencing. In this study, we employ shotgun metagenomic sequencing to compare the vaginal microbiota of women at high risk of preterm birth ($n = 35$) vs. a low-risk control group ($n = 14$). Although microbial diversity and richness did not differ between groups, there were significant differences in terms of individual species. In particular, *Lactobacillus crispatus* was associated with samples from a full-term pregnancy, whereas one community state-type was associated with samples from preterm pregnancies. Furthermore, by predicting gene functions, the functional potential of the preterm microbiota was different from that of full-term equivalent. Taken together, we observed a discrete structural and functional difference in the microbial composition of the vagina in women who deliver preterm. Importance: with an estimated 15 million cases annually, spontaneous preterm birth (PTB) is the leading cause of death in infants under the age of five years. The ability to accurately identify pregnancies at risk of spontaneous PTB is therefore of utmost importance. However, no single cause is attributable. Microbial infection is a known risk factor, yet the role of vaginal microbes is poorly understood. Using high-resolution DNA-sequencing techniques, we investigate the microbial communities present in the vaginal tracts of women deemed high risk for PTB. We confirm that *Lactobacillus crispatus* is strongly linked to full-term pregnancies, whereas other microbial communities associate with PTB. Importantly, we show that the specific functions of the microbes present in PTB samples differs from FTB samples, highlighting the power of our sequencing approach. This information enables us to begin understanding the specific microbial traits that may be influencing PTB, beyond the presence or absence of microbial taxa.

npj Biofilms and Microbiomes (2020)6:50; <https://doi.org/10.1038/s41522-020-00162-8>

INTRODUCTION

Preterm birth (PTB) is the leading cause of mortality in infants under the age of 5 years¹. The rate of spontaneous PTB, defined as the onset of labour and subsequent delivery before 37 weeks of gestation, is reported to be at one in ten pregnancies (~15 million PTB annually) worldwide^{2,3}. In addition to increased mortality, prematurity is associated with significant morbidity in terms of chronic lung disease, increased rates of neurodevelopmental delay, and long-term health problems^{4–6}. The economic burden through healthcare costs related to prematurity is estimated to be \$26 billion in the United States alone⁷.

Several studies have investigated links between maternal health status including diet⁸, smoking⁹, obesity¹⁰, stress¹¹, age¹², and previous history of obstetric complications¹³ as indicators of PTB. Microbial infection, specifically urinary tract infections, vaginitis, bacterial vaginosis (BV), and even periodontal disease have also been associated with an increased risk of spontaneous preterm delivery^{14–16}. The composition of the microbiota of the mother, and in particular the mother's vaginal microbiome, has more recently been linked to spontaneous PTB³. In general, the vaginal microbiota is stable throughout pregnancy with the dominant species rarely changing^{17,18}; however, biogeography and ethnicity

can play a considerable role in determining the apparent 'normal' microbiota¹⁹. A study of 90 women, mainly of African American descent, found that there was no association between the vaginal microbiota and PTB²⁰. In contrast, DiGiulio et al.²¹ found that changes in the vaginal populations of Caucasian women were correlated with PTB. Further studies have provided additional evidence that race is a key factor when identifying patterns linking PTB with vaginal microbiota composition^{22,23}.

The common means of categorizing the vaginal microbiome has been to group them into community state types (CSTs). CSTs dominated by *Lactobacillus* species have generally been associated with positive health states for the woman²⁴. In contrast, CST-4, which is reduced in lactobacilli and contains an increased abundance of mixed species, has been associated with poorer health outcomes. These mixed species, including *Gardnerella vaginalis* (assigned as *Bifidobacterium vaginae* in the Genome Taxonomy Database (GTDB)), *Atopobium vaginae* (assigned as *Fannyhessea vaginae* in the GTDB), *Mobiluncus* sp., *Prevotella* sp., and others, are most often associated with BV. Multiple studies have reported a twofold increased risk for PTB in women suffering from BV^{25,26}. Despite this, large cohort studies have reported that CST-4 can also be present in a high percentage of the healthy

¹Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Ireland. ²APC Microbiome Ireland, University College Cork, Cork, Ireland. ³UCD Perinatal Research Centre, Obstetrics and Gynaecology, School of Medicine, University College Dublin, National Maternity Hospital, Dublin 2, Ireland. ⁴These authors jointly supervised this work: Fionnuala M. McAuliffe, Paul D. Cotter. ✉email: fionnuala.mcauliffe@ucd.ie; paul.cotter@teagasc.ie

Table 1. Descriptive statistics of study participants.

	Risk		Control	
	risk_PTBA ^a (n = 7)	risk_FTBB ^b (n = 28)	non-risk PTB ^c (n = 1)	non-risk FTB ^d (n = 13)
Age (years)	31.43	33.14	42	34.08
BMI (kg/m ²)	26.07	25.41	32.02	23.6
Gestational age at delivery (weeks)	33.06	39.71	36.7	39.67
<28 Weeks	1	0	0	0
28–32 Weeks	0	0	0	0
32–37 Weeks	6	0	1	1
Ethnicity				
White Irish	6	24	1	12
White Caucasian	1	2	0	0
Asian	0	1	0	0
African	0	1	0	1
Smokers	1	3	0	0
Birth weight	2143.57	3666.43	2840	3602.38
LLETZ	2	9	0	0
Previous PTB	6	19	0	0
ABX within previous 6 months	3	6	0	2

^arisk_PTBA preterm birth refers to those women who delivered prior to 37 weeks' gestation.
^brisk_FTBB refers to those women with risk factors for preterm birth but who delivered full term.
^cnon-risk PTB refers to low-risk women who delivered prior to 37 weeks' gestation.
^dnon-risk FTB refers to low-risk women who delivered at term.

population with no symptoms of BV²⁷. Indeed, *B. vaginalis* has also been frequently isolated from asymptomatic women²⁸.

Determining a role for the vaginal microbiota in influencing pregnancy outcome has been limited by the ability to fully differentiate taxonomic groups. This resolution may prove important if discrete communities, species, or strains prove to represent a risk factor. In light of this, it is important to consider that the use of standard 16S rRNA targeted amplicon sequencing has been reported to under-report certain subgroups of *B. vaginalis*²⁹, while targeting the *cpn60* gene, has shown an improvement in terms of identifying *Bifidobacterium*³⁰. Shotgun metagenomics presents an opportunity to overcome many of the limitations of amplicon sequencing, with the additional benefit of understanding the functional potential of a community. An increasing number of studies have now used this approach to elucidate the vaginal microbial communities. In particular, functional distinctions among different metagenomic assembled *B. vaginalis* isolated from the same sample³¹ and an ability to sub-speciate important taxa³² provide examples for the power of this approach.

In a cohort of women with a predisposed risk for PTB, we aimed to use shotgun metagenomics to distinguish any vaginal microbial signatures different to a low-risk control group.

RESULTS

Participant data

A total of 57 participants with singleton pregnancy were successfully recruited over the study period, 20 at low risk of PTB and 37 with risk factors for PTB. Multiple swabs were collected

from some women depending on clinic visit frequency, accounting for 89 total samples. To control for multiple sampling of some woman and different trimester timepoints, we focused our analysis on single samples for each woman in the second trimester of pregnancy ($n = 49$). Of these, 8 pregnancies ultimately were preterm (PTB) and the remaining 41 were full-term birth (FTB). In addition, for the 35 participants initially considered at risk of PTB, 7 women delivered before 37 weeks' gestation and were grouped into the PTB group (risk_PTBA). The remainder were grouped into a risk but full-term group (risk_FTBB). For the 14 women at low risk of PTB, 13 delivered at term (no-risk_FTBB). The single sample from this control group that delivered prior to 37 weeks was subsequently provided a distinct group of no-risk_PTBA. Within the PTB samples, six deliveries were late preterm (32–37 weeks), one was a moderate preterm (31.7 weeks), and one was a preterm at 26 weeks (Table 1). None of the women had preterm premature rupture of membranes. There were no significant differences in the age, race, body mass index, or smoking status of the participants by either the grouping of risk_PTBA, risk_FTBB, no-risk_FTBB, and no-risk_PTBA or the FTB and PTB grouping. Patient demographics are outlined in Table 1. There was a lower birth weight for infants from the PTB compared to FTB (2230.62 g vs. 3646.12 g, respectively; $p < 0.001$ Student's *t*-test).

Taxonomic analysis

Following quality filtering, a mean of 828,950 high-quality microbial sequencing reads were obtained per sample. The median number of observed species for either the FTB or PTB groups was 168 and 185.5, respectively (Fig. 1a), with the highest number of species observed at 15 for a risk_PTBA sample. There was no significant difference in α -diversity measure between either the FTB or PTB groups (Fig. 1a); however, there was an increased diversity of the risk_FTBB group compared to the no-risk_FTBB group using both Shannon and Simpson measures (Fig. 1b; $p = 0.01$ and $p = 0.003$, respectively). In terms of β -diversity, the risk_PTBA communities were significantly dissimilar from both the no-risk_FTBB and risk_FTBB samples (Fig. 1b; $p = 0.02$).

Taxonomic classification revealed that lactobacilli were dominant across all groups at 59.13% mean relative abundance (Fig. 2a), with a greater proportion observed in full-term pregnancies (61.86%) compared to preterm (45.11%). This dominance was greatest in the no-risk_FTBB group (71.32%) where both *Lactobacillus crispatus* (50.86%) and *Lactobacillus iners* (21.27%) were dominant. Within the risk_FTBB group, *L. crispatus* (25.83%), *L. iners* (14.83%), *Lactobacillus gasseri* (11.70%), and *L. gasseri* A (7.30%) were the dominant species detected. *L. crispatus* was never >0.03% relative abundance in the risk_PTBA group. In addition, *L. gasseri* was only detected in two samples in the risk_PTBA group, accounting for just 0.01% mean relative abundance. Although *L. iners* (30.82%) accounted for a greater proportion of the mean population in the risk_PTBA group compared to both FTB groups, this increase was not statistically significant. Due to the large variation in both the abundance and number of species observed between samples, *L. crispatus* ($p < 0.001$), *L. gasseri* ($p = 0.028$), and *Bifidobacterium breve* ($p = 0.036$) were the only species that significantly differed between full-term and preterm groups, with a higher respective mean relative abundance (Fig. 2b). A significant negative correlation was observed between PTB samples and both *L. crispatus* and *L. gasseri* (p -value 0.03 and 0.03, respectively), and after correcting for multiple comparisons, these q -values were below 0.25 (Fig. 2c, Table 2, and Supplementary Table 2). Overall, the dominance of BV-associated bacteria was low across all samples. Using the GTDB database, *B. vaginalis* (this database assigns *G. vaginalis* as *B. vaginalis*) was assigned as multiple sub-species. The two most abundant BV-associated species across all samples were *B. vaginalis* and *B. vaginalis_G*; however, neither were significantly increased in either

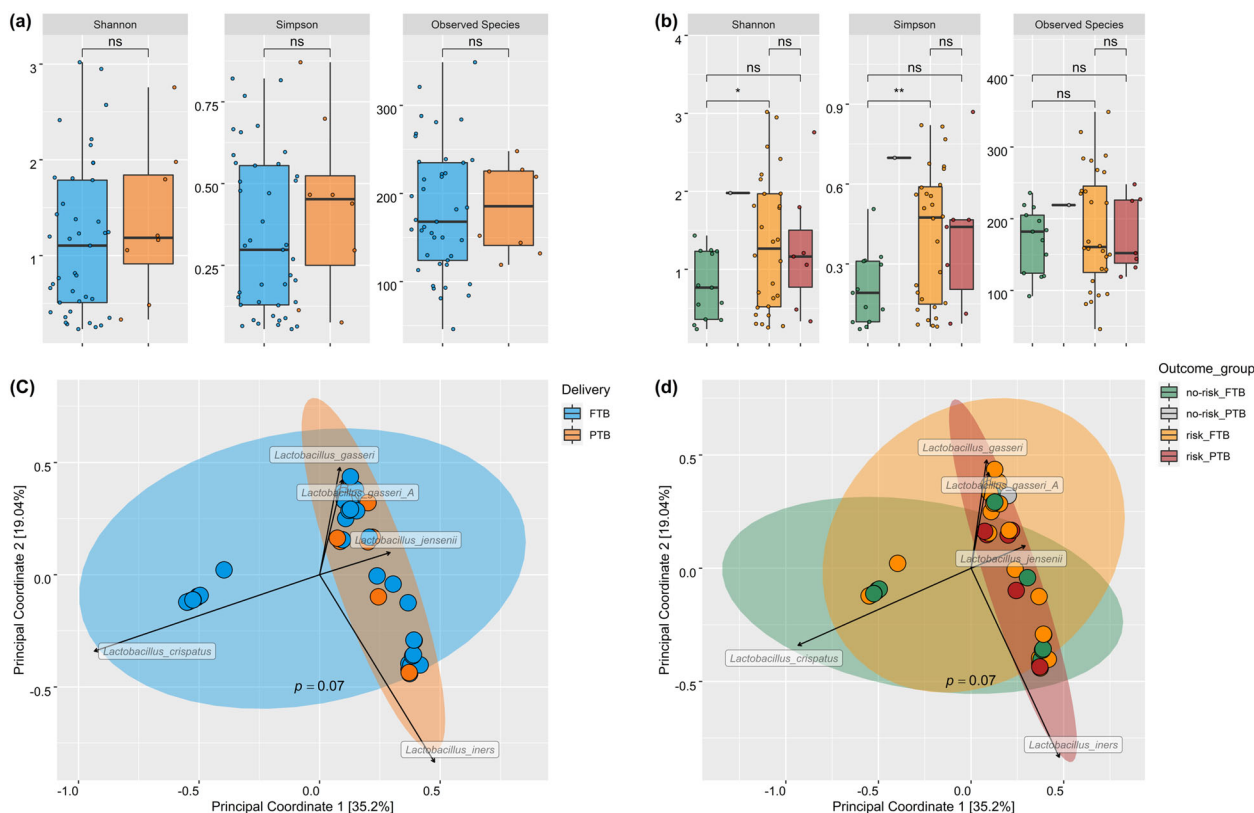


Fig. 1 Species diversity between study groups. α -Diversity comparison as measured by Shannon and Simpson index, including the total observed species for either delivery outcome (a) or risk grouping (b). Significant differences as calculated by Wilcoxon's test are noted with * $p < 0.05$ or ** $p < 0.02$. The species level community dissimilarity as measured by Bray–Curtis and visualized using PCoA for either delivery outcome (c) or risk grouping (d). Ellipses are generated using `stat_ellipse` function in R. Significance of group dissimilarity as calculated by PERMANOVA is identified by the given p -value. The top five lactobacilli across all samples are labelled to highlight drivers of variation for clusters with black arrows to indicate the directionality.

FTB or PTB samples (Fig. 2b). A significant correlation with PTB samples was observed for *B. vaginalis*_D, _E, and _F (Fig. 2c, Table 2, and Supplementary Table 2; q -value 0.097, 0.006, and 0.002, respectively). In addition, *F. vaginae* and *F. vaginae_A* (this database assigns *A. vaginae* as *F. vaginae*) were positively correlated with PTB samples (q -value 0.137 and 0.008, respectively; Fig. 2c, Table 2, and Supplementary Table 2).

Resolution of the microbial species into CSTs revealed that six distinct CSTs were present across all swabs (Fig. 3). CSTs were determined by the most dominant species in a cluster as previously described³³ with the definition of CST-8 according to Brooks et al.²⁴. CST-1, dominated by *L. crispatus*, had a weak association overall with delivery outcome (Supplementary Table 1; $p = 0.09$). There was a weak association between samples from the risk_PTb groups and CST-5 (Supplementary Table 1; $p = 0.09$). There were five observations of the BV-associated CST-4, three of which were from the risk_PTb group. This CST had a positive association with samples from the risk_PTb group ($p = 0.02$). Neither CST-3 nor CST-8 were associated with any pregnancy outcome.

Functional pathways analysis

A total of 2928 gene functions were identified across all samples. Stratification of the data into the three Gene Ontology classifications of cellular component (CC), biological process (BP), and molecular functions (MFs) revealed functional diversity between the risk groupings. Both Shannon and Simpson diversity indexes determined there was a significant increase in the diversity of gene functions for both BPs and metabolic functions for PTB

samples compared to FTB (Fig. 4a). When samples were stratified into risk groupings, this difference in functional diversity was not observed between the risk_FTb and no-risk_FTb groups (Supplementary Fig. 1a). The FTB and PTB samples were also functionally dissimilar to each other by Bray–Curtis measure both in the BP and MF classes (Fig. 4b; p -value 0.038 and 0.045, respectively). From ADONIS analysis, only *L. gasseri* was significantly influencing the variation in gene function across all three categories (p -value 0.017 (MF), 0.027 (CC), and 0.013 (BP)). Stratification of the samples into the risk groupings did not show any significant differences in diversity of function (Supplementary Fig. 1b). Using multivariate analysis, 22 CCs, 154 BPs, and 299 MFs were significantly correlated to either women who had a previous PTB, large loop excisions of the transformation zone (LLETZ), were categorized with a prior risk for PTB, subgrouped to the four risk categories according to this study, or ultimately delivered preterm (Fig. 5 and Supplementary Table 3; q -value < 0.25 after multiple correction). For gene functions with significant association to preterm delivery, the MF category contained the most differential features. Among these were genes involved in '2' 3'-bisphosphoglycerate-independent phosphoglycerate mutase activity', 'receptor activity', 'aldose 1 epimerase activity', 'carboxyl or carbamoyltransferase activity', and 'copper exporting ATPase activity' (Fig. 5 and Supplementary Table 3). The top five gene functions most associated with preterm delivery within the BP category included 'respiratory electron transport chain', 'alginate biosynthetic process', 'ATP hydrolysis coupled proton transport', 'folic acid biosynthetic process', and 'glucose catabolic process' (Fig. 5 and Supplementary Table 3).

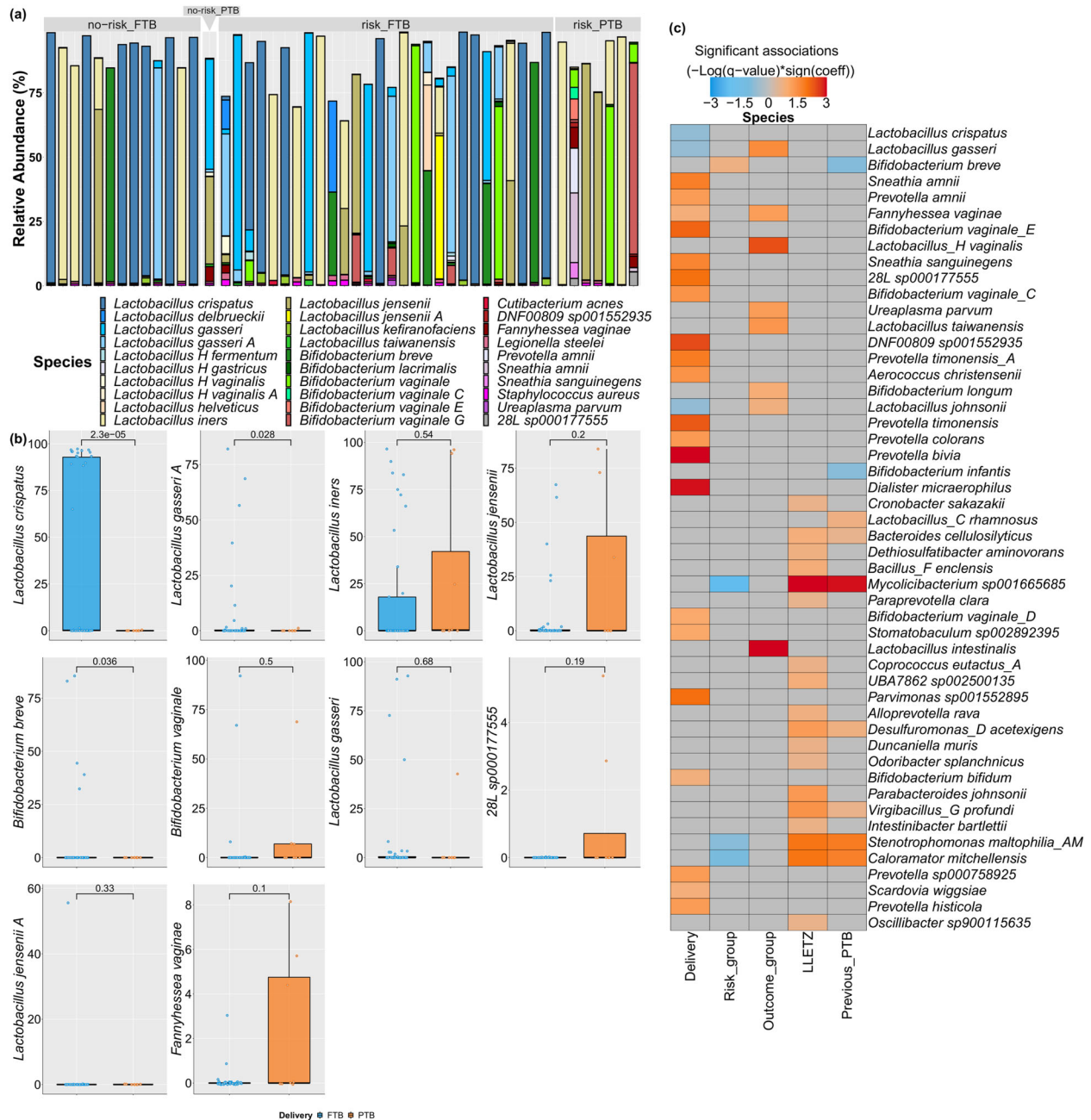


Fig. 2 Species composition across study groups. **a** The relative abundance for the top 30 species across all samples for each of the four risk groupings. **b** Comparative analysis within the delivery outcomes for the ten most abundant species across all samples. Significant differences were calculated using Student's *t*-test. **c** MaAsLiN analysis correlating species multiple metadata fixed effects. Heatmap displays the top 50 species with a significant association to either fixed effect with *q*-value < 0.25.

DISCUSSION

In summary, using a shotgun metagenomics approach, this study has confirmed the association of a vaginal microbiome dominated by *L. crispatus* with full-term pregnancies. In addition, evidence is provided that CST-4 is more associated with the vaginal microbiome from a spontaneous PTB. This study also identified functional differences in the vaginal microbiome of women who subsequently deliver preterm.

To date, numerous microbiome sequencing studies have been conducted with the aim of understanding the vaginal microbiota and its association with PTB. A limiting factor in several of these studies has been the use of targeted amplicon sequencing. In particular, it has been shown that certain key vaginal taxa can be

underrepresented or not detected, depending on the gene region targeted. Of note, BV has been determined as a risk factor for PTB, yet the detection and differentiation of *B. vaginae* (*G. vaginalis*) subtypes can be difficult with the potential to omit subtle nuances when associating vaginal communities with PTB.

In our study, we have used shotgun metagenomics to determine the vaginal communities in women at high risk of PTB. This approach has enabled a higher resolution of the species of bacteria associated with high-risk pregnancies compared to previous approaches. The relatively high microbial read depth achieved, coupled with the recently updated GTDB has revealed a richness of species that are present at low abundance (<5%; Fig. 3). With use of the GTDB classification database, several

Table 2. MaAsLiN2 multivariate correlation analysis of most abundant microbial species and sample groupings.

Species	Group	Value	Coef	stderr	N	p-Value	q-Value
<i>B. breve</i>	Previous_PTB	Yes	-1.8149	0.7393	49	0.0182	0.1366
<i>B. breve</i>	Risk_group	Risk	1.8357	0.8000	49	0.0267	0.1810
<i>B. vaginale_C</i>	Delivery	PTB	0.6202	0.2028	49	0.0038	0.0353
<i>B. vaginale_E</i>	Delivery	PTB	0.7502	0.2001	49	0.0005	0.0064
<i>DNF00809 sp001552935</i>	Delivery	PTB	0.6983	0.1710	49	0.0002	0.0032
<i>F. vaginae</i>	Outcome_group	no-risk_PTB	1.3714	0.4852	49	0.0071	0.0580
<i>F. vaginae</i>	Delivery	PTB	0.4585	0.1870	49	0.0184	0.1373
<i>L. crispatus</i>	Delivery	PTB	-1.6480	0.7347	49	0.0301	0.2030
<i>L. gasseri</i>	Outcome_group	no-risk_PTB	4.5300	1.4246	49	0.0027	0.0266
<i>L. gasseri</i>	Delivery	PTB	-1.2000	0.5492	49	0.0344	0.2284
<i>Lactobacillus taiwanensis</i>	Outcome_group	no-risk_PTB	2.6198	0.8731	49	0.0045	0.0407
<i>Lactobacillus_H vaginalis</i>	Outcome_group	no-risk_PTB	2.6170	0.6470	49	0.0002	0.0036
<i>Prevotella amnii</i>	Delivery	PTB	0.0496	0.0173	49	0.0062	0.0513
<i>Sneathia amnii</i>	Delivery	PTB	0.2996	0.0878	49	0.0014	0.0160
<i>Sneathia sanguinegens</i>	Delivery	PTB	0.5417	0.1629	49	0.0018	0.0193
<i>Ureaplasma parvum</i>	Outcome_group	no-risk_PTB	1.5612	0.5485	49	0.0067	0.0553
<i>28 L sp000177555</i>	Delivery	PTB	0.8062	0.2222	49	0.0008	0.0090

species of *B. vaginale* were detected with each determined to correlate differently with pregnancy outcome (Table 2 and Fig. 2c). Importantly, these sub-species were not the dominant *B. vaginale*, which when present (18/49 occurrences) had a mean relative abundance of 14%. In contrast, *B. vaginale_D* (9/49 occurrences), *B. vaginale_E* (8/49 occurrences), and *B. vaginale_F* (8/49 occurrences), which significantly correlated with PTB outcome, were never >0.2% relative abundance. At such low abundance, it remains to be determined whether these species can exert a meaningful biological influence within the microbiota. Nonetheless, this difference within the *B. vaginale* group highlights the importance of stratifying this species when investigating associations with vaginal health and perinatal outcome. Another important member of BV-associated bacteria, *F. vaginae* (*A. vaginae*), was identified in three of the PTB samples at relatively high abundance (4.42–8.10%), yet was rare in full-term samples at a maximum relative abundance of 3% (in 3/41 samples). This species has been shown to develop strong biofilms with *B. vaginale*^{34,35}. Previous studies identified high loads of this species as a risk for PTB^{36,37}. Of interest in our study, there was an increased relative abundance of *F. vaginae* in the control group sample that had no perceived risk of PTB, yet delivered preterm (Fig. 3).

The species *Sneathia amnii* and *Prevotella amnii* have recently been identified as emerging candidates for poor pregnancy outcomes^{38,39}. In our study, both of these were increased and had a significant association with PTB (*q*-value 0.02 and 0.05, respectively; Table 2). Although BVAB-1 (assigned UBA629 sp005465875 in GTDB) has been identified in microbiome studies as a potential problematic species in terms of vaginal health, this species was not identified in this study, perhaps reflective of the cohort studied as it has predominantly been reported in North American cohort studies^{29,40,41}. Overall, however, the BV-associated CST, CST-4, was positively associated with PTB, a finding that is in agreement with previous studies in this area^{21,42}.

Unlike previous reports, we did not see any significant differences in α -diversity between the samples from preterm and full-term pregnancies; however, there was a weak correlation between *L. crispatus* and full-term pregnancy, and an overall increased abundance of this species in samples from full-term pregnancies. This association is in agreement with several other studies and supports the concept that this species is a beneficial component of a healthy vaginal microbiome. Taken together,

these studies may suggest a role of *L. crispatus* in providing protection from PTB, particularly in a Caucasian population. However, a mechanistic basis for such a role has yet to be revealed. Nonetheless, *L. crispatus* has previously been identified as a species that dominates a healthy vaginal tract, specifically in subjects free from BV⁴³. It has also been linked to the stability of the vaginal microbiota during pregnancy⁴⁴. Given that inflammation of the uterus has been linked with PTB⁴⁵, there may be merit to assessing the ability of *L. crispatus* to provide a protective effect by suppressing inflammation through H₂O₂ signalling to control nuclear factor- κ B activity⁴⁶ or, indirectly, whereby the inhibition of BV-associated bacteria prevents the occurrence of a proinflammatory response to the vaginosis⁴⁷. Importantly, a strong correlation with proinflammatory cytokines and BV-associated bacteria was observed by Fettweis et al.³⁹ in a preterm cohort, further highlighting a potential microbial inflammatory mechanism for preterm onset. In general, lactobacilli have been regarded as a marker of a healthy vaginal microbiome and have been seen to improve pregnancy outcome even in the presence of risk associated taxa³⁷. An exception to this is *L. iners*, which has previously been shown to associate with PTB in a similar cohort⁴⁸. This observation was not repeated in the study.

Although these findings are important indications that merit repeated investigation, there are limitations to this dataset. In particular, due to the relatively low ratio of PTB samples, a much larger sample size will be necessary to investigate the link between the CSTs, species, and preterm outcome. Moreover, this study's findings are specific to a Caucasian population, whilst ethnicity has been previously shown as major confounder in describing the vaginal microbiome. Notably, Romero et al.²⁰ found that the composition of the vaginal microbiota did not differ between PTB and FTB across a cohort of predominately African American women. The difference in ethnicity in terms of vaginal microbiota and PTB has been further evidenced recently in a study that indicated that the frequency of *Lactobacillus*, *Gardnerella*, and *Ureaplasma* varies significantly between preterm and full-term deliveries within a Caucasian cohort but not within and African American cohort⁴⁹. Unfortunately, socioeconomic data was not collected as part of this work, and as has been previously reported^{50,51} could be a confounder in the different microbiome profile observed. In addition, there were women in both the risk_FT B and risk_PT B groups that had received antibiotic treatment in the 6 months prior to sample

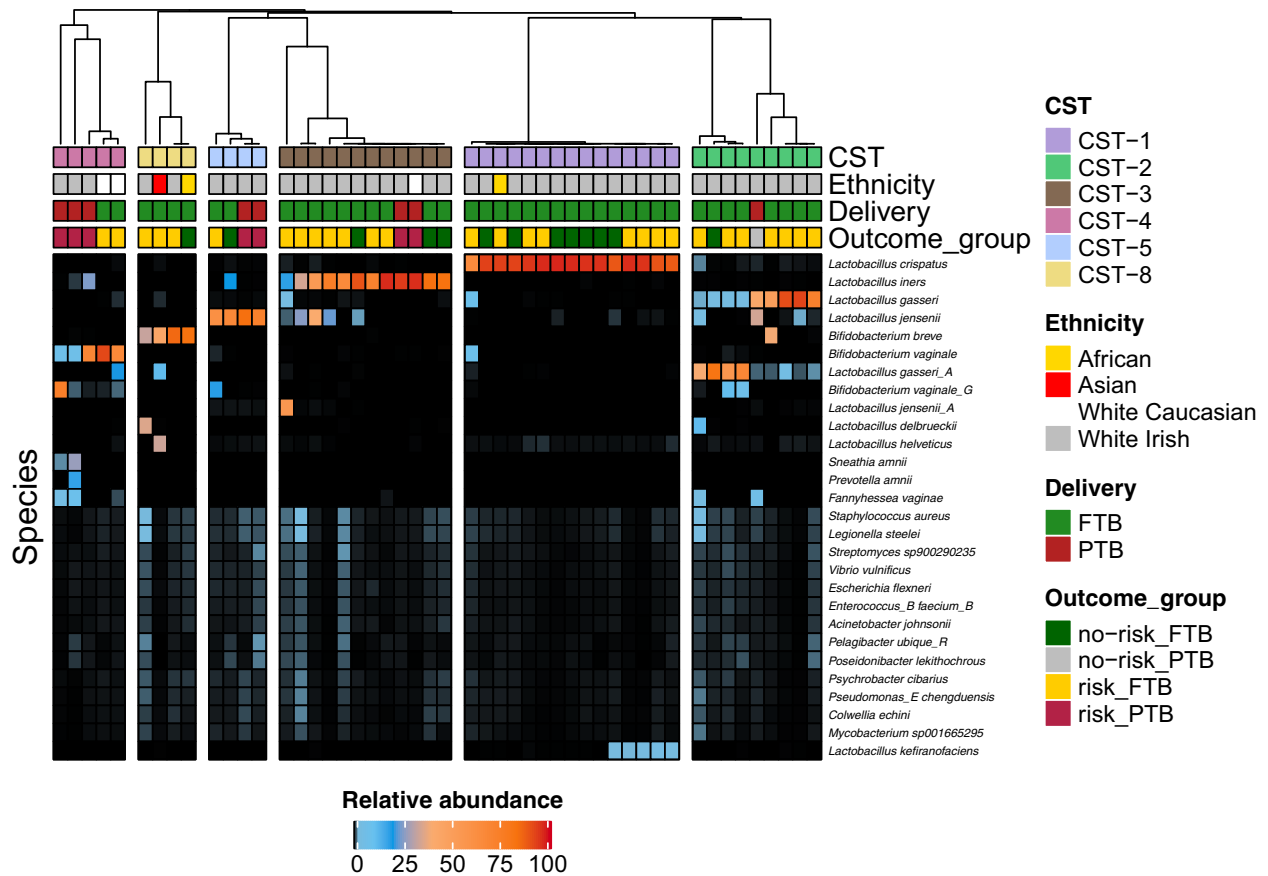


Fig. 3 Community state types of all samples. Relative heatmap intensity comparison for all species with an across-sample mean relative abundance > 0.2 , with sample clustering of similar samples by Pearson's clustering. Community state types are defined based on the most abundant species per sample and are indicated by the top bar of the figure.

collection, yet due to the low numbers it was not possible to determine the influence of this.

Although there is growing evidence for a microbiota-related role in spontaneous PTB, and indeed numerous descriptions of the importance for the microbiota with general vaginal health, there has been a limited understanding of the functional capacity of these microbial communities. As suggested by Heintz-Buschart and Wilmes⁵², a functional insight is required to develop and test hypothesis for mechanisms of action for host-microbe interactions. Indeed, the sensitivity of community analysis using functional metagenomics has already been shown to distinguish healthy populations from cases of inflammatory bowel disease^{53,54}, type 2 diabetes⁵⁵, and obesity⁵⁶. Within our preterm cohort, there was a clear distinction with respect to the functional profile of genes involved in both metabolic functions and BPs. The PTB samples had an increased abundance of functions relating to methionine/homocysteine and folate metabolism in addition to purine metabolism. Notably, maternal folate and homocysteine levels have been associated with poor pregnancy outcomes including PTB⁵⁷. Moreover, there has been reductions in the concentrations of cysteine as measured in cervico-vaginal fluid from preterm samples in a previous study⁵⁸. Taken together, it would suggest that further research into methionine/folate/cysteine homeostasis is merited in terms of a role for microbial metabolism and host interaction.

Overall, this study demonstrates the benefits of using shotgun metagenomics to understand the vaginal microbiome of women at risk of PTB. Moreover, we have shown that use of a microbial DNA enrichment kit is a feasible method to increase high-quality microbial sequencing reads to the upper limit of what has been achieved in previous studies (Supplementary Table 4) and

overcome the notable problem of host contamination^{31,59}. A distinct variation in both the taxonomic and functional potential of the preterm linked vaginal microbiome reinforces the concept of a microbiome role in PTB. However, the inability of both this study and others to determine a strict signature highlights the need for much larger controlled studies with the ability to examine confounding factors like ethnicity, biogeography, and predetermined risk factors using high-resolution, strain level analysis. This study provides evidence for the functional role of the microbiota in spontaneous PTB. Direct gene expression studies are now required if we are to fully elucidate any microbial-host interactions, which are involved in the spontaneous onset of PTB.

METHODS

Study participants and sampling

This was a prospective cohort study with institutional ethics approval by the National Maternity Hospital Research Ethics Committee and maternal written consent. Women were determined to be at risk of spontaneous PTB if they had either a history of previous spontaneous PTB, and therefore at high risk of subsequent spontaneous PTB ($n = 29$), or had two previous LLETZ ($n = 11$). High-risk participants in the study were recruited from women attending the PTB clinic at The National Maternity Hospital Dublin, Ireland. Anonymized patient data were collected from patient charts by an independent researcher and pregnancy outcome data collected from patient charts and a computerized database in the National Maternity Hospital, Dublin. Low-risk controls ($n = 14$) were women who had previously delivered a full-term baby with no prior history of PTB or LLETZ, and were attending routine antenatal care. Inclusion criteria were women over 18 years of age and pregnant. Exclusion criteria were women currently on antibiotic treatment. Following informed consent, a high vaginal swab was taken using a speculum from the posterior vaginal fornix

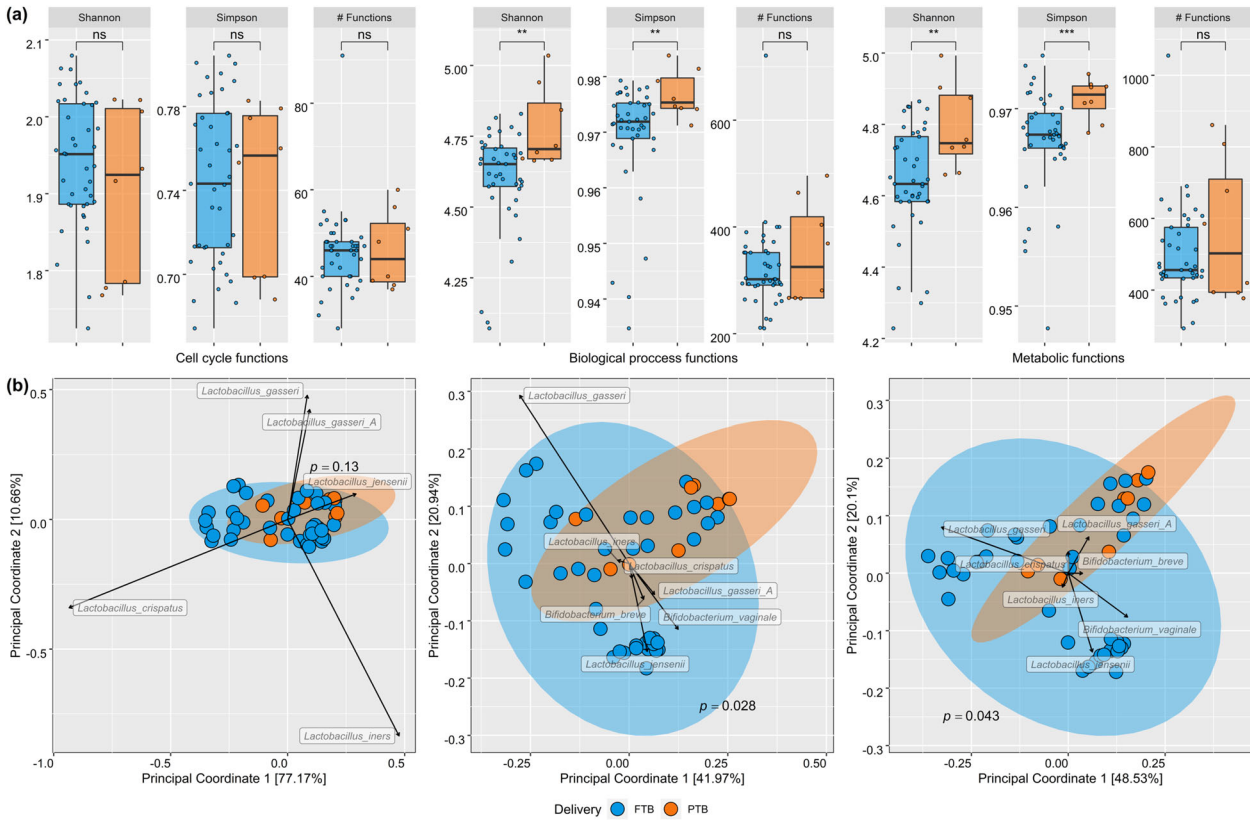


Fig. 4 Functional diversity of the vaginal microbiome. Gene functions are divided into three higher-level functional categories based on Gene Ontology classification with both α - (a) and β -diversity (b) measures of identified content shown. Significant differences are highlighted by an asterisk where p -value was <0.05 as calculated by Kruskal–Wallis. Significant differences in β -diversity were determined by PERMANOVA analysis. Ellipses are generated using `stat_ellipse` function in R. The top seven species across all samples are labelled to highlight drivers of variation for clusters with black arrows to indicate the directionality.

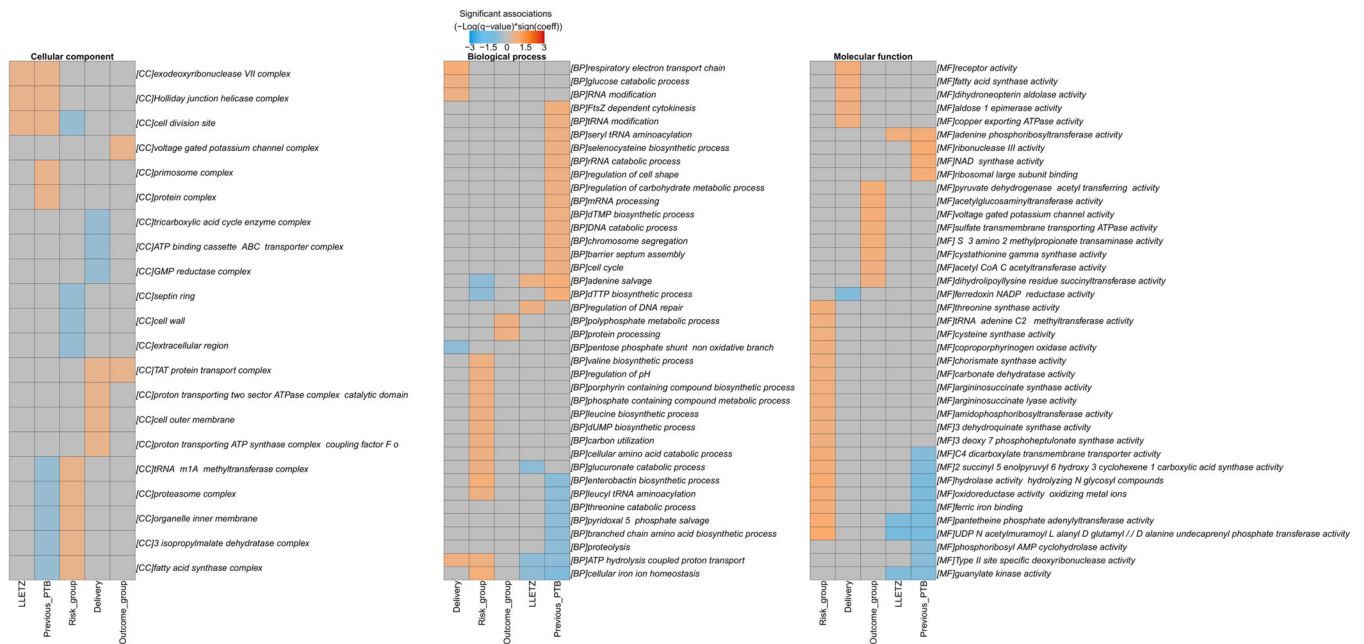


Fig. 5 Multivariate association analysis. Each heatmap presents the independent significant associations of species to grouping as determined by MaAsLiN2 with q -value <0.25 . For biological process and molecular function on the top 50 associations are presented.

and external orifice of the cervix prior to transvaginal ultrasonography using a dry cotton swab and delivered to the laboratory within 24 h.

DNA extraction/purification

One millilitre of sterile phosphate-buffered saline was added to each swab followed by rigorous vortexing for 1 min. Total microbial DNA was extracted using the MoBio PowerFood Microbial DNA isolation kit and manufacturer's instructions. Briefly, 700 µl of swab material was centrifuged and the resulting pellet was disrupted by mechanical and enzymatic treatment. DNA was eluted at 100 µl in elution buffer and samples were stored at -20°C. DNA concentrations were determined using the Qubit high sensitivity kit as per manufacturer's instructions. To reduce sequencing reads mapping to eukaryotic DNA downstream, DNA from each swab was treated with the Microbiome Enrichment kit (NEB) and purified with AMPure magnetic beads (Beckman Coulter). A negative sterile water control was included during the enrichment and carried forward through subsequent sequencing steps. Finally, all samples were normalized to 0.2 ng µl⁻¹.

Shotgun metagenomic sequencing

Libraries of DNA were prepared according to standard Illumina protocols. Briefly, DNA was sheared by heating to 55 °C for 7 min. Paired-end indexes were added and amplification occurred for 12 cycles before samples were purified with AMPure magnetic beads. DNA was quantified by Qubit dsDNA HS assay kit and the Agilent Bioanalyser 2100 with high sensitivity DNA chips before being pooled to 2 mM. Quality of the sample pool was confirmed by quantitative PCR. A sample of sterile water was processed in parallel with the DNA during library preparation to act as a negative control (Supplementary Table 5). Libraries were sequenced using 2 × 150 bp paired-end kit on the Illumina NextSeq platform.

Bioinformatic analysis

Raw sequencing data were base-called using Illumina's bcl2fastq software (v 2.19) (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html). TrimGalore (v 0.6.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), a perl wrapper for Cutadapt (v. 1.18)⁶⁰ and FastQC (v. 0.11.8) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), was used to remove adapter sequences and low-quality sequences using default parameters. Removal of human DNA contamination was performed by aligning all high-quality paired-end reads to the latest draft of the human genome (hg38) using Bowtie2 (v. 2.3.4)⁶¹. The resulting SAM files were converted to BAM format and filtered to keep only unmapped paired-end reads using SAMtools⁶². Bedtools⁶³ was used to convert the remaining reads from BAM to FASTQ format. Taxonomic assignment of paired-end reads was performed using Kraken2⁶⁴ alignment against the GTDB_54k database created by Méric et al.⁶⁵ (<https://github.com/rrwick/Metagenomics-Index-Correction>). Functional profiling was performed using the HUMAnN2 pipeline (v. 2.8.1)⁶⁶. The gene families output were renormalized as copies per million reads and regrouped according to Gene Ontology terms.

Data were visualized using both Graphpad Prism 6 and RStudio (R v.3.6.0). Heatmaps were generated using the 'ComplexHeatmap' package⁶⁷ with samples clustered using the Pearson's distance metric and columns split by *k*-means clustering to visualize CSTs, assigned based on previous reported definitions²⁴. Plots for diversity analysis were generated using the ggplot2 package (<https://cran.r-project.org/web/packages/ggplot2/index.html>). Statistical analysis was carried out in R using the vegan package (<https://cran.r-project.org/web/packages/vegan/index.html>) and RVAideMemoire (<https://cran.r-project.org/web/packages/RVAideMemoire/index.html>). Multivariate Association with Linear Models 2 (MaAsLiN2, R V.1.2.0) was used to determine independent associations of species and functions with metadata factors. A *q*-value of >0.25 was considered significant in our analysis.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Sequence data have been deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB34536.

Received: 23 April 2020; Accepted: 16 October 2020;

Published online: 12 November 2020

REFERENCES

- Liu, L. et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. *Lancet* **385**, 430–440 (2015).
- Blencowe, H. et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* **379**, 2162–2172 (2012).
- Vinturache, A. E., Gyamfi-Bannerman, C., Hwang, J., Mysorekar, I. U. & Jacobsson, B. Maternal microbiome - a pathway to preterm birth. *Semin. Fetal Neonatal Med.* **21**, 94–99 (2016).
- Khashu, M., Narayanan, M., Bhargava, S. & Osiovich, H. Perinatal outcomes associated with preterm birth at 33 to 36 weeks' gestation: a population-based cohort study. *Pediatrics* **123**, 109–113 (2009).
- Pike, K. C. & Lucas, J. S. A. Respiratory consequences of late preterm birth. *Pae-diatr. Respir. Rev.* **16**, 182–188 (2015).
- Twilhaar, E. S. et al. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: a meta-analysis and meta-regression. *JAMA Pediatr.* **172**, 361–367 (2018).
- Frey, H. A. & Klebanoff, M. A. The epidemiology, etiology, and costs of preterm birth. *Semin. Fetal Neonatal Med.* **21**, 68–73 (2016).
- Englund-Ögge, L. et al. Maternal dietary patterns and preterm delivery: results from large prospective cohort study. *BMJ* **348**, g1446 (2014).
- Moore, E., Blatt, K., Chen, A., Van Hook, J. & DeFranco, E. A. Relationship of trimester-specific smoking patterns and risk of preterm birth. *Am. J. Obstet. Gynecol.* **215**, 109–e1 (2016).
- Faucher, M. A., Hastings-Tolsma, M., Song, J. J., Willoughby, D. S. & Bader, S. G. Gestational weight gain and preterm birth in obese women: a systematic review and meta-analysis. *BJOG* **123**, 199–206 (2016).
- Dole, N. et al. Maternal stress and preterm birth. *Am. J. Epidemiol.* **157**, 14–24 (2003).
- Waldenström, U., Cnattingius, S., Vixner, L. & Norman, M. Advanced maternal age increases the risk of very preterm birth, irrespective of parity: a population-based register study. *BJOG* **124**, 1235–1244 (2017).
- Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75–84 (2008).
- Jeffcoat, M. K. et al. Periodontal infection and preterm birth: results of a prospective study. *J. Am. Dent. Assoc.* **132**, 875–880 (2001).
- Romero, R., Dey, S. K. & Fisher, S. J. Preterm labor: one syndrome, many causes. *Science* **345**, 760–765 (2014).
- Romero, R. et al. The role of inflammation and infection in preterm birth. in *Semin. Reprod. Med.* **25**, 21–39 (2007). Copyright © 2007 by Thieme Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
- Walther-Antônio, M. R. S. et al. Pregnancy's stronghold on the vaginal microbiome. *PLoS ONE* **9**, 1–10 (2014).
- Romero, R. et al. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* **2**, 1–19 (2014).
- MacIntyre, D. A. et al. The vaginal microbiome during pregnancy and the post-partum period in a European population. *Sci. Rep.* **5**, 1–9 (2015).
- Romero, R. et al. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* **2**, 1–15 (2014).
- DiGiulio, D. B. et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl Acad. Sci. USA* **112**, 11060–11065 (2015).
- Brown, R. et al. Role of the vaginal microbiome in preterm prelabour rupture of the membranes: an observational study. *Lancet* **387**, S22 (2016).
- Nelson, D. B., Shin, H., Wu, J. & Dominguez-Bello, M. G. The gestational vaginal microbiome and spontaneous preterm birth among nulliparous African American women. *Am. J. Perinatol.* **33**, 887–893 (2016).
- Brooks, J. P. et al. Changes in vaginal community state types reflect major shifts in the microbiome. *Microb. Ecol. Health Dis.* **28**, 1303265 (2017).
- Leitch, H. et al. Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *Am. J. Obstet. Gynecol.* **189**, 139–147 (2003).
- Leitch, H. & Kiss, H. Asymptomatic bacterial vaginosis and intermediate flora as risk factors for adverse pregnancy outcome. *Best. Pract. Res. Clin. Obstet. Gynaecol.* **21**, 375–390 (2007).

27. Ravel, J. et al. Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108**, 4680–4687 (2011).
28. Janulatiene, M. et al. Prevalence and distribution of *Gardnerella vaginalis* subgroups in women with and without bacterial vaginosis. *BMC Infect. Dis.* **17**, 1–9 (2017).
29. Albert, A. Y. K. et al. A study of the vaginal microbiome in healthy Canadian women utilizing cpn 60-based molecular profiling reveals distinct *Gardnerella* subgroup community state types. 1–21, <https://doi.org/10.1371/journal.pone.0135620> (2015).
30. Chaban, B. et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome* **2**, 1–12 (2014).
31. Goltsman, D. S. A. et al. Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res.* **28**, 1467–1480 (2018).
32. Serrano, M. G. et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. *Nat. Med.* **25**, 1001–1011 (2019).
33. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019).
34. Hardy, L. et al. A fruitful alliance: the synergy between *Atopobium vaginae* and *Gardnerella vaginalis* in bacterial vaginosis-associated biofilm. *Sex. Transm. Infect.* **92**, 487–491 (2016).
35. Hardy, L. et al. Unravelling the bacterial vaginosis-associated biofilm: a multiplex *Gardnerella vaginalis* and *Atopobium vaginae* fluorescence in situ hybridization assay using peptide nucleic acid probes. *PLoS ONE* **10**, 1–16 (2015).
36. Bretelle, F. et al. High *Atopobium vaginae* and *Gardnerella vaginalis* vaginal loads are associated with preterm birth. *Clin. Infect. Dis.* **60**, 860–867 (2015).
37. Elovitz, M. A. et al. Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. *Nat. Commun.* **10**, 1–8 (2019).
38. Gentile, G. L. et al. Identification of a cytopathogenic toxin from *Sneathia amnii*. *J. Bacteriol.* **202**, 1–11 (2020).
39. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019).
40. Holm, J. B. et al. Ultrahigh-throughput multiplexing and sequencing of >500-base-pair amplicon regions on the Illumina HiSeq 2500 platform. *mSystems* **4**, 1–10 (2019).
41. Holm, J. B. et al. Comparative metagenome-assembled genome analysis of *Lachnospirillum* genomospecies, formerly known as BVAB1. Preprint at <https://doi.org/10.1101/657197>, <https://www.biorxiv.org/content/10.1101/657197v1.full> (2019).
42. Freitas, A. C., Bocking, A., Hill, J. E. & Money, D. M. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome* **6**, 117 (2018).
43. Srinivasan, S. et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS ONE* **5**, e10197 (2010).
44. Verstraelen, H. et al. 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.* **9**, 116 (2009).
45. Elovitz, M. A., Wang, Z., Chien, E. K., Rychlik, D. F. & Phillippe, M. A new model for inflammation-induced preterm birth: the role of platelet-activating factor and Toll-like receptor-4. *Am. J. Pathol.* **163**, 2103–2111 (2003).
46. Voltan, S. et al. *Lactobacillus crispatus* M247-derived H₂O₂ acts as a signal transducing molecule activating peroxisome proliferator activated receptor-γ in the intestinal mucosa. *Gastroenterology* **135**, 1216–1227 (2008).
47. Mirmonsef, P. et al. The effects of commensal bacteria on innate immune responses in the female genital tract. *Am. J. Reprod. Immunol.* **65**, 190–195 (2011).
48. Kindinger, L. M. et al. The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. *Microbiome* **5**, 1–14 (2017).
49. Callahan, B. J. et al. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. *Proc. Natl Acad. Sci. USA* **114**, 9966–9971 (2017).
50. Virtanen, S. et al. Vaginal microbiota composition correlates between Pap smear microscopy and next generation sequencing and associates to socioeconomic status. *Sci. Rep.* **9**, 1–9 (2019).
51. Dunlop, A. L. et al. Stability of the vaginal, oral, and gut microbiota across pregnancy among African American women: the effect of socioeconomic status and antibiotic exposure. *PeerJ* **2019**, e8004 (2019).
52. Heintz-Buschart, A. & Wilmes, P. Human Gut Microbiome: Function Matters. *Trends Microbiol.* **26.7**, 563–574 (2018).
53. Armour, C. R., Nayfach, S., Pollard, K. S. & Sharpton, T. J. A metagenomic meta-analysis reveals functional signatures of health and disease in the human gut microbiome. *mSystems* **4**, 1–15 (2019).
54. Schirmer, M. et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* **3**, 337–346 (2017).
55. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
56. Chatelier, E. Le et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546, <https://doi.org/10.1038/nature12506> (2013).
57. Bergen, N. E. et al. Homocysteine and folate concentrations in early pregnancy and the risk of adverse pregnancy outcomes: the Generation R Study. *BJOG* **119**, 739–751, <https://doi.org/10.1111/j.1471-0528.2012.03321.x> (2012).
58. Ghartey, J., Bastek, J. A., Brown, A. G., Anglim, L. & Elovitz, M. A. Women with preterm birth have a distinct cervicovaginal metabolome. *Am. J. Obstet. Gynecol.* **212**, 776.e1–776.e12 (2015).
59. Marotz, C. A. et al. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **6**, 1–9 (2018).
60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
62. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
64. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 1–13 (2019).
65. Méric, G. et al. Correcting index databases improves metagenomic studies. *bioRxiv* (2019).
66. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
67. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

ACKNOWLEDGEMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273 and 16/SP/3827, and an Irish Research Council fellowship (EPSPD/2016/25) with Alimentary Health Ltd.

AUTHOR CONTRIBUTIONS

Conceptualization: C.F., D.C., S.H., F.M., and P.C. Ethical committee application: D.C., S.H., and F.M. Experimental investigation: C.F. and E.L. Bioinformatics and statistics: C.F., C.W., and P.C. Writing—original draft: C.F. Writing—review and editing: C.F., D.C., C.W., S.H., F.M., and P.C. Samples collection and data curation: C.F., D.C., and S.H. Figures preparation: C.F. Funding acquisition: C.F., F.M., and P.C. F.M. and P.C. are joint senior authors. All authors discussed the results and commented on the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41522-020-00162-8>.

Correspondence and requests for materials should be addressed to F.M.M. or P.D.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.