Vaginal microbiota in ethnically diverse young women who did or did not develop pelvic inflammatory disease: community-based prospective study

Sarah Kerry-Barnard,1 Liqing Zhou,2 Laura Phillips,2 Martina Furegato,2 Adam A Witney,3,4 S Tariq Sadiq,2,4 Pippa Oakeshott1

ABSTRACT

Objectives: A lactobacilli-dominated vaginal microbiome may protect against pelvic inflammatory disease (PID), but one dominated by Gardnerella species might increase susceptibility. Not all lactobacilli are equally protective. Recent research suggests that D(−) isomer lactic acid producing lactobacilli (Lactobacillus crispatus, Lactobacillus jensenii and Lactobacillus gasseri) may protect against infection with Chlamydia trachomatis, an important cause of PID. Lactobacillus iners, which produces L(+) isomer lactic acid, may be less protective. We investigated the microbiome in stored vaginal samples from participants who did or did not develop PID during the prevention of pelvic infection (POP) chlamydia screening trial.

Methods: Long-read 16S rRNA gene nanopore sequencing was used on baseline vaginal samples (one per participant) from all 37 women who subsequently developed clinically diagnosed PID during 12-month follow-up, and 111 frequency matched controls who did not, matched on four possible risk factors for PID: age <2 years, ≥20 black ethnicity versus other ethnicity, chlamydia positive versus negative at baseline and ≥2 sexual partners in the previous year versus 0–1 partners.

Results: Samples from 106 women (median age 19 years, 40% black ethnicity, 22% chlamydia positive, 54% reporting multiple partners) were suitable for analysis. Three main taxonomic clusters were identified dominated by L iners, L crispatus and Gardnerella vaginalis. There was no association between a more diverse, G vaginalis dominated microbiome and subsequent PID, although increased Shannon diversity was associated with black ethnicity (p=0.002) and bacterial vaginosis (diagnosed by Gram stain p<0.0001). Women who developed PID had similar relative abundance of protective D(−) isomer lactic acid producing lactobacilli to women without PID, but numbers of PID cases were small.

Conclusions: In the first-ever community-based prospective study of PID, there was no clear association between the vaginal microbiome and subsequent development of PID. Future studies using serial samples may identify vaginal microbial communities that may predispose to PID.

BACKGROUND

Pelvic inflammatory disease (PID) is an infection-induced inflammation of the female upper reproductive tract that can lead to infertility and ectopic pregnancy. Chlamydia trachomatis and Neisseria gonorrhoeae cause up to 30% of PID, but in many cases, the aetiology is polymicrobial or unclear.1 In addition, we do not know why only some women with genital C. trachomatis infection develop upper tract infection. A healthy, low diversity, vaginal microbiome dominated by lactobacilli may protect against disease.2–5 However, a high diversity microaerophilic/anaerobic microbiome with fewer lactobacilli (as may be found in bacterial vaginosis often with Gardnerella vaginosis and/or Atopobium vaginae) may increase susceptibility to STIs by causing cervicovaginal barrier disruption and epithelial portals of entry for ascending infection.6–8

Not all lactobacilli are equal.3 A recent study suggested that lactobacilli that produce mainly D(−) isomer lactic acid (Lactobacillus crispatus, L. jensenii and L. gasseri) may be associated with long-term protection against C. trachomatis infection, possibly by reducing epithelial cell recycling.7 By contrast, a vaginal microbiome dominated by L iners, which mainly produces L(+) isomer lactic acid, may be associated with increased susceptibility to infection.7–9

A recent systematic review called for longitudinal studies of vaginal microbiota and infections to lay the foundation for possible prevention and treatment strategies.3 We used stored baseline self-taken vaginal swabs from a cohort of 2357 ethnically diverse young female students recruited from public areas at 22 London colleges to the prevention of pelvic infection (POP) chlamydia screening trial in 2004–2006.10 Previously, 16S rRNA sequencing of 20 samples using a long-read PacBio sequencing platform had confirmed adequate DNA integrity in these samples.11 Conventional short-read 16S rRNA gene sequencing is constrained in its ability to comprehensively speciate pathogens within a sample because of the limited 16S variable regions that are usually sequenced. Long-read sequencing platforms such as PacBio and Oxford Nanopore Technologies’ MinION allow for full sequencing of the 16S rRNA gene for microbiome studies.12 13

We used long-read 16S rRNA gene nanopore sequencing to investigate bacterial communities in 37 women who developed clinical PID within the following year, and 111 frequency matched control women who did not. Since D(−) isomer lactic acid producing lactobacilli may protect against C. trachomatis,4 7 a major cause of PID,
2004-6
2357 women completed sexual health questionnaires and provided baseline vaginal samples which were stored and subsequently tested for C. trachomatis by PCR and bacterial vaginosis by Gram stain.

2005-7
2225 (94%) women were followed up after a year by questionnaire and/or medical record search for incidence of clinical pelvic inflammatory disease (PID) over 12 months diagnosed by 3 blinded sexual health physicians

2019-20
Sequencing of baseline samples from all 37 women subsequently diagnosed with PID in the next year, and 111 control women without PID matched on age, ethnicity, chlamydia status at baseline and number of sexual partners

Figure 1  POPI vaginal microbiota study flow chart. PID, pelvic inflammatory disease; POPI, prevention of pelvic infection.

we also explored the relative abundance of D(−) isomer lactic acid producing lactobacilli in women with and without subsequent PID, and in women with and without concurrent C. trachomatis infection or bacterial vaginosis (BV).14

METHODS
Study samples
One hundred and forty-eight frozen, baseline vaginal samples (one sample per participant) were retrieved from all 37 women who subsequently developed clinically diagnosed PID in the next year and 111 controls with no PID, as part of the POPI trial (figure 1). At recruitment, participants had provided self-taken vaginal swabs in the college toilets, rolled the swab over a glass slide for future Gram stain analysis for BV (Nugent score ≥7) and placed the swab in a tube of Apta transport medium (Gen Probe), which was frozen within 24 hours. Controls (three per index case of PID) were frequency matched on possible risk factors for PID15: age <20 versus ≥20, black ethnicity versus other ethnicity, chlamydia positive versus negative at baseline and ≥2 sexual partners in the previous year versus 0–1 partners and bacterial vaginosis by Gram stain versus no bacterial vaginosis. P values were adjusted for multiple comparisons using the Benjamini-Hochberg method.

The relative abundance of the dominant species was explored (including a combination of D(−) isomer lactic acid producing lactobacillic L. crispatus, L. jensenii and L. gasseri), for three clinical diagnoses: pelvic inflammatory disease, bacterial vaginosis or C. trachomatis. Finally, we investigated beta diversity using non-metric multidimensional scaling (NMDS) of Bray-Curtis distances.

The study was constrained by the number of women diagnosed with PID. We estimated that a sample size of 148 women (37 who developed PID and 111 who did not develop PID) would detect a significant difference (p<0.05, power 80%) if vaginal samples from 35% of women with PID had Gardnerella vaginalis16 versus ≤11% of women without PID. (We used G. vaginalis as a surrogate for bacterial vaginosis.)

Excluded samples
An initial heatmap representation of the relative abundances (online supplemental figure 1) showed four major clusters, one of which was dominated by the plant pathogen Burkholderia gladioli. Although a rare cause of opportunistic infection in humans,20 the presence of B. gladioli is unusual, and the Burkholderia genus has been associated with contamination in other microbiome studies.21 Further investigation showed that the majority of samples with high levels of B. gladioli were collected within a specific time period at the end of the study (online supplemental figure 2), suggesting a systematic contamination within the samples collected at the time. A similar plot for L. crispatus showed no chronological bias (data not shown). In light of this, all 42 samples containing more than 1% B. gladioli were removed from further analysis. (This included 14 samples with PID, 13 with BV and 9 with C. trachomatis.) Inclusion or exclusion of these samples had little effect on the conclusions of this study.

RESULTS
Study population
Samples from 106 participants (23 with PID) had adequate sequencing data and were included in the analysis. The median age of the women in the cohort was 19 years (range 16–27), 40% (42) described their ethnic group as black and 54% (57) reported ≥2 sexual partners in the previous year. At baseline, 22% of women (23) had C. trachomatis, 21% (22) had bacterial vaginosis (diagnosed by Nugent’s score 7–10 on Gram stain, with a further three classified as intermediate), 4% (4) had Mycoplasma genitalium and 2% (2) Neisseria gonorrohoeae (diagnosed by Gen-Probe PCR). The estimated mean time from baseline sample to diagnosis of PID, where data were available (in 22 of 23 women), was 35 weeks (range 13–52 weeks).

Microbiome comparison and statistical analysis
The sequence analysis protocol was based on previously published methods.17 Briefly, reads were first filtered for length >1200 and <1800 bases, before chimeric reads were removed with yacrd.18 Remaining reads were mapped to the National Center for Biotechnology Information (NCBI) 16S Microbial database using minimap2.19 Aligned reads were merged by species using R, counted and normalised using decostand (R vegan package V4.0.4, 2021-02-15) by dividing by the mean total. The normalised number of reads mapping to each species against that sample was shown in a heat map. Shannon alpha diversity metrics were generated with vegan, and statistical significance was determined using a Kruskal-Wallis rank sum test for six clinical characteristics: PID versus no PID, age <20 versus ≥20, black ethnicity versus other ethnicity, chlamydia positive versus negative at baseline and ≥2 sexual partners in the previous year versus 0–1 partners and bacterial vaginosis by Gram stain versus no bacterial vaginosis. P values were adjusted for multiple comparisons using the Benjamini-Hochberg method.

DNA preparation and nanopore sequencing
DNA was isolated from vaginal samples using QIAamp DNA Mini Kit (Qiagen), following the manufacturer’s protocol for Gram-positive and difficult-to-lyse bacteria, and quantified using Qubit 3.0 Fluorometer. 16S rRNA genes in each sample were amplified with the primers, B27F (TTCCTGTTGCTGATATTCAG RTTTYGATYMTGGCTCAQ) and B1492R (ACTTGCCTGTCGCTCTACGTGACACTT), which were modified from S-D-bact-0008-c-S20 and S-y-Univ-1492-a-A-19.16 DNA libraries, each containing 12 different clinical samples, were prepared using PCR Barcoding Kit - EXP-PBC096 and Ligation Sequencing Kit 1D - SQK-LSK108, and sequenced on MinION MK I with SpotON Flow Cell Mk I FLO-MIN106 R9.4 (Oxford Nanopore Technologies). Sequencing reads were live base called locally using the MinKNOW protocol - NC_48HR_Sequencing_Run_FLO-MIN106_SQK-LSK108.py. Sequence data have been submitted to the ENA database with accession number PRJEB41336.
Vaginal microbiota show three main clusters

Figure 2 shows three main taxonomic clusters dominated by L. iners, L. crispatus and G. vaginalis, with a smaller number of samples dominated by L. jensenii and L. gasseri. The lactobacillus dominated clusters showed lower Shannon diversity than the G. vaginalis cluster.

No association between increasing diversity and PID

Comparing Shannon alpha diversity (Figure 3) across six risk factors (PID, C. trachomatis, BV, age, ethnicity and multiple partners) showed no association between a more diverse microbiome and subsequent PID (Kruskal-Wallace rank sum test $p=0.53$) or concurrent C. trachomatis infection ($p=0.38$). However, there was a significantly higher diversity in samples from participants aged ≥20 years ($p=0.048$), participants of black ethnic group ($p=0.002$) and those with a diagnosis of bacterial vaginosis (BV, $p<0.0001$).

PID was not associated with lower abundance of D(−) isomer lactic acid producing lactobacilli

Figure 4 demonstrates the relative abundance of the five main species Lactobacillus iners, Lactobacillus crispatus, Gardnerella vaginalis, Lactobacillus jensenii and Lactobacillus gasseri (including a combination of D(−) isomer lactic acid producing lactobacilli L. crispatus, L. jensenii and L. gasseri), related to the three main clinical diagnoses:

1. BV at baseline.
2. PID within the next 12 months.
3. C. trachomatis at baseline.
The present study failed to detect an association between a more diverse vaginal microbiota and the occurrence of PID, as we found no difference in beta diversity between PID-affected and PID-unaffected women. This is in contrast to previous studies that identified BV as a strong predictor of PID.

### DISCUSSION

#### Principal findings

In the first ever community-based, frequency-matched prospective study of vaginal microbiota in women who did or did not develop PID, we found no association between a more diverse microbiota and subsequent pelvic inflammatory disease. Women with bacterial vaginosis had lower relative abundance of protective D(−) isomer lactic acid producing lactobacilli, but this did not apply to women who developed PID.

#### Strengths and weaknesses

This study is unique as it is the only community-based prospective study of the vaginal microbiome and PID, and the first to explore the possible association of PID with low levels of vaginal D(−) isomer lactic acid producing lactobacilli. Previous studies were based in hospitals or clinics, or were cross-sectional rather than prospective. Forty per cent of participants were of black ethnicity, a group who may have higher rates of bacterial vaginosis compared with other ethnic groups. This is supported by our finding that the G. vaginalis taxonomic cluster was associated with black ethnicity, an association that lends internal validity to this study. The project also responds to recent calls for vaginal microbiome research to include prospective studies and data on sexual behaviour.

To our knowledge, this is also the first study to use long-read sequencing on self-taken vaginal swabs. Combined with evidence...
Figure 4  Relative abundance of the dominant species, including a combination of D(−) isomer lactic acid producing lactobacilli (Lactobacillus crispatus, L. jensenii and L. gasseri), compared for the three main clinical diagnoses: box A: bacterial vaginosis (BV); box B: pelvic inflammatory disease (PID); box C: chlamydia trachomatis at baseline. Red: has clinical diagnosis. Green: does not have clinical diagnosis. BV (A) was associated with lower abundances of D(−) isomer lactic acid producing lactobacilli species ($p = 4.05 \times 10^{-6}$) and higher abundance of Gardnerella vaginalis ($p = 3.06 \times 10^{-9}$), but this did not apply to PID (B) or Chlamydia.
that unsupervised self-collected vaginal swabs give similar microbiota results to clinician-collected swabs, this study supports feasibility of larger scale community-based projects. Encouragingly, the clusters we identified using long-read sequencing were similar to those found in earlier studies with three main clusters: one dominated by L. iners, one by L. crispatus and a more diverse cluster dominated by anaerobes. Further strengths include our pilot work confirming the integrity of stored samples and the availability of detailed information on the demographics and baseline infection status of all participants, with sequence analysis being conducted blind to these characteristics. Finally, the large number of participants (>2300) in the POPI trial ensured that controls were well matched.

The main limitation is that this was an exploratory study with a small number of PID cases. However, it is similar in size to earlier studies and contributes to the very limited available data on vaginal microbiota composition in relation to PID. The small sample size also meant we could not do complex epidemiological analyses. Another major weakness is the absence of negative and positive controls. The lack of laboratory negative extraction controls meant we could not be sure that B. adolescentis was a contaminant. Removal of affected samples did not seem to influence the results, although it did reduce the power of the study. PID is a clinical diagnosis of low specificity, with around a third of diagnoses not confirmed on laparoscopy. This could also weaken the power of the study. However, diagnosis was confirmed by three experienced sexual health physicians. Although we found a significant association between BV and lower abundance of D-isomer lactic acid producing lactobacilli, we could not attribute causality or measure D(−) isomer lactic acid concentrations in these samples. We only assessed samples at one time point, and there was an estimated time gap of 3–12 months between obtaining the samples and diagnosis of PID, with only two women diagnosed with PID within 4 months of sampling. The vaginal microbiota may have changed over time in a significant proportion of women, potentially weakening the power of the study. However, this was unavoidable in a prospective study of PID.

Another weakness is that the sensitivity of samples stored for 16 years may be reduced. Nonetheless, other sequencing studies using stored samples have shown reliable results. Although commonly used, the Kruskal-Wallis rank sum test is not generally considered appropriate for differential analysis of compositional data. A differential abundance analysis approach that is suited to compositional data would have been preferable. We used 16S rRNA gene sequencing, which does not cover fungi, protozoa and viruses, and may not reliably identify low loads of C. trachomatis or N. gonorrhoeae. However, samples had already been tested for C. trachomatis, N. gonorrhoeae and M. genitalium by PCR. We could not evaluate three recently identified novel Gardnerella species, as it is unclear whether full length 16S rRNA sequencing adequately differentiates the species within the Gardnerella genus, and they are not yet included in NCBI and other 16S rRNA databases. In addition, the taxon for bacterial vaginosis associated bacteria such as BVAB-1 (Candidatus Lachnocurva vaginae) may be misclassified as Shuttleworthia using the NCBI database. Finally, we analysed by the dominant species within a community state type rather than by actual community state type, but this would likely give similar results.

Comparison with other studies
There is a dearth of studies of vaginal microbiota in women with and without PID, and like ours, none have shown a clear relationship between the vaginal microbiome and subsequent PID. A cross-sectional study from China used high-throughput sequencing on pelvic and cervical samples from 38 women with PID and 19 controls. They found microbiota in PID could be dominated either by a single organism or by polymicrobial infection. Analyses of bacterial vaginosis and incident PID in the Gynecologic Infection Follow-Through study were also inconclusive, with one analysis showing no association and others that women with bacterial vaginosis-associated bacteria had increased risk of PID. Interestingly, microbiological analysis of POPI samples found bacterial vaginosis was not significantly associated with PID after adjustment for baseline C. trachomatis infection. Most recently, Trent and colleagues analysed samples from 26 women aged 13–25 years with PID who were enrolled in the Technology Enhanced Community Health Nursing study. Over half of participants had low abundance of Lactobacillus species indicative of bacterial vaginosis.

Other studies of the vaginal microbiome in healthy ethnically diverse women have found similar main clusters/community state types to those in our study, two dominated by lactobacilli and one with higher proportions of anaerobic organisms. A recent study of the vaginal microbiome during genital infections found that L. crispatus was progressively replaced by L. iners in the shift from a healthy to an infected microbiome. As in our study, this was mainly characterised by anaerobic genera such as Gardnerella, Prevotella, Megaesophaga, Roseburia and Atopobium. We did not find an association between C. trachomatis and L. iners dominated vaginal microbiota, but numbers with C. trachomatis infection were small. Finally, our finding of lower rates of protective D(+) isomer lactic acid producing lactobacilli in women with (vs without) concurrent bacterial vaginosis confirms other studies. The cervicovaginal microbiota could modulate host functions to protect against infection.

Implications
This study supports the feasibility of larger scale community-based projects, the use of self-collected vaginal swabs and the integrity of stored samples. The significantly higher abundance of D(−) isomer lactic acid producing lactobacilli in women without bacterial vaginosis is in line with other studies suggesting a protective role for these lactobacilli. Finally, negative studies are important. The key impact from this study is the need for more work, particularly serial sampling studies, to gain knowledge of predisposing vaginal communities that may lead to PID.

Handling editor Jo Gibbs
Acknowledgements We would like to thank Professors David Martin and David Strachan for advice on study design.

Key messages
- There are few data on vaginal microbiota and subsequent development of pelvic inflammatory disease (PID).
- A lactobacilli-dominated vaginal microbiome may protect against pelvic inflammatory disease, but one dominated by Gardnerella species might increase susceptibility.
- In this first ever community-based prospective study of PID, there was no clear association between vaginal microbiota and development of PID in the next 12 months.
- In line with previous studies, findings highlight the need for large serial sampling studies to identify vaginal microbiota that might predispose to PID.
Contributors STS, PO and SKB designed the study; LZ and LP conducted the long-read 16S rRNA gene nanopore sequencing; AAW conducted the bioinformatics analysis; MF advised on statistics; PO and SKB wrote the first draft; and all authors contributed to the final version. PO, STS and AAW are responsible for the overall content as guarantor(s).

Funding The Wellcome Trust Institutional Strategic Support Fund grant number 204809/2/16/2 funded the sequencing of samples. Additional support came from National Institute for Health Research, Innovation for Innovation (i4i) grant: ‘A Point of Care Antimicrobial Resistance Test for Neisseria gonorrhoeae and Mycoplasma genitalium infection - Ensuring accurate therapy and antibiotic stewardship in sexual health medicine’ (II-LB-0214–20005). The Prevention of Pelvic Infection trial was funded by the BUPA Foundation grant number 684/GB14B and the UK Medical Research Council grant 80 280.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Bromley research ethics committee reference 07/Q0705/16. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplemental information. Not applicable.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Adam A Witney http://orcid.org/0000-0003-4561-7170
S Tariq Sadiq http://orcid.org/0000-0002-0608-3818
Pippa Oakeshott http://orcid.org/0000-0003-2568-8211

REFERENCES
Supplementary Figure 1. Heat map of relative abundances of microbial taxa in the vaginal bacterial communities of 148 women who did or did not develop pelvic inflammatory disease (PID) in the next 12 months.

A) Hierarchical clustering based on the Pearson correlation dissimilarity of the relative abundances within each sample. Clusters are labelled I-V according to those defined in Ravel et al \( ^{16} \) with the addition of a \textit{B. gladioli} cluster (BG).

B) Clinical characteristics: Black filled box indicates PID, Chlamydia positive at baseline, >1 sexual partner in the past year, Age < 20 years, Self-assigned black ethnic group.

C) Heatmap of relative abundances showing only the most common species.

D) Shannon diversity for each sample: red dashed line shows the median diversity for each cluster, purple dashed lines show + the standard deviation.
**Supplementary Figure 2.** Relative abundance of *B. gladioli* for each sample, ordered by the date of swab collection. Blue lines separate years of the study. The red line shows that swabs collected after March 2006 had a high percentage of *B. gladioli* representative sequences in the samples suggesting this was a contaminant.
Supplementary Figure 3. Relative abundance of the dominant species in the three major clusters compared for each risk factor.
A) PID, B) Chlamydia at baseline, C) Age < 20, D) Black ethnicity, E) ≥2 sexual partners, F) BV at baseline
Red: Has risk factor. Green: Does not have risk factor
BV is associated with low levels of *L. crispatus* (p = 3.96x10^-3) and high levels of *G. vaginalis* (p = 3.06x10^-9 Box F). *G. vaginalis* is also associated with black ethnicity (p = 0.004 Box D)
Supplementary Figure 4. Non Metric Multi-dimensional Scaling (NMDS) of Bray-Curtis distances for Bacterial vaginosis (BV), Chlamydia at baseline, and Pelvic inflammatory disease

Coding: Boxes A and B BV (red), no BV (green) or intermediate BV (blue). Box C PID (pelvic inflammatory disease): Red subsequent PID, Green no PID. Box D Chlamydia: Red Chlamydia positive, Green no chlamydia.

Box A Species fitted; Boxes B, C and D Ellipses drawn at the SD of the mean of each group.

For Bacterial vaginosis (Boxes A and B) the three clusters correlated with the dominant species in Figure 1: L. iners, L. crispatus, G. vaginalis. Similarly dimensions 1 and 2 distinguish BV versus no BV (Box B), but this did not apply to PID (Box C) or to Chlamydia (Box D).