






OPEN ACCESS

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

Sarah Kerry-Barnard,<sup>1</sup> Liqing Zhou,<sup>2</sup> Laura Phillips,<sup>2</sup> Martina Furegato,<sup>2</sup> Adam A Witney ,<sup>3</sup> S Tariq Sadiq ,<sup>2,4</sup> Pippa Oakeshott <sup>1</sup>

<sup>1</sup>Population Health Research Institute, St George's, University of London, London, UK

<sup>2</sup>Applied Diagnostic Research and Evaluation Unit, St George's, University of London, London, UK

<sup>3</sup>Institute for Infection and Immunity, St George's, University of London, London, UK

<sup>4</sup>Clinical Academic group in Infection and Immunity, St George's University Hospitals NHS Foundation Trust, London, UK

## Correspondence to

Professor Pippa Oakeshott, St George's University of London, London, UK; oakeshot@sgul.ac.uk

SK-B, LZ, AAW, STS and PO contributed equally.

Received 24 August 2021  
Accepted 26 November 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Kerry-Barnard S, Zhou L, Phillips L, et al. *Sex Transm Infect* Epub ahead of print: [please include Day Month Year]. doi:10.1136/sextrans-2021-055260

## 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 (POPI) 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 <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.

**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.<sup>1</sup> 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.<sup>2–5</sup> However, a high diversity microaerophilic/anaerobic microbiome with fewer lactobacilli (as may be found in bacterial vaginosis often with *Gardnerella vaginalis* and/or *Atopobium vaginae*) may increase susceptibility to STIs by causing cervicovaginal barrier disruption and epithelial portals of entry for ascending infection.<sup>3,6</sup>

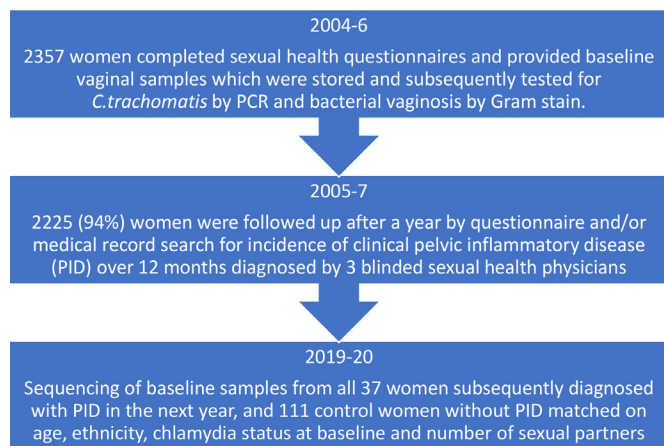
Not all lactobacilli are equal.<sup>3</sup> A recent study suggested that lactobacilli that produce mainly D(−) isomer lactic acid (*Lactobacillus crispatus*, *L. jensenii* and *L. gasseri*) may be associated with long-term protection against *C. trachomatis* infection, possibly by reducing epithelial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

A recent systematic review called for longitudinal studies of vaginal microbiota and infections to lay the foundation for possible prevention and treatment strategies.<sup>5</sup> 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 (POPI) chlamydia screening trial in 2004–2006.<sup>10</sup> Previously, 16S rRNA sequencing of 20 samples using a long-read PacBio sequencing platform had confirmed adequate DNA integrity in these samples.<sup>11</sup>

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.<sup>12,13</sup>

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*,<sup>5,7</sup> a major cause of PID,

## Original research



**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).<sup>14</sup>

## 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<sup>10</sup> (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's score  $\geq 7$ ) and placed the swab in a tube of Aptima 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 PID<sup>15</sup>: age  $< 20$  versus  $\geq 20$ , black ethnicity versus other ethnicity, chlamydia positive versus negative at baseline and  $\geq 2$  sexual partners in the previous year versus 0–1 partners. The diagnosis of PID was made by three sexual health physicians (blinded to chlamydia status) based on medical records and participant questionnaires after 12 months and using modified Hager's criteria: pelvic pain, cervical motion tenderness and/or adnexal tenderness.<sup>10</sup>

### 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 (TTTCTGTTGGTGCTGATATTGCAGRGTTYGATYMTGGCTCAG) and B1492R (ACTTGCCTGTCGCTCTATCTTCRGYTACCTTGTTACGACTT), which were modified from S-D-bact-0008-c-S20 and S\*-Univ-1492-a-A-19.<sup>16</sup> 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.

### Microbiome comparison and statistical analysis

The sequence analysis protocol was based on previously published methods.<sup>17</sup> Briefly, reads were first filtered for length  $> 1200$  and  $< 1800$  bases, before chimeric reads were removed with yacr. Remaining reads were mapped to the National Center for Biotechnology Information (NCBI) 16S Microbial database using minimap2.<sup>19</sup> 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 margin 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  $\geq 20$ , black ethnicity versus other ethnicity, chlamydia positive versus negative at baseline,  $\geq 2$  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 lactobacilli *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 vaginalis*<sup>15</sup> versus  $\leq 11\%$  of women without PID. (We used *G. vaginalis* as a surrogate for bacterial vaginosis.)

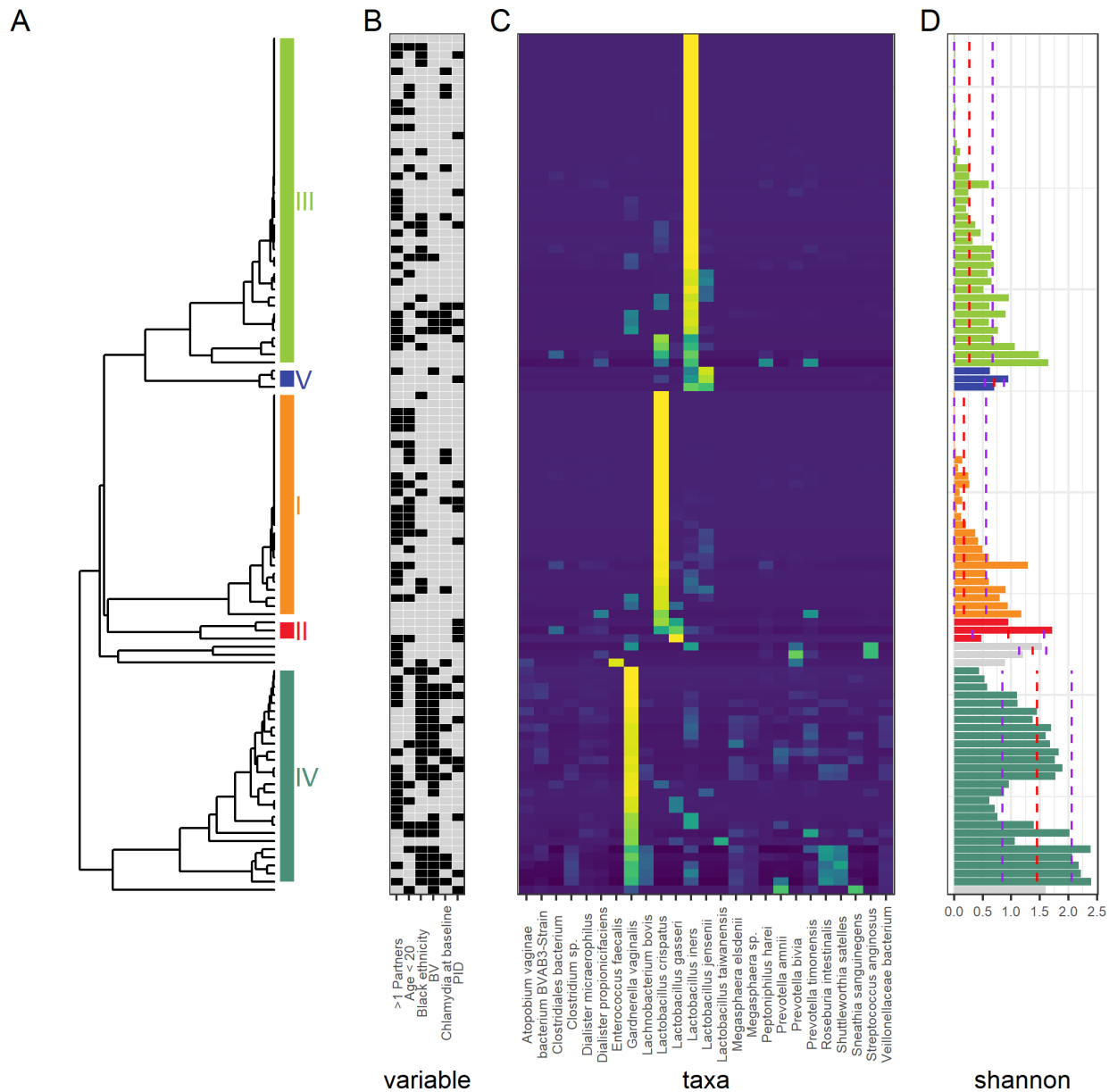
### Excluded samples

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,<sup>20</sup> the presence of *B. gladioli* is unusual, and the *Burkholderia* genus has been associated with contamination in other microbiome studies.<sup>21</sup> 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  $\geq 2$  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 gonorrhoeae* (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).



**Figure 2** Heat map of relative abundances of microbial taxa in the vaginal bacterial communities of 106 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<sup>16</sup>. (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  $\pm$  the SD.

### 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

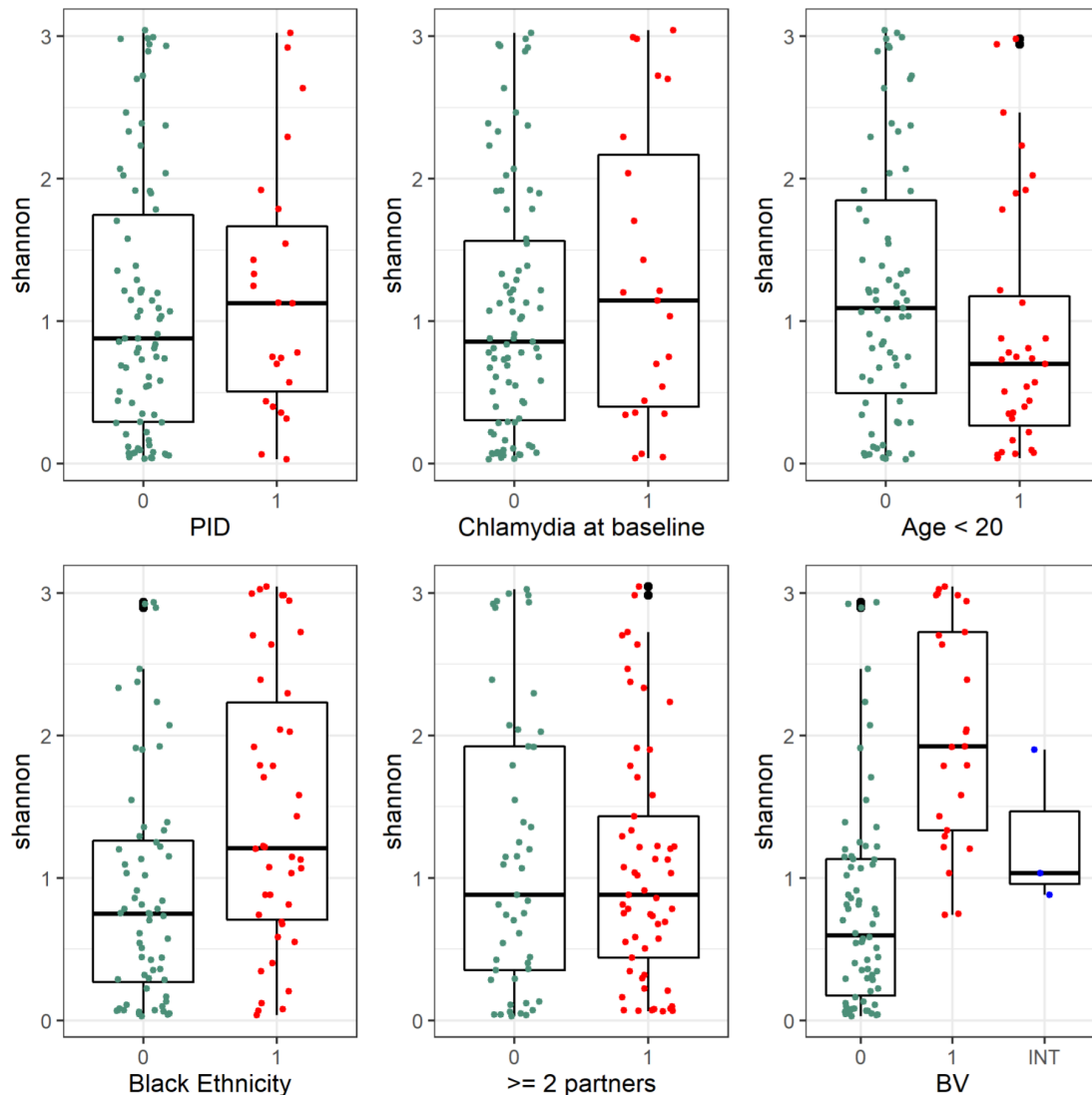
Comparing Shannon alpha diversity (figure 3) across six risk factors (PID, *C. trachomatis*, BV, age, ethnicity and multiple partners) showed no association between a more diverse microbiome and subsequent PID (Kruskal-Wallis rank sum test  $p=0.53$ ) or concurrent *C. trachomatis* infection ( $p=0.38$ ). However, there was a significantly higher diversity in samples from participants aged  $\geq 20$  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.



**Figure 3** Shannon diversity plotted for each clinical risk factor. Kruskal-Wallis rank sum test shows differences in Shannon diversity between groups for bacterial vaginosis (BV;  $n=22$ ,  $p=5.4 \times 10^{-8}$ ), black ethnic group ( $n=42$ ,  $p=0.002$ ), age <20 ( $n=33$ ,  $p=0.048$  less diverse), but no differences for pelvic inflammatory disease ( $n=23$ , PID:  $p=0.53$ ), Chlamydia at baseline ( $n=23$ ,  $p=0.38$ ),  $\geq 2$  sexual partners ( $n=57$ ,  $p=0.88$ ). 1 (red)=Has risk factor, 0 (green)=does not have risk factor.

This shows that unlike BV, neither PID nor *C. trachomatis*, was associated with lower abundance of D(-) isomer lactic acid producing lactobacilli species ( $p=0.77$  and  $0.33$ , respectively). Online supplemental figure 3 shows similar results.

### Beta diversity shows no association with PID

Visualisation of Bray-Curtis based NMDS dimensions one and two (stress=0.09 after 20 iterations (online supplemental figure 4) showed no association with subsequent PID or concurrent *C. trachomatis*. However for BV, three clusters correlated with the dominant species shown in figure 1 (*L. iners*, *L. crispatus* and *G. vaginalis*), with dimensions one and two distinguishing BV versus no BV (online supplemental figure 4A,B, respectively).

## DISCUSSION

### Principal findings

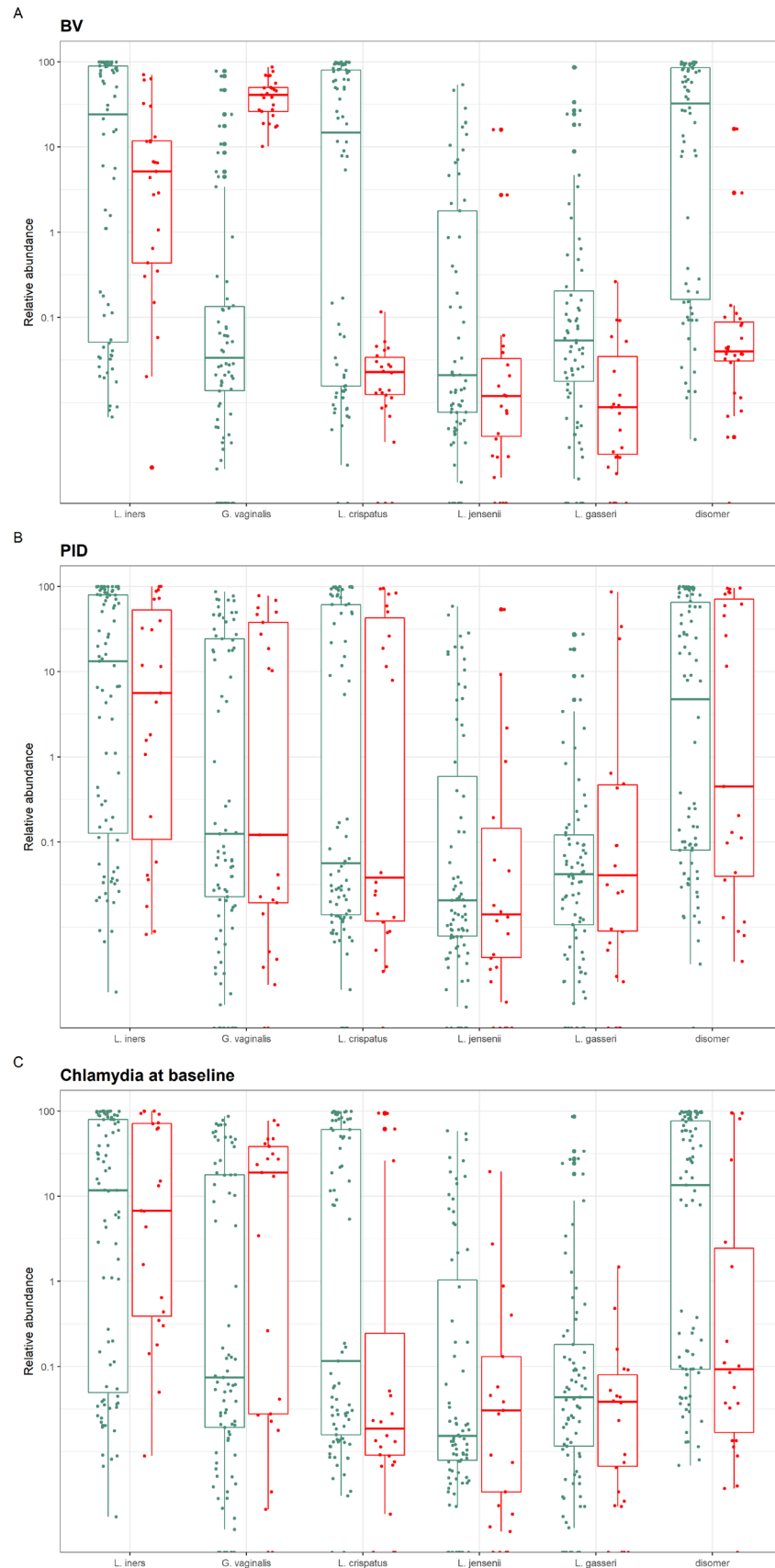
In the first ever community-based, frequency-matched prospective study of vaginal microbiota in women who did or did not develop PID, we found no association between a more diverse microbiota and subsequent pelvic inflammatory disease. Women

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,<sup>5 9 22</sup> or were cross-sectional rather than prospective.<sup>23</sup> Forty per cent of participants were of black ethnicity, a group who may have higher rates of bacterial vaginosis<sup>8 24</sup> and STIs<sup>10</sup> 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.<sup>8</sup> The project also responds to recent calls for vaginal microbiome research to include prospective studies<sup>5</sup> and data on sexual behaviour.<sup>2</sup>

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,<sup>25</sup> 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<sup>9,24</sup> 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<sup>11</sup> 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<sup>7,9,23</sup> 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. gladioli* 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.<sup>10</sup> 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.<sup>6</sup> However, this was unavoidable in a prospective study of PID.<sup>22</sup>

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.<sup>11,26</sup> 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*.<sup>3</sup> However, samples had already been tested for *C. trachomatis*, *N. gonorrhoeae* and *M. genitalium* by PCR.<sup>10,15</sup> We could not evaluate three recently identified novel *Gardnerella* species,<sup>27</sup> 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<sup>9</sup> rather than by actual community state type,<sup>28</sup> 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,<sup>3,23,29</sup> and like ours, none have shown a clear

### 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.

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.<sup>23</sup> Analyses of bacterial vaginosis and incident PID in the Gynecologic Infection Follow-Through study were also inconclusive, with one analysis showing no association<sup>22</sup> and others that women with bacterial vaginosis-associated bacteria had increased risk of PID.<sup>6,14</sup> Interestingly, microbiological analysis of POPI samples<sup>15</sup> 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.<sup>29</sup> 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.<sup>9,24</sup> A recent study of the vaginal microbiome during genital infections<sup>4</sup> 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*, *Megasphaera*, *Roseburia* and *Atopobium*. We did not find an association between *C. trachomatis* and *L. iners* dominated vaginal microbiota,<sup>9</sup> 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.<sup>7</sup> The cervicovaginal microbiota could modulate host functions to protect against infection.<sup>7</sup>

### 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.<sup>3,7,30</sup> 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.

**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 Institution Strategic Support Fund grant number 204809/Z/16/Z funded the sequencing of samples. Additional support came from National Institute for Health Research, Invention 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

- Haggerty CL, Totten PA, Tang G, *et al.* Identification of novel microbes associated with pelvic inflammatory disease and infertility. *Sex Transm Infect* 2016;92:441–6.
- Lewis FMT, Bernstein KT, Aral SO. Vaginal microbiome and its relationship to behavior, sexual health, and sexually transmitted diseases. *Obstet Gynecol* 2017;129:643–54.
- van de Wijgert JHHM, JHHM vandeW. The vaginal microbiome and sexually transmitted infections are interlinked: consequences for treatment and prevention. *PLoS Med* 2017;14:e1002478.
- Ceccarani C, Foschi C, Parolin C, *et al.* Diversity of vaginal microbiome and metabolome during genital infections. *Sci Rep* 2019;9:14095.
- Tamarelle J, Thiébaud ACM, de Barbeyrac B, *et al.* The vaginal microbiota and its association with human papillomavirus, *Chlamydia trachomatis*, *Neisseria gonorrhoeae* and *Mycoplasma genitalium* infections: a systematic review and meta-analysis. *Clin Microbiol Infect* 2019;25:35–47.
- Ness RB, Kip KE, Hillier SL, *et al.* A cluster analysis of bacterial vaginosis-associated microflora and pelvic inflammatory disease. *Am J Epidemiol* 2005;162:585–90.
- Edwards VL, Smith SB, McComb EJ, *et al.* The cervicovaginal Microbiota-Host interaction modulates *Chlamydia trachomatis* infection. *mBio* 2019;10. doi:10.1128/mBio.01548-19. [Epub ahead of print: 13 08 2019].
- Borgdorff H, van der Veer C, van Houdt R, *et al.* The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLoS One* 2017;12:e0181135.
- van Houdt R, Ma B, Bruisten SM, *et al.* *Lactobacillus iners*-dominated vaginal microbiota is associated with increased susceptibility to *Chlamydia trachomatis* infection in Dutch women: a case-control study. *Sex Transm Infect* 2018;94:117–23.
- Oakeshott P, Kerry S, Aghaizu A, *et al.* Randomised controlled trial of screening for *Chlamydia trachomatis* to prevent pelvic inflammatory disease: the POPI (prevention of pelvic infection) trial. *BMJ* 2010;340:1642.
- Wagner J, Kerry-Barnard S, Sadiq ST, *et al.* Research benefits of storing genitourinary samples: 16S rRNA sequencing to evaluate vaginal bacterial communities. *Int J STD AIDS* 2017;28:315–7.
- Shin J, Lee S, Go M-J, *et al.* Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci Rep* 2016;6:29681.
- Quan L, Dong R, Yang W, *et al.* Simultaneous detection and comprehensive analysis of HPV and microbiome status of a cervical liquid-based cytology sample using Nanopore MinION sequencing. *Sci Rep* 2019;9:19337.
- Haggerty CL, Ness RB, Totten PA, *et al.* Presence and concentrations of select bacterial vaginosis-associated bacteria are associated with increased risk of pelvic inflammatory disease. *Sex Transm Dis* 2020;47:344–6.
- Hay PE, Kerry SR, Normansell R, *et al.* Which sexually active young female students are most at risk of pelvic inflammatory disease? A prospective study. *Sex Transm Infect* 2016;92:63–6.
- Klindworth A, Pruesse E, Schweer T, *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;41:e1.
- Cuscó A, Catozzi C, Viñes J, *et al.* Microbiota profiling with long amplicons using nanopore sequencing: full-length 16S rRNA gene and whole *rrn* operon. *F1000Res* 2018;7:1755.
- Marijon P, Chikhi R, Varré J-S. *yacr* and *fpa*: upstream tools for long-read genome assembly. *Bioinformatics* 2020;36:3894–6.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–100.
- Imataki O, Kita N, Nakayama-Imaohji H, *et al.* Bronchiolitis and bacteraemia caused by *Burkholderia gladioli* in a non-lung transplantation patient. *New Microbes New Infect* 2014;2:175–6.
- Salter SJ, Cox MJ, Turek EM, *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;12:87.
- Ness RB, Hillier SL, Kip KE, *et al.* Bacterial vaginosis and risk of pelvic inflammatory disease. *Obstet Gynecol* 2004;104:761–9.
- Wang Y, Zhang Y, Zhang Q, *et al.* Characterization of pelvic and cervical microbiotas from patients with pelvic inflammatory disease. *J Med Microbiol* 2018;67:1519–26.
- Ravel J, Gajer P, Abdo Z, *et al.* Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 2011;108(Suppl 1):4680–7.
- Wylie KM, Blankenship SA, Tuuli MG, *et al.* Evaluation of patient- versus provider-collected vaginal swabs for microbiome analysis during pregnancy. *BMC Res Notes* 2018;11:706.
- Bai G, Gajer P, Nandy M, *et al.* Comparison of storage conditions for human vaginal microbiome studies. *PLoS One* 2012;7:e36934.
- Hill JE, Albert AYK, VOGUE Research Group. Resolution and Cooccurrence patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. piotii*, and *G. vaginalis* within the vaginal microbiome. *Infect Immun* 2019;87. doi:10.1128/IAI.00532-19. [Epub ahead of print: 18 11 2019].
- France MT, Ma B, Gajer P, *et al.* VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* 2020;8:166.
- Trent ME, Perin J, Matson P. Vaginal microbiota among adolescent and young adult women with pelvic inflammatory disease. *Sex Transm Infect* 2019:A265.
- Cohen CR, Wierzbicki MR, French AL, *et al.* Randomized trial of Lactin-V to prevent recurrence of bacterial vaginosis. *N Engl J Med* 2020;382:1906–15.

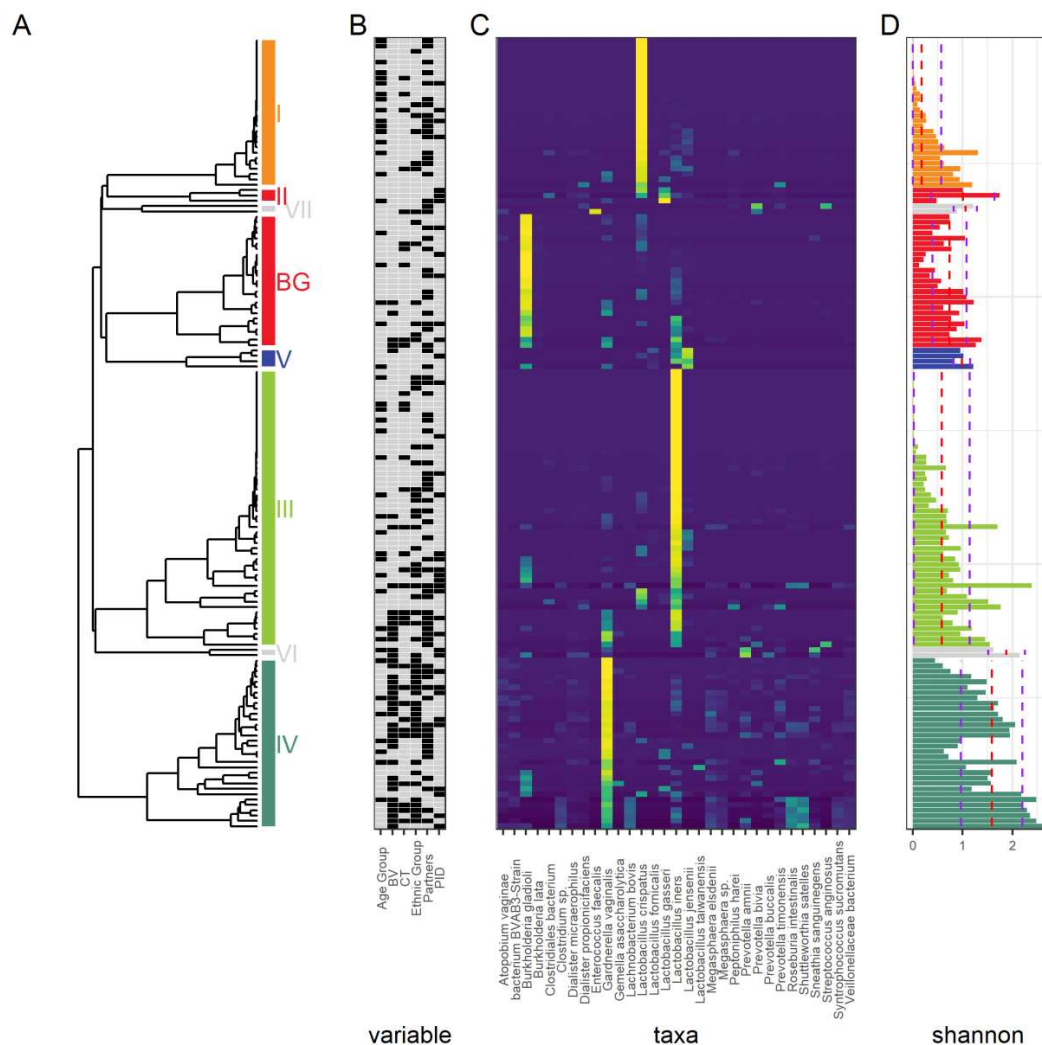
**Supplementary Figure 1. Heat map of relative abundances of microbial taxa in the vaginal bacterial communities of 148 women who did or did not develop pelvic inflammatory disease (PID) in the next 12 months.**

**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<sup>16</sup> with the addition of a *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

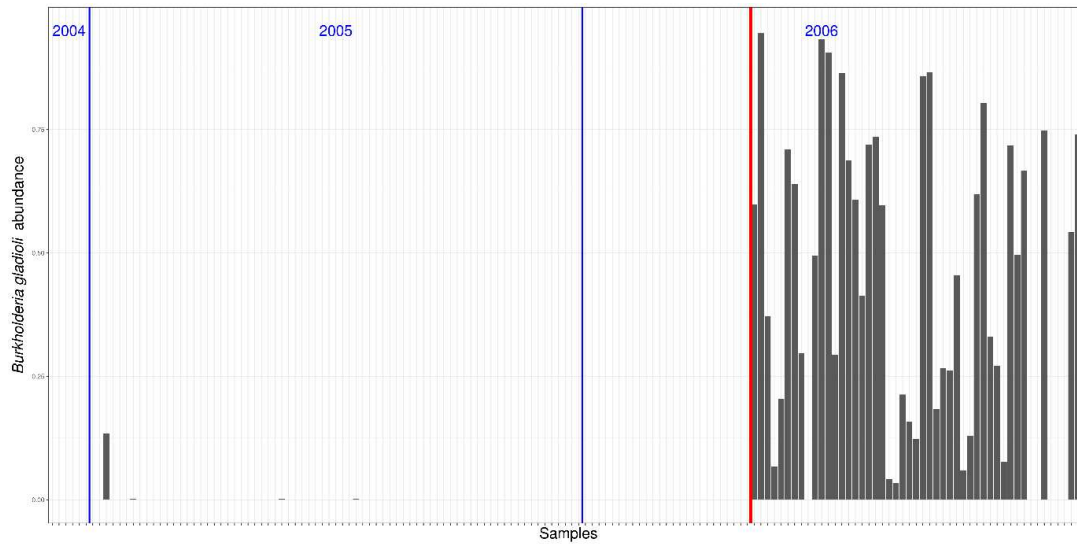
**C) Heatmap** of relative abundances showing only the most common species

**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.

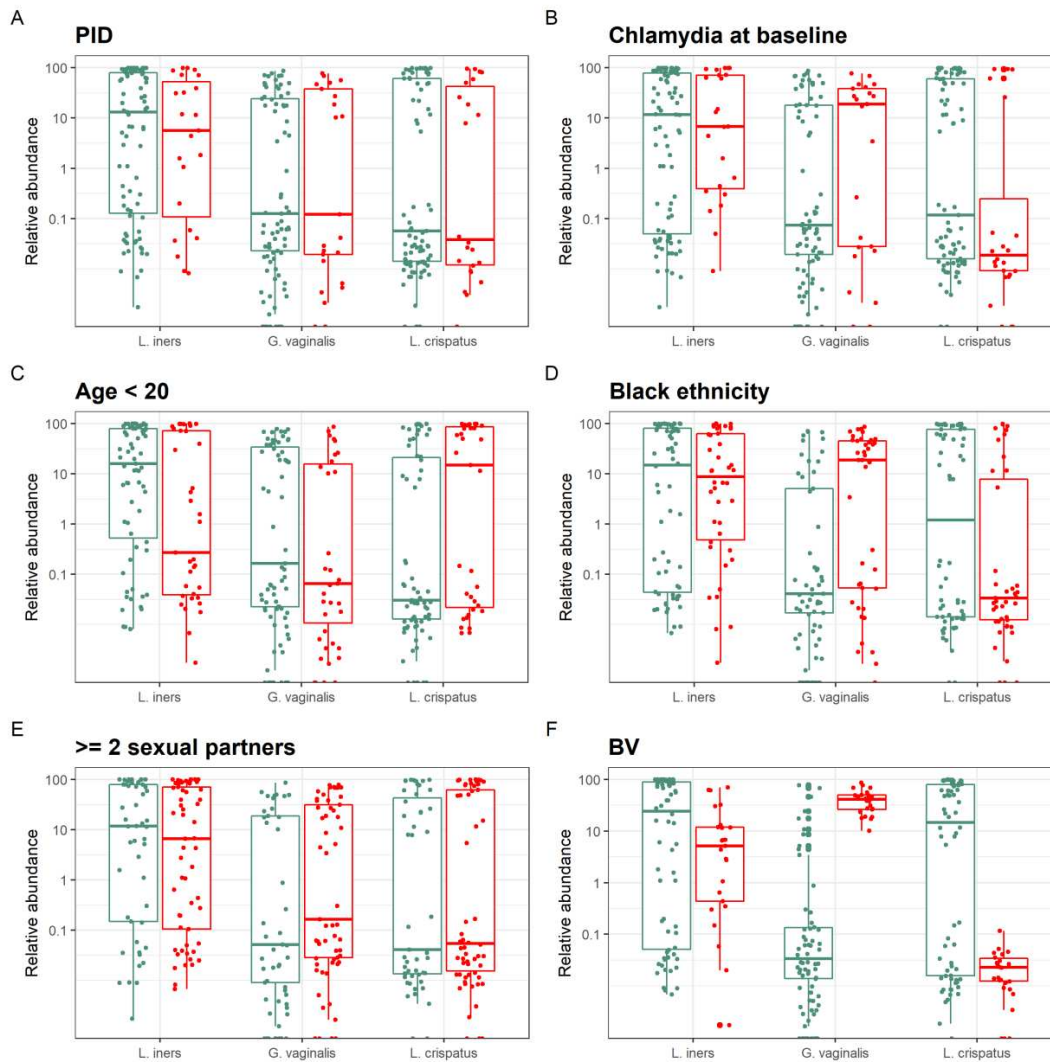


**Supplementary Figure 3. Relative abundance of the dominant species in the three major clusters compared for each risk factor.**

A) PID, B) Chlamydia at baseline, C) Age < 20, D) Black ethnicity, E)  $\geq 2$  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* ( $p = 3.96 \times 10^{-3}$ ) and high levels of *G. vaginalis* ( $p = 3.06 \times 10^{-9}$  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).

