The effect of secretor status and the vaginal microbiome on birth outcome


1March of Dimes European Prematurity Research Centre, Imperial College London
2Institute for Reproductive and Developmental Biology, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, United Kingdom
3Institute for Women’s Health, University College London, London, UK
4Department of Life Sciences, Imperial College London, London, UK
5Glycosciences Laboratory, Department of Metabolism Digestion and Reproduction, Imperial College London, London, UK
6Division of Digestive Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, UK.
*Correspondence to Dr D.A. MacIntyre (d.macintyre@imperial.ac.uk)

Summary

Mutations in the FUT2 gene that result in a lack of expression of histo-blood group antigens on secreted glycoproteins may shape the vaginal microbiota with consequences for birth outcome. To test this, we analysed the relationship between secretor status, vaginal microbiota and gestational length in an ethnically diverse cohort of 313 pregnant women, including 91 who delivered prematurely. Lactobacillus species were found to co-occur less often with other microbial taxa in non-secretors. Moreover, non-secretors with Lactobacillus spp. depleted vaginal microbiota in early

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
pregnancy had significantly shorter gestational length than *Lactobacillus* spp.
dominated non-secretors (mean of 245.5 (SD=44.5) versus 265.9 (23.6)); p=0.045),
but not compared to *Lactobacillus* spp. dominated (261.8 (27.5)) and depleted (264.3
days (21.2)) secretors. In identifying a relationship between blood-group antigen
expression and vaginal microbiota-host interactions, our results point towards
stratification by secretor status as an important factor for considering preterm birth risk
and prevention.

**Keywords**

Preterm birth, Secretor, blood-type antigens, vaginal microbiome, *Lactobacillus*
dominance
Introduction

Histo-blood group antigens (HBGA), ABH and Lewis, are carbohydrate sequences found on the surface of a range of cell types and in secretions such as cervico-vaginal fluid. These antigens can serve as attachment sites and energy sources for microbial organisms \(^1\text{-}^3\). The \textit{FUT2} gene encodes the \(\alpha(1,2)\)-fucosyltransferase enzyme, which adds a fucose residue to the terminal galactose on a type 1 glycan precursor forming the H antigen. Four mutations in the \textit{FUT2} gene have been identified (alleles \textit{se}\textsuperscript{302}, \textit{se}\textsuperscript{385}, \textit{se}\textsuperscript{428} and \textit{se}\textsuperscript{571}) that result in non-secretion of these antigens from mucosal surfaces, in homozygotes, and up to 20\% of human populations express this non-secretor phenotype \(^4\text{-}^7\). Non-secretors will present as either Le(a+b\textsuperscript{-}) or Le(a-b\textsuperscript{-}), depending upon the presence of mutations in their \textit{FUT3} gene. Secretor status has been associated with increased and decreased susceptibility to a range of bacterial and viral pathogens as well as gut microbiota composition \(^8\text{-}^{14}\).

The vaginal microbiome plays an important role in influencing pregnancy outcome. Dominance of the vaginal niche by \textit{Lactobacillus} species has been widely reported to associate with healthy, term pregnancy \(^{15,16}\). In contrast, high diversity microbial communities depleted of \textit{Lactobacillus} species and enriched with pathobionts and/or bacterial vaginosis (BV)-associated bacteria, increase the risk of adverse pregnancy outcomes including miscarriage and preterm birth \(^{17\text{-}19}\). Preterm birth represents a significant global health burden – it is the leading cause of death in children under age five and associates with serious short and long-term morbidities in survivors \(^{20\text{-}22}\).

Significant risk factors for spontaneous preterm birth in our populations include history of late miscarriage or preterm birth, and previous cervical excisional treatment, an operation to treat cervical pre-cancer \(^{23,24}\). Both of these risk factors have been linked with characteristic compositions of the vaginal microbiota \(^{15,16,25,26}\).
In a recent study of 300 pregnant women (of whom 28 experienced preterm birth), maternal secretor status was reported as an independent risk factor for preterm delivery (Caldwell et al., 2020). Specific human milk oligosaccharides (HMOs) in blood and urine have also been recently associated with vaginal microbiota in a small cohort (n=60) of women presenting with threatened preterm labour (Pausen et al., 2020). We postulated that the composition of the vaginal microbiota in pregnancy might be influenced by maternal secretor status, and that this may influence pregnancy duration. To characterise the effect of secretor status on the vaginal microbiome and gestational length, we sequenced the second exon of the FUT2 gene to infer secretor status and undertook metataxonomic analysis of vaginal samples collected longitudinally from a cohort of 313 pregnant women, of which 91 delivered prematurely.
Results/Discussion

Distribution of non-secretors in our cohort

A total of 313 women were included in our study (Supplementary Table S1), of which 91 experienced spontaneous preterm birth before 37 weeks (259 days) of gestation. The ethnic distribution of the cohort was similar to the expected background prevalence within the clinical population. In total, 87 women were identified as non-secretors and 141 as heterozygous at these loci. Although nine women were heterozygous at multiple non-secretor loci, we did not identify any who were multiply homozygous. All four known non-secretor mutations were identified in our cohort. The se428 nonsense mutation accounted for 89.7% of the heterozygotes and 94% of the non-secretor homozygotes and was present across all ethnicities except East Asian patients. The remaining mutations had more restricted distributions: five women of East Asian ethnicity were homozygous for the se385 non-secretor allele (and a further five who were heterozygous) and only one se302 homozygote was identified in a woman of Central-South Asian descent (ten other women, of similar ethnicity, were heterozygous). We did not identify any homozygotes of se571. These patterns are consistent with previously documented distributions of non-secretor genetic diversity.

Overall, 28% of the cohort were identified as non-secretor with the highest proportion in Afro-Caribbean women, 34.6%, compared with 26.8% and 19.2% in European and Central-South Asian ethnicities respectively. For all subsequent analyses secretor homozygotes were combined with heterozygotes into a single group since they are phenotypically similar.

Lactobacilli are more refractory to co-colonisation in non-secretors

Vaginal microbiota were classified into one of five Community State Types (CST) using the 'predominant taxon' rule, which is concordant with clustering based methods and
allows comparisons across datasets\textsuperscript{28,29}. The most prevalent CST observed in our cohort was CST 1 (majority \textit{L. crispatus}) in 39.7, 45.9 and 41.1\% of the patients at early, mid and late gestation respectively, followed by CST 3 (majority \textit{L. iners}) in 30.0, 23.1 and 27.7\%, CST 4 (diverse) in 14.3, 12.8 and 16.6\%, CST 2 (majority \textit{L. gasseri}) in 10.1, 10.7 and 7.5\% and CST 5 (majority \textit{L. jensenii}) in 5.9, 7.5 and 7.1\%.

In samples taken during early pregnancy, prior to any interventions, we observed a similar prevalence of major bacterial genera across secretor and non-secretor women, including those classically associated with BV (a clinical syndrome of vaginal discharge and odour characterised by polymicrobial overgrowth), e.g., \textit{Gardnerella}, \textit{Atopobium}, \textit{Anaerococcus} and \textit{Sneathia}, (Figure 1A). As expected, most women had a high relative abundance of \textit{Lactobacillus}, though in the secretors there was evidence of a greater spread (Figure 1B). At species level, an enrichment for CST 3 in non-secretors and CST 4 in secretors was observed but these differences were not statistically significant (Fisher’s Exact test \(p=0.163\), Figure 1C).

We observed no difference in bacterial diversity (Shannon Index) between non-secretors and secretors in early pregnancy (\(p\)-value=0.409) nor any differential classification in principal components analysis (PCA) (\(p\)-value=0.303). Differential abundance analyses (DAA) suggested no significant differences of relative abundance in species on the basis of secretor phenotype (Supplementary Figure S1) however, the low prevalence of \textit{Lactobacillus} depleted patients (26.8\% of all early pregnancy samples) in our cohort may have limited our power to detect such associations. Low prevalence of \textit{Lactobacillus} depleted vaginal communities is likely attributable to increased relative abundance of \textit{Lactobacillus} species in the vaginal microbiome during pregnancy and consistent with our clinical population: mostly women with European ancestry who, compared to other ethnicities, have higher prevalence of
Lactobacillus-dominated vaginal microbiomes \(^\text{28,30}\). A recent study of 60 women presenting with suspected preterm labour reported positive associations between specific urinary and blood human milk oligosaccharides, and some vaginal taxa and clinical outcomes including preterm birth \(^\text{31}\). However, it is difficult to compare these findings to our own given that over half of their patient cohort received tocolysis to inhibit uterine contractions and prevent preterm birth (only a small number of women in their study cohort experienced preterm birth, \(n=11\)).

To investigate whether vaginal microbial community structure might differ between secretors and non-secretors, we compared co-occurrence networks inferred from correlation matrices based on the 16S relative abundance data (Figure 2) as previously described \(^\text{32}\). For this analysis, non-Lactobacillus taxa were classified as ‘pathobionts’, ‘BV-associated’ or ‘other’ using definitions from Wijgert et al \(^\text{33}\). In both the secretor and non-secretor microbial networks, nodes representing BV-associated bacteria tended to be connected via positive edges, reflecting positive correlations among these microbes (Figure 2A). Analysis of node importance, using Expected Influence (EI) \(^\text{34}\) as a metric, confirmed this pattern, i.e., that in both networks, nodes with the most positive EI scores were BV-associated bacteria and pathobionts, such as \textit{Prevotella timonensis} and \textit{Finegoldia magna} (Supplementary Figure S2). These patterns mirror previous studies showing that these microbes establish interactions based on mutualistic relationships \(^\text{35,36}\). In non-secretors \textit{L. crispatus}, \textit{L. jensenii} and \textit{L. iners} were the most negatively scored nodes for EI, but in secretors, a greater proportion of positive edges from these lactobacilli to other BV-associated microbes was observed (Figure 2B and Supplementary Figure S2). These results suggest that \textit{Lactobacillus} species, particularly \textit{L. crispatus}, are more refractory to co-colonisation in non-secretors where they may offer a greater “protective” effect via competitive
exclusion. These findings are consistent with comparative genomics studies that have highlighted bacterial cell surface glycoconjugates and carbohydrate-binding proteins as potentially important mediators of vaginal microbiota-host \(^3^7\).

**Non-secretors with Lactobacillus depleted microbiomes are associated with shorter gestational length**

On the basis of the distribution of *Lactobacillus* abundance (Figure 1C) and previous studies \(^3^8^,^3^9\) we classified the vaginal microbiome into *Lactobacillus* dominated (>90% *Lactobacillus* spp.) and depleted states (<90%). Non-secretors with a *Lactobacillus* depleted microbiome in early pregnancy had a shorter mean gestational length 245.5 days (SD=44.5) compared to *Lactobacillus* dominated non-secretors, 265.9 days (SD=23.6) and to both *Lactobacillus* depleted and dominated secretors, 264.3 (SD=21.2) and 261.8 (SD=27.5) days, respectively (Figure 3A). However, the association between *Lactobacillus* status in non-secretors on gestational outcome was diminished by mid pregnancy. To confirm these observations, for each sampling timepoint we modelled gestational length on secretor status and the vaginal microbiome (*Lactobacillus* status) together with BMI, age, cervical cerclage status and previous pregnancy history, i.e. cervical excisional treatment (LLETZ) and a combined previous preterm birth (PTB) and/or mid-trimester loss (MTL) covariate using generalised linear mixed effects modelling (GLMM) with ethnicity as a random effect (Table 1). As expected, previous pregnancy histories were significant explanatory variables of gestational length (Supplementary Table S3). Comparison of the GLMMs with generalised linear models (GLM) showed similar model fits (for example, AIC 2495.225 and 2497.225 for the GLM and GLMM, respectively, for the early pregnancy modelling) indicating that ethnicity was not a significant factor in the model. However, the early pregnancy GLMM showed that the secretor status and *Lactobacillus* status
interaction term is a significant explanatory variable of gestational length (p=0.02). This significance disappeared by mid-pregnancy supporting our earlier observation that non-secretors with a *Lactobacillus* depleted vaginal microbiome in early pregnancy, but not mid-pregnancy, are associated with shorter gestational lengths. *Post hoc* testing using the estimated marginal means confirmed this significant comparison (p-value=0.045), but also indicated that a difference in gestational length may exist between *Lactobacillus* depleted non-secretors (mean gestation of 245.5 days) and *Lactobacillus* depleted secretors (mean gestation of 264.3 days) (p-value=0.068). No significant difference between *Lactobacillus* dominated and depleted secretors was found (p-value=0.283).

When vaginal microbiomes were classified into CSTs, non-secretors with CST 2, 3 or 4 in early pregnancy had shorter gestations, mean (±SD) of 249.3 (34.1), 260.5 (26.9) and 232.2 (61.4) days, respectively, compared to CST 1 (269.4 days (19.7)) (Figure 3B). A similar pattern was observed with CST 5 (mean of 278 days), however sample size was limited to only n=4. By contrast, gestational length in secretors was broadly similar across all CSTs: mean gestational length (±SD) of 264.0 (28.4), 264.7 (20.5), 258.0 (29.4), 264.1 (18.2) and 262.6 (26.0) respectively in CST 1, 2, 3, 4 and 5. In line with these findings, modelling gestational length with early pregnancy CST (Table 1) showed that the interaction term between CST and secretor status was statistically significant (p-value=0.017). By mid-pregnancy, this interaction term was not statistically significant (Supplementary Table S4). The estimated marginal means identified the gestational length differences in the non-secretors between CST 1 versus CST 2, and CST 1 versus CST 4 to be statistically significant (respectively, p-value=0.037 and 0.033). Trend differences were also observed in the non-secretors between CST 2 and 5 (p-value=0.059), and CST 4 and 5 (p-value=0.051). These post
hoc tests also identified a difference in gestational length between secretors and non-secretors with CST 1 vaginal microbiomes in early pregnancy (p-value=0.017).

Our findings are consistent with a previous study by Lurie et al.\(^40\) who identified a significantly elevated proportion of non-secretors in a cohort of 28 patients who suffered preterm premature rupture of the membranes (PPROM) compared with term patients. PPROM precedes approximately 30% of all preterm birth cases and has been associated with ascending infection in the vagina\(^18,41\).

Caldwell et al.\(^42\) also reported that maternal secretor status is a potential biomarker for prematurity. Our data supports these findings but demonstrates that this relationship is likely shaped by the vaginal microbiota in early pregnancy. Dominance of the vaginal niche by\(Lactobacillus\) spp. is considered to be a hallmark of vaginal health due to the role that these species play in preventing colonisation of other microbes through the promotion of a hostile, acidic mucosal environment enriched by bacteriocins and other antimicrobial compounds\(^43,44\). Our results support previous findings in highlighting this protective role extending into pregnancy, but demonstrate that this is nuanced by host secretor status. Studies identifying an association between vaginal\(Lactobacillus\) spp. depletion and increased risk of preterm birth\(^19,28,45,46\) also emphasise that high diversity vaginal bacterial communities do not always result in preterm birth (as is the case of the secretors). An enhanced inflammatory response, previously observed in non-secretors\(^47\), may provide a mechanism linking gestational length to secretor status and vaginal microbiota but this requires further investigation.

It is possible that ABO blood groups influence the relationship between secretor status and the microbiota by affecting terminal glycan structures. To address this, we incorporated ABO status into our models and observed that there may be evidence of
shorter gestational length in blood group B women compared to blood group A, though this was only observed at mid-pregnancy (Supplementary Table S5).

Early developmental stages in pregnancy have lasting effects on pregnancy gestation and outcome. Our results indicate that the relationship between Lactobacillus depletion and reduced gestational length in non-secretors is altered as pregnancy progresses and points towards the microbiota-host interactions in early pregnancy being particularly key in shaping preterm birth risk. This result is consistent with findings from Stout and colleagues who showed that, in a predominantly African-American cohort, early gestation is an ecologically important time for events that predict subsequent term and preterm birth. Similarly, Tabatabaei and coworkers identified an association between Lactobacillus depleted vaginal microbiota in early pregnancy and increased risk of early preterm birth in a cohort of Canadian women of predominantly white European origin. Lactobacillus depletion is also a risk factor for miscarriage, which collectively suggests a potential effect of the vaginal microbiota upon decidual function and placentation.

Concluding remarks

In this study we identify a subset of women, non-secretors with Lactobacillus depleted vaginal microbiomes, who are predisposed to shorter gestational length. This stratification could allow a more targeted intervention of “at-risk” pregnancies, although these results should be verified with larger numbers of Lactobacillus depleted samples. We find evidence of a more “protective” role of Lactobacillus species, especially L. crispatus, in non-secretors. Whilst further studies are required to elucidate the mechanisms underlying these interactions in secretors and non-secretors, the data supports the evolving concept that L. crispatus offers optimal
protection against preterm birth and is a candidate for development of live biotherapeutic therapies during pregnancy.
Resource Availability

Further information and requests for supporting data, resources, and reagents should be directed to the Lead Contact: David MacIntyre (d.macintyre@imperial.ac.uk).

Materials Availability

Reagents from this study are available upon request.

Data and Code Availability

All sequencing data used in this study has been made available in ENA database (Accession numbers ERR4864561-ERR4865307). Python and R scripts used in support of the analyses are available at: https://github.com/samitkundu/FUT2.

Acknowledgements

The study was funded by the March of Dimes. The funder had no role in the study design, data collection and analysis, and preparation of the manuscript.

Author Contributions


Competing Interests

The authors declare no competing interests.
Methods

Ethics Statement

The study was conducted with approval of the NHS National Research Ethics Service (NRES) Committees London - City and East (REC 12/LO/2003) and London–Stanmore (REC 14/LO/0328), and by the North of Scotland Research Ethics Service (REC 14/NS/1078). Written informed consent was obtained from all patients prior to sampling and experiments were performed in accordance with the approved institutional guidelines.

Patient sampling

Recruitment and sampling were performed at Imperial College Healthcare NHS Trust Hospitals (Queen Charlotte’s and Chelsea and St Mary’s Hospitals), London, UK, at Chelsea & Westminster Hospital (NHS Trust, London, UK), University College London Hospital (NHS Foundation Trust, London, UK) and the Royal Infirmary of Edinburgh, Scotland, UK. Eligibility criteria included singleton pregnancies, with and without risk factors for preterm birth. Exclusion criteria included women under 18 years of age, those who had sexual intercourse within 72 h of sampling, vaginal bleeding in the preceding week, antibiotic use in the preceding 2 weeks, multiple pregnancies, HIV or Hepatitis C positive status. Detailed maternal clinical metadata and birth outcome data was collected for all participants. Cervicovaginal fluid swab samples were collected from the posterior fornix using BBL CultureSwab MaxV Liquid Amies swabs (Becton, Dickinson and Company, Oxford, UK) at up to three timepoints throughout pregnancy; early (63 to 130 days), mid (118-180 days) and late (163-252 days). Swabs were immediately placed in Amies transport media and stored at -80°C.

Determination of secretor status
Extraction of DNA from vaginal swabs was performed as previously described.  
Individuals were genotyped by sequencing the coding part of exon 2 of the FUT2 gene. The exon was amplified using primers 5'-CCATATCCCAGCTAACGTGTCC-3' and 5'-GGGAGGCAGAGAAGGAAAGG-3 and the amplicons sequenced with the PacBio Sequel system. Primer sequences were removed from the CCS reads using Dada2 and the trimmed sequences mapped to a human reference FUT2 sequence (derived from HG38) using Minimap. To limit the effect of any reference bias, we generated consensus sequences using the “ALT” allele from the mapping and reads were remapped to this sequence. Finally, variants were called using Freebayes and phased using WhatsHap.

Individuals that were homozygous for any one of the four known non-secretor mutations (se302 (P101L, rs200157007), se385 (I129F, rs1047781), se428 (W413X, rs601338) and se571(R191X, rs18000028)) were inferred to be phenotypically non-secretor. Heterozygotes and wildtype homozygotes were classified as secretors.

**Sequencing of 16S rRNA gene amplicons and assembly**

The V1-V2 hyper variable regions of bacterial 16S rRNA genes were amplified with using the forward primer set (28f-YM) consisting of a mixture of the following primers mixed at a 4:1:1:1 ratio; 28F-Borrelia GAGTTTGATCCTGGCTTAG; 28F-Chlorflex GAATTTGATCCTGGCTCAG; 28F-Bifido GGGTTCGATTCTGGCTCAG; 28F GAGTTTGATCCTGGCTCAG. The reverse primer consisted of; 388R TGCTGCTCCGCTTAGGAGT. Primer sequences were trimmed using Cutadapt, performed QC using FastQC and calculated ASV counts per sample using the Qiime2 pipeline. We used DADA2 for denoising and taxonomically classified sequences to species level using the STIRRUPS reference database.
Data analyses

Generalised linear mixed effects gamma regression models for each timepoint in pregnancy were used to regress gestational length (in days) on secretor status and the vaginal microbiome (Lactobacillus status or CST) as well as age, BMI, previous PTB/MTL, cervical stitch and Previous cervical excisional treatment in glmmTMB. We used ethnicity as a random effect: as the sample size for the East Asian and Middle East North African and Mixed/Other groups was small (less than 10) these were combined into a single group (Mixed/Other). We applied a reflection transformation to the gestational length prior to fitting the gamma GLMM as this variable was negatively skewed. Model diagnostics were examined with scaled residuals simulated from the fitted models using DHARMa. We used the emmeans package to calculate the estimated marginal means from the models to look for statistically significant comparisons between groups: this post hoc method automatically adjusts for multiple comparisons.

We retained taxa that occurred at greater than 0.5% abundance in two or more samples. Differential abundance analyses were performed using ALDEx2 to compare taxon abundances in the vaginal microbiome in early pregnancy between secretors and non-secretors. This method uses a compositionally robust approach to perform DAA: microbiome count data are considered compositional and thus have to be transformed to permit analysis by multivariate statistics. Count data were transformed with the centred log-ratio (CLR) and we ran 1000 Monte-Carlo instances to estimate effect sizes and perform a Welch’s t test to compare our two conditions. We also analysed the compositional differences between phenotypes using PCAs: zeros in the count data were imputed using a Bayesian multiplicative approach in the zCompositions package followed by CLR-transformation. To test whether the
centroids of the two secretor groups differ we performed a PERMANOVA on the CLR-
transformed distance matrix (after checking for homogeneity of dispersion) using the
Vegan package \textsuperscript{70}.

Finally, we used BAnOCC to infer co-occurrence networks for the microbial taxa
observed in secretors and non-secretors in early pregnancy \textsuperscript{71}. This program uses a
Bayesian framework to analyse compositional covariance. We ran the MCMC for
10000 generations and four chains to reach convergence. We used a 95\% credible
interval to identify significant correlations. Qgraph was used to draw the networks
based on these Banocc-derived adjacency matrices and to calculate the Expected
Influence network statistic, which is a measure of degree centrality for signed networks
\textsuperscript{34,72}.


occurrence network reveals interconnection pattern across microbiomes.

Microbiome. 2020 Dec 4;8(1):82.


533  Bioinformatics. 2018 Sep 15;34(18):3094–100.
534  55. Garrison E, Marth G. Haplotype-based variant detection from short-read
535  sequencing. 2012 Jul 17;
537  WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing
540  Molecular basis for erythrocyte Le(a+b+) and salivary ABH partial-secretor
541  phenotypes: expression of a FUT2 secretor allele with an A-T mutation at
542  nucleotide 385 correlates with reduced a(1,2) fucosyltransferase activity.
544  58. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical
545  Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S
547  59. Martin M. Cutadapt removes adapter sequences from high-throughput
548  sequencing reads. EMBnet.journal. 2011 May 2;17(1):10.
549  60. Andrew S. FastQC: A Quality Control Tool for High Throughput Sequence
550  Data. 2010.
552  Reproducible, interactive, scalable and extensible microbiome data science
554  62. Fettweis JM, Serrano MG, Sheth NU, Mayer CM, Glascock AL, Brooks JP, et
555  al. Species-level classification of the vaginal microbiome. BMC Genomics.
556  2012 Dec 17;13(S8):S17.


68. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8(NOV):1–6.

69. Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst. 2015 Apr;143:85–96.


Figure Legends

Figure 1: A. Distribution of the top 10 most abundant bacterial genera in the vaginal microbiota of secretors and non-secretors in early, mid and late pregnancy. B. The distribution of the proportion of *Lactobacillus* in secretors and non-secretors in early, mid and late pregnancy was similar. C. Frequencies of each Community State Type (CST) through pregnancy in secretors and non-secretors.

Figure 2: A. Co-occurrence networks showing the vaginal microbiota in early pregnancy in secretors (left) and non-secretors (right). Nodes represent bacterial taxa present in the 16S relative abundance data (above our low abundance threshold of 0.5% in two or more samples), and have been grouped into *L. crispatus*, *L. gasseri*, *L. iners*, *L. jensenii*, other *Lactobacillus*, BV-associated bacteria, Pathobionts, and Other, as defined by previously. Statistically significant edges, indicating correlations, are coloured and weighted according to their r value: red and green, respectively indicate negatively and positively correlations. In both secretors and non-secretors, many of the positive associations occur between BV-associated microbes, reflecting the typically polymicrobial nature of BV. B. Flow diagrams (derived from the network analyses) centred on the *L. crispatus*, *L. gasseri*, *L. iners* and *L. jensenii* nodes indicate that these tend to be more negatively correlated with other bacteria in the vaginal microbiota of the non-secretors compared with the secretors with *L. crispatus* having the greatest proportion of negative edges, particularly in the non-secretors.

Figure 3: The association between secretor status, the vaginal microbiome and gestational length (days). A red dashed line highlights the 37 weeks (259 days) threshold for prematurity and sample sizes are indicated at the top of the plots. A. The distribution of gestational length by secretor status and vaginal microbiome shows non-secretors with *Lactobacillus* deplete microbiomes in early pregnancy tend to have
shorter gestation but this relationship disappears by mid pregnancy. B. Classifying the microbiome by Community State Type (CST) rather than by *Lactobacillus* status indicates that reduced gestation in *Lactobacillus* deplete non-secretors is mainly driven by women with CST 2 and high diversity CST 4 microbiomes.
Table 1. Unstandardised coefficients ($b$), standard error (SE) and 95% confidence intervals (CI) from gamma generalised linear mixed effects modelling (GLMM) of gestational length (in days) over three time points in pregnancy (with ethnicity as a random effect). Early, Mid and Late pregnancy GLMMs using Lactobacillus status (dominated/depleted), as well as an Early Community State Type (CST) GLMM, which replaces the Lactobacillus covariate with CST. Secretors with Lactobacillus dominated microbiota (or CST 1) are baseline in the model. *1 Preterm birth. *2 Mid-trimester loss. * p<0.05. ** p<0.01. *** p<0.001.
<table>
<thead>
<tr>
<th>Secretor × CST 4</th>
<th>NA</th>
<th>NA</th>
<th>NA</th>
<th>NA</th>
<th>NA</th>
<th>NA</th>
<th>NA</th>
<th>0.85 (0.35)†</th>
<th>0.17, 1.52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretor × CST 5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-0.11 (0.43)</td>
<td>-0.95, 0.73</td>
</tr>
</tbody>
</table>
Figure 1
Figure 2
Figure 3