Supporting Information

Variations in the Cervicovaginal Microbiota and Metabolite Profiles of Pregnant Women with and without Short Cervix

Xiaoxiu Huang MD[1#], Miaomiao Wang PhD[1#], Meng Hong PhD[2#], Mengni Wang PhD[1], Na Li MD[1], Baohua Li PhD[1,3*]

1 Department of Obstetrics, Women's Hospital, Zhejiang University School of Medicine, Xueshi Rd No. 1, Hangzhou 310006 Zhejiang, People's Republic of China

2 State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, National Medical Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

3 Zhejiang Key Laboratory of Maternal and Infant Health, Hangzhou, Zhejiang Province.

Emails:

Xiaoxiu Huang: 5517017@zju.edu.cn

Miaomiao Wang: 11618227@zju.edu.cn

Meng Hong: hongmeng@zju.edu.cn

Mengni Wang: 11618394@zju.edu.cn

Na Li: lina83875170@zju.edu.cn

Baohua Li: lbh19787@zju.edu.cn

[#] These authors contributed to the work equally and should be regarded as co-first authors

***Corresponding to**: Baohua li, PhD, Department of Obstetrics, Women's Hospital, School of Medicine, Zhejiang University, Xueshi Rd No. 1, Hangzhou 310006 Zhejiang, People's Republic of China. Email: lbh19787@zju.edu.cn. Tel: (+86) 571-87061501. Fax: (+86) 571-87061878. https://orcid.org/0000-0002-2205–3248.

Supplemental Methods

Exclusion criteria

The exclusion criteria were applied to all groups, including women with prior cervical surgery, evidence of premature rupture of membranes, uterine malformation, the presence of fetal anomalies, intrauterine fetal demise, recent drug use (<4 weeks) including chronic steroid use, probiotic therapy, antibiotic treatment, sexual activity, vaginal symptoms consistent with infection at the time of recruitment, multiple gestations, and symptoms such as persistent contractions or vaginal bleeding.

Protocol for managing patients with a short cervix

Women with a short cervix were observed (n = 2) or underwent cervical cerclage (n = 16) and were prescribed vaginal progesterone (n = 15) for sPTB prevention at the discretion of the attending physician. Vaginal progesterone therapy included the use of soft progesterone capsules twice daily (in the morning and at night; each pill contained 100 mg of progesterone; Besins Manufacturing Belgium, France) until >34 wks. All cerclage procedures were performed by Doctor Li using the McDonald

technique (transvaginal cerclage with a double 10 silk suture placed at the cervicovaginal junction without bladder mobilization). Cervical cerclages were removed at 36–37 wks; however, if patients presented with cervical changes, painful contractions, or progressive vaginal bleeding, the cerclage was removed at that time.

***DNA* extraction**

DNA extraction was performed within one month after sample collection using the QIAamp DNA Mini Kit (QIAGEN, Germany) according to the manufacturer's instructions. The DNA concentration and purity were monitored on 1% agarose gels. The concentration of bacterial DNA was measured using a Nanodrop 2000 (Thermo Scientific, USA). The extracted DNA was stored at -80 °C until analysis by PCR.

PCR amplification and 16S rDNA sequencing

Bacterial genomic DNA was amplified with the primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3'), which are specific to the V3-V4 hypervariable regions of the 16S rRNA gene. The 5' ends of the primers were tagged with barcodes unique to each sample, and sequencing was performed with universal primers. PCR amplification was performed in a total reaction volume of 25 μL containing 25 ng of template DNA, 12.5 μL of PCR Premix, 2.5 μL of each primer, and the remaining volume of PCR-grade water. The PCR conditions used to amplify the prokaryotic 16S fragments consisted of initial denaturation at 98 °C for 30 s; 32 cycles of denaturation at 98 °C for 10 s, annealing at 54 °C for 30 s, and extension at 72 °C for 45 s; and a final extension step at 72 °C for 10 min. The PCR products were confirmed via 2% agarose gel electrophoresis. Throughout the DNA extraction process, ultrapure water was used instead of a sample mixture to exclude the possibility of false-positive results as a negative control. The PCR products were purified with AMPure XP beads (Beckman Coulter Genomics, Danvers, MA, USA) and quantified by Qubit (Invitrogen, USA). The amplicon pools were prepared for sequencing, and the size and quantity of the amplicon library were assessed on an Agilent 2100 Bioanalyzer (Agilent, USA) with a Library Quantification Kit for Illumina (Kapa Biosciences, Woburn, MA, USA). The libraries were sequenced on a NovaSeq PE250 platform.

Sequencing, paired-end read assembly and quality control

The samples were sequenced on an Illumina NovaSeq 6000 platform according to the manufacturer's recommendations (LC-Bio). Paired-end reads were assigned to the samples on the basis of their unique barcode and then truncated by removing the barcode and primer sequences. Paired-end reads were merged using FLASH. Quality filtering of the raw reads was performed under specific filtering conditions to obtain high-quality clean tags using fqtrim (v0.94). Chimeric sequences were filtered using Vsearch software (v2.3.4). After dereplication using DADA2, a feature table and feature sequence were generated. Alpha and beta diversity were calculated via random normalization to the same sequences.

Metabolite extraction

The frozen samples were thawed on ice, extracted with 1 ml of precooled 50% methanol, vortexed for 1 min, and incubated at room temperature for 10 min; then, the

extraction mixture was stored overnight at -20 °C. After centrifugation at 4,000 × g for 20 min, the supernatants were transferred to 96-well plates. Additionally, pooled quality control (QC) samples were prepared by combining 10 μL of each extraction mixture.

LC conditions

All samples were analyzed by LC–MS according to the instructions of the instrument. Reversed-phase chromatographic separation was performed using a Vanquish Flex UHPLC system (Thermo Fisher Scientific, Bremen, Germany) and an ACQUITY UPLC T3 column (100 mm×2.1 mm, 1.8 μm, Waters, Milford, USA). Additionally, the column oven was maintained at 35 °C, and the flow rate was 0.4 ml/min. The mobile phase consisted of solvent A (water, 0.1% formic acid) and solvent B (acetonitrile, 0.1% formic acid), and gradient elution was performed as follows: 0–0.5 min, 5% B; 0.5–7 min, 5% to 100% B; 7–8 min, 100% B; 8–8.1 min, 100% to 5% B; and 8.1–10 min, 5% B.

MS conditions

Chromatographically separated metabolites were subjected to MS detection using a Q Exactive high-resolution tandem mass spectrometer (Thermo Scientific) in both positive and negative ion modes. Precursor spectra (70–1050 m/z) were collected at a resolution of 70,000 to reach an automatic gain control (AGC) target of 3e6. The maximum injection time was 100 ms, and data were acquired with a top 3 configuration in data-dependent acquisition (DDA) mode. Fragment spectra were collected at a resolution of 17,500 to reach an AGC target of 1e5 with a maximum injection time of 80 ms. To evaluate the stability of the LC–MS system throughout the experiment, a pooled QC sample was analyzed after every 10 samples.

Statistical analyses

Alpha diversity was used to analyze the complexity of species diversity in a sample using the Chao1, Shannon and Simpson indices, each of which were calculated with QIIME 2 (version 2022.11) based on rarefied ASV counts and displayed using R (version 3.5.2). Beta diversity analysis was used to evaluate differences in species complexity between samples, and beta diversity (using the Bray–Curtis distance) was calculated using QIIME 2 on the basis of the rarefied ASV counts. PCoA was performed to determine the principal coordinates and visualize complex and multidimensional data. The PCoA results were displayed using the WGCNA, stats and ggplot2 packages in R. Differences between the two groups were assessed on the basis of a distance matrix using the nonparametric multivariate analysis test ANOSIM with the Vegan package in R.

The features of the cervicovaginal microbiota differentiating the groups were characterized via LEfSe for biomarker discovery. On the basis of the normalized relative abundance matrix, the Kruskal–Wallis rank-sum test was used to identify specific taxa with significantly different abundance, and LDA was performed to estimate the effect size of each feature. An significance level (alpha value) of 0.05 and an effect size threshold of 3 were applied for all biomarkers in this study. The differentially abundant taxa were analyzed at the phylum, class, order, family, genus and species levels using the Wilcoxon rank-sum test. The identified differentially

abundant genera and metabolite were determined to be significantly correlated by Spearman's rank correlation analysis (Student's t test, $p < 0.05$, |correlation coefficient| > 0.6). The correlations and corresponding attributes within the heatmap and network model were visualized with Cytoscape (version 3.2.1).

The metabolite data were preprocessed conducted using XCMS software. The LC−MS raw data files were converted into mzXML format and then processed by the XCMS, CAMERA and metaX toolbox in R. Each ion was identified by a combination of the retention time (RT) and m/z data. The intensity of each peak was recorded, and three-dimensional matrices containing arbitrarily assigned peak indices (retention time–m/z pairs), sample names (observations) and ion intensity information (variables) were generated. Metabolites were annotated using the KEGG and HMDB by matching the molecular mass (m/z) of the samples with the values in the databases. Metabolites with a mass difference less than 10 ppm from that reported in the database were annotated, and the molecular formula was validated according to isotopic distribution. An in-house fragment spectrum library was also used to confirm metabolite identification. The peak intensities were further preprocessed by metaX. Features detected in less than 50% of the QC samples or 80% of the biological samples were removed, and the remaining peaks with missing values were imputed with the k-nearest neighbor algorithm to further improve the data quality. Quality control-based robust LOESS signal correction was applied to the QC data with respect to the order of injection to minimize signal intensity drift over time. In addition, the relative standard deviations of the metabolic features were calculated across all the QC samples, and those >30% were removed. Supervised PLS-DA was used to visualize the maximum differences between the two groups, with the goodness of fit quantified by R^2Y and the predictive ability expressed by Q^2. Independent samples t tests were performed to assess the significance of each metabolite. Finally, VIP values were used to select important features, with a cutoff of 1.0.

Table S1. Differentially abundant cervicovaginal metabolites identified on the basis of a *p* value < 0.05, VIP ≥ 1, and FC threshold of 2 between patients with and without short cervix

No.	Metabolites	VIP	*ratio*	*p*-value
1	(R)-3-Hydroxybutyric acid (neg-M103T115)	2.68	4.49	0.001
2	(Z)-9-Cycloheptadecen-1-one (pos-M251T308)	1.38	0.38	0.046
3	.alpha.,.beta.-Trehalose (neg-M387T51)	1.54	2.08	0.016
4	12,13-DiHODE (neg-M311T255)	1.57	4.82	0.029
5	12-Hydroxystearic acid (neg-M299T329)	2.20	0.27	0.037
6	13,14-Dihydro PGF2a (neg-M355T221)	2.30	3.93	0.010
7	**13,14-Dihydro-15-ketoprostaglandin A2 (pos-M335T254)**	**1.82**	**3.14**	**0.007**
8	13-HODE (neg-M295T288)	1.91	10.84	0.028
9	16-Hydroxyhexadecanoic acid (neg-M271T285)	2.22	0.21	0.017
10	17,18-DiHETE (neg-M335T242)	2.47	2.46	<0.001
11	2',6'-Dihydroxy-4-methoxychalcone-4'-O-neohesperid (neg-M593T65)	1.56	2.19	0.048

12	20-Hydroxyarachidonic acid (pos-M321T323)	1.80	2.09	0.014
13	2'-Deoxyuridine 5'-monophosphate (neg-M307T63)	1.59	0.40	0.006
14	3-(8,11,14-Pentadecatrienyl)phenol (neg-M343T297)	2.63	5.42	0.004
15	3.alpha.,4.beta.-Galactotriose, neg-M503T67)	2.03	3.26	0.005
16	3-carboxy-4-methyl-5-pentyl-2-furanpropanoic acid (neg-M267T236)	1.44	2.02	0.009
17	3-Hydroxybenzoic acid (neg-M137T208)	5.39	0.04	<0.001
18	4-Phenyl-1H,3H-naphtho[1,8-cd]pyran-1,3-dione (pos-M275T90)	2.33	0.26	0.027
19	5.alpha.-Androstan-3.beta.-ol-17-one sulfate (neg-M369T255)	2.39	4.26	0.002
20	5.alpha.-Pregnan-3.alpha.,17-diol-20-one 3-sulfate (neg-M413T208)	2.84	3.21	<0.001
21	5-Hexyltetrahydro-2-furanoctanoic acid (neg-M297T304)	1.85	0.43	0.037
22	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid (neg-M319T302)	2.77	4.90	0.002
23	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid, 1,5-lactone (pos-M303T302)	2.38	6.33	0.010
24	5-KETE (neg-M317T310)	2.42	3.01	0.008
25	8Z,14Z-Eicosadienoic acid (pos-M331T333)	2.37	5.93	0.007
26	9,10,13-TriHOME (neg-M329T207)	1.46	3.12	0.023
27	9-Oxo-10(E),12(E)-octadecadienoic acid (neg-M293T299)	1.77	5.03	0.018
28	9-Oxo-10(E),12(E)-octadecadienoic acid (pos-M295T252)	1.39	2.73	0.031
29	9-Oxo-10(E),12(E)-octadecadienoic acid (pos-M295T207)	1.28	2.22	0.026
30	9S,11R,15S-Trihydroxy-20a,20b-dihomo-5Z,13E-prostadienoic acid (neg-M381T234)	2.42	6.31	0.005
31	Acylcarnitine 10:1 (pos-M314T192)	2.25	3.99	0.009
32	Asp-Val (neg-M231T115)	1.91	0.36	0.004
33	Cannabidiolic acid (pos-M359T272)	2.17	3.15	0.010
34	Cardanoldiene (neg-M345T342)	1.90	2.28	0.016
35	Cardanolmonoene (neg-M347T331)	2.59	7.70	0.009
36	cis-5,8,11,14-Eicosatetraenoic acid (pos-M305T313)	2.19	4.61	0.014
37	cis-9-Hexadecenoic acid (pos-M255T285)	2.01	0.30	0.020
38	Cytokinin B (pos-M248T163)	2.35	2.95	<0.001
39	D-erythro-Sphingosine C-20 (pos-M328T426)	1.54	2.15	0.024
40	DP7 (pos-M1153T53)	1.93	3.37	0.008
41	D-Ribulose 5-phosphate (neg-M229T49)	1.59	2.06	0.022
42	D-Sorbitol (pos-M183T52)	2.49	0.34	0.002
43	Ethyl 4-hydroxybenzoate (pos-M167T209)	2.41	0.27	0.006
44	Ethyl formate (neg-M73T66)	1.44	0.46	0.049
45	Ethyl salicylate (neg-M165T209)	8.24	0.00	<0.001
46	Ginkgolic acid I (neg-M345T307)	1.69	3.53	0.026
47	Glycocholic acid (neg-M464T309)	2.36	0.43	0.001
48	Hepoxilin B3 (neg-M335T271)	2.61	4.94	0.007

49	Hippuric acid (neg-M178T167)	1.95	12.56	0.031
50	Indoxyl sulfate (neg-M212T167)	2.12	7.30	0.023
51	LysoPI 16:1; LysoPI 16:1 (neg-M569T362)	1.75	2.48	0.031
52	Maltitol (neg-M343T51)	2.02	0.43	0.030
53	Maltitol (pos-M345T51)	2.48	0.35	0.006
54	Maltohexaose (pos-M991T53)	1.68	3.01	0.021
55	Maltopentaose (pos-M829T53)	1.58	2.51	0.032
56	Maltotetraose (pos-M667T52)	2.06	3.53	0.010
57	Maltotetraose (pos-M684T75)	1.99	3.23	0.016
58	Maltotetraose (pos-M667T78)	1.83	2.81	0.029
59	Maltotetraose (pos-M684T53)	1.78	2.50	0.004
60	Maltotriose (neg-M549T67)	2.12	5.05	0.011
61	Maltotriose (neg-M549T52)	2.18	3.34	0.003
62	Maltotriose (pos-M505T51)	1.74	2.77	0.024
63	Mannitol (neg-M181T50)	3.92	0.16	<0.001
64	Methionine sulfoxide (pos-M166T50)	1.96	0.43	0.008
65	Nadolol (pos-M310T183)	2.31	5.44	0.008
66	Nebularine (neg-M251T92)	2.42	0.28	0.032
67	Nebularine (neg-M251T54)	2.95	0.21	0.003
68	O-methoxycatechol-O-sulphate (neg-M203T179)	1.42	2.43	0.044
69	Palatinose (neg-M341T51)	1.68	2.26	0.007
70	p-Cresol (neg-M107T180)	1.89	4.53	0.011
71	p-Cresol sulfate (neg-M187T180)	2.52	9.73	0.016
72	Phenol (neg-M93T208)	4.74	0.09	<0.001
73	Phenol sulphate (neg-M173T164)	2.20	6.55	0.032
74	Phenylacetyl-L-glutamine (neg-M263T164)	2.22	5.93	0.012
75	Phenylacetyl-L-glutamine (pos-M265T164)	2.18	4.97	0.010
76	Propionic acid (neg-M73T105)	1.57	0.46	0.042
77	Raffinose (pos-M522T52)	1.82	2.33	0.004
78	Succinic acid (neg-M117T104)	1.62	0.47	0.035
79	Succinic acid (neg-M117T83)	1.58	0.45	0.044
80	Succinic acid (neg-M117T66)	1.56	0.37	0.036
81	Tetradecylamine (pos-M214T214)	2.98	2.53	<0.001
82	trans-.DELTA.2-11-Methyldodecenoic acid (pos-M213T333)	1.58	0.46	0.026
83	trans-Vaccenic acid (pos-M283T333)	2.17	0.31	0.031
84	Trehalose (pos-M360T50)	1.74	2.97	0.020
85	Trehalose (pos-M343T51)	1.45	2.19	0.044
86	Uric acid (neg-M167T80)	1.86	2.09	0.017
87	Xanthine (pos-M153T91)	2.26	3.71	0.031
88	Xanthine (neg-M151T91)	2.71	3.40	0.006

161
162

Table S2. KEGG enrichment analysis of the differentially abundant cervicovaginal metabolites between patients with and without short cervix

Super pathway	Sub pathway	Differential metabolites	*p*-value
Amino acid metabolism	Phenylalanine metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M178T167;neg-M263T164;pos-M265T164	0.001
	Tyrosine metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M93T208	0.013
	Alanine, aspartate and glutamate metabolism	neg-M117T104;neg-M117T66;neg-M117T83	0.014
	Phenylalanine, tyrosine and tryptophan biosynthesis	neg-M137T208	0.021
Biosynthesis of other secondary metabolites	Caffeine metabolism	neg-M151T91;pos-M153T91	0.008
Carbohydrate metabolism	Butanoate metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M103T115	0.002
	Fructose and mannose metabolism	neg-M181T50;pos-M183T52	0.005
	Galactose metabolism	pos-M522T52;pos-M183T52	0.003
	Propanoate metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M73T105	0.003
	Citrate cycle (TCA cycle)	neg-M117T104;neg-M117T66;neg-M117T83	0.007
	Pentose phosphate pathway	neg-M229T49	0.021
	Pyruvate metabolism	neg-M117T104;neg-M117T66;neg-M117T83	0.017
	Starch and sucrose metabolism	neg-M387T51;pos-M343T51;pos-M360T50	0.024
Energy metabolism	Oxidative phosphorylation	neg-M117T104;neg-M117T66;neg-M117T83	0.005
	Sulfur metabolism	neg-M117T104;neg-M117T66;neg-M117T83	0.018
	Carbon metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M229T49	0.035
Lipid metabolism	**Arachidonic acid metabolism**	pos-M305T313;neg-M317T310;pos-M321T323;neg-M335T271;	<0.001
	Linoleic acid metabolism	pos-M305T313;neg-M329T207	0.001
	Biosynthesis of unsaturated fatty acids	pos-M305T313	0.047
	Fatty acid biosynthesis	pos-M255T285	0.041
	Glycerophospholipid metabolism	neg-M569T362	0.044
	Primary bile acid biosynthesis	neg-M464T309	0.037
	Synthesis and degradation of ketone	neg-M103T115	0.001

	bodies		
Metabolism of cofactors and vitamins	Nicotinate and nicotinamide metabolism	neg-M117T104;neg-M117T66;neg-M117T83;neg-M73T105	0.005
	Riboflavin metabolism	neg-M229T49	0.007
	Vitamin B6 metabolism	neg-M229T49	0.014

163

164

Table S3. Differential relative abundance of the cervicovaginal microbiota at the genus level between patients with and without short cervix (p value < 0.05)

No.	Genes	Log2 fold change	p-value
1	g__Ruminococcus]_gnavus_group	Inf	<0.001
2	g__Bacteroides	3.14	<0.001
3	g__Lachnospiraceae_NK4A136_group	-Inf	<0.001
4	g__Clostridia_vadinBB60_group_unclassified	-5.43	<0.001
5	g__Ruminococcaceae_unclassified	-Inf	<0.001
6	g__Arthrobacter	Inf	<0.001
7	g__Clostridia_UCG-014_unclassified	-3.43	<0.001
8	g__Muribaculaceae_unclassified	-Inf	<0.001
9	g__Negativibacillus	-5.66	<0.001
10	g__Tannerellaceae_unclassified	-5.92	<0.001
11	g__Flavonifractor	4.95	<0.001
12	g__Lachnospiraceae_unclassified	-5.63	<0.001
13	g__Pyramidobacter	Inf	<0.001
14	g__Rikenella	-Inf	<0.001
15	g__Ruminococcus]_torques_group	Inf	<0.001
16	g__Christensenellaceae_unclassified	-Inf	<0.001
17	g__Enterobacter	2.87	<0.001
18	g__Enterococcus	4.04	<0.001
19	g__Eubacterium]_coprostanoligenes_group_unclassified	-2.74	<0.001
20	g__Eisenbergiella	-Inf	<0.001
21	g__Elizabethkingia	Inf	<0.001
22	g__Ellin6067	Inf	<0.001
23	**g__Ureaplasma**	**6.45**	**0.01**
24	g__Fusobacterium	0.51	0.01
25	g__Firmicutes_unclassified	-3.64	0.01
26	g__UCG-010_unclassified	-Inf	0.01
27	g__Incertae_Sedis	2.18	0.01
28	g__Subdoligranulum	-4.21	0.01
29	g__Lactococcus	5.05	0.01
30	g__Weissella	4.66	0.01
31	g__4-29-1_unclassified	2.09	0.01
32	g__Parasutterella	-3.52	0.01

33	g__Rothia	-0.90	0.02
34	g__Peptococcus	-Inf	0.02
35	g__BSV26_unclassified	Inf	0.02
36	g__Lachnoclostridium	2.43	0.02
37	g__Anaerococcus	3.60	0.02
38	g__Thermodesulfovibrionia_unclassified	2.77	0.02
39	g__Trichococcus	-1.16	0.02
40	g__Ligilactobacillus	-5.31	0.02
41	g__Escherichia-Shigella	1.16	0.02
42	g__Veillonella	-0.65	0.03
43	g__UCG-002	1.42	0.04
44	g__Moraxella	-Inf	0.04
45	g__Christensenellaceae_R-7_group	-Inf	0.04
46	g__Candidatus_Koribacter	Inf	0.04
47	g__Acidibacter	Inf	0.04
48	g__Peptococcaceae_unclassified	-Inf	0.04
49	g__Bacillus	-Inf	0.04
50	g__ADurb.Bin063-1	-Inf	0.04
51	g__SC-I-84_unclassified	2.91	0.04
52	g__Burkholderia-Caballeronia-Paraburkholderia	1.96	0.04
53	g__Ralstonia	1.66	0.04
54	g__Streptococcus	-2.72	0.04
55	g__Stenotrophomonas	2.61	0.04

Table S4. Spearman's correlation analysis between the differentially abundant genera and cervicovaginal metabolites, with a correlation coefficient of $|r| > 0.6$ and $p < 0.05$

Cervicovaginal microbiota	Cervicovaginal metabolites	rho	*p*-value	relation
g__Clostridia_UCG-014_unclassified	Cytokinin B	-0.61	< 0.001	negative
g__Ruminococcus]_torques_group	17,18-DiHETE	0.63	< 0.001	positive
g__Ruminococcus]_torques_group	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid, 1,5-lactone	0.63	< 0.001	positive
g__Ruminococcus]_torques_group	Ginkgolic acid I	0.60	< 0.001	positive
g__Ruminococcus]_torques_group	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid	0.63	< 0.001	positive
g__Ruminococcus]_torques_group	8Z,14Z-Eicosadienoic acid	0.62	< 0.001	positive
g__Ruminococcus]_torques_group	Cardanolmonoene	0.60	< 0.001	positive
g__Enterobacter	4-Phenyl-1H,3H-naphtho[1,8-cd]pyran-1,3-dione	-0.63	< 0.001	negative

g__Ureaplasma	**13,14-Dihydro-15-ketoprostaglandin A2**	**0.63**	**< 0.001**	**positive**
g__Rothia	3-Hydroxybenzoic acid	0.61	< 0.001	positive
g__Rothia	Ethyl salicylate	0.62	< 0.001	positive
g__Trichococcus	Cannabidiolic acid	-0.65	< 0.001	negative
g__Trichococcus	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid, 1,5-lactone	-0.60	< 0.001	negative
g__Trichococcus	Ginkgolic acid I	-0.64	< 0.001	negative
g__Trichococcus	5-Hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid	-0.64	< 0.001	negative
g__Trichococcus	20-Hydroxyarachidonic acid	-0.61	< 0.001	negative
g__Ralstonia	Glycocholic acid	-0.64	< 0.001	negative

167

Table S5. Patients' demographic and clinical characteristics of patients with a short cervix

	All Women	Patients with sPTB <34 weeks (N=5)	Patients with gestations ≥34 weeks (N=13)	*p*-Value
Age at sampling (years), mean ± SD	31.4±4.4	29.6±4.5	31.2±4.3	0.283
Pre-gestational BMI, mean ± SD	21.62±2.98	20.85±2.01	21.91±3.30	0.512
BMI at sampling, mean ± SD	23.25±2.54	22.08±1.76	23.7±2.71	0.235
Smoking, n (%)	1 (5.6)	1 (20.0)	0(0.0)	0.278
Nulliparity, n (%)	12 (66.7)	4 (80.0)	8(61.5)	0.615
History of sPTB or mid-trimester pregnancy loss	6 (33.3)	1 (20%)	5 (38.5)	0.615
Gestational age at sampling (weeks), median (IQR)	20 (19, 22)	22 (19, 22)	20 (19,22.5)	0.801
Length of the cervix at sampling (mm), mean ± SD	10.4±7.7	9.4±8.6	10.7±7.6	0.771
Gestational diabetes mellitus, n (%)	5 (27.8)	1 (20.0)	4(30.8)	1.000
Preeclampsia, n (%)	3 (16.7)	0 (0.0)	3(23.1)	0.522
Cerclage, n (%)	16 (88.9)	4 (80)	12 (92.3)	0.490

Vaginal progesterone use, No. (%)	15(83.3)	4 (80.0)	11(84.6)	1.000
Gestational age at delivery (weeks), median (IQR)	36.5 (32, 38)	30 (26, 32)	37 (36, 38)	0.001
Birth weight (g), (IQR)	2890 (1770, 3263)	1590 (685, 1740)	3180 (2640, 3285)	0.001
Apgar score at 1 min, median (IQR)	10 (9, 10)	8 (0, 9)	10 (10, 10)	0.001
Apgar score at 5 min, median (IQR)	10 (9, 10)	9 (0, 10)	10 (10, 10)	0.015
NICU, n (%)	6 (33.3)	3 (60.0)	3 (23.1)	0.268

BMI, body mass index; sPTB, spontaneous preterm delivery; NICU, Neonatal intensive care unit; IQR, interquartile range.

Fig S1

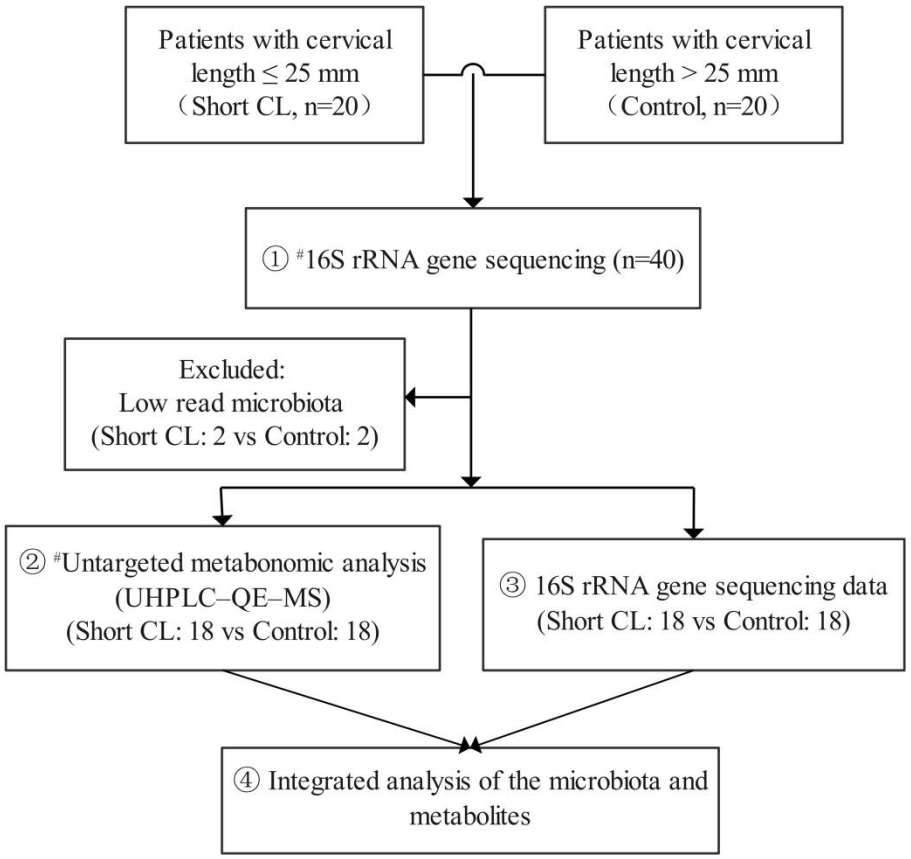

Fig S1. Flowchart explaining the recruitment of patients with a short cervix and controls for this study. ① To investigate the cervicovaginal microbiota and metabolites in patients with a short cervix, we analyzed 40 cervicovaginal samples using 16S rRNA gene sequencing, and 4 subjects were excluded from subsequent analysis because the number of reads obtained after sequencing was low (n = 2 and n = 2, Short CL vs Control). Ultimately, 36 subjects (n = 18 and n = 18, Short CL vs Control) remained for subsequent statistical analyses. ② To fully understand the functional activity of the cervicovaginal microbiota, we conducted untargeted metabonomic analysis on the 36 cervicovaginal samples. ③ The 16S rRNA gene sequencing data of the same 36 cervicovaginal samples were used for untargeted metabonomic analysis, subjected to statistical analyses of the microbial composition and prepared for further integrated analysis of the microbiota and metabolites. ④ Finally, an integrated analysis of the microbiota and metabolites of the 36 cervicovaginal samples (n=18 and n=18, Short CL vs Control) was performed. Short CL, patients with a cervical length ≤25 mm; Control, patients with a cervical length >25 mm; #, 16S rRNA gene sequencing and the metabolomics study were carried out at the same time.

Fig S2

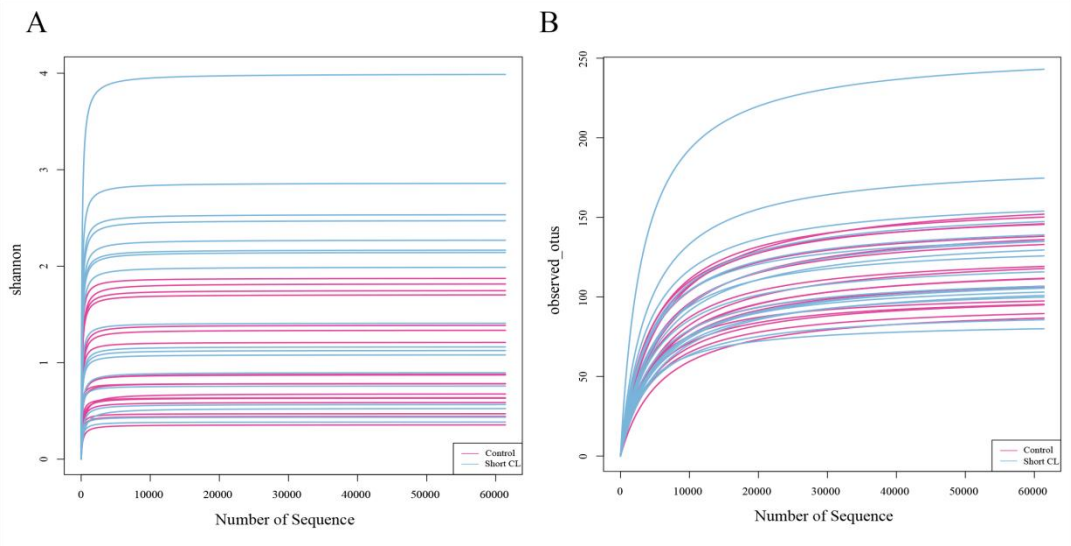

Fig S2. Assessment of 16S rRNA gene sequencing coverage. Rarefaction curves were used to evaluate whether the sequencing coverage was sufficient to cover all the taxonomic groups optimally; if the sequencing depth is sufficient, the curves will reach a plateau. The Shannon index and Good's coverage rarefaction curves plateaued, indicating adequate sequencing depth and capture of the majority of the microbial diversity in the samples.

Fig S3

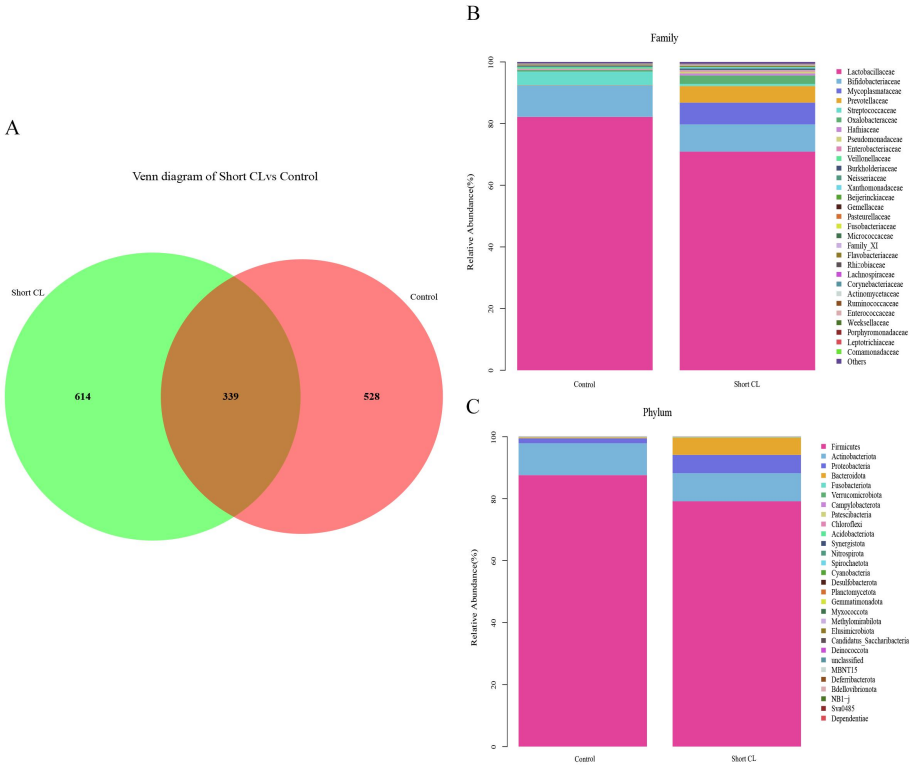

Fig S3. (A) Venn diagram illustrating the numbers of ASVs in patients with a short cervix and control patients. (B, C) The cervicovaginal bacterial families and phyla clustered into different groups on the basis of relative abundance, and the microbiota compositions significantly differed. Only the top 30 communities were included in this analysis.

Fig S4

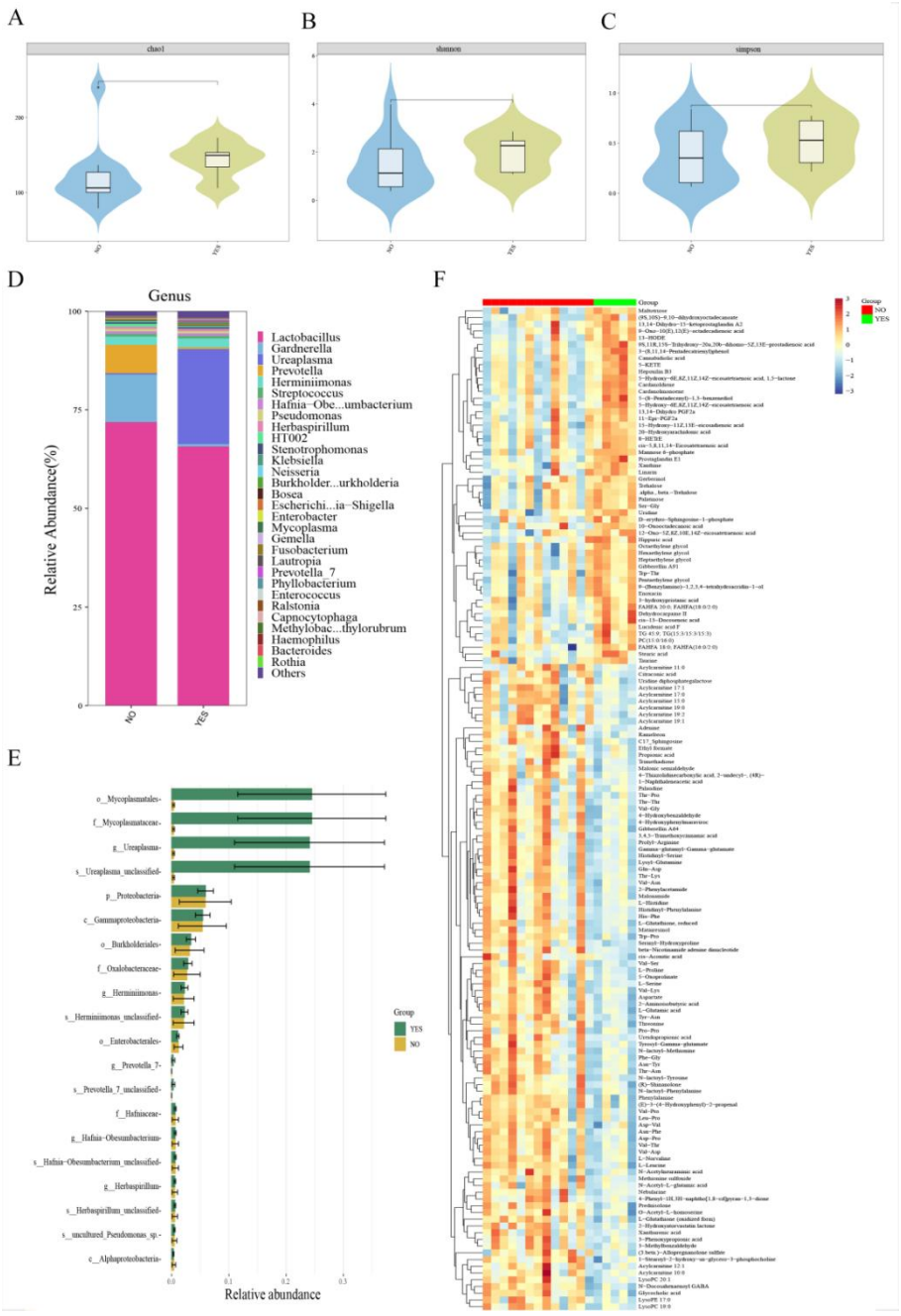

Fig S4. Comparison of cervicovaginal microbiota and metabolites between patients with a short cervix delivering at <34 weeks vs. ⩾34 weeks of gestation. (A, B, C) The Chao 1, Shannon, and Simpson indices were used to estimate the richness and diversity of the cervicovaginal microbiota in the two groups. (D) The relative abundance of bacterial genera clustered into two groups, and the microbiota compositions significantly differed. This analysis included only the top 30 communities. (E) The relative abundance of the cervicovaginal microbiota between the two groups was compared using box maps at the genus, family, order, class and species levels. (F) Metabolic profiles of patients with a short cervix who delivered at <34 wks and control patients. The heatmap shows the scaled relative intensity (Lg) of 149 differential metabolites (VIP ≥ 1.0, $p < 0.05$, and FC threshold of 1.5).

Fig S5

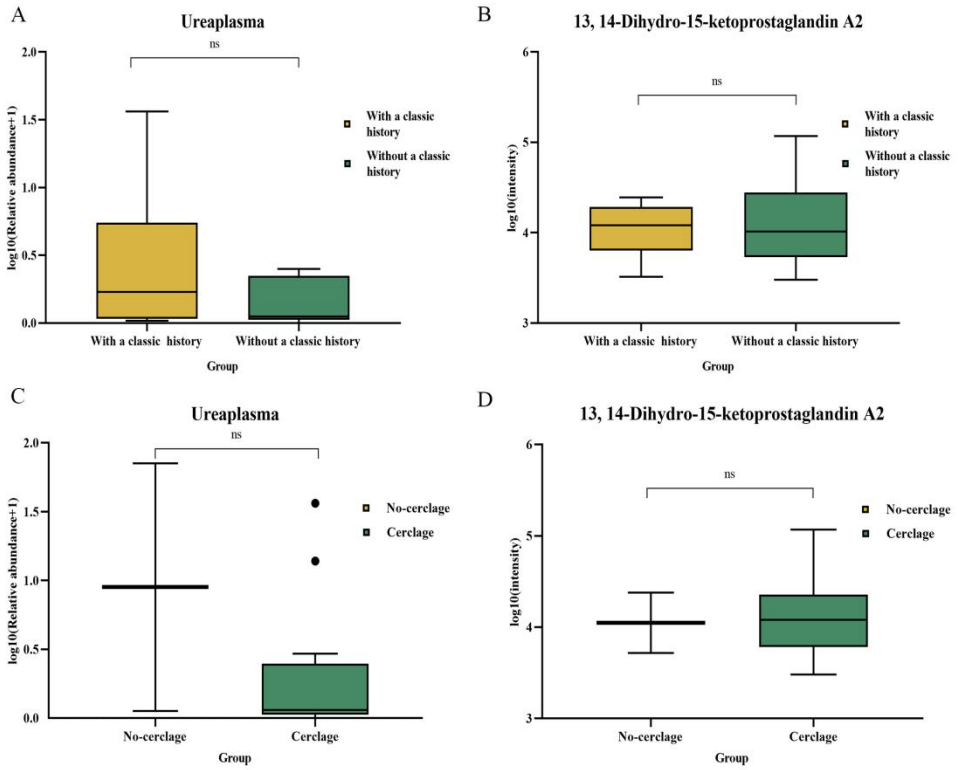

Fig S5. The relative abundance of *Ureaplasma* and 13,14-dihydro-15-ketoprostaglandin A2 differed in patients with a short cervix. (A, B) The patients were divided into two groups: those with a history of sPTB or mid-trimester pregnancy loss (n = 6) and those without such a history (n = 12). (C, D) The patients were divided into two groups: those who had undergone cerclage (n = 16) and those who had not (n = 2). ns, no statistically significant difference between the two groups. YES, patients with sPTB at <34 wks; NO, patients with gestations ≥34 wks.