Fig. S1

Alpha rarefaction curve

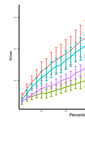

Fig. S1
The curve of all four groups were rising in the beginning of the curve and showed horizontal trend in the end of curve, which means the richness is still maximal and it will not rise if we enlarge the sample size further.
Vertical segments are 95% confidential interval of each point. This alpha rarefaction curve is aimed to estimate whether sample size is big enough to conduct alpha diversity analysis.

Fig. S2

Beta-diversity

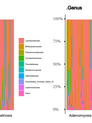

Fig. S2
The column was integrated by each samples after the normalization, both Family and Genus level were shown. Top 10 (abundance) OTUs in all groups had been selected. Lactobacillus is still dominant mostly.
It should be noticed that some OTUs such as Gardnerella, Atopobium and Prevotella were dominant microbiota in several single cases, these participants had no sign and symptom of vaginitis.

Fig. S3

Statisical difference of microbiota

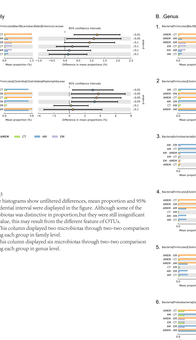

Fig. S3
These histograms show unfiltered differences, mean proportion and 95% confidential interval were displayed in the figure. Although some of the microbiota was distinctive in proportion,but they were still insignificant in statistical test because of too many different feature of OTUs.
(A) This column displayed two microbiotas through two-two comparison among each group in family level.
(B) This column displayed six microbiotas through two-two comparison among each group in genus level.

Fig. S4

3 dimensional PCA analysis

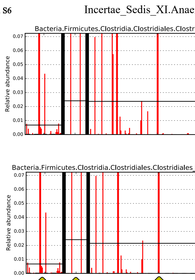

Fig. S4
Principle Component Analysis (PCA). Multi-dimensional reduction of matrix distance, three-dimensional figure is displayed, subjected to Benjamini-Hochberg false discovery rate (FDR) correction. Every dot represents a single sample and different colors represent different disease groups.

Fig. S5

LDA Effect Size(LEfSe) Analysis

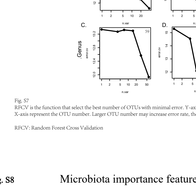

Fig. S5
LEfSe method is used for biomarker discovery. With OTU abundance >0.01% selection, the analysis emphasized on both statistical significance and biological relevance. Clostridiales_Incertae_Sedis_XI.Anaerococcus was the possible biomarker of adenomyosis-endometriosis.
(A)Effect size threshold 4 (log 10) and p=0.05 were used. This histogram
shows as this bacterial is identified as a biomarker both in family and genus level, both of them were LDA score >4.
(B) From the center to the edge of the circle, 7 layers of nodes represent 7 levels of evolution. Cladogram shows the significant family and genus with red mark. All the other nodes that had not been marked were insignificant species.

Fig. S6

Incertae_Sedis_XI.Anaerococcus

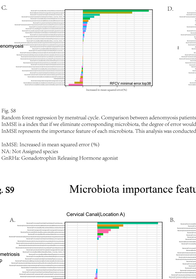

Fig. S6
The relative abundance of Clostridiales_Incertae_Sedis_XI.Anaerococcus in each sample, both in family and genus level.

Fig. S7

Random Forest Cross Validation

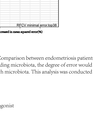

Fig. S7
RFCV is the function that select the best number of OTUs with minimal error. Y-axis represents error rate by cross validation, X-axis represent the OTU number. Larger OTU number may increase error rate, the best number is the lowest point of curve.

RFCV: Random Forest Cross Validation

Fig. S8

Microbiota importance feature in adenomyosis

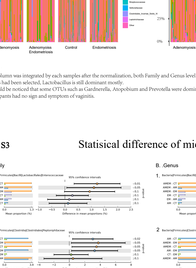

Fig. S8
Random forest regression by menstrual cycle. Comparison between adenomyosis patients and non-adenomyosis patients.
IncMSE is a index that if we eliminate corresponding microbiota, the degree of error would increase in final result.
IncMSE represents the importance feature of each microbiota. This analysis was conducted in family level. GnRHa treated patients were excluded.

IncMSE: Increased in mean squared error (%)
NA: Not Assigned species
GnRHa: Gonadotrophin Releasing Hormone agonist

Fig. S9

Microbiota importance feature in endometriosis

Fig. S9
Random forest regression by menstrual cycle. Comparison between endometriosis patients and non-endometriosis patients.
IncMSE is a index that if we eliminate corresponding microbiota, the degree of error would increase in final result.
IncMSE represents the importance feature of each microbiota. This analysis was conducted in family level. GnRHa treated patients were excluded.

IncMSE: Increased in mean squared error (%)
NA: Not Assigned species
GnRHa: Gonadotrophin Releasing Hormone agonist