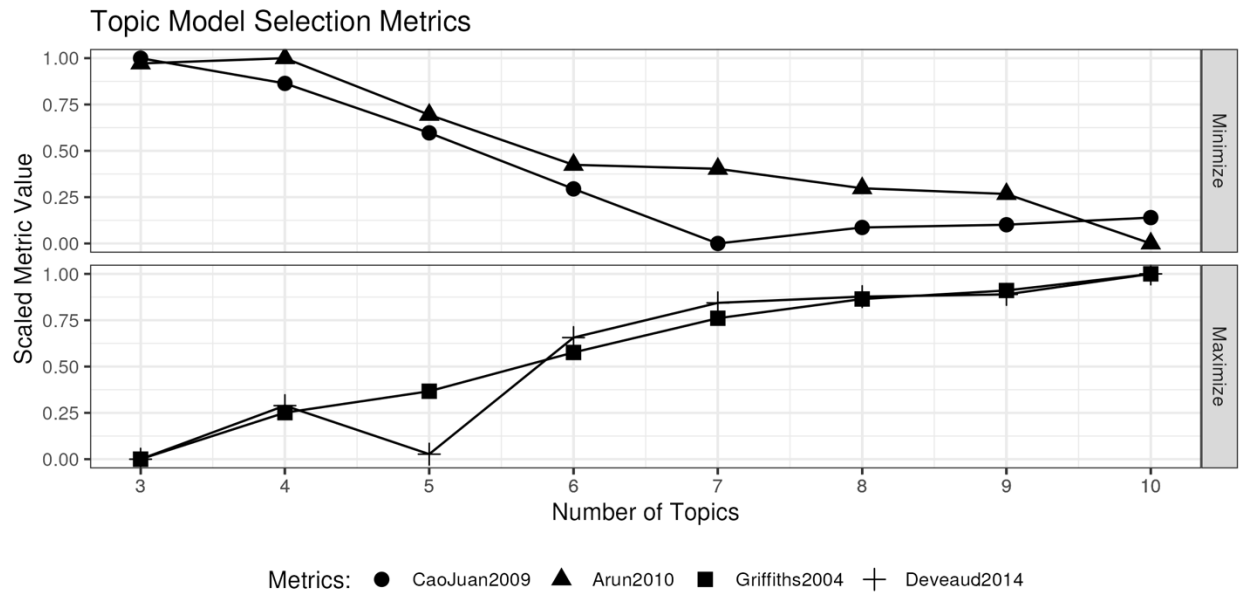


**Supplemental Figure 1. Vaginal microbiome alpha and beta diversity metrics by categorical variables.** Shannon diversity, an alpha diversity metric that accounts for both richness and evenness, is shown by **(A)** self-reported race ( $n=22$  for NH-Black, 110 for NH-White), **(C)** stage ( $n=41$  for early, 91 for late) and **(E)** histology group ( $n=52$  for Type I epithelial and others, and 80 for Type II epithelial). Weighted UniFrac distance, a beta diversity metric that accounts for phylogenetic distance and relative abundance, is shown by **(B)** self-reported race, **(D)** stage, and **(F)** histology group. Significance is based on ANOVA (A, C, E) or PERMANOVA (B, D, F). *ns*, not significant.



**Supplemental Figure 2. Optimal topic number determination for vaginal microbiome LDA model.** Evaluation metrics from Latent Dirichlet Allocation (LDA) topic modeling are plotted to identify the optimal number of latent topics representing the vaginal microbiome community structure. Four complementary metrics are shown: CaoJuan2009 and Arun2010 (minimized values indicate better model fit, top panel), and Griffiths2004 and Deveaud2014 (maximized values indicate better model fit, bottom panel) across a range of topic counts from 3 to 10 topics. Convergence of multiple optimization metrics is observed at 7 topics, indicating that this number provides the best balance between model complexity and biological interpretability. Based on this multi-metric analysis, 7 latent topics were selected for the final LDA model. Each topic corresponds to a distinct co-occurring taxa pattern in the vaginal microbiome.