

Presenting Author Name

Kay Durojaiye

Phonetic spelling of your (full) name

Kay Doo-row-jai_yey

Presenting Author Category

Undergraduate Student

Supervisor Name

Alicia Berard

Research Category

Basic Science

Abstract Title

Developing Reproducible Pipelines for Metaproteomic and 16S rRNA Analysis

Authors

Name	Role	Profession
Kehinde Durojaiye	Presenting Author	Undergraduate Student
Samantha Knodel	Co Author	Biostatistician
Adam Burgener	Co Author	Associate Professor
Versha Banerji	Co Author	Assistant Professor
Alicia Berard	Co Author	Assistant Professor

Background

The human microbiome is associated with health and disease across the lifespan, including maternal and child health, making it an important factor in many research studies. Molecular tools are used to understand the composition and function of the microbiome, including 16S sequencing and mass spectrometry (MS) metaproteomics, however expensive proprietary software to identify the microbiome with these tools has been a limitation.

Objective

To develop a standardized and reproducible microbiome analysis pipeline using open-source tools for metaproteomic protein identification and appropriate statistical methods for 16S rRNA and metaproteomic data, providing an analytical framework that supports microbiome research.

Methods

Four vaginal swabs collected from our local THRIVE cohort were processed for MS analysis and run on two different instruments. Multiple data-processing platforms were evaluated, comparing databases (SwissProt, VIRGO2) and protein identification pipelines (FragPipe, Scaffold/Mascot, MaxQuant, MSGF+).

16S rRNA data from nine vaginal samples were analyzed for alpha and beta diversity, PERMANOVA, differential abundance analysis, and non-parametric testing. Analytical workflows were implemented in R and integrated into interactive Apps to support standardized analysis, visualization, and integration with clinical metadata.

Results

Of the open-source protein identification software, FragPipe produced the same results as our gold-standard paid software (Scaffold/Mascot) using LFQ-MBR, tryptic digestion, and decoy database searching. VIRGO2 and human Swiss-Prot were used instead of full UniProt to reduce FDR and computational time. Different instruments did not significantly change the microbiome composition, indicating reproducibility of our data. We developed an analysis workflow integrating data cleaning, quality control, statistical analysis, and visualization into a standardized analysis pipeline, which will be used for both MS and 16S data.

Conclusion

Our new metaproteomics identification and microbiome analysis pipeline provides a standardized and reproducible framework, enabling our team to focus on biological interpretation for our ongoing maternal and child microbiome studies.