Managing and integrating multiple pathogen and microbiome datasets

Liz Knorr

Background

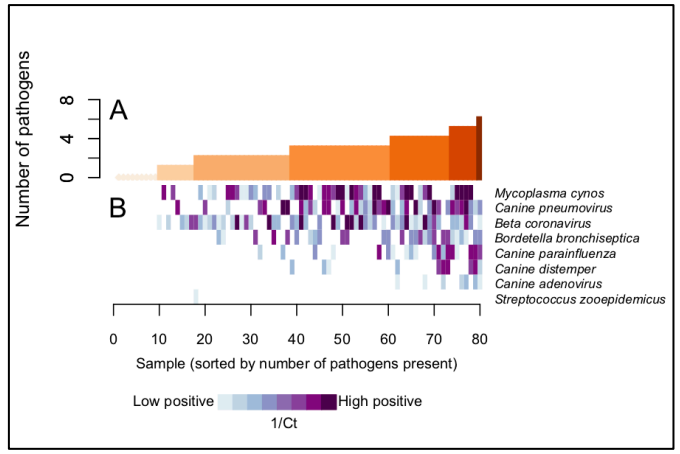
Host Data (location, health status, age)

Pathogen prevalence data

CGRB: 16s microbiome data

?

Analysis and Visualization



Problem

- Multiple datasets I need to combine and be able to analyze in different ways
  1. Sample collection
  2. Respiratory pathogen diagnostic panels for each sample
  3. 16s microbiome data for each sample
  4. Demographic data for each sample
- Experienced trouble identifying and matching samples; no checkpoints, hardcoding
- Multiple copies of raw data

More data collection coming soon!

Potential Approaches

- Best-practices for data management
  1. Data repository
  2. Combine data sets
  3. Clean data sets
  4. Reproducibility checkpoints

End Result

- Reproducible workflow that allows me to integrate, manipulate, and analyze datasets
  - Pathogen data analysis
  - Microbiome data analysis
  - Host/demographic data
- Workflow that I can use for future pathogen and microbial abundance data