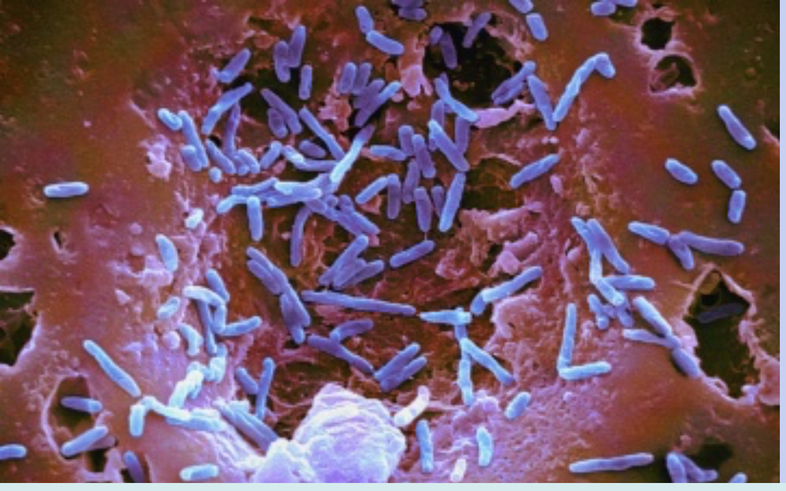


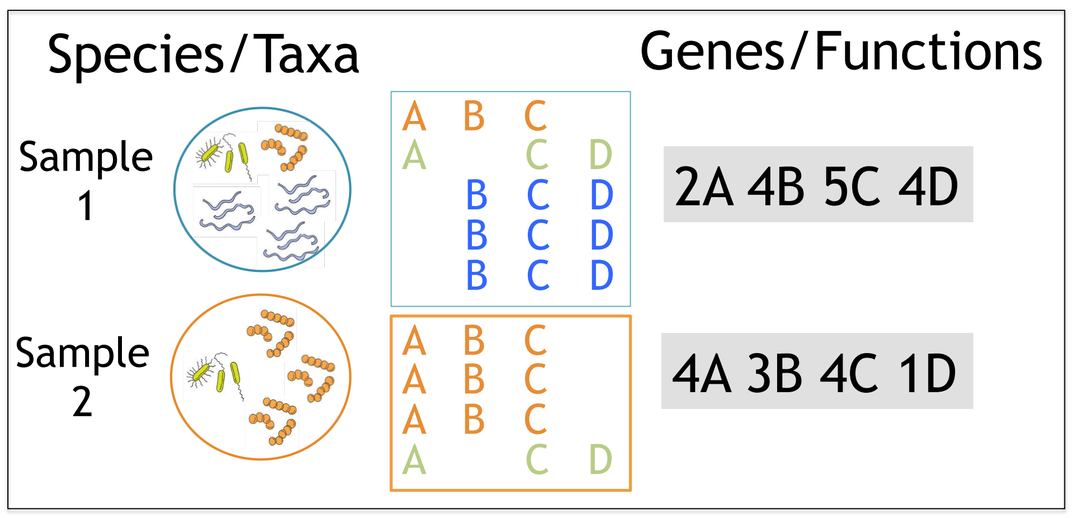
# Exploratory data analysis of taxonomic and functional metagenomic data

Alex Eng, Will Gagne-Maynard,
Colin McNally, Cecilia Noecker

## Problem and Motivation

New technologies enable detailed profiling of the taxonomic and functional composition of microbial communities and the variation in these across different environments. Microbiome researchers want to explore relationships between these two types of data, yet no method is currently available to facilitate this exploration.

### The data:




### Typical questions to ask about these datasets:

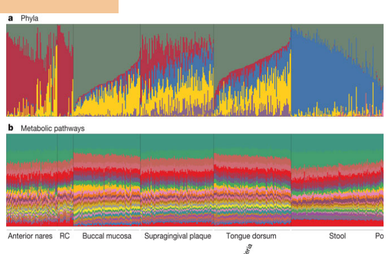
- What are the differences in function abundance across samples?
- What variation in taxonomic abundances accounts for those differences?
- Are taxa or functions more variable?
- Are particular taxa or functions associated with different study groups?

**Our tool will allow researchers to upload a dataset and interactively explore these questions and others.**

## Challenges and Previous Work

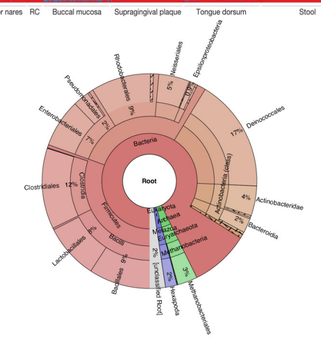
### Compositional data

Stacked bar plots are a standard visualization technique



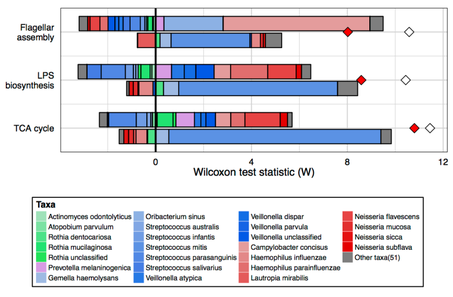

### Summarizing by hierarchies of taxa and functions

Previous example: Krona plots



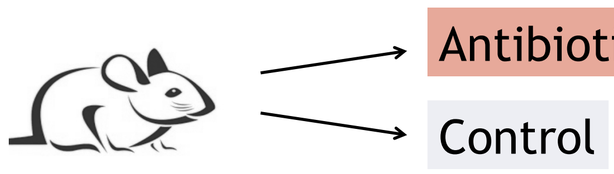

### Displaying complex relationships between taxa and functions

Previous example: FishTaco differential contributions




## Approach and Results

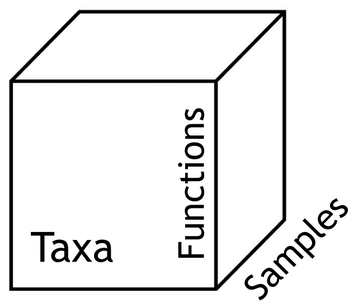
**Example dataset:**
Metagenomic sequence data from the gut microbiota of mice following antibiotic treatment




Samples from 2 days and 6 weeks following treatment.

### Data structure:

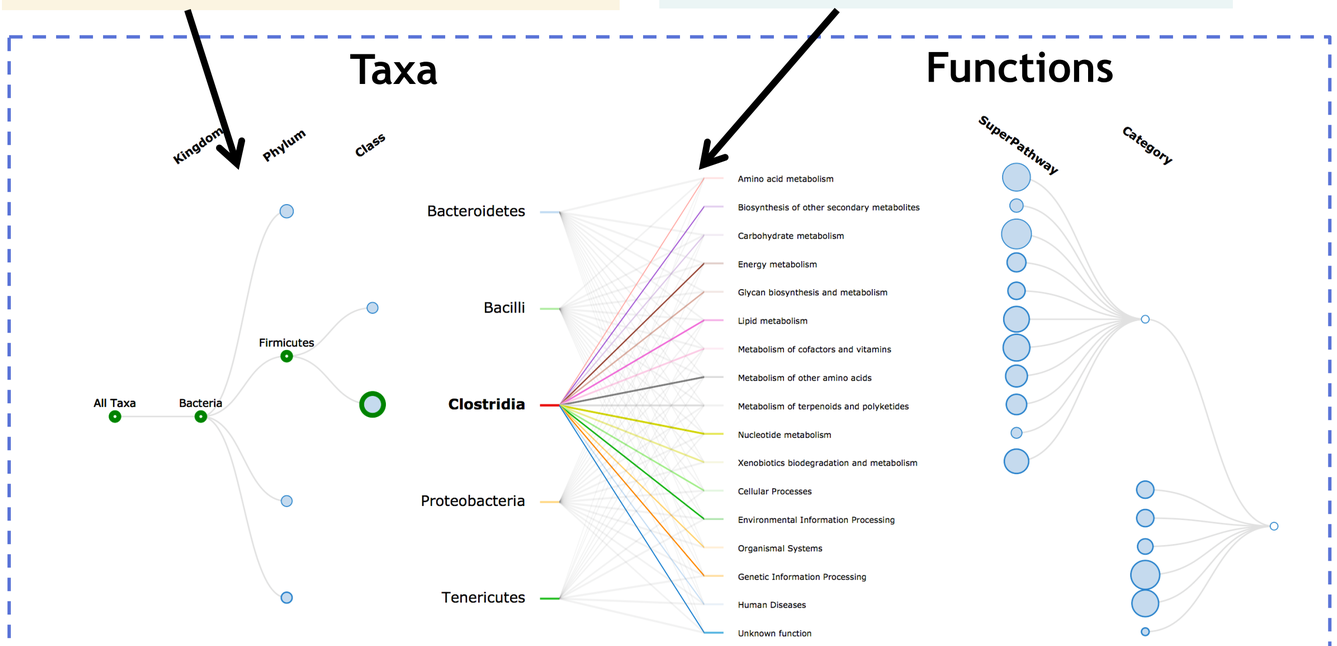
Dynamically updated 3D matrix of functions contributed by taxa to samples. Allows efficient operations to calculate visualized data after level of detail changes.




### Our Visualization:

Trees show and control hierarchical structure

Bipartite graph shows taxon-function links




Stacked bar charts enable comparisons across samples

Interactions reveal taxon-function contributions across samples

## Future Work

- Allow the user to view subsets of taxa and functions
- Allow the user to select which samples are displayed and facilitate better comparison between sample groups.
- Restructure the code to allow server-side computation and database query so that larger data sets can be supported.
- Support user-uploaded data

## References

Theriot et al (2014) *Nature Communications*

HMP Consortium (2012) *Nature*

Treangen et al (2013) *Genome Biology*

Manor et al (2015) in preparation