Approaches to Microbiome Analysis

Justin Gibbons, PhD

Postdoc, Jiang lab, USF Genomics Program

Consultant, USF Omics Hub

What is the microbiome?

- All of the microbes that are found on Earth
  - Bacteria
  - Archaea
  - Viruses
  - Fungi
- For practicality refer to the microbes of defined environments
  - The benthic zone of a lake
  - Gut
  - Vagina
  - Lungs
  - Many others…
- Impact ecosystem processes and health

How is the microbiome measured?

1. Marker gene studies: Taxa composition

2. Metagenomics: Gene composition

3. Metatranscriptomics: Gene expression

Types of analysis performed

- Diversity
  - Is there a relationship between health status and microbiome diversity?

- Sample ordination
  - Do changes in health status, diet or drug exposure result in characteristic microbiomes?

- Generalized linear model (GLM)
  - Is there an association between the microbiome and:
    - Gene expression
    - Metabolite concentration
    - Disease state

What is diversity?

- How many taxa are present and how evenly distributed are they?
- Measures of the unpredictability of the species identity of a randomly chosen individual
- There are many ways to measure diversity. None are perfect
- Do not filter data before calculating diversity!
  - Many diversity measures model the probability of low abundance species being shared between samples

Antibiotic use decreases gut microbiome diversity in infants

Gasparrini et al 2019
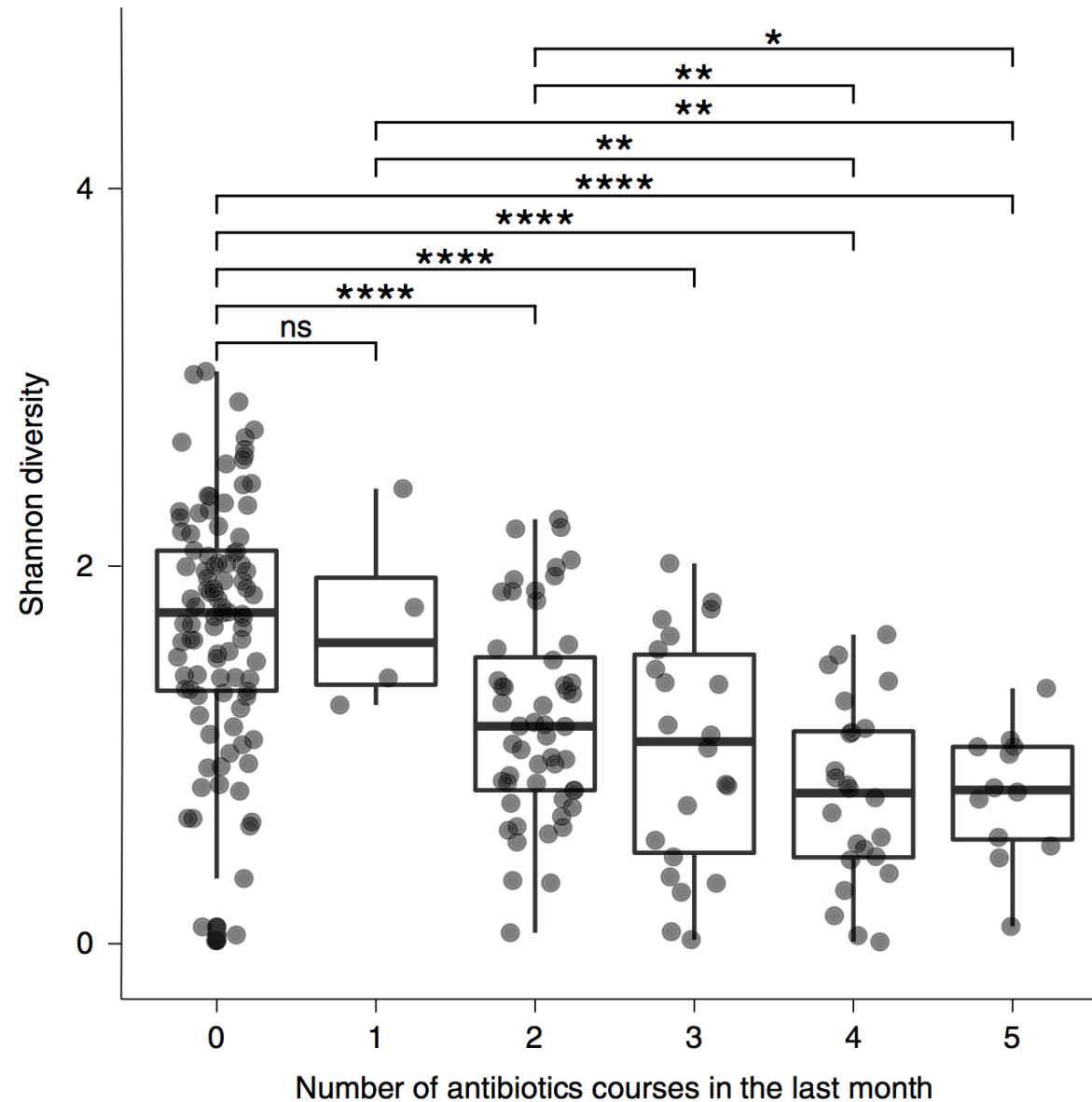
Sample Ordination

- Microbiome data has high dimensionality
  - There are many different taxa
- This makes it difficult to determine sample similarity
- Ordination techniques group similar samples together in a lower dimensional space (i.e. 2D or 3D)
  - PCA is a commonly used ordination technique, but does not work well with microbiome data
- Commonly used ordination methods in microbiome analysis
  - PCoA—Principal coordinates analysis
    - PCA modified to work with microbiome type data
    - Sometimes produces results that cannot be graphed
  - NMDS—Non-metric multidimensional scaling
    - Compares samples based on rank similarity
    - Can be used if PCoA does not work

Gut microbiome samples cluster by nationality

density

3
2
1

nationality

● AAM
● AFR
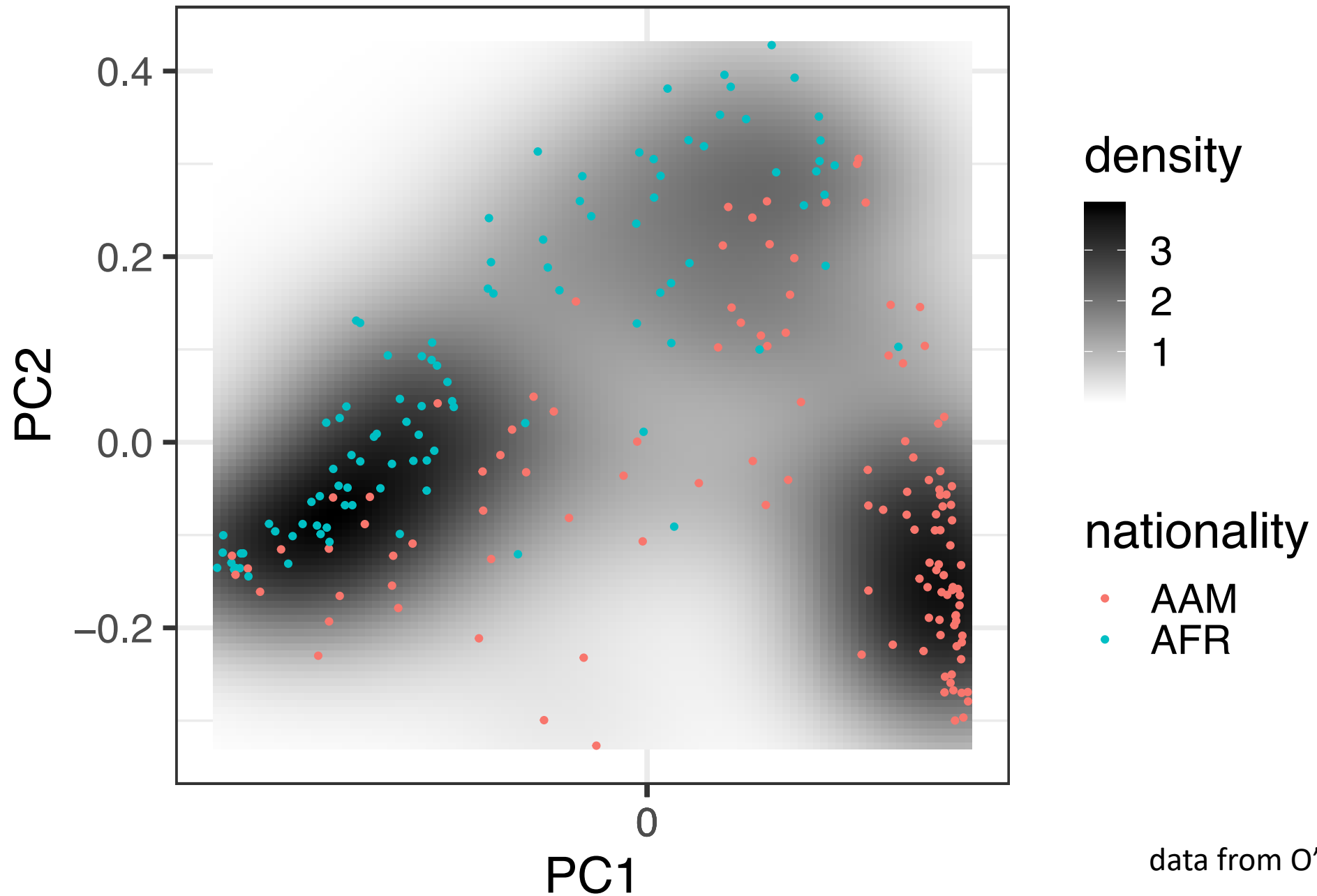
data from O'Keefe et al 2015

Generalized Linear Model

- Generalized linear models that can be used when:
  - The range of the measurement is restricted (e.g. binary or count)
  - The variance of the measurement depends on the mean

- DESeq2: Differential analysis of count data
  - Developed for RNA-seq
  - Statistical model works with any type of count data
  - Can be used to:
    - Identify differences in microbe abundance between conditions
    - Test for association between microbe level and:
      - Gene expression
      - Metabolite concentration
      - Disease state
      - Other…

Abundance of microbiome taxa is associated with specific host gene expression changes.
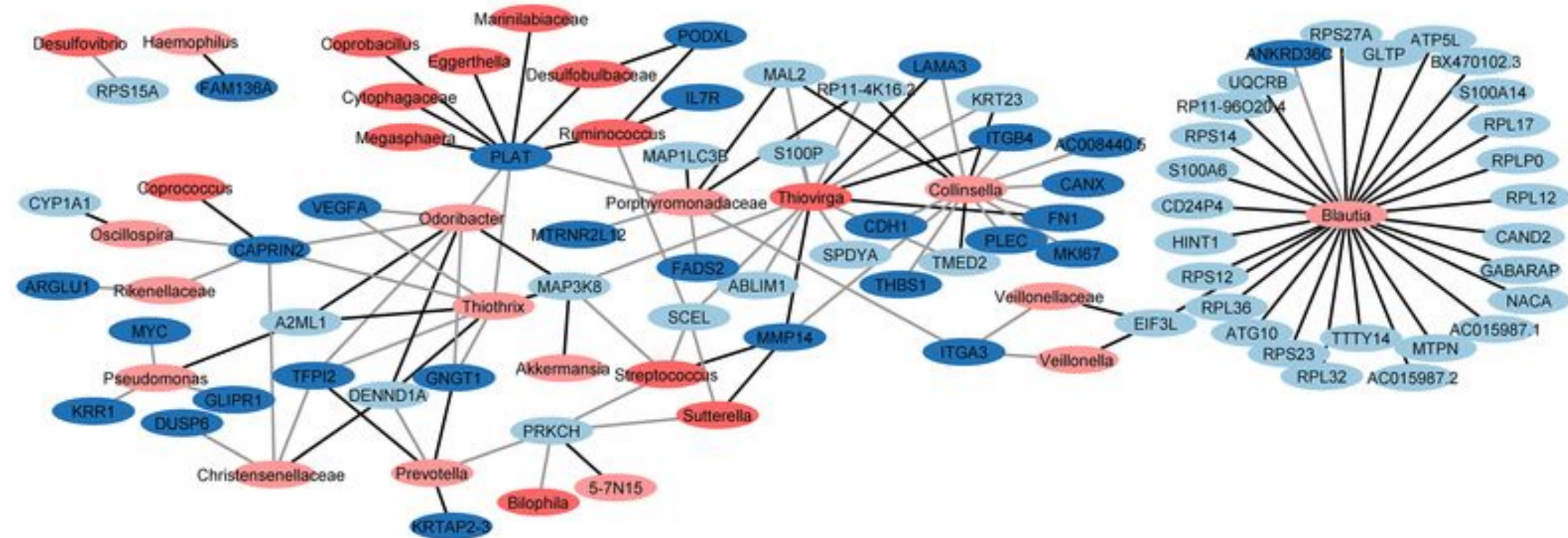
Summary

- The microbiome is an important component of ecosystems and health

- 3 of the main analysis techniques used are:
    1. Diversity analysis
    2. Sample ordination
    3. Generalized linear models

- The gene content of the microbiome is more important than who is there
    - PICRUSt: predict gene composition from taxa composition
    - Metagenomics: sequencing of microbiome genomes