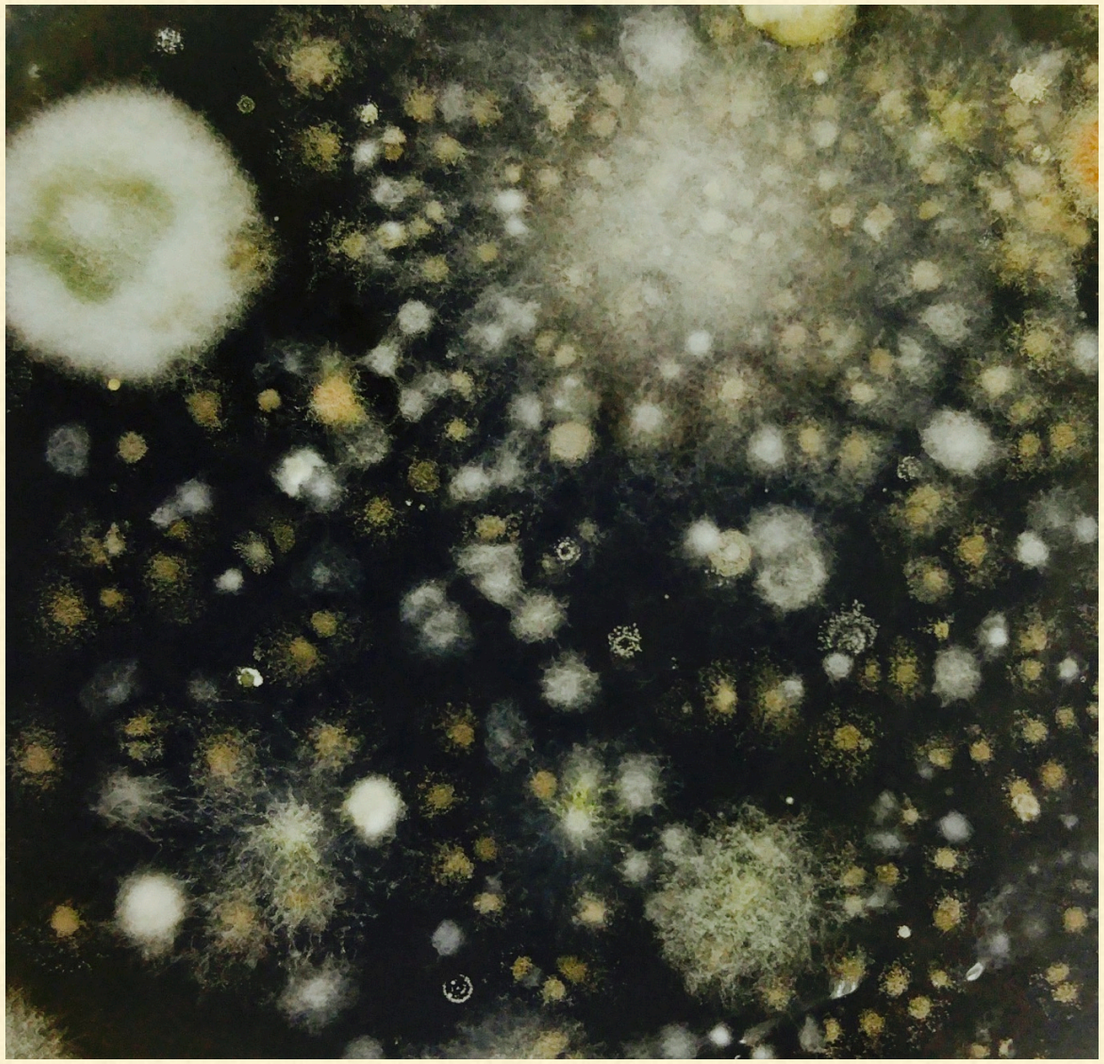


STATISTICAL CONSIDERATIONS FOR MICROBIOME RESEARCH

Amy D Willis, PhD

Assistant Professor, Department of Biostatistics, UW

@AmyDWillis adwillis@uw.edu

Slides: github.com/adw96/presentations

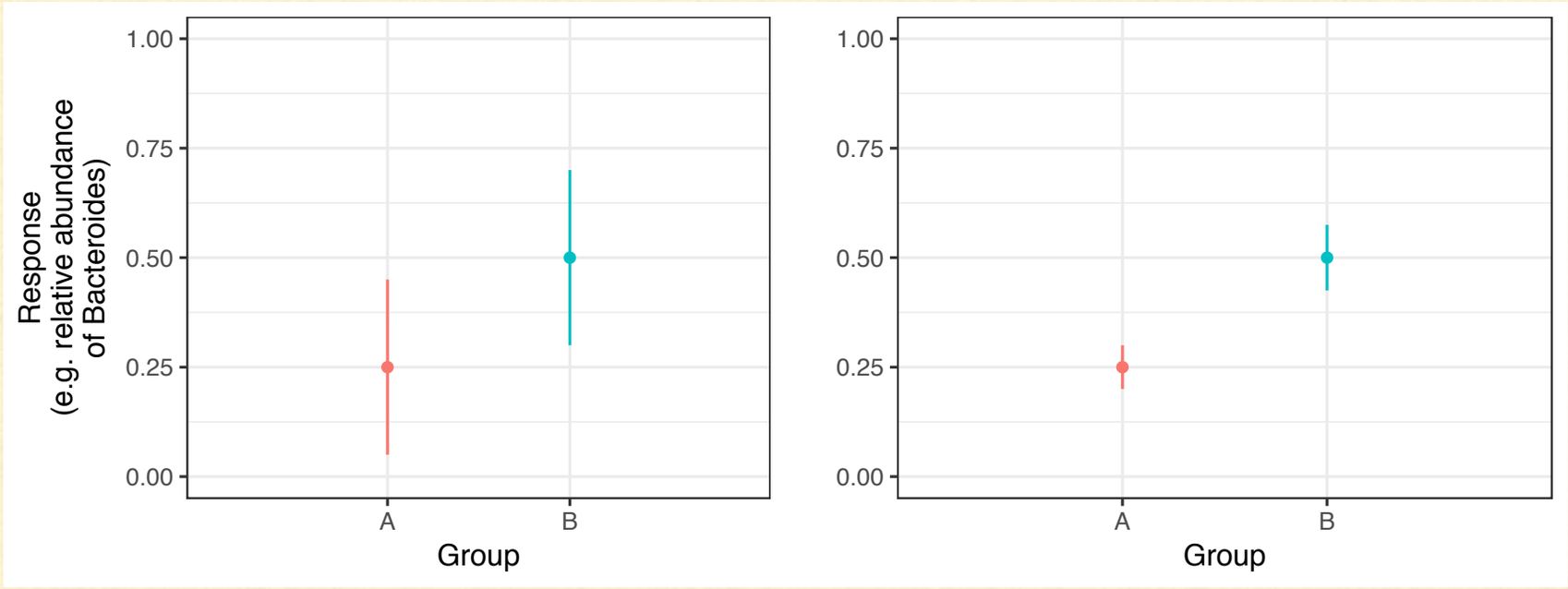
Photo credit: T.D. Berry, Whitman lab, UW Madison

HYPOTHESIS TESTING

- Almost all statistical tests of interest to microbiome researchers look like

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{Var(\bar{X}_1) + Var(\bar{X}_2)}} \sim t \text{ or } \mathcal{N}$$

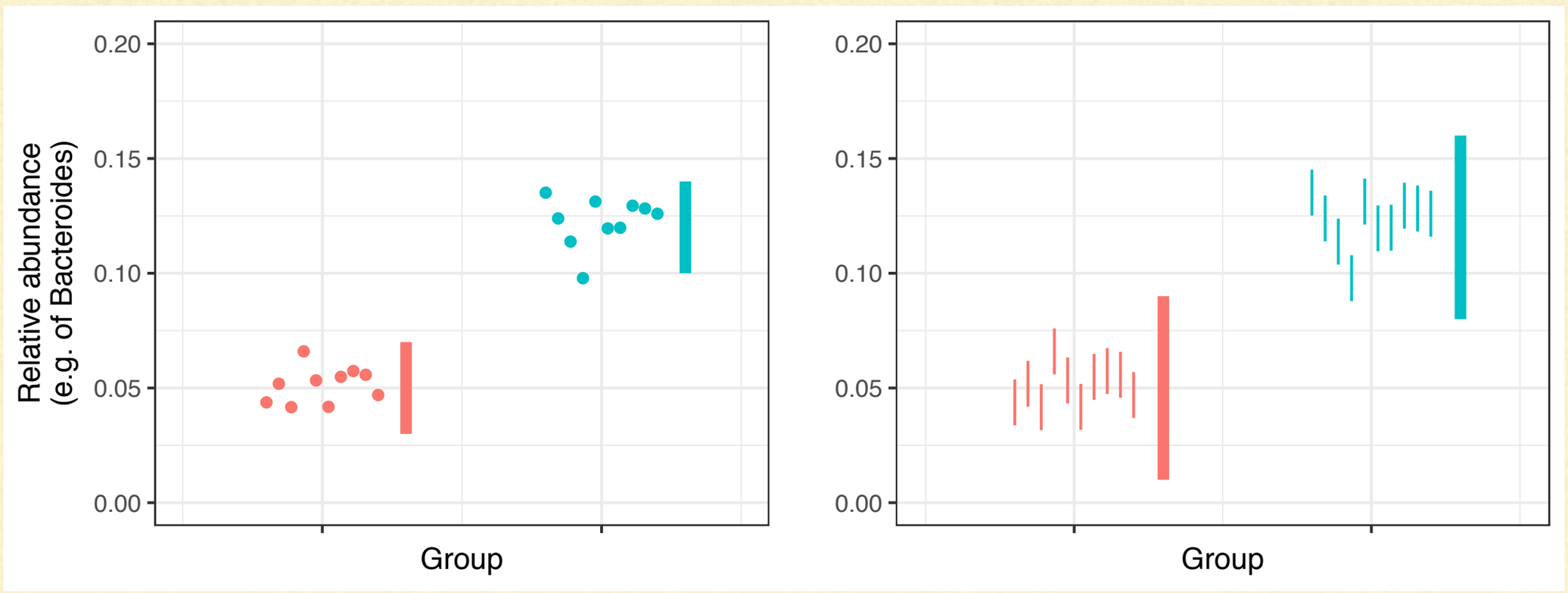
- The challenge in microbiome research is finding $Var(\bar{X}_i)$



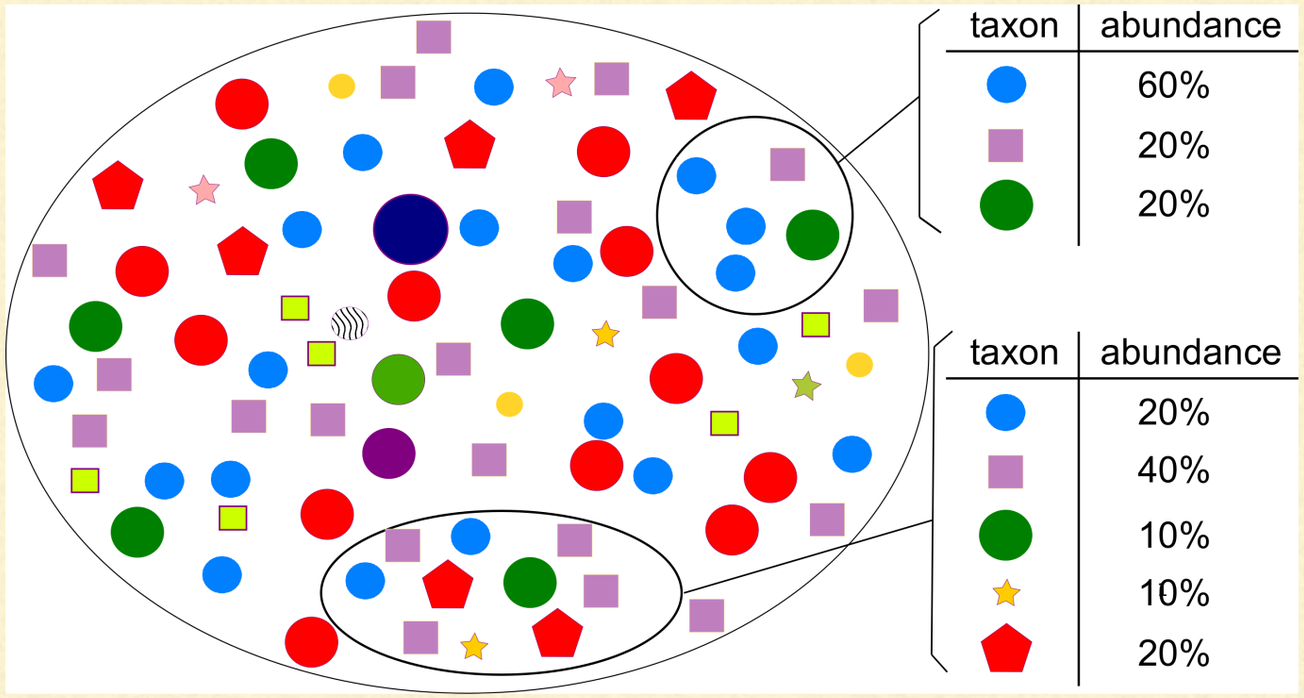
Slides: github.com/adw96/presentations

CHALLENGE: INEXHAUSTIVE SEQUENCING

- You didn't see every microbe in the community, so you have *measurement error* in every test that you do



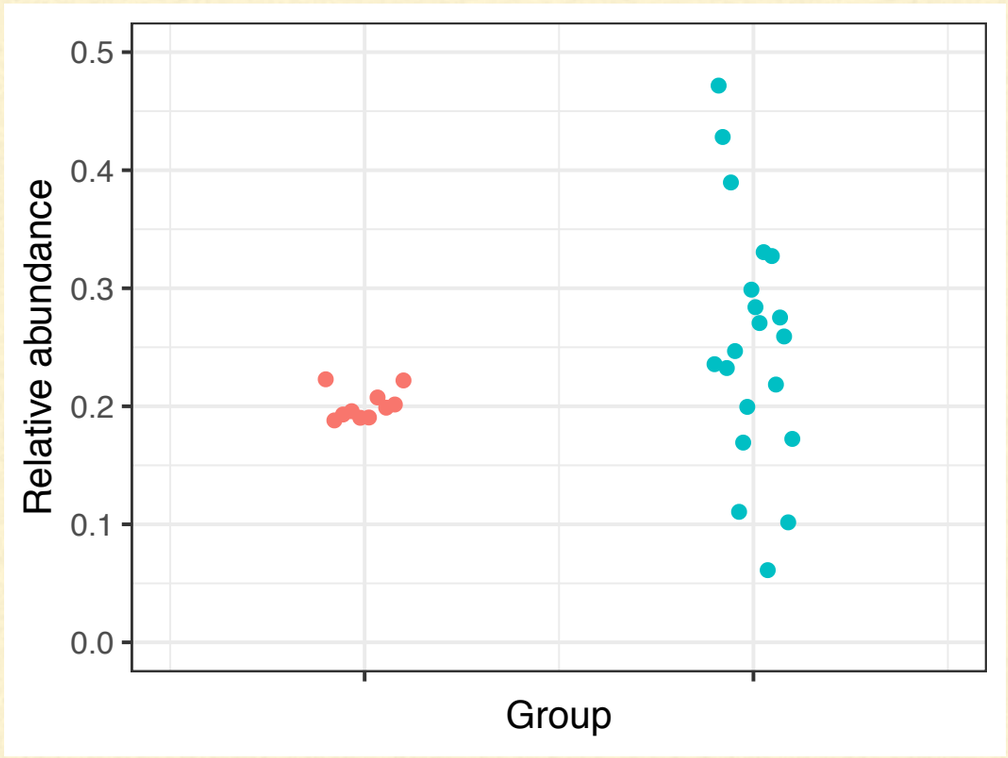
SOLUTION: LATENT VARIABLE MODELS



- Latent variable models for the microbiome explicitly accounts for inexhaustive sequencing
- Old idea: Aitchison (1986)
- Rebranded as "ANCOM": Mandal et al. (2015)

LATENT VARIABLE MODELS

- ~~ANCOM~~ Aitchison models (and its variance) are fine for testing means, but interest may be in variance differences
- No existing statistical methods permit hypothesis testing for variance (dysbiosis)



TESTING VARIANCES

- Null hypothesis: the instability in the relative abundance of Microbe X in group 1 is the same as the instability in group 2

- Martin, Witten & Willis (new work)
  - Beta-binomial model for relative abundance and variance of relative abundance
  - Adjusts for different library sizes
  - Independence not required!
  - Suitable for longitudinal/time-series/cross-sectional studies!
  - Manuscript coming soon, software available now on request

Bryan Martin, UW Statistics

Daniela Witten, UW Statistics

ALPHA DIVERSITY

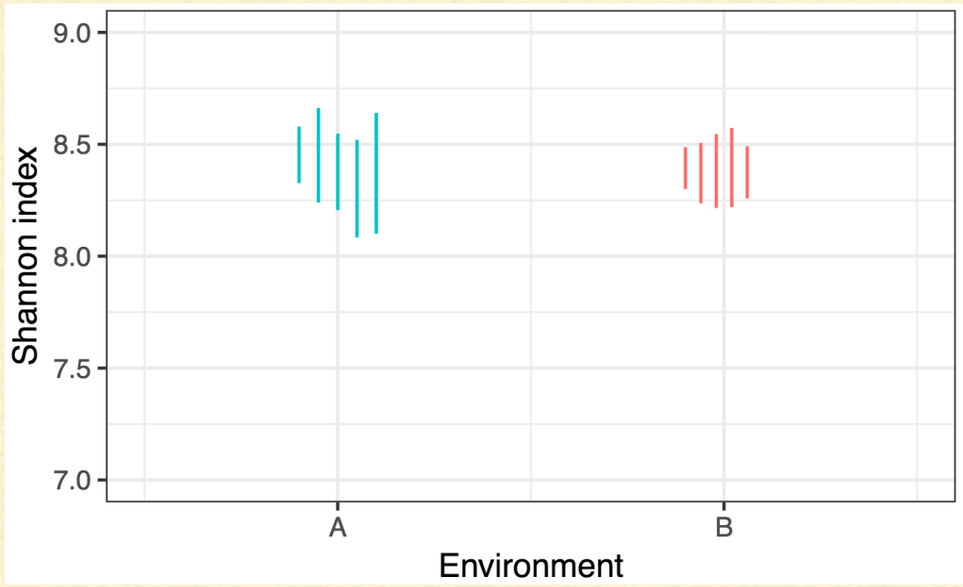
- Species richness
  - Hard to avoid assumption that microbes behave independently
  - Only method that does: Willis & Bunge (2015)
  - breakaway()
- Shannon and Simpson diversity
  - estimate_shannon() and estimate_simpson()
- All functions in R package breakaway: github.com/adw96/breakaway

ALPHA DIVERSITY

- Never, ever rarefy your data
  - Much better ways to do the same thing (Willis 2017)
- Never, ever extrapolate a rarefaction curve
  - Use species richness estimates instead (breakaway, CatchAll...)

TESTING ALPHA DIVERSITY HYPOTHESES

- Once you have good alpha diversity estimates, what hypothesis test accounts for the uncertainty?



- betta() is the correct t-test generalization for alpha diversity
  - Longitudinal, cross-sectional, time-series, multiple covariates, random effects...

FURTHER DETAILS

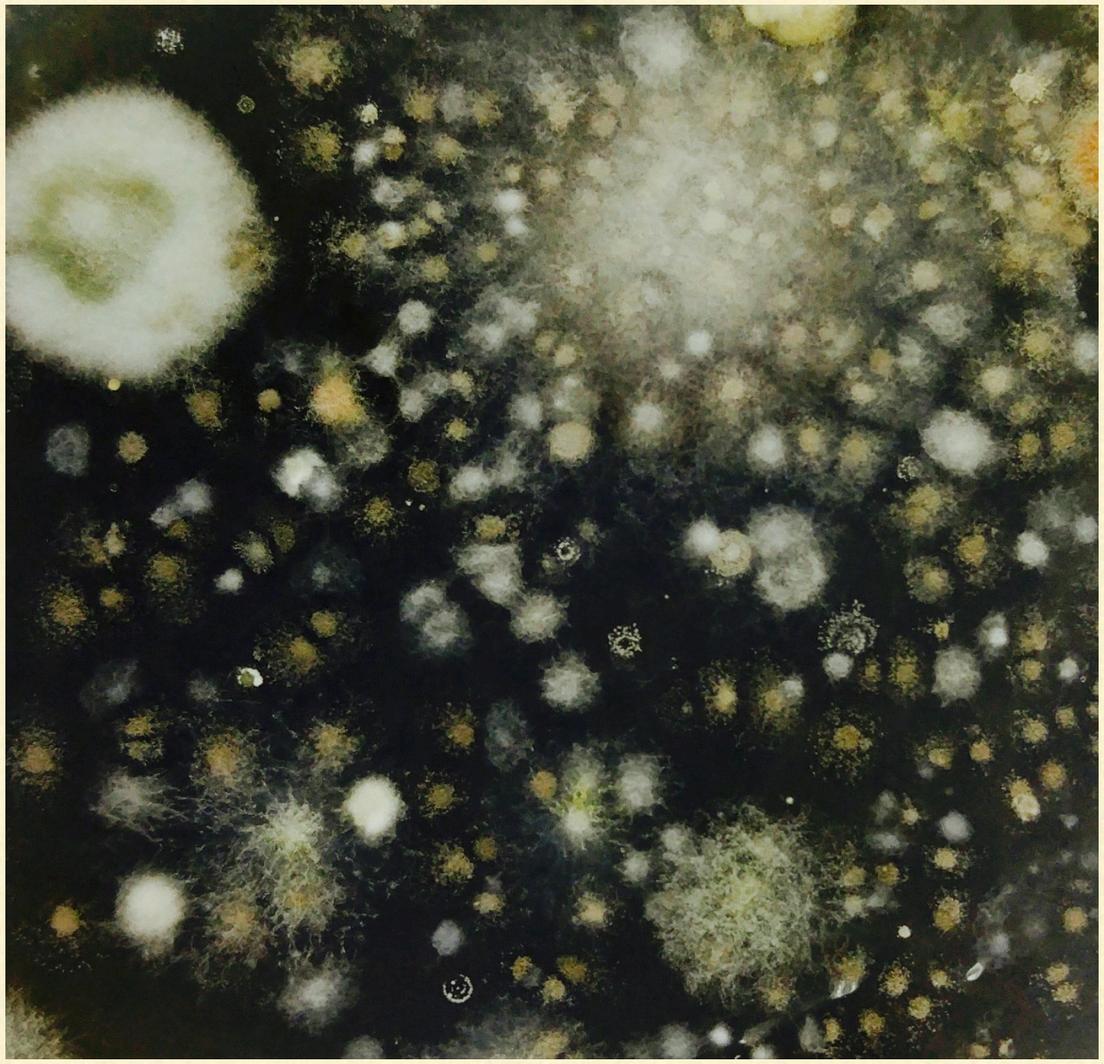
- These slides: github.com/adw96/presentations
- Testing dysbiosis: Martin, Witten & Willis (In Prep)
- breakaway(): Willis & Bunge (Biometrics, 2015)
- betta(): Willis, Bunge & Whitman (JRSS-C, 2017)
- Don't rarefy: Willis (bioarxiv, 2017)

REFERENCES: SOFTWARE

- breakaway: github.com/adw96/breakaway
- CatchAll: northeastern.edu/catchall/
- ANCOM: Mandal et al. (*Microb Ecol Health Dis*, 2015)
- Beta-binomial: github.com/martinbryan

RESOURCES

- UW Department of Biostatistics, Statistics
- Statistical Diversity Lab @ UW
  - faculty.washington.edu/adwillis/
    - new site coming soon...
- STAMPS course @ MBL, applications due April 6
  - mbl.edu/education/courses/stamps/



STATISTICAL CONSIDERATIONS FOR MICROBIOME RESEARCH

Amy D Willis, PhD

Assistant Professor, Department of Biostatistics, UW

@AmyDWillis adwillis@uw.edu

Slides: github.com/adw96/presentations

Photo credit: T.D. Berry, Whitman lab, UW Madison