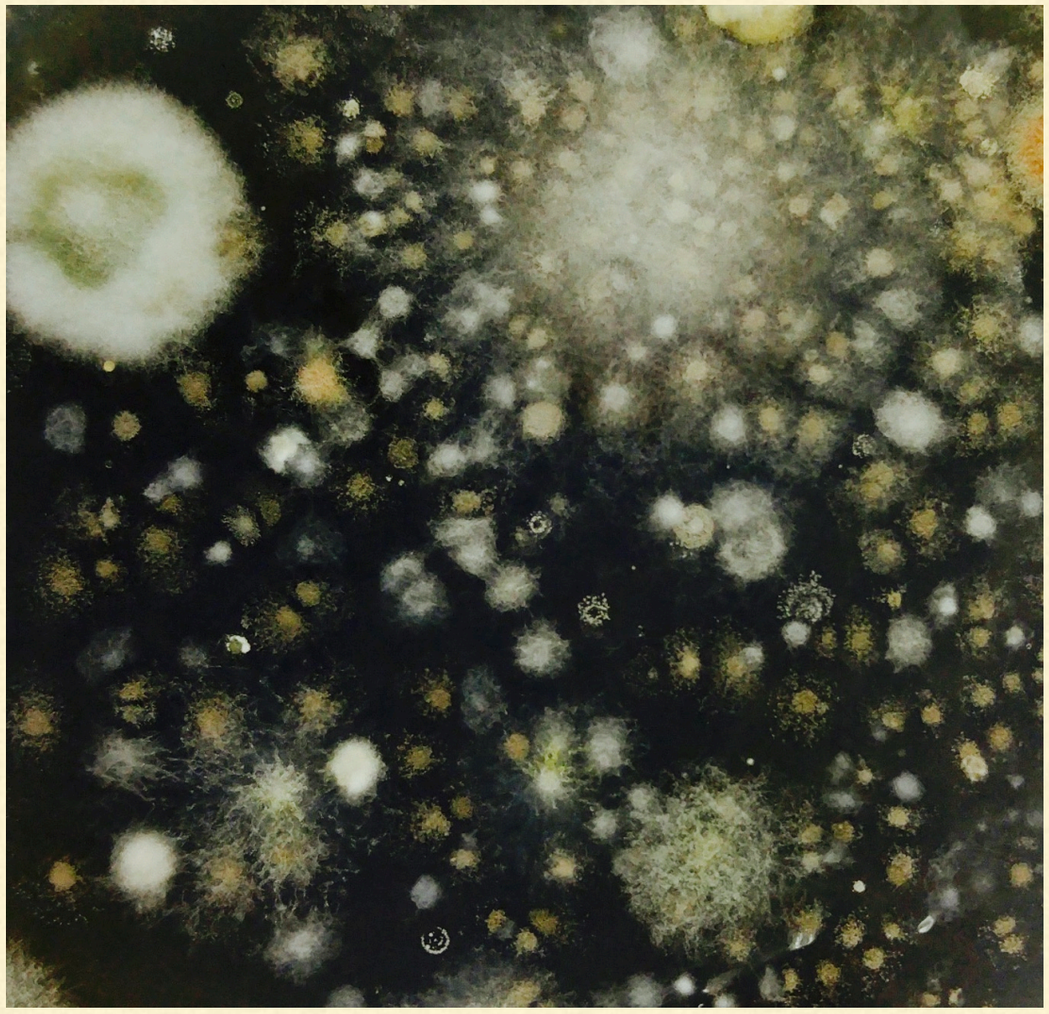


ADJUSTING FOR MULTIPLE TESTING IN MICROBIOME DATA ANALYSIS

Research Group: Statistical Diversity Lab (new!)

PI: Amy D Willis PhD, Assistant Professor, Department of Biostatistics, UW

@AmyDWillis adwillis@uw.edu

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

GOAL OF SCIENCE

- Find "truth"
  - Find an interesting result that stands up to replication
- Get papers published?

APPROACHES

Exploratory

- hypothesis generating
- "I wonder if [diet] affects [the microbiome], and if so, how..."

Confirmatory

- hypothesis confirming
- "Does the abundance of [firmicutes] change with diet?"

HYPOTHESIS TESTING

- Tests that use data to assess the "statistical significance" of a result

- Definition: the *p-value* of a test is the probability of observing a more extreme result than we did if the null hypothesis were true.

- Idea: If our results were extreme under the assumption of the null hypothesis, then maybe the hypothesis isn't supported by the data.

TESTING MULTIPLE HYPOTHESES

- Extreme things happen occasionally: the longer you look, the more likely they are

- If you have no signal in your experiment, what's the probability of a p-value < 0.01?

TESTING MULTIPLE HYPOTHESES

- If you have no signal in your experiment, the probability of a p-value < 0.01 is 0.01

- If you do 2 hypothesis tests, the probability that one or more has p-value < 0.01 is 1.99%

- 3 tests = 3%; 10 tests = 10%, 20 tests =18%, 100 tests = 63%

TESTING MULTIPLE HYPOTHESES

- Does 100 tests seem ridiculous?

Diet 1

6 weeks

Diet 1

8 weeks

Diet 2

8 weeks

Diet 3

8 weeks

- 128 families + Simpson + Shannon:130 x 4 = 520 tests = 99.5% chance of finding something at 1% level *when there is no difference*

- 10 phyla + 20 families + Simp. + Shann.: 32 x 4 = 128 tests = 73% chance

TESTING MULTIPLE HYPOTHESES

- Two strategies for correcting for multiple comparisons
  - Family wise error rate control
  - False discovery rate control
- Both either decrease your significance threshold OR increase your p-value
- Both start with a protocol: a list of every hypothesis you're interested in

PROTOCOL

- ~~"Characterize the taxonomic composition of the gut microbiome before, during and after the dietary intervention..."~~

- Write down a list of every single hypothesis you're interested in
  - Confirmatory study: maybe 10, definitely < 20
  - Exploratory study: could be hundreds, thousands

FAMILY WISE ERROR RATE CONTROL

- Idea: Want to guard against any false positives
  - Very, very strict standard
- Many, many methods exist!
- Simplest method is Bonferroni: new p-values are

$$p_{new} = \min(\#\text{ tests} \times p_{old}, 1)$$

FALSE DISCOVERY RATE CONTROL

- Idea: False positives inevitable, try to limit the number
  - Less strict than FWERC
- Many, many methods exist!
- Simplest method is Benjamini-Hochberg

FALSE DISCOVERY RATE CONTROL

- Benjamini-Hochberg: Control percentage of false discoveries at 10%

- Order p-values $p_1 \leq p_2 \leq \ldots \leq p_m$

- $p_1,\ldots, p_j$ are significant for largest j such that $p_j \leq 0.1 \times j/m$

RANK	1	2	3	4	5	6	7	8	9	10
p-value	0.0008	0.009	0.165	0.205	0.396	0.450	0.641	0.781	0.9	0.993
0.1j/m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1

EXAMPLE METHODS SECTION

- Main: The large number of inferential tests performed in this article necessitated a multiple comparisons adjustment. Details of the adjustment procedure are available in Supplementary Statistical Methods. The procedure implies that a significance level of $\alpha = 0.0244$ should be used to assess significant hypotheses. Throughout the article, only hypotheses that meet this threshold are described as significant.

- Supp: ...to control false discovery rate (FDR) at 5%, we employ the Benjamini-Hochberg procedure on all 38 hypotheses investigated. This count includes hypotheses that were investigated prior to preparing the manuscript...

Chan, Willis, *et al.* (2016) Clinical Cancer Research

DEMONSTRATION

- DePaolo Lab study on non-alcoholic fatty liver disease (NAFLD)
- Mice fed a non-alcoholic fatty liver disease (NAFLD) inducing diet for 6 weeks
- Then assigned for 8 weeks to either a NAFLD diet, a HFF diet or a LFF diet
- 2 cohorts
- Ileum & fecal matter
- *Amy's analysis: "We found a significant alteration of Firmicutes in both fecal and ileum samples ($p = 0.01$ and $p = 0.04$), and a significant alteration of Verrucomicrobia in both fecal and ileum samples ($p = 0.02$, $p = 0.03$). We also found changes in Actinobacteria abundance in fecal matter ($p = 0.01$)."*

DEPAOLO LAB STUDY

NAFLD
6 weeks

2 + 2

NAFLD
8 weeks
6 + 2

LFF
8 weeks
4 + 2

HFF
8 weeks
3 + 2

AMY'S TWEAK

NAFLD

6 weeks

2 + 2

NAFLD

8 weeks

LFF

8 weeks

HFF

8 weeks

6 + 2

move 2

move 1

4 + 2

move 1

move 1

3 + 2

move 3

move 1

ANALYZING SCRAMBLED DATA

- First fit logistic regression to all phyla abundances with diet and cohort as fixed effects… nothing
- Repeat at family level… nothing
- Then fit logistic regression with only diet… nothing
- Then fit linear model with only diet…

"We found a significant alteration of Firmicutes in both fecal and ileum samples ($p = 0.01$ and $p = 0.04$), and a significant alteration of Verrucomicrobia in both fecal and ileum samples ($p = 0.02$, $p = 0.03$). We also found changes in Actinobacteria abundance in fecal matter ($p = 0.01$)."

KEEP IN MIND

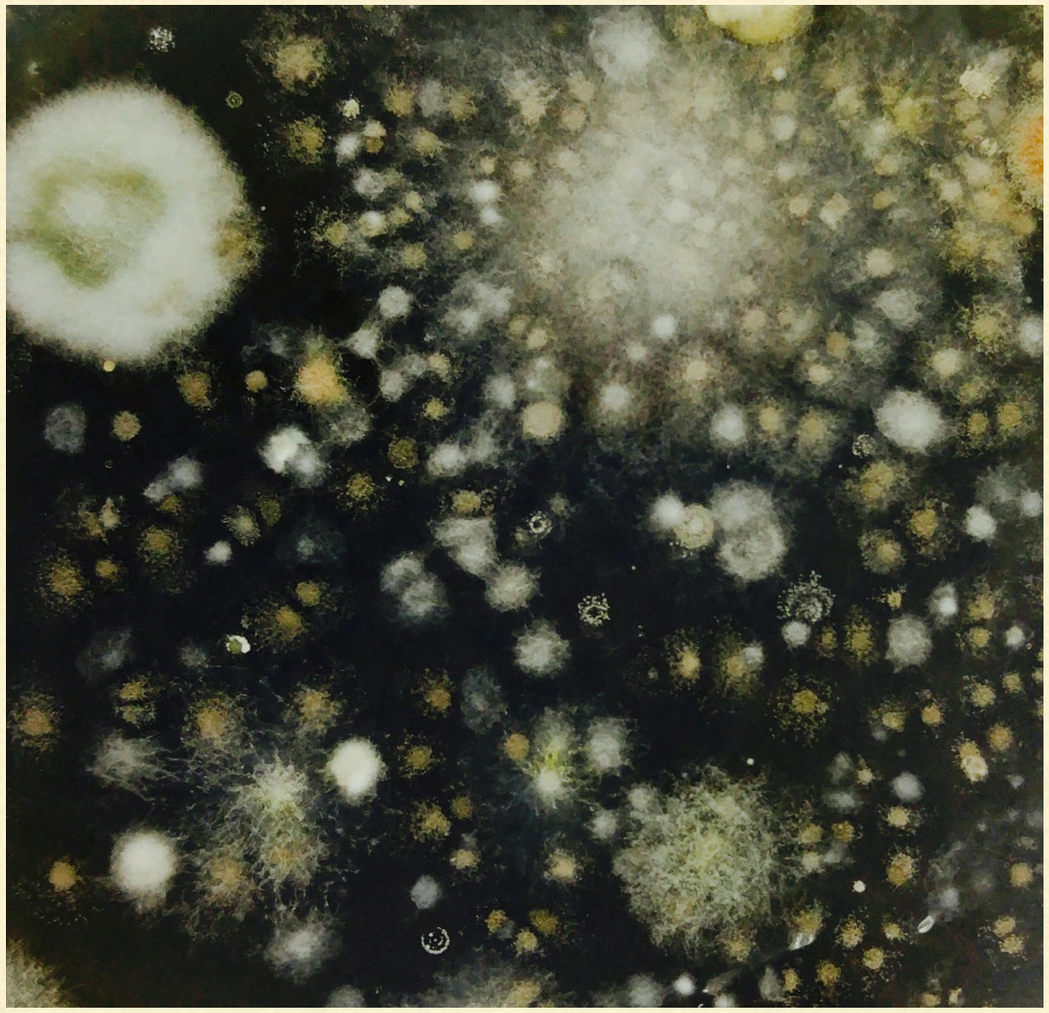
- If you don't know every statistical analysis that was involved in preparing a paper, you cannot make an informed decision about the "significance" of its results

- Fishing: Many tests conducted but only "interesting" results reported

  - Withholds information necessary to adjust for multiple comparisons

RESOURCES

- Departments of Biostatistics and Statistics @ UW provide consulting services
  - stat.washington.edu/consulting/
- The new Statistical Diversity Lab @ UW
  - http://faculty.washington.edu/adwillis/
    - new site coming soon...
- Any STAT/BIOST class in any university...



ADJUSTING FOR MULTIPLE TESTING IN MICROBIOME DATA ANALYSIS

Research Group: Statistical Diversity Lab (new!)

PI: Amy D Willis PhD, Assistant Professor, Department of Biostatistics, UW

@AmyDWillis adwillis@uw.edu

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