

MODELING ABUNDANCES

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



@AmyDWillis



adwillis@uw.edu

Bryan D Martin, PhD Candidate, Department of Statistics, UW

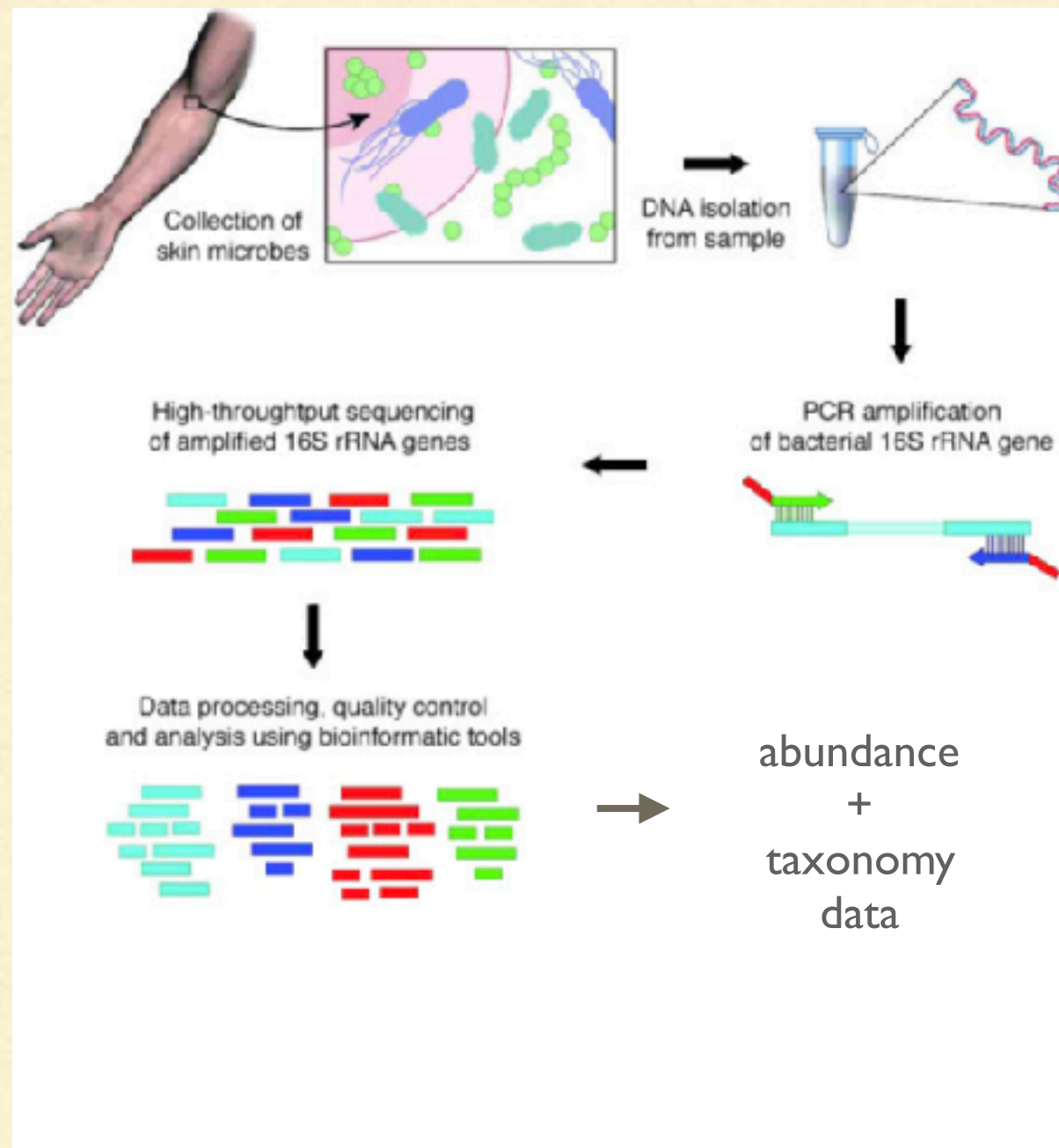


@BryanDMartin_



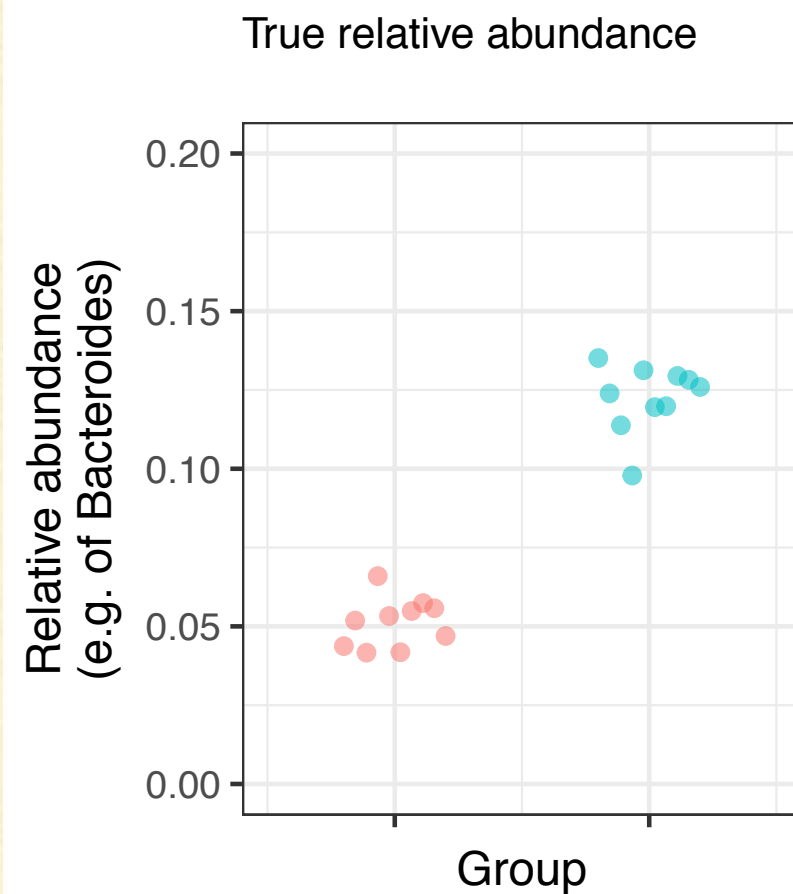
bmartin6@uw.edu

EXPERIMENTAL WORKFLOW

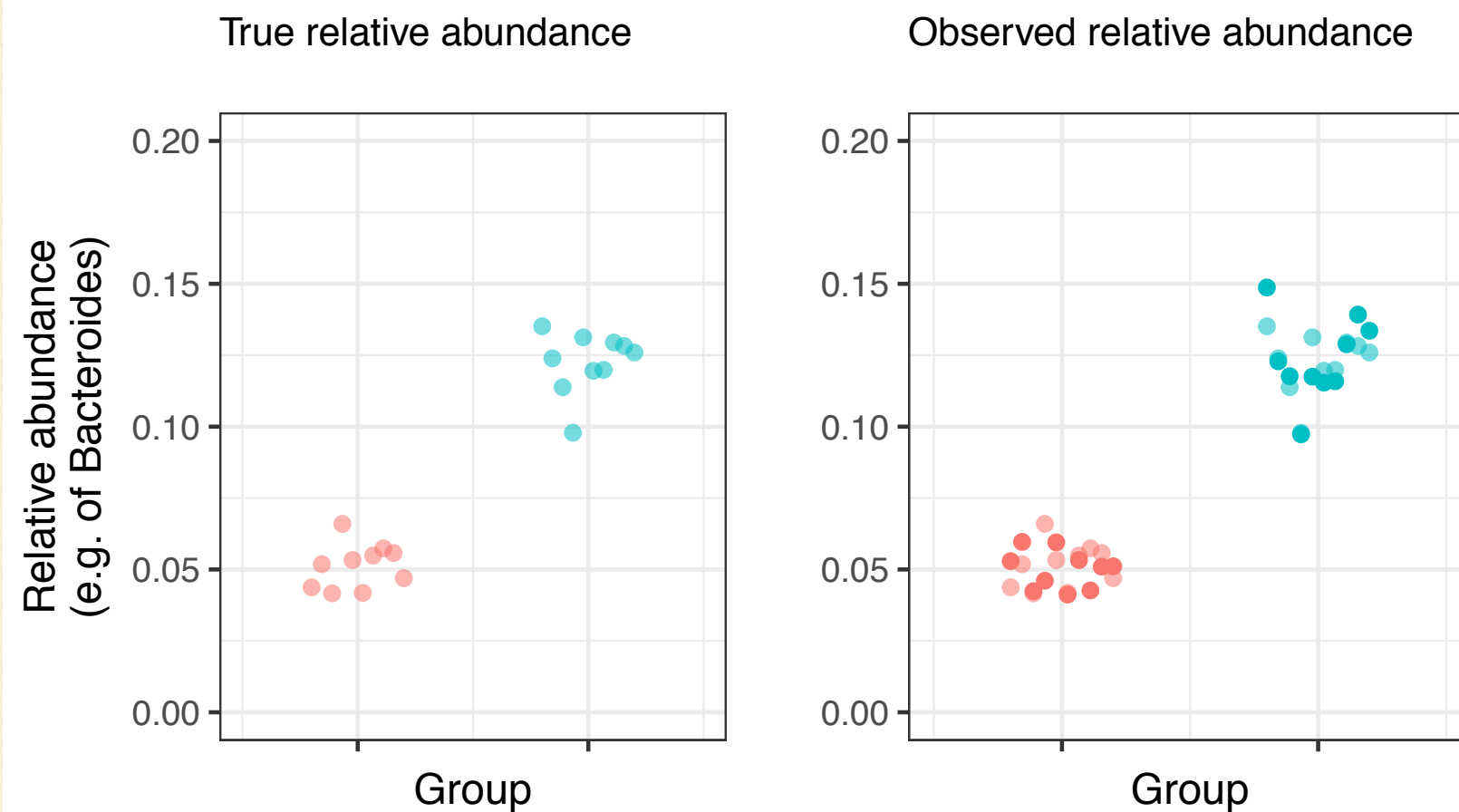


What ecosystem
are we interested
in?

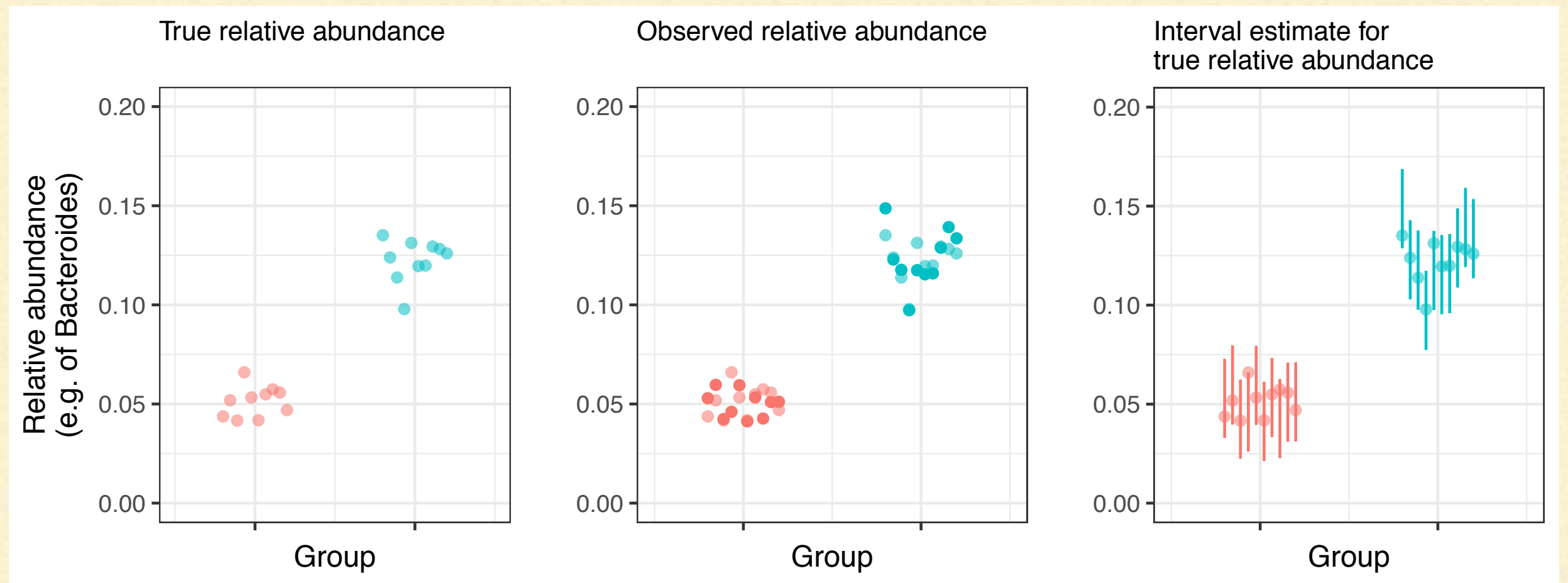
SAMPLE VS POPULATION



SAMPLE VS POPULATION

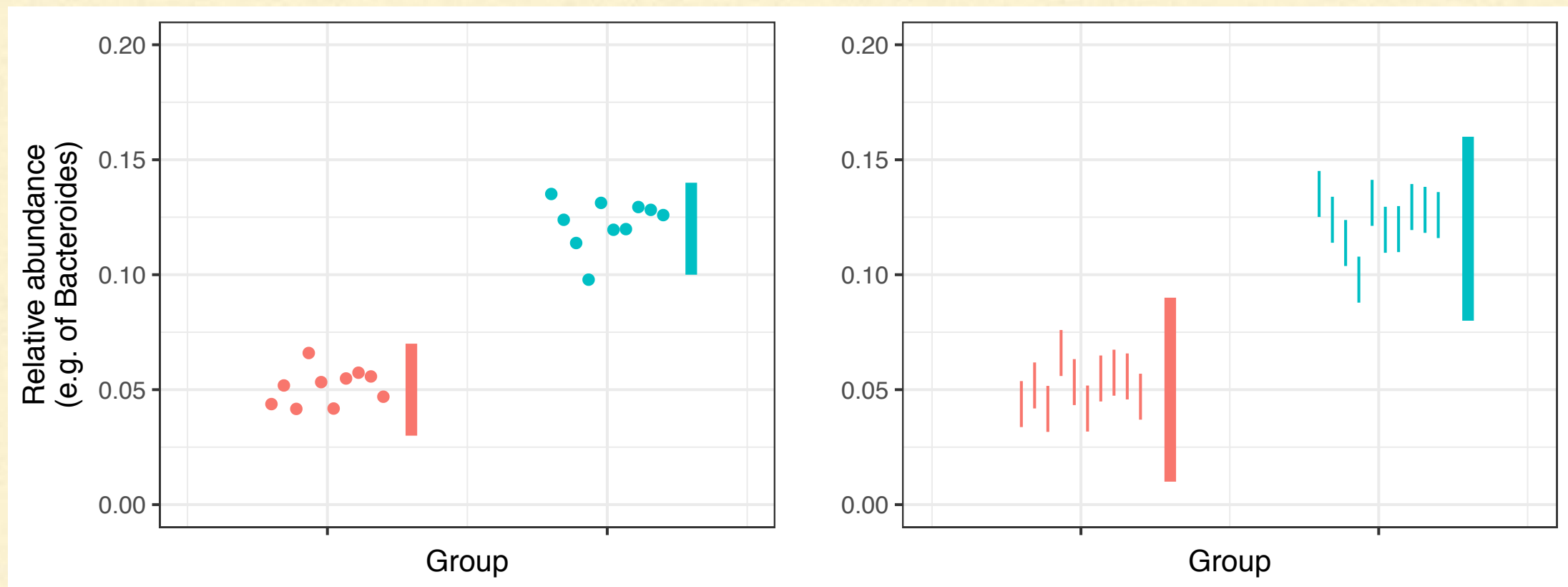


SAMPLE VS POPULATION



SAMPLE \neq POPULATION

- Observed relative abundance \neq true relative abundance
- Any statistical test for the microbiome needs to account for this *measurement error*



CORNCOB

COmpositional RegressioN for Correlated Observations with the Beta-binomial



- Latent variable model for comparing **relative abundances**
- Addresses measurement error issue
- Adjusts for different library sizes
- Suitable for longitudinal/time-series/cross-sectional studies; handles multiple covariates
- Mean and variance ("dysbiosis") testing

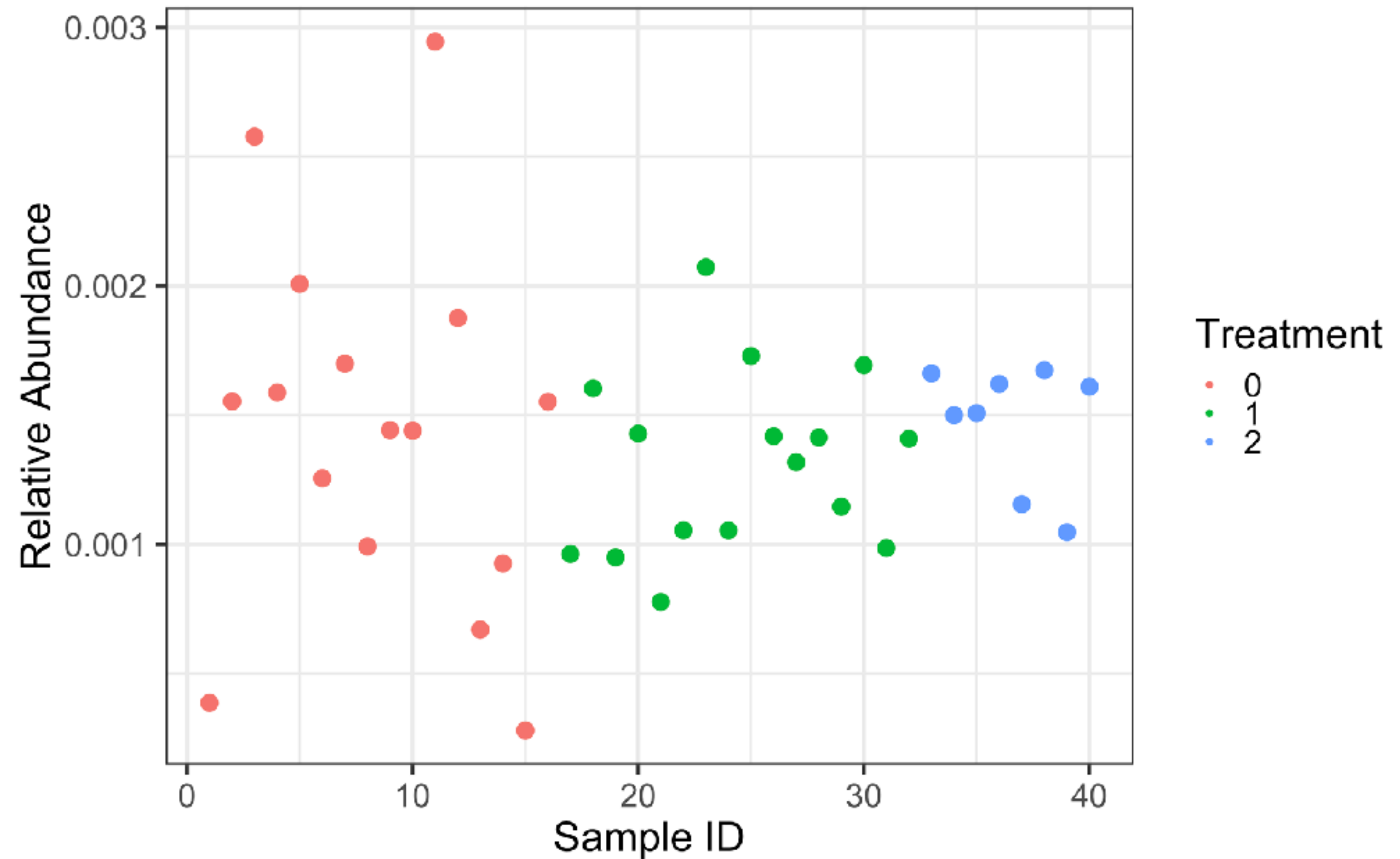


Bryan Martin, UW Statistics

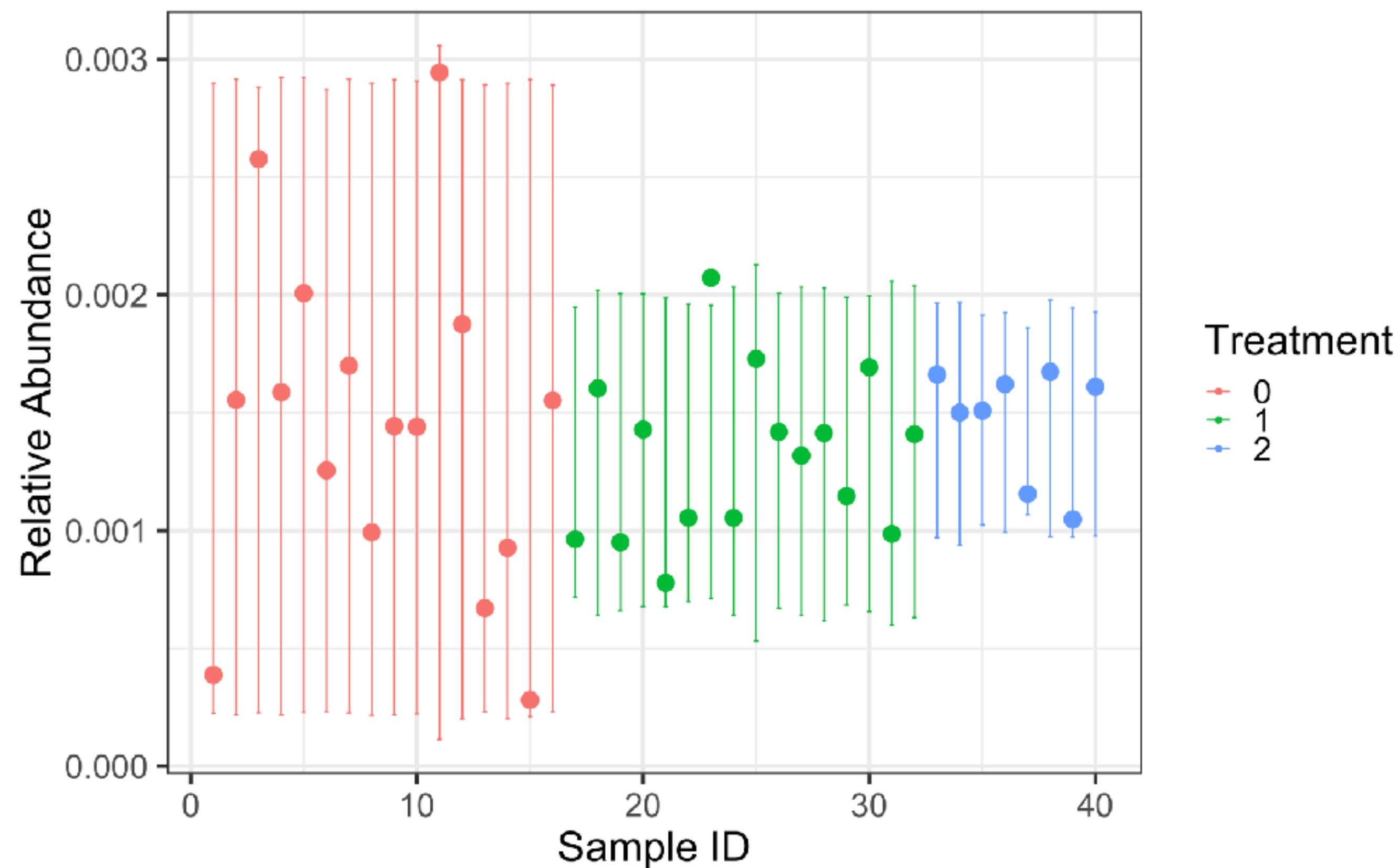


Daniela Witten, UW Statistics

Soil by Fertilizer Treatment



Soil by Fertilizer Treatment



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-6.54691	0.13182	-49.667	<2e-16	***
DayAmdmt1	-0.09160	0.14827	-0.618	0.5409	
DayAmdmt2	0.01372	0.14484	0.095	0.9251	
phi.(Intercept)	-7.92054	0.40688	-19.467	<2e-16	***
phi.DayAmdmt1	-1.67400	0.61916	-2.704	0.0106	*
phi.DayAmdmt2	-3.00787	1.13831	-2.642	0.0124	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -167.78

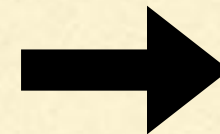
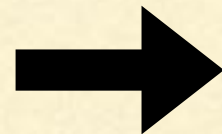
We found no significant difference in the predicted relative abundance of this taxon with different soil treatments.

CORNCOB



Abundance table
+
Sample data

(e.g. disease status, diet,
treatment, BMI, age, ...)



1. List of differentially abundant taxa
(with p-values)
2. List of differentially variable taxa
(with p-values)

BETA-BINOMIAL DISTRIBUTION

$$W_i | Z_i, M_i \sim \text{Binomial}(M_i, Z_i),$$

$$Z_i \sim \text{Beta}(a_{1,i}, a_{2,i})$$

- **n** = samples, indexed by **i** = 1, ..., n
- **W_i** = # of individuals observed in the taxon of interest
- **M_i** = total # of individuals observed
- **Z_i** = the (latent) relative abundance

LINKING ABUNDANCE TO COVARIATES

1. Parameters

$$\mu_i = \frac{a_{1,i}}{a_{1,i} + a_{2,i}}, \quad \text{“(latent) relative abundance”}$$

$$\phi_i = \frac{1}{a_{1,i} + a_{2,i} + 1} \quad \begin{array}{l} \text{“within sample correlation”} \\ \text{“absolute abundance overdispersion”} \end{array}$$

2. Link to covariates

μ_i is a function of \mathbf{X}_i, β

ϕ_i is a function of \mathbf{X}_i^*, β^*

Hypothesis Testing

$$H_0 : \beta = 0$$

- Does the healthy group have a different mean relative abundance?
- Is a high-fat diet associated with changes in the relative abundance of *Firmicutes*?

$$H_0 : \beta^* = 0$$

- Are changes in diet associated with a change in within-taxon variability?
- Is a high-fat diet associated with changes in the stability of *Firmicutes*?

CORNCOB

COmpositional RegressioN for Correlated Observations with the Beta-binomial



- CORNCOB does not assume that microbes behave independently
- Parameter ϕ controls cooccurrence of taxa of the same group
- Same idea as DivNet
 - CORNCOB reflects structure in microbial communities
 - (Bonus: Very cool urn model interpretation of microbial reproduction)

CORNCOB AND DESEQ2

- **Similarities**

- Easy to use with phyloseq
 - Use un-normalized counts to assess precision of estimates
 - Benjamini-Hochberg adjustment for multiple comparisons
 - Dispersion parameter for overdispersion
 - Tests for differential abundance
-

CORNCOB AND DESEQ2

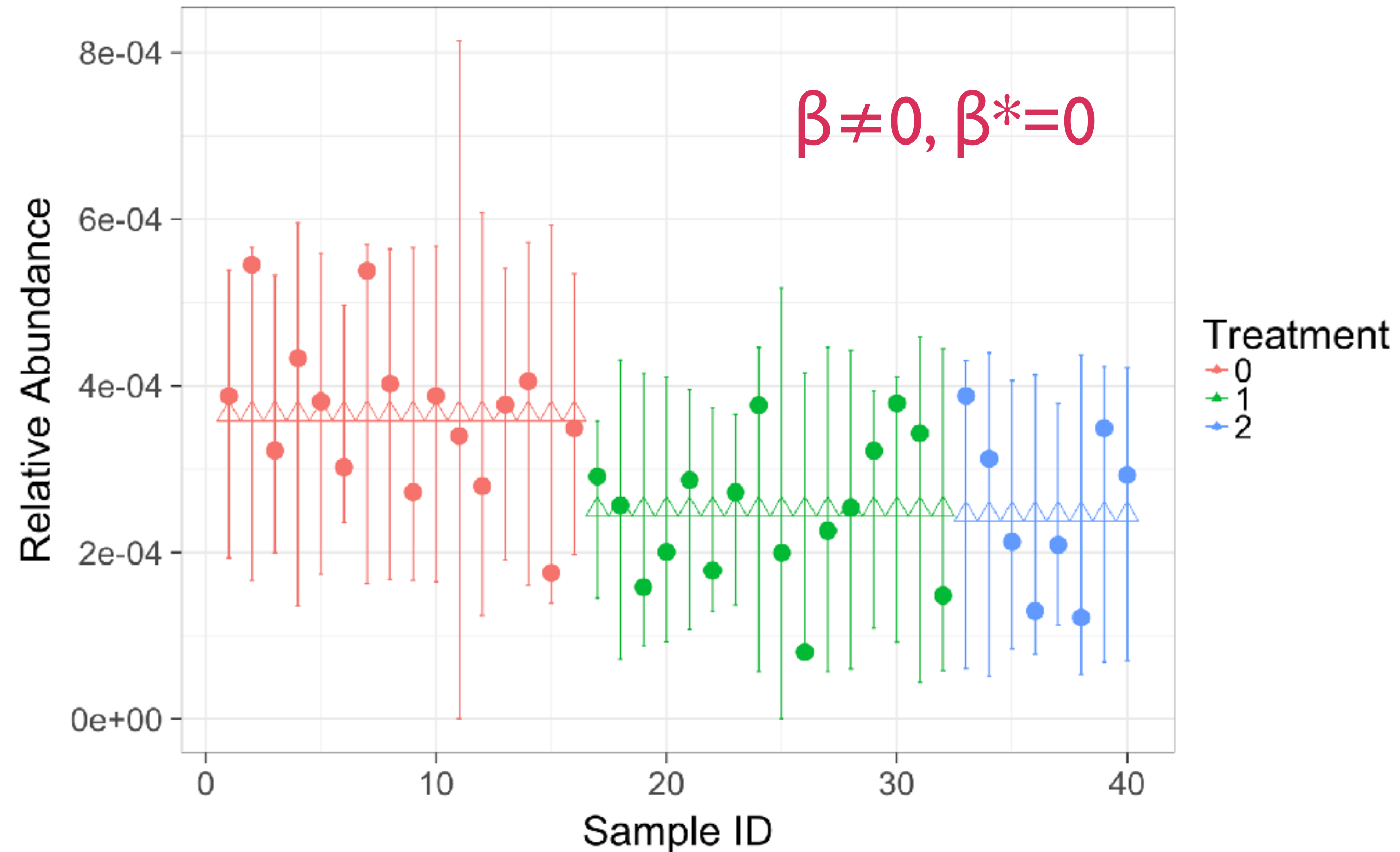
corncob

- Designed for marker gene (compositional) data
- Models relative abundance, overdispersion, and correlation parameters
- Different structure for different taxa
- Uses within-taxon correlation to model zeros

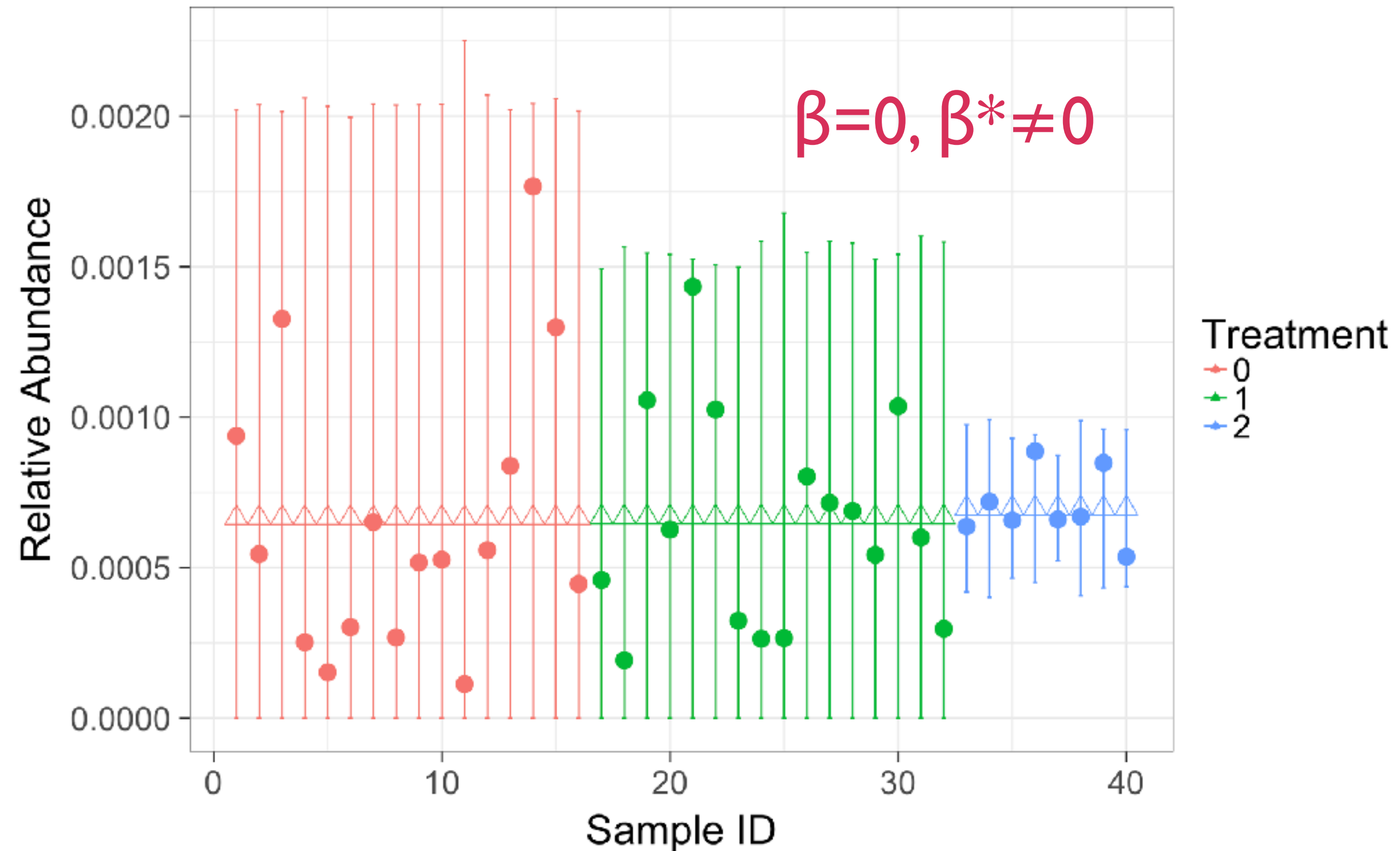
DESeq2

- Designed for RNAseq (different data structure)
 - Tests changes in abundance
 - Constrained dispersion
 - Individual microbes are assumed independent
-

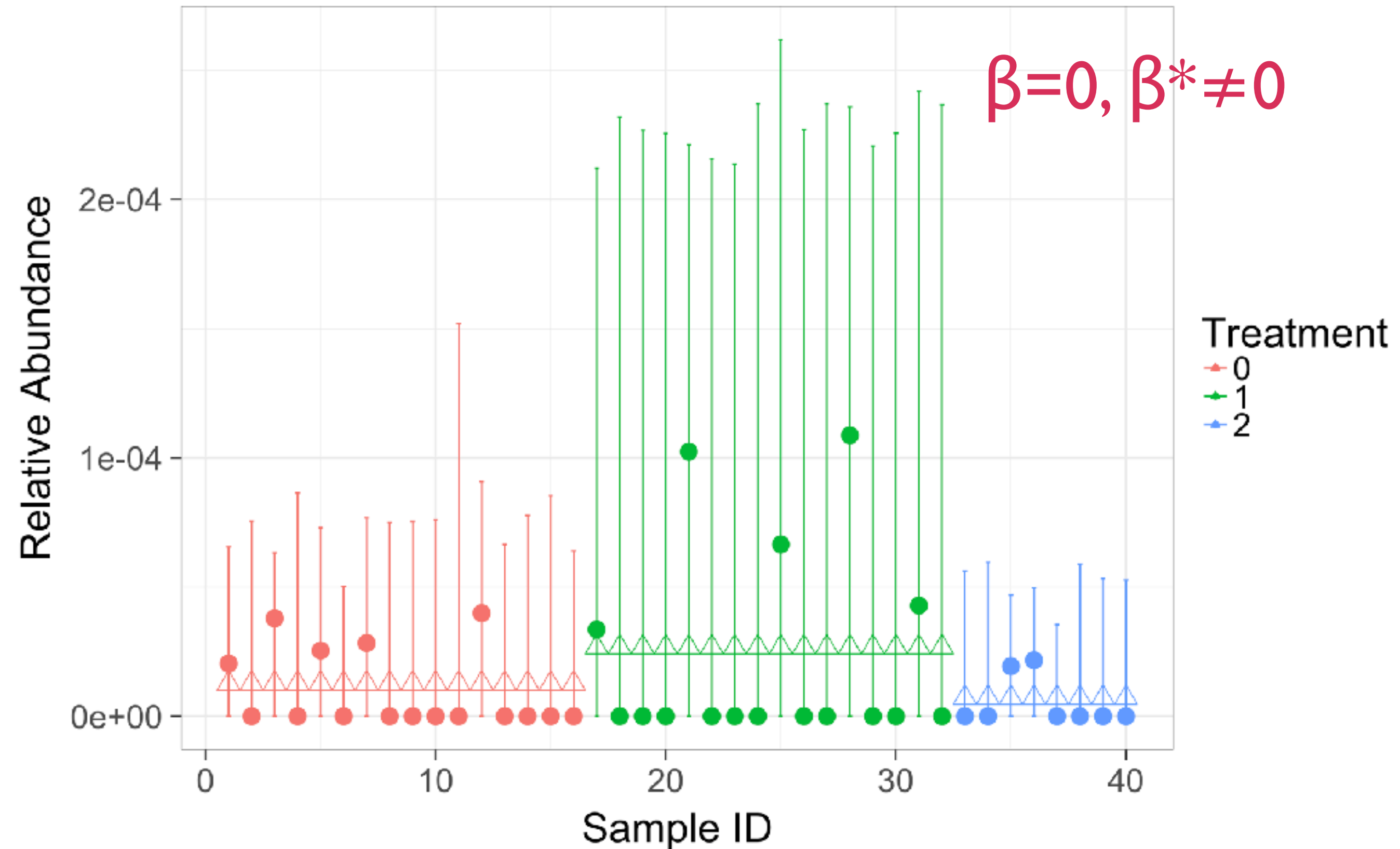
Rare taxon, different means, same variance



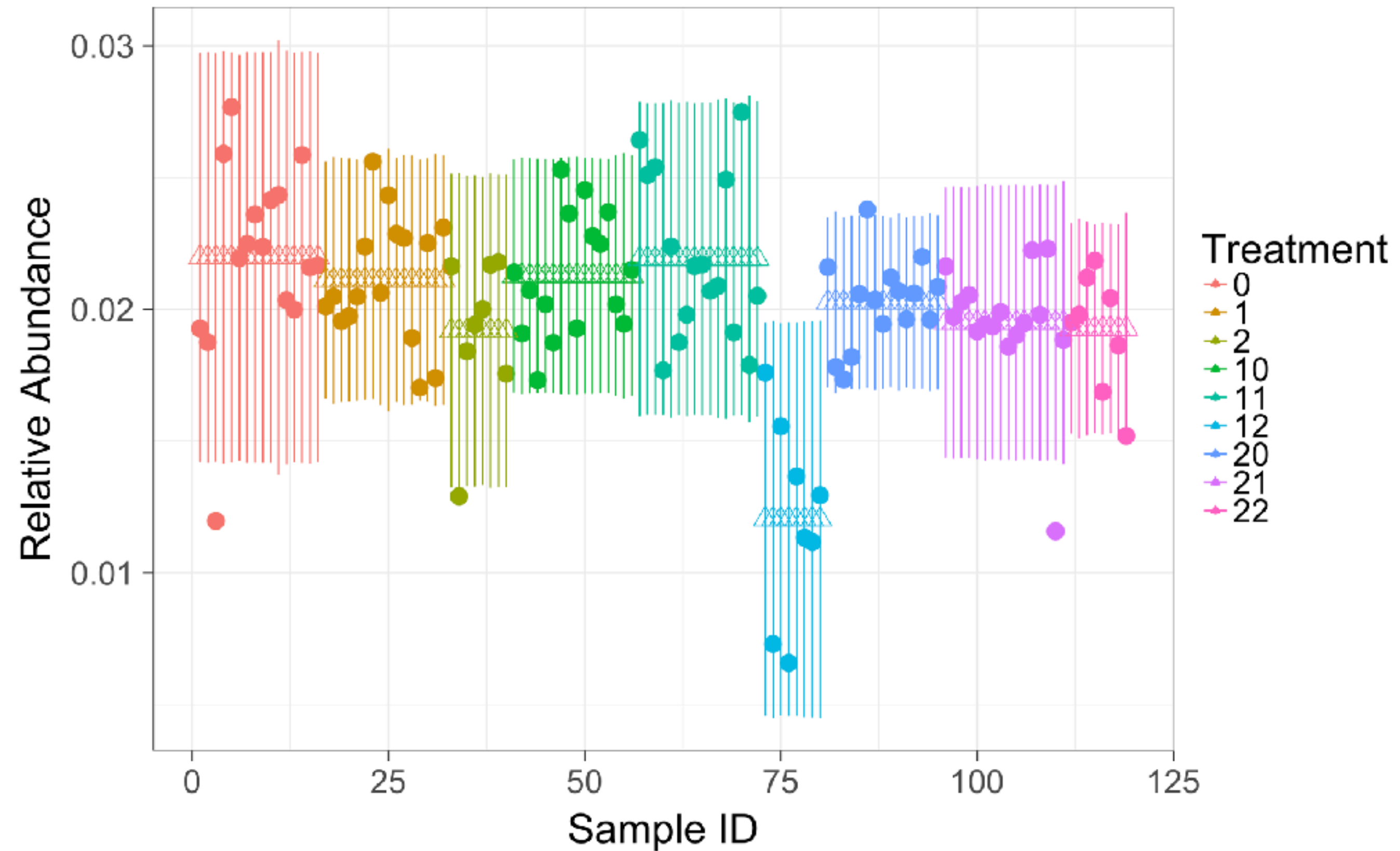
Different variance, same mean abundance



Rare taxon with high variability

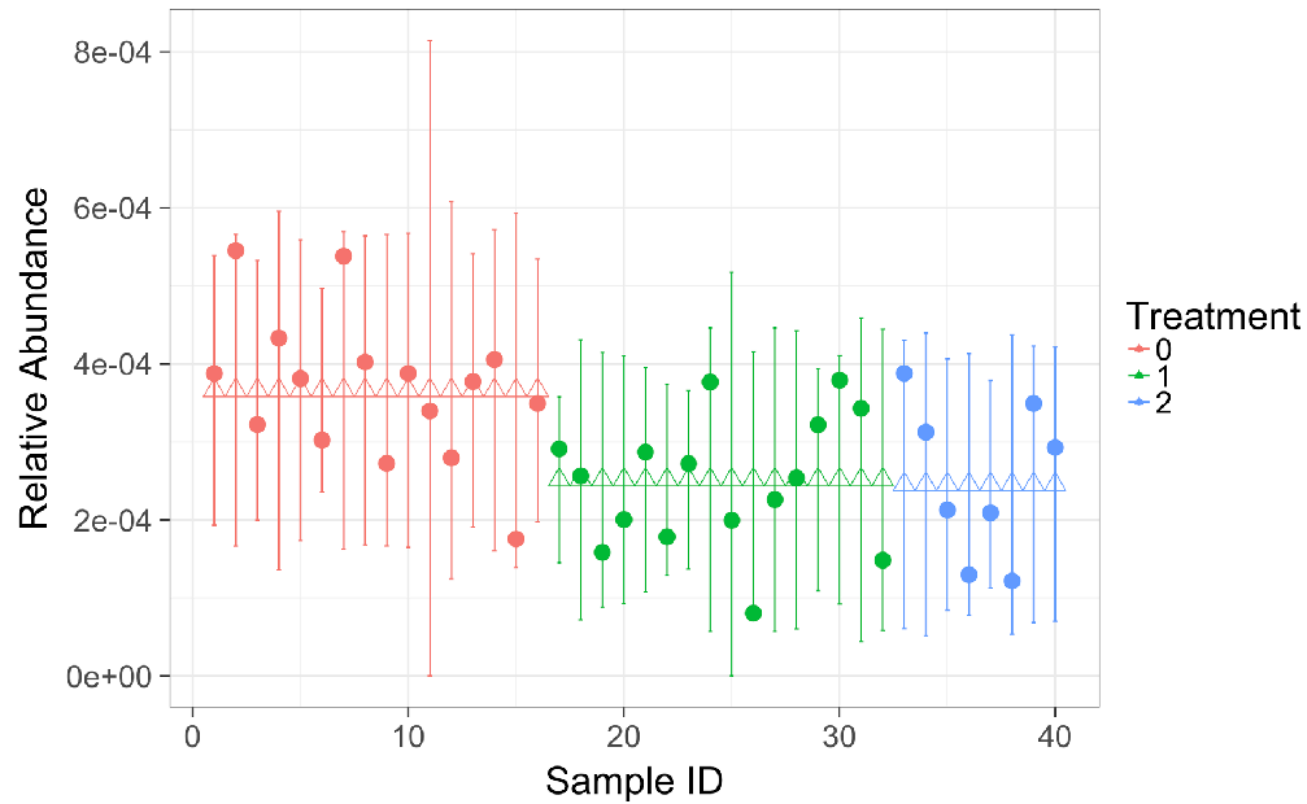


Many classes

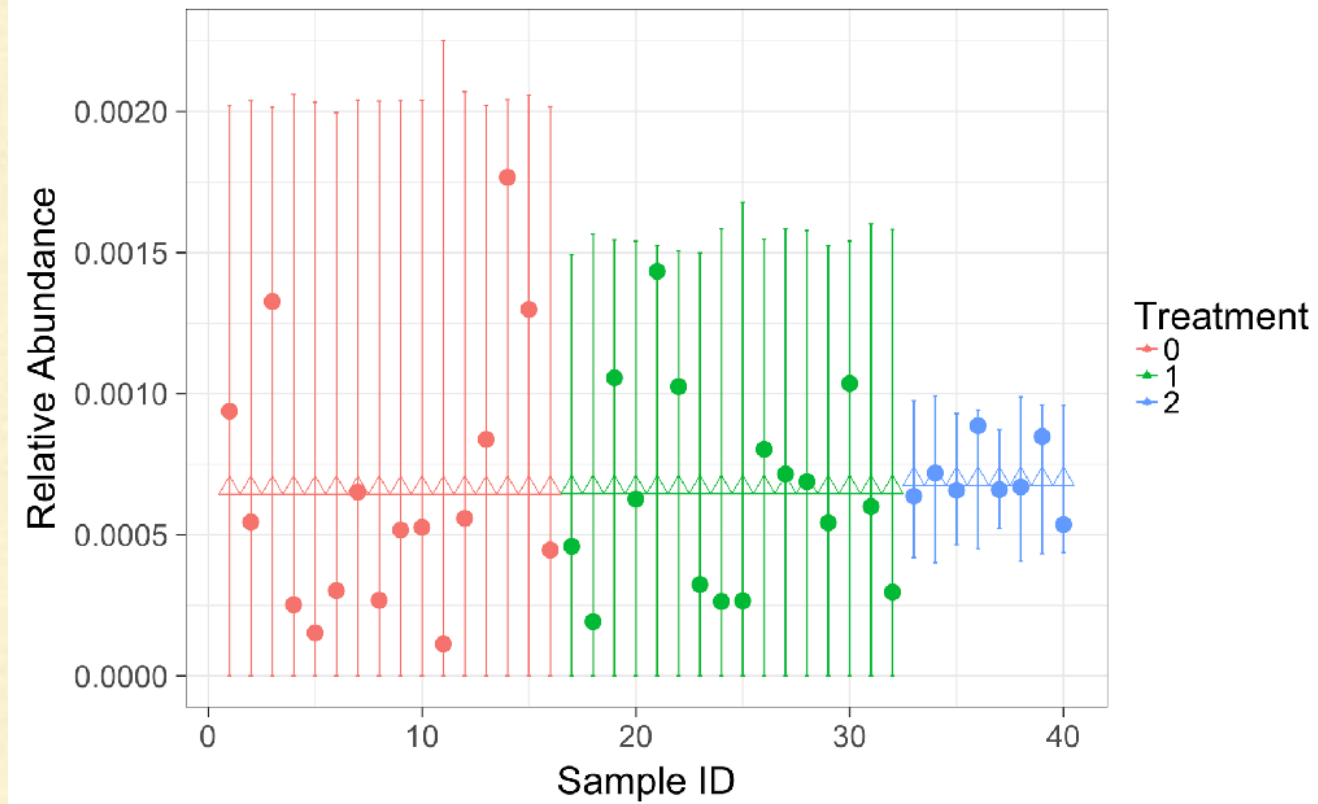


Model Fit

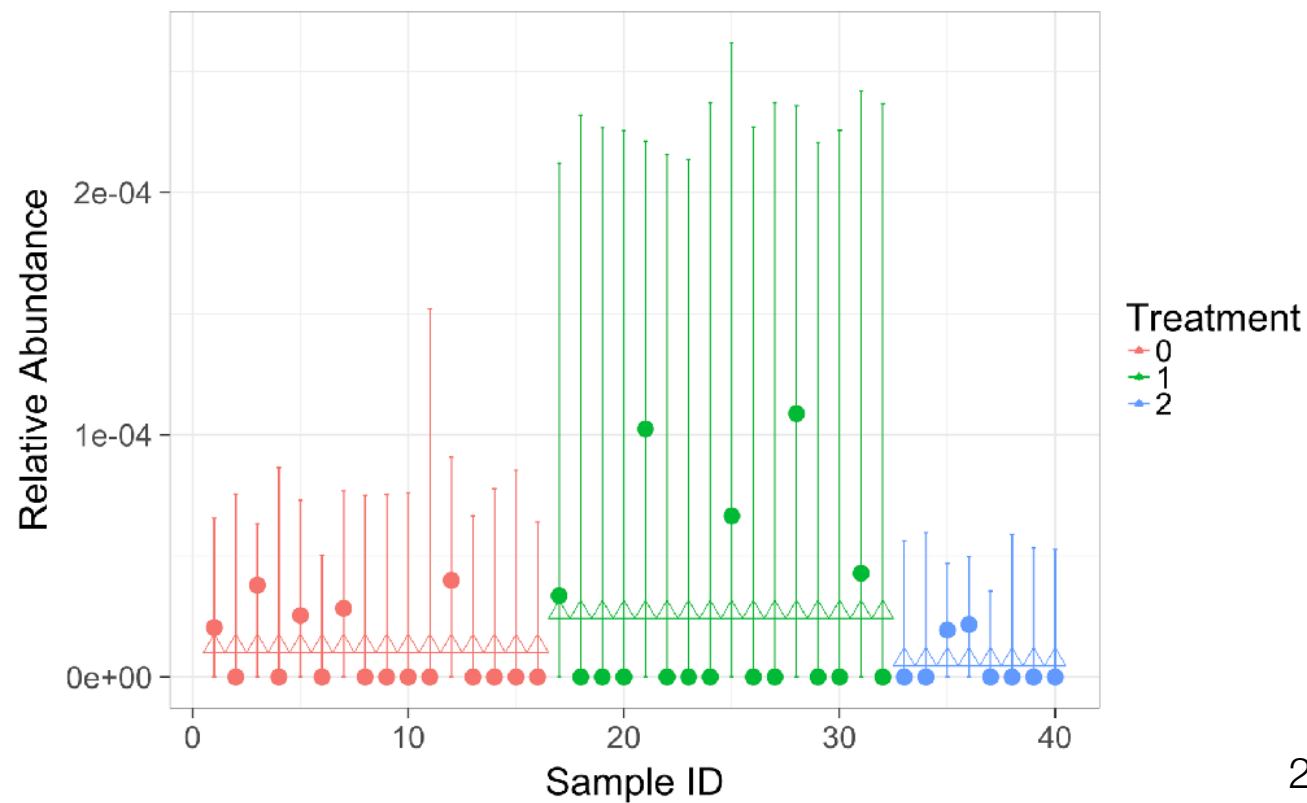
Rare taxon, different means, same variance



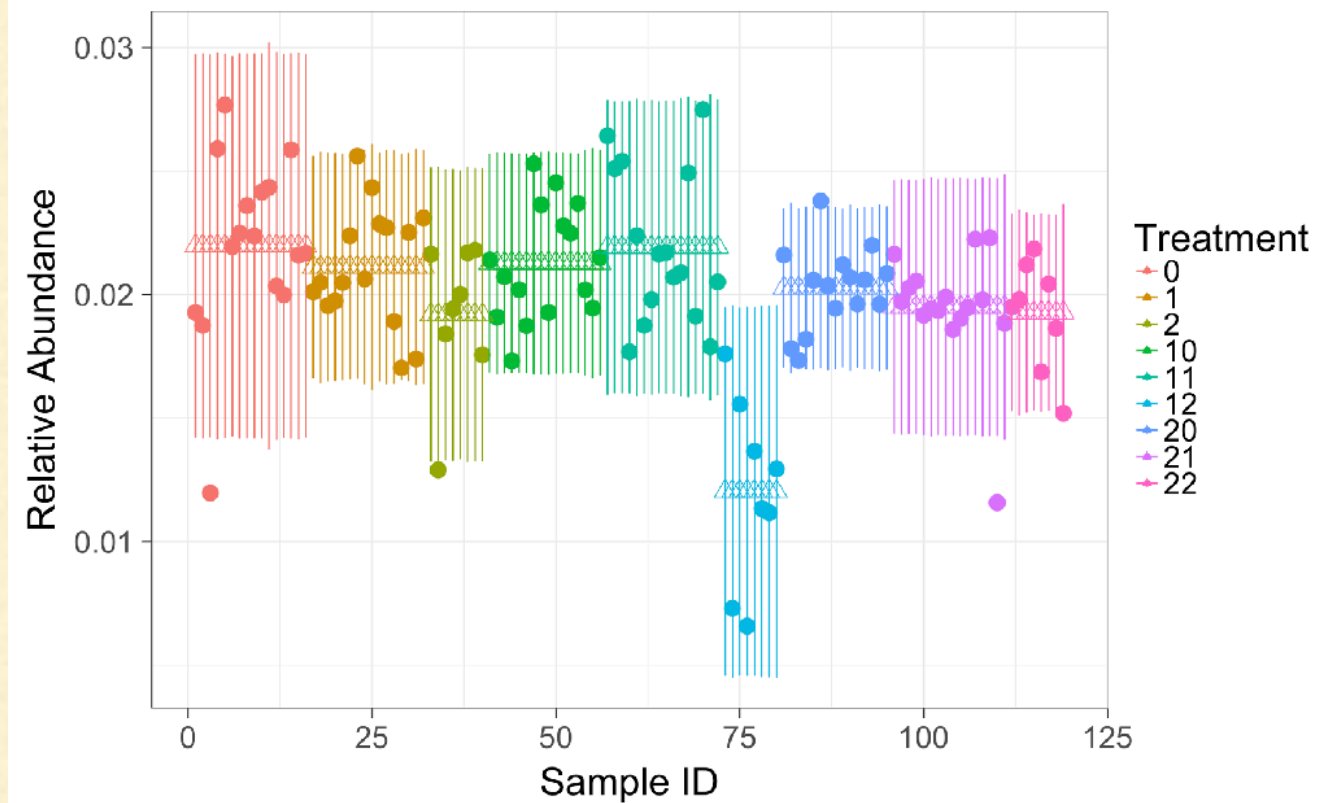
Different variance, same mean abundance



Rare taxon with high variability



Many classes



DETAILS

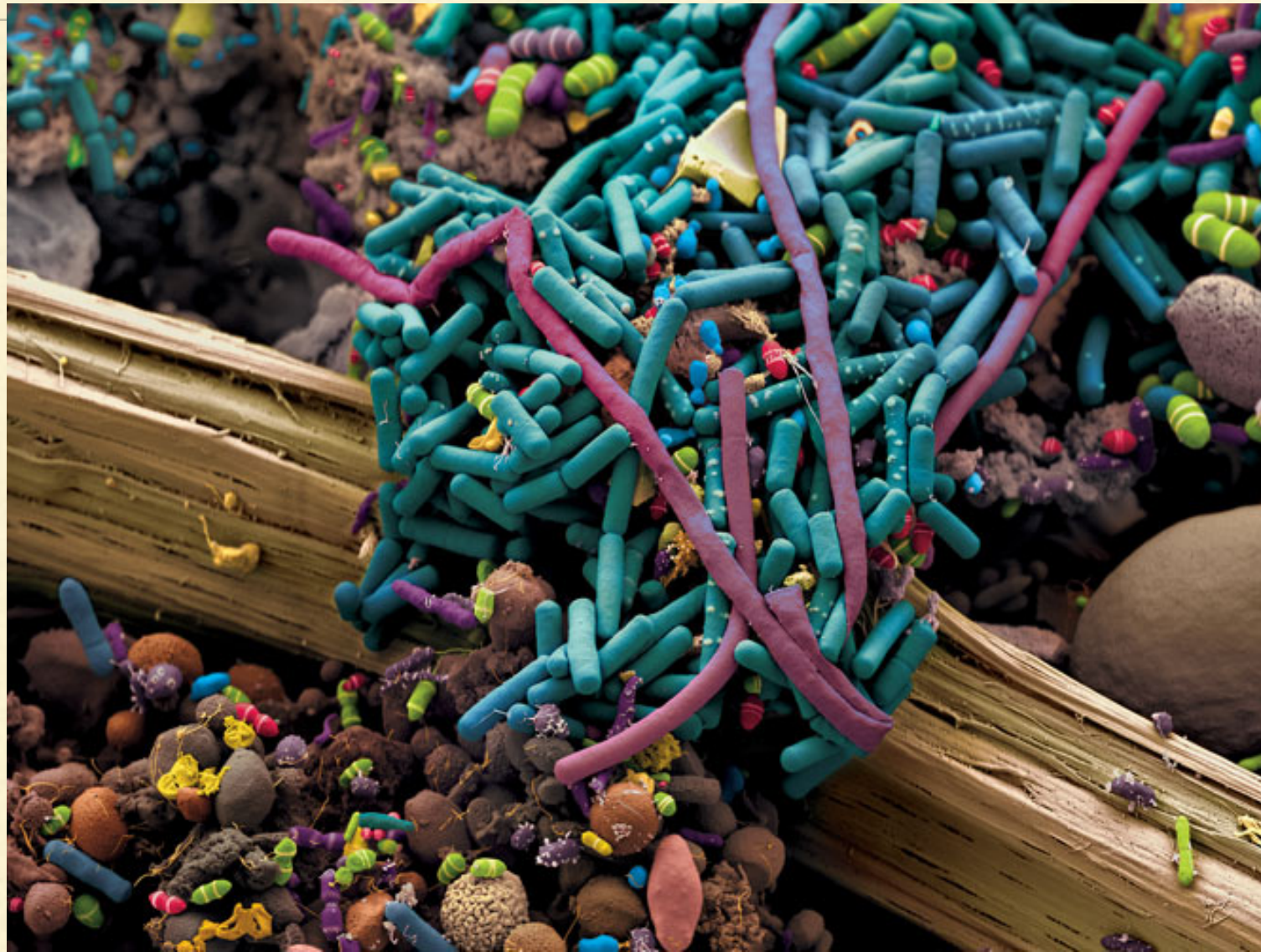


- Analytic gradient and Hessian to optimize likelihood (Fast!)
- Parametric bootstrap hypothesis testing framework
 - More samples better, but works with few samples
- Handles zero counts well
- Available at github.com/bryandmartin/CORNCOB
 - Martin, Witten & Willis, 2018+, *In Prep*

SUMMARY: CORNCOB



- Modeling and testing relative abundances using marker gene abundances
- Adjusts for sequencing depth
- Hypothesis testing for mean and overdispersion
- Allows for valid hypothesis testing with small sample sizes
- False discovery rate control for testing many taxa



MODELING ABUNDANCES

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



@AmyDWillis



adwillis@uw.edu

Bryan D Martin, PhD Candidate, Department of Statistics, UW



@BryanDMartin_



bmartin6@uw.edu

ABUNDANCES LAB

- github.com/adw96/stamps2018/
- Click on **estimation**
- Click on **corncob-tutorial.R**
- Click on **raw**
- Copy and paste into an R script and start reading and working through
 - ~30 minutes