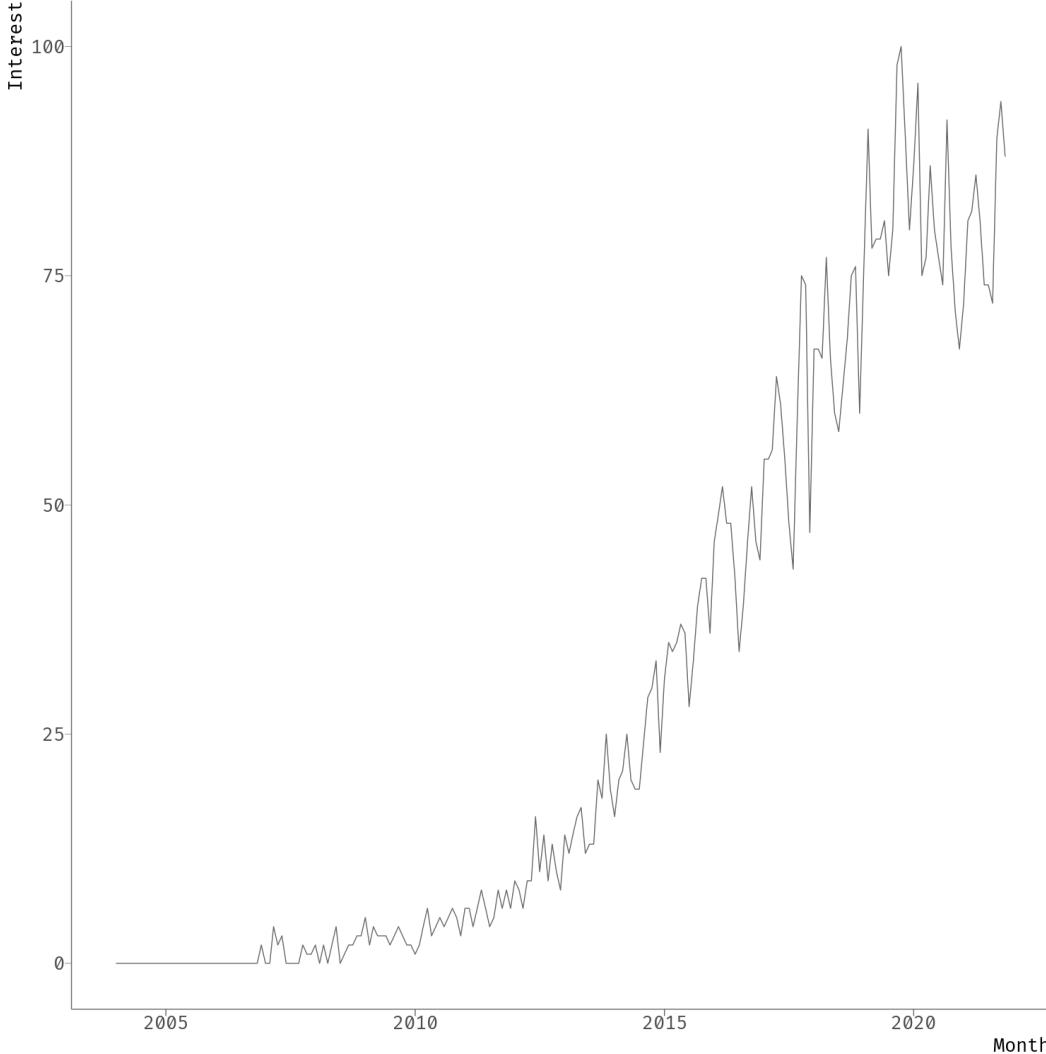




COMPARATIVE META-TAXONOMY

Roberto Siani,
24.11.2021



MICROBIOME RESEARCH IS A GROWING TREND

“the increasing amount and dimension of microbiome data may be drowning us in “significant” but irreproducible findings” (Willis, 2019)

META-GENOMICS COMES WITH MAYOR CHALLENGES

(Grimm, 2021)

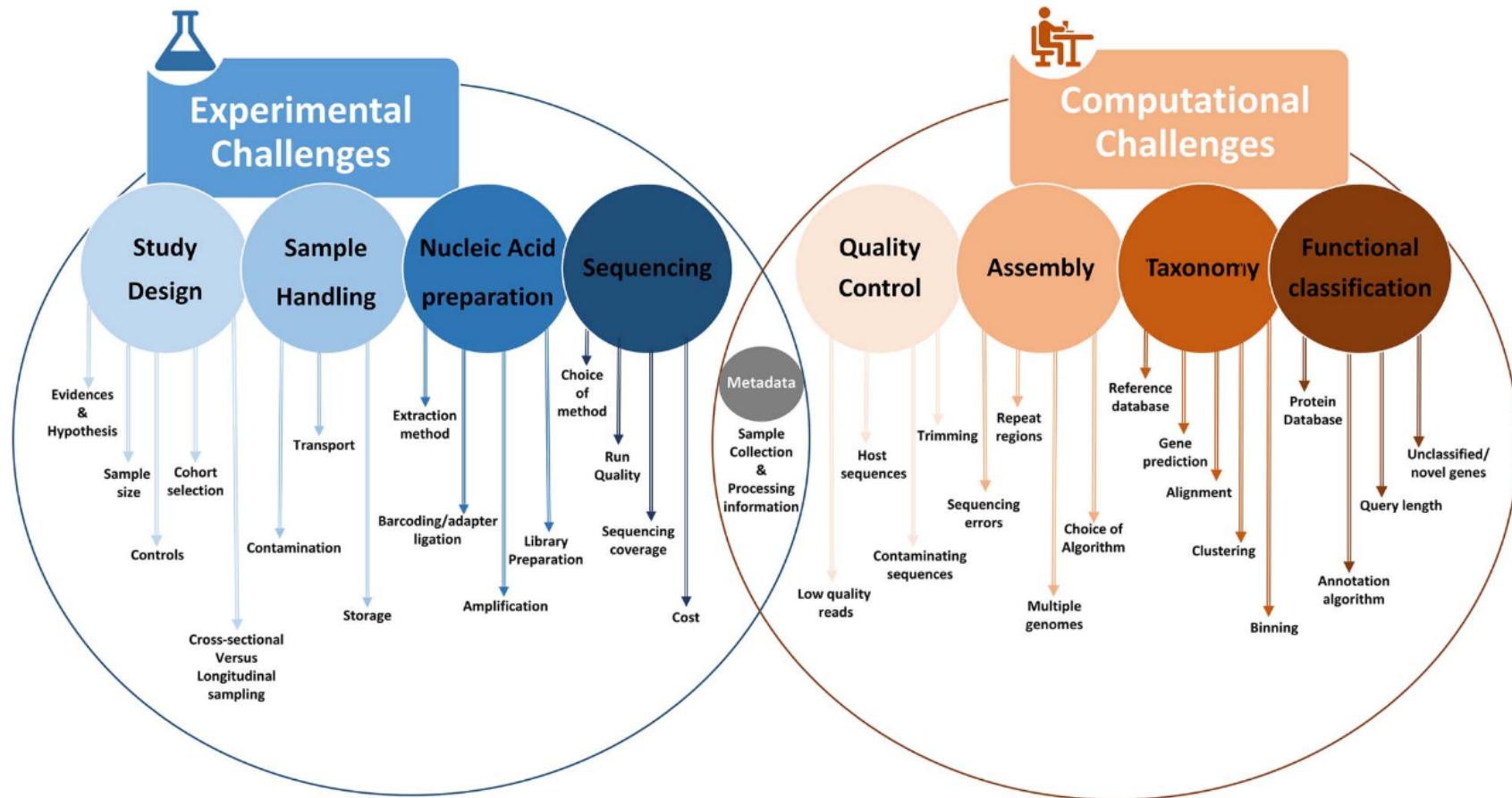
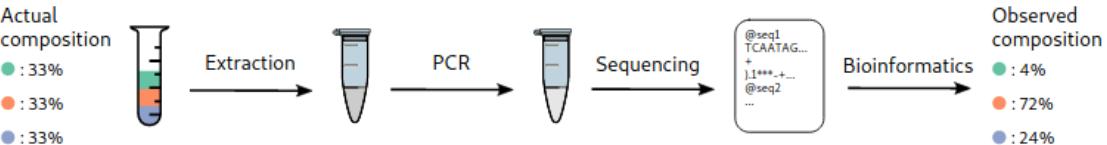


Figure 2. A schematic overview outlining various experimental and computational challenges associated with 16S rRNA-based and shotgun metagenomic sequencing.

SOME OF THE ISSUES:

A



B

Actual composition	Extraction bias	PCR bias	Sequencing bias	Bioinformatics bias	Total bias	Normalize to 100%	Observed composition
Green: 33%	1	x1	x1	x1	x1 (= 1x1x1x1)		Green: 4%
Red: 33% or 1	x 4	x 6	x 0.5	x 1.5	x 18 (= 4x6x0.5x1.5)		Red: 72% or 18
Blue: 33%	1	x 15	x 2	x 0.25	x 6 (= 15x2x0.25x0.8)		Blue: 24% or 6

$\mathbf{A} \quad \mathbf{B}^{(P_1)} \quad \mathbf{B}^{(P_2)} \quad \mathbf{B}^{(P_3)} \quad \mathbf{B}^{(P_4)} \quad \mathbf{B}^{(P)} \sim \mathbf{B}^{(P_1)} \cdot \dots \cdot \mathbf{B}^{(P_4)}$

$\mathbf{O} \sim \mathbf{A} \cdot \mathbf{B}^{(P)}$

- Several random subsampling steps (McLaren, 2019)

- Compositionality

- Interdependence

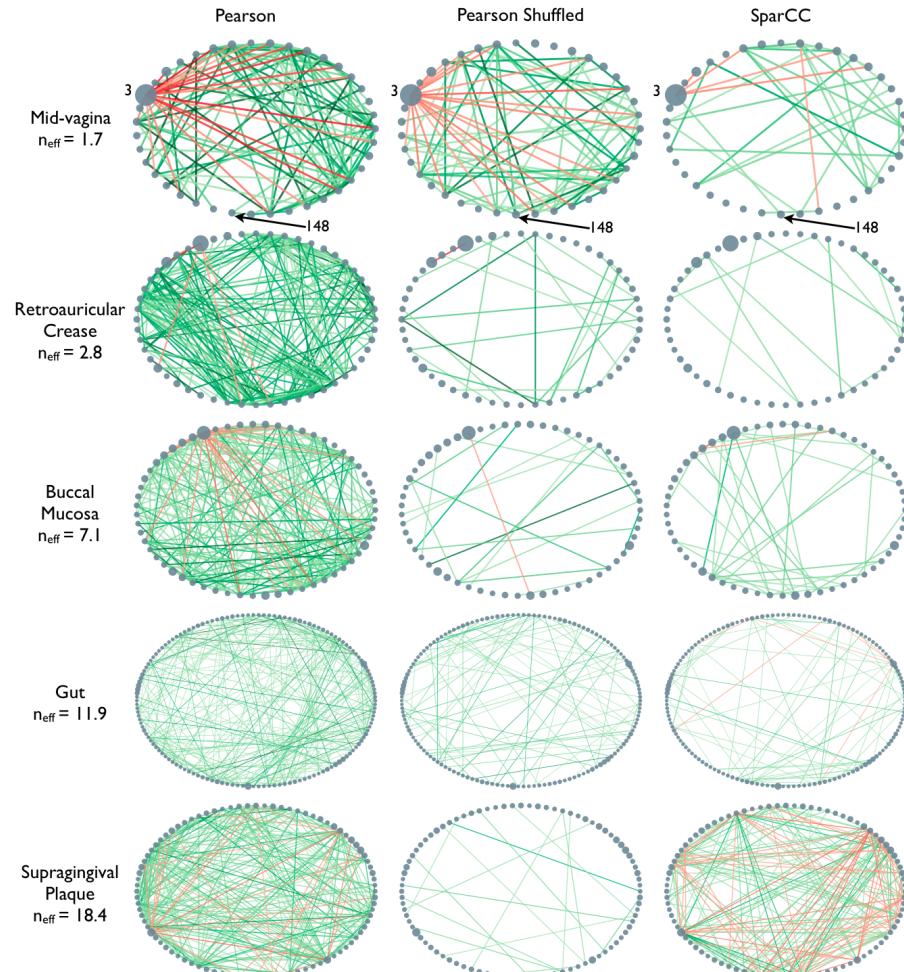
- Structural and sampling zeroes

- Sparseness

SOME OF THE ISSUES:

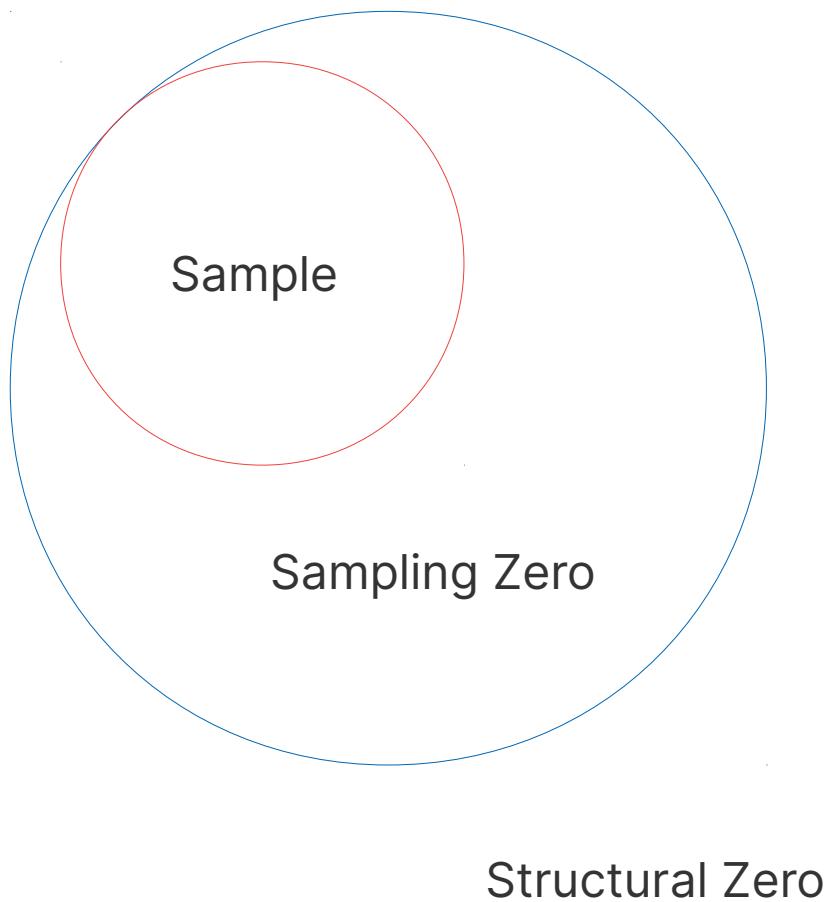


- Several random subsampling steps
- **Compositionality**^(Aitchison, 1982)
- Interdependence
- Structural and sampling zeroes
- Sparseness



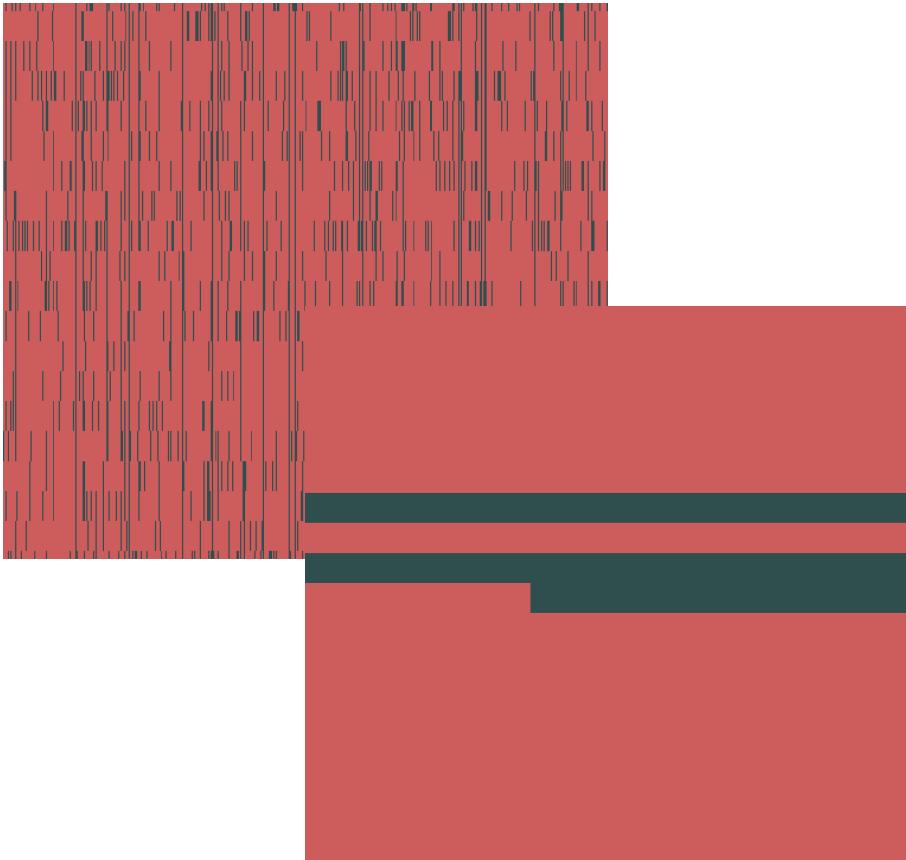
SOME OF THE ISSUES:

- Several biased subsampling steps
- Compositionality
- **Interdependence**^(Friedman, 2012)
- Structural and sampling zeroes
- Sparseness



SOME OF THE ISSUES:

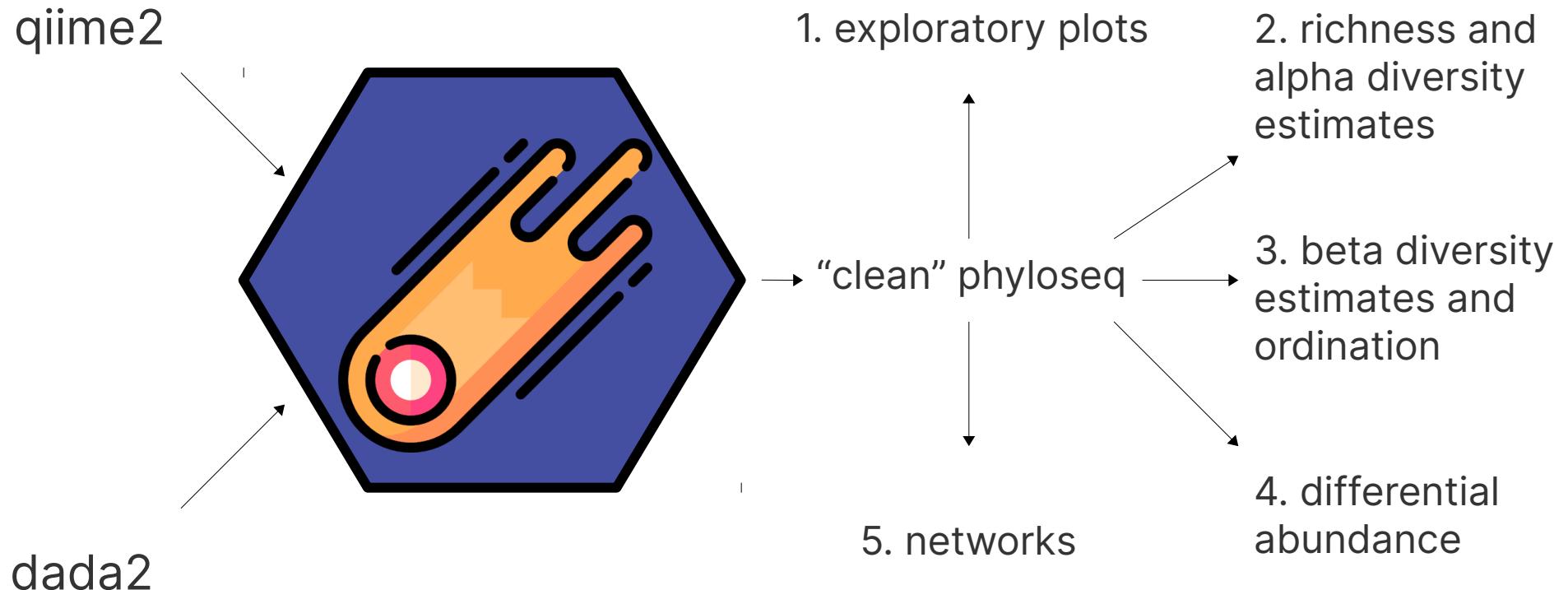
- Several random subsampling steps
- Compositionality
- Interdependence
- **Structural and sampling zeroes** (Blasco-Moreno, 2019)
- Sparseness

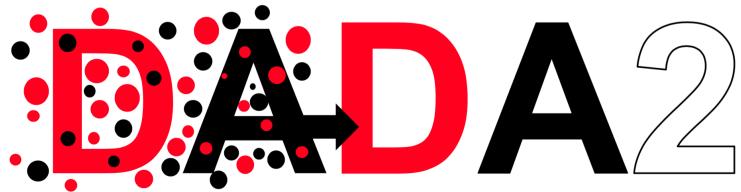


SOME OF THE ISSUES:

- Several random subsampling steps
- Compositionality
- Interdependence
- Structural and sampling zeroes
- **Sparseness** (Xu, 2015)

CoMet: R pipeline for comparative metataxonomy



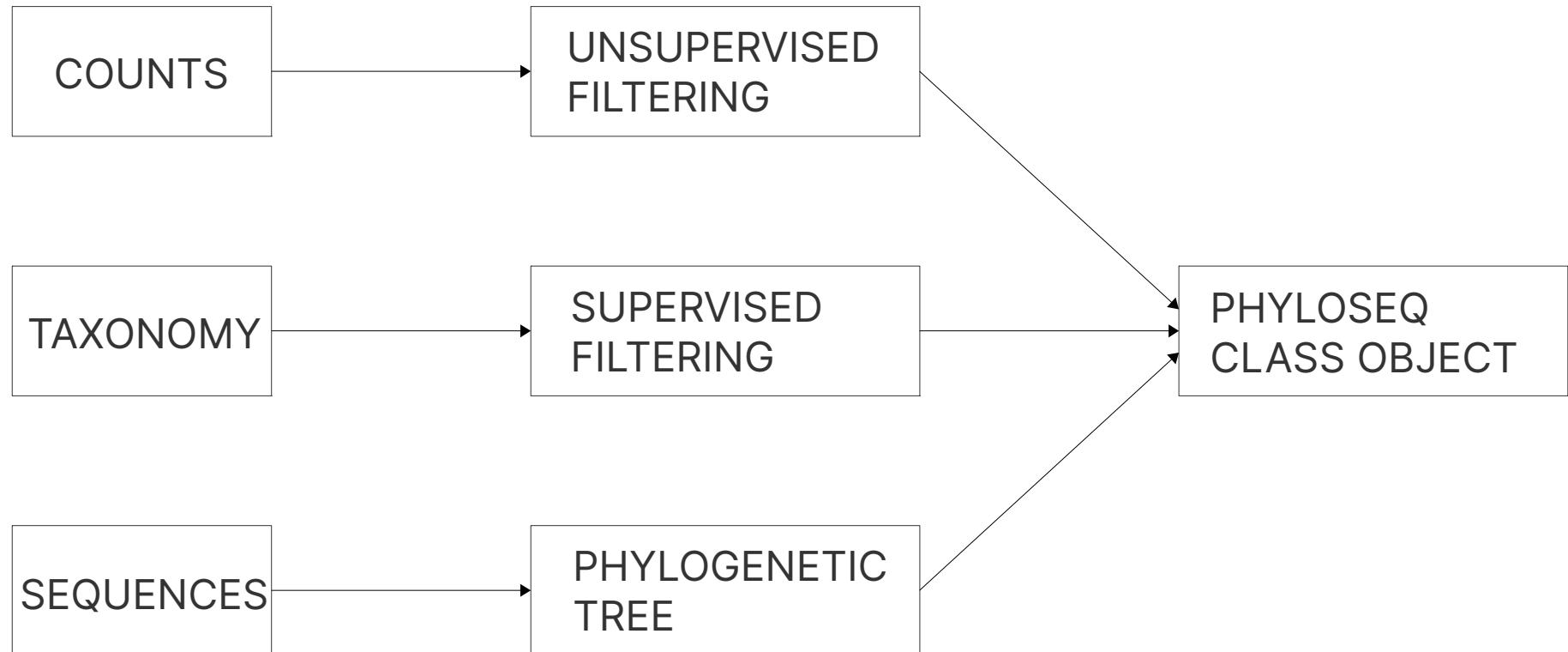


0. INPUT

Counts table*, taxonomic assignment*, metadata, representative sequences

*qza from qiime2 or data tables from dada2

0. PRE-PROCESSING



0. NORMALIZATION?

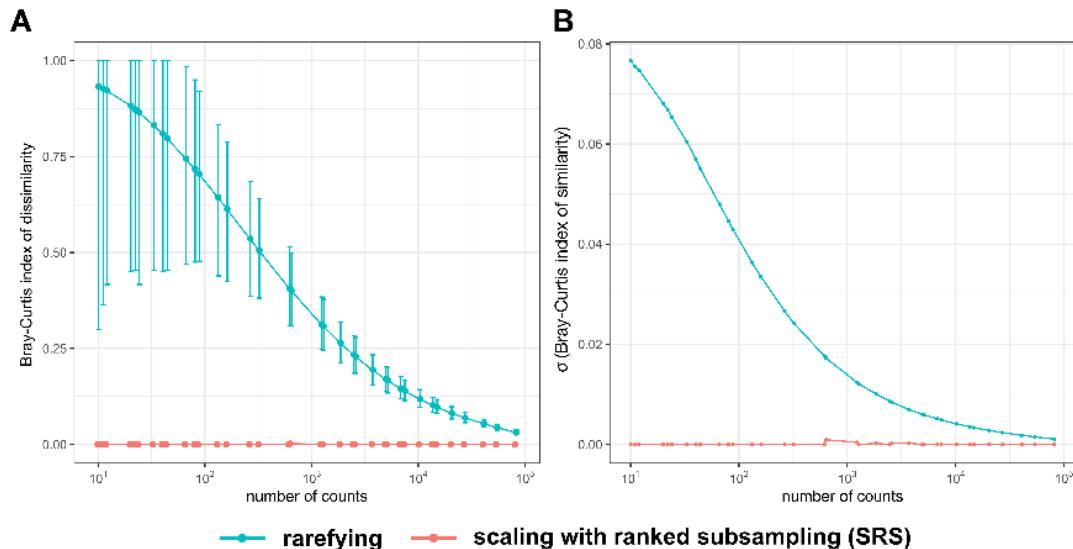
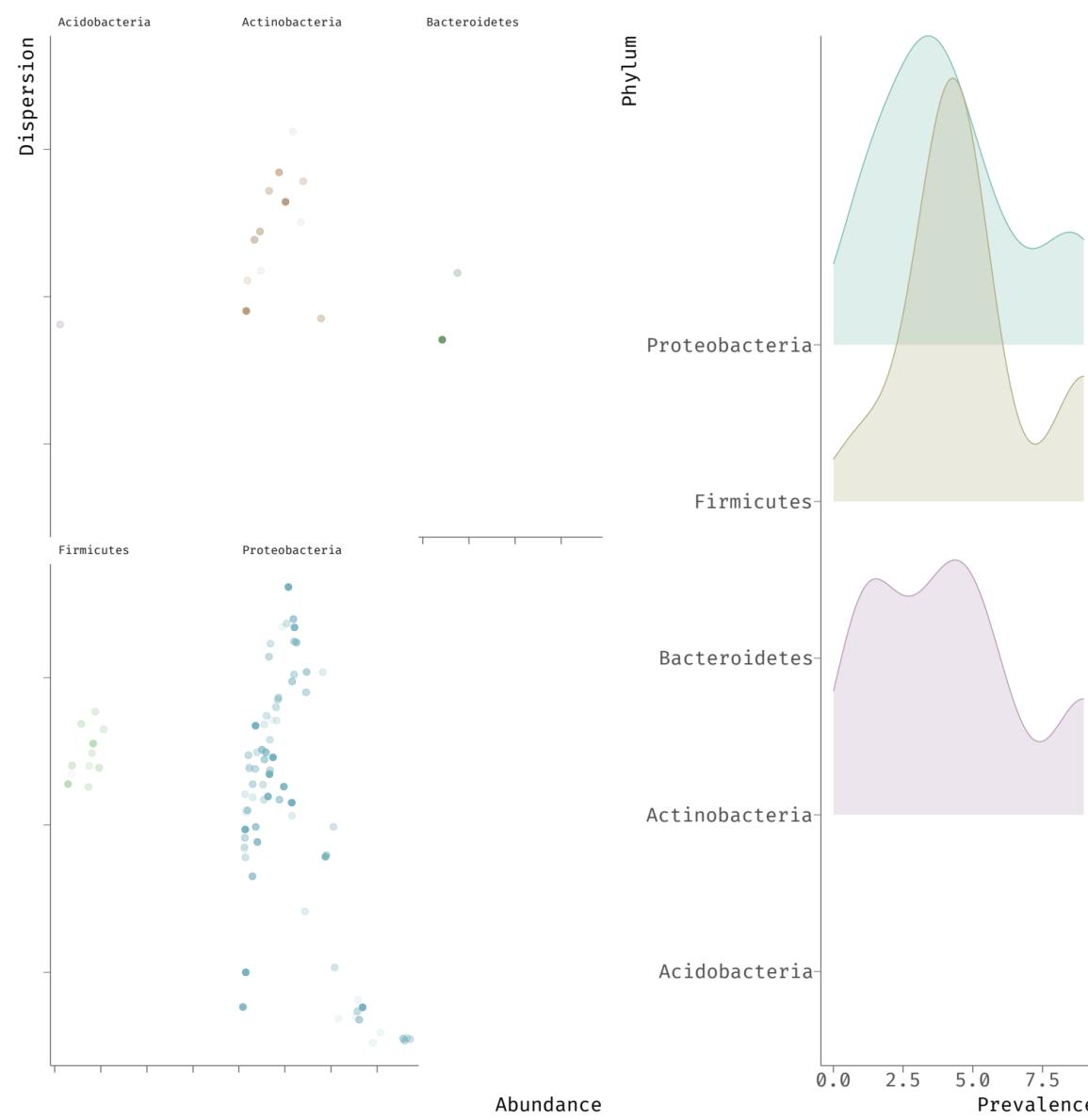


Figure 3 Bray-Curtis index of dissimilarity (A) among 10,000 replications of the normalized test library and its standard deviation (σ) (B) normalized by rarefying or SRS. The sampled numbers of

Rarefaction^(Beule, 2020):

- 1) loss of information
- 2) additional uncertainty
- 3) arbitrary library size

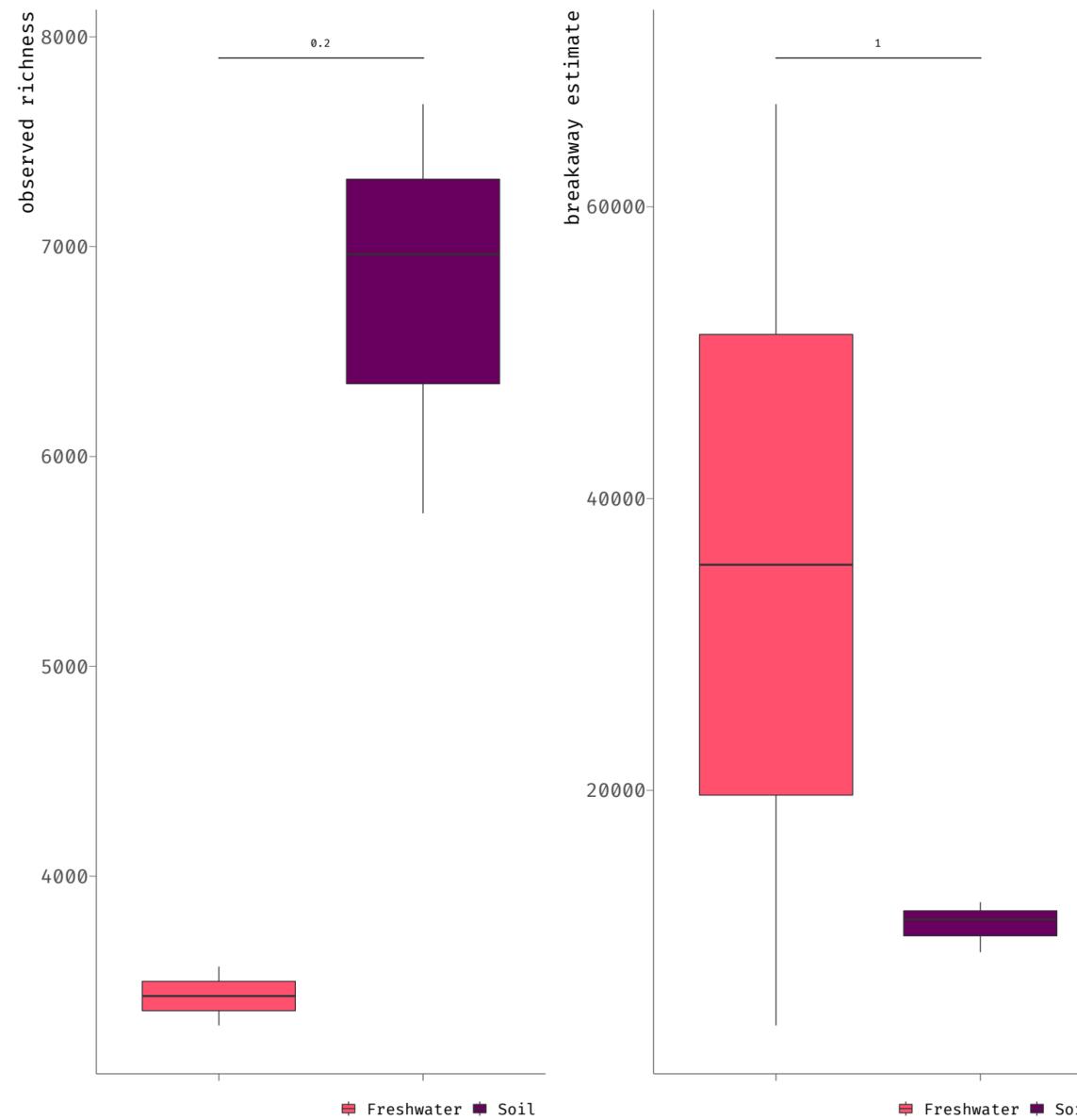


1. EXPLORATORY PLOTS

Taxa abundance vs dispersion

Prevalence distribution across phyla

Tree, relative abundance can also be informative



2. RICHNESS AND DIVERSITY ESTIMATES

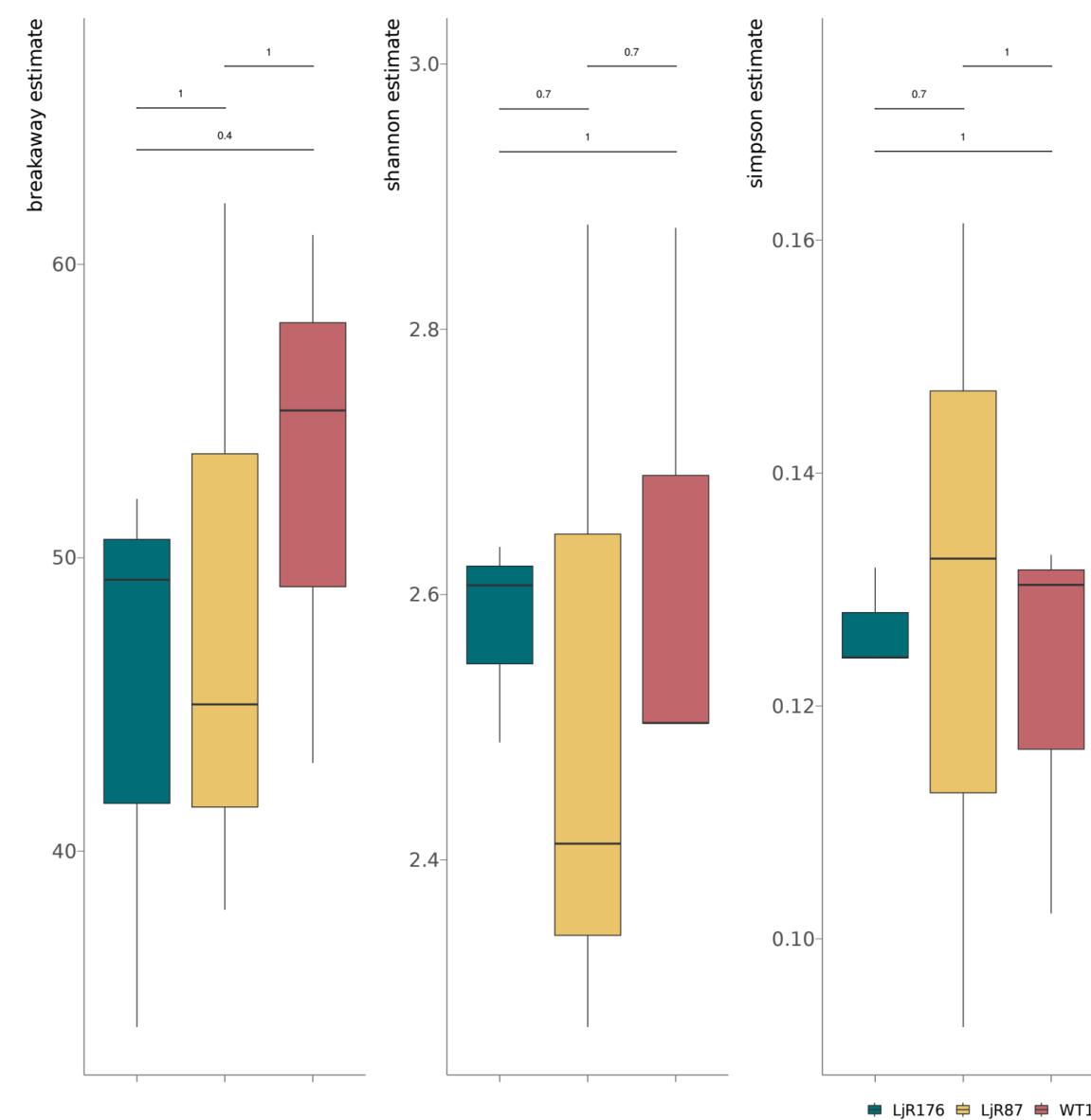
Breakaway^(Willis, 2015) for species richness and modeling

DivNet^(Willis, 2020) estimates shannon and simpson diversity indexes incorporating network structure

2. RICHNESS AND DIVERSITY ESTIMATES

Breakaway^(Willis, 2015) for species richness and modeling

DivNet^(Willis, 2020) estimates shannon and simpson diversity indexes incorporating network structure

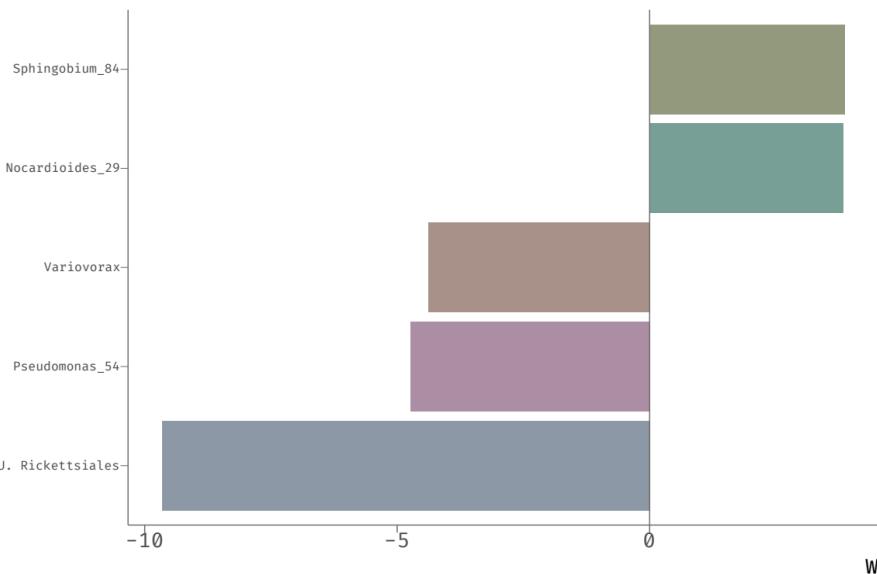
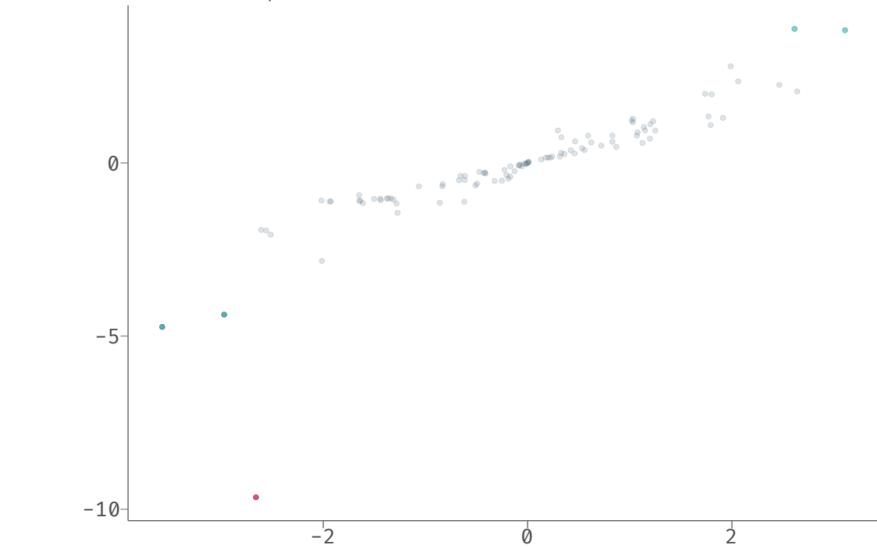


3. BETA DIVERSITY ESTIMATES AND ORDINATION

DivNet also models Bray-Curtis dissimilarities and Aitchison distance (euclidean distance in the simplex)

Visualisation is done by singular value decomposition (Gloor, 2017) after a centred log-ratio transform

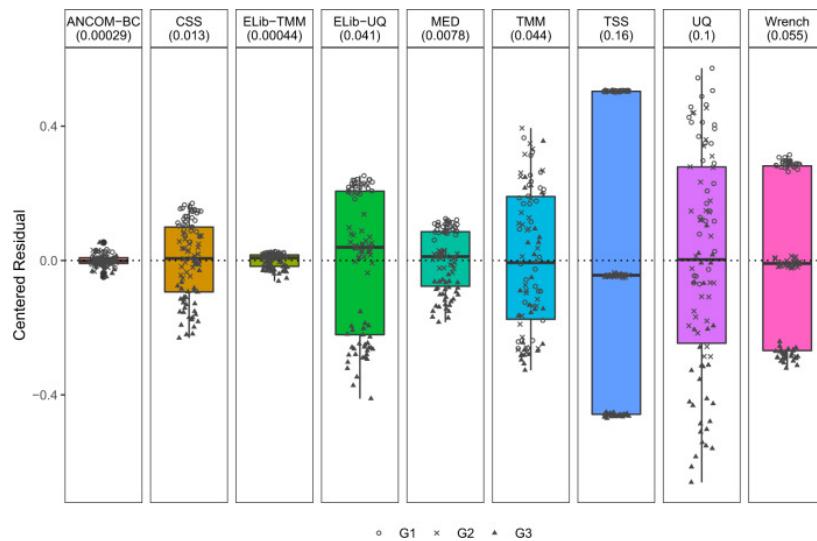
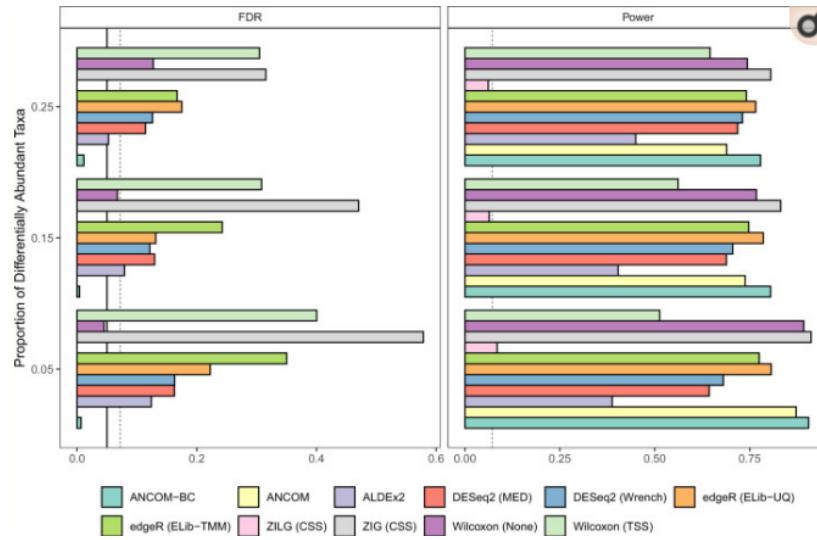
← Control | Treatment →



4. DIFFERENTIAL ABUNDANCE

Analysis of Compositions of Microbiomes with Bias Correction^(Lin, 2020a):

- estimates sampling fraction
- linear regression
- confidence intervals and adjusted p-values

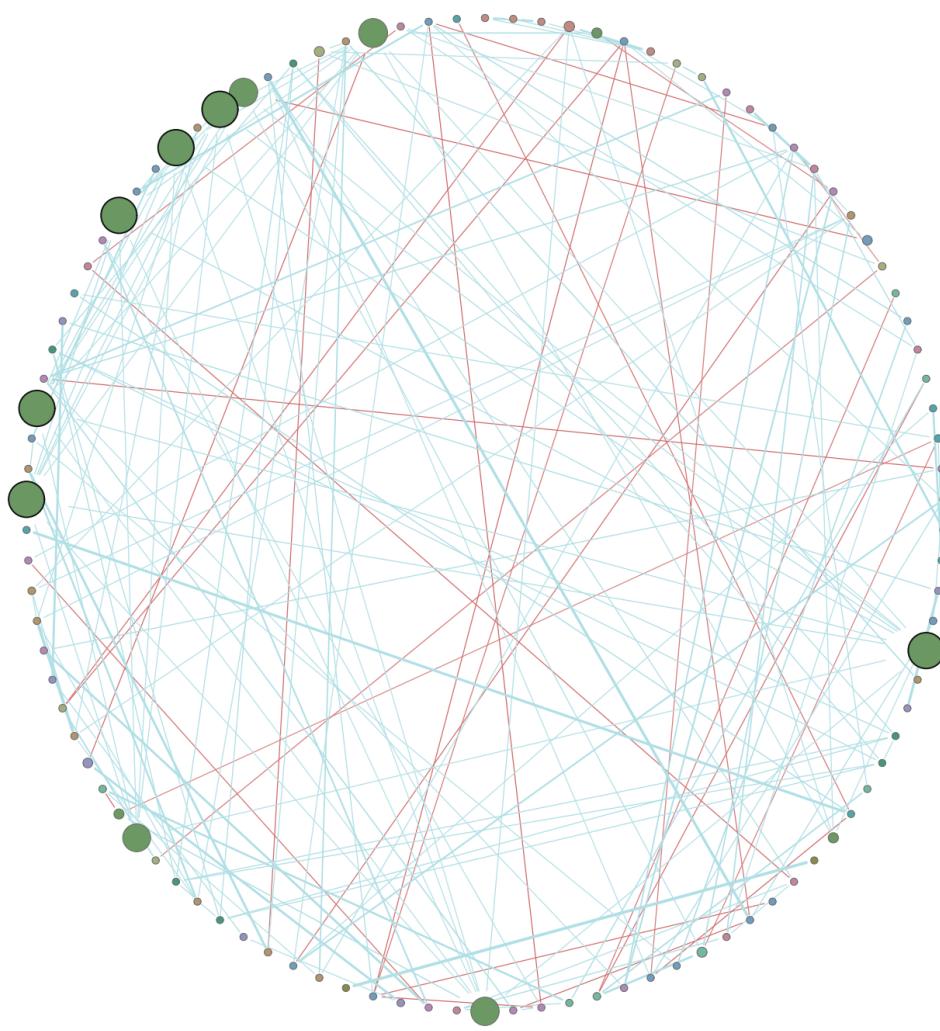


4. DIFFERENTIAL ABUNDANCE

Analysis of Compositions of Microbiomes with Bias Correction^(Lin, 2020b):

- estimates sampling fraction
- linear regression
- confidence intervals and adjusted p-values

Network on OTU level with SPRING associations



5. NETWORKS

NetCoMi (Peschel, 2020) provides an elastic framework to use novel tools such as SPRING (Yoon, 2019) and SparCC (Kurtz, 2015)

CoMet:

- will be available at <https://github.com/rsiani/CoMet>
- follows modern statistical frameworks
- incorporates state of the art tools
- provides a skeleton for your microbiome analyses, but does not cover every possible cases
- is mostly meant for “in house” use, but might be further developed in a R package
- must be approached with skepticism!**



THANKS!

Coming soon(ish):
relative abundance
estimates, improved
documentation, self-
contained packaging, tools
for multivariate analyses
(hypothesis testing,
variable selection, machine
learning)