16S ANALYSIS: FROM OTU TO ASVS

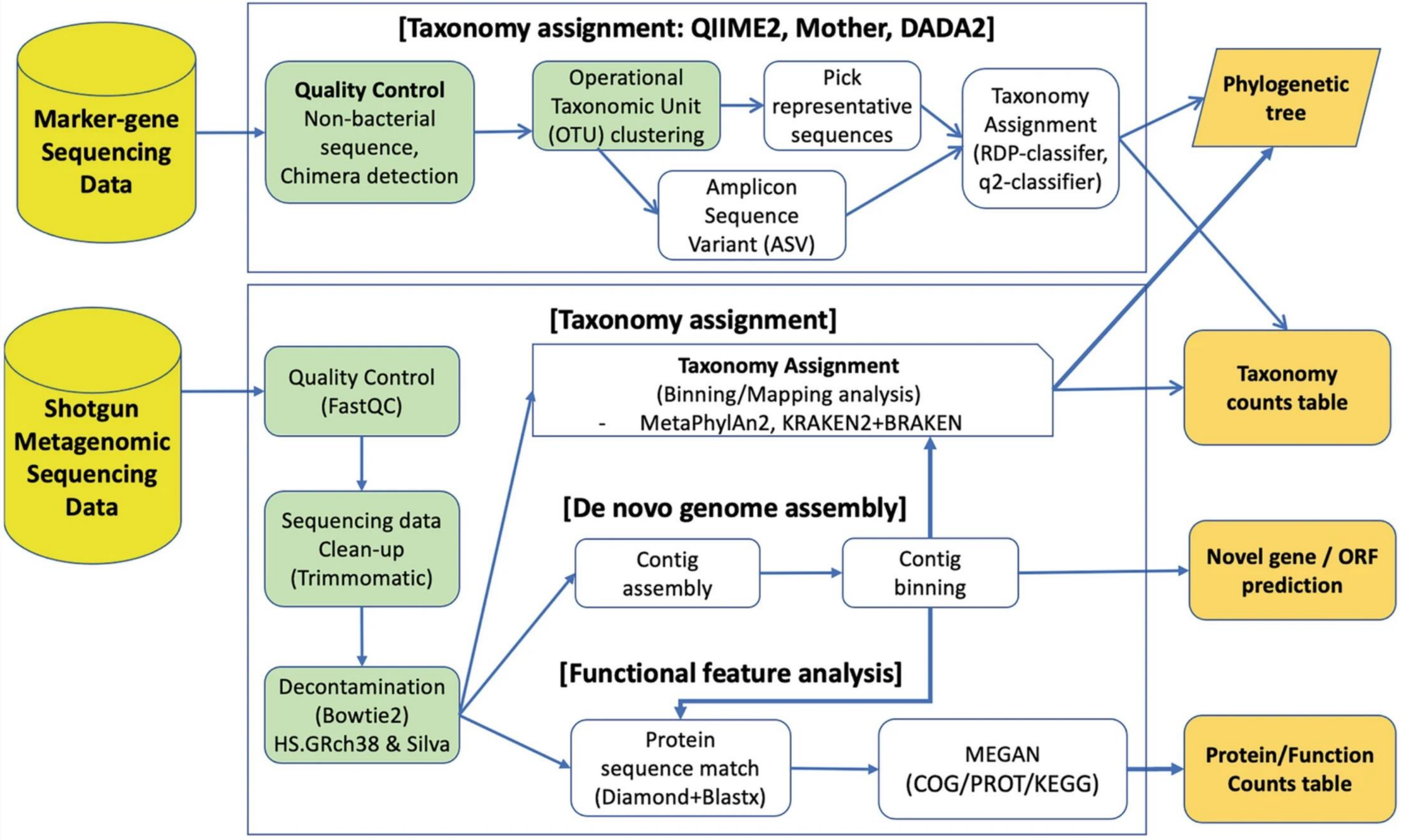
Dr. Brigida Rusconi

Department of Pediatrics, Washington University School of Medicine

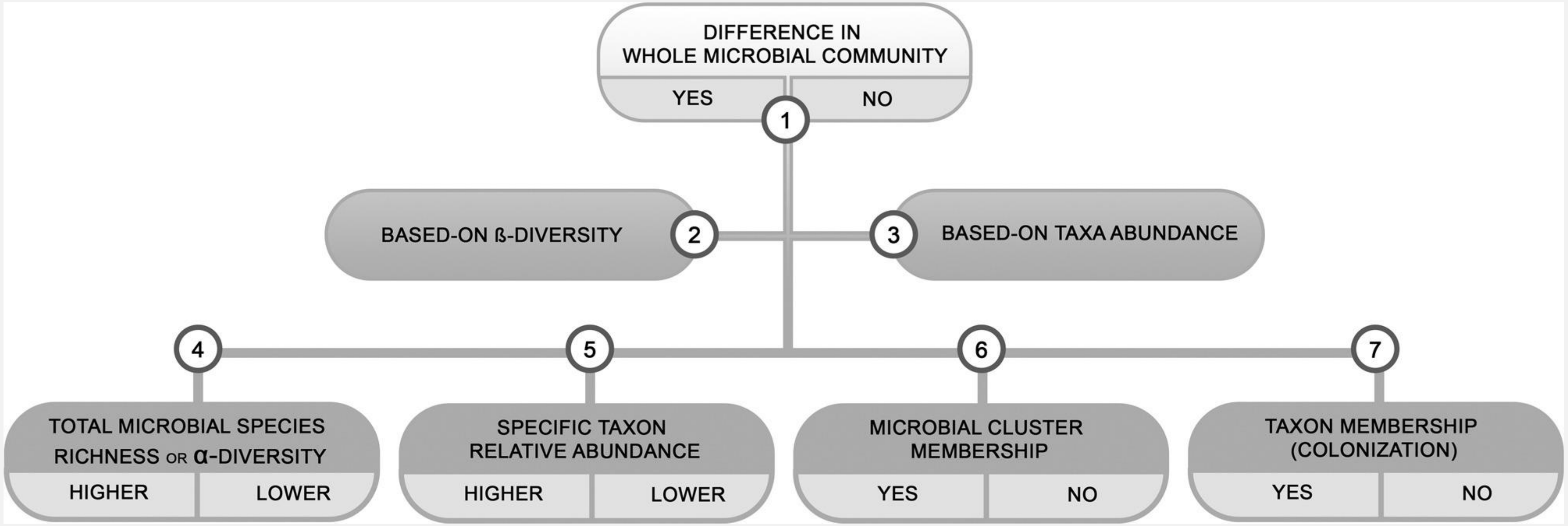
THE ROLE OF THE MICROBIOTA IN HEALTH AND DISEASE

- Many disease cannot be explained by genetics alone
  - Environmental variables that exacerbate or induce common human diseases
  - Sometimes the microbiota alone can be the trigger for disease
- Germ-free mice have an altered immune system and gut function
  - Colonization with gut microbes is required for tolerance against food antigens
  - Skews the immune system towards inflammation
- Perturbations in early life to the microbiota have life-long consequences
- Gut microbiota has systemic effects not just local

From: Bioinformatic and Statistical Analysis of Microbiome Data


[Taxonomy assignment: QIIME2, Mother, DADA2]
Marker-gene Sequencing Data
Quality Control
Non-bacterial sequence, Chimera detection
Operational Taxonomic Unit (OTU) clustering
Pick representative sequences
Amplicon Sequence Variant (ASV)
Taxonomy Assignment (RDP-classifer, q2-classifier)
Phylogenetic tree
Shotgun Metagenomic Sequencing Data
[Taxonomy assignment]
Quality Control (FastQC)
Sequencing data Clean-up (Trimmomatic)
Decontamination (Bowtie2) HS.GRch38 & Silva
Taxonomy Assignment
(Binning/Mapping analysis)
- MetaPhylAn2, KRAKEN2+BRAKEN
Taxonomy counts table
[De novo genome assembly]
Contig assembly
Contig binning
Novel gene / ORF prediction
[Functional feature analysis]
Protein sequence match (Diamond+Blastx)
MEGAN (COG/PROT/KEGG)
Protein/Function Counts table

SAMPLE SIZE


DIFFERENCE IN
WHOLE MICROBIAL COMMUNITY
YES
NO
1
BASED-ON ß-DIVERSITY
2
3
BASED-ON TAXA ABUNDANCE
4
5
6
7
TOTAL MICROBIAL SPECIES RICHNESS OR α-DIVERSITY
HIGHER
LOWER
SPECIFIC TAXON RELATIVE ABUNDANCE
HIGHER
LOWER
MICROBIAL CLUSTER MEMBERSHIP
YES
NO
TAXON MEMBERSHIP (COLONIZATION)
YES
NO

Ferdous, 2022, Muc Imm

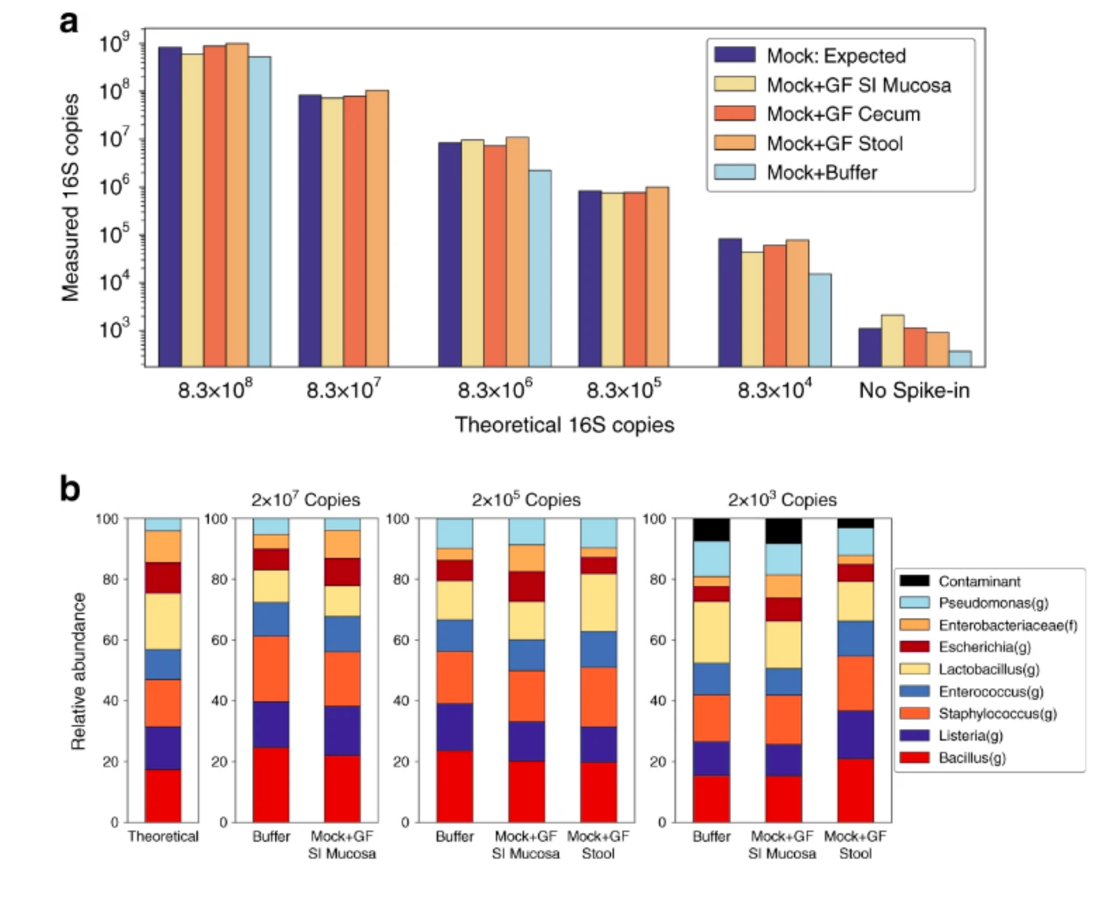
16S OR SHOTGUN SEQUENCING?

- Depends on the budget and the scientific question
- Do you want to assess changes to community structure or potential function?
- Are you interested in rare or low abundance species?
- What is the source of your specimens (host or environment contaminants)?
- Do you need species or strain level resolution?
- If you are interested in function shotgun sequencing will provide better resolution
  - For functional information you will need higher read depth
  - Less samples per run
- Intermediate approach is to do shallow shotgun sequencing (500k/sample)
  - Increased taxonomic resolution
  - Computationally more demanding
  - Shallow shotgun sequencing cannot be used for novel gene and genome assembly

LOW ABUNDANCE SAMPLES

- What can I do to improve analysis of low abundance samples?
  - Enrich target DNA by degrading free DNA prior to DNA extraction (Stinson et al, 2019, Lett Appl Microbiol., Li et al., 2017, Scientific Reports)
  - Include an extraction control to get background amplification
    - This can then be used to remove background with package "decontam" (Davis et al,2018, Microbiome)
  - Novel approach to prepare the samples to reduce host DNA and increase microbial reads
    - Microbial-enrichment method (MEM), has been validated on a wide range of sample types, including saliva, stool, intestinal scrapings, and intestinal mucosal biopsies (Natalie J. Wu-Woods, Nature Methods, 2023)
    - Recommended kits that have low contamination of bacterial DNA
      - Qiagen Dneasy Power soil Pro PN or MagAttract Power soil pro DNA kit,
      - Qiagen sells Microbial Free Water

LOW ABUNDANCE AND HOST DNA



Limit of detection

4.2×10^5 16S rRNA gene copies per gram for stool/cecum

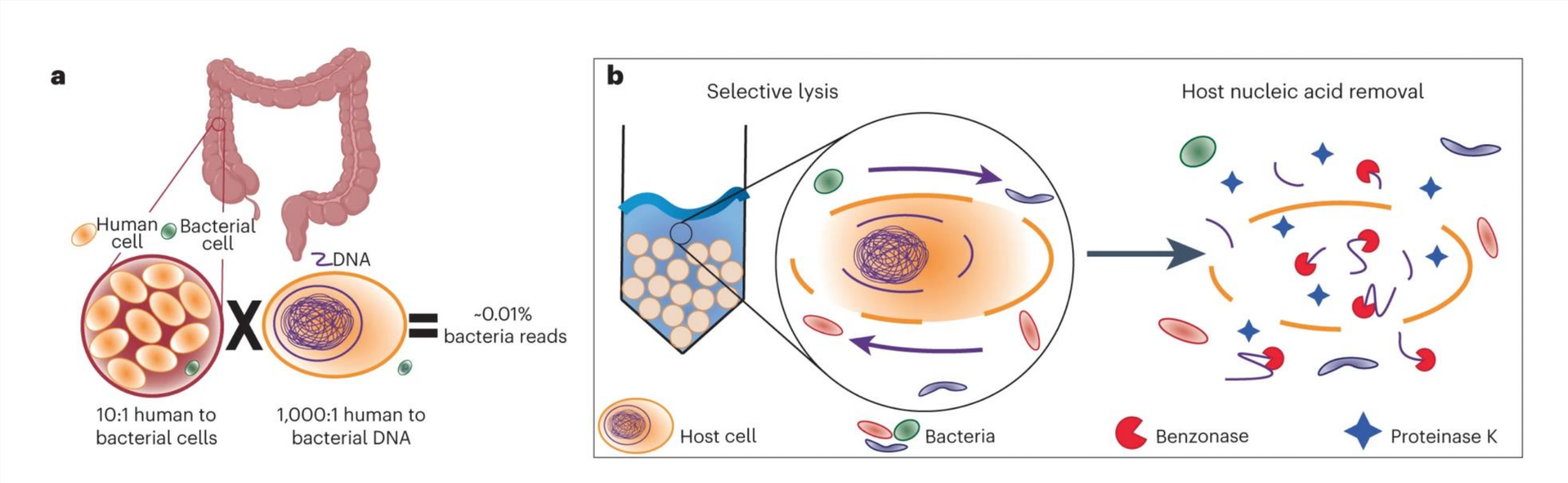
1×10^7 16S rRNA gene copies per gram for mucosa

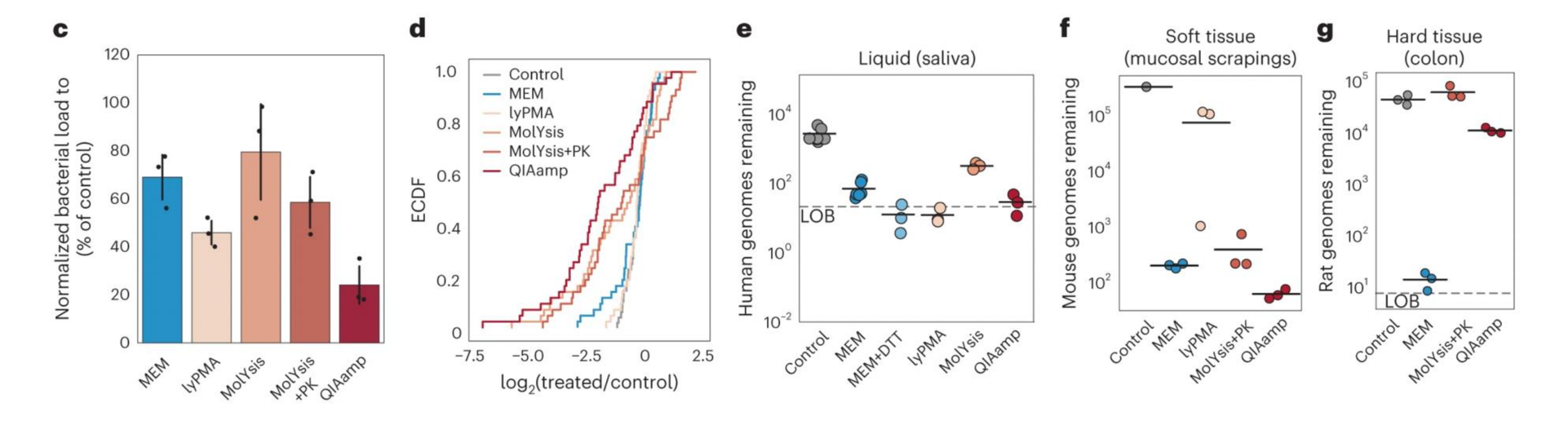
Usually extract
50-200mg stool/caecum
10mg of scraping

gDNA extraction columns cannot differentiate between host or bacterial DNA,
Columns can get saturated with host DNA which can lower the limit of detection

Barlow, 2020, Nat Comm

IMPROVED ISOLATION FROM SPECIMENS





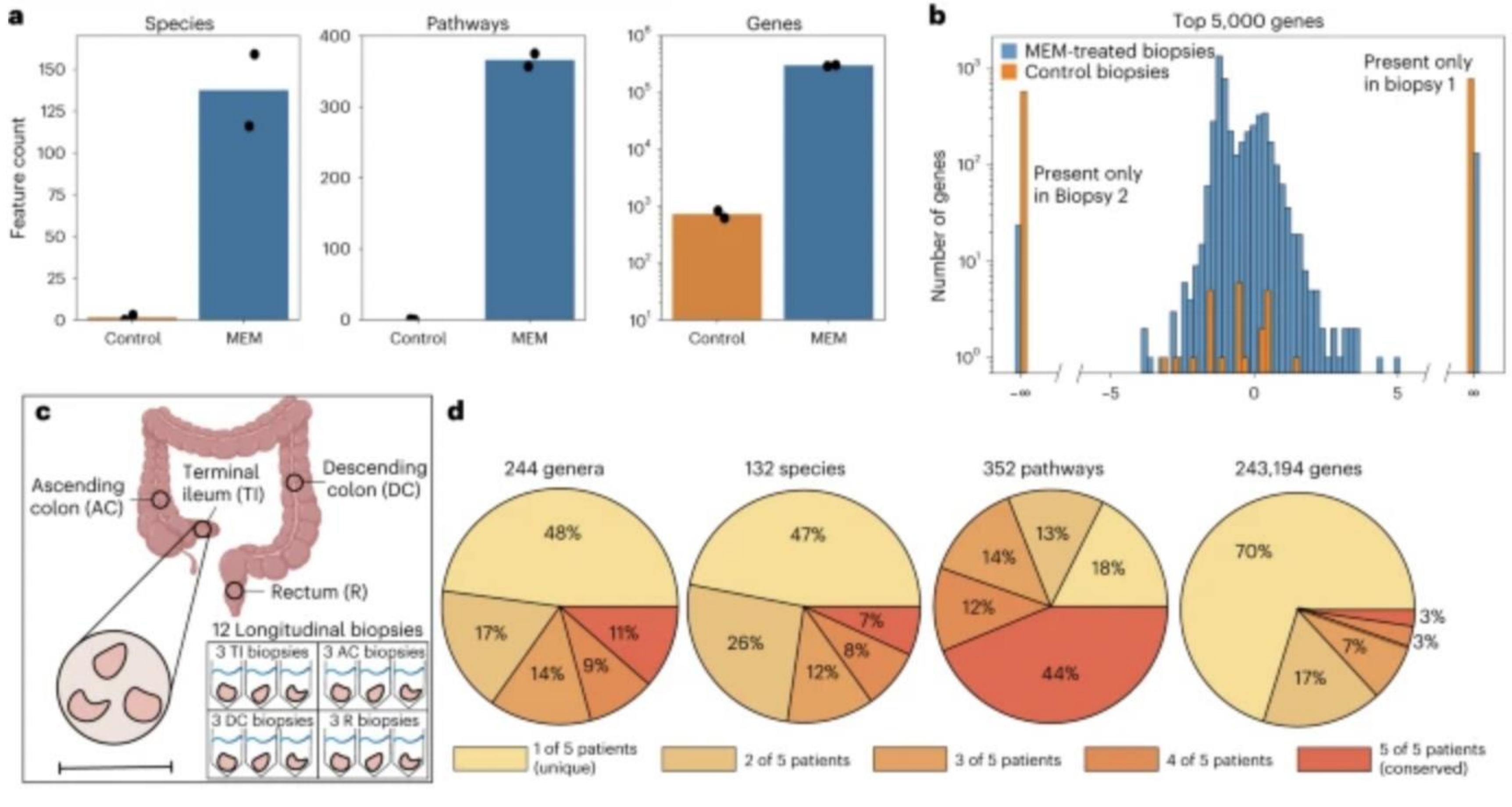
Natalie J. Wu-Woods, Nature Methods, 2023

INCREASED FUNCTIONAL RESOLUTION

Classification of taxa present at 0.005% compared to 10% without

10^4 16S copies per mg

Fig. 4: Shotgun sequencing of MEM-treated human intestinal biopsies.



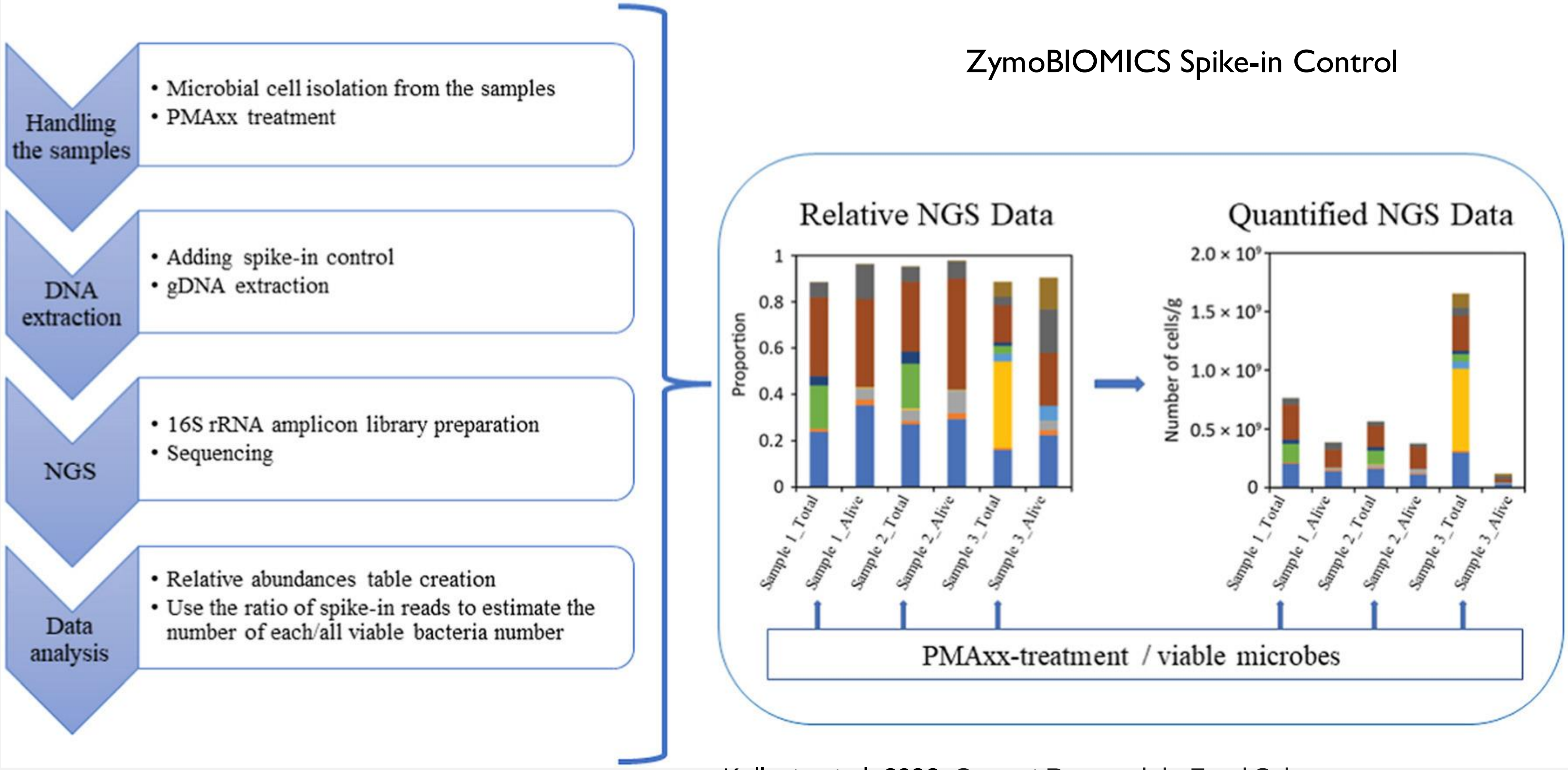
Natalie J. Wu-Woods, Nature Methods, 2023

WHAT 16S CAN AND CANNOT DO

- Can
  - Rapid snapshot of bacterial community composition
  - Scalable for many samples

- Cannot
  - Full taxonomic resolution (species or strain)
  - Function of taxa present (strain differences, etc)
  - Absolute quantification
  - Only association without follow-up validation in animal studies

SPIKE-IN FOR ABSOLUTE 16S QUANTIFICATION



Handling the samples
- Microbial cell isolation from the samples
- PMAxx treatment

DNA extraction
- Adding spike-in control
- gDNA extraction

NGS
- 16S rRNA amplicon library preparation
- Sequencing

Data analysis
- Relative abundances table creation
- Use the ratio of spike-in reads to estimate the number of each/all viable bacteria number

ZymoBIOMICS Spike-in Control

Kallastu et al, 2023, Current Research in Food Science

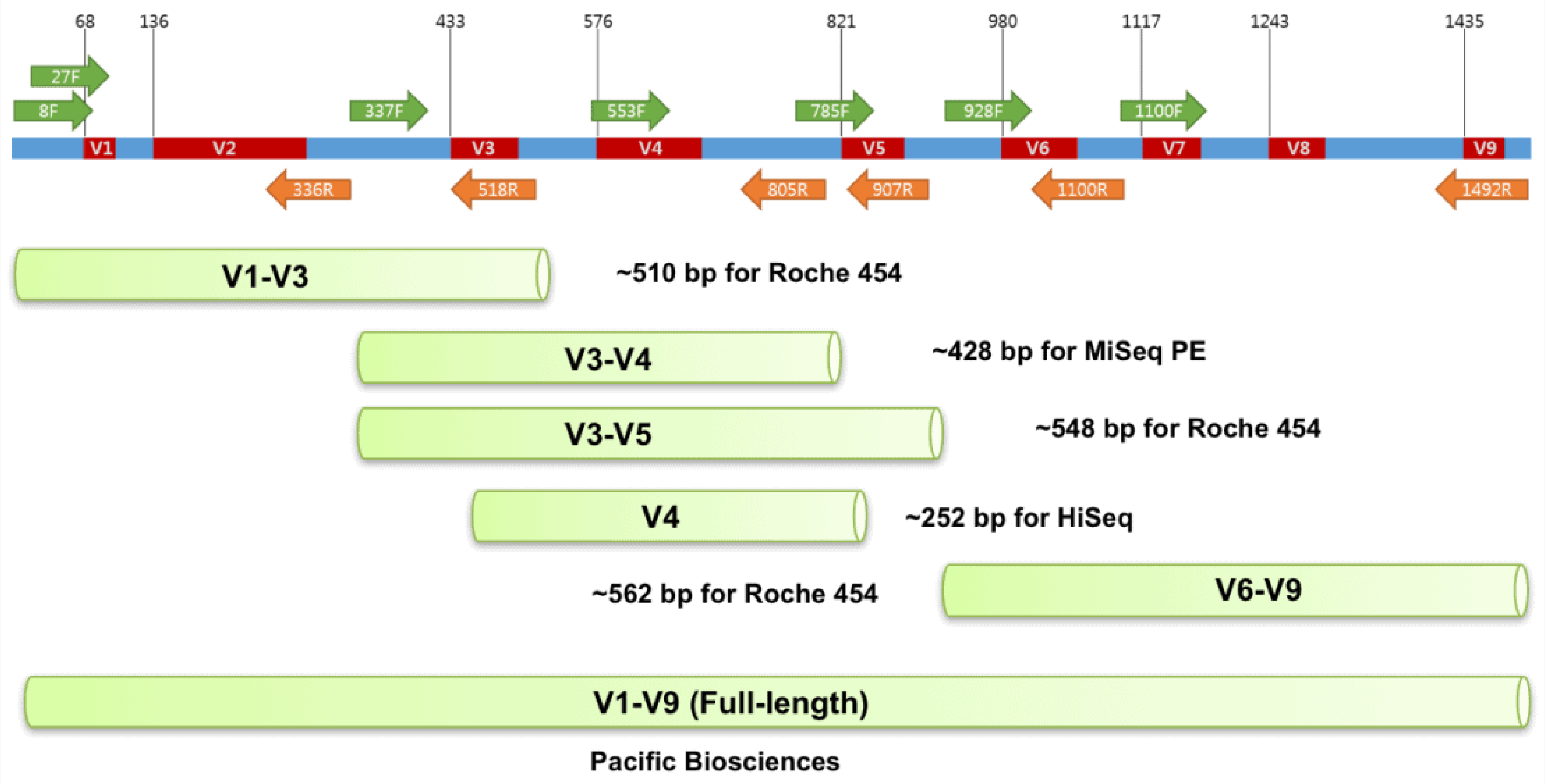
STRENGTHS OF 16S SEQUENCING

- Cheap and rapid data generation
- Can accommodate many samples on the same run
- Ideal to study conditions that change bacterial community patterns
- Multiple detailed tutorials and tools available to analyze the data
- Can compare datasets to previous publications (if you reanalyze the data with same parameters and if it's the same sequencing technology)
- Not as computationally intense as shotgun sequencing
- Low biomass ok (e.g. swabs, tissue, etc)

COMMON ISSUES WITH 16S

- Imperfect recall (not all sequences or taxa are detected)
  - Bias in extraction (universal)
  - Inefficiency of universal primers to hybridize all the templates
    - V4 sub-region has complete overlap of paired-end sequences
- 16S rRNA genes that do not differ in their amplified sequence cannot be resolved
  - Can only resolve on the family or genus level (Bukin et al, 2019, Scientific Data)
  - *Enterobacteriaceae* and the *Clostridiales* order are known to be poorly resolved using V4 amplicons
  - V2-V3 is better at resolving beyond the Genus level

WHICH VARIABLE REGION SHOULD I SEQUENCE?



EZ BioCloud

WHICH VARIABLE REGION SHOULD I SEQUENCE?

- Depends on the source of your sample
- Most cohort studies of the past have used V4 for profiling
- 515f/806r primer pair (V4) originally described by Caporaso et al, 2011 has been used in over 2000 publications

DO IT IN HOUSE OR SEND OUT?

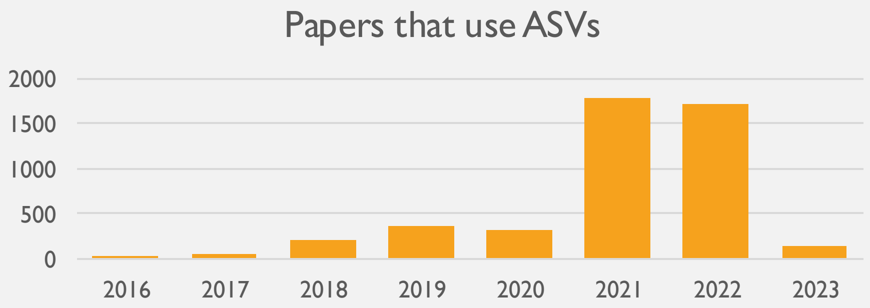
- Budget
- Time-frame
- Sample #
- How complex is the analysis?
- Services available at WashU:
  - Spike-in initiative
  - GTAC

WHY WE HAVE MOVED TO ASV OVER OTU

- Clustering sequences masks biological variation and produces artifacts
- ASVs are considered to be a more detailed view of the bacteria present compared to OTUs
- DADA2 is considered to provide the best compromise of specificity and sensitivity
- Always make raw sequences available when

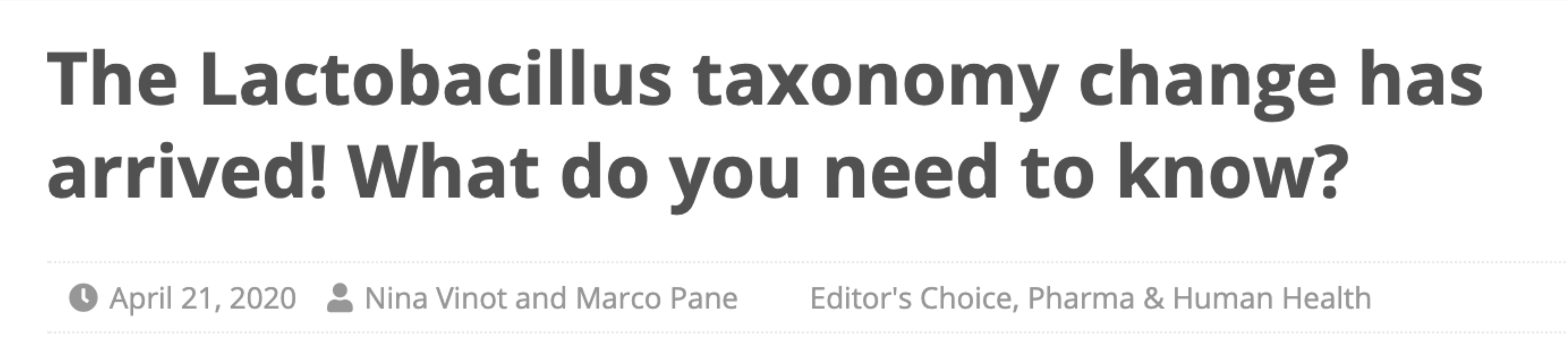
publishing and consider including table with

sequence of important bacteria



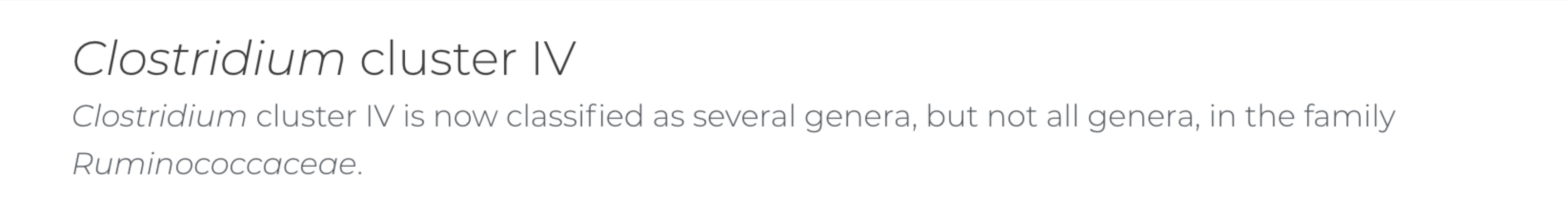
APPROACHES TO ANALYZE 16S

- Clustering of sequencing reads to obtain operational taxonomic units (OTUs) (USEARCH)
  - ≥97% is often used as a cutoff to cluster reads
- Generation of amplicon sequence variants (ASVs) using error-corrected reads (DADA2, Deblur)
- Direct taxonomic classification of raw reads (Kraken, Bracken)
  - Classifies sequencing reads into taxonomic bins
  - Bracken extension now also offers read count
  - Read that has identical matches to two species will be classified at the genus level
    - Multiple reads that can only be assigned at the genus level will be combined
  - Alignment-free algorithm makes it really fast (Lu et al, 2020, Mirobiome)



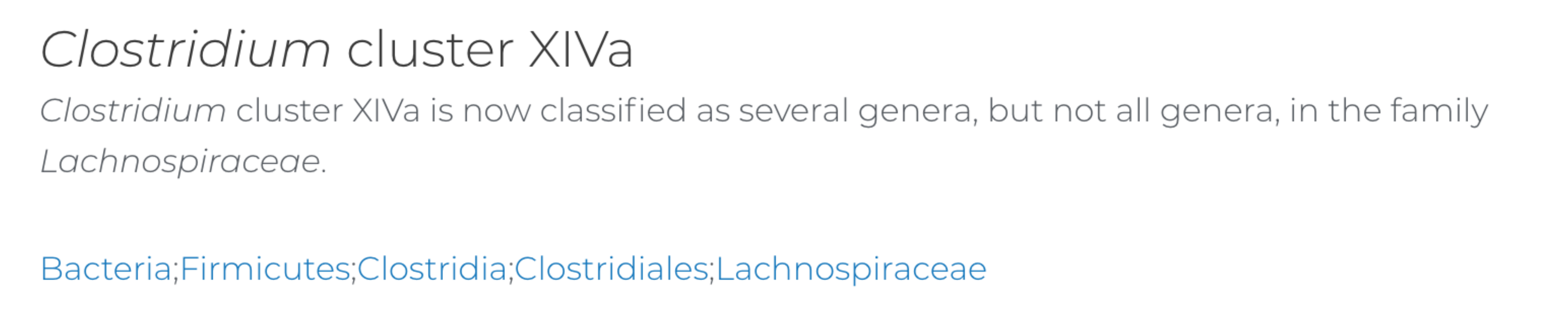
The Lactobacillus taxonomy change has arrived! What do you need to know?

April 21, 2020 Nina Vinot and Marco Pane Editor's Choice, Pharma & Human Health



Clostridium cluster IV

Clostridium cluster IV is now classified as several genera, but not all genera, in the family *Ruminococcaceae*.



Clostridium cluster XIVa

Clostridium cluster XIVa is now classified as several genera, but not all genera, in the family *Lachnospiraceae*.

Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae

16S DATA STRUCTURE

- More features (ASV) than samples (high dimensional)
- Many features with 0 (sparse)
- Number of reads between samples can vary
- Compositional (we only get a snapshot of the whole community)
    - Unless we use spike ins

PRE-REQUISITES FOR ANALYSIS

- Samples have been demultiplexed, i.e. split into individual per-sample fastq files
- Non-biological nucleotides have been removed, e.g. primers, adapters, linkers, etc
- If paired-end sequencing data, the forward and reverse fastq files contain reads in matched order
- Most sequencing services and cores will provide data in this format
- Have access to RIS or own machine with enough memory to run either RStudio or Qiime2 docker image

COMMON STEPS IN AMPLICON DENOISING PIPELINES

Filter Low Quality Reads

- dada2::filterAndTrim

Alignment and Dereplication

- dada2::nwalign and dada2::derepFastq

Error Correction

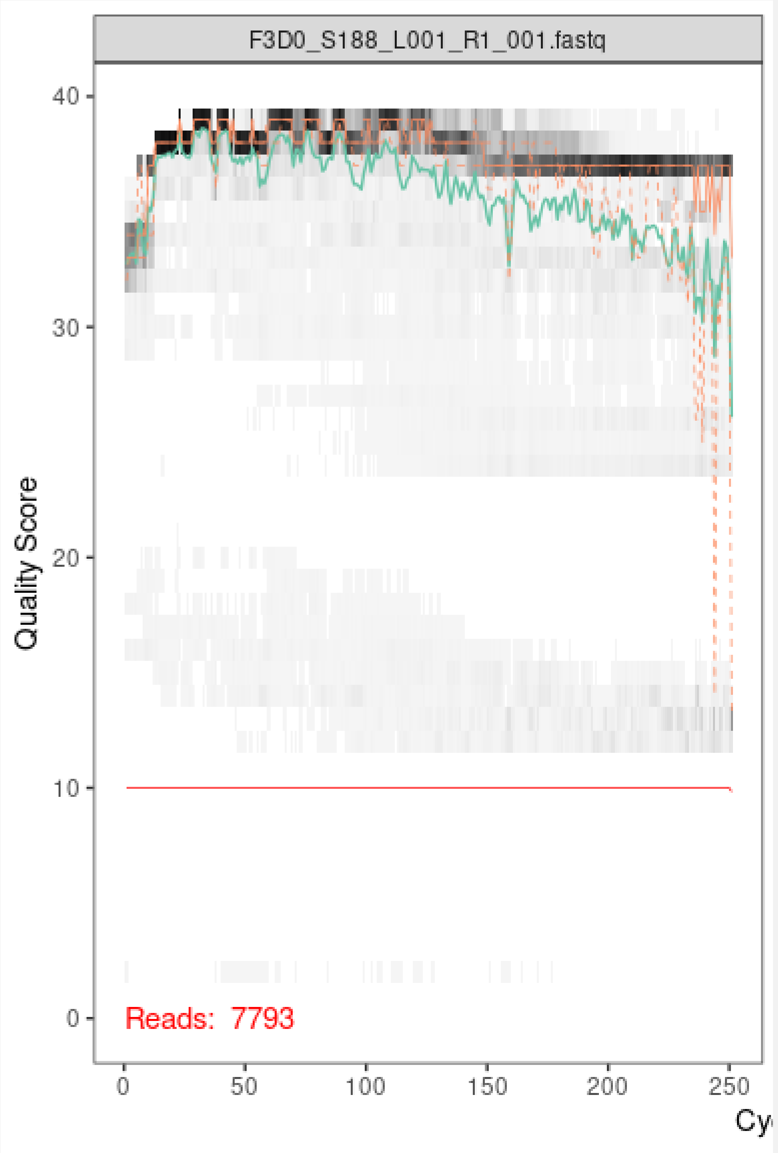
- dada2::learnErrors and dada2::dada

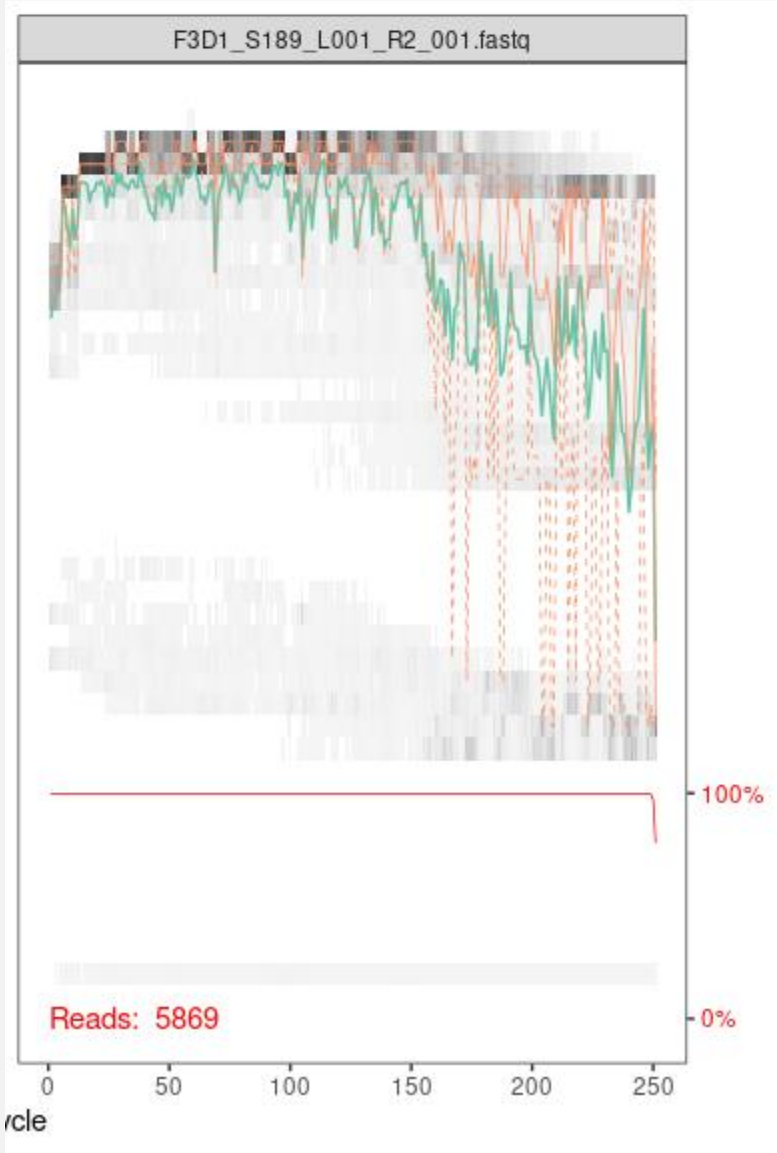
Chimera Removal

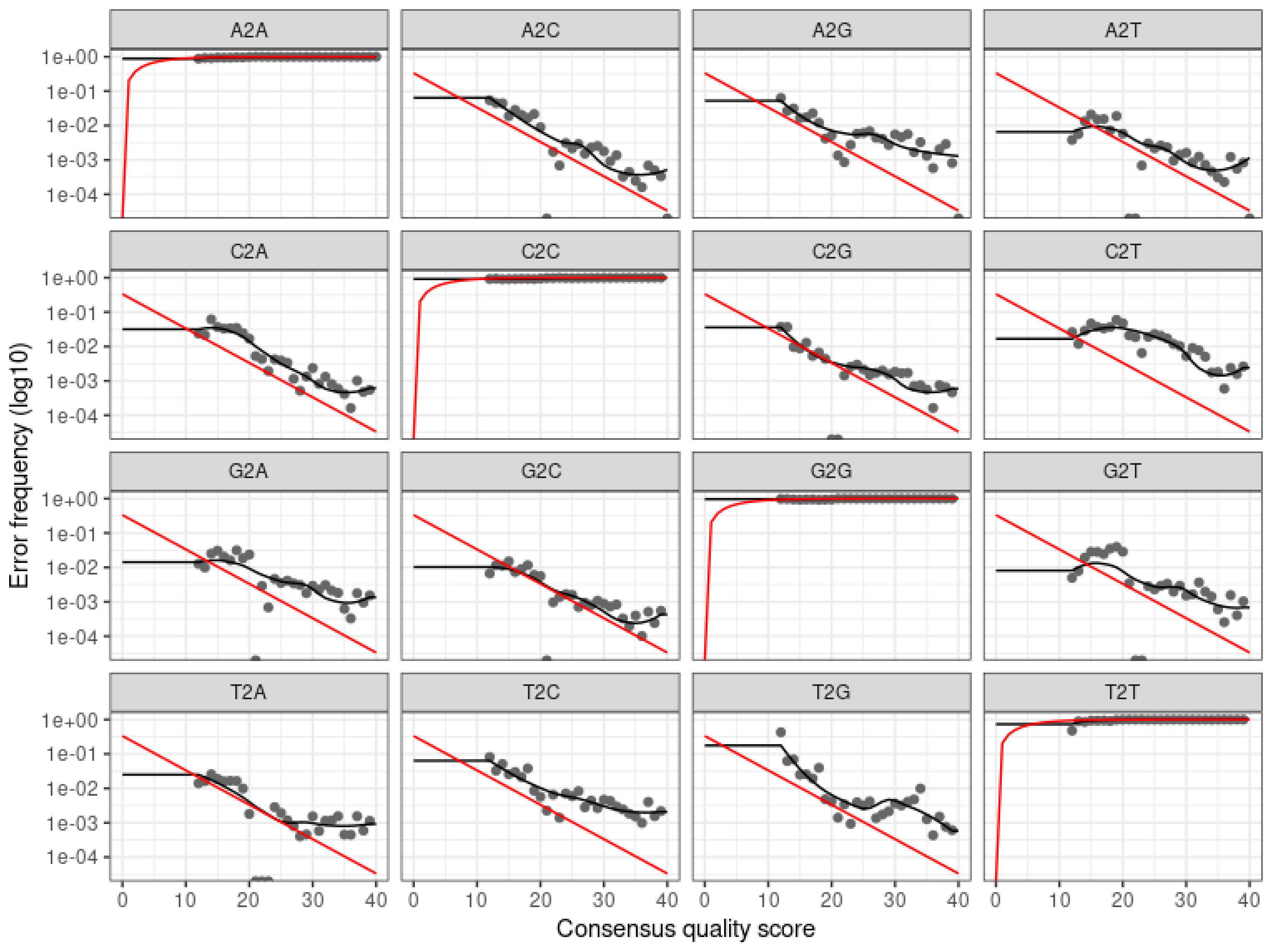
- dada2::removeBimeraDenovo

Taxonomic Classification

- dada2:assignTaxonomy and dada2::addSpecies


F3D0_S188_L001_R1_001.fastq
Quality Score
40
30
20
10
0
Reads: 7793
0
50
100
150
200
250
Cy


F3D1_S189_L001_R2_001.fastq
100%
0%
Reads: 5869
0
50
100
150
200
250
ycle


A2A
A2C
A2G
A2T
C2A
C2C
C2G
C2T
G2A
G2C
G2G
G2T
T2A
T2C
T2G
T2T
1e+00
1e-01
1e-02
1e-03
1e-04
0
10
20
30
40
Error frequency (log10)
Consensus quality score

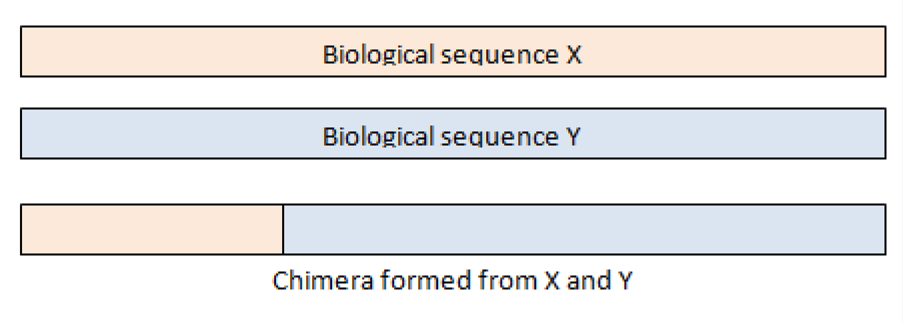
OVERVIEW DADA2 OUTPUT

	input	filtered	denoisedF	denoisedR	merged
F3D0	7793	7113	6976	6979	6540
F3D1	5869	5299	5227	5239	5028
F3D141	5958	5463	5331	5357	4986
F3D142	3183	2914	2799	2830	2595
F3D143	3178	2941	2822	2868	2553
F3D144	4827	4312	4151	4228	3646

CHIMERIC SEQUENCES

- Chimeras are sequences formed from two (or more) biological sequences joined together

- Typically form during PCR when:
  - Extension of an amplicon is aborted
  - Serves as a primer in the next PCR cycle
  - Aborted product anneals to the wrong template
  - Makes a new sequence from two different template

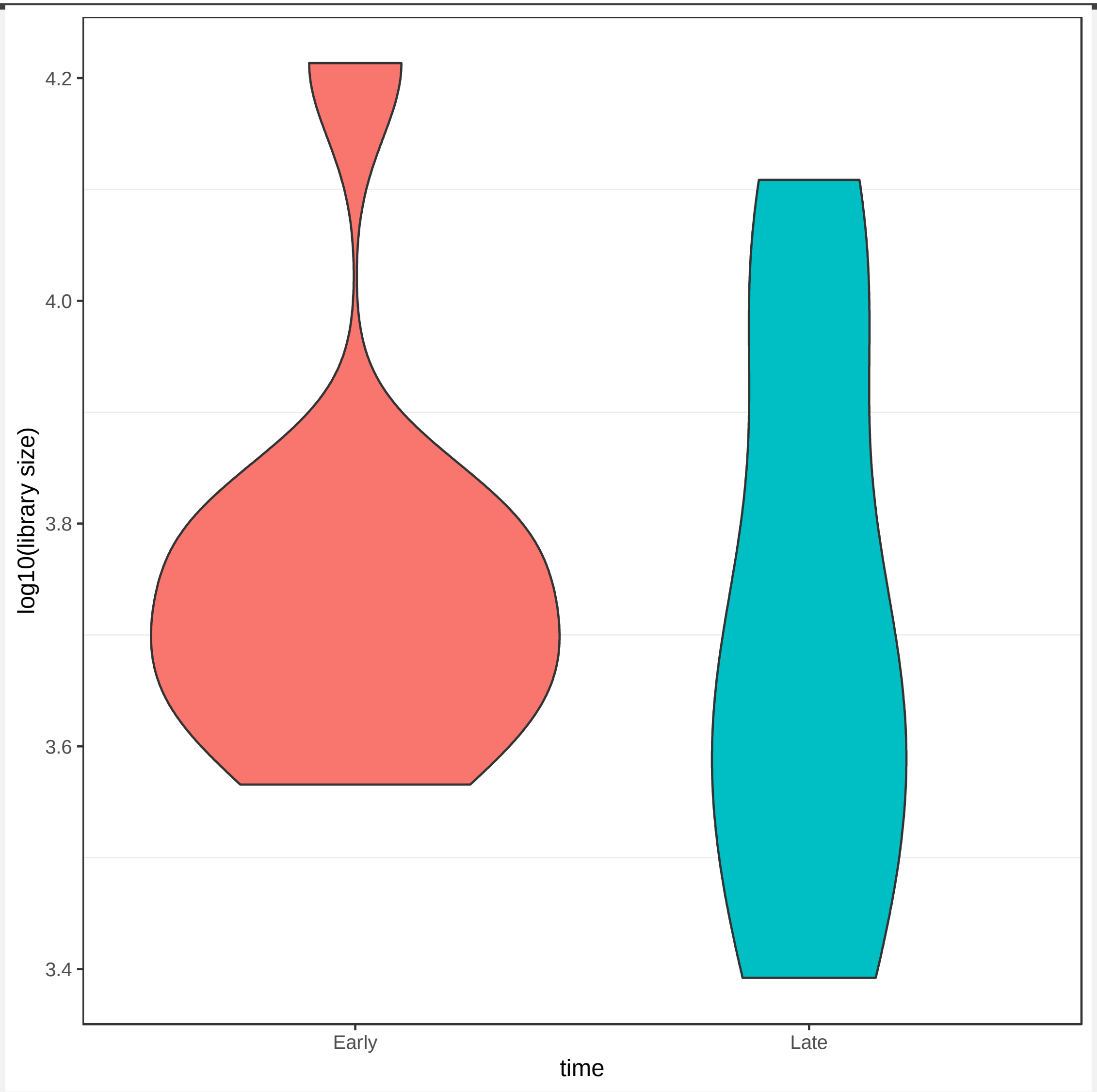


- Typically, small fraction of reads are chimeric
  - However, a large fraction of ASVs/OTUs are often chimeric

TAXONOMY

- Taxonomic assignment considerations
- ASVs will only get species level annotation if there is an exact 100% match to the database
- There are multiple species that have exact same variable regions (e.g. Escherichia coli and Shigella)
  - Change default parameters to allow multiple species names if you want to capture all matching species names (see option in Rmd file)
- Which database to compare to
  - Silva (most comprehensive database has a lot of environmental sequences)
  - RDP (mainly based on human microbiome data)
  - Greengenes (highest taxonomic accuracy but outdated naming 2012-2013)
- What classifier to use
  - *RDP-classifier* can predict closest relative of unknown taxa up to genus level (80% accuracy in genus-level assignments)
- Before starting a new project or before publication I recommend checking if there is an updated in the reference database to have the most up-to-date nomenclature

QUALITY CONTROL

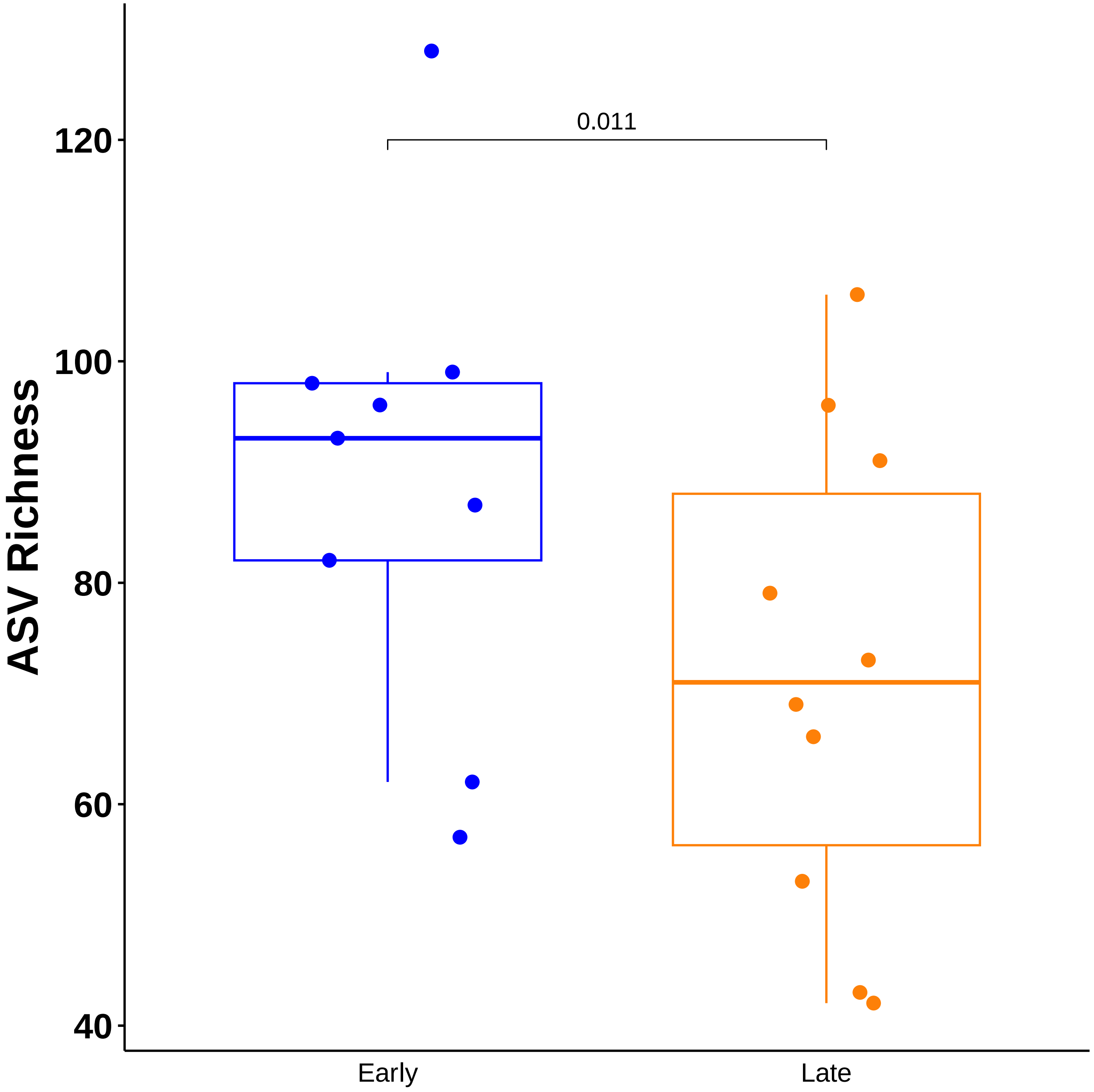


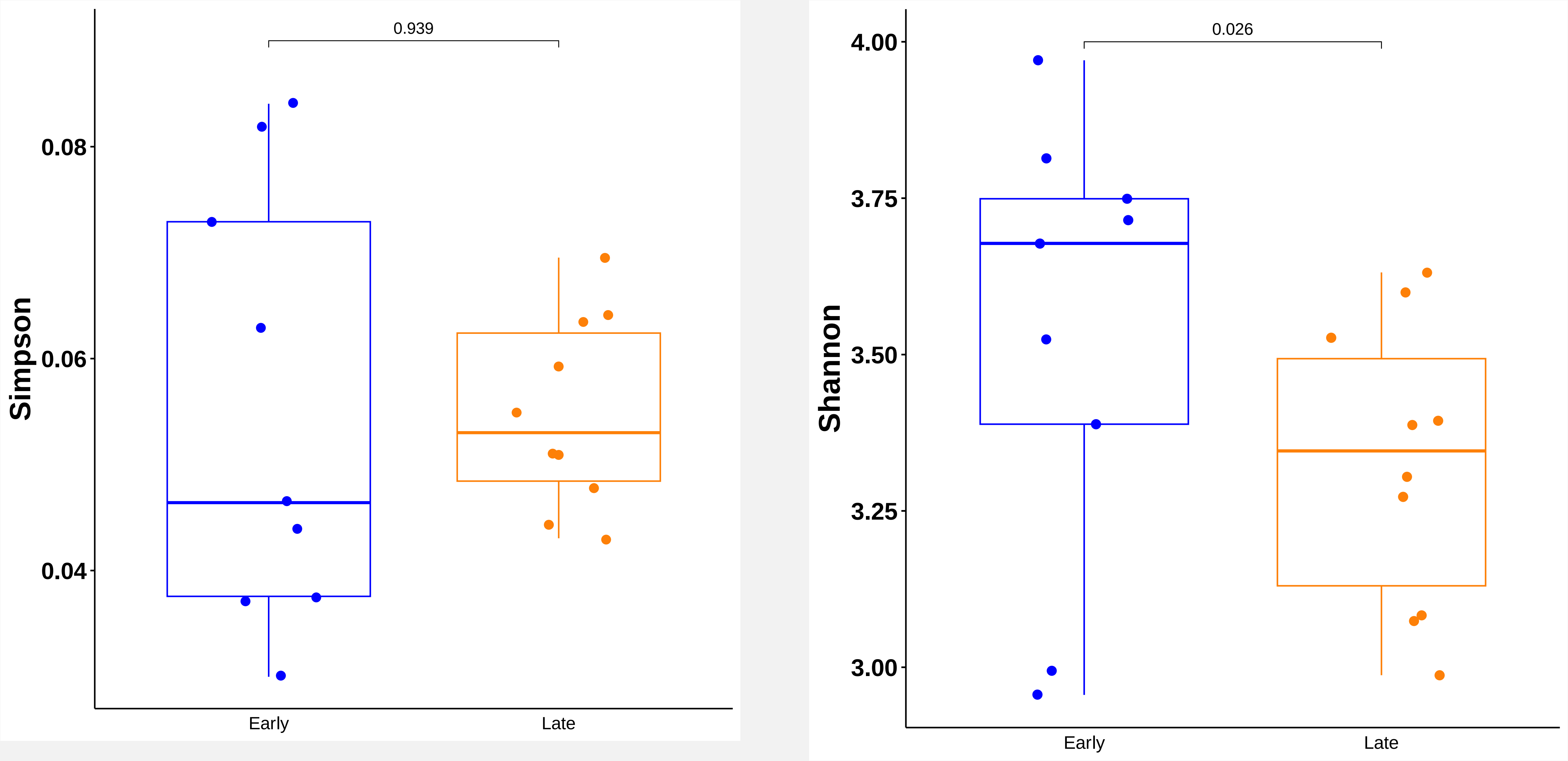
SHOULD I USE RAREFACTION?

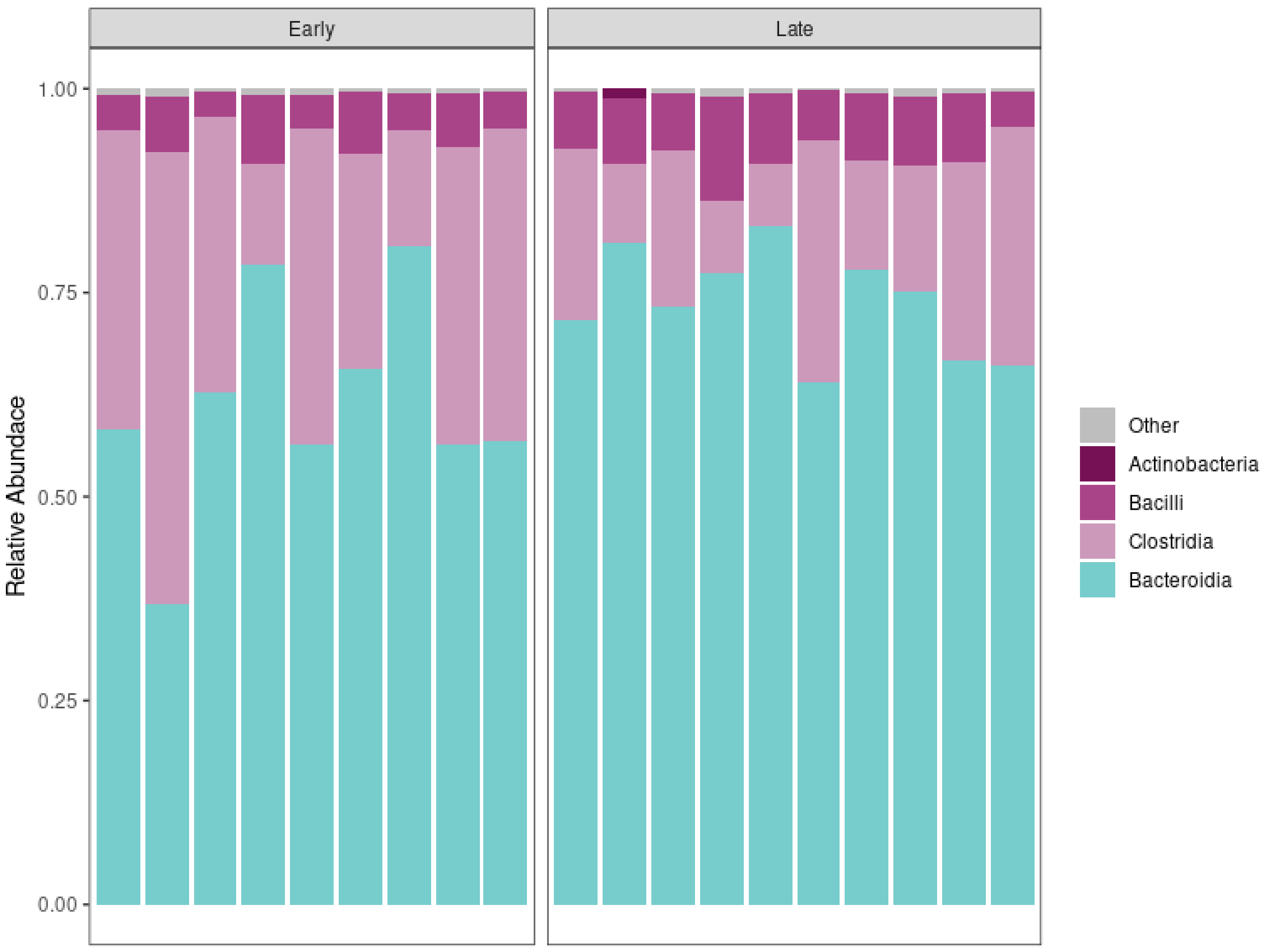
- Issue: Random subsampling that results in loss of data and generation of artificial variation
- There are now new approaches for diversity metrics that do not require rarefaction
- For differential abundance do not use rarefied data as most approaches have their own normalization techniques that take library size into account

ALPHA DIVERSITY MEASURES

- Focus on within sample diversity
  - Measure of richness and evenness
- New packages developed by Amy Willis allow to calculate richness estimates so that we don’t need to worry about differences in sequencing depth between samples
- You can calculate Shannon, Simpson, and richness with her packages and fit models with fixed variables, you can even add random effects (e.g. cage) to account for changes in richness
- More here https://adw96.github.io/breakaway/articles/breakaway.html


0.011
ASV Richness
120
100
80
60
40
Early
Late


0.939
Simpson
0.08
0.06
0.04
Early
Late
0.026
Shannon
4.00
3.75
3.50
3.25
3.00
Early
Late


Early
Late
Relative Abundace
1.00
0.75
0.50
0.25
0.00
Other
Actinobacteria
Bacilli
Clostridia
Bacteroidia

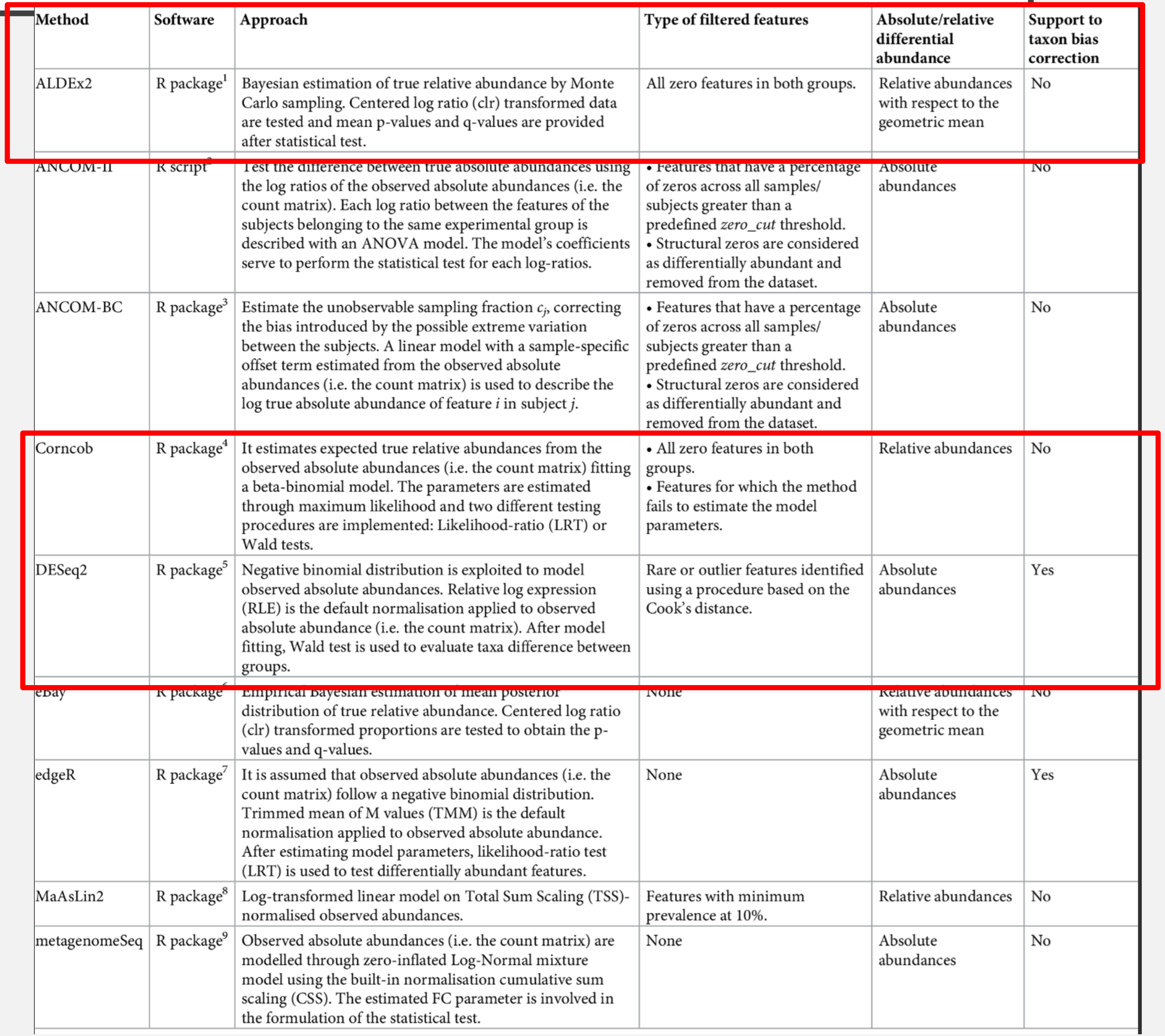
ZERO-INFLATION

- Zero inflated models should be considered for differential abundance:
  - If you compare samples from very different sourcing sites
  - If you do a treatment or intervention that would cause the depletion of given taxa
- This is important if you think that one of the groups will lack taxa not because you didn't sequence deeply enough, but because they are truly gone
- More information on https://www.nicholas-ollberding.com/post/observation-weights-for-differential-abundance-of-zero-inflated-microbiome-data-with-deseq2/
- Calgaro et al, 2020, Genome Biology

DIFFERENTIAL ABUNDANCE

- Sample size is the main driver of false discovery issues
  - You need enough samples to overcome the noise, especially in human data (>50 samples per group)
- It's best to try different approaches and see which taxa consistently come up with all of them

DIFFERENTIAL ABUNDANCE



Method	Software	Approach	Type of filtered features	Absolute/relative differential abundance	Support to taxon bias correction
ALDEx2	R package[1]	Bayesian estimation of true relative abundance by Monte Carlo sampling. Centered log ratio (clr) transformed data are tested and mean p-values and q-values are provided after statistical test.	All zero features in both groups.	Relative abundances with respect to the geometric mean	No
ANCOM-II	R script[2]	Test the difference between true absolute abundances using the log ratios of the observed absolute abundances (i.e. the count matrix). Each log ratio between the features of the subjects belonging to the same experimental group is described with an ANOVA model. The model's coefficients serve to perform the statistical test for each log-ratios.	• Features that have a percentage of zeros across all samples/ subjects greater than a predefined *zero_cut* threshold. • Structural zeros are considered as differentially abundant and removed from the dataset.	Absolute abundances	No
ANCOM-BC	R package[3]	Estimate the unobservable sampling fraction c_j, correcting the bias introduced by the possible extreme variation between the subjects. A linear model with a sample-specific offset term estimated from the observed absolute abundances (i.e. the count matrix) is used to describe the log true absolute abundance of feature *i* in subject *j*.	• Features that have a percentage of zeros across all samples/ subjects greater than a predefined *zero_cut* threshold. • Structural zeros are considered as differentially abundant and removed from the dataset.	Absolute abundances	No
Corncob	R package[4]	It estimates expected true relative abundances from the observed absolute abundances (i.e. the count matrix) fitting a beta-binomial model. The parameters are estimated through maximum likelihood and two different testing procedures are implemented: Likelihood-ratio (LRT) or Wald tests.	• All zero features in both groups. • Features for which the method fails to estimate the model parameters.	Relative abundances	No
DESeq2	R package[5]	Negative binomial distribution is exploited to model observed absolute abundances. Relative log expression (RLE) is the default normalisation applied to observed absolute abundance (i.e. the count matrix). After model fitting, Wald test is used to evaluate taxa difference between groups.	Rare or outlier features identified using a procedure based on the Cook's distance.	Absolute abundances	Yes
eBay	R package[6]	Empirical Bayesian estimation of mean posterior distribution of true relative abundance. Centered log ratio (clr) transformed proportions are tested to obtain the p-values and q-values.	None	Relative abundances with respect to the geometric mean	No
edgeR	R package[7]	It is assumed that observed absolute abundances (i.e. the count matrix) follow a negative binomial distribution. Trimmed mean of M values (TMM) is the default normalisation applied to observed absolute abundance. After estimating model parameters, likelihood-ratio test (LRT) is used to test differentially abundant features.	None	Absolute abundances	Yes
MaAsLin2	R package[8]	Log-transformed linear model on Total Sum Scaling (TSS)-normalised observed abundances.	Features with minimum prevalence at 10%.	Relative abundances	No
metagenomeSeq	R package[9]	Observed absolute abundances (i.e. the count matrix) are modelled through zero-inflated Log-Normal mixture model using the built-in normalisation cumulative sum scaling (CSS). The estimated FC parameter is involved in the formulation of the statistical test.	None	Absolute abundances	No

Cappellato M, Baruzzo G, Di Camillo B (2022) Investigating differential abundance methods in microbiome data: A benchmark study. PLOS Computational Biology 18(9): e1010467. https://doi.org/10.1371/journal.pcbi.1010467
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010467

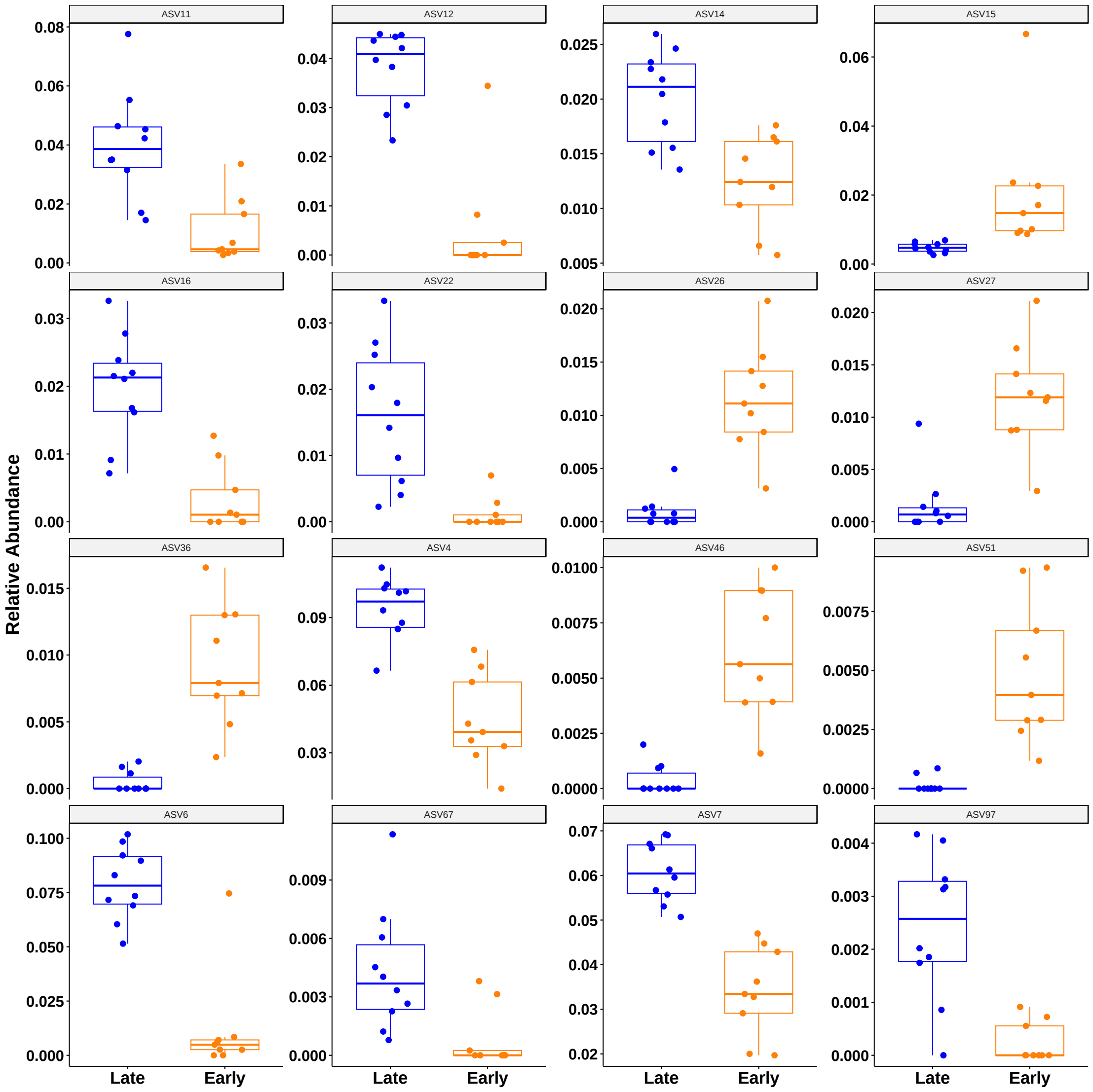
CORNCOB

- Corncob is a beta-binomial model
- allows for the overdispersion in the taxon's counts to be associated with covariates of interest
- Differential relative abundance, but also for differential variability

Bryan D. Martin

Bryan D. Martin,2020, Annals of Applied Statistics

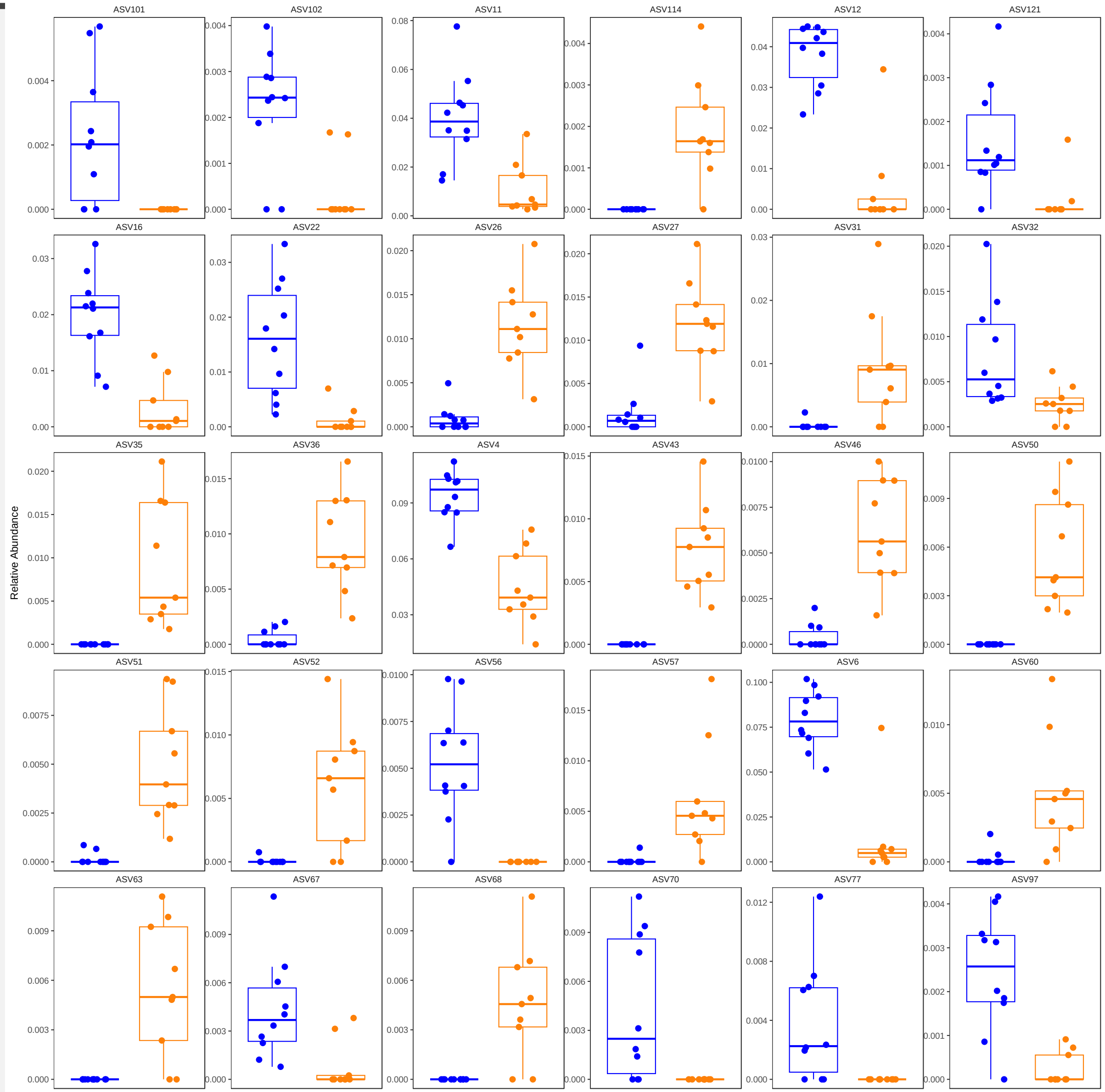
CORNCOB


Relative Abundance
ASV11
0.08
0.06
0.04
0.02
0.00
ASV12
0.04
0.03
0.02
0.01
0.00
ASV14
0.025
0.020
0.015
0.010
0.005
ASV15
0.06
0.04
0.02
0.00
ASV16
0.03
0.02
0.01
0.00
ASV22
0.03
0.02
0.01
0.00
ASV26
0.020
0.015
0.010
0.005
0.000
ASV27
0.020
0.015
0.010
0.005
0.000
ASV36
0.015
0.010
0.005
0.000
ASV4
0.09
0.06
0.03
ASV46
0.0100
0.0075
0.0050
0.0025
0.0000
ASV51
0.0075
0.0050
0.0025
0.0000
ASV6
0.100
0.075
0.050
0.025
0.000
ASV67
0.009
0.006
0.003
0.000
ASV7
0.07
0.06
0.05
0.04
0.03
0.02
ASV97
0.004
0.003
0.002
0.001
0.000
Late
Early
Late
Early
Late
Early
Late
Early

ALDEX2

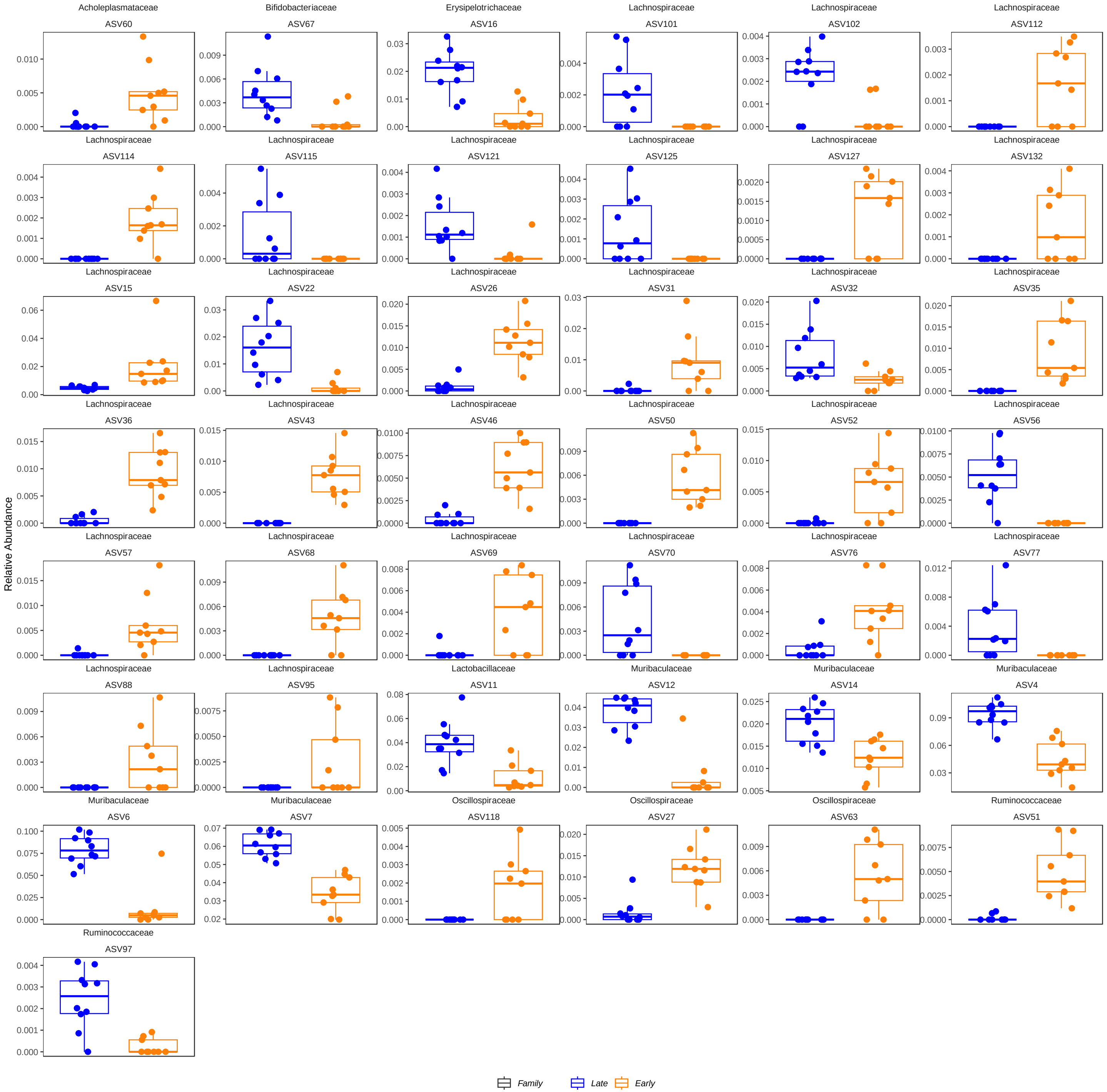
- Initially developed for RNA-Seq
- Produce the most consistent results across studies
- Assumes the data is compositional which sometimes might not work if your treatment completely depletes a population (e.g. antibiotics)
- Normalizes to central log ratios

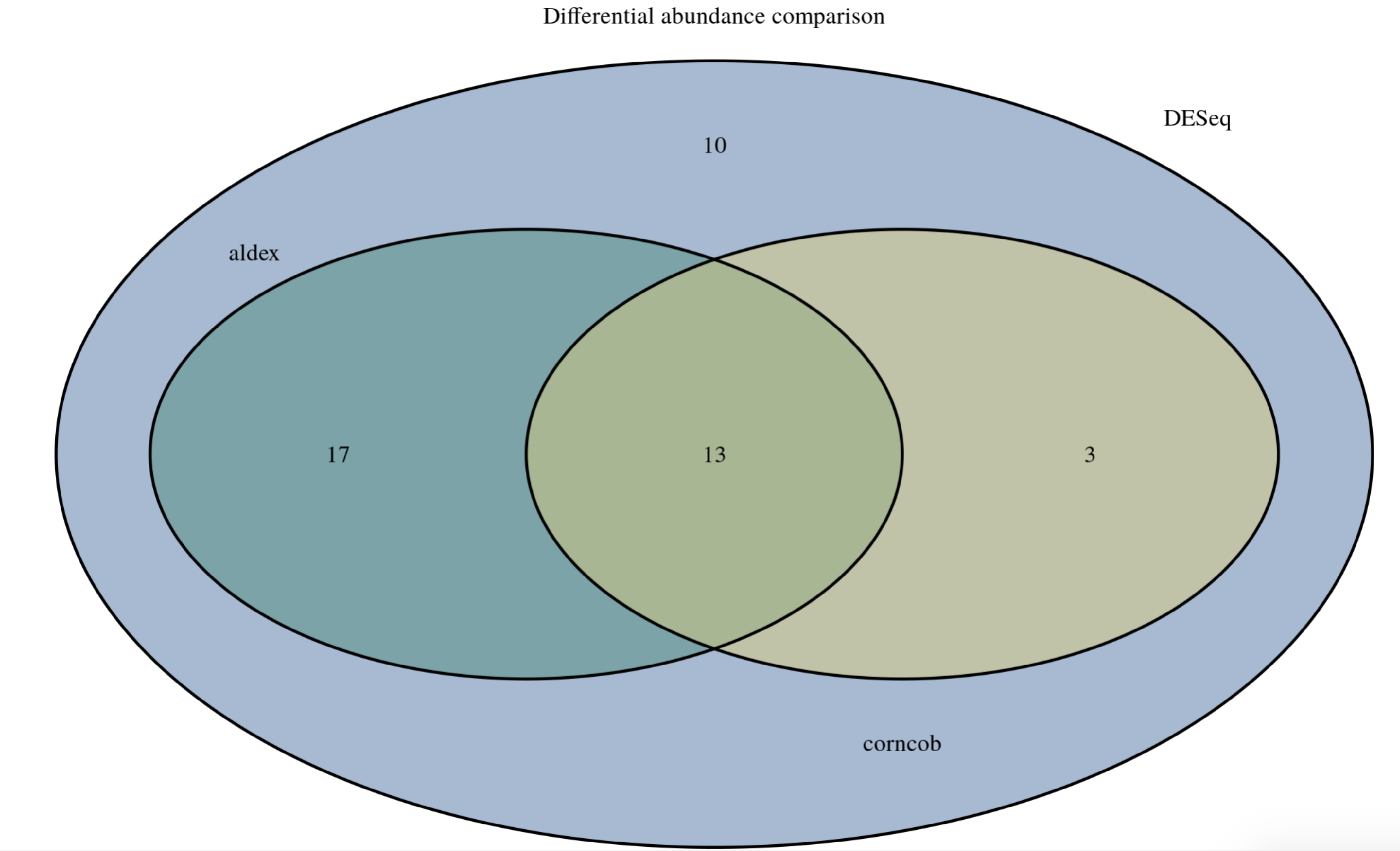
ALDEX2


Relative Abundance
ASV101
0.000
0.002
0.004
ASV102
0.000
0.001
0.002
0.003
0.004
ASV11
0.00
0.02
0.04
0.06
0.08
ASV114
0.000
0.001
0.002
0.003
0.004
ASV12
0.00
0.01
0.02
0.03
0.04
ASV121
0.000
0.001
0.002
0.003
0.004
ASV16
0.00
0.01
0.02
0.03
ASV22
0.00
0.01
0.02
0.03
ASV26
0.000
0.005
0.010
0.015
0.020
ASV27
0.000
0.005
0.010
0.015
0.020
ASV31
0.00
0.01
0.02
0.03
ASV32
0.000
0.005
0.010
0.015
0.020
ASV35
0.000
0.005
0.010
0.015
0.020
ASV36
0.000
0.005
0.010
0.015
ASV4
0.03
0.06
0.09
ASV43
0.000
0.005
0.010
0.015
ASV46
0.0000
0.0025
0.0050
0.0075
0.0100
ASV50
0.000
0.003
0.006
0.009
ASV51
0.0000
0.0025
0.0050
0.0075
ASV52
0.000
0.005
0.010
0.015
ASV56
0.0000
0.0025
0.0050
0.0075
0.0100
ASV57
0.000
0.005
0.010
0.015
ASV6
0.000
0.025
0.050
0.075
0.100
ASV60
0.000
0.005
0.010
ASV63
0.000
0.003
0.006
0.009
ASV67
0.000
0.003
0.006
0.009
ASV68
0.000
0.003
0.006
0.009
ASV70
0.000
0.003
0.006
0.009
ASV77
0.000
0.004
0.008
0.012
ASV97
0.000
0.001
0.002
0.003
0.004

DESEQ2

- If each ASV has the same chance of being selected for sequencing, and selected independently, then the number of read counts for a given ASV should follow a Poisson distribution
  - In most experiments the assumption of independence fails so we have to use a different type of model
  - Taxa and samples can be correlated which can lead to overdispersion
- In DESeq library size for each sample is accounted for with the size factor
  - Used to transform the counts to a common scale
- DESeq2 assumes that ASVs of similar average count have a similar dispersion and fits the ASV-specific dispersion towards the average dispersion
- ASVs are transformed to logarithmic fold change
- Wald tests for differential expression with multiple testing correction
- You can add additional factors to correct for interactions (batch effect, etc)

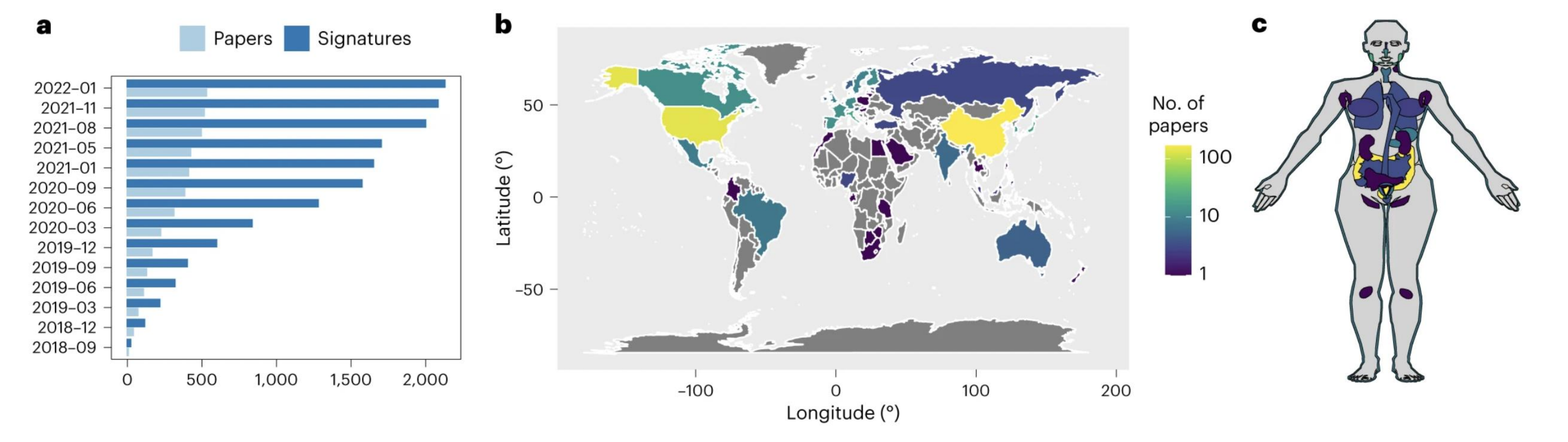

Acholeplasmataceae
ASV60
0.010
0.005
0.000
Bifidobacteriaceae
ASV67
0.009
0.006
0.003
0.000
Erysipelotrichaceae
ASV16
0.03
0.02
0.01
0.00
Lachnospiraceae
ASV101
0.004
0.002
0.000
Lachnospiraceae
ASV102
0.004
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV112
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV114
0.004
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV115
0.004
0.002
0.000
Lachnospiraceae
ASV121
0.004
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV125
0.004
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV127
0.0020
0.0015
0.0010
0.0005
0.0000
Lachnospiraceae
ASV132
0.004
0.003
0.002
0.001
0.000
Lachnospiraceae
ASV15
0.06
0.04
0.02
0.00
Lachnospiraceae
ASV22
0.03
0.02
0.01
0.00
Lachnospiraceae
ASV26
0.020
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV31
0.03
0.02
0.01
0.00
Lachnospiraceae
ASV32
0.020
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV35
0.020
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV36
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV43
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV46
0.0100
0.0075
0.0050
0.0025
0.0000
Lachnospiraceae
ASV50
0.009
0.006
0.003
0.000
Lachnospiraceae
ASV52
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV56
0.0100
0.0075
0.0050
0.0025
0.0000
Relative Abundance
Lachnospiraceae
ASV57
0.015
0.010
0.005
0.000
Lachnospiraceae
ASV68
0.009
0.006
0.003
0.000
Lachnospiraceae
ASV69
0.008
0.006
0.004
0.002
0.000
Lachnospiraceae
ASV70
0.009
0.006
0.003
0.000
Lachnospiraceae
ASV76
0.008
0.006
0.004
0.002
0.000
Lachnospiraceae
ASV77
0.012
0.008
0.004
0.000
Lachnospiraceae
ASV88
0.009
0.006
0.003
0.000
Lachnospiraceae
ASV95
0.0075
0.0050
0.0025
0.0000
Lactobacillaceae
ASV11
0.08
0.06
0.04
0.02
0.00
Muribaculaceae
ASV12
0.04
0.03
0.02
0.01
0.00
Muribaculaceae
ASV14
0.025
0.020
0.015
0.010
0.005
Muribaculaceae
ASV4
0.09
0.06
0.03
Muribaculaceae
ASV6
0.100
0.075
0.050
0.025
0.000
Muribaculaceae
ASV7
0.07
0.06
0.05
0.04
0.03
0.02
Oscillospiraceae
ASV118
0.005
0.004
0.003
0.002
0.001
0.000
Oscillospiraceae
ASV27
0.020
0.015
0.010
0.005
0.000
Oscillospiraceae
ASV63
0.009
0.006
0.003
0.000
Ruminococcaceae
ASV51
0.0075
0.0050
0.0025
0.0000
Ruminococcaceae
ASV97
0.004
0.003
0.002
0.001
0.000
Family
Late
Early



Differential abundance comparison

DEVELOPMENTS

- Moving away from 16S to shotgun or long read technology

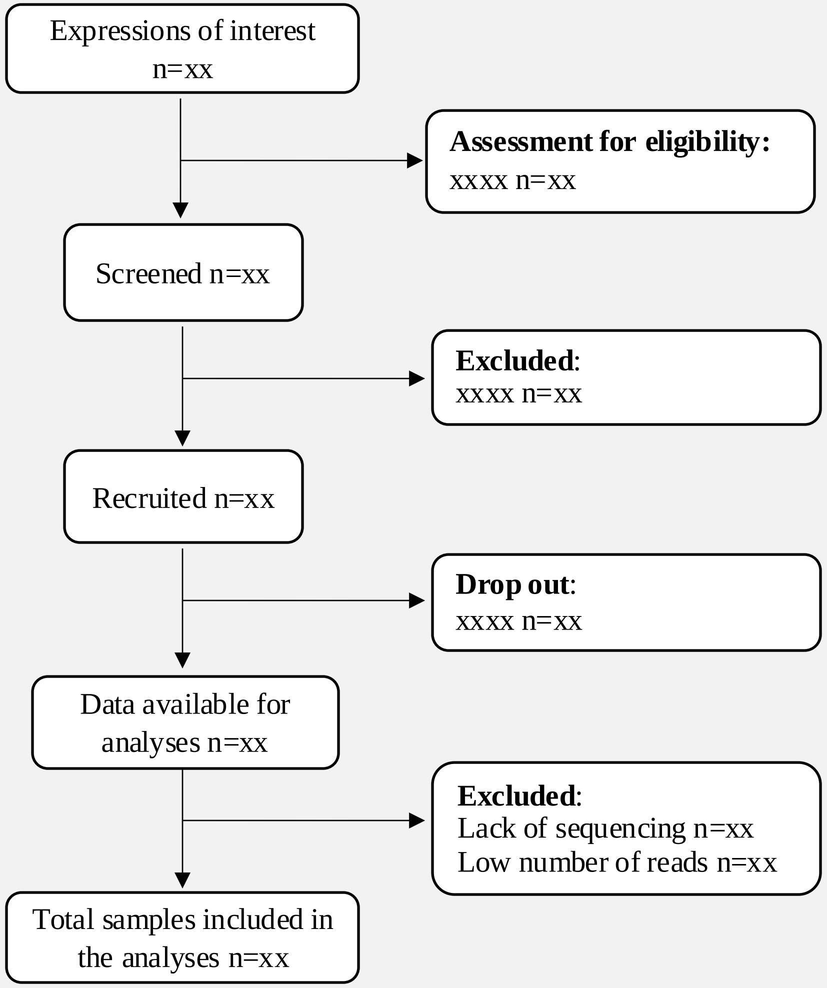


Geistlinger et al, Nature Biotech, 2023

https://bugsigdb.org/Main_Page

STORMS analytic sample size flowcharts (item 3.6)

Observational



Expressions of interest n=xx

Assessment for eligibility: xxxx n=xx

Screened n=xx

Excluded: xxxx n=xx

Recruited n=xx

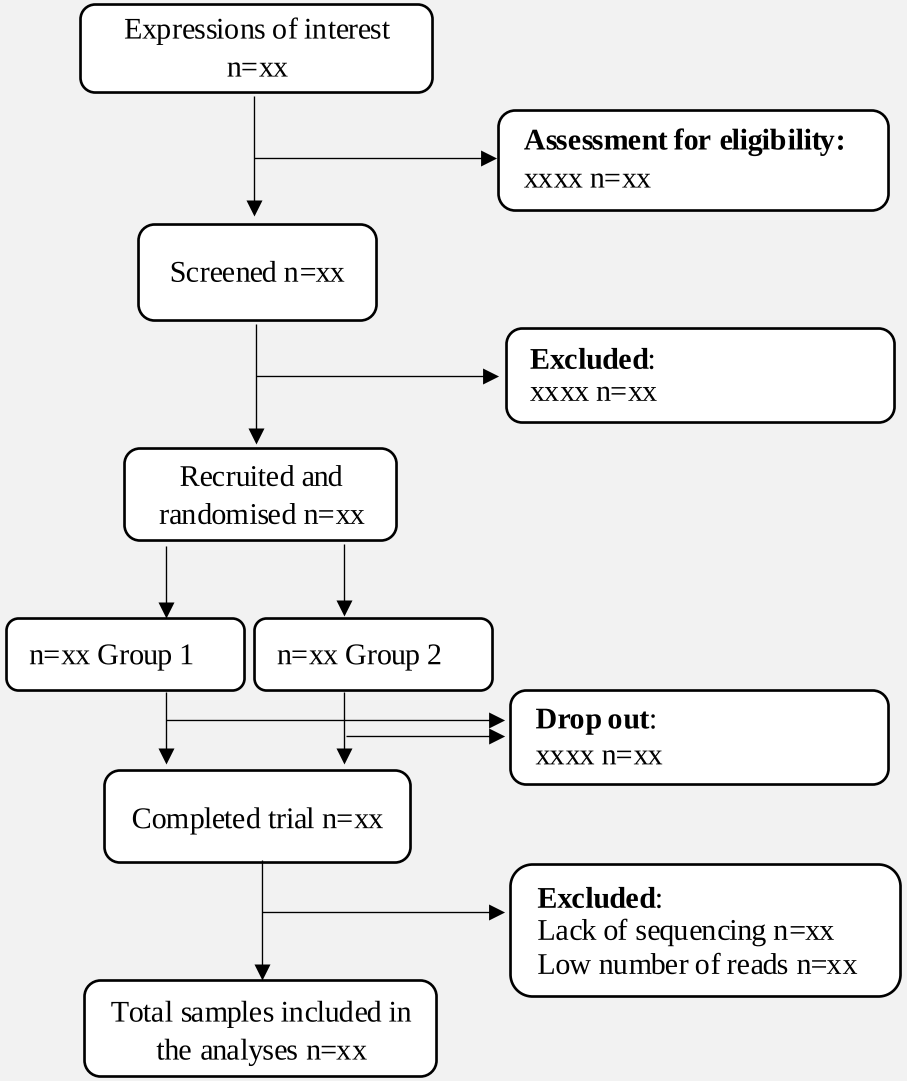
Drop out: xxxx n=xx

Data available for analyses n=xx

Excluded:
Lack of sequencing n=xx
Low number of reads n=xx

Total samples included in the analyses n=xx

Experimental



Expressions of interest n=xx

Assessment for eligibility: xxxx n=xx

Screened n=xx

Excluded: xxxx n=xx

Recruited and randomised n=xx

n=xx Group 1

n=xx Group 2

Drop out: xxxx n=xx

Completed trial n=xx

Excluded:
Lack of sequencing n=xx
Low number of reads n=xx

Total samples included in the analyses n=xx

Mirzayi et al, Nature Medicine, 2021

https://www.stormsmicrobiome.org/