

# micRobiome analysis

Kirsty Lee Garson  
Katie Lennard

## The microbiome is an essential part of human physiology

- more than half of the genes which constitute the human body are microbial
- this microbial gene pool is diverse and highly dynamic:
  - sensitive to changes in diet and behaviour
  - regular exchange of genes between microbes occurs
  - mutations occur more frequently than in the human genome

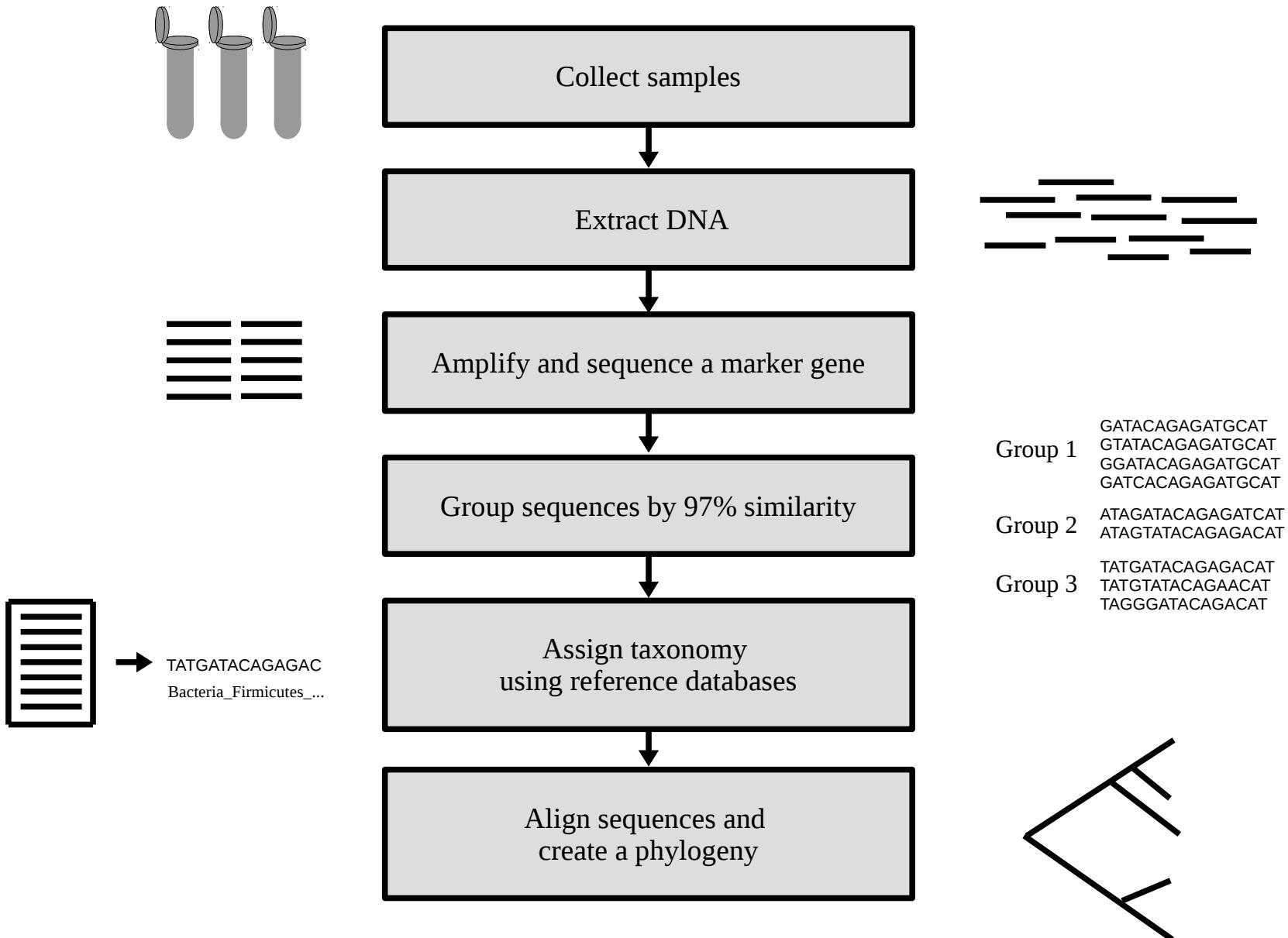
## The microbiome is an essential part of human physiology

- the human microbiome plays a role in numerous physiological functions, including immunity and nutrition
- alterations in the composition of the human microbiome have been linked to a wide variety of diseases

## Microbiome research is developing rapidly

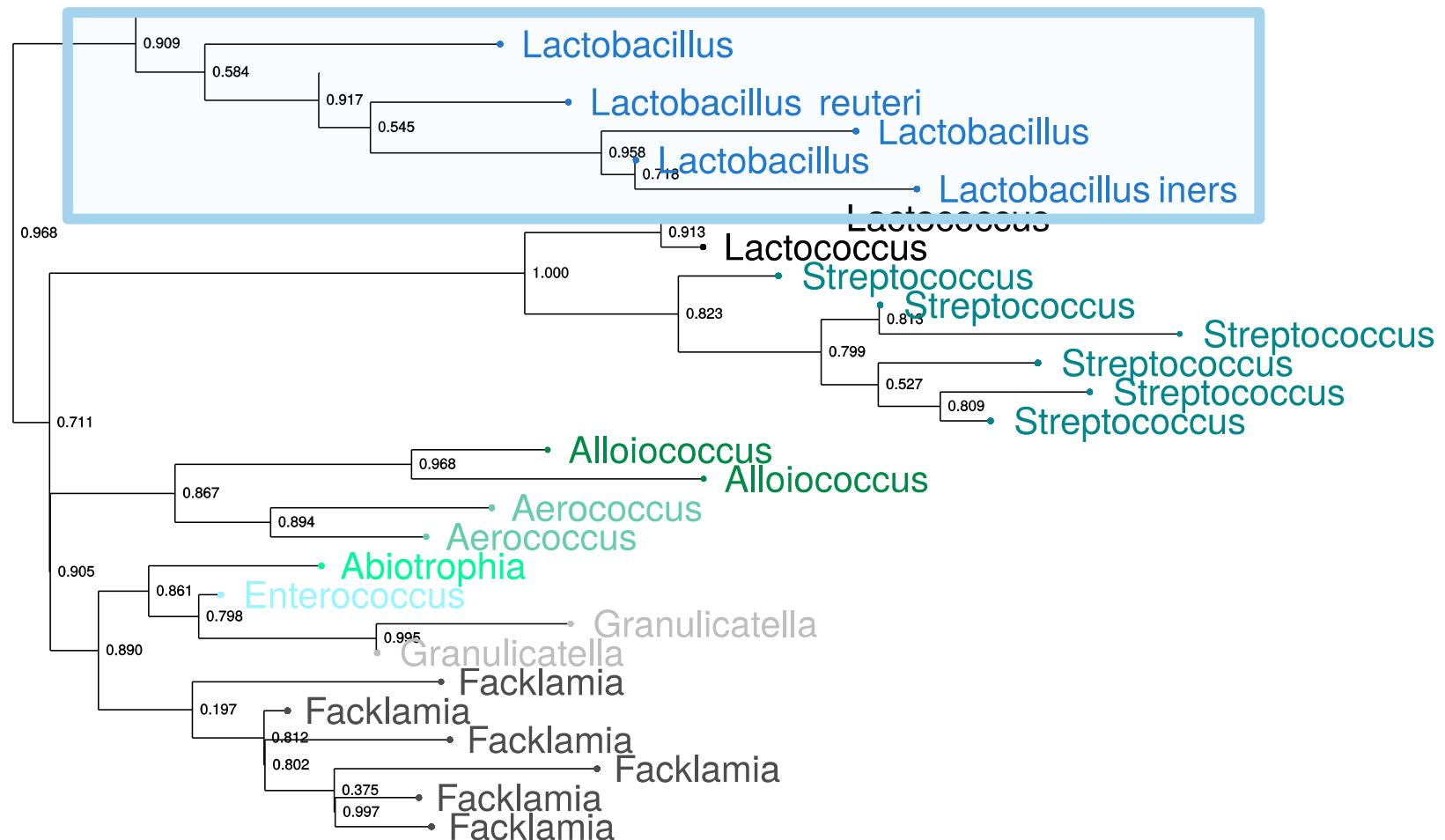
- this is as result of recent advances in sequencing technology  
*(next generation sequencing)*
- there has been a corresponding surge in the development of analysis tools, many of which are implemented in R

# Aim One: Derive sequence data from clinical samples



## Aim One: Derive sequence data from clinical samples

Kingdom	Phylum	Class	Order	Family	Genus	Species
"Bacteria"	"Firmicutes"	"Bacilli"	"Lactobacillales"	"Lactobacillaceae"	"Lactobacillus"	"iners"



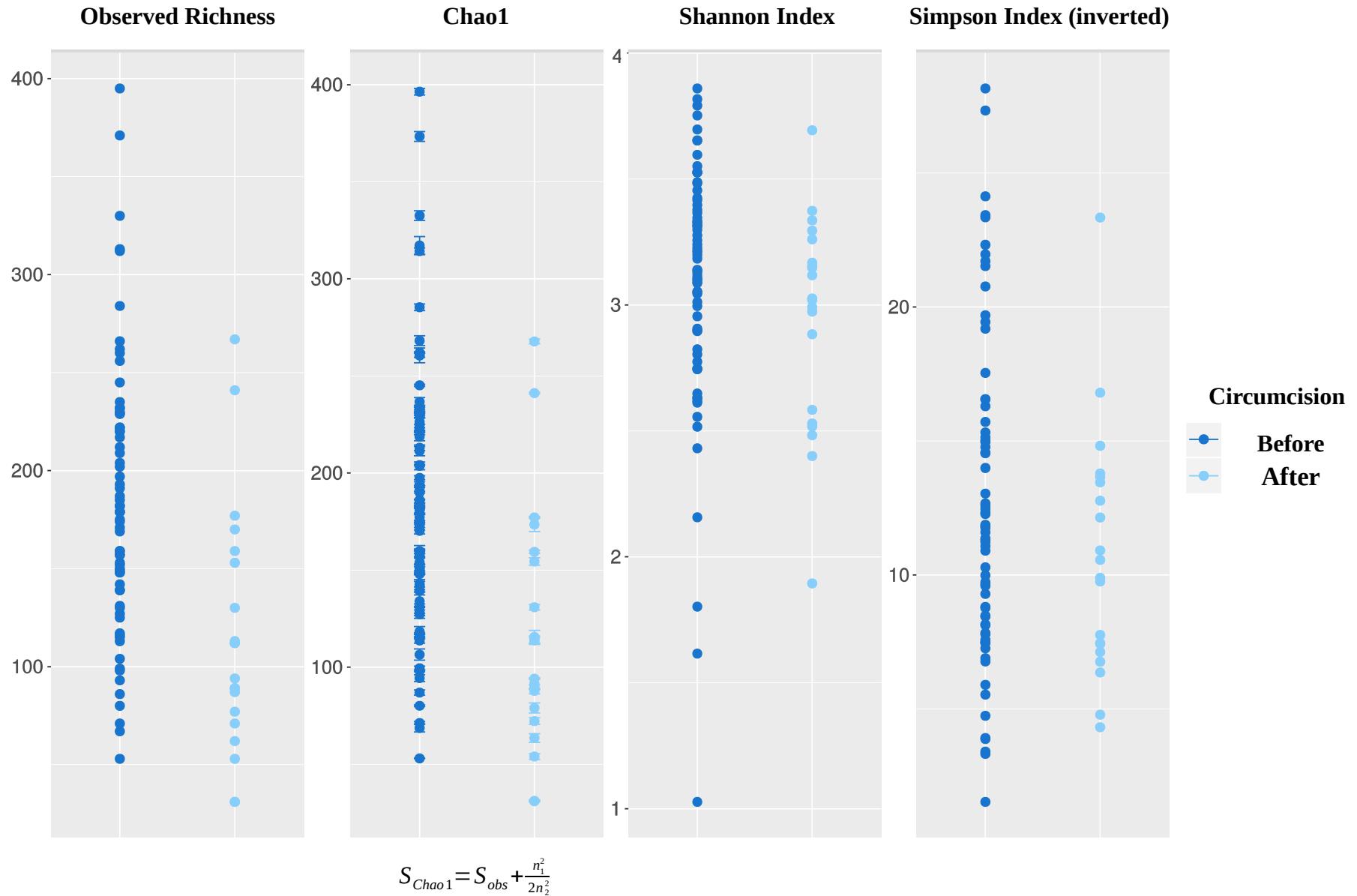
## **Aim Two:** Analyse the microbial composition of samples

What is the composition of these bacterial communities?

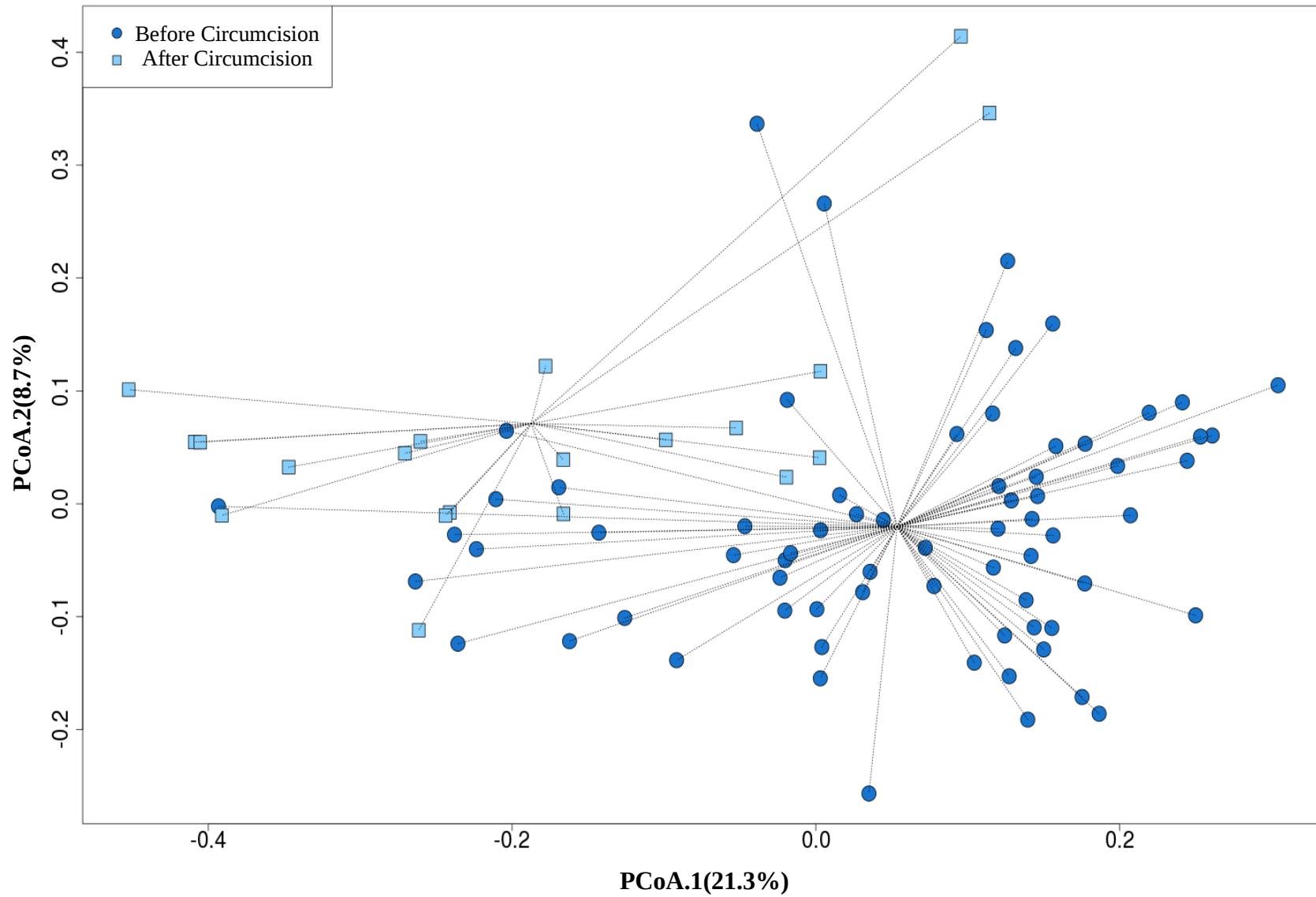
How does it differ between healthy individuals and those with disease?

**(marker gene analysis)**

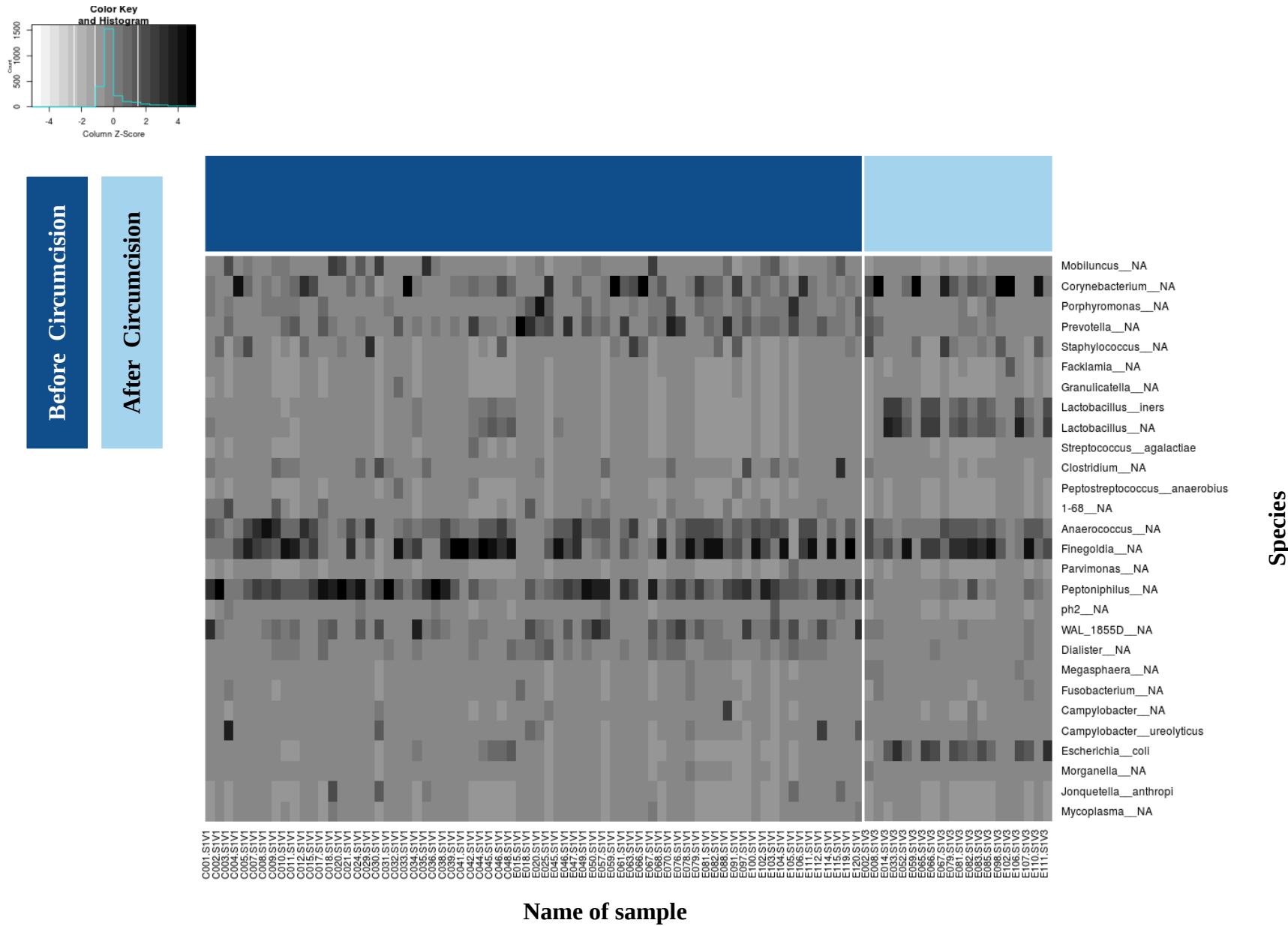
## 2.1 Measure diversity within samples



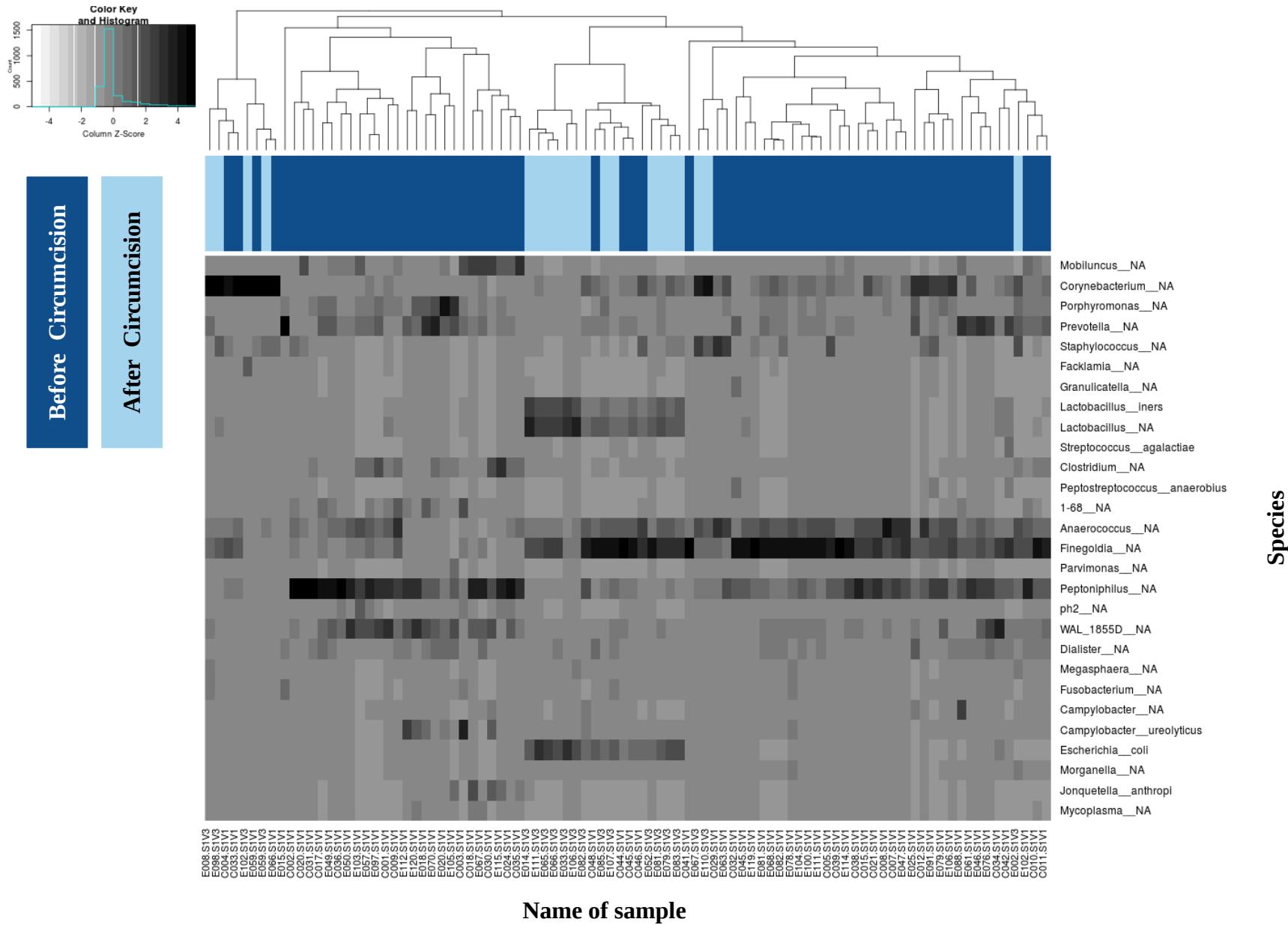
## 2.2 Measure diversity across samples



## 2.2 Measure diversity across samples



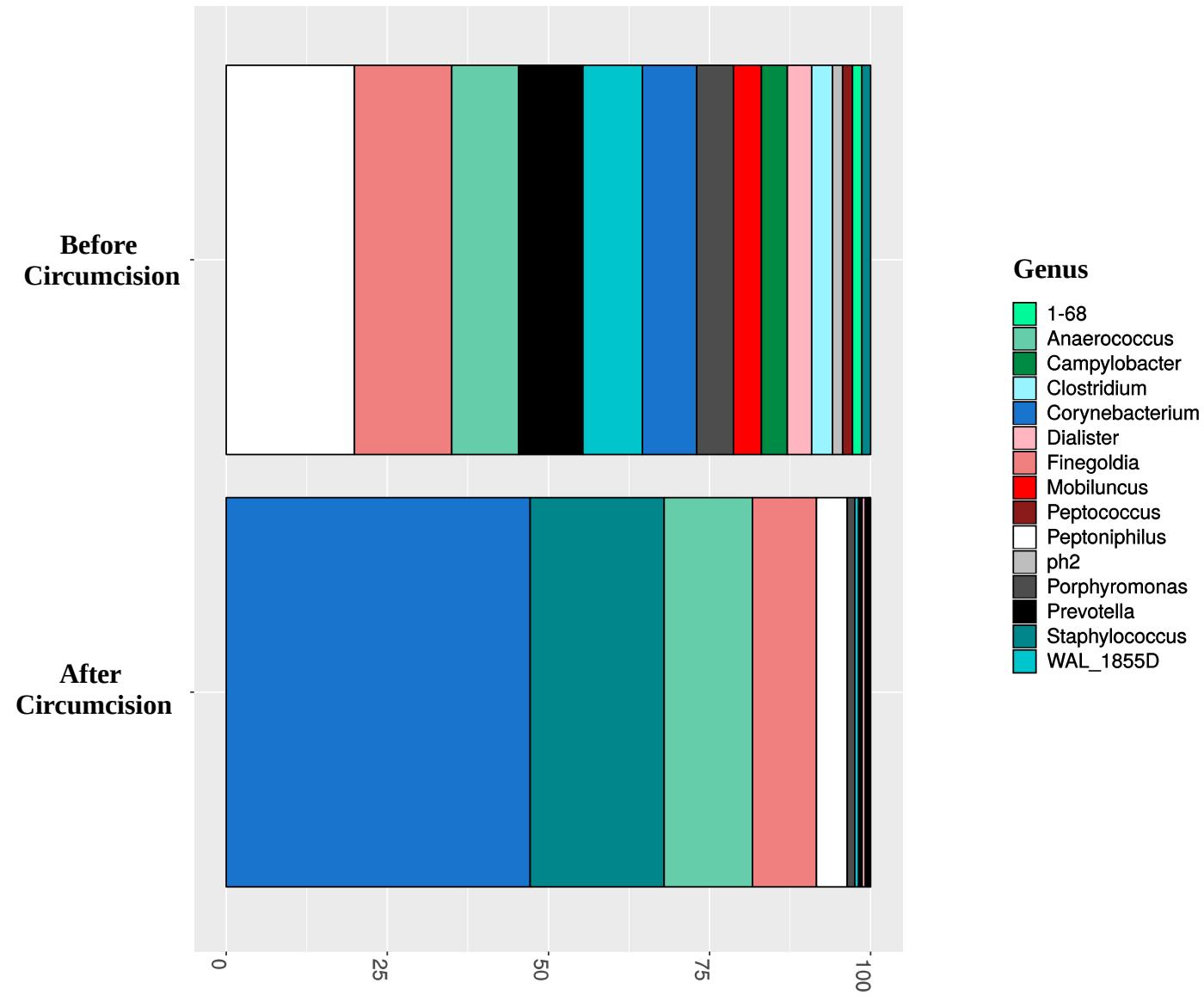
## 2.2 Measure diversity across samples



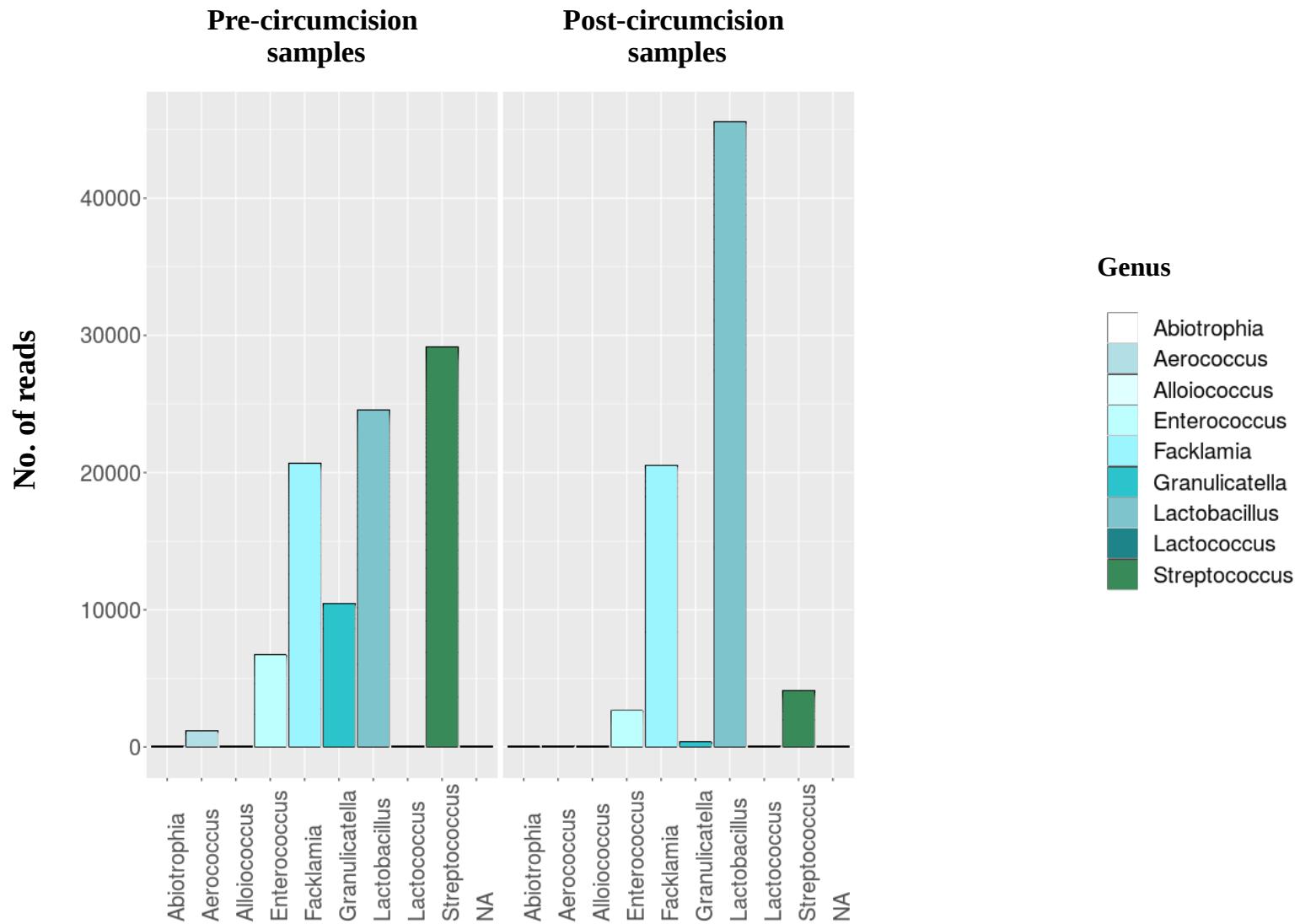
## 2.3 Assess sample composition



## 2.4 Compare sample composition across groups



## 2.5 Compare the number of times microbes of interest occur across groups



## **Aim Three:** Gain insight into the role these microbes play

Which genes are present?

What is the microbial community capable of?

**(functional prediction, shotgun metagenomics)**



Which genes are being expressed?

Which capabilities are currently in use?

**(transcriptomics, proteomics)**



Which metabolic processes are underway?

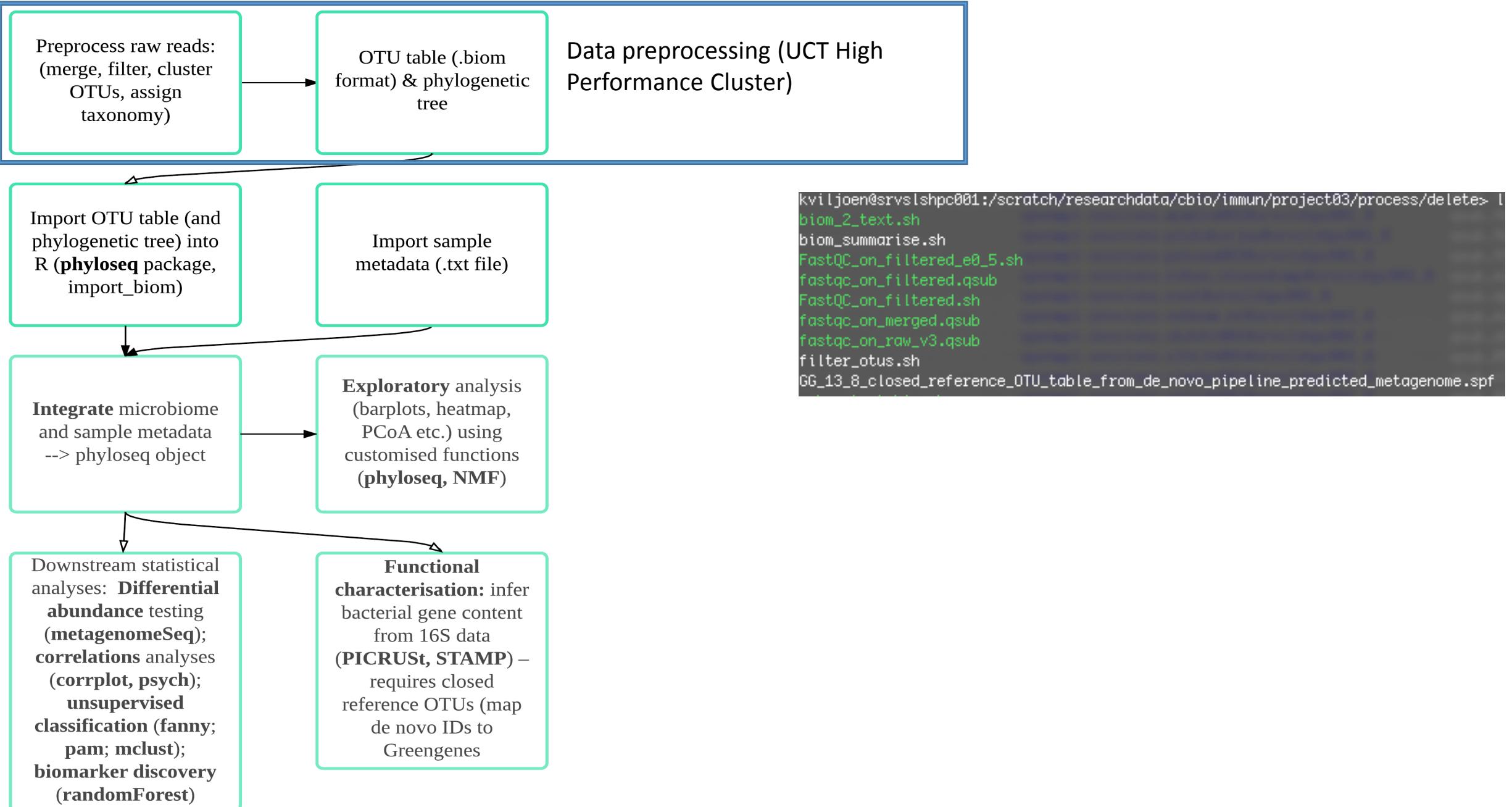
How are these capabilities being used to influence the human-microbial environment?

**(metabolomics)**

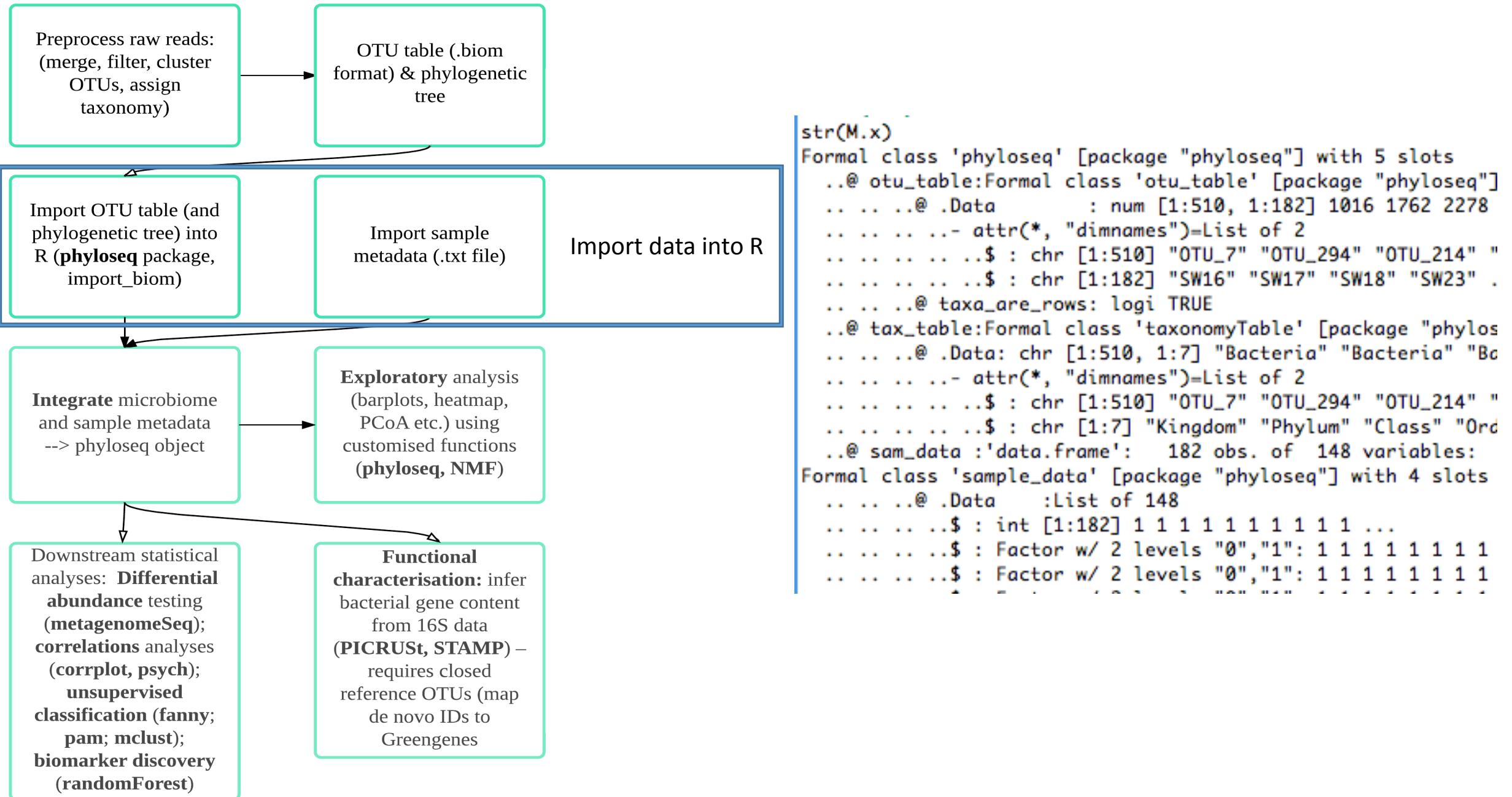
# Cbio 16S analysis pipeline

Katie Lennard

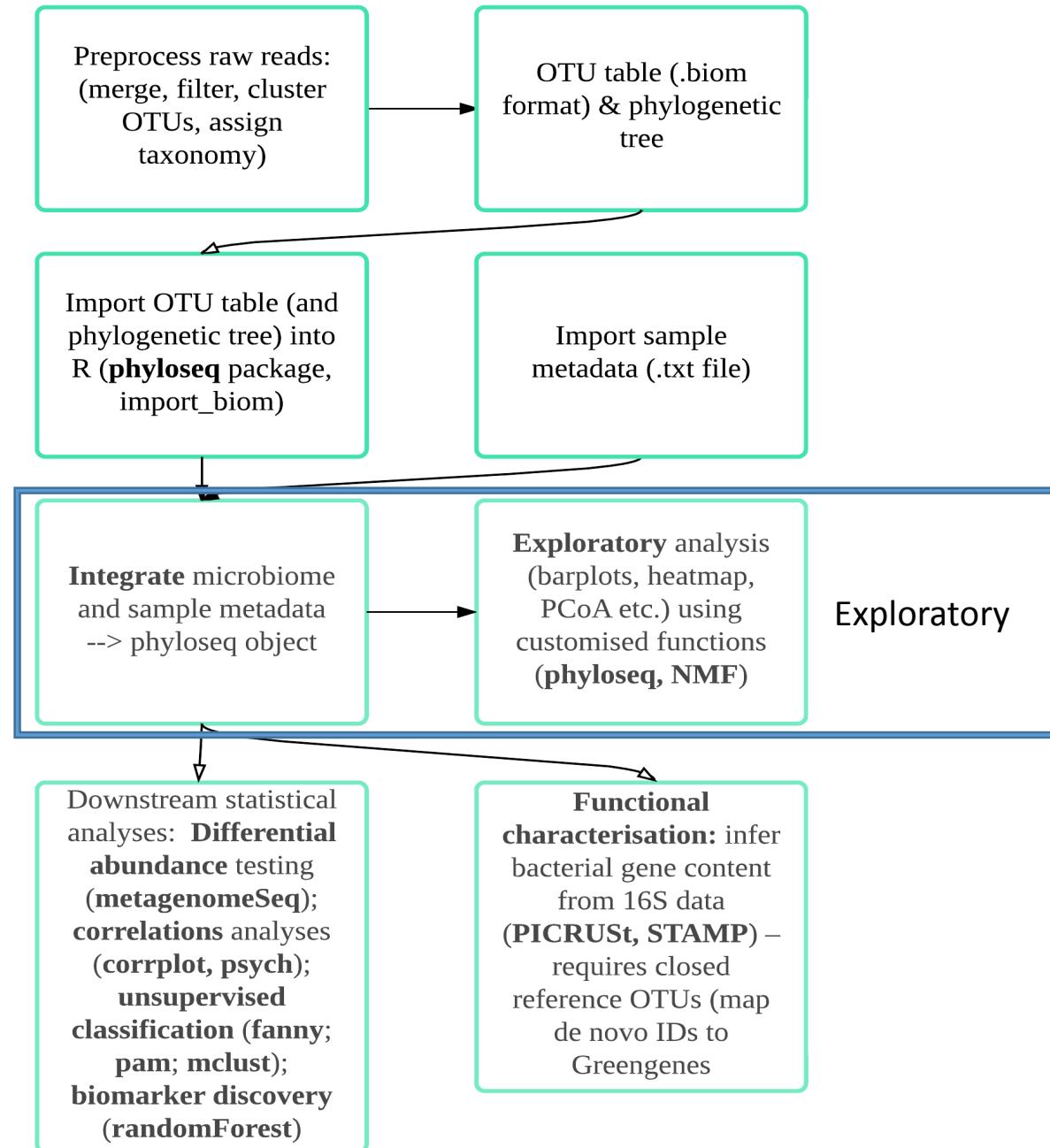
# Microbiome analysis workflow



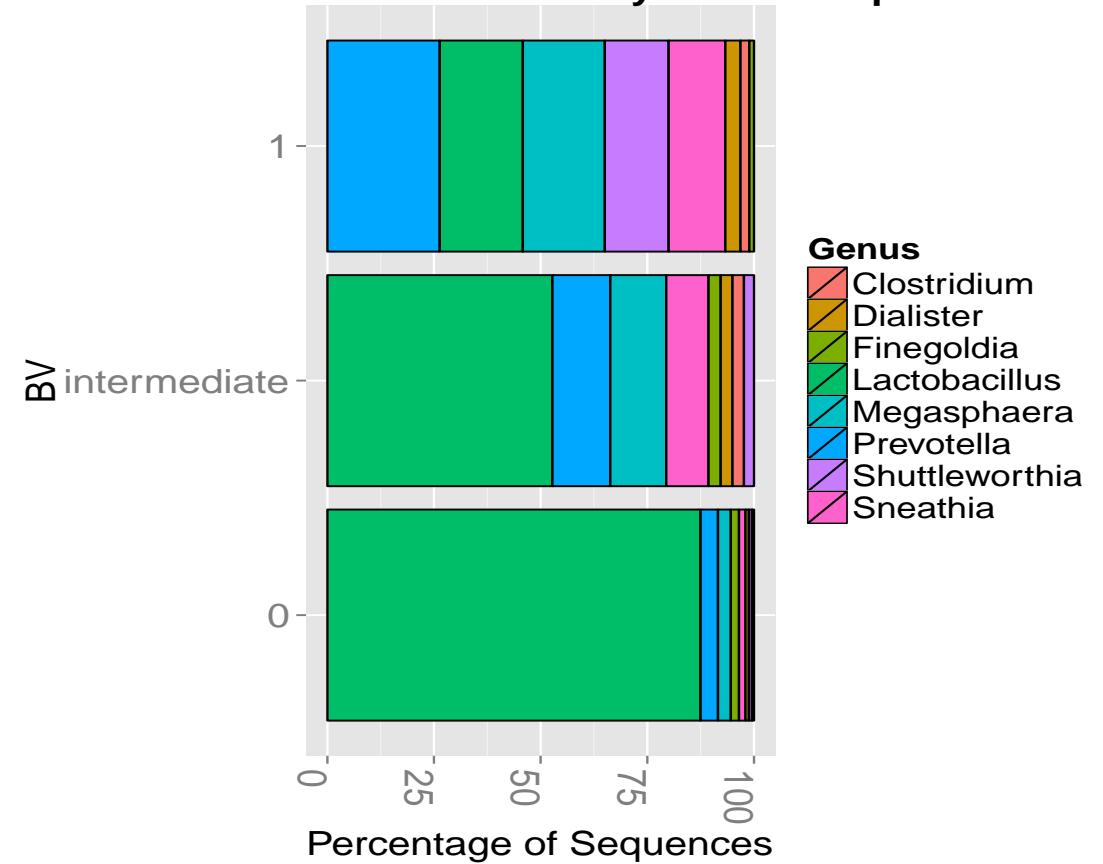
# Microbiome analysis workflow



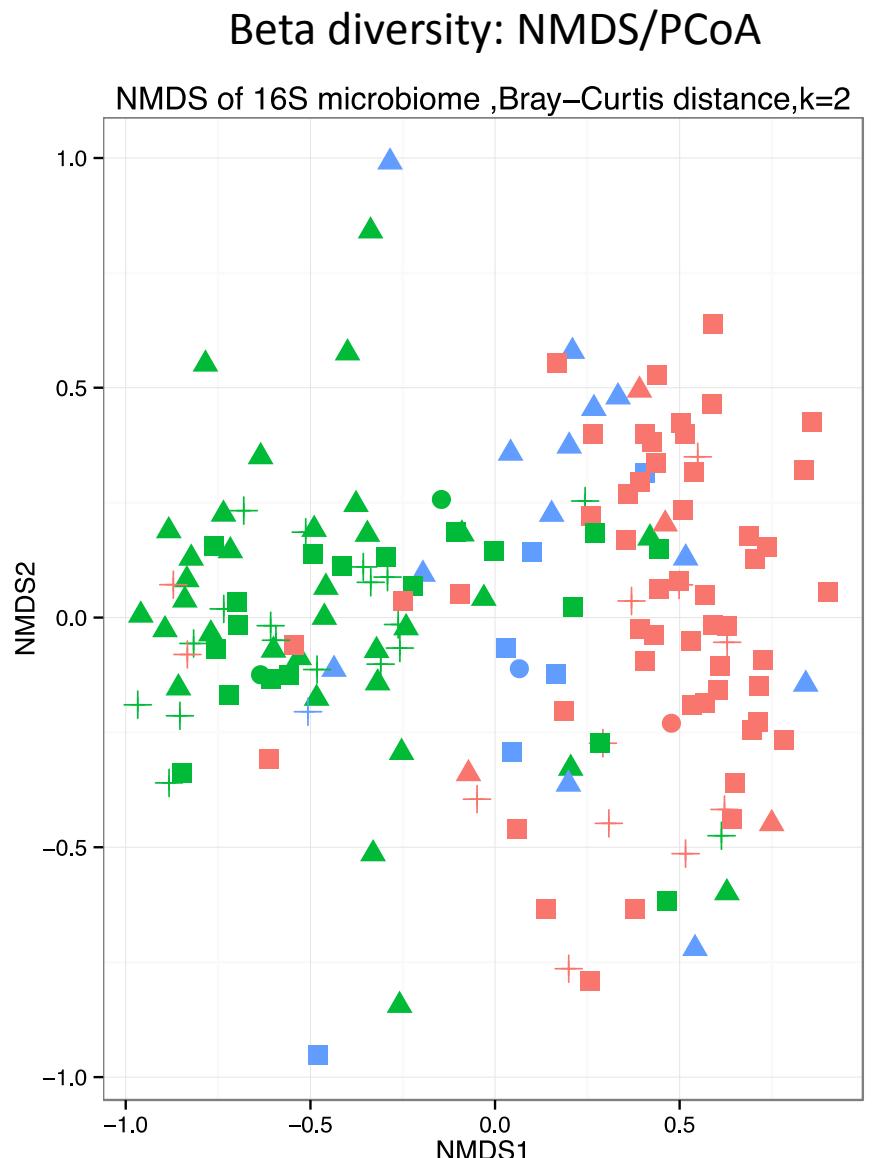
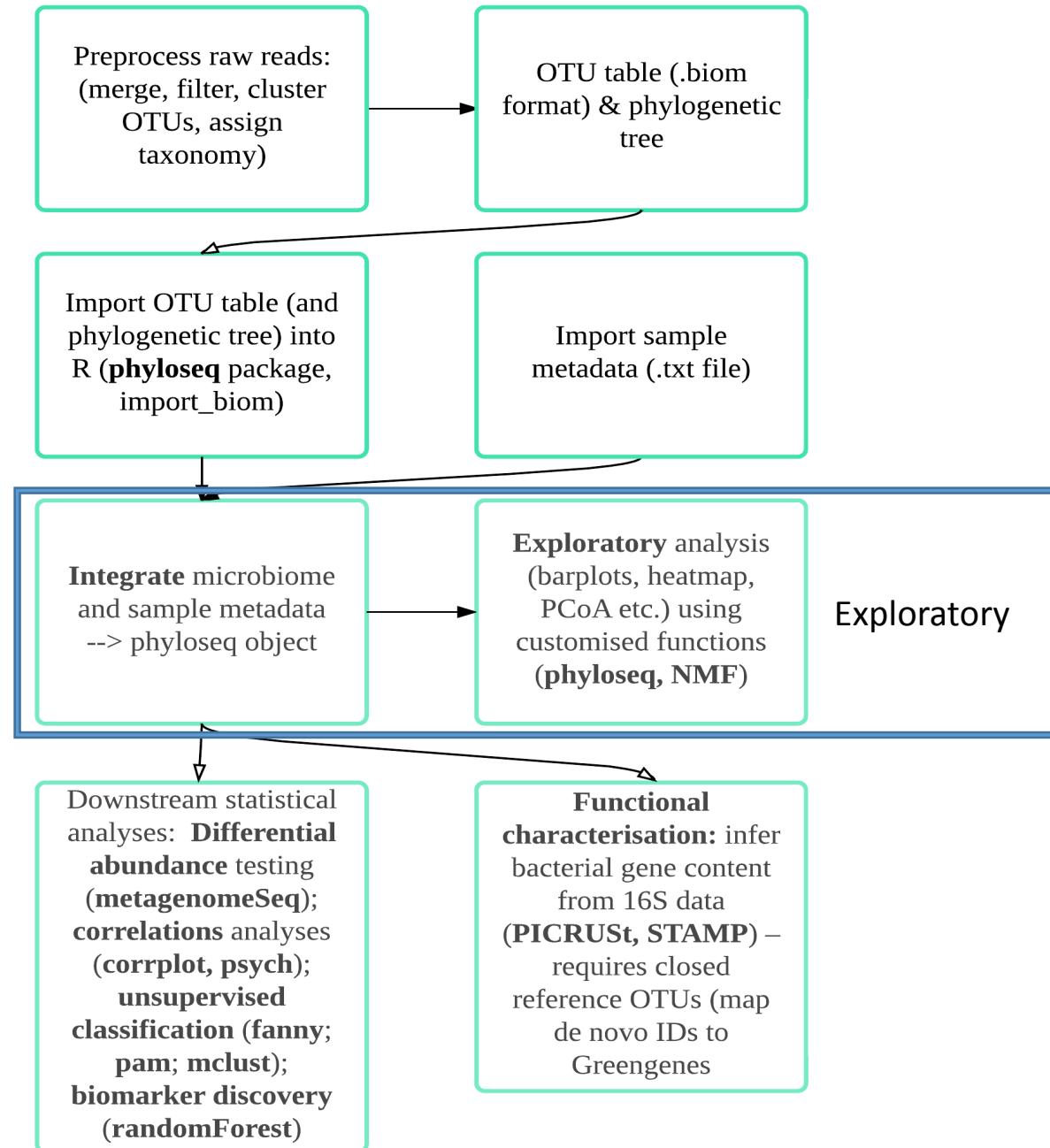
# Microbiome analysis workflow



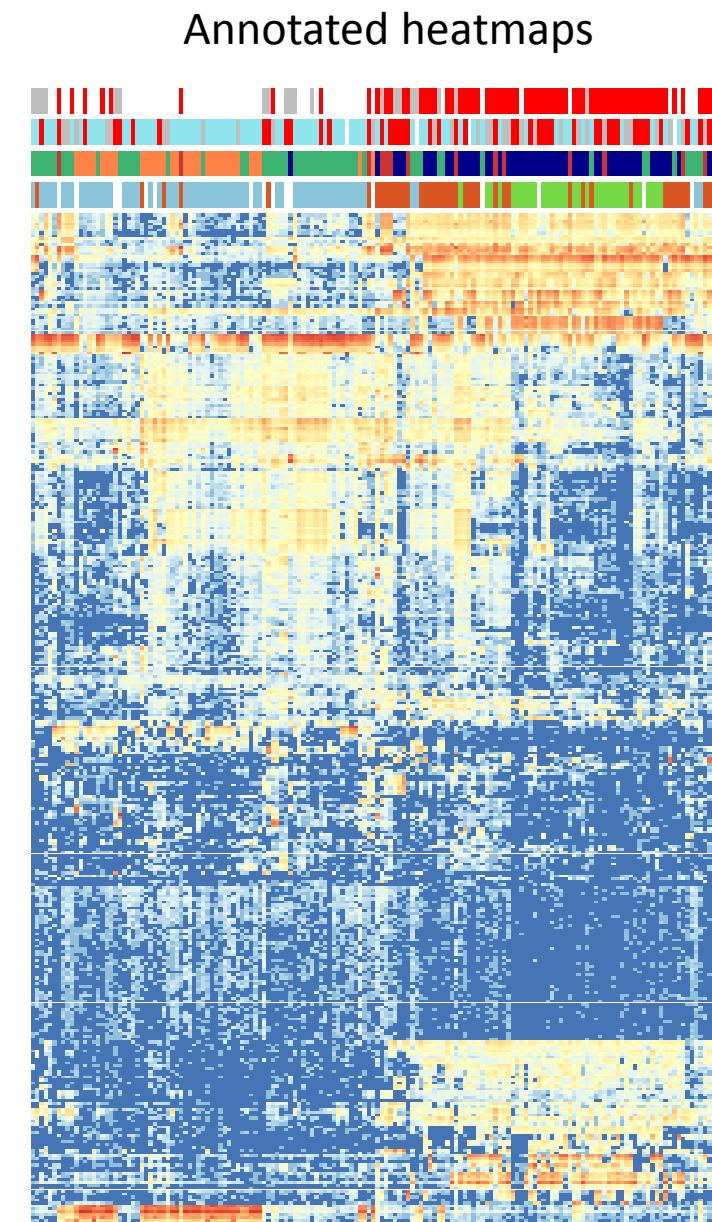
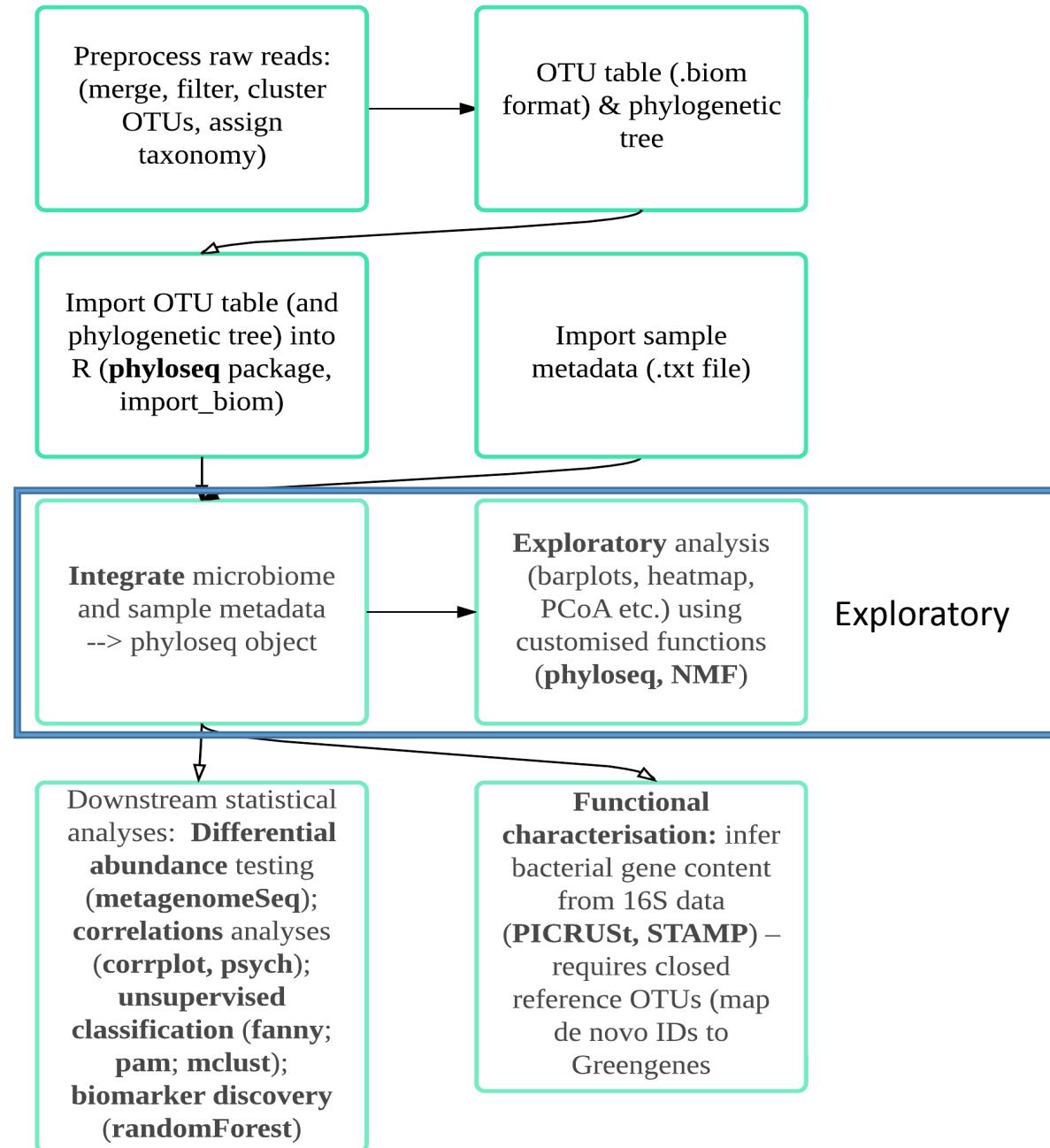
Summary barplots  
Genus-level abundance by BV all samples



# Microbiome analysis workflow



# Microbiome analysis workflow



# Microbiome analysis workflow

Preprocess raw reads:  
(merge, filter, cluster  
OTUs, assign  
taxonomy)

OTU table (.biom  
format) & phylogenetic  
tree

Import OTU table (and

## Differential abundance testing

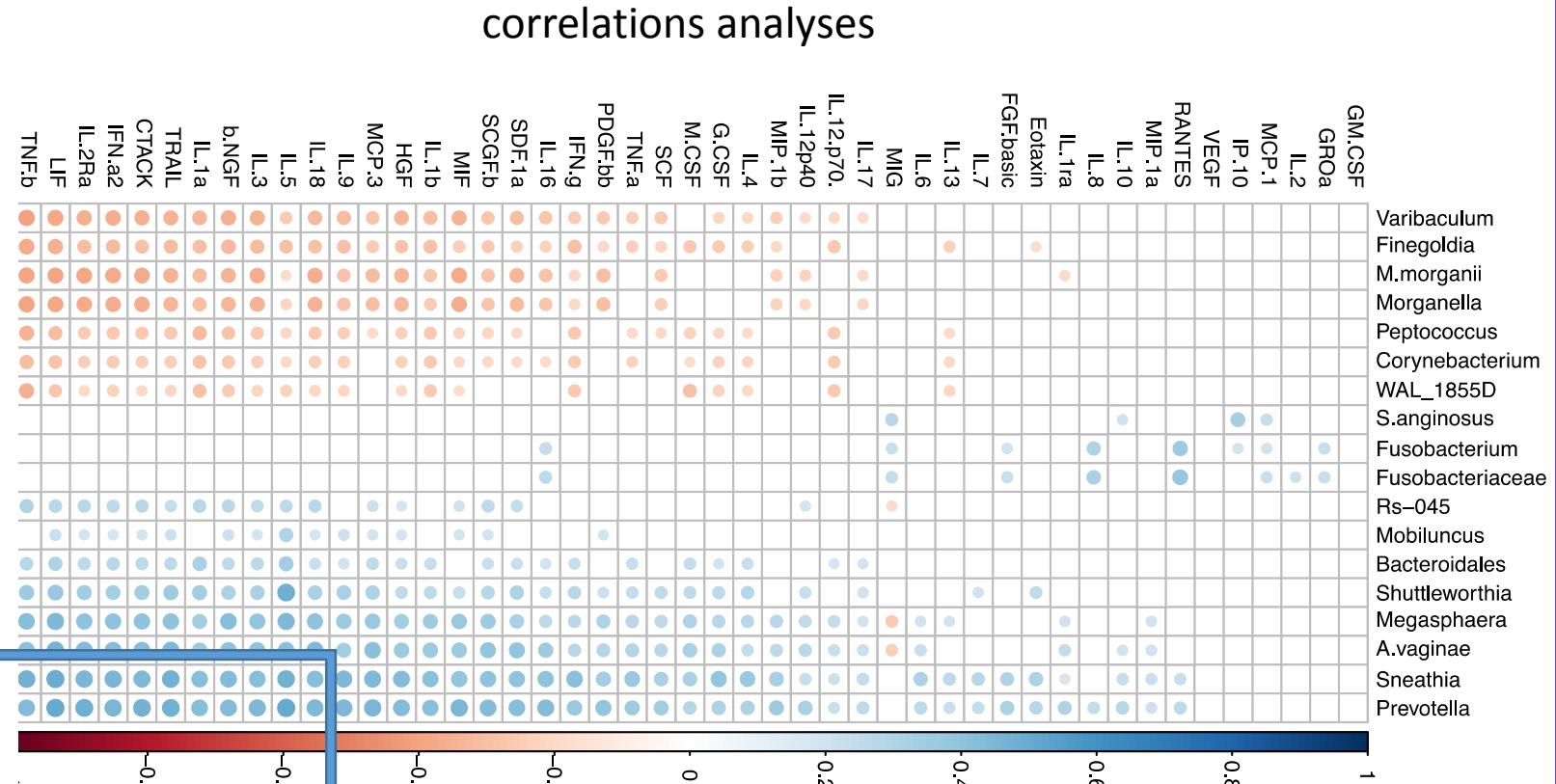
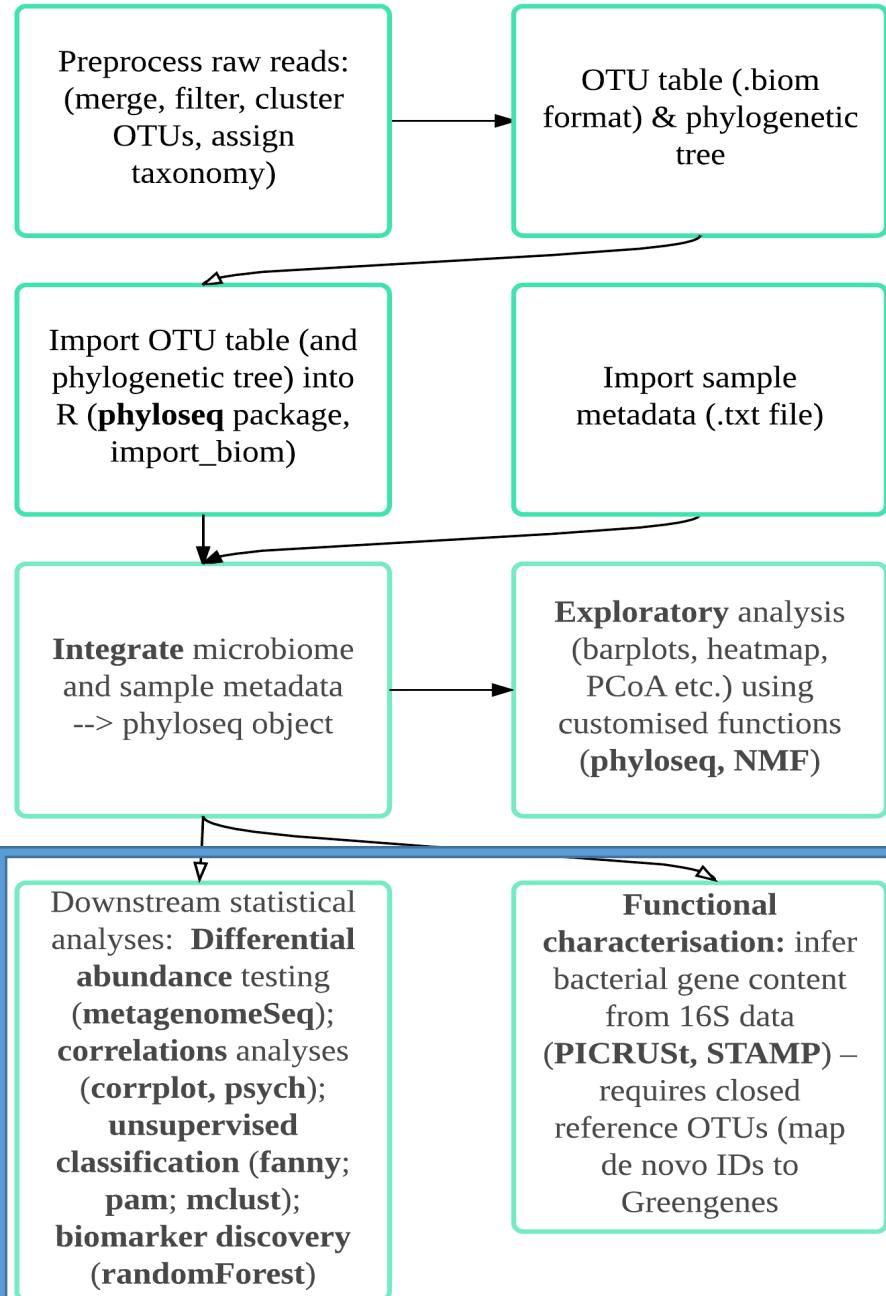
	percent_positive_group0	percent_positive_group1	+samples in group 0	+samples in group 1	mean_positive_group0	mean_positive_group1	oddsRatio	lower	upper	fisherP	fisherAdjP	coeff	pvalues	adjPvalues	Kingdom
OTU_254	100	100	89	72	9139	2602	0	0	Inf	1	1	-4.3814786	3.50E-08	1.18E-07	Bacteria
OTU_59	97.8	93.1	87	67	4243	1414	3.2233077	0.50855097	34.8420058	0.24392345	0.39872103	-4.3369	3.66E-08	1.23E-07	Bacteria
OTU_27	97.8	91.7	87	66	3220	968	3.92244426	0.67377754	40.9629286	0.14084917	0.25654671	-4.3022808	2.81E-08	9.67E-08	Bacteria
OTU_314	19.1	4.2	17	3	67	46	5.38149246	1.46306825	29.9353159	0.00398916	0.01667601	-4.2805204	5.79E-27	1.96E-25	Bacteria
OTU_417	87.6	81.9	78	59	1494	639	1.55799049	0.59630478	4.14218949	0.37566912	0.54121822	-3.9347862	5.41E-08	1.75E-07	Bacteria
OTU_9	93.3	90.3	83	65	1266	456	1.48599802	0.40568486	5.63223074	0.56702008	0.71402529	-3.8579988	6.56E-08	2.07E-07	Bacteria
OTU_148	82	63.9	73	46	1064	119	2.56313091	1.18007331	5.71388377	0.01150991	0.03578633	-3.2256537	1.20E-07	3.51E-07	Bacteria
OTU_2	76.4	47.2	68	34	389	219	3.58801854	1.75318671	7.52114879	0.00015686	0.00114286	-3.0135063	1.98E-06	4.97E-06	Bacteria
OTU_255	84.3	72.2	75	52	1549	141	2.05102697	0.8932306	4.82550329	0.08049993	0.16689011	-2.9487617	1.12E-05	2.56E-05	Bacteria
OTU_458	93.3	87.5	83	63	420	146	1.96779079	0.59001023	7.0899145	0.27742903	0.43940622	-2.8899709	1.89E-09	8.35E-09	Bacteria
OTU_178	100	100	89	72	10637	3687	0	0	Inf	1	1	-2.8716545	4.26E-06	1.04E-05	Bacteria
OTU_16	100	100	89	72	7149	2672	0	0	Inf	1	1	-2.6658928	2.35E-05	5.19E-05	Bacteria
OTU_307	100	98.6	89	71	4700	1956	Inf	0.03169487	Inf	0.44720497	0.61808817	-2.5817595	4.50E-05	9.62E-05	Bacteria

Downstream statistical analyses: **Differential abundance testing** (`metagenomeSeq`); **correlations** analyses (`corrplot`, `psych`); **unsupervised classification** (`fanny`; `pam`; `mclust`); **biomarker discovery** (`randomForest`)

**Functional characterisation:** infer bacterial gene content from 16S data (**PICRUSt**, **STAMP**) – requires closed reference OTUs (map de novo IDs to Greengenes)

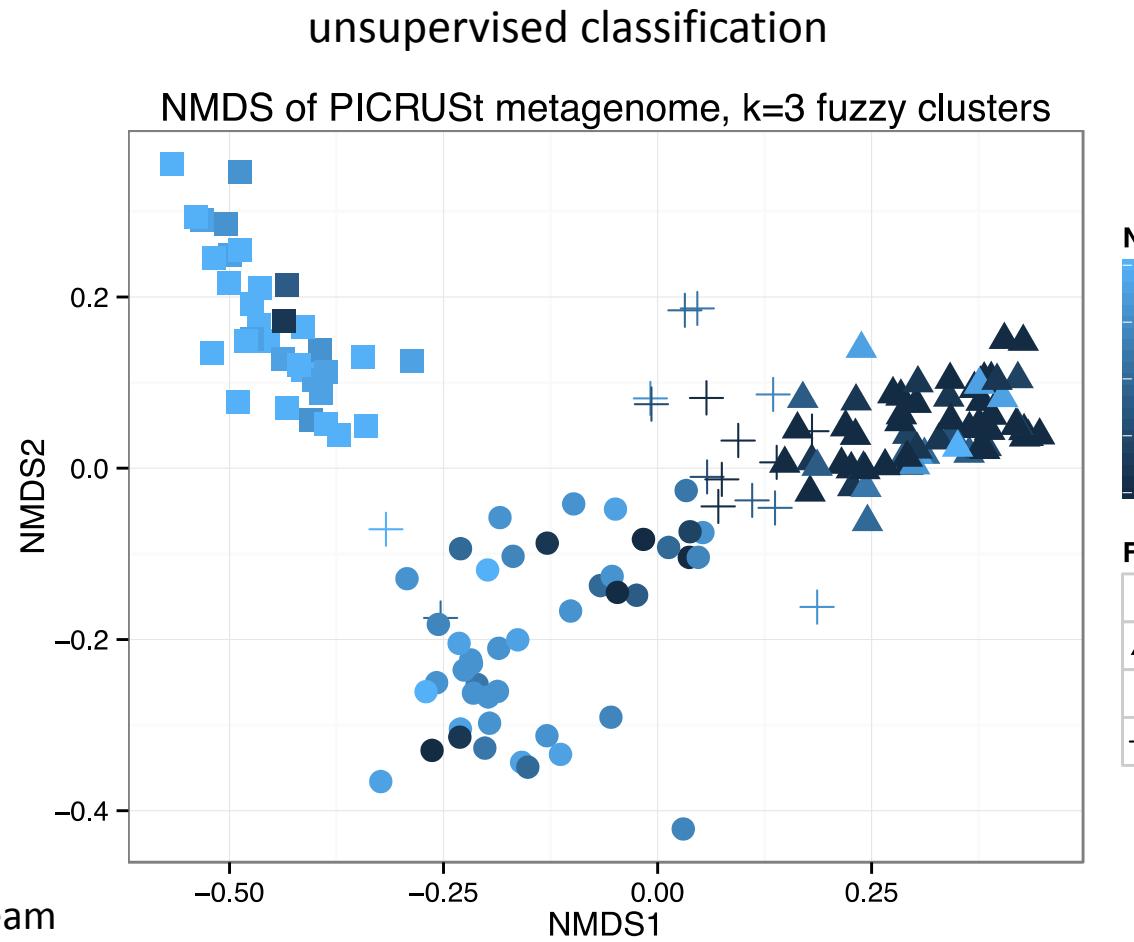
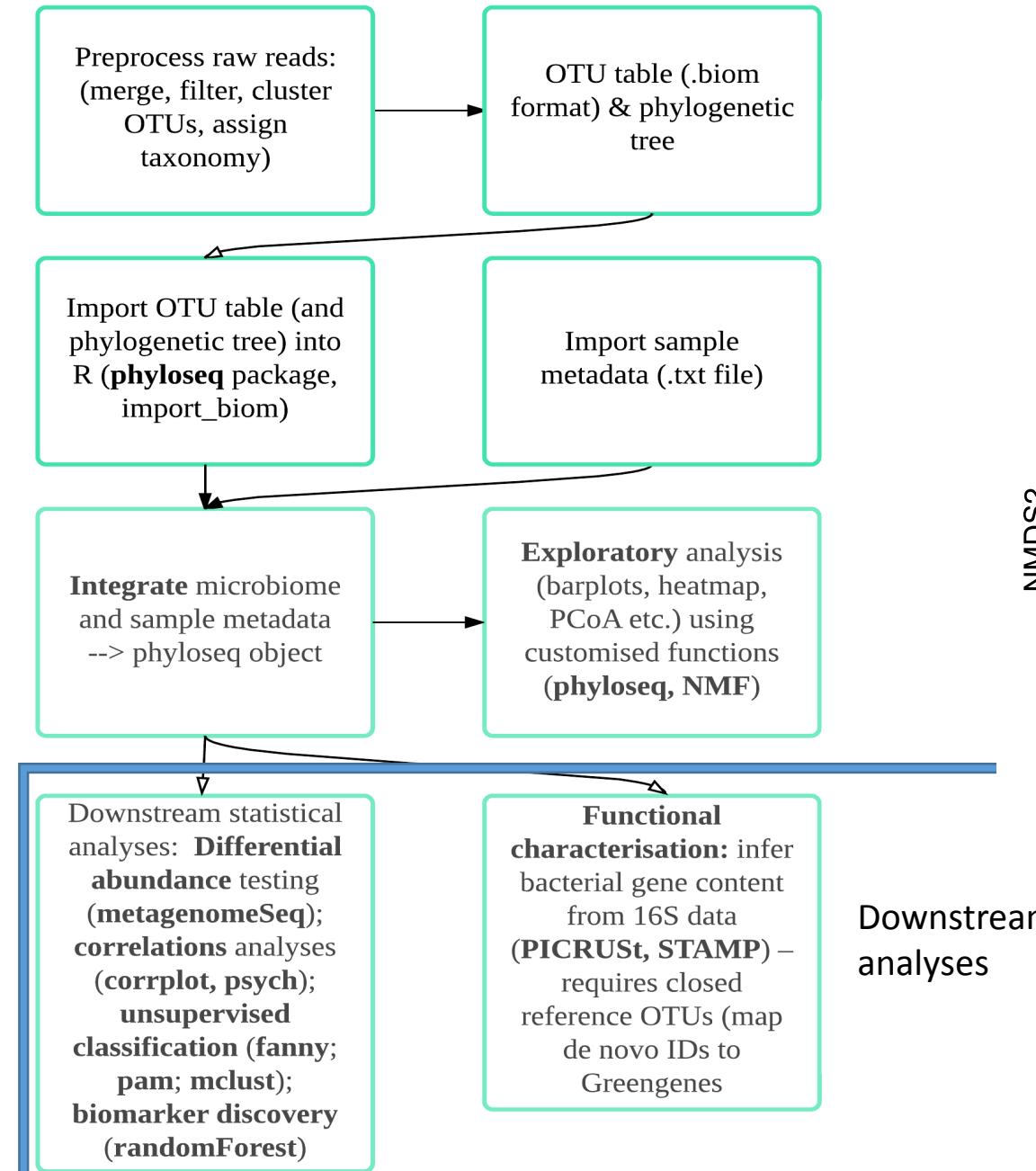
Downstream analyses

# Microbiome analysis workflow

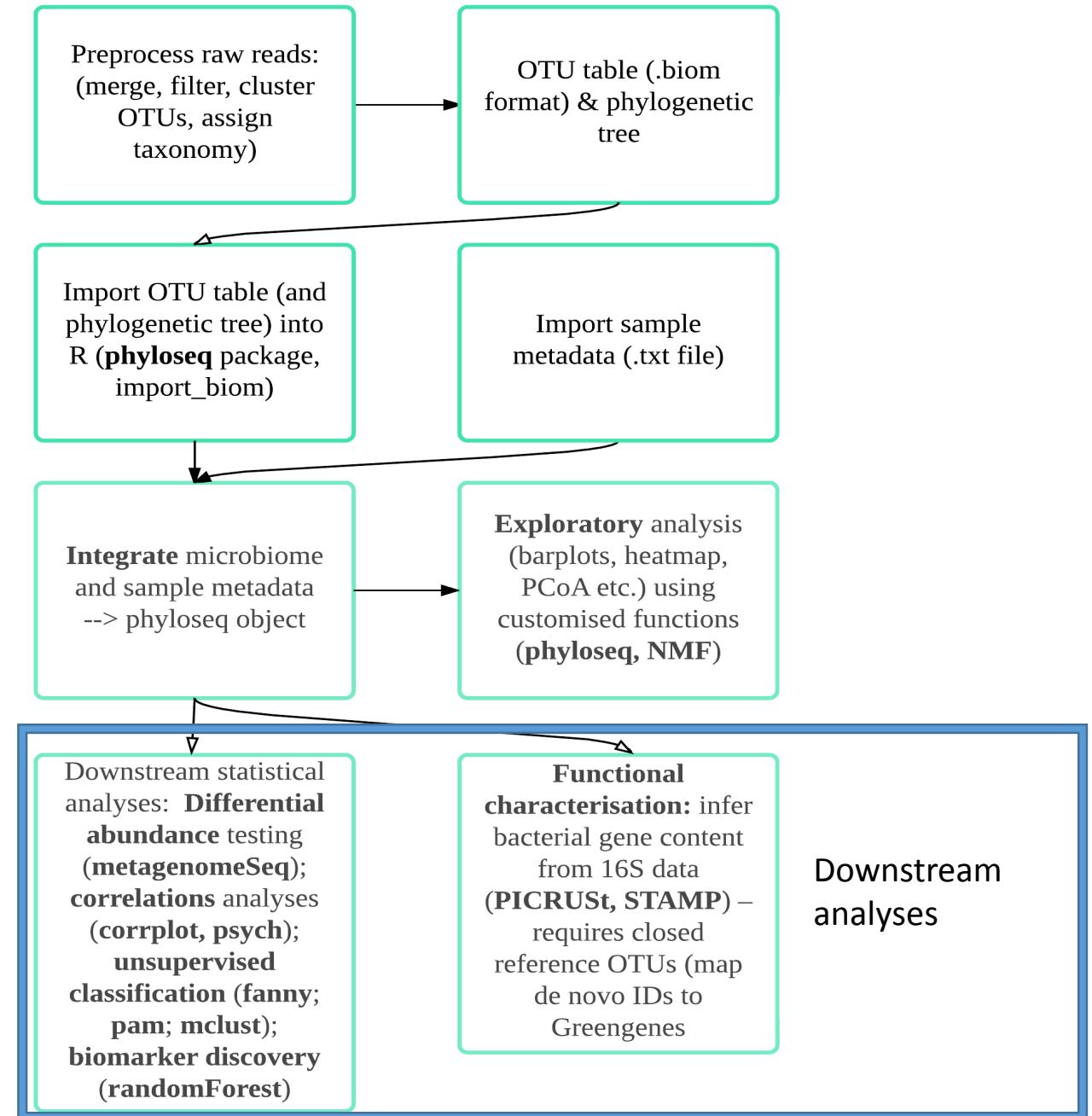


Downstream analyses

# Microbiome analysis workflow



# Microbiome analysis workflow



Biomarker discovery: random forests

ROC Curve taxa Predictors

