

## Contamination Study

Investigate Lab Intrinsic Contamination and PDAC intratumor microbiome

Linh Dang

# Introduction to the Project

This presentation covers the following sub-projects.

- **Contamination Survey:** Collection of negative controls (NCT) in the lab over period 2021-2024.
- **NCT Profile:** Investigate bacterial profile of NCT samples under various environments.
- **Microbiome in intratumor PDAC:** Benchmarking various decontamination approaches for investigation of microbiome in PDAC
- **Suggested Pipeline:**

# Contamination Survey

## *Original NCT collection:*

- **93** paraffin controls,
- **133** buffer controls,
- **11** PCR controls,
- **3** sequencing controls.

Resulting in **6128** taxa.

## *Trivial Filtering:*

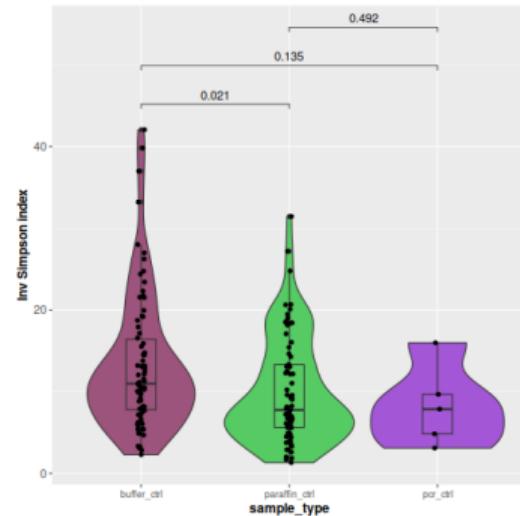
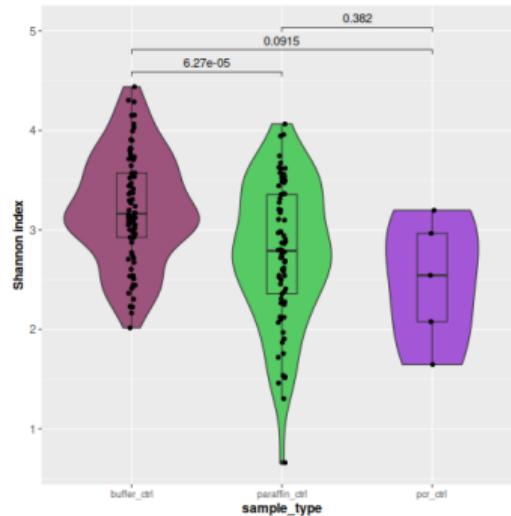
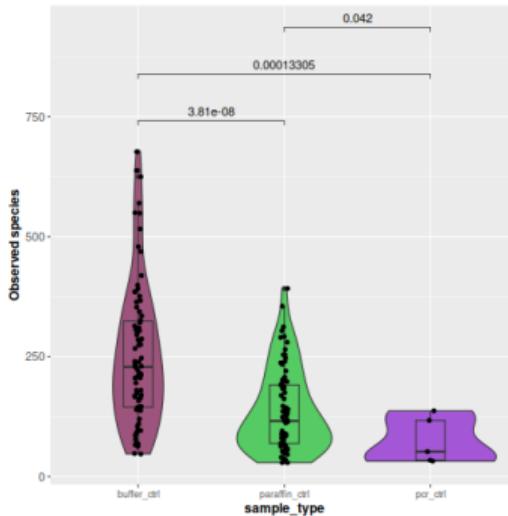
- Remove NCT samples with insufficient reads (**750**).
- Taxa filtering with following criteria:
  - threshold for prevalence detection: 2 reads
  - Minimum prevalence: 2 samples
  - Minimum abundance:  $1e-6$

Resulting in **1775** taxa, consist of:

- **84** paraffin controls,
- **133** buffer controls,
- **6** PCR controls.

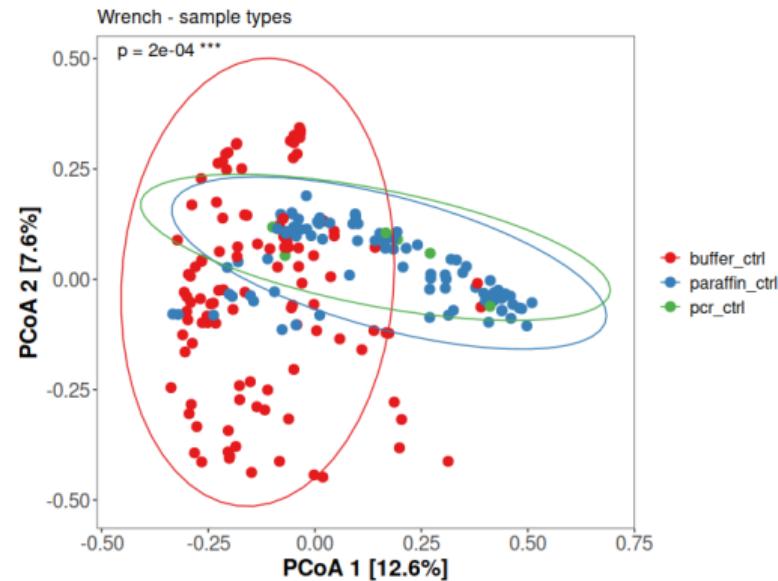
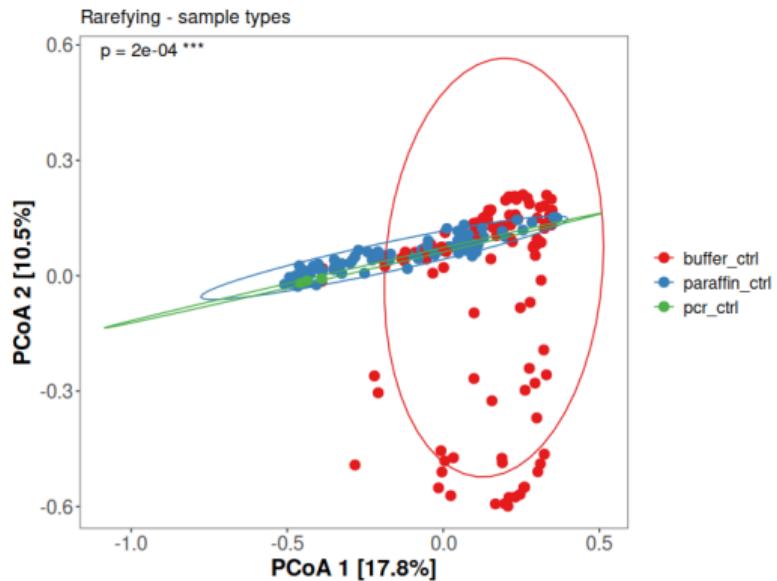
# Diversity Analyses

## Alpha Diversity w.r.t Sample Types



# Beta diversity

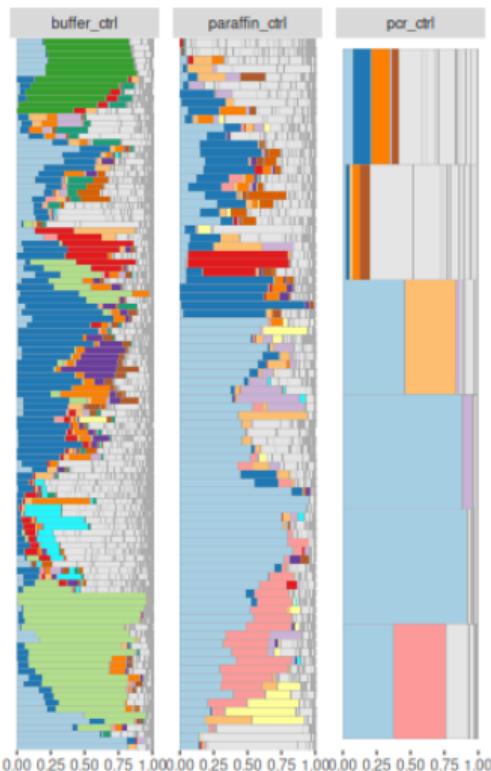
## Beta diversity with rarefaction and Wrench as normalization



# Bacterial Profile - Barplot

## Top 15 by abundance

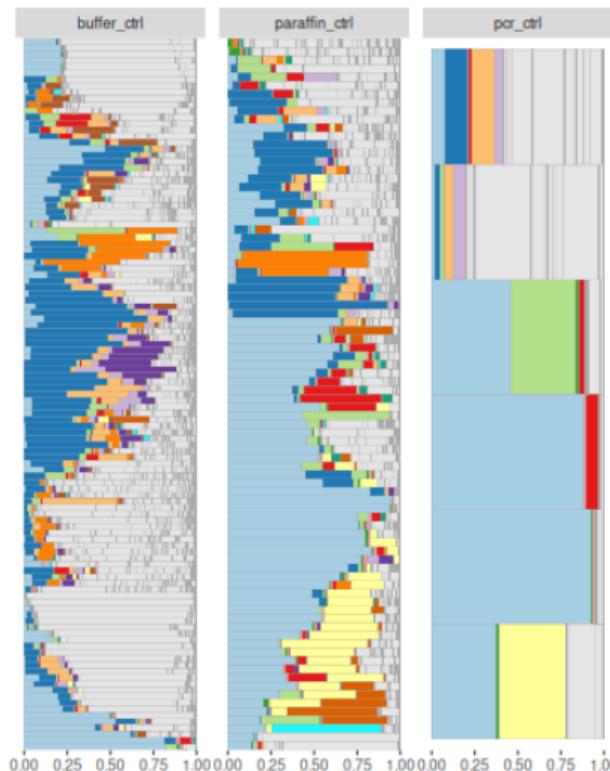
Sample types, order by sum



Linh Dang

## Top 15 by prevalence

Sample types, order by prevalent

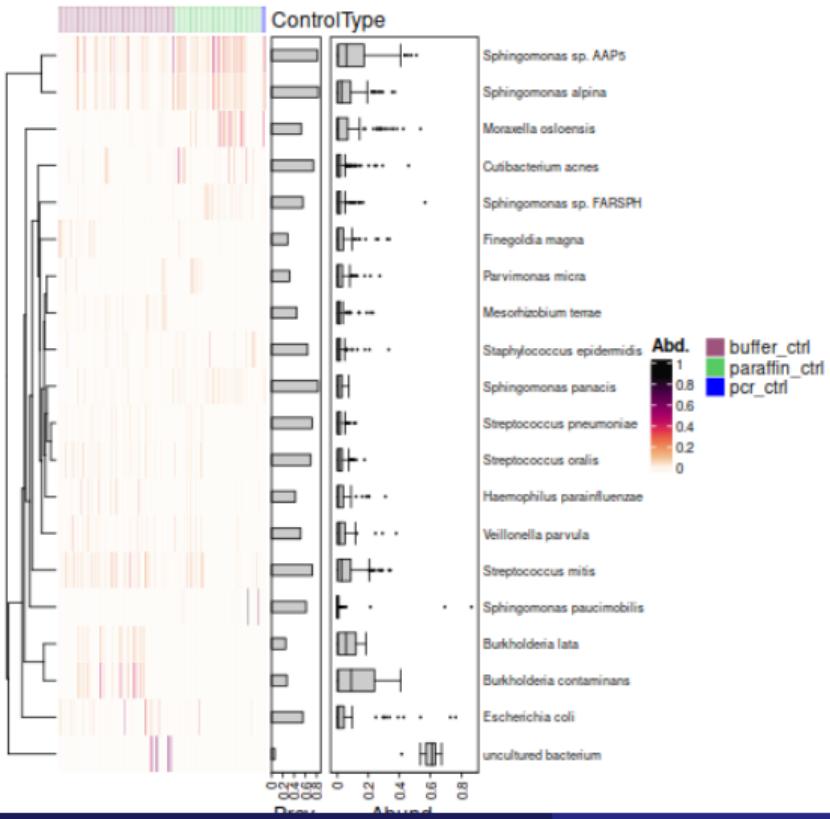


Contamination Study

6 / 36

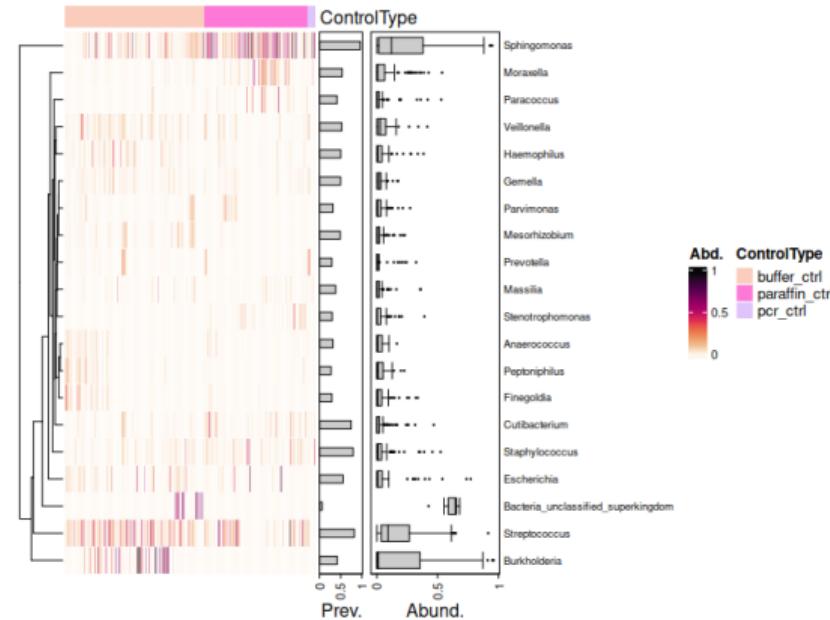
# Bacterial Profile - Heatmap

## Top 20 species order by abundance



Linh Dang

## Top 20 genus order by abundance

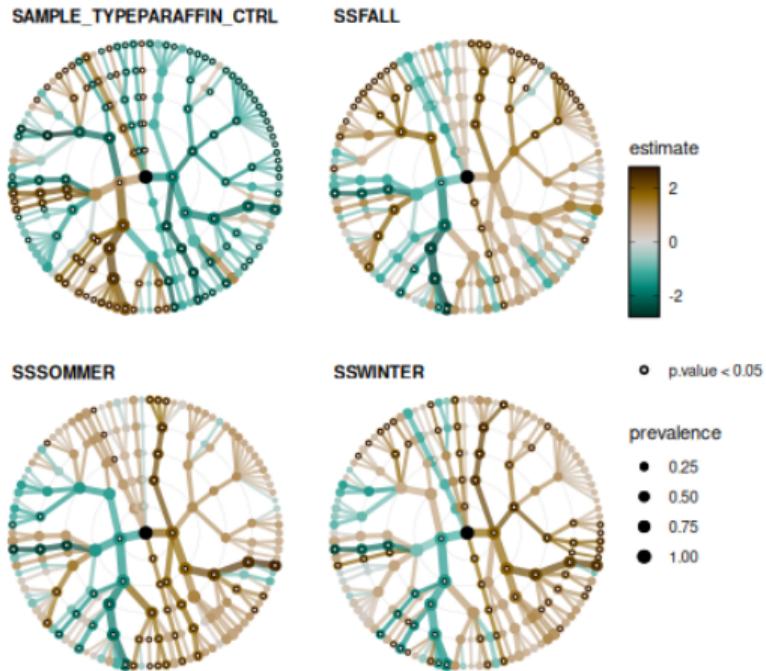


Contamination Study

7 / 36

# Trend in Contamination

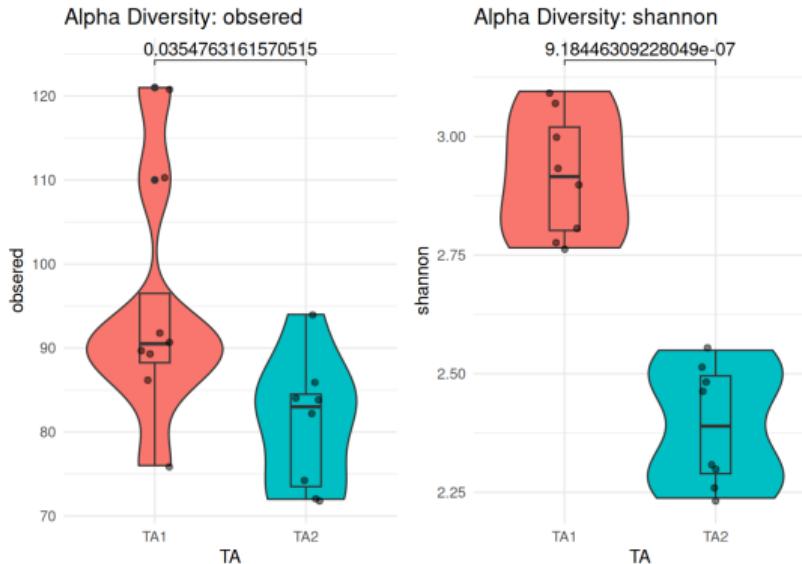
We applied DAA to find trends in contamination.



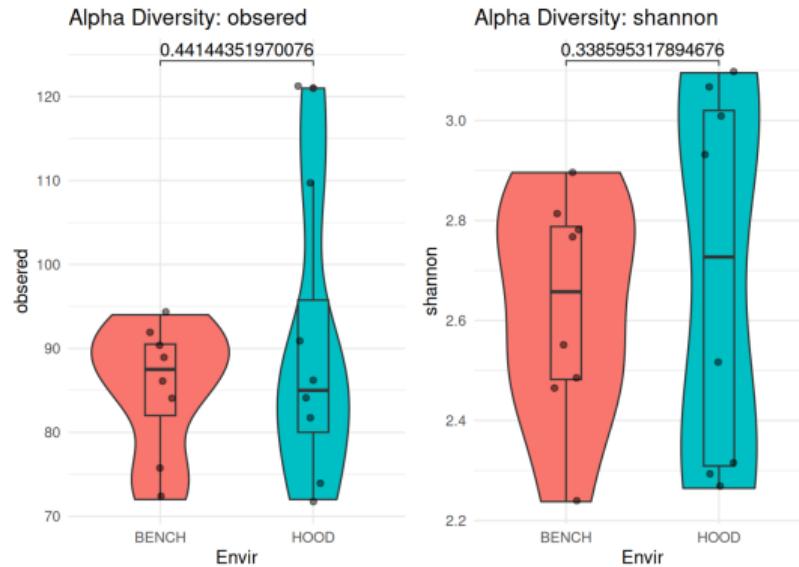
## Experiment Setup

# Hood vs. Bench - Alpha Diversity

## Between technicians

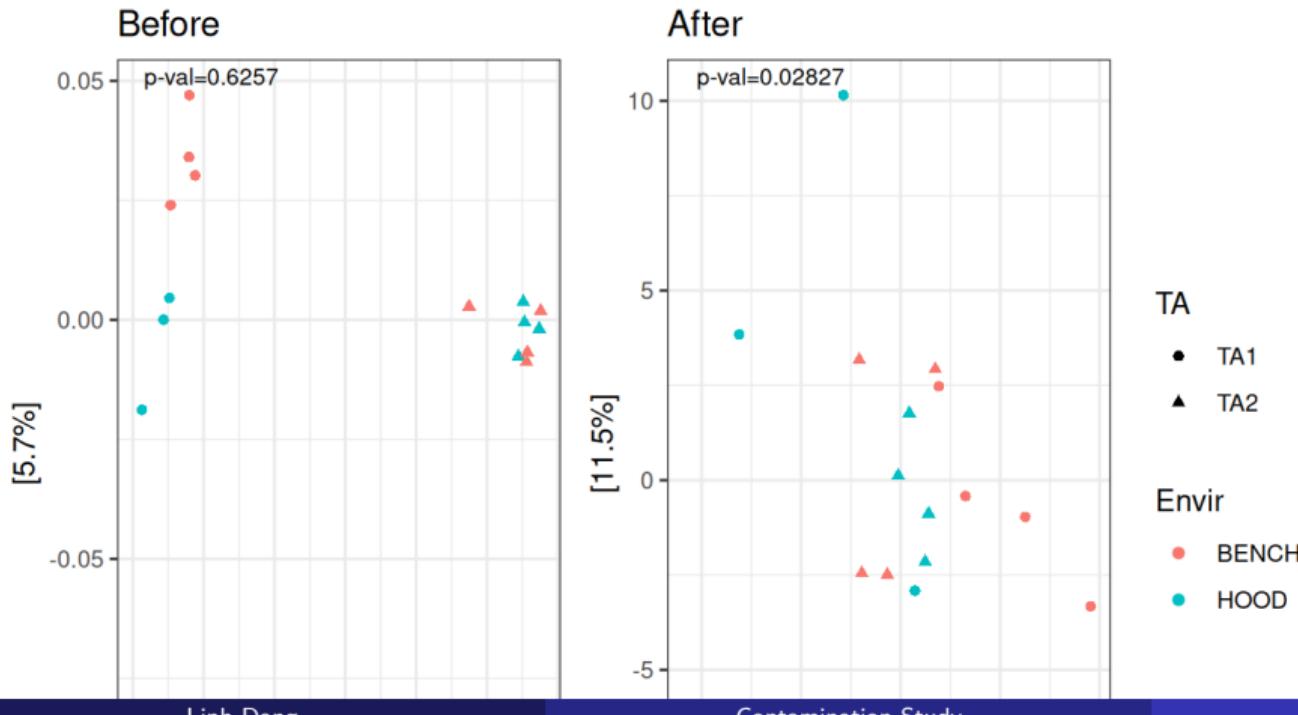


## Between environments after accounting for technicians



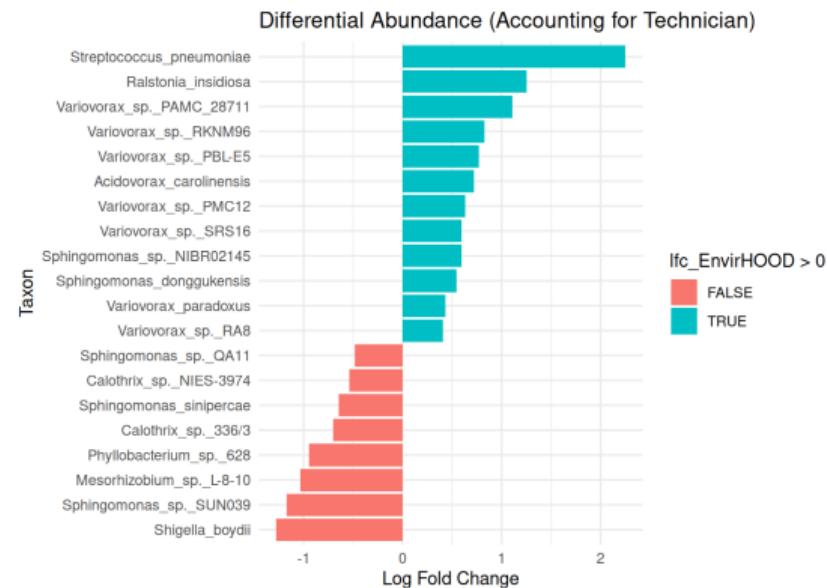
# Hood vs. Bench - Beta Diversity

NCT samples before and after accounting for the effect of technician by constrains permutation.

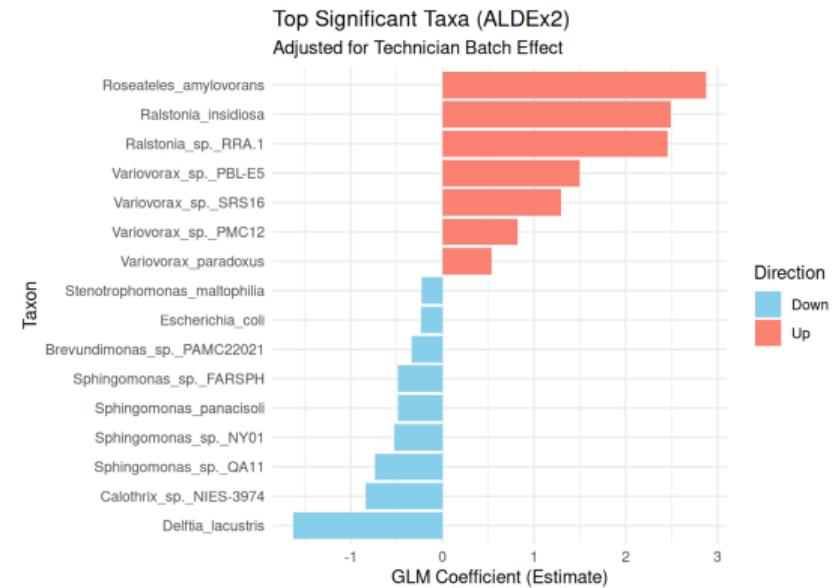


# Hood vs. Bench - DAA

By ANCOM-BC2 with p-val 0.1



By ALDEx2 with p-val 0.2



## Approaches:

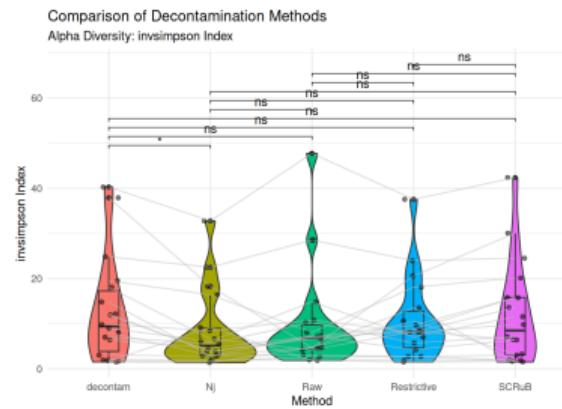
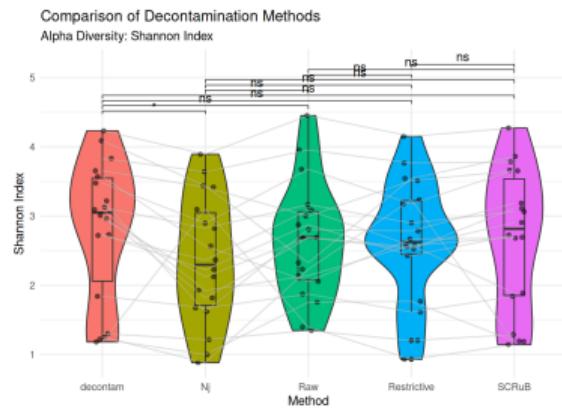
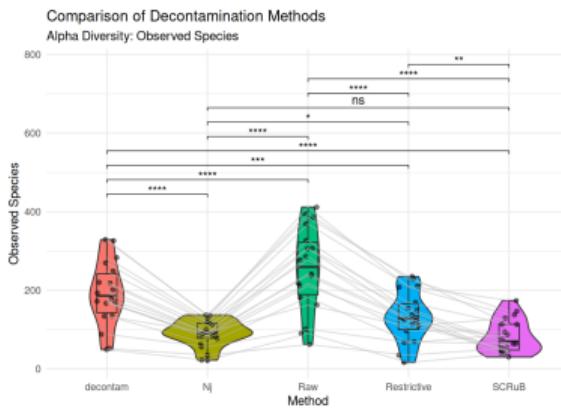
- Raw (baseline)
- Restrictive
- decontam (R package)
- SCRuB
- Nejman approach.

## Metrics for evaluation:

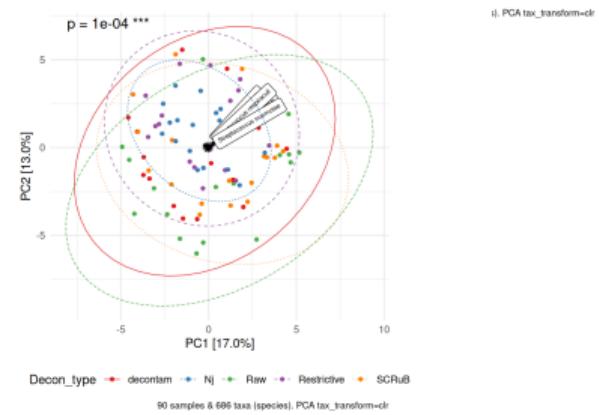
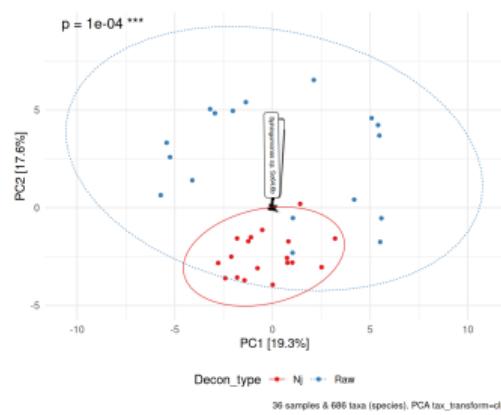
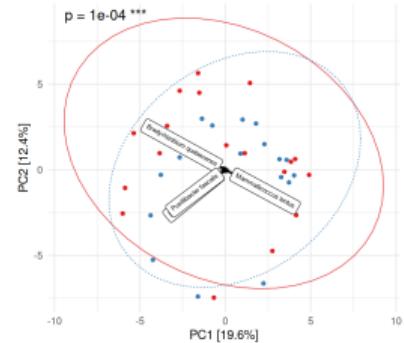
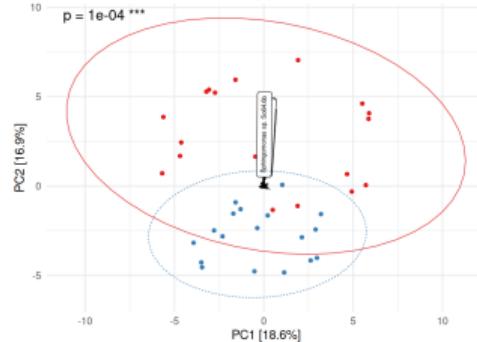
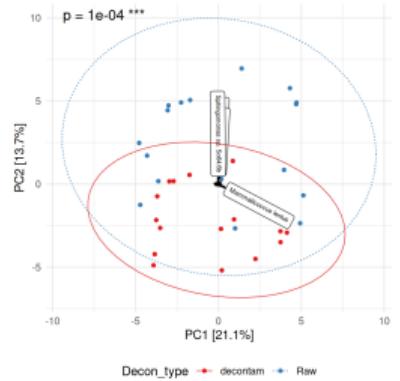
- Assessment of batch effect removal,
- Divergence from Negative Controls,
- Biologically meaningful of Consensus taxa,
- Overlap with known true taxa / contaminants

# Are decontamination approaches different?

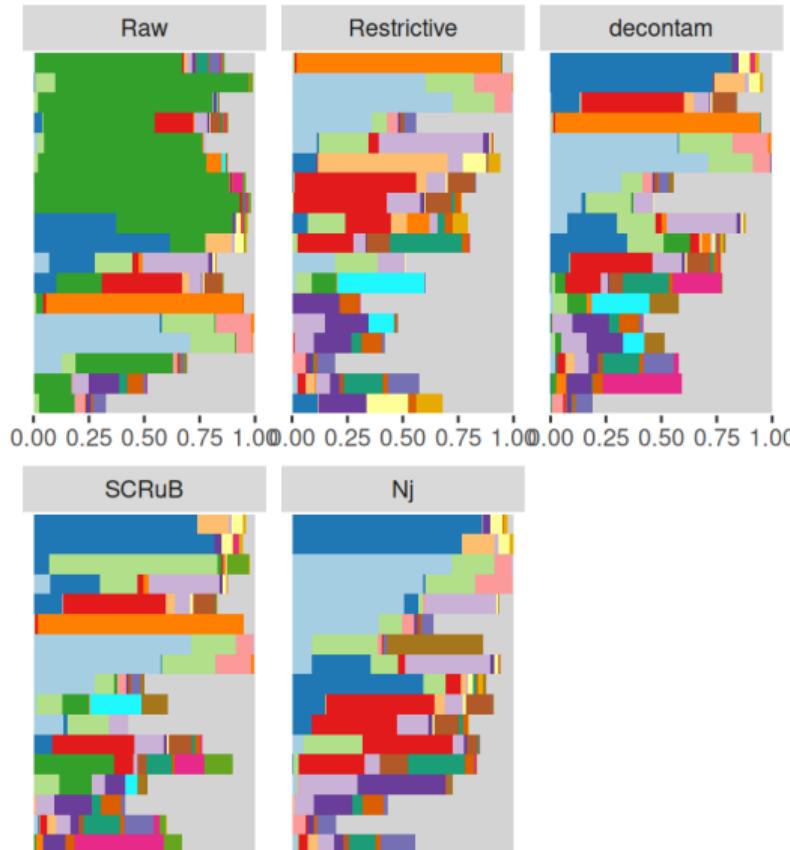
## Alpha Diversities



# Beta Diversity



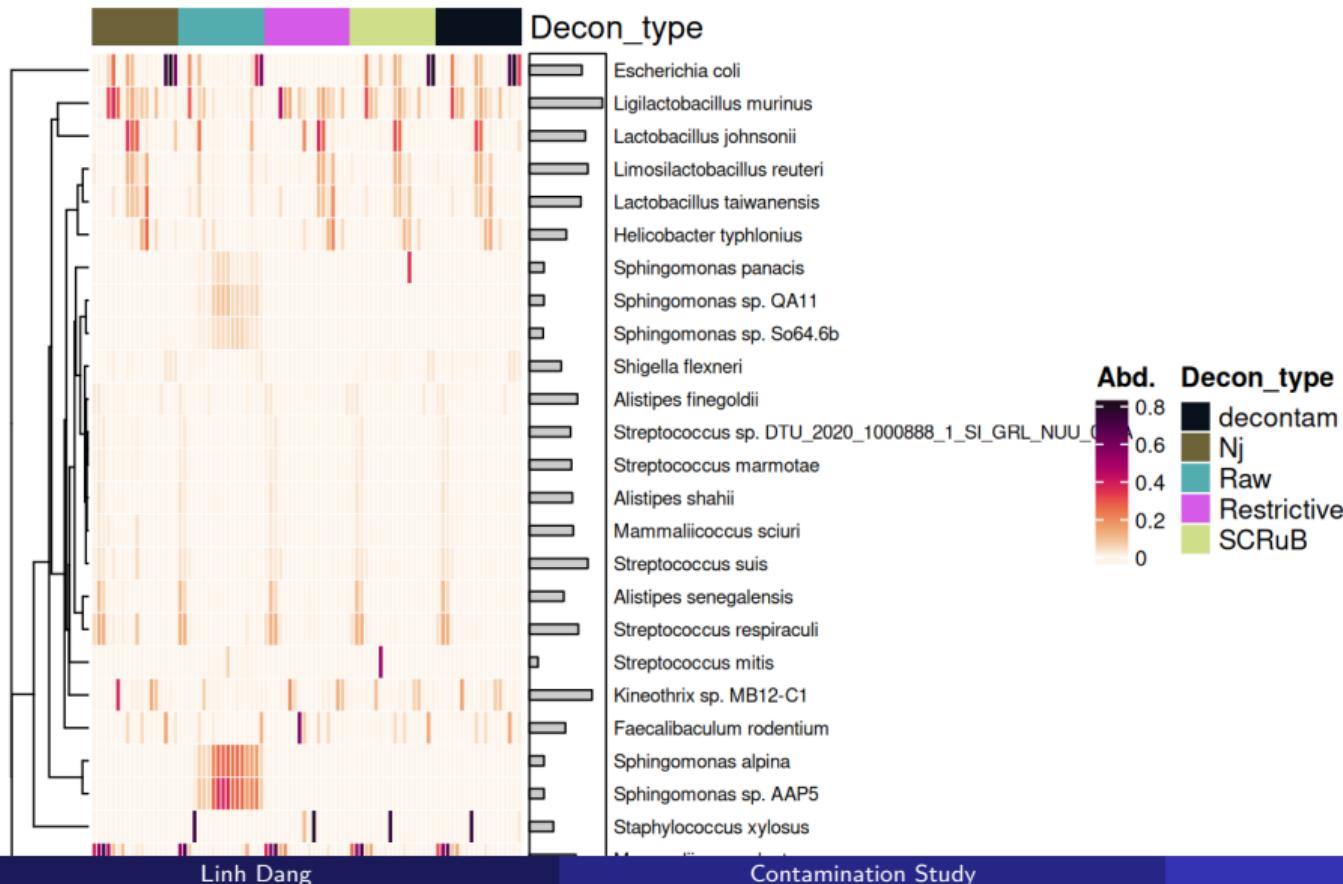
# Bacterial Profile - Barplot



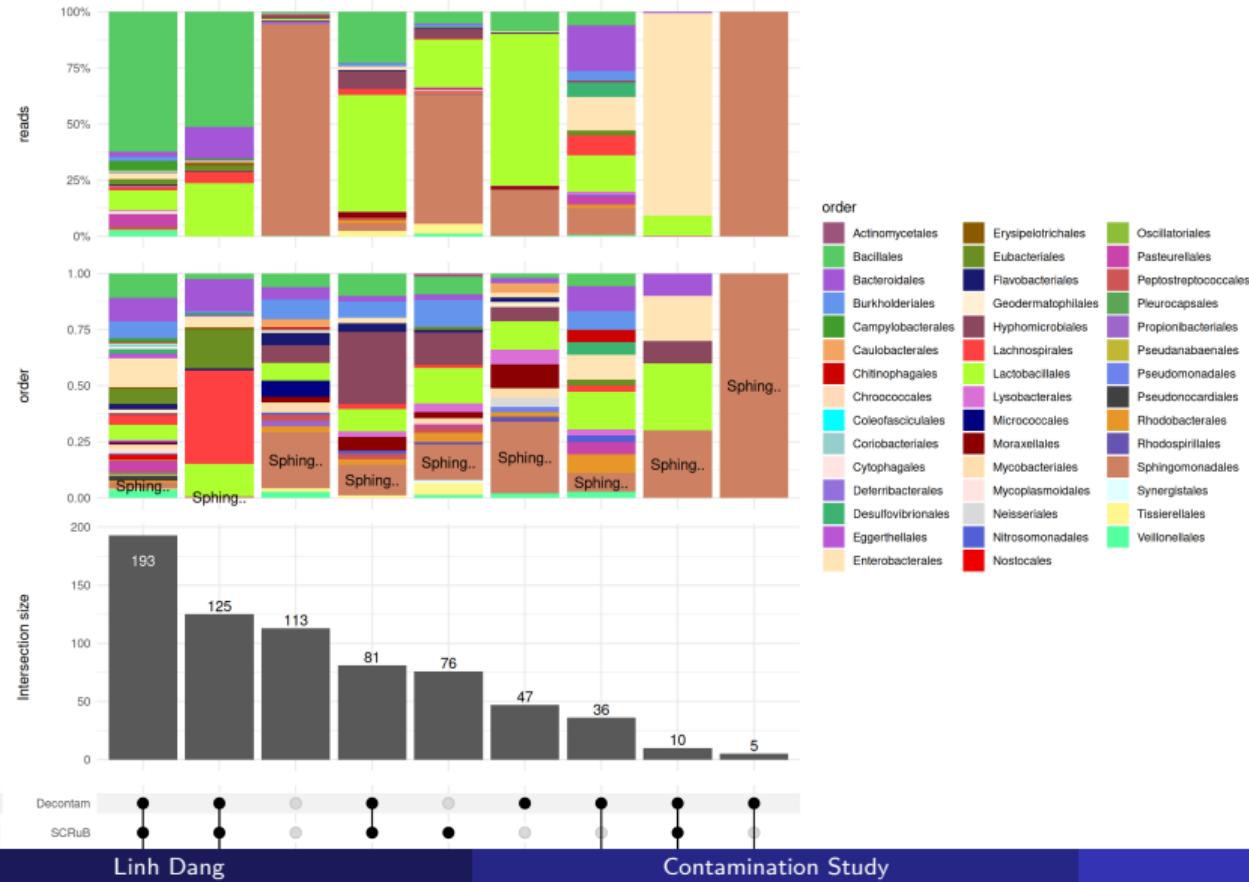
## genus

Mammaliicoccus	Limosilactobacillus
Escherichia	Dialister
Streptococcus	Helicobacter
Sphingomonas	Simiaoa
Alistipes	Duncaniella
Lactobacillus	Bradyrhizobium
Faecalibaculum	Sphingosinicella
Staphylococcus	Salmonella
Ligilactobacillus	Sphingobium
Kineothrix	Other
Shigella	

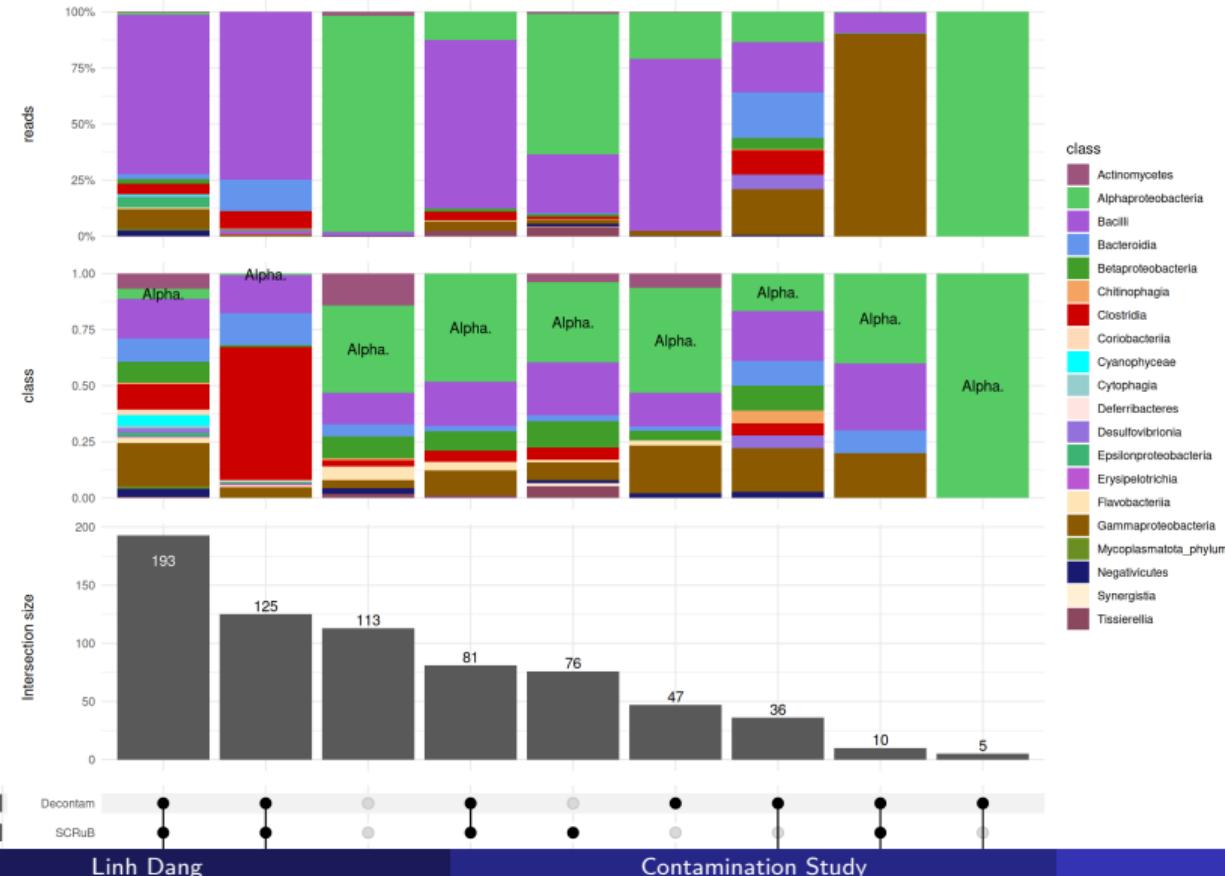
# Bacterial Profile - Heatmap



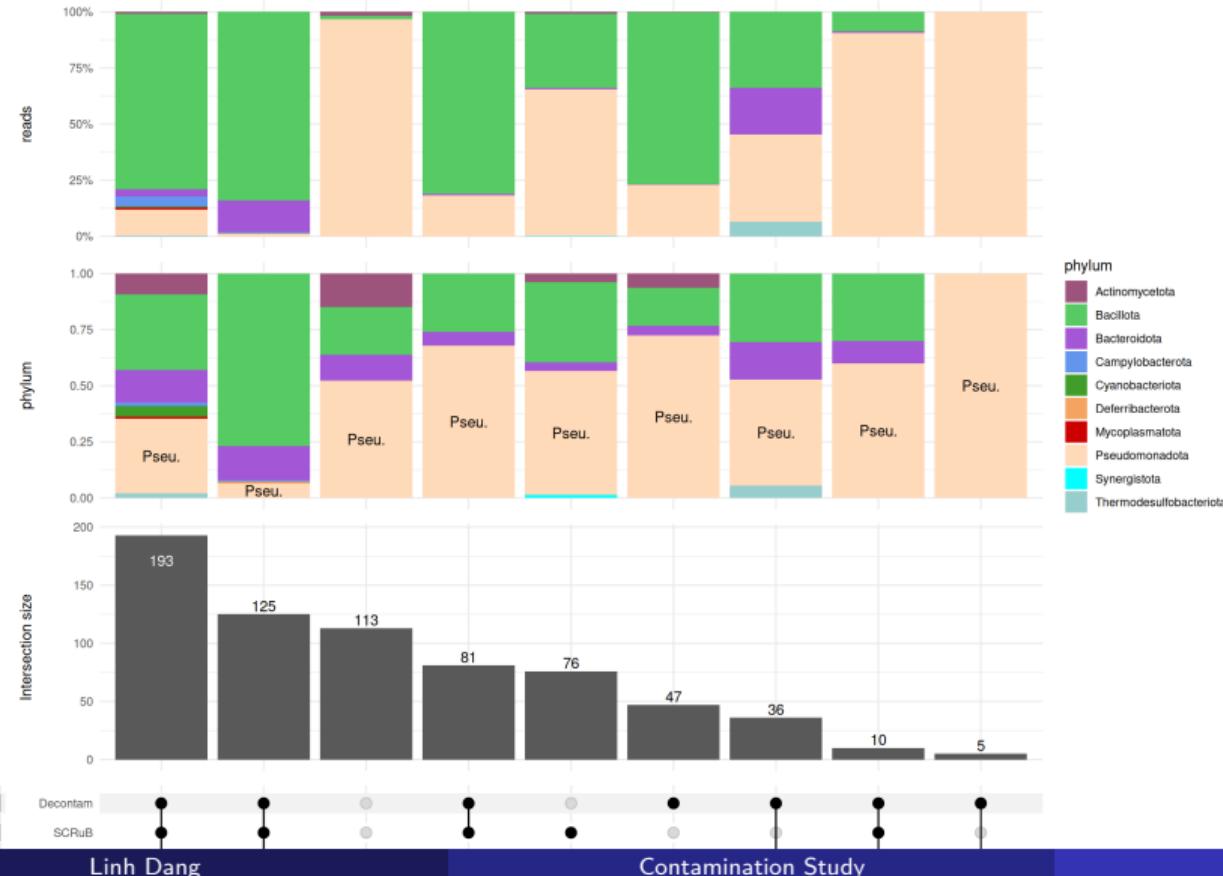
# Overlap Taxa among Approaches - Order



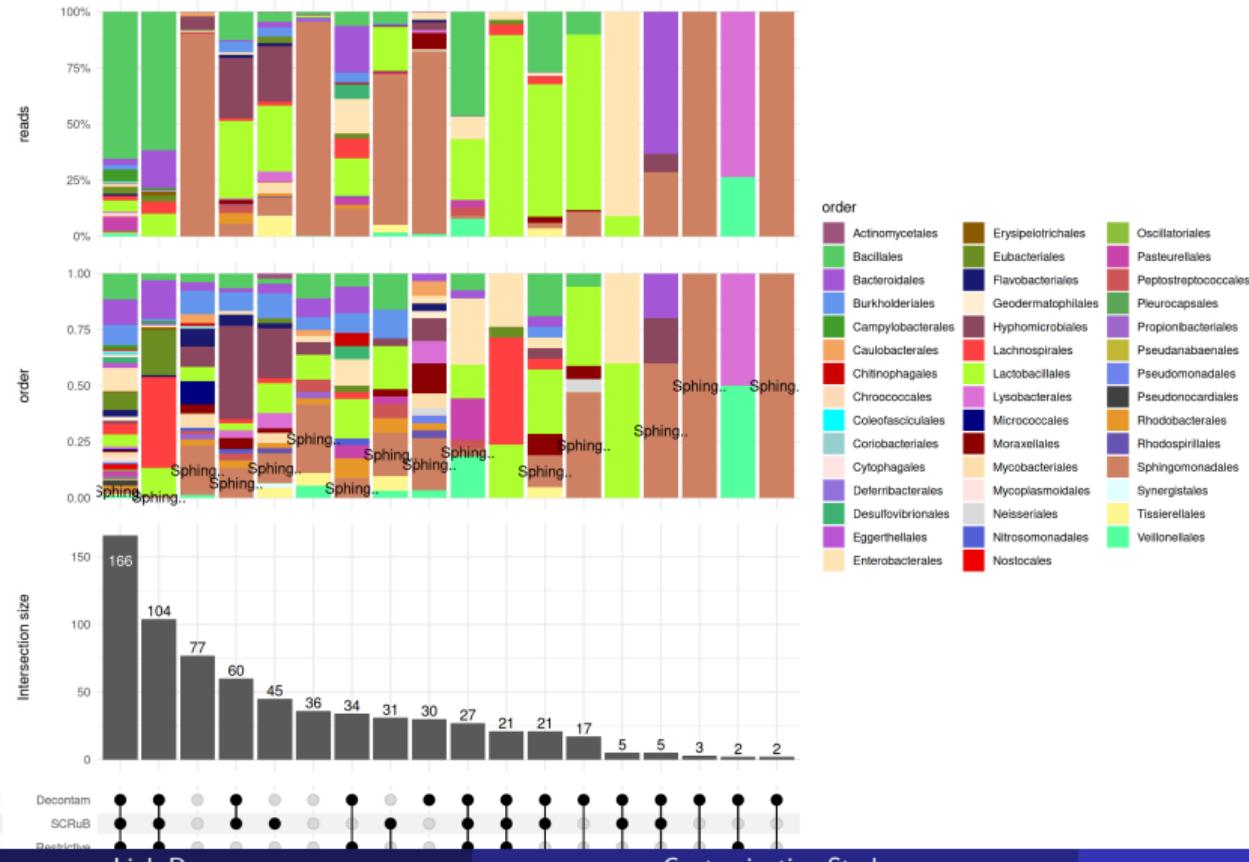
# Overlap Taxa among Approaches - Class



# Overlap Taxa among Approaches - Phylum



# Overlap with Contaminants in NCT (prev 20%)



# Quantify Assessment of Decontamination Approaches

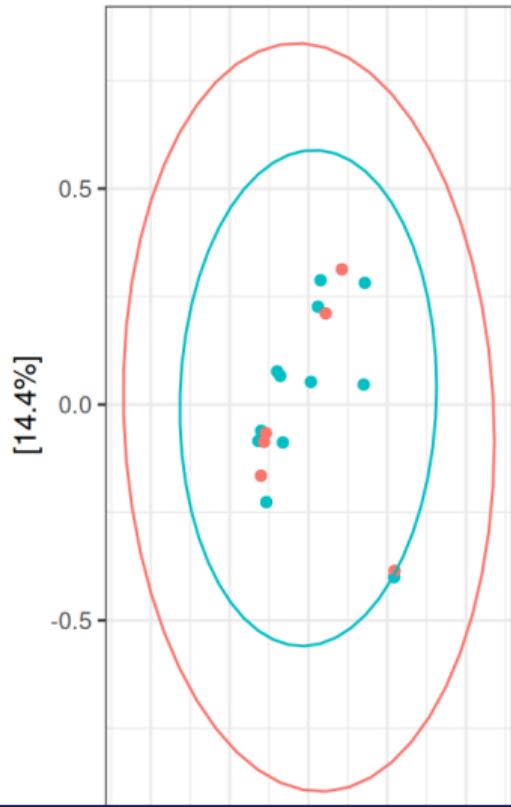
- **Assessment of batch effect removal**
- Divergence from Negative Controls
- Biologically meaningful of Consensus taxa
- Overlap with known true taxa / contaminants

# Assessment of batch effect removal

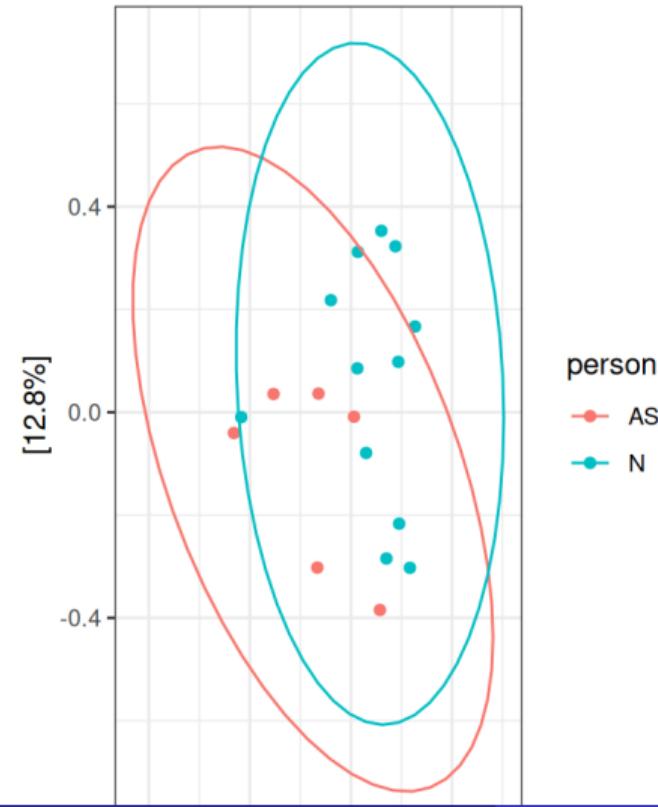
- **Main idea:** A good decontamination method should reduce technical variation. In this case is the technician.
- **Approach:** Perform Ordination (PCoA) and PERMANOVA (adonis2).
- **Metric:** Check if the variance explained ( $R^2$ ) by “Sequencing Run” or “Extraction Batch” decreases after decontamination.
- **Logic:** Real tumor microbiomes shouldn’t cluster by which technician sequenced.

# Assessment of batch effect removal - Restrictive

Raw - Before Decontamination

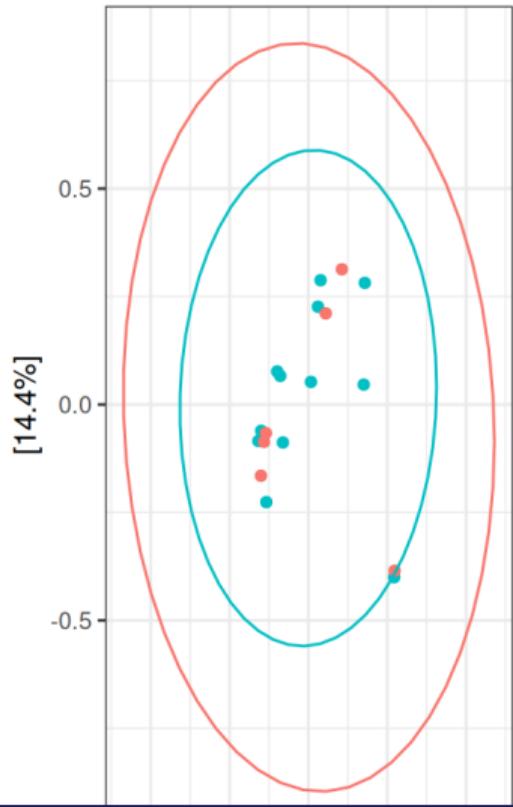


After Decon. - Restrictive

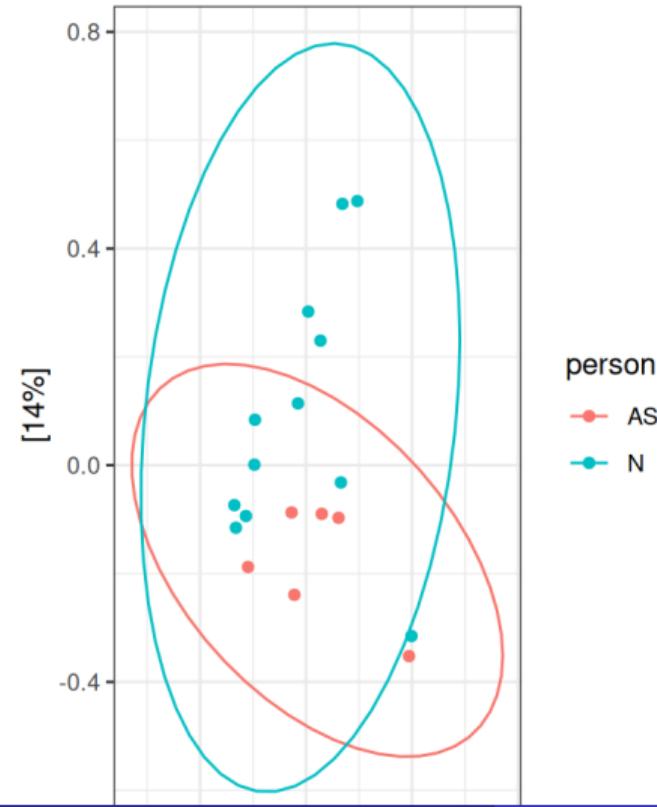


# Assessment of batch effect removal - Decontam

Raw - Before Decontamination

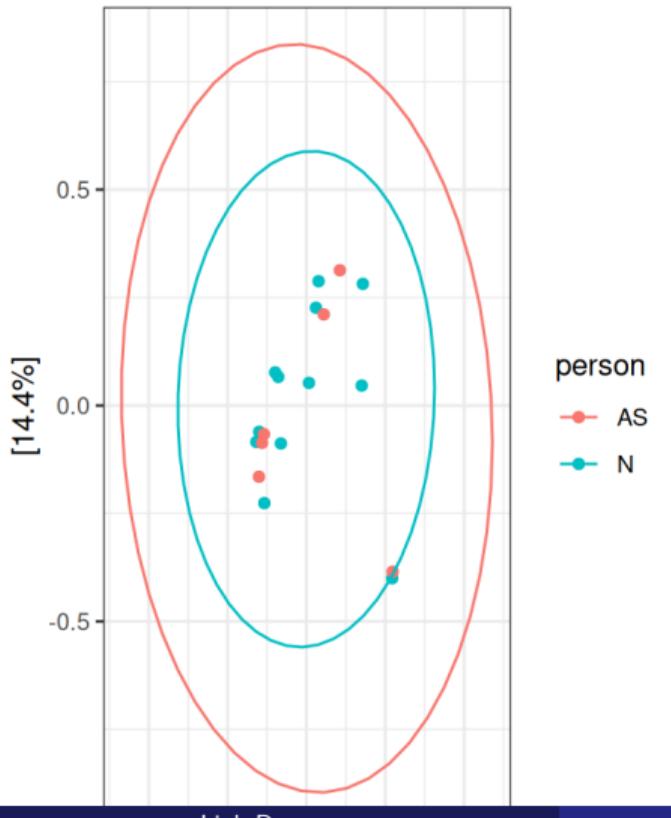


After Decon. - Decontam

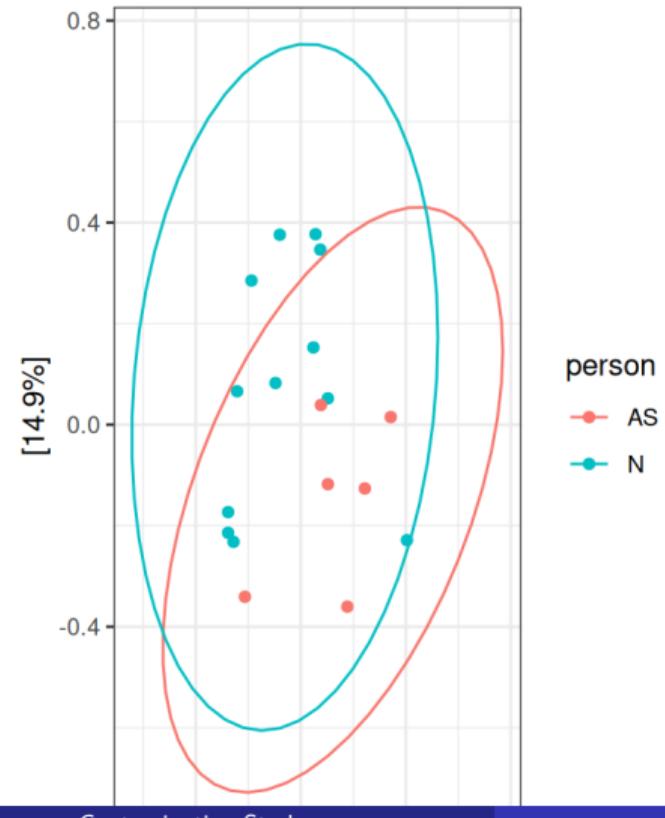


# Assessment of batch effect removal - Nejman

Raw - Before Decontamination



After Decon. - Nejman



# Assessment of batch effect removal - Overall

	R2	p-val
<i>Raw</i>	7.53%	0.1776
<i>Restrictive</i>	9.74%	0.0195
<i>Decontam</i>	8.88%	0.0352
<i>SCRuB</i>	9.2%	0.0291
<i>Nejman</i>	9.17%	0.0486

## Remarks:

- Before decontamination: samples are not significantly clustered w.r.t technicians/years.
- After decontamination: samples are significantly clustered w.r.t technicians/years.

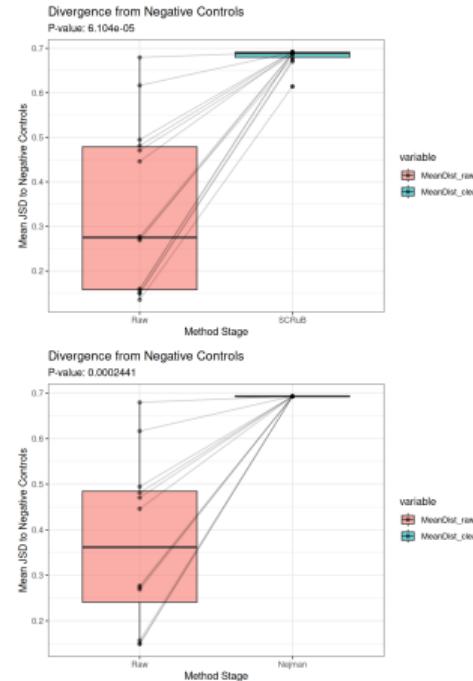
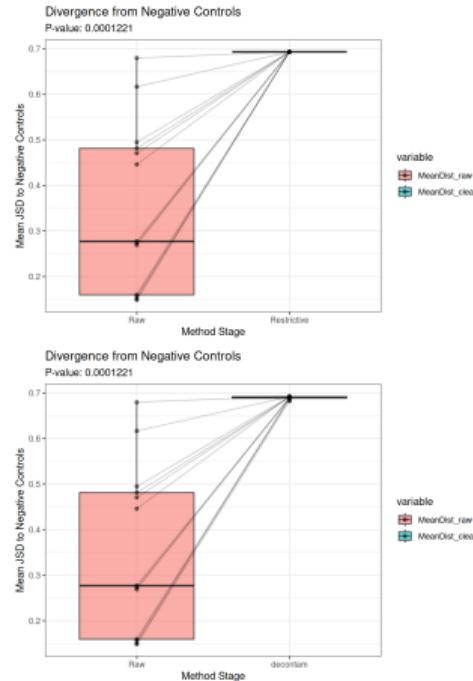
# Quantify Assessment of Decontamination Approaches

- Assessment of batch effect removal
- **Divergence from Negative Controls**
- Biologically meaningful of Consensus taxa
- Overlap with known true taxa / contaminants

# Divergence from Negative Controls

- **Approach:** perform Inter-group dissimilarity and Wilcox paired test, following by visualization.
- **Metric:** Calculate the Jensen-Shannon Divergence (or Bray-Curtis) between samples and the pool of negative controls.
- **Logic:** the decontamination method should maximize the dissimilarity between true PDAC samples and NCTs.

# Divergence from Negative Controls - Plots



## Remarks

- Restrictive method: remove all taxa belongs to negative controls. Thus the taxa profile of decontaminated samples and negative controls are always different.
- All methods seem not to be differentiated.

# Quantify Assessment of Decontamination Approaches

- Assessment of batch effect removal
- Divergence from Negative Controls
- Biologically meaningful of Consensus taxa
- **Overlap with known true taxa / contaminants**

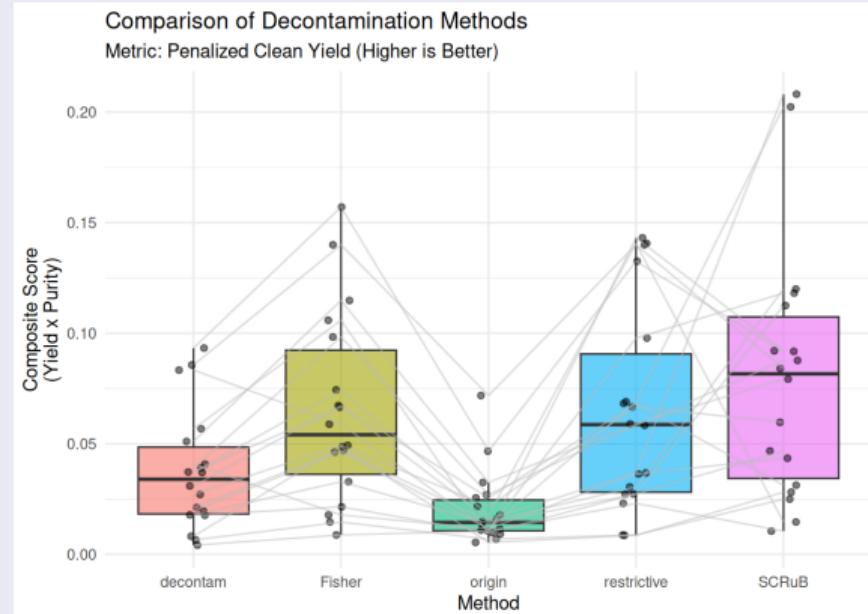
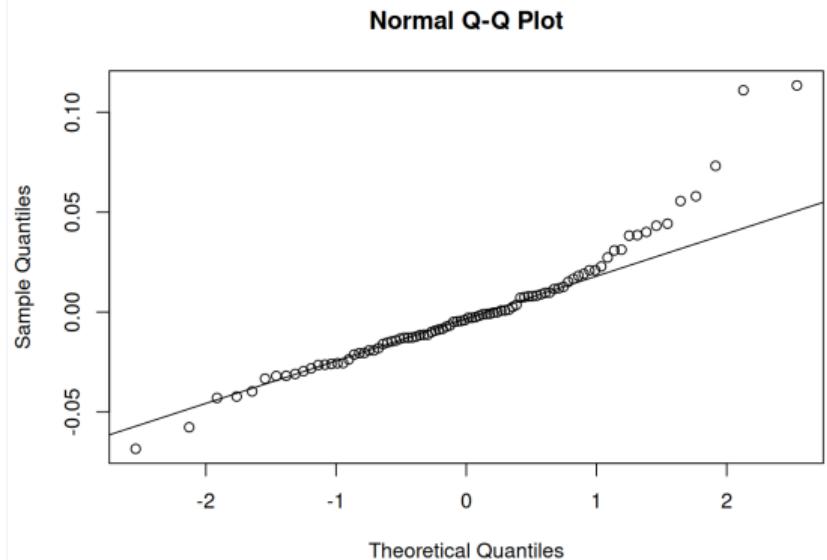
## Overlap with known true taxa / contaminants

- **Approach:** design a measure quantifying the quality of decontamination approach by balancing the yield and purity explained below.
- **Metric:**
  - *Yield* → number of putative significant true taxa not in NCTs / number of observed species.
  - *Purity* → number of putative significant true taxa not in NCTs / whole number of putative contaminants.
- **Logic:** a good decontamination methods should give results whose taxa is less overlap with known high prevalence contaminants in large set of NCTs, while not over-removing true taxa.

Putative significant true taxa: by prevalence, by abundance, combination.

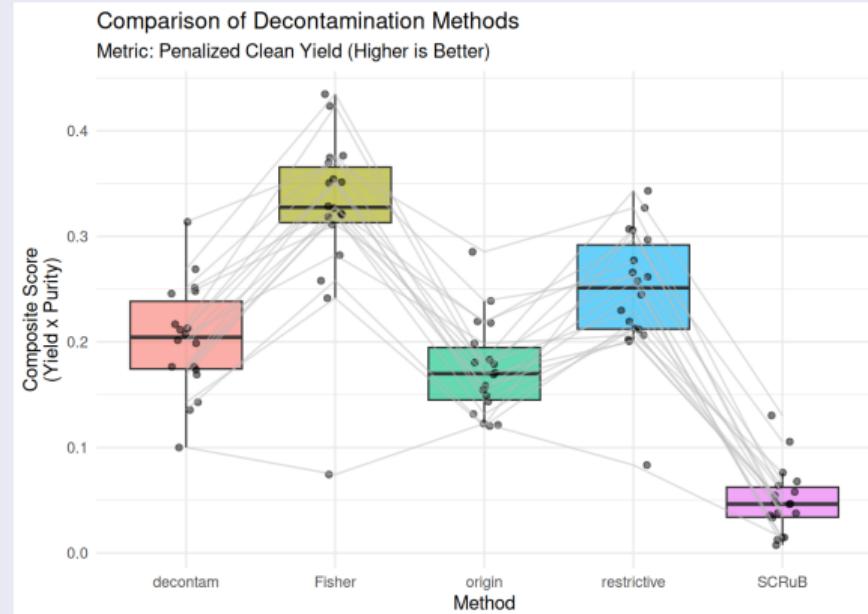
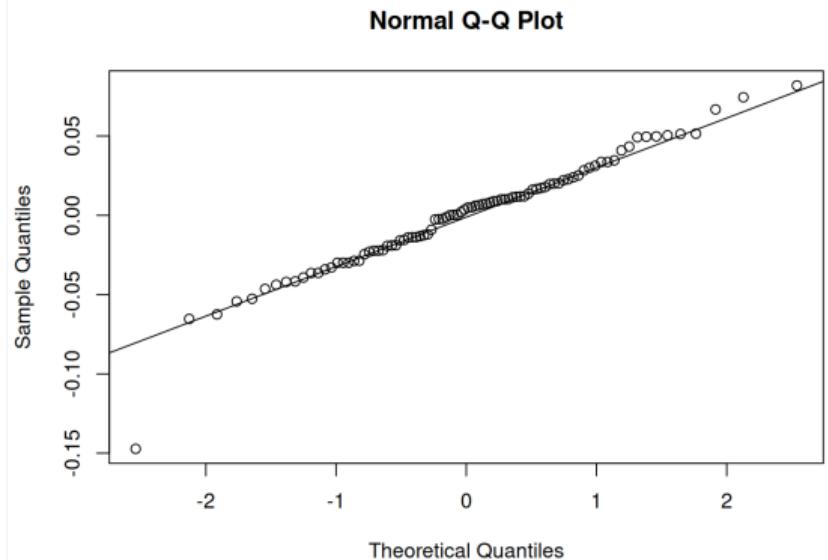
# Abundance-based Approach

## Check normality assumption



# Prevalence-based Approach

## Check normality assumption



# Combination of Abundance and Prevalence

Check normality assumption

