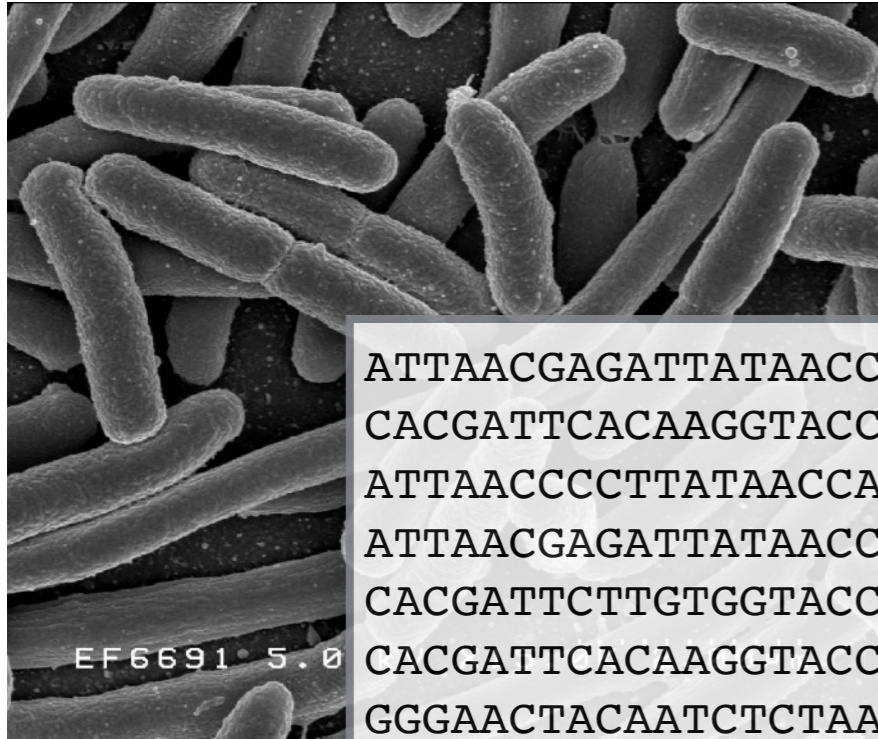


Contamination, controls and  
**accurate** sequencing-based **measurement**  
of microbial communities

# A Microbial Census



```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGA ACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGA
CACGATTCACAAGGTACCACA
ATTAACGAGATTATAACCAGA
```

<i>Lactobacillus crispatus</i>	1300	5	0	882	596
<i>Ureaplasma urealytica</i>	15	0	220	0	0
<i>Gardnerella vaginalis</i>	22	0	1	0	412
<i>Prevotella intermedia</i>	0	0	8	12	0
...	...	...	...	...	...

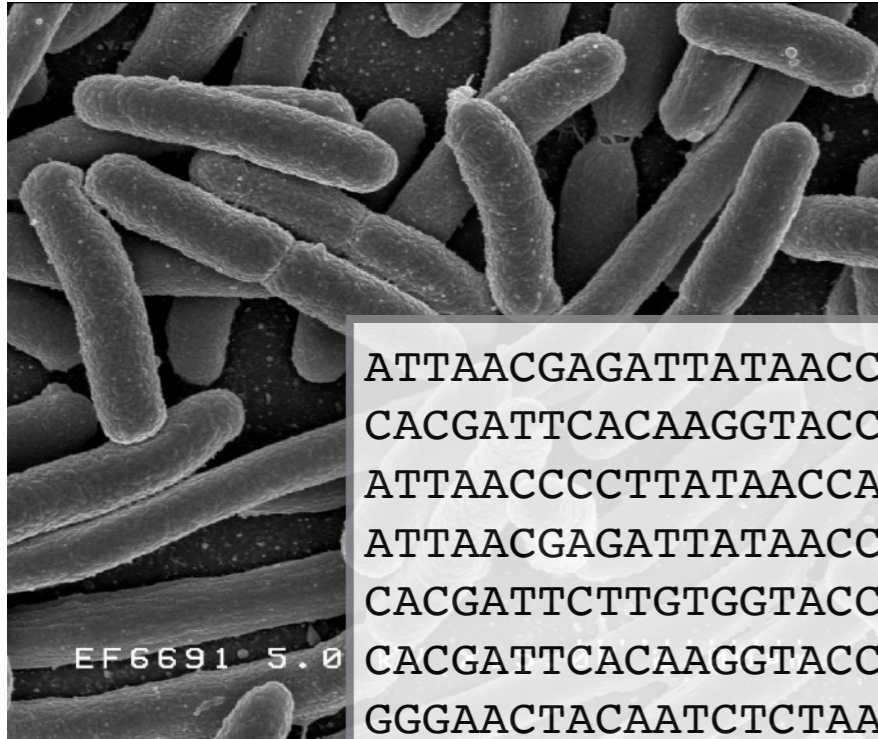
→ Inference

↓ Visualization

↘ Exploration

# A Microbial Census

## Marker-gene or Metagenomics Sequencing (MGS)



```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGA ACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGA
CACGATTCACAAGGTACCACA
ATTAACGAGATTATAACCAGA
```

<i>Lactobacillus crispatus</i>	1300	5	0	882	596
<i>Ureaplasma urealytica</i>	15	0	220	0	0
<i>Gardnerella vaginalis</i>	22	0	1	0	412
<i>Prevotella intermedia</i>	0	0	8	12	0
...	...	...	...	...	...

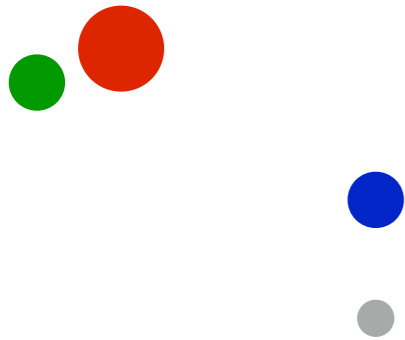
→ Inference

↓ Visualization

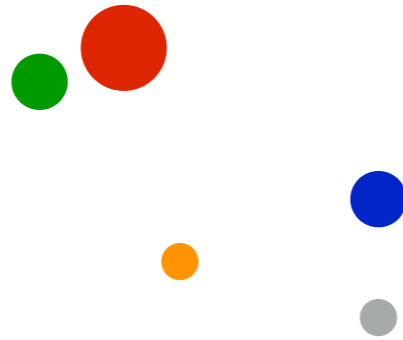
↘ Exploration

# MGS: What is really there?

**Sample  
Sequences**

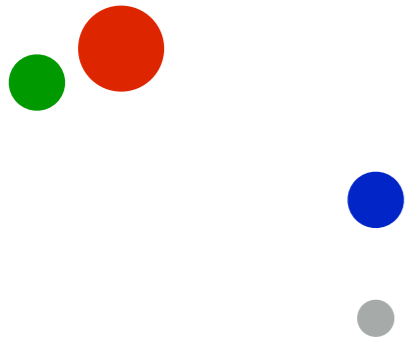


**Extraction**

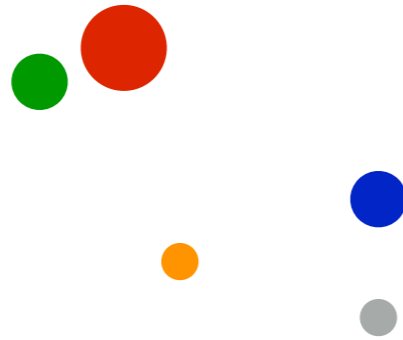


# MGS: What is really there?

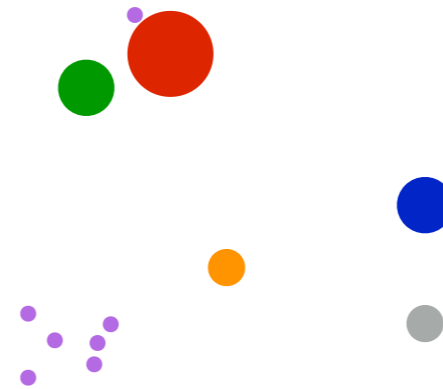
**Sample  
Sequences**



**Extraction**

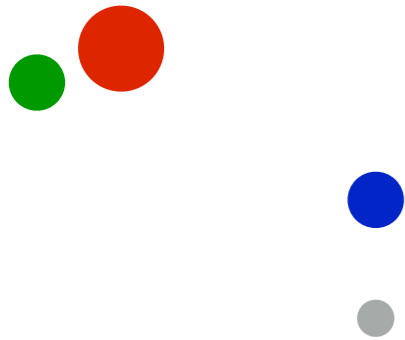


**Library Prep**

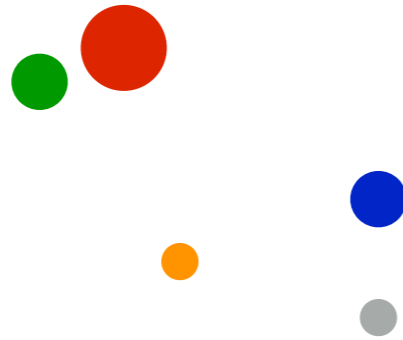


# MGS: What is really there?

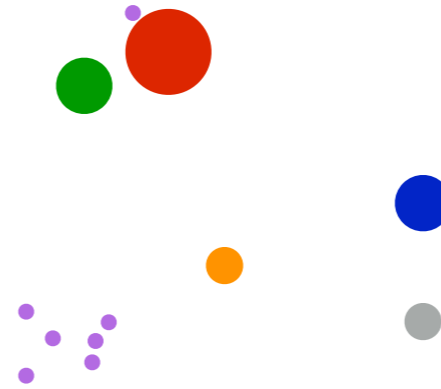
Sample  
Sequences



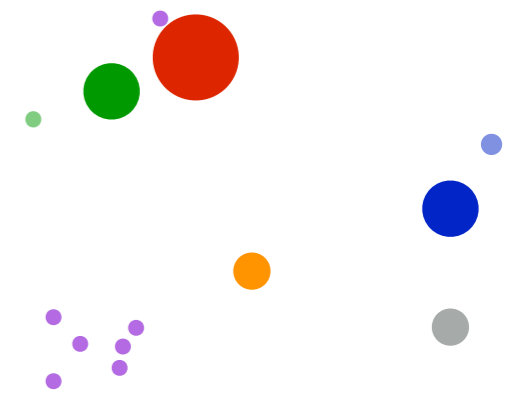
Extraction



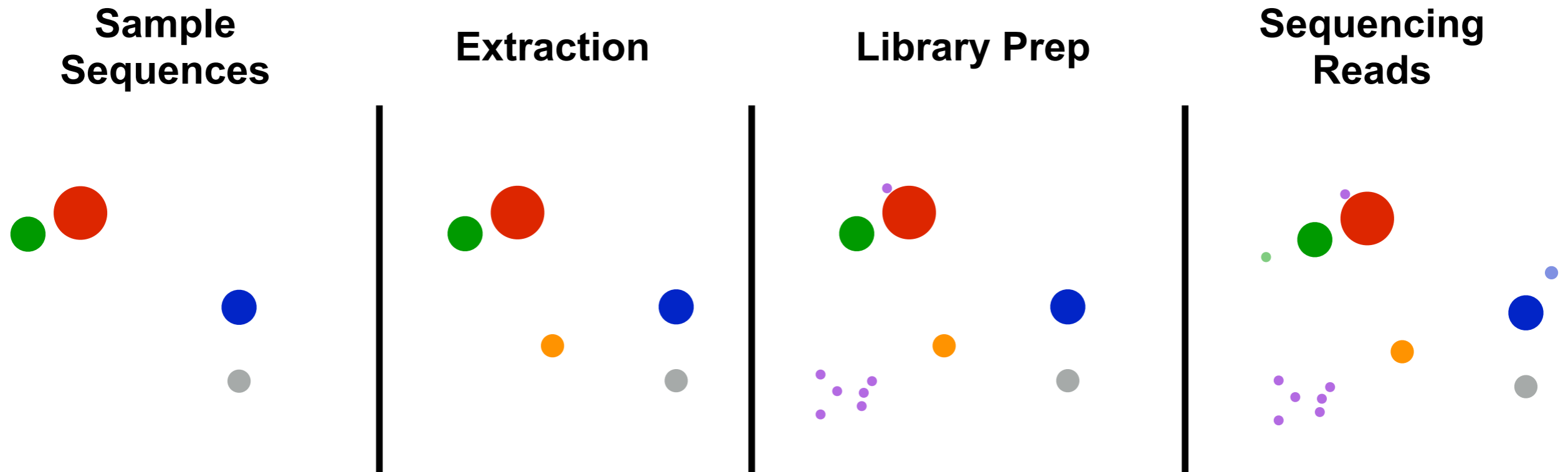
Library Prep



Sequencing  
Reads

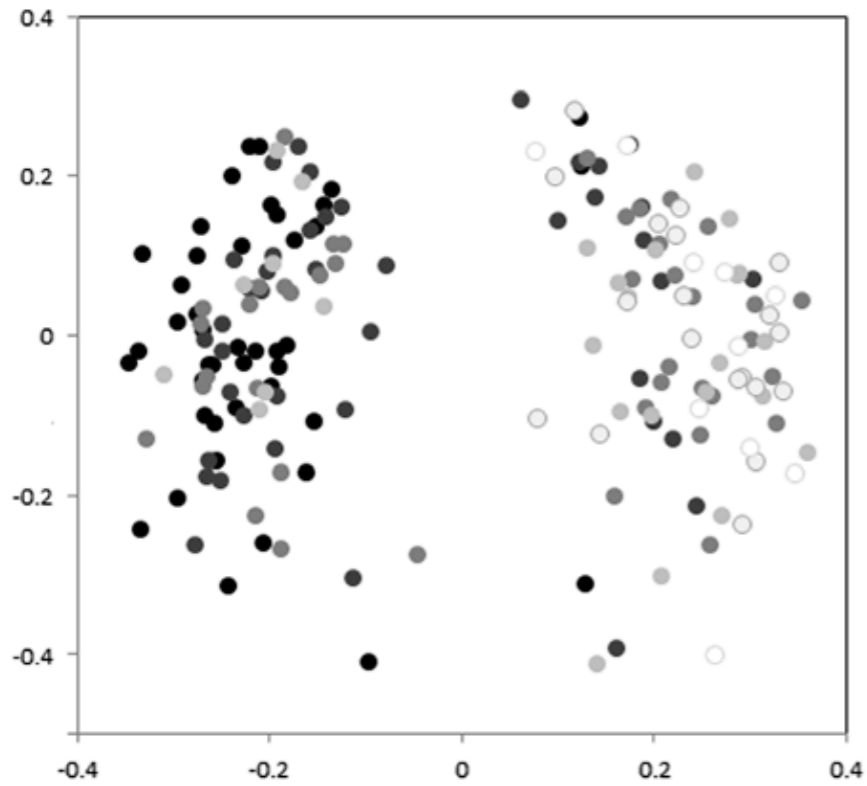


# Problem: Contamination



**Contaminants** — DNA sequences from organisms not truly present in the sample.

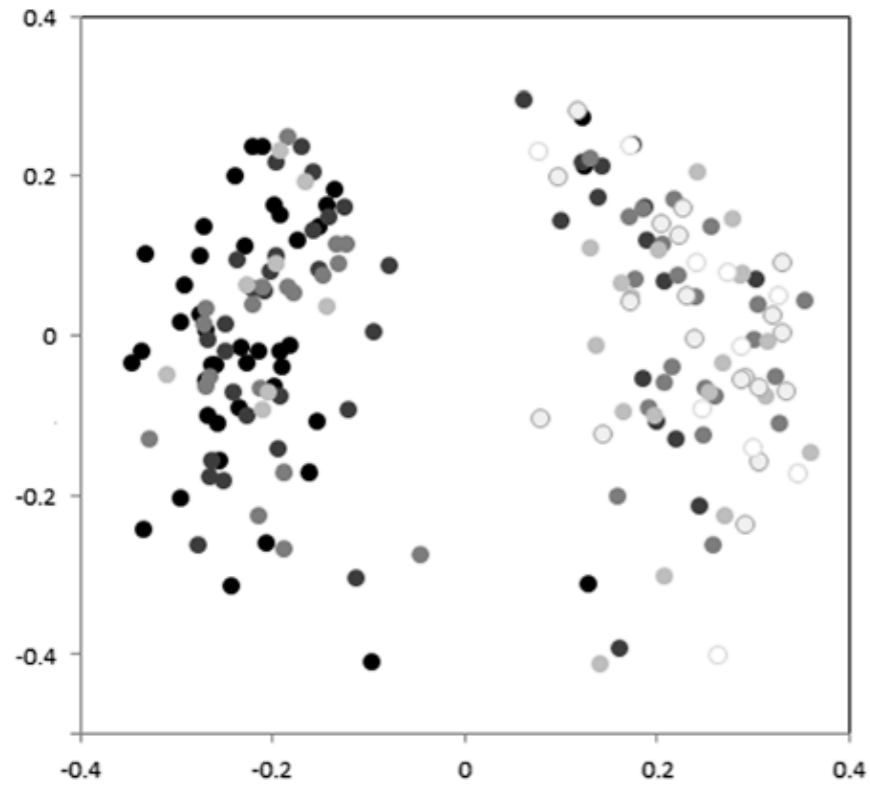
# Problem: Contamination



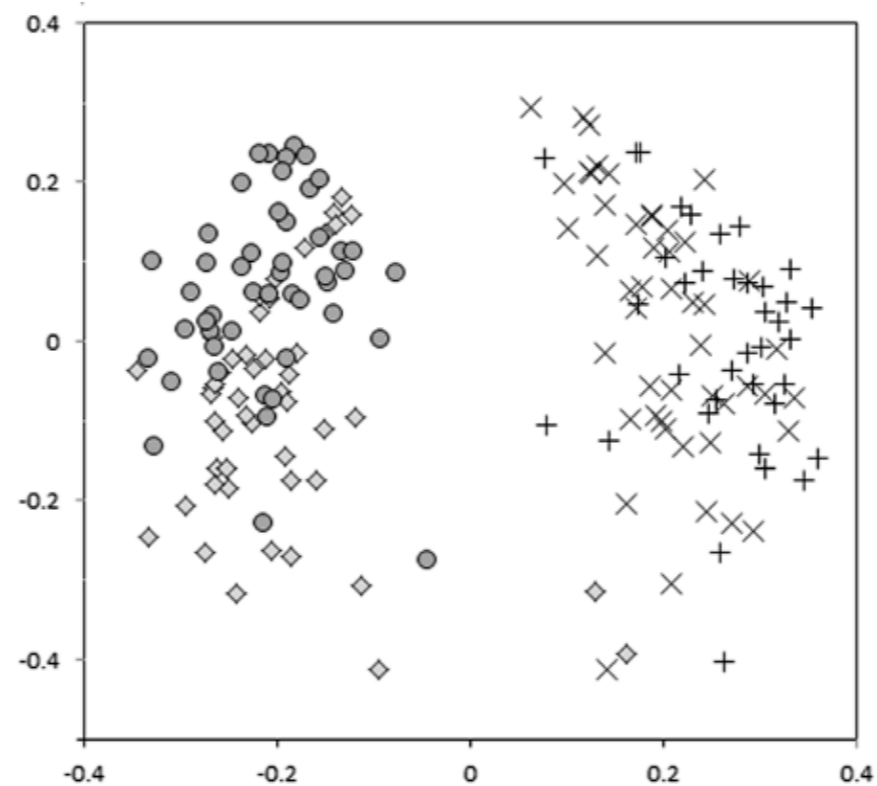
a) Full data coloured by age (months)



# Problem: Contamination

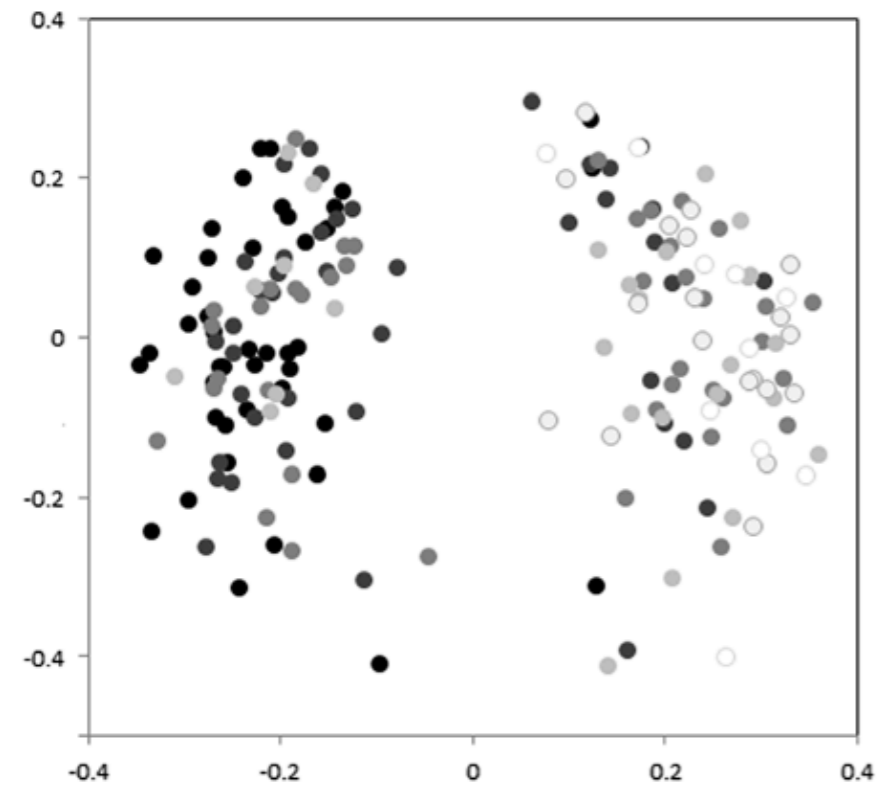


a) Full data coloured by age (months)

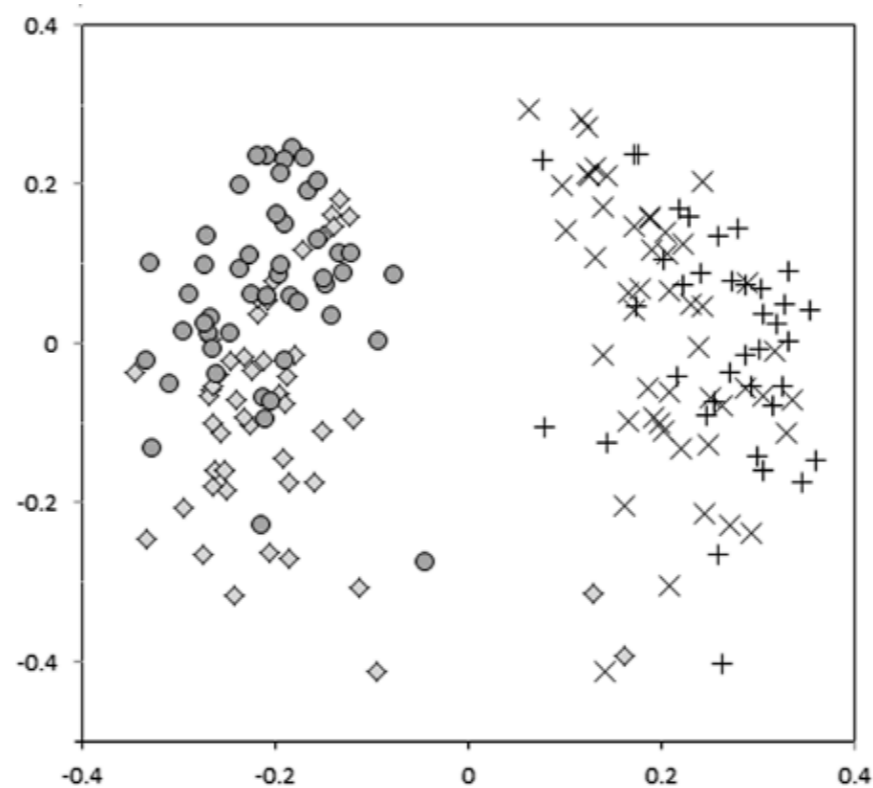


b) Full data shaped by extraction kit

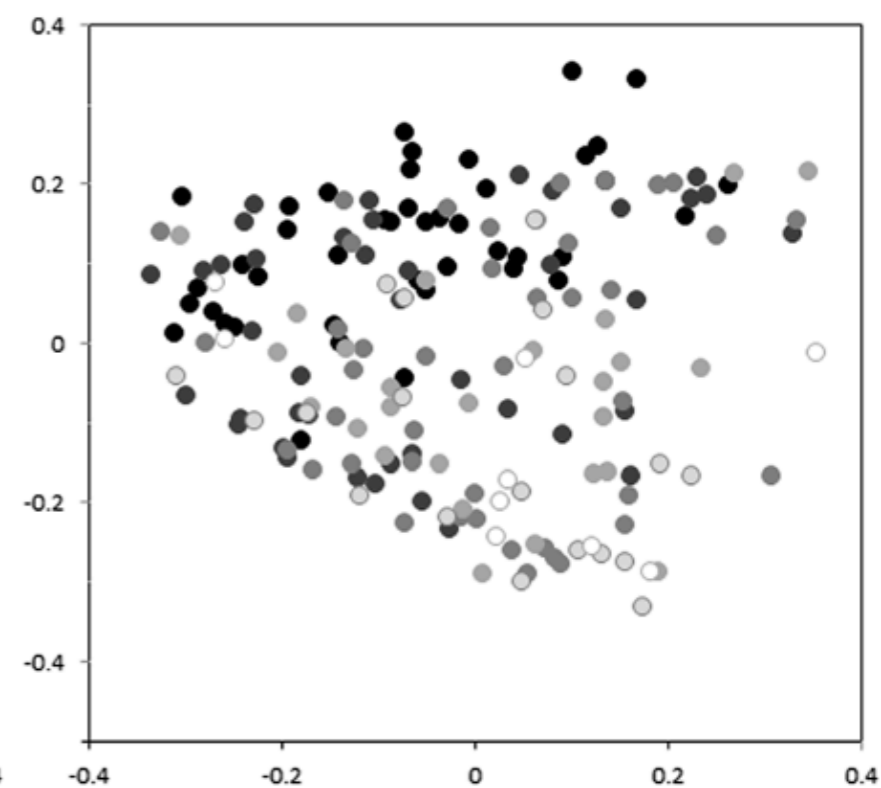
# Problem: Contamination



a) Full data coloured by age (months)



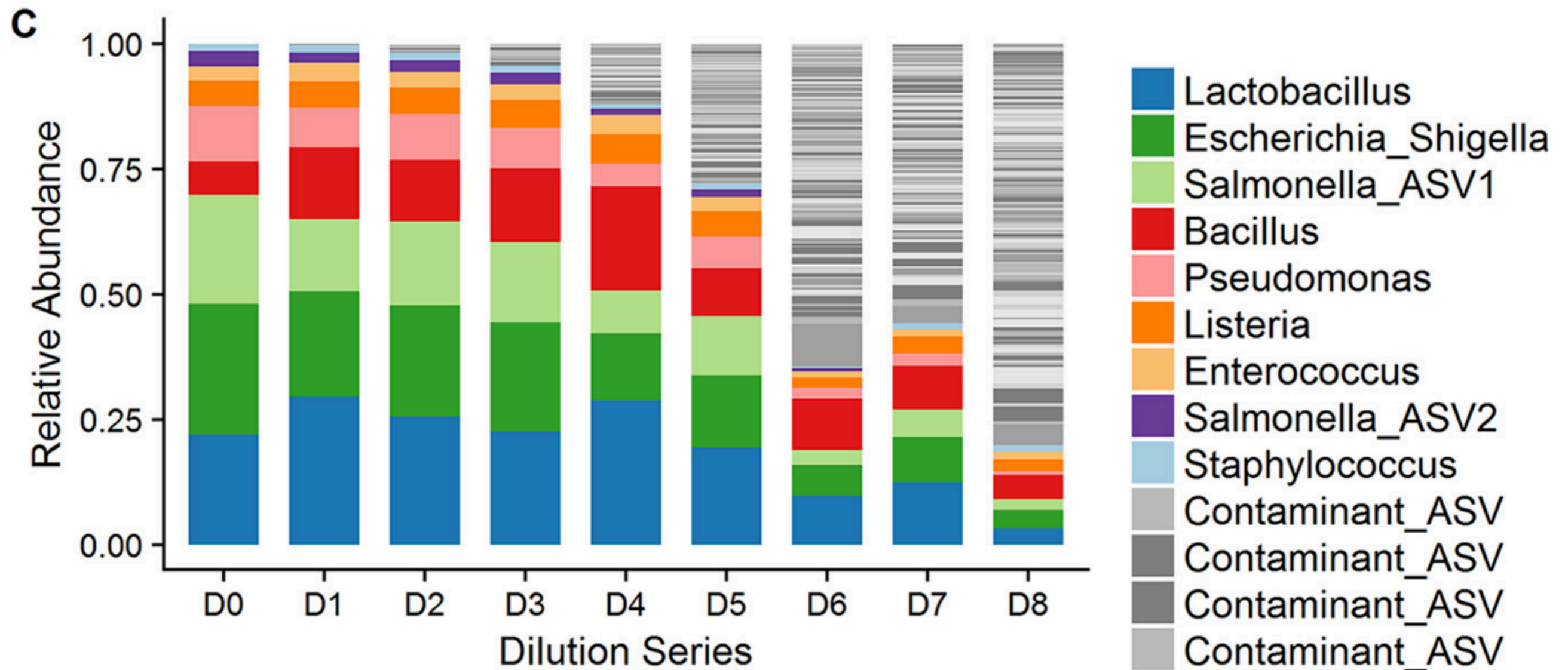
b) Full data shaped by extraction kit



c) Contaminant OTUs removed

**Spurious signal driven by contaminants!**

# Problem: Contamination





**Figure:** Karstens, et al. mSystems, 2018.

# Problem: Contamination

Article | [Published: 31 July 2019](#)

## Human placenta has no microbiome but can contain potential pathogens

[Marcus C. de Goffau](#), [Susanne Lager](#), [Ulla Sovio](#), [Francesca Gaccioli](#), [Emma Cook](#), [Sharon J. Peacock](#), [Julian Parkhill](#) , [D. Stephen Charnock-Jones](#) & [Gordon C. S. Smith](#) 

[Nature](#) **572**, 329–334 (2019) | [Cite this article](#)



**27k** Accesses | **326** Citations | **643** Altmetric

EDITORIAL | [VOLUME 220, ISSUE 3, P213-214, MARCH 01, 2019](#)

## De-Discovery of the Placenta Microbiome

[Frederic D. Bushman, PhD](#)  

## Lack of detection of a human placenta microbiome in samples from preterm and term deliveries

[Jacob S. Leiby](#), [Kevin McCormick](#), [Scott Sherrill-Mix](#), [Erik L. Clarke](#), [Lyanna R. Kessler](#), [Louis J. Taylor](#), [Casey E. Hofstaedter](#), [Aoife M. Roche](#), [Lisa M. Mattei](#), [Kyle Bittinger](#), [Michal A. Elovitz](#), [Rita Leite](#), [Samuel Parry](#)  & [Frederic D. Bushman](#) 

[Microbiome](#) **6**, Article number: 196 (2018) | [Cite this article](#)

**8898** Accesses | **143** Citations | **110** Altmetric | [Metrics](#)

# Modeling Contaminants

**T = S + C**, where **C** is constant

*hence*

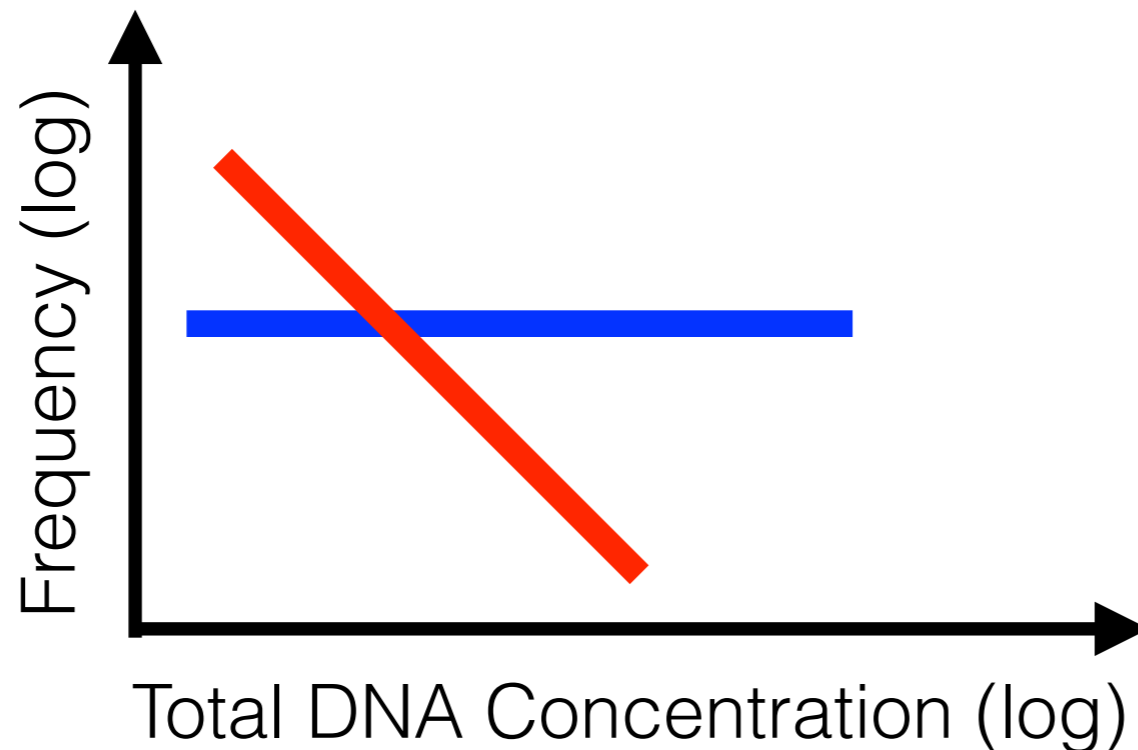
**f<sub>c</sub> = C/(S+C) ~ 1/T**, where **C ≪ S**

# Modeling Contaminants

$T = S + C$ , where  $C$  is constant

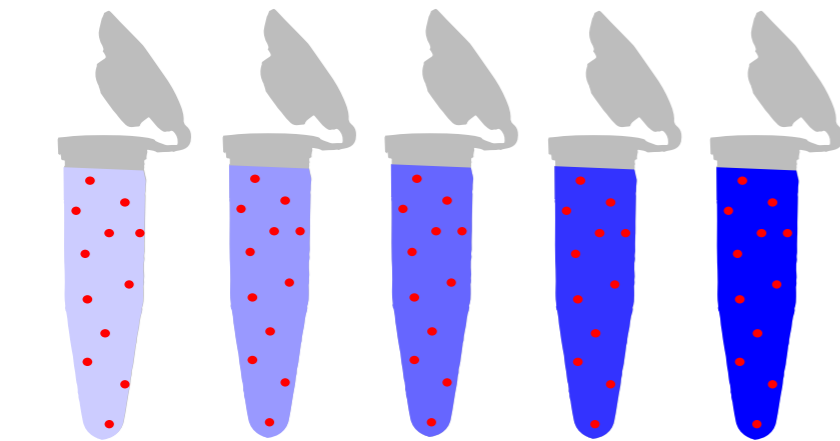
*hence*

$f_c = C/(S+C) \sim 1/T$ , where  $C \ll S$




**Sample Sequence**  
**Contaminant**

# Modeling Contaminants

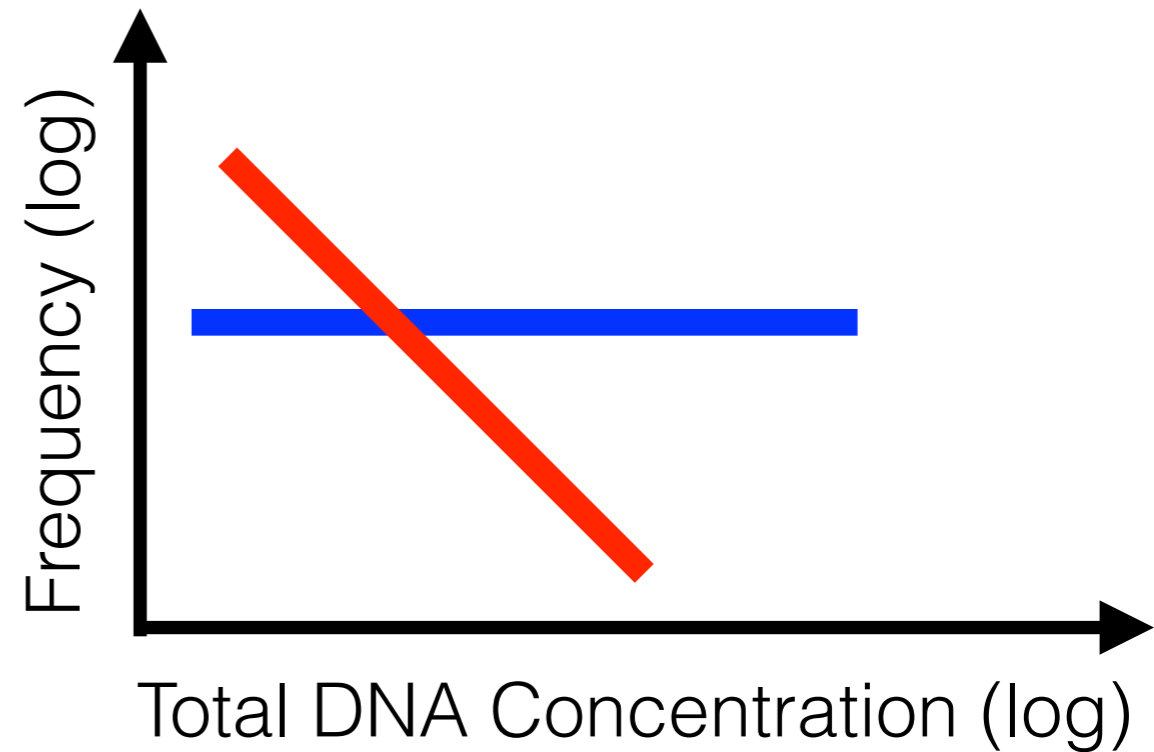


sample [DNA]

 equal, low-level  
contaminating DNA



sequence  
equimolar  
amounts  
well-mixed  
total DNA



**Sample Sequence**  
**Contaminant**

# Decontam Method

## Frequency

Input: DNA concentrations,  
Feature table w/ abundances.

Output: Score 0 (contaminant) - 1 (non-contaminant),  
Binary classification based on threshold.



# Decontam Method

## Frequency

Input: DNA concentrations,  
Feature table w/ abundances.

Output: Score 0 (contaminant) - 1 (non-contaminant),  
Binary classification based on threshold.

## Prevalence

Input: Categorization of samples as negative controls,  
Feature table w/ abundances or presences.

Output: Score 0 (contaminant) - 1 (non-contaminant)  
Binary classification based on threshold.

# Decontam Method

## Frequency

*Needs range of DNA concentrations*

Input: DNA concentrations,  
Feature table w/ abundances.

Output: Score 0 (contaminant) - 1 (non-contaminant),  
Binary classification based on threshold.

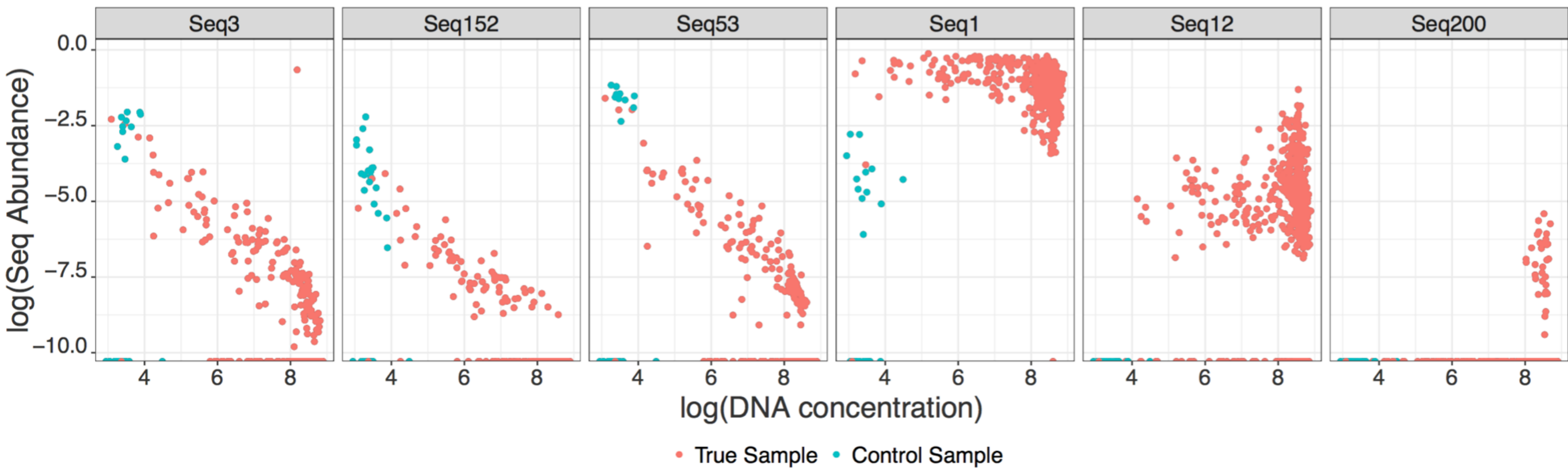
## Prevalence

*Needs multiple (5+) sequenced negative controls*

Input: Categorization of samples as negative controls,  
Feature table w/ abundances or presences.

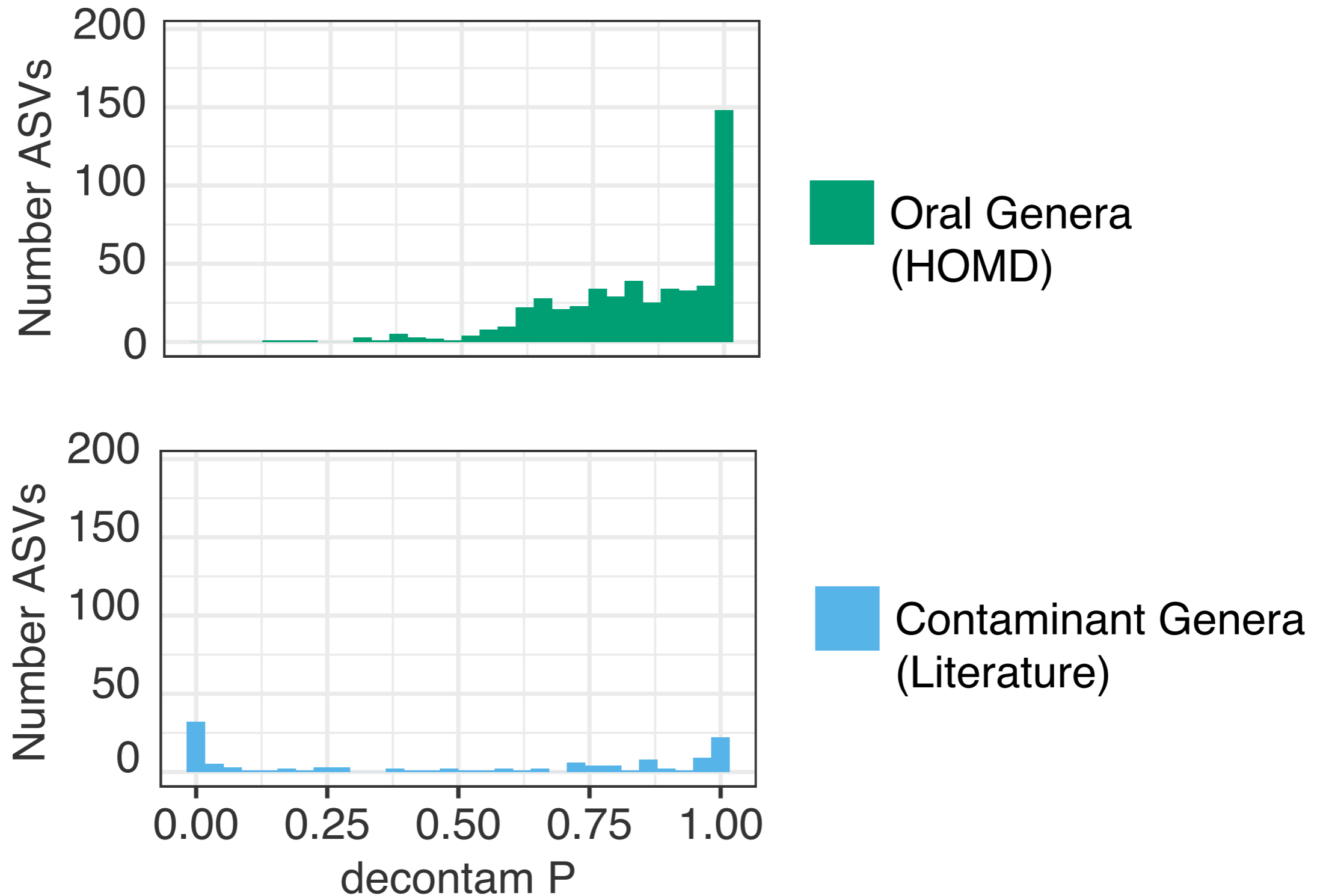
Output: Score 0 (contaminant) - 1 (non-contaminant)  
Binary classification based on threshold.

# Validating the Model



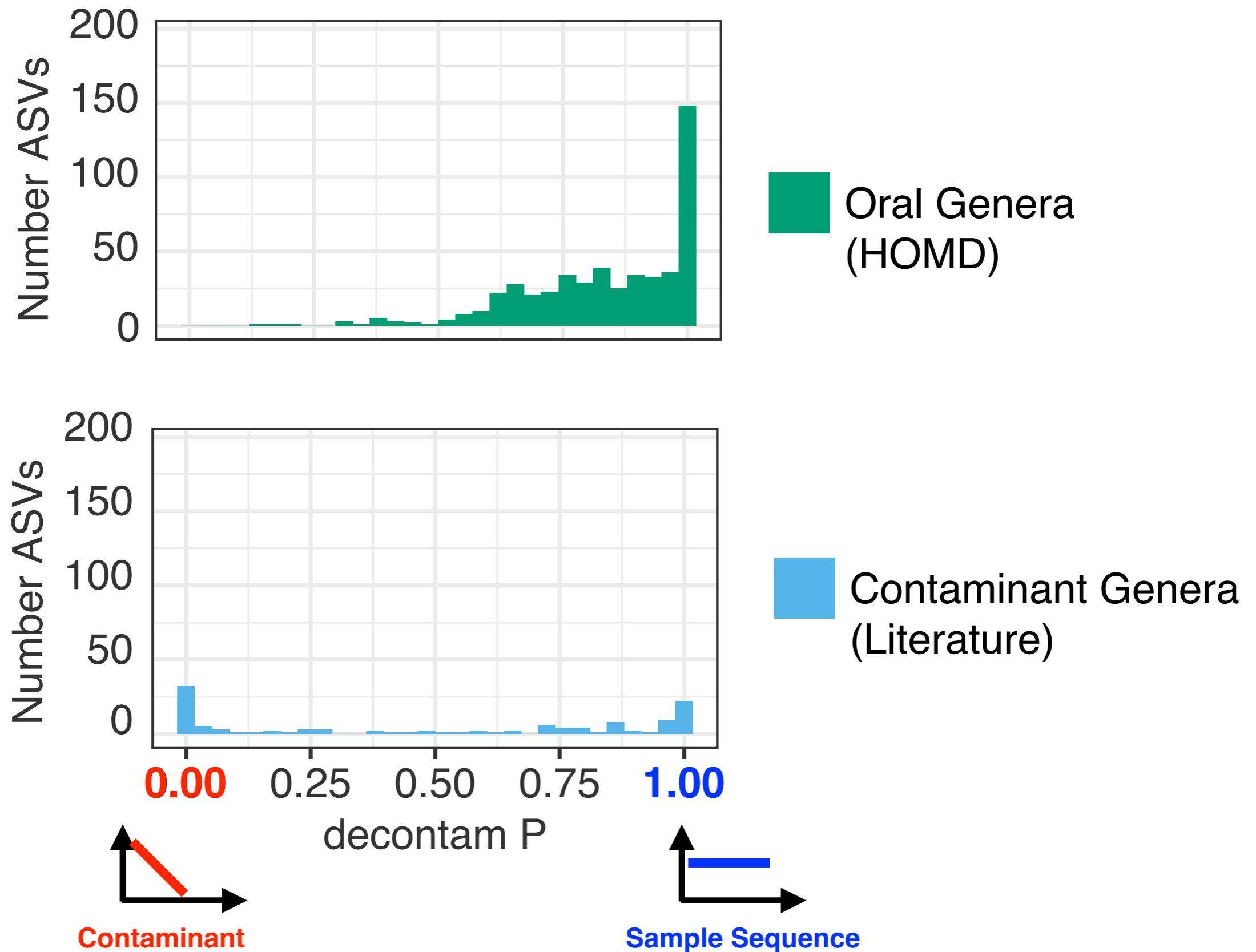
# Validating the Model

## Oral Mucosal Dataset



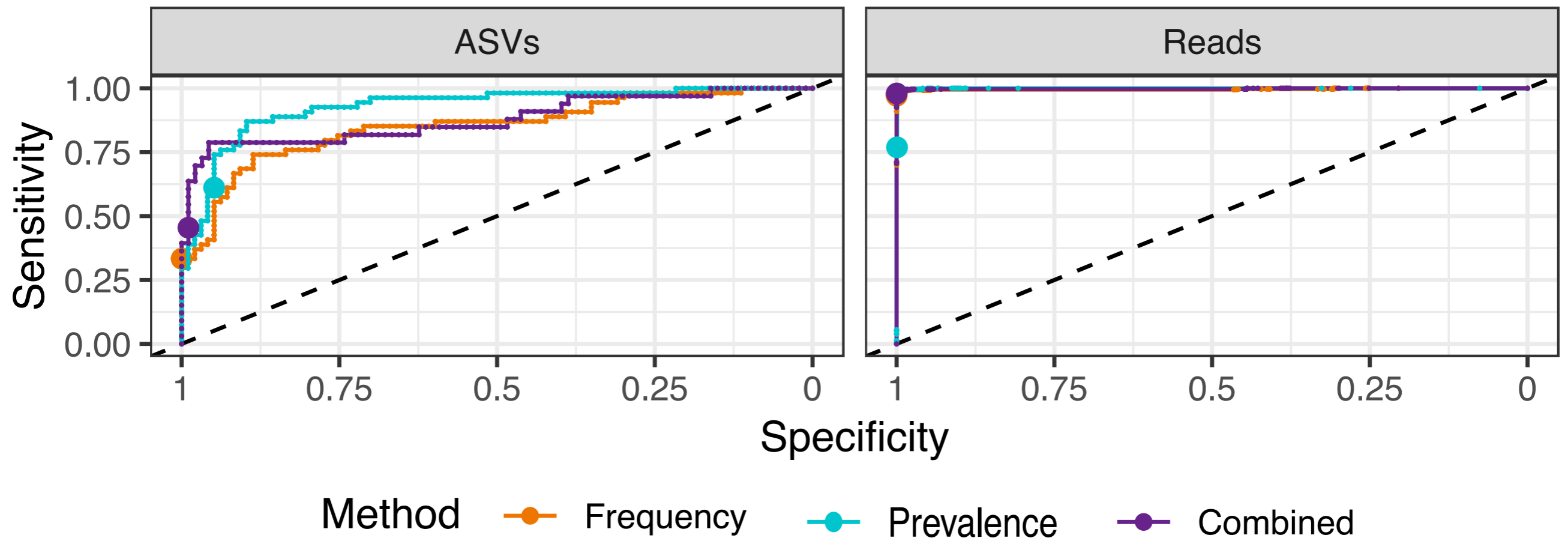
# Validating the Model

## Oral Mucosal Dataset



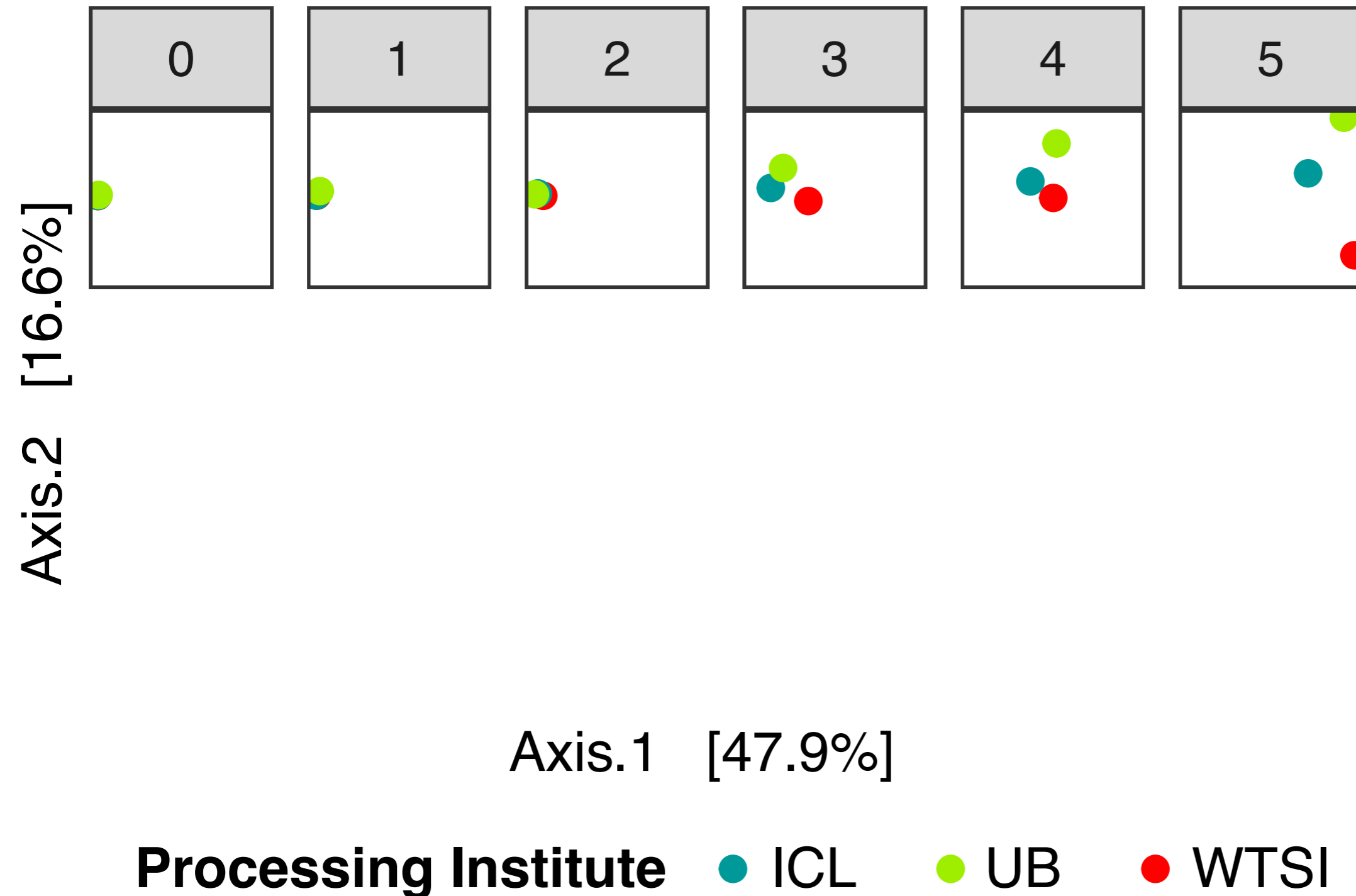
# Classification Accuracy

## *Oral Mucosal Dataset*



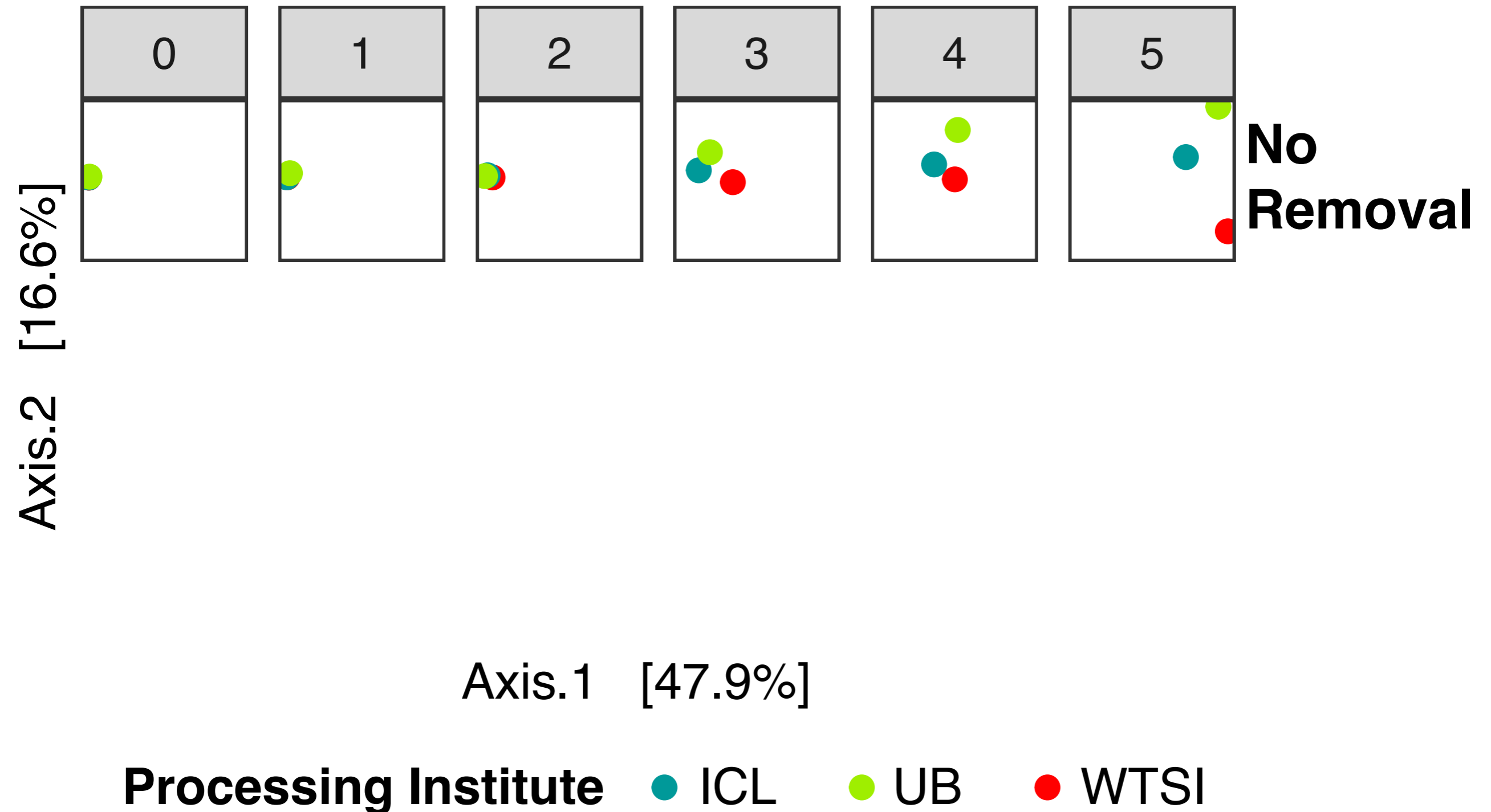
# Reducing Technical Variation

## *Salmonella bongori: Ten-fold dilutions*



# Reducing Technical Variation

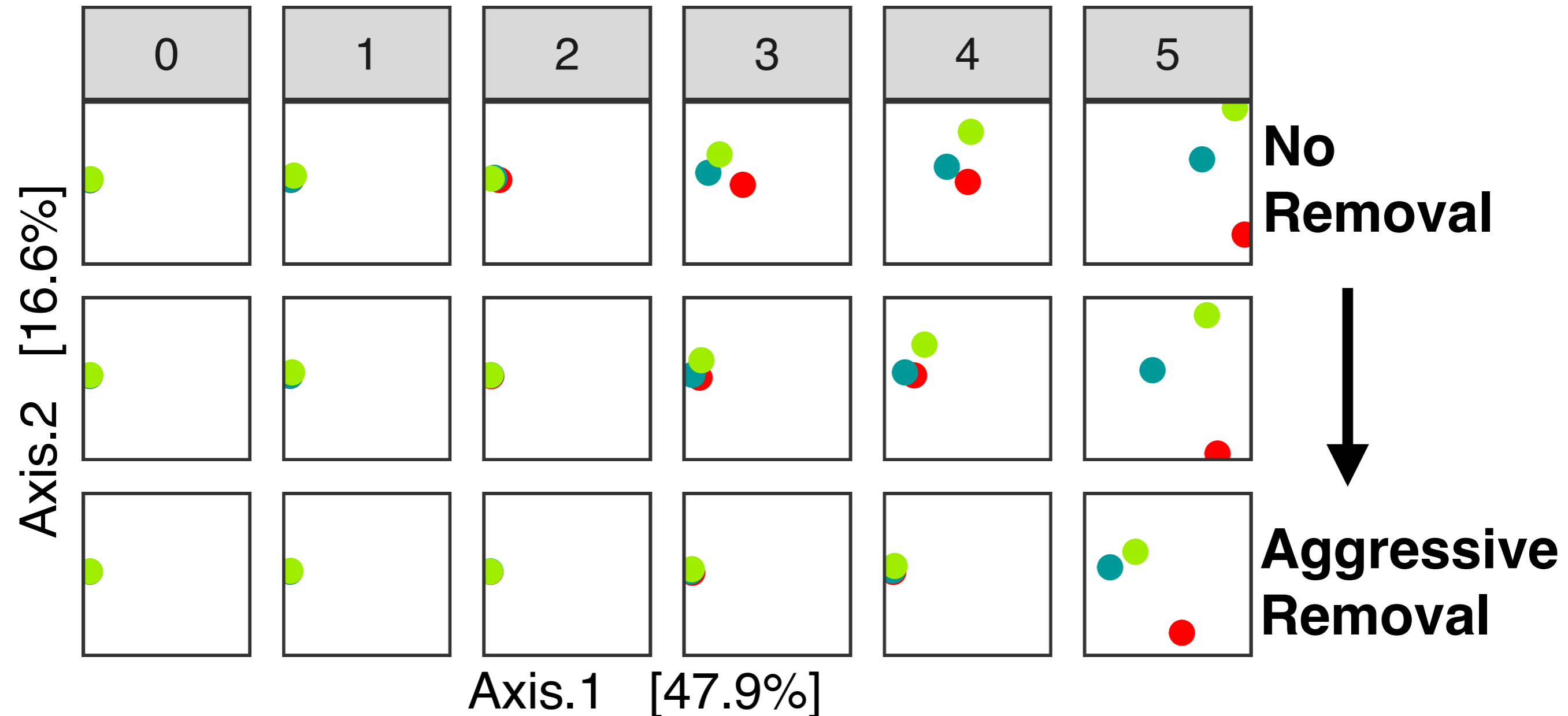
## *Salmonella bongori: Ten-fold dilutions*





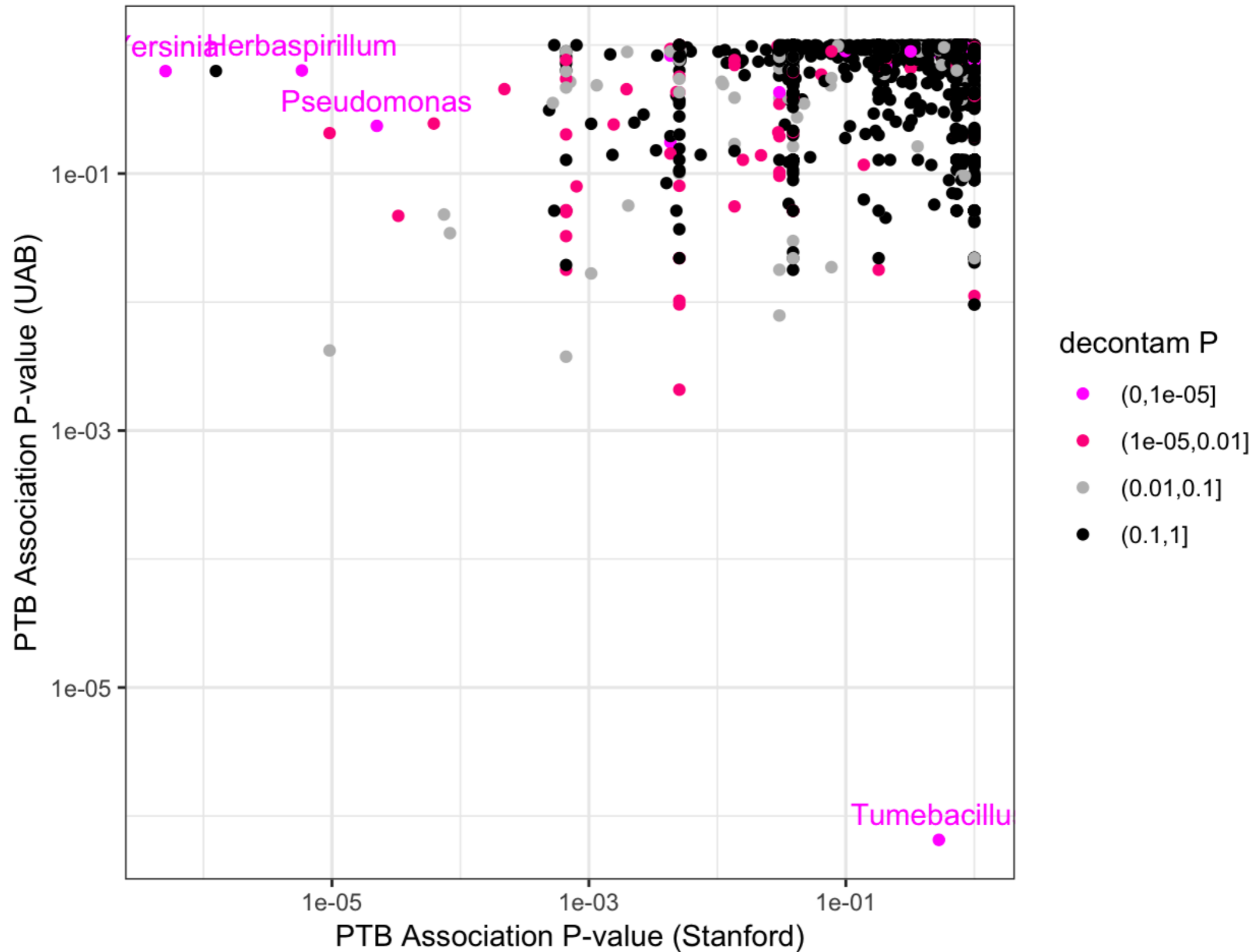
# Reducing Technical Variation

## *Salmonella bongori: Ten-fold dilutions*



**Processing Institute** ● ICL ● UB ● WTSI

# Avoiding Spurious Results



# Available now...

Methodology | **Open Access**

## Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data

Nicole M. Davis, Diana M. Proctor, Susan P. Holmes, David A. Relman and Benjamin J. Callahan  

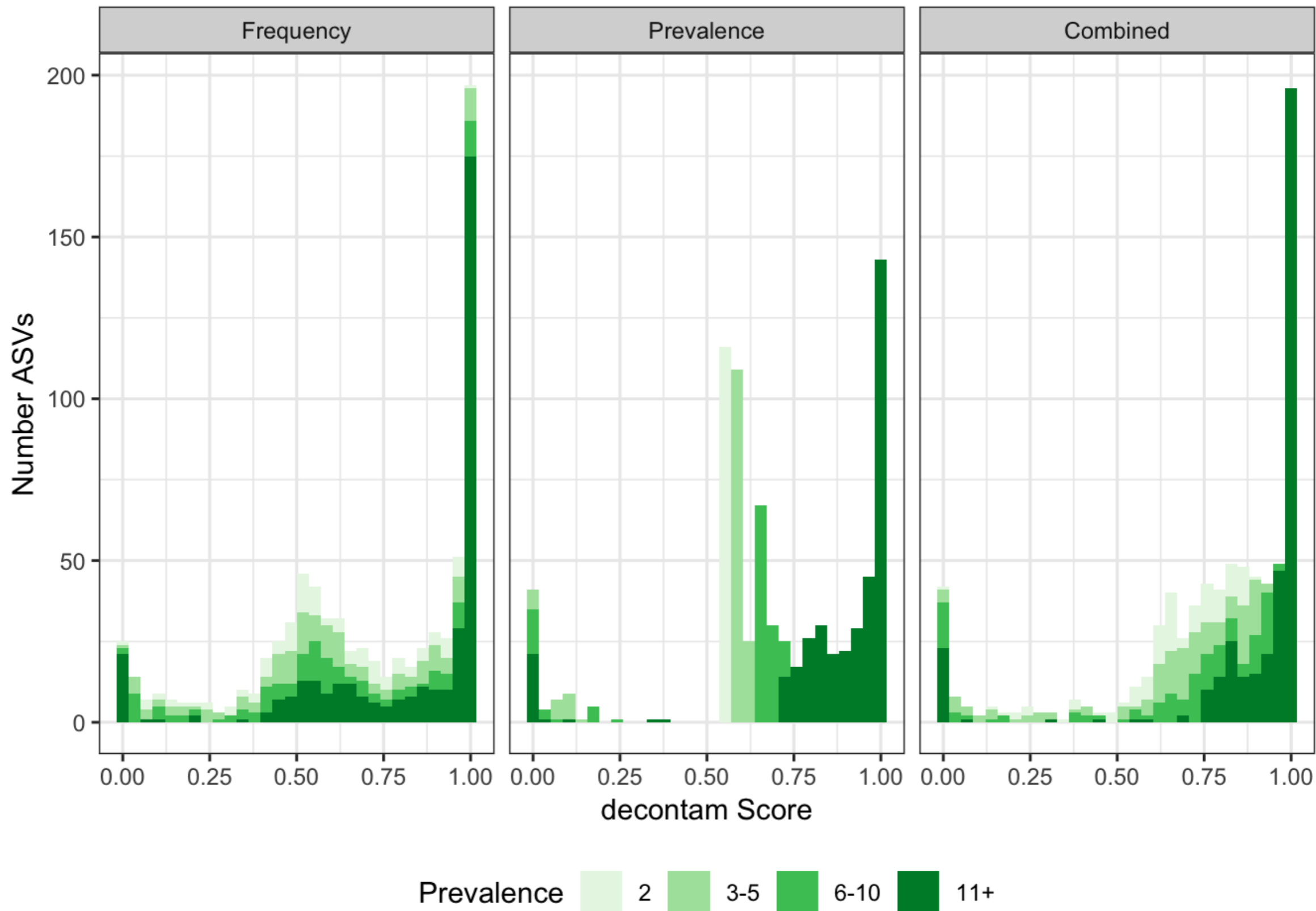
*Microbiome* 2018 6:226



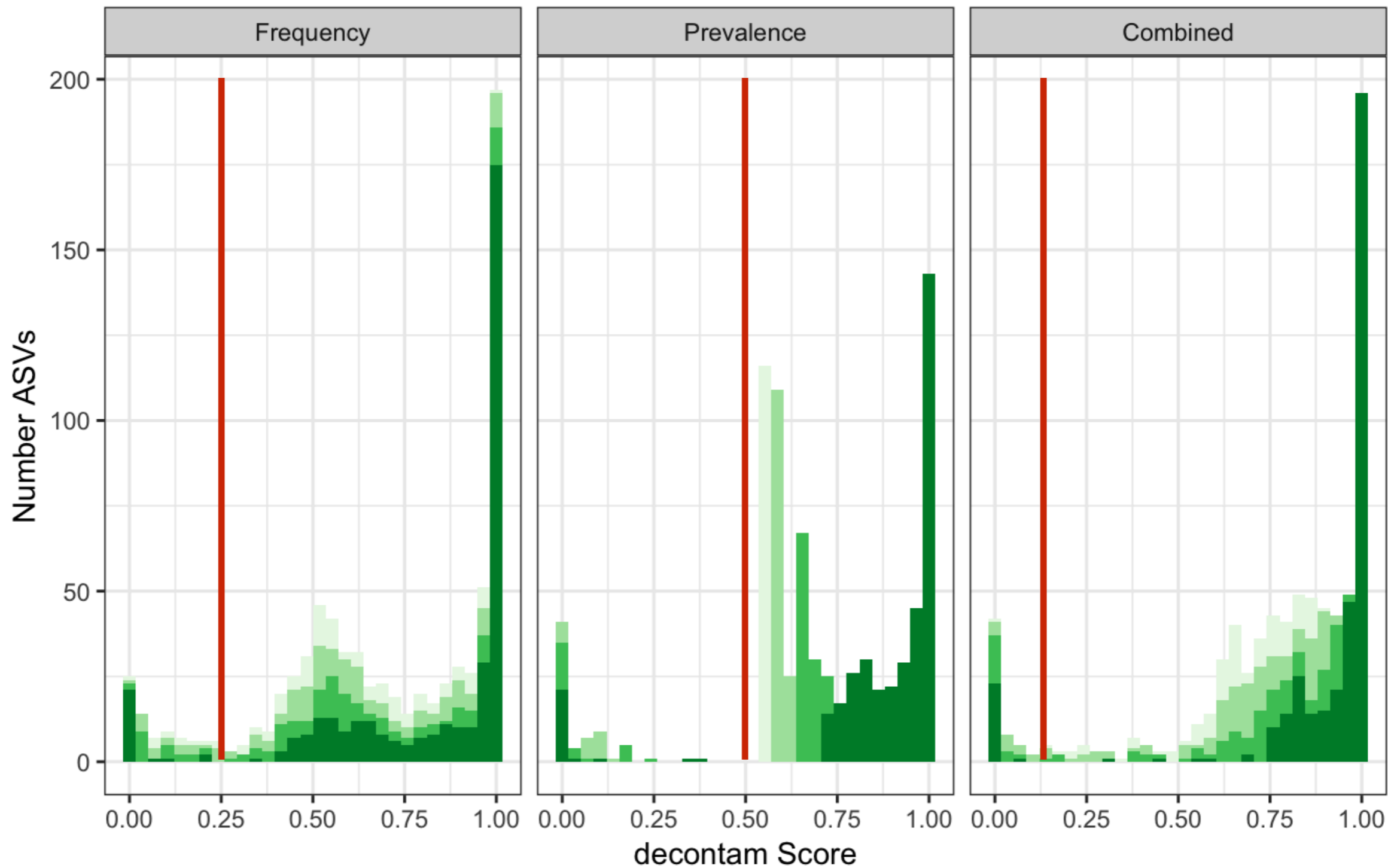
DECONTAM

- Open-source
- Well documented
- R package
- 16S or shotgun

# Beyond defaults

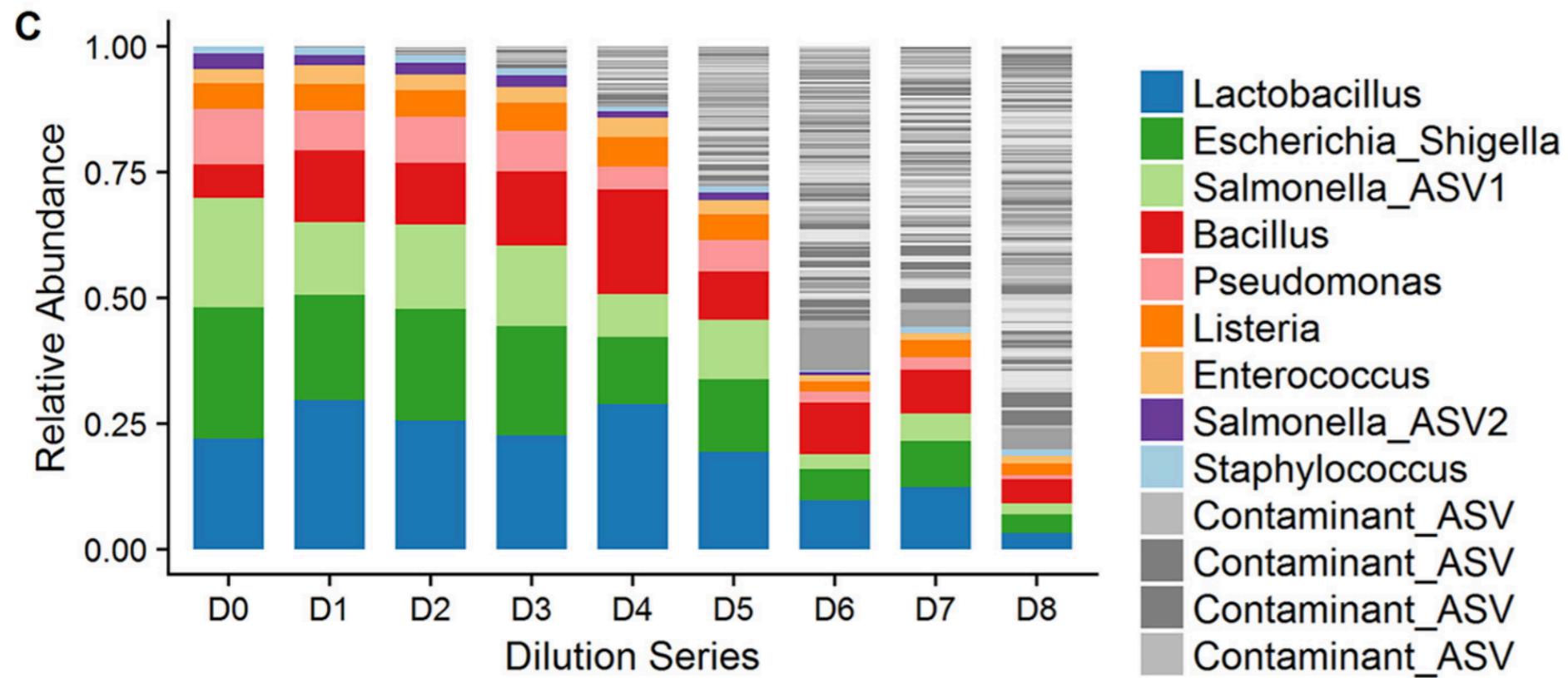


# Beyond defaults



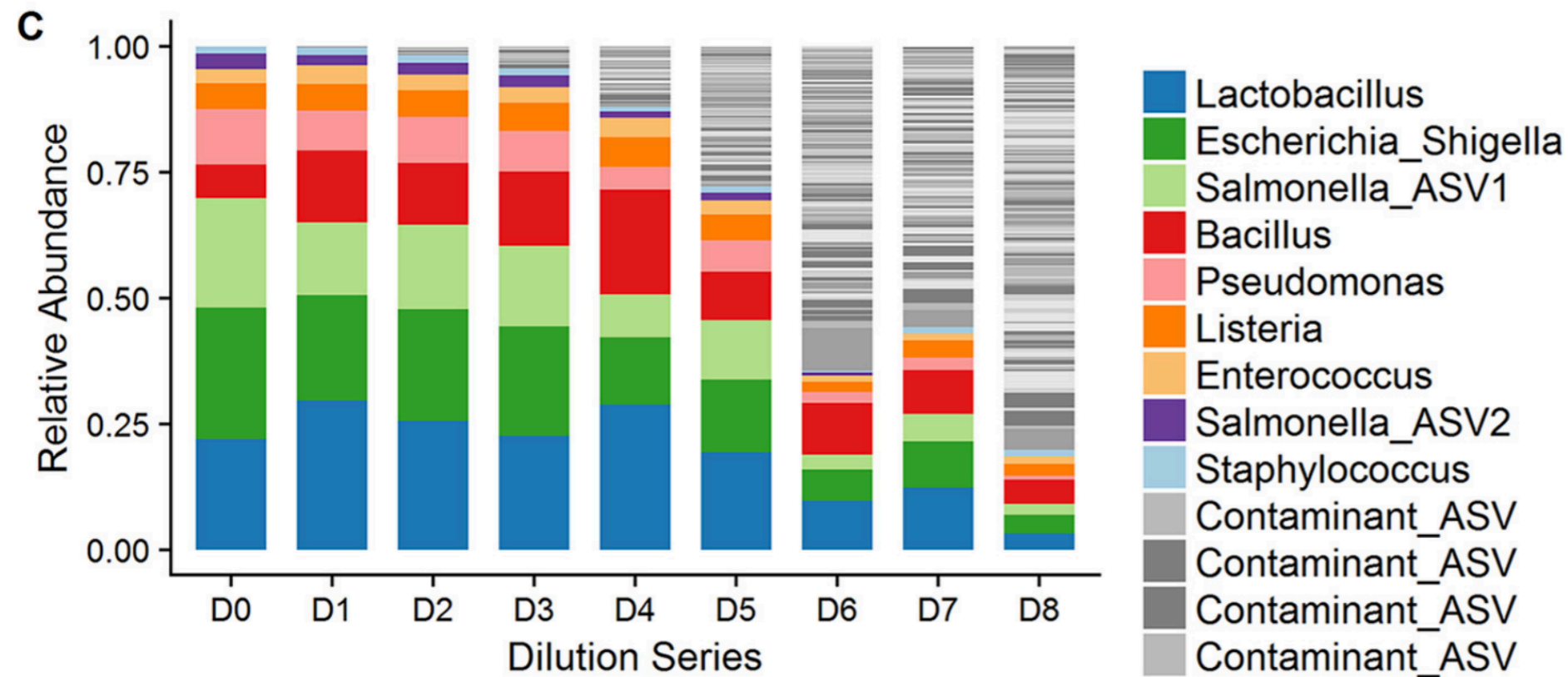
Prevalence 2 3-5 6-10 11+

# When to care?



**Figure:** Karstens, et al. mSystems, 2018.

# When to care?



Aerobiome, placenta, internal tissues,  
rare stuff matters, dry surfaces, parchment...

# Recommendations...

- **There is no substitute for clean lab practices**



# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!

# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!
- Consider dilution series of a positive control

# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!
- Consider dilution series of a positive control
- Record DNA concentrations

# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!
- Consider dilution series of a positive control
- Record DNA concentrations
- *In silico* decontamination (at high resolution)

# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!
- Consider dilution series of a positive control
- Record DNA concentrations
- *In silico* decontamination (at high resolution)
- Be skeptical of unexpected or implausible taxa

# Recommendations...

- **There is no substitute for clean lab practices**
- **Sequence** *multiple* full-process negative controls!
- Consider dilution series of a positive control
- Record DNA concentrations
- *In silico* decontamination (at high resolution)
- Be skeptical of unexpected or implausible taxa
- Report taxa in negative controls

# Recommendations...

**Sequencing-based evidence of a microbiome in locations previously thought “sterile” is not conclusive on its own!**

**What additional evidence could make it convincing?**

# Acknowledgements



Susan Holmes



Nicole Davis

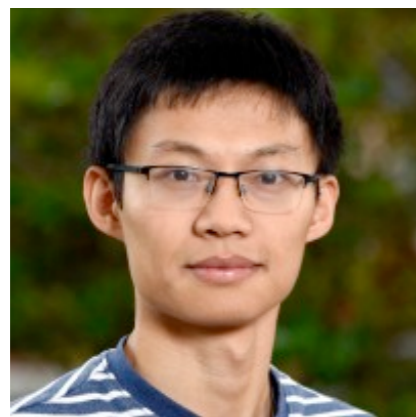


Diana Proctor



David Relman

*And new developments coming soon....*



Caizhi "David" Huang



Jordan Rabasco





# Available now...

Methodology | **Open Access**

## Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data

Nicole M. Davis, Diana M. Proctor, Susan P. Holmes, David A. Relman and Benjamin J. Callahan  

*Microbiome* 2018 6:226



DECONTAM

- Open-source
- Well documented
- R package
- 16S or shotgun