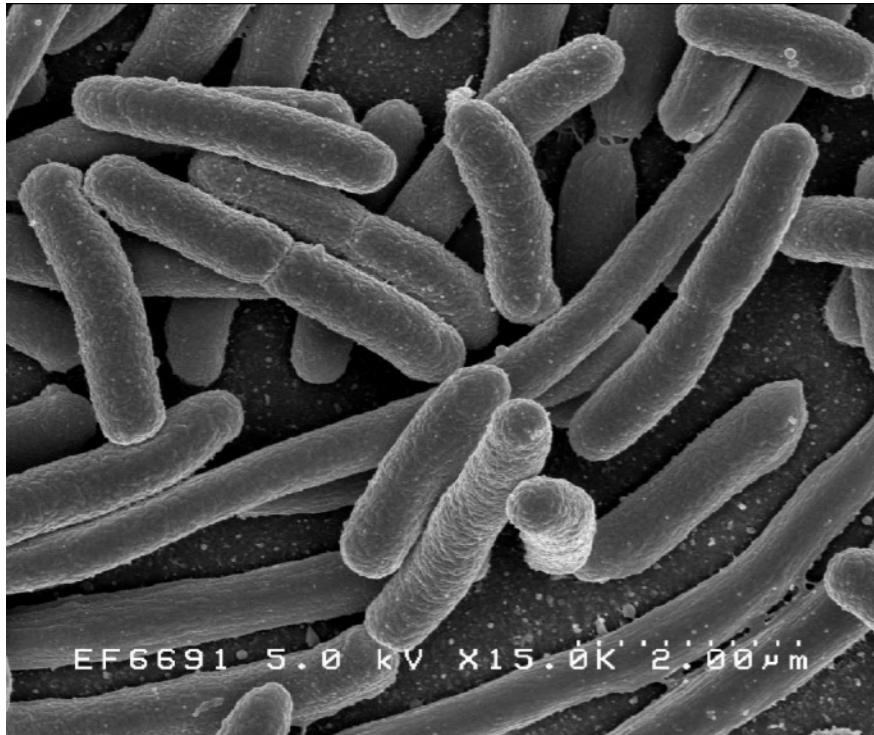


# Contamination, controls and accurate sequencing-based measurement of microbial communities

# A Microbial Census

## **Marker-gene or Metagenomics Sequencing (MGS)**



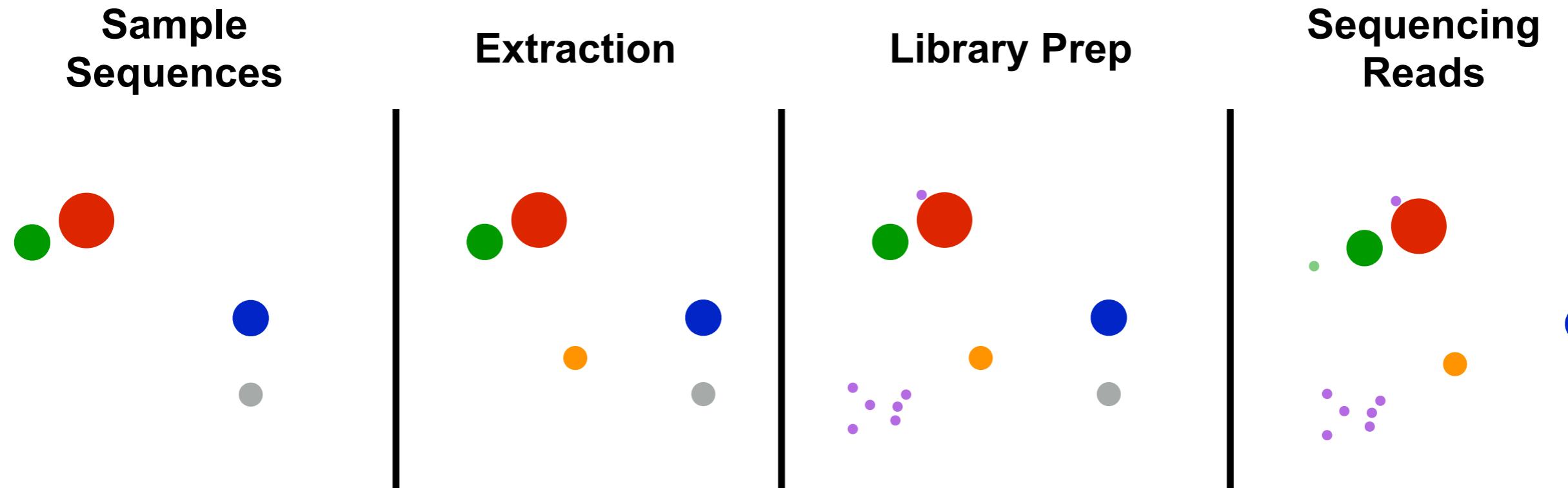
<i>Lactobacillus crispatus</i>	1300	5	0	882	596
<i>Ureaplasma urealytica</i>	15	0	220	0	0
<i>Gardnerella vaginalis</i>	22	0	1	0	412
<i>Prevotella intermedia</i>	0	0	8	12	0
...	...	...	...	...	...

Visualization

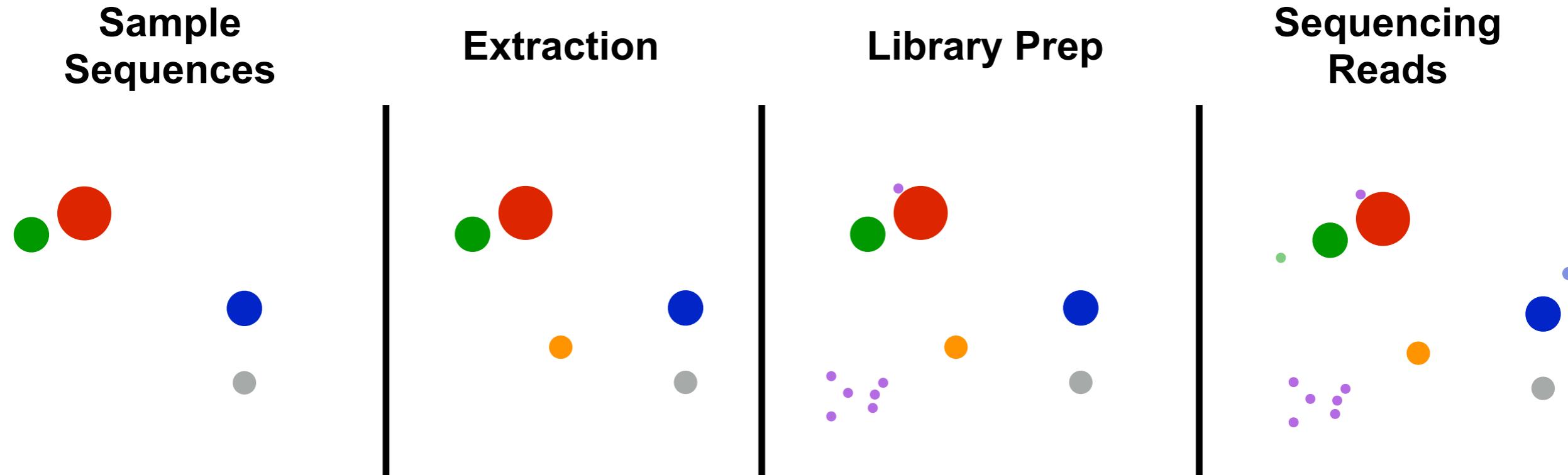
Inference

Exploration

# MGS: What is really there?

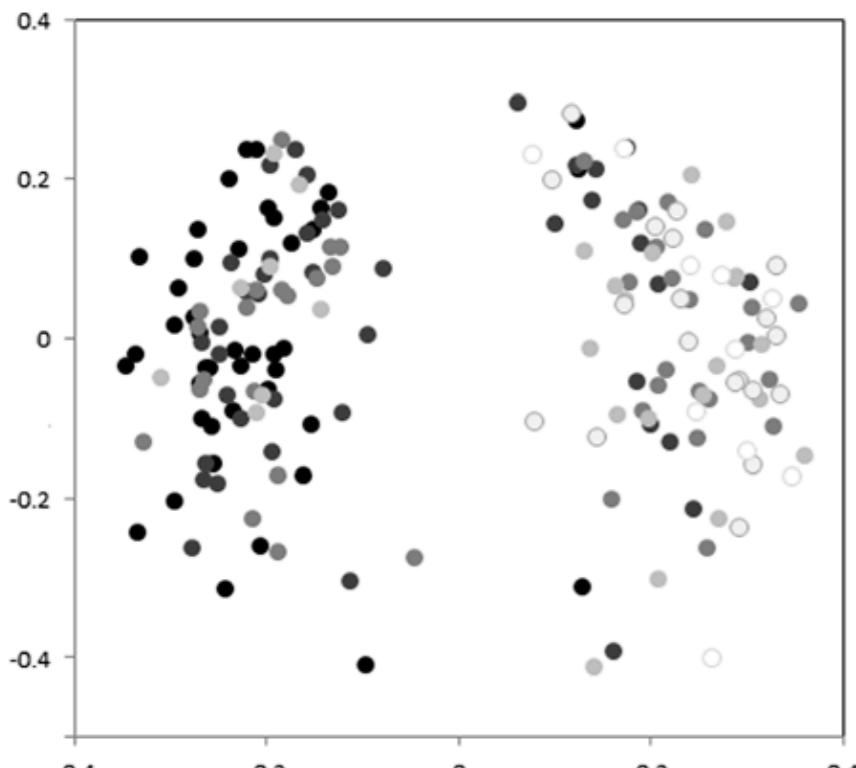


# Problem: Contamination



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

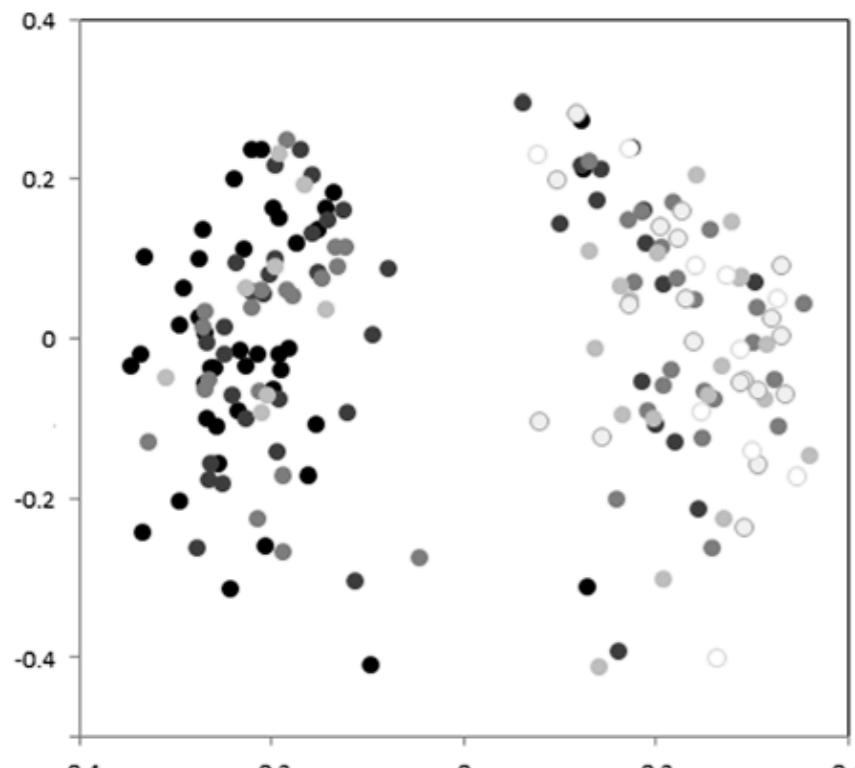
# Problem: Contamination



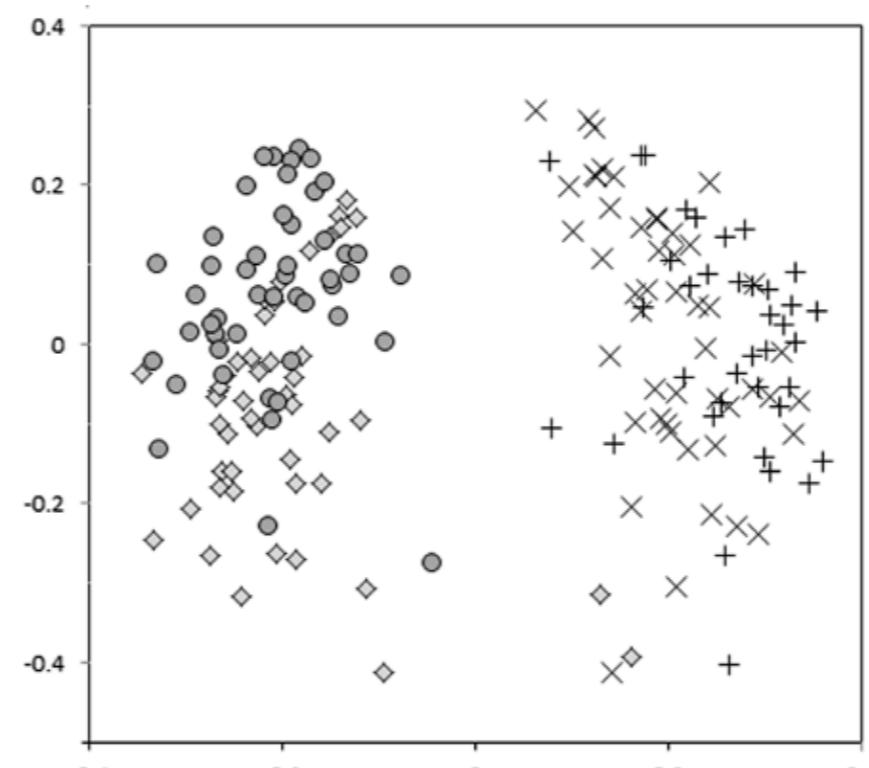
a) Full data coloured by age (months)

**Figure:** Salter, et al. BMC Biology, 2014.

# Problem: Contamination

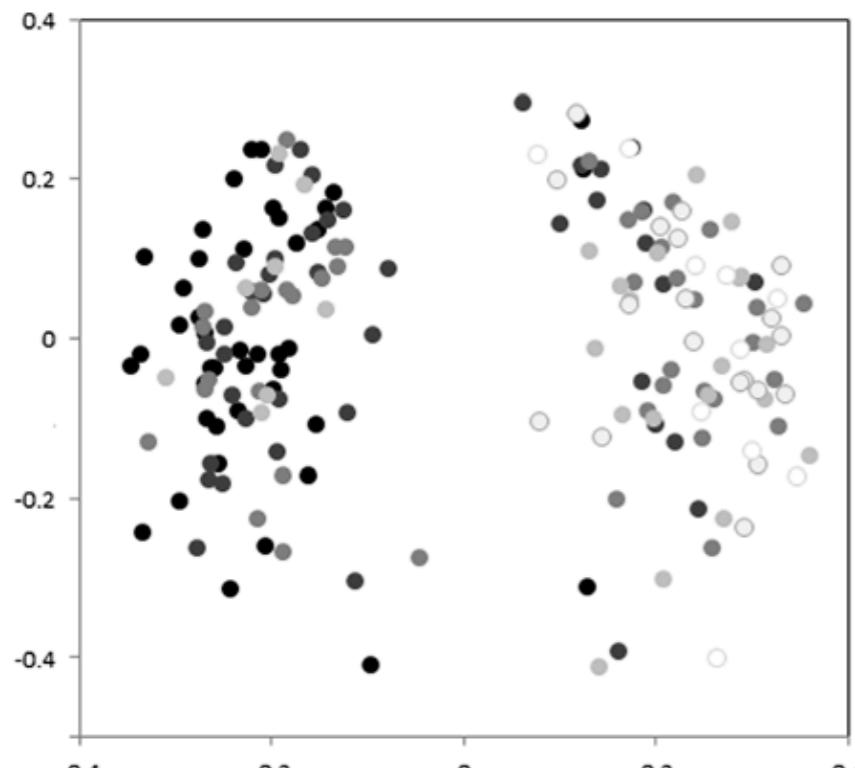


a) Full data coloured by age (months)

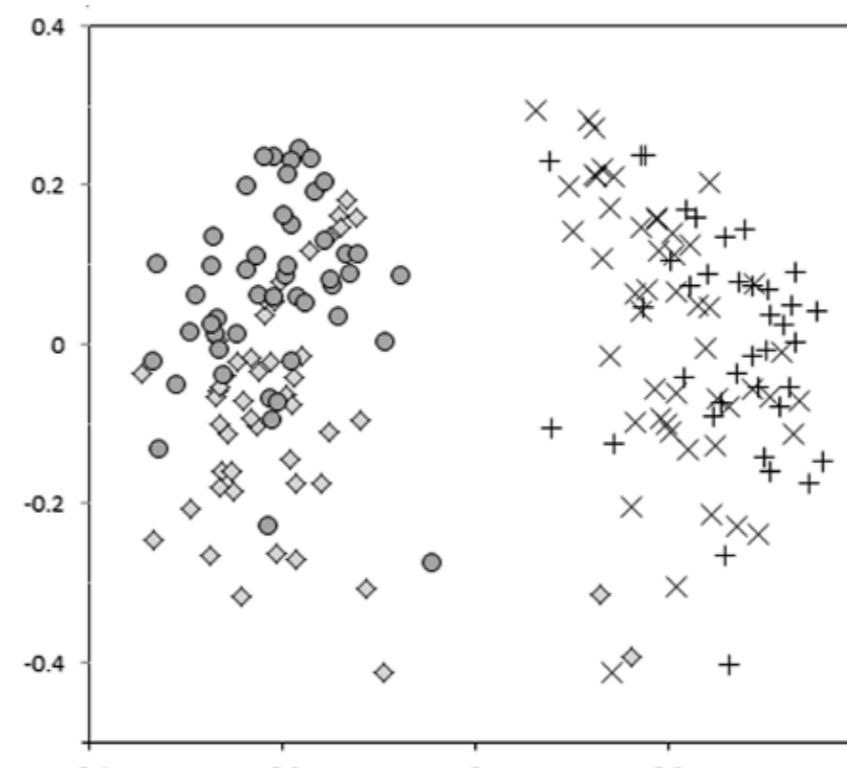


b) Full data shaped by extraction kit

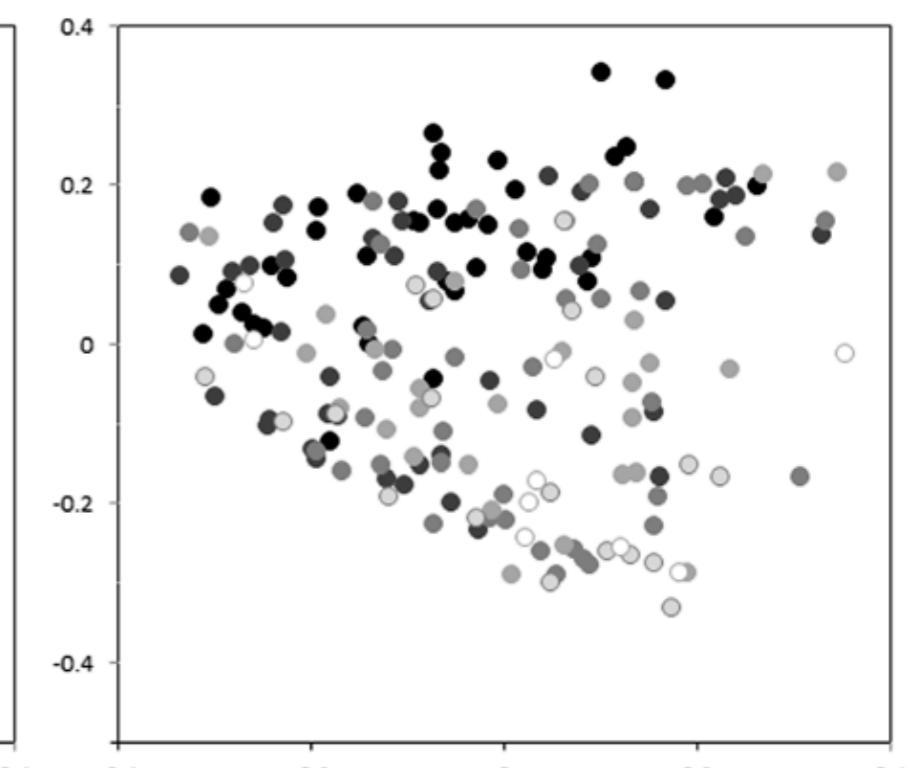
# Problem: Contamination



a) Full data coloured by age (months)



b) Full data shaped by extraction kit



c) Contaminant OTUs removed

**Spurious signal driven by contaminants!**

**Figure:** Salter, et al. BMC Biology, 2014.

# Problem: Contamination

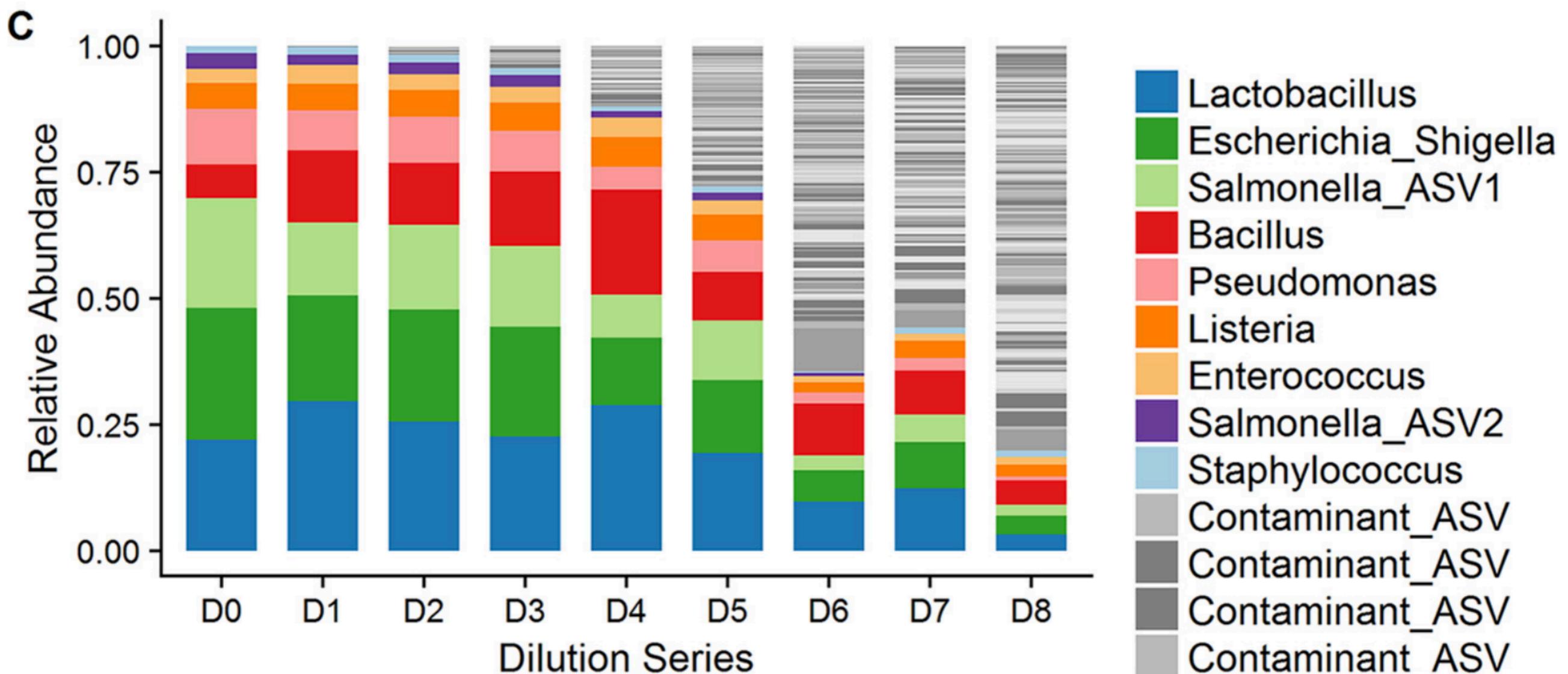


Figure: Karstens, et al. mSystems, 2019.

# Problem: Contamination

## Science Translational Medicine

Current Issue First

RESEARCH ARTICLE | MICROBIOME



### The Placenta Harbors a Unique Microbiome

KJERSTI AAGAARD, JUN MA, KATHLEEN M. ANTONY, RADHIKA GANU, JOSEPH PETROSINO, AND JAMES VERSALOVIC [Authors Info & Affiliations](#)

SCIENCE TRANSLATIONAL MEDICINE • 21 May 2014 • Vol 6, Issue 237 • p. 237ra65 • DOI: 10.1126/scitranslmed.3008599

2,080 1,233



#### Bacteria in Healthy Placentas

Contrary to the prevailing idea of a “sterile” intrauterine environment, Aagaard and coauthors demonstrated the consistent presence of a microbiome in placentas from healthy pregnancies. This microbiome was consistently different from those reported in other parts of the body, including the skin and urogenital tract. The placental microbiome was most similar to that of the oral cavity, but the clinical implications of this finding remain to be explored. In addition, the authors identified associations between the composition of the placental microbiome and a history of remote antenatal infection, as well as preterm birth, raising the possibility that the placental microbiome may play a role in these events.

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

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

### 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](#)

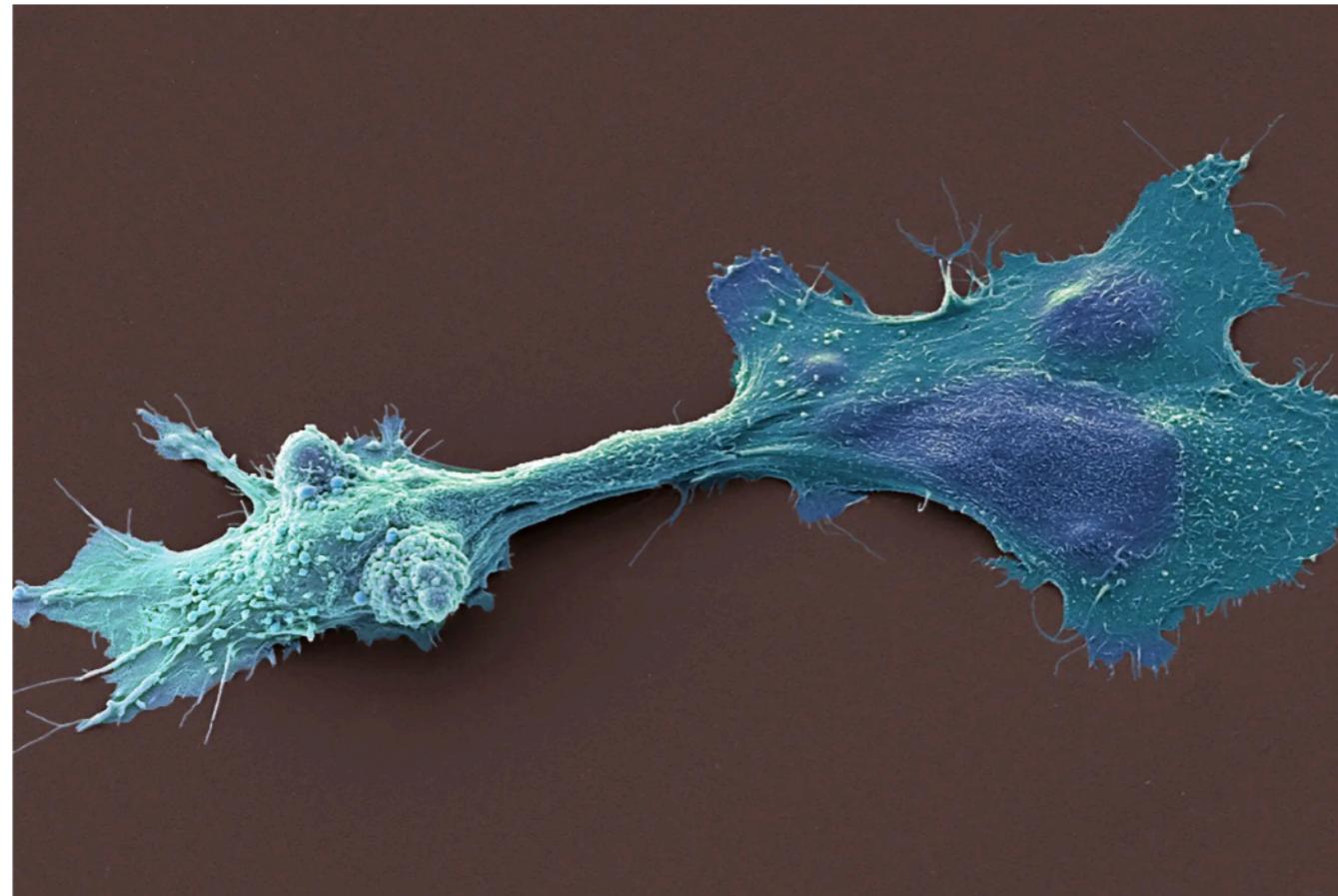
**8898** Accesses | **143** Citations | **110** Altmetric | [Metrics](#)

# Problem: Contamination

The New York Times

## *Researchers Dispute High-Profile Discoveries of Cancer Microbes*

Three studies in elite journals found that tumors are littered with microbes. But several teams have been unable to replicate the work.



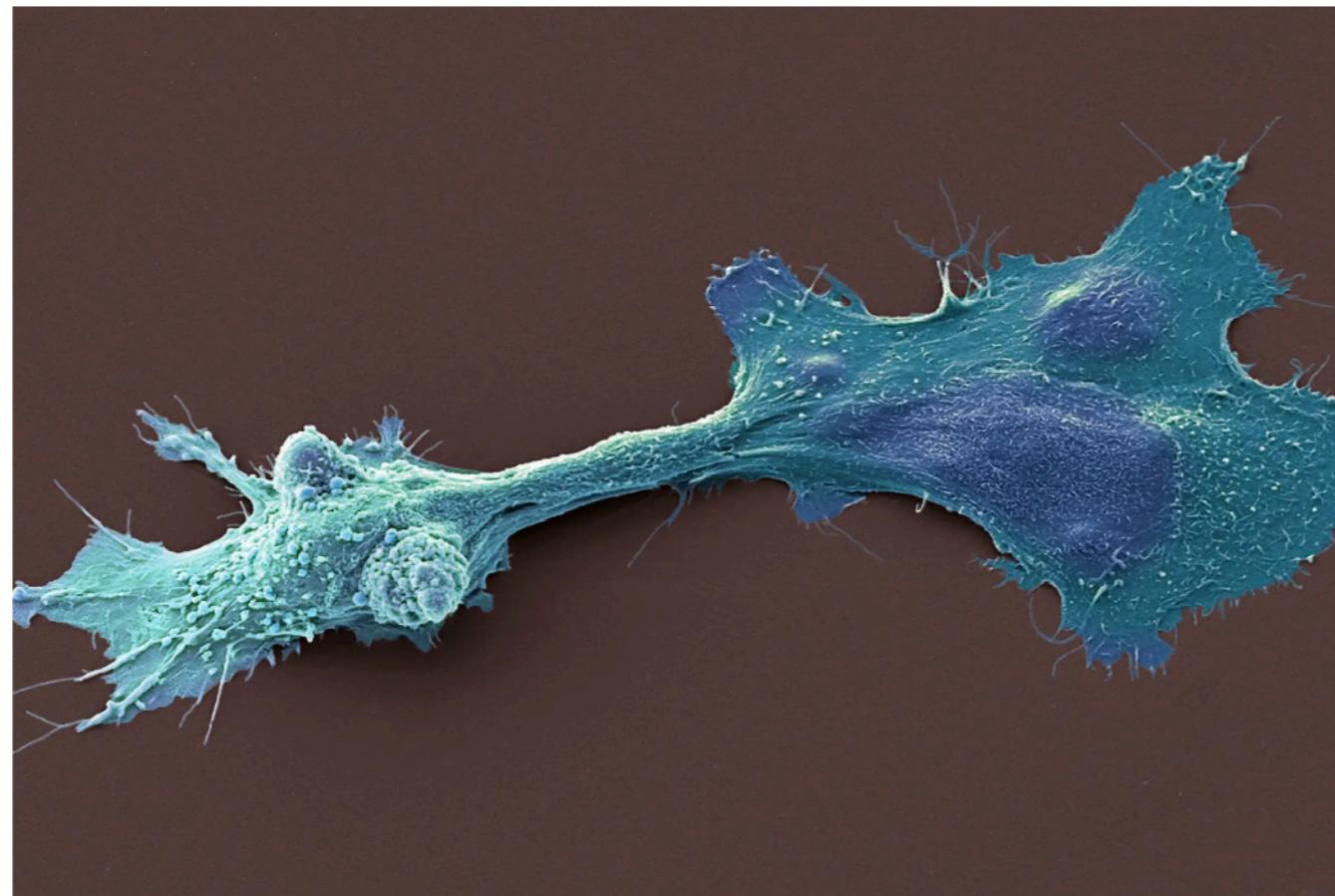
A scanning electron micrograph of pancreatic cancer cells. Several research teams have cast doubt on the science of tumor-causing microbes, saying they were most likely mirages or the result of contamination. Steve Gschmeissner/Science Source

# Problem: Contamination

The New York Times

## *Researchers Dispute High-Profile Discoveries of Cancer Microbes*

Three studies in elite journals found that tumors are littered with microbes. But several teams have been unable to replicate the work.



A scanning electron micrograph of pancreatic cancer cells. Several research teams have cast doubt on the science of tumor-causing microbes, saying they were most likely mirages or the result of contamination. Steve Gschmeissner/Science Source

But what kind of “contamination” is this exactly?

# Problem: Contamination

**Now what?**

# Modeling Contaminants

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

*hence*

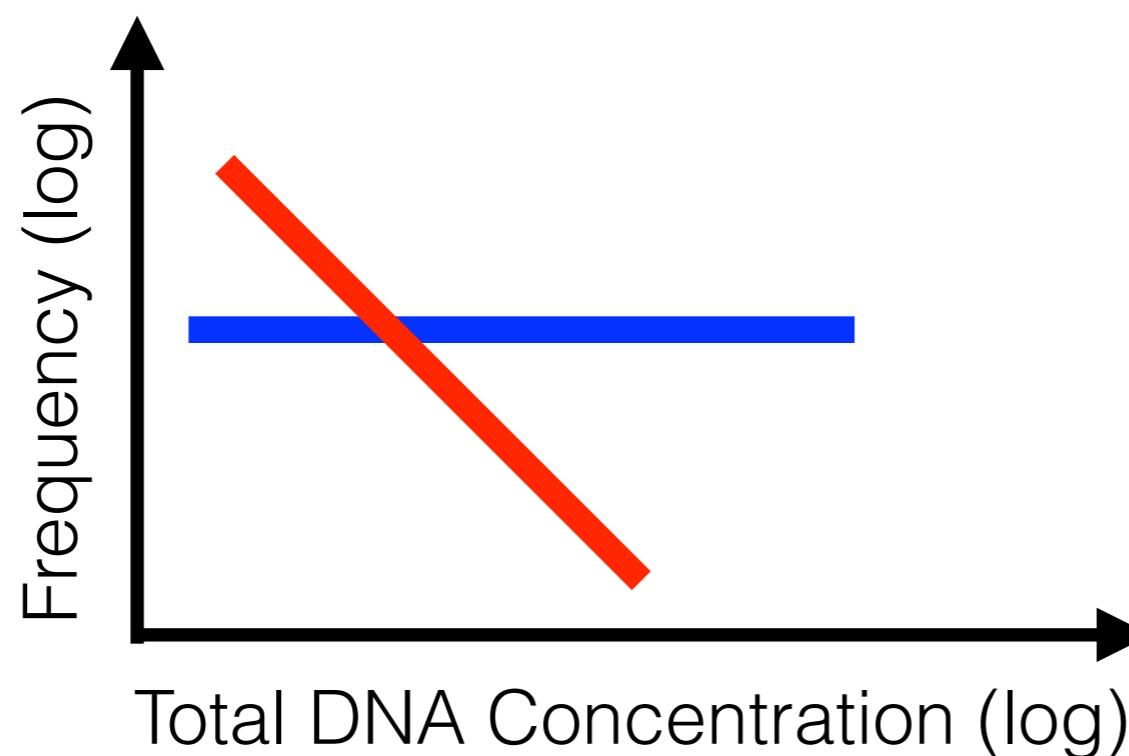
$f_C = C/(S+C) \sim 1/T$ , where  $C \ll S$

---

# Modeling Contaminants

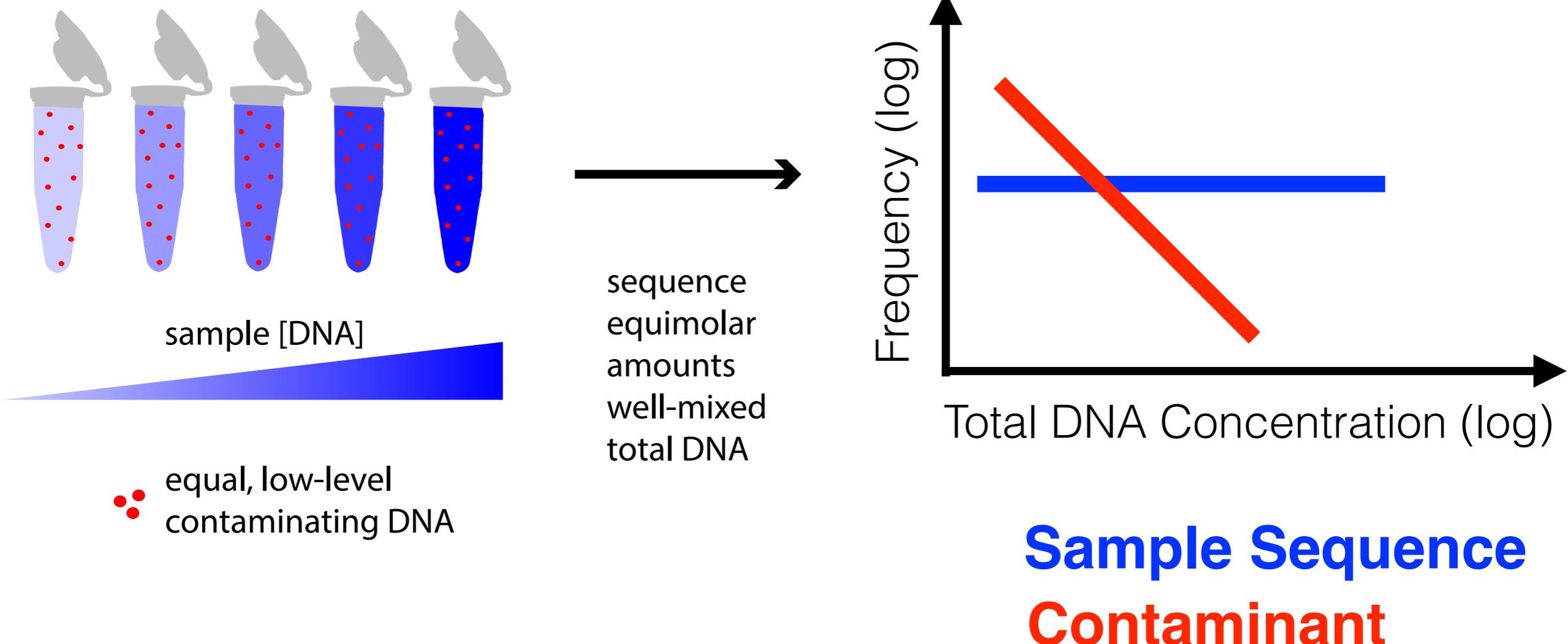
$T = S + C$ , where  $C$  is constant  
*hence*

$$f_C = C/(S+C) \sim 1/T, \text{ where } C \ll S$$



**Sample Sequence**  
**Contaminant**

# Modeling Contaminants



# 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

Input: DNA concentrations,  
Feature table w/ abundances.

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

## Prevalence

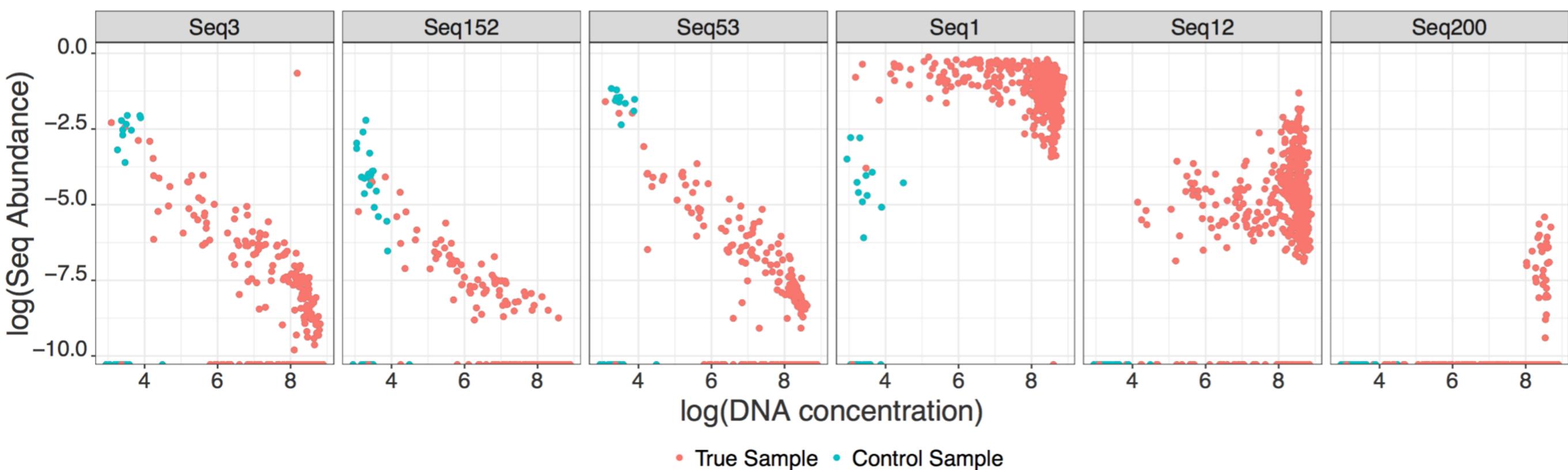
Input: Categorization of samples as negative controls,  
Feature table w/ abundances or presences.

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

*Needs range of DNA concentrations*

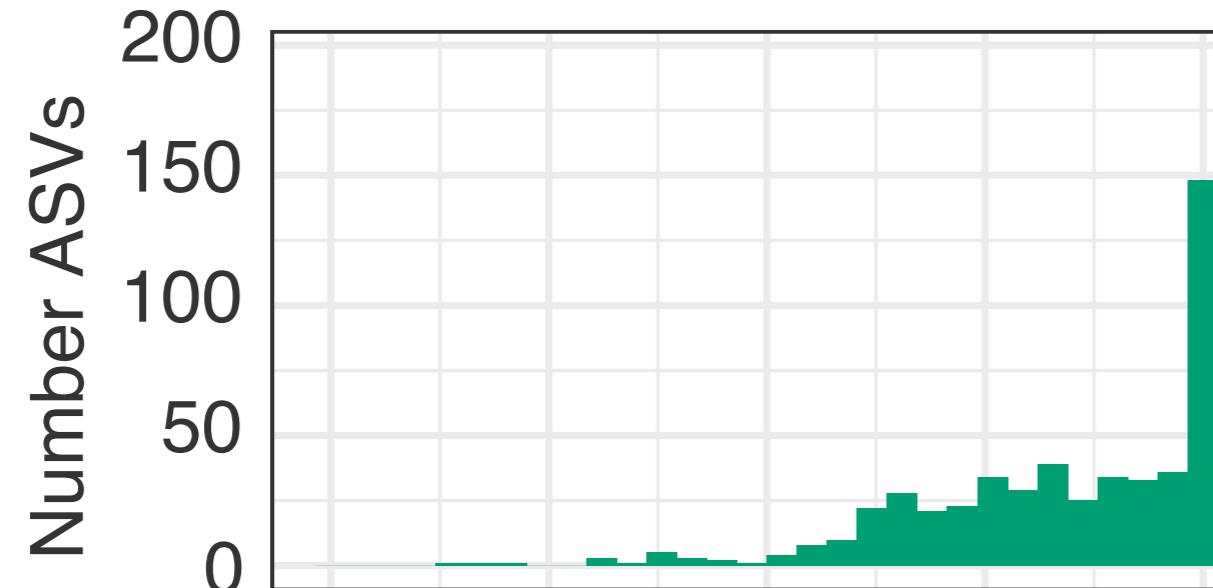
*Needs multiple (5+) sequenced negative controls*

# Validating the Model

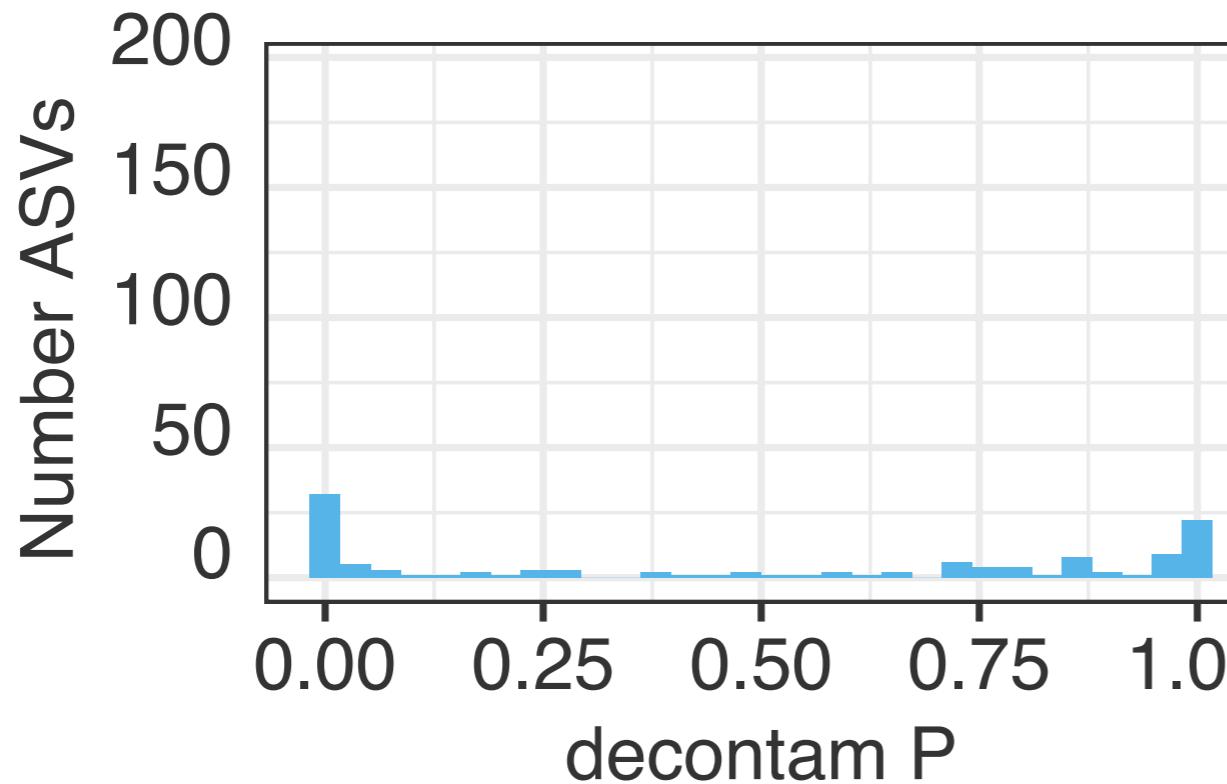


# Validating the Model

## *Oral Mucosal Dataset*



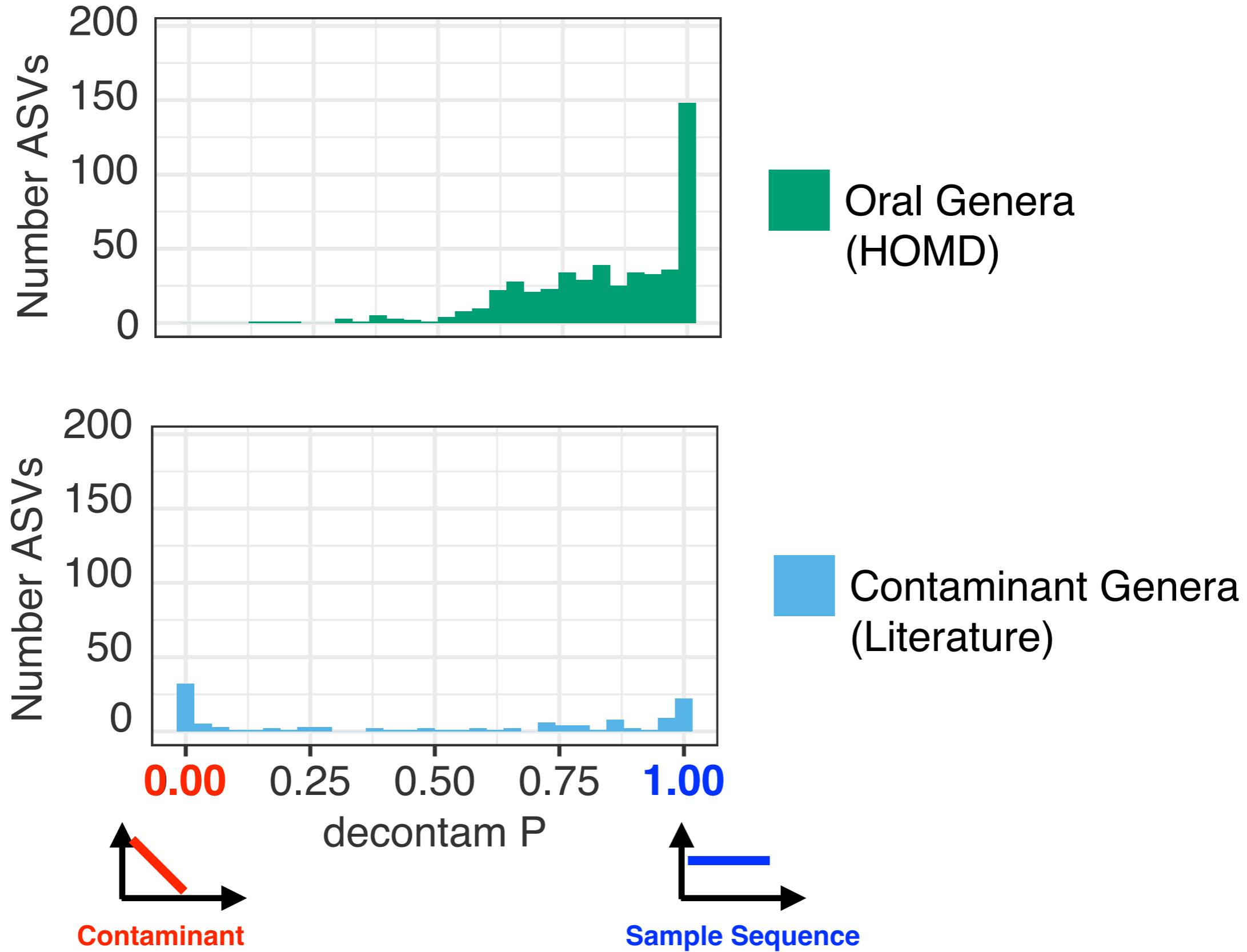
Oral Genera  
(HOMD)



Contaminant Genera  
(Literature)

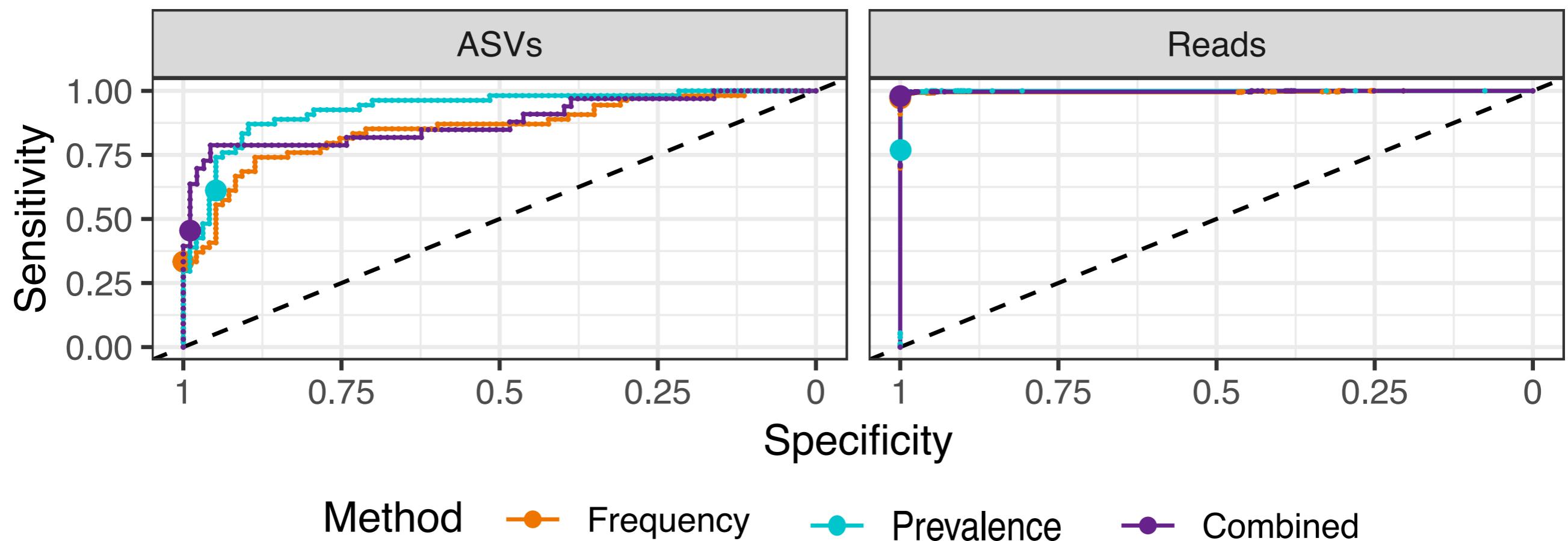
# Validating the Model

## *Oral Mucosal Dataset*



# Classification Accuracy

## *Oral Mucosal Dataset*

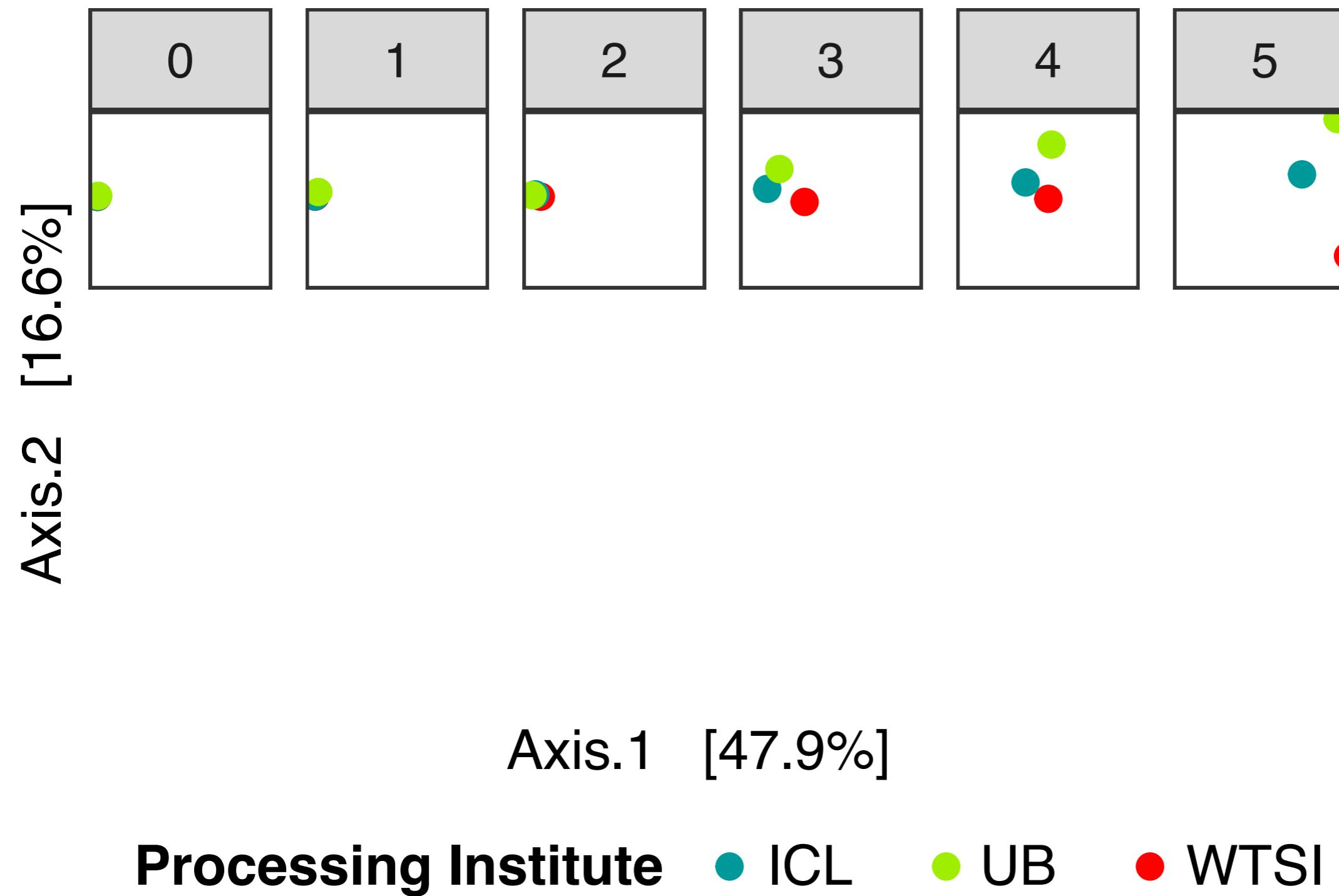


# Classification Accuracy

**High accuracy,  
if data fits  
assumptions.**

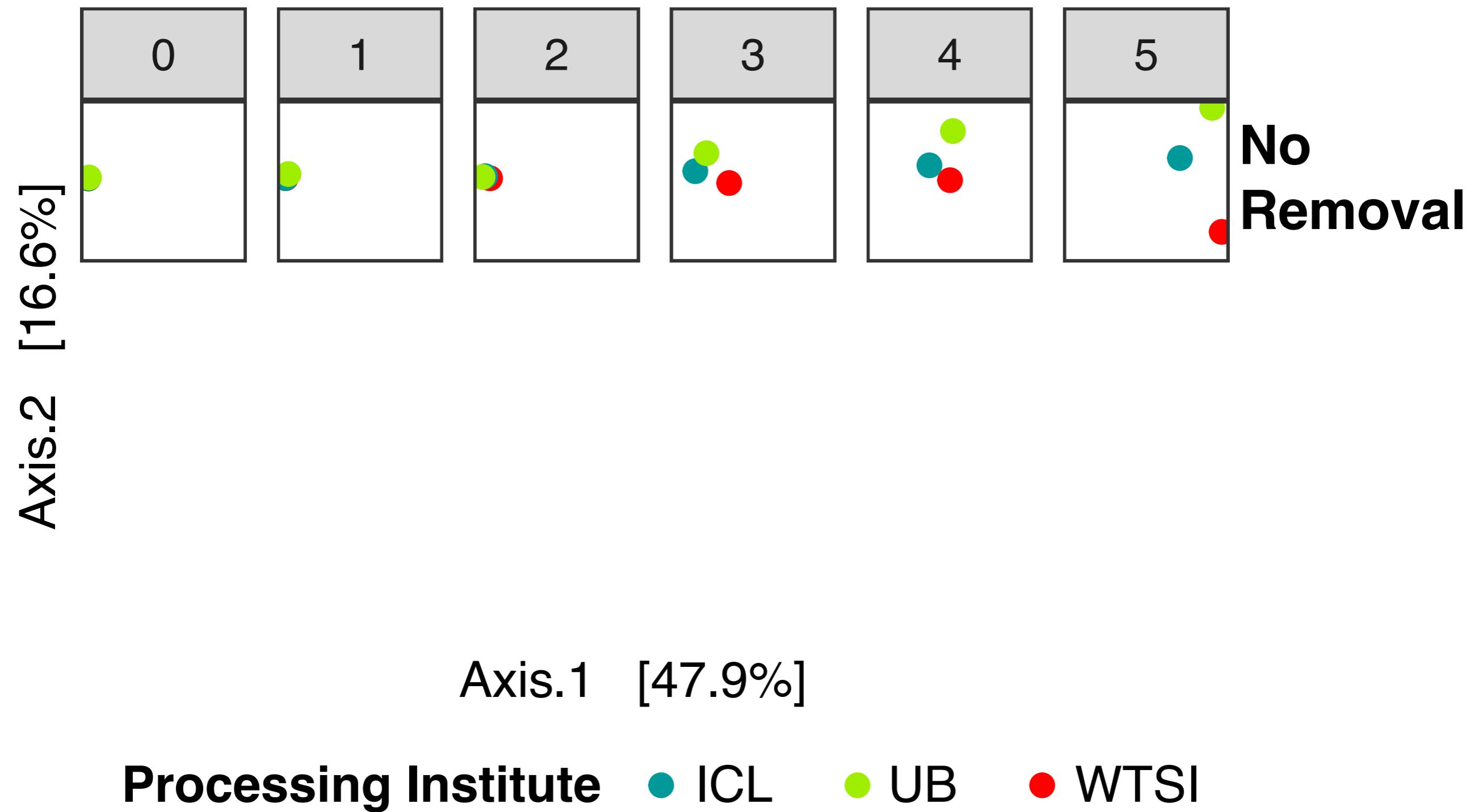
# Reducing Technical Variation

## *Salmonella bongori: Ten-fold dilutions*



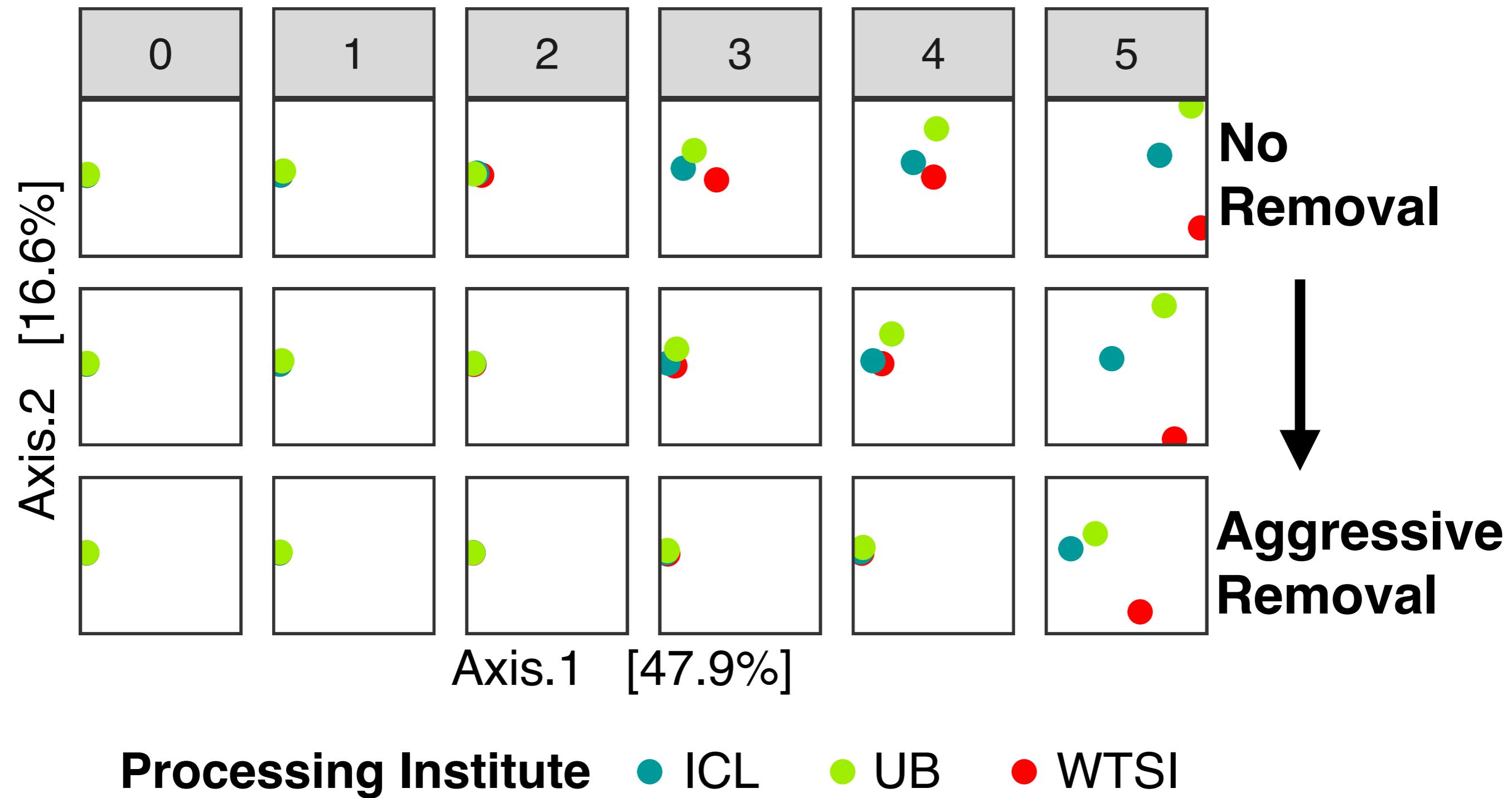
# Reducing Technical Variation

## *Salmonella bongori: Ten-fold dilutions*

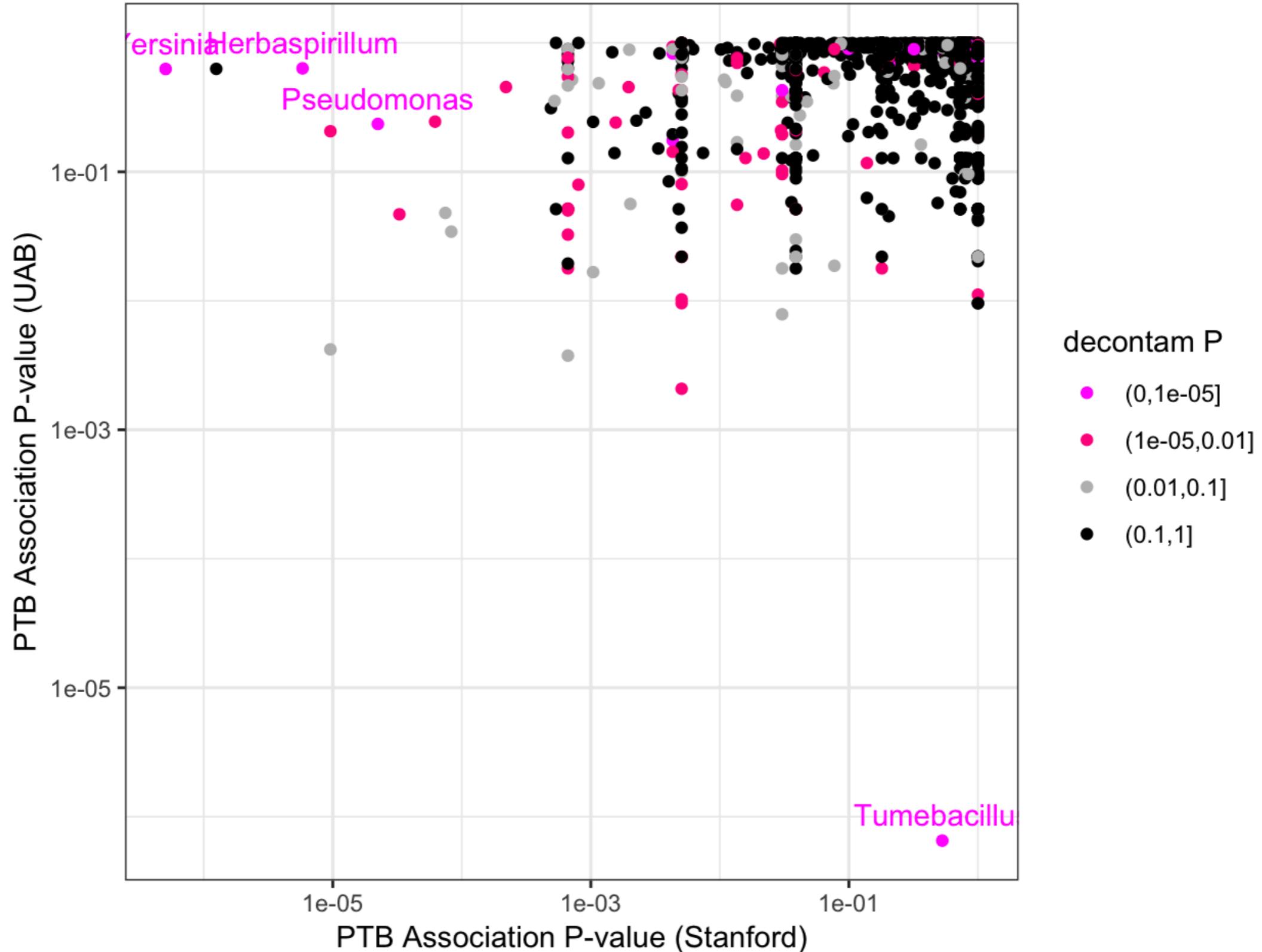


# Reducing Technical Variation

## *Salmonella bongori: Ten-fold dilutions*



# Avoiding Spurious Results



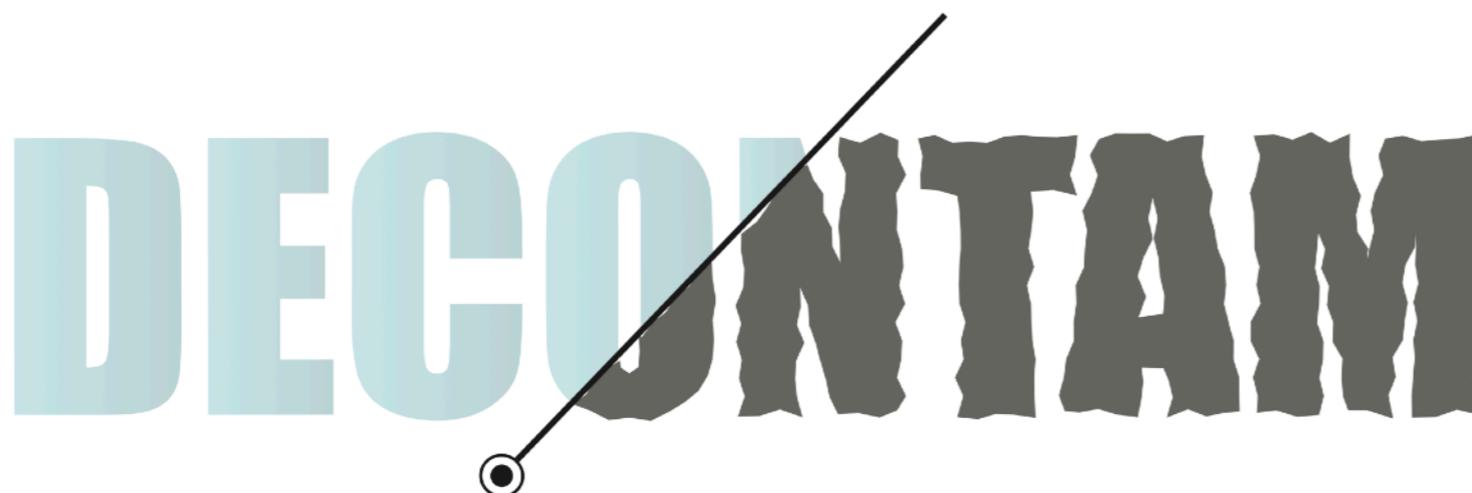
# Available now...

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 [✉](#) [ID](#)

*Microbiome* 2018 6:226



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

# Decontam Method

## Frequency

Input: DNA concentrations,  
Feature table w/ abundances.

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

```
contam <- isContaminant(seqtab, conc, threshold)
```

# Decontam Method

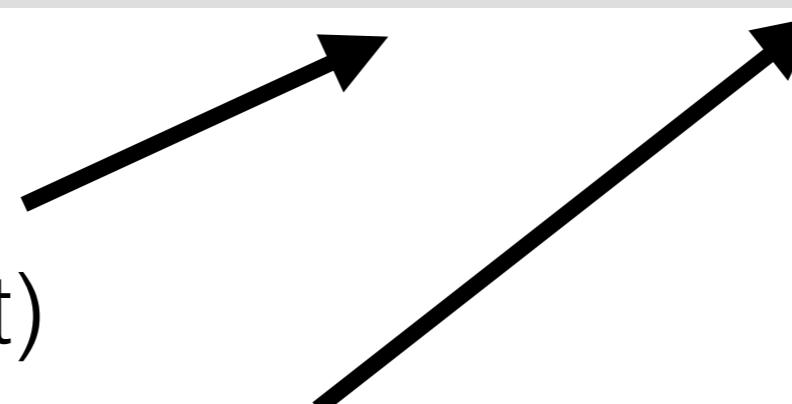
## Frequency

Input: DNA concentrations,  
Feature table w/ abundances.

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

```
contam <- isContaminant(seqtab, conc, threshold)
```

ASV or OTU table  
(or phyloseq object)



Vector of DNA concentrations  
(or phyloseq variable name)

Number: 0 to 1  
(default 0.5)

# Decontam Method

## 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

## 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.

```
contam <- isContaminant(seqtab, neg, threshold)
```

# Decontam Method

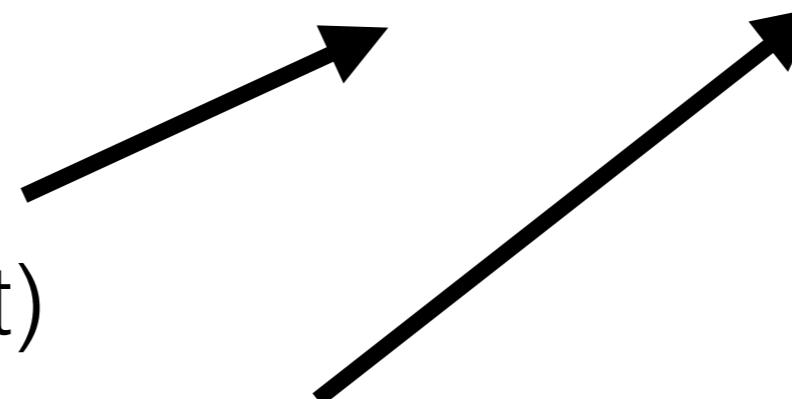
## Prevalence

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.

```
contam <- isContaminant(seqtab, neg, threshold)
```

ASV or OTU table  
(or phyloseq object)



Number: 0 to 1  
(default 0.5)

Vector: True if neg control, False otherwise  
(or phyloseq variable name)

# Decontam Method

## Output

```
> class(contam)
## [1] "data.frame"

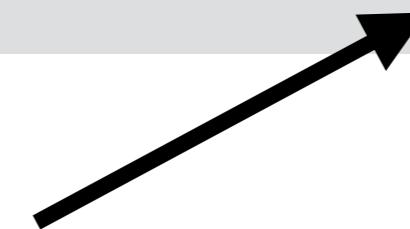
> head(contam)

##          freq  prev      p.freq      p.prev          p contaminant
## Seq1 0.323002694 549 1.000000e+00       NA 1.000000e+00 FALSE
## Seq2 0.098667396 538 1.000000e+00       NA 1.000000e+00 FALSE
## Seq3 0.003551358 160 1.135975e-18       NA 1.135975e-18 TRUE
## Seq4 0.067588419 519 9.999998e-01       NA 9.999998e-01 FALSE
## Seq5 0.045174743 354 1.000000e+00       NA 1.000000e+00 FALSE
## Seq6 0.040417101 538 1.000000e+00       NA 1.000000e+00 FALSE
```

Score: 0 to 1

(0: contaminant-like,  
1: non-contaminant-like)

Classification: T/F  
(score < threshold)



# Decontam Method

## Output

```
> class(contam)  
## [1] "data.frame"  
  
> head(contam)  
  
##          freq  prev      p.freq      p.prev          p contaminant  
## Seq1 0.323002694 549 1.000000e+00       NA 1.000000e+00 FALSE  
## Seq2 0.098667396 538 1.000000e+00       NA 1.000000e+00 FALSE  
## Seq3 0.003551358 160 1.135975e-18       NA 1.135975e-18 TRUE  
## Seq4 0.067588419 519 9.999998e-01       NA 9.999998e-01 FALSE  
## Seq5 0.045174743 354 1.000000e+00       NA 1.000000e+00 FALSE  
## Seq6 0.040417101 538 1.000000e+00       NA 1.000000e+00 FALSE
```

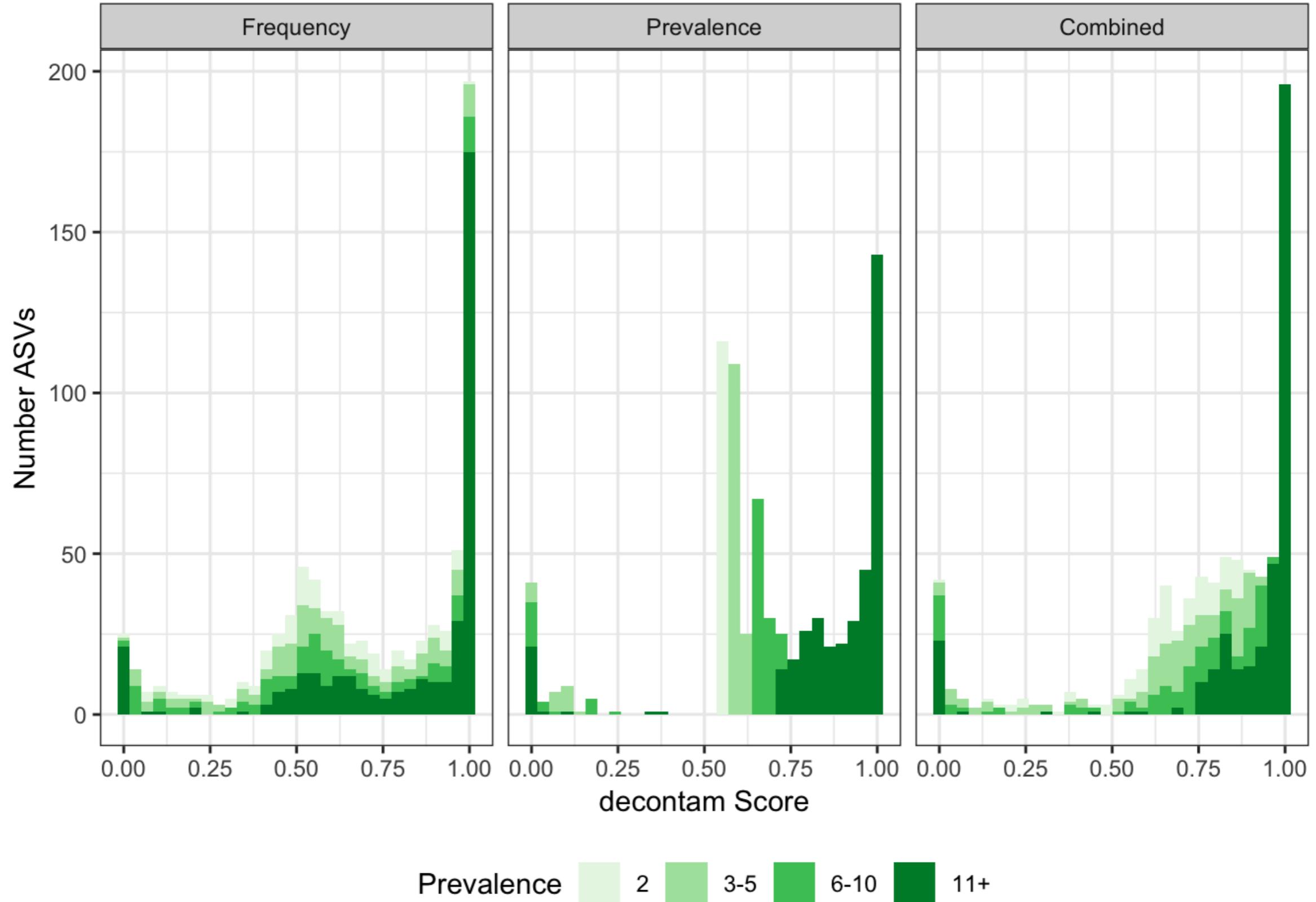
**Should not have called  
the decontam score "p"**

Score: 0 to 1  
(0: contaminant-like,  
1: non-contaminant-like)

Classification: T/F  
(score < threshold)

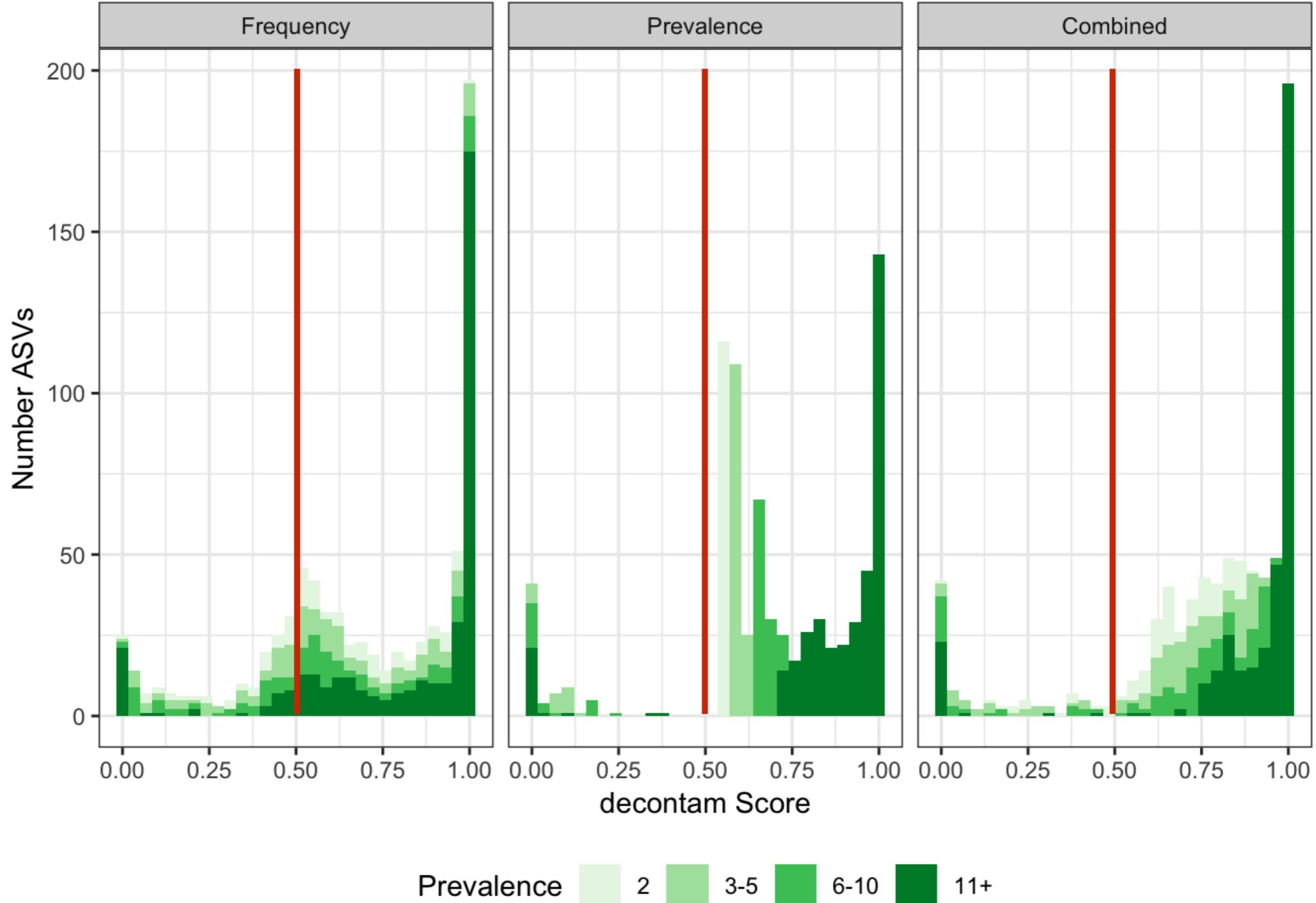


# Beyond defaults



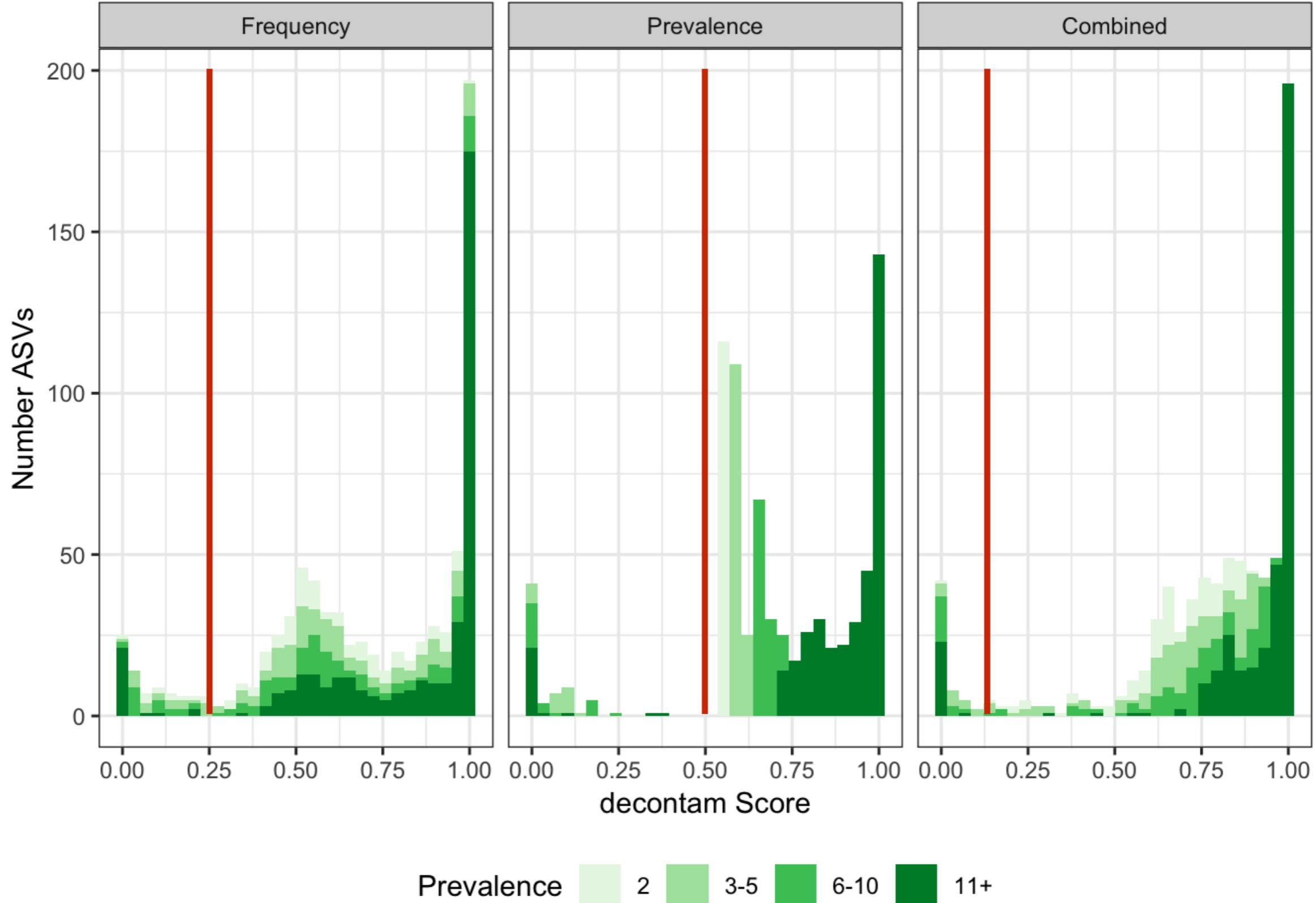
# Beyond defaults

## Default

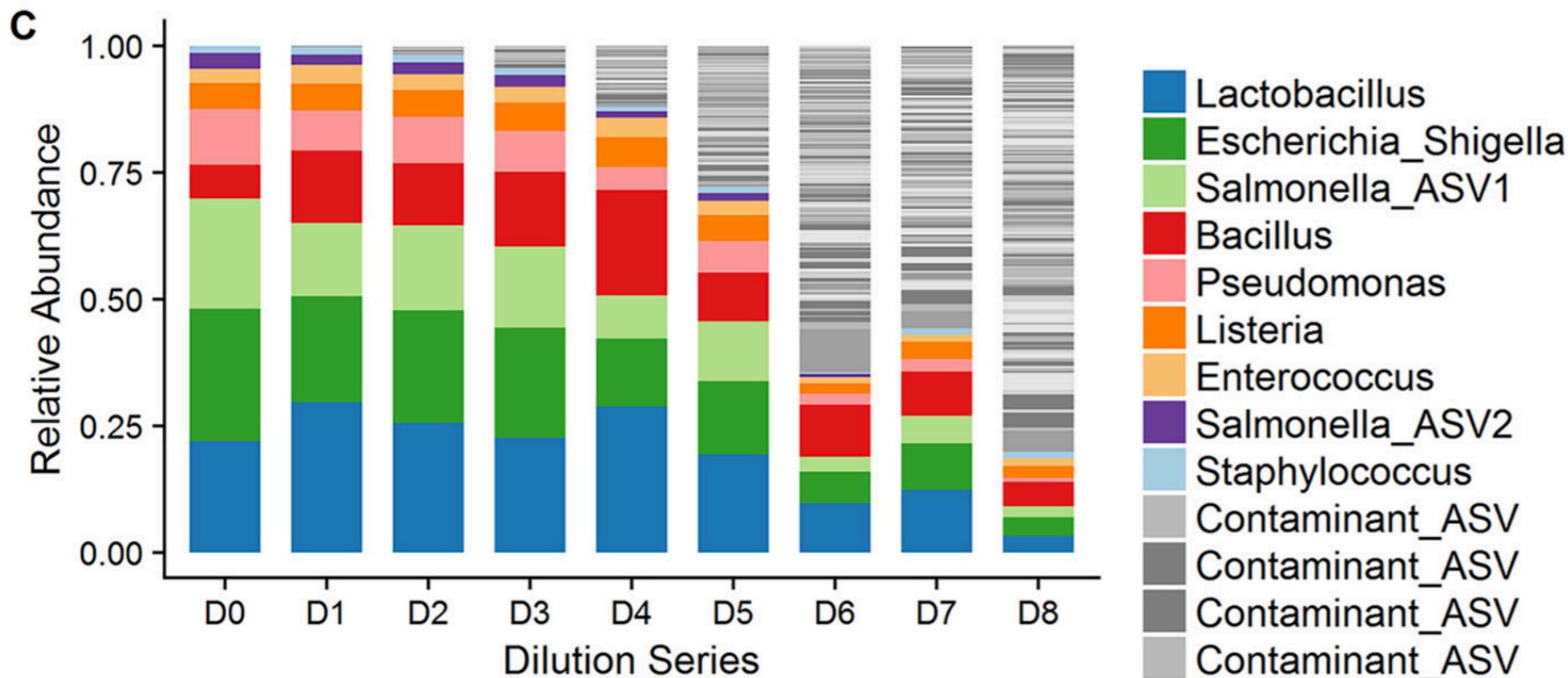


# Beyond defaults

**Better**

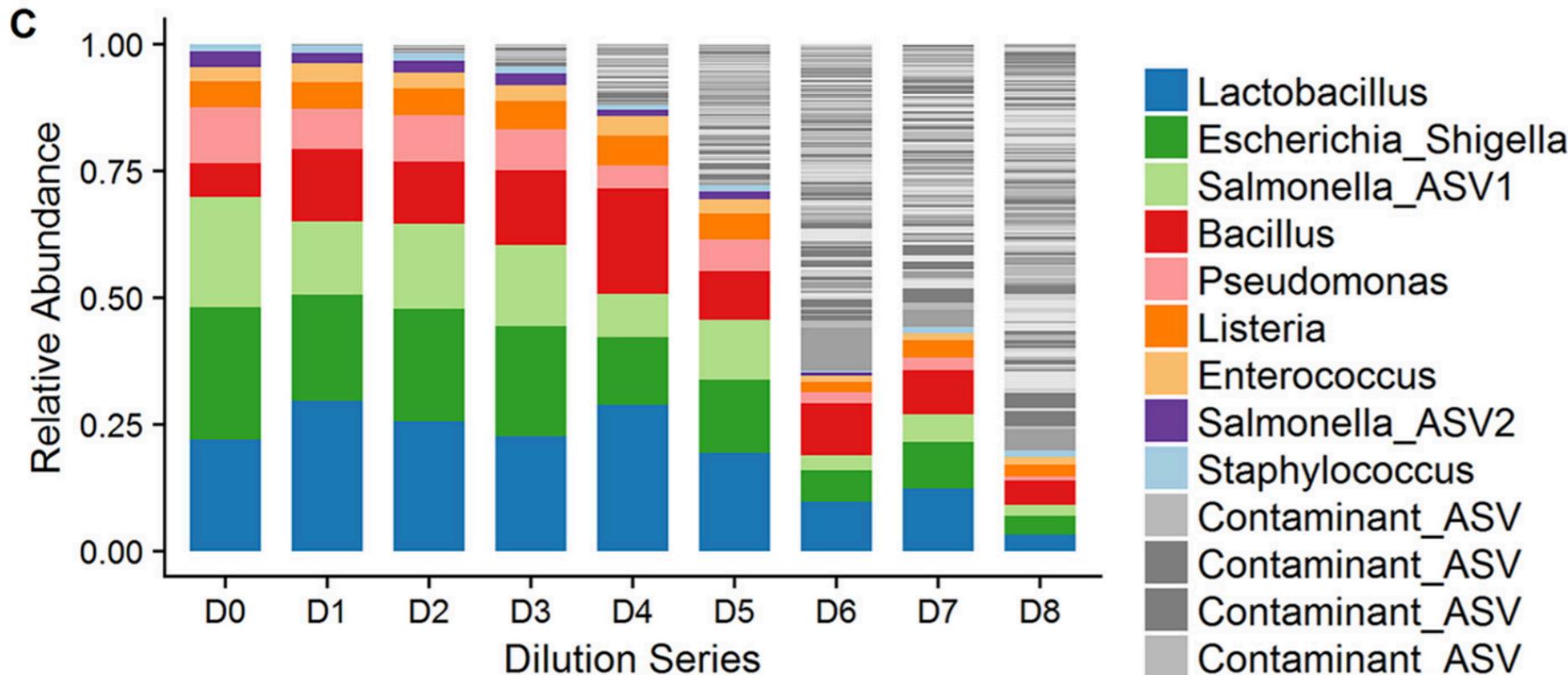


# When to care?



**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!**

# 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?**

# Growing options

Article | [Open Access](#) | Published: 10 November 2022

## De novo identification of microbial contaminants in low microbial biomass microbiomes with Squeegee

[Yunxi Liu](#), [R. A. Leo Elworth](#), [Michael D. Jochum](#), [Kjersti M. Aagaard](#) & [Todd J. Treangen](#) 

[Nature Communications](#) 13, Article number: 6799 (2022) | [Cite this article](#)

## Environmental DNA

[Open Access](#)

Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE | [Open Access](#) | 

**microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies**

[Donald T. McKnight](#)  [Roger Huerlimann](#), [Deborah S. Bower](#), [Lin Schwarzkopf](#), [Ross A. Alford](#), [Kyall R. Zenger](#)

First published: 16 May 2019 | <https://doi.org/10.1002/edn3.11> | Citations: 68

Article | Published: 16 March 2023

## Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data

[George I. Austin](#), [Heekuk Park](#), [Yoli Meydan](#), [Dwayne Seeram](#), [Tanya Sezin](#), [Yue Clare Lou](#), [Brian A. Firek](#), [Michael J. Morowitz](#), [Jillian F. Banfield](#), [Angela M. Christiano](#), [Itsik Pe'er](#), [Anne-Catrin Uhlemann](#), [Liat Shenhav](#)  & [Tal Korem](#) 

[Nature Biotechnology](#) (2023) | [Cite this article](#)

# Growing options

Article | [Open Access](#) | Published: 10 November 2022

## De novo identification of microbial contaminants in low microbial biomass microbiomes with Squeegee

[Yunxi Liu](#), [R. A. Leo Elworth](#), [Michael D. Jochum](#), [Kjersti M. Aagaard](#) & [Todd J. Treangen](#) 

[Nature Communications](#) 13, Article number: 6799 (2022) | [Cite this article](#)

## Environmental DNA

[Open Access](#)

Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE | [Open Access](#) | 

**microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies**

Donald T. McKnight , Roger Huerlimann, Deborah S. Bower, Lin Schwarzkopf, Ross A. Alford, Kyall R. Zenger

First published: 16 May 2019 | <https://doi.org/10.1002/edn3.11> | Citations: 68

Article | Published: 16 March 2023

## Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data

[George I. Austin](#), [Heekuk Park](#), [Yoli Meydan](#), [Dwayne Seeram](#), [Tanya Sezin](#), [Yue Clare Lou](#), [Brian A. Firek](#), [Michael J. Morowitz](#), [Jillian F. Banfield](#), [Angela M. Christiano](#), [Itsik Pe'er](#), [Anne-Catrin Uhlemann](#), [Liat Shenhav](#)  & [Tal Korem](#) 

[Nature Biotechnology](#) (2023) | [Cite this article](#)

What assumptions are these methods making?  
What additional data do these methods require?  
When is it appropriate to use these methods?

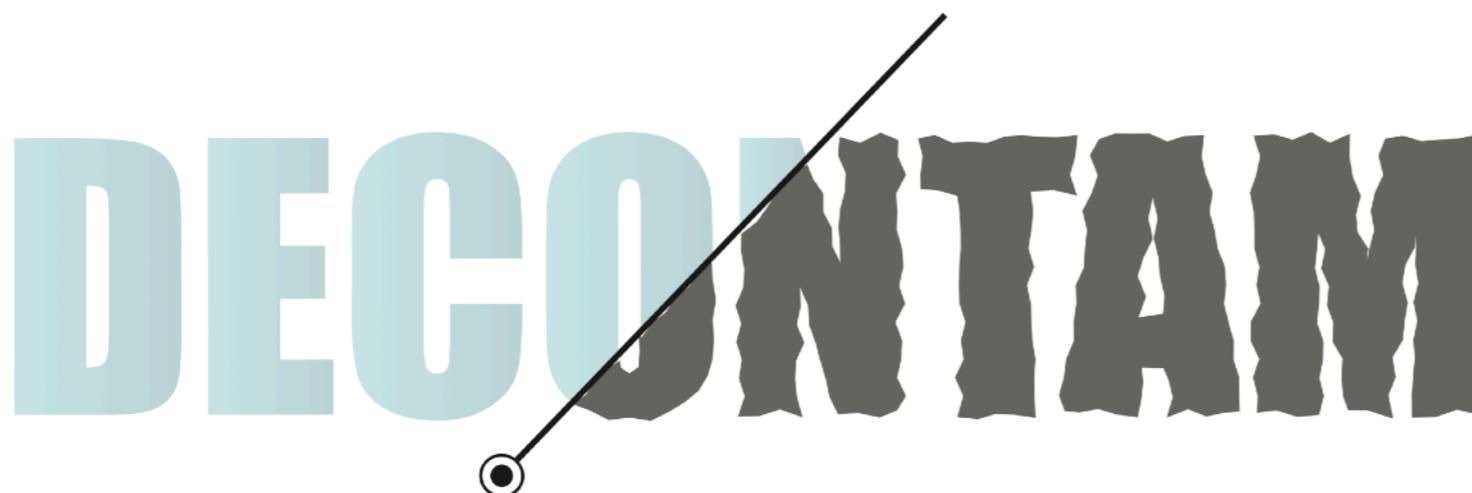
# Available now...

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 [✉](#) [ID](#)

*Microbiome* 2018 6:226



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

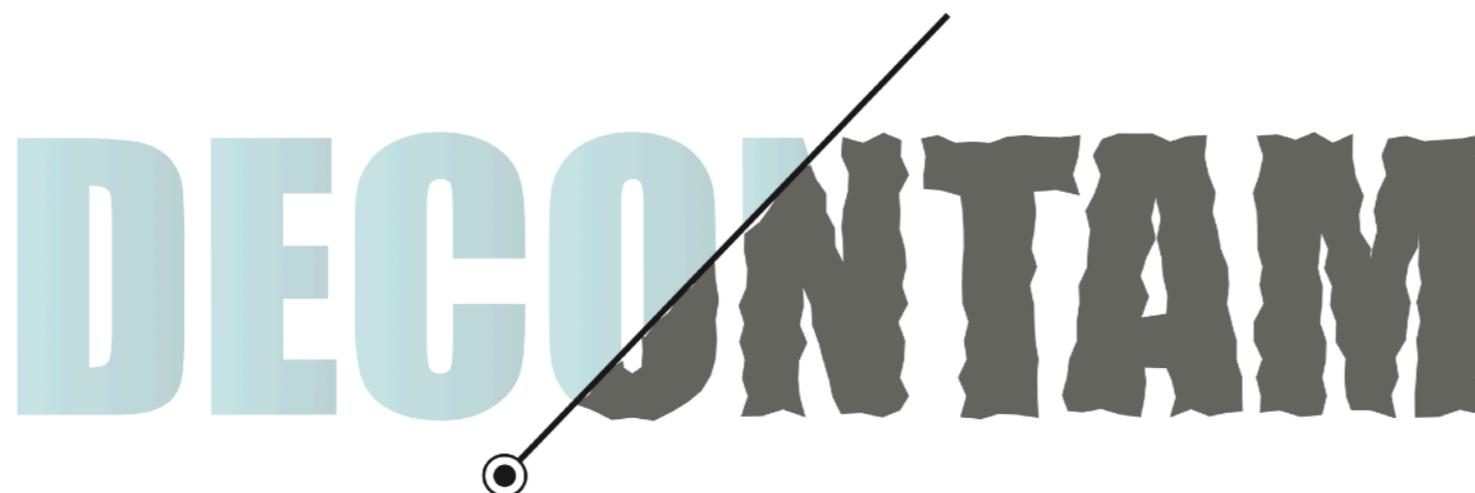
# Available now...

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 [✉](#) [ID](#)

*Microbiome* 2018 6:226



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

\*and recently, also via QIIME2!

# Package resources

## Manuscript

<https://doi.org/10.1186/s40168-018-0605-2>

## Accompanying analyses in R

<https://github.com/benjjneb/decontammanuscript>

## Vignette

[https://benjjneb.github.io/decontam/vignettes/decontam\\_intro.html](https://benjjneb.github.io/decontam/vignettes/decontam_intro.html)

## Github (and use Issues for support)

<https://github.com/benjjneb/decontam>

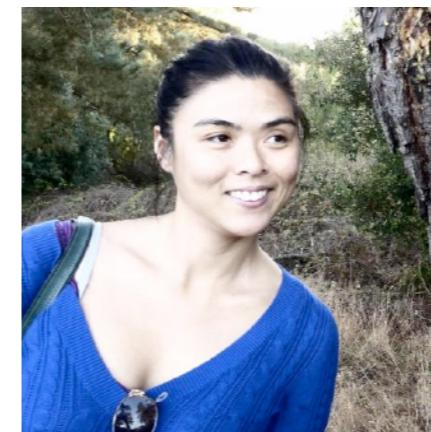
# Acknowledgements



Susan Holmes



Nicole Davis

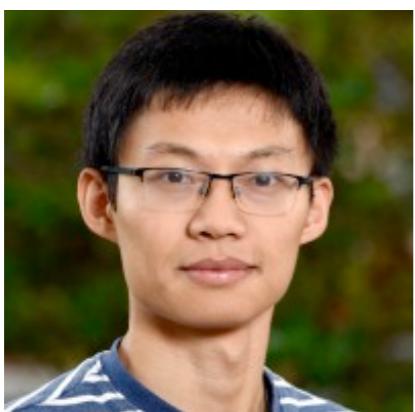


Diana Proctor



David Relman

*And more recent developments...*



Caizhi "David" Huang



Jorden Rabasco

