INTRODUCING THE HUMAN CONTAMINOME

Lisa Karstens, PhD

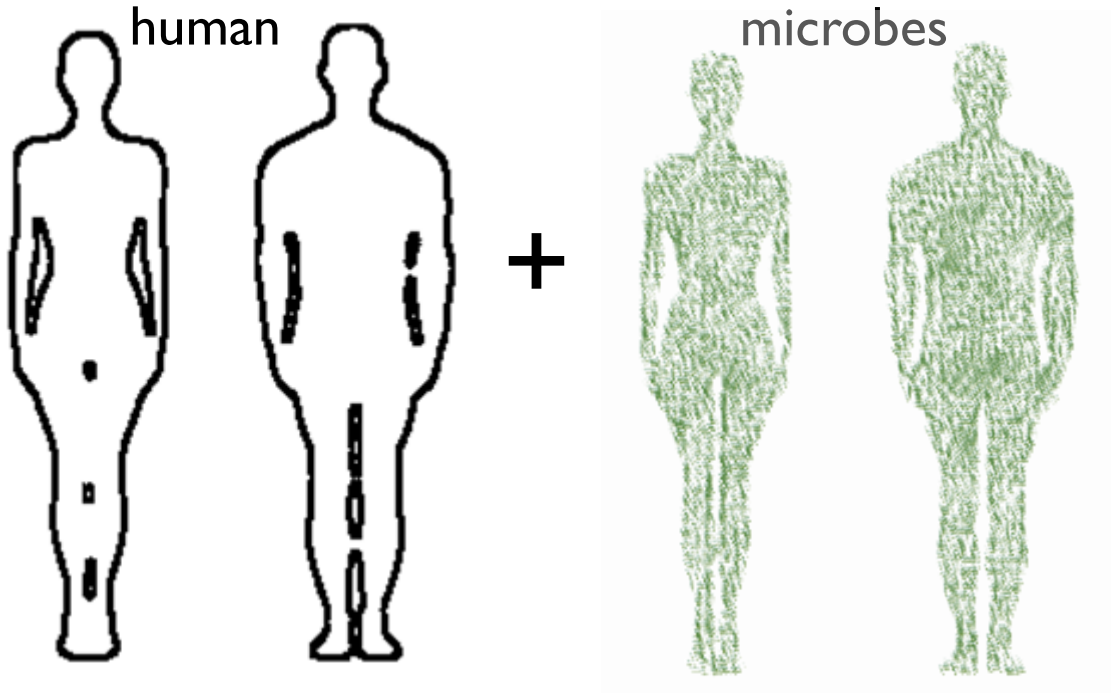
International Conference on Clinical Metagenomics

16 November 2023

CHALLENGE: MICROBES ARE EVERYWHERE!

- Overview of DNA contaminants
- Example experimental design for evaluating contaminants
- Example of contaminant removal
- Key take aways

Part human, part microbes

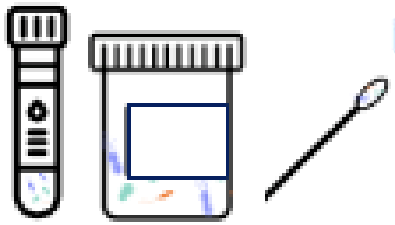


10^{14} human cells
23,000 human genes

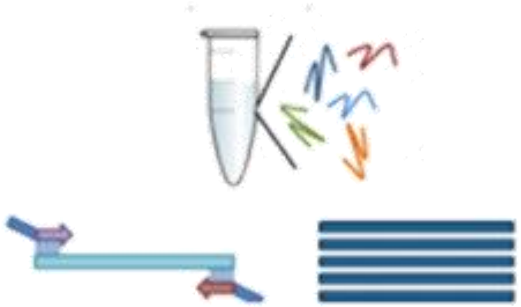
> 10^{14} microbial cells
2 – 10 million microbial genes

→ Without understanding the interactions between microbes and humans, it is impossible to obtain a complete picture of our biology

Sequencing microbial communities directly from patient samples



Patient sample



DNA extraction

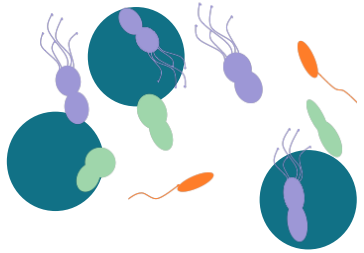
Sequencing

	S1
Bacteria 1	4
Bacteria 2	43
Bacteria 3	56
...	

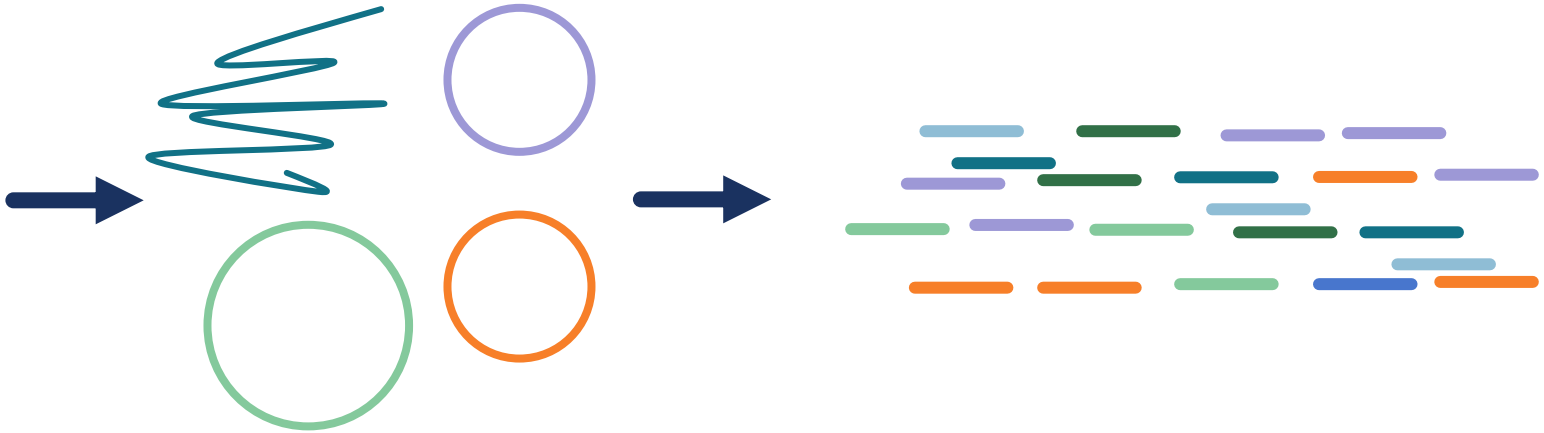
Annotation

Sequencing microbial communities directly from patient samples

Shotgun Metagenomics



Mixed microbial community



What it detects	What it reveals	Advantage	Disadvantage
Bacteria, fungi, viruses, human	Taxonomy All genes	Breadth, function	$ $ $ host depletion complex bioinformatics

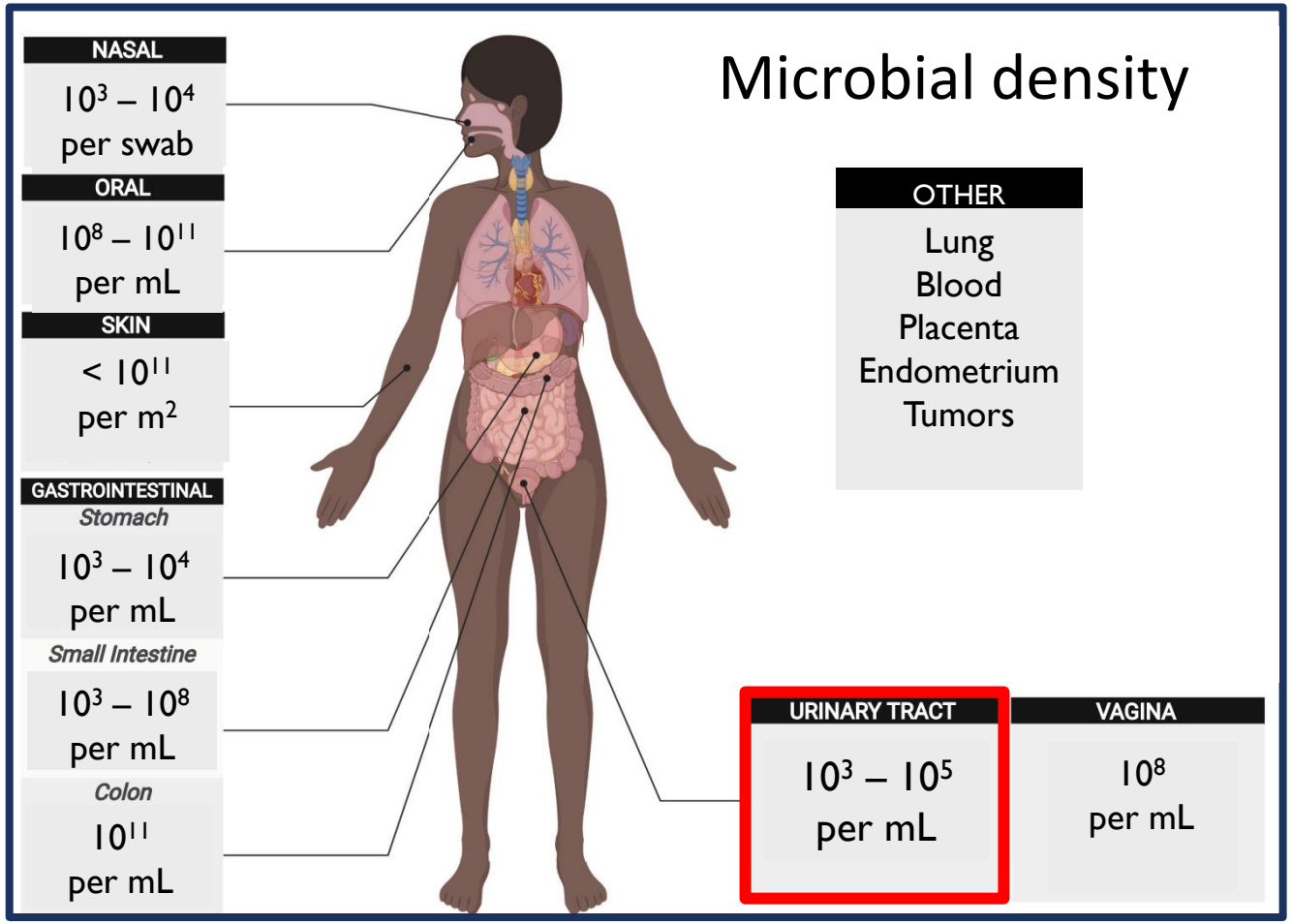
Sequencing microbial communities directly from patient samples

Marker gene sequencing

Mixed microbial community

What it detects	What it reveals	Advantage	Disadvantage
Bacteria only (16S rRNA gene) Fungi only (ITS or 18S rRNA gene)	Taxonomy	Low cost Increased depth No/Little host DNA	Limited view of the microbiome, Amplification bias Compositionality

Microbial load varies across body sites



Adapted from Neugent et al. mBio. (2020) PMID: 32345639

Gut = 10^{11}/mL

Vagina = 10^8/mL

Urinary tract = 10^3 -10^5/mL

Need to evaluate methods with microbial biomass load in mind

Contaminant bacterial DNA in sequenced-based microbiome studies

- Contaminant DNA is any microbial DNA that did not originate from the sample
- Several studies have reported bacteria found in negative controls
  - Extremophiles
    - *Deinococcus, Rubrobacter, …*
  - Lab-associated
    - *Pseudomonas, Stenotrophomonas, …*
  - Human-associated
    - *Lactobacillus, Escherichia, …*

Bacteria listed reported in: Salter et al 2014 PMID 25387460; Eisenhorf et al 2019, PMID: 30497919; Karstens et al 2019

Types of contaminant bacterial DNA in sequenced-based microbiome studies

Environmental Contaminants

- Bacteria from surfaces, materials, people when collecting and handling specimens
- Common taxa: *Propionibacterium acnes*, *Staphylococcus* …

Controls to detect:

- Empty tube / collection material at point of sample collection

Minimize by:

- Use best practices, wipe down surfaces, use gloves, laboratory coats

How to detect and remove:

- Evaluation of controls, associations of taxa with key metadata, decontam, other software

Types of contaminant bacterial DNA in sequenced-based microbiome studies

'Kitome' Contaminants

- Bacterial DNA from reagents and materials used in laboratory processing
- Common taxa: *Deinococcus*, *Pseudomonas*, *Stenotrophomonas*, *Xanthomonas*, …

Controls to detect:

- Extraction blanks, Mock microbial dilution series

Minimize by:

- Use DNA-free reagents and kits designed for low microbial biomass

How to detect and remove:

- Evaluation of controls, batch effect correction, decontam, SCRuB, other software

Types of contaminant bacterial DNA in sequenced-based microbiome studies

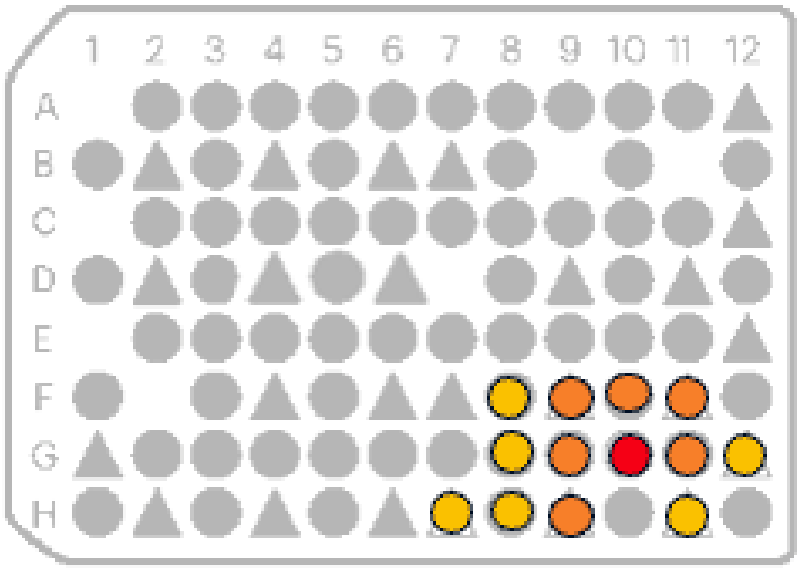


Image modified from Austin et al.
Nat Biotechnol 2023 Mar 16
PMID: 36928429

Well-to-well Contaminants

- Bacteria from other samples in adjacent samples due to sampling / sample handling / spillover
- Common taxa: Any in the dataset

Controls to detect:

- Mock community / Positive control

Minimize by:

- Process similar biomass specimens together
- Use a block design on 96-well plates

How to detect and remove:

- Individual taxa visualization, SCRuB

Types of contaminant bacterial DNA in sequenced-based microbiome studies

Image from: www.illumina.com/techniques/sequencing/ngs-library-prep/multiplexing/index-hopping.html

Index hopping

- Sequencing artifact
- Common taxa: Any in the dataset

Controls to detect:

- Mock community / Positive control

Minimize by:

- Use unique dual indexing pooling combos
- Remove free adapters from library preps

How to detect and remove:

- Abundance filter, Barcode error correction

Tools for evaluating contaminants

Amplicon

- Decontam

  Davis N et. al. *Microbiome* 226 (2018) PMID: 30558668

- microDecon

  McKnight, DT, et al *Environmental DNA*. 2019; 1: 14–25

- SCRuB: Source-tracking for Contamination Removal in *micro*Biomes

  Austin et al. *Nat Biotechnol* 2023 Mar 16 PMID: 3692842

- ConQuR

  Ling W et al. Nature Comm 2022 Sep 15;13(1):5418 PMID: 36109499

Shotgun Metagenomics

- Squeegee

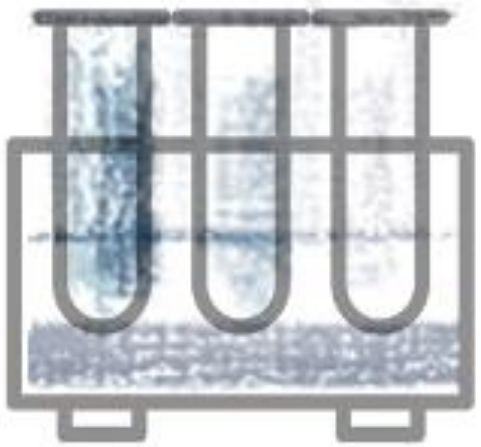
  Lui et al *Nature Comm* 2022 Nov 10;13(1) PMID: 36357382

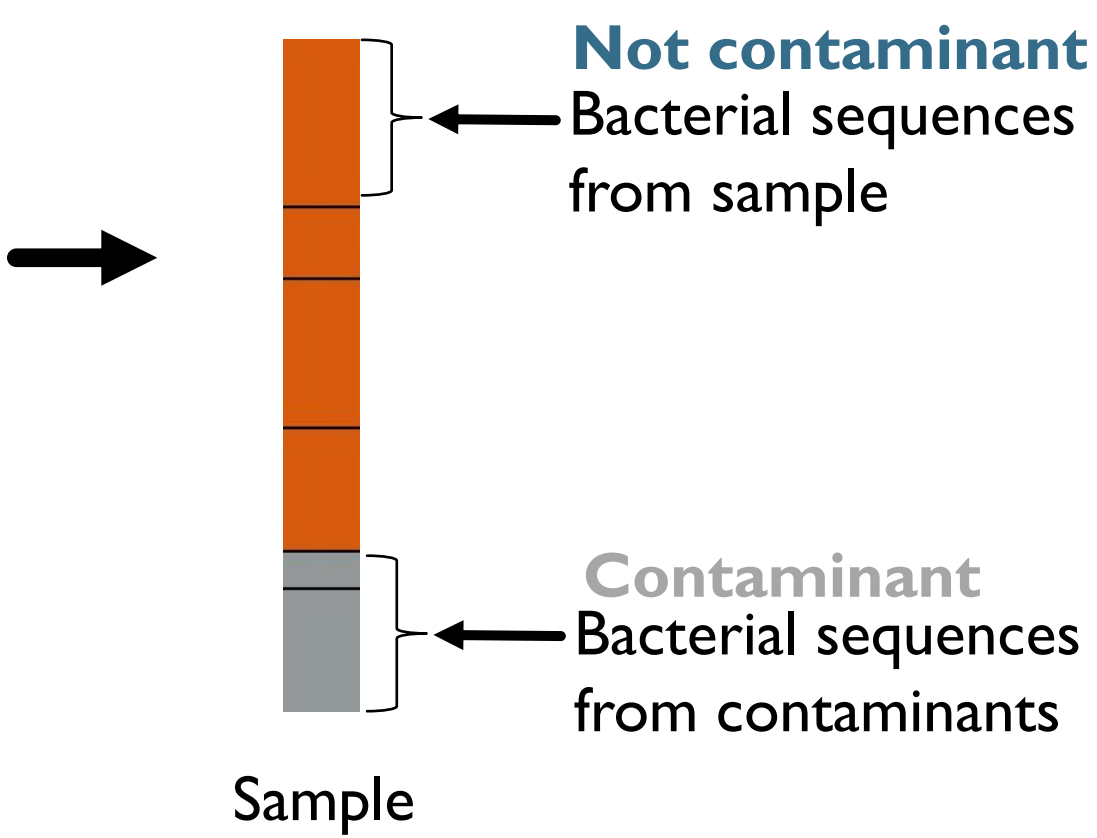
Visualization

- GRIMER –

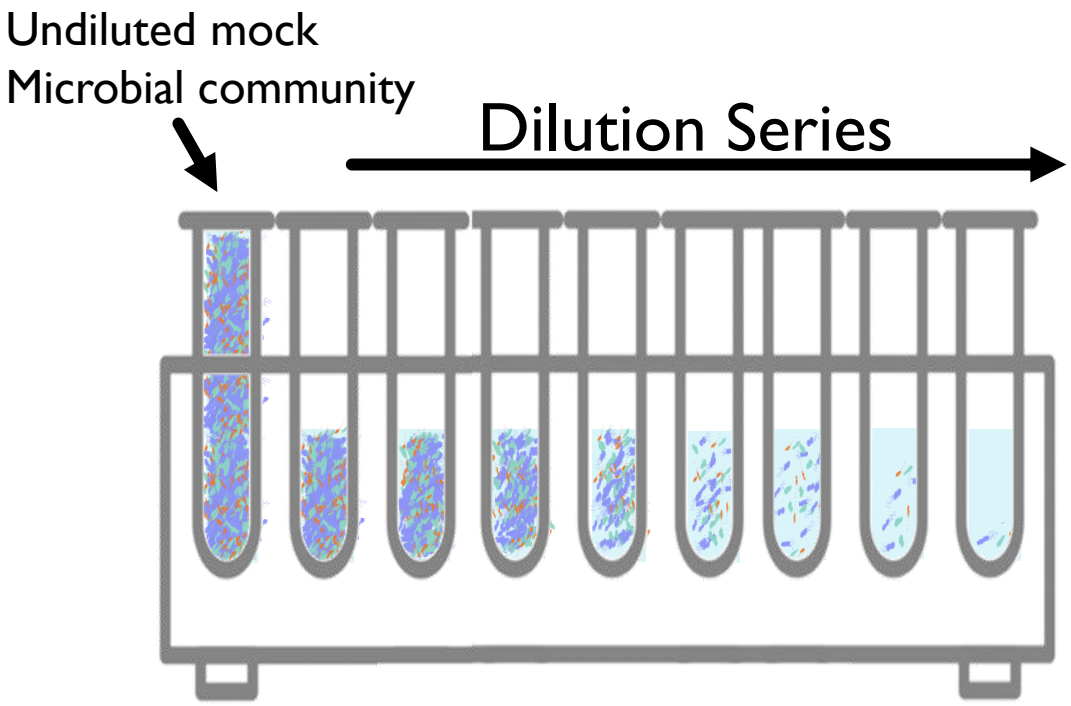
  Piro & Renard, GigaScience, Volume 12, 2023, giad017 https://pirovc.github.io/grimer/

Experimental design to detect contaminants





Computational identification of contaminant DNA

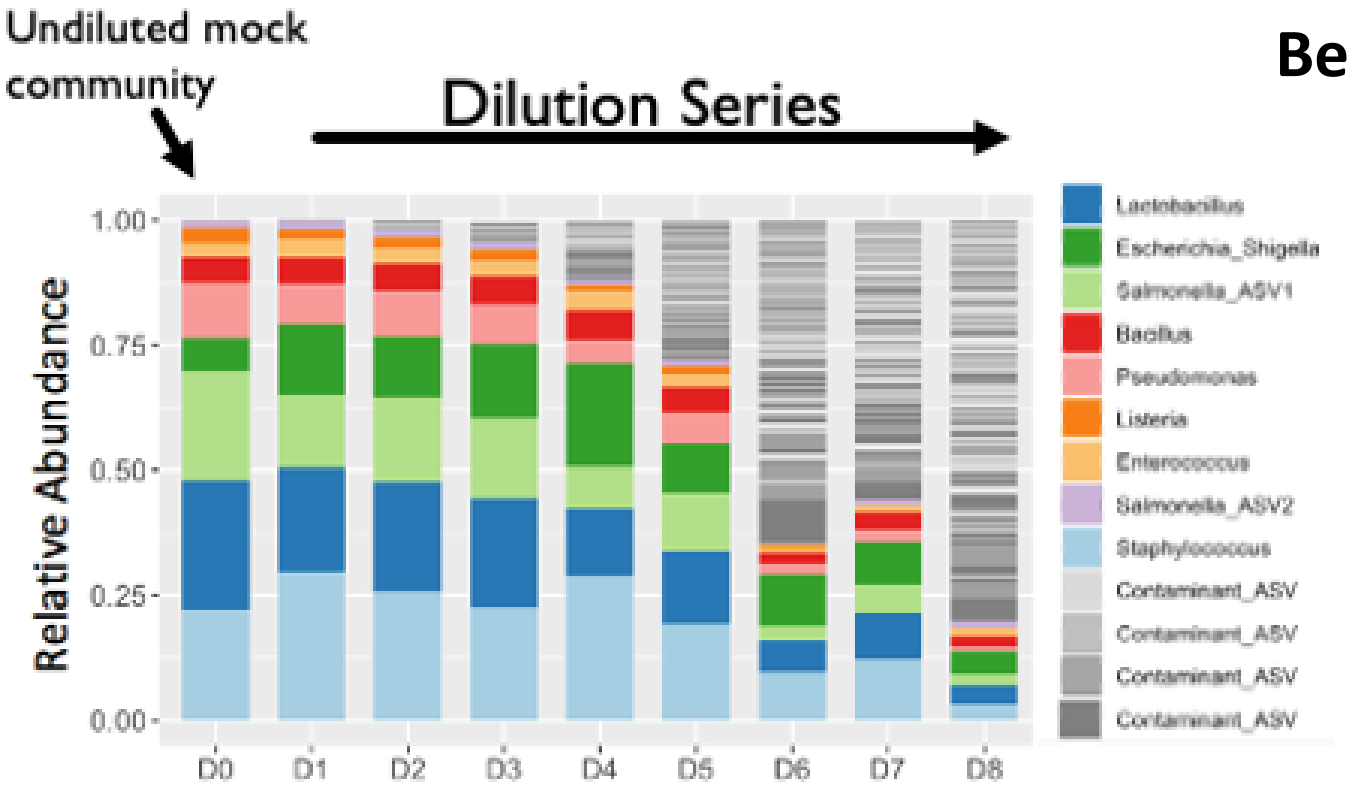


Benchmarking study:

- Mock microbial dilution series
- Extracted DNA and sequenced along with biological samples
- Assess accuracy
- Guide bioinformatics

Karstens et al **mSystems** 4(4): e00290-19 (2019))

Computational identification of contaminant DNA

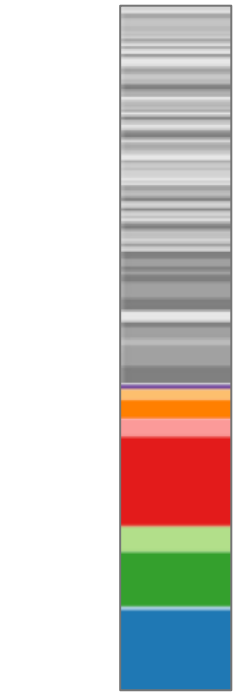


Benchmarking study:

- Unexpected sequences increase as microbial biomass decreases
- Has a major impact on diversity measures and bacterial abundances

Karstens et al **mSystems** 4(4): e00290-19 (2019))

Computational identification of contaminant DNA



Experimental sample

Negative control filtering

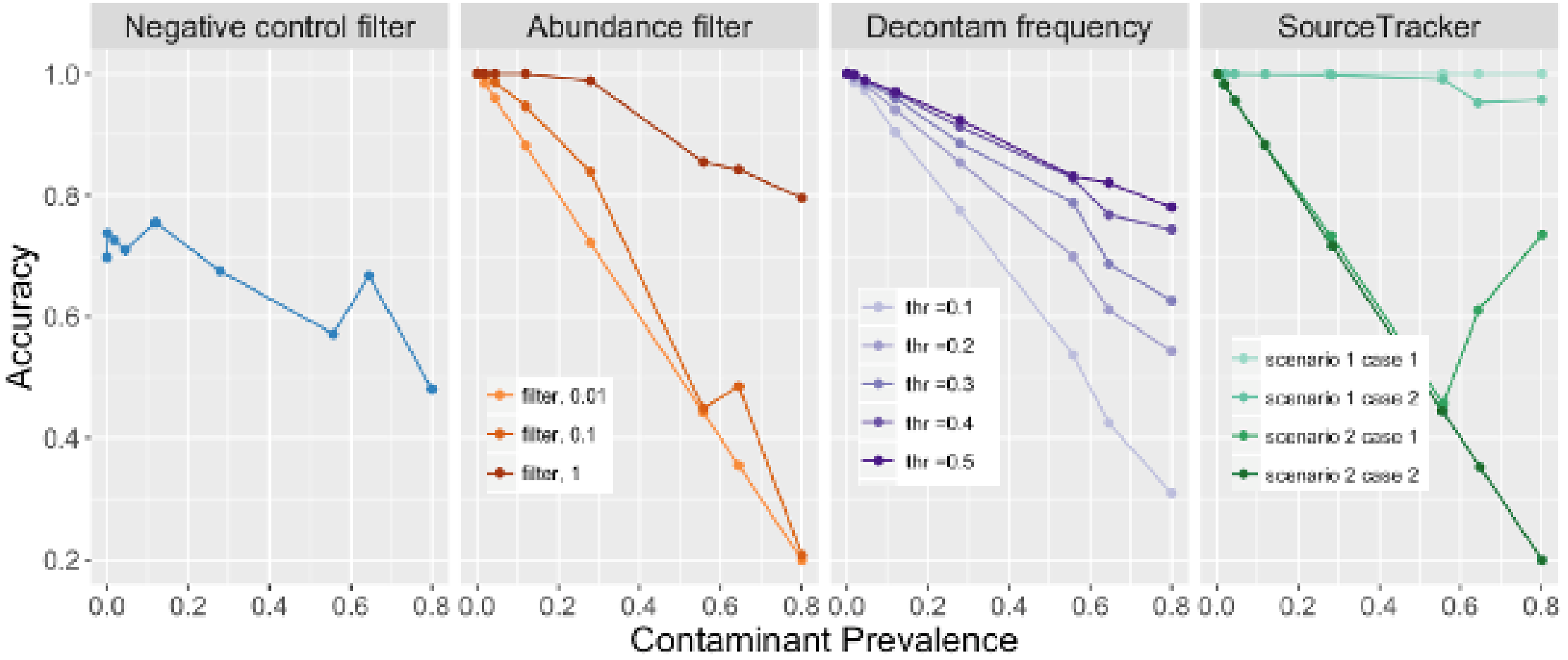
Abundance filtering

Decontam

Sourcetracker

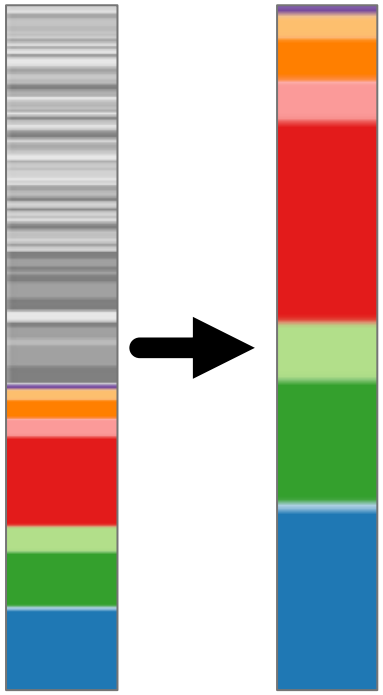
Corrected sample

Computational identification of contaminant DNA



Karstens et al **mSystems** 4(4): e00290-19 (2019))

Computational identification of contaminant DNA



No method is perfect.

Best approaches require additional information

- Decontam (frequency) requires DNA measured for each sample
- Decontam (prevalence) requires at least 5 negative controls
- SourceTracker requires a well-defined communities

Likely need more than one approach is needed

No current available method accounts for all types of contaminants

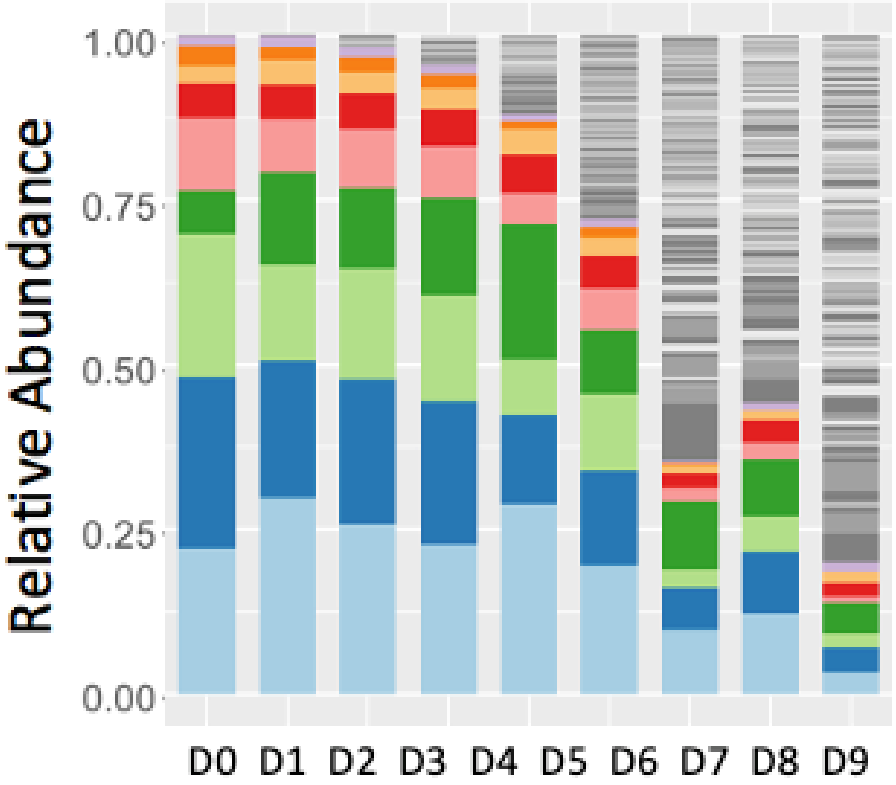
Parameterization is experiment dependent

- Need controls and exploratory data analysis to evaluate
- Need to prioritize goals of the study

Karstens et al **mSystems** 4(4): e00290-19 (2019))

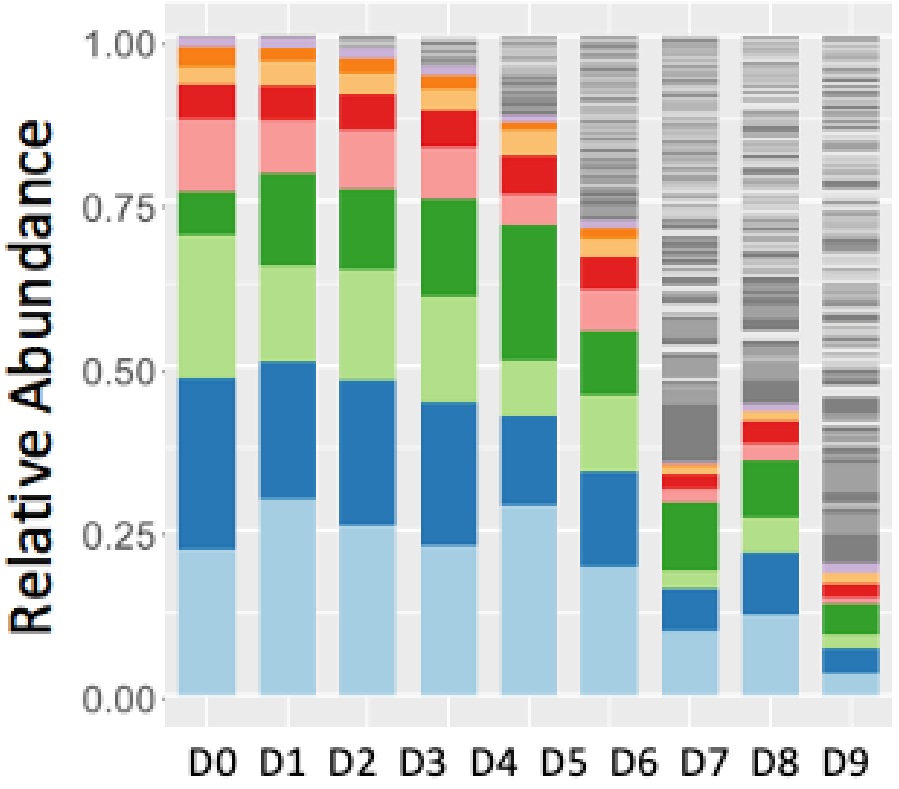
Take home points

Many methods exist and are being developed to handle contaminants



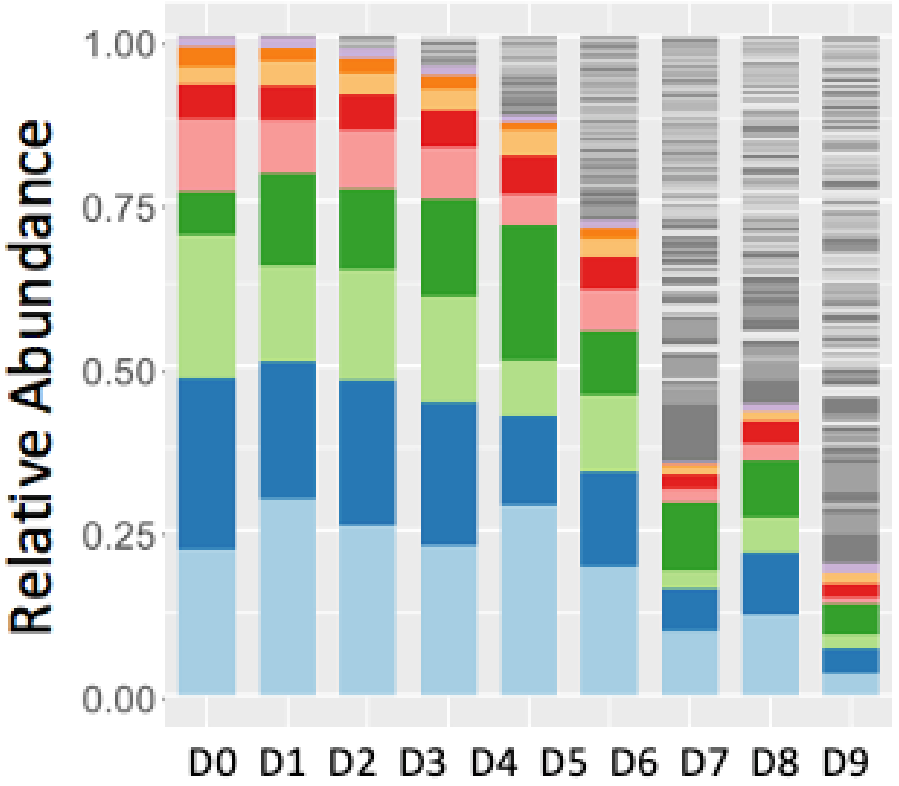
- Appropriately applying these methods requires:
  - Good experimental design (Positive AND Negative controls)
  - Thorough exploratory data analysis
  - A team effort (reviewing/discussing results)
  - Patience

Take home points



- **Contaminants can be problematic for sequenced-based microbiome studies.** Especially when:
  - Sample type has a wide range of possible biomass
  - Biomass is linked to the outcome of interest
- **Interpretation is key**
  - Association with contamination indicates a change in microbial biomass / change in microbiome structure
  - Presence of DNA is NOT always evidence of a microbiome BUT can still be meaningful

Steps to improve rigor of sequence-based microbiome studies



- **Transparency when publishing findings is key**
  - Critical assessment is needed by both authors and reviewers
  - Limitations should be clearly stated
- **Use community developed checklists:**
  - STORMS (Mirzayi et al 2021, PMID 34789871; www.stormsmicrobiome.org)
  - RIDE (Eisenhorf et al 2019, PMID: 30497919)
- **Share data from negative and positive controls**
  - Enable re-use by method developers
  - Enable large data libraries for AI/ML

Thank You

Acknowledgements

Students & Staff

Current:

Kate Bowie (PhD candidate)

Rebecca Wetzel (MS candidate)

Past:

Alex Rouhier (MS)

Matt Hammer (MS)

JP Gourdine, PhD

Erin Dahl

Carter Hoffman (MS)

Vincent Caruso (MS)

Eric Leung (MS)

Alec Barstad (EXITO)

Fatoumata Jallow (EXITO)

Emory Neer (EXITO)

Collaborators

Mark Asquith, PhD

Alan Wolfe, PhD

Linda Brubaker, MD

Nazema Siddiqui, MD

Rahel Nardos, MD, MCR

Ian Fields, MD

Tanya Sysoeva, PhD

Maryellen Kelly, DNP

Shannon McWeeney, PhD

Jim Rosenbaum, MD

Tom Gregory, MD

Jonathan Braun, MD PhD

Asquith/Rosenbaum Lab

Women's Health Research Unit

Funding

ISAC INNOVATIVE SCIENCE ACCELERATOR PROGRAM

NIH National Institute of Diabetes and Digestive and Kidney Diseases

K01 DK116706

Center for WOMEN'S HEALTH Circle of Giving Driving Innovation in Women's Health

K12 HD043488

Oregon birch W NIH Interdisciplinary Research Careers in Women's Health

Collins Medical Trust

microshades

karstens@ohsu.edu

@lisakarstens

https://bit.ly/KarstensLab