

# Designing a microbiome study: ecological and technical considerations

ANSC 516

# Today's goals

- How to design and carry out a microbiome study?
  - Experimental treatment groups
  - Biological and technical controls
- Sequencing library preparation
- The “how” of microbiome studies.
- Community succession



# Controls



REVIEW ARTICLE

## Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design

Debby Laukens<sup>1,†</sup>, Brigitte M. Brinkman<sup>2,3,†</sup>, Jeroen Raes<sup>4,5</sup>, Martine De Vos<sup>1</sup> and Peter Vandenebeele<sup>2,3,6,\*</sup>

<sup>1</sup>Department of Gastroenterology, Ghent University, B-9000 Ghent, Belgium. <sup>2</sup>Inflammation Research Center,

Downloaded from https://academic.oup.com/femsre/article/40/1/117/2866036 by guest on 08 August 2016

## PERSPECTIVE

## Unlocking the potential of metagenomics through replicated experimental design

Rob Knight<sup>1</sup>, Janet Jansson<sup>2–4</sup>, Dawn Field<sup>5</sup>, Noah Fierer<sup>6</sup>, Narayan Desai<sup>7</sup>, Jed A Fuhrman<sup>8</sup>, Phil Hugenholtz<sup>9</sup>, Daniel van der Lelie<sup>10</sup>, Folker Meyer<sup>7,11</sup>, Rick Stevens<sup>7,11</sup>, Mark J Bailey<sup>5</sup>, Jeffrey I Gordon<sup>12</sup>, George A Kowalchuk<sup>13,14</sup> & Jack A Gilbert<sup>7,15</sup>

## Skin Microbiome Surveys Are Strongly Influenced by Experimental Design

Jacquelyn S. Meisel<sup>1</sup>, Geoffrey D. Hannigan<sup>1</sup>, Amanda S. Tyldsley<sup>1</sup>, Adam J. SanMiguel<sup>1</sup>, Brendan P. Hodkinson<sup>1</sup>, Qi Zheng<sup>1</sup> and Elizabeth A. Grice<sup>1</sup>

Brooks et al. *BMC Microbiology* (2015) 15:66  
DOI 10.1186/s12866-015-0351-6



## METHODOLOGY ARTICLE

## Open Access

## The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies

J Paul Brooks<sup>1,2\*</sup>, David J Edwards<sup>1</sup>, Michael D Harwiche Jr<sup>3</sup>, Maria C Rivera<sup>4</sup>, Jennifer M Fettweis<sup>3</sup>, Myrna G Serrano<sup>2,3</sup>, Robert A Reris<sup>1</sup>, Nihar U Sheth<sup>2</sup>, Bernice Huang<sup>3</sup>, Philippe Girerd<sup>5</sup>, Vaginal Microbiome Consortium (additional members), Jerome F Strauss III<sup>5</sup>, Kimberly K Jefferson<sup>2,3</sup> and Gregory A Buck<sup>2,3</sup>

## ANALYSIS

## Towards standards for human fecal sample processing in metagenomic studies

Paul I Costea<sup>1</sup>, Georg Zeller<sup>1</sup>, Shinichi Sunagawa<sup>1,2</sup>, Eric Pelletier<sup>3–5</sup>, Adriana Alberti<sup>3</sup>, Florence Levenez<sup>6</sup>, Melanier Tramontano<sup>1</sup>, Marja Driessens<sup>1</sup>, Rajna Hercog<sup>1</sup>, Ferris-Elias Jung<sup>1</sup>, Jens Roat Kultima<sup>1</sup>, Matthew R Hayward<sup>1</sup>, Luis Pedro Coelho<sup>1</sup>, Emma Allen-Vercoe<sup>7</sup>, Laurie Bertrand<sup>3</sup>, Michael Blaut<sup>8</sup>, Jillian R M Brown<sup>9</sup>, Thomas Carton<sup>10</sup>, Stéphanie Cools-Portier<sup>11</sup>, Michelle Daigneault<sup>6</sup>, Muriel Derrien<sup>11</sup>, Anne Druesne<sup>11</sup>, Willem M de Vos<sup>12,13</sup>, Brett Finlay<sup>14</sup>, Harry J Flint<sup>15</sup>, Francisco Guarner<sup>16</sup>, Masahira Hattori<sup>17,18</sup>, Hans Heilig<sup>12</sup>, Ruth Ann Luna<sup>19</sup>, Johan van Hylckama Vlieg<sup>11</sup>, Jana Junick<sup>8</sup>, Ingeborg Klymiuk<sup>20</sup>, Philippe Langella<sup>6</sup>, Emmanuelle Le Chatelier<sup>6</sup>, Volker Mai<sup>21</sup>, Chaysavanh Manichanh<sup>16</sup>, Jennifer C Martin<sup>15</sup>, Clémentine Mery<sup>10</sup>, Hidetoshi Morita<sup>22</sup>, Paul W O'Toole<sup>9</sup>, Céline Orvain<sup>3</sup>, Kiran Raosaheb Patil<sup>1</sup>, John Penders<sup>23</sup>, Søren Persson<sup>24</sup>, Nicolas Pons<sup>6</sup>, Milena Popova<sup>10</sup>, Anne Salonen<sup>13</sup>, Delphine Saulnier<sup>8</sup>, Karen P Scott<sup>15</sup>, Bhagirath Singh<sup>25</sup>, Kathleen Slezak<sup>8</sup>, Patrick Veiga<sup>11</sup>, James Versalovic<sup>19</sup>, Liping Zhao<sup>26</sup>, Erwin G Zoetendal<sup>12</sup>, S Dusko Ehrlich<sup>6,27</sup>, Joel Dore<sup>6</sup> & Peer Bork<sup>1,28–30</sup>



# Controls



REVIEW ARTICLE

## Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design

Debby Laukens<sup>1,†</sup>, Brigitte M. Brinkman<sup>2,3,†</sup>, Jeroen Raes<sup>4,5</sup>, Martine De Vos<sup>1</sup> and Peter Vandenebeele<sup>2,3,6,\*</sup>

<sup>1</sup>Department of Gastroenterology, Ghent University, B-9000 Ghent, Belgium. <sup>2</sup>Inflammation Research Center,

- Age
- Diet
- Sex
- Antibiotic use

FEMS Microbiology Reviews, fuv036, 40, 2016, 117–132

doi: 10.1093/femsre/fuv036

Advance Access Publication Date: 30 August 2015  
Review Article

Downloaded from https://academic.oup.com/femsre

## PERSPECTIVE

## Unlocking the potential of metagenomics through replicated experimental design

Rob Knight<sup>1</sup>, Janet Jansson<sup>2–4</sup>, Dawn Field<sup>5</sup>, Noah Fierer<sup>6</sup>, Narayan Desai<sup>7</sup>, Jed A Fuhrman<sup>8</sup>, Phil Hugenholtz<sup>9</sup>, Daniel van der Lelie<sup>10</sup>, Folker Meyer<sup>7,11</sup>, Rick Stevens<sup>7,11</sup>, Mark J Bailey<sup>5</sup>, Jeffrey I Gordon<sup>12</sup>, George A Kowalchuk<sup>13,14</sup> & Jack A Gilbert<sup>7,15</sup>

## Skin Microbiome Surveys Are Strongly Influenced by Experimental Design

Jacquelyn S. Meisel<sup>1</sup>, Geoffrey D. Hannigan<sup>1</sup>, Amanda S. Tyldsley<sup>1</sup>, Adam J. SanMiguel<sup>1</sup>, Brendan P. Hodkinson<sup>1</sup>, Qi Zheng<sup>1</sup> and Elizabeth A. Grice<sup>1</sup>

Brooks et al. *BMC Microbiology* (2015) 15:66  
DOI 10.1186/s12866-015-0351-6



## METHODOLOGY ARTICLE

## Open Access

## The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies

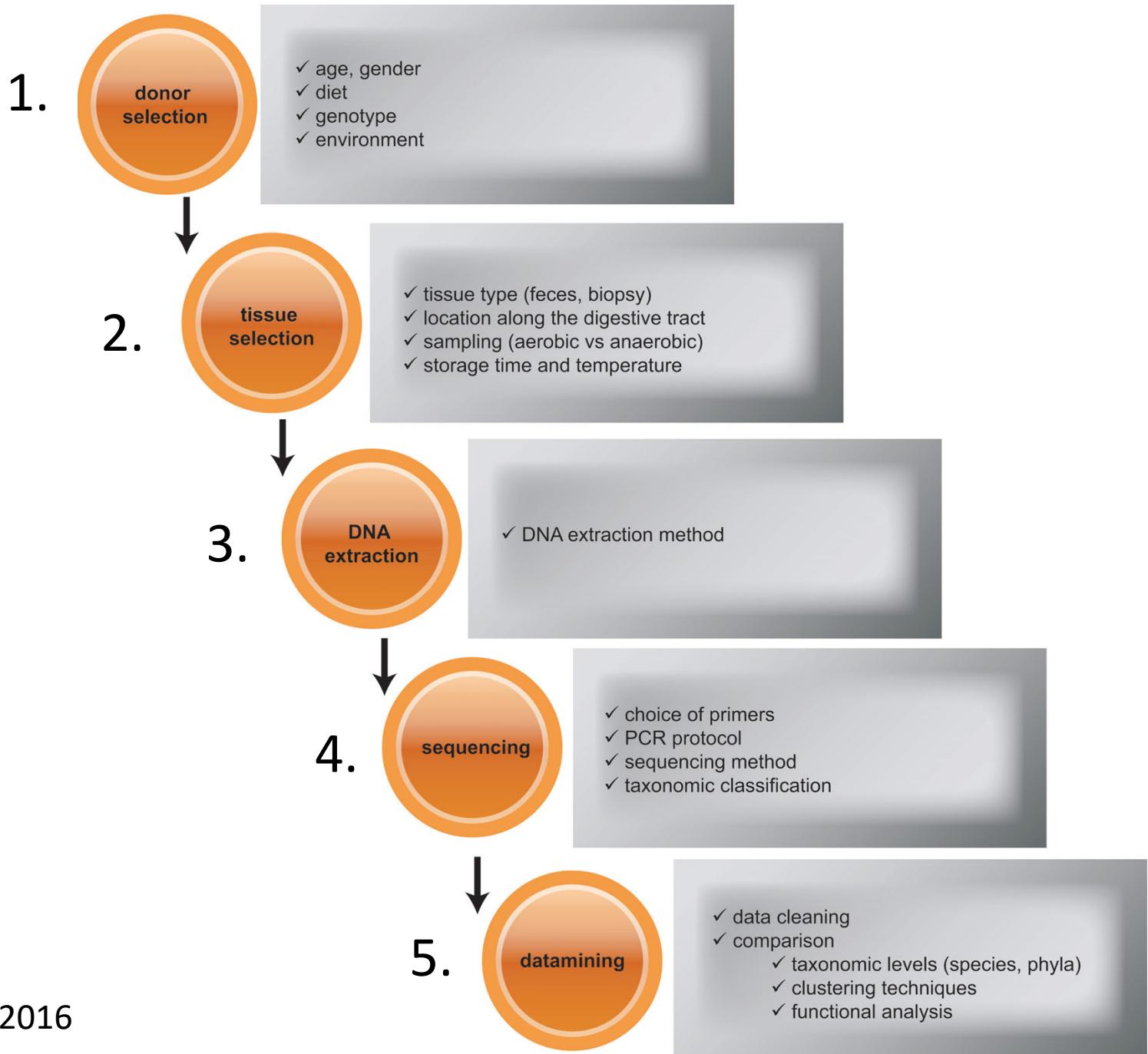
J Paul Brooks<sup>1,2\*</sup>, David J Edwards<sup>1</sup>, Michael D Harwiche Jr<sup>3</sup>, Maria C Rivera<sup>4</sup>, Jennifer M Fettweis<sup>3</sup>, Myrna G Serrano<sup>2,3</sup>, Robert A Reris<sup>1</sup>, Nihar U Sheth<sup>2</sup>, Bernice Huang<sup>3</sup>, Philippe Girerd<sup>5</sup>, Vaginal Microbiome Consortium (additional members), Jerome F Strauss III<sup>5</sup>, Kimberly K Jefferson<sup>2,3</sup> and Gregory A Buck<sup>2,3</sup>

## ANALYSIS

## Towards standards for human fecal sample processing in metagenomic studies

Paul I Costea<sup>1</sup>, Georg Zeller<sup>1</sup>, Shinichi Sunagawa<sup>1,2</sup>, Eric Pelletier<sup>3–5</sup>, Adriana Alberti<sup>3</sup>, Florence Levenez<sup>6</sup>, Melanie Tramontano<sup>1</sup>, Marja Driessens<sup>1</sup>, Rajna Hercog<sup>1</sup>, Ferris-Elias Jung<sup>1</sup>, Jens Roat Kultima<sup>1</sup>, Matthew R Hayward<sup>1</sup>, Luis Pedro Coelho<sup>1</sup>, Emma Allen-Vercoe<sup>7</sup>, Laurie Bertrand<sup>3</sup>, Michael Blaut<sup>8</sup>, Jillian R M Brown<sup>9</sup>, Thomas Carton<sup>10</sup>, Stéphanie Cools-Portier<sup>11</sup>, Michelle Daigneault<sup>6</sup>, Muriel Derrien<sup>11</sup>, Anne Druesne<sup>11</sup>, Willem M de Vos<sup>12,13</sup>, Brett Finlay<sup>14</sup>, Harry J Flint<sup>15</sup>, Francisco Guarner<sup>16</sup>, Masahira Hattori<sup>17,18</sup>, Hans Heilig<sup>12</sup>, Ruth Ann Luna<sup>19</sup>, Johan van Hylckama Vlieg<sup>11</sup>, Jana Junick<sup>8</sup>, Ingeborg Klymiuk<sup>20</sup>, Philippe Langella<sup>6</sup>, Emmanuelle Le Chatelier<sup>6</sup>, Volker Mai<sup>21</sup>, Chaysavanh Manichanh<sup>16</sup>, Jennifer C Martin<sup>15</sup>, Clémentine Mery<sup>10</sup>, Hidetoshi Morita<sup>22</sup>, Paul W O'Toole<sup>9</sup>, Céline Orvain<sup>3</sup>, Kiran Raosaheb Patil<sup>1</sup>, John Penders<sup>23</sup>, Søren Persson<sup>24</sup>, Nicolas Pons<sup>6</sup>, Milena Popova<sup>10</sup>, Anne Salonen<sup>13</sup>, Delphine Saulnier<sup>8</sup>, Karen P Scott<sup>15</sup>, Bhagirath Singh<sup>25</sup>, Kathleen Slezak<sup>8</sup>, Patrick Veiga<sup>11</sup>, James Versalovic<sup>19</sup>, Liping Zhao<sup>26</sup>, Erwin G Zoetendal<sup>12</sup>, S Dusko Ehrlich<sup>6,27</sup>, Joel Dore<sup>6</sup> & Peer Bork<sup>1,28–30</sup>

# Experimental Design



# Animal housing guidelines

1.



- ✓ age, gender
- ✓ diet
- ✓ genotype
- ✓ environment

Table 1. Guidelines to control variations in microbiota composition in mice.

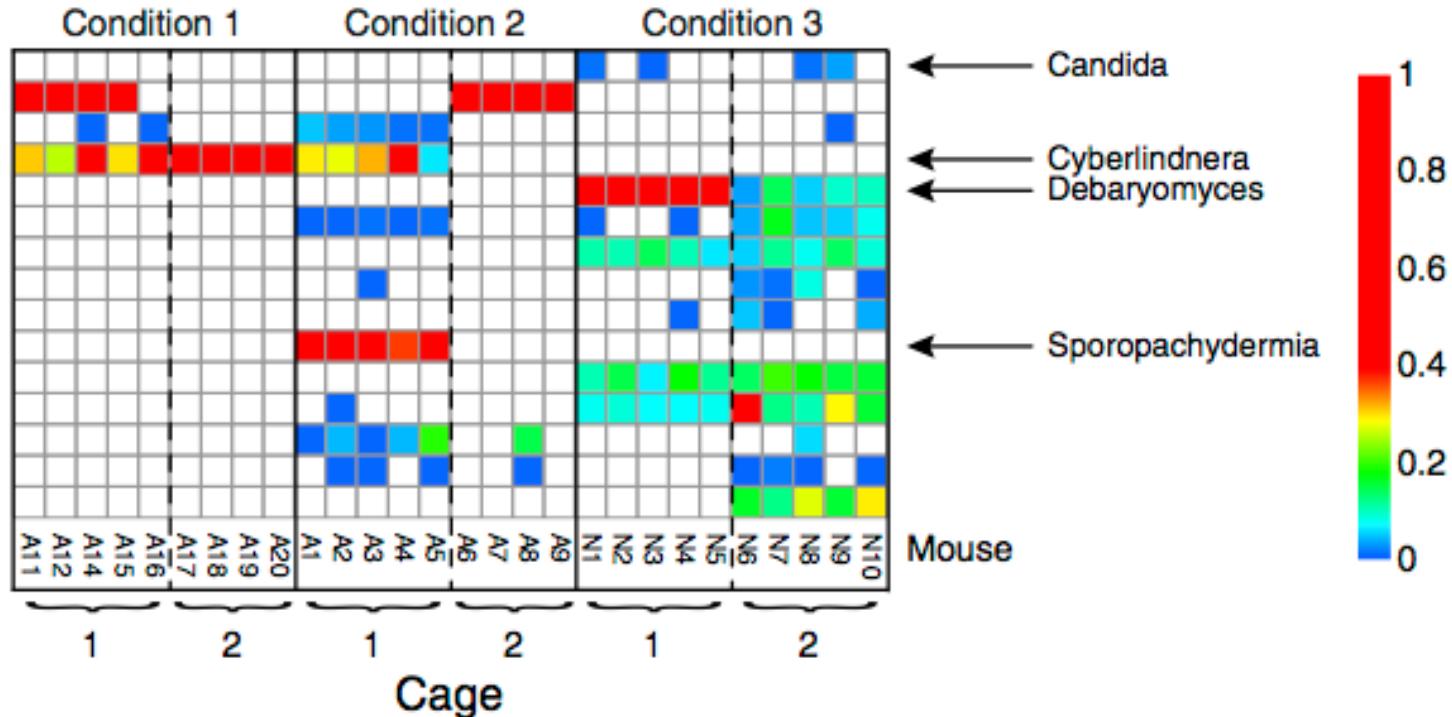
Guideline	Variable that influences microbiota composition	Possible complication
<b>General guidelines</b>		
Selective breeding of siblings over several generations	Maternal transmission	Genetic drift
Standardize diet and food autoclaving parameters	Diet	Standardization fallacy
Keep mice together (same room, same rack) and do not relocate cages	Environment, stress	Logistic problems
Minimize noise, handling time, stress to set hierarchy	Stress	
Maximize number of cages	Cage effect	
Collect tissue or fecal pellets for microbiome analysis		
<b>Study the effect of treatment (excluding probiotics)</b>		
If possible, use isobiotic mice and keep in individually ventilated cages	Origin of mice	
Or homogenize the flora by cohousing mice 3–4 weeks before the start of the experiment	Cage effect	
Mix treatment groups within each cage	Cage effect	
<b>Study the effect of probiotics</b>		
Homogenize the flora and minimize genetic influences by switching mice and distribute littermates over cages just before the start of the experiment	Cage effect	
Maximize the number of cages	Cage effect	
<b>Study the effect of host genetics on a common microbial background</b>		
Use heterozygous littermates as reference group	Maternal transmission	
Homogenize the flora by cohousing mice 3–4 weeks before the start of the experiment	Cage effect	
<b>Study the effect of host genetics on microbiota composition</b>		
Use heterozygous littermates as reference group	Maternal transmission	
Separate litters according to genotype after weaning, divide over several cages (in case of subtle differences)	Cage effect, age effect	Synchronization of microbiota
Cohouse wild type and mutant (in case of profound differences)	Cage effect	Synchronization of microbiota

# Pen effects

1.



- ✓ age, gender
- ✓ diet
- ✓ genotype
- ✓ environment



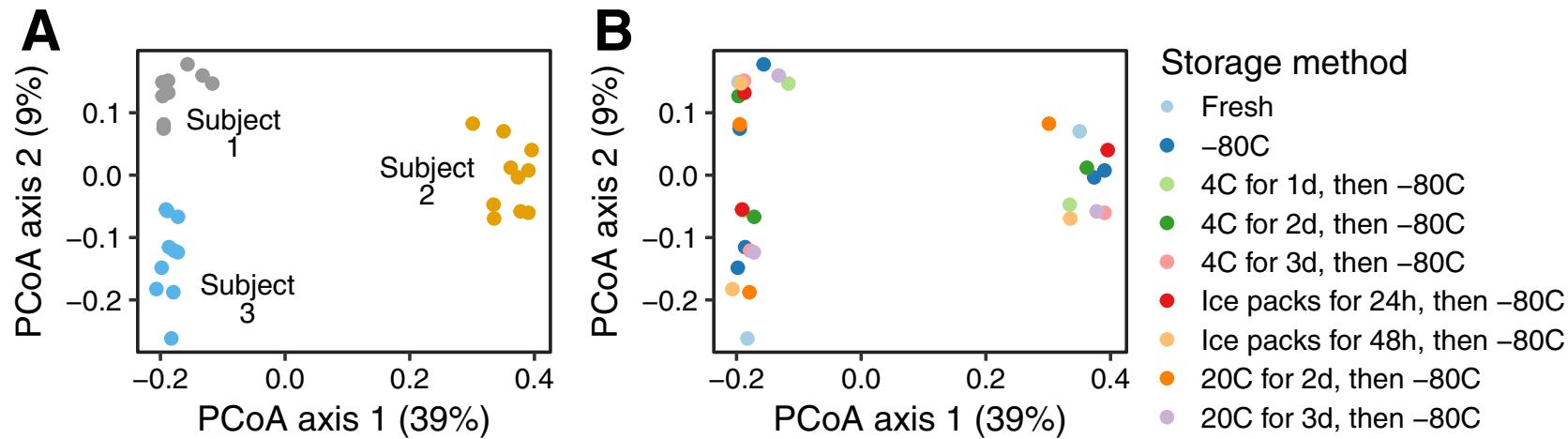
**Fig. 1** Example of cage effects dominating a mouse study of fungal communities. Fungal lineages in the murine gut were inferred from ITS rRNA gene sequencing of pellets [87]. The heat maps summarize taxonomic assignments derived from the sequence data. The color scale to the right indicates the proportions of each lineage; white indicates not detected. Caging dominated over treatment in this study. The three conditions studied were continuous exposure to antibiotics (Condition 1), short-term exposure to antibiotics (Condition 2), and no exposure to antibiotics (Condition 3). For details see [87]

# Sample storage

2.



- ✓ tissue type (feces, biopsy)
- ✓ location along the digestive tract
- ✓ sampling (aerobic vs anaerobic)
- ✓ storage time and temperature



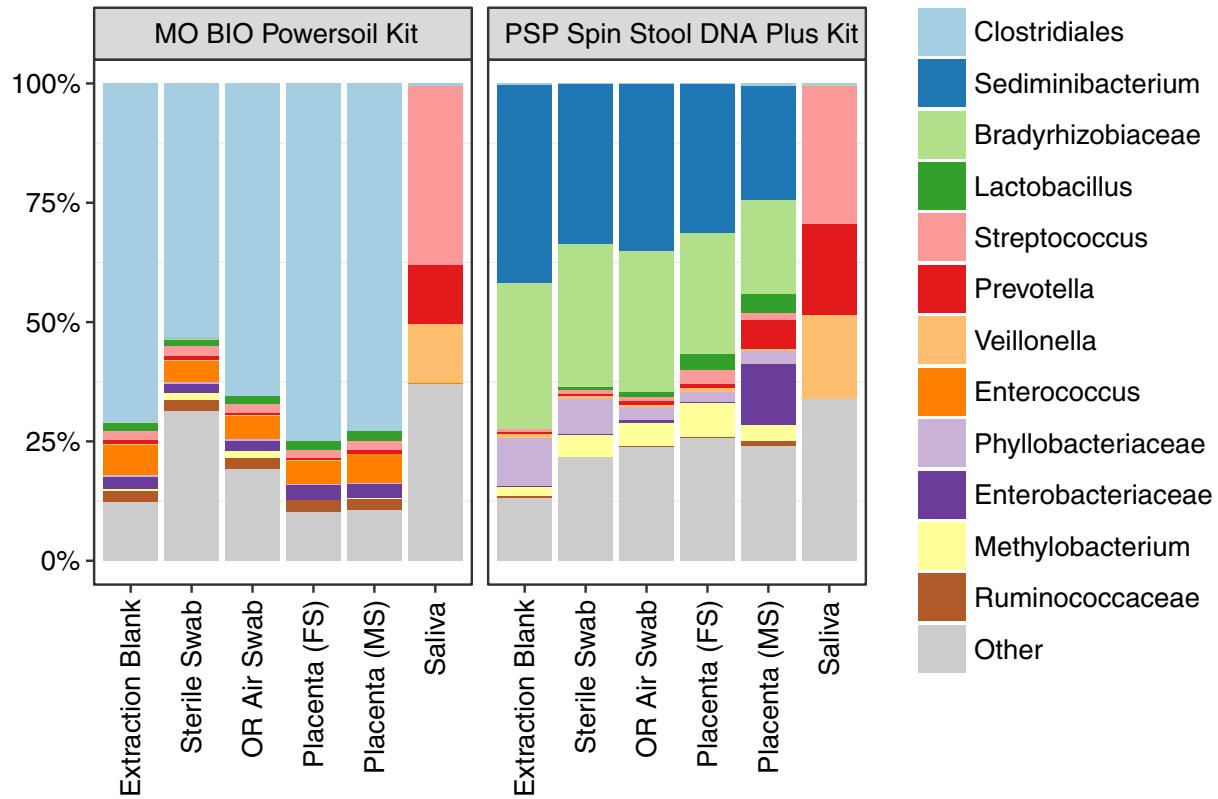
**Fig. 2** Effects of sample storage methods on community structure inferred for oral swabs. Oral swab samples were acquired from three human individuals and DNA extracted. DNAs were amplified using 16S rRNA gene primers binding to the V1-V2 region then sequenced using the Illumina platform using our standard procedures [88]. Unweighted Unifrac (C [129]) was used to generate distances between all pairs of samples then results were displayed using Principal Coordinate Analysis (PCoA). **a** Samples from each of the three subjects are color coded (red, blue, and green). **b** Nine storage conditions were compared, indicated by the different colors. The key to storage conditions is at the right

# Low biomass samples

2.



- ✓ tissue type (feces, biopsy)
- ✓ location along the digestive tract
- ✓ sampling (aerobic vs anaerobic)
- ✓ storage time and temperature



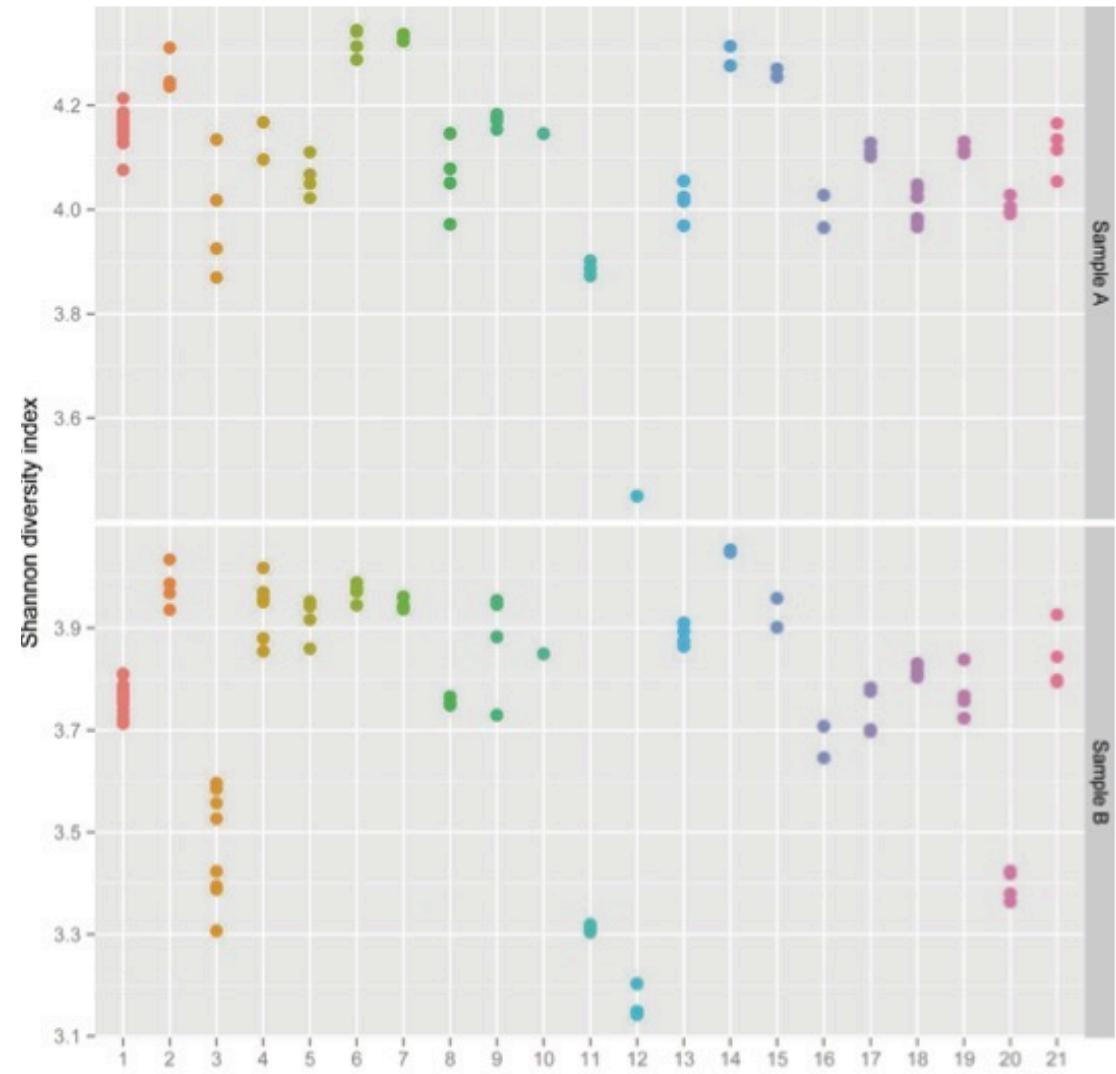
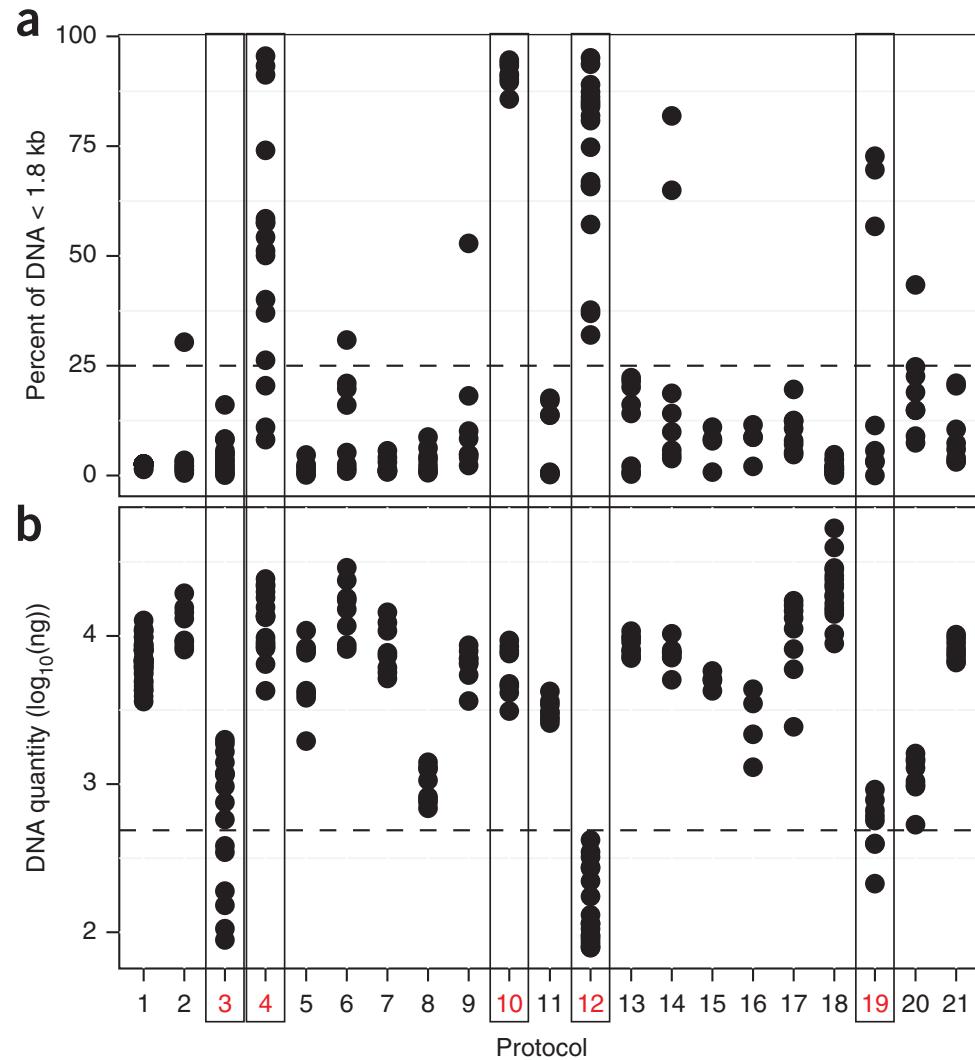
**Fig. 3** Wrestling with kit contamination—similar bacterial composition in placental samples and negative controls. Relative abundances of bacterial lineages were inferred from 16S V1-V2 rRNA marker gene sequence information [22]. Samples studied included negative controls, fetal side (FS) placental swabs, maternal side (MS) placental swabs, saliva, and vaginal swabs. Replicates of each sample were extracted using two different kits—the kit type is indicated above each panel. Operating room (OR) air swabs are swabs that were waved in the air at the time of sample collection to be used as negative controls. Saliva samples, which are high in microbial biomass, showed similar compositions for each of the two extractions; placental samples resemble the kit-specific negative controls

# DNA extraction – 2017

3.

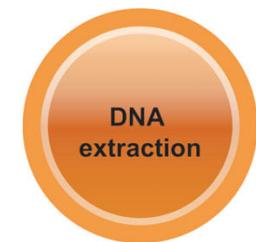


✓ DNA extraction method

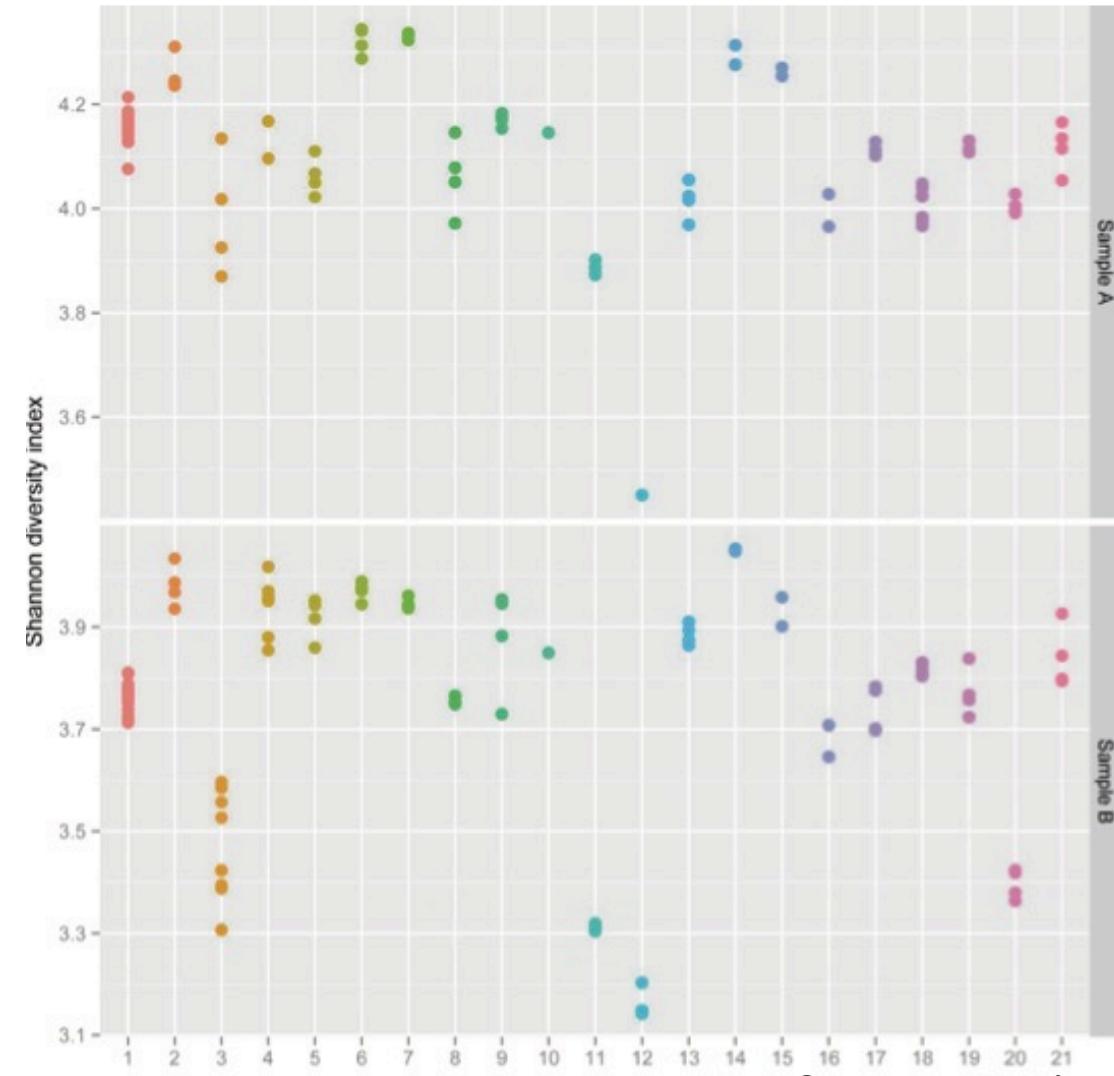
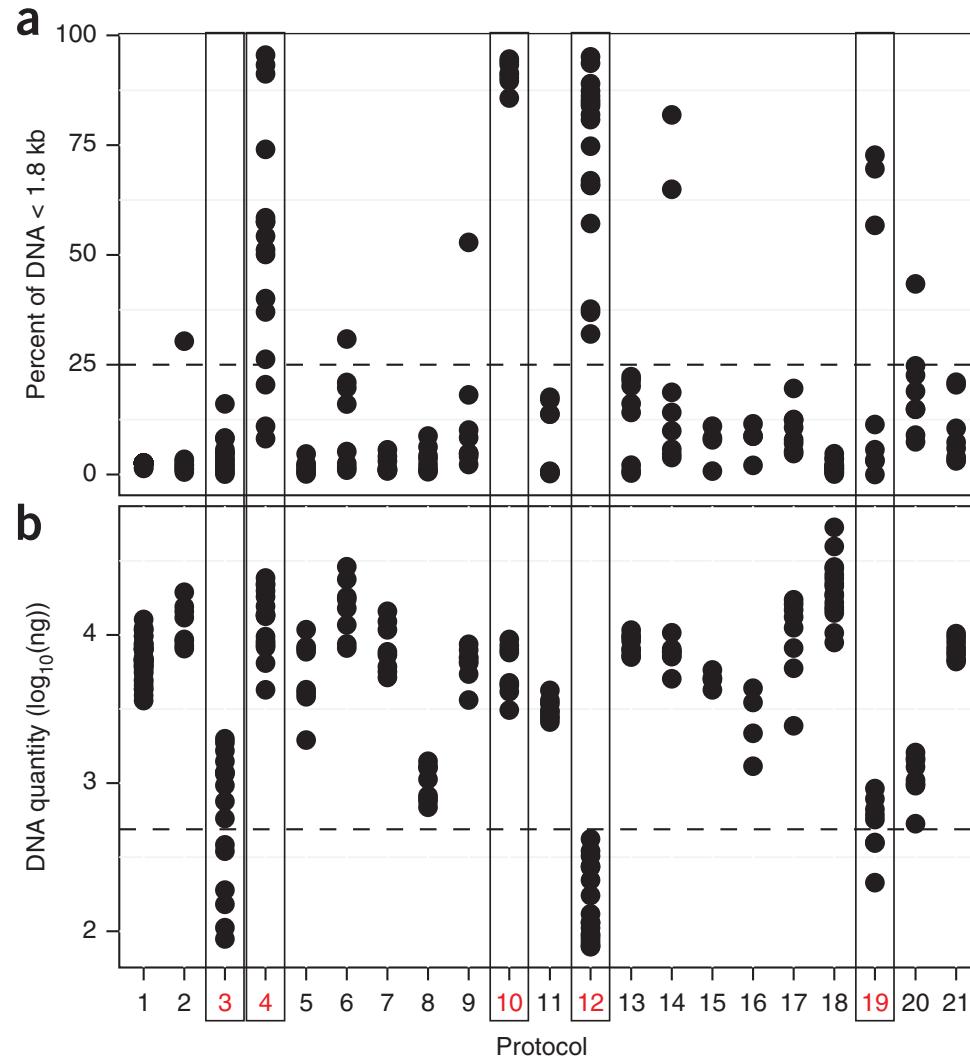


# DNA extraction – 2017

3.

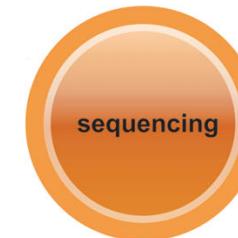


- #3 HMP MoBio PowerSoil
- #12 TissueLyzer, easymag kit
- #18 Phenol chloroform, no beads
- #19 MagnaPure kit

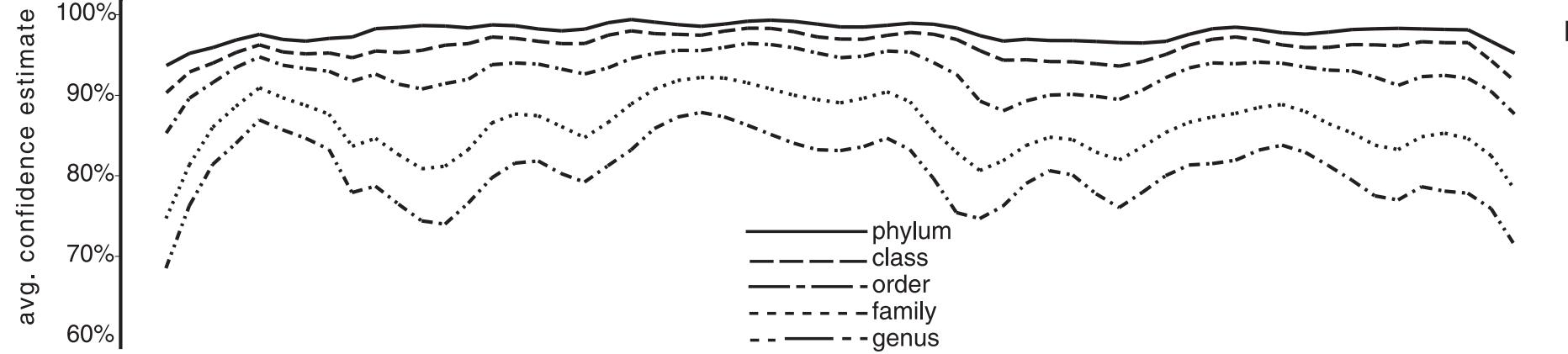
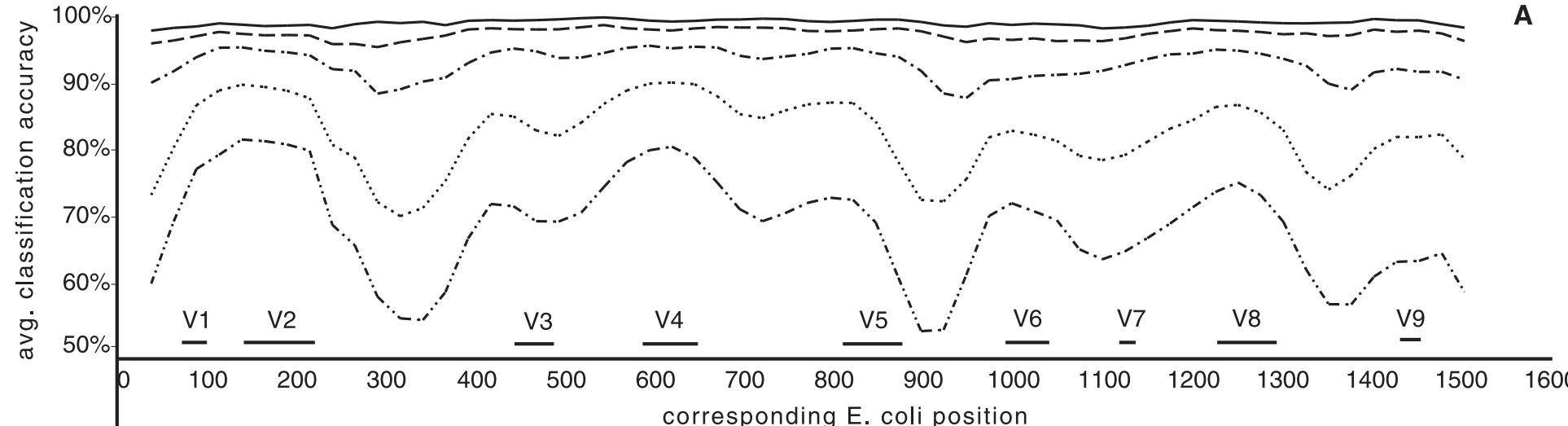


# PCR bias

4.

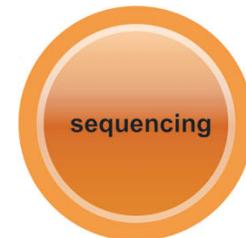


- ✓ choice of primers
- ✓ PCR protocol
- ✓ sequencing method
- ✓ taxonomic classification

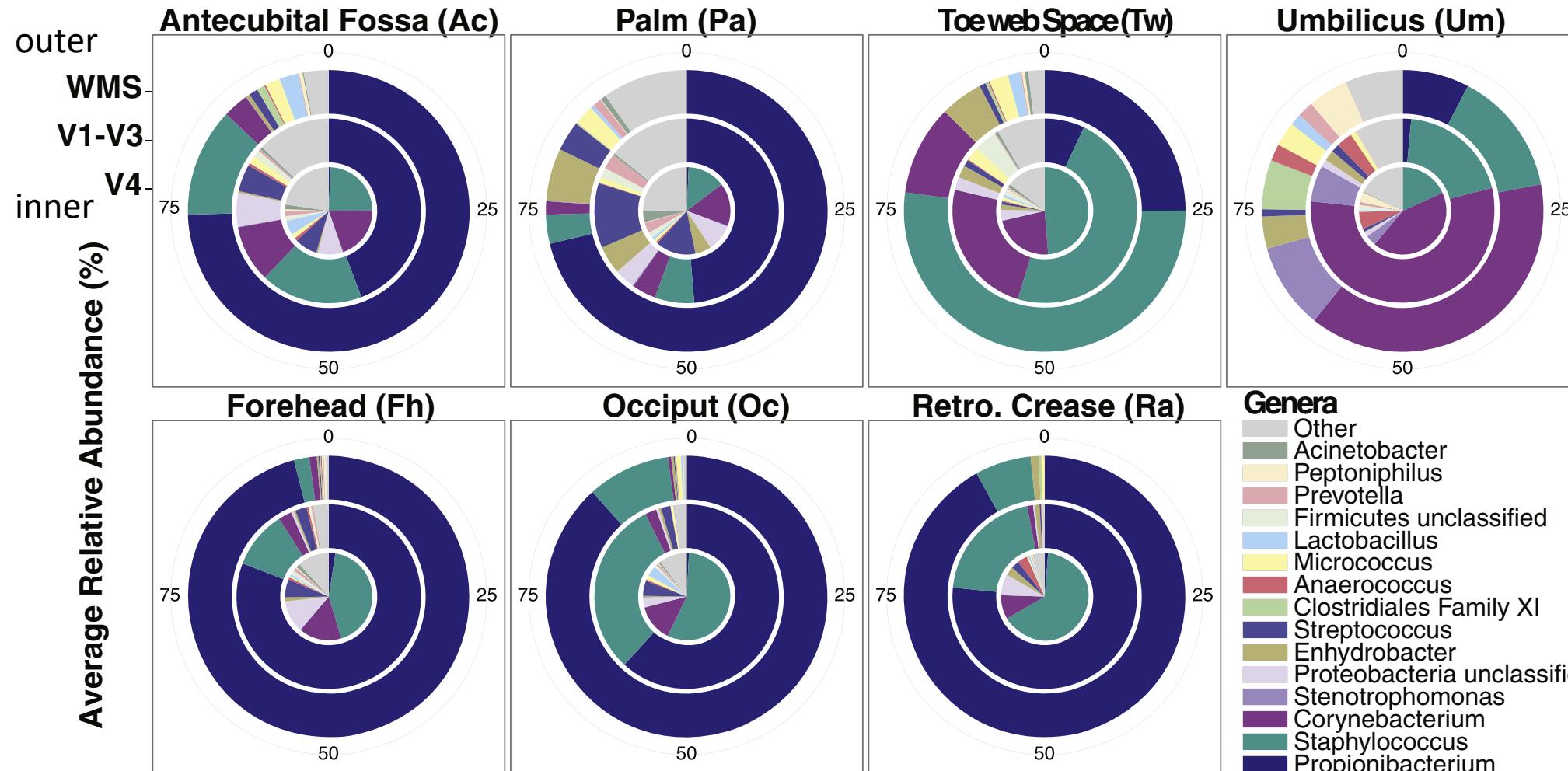


# PCR bias: skin microbiome

4.



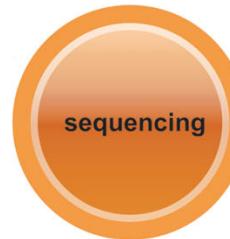
- ✓ choice of primers
- ✓ PCR protocol
- ✓ sequencing method
- ✓ taxonomic classification



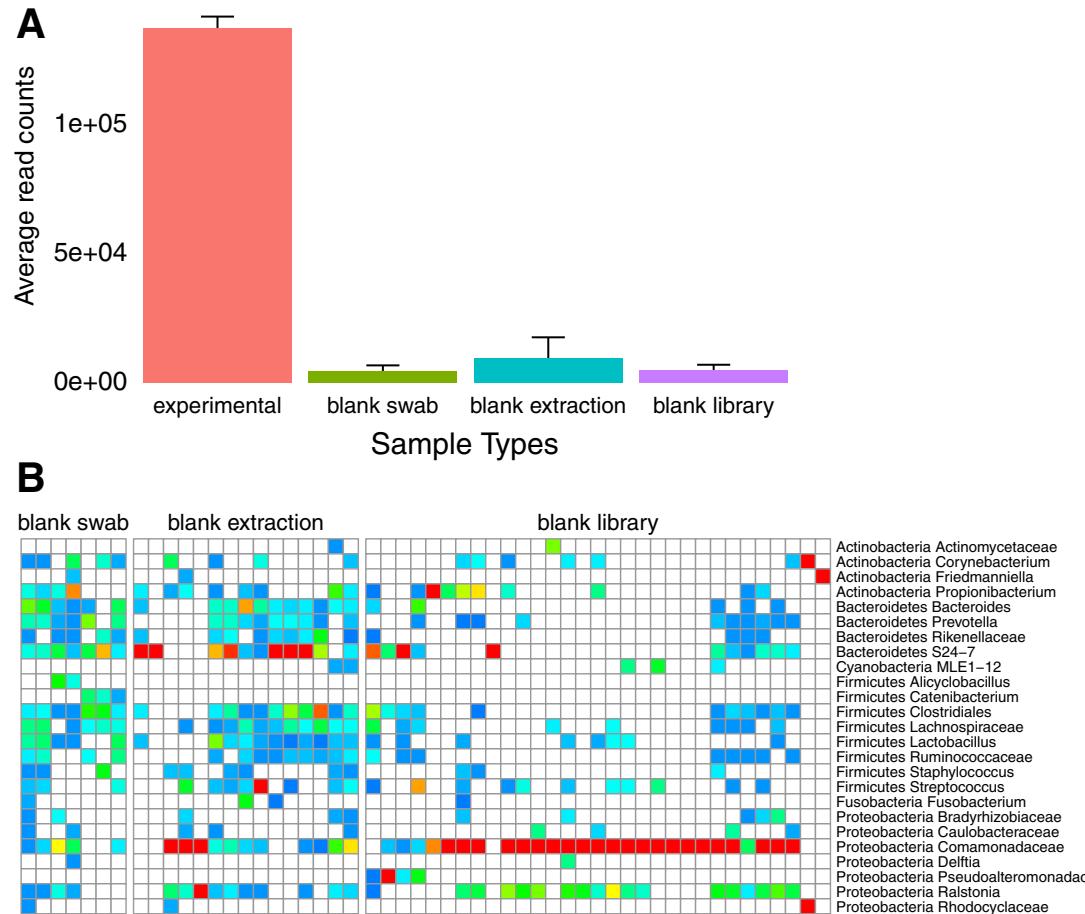
WMS = Whole metagenome shotgun. They are considering this to be the ground truth, because there is no PCR bias.

# Kit-ome?

4.

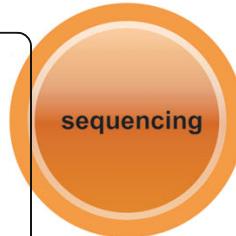


- ✓ choice of primers
- ✓ PCR protocol
- ✓ sequencing method
- ✓ taxonomic classification



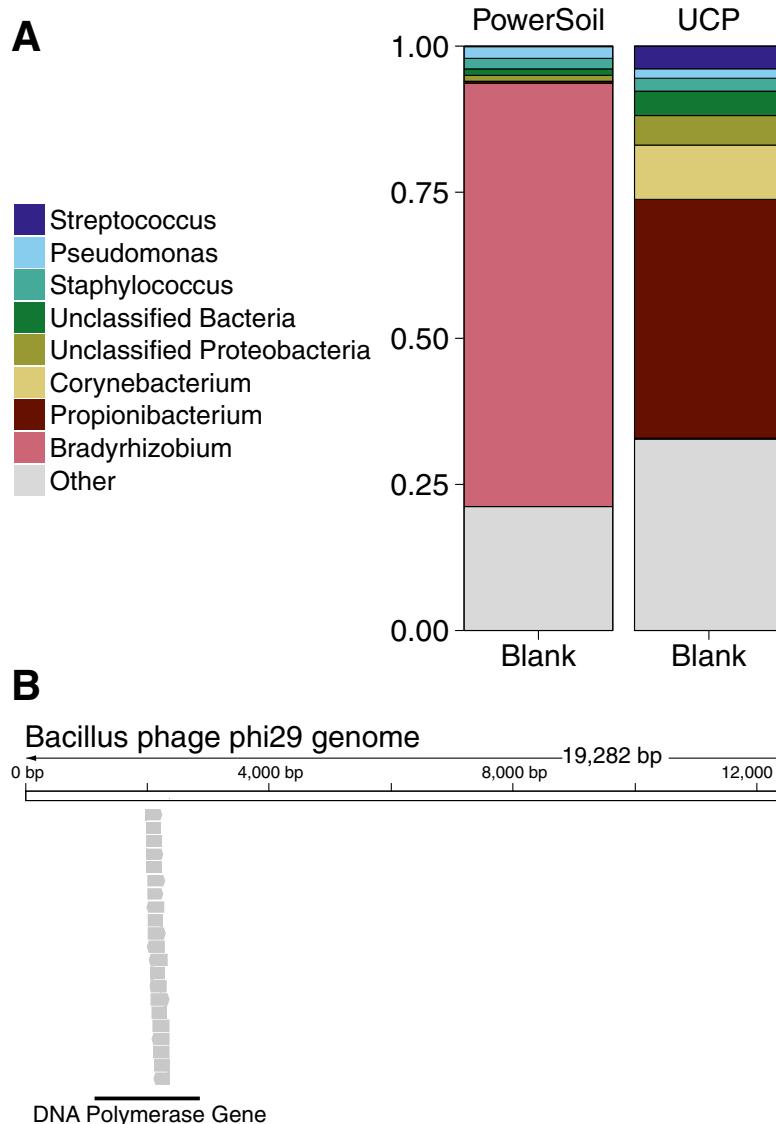
**Fig. 4** Analysis of three negative control sample types reveals contaminating taxa. Data for negative controls was acquired using 16S V1-V2 rRNA marker gene sequencing analyzed on the Illumina MiSeq platform. Data from 11 experiments were pooled. **a** Comparison of average read counts. Experimental samples had an average read count of 137,243 and negative control samples had an average read count of 6613. **b** Heat map summary of bacterial lineages present in negative control samples. Different OTUs are present in DNA-extraction controls ("blank extraction" and "blank swab") and library preparation controls ("library blank") collected over multiple sequencing runs

4.



- ✓ choice of primers
- ✓ PCR protocol
- ✓ sequencing method
- ✓ taxonomic classification

# Kit-ome



**Fig. 6** Contamination in shotgun metagenomic data. **a** Lineages observed in shotgun metagenomic sequencing of negative control samples using standard (DNeasy PowerSoil) and low-contaminant (QiaAmp UCP Pathogen) kits. **b** Detecting *Bacillus* phage phi29 polymerase reads in a blank sample. Twenty-one reads from a blank sample aligned to the DNA polymerase gene (1145 to 2863 bp) of *Bacillus* phage phi29. The protein was purchased as a reagent from a commercial supplier, suggestive of contamination of the protein with cloned DNA encoding the polymerase gene used in protein over-expression

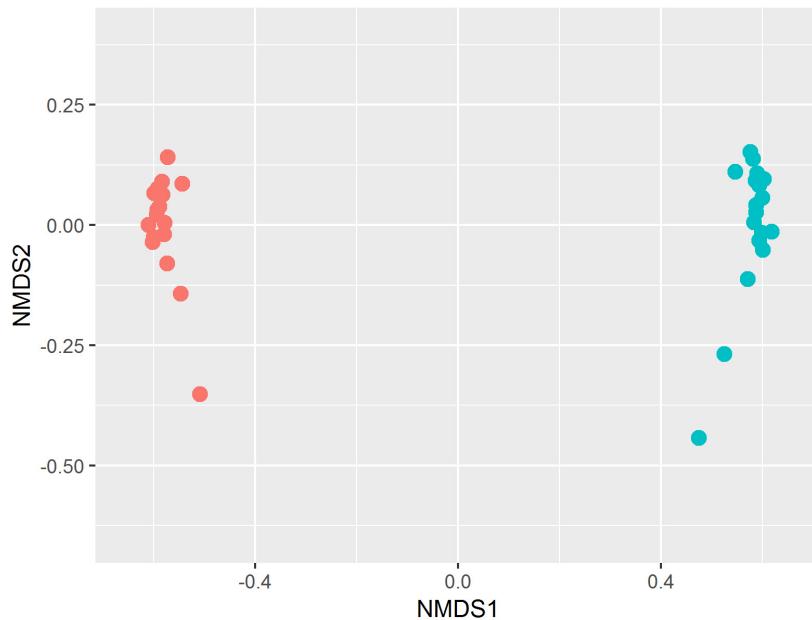
# Analysis pipeline

5.

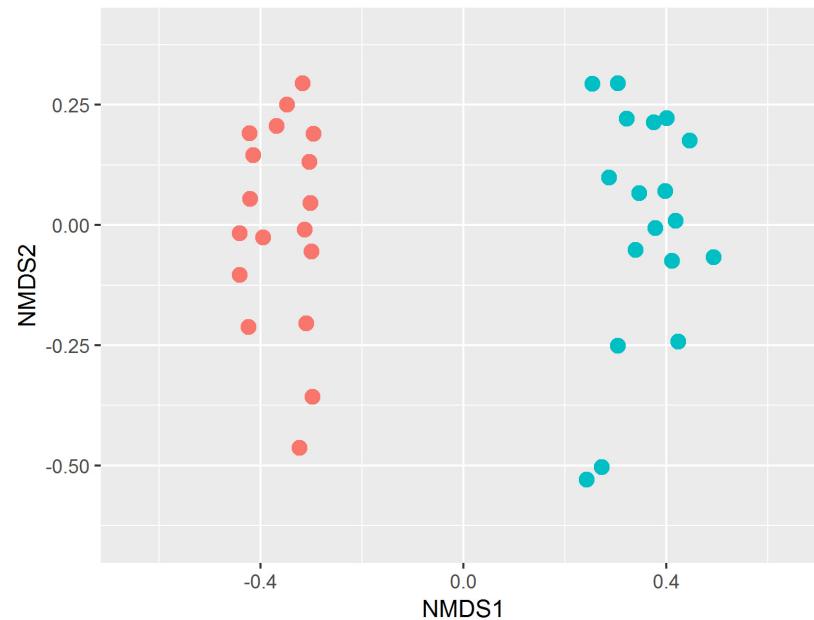


- ✓ data cleaning
- ✓ comparison
  - ✓ taxonomic levels (species, phyla)
  - ✓ clustering techniques
  - ✓ functional analysis

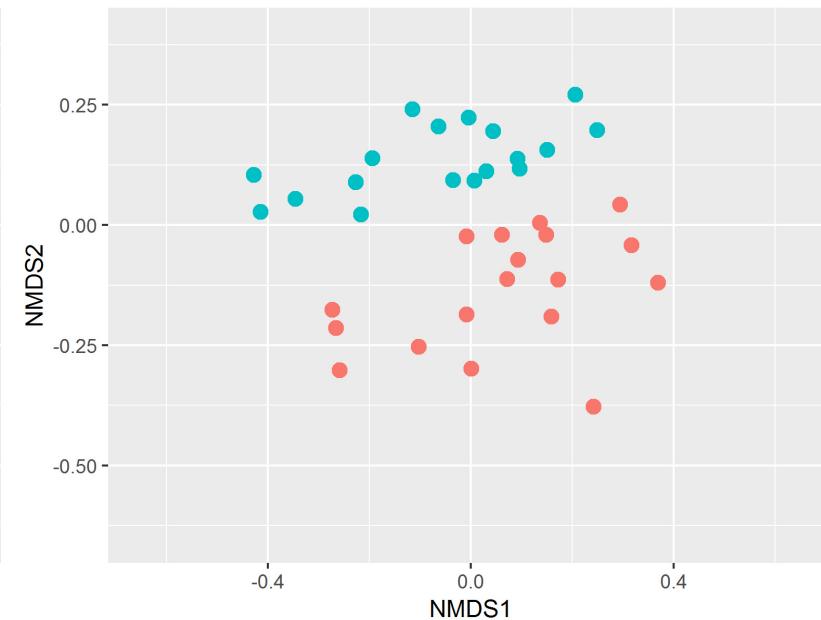
SILVA: Taxa grouped by Genus



SILVA: Taxa grouped by Family



SILVA: Taxa grouped by Phylum



Compared ruminal contents from 18 dairy cattle  
Results from mothur and qiime were different, but...

# Analysis pipeline

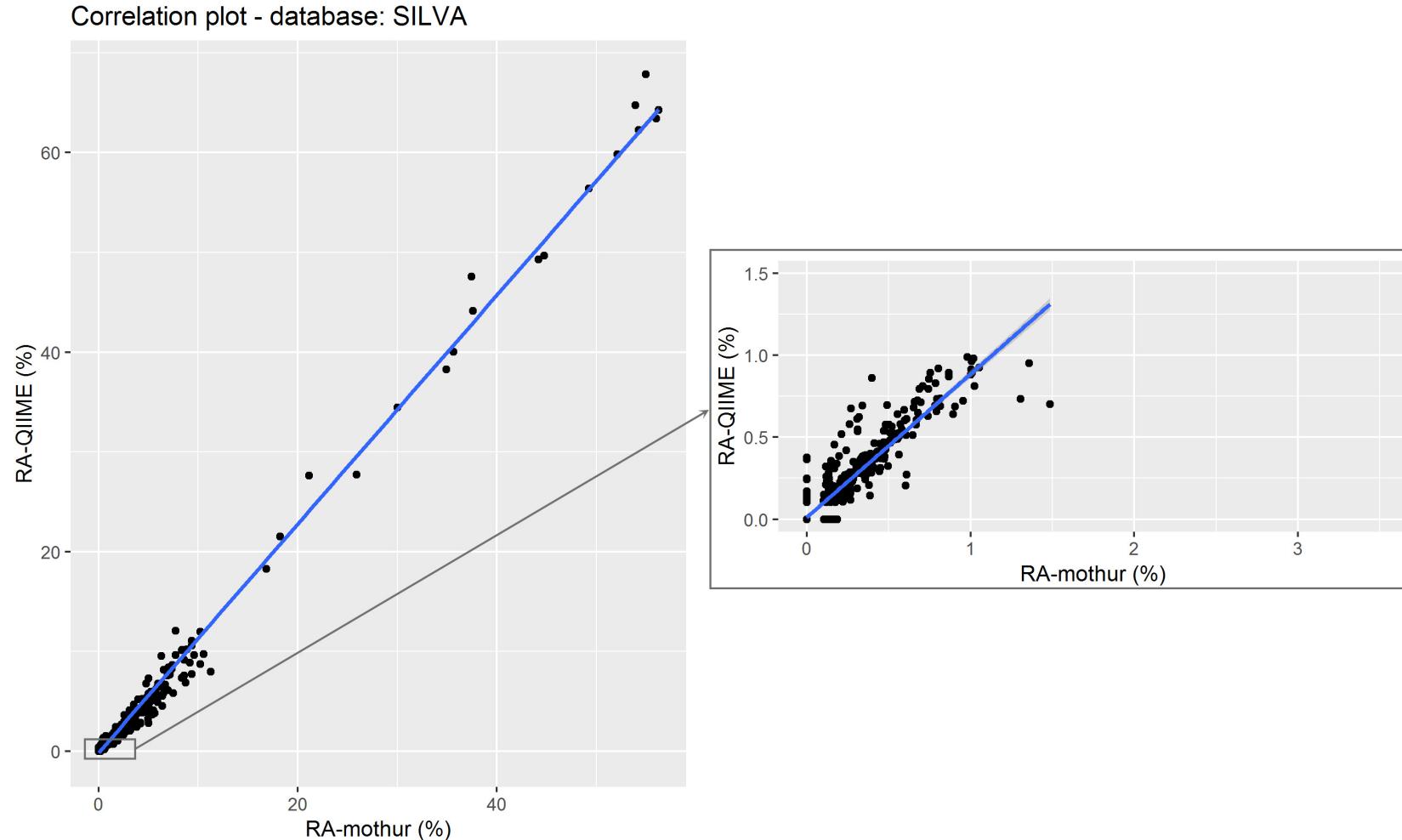
5.



- ✓ data cleaning
- ✓ comparison
  - ✓ taxonomic levels (species, phyla)
  - ✓ clustering techniques
  - ✓ functional analysis

## Take aways:

1. Mothur and qiime performed similarly to estimate relative abundance (RA).
2. Meta-analysis incorporating datasets generated from both might introduce errors.
3. GreenGenes is out of date



**FIGURE 5 |** Relative abundance of the different microorganisms (by genera) detected by QIIME and mothur within the 18 samples using SILVA as reference data set. The subset shows the correlation between data with QIIME RA < 1%. Points represent individual RA within each sample.

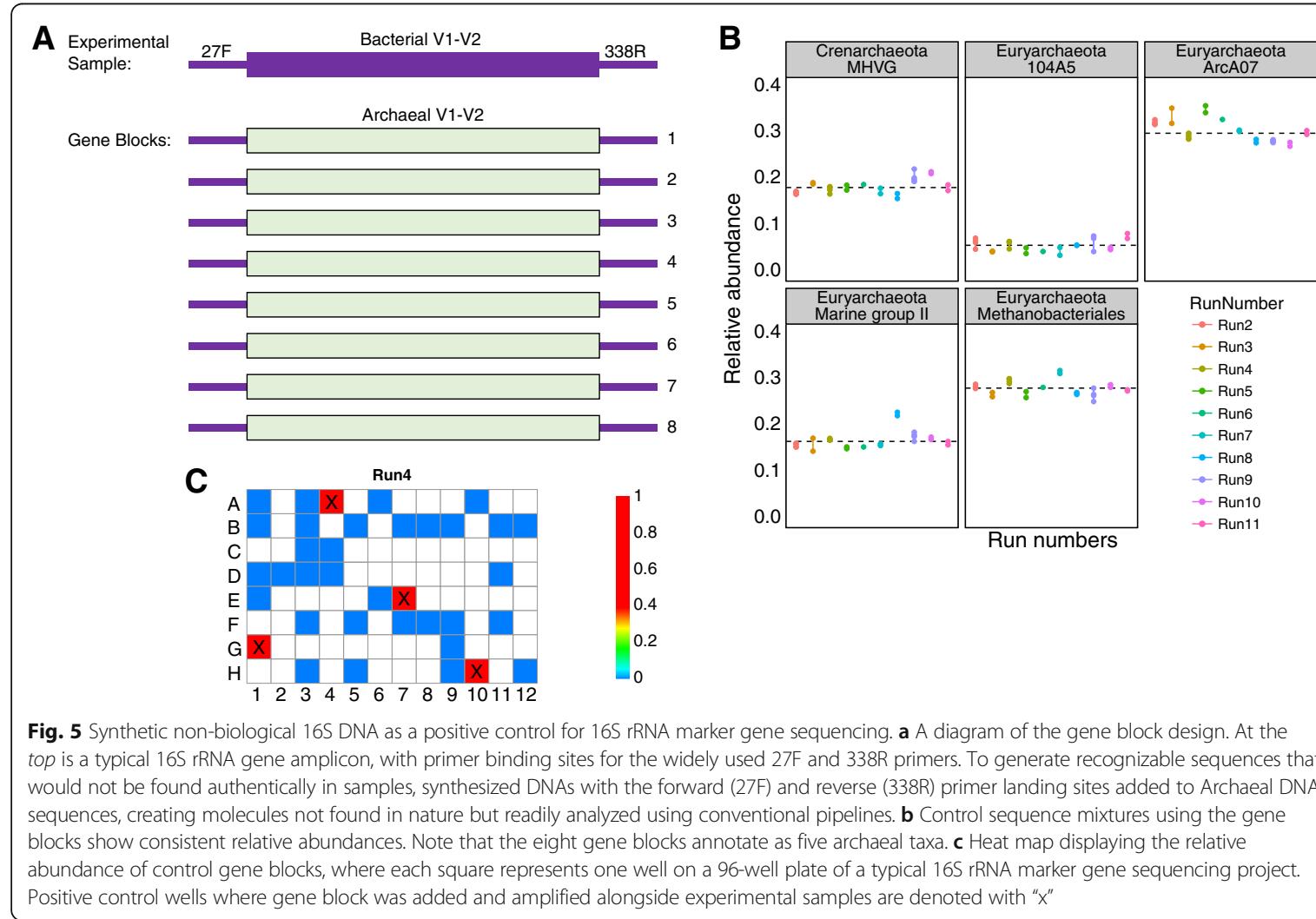
# Attempt at meta-analysis-which are most important factors?

**TABLE 2** Factors associated with the community structure of the swine gut microbiota as measured using PERMANOVA with the adonis function (9,999 permutations) of the weighted and unweighted UniFrac distances and Bray-Curtis dissimilarities<sup>a</sup>

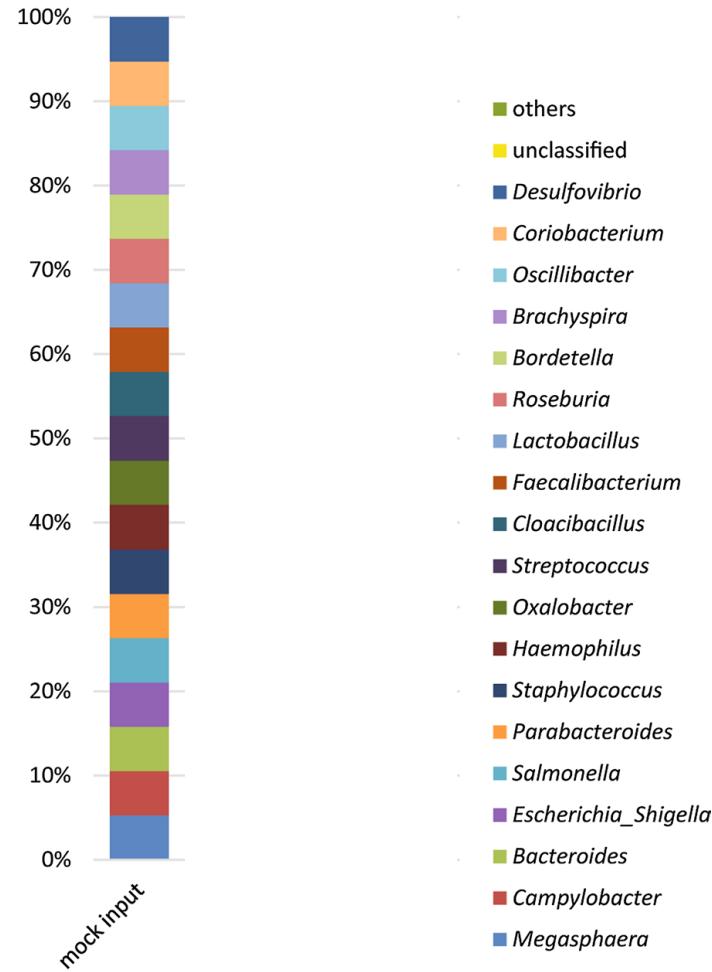
Parameter	Value									
	Weighted UniFrac			Unweighted UniFrac			Bray-Curtis			
	Pseudo-F		P value	Pseudo-F		P value	Pseudo-F		P value	
	ratio	R <sup>2</sup>		ratio	R <sup>2</sup>		ratio	R <sup>2</sup>		
Study	28.2	0.37	0.0001	18.6	0.28	0.0001	24.4	0.34	0.0001	
GI sampling location	35.4	0.29	0.0001	12.9	0.13	0.0001	12.1	0.13	0.0001	
Age	13.9	0.21	0.0001	8.5	0.14	0.0001	10.6	0.16	0.0001	
Country of origin	25.4	0.18	0.0001	15.6	0.12	0.0001	20.4	0.15	0.0001	
Hypervariable region sequenced	34.5	0.16	0.0001	22.7	0.11	0.0001	18.9	0.13	0.0001	
Sequencing platform	20.4	0.04	0.0001	19.0	0.04	0.0001	25.6	0.05	0.0001	

<sup>a</sup>For the age category, one study was excluded due to a lack of information about the age of the pigs used. PERMANOVA, permutational multivariate analysis of variance.

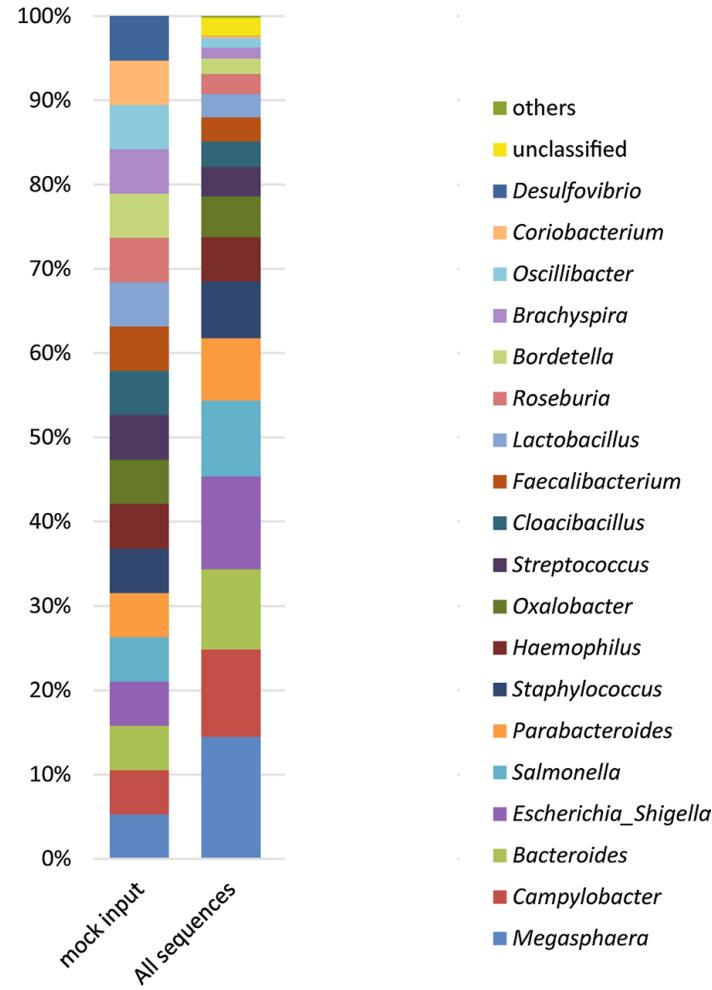
# Positive control samples



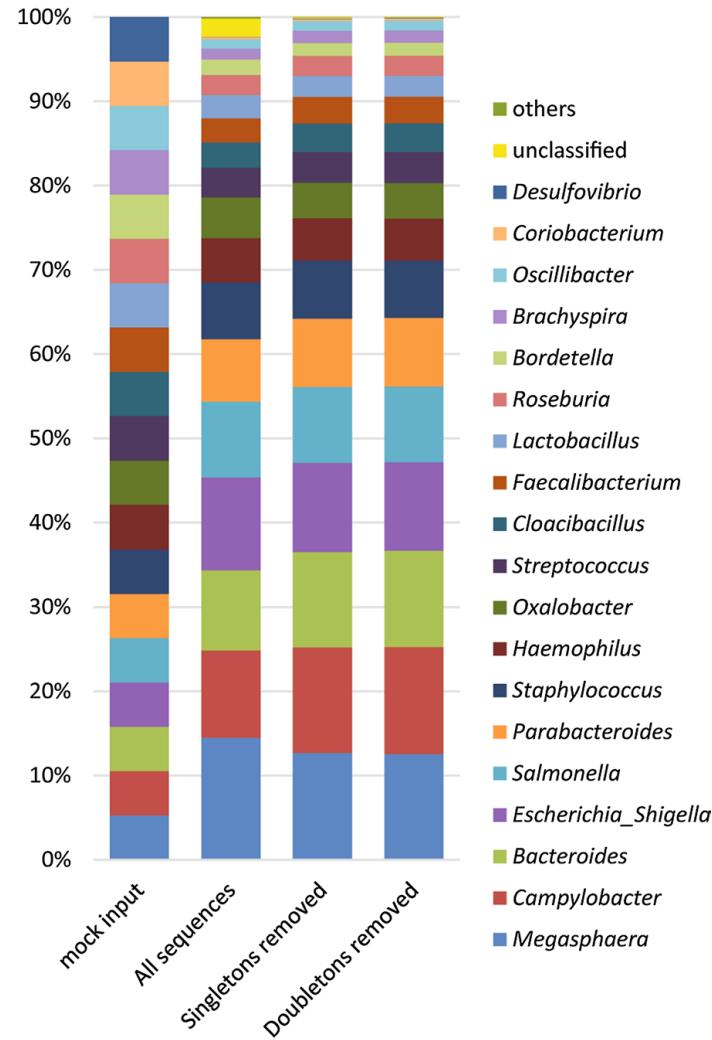
# Positive controls – mock community



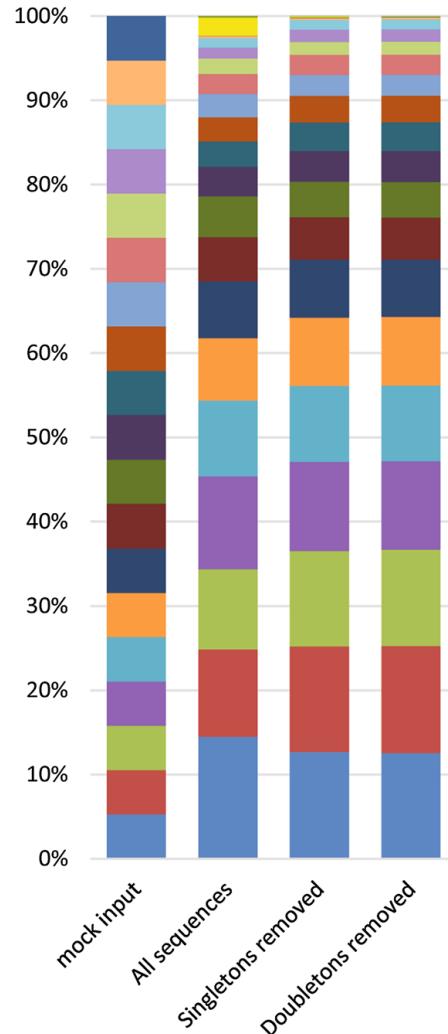
# Positive controls – mock community



# Positive controls – mock community



# Positive controls – mock community



**Table 3 Average diversity estimates of the mock community ( $n = 12$ , rarified to 6654 sequences per sample) with and without removing low-frequency sequences**

Mock community	Actual number of OTUs <sup>a</sup>	Observed number of OTUs	Estimated total number of OTUs <sup>b</sup>	Chao diversity index	Shannon diversity index	Inverse Simpson index	Error rate (%)	File size (Gb) <sup>c</sup>
All sequences	20	$734 \pm 56$	$374,770 \pm 214,807$	$21,676 \pm 3273$	$3.6 \pm 0.1$	$18 \pm 0.8$	3.6	41
Singletons removed	20	$28 \pm 0.8$	$68 \pm 13$	$41 \pm 3$	$2.7 \pm 0.02$	$12 \pm 0.3$	1.4	21
Single and doubleton removed	20	$22 \pm 0.3$	$22 \pm 0.3$	$23 \pm 0.7$	$2.6 \pm 0.02$	$12 \pm 0.3$	1.3	3

Average diversity estimates: plus or minus ( $\pm$ ) the standard error of the mean, where appropriate

<sup>a</sup> *Haemophilus parasuis* has two divergent copies of the 16S rRNA gene that cluster separately

<sup>b</sup> The estimated total number of OTUs is the number of OTUs predicted to be in the sample based on the number of OTUs observed in the sequences. The program Catchall was used to make the estimates [14]

<sup>c</sup> Size of the distance matrix file

# Experimental Design

- From the top of the field in 2012
- Centralization
- Metadata
  - Minimum information checklists
  - MIGS
  - MIMS

**Table 1 Key factors to take into account when designing a metagenomics pipeline**

Challenge	Decision	Pitfall	Consequence
Biological and technical replicates are expensive and time-consuming	Whether to perform replication, or gamble that a single sample in each group is informative with sufficiently well-described ecosystem parameters	Often nonreplicated designs are not interpretable, or are overinterpreted (e.g., attributing differences in a single healthy versus single diseased person to the disease)	Conclusions cannot be replicated by other researchers, and may not be generalizable beyond the specific samples analyzed
A fixed sequencing budget must be divided among some number of samples (e.g., by multiplexing at some level)	Whether to sequence few samples deeply, or many samples more shallowly	The appropriate number of samples and sequencing depth are unknown	Few samples may be uninformative and may preclude informative analysis of variation in the system and/or replication; shallow sequencing may miss rare but important taxa or functions
Experimental challenges due to low yield of DNA and/or high community diversity	Whether to adopt new protocols for improved DNA extraction, amplification and/or assembly	DNA extraction and manipulation steps all introduce biases that may make it difficult to compare between studies	For unique or rare samples that require special treatment, it is essential that all steps in the treatment are considered if comparing results to those from other studies
Defining the dimensions of variation that matter in a given system is challenging, and often is the purpose of the study itself	Which scales and parameters to select, and how much variation to cover	'Extremes' of variation in the system being studied are expensive and difficult to obtain (tail of distribution) and may not even be extreme from the microbes' perspective; relevant variation often unknown	Conclusions from one population or study site inappropriately generalized to other populations or study sites; relevant variation in system undiscovered; extreme efforts to obtain exotic samples are unrewarded
Must choose a sequencing platform	Trade-off between read length and number of sequences; must decide when to adopt new technology	All sequencing technologies and processing pipelines have drawbacks, not all of which are widely advertised; technology changes rapidly	Sequences may be too short, too few or too error-prone to interpret or too passé to publish
Interpretation of sequence data	Must decide whether to use reference-based or <i>de novo</i> methods for assembly, taxonomy and functional assignment, and if so which reference to use	Different reference databases give different results; <i>de novo</i> is unbiased but far less powerful when appropriate references exist; analyses differ as reference databases update rapidly, limiting comparisons between studies. Current assembly algorithms are insufficient for highly complex metagenome data.	Incorrect and/or hard-to-reconcile functional and taxonomic assignments
Metadata collection	Must decide what metadata (that is, sample or site data) to collect and associate with sample	Too complex to be implemented; fields inconsistent with previous studies due to lack of standards compliance; data model can't accommodate	Chaos!
Centralization	Whether to centralize sample collection, metadata curation, DNA extraction, sequencing, data storage and data analysis	Decentralization can lead to inconsistencies that make data difficult to interpret; centralization can lead to delays while funding is acquired or capacity is built, and can limit creativity	Either the data set may be vast but too inconsistent to interpret, or it may be extremely consistent but limited in scope and/or interpretation. Specific considerations apply to each stage; the EMP currently favors decentralized sample collection and centralization of other steps on a case-by-case basis.

# Take aways

- Everything matters in experimental design and sample processing,
- BUT, DNA extraction and primer choice are especially important.
- Keep everything consistent within a study.
- Look for methods in your field and if you want to compare with another study use the same methods.
- Always use a mock community and kit negative control and vigorously analyze and share those results.

# Need for a universal standard operating procedure

- One option is the method described by the Schloss lab
- [https://github.com/SchlossLab/MiSeq\\_WetLab\\_SOP/blob/master/MiSeq\\_WetLab\\_SOP.md](https://github.com/SchlossLab/MiSeq_WetLab_SOP/blob/master/MiSeq_WetLab_SOP.md)

## Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform

James J. Kozich,<sup>a</sup> Sarah L. Westcott,<sup>a</sup> Nielson T. Baxter,<sup>a</sup> Sarah K. Highlander,<sup>b</sup> Patrick D. Schloss<sup>a</sup>

Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA<sup>a</sup>; Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA<sup>b</sup>



# Universal SOP

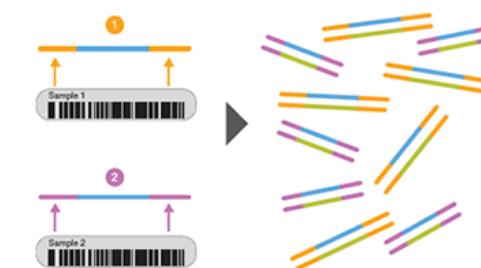
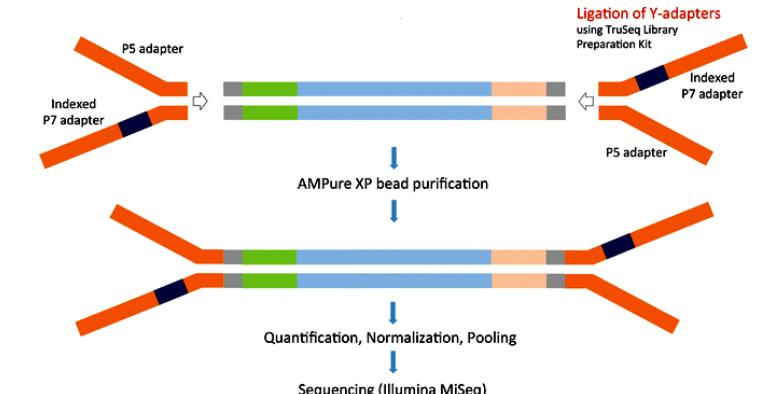
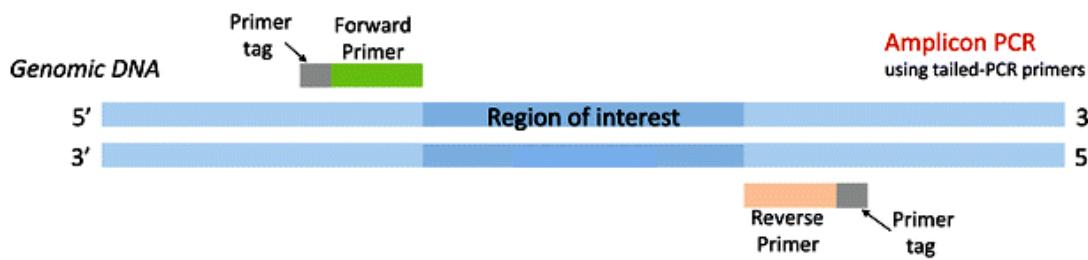
- DNA extraction
- PCR with V4 primers
- PCR QC
- PCR normalization
- Plate pooling and QC
- Library pooling

# Universal SOP: DNA extraction

- They leave this up to you.
- Is a balance between quality and throughput.

# Universal SOP: PCR with V4 primers

- In 96 well plate format
- Use barcode indexed primers
  - 16 forwards, 24 reverses ( $16 \times 24 = 384$ )
- AccuPrime Pfx Supermix



# Universal SOP: PCR QC and normalization

- Run a few products on gel
- Cleanup and normalize with SequalPrep Normalization Kit
- No gel extraction!
- Good for throughput and better at normalizing. (Harris, AEM, 2010)
- Then put all 96 samples from the plate into one tube

# Universal SOP: Library pooling

- Fragment length determination: Agilent Bioanalyzer
- Library quantification: KAPA library quantification kit, a qPCR method
- Mix up to 4 pools into a single tube

# Ecological Principles for ecosystem colonization

- Dispersal
  - movement through space
  - ex-host survivability and transmission
- Diversification
  - species change by recombination or mutation
  - Favored in host: high density, temperature, nutrients
- Environmental selection
  - colonization of niches
  - host is “habitat filter” or an interactive biome
- Ecological drift
  - demographic stochasticity

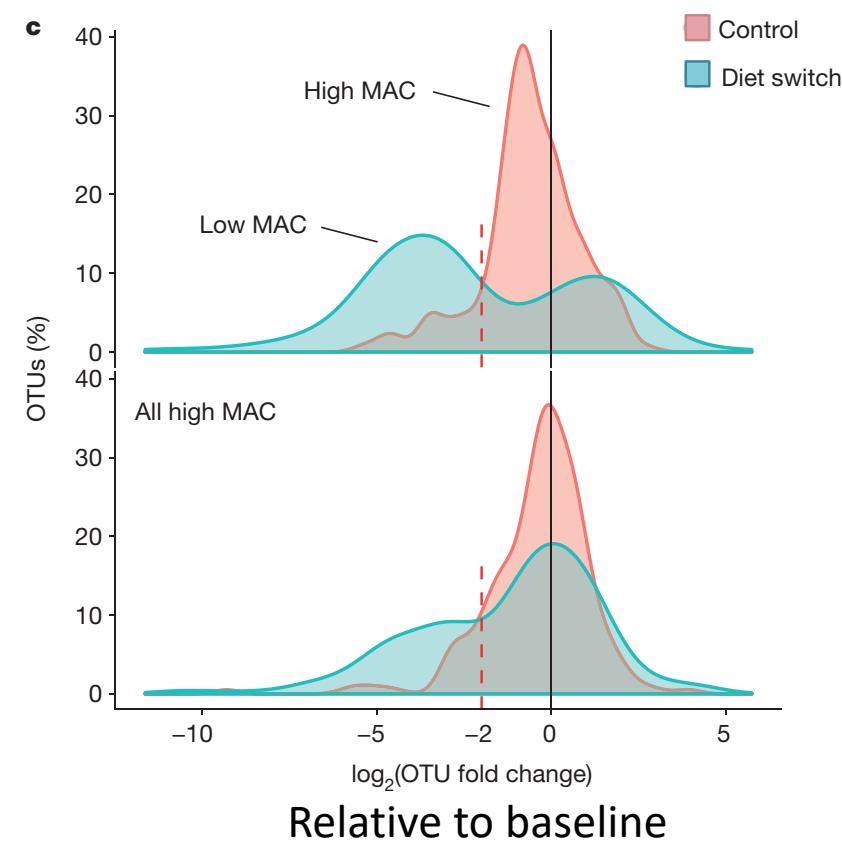
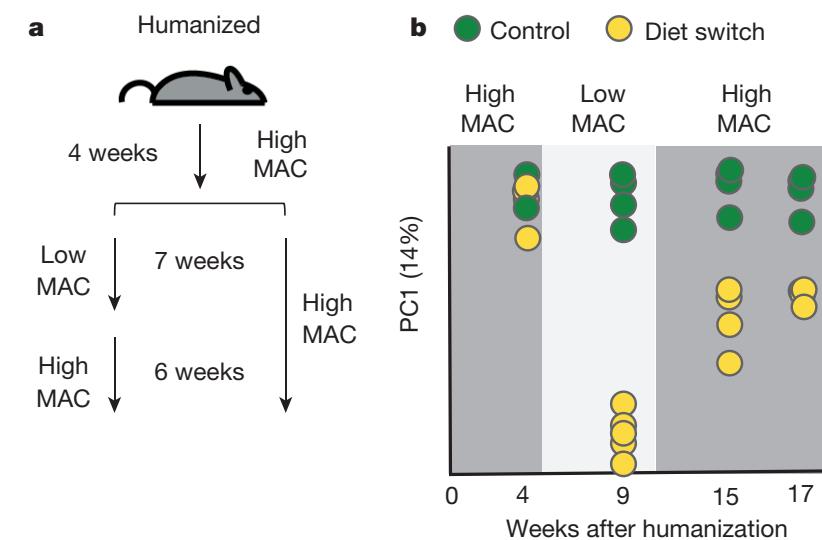
# Generational succession

LETTER

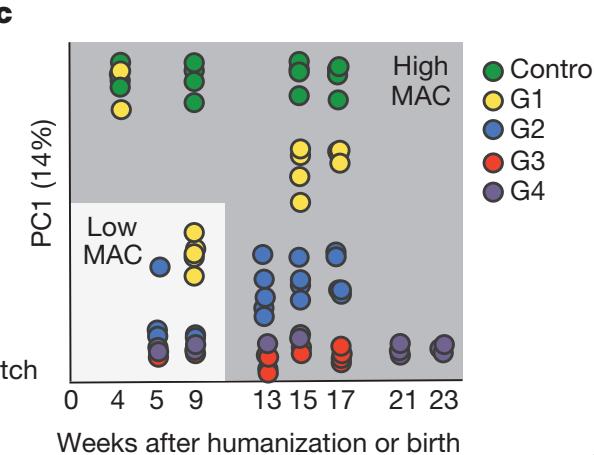
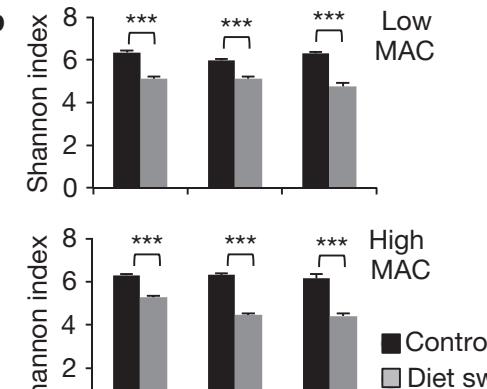
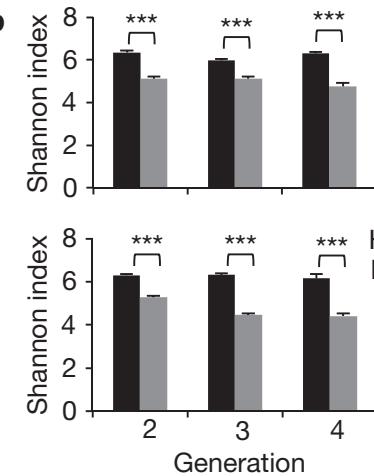
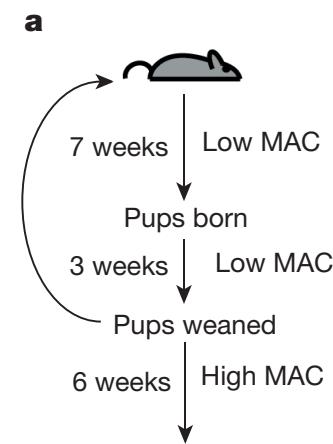
doi:10.1038/nature16504

## Diet-induced extinctions in the gut microbiota compound over generations

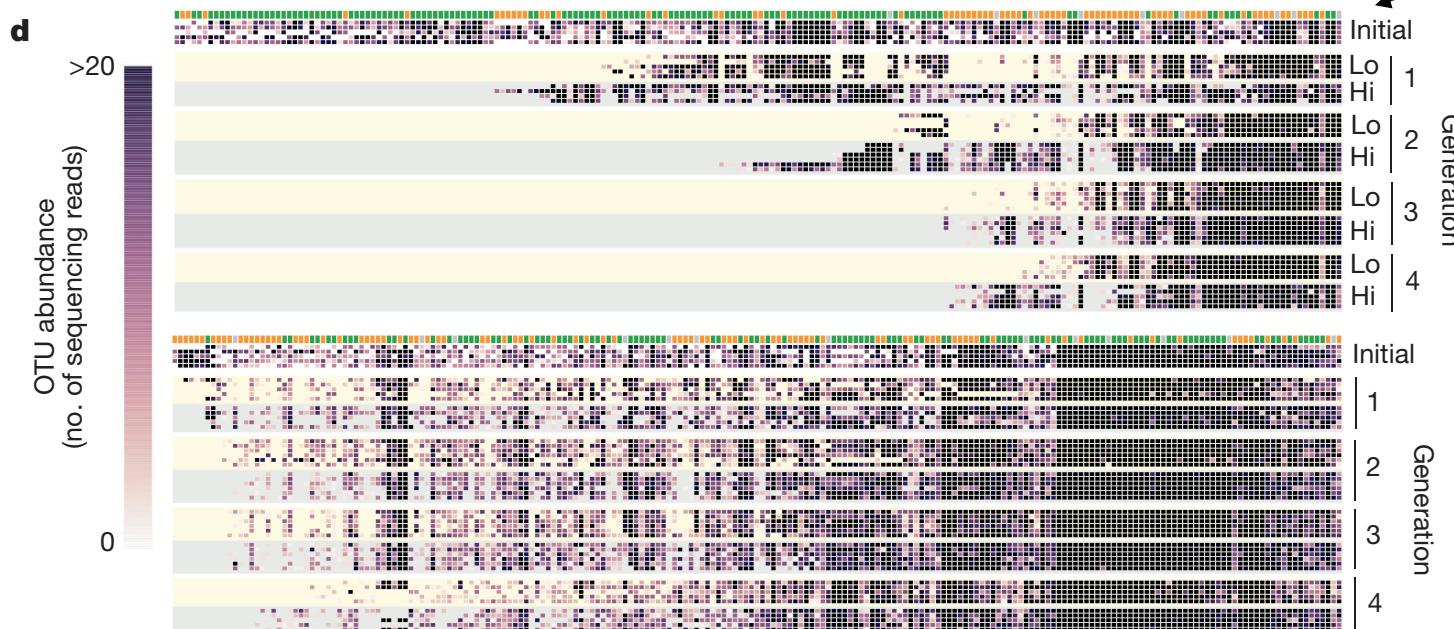
Erica D. Sonnenburg<sup>1\*</sup>, Samuel A. Smits<sup>1\*</sup>, Mikhail Tikhonov<sup>2,3</sup>, Steven K. Higginbottom<sup>1</sup>, Ned S. Wingreen<sup>4,5</sup> & Justin L. Sonnenburg<sup>1</sup>



# Generational succession



*Bacteriodes*, green  
Firmicutes, orange

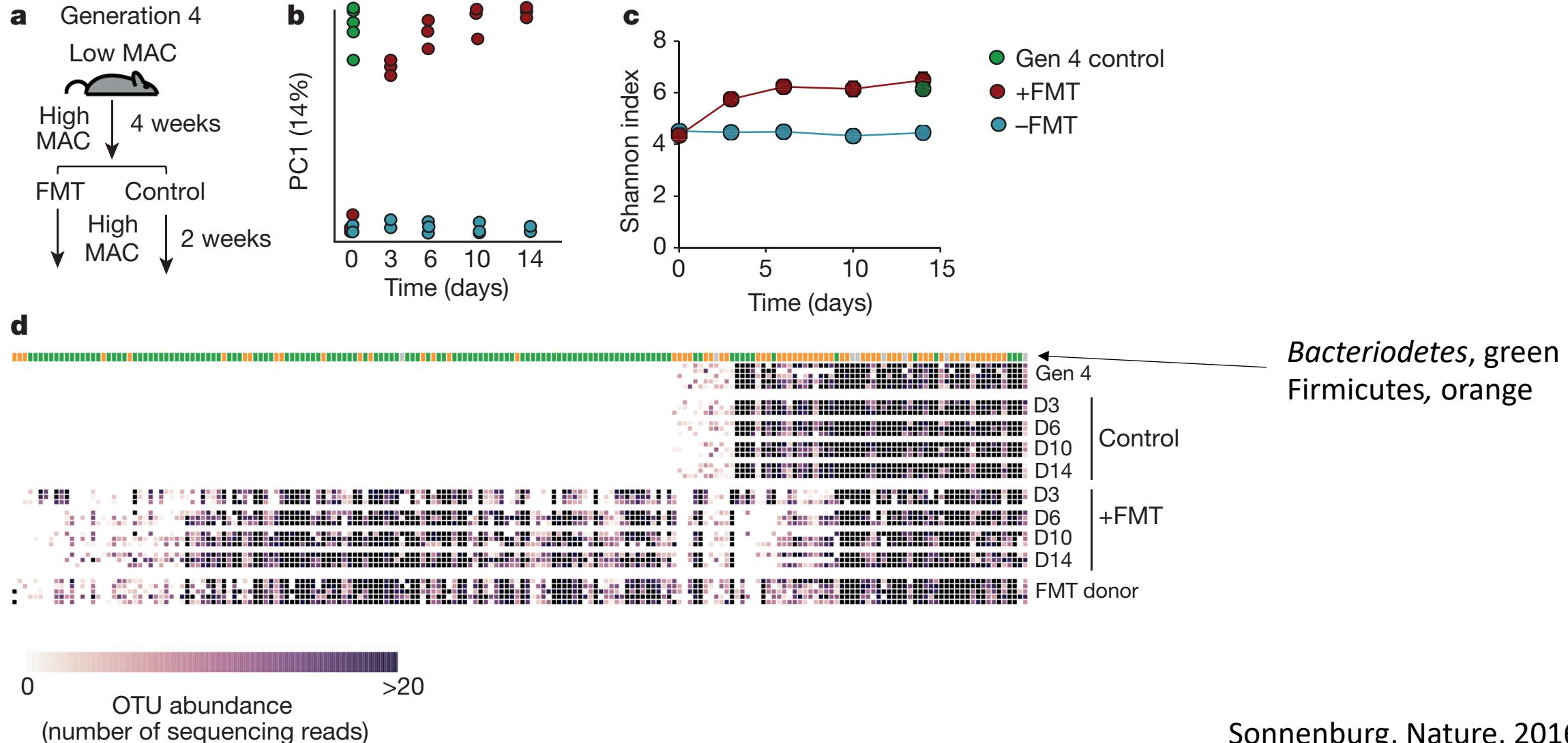


## Diet switching group

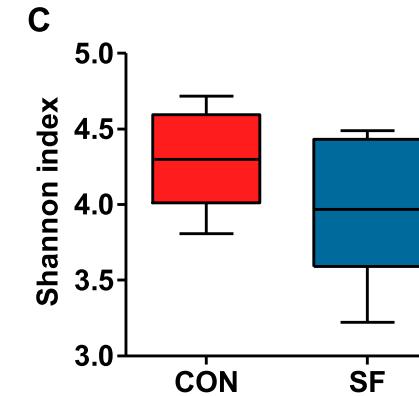
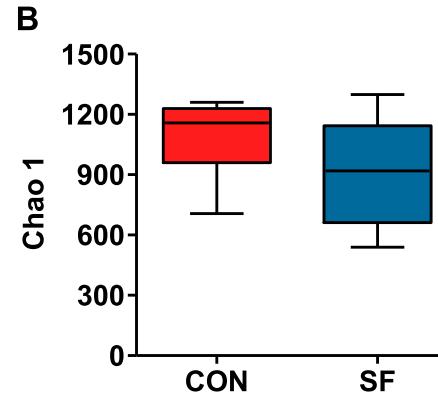
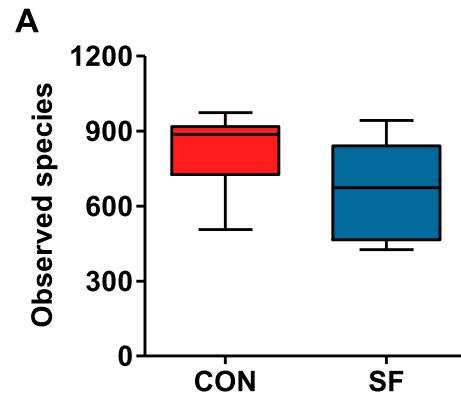
## Control group

Sonnenburg, Nature, 2016

# Generational succession: fecal transplant

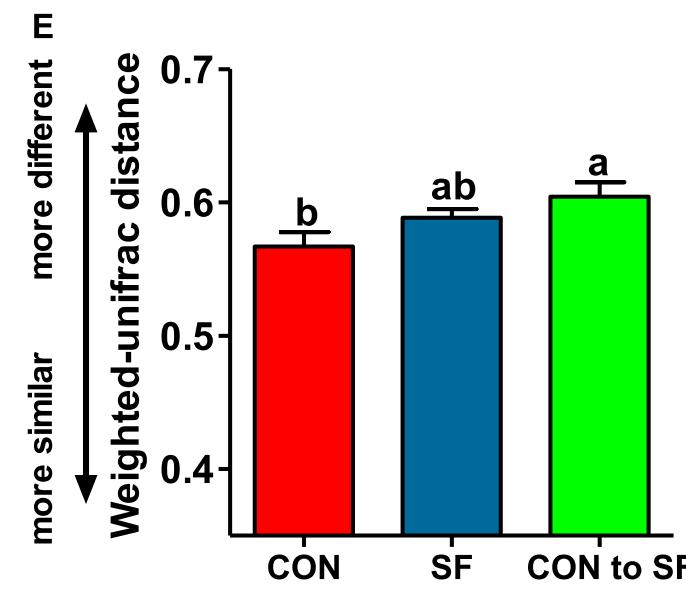
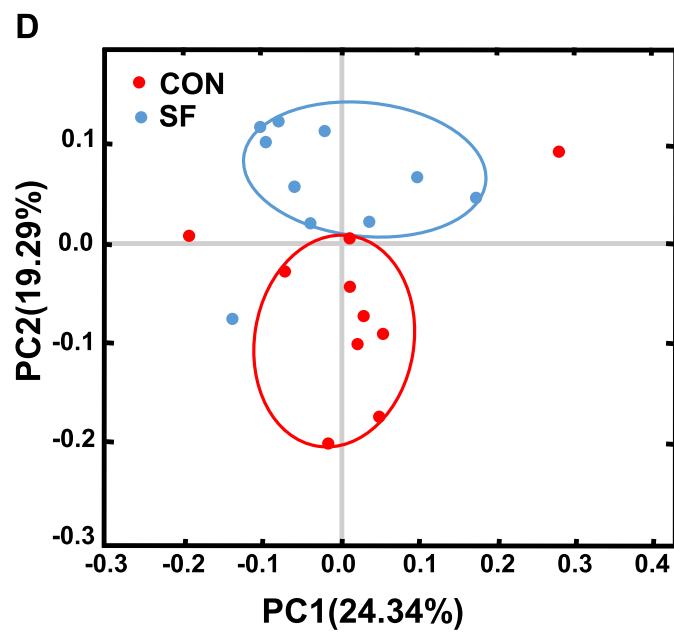


# Can we control the generational succession?



Sows given 2% soluble fiber (SF) during gestation

14d old suckling piglets

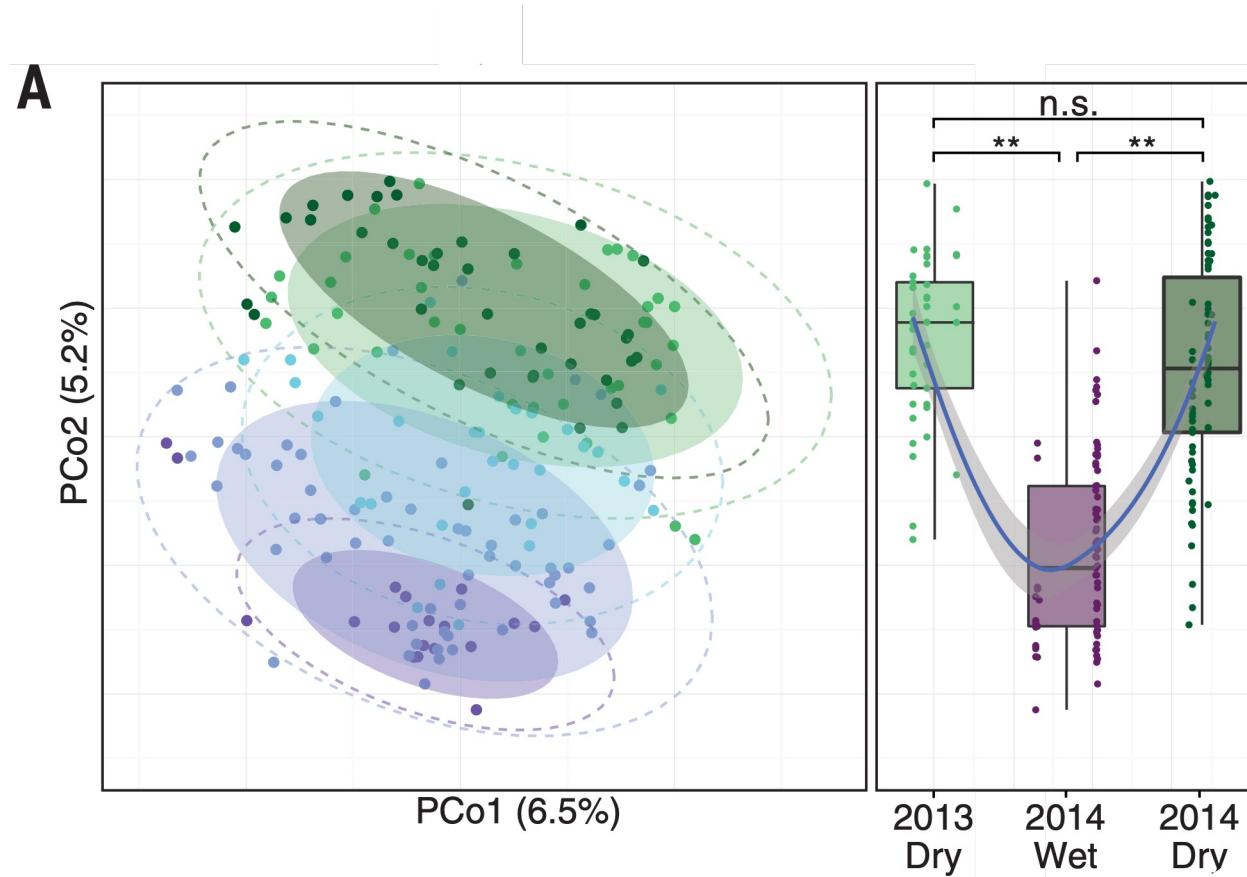


# Food for the microbiome: they are what we feed them

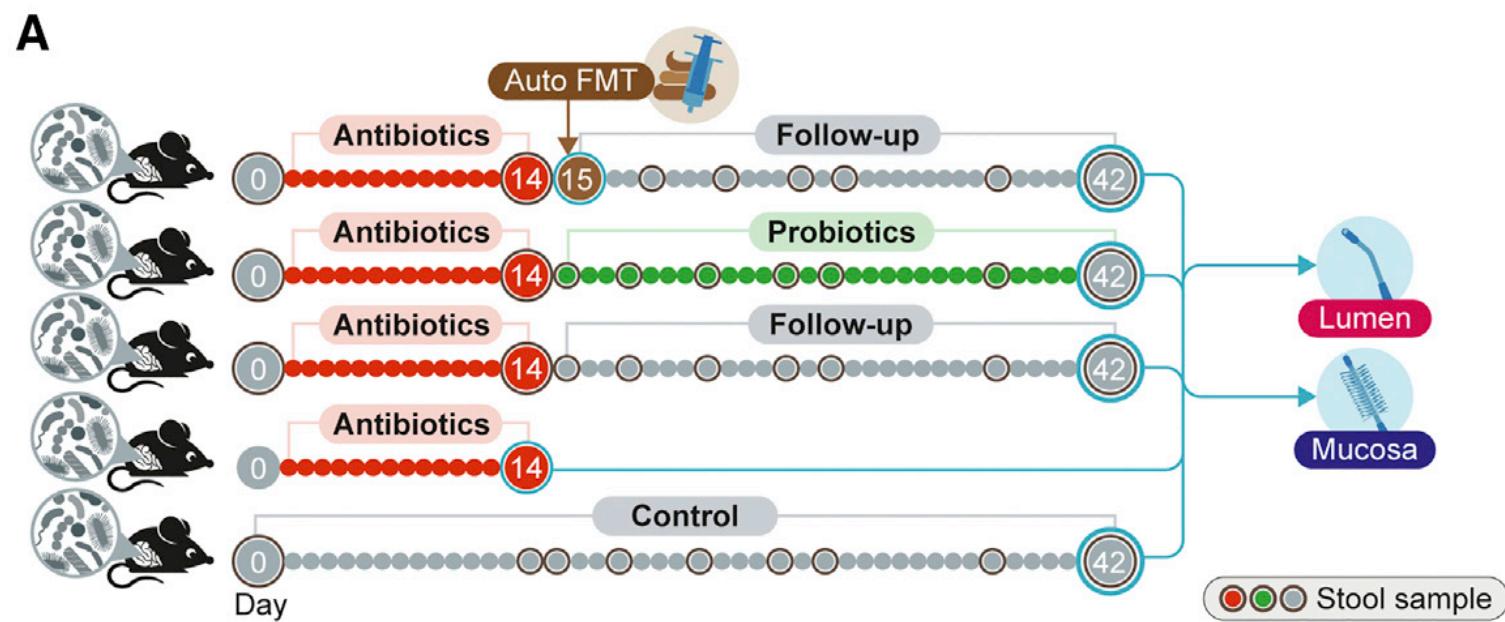
## Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania

Samuel A. Smits,<sup>1\*</sup> Jeff Leach,<sup>2,3\*</sup> Erica D. Sonnenburg,<sup>1</sup>  
Carlos G. Gonzalez,<sup>4</sup> Joshua S. Lichtman,<sup>4</sup> Gregor Reid,<sup>5</sup> Rob Knight,<sup>6</sup>  
Alphaxard Manjurano,<sup>7</sup> John Changalucha,<sup>7</sup> Joshua E. Elias,<sup>4</sup>  
Maria Gloria Dominguez-Bello,<sup>8</sup> Justin L. Sonnenburg<sup>1†</sup>

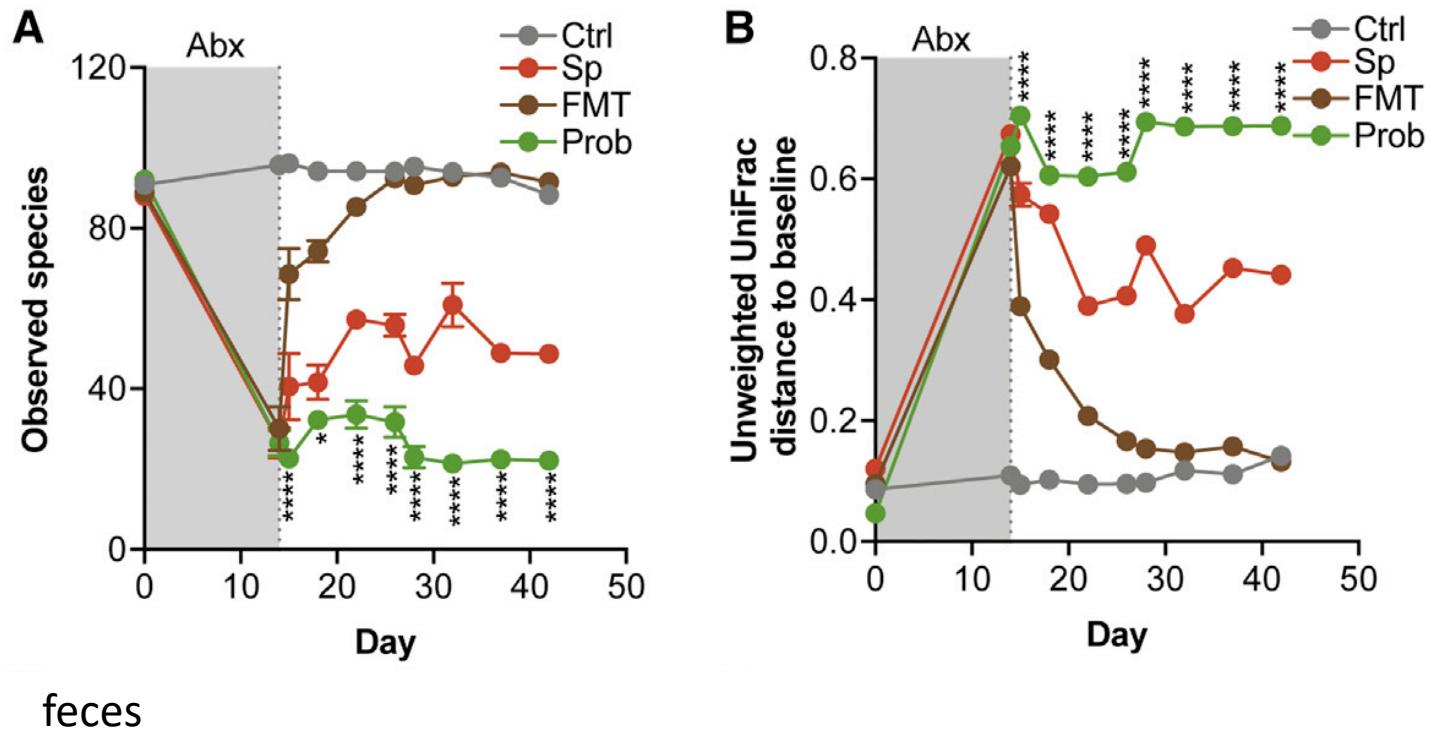
- Diet is determined by season
- Fiber-rich tubers and baobab year round
- Wet (Nov – April)
  - Berries
  - Honey
- Dry (May – Oct)
  - Hunting



# Historical contingency



# Historical contingency



# Take home messages

- Colonization depends on dispersal, inoculum, selection pressures, historical contingencies and random factors (not yet understood factors)
- If healthy, the microbiome is resilient to change
- There are microbe-pathogen-host interactions
- We must be aware of the possibilities and limitations of our methodologies and statistics.