

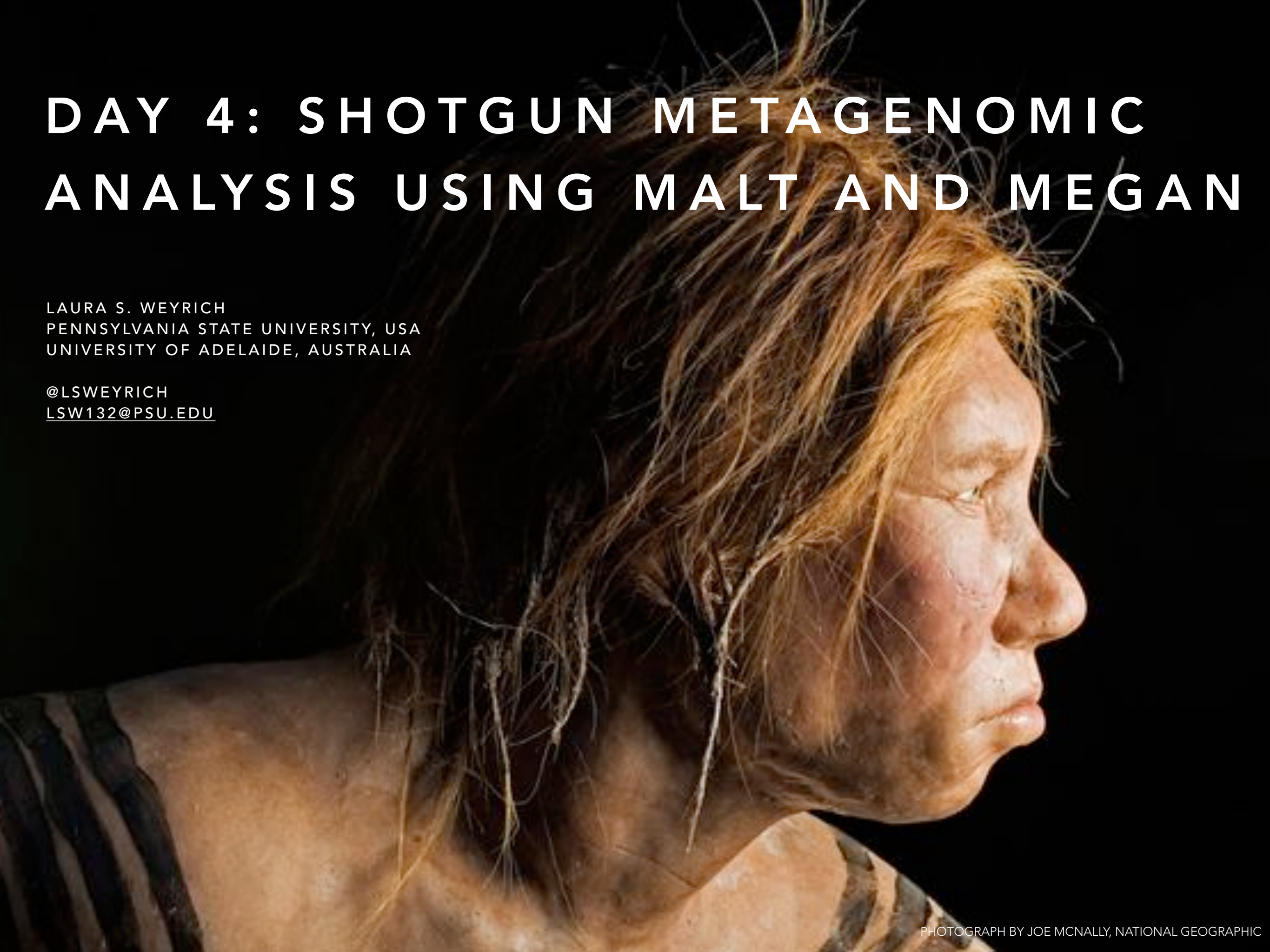
# TA Introductions

Emily Bean  
Mara Cloutier  
Diana Ayala

# DAY 4: SHOTGUN METAGENOMIC ANALYSIS USING MALT AND MEGAN

LAURA S. WEYRICH  
PENNSYLVANIA STATE UNIVERSITY, USA  
UNIVERSITY OF ADELAIDE, AUSTRALIA

@LSWEYRICH  
[LSW132@PSU.EDU](mailto:LSW132@PSU.EDU)



PHOTOGRAPH BY JOE MCNALLY, NATIONAL GEOGRAPHIC



# Microbiome NGS analysis

Metagenomic sample



**Amplicon sequencing (meta-barcoding):** Identify organisms within a sample via one 'barcode' gene

DNA isolation



Metagenomic library construction



Amplicon libraries – constructed via PCR

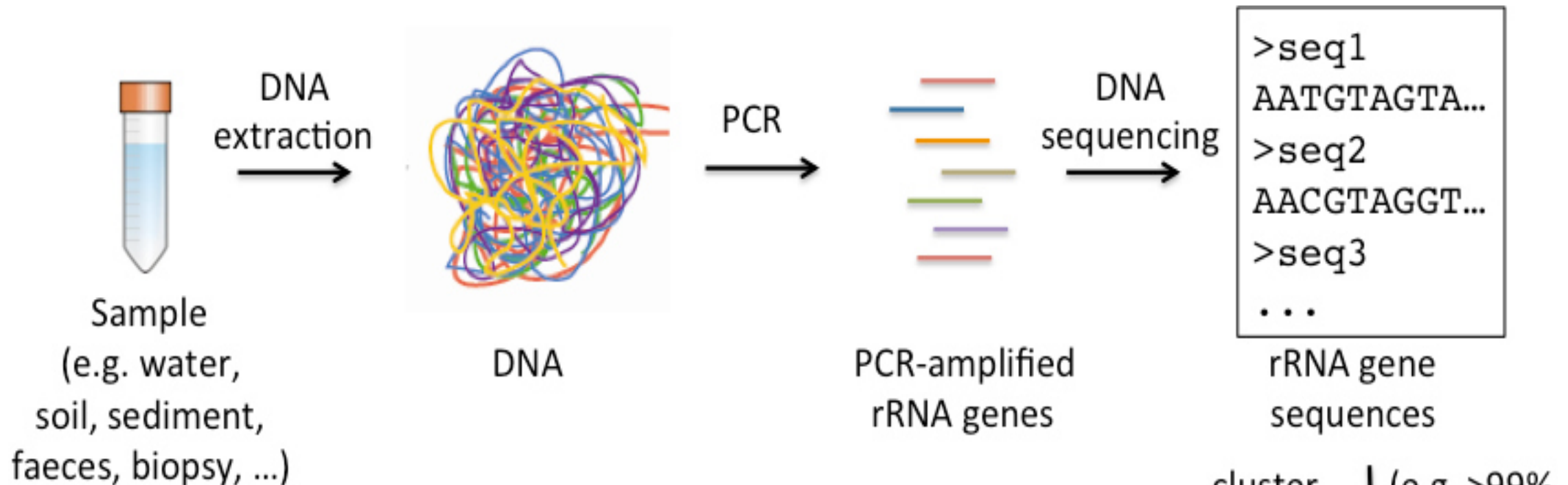
Sequencing

Genus ID

Sequence analysis

# Amplicon Based Sequencing

Wet Lab



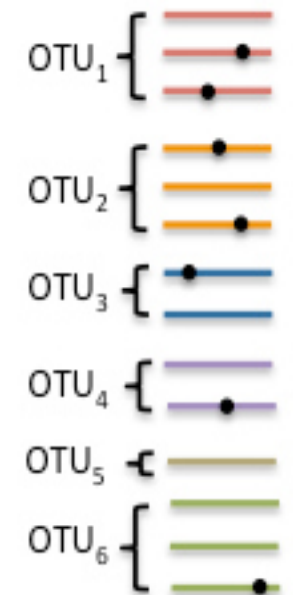
Bioinformatics

OTU	Species	Sample1	Sample2	Sample3
1	E.coli	17	0	335
2	S.aurus	231	11800	45
3	unknown	30	0	0
...	...	...	...	...

Counts of OTUs  
per sample

BLAST-search  
rRNA sequence  
database  
←  
with millions  
of taxonomically  
classified  
rRNA sequences  
(e.g. RDP, Silva)

Operational  
Taxonomic  
Units  
(OTUs)



cluster  
sequences ↓ (e.g. >99%  
Identity)

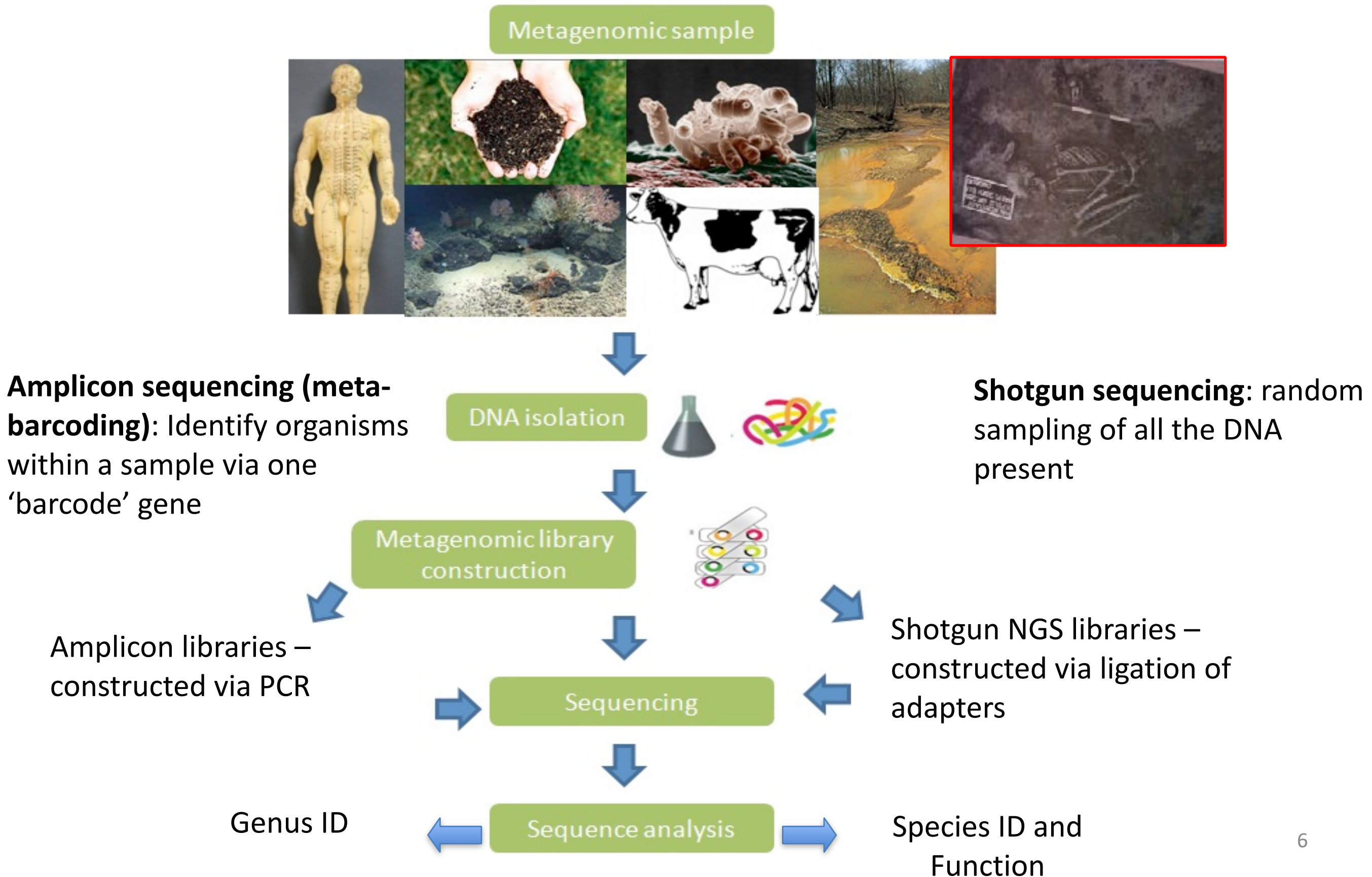


# Issues with amplicon analysis

- Sample Preservation (DNA size)
- Laboratory, reagents, and technician contamination can significantly alter findings.
- Biases from amplification.
- Species IDs are difficult.
- Limited information during bioinformatic analysis.

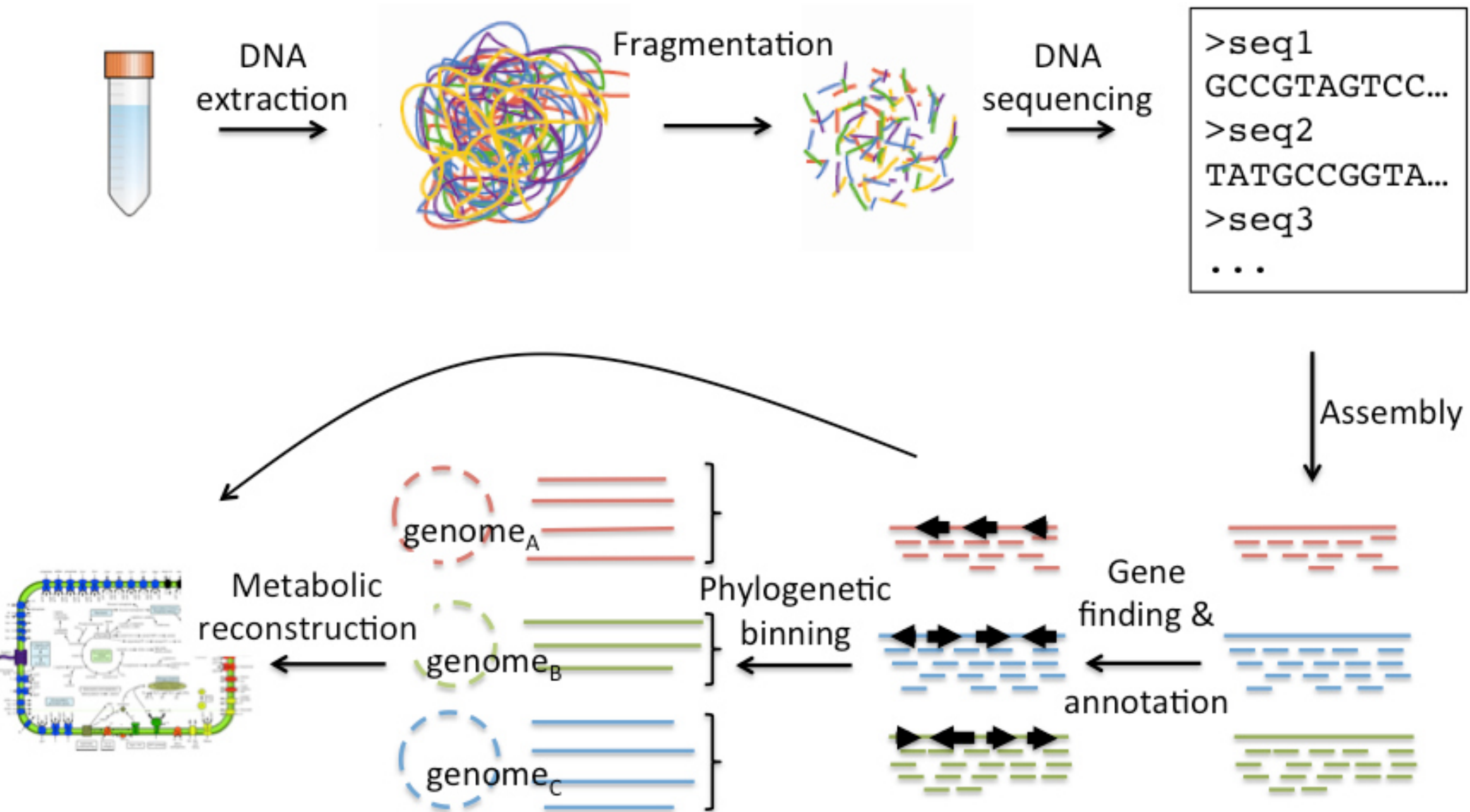


# Microbiota/Microbiome Analysis: NGS





# Shotgun Sequencing Metagenomics



# Shotgun Pipelines for Taxonomic Assignment

- **Align to a single reference**

- Example: BWA

<http://bio-bwa.sourceforge.net/>

- **Align to genomes**

- Example: MALT(DIAMOND)/MEGAN

<http://megan.informatik.uni-tuebingen.de/>

- **Using marker genes**

- Example: HUMAnN 2.0

<https://huttenhower.sph.harvard.edu/humann>

- **Binning sequences de novo**

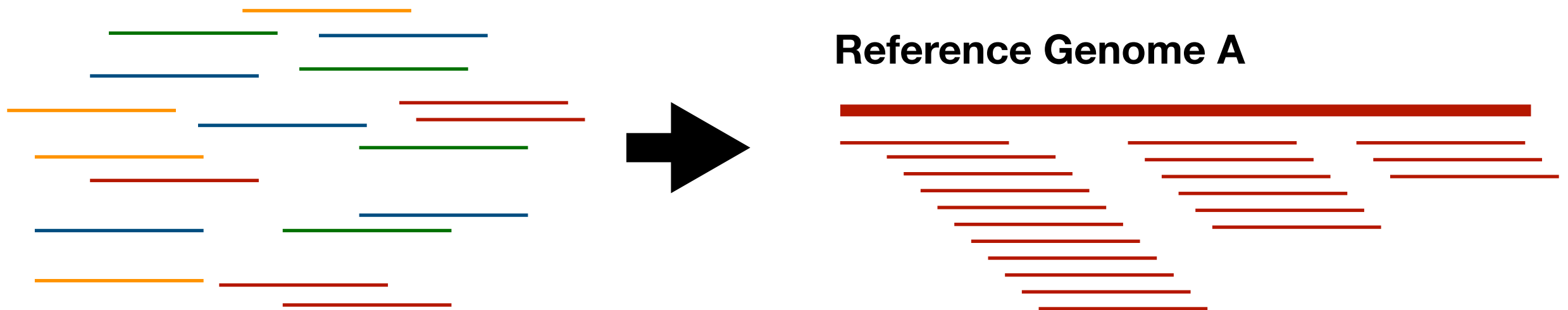
- Example: GroopM

<https://ecogenomics.github.io/GroopM/>



# Aligning to a Single Reference

- **Align to a single reference** <http://bio-bwa.sourceforge.net/>
  - Example: BWA



## Limitations:

Only a single genome  
Must know reference a priori

## Benefits:

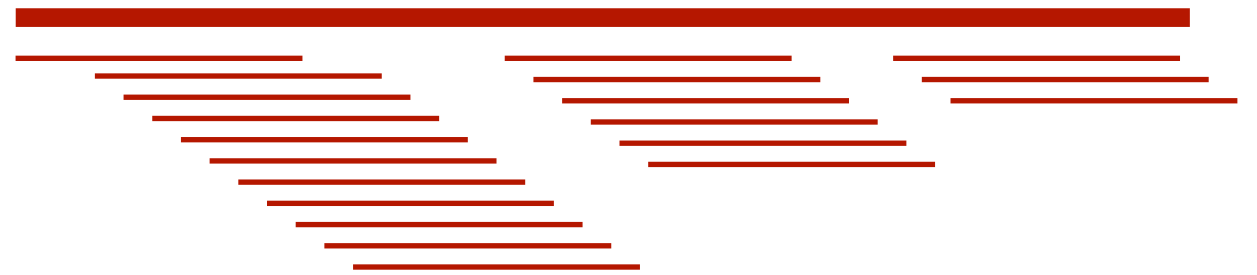
Very accurate species/strain ID  
Can use genomes for downstream functions

# Aligning to multiple genomes

- **Align to multiple genomes**
  - MALT(DIAMOND)/MEGAN

<http://megan.informatik.uni-tuebingen.de/>

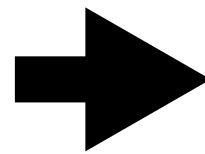
**Reference Genome A**



**Reference Genome B**



**Reference Genome XYZ...**



## **Limitations:**

**Lots of memory needed**  
**Need to know reference genomes**

## **Benefits:**

**Accurate species/strain ID for lots of taxa**  
**Can use genomes for downstream functions**

# Using Marker Genes

- Using marker genes

- Example: HUMAnN2 <https://huttenhower.sph.harvard.edu/humann>



Franzosa, et al., *Nature Methods*, 2018

## Limitations:

- Need to know reference genomes
- Markers may miss certain strains

## Benefits:

- Accurate species/strain ID for lots of taxa
- Can use genomes for downstream functions
- Faster than whole genome alignments



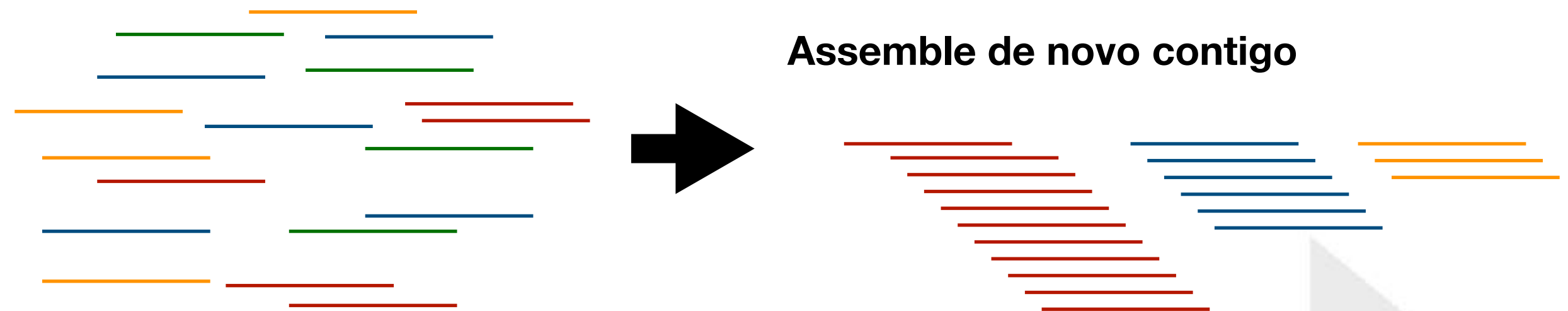
# Binning de novo

- Binning sequences de novo

- GroopM

<https://ecogenomics.github.io/GroopM/>

## Assemble de novo contigo



## Limitations:

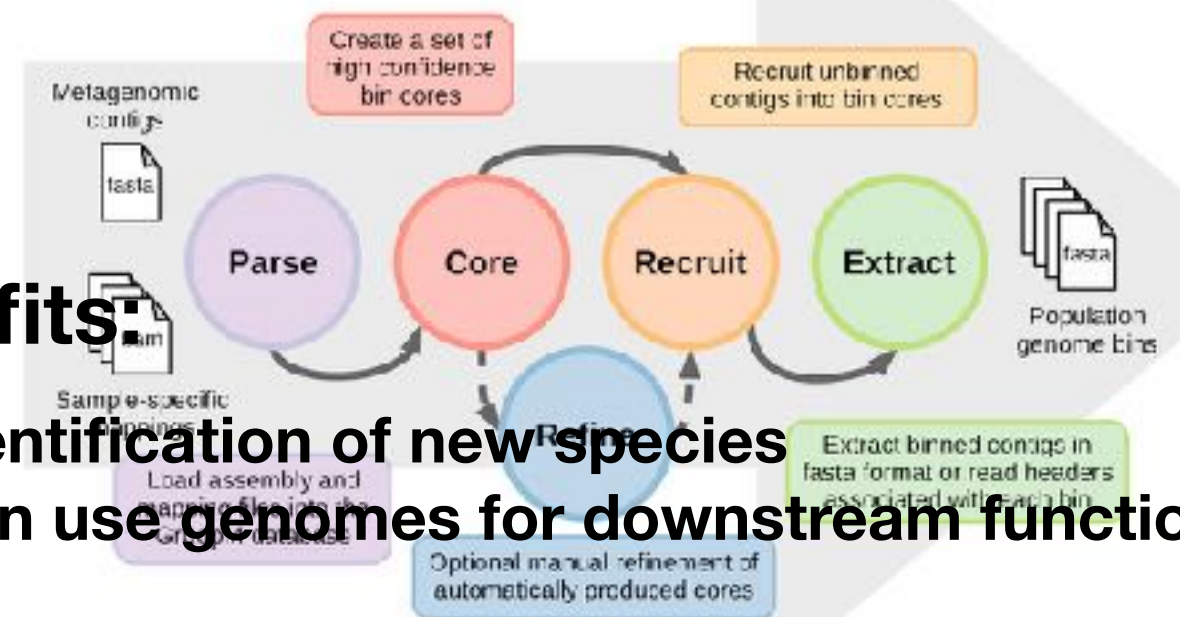
Need long sequences

May not be able to bin all contigs

## Benefits:

Identification of new species

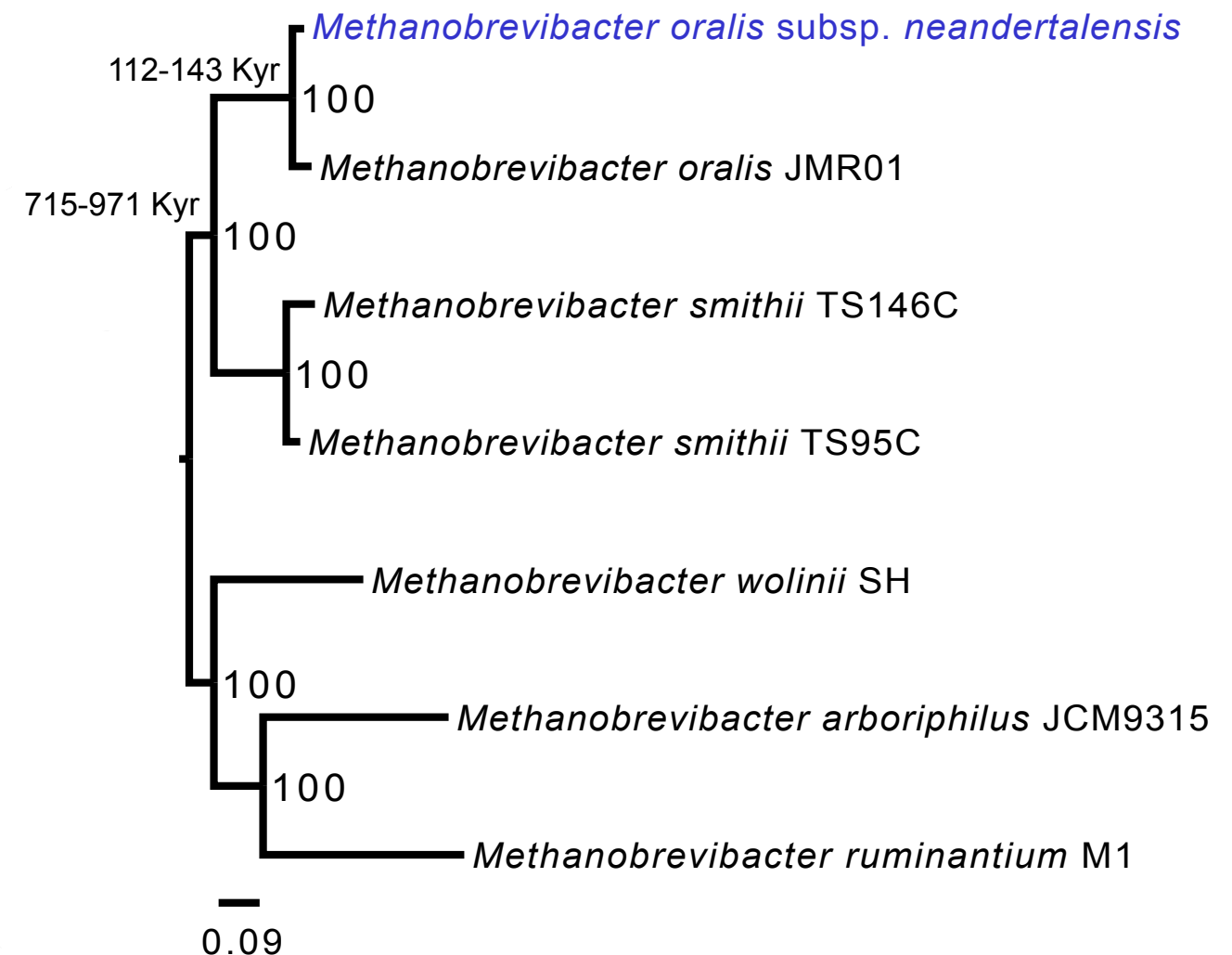
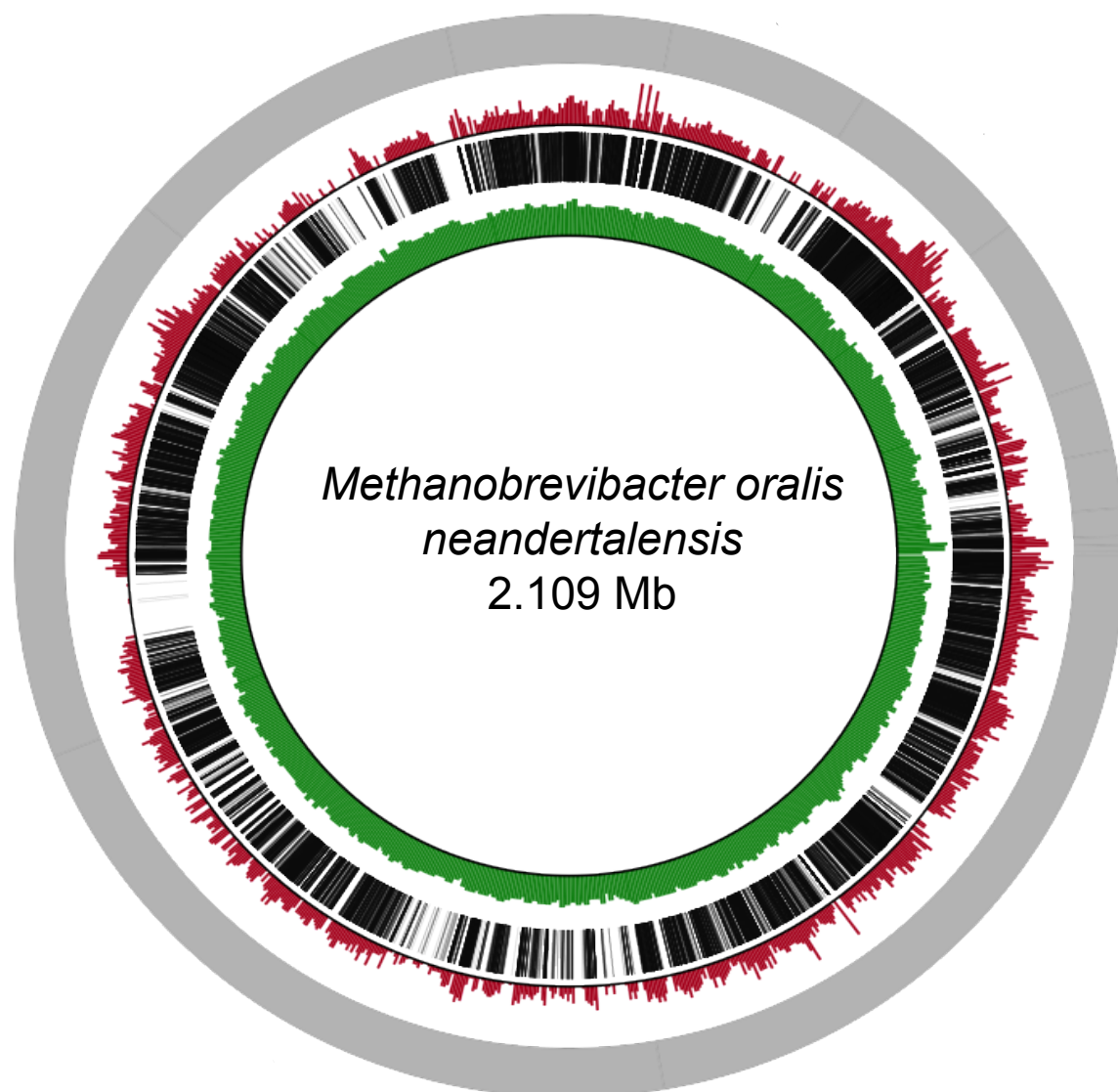
Can use genomes for downstream functions



# What can you do?

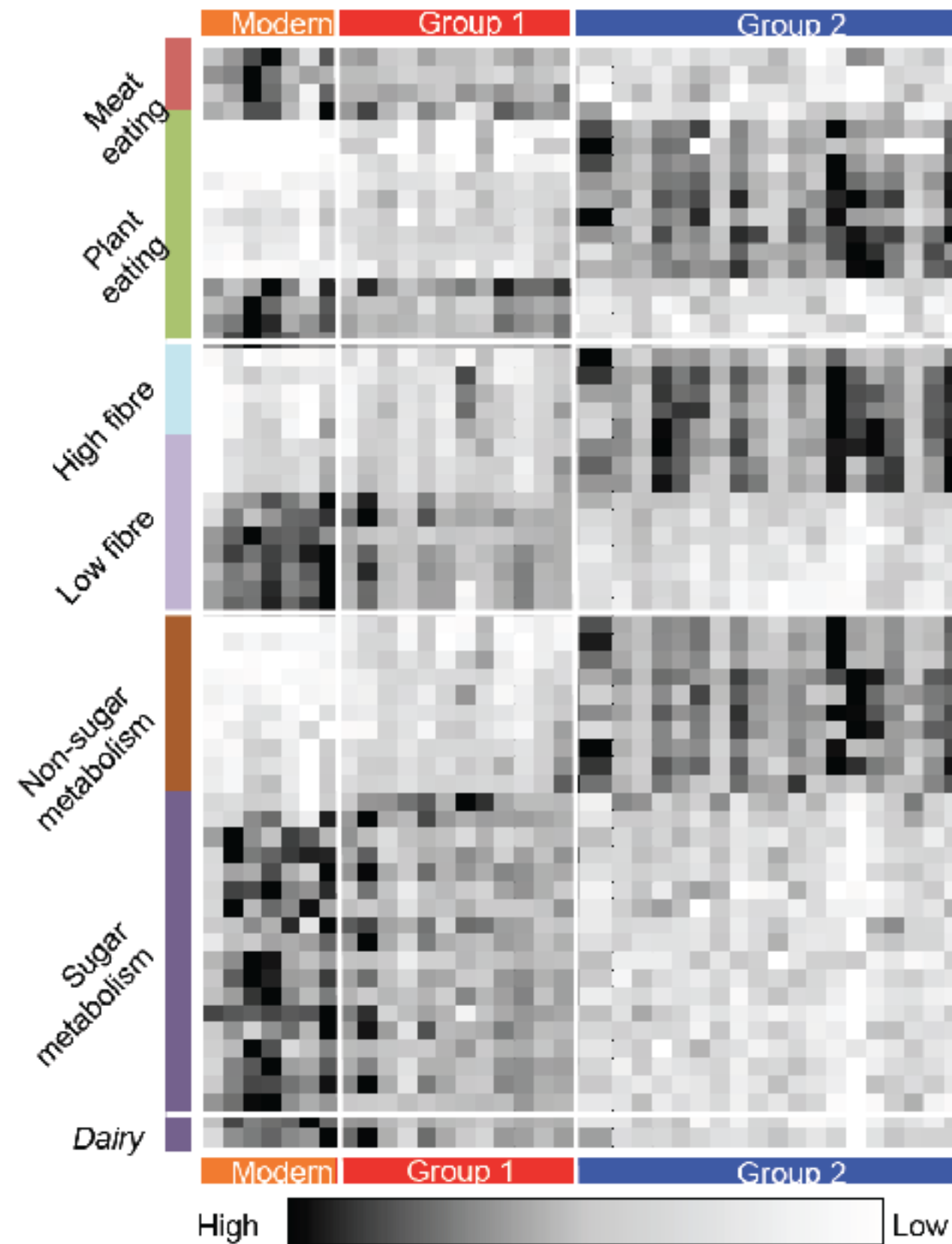
## Taxonomic, Functional, and Genomic Analysis

Reference Genome				Mapped Reads								
Reference Genome	Sequence Reference Number	Length (Mbps)	GC content (%)	Base Covered	Unique	Depth	Average					
				(Mbps)	Hits	(Avg coverage)	Read Length	5p-C-T	3p-G-A	DeltaD	DeltaS	Lambda
Methanobrevibacter oralis JMR01	NZ_CBWS000000000	2.107	27.8	0.941	370115	15.16	58.67	0.33	0.36	0.05	1	0.38
Candidatus Saccharibacteria oral TM7	NZ_CP007496.1	0.705	44.5	0.131	108919	5.83	52.46	0.37	0.41	0.01	1	0.38
Campylobacter gracilis ATCC 33236	NZ_CP012196.1	2.282	46.6	1.199	94472	2.40	51.7	0.38	0.41	0.01	1	0.36
Propionibacterium propionicum F0230a	NZ_018142.1	3.449	66.1	2.083	130748	1.89	48.85	0.37	0.43	0	1	0.43
Fretibacterium fastidiosum	gi_296110870	2.728	55.5	1.466	121822	2.43	48	0.39	0.43	0	1	0.41
Eubacterium infirmum F0142	NZ_AGWI000000000	1.9	40.1	0.176	52170	10.73	51.53	0.33	0.38	0.02	1	0.41
Peptostreptococcus stomatis DSM 17678	GCF_000147675.1	1.988	36.7	1.222	94743	2.90	54.62	0.36	0.4	0.02	1	0.38
Eubacterium sphenum ATCC 49989	NZ_GG688422.1	1.084	40.6	0.261	23124	3.46	52.87	0.37	0.41	0.03	1	0.36



# What can you do?

## Taxonomic, Functional, and Genomic Analysis





# Pros/Cons of NGS Approaches

## Amplicon Sequencing

## Shotgun Sequencing

### Pros

Cheap  
Rapid  
Technically Easy  
Less Complex

Functional Information  
Less amplification Bias  
Species/Strain Level ID  
Genome Assembly

### Cons

No Functional Data  
PCR Biases  
ID limited to genera

More expensive  
Labor intensive  
Limited Analysis Programs  
Need Deep Coverage?

**Questions?**

**What types of analyses  
would we want to do with  
metagenomic data?**

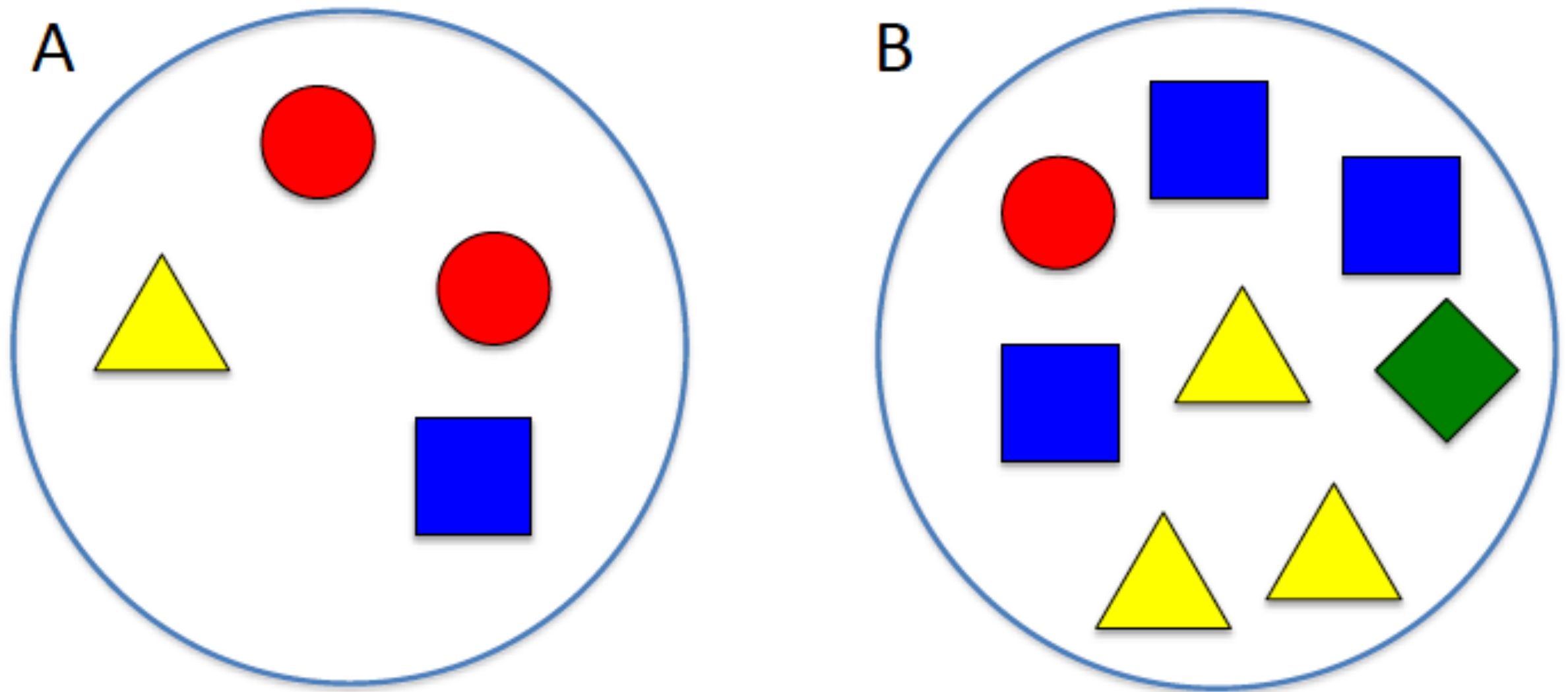


# Analyses

- Taxonomic and Functional Alpha diversity
- Taxonomic and Functional Beta diversity
- Species Identification
  - Phylogenetic analyses
  - Individual adaptations
  - Evolutionary analyses

**Alpha diversity**

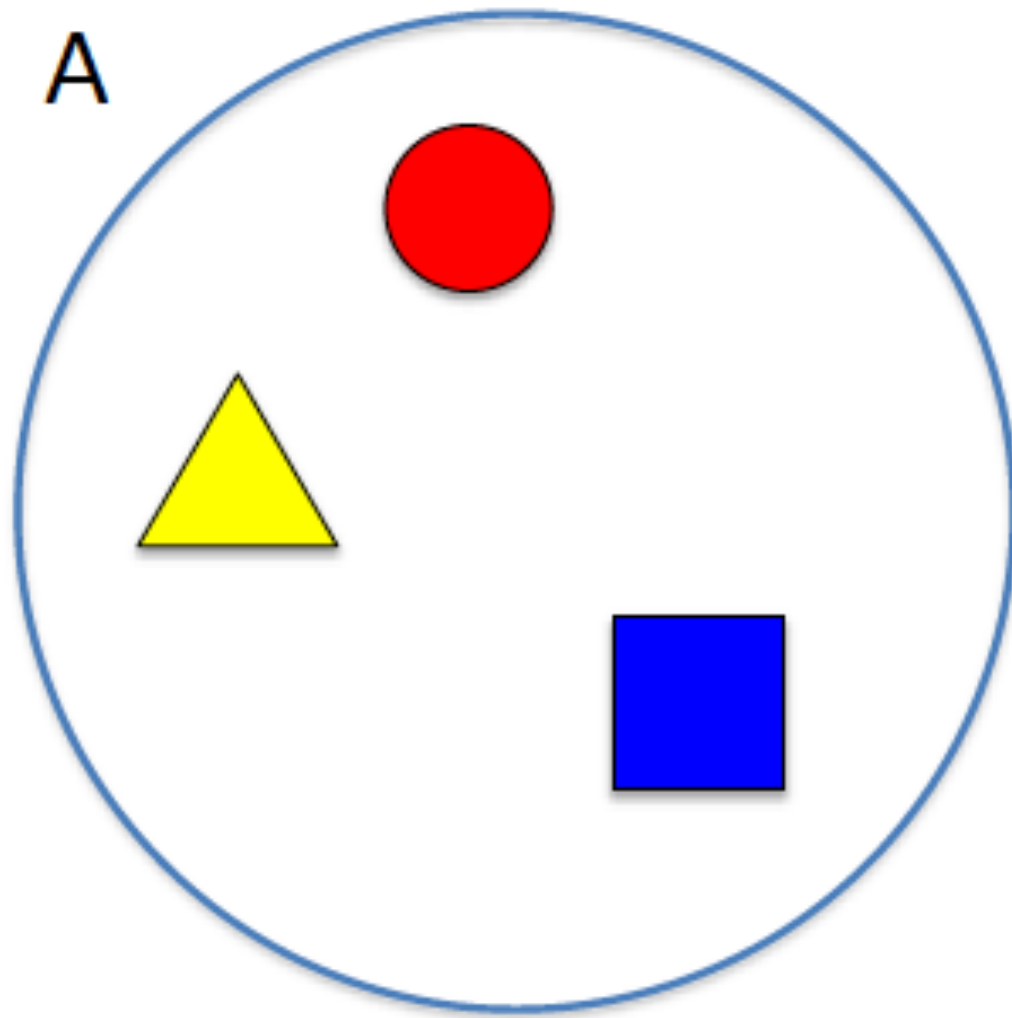
# Alpha diversity



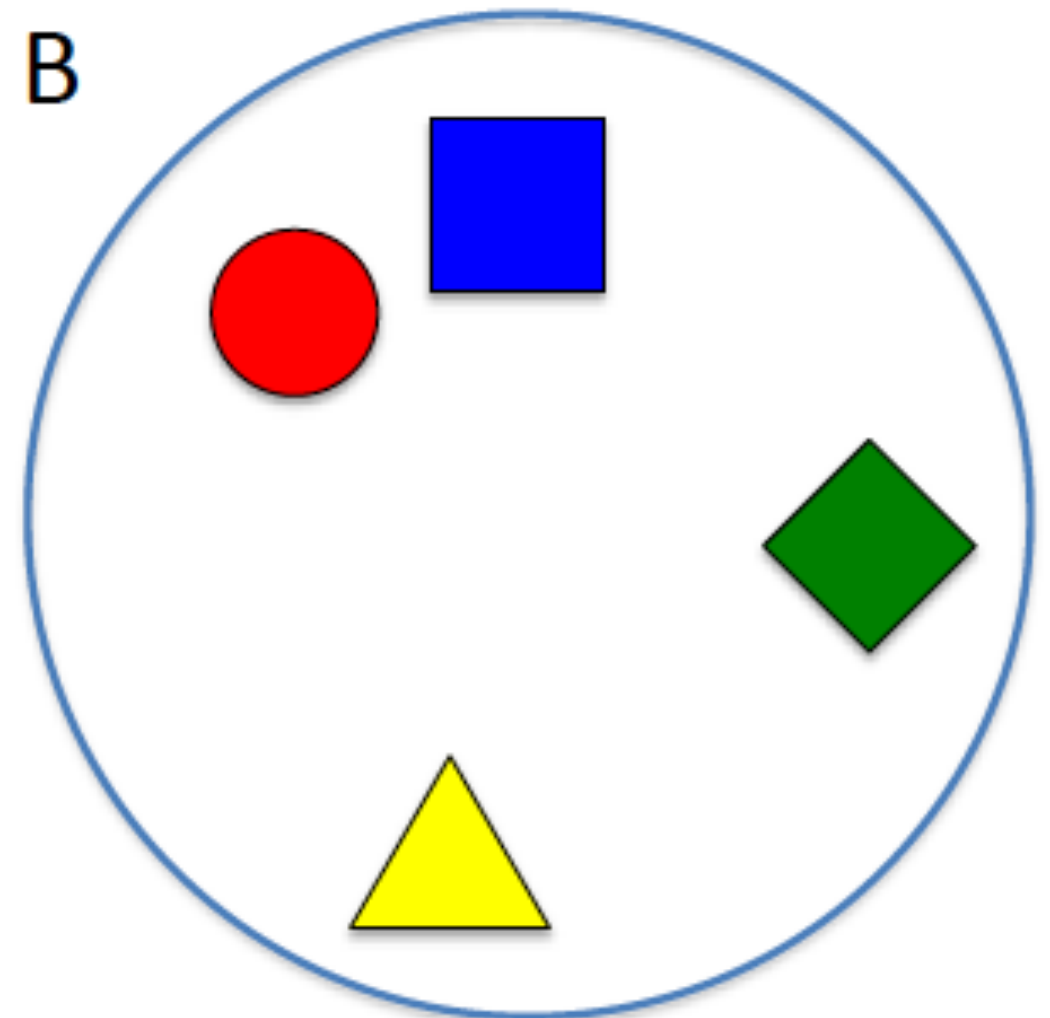
Ref: Daniel McDonald; American Gut Project



# Alpha diversity



Alpha diversity = 3

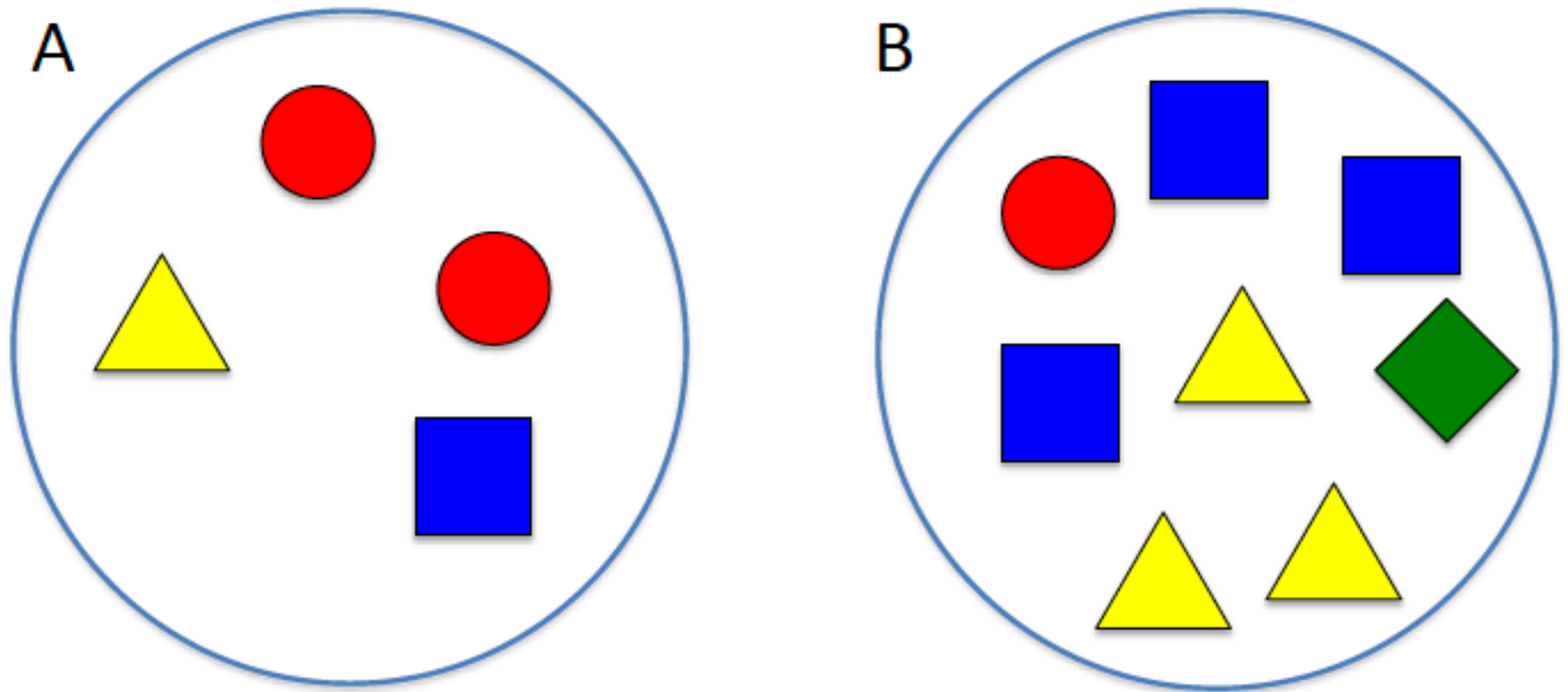


Alpha diversity = 4

Sample B is more diverse than sample A

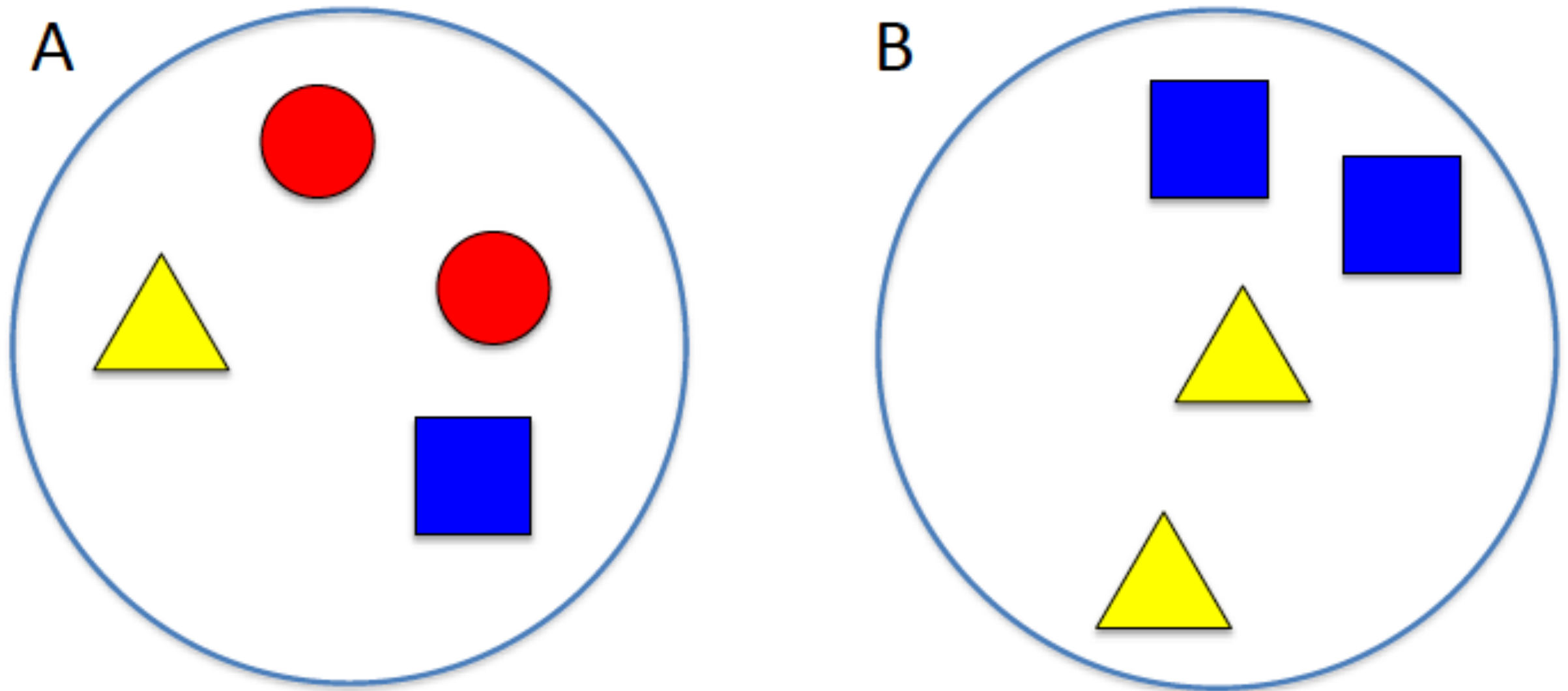
Ref: Daniel McDonald; American Gut Project

# Alpha diversity



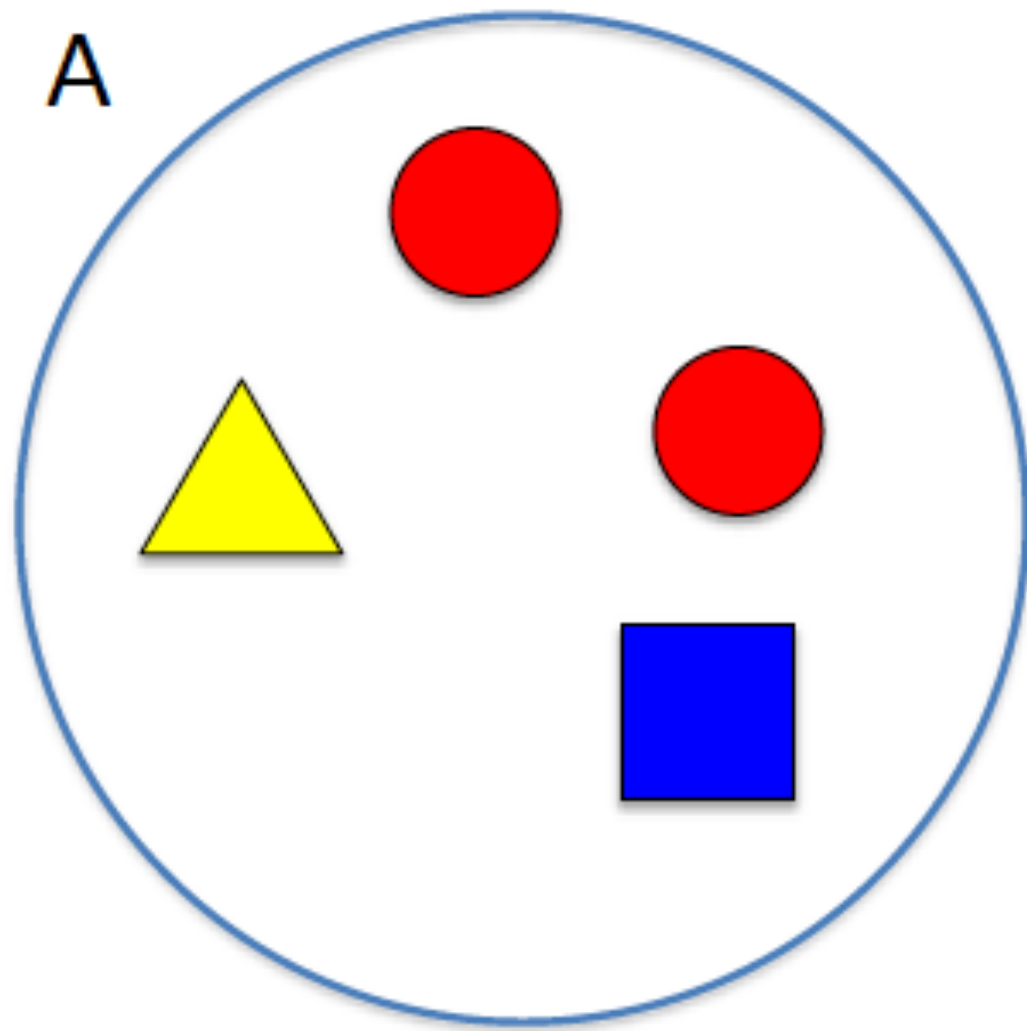
Ref: Daniel McDonald; American Gut Project

# Rarefaction

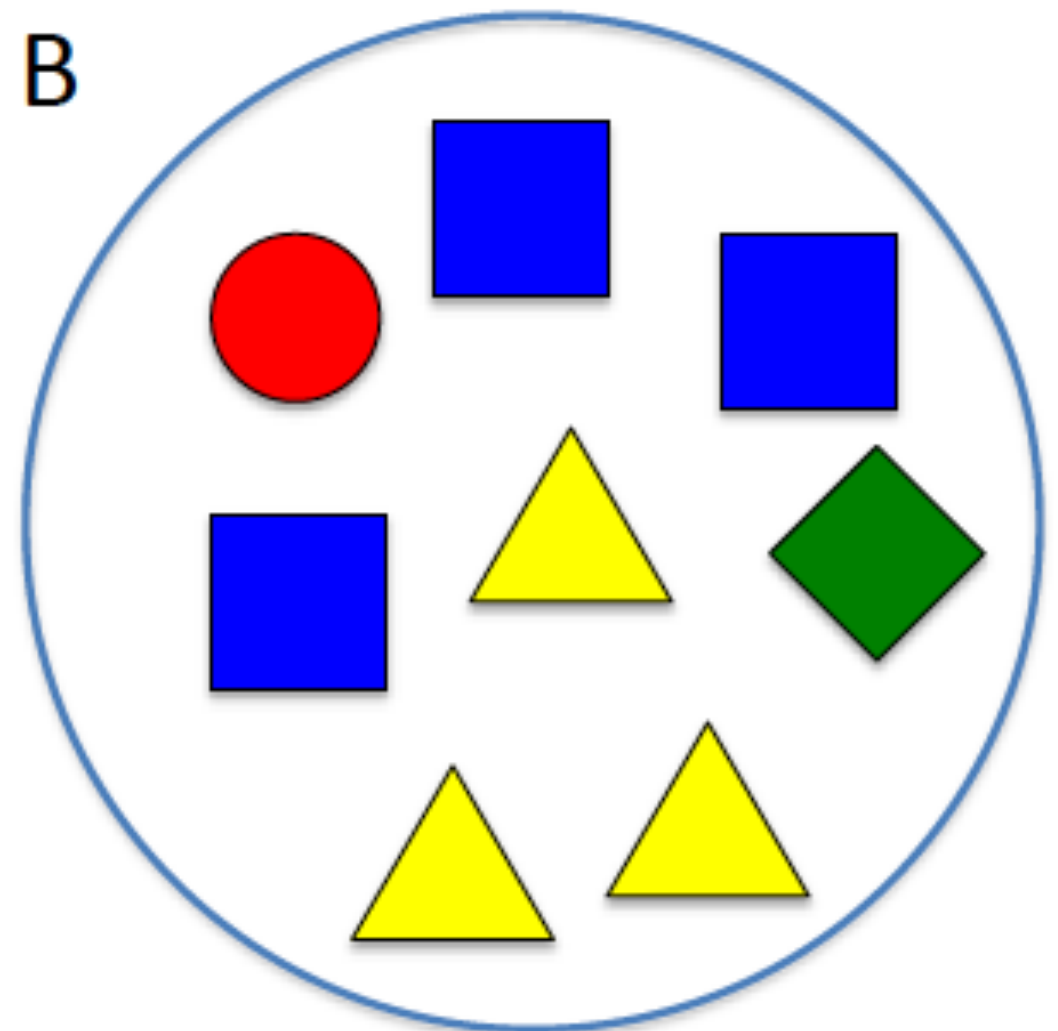


Ref: Daniel McDonald; American Gut Project

# Rarefaction



Alpha diversity = 3  
Total sequences = 4

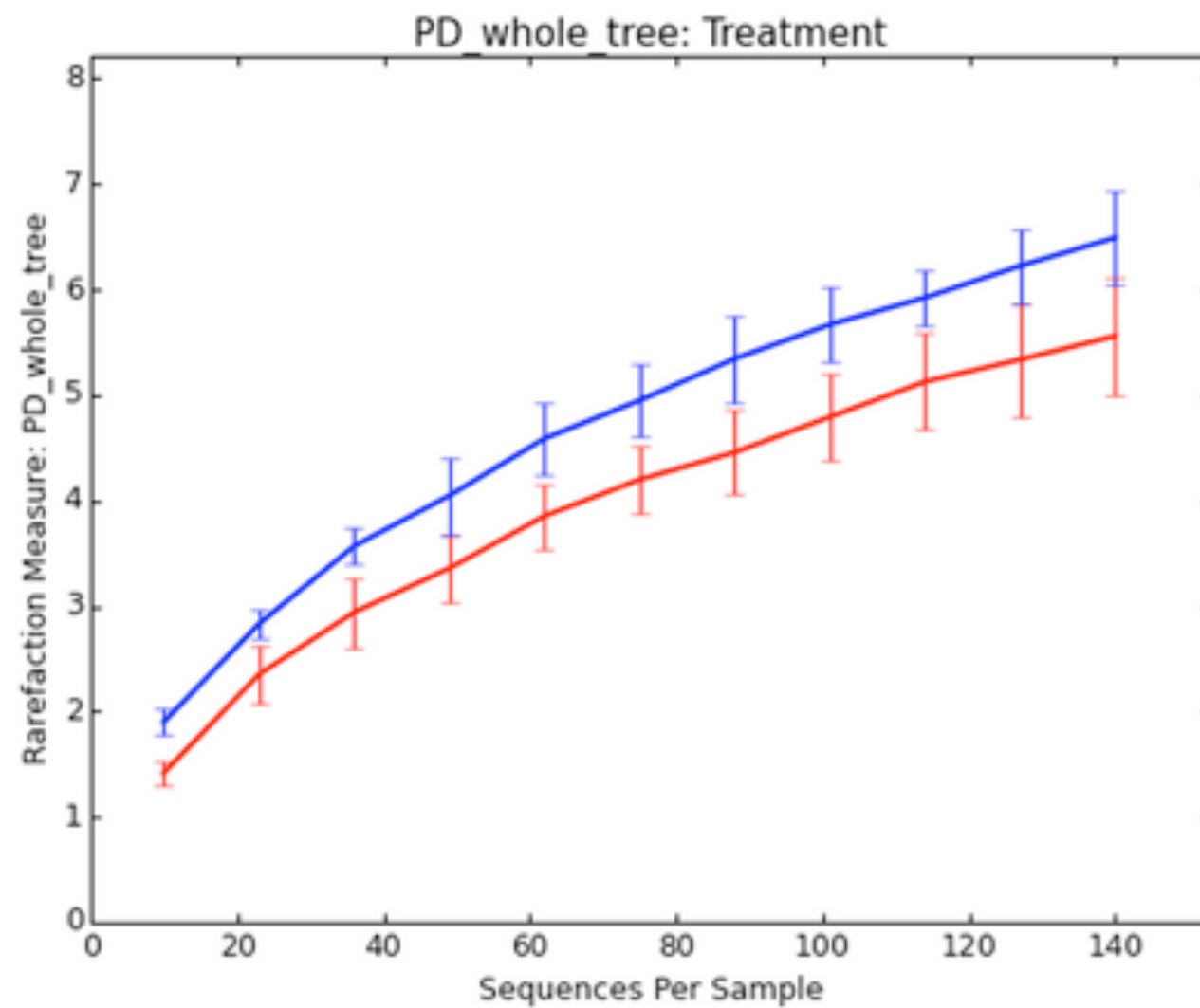


Alpha diversity = 4  
Total sequences = 8

Rarefy to 4 sequences

Ref: Daniel McDonald; American Gut Project

Select a Metric:  Select a Category:



Show Categories:

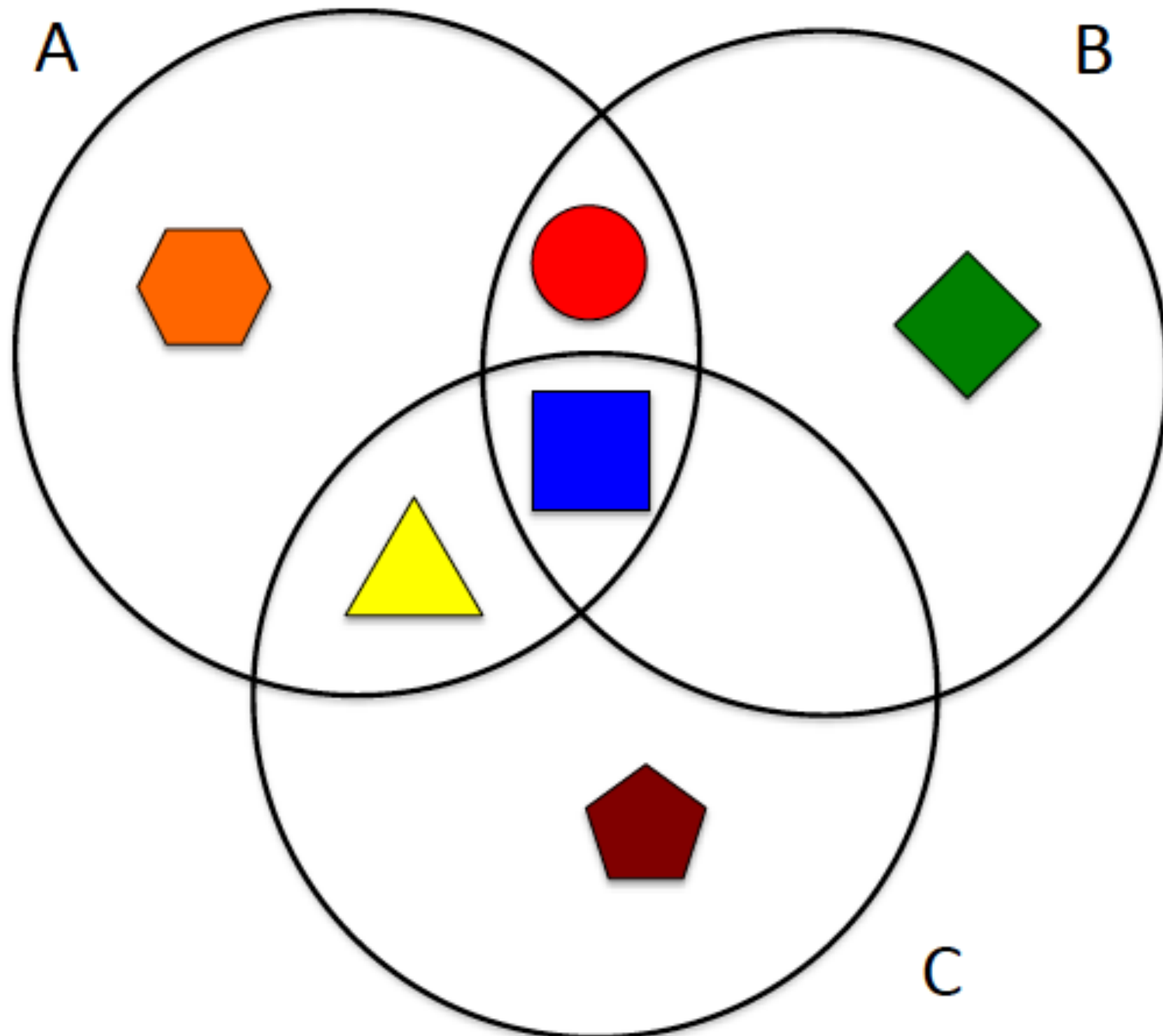
#### Legend

- Control
- Fast



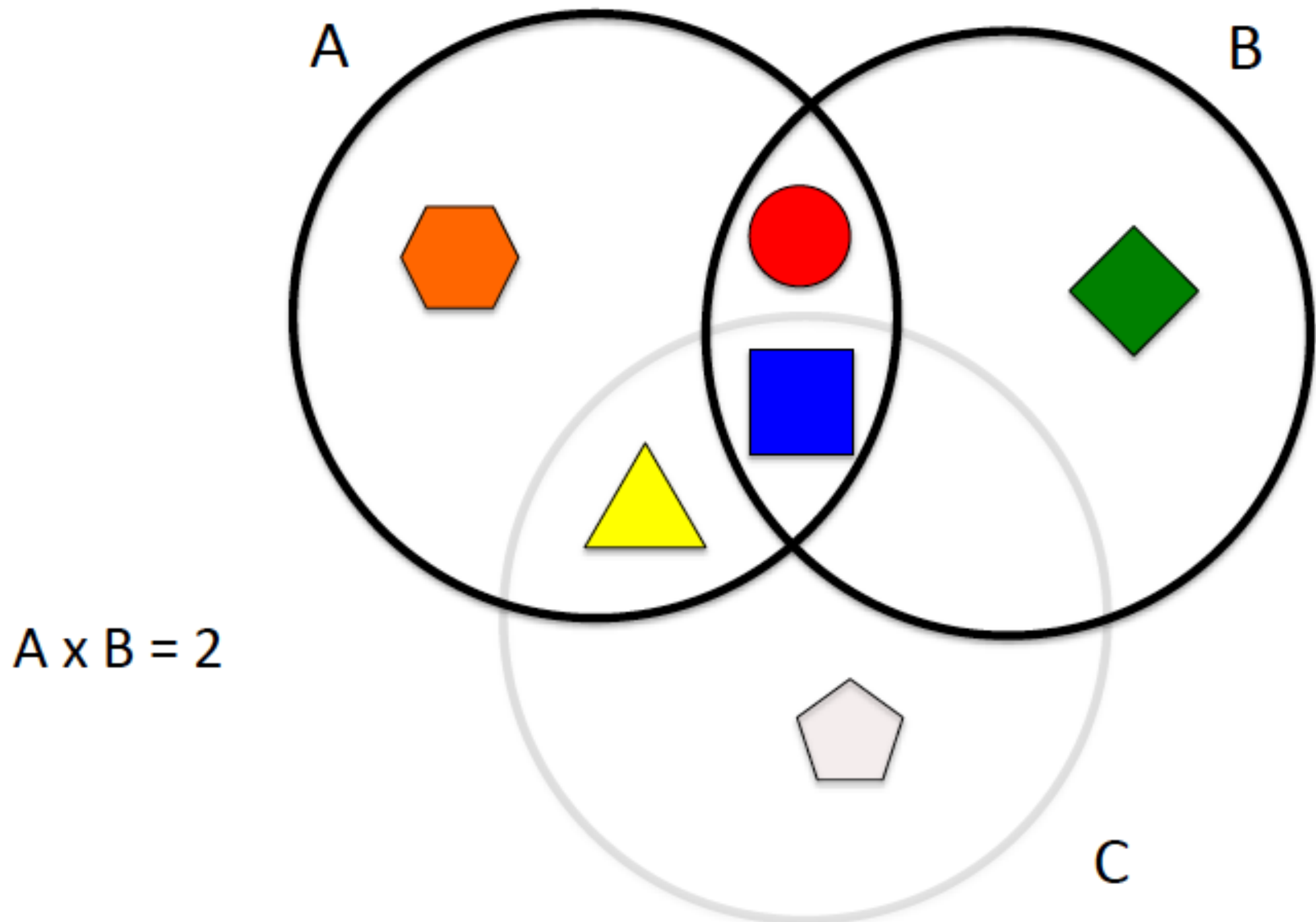
**Beta diversity**

# Beta diversity



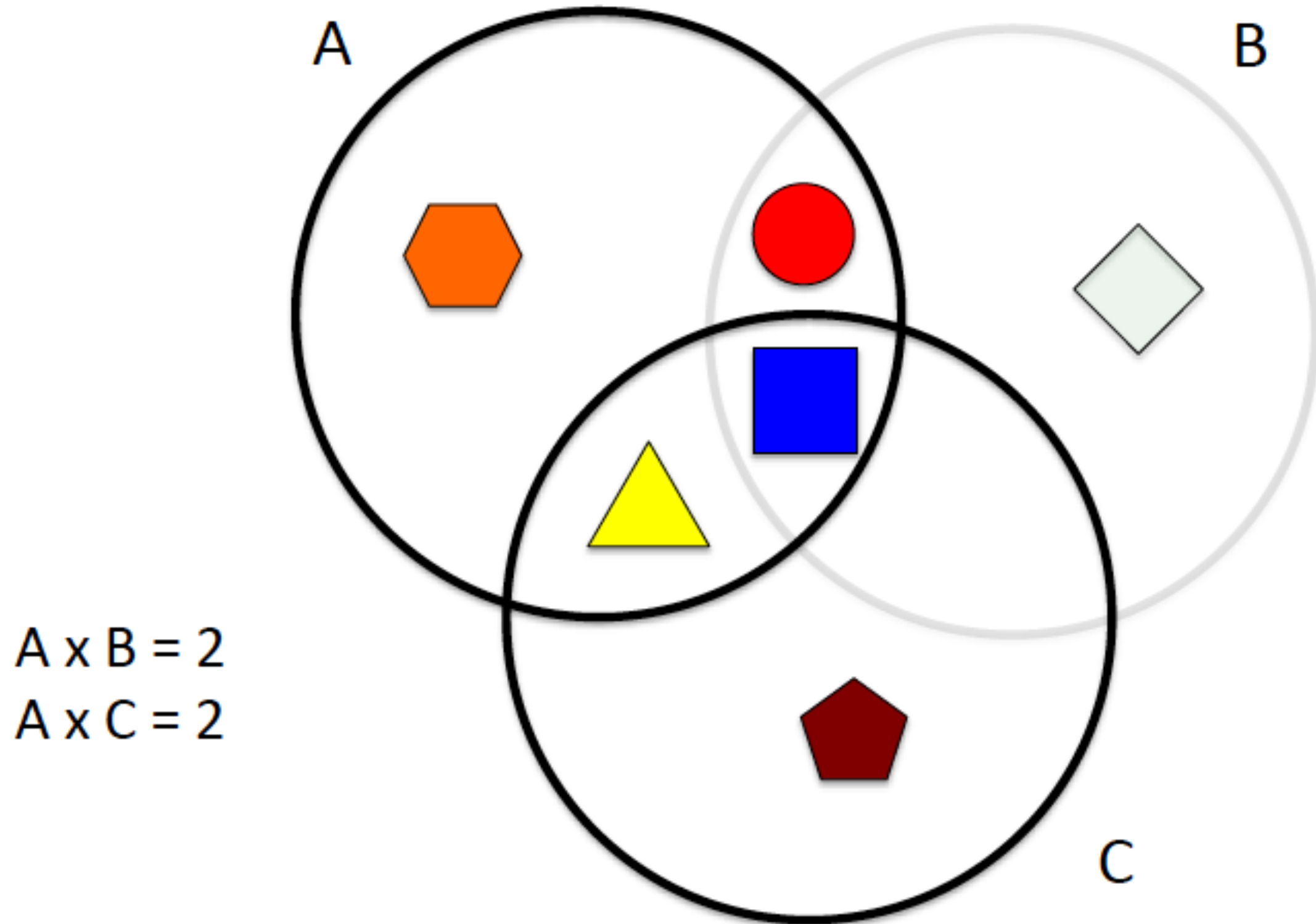
Ref: Daniel McDonald; American Gut Project

# Beta diversity



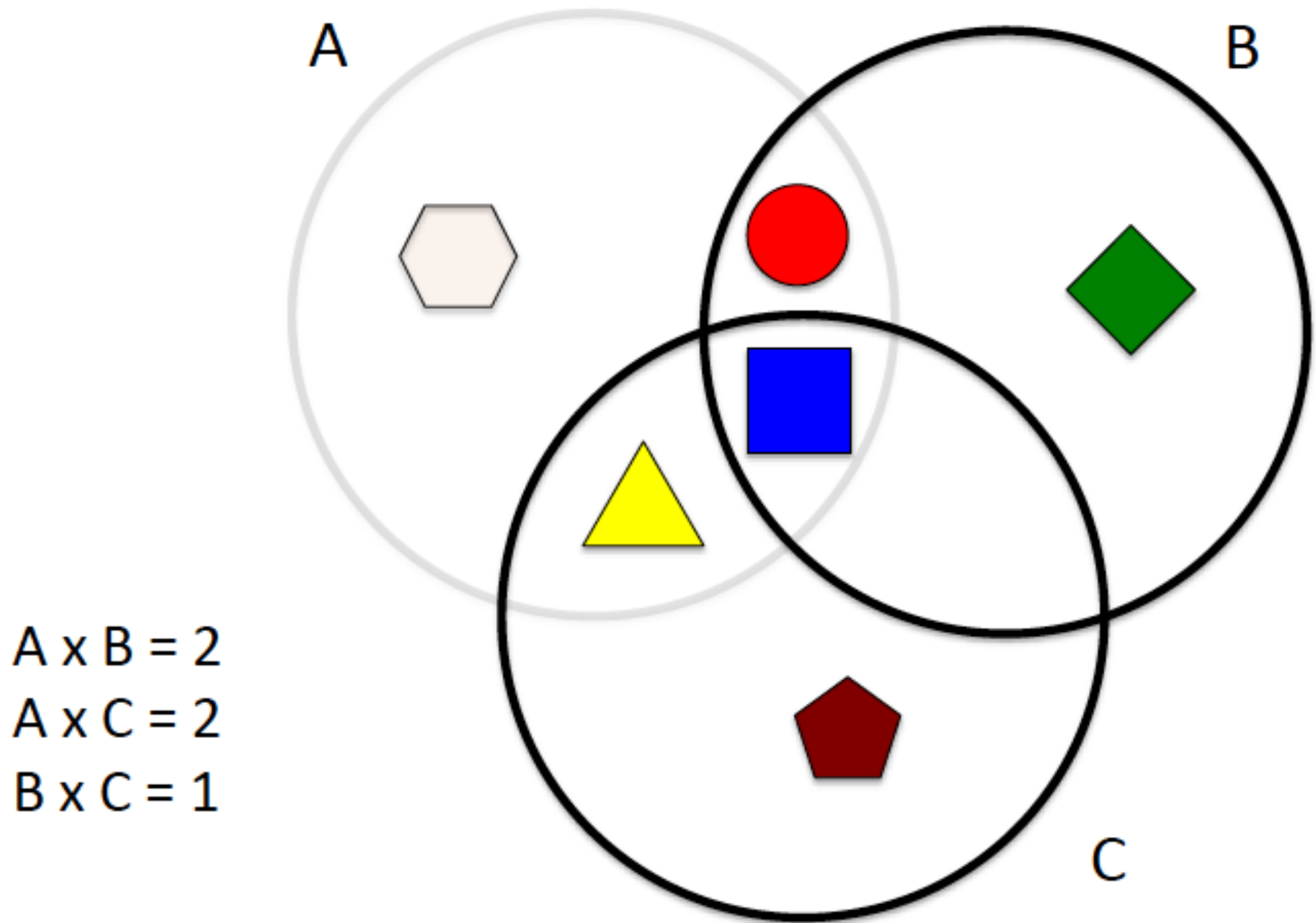
Ref: Daniel McDonald; American Gut Project

# Beta diversity



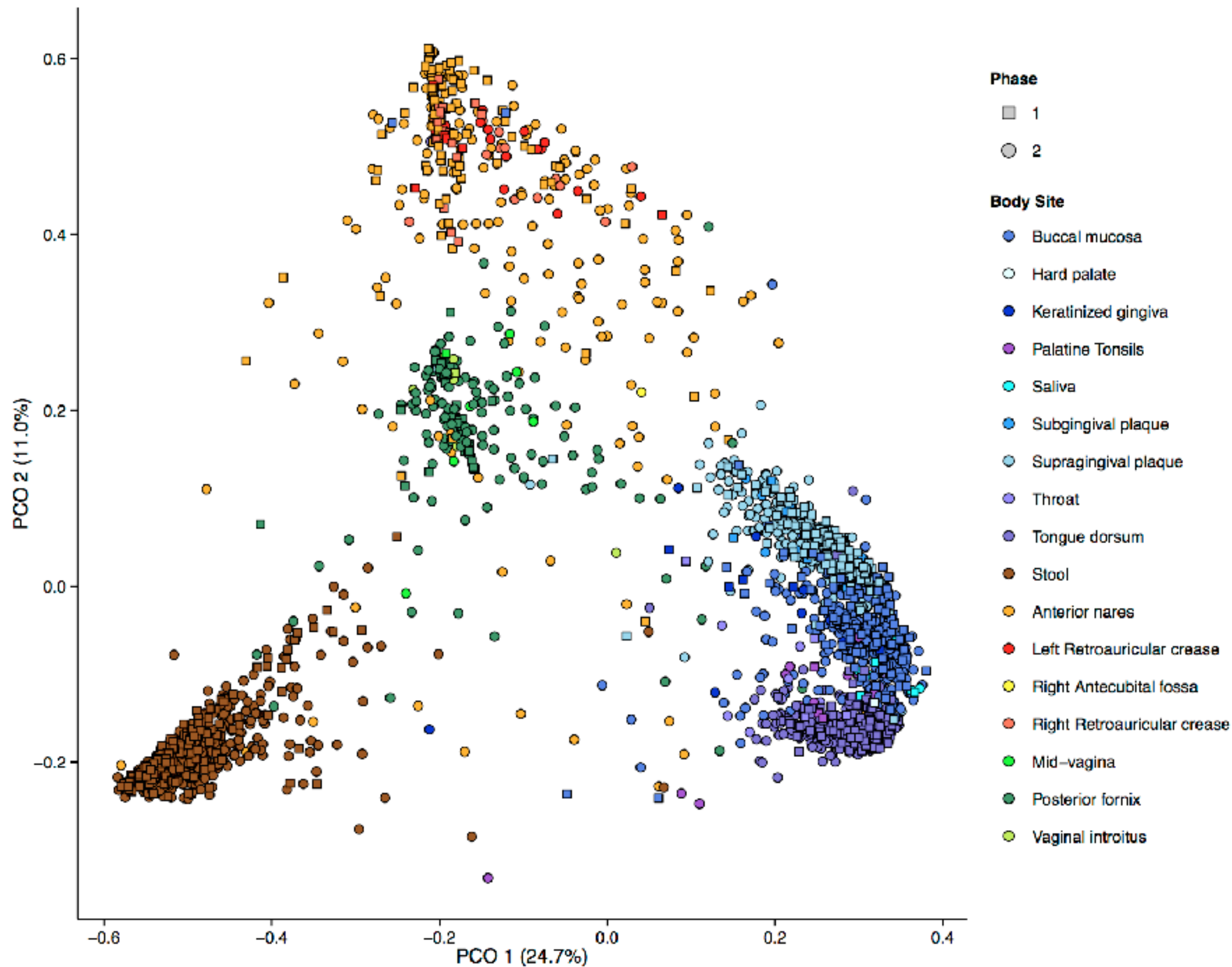
Ref: Daniel McDonald; American Gut Project

# Beta diversity

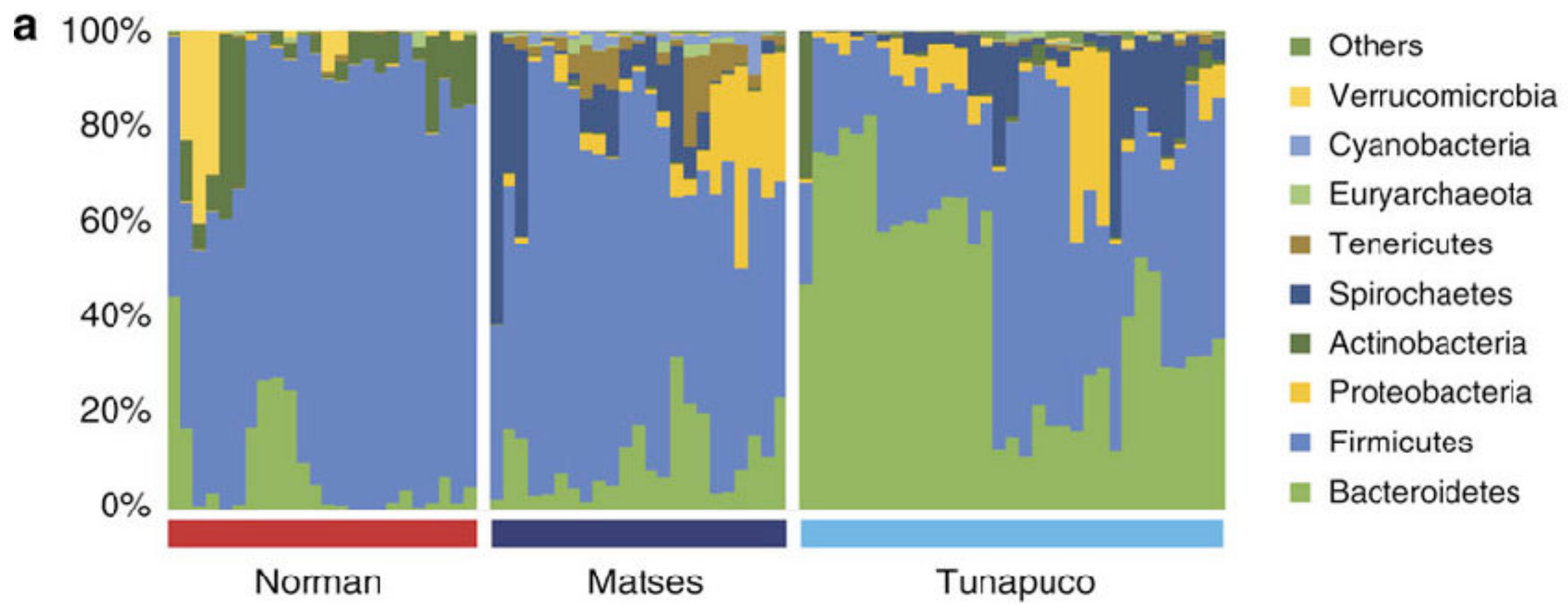


Ref: Daniel McDonald; American Gut Project





# Species Analyses



A. Obregon-Tito, et al. Nat Comms, 2015

**Questions?**