Evolution
&
the Microbiome

8th Annual Workshop on Genomics
Český Krumlov, Czech Republic
16 January 2018

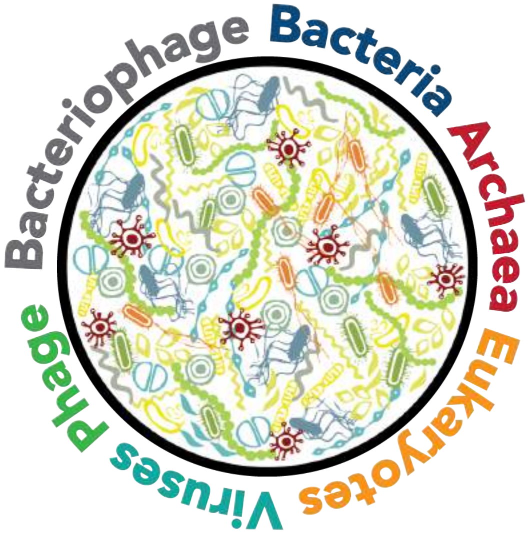
Sarah M. Hird

@sarahmhird
hirdlab.com
sarah.hird@uconn.edu

UCONN
Molecular and Cell Biology

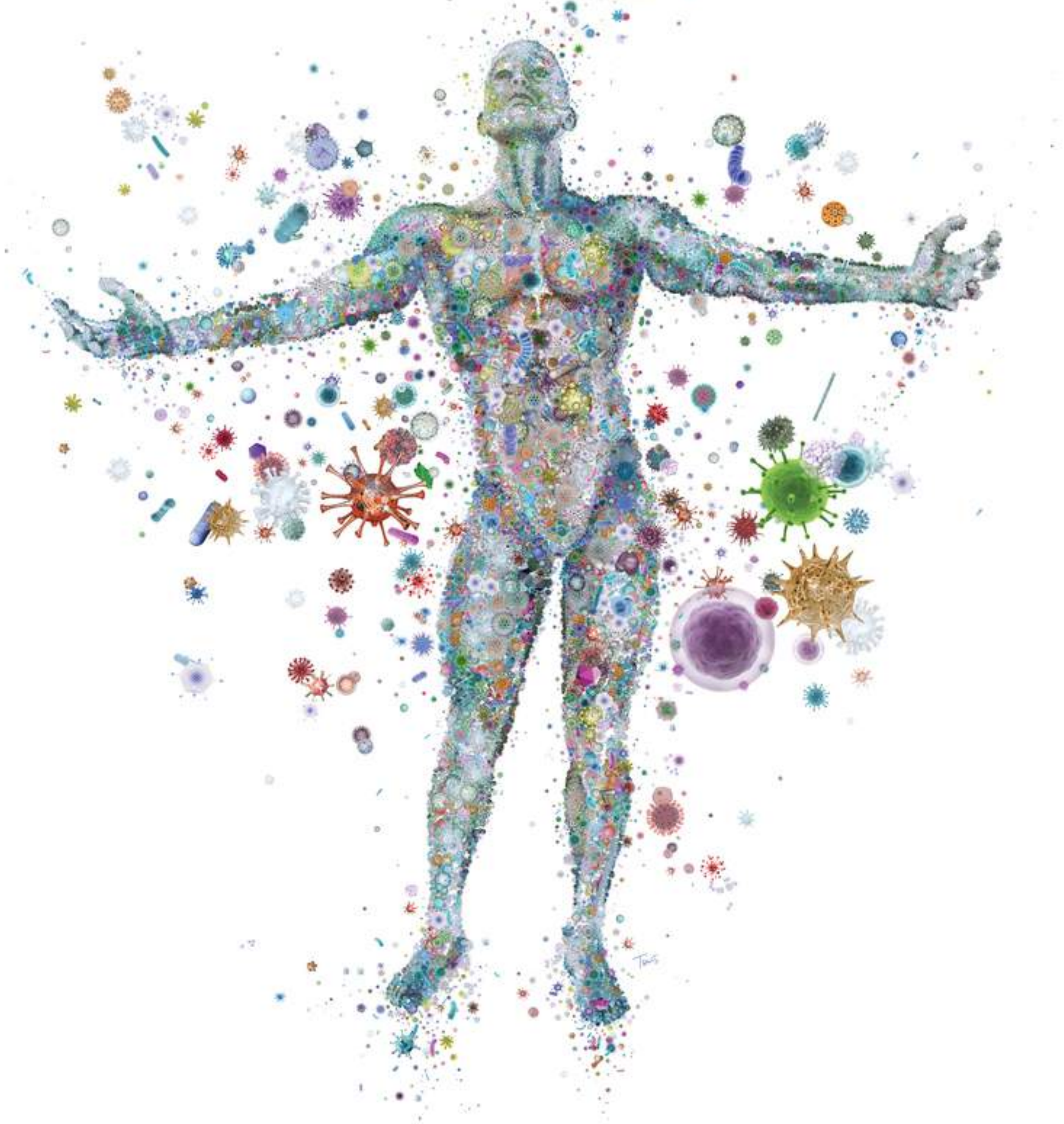
What is the microbiome?

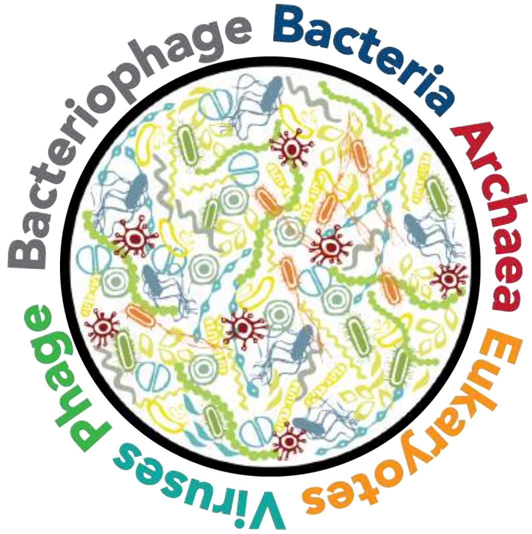
Microbiome (*n*): A characteristic microbial community, found in a particular environment.



What is the microbiome?

Microbiome (*n*): A characteristic microbial community, found in a particular environment.





Extremely detailed talk outline

1. Introduction
2. Methods
3. Research

Microbes are everywhere

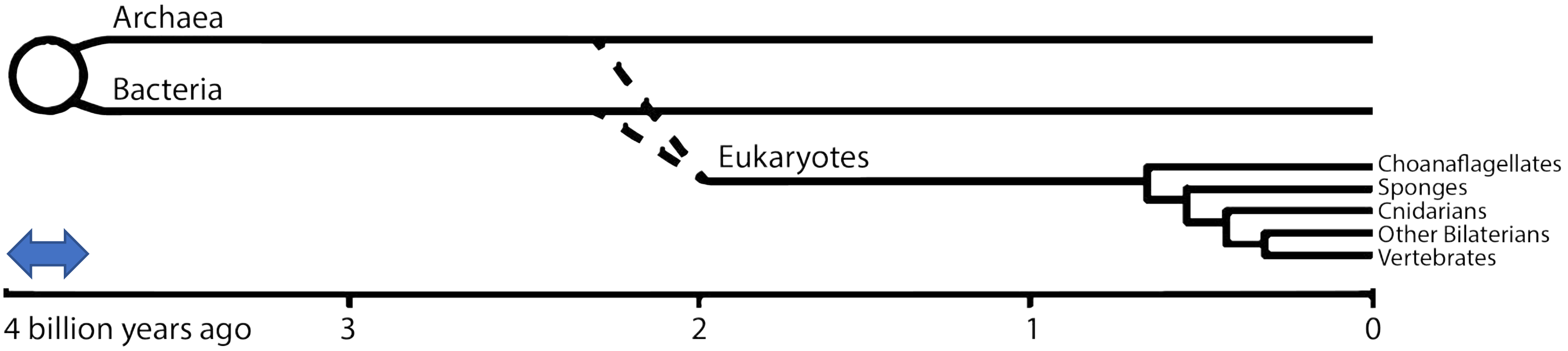
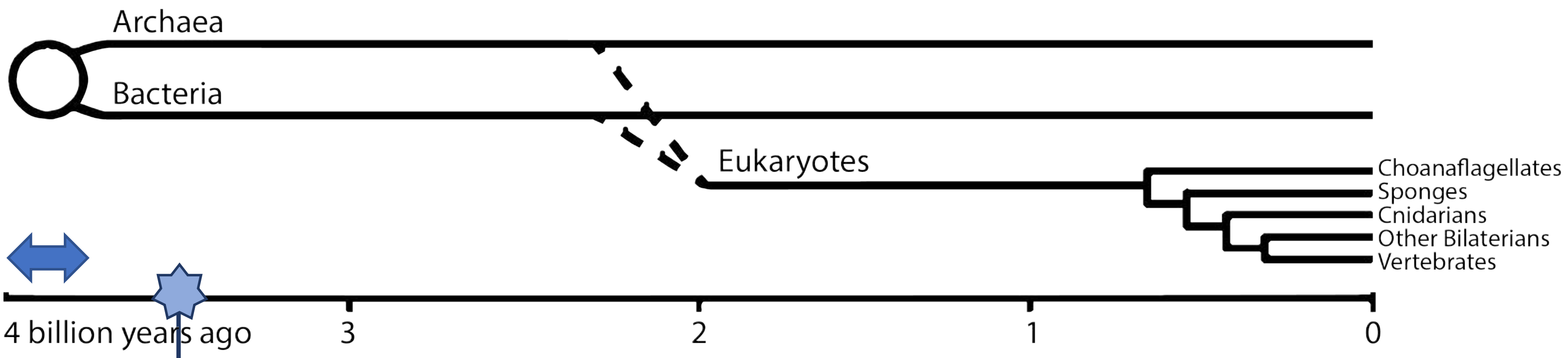
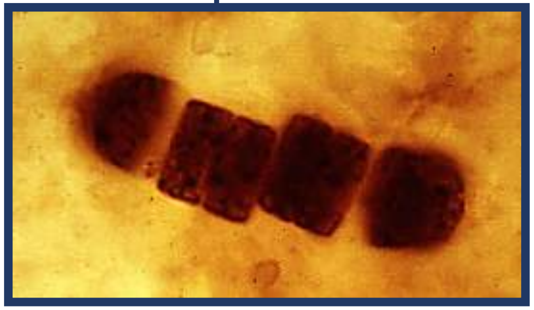


Figure modified from McFall-Ngai et al. 2013. *PNAS*.

@sarahmhird

Microbes are everywhere

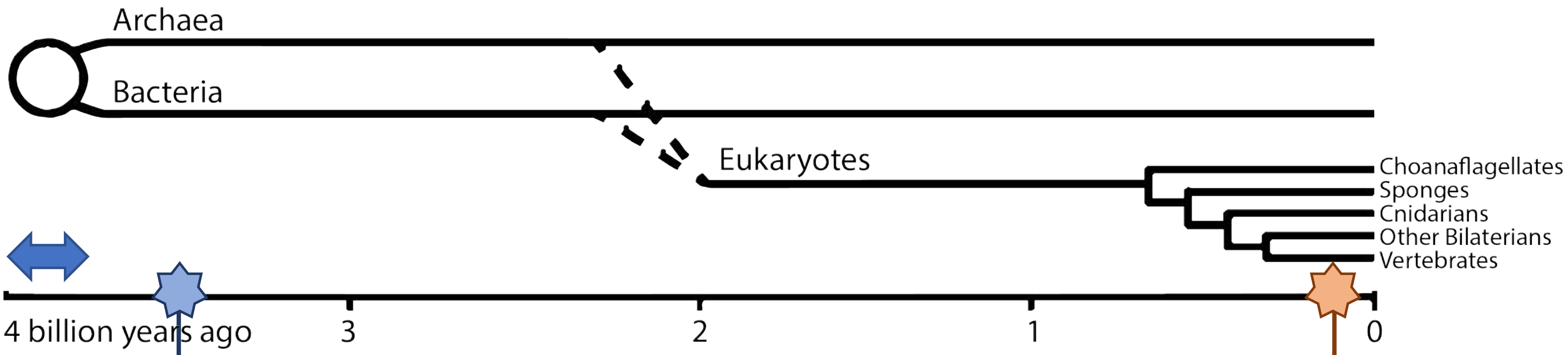




Cyanobacteria ~3.5 BYA

Figure modified from McFall-Ngai et al. 2013. *PNAS*.

Animals evolved in a microbial world.




Cyanobacteria ~3.5 BYA

Archaeopteryx ~150MYA

Figure modified from McFall-Ngai et al. 2013. *PNAS.*

@sarahmhird

Microbes are everywhere

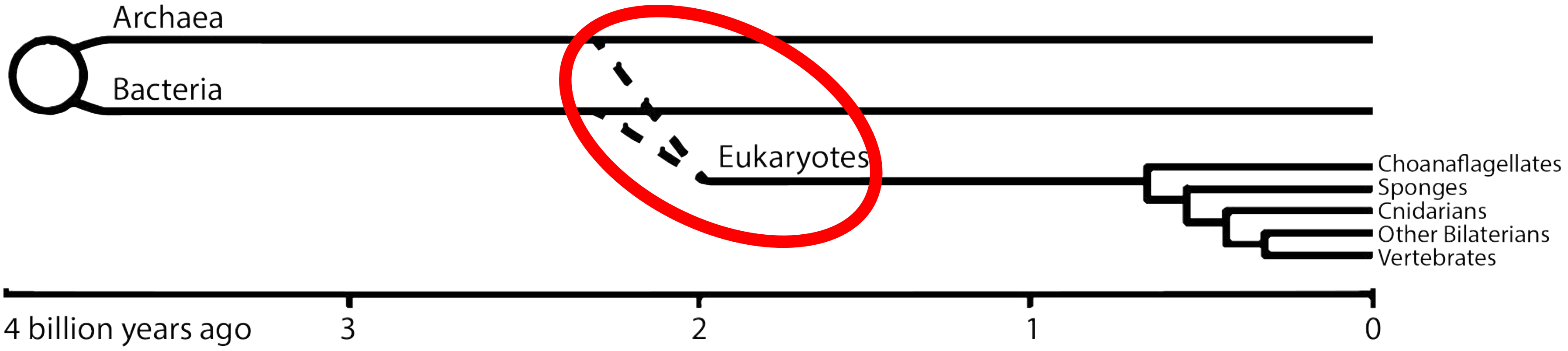
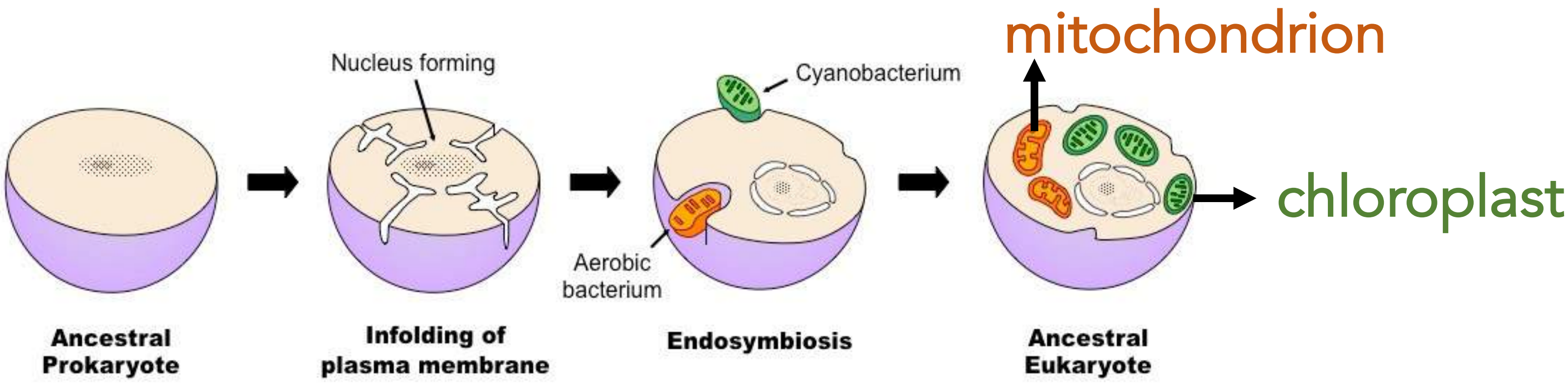


Figure modified from McFall-Ngai et al. 2013. *PNAS*.

@sarahmhird

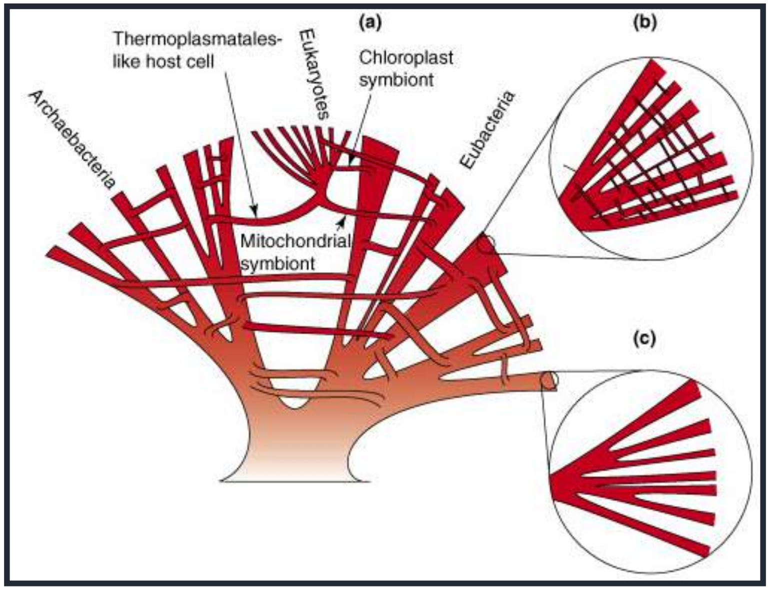
Endosymbiont Theory (Margulis 1981)



(NOTE: Not simultaneous.)

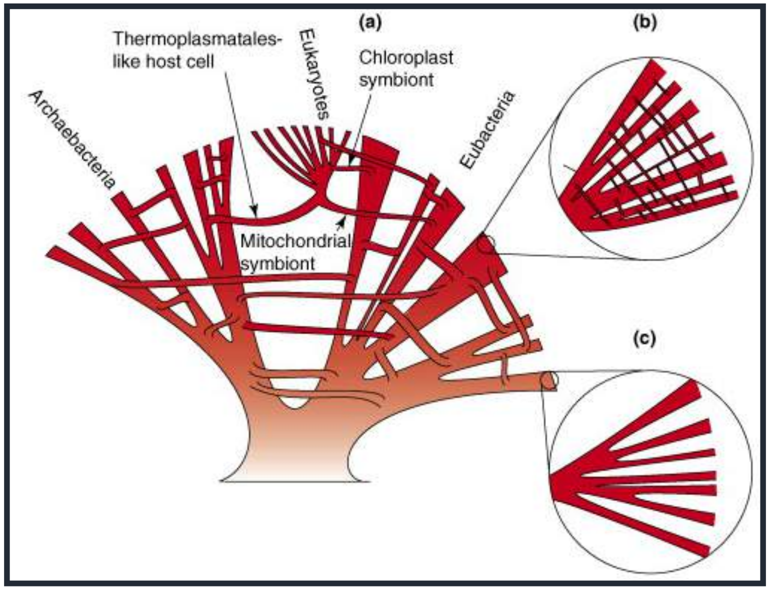
ib.bioninja.com.au

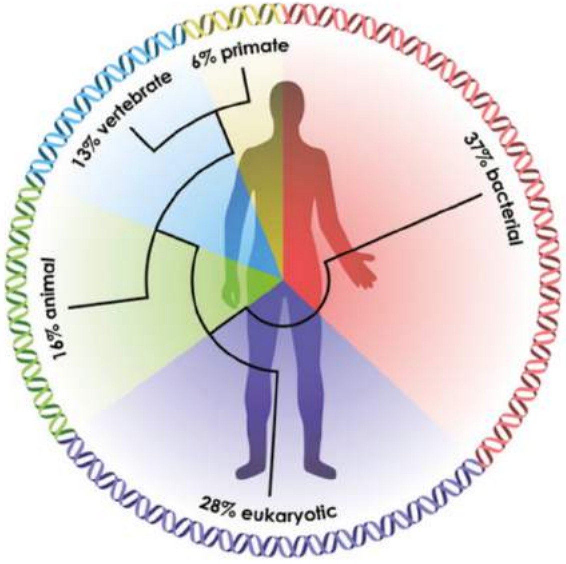
Microbial genes are everywhere.



McInerney et al. 2008. *Trends in Ecology and Evolution.*

Microbial genes are everywhere.

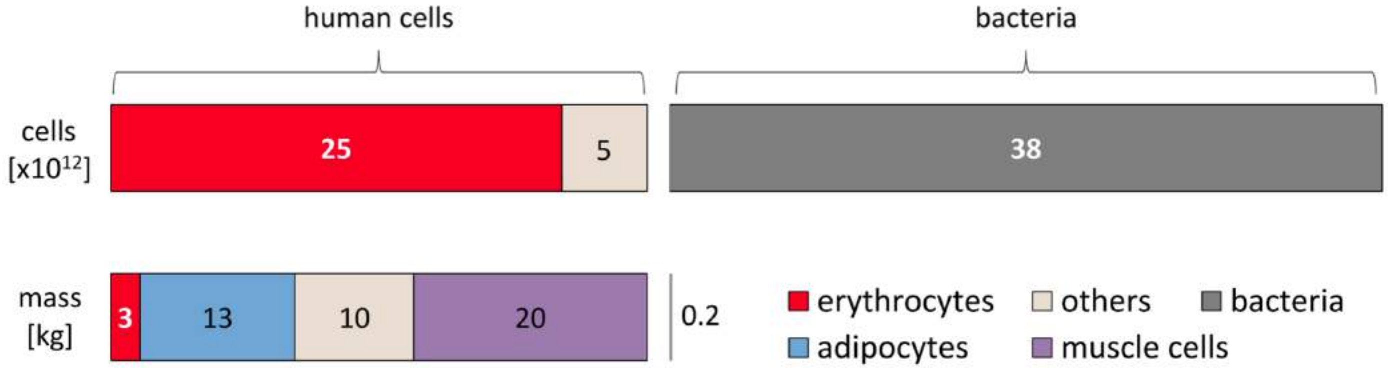




McFall-Ngai et al. 2013. *PNAS.*
McInerney et al. 2008. *Trends in Ecology and Evolution.*

We are holobionts.

"assemblages of different species that form ecological units"



Sender et al. 2016. *PLOS Biology.*

@sarahmhird

So what?

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny
2. Understand natural history, lifestyle and traits
3. Elucidate the forces and processes affecting natural history and phylogeny

Understanding the microbiome contributes to all three of these goals!

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny

Hello, Hoatzin.

©BBC

@sarahmhird

Hoatzin crop microbiome convergent to cow rumen microbiome.

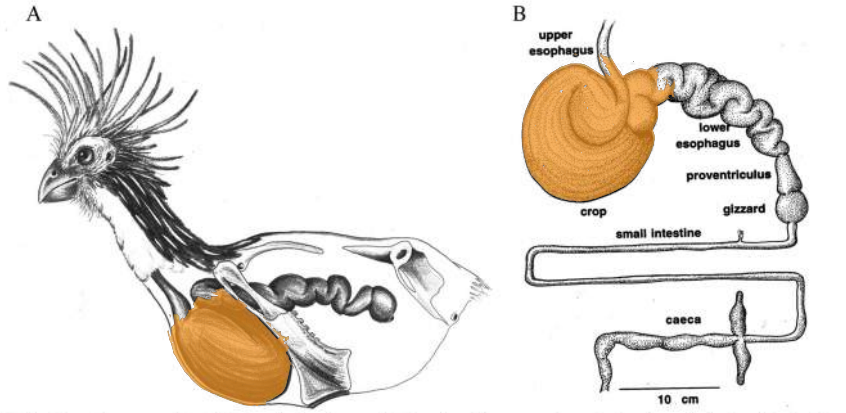


FIG. 1. Schematic representation of the hoatzin digestive tract. (A) Location of the crop and expanded esophagus in the hoatzin body. The anterior sternum is much reduced to make room for the large crop. (Reprinted from *Natural History* [[17] with permission of the publisher.) (B) Extended complete digestive tract of the hoatzin. (Courtesy of Alejandro Grajal.)

Godoy-Vitorino et al. 2008. *Applied & Environmental Microbiology.*

Hoatzin crop microbiome convergent to cow rumen microbiome.

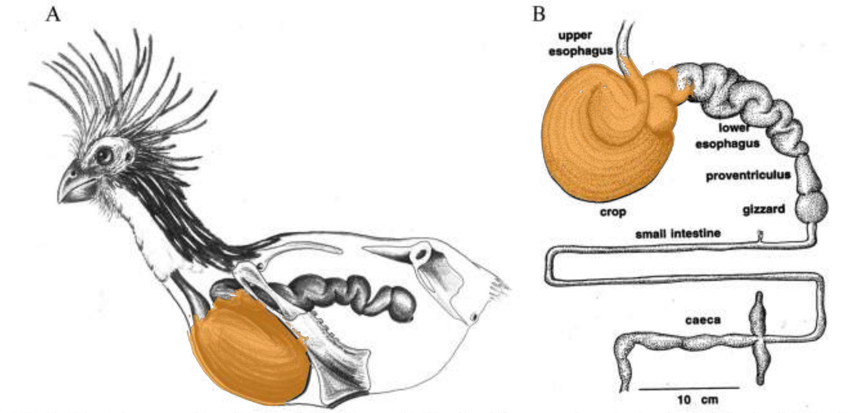
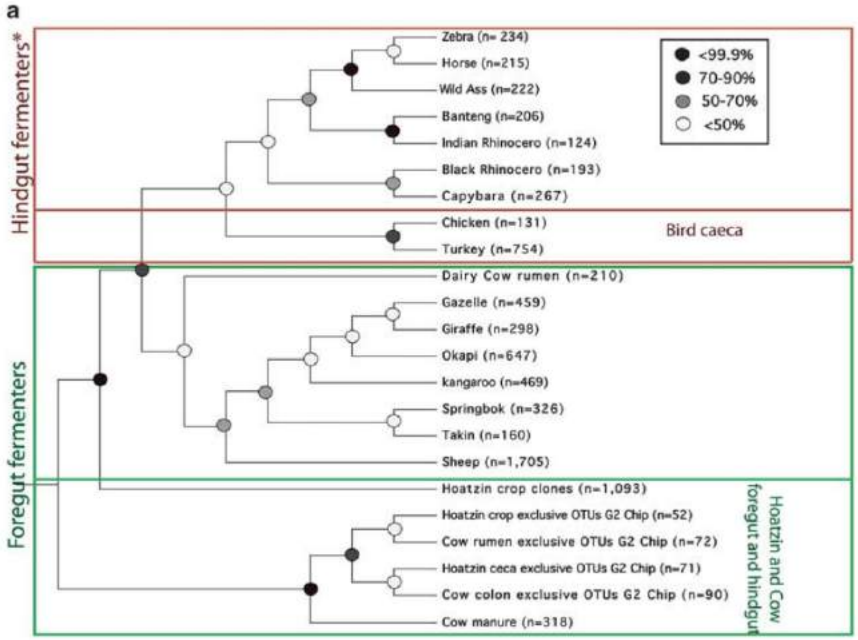


FIG. 1. Schematic representation of the hoatzin digestive tract. (A) Location of the crop and expanded esophagus in the hoatzin body. The anterior sternum is much reduced to make room for the large crop. (Reprinted from *Natural History* [[17] with permission of the publisher.) (B) Extended complete digestive tract of the hoatzin. (Courtesy of Alejandro Grajal.)



Godoy-Vitorino et al. 2010. *The ISME Journal.*
Godoy-Vitorino et al. 2008. *Applied & Environmental Microbiology.*

Hoatzin crop microbiome contains novel biodiversity.

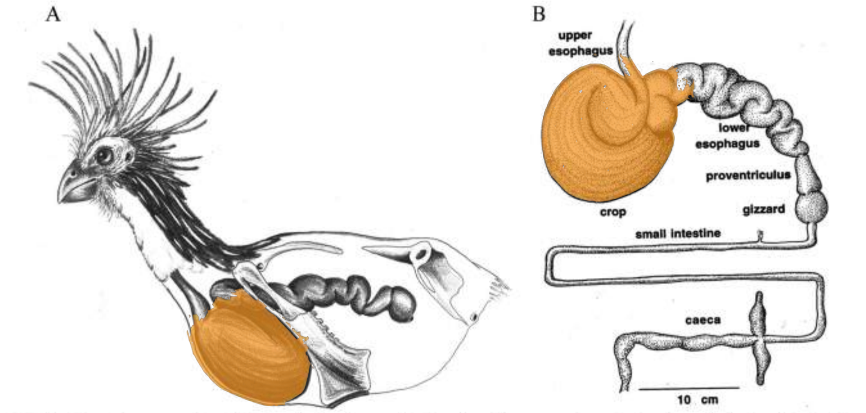
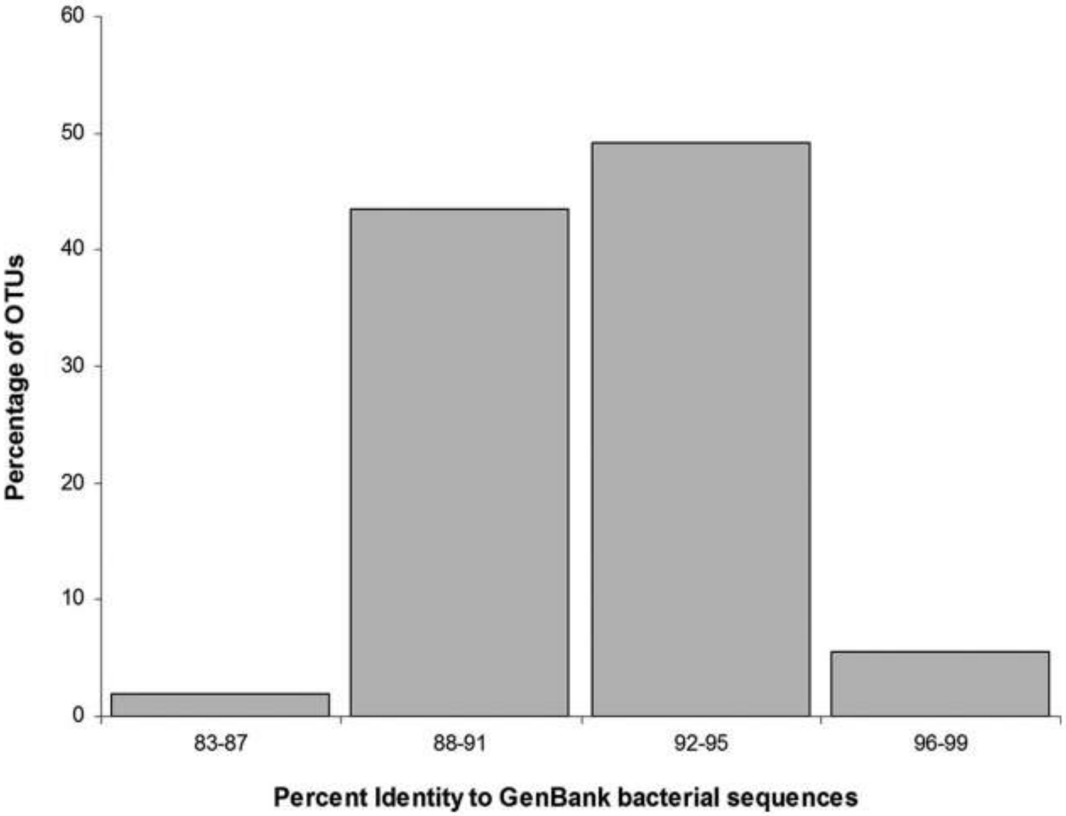


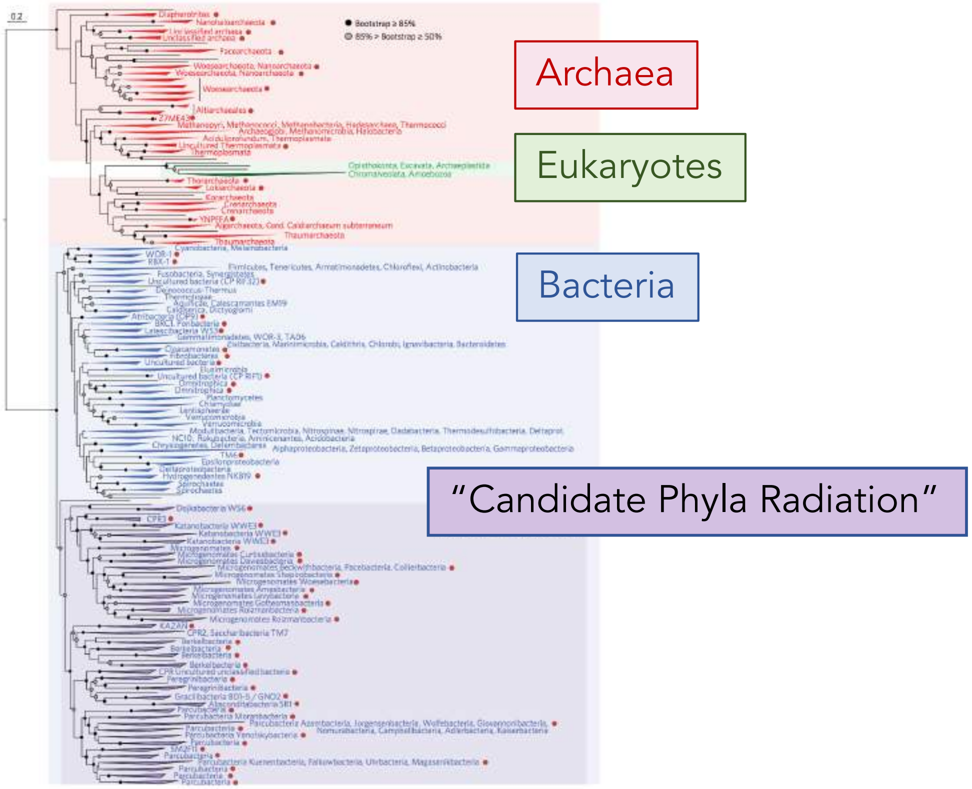
FIG. 1. Schematic representation of the hoatzin digestive tract. (A) Location of the crop and expanded esophagus in the hoatzin body. The anterior sternum is much reduced to make room for the large crop. (Reprinted from *Natural History* [[17] with permission of the publisher.) (B) Extended complete digestive tract of the hoatzin. (Courtesy of Alejandro Grajal.)



Godoy-Vitorino et al. 2008. *Applied & Environmental Microbiology.*

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny



Hug et al. 2016. *Nature Microbiology*. A new view of the tree of life.

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny
2. Understand natural history, lifestyle and traits

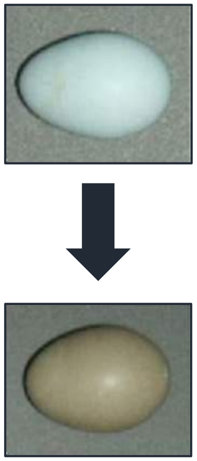
@sarahmhird

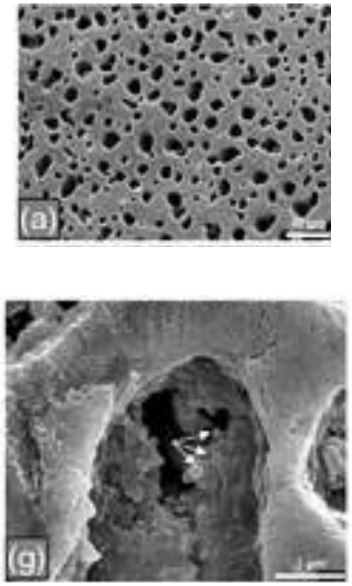
Hello, Hoopoe.

Martin-Vivaldi et al. 2014. *Journal of Animal Ecology.*

fromthegrapevine.com/nature/10-things-you-didnt-know-about-hoopoe

Hello, Hoopoe.





Soler et al. 2014. *Naturwissenschaften.*

Martin-Vivaldi et al. 2014. *Journal of Animal Ecology.*

fromthegrapevine.com/nature/10-things-you-didnt-know-about-hoopoe

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny
2. Understand natural history, lifestyle and traits
3. Elucidate the forces and processes affecting natural history and phylogeny

Evolutionary biology has 3 main goals:

1. Discover and describe biodiversity, including phylogeny
2. Understand natural history, lifestyle and traits
3. Elucidate the forces and processes affecting natural history and phylogeny

COMING SOON!

Evolutionary biology needs wild microbiomes!

Hird. 2017.

frontiers in Microbiology

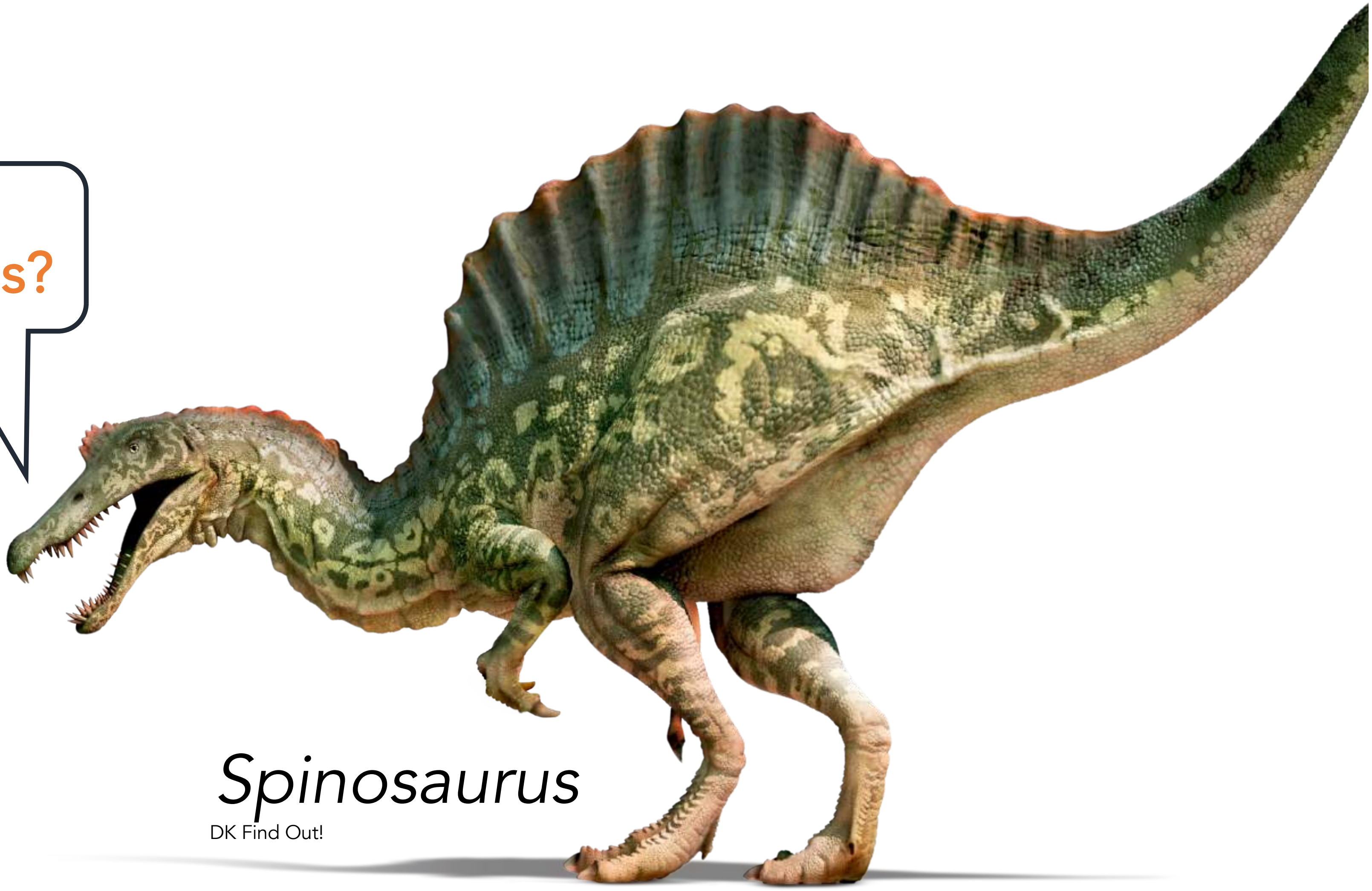
Extremely detailed talk outline

1
Introduction

2
Methods

3
Research

Any questions?



Spinosaurus

DK Find Out!

@sarahmhird

Extremely detailed talk outline

1

Introduction

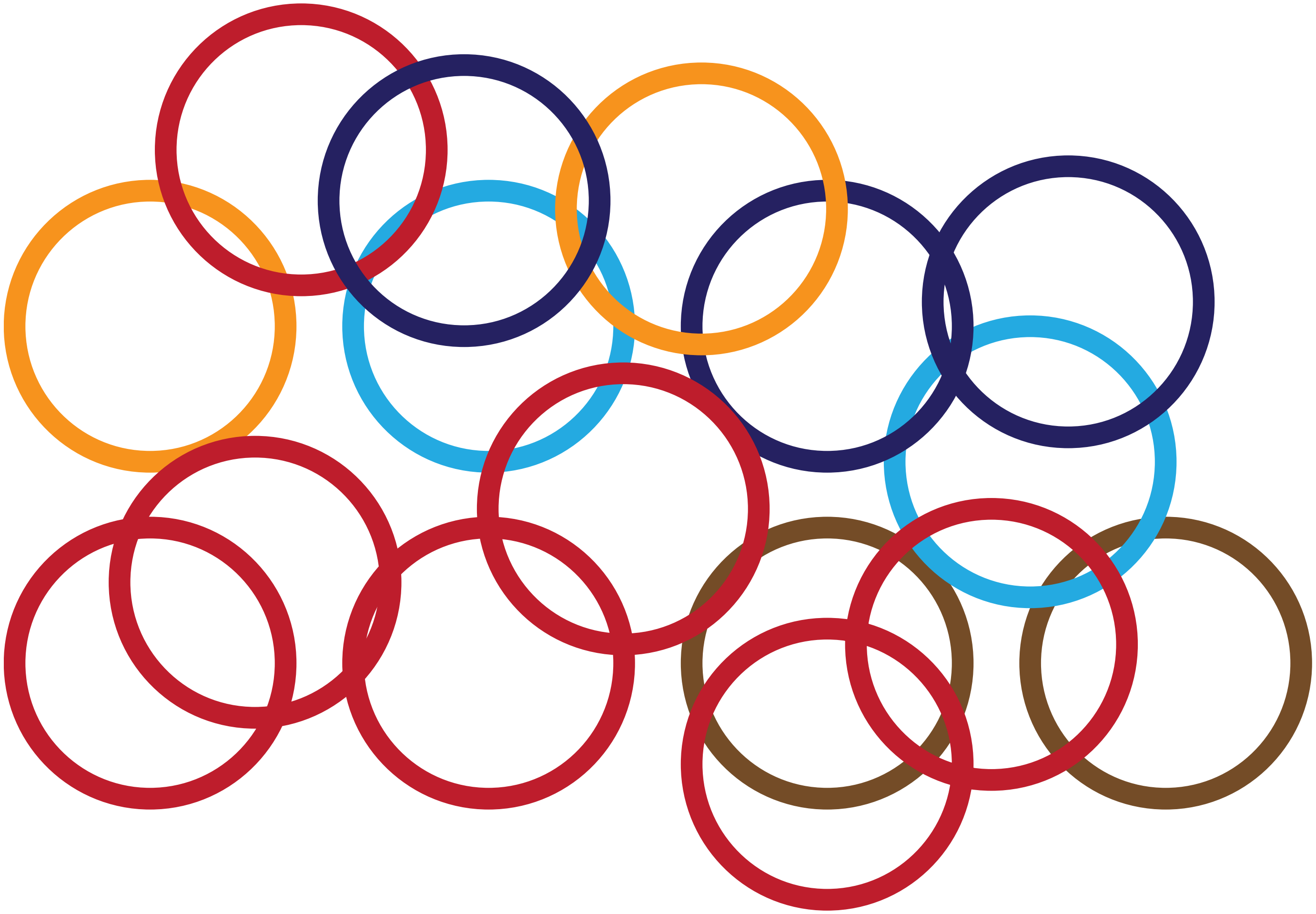
2

Methods

3

Research

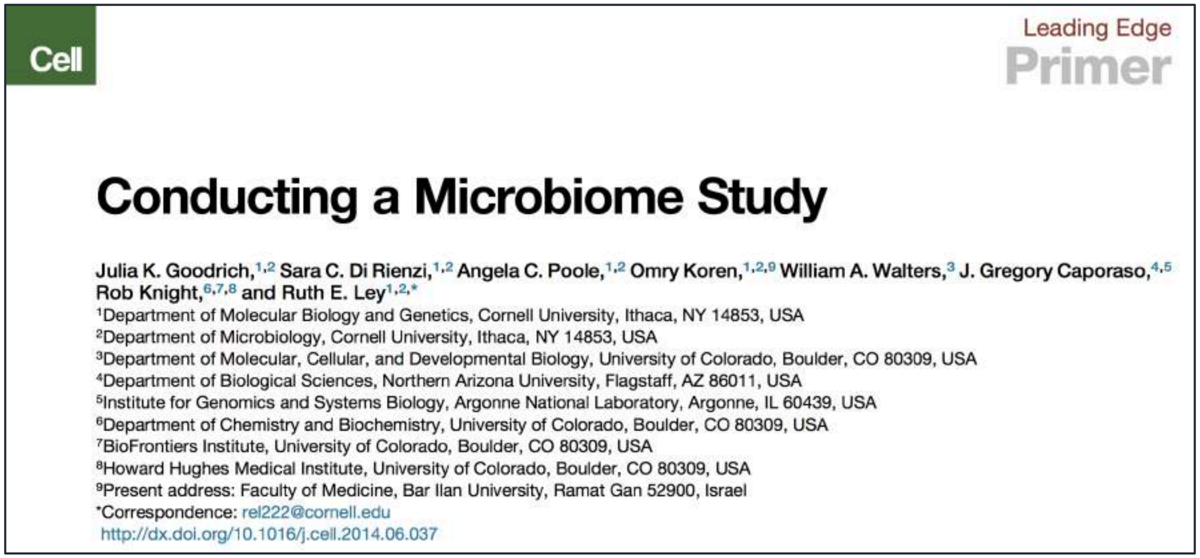
A microbial community



@sarahmhird

Microbiome Methods

1. Develop a question
2. Design the study
3. Conduct the study
4. Analyze the results ★
5. Interpret the results
6. Publish; share data



Cell

Leading Edge
Primer

Conducting a Microbiome Study

Julia K. Goodrich,[1,2] Sara C. Di Rienzi,[1,2] Angela C. Poole,[1,2] Omry Koren,[1,2,9] William A. Walters,[3] J. Gregory Caporaso,[4,5] Rob Knight,[6,7,8] and Ruth E. Ley[1,2,*]
[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[4]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA
[5]Institute for Genomics and Systems Biology, Argonne National Laboratory, Argonne, IL 60439, USA
[6]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[7]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[8]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
[9]Present address: Faculty of Medicine, Bar Ilan University, Ramat Gan 52900, Israel
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.cell.2014.06.037

1. Develop a question

Do obese and lean pet dogs have different microbiomes?

2. Design the study

Do obese and lean pet dogs have different microbiomes?

How many dogs (sample size)?

2. Design the study

Do obese and lean pet dogs have different microbiomes?

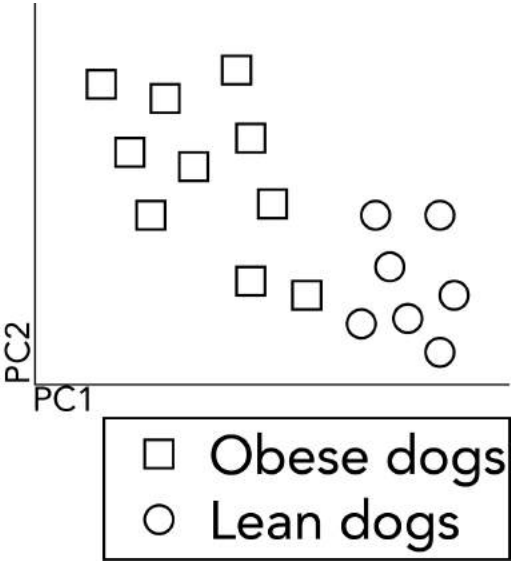
How many dogs (sample size)?

What are they eating?
How much exercise are they getting?
Where do they live?
How much time did they spend with their birth mother?
Do the dogs have any medical conditions?

How many dogs (sample size)? Re-evaluate

2. Design the study

Do obese and lean pet dogs have different microbiomes?

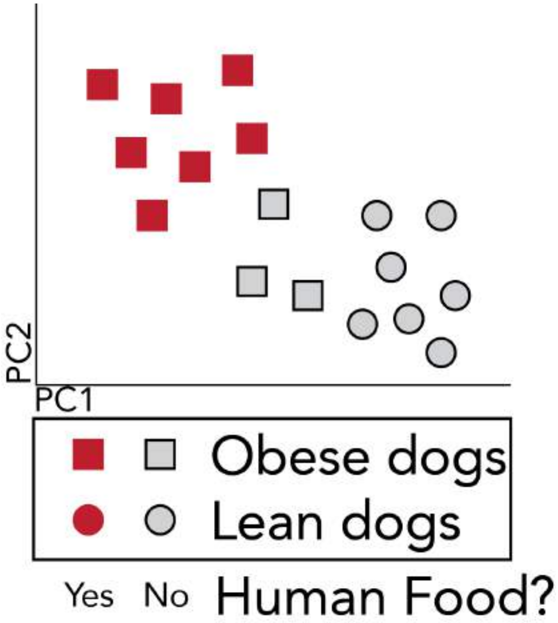


So – how do we interpret this?

Modified from Goodrich et al. 2014. Conducting a microbiome study. *Cell.*
DOI: 10.1016/j.cell.2014.06.037

2. Design the study

Do obese and lean pet dogs have different microbiomes?



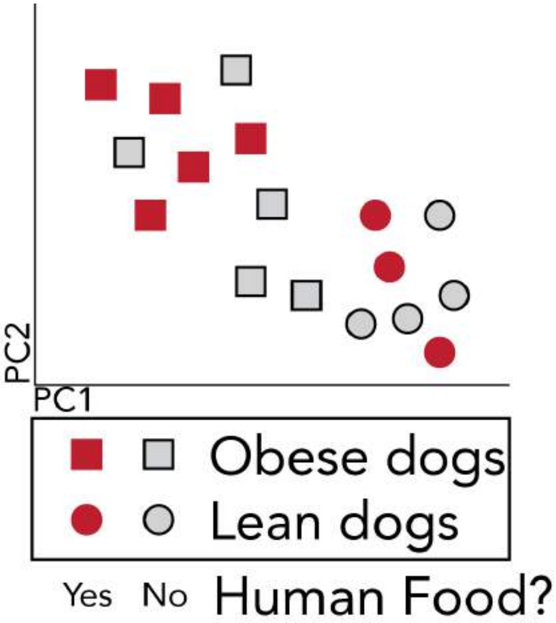
So – how do we interpret this?

Modified from Goodrich et al. 2014. Conducting a microbiome study. *Cell.*

DOI: 10.1016/j.cell.2014.06.037

2. Design the study

Do obese and lean pet dogs have different microbiomes?



So – how do we interpret this?

Modified from Goodrich et al. 2014. Conducting a microbiome study. *Cell.*

DOI: 10.1016/j.cell.2014.06.037

2. Design the study

Do obese and lean pet dogs have different microbiomes?

How many dogs (sample size)?

What are they eating?
How much exercise are they getting?
Where do they live?
How much time did they spend with their birth mother?
Do the dogs have any medical conditions?

How many dogs (sample size)? Re-evaluate

You may not be able to do anything about some of these variables, but you need to RECORD RELEVANT METADATA to understand observed patterns.

2. Design the study

Do obese and lean pet dogs have different microbiomes?

The Bottom Lines For designing a microbiome study:

- Sample design
- Sample size
- Sequences needed
- Confounding variables / associated metadata

3. Conduct the study

Sequence the samples

What do I sequence?

Two main avenues for sequence-based microbiome research (right now):

- **Amplicon-based** studies ("16S")
- **Shotgun metagenomics**

These are very different.

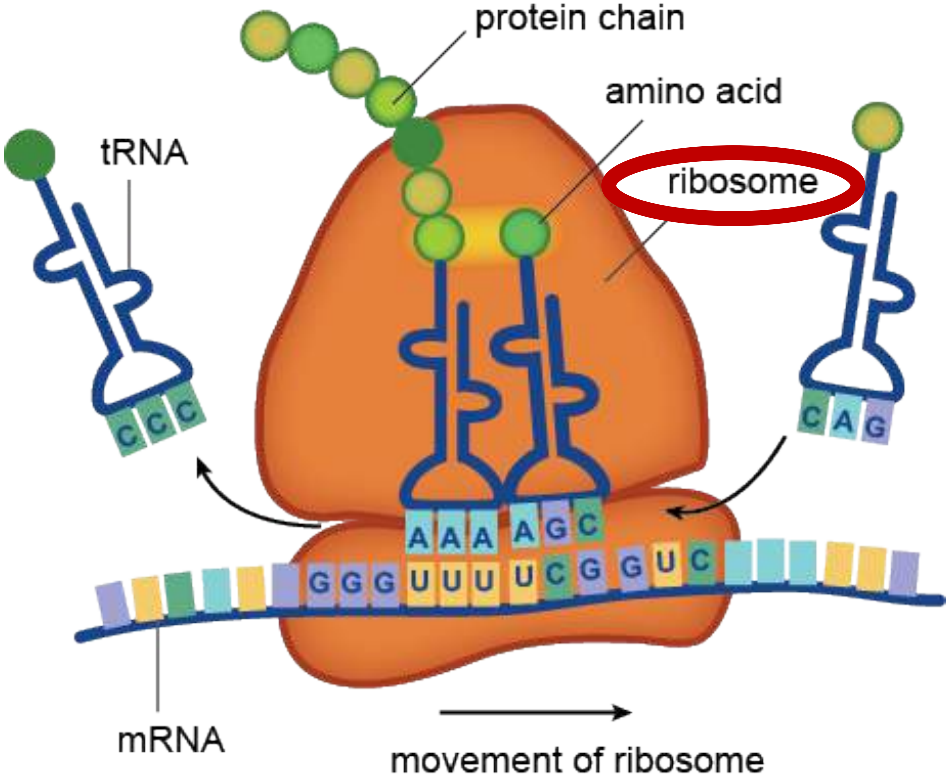
(A microbial community)

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

16S rRNA studies

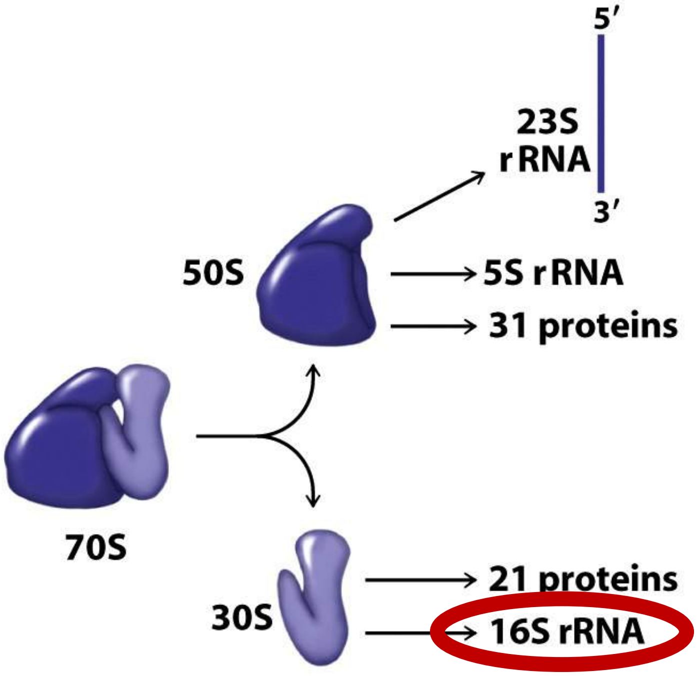
PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.



@sarahmhird

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.



Prokaryote

Figure 22-12 Principles of Biochemistry, 4/e
© 2006 Pearson Prentice Hall, Inc.

(18S for microbial eukaryotes)

@sarahmhird

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

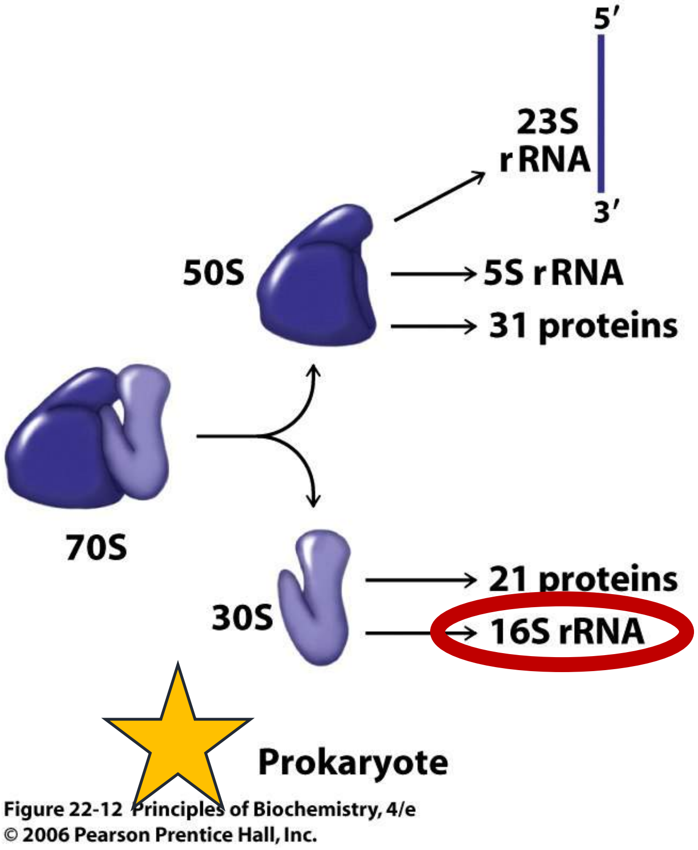


Figure 22-12 Principles of Biochemistry, 4/e
© 2006 Pearson Prentice Hall, Inc.

(18S for microbial eukaryotes)

The power of the ribosome

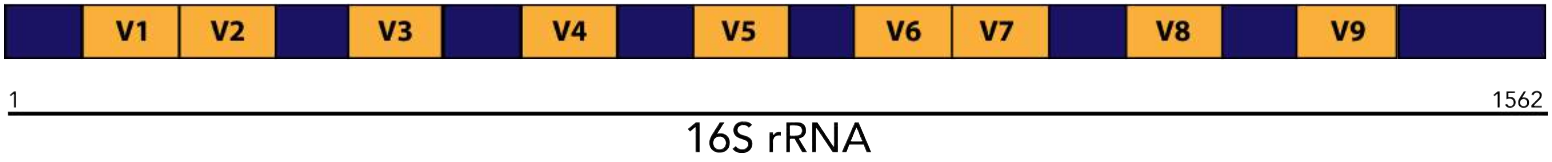


Figure modified from Del Chierico et al. 2015. Methods in Molecular Biology (1231).

@sarahmhird

The power of the ribosome

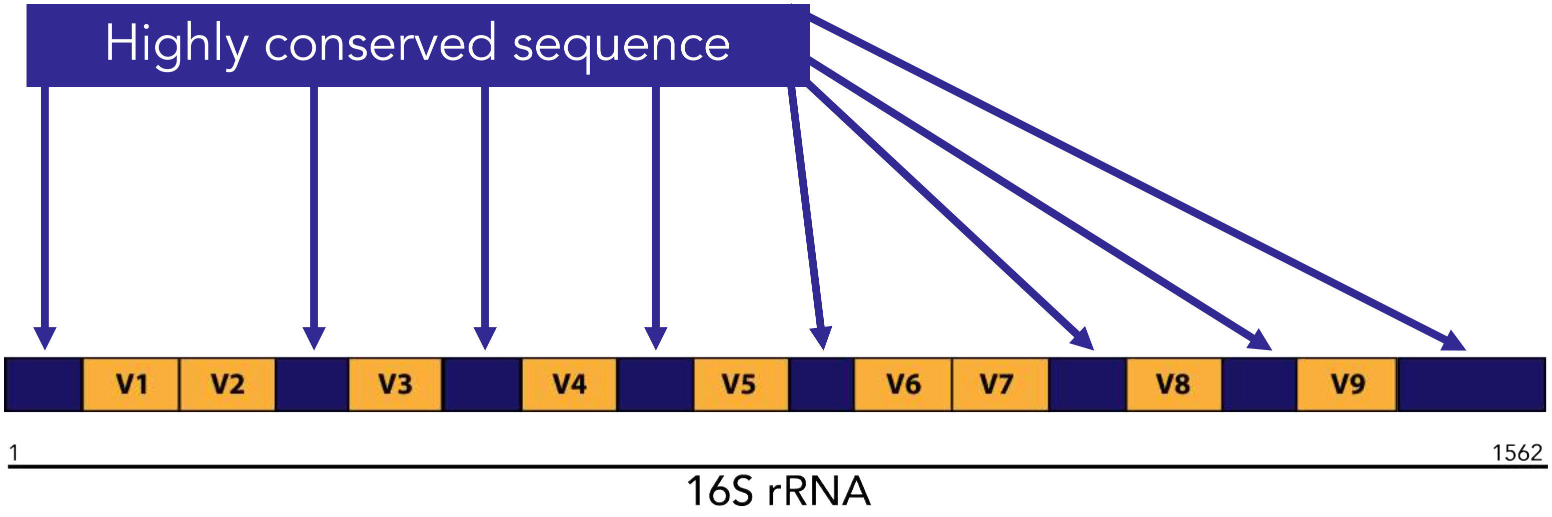
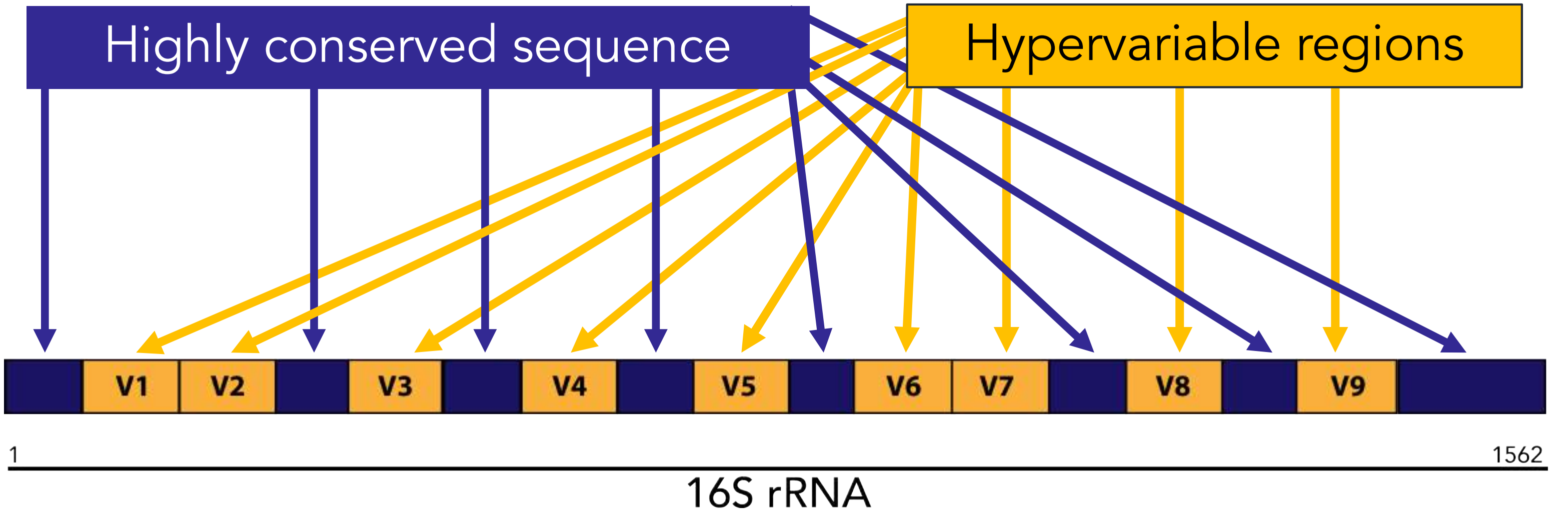


Figure modified from Del Chierico et al. 2015. Methods in Molecular Biology (1231).

The power of the ribosome



Highly conserved sequence

Hypervariable regions

16S rRNA

Figure modified from **Del Chierico** et al. 2015. Methods in Molecular Biology (1231).

@sarahmhird

The power of the ribosome

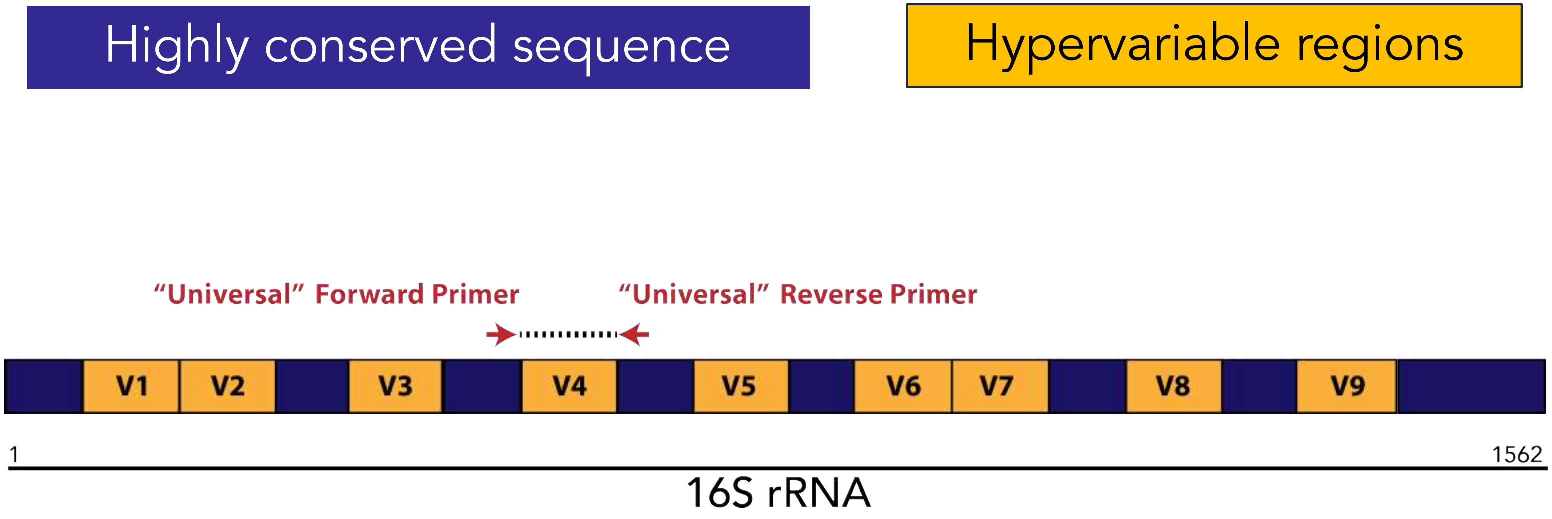
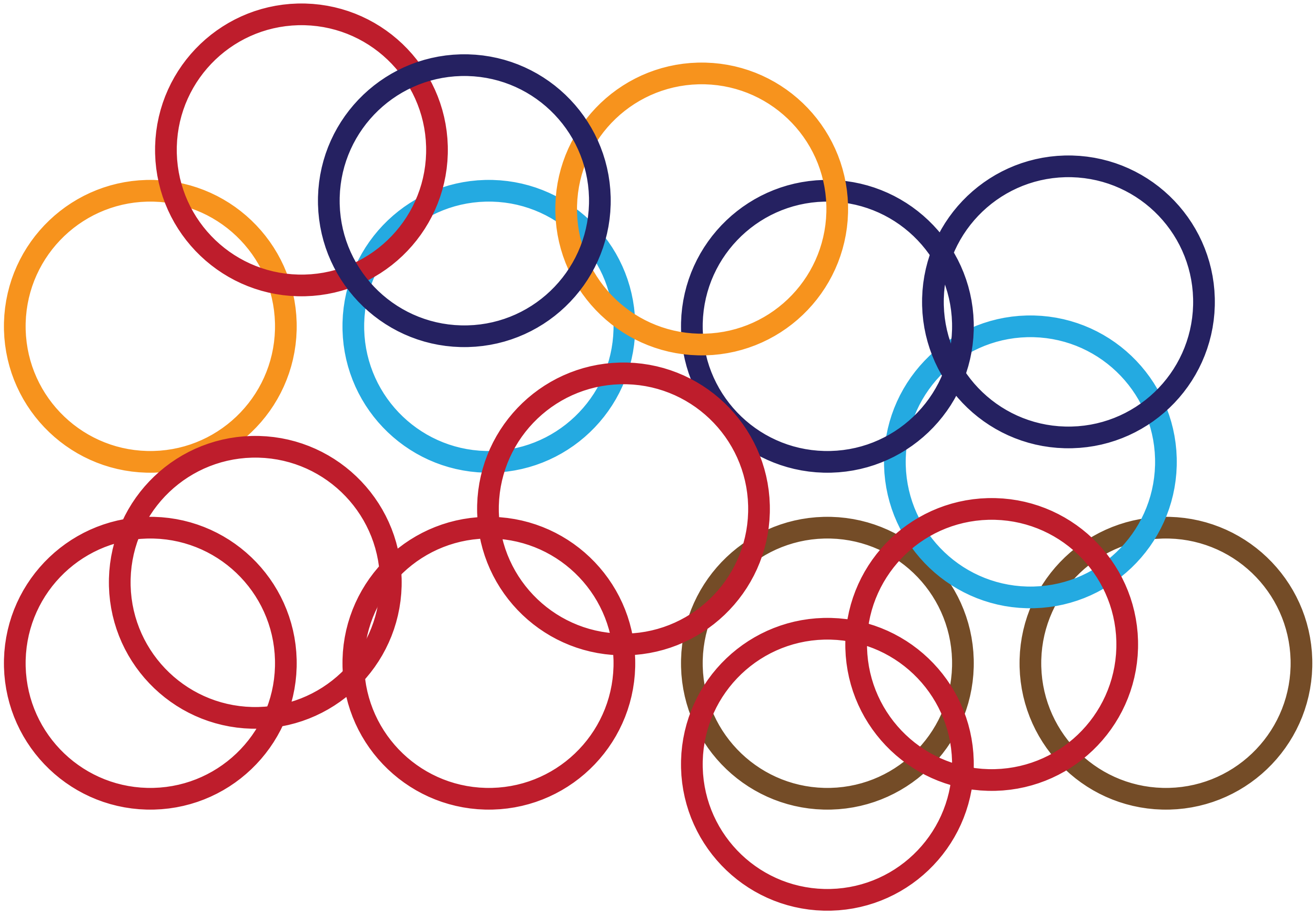
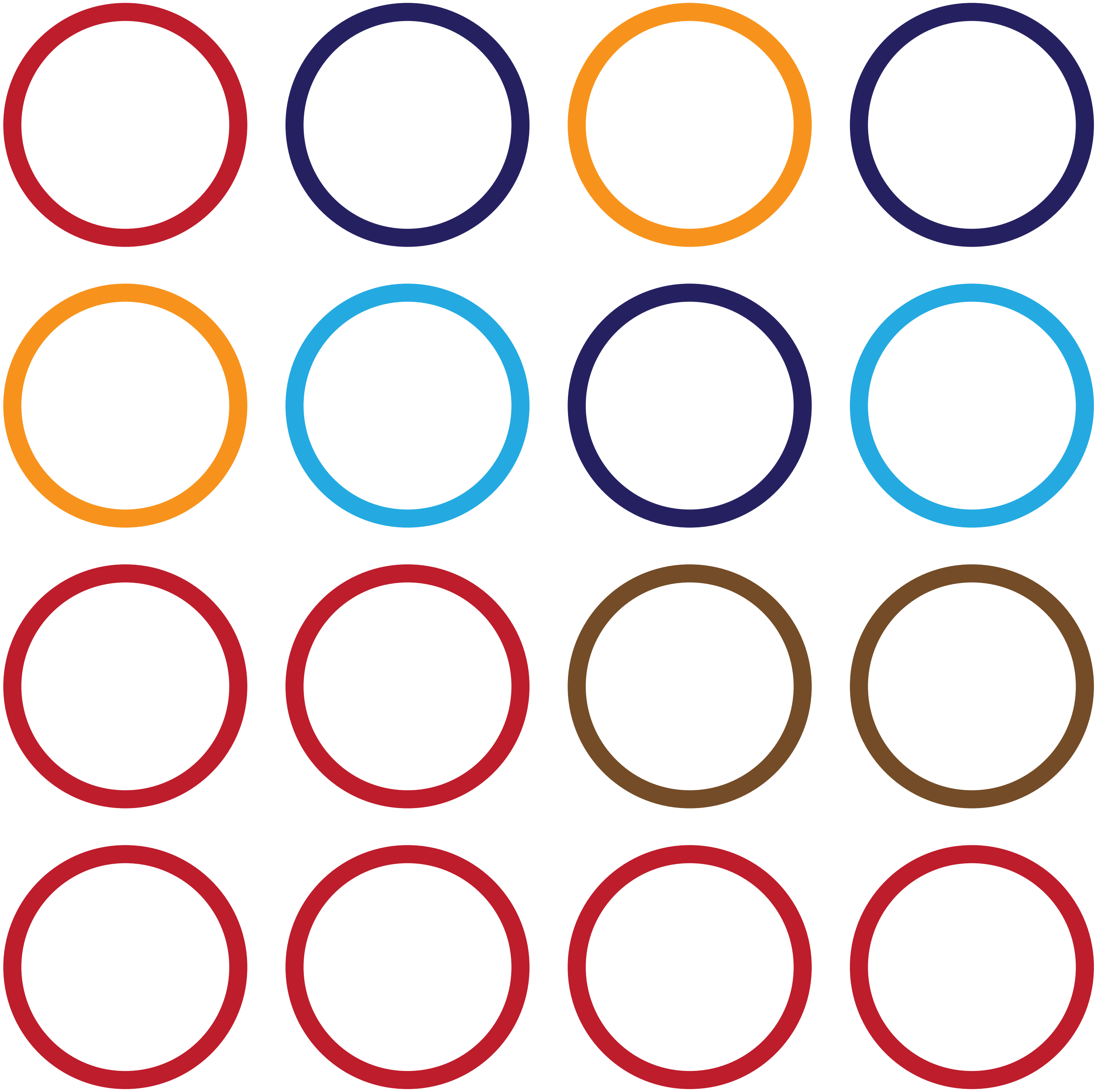


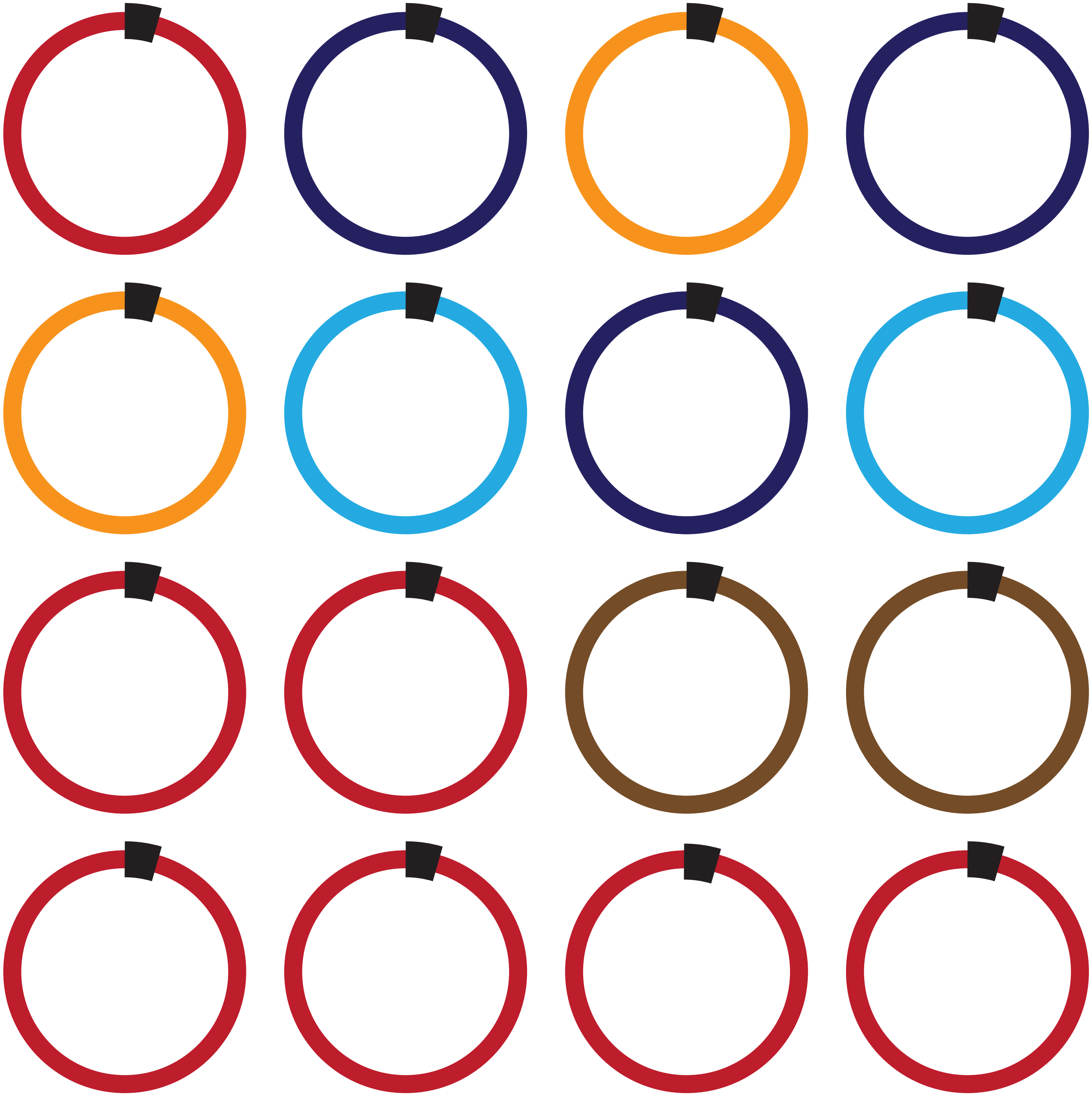
Figure modified from Del Chierico et al. 2015. Methods in Molecular Biology (1231).

A community

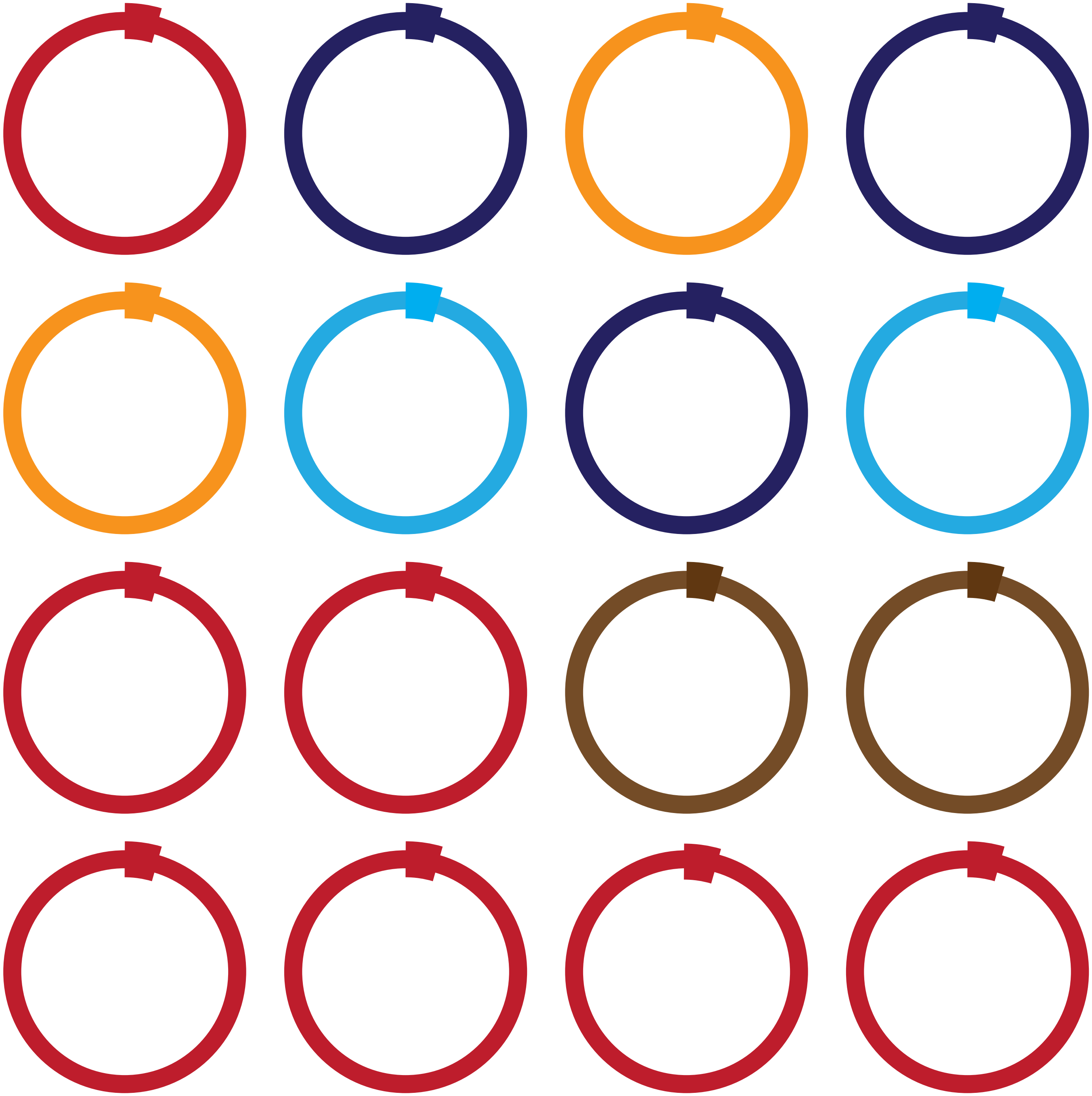




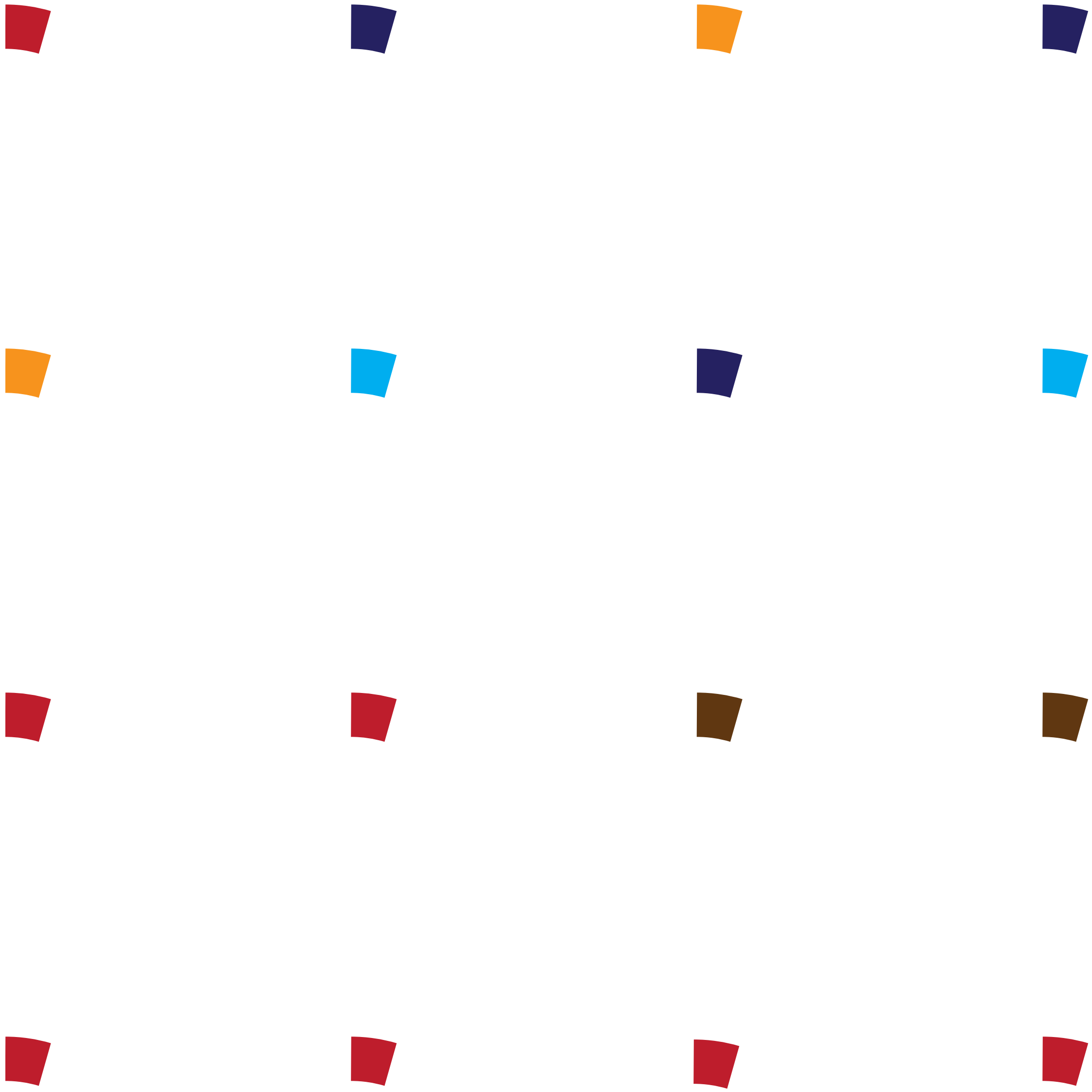
@sarahmhird



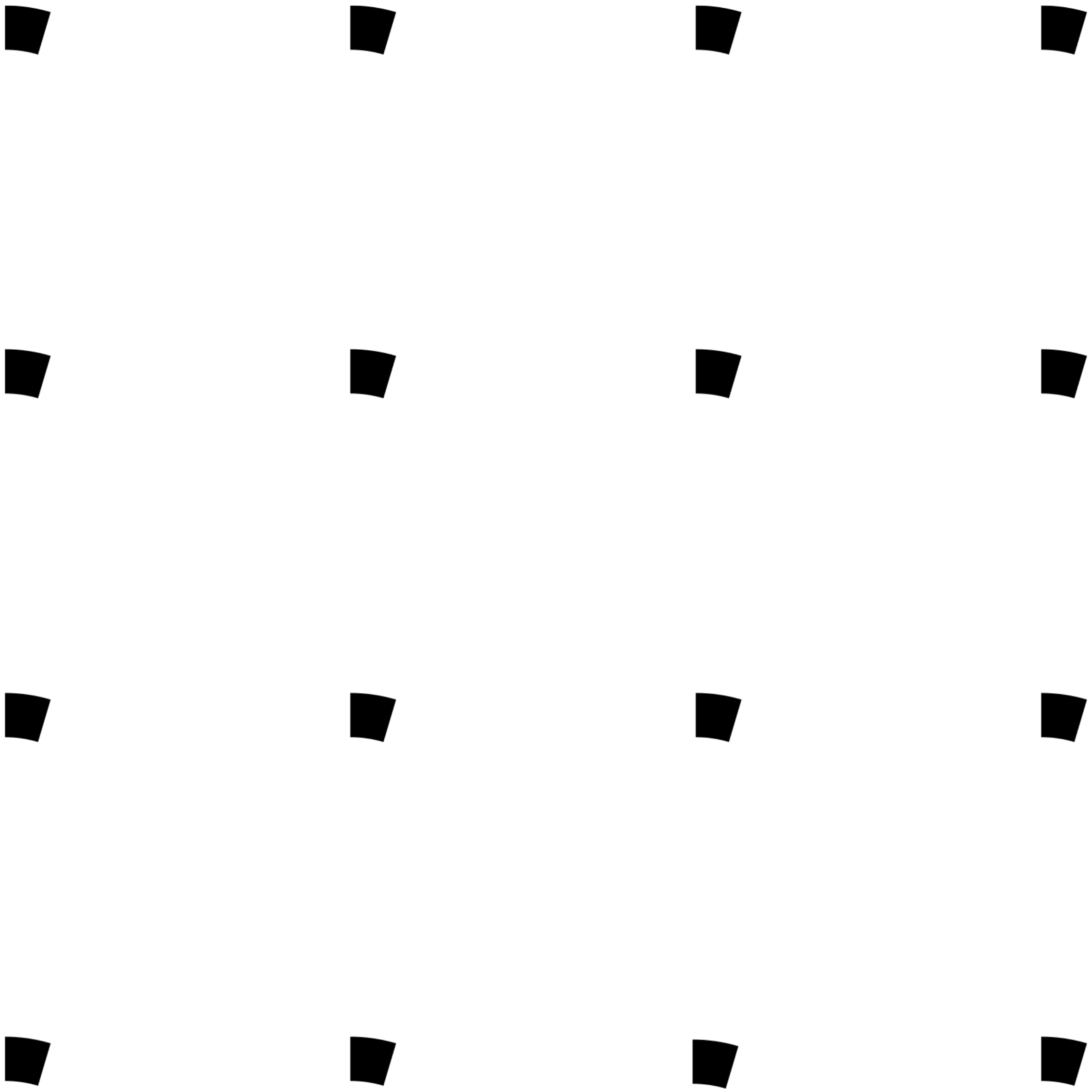
@sarahmhird



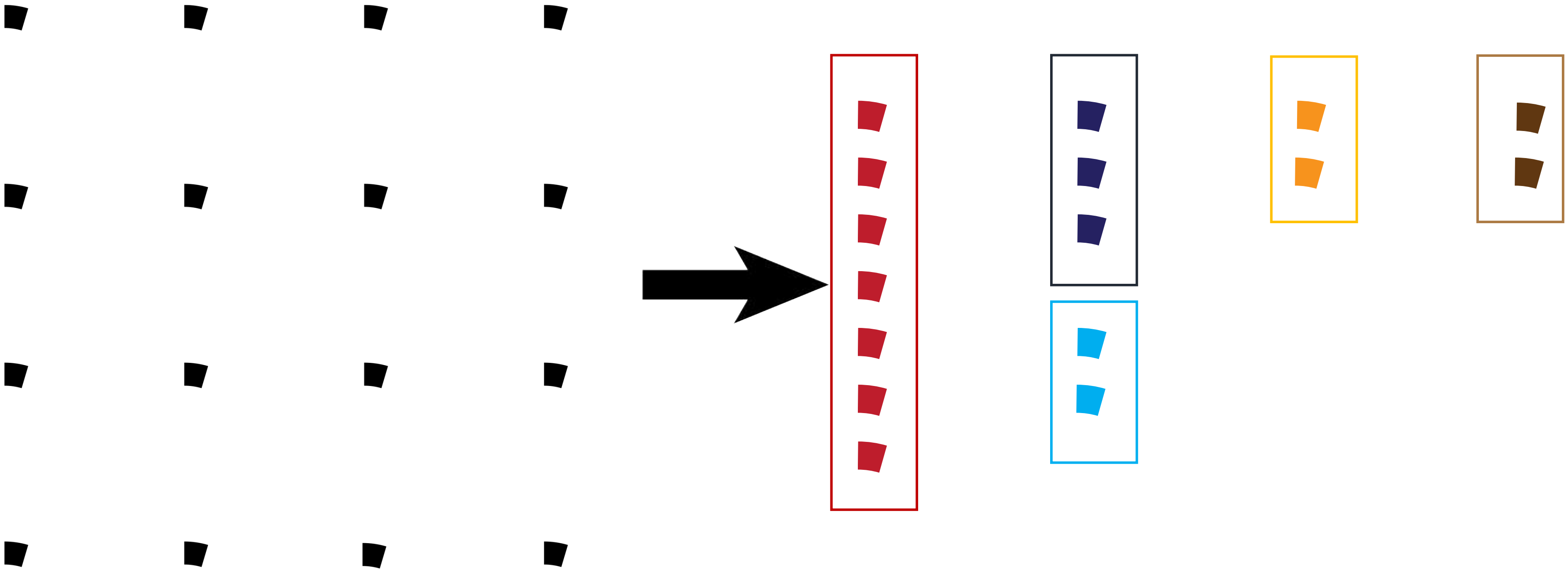
@sarahmhird



@sarahmhird



@sarahmhird



@sarahmhird

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

By far the most popular method for microbiome studies. Because:

- cheap
- well developed molecular and analytical methods
- easy to perform
- "good results"

CONS:

- Limited information
- Database dependent
- PCR
- Single locus

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

By far the most popular method for microbiome studies. Because:

- cheap
- well developed molecular and analytical methods
- easy to perform
- "good results"

CONS:

- Limited information
- Database dependent
- PCR
- Single locus

Shotgun Metagenomics

Shearing (and sequencing) of RANDOM DNA from all organisms in a sample.

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

By far the most popular method for microbiome studies. Because:

- cheap
- well developed molecular and analytical methods
- easy to perform
- "good results"

CONS:

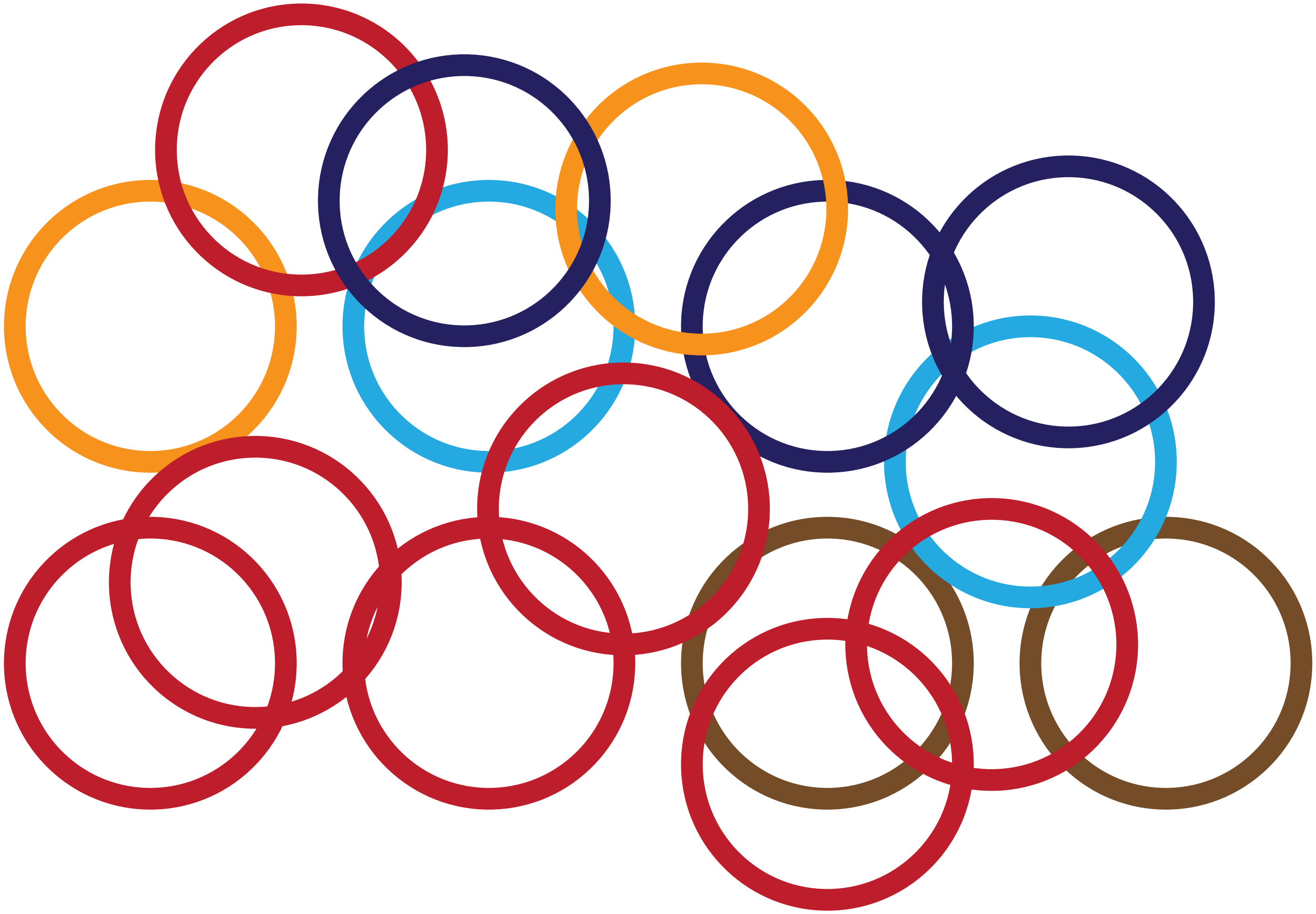
- Limited information
- Database dependent
- PCR
- Single locus

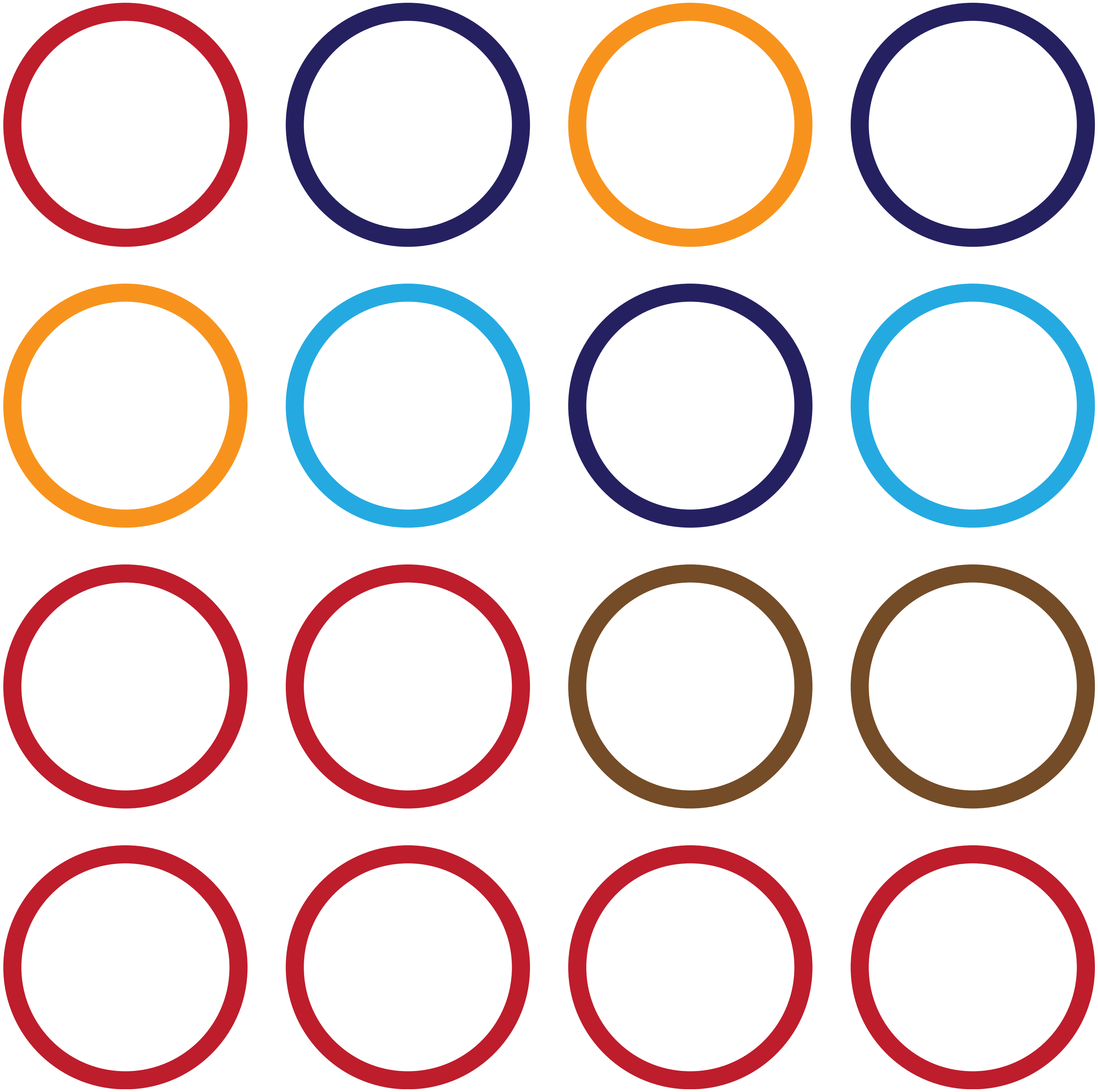
Shotgun Metagenomics

Shearing (and sequencing) of RANDOM DNA from all organisms in a sample.

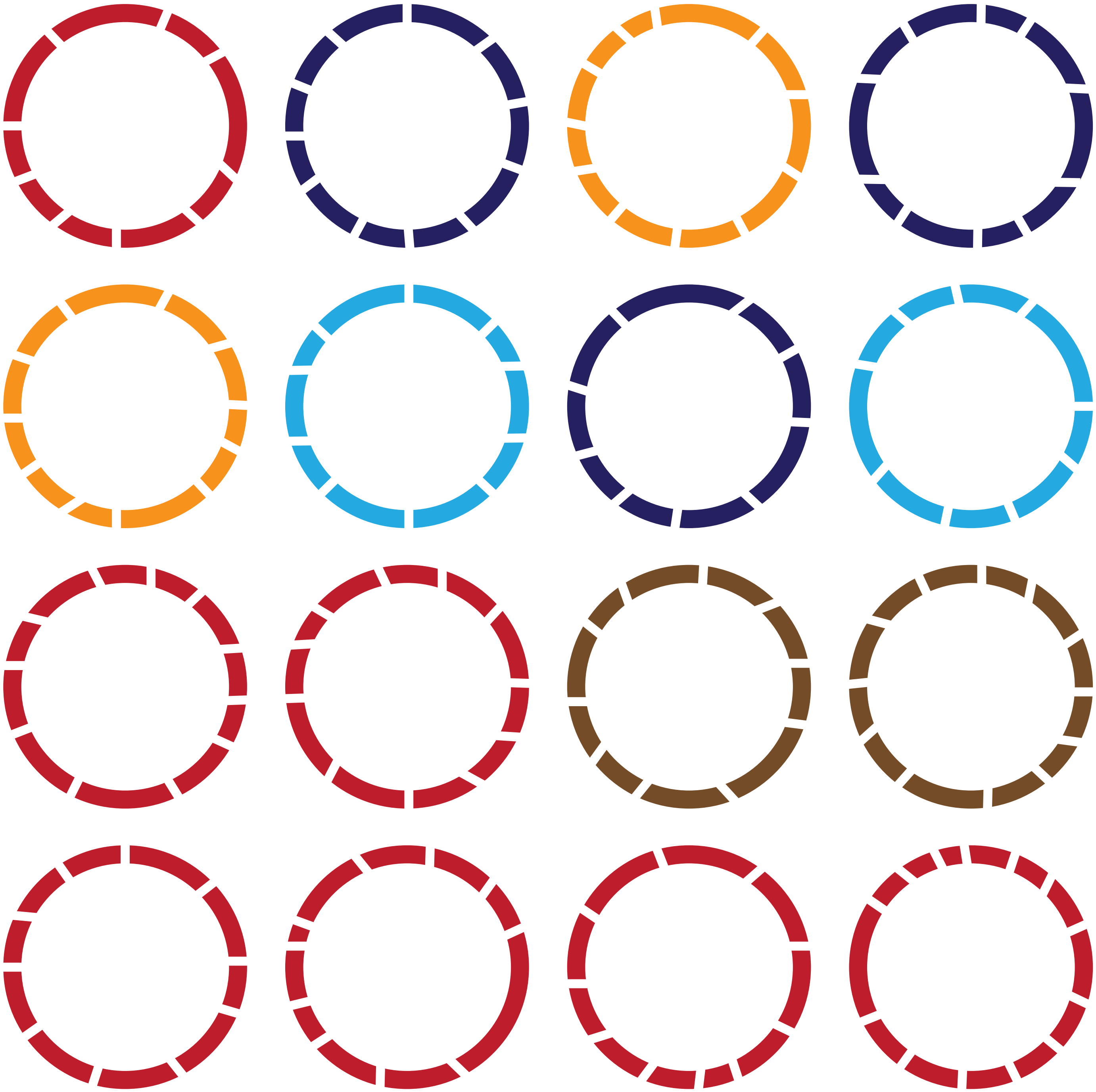
("Metagenome" = all the genomes in a sample)

A community

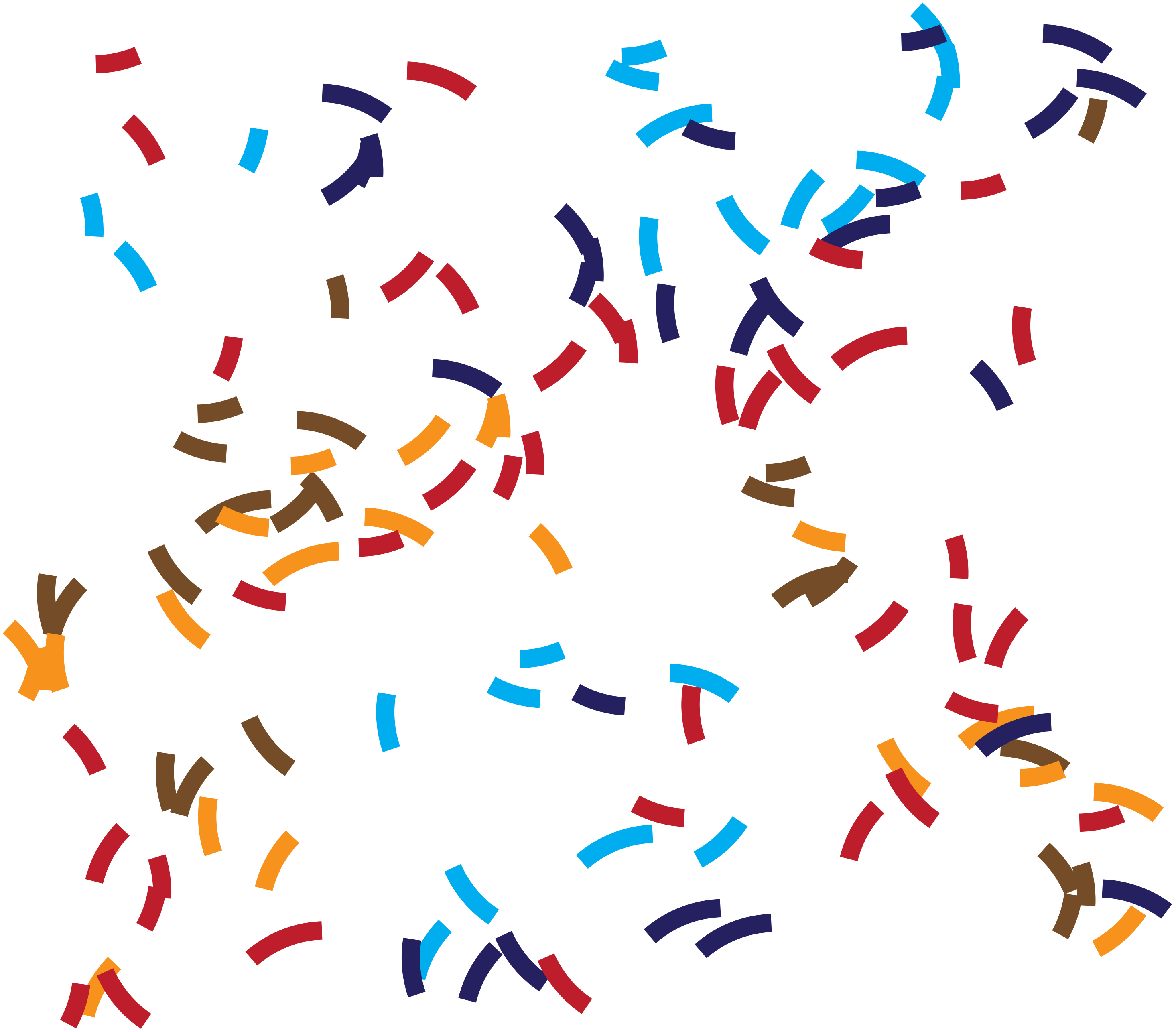




@sarahmhird

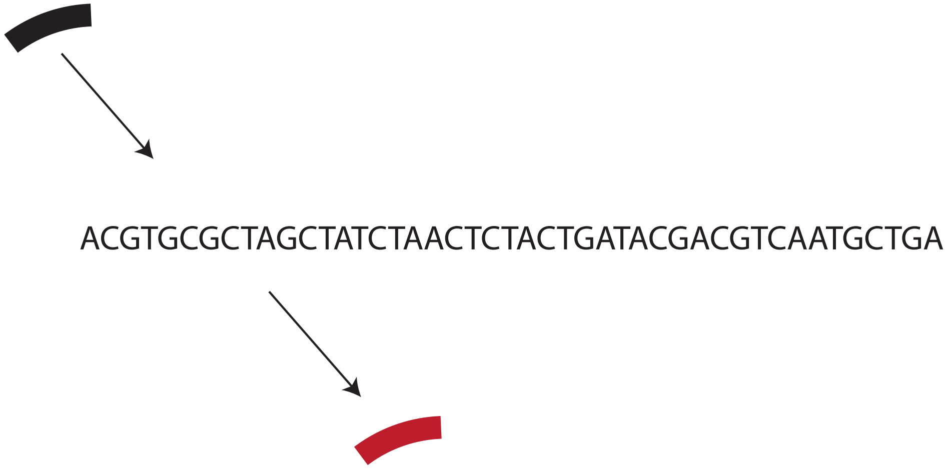


@sarahmhird



@sarahmhird

@sarahmhird



ACGTGCGCTAGCTATCTAACTCTACTGATACGACGTCAATGCTGA

16S rRNA studies

PCR-based amplification (and sequencing) of a SINGLE HOMOLOGOUS MARKER in all organisms in a sample.

By far the most popular method for microbiome studies. Because:
- cheap
- well developed molecular and analytical methods
- easy to perform
- "good results"

CONS:
- Limited information
- Database dependent
- PCR
- Single locus

Shotgun Metagenomics

Shearing (and sequencing) of a RANDOM DNA from all organisms in a sample.

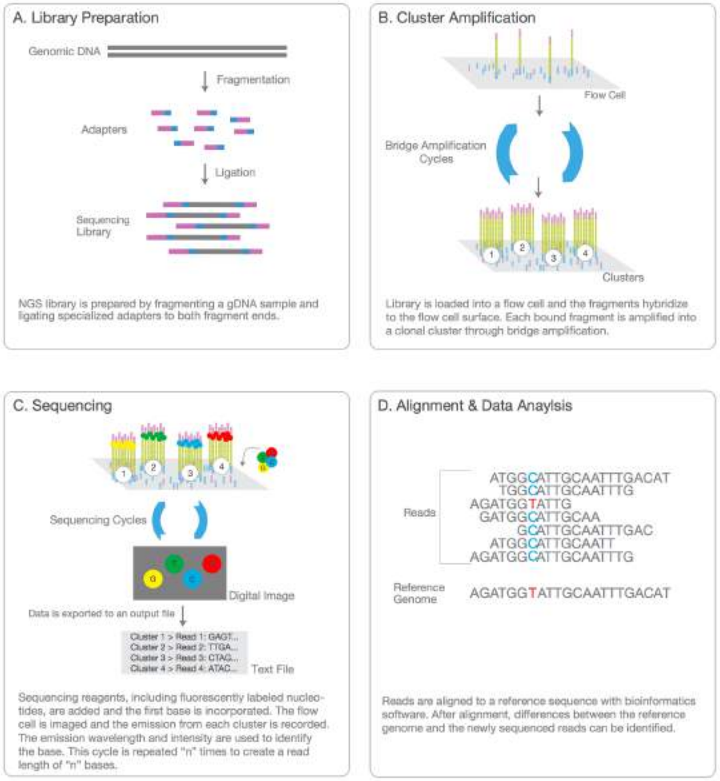
Second most popular method for microbiome studies. Because:
- Gene content!
- Genome assembly (?)
- No PCR
- "Multi-locus"

CONS:
- Limited information
- Data intensive
- Database dependent
- Expensive!

@sarahmhird

The Big One:

Illumina is a favorite method for high throughput sequencing.



NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

Library is loaded into a flow cell and the fragments hybridize to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

```
Cluster 1 > Read 1: GAGT...
Cluster 2 > Read 2: TTGA...
Cluster 3 > Read 3: CTAG...
Cluster 4 > Read 4: ATAC...
```

Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

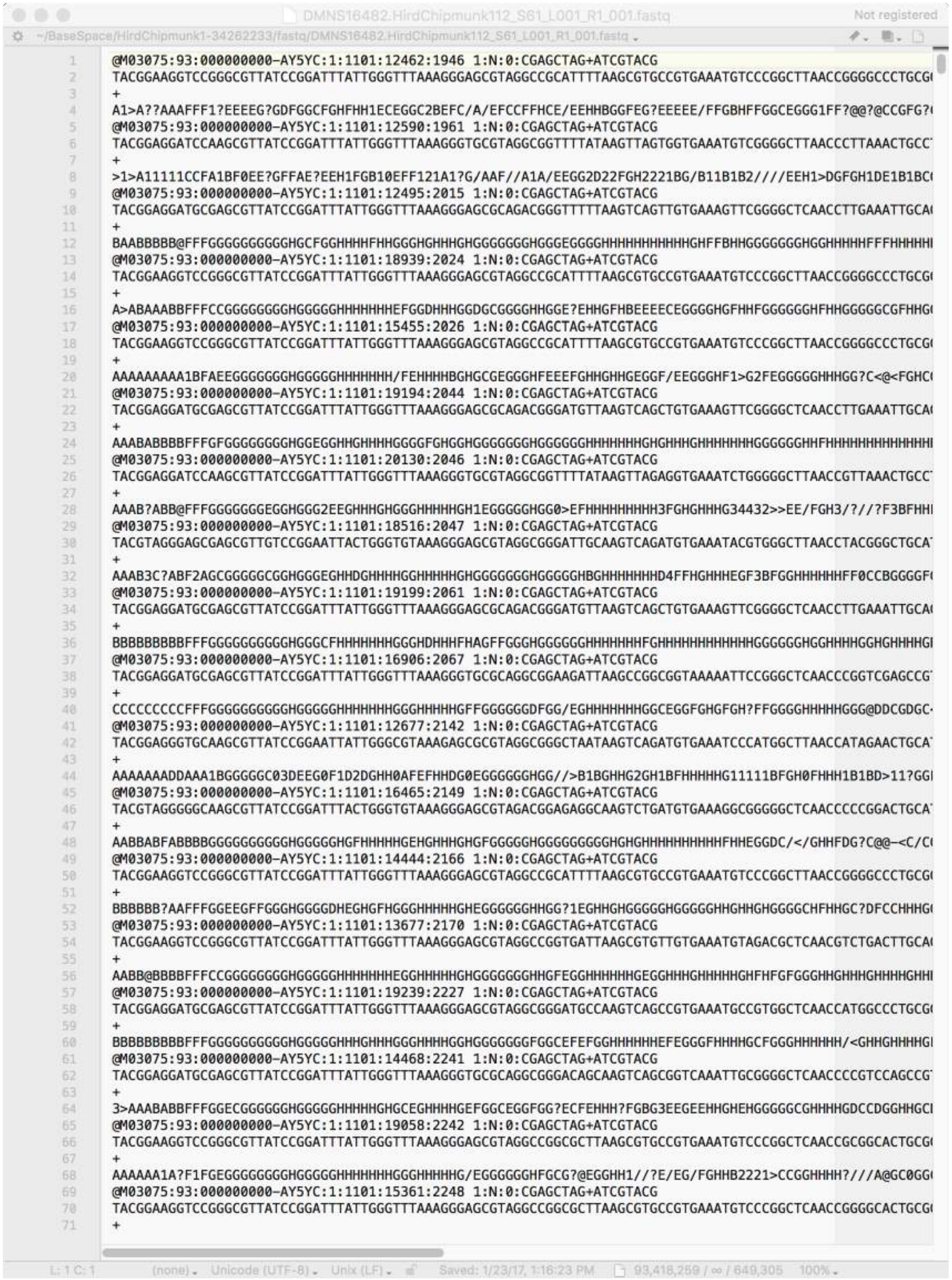
```
  ATGGCATTGCAATTTGACAT
   TGGCATTGCAATTTG
AGATGGTATTG
 GATGGCATTGCAA
     GCATTGCAATTTGAC
  ATGGCATTGCAATT
AGATGGCATTGCAATTTG

AGATGGTATTGCAATTTGACAT
```

Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

The End Result:

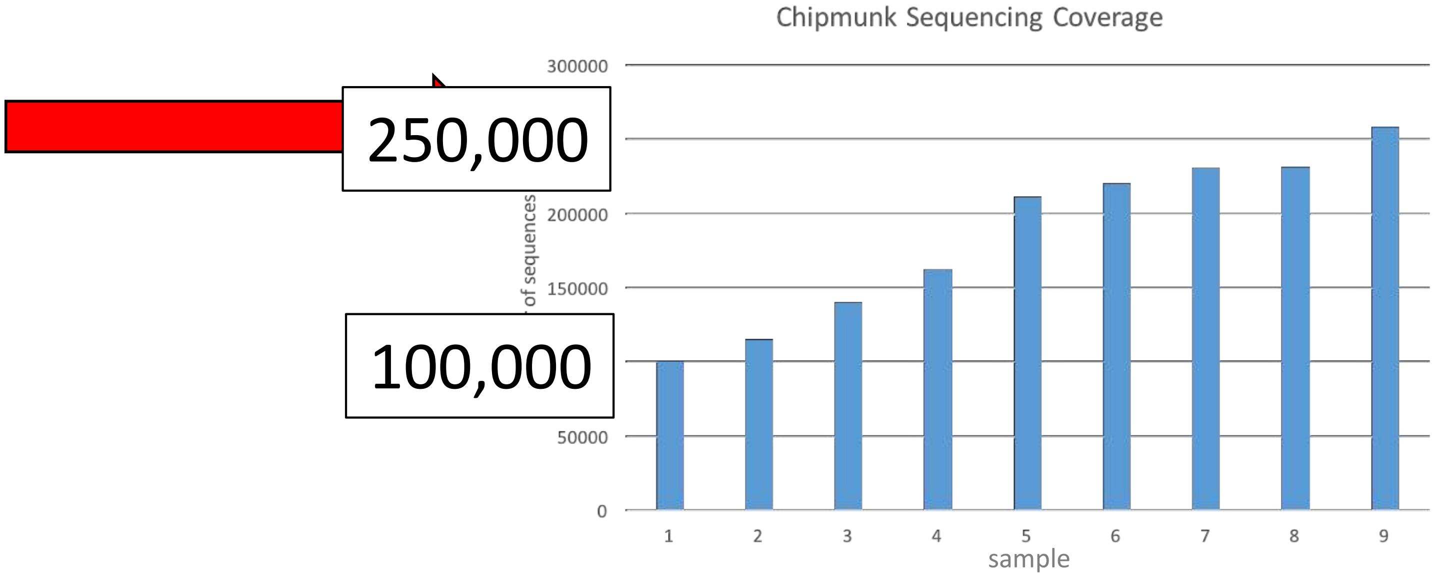
GASP! Sequences!



```
@M03075:93:000000000-AY5YC:1:1101:12462:1946 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGCATTTTAAGCGTGCCGTGAAATGTCCCGGCTTAACCGGGGCCCTGCG
+
A1>A??AAAFFF1?EEEEG?GDFGGCFGHFHH1ECEGGC2BEFC/A/EFCCFFHCE/EEHHBGGFEG?EEEEE/FFGBHFFGGCEGGG1FF?@@?@CCGFG?
@M03075:93:000000000-AY5YC:1:1101:12590:1961 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATCCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTTATAAGTTAGTGGTGAAATGTCGGGGCTTAACCCTTAAACTGCC
+
>1>A11111CCFA1BF0EE?GFFAE?EEH1FGB10EFF121A1?G/AAF//A1A/EEGG2D22FGH2221BG/B11B1B2////EEH1>DGFGH1DE1B1BC
@M03075:93:000000000-AY5YC:1:1101:12495:2015 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGCAGACGGGTTTTTAAGTCAGTTGTGAAAGTTCGGGGCTCAACCTTGAAATTGCA
+
BAABBBBB@FFFGGGGGGGGGGHGCFGGHHHHFHHGGGHGHHHGHGGGGGGGHGGGEGGGGHHHHHHHHHHHGHFFBHHGGGGGGGHGGHHHHHFFFHHHHH
@M03075:93:000000000-AY5YC:1:1101:18939:2024 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGCATTTTAAGCGTGCCGTGAAATGTCCCGGCTTAACCGGGGCCCTGCG
+
A>ABAAABBFFFCCGGGGGGGGHGGGGGHHHHHHHEFGGDHHHGGDGCGGGGHHGGE?EHHGFHBEEEECEGGGGHGFHHFGGGGGGHFHHGGGGGCGFHHG
@M03075:93:000000000-AY5YC:1:1101:15455:2026 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGCATTTTAAGCGTGCCGTGAAATGTCCCGGCTTAACCGGGGCCCTGCG
+
AAAAAAAAA1BFAEEGGGGGGGHGGGGGHHHHHHH/FEHHHHBGHGCGEGGGHFEEEFGHHGHHGEGGF/EEGGGHF1>G2FEGGGGGHHHGG?C<@<FGHC
@M03075:93:000000000-AY5YC:1:1101:19194:2044 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGCAGACGGGATGTTAAGTCAGCTGTGAAAGTTCGGGGCTCAACCTTGAAATTGCA
+
AAABABBBBFFFGFGGGGGGGGHGGEGGHHGHHHHGGGGFGHGGHGGGGGGGHGGGGGHHHHHHHGHGHHHGHHHHHHHGGGGGHHFHHHHHHHHHHHHHH
@M03075:93:000000000-AY5YC:1:1101:20130:2046 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATCCAAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTTTTATAAGTTAGAGGTGAAATCTGGGGGCTTAACCGTTAAACTGCC
+
AAAB?ABB@FFFGGGGGGGEGGHGGG2EEGHHHGHGGGHHHHHGH1EGGGGGHGG0>EFHHHHHHHHH3FGHGHHHG34432>>EE/FGH3/?//?F3BFHH
@M03075:93:000000000-AY5YC:1:1101:18516:2047 1:N:0:CGAGCTAG+ATCGTACG
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGATTGCAAGTCAGATGTGAAATACGTGGGCTTAACCTACGGGCTGCA
+
AAAB3C?ABF2AGCGGGGGCGGHGGGEGHHDGHHHHGGHHHHHGHGGGGGGGHGGGGGHBGHHHHHHHD4FFHGHHHEGF3BFGGHHHHHHFF0CCBGGGGF
@M03075:93:000000000-AY5YC:1:1101:19199:2061 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGCAGACGGGATGTTAAGTCAGCTGTGAAAGTTCGGGGCTCAACCTTGAAATTGCA
+
BBBBBBBBBFFFGGGGGGGGGGHGGGCFHHHHHHHGGGHDHHHFHAGFFGGGHGGGGGGHHHHHHHFGHHHHHHHHHHHHGGGGGHGGHHHHGGHGHHHHG
@M03075:93:000000000-AY5YC:1:1101:16906:2067 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGCAGGCGGAAGATTAAGCCGGCGGTAAAAATTCCGGGCTCAACCCGGTCGAGCCG
+
CCCCCCCCCFFFGGGGGGGGGGHGGGGGHHHHHHHGGGHHHHHGFFGGGGGGDFGG/EGHHHHHHHGGCEGGFGHGFGH?FFGGGGHHHHHGGG@DDCGDGC
@M03075:93:000000000-AY5YC:1:1101:12677:2142 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGGTGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCGCGTAGGCGGGCTAATAAGTCAGATGTGAAATCCCATGGCTTAACCATAGAACTGCA
+
AAAAAAADDAAA1BGGGGGC03DEEG0F1D2DGHH0AFEFHHDG0EGGGGGGHGG//>B1BGHHG2GH1BFHHHHHG11111BFGH0FHHH1B1BD>11?GG
@M03075:93:000000000-AY5YC:1:1101:16465:2149 1:N:0:CGAGCTAG+ATCGTACG
TACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGACGGAGAGGCAAGTCTGATGTGAAAGGCGGGGGCTCAACCCCCGGACTGCA
+
AABBABFABBBBGGGGGGGGGGHGGGGGHGFHHHHHGEHGHHHGHGFGGGGGHGGGGGGGGGHGHGHHHHHHHHHHFHHEGGDC/</GHHFDG?C@@-<C/C
@M03075:93:000000000-AY5YC:1:1101:14444:2166 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGCATTTTAAGCGTGCCGTGAAATGTCCCGGCTTAACCGGGGCCCTGCG
+
BBBBBB?AAFFFGGEEGFFGGGHGGGGDHEGHGFHGGGHHHHHGHEGGGGGGHHGG?1EGHHGHGGGGGHGGGGGHHGHHGHGGGGCHFHHGC?DFCCHHHG
@M03075:93:000000000-AY5YC:1:1101:13677:2170 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGGTGATTAAGCGTGTTGTGAAATGTAGACGCTCAACGTCTGACTTGCA
+
AABB@BBBBFFFCCGGGGGGGGHGGGGGHHHHHHHEGGHHHHHGHGGGGGGGHHGFEGGHHHHHHGEGGHHHGHHHHHGHFHFGFGGGHHGHHHGHHHHGHH
@M03075:93:000000000-AY5YC:1:1101:19239:2227 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCGGGATGCCAAGTCAGCCGTGAAATGCCGTGGCTCAACCATGGCCCTGCG
+
BBBBBBBBBFFFGGGGGGGGGGHGGGGGHHHGHHHGGGHHHHGGHGGGGGGGFGGCEFEFGGHHHHHHEFEGGGFHHHHGCFGGGHHHHHH/<GHHGHHHHG
@M03075:93:000000000-AY5YC:1:1101:14468:2241 1:N:0:CGAGCTAG+ATCGTACG
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGCAGGCGGGACAGCAAGTCAGCGGTCAAATTGCGGGGCTCAACCCCGTCCAGCCG
+
3>AAABABBFFFGGECGGGGGGHGGGGGHHHHHGHGCEGHHHHGEFGGCEGGFGG?ECFEHHH?FGBG3EEGEEHHGHEHGGGGGCGHHHHGDCCDGGHHGC
@M03075:93:000000000-AY5YC:1:1101:19058:2242 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGGCGCTTAAGCGTGCCGTGAAATGTCCCGGCTCAACCGCGGCACTGCG
+
AAAAAA1A?F1FGEGGGGGGGGHGGGGGHHHHHHHGGGHHHHHG/EGGGGGGHFGCG?@EGGHH1//?E/EG/FGHHB2221>CCGGHHHH?///A@GC0GG
@M03075:93:000000000-AY5YC:1:1101:15361:2248 1:N:0:CGAGCTAG+ATCGTACG
TACGGAAGGTCCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCCGGCGCTTAAGCGTGCCGTGAAATGTCCCGGCTCAACCGGGGCACTGCG
+
```

Quality Control: Sequence Coverage

Good



250,000

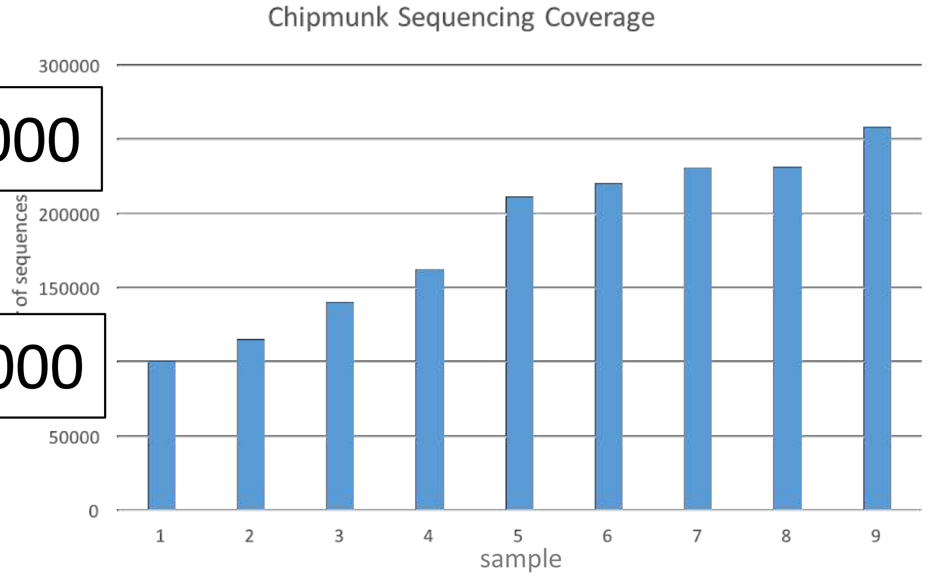
100,000

Quality Control: Sequence Coverage

Good

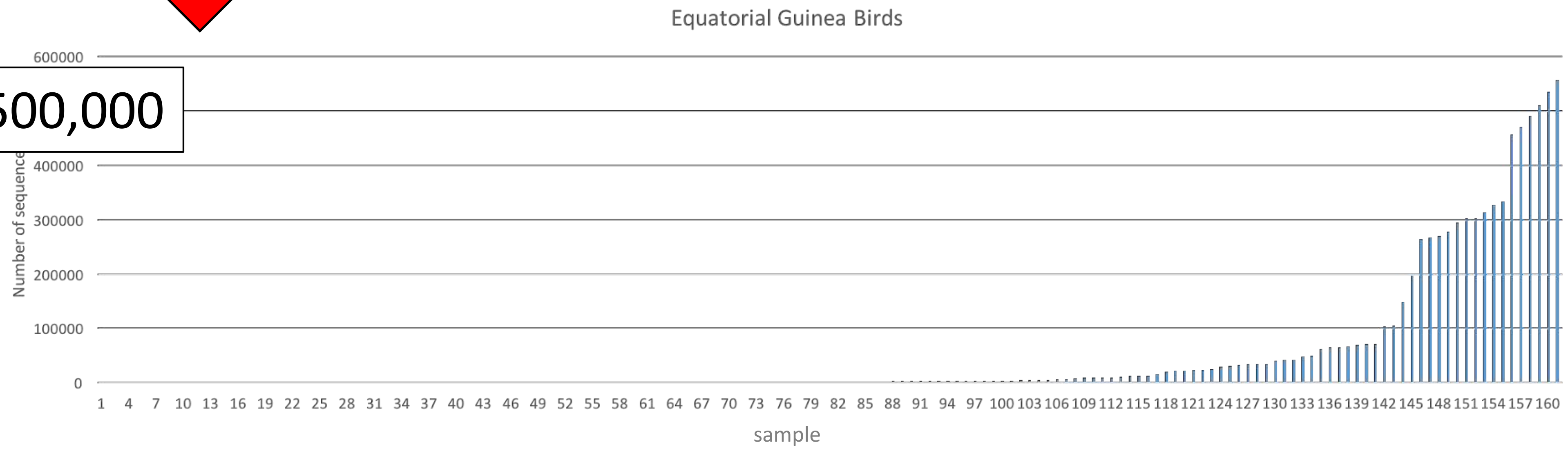
250,000

100,000

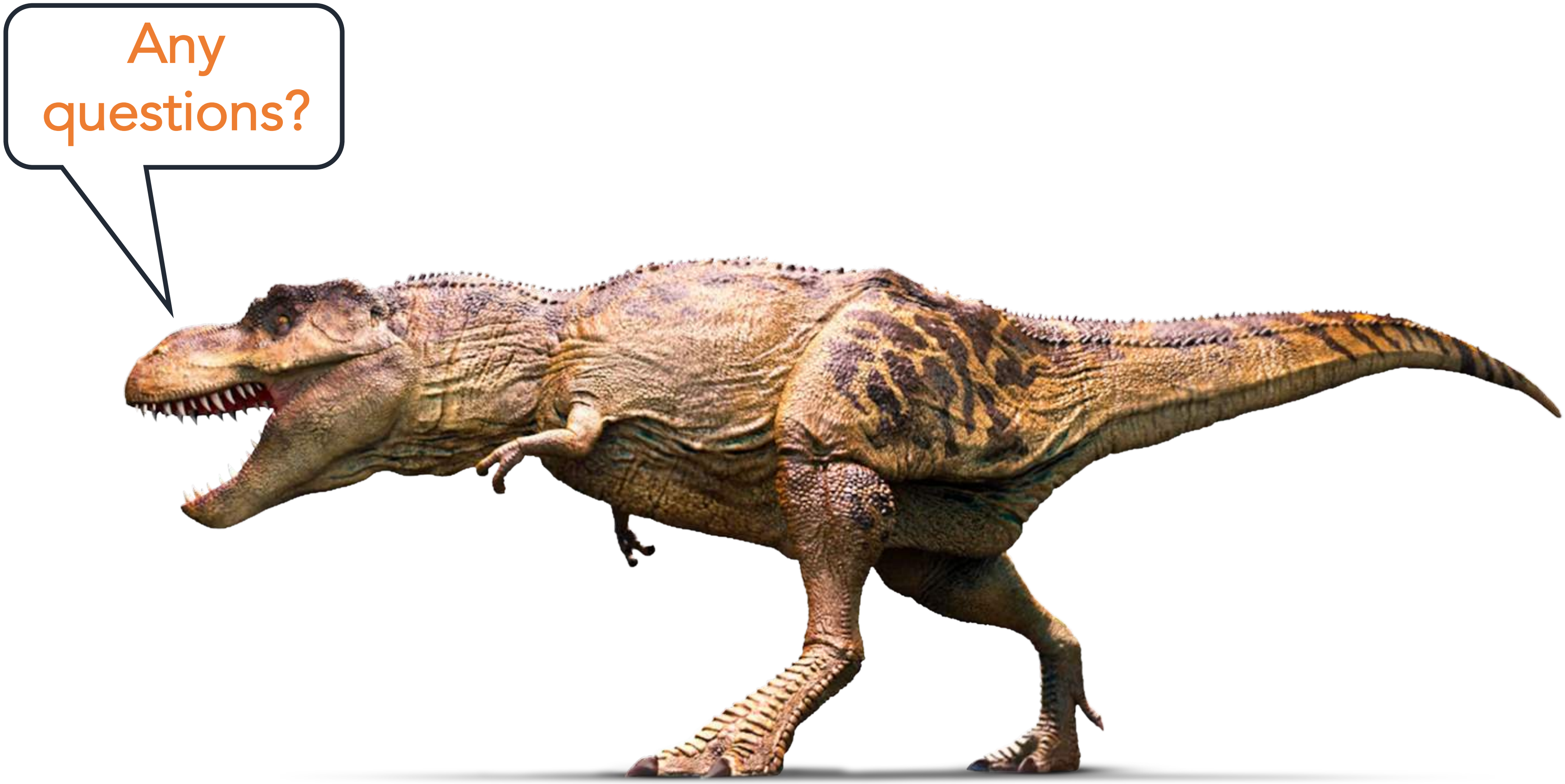


Boohoo

500,000



@sarahmhird



Tyrannosaurus

DK Find Out!

4. Analyze the results (16S)

Analyses of 16S rRNA sequence data usually consist of:
- **Identify** “who” is there
- **Describe**/measure/quantify diversity (alpha and beta)
- **Compare** categories

The Big Two:

There are two excellent, free analysis packages for analyzing microbiome (16S) data:

Mothur

QIIME (or QIIME2)

The Big Two:

There are two excellent, free analysis packages for analyzing microbiome (16S) data:

Mothur

QIIME (or QIIME2)

Analyses of 16S rRNA sequence data usually consist of:
- **Identify** “who” is there
- **Describe**/measure/quantify diversity (alpha and beta)
- **Compare** categories

The Big Two:

There are two excellent, free analysis packages for analyzing microbiome (16S) data:

Mothur

QIIME (or QIIME2)

phyloseq: Import, share, and analyze microbiome census data using R

Analyses of 16S rRNA sequence data usually consist of:
- **Identify** “who” is there
- **Describe**/measure/quantify diversity (alpha and beta)
- **Compare** categories

In order to analyze ANYTHING – we must have **BIOLOGICAL UNITS.**

These include:

Species

Genera (etc)

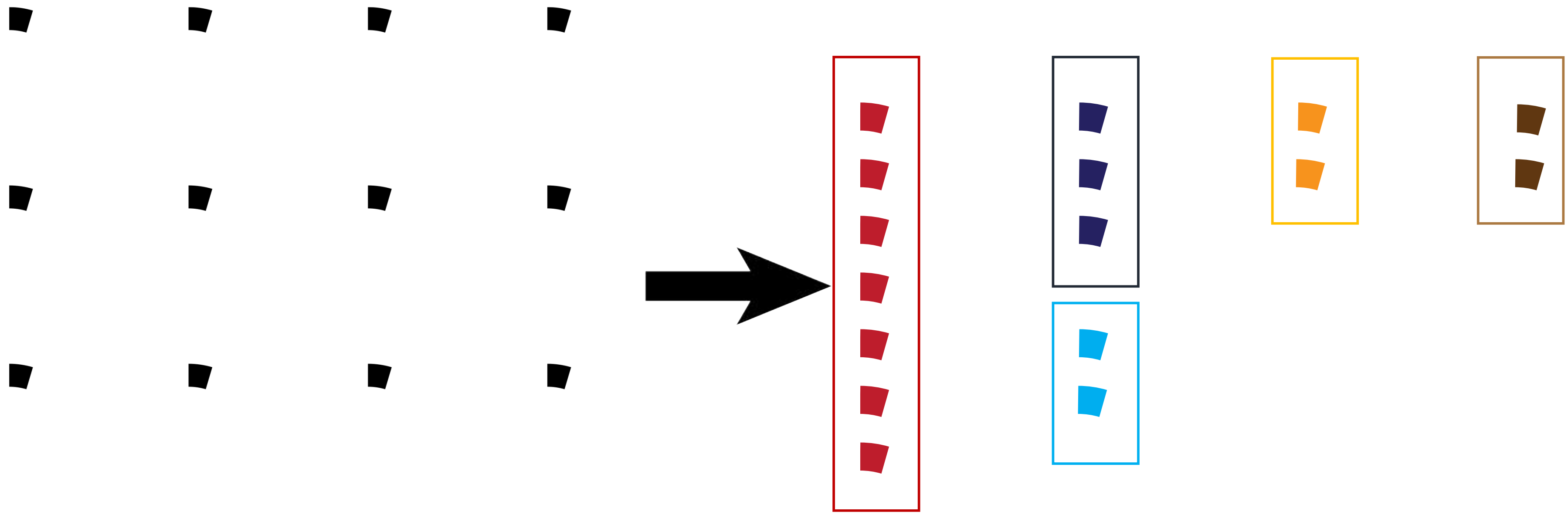
OTUs

This is a fundamental issue in microbiome research!

OTU picking

"Operational Taxonomic Unit"

Given a set of sequences: how do you sort them into "species"?



OTU picking

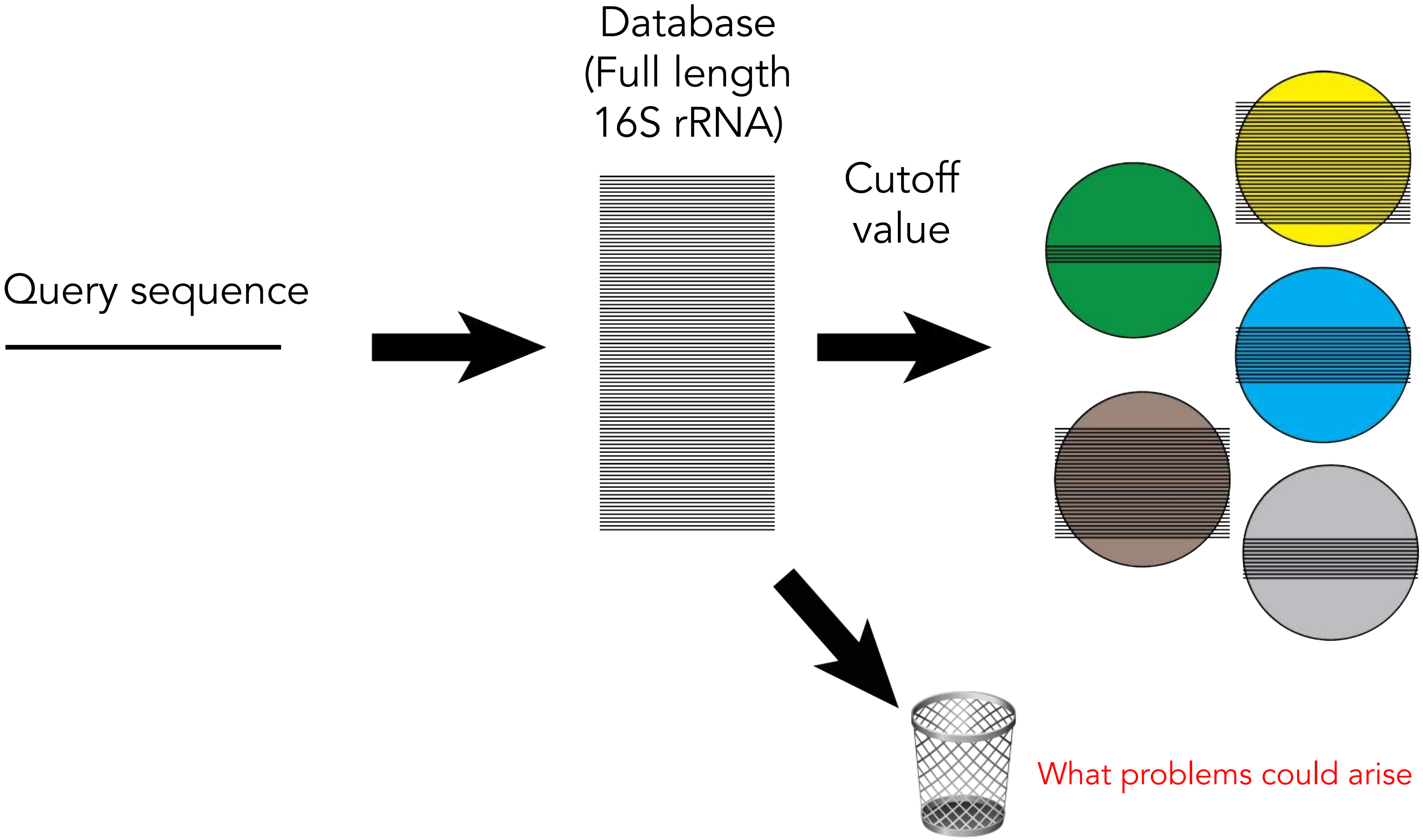
"Operational Taxonomic Unit"

Given a set of sequences: how do you sort them into "species"?

Three (classes of) methods:

- Closed reference
- De novo
- Exact Sequence Variants (ESV)

OTU picking: Closed reference



Query sequence

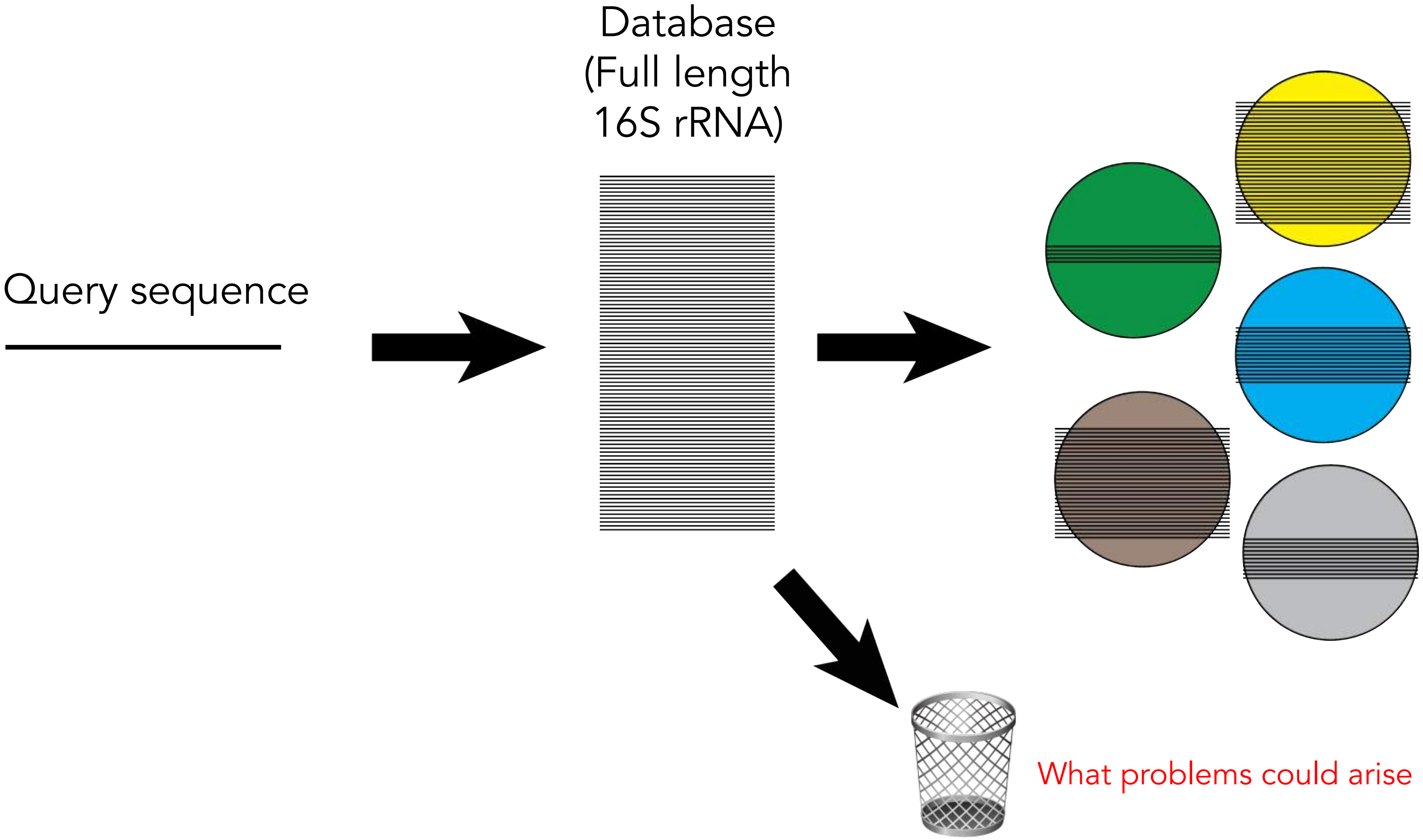
Database
(Full length
16S rRNA)

Cutoff
value

What problems could arise here?

OTU picking: Closed reference

Justification = Using database allows for comparison across datasets



Query sequence

Database
(Full length
16S rRNA)

What problems could arise here?

OTU picking: Closed reference

✓ PEER-REVIEWED

Accuracy of microbial community diversity estimated by closed- and open-reference OTUs

Robert C. Edgar

Published October 4, 2017 PubMed 29018622

"Biological inferences obtained using these methods are therefore not reliable."

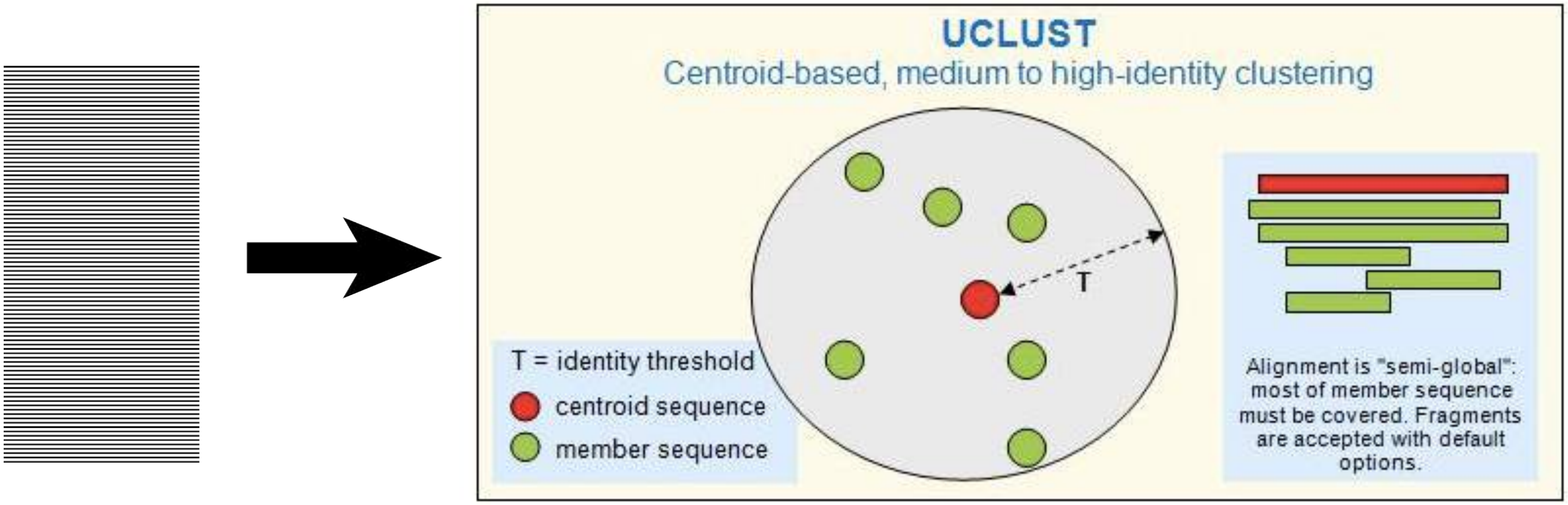
OTU picking: De novo (UCLUST)

Volume 26, Issue 19
1 October 2010

Search and clustering orders of magnitude faster than BLAST

Robert C. Edgar

Bioinformatics (2010) 26 (19): 2460-2461.
DOI: https://doi.org/10.1093/bioinformatics/btq461
Published: 12 August 2010 **Article history**



Alignment is "semi-global": most of member sequence must be covered. Fragments are accepted with default options.

http://drive5.com/usearch/manual/uclust_algo.html

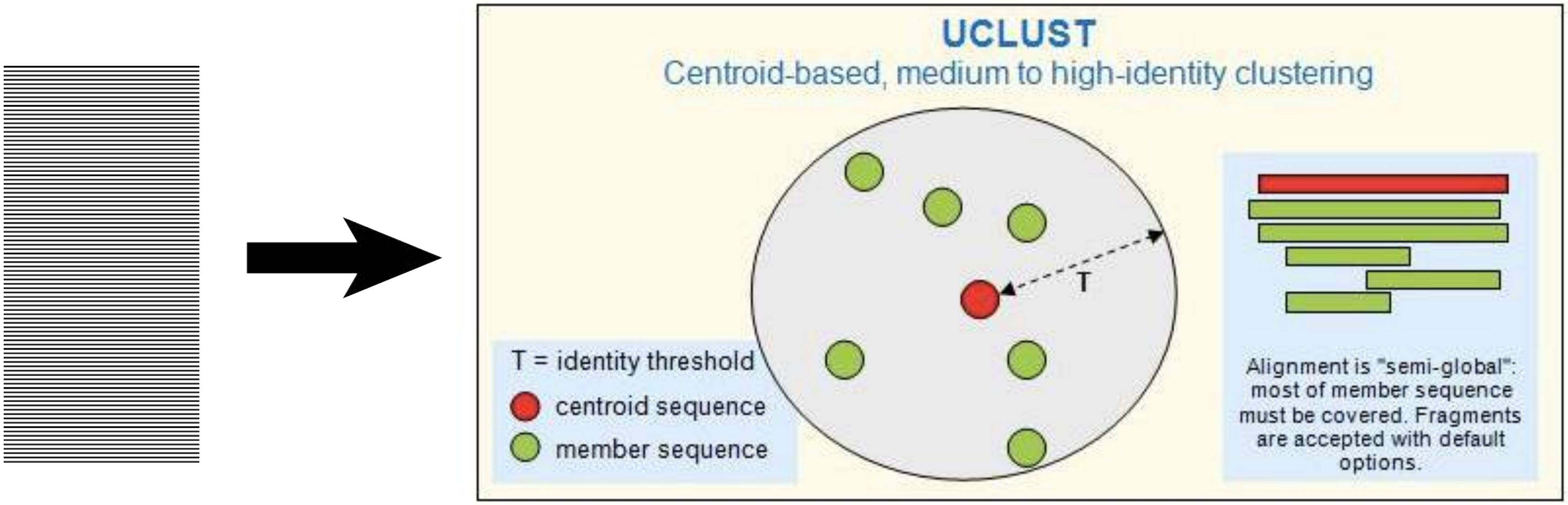
OTU picking: De novo (UCLUST)

Volume 26, Issue 19
1 October 2010

Search and clustering orders of magnitude faster than BLAST

Robert C. Edgar

Bioinformatics (2010) 26 (19): 2460-2461.
DOI: https://doi.org/10.1093/bioinformatics/btq461
Published: 12 August 2010 **Article history ▾**



UCLUST

Centroid-based, medium to high-identity clustering

Alignment is "semi-global": most of member sequence must be covered. Fragments are accepted with default options.

http://drive5.com/usearch/manual/uclust_algo.html

PROS: Fast, efficient, pairwise (no global alignment step)
CONS: threshold, input order dependent, overlapping clusters

OTU picking: De novo (swarm)

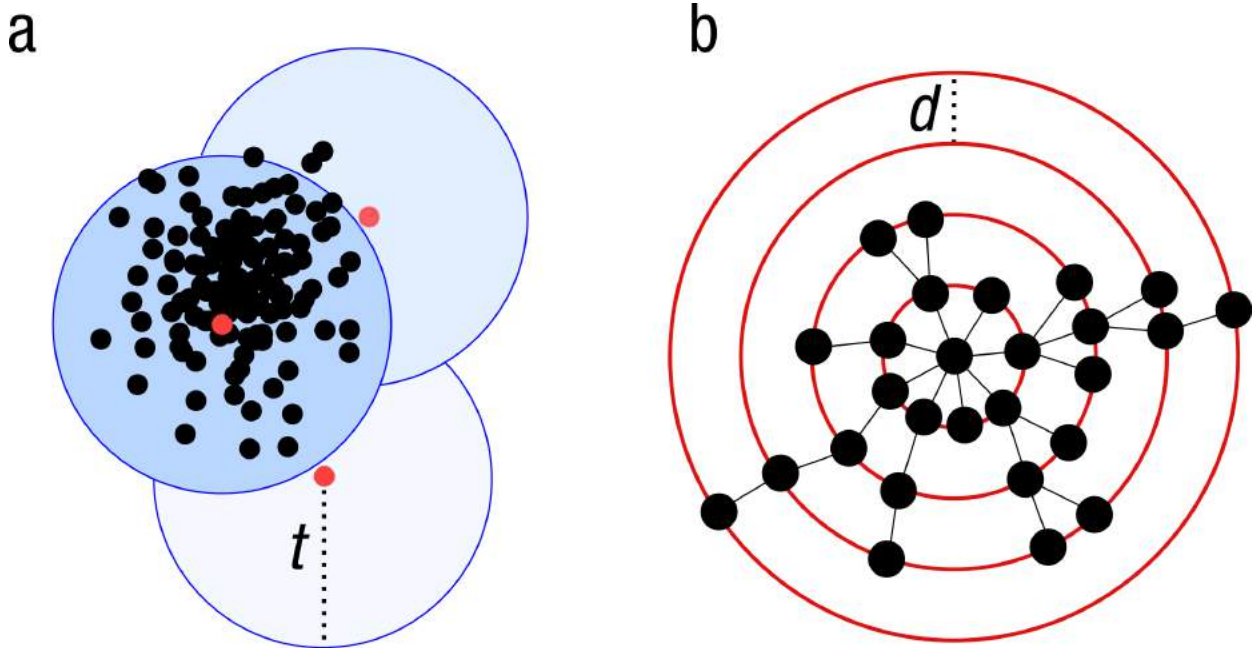
✓ PEER-REVIEWED

Swarm: robust and fast clustering method for amplicon-based studies

Biodiversity | Bioinformatics | Ecology | Microbiology | Molecular Biology

Frédéric Mahé[1,2,3], Torbjørn Rognes[4,5], Christopher Quince[6], Colomban de Vargas[1,2], Micah Dunthorn[3]

Published September 25, 2014



OTU picking: 97%

- Frequently cited as a good species threshold for microbes because it was established that it worked well for some known species using the full 16S gene (~1500bp)
- Highly debated / disputed value
- No single cutoff works for all life.

OTU picking: ASVs

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Altmetric: 101 Citations: 2 **More detail »**

Perspective | OPEN

Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan, Paul J McMurdie & Susan P Holmes

The ISME Journal **11**, 2639–2643 (2017)
doi:10.1038/ismej.2017.119
Download Citation

Metagenomics Microbiome
Next-generation sequencing

Received: 21 March 2017
Revised: 20 May 2017
Accepted: 07 June 2017
Published online: 21 July 2017

First biological unit = AAACTCTATCTATCTACTCTCGCGCGTACGCGTCAT

Second biological unit = AACCCCTCGCACGACCAGCACAACACAACTACCA

Third biological unit = AACTCCGTAAAACTACAACTACTACTACCATACACG

etc.

OTU picking: ASVs

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Altmetric: 101 Citations: 2 More detail »

Perspective | OPEN

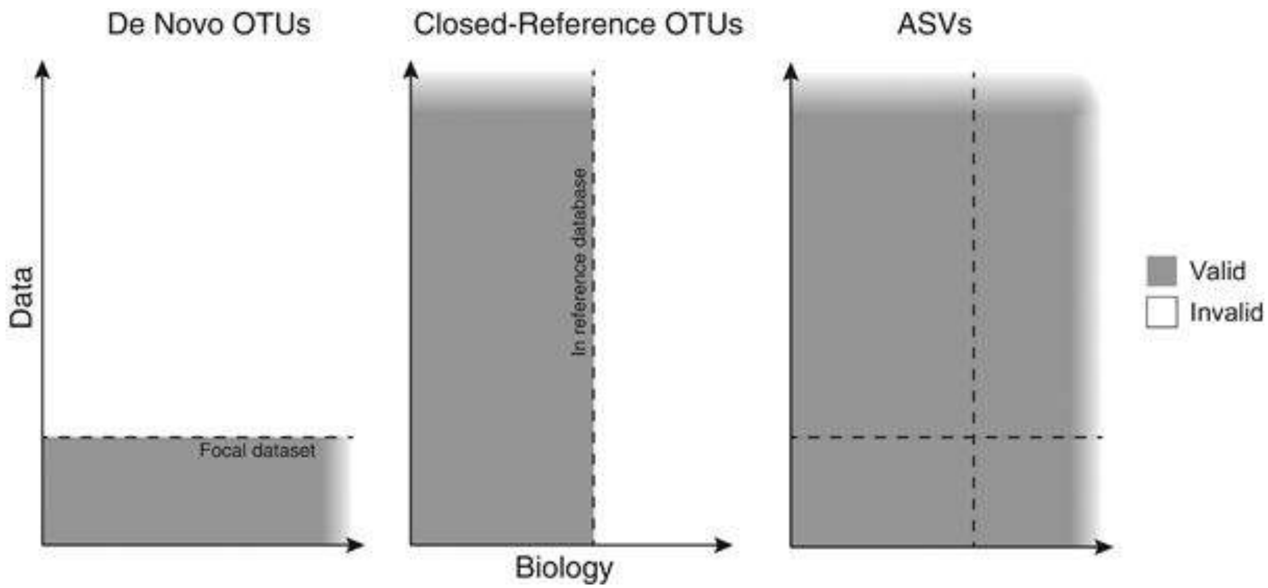
Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan, Paul J McMurdie & Susan P Holmes

The ISME Journal 11, 2639–2643 (2017)
doi:10.1038/ismej.2017.119
Download Citation

Metagenomics Microbiome
Next-generation sequencing

Received: 21 March 2017
Revised: 20 May 2017
Accepted: 07 June 2017
Published online: 21 July 2017



First biological unit = AAACTCTATCTATCTACTCTCGCGCGTACGCGTCAT

Second biological unit = AACCCCTCGCACGACCAGCACAACACAACTACCA

Third biological unit = AACTCCGTAAAACTACAACTACTACTACCATACACG

etc.

OTU picking: ASVs

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Altmetric: 101 Citations: 2 More detail »

Perspective | OPEN

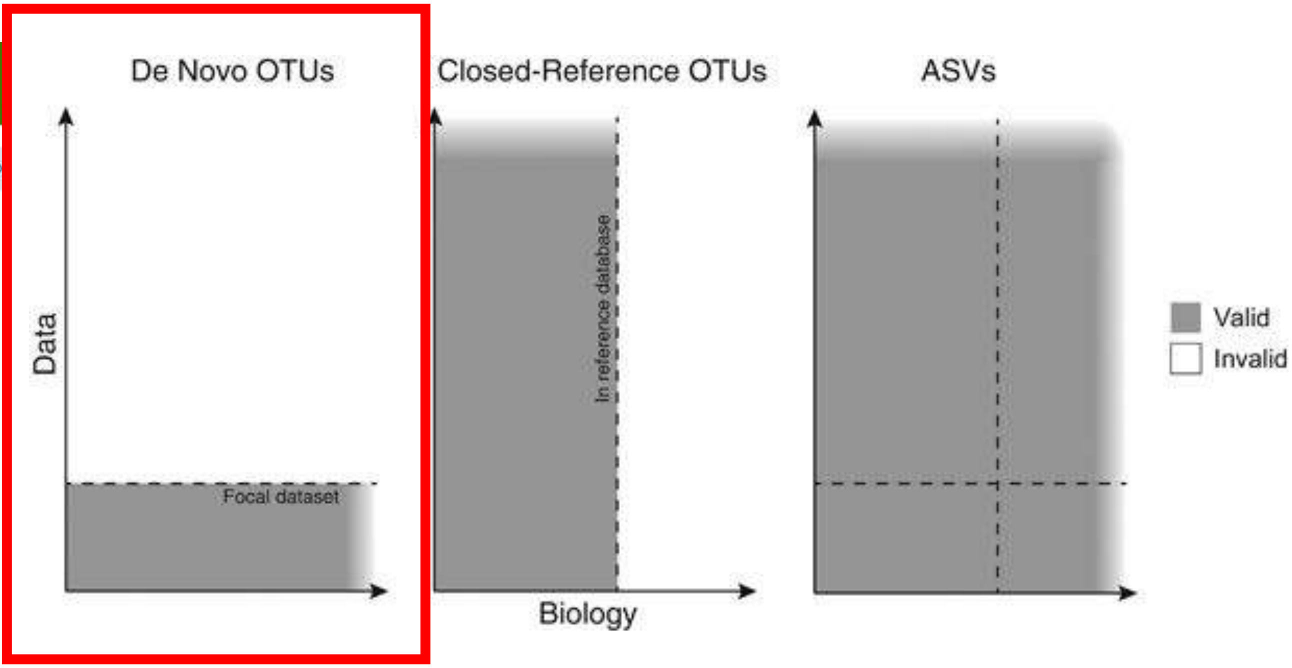
Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan, Paul J McMurdie & Susan P Holmes

The ISME Journal **11**, 2639–2643 (2017)
doi:10.1038/ismej.2017.119
Download Citation

Metagenomics Microbiome
Next-generation sequencing

Received: 21 March 2017
Revised: 20 May 2017
Accepted: 07 June 2017
Published online: 21 July 2017



First biological unit = AAACTCTATCTATCTACTCTCGCGCGTACGCGTCAT

Second biological unit = AACCCCTCGCACGACCAGCACAACACAACTACCA

Third biological unit = AACTCCGTAAAACTACAACTACTACTACCATACACG

etc.

OTU picking: ASVs

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Altmetric: 101 Citations: 2 More detail »

Perspective | OPEN

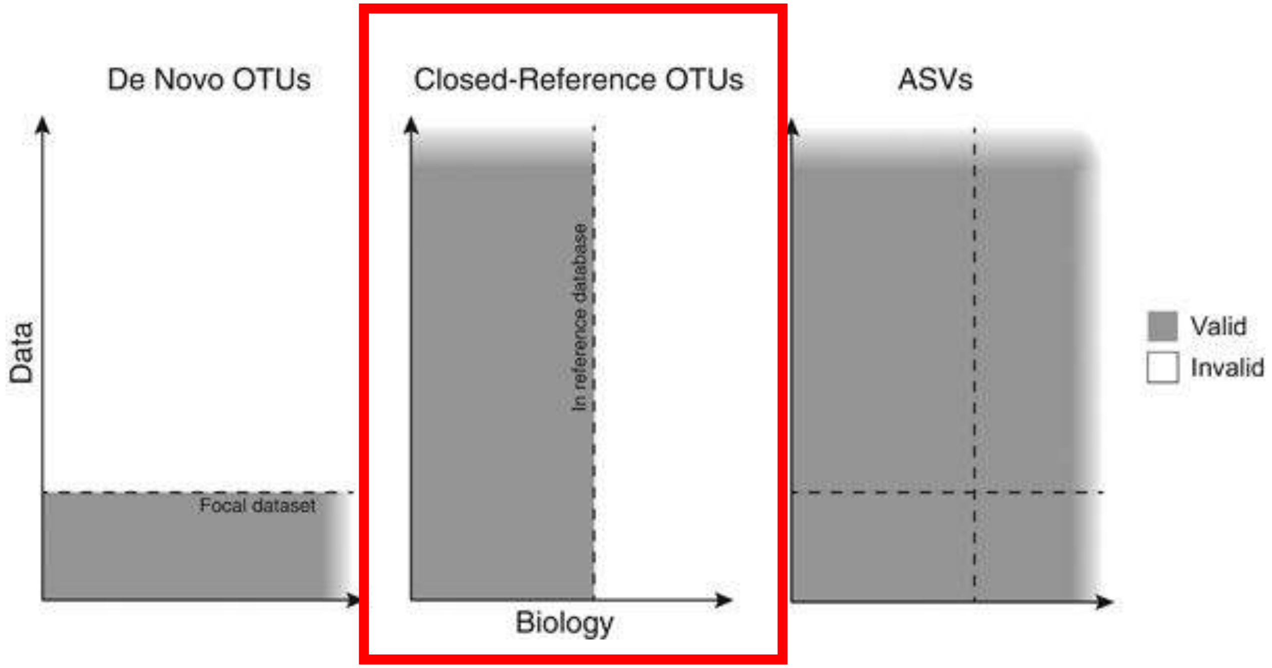
Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan, Paul J McMurdie & Susan P Holmes

The ISME Journal **11**, 2639–2643 (2017)
doi:10.1038/ismej.2017.119
Download Citation

Metagenomics Microbiome
Next-generation sequencing

Received: 21 March 2017
Revised: 20 May 2017
Accepted: 07 June 2017
Published online: 21 July 2017



First biological unit = AAACTCTATCTATCTACTCTCGCGCGTACGCGTCAT

Second biological unit = AACCCCTCGCACGACCAGCACAACACAACTACCA

Third biological unit = AACTCCGTAAAACTACAACTACTACTACCATACACG

etc.

OTU picking: ASVs

The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Altmetric: 101 Citations: 2 More detail »

Perspective | OPEN

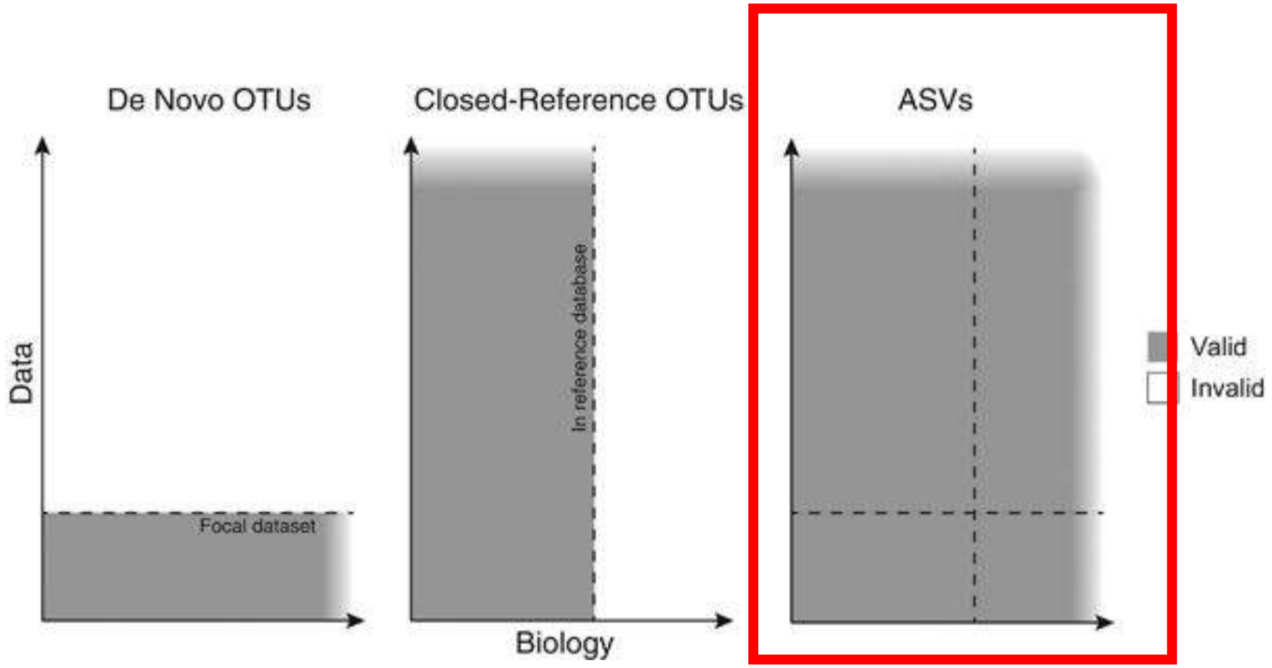
Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan, Paul J McMurdie & Susan P Holmes

The ISME Journal **11**, 2639–2643 (2017)
doi:10.1038/ismej.2017.119
Download Citation

Metagenomics Microbiome
Next-generation sequencing

Received: 21 March 2017
Revised: 20 May 2017
Accepted: 07 June 2017
Published online: 21 July 2017



First biological unit = AAACTCTATCTATCTACTCTCGCGCGTACGCGTCAT

Second biological unit = AACCCCTCGCACGACCAGCACAACACAACTACCA

Third biological unit = AACTCCGTAAAACTACAACTACTACTACCATACACG

etc.

OTU picking: ASVs (DADA2)

Does quality filtering of each sequence (So must have fastq files)

Processes data within each sample (an error in one sample doesn't make something an error in another sample)

```
plotQualityProfile(fnFs[1:2])
```

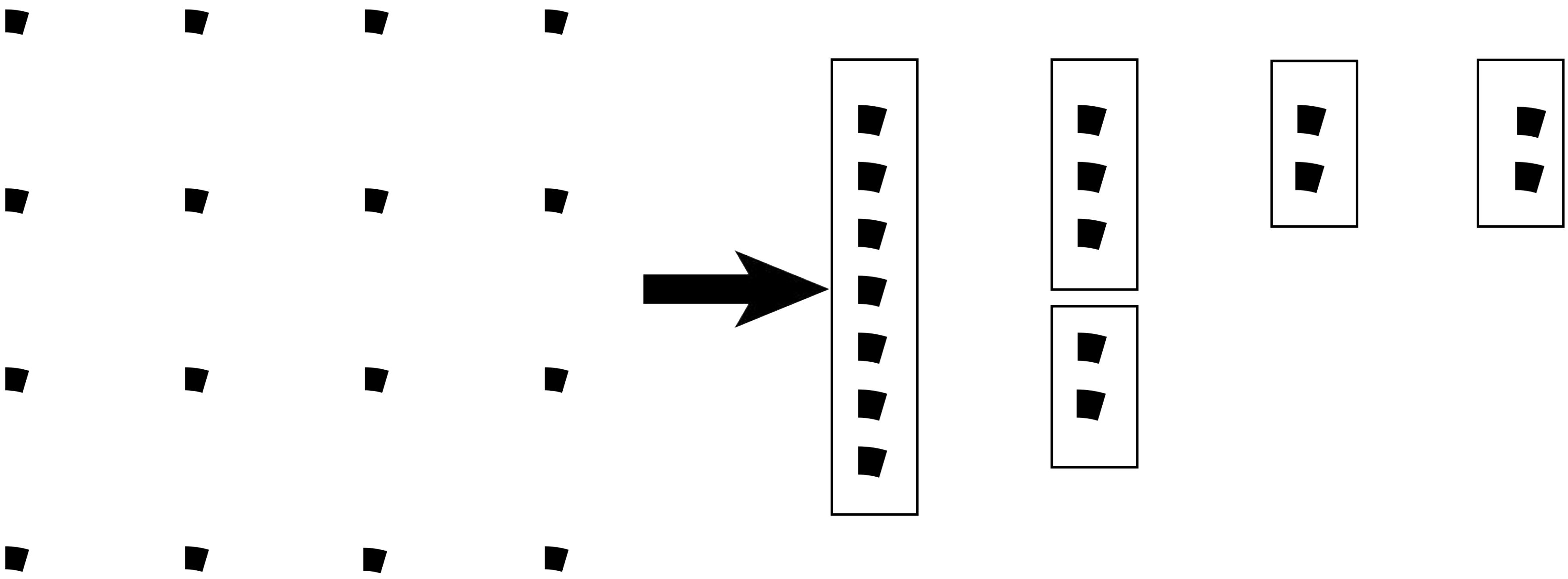
The forward reads are good quality. We generally advise trimming the last few nucleotides to avoid less well-controlled errors that can arise there. These quality profiles do not suggest that any additional trimming is needed, so we will truncate the forward reads at position 240 (trimming the last 10 nucleotides).

Now we visualize the quality profile of the reverse reads:

```
plotQualityProfile(fnRs[1:2])
```

https://benjjneb.github.io/dada2/tutorial.html

16S Analytical Methods



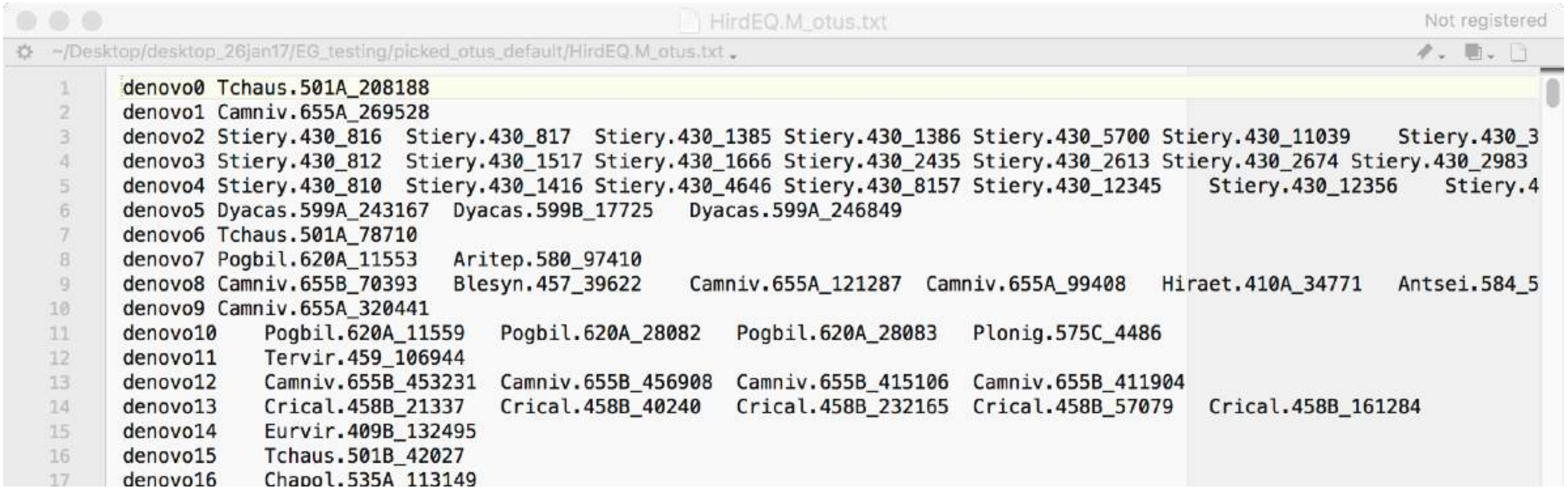
OTU picking creates a list

Sample 1:

- OTU001 = Sequence001, Sequence098, Sequence10232
- OTU002 = Sequence004
- OTU003 = Sequence002, Sequence003, Sequence007… (contains 50,000 sequences)
- OTU004 = Sequence006, Sequence45, Sequence15601, Sequence48973
- Etc.

Many samples together:

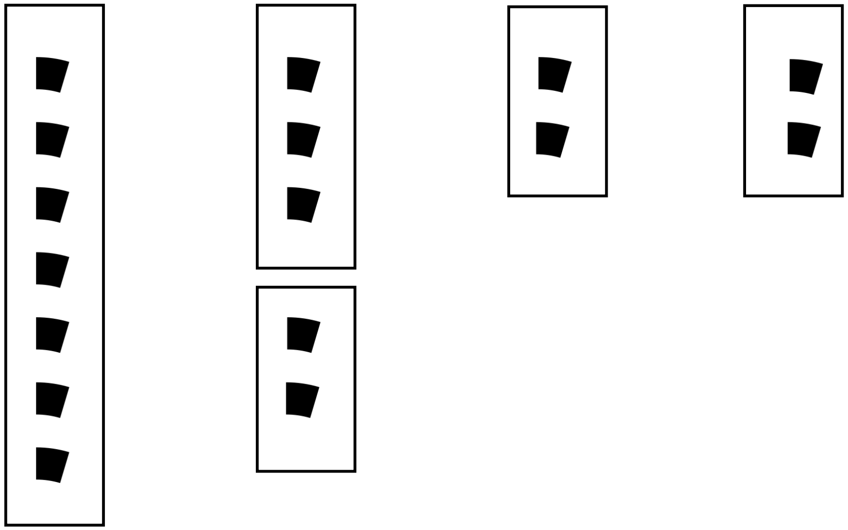
OTU001 = SampleA_Seq001, SampleA_Seq098, SampleB_Seq10232
OTU002 = SampleA_Seq004
OTU003 = SampleB_Seq002, SampleC_Seq003, SampleC_Seq005 … (contains 50,000 sequences)
OTU004 = SampleA_Seq006, SampleA_Seq45, SampleB_Seq15601, SampleC_Seq48973
Etc.



```
denovo0 Tchaus.501A_208188
denovo1 Camniv.655A_269528
denovo2 Stiery.430_816  Stiery.430_817  Stiery.430_1385 Stiery.430_1386 Stiery.430_5700 Stiery.430_11039    Stiery.430_3
denovo3 Stiery.430_812  Stiery.430_1517 Stiery.430_1666 Stiery.430_2435 Stiery.430_2613 Stiery.430_2674 Stiery.430_2983
denovo4 Stiery.430_810  Stiery.430_1416 Stiery.430_4646 Stiery.430_8157 Stiery.430_12345    Stiery.430_12356    Stiery.4
denovo5 Dyacas.599A_243167  Dyacas.599B_17725   Dyacas.599A_246849
denovo6 Tchaus.501A_78710
denovo7 Pogbil.620A_11553   Aritep.580_97410
denovo8 Camniv.655B_70393   Blesyn.457_39622    Camniv.655A_121287  Camniv.655A_99408   Hiraet.410A_34771   Antsei.584_5
denovo9 Camniv.655A_320441
denovo10    Pogbil.620A_11559   Pogbil.620A_28082   Pogbil.620A_28083   Plonig.575C_4486
denovo11    Tervir.459_106944
denovo12    Camniv.655B_453231  Camniv.655B_456908  Camniv.655B_415106  Camniv.655B_411904
denovo13    Crical.458B_21337   Crical.458B_40240   Crical.458B_232165  Crical.458B_57079   Crical.458B_161284
denovo14    Eurvir.409B_132495
denovo15    Tchaus.501B_42027
denovo16    Chapol.535A_113149
```

16S Analytical Methods

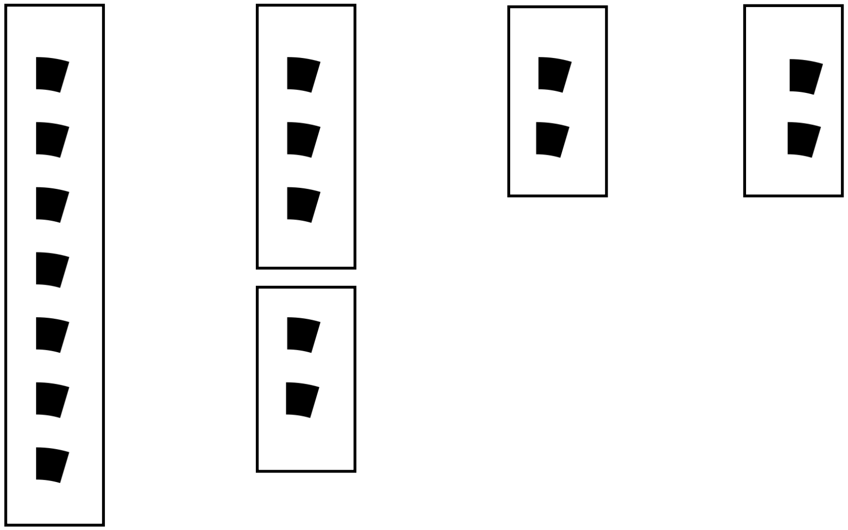
Just by grouping sequences into OTUs, we already have some information about the diversity of this sample:



SampleA: 5 OTUs

16S Analytical Methods

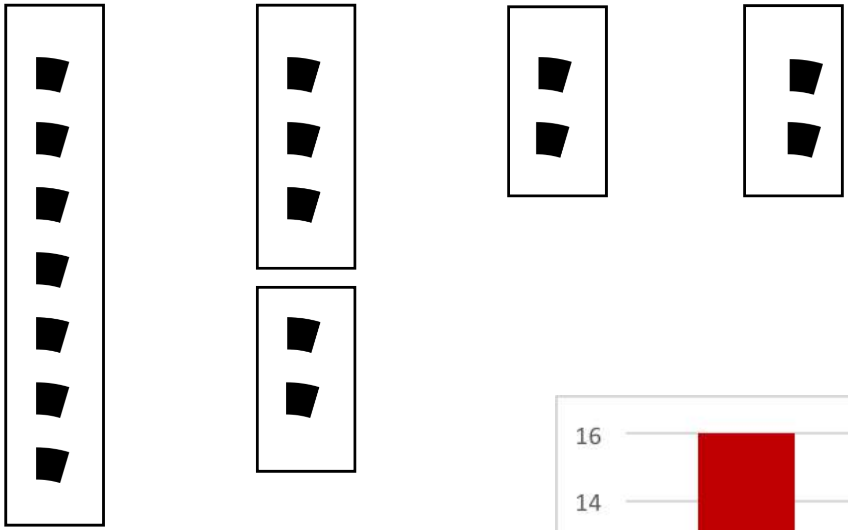
Just by grouping sequences into OTUs, we already have some information about the diversity of this sample:



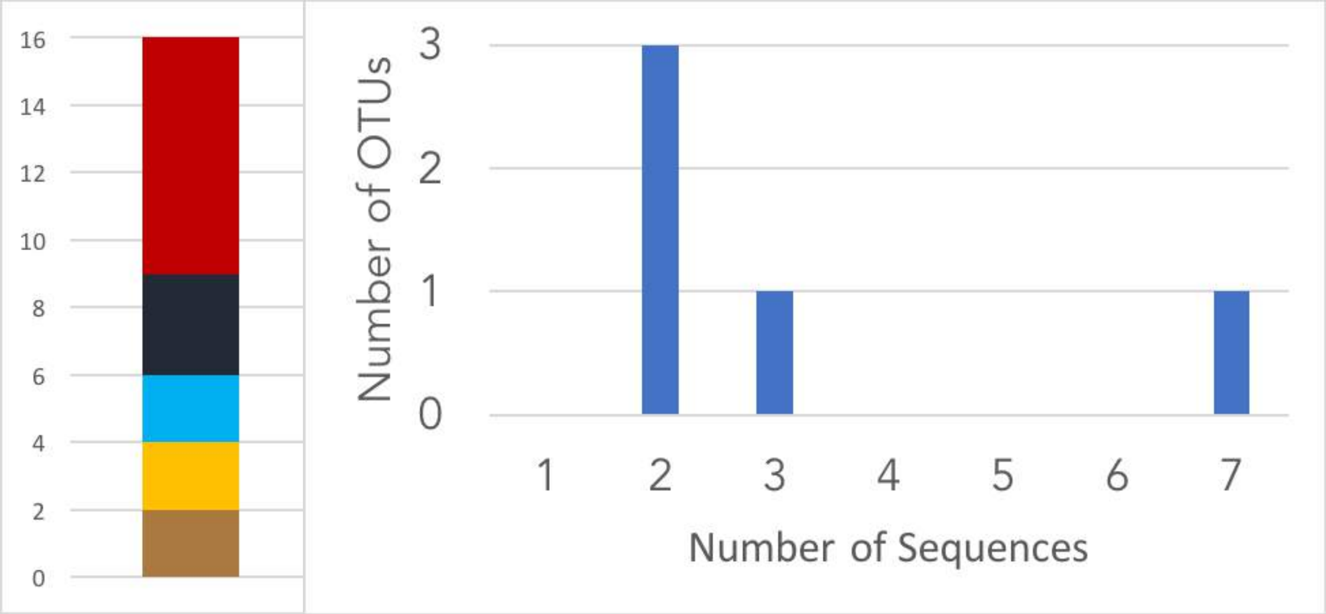
SampleA: 5 OTUs

16S Analytical Methods

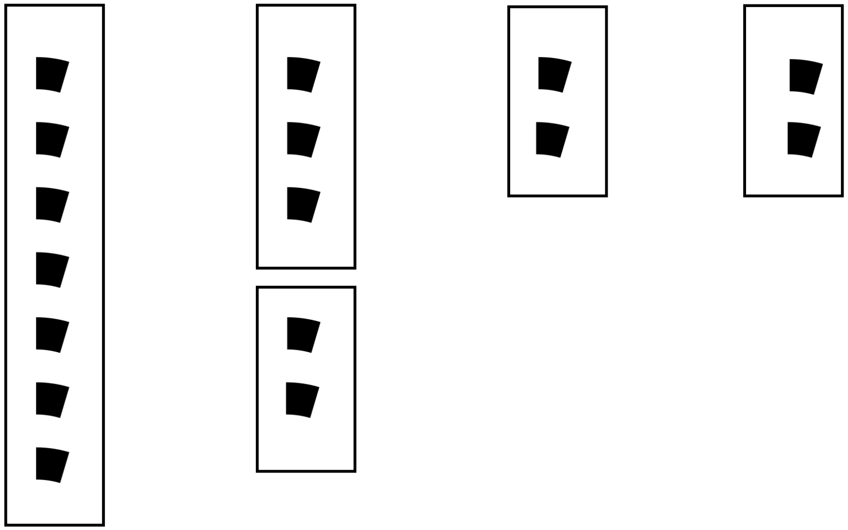
Just by grouping sequences into OTUs, we already have some information about the diversity of this sample:



SampleA: 5 OTUs



Alpha diversity



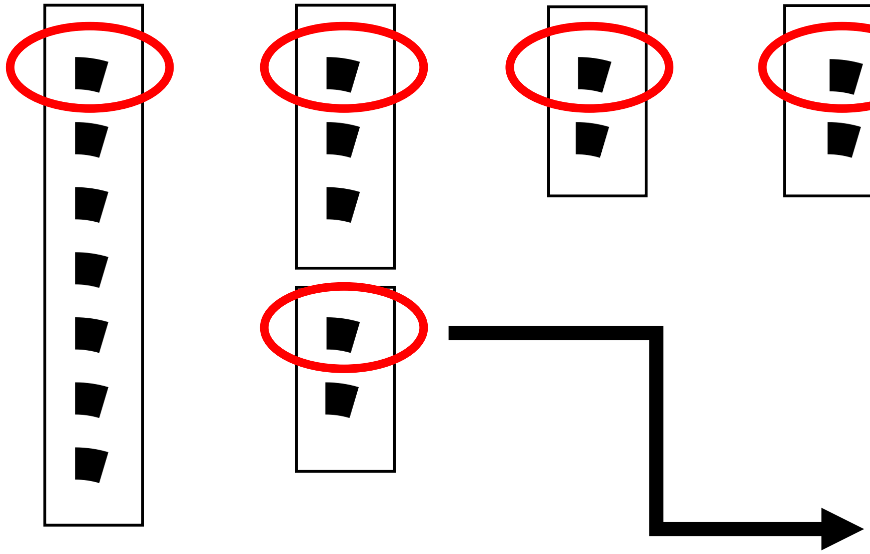
How many (and what) things are in a single community?

✓Richness (how many)

✓Evenness (how distributed / abundance)

- Composition (who)

Assign taxonomy



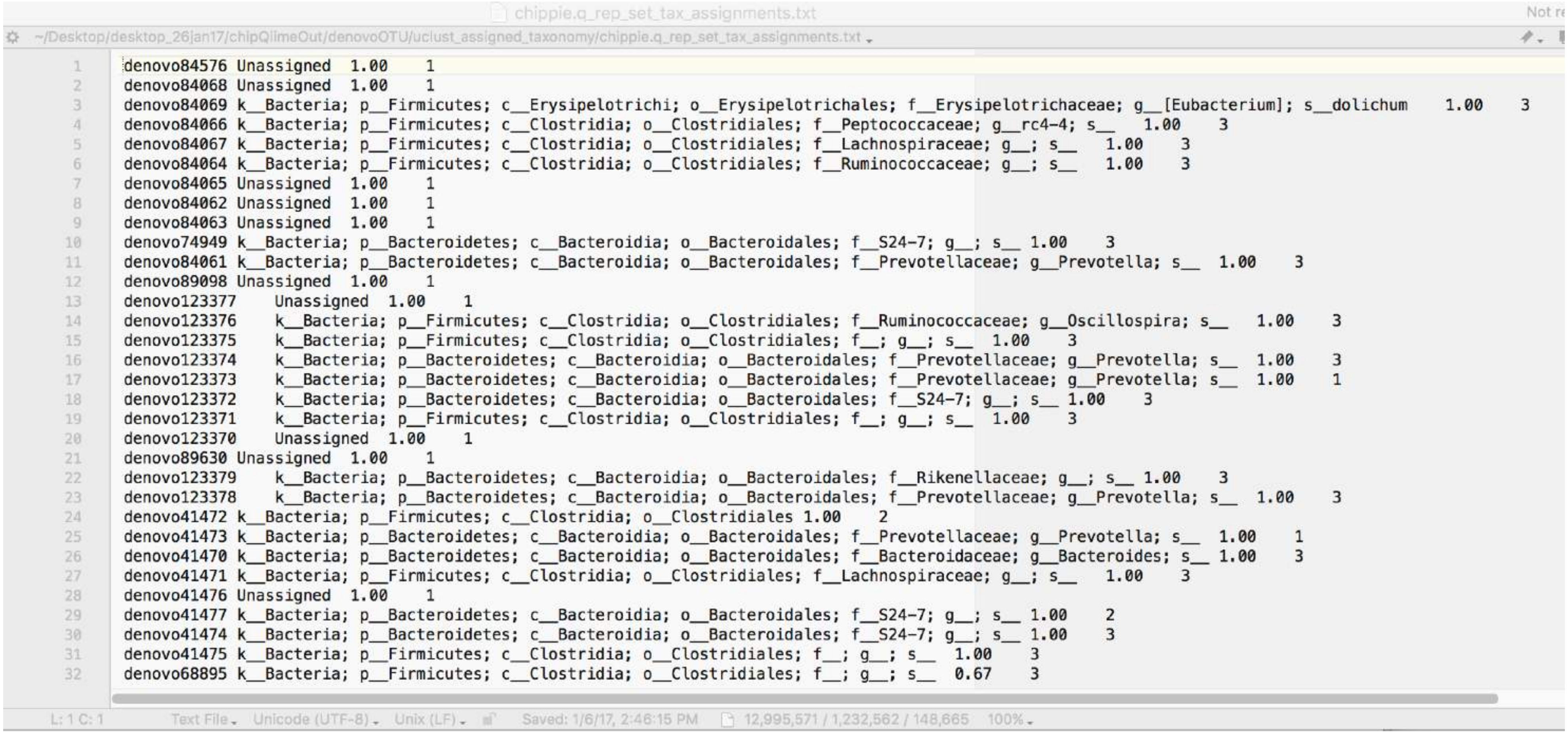
Use A Database (and a cut-off)

rdp

silva
high quality ribosomal RNA databases

GREENGENES
The 16S rRNA Gene Database and Tools

Assign taxonomy



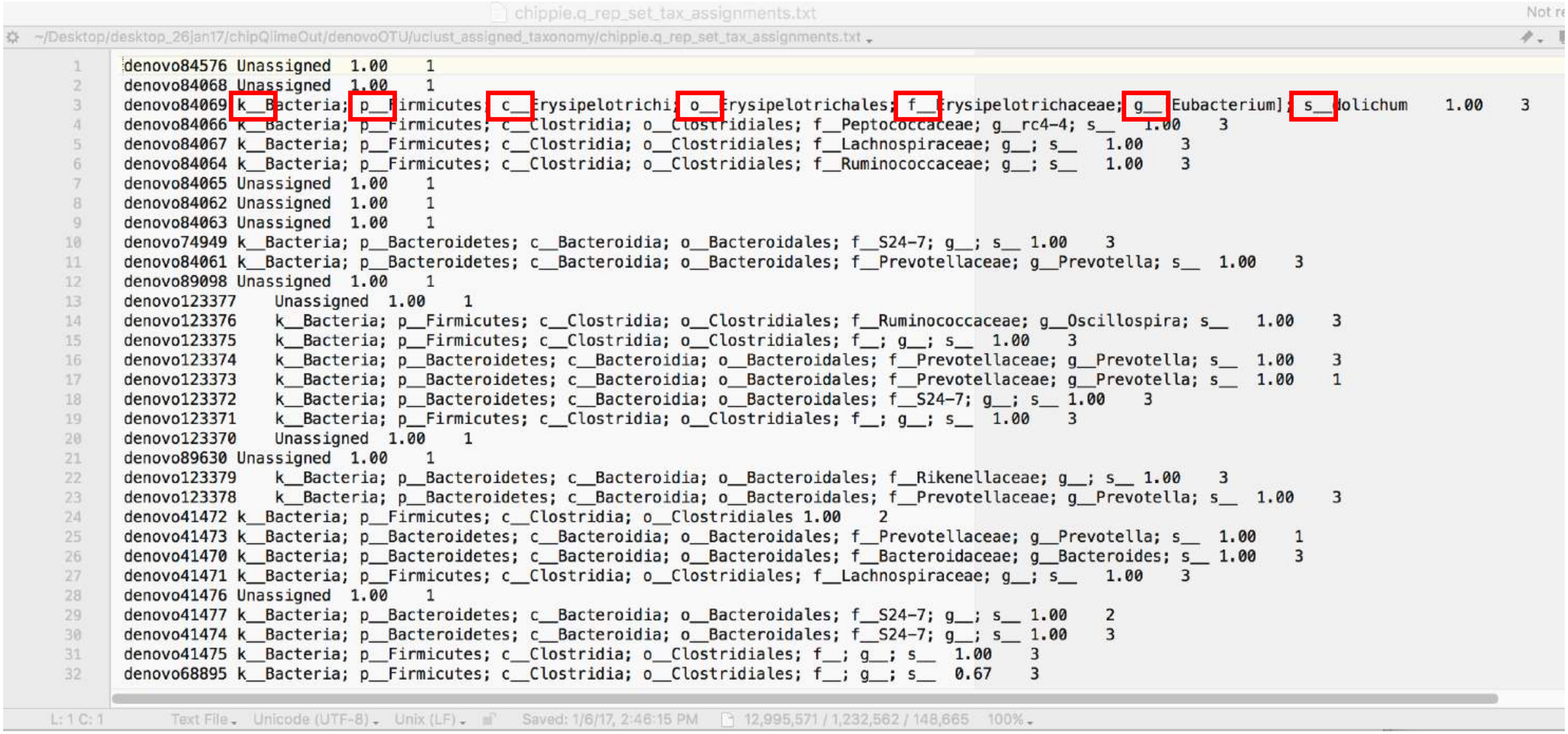
```
denovo84576 Unassigned  1.00    1
denovo84068 Unassigned  1.00    1
denovo84069 k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]; s__dolichum    1.00    3
denovo84066 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptococcaceae; g__rc4-4; s__   1.00    3
denovo84067 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo84064 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__   1.00    3
denovo84065 Unassigned  1.00    1
denovo84062 Unassigned  1.00    1
denovo84063 Unassigned  1.00    1
denovo74949 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo84061 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo89098 Unassigned  1.00    1
denovo123377    Unassigned  1.00    1
denovo123376    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__  1.00    3
denovo123375    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123374    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo123373    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo123372    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo123371    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123370    Unassigned  1.00    1
denovo89630 Unassigned  1.00    1
denovo123379    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__ 1.00    3
denovo123378    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo41472 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales 1.00    2
denovo41473 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo41470 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    3
denovo41471 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo41476 Unassigned  1.00    1
denovo41477 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    2
denovo41474 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo41475 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo68895 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
```

rdp

GREENGENES
The 16S rRNA Gene Database and Tools

silva
high quality ribosomal RNA databases

Assign taxonomy



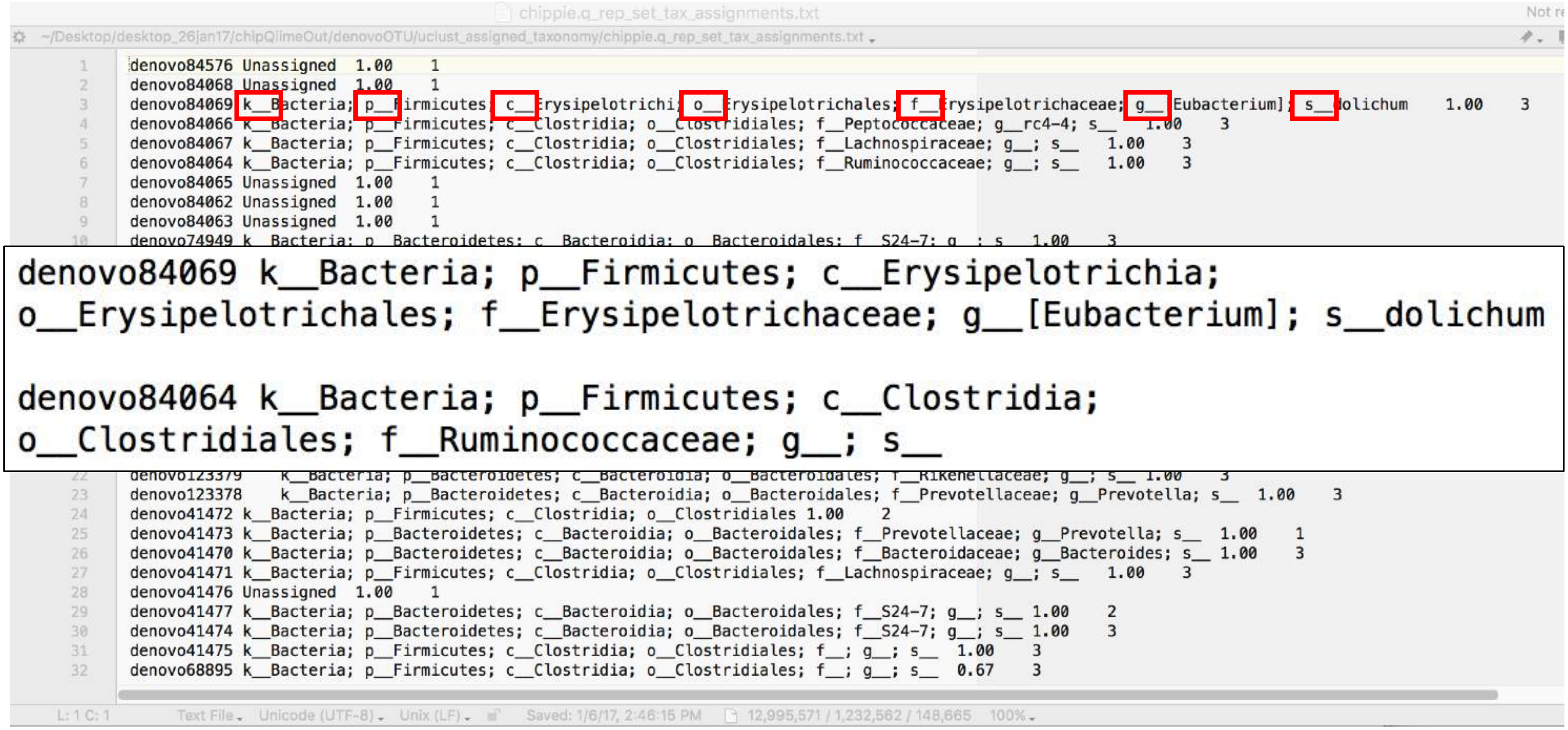
```
denovo84576 Unassigned  1.00    1
denovo84068 Unassigned  1.00    1
denovo84069 k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]; s__dolichum    1.00    3
denovo84066 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptococcaceae; g__rc4-4; s__   1.00    3
denovo84067 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo84064 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__   1.00    3
denovo84065 Unassigned  1.00    1
denovo84062 Unassigned  1.00    1
denovo84063 Unassigned  1.00    1
denovo74949 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo84061 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo89098 Unassigned  1.00    1
denovo123377    Unassigned  1.00    1
denovo123376    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__   1.00    3
denovo123375    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123374    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo123373    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo123372    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo123371    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123370    Unassigned  1.00    1
denovo89630 Unassigned  1.00    1
denovo123379    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__ 1.00    3
denovo123378    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo41472 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales 1.00    2
denovo41473 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo41470 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    3
denovo41471 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo41476 Unassigned  1.00    1
denovo41477 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    2
denovo41474 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo41475 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo68895 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
```

rdp

GREENGENES
The 16S rRNA Gene Database and Tools

silva
high quality ribosomal RNA databases

Assign taxonomy



chippie.q_rep_set_tax_assignments.txt

~/Desktop/desktop_26jan17/chipQiimeOut/denovoOTU/uclust_assigned_taxonomy/chippie.q_rep_set_tax_assignments.txt

```
denovo84576 Unassigned  1.00    1
denovo84068 Unassigned  1.00    1
denovo84069 k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]; s__dolichum    1.00    3
denovo84066 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptococcaceae; g__rc4-4; s__  1.00    3
denovo84067 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__  1.00    3
denovo84064 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__  1.00    3
denovo84065 Unassigned  1.00    1
denovo84062 Unassigned  1.00    1
denovo84063 Unassigned  1.00    1
denovo74949 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
```

```
denovo84069 k__Bacteria; p__Firmicutes; c__Erysipelotrichia;
o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]; s__dolichum

denovo84064 k__Bacteria; p__Firmicutes; c__Clostridia;
o__Clostridiales; f__Ruminococcaceae; g__; s__
```

```
denovo123379    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__ 1.00    3
denovo123378    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo41472 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales 1.00    2
denovo41473 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo41470 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    3
denovo41471 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__  1.00    3
denovo41476 Unassigned  1.00    1
denovo41477 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    2
denovo41474 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo41475 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo68895 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
```

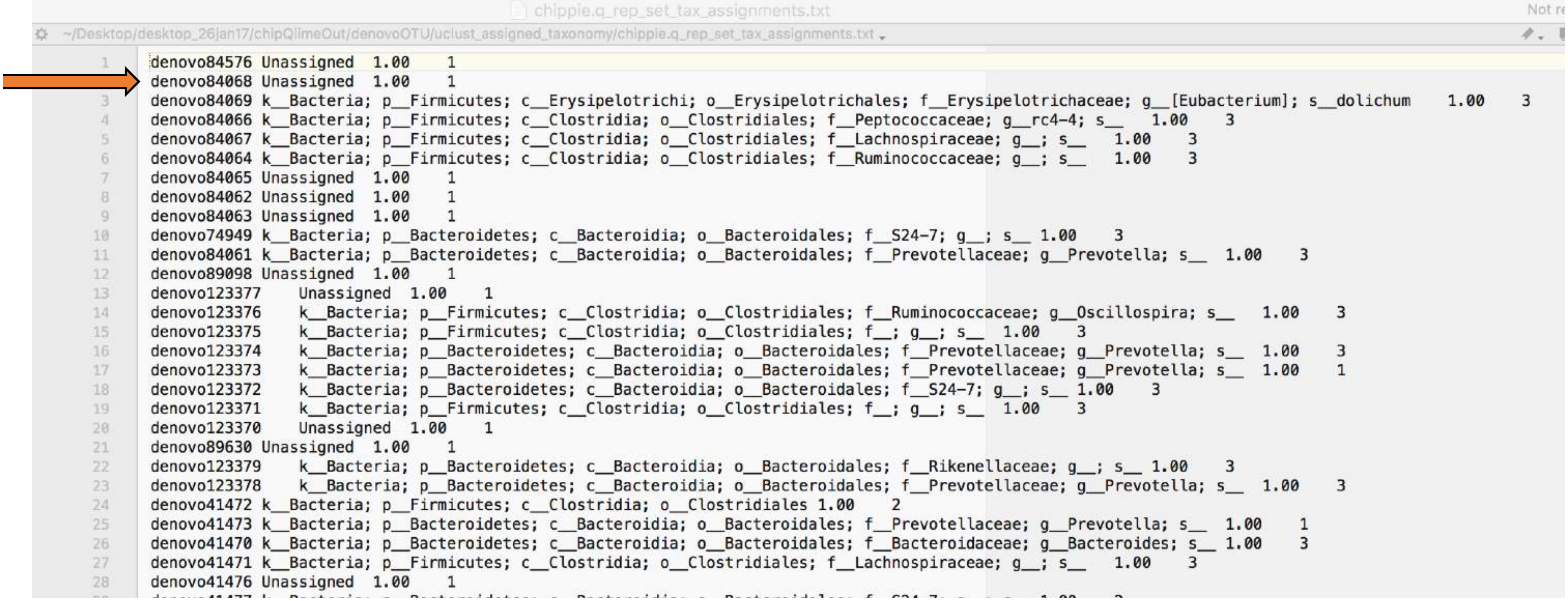
rdp

GREENGENES
The 16S rRNA Gene Database and Tools

silva
high quality ribosomal RNA databases

More QC: Throwing out data

Taxonomic assignment: Unassigned



```
denovo84576 Unassigned  1.00    1
denovo84068 Unassigned  1.00    1
denovo84069 k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]; s__dolichum    1.00    3
denovo84066 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptococcaceae; g__rc4-4; s__   1.00    3
denovo84067 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo84064 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__; s__   1.00    3
denovo84065 Unassigned  1.00    1
denovo84062 Unassigned  1.00    1
denovo84063 Unassigned  1.00    1
denovo74949 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo84061 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo89098 Unassigned  1.00    1
denovo123377    Unassigned  1.00    1
denovo123376    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__   1.00    3
denovo123375    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123374    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo123373    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo123372    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo123371    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo123370    Unassigned  1.00    1
denovo89630 Unassigned  1.00    1
denovo123379    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__; s__ 1.00    3
denovo123378    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
denovo41472 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales 1.00    2
denovo41473 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo41470 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    3
denovo41471 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo41476 Unassigned  1.00    1
```

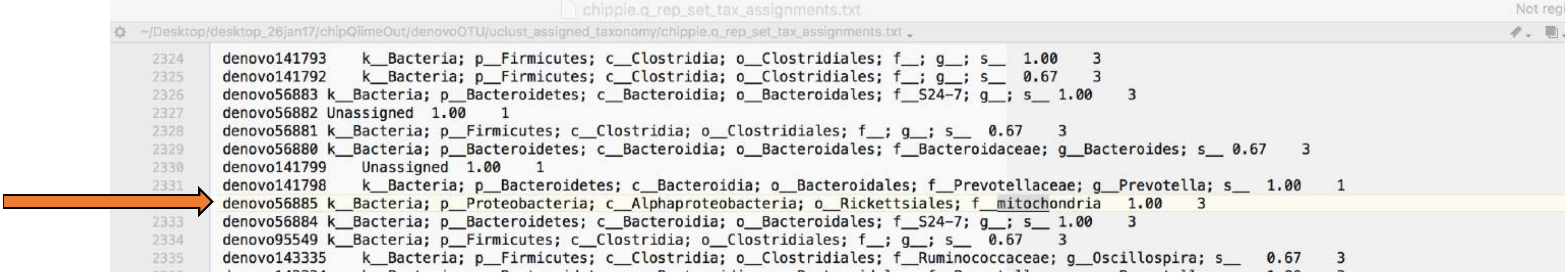
rdp

GREENGENES
The 16S rRNA Gene Database and Tools

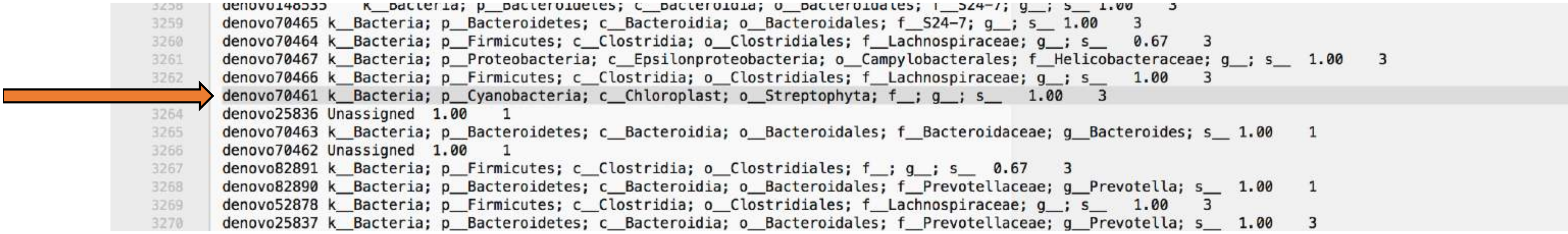
silva
high quality ribosomal RNA databases

More QC: Throwing out data

Taxonomic results: Mitochondria and Chloroplast



```
denovo141793    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo141792    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
denovo56883 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo56882 Unassigned  1.00    1
denovo56881 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
denovo56880 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 0.67    3
denovo141799    Unassigned  1.00    1
denovo141798    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo56885 k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria   1.00    3
denovo56884 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo95549 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
denovo143335    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__   0.67    3
```



```
denovo148535    k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo70465 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__; s__ 1.00    3
denovo70464 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   0.67    3
denovo70467 k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__; s__  1.00    3
denovo70466 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo70461 k__Bacteria; p__Cyanobacteria; c__Chloroplast; o__Streptophyta; f__; g__; s__   1.00    3
denovo25836 Unassigned  1.00    1
denovo70463 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    1
denovo70462 Unassigned  1.00    1
denovo82891 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
denovo82890 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo52878 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo25837 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
```

rdp

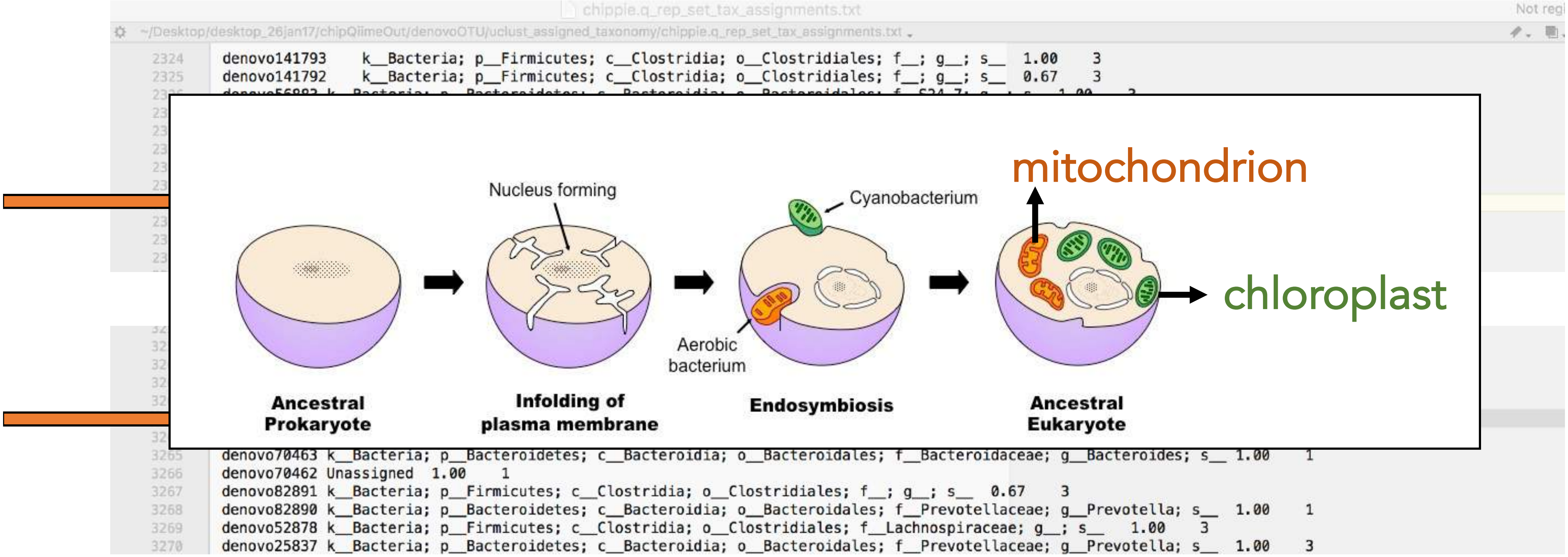
GREENGENES
The 16S rRNA Gene Database and Tools

silva
high quality ribosomal RNA databases

@sarahmhird

More QC: Throwing out data

Taxonomic results: Mitochondria and Chloroplast



```
denovo141793    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  1.00    3
denovo141792    k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
```

```
denovo70463 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides; s__ 1.00    1
denovo70462 Unassigned  1.00    1
denovo82891 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__; s__  0.67    3
denovo82890 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    1
denovo52878 k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__   1.00    3
denovo25837 k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella; s__  1.00    3
```

rdp

GREENGENES
The 16S rRNA Gene Database and Tools

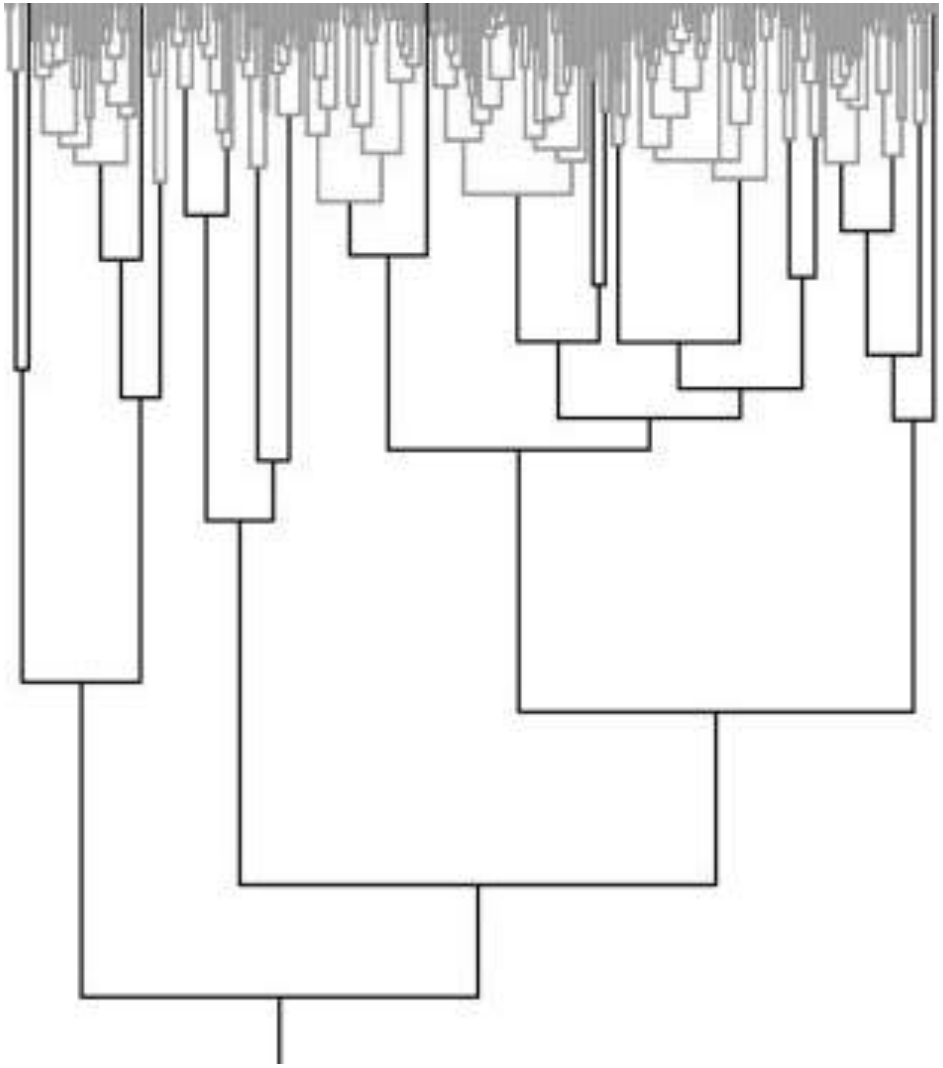
silva
high quality ribosomal RNA databases

Alpha diversity

- Richness (how many)
- Evenness (how distributed / abundance)
- Composition (who)
- Phylogenetic diversity

Phylogenetic diversity

Faith's phylogenetic diversity = minimum total branch length that includes all OTUs



Other alpha diversity metrics

- Chao1 = richness estimate; uses singletons to account for undiscovered species; how likely are undiscovered species?

- Can also use abundance information to estimate diversity

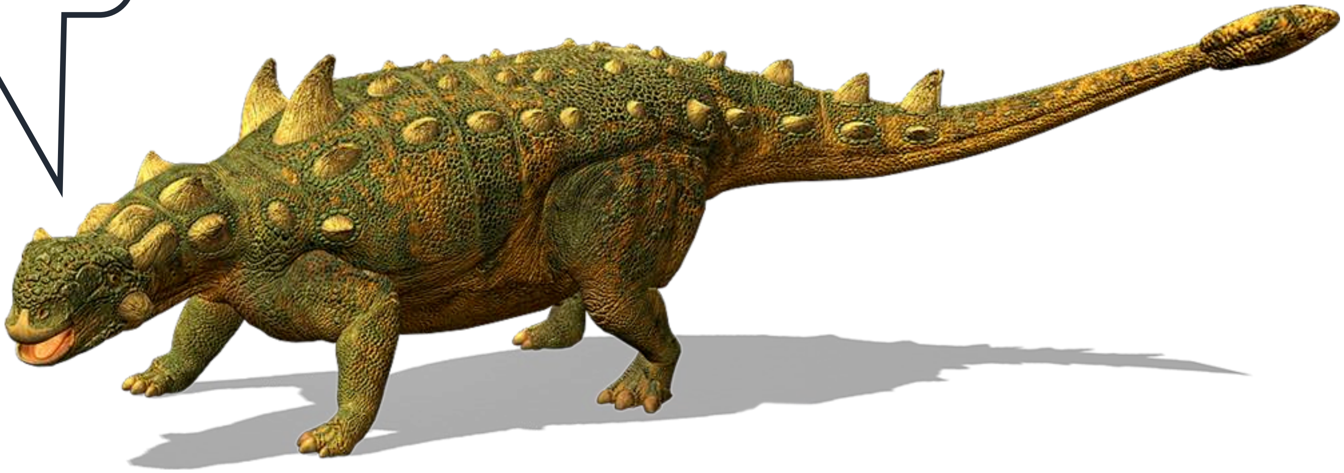
$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

Shannon Diversity Index = complicated equation, accounts for evenness; how likely are you to predict the next species found?

$$l = \frac{\sum_{i=1}^{R} n_i(n_i - 1)}{N(N-1)}$$

Simpson’s Index = simple equation, accounts for evenness; what is the probability that two sequences belong to the same species?

Any questions?

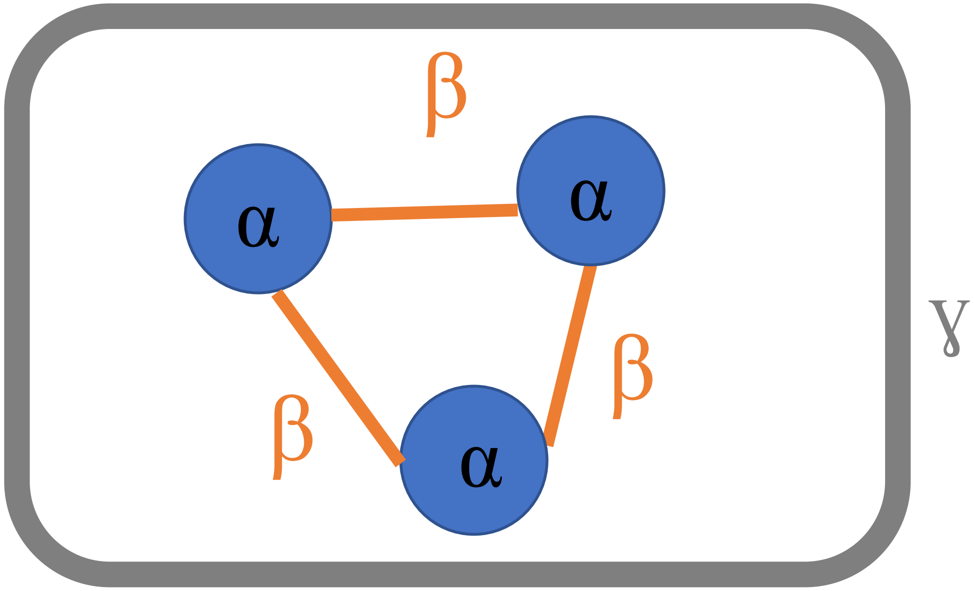


Euoplocephalus

walkingwith.wikia.com

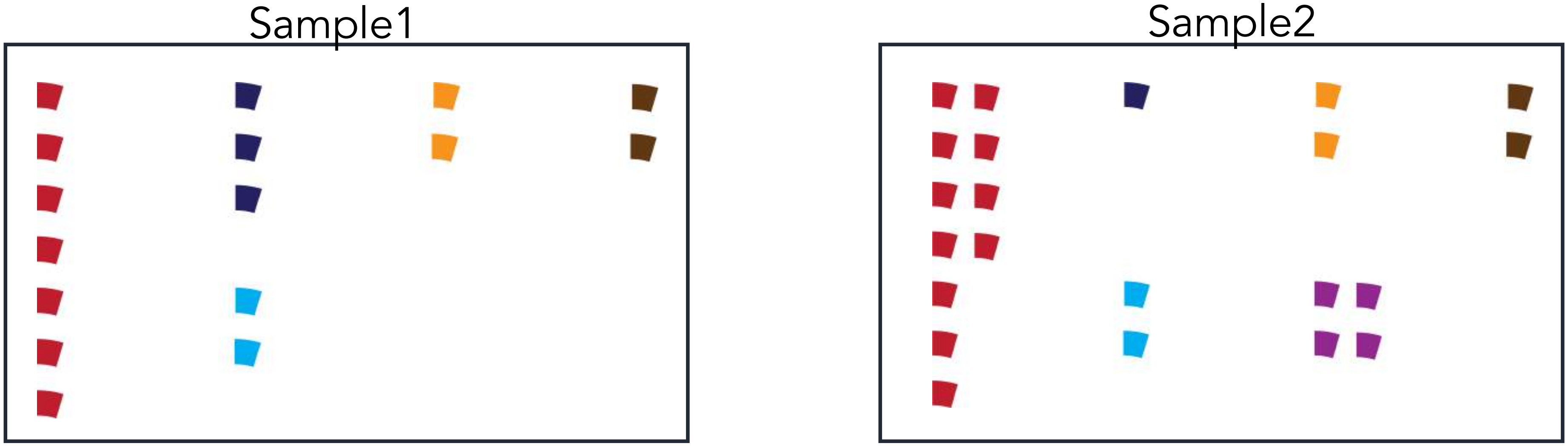
Measures of species diversity

- ✓Alpha diversity: diversity within a site
- Beta diversity: diversity between sites
- Gamma diversity: total diversity of a region



Beta diversity

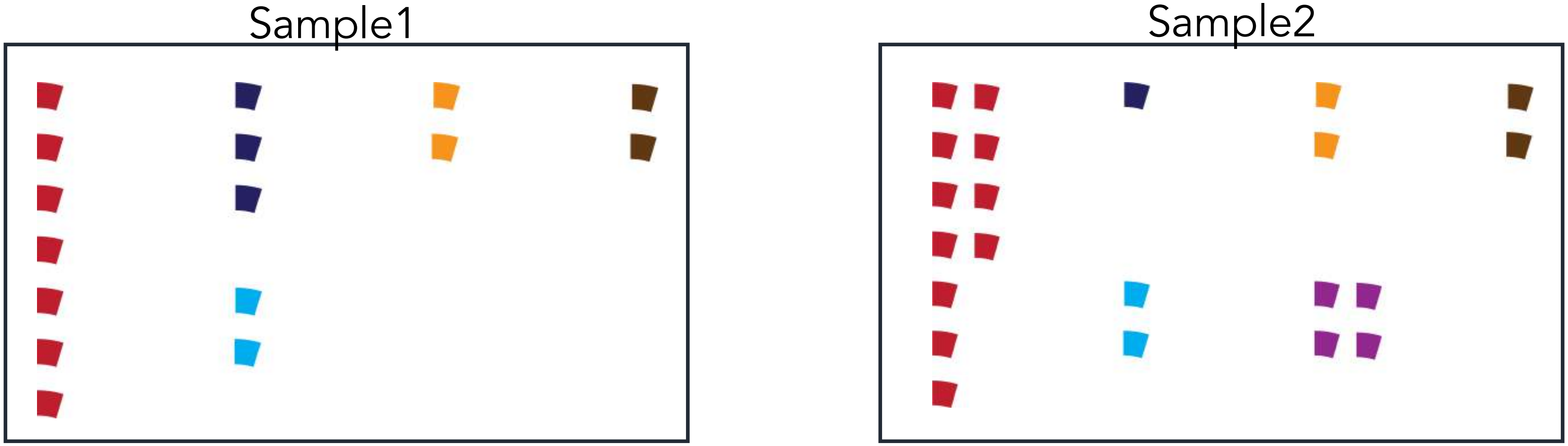
- How similar are these communities?



Beta diversity

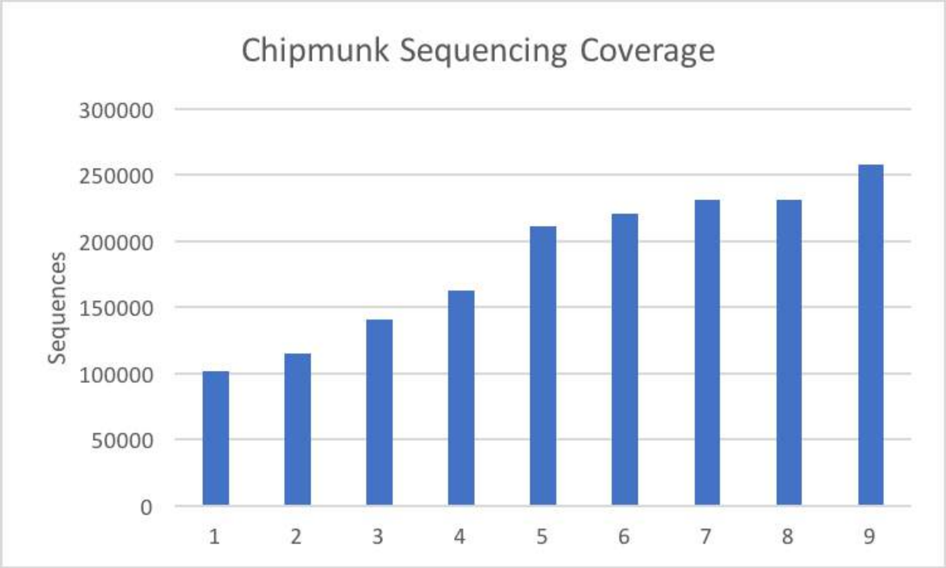
- How similar are these communities?

- Usually need <u>equal sampling effort</u>.
  - Rarefy the data.



Rarefying data

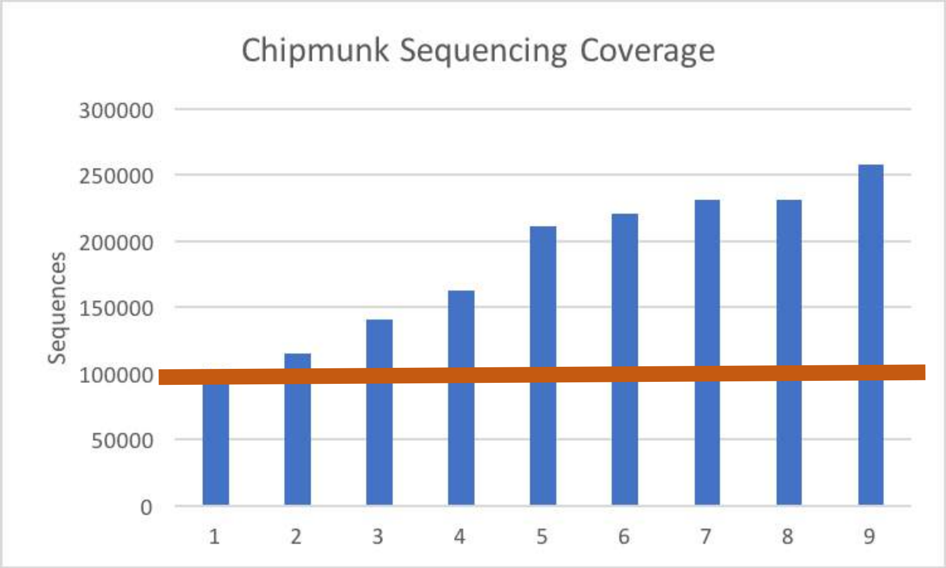
Rarefying refers to making all samples in a dataset have the same number of sequences.



"We rarefied our dataset to 101242 reads, which is the lowest sequence coverage in our dataset."

Rarefying data

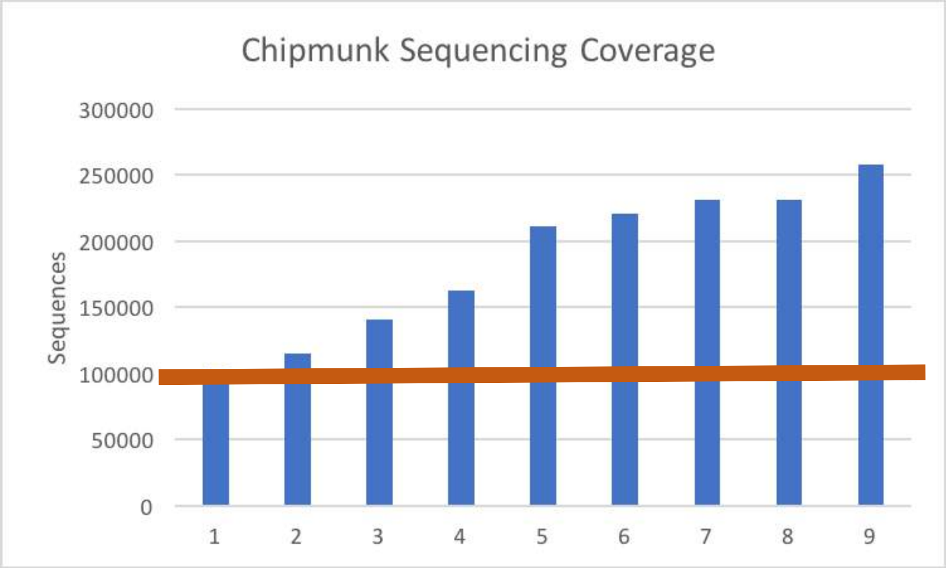
Rarefying refers to making all samples in a dataset have the same number of sequences.



"We rarefied our dataset to 101242 reads, which is the lowest sequence coverage in our dataset."

Rarefying data

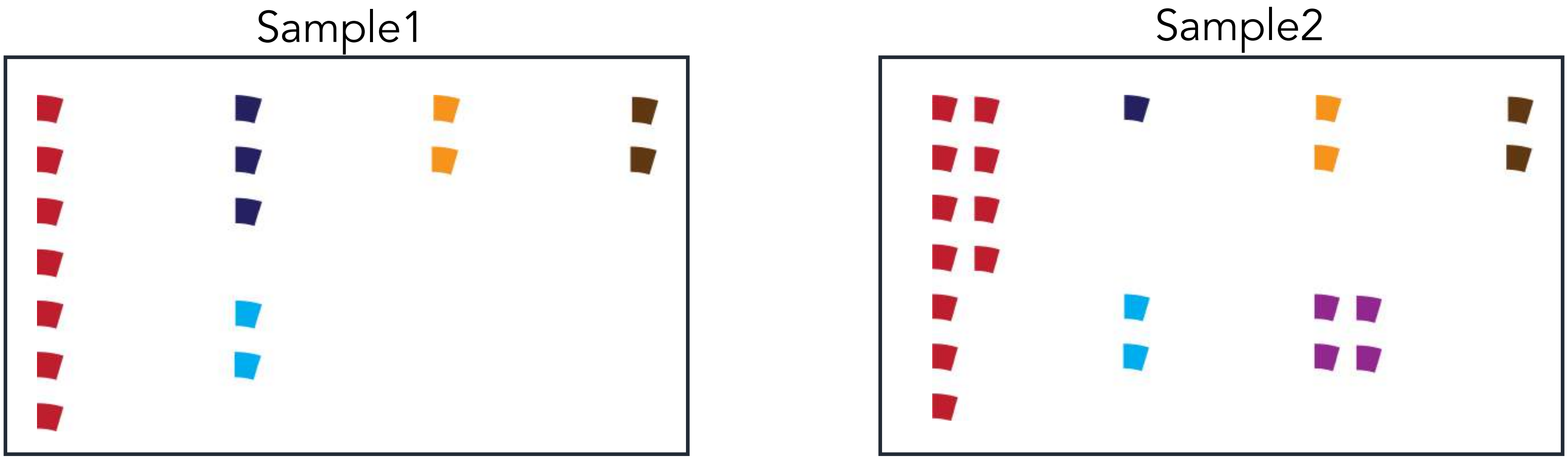
Rarefying refers to making all samples in a dataset have the same number of sequences.



Note: it is painful to throw away good data.

Beta diversity

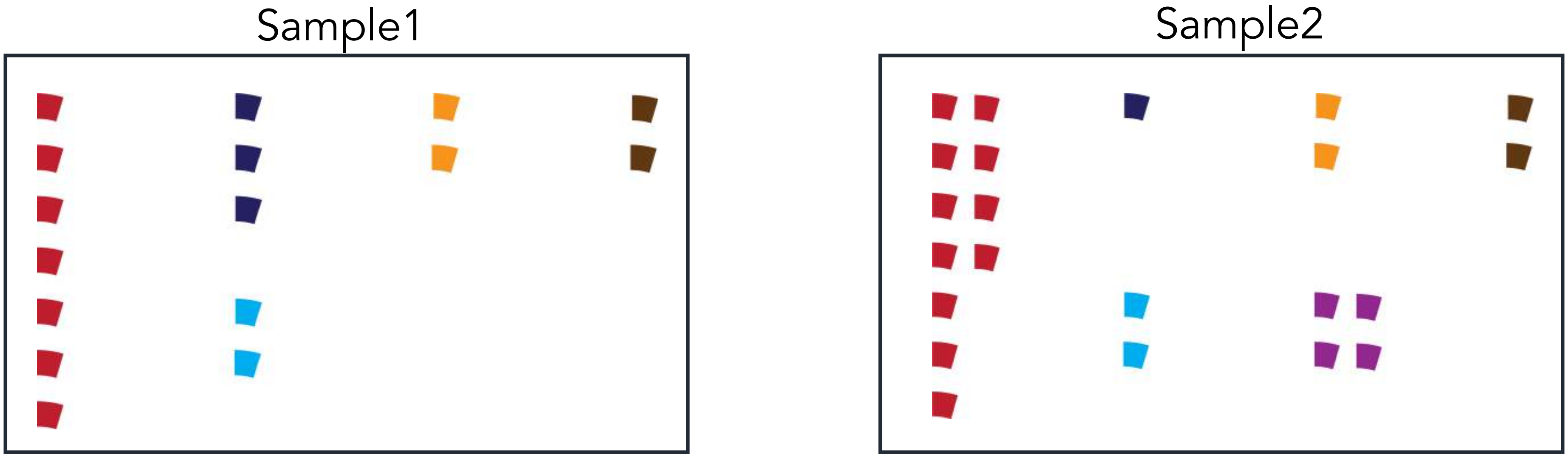
- Non-phylogenetic metrics
  - Bray-Curtis
  - Jaccard
- Phylogenetic metrics
  - UniFrac
  - KR Distance



Beta diversity

- Non-phylogenetic metrics
  - Bray-Curtis
  - Jaccard
- Phylogenetic metrics
  - UniFrac
  - KR Distance

	Sample1	Sample2
dark red	7	11
dark blue	3	1
light blue	2	2
orange	2	2
brown	2	2
purple	0	4



Bray-Curtis Dissimilarity

- Based on counts in each sample
- Values vary from 0 (identical) to 1 (no overlap)

C_{ij} = sum of the lesser value of all common species

$$\boxed{BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}}$$

S_i = total number of individuals at site I
S_j = total number of individuals at site j

S_i+S_j = total observed individuals

	Sample1	Sample2
dark red	7	11
dark blue	3	1
light blue	2	2
orange	2	2
brown	2	2
purple	0	4

Bray-Curtis Dissimilarity

- Based on <u>counts</u> in each sample
- Values vary from 0 (identical) to 1 (no overlap)

C_{ij} = sum of the lesser value of all common species

$$\boxed{BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}}$$

S_i = total number of individuals at site I
S_j = total number of individuals at site j

S_i+S_j = total observed individuals

	Sample1	Sample2
dark red	7	11
dark blue	3	1
light blue	2	2
orange	2	2
brown	2	2
purple	0	4

16 22

Bray-Curtis Dissimilarity

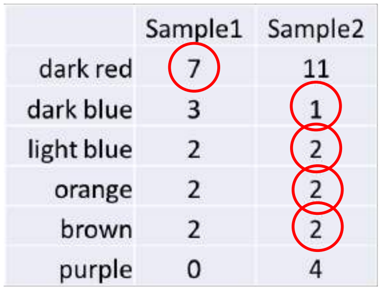
- Based on <u>counts</u> in each sample
- Values vary from 0 (identical) to 1 (no overlap)

C_{ij} = sum of the lesser value of all common species

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$$

S_i = total number of individuals at site I
S_j = total number of individuals at site j

$S_i + S_j$ = total observed individuals



	Sample1	Sample2
dark red	7	11
dark blue	3	1
light blue	2	2
orange	2	2
brown	2	2
purple	0	4

16 22

○ = 14

Bray-Curtis Dissimilarity

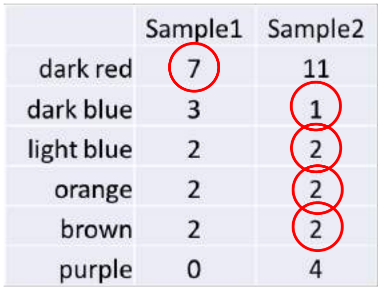
- Based on <u>counts</u> in each sample
- Values vary from 0 (identical) to 1 (no overlap)

C_{ij} = sum of the lesser value of all common species

$$\boxed{BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}}$$

S_i = total number of individuals at site I
S_j = total number of individuals at site j

$S_i + S_j$ = total observed individuals



	Sample1	Sample2
dark red	7	11
dark blue	3	1
light blue	2	2
orange	2	2
brown	2	2
purple	0	4

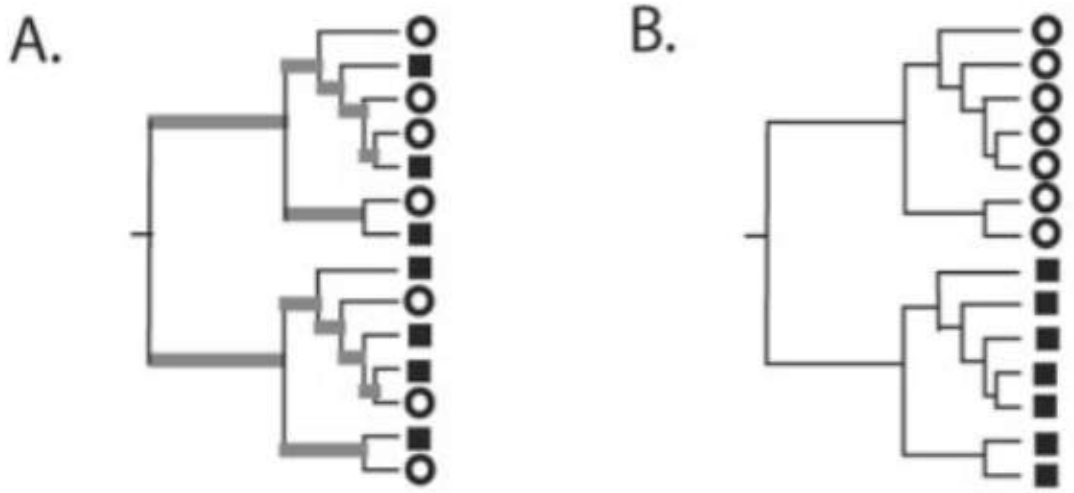
16 22

BC = 1 – [(2 * 14) / (16 + 22)]
BC = 1 – [28 / 38]
BC = 0.2631

○ = 14

UniFrac distances

Phylogenetic metric. Calculates the
UNIque FRACtion
of the tree belonging to each community

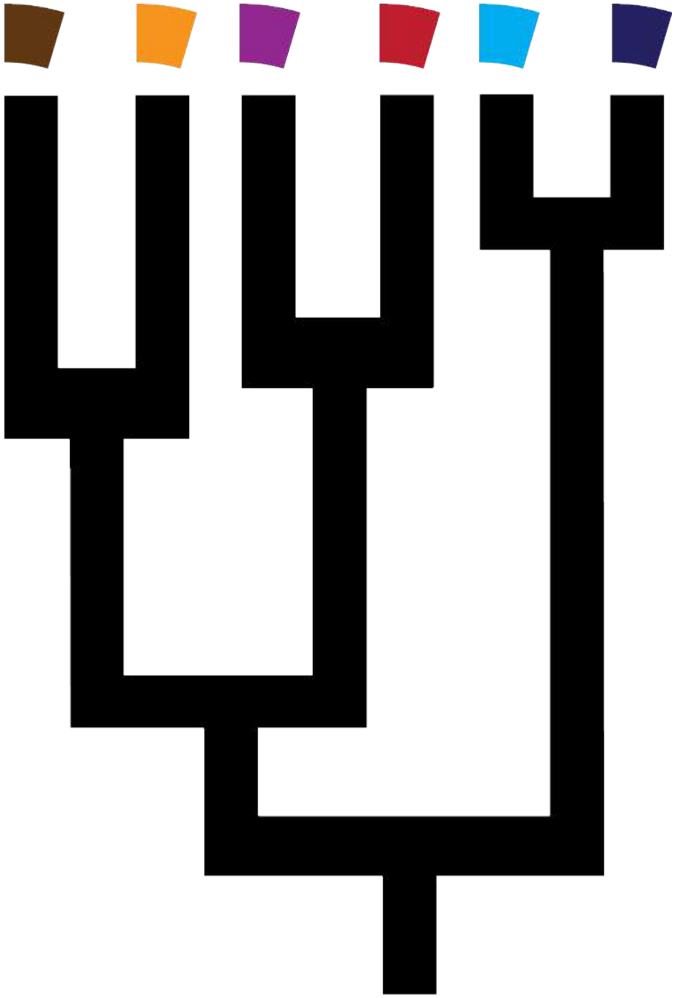


UF = ~0.5

UF = 1.0

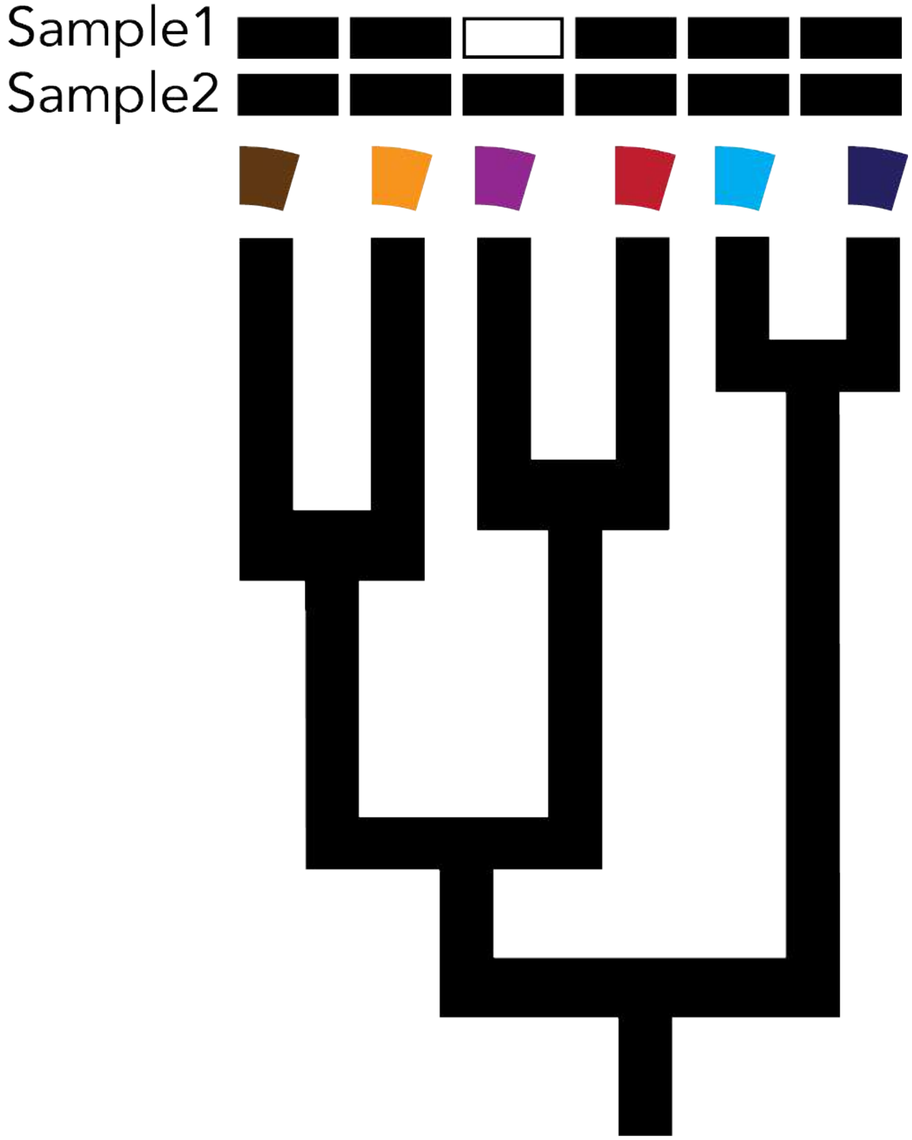
Lozupone and Knight. 2005. *Applied and Environmental Microbiology*. UniFrac: a new phylogenetic method for comparing microbial communities.

UniFrac distances



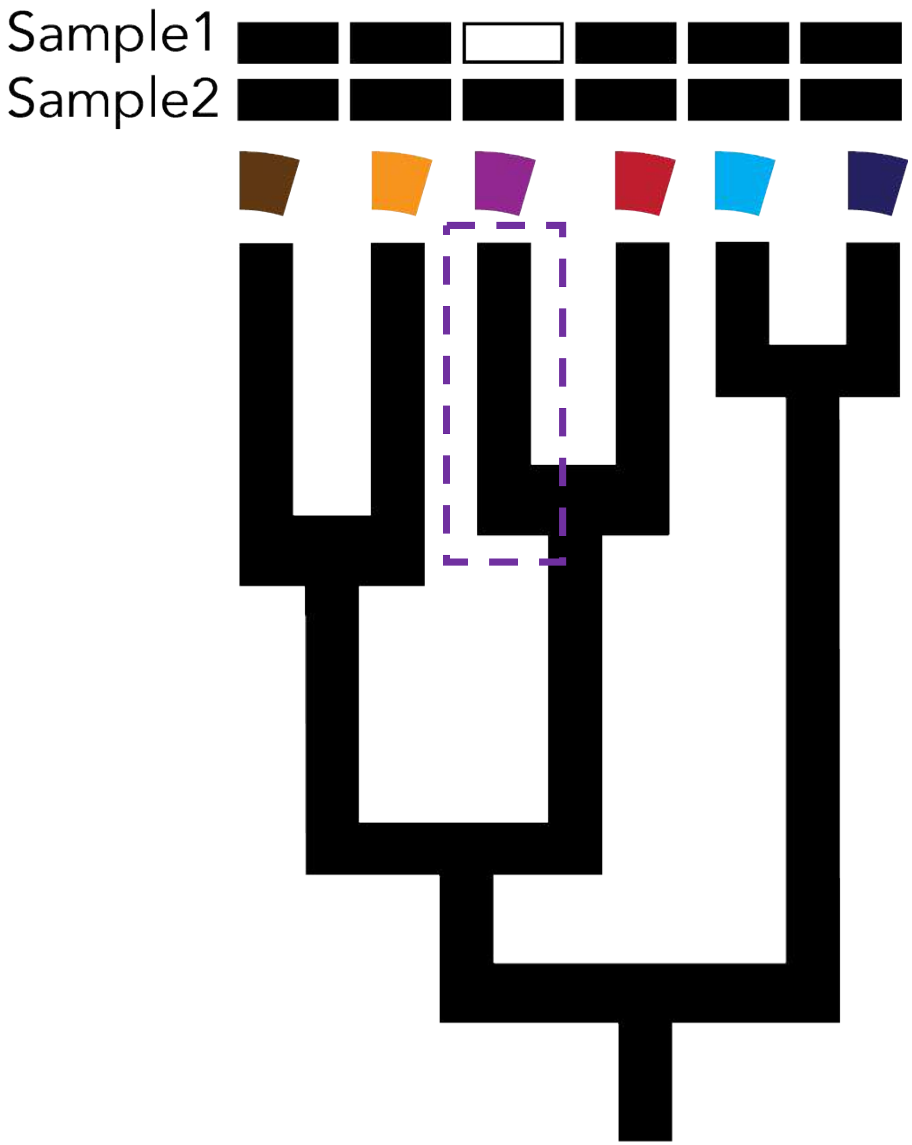
UniFrac distances

Unweighted: sum of branch lengths unique to one environment or the other based on presence/absence of lineages in a community



UniFrac distances

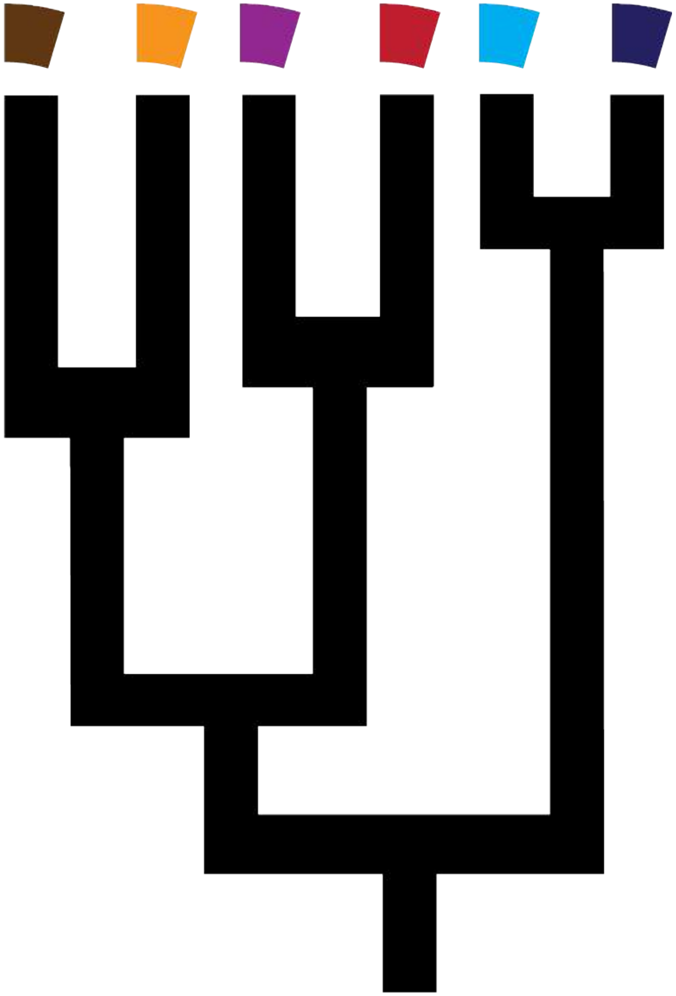
Unweighted: sum of branch lengths unique to one environment or the other based on presence/absence of lineages in a community



UniFrac distances

Weighted: sum of branch lengths weighted by the relative abundance of the sequences

Sample1	0.125	0.125	0	0.4375	0.125	0.1875
Sample2	0.090	0.090	0.182	0.5	0.090	0.045



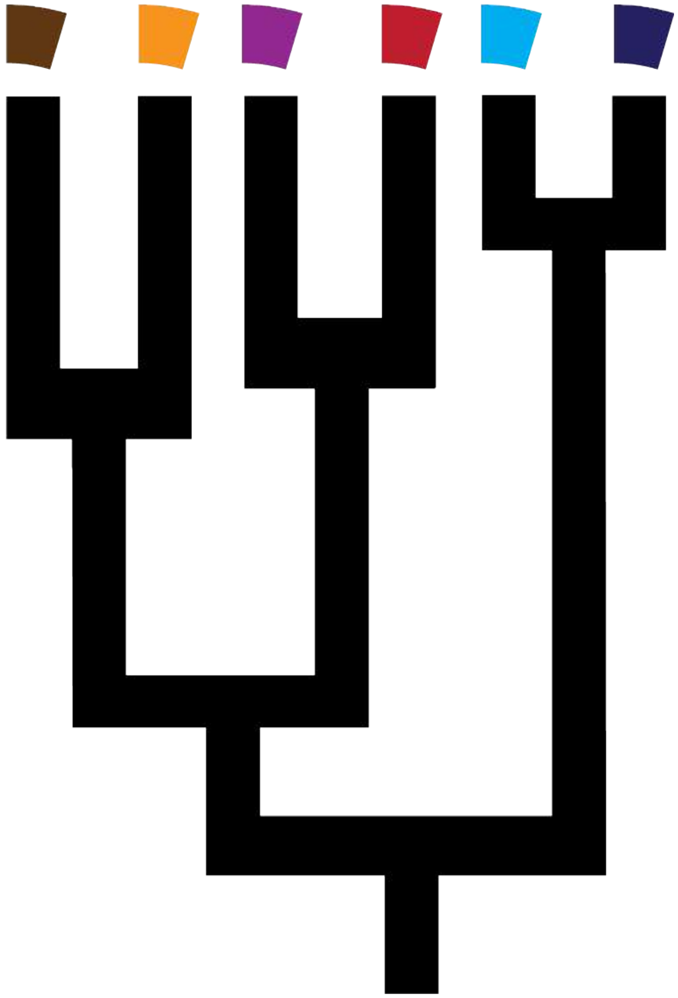
UniFrac distances

Weighted: sum of branch lengths weighted by the relative abundance of the sequences

$$W = \frac{\sum_{i=1}^{N} li \left| \frac{A_i}{A_T} - \frac{B_i}{B_T} \right|}{\sum_{j=1}^{S} L_j}$$

N=nodes
S = total sequences
li = branch length of i
Lj = total branch length
Ai and Bi = sequences at leaf i from A or B
AT and BT = total sequences from A or B

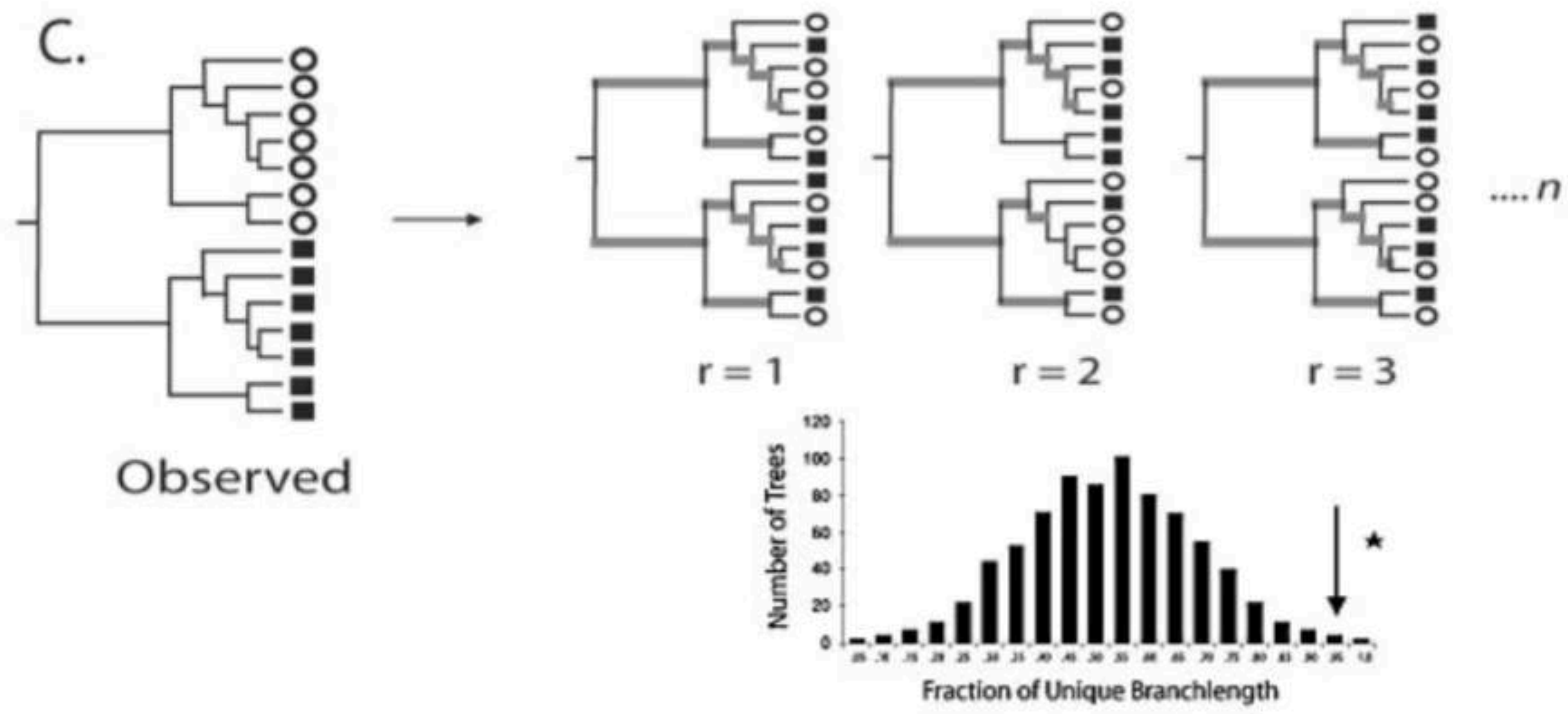
Sample1	0.125	0.125	0	0.4375	0.125	0.1875
Sample2	0.090	0.090	0.182	0.5	0.090	0.045



mothur.org

UniFrac distances

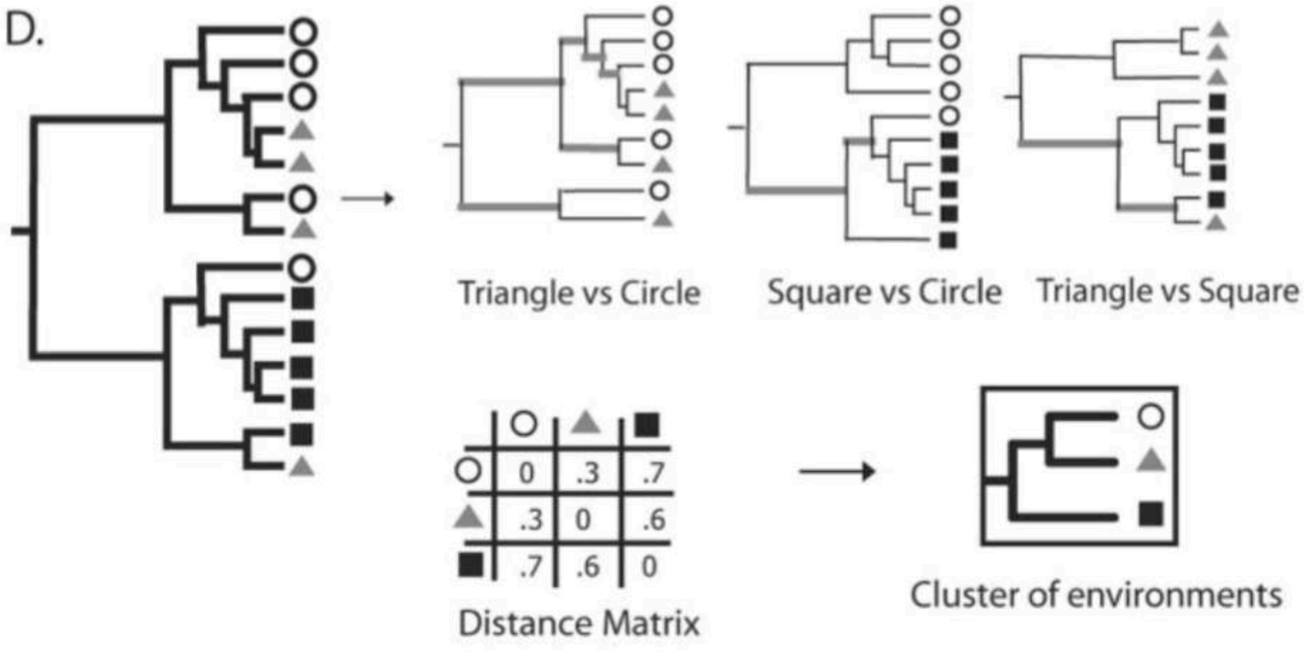
Calculates the
UNIque FRACtion
of the tree belonging to each community



Lozupone and Knight. 2005. *Applied and Environmental Microbiology*. UniFrac: a new phylogenetic method for comparing microbial communities.

UniFrac distances

Calculates the
UNIque FRACtion
of the tree belonging to each community



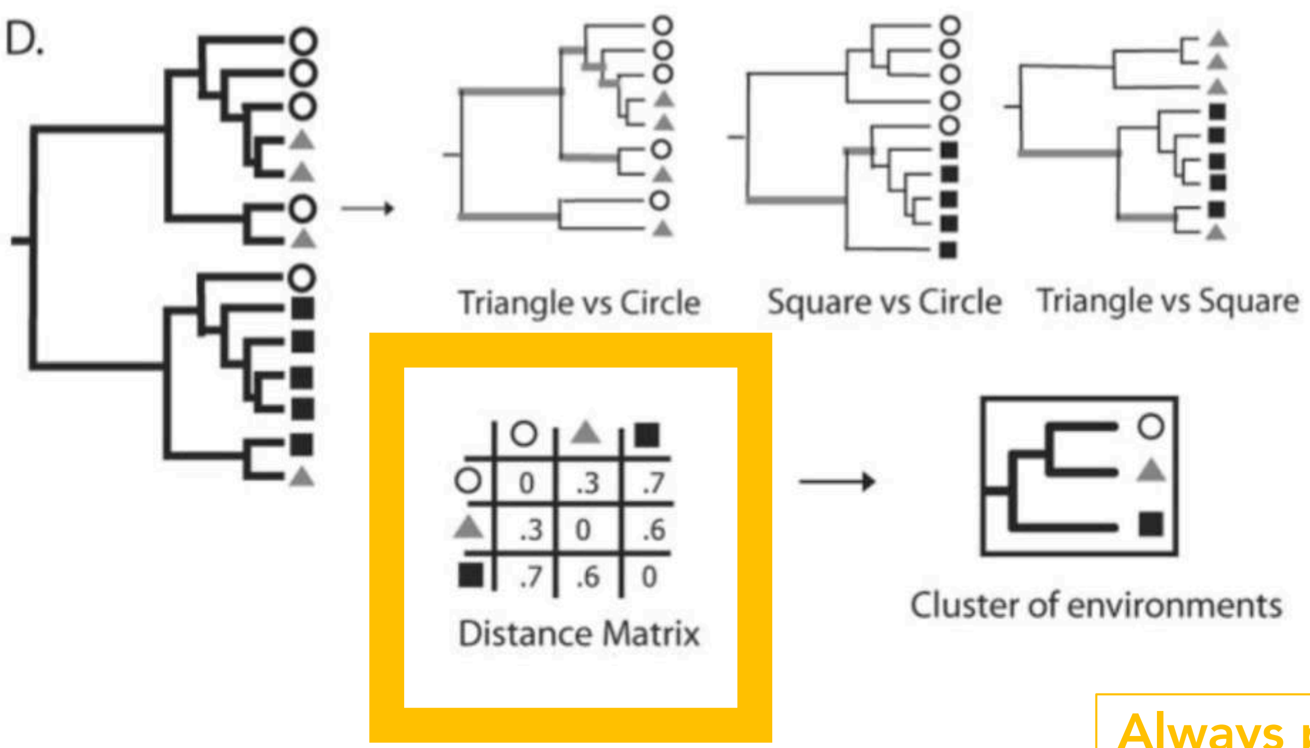
	○	▲	■
○	0	.3	.7
▲	.3	0	.6
■	.7	.6	0

Always pairwise

Lozupone and Knight. 2005. *Applied and Environmental Microbiology*. UniFrac: a new phylogenetic method for comparing microbial communities.

UniFrac distances

Calculates the
UNIque FRACtion
of the tree belonging to each community



	○	▲	■
○	0	.3	.7
▲	.3	0	.6
■	.7	.6	0

Always pairwise

Lozupone and Knight. 2005. *Applied and Environmental Microbiology*. UniFrac: a new phylogenetic method for comparing microbial communities.

Comparing data

Distance matrix:

	WATER_1	WATER_2	WATER_3	SEDIMENT_1	SEDIMENT_2	SEDIMENT_3
WATER_1	-	0.34	0.36	0.75	0.76	0.70
WATER_2		-	0.42	0.77	0.76	0.71
WATER_3			-	0.80	0.79	0.81
SEDIMENT_1				-	0.47	0.50
SEDIMENT_2					-	0.45
SEDIMENT_3						-

Comparing data

Distance matrix:

	WATER_1	WATER_2	WATER_3	SEDIMENT_1	SEDIMENT_2	SEDIMENT_3
WATER_1	-	0.34	0.36	0.75	0.76	0.70
WATER_2		-	0.42	0.77	0.76	0.71
WATER_3			-	0.80	0.79	0.81
SEDIMENT_1				-	0.47	0.50
SEDIMENT_2					-	0.45
SEDIMENT_3						-

How do we synthesize and visualize these data?

- Clustering
- “Trees”
- Categories
- Networks
- Significant OTUs

Comparing data

Distance matrix:

	WATER_1	WATER_2	WATER_3	SEDIMENT_1	SEDIMENT_2	SEDIMENT_3
WATER_1	-	0.34	0.36	0.75	0.76	0.70
WATER_2		-	0.42	0.77	0.76	0.71
WATER_3			-	0.80	0.79	0.81
SEDIMENT_1				-	0.47	0.50
SEDIMENT_2					-	0.45
SEDIMENT_3						-

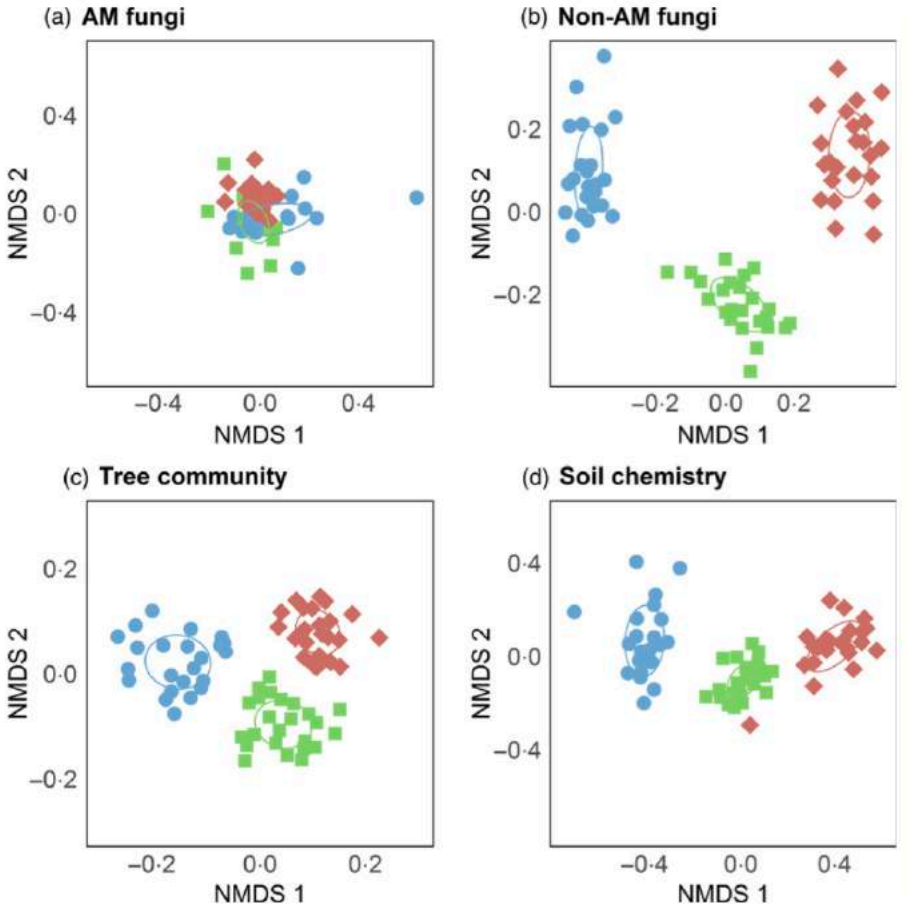
How do we synthesize and visualize these data?

- Clustering
- "Trees"
- Categories
- Networks
- Significant OTUs

A lot of these apply to shotgun metagenomic data too…or any distance matrix.

Clustering (or Ordination)

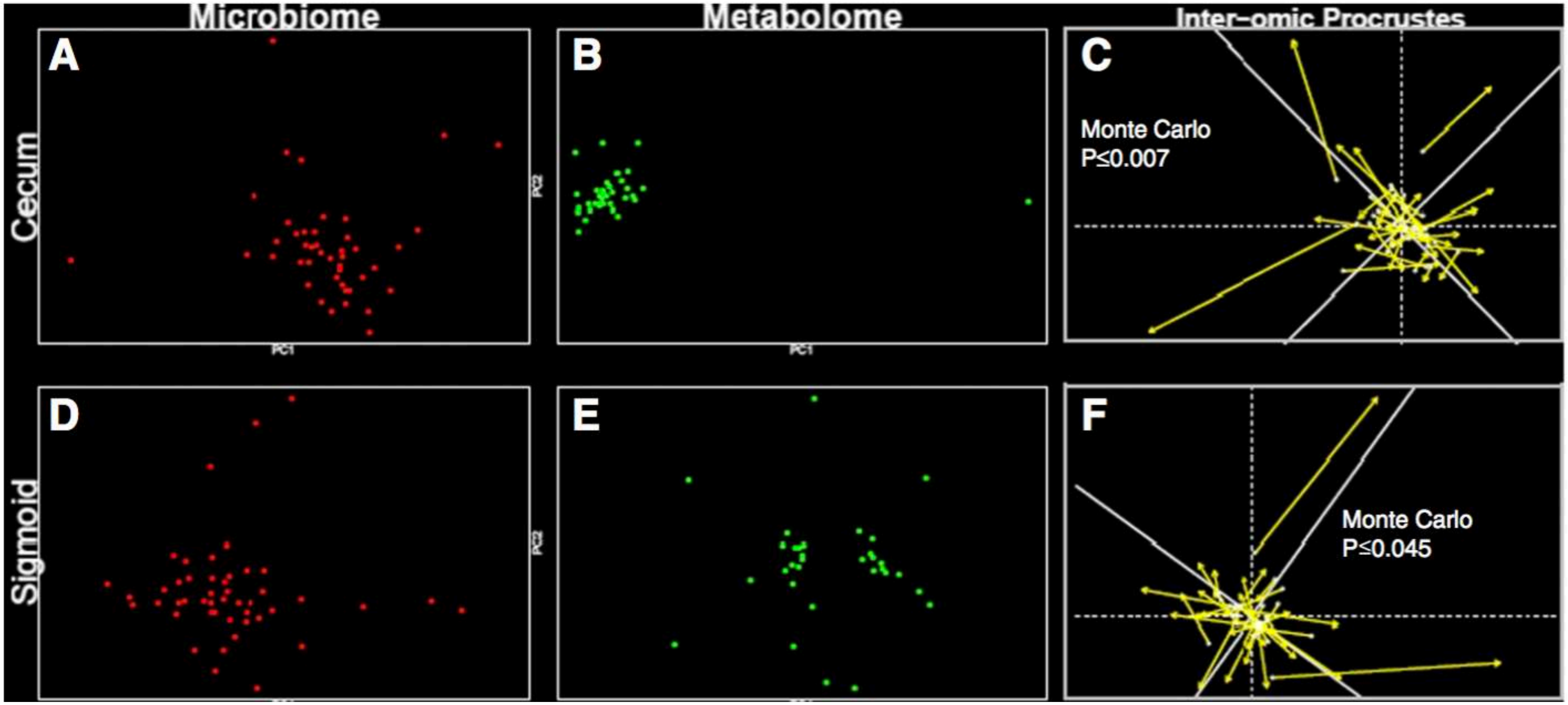
- NMDS (Nonmetric multidimensional scaling)
- PCA (Principle Components Analysis)
- PCoA (Principle Coordinates Analysis)



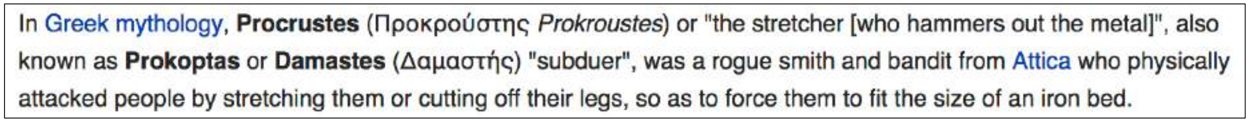
Schappe et al. 2017. Journal of Ecology. The role of soil chemistry and plant neighbourhoods in structuring fungal communities in three Panamanian rainforests.

Comparing ordinations: Procrustes rotation

Do we draw the same conclusions using different methods/metrics?



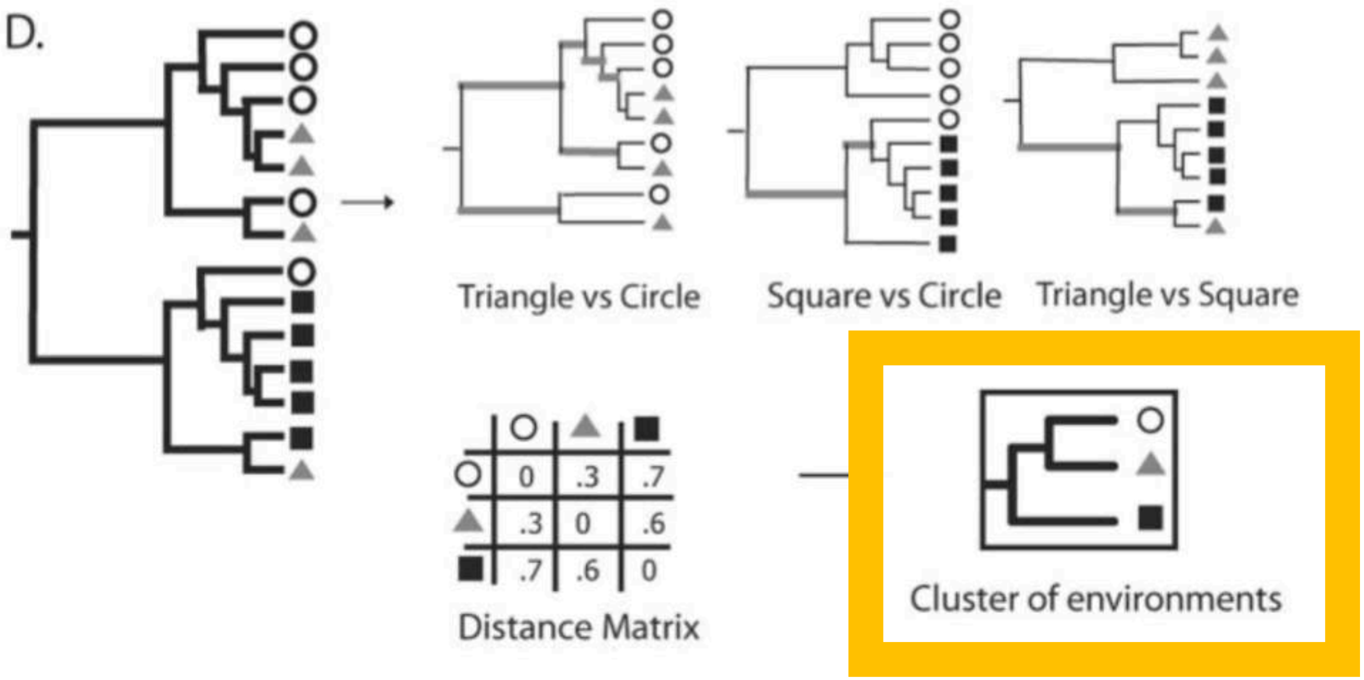
Fun fact!
(Wikipedia)


In Greek mythology, **Procrustes** (Προκρούστης *Prokroustes*) or "the stretcher [who hammers out the metal]", also known as **Prokoptas** or **Damastes** (Δαμαστής) "subduer", was a rogue smith and bandit from Attica who physically attacked people by stretching them or cutting off their legs, so as to force them to fit the size of an iron bed.

McHardy et al. 2013. Microbiome. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships

"Trees"

UPGMA = Unweighted Pair Group Method with Arithmetic Mean



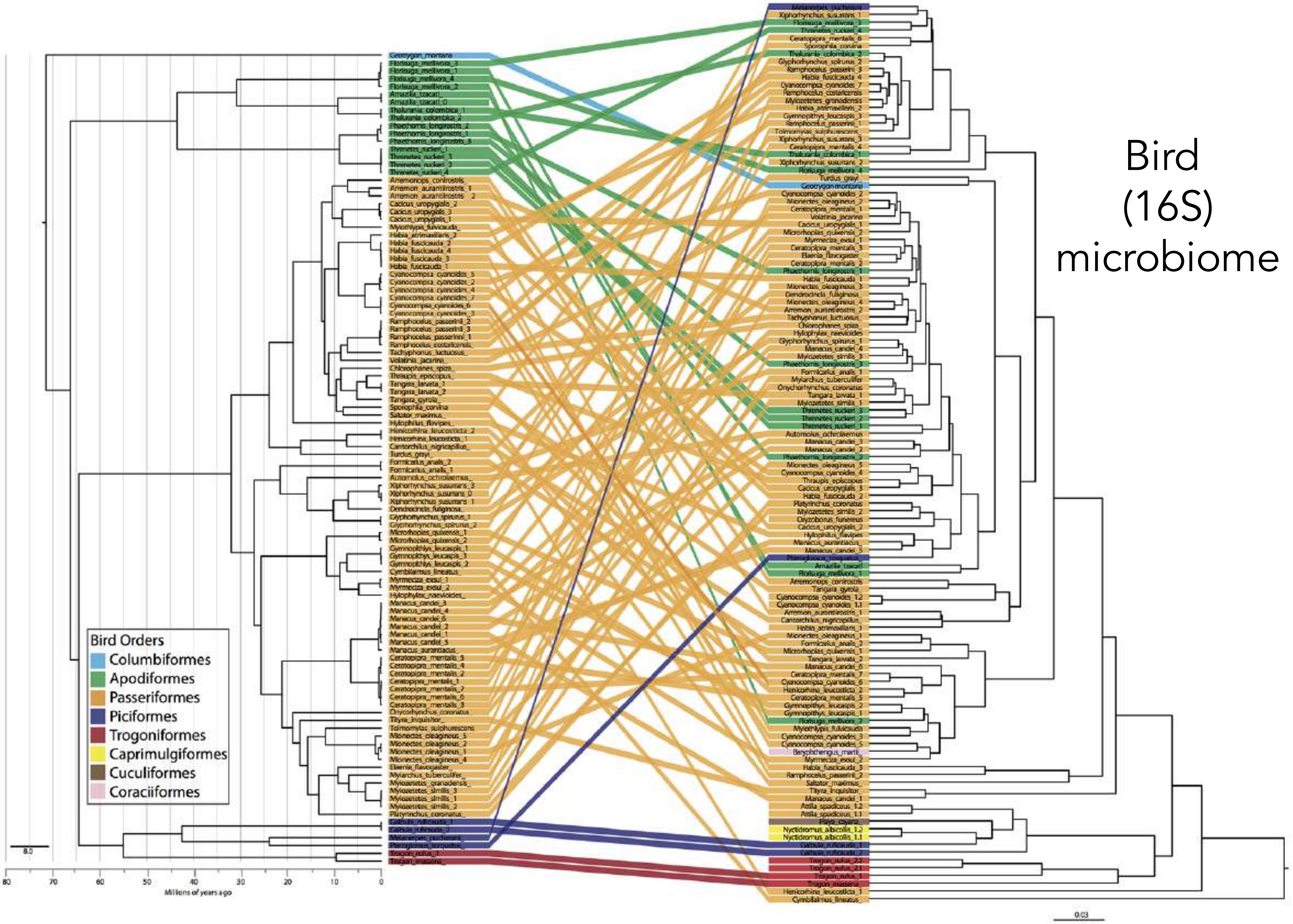
	○	▲	■
○	0	.3	.7
▲	.3	0	.6
■	.7	.6	0

Hird et al. 2015. Frontiers in Microbiology. Comparative gut microbiota of 59 Neotropical bird species.

"Trees"

UPGMA = Unweighted Pair Group Method with Arithmetic Mean

Bird
mtDNA



Bird
(16S)
microbiome

Hird et al. 2015. Frontiers in Microbiology. Comparative gut microbiota of 59 Neotropical bird species.

Categories

Statistical comparison of groups within your data

- Are the red-tailed chipmunk samples different from the yellow-pine chipmunk samples?
- Is the right hand of right handed people different from the left hand of left handed people?

Categories

Statistical comparison of groups within your data

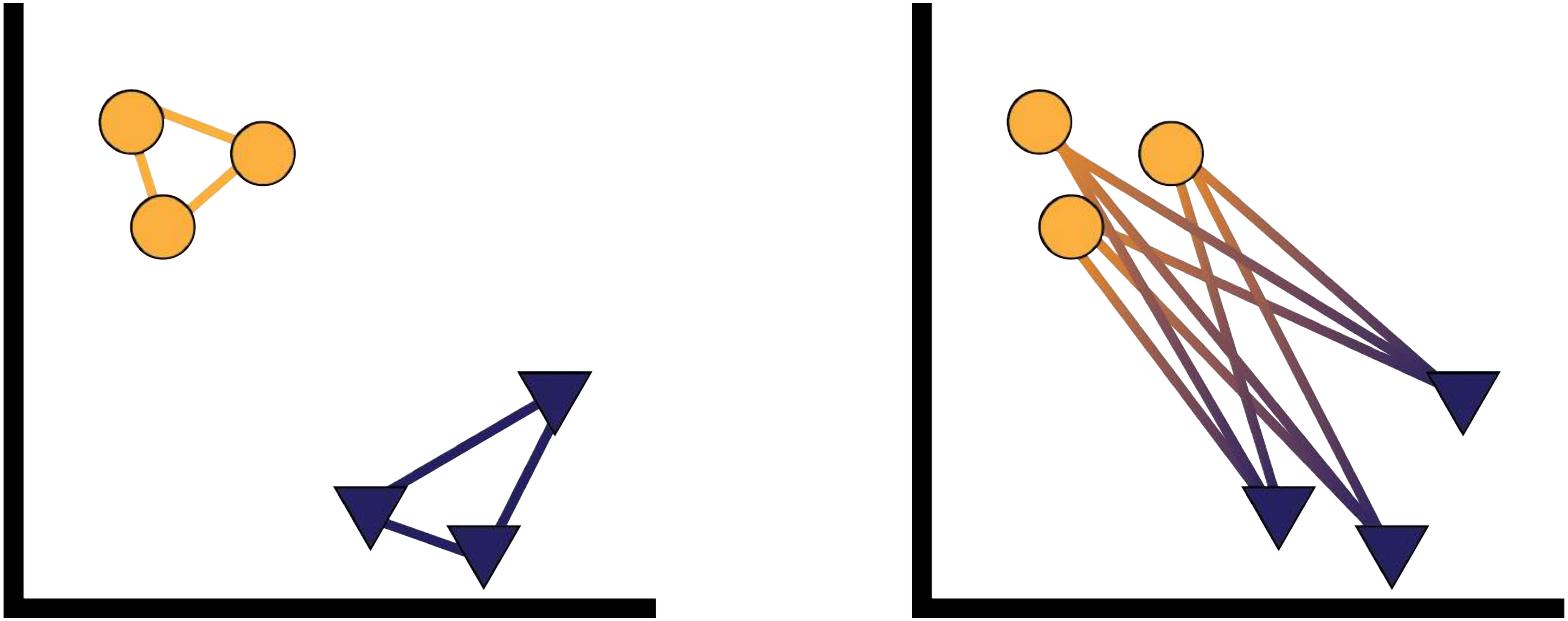
- Are the red-tailed chipmunk samples different from the yellow-pine chipmunk samples?
- Is the right hand of right handed people different from the left hand of left handed people?

Categorical tests

- ANOSIM
- Adonis (PERMANOVA)

Categories

Are distances WITHIN categories smaller than distances BETWEEN categories?



Categories

Statistical comparison of groups within your data

Categorical tests
- ANOSIM
- Adonis (PERMANOVA)

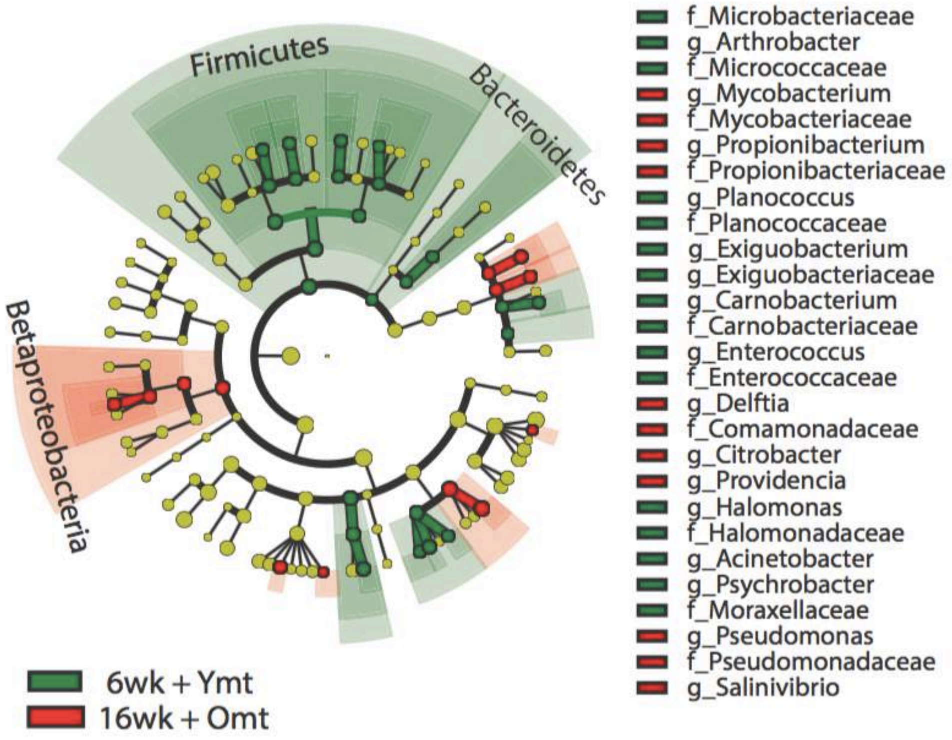
Calculate a significance and effect size of the variables

Significant OTUs

Which OTUs change between variables or treatments?

Significant OTUs

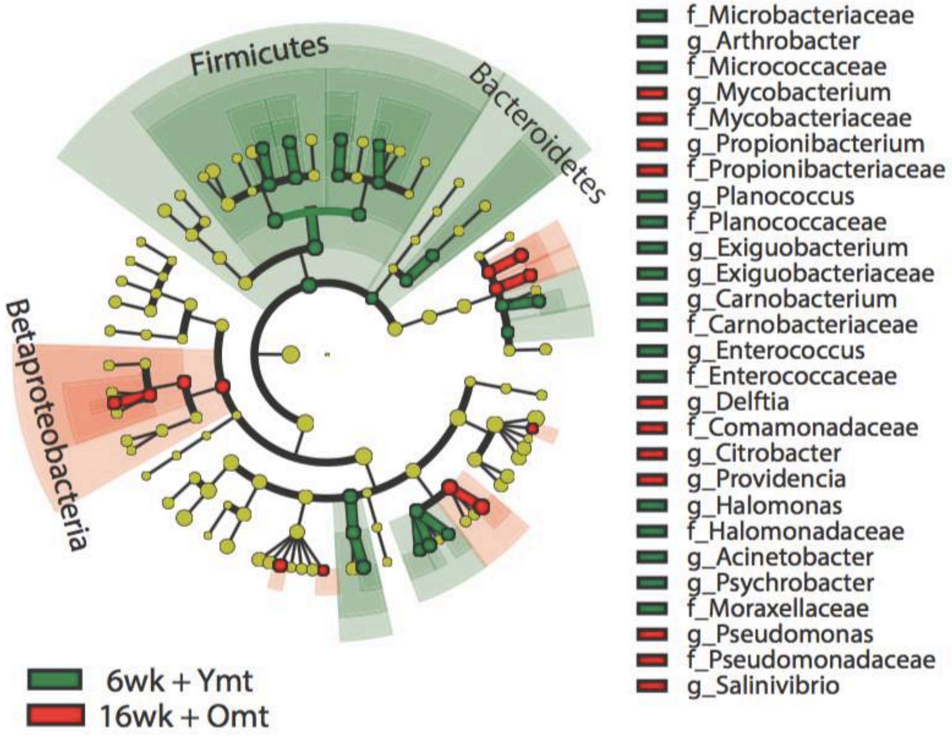
Which OTUs change between variables or treatments?

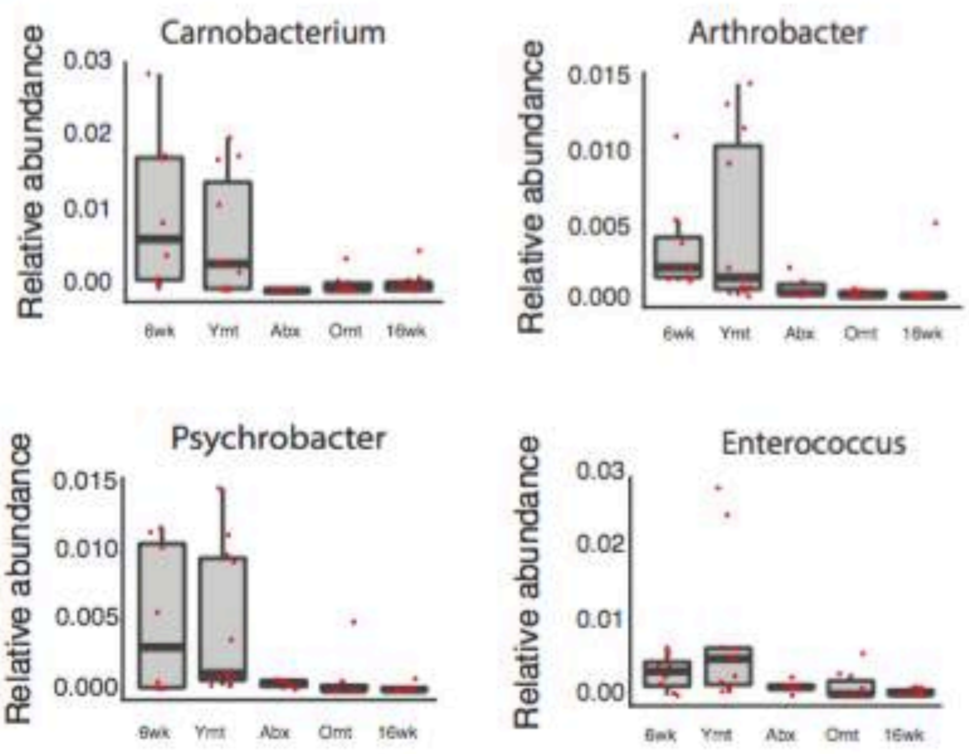


Smith et al. 2017. *eLife*. Regulation of life span by the gut microbiome in the short-lived African turquoise killifish

Significant OTUs

Which OTUs change between variables or treatments?





Smith et al. 2017. *eLife*. Regulation of life span by the gut microbiome in the short-lived African turquoise killifish

Networks

Which OTUs **always** / sometimes / **never** occur together?

Networks

Which OTUs **always** / sometimes / **never** occur together?

	Obese1	Obese2	Obese3	Obese4	Obese5	Lean1	Lean2	Lean3	Lean4	Lean5
A	0.2	0.3	0	0.3	0.3	0	0	0	0	0
B	0.35	0.35	0.65	0.35	0.35	0.35	0.35	0.35	0.35	0.35
C	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.3
D	0.1	0	0	0	0	0.3	0.3	0.3	0.3	0.35

Networks

Which OTUs **always** / sometimes / **never** occur together?

	Obese1	Obese2	Obese3	Obese4	Obese5	Lean1	Lean2	Lean3	Lean4	Lean5
A	0.2	0.3	0	0.3	0.3	0	0	0	0	0
B	0.35	0.35	0.65	0.35	0.35	0.35	0.35	0.35	0.35	0.35
C	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.3
D	0.1	0	0	0	0	0.3	0.3	0.3	0.3	0.35

SUMMARIZE
(percentage of
samples in a group
that contain both taxa)

	Obese	Lean
AB	80	0
AC	80	0
AD	20	0
BC	100	100
BD	20	100
CD	20	100

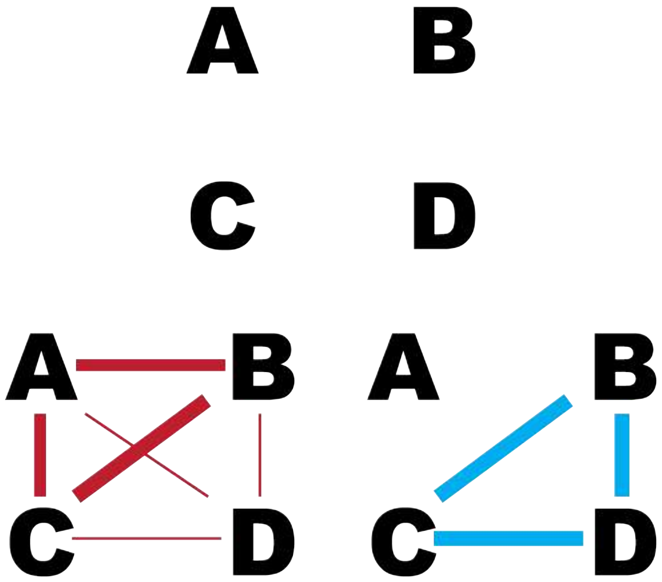
Networks

A **B**

C **D**

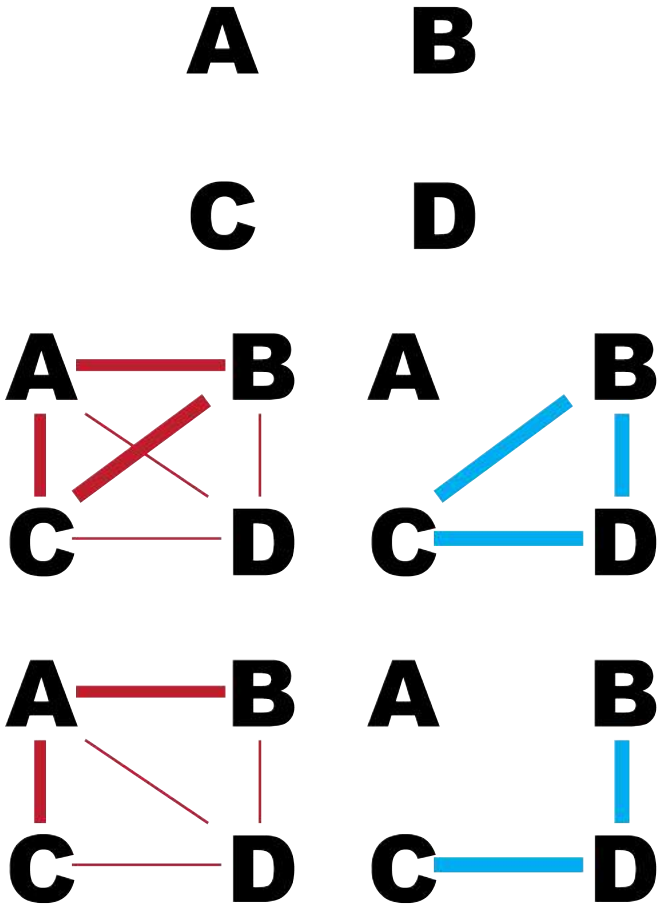
	Obese	Lean
AB	80	0
AC	80	0
AD	20	0
BC	100	100
BD	20	100
CD	20	100

Networks



	Obese	Lean
AB	80	0
AC	80	0
AD	20	0
BC	100	100
BD	20	100
CD	20	100

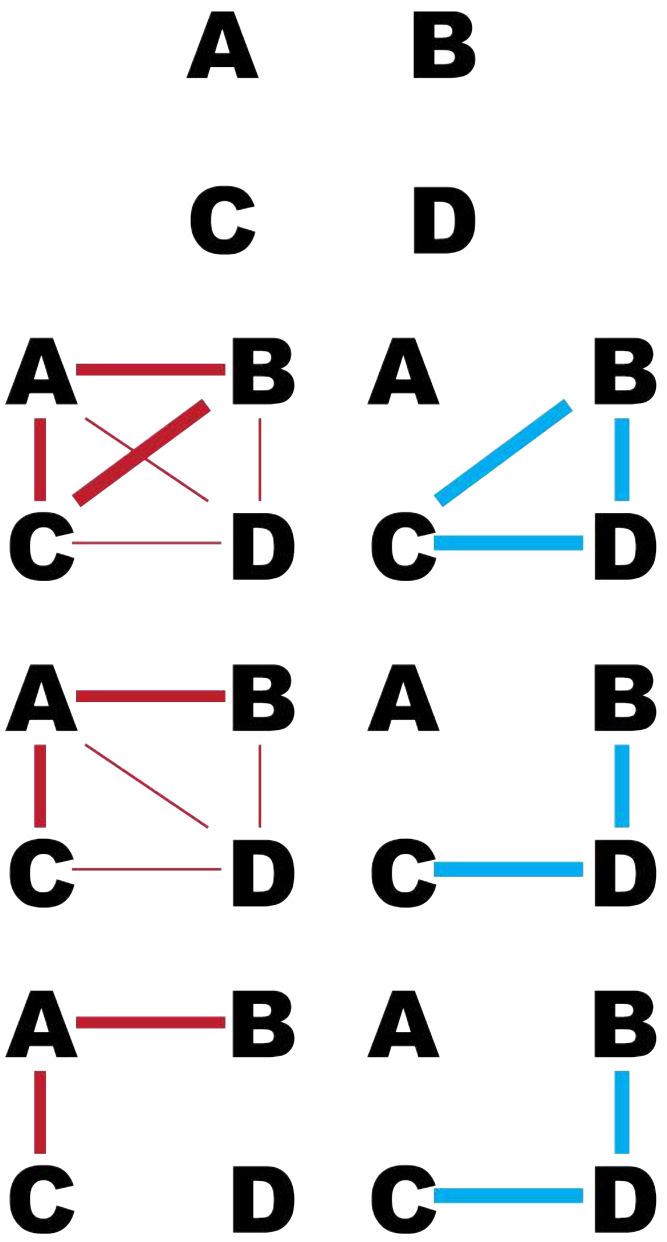
Networks



	Obese	Lean
AB	80	0
AC	80	0
AD	20	0
BC	100	100
BD	20	100
CD	20	100

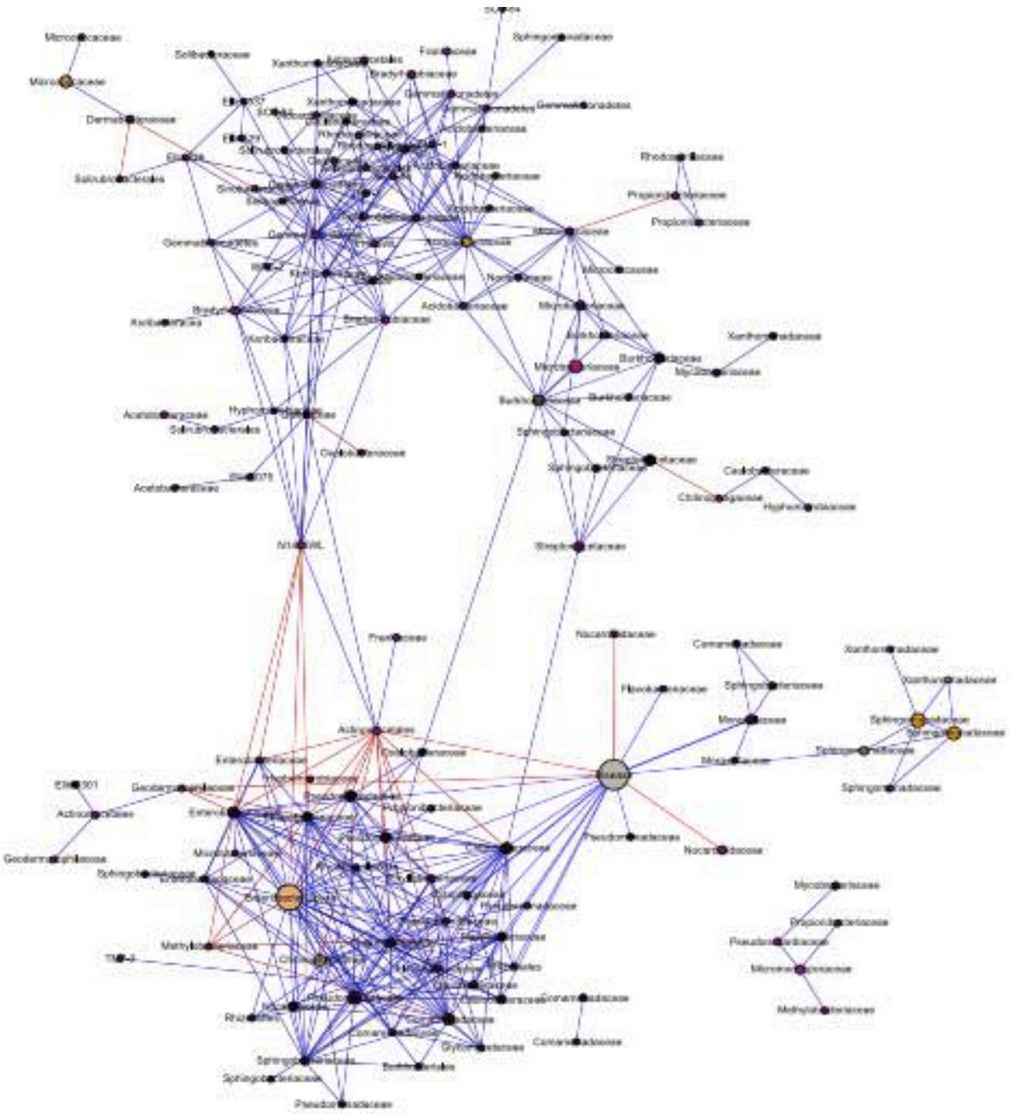
Networks

Which OTUs **always** / sometimes / **never** occur together?



- A & D are mutually exclusive
- A & D are connected to identical OTUs
- Phylogenetic information
- Function

Networks



(Intentionally no citation)

4. Analyze the results (16S)

Analyses of 16S rRNA sequence data usually consist of:
- **Identify** “who” is there
- **Describe**/measure/quantify diversity (alpha and beta)
- **Compare** categories

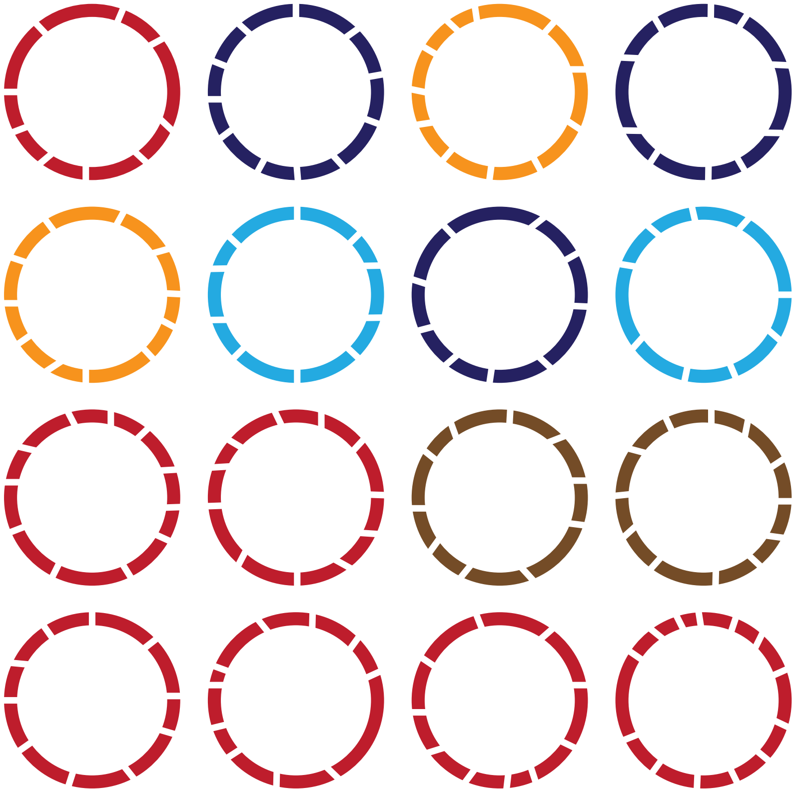
Any questions?

Apatosaurus

DK Find Out!

4. Analyze the results (Shotgun metagenomics)

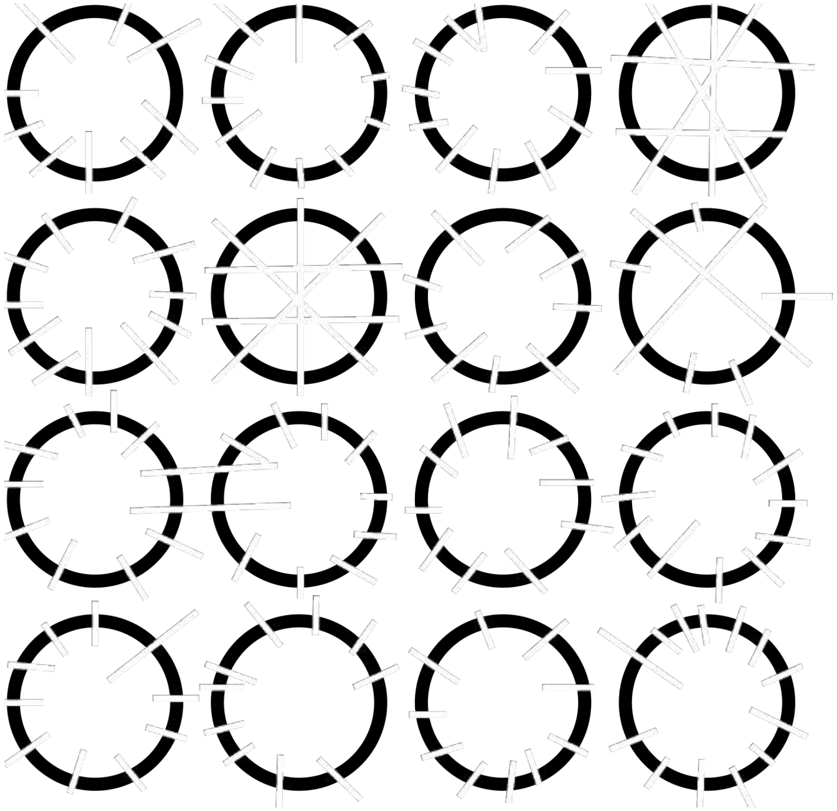
Analyses of shotgun metagenomic sequence data usually consist of:
- Identify "who" is there
- Identify function



@sarahmhird

Bacterial Genomes

(Generally) Circular chromosome
Plasmids
Size (130kb – 14Mb)
Horizontal gene transfer



Bacterial Genomics

(Generally) Circular chromosome

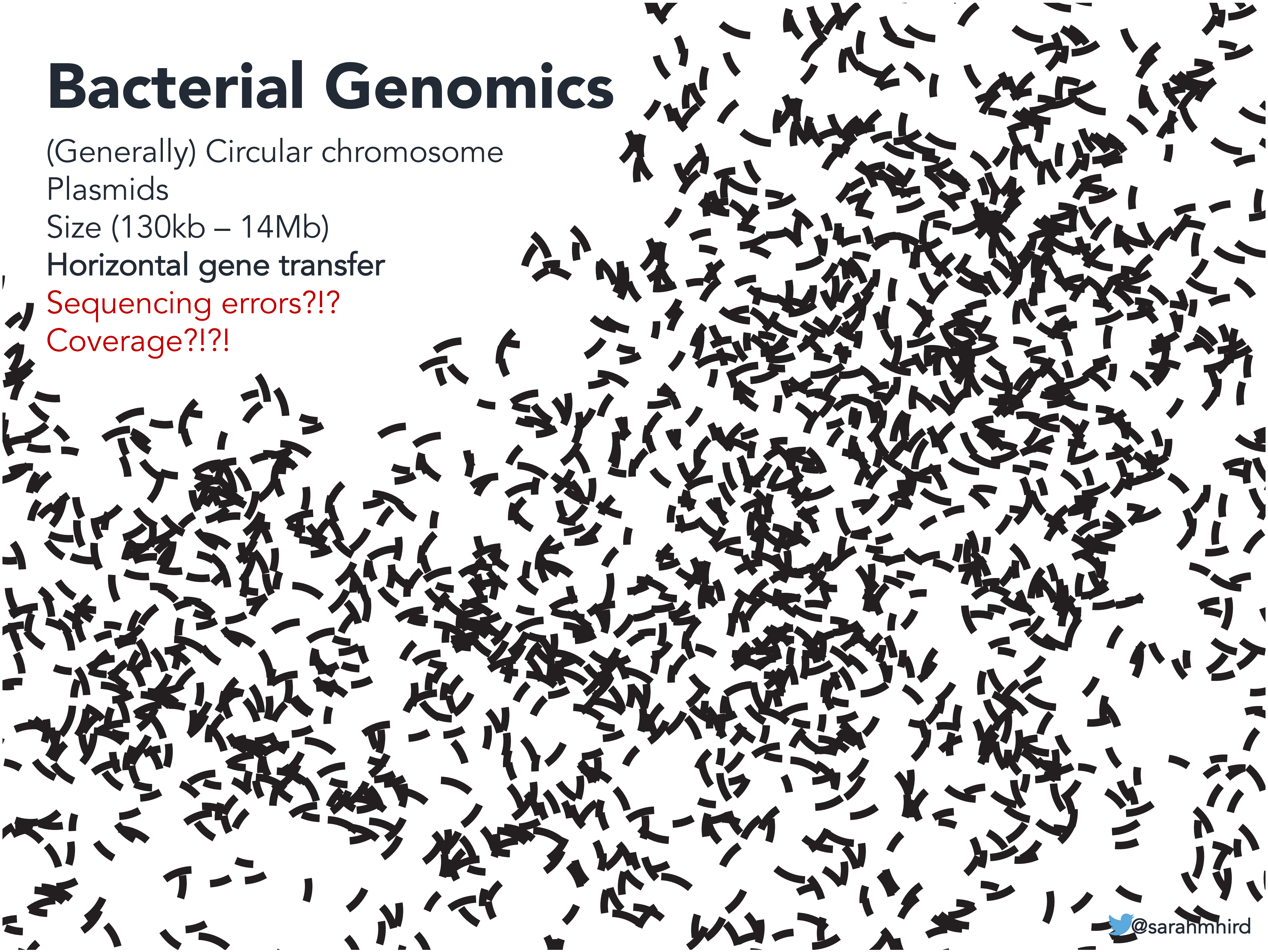
Plasmids

Size (130kb – 14Mb)

Horizontal gene transfer

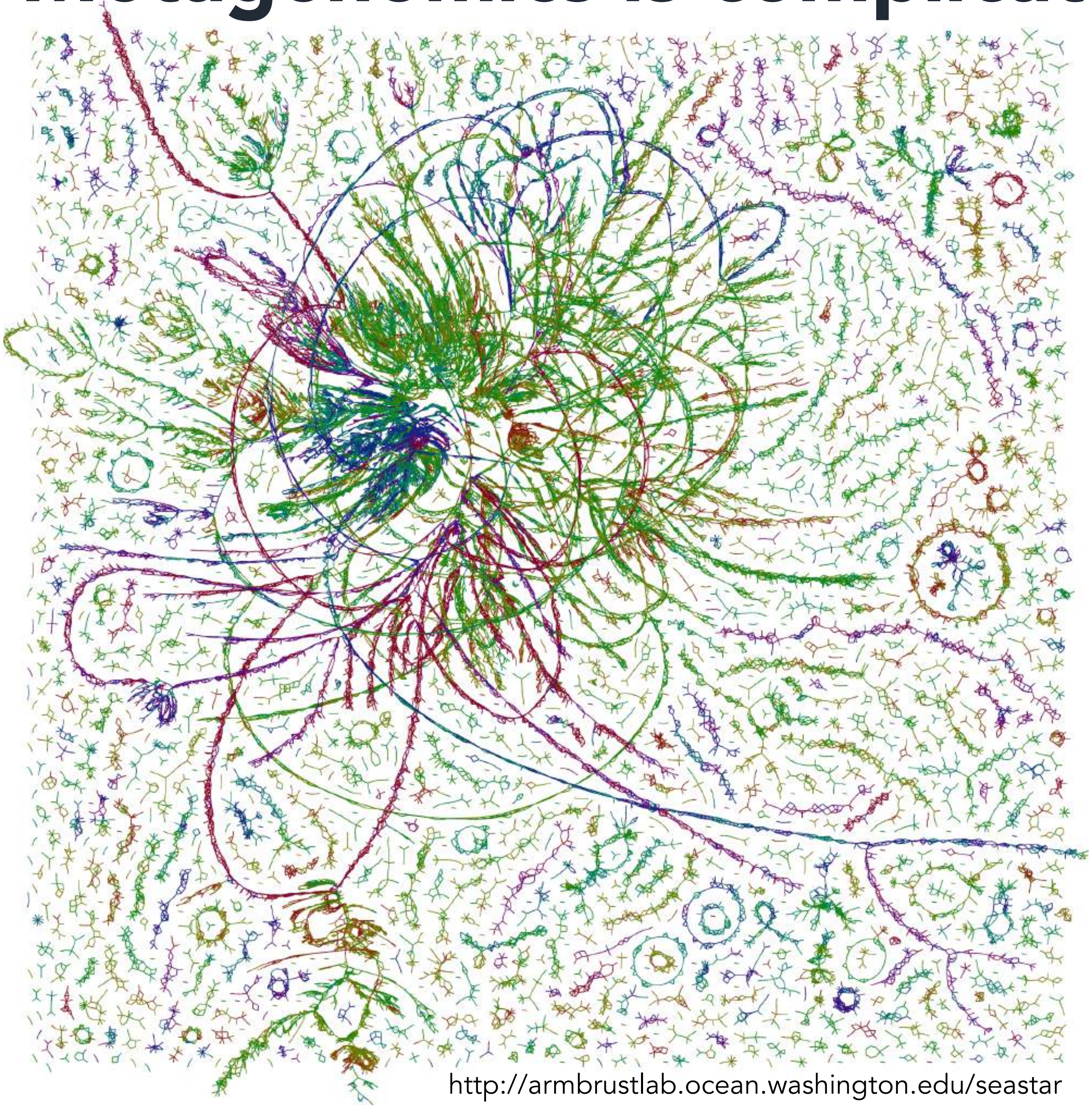
Sequencing errors?!?

Coverage?!?!



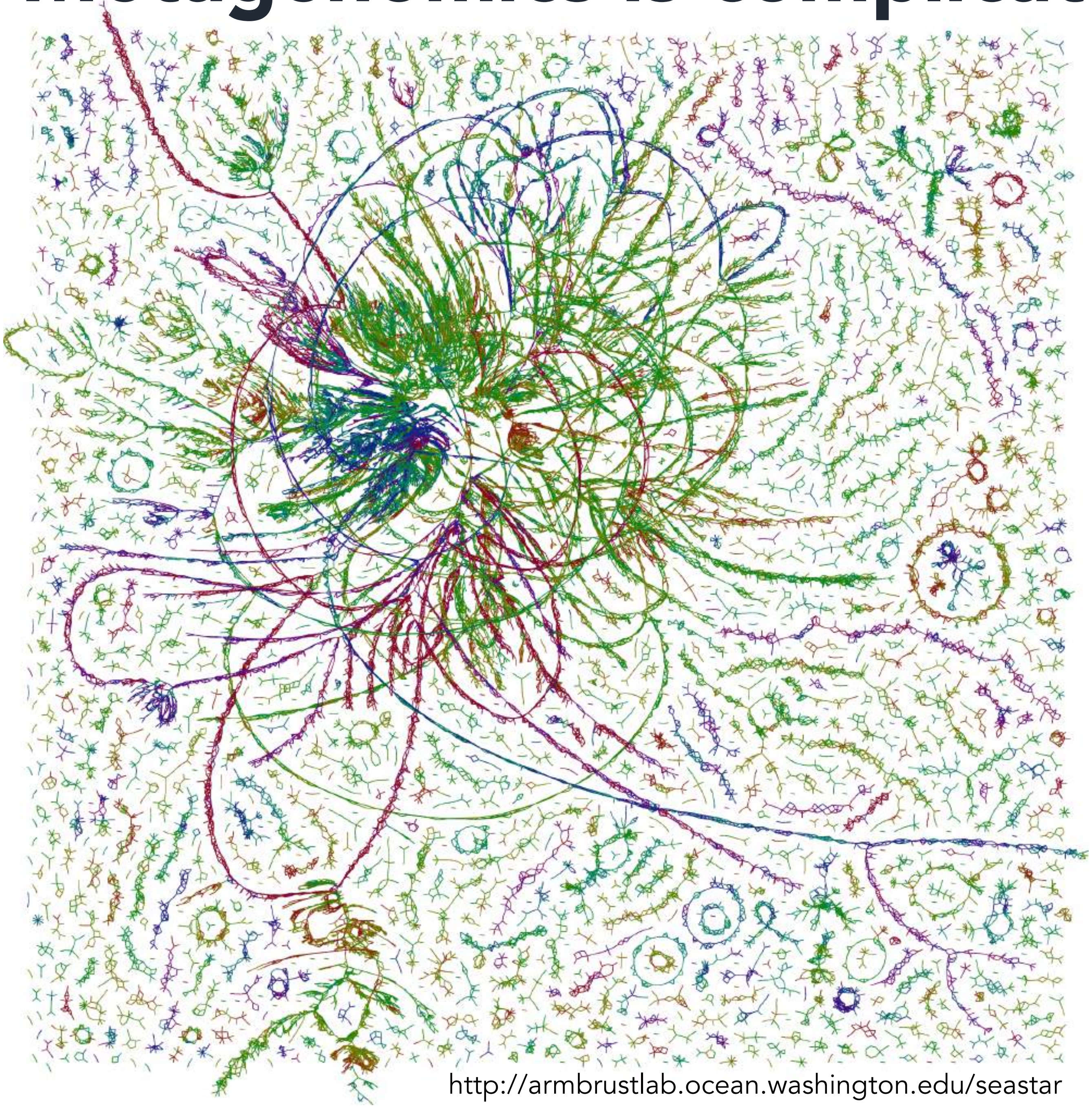
@sarahmhird

Metagenomics is complicated.



http://armbrustlab.ocean.washington.edu/seastar

Metagenomics is complicated.



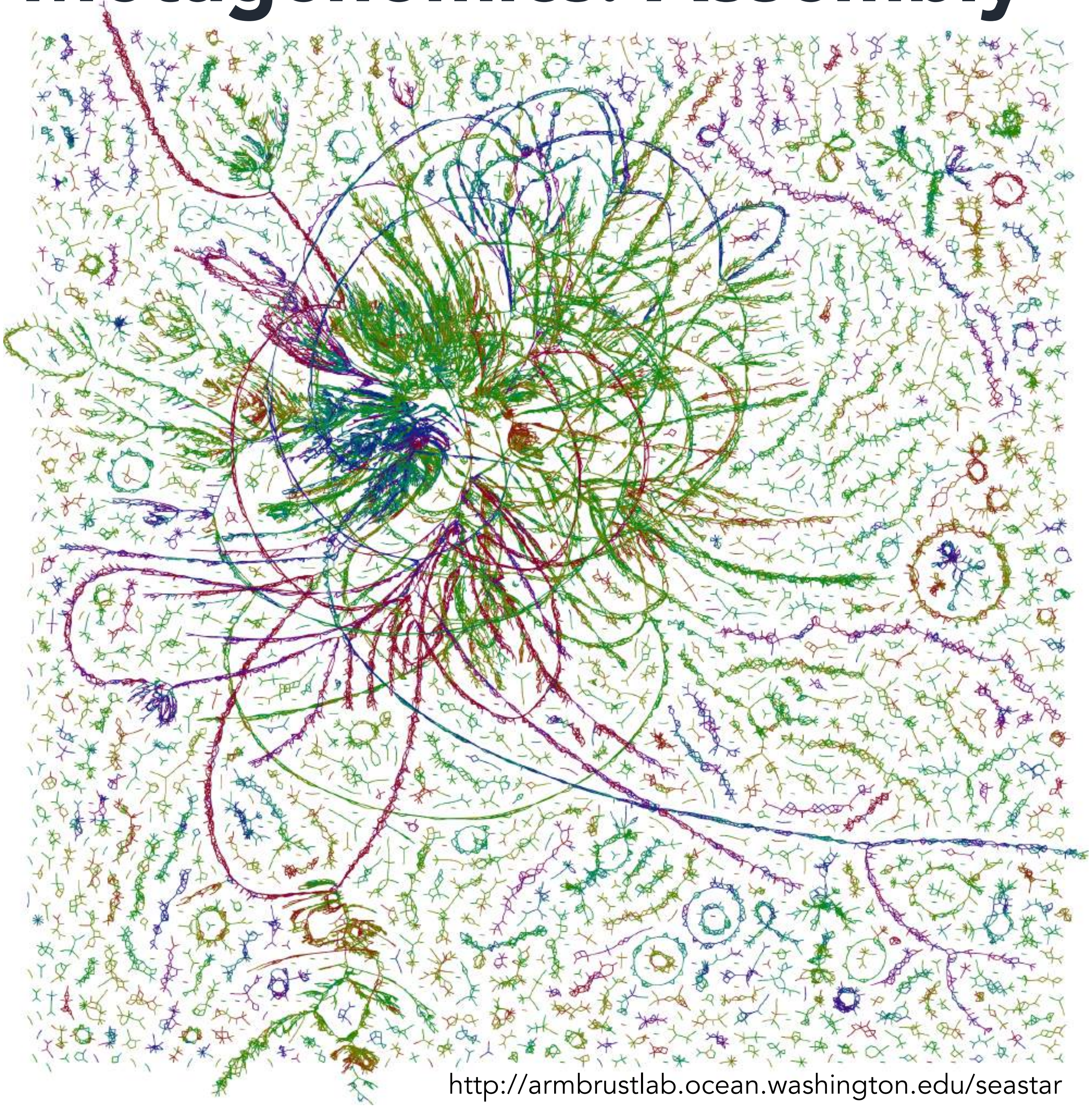
http://armbrustlab.ocean.washington.edu/seastar

Assemble (or not)

Align and Assign

Analyze

Metagenomics: Assembly



http://armbrustlab.ocean.washington.edu/seastar

Similar to genome assembly

Repetitive sequences and homologous regions make things difficult

deBruijn graphs are frequently utilized

@sarahmhird

Metagenomics: Align and Assign

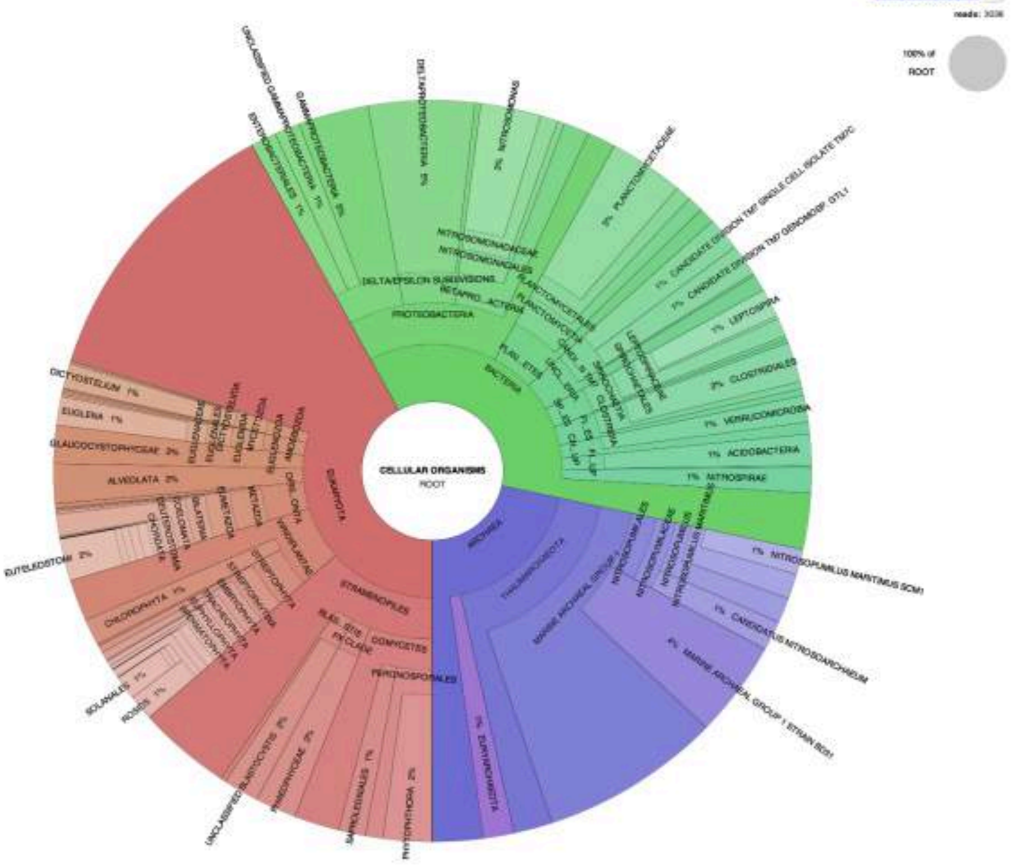
WHO is there?

Marker Gene Analysis – look for specific marker genes that will tell you *who* is in the sample

PROGRAMS

PhyloSift: Uses database of ELITE markers; will only use about 1% of your data; taxonomic identification

- Low copy number
- "Universal"

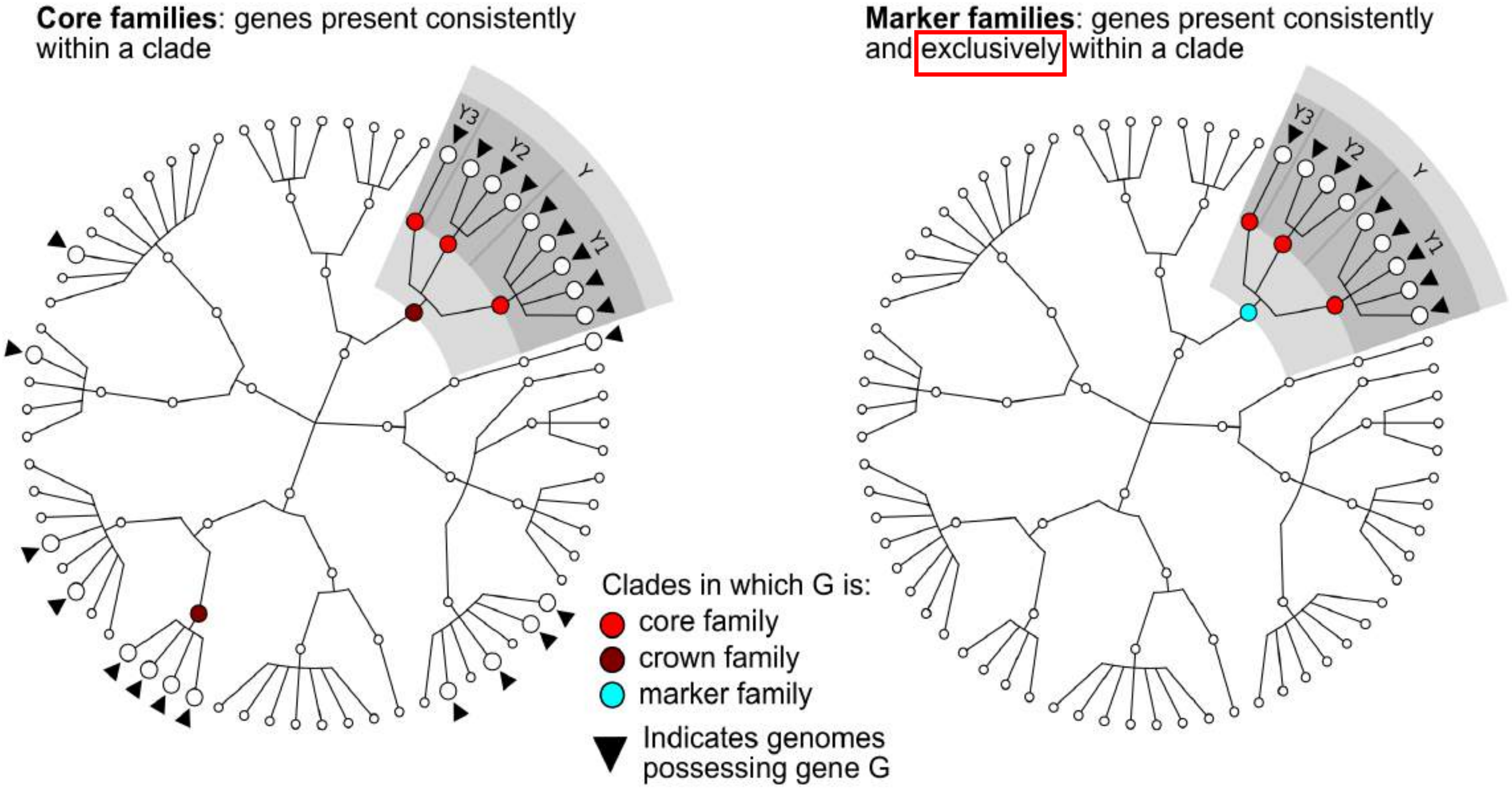


@sarahmhird

Metagenomics: Align and Assign

WHO is there?
Marker Gene Analysis – look for specific marker genes that will tell you *who* is in the sample

PROGRAMS
metaphlan2: Uses database of CLADE SPECIFIC markers; taxonomic identification



Metagenomics: Align and Assign

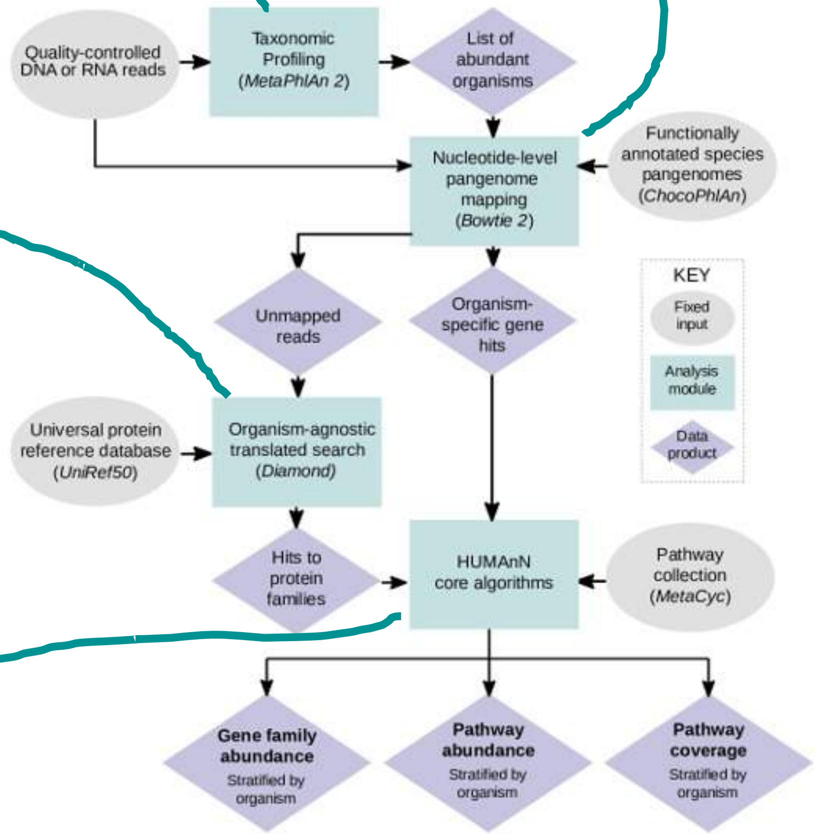
Functional Profiling still requires databases.

WHAT are the microbes (genes) doing?

PROGRAMS

humann2:

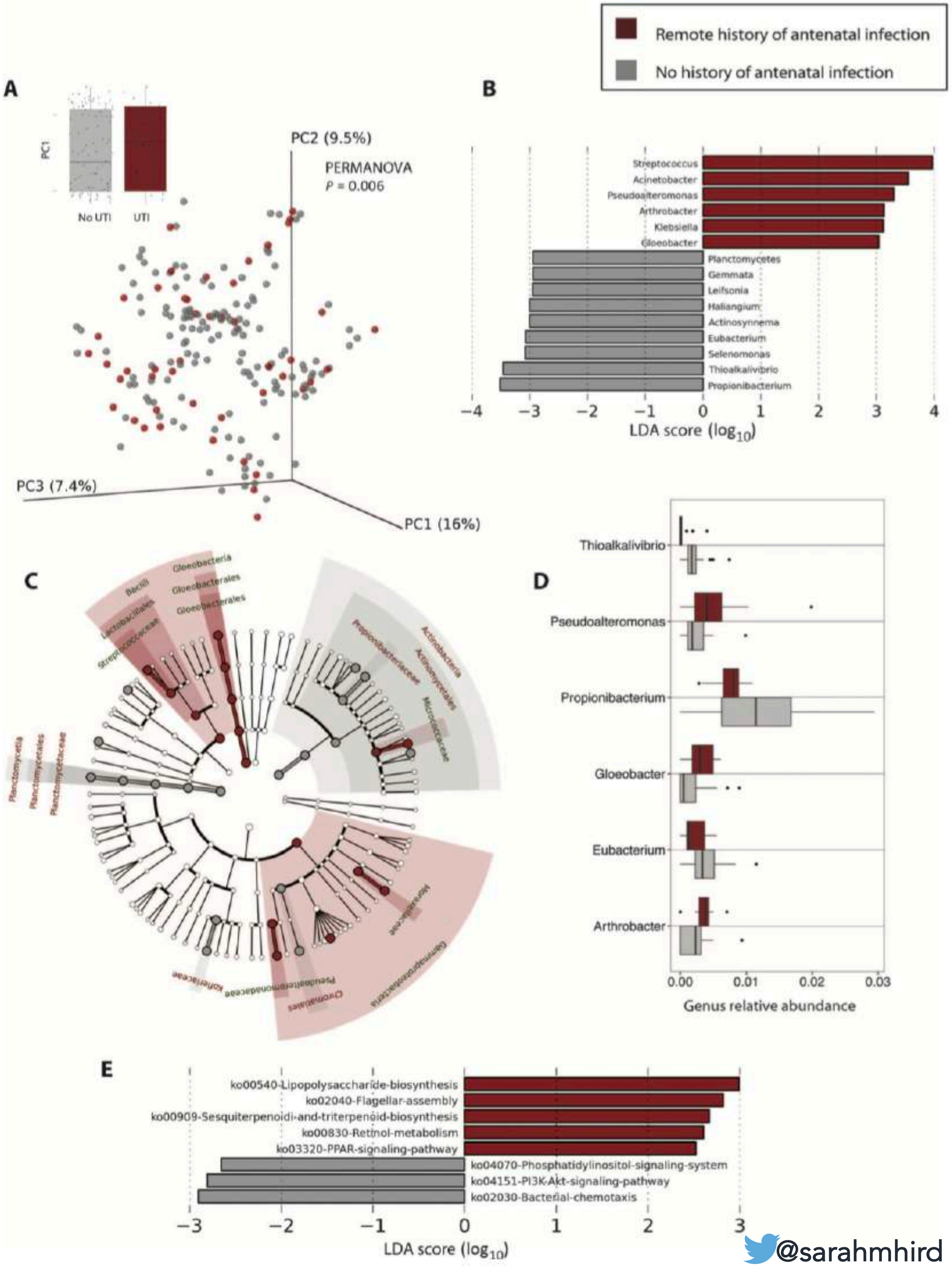
1. Identify who is there and specific genes
2. Identify anonymous genes
3. Summarize



Metagenomics: Analyze

Are two samples (classes) different? How?

PROGRAMS
Lefse (seen here)
R
Lots of others…



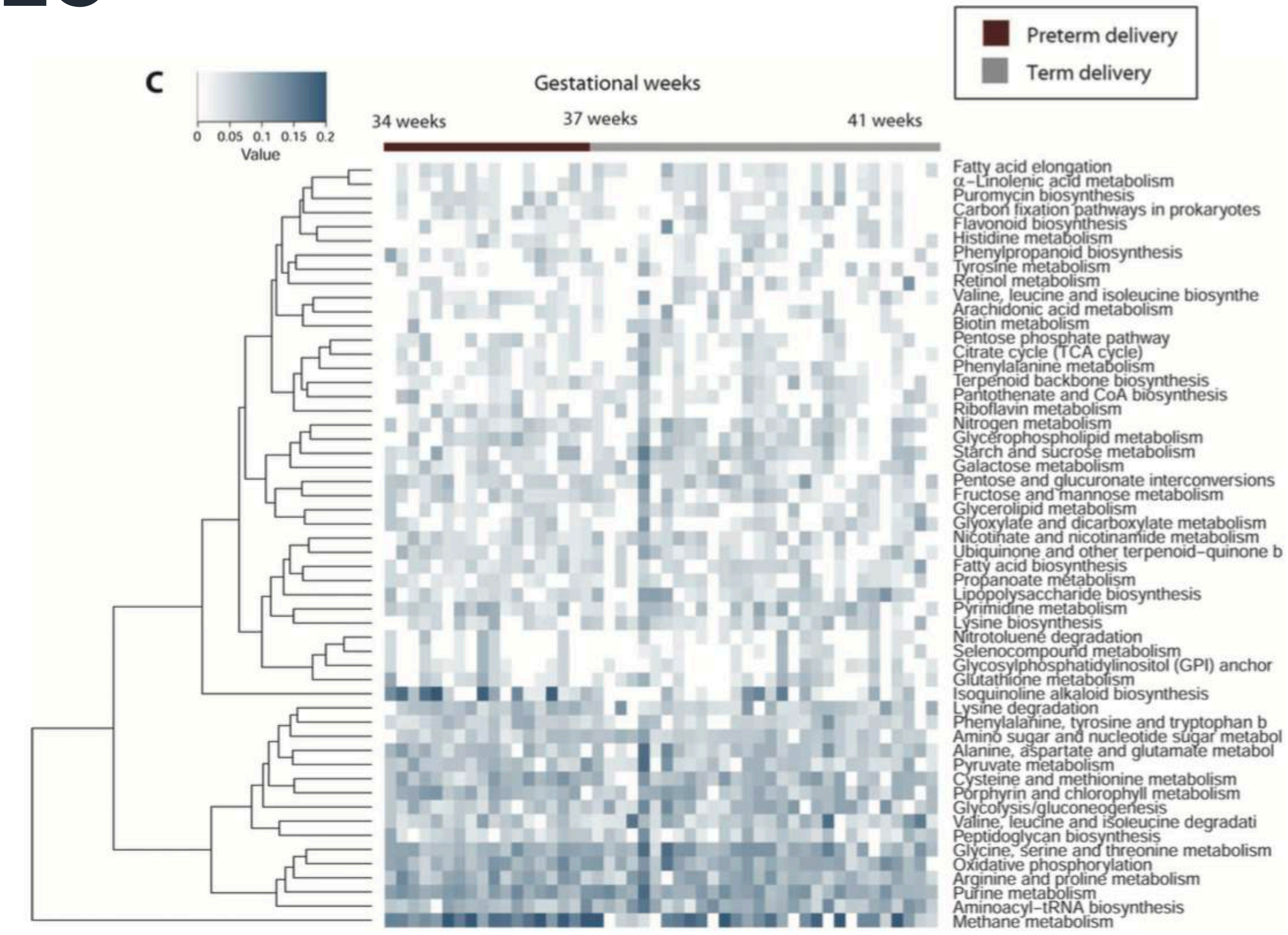
MICROBIOME

The Placenta Harbors a Unique Microbiome

Kjersti Aagaard,[1,2,3]* Jun Ma,[1,2] Kathleen M. Antony,[1] Radhika Ganu,[1] Joseph Petrosino,[4] James Versalovic[5]

@sarahmhird

Metagenomics: Analyze



MICROBIOME

The Placenta Harbors a Unique Microbiome

Kjersti Aagaard,[1,2,3]* Jun Ma,[1,2] Kathleen M. Antony,[1] Radhika Ganu,[1] Joseph Petrosino,[4] James Versalovic[5]

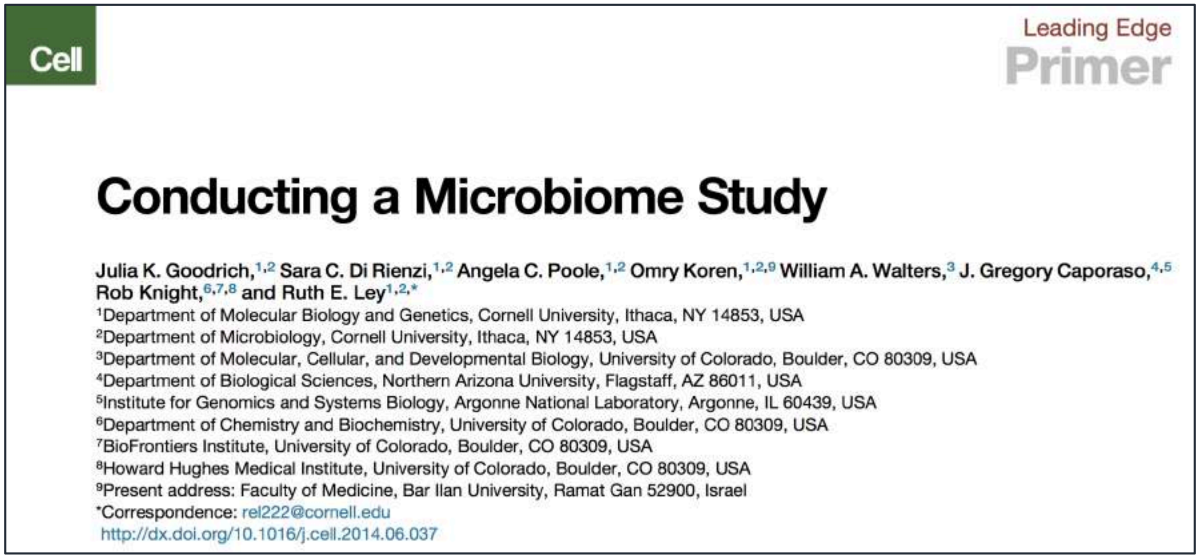
Pachycephalosaurus

DK Find Out!

5. **Interpret the results**
6. **Publish! Share data!**

Microbiome Methods

1. Develop a question
2. Design the study
3. Conduct the study
4. Analyze the results
5. Interpret the results
6. Publish; share data



Cell

Leading Edge
Primer

Conducting a Microbiome Study

Julia K. Goodrich,[1,2] Sara C. Di Rienzi,[1,2] Angela C. Poole,[1,2] Omry Koren,[1,2,9] William A. Walters,[3] J. Gregory Caporaso,[4,5] Rob Knight,[6,7,8] and Ruth E. Ley[1,2,*]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[4]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA
[5]Institute for Genomics and Systems Biology, Argonne National Laboratory, Argonne, IL 60439, USA
[6]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[7]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[8]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
[9]Present address: Faculty of Medicine, Bar Ilan University, Ramat Gan 52900, Israel
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.cell.2014.06.037

Extremely detailed talk outline

1

Introduction

2

Methods

3

Research

Any questions?

Anchiornis

Zhao Chuang; Peking Natural Science Organization

Extremely detailed talk outline

1

Introduction

2

Methods

3

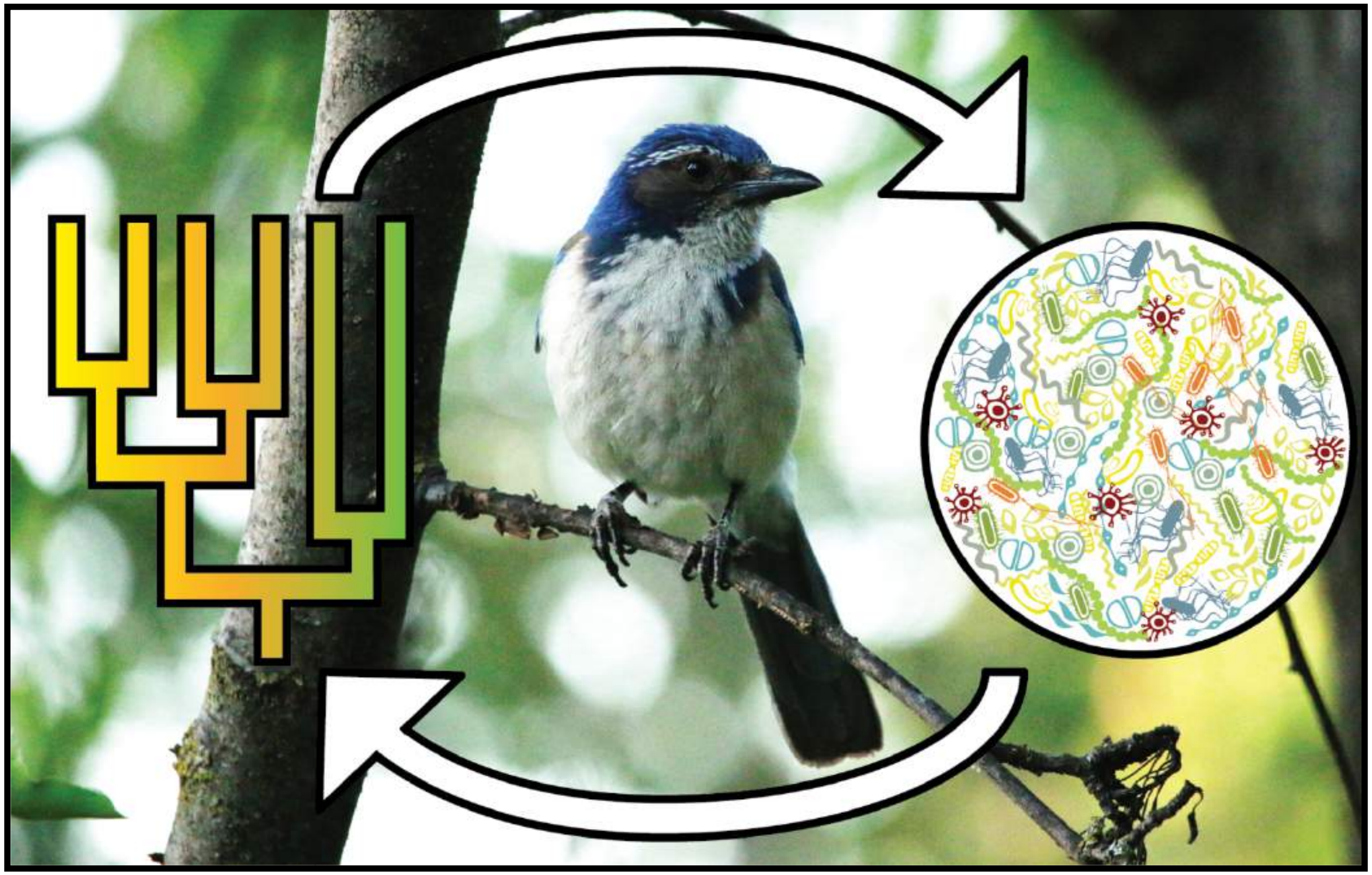
Research

INTRODUCTION

My Research Interests

How does the microbiome affect the evolution of hosts (birds)?

How does the host affect the evolution of the microbiome?



@sarahmhird

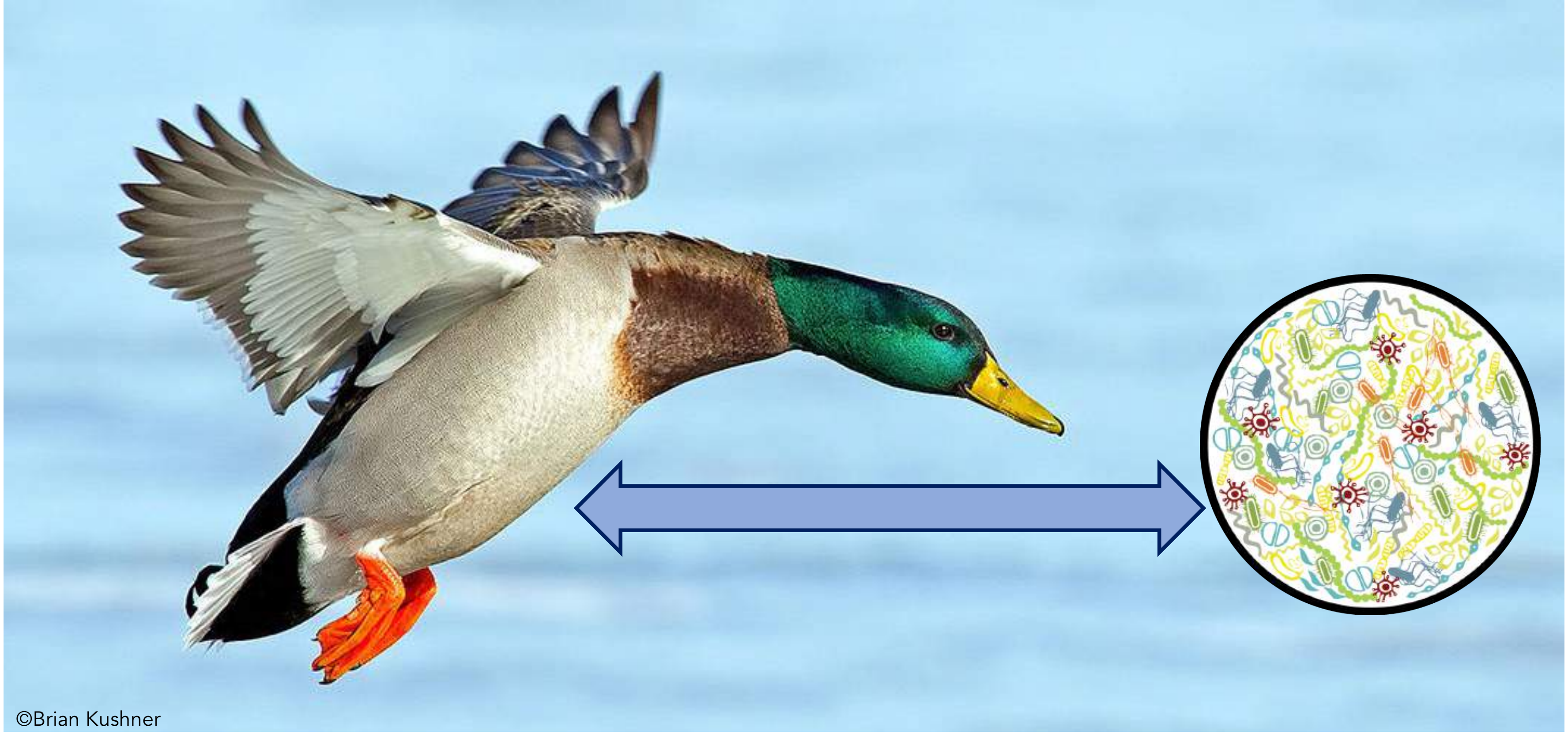
INTRODUCTION

The mallard

©Brian Kushner

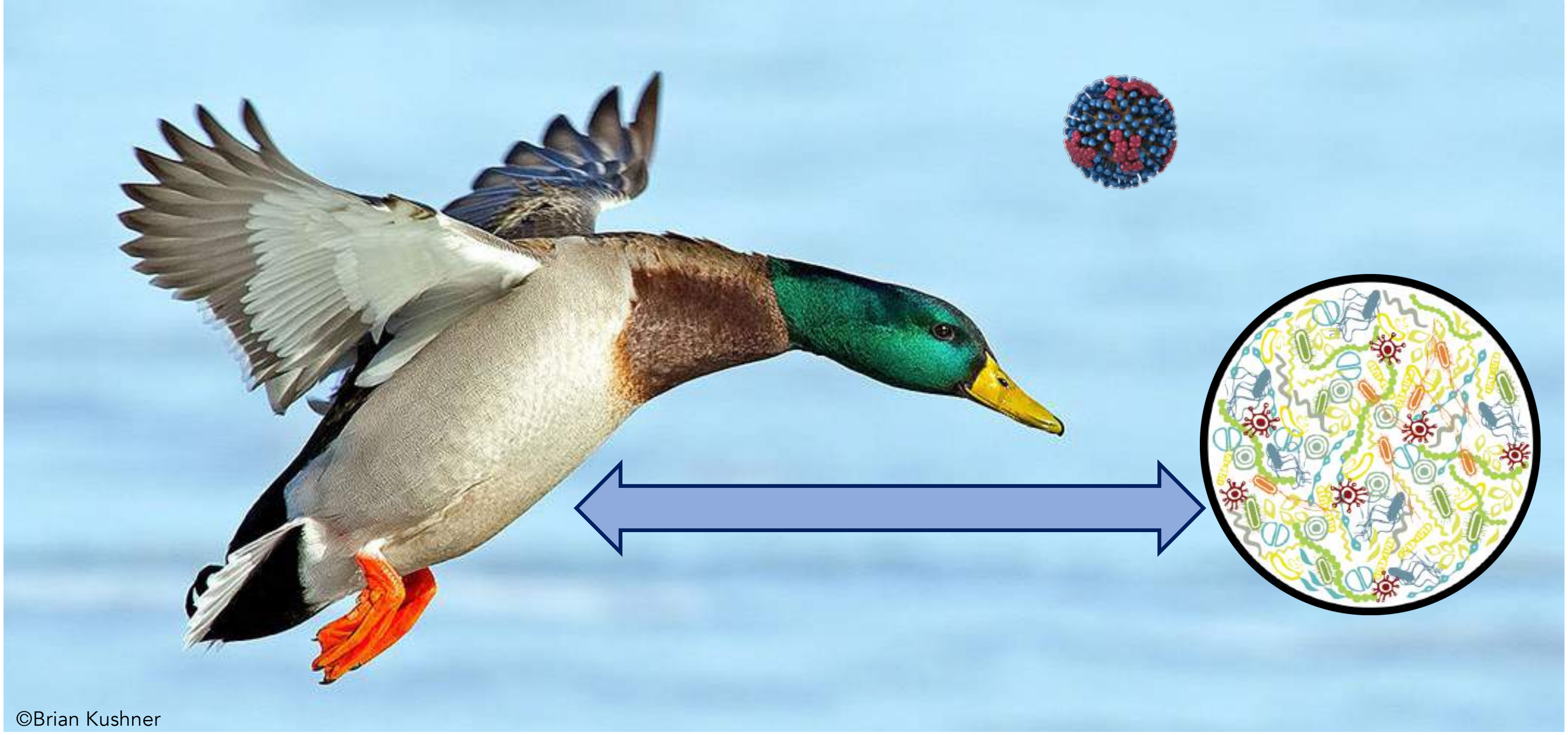
INTRODUCTION

The mallard


©Brian Kushner

INTRODUCTION

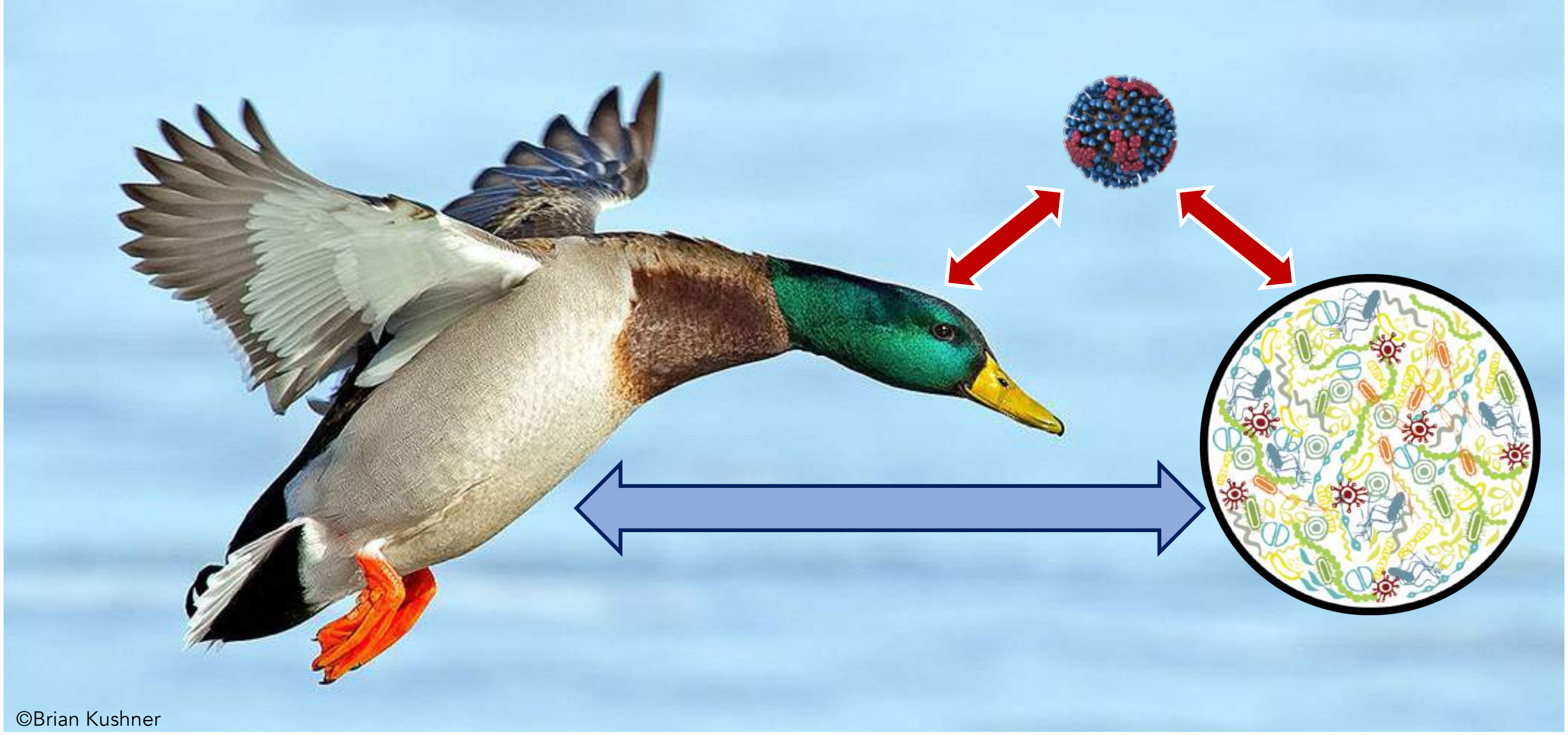
Scaling up: More microbes



©Brian Kushner

INTRODUCTION

Scaling up: More microbes

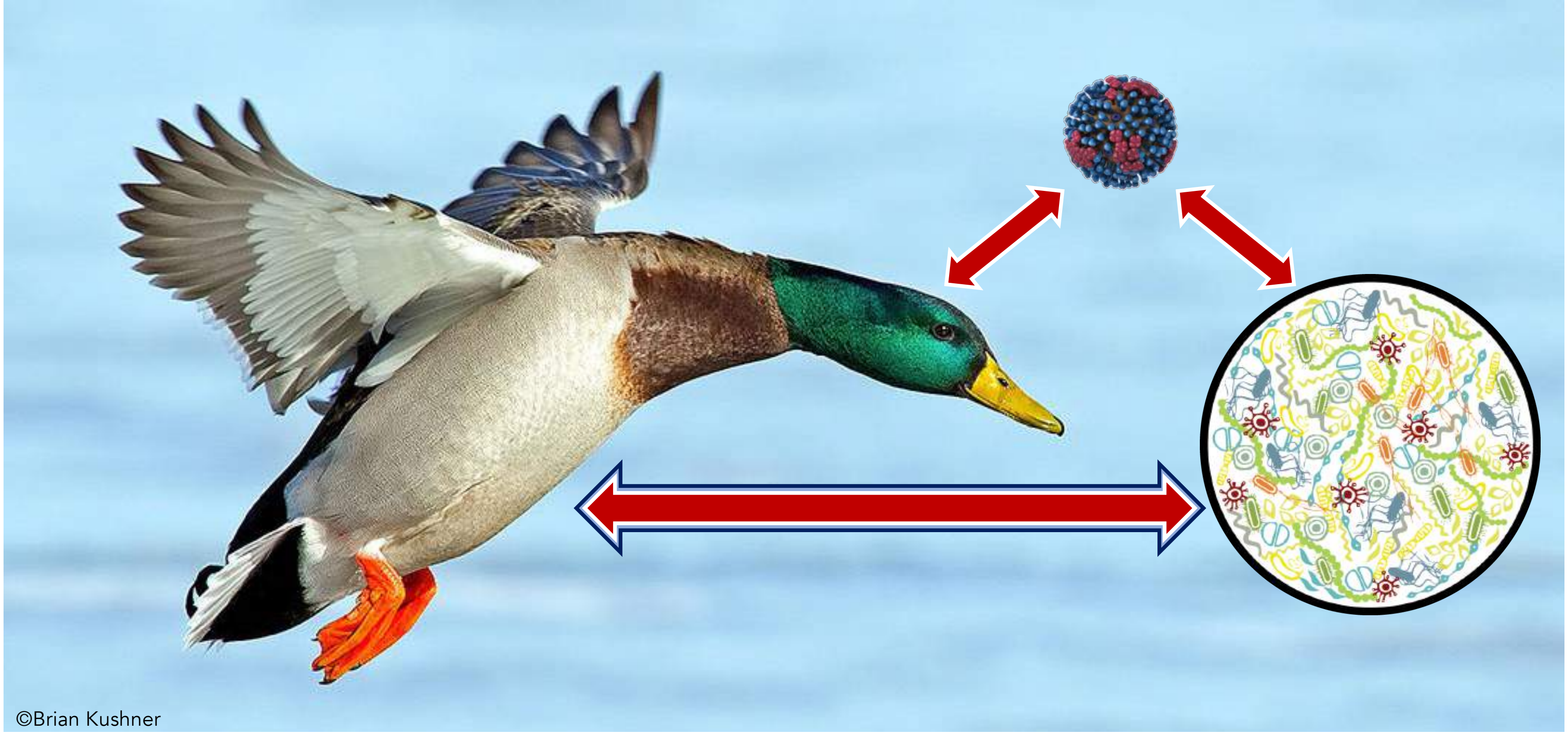


©Brian Kushner

@sarahmhird

INTRODUCTION

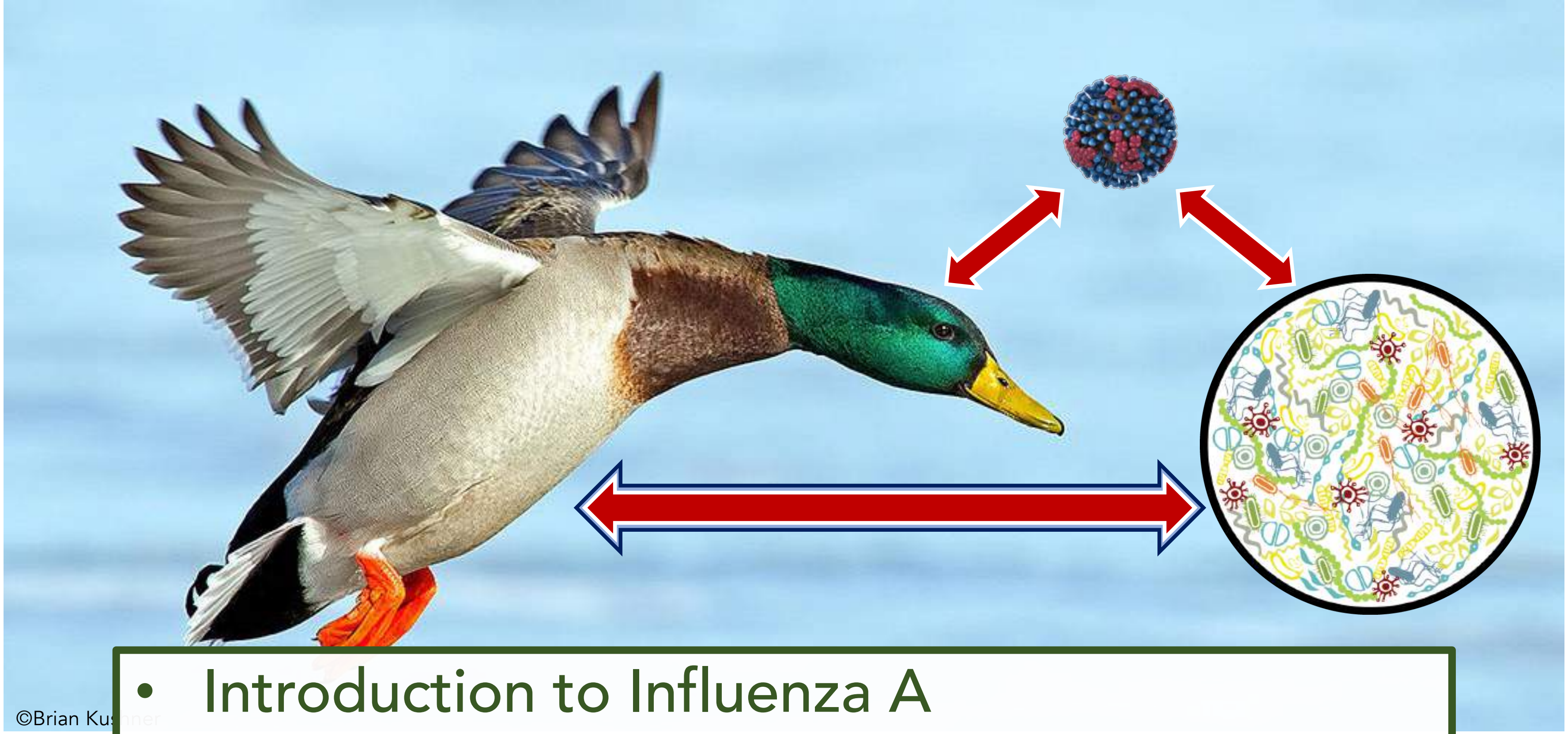
Scaling up: More microbes



©Brian Kushner

INTRODUCTION

Mini-talk outline

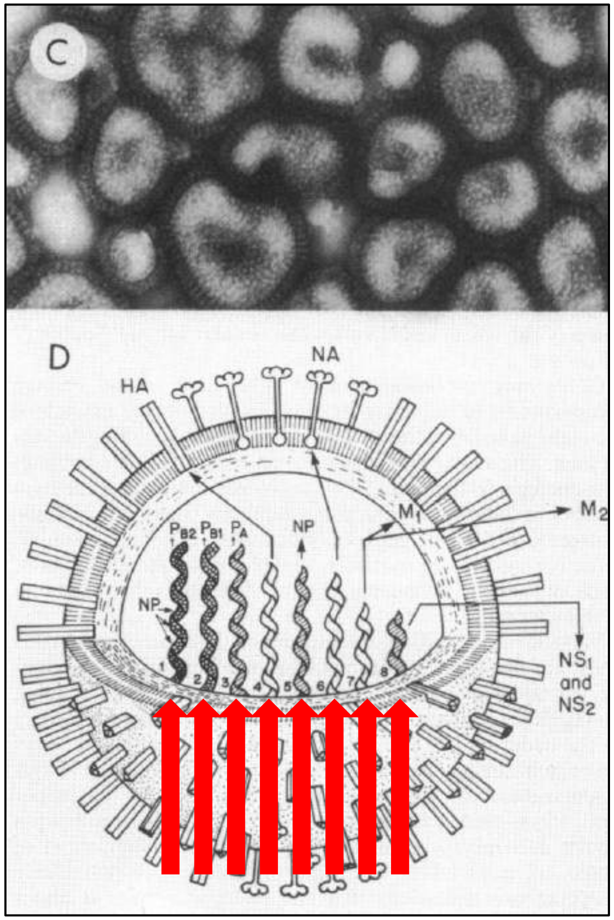

©Brian Kushner

- Introduction to Influenza A
- What's the big deal about Bird Flu?
- Research Question 1 - Mallards
- Research Question 2 – Five Duck Species

INTRODUCTION

Influenza A virus (IAV)

- RNA virus of family Orthomyxoviridae
- Can infect many birds and mammals
- 8 segments of ssRNA
  - code for up to 14 proteins



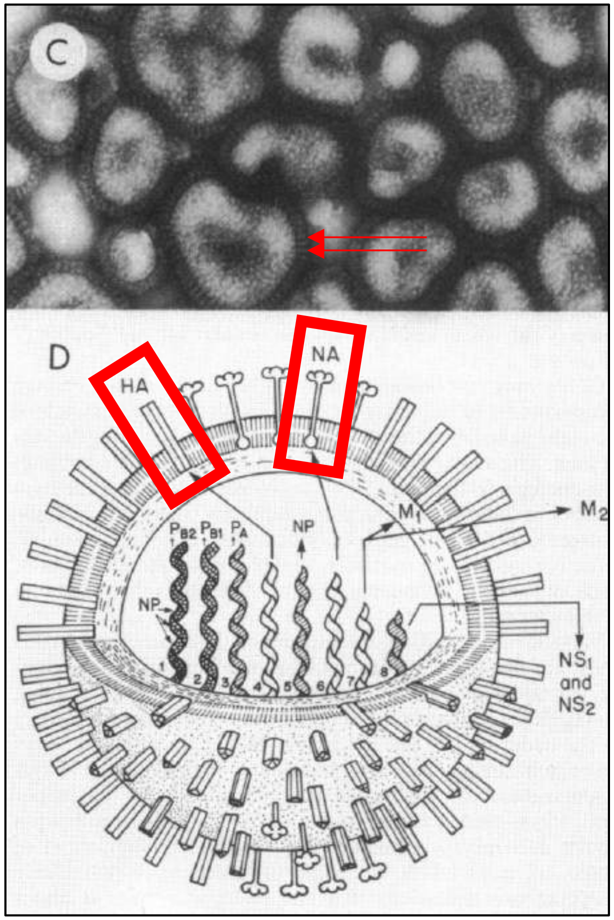
Webster et al. 1992

INTRODUCTION

Influenza A virus (IAV)

- RNA virus of family Orthomyxoviridae
- Can infect many birds and mammals
- 8 segments of ssRNA
    - code for up to 14 proteins

- 2 main antigenic glycoproteins: **hemagglutinin (HA)** and **neuraminidase (NA)**

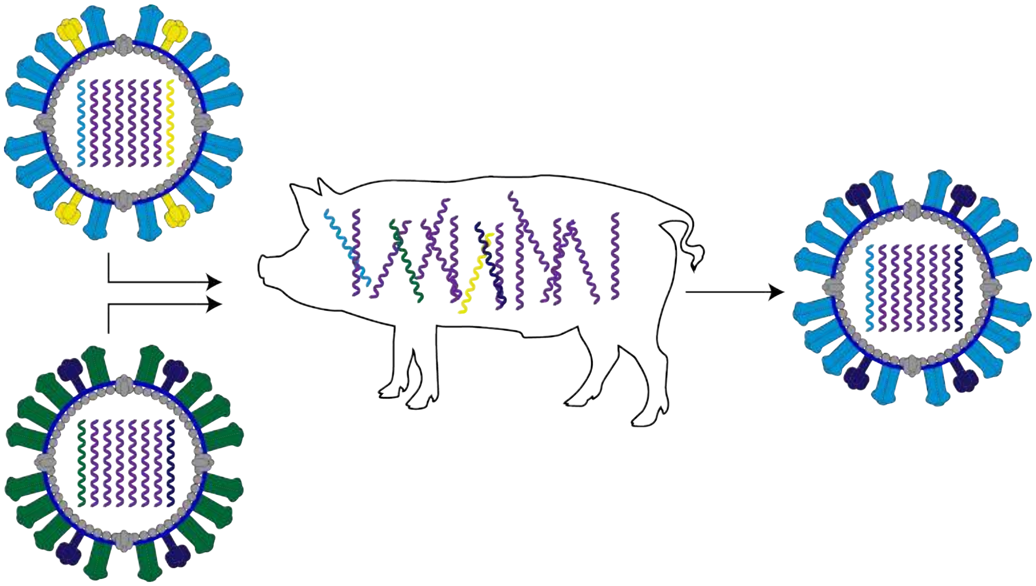


Webster et al. 1992

INTRODUCTION

Influenza A virus (IAV)

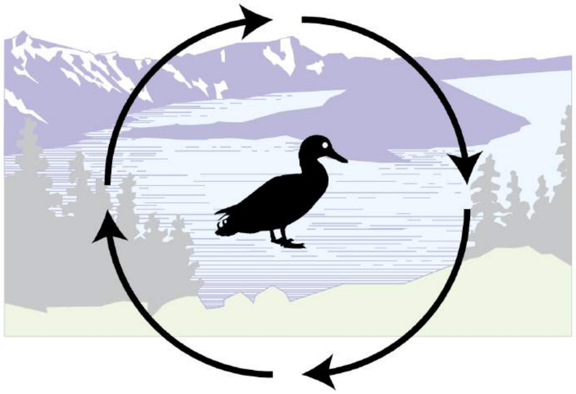
- RNA virus of family Orthomyxoviridae
- Can infect many birds and mammals
- 8 segments of ssRNA
  - code for up to 14 proteins
- 2 main antigenic glycoproteins: **hemagglutinin (HA)** and **neuraminidase (NA)**
- **Rapidly evolving**



@sarahmhird

What’s the big deal about Bird Flu?

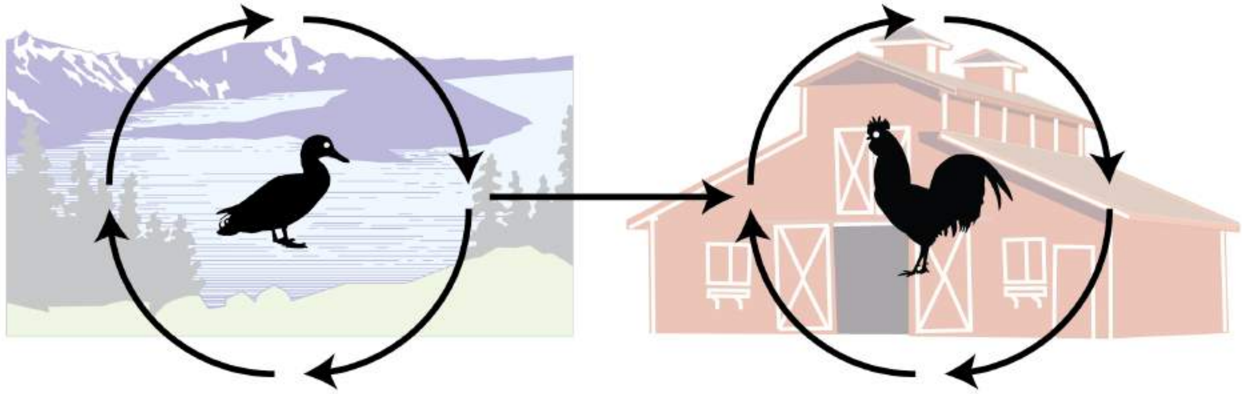
“Low Path” AI



16 / 18 known HAs
10 / 11 known NAs

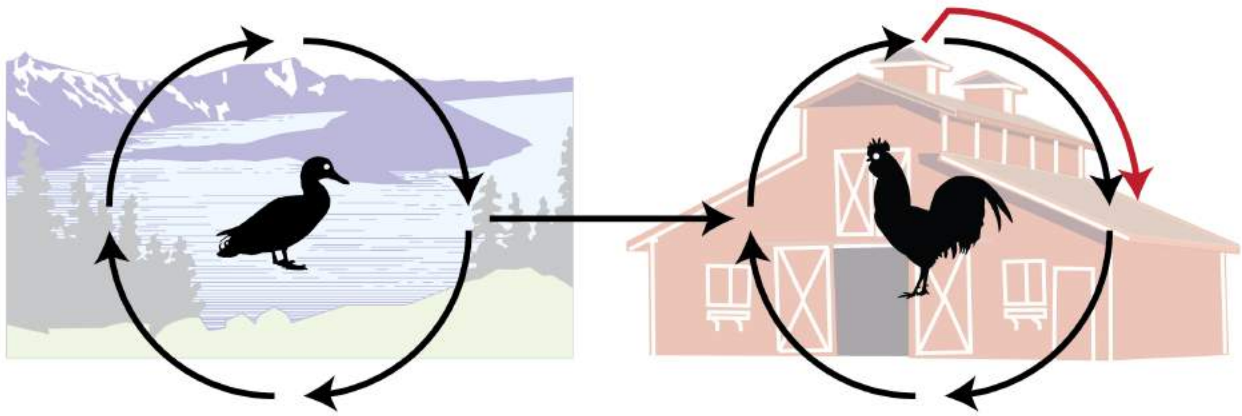
Infected ducks are “asymptomatic”

“Low Path” AI



“Low Path” AI

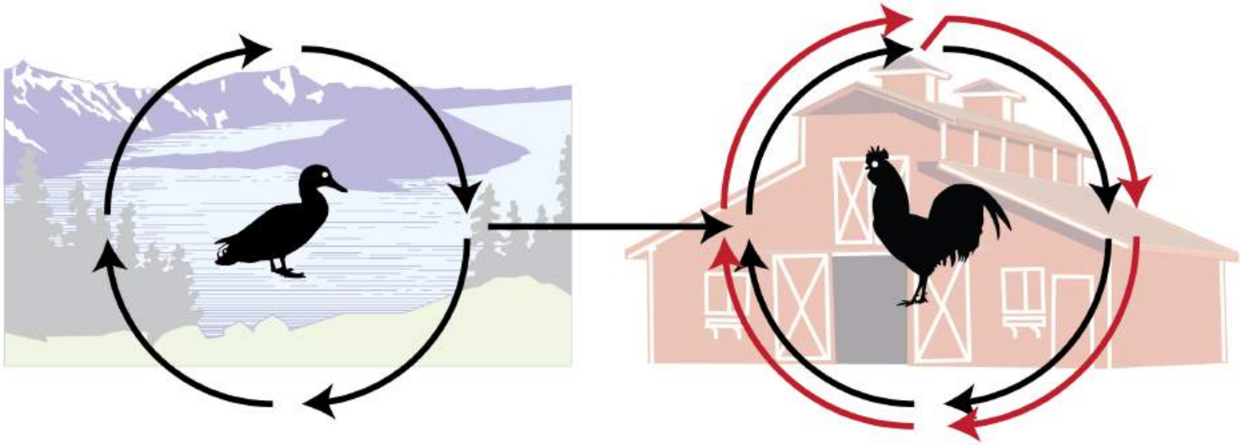
“High Path” AI



@sarahmhird

“Low Path” AI

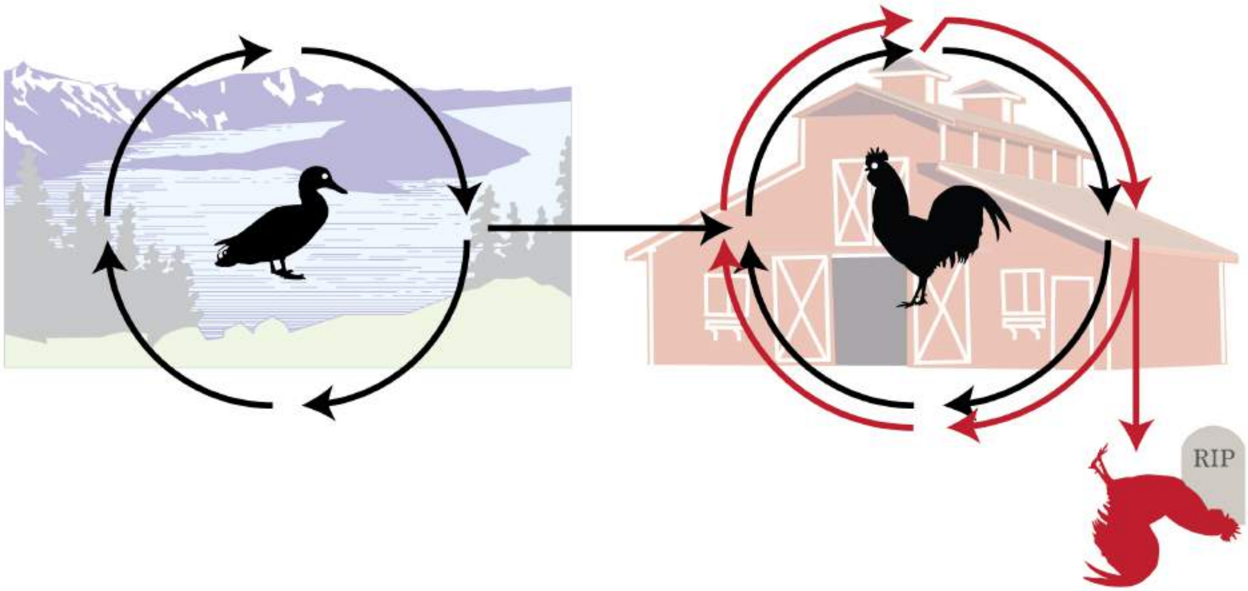
“High Path” AI



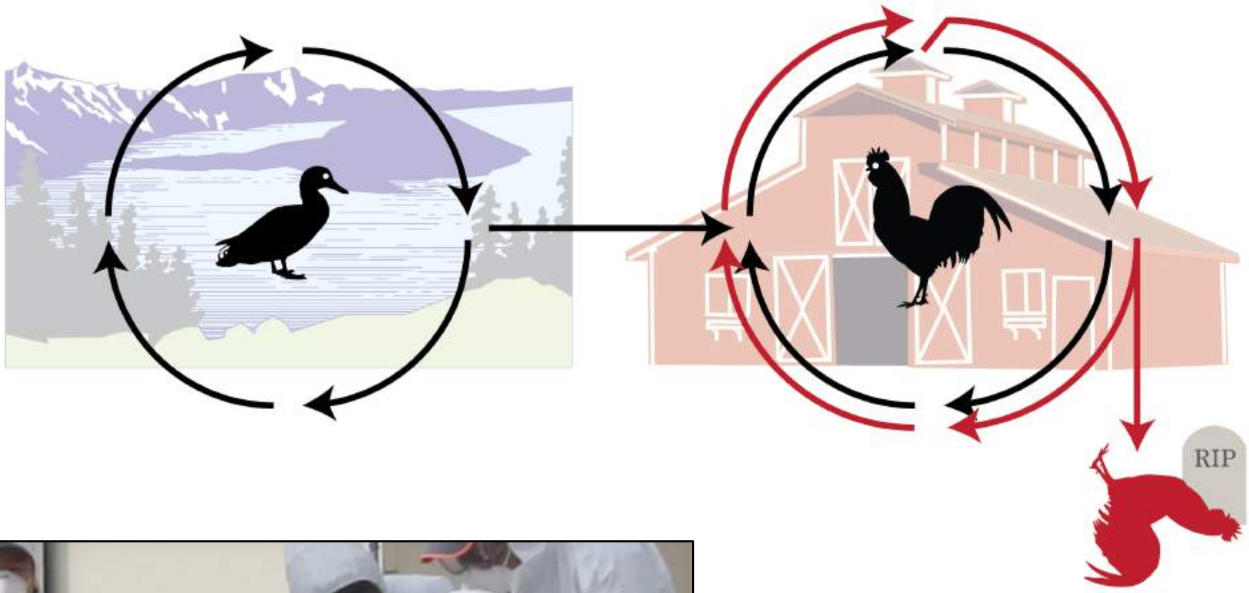
@sarahmhird

“Low Path” AI

“High Path” AI



“Low Path” AI
“High Path” AI




NEWS / SOUTH KOREA

Bird flu forces cull of 22 million poultry

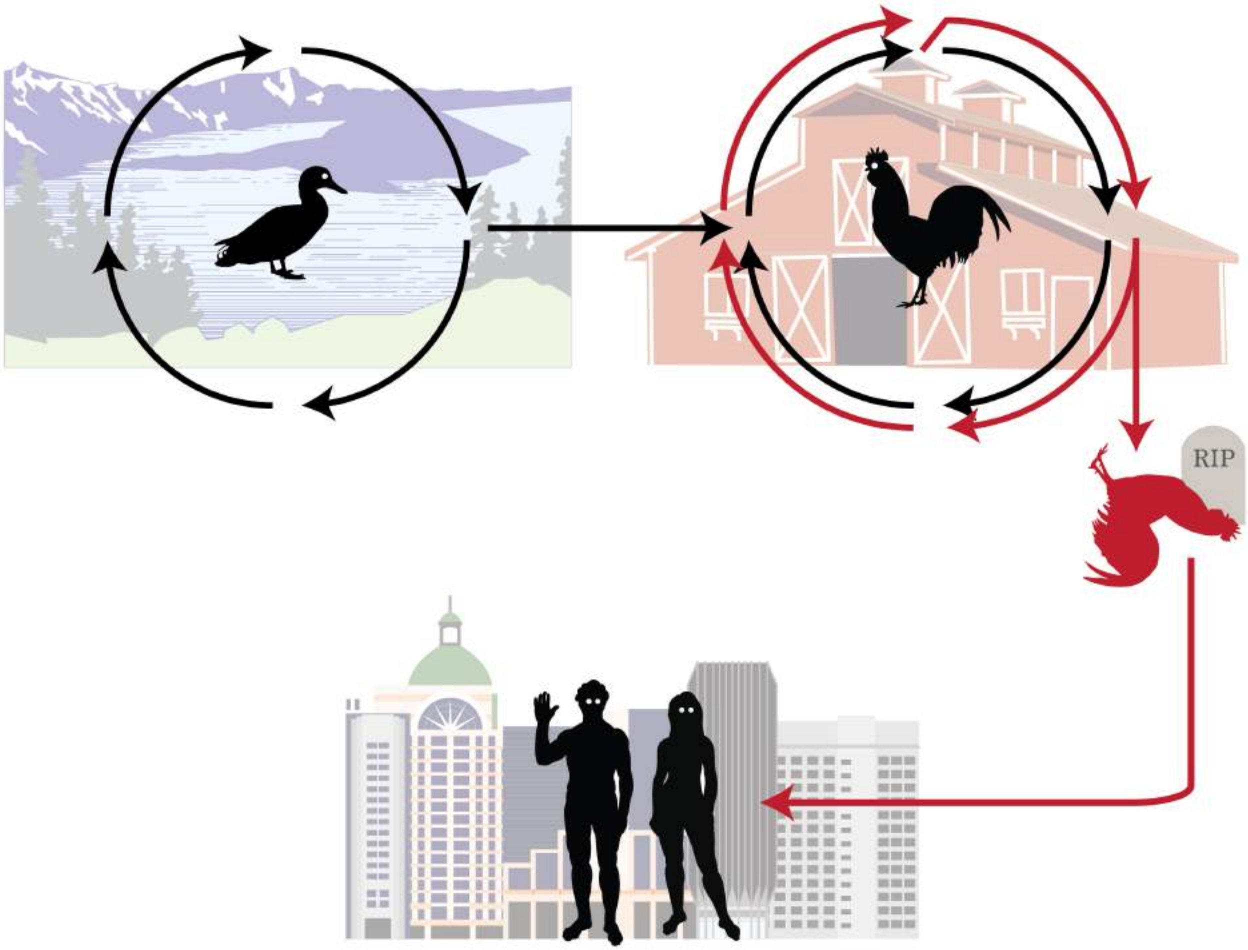
24 Dec 2016

South Korean authorities have culled more than

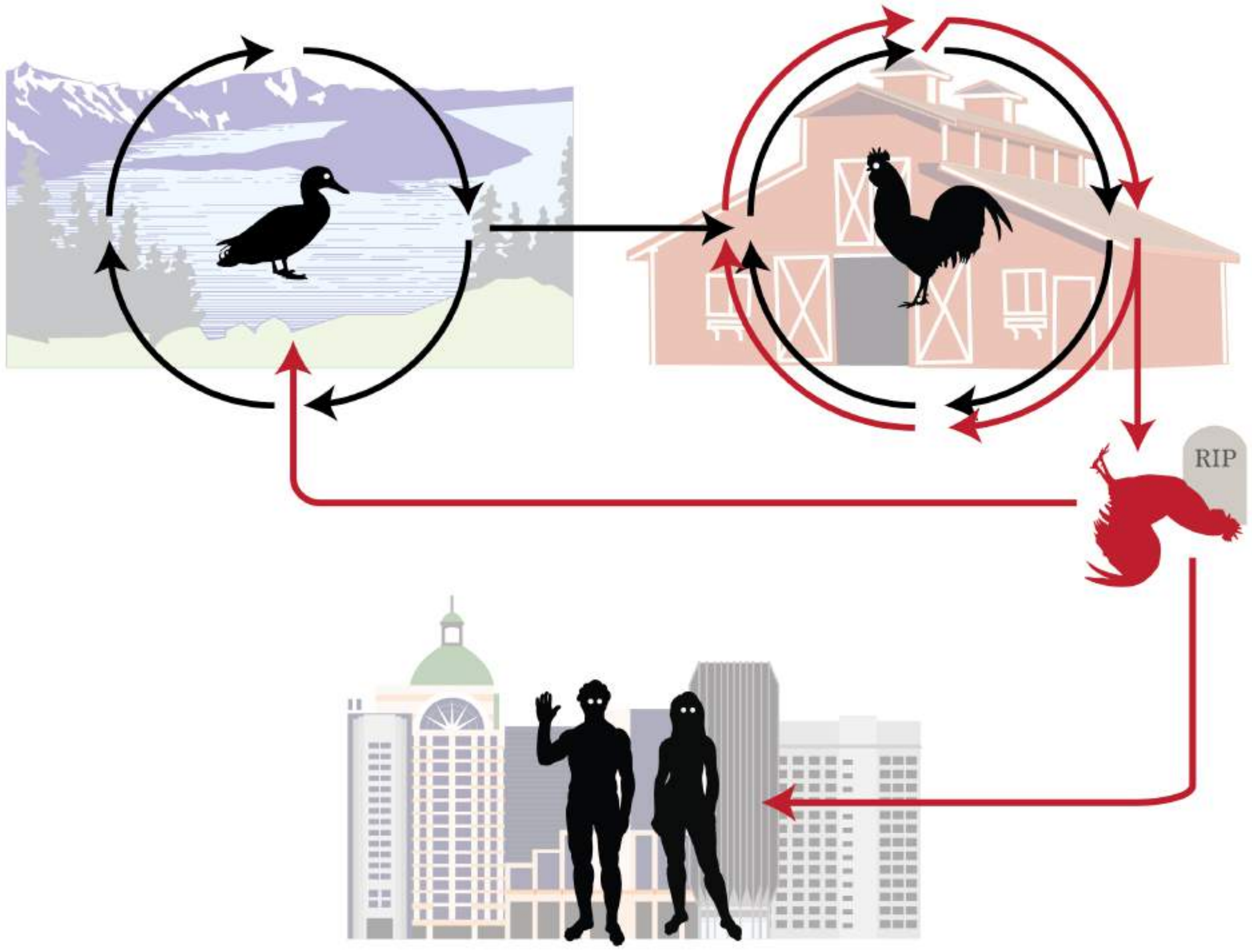
http://www.aljazeera.com/news/2016/12/birde-flu-force-cull-22-million-poultry-pieces-161224140943994.html

“Low Path” AI

“High Path” AI

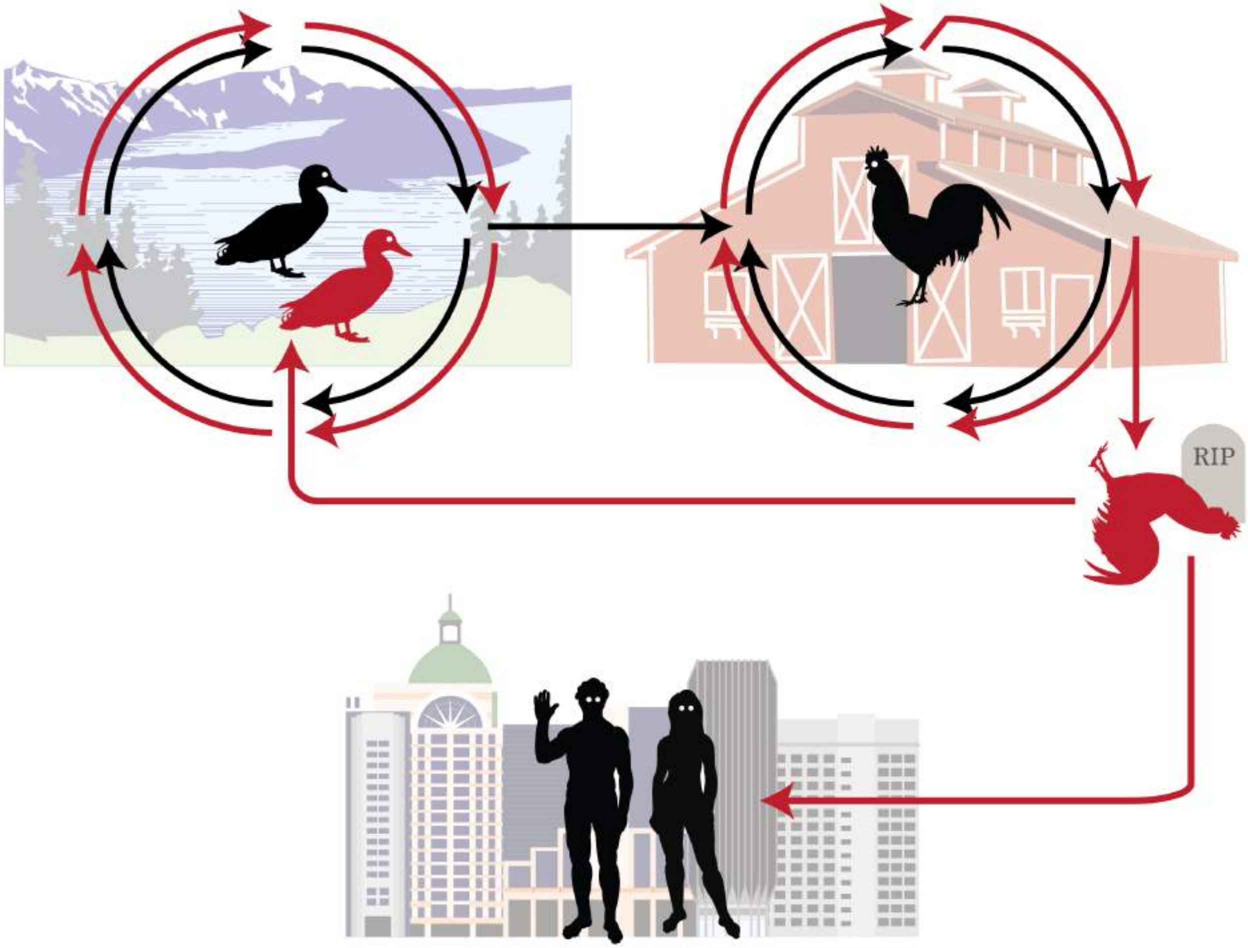


“Low Path” AI
“High Path” AI



@sarahmhird

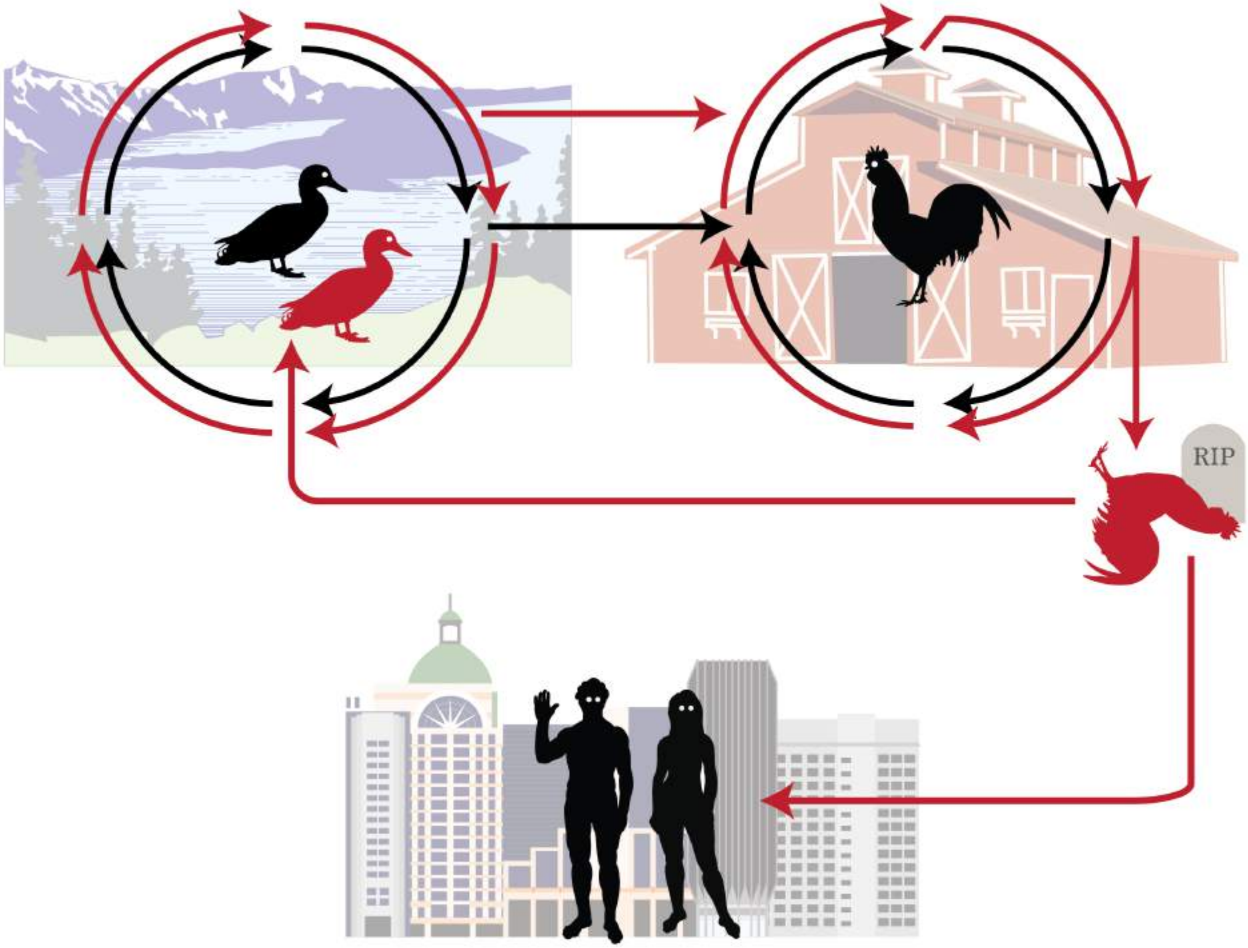
"Low Path" AI
"High Path" AI



@sarahmhird

"Low Path" AI

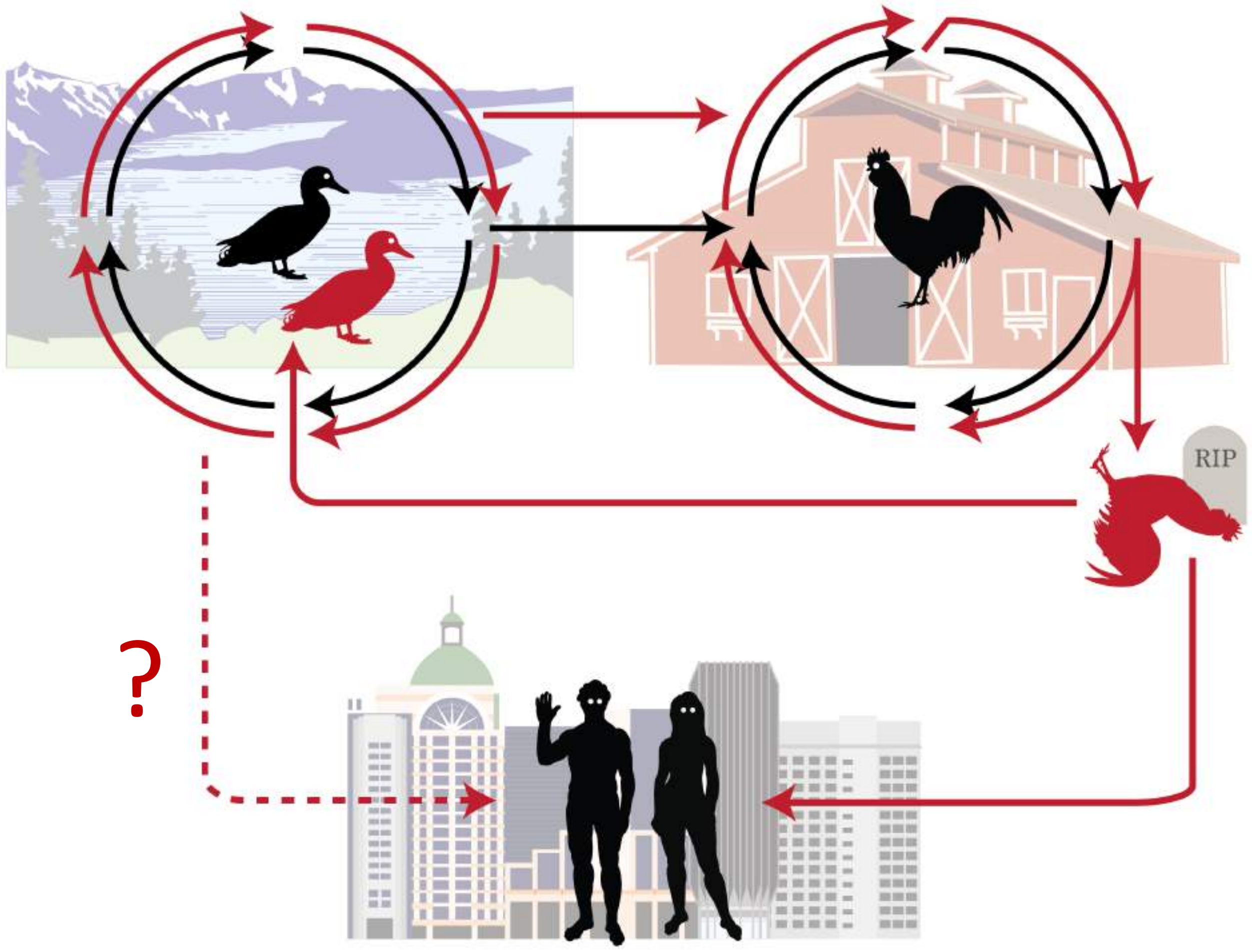
"High Path" AI



@sarahmhird

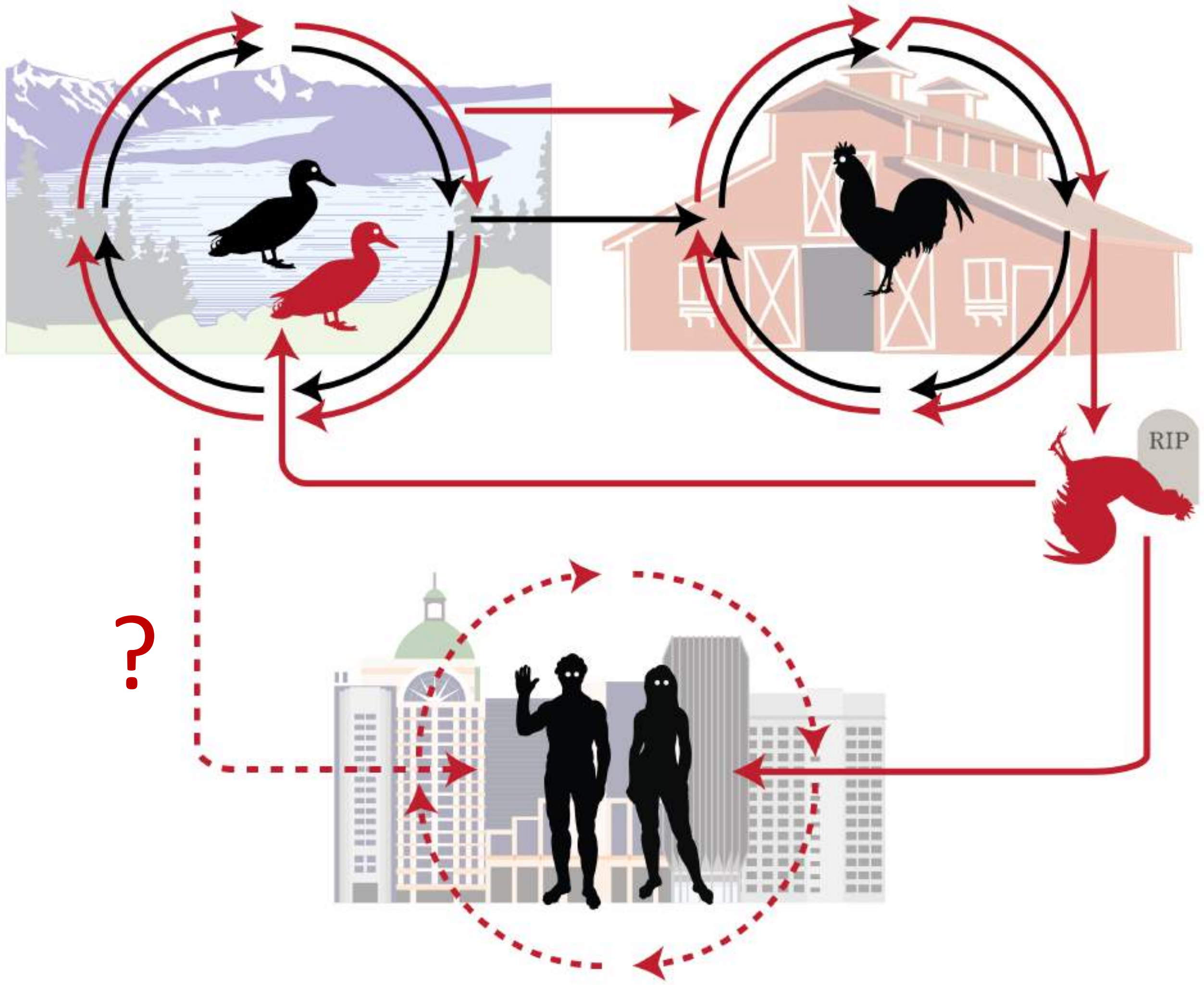
“Low Path” AI

“High Path” AI



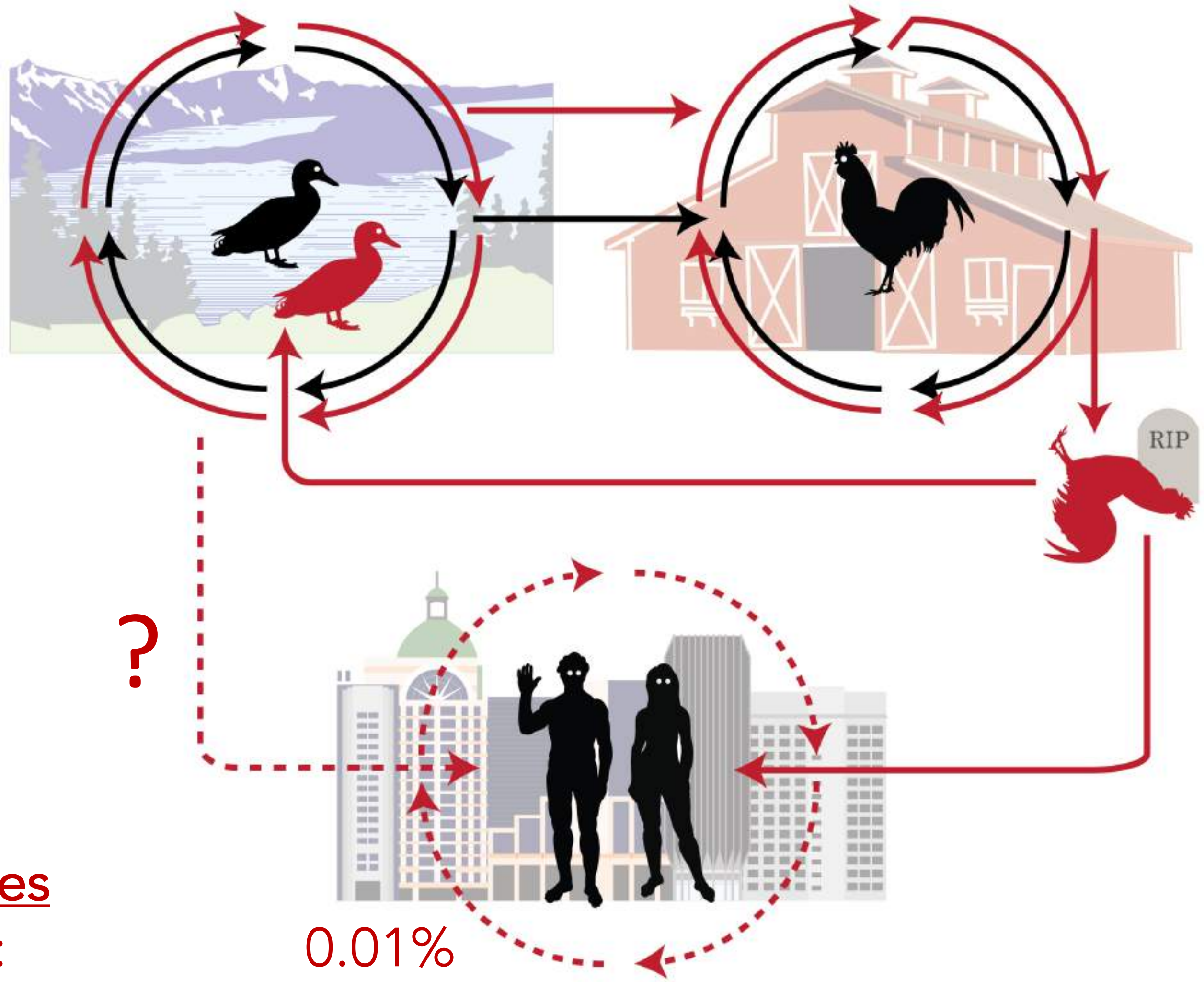
@sarahmhird

“Low Path” AI
“High Path” AI



@sarahmhird

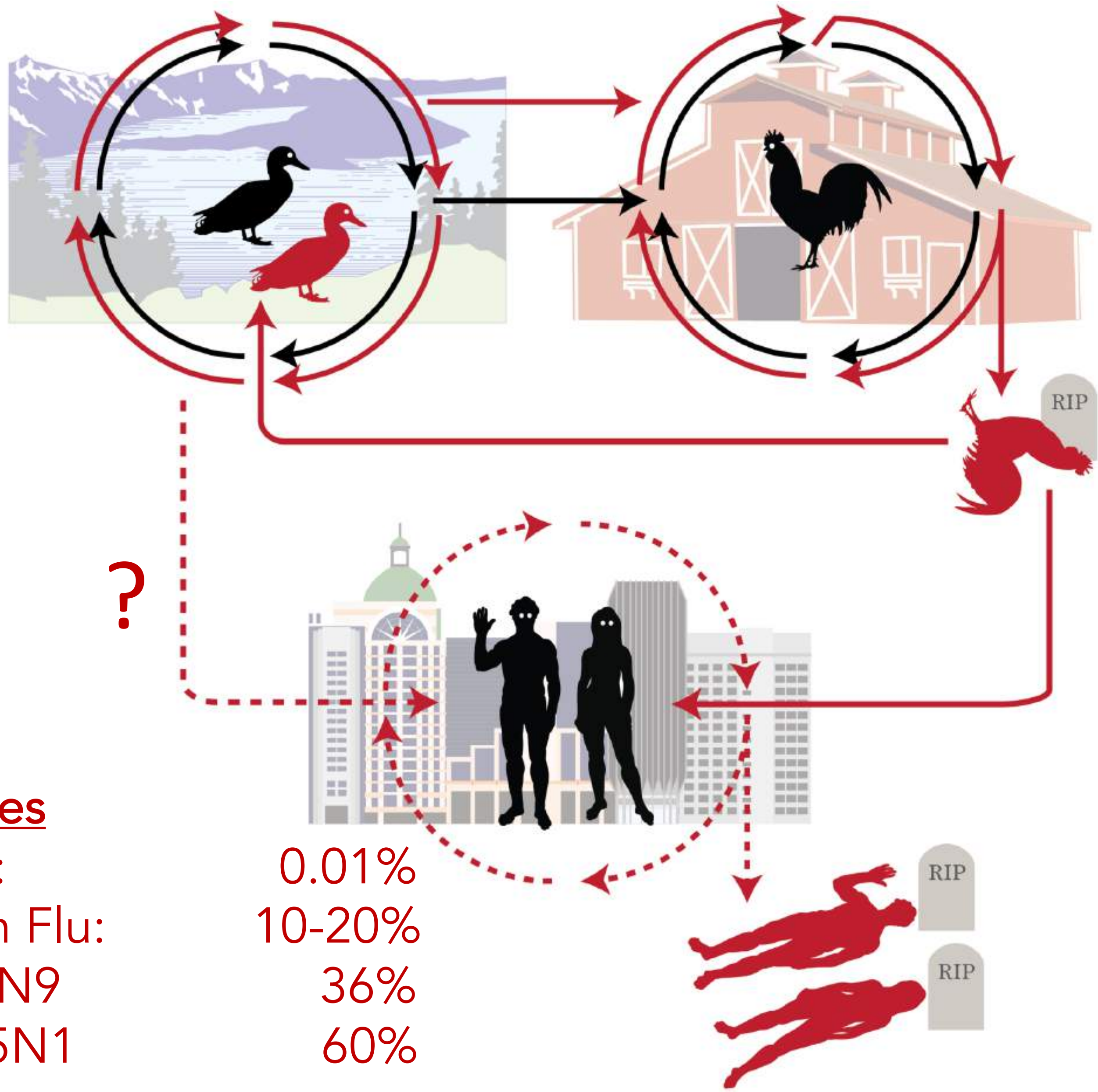
“Low Path” AI
“High Path” AI



Mortality Rates

Seasonal Flu:	0.01%
1918 Spanish Flu:	10-20%

“Low Path” AI
“High Path” AI



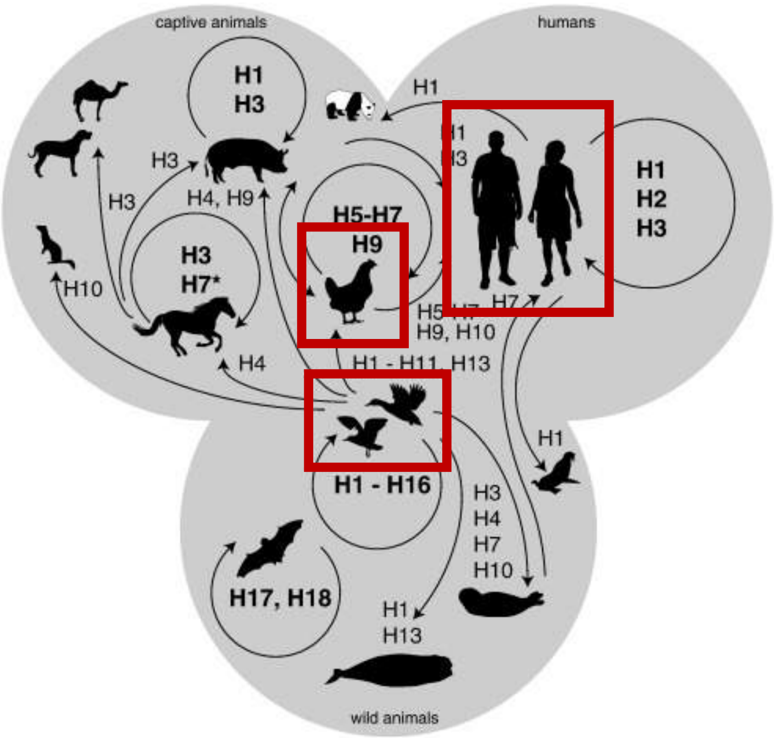
Mortality Rates

Seasonal Flu:	0.01%
1918 Spanish Flu:	10-20%
Low Path H7N9	36%
High Path H5N1	60%

http://www.who.int/influenza/human_animal_interface/EN_GIP_20150106CumulativeNumberH5N1cases_corrected.pdf?ua=1

INTRODUCTION

IAV is very infectious



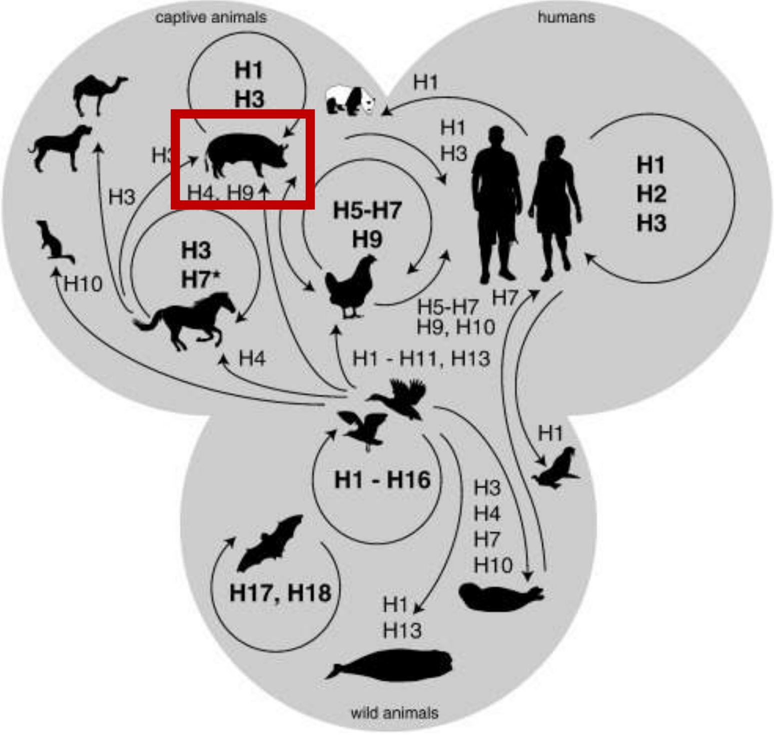
One health, multiple challenges: The inter-species transmission of influenza A virus

Kirsty R. Short [a,b], Mathilde Richard [a], Josanne H. Verhagen [a], Debby van Riel [a], Eefje J.A. Schrauwen [a], Judith M.A. van den Brand [a], Benjamin Mänz [a], Rogier Bodewes [a], Sander Herfst [a,*]

[a] *Department of Viroscience, Erasmus Medical Centre, The Netherlands*
[b] *School of Biomedical Sciences, University of Queensland, Brisbane, Australia*

INTRODUCTION

IAV is very infectious



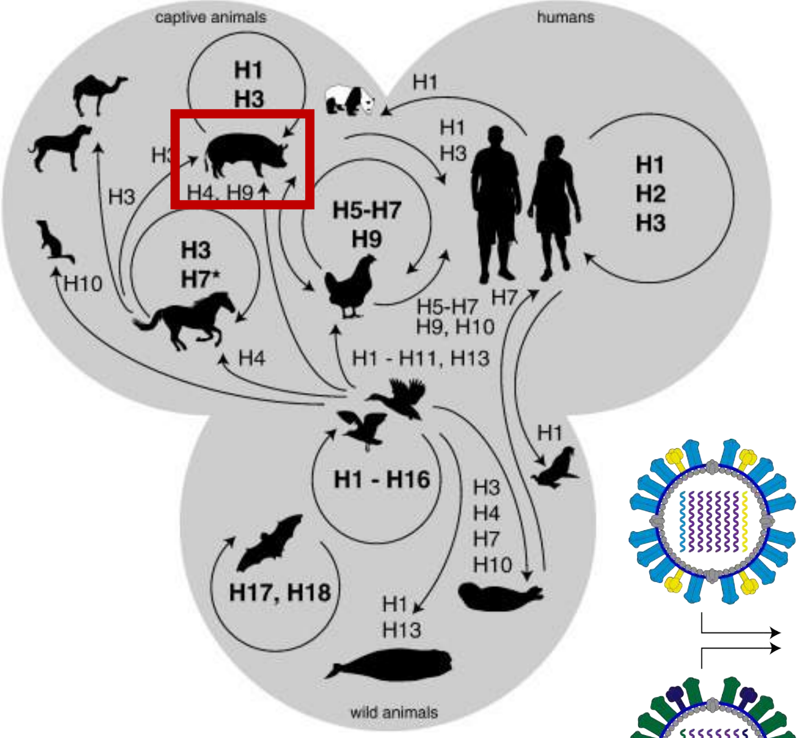
One health, multiple challenges: The inter-species transmission of influenza A virus

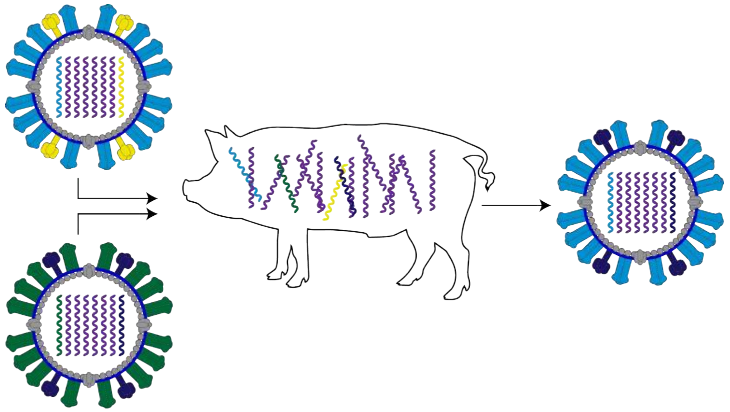
Kirsty R. Short [a,b], Mathilde Richard [a], Josanne H. Verhagen [a], Debby van Riel [a], Eefje J.A. Schrauwen [a], Judith M.A. van den Brand [a], Benjamin Mänz [a], Rogier Bodewes [a], Sander Herfst [a,*]

[a] *Department of Viroscience, Erasmus Medical Centre, The Netherlands*
[b] *School of Biomedical Sciences, University of Queensland, Brisbane, Australia*

INTRODUCTION

IAV is very infectious





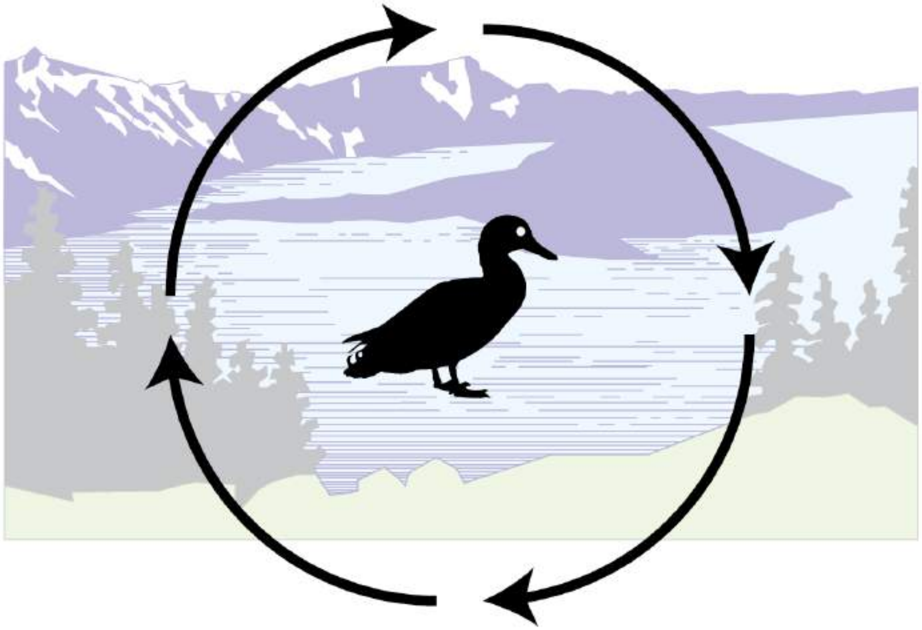
One health, multiple challenges: The inter-species transmission of influenza A virus

Kirsty R. Short [a,b], Mathilde Richard [a], Josanne H. Verhagen [a], Debby van Riel [a], Eefje J.A. Schrauwen [a], Judith M.A. van den Brand [a], Benjamin Mänz [a], Rogier Bodewes [a], Sander Herfst [a,*]

[a] *Department of Viroscience, Erasmus Medical Centre, The Netherlands*
[b] *School of Biomedical Sciences, University of Queensland, Brisbane, Australia*

INTRODUCTION

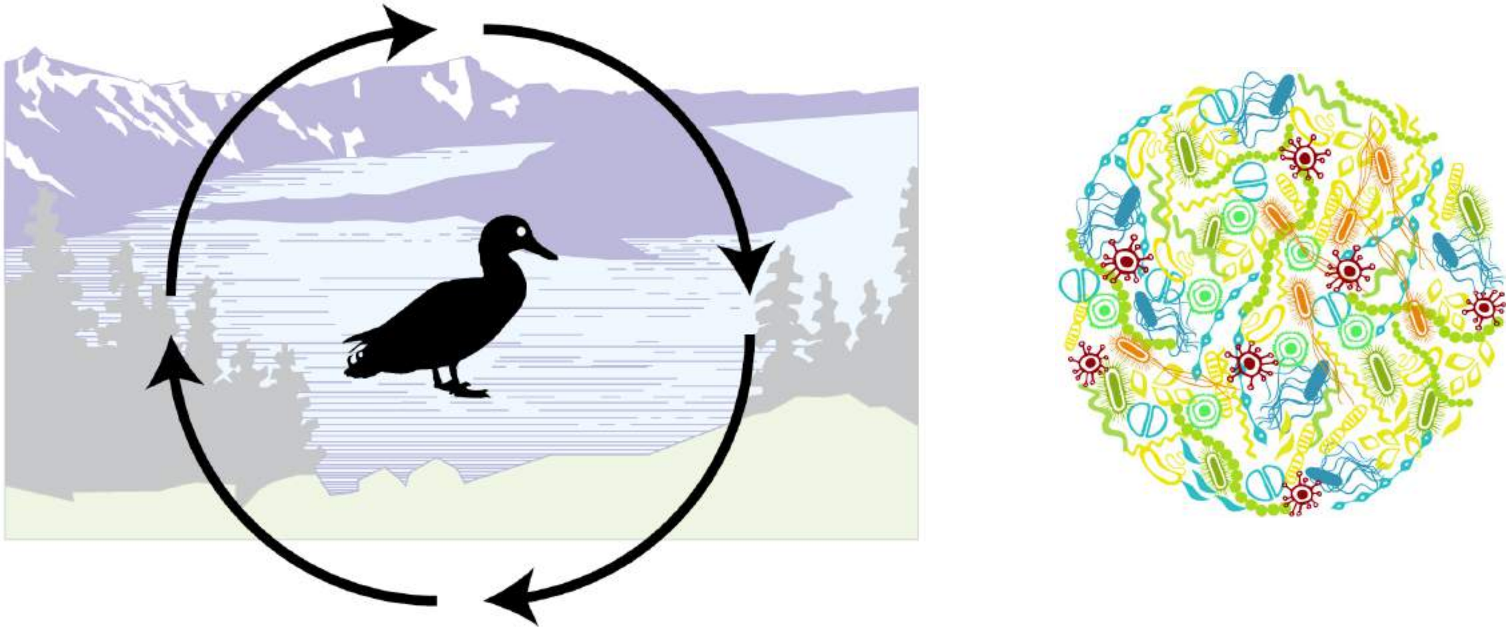
My focus: Wild birds in their natural habitat.



@sarahmhird

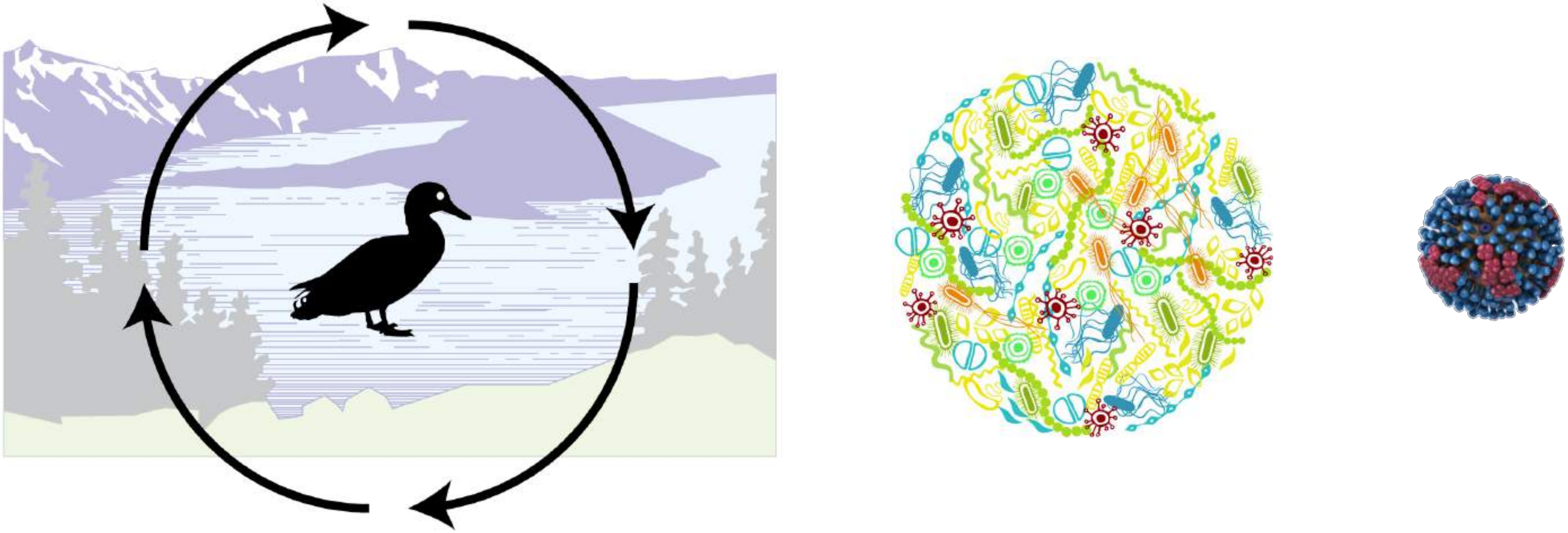
INTRODUCTION

My focus: Wild birds in their natural habitat.



INTRODUCTION

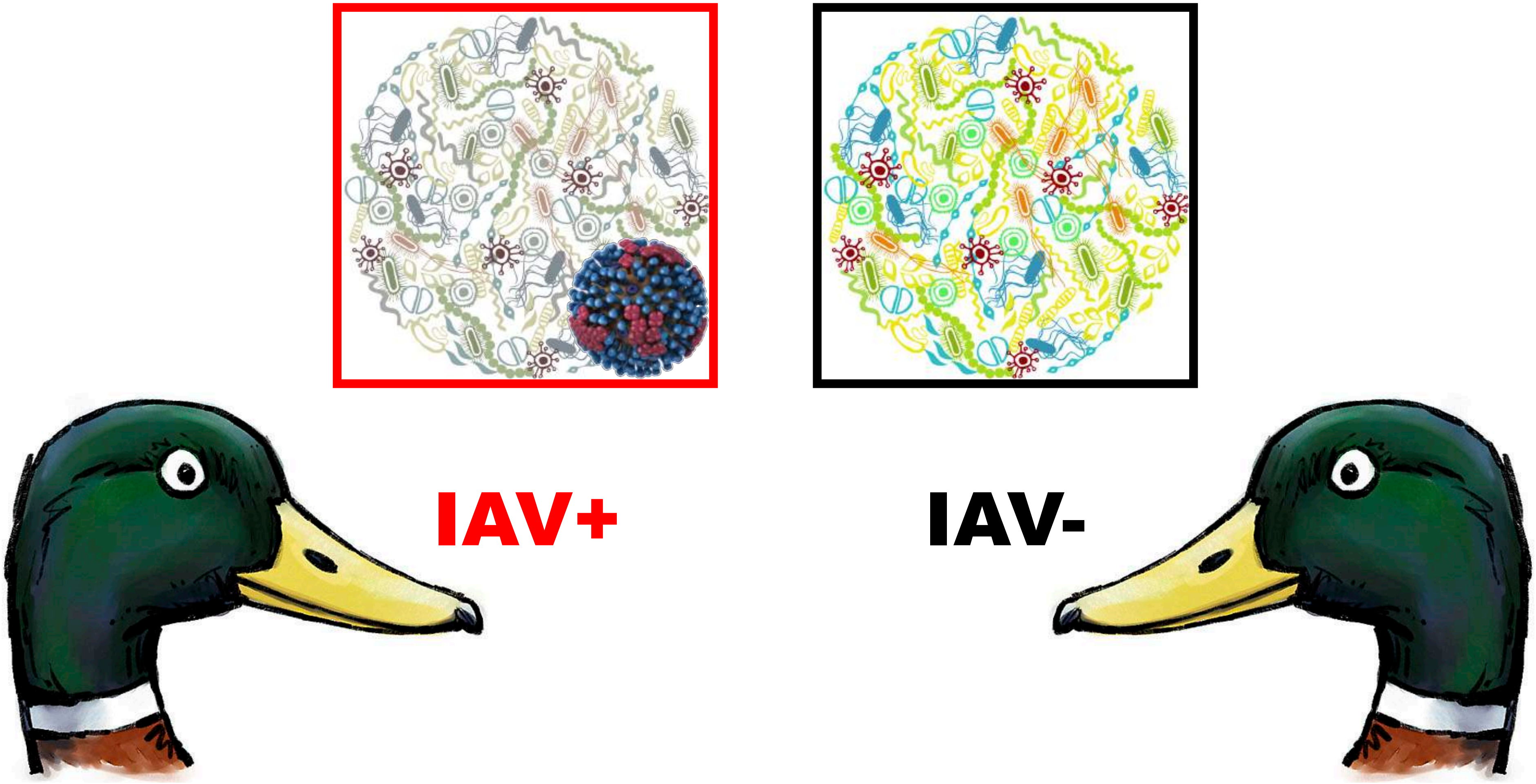
My focus: Wild birds in their natural habitat.



INTRODUCTION

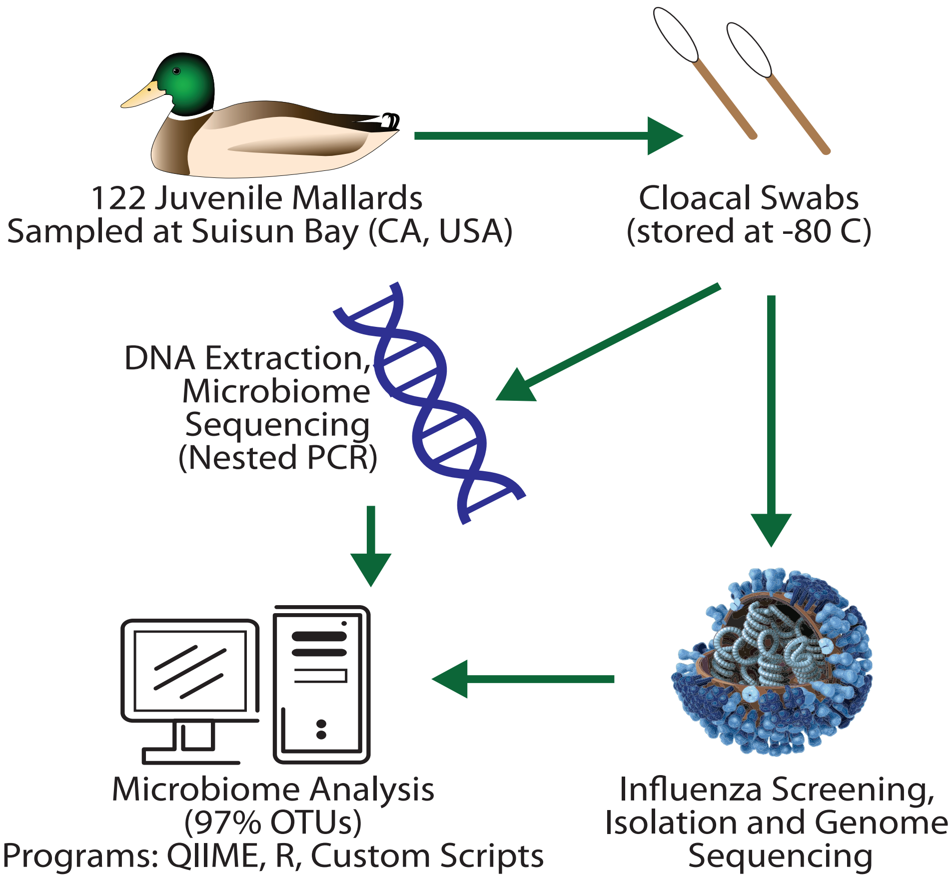
Question 1

Does IAV affect the microbiome of wild ducks?



Q1: METHODS

Mallards



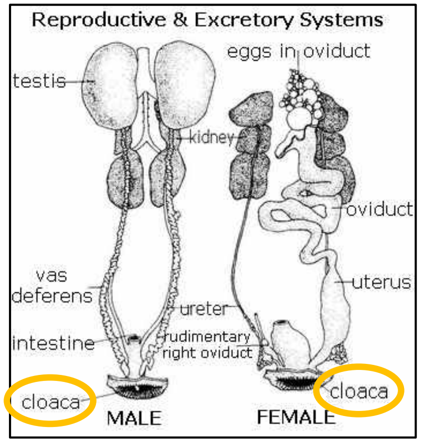
122 Juvenile Mallards
Sampled at Suisun Bay (CA, USA)

Cloacal Swabs
(stored at -80 C)

DNA Extraction,
Microbiome
Sequencing
(Nested PCR)

Microbiome Analysis
(97% OTUs)
Programs: QIIME, R, Custom Scripts

Influenza Screening,
Isolation and Genome
Sequencing



Reproductive & Excretory Systems

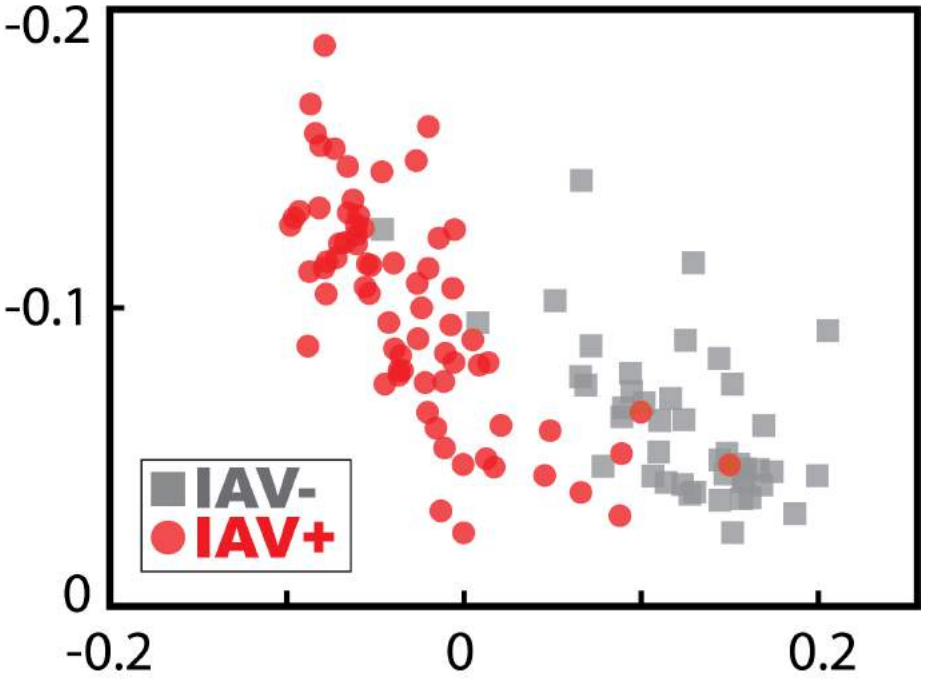
Dr. Walter
Boyce

UCDAVIS
VETERINARY MEDICINE

@sarahmhird

Q1: RESULTS

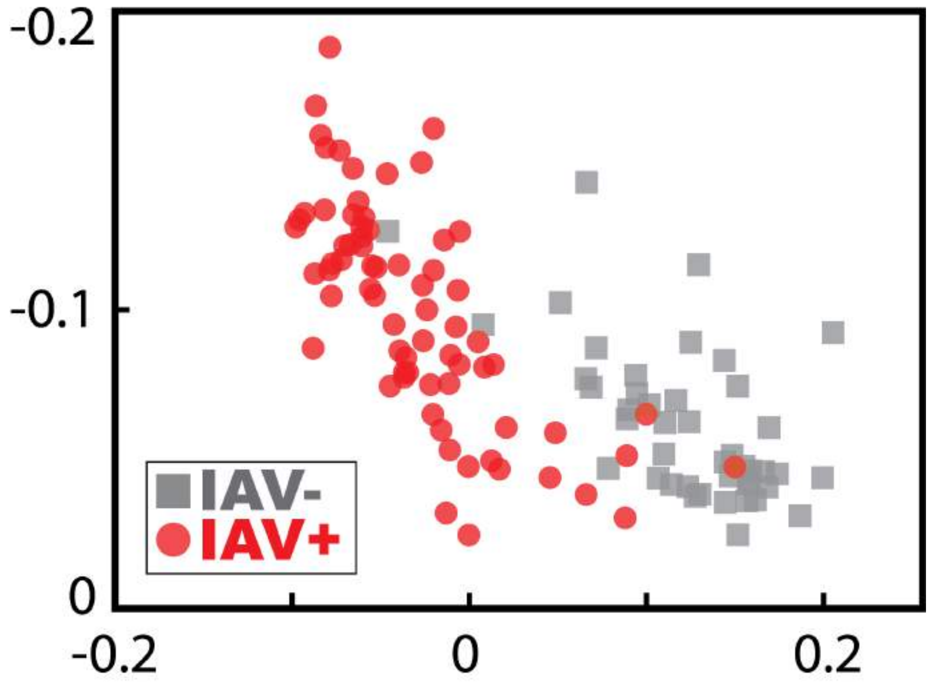
IAV+ and **IAV-** bird microbiomes are distinct.

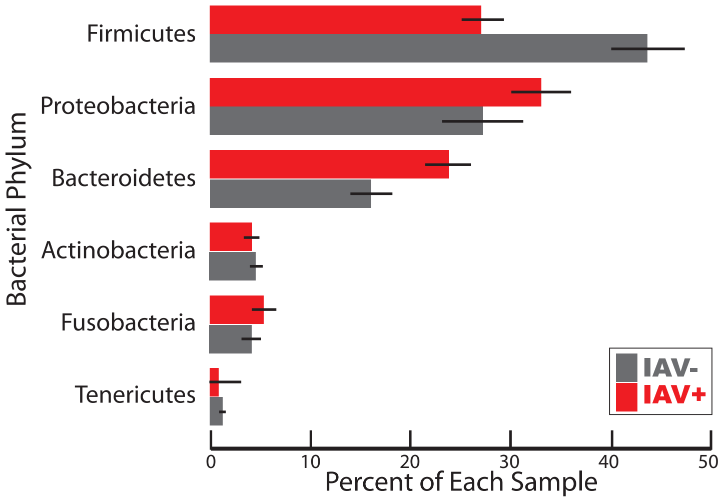


@sarahmhird

Q1: RESULTS

IAV+ and IAV- bird microbiomes are distinct.

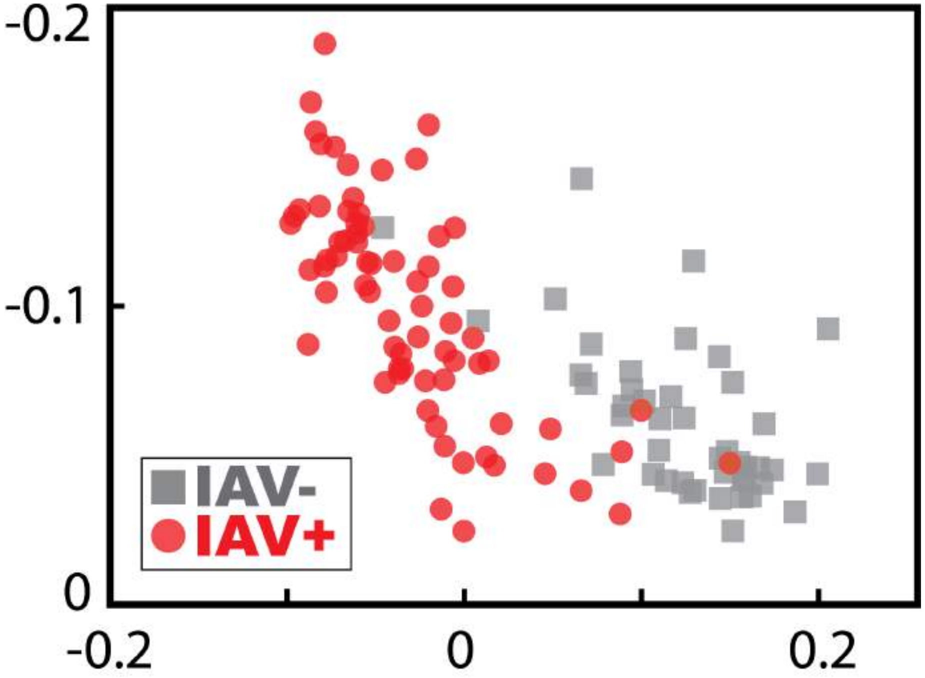


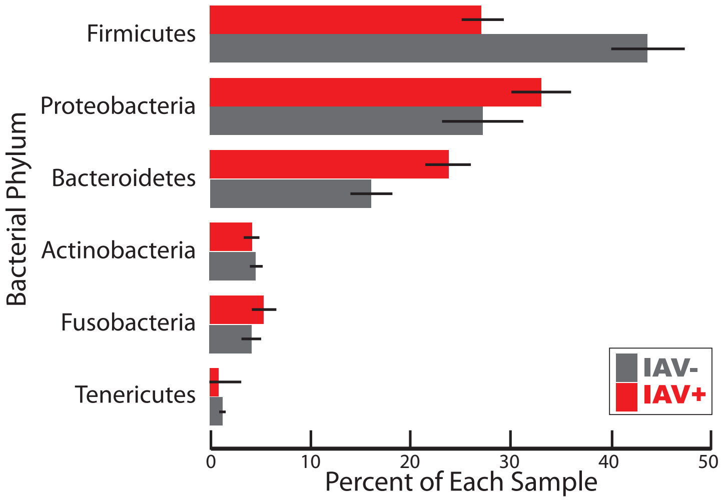


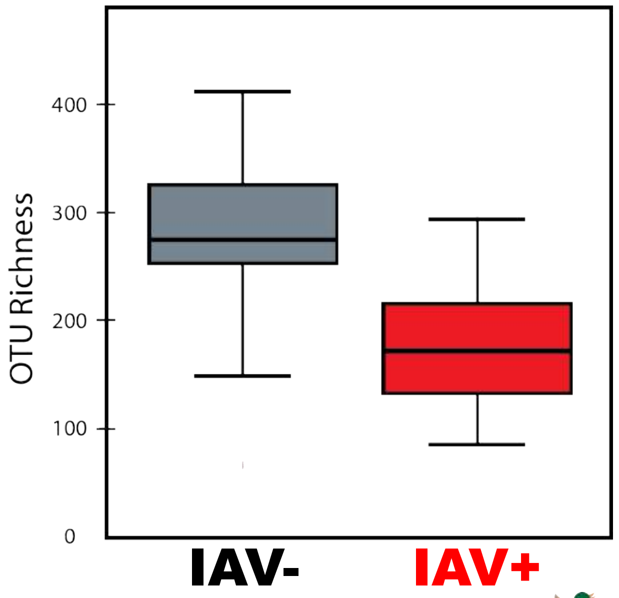
@sarahmhird

Q1: RESULTS

IAV+ and IAV- bird microbiomes are distinct.



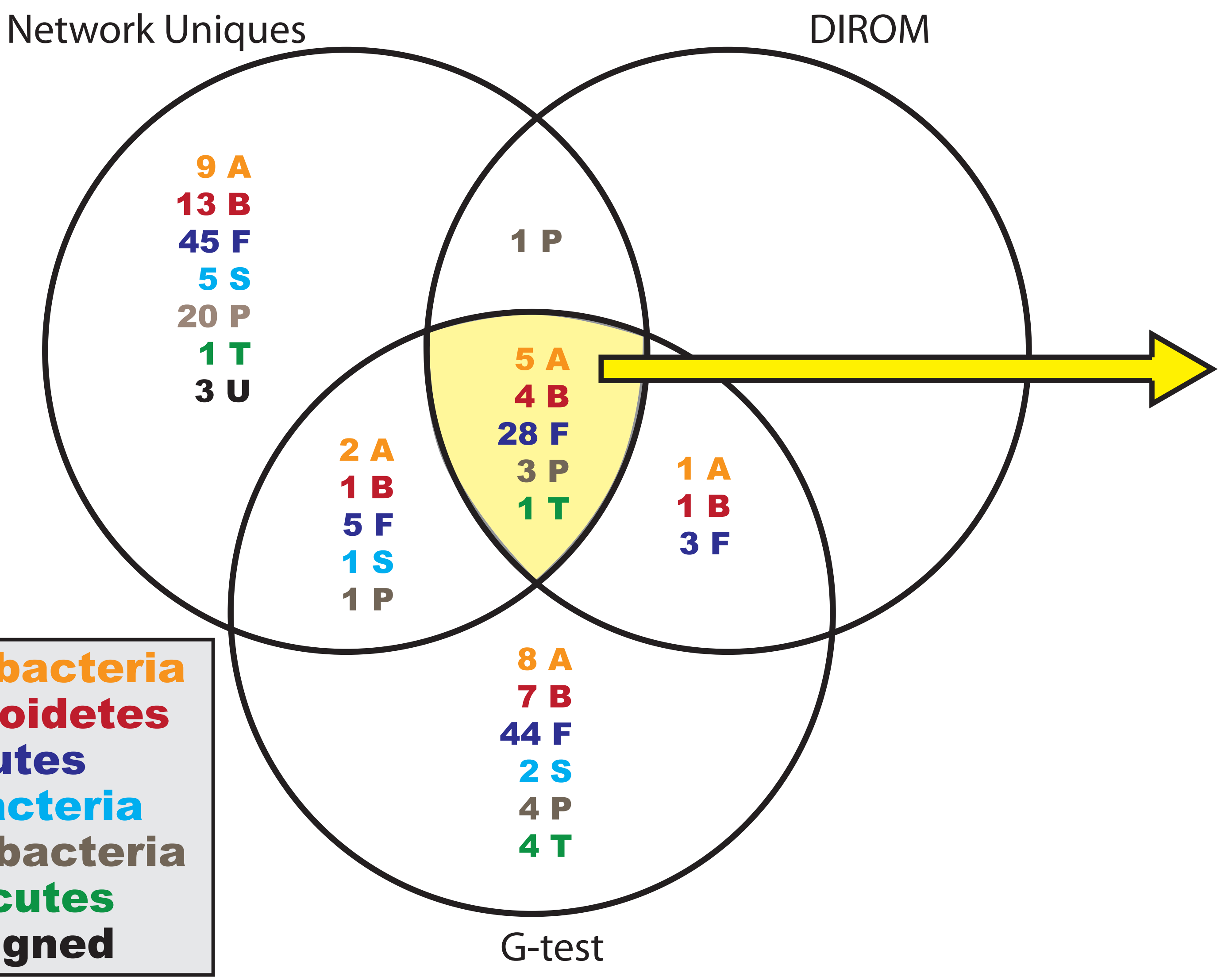




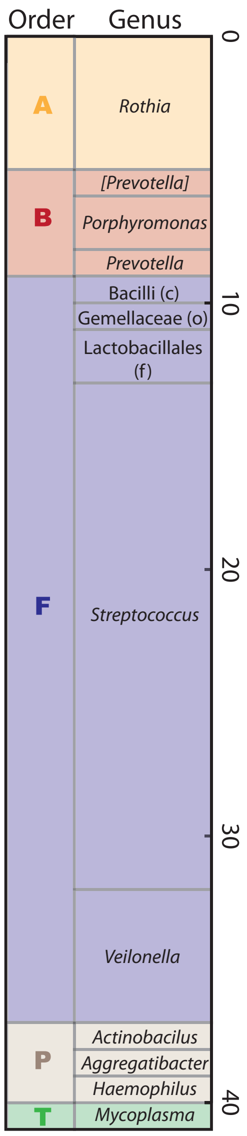
@sarahmhird

Q1: RESULTS

41 OTUs were highly significant.



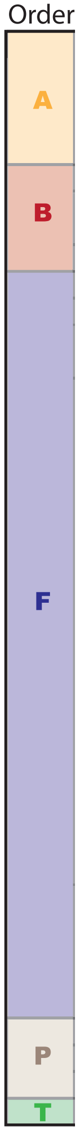
A - Actinobacteria
B - Bacteroidetes
F - Firmicutes
S - Fusobacteria
P - Proteobacteria
T - Tenericutes
U - Unassigned



@sarahmhird

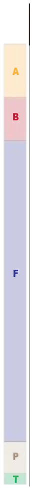
Q1: RESULTS

41 OTUs were highly significant.



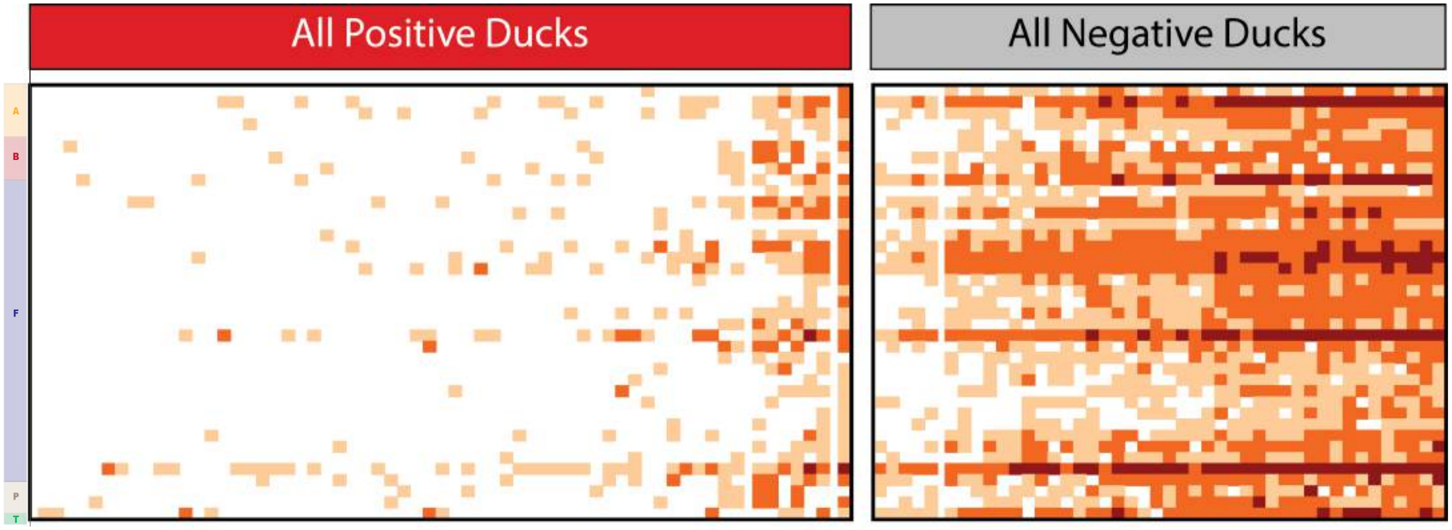
Q1: RESULTS

41 OTUs were highly significant.



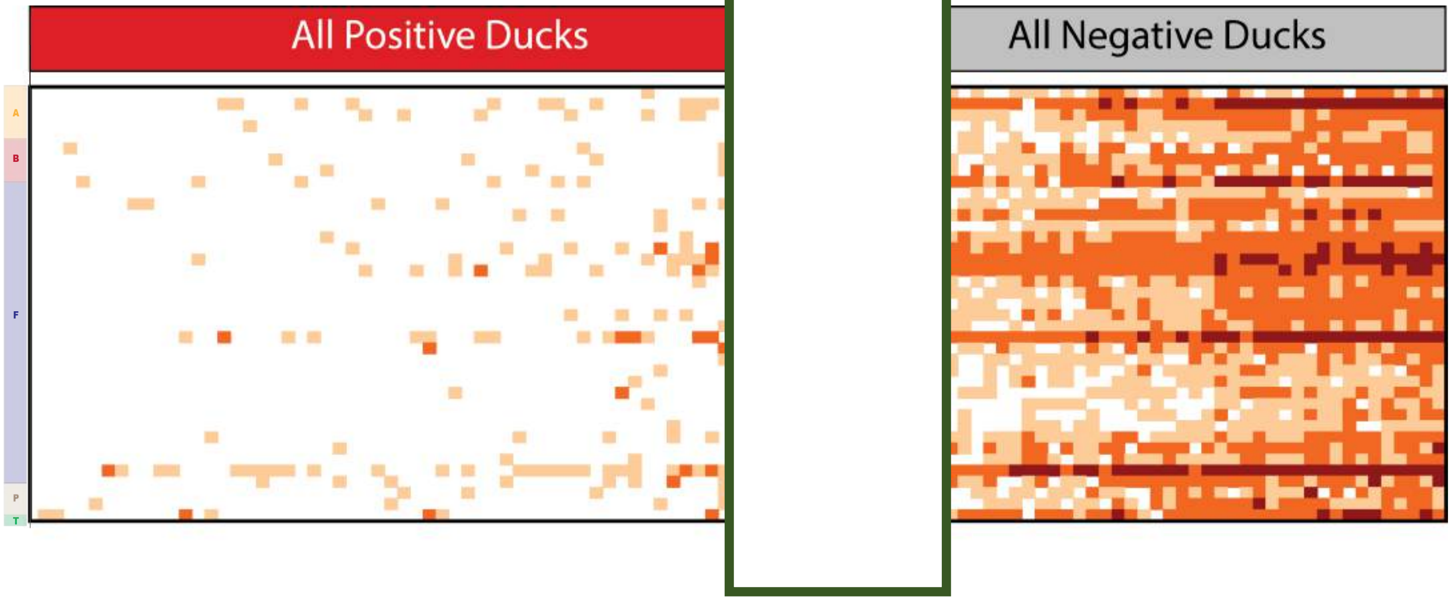
Q1: RESULTS

Fairly distinct groups



Q1: RESULTS

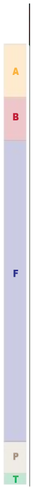
Fairly distinct groups

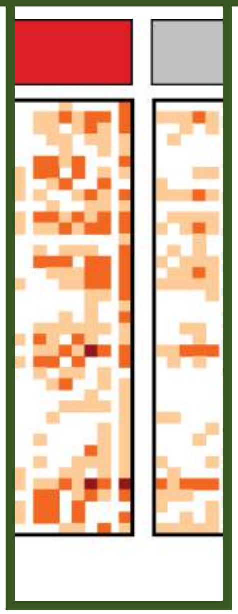


@sarahmhird

Q1: RESULTS

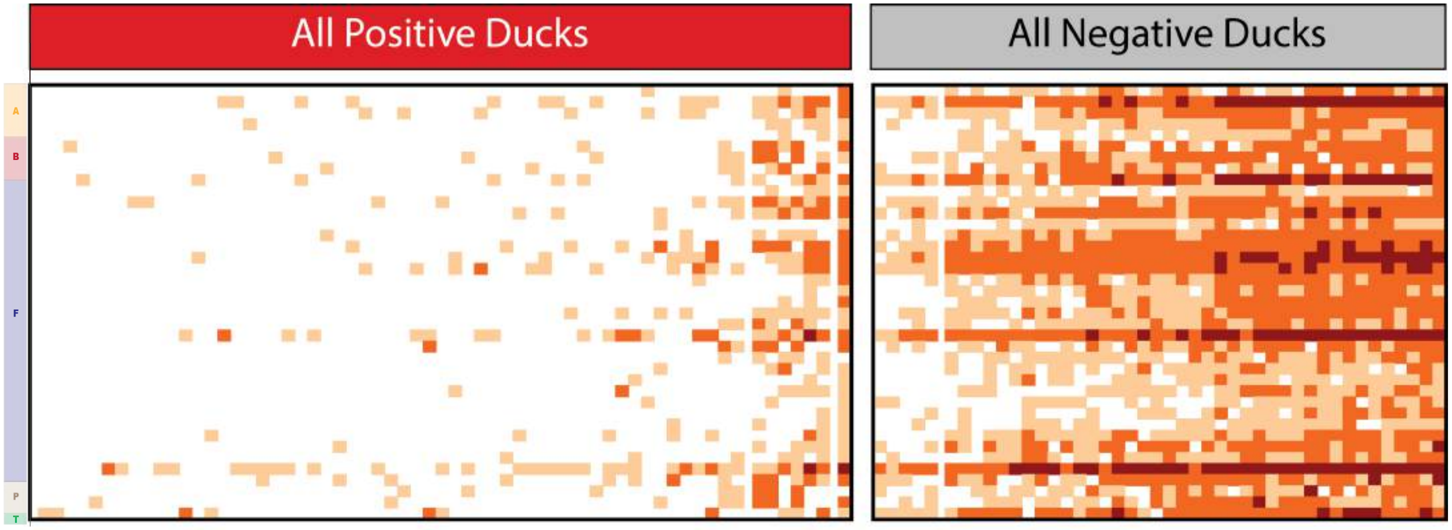
Fairly distinct groups

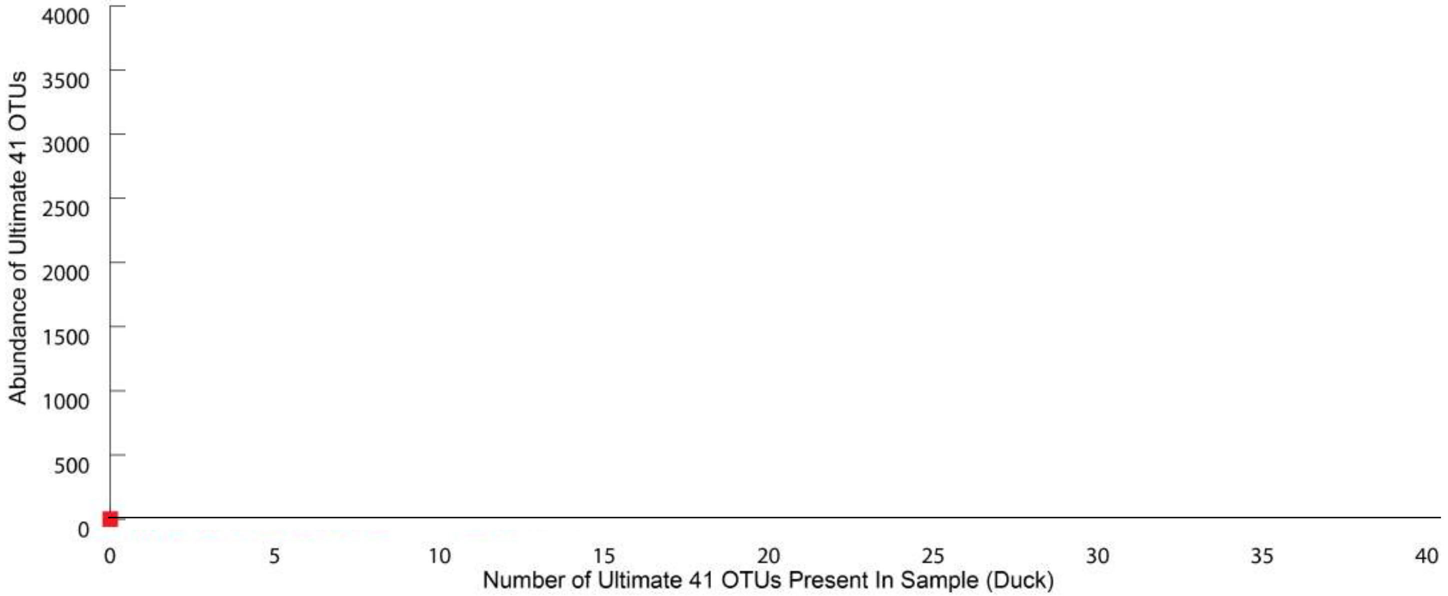




Q1: RESULTS

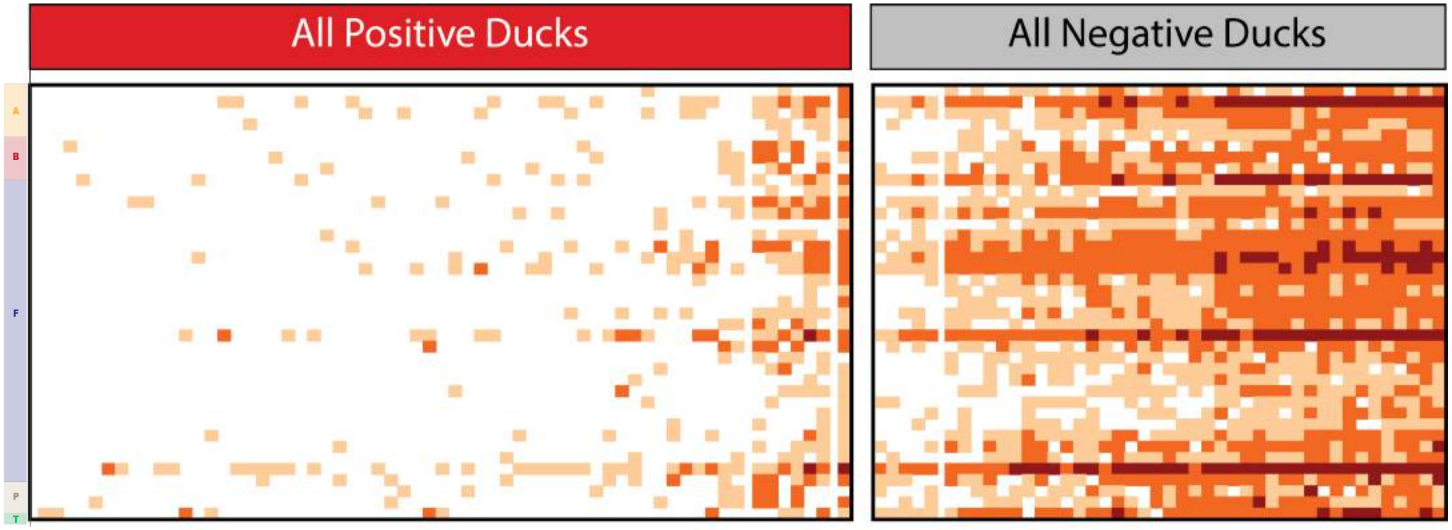
Abundance of 41 OTUs

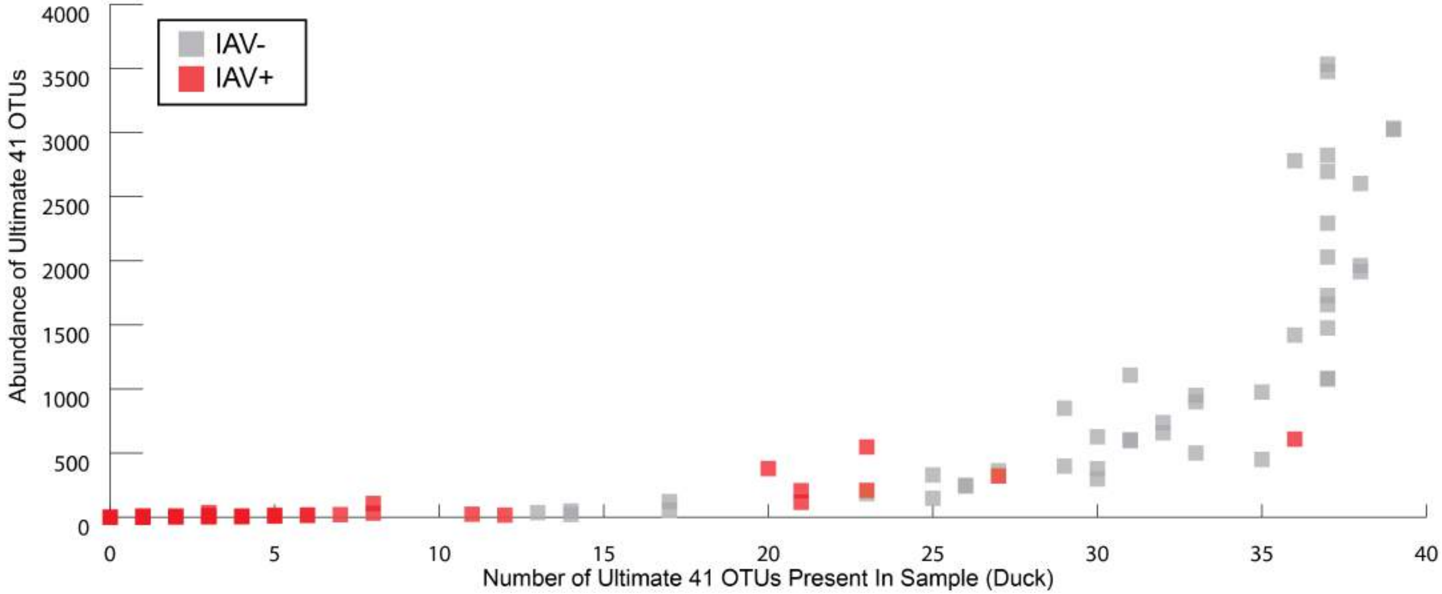




Q1: RESULTS

Abundance of 41 OTUs

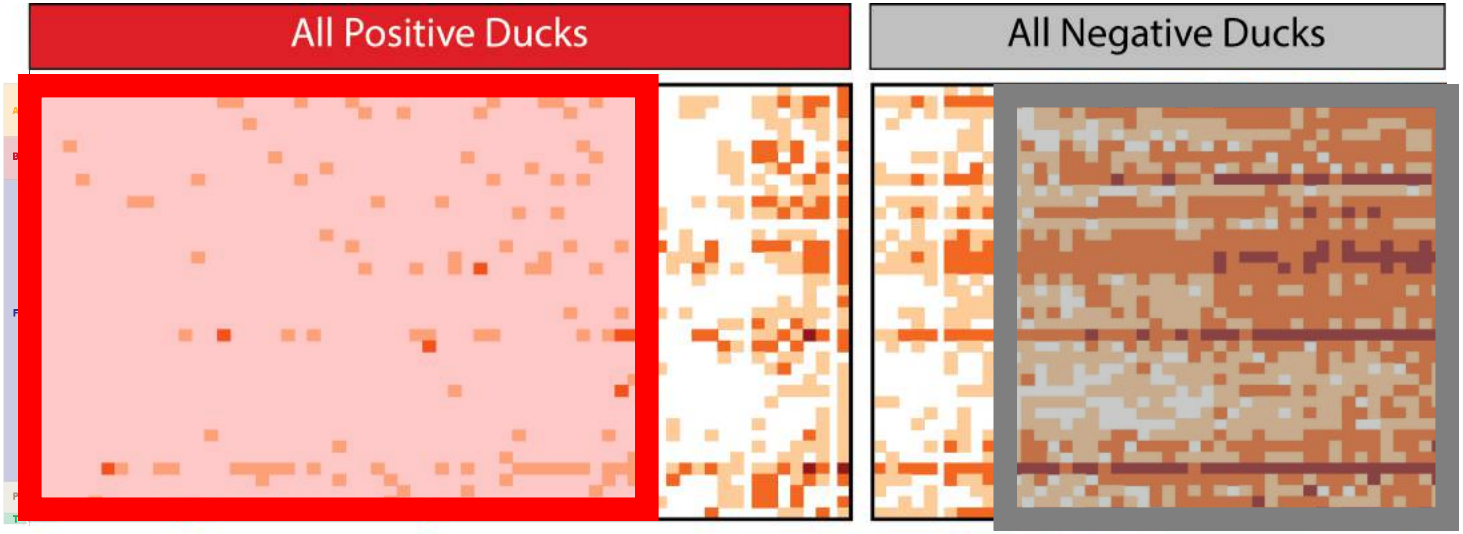


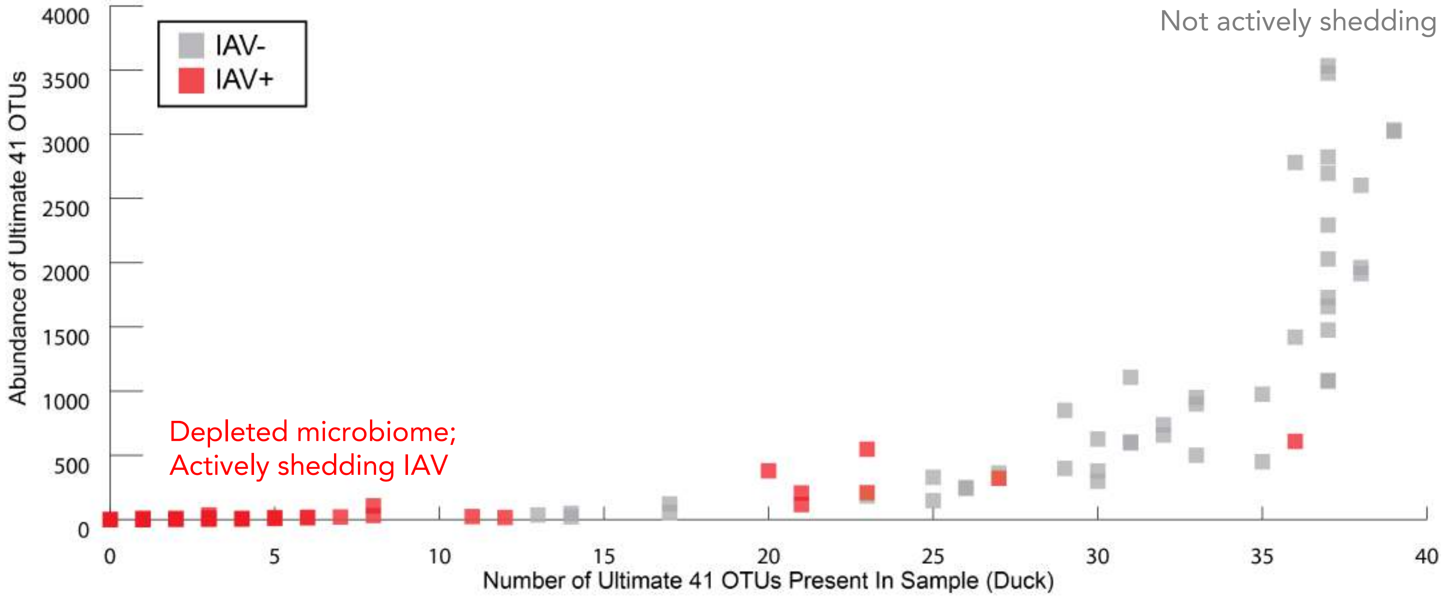


@sarahmhird

Q1: RESULTS

Abundance of 41 OTUs



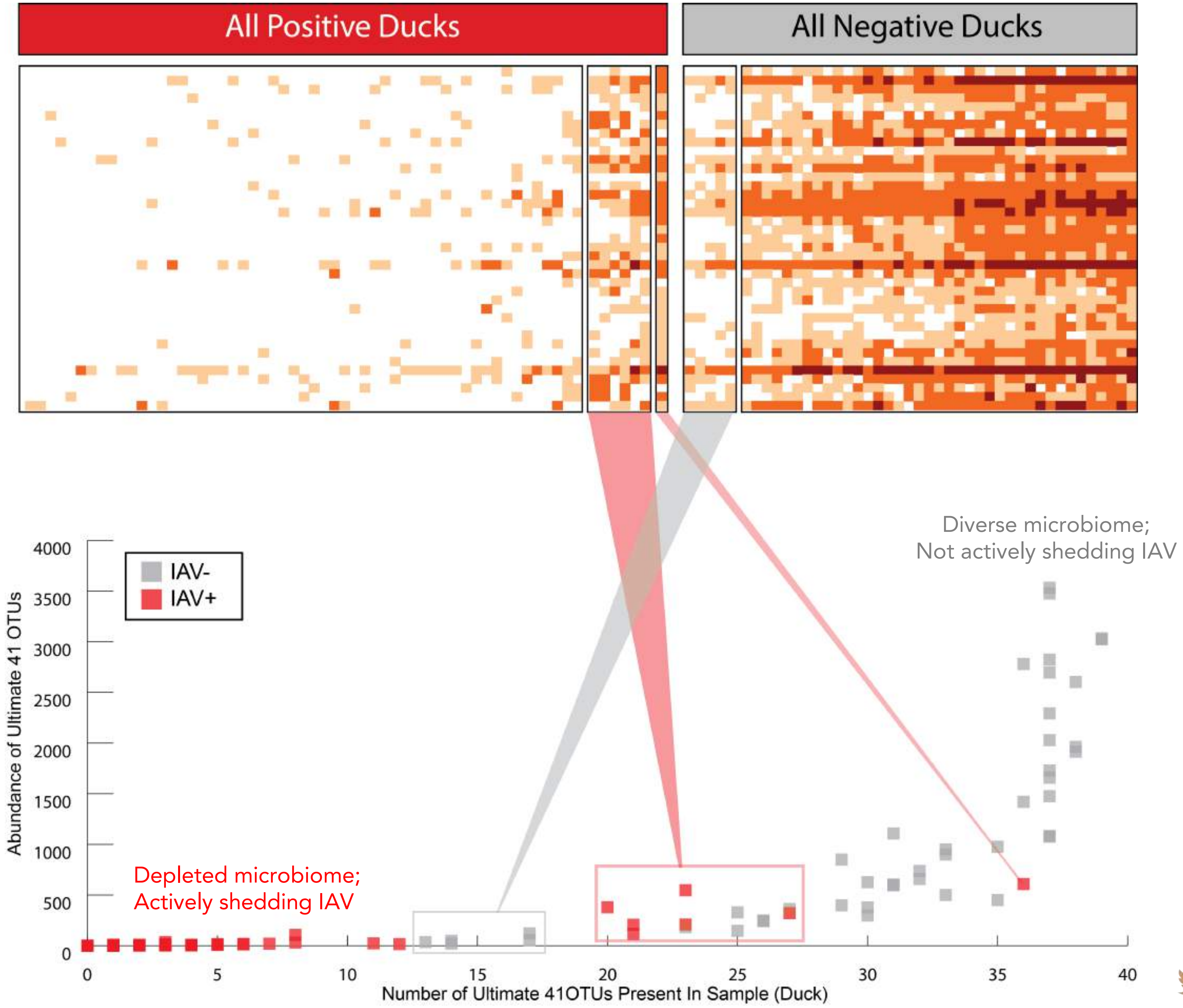


Diverse microbiome;
Not actively shedding IAV

Depleted microbiome;
Actively shedding IAV

Q1: RESULTS

Abundance of 41 OTUs



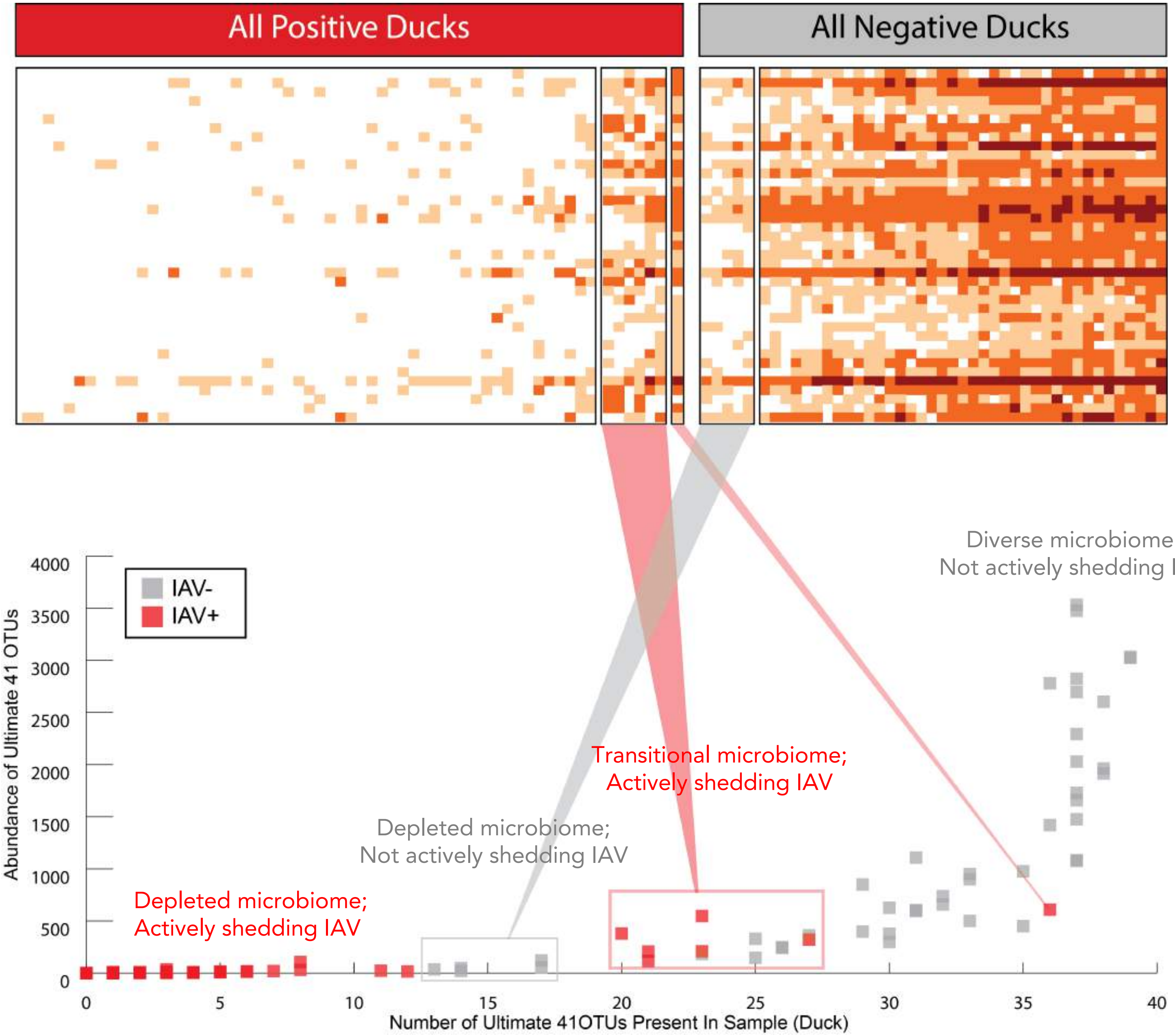
Diverse microbiome;
Not actively shedding IAV

Depleted microbiome;
Actively shedding IAV

@sarahmhird

Q1: RESULTS

Hypothesis: A spectrum of infection response



Diverse microbiome;
Not actively shedding IAV

Transitional microbiome;
Actively shedding IAV

Depleted microbiome;
Not actively shedding IAV

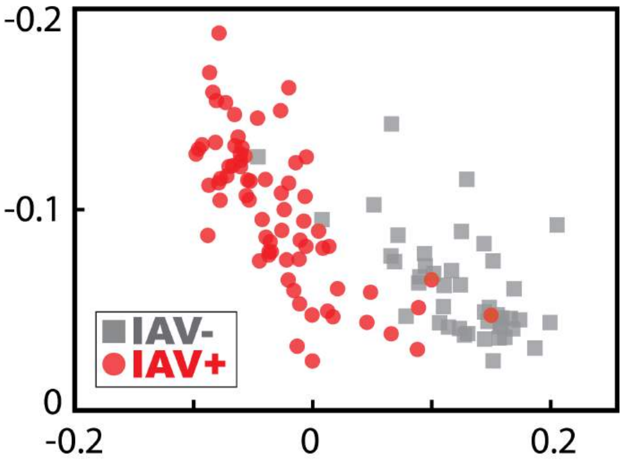
Depleted microbiome;
Actively shedding IAV

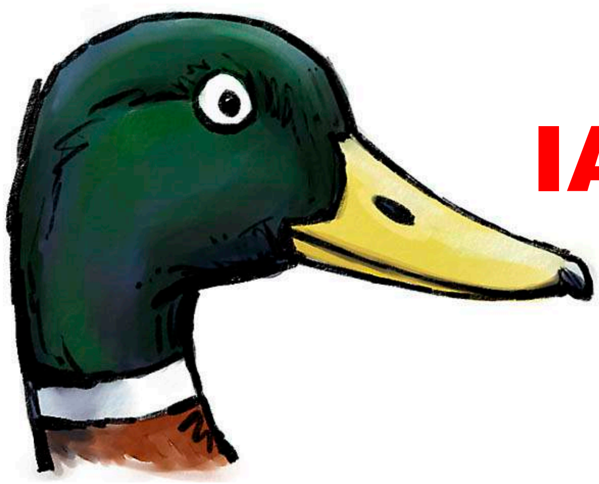
CONCLUSION

Question 1

Does IAV affect the microbiome of wild ducks?

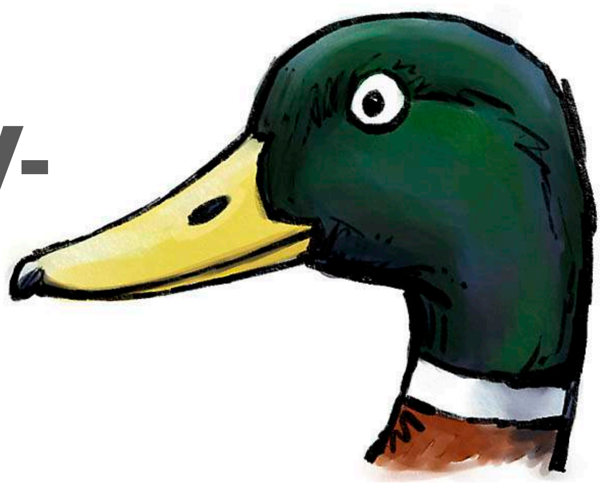
Yes. And…





IAV+

IAV-

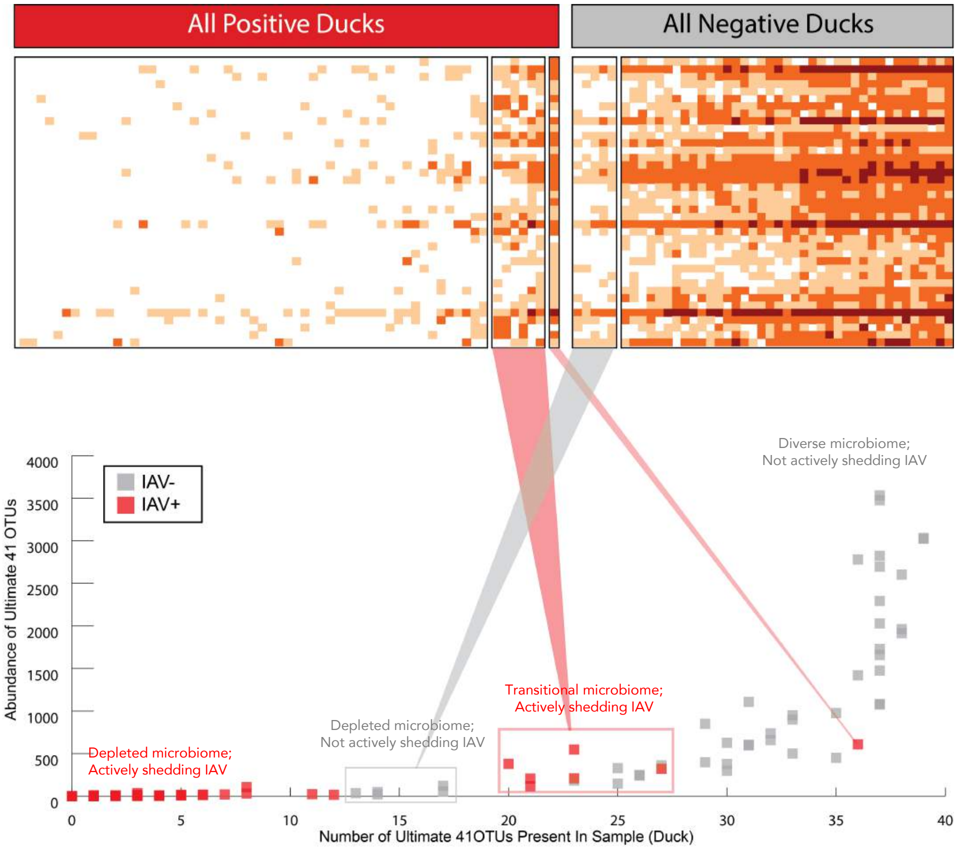


CONCLUSION

Question 1

Does IAV affect the microbiome of wild ducks?

Yes. And…

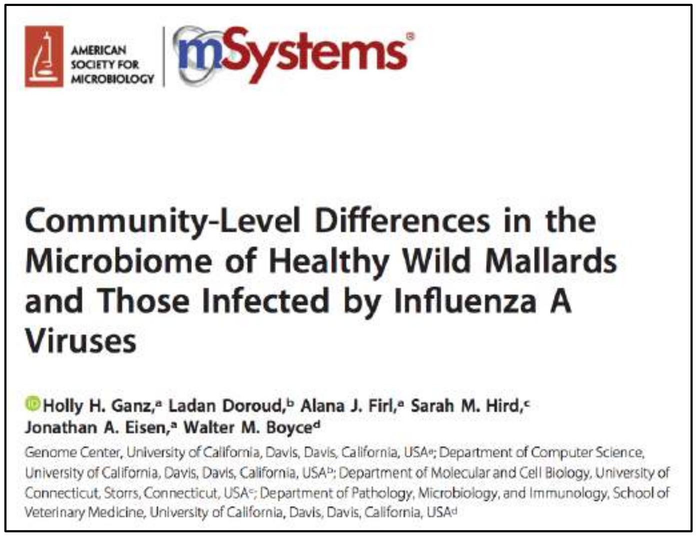


CONCLUSION

Question 1

Does IAV affect the microbiome of wild ducks?

Yes. And…



AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

Community-Level Differences in the Microbiome of Healthy Wild Mallards and Those Infected by Influenza A Viruses

Holly H. Ganz,[a] Ladan Doroud,[b] Alana J. Firl,[a] Sarah M. Hird,[c] Jonathan A. Eisen,[a] Walter M. Boyce[d]

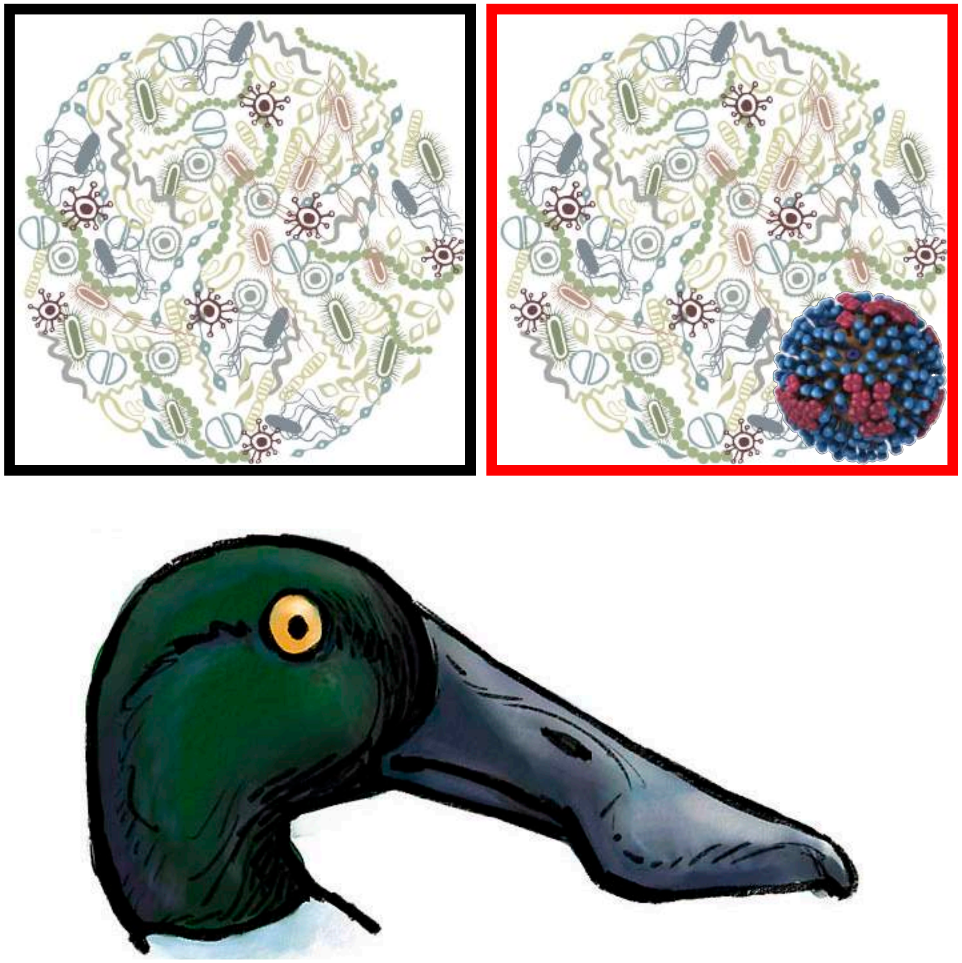
Genome Center, University of California, Davis, Davis, California, USA[a]; Department of Computer Science, University of California, Davis, Davis, California, USA[b]; Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA[c]; Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, Davis, California, USA[d]

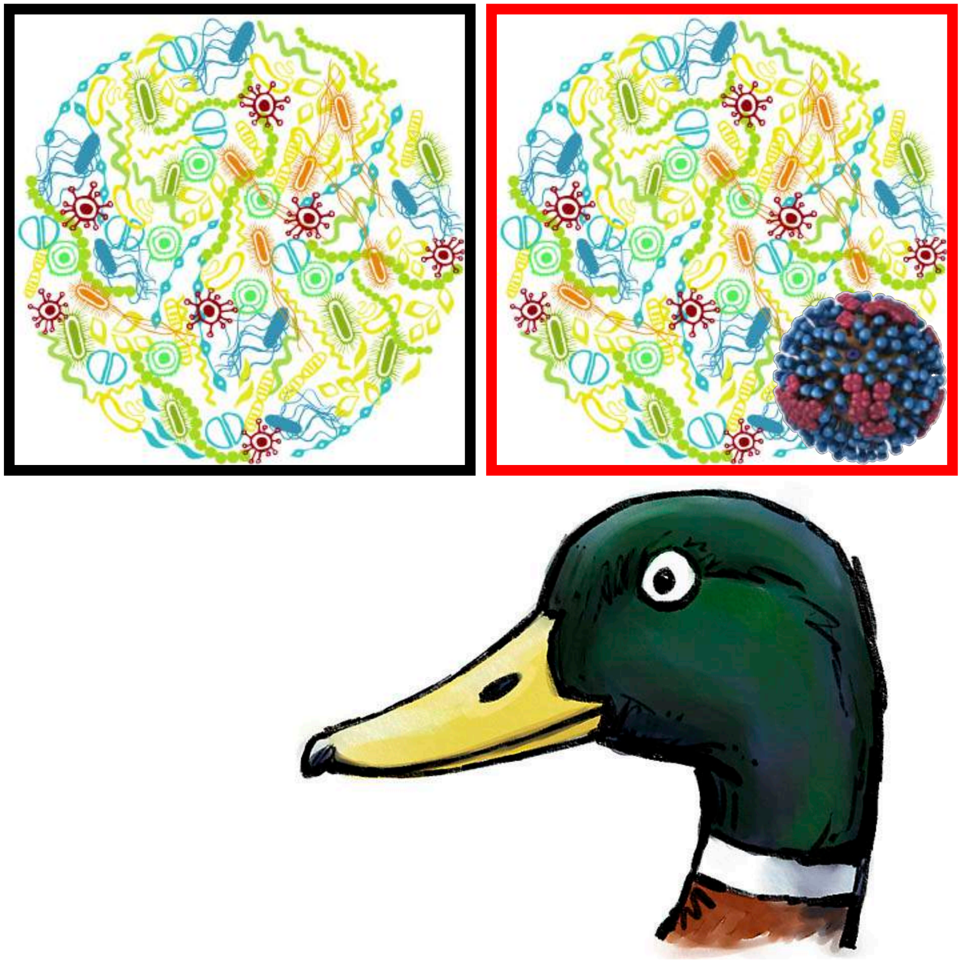
@sarahmhird

INTRODUCTION

Question 2

Does IAV affect the microbiomes of different wild duck species in the same way?





Q2: METHODS

Meet the ducks (*Anas*)

A. acuta (Northern Pintail)

20 IAV-
11 IAV+

A. americana (American Wigeon)

10 IAV-
15 IAV+

A. carolinensis (Green-winged Teal)

10 IAV-
9 IAV+

A. clypeata (Northern Shoveler)

28 IAV- 28 IAV+

A. platyrhynchos (Mallard)

55 IAV-
120 IAV+

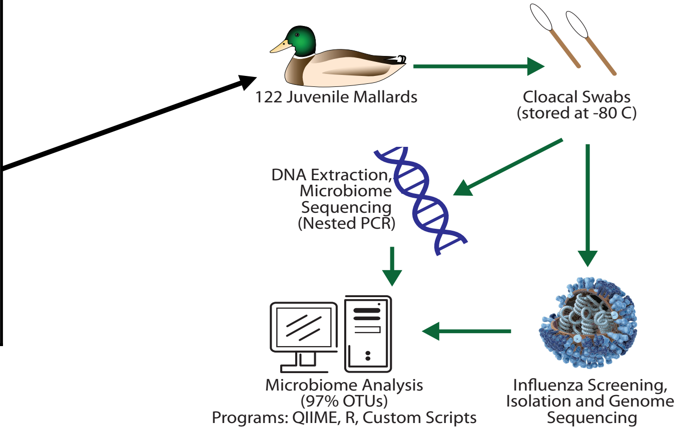
Q2: METHODS

Meet the ducks (*Anas*)


Jim Gain

		SAMPLING LOCALITY				
	N	CR	GIWA	SNWR	MI	SMAR
All Ducks	**300**	**46**	**110**	**72**	**19**	**29**
A. acuta	31	14	3	14	0	0
A. americana	25	0	1	0	0	0
A. carolinensis	19	13	0	6	0	0
A. clypeata	57	0	5	52	0	0
A. platyrhynchos	168	19	101	0	19	29

California, USA


122 Juvenile Mallards

Cloacal Swabs (stored at -80 C)

DNA Extraction, Microbiome Sequencing (Nested PCR)

Influenza Screening, Isolation and Genome Sequencing

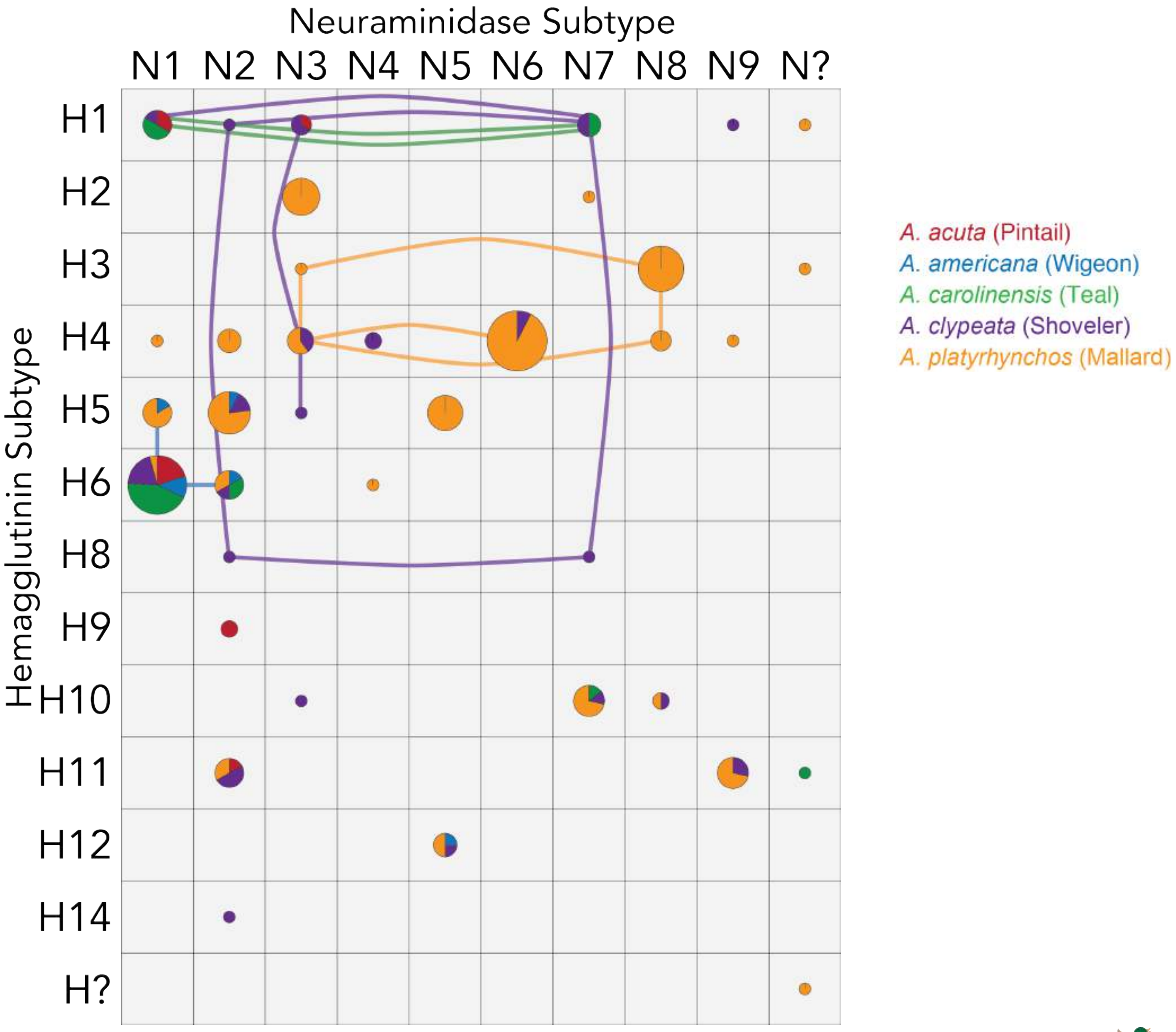
Microbiome Analysis (97% OTUs)
Programs: QIIME, R, Custom Scripts

@sarahmhird

Q2: METHODS

Meet the flu

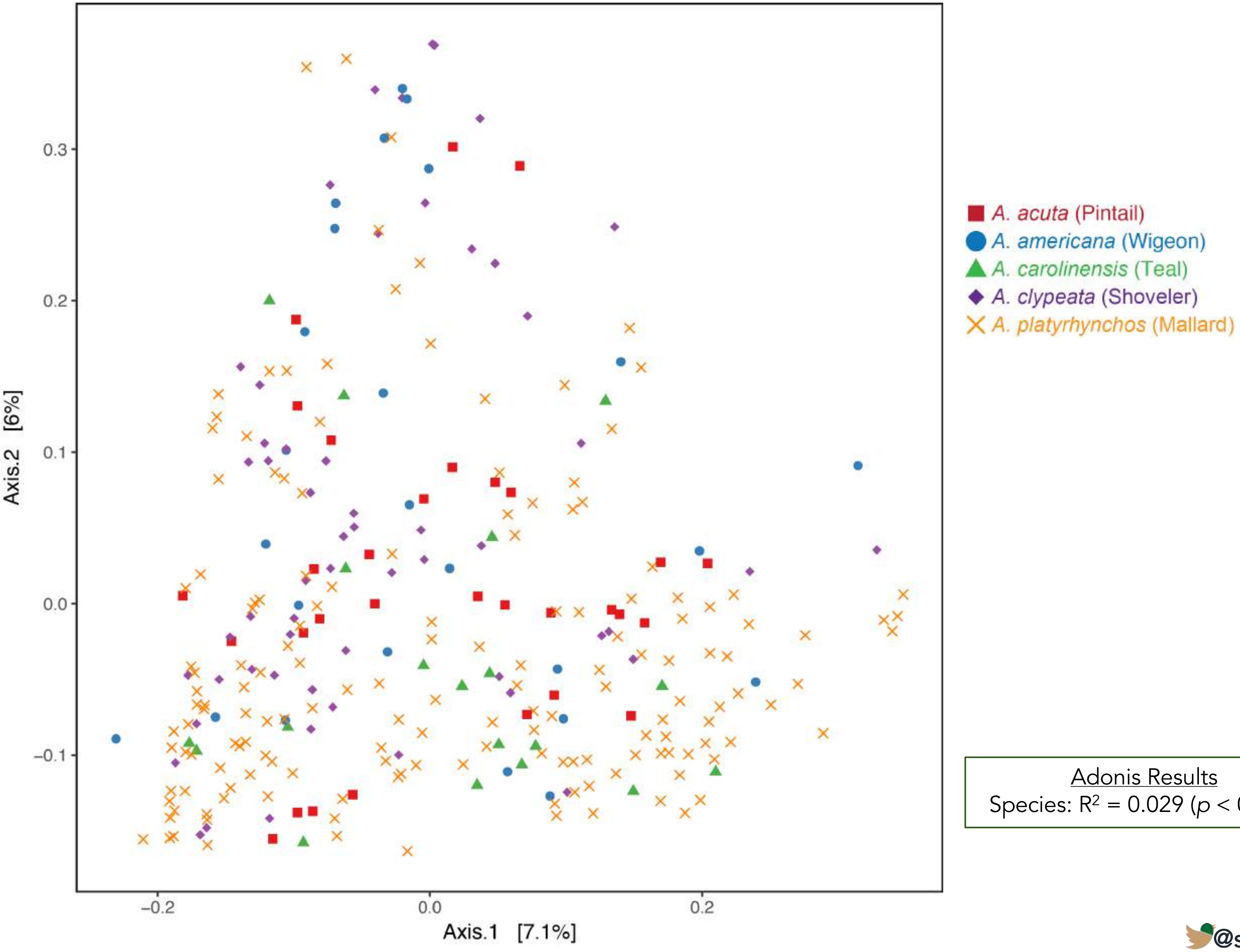
Lots of strain-level diversity



@sarahmhird

Q2: RESULTS (BETA DIVERSITY)

PCoA (Unweighted UniFrac Distances)



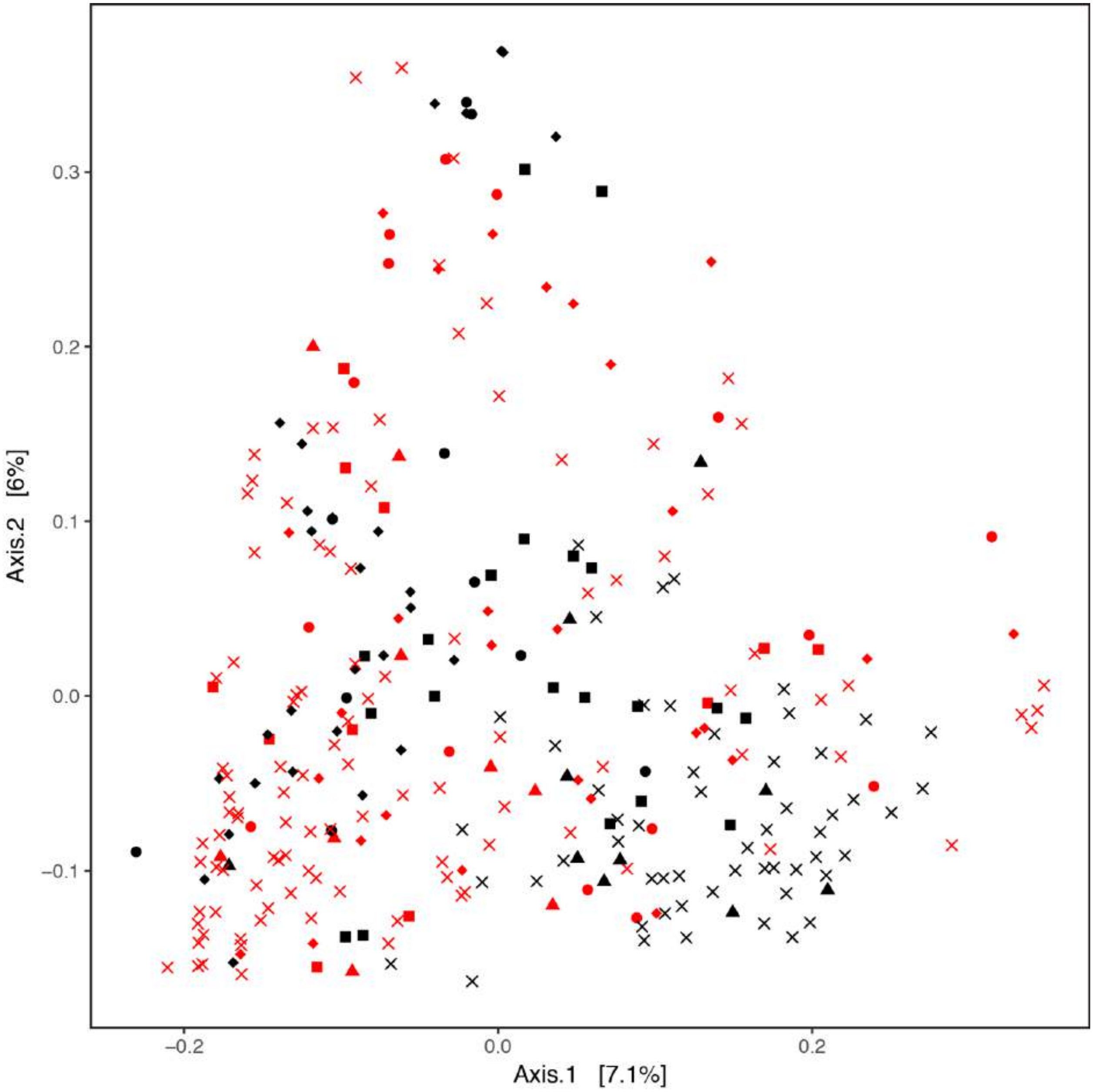
Adonis Results

Species: $R^2 = 0.029$ ($p < 0.001$)

@sarahmhird

Q2: RESULTS (BETA DIVERSITY)

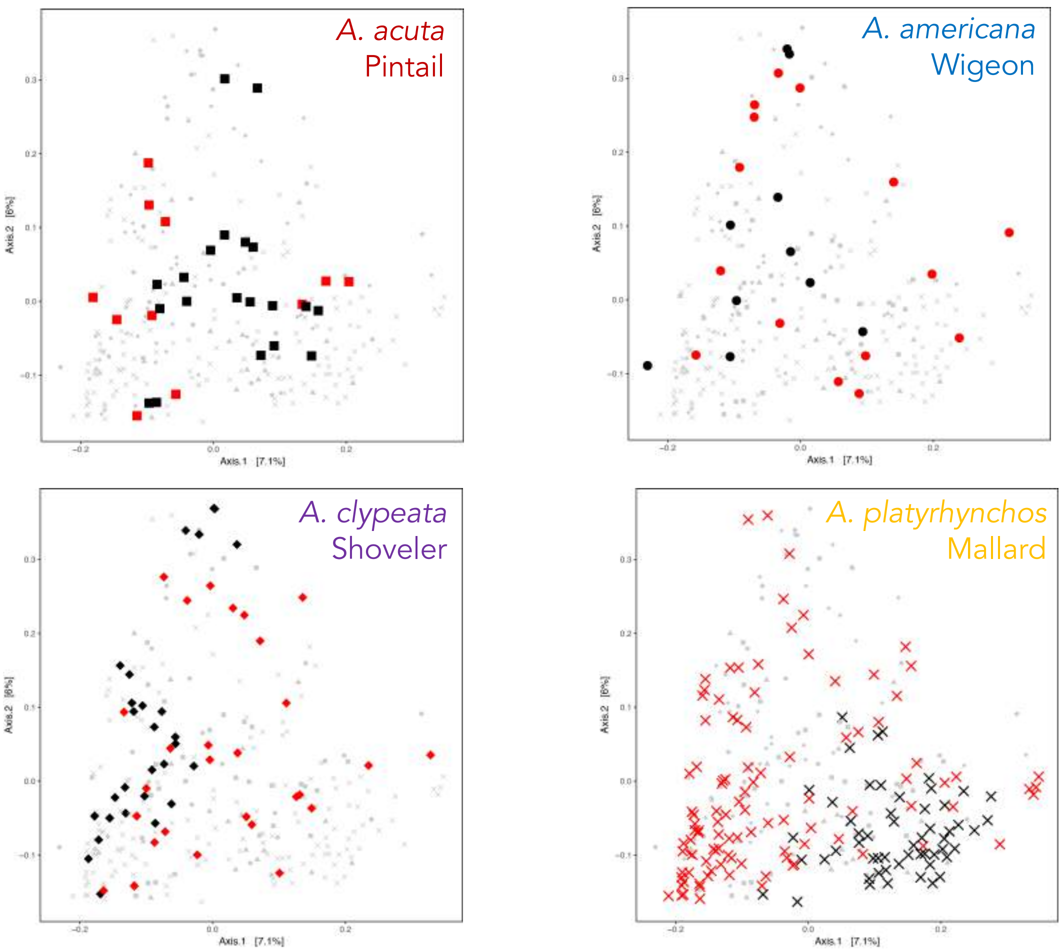
PCoA (Unweighted UniFrac Distances)



Adonis Results
Species: $R^2 = 0.029$ ($p < 0.001$)
Influenza: $R^2 = 0.013$ ($p < 0.001$)

Q2: RESULTS (BETA DIVERSITY)

PCoA (Unweighted UniFrac Distances)



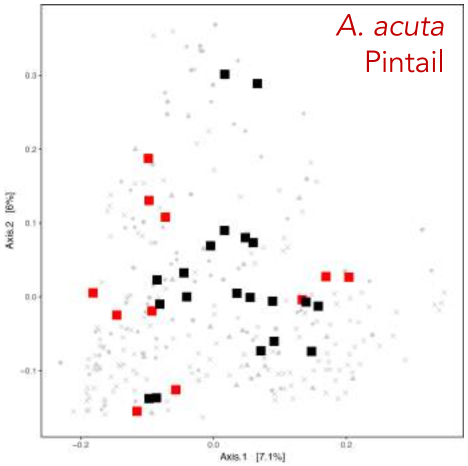
IAV- (Negative)
IAV+ (Positive)

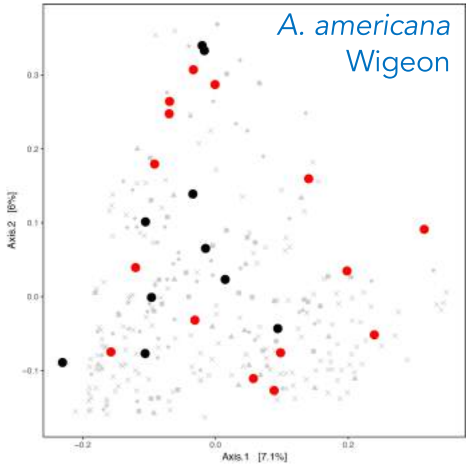
Adonis Results (Influenza)

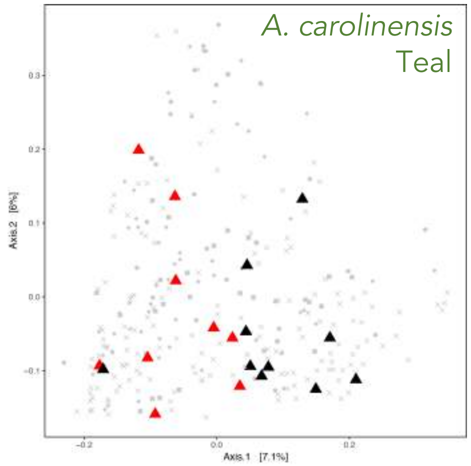
A. acuta (Pintail): $R^2 = 0.044$ ($p = 0.012$)
A. americana (Wigeon): $R^2 = 0.062$ ($p = 0.014$)
A. carolinensis (Teal): $R^2 = 0.078$ ($p = 0.006$)
A. clypeata (Shoveler): $R^2 = 0.046$ ($p < 0.001$)
A. platyrhynchos (Mallard): $R^2 = 0.041$ ($p < 0.001$)

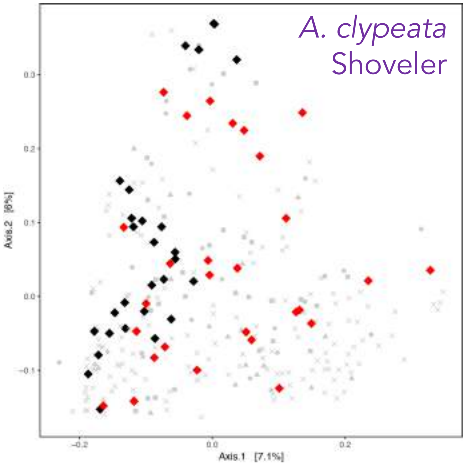
Q2: RESULTS (BETA DIVERSITY)

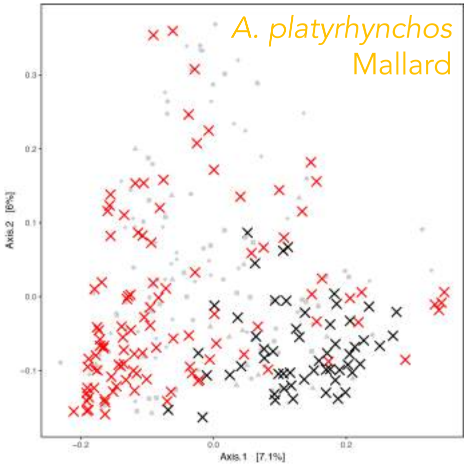
PCoA (Unweighted UniFrac Distances)











IAV- (Negative)
IAV+ (Positive)

Adonis Results (Influenza)

A. acuta (Pintail):	$R^2 = 0.044$ ($p = 0.012$)
A. americana (Wigeon):	$R^2 = 0.062$ ($p = 0.014$)
A. carolinensis (Teal):	$R^2 = 0.078$ ($p = 0.006$)
A. clypeata (Shoveler):	$R^2 = 0.046$ ($p < 0.001$)
A. platyrhynchos (Mallard):	$R^2 = 0.041$ ($p < 0.001$)

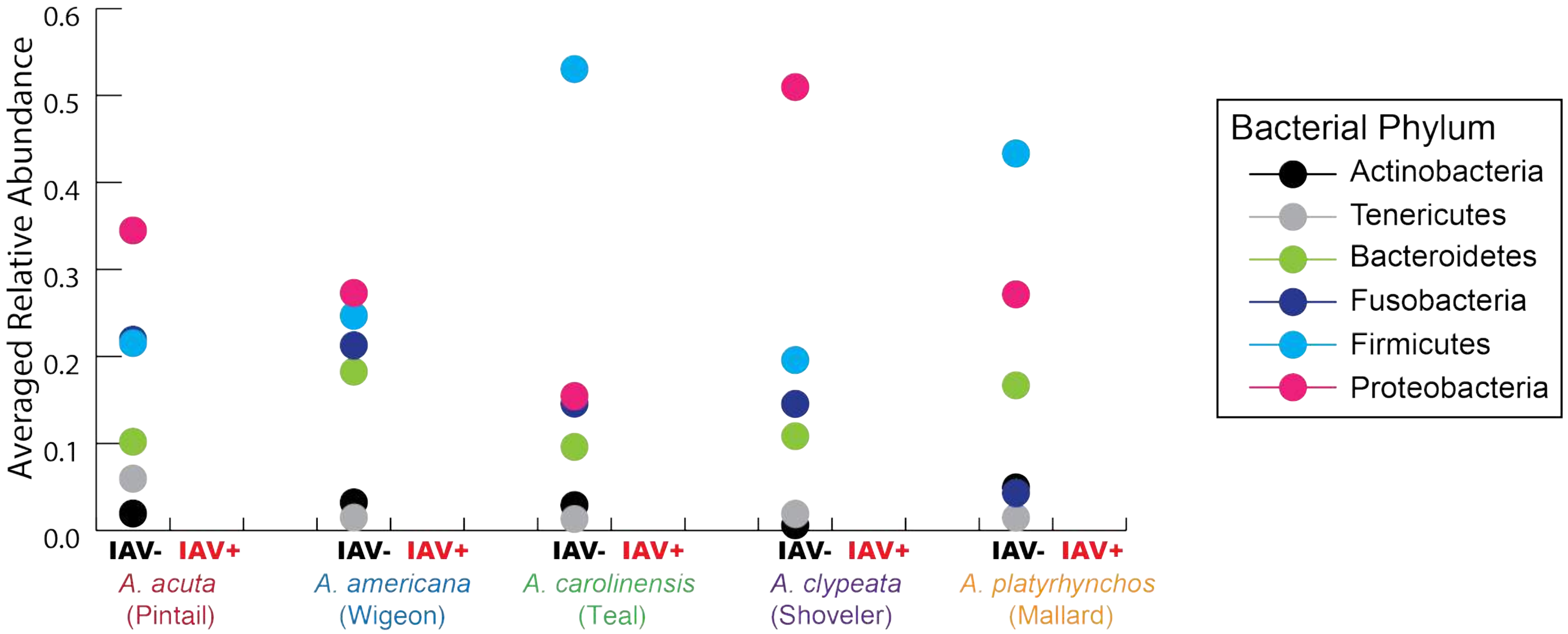
Flu is significantly associated with the cloacal microbiome within species but the effect size is small.

@sarahmhird

What do the communities look like?

Q2: RESULTS (ALPHA DIVERSITY)

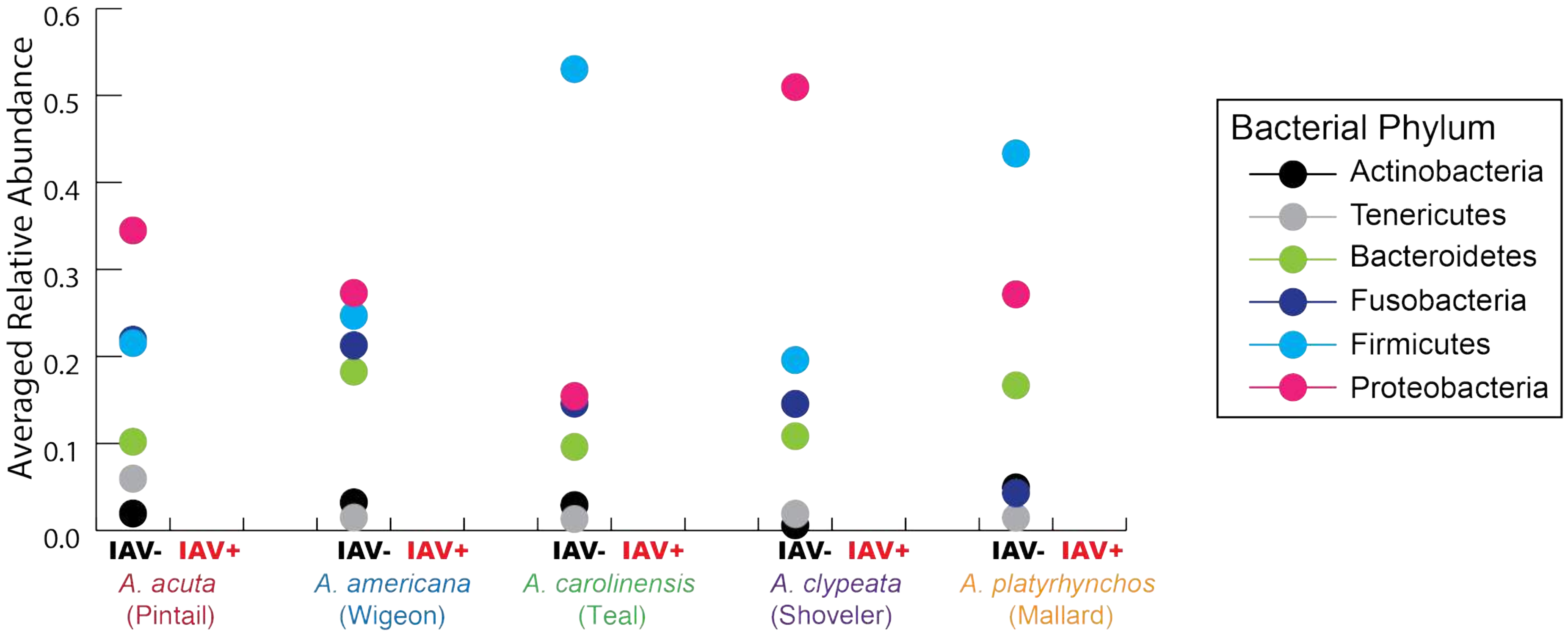
Bacterial Taxonomic Composition



@sarahmhird

Q2: RESULTS (ALPHA DIVERSITY)

Bacterial Taxonomic Composition

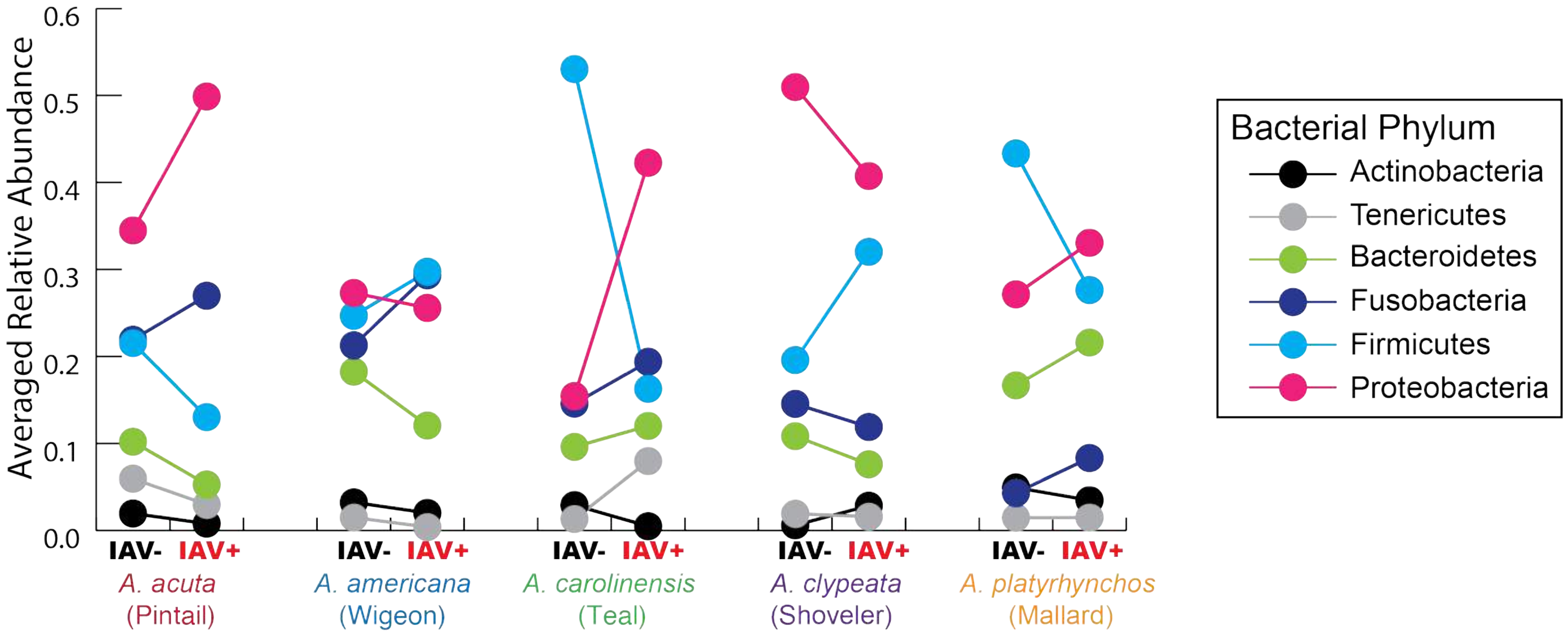


The same six bacterial phyla dominate but at differing relative abundances.

@sarahmhird

Q2: RESULTS (ALPHA DIVERSITY)

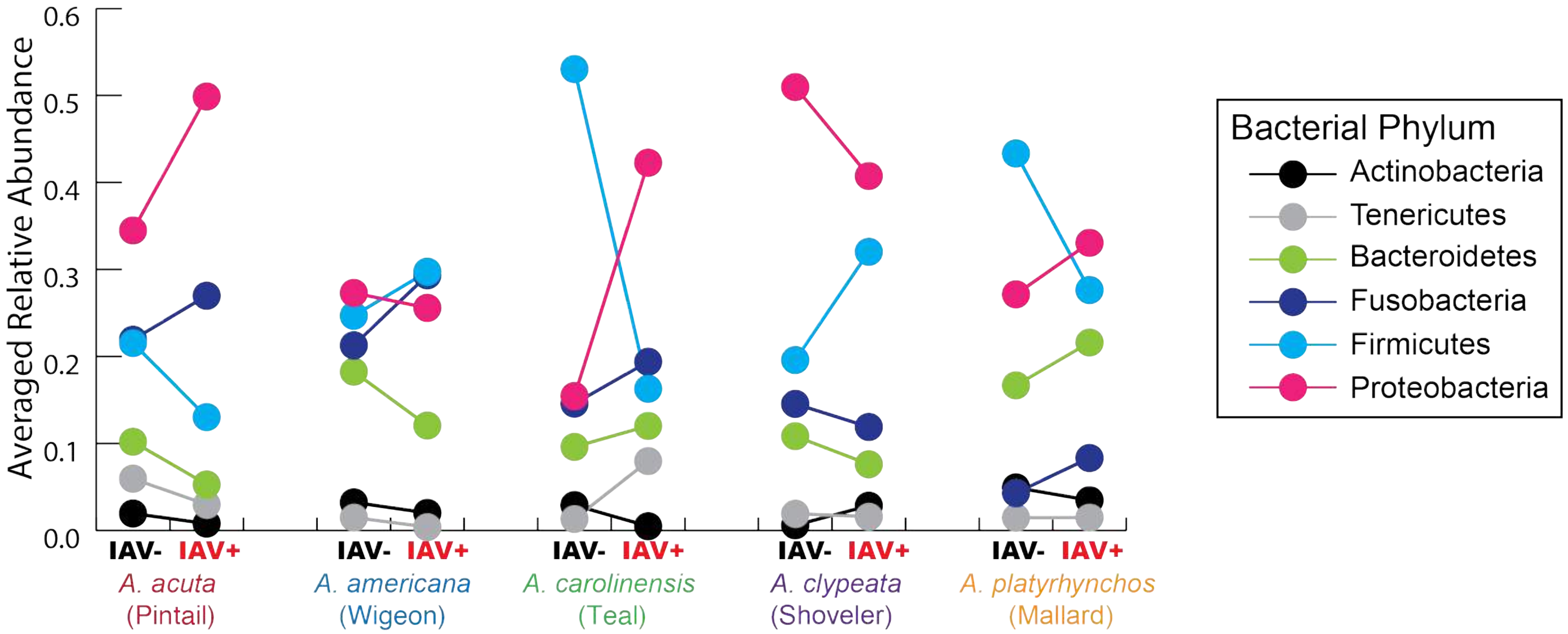
Bacterial Taxonomic Composition



@sarahmhird

Q2: RESULTS (ALPHA DIVERSITY)

Bacterial Taxonomic Composition

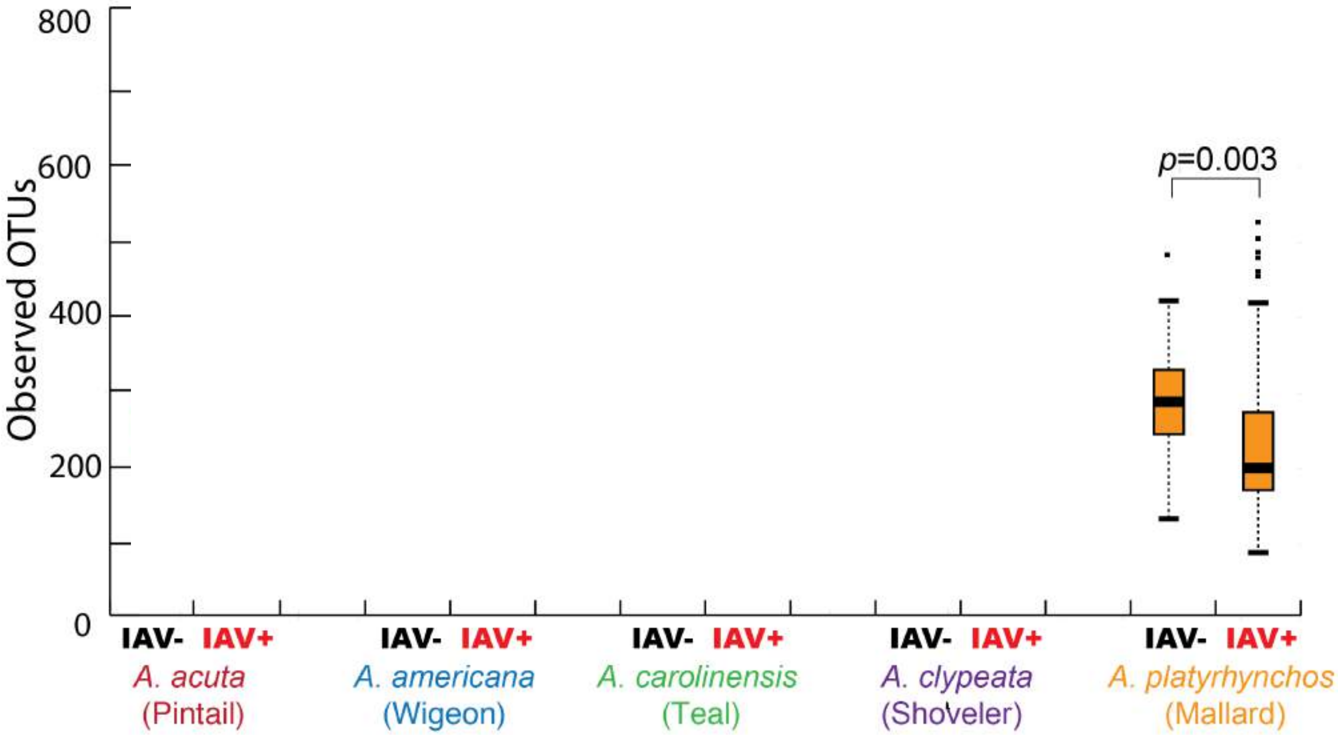


No pattern to changes in microbiome taxonomy across species.

@sarahmhird

Q2: RESULTS

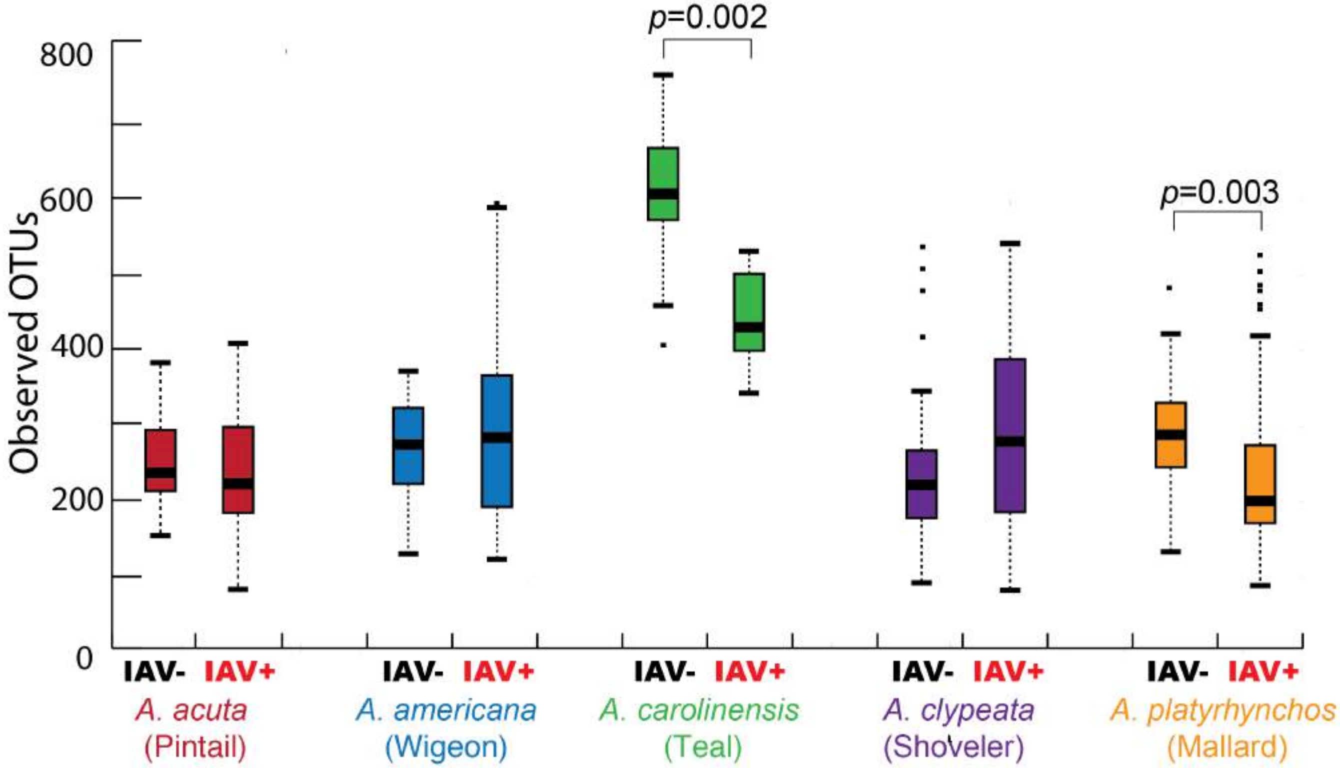
Community Richness



@sarahmhird

Q2: RESULTS

Community Richness

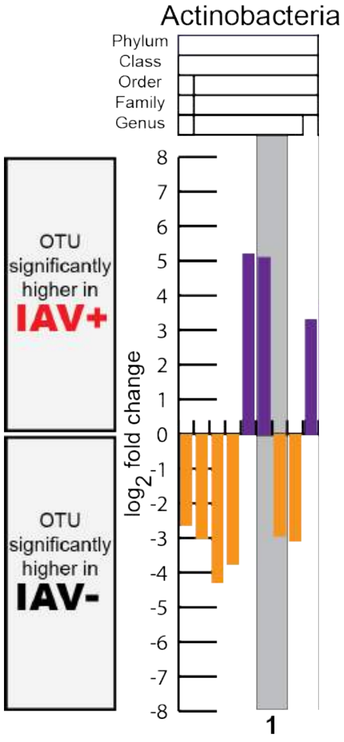


No pattern in community richness across species.

@sarahmhird

Q2: RESULTS

Significantly different OTUs



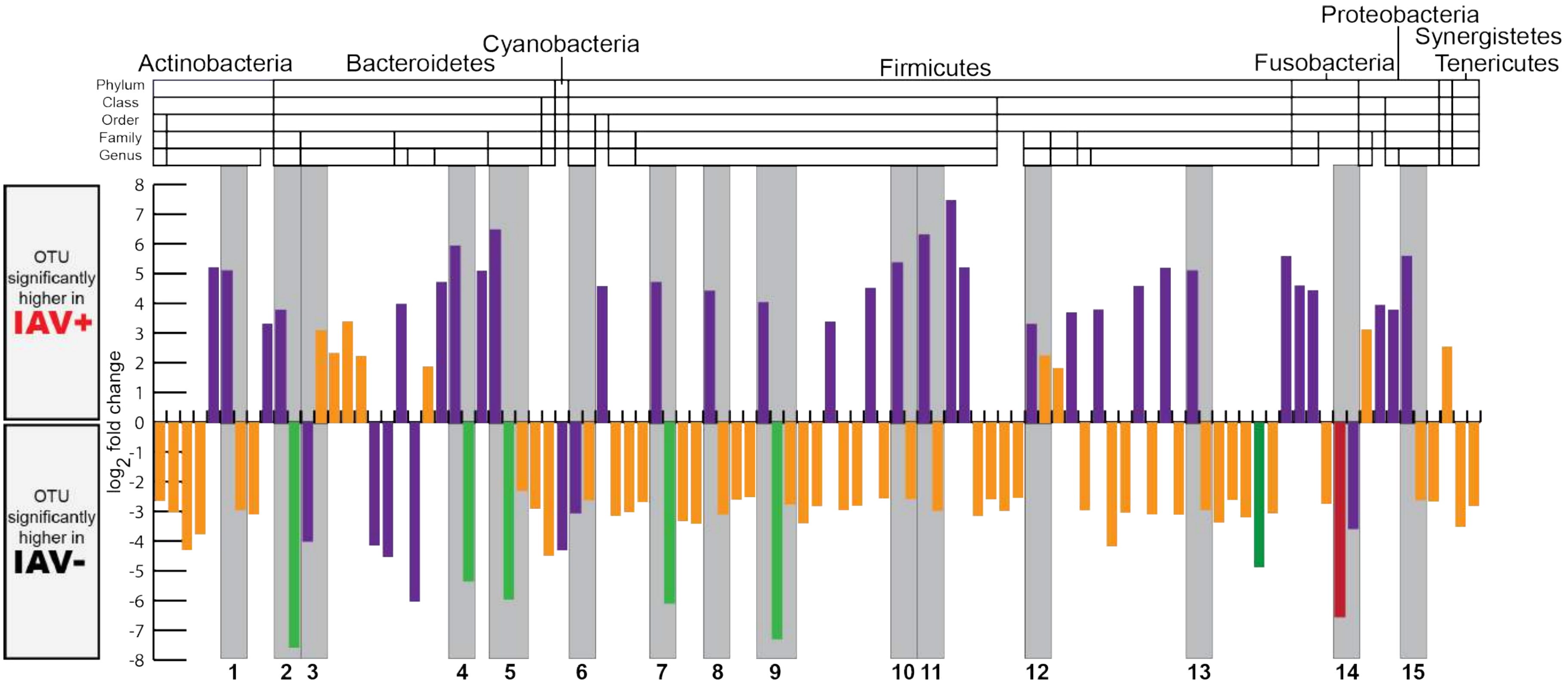
1 *Rothia mucilaginosa*

A. acuta (Pintail)
A. americana (Wigeon)
A. carolinensis (Teal)
A. clypeata (Shoveler)
A. platyrhynchos (Mallard)

@sarahmhird

Q2: RESULTS

Significantly different OTUs



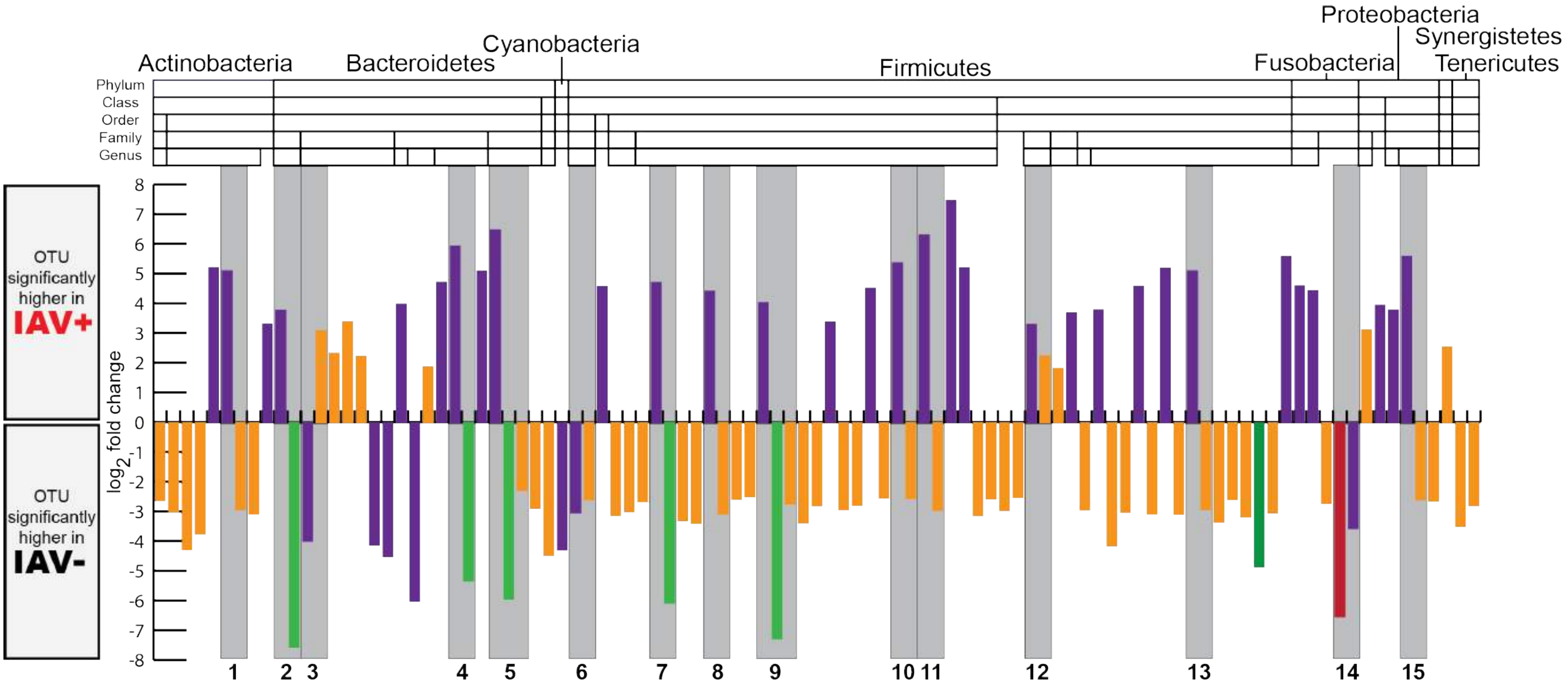
1 *Rothia mucilaginosa*
2 *[Prevotella]*
3 *Bacteroides*
4 *Porphyromonas*
5 *Prevotella melaninogenica*
6 *Staphylococcus*
7 - 11 *Streptococcus*
12 *Candidatus Arthromitus*
13 *Veillonella dispar*
14 Leptotrichiaceae (family)
15 *Haemophilus*

A. acuta (Pintail)
A. americana (Wigeon)
A. carolinensis (Teal)
A. clypeata (Shoveler)
A. platyrhynchos (Mallard)

@sarahmhird

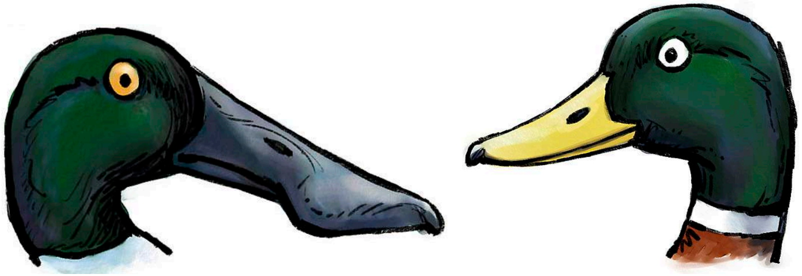
Q2: RESULTS

Significantly different OTUs



1 *Rothia mucilaginosa*
2 *[Prevotella]*
3 *Bacteroides*
4 *Porphyromonas*
5 *Prevotella melaninogenica*
6 *Staphylococcus*
7 - 11 *Streptococcus*
12 *Candidatus Arthromitus*
13 *Veillonella dispar*
14 Leptotrichiaceae (family)
15 *Haemophilus*

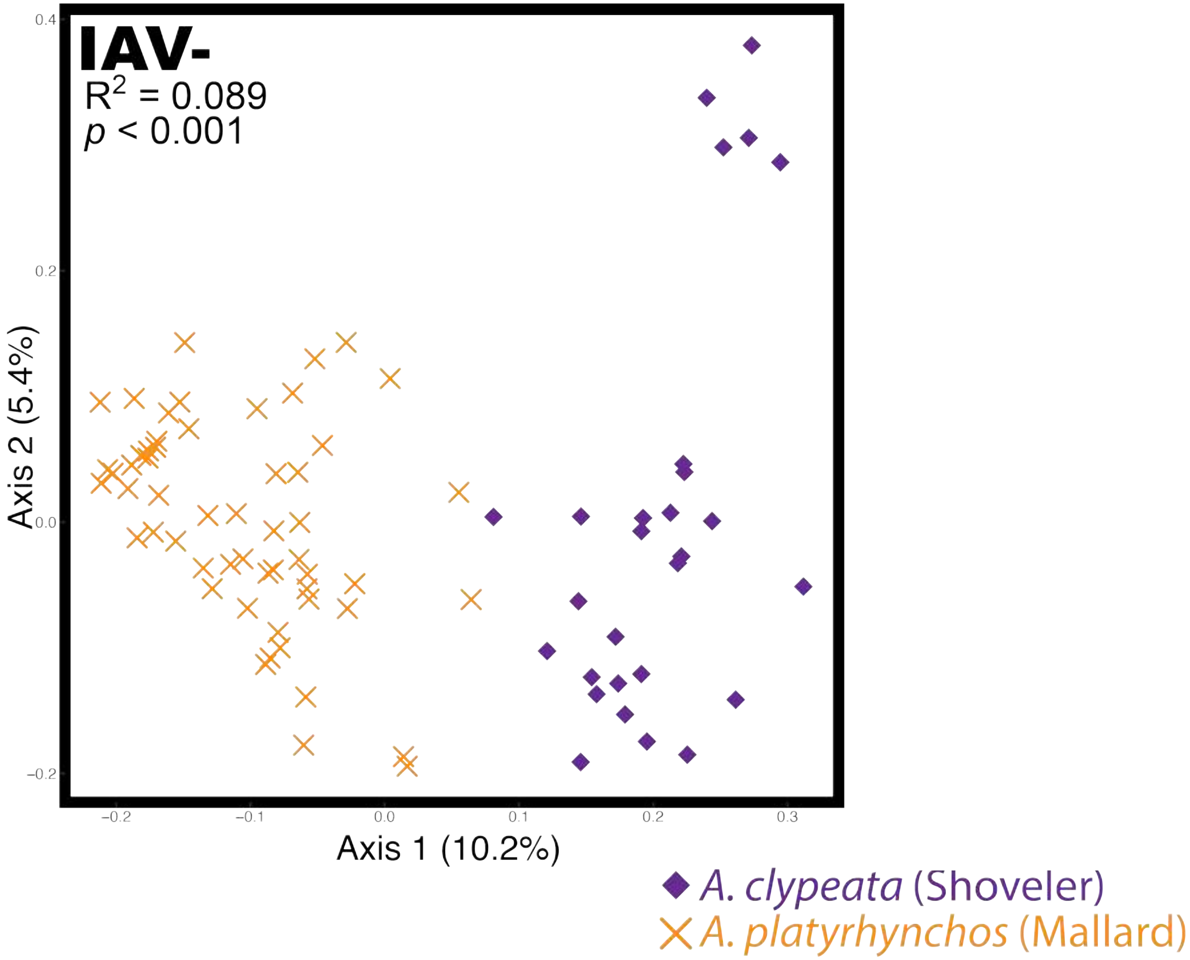
A. acuta (Pintail)
A. americana (Wigeon)
A. carolinensis (Teal)
A. clypeata (Shoveler)
A. platyrhynchos (Mallard)

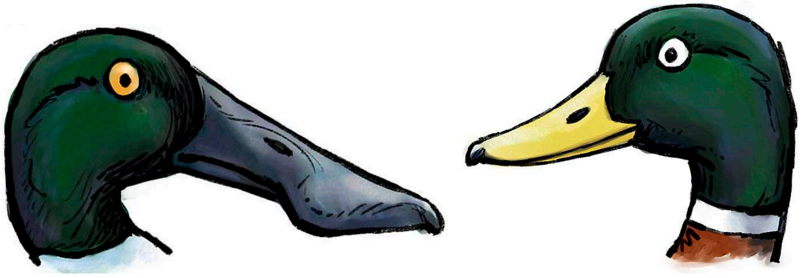


@sarahmhird

Q2: RESULTS (BETA DIVERSITY)

PCoA (Unweighted UniFrac Distances)

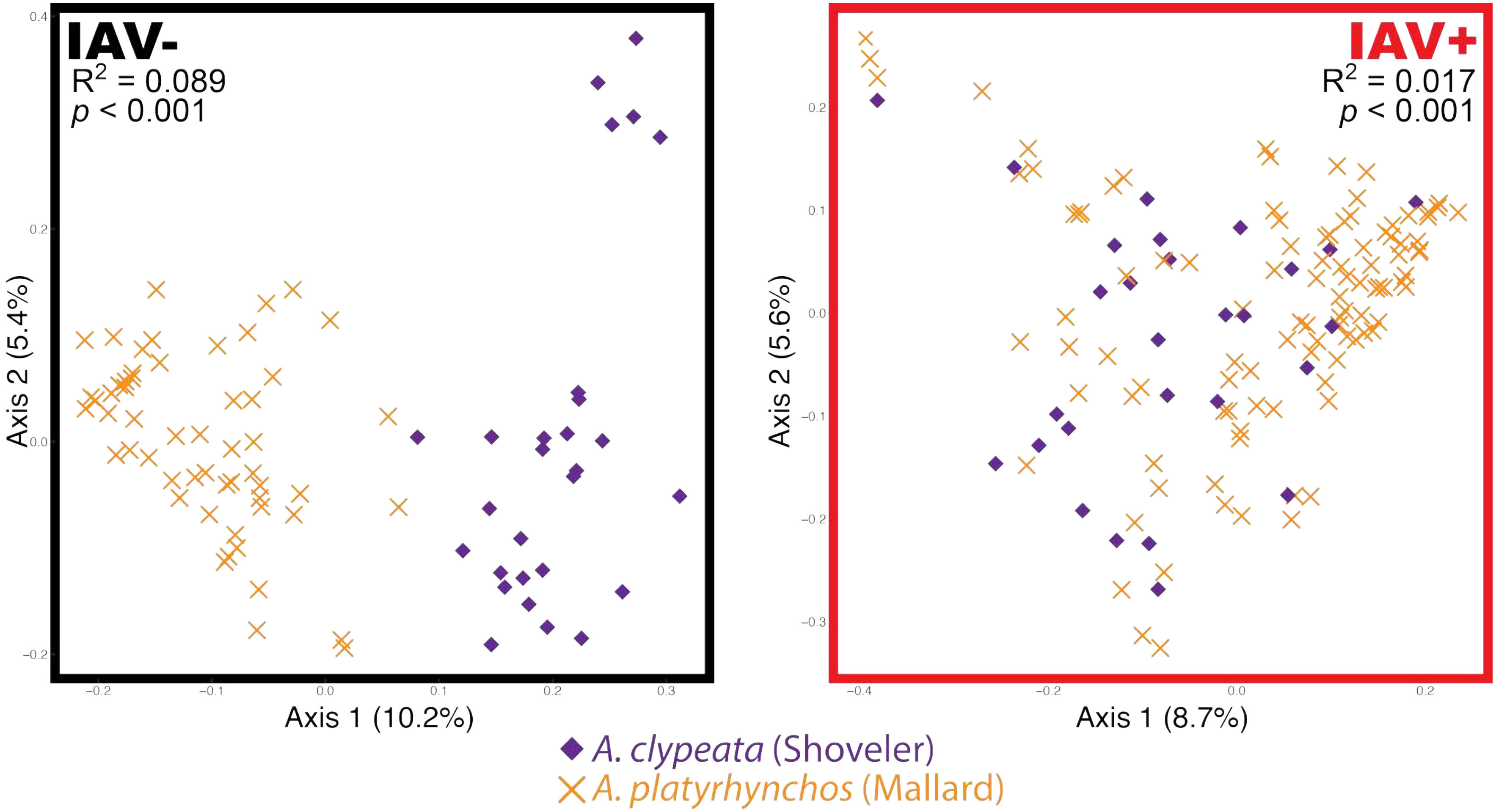


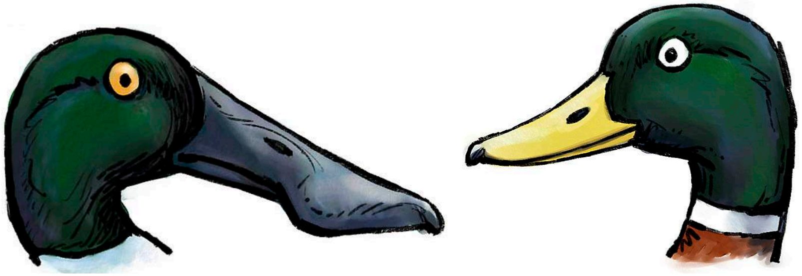


@sarahmhird

Q2: RESULTS (BETA DIVERSITY)

PCoA (Unweighted UniFrac Distances)

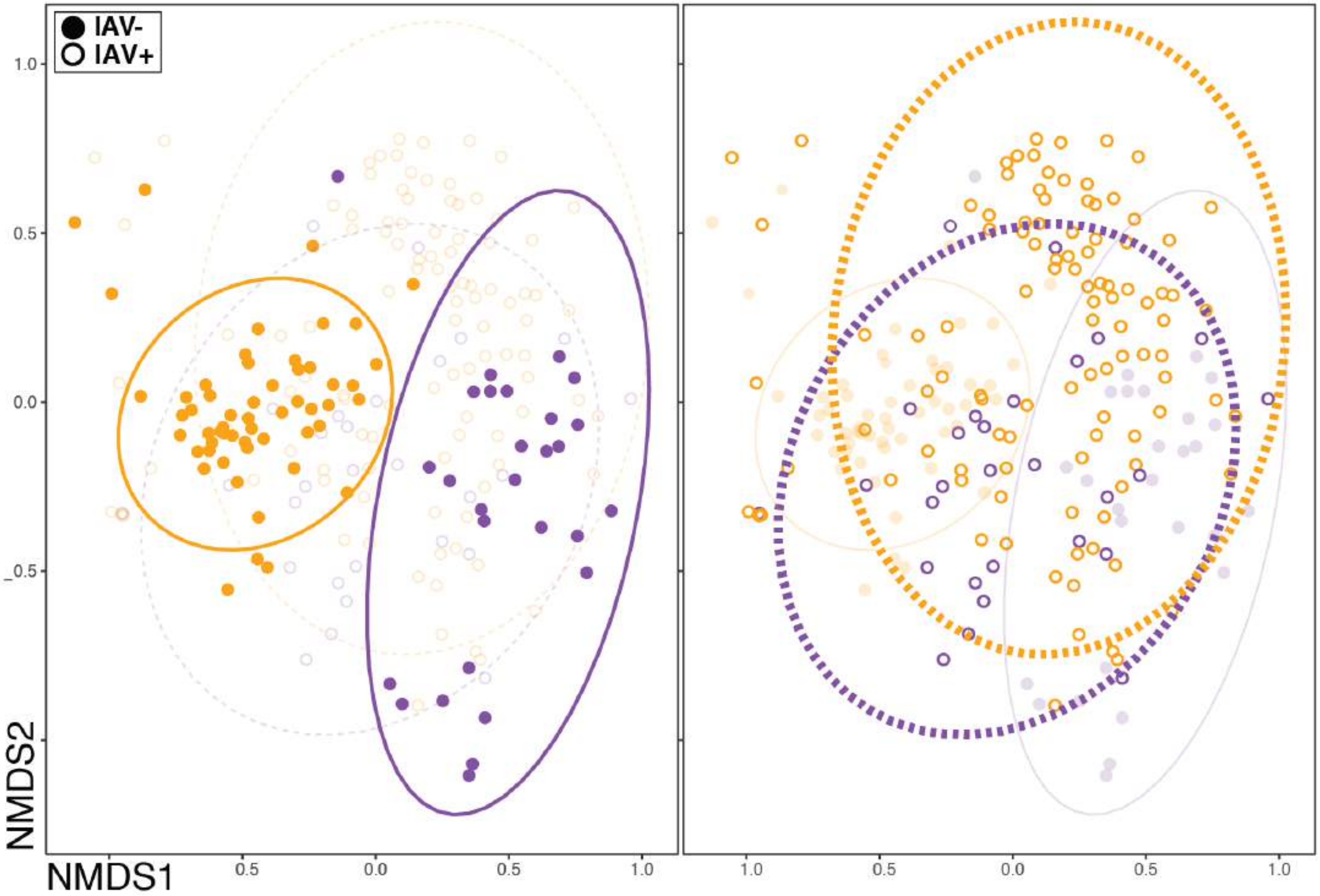


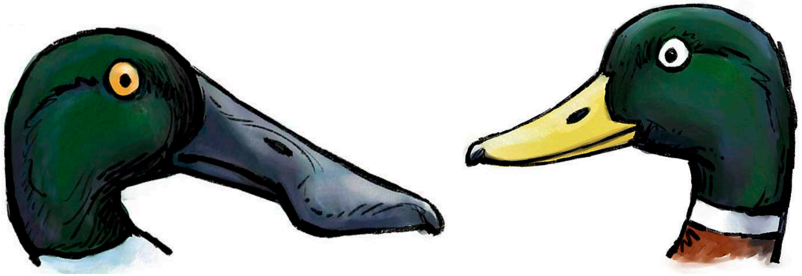


@sarahmhird

Q2: RESULTS (BETA DIVERSITY)

PCoA (Unweighted UniFrac Distances)





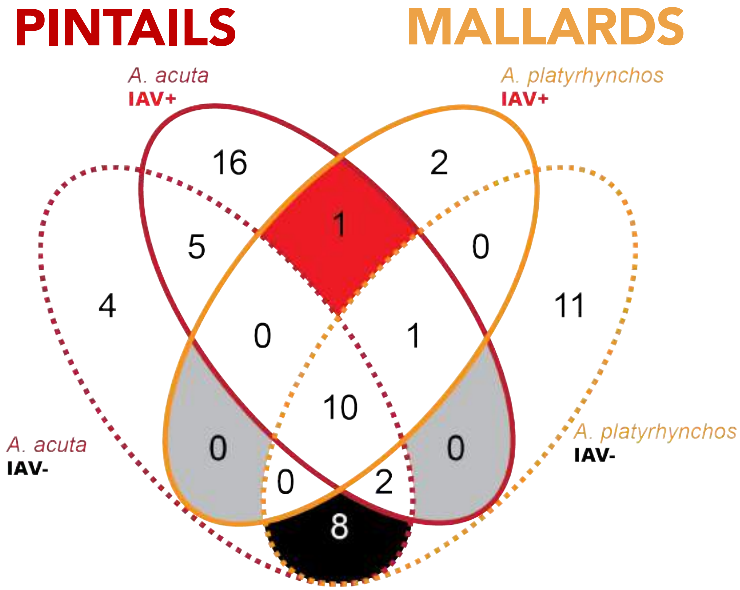
@sarahmhird

Q2: RESULTS

Shared "core" microbes (in 90% of samples)

IAV+

IAV-



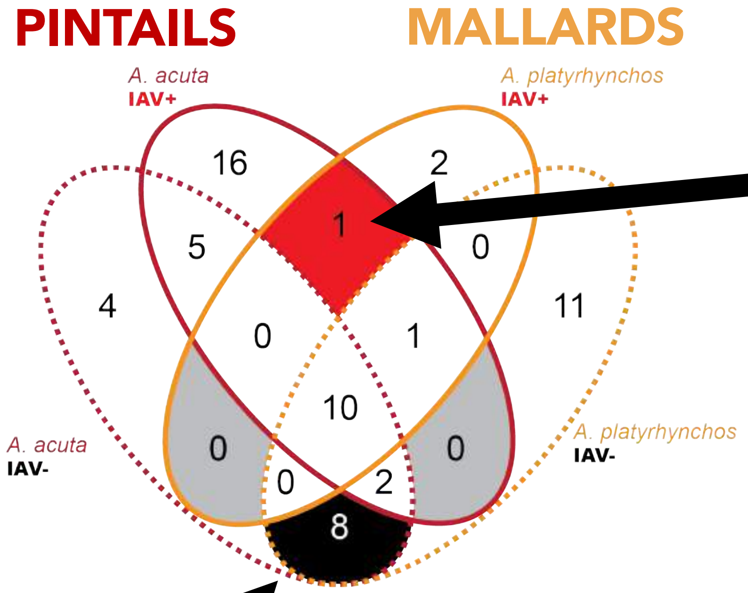
@sarahmhird

Q2: RESULTS

Shared "core" microbes (in 90% of samples)

IAV+

IAV-



IAV+ birds of both species always have these OTUs

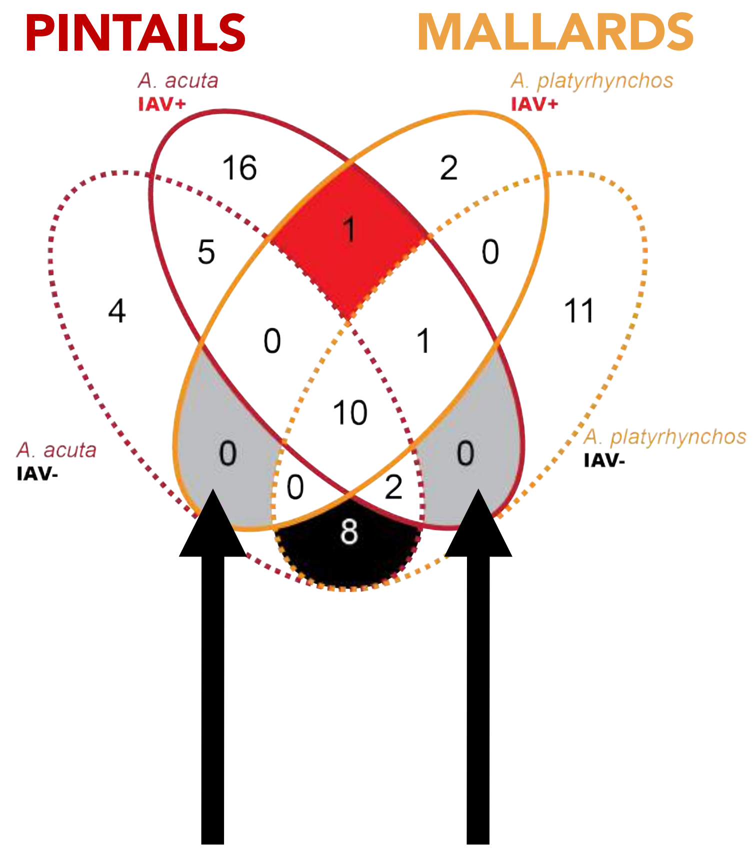
IAV- birds of both species always have these OTUs

Q2: RESULTS

Shared "core" microbes (in 90% of samples)

IAV+

IAV-



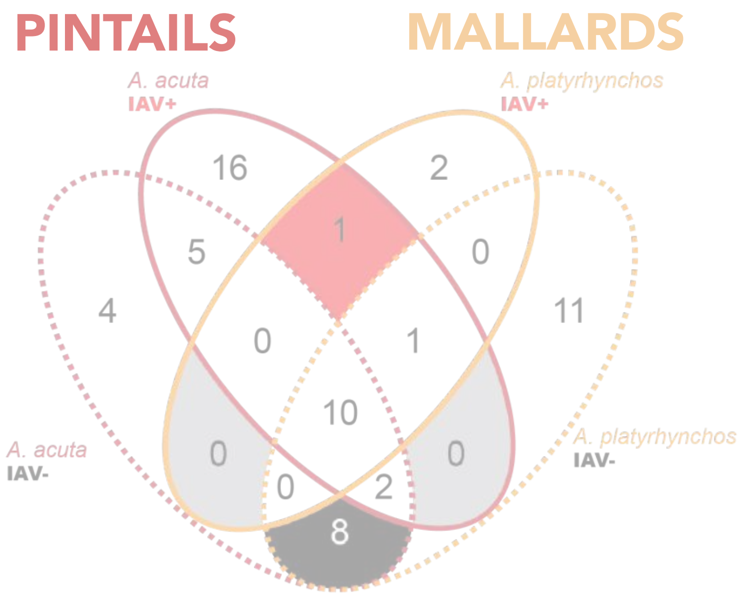
No OTUs shared by **IAV-** birds of one species
and **IAV+** birds of the other

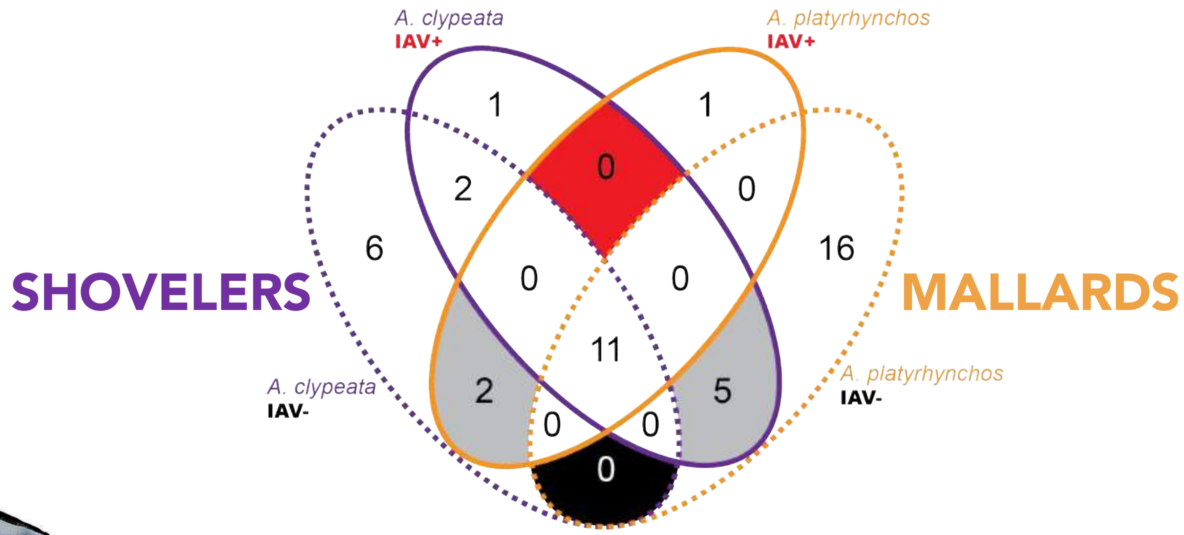
Q2: RESULTS

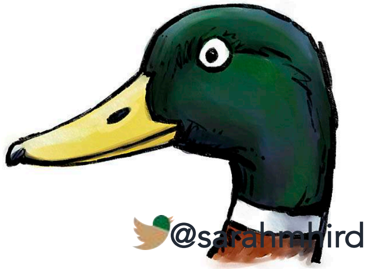
Shared "core" microbes (in 90% of samples)

IAV+

IAV-







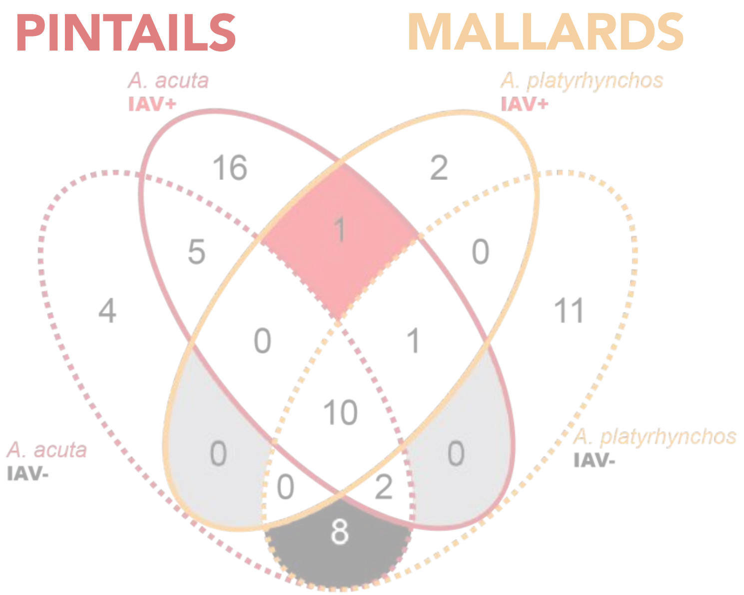
@sarahmbird

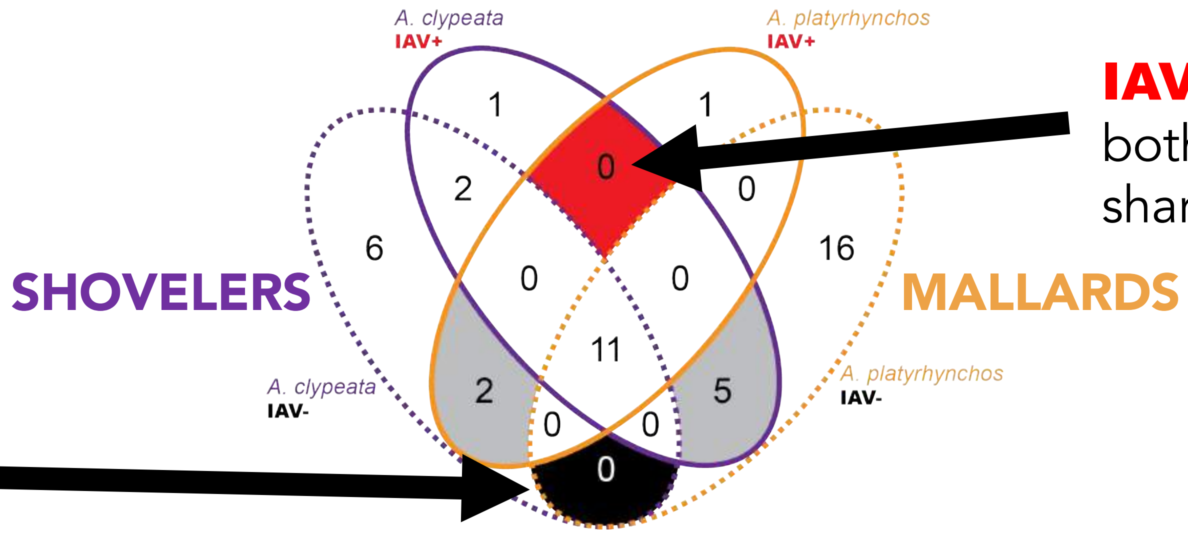
Q2: RESULTS

Shared "core" microbes (in 90% of samples)

IAV+

IAV-





IAV+ birds of both species share 0 OTUs

IAV- birds of both species exclusively share 0 OTUs

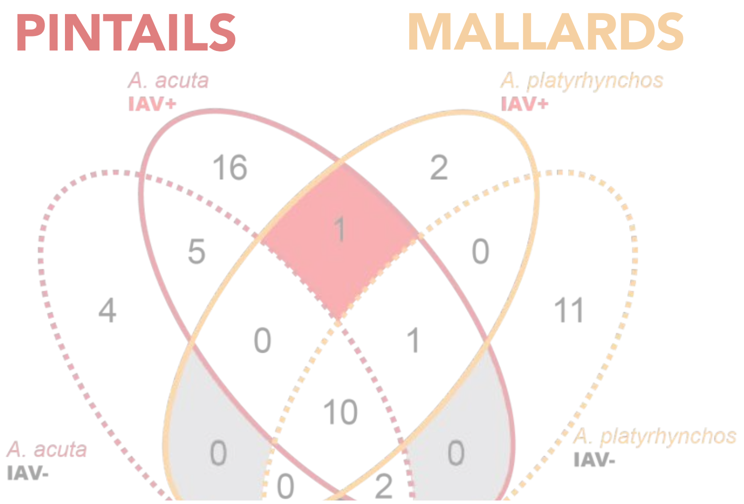
@sarahmhird

Q2: RESULTS

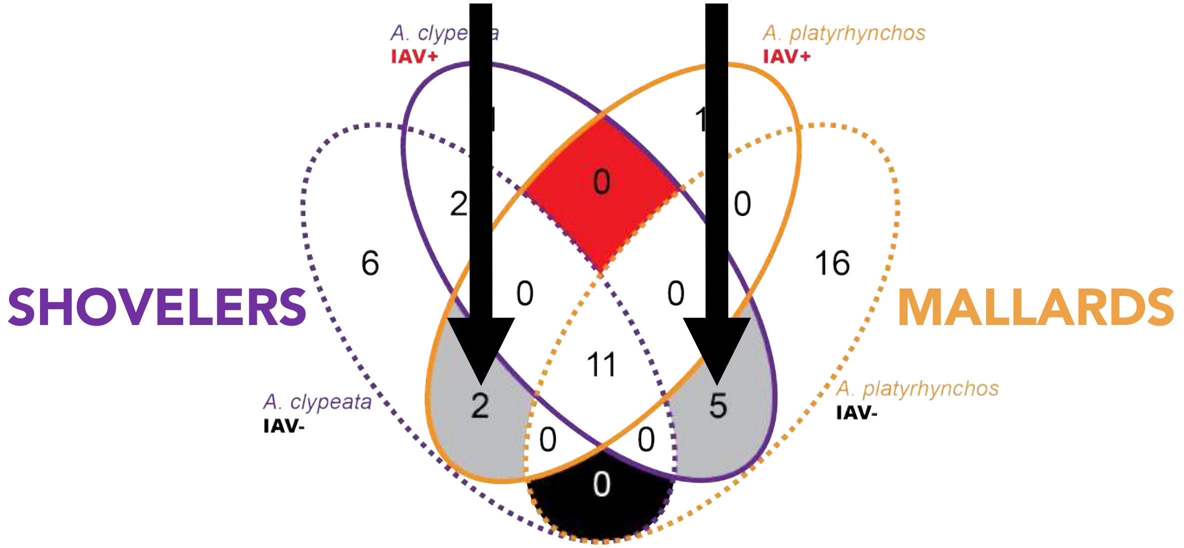
Shared "core" microbes (in 90% of samples)

IAV+

IAV-



OTUs shared by **IAV-** birds of one species and **IAV+** birds of the other

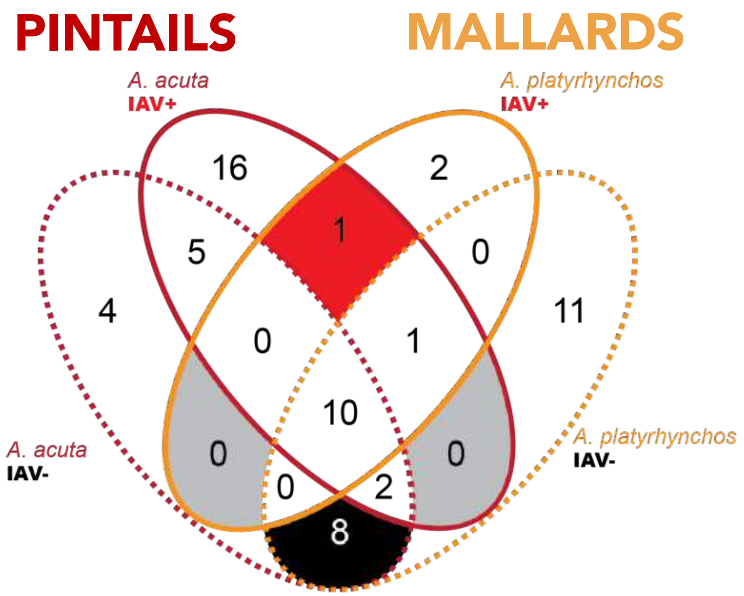


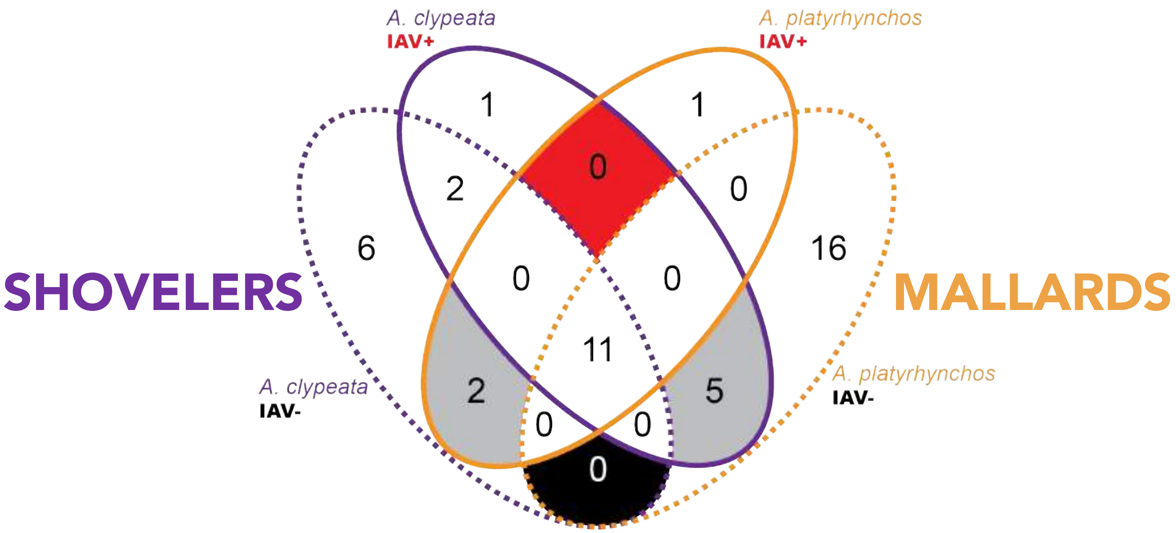
Q2: RESULTS

Shared "core" microbes (in 90% of samples)

IAV+

IAV-

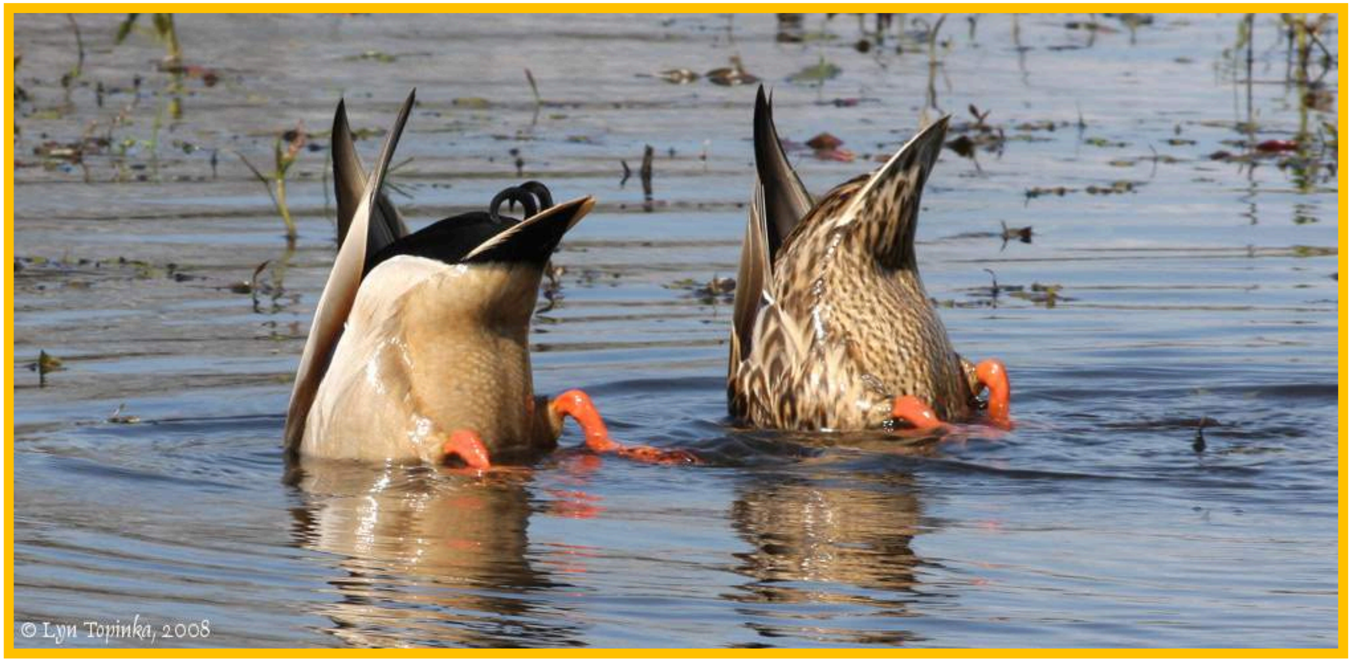




@sarahmhird

Q2: INTERPRETATION

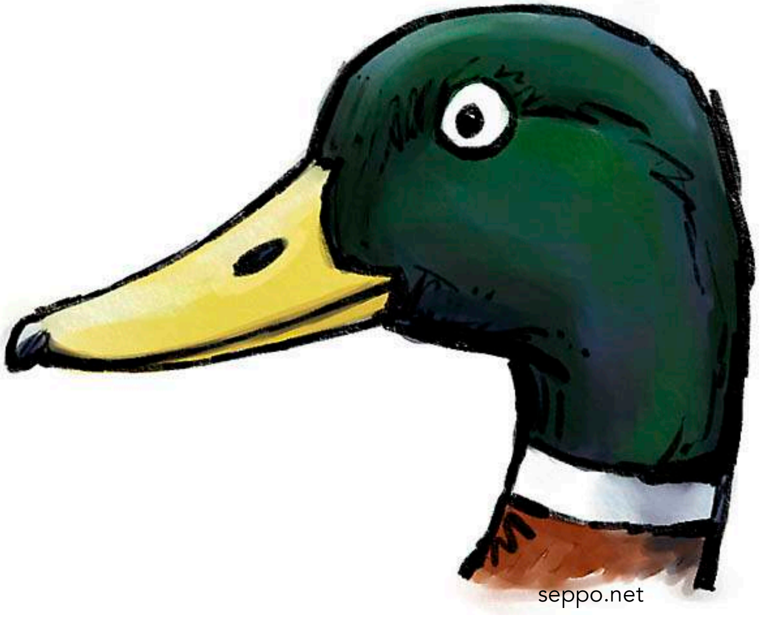
A role for ecology?



© Lyn Topinka, 2008

@sarahmhird

So what’s important?



Q2: *Adonis* TESTS

Possibly important variables

IAV: IAV+ vs IAV-

Species: 5 species

Location: 5 sampling localities

Age: Adult, Hatchyear and Unknown

Season & Year: When collected

HA Subtype

NA Subtype

HANA combination

Sex

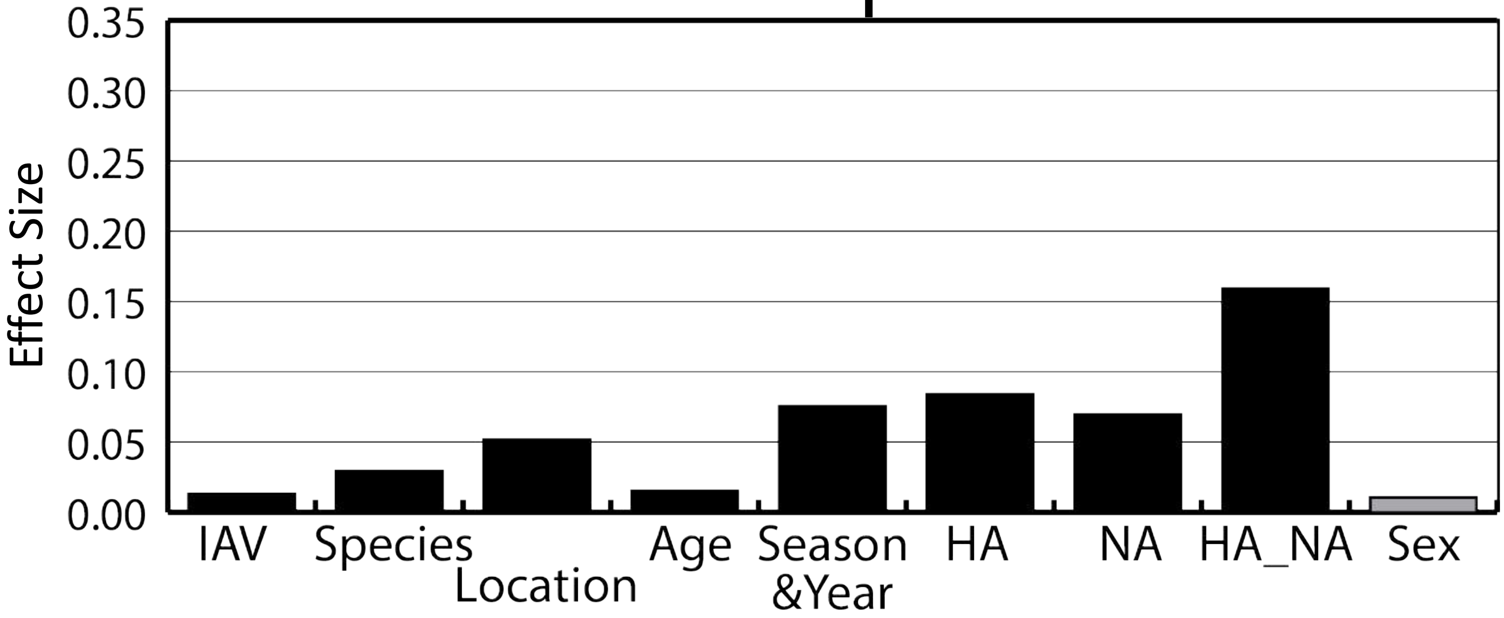
(Lots of these are highly confounded)

Q2: *Adonis* TESTS

What is associated with the microbiome?

black: $p \leq 0.001$
gray: $p > 0.001$

All samples



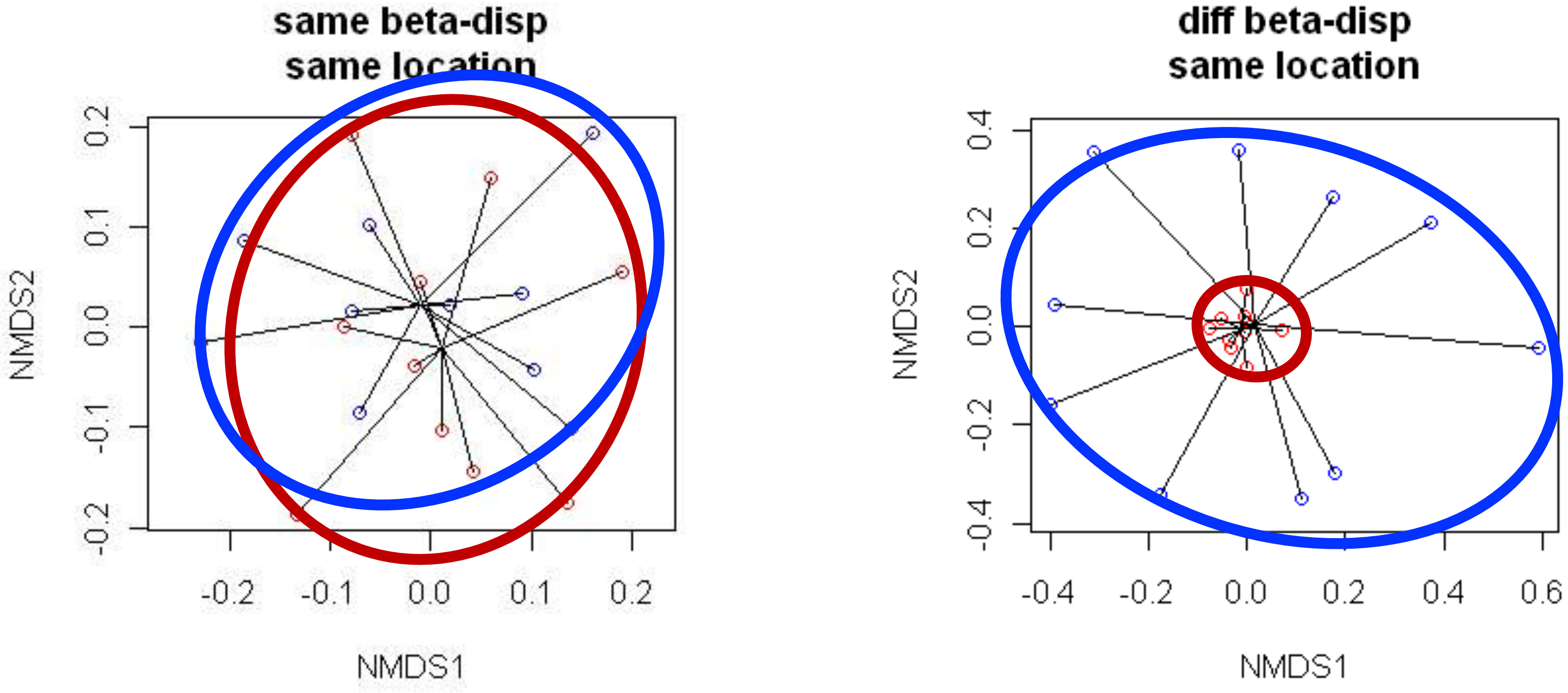
Everything is significant!
(Although some with very low effect sizes)

@sarahmhird

Q2: *Adonis* TESTS

What is associated with the microbiome?

But wait…must test for difference in dispersions.



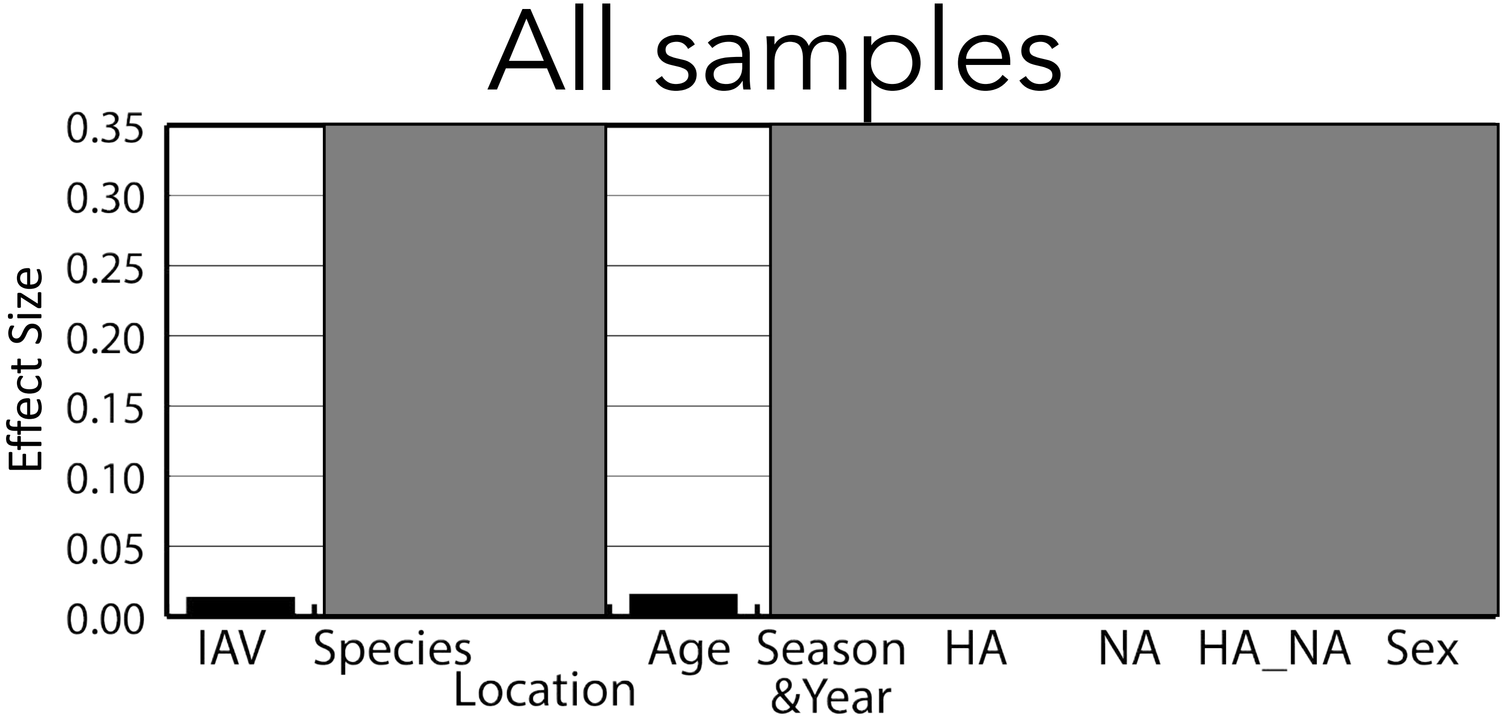
Figures modified from ©Kay Cichini

@sarahmhird

Q2: *Adonis* TESTS

What is associated with the microbiome?

black: $p \leq 0.001$
gray: $p > 0.001$



IAV and Age are legitimately significant.
(With <u>very</u> low effect sizes)

@sarahmhird

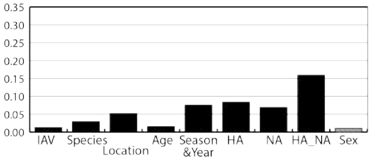
Q2: *Adonis* TESTS

What is associated with the microbiome?

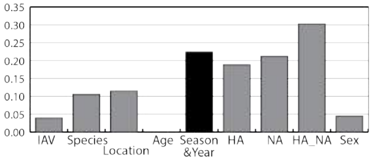
black: $p \leq 0.001$

gray: $p > 0.001$

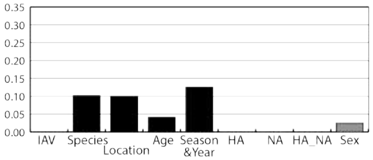
AWF



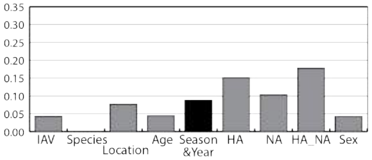
adult



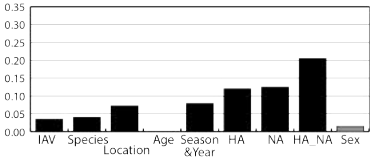
fluneg



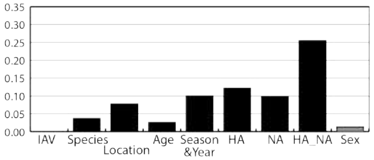
acuta



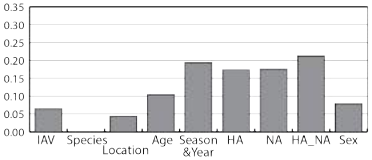
hatchyear



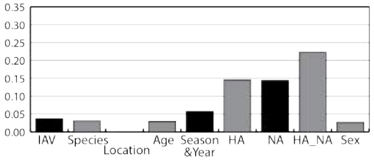
flupos



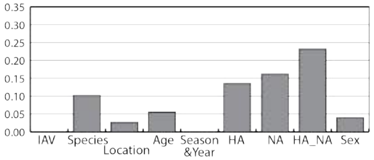
americana



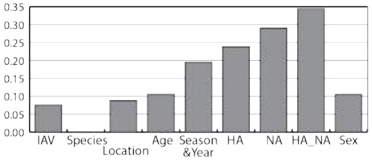
snwr



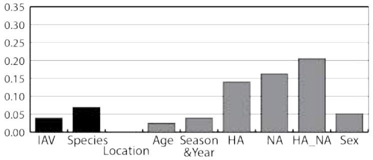
fw0809



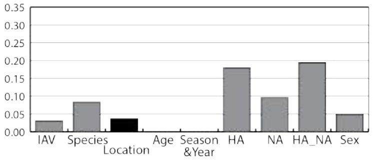
carolinensis



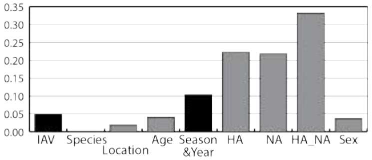
conranch



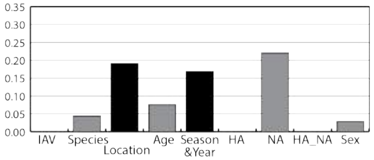
fw1011



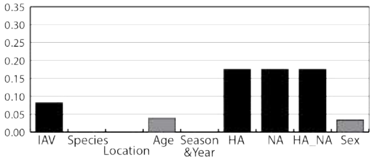
clypeata



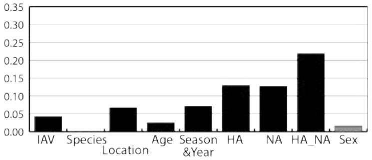
h4



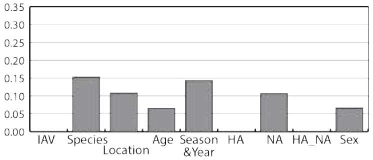
s13



platy



h6



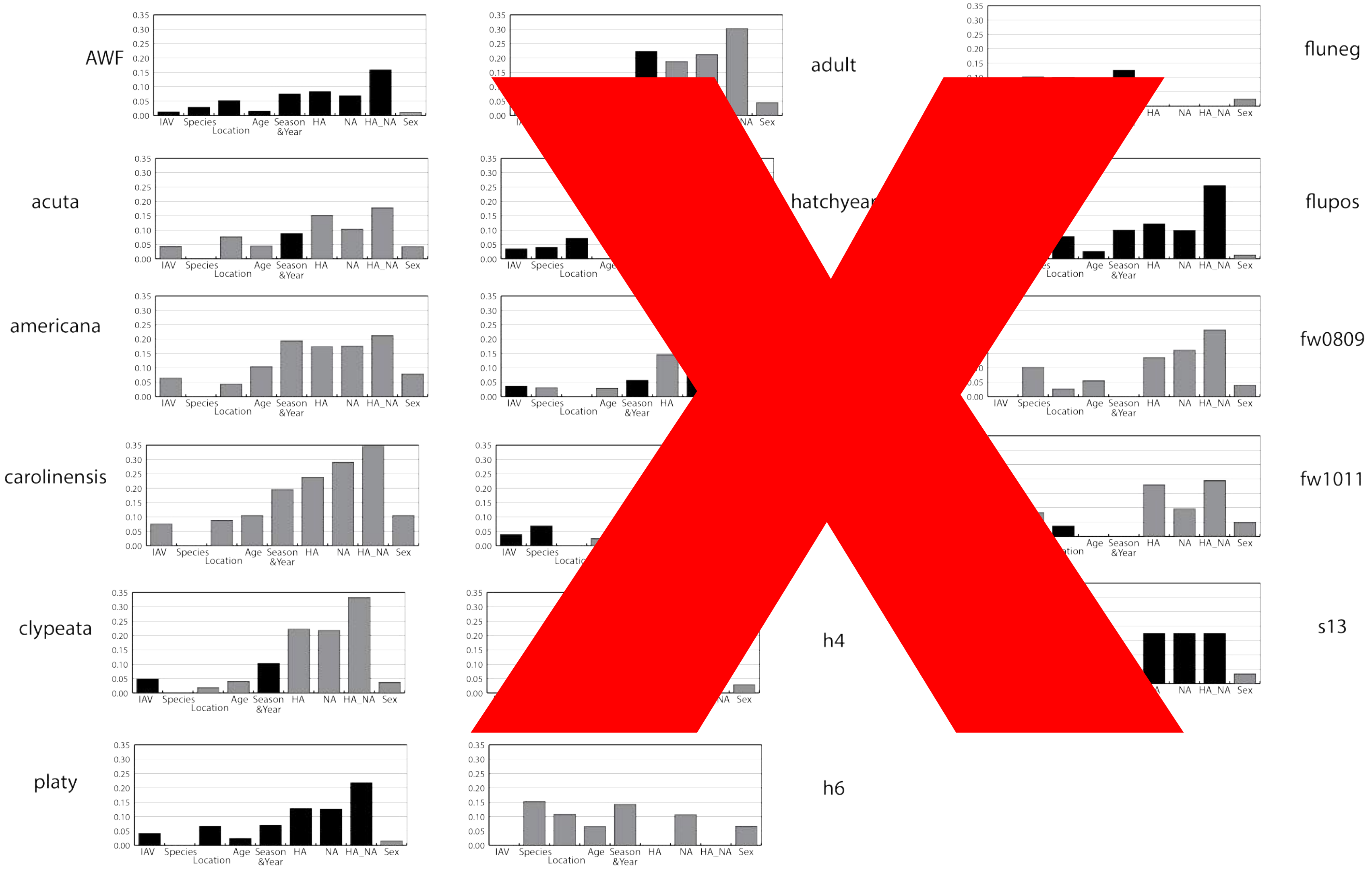
@sarahmhird

Q2: *Adonis* TESTS

What is associated with the microbiome?

black: $p \leq 0.001$

gray: $p > 0.001$

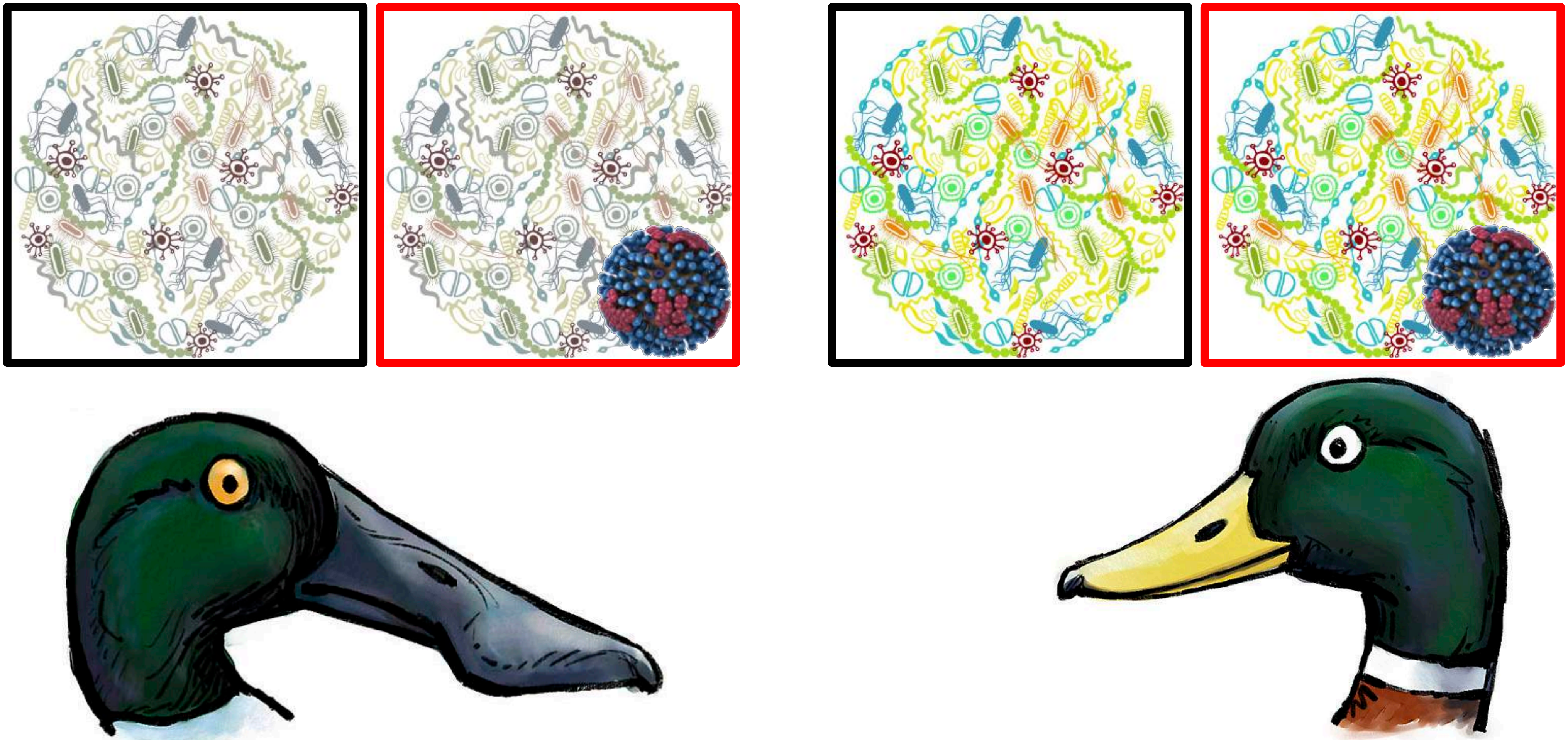


@sarahmhird

CONCLUSION

Question 2

Does IAV affect the microbiomes of different wild duck species in the same way?

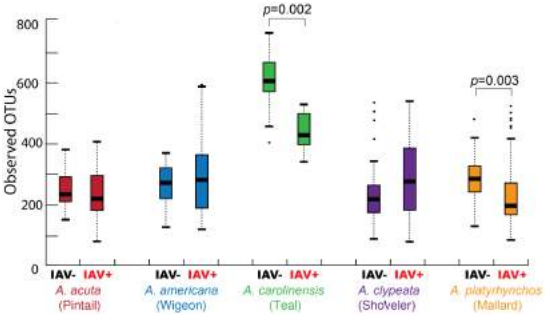


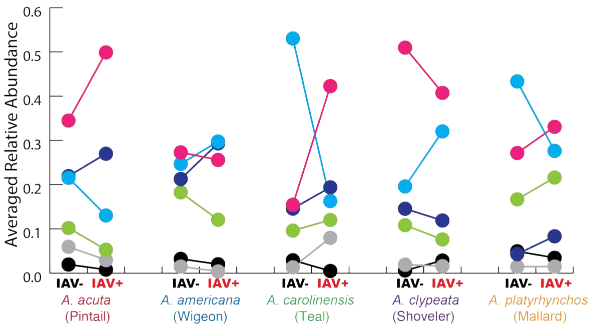
CONCLUSION

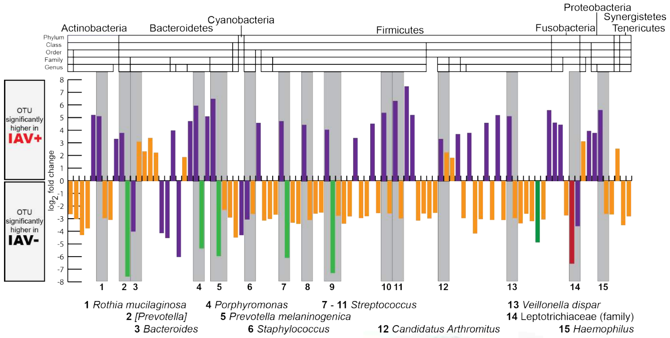
Question 2

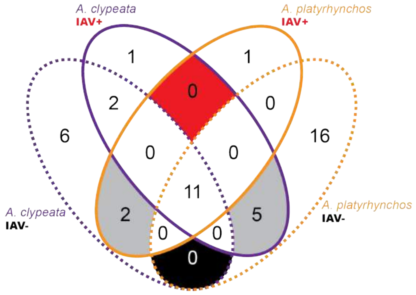
Does IAV affect the microbiomes of different wild duck species in the same way?

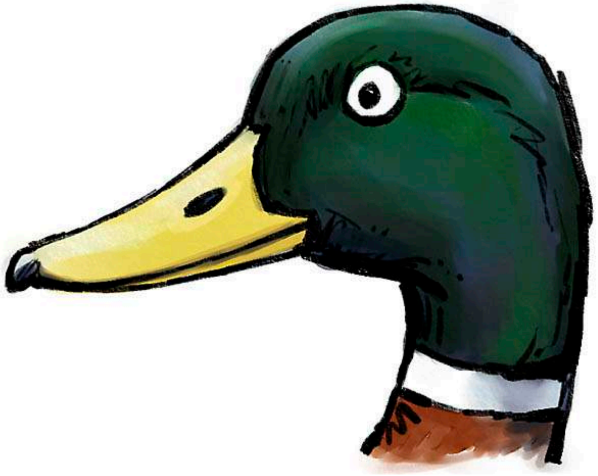
No. But…











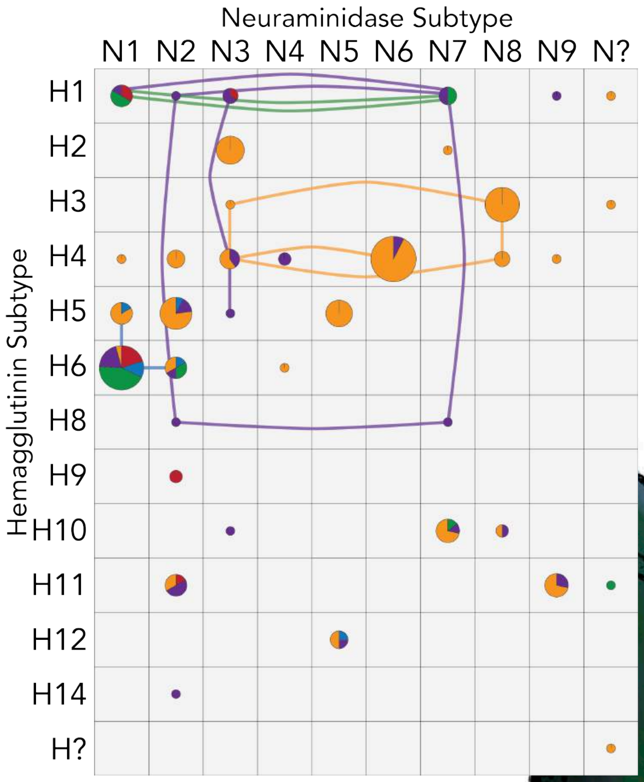
CONCLUSION

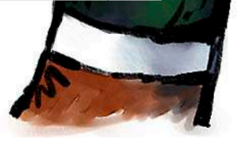
Question 2

Does IAV affect the microbiomes of different wild duck species in the same way?

No. But…

		SAMPLING LOCALITY				
	N	CR	GIWA	SNWR	MI	SMAR
All Ducks	**300**	**46**	**110**	**72**	**19**	**29**
A. acuta	31	14	3	14	0	0
A. americana	25	0	1	0	0	0
A. carolinensis	19	13	0	6	0	0
A. clypeata	57	0	5	52	0	0
A. platyrhynchos	168	19	101	0	19	29





Therizinosaurus

DK Find Out!

sarah.hird@uconn.edu
@sarahmhird
HIRDLAB.COM

Thank you.

Collaborators and Contributors
Walter Boyce
Holly Ganz
Jonathan Eisen
Magdalena Plancarte
Matt Rolston
Alana Firl
Ladan Doroud
Noah Reid

Funding
NIH:CEIRS - HHSN272201400008C (WMB)
NIH: CEIRS - HHSN266200700010C (WMB)
UC Davis Chancellor's Postdoctoral Fellowship (SMH)
University of Connecticut

UCONN
Molecular and Cell Biology

NIH

UCDAVIS
UNIVERSITY OF CALIFORNIA

UCDAVIS
VETERINARY MEDICINE

My lab

Dr. Kirsten Grond
(me) Darien Capunitan
Beth Herder

My family

Dr. Noah Reid

Thank you.

sarah.hird@uconn.edu
@sarahmhird
HIRDLAB.COM

Collaborators and Contributors
Walter Boyce
Holly Ganz
Jonathan Eisen
Magdalena Plancarte
Matt Rolston
Alana Firl
Ladan Doroud
Noah Reid

Funding
NIH:CEIRS - HHSN272201400008C (WMB)
NIH: CEIRS - HHSN266200700010C (WMB)
UC Davis Chancellor's Postdoctoral Fellowship (SMH)
University of Connecticut

UCONN
Molecular and Cell Biology

NIH

UCDAVIS
UNIVERSITY OF CALIFORNIA

UCDAVIS
VETERINARY MEDICINE

Any questions? Or comments!

Archaeopteryx
DK Find Out!

My lab

Dr. Kirsten Grond
Beth Herder
(me) Darien Capunitan

My family

Dr. Noah Reid