

Amplicon Sequencing: *OTUs and ASVs*

A Microbial Census

Metabarcoding or Marker-gene Sequencing

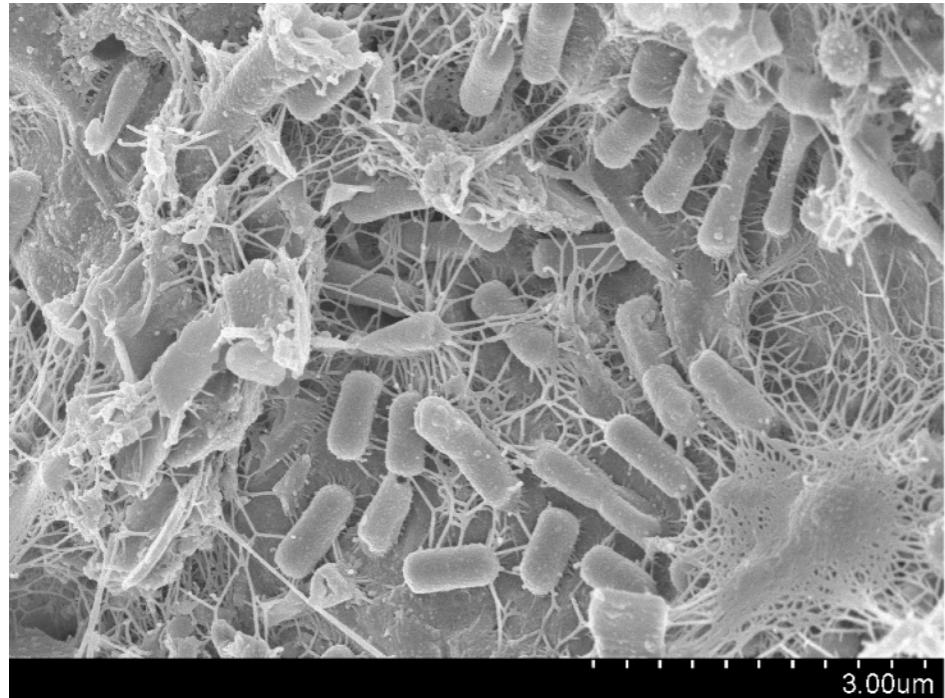


Image: Anthony D'Onofrio, Northeastern University, 2009

A Microbial Census

Metabarcoding or Marker-gene Sequencing

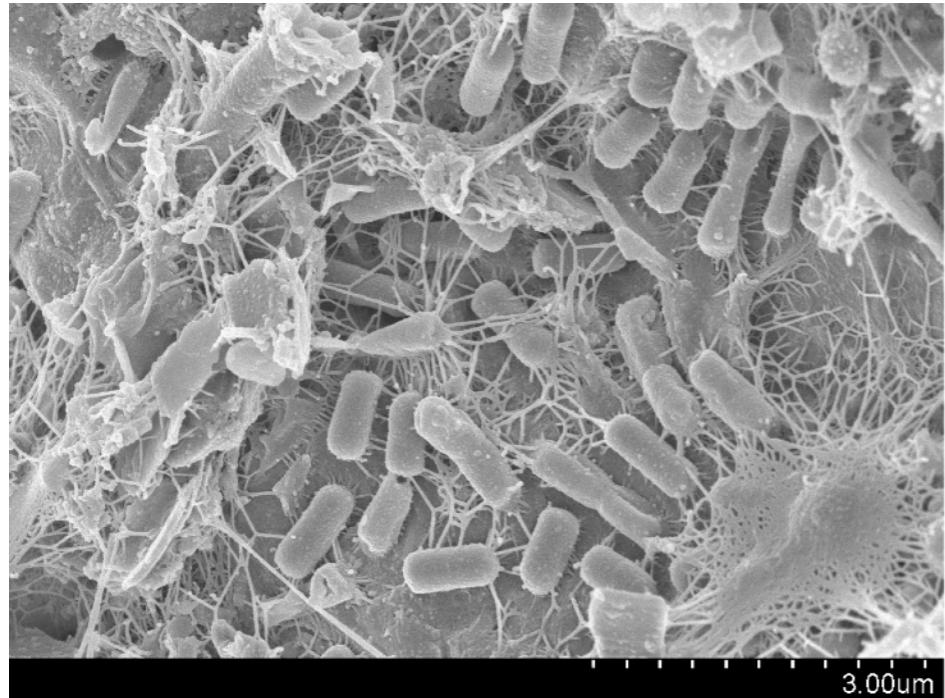
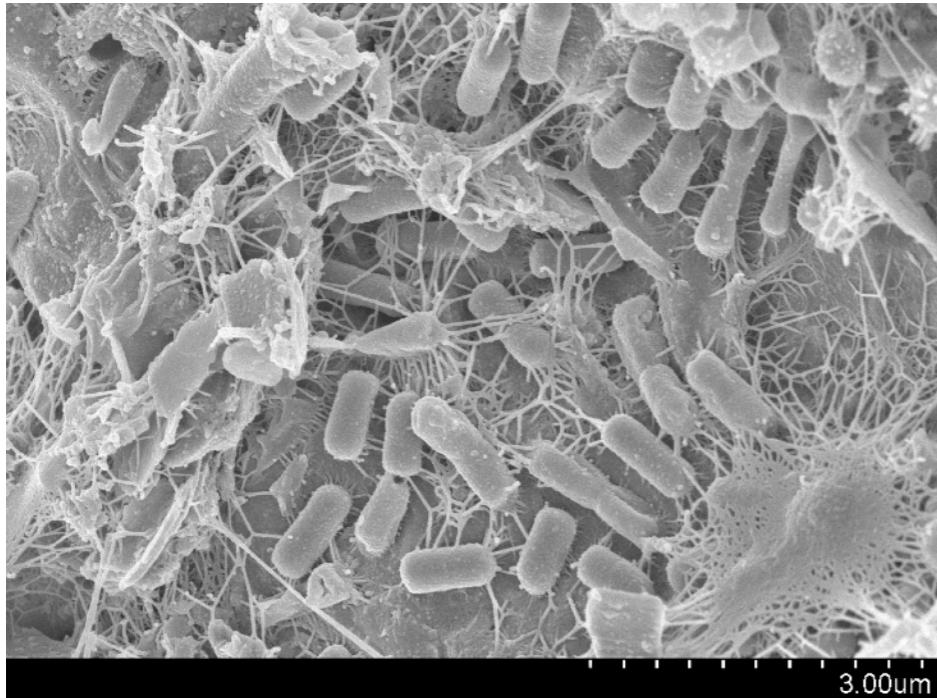


Image: Anthony D'Onofrio, Northeastern University, 2009

A Microbial Census

Metabarcoding or Marker-gene Sequencing

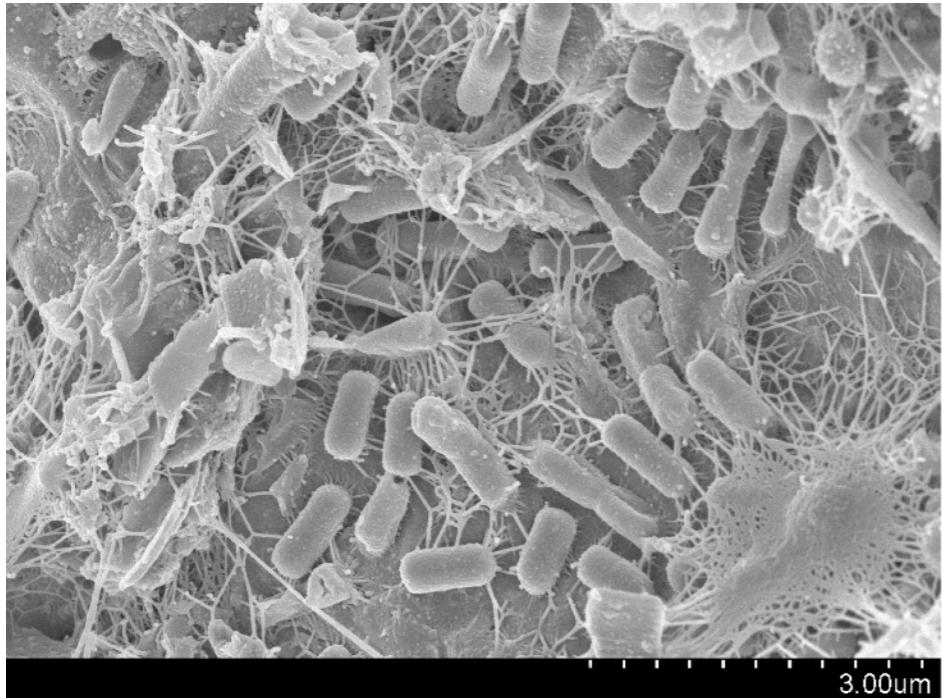


<i>Lactobacillus crispatus</i>	1300	5	0	882	596
<i>Ureaplasma urealytica</i>	15	0	220	0	0
<i>Gardnerella vaginalis</i>	22	0	1	0	412
<i>Prevotella intermedia</i>	0	0	8	12	0
...

Image: Anthony D'Onofrio, Northeastern University, 2009

A Microbial Census

Metabarcoding or Marker-gene Sequencing



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

Visualization

Inference

Exploration

Marker-gene Sequencing

300 bp

2,000,000 bp

Marker-gene Sequencing

300 bp

2,000,000 bp

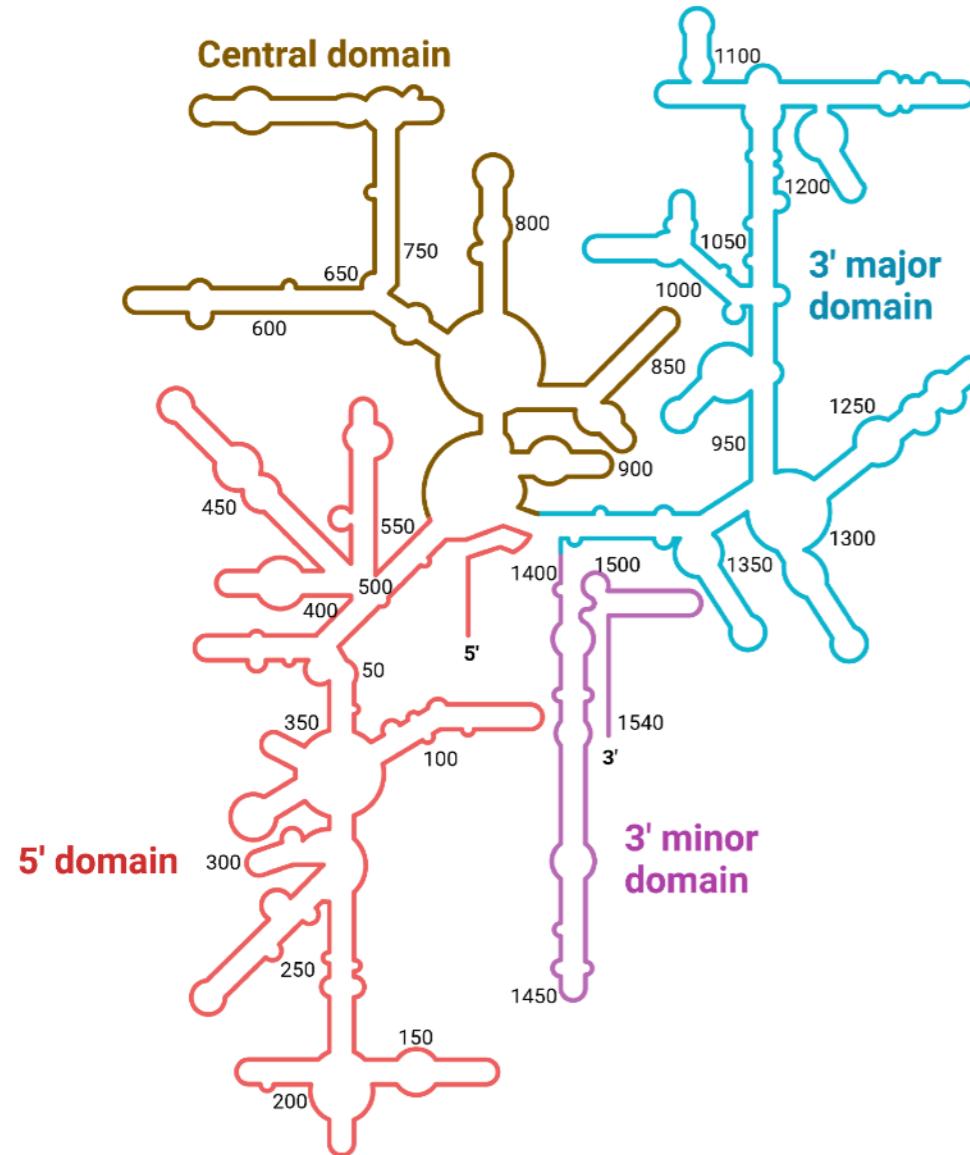


Selected by choice of PCR primers

**Example:
16S rRNA gene sequencing**

A barcode of life

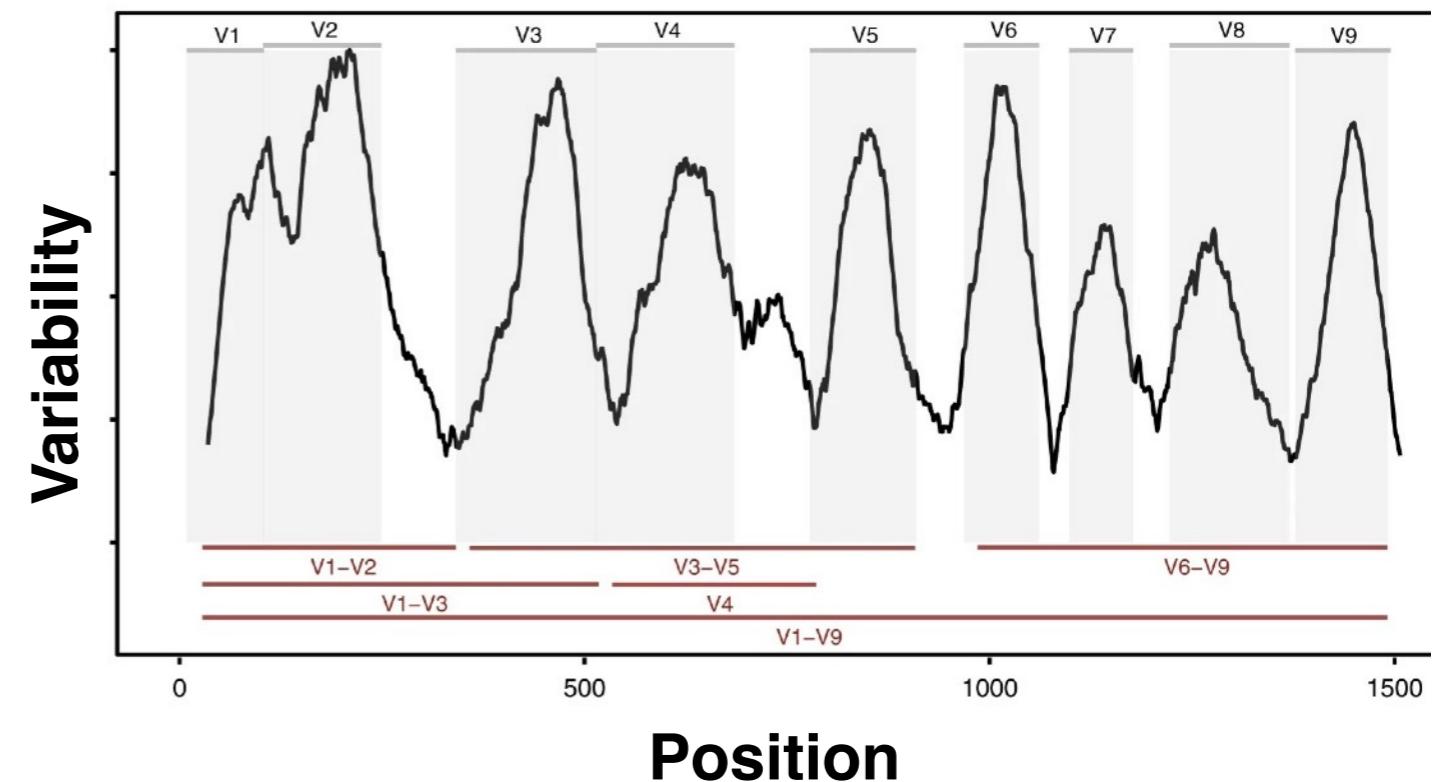
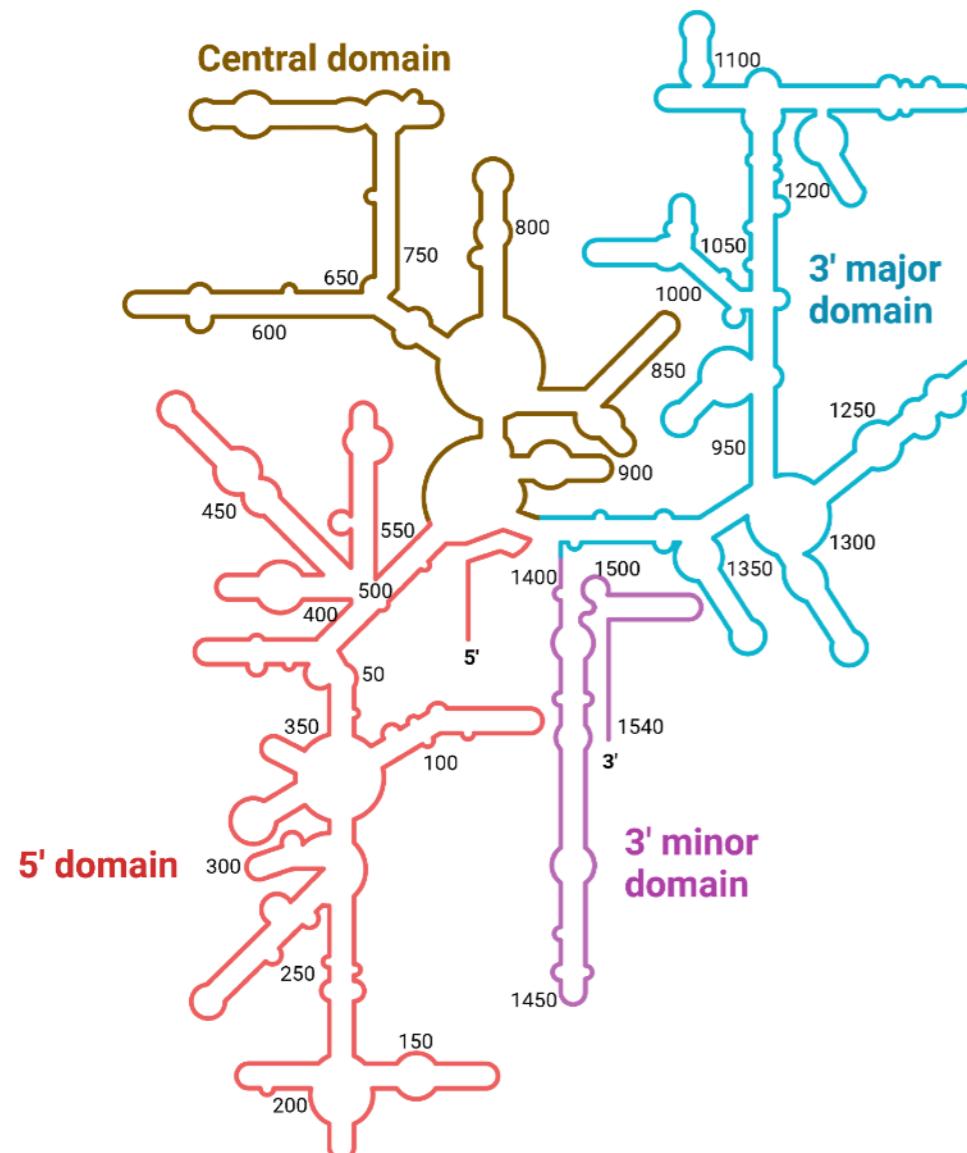
Secondary Structure of 16S rRNA



Images: Biorender; Johnson et al., Nat Commun, 2019.

A barcode of life

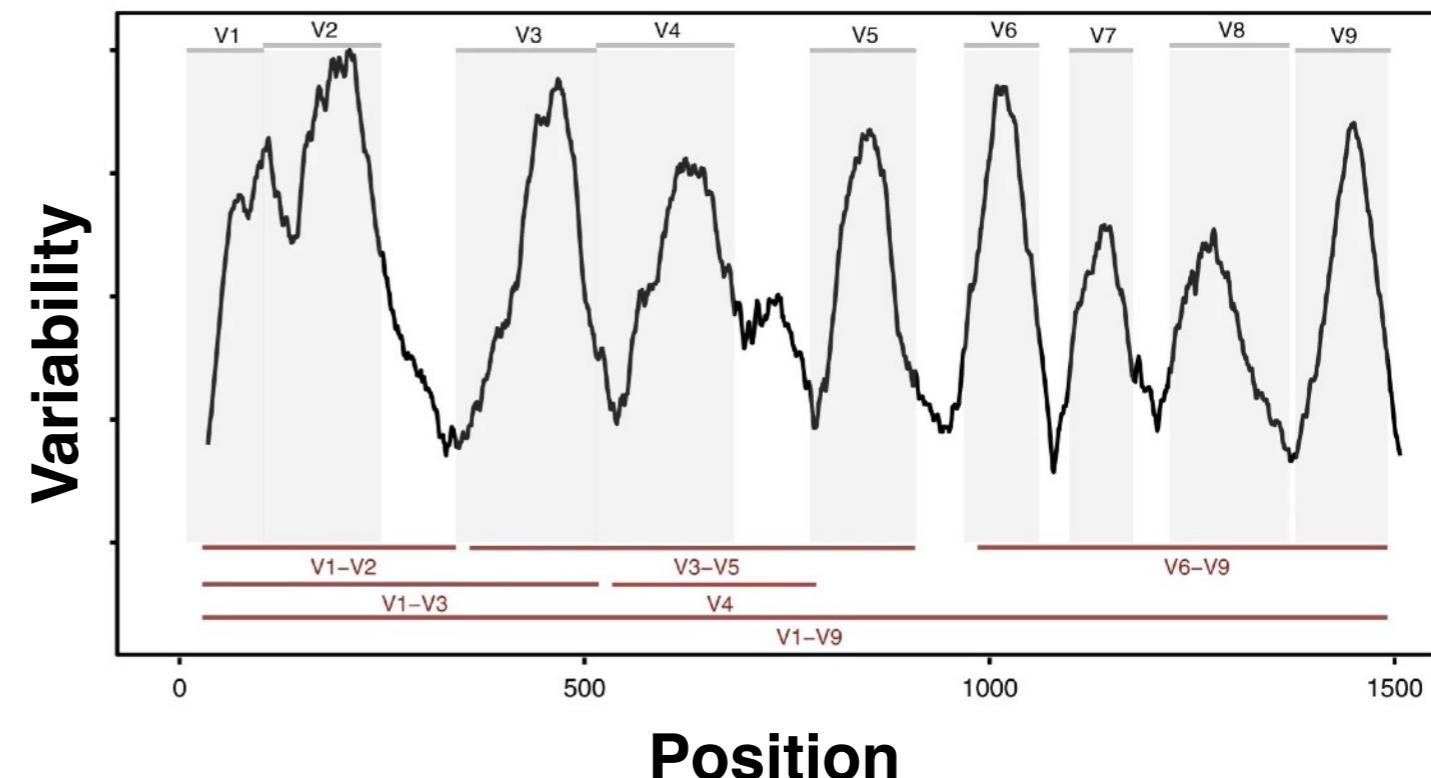
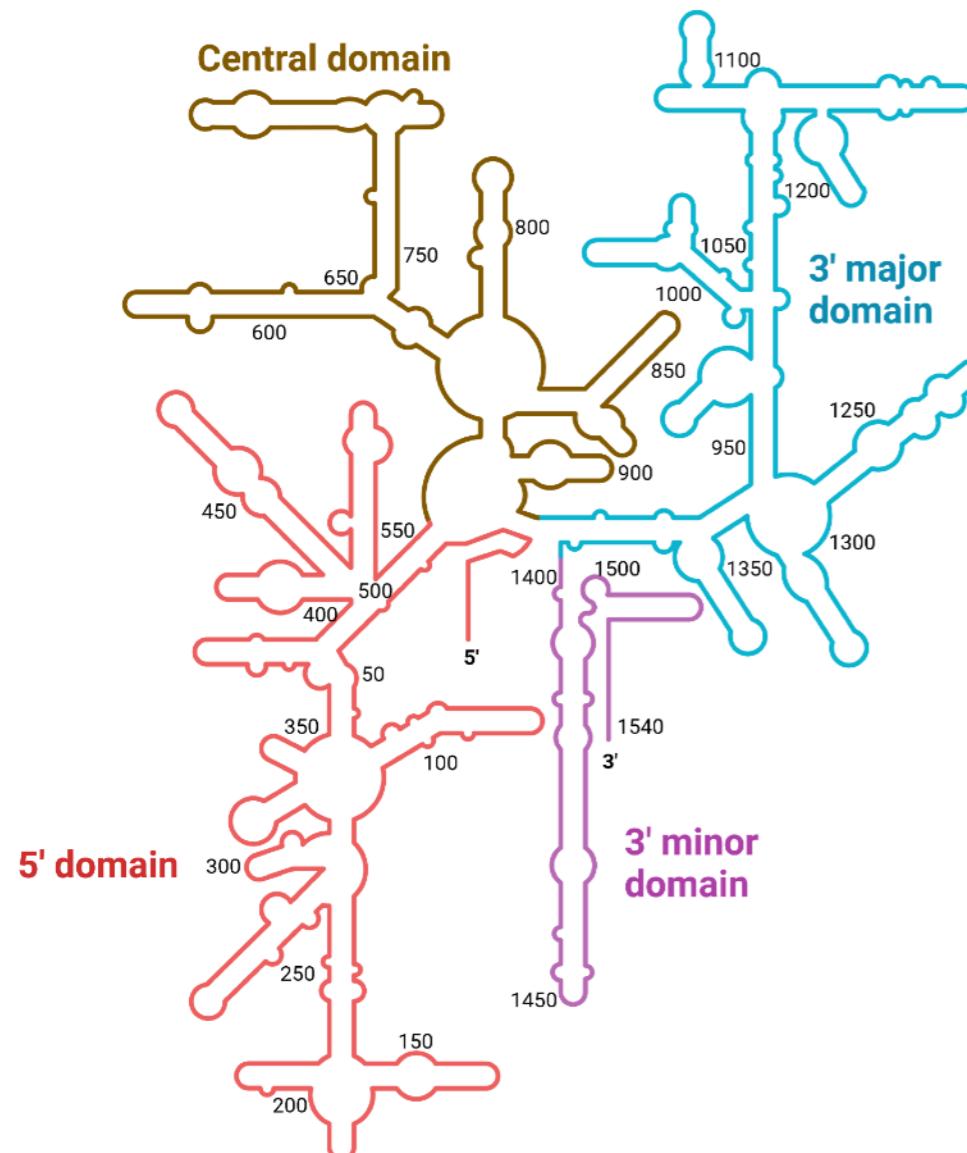
Secondary Structure of 16S rRNA



Images: Biorender; Johnson et al., Nat Commun, 2019.

A barcode of life

Secondary Structure of 16S rRNA



Why 16S? Universal. Conserved primer sites flanking variable “barcode” regions.

Barcodes of life

Common barcodes

16S: Bacteria, Archaea

CO1: Animals

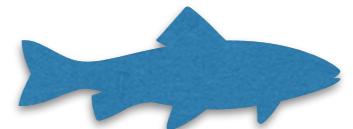
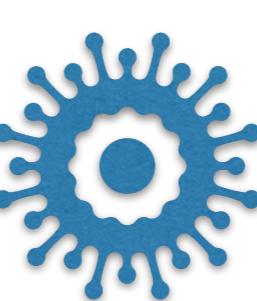
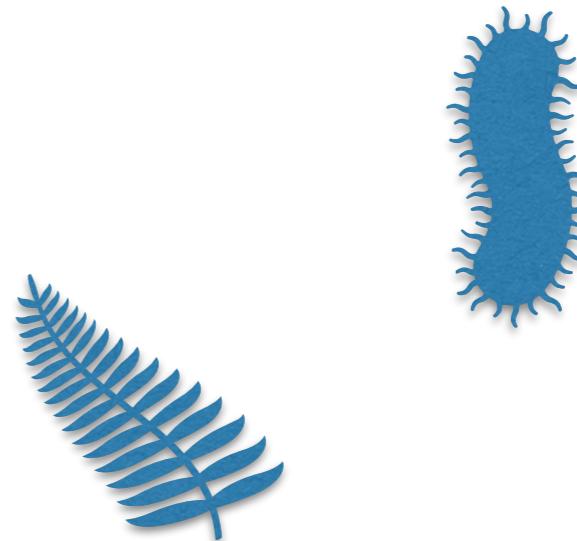
ITS: Plants, Fungi

trnL: Plants, Food

12S: Fish, Food

18S: Eukaryotes

And many, many more!



Marker-gene Sequencing

300 bp

2,000,000 bp

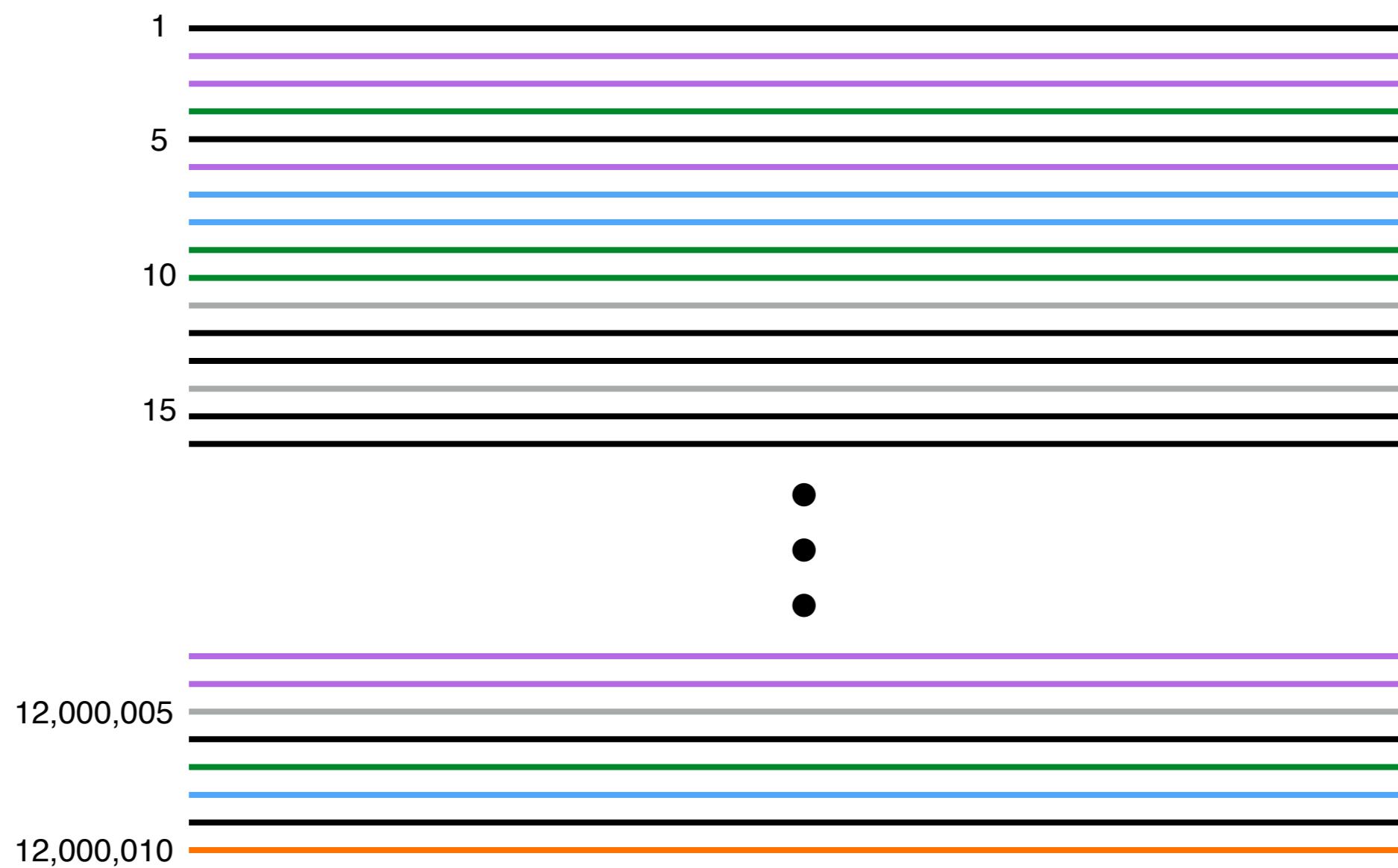


Selected by choice of PCR primers

Marker-gene Sequencing

300 bp

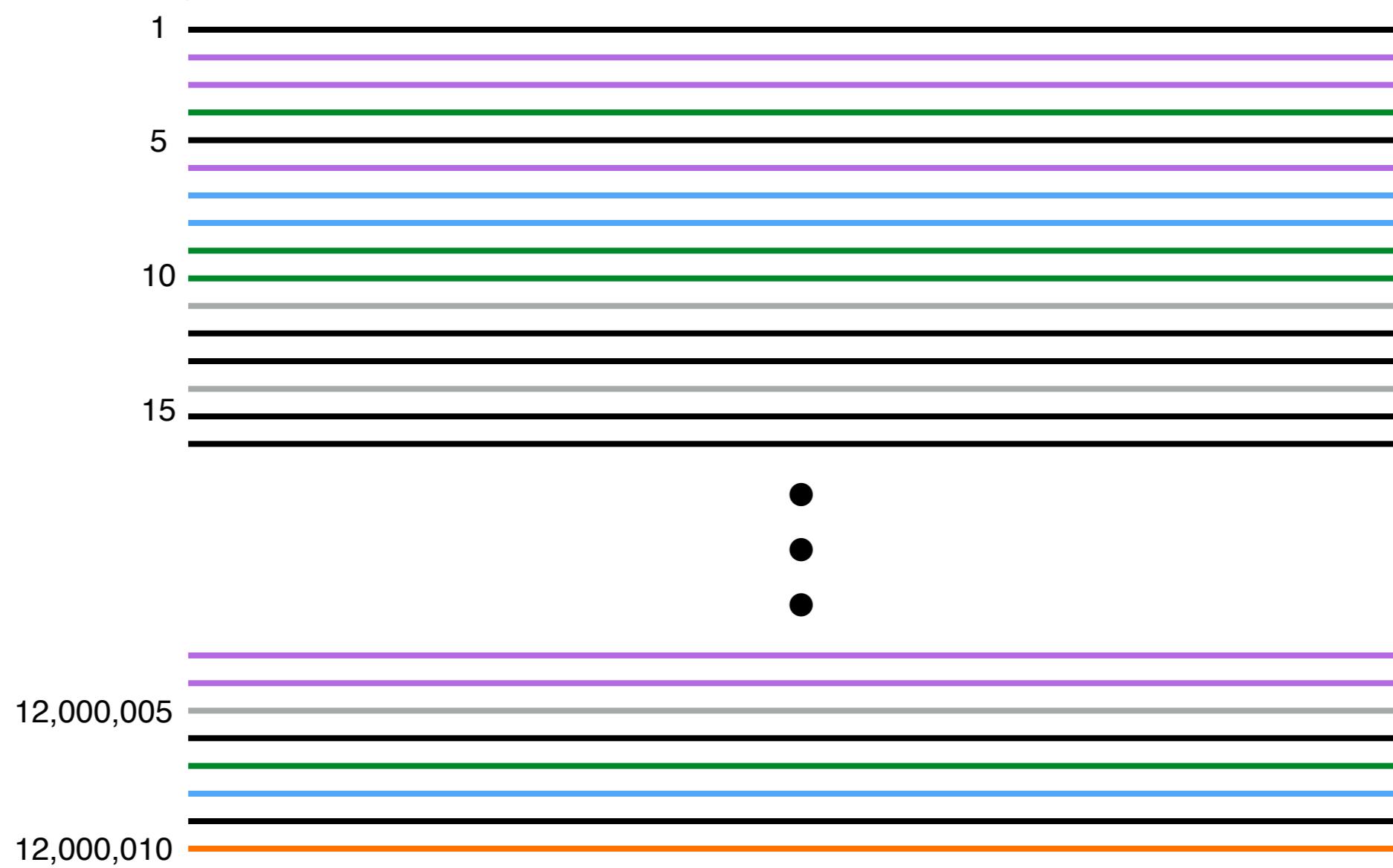
2,000,000 bp



Marker-gene Sequencing

300 bp

2,000,000 bp

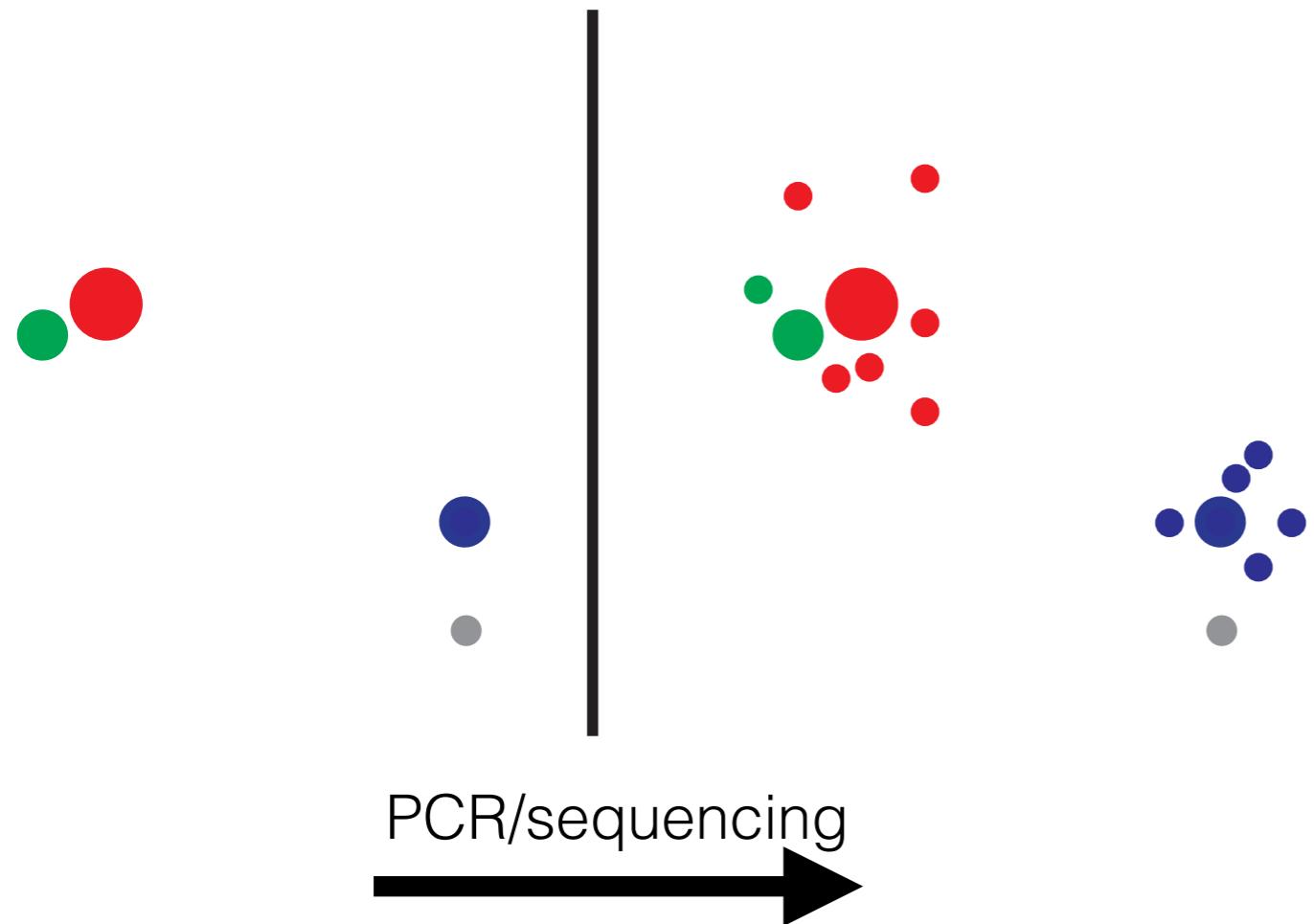


Thousands to Billions of “Barcodes”

OTUs and ASVs

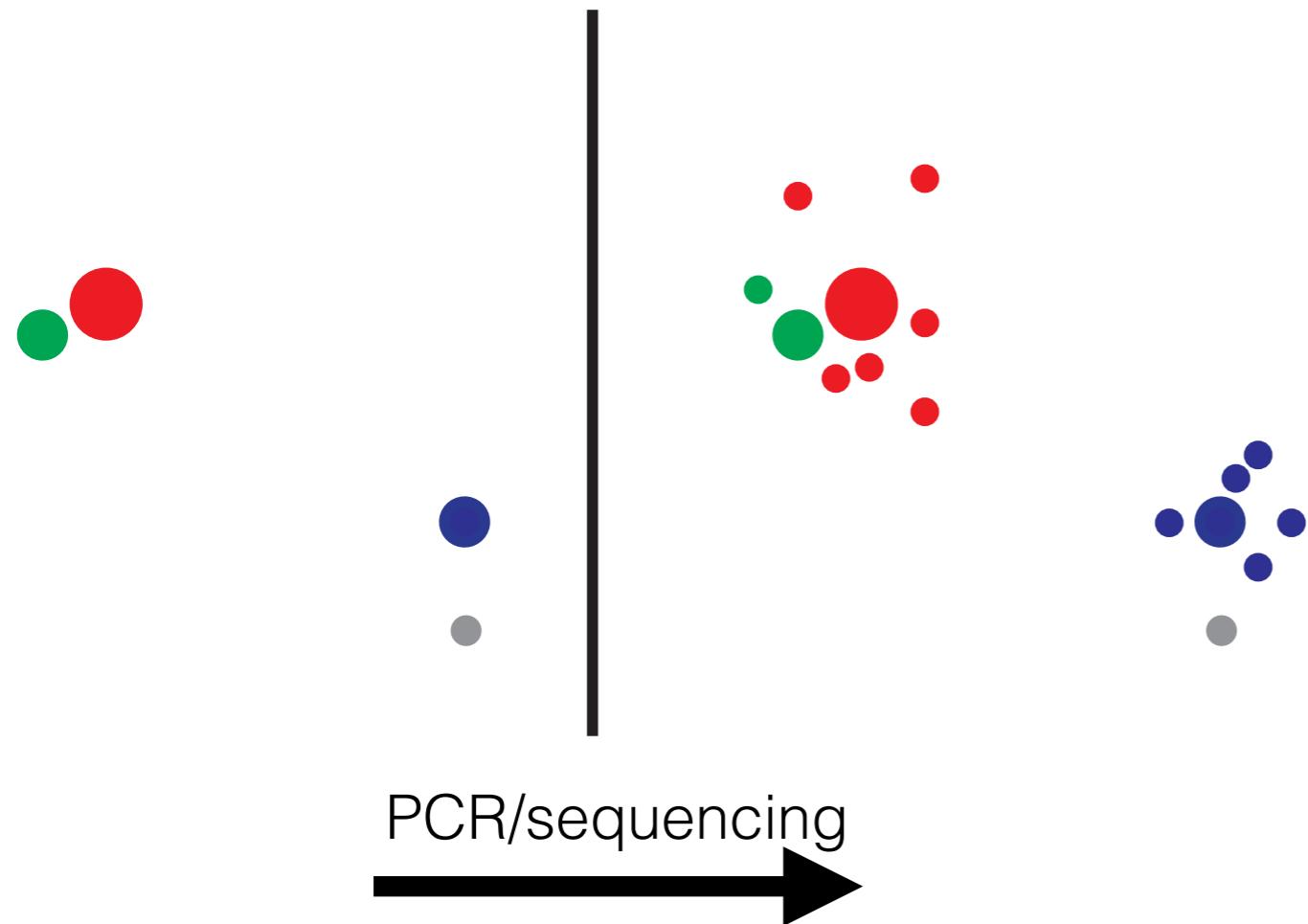
Sample Sequences

Amplicon Reads



Sample Sequences

Amplicon Reads

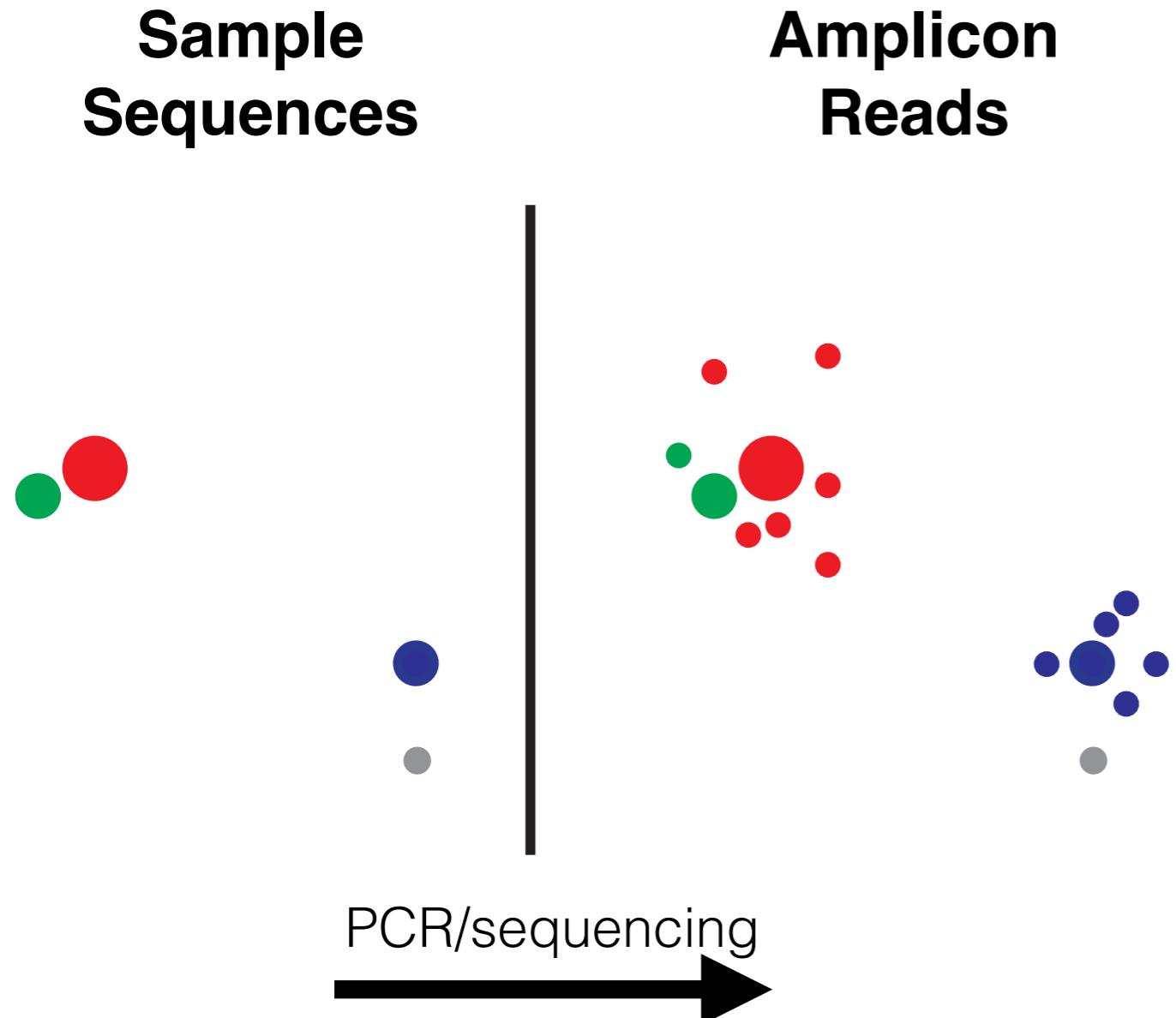


Two Challenges

How do we deal with errors?

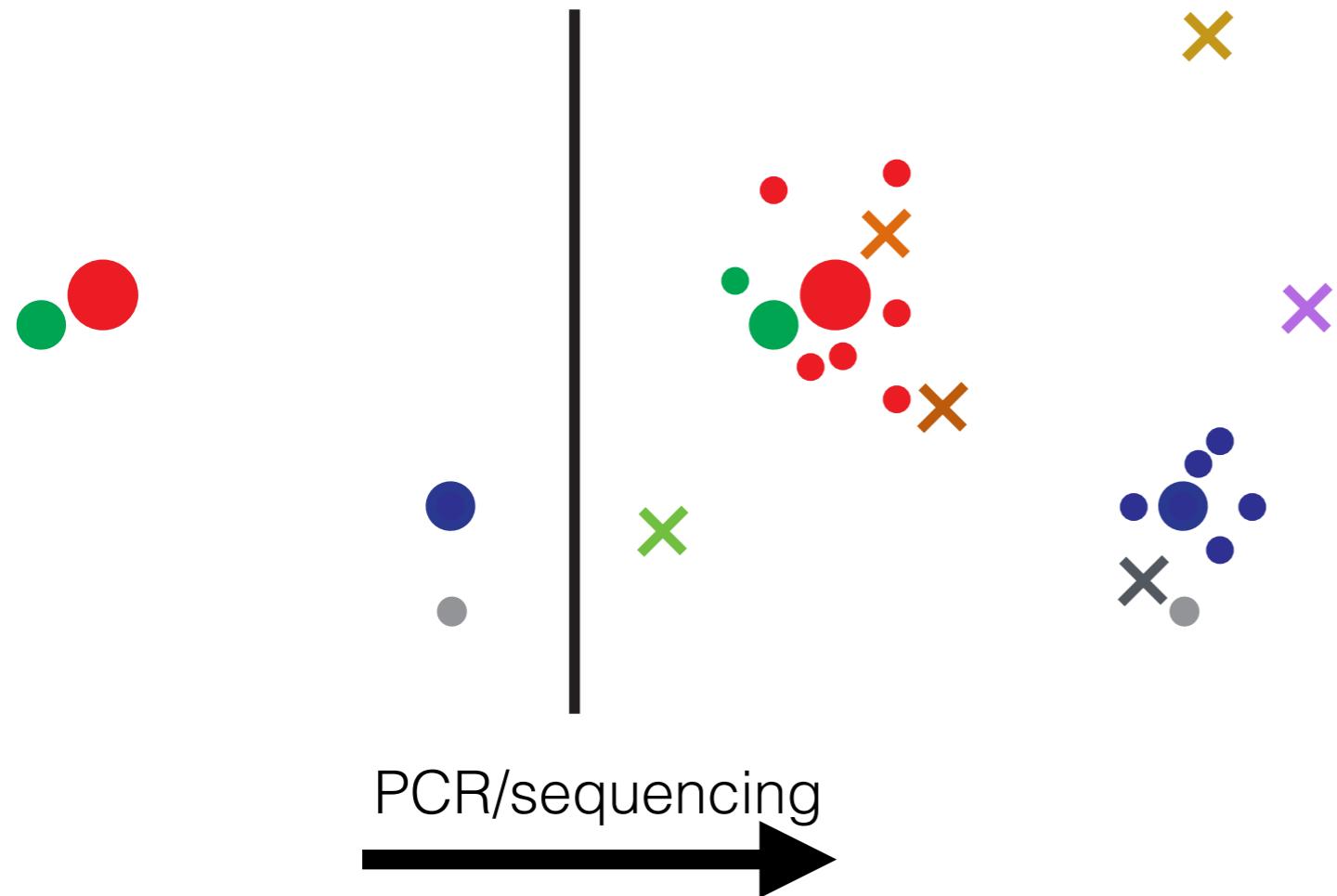
How do we define the units of our analysis?

Closed reference OTUs

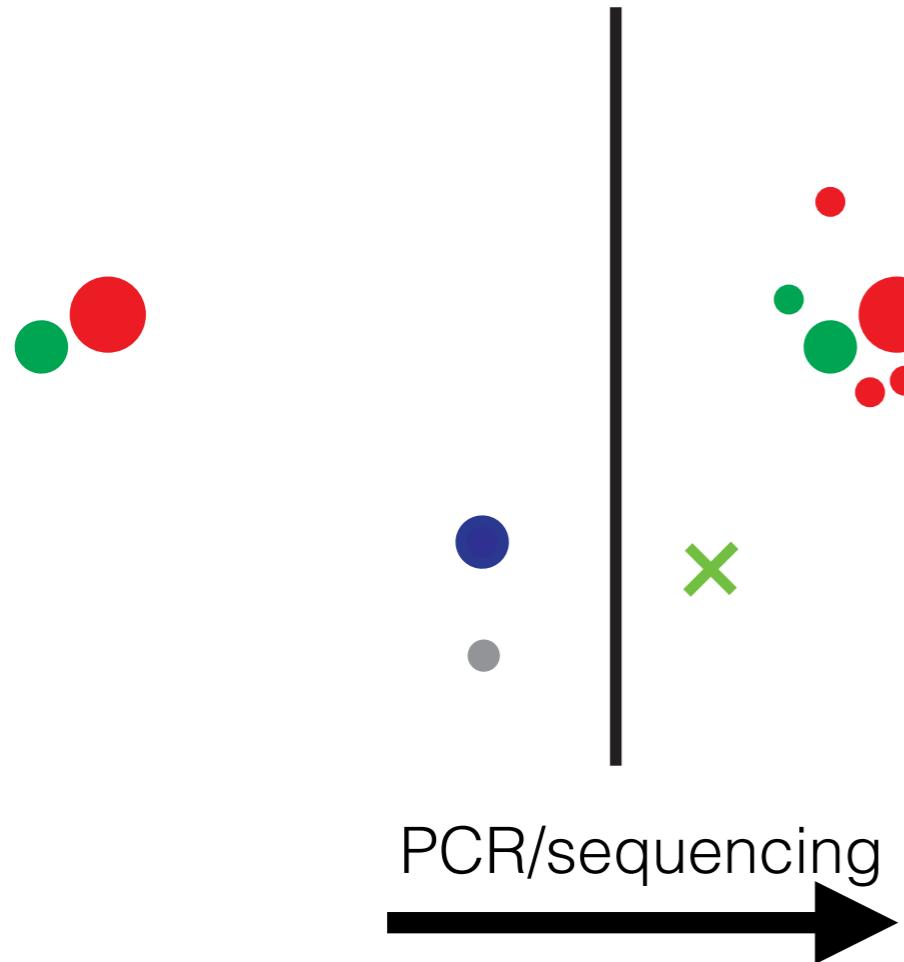


Sample Sequences

Amplicon Reads



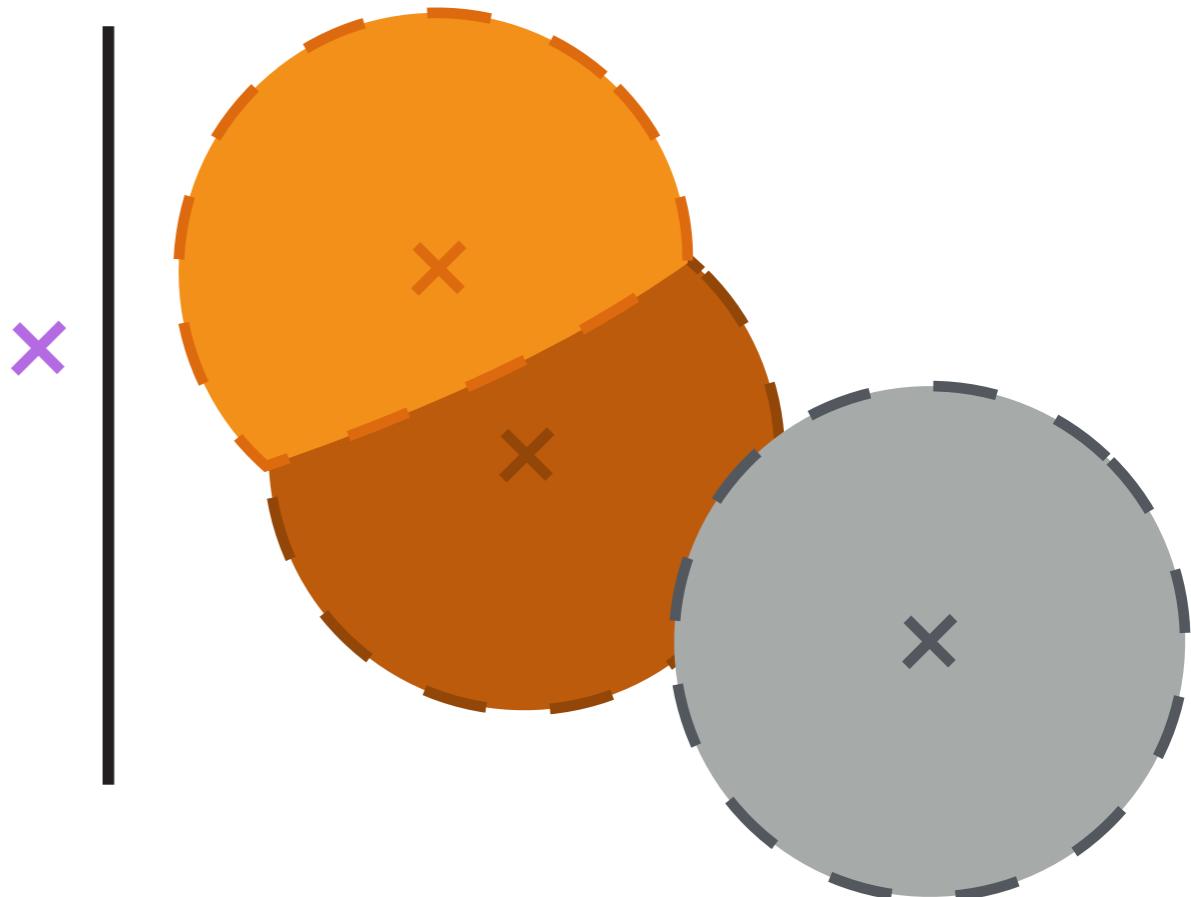
Sample Sequences



Amplicon Reads

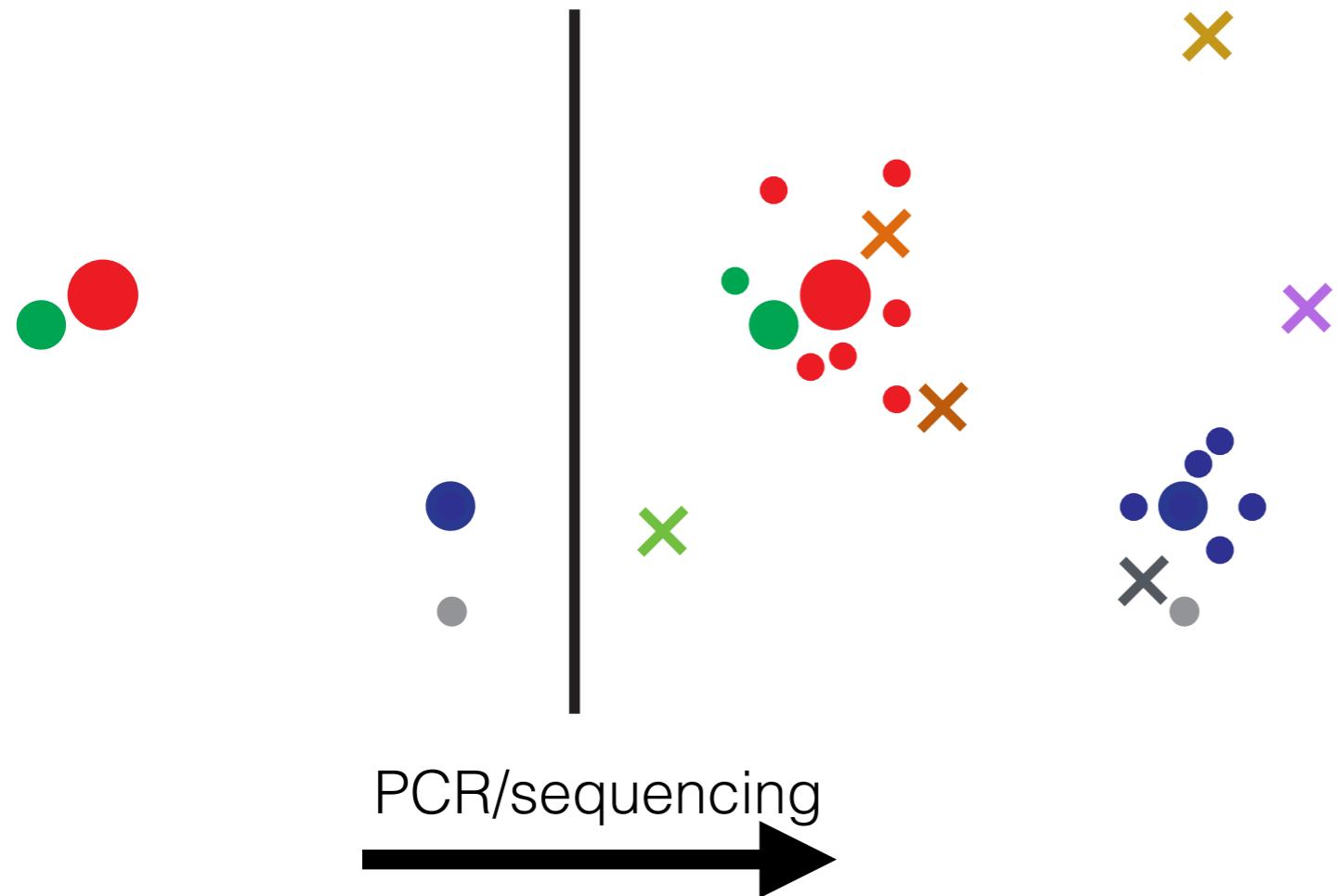


Operational Taxonomic Units (closed reference)



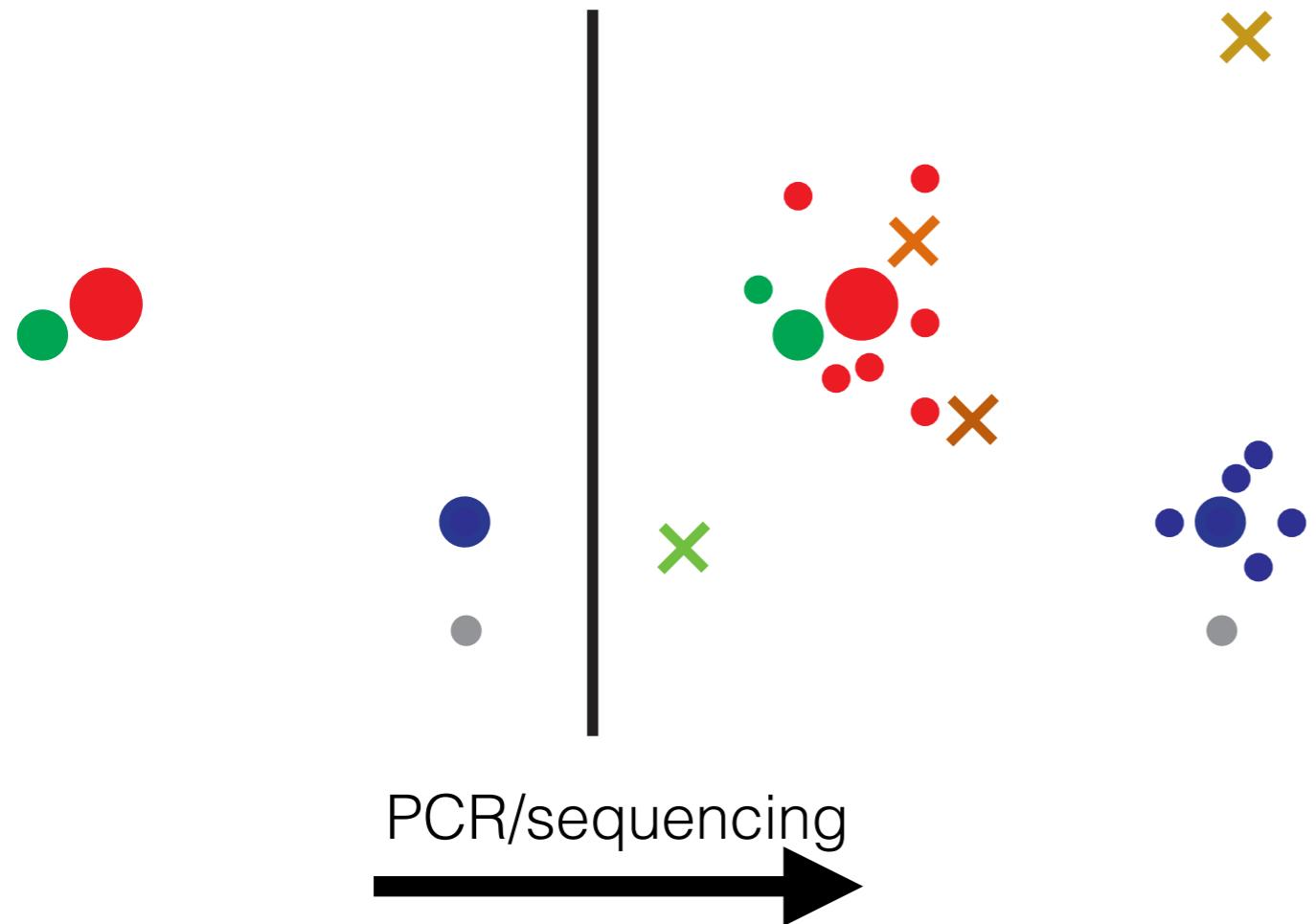
Sample Sequences

Amplicon Reads



Sample Sequences

Amplicon Reads



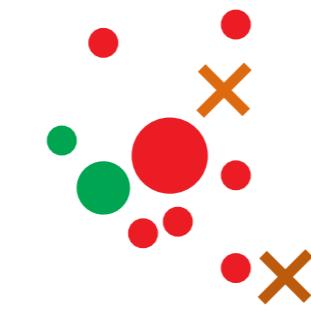
Sample Sequences



PCR/sequencing



Amplicon Reads

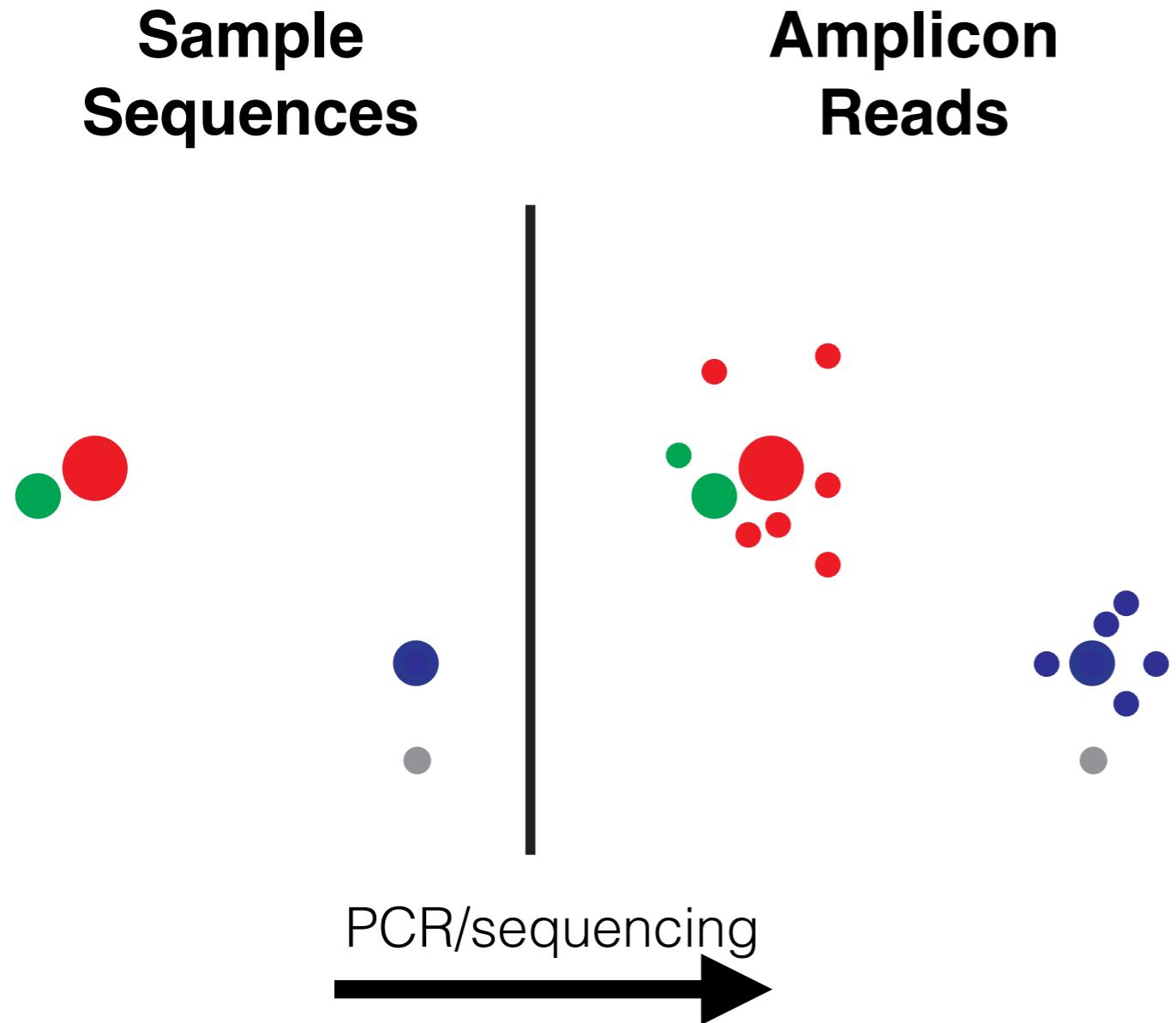


Operational Taxonomic Units (closed reference)



Incomplete OTU table!

de novo OTUs



Sample Sequences

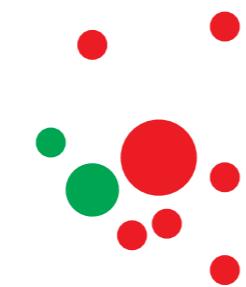


⋮

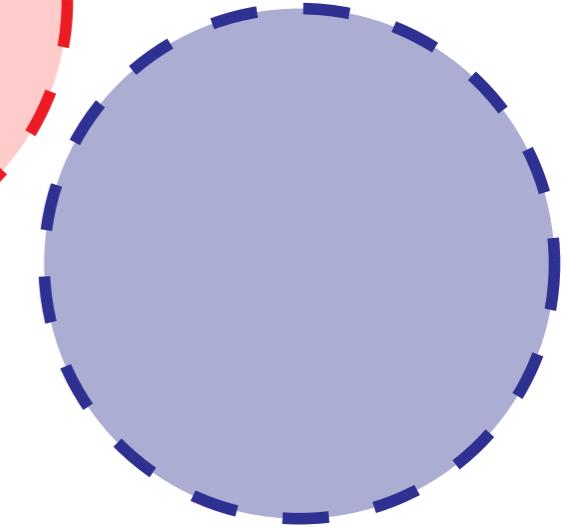
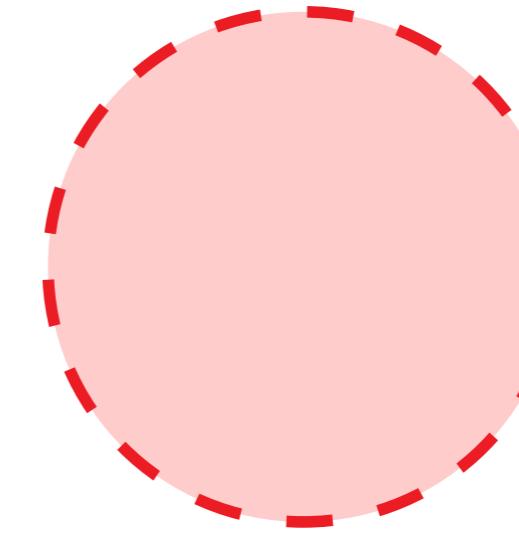
PCR/sequencing



Amplicon Reads



Operational Taxonomic Units (de novo)

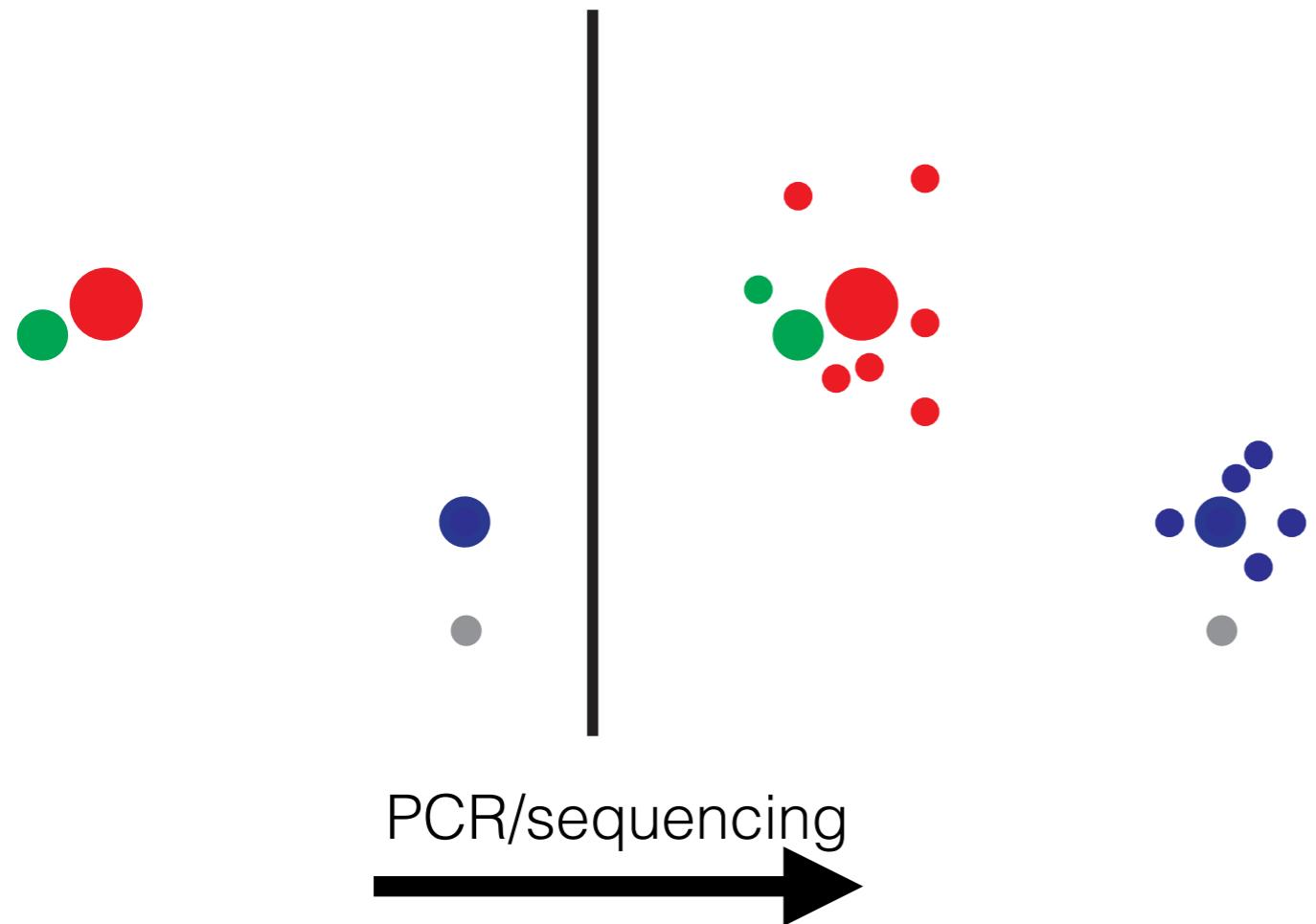


Pick OTUs



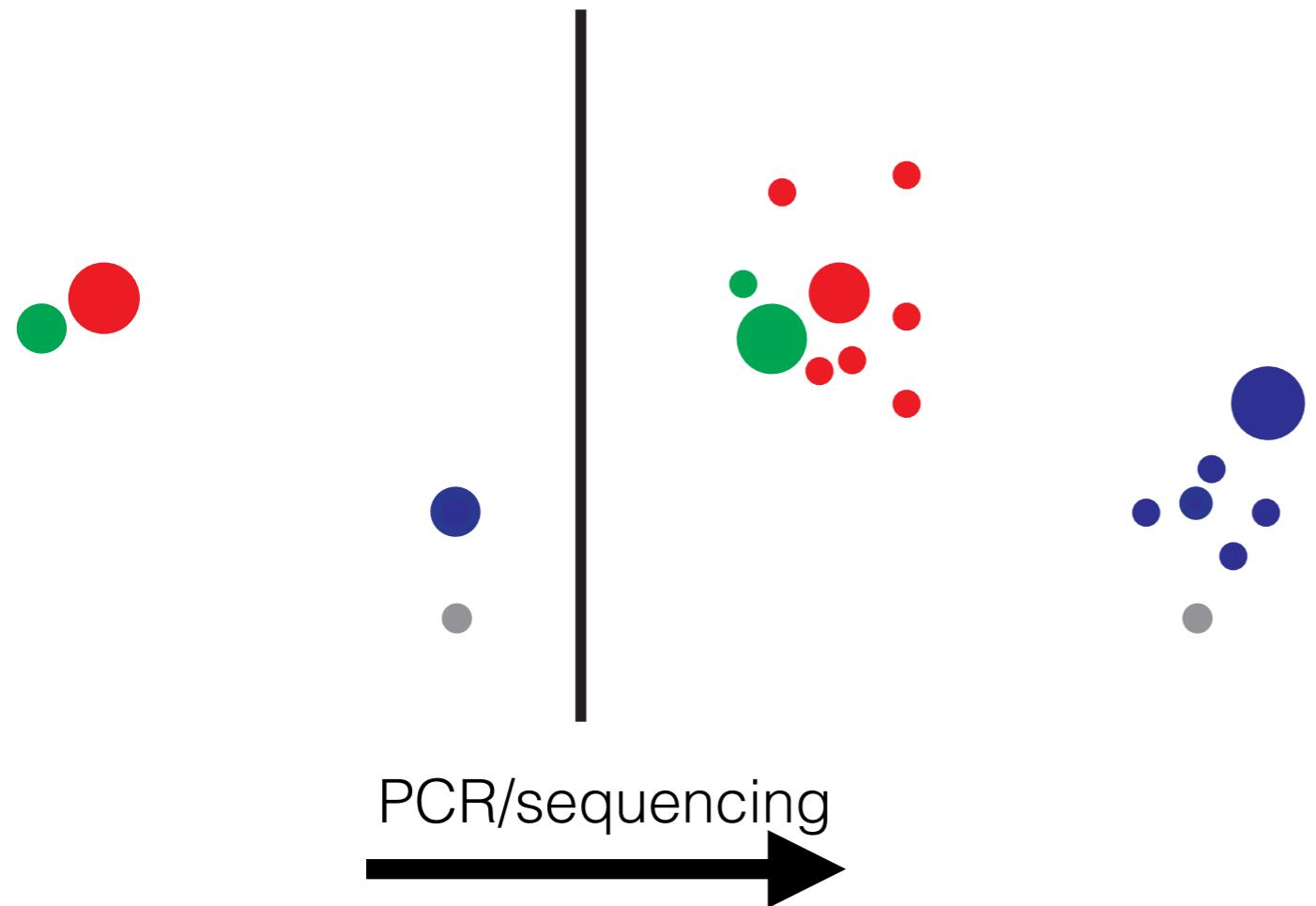
Sample Sequences

Amplicon Reads



Sample Sequences

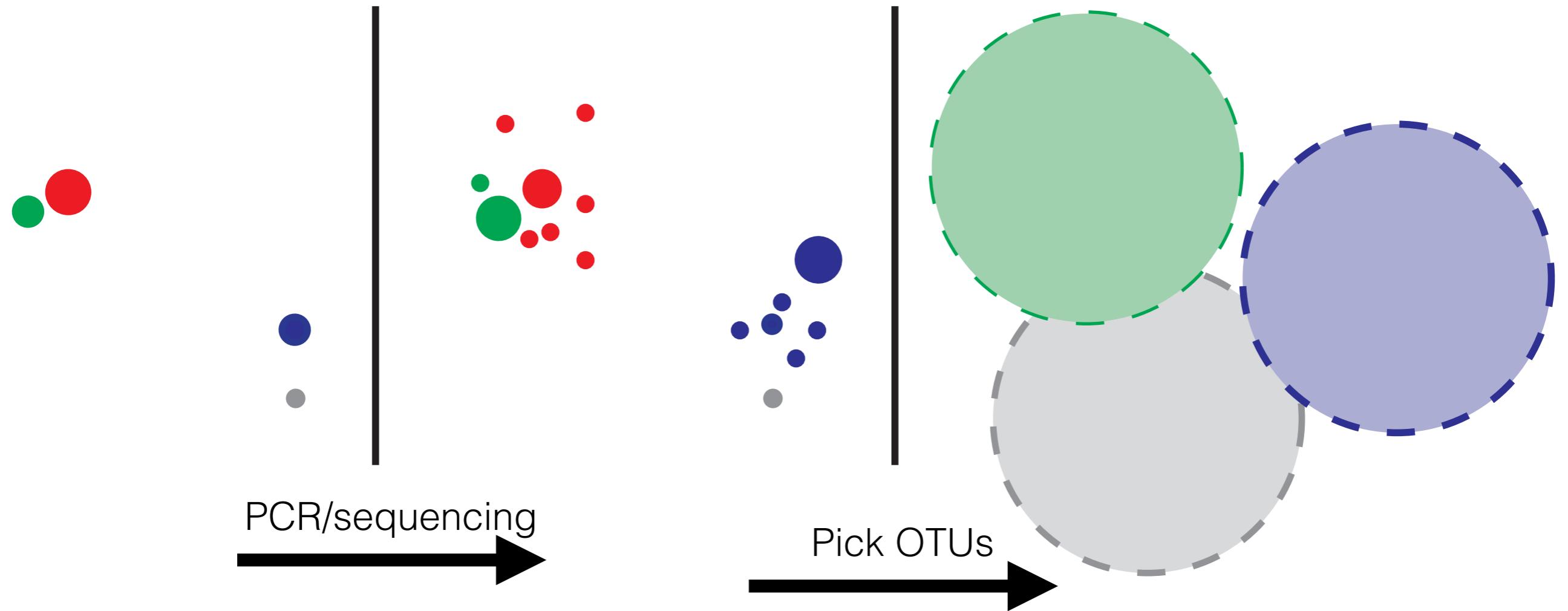
Amplicon Reads



Sample Sequences

Amplicon Reads

Operational Taxonomic Units (de novo)



Different OTUs!

(In)consistent labels

OTU85 is *not a consistent label*

OTU85 is *predictive of a disease? Not in future data!*

OTU85 is *associated w/ X and Y? Can't be tested!*

OTU85 is *in this community? OTUs don't exist in nature!*

Exact Sequence Variants...

OPEN

The ISME Journal (2017), 1–5

www.nature.com/ismej

PERSPECTIVE

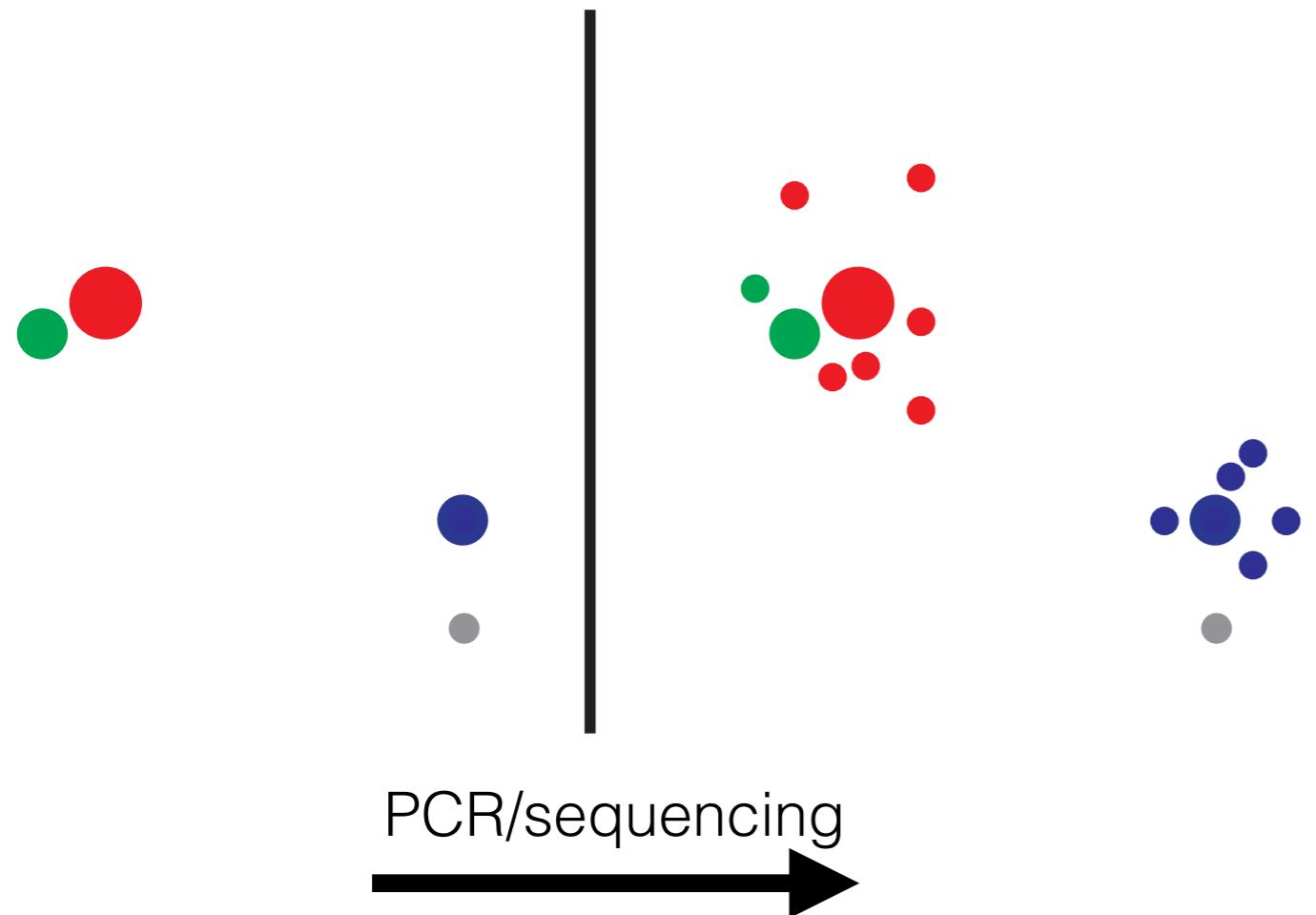
Exact sequence variants should replace operational taxonomic units in marker-gene data analysis

Benjamin J Callahan¹, Paul J McMurdie² and Susan P Holmes³

¹*Department of Population Health and Pathobiology, NC State University, Raleigh NC, USA;* ²*Whole Biome Inc, San Francisco CA, USA* and ³*Department of Statistics, Stanford University, Stanford CA, USA*

Sample Sequences

Amplicon Reads



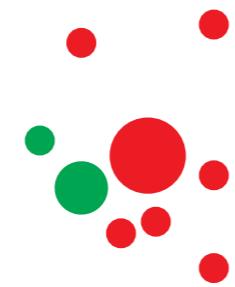
Sample Sequences



PCR/sequencing



Amplicon Reads



Remove/correct errors



Amplicon Sequence Variants (ASVs)



Exact Sequence Variants

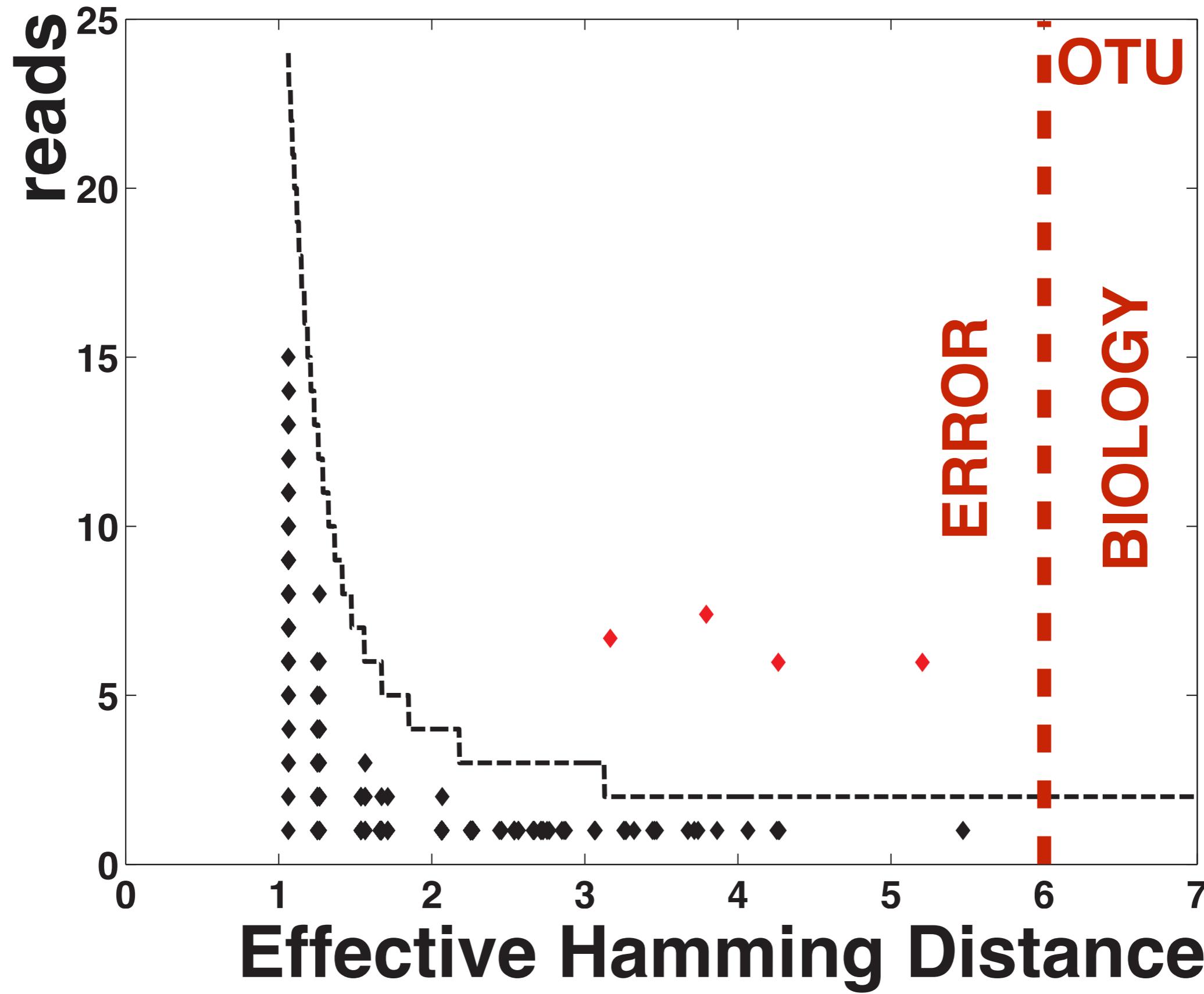
OTU85 is *not a consistent label, but...*

ATTAACGAGATTATAACCAGAGTACGAATA...

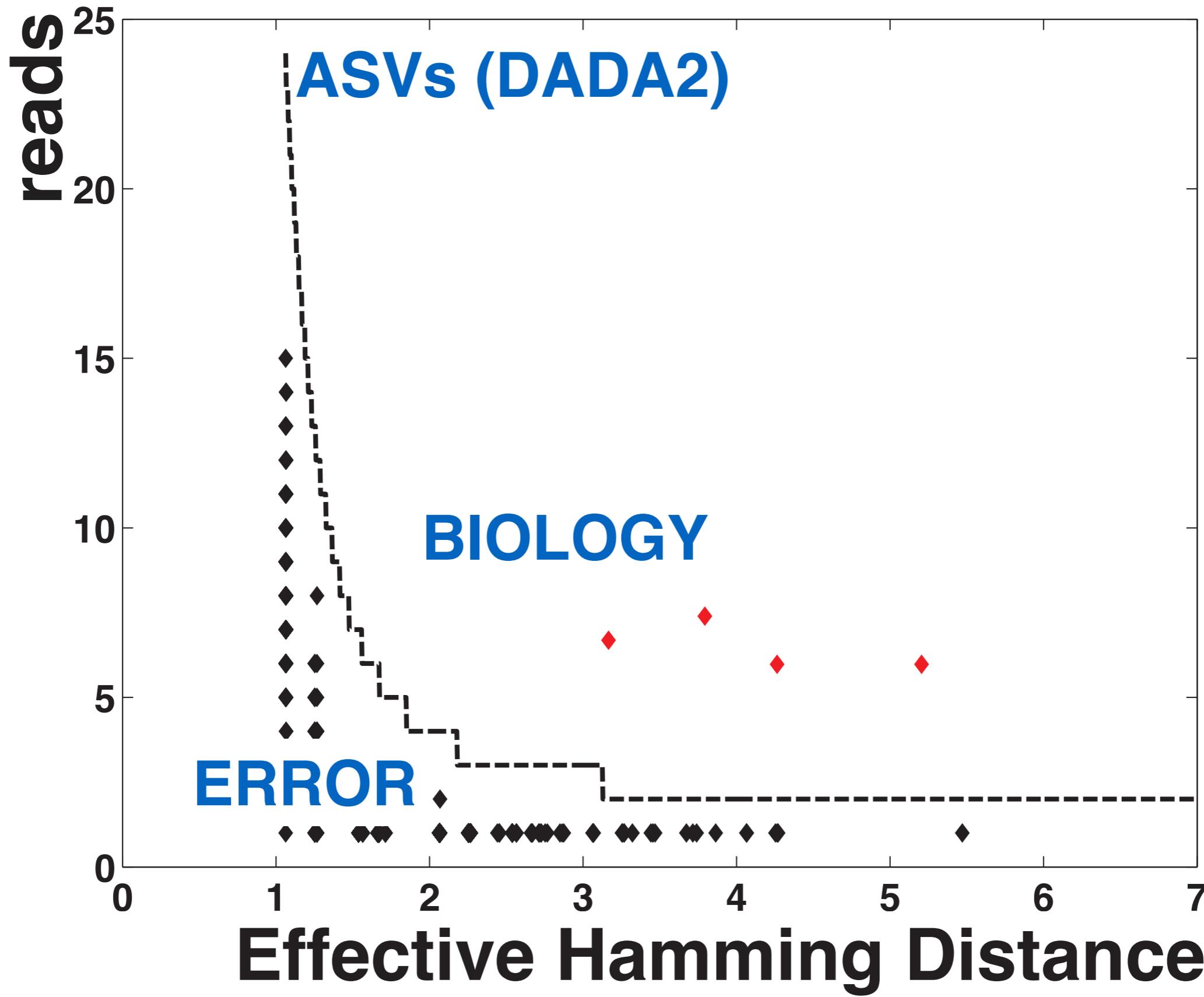
is consistent!



Signal from Noise: OTUs

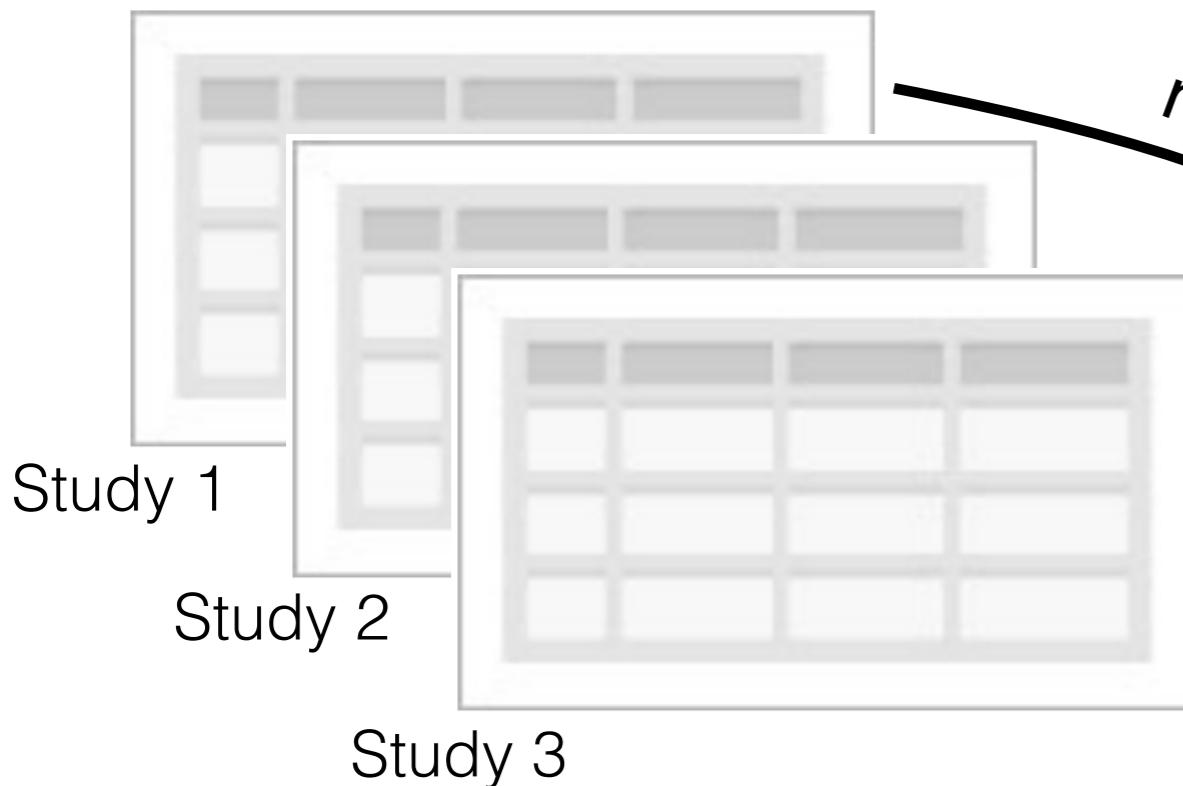


Signal from Noise: ASVs



The Sequence is the Label

Sequence Tables

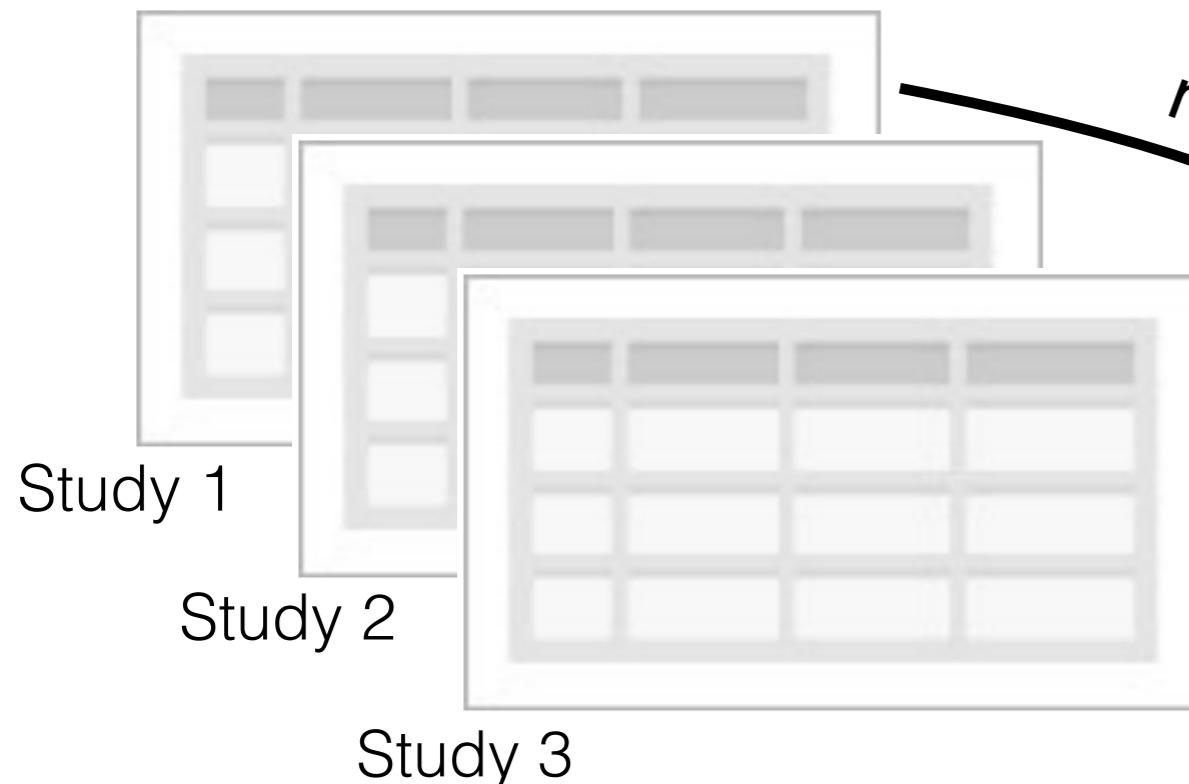


Cross-study comparison

Eliminates need for joint reprocessing of raw data.

The Sequence is the Label

Sequence Tables



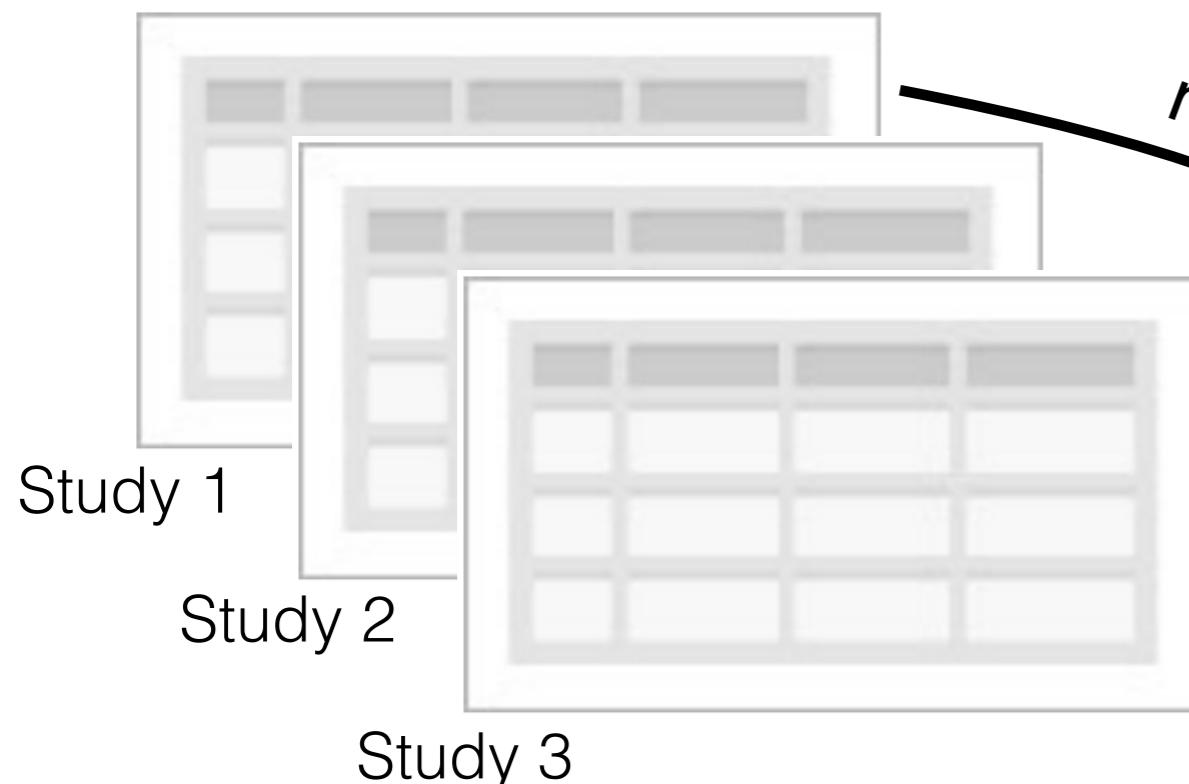
Cross-study comparison

Eliminates need for joint reprocessing of raw data.

Continuous data integration. Unlimited dataset size.

The Sequence is the Label

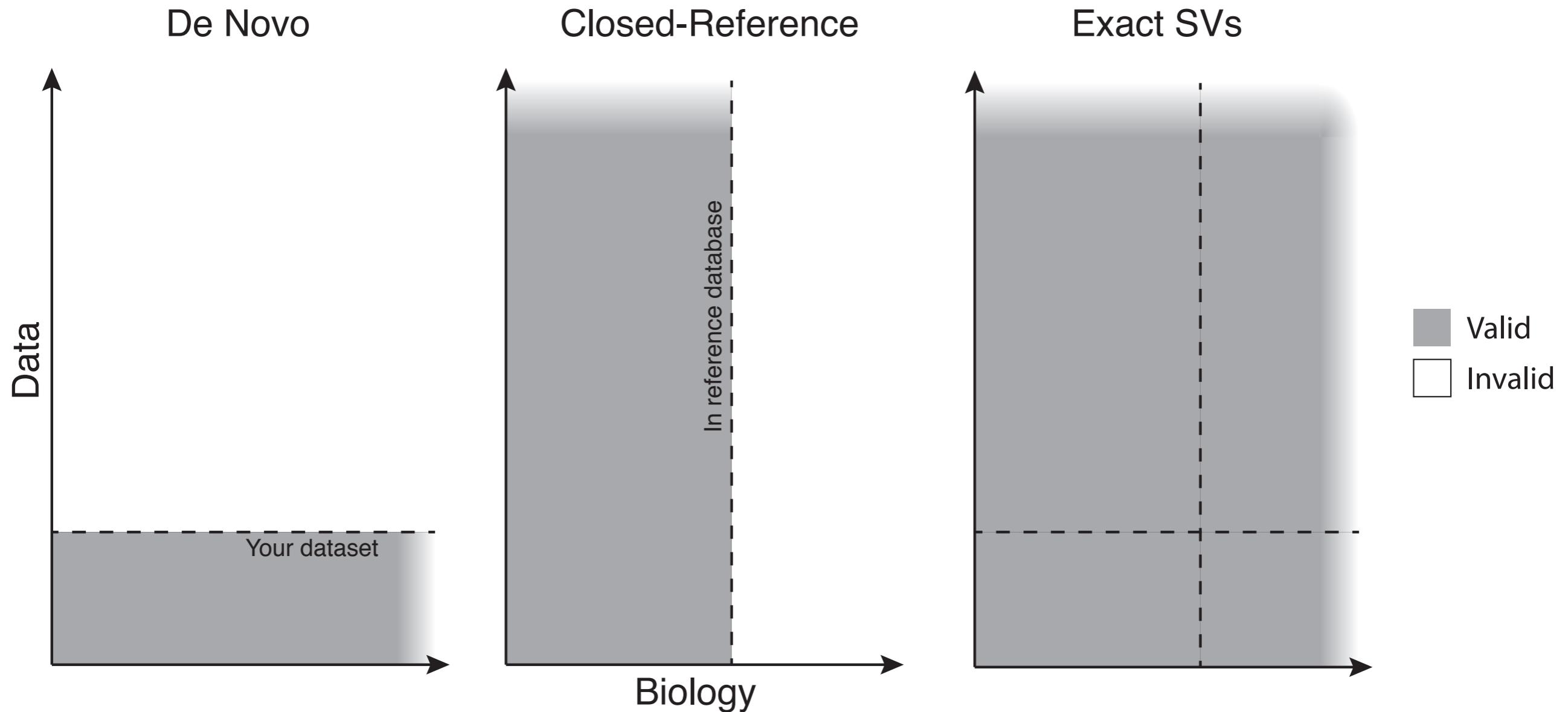
Sequence Tables



Cross-study comparison

Eliminates need for joint reprocessing of raw data.
Continuous data integration. Unlimited dataset size.
You in 2 years can work directly with you today.

The Sequence is the Label



“Replacing OTUs with ASVs makes marker-gene sequencing more precise, reusable, reproducible and comprehensive.”

ASV Methods and Terminology

ASV Methods

- DADA2
- Deblur
- Unoise2
- Oligotyping/MED
- AmpliCI
- ...

ASV Methods and Terminology

ASV Methods

- DADA2
- Deblur
- Unoise2
- Oligotyping/MED
- AmpliCI
- ...

Synonyms

- Amplicon Sequence Variant (ASV)
- Exact Sequence Variant (ESV)
- subOTU (sOTU)
- Zero-radius OTU (zOTU)
- Haplotypes, oligotypes, ...

ASVs vs. OTUs

	ASVs	De novo	Closed-ref
Precise	✓	~	~
Tractable	✓	~	✓
Reproducible	✓	✗	✓
Comprehensive	✓	✓	✗

ASVs vs. OTUs

	ASVs	De novo	Closed-ref
Precise	✓	~	~
Tractable	✓	~	✓
Reproducible	✓	✗	✓
Comprehensive	✓	✓	✗

Are OTUs *wrong*?

ASVs vs. OTUs

	ASVs	De novo	Closed-ref
Precise	✓	~	~
Tractable	✓	~	✓
Reproducible	✓	✗	✓
Comprehensive	✓	✓	✗

Are OTUs *wrong?* No.

ASVs vs. OTUs

	ASVs	De novo	Closed-ref
Precise	✓	~	~
Tractable	✓	~	✓
Reproducible	✓	✗	✓
Comprehensive	✓	✓	✗

Are OTUs *wrong*? No.

Are ASVs *always* the best unit of analysis?

ASVs vs. OTUs

	ASVs	De novo	Closed-ref
Precise	✓	~	~
Tractable	✓	~	✓
Reproducible	✓	✗	✓
Comprehensive	✓	✓	✗

Are OTUs *wrong*? No.

Are ASVs *always* the best unit of analysis? No.

Phylogenetic Scale

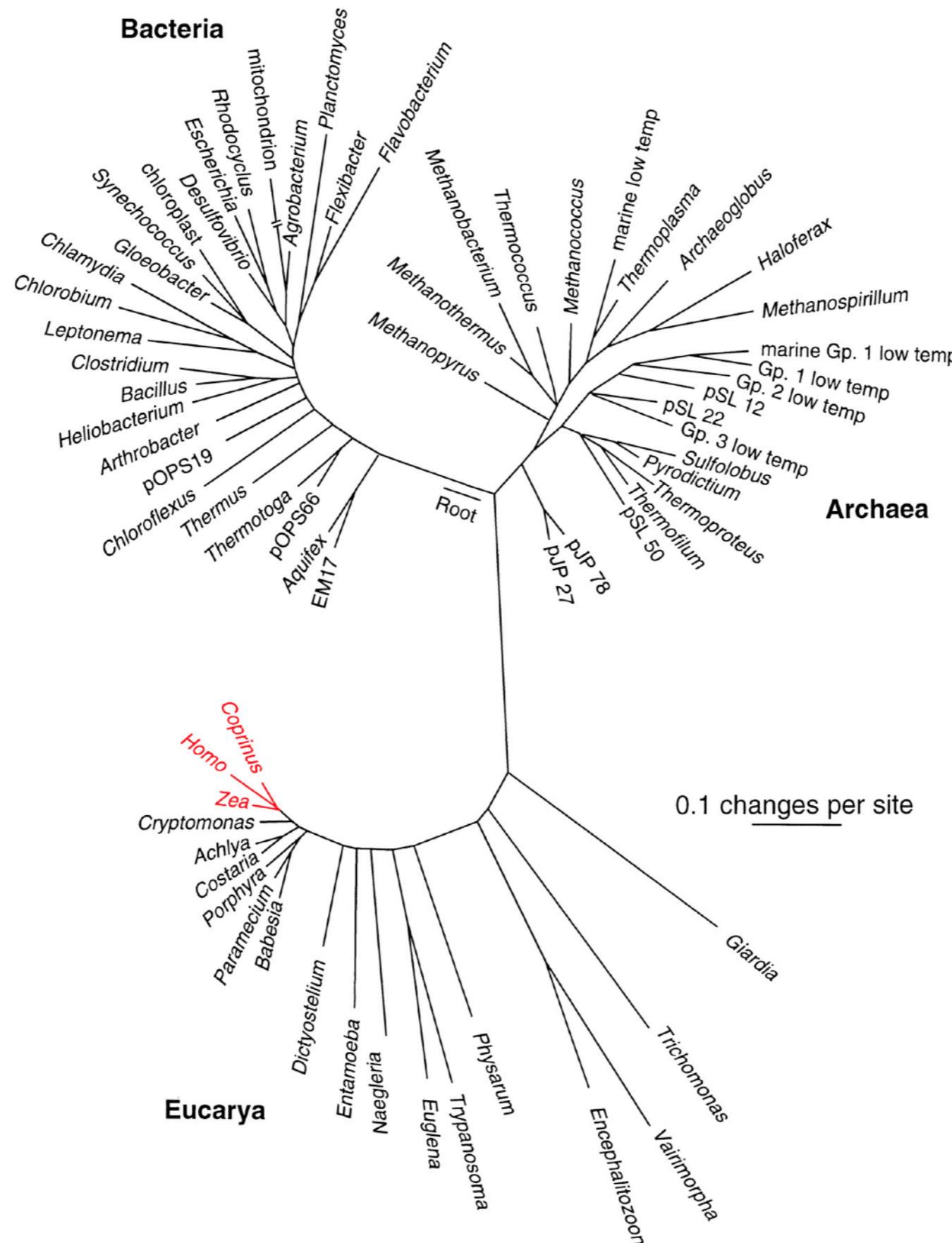


Image: Norman Pace

Phylogenetic Scale

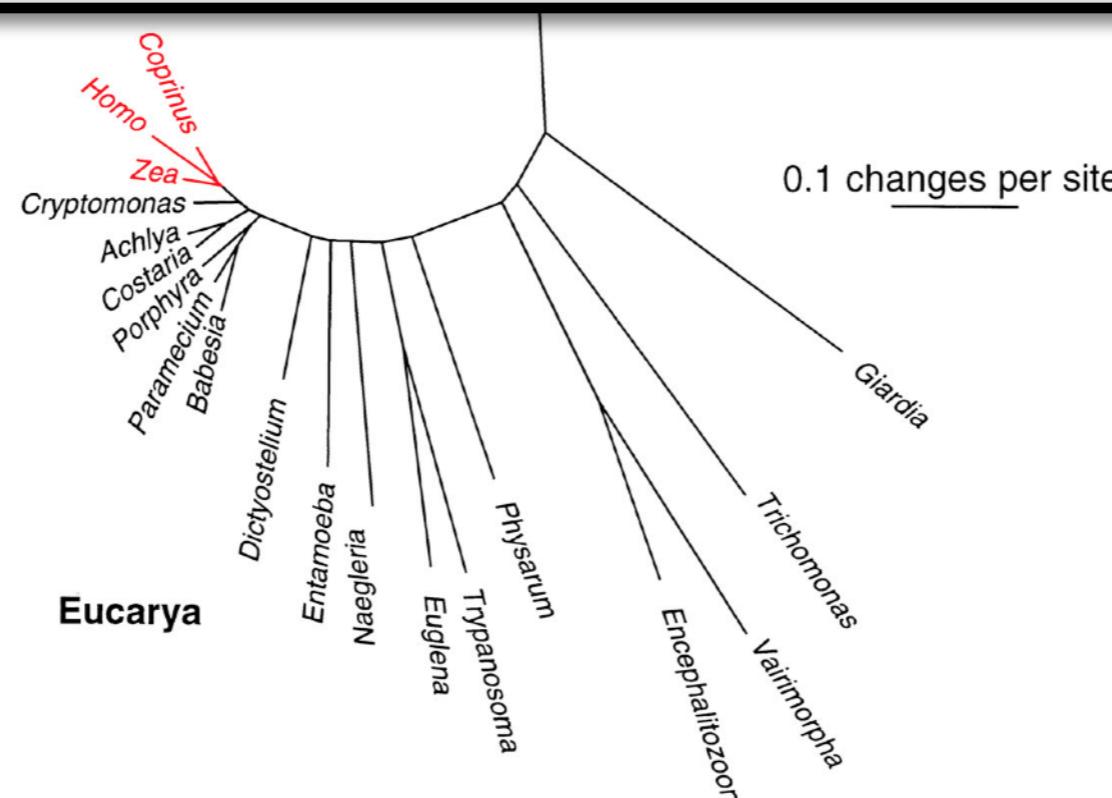
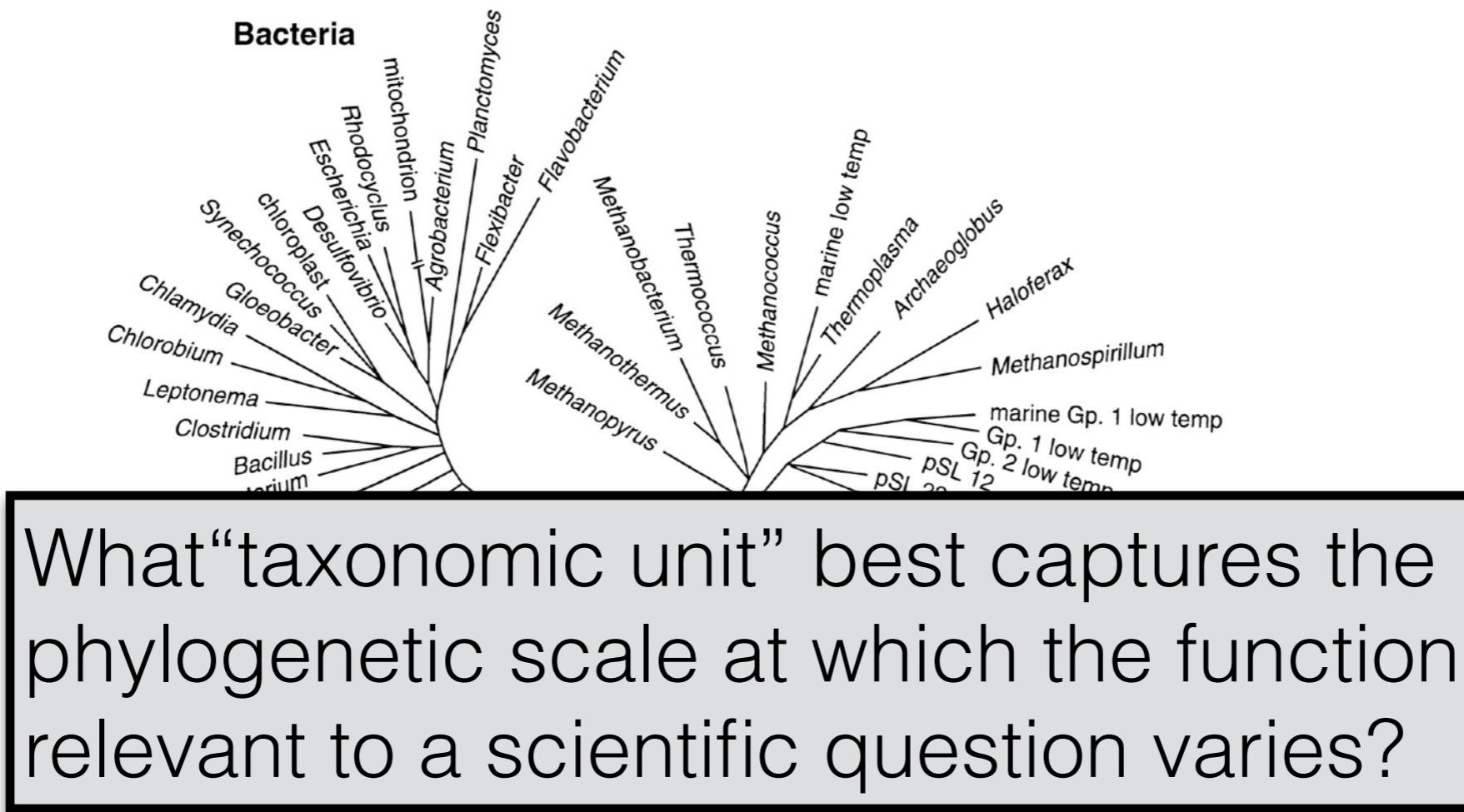
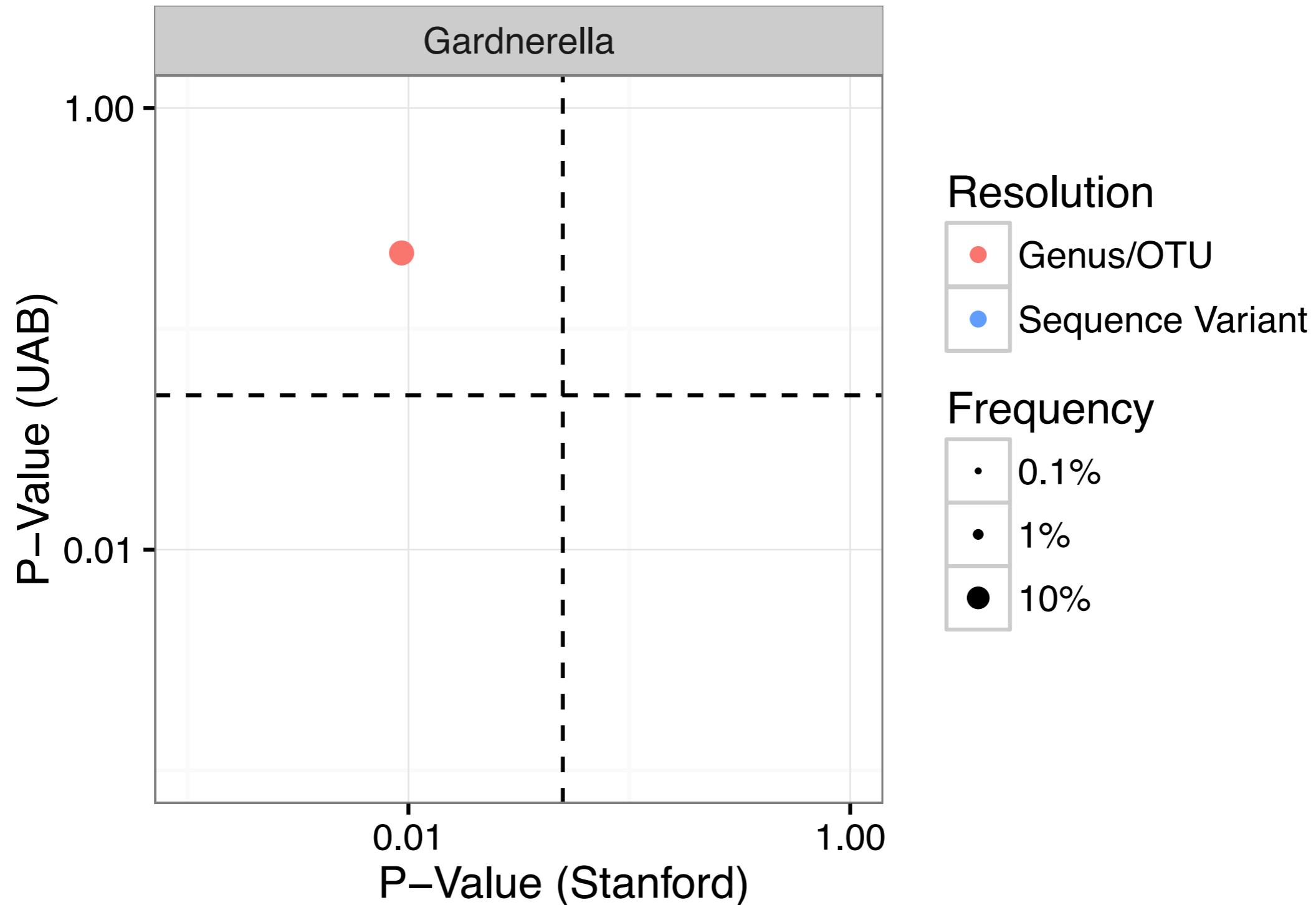
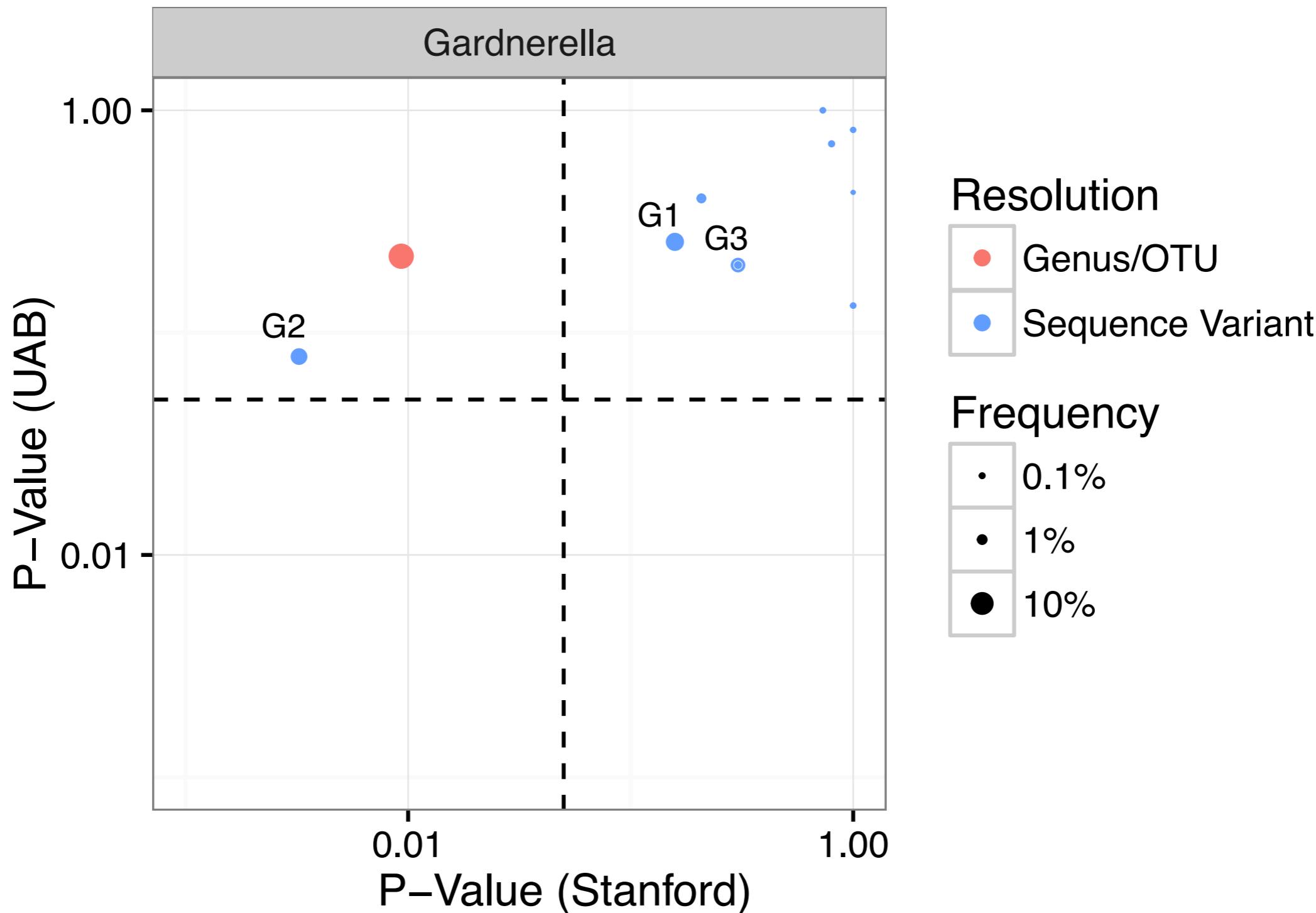


Image: Norman Pace

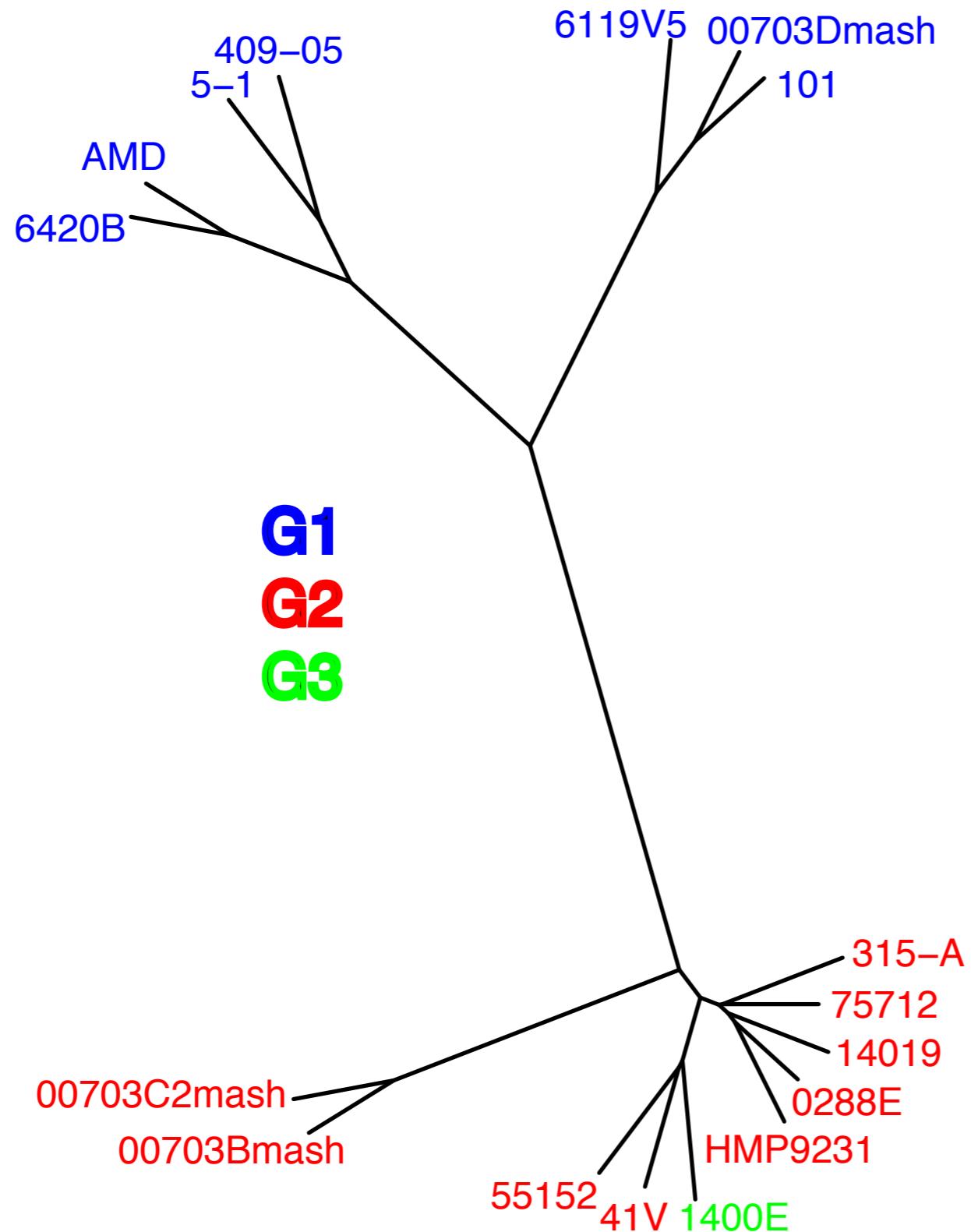
Resolution: Preterm Birth



Resolution: Preterm Birth

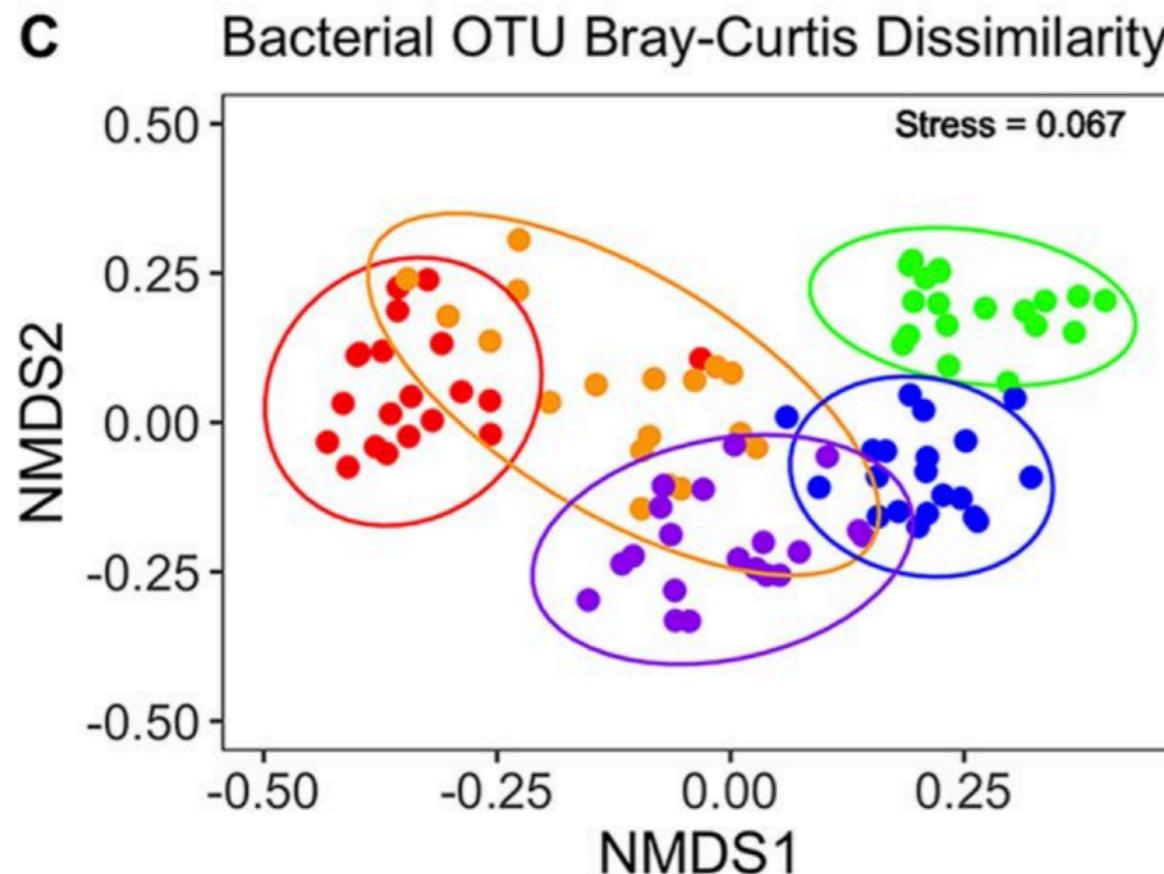


Resolution: Preterm Birth

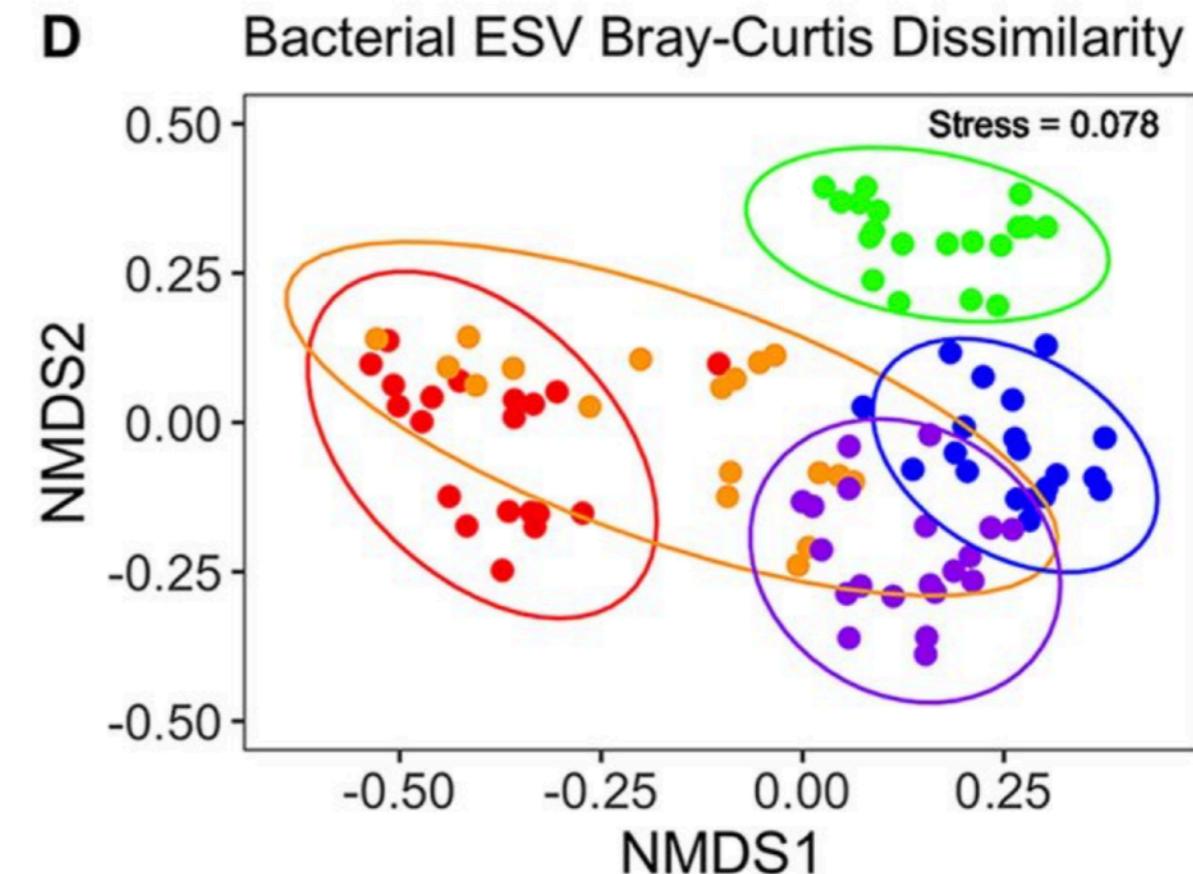
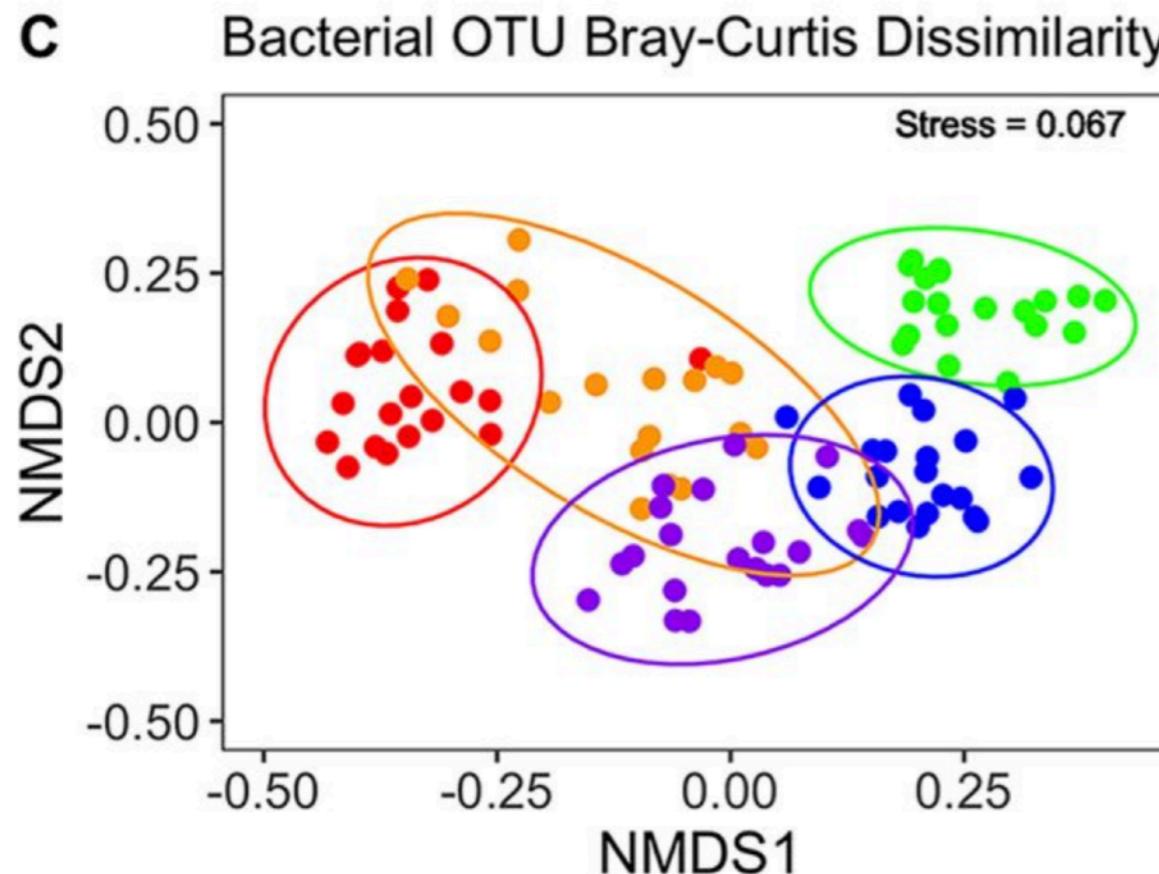


Tree from
whole genomes

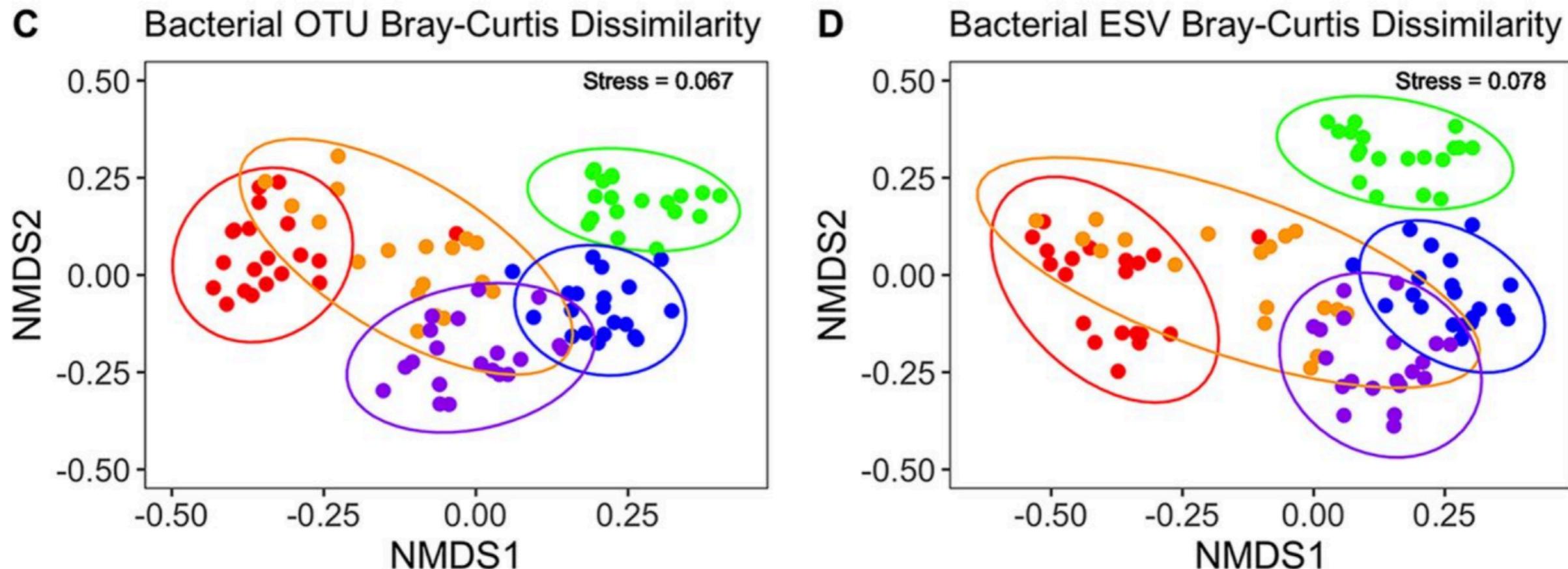
Resolution: Broadscale Patterns



Resolution: Broadscale Patterns



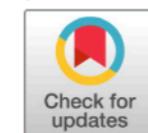
Resolution: Broadscale Patterns



AMERICAN
SOCIETY FOR
MICROBIOLOGY



OBSERVATION
Ecological and Evolutionary Science



Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units

Sydney I. Glassman,^{a,b} Jennifer B. H. Martiny^a

To species and beyond?

Short-read 16S ASVs

Typically will discriminate pairs of species in a genus.

Typically can't unambiguously identify a species.

To species and beyond?

Short-read 16S ASVs

Typically will discriminate pairs of species in a genus.

Typically can't unambiguously identify a species.

Long-read 16S/+ ASVs

Almost always discriminate pairs of species in a genus.

Typically can unambiguously identify a species.

To species and beyond?

Short-read 16S ASVs

Typically will discriminate pairs of species in a genus.
Typically can't unambiguously identify a species.

Long-read 16S/+ ASVs

Almost always discriminate pairs of species in a genus.
Typically can unambiguously identify a species.

Other phylogenetic marker ASVs

It depends! Some markers can go below species.
Often a resolution/taxonomic-breadth tradeoff.

To species and beyond?

Short-read 16S ASVs

Typically will discriminate pairs of species in a genus.
Typically can't unambiguously identify a species.

Long-read 16S/+ ASVs

Almost always discriminate pairs of species in a genus.
Typically can unambiguously identify a species.

Other phylogenetic marker ASVs

It depends! Some markers can go below species.
Often a resolution/taxonomic-breadth tradeoff.

* All of this assumes a comprehensive reference database.

This speaker recommends...

- **You should start with ASVs**
- Consider analysis at ***multiple*** taxonomic levels
- **Reproducible**: ASVs, taxonomy. **Not**: *de novo* OTUs.
- **Comprehensive**: ASVs, *de novo* OTUs. **Not**: taxonomy.
- Use domain knowledge on relevant phylogenetic scale
- Deposit and share your ASV table

This speaker recommends...

- **You should start with ASVs**
- Consider analysis at ***multiple*** taxonomic levels
- **Reproducible**: ASVs, taxonomy. **Not**: *de novo* OTUs.
- **Comprehensive**: ASVs, *de novo* OTUs. **Not**: taxonomy.
- Use domain knowledge on relevant phylogenetic scale
- Deposit and share your ASV table

Can agglomerate up, but can't divide down.

Acknowledgements

Amplicon Sequence Variants



...for ITS

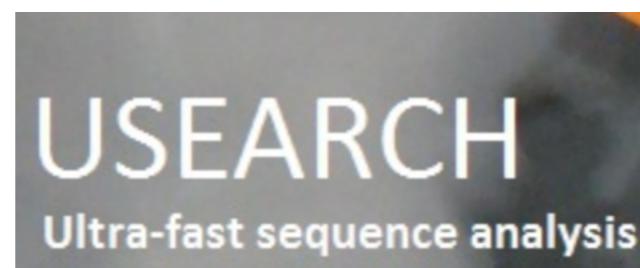
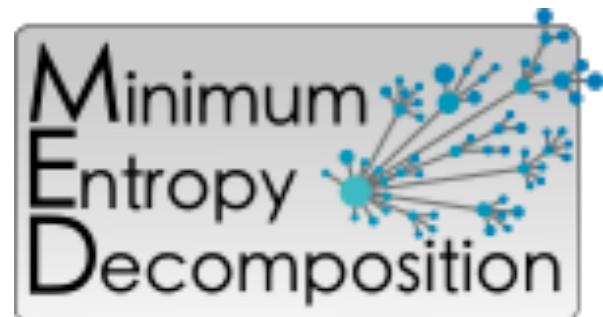


Susan Holmes

Joey McMurdie

Michael Rosen

Naga Betrapally



Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns

