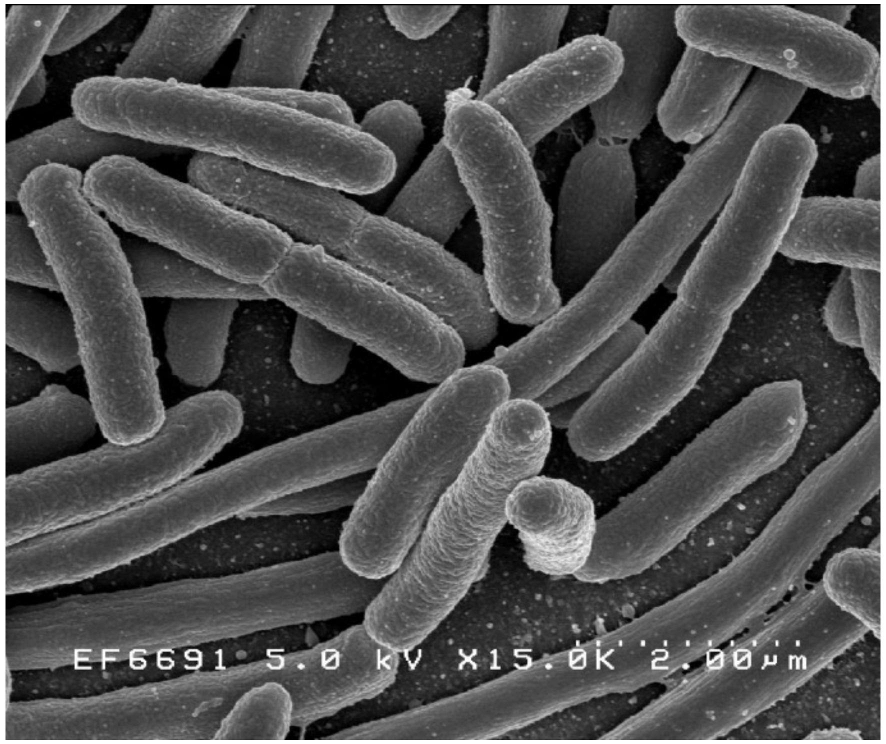
Amplicon Sequencing

OTUs and ASVs

STAMPS 2022

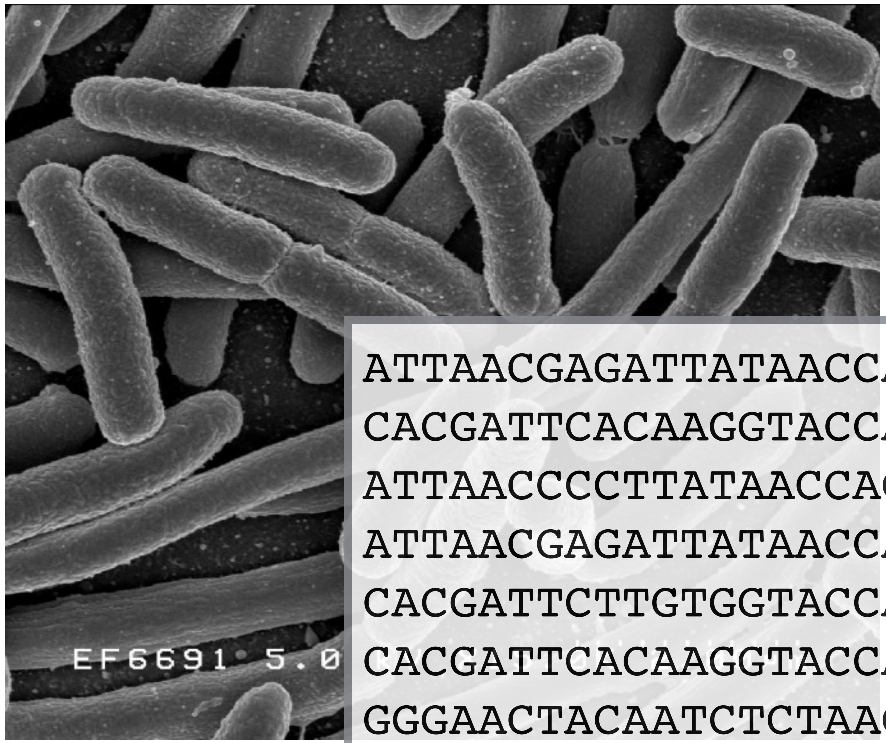
A Microbial Census

Metabarcoding or Marker-gene Sequencing



A Microbial Census

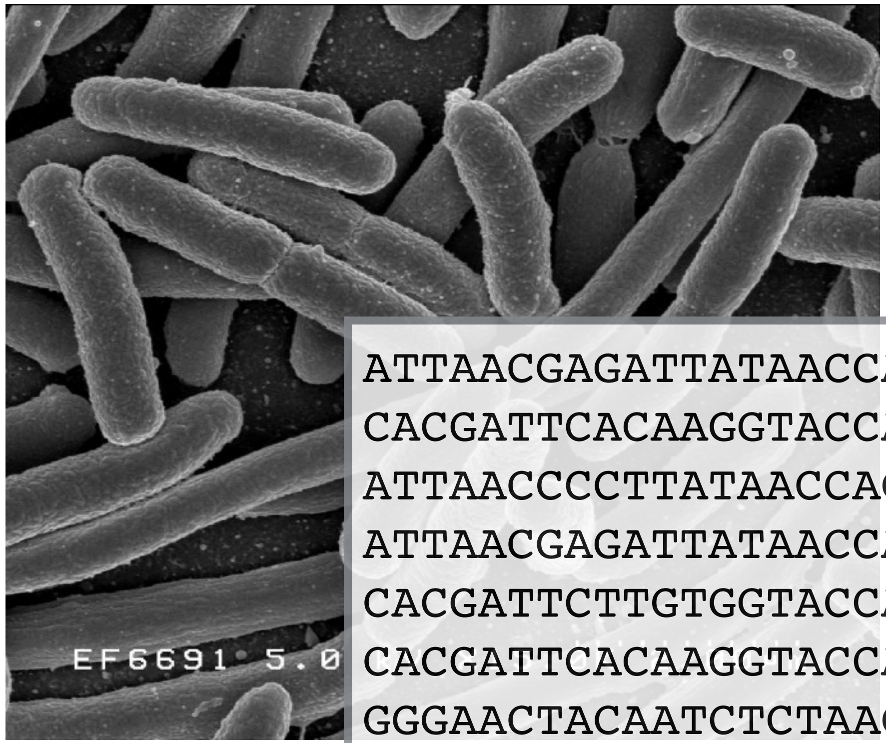
Metabarcoding or Marker-gene Sequencing



ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGAACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC

A Microbial Census

Metabarcoding or Marker-gene Sequencing

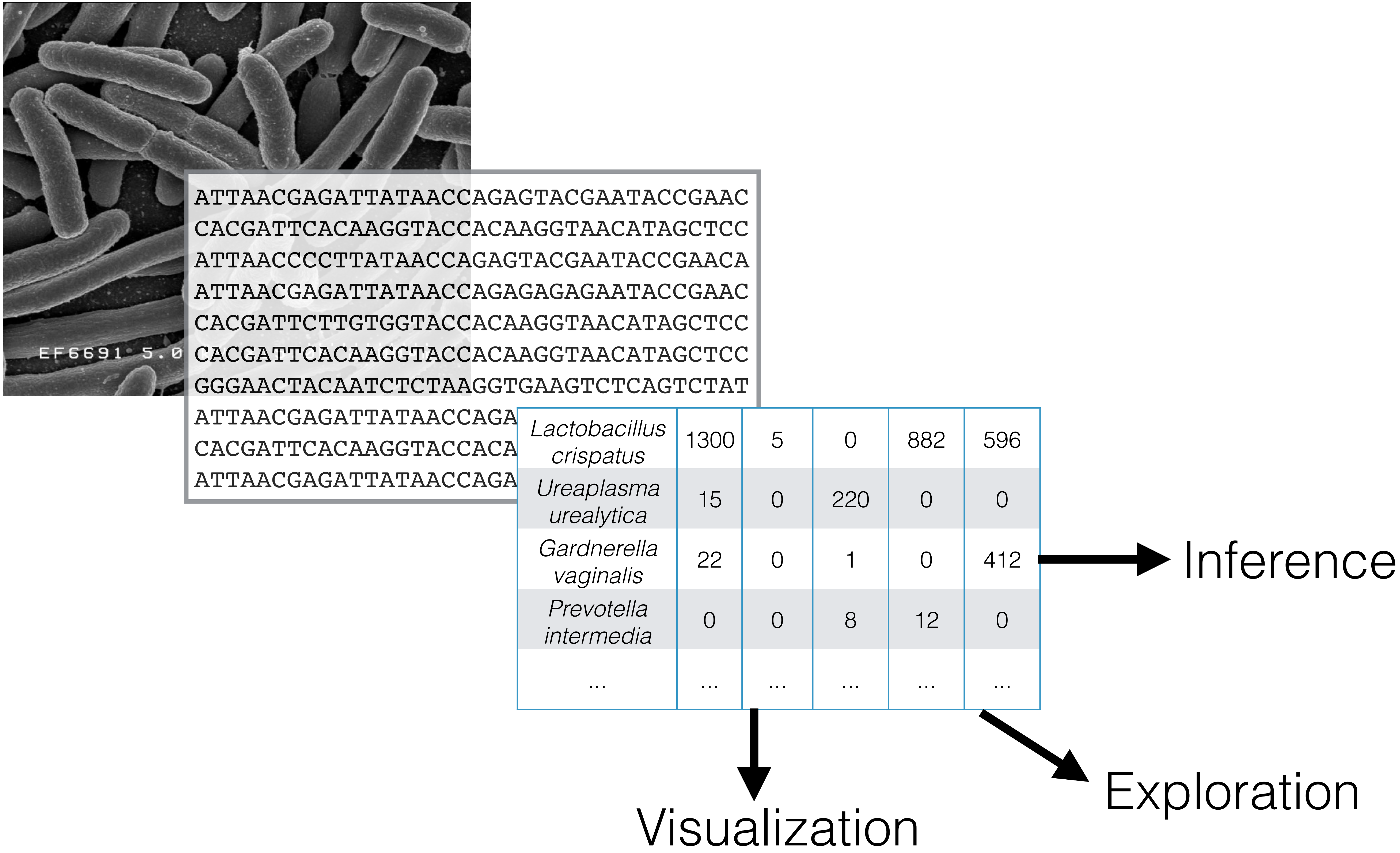


ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGAACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGA
CACGATTCACAAGGTACCACA
ATTAACGAGATTATAACCAGA

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

A Microbial Census

Metabarcoding or Marker-gene Sequencing



```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGAACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGA
CACGATTCACAAGGTACCACA
ATTAACGAGATTATAACCAGA
```

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

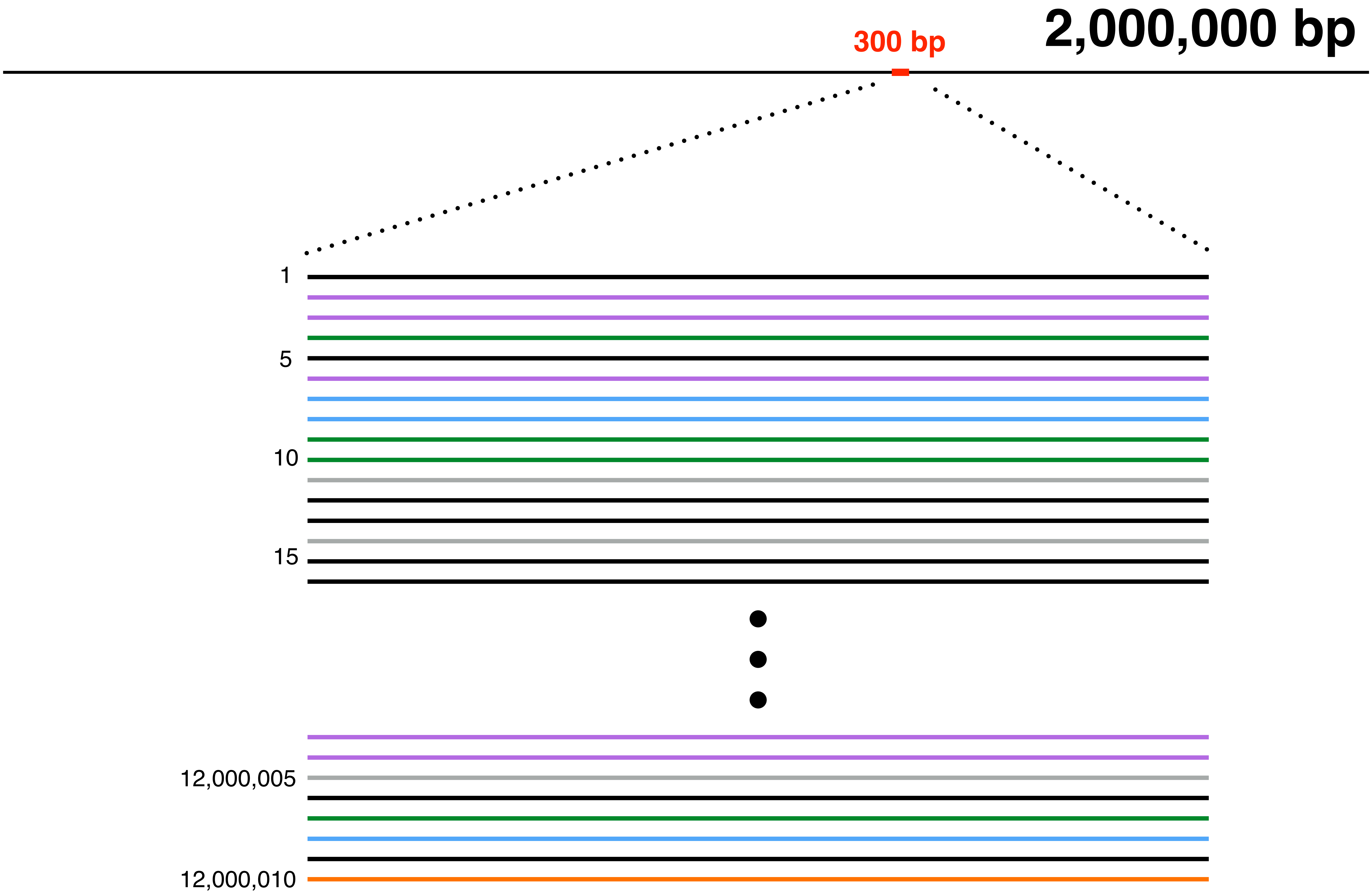
Inference

Visualization

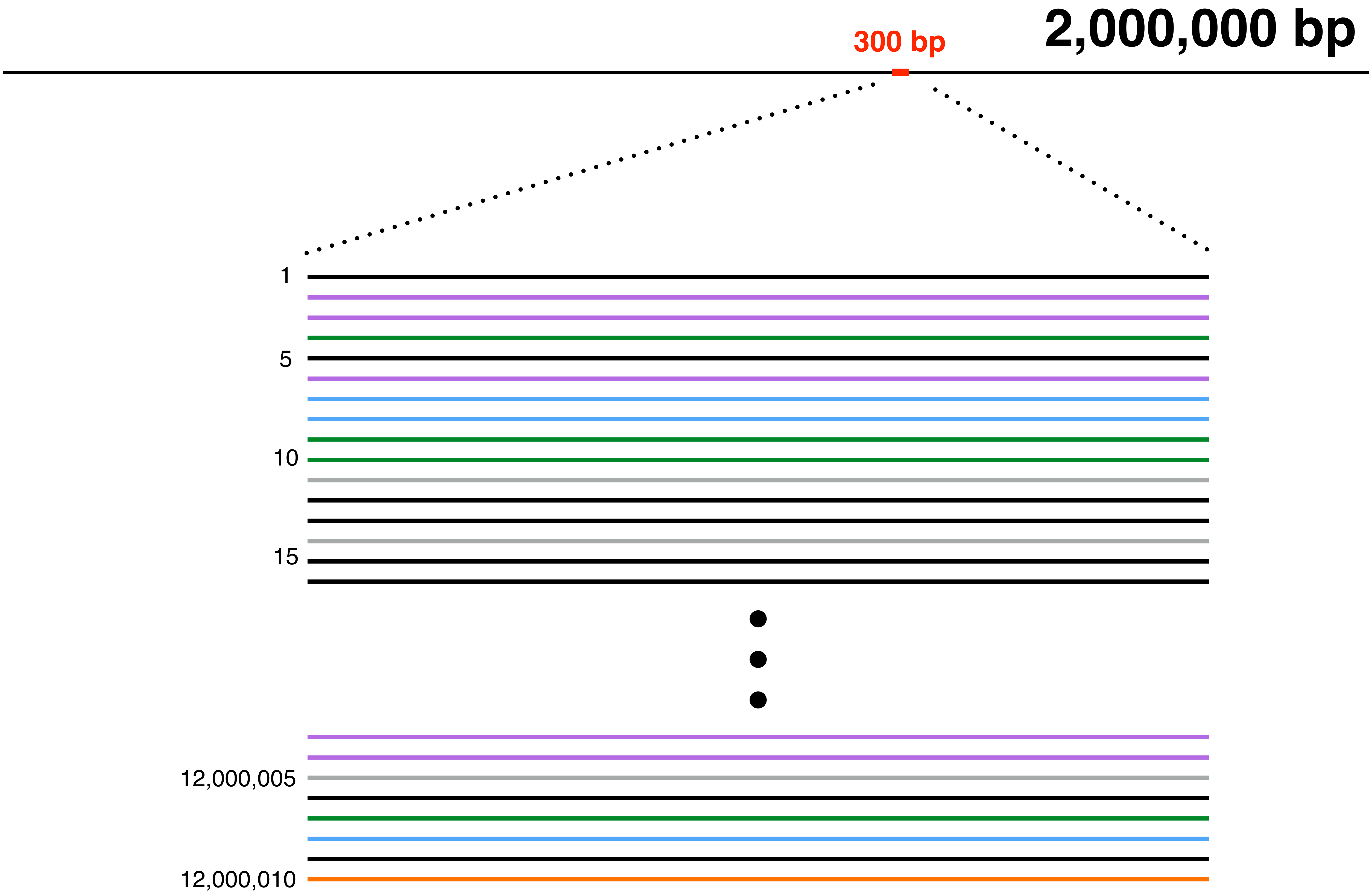
Exploration

Marker-gene Sequencing

Marker-gene Sequencing



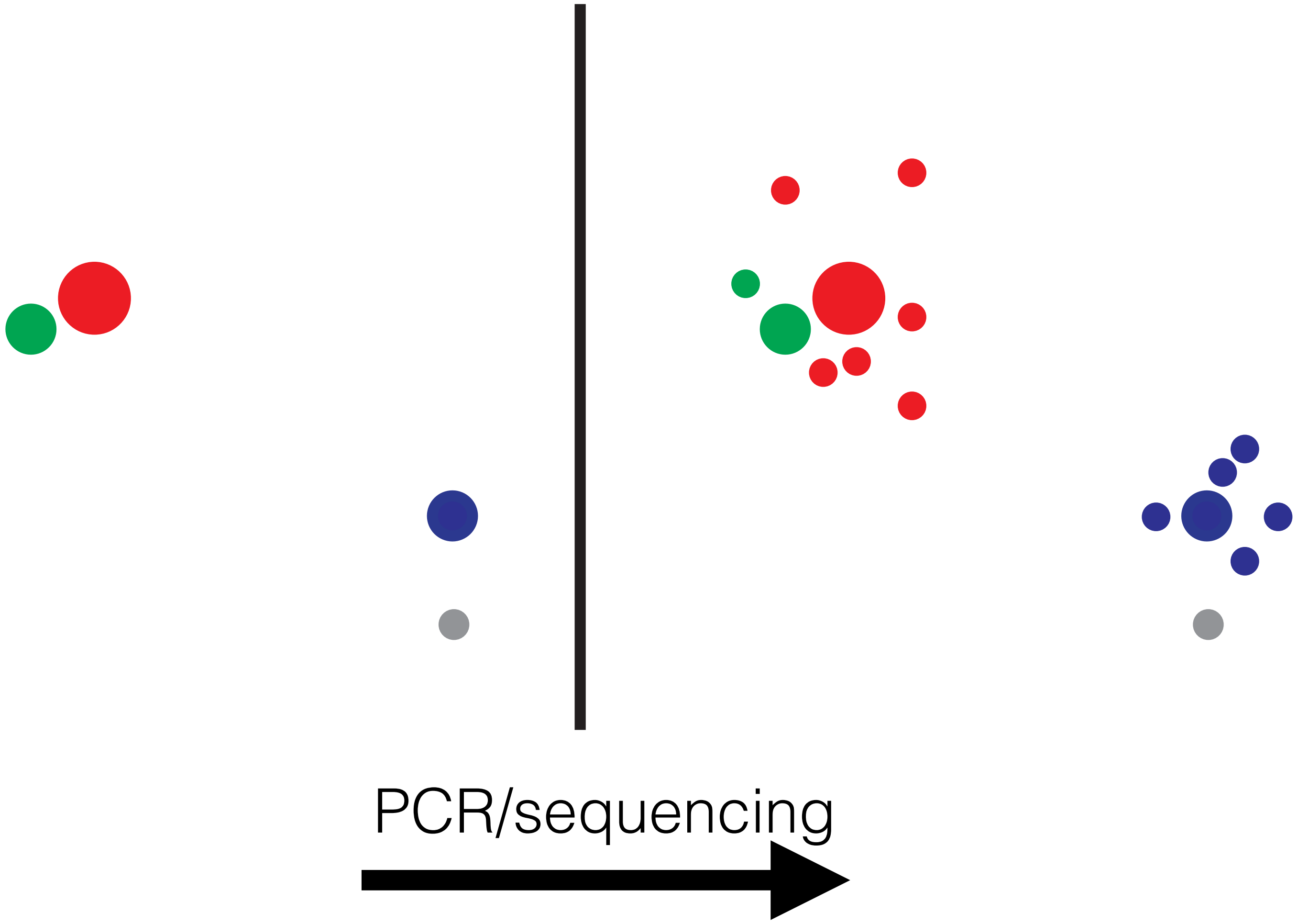
Marker-gene Sequencing



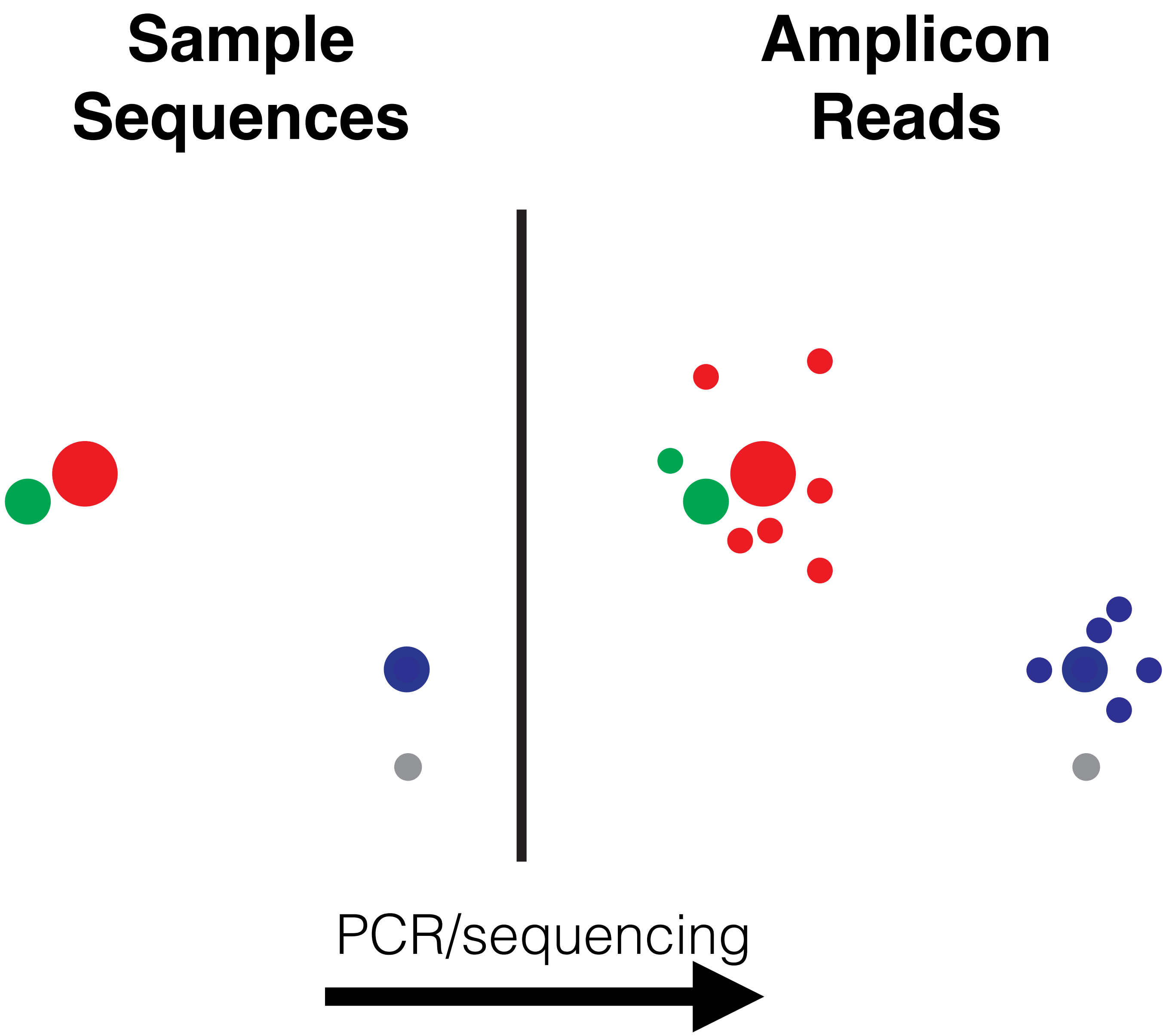
Thousands to Billions of “Barcodes”

Sample Sequences

Amplicon Reads



PCR/sequencing



Sample Sequences

Amplicon Reads

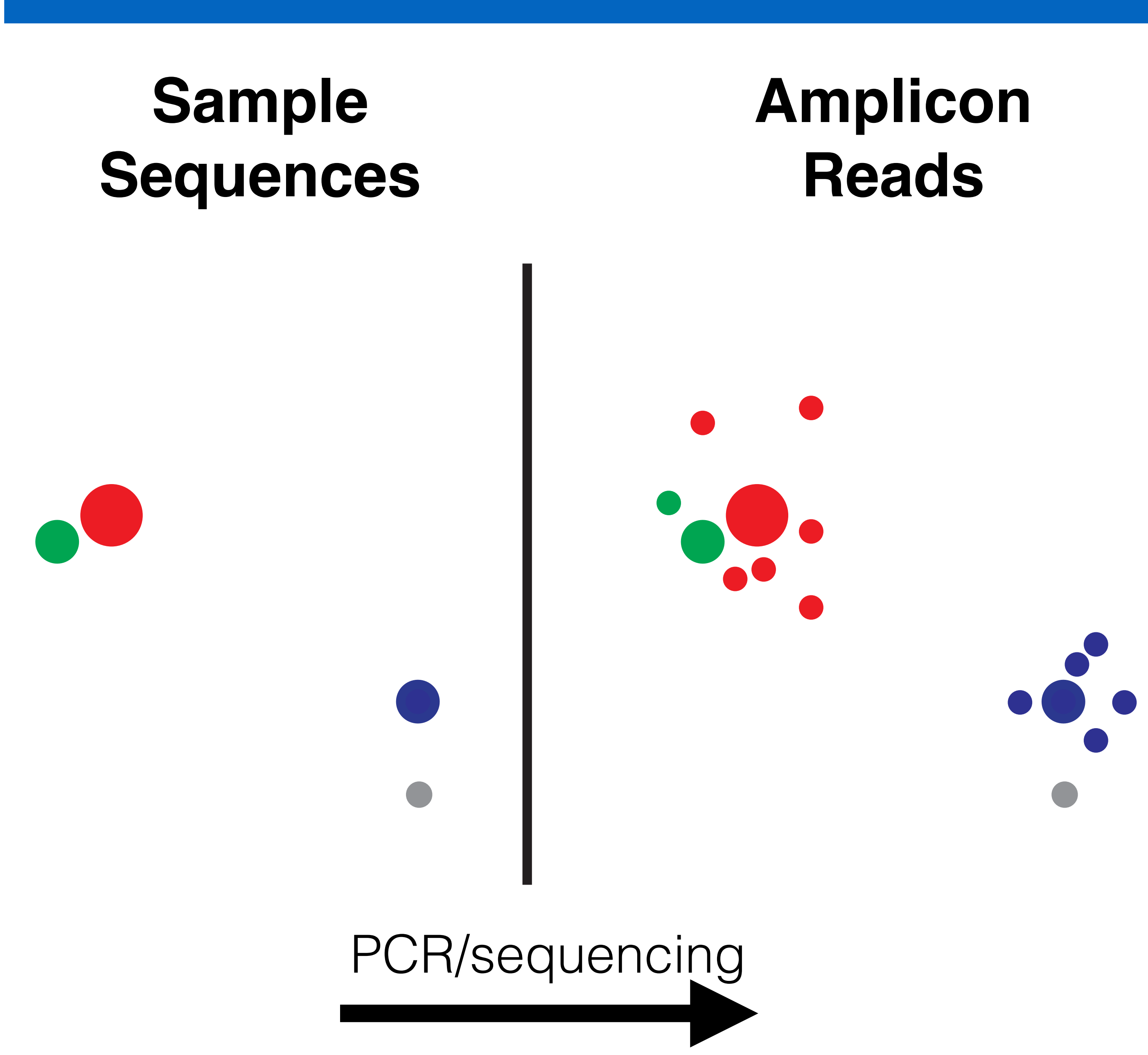
PCR/sequencing

Two Challenges

How do we deal with errors?

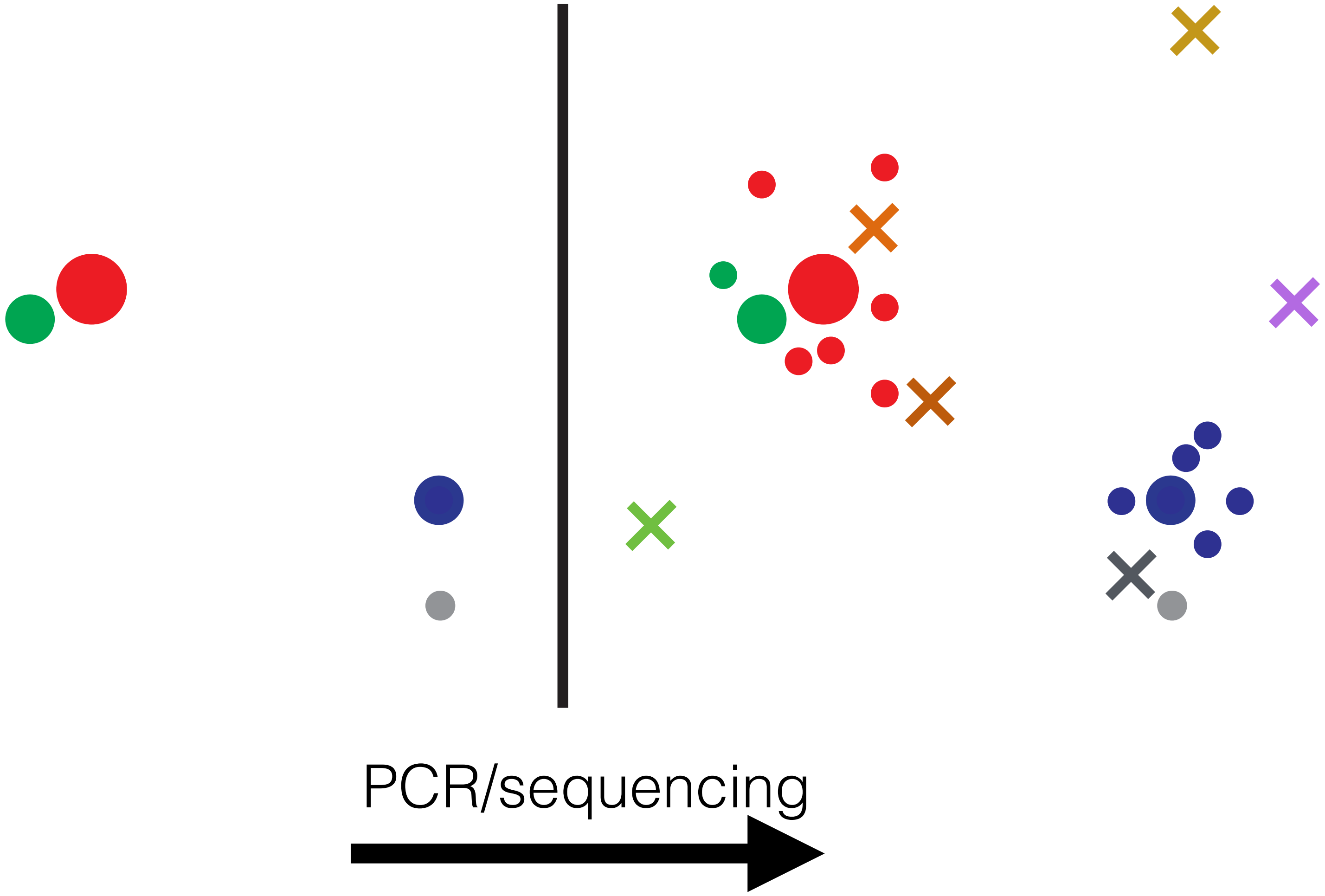
How do we define the units of our analysis?

Closed reference OTUs

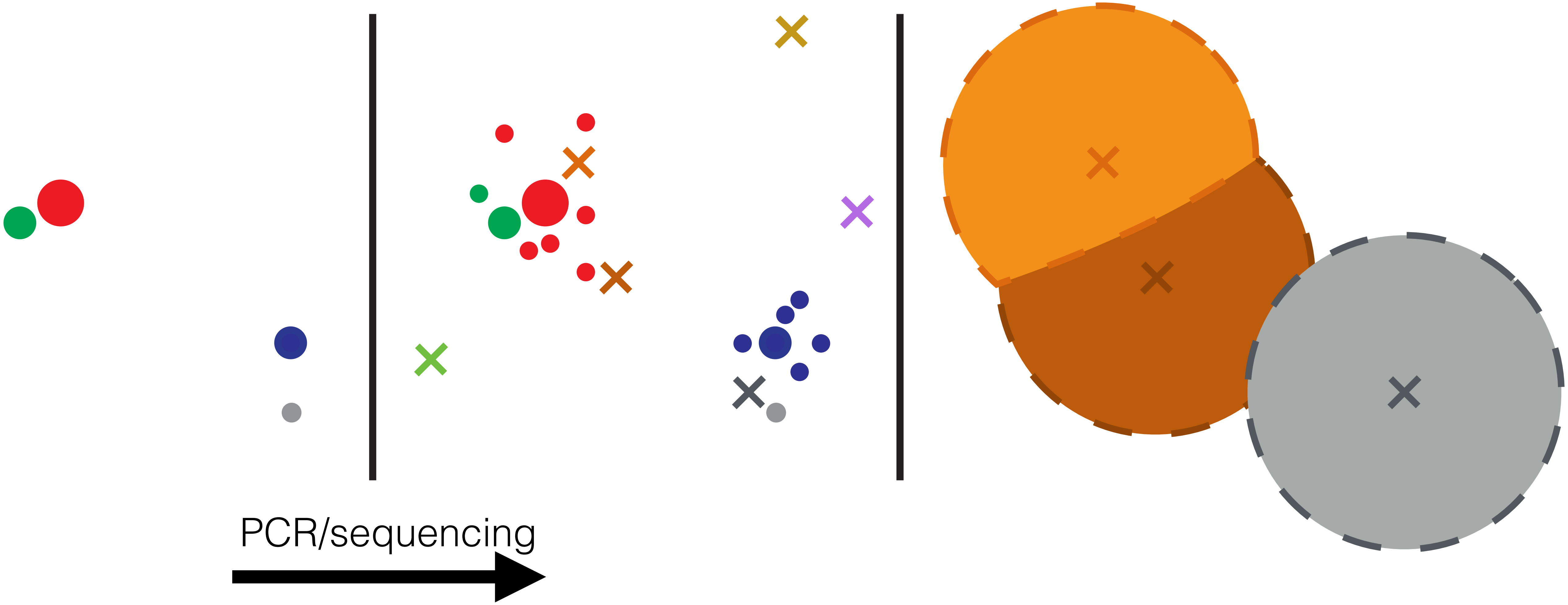


Sample Sequences

Amplicon Reads



PCR/sequencing



Sample Sequences

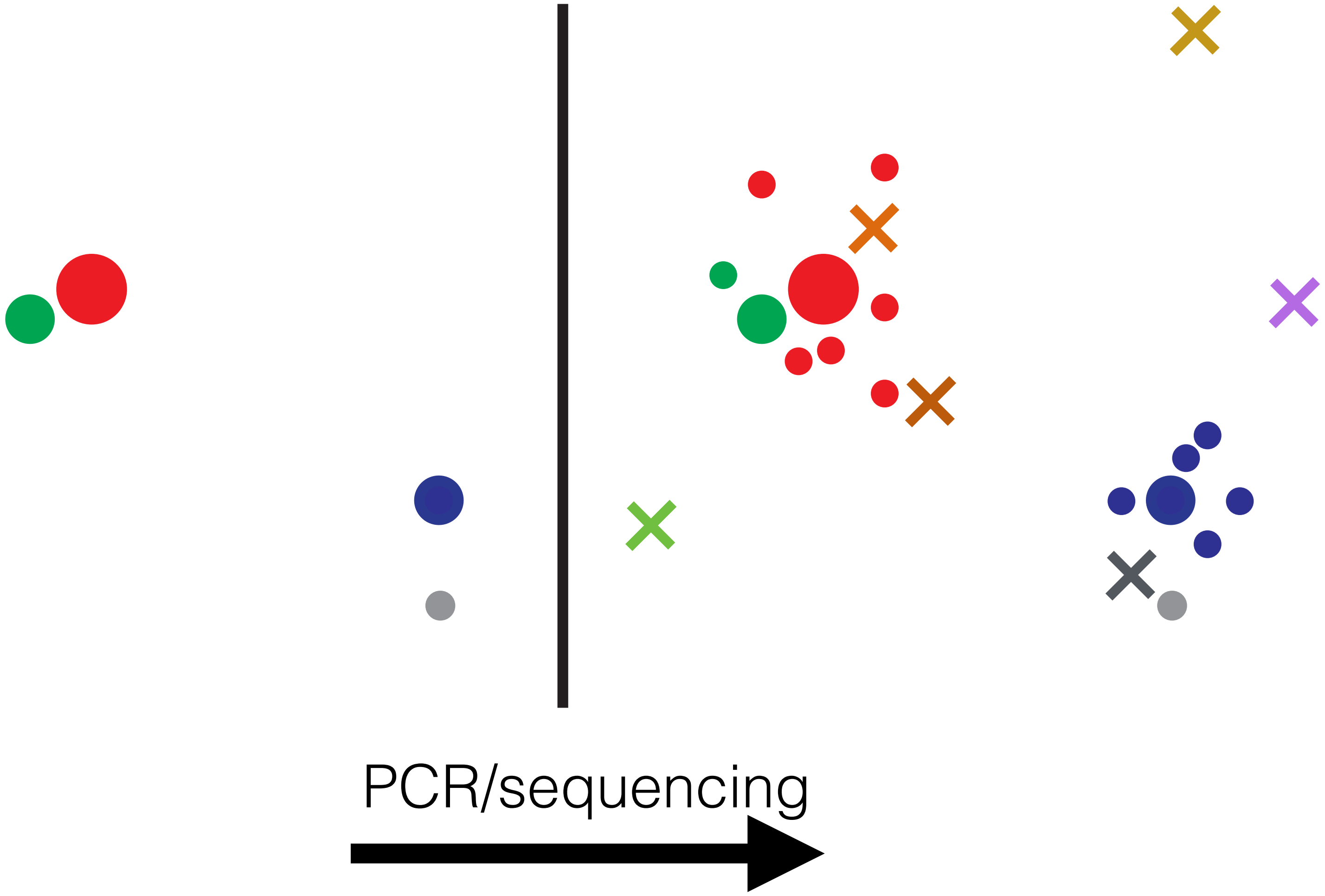
Amplicon Reads

Operational Taxonomic Units (closed reference)

PCR/sequencing

Sample Sequences

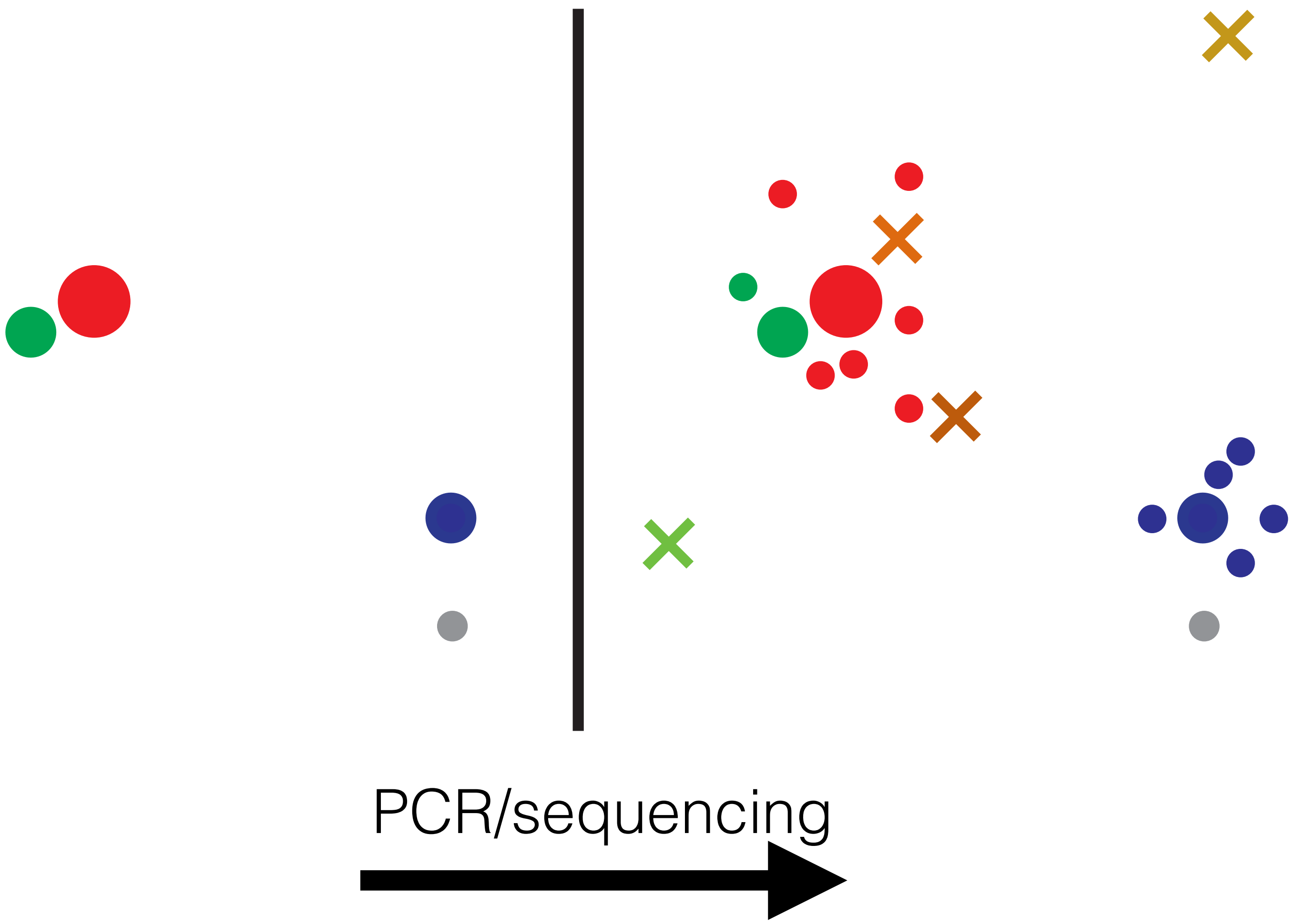
Amplicon Reads



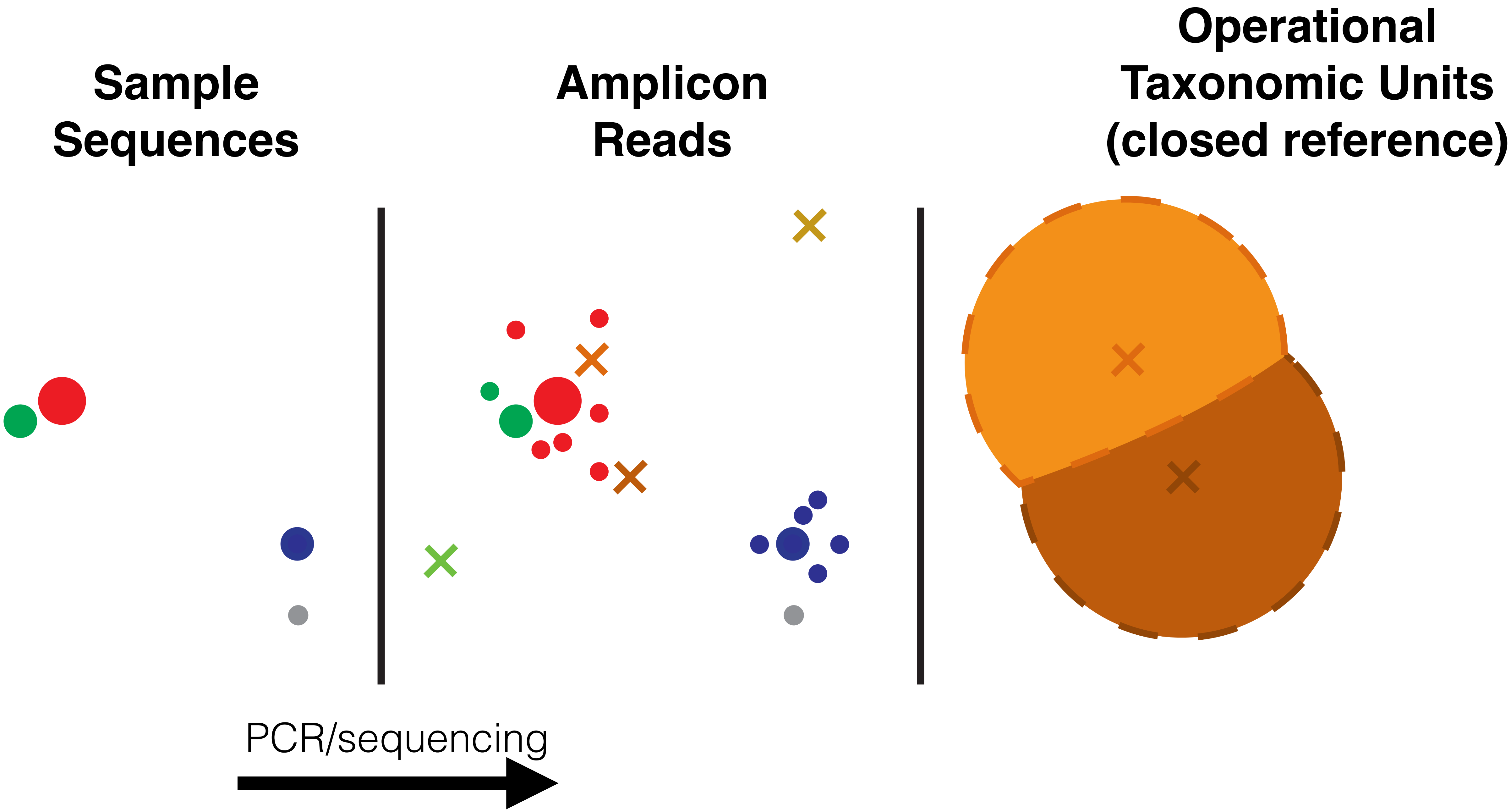
PCR/sequencing

Sample Sequences

Amplicon Reads



PCR/sequencing



Sample Sequences

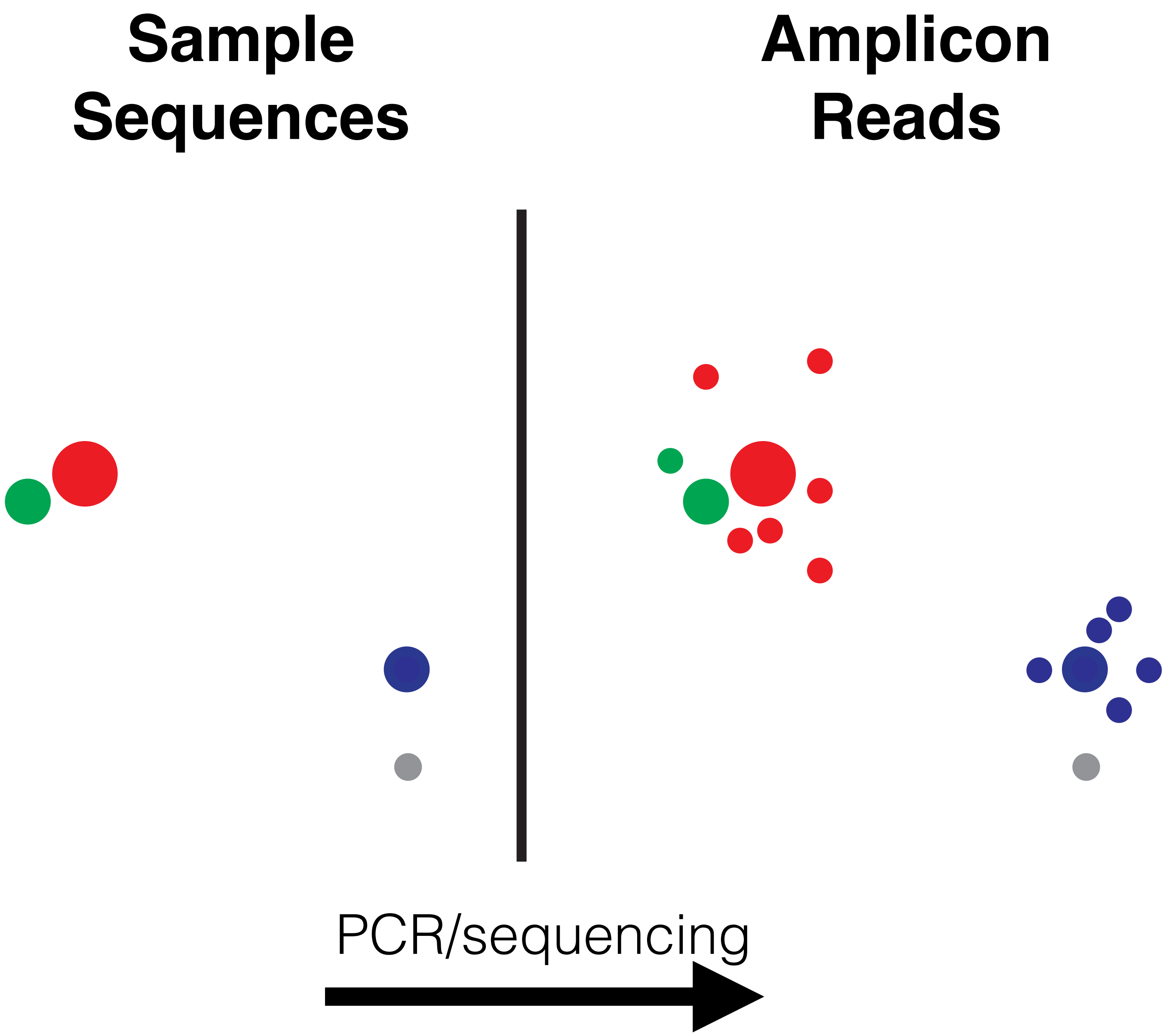
Amplicon Reads

Operational Taxonomic Units (closed reference)

PCR/sequencing

Incomplete OTU table!

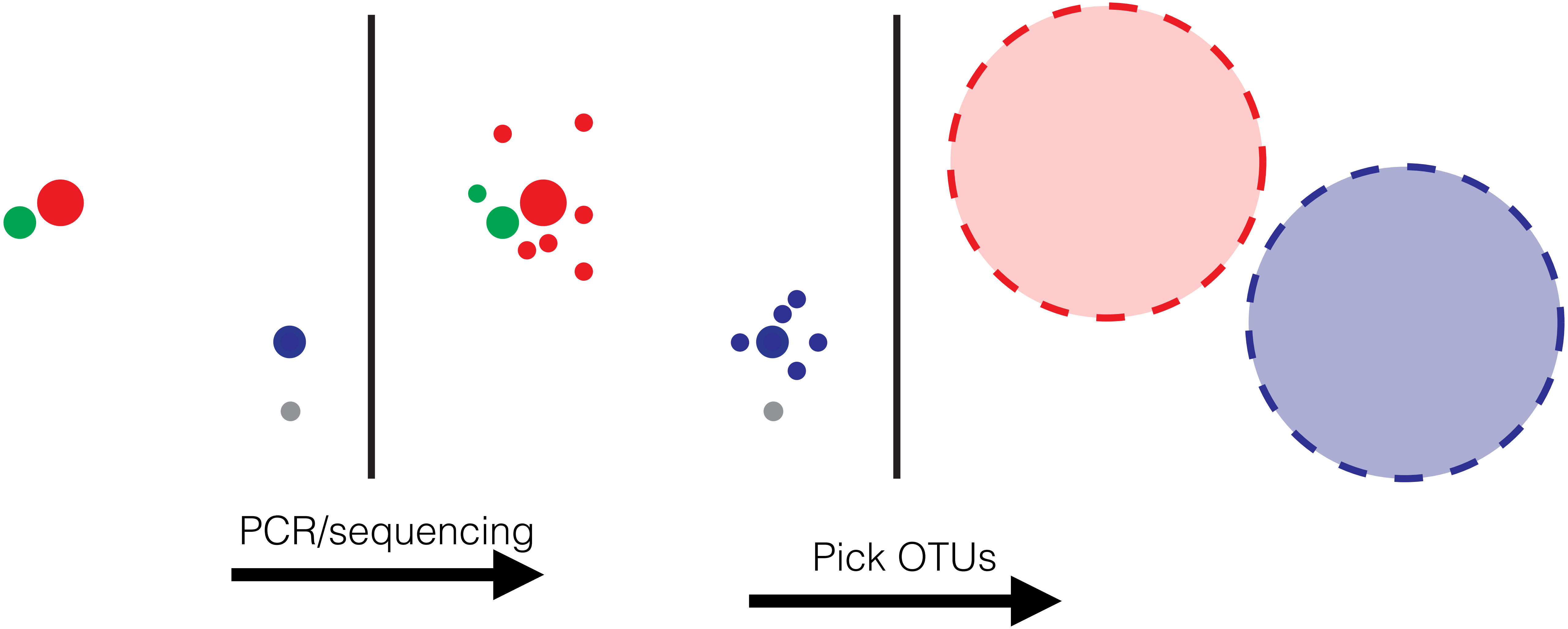
de novo OTUs



Sample Sequences

Amplicon Reads

PCR/sequencing



Sample Sequences

Amplicon Reads

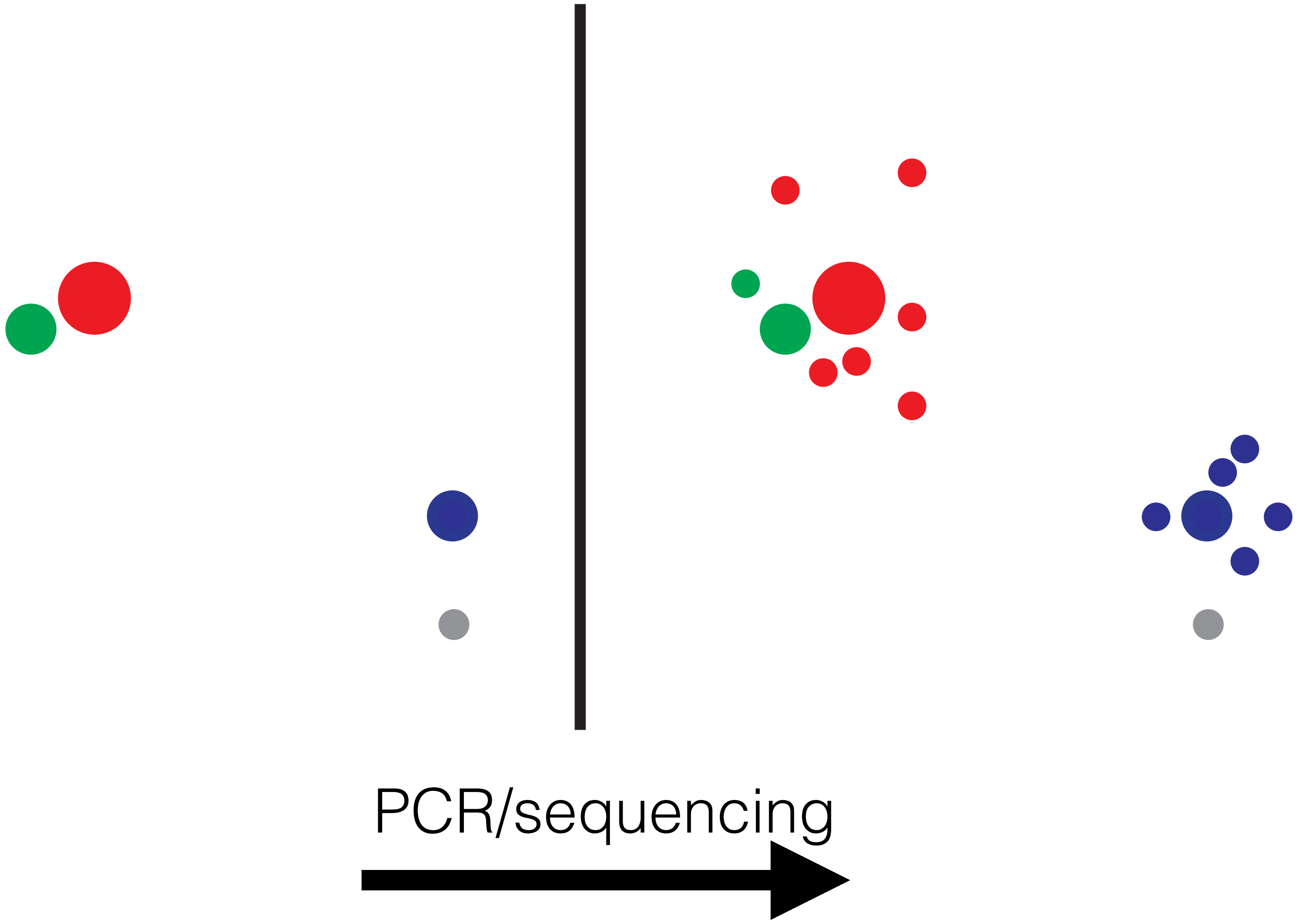
Operational Taxonomic Units (de novo)

PCR/sequencing

Pick OTUs

Sample Sequences

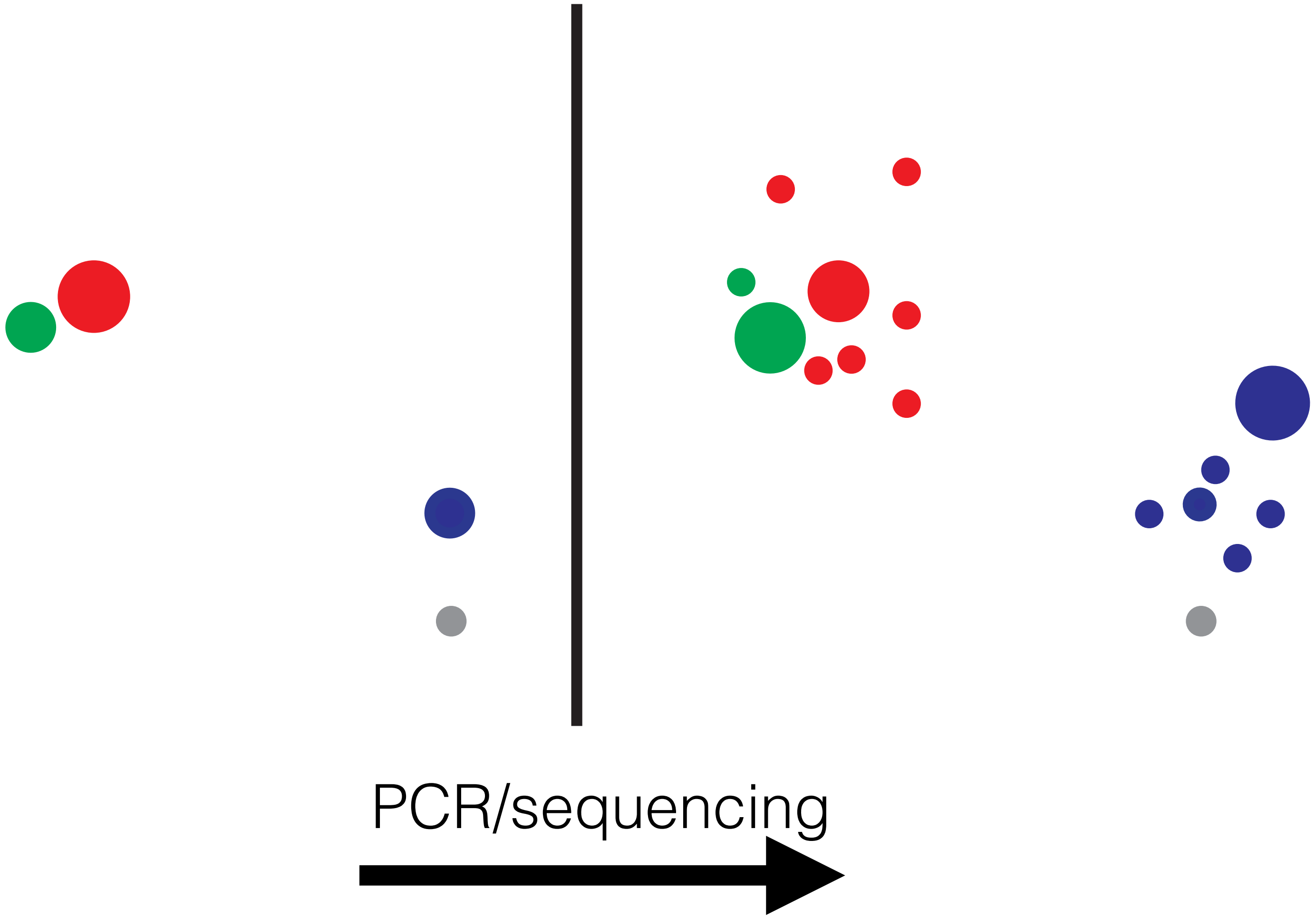
Amplicon Reads



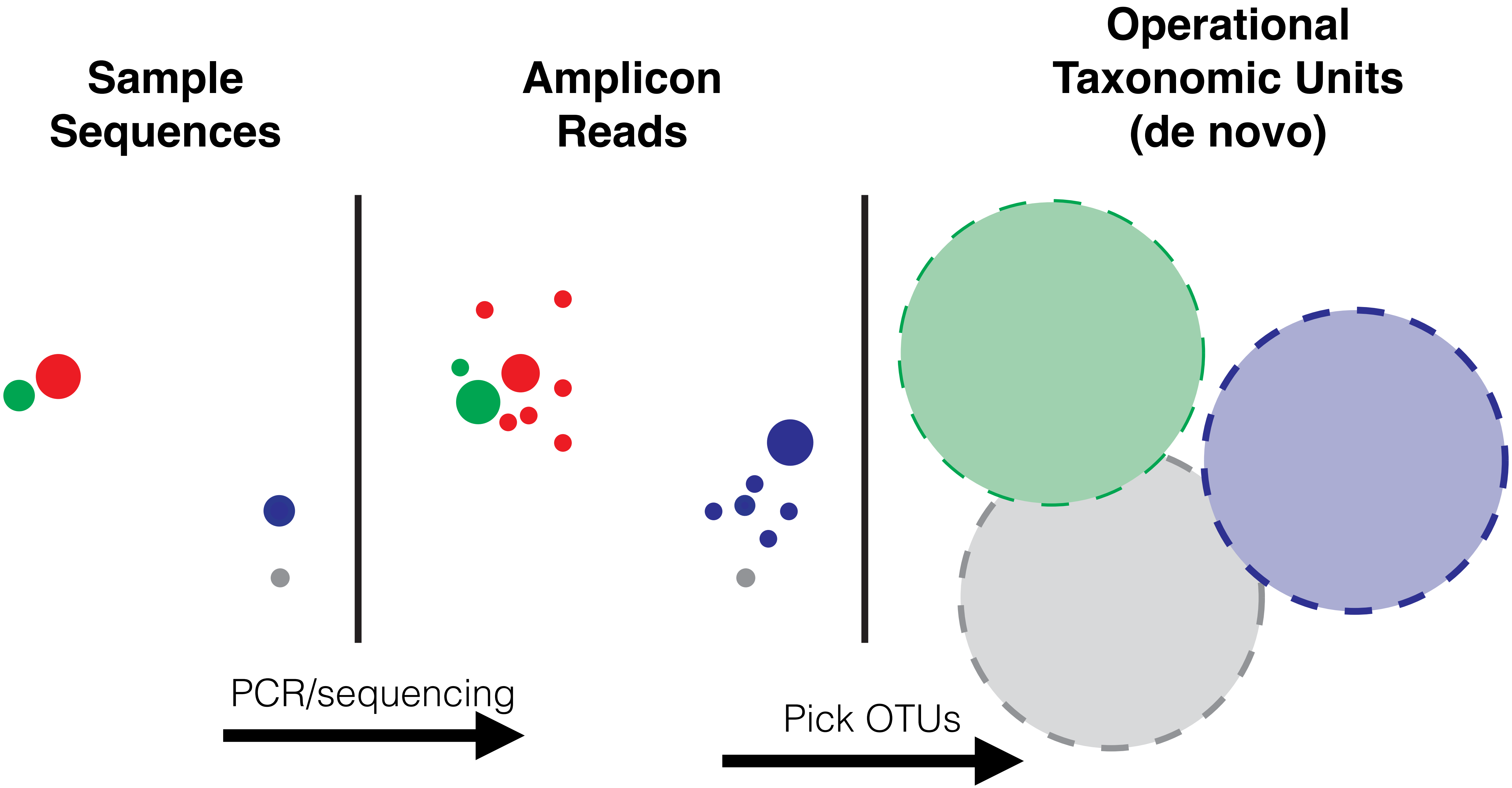
PCR/sequencing

Sample Sequences

Amplicon Reads



PCR/sequencing



Sample Sequences

Amplicon Reads

Operational Taxonomic Units (de novo)

PCR/sequencing

Pick OTUs

Different OTUs!

(In)consistent labels

`OTU85` is ***not a consistent label***

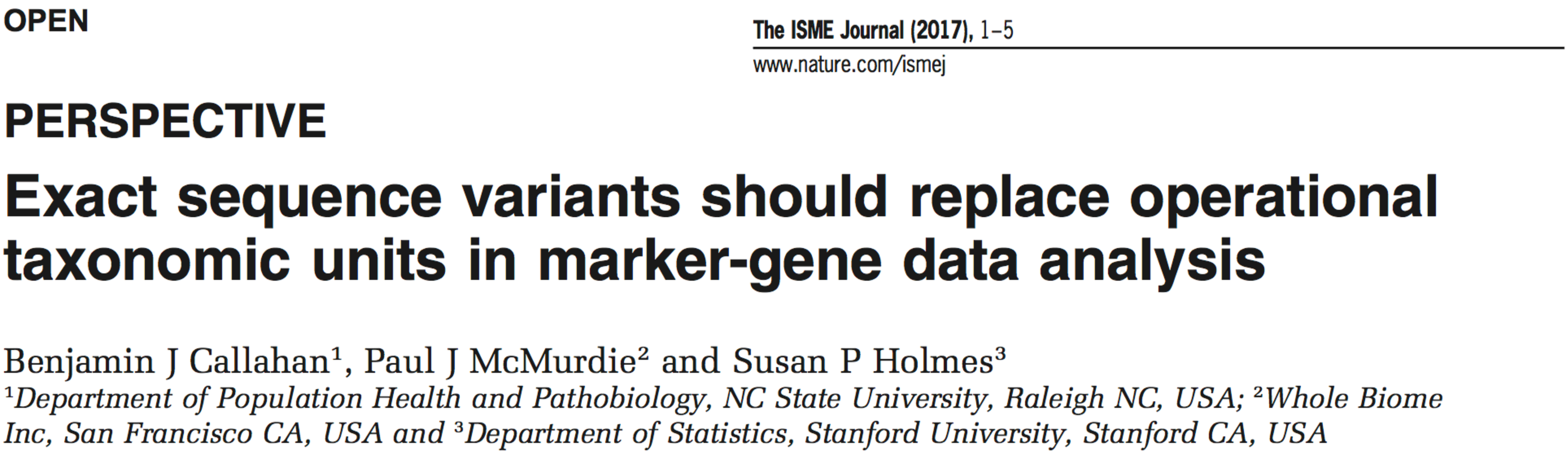
`OTU85` is *predictive of a disease?* **Not in future data!**

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

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

McLaren & Callahan, mBio, 2018.

Exact Sequence Variants…



OPEN

The ISME Journal (2017), 1–5
www.nature.com/ismej

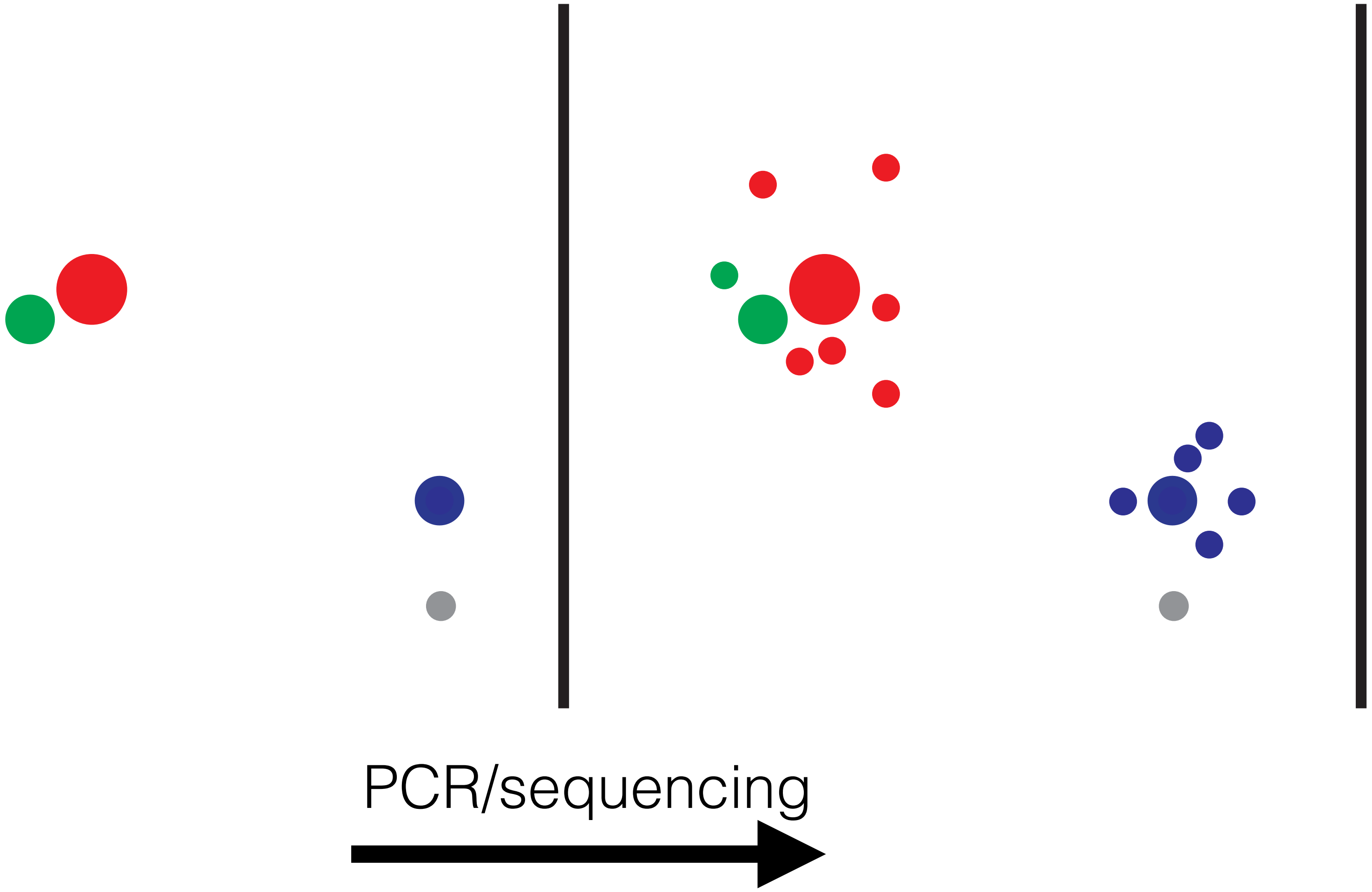
PERSPECTIVE

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

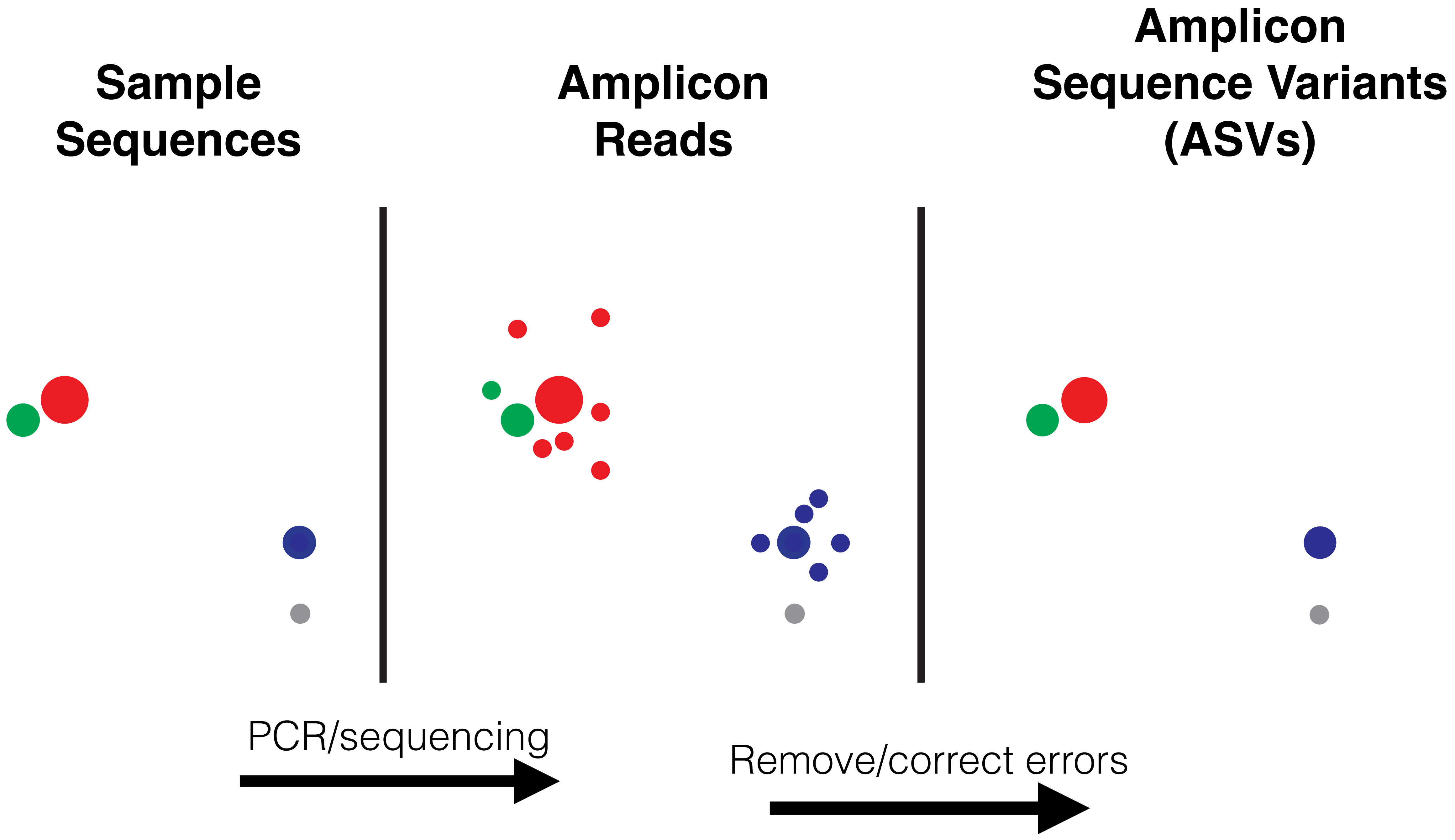
Benjamin J Callahan[1], Paul J McMurdie[2] and Susan P Holmes[3]
[1]Department of Population Health and Pathobiology, NC State University, Raleigh NC, USA; [2]Whole Biome Inc, San Francisco CA, USA and [3]Department of Statistics, Stanford University, Stanford CA, USA

Sample Sequences

Amplicon Reads



PCR/sequencing



Sample Sequences

Amplicon Reads

Amplicon Sequence Variants (ASVs)

PCR/sequencing

Remove/correct errors

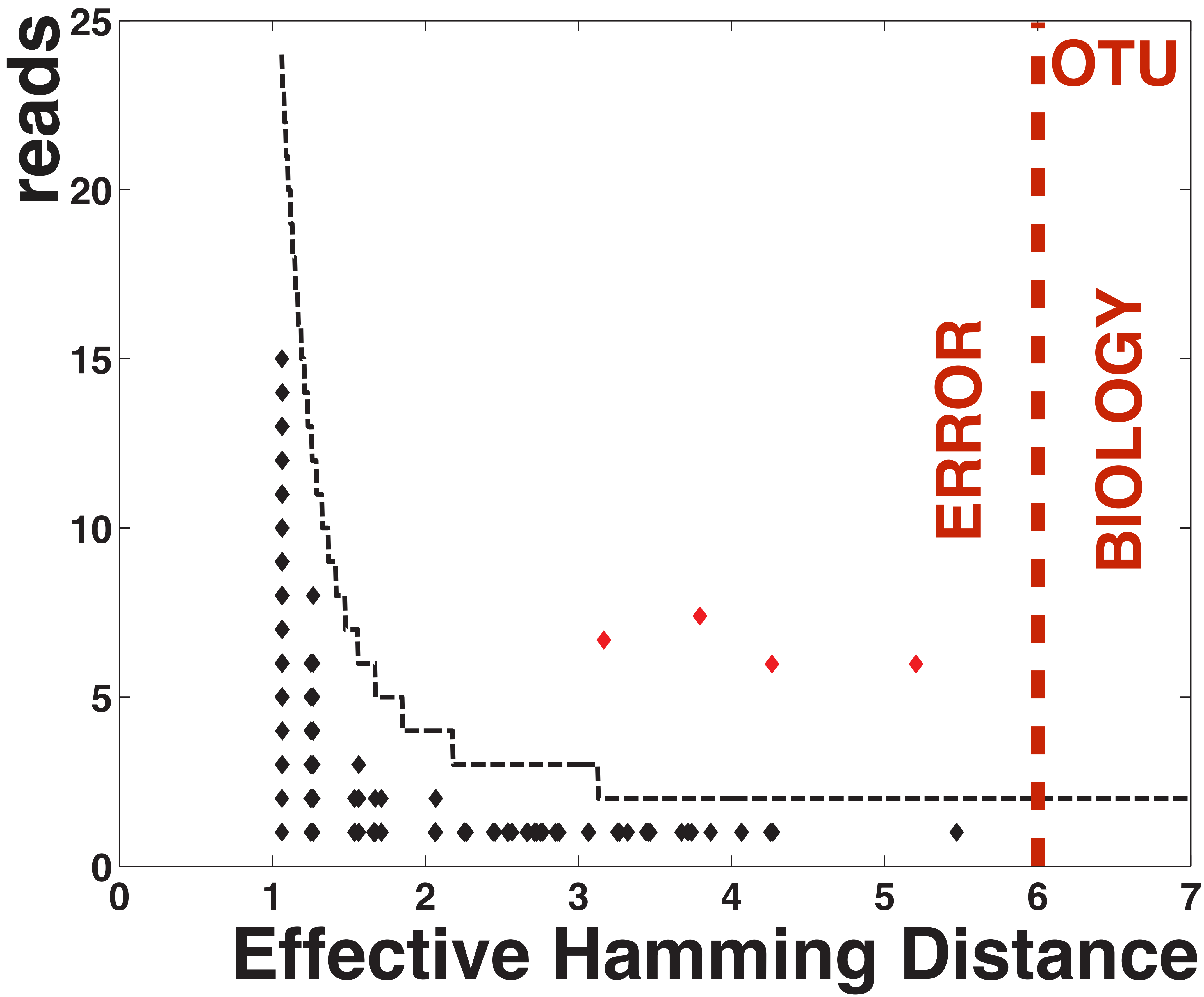
Exact Sequence Variants

`OTU85` is ***not* a consistent label,** *but…*

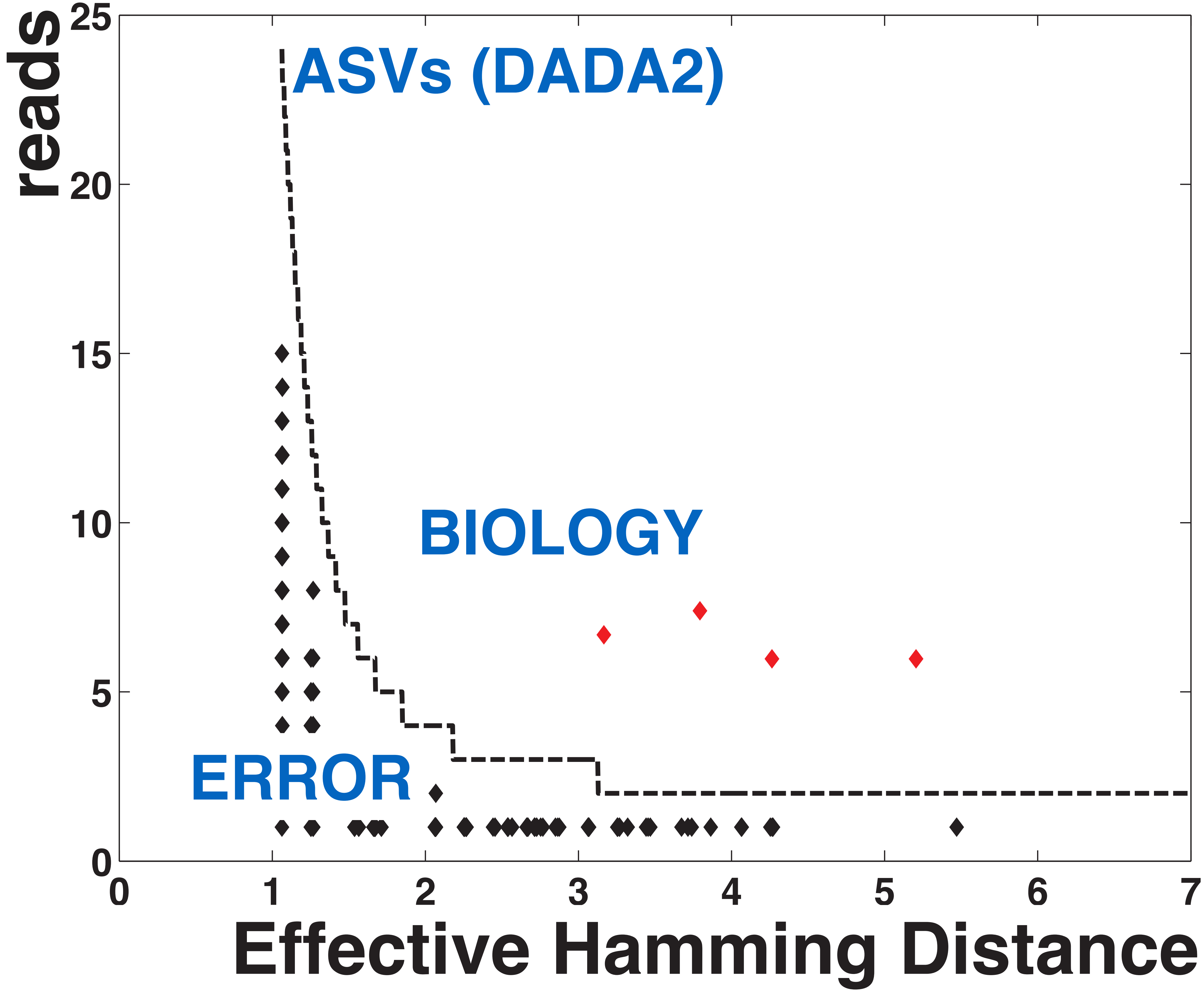
ATTAACGAGATTATAACCAGAGTACGAATA…

is consistent!

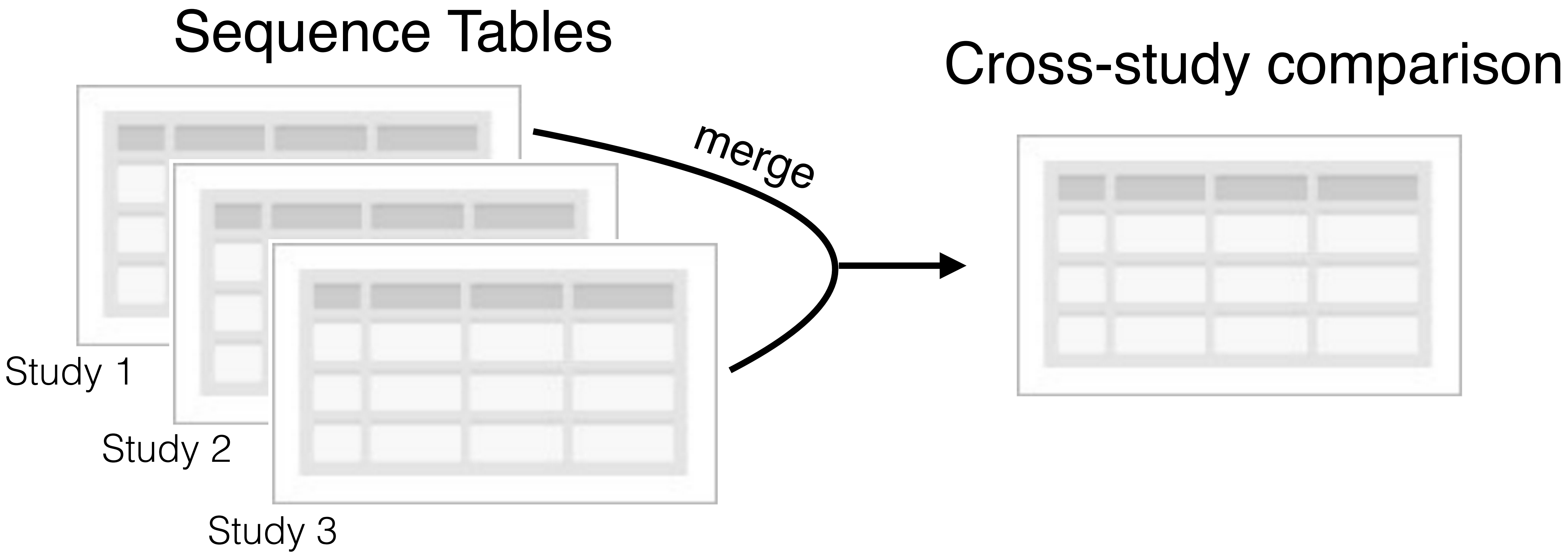
Signal from Noise: OTUs



Signal from Noise: ASVs



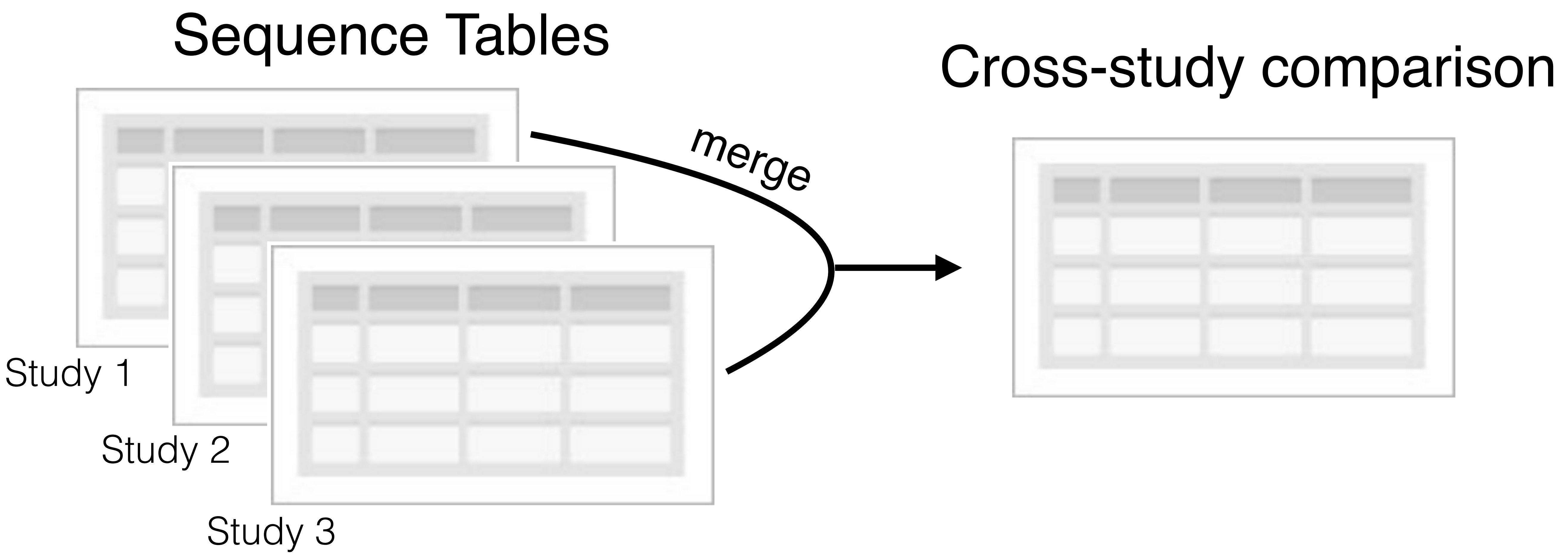
The Sequence is the Label



Eliminates need for joint reprocessing of raw data.

Callahan, et al. ISMEJ, 2017.

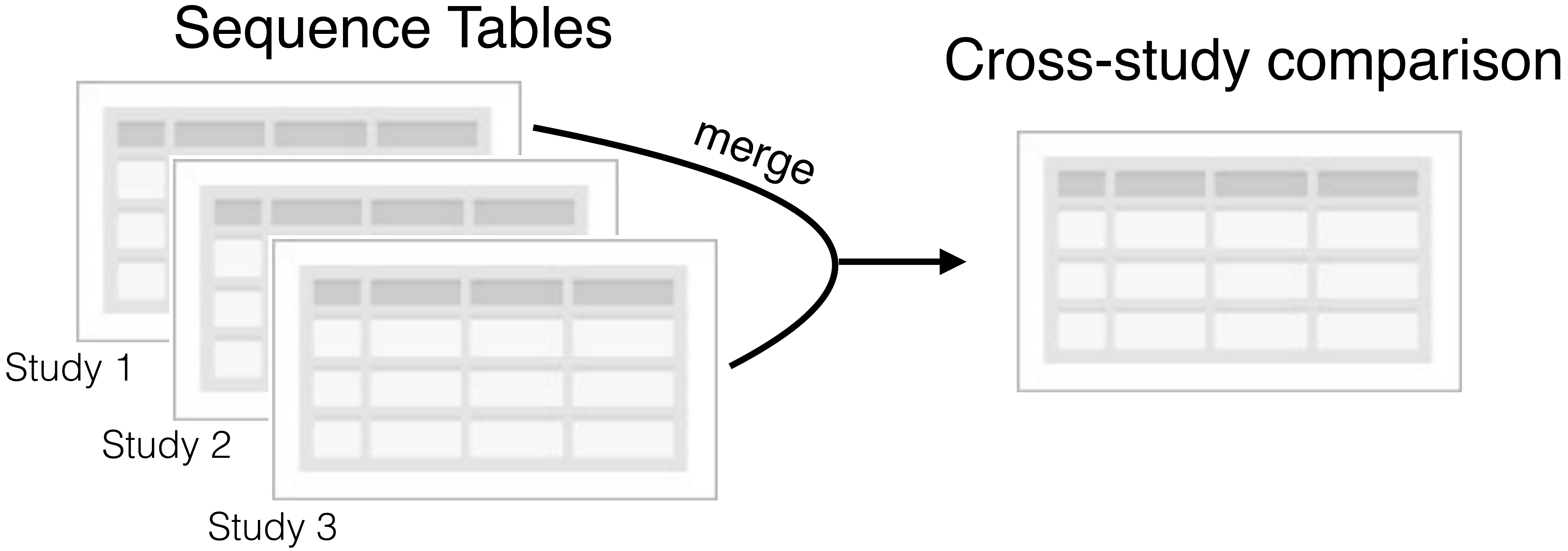
The Sequence is the Label



Eliminates need for joint reprocessing of raw data.
Continuous data integration. Unlimited dataset size.

Callahan, et al. ISMEJ, 2017.

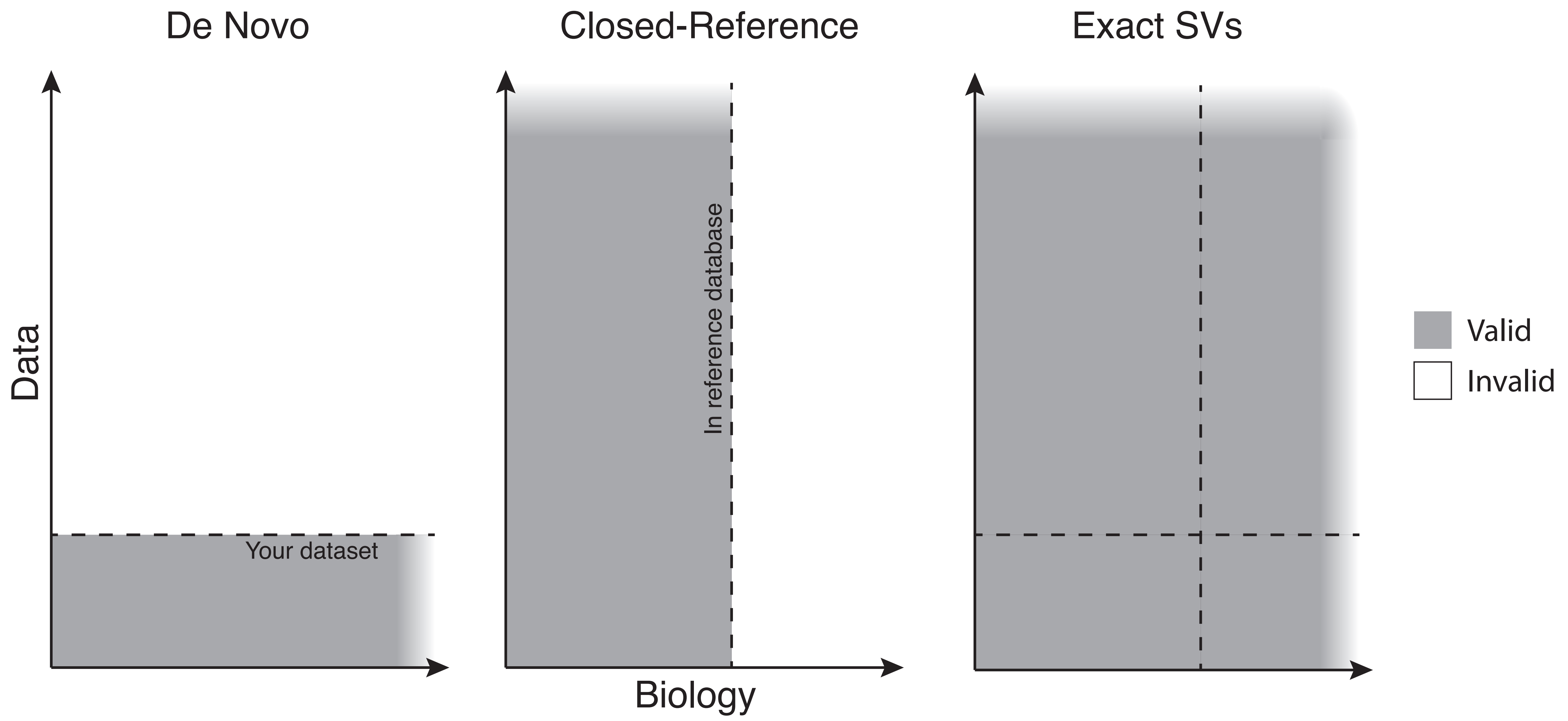
The Sequence is the Label



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

Callahan, et al. ISMEJ, 2017.

The Sequence is the Label



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

Callahan, et al. ISMEJ, 2017.

ASV Methods and Terminology

ASV Methods

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

ASV Methods and Terminology

ASV Methods

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

Synonym acronyms

- 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	✓	X	✓
Comprehensive	✓	✓	X

Callahan, et al. ISMEJ, 2017.


I've made a robot that lumps similar sequences together!
97% OTUs!!
WHY?
OTU

ASVs vs. OTUs

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

Callahan, et al. ISMEJ, 2017.

ASVs vs. OTUs

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

Are OTUs *wrong*?

Callahan, et al. ISMEJ, 2017.

ASVs vs. OTUs

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

Are OTUs *wrong*? No.

Callahan, et al. ISMEJ, 2017.

ASVs vs. OTUs

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

Are OTUs *wrong*? No.

Are ASVs *always* the best unit of analysis?

Callahan, et al. ISMEJ, 2017.

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.

Callahan, et al. ISMEJ, 2017.

Phylogenetic Scale

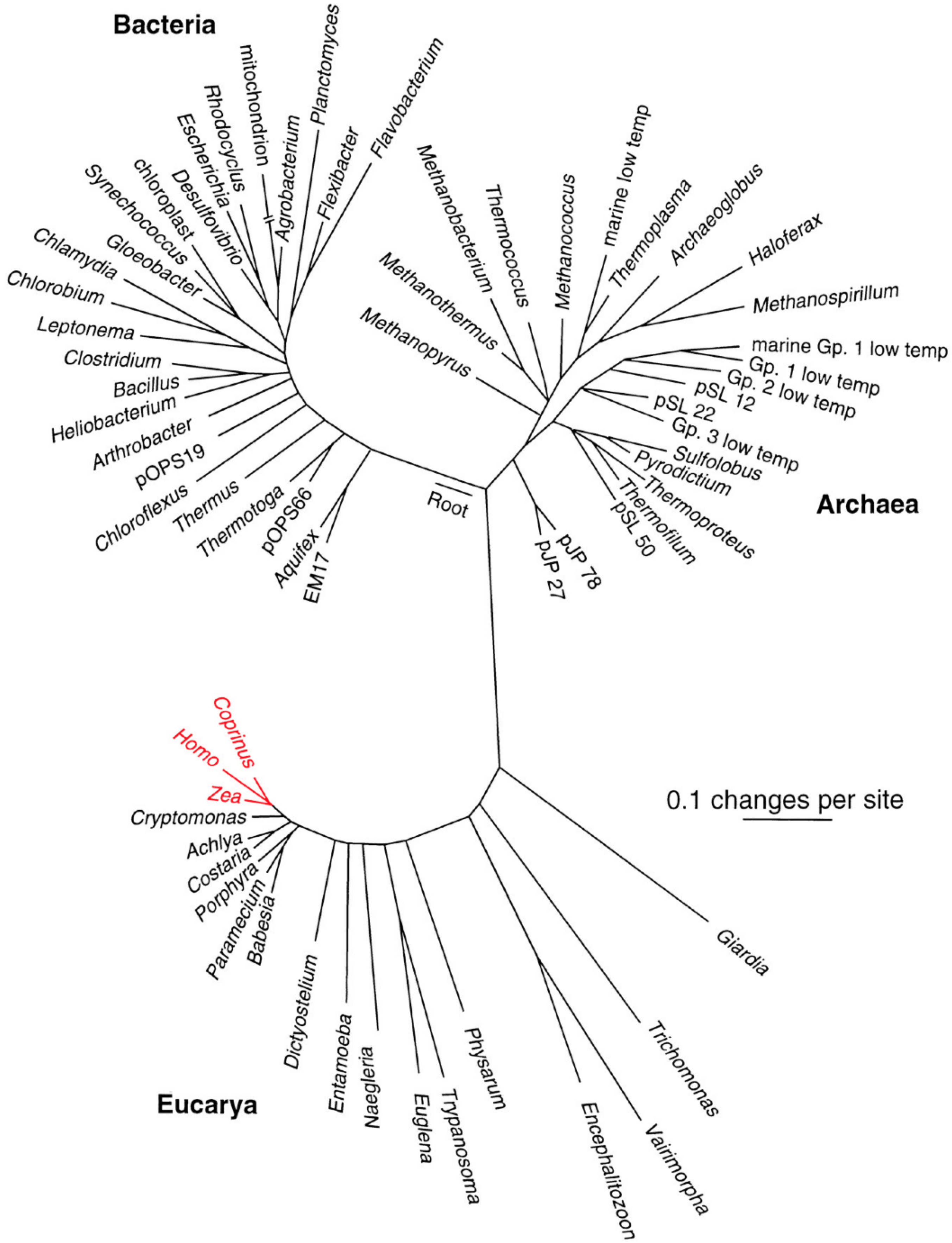
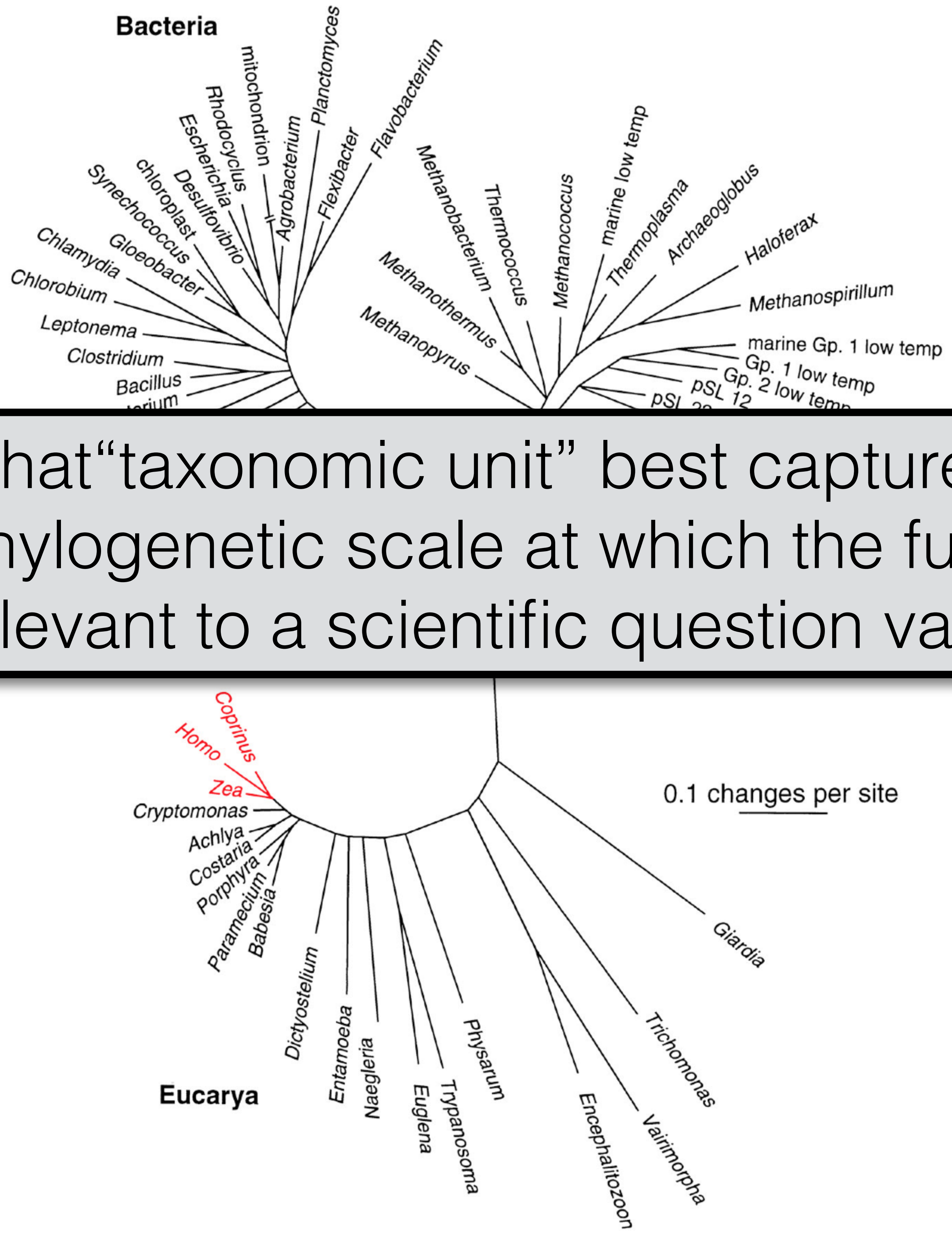


Image: Norman Pace

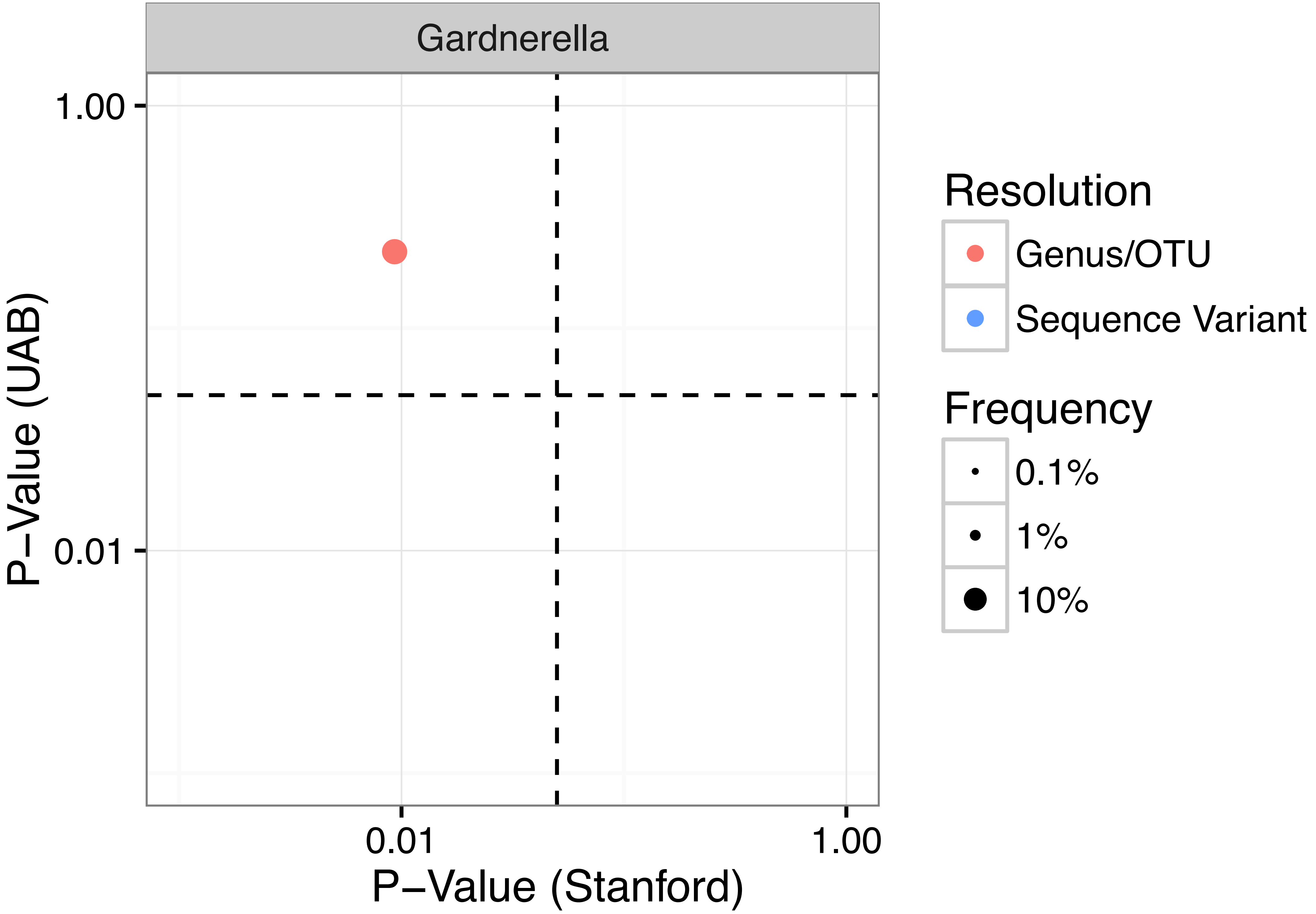
Phylogenetic Scale



What“taxonomic unit” best captures the phylogenetic scale at which the function relevant to a scientific question varies?

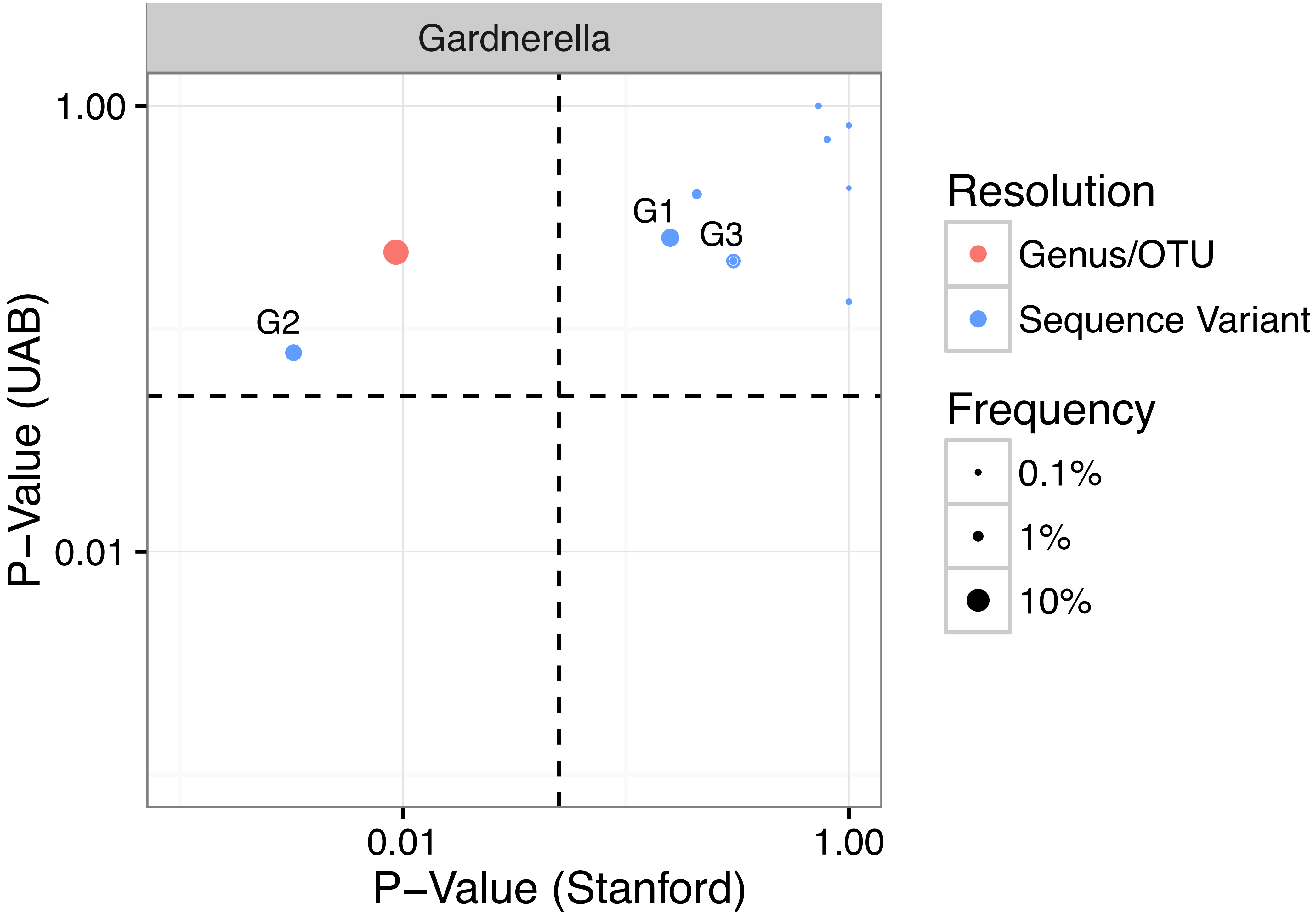
Image: Norman Pace

Resolution: Preterm Birth



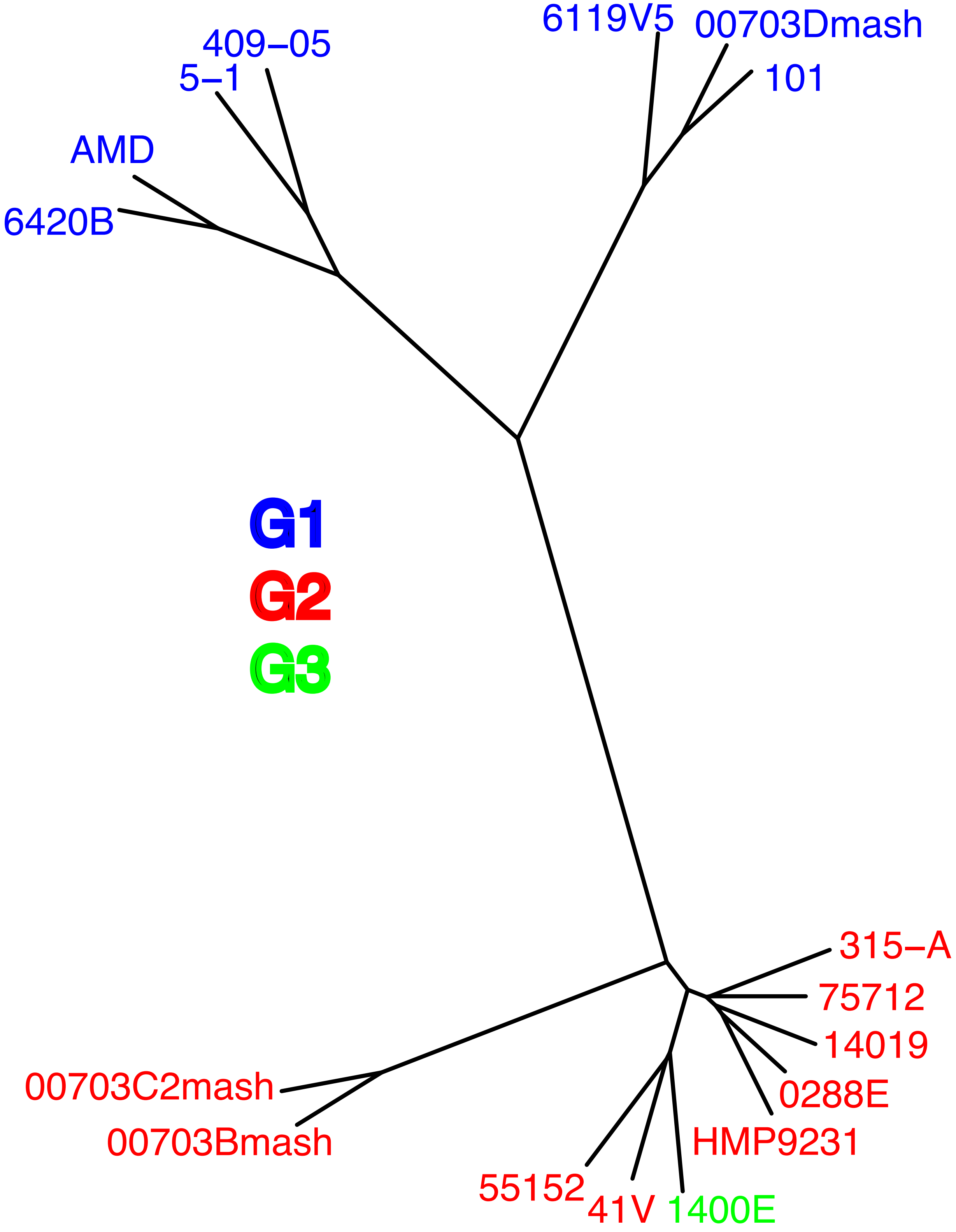
Callahan et al. PNAS, 2017.

Resolution: Preterm Birth



Callahan et al. PNAS, 2017.

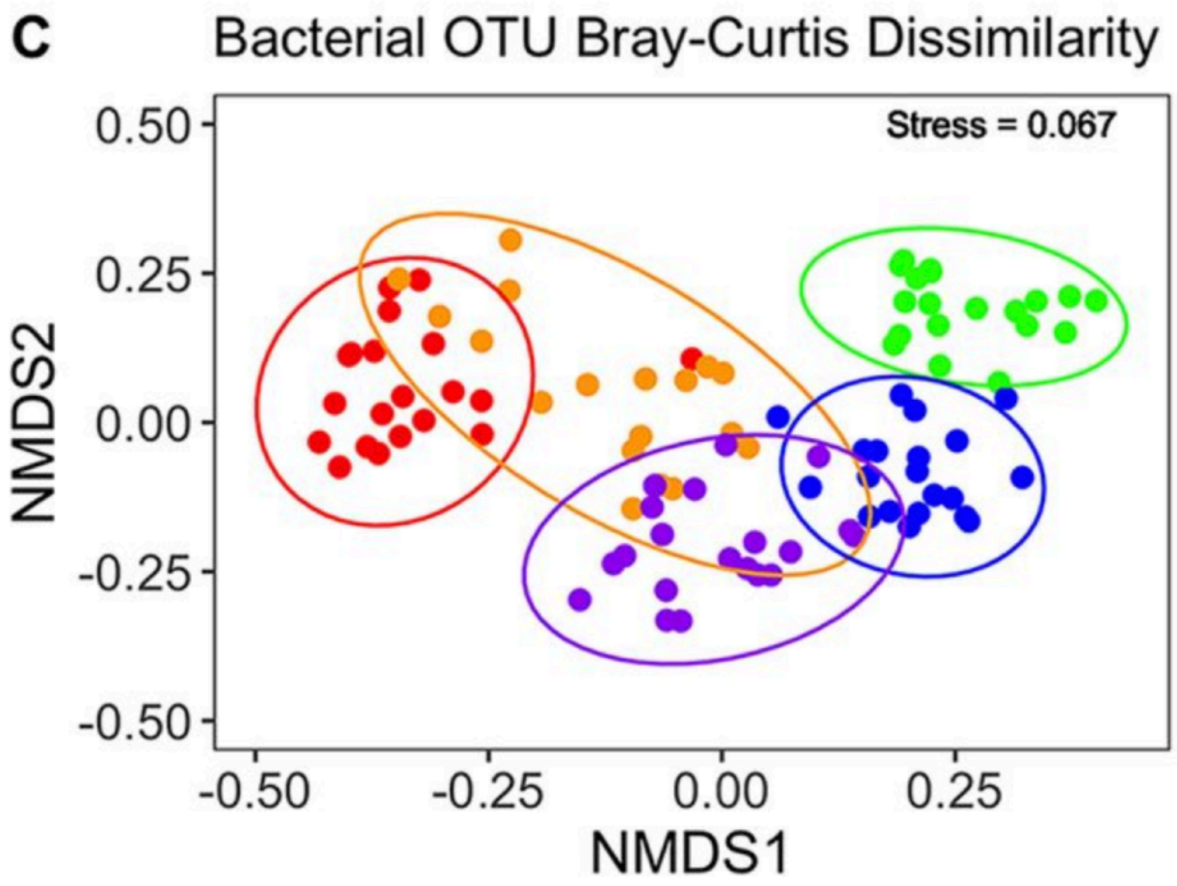
Resolution: Preterm Birth



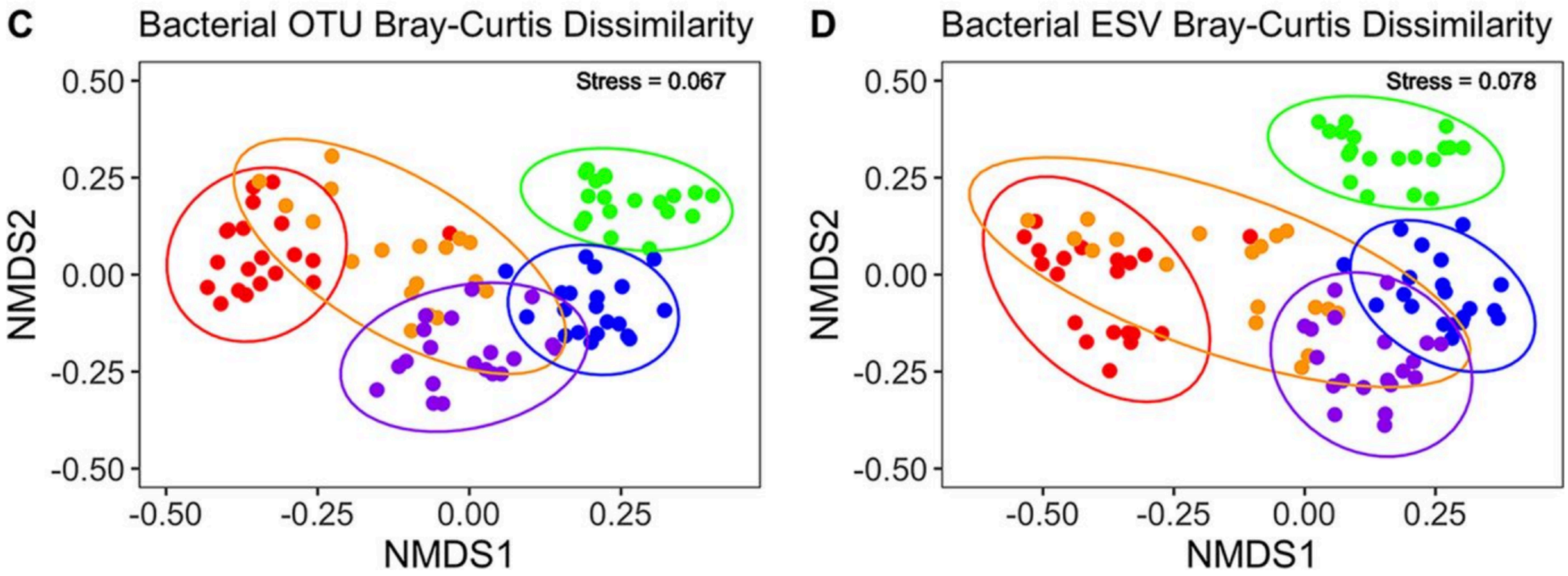
Tree from
whole genomes

Callahan et al. PNAS, 2017.

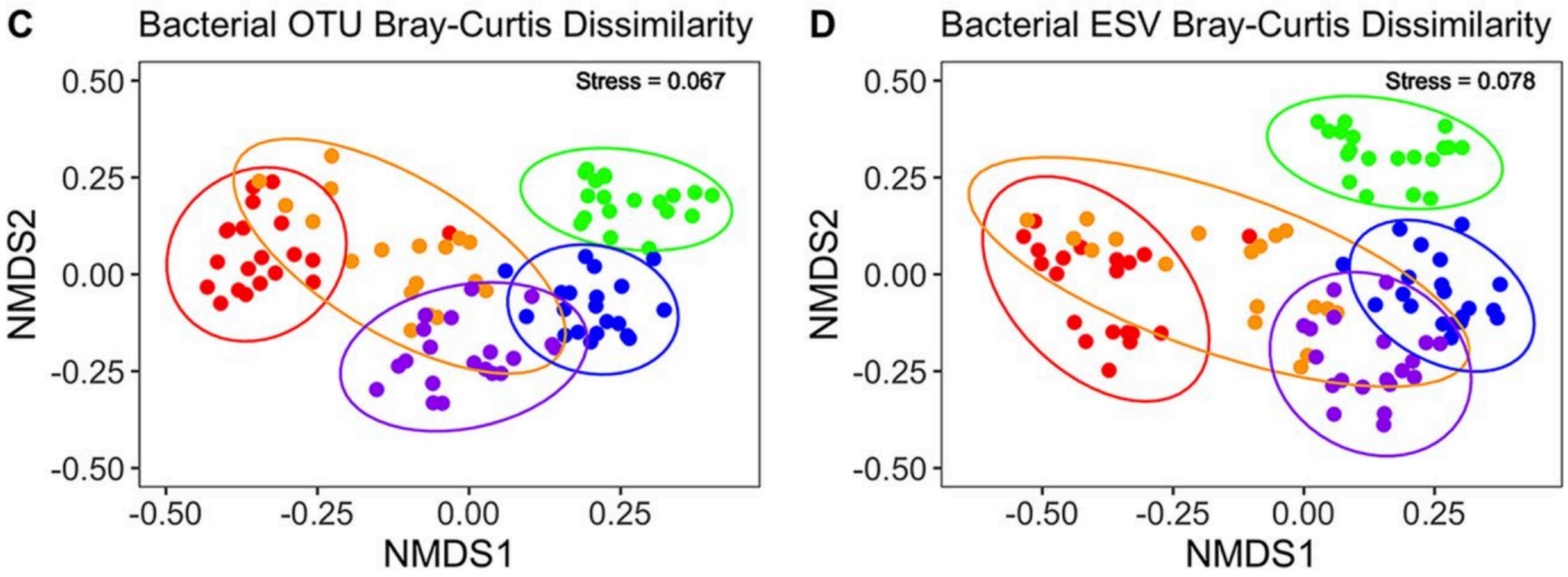
Resolution: Broadscale Patterns

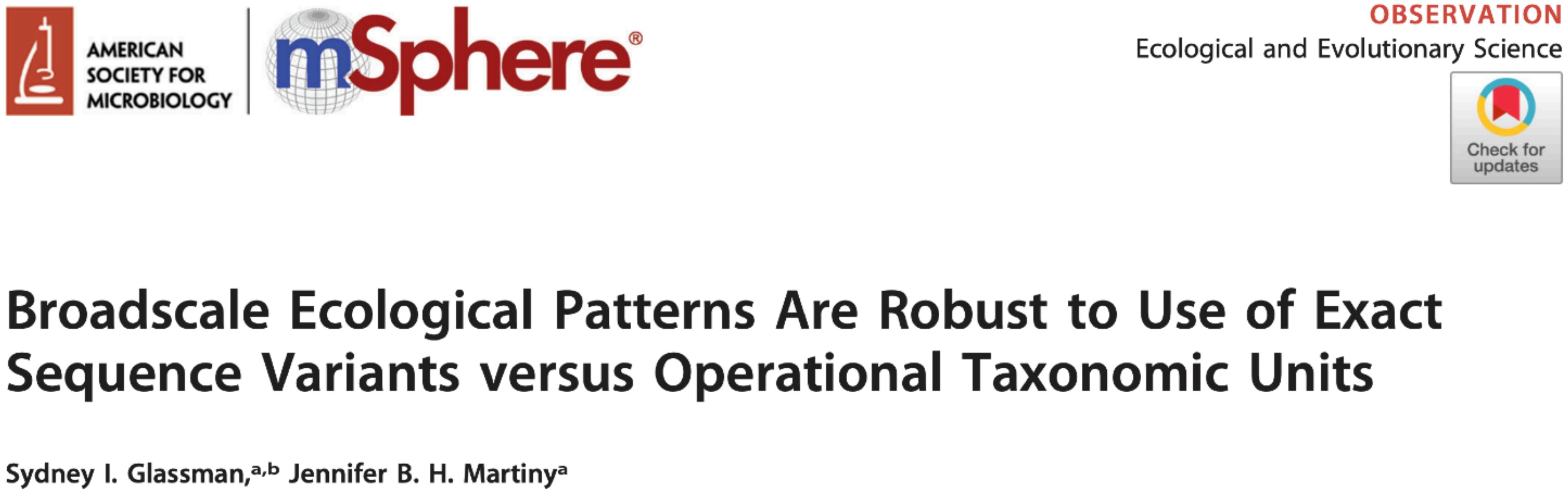


Resolution: Broadscale Patterns



Resolution: Broadscale Patterns




AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

OBSERVATION
Ecological and Evolutionary Science

Check for updates

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

<u>Typically will</u> discriminate pairs of species in a genus.

<u>Typically can’t</u> 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

Minimum Entropy Decomposition

USEARCH
Ultra-fast sequence analysis

Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns

NIH National Institutes of Health

BIO-X

NSF

NC STATE UNIVERSITY Plant Soil Microbial Community Consortium