

Exact Amplicon Sequence Variants

Diversity of Life

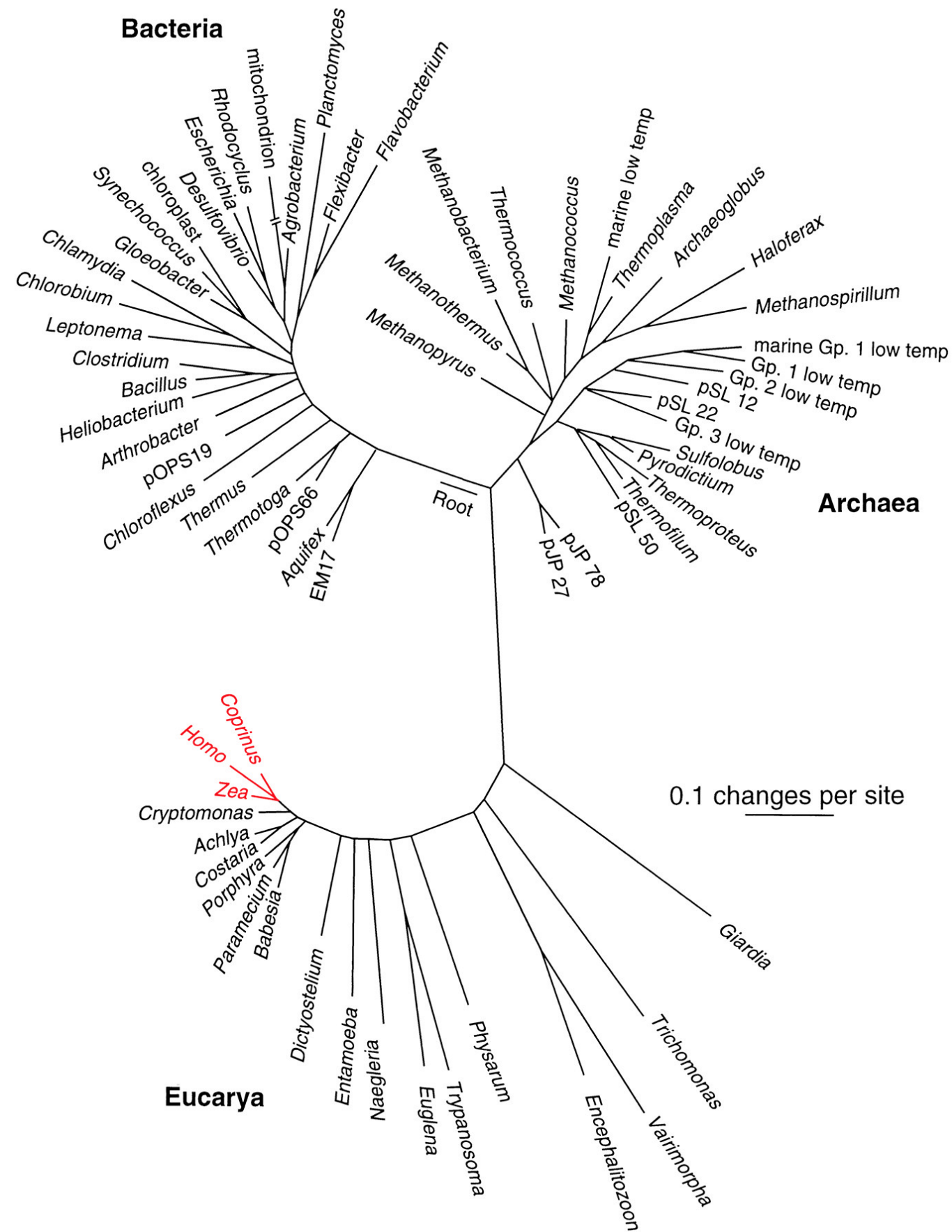
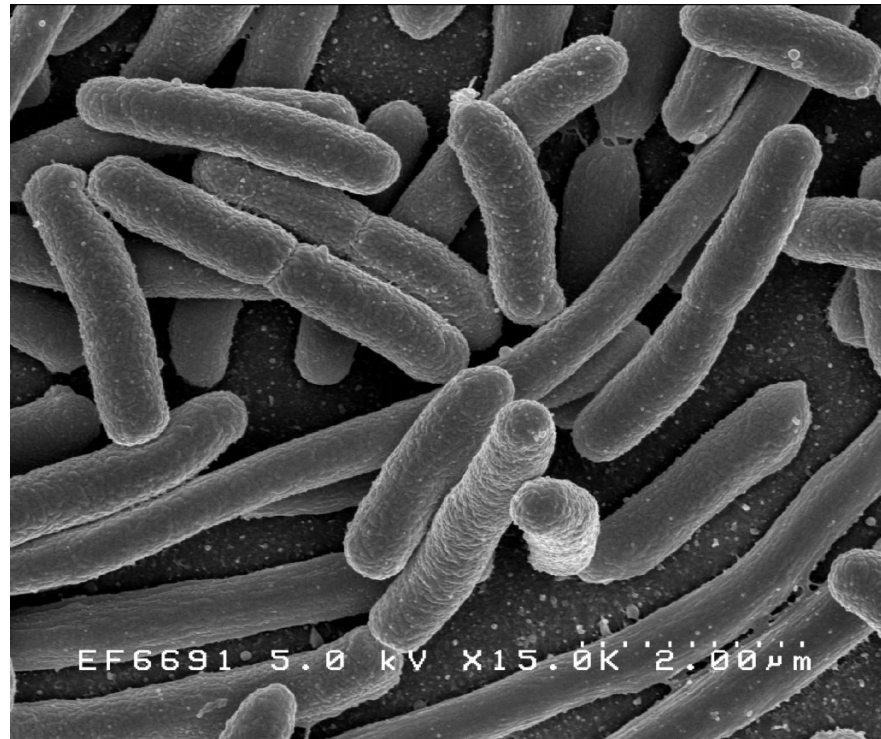


Image: Norman Pace

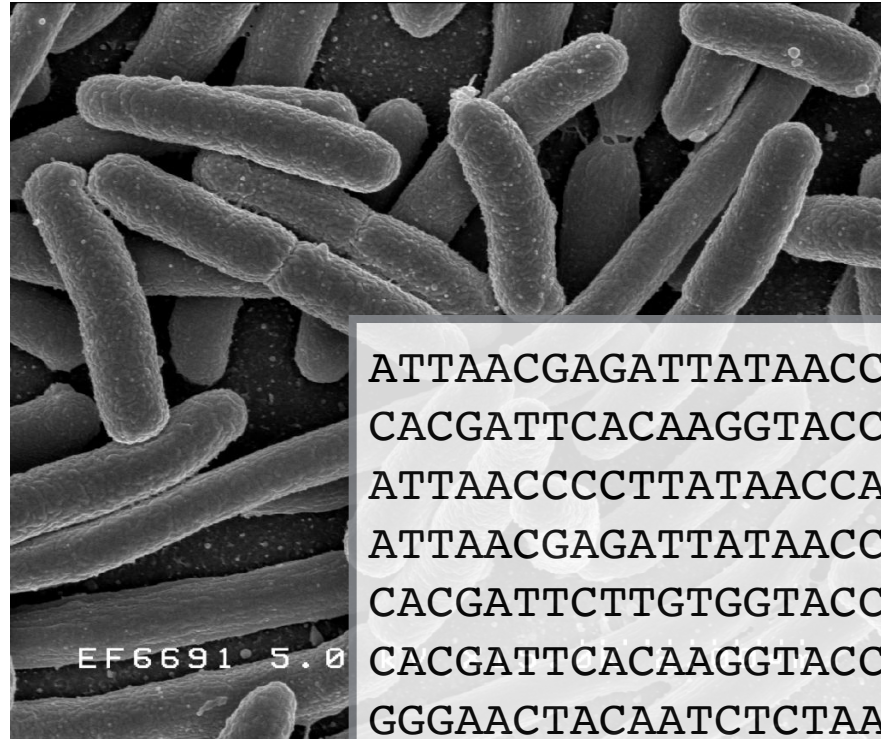
A Molecular Census

Metabarcoding or Marker-gene Sequencing



A Microbial Census

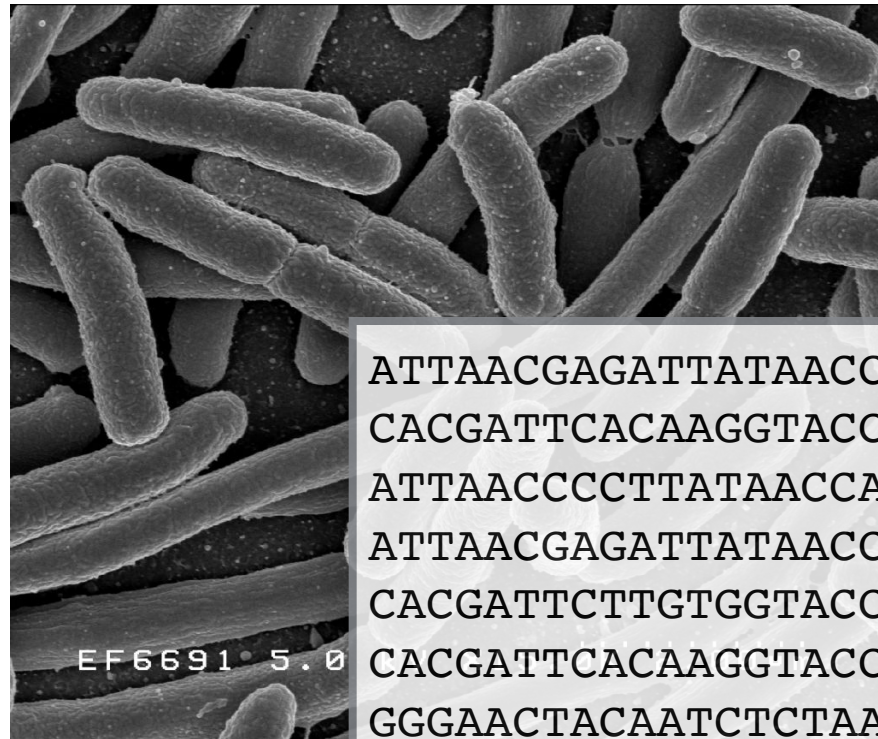
Metabarcoding or Marker-gene Sequencing



```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC  
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC  
ATTAACCCCTTATAACCAGAGTACGAATACCGAAC  
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC  
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC  
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC  
GGGA ACTACAATCTCTAAGGTGAAGTCTCAGTCTAT  
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC  
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC  
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
```

A Microbial Census

Metabarcoding or Marker-gene Sequencing

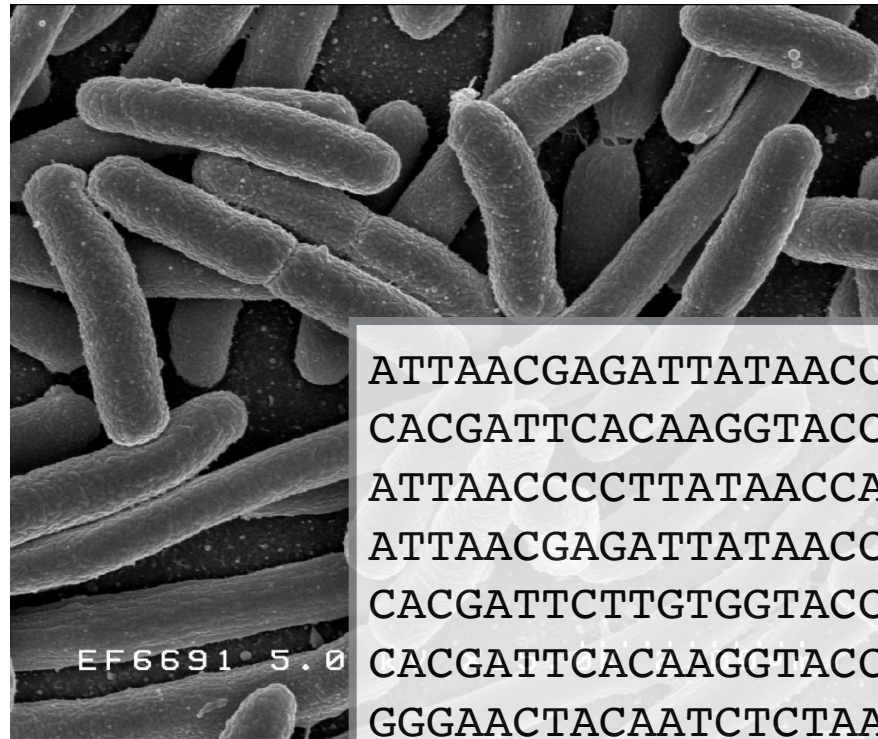


ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAAC
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGA ACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC

<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
...

A Microbial Census

Metabarcoding or Marker-gene Sequencing



ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGA ACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC

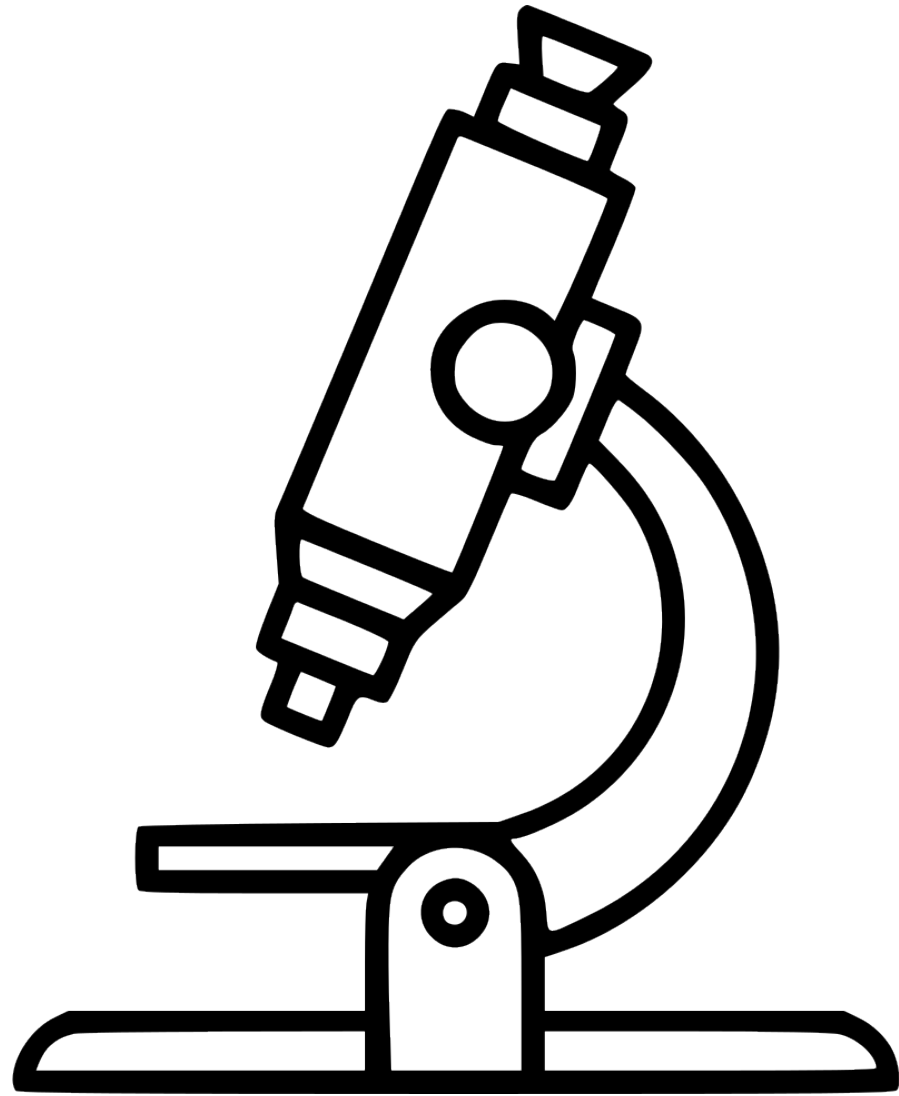
<i>Lactobacillus crispatus</i>	1300	5	0	882	596
<i>Ureaplasma urealytica</i>	15	0	220	0	0
<i>Gardnerella vaginalis</i>	22	0	1	0	412
<i>Prevotella intermedia</i>	0	0	8	12	0
...

Visualization

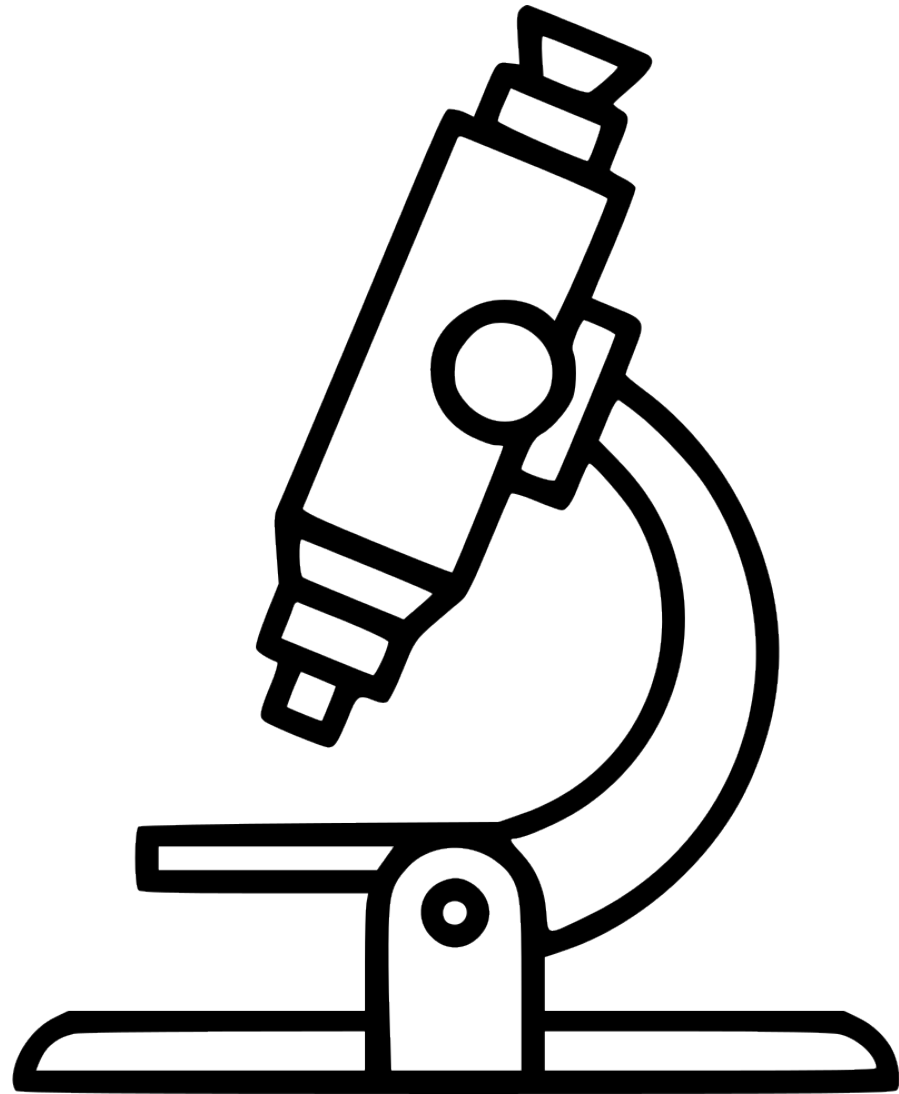
Exploration

Inference

The MGS Microscope



The MGS Microscope



- **sample collection**
- **sample storage**
- **DNA extraction**
- **library preparation**
- **PCR**
- **sequencing instrument**
- **bioinformatics**

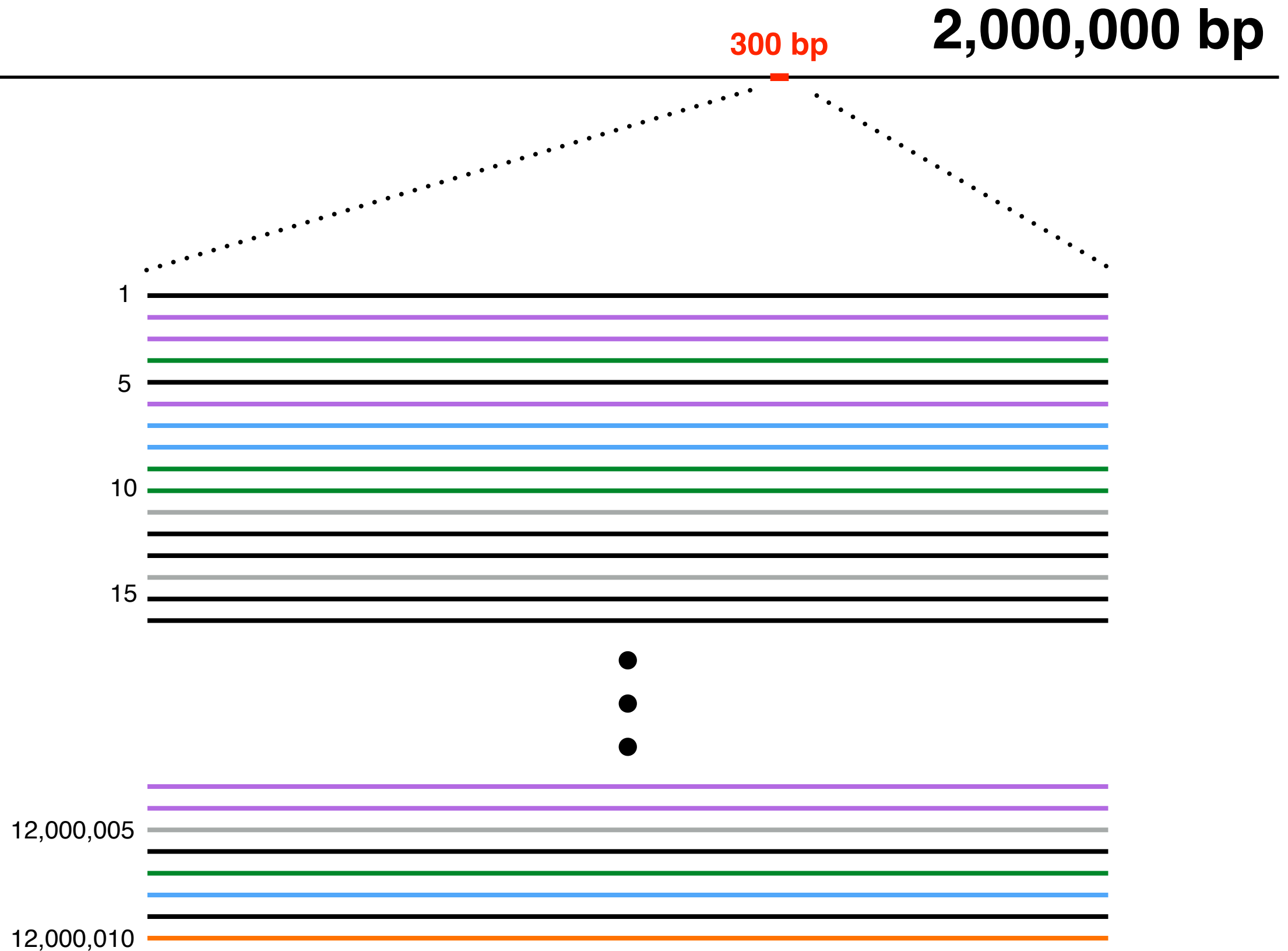
Marker-gene Sequencing

300 bp

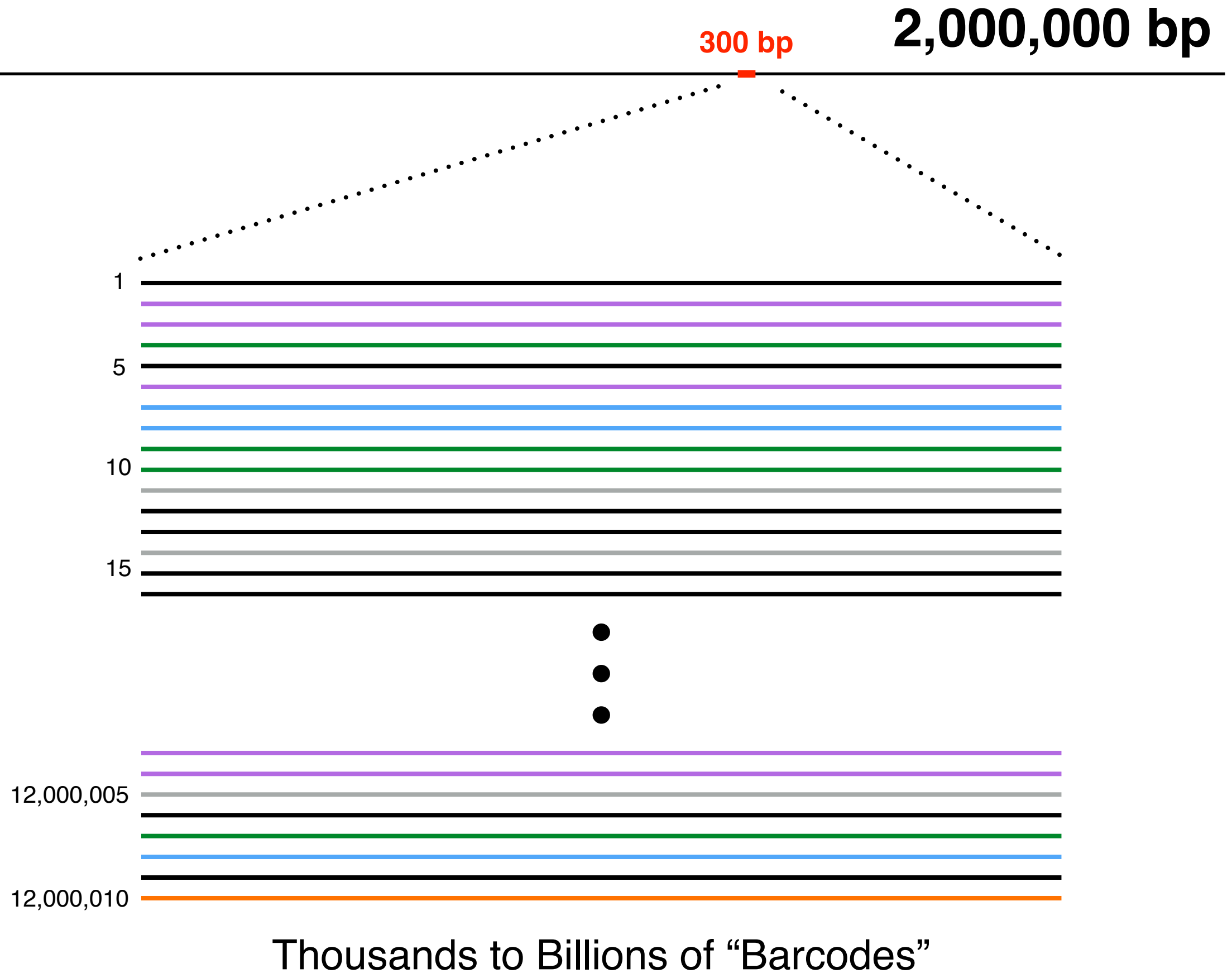
2,000,000 bp



Marker-gene Sequencing

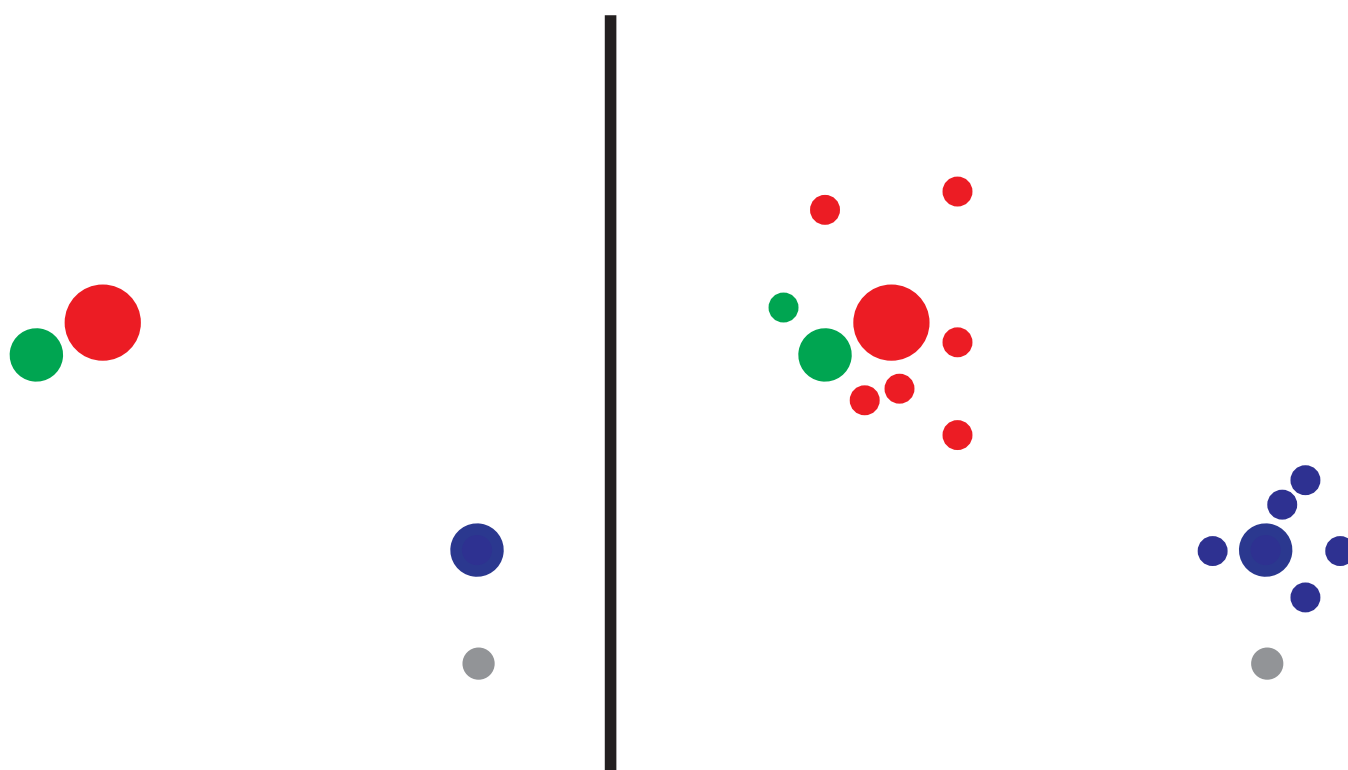


Marker-gene Sequencing



**Sample
Sequences**

**Amplicon
Reads**



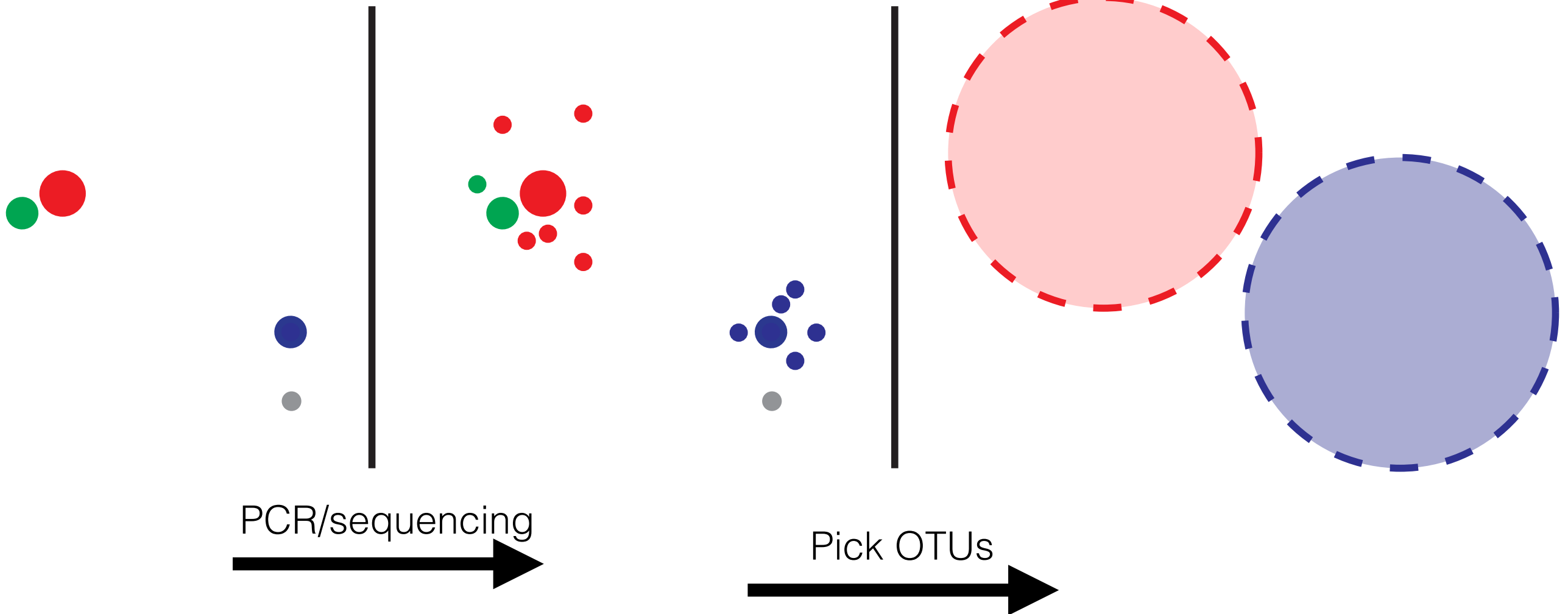
PCR/sequencing



Sample Sequences

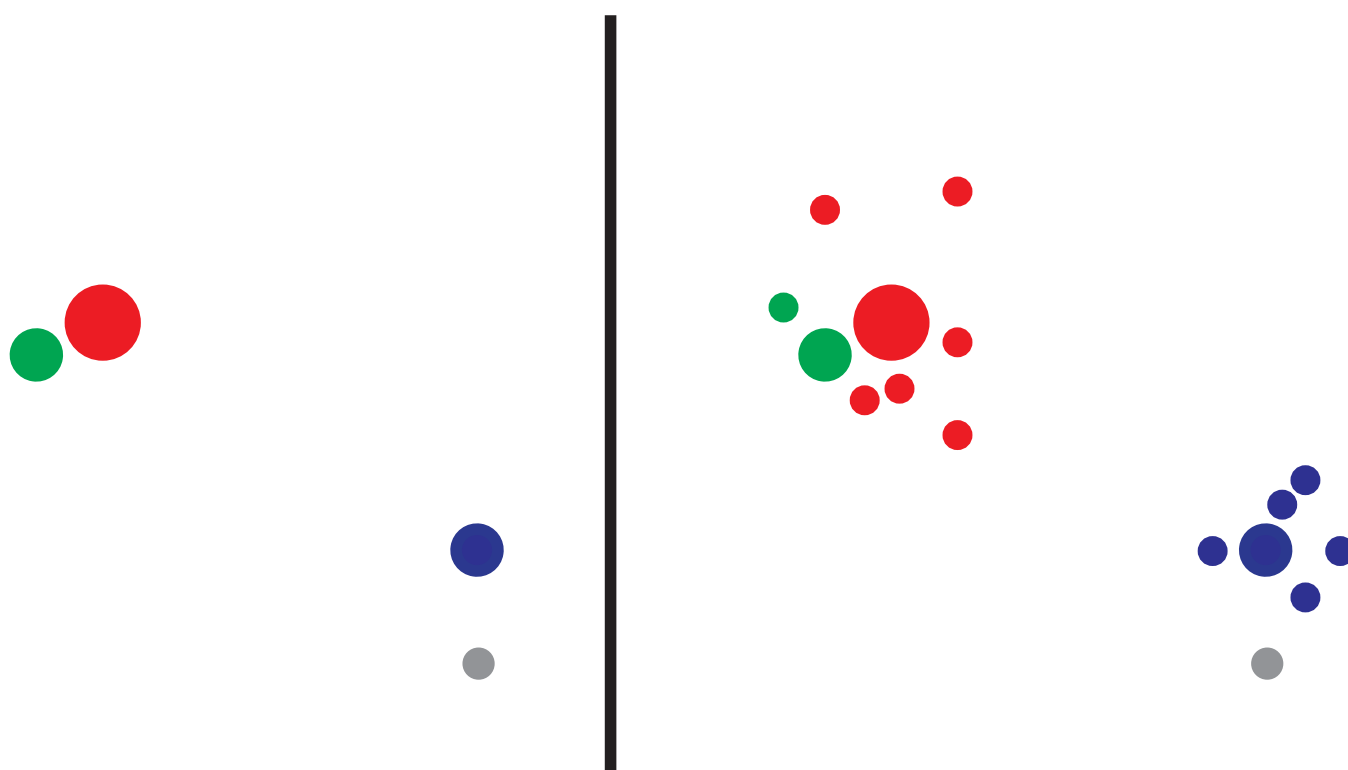
Amplicon Reads

Operational Taxonomic Units



**Sample
Sequences**

**Amplicon
Reads**

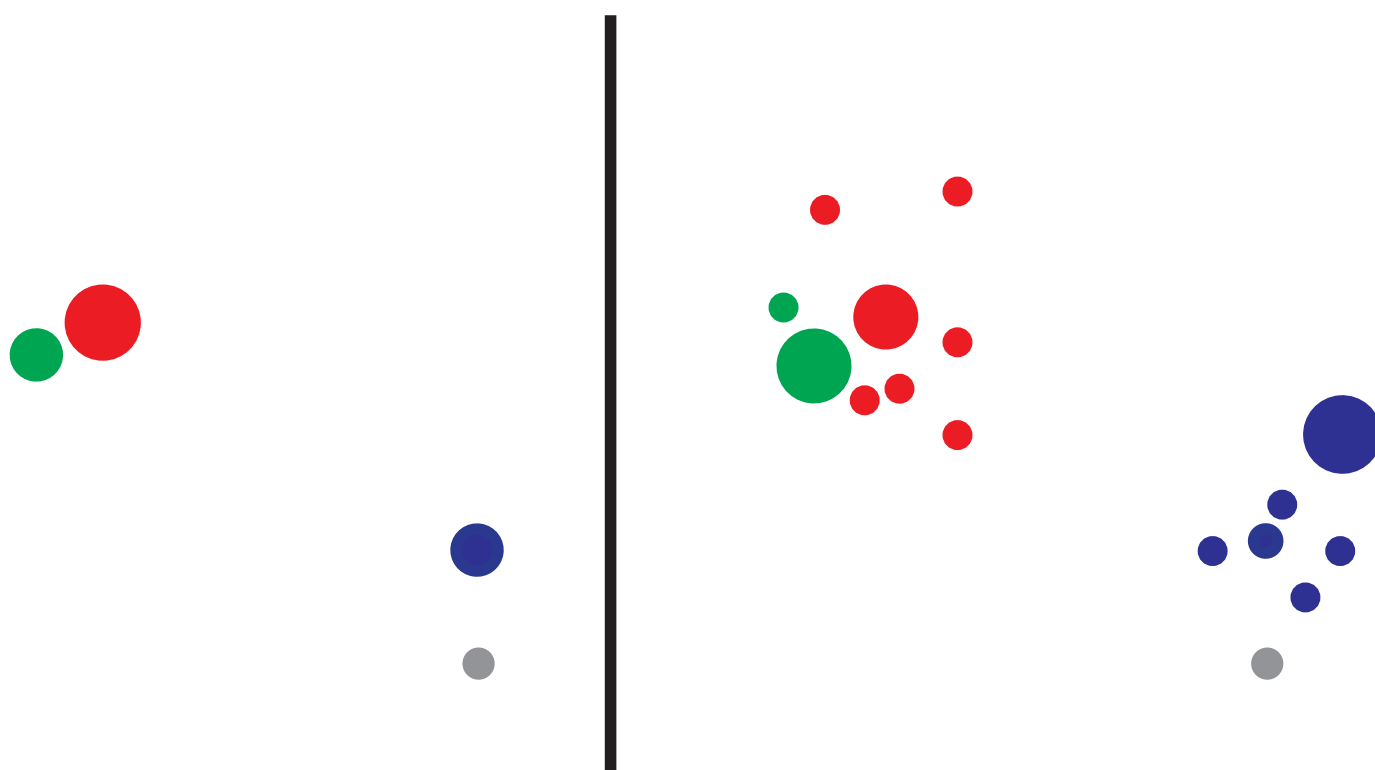


PCR/sequencing



**Sample
Sequences**

**Amplicon
Reads**



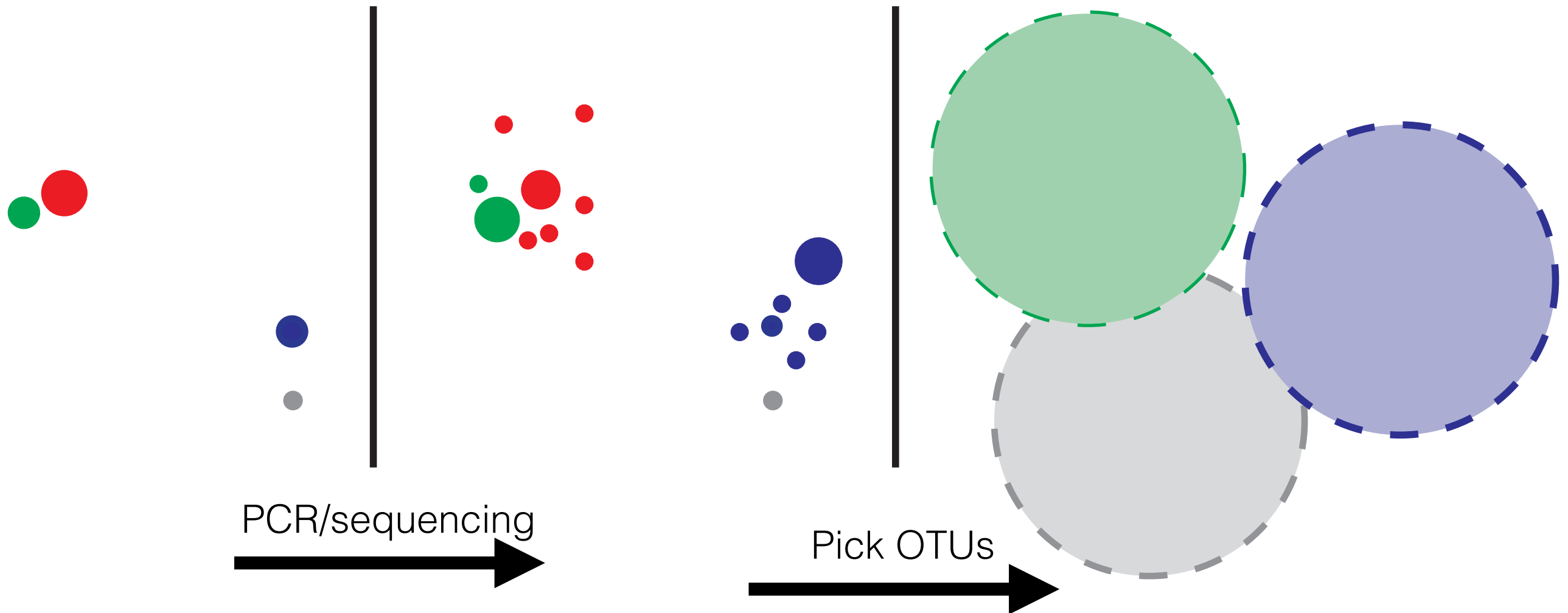
PCR/sequencing



**Sample
Sequences**

**Amplicon
Reads**

**Different Operational
Taxonomic Units**



Uh oh!

Exact Sequence Variants...

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

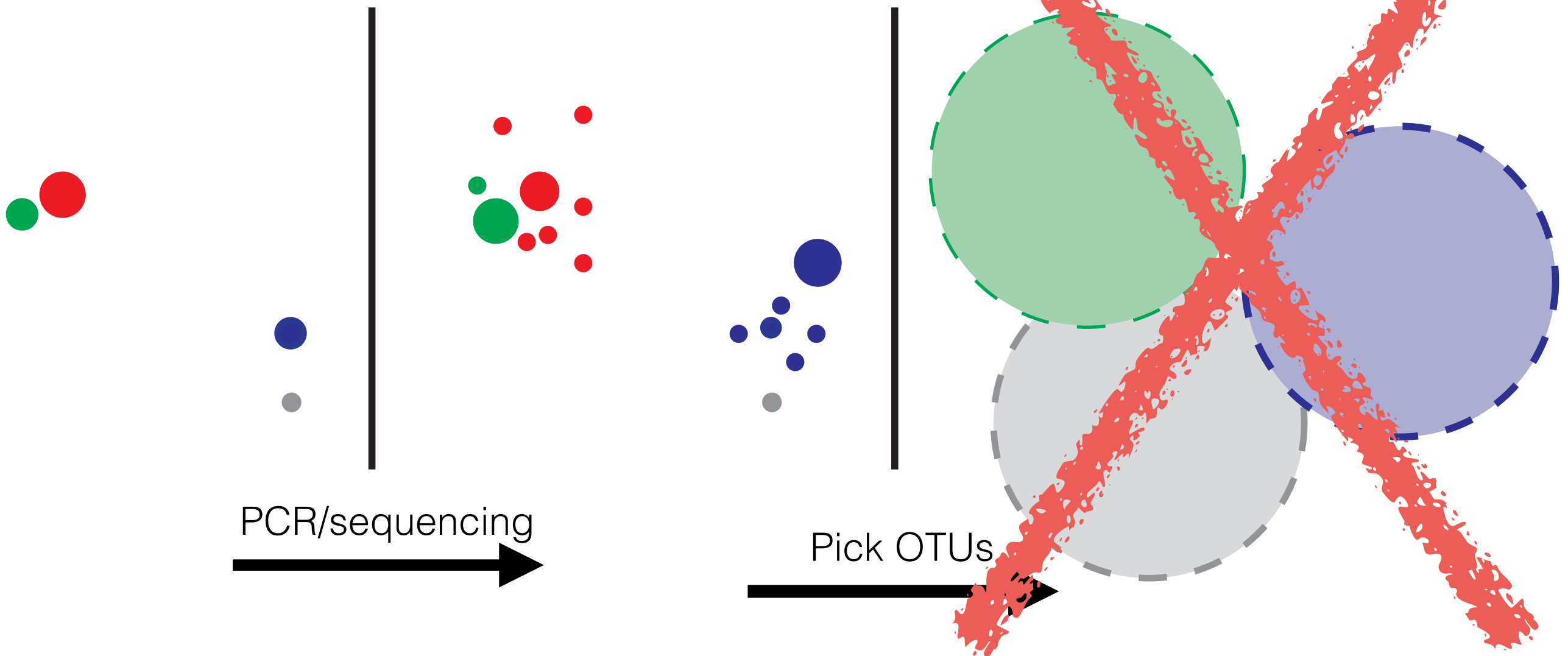
Benjamin J Callahan¹, Paul J McMurdie² and Susan P Holmes³

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**

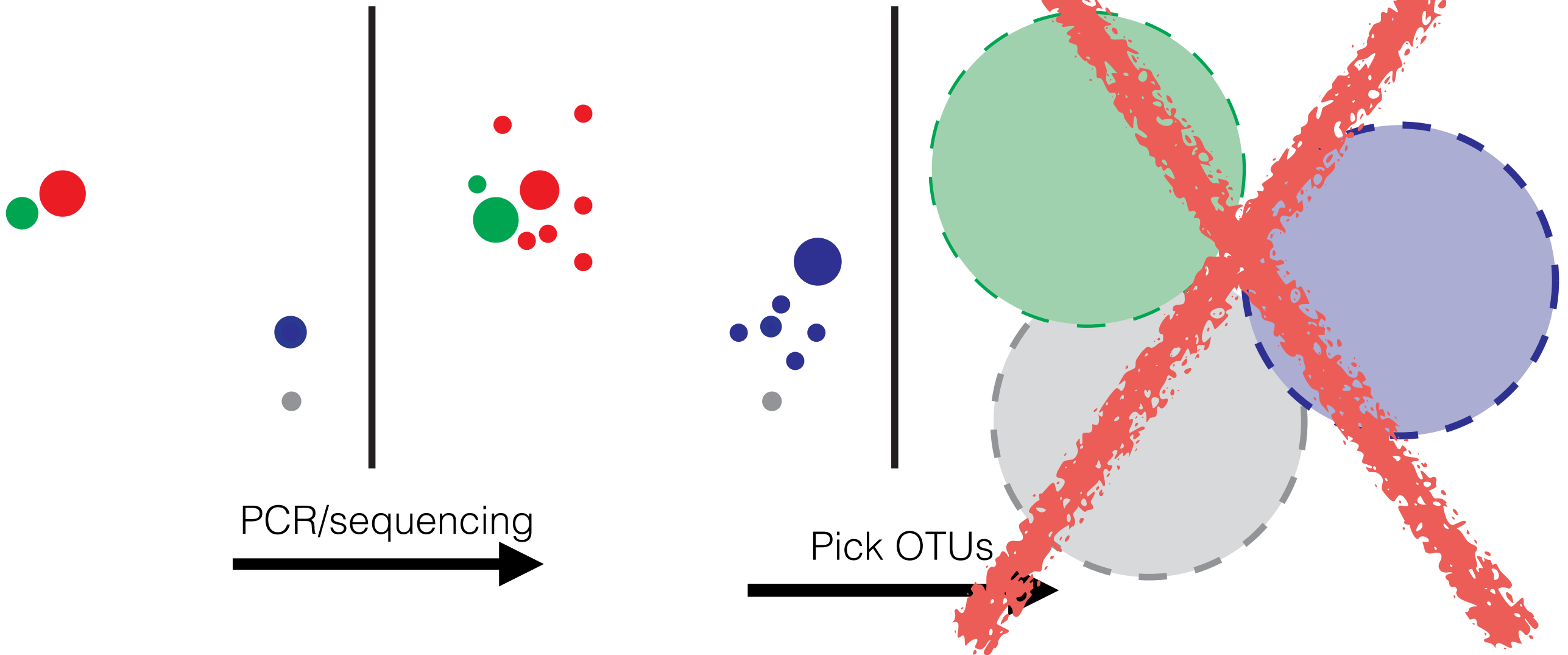
**Different Operational
Taxonomic Units**



**Sample
Sequences**

**Amplicon
Reads**

**Different Operational
Taxonomic Units**



PCR/sequencing

Pick OTUs

Infer Sample Sequences

Exact Sequence Variants...

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

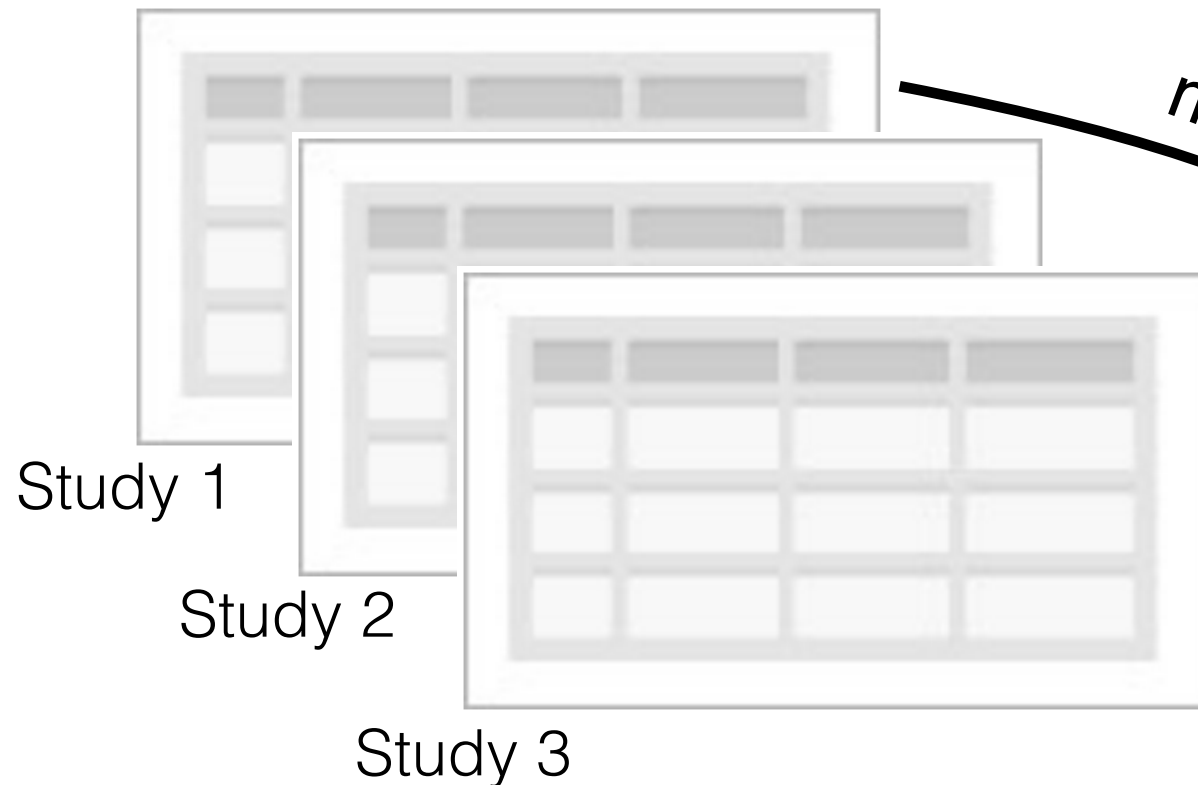
ATTAAACGAGATTATAACCCAGAGTACGAATA...

is consistent!



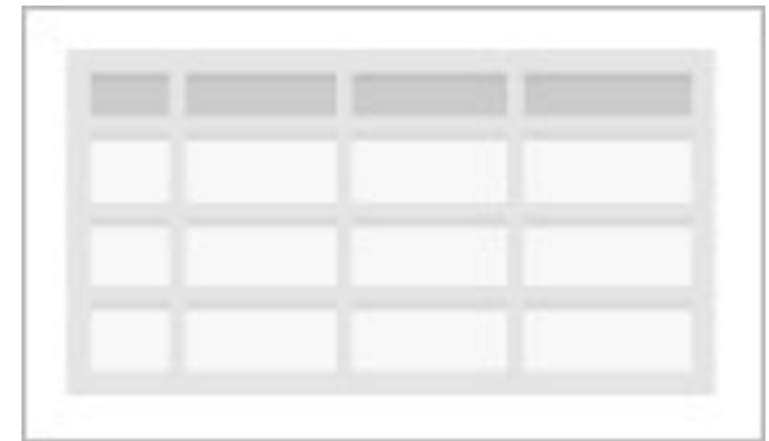
...the sequence is the label

Sequence Tables



merge

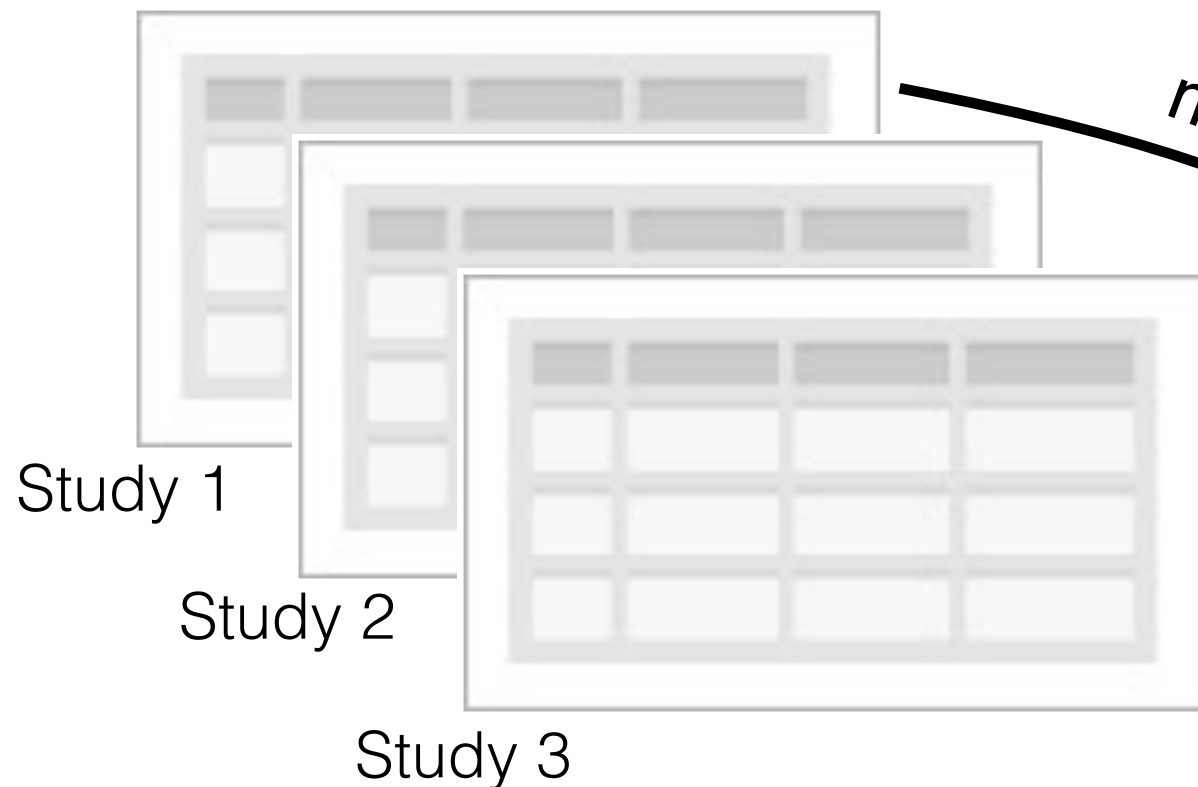
Cross-study comparison



Eliminates need for joint reprocessing of raw data.

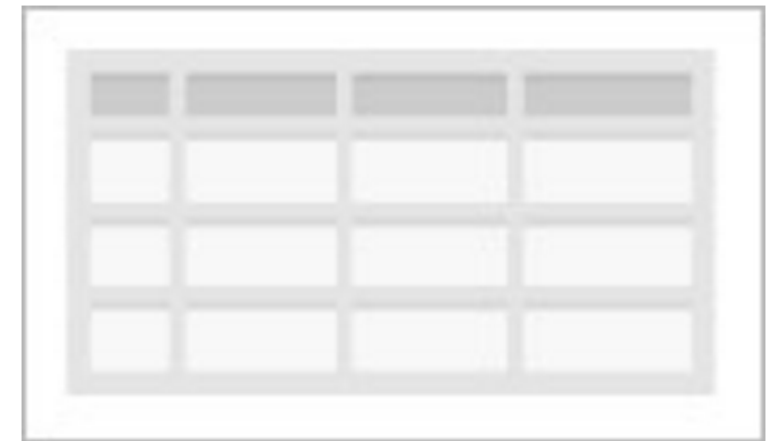
...the sequence is the label

Sequence Tables



merge

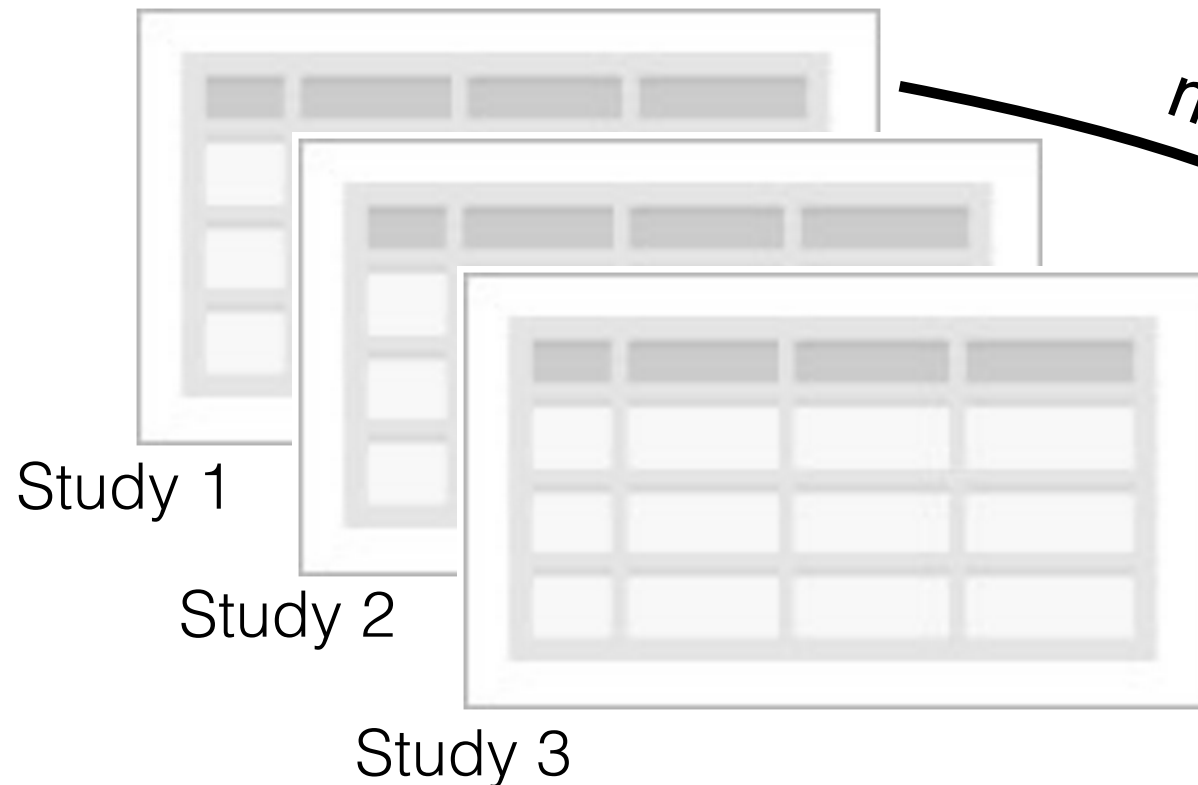
Cross-study comparison



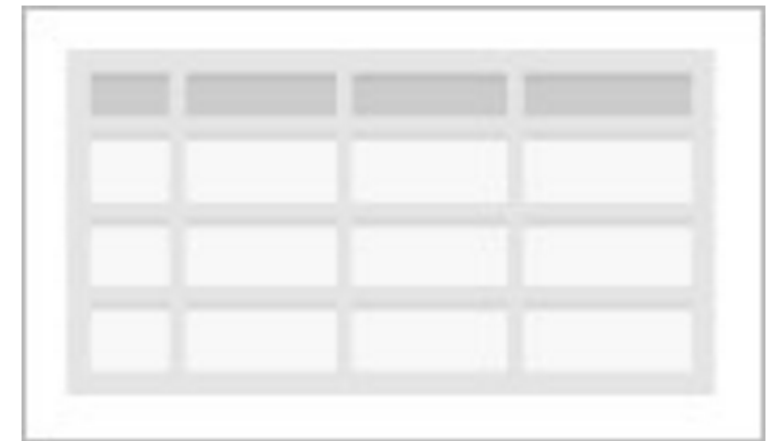
Eliminates need for joint reprocessing of raw data.
Continuous data integration. Unlimited dataset size.

...the sequence is the label

Sequence Tables

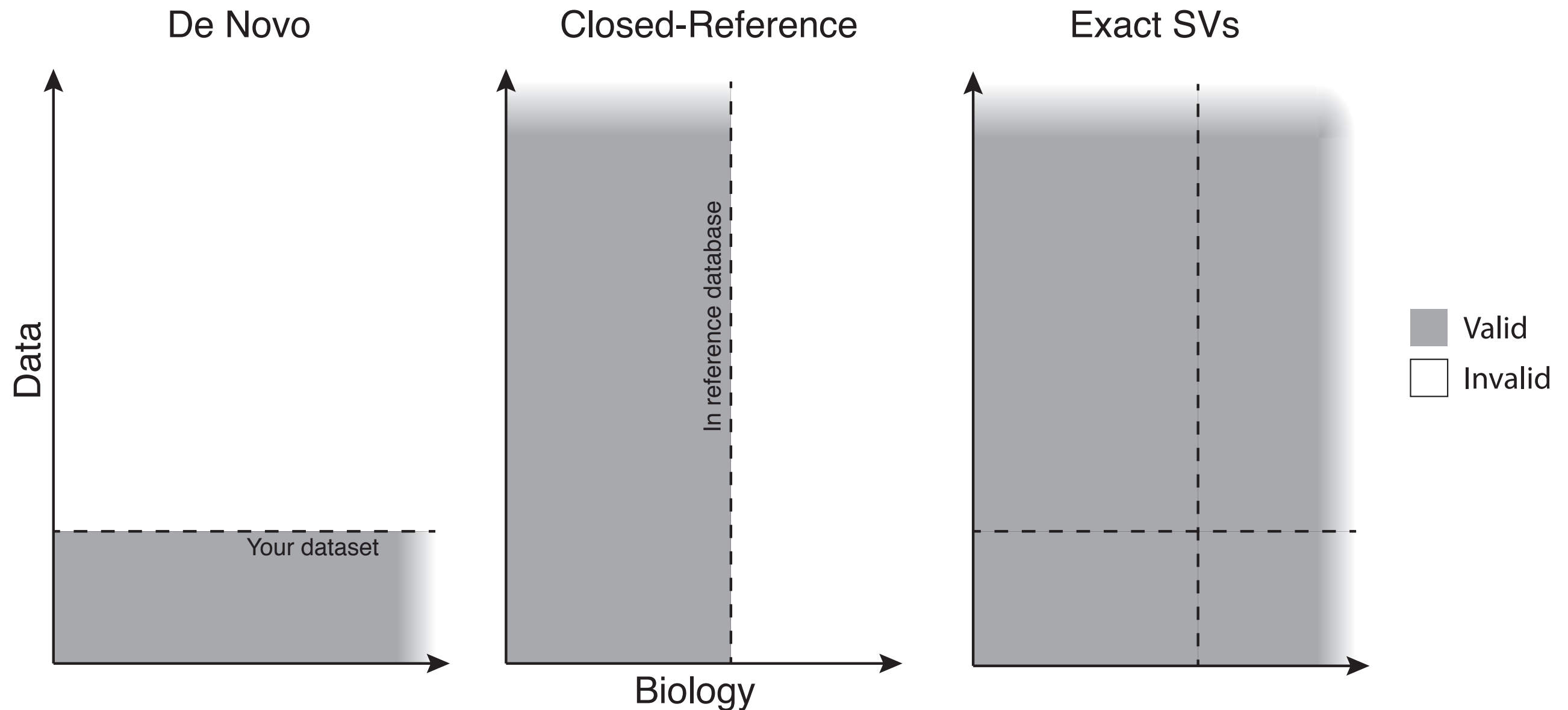


Cross-study comparison



Eliminates need for joint reprocessing of raw data.
Continuous data integration. Unlimited dataset size.
You in 2020 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.”

A rose by another name...

Amplicon Sequence Variants (ASVs)

- Needham et al. 2017
- Callahan et al. 2017

Exact Sequence Variants (ESVs)

- Callahan et al. 2017 (by accident)

sub-OTUs (sOTUs)

- Amir et al. 2017

Zero radius OTUs (zOTUs)

- Edgar 2017

Haplotypes, oligotypes, ...

A rose by another name...

Amplicon Sequence Variants (ASVs)

- Needham et al. 2017
- Callahan et al. 2017

Exact Sequence Variants (ESVs)

- Callahan et al. 2017 (by accident)

sub-OTUs (sOTUs)

- Amir et al. 2017

Zero radius OTUs (zOTUs)

- Edgar 2017

Haplotypes, oligotypes, ...

All the same thing! All the same (qualitative) benefits!

Exact Sequence Variants...

Precise
Tractable
Reproducible
Comprehensive

Callahan, et al. ISMEJ, 2017.

Exact Sequence Variants...

NATURE | ARTICLE **OPEN**

A communal catalogue reveals Earth's multiscale microbial diversity

Luke R. Thompson, Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, *et al.*

Nature (2017) | doi:10.1038/nature24621

Received 13 March 2017 | Accepted 10 October 2017 | Published online 01 November 2017

**Precise
Tractable
Reproducible
Comprehensive**

Callahan, et al. ISMEJ, 2017.

... are replacing OTUs

NATURE | ARTICLE **OPEN**

A communal catalogue reveals Earth's multiscale microbial diversity

Luke R. Thompson, Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, *et al.*

Nature (2017) | doi:10.1038/nature24621

Received 13 March 2017 | Accepted 10 October 2017 | Published online 01 November 2017

*“the use of **exact sequences instead of clustered operational taxonomic units**, enable bacterial and archaeal ribosomal RNA gene sequences to be followed across multiple studies and allow us to explore patterns of diversity at an **unprecedented scale**...for many sample types, especially plant-associated and free-living communities, one-third of reads or more **could not be mapped** to [closed-reference OTUs]...Because exact sequences are stable identifiers, unlike OTUs, they can be compared to any 16S rRNA or genomic database now and in the future, thereby promoting **reusability** [22]... An advantage of using exact sequences is that they enable us to observe and analyse microbial distribution patterns at **finer resolution** than is possible with traditional OTUs...”*

**Precise
Tractable
Reproducible
Comprehensive**

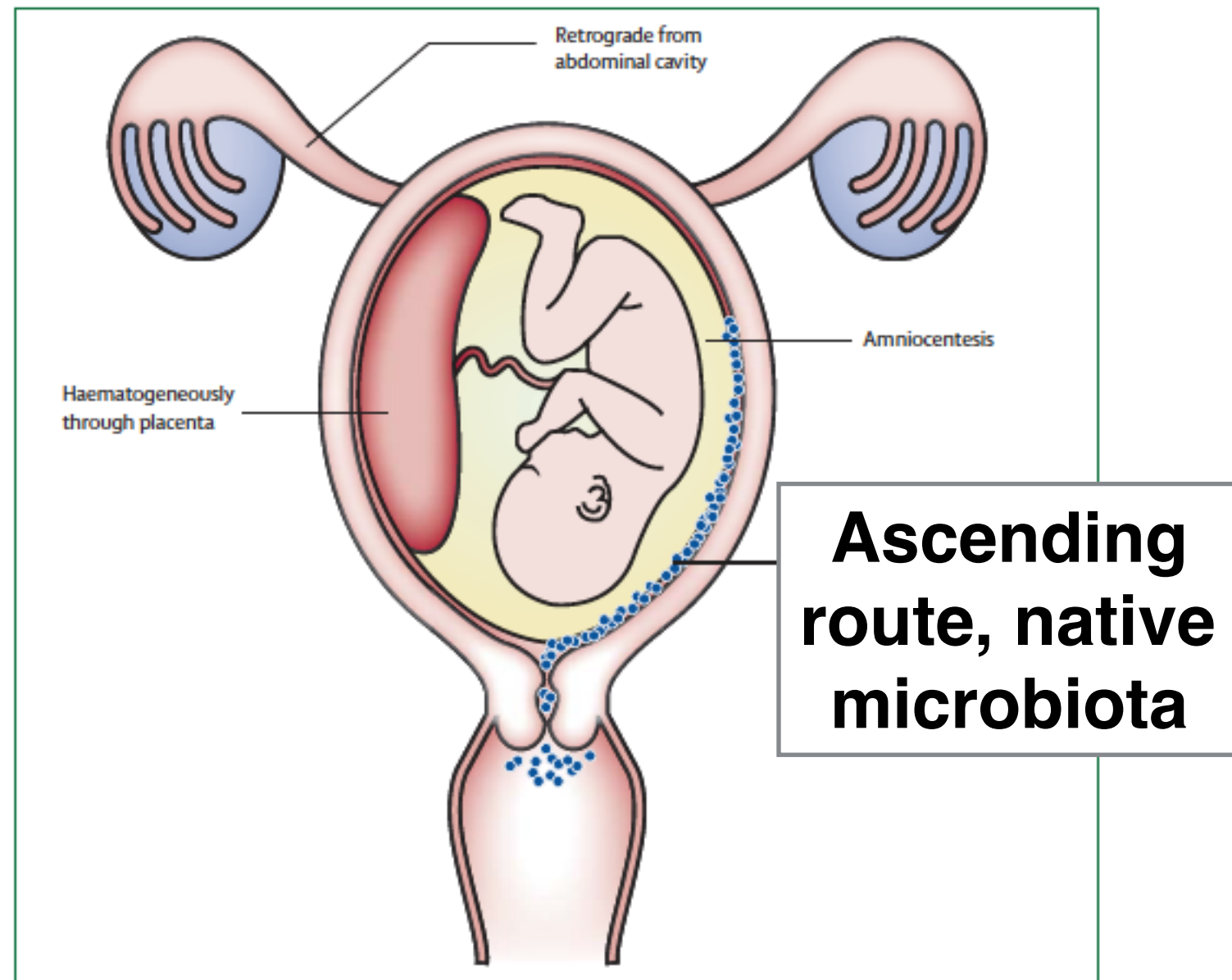
Callahan, et al. ISMEJ, 2017.

Resolution: Preterm Birth

Preterm birth is the leading cause of child mortality in the US and worldwide.

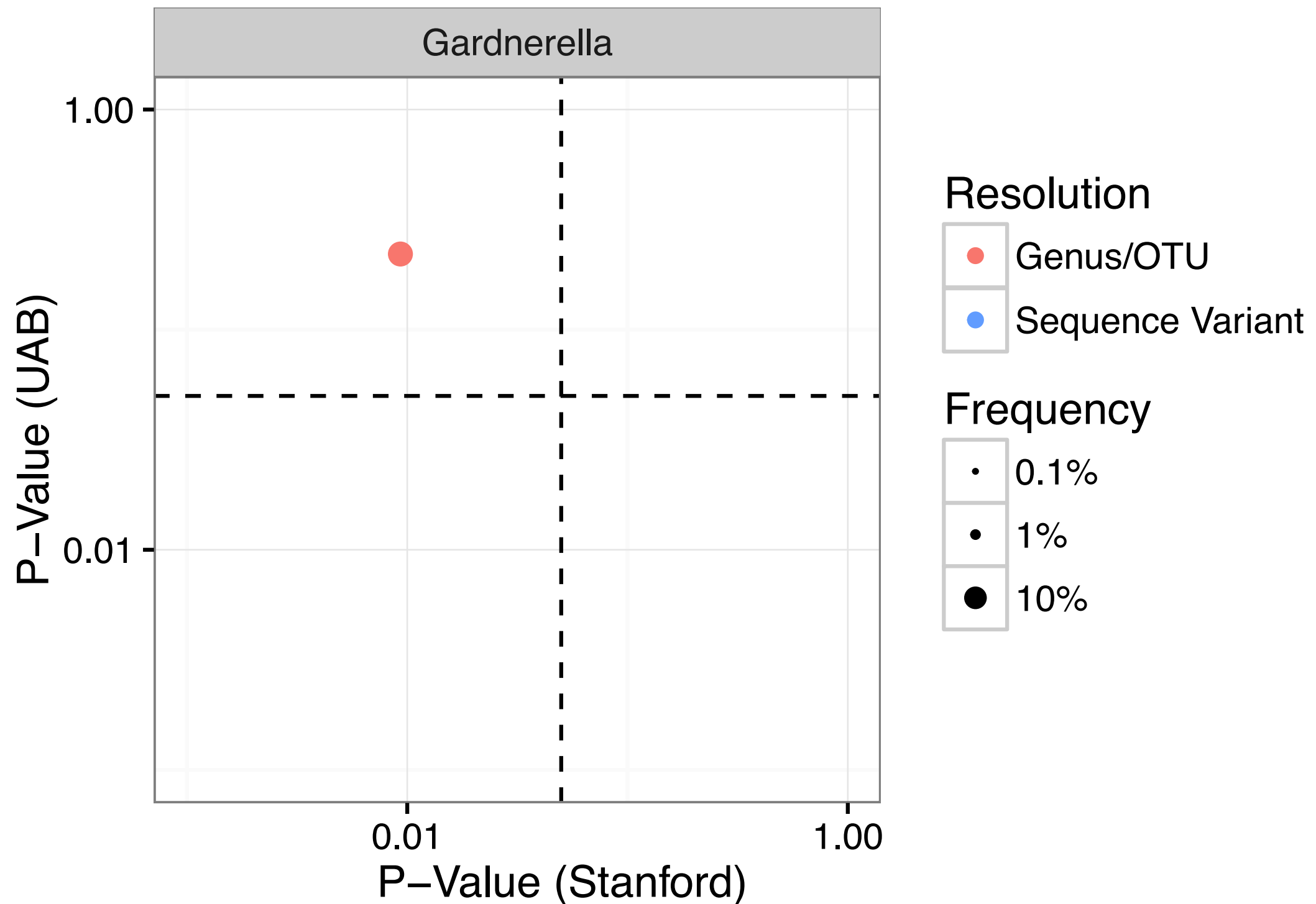
Intrauterine infection is the most common mechanism, estimated to occur in **25-40%** of cases. This may be an underestimate.

Bacterial vaginosis is a known risk factor (2-4x). Perhaps **periodontitis** as well.

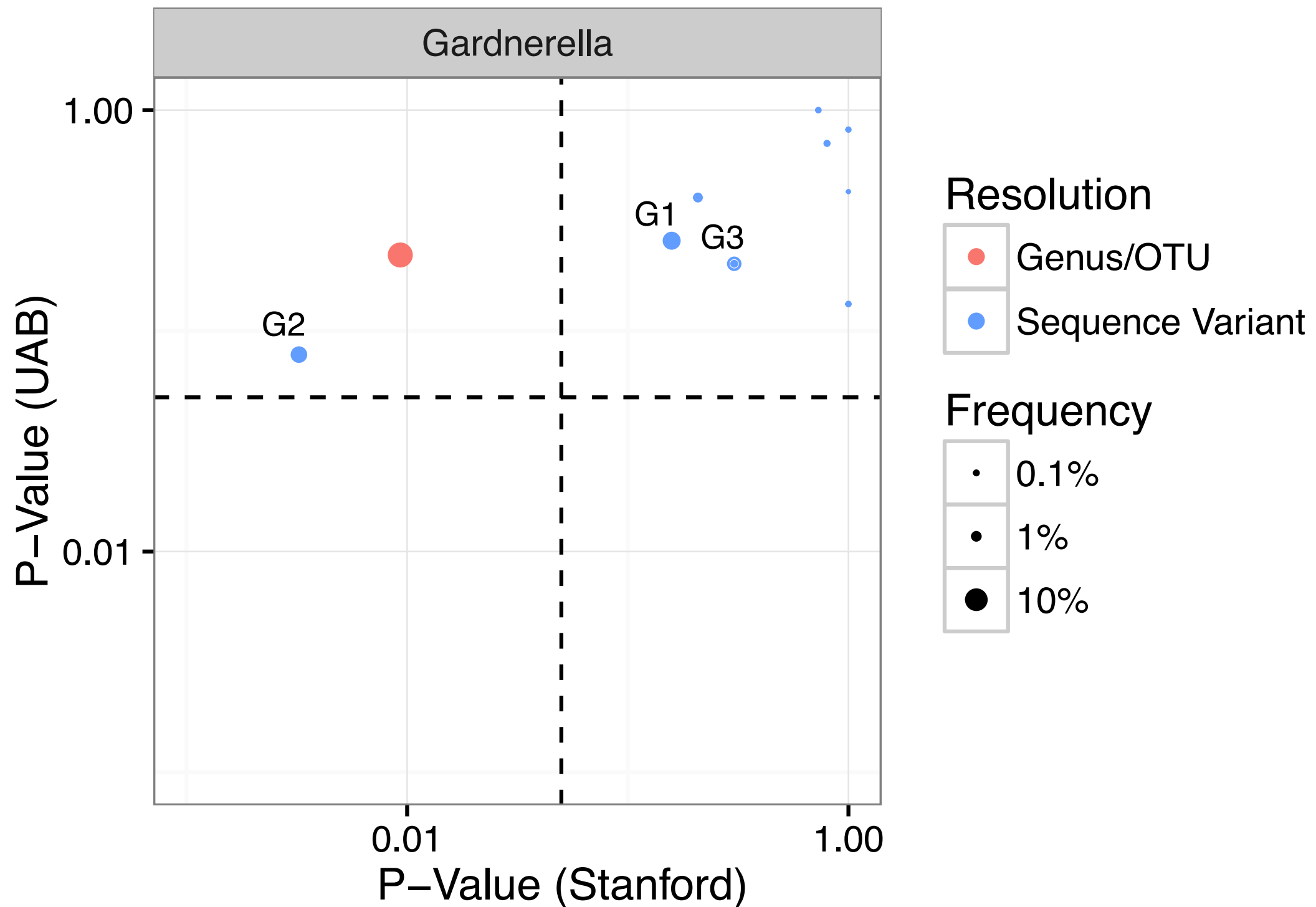


Potential routes of intrauterine infection

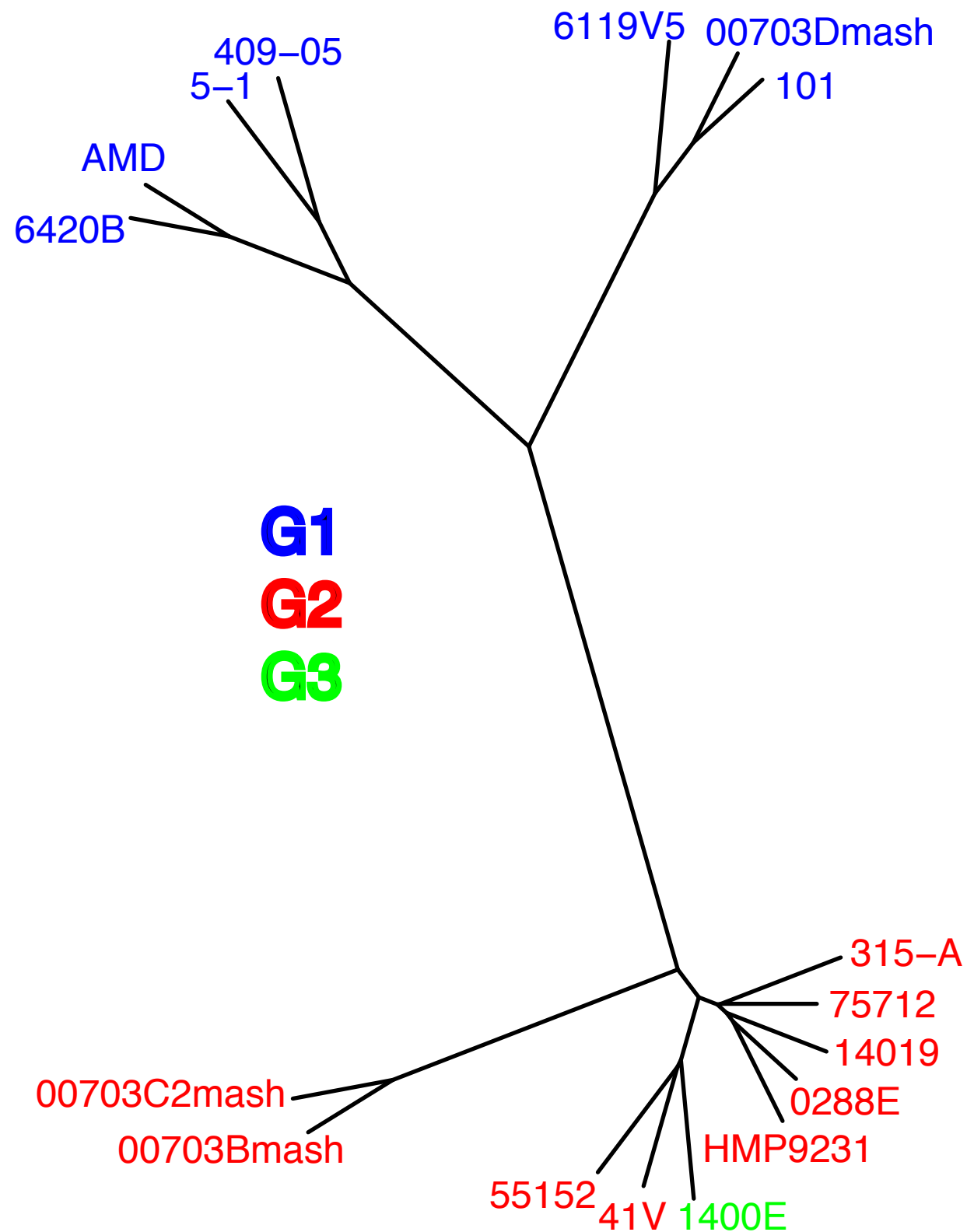
Resolution: Preterm Birth



Resolution: Preterm Birth

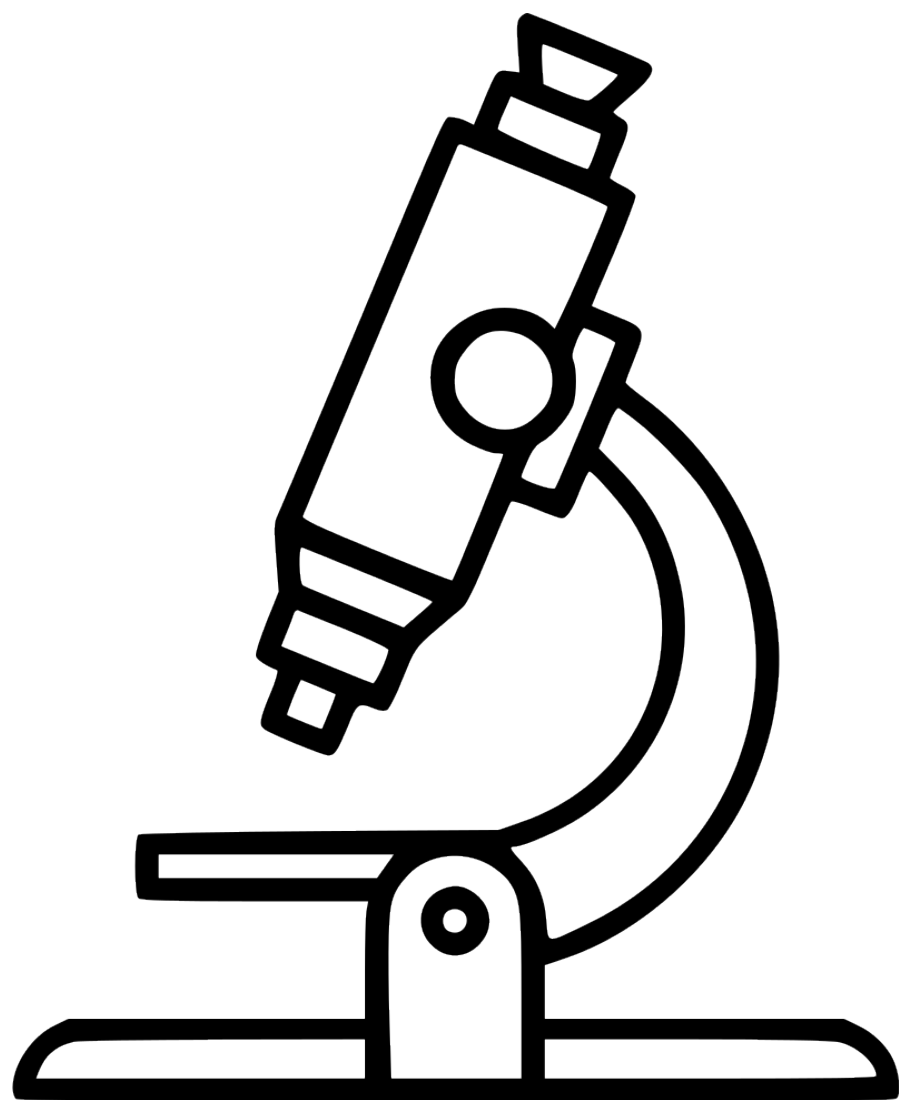


Resolution: Preterm Birth



Tree from
whole genomes

Reproducible Measurements



?

=

