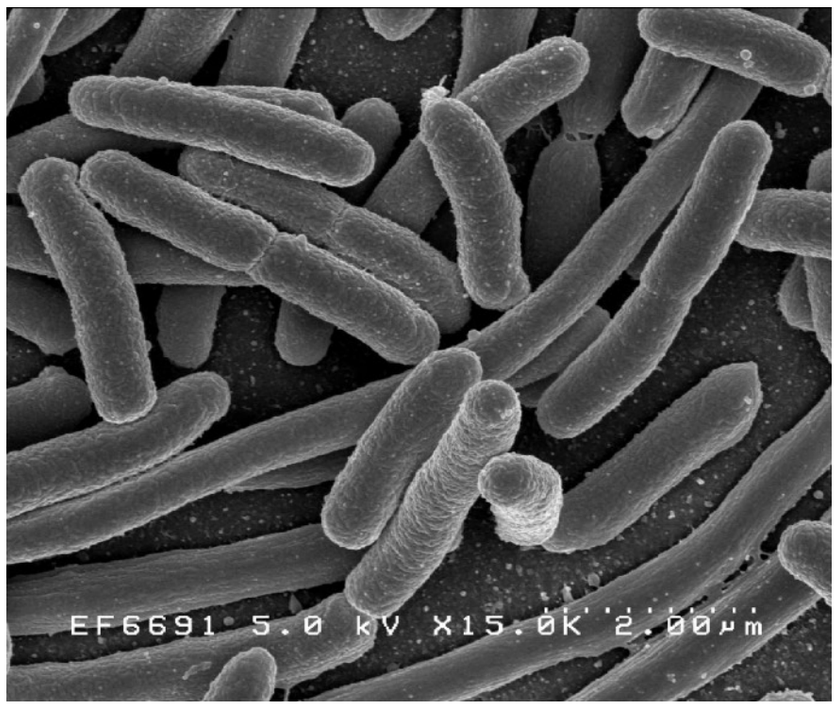
Exact Sequences from Marker-gene sequencing

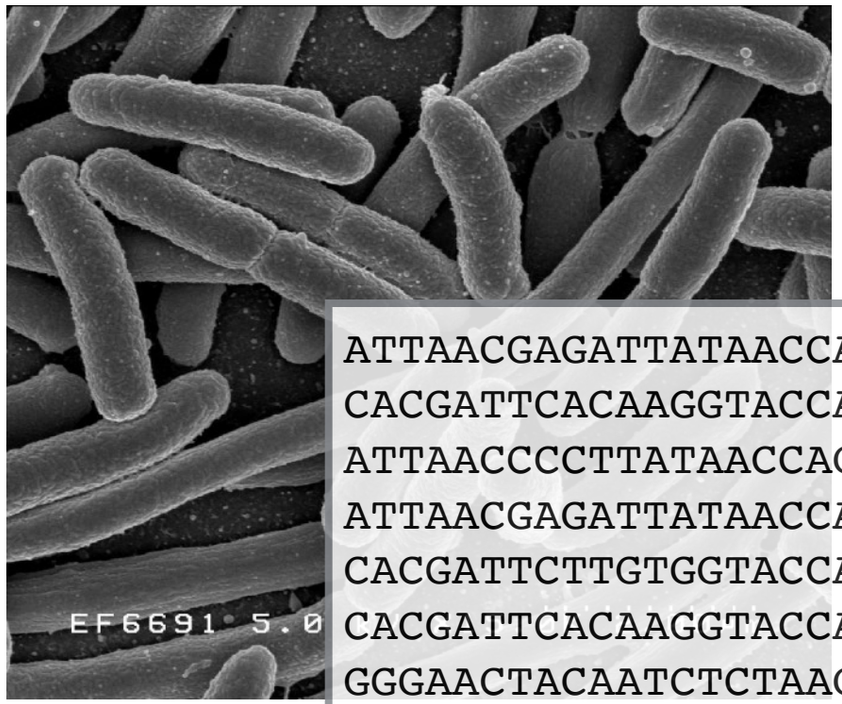
A Microbial Census

Marker-gene or Metagenomics Sequencing ***(MGS)***



A Microbial Census

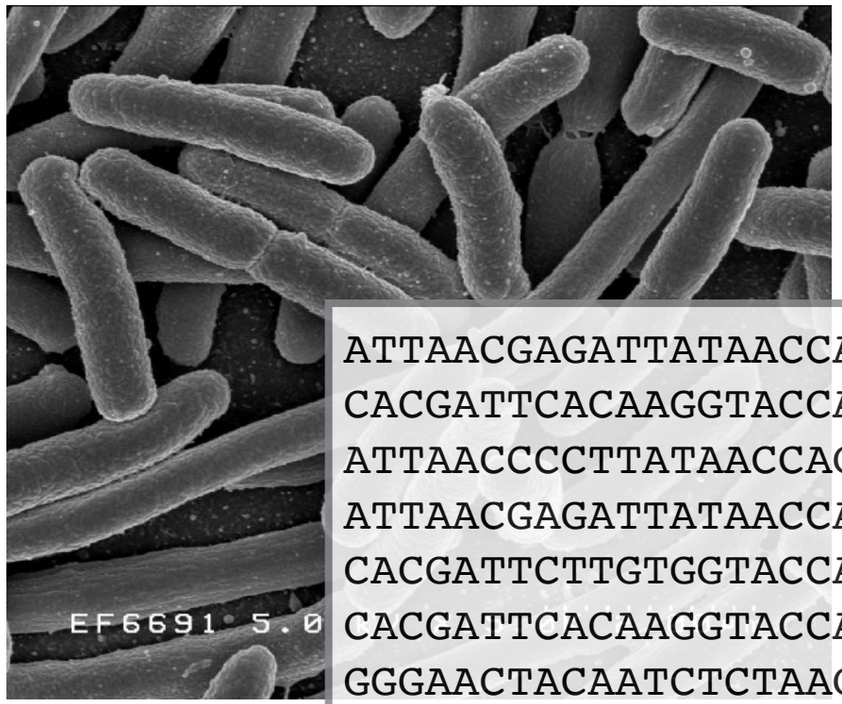
Marker-gene or Metagenomics Sequencing ***(MGS)***




```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGAACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
```

A Microbial Census

Marker-gene or Metagenomics Sequencing ***(MGS)***

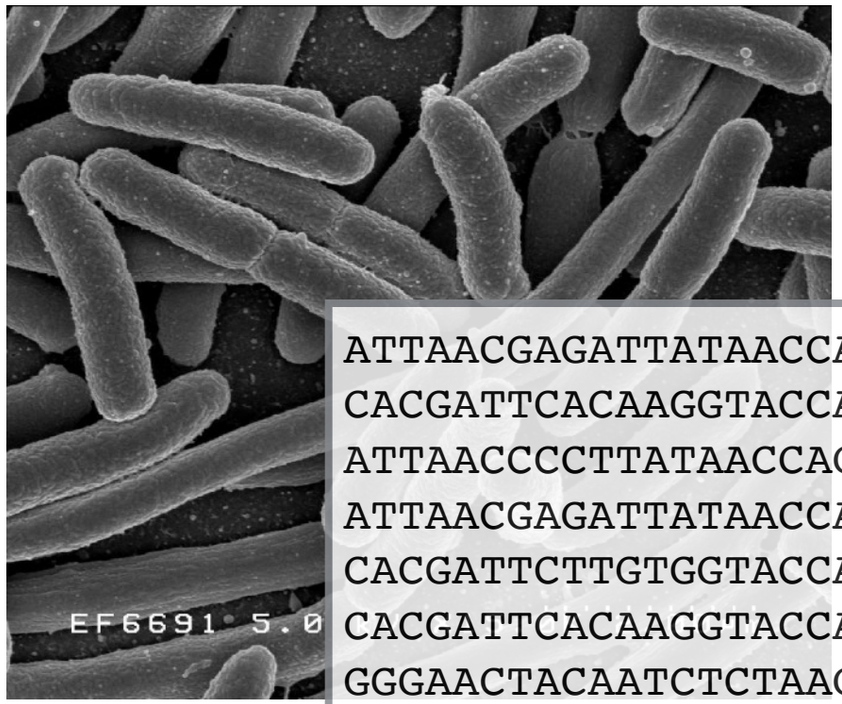



```
ATTAACGAGATTATAACCAGAGTACGAATACCGAAC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
ATTAACCCCTTATAACCAGAGTACGAATACCGAACA
ATTAACGAGATTATAACCAGAGAGAGAATACCGAAC
CACGATTCTTGTGGTACCACAAGGTAACATAGCTCC
CACGATTCACAAGGTACCACAAGGTAACATAGCTCC
GGGAACTACAATCTCTAAGGTGAAGTCTCAGTCTAT
ATTAACGAGATTATAACCAGA
CACGATTCACAAGGTACCACA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

Marker-gene or Metagenomics Sequencing ***(MGS)***




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

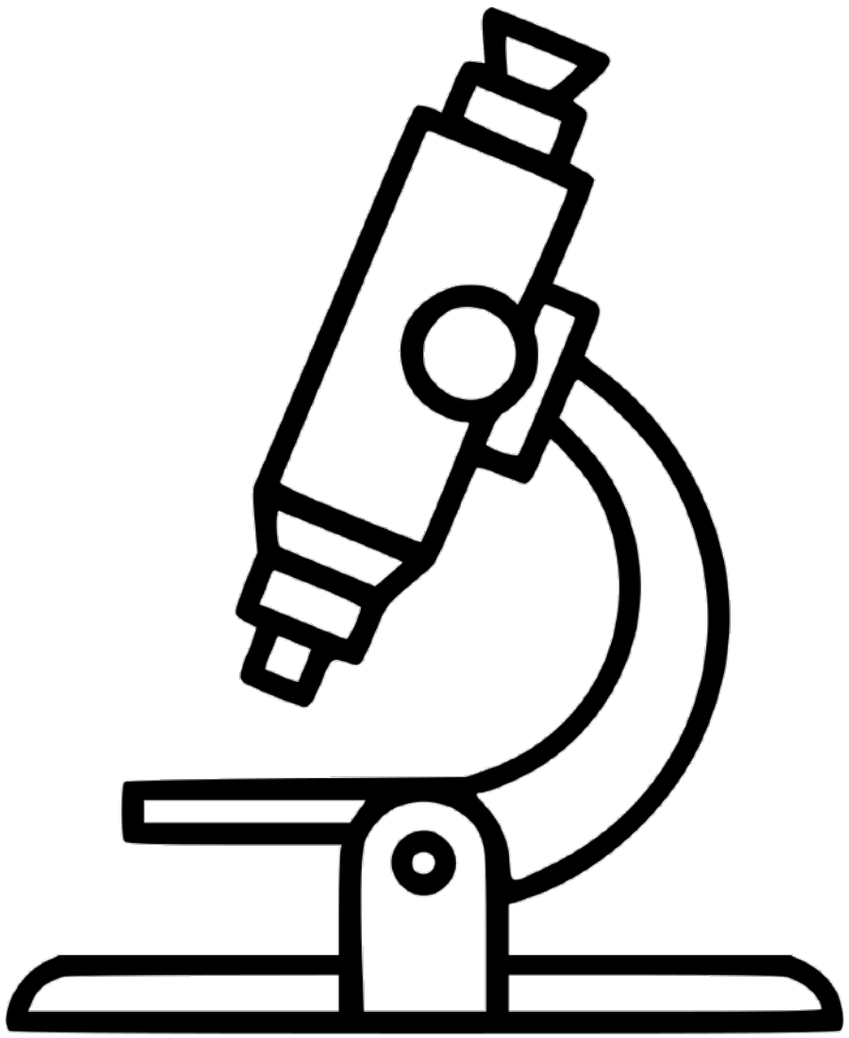
Lactobacillus crispatus	1300	5	0	882	596
Ureaplasma urealytica	15	0	220	0	0
Gardnerella vaginalis	22	0	1	0	412
Prevotella intermedia	0	0	8	12	0
...	...	...	...	...	...

Inference

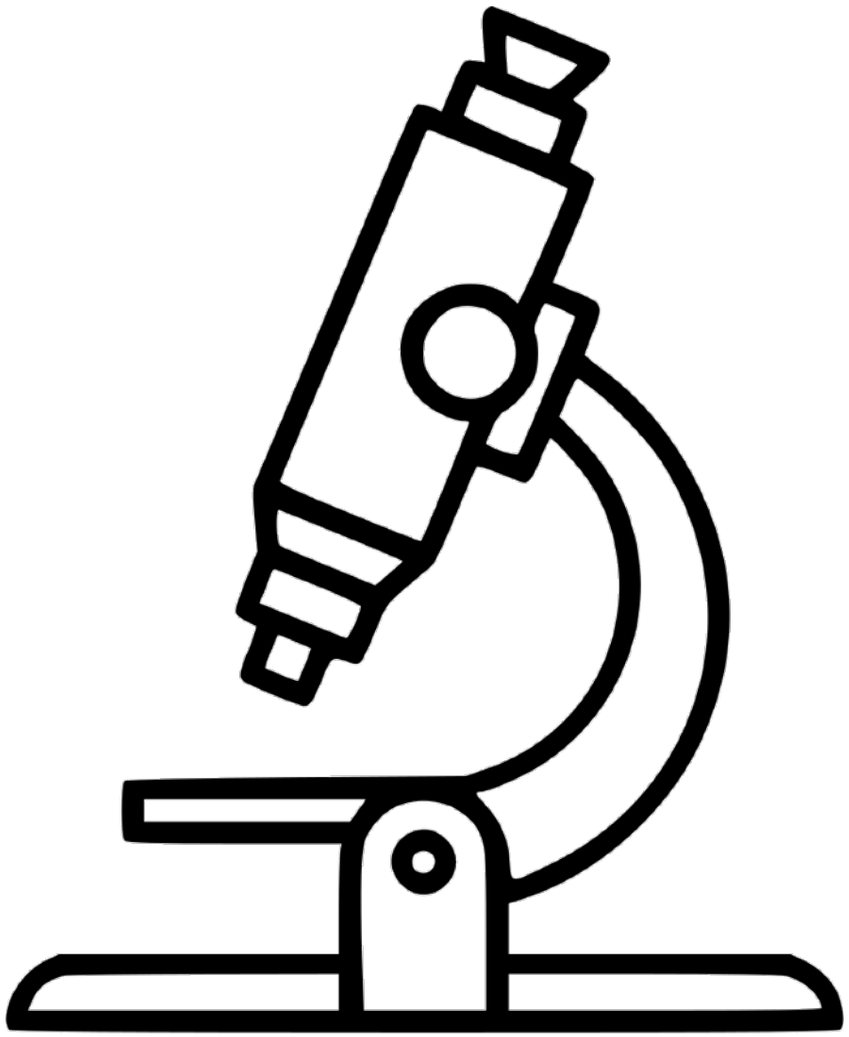
Visualization

Exploration

The MGS Microscope

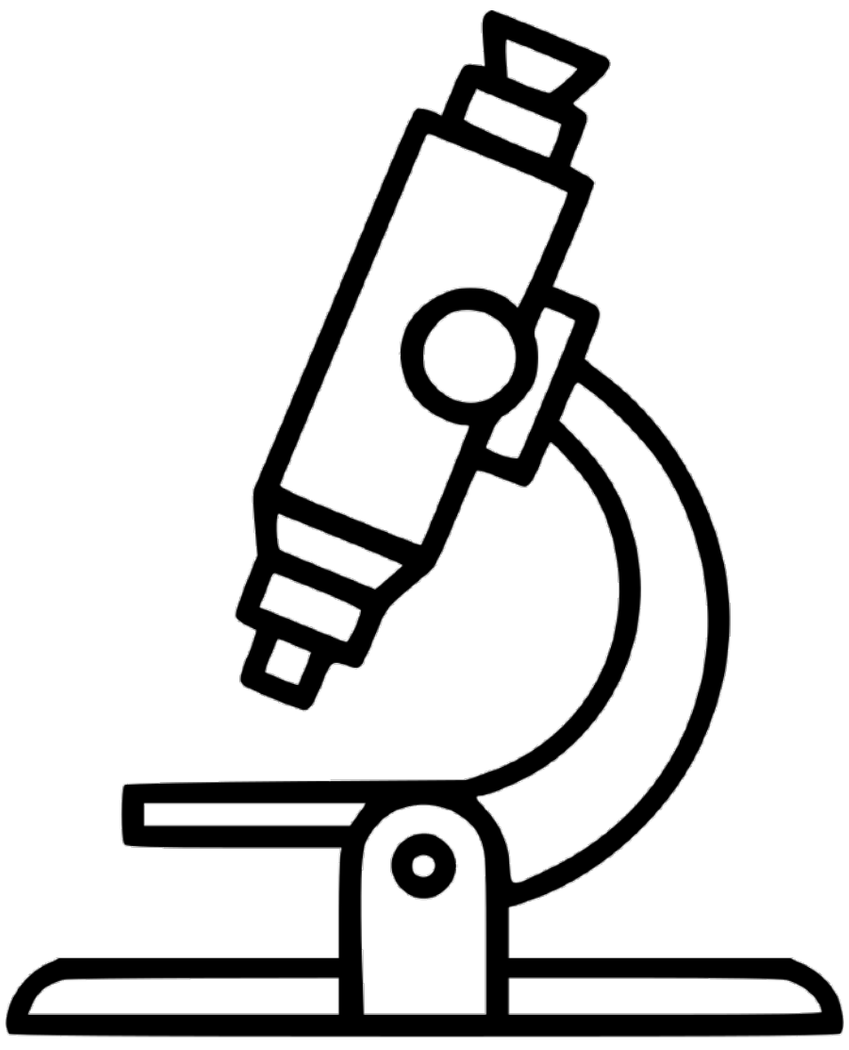


The MGS Microscope



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

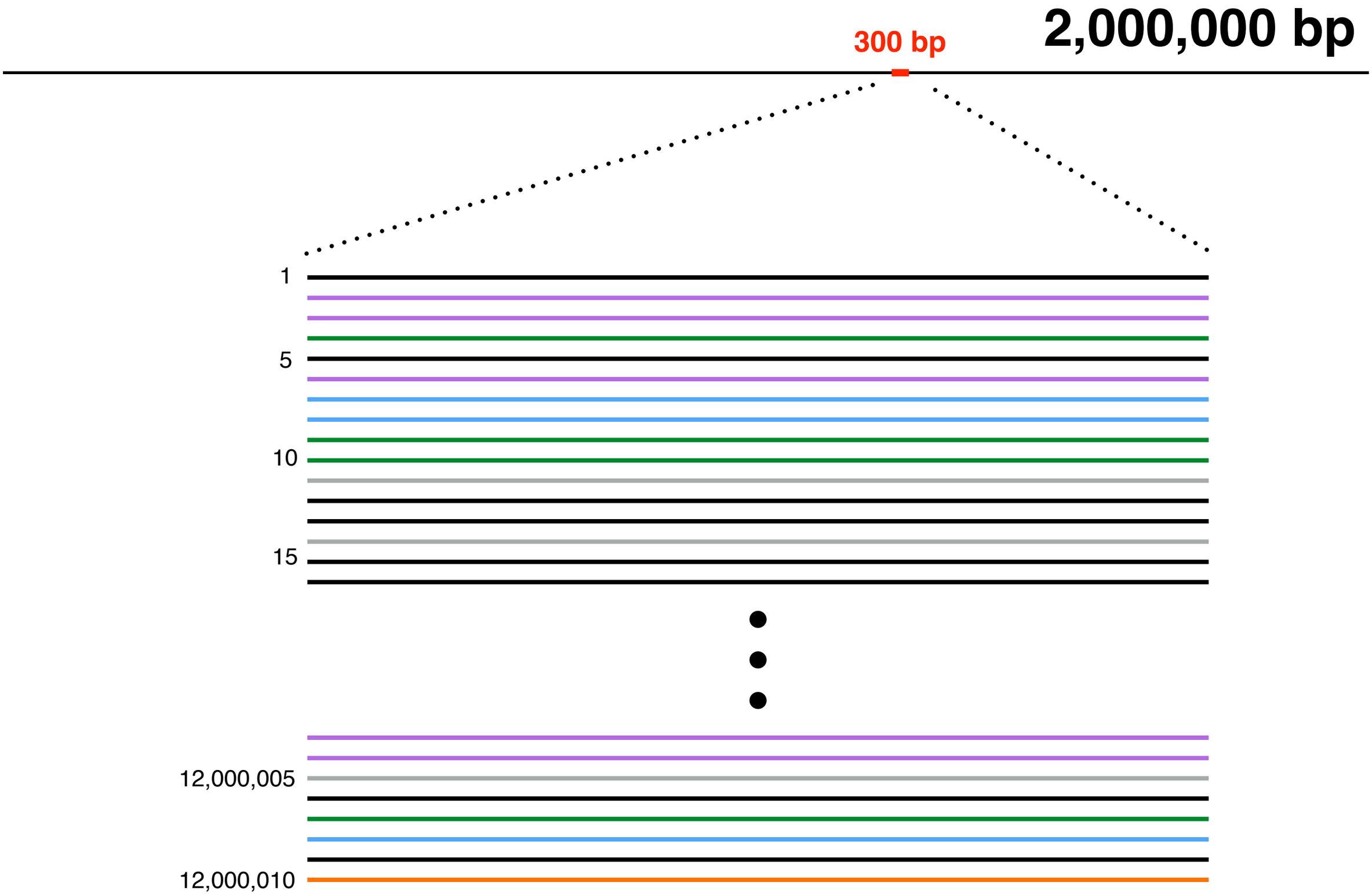
The MGS Microscope



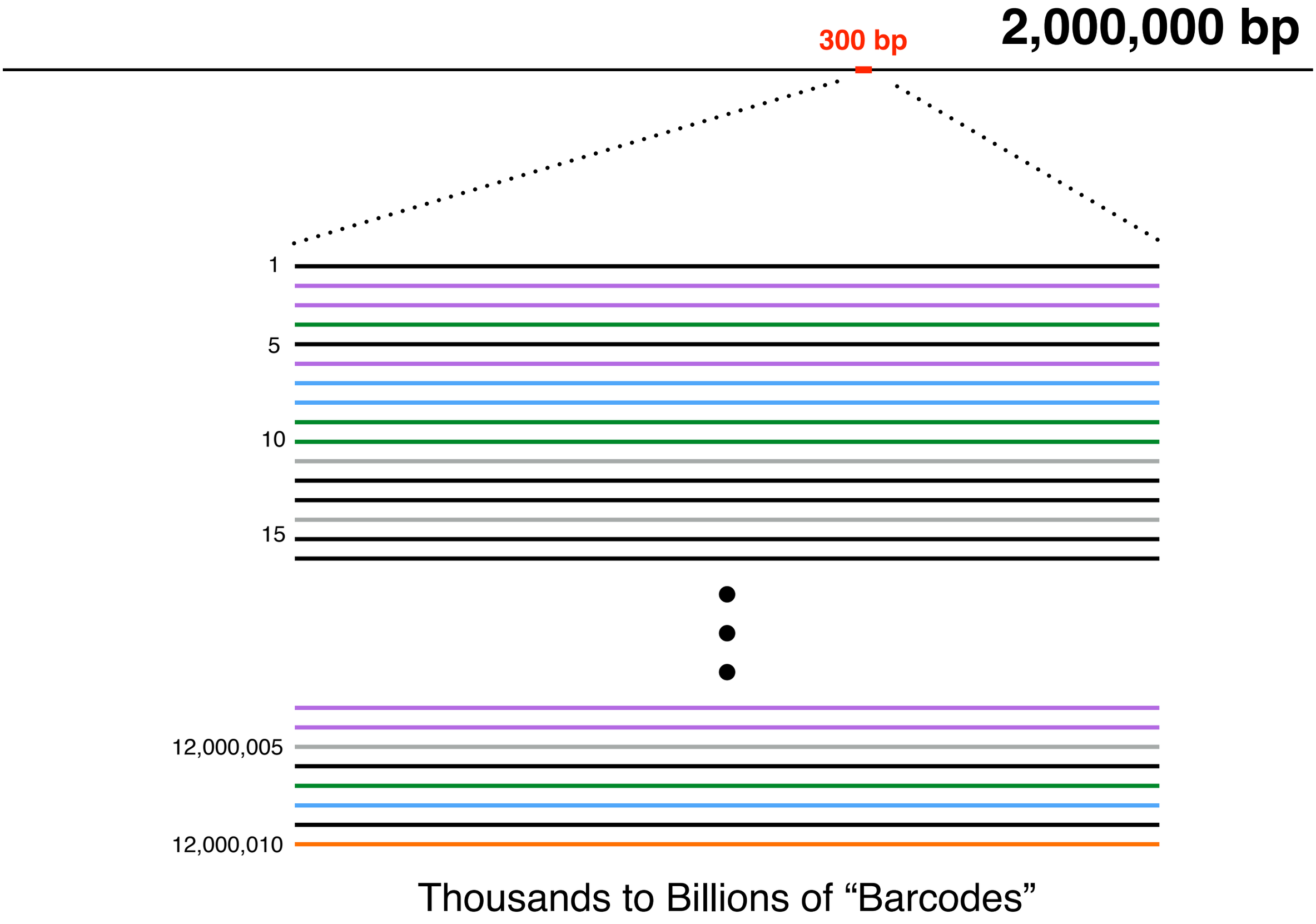
?

Marker-gene Sequencing

Marker-gene Sequencing

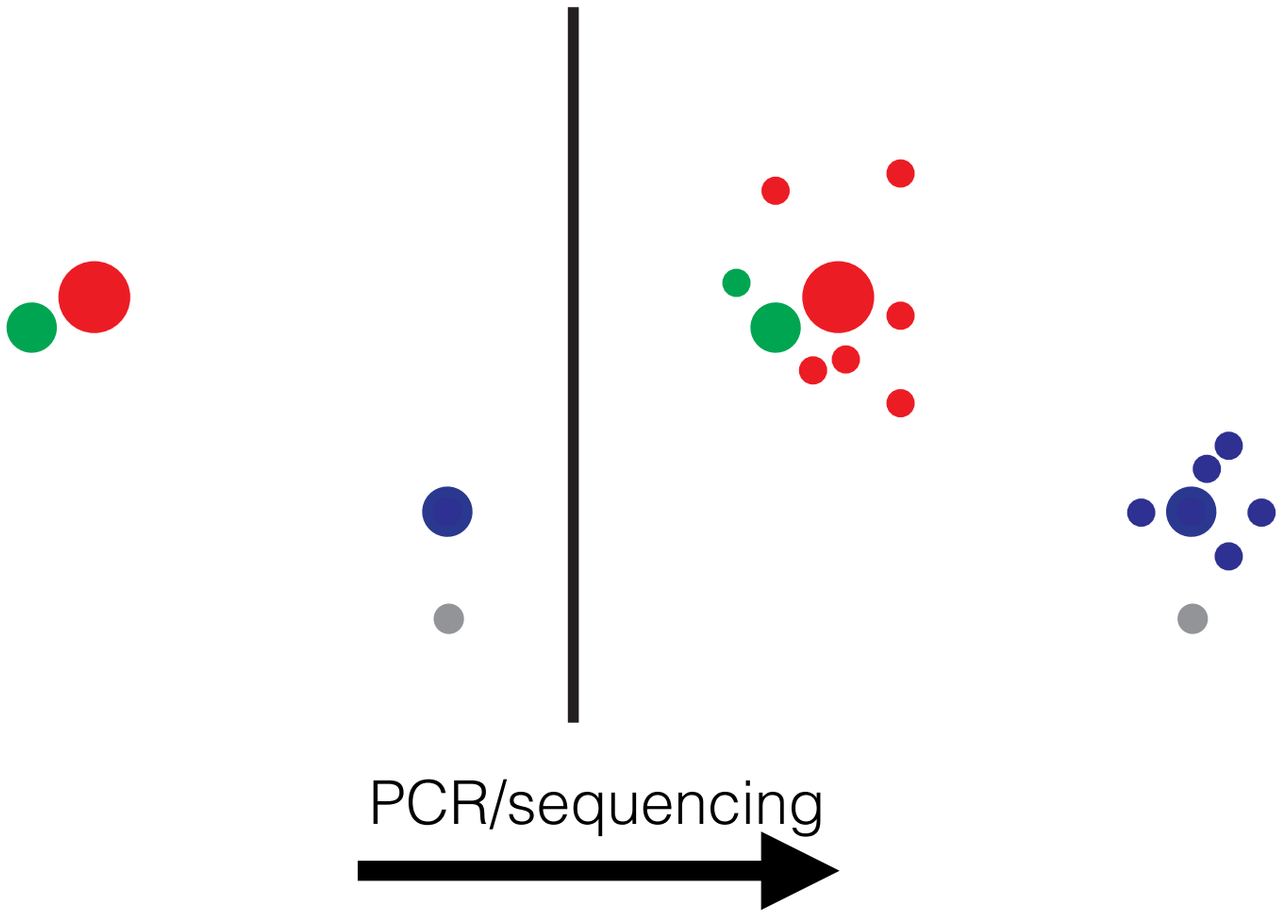


Marker-gene Sequencing



Sample Sequences

Amplicon Reads



PCR/sequencing

Sample Sequences

Amplicon Reads

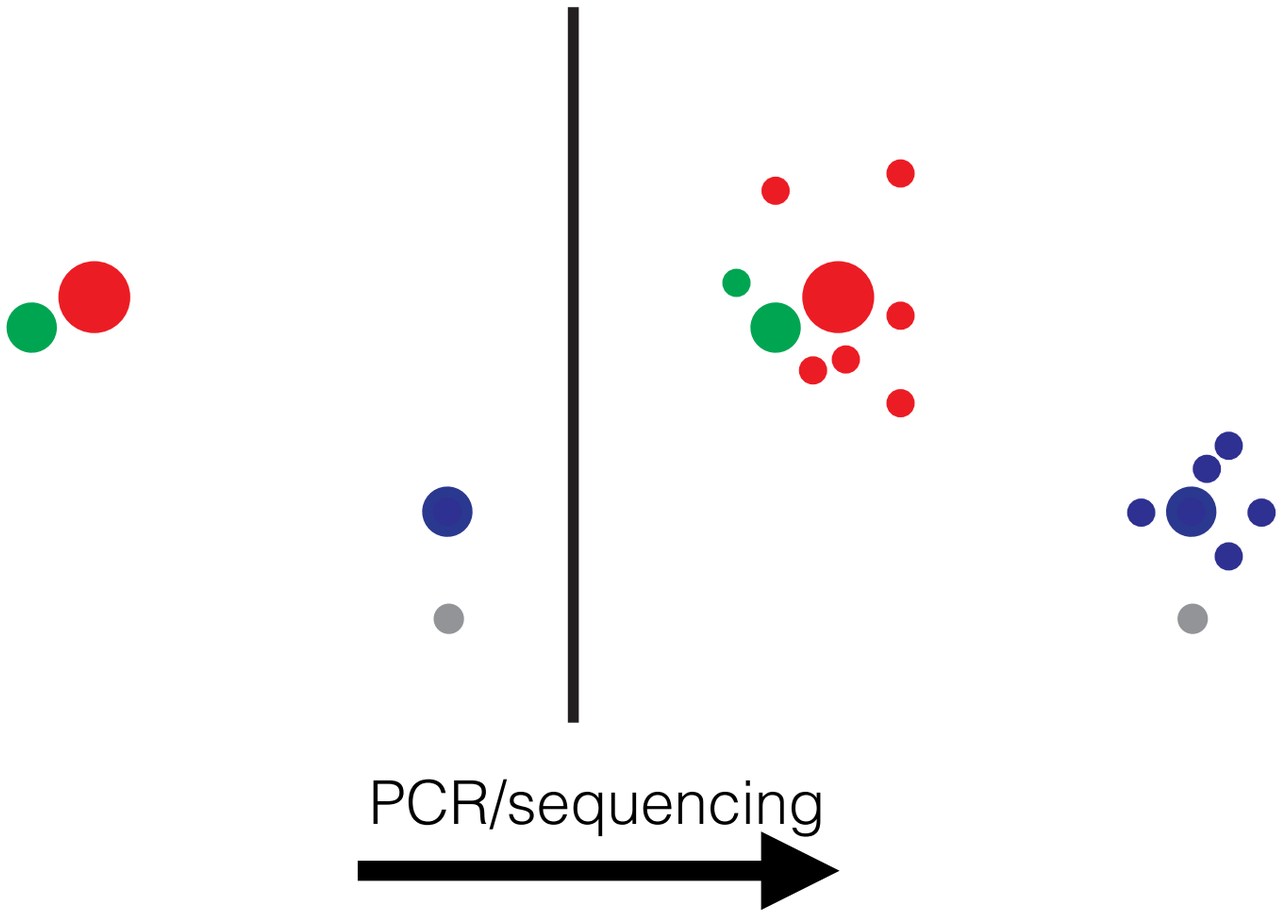
Operational Taxonomic Units

PCR/sequencing

Pick OTUs

Sample Sequences

Amplicon Reads



PCR/sequencing

Sample Sequences

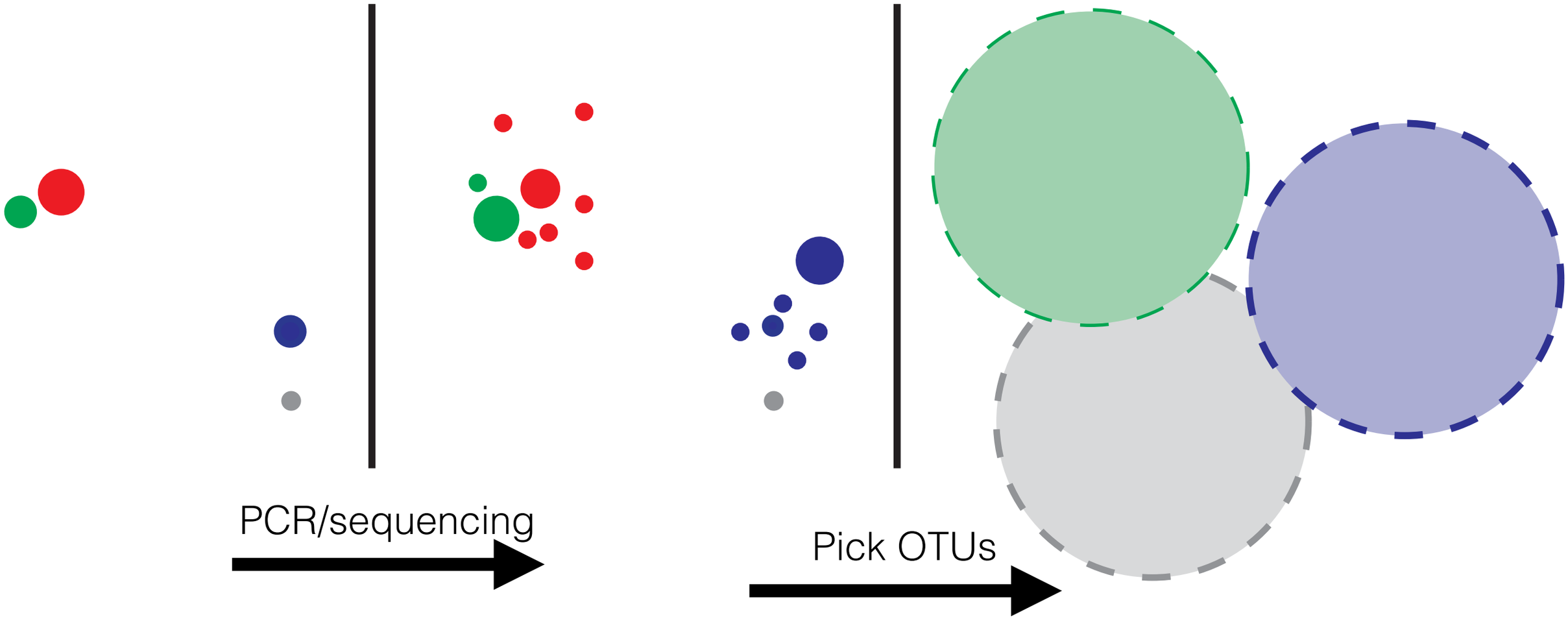
Amplicon Reads

PCR/sequencing

Sample Sequences

Amplicon Reads

<u>Different</u> 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!**

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

Different Operational Taxonomic Units

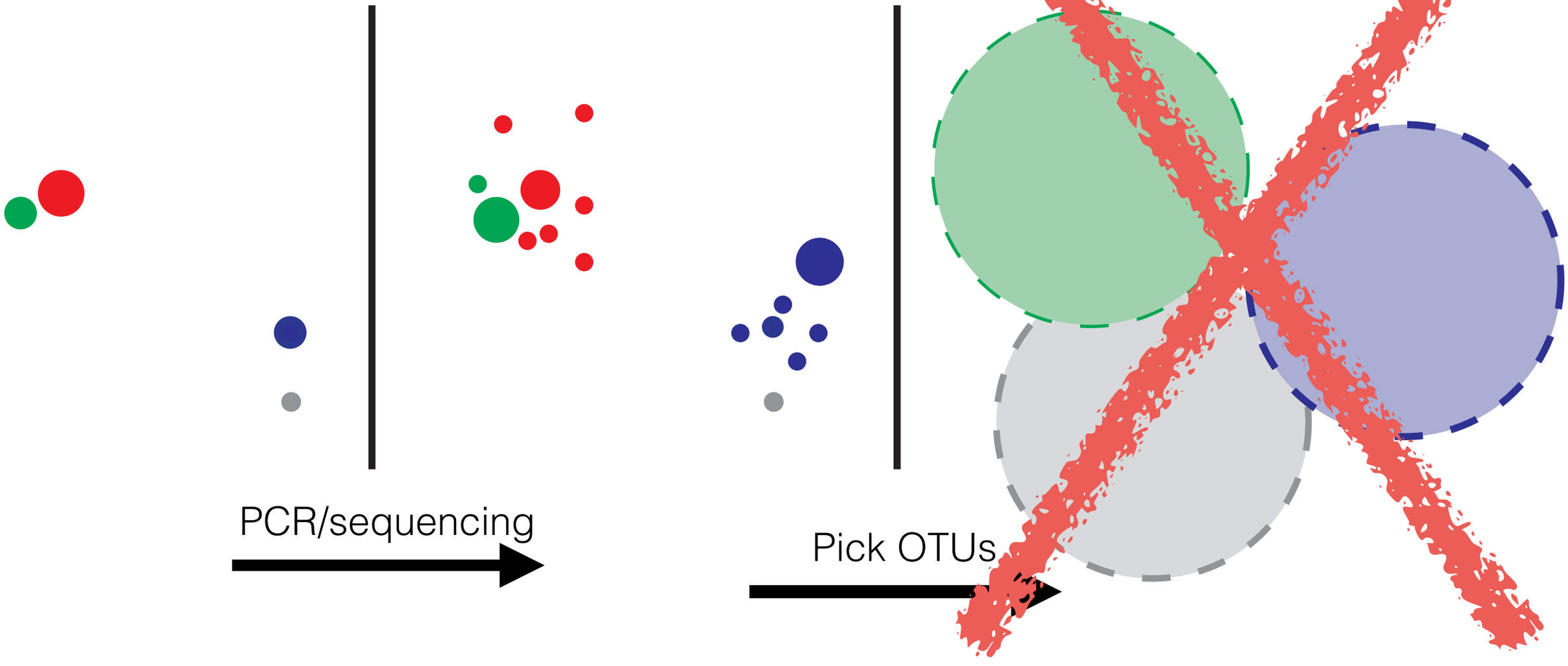
PCR/sequencing

Pick OTUs

Sample Sequences

Amplicon Reads

Different Operational Taxonomic Units



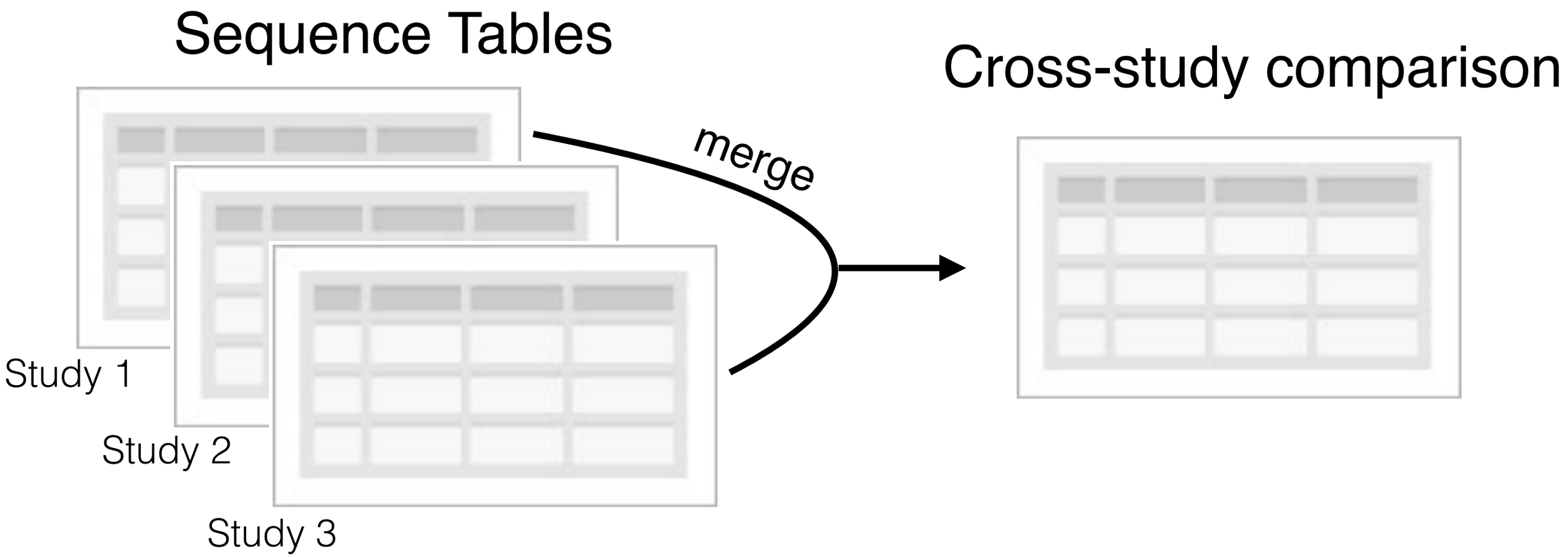
Exact Sequence Variants…

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

ATTAACGAGATTATAACCAGAGTACGAATA…

is consistent!

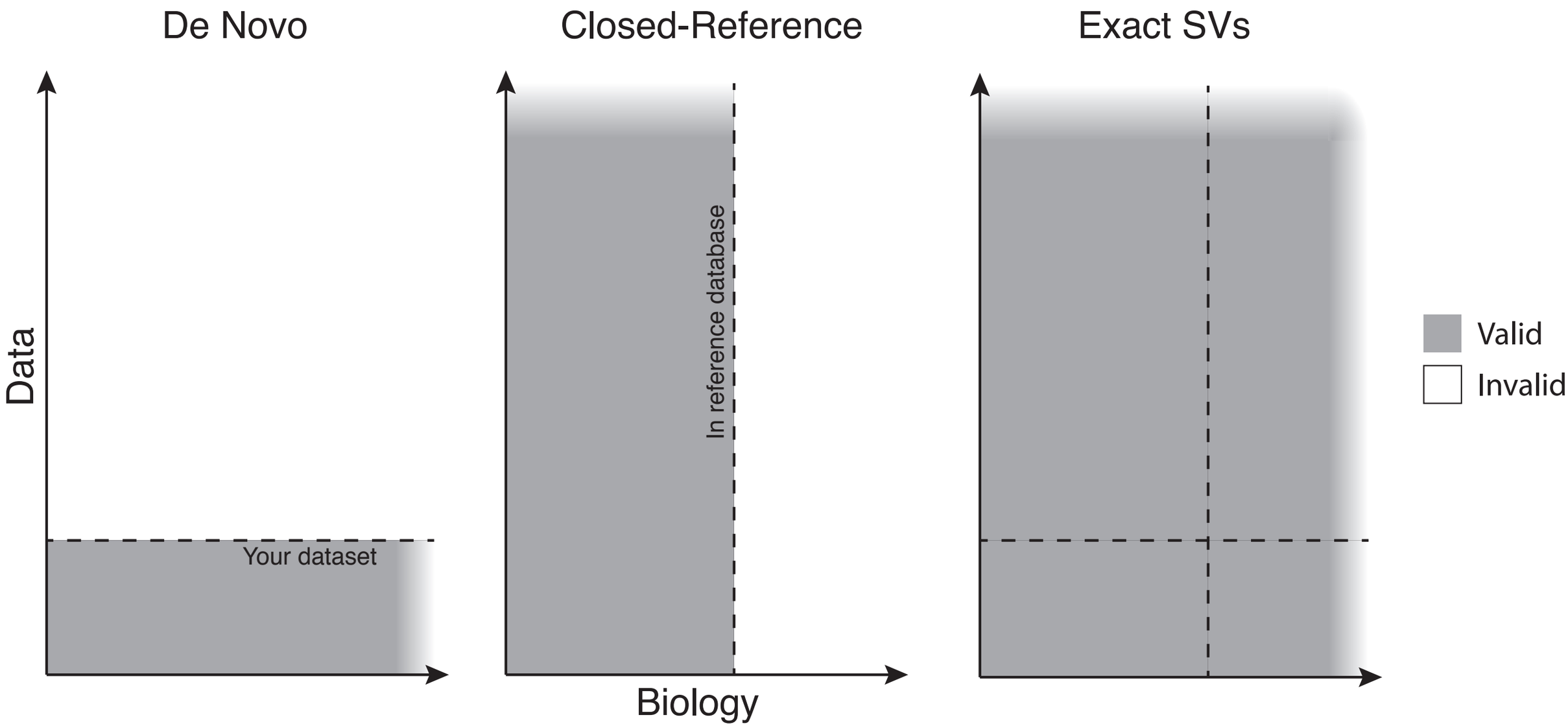
…the sequence is the label



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

Callahan, et al. ISMEJ, 2017.

…the sequence is the label



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

Callahan, et al. ISMEJ, 2017.