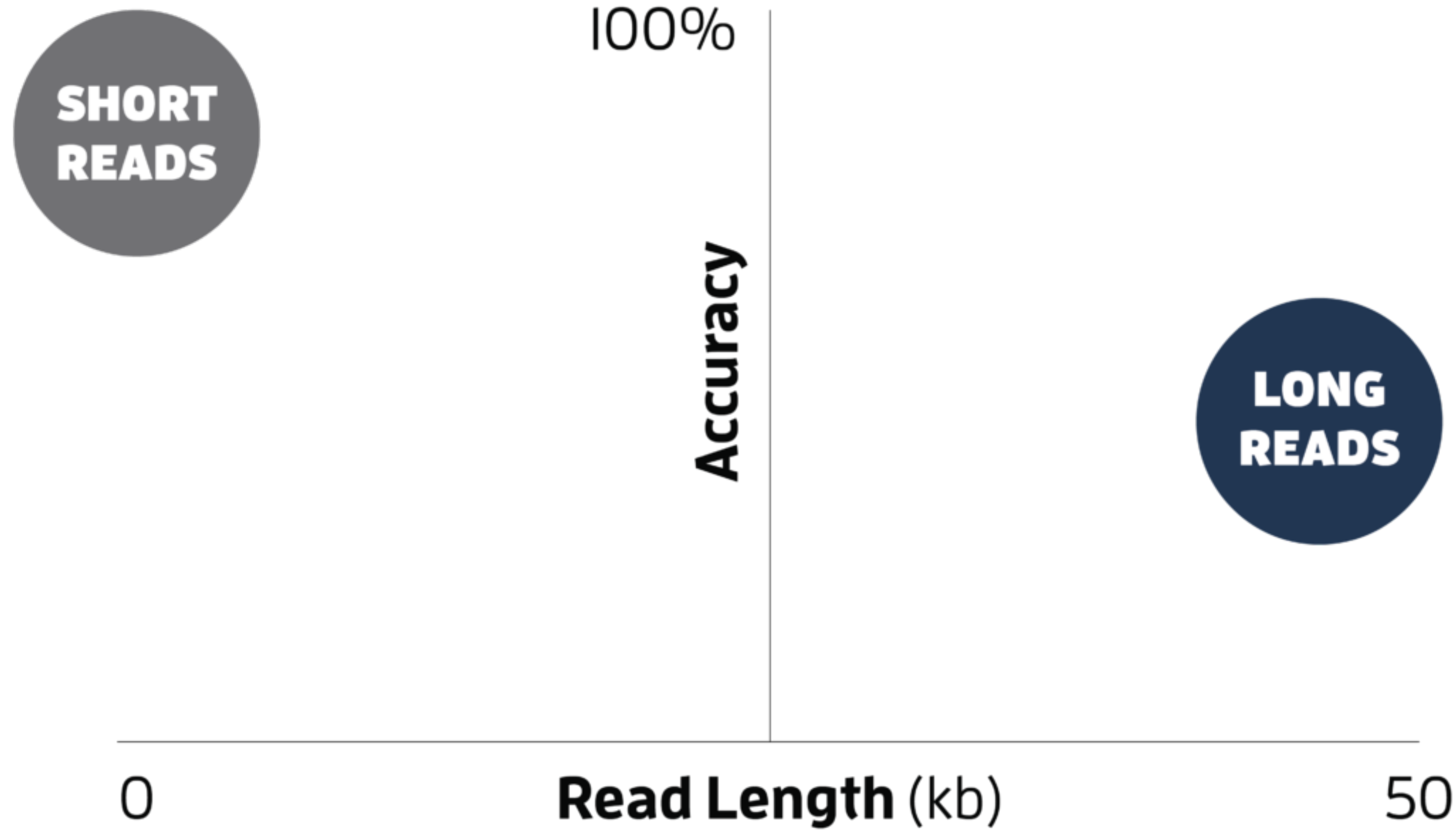
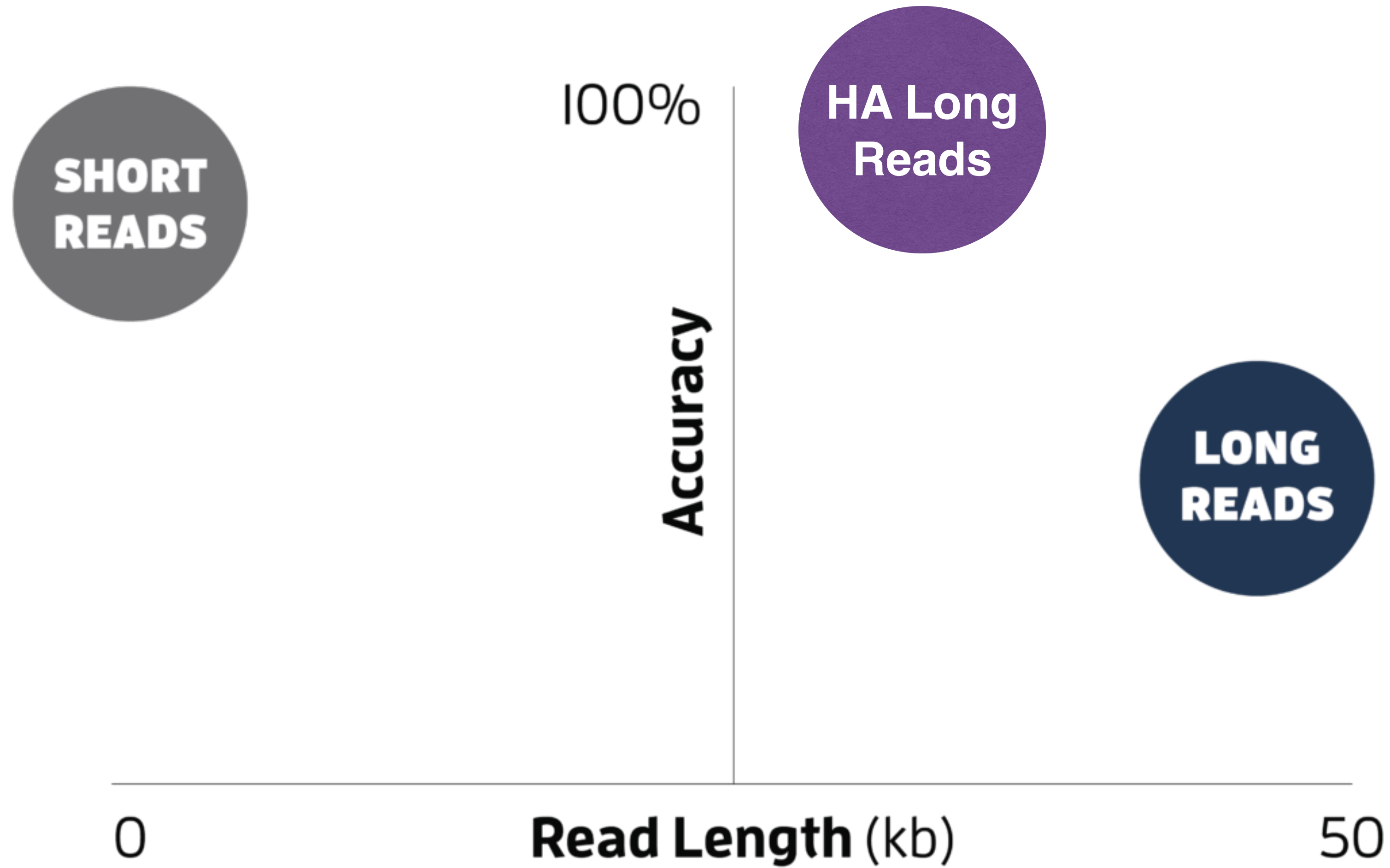


Maximizing Resolution with  
DADA2  
and Long-read Amplicon Sequencing

# Highly-Accurate Long Reads



# Highly-Accurate Long Reads



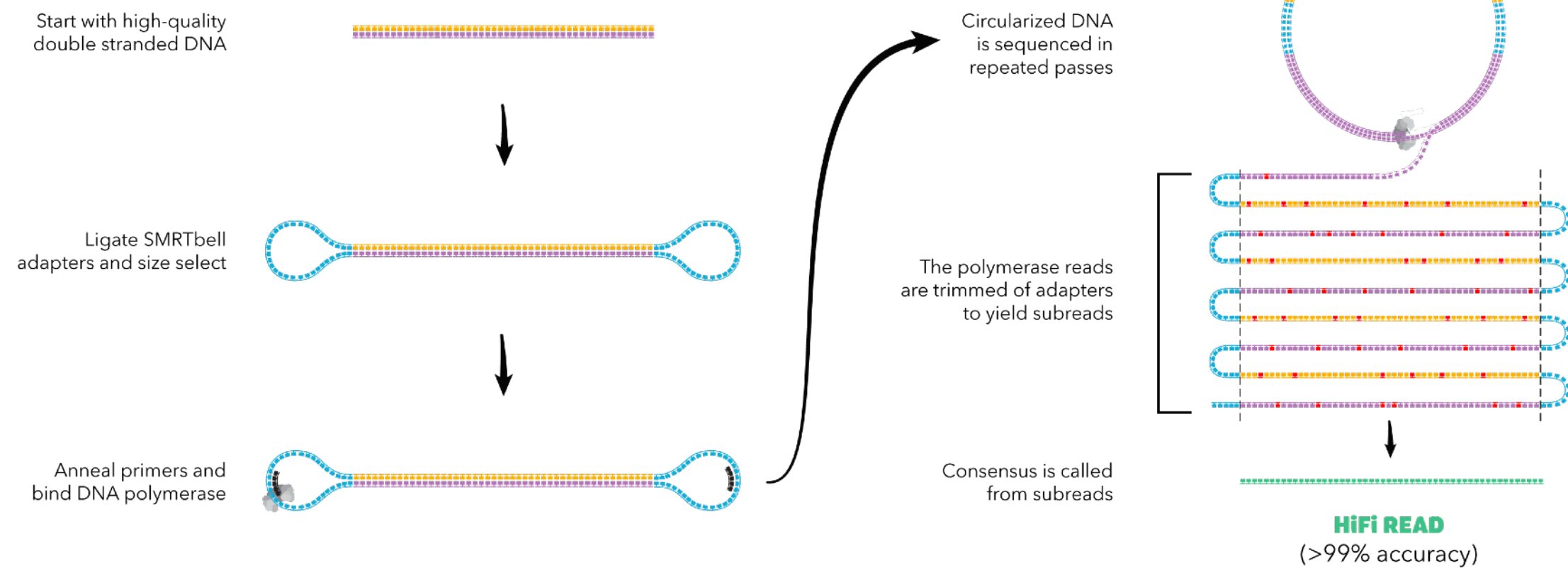
# Highly-Accurate Long Reads

**Strategy:** Build Consensus from Multiple Reads of a Long DNA Molecule

# Highly-Accurate Long Reads

**Strategy:** Build Consensus from Multiple Reads of a Long DNA Molecule

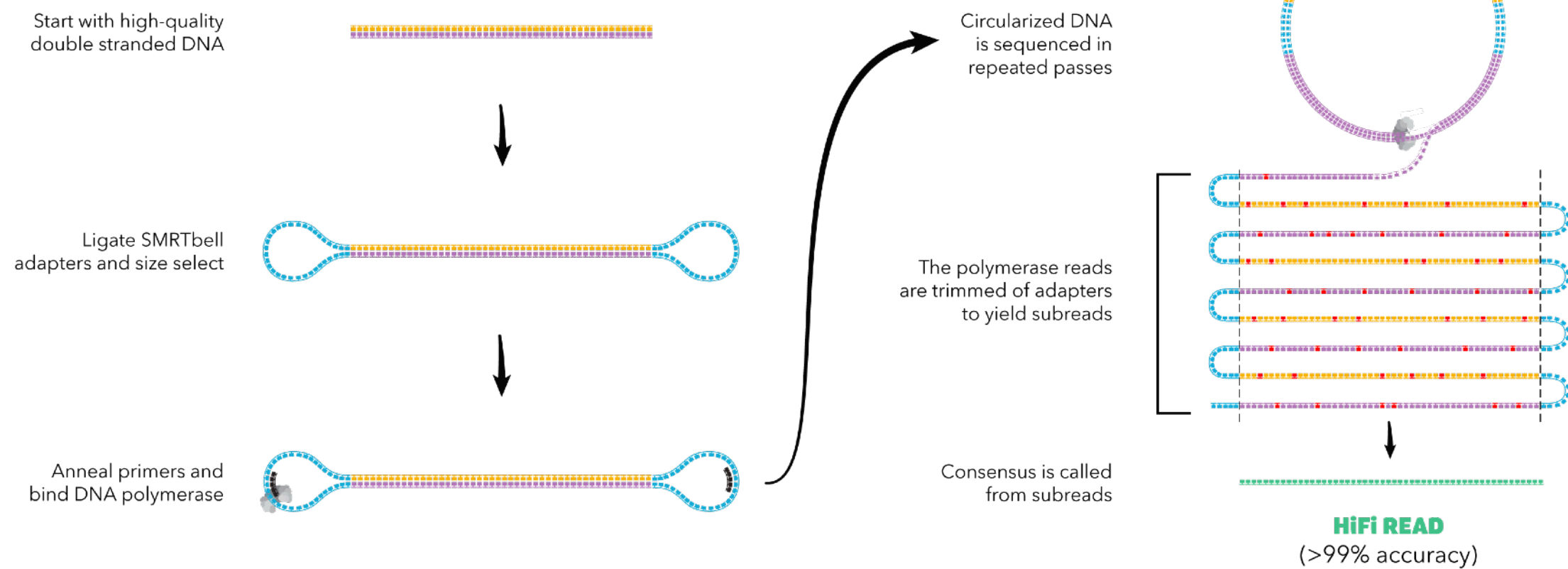
## PacBio HiFi Sequencing



# Highly-Accurate Long Reads

**Strategy:** Build Consensus from Multiple Reads of a Long DNA Molecule

## PacBio HiFi Sequencing

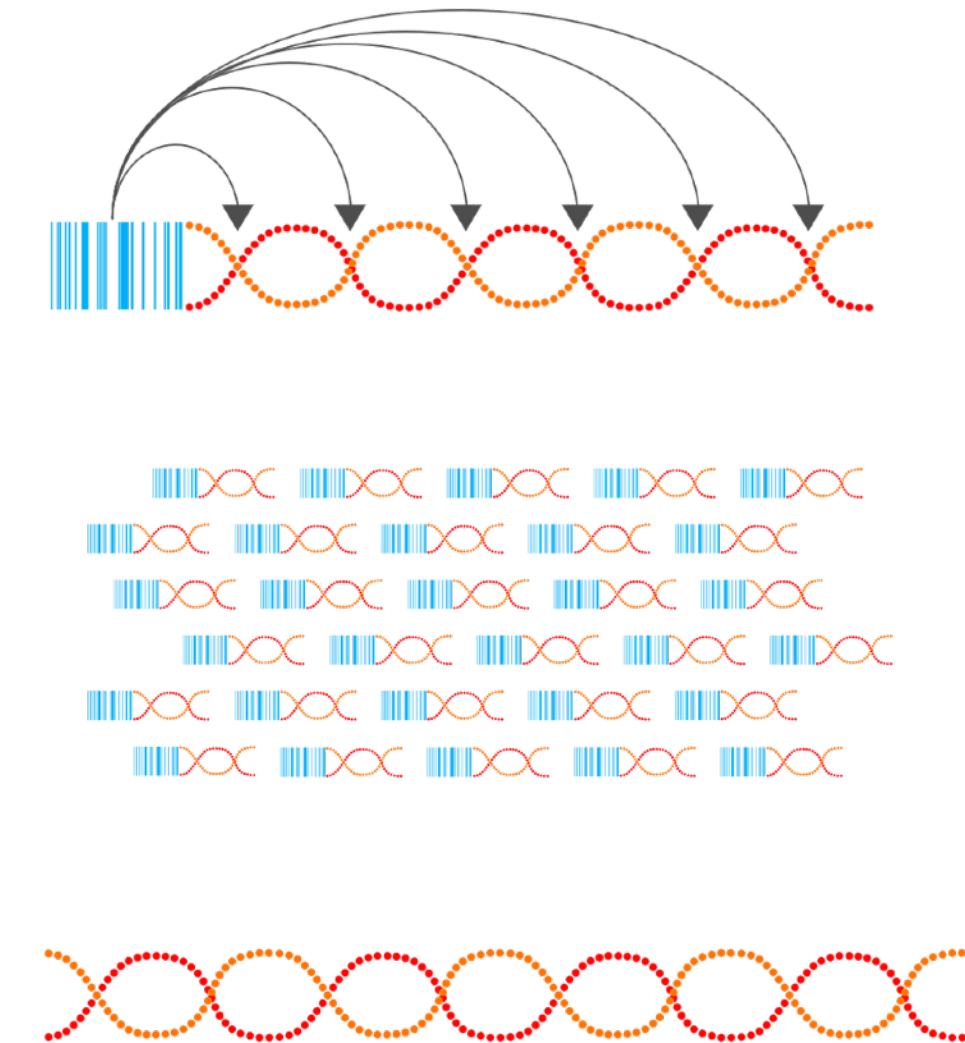


## Loop Genomics Synthetic Long Reads (SLRs)

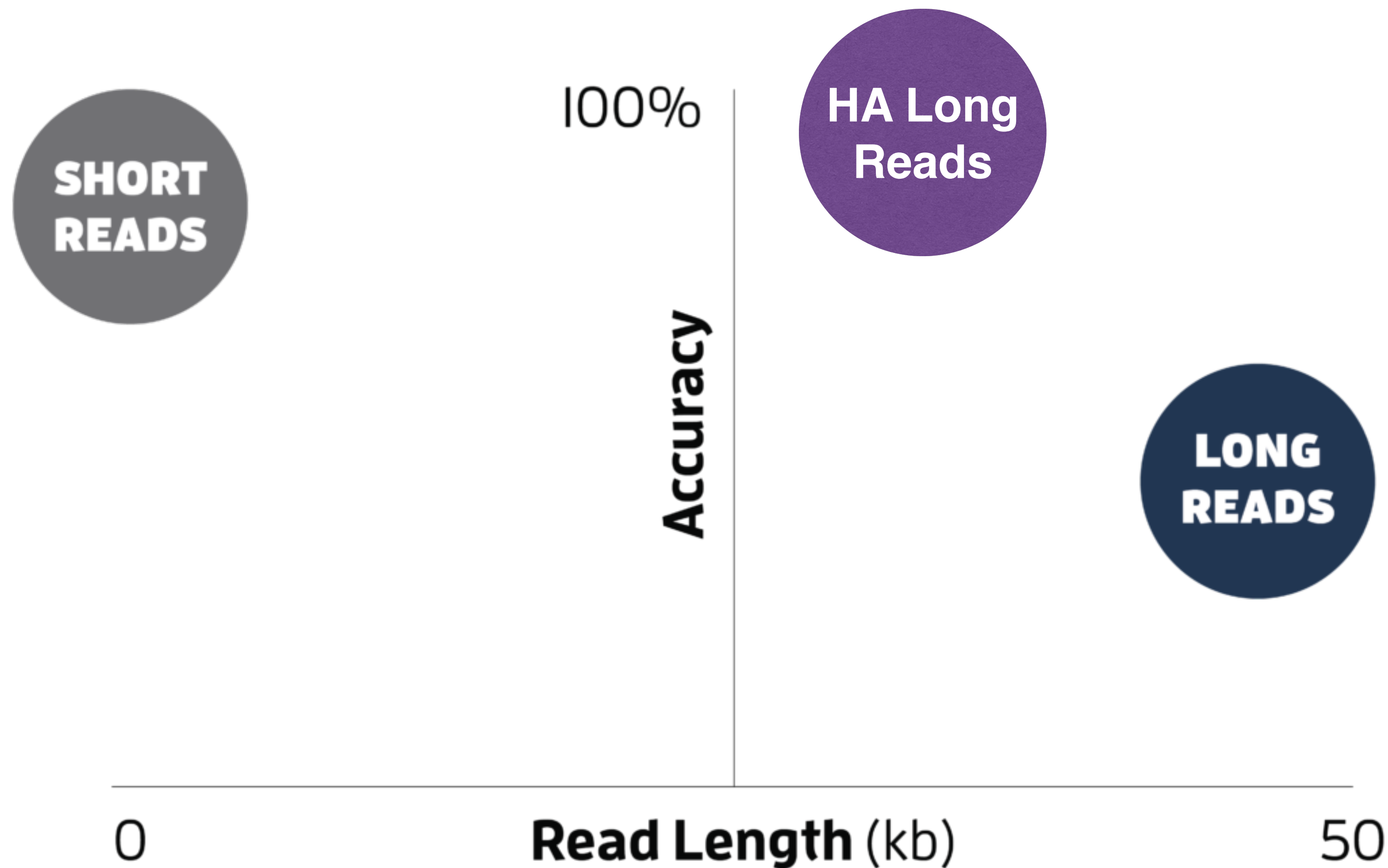
Distribute UMI throughout long DNA molecule

Short-read sequencing

Consensus assembly



# Highly-Accurate Long Reads



## Short Reads (Illumina)

Read length: **~250 nts**

Per-base error-rate: **0.1 - 0.5%**

## Long Reads (PacBio, Oxford)

Read length: **1 - 200 kilobases**

Per-base error-rate: **2 - 10%**

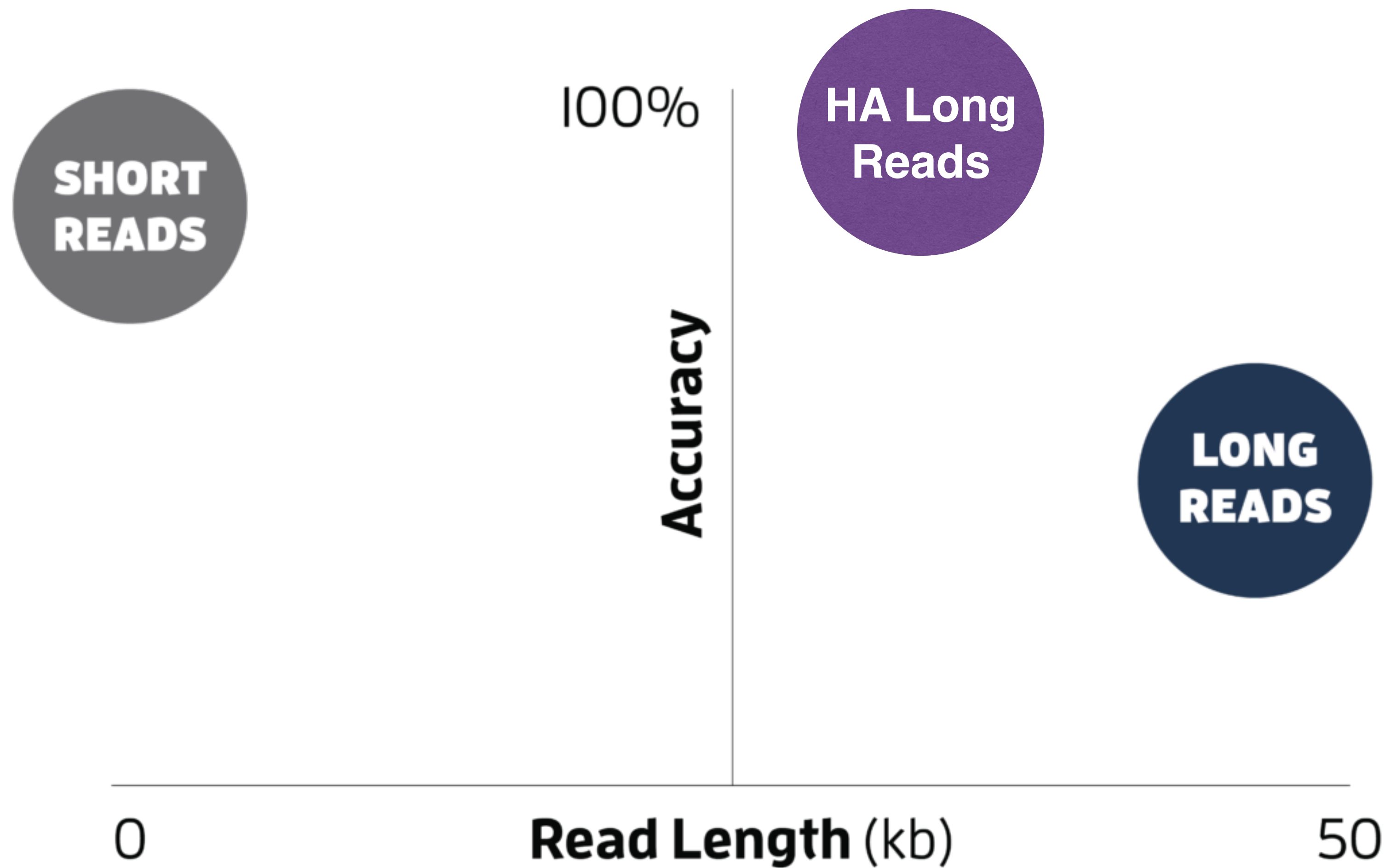
## HA Long Reads (HiFi, LoopSeq)

Read length: **1 - 20 kilobases**

Per-base error-rate: **< 0.1%**



# Highly-Accurate Long Reads



**Highly-accurate long reads are effectively a new sequencing technology.**

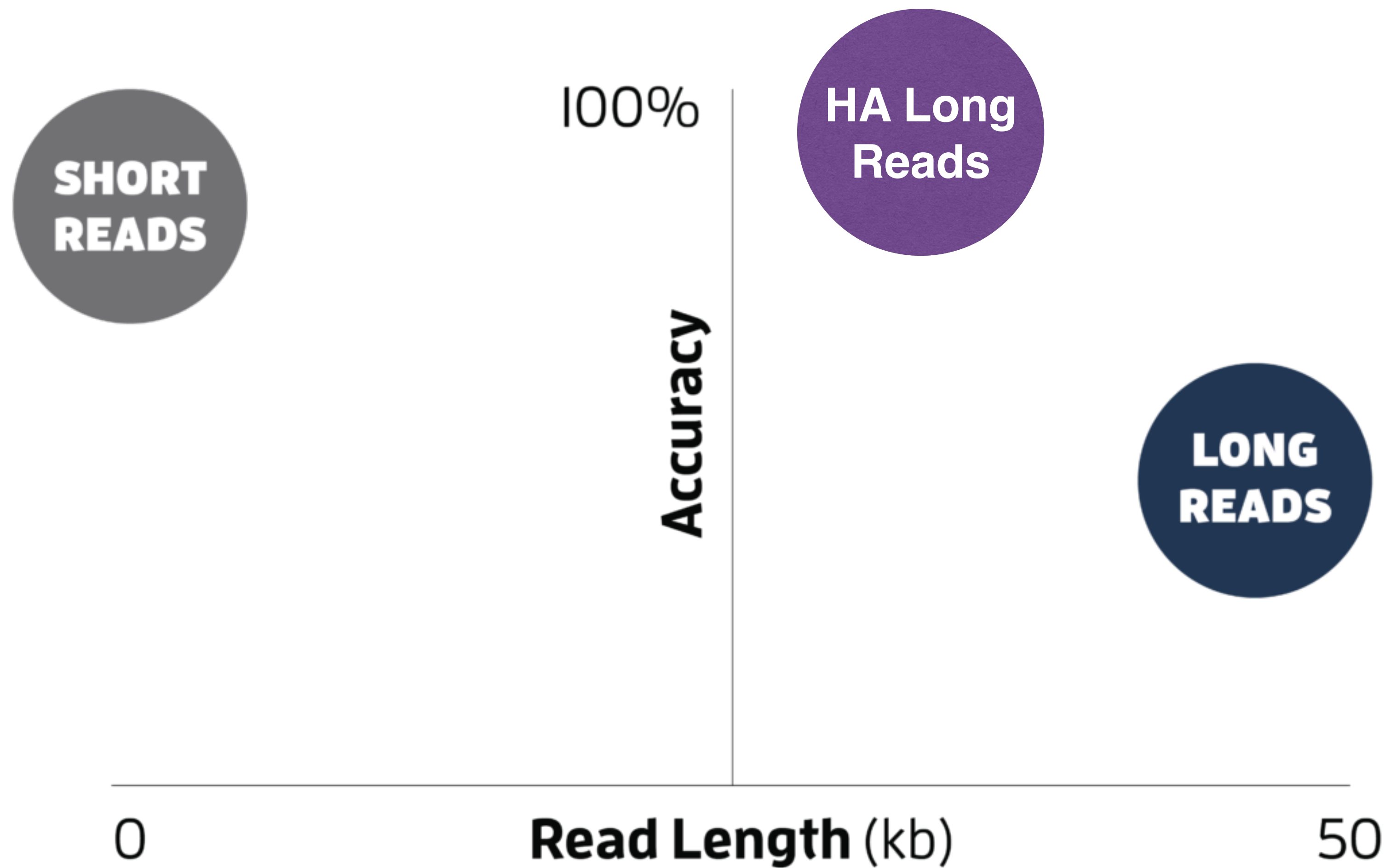
## HA Long Reads (HiFi, LoopSeq)

Read length: **1 - 20 kilobases**

Per-base error-rate: **< 0.1%**



# Highly-Accurate Long Reads



Highly-accurate long reads are effectively a new sequencing technology. **But do they really work for amplicon sequencing?**

## HA Long Reads (HiFi, LoopSeq)

Read length: **1 - 20 kilobases**

Per-base error-rate: **< 0.1%**

## Zymo Mock Community

8 Bacterial Strains at  
nominally equal concentrations

- *Pseudomonas aeruginosa*
- *Escherichia coli*
- *Salmonella enterica*
- *Lactobacillus fermentum*
- *Enterococcus faecalis*
- *Staphylococcus aureus*
- *Listeria monocytogenes*
- *Bacillus subtilis*

## Zymo Mock Community

8 Bacterial Strains at  
nominally equal concentrations

- *Pseudomonas aeruginosa*
- *Escherichia coli*
- *Salmonella enterica*
- *Lactobacillus fermentum*
- *Enterococcus faecalis*
- *Staphylococcus aureus*
- *Listeria monocytogenes*
- *Bacillus subtilis*



**PacBio  
HiFi +  
DADA2**

# Full-Length, Full-Complement 16S

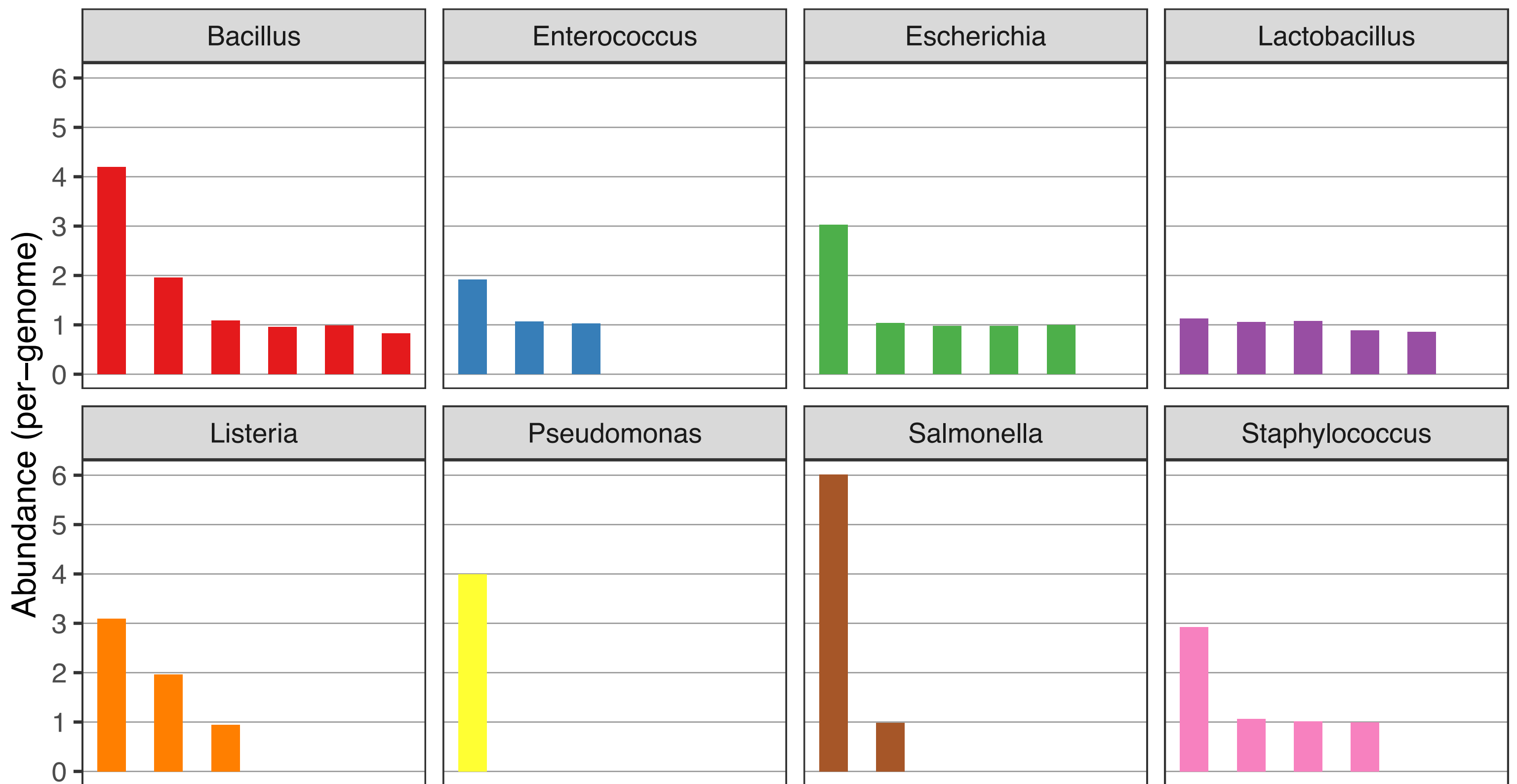
## Zymo Mock Community

8 Bacterial Strains at  
nominally equal concentrations

- *Pseudomonas aeruginosa*
- *Escherichia coli*
- *Salmonella enterica*
- *Lactobacillus fermentum*
- *Enterococcus faecalis*
- *Staphylococcus aureus*
- *Listeria monocytogenes*
- *Bacillus subtilis*



**PacBio  
HiFi +  
DADA2**



Full-length 16S Sequence Variants

# Full-Length, Full-Complement 16S

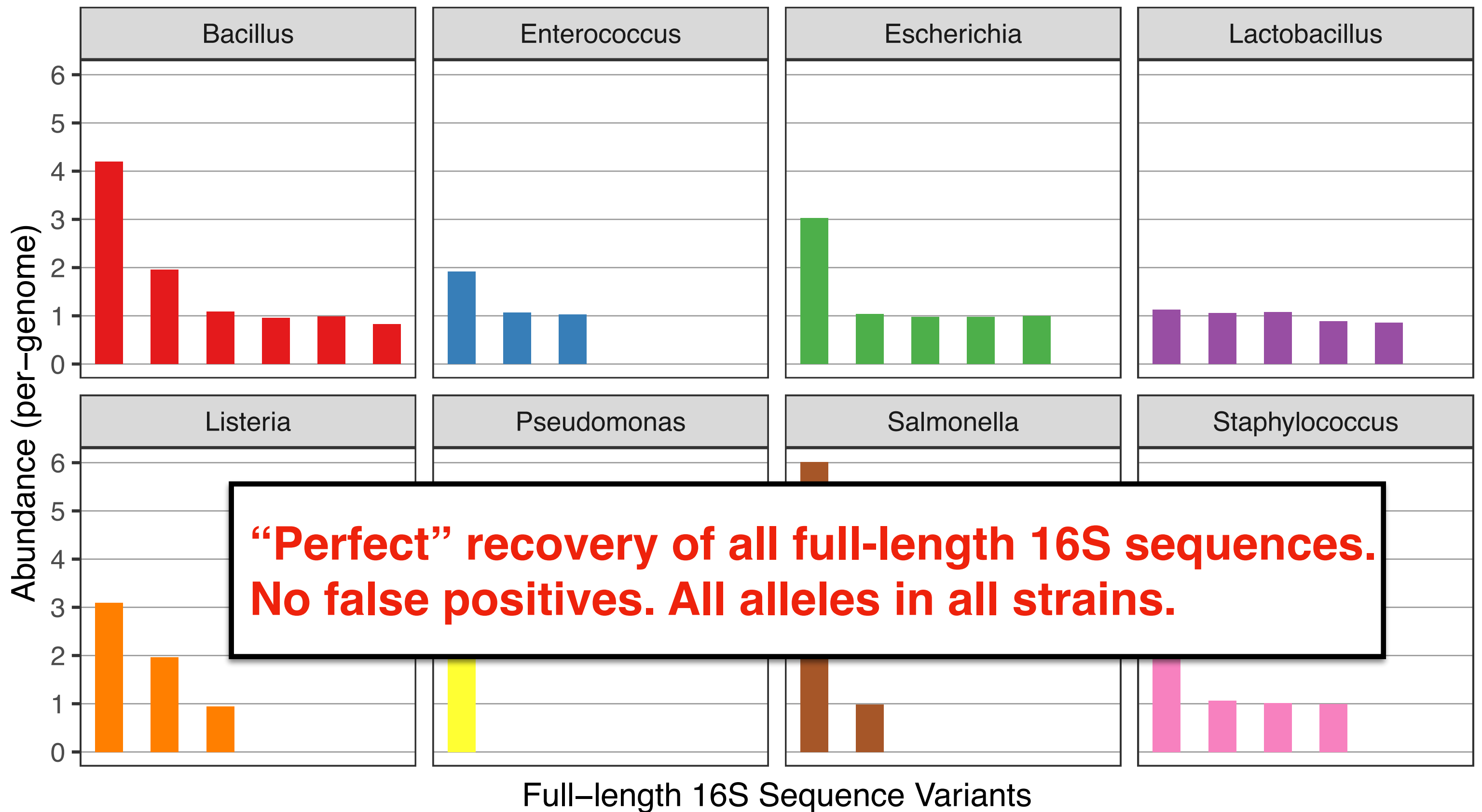
## Zymo Mock Community

8 Bacterial Strains at nominally equal concentrations

- *Pseudomonas aeruginosa*
- *Escherichia coli*
- *Salmonella enterica*
- *Lactobacillus fermentum*
- *Enterococcus faecalis*
- *Staphylococcus aureus*
- *Listeria monocytogenes*
- *Bacillus subtilis*



**PacBio  
HiFi +  
DADA2**



# Full-Length, Full-Complement 16S

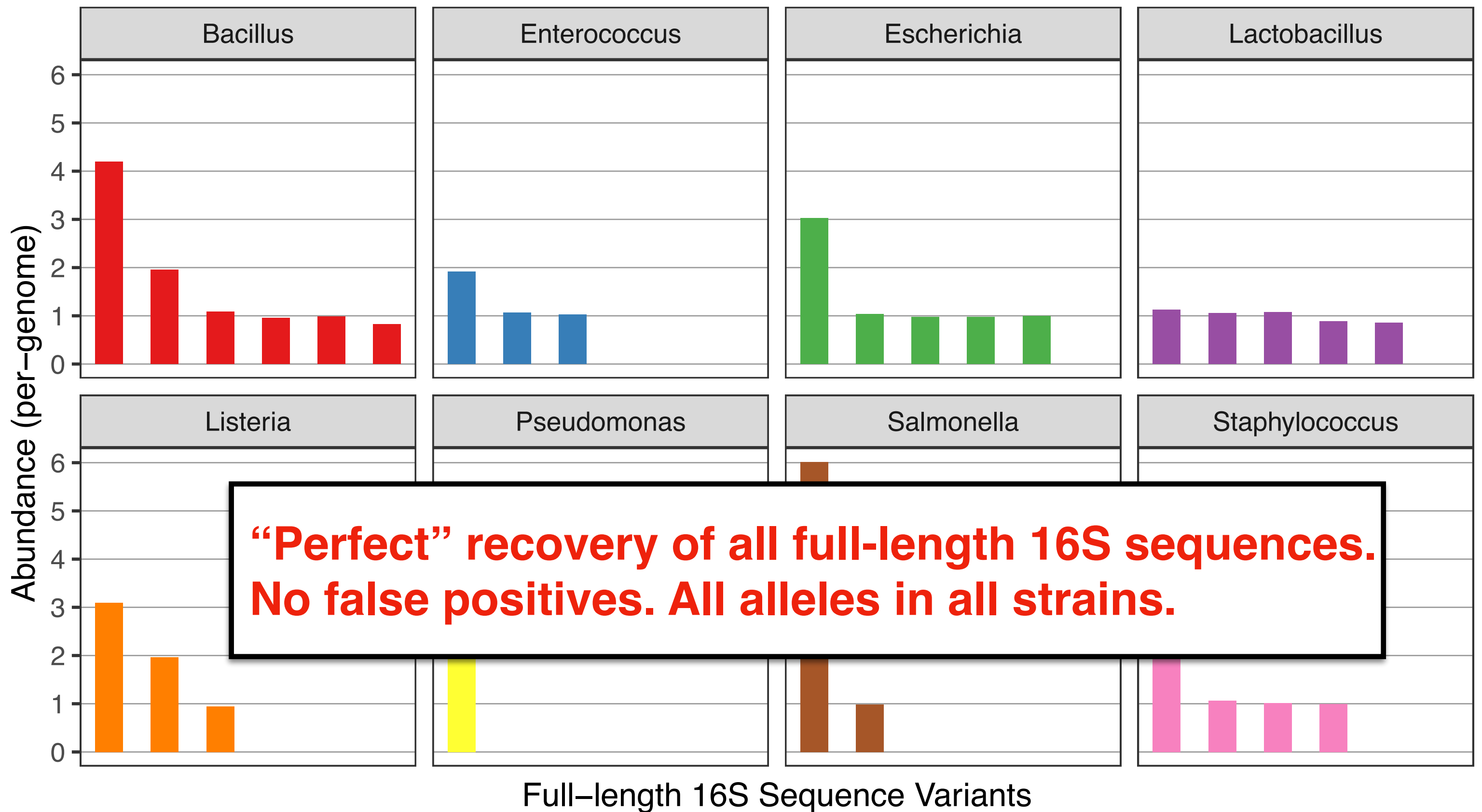
## Zymo Mock Community

8 Bacterial Strains at nominally equal concentrations

- *Pseudomonas aeruginosa*
- *Escherichia coli*
- *Salmonella enterica*
- *Lactobacillus fermentum*
- *Enterococcus faecalis*
- *Staphylococcus aureus*
- *Listeria monocytogenes*
- *Bacillus subtilis*



**PacBio  
HiFi +  
DADA2\***




\* Modestly modified workflow for long-read amplicon sequencing.



## PacBio HiFi

### High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution

[Benjamin J Callahan](#) , [Joan Wong](#), [Cheryl Heiner](#), [Steve Oh](#), [Casey M Theriot](#), [Ajay S Gulati](#), [Sarah K McGill](#), [Michael K Dougherty](#)

*Nucleic Acids Research*, Volume 47, Issue 18, 10 October 2019, Page e103,

<https://doi.org/10.1093/nar/gkz569>

**Published:** 03 July 2019 **Article history** ▼

## LoopSeq

### Ultra-accurate microbial amplicon sequencing with synthetic long reads


[Benjamin J. Callahan](#) , [Dmitry Grinevich](#), [Siddhartha Thakur](#), [Michael A. Balamotis](#) & [Tuval Ben Yehezkel](#)

*Microbiome* **9**, Article number: 130 (2021) | [Cite this article](#)

**5717** Accesses | **11** Citations | **44** Altmetric | [Metrics](#)

## PacBio HiFi

High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution 

[Benjamin J Callahan](#) , [Joan Wong](#), [Cheryl Heiner](#), [Steve Oh](#), [Casey M Theriot](#), [Ajay S Gulati](#), [Sarah K McGill](#), [Michael K Dougherty](#)

*Nucleic Acids Research*, Volume 47, Issue 18, 10 October 2019, Page e103,

<https://doi.org/10.1093/nar/gkz569>

**Published:** 03 July 2019 [Article history](#) ▼

## LoopSeq

Ultra-accurate microbial amplicon sequencing with synthetic long reads

[Benjamin J. Callahan](#) , [Dmitry Grinevich](#), [Siddhartha Thakur](#), [Michael A. Balamotis](#) & [Tuval Ben Yehezkel](#)


*Microbiome* **9**, Article number: 130 (2021) | [Cite this article](#)

**5717** Accesses | **11** Citations | **44** Altmetric | [Metrics](#)

**Highly-accurate long reads  
are effectively a new  
sequencing technology.  
But do they really  
work for amplicon sequencing?**

## PacBio HiFi

High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution 

Benjamin J Callahan , Joan Wong, Cheryl Heiner, Steve Oh, Casey M Theriot, Ajay S Gulati, Sarah K McGill, Michael K Dougherty

*Nucleic Acids Research*, Volume 47, Issue 18, 10 October 2019, Page e103,

<https://doi.org/10.1093/nar/gkz569>

Published: 03 July 2019 [Article history](#) ▼

## LoopSeq

Ultra-accurate microbial amplicon sequencing with synthetic long reads

Benjamin J. Callahan , Dmitry Grinevich, Siddhartha Thakur, Michael A. Balamotis & Tuval Ben Yehzekel

*Microbiome* **9**, Article number: 130 (2021) | [Cite this article](#)

5717 Accesses | 11 Citations | 44 Altmetric | [Metrics](#)

**Highly-accurate long reads  
are effectively a new  
sequencing technology.  
But do they really  
work for amplicon sequencing?**

**Yes!**  
Multiple technologies  
Out to 6+ kilo bases  
>99.9% accuracy



*We support long-read amplicon sequencing!*



*We support long-read amplicon sequencing!*

Fastq

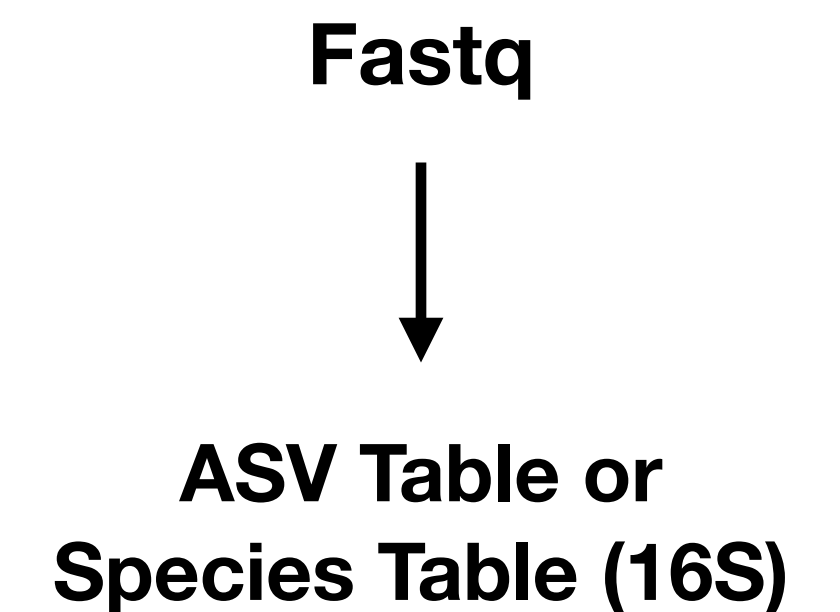


**ASV Table or  
Species Table (16S)**





*We support long-read amplicon sequencing!*



---

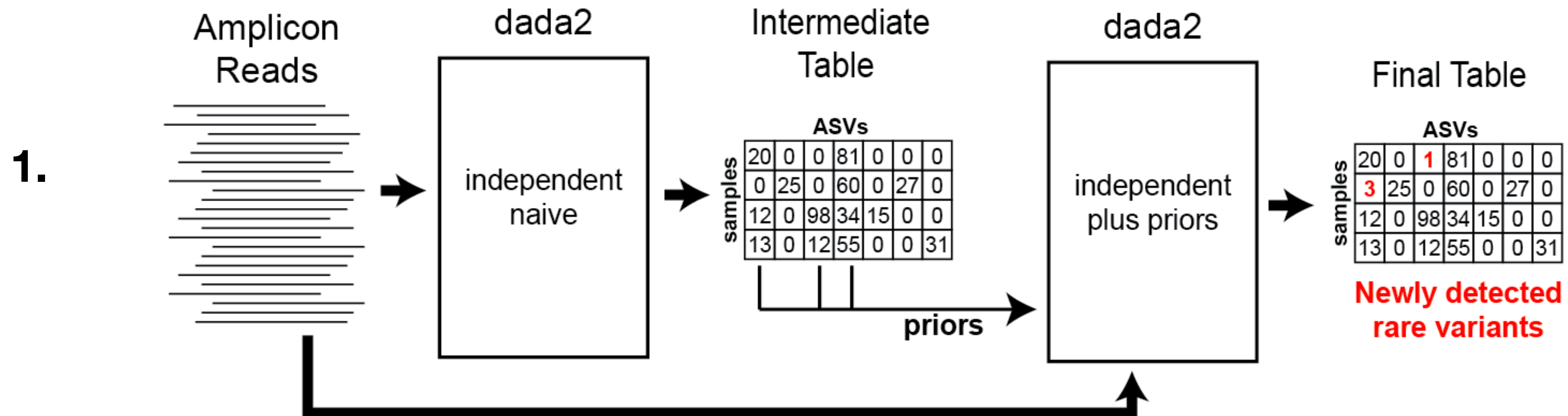
## Long-read specific documentation

**PacBio HiFi:** <https://github.com/benjjneb/LRASmanuscript>

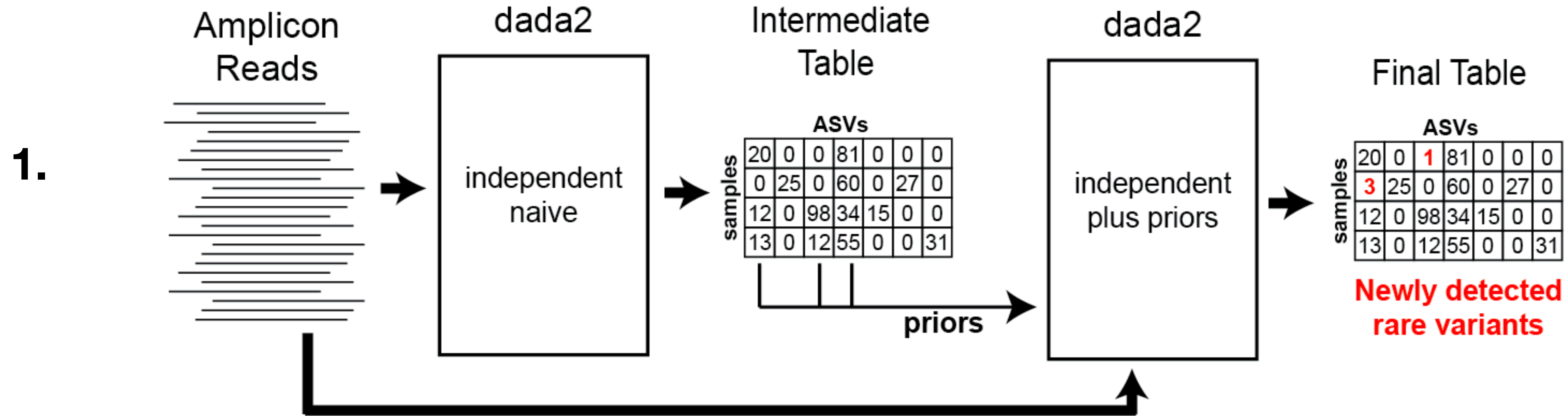
**LoopSeq:** <https://github.com/benjjneb/LoopManuscript>



## Pseudo-Pooling

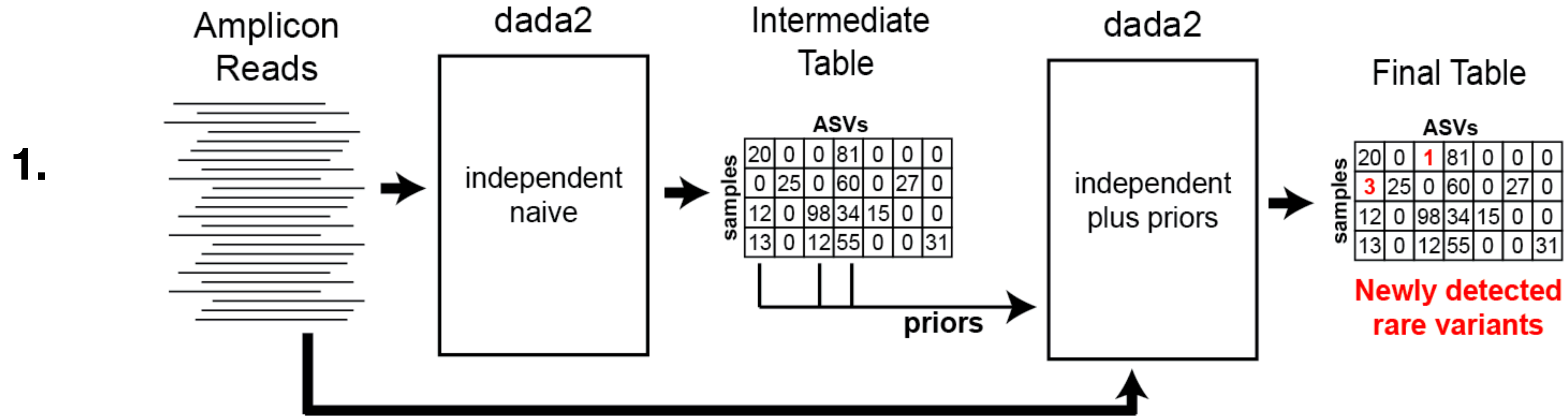


## Pseudo-Pooling



2. DETECT\_SINGLETONS=TRUE

## Pseudo-Pooling

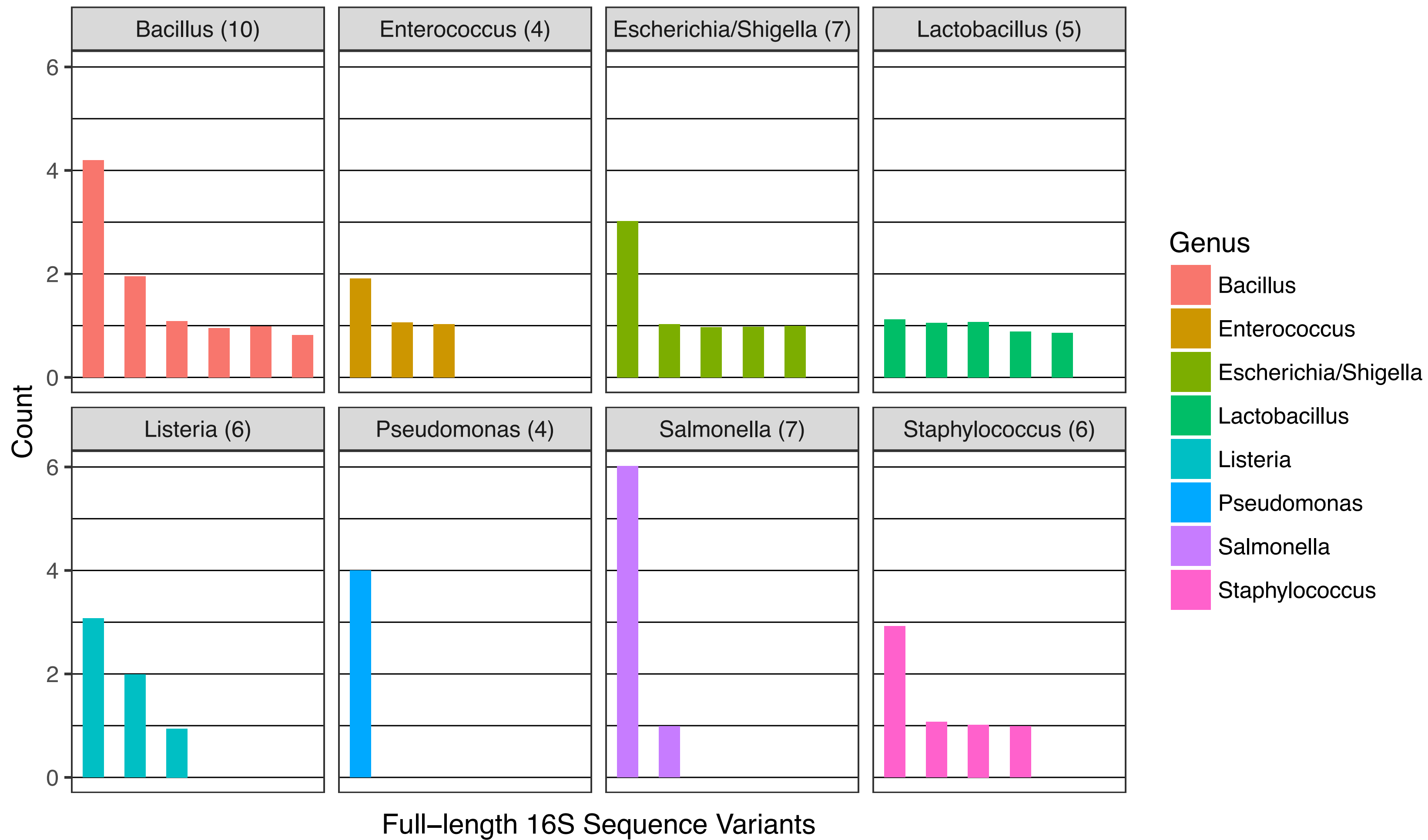


2. DETECT\_SINGLETONS=TRUE

3. Long-read specific parameters.

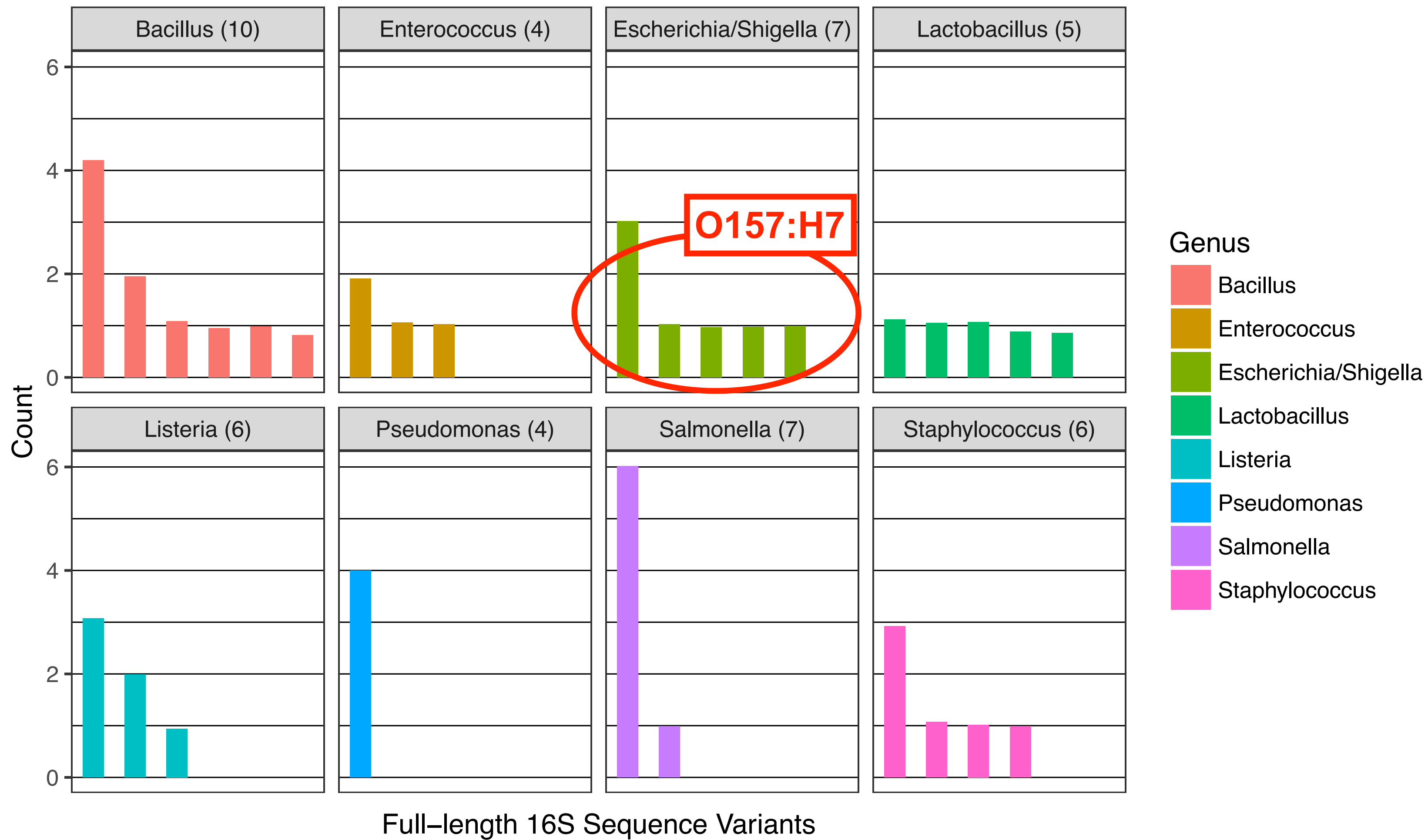
# Sub-species bacterial profiling

*Full-length 16S from Zymo mock community*



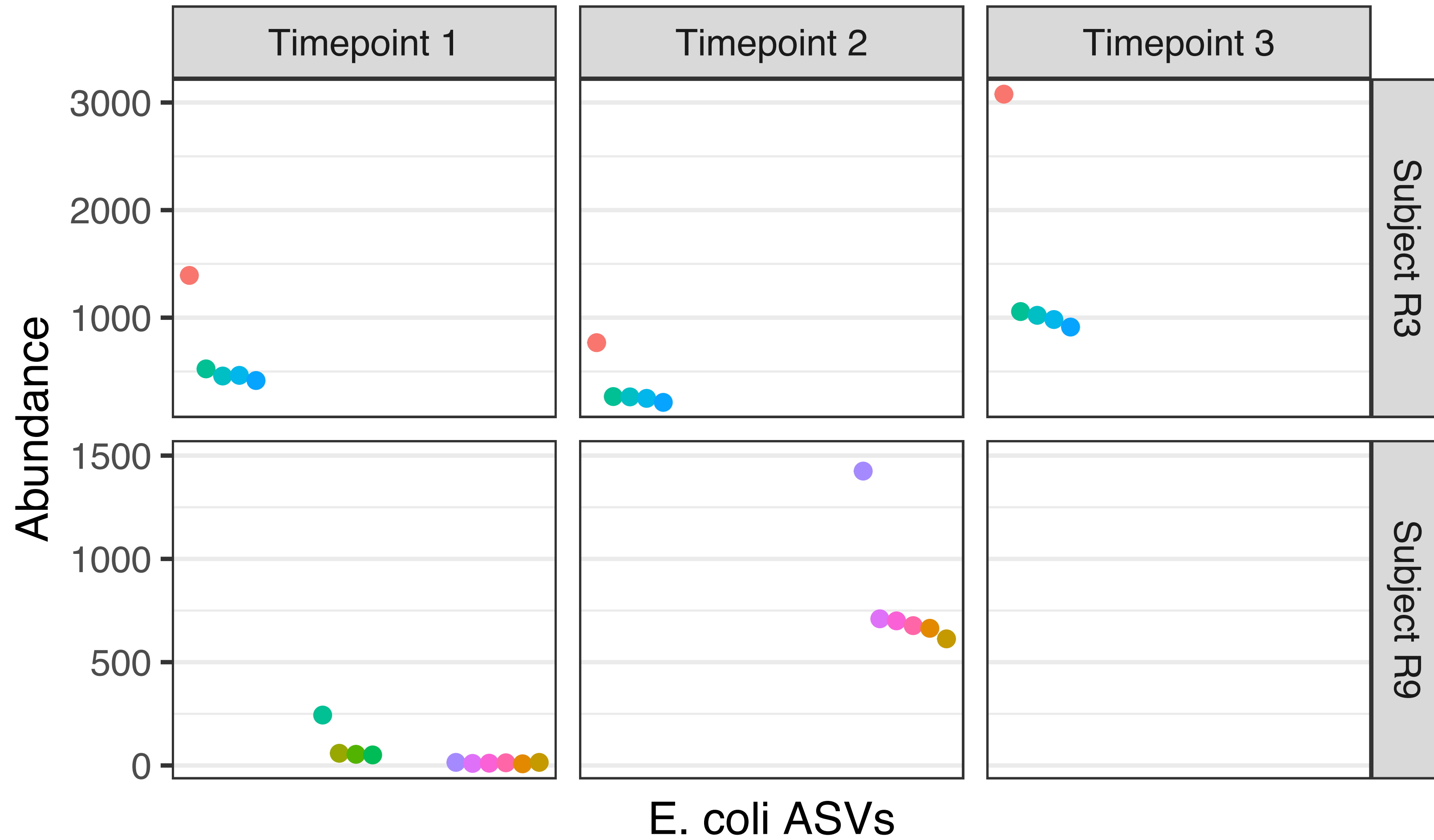
# Sub-species bacterial profiling

*Full-length 16S from Zymo mock community*



# Sub-species bacterial profiling

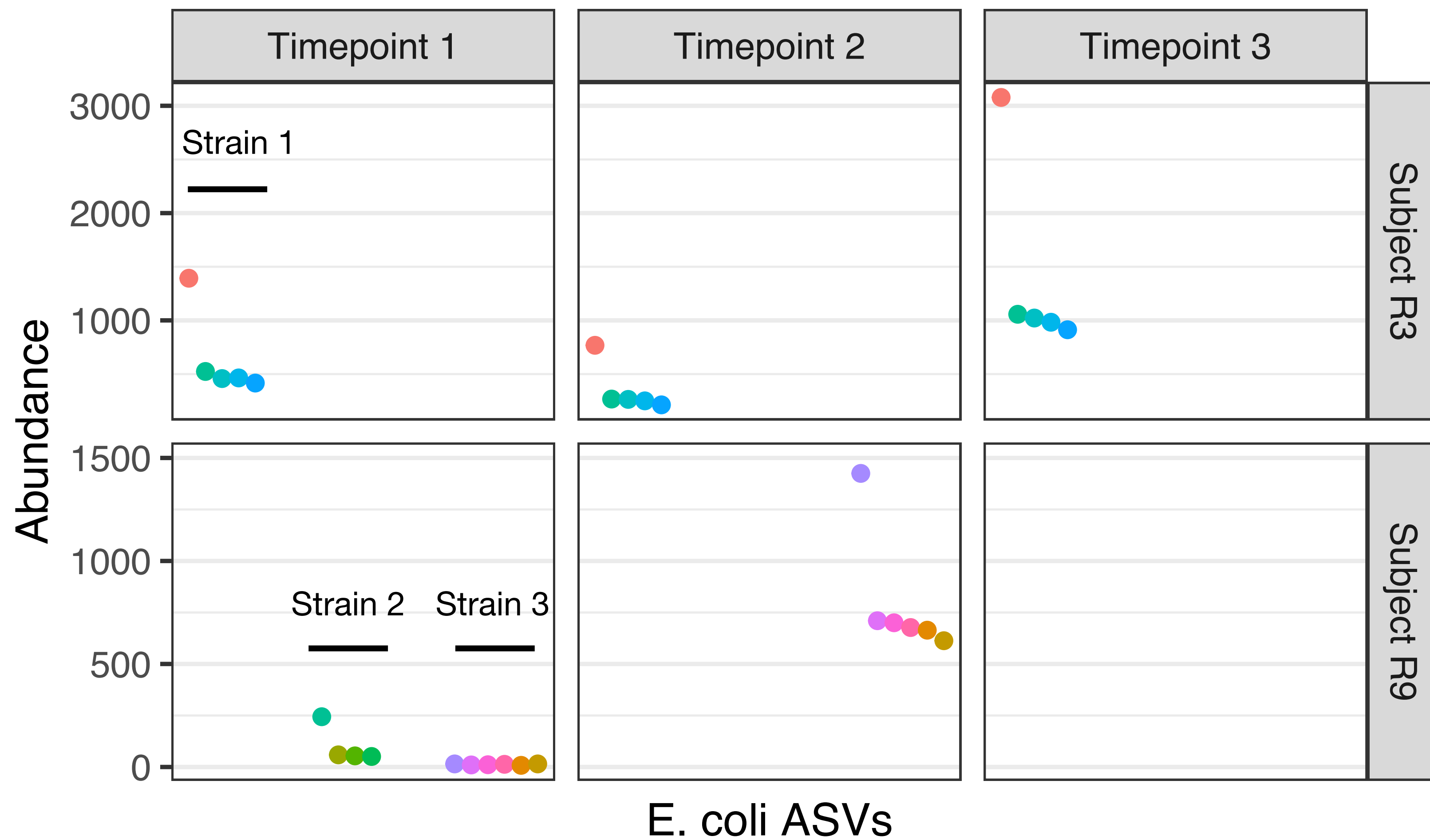
*Full-length 16S from human fecal samples*





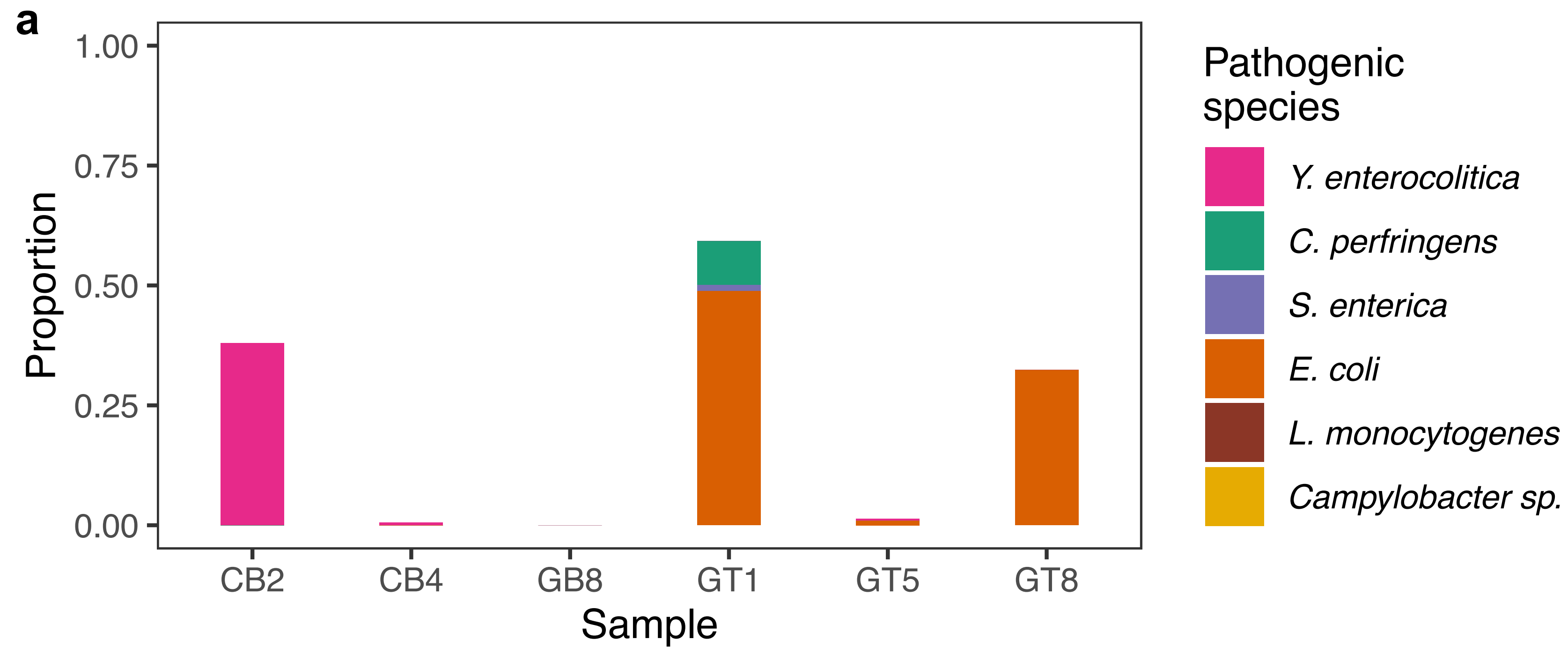
# Sub-species bacterial profiling

*Full-length 16S from human fecal samples*



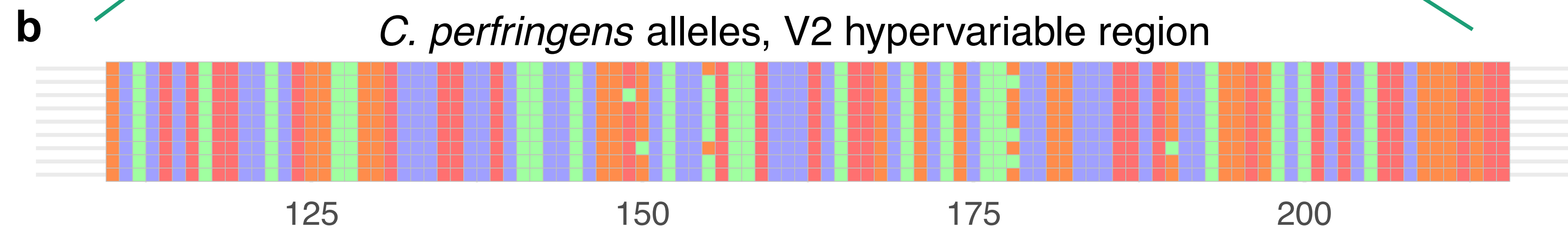
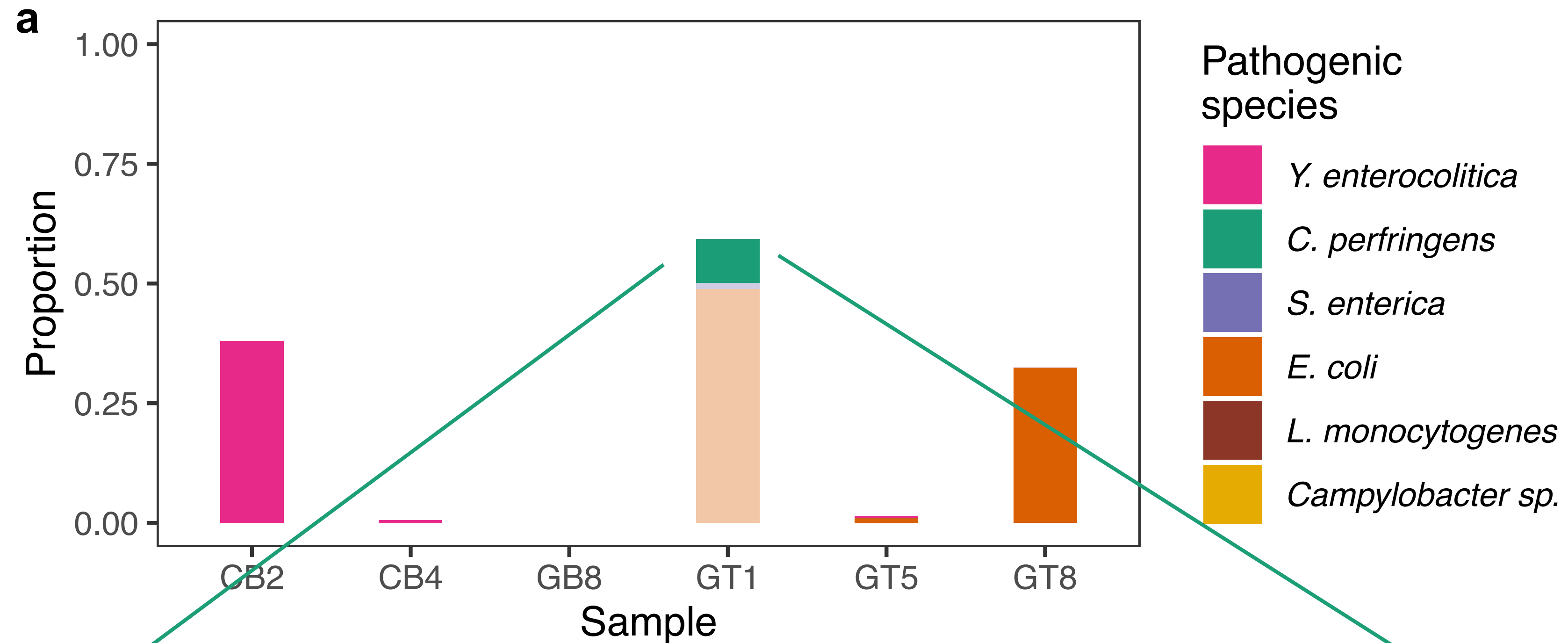
# Sub-species bacterial profiling

*Full-length 16S from retail meat samples*



# Sub-species bacterial profiling

Full-length 16S from retail meat samples

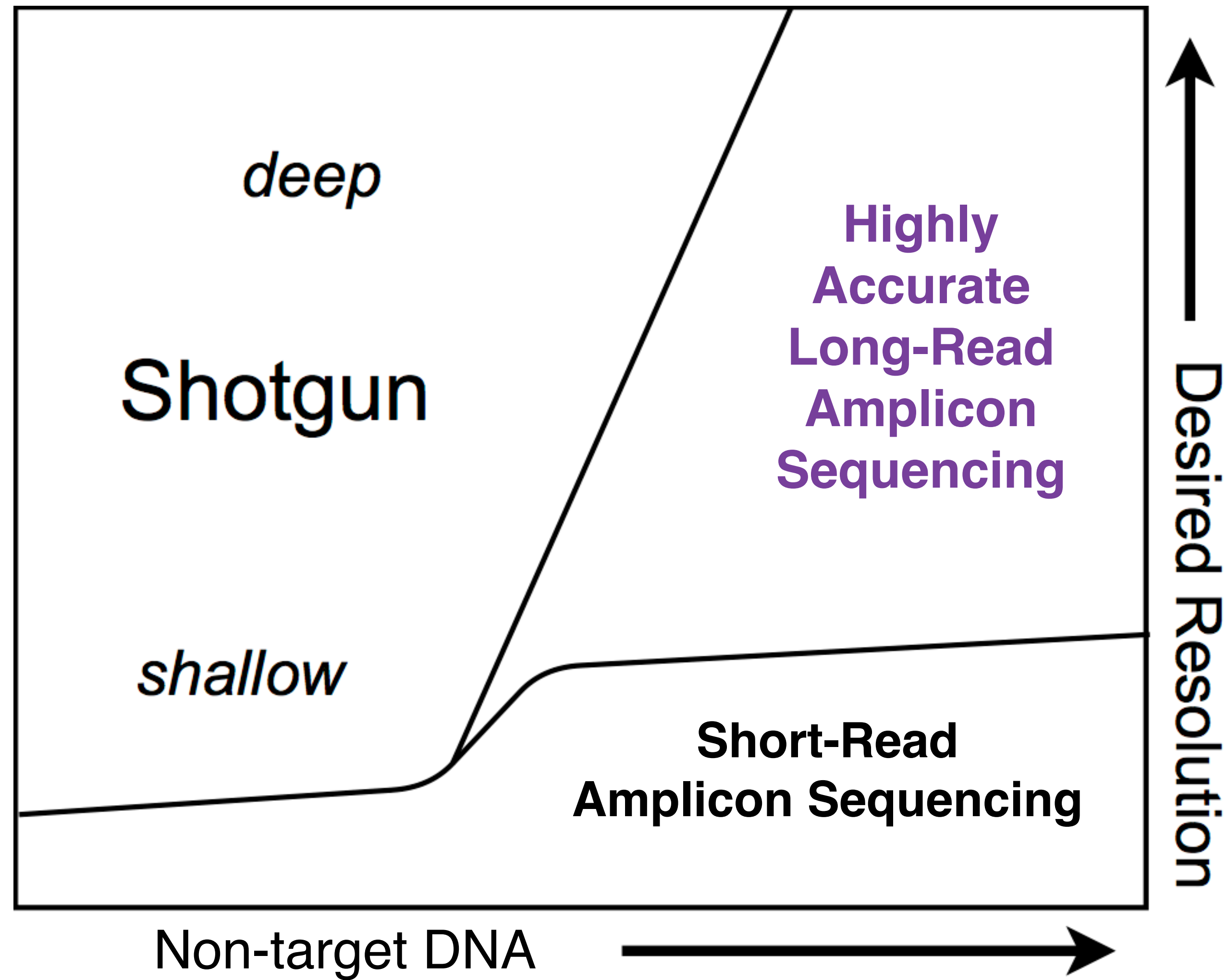


Full genomic complement of 16S alleles

## Limitations

- PCR Biases
- PCR Length Scaling (past ~5 kilo bases)
- Cost-per-base
- Chimeras
- Tooling (especially sub-species assignment)
- ...

# Highly Accurate Long-Read Amplicon Seq



## Limitations

- PCR Biases
- PCR Length Scaling (past ~5 kilo bases)
- Cost-per-base
- Chimeras
- **Tooling (especially sub-species assignment)**
- ...





New Results

 [Follow this preprint](#)

## Serovar-level Identification of Bacterial Foodborne Pathogens From Full-length 16S rRNA Gene Sequencing

Dmitry Grinevich, Lyndy Harden, Siddhartha Thakur,  Benjamin J Callahan

**doi:** <https://doi.org/10.1101/2023.06.28.546915>

This article is a preprint and has not been certified by peer review [what does this mean?].



**Dmitry Grinevich**



Sid Thakur and lab



National Institute of  
Food and Agriculture

## **Bioinformatics**

Susan Holmes  
Joey McMurdie  
Michael Rosen  
Dmitry Grinevich

## **Applications**

Dmitry Grinevich  
Casey Theriot  
Siddhartha Thakur



## **Technology (PacBio)**

Joan Wong  
Cheryl Heiner  
Steve Oh



## **Technology (Loop Genomics)**

Tuval Ben Yehezkel  
Michael Balamotis

