# Maximizing Resolution with DADA2 and Long-read Amplicon Sequencing

# Highly-Accurate Long Reads

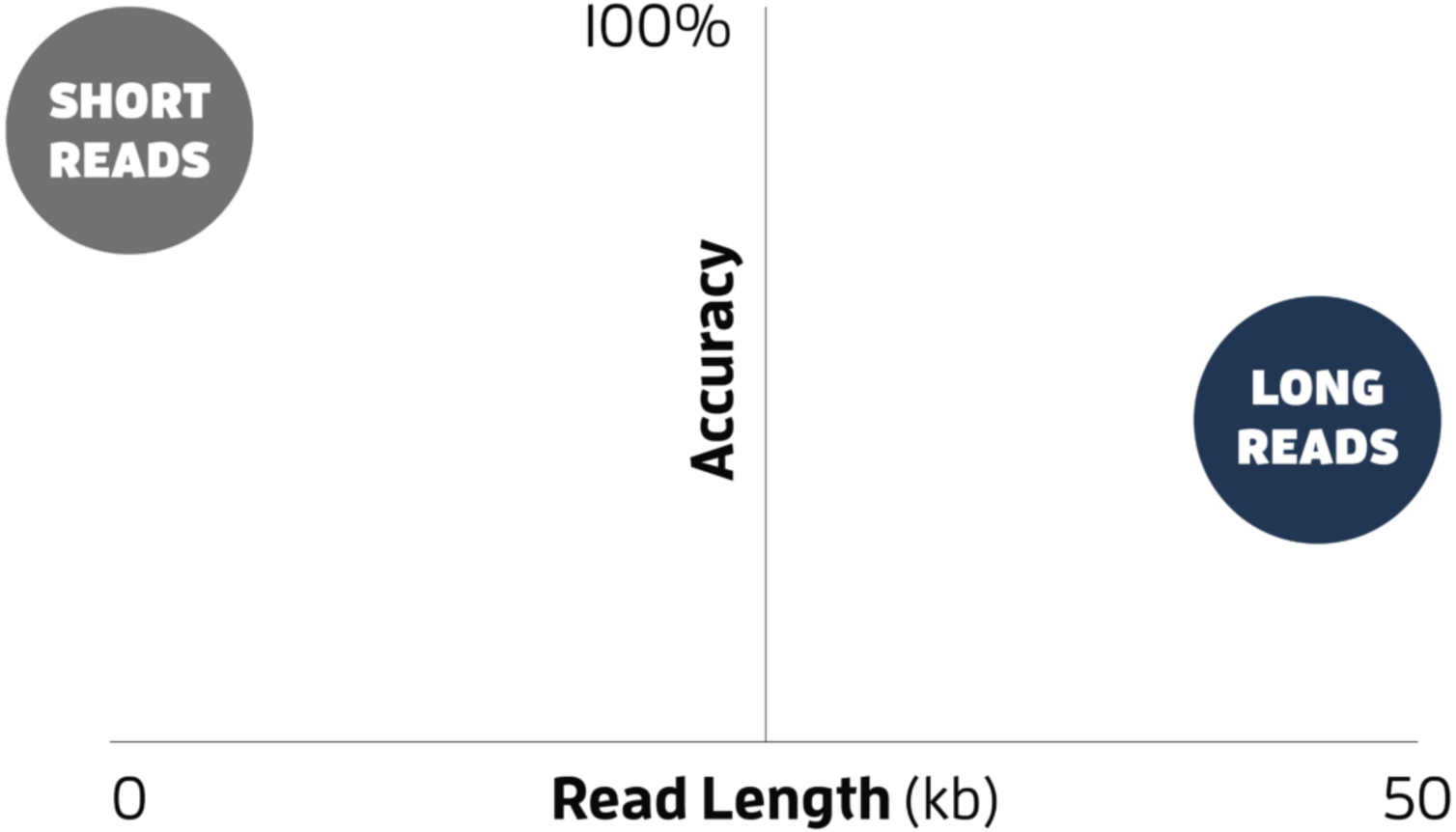



**Image modified from:** PacBio (pacb.com)

# Highly-Accurate Long Reads

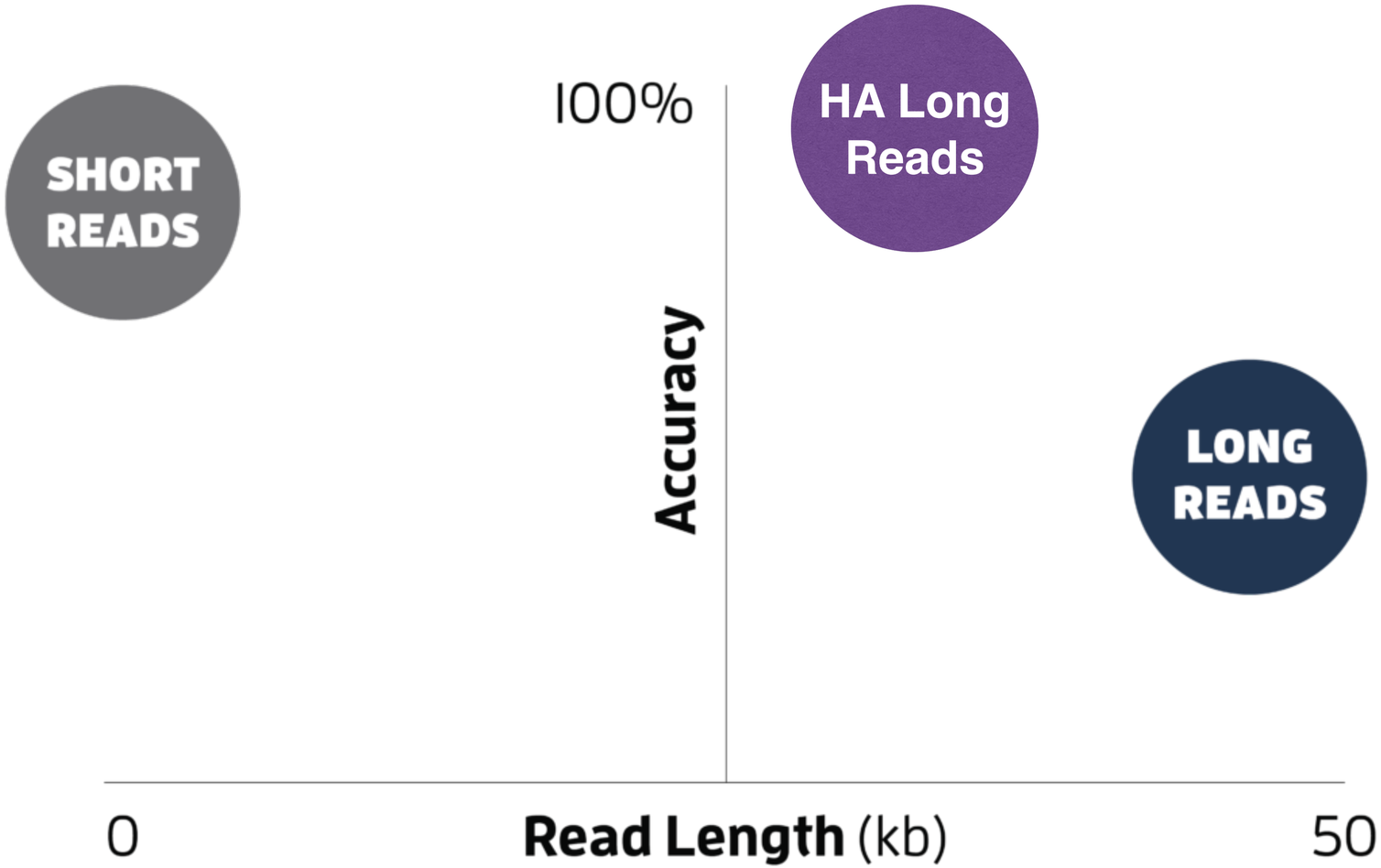



**Image modified from:** PacBio (pacb.com)

# Highly-Accurate Long Reads

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

**Image Credit:** PacBio (pacb.com), Loop Genomics (loopgenomics.com)

# Highly-Accurate Long Reads

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

## PacBio HiFi Sequencing

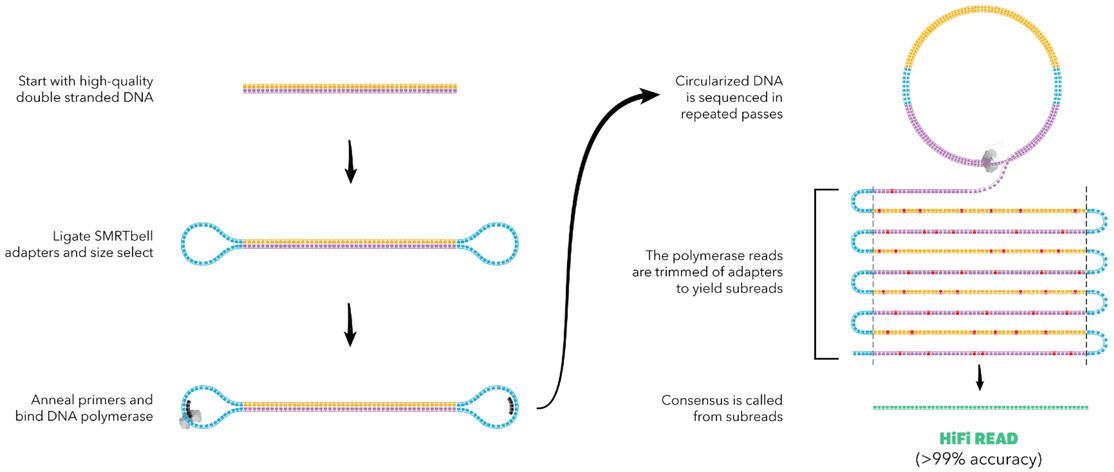
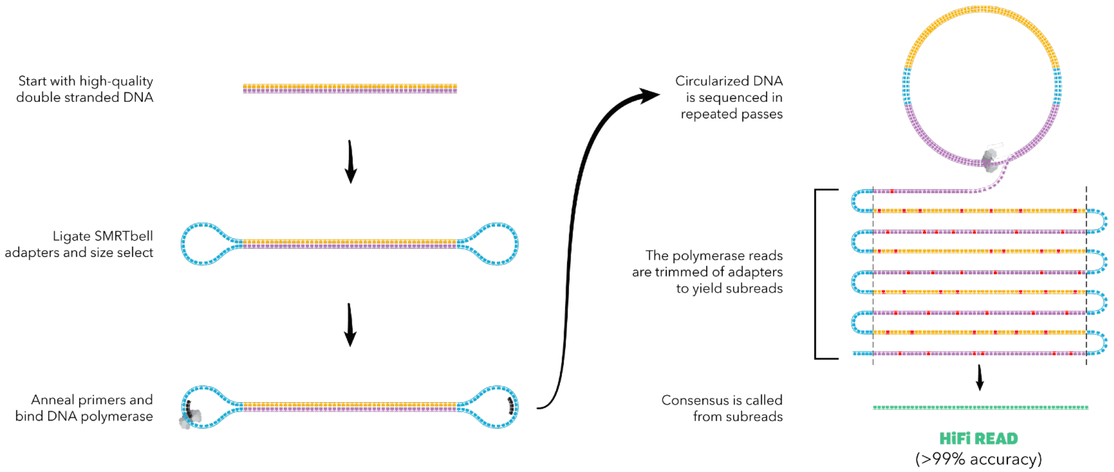



**Image Credit:** PacBio (pacb.com), Loop Genomics (loopgenomics.com)

# Highly-Accurate Long Reads

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

## PacBio HiFi Sequencing




## Element Biosciences Synthetic Long Reads (SLRs)

Distribute UMI throughout long DNA molecule

↓

Short-read sequencing

↓

Consensus assembly

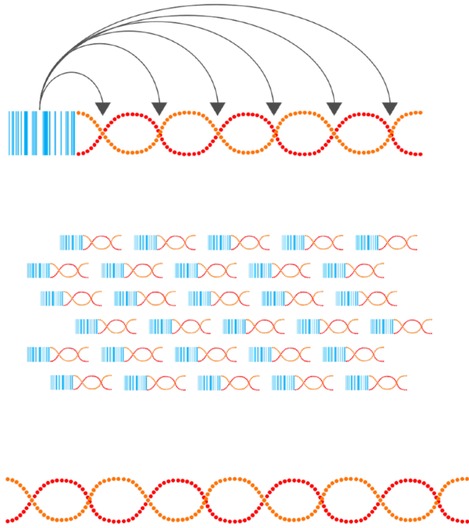
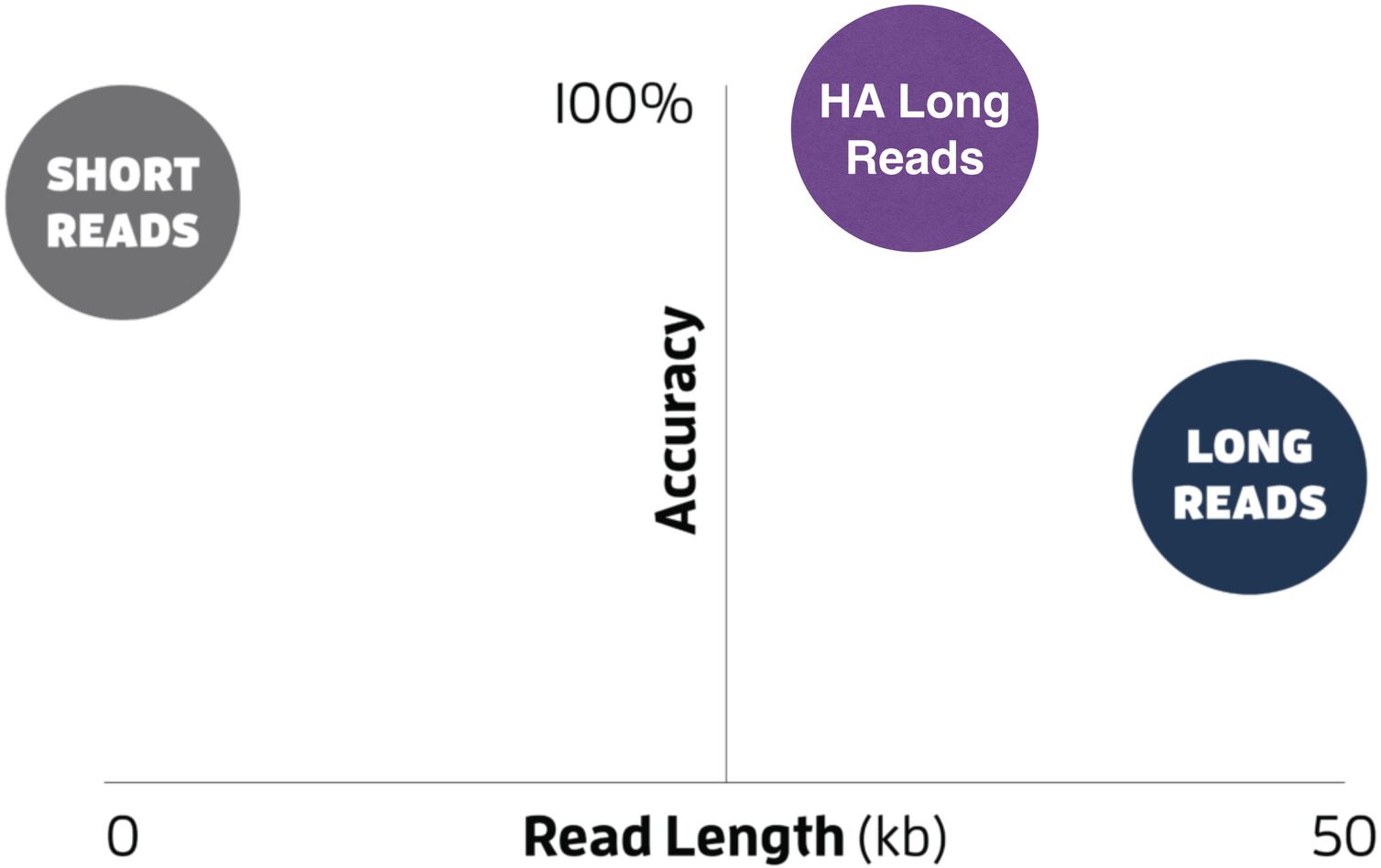


**Image Credit:** PacBio (pacb.com), Loop Genomics (loopgenomics.com)

# 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 *kilo*bases**

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

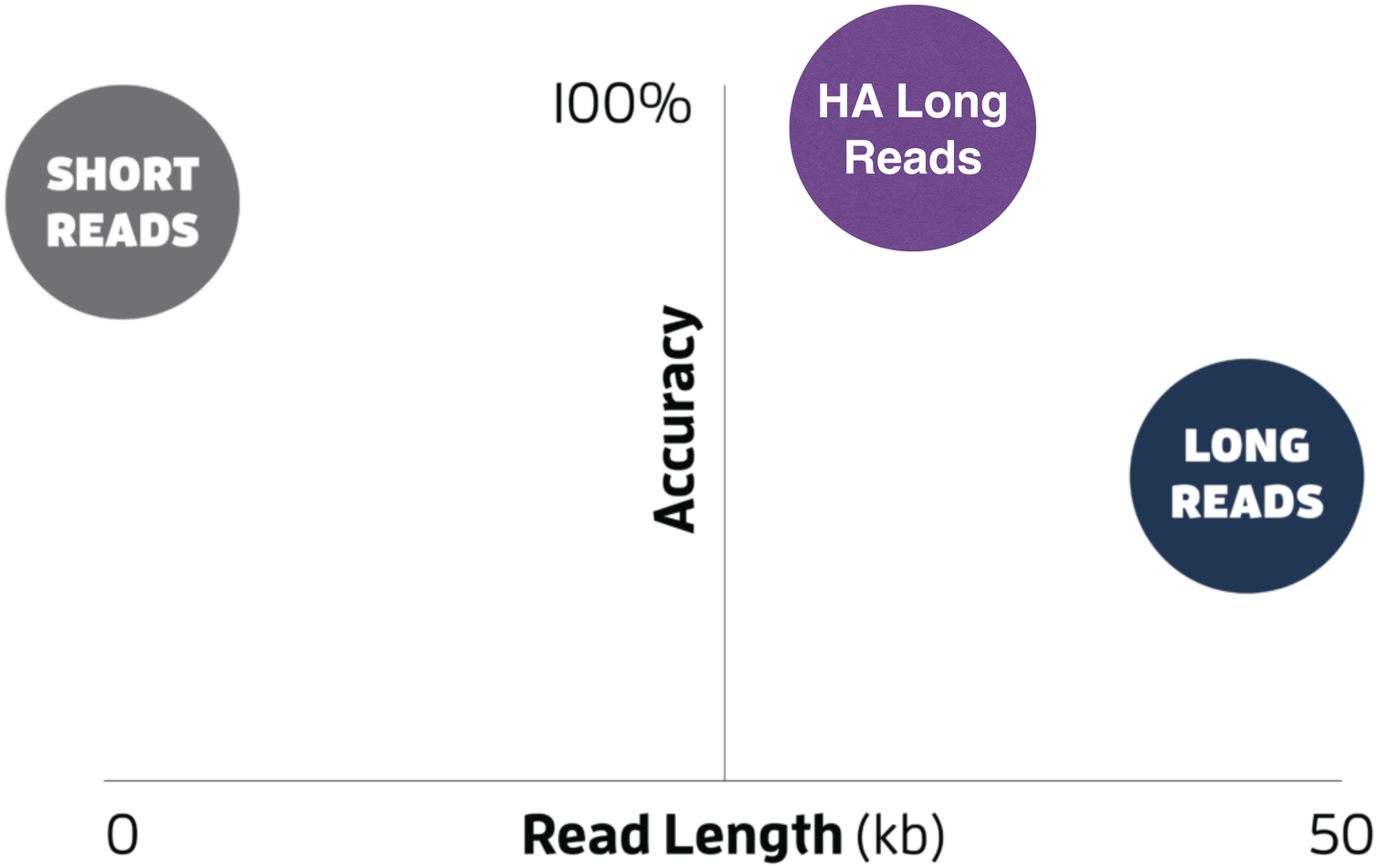
**HA Long Reads (HiFi, LoopSeq)**

Read length: **1 - 20 *kilo*bases**

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

**Image modified from:** PacBio (pacb.com)

# Highly-Accurate Long Reads




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

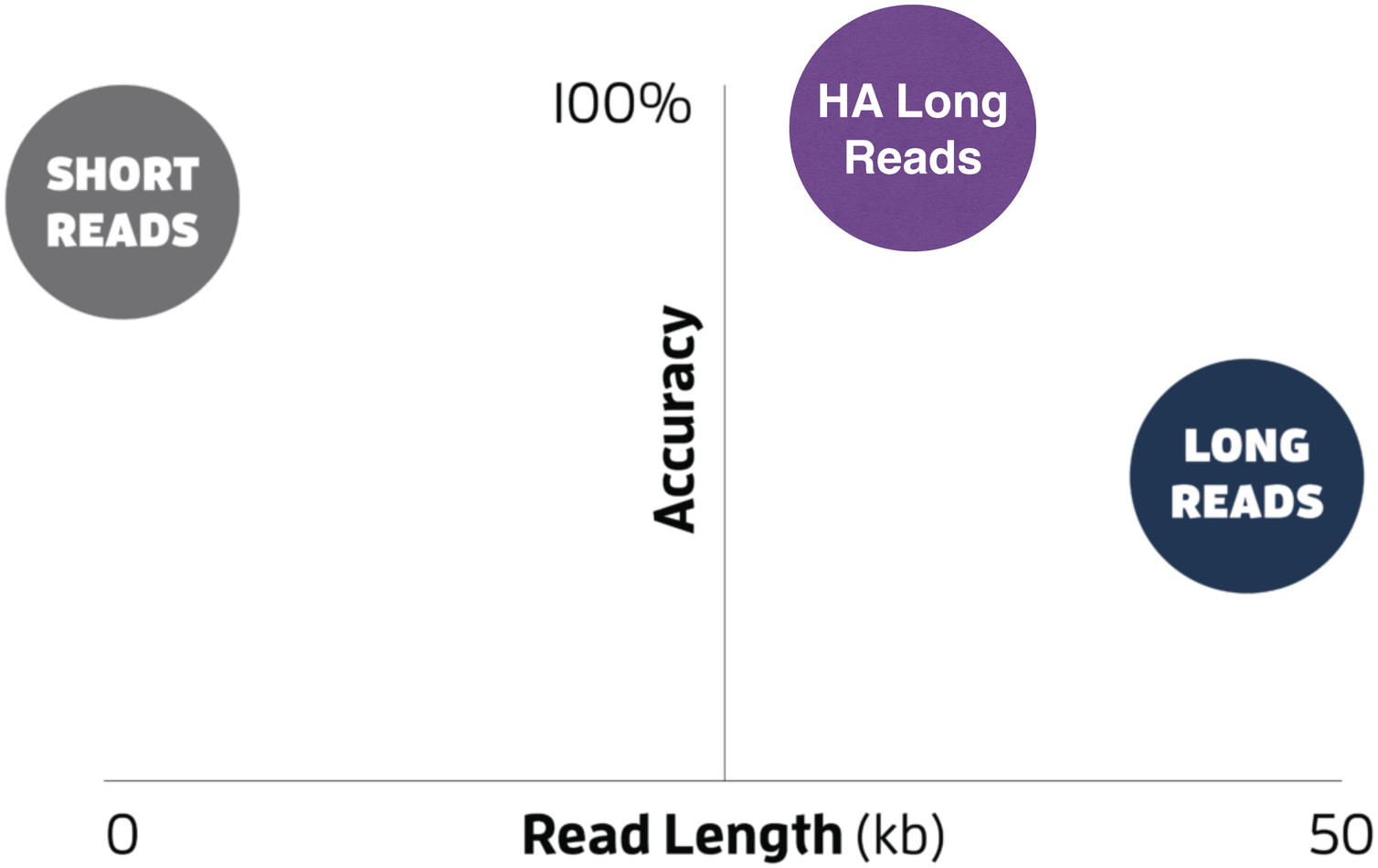
**HA Long Reads (HiFi, LoopSeq)**

Read length: **1 - 20 *kilo*bases**

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

**Image modified from:** PacBio (pacb.com)

# 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 *kilo*bases**

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

**Image modified from:** PacBio (pacb.com)

# 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*

Callahan et al, Nucleic Acids Research, 2019.

# 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**

Callahan et al, Nucleic Acids Research, 2019.

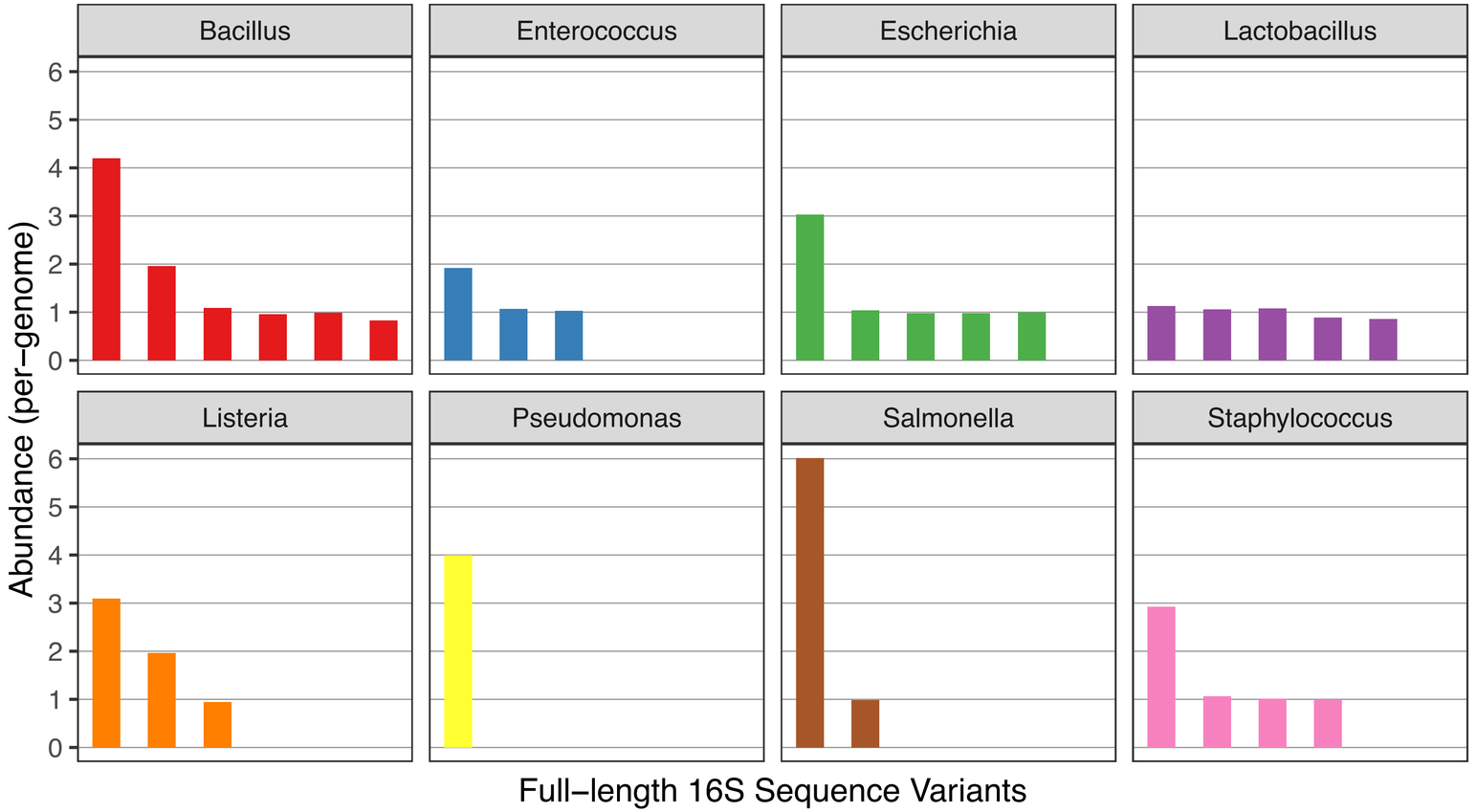
# Full-Length, Full-Complement 16S

## 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**




Callahan et al, Nucleic Acids Research, 2019.

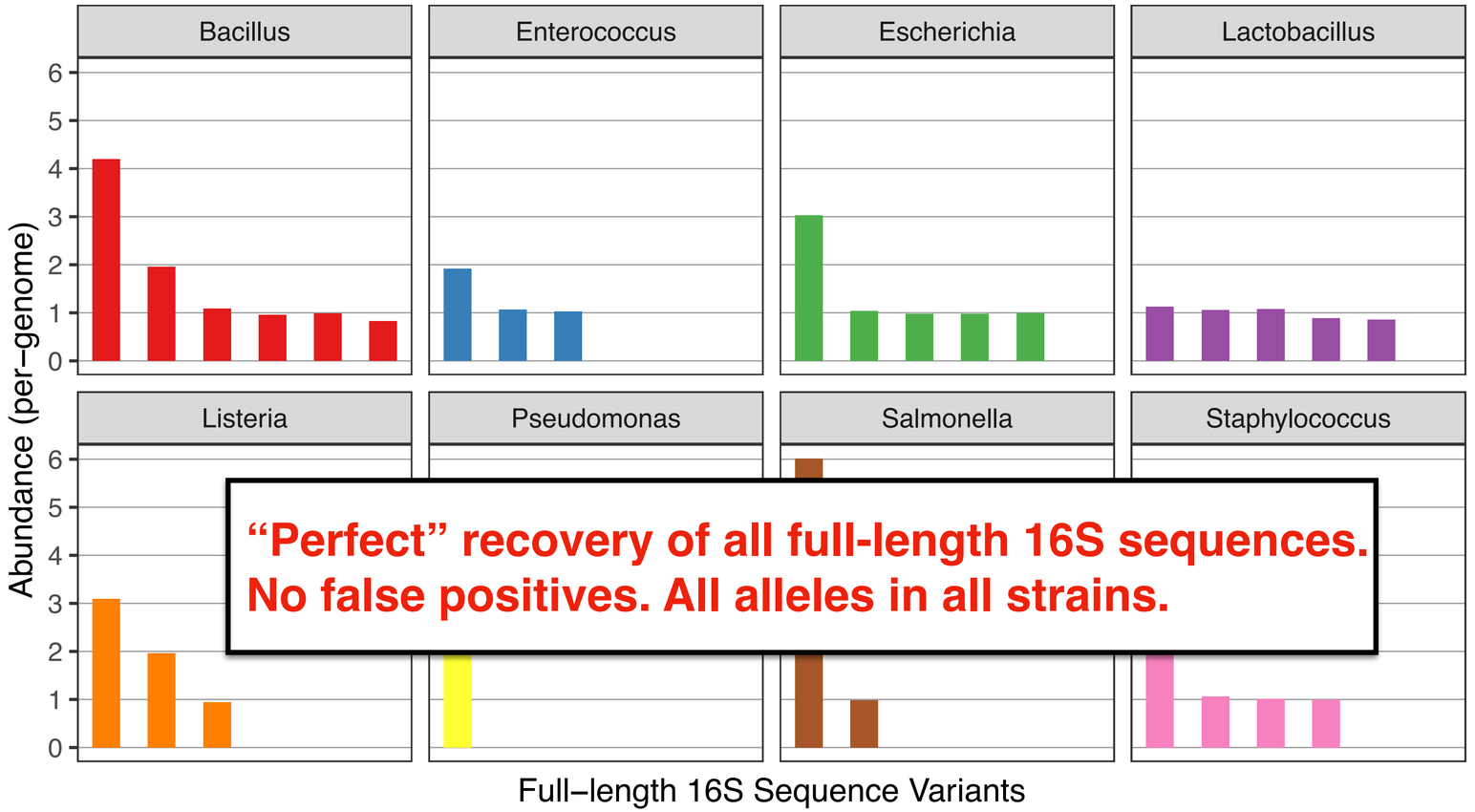
# Full-Length, Full-Complement 16S

## 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**




Callahan et al, Nucleic Acids Research, 2019.

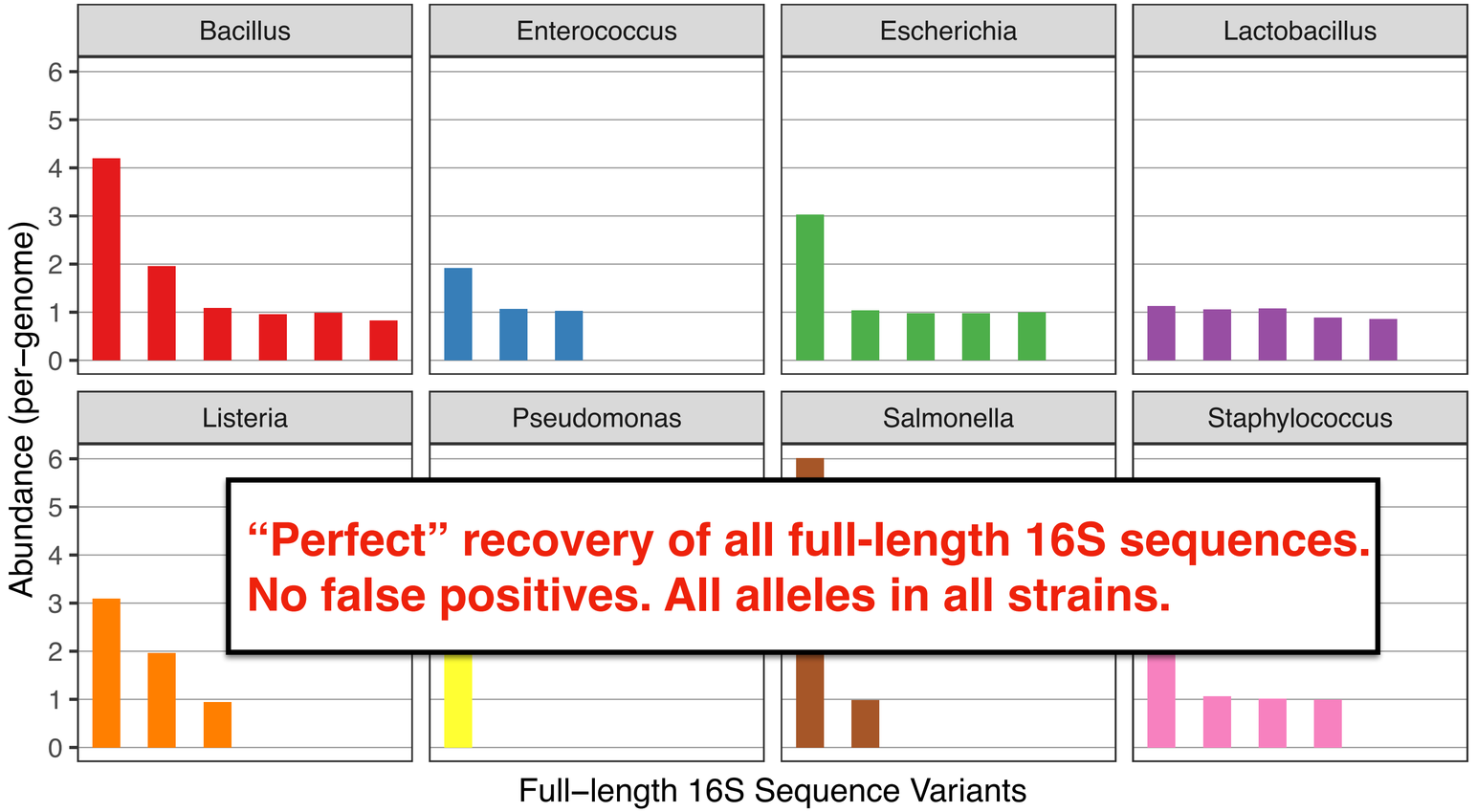
# Full-Length, Full-Complement 16S

## 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***




**"Perfect" recovery of all full-length 16S sequences.**
**No false positives. All alleles in all strains.**

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

Callahan et al, Nucleic Acids Research, 2019.

# Validating Highly-Accurate Long-Read Amplicon Seq

## 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

Callahan et al, Nucleic Acids Research, 2019. Callahan et al, Microbiome, 2021.

# Validating Highly-Accurate Long-Read Amplicon Seq

## 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?**

Callahan et al, Nucleic Acids Research, 2019. Callahan et al, Microbiome, 2021.

# Validating Highly-Accurate Long-Read Amplicon Seq

## 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?**

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

Callahan et al, Nucleic Acids Research, 2019. Callahan et al, Microbiome, 2021.

# Software for Highly-Accurate Long-Read Amplicon Seq

*We support long-read amplicon sequencing!*

# Software for Highly-Accurate Long-Read Amplicon Seq

Fastq

↓

ASV Table or
Species Table (16S)

*We support long-read amplicon sequencing!*

# Software for Highly-Accurate Long-Read Amplicon Seq

Fastq

↓

ASV Table or
Species Table (16S)

*We support long-read amplicon sequencing!*

---

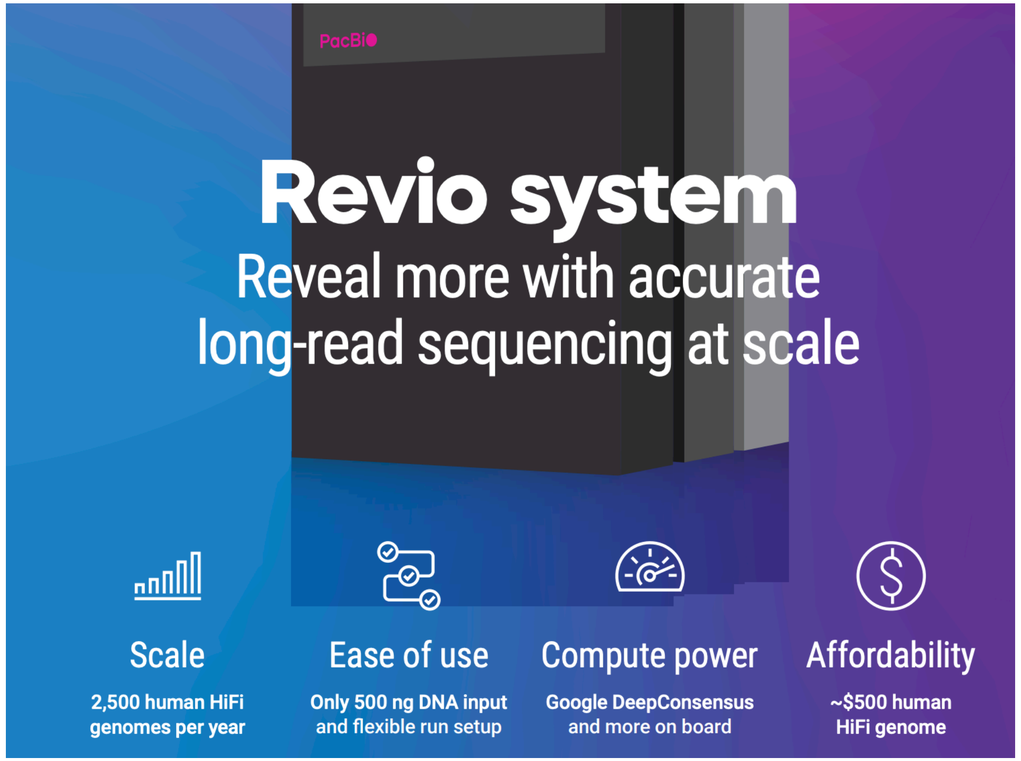
## Long-read specific documentation

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

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

# PacBio at scale: Revio and Kinnex

## Sequencing instrument

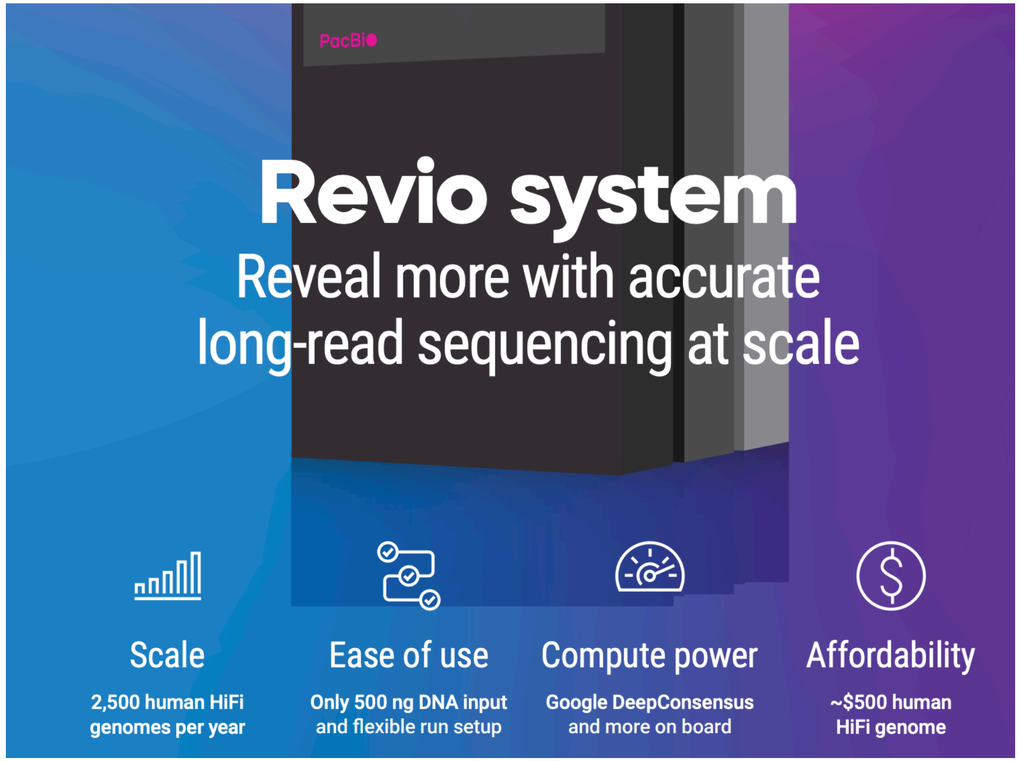



- Higher throughput
- Lower costs
- Binned quality scores

**Image:** PacBio (pacb.com)

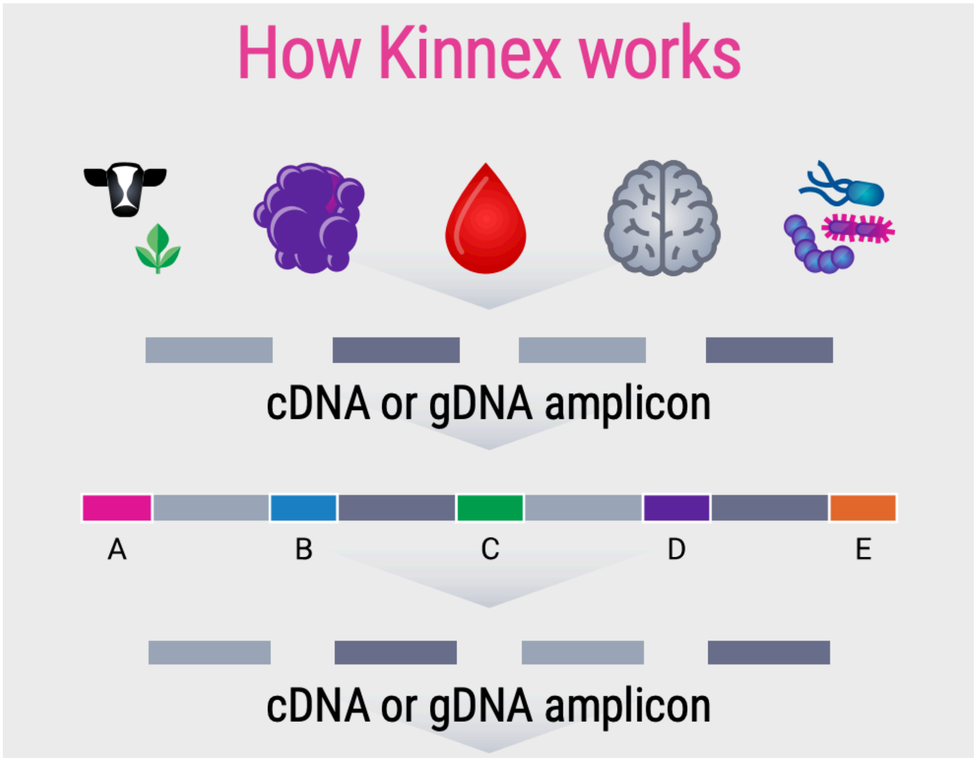
# PacBio at scale: Revio and Kinnex

## Sequencing instrument




- Higher throughput
- Lower costs
- Binned quality scores

## Library preparation

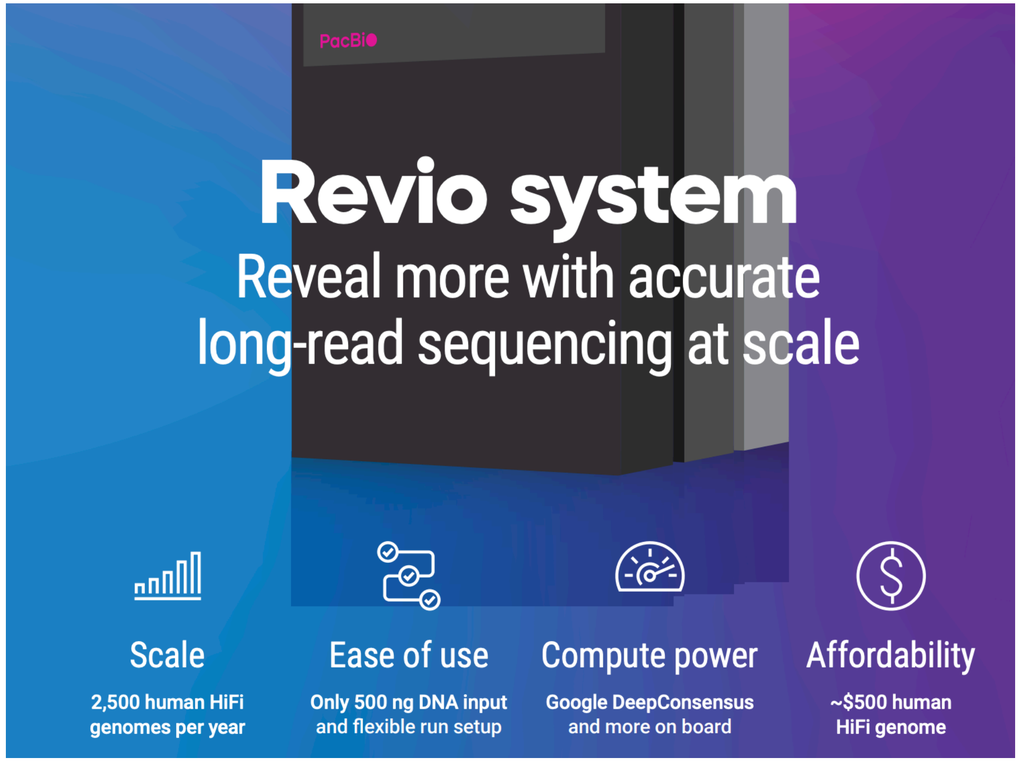



- Higher throughput
- Lower costs
- Lower quality (16S)

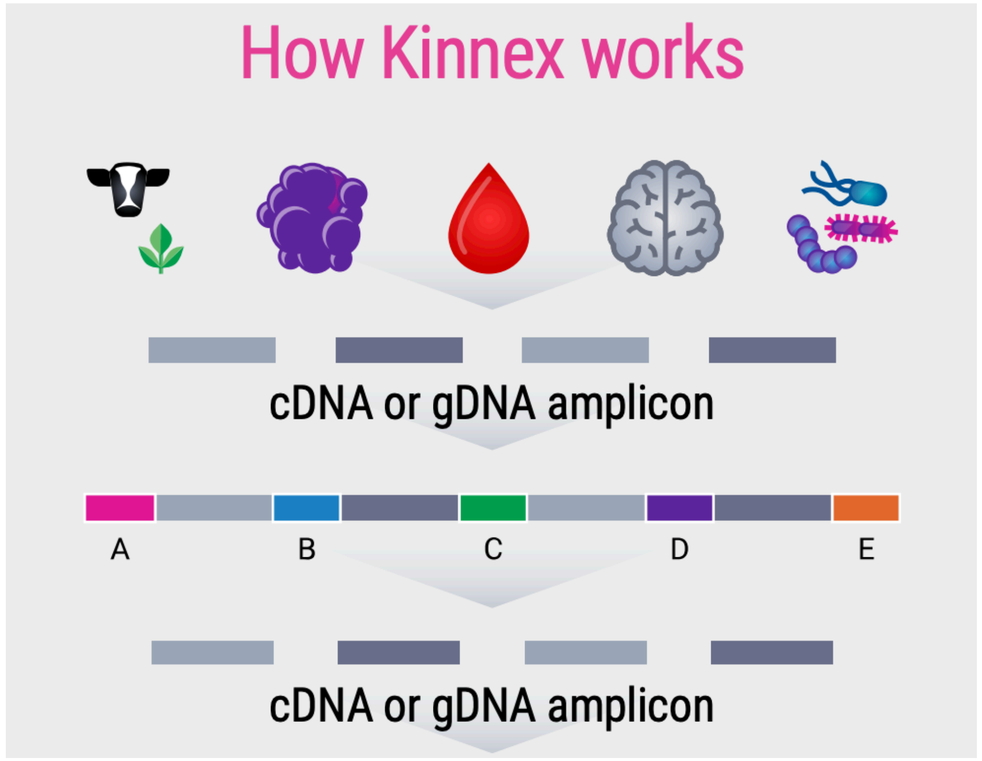
**Image:** PacBio (pacb.com)

# PacBio at scale: Revio and Kinnex

## Sequencing instrument




## Library preparation




- **Higher throughput**
- Lower costs
- **Binned quality scores**

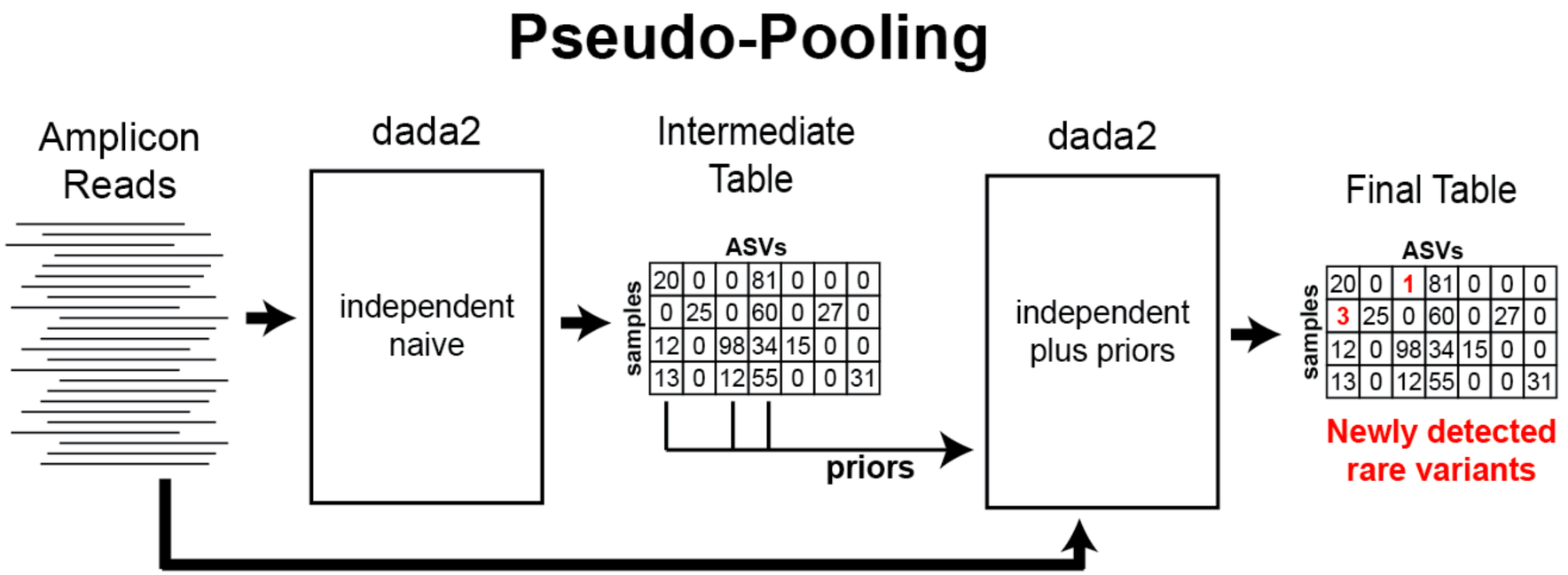
- Higher throughput
- Lower costs
- **Lower quality (16S)**

**Updated methods and guidance for DADA2 worfklow (in progress)**

**Image:** PacBio (pacb.com)

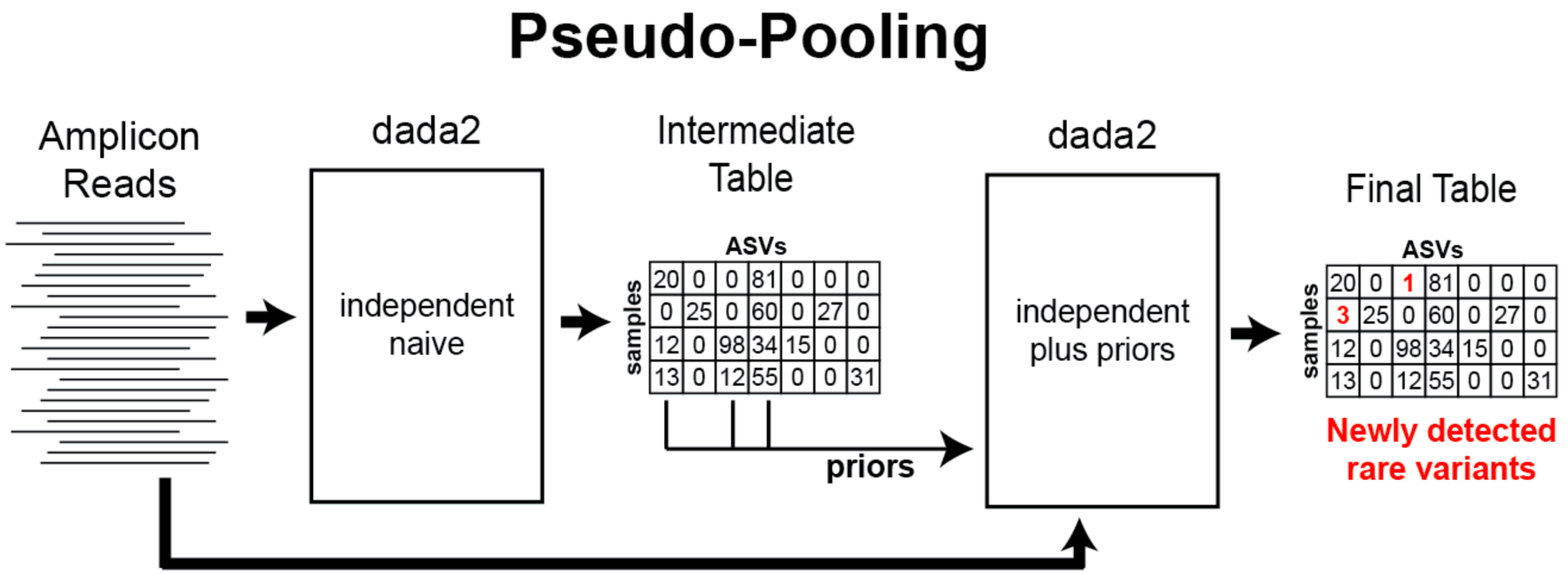
# Software for Highly-Accurate Long-Read Amplicon Seq

1.



# Software for Highly-Accurate Long-Read Amplicon Seq

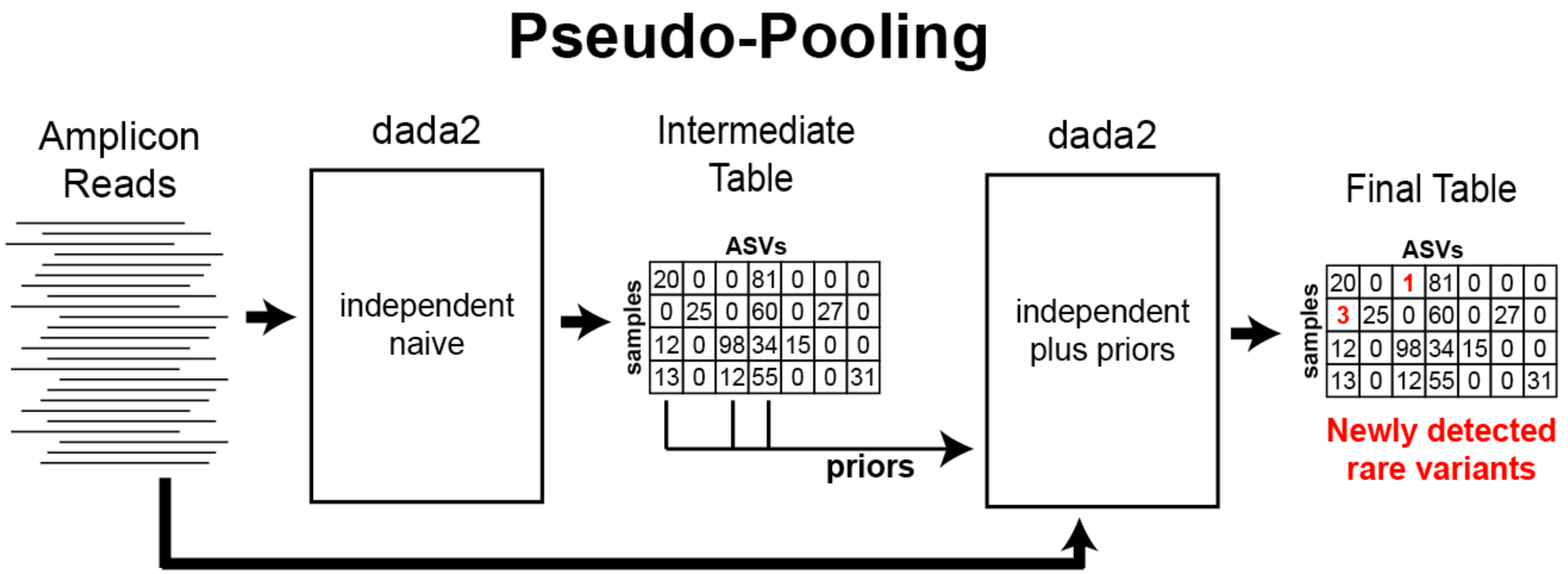
1.




**2. DETECT_SINGLETONS=TRUE**

# Software for Highly-Accurate Long-Read Amplicon Seq

1.

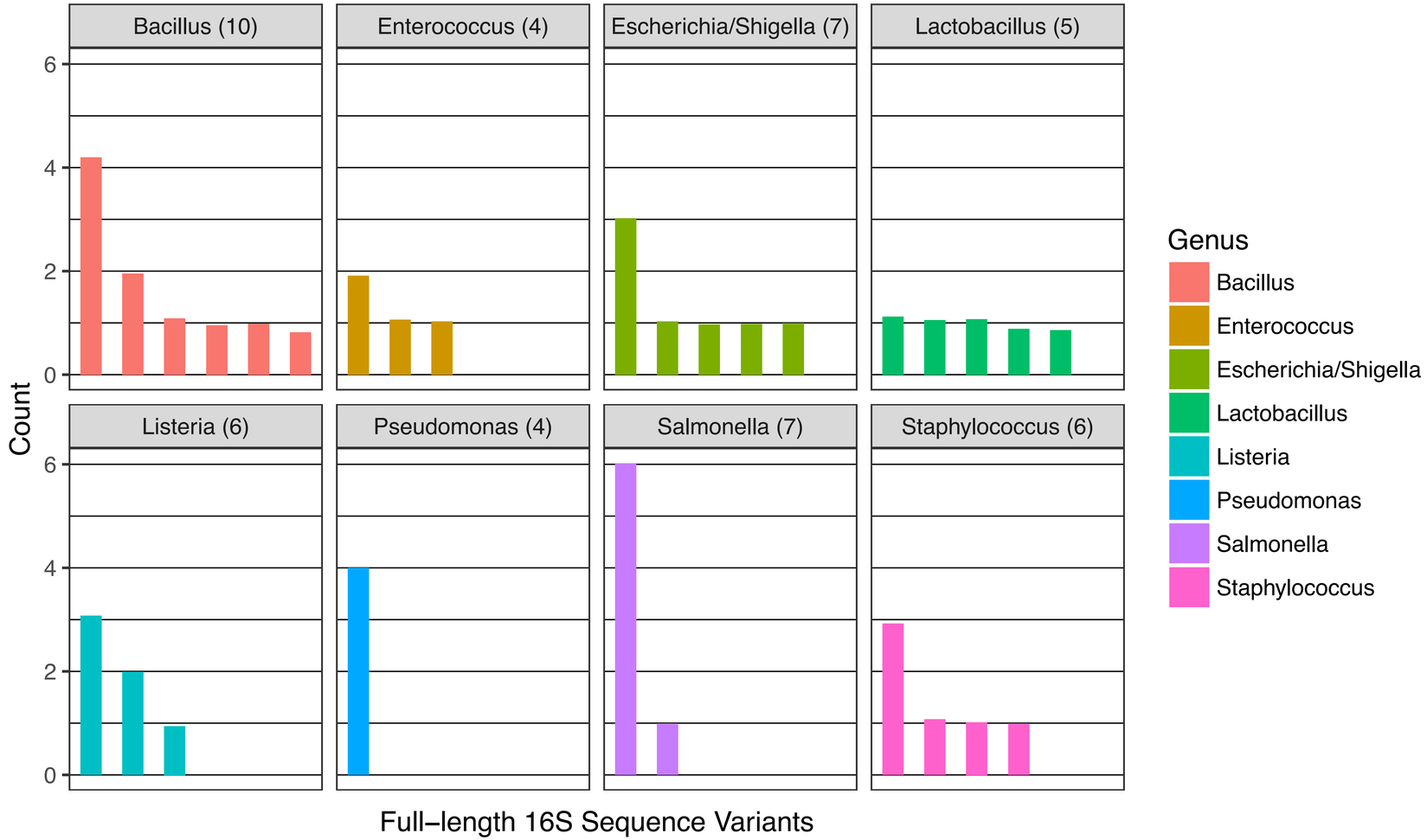



2. DETECT_SINGLETONS=TRUE

3. Long-read specific parameters.

# Sub-species bacterial profiling

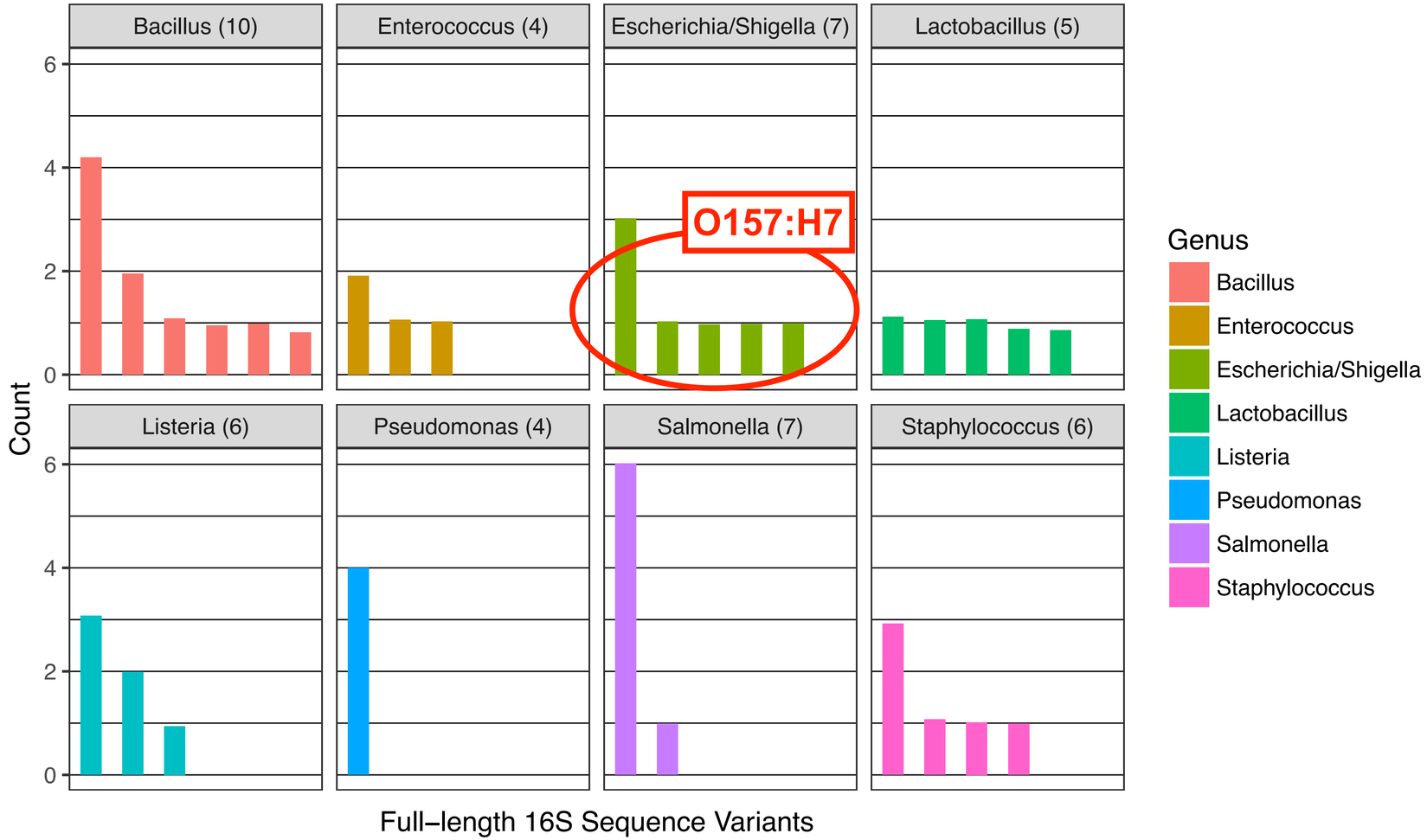
*Full-length 16S from Zymo mock community*




Callahan et al, Nucleic Acids Research, 2019.

# Sub-species bacterial profiling

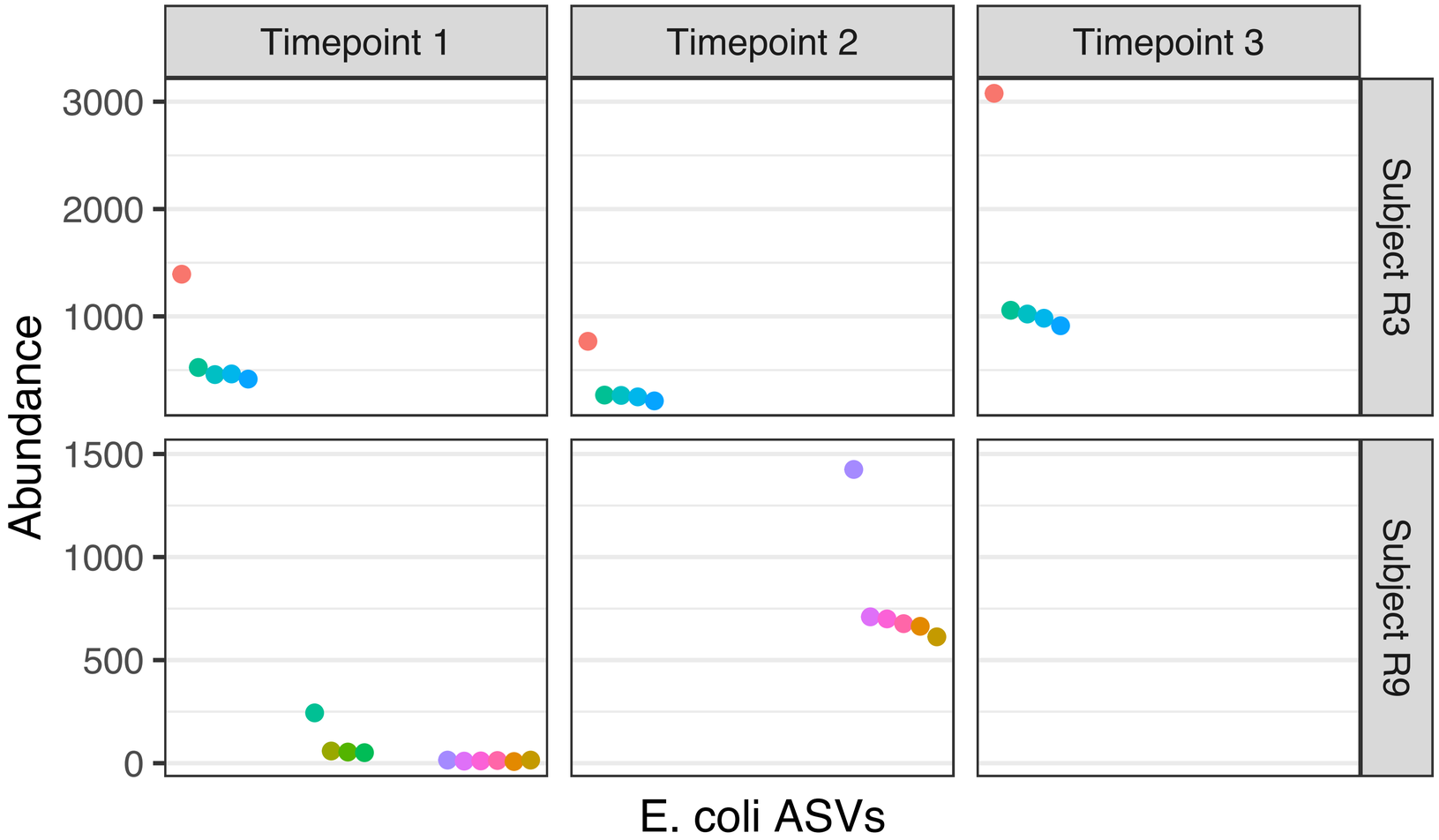
*Full-length 16S from Zymo mock community*




Callahan et al, Nucleic Acids Research, 2019.

# Sub-species bacterial profiling

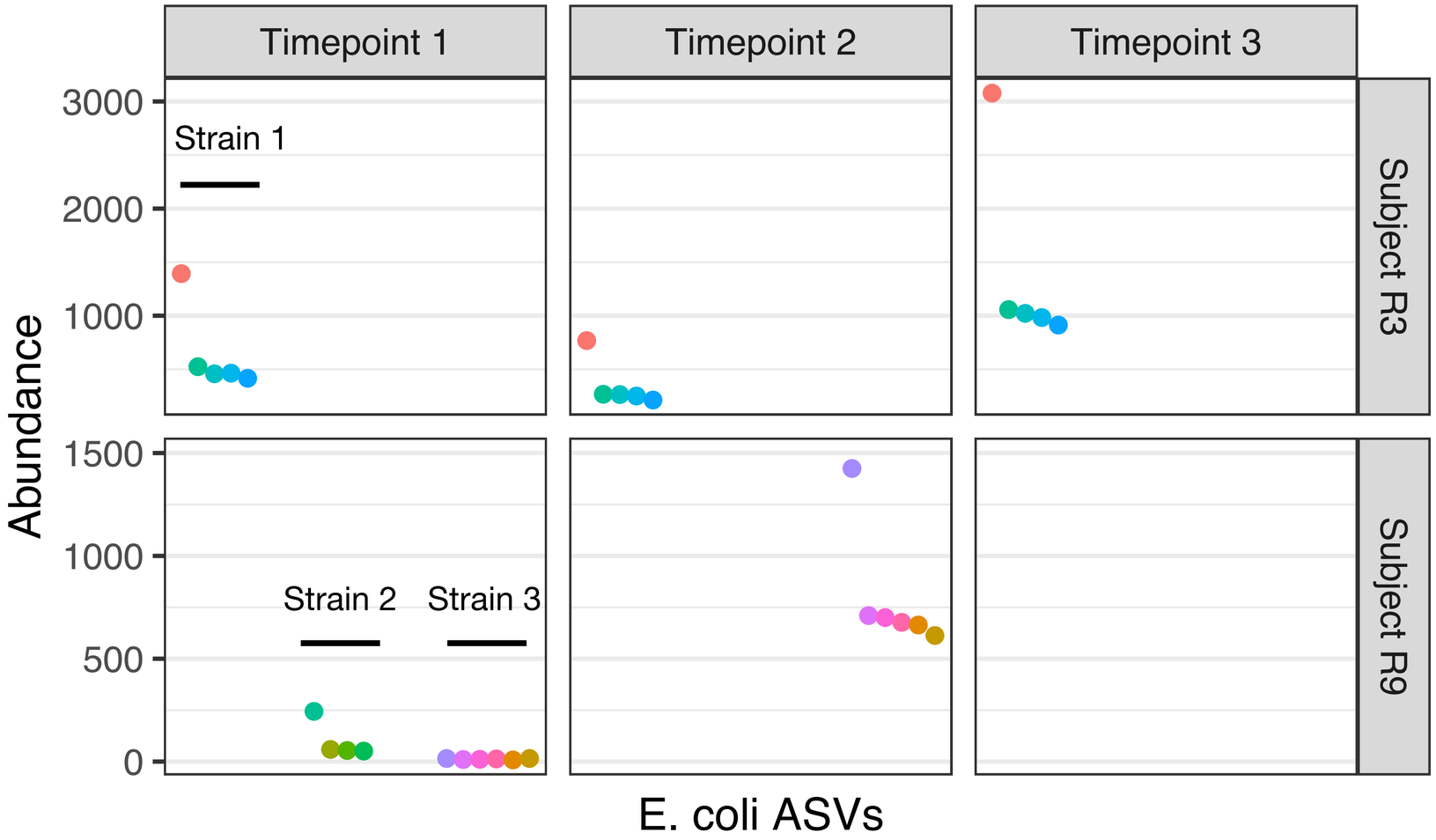
*Full-length 16S from human fecal samples*




Callahan et al, Nucleic Acids Research, 2019.

# Sub-species bacterial profiling

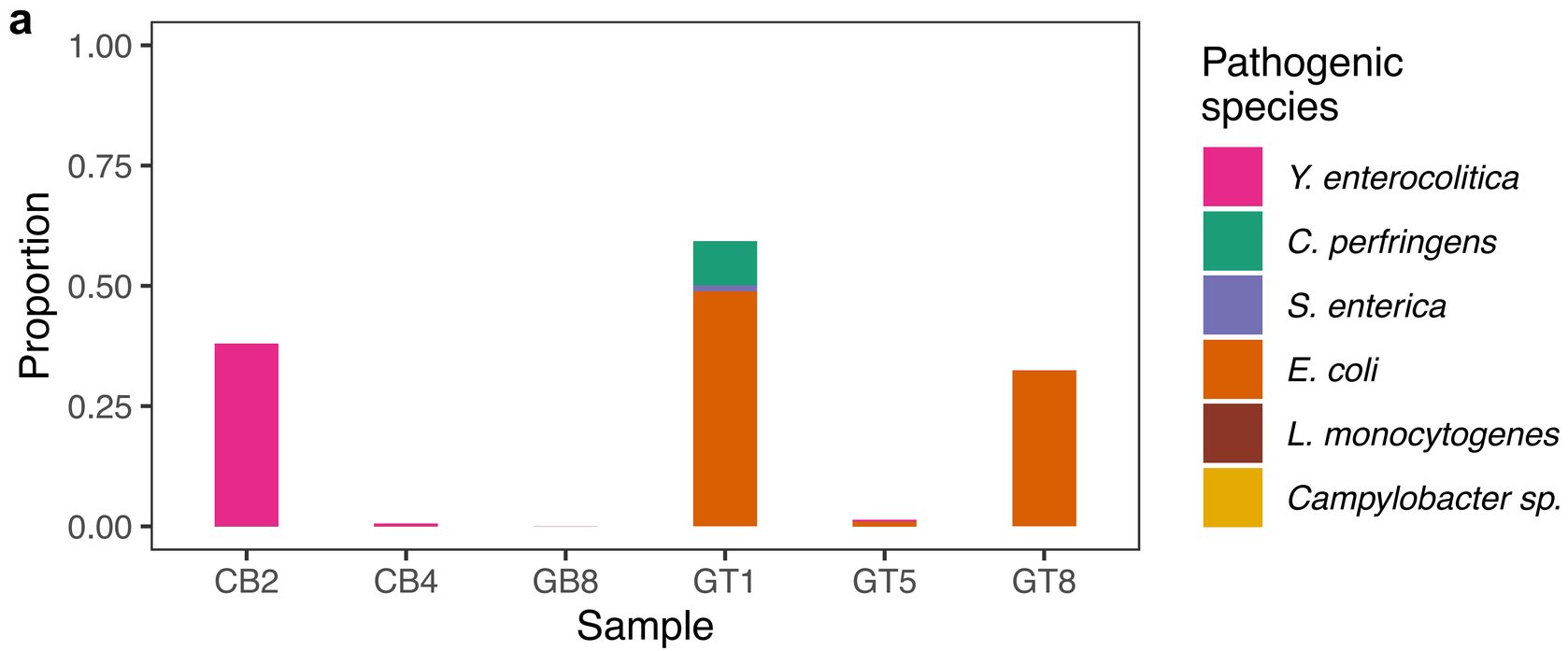
*Full-length 16S from human fecal samples*




Callahan et al, Nucleic Acids Research, 2019.

# Sub-species bacterial profiling

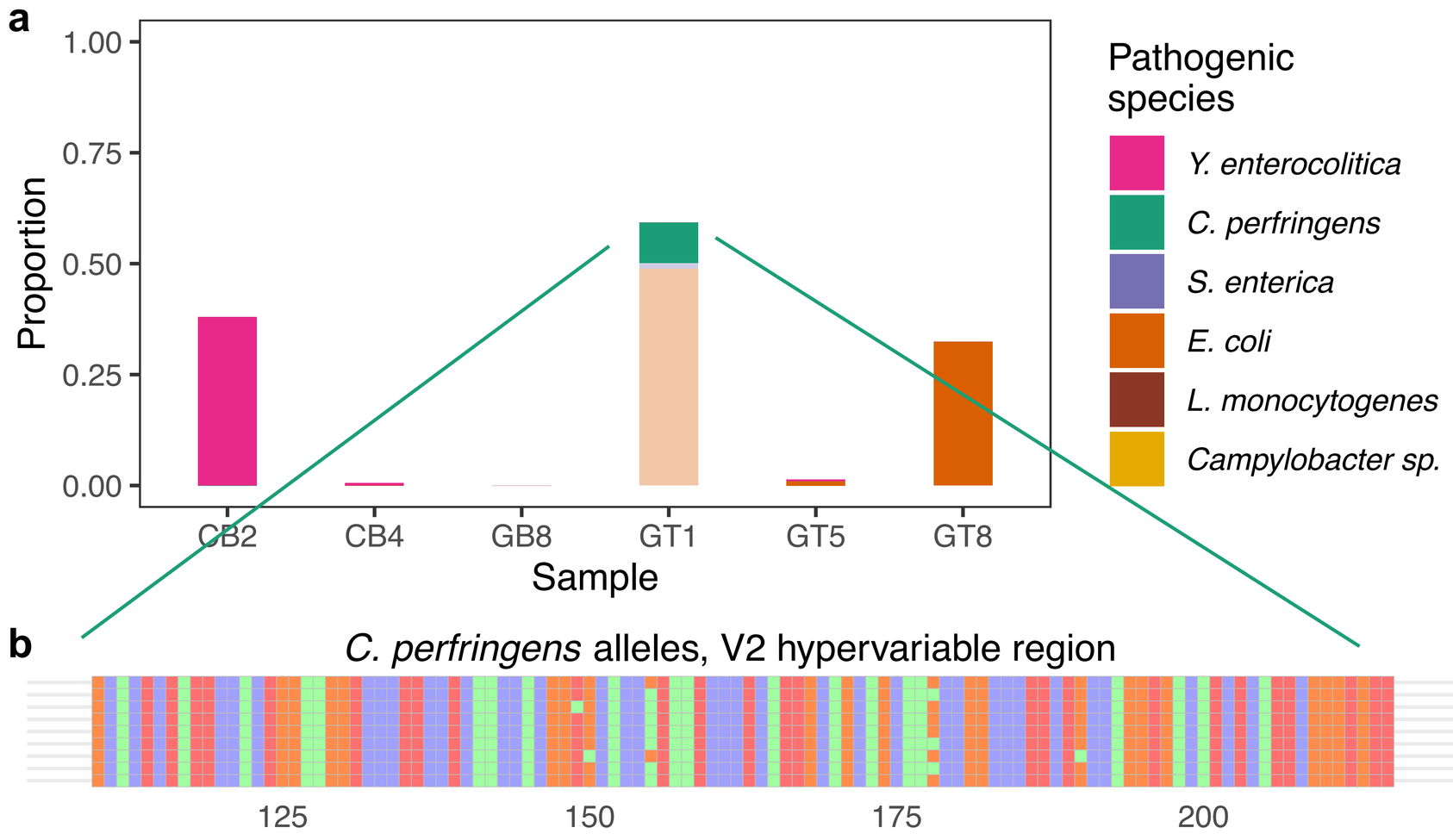
*Full-length 16S from retail meat samples*




Callahan et al, Microbiome, 2021.

# Sub-species bacterial profiling

*Full-length 16S from retail meat samples*




**Full genomic complement of 16S alleles**

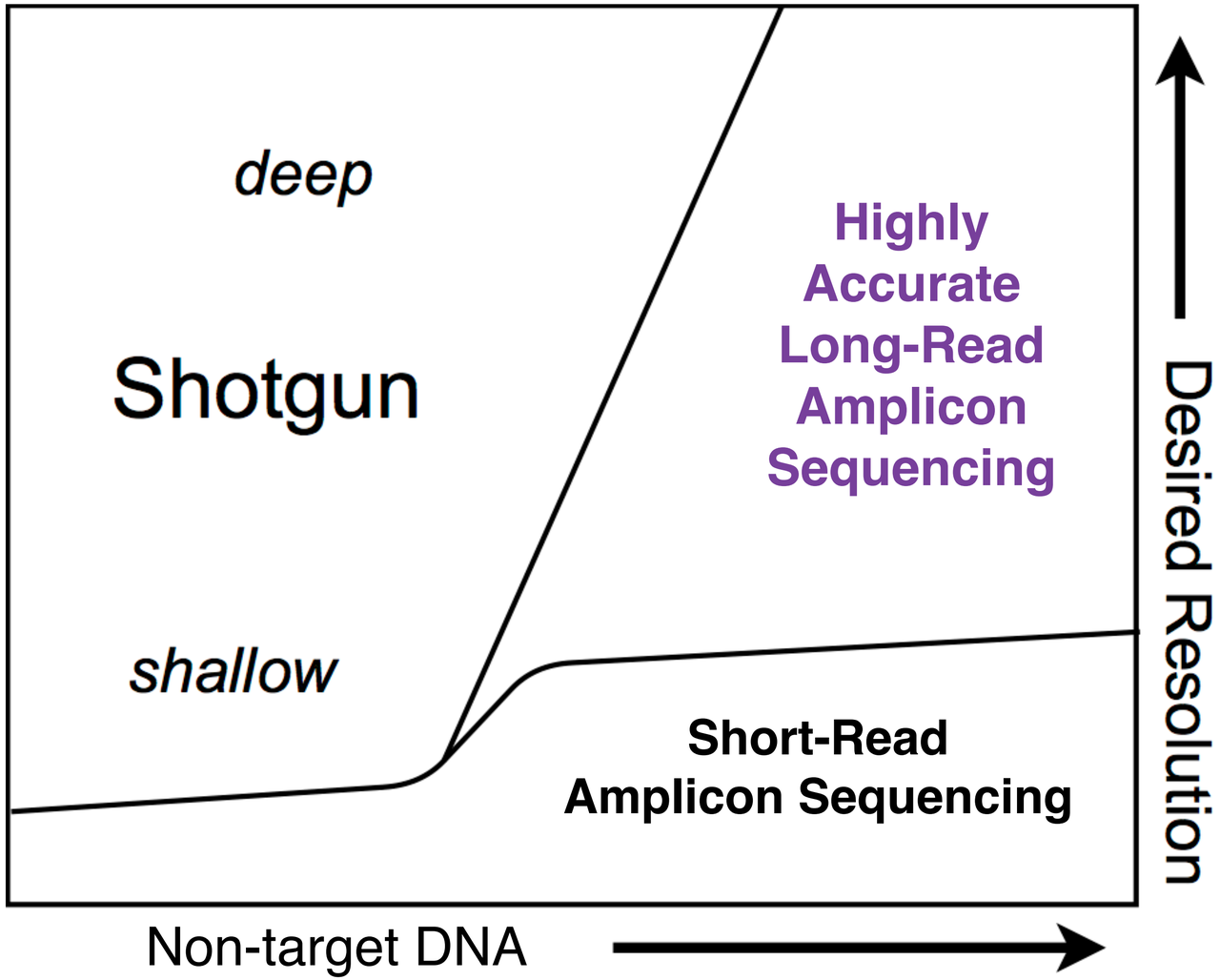
Callahan et al, Microbiome, 2021.

# 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)
- …

# Highly Accurate Long-Read Amplicon Seq



# Highly Accurate Long-Read Amplicon Seq

**Limitations**

- PCR Biases
- PCR Length Scaling (past ~5 kilo bases)
- Cost-per-base (big improvements recently!)
- Chimeras
- **Tooling (especially sub-species assignment)**
- …

# Highly Accurate Long-Read Amplicon Seq

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

Check for updates

| Computational Biology | Research Article

## Serovar-level identification of bacterial foodborne pathogens from full-length 16S rRNA gene sequencing

Dmitry Grinevich,[1] Lyndy Harden,[1] Siddhartha Thakur,[1] Benjamin Callahan[1,2]

**Dmitry Grinevich**

Sid Thakur and lab

USDA

National Institute of Food and Agriculture

# Acknowledgements


**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