PACIFIC BIOSCIENCES®

Introduction to PacBio HiFi data and it's applications

Kristin Mars, Sequencing Applications Specialist

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2018 by Pacific Biosciences of California, Inc. All rights reserved.

PACBIO®

PacBio long-read sequencing: How it works

PACBIO LONG-READ SEQUENCING

Underlying technology: Single Molecule, Real-Time (SMRT) Sequencing

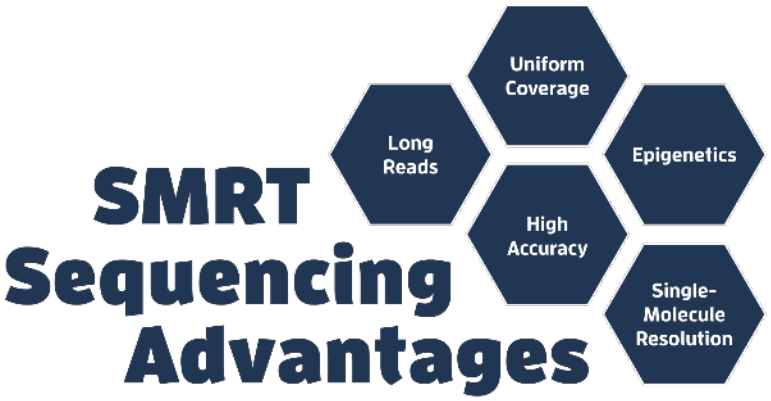
Long Reads
- Tens of kilobases
- Sequence from 500 bp to >50,000 bp inserts

High Accuracy
- Free of systematic errors
- Achieves >99.999% (Q50) consensus accuracy

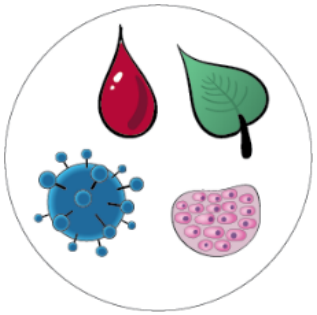
Single-Molecule Resolution
- Sequence DNA or RNA
- Long reads with ≥Q20 (99%) single-molecule accuracy

Uniform Coverage
- No DNA amplification
- Least GC content and sequence complexity bias

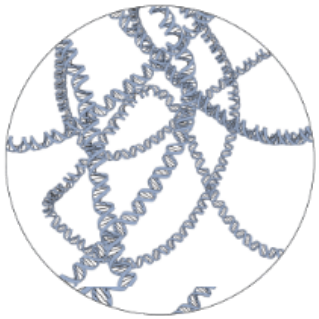
Simultaneous Epigenetic Detection
- Characterize epigenome
- No separate sample preparation required



FROM SAMPLE TO SMRT SEQUENCING



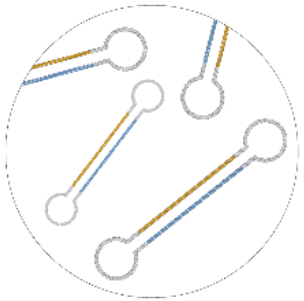
From viruses to vertebrates



Isolate DNA or RNA

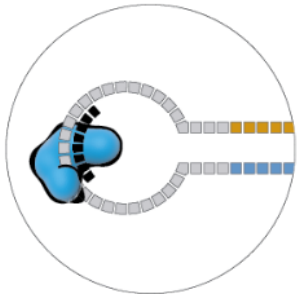
Ligate adapters

\+



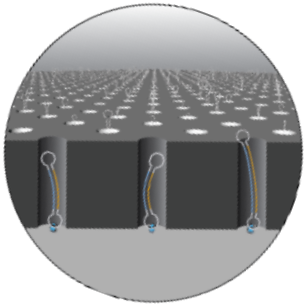
Generate SMRTbell libraries

\+ Primer & Polymerase



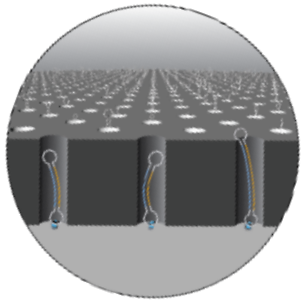
Prepare sequencing reaction

Use PacBio Sequel Systems to sequence genomes, transcriptomes, and epigenomes

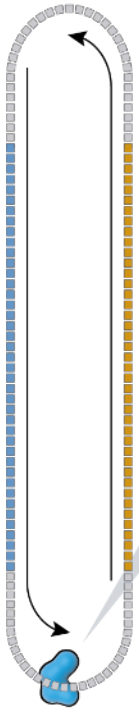


SMRT Cells contain millions of zero-mode waveguides (ZMWs)

OUR CORE TECHNOLOGY



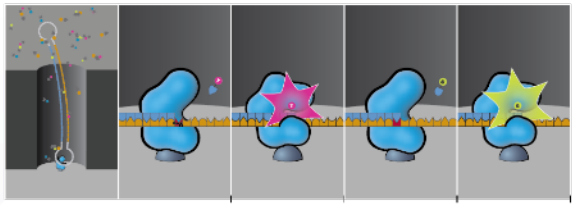
SMRTbell templates enable repeated sequencing of circular template with real-time base incorporation



Single-Molecule Resolution

A single molecule of DNA is immobilized in each ZMW

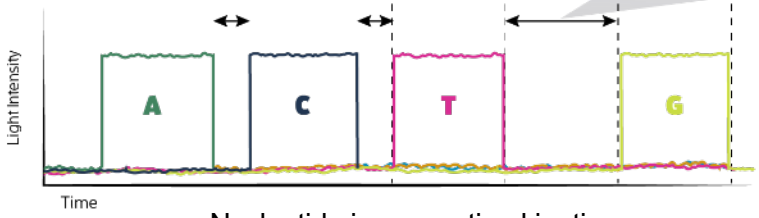
+ Phospholinked nucleotides



As anchored polymerases incorporate labeled bases, light is emitted

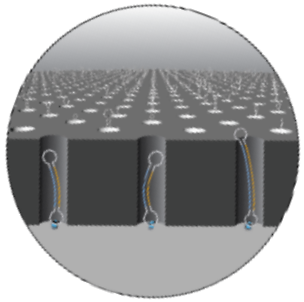
Epigenetics

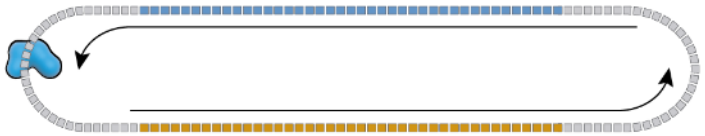
Directly detect DNA modifications during sequencing



Nucleotide incorporation kinetics are measured in real time

GENERATE HIGHLY ACCURATE LONG READS





Long Reads

Long reads space large regions for improved assembly, variant detection and haplotype phasing

Polymerase read

Sub reads

High Accuracy

Calling consensus from subreads increases accuracy

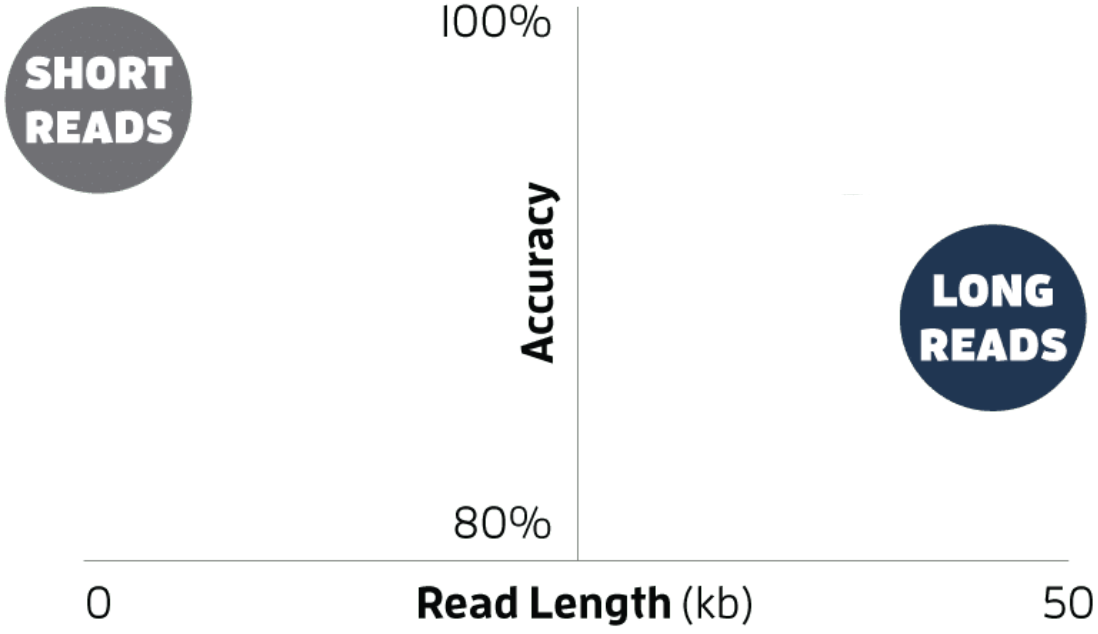
Uniform Coverage

Uniform sequencing coverage through low-complexity regions with no amplification bias

Highly Accurate Long Read

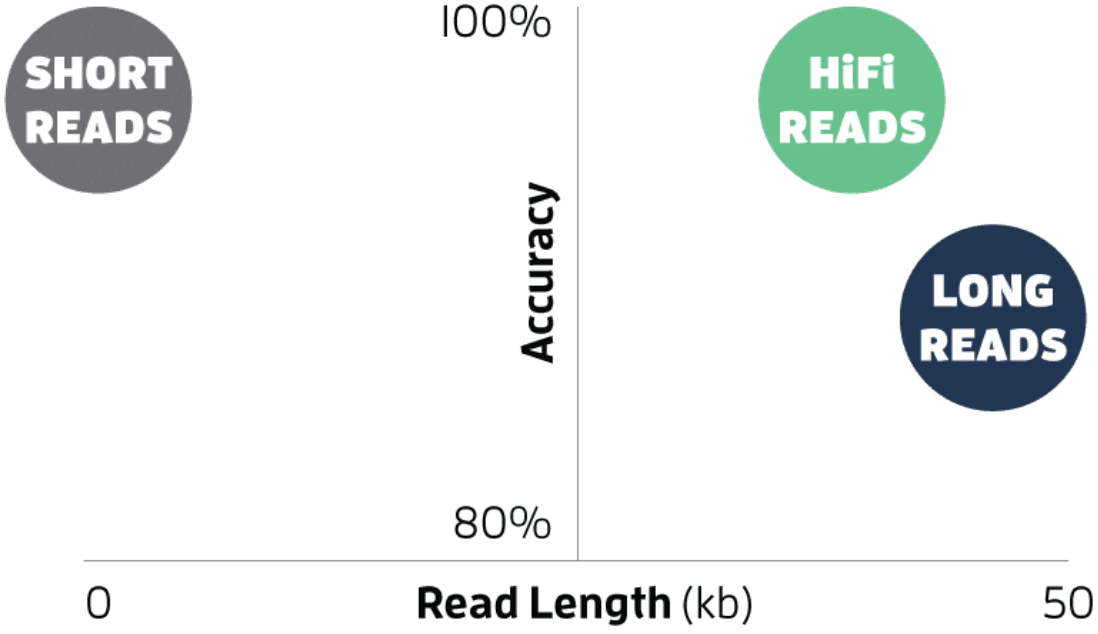
THE OLD PARADIGM:

DNA Sequence Reads are Long OR Accurate

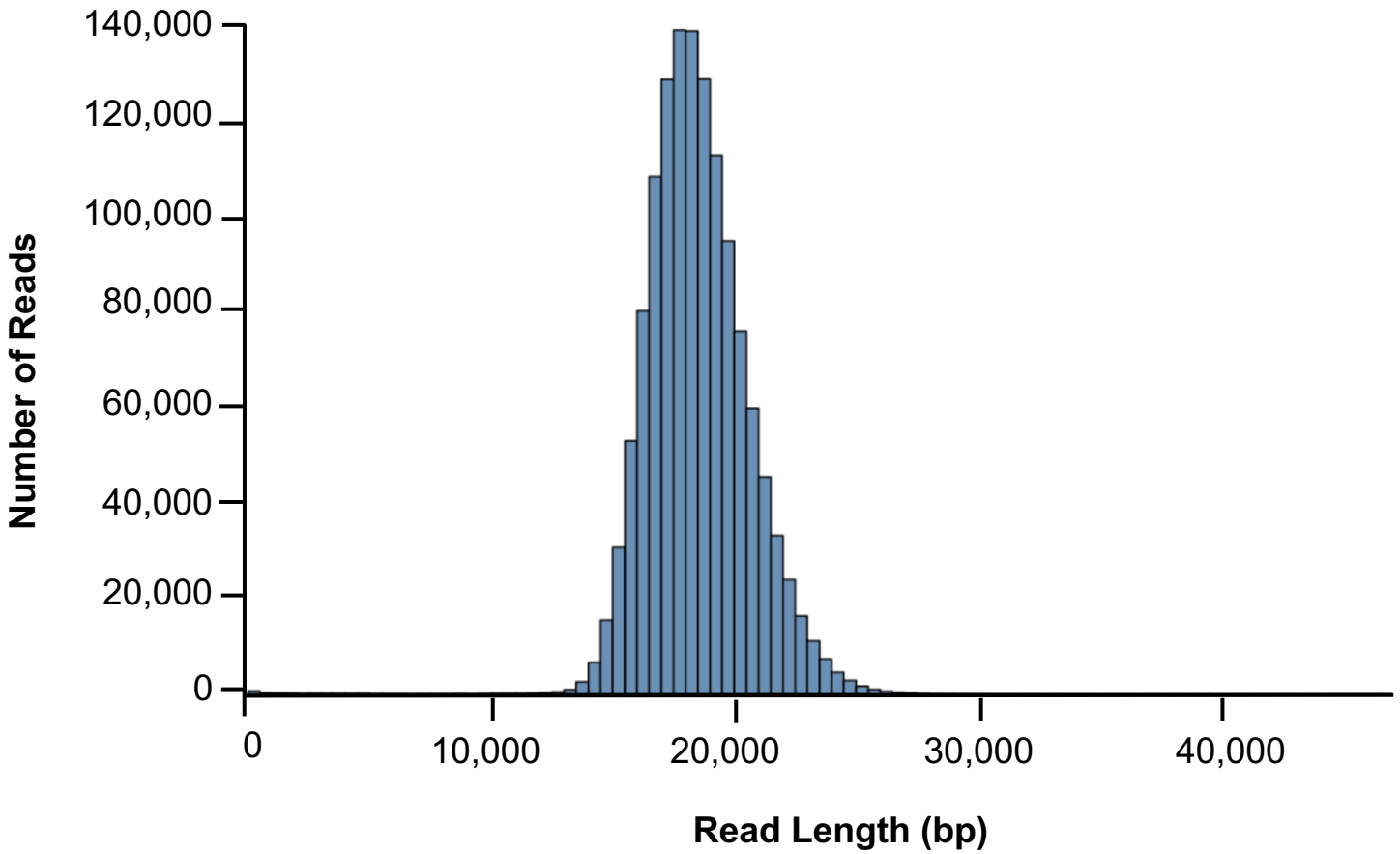


THE NEW PARADIGM:

HiFi Reads are Long AND Accurate

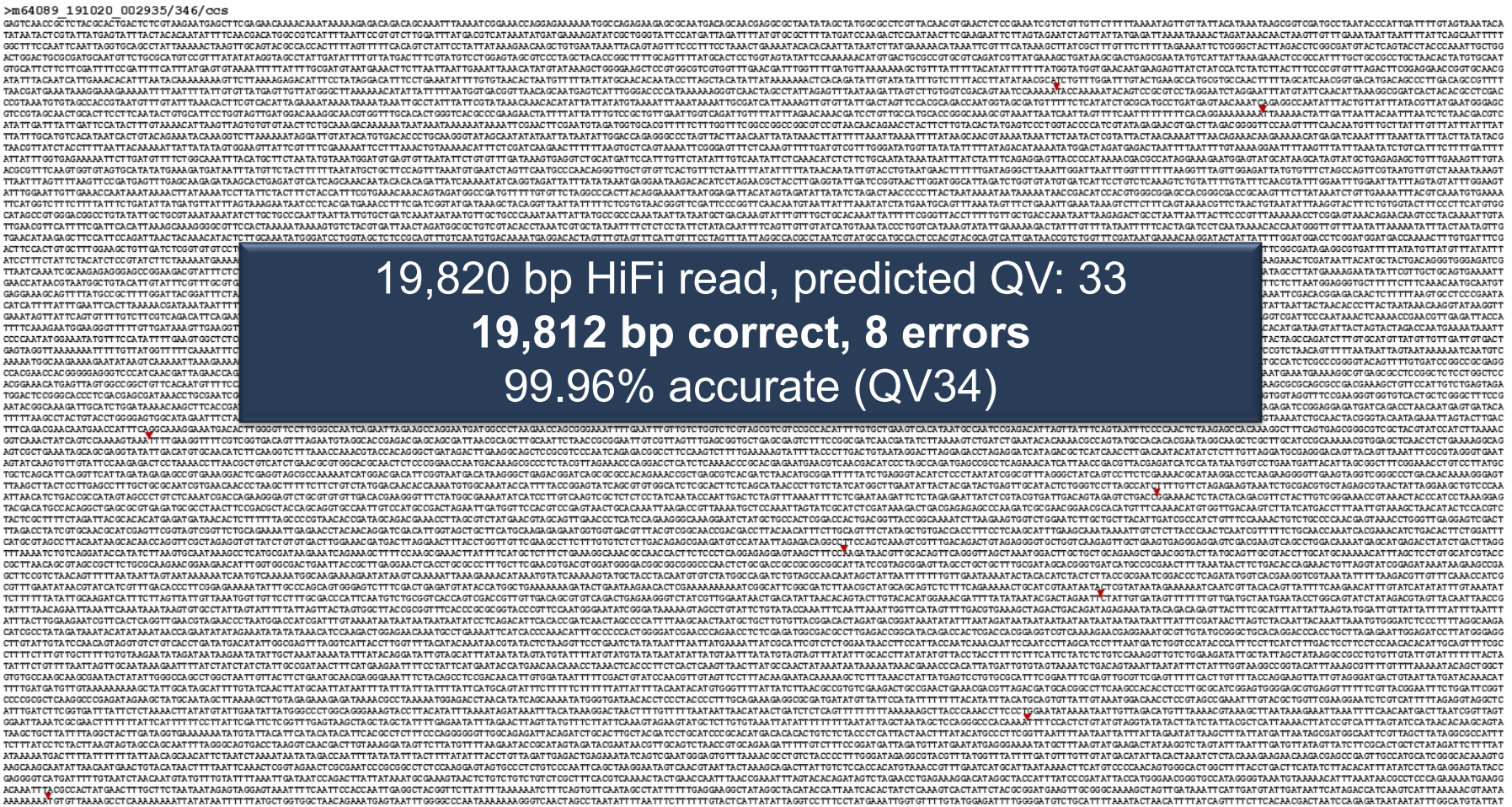


HOW LONG ARE HIFI READS?



Data shown above from a 20 kb size-selected human library using the SMRTbell Template Prep Kit on a Sequel II System (2.0 Chemistry, Sequel II System Software v8.0, 30-hour movie). Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

WHY DOES ACCURACY MATTER?



WHY DOES ACCURACY MATTER?

Base-level accuracy allows you to differentiate between real variation and errors.

Human (HG001, HG002)	Accuracy	Complete RefSeq Genes (n=19,313)
PacBio HiFi Reads 22-fold	**Q51**	**99.5%**
Nanopore 47-fold	Q25	44.5%
Nanopore 30-fold (+ short read polishing)	Q34	98.6%

Rice (MH63, Basmati)	BUSCO Complete Genes (n=1,440)	Frameshifted Genes (n=35,666)
PacBio HiFi Reads 20-fold	**98.7%**	**444**
Nanopore 62-fold (+ short read polishing)	97.6%	1,379

Only HiFi reads give you the accuracy needed to see more complete and frameshift-free genes

WHAT CAN YOU DO WITH HIGHLY ACCURATE LONG READS (HIFI READS)?

- Highly accurate *de novo* assembly
- Detect all variants types with high precision and recall
  - Detect 5% more variants in “medical exome”
- Phase variants into haplotypes
- Sequence full-length transcripts
- Explore metagenomes in high resolution

WHOLE GENOME SEQUENCING

VARIANT DETECTION

RNA SEQUENCING

COMPLEX POPULATIONS

TARGETED SEQUENCING

EPIGENETICS

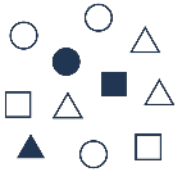
HIFI SEQUENCING APPLICATIONS

Many applications can be completed with a **single SMRT Cell 8M**

WHOLE GENOME SEQUENCING

RNA SEQUENCING

TARGETED SEQUENCING



COMPLEX POPULATIONS

One SMRT Cell 8M

pacb.com/OneSMRTCell

PACBIO®

Comprehensive Whole Genome *De Novo* Assemblies

Resolving repetitive regions and heterozygosity

UNIQUE CHALLENGES OF PLANT & ANIMAL GENOMES

Size and Complexity

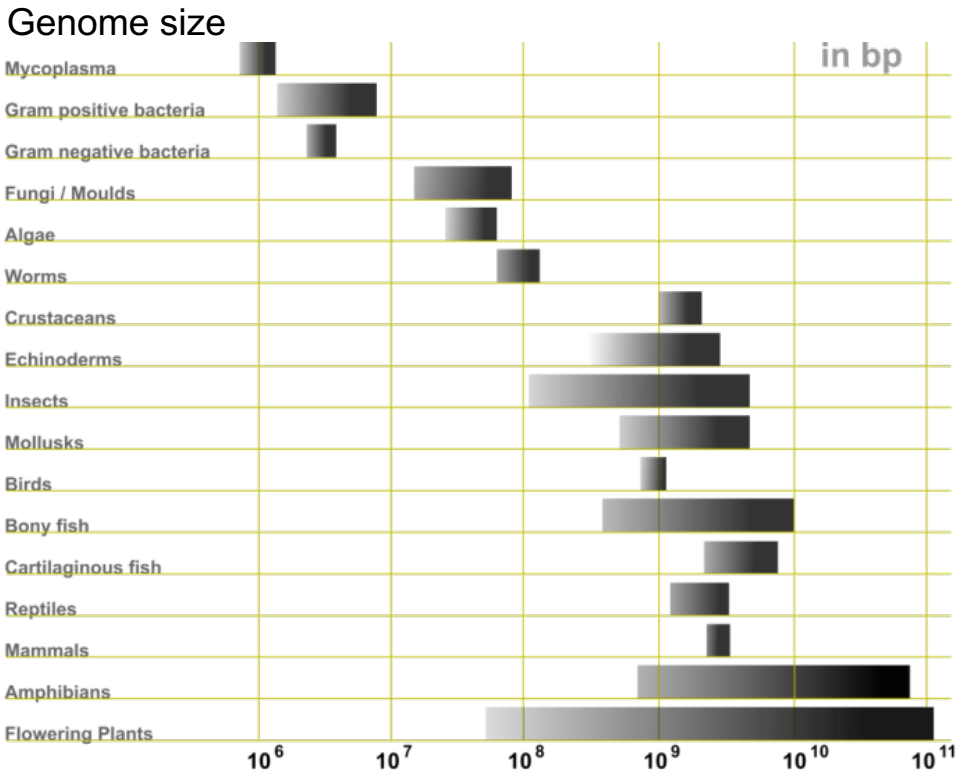
- Loblolly pine, 21 Gb (>6-fold human genome)
- Wheat, 17 Gb, hexaploid
- Sugarcane, 10- or 12-ploid

Extreme Repeat Content

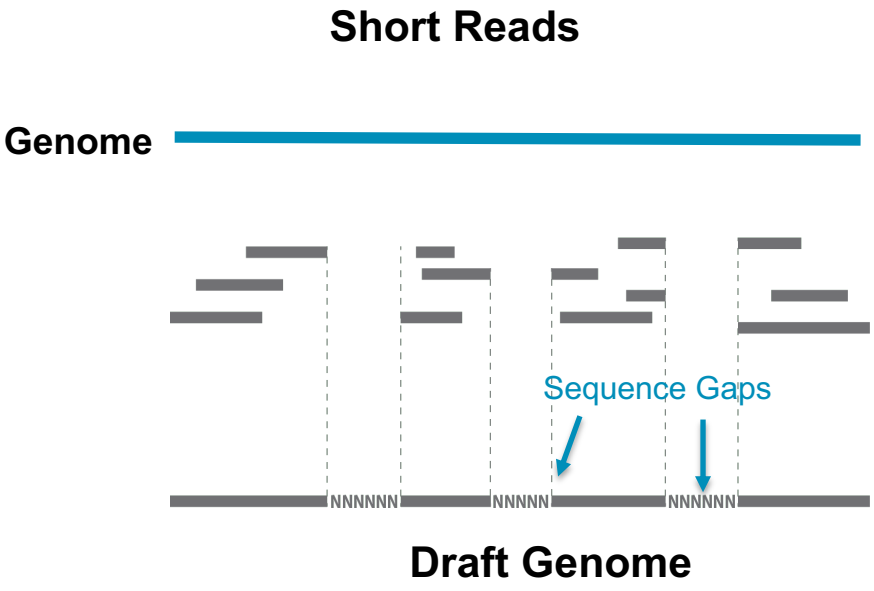
- Maize >60%
- Wheat >80%

Each Project is Unique

- Ranges in size, ploidy, and repeat content
- Custom strategy is commonly needed



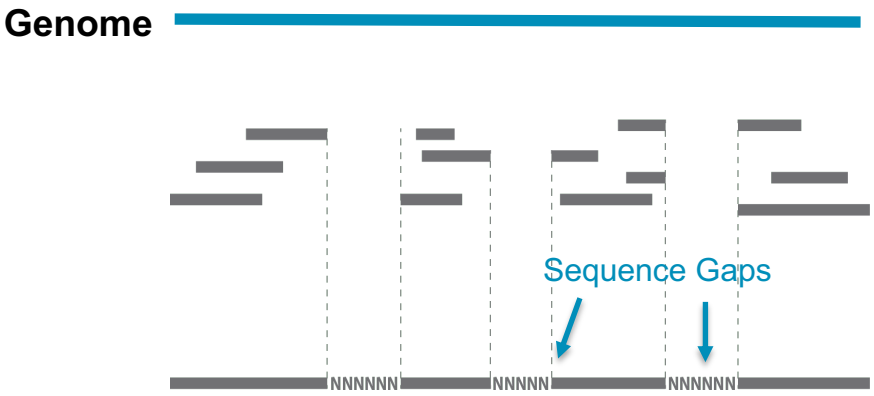
DRAFT VS COMPLETE GENOME ASSEMBLY



Missing sequencing leads to missed genes and limits biological interpretation

DRAFT VS COMPLETE GENOME ASSEMBLY

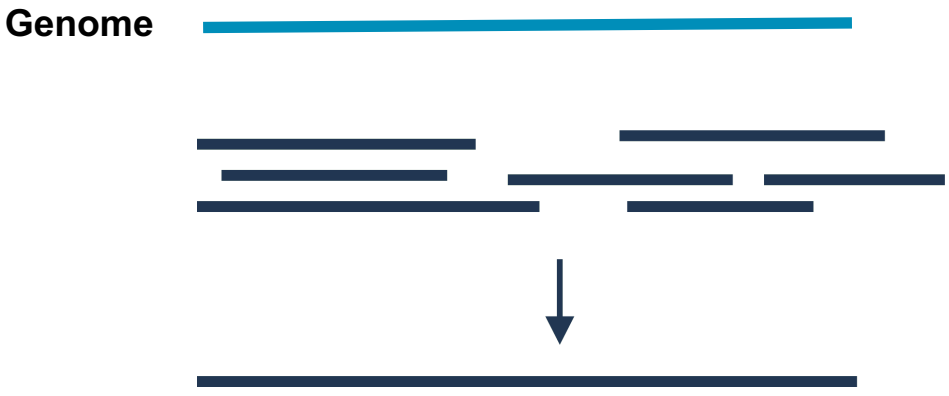
Short Reads



Draft Genome

Missing sequencing leads to missed genes and limits biological interpretation

HiFi Reads

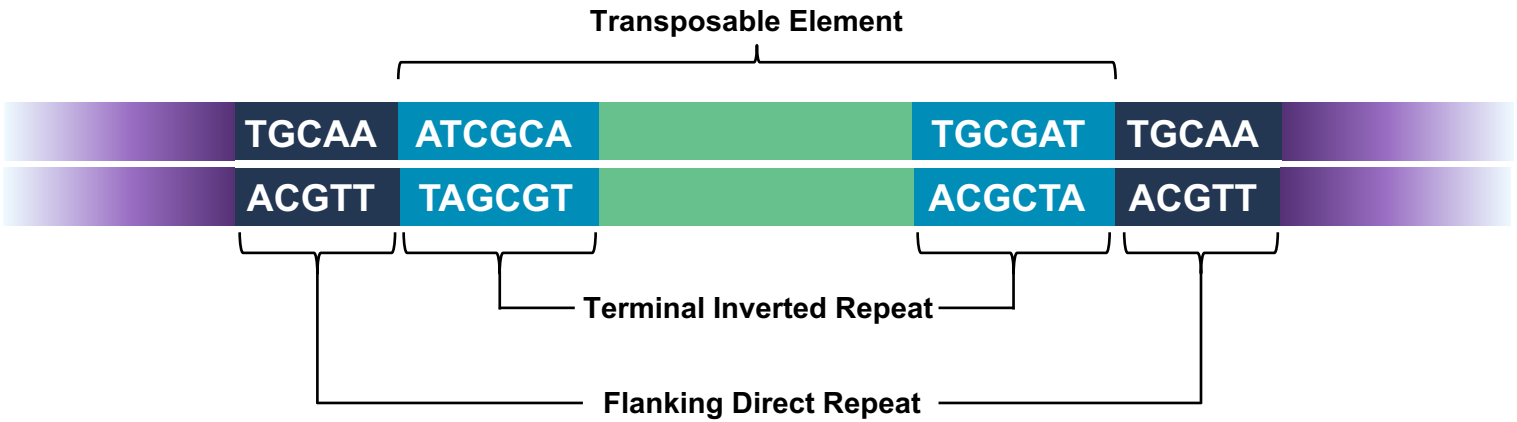


Complete Genome

A comprehensive structural, functional and organizational picture of the genome

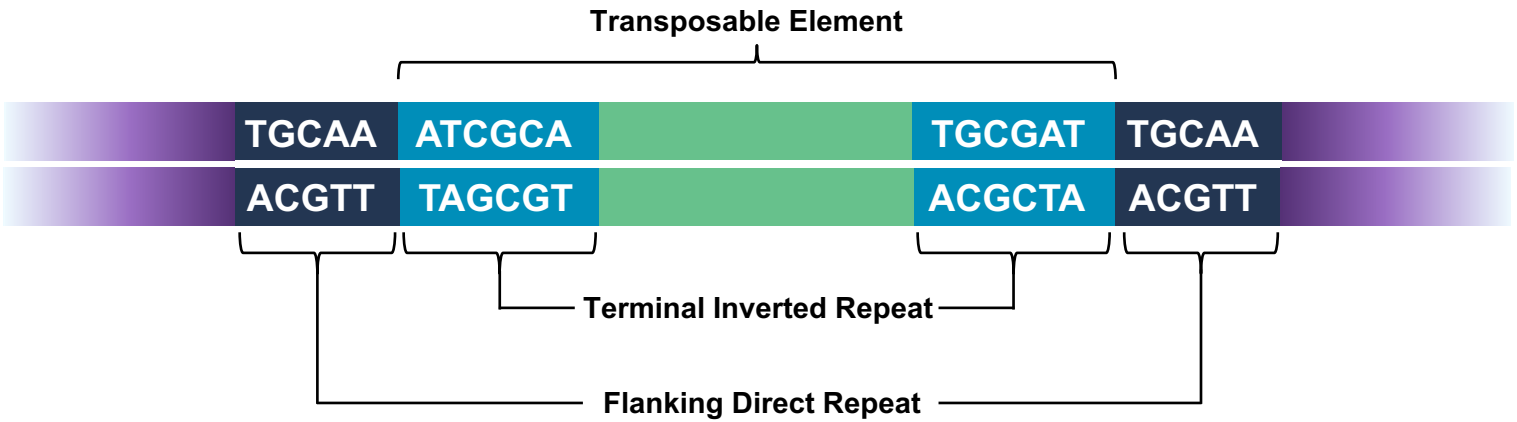
HIFI READS ALLOW YOU TO:

Span repetitive elements



HIFI READS ALLOW YOU TO:

Span repetitive elements



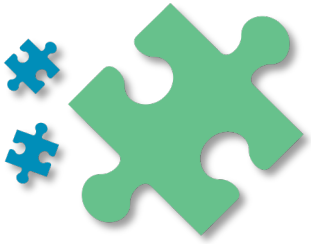
Phase haplotypes

Variation collapsed into a single mixed sequence

Each haplotype assembled separately

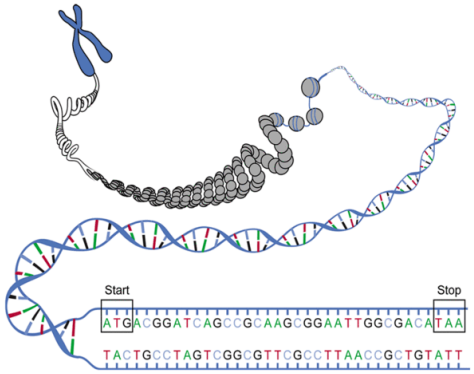
WHAT METRICS MAKE A GENOME HIGH QUALITY?

Contiguity



- Contig N50

Completeness



- Assembly size
- Gene completeness (BUSCO)

Correctness

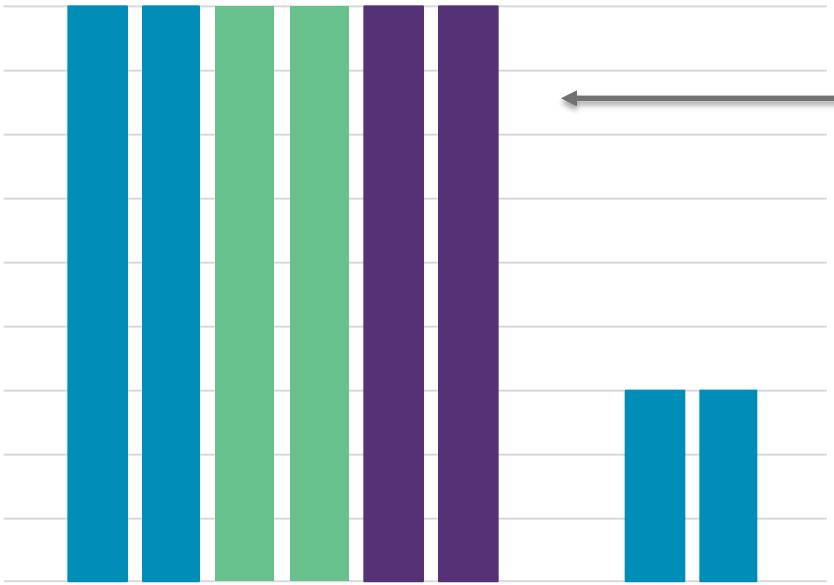
- Accuracy
- Genes in frame
- Phasing

THE CALIFORNIA REDWOOD GENOME

Sequoia sempervirens

- One of the world's fastest-growing conifers
- Live for thousands of years
- Only 5% of the original old-growth coast redwood forest remains
- 27 Gb hexaploid genome
- Genome assemblies by ONT in 2019 and PacBio in 2020

THE REDWOOD GENOME IS LARGE AND COMPLEX



54 Gb of DNA content!

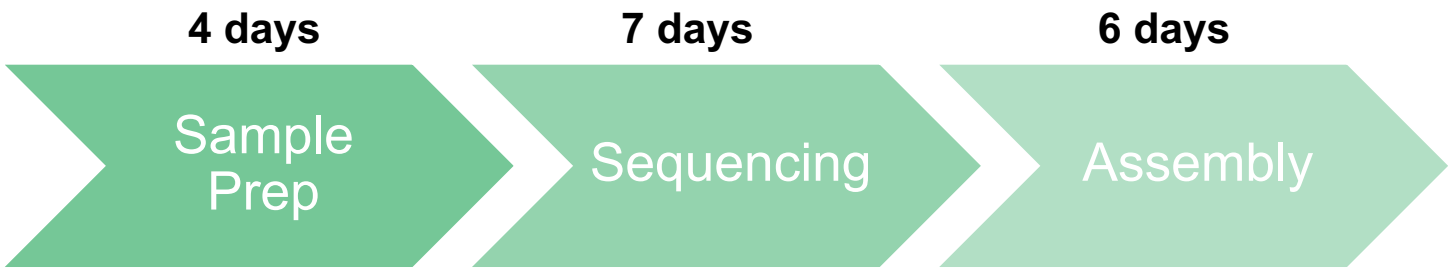
9 Gb hexaploid

9x the size of the human genome!

3 Gb diploid

6 Gb of DNA content!

THE PROJECT WORKFLOW



- Collected ~80 g of needles and froze
- Extracted 56 µg of DNA with Circulomics kit
- Prepped 2 HiFi libraries at ~25 kb

- Ran 31 SMRT Cells 8M across 9 instruments in 7 days
- Streamed data for immediate CCS analysis conversion to HiFi reads

- Used Hifiasm for quick, haplotype aware assembly

PACIFIC BIOSCIENCES® CONFIDENTIAL

17 DAYS

GENOME ASSEMBLY QUALITY

HiFi exceeds results of ONT + short reads for all three C's of genome quality – Contiguity, Completeness, and Correctness

California Redwood Genome Assembly Results		
Methodology	**PacBio HiFi**	ONT + short reads[1]
Genome Coverage	**22-fold**	23-fold + 122-fold
Assembly Size (Gb)	**47.7**	26.5
Contig N50 (Mb)	**1.92**	0.11
BUSCO Complete	**59%**	56%
Mapped transcripts with frameshift errors[2]	**0.12%**	1.97%

~2N assembly resolving recent autopolyploidy event

Significantly more transcripts with frameshift errors, impeding downstream analysis

1. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes
2. Transcript set of *Abies alba* from Neale, D. et al. Varying number of transcripts aligned to each genome (4,958 mapped to PacBio HiFi redwood, 4,760 mapped to ONT redwood)

OVERALL EFFORT

PacBio HiFi[1]

- DNA extraction from needles of a tree found locally
- HiFi sequencing done in 7 days
- First assembly done in 6 days

17 days vs ***>1 year*** for completion of respective projects

ONT + short reads[2]

- DNA extraction for short reads done from seed to get haploid DNA
- DNA extraction for ONT reads done on needles
- Short read sequencing done in December 2017
- ONT sequencing done in mid 2018
- Initial assembly done after several iterations in March 2019

PACIFIC BIOSCIENCES® CONFIDENTIAL

1. A Genome Fit for a Giant – Sequencing the California Redwood
2. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes

CAN WE DO BETTER WITH HIGHER COVERAGE?

- Additional HiFi coverage increased contiguity, completeness, and accuracy of resulting assembly

California Redwood Genome Assembly Results			
Methodology	**PacBio HiFi**	PacBio HiFi	ONT + short reads[1]
Genome Coverage	**33-fold**	22-fold	23-fold + 122-fold
Assembly Size (Gb)	**48.47**	47.7	26.5
Contig N50 (Mb)	**3.76**	1.92	0.11
BUSCO Complete	**60.6%**	59%	56%
Mapped transcripts with frameshift errors[2]	**0.11%**	0.12%	1.97%

1. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes
2. Transcript set of *Abies alba* from Neale, D. et al. Varying number of transcripts aligned to each genome (4,958 mapped to PacBio HiFi redwood, 4,760 mapped to ONT redwood)

RECENT LARGE AND/OR COMPLEX PLANT HIFI ASSEMBLIES

	Diploid plant 1	Diploid plant 2	Maize	Oat
Genome size	3.2 Gb	3.2 Gb	2.5 Gb	11 Gb
Library size	20 kb	20 kb	17 kb	17 kb
Coverage	21-fold	16-fold	20-fold	22-fold
Contig N50	12 Mb	7 Mb	14.7 Mb	20.3 Mb
Assembly time	<1 day	<1 day	6 hours	12 hours

We see **consistently good results** across a wide array of **complex plant genomes** with assemblies complete in less than a day!

HIFI FOR GENOME ASSEMBLY OF PLANTS & ANIMALS

With HiFi reads you can assemble reference-quality genomes with one technology

- Reach high contiguity, completeness, and correctness ensuring downstream utility
- Phase haplotypes for allele-specific genomic information
- Generate complete genomes in half the assembly time of traditional long reads

Small-bodied Species

- 150 Mb genome
- 14.4 Mb contig N50
- 99.999% accuracy (Q50)

Newly Sequenced Species

- 800 Mb genome
- 26.5 Mb contig N50
- 98.1% of genome phased

Large, Complex Species

- 11 Gb genome
- 20 Mb contig N50
- Assembly in 12 hours

PACIFIC BIOSCIENCES® CONFIDENTIAL

PACBIO®

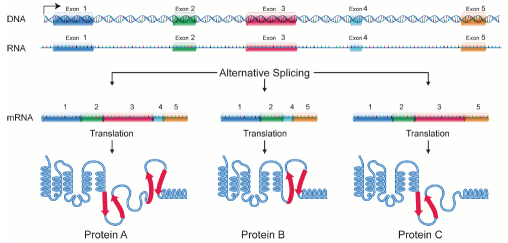
Iso-Seq Method

ISO-SEQ: FULL-LENGTH RNA SEQUENCING

RNA SEQUENCING

Iso-Seq is...

- Full-Length cDNA sequencing – no assembly required
- Targeted or whole transcriptome



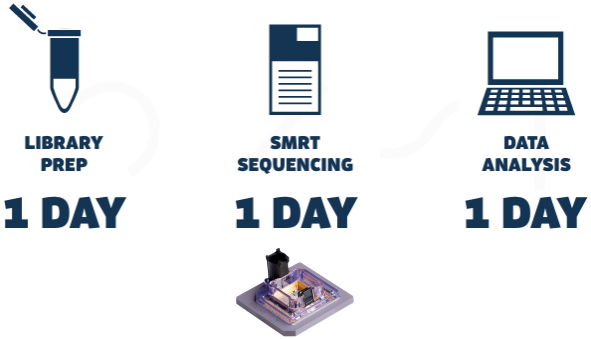
Iso-Seq can...

- Discover novel genes and isoforms
- Improve genome annotation, with or without reference genome
- Increase the accuracy of RNA-seq quantification at isoform-level resolution

ANNOTATION

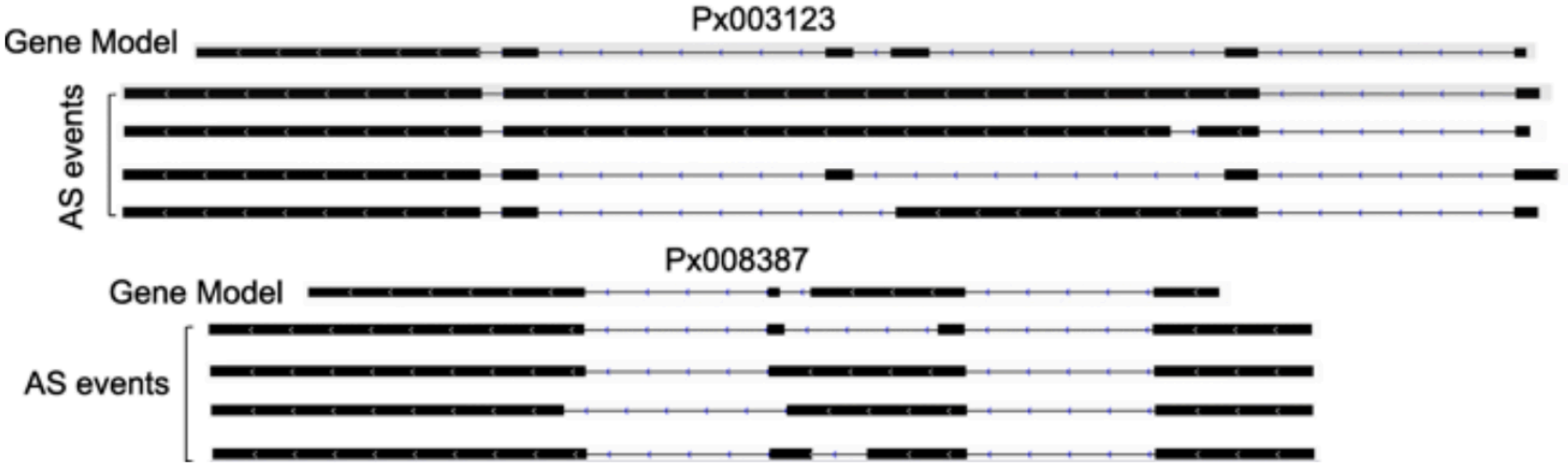
You can do Iso-Seq with...

- 1-Day Library Prep
- 1 SMRT Cell 8M
- Full bioinformatics solution



IMPROVED GENOME ANNOTATION WITH ISO-SEQ

Plutella xylostella
Diamondback moth

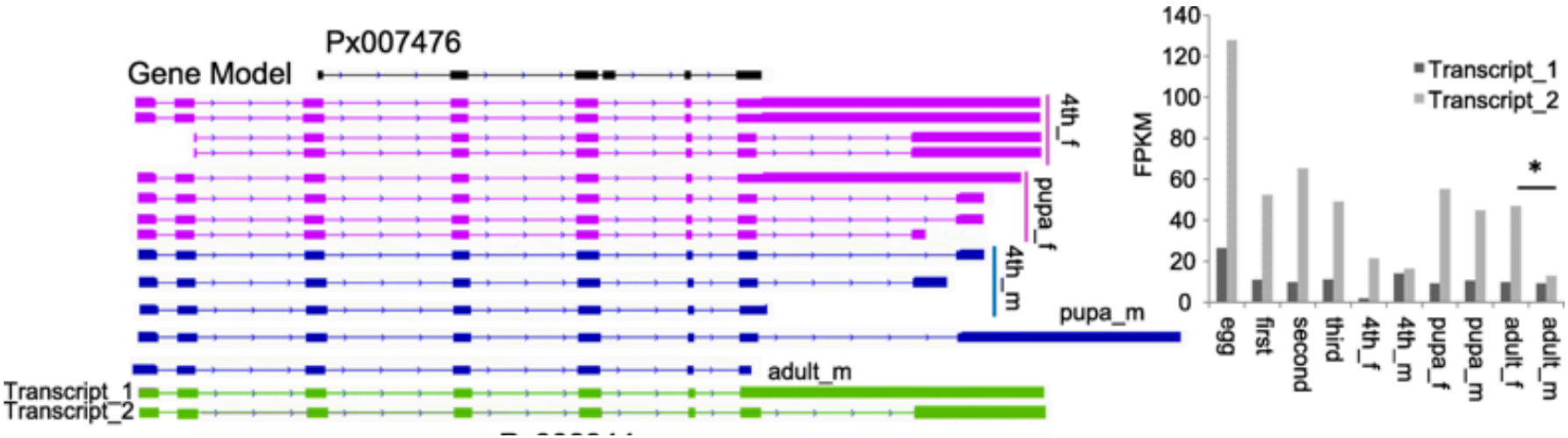


Alternative splicing contributes to a **much more diverse transcript set** compared to original gene model

Zhao, Q. et al., 2019. Genome-wide profiling of the alternative splicing provides insights into development in Plutella xylostella, *BMC Genomics*, 20:463.

IDENTIFYING GENES WITH SEX-SPECIFIC ALTERNATIVE SPLICING

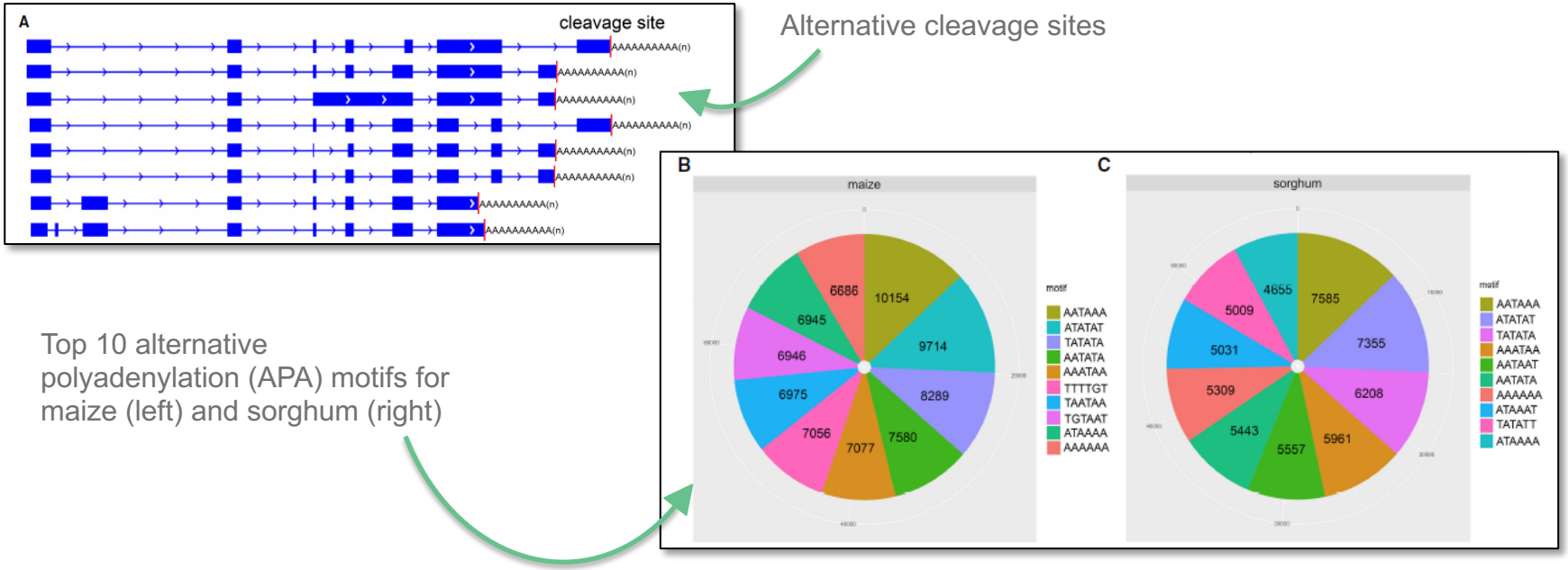
Plutella xylostella
Diamondback moth



Identified **156 genes** with sex-differentiated alternative splicing events to be further explored

Zhao, Q. et al., 2019. Genome-wide profiling of the alternative splicing provides insights into development in Plutella xylostella, *BMC Genomics*, 20:463.

GAIN POLY-ADENYLATION SITE INFORMATION

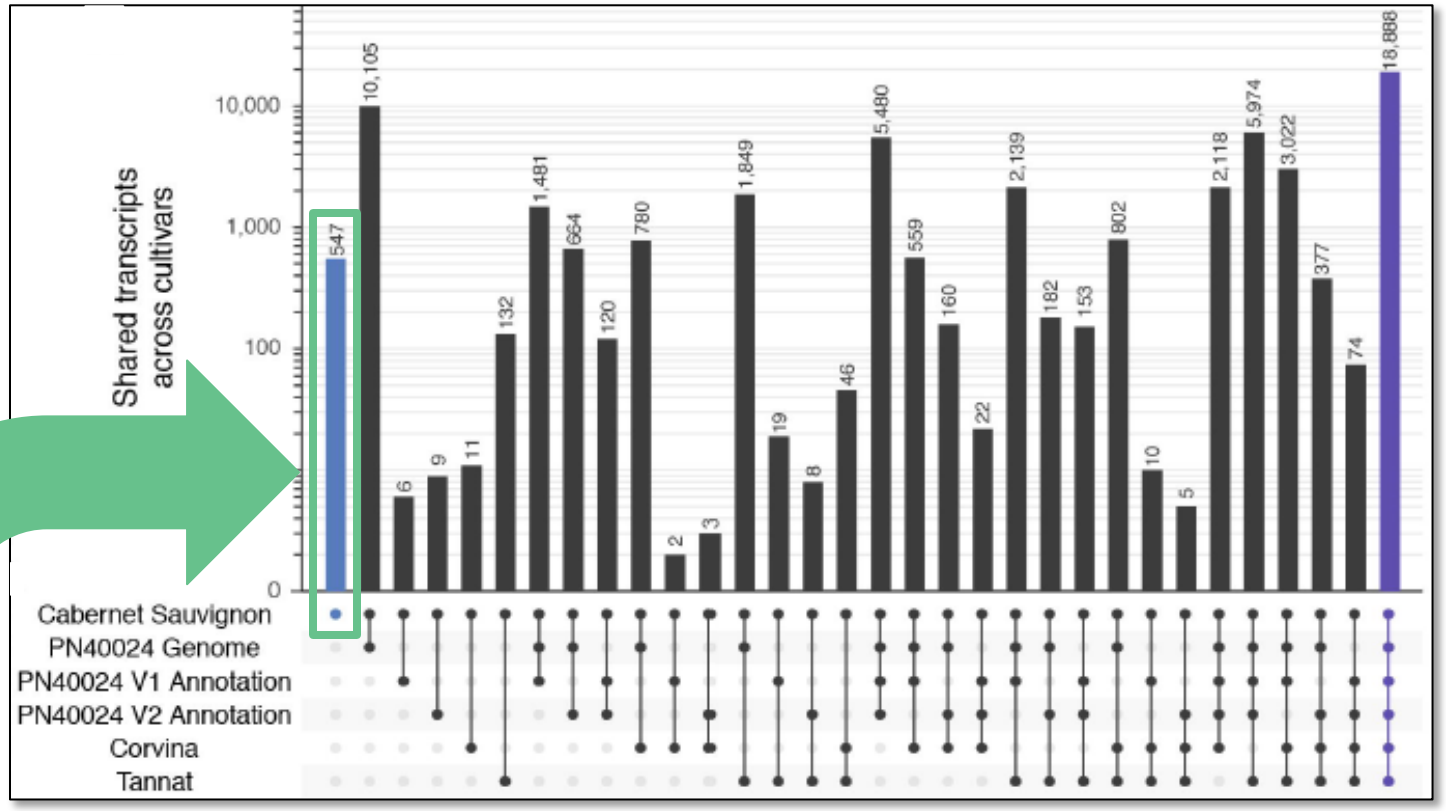


Top 10 alternative polyadenylation (APA) motifs for maize (left) and sorghum (right)

"We generated ***comprehensive*** and ***high-resolution*** maps of genome-wide poly(A) sites, allowing ***systematic characterization*** of the role of APA in… agronomically important species."

Wang, B. et al. (2018) A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. Genome Research 28:1-12

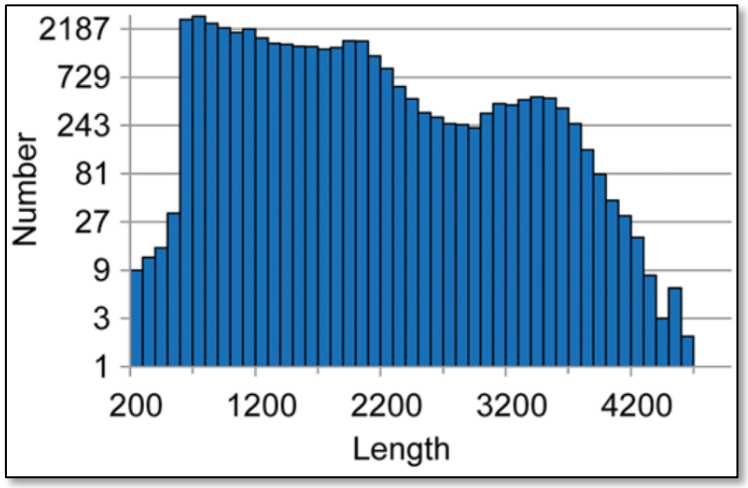
IMPROVE SHORT-READ RNA-SEQ ISOFORM QUANTIFICATION

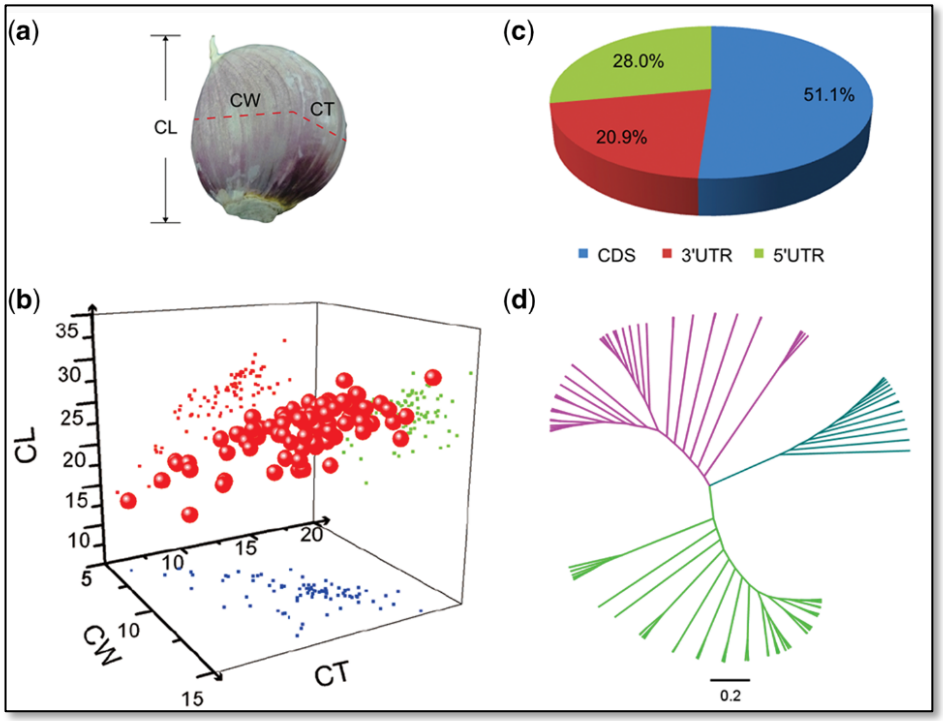


>500 Cabernet Sauvignon-specific transcripts were found when the transcriptome was compared other grape cultivars

Minio, A. et al. (2018) Isoform-scale annotation and expression profiling of the Cabernet Sauvignon transcriptome using single-molecule sequencing of full-length cDNA. bioRxiv

INVESTIGATE TRANSCRIPTOMES WITHOUT REFERENCE GENOMES





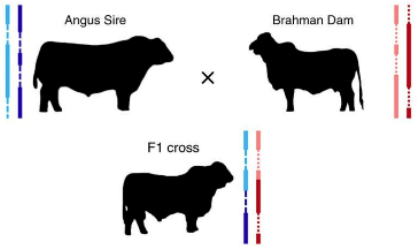
"A large number of transcripts in ***[previous] transcriptomes were incomplete***…therefore, we used single-molecule long-read sequencing technology for RNA sequencing, which ***significantly improved the transcriptome quality***."

Chen, X. et al. (2018) Transcriptome-referenced association study of clove shape traits in garlic. DNA Research dsy027.

HIFI FOR RNA SEQUENCING OF PLANTS & ANIMALS

With HiFi reads you can sequence full-length cDNA sequences – from 5’ end to the poly-A tail

- Discover novel genes and isoforms
- Improve genome annotation, with or without reference genome
- Increase the accuracy of RNA-seq quantification at isoform-level resolution

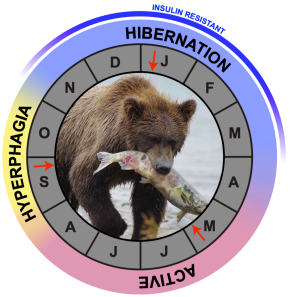


Brahman x Angus F1 Cattle

- Allele-specific isoform expression
- Tissue-specific isoform expression

Cannabis

- Tissue-specific transcripts associated w/ THC & CBD synthesis
- Chr Y gene annotation



Grizzly Bears

- Tissue-specific alternative splicing
- Hibernation vs active state

PACBIO®

Targeted Sequencing

TARGETED SEQUENCING ON SEQUEL II SYSTEM

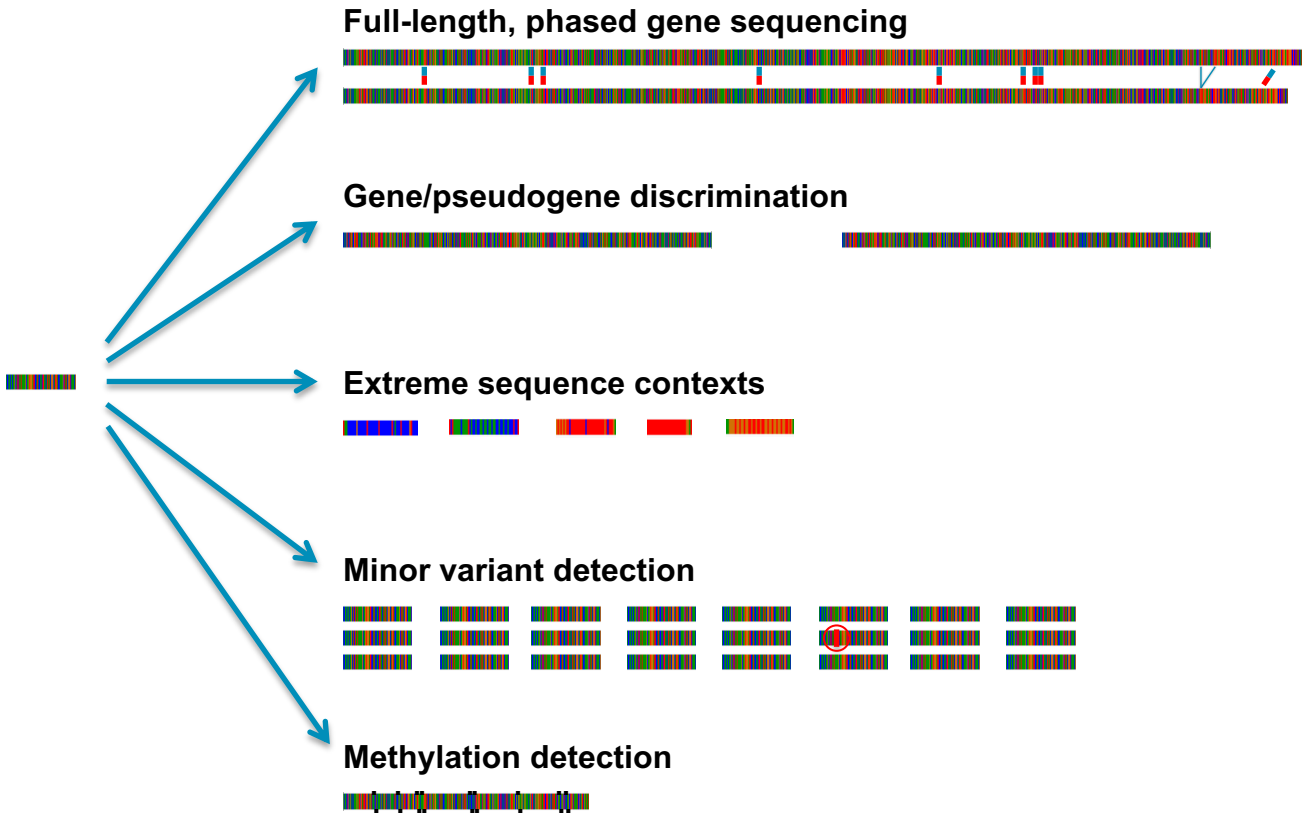
Obtain uniform coverage and accurately sequence through high- or low-GC content for targeted regions of the genome.

SMRT Sequencing offers a flexible solution to deliver the long read length and accuracy needed to:

- Efficiently multiplex large amplicons
- Discover haplotype-specific markers
- Reassemble multi-megabase regions of a genome
- Confirm insertions sites of transgenes and validate gene editing events
- Capture complete genes

www.pacb.com/applications/targeted-sequencing

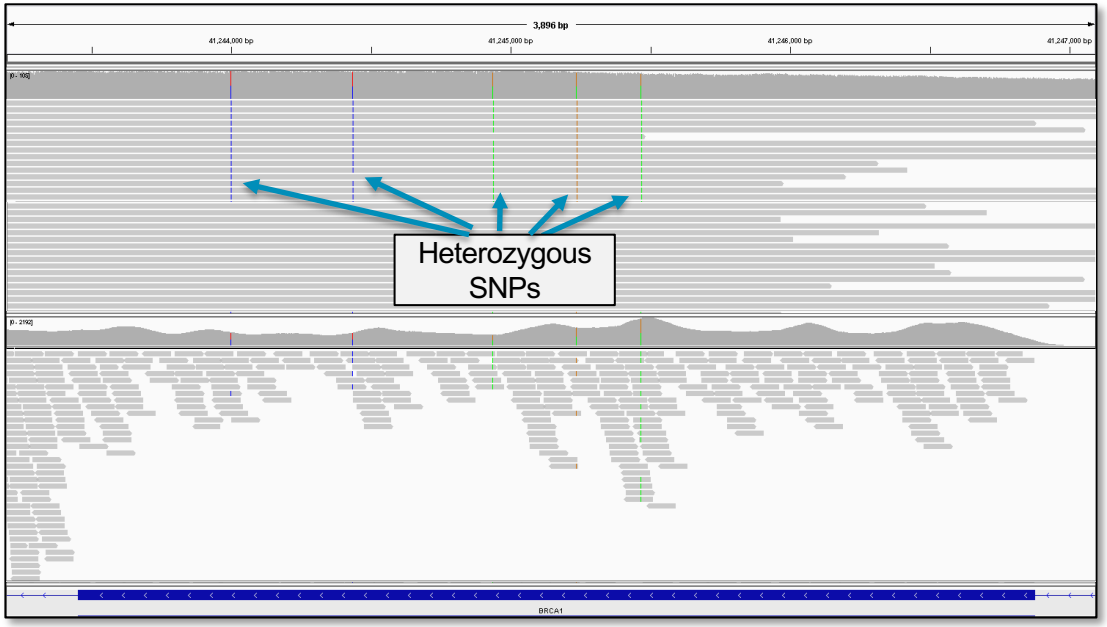
ADVANTAGES OF TARGETED LONG-READ SMRT SEQUENCING



PHASING VARIANTS OVER LARGE DISTANCES: *BRCA1*, EXON 10

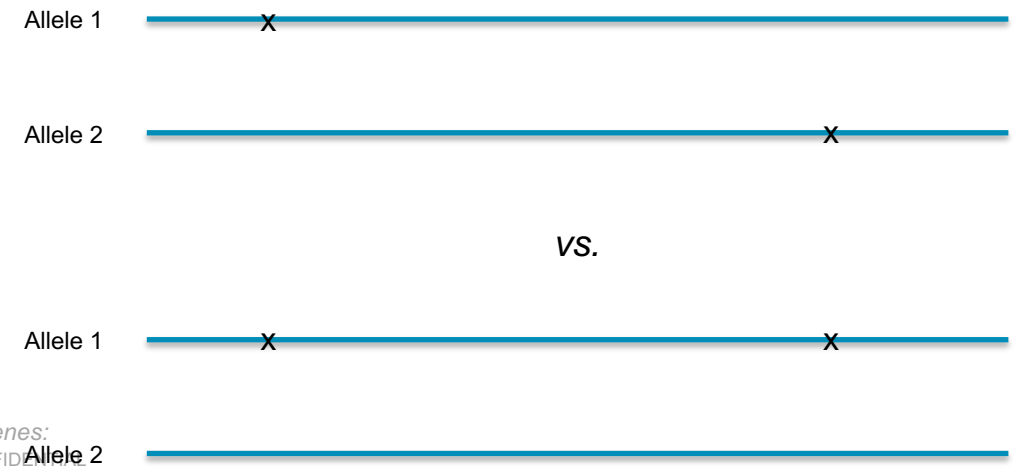
PacBio (~5 kb fragments)

MiSeq (200 bp fragments)



PACIFIC BIOSCIENCES® CONFIDENTIAL

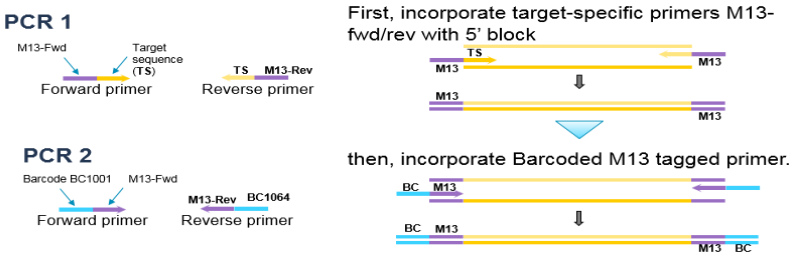
IMPORTANCE OF VARIANT PHASING



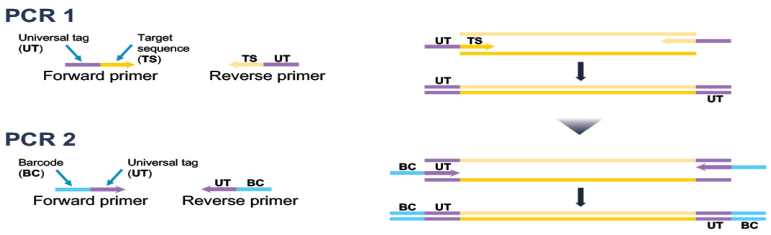
e.g., tumor suppressor genes:

FLEXIBLE MULTIPLEXING AND BARCODING SOLUTIONS FOR AMPLICONS

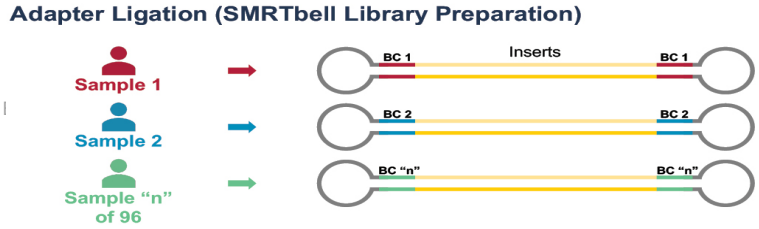
Barcoded M13 Primers



Barcoded Universal Primers



Barcoded Adapters

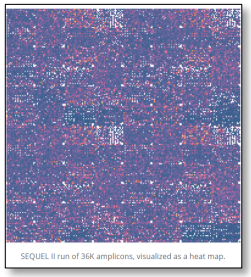


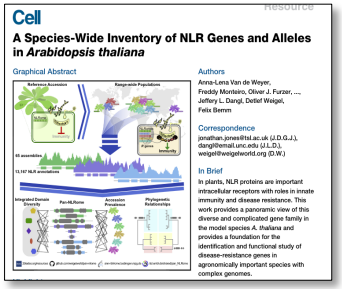
Barcoded Primers

HIFI FOR TARGETED SEQUENCING OF PLANTS & ANIMALS

With HiFi reads you can view specific genomic regions of interest, regardless of size

- Rapidly screen and identify all variants
- Discover haplotype-specific markers
- Resolve difficult-to-sequence regions





Massive Parallel Sequencing

- Catalog biodiversity via barcoding targeted regions
- Barcoding up to 40,000 species in a single SMRT Cell 8M

Building an NLR-ome

- Target NLR genes across 65 accessions
- Catalog diversity in resistance-associated genes

Biomonitoring of Herbal Supplements

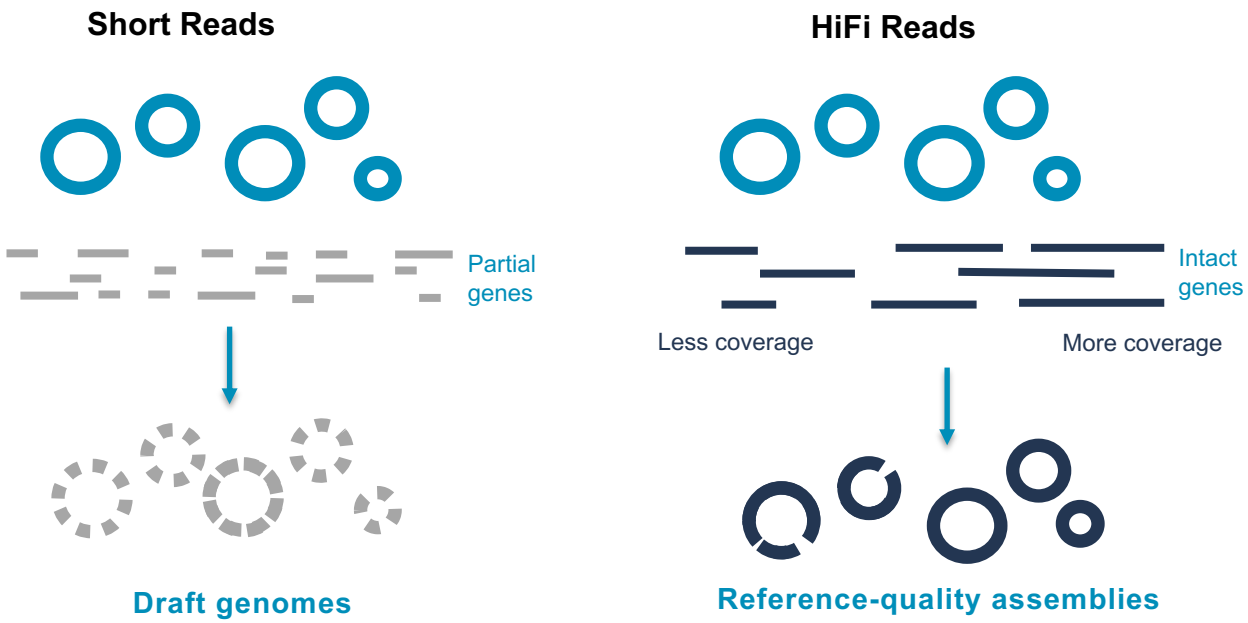
- Ensure supplements contain the expected biological composition of herbs

PACIFIC BIOSCIENCES® CONFIDENTIAL

PACBIO®

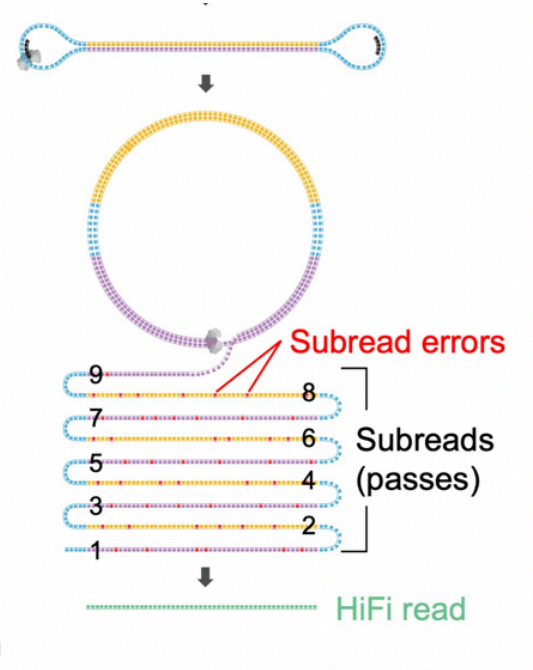
Metagenomics

METAGENOMIC SEQUENCING OF COMPLEX POPULATIONS



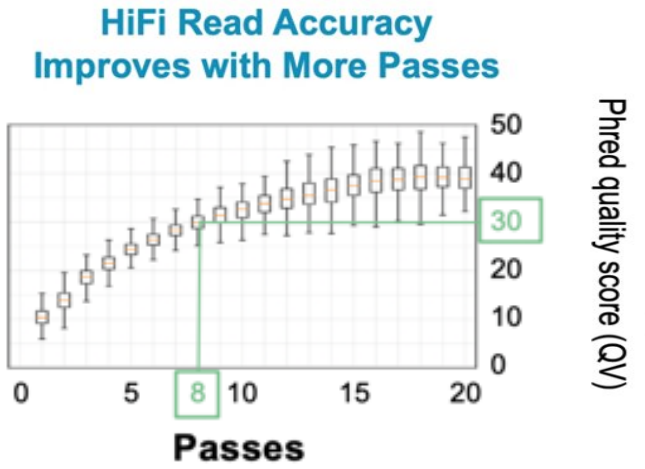
PACBIO HIFI READS COMBINE LONG READ LENGTHS WITH HIGH ACCURACY

HiFi READS



PACIFI

- **16S**: Up to 3.6 M Q30 full-length 16S sequences
- **Shotgun**: Up to 2.4 M Q20 reads from 10 kb libraries

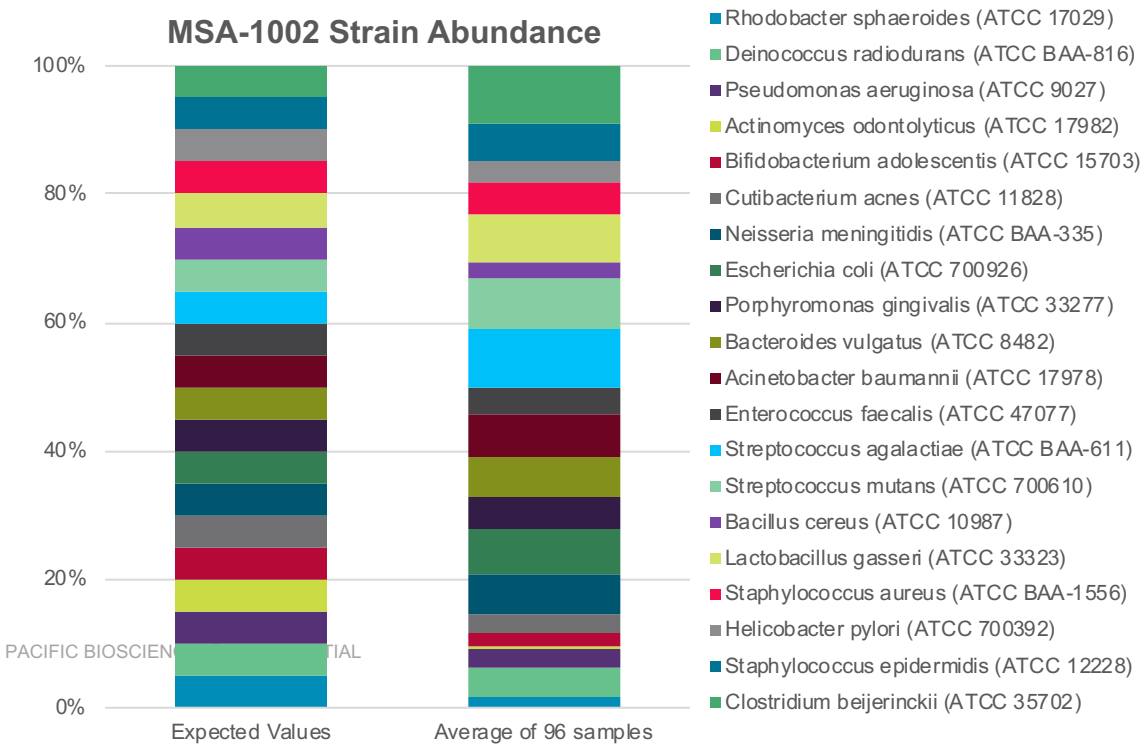


PACBIO HIFI SEQUENCING ON THE SEQUEL II SYSTEM: HIGH THROUGHPUT, HIGH RESOLUTION

Full-length 16S reveals the true diversity of your sample and generates more testable hypotheses

- Distinguish keystone or critical *species* from genus-level noise
- All-in-one kit (extract, amplify, analyze) from our partner Shoreline Biome OR
- PacBio 1-step, low chimera protocol for 96-plex sequencing with 20 self-ordered primers

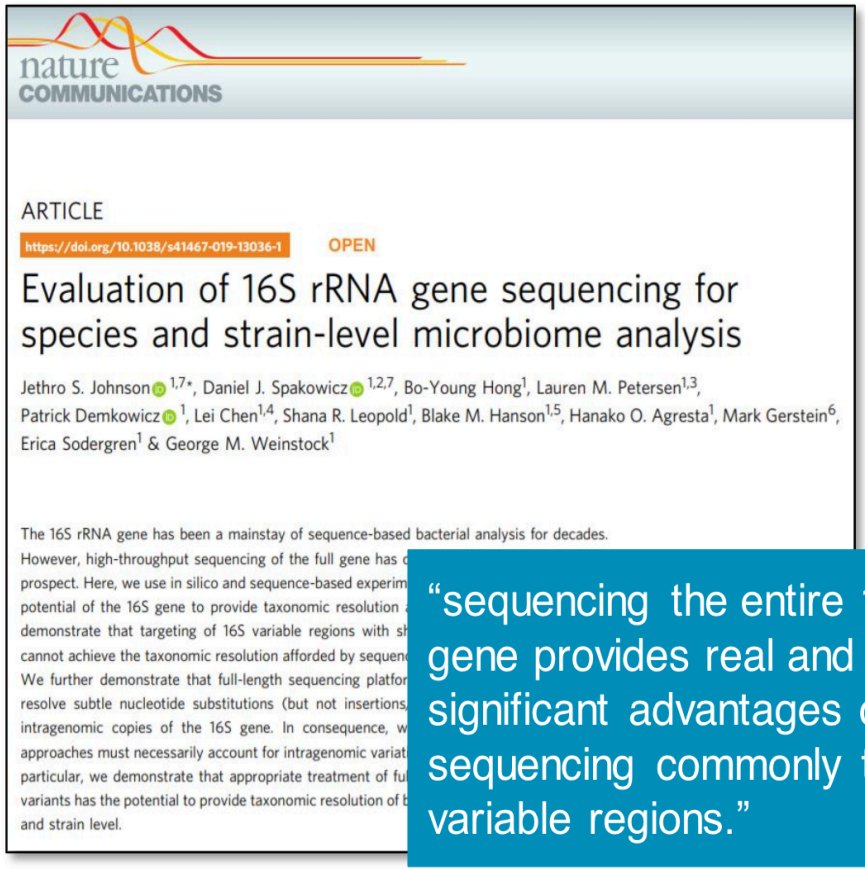
PACBIO 16S SEQUENCING FAITHFULLY REPRESENTS A KNOWN MOCK COMMUNITY SAMPLE



- Download and explore Sequel II system 16S data for yourself

V1-V9 amplicons were sequenced on a single SMRT Cell 8M at 96-plex

SUB-REGION SEQUENCING SHOWS BIAS IN PROFILING TAXA



nature COMMUNICATIONS

ARTICLE

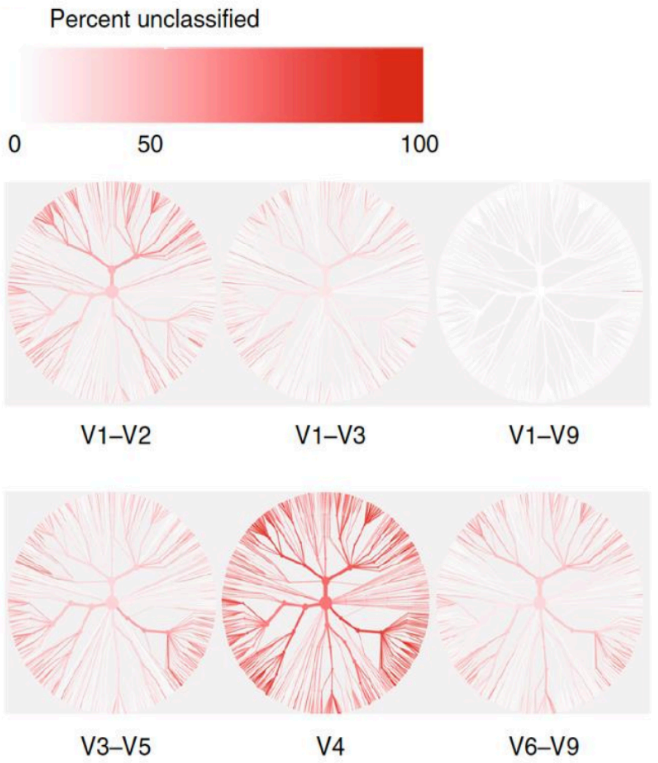
https://doi.org/10.1038/s41467-019-13036-1 OPEN

Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis

Jethro S. Johnson[1,7*], Daniel J. Spakowicz[1,2,7], Bo-Young Hong[1], Lauren M. Petersen[1,3], Patrick Demkowicz[1], Lei Chen[1,4], Shana R. Leopold[1], Blake M. Hanson[1,5], Hanako O. Agresta[1], Mark Gerstein[6], Erica Sodergren[1] & George M. Weinstock[1]

The 16S rRNA gene has been a mainstay of sequence-based bacterial analysis for decades. However, high-throughput sequencing of the full gene has c prospect. Here, we use in silico and sequence-based experim potential of the 16S gene to provide taxonomic resolution demonstrate that targeting of 16S variable regions with sh cannot achieve the taxonomic resolution afforded by sequenc We further demonstrate that full-length sequencing platfor resolve subtle nucleotide substitutions (but not insertions intragenomic copies of the 16S gene. In consequence, w approaches must necessarily account for intragenomic variat particular, we demonstrate that appropriate treatment of ful variants has the potential to provide taxonomic resolution of b and strain level.

"sequencing the entire 16S gene provides real and significant advantages over sequencing commonly targeted variable regions."



Johnson JS, et. al. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level miocrbiome analysis. Nature, 10; 5029.

PACBIO HIFI SEQUENCING ON THE SEQUEL II SYSTEM: HIGH THROUGHPUT, HIGH RESOLUTION

Full-length 16S reveals the true diversity of your sample and generates more testable hypotheses

- Distinguish keystone or critical *species* from genus-level noise
- All-in-one kit (extract, amplify, analyze) from our partner Shoreline Biome OR
- PacBio 1-step, low chimera protocol for 96-plex sequencing with 20 self-ordered primers

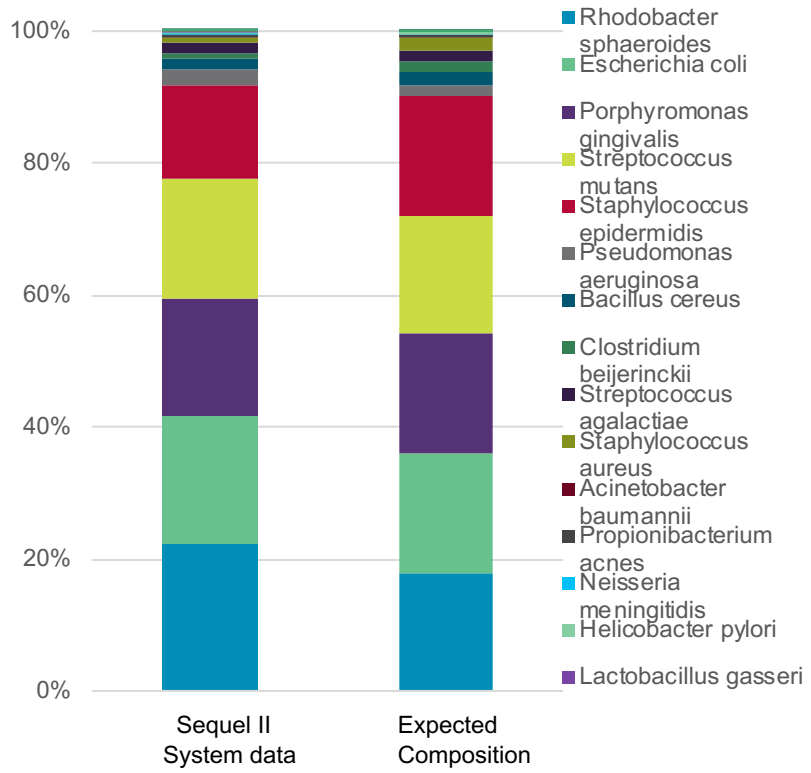
HiFi Shotgun profiling reveals intact genes and operons without assembly

- Eliminate reliance on the step that wastes 30-70% of your raw data
- See the metabolic functions of even low-abundance species without enough coverage for assembly or error correction

SHOTGUN SEQUENCING ON THE SEQUEL II SYSTEM FAITHFULLY RECAPITULATES THE MSA-1003 MOCK COMMUNITY

- Successful detection of species down to 0.018 % abundance

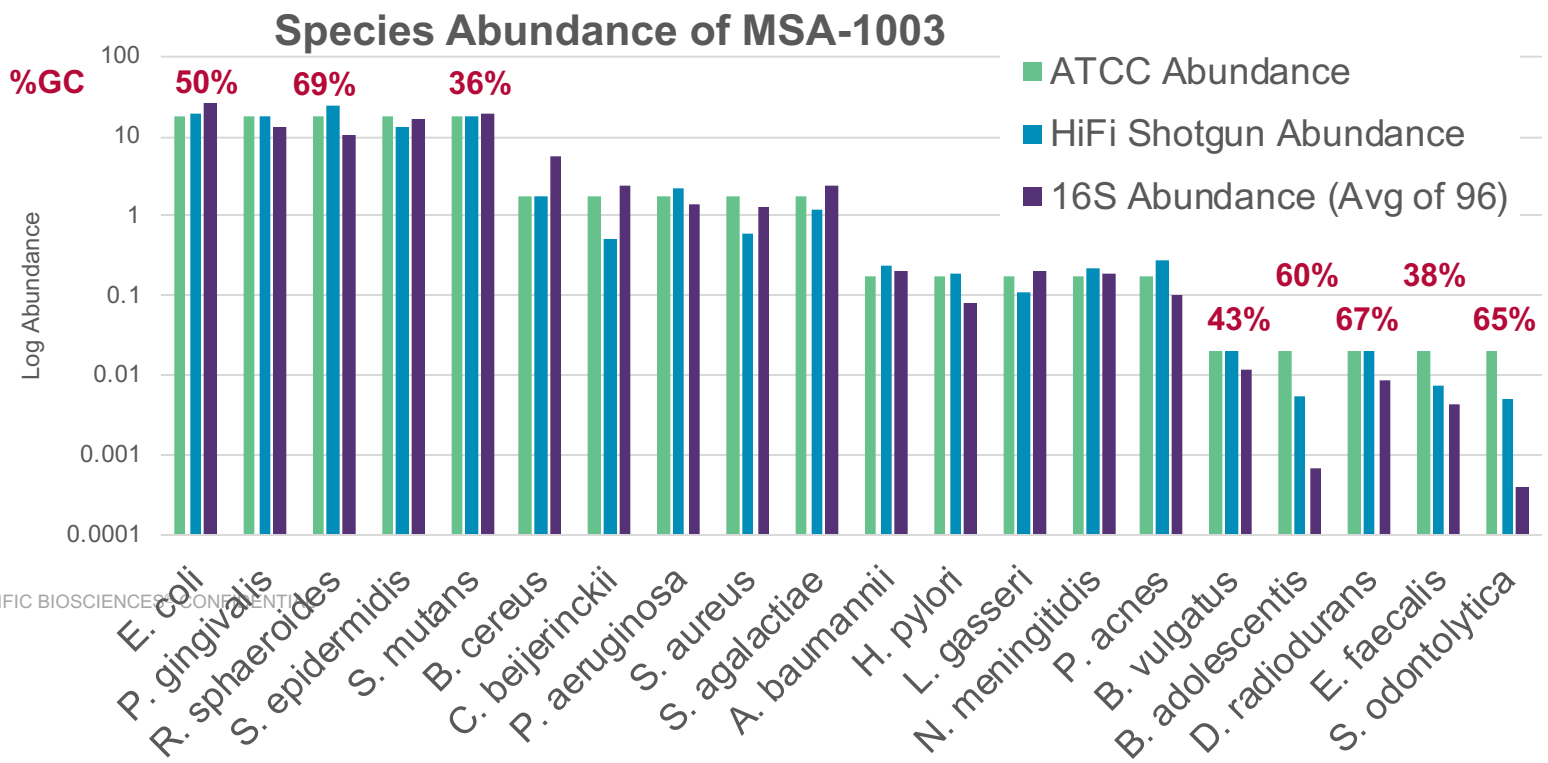
- Download and explore MSA-1003 Mock Community shotgun data



PACIFIC BIOSCIENCES® CONFIDENTIAL

PACBIO SHOTGUN SEQUENCING IS FREE FROM GC BIAS, ENABLING ACCURATE REPRESENTATION OF DIVERSITY

HiFi READS



PACIFIC BIOSCIENCES CONFIDENTIAL

HIFI METAGENOMICS SHOTGUN PERFORMANCE ON THE SEQUEL II SYSTEM

Shotgun	>Q20 reads	>Q20 bases	Avg read length	>Q20 QV
Human fecal 1	2,485,902	21,892,869,069	8,806	Q39
Human fecal 2	2,634,276	24,359,683,697	9,247	Q37
Human fecal 3	2,371,437	20,325,373,131	8,570	Q39
Human fecal 4	2,133,478	21,557,465,918	10,104	Q36
Human fecal 5	2,037,230	19,855,203,301	9,746	Q37
Human fecal 6	2,230,353	19,784,876,972	8,870	Q39
Human fecal 7	2,796,697	22,710,850,840	8,120	Q40
Human fecal 8	1,977,870	17,034,971,133	8,612	Q40
Human fecal 9	2,529,830	21,908,484,087	8,660	Q39

- The median read QV and read length of HiFi data outperforms many metagenome assembly quality metrics

LONG READS + HIGH ACCURACY MEANS GENE DISCOVERY CAN BE DONE *DIRECTLY* ON HIFI READS, WITHOUT ASSEMBLY

Sample	Number of Predicted Genes	Mean Length (bp)	Mean Predicted Genes / Read	Clustered Genes (99% ID)
Human fecal 1	19,639,322	1,005	7.9	1,012,982
Human fecal 2	22,064,417	1,001	8.4	1,141,123
Human fecal 3	18,059,181	1,024	7.6	1,154,341
Human fecal 4	19,844,033	978	9.3	1,250,711
Human fecal 5	18,396,237	970	9.0	1,087,015

- 30-70% of short data will not map to a metagenome assembly, and are therefore not useful for gene finding
- With HiFi sequencing, error-free genes can be found even from species with too little coverage for assembly
- High accuracy means existing NGS tools and pipelines can be used without modification

PACBIO HIFI SEQUENCING ON THE SEQUEL II SYSTEM: HIGH THROUGHPUT, HIGH RESOLUTION

Full-length 16S reveals the true diversity of your sample and generates more testable hypotheses

- Distinguish keystone or critical *species* from genus-level noise
- All-in-one kit (extract, amplify, analyze) from our partner Shoreline Biome OR
- PacBio 1-step, low chimera protocol for 96-plex sequencing with 20 self-ordered primers

HiFi Shotgun profiling reveals intact genes and operons without assembly

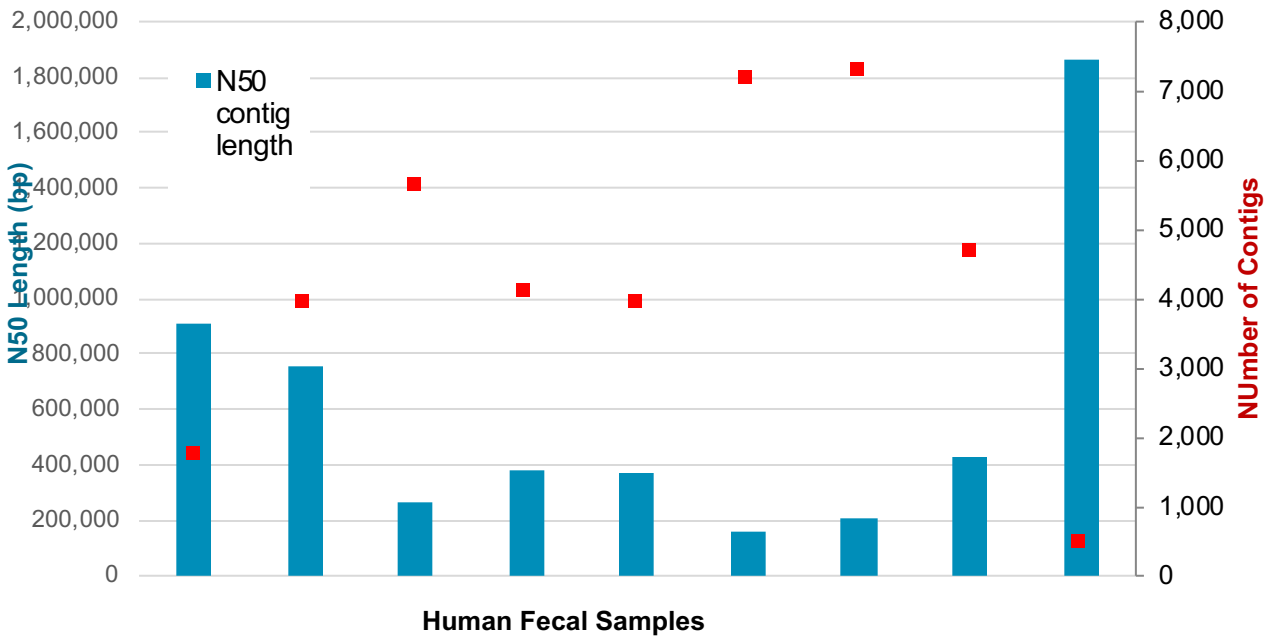
- Eliminate reliance on the step that wastes 30-70% of your raw data
- See the metabolic functions of even low-abundance species without enough coverage for assembly or error correction

Metagenome assembly with HiFi reads generates new references for unculturable species

- >99% accurate, complete genomes with ~15-20x coverage from a single technology
- Leverage epigenomic data to cluster contigs and plasmids from the same strain

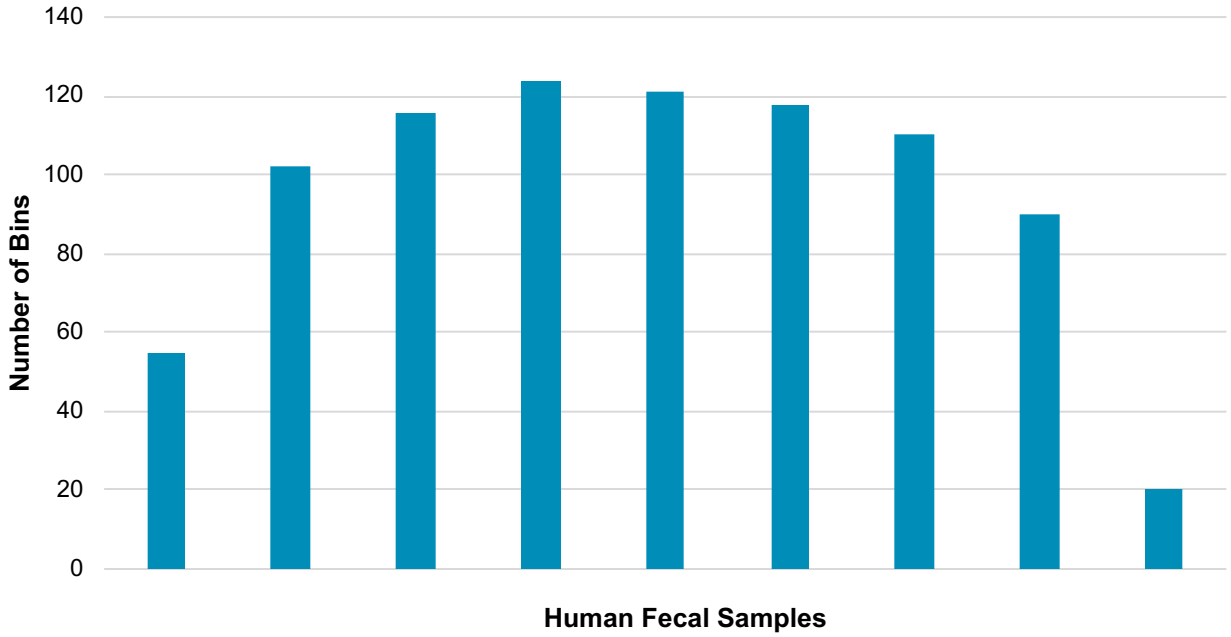
HIFI READS CAN BE ASSEMBLED WITH CANU TO PRODUCE NOVEL REFERENCES FOR UNCULTURABLE SPECIES

HiFi READS



CONTIGS PRODUCED BY CANU CAN BE BINNED BY SPECIES WITH TOOLS LIKE PATRIC RBS

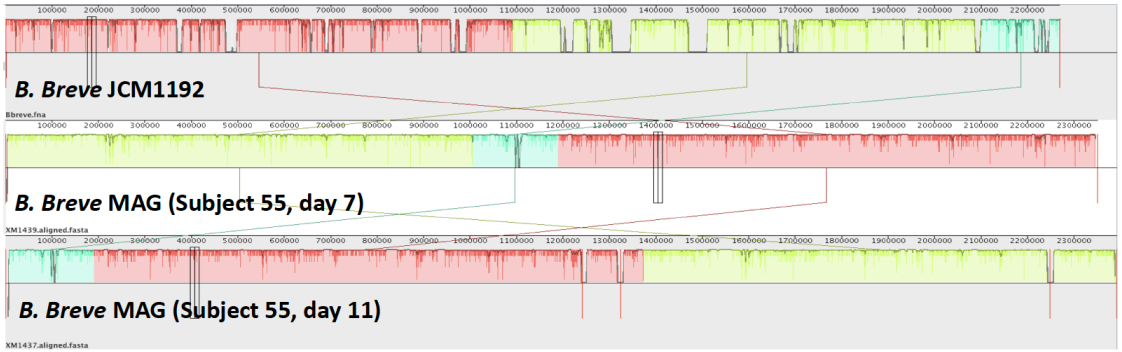
HiFi READS



Analyses performed at https://patricbrc.org/ on the combined Canu contigs and unassembled reads for each sample.

PACBIO METAGENOME DATA CAN PROVIDE REFERENCE QUALITY ASSEMBLIES OF UNCULTURABLE STRAINS

Uncover mechanisms with a highly resolved, functional view of your metagenome



- Closed, complete or nearly complete *B. breve* genomes from preterm neonate gut microbiome samples, with sequencing coverage between 7- to 115-fold.

- The *B. breve* strains, associated with healthy gut development, possess diverse carbohydrate metabolism capabilities, including a “bifid shunt” that can convert human milk oligosaccharides (HMO) to short chain fatty acids (SCFAs).

PACIFIC BIOSCIENCES® CONFIDENTIAL

MARYLAND GENOMICS

McComb, E. (2019). High-resolution evaluation of gut microbiota associated with intestinal maturation in early preterm neonates. ASM Microbe poster presentation.

CALCULATING THE ESTIMATED COVERAGE OF RARE SPECIES AT DIFFERENT MULTIPLEX LEVELS

	1 / SMRT Cell	2 / SMRT Cell	3 / SMRT Cell
Assignable Q20 reads / cell*	2.4 M	2.4 M	2.4 M
Reads / sample	2.4 M	1.2 M	800,000
1% of reads	assembly 24,000	assembly 12,000	profiling 8,000
0.2% of reads	profiling 4,800	detection 2,400	detection 1,600

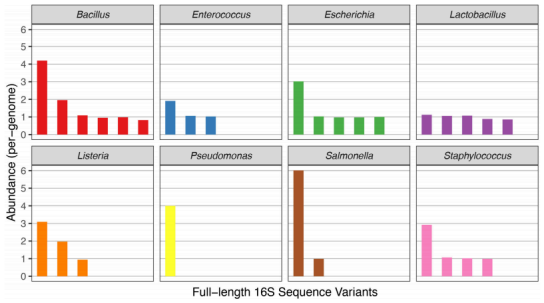
*99.5% of HiFi reads have recoverable barcodes

- The average read length for metagenomics samples is 8.5 kb when following the recommended protocol with high molecular weight DNA.
- Choose your multiplex level depending on how many reads per rarest-OTU of interest you require for your analysis plan.

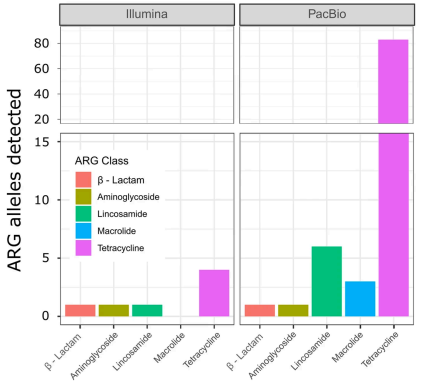
HIFI FOR METAGENOMIC SEQUENCING OF ECOSYSTEMS

With HiFi reads you can fully characterize microbes and their communities with one technology

- Obtain strain-level resolution of complex populations
- Uncover key community functions by recovering complete genes and operons
- Discover novel genes and gene clusters by reconstructing long contigs



Sample	Number of Predicted Genes	Mean Length (bp)	Mean Predicted Genes / Read
Human fecal 1	19,639,322	1,005	7.9
Human fecal 2	22,064,417	1,001	8.4
Human fecal 3	18,059,181	1,024	7.6
Human fecal 4	19,844,033	978	9.3
Human fecal 5	18,396,237	970	9.0
Average	**19 M**	**996 bp**	**8 genes**



16S Community Cataloging

- Uncover the full complement of full-length 16S sequence variants

Shotgun Metagenomic Profiling

- Recover multiple genes per HiFi read for optimal utility

Metagenomic Assembly

- Assign virus and antimicrobial resistance genes to microbial hosts

PACIFIC BIOSCIENCES® CONFIDENTIAL

PACBIO®

Conclusions

HIFI SEQUENCING ENABLES PLANT & ANIMAL RESEARCH

- Uncover new biology to improve crop or animal health, enhance breeding efficiency, and protect biodiversity
  - Reference-quality genome assemblies
  - Isoform-level genome annotation
  - Targeted gene sequencing, regardless of size
  - Strain-level community analysis of complex populations

2020 SMRT GRANT PROGRAMS

The 2020 Plant and Animal Sciences SMRT Grant Program is Now Open!

Explore Earth's Biodiversity with HiFi Sequencing

Show us in 90-seconds how highly accurate long reads will help you understand any organism from earth's many ecosystems for a chance to win free sequencing.

APPLY NOW

And don't forget to submit your video abstract by **Friday, July 24, 2020**.

PACBIO®

www.pacb.com

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx.

All other trademarks are the sole property of their respective owners.