NORWEGIAN SEQUENCING CENTRE

PacBio long read sequencing

Ave Tooming-Klunderud, NCS/CEES/UiO,

ave.tooming-klunderud@ibv.uio.no

Oslo universitetssykehus

The Research Council of Norway

CEES Centre for Ecological and Evolutionary Synthesis

HELSE SØR-ØST

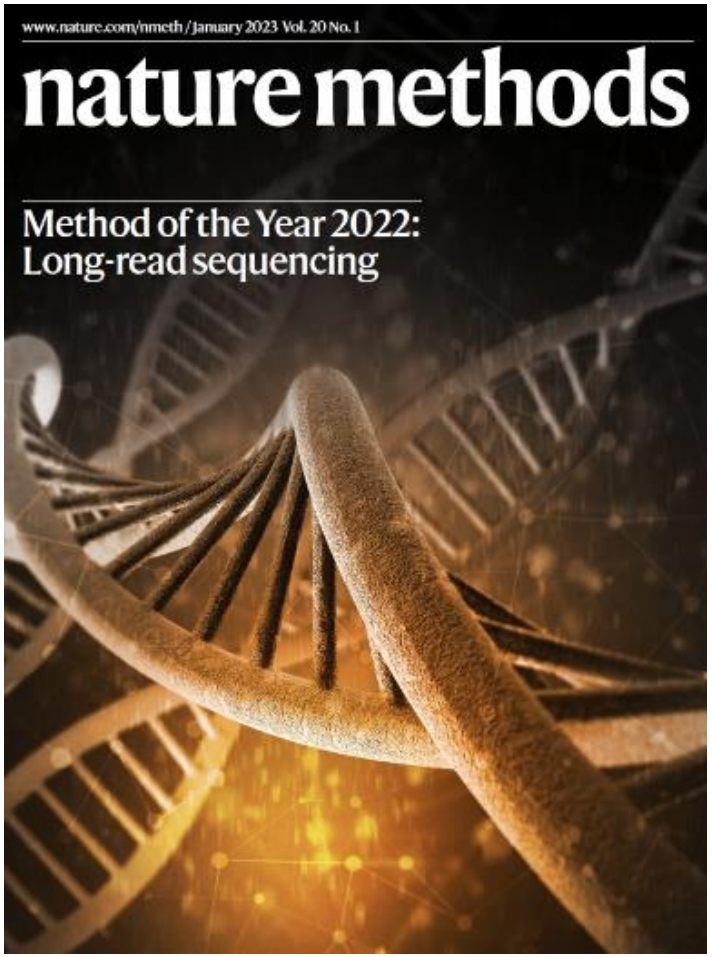
UNIVERSITAS OSLOENSIS MDCCCXI

Editorial | Published: 12 January 2023

Method of the Year 2022: long-read sequencing

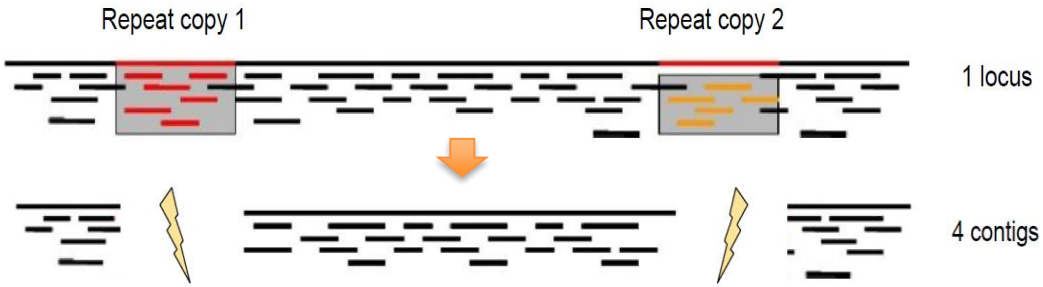
Nature Methods **20**, 1 (2023) | Cite this article

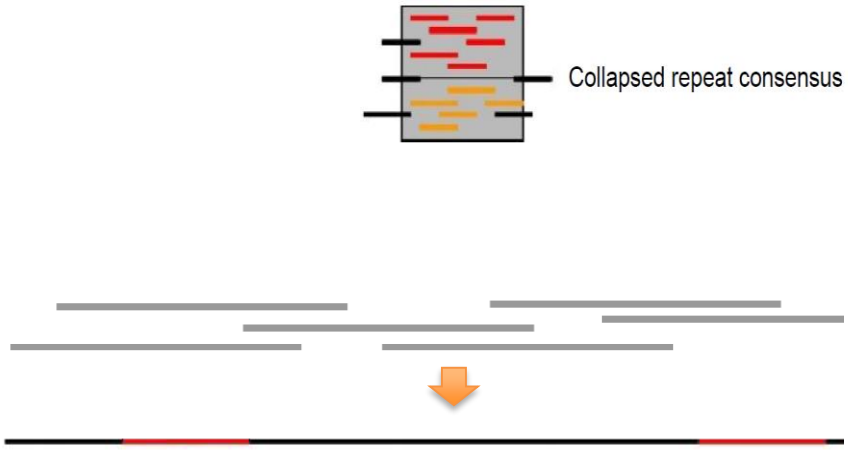
Long-read sequencing powers a more complete reading of genomic information.



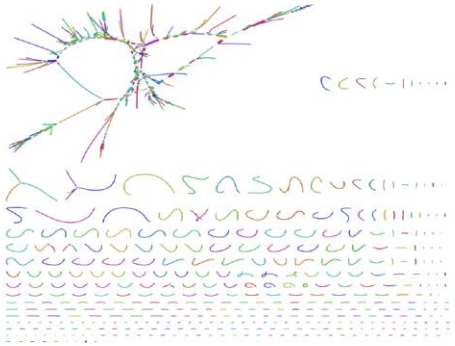
Why long reads?

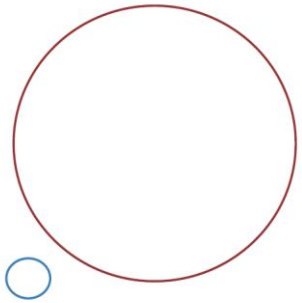
Long reads will span repeats





Complete genomes (small genomes)





Complete genomes (large genomes)

Contig Type	# Polished Contigs	Maximum Contig Length	N50 Contig Length	L50	Sum of Contig Lengths
Primary Contigs	698	29,198,342	12,301,043	20	684,461,580

Draft vs reference quality genome

Type to search GoaT taxon index (e.g. Canidae)

Rangifer tarandus

include descendants: Off | include estimates: On | result columns | query builder | clear all

taxon record 9870

Rangifer tarandus (species) 9870

gc_percent	assembly_span	assembly_level	bioproject	biosample	contig_n50	assembly_date	scaffold_n50	nohit
43.2 median, n=2	2.86G primary, n=2	Scaffold primary, n=2	PRJEB35834 ... list, n=3	SAMEA6417239 ... list, n=3	129k primary, n=2	2022-03-03 primary, n=2	986k primary, n=2	60.2 median, n=2

target	mitochondrion_assembly_span	mitochondrion_gc_percent	haploid_number	chromosome_number	ploidy	genome_size	c_value	long_list
98.2 median, n=2	16.4k median, n=1	36.2 median, n=1	35 median_high, n=1	70 median_high, n=1	2 median_high, n=1	3.33G primary, n=1	3.41 primary, n=1	CANBP ... list, n=2

other_priority	sequencing_status	sequencing_status_zoonomia	sample_collected	sample_acquired	in_progress
CANBP ... list, n=2	published enum, n=3	published enum, n=1	ZOONOMIA list, n=1	ZOONOMIA list, n=1	ZOONOMIA list, n=1

Lineage

Eukaryota Opisthokonta Metazoa Eumetazoa Bilateria Deuterostomia Chordata Craniata Vertebrata Gnathostomata Sarcopterygii Dipnotetrapodomorpha Tetrapoda Amniota Mammalia Theria Eutheria Boreoeutheria Laurasiatheria A Cervidae Odocoileinae Rangifer

Names

caribou - common name | Rangifer spitzbergensis - synonym | Rangifer tarar

ASM2245718v1

Organism name: Rangifer tarandus (reindeer)
Isolate: DF-B-001
BioSample: SAMN07274499
BioProject: PRJNA438286
Submitter: Northwestern Polytechnical University
Date: 2022/03/03
Assembly level: Scaffold
Genome representation: full
GenBank assembly accession: GCA_022457185.1 (latest)
RefSeq assembly accession: n/a
RefSeq assembly and GenBank assembly identical: n/a
WGS Project: JAJJMQ01
Assembly method: SOAPdenovo v. 2
Expected final version: yes
Genome coverage: 200.0x
Sequencing technology: Illumina

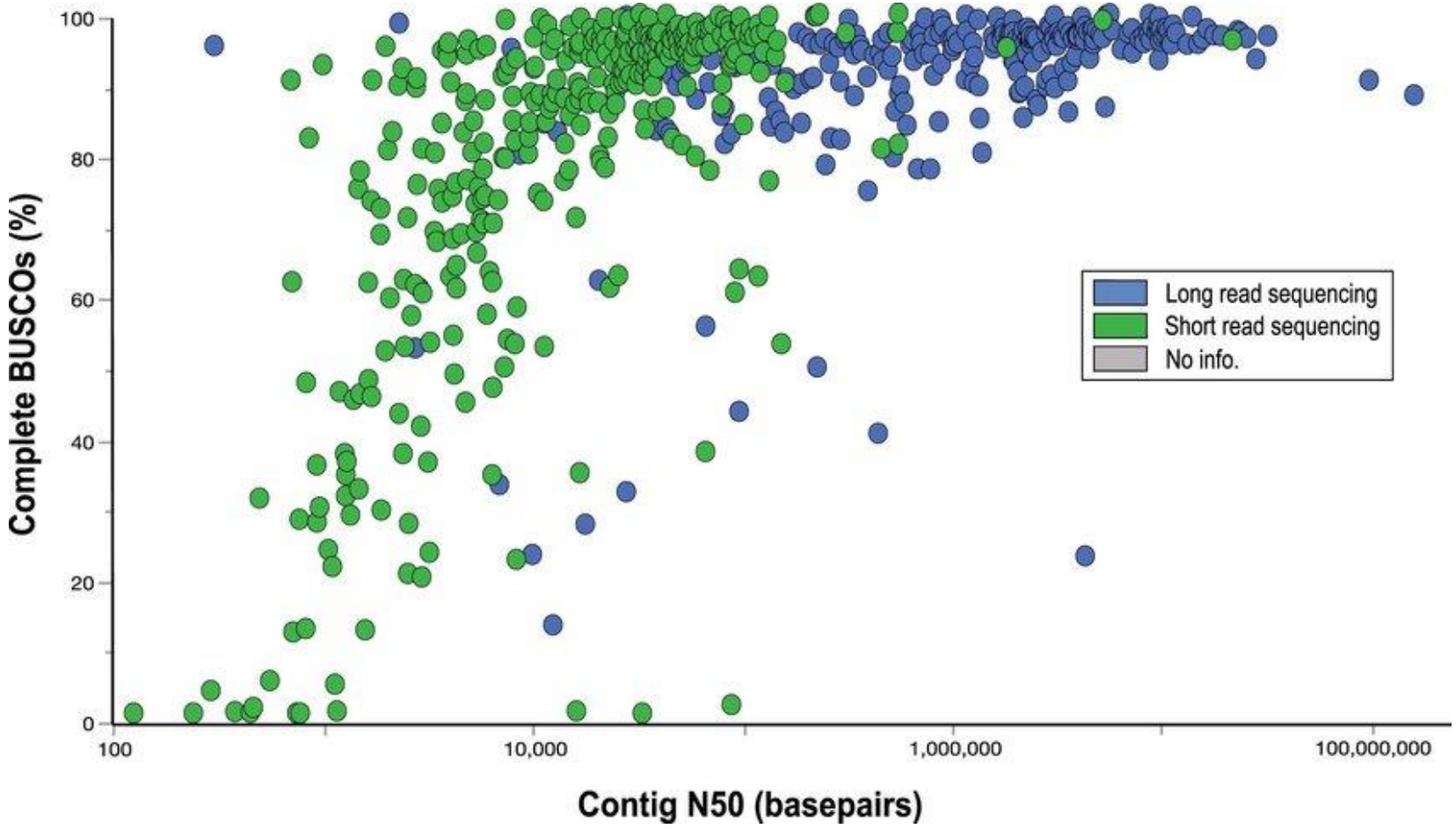
Draft vs reference quality genome II

EBP-Nor sequencing of reindeer from Svalbard:

- 30x PacBio HiFi reads
- 50x Arima Hi-C reads

Assembly	Reinsdyr hap1 (incl X and Y chromosomes)	Reinsdyr hap2
# scaffolds	1395	1291
Total scaffold length:	2.97 Gb	2.82 Gb
Contig N50:	22.48 Mb	25.53 Mb
Scaffold N50:	69.83 Mb	64.92 Mb
Largest scaffold:	157.94 Mb	119.78 Mb
Scaffolds placed in chromosomes (%)	89.22%	82.25%
BUSCOs percentage complete	96.3%	94.1%
BUSCOs complete	8883	8686
BUSCOs single	8548	8381
BUSCOs duplicated	335	305
BUSCOs fragmented	89	81
BUSCOs missing	254	459
BUSCOs total	9226	9226

Relationship between assembly contiguity and the percentage of complete BUSCOs

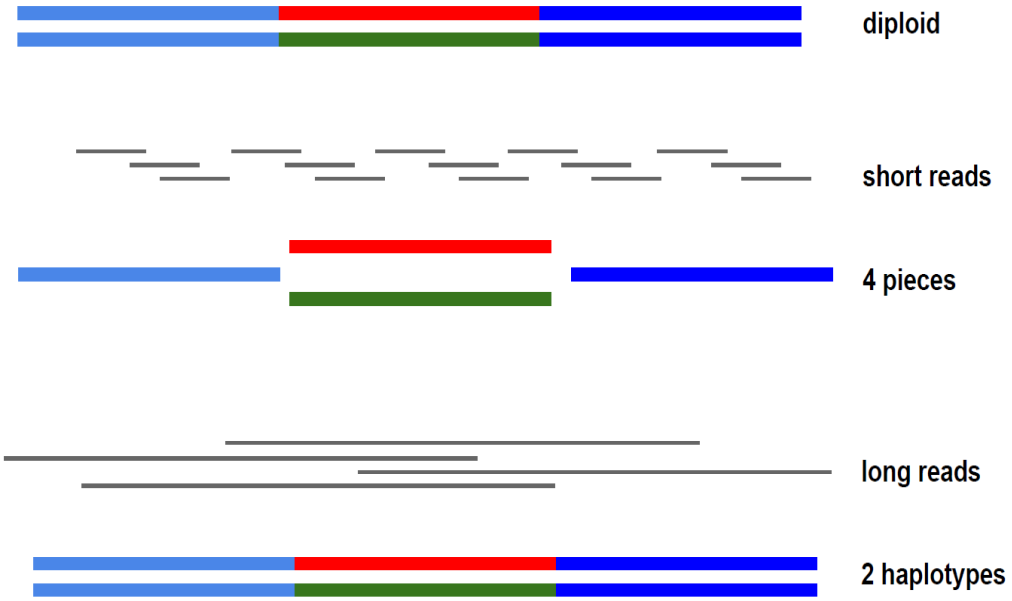


https://www.nature.com/articles/s41477-021-01031-8

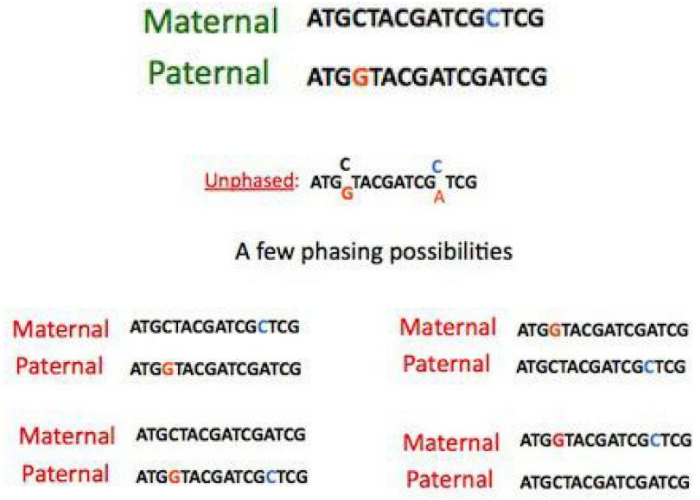
Why long reads?

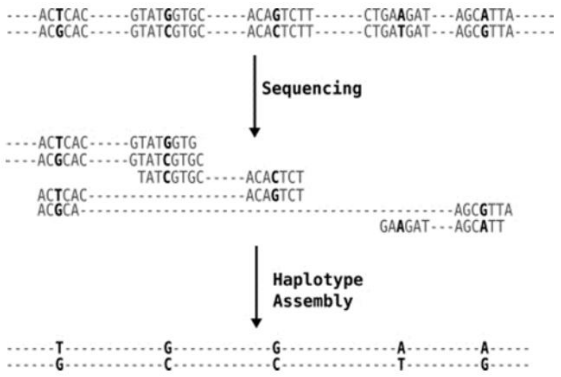
Phased haplotypes

- Genome assembly



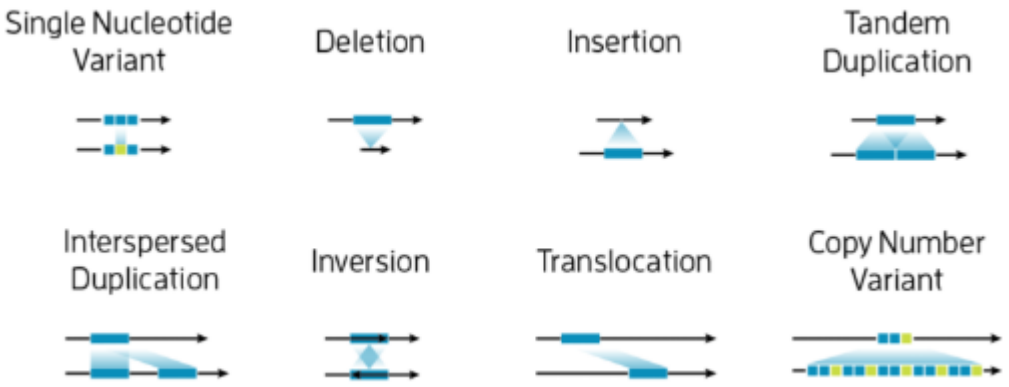
- Shorter regions (for example HLA typing)

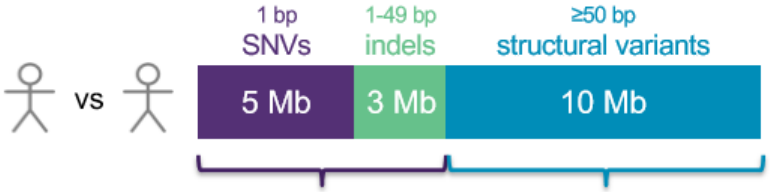




Why long reads?

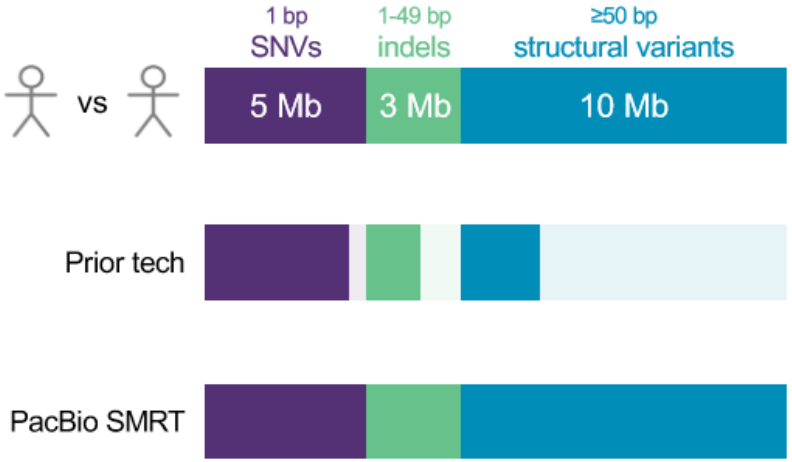
Structural variation – the missing heritability, not just SNVs





"Small variants":
- Single Nucleotide Variants (SNVs)
- Indels <50 bp

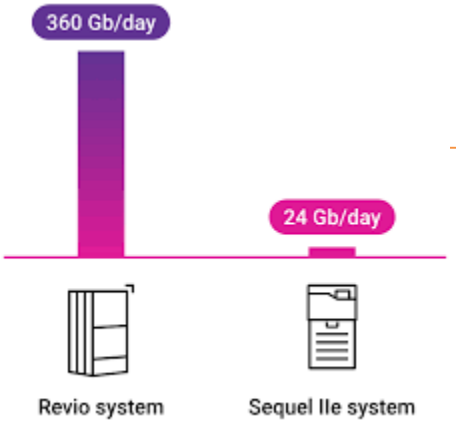
Structural Variants (SVs):
- Indels ≥50 bp
  - Duplications
  - Copy Number Variants (CNVs)
- Translocations
- Inversions

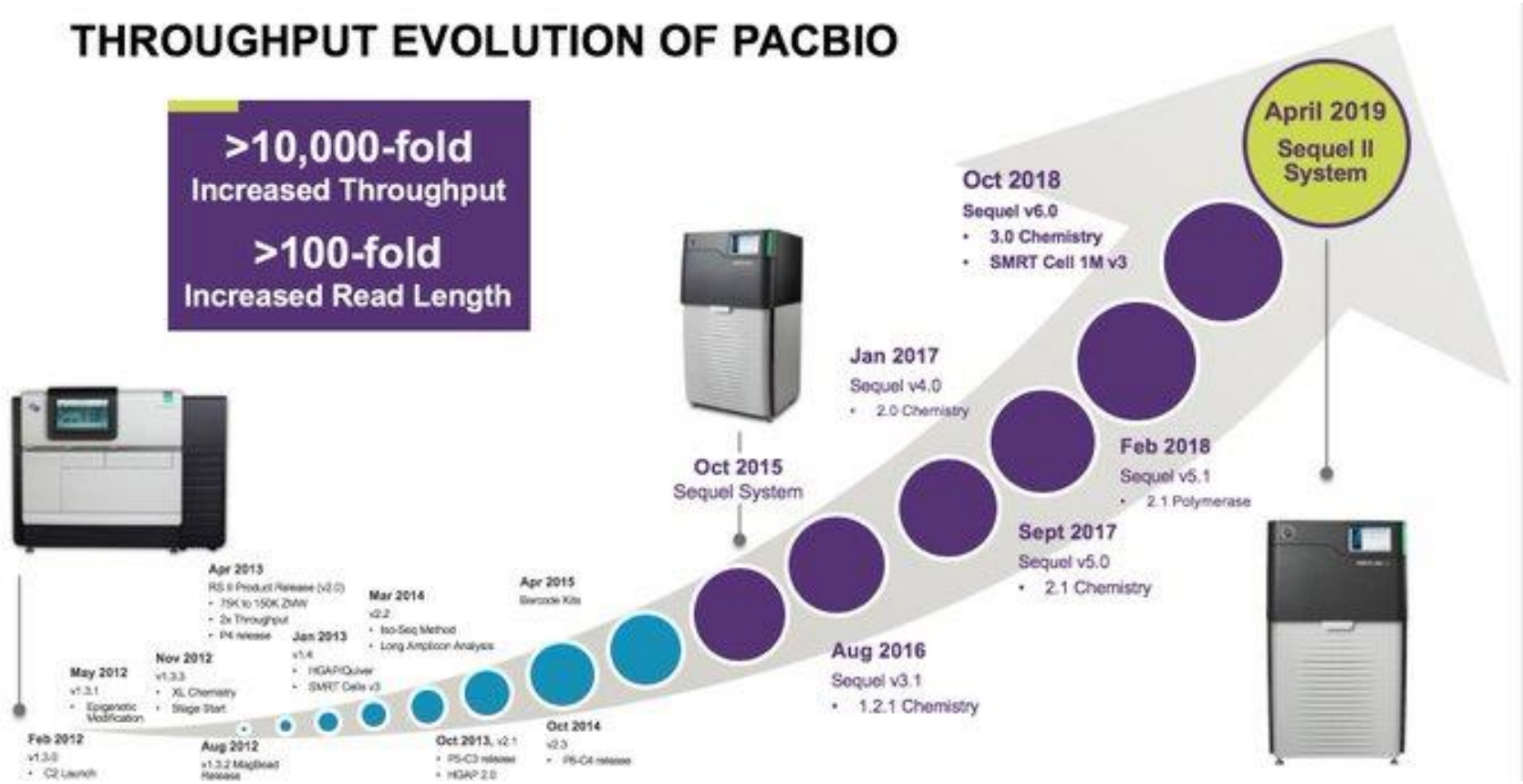


Short read vs long read sequencing

Short read sequencing	Long read sequencing
Amplification during sequencing	No amplification involved
High read accuracy	High consensus read accuracy
DNA requirements	
Works with almost any DNA sample	gDNA: pure HMW DNA needed
Fragmented DNA	DNA fragments at least 40-50 kb long
Low amount of DNA	High amount of DNA
	Low/Ultra-Low DNA input protocols available
Per base price	
Low	Medium/high

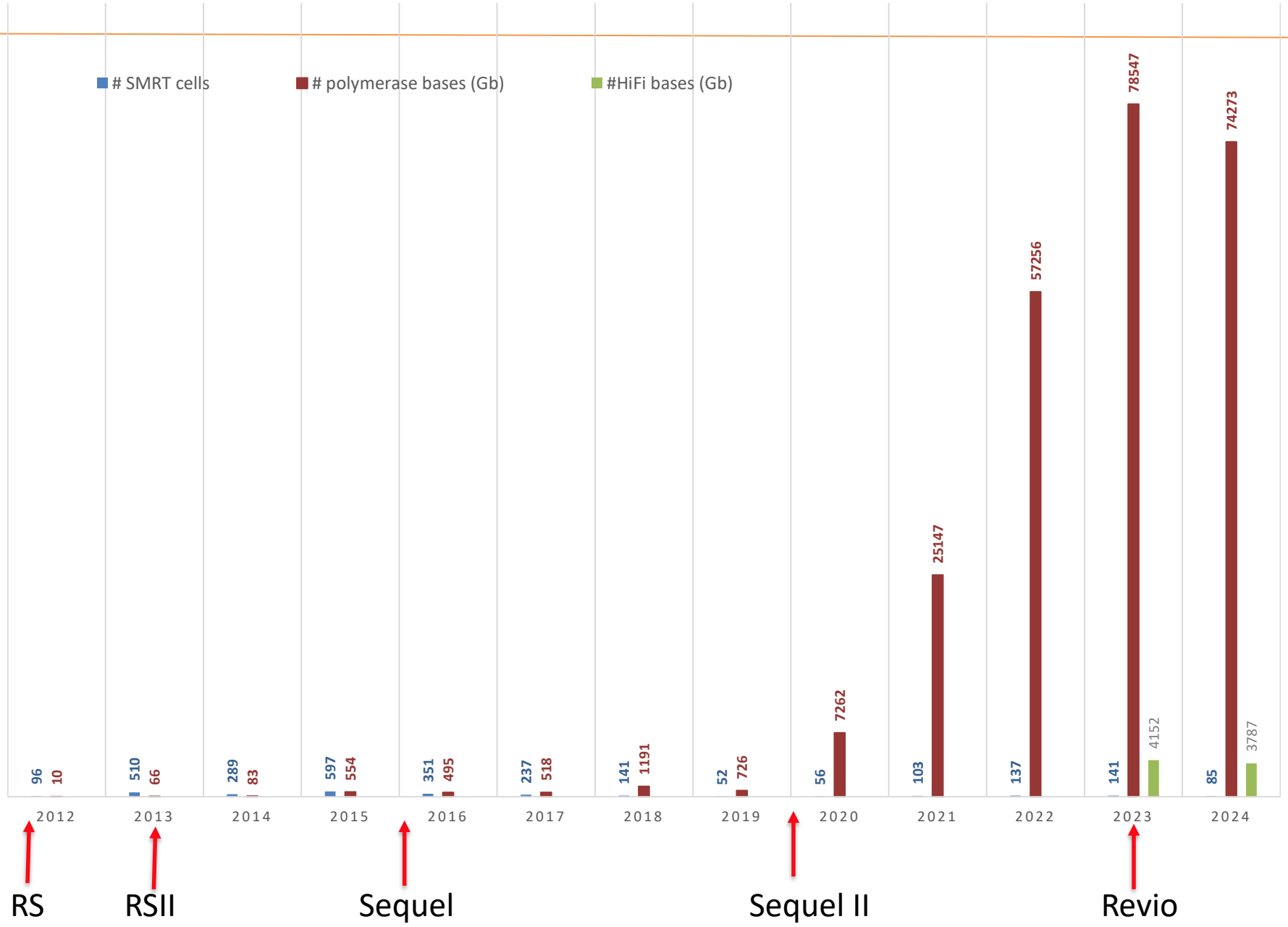
PacBio sequencing since 2012





Courtesy of Pacific Biosciences of California, Inc., Menlo Park, CA, USA

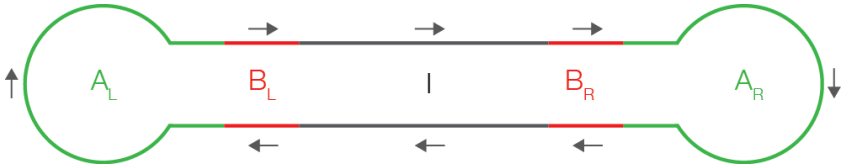
Throughput evolution at NSC



The PacBio sequencing technology

PACIFIC BIOSCIENCES™

SMRTbell library



How SMRT Sequencing Works

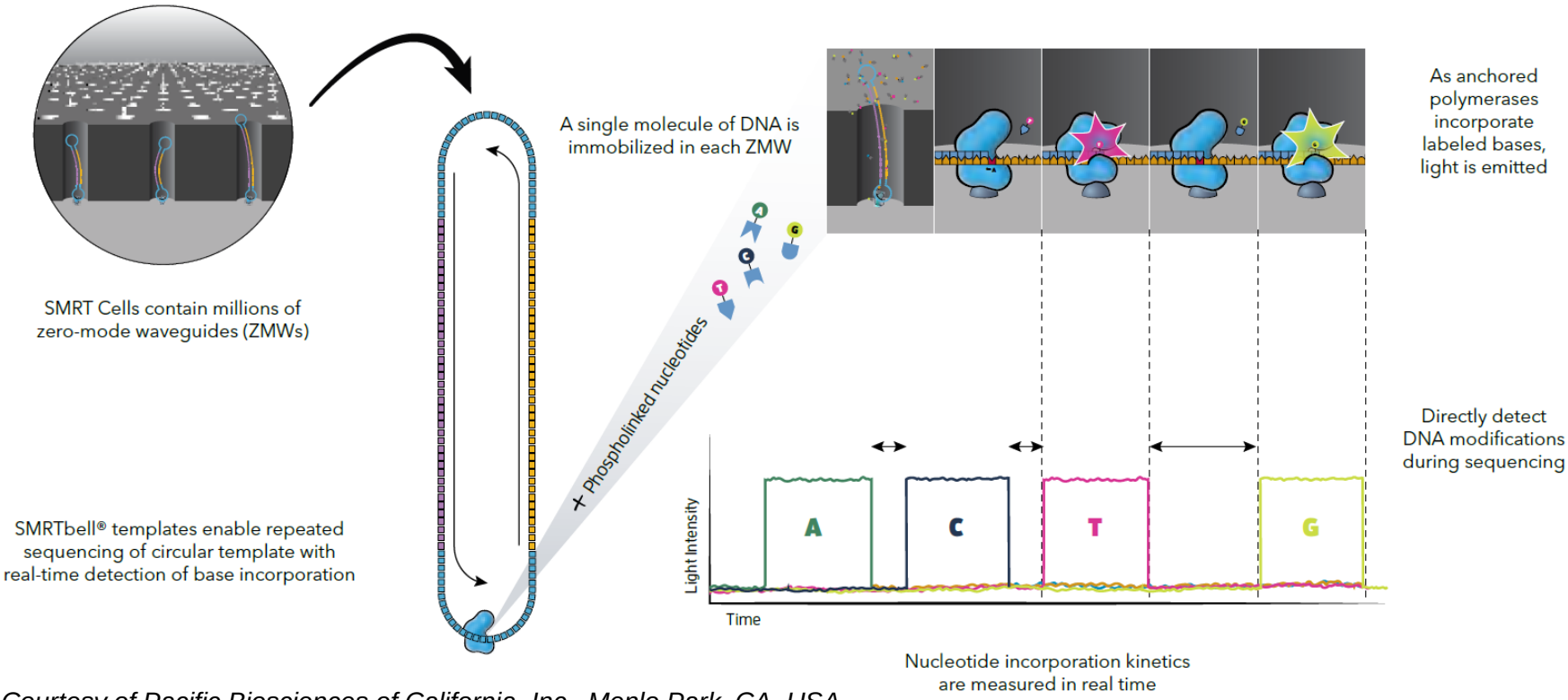


Image: Courtesy of Pacific Biosciences of California, Inc., Menlo Park, CA, USA

Measuring DNA methylation

SMRTbell® library

PacBio® long-read systems

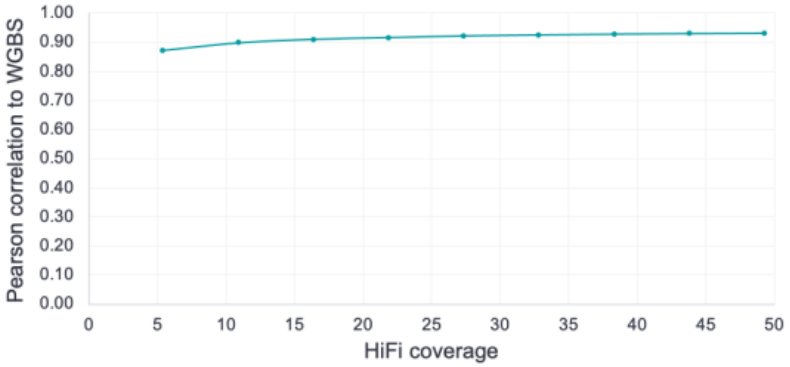
Nucleotide incorporation kinetics are measured in real time

5-base HiFi sequencing with A, C, G, T, +5mC

5mC encoded with standard BAM tags[3]
MM:Z:C+m,4,12,16,4,16,19,44,10
ML:B:C,249,4,247,177,210,228,245,244

The PacBio long-read systems directly output long, highly accurate HiFi reads with annotation of 5mC methylation at all CpG sites.
No special library preparation like bisulfite treatment is required.

Coverage



Correlation of methylation calling in HiFi reads to whole-genome bisulfite sequencing (WGBS) of the human sample HG002.[4,5,6]

Applicability

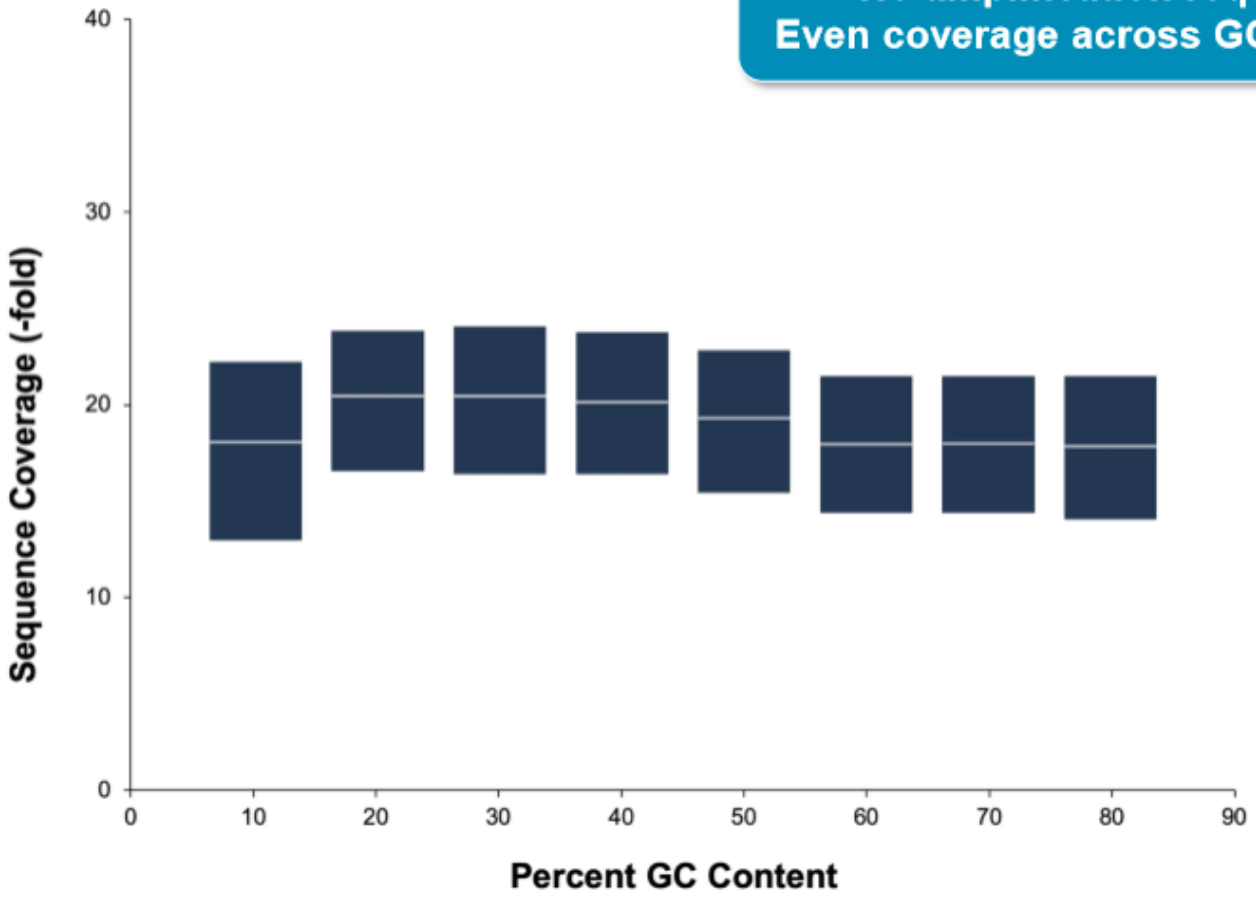
Methylation	Species	5-base HiFi sequencing
5mC at CpG sites	Human and other vertebrates	✓
5mC at various motifs	Other eukaryotes, including plants	✓ Useful though partial view
4mC and 6mA	Microbes	Enabled through SMRT® Link microbial genome analysis

PacBio

Sequence performance: uniformity

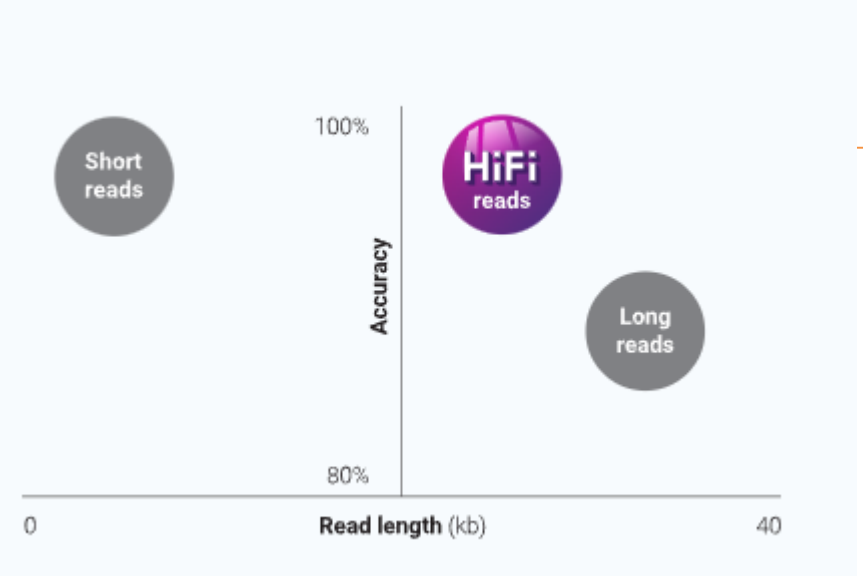
UNIFORM COVERAGE

No amplification required
Even coverage across GC content

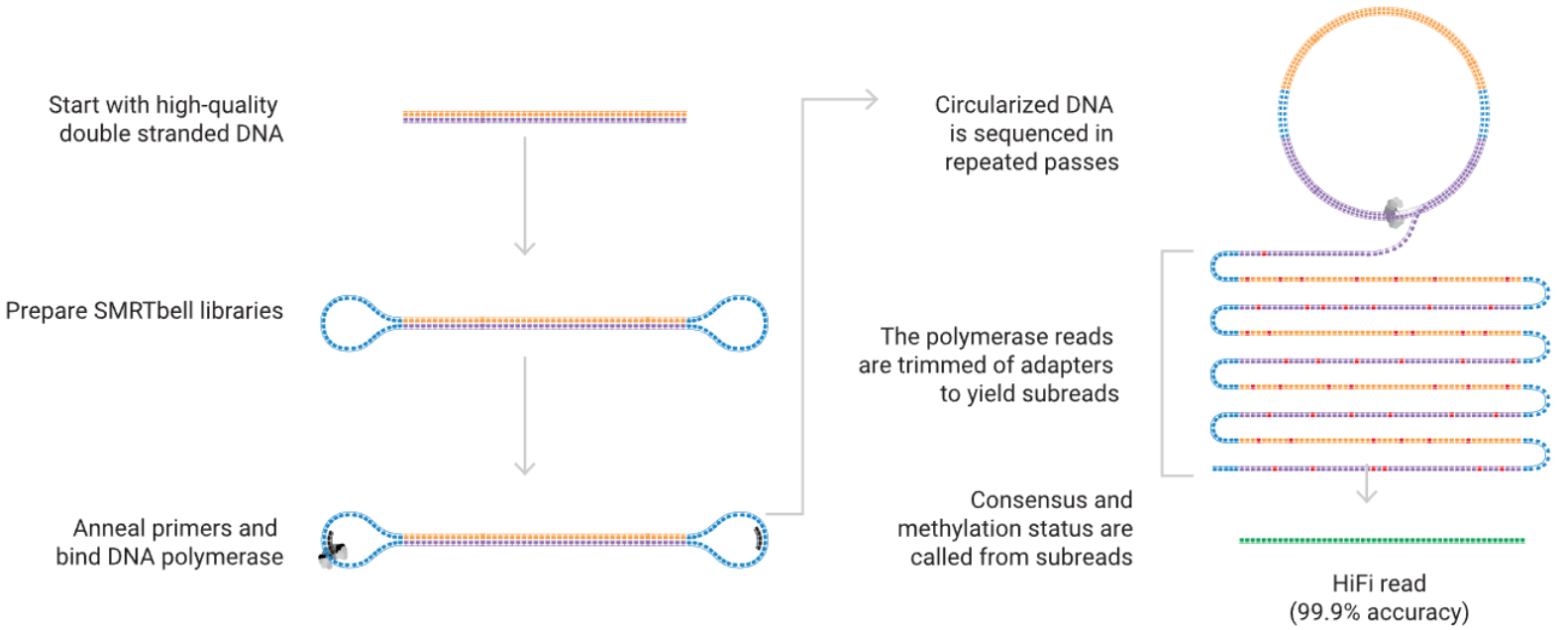


Mean coverage per GC window across a human sample. Data generated with a 15 kb human HiFi library on a Sequel II System using 2.0 Chemistry and Sequel II System Software v8.0

HiFi sequencing:

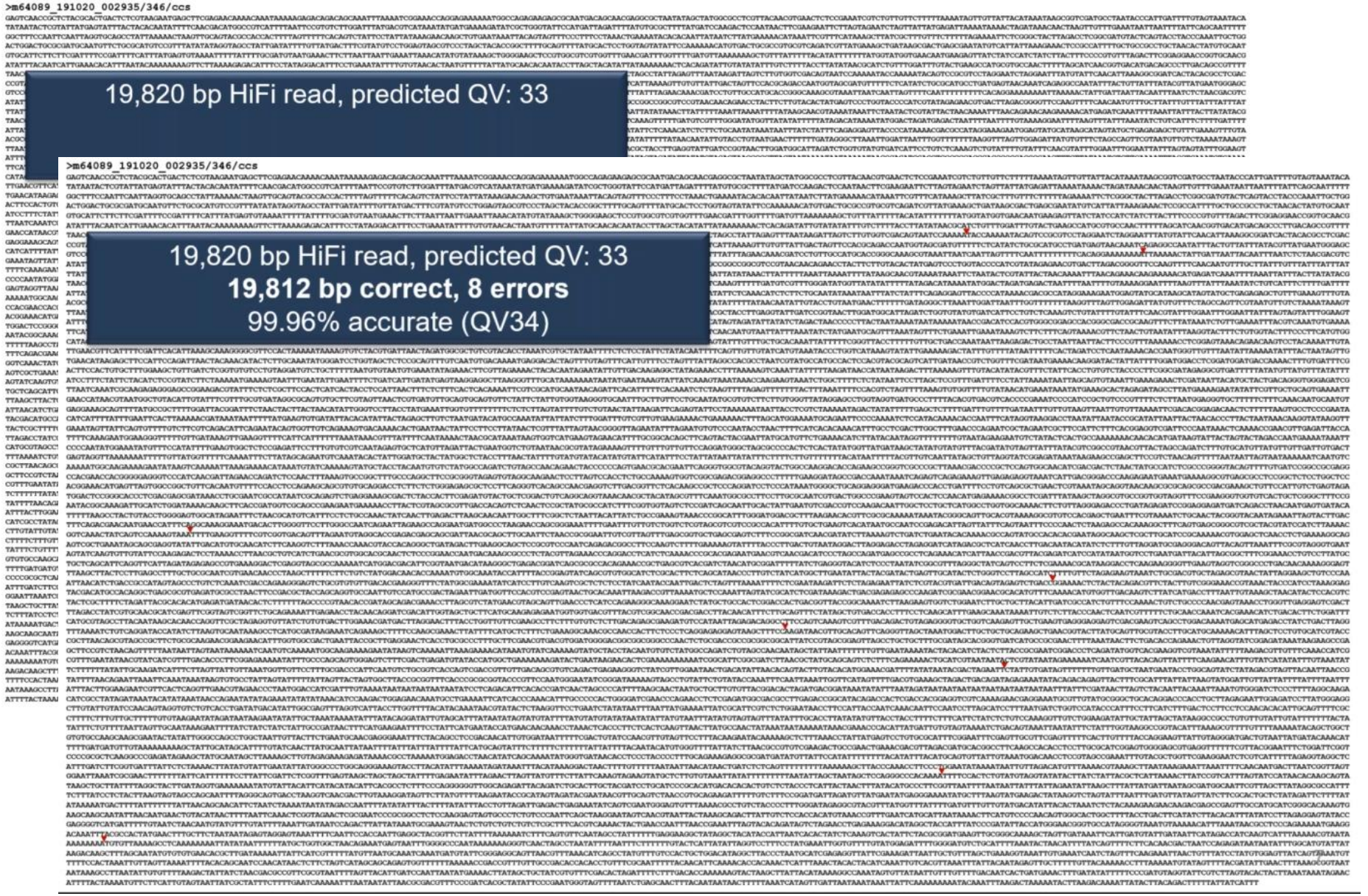


How are HiFi reads generated?

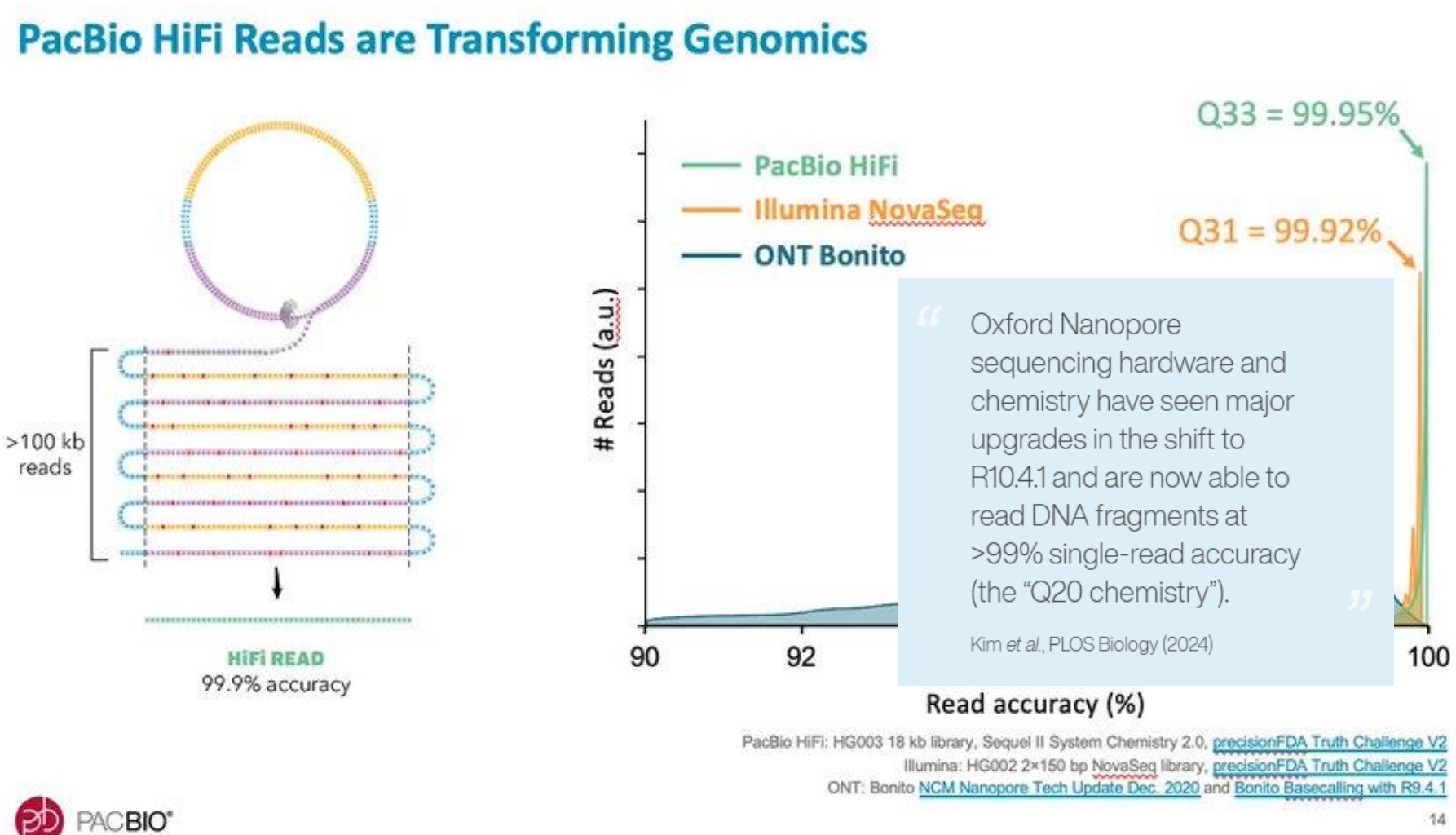


Courtesy of Pacific Biosciences of California, Inc., Menlo Park, CA, USA

HiFi read – long and accurate!

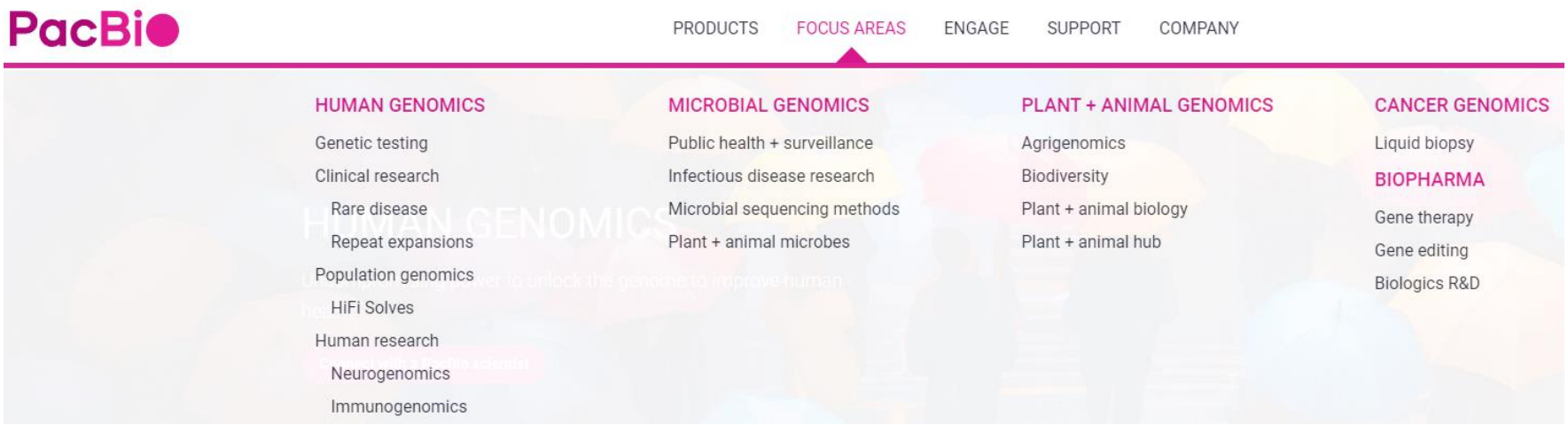


Read accuracy comparison between different sequencing platforms:



Courtesy of Pacific Biosciences of California, Inc., Menlo Park, CA, USA

PacBio long read applications



Whole genome sequencing

PacBio

	Application	Whole genome sequencing			
		De novo genome assembly	Variant detection	Microbial *de novo* genome assembly	*De novo* genome assembly w/ ultralow input
Experimental design	**Value proposition**	Produce reference-quality, haplotype-phased genome assemblies including 5mC methylation profiles	Detect and phase variants: SNVs, indels, SVs, tandem repeats, DNA methylation	Produce accurate, closed assemblies of chromosomes and plasmids. Detect DNA methylation profiles	Produce high-quality, haplotype-phased genome assemblies.
	Coverage	15X / haplotype	10X for SVs; 30X for all variant classes	15X / microbe	15X / haplotype
	Library insert size	15 - 20 kb	15 - 20 kb	7 - 10 kb	10 - 12 kb
	Multiplexing: Sequel II/IIe SMRT Cell	1 Gb of genome	n/a	96 microbes up to 375 Mb of total genome	1 Gb of genome
	Multiplexing: Revio SMRT Cell	3 Gb of genome	SV: 3 humans All variants: 1 human	96 microbes up to 1 Gb of total genome	3 Gb of genome
Data analysis tools	**SMRT® Link workflows**	Genome Assembly	Variant Calling	Microbial Genome Analysis	Genome Assembly
	Community tools	hifiasm	DeepVariant	hifiasm	hifiasm

WHOLE GENOME SEQUENCING — HOW PACBIO COMPARES

	PacBio HiFi	Illumina	Oxford Nanopore
Average read length[1]	15–20 kb	2 x 150 bp	10–100 kb
Average read accuracy[1]	99.95% (Q33)	99.92% (Q31)	99.26% (Q21)
Coverage[2]	Unbiased	Reduced at low and high [GC]	Reduced in low-complexity runs
Variant calling: SNVs	✓	✓	✓
Variant calling: indels	✓	✓	✗
Variant calling: SVs	✓	✗	✓
Genome assembly: contiguity	✓	✗	✓
Genome assembly: accuracy	✓	✓	✗
Epigenetics: 5mC	✓	✗	✓

1. PacBio HiFi: HG003 18 kb library, Sequel II system chemistry 2.0, precisionFDA *Truth Challenge* V2 (https://doi.org/10.1101/2020.11.13.380741), Illumina: HG002 2×150 bp NovaSeq library, precisionFDA *Truth Challenge* V2 (https://doi.org/10.1101/2020.11.13.380741), ONT: Q20+ chemistry (R10.4, Kit 12), Oct 2021 GM24385 Q20+ Simplex Dataset Release (https://labs.epi2me.io/gm24385_q20_2021.10/)

2. HiFi+ONT: Nurk 2021 https://doi.org/10.1101/2021.05.26.445798, HiFi+Illumina: Logsdon 2020 https://doi.org/10.1038/s41576-020-0236-x, ONT: Tan 2022 https://doi.org/10.1101/2022.01.11.475254

Whole genome sequencing for *De novo* assembly

SAMPLE & PROJECT CONSIDERATIONS	STANDARD HIFI SEQUENCING	LOW DNA INPUT SEQUENCING (2-PLEX)	LOW DNA INPUT SEQUENCING (SINGLE SAMPLE)	ULTRA-LOW DNA INPUT SEQUENCING
Minimum DNA Input	≥5 µg for a 3-Gb genome	300 ng for each genome	400 ng	5 ng
Amplification Based?	No	No	No	Yes
Genome Size Limit	N/A	600 Mb for each genome	1 Gb	500 Mb
Supported Applications	*De novo* Assembly Human Variant Detection	*De novo* Assembly	*De novo* Assembly	*De novo* Assembly Human Variant Detection

Ultra-Low DNA Input: SUPPORTED APPLICATIONS

☑ **ASSEMBLY**

De novo assembly of insect/arthropod genomes (Up to 500 Mb)

☑ **VARIANT DETECTION**

Variant detection (SNPs, Indels, SVs) in human genomes (3 Gb)

Ultra-Low DNA Input: UNSUPPORTED APPLICATIONS

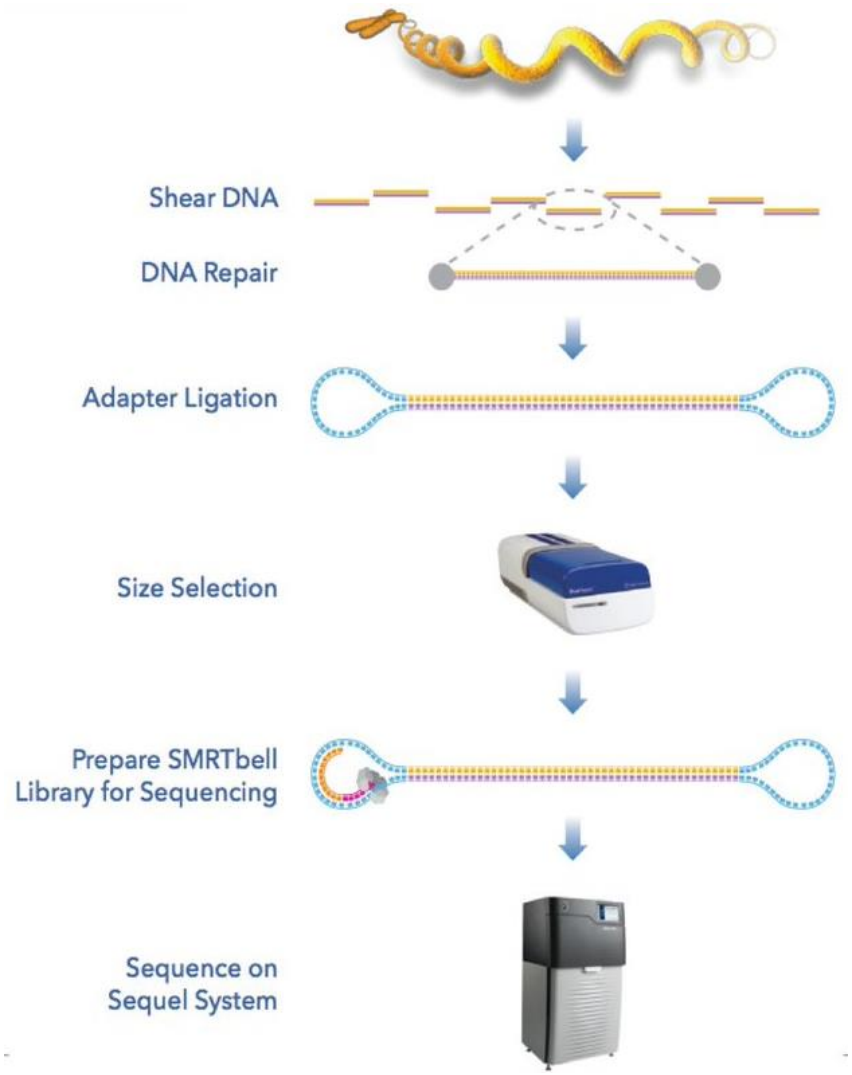
☒ **ASSEMBLY**

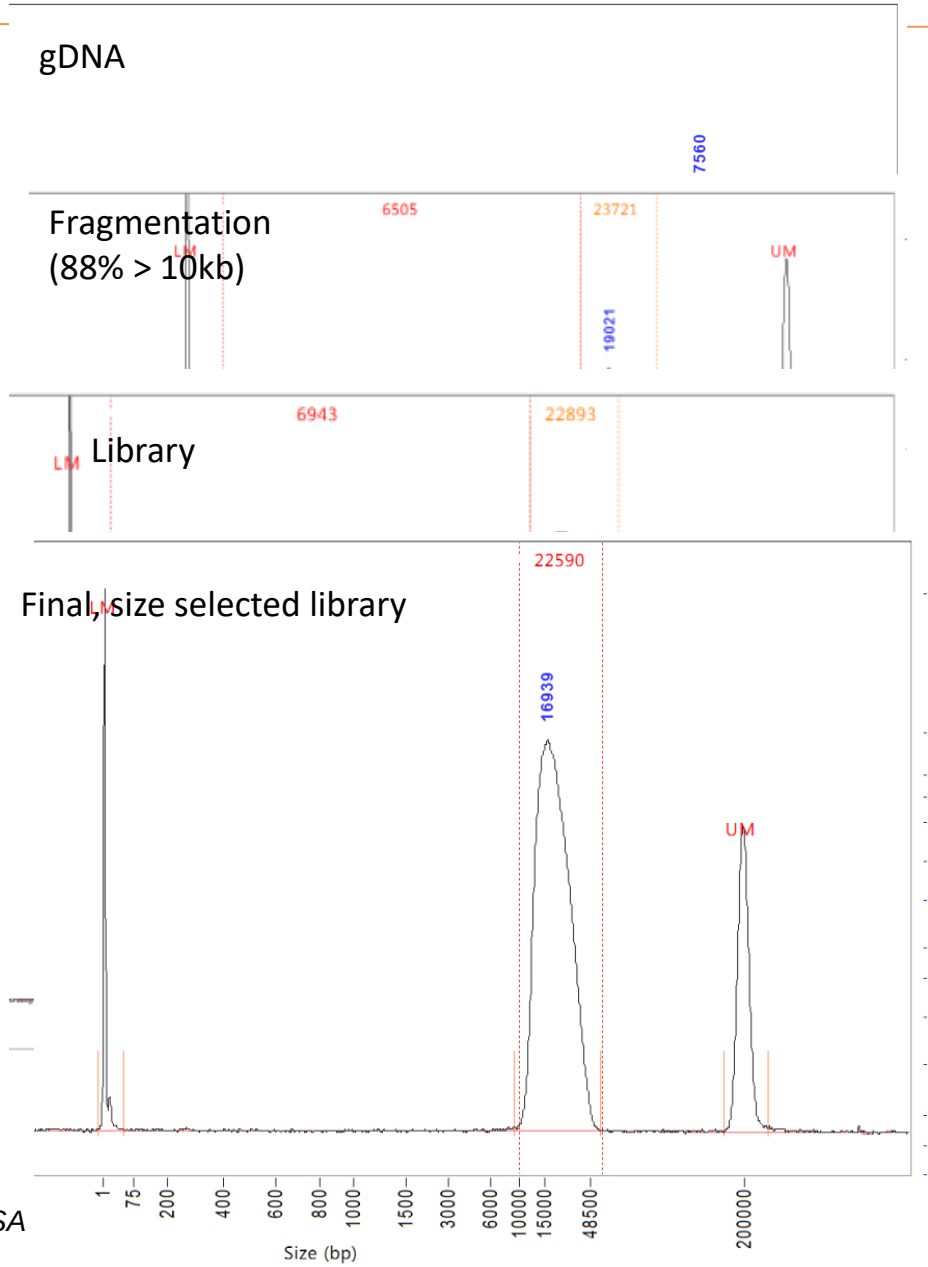
De novo assembly for microbes, plants, vertebrates, or other non-DNA limited sample types

☒ **COMPLEX POPULATIONS**

Metagenomics sequencing

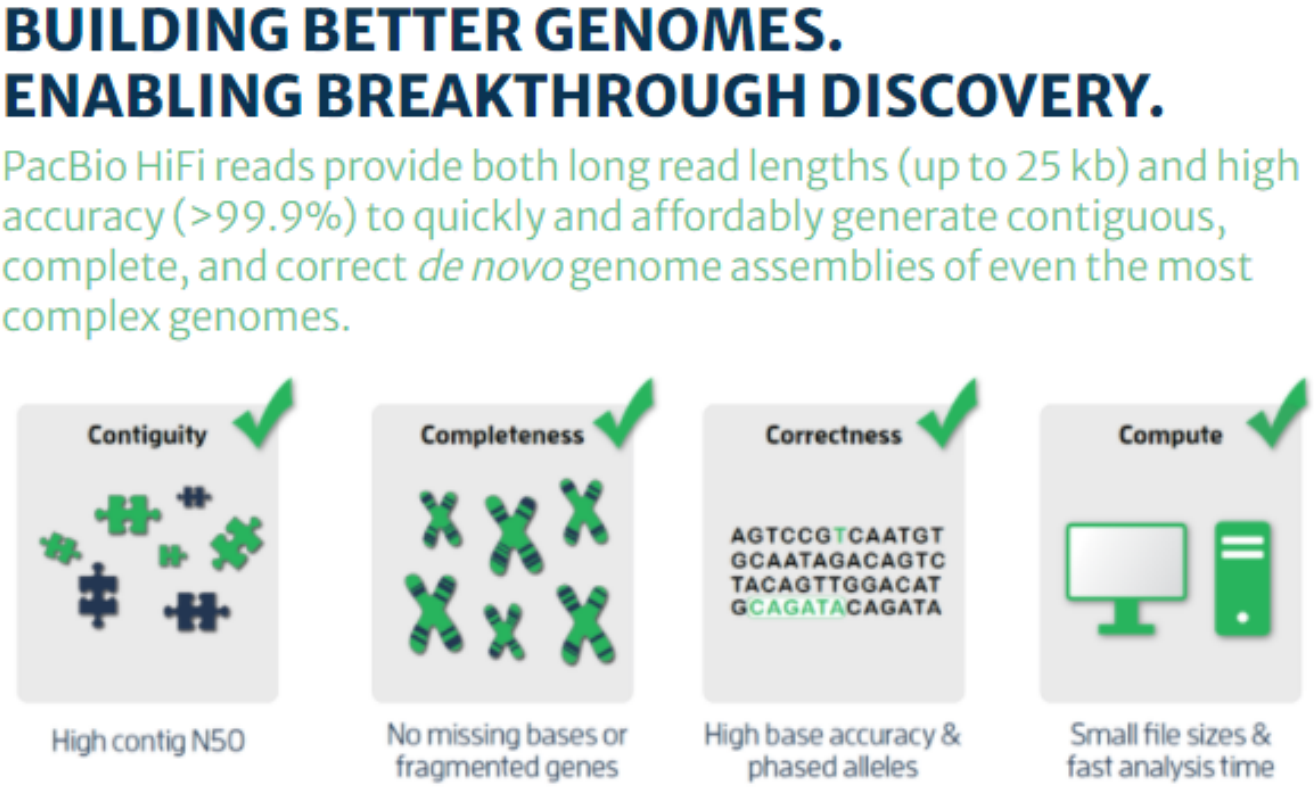
Library prep for reference quality assemblies





Courtesy of Pacific Biosciences of California, Inc., Menlo Park, CA, USA

Whole genome sequencing for *De novo* assembly



What about hybrid approaches?

- ✗ **HiFi reads + short reads:** no benefit for contig building or polishing
- ✗ **HiFi reads + long reads:** may have marginal benefit to contiguity, but no readily available tools
- ✓ **HiFi + scaffolding:** technologies like optical maps and HiC help assign your high-quality HiFi genome assemblies into chromosomes

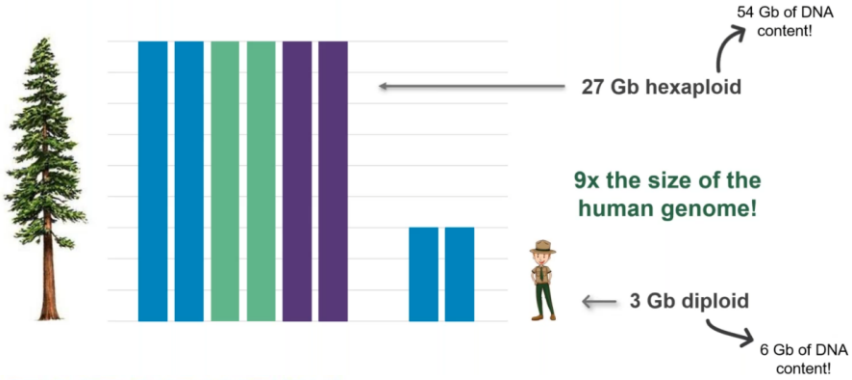
HiFi assembly of large genomes - redwood

THE CALIFORNIA (COASTAL) 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



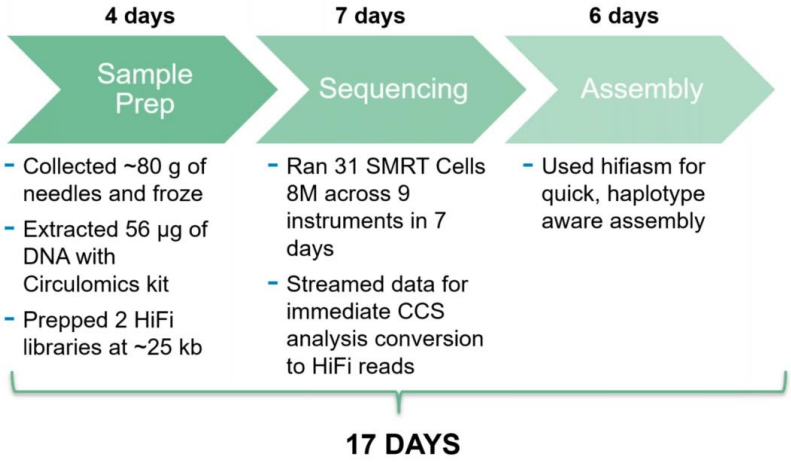
RESULTING GENOME ASSEMBLY

- Standard running parameters – no iteration
- Run on 64 cores with 512 Gb of RAM – no specialized or particularly large compute cluster

California Redwood Genome Assembly Results		
Methodology	**PacBio HiFi reads**	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%

BUSCO does not work well in conifers due to very long introns

THE PROJECT WORKFLOW

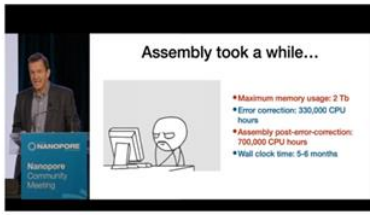


PacBio HiFi reads[1]

- 64 cores with 512 Gb of RAM
- ~46,000 CPU hours for HiFi generation ("error correction")
- 6 days wall time, ~7,200 CPU hours for assembly

6 days vs ***5-6 months*** of wall time for just genome assembly

ONT + short reads[2]



1. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes
2. Using transcript set of *Abies alba* from Neale, D. et al. consisting of 22,561 transcript sequences

Courtesy of Pacific Biosciences of California, Inc.

Procedure & Checklist - Preparing HiFi SMRTbell® Libraries from Ultra-Low DNA Input

Required gDNA Input Amount	Required Quality of Input gDNA	gDNA Shearing Method	Target Sheared Fragment Size Distribution Mode	Amplification Target Size Distribution Mode	Total Mass of Pooled PCR Product Required for Library Construction	Required SMRTbell Library Input for BluePippin Size-Selection
5-20 ng	Majority of gDNA >20 kb	Megaruptor or g-TUBE	10 kb sheared DNA is optimal	8-10 kb	≥500 ng	≥400ng

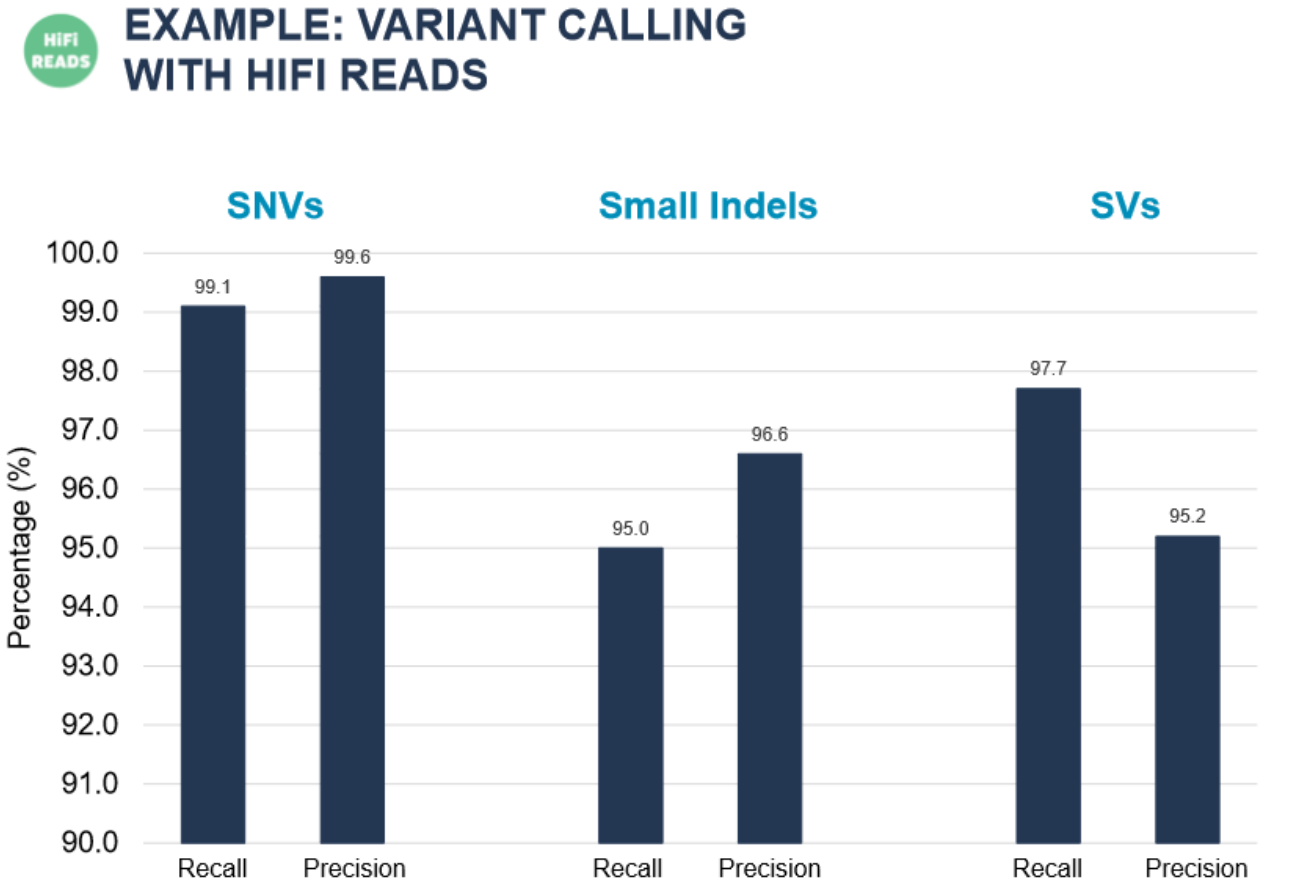
Research | Open access | Published: 07 October 2024

Chromosome-level genome assembly of the sacoglossan sea slug *Elysia timida* (Risso, 1818)

Lisa Männer, Tilman Schell, Julia Spies, Carles Galià-Camps, Damian Baranski, Alexander Ben Hamadou, Charlotte Gerheim, Kornelia Neveling, Eric J. N. Helfrich & Carola Greve

BMC Genomics **25**, Article number: 941 (2024) | Cite this article

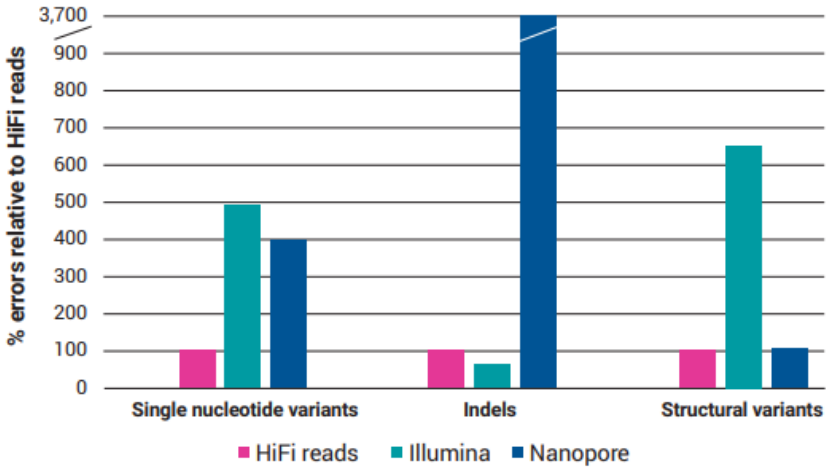
Variant calling



Variant calls from ~15-fold HiFi read coverage of a human genome (HG002) were measured against the Genome in a Bottle small variant benchmark (v3.3.2) for SNVs and indels using Deep Variant and SMRT Link 8.0 for SVs. Libraries were generated using a 15 kb insert and sequenced using Chemistry 2.0.

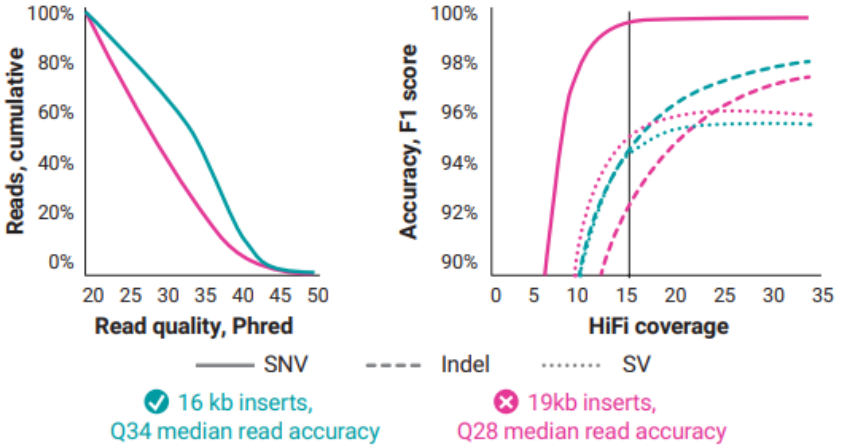
Variant calling

Low error rates for all variant types



Variant calling performance against *Genome in a Bottle* benchmarks for PacBio HiFi reads (35-fold, Sequel II system, 2.0 chemistry); Illumina (35-fold, NovaSeq); Oxford Nanopore (60-fold, PromethION R9.4.1).

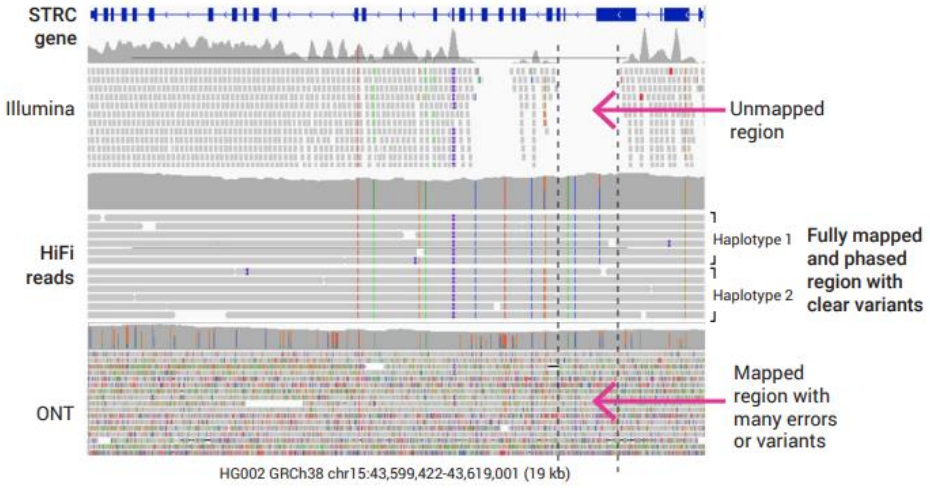
High precision and recall at 15-fold coverage



Variant calling performance against *Genome in a Bottle* benchmark v4.2 (Sequel II system, 2.2 chemistry; DeepVariant v1.1). Read lengths 15–18 kb are recommended to achieve the highest read and variant calling accuracy.

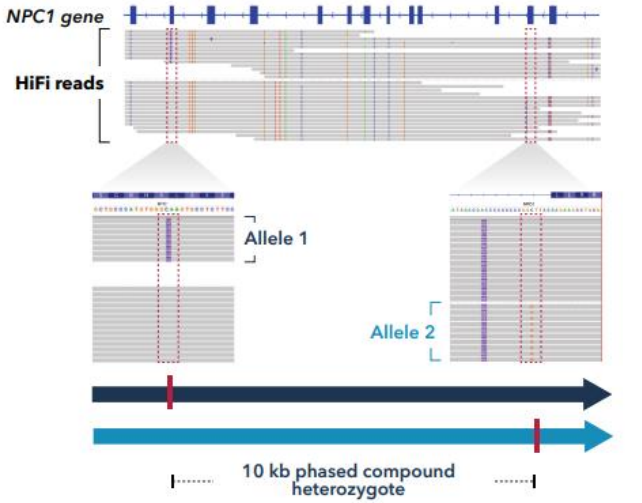
Variant calling

Detect more variants in medically relevant genes[2]



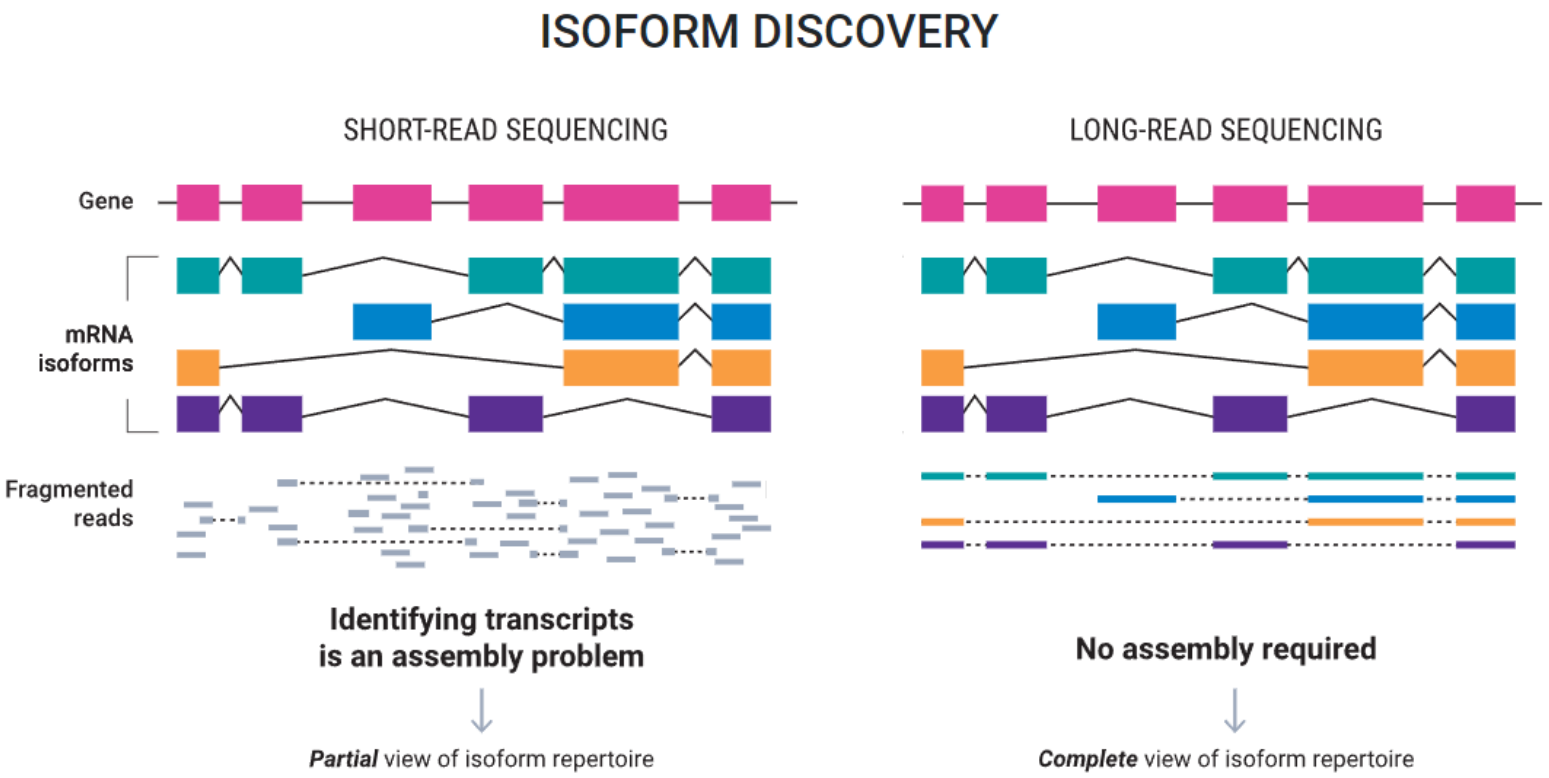
STRC gene alignments from *Genome in a Bottle* (GIAB), HG002_NA24385_son.

Phase all variants into haplotypes[3]



NPC1 gene showing a phased compound heterozygote.

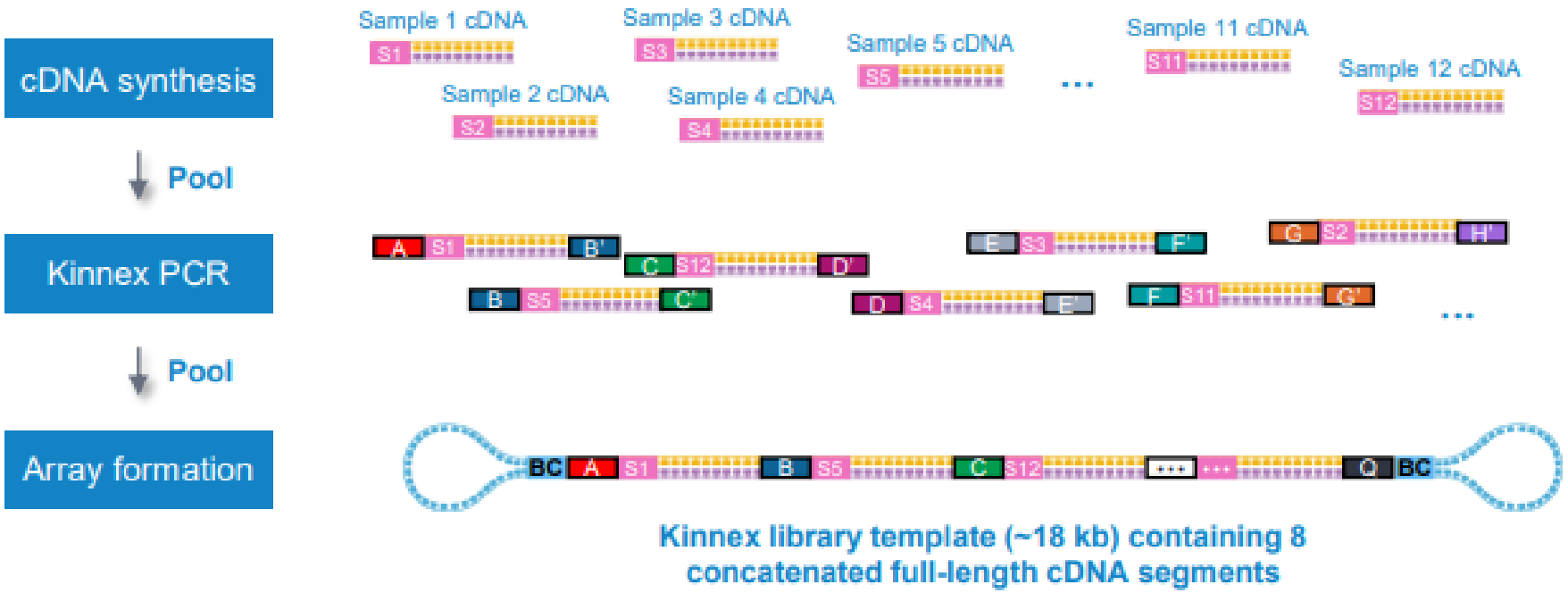
Long read RNA sequencing



- Characterize alternative splicing (AS) events, including alternative start sites, end sites, intron retention, and exon-skipping events
- Find gene fusions in tumor samples
- Identify allele-specific isoforms
- Detect differentially expressed isoforms and isoform switching events
- Predict functional impact of novel isoforms through open reading frame (ORF) prediction

Higher output with Kinnex library prep

Input: 300 ng total RNA (RIN >7)
40 mill reads from a Revio 25M SMRT cell



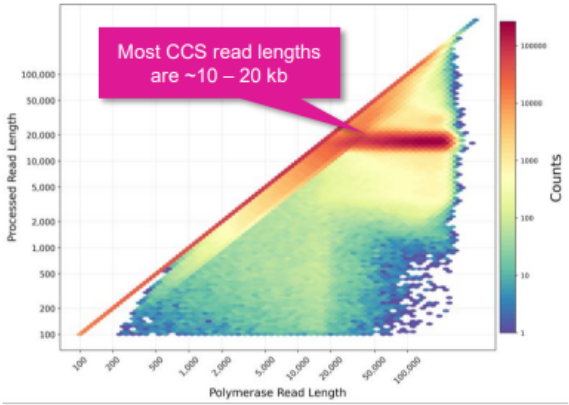
Single cell Kinnex

- 10x Chromium Single Cell 3’ kit (v3.1) and 5’ kit (v2)
- 15–75 ng cDNA input
- 3,000 to 10,000 target cell recovery

Example sequencing performance for Kinnex single-cell RNA libraries prepared with human cDNA

Revio system example data[1] – Kinnex single-cell RNA 3’ library sample

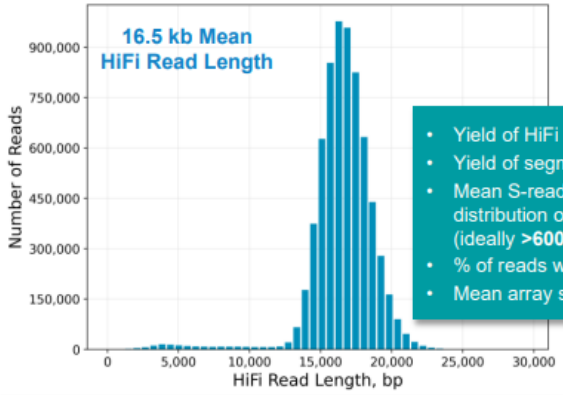
Raw Data Report



Raw Base Yield	1,289 Gb
Mean Polymerase Read Length	73.16 kb
P0	27%
P1	70%
P2	3%

Example sequencing metrics for a human Kinnex single-cell RNA 3’ library sample run on a Revio system with Revio polymerase kit / 130 pM on-plate loading concentration (OPLC) / 24-hrs movie time.

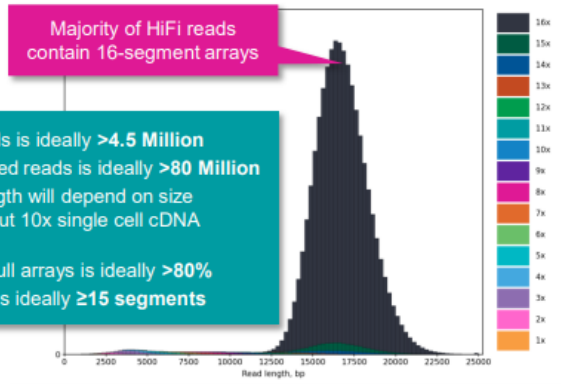
HiFi Read Length



HiFi Reads	6.7 M
HiFi Base Yield	111.24 Gb
Mean HiFi Read Length	16.55 kb
Median HiFi Read Quality	Q28
HiFi Read Mean # of Passes	8

For human Kinnex single-cell RNA libraries, per-Revio SMRT Cell HiFi read counts were typically ~4 – 7 Million depending on the final library insert size and *P1* loading performance.

Read Segmentation Metrics



- Yield of HiFi reads is ideally **>4.5 Million**
- Yield of segmented reads is ideally **>80 Million**
- Mean S-read length will depend on size distribution of input 10x single cell cDNA (ideally **>600 bp**)
- % of reads with full arrays is ideally **>80%**
- Mean array size is ideally **≥15 segments**

Input HiFi Reads	6,673,602
Segmented reads (S-reads)	104,869,257
Mean length of S-reads	1,031 bp
Percent of reads with full arrays	93.89%
Mean array size (concentration factor)	15.71

For Kinnex single-cell RNA libraries, per-Revio SMRT Cell segmentation read counts were typically >80 Million.

THE RIGHT SOLUTION FOR YOUR RNA APPLICATIONS

	Genome annotation	Whole transcriptome	Single-cell transcriptome
Goal	Comprehensive, high-quality genome annotation for plant + animal organisms	- Isoform discovery in disease cohorts - Differential isoform expression analysis in disease vs normal samples	Cell-type specific, allele-specific isoform and variant characterization in single-cell studies
Library prep	Kinnex full-length RNA kit	Kinnex full-length RNA kit	Kinnex single-cell RNA kit
Sequencing recommendation	1 Revio SMRT Cell for 40 million reads total (5-10 million reads per tissue)	1 Revio SMRT Cell for 40 million reads total (5-10 million reads per sample)	1 Revio SMRT Cell for 80-100 million reads for one single-cell library
Analysis	SMRT Link followed by tertiary tools	SMRT Link followed by tertiary tools	SMRT Link followed by tertiary tools

Systematic assessment of long-read RNA-seq methods for transcript identification and quantification

Francisco J. Pardo-Palacios, Dingjie Wang, Fairlie Reese, Mark Diekhans, Sílvia Carbonell-Sala, Brian Williams, Jane E. Loveland, Maite De María, Matthew S. Adams, Gabriela Balderrama-Gutierrez, Amit K. Behera, Jose M. Gonzalez Martinez, Toby Hunt, Julien Lagarde, Cindy E. Liang, Haoran Li, Marcus Jerryd Meade, David A. Moraga Amador, Andrey D. Prjibelski, Inanc Birol, Hamed Bostan, Ashley M. Brooks, Muhammed Hasan Çelik, Ying Chen, ... Angela N. Brooks + Show authors

Nature Methods **21**, 1349–1363 (2024) | Cite this article

Metagenomics

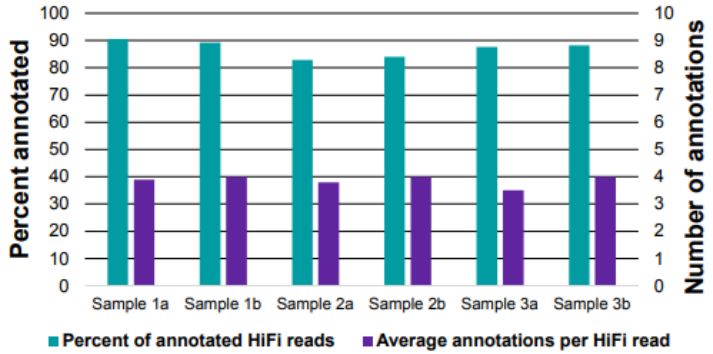
PacBio

	Application	Metagenomics	
		Full-length 16S rRNA	Shotgun profiling or assembly
Experimental design	Value proposition	Obtain strain level resolution from 16S rRNA	Generate near-complete assemblies of high-complexity samples (e.g. gut microbiome)
	Coverage	8,000 reads / sample	See Best Practices Guide
	Library insert size	1.5 kb amplicon, 15 - 20 kb Kinnex array	10 kb
	Multiplexing: Sequel II/IIe SMRT Cell	Up to 1,536 samples	Profile: 48 communities Assemble: 4 communities
	Multiplexing: Revio SMRT Cell	Up to 1,536 samples	Profile: 96 communities Assemble: 12 communities

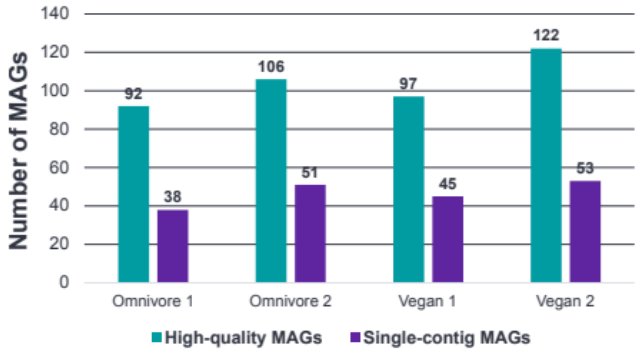
Metagenomics

HiFi metagenomics

Generate precise species characterization, more functional annotations, more HQ MAGs, and more circular MAGs, even at lower coverage. HiFi metagenomes generated with the HiFi plex prep kit 96[3] or HiFi prep kit 96[4] are cost competitive with short-read metagenomes.



With up to nine complete genes per HiFi read, PacBio® data provides rich functional information; nearly every read contributes to your understanding of the biological functions present in your microbial community.[5]



The unique combination of long read lengths and high accuracy overcomes many challenges involved with metagenome assembly such as distinguishing closely related strains in the same sample and yielding single-contig MAGs. Analysis of four human gut microbiome samples from *The BioCollective*[6] shows HQ MAGs ≥70% completeness, <10% contamination, <10 contigs.

Obtain more and richer metagenome functional information

- ~80–90% of HiFi reads are functionally annotatable
- Each HiFi read typically has an average of four functional annotations

Achieve standout metagenome assemblies

- ~90–125 HQ MAGs per sample with ~17 Gb data; many are single-contig with ~50 per sample
- 417 HQ MAGs in total across four samples

PacBio

Metagenomics: 16S rRNA

Full-length 16S rRNA sequencing

Achieve species- and strain-level phylogenetic resolution with the Kinnex™ 16S rRNA kit[1] at a highly competitive cost relative to short-read partial 16S sequencing.

The proportion of 16S sequences from each bacterial genus that cannot be identified at the species level varies significantly depending on which variable region is used. Since the human gut can harbor a broad diversity of bacterial clades, only full-length sequences (V1–V9) can provide unbiased resolution of all the species that may be present.[2]

Kinnex 16S rRNA application use case recommendations for PacBio systems

	Sequel II and IIe systems		Revio system	
Experimental goal	Determine the microbial diversity (phylogeny and taxonomy) of bacteria in a metagenomic sample			
Sample multiplexing[1]	Up to 384 samples per SMRT Cell 8M (384-plex)		Up to 1,536 samples per Revio SMRT Cell (1536-plex)	
Expected coverage per sample[2]	**96-plex**	**260 K**	**96-plex**	**625 K**
	192-plex	**130 K**	**192-plex**	**313 K**
	384-plex	**65 K**	**384-plex**	**156 K**
	768-plex	33 K	**768-plex**	**78 K**
	1,536-plex	16 K	**1,536-plex**	**39 K**
Kinnex library prep protocol	Procedure & checklist – Preparing Kinnex libraries from 16S rRNA amplicons (103-238-800)			
Metagenomic DNA input amount input into 16S gene amplification	1-2 ng of input gDNA per metagenomic sample			
16S amplicon DNA input into Kinnex library prep workflow	35 ng of purified pooled 16S amplicon DNA			
SMRT Link data analysis workflows	Read Segmentation			
Community data analysis tools	pb-16S-nf			

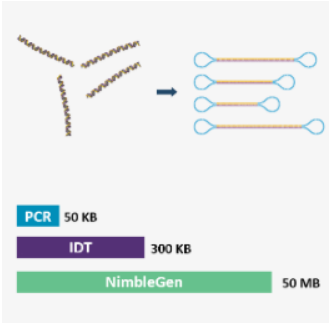
[1] Kinnex concatenation kit (103-071-800) can support up to 1,536-plex sample multiplexing through the combined use of 12 different 16S barcoded Forward PCR primers + 32 different 16S barcoded Reverse PCR primers and 4 different barcoded Kinnex terminal SMRTbell adapters during Kinnex 16s rRNA library construction.

[2] With proper full array formation and adequate sequencing, one SMRT Cell on the Sequel II, IIe, and Revio systems are expected to achieve 20–25 million and 50–60 million 16S sequences, respectively. For most 16S analysis applications, typically aim for ~30-50 K reads/sample.

PacBio

Sequel II/IIe applications – targeted sequencing

PacBio	Application	Targeted sequencing: Amplicons	Targeted sequencing: Target enrichment
Experimental design	Value proposition	Generate sequences of complete long-range amplicons	Detect all classes of variant at scale for genes of interest
	Coverage	50X / locus	50X / locus
	Library insert size	500 bp - 15 kb	3 - 8 kb
	Multiplexing: Sequel II/IIe SMRT Cell	Up to 1,000+ samples	Large 20 Mb panel: 4 samples Medium 2 Mb panel: 24 samples Small 100 kb panel: 96 samples
	Multiplexing: Revio SMRT Cell	Up to 1,000+ samples	Large 20 Mb panel: 12 samples Medium 2 Mb panel: 72 samples Small 100 kb panel: 288 samples

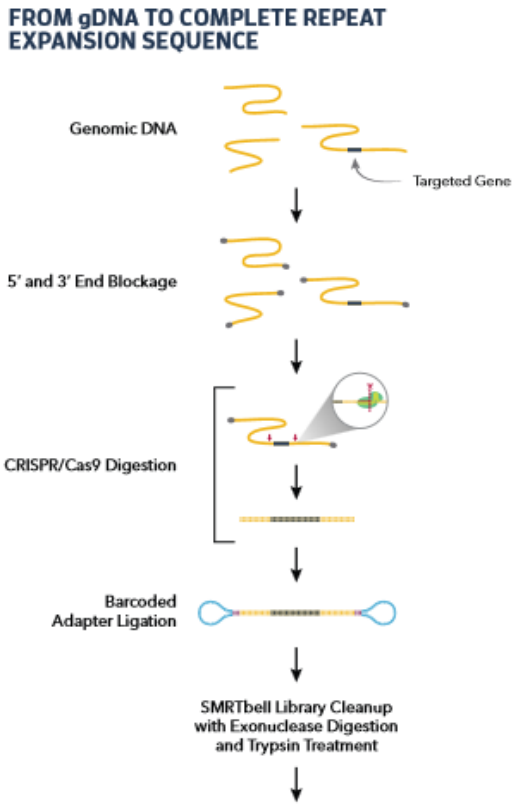


When targeting >50 kb genomic regions – use probe-based capture using DNA oligo hybridization. Protocols available for:

- IDT xGen Lockdown probes
- Nimbelgen SeqCap EZ

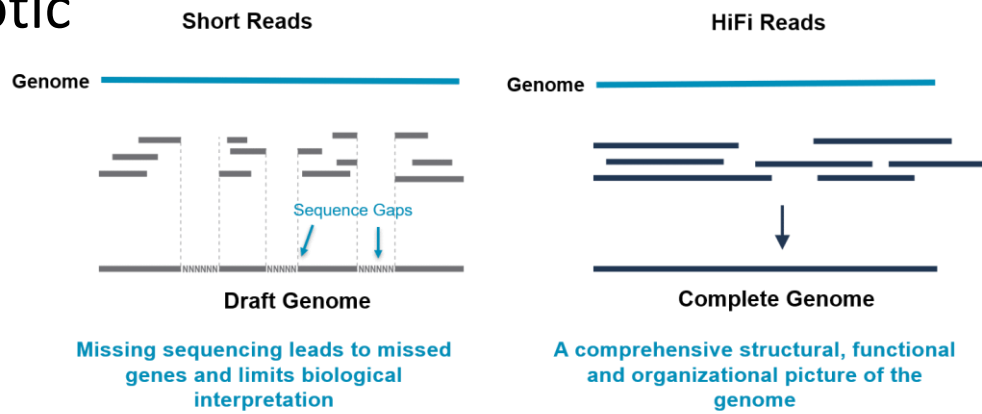
No-amplification targeted sequencing using CRISPR/Cas9 system:

- Challenging regions for PCR amplification (repeat expansions, low complexity regions)
- No PCR bias
- Preserves epigenetic modification signals



PacBio applications at NSC

- *de novo* sequencing of prokaryotic and eukaryotic organisms –
  - Multiplexing up to 96 bacterial samples
  - Mostly HiFi library prep and sequencing for large genomes



- Sequencing of full-length transcriptomes – IsoSeq
  - Multiplexing up to 12 samples for genome annotation

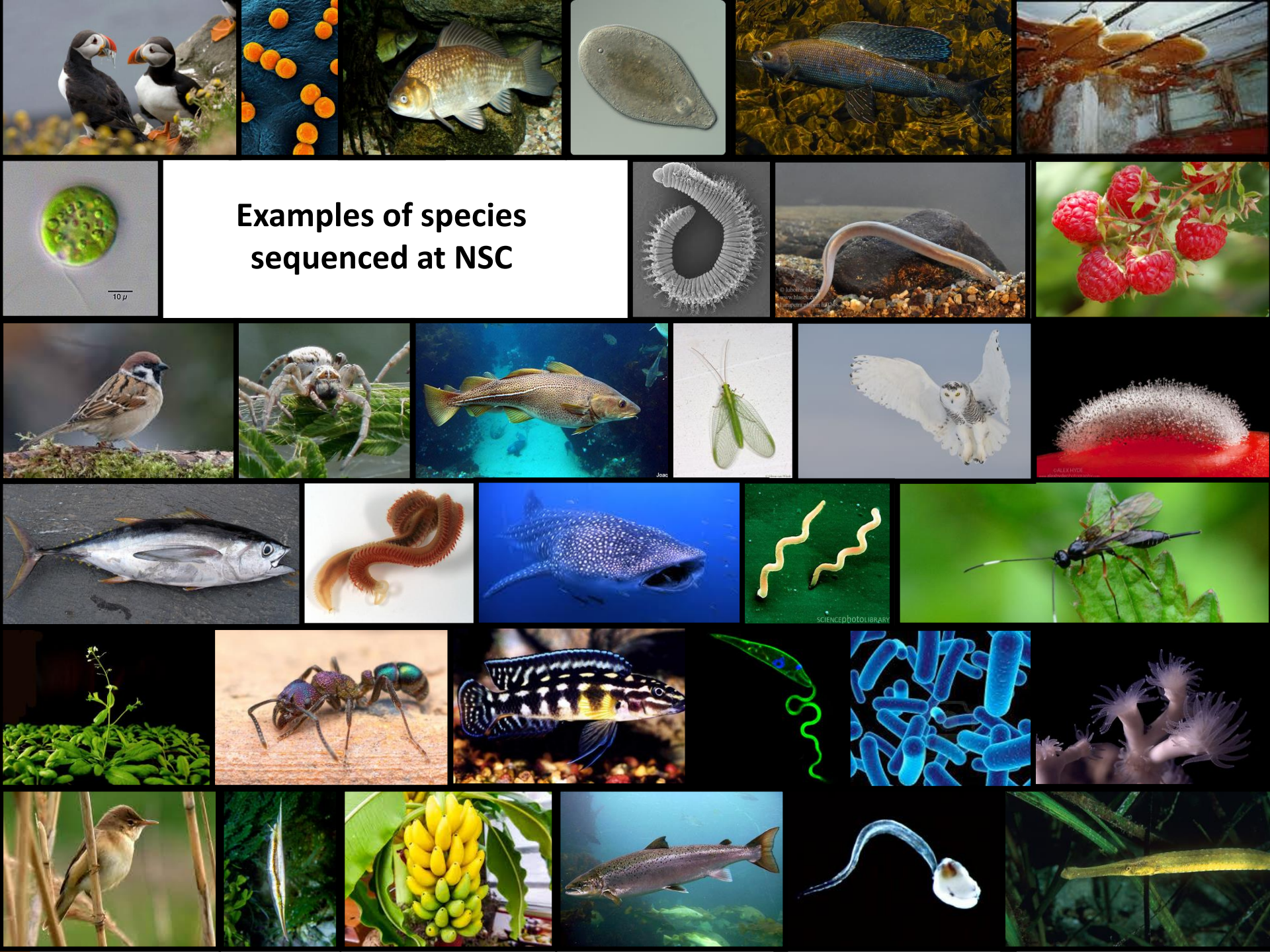
"The way we do RNA-seq now is… you take the transcriptome, you **blow it up into pieces** and then you try to figure out **how they all go back together again**… If you think about it, it's kind of a **crazy way to do things**"

Michael Synder
Professor and Chair of Genetics
Stanford University

Tal Nawy, End to end RNA Sequencing, *Nature Methods*, v10, n10, Dec . 2013, p1144–1145

- Targeted sequencing – amplicons and sequence capture

post@sequencing.uio.no
https://www.sequencing.uio.no/


Examples of species sequenced at NSC

Largest ongoing project: EBP-Nor














EARTH BIOGENOME PROJECT

NORWAY

EBP-Nor

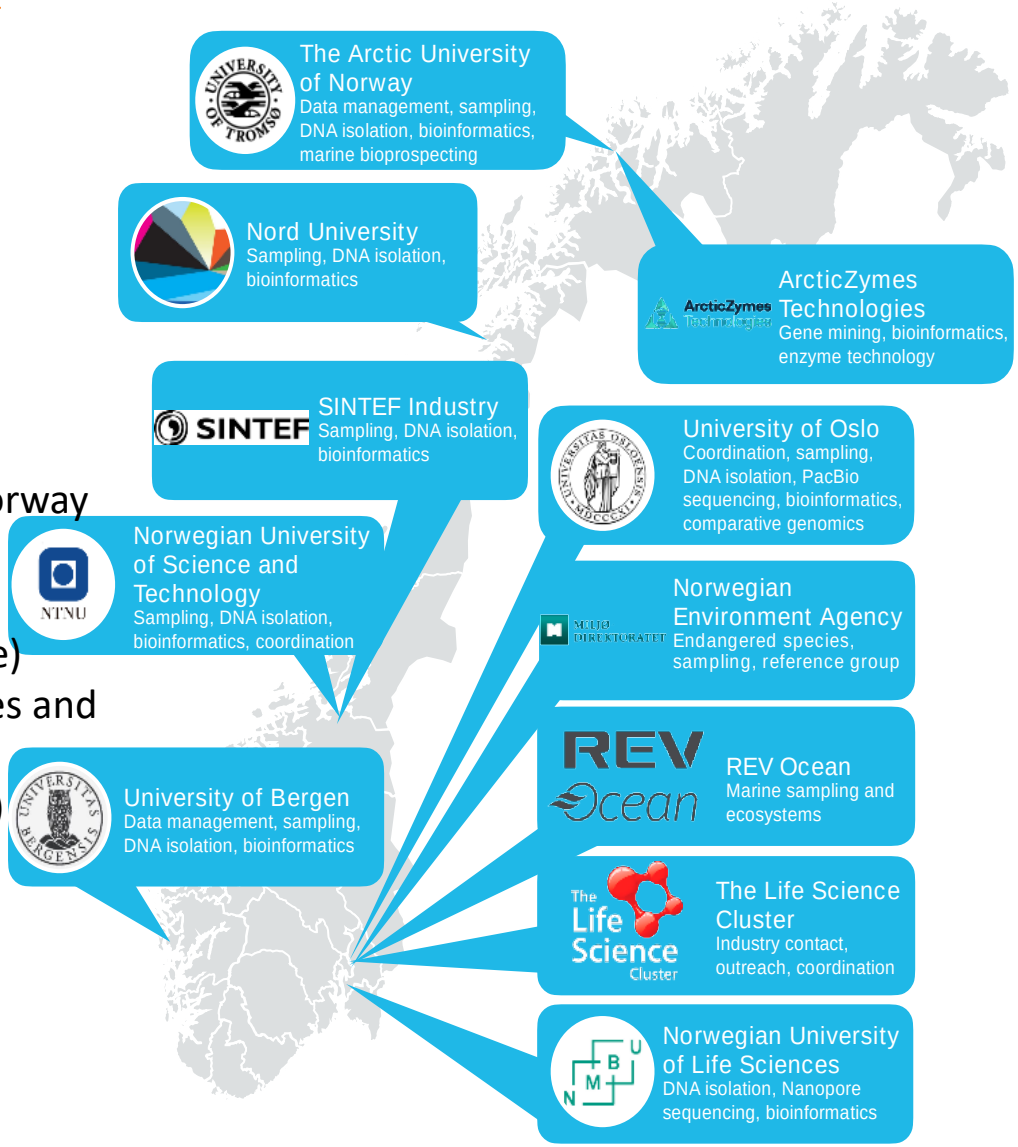
The Norwegian EBP initiative (EBP-Nor)

Planned in 3 phases

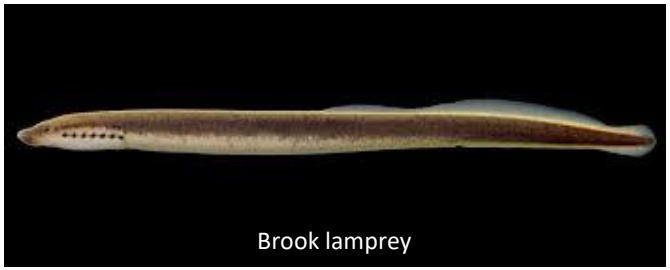
Phase 1 2021-2024 (30 million NOK)

- Funded by the Research Council of Norway
- Planned 150 species
- Norwegian and arctic species
- Marine species (sampling competence)
- Coordination with the Nordic countries and ERGA (and EBP, VGP, DToL etc..)
- Several genomes underway (HiFI, HiC)

Preparation for 2 phase has begun



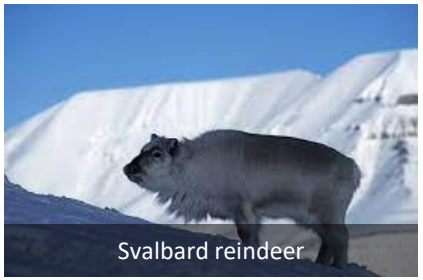
The first EBP-Nor genomes are published


Brook lamprey


River lampreye


25 fungal species


4 moss species


Svalbard reindeer


Snowy owl

https://www.ebpnor.org/

Published Oct. 25, 2024 2:54 PM

From Snowy Owl to Bullfinch: EBP-Nor is mapping the genomes of Norwegian birds

Published Oct. 20, 2023 11:38 AM

Two New Publications Using Complete Reference Genomes and Historical Specimens to Study Charismatic Arctic Organisms

Published Aug. 8, 2023 3:46 PM

An Arctic Icon - The Svalbard reindeer genome is out

In progress: insects (bumble bee, lacewing), Atlantic puffin, cod and salmon (improved), bird cherry (*Prunus padus*), cloudberry ++