millennium science®

PACBIO®

PacBio Virtual Library Prep Workshop

Agenda		
Welcome and Introduction	James Miller, Millennium Science	
HiFi SMRT sequencing current technology summary	James Miller, Millennium Science	15 mins
Library preparation essentials for key applications	Paul Gooding , Millennium Science	60 mins
Having your samples submitted for sequencing at our certified service provider	David Hawkes, AGRF	5 mins

millennium science®

PACBIO®

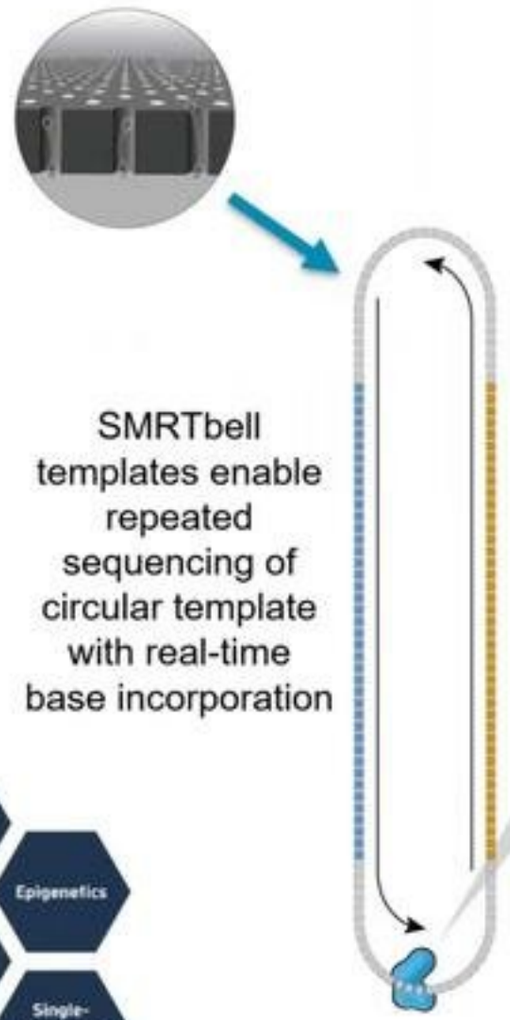
HiFi SMRT sequencing current technology summary

James Miller, Millennium Science

millennium science®

PACBIO®

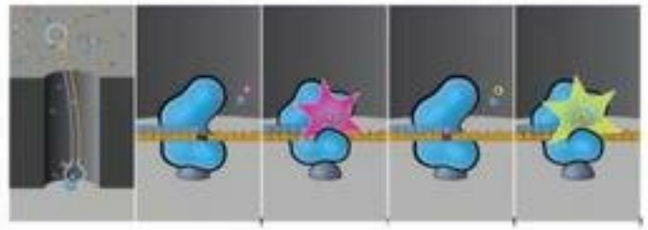
Our Core Technology : SMRT Sequencing



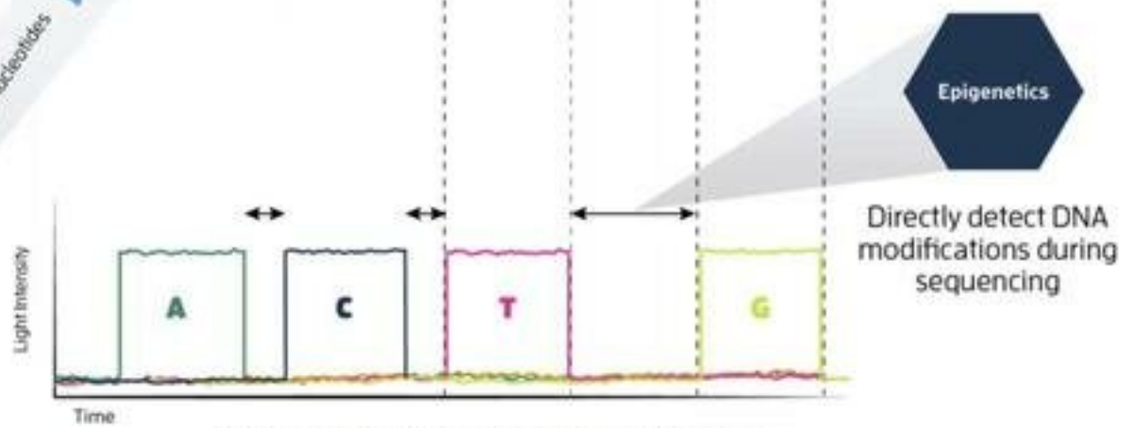
SMRTbell templates enable repeated sequencing of circular template with real-time base incorporation

Single-Molecule Resolution

A single molecule of DNA is immobilized in each ZMW



As anchored polymerases incorporate labeled bases, light is emitted



Nucleotide incorporation kinetics are measured in real time

SMRT Sequencing Advantages

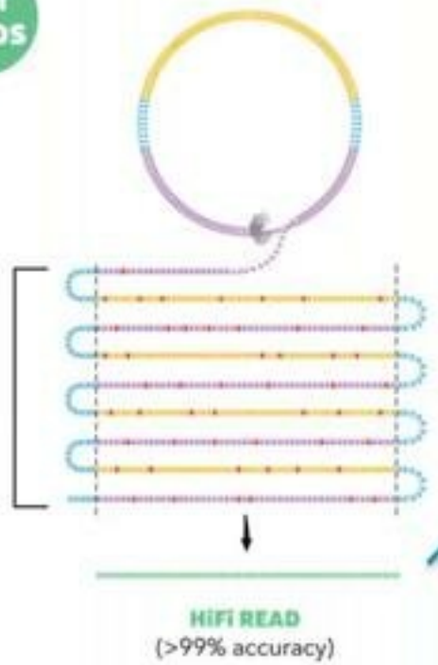
millennium science®

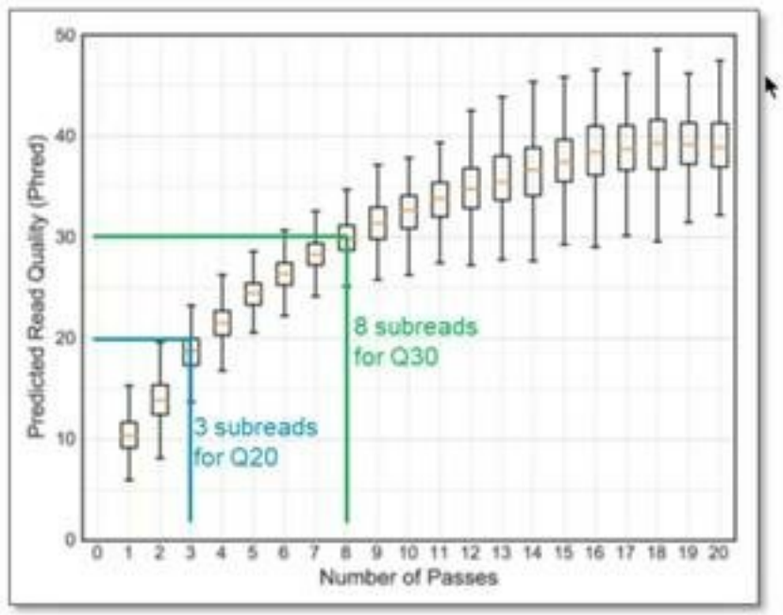
PACBIO®

Introducing highly accurate long reads

- Longer reads allow for multiple passes of the same single native DNA molecule; unbiased coverage results in highly accurate (**HiFi**) long-read sequencing data.
- Significant increase in polymerase read length in the Sequel System 6.0 release increases the number of high fidelity, long reads (>Q20 single-molecule read accuracy) for **inserts up to 10-20 kb**

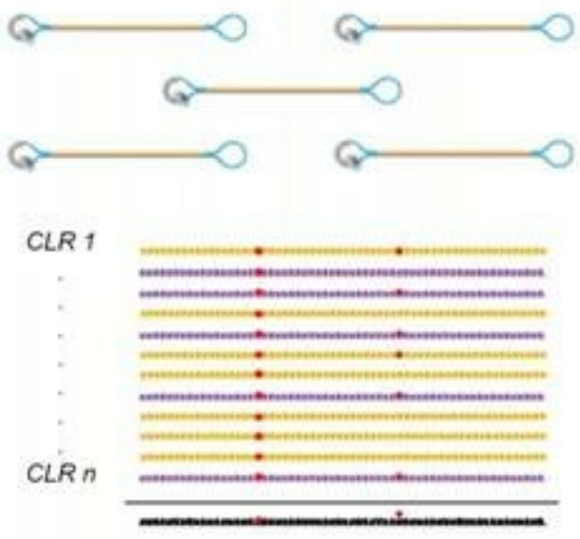
HIFI READS





LONG READS

Continuous Long Reads (CLR)



Consensus sequence

THROUGHPUT EVOLUTION OF PACBIO

>10,000-fold
Increased Throughput

>100-fold
Increased Read Length

Feb 2010
RS Announcement

Aug 2012
v1.3.2 MagBead Release

Feb 2012
v1.3.0
- C2 Launch

Nov 2012
v1.3.3
- XL Chemistry
- Stage Start

Apr 2013
RS II Product Release (v2.0)
- 75K to 150K ZMW
- 2x Throughput
- P4 release

Jan 2013
v1.4
- HGAP/Quiver
- SMRT Cells v3

Oct 2013, v2.1
- P5-C3 release
- HGAP 2.0

Mar 2014
v2.2
- Iso-Seq Method
- Long Amplicon Analysis

Oct 2014
v2.3
- P6-C4 release

Apr 2015
Barcode Kits

Oct 2015
Sequel System

Aug 2016
Sequel v3.1
- 1.2.1 Chemistry

Jan 2017
Sequel v4.0
- 2.0 Chemistry

Sept 2017
Sequel v5.0
- 2.1 Chemistry

Feb 2018
Sequel v5.1
- 2.1 Polymerase

Oct 2018
Sequel v6.0
- 3.0 Chemistry
- SMRT Cell 1M v3

April 2020
Sequel II System

Oct 2020
Sequel II
- 2.0 Chemistry
- v8.0 Software

millennium science®

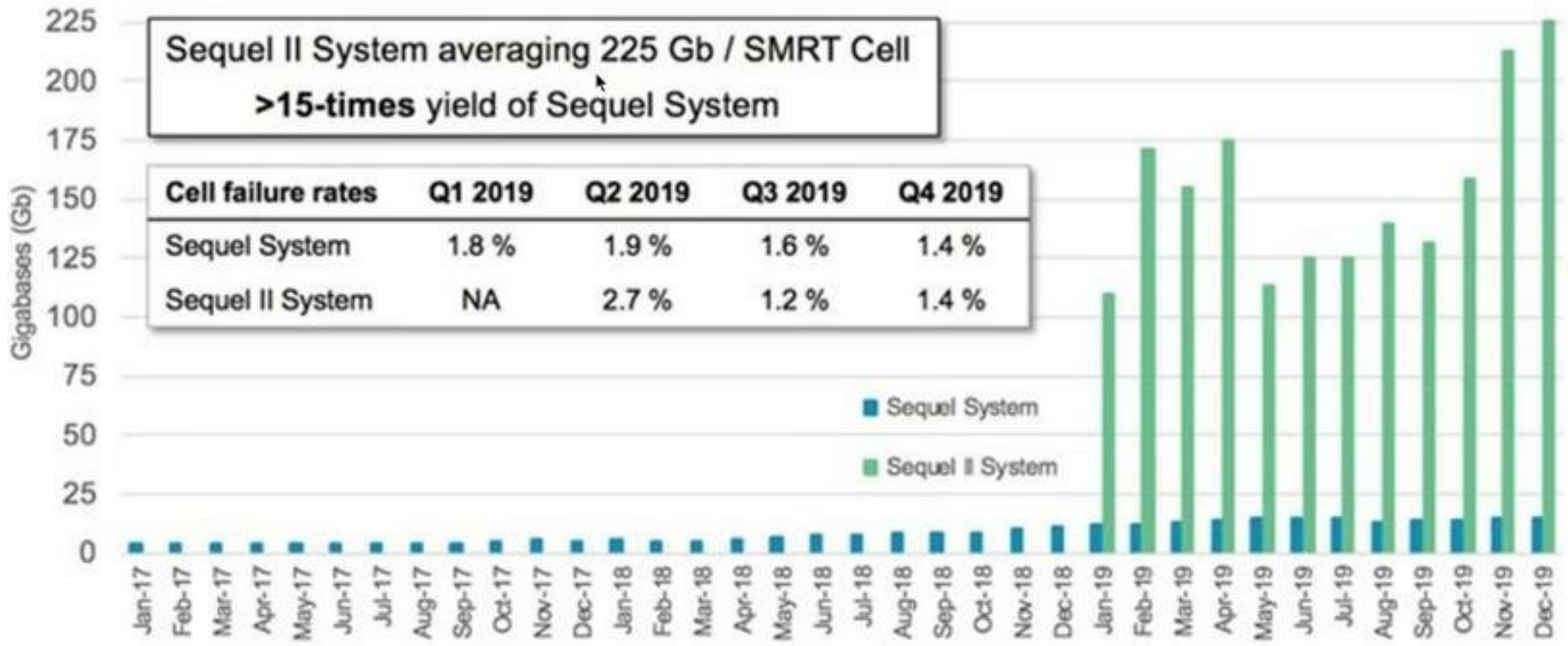
PACBIO®

PERFORMANCE & RELIABILITY

Average Customer Yield per SMRT Cell

Sequel II System averaging 225 Gb / SMRT Cell
>15-times yield of Sequel System

Cell failure rates	Q1 2019	Q2 2019	Q3 2019	Q4 2019
Sequel System	1.8 %	1.9 %	1.6 %	1.4 %
Sequel II System	NA	2.7 %	1.2 %	1.4 %



millennium science®

PACBIO®

SMRT Sequencing Key Applications

WHOLE GENOME SEQUENCING

RNA SEQUENCING

TARGETED SEQUENCING

COMPLEX POPULATIONS

millennium science

PACBIO

WHAT CAN YOU DO WITH ONE SMRT CELL 8M?
WHAT CAN YOU DO WITH ONE SMRT CELL 8M? - WGS

WHOLE GENOME SEQUENCING

SMRT Sequencing Applications	Number of SMRT Cells 8M*
***De Novo* Assembly** Produce reference-quality assemblies for genomes up to 2 Gb	1
Microbial *De Novo* Assembly Generate reference-quality assemblies for up to 48 microbial isolates	1
Variant Detection Call single nucleotide, indel, and structural variants in a ~3 Gb genome (eg Human Genome)	2

millennium science®

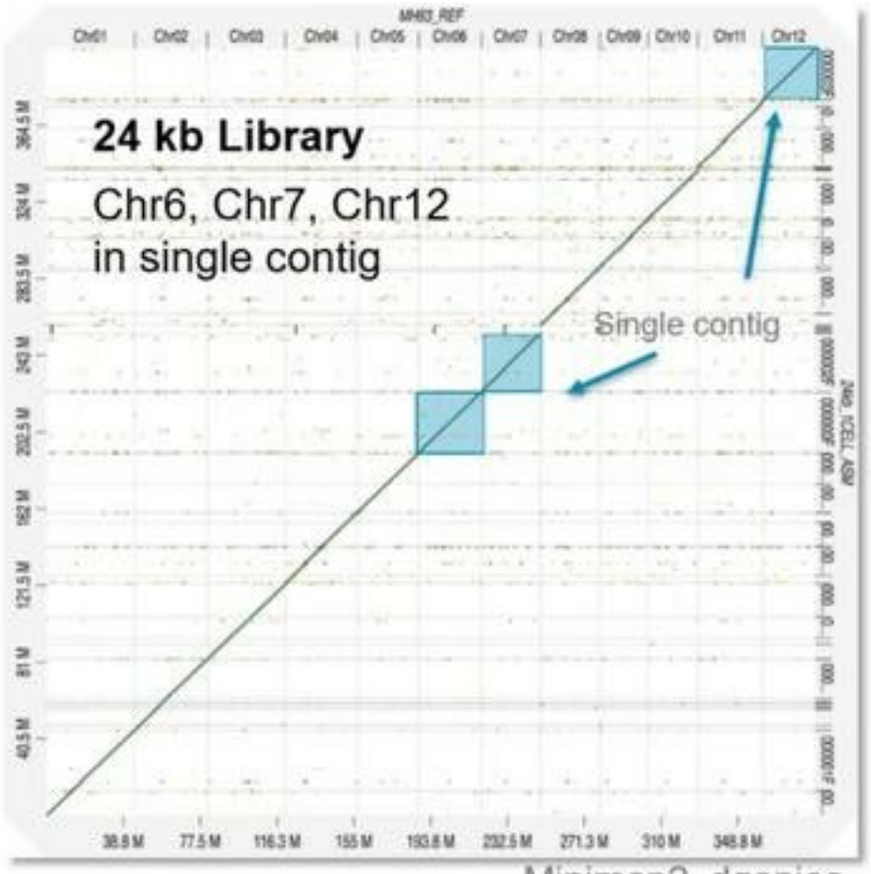
PACBIO®

EXAMPLE: DE NOVO ASSEMBLY WITH HIFI READS

THE 4 Cs of Genome assembly

- ✓ – Contiguity
- ✓ – Completeness
- ✓ – Correctness
- ✓ – Compute

Library Size	17 kb	24 kb
Raw Yield	218 Gb	366 Gb
1 SMRT Cell Yield	15 Gb	25 Gb
Median Read Length	16 kb	21 kb
Coverage (400Mb)	38-fold	63-fold
FALCON Asm Length	403 Mb	405 Mb
N Contigs	209	211
Contig N50	14 Mb	20 Mb
N Chrom in 1 Contig	0	**3**



Minimap2, dgenies

Rice

Data shown above from a 17 kb & 24 kb size-selected rice library using the SMRTbell Template Prep Kit on a Sequel II System (2.0 Chemistry, Sequel II System Software v8.0, 30-hour movie). Falcon assembly was performed post CCS analysis. Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.

millennium science®

PACBIO®

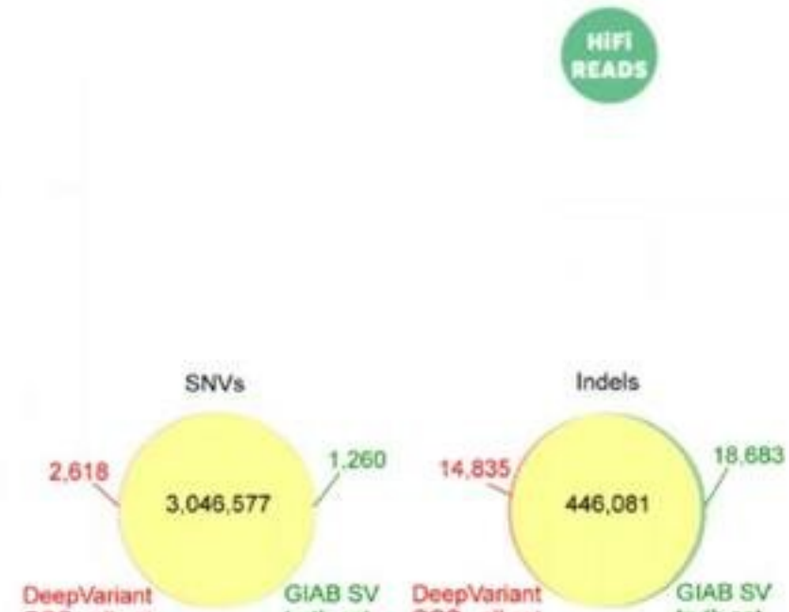
nature biotechnology

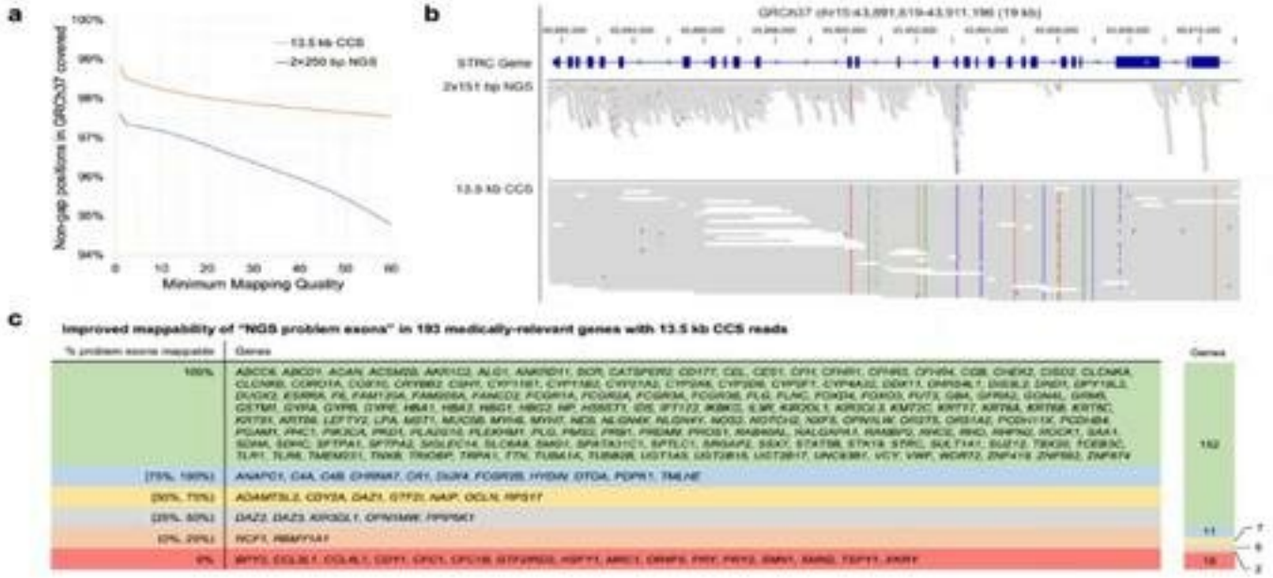
Explore our content ▾ Journal information ▾

nature › nature biotechnology › articles › article

Article | Published: 12 August 2019

Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome





millennium science

PACBIO

Human Genome in a bottle v4 benchmarks (2020)

nature biotechnology

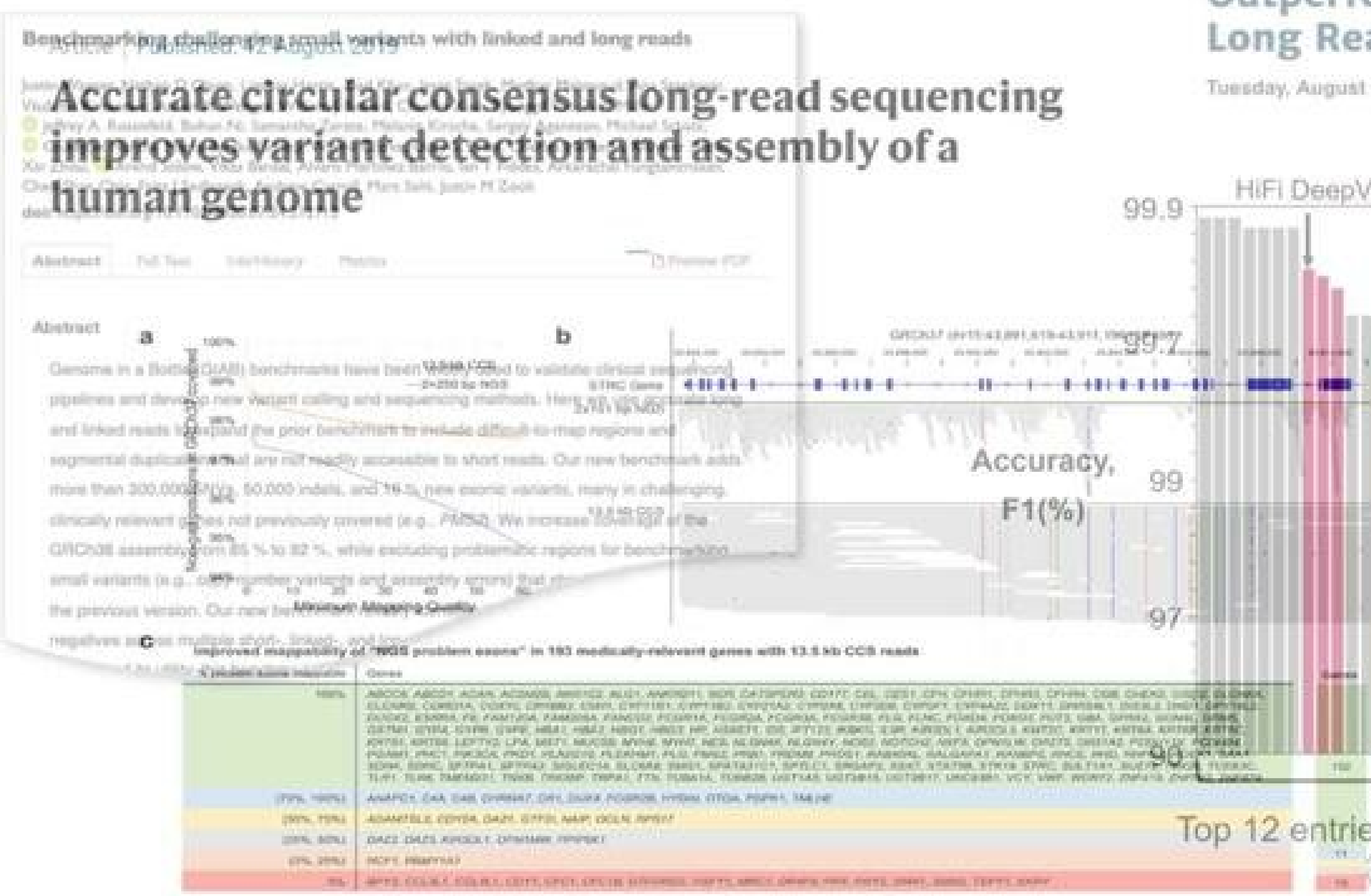
Explore our content ˅ Journal information ˅

nature > nature biotechnology > articles > article

Article | Published: 12 August 2019

Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome

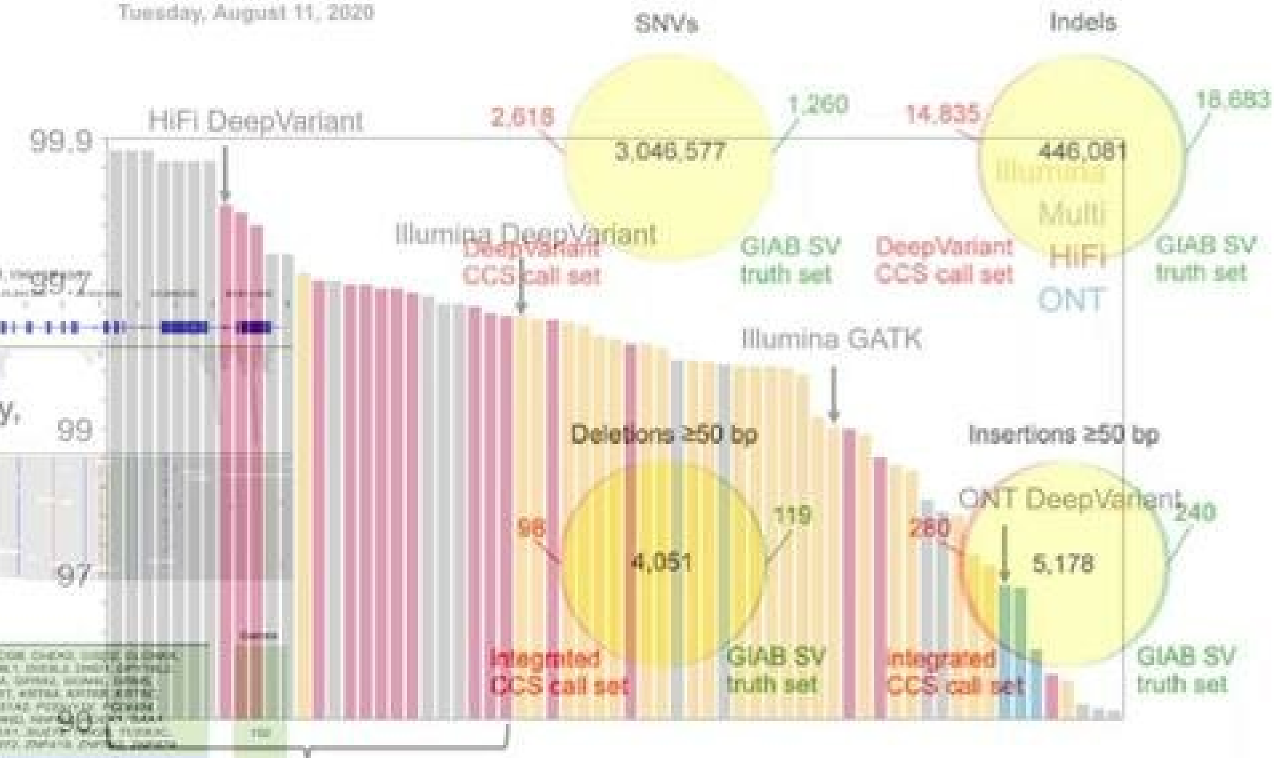
Benchmarking challenging small variants with linked and long reads



HiFi READS

In precisionFDA Challenge, PacBio HiFi Reads Outperform Both Short Reads and Noisy Long Reads

Tuesday, August 11, 2020



Top 12 entries and 25 of top 26 use PacBio HiFi reads

millennium science®

PACBIO®

Human Genome in a bottle v4 benchmarks (2020)

Benchmarking challenging small variants with linked and long reads

Justin Wagner, Nathan D Olson, Lindsay Harris, Ziad Khan, Jesse Farek, Medhat Mahmoud, Ana Stankovic, Vladimir Kovacevic, Aaron M Wenger, William J Rowell, Chunlin Xiao, Byunggil Yoo, Neil Miller, Jeffrey A. Rosenfeld, Bohan Ni, Samantha Zarate, Melanie Kirsche, Sergey Aganezov, Michael Schatz, Giuseppe Narzisi, Marta Byrska-Bishop, Wayne Clarke, Uday S. Evani, Charles Markello, Kishwar Shafin, Xin Zhou, Arend Sidow, Vikas Bansal, Alvaro Martinez Barrio, Ian T Fiddes, Arkarachai Fungtammasan, Chen-Shan Chin, Fritz J Sedlazeck, Andrew Carroll, Marc Salit, Justin M Zook

doi: https://doi.org/10.1101/2020.07.24.212712

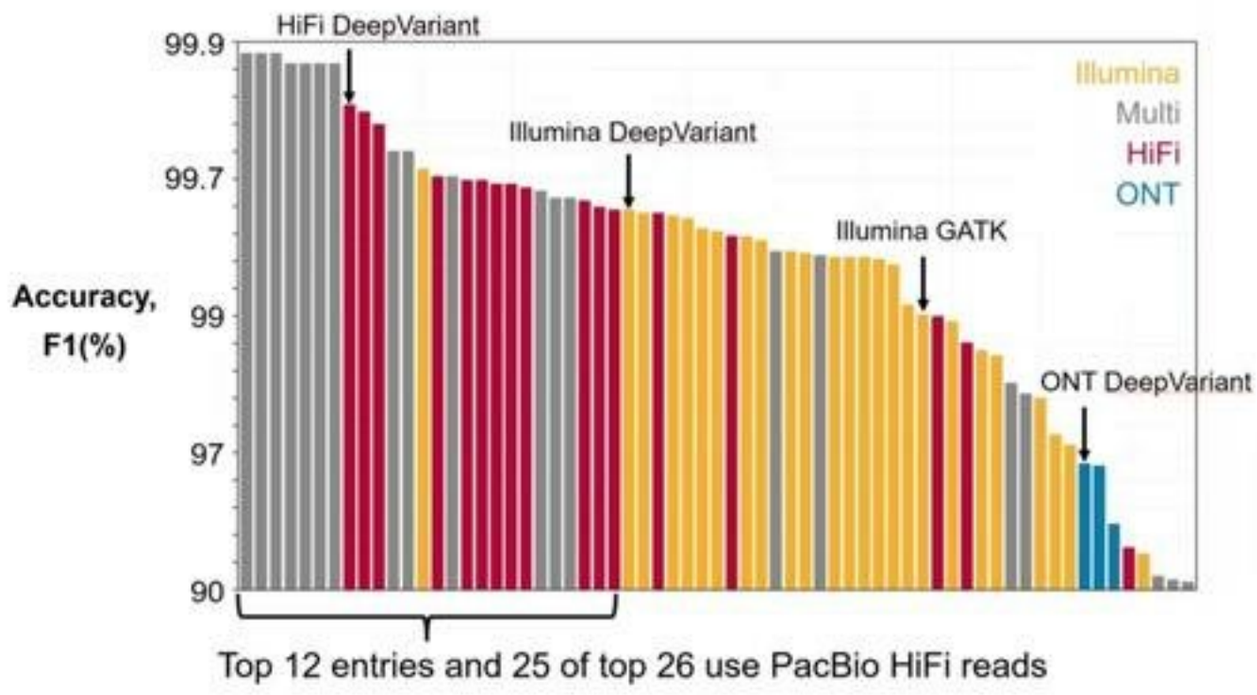
Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

Genome in a Bottle (GIAB) benchmarks have been widely used to validate clinical sequencing pipelines and develop new variant calling and sequencing methods. Here we use accurate long and linked reads to expand the prior benchmark to include difficult-to-map regions and segmental duplications that are not readily accessible to short reads. Our new benchmark adds more than 300,000 SNVs, 50,000 indels, and 16 % new exonic variants, many in challenging, clinically relevant genes not previously covered (e.g., *PMS2*). We increase coverage of the GRCh38 assembly from 85 % to 92 %, while excluding problematic regions for benchmarking small variants (e.g., copy number variants and assembly errors) that sh... the previous version. Our new benchmark reliably identifies... negatives across multiple short-, linked-, and lo... the utility this benchmark...

In precisionFDA Challenge, PacBio HiFi Reads Outperform Both Short Reads and Noisy Long Reads

Tuesday, August 11, 2020



millennium science®

PACBIO®

WHAT CAN YOU DO WITH ONE SMRT CELL 8M? - TRANSCRIPTOME

RNA SEQUENCING

SMRT Sequencing Applications	Number of SMRT Cells 8M*
Whole Transcriptome Characterize alternative splicing with full-length transcripts	1
Genome Annotation Sequence full-length transcripts and multiplex up to 8 tissues	1

millennium science®

PACBIO®

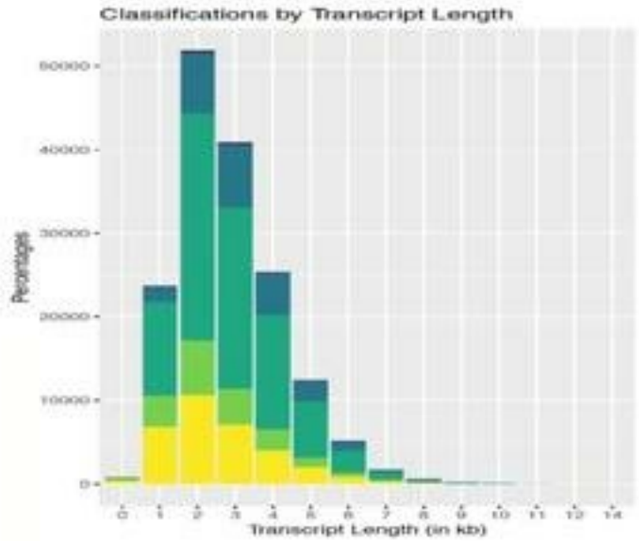
HiFi READS

ISO-SEQ PERFORMANCE

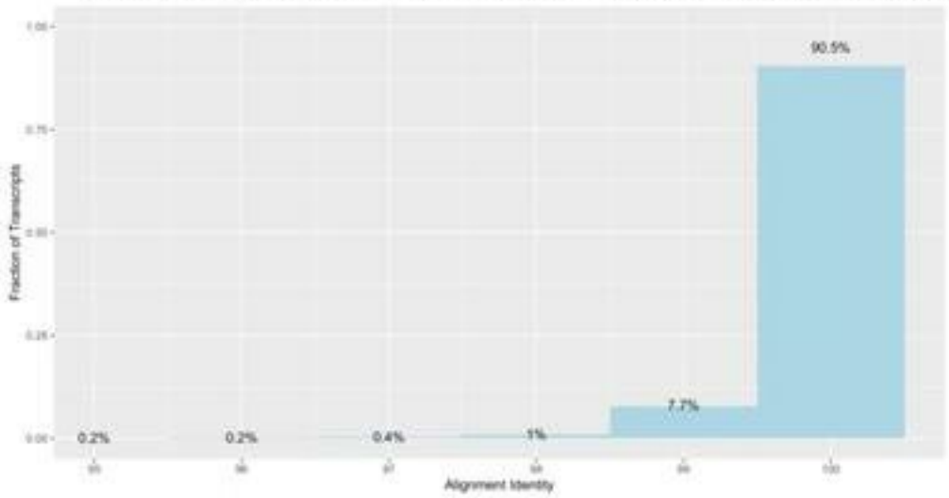
- **Comprehensive**
- **Full-Length**
- **Highly accurate**

162,290 transcripts

80 – 14,288 bp
(mean: 3.3 kb)



> 99% of Iso-Seq output transcript is >99% accurate



Category	Count	CAGE peak within 50 bp	polyA Motif Detected
FSM	32,649	70%	72%
ISM	19,011	37%	62%
NIC	84,610	36%	55%
NNC	25,323	57%	72%
Antisense	321	24%	43%
Intergenic	376	24%	38%

Dataset: Alzheimer brain on 1 SMRT Cell 8M

millennium science®

PACBIO®

WHAT CAN YOU DO WITH ONE SMRT CELL 8M? - MICROBIOME

COMPLEX POPULATIONS

SMRT Sequencing Applications	Number of SMRT Cells 8M*
Full-length 16S Gain strain-level resolution by multiplexing up to 96 samples	1
Metagenomic Functional Profiling Examine up to 3 low-complexity samples with multiplexing	1
Shotgun Metagenomic Assembly Generate near-complete assemblies of high-complexity samples (e.g. gut microbiome)	1

millennium science®

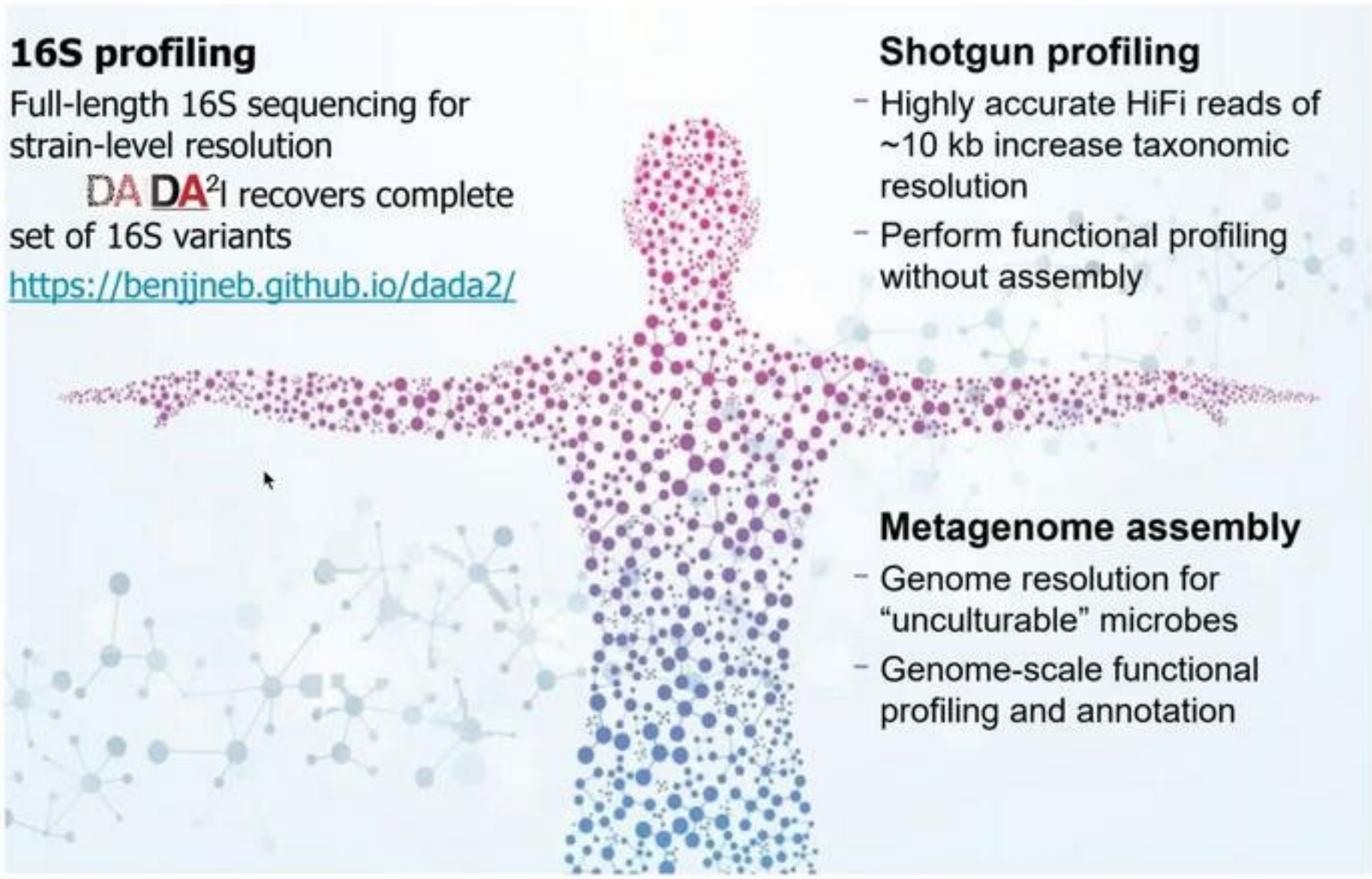
PACBIO®

Metagenomics approaches using PacBio HiFi data

16S profiling

- Full-length 16S sequencing for strain-level resolution
- DADA2 recovers complete set of 16S variants

https://benjjneb.github.io/dada2/



Shotgun profiling

- Highly accurate HiFi reads of ~10 kb increase taxonomic resolution
- Perform functional profiling without assembly

Metagenome assembly

- Genome resolution for “unculturable” microbes
- Genome-scale functional profiling and annotation

millennium science®

PACBIO®

WHAT CAN YOU DO WITH ONE SMRT CELL 8M? – TARGETED SEQUENCING

TARGETED SEQUENCING

SMRT Sequencing Applications	Number of SMRT Cells 8M*
Amplicon Sequencing Detect variation in specific regions by multiplexing	1
No-Amp Sequencing Enrich hard-to-amplify targets and multiplex up to 10 samples	1

millennium science®

PACBIO®

ONE KIT FOR ALL APPLICATIONS

- Single-tube
- Additive workflow
- ~4 hrs library construction time

SMRTbell Express Template Prep Kit 2.0

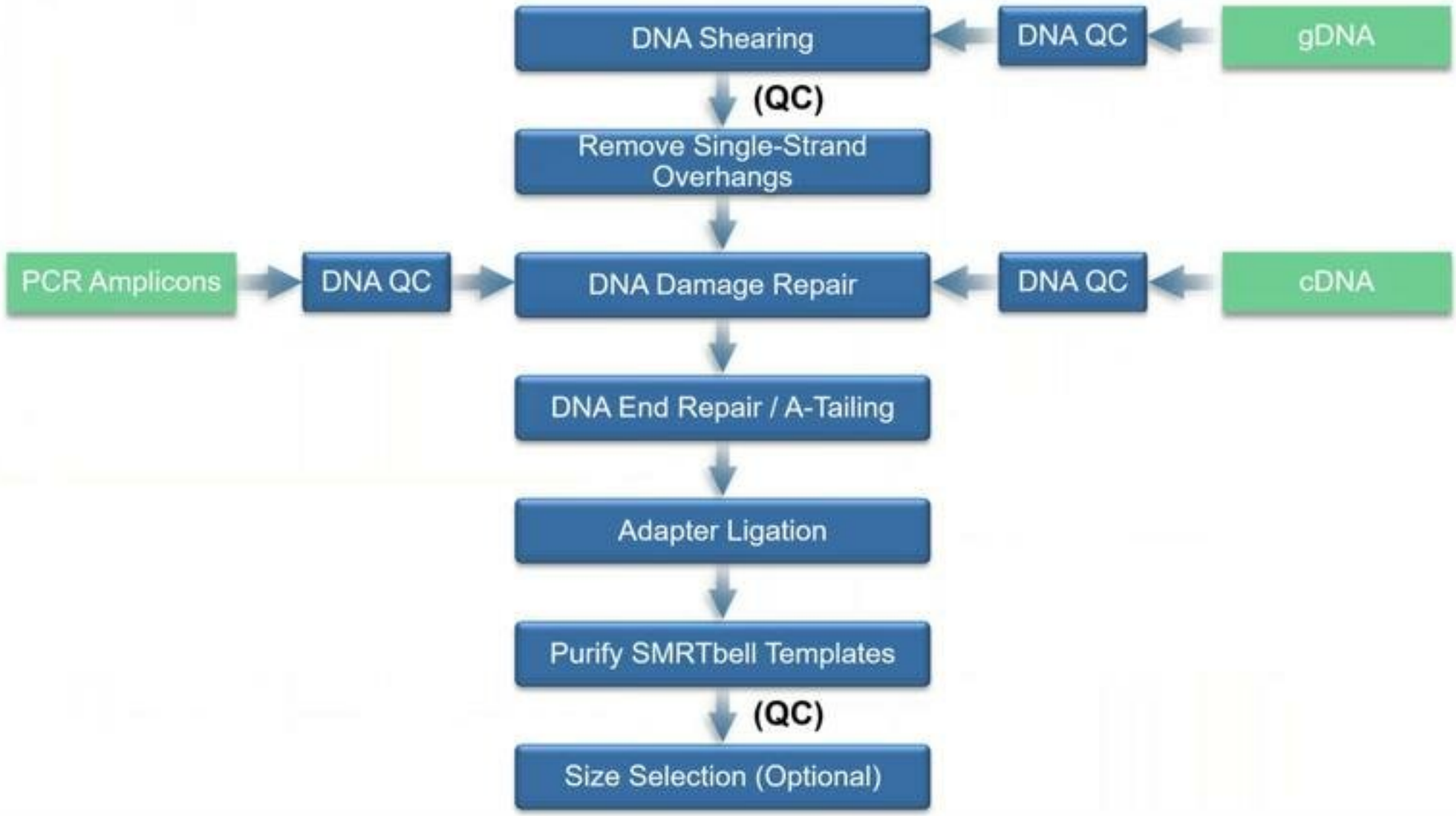
- **Compatibility with ALL libraries.**
- **Improved formulation** to accommodate all library insert sizes from 500 bp insert to >30 kb size-selected libraries.
- **Streamlined, consolidated protocols** for improved ease of use for library construction in ~ 4 hours.
- **Lower input requirements** while minimizing handling induced DNA damage
- **Updated multiplexing solutions** with Express workflow for microbial genomes.
- **Amplicon and Iso-Seq compatible**

One kit supports 18 large-insert libraries, 48 microbial genomes, and 96 amplicon template preparations.

millennium science®

PACBIO®

SMRTbell Express library construction workflow Summary and Sample input Types



millennium science®

PACBIO®

Supporting Reagent Kits

cDNA: Iso-Seq Express for Faster RNA Sequencing

- RT-PCR kit + new Iso-Seq sample prep protocol

WGS: SMRTbell gDNA Sample Amplification Kit for low input DNA

- DNA sample amplification in the Ultra-Low DNA input procedure (5-20ng)

millennium science®

PACBIO®

AutoSave Off | Library Pr... | Search | Paul Gooding PG

File Home Insert Design Transitions Animations Slide Show Review View Help | Share | Comments

Paste | New Slide | Reuse Slides | Layout | Reset | Section | Clipboard | Slides | Font | Paragraph | Shapes | Arrange | Quick Styles | Shape Fill | Shape Outline | Shape Effects | Drawing | Find | Replace | Select | Editing | Dictate | Voice | Design Ideas | Designer






PACBIO

millennium science

Paul Gooding PhD
Field Applications Scientist
Australia and New Zealand
pgooding@mscience.com.au

Virtual Library Preparation Workshop

24/09/2020

Click to add notes

Slide 1 of 90 English (Australia) | Notes | 57%

Type here to search | ENG | 10:23 AM 24/09/2020

Paul Gooding PhD

Field Applications Scientist

Australia and New Zealand

pgooding@mscience.com.au

Virtual Library Preparation Workshop

24/09/2020

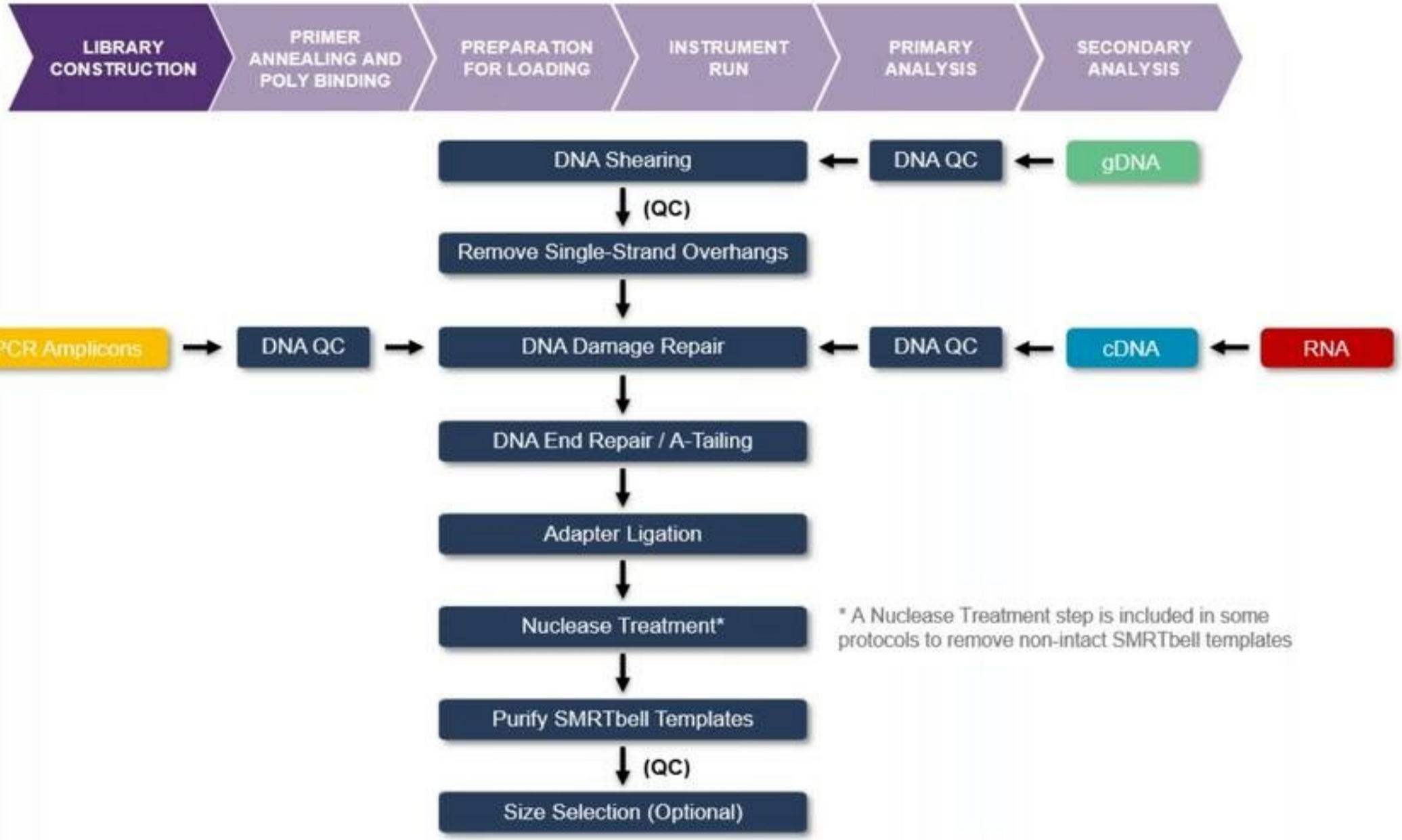
PACBIO SEQENCING APPLICATIONS

There are many similarities between Library Preparations for various applications, but differences in starting material and subtle variations in workflow

1. Library Preparation Workflow
   - Input DNA, QC, sizing, shearing
   - SMRTbell Express TPK 2.0
2. Whole Genome Sequencing (WGS)
   - de novo
   - variant calling
3. Iso-Seq (Long-read RNA Sequencing)
4. Amplicons
5. Metagenomics

SMRTBELL EXPRESS LIBRARY CONSTRUCTION WORKFLOW SUMMARY AND SAMPLE INPUT TYPES

PACBIO



* A Nuclease Treatment step is included in some protocols to remove non-intact SMRTbell templates

PACBIO®

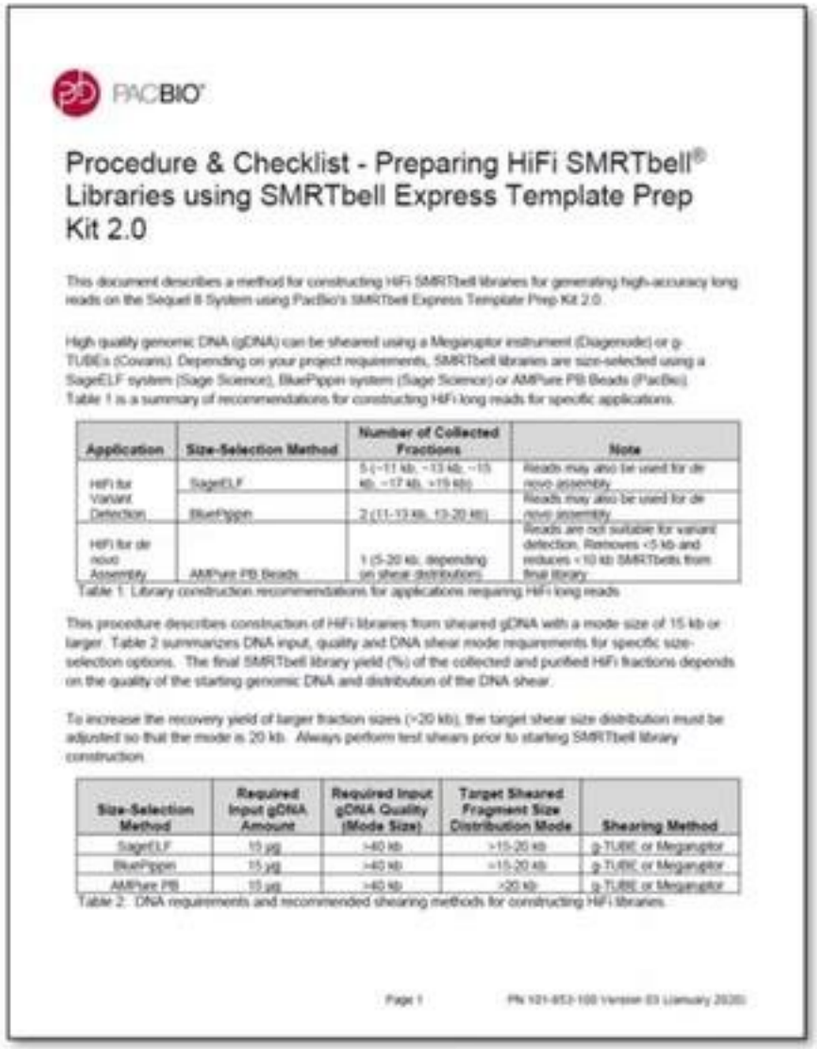
millennium science

SMRTbell Express Library Construction Workflow

PREPARING HIFI SMRTBELL LIBRARIES USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0

- Document (PN 101-853-100) describes a method for constructing SMRTbell libraries ~11 – 20 kb (or larger) that are suitable for generating high-accuracy long reads on the Sequel II System using SMRTbell Express TPK 2.0 for *de novo* assembly and variant detection applications
- Protocol document contains:

1. Recommendations for gDNA QC and quantification
2. Recommendations for shearing gDNA to the desired target mode size using either the Megaruptor System (Diagenode) or g-Tubes (Sage Science)
3. Enzymatic steps for preparation of a HiFi SMRTbell library using SMRTbell Express TPK 2.0
4. Instructions for size-selection of the HiFi SMRTbell library using either the SageELF System (Sage Science) or BluePippin System (Sage Science), and also includes protocol reference for performing AMPure BP Size Selection method for *de novo* assembly applications using HiFi reads
5. Sample setup guidance for preparing HiFi libraries for sequencing on the Sequel II System



PACBIO

Procedure & Checklist - Preparing HiFi SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0

This document describes a method for constructing HiFi SMRTbell libraries for generating high-accuracy long reads on the Sequel II System using PacBio's SMRTbell Express Template Prep Kit 2.0

High quality genomic DNA (gDNA) can be sheared using a Megaruptor instrument (Diagenode) or g-TUBEs (Covaris). Depending on your project requirements, SMRTbell libraries are size-selected using a SageELF system (Sage Science), BluePippin system (Sage Science) or AMPure PB Beads (PacBio). Table 1 is a summary of recommendations for constructing HiFi long reads for specific applications.

Application	Size-Selection Method	Number of Collected Fractions	Note
HiFi for Variant Detection	SageELF	5 (~11 kb, ~13 kb, ~15 kb, ~17 kb, >19 kb)	Reads may also be used for de novo assembly
	BluePippin	2 (11-13 kb, 13-20 kb)	Reads may also be used for de novo assembly
HiFi for de novo Assembly	AMPure PB Beads	1 (5-20 kb, depending on shear distribution)	Reads are not suitable for variant detection. Removes <5 kb and reduces <10 kb SMRTbells from final library

Table 1: Library construction recommendations for applications requiring HiFi long reads

This procedure describes construction of HiFi libraries from sheared gDNA with a mode size of 15 kb or larger. Table 2 summarizes DNA input, quality and DNA shear mode requirements for specific size-selection options. The final SMRTbell library yield (%) of the collected and purified HiFi fractions depends on the quality of the starting genomic DNA and distribution of the DNA shear.

To increase the recovery yield of larger fraction sizes (>20 kb), the target shear size distribution must be adjusted so that the mode is 20 kb. Always perform test shears prior to starting SMRTbell library construction.

Size-Selection Method	Required Input gDNA Amount	Required Input gDNA Quality (Mode Size)	Target Sheared Fragment Size Distribution Mode	Shearing Method
SageELF	15 µg	>40 kb	>15-20 kb	g-TUBE or Megaruptor
BluePippin	15 µg	>40 kb	>15-20 kb	g-TUBE or Megaruptor
AMPure PB	15 µg	>40 kb	>20 kb	g-TUBE or Megaruptor

Table 2: DNA requirements and recommended shearing methods for constructing HiFi libraries

Page 1 PN 101-853-100 Version 03 (January 2020)

https://www.pacb.com/support/documentation/

METHODS FOR EVALUATION OF GENOMIC DNA QUALITY

PACBIO

Starting with high-quality genomic DNA will result in longer libraries and better *de novo* assembly performance

- Input genomic DNA must be carefully QC'ed to assess integrity
- PFGE/FIGE or Femto Pulse sizing tool is highly recommended
- High-quality, high-molecular weight DNA → longer read lengths
- Low-quality, degraded/damaged DNA → shorter read lengths, lower library synthesis yields (dependent on BluePippin size selection parameters)
- DNA purity can be determined by using a NanoDrop instrument or other spectrophotometer device
- PacBio highly recommends using the Qubit High Sensitivity fluorometric assay for accurate dsDNA quantitation

A. Genomic DNA Sizing Characterization

DNA QC

PACBIO

Recommended methods for determining gDNA size distribution:

Starting with high-quality, high molecular weight (HMW) genomic DNA will result in longer libraries and better *de novo* assembly performance

PippinPulse System (Sage Science)

Resolves up to ~80 kb
Requires ≥50 ng of sample
~16-hour analysis time

http://www.sagescience.com/products/pippin-pulse/

Femto Pulse System (Agilent Technologies)

Highly Recommended

Resolves up to ~165 kb
Requires <1 ng of sample
<1.5-hour analysis time

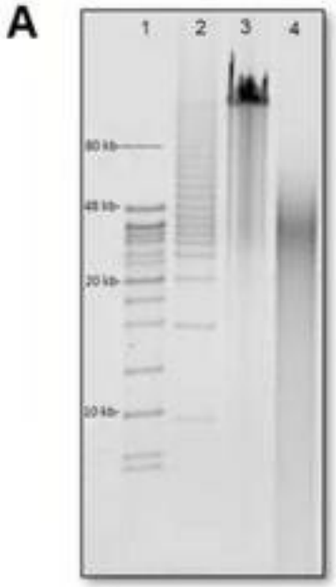
https://www.agilent.com/en/product/automated-electrophoresis/femto-pulse-systems

CHEF Mapper XA System (Bio-Rad)

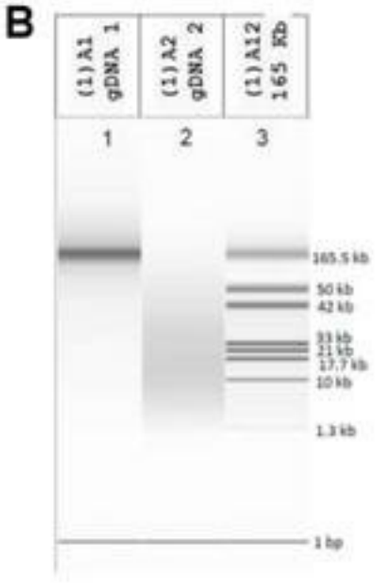
Resolves up to ~10 Mb
Requires ≥100 ng of sample
~16-h analysis time

http://www.bio-rad.com/en-us/category/pulsed-field-gel-electrophoresis-systems

Technical Note TN101-040518: Preparing DNA for PacBio Whole Genome Sequencing for *de novo* Assembly: Quality Control and Shearing L



Lane 1: 8-48 kb Ladder (Bio-Rad)
Lane 2: 5 kb Ladder (Bio-Rad)
Lane 3: HMW gDNA
Lane 4: Degraded gDNA



Lane 1: High MW gDNA
Lane 2: Degraded gDNA
Lane 3: 165 kb Ladder

Evaluation of gDNA quality using A) Bio-Rad CHEF Mapper System and B) Femto Pulse System. Lanes A3 and B1 are examples of high quality, high-molecular weight genomic DNA. Lanes A4 and B2 are examples of degraded gDNA.

B. DNA Purity Determination

- DNA purity can be determined by using a **NanoDrop®** instrument or other spectrophotometers.
- For ultrapure gDNA, A260/280 ratio is typically between ~1.8 - 2.0 and A260/230 ratio is ≥2.0.
- If A260/280 and A260/230 readings are out of the range specified above, PacBio recommends performing an AMPure® purification step followed by re-assessment of quantity and purity of the gDNA sample.

260/280 Ratio

- A low A260/A280 ratio may indicate the presence of protein, phenol, or other contaminants that absorb strongly at or near 280 nm. Sometimes it may be caused by a very low concentration of nucleic acid.
- High 260/280 ratios are not indicative of an issue.

260/230 Ratio

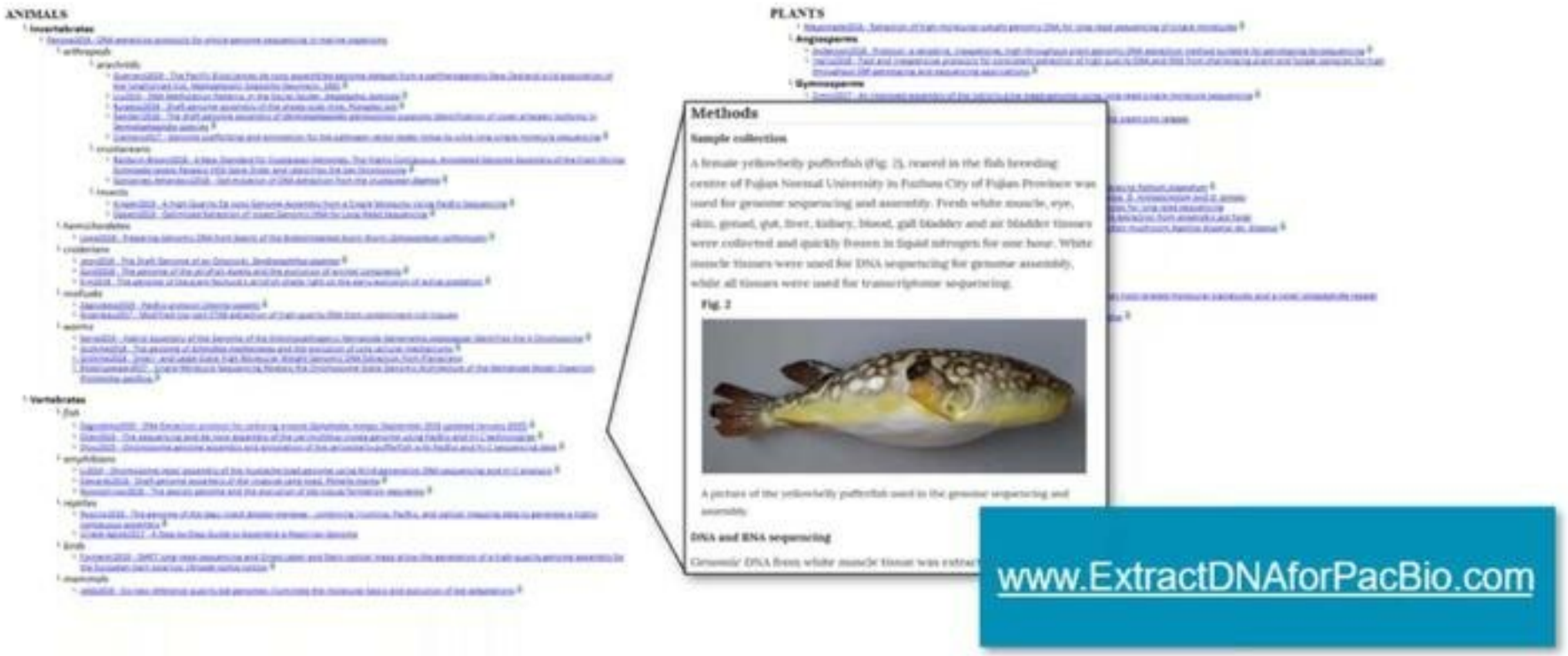
- A low A260/A230 ratio may be the result of:
  - Carbohydrate carryover (often a problem with plants).
  - Residual phenol from nucleic acid extraction.
  - Residual guanidine (often used in column-based kits).
  - Glycogen used for precipitation.
- A high A260/A230 ratio may be the result of:
  - Making a blank measurement on a dirty pedestal of a Nanodrop instrument.
  - Using an inappropriate solution for the blank measurement.

C. DNA Quantification

- Accurate quantitation of DNA concentration is critical for PacBio template preparation procedures.
  - Specifically, it is critical to determine the concentration of the double-stranded DNA, since only double-stranded DNA will be converted into sequencing templates.
- PacBio highly recommends using a **Qubit fluorometer** tool and **Qubit dsDNA High Sensitivity (HS) Assay Kit** for routine DNA quantitation during SMRTbell library construction.
- When assessing gDNA QC, PacBio recommends using both fluorometric and spectrophotometric methods – for example, using both the Qubit and NanoDrop instruments
  - If the sample is pure gDNA, free of any RNA contaminants and other small molecules, the two methods should converge to similar DNA concentration measurement values
- If the measured NanoDrop concentration is significantly different (>50%) from the Qubit measurement, PacBio recommends doing an AMPure PB Bead purification step (as specified by your chosen library preparation protocol), followed by a re-measurement with both methods. Typically, a single AMPure PB Bead purification step resolves the discrepancy.
  - If the concentration measurement discrepancy after one AMPure PB Bead purification step is not reduced, we recommend trying another cleanup approach before a re-measurement with both methods.

NEW SAMPLE PREPARATION ONLINE RESOURCE

- Literature resource for sample collection and DNA extraction protocol references



PacBio does not assume responsibilities/guarantees for these external publications/protocols, but we are happy to help as best as we can to guide/connect.
Please contact ExtractDNA@pacb.com for more discussions around your particular species & sequencing project!

METHODS FOR DNA SHEARING

Input gDNA must be sheared carefully so that the average size of fragmented DNA remains well above the desired size selection cut-off

Recommended methods for shearing genomic DNA to the desired target fragment size:

Covaris g-TUBE Devices (6 kb – 20 kb)

- Single-use device that uses centrifugal force to pass solutions through an engineered ruby orifice to shear genomic DNA into selected fragment sizes ranging from ~6 kb to 20 kb in <5 minutes
- The only equipment needed is a compatible bench-top centrifuge; the rate of centrifugation determines the degree of DNA fragmentation

https://covaris.com/products/g-tube/

Needle Shearing with 26 G Needles (>30 kb)

Highly Recommended

- This low-tech method relies on shearing forces created by passing genomic DNA through a small-gauge needle.
- Can be a useful low-cost option for generation of fragments tens of kb in length; however, variable shearing performance may be observed for different DNA sample types

https://www.sai-infusion.com/products/blunt-needles

Diagenode Megaruptor® System (2 – 75 kb)

- A syringe pump generates hydrodynamic shear forces within a disposable tube; the size of the constriction and the flow rate of the liquid determine the DNA fragment size.
- This method is useful for generation of fragments ~2 to 75 kb in a highly reproducible manner across different DNA samples types

https://www.diagenode.com/en/categories/megaruptor

DNA Fragmentation Using the Diagenode Megaruptor System

PACBIO

- Large gDNA must be sheared carefully
- Test Shears help determine the best shearing parameters
- Evaluation with PFGE or Agilent Femto Pulse is recommended
- The average library size depends on the fragment distribution post shearing

A

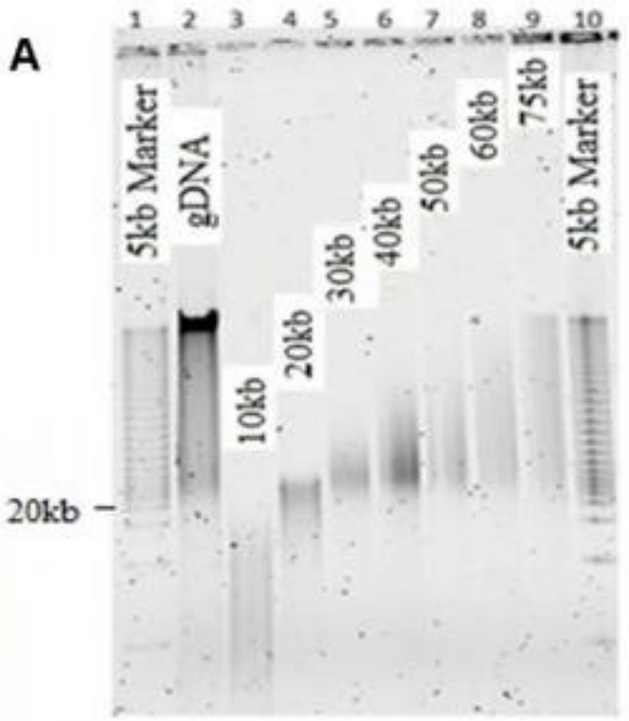


Figure A. PFGE gel image of gDNA sheared to various target fragment sizes using a Diagenode Megaruptor 1 System.

B

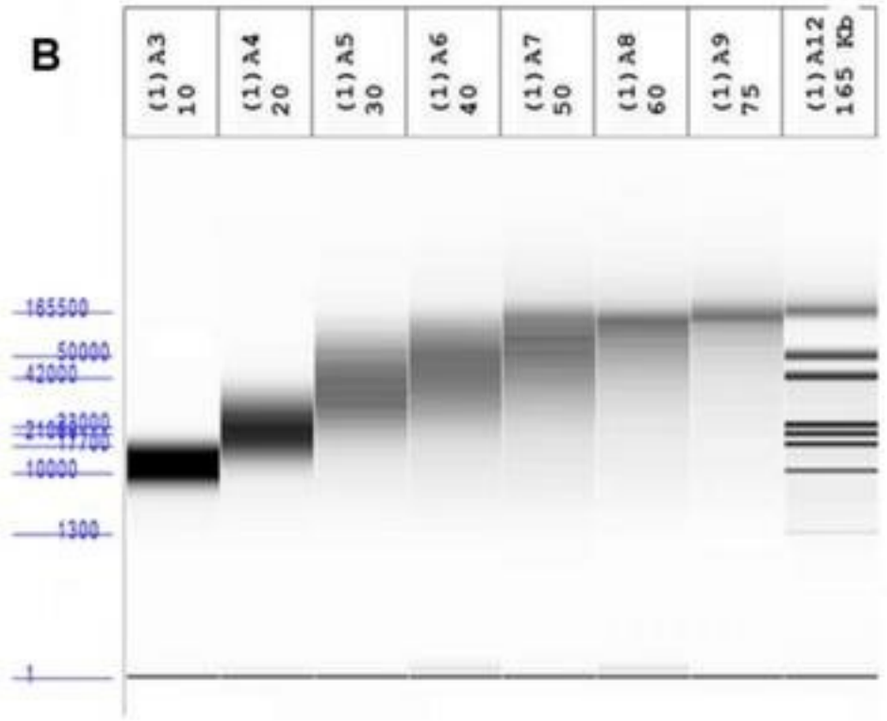
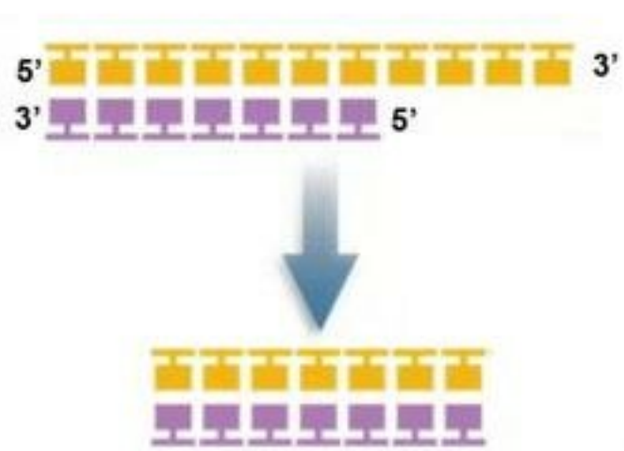


Figure B. Femto Pulse digital gel image of gDNA sheared to various target fragment sizes using a Diagenode Megaruptor 1 System.

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

REMOVE SINGLE-STRANDED DNA OVERHANGS

- Removes 3' overhangs
  - If not removed, overhang can loop and anneal to itself to form a 'hairpin-like' structure (which can lead to palindromic read structure artifacts)
- Highly recommended when constructing sheared genomic DNA libraries for whole genome sequencing *de novo* assembly applications



Why Remove Single-Stranded DNA Overhangs?

Remove SS Overhangs

PACBIO

- This pre-treatment step can help reduce the occurrence of missing adapters in sheared gDNA library data
- Missing adapters can result in chimeric reads and may cause problems in *de novo* assembly applications
- Enzymatic treatment results in digestion of both 5´→ 3´ and 3´→ 5´ overhangs, but is not active on dsDNA

How are molecular missing adapters created?

1. Shearing may create long overhangs…

2. …which can self-anneal:

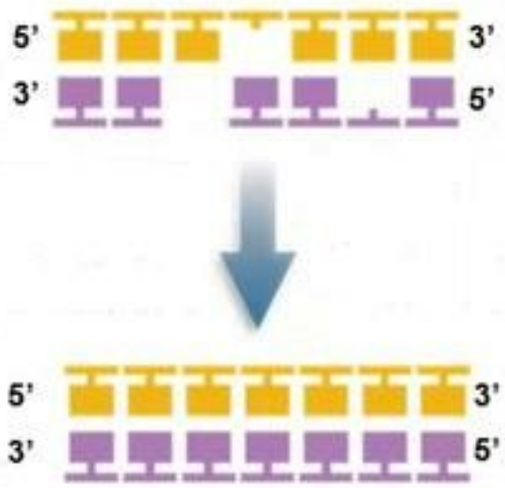
3. And get filled in during damage repair:

4. Ligation adds only <u>one</u> adapter and closes any open gaps

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

DNA DAMAGE REPAIR

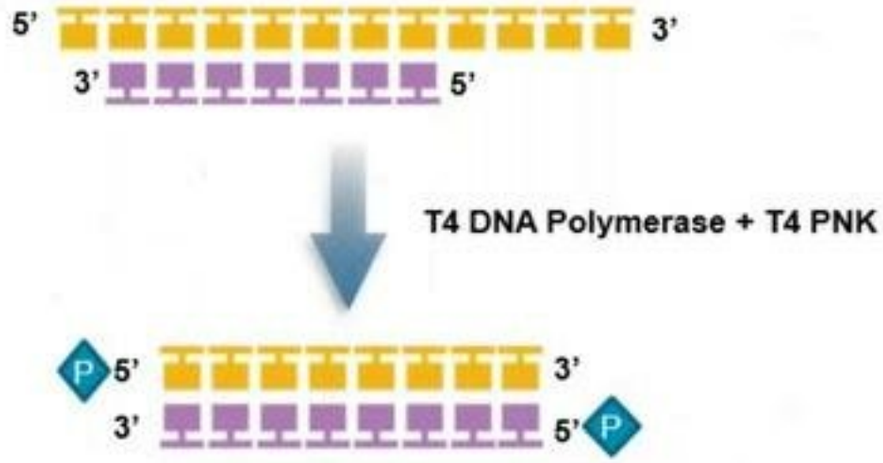
- Recommended for all library insert sizes
- DNA Damage Repair enzymes and buffer are included in the SMRTbell Express Template Prep Kit 2.0
- Repairs abasic sites, nicks, thymine dimers, blocked 3'-ends, oxidized guanines / pyrimidines, deaminated cytosine



DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

DNA END REPAIR / A-TAILING

1. DNA End Repair reaction polishes ends of fragments prior to the A-Tailing reaction:
   - 5' overhangs are filled-in by T4 DNA Polymerase
   - 3' overhangs are removed by T4 DNA Polymerase
   - T4 PNK phosphorylates the 5' hydroxyl group

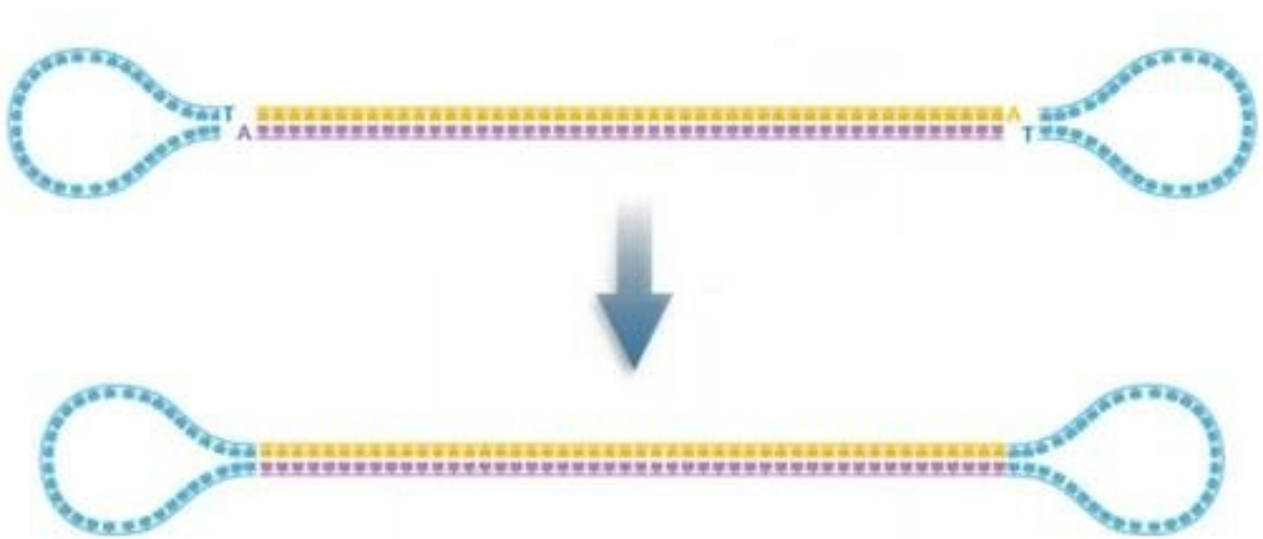


2. After DNA End Repair, an A-Tailing reaction is performed to generate single-nucleotide overhangs (A-tail):

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

ADAPTER LIGATION

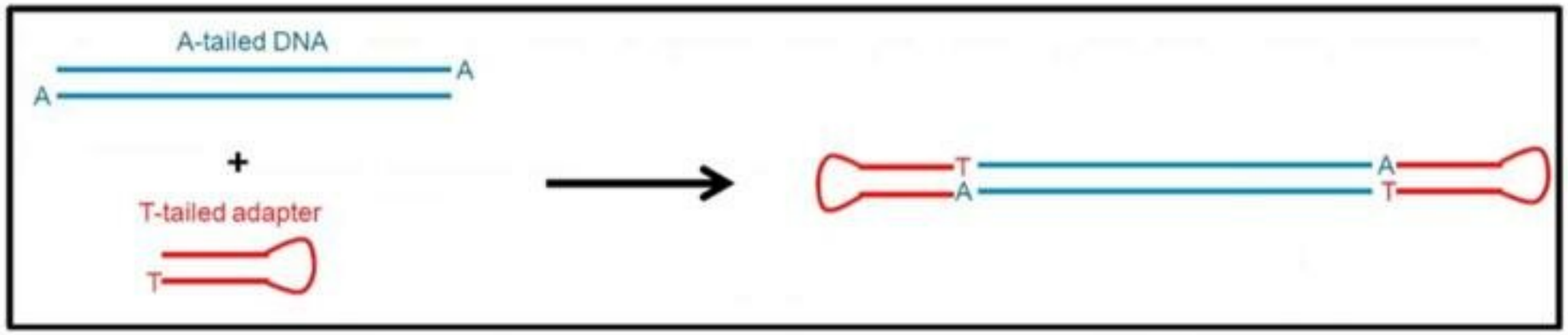
- SMRTbell hairpin overhang adapters are ligated to repaired and A-tailed dsDNA ends
- Pre-annealed hairpin overhang adapters are included in the SMRTbell Express Template Prep Kit 2.0
- Overhang adapter ligation reaction incubation time is typically 1 hour



- After Adapter Ligation, AMPure PB bead purification is performed

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0 USES A/T LIGATION INSTEAD OF BLUNT END LIGATION FOR (BARCODED AND NON-BARCODED) ADAPTER LIGATION REACTIONS



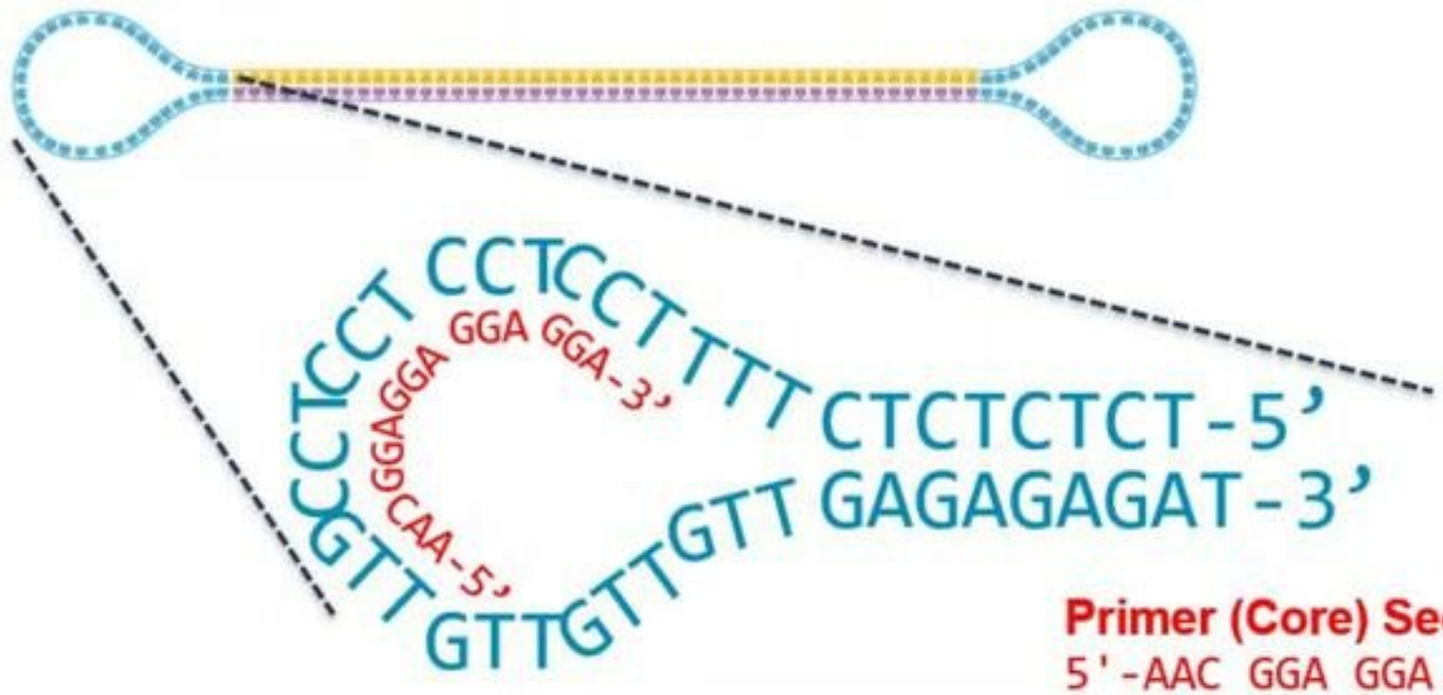
Why?

1. Reduces chimera formation

2. Reduces adapter dimerization (requires fewer AMPure cleanup steps)

SMRTbell Overhang Adapter Sequence

- Overhang adapter hairpin is 44 bases in length



Primer (Core) Sequence

5'-AAC GGA GGA GGA GGA-3’

Overhang Adapter Sequence

5’-TCT CTC TCT TTT CCT CCT CCT CCG TTG TTG TTG TTG AGA GAG AT-3’

3’-AGA GAG AGA AAA GGA GGA GGA GGC AAC AAC AAC AAC TCT CTC TA-5’

Reverse Complement

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

NUCLEASE TREATMENT *

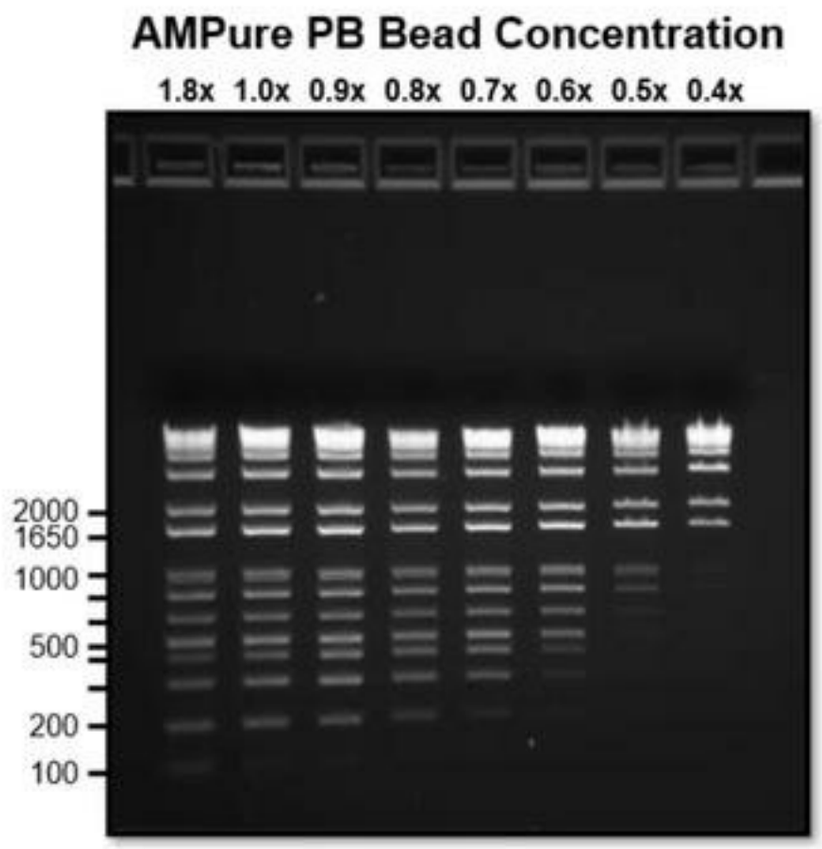
- For selected protocols*, a nuclease treatment step is performed with the SMRTbell Enzyme Cleanup Kit to reduce levels of templates molecules that are not intact, circular SMRTbell templates
- SMRTbell Enzyme Clean Up Kit contains 4 different enzyme tubes plus a buffer solution

Used for: HiFi TPK 2.0 (CCS), Multiplexed SV detection TPK 2.0 (CLR), No-Amp (CCS), Metagenomics Shotgun (10 kb; CCS), Low DNA input (With Multiplexing; CCS)

Not currently used for: gDNA CLR Library (CCS), Iso-Seq (CCS), Single-Cell Iso-Seq (CCS), Full-length 16S (CCS), Multiplexed BOA amplicon (CCS), Multiplexed BUP amplicon (CCS), Microbial Multiplexing (CLR), Low DNA input (Without Multiplexing; CCS)

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

AMPURE PB BEADS FOR DNA CONCENTRATION AND PURIFICATION THROUGHOUT SMRTBELL LIBRARY CONSTRUCTION*



- AMPure PB Beads selectively purify SMRTbell templates
- AMPure PB Beads are specifically formulated for SMRT Sequencing
- After the final AMPure PB Bead purification, the SMRTbell library is ready for primer annealing and polymerase binding
- Purified SMRTbell library can be stored at -20°C for several months

** Some library construction protocols (e.g., Iso-Seq Express TPK 2.0 libraries) require the use of ProNex Beads instead of AMPure beads for SMRTbell library purification – refer to the appropriate Procedure & Checklist document for specific instructions*

SMRTBELL EXPRESS TPK 2.0 WORKFLOW TIME SUMMARY

PACBIO

Example times shown are for constructing a >15 kb (BluePippin) size-selected large insert HiFi library starting with sheared and purified gDNA

Protocol Step	H	W
Remove Single-Strand Overhangs	5 m	15 m
DNA Damage Repair	5 m	30 m
DNA End Repair / A-Tailing	5 m	40 m
Adapter Ligation	5 m	70 m
Nuclease Treatment	5 m	60 m
Purify SMRTbell Templates	5 m	60 m
Size Selection (Optional)	0.5 h	>5 h

H Hands-on Time

W Walk-away Time

~ 5 h Library Construction Time

Overnight Size-Selection

SMRTbell Library QC and Size Selection

SMRTBELL LIBRARY QUALITY CONTROL (QC)

It is important to determine the size distribution and DNA concentration of the final purified SMRTbell library for accurate input into SMRT Link Sample Setup when preparing samples for sequencing

Technical Note TN102-040518: Preparing Libraries for PacBio Whole Genome Sequencing for de novo Assembly: Quality Control and Size Selection

A. DNA Sizing Characterization

Recommended methods for determining SMRTbell library size distribution:

1. **CHEF Mapper XA System (Bio-Rad):**
   - 100 bp up to 10 Mb

2. **Femto Pulse System (Agilent)**
   - Up to 165 kb

3. **Pippin Pulse System (Sage Science):**
   - Up to 100 kb

4. **Fragment Analyzer (Agilent):**
   - Up to 50 kb

5. **Bioanalyzer (Agilent):**
   - Up to 17 kb

B. DNA Quantification

After library preparation, measure the concentration of the SMRTbell library present using fluorometric quantitation (e.g., Qubit dsDNA HS Assay).

- Typical SMRTbell library construction yield recoveries (before size selection) are >50% for SMRTbell Express libraries

SMRTBELL LIBRARY SIZE-SELECTION

Size Selection (Optional)

PACBIO®

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

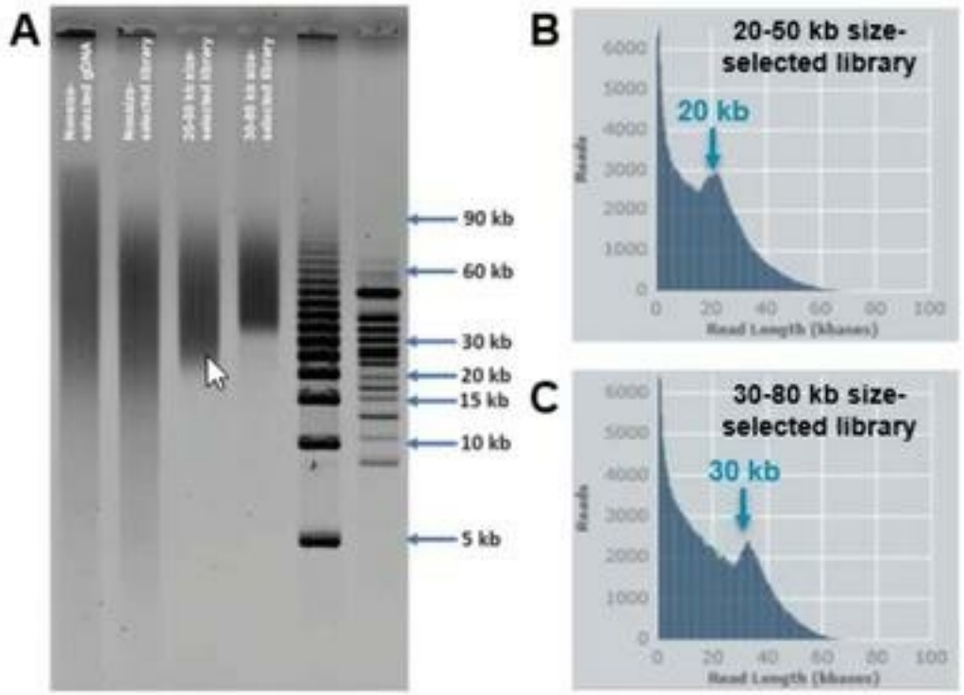
BLUEPIPPIN SYSTEM EXAMPLE

Selecting the highest-attainable size cutoff will result in optimal read lengths and yield for WGS *de novo* assembly applications

Nominal Large Insert Library Size	Size Selection Lower Cut-off	Target gDNA Shear Size
10 kb	6 kb	≥10 kb
20 kb	15 kb	≥20 – 30 kb
30 kb or 40 kb	20 kb	≥30 – 50 kb
50 kb	30 kb	≥50 – 60 kb
60 kb	40 kb	≥60 – 75 kb

The size selection protocol to be used will depend on the size distribution of the SMRTbell library

- Typical post-Blue Pippin size selection SMRTbell library recovery yields are ~15-25% (per lane)



Example SMRTbell library size selection and sequencing results. A. Pippin Pulse gel showing non size-selected genomic DNA (gDNA) in Lane 1, non-size selected library in **Lane 2**, 20-50 kb size-selected library in **Lane 3**, 30-80 kb size-selected library in **Lane 4**, and ladders in **Lanes 5** and **6**. **B.** Results of sequencing one SMRT Cell with the 20-50 kb size-selected library from **Lane 3** of **A**. **C.** Results of sequencing one Sequel SMRT Cell 1M with the 30-80 kb size-selected library from **Lane 4** of **A**.

SMRTBELL LIBRARY SIZE-SELECTION

Size Selection (Optional)

PACBIO

DNA Shearing
↓ (QC)
Remove Single-Strand Overhangs
↓
DNA Damage Repair
↓
DNA End Repair / A-Tailing
↓
Adapter Ligation
↓
Nuclease Treatment*
↓
Purify SMRTbell Templates
↓ (QC)
Size Selection (Optional)

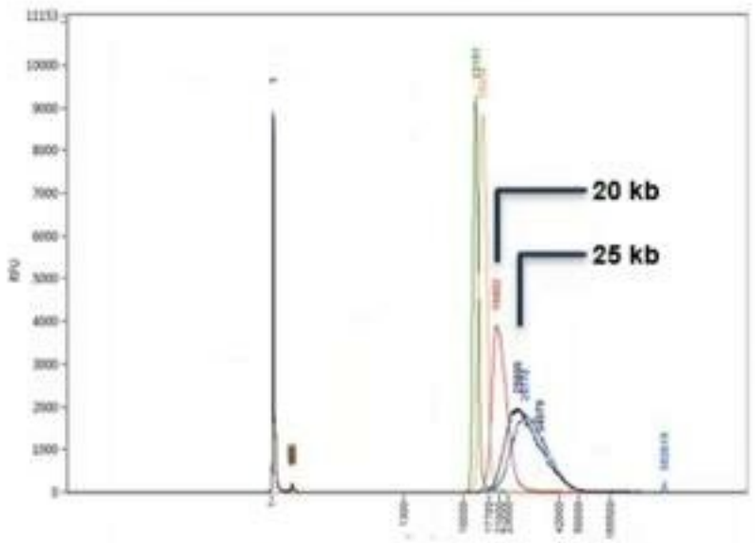
SAGEELF SYSTEM EXAMPLE

Performing library size selection will result in optimal read lengths and data yield for *de novo* assembly and variant detection applications using whole genome sequencing

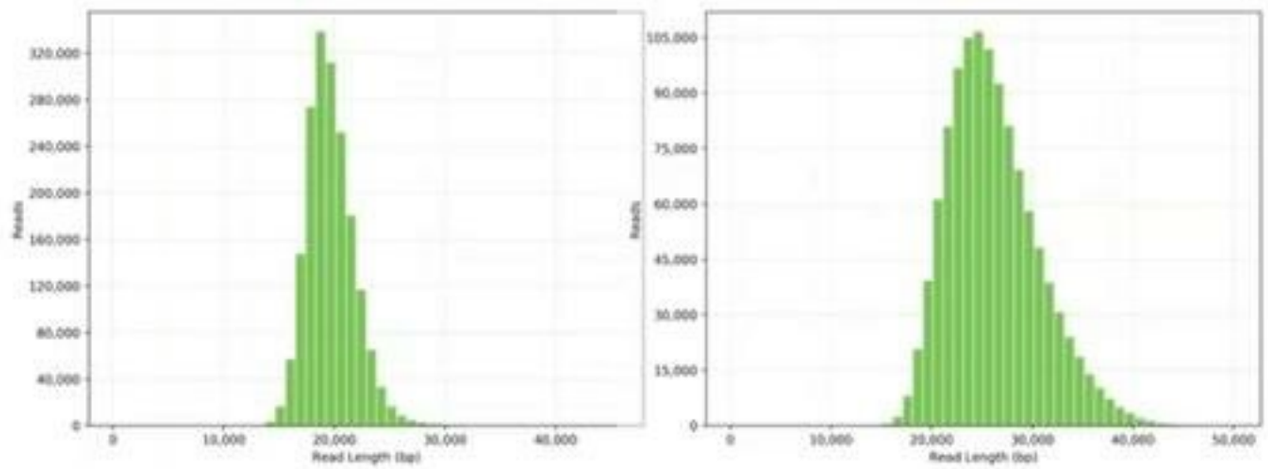
- Depending on project requirements, SMRTbell libraries may be automatically size-selected using a Sage Science SageELF system

SageELF System: Separates DNA samples by size, and then simultaneously fractionates the whole sample into 12 fractions. One sample is fractionated on a single precast agarose cassette, and one or two cassettes may be processed at one time

Example sizing QC analysis and sequencing performance metrics for a HiFi SMRTbell library size-selected using a SageELF system.



HIFI (≥Q20 CCS) INSERT READ LENGTH (MEAN)	
20 KB HUMAN HIFI LIBRARY	25 KB HUMAN HIFI LIBRARY
19,683 bp	26,149 bp



ADDITIONAL CONSIDERATIONS FOR SMRTBELL LIBRARY PREPARATION

PACBIO

Storage

- If planning to sequence within ~1 week of library generation, storing the SMRTbell library at 4°C in Elution Buffer (EB) is recommended
- For storage longer than 1 week, aliquot the SMRTbell library and store at -20°C
- Reduce or eliminate freeze/thaw cycles of your SMRTbell library to prevent damage

Shipping

- Heat exposure to DNA should be minimized or eliminated during transport (incubation at 37° C for 1 hour has been shown to cause DNA damage that may result in impaired sequencing performance)
- Lyophilized DNA may be used as long as heat is not applied during the process
- PacBio recommends shipping genomic DNA and SMRTbell libraries in a frozen state on dry ice with overnight shipping priority
    - Place the primary sample tube(s) inside a secondary form of containment like a 50 mL conical tube and surround it with bubble wrap to help ensure that the primary sample tube does not become damaged during transport
- For transporting liquid DNA samples overseas or in situations where the environment is uncontrolled, adding DNAstable® Plus (Biomatrica) preservative to samples is recommended prior to shipment on dry ice to help minimize breakdown during transport (DNAstable Plus preservative will not affect SMRTbell library preparation)

https://www.pacb.com/wp-content/uploads/Pacbio-Guidelines-DNA-Shipping-Handling-v2.0.pdf

SEQUEL II SYSTEM REAGENT KITS AND CONSUMABLES OVERVIEW

PACBIO

DNA Template Preparation

SMRTbell Express Template Prep Kit 2.0

SMRTbell Enzyme Clean Up Kit

Barcoding Reagents and Kits

Other Application-Specific Accessory Kits

SMRTbell library can be stored up to 6 months at -20° C

https://www.pacb.com/products-and-services/consumables/

SEQUEL II SYSTEM REAGENT KITS AND CONSUMABLES OVERVIEW

PACBIO

DNA Template Preparation	Polymerase Binding	Sample Cleanup	Sequencing
			
SMRTbell Express Template Prep Kit 2.0 SMRTbell Enzyme Clean Up Kit Barcoding Reagents and Kits Other Application-Specific Accessory Kits	Sequel II Binding and Internal Control Kit	AMPure PB Beads	Sequel II Sequencing Kit SMRT Cell 8M SMRT Cell Oil
SMRTbell library can be stored up to 6 months at -20° C	Polymerase-bound complex can be stored up to 7 days at 4° C	AMPure PB Bead cleaned polymerase-bound complex can be stored up to 7 days at 4° C	Partially used DNA sequencing kit can be stored at 4° C up to 100 h Unused SMRT Cells can be stored at room temperature

https://www.pacb.com/products-and-services/consumables/

Whole Genome Sequencing Applications

Whole Genome Sequencing Applications

WHOLE GENOME SEQUENCING ON THE SEQUEL II SYSTEM

PACBIO

ASSEMBLY

De novo assembly of ethnicity-specific reference genomes with **high contiguity, correctness, and completeness**

1 Human Genome per 2 to 3 SMRT Cells 8M
(~70-100 samples / year / Sequel II)

HiFi READS

VARIANT DETECTION

High precision and recall for single-nucleotide variants, indels, and structural and copy number variants with HiFi reads

1 Human Genome per 2 SMRT Cells 8M
(~100 samples / year / Sequel II)

HiFi READS

STRUCTURAL VARIANT DETECTION

High-throughput solution for **high precision and recall** of structural and copy number variants

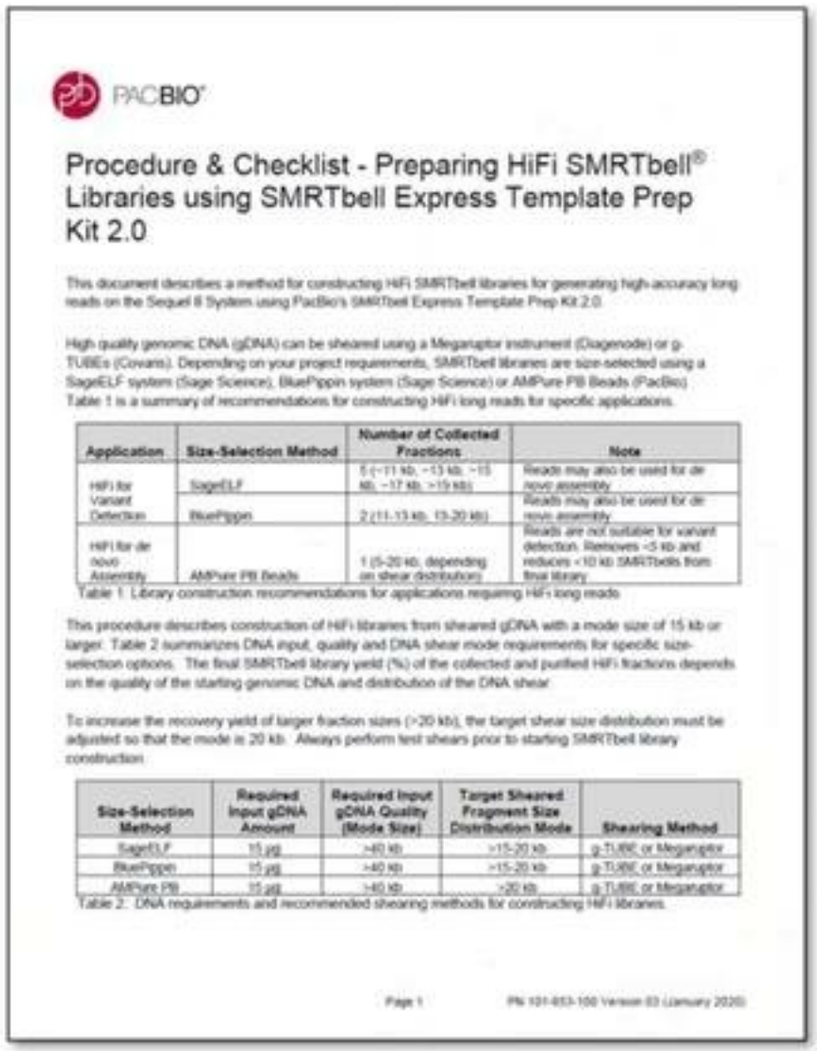
1 to 2 Human Genomes per SMRT Cell 8M
(~1000 samples / year / Sequel II)

LONG READS

PREPARING HIFI SMRTBELL LIBRARIES USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0

- Document (PN 101-853-100) describes a method for constructing SMRTbell libraries ~11 – 20 kb (or larger) that are suitable for generating high-accuracy long reads on the Sequel II System using SMRTbell Express TPK 2.0 for *de novo* assembly and variant detection applications
- Protocol document contains:

1. Recommendations for gDNA QC and quantification
2. Recommendations for shearing gDNA to the desired target mode size using either the Megaruptor System (Diagenode) or g-Tubes (Sage Science)
3. Enzymatic steps for preparation of a HiFi SMRTbell library using SMRTbell Express TPK 2.0
4. Instructions for size-selection of the HiFi SMRTbell library using either the SageELF System (Sage Science) or BluePippin System (Sage Science), and also includes protocol reference for performing AMPure BP Size Selection method for *de novo* assembly applications using HiFi reads
5. Sample setup guidance for preparing HiFi libraries for sequencing on the Sequel II System



PACBIO

Procedure & Checklist - Preparing HiFi SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0

This document describes a method for constructing HiFi SMRTbell libraries for generating high-accuracy long reads on the Sequel II System using PacBio's SMRTbell Express Template Prep Kit 2.0.

High quality genomic DNA (gDNA) can be sheared using a Megaruptor instrument (Diagenode) or g-TUBEs (Covaris). Depending on your project requirements, SMRTbell libraries are size-selected using a SageELF system (Sage Science), BluePippin system (Sage Science) or AMPure PB Beads (PacBio). Table 1 is a summary of recommendations for constructing HiFi long reads for specific applications.

Application	Size-Selection Method	Number of Collected Fractions	Note
HiFi for Variant Detection	SageELF	5 (~11 kb, ~13 kb, ~15 kb, ~17 kb, >19 kb)	Reads may also be used for de novo assembly
	BluePippin	2 (11-13 kb, 13-20 kb)	Reads may also be used for de novo assembly
HiFi for de novo Assembly	AMPure PB Beads	1 (5-20 kb, depending on shear distribution)	Reads are not suitable for variant detection. Removes <5 kb and reduces <10 kb SMRTbells from final library

Table 1: Library construction recommendations for applications requiring HiFi long reads

This procedure describes construction of HiFi libraries from sheared gDNA with a mode size of 15 kb or larger. Table 2 summarizes DNA input, quality and DNA shear mode requirements for specific size-selection options. The final SMRTbell library yield (%) of the collected and purified HiFi fractions depends on the quality of the starting genomic DNA and distribution of the DNA shear.

To increase the recovery yield of larger fraction sizes (>20 kb), the target shear size distribution must be adjusted so that the mode is 20 kb. Always perform test shears prior to starting SMRTbell library construction.

Size-Selection Method	Required Input gDNA Amount	Required Input gDNA Quality (Mode Size)	Target Sheared Fragment Size Distribution Mode	Shearing Method
SageELF	15 µg	>40 kb	>15-20 kb	g-TUBE or Megaruptor
BluePippin	15 µg	>40 kb	>15-20 kb	g-TUBE or Megaruptor
AMPure PB	15 µg	>40 kb	>20 kb	g-TUBE or Megaruptor

Table 2: DNA requirements and recommended shearing methods for constructing HiFi libraries

Page 1 PN 101-853-100 Version 03 (January 2020)

https://www.pacb.com/support/documentation/

SIMPLE WORKFLOW FOR *DE NOVO* GENOME ASSEMBLY USING HIFI READS

PACBIO

1 LIBRARY PREP

- Start with >15 µg DNA ~40 kb
- Use SMRTbell Express Template Prep Kit 2.0
- Size select at either >15 kb or >20 kb using SageELF or BluePippin System

2 SMRT SEQUENCING

- Sequence using Circular Consensus Sequencing (CCS) mode
- Recommend 10-fold HiFi read coverage per haplotype for phased assembly

3 DATA ANALYSIS

- Use SMRT Link CCS analysis application to generate HiFi reads from sequencing data
- Assemble and phase HiFi reads using FALCON & FALCON-Unzip or a wide array of community tools

GENERATE CONTIGUOUS *DE NOVO* ASSEMBLIES USING HIFI READS

PACBIO

DATASET	RICE	DROSOPHILA	HUMAN
HiFi Library Insert Size	17 kb	19 kb	15 kb
HiFi Coverage	20-fold	20-fold	22-fold
Contig N50 (Mb)	10.7	6.5	30.5

Obtain **high contiguity** assemblies with HiFi data; **high accuracy** allows assembly through very similar repeats

GENERATE COMPLETE & ACCURATE *DE NOVO* ASSEMBLIES USING HIFI READS

PACBIO

DATASET	RICE	DROSOPHILA	HUMAN
Assembly size (Gb)	0.400	0.150	2.92
Base pair accuracy (Phred/Percentage)	Q50 / 99.999%	Q50 / 99.999%	Q49 / 99.9987%
BUSCO complete	N=1,440 98.7%	N=2,799 98.9%	N=4,104 94.9%
Species-specific genes in frame	N=19,313 98.7%	N=19,947 98.9%	N=35,666 99.5%

Accuracies **>Q40** (99.99%)
>98% of genes in frame

GENERATE *DE NOVO* ASSEMBLIES QUICKLY WITH HIFI READS

PACBIO

Compute times for *de novo* assembly of a human genome

DATA TYPE		HIFI READS	LONG READS
File Type		CCS.FASTQ.GZ	SUBREADS.BAM
File Size (GB)		48	323
Read Correction Method		CCS Analysis	Pre-assembly
Time to Results (Hours)	Read Correction	17.5	43.5
	Contig Assembly	13.7	18.9

Analyses run with PacBio recommended compute infrastructure

~31 hrs HiFi READS

~62 hrs LONG READS

Assembly with HiFi Reads takes **half the time** of assembly with Long Reads.

"Maize Reference Genome Assembly in a Day"

Kevin Fengler, Corteva Agriscience

PAG 2020

REDWOOD GENOME ASSEMBLY

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

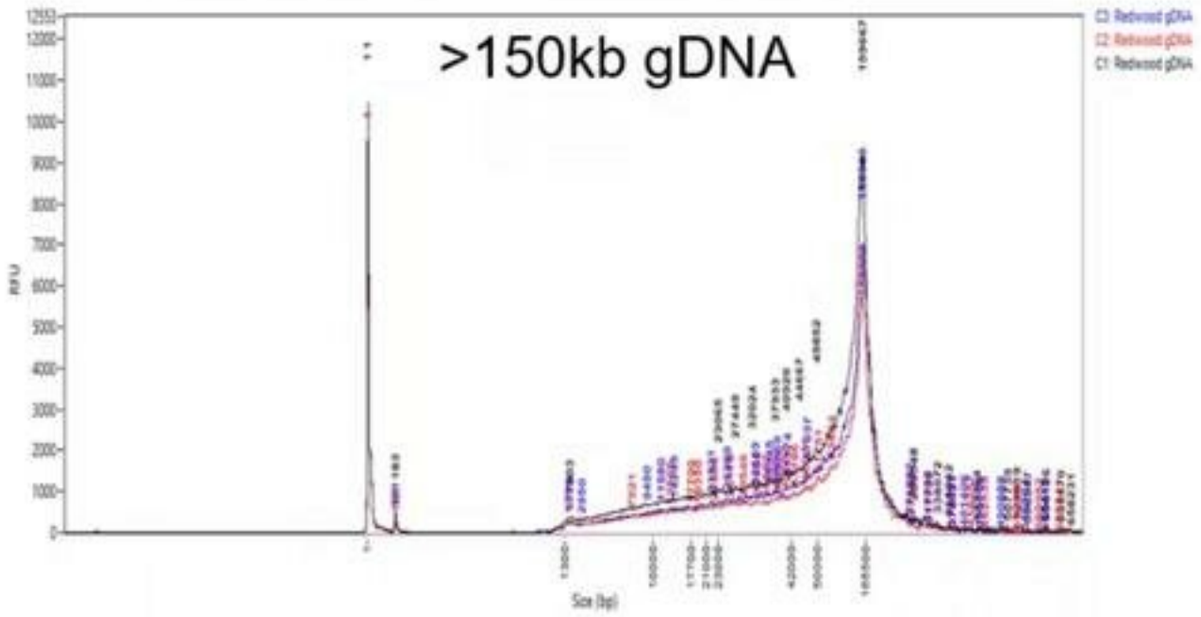
California Redwood Genome Assembly Results		
Methodology	**PacBio HiFi**	ONT + short reads[1]
Genome Coverage	**22-fold**	23-fold + 122-fold
Assembly Size (Gb)	**47.7**	26.5
Contig N50 (Mb)	**1.92**	0.11
BUSCO Complete	**59%**	56%
Mapped transcripts with frameshift errors[2]	**0.11%**	1.97%

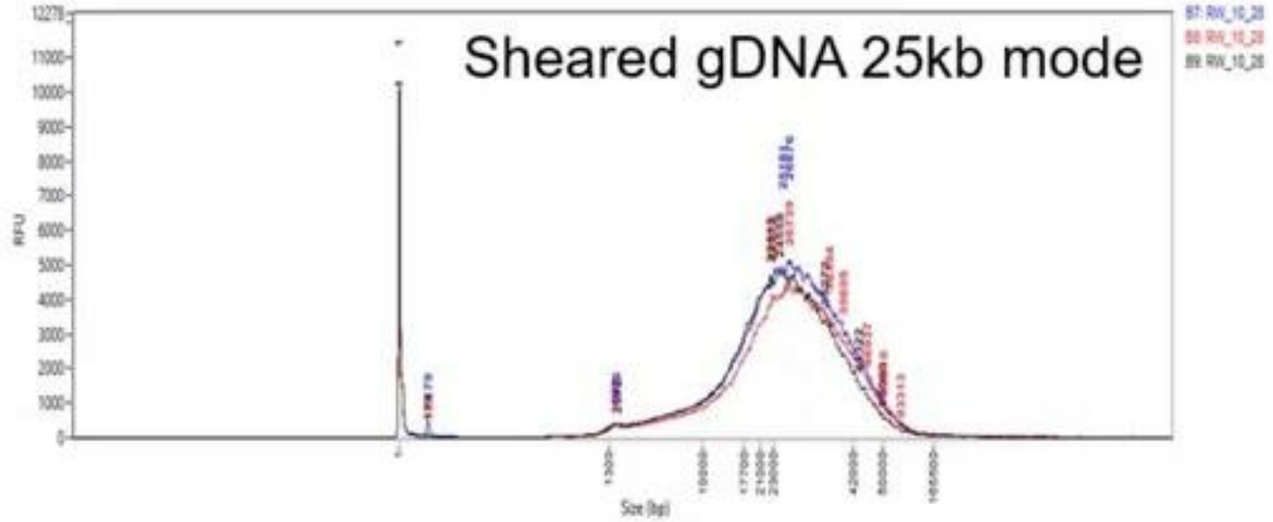
~2N assembly resolving polyploidization event

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)

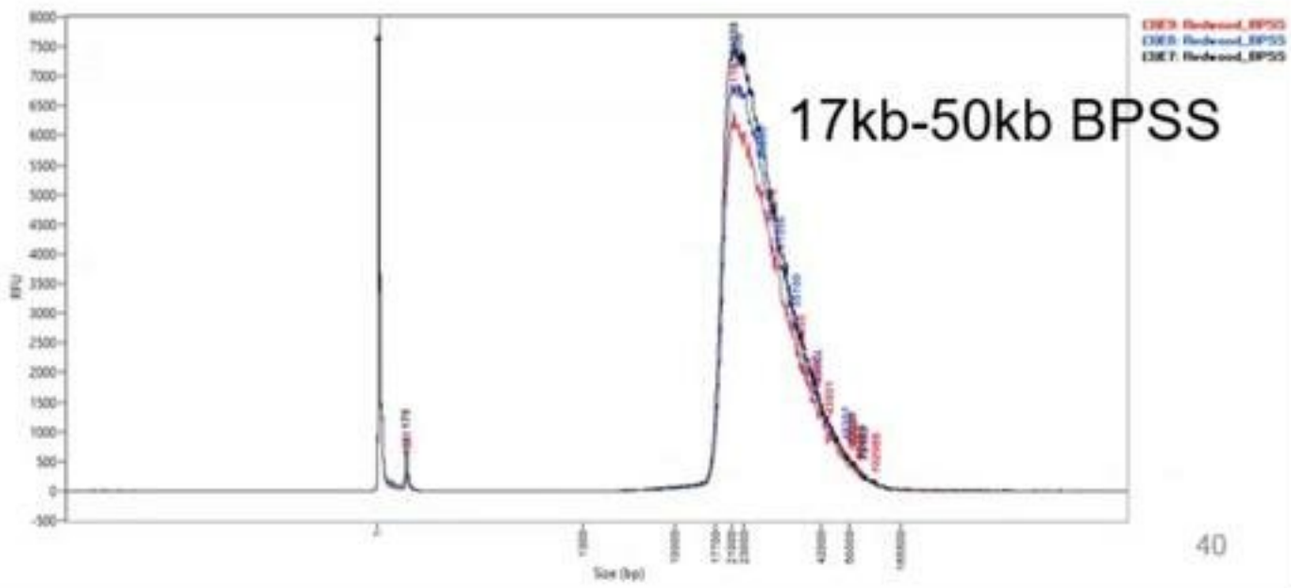
CONSTRUCTING REDWOOD HIFI LIBRARY



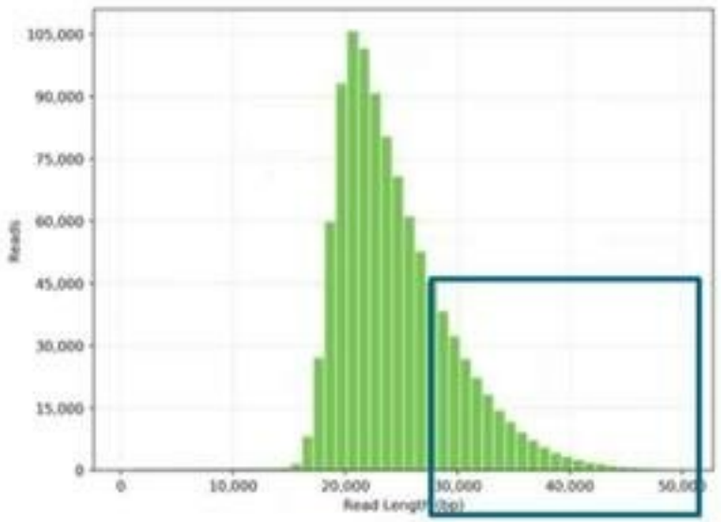


WGS for de novo assembly of large genomes:

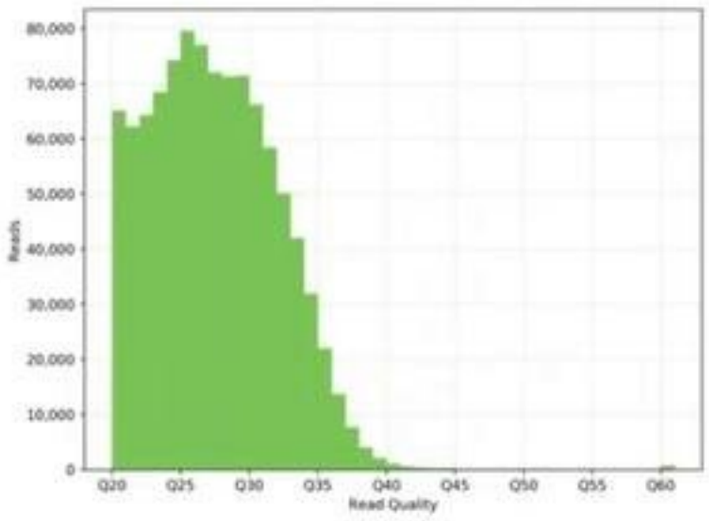
- Longest length starting DNA needed
- Recommend minimal shearing
- Size selection with BluePippin to achieve a range of fragments (20-25 kb for HiFi data but some larger fragments out to >40 kb to help improve genome contiguity)



REDWOOD HIFI READS – READ LENGTH AND QUALITY DISTRIBUTION



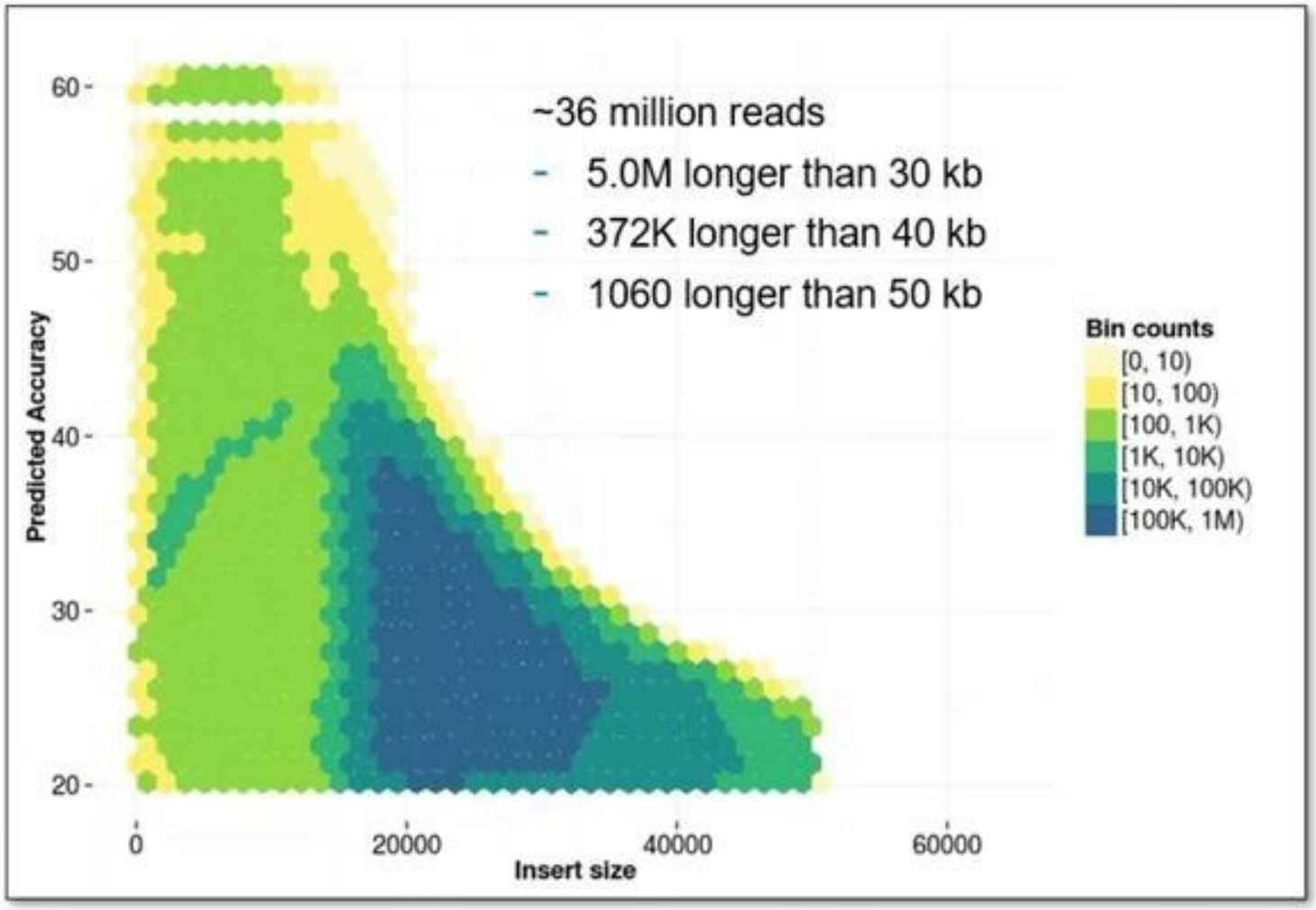
- Mean read length of CCS reads = 24kb
- >30 kb HiFi reads help assembly contiguity
- Filter = 3 passes, >Q20



- Read Quality distribution (Mean) = 27
- Filter = 3 passes, >Q20

REDWOOD HIFI DATA: READ ACCURACY

33-fold Dataset



Similar accuracies as 22-fold dataset: >Q20 even out to 50 kb with a bulk of data >Q30 (99.9%)

https://github.com/armintoepfer/ggeyecandy.git

REDWOOD GENOME ASSEMBLY TIME – THE 4TH C (COMPUTE)

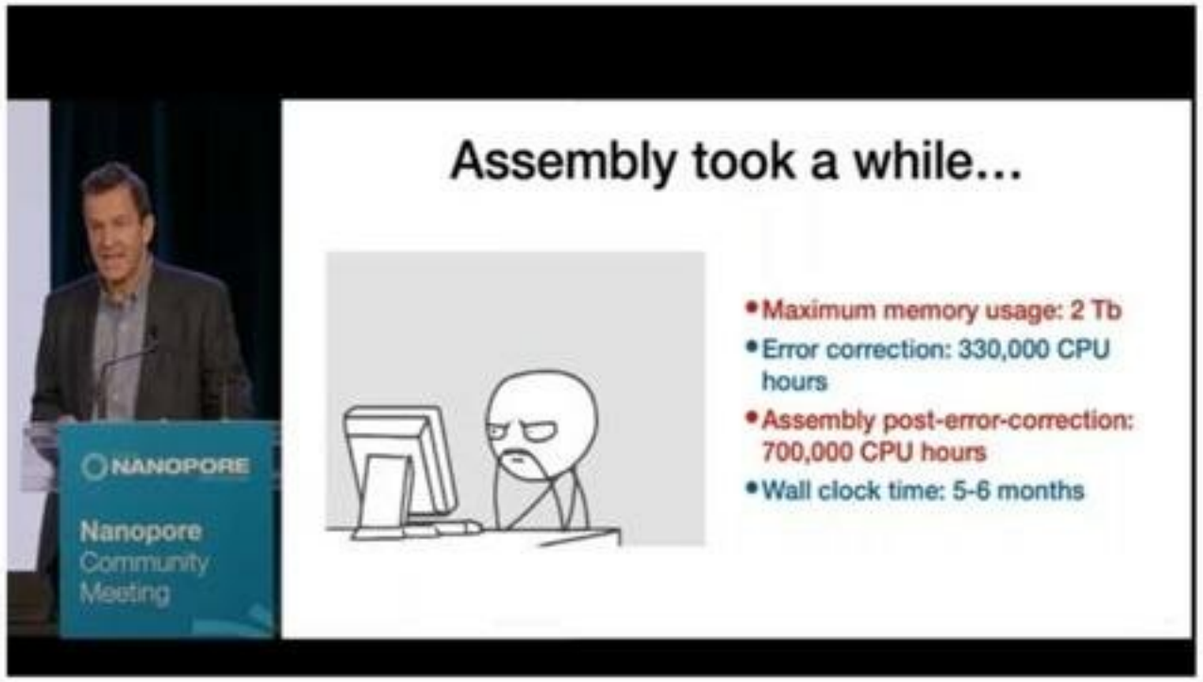
PACBIO

PacBio HiFi (22-fold)[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]

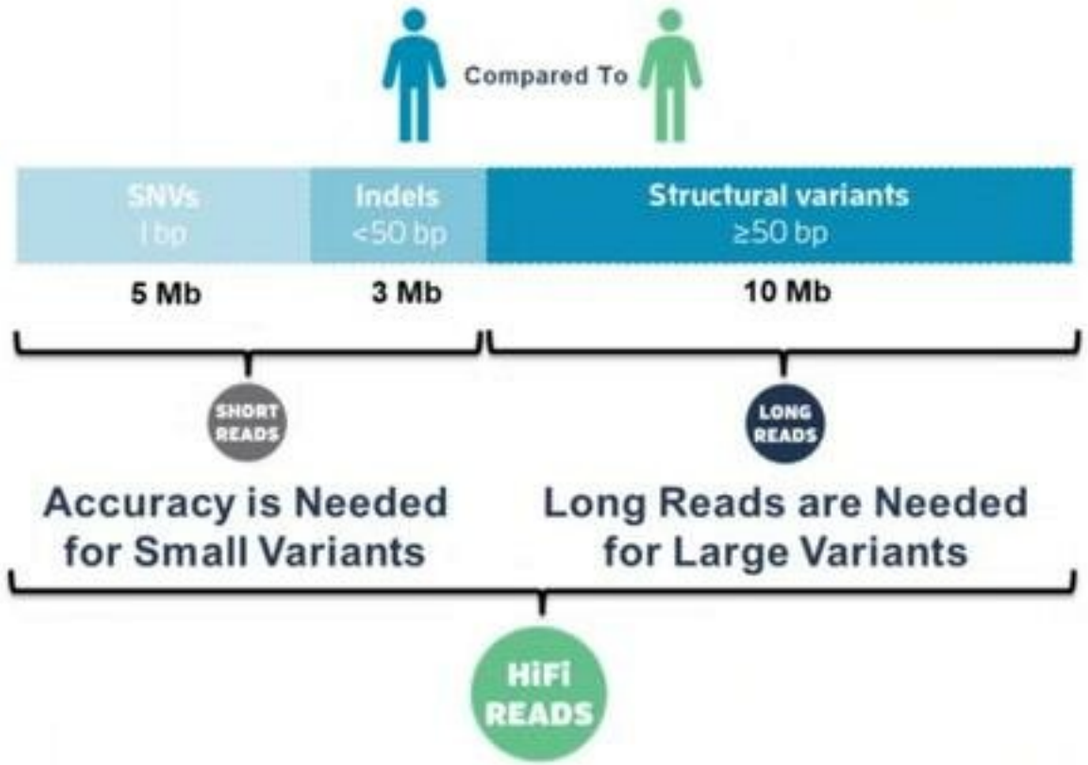


1. Sequencing and assembling mega-genomes of mega-trees: the giant sequoia and coast redwood genomes
2. Using transcript set of *Abies alba* from Neale, D. et al. consisting of 22,561 transcript sequences

COMPREHENSIVE VARIANT DETECTION WITH HIFI READS

PACBIO

PacBio Single Molecule, Real-Time (SMRT) Sequencing provides comprehensive detection of all variant types in a genome. Increased variant detection improves power to link genetics to phenotypes of interests for novel discovery of genes and causative variants.



The NEW ENGLAND JOURNAL of MEDICINE

Table 1. Classes of Human Genetic Variation.*

Class	Size of Variant	No. per Genome†	Size of Region Affected	Percent of Genome
	bp		*Mbp*	
Single-nucleotide variants	1	4,000,000–5,000,000	4–5	0.078
Insertions–deletions	1–49	700,000–800,000	3–5	0.069
Structural variants	>50	23,000–28,000	10–12	0.19
Inversions	>50	153‡	23‡	0.397
Multi-copy-number variants§	>1000	Approximately 500	12–15	0.232

Wenger, A. et al. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome *Nature Biotechnology*
Eichler, EE (2019) Genetic Variation, Comparative Genomics, and the Diagnosis of Disease, N Engl J Med. DOI: 10.1056/NEJMra1809315

MULTIPLEXING SMRTBELL LIBRARIES USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0 FOR STRUCTURAL VARIATION DETECTION

- Document (PN 101-88-600) describes how to prepare multiplexed SMRTbell libraries for structural variation (SV) detection using the Sequel II System
- Up to two genomic DNA samples can be pooled for sequencing on one SMRT Cell 8M
- Protocol document contains:

1. General laboratory best practices recommendations
2. Recommendations for gDNA QC and quantification
3. Guidelines for evaluation of gDNA samples for multiplexed SV library construction
4. Enzymatic steps for preparation of multiplexed SV SMRTbell libraries
5. Instructions for size-selection of multiplexed SV libraries using the BluePippin System, and also includes a protocol reference for performing the AMPure BP Size Selection method
6. Sample setup guidance for preparing multiplexed SV libraries for sequencing on the Sequel II System


PACBIO

Procedure & Checklist - Multiplexing SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0 for Structural Variation Detection

This document describes a procedure for constructing SMRTbell libraries for structural variant detection using the Sequel II System. Two genomic samples can be multiplexed on one SMRT Cell 8M.

High quality genomic DNA (gDNA) is sheared using a Megaruptor instrument, constructed to a SMRTbell library using the SMRTbell Express Template Prep Kit 2.0 and then size-selected. Note that the size distribution of the sheared DNA is critical for generating sufficient unique molecular coverage for structural variant detection.

Size-selection of SMRTbell libraries can be performed using either a BluePippin System or AMPure® PB beads (depending on the desired level of stringency for removing short inserts). The BluePippin System removes short insert SMRTbell libraries <15 kb efficiently and is, therefore, the preferred method for size-selection for Structural Variant multiplexing. Genomic DNA should be sheared so that the size distribution mode is between 20 kb to 40 kb (larger size is preferred). Using this method, short inserts (<15 kb) are removed during the size-selection process using a 15 kb lower cutoff.

To use AMPure PB beads for size-selection, the same sheared gDNA size distribution mode (20 kb - 40 kb) is required; however, it is important to avoid generating fragments <10 kb since AMPure PB beads are effective in removing fragments less than 5 kb only. Therefore, fragments >5 kb are carried through library construction and sequencing, which can result in shorter mean read length and lower unique molecular coverage compared to samples size-selected using the BluePippin System.

Always perform test shears to determine the best parameters to meet the above size distribution mode requirements. Additionally, the response of individual gDNA samples to shearing parameters may differ, so small-scale test shears are always required. If, after shearing, the gDNA sample contains an excess of <10 kb fragments, it is best to use the BluePippin System for size-selection.

For multiplexing, Barcoded Overhang Adapters are required for the ligation step. Any of the barcoded overhang adapters included with the Barcoded Overhang Adapter Kit 8A or Barcoded Overhang Adapter Kit 8B are suitable for use with this procedure.

Page 1 PN 101-891-600 Version 01 (November 2019)

https://www.pacb.com/support/documentation/

MORE VARIANTS DETECTED IN MEDICALLY-RELEVANT GENES

PACBIO®

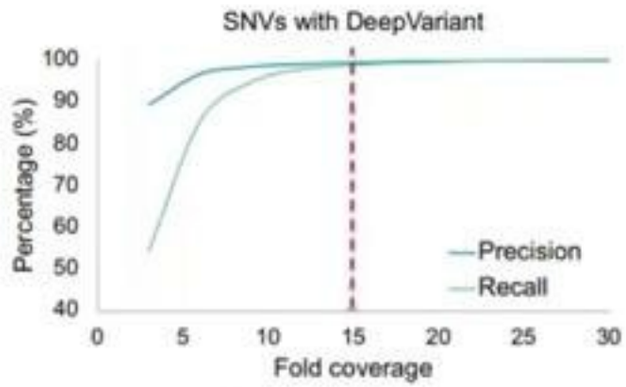
- High precision and recall for SNVs, indels, and SVs
- Detect 5% more variants in "medical exome"
- Detect structural variants
- Phase variants into haplotypes

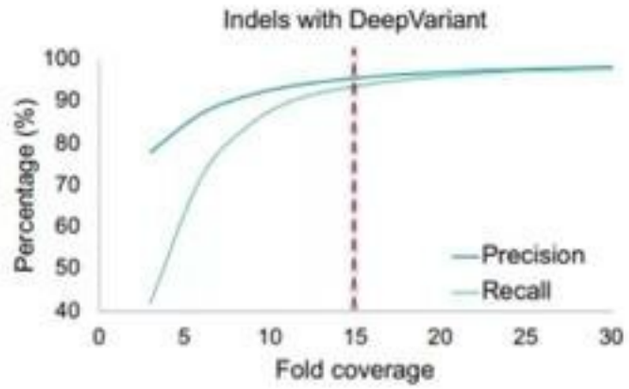
% problem exons resolved	Genes
100%	*ABCC6, ABCD1, ACAN, ACSM2B, AKR1C2, ALG1, ANKRD11, BCR, CATSPER2, CD177, CEL, CES1, CFH, CFHR1, CFHR3, CFHR4, CGB, CHEK2, CISD2, CLCNKA, CLCNKB, CORO1A, COX10, CRYBB2, CSH1, CYP11B1, CYP11B2, CYP21A2, CYP2A6, CYP2D6, CYP2F1, CYP4A22, DDX11, DHRS4L1, DIS3L2, DND1, DPY19L2, DUOX2, ESRRA, F8, FAM120A, FAM205A, FANCD2, FCGR1A, FCGR2A, FCGR3A, FCGR3B, FLG, FLNC, FOXD4, FOXO3, FUT3, GBA, GFRA2, GON4L, GRM5, GSTM1, GYPA, GYPB, GYPE, HBA1, HBA2, HBG1, HBG2, HP, HS6ST1, IDS, IFT122, IKBKG, IL9R, KIR2DL1, KIR2DL3, KMT2C, KRT17, KRT6A, KRT6B, KRT6C, KRT81, KRT86, LEFTY2, LPA, MST1, MUC5B, MYH6, MYH7, NEB, NLGN4X, NLGN4Y, NOS2, NOTCH2, NXF5, OPN1LW, OR2T5, OR51A2, PCDH11X, PCDHB4, PGAM1, PHC1, PIK3CA, PKD1, PLA2G10, PLEKHM1, PLG, PMS2, PRB1, PRDM9, PROS1, RAB40AL, RALGAPA1, RANBP2, RHCE, RHD, RHPN2, ROCK1, SAA1, SDHA, SDHC, SFTPA1, SFTPA2, SIGLEC14, SLC6A8, SMG1, SPATA31C1, SPTLC1, SRGAP2, SSX7, STAT5B, STK19, STRC, SULT1A1, SUZ12, TBX20, TCEB3C, TLR1, TLR6, TMEM231, TNXB, TRIOBP, TRPA1, TTN, TUBA1A, TUBB2B, UGT1A5, UGT2B15, UGT2B17, UNC93B1, VCY, VWF, WDR72, ZNF419, ZNF592, ZNF674*
[75%, 100%)	*ANAPC1, C4A, C4B, CHRNA7, CR1, DUX4, FCGR2B, HYDIN, OTOA, PDPK1, TMLHE*
[50%, 75%)	*ADAMTSL2, CDY2A, DAZ1, GTF2I, NAIP, OCLN, RPS17*
[25%, 50%)	*DAZ2, DAZ3, KIR3DL1, OPN1MW, PPIP5K1*
(0%, 25%)	*NCF1, RBMY1A1*
0%	*BPY2, CCL3L1, CCL4L1, CDY1, CFC1, CFC1B, GTF2IRD2, HSFY1, MRC1, OR4F5, PRY, PRY2, SMN1, SMN2, TSPY1, XKRY*

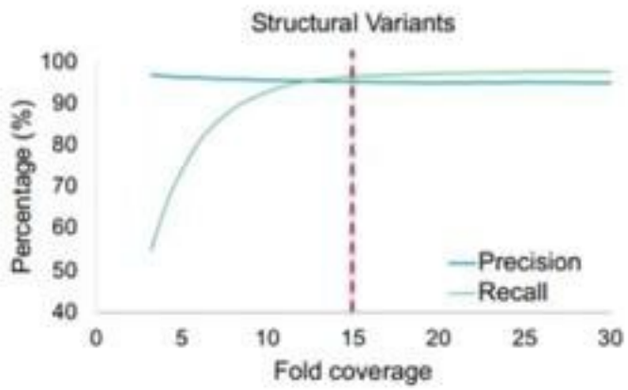
Table shows improvement in mappability with 13.5 kb HiFi (CCS) reads for 193 human genes previously reported as medically relevant and problematic to map with NGS reads

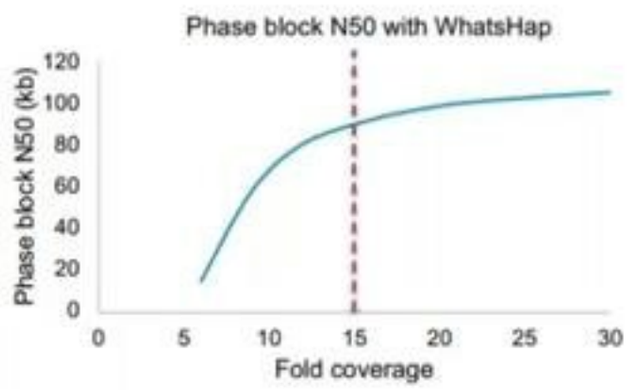
Wenger, A. et al. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*.

15-FOLD HIFI READ COVERAGE RECOMMENDATION FOR COMPREHENSIVE VARIANT DETECTION APPLICATIONS









15-fold HiFi (≥Q20) Coverage
[2 SMRT Cells 8M for a 3 Gb genome] provides a good trade-off between cost and results

WGS for human variant calling:

- Longest length starting material required
- Recommend shearing to 20 kb
- Size selection with SageELF for tight size distribution around 20 kb
- Will result in the most amount of HiFi data possible
- Important for comprehensive variant calling of SNVs, indels and SVs using DeepVariant or other tools

Wenger, A. et al., Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. 2019. *Nature Biotechnology.*

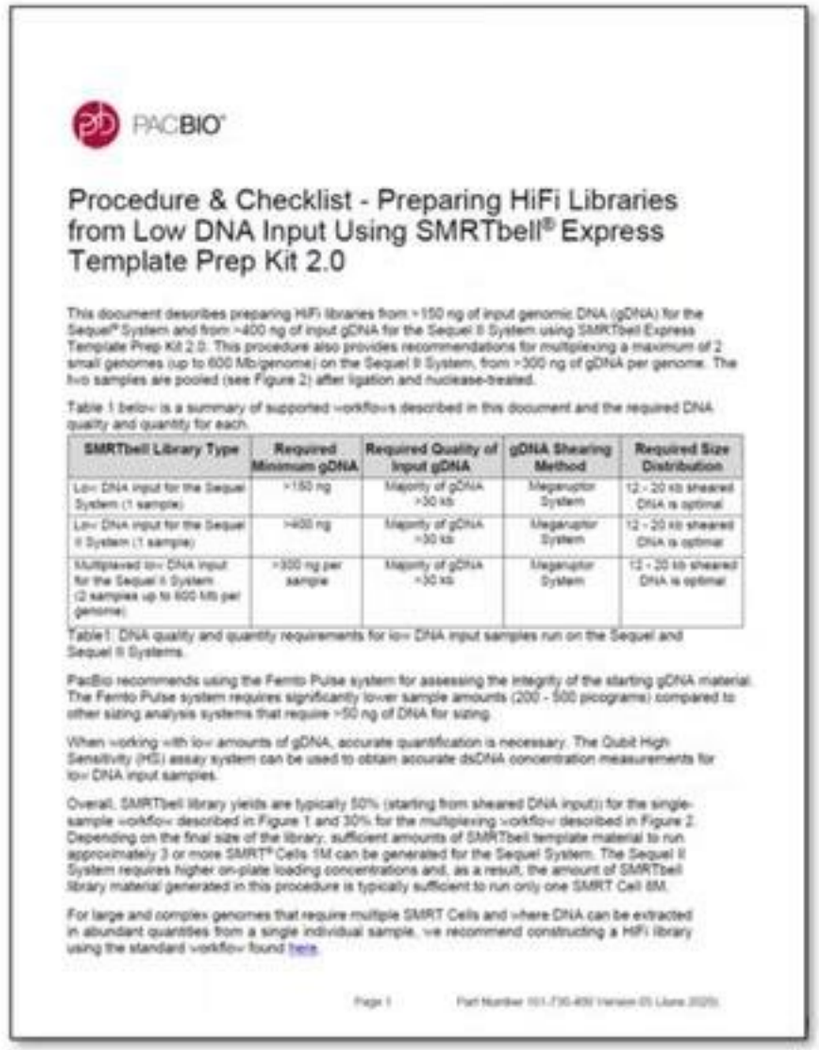
PREPARING HIFI LIBRARIES FROM LOW DNA INPUT USING SMRTBELL EXPRESS TPK 2.0

- Document (PN 101-730-400) describes preparing HiFi SMRTbell libraries for a single sample using Express Template Prep Kit 2.0 with low DNA input for sequencing on the Sequel and Sequel II Systems.
- Also describes preparing multiplexed HiFi SMRTbell libraries with a maximum of 2 pooled low DNA input samples (up to 600 Mb per genome) for sequencing on the Sequel II System.

DNA QUALITY AND QUANTITY REQUIREMENTS FOR LOW DNA INPUT SMRTBELL LIBRARY TYPES SUPPORTED IN THIS PROCEDURE

SMRTbell Library Type	Recommended Starting Input gDNA Amount	Required Input gDNA Quality	gDNA Shearing Method	Required gDNA Size Distribution
Low DNA input for the Sequel System (1 sample)*	>150 ng	Majority of gDNA >30 kb	Megaruptor System	12-20 kb sheared gDNA mean fragment size is optimal
Low DNA input for the Sequel II System (1 sample)	>400 ng	Majority of gDNA >30 kb	Megaruptor System	12-20 kb sheared gDNA mean fragment size is optimal
Multiplexed Low DNA input for the Sequel II System (2 samples up to 600 Mb per genome)	>300 ng per sample	Majority of gDNA >30 kb	Megaruptor System	12-20 kb sheared gDNA mean fragment size is optimal

*** For the Sequel System:** For genomes larger than 300 Mb, it may be necessary to increase the amount of gDNA input in proportion to the genome size to obtain the required data amount for genome assembly.



PACBIO

Procedure & Checklist - Preparing HiFi Libraries from Low DNA Input Using SMRTbell® Express Template Prep Kit 2.0

This document describes preparing HiFi libraries from >150 ng of input genomic DNA (gDNA) for the Sequel® System and from >400 ng of input gDNA for the Sequel II System using SMRTbell Express Template Prep Kit 2.0. This procedure also provides recommendations for multiplexing a maximum of 2 small genomes (up to 600 Mb/genome) on the Sequel II System, from >300 ng of gDNA per genome. The two samples are pooled (see Figure 2) after ligation and nuclease-treated.

Table 1 below is a summary of supported workflows described in this document and the required DNA quality and quantity for each.

SMRTbell Library Type	Required Minimum gDNA	Required Quality of Input gDNA	gDNA Shearing Method	Required Size Distribution
Low DNA input for the Sequel System (1 sample)	>150 ng	Majority of gDNA >30 kb	Megaruptor System	12 - 20 kb sheared DNA is optimal
Low DNA input for the Sequel II System (1 sample)	>400 ng	Majority of gDNA >30 kb	Megaruptor System	12 - 20 kb sheared DNA is optimal
Multiplexed low DNA input for the Sequel II System (2 samples up to 600 Mb per genome)	>300 ng per sample	Majority of gDNA >30 kb	Megaruptor System	12 - 20 kb sheared DNA is optimal

Table1: DNA quality and quantity requirements for low DNA input samples run on the Sequel and Sequel II Systems.

PacBio recommends using the Femto Pulse system for assessing the integrity of the starting gDNA material. The Femto Pulse system requires significantly lower sample amounts (200 - 500 picograms) compared to other sizing analysis systems that require >50 ng of DNA for sizing.

When working with low amounts of gDNA, accurate quantification is necessary. The Qubit High Sensitivity (HS) assay system can be used to obtain accurate dsDNA concentration measurements for low DNA input samples.

Overall, SMRTbell library yields are typically 50% (starting from sheared DNA input) for the single-sample workflow described in Figure 1 and 30% for the multiplexing workflow described in Figure 2. Depending on the final size of the library, sufficient amounts of SMRTbell template material to run approximately 3 or more SMRT® Cells 1M can be generated for the Sequel System. The Sequel II System requires higher on-plate loading concentrations and, as a result, the amount of SMRTbell library material generated in this procedure is typically sufficient to run only one SMRT Cell 8M.

For large and complex genomes that require multiple SMRT Cells and where DNA can be extracted in abundant quantities from a single individual sample, we recommend constructing a HiFi library using the standard workflow found here.

Page 1 Part Number 101-730-400 Version 01 (June 2020)

https://www.pacb.com/support/documentation/

ULTRA-LOW DNA INPUT LIBRARY PREP

PACBIO

For some samples (small organisms), even ~100 ng HMW gDNA is not possible

- Requires only **5 ng** of input genomic DNA
- Simple, all-in-one kit for whole genome amplification
- Early access program with Sequel II System users in progress

PACBIO

Customer Collaboration - Preparing SMRTbell® Libraries from Ultra-Low DNA Input

This document describes preparing SMRTbell libraries from 5 ng of input genomic DNA (gDNA) for the Sequel II System using SMRTbell Express Template Prep Kit 2.0. Genomic DNA is sheared to approximately 10 kb using g-TUBE or Megaruptor, amplified by PCR, subsequently constructed to a SMRTbell library and size-selected using the Blue Pippin system.

PacBio recommends using the Femto Pulse for assessing the integrity of your starting gDNA material. The Femto Pulse system requires significantly lower sample amounts (200-500 picograms) compared to other systems that require >50 ng of DNA for sizing.

When working with low amounts of gDNA, accurate quantification is required. The Qubit High Sensitivity (HS) assay system can be used to obtain accurate dsDNA concentration measurements for low DNA input samples.

Required Materials

Item	Vendor	Part Number
DNA QC		
Femto Pulse®	Agilent	P-0003-0817
DNA Quantitation		
Qubit™ Fluorometer	ThermoFisher Scientific	Q33226
Qubit™ 1X dsDNA HS Assay Kit	ThermoFisher Scientific	Q33230
DNA Shearing		
Megaruptor	Diagenode	B06010001
Long Hydropores	Diagenode	E07010002
Hydrotubes	Diagenode	C30010018
g-TUBE	Covaris	520104
SMRTbell Library Preparation		
SMRTbell Express Template Prep Kit 2.0	PacBio	100-938-900
SMRTbell gDNA Sample Amplification Kit	Pacbio	101-868-400
Eppendorf MiniSpin Plus or other equivalent benchtop centrifuge model	Eppendorf	22620100

PACBIO®

millennium science

Long-Read RNA Sequencing Applications

PACBIO ISO-SEQ METHOD: FULL-LENGTH RNA SEQUENCING

PACBIO

Iso-Seq is...

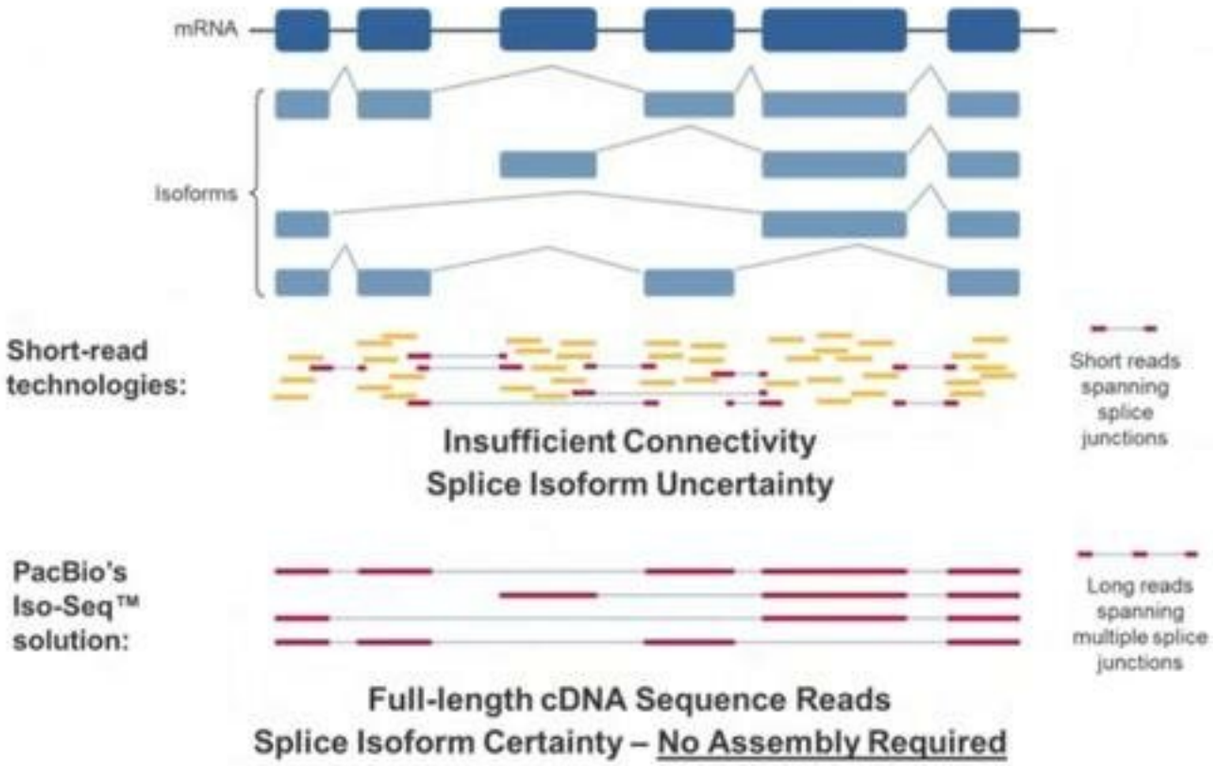
- Full-Length cDNA sequencing – *no assembly required*
- Targeted or whole transcriptome

Iso-Seq can...

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

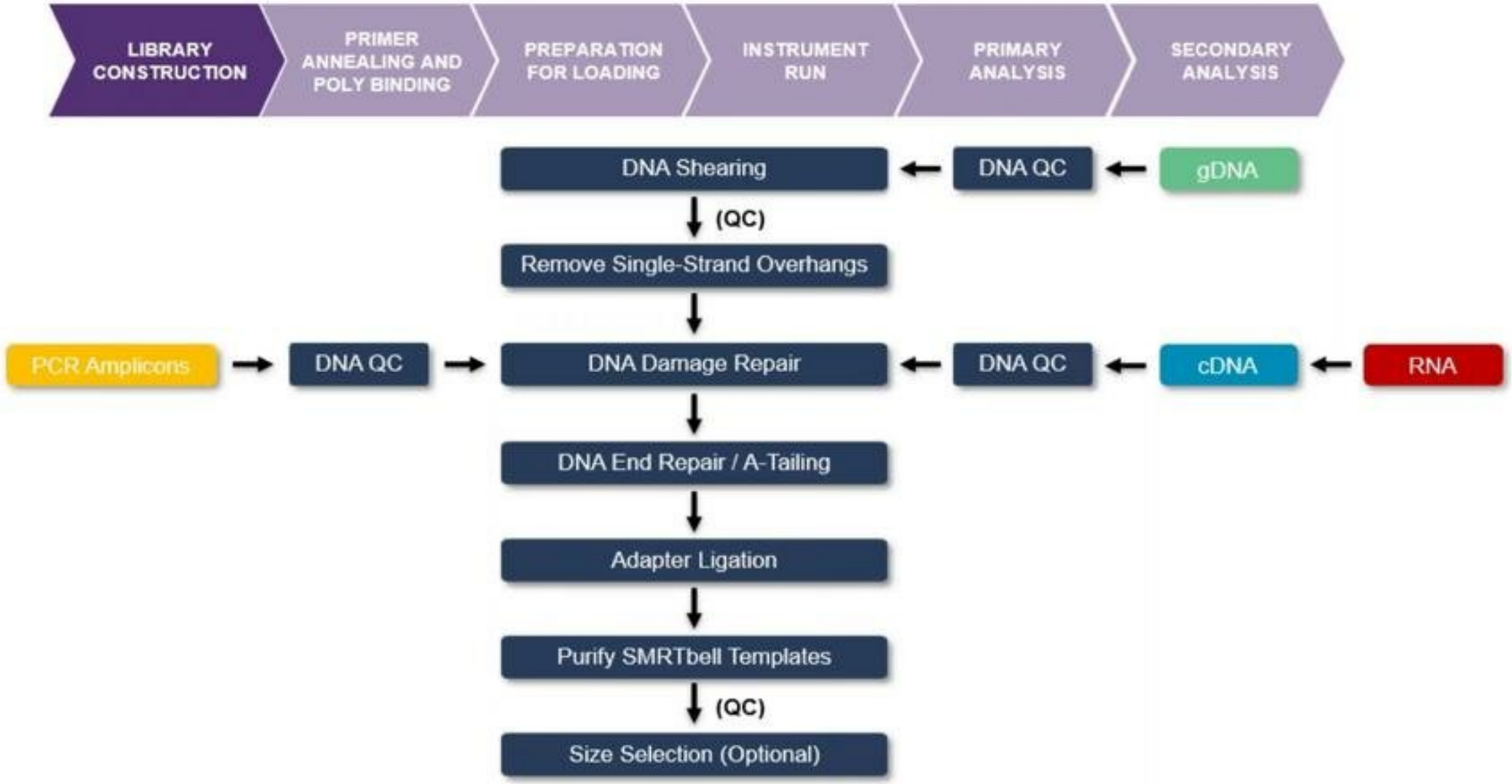
You can do Iso-Seq with...

- 1-Day sample library prep
- 1 Sequel II System SMRT Cell 1M to perform genome annotation for up to 8 multiplexed samples
- Full bioinformatics analysis solution available



SMRTBELL EXPRESS LIBRARY CONSTRUCTION WORKFLOW SUMMARY AND SAMPLE INPUT TYPES

PACBIO



ISO-SEQ EXPRESS PROTOCOL FOR FASTER RNA SEQUENCING

PACBIO

Streamlined and accelerated workflow for constructing Iso-Seq transcriptome SMRTbell libraries in one day

Total RNA (≥300 ng)

↓ Reverse Transcription

Full Length 1st Strand cDNA

↓ Large-scale Amplification

Amplified cDNA (Barcoding Option Available for Multiplexed Iso-Seq)

↓

Combined SMRTbell Library

↓ SMRT Sequencing on Sequel II System

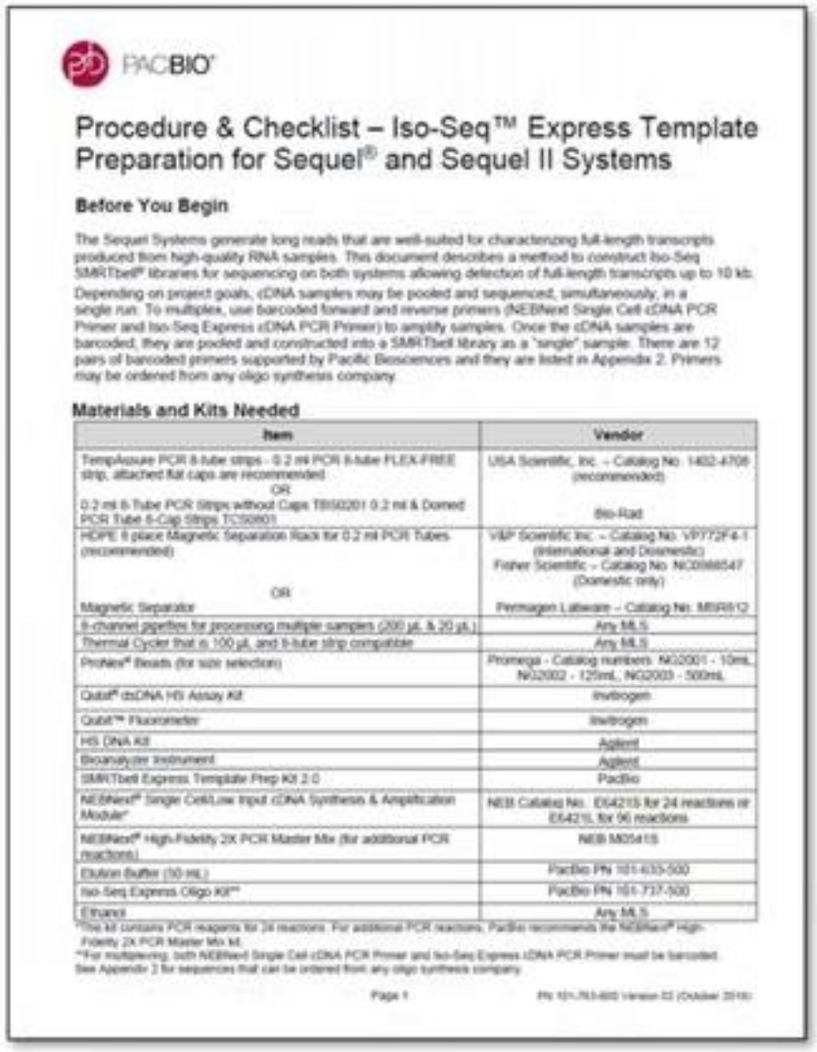
- Use PacBio RT-PCR accessory kit (PN 101-737-500) with Iso-Seq Express sample prep protocol (PN 101-763-800)
- Go from total RNA to SMRTbell library in 1 day
- Total RNA input requirement is 300 ng

- Generate up to 4 million full-length, non-concatemer (FLNC) reads per SMRT Cell 8M*
- Surveying transcript diversity can be done either broadly (**whole transcriptome**) or in a **targeted** fashion

** Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size.*

ISO-SEQ EXPRESS TEMPLATE PREPARATION FOR SEQUEL AND SEQUEL II SYSTEMS

- Document (PN 101-763-800) describes a method to construct Iso-Seq SMRTbell libraries for sequencing on Sequel and Sequel II Systems allowing detection of full-length transcripts up to 10 kb
- Streamlined and accelerated workflow for constructing Iso-Seq libraries in 1 Day
- For multiplexed Iso-Seq analysis, barcoded cDNA samples may be pooled and constructed into a SMRTbell library as a "single" sample. The pooled library sample can then be sequenced on a single SMRT Cell.
- Protocol document contains:
  1. General laboratory best practices and input RNA QC recommendations
  2. Instructions for performing first-strand cDNA synthesis and amplification of cDNA products prior to SMRTbell library construction
  3. Instructions for constructing SMRTbell libraries using amplified cDNA products and SMRTbell Express Template Prep Kit 2.0



PACBIO

Procedure & Checklist – Iso-Seq™ Express Template Preparation for Sequel® and Sequel II Systems

Before You Begin

The Sequel Systems generate long reads that are well-suited for characterizing full-length transcripts produced from high-quality RNA samples. This document describes a method to construct Iso-Seq SMRTbell® libraries for sequencing on both systems allowing detection of full-length transcripts up to 10 kb. Depending on project goals, cDNA samples may be pooled and sequenced, simultaneously, in a single run. To multiplex, use barcoded forward and reverse primers (NEBNext Single Cell cDNA PCR Primer and Iso-Seq Express cDNA PCR Primer) to amplify samples. Once the cDNA samples are barcoded, they are pooled and constructed into a SMRTbell library as a "single" sample. There are 12 pairs of barcoded primers supported by Pacific Biosciences and they are listed in Appendix 2. Primers may be ordered from any oligo synthesis company.

Materials and Kits Needed

Item	Vendor
TempAssure PCR 8-tube strips - 0.2 ml PCR 8-tube FLEX-FREE strip, attached flat caps are recommended OR 0.2 ml 8-Tube PCR Strips without Caps TS50201 0.2 ml 8-Domed PCR Tube 8-Cap Strips TCS0801	USA Scientific, Inc. – Catalog No. 1402-4708 (recommended) Bio-Rad
HDPE 8 place Magnetic Separation Rack for 0.2 ml PCR Tubes (recommended) OR Magnetic Separator	V&P Scientific Inc. – Catalog No. VP772F4-1 (International and Domestic) Fisher Scientific – Catalog No. NC9988547 (Domestic only) Permagen Labware – Catalog No. MSR812
8-channel pipettes for processing multiple samples (200 µL & 20 µL)	Any MLS
Thermal Cycler that is 100 µL and 8-tube strip compatible	Any MLS
ProNex® Beads (for size selection)	Promega - Catalog numbers: NG2001 - 10mL, NG2002 - 125mL, NG2003 - 500mL
Qubit® dsDNA HS Assay Kit	Invitrogen
Qubit™ Fluorometer	Invitrogen
HS DNA Kit	Agilent
Bioanalyzer Instrument	Agilent
SMRTbell Express Template Prep Kit 2.0	PacBio
NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module*	NEB Catalog No. E6421S for 24 reactions or E6421L for 96 reactions
NEBNext® High-Fidelity 2X PCR Master Mix (for additional PCR reactions)	NEB M0541S
Elution Buffer (10 mL)	PacBio PN 101-633-500
Iso-Seq Express Oligo Kit**	PacBio PN 101-737-500
Ethanol	Any MLS

*The kit contains PCR reagents for 24 reactions. For additional PCR reactions, PacBio recommends the NEBNext® High-Fidelity 2X PCR Master Mix kit.
**For multiplexing, both NEBNext Single Cell cDNA PCR Primer and Iso-Seq Express cDNA PCR Primer must be barcoded. See Appendix 2 for sequences that can be ordered from any oligo synthesis company.

Page 1 PN 101-763-800 Version 02 (October 2019)

https://www.pacb.com/support/documentation/

ISO-SEQ EXPRESS WORKFLOW KEY FEATURES & BENEFITS

PACBIO

New streamlined and accelerated workflow for Iso-Seq transcriptome SMRTbell libraries in one day

Iso-Seq Express Workflow Using SMRTbell Express Template Prep Kit 2.0

- **Improved formulation** for reduced Total RNA input requirement at 300 ng
- **Significantly faster workflow** from RNA to SMRTbell library in one day
- **No size-selection required**
- **Minimal handling-induced DNA damage**
- **Capture full-length transcriptomes** in a single SMRT Cell
- **Supports multiplexing up to 12 Iso-Seq library samples** per SMRT Cell

One SMRTbell Express Template Prep Kit 2.0 supports preparation of up to 18 large-insert gDNA and Iso-Seq transcriptome libraries, 48 microbial gDNA libraries or 96 amplicon template preparations

ISO-SEQ EXPRESS WORKFLOW SUMMARY OVERVIEW

PACBIO

1. **Input RNA QC**
   - ≥300 ng of Total RNA input recommended
   - RNA integrity number ≥7.0 (ideally ≥8.0)

2. **First Stand cDNA Synthesis, cDNA Amplification & Pooling**
   - NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module
   - Multiplex up to 12 samples for sequencing on a single SMRT Cell
   - Purify amplified samples with ProNex Beads followed by equimolar pooling

~3.5 h

3. **SMRTbell Express 2.0 Library Construction**
   - Single-tube, addition-only reactions
   - No size selection required
   - Use ProNex Beads for purification steps
   - Typical library yield ≥50%

~4 h

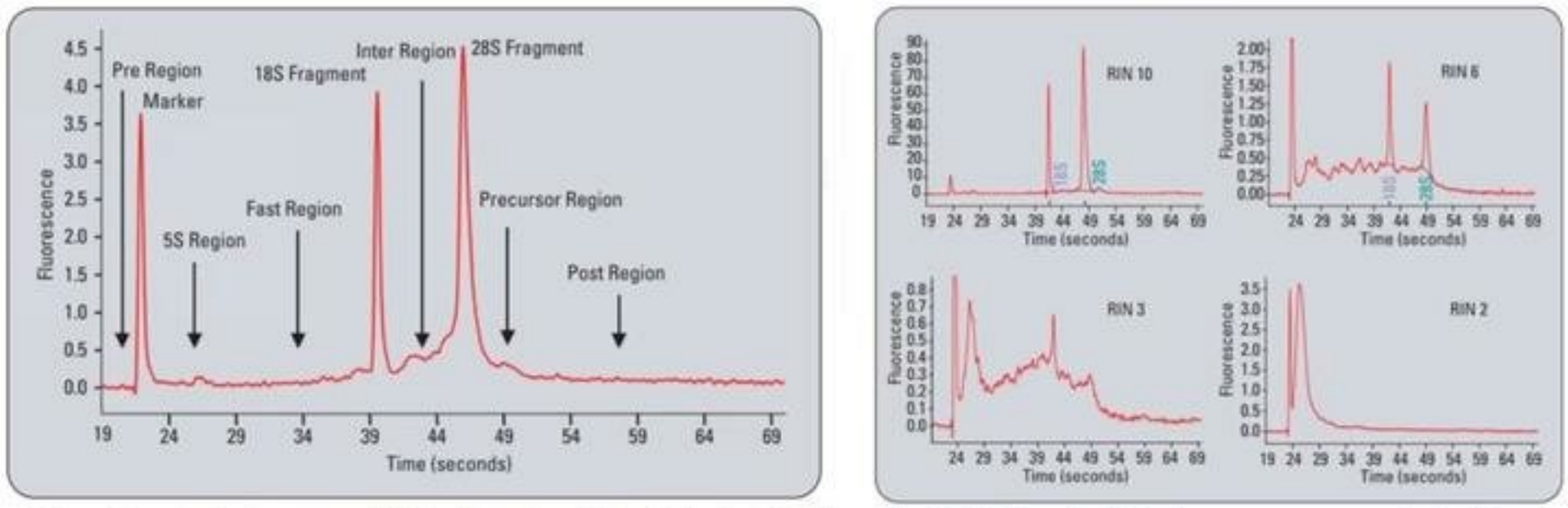
4. **Sample A/B/C & Sequence**
   - Anneal v4 Primer, Bind Polymerase, and perform ProNex Bead Complex Cleanup
   - # of SMRT Cells per library prep: >3 Sequel SMRT Cell 1M; >1 Sequel II SMRT Cell 8M
   - Pre-extension time: 4 h (Sequel System); 2 h ((Sequel II System)
   - Movie collection time: 20 h (Sequel System); 24 h (Sequel II System)

5. **Analyze**
   - Use the Iso-Seq analysis application in SMRT Link v7.0 GUI to output high-quality, full-length transcript FASTA sequences, with no assembly required, to characterize transcripts and splice variants

ISO-SEQ EXPRESS SAMPLE QC REQUIREMENTS

Evaluation of Input Total RNA Sample Integrity

- Sample QC of input Total RNA samples should be assessed by measuring the RNA Integrity Number (RIN) using a Bioanalyzer 2100 instrument (Agilent Technology)
- RIN score (1 to 10) is related to the ratio of the area under the 28s and 18s fragment peaks and also takes into account the signal intensity above the baseline in the Inter-Region and Fast Region since this is where degradation products appear
- Higher RIN numbers are correlated with better overall sample quality and lower degradation



Left: Bioanalyzer electropherogram detailing the regions that are indicative of RNA quality. **Right:** Sample electropherograms corresponding to different RNA Integrity Number (RIN) scores. Samples range from intact (RIN 10), to degraded (RIN 2). Images from *Agilent Application Note: RNA Integrity Number (RIN) – Standardization of RNA Quality Control* (https://www.agilent.com/cs/library/applications/5989-1165EN.pdf)

A RIN ≥7.0 (ideally ≥8.0) is sufficient for the Iso-Seq protocol. *Samples with a RIN <7.0 can be processed, but the risk of significant underperformance or even failure is greatly increased.*

ISO-SEQ EXPRESS SAMPLE QC REQUIREMENTS (CONT)

PACBIO

Evaluation of Input Total RNA Sample Purity

- RNA purity can be assessed through UV-spectrophotometry using a Nanodrop spectrophotometer (Thermo Scientific)
- For pure RNA, A260/280 ratio is typically ~2.0 and A260/230 ratio is ≥2.0.
- For samples with ratios that fall outside the expected optimal values, refer to the manufacturer of the RNA isolation kit for additional information regarding protocol optimization and troubleshooting.

260/280 Ratio

- A low A260/A280 ratio may indicate the presence of protein, phenol, or other contaminants that absorb strongly at or near 280 nm. Sometimes it may be caused by a very low concentration of nucleic acid.
- High 260/280 ratios are not indicative of an issue.

260/230 Ratio

- A low A260/A230 ratio may be the result of:
  - Carbohydrate carryover (often a problem with plants).
  - Residual phenol from nucleic acid extraction.
  - Residual guanidine (often used in column-based kits).
  - Glycogen used for precipitation.

PacBio recommends only proceeding with RNA samples that have an absorbance A260/A280 ratio between 1.8 and 2.0 (or higher) and a A260/A230 between 2.0 and 2.5.

ISO-SEQ EXPRESS SAMPLE QC REQUIREMENTS (CONT)

PACBIO

Minimize Genomic DNA Contamination

- It is best to use extraction methods that selectively precipitate RNA and minimize contaminating genomic DNA.
- DNase I treatments can be used to remove contaminating DNA, but before performing a treatment we recommend assessing the risk it poses to RNA integrity.
  - For example, only use RNase-free DNase and avoid the heat inactivation methods which can degrade RNA in the presence of metal ions.
  - If you do use a DNase treatment, PacBio recommends using one of the commercially available kits that includes a purification method that does not involve heat inactivating the DNase I enzyme.
- In most circumstances, low-level residual genomic DNA contamination is not problematic for the Iso-Seq application.
  - This is because of the use of the oligo-dT primer in combination with the 5' template-switching oligo (TSO) during cDNA synthesis.
  - Moreover, the subsequent PCR using primers annealing to the sequences on the 5' TSO and 3' dT primer further selects against any contaminating DNA fragments.

BARCODING SAMPLES FOR MULTIPLEXED ISO-SEQ ANALYSES

cDNA samples may be barcoded and pooled together prior to construction into a SMRTbell library as a "single" sample

- To multiplex, use barcoded forward and reverse primers (i.e., barcoded NEBNext Single Cell cDNA PCR Primer and barcoded Iso-Seq Express cDNA PCR Primer) to amplify cDNA samples
- Once the amplified cDNA samples are barcoded, they are purified using ProNex Beads, pooled together and then constructed into a SMRTbell library as a "single" sample.
- There are 12 pairs of barcoded primers supported by PacBio and they are listed in **Appendix 2** of the protocol
- Barcoded forward and reverse primers may be ordered from any oligo synthesis company
- The oligos must be diluted to 12 µM concentration for use in the "cDNA Amplification" section of the procedure. (Use 10 mM Tris, 0.1 mM EDTA for diluting oligos)

Appendix 2: Recommended Barcoded NEBNext Single Cell cDNA PCR Primer and Iso-Seq Express cDNA PCR Primer Sequences

Name	Sequence	Scale	Purification
bc1001-F	CACATATCAGAGTGCGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1001-R	CACATATCAGAGTGCGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1002-F	ACACACAGACTGTGAGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1002-R	ACACACAGACTGTGAGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1003-F	ACACATCTCGTGAGAGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1003-R	ACACATCTCGTGAGAGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1004-F	CACGCACACACGCGCGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1004-R	CACGCACACACGCGCGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1005-F	CACTCGACTCTCGCGTGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1005-R	CACTCGACTCTCGCGTAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1006-F	CATATATATCAGCTGTGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1006-R	CATATATATCAGCTGTAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1008-F	ACAGTCGAGCGCTGCGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1008-R	ACAGTCGAGCGCTGCGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1012-F	ACACTAGATCGCGTGTGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1012-R	ACACTAGATCGCGTGTAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1018-F	TCACGTGCTCACTGTGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1018-R	TCACGTGCTCACTGTGAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1019-F	ACACACTCTATCAGATGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1019-R	ACACACTCTATCAGATAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1020-F	CACGACACGACGATGTGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1020-R	CACGACACGACGATGTAAGCAGTGGTATCAACGCAGAGT	25nm	STD
bc1023-F	CAGAGAGATATCTCTGGCAATGAAGTCGCAGGGTTG	25nm	STD
bc1023-R	CAGAGAGATATCTCTGAAGCAGTGGTATCAACGCAGAGT	25nm	STD

Oligo Order Sheet for 12 Barcoded Iso-Seq primers:

https://www.pacb.com/wp-content/uploads/IsoSeq_Primers_12_Barcodes_v1_Ordering_Sheet.xlsx

PURIFICATION OF AMPLIFIED CDNA PRODUCTS

The specific method chosen to purify the amplified cDNA depends on the goal of the experiment and the expected size distribution of transcripts.

- Use **Pronex Beads** for purification of amplified cDNA products according to the table below:

Workflow	Goal of Experiment	Pronex Bead Volume
Standard	Sample is composed primarily of transcripts centered at ~2 kb	86 µL
Short Transcripts	Sample is composed primarily of transcripts <2 kb; or Transcripts of research interest are primarily <2 kb; or Sample is degraded and shows a low RIN number	95 µL
Long Transcripts	To obtain material enriched for longer transcripts >3 kb	82 µL

- After purification, perform a sizing QC by running 1 µL of the purified cDNA products on a Bioanalyzer using a High Sensitivity DNA kit.
- Examining the amplified cDNA on a Bioanalyzer prior to PacBio library construction is an excellent quality control step to ensure that the amplified cDNA material has the expected size distribution.

QUANTIFICATION OF AMPLIFIED CDNA PRODUCTS

You must have the required mass of purified cDNA to proceed with SMRTbell library construction

Instrument	Min. cDNA Amount for 1 sample	Min. cDNA Amount for Multiplexed Sample	Recommendation for Samples with Low Yield
Sequel	80-500 ng	80-500* ng	Go to **Appendix 1** if total mass is <80 ng (<1.75 ng/µL)
Sequel II	160-500 ng	160-500* ng	Go to **Appendix 1** if total mass is <160 ng (<3.5 ng/µL)

* This refers to the required total mass of the pooled cDNA samples (not individual samples of the pool). See "Sample Pooling" section for additional information.

- **Appendix 1:** Recommendations for Additional cDNA Amplification by PCR for Samples with a Lower Yield or to Enrich for Longer Transcripts
  - The Sequel and Sequel II Systems require different amounts (ng) of cDNA for SMRTbell library construction. The Sequel System requires >80 ng of DNA, while the Sequel II System requires >160 ng DNA.
  - If there is not enough DNA to proceed with library construction, this Appendix section describes a workflow for enriching cDNA by PCR.
- Note: Over-amplification can result in sub-optimal data.
  - For high-yield samples with concentrations >40 ng/µL, optimal libraries may be obtained by repeating cDNA generation with less RNA input or by decreasing the number of PCR cycles.

If you want to **enrich for longer transcripts (>3 kb),** additional cDNA amplification (as described in **Appendix 1**) is **required**

SAMPLE POOLING FOR MULTIPLEXED ISO-SEQ ANALYSES

Equal molar pooling of barcoded cDNA samples is necessary to generate good representation of samples that are being multiplexed.

1. Use the concentration and average library size* from the Bioanalyzer trace to determine the molarity of each sample. Use the following equation to determine Molarity:

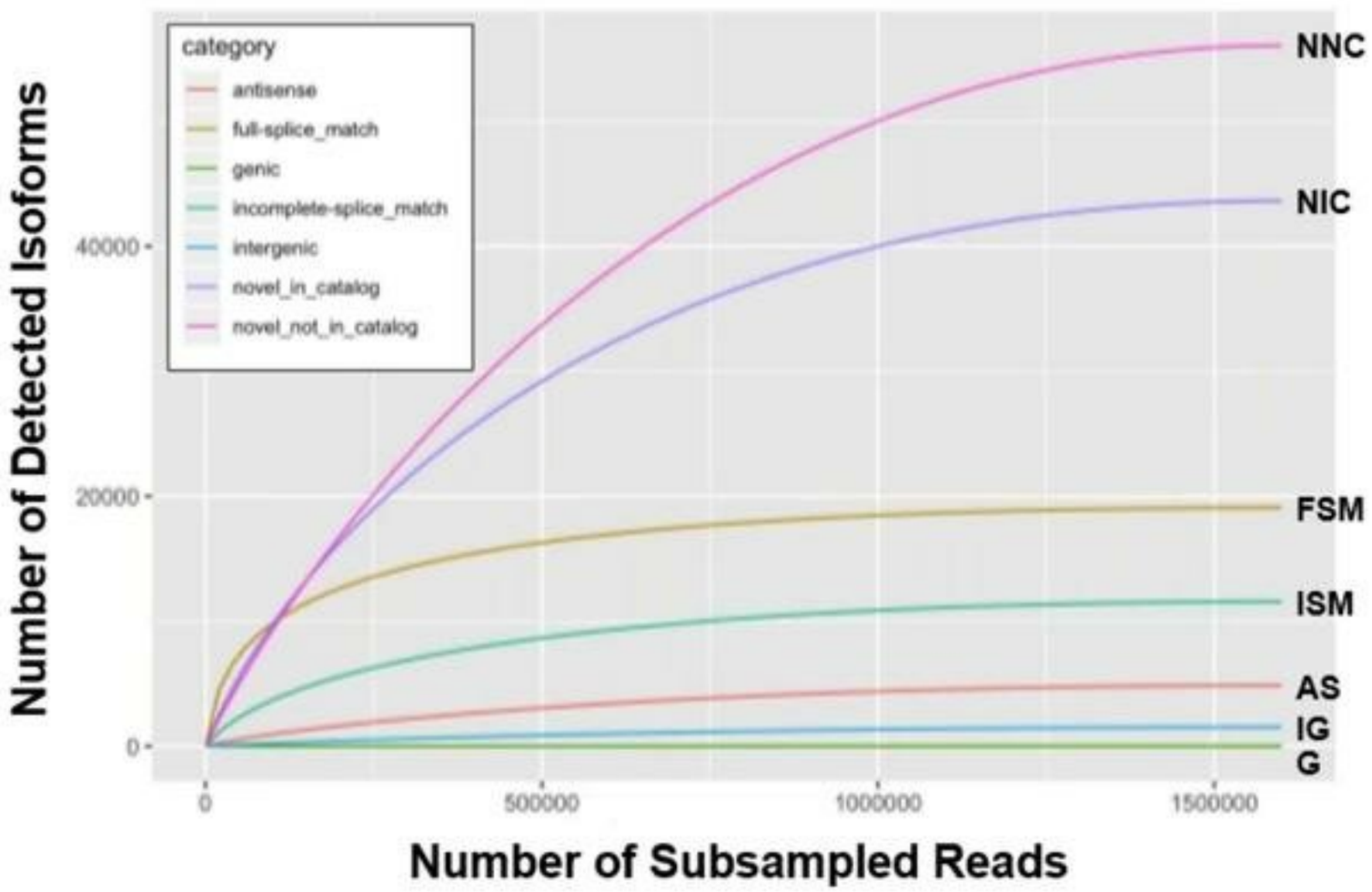
$$Concentration\ in\ nM = \frac{(DNA\ Concentration\ in\ ng\ \mu L^{-1})\ x\ 10^6}{(660\ g\ mol^{-1}\ x\ Average\ Library\ Size\ in\ bp)}$$

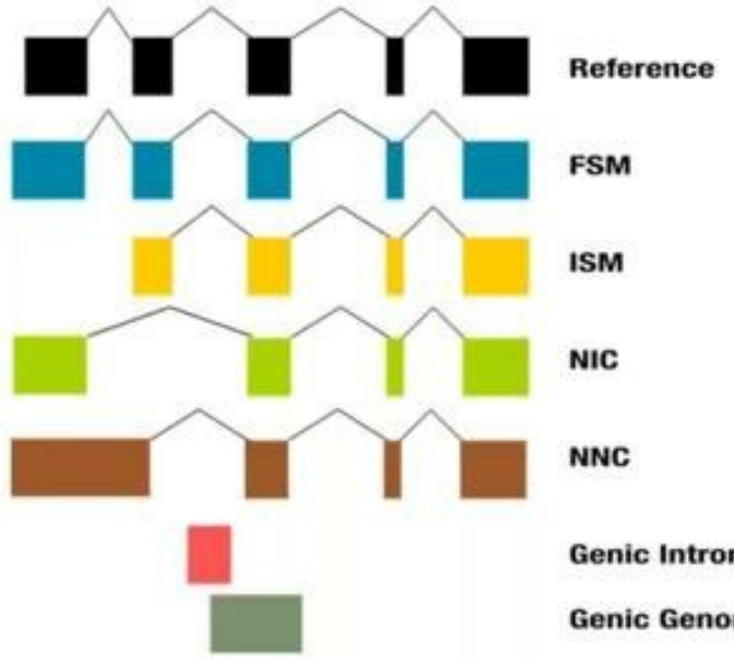
*To determine the average library size using a Bioanalyzer System, select the region of interest by defining the start of the smear at 200 bp and the end point at 9500 bp (when using a High Sensitivity DNA assay kit).

2. Pool equal molar quantities of the barcoded cDNA.
   - Use the maximum total combined mass possible without exceeding 500 ng in 47.4 µL.
   - The total combined mass must be >80 ng for Sequel and >160 ng for Sequel II to proceed to DNA Damage Repair.
   - If the volume required to achieve the minimum mass of the pooled cDNA exceeds 47.4 µL, concentrate the pooled cDNA by performing a 1X volume of ProNex beads and elute it in 48 µL. To account for potential losses during concentration at this step, start with ≥100 ng for Sequel and ≥200 ng for Sequel II.
3. The pooled cDNA can now be constructed into a SMRTbell library as a single sample. Proceed to the DNA Damage Repair step.

COMPREHENSIVE HUMAN TRANSCRIPTOME CHARACTERIZATION WITH ONE SEQUEL II SMRT CELL 8M

PACBIO





FSM = Full Splice Match, matches reference perfectly

ISM = Incomplete Splice Matches, matches reference partially

NIC = Novel In Catalog, novel isoform using known junctions

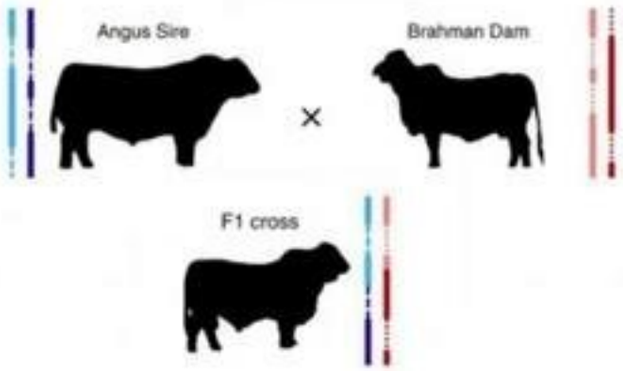
NNC = Novel Not in Catalog, novel isoforms using novel junction

Genic Intron = Within intron

Genic Genomic = Overlap with intron and exons

Tardaguila, M. *et al.* SQANTI: extensive characterization of long read transcript sequences for quality control in full-length transcriptome identification and

ISO-SEQ METHOD FOR PLANT & ANIMAL SCIENCES



Brahman x Angus F1 Cattle

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

Low et al., Haplotype-Resolved Cattle Genomes Provide Insights Into Structural Variation and Adaptation. 2019. bioRxiv

Cannabis

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

McKernan et al., Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. 2020. bioRxiv

Grizzly Bear

- Tissue-specific alternative splicing
- Hibernation vs active state

Trojahn, Kelley, et al. Presentation at UGM: https://www.pacb.com/videos/user-group-meeting-using-grizzly-bears-to-unlock-the-biomedical-promise-of-hibernation/

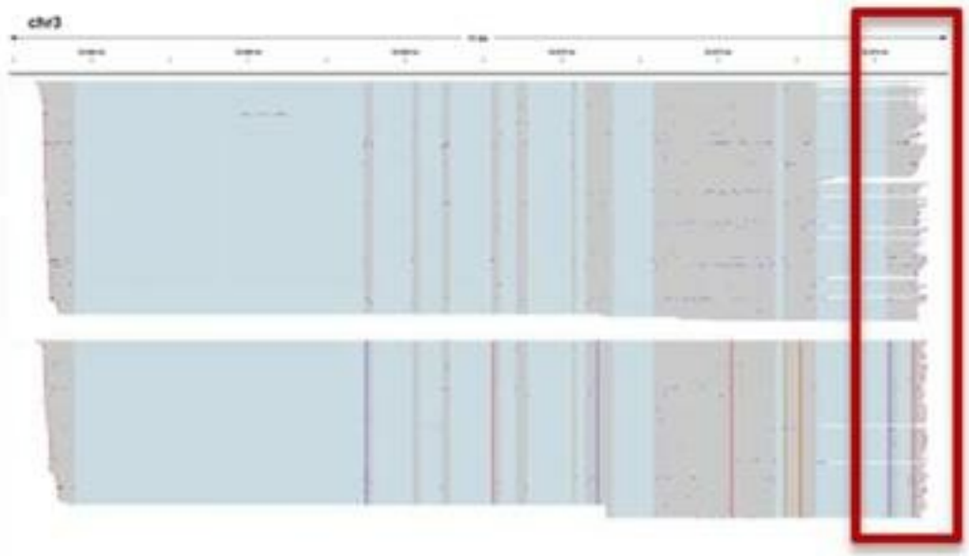
ISO-SEQ ANALYSIS OF ALLELE-SPECIFIC ISOFORM EXPRESSION IN MAIZE F1 HYBRID OFFSPRING


B73


Ki11


male B73
X
female Ki11


male Ki11
X
female B73

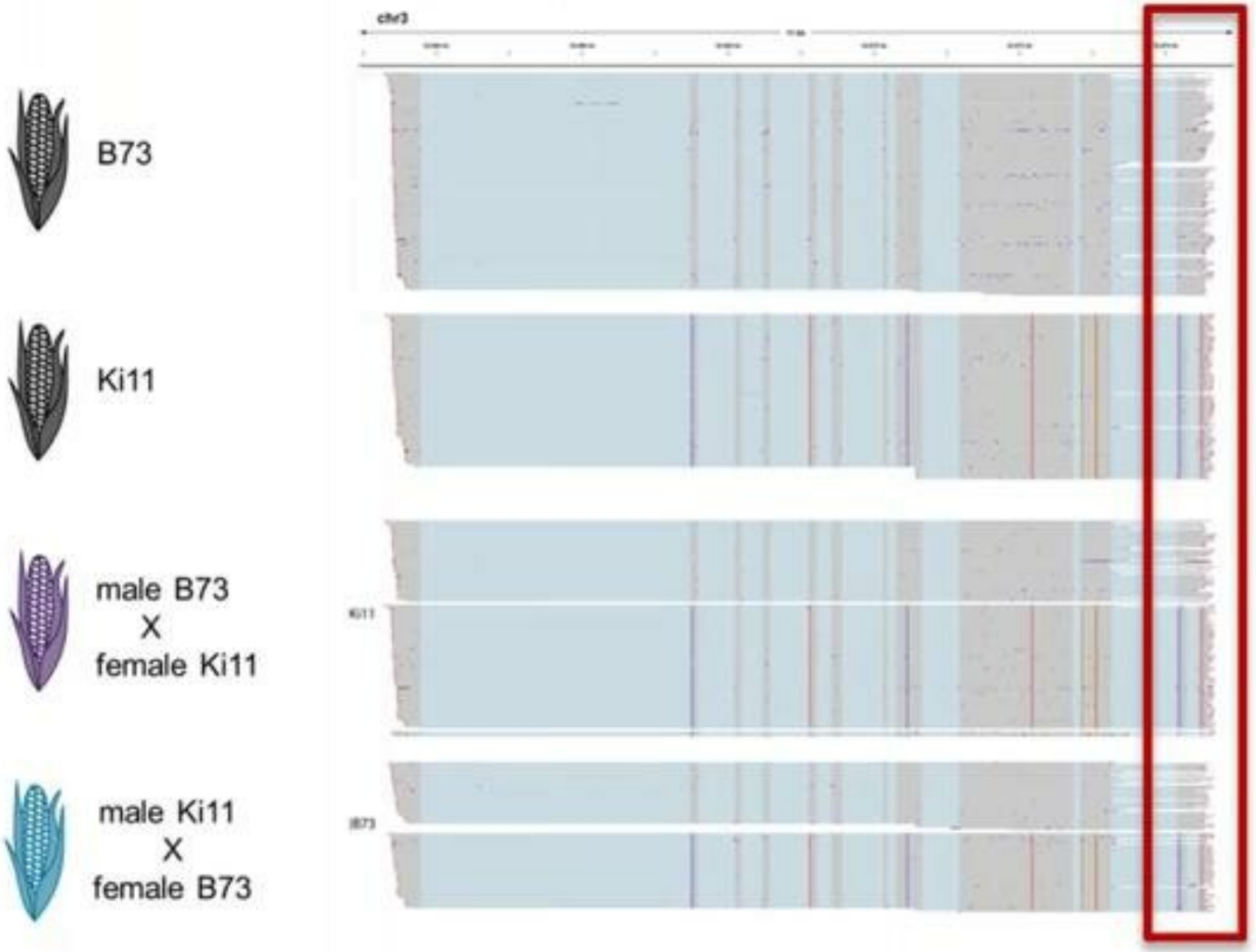


Parent B73 and Ki11 express different isoforms (3' exon difference)

- Two dominant isoforms PB.8517.4 and PB.8517.1
- PB.8517.4 is the canonical isoform and has 11 exons
- PB.8517.1 is a novel isoform with the last exon spliced
- B73 **only** expresses PB.8517.4 (**unspliced** 3' exon)
- Ki11 **only** expresses PB.8517.1 (**spliced** 3' exon)

Wang, B. et al. Variant Phasing and Haplotypic Expression from Single-molecule Long-read Sequencing in Maize. BioRxiv doi: https://doi.org/10.1101/654533

ISO-SEQ ANALYSIS OF ALLELE-SPECIFIC ISOFORM EXPRESSION IN MAIZE F1 HYBRID OFFSPRING

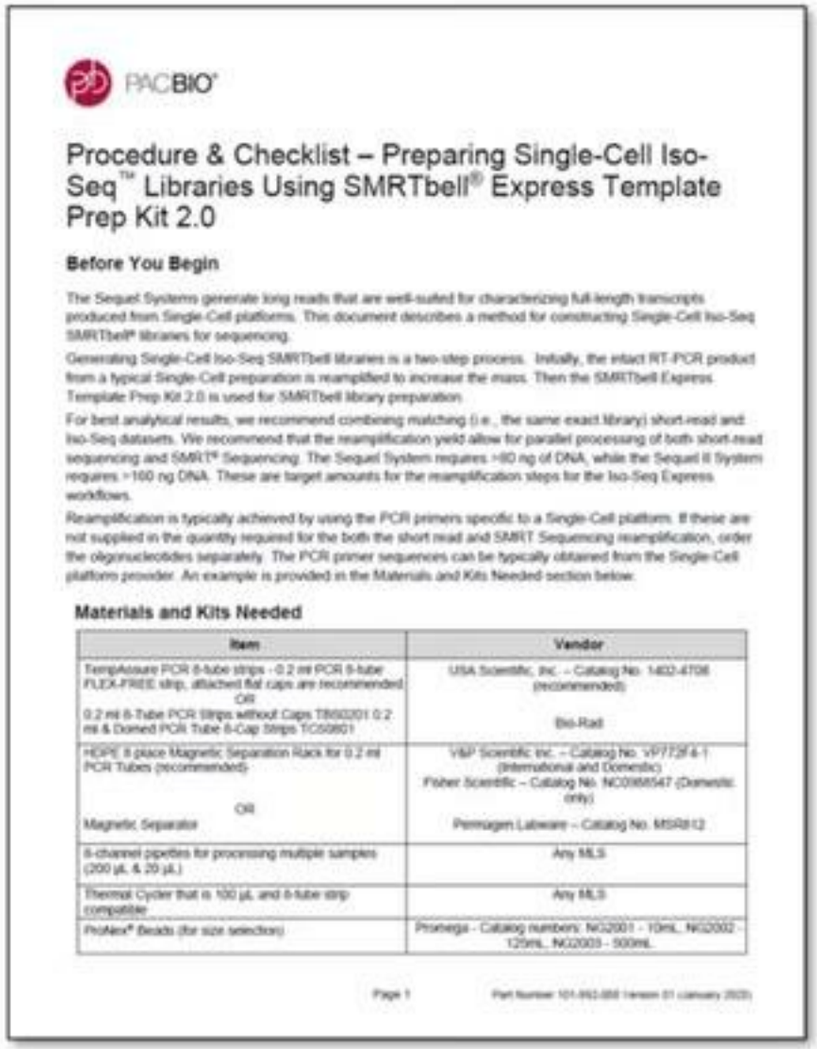


Both F1s inherit the allele-specific isoform expression

Wang, B. et al. Variant Phasing and Haplotypic Expression from Single-molecule Long-read Sequencing in Maize. BioRxiv doi: https://doi.org/10.1101/654533

PREPARING SINGLE-CELL ISO-SEQ LIBRARIES USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0

- Document (PN 101-892-000) describes a method to construct Single-Cell Iso-Seq SMRTbell libraries for sequencing. for sequencing on Sequel and Sequel II systems allowing detection of full-length transcripts up to 10 kb
- Generating Single-Cell Iso-Seq SMRTbell libraries is a two-step process. Initially, the intact RT-PCR product from a typical Single-Cell preparation is reamplified to increase the mass. Then the SMRTbell Express Template Prep Kit 2.0 is used for SMRTbell library preparation.
- Protocol document contains:
    1. General laboratory best practices
    2. Instructions for performing re-amplification of cDNA products from a Single-Cell preparation prior to SMRTbell library construction
    3. Instructions for constructing SMRTbell libraries using re-amplified cDNA products and SMRTbell Express Template Prep Kit 2.0



PACBIO®

Procedure & Checklist – Preparing Single-Cell Iso-Seq™ Libraries Using SMRTbell® Express Template Prep Kit 2.0

Before You Begin

The Sequel Systems generate long reads that are well-suited for characterizing full-length transcripts produced from Single-Cell platforms. This document describes a method for constructing Single-Cell Iso-Seq SMRTbell® libraries for sequencing.

Generating Single-Cell Iso-Seq SMRTbell libraries is a two-step process. Initially, the intact RT-PCR product from a typical Single-Cell preparation is reamplified to increase the mass. Then the SMRTbell Express Template Prep Kit 2.0 is used for SMRTbell library preparation.

For best analytical results, we recommend combining matching (i.e., the same exact library) short-read and Iso-Seq datasets. We recommend that the reamplification yield allow for parallel processing of both short-read sequencing and SMRT® Sequencing. The Sequel System requires >80 ng of DNA, while the Sequel II System requires >160 ng DNA. These are target amounts for the reamplification steps for the Iso-Seq Express workflows.

Reamplification is typically achieved by using the PCR primers specific to a Single-Cell platform. If these are not supplied in the quantity required for the both the short read and SMRT Sequencing reamplification, order the oligonucleotides separately. The PCR primer sequences can be typically obtained from the Single-Cell platform provider. An example is provided in the Materials and Kits Needed section below.

Materials and Kits Needed

Item	Vendor
TempAssure PCR 8-tube strips - 0.2 ml PCR 8-tube FLEX-FREE strip, attached flat caps are recommended OR 0.2 ml 8-Tube PCR Strips without Caps TS02201 0.2 ml 8-Domed PCR Tube 8-Cap Strips TCS0801	USA Scientific, Inc. – Catalog No. 1402-4708 (recommended) Bio-Rad
HDPE 8 place Magnetic Separation Rack for 0.2 ml PCR Tubes (recommended) OR Magnetic Separator	V&P Scientific Inc. – Catalog No. VP772F4-1 (International and Domestic) Fisher Scientific – Catalog No. NC0958547 (Domestic only) Permagen Labware – Catalog No. MSR812
8-channel pipettes for processing multiple samples (200 µL & 20 µL)	Any MLS
Thermal Cycler that is 100 µL and 8-tube strip compatible	Any MLS
ProNex® Beads (for size selection)	Promega - Catalog numbers: NG2001 - 10mL, NG2002 - 125mL, NG2003 - 500mL

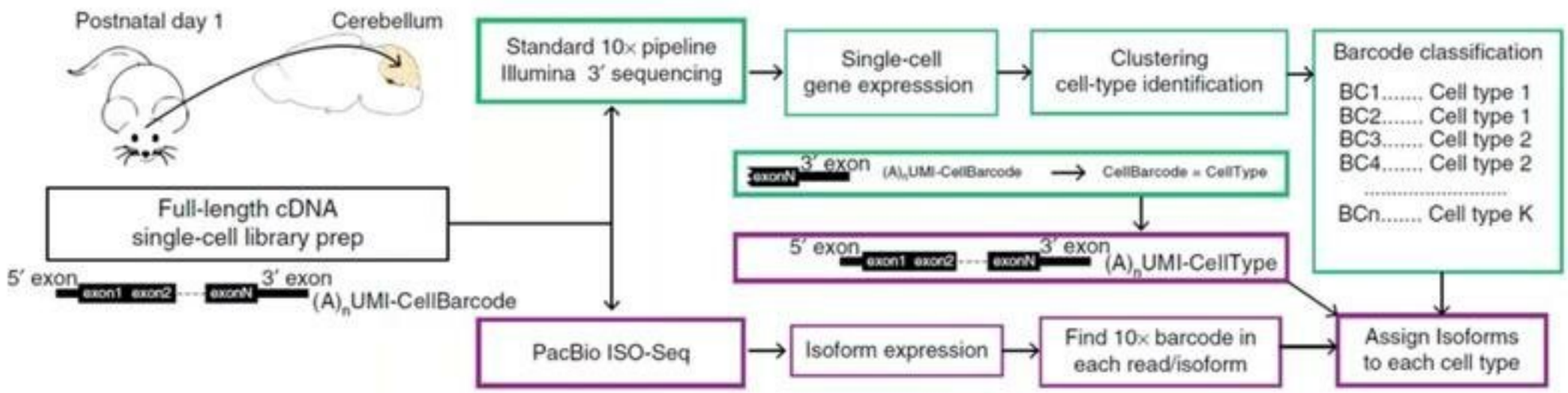
Page 1

https://www.pacb.com/support/documentation/

SINGLE-CELL ISO-SEQ ANALYSIS EXAMPLE

PACBIO

*"We used **ScISOr-Seq** to improve genome annotation in mouse Gencode v10 by determining the **cell-type-specific expression of 18,173 known and 16,872 novel isoforms**"*



Gupta et al. (2018) Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nature Biotechnology 36: 1197

Targeted Sequencing Applications

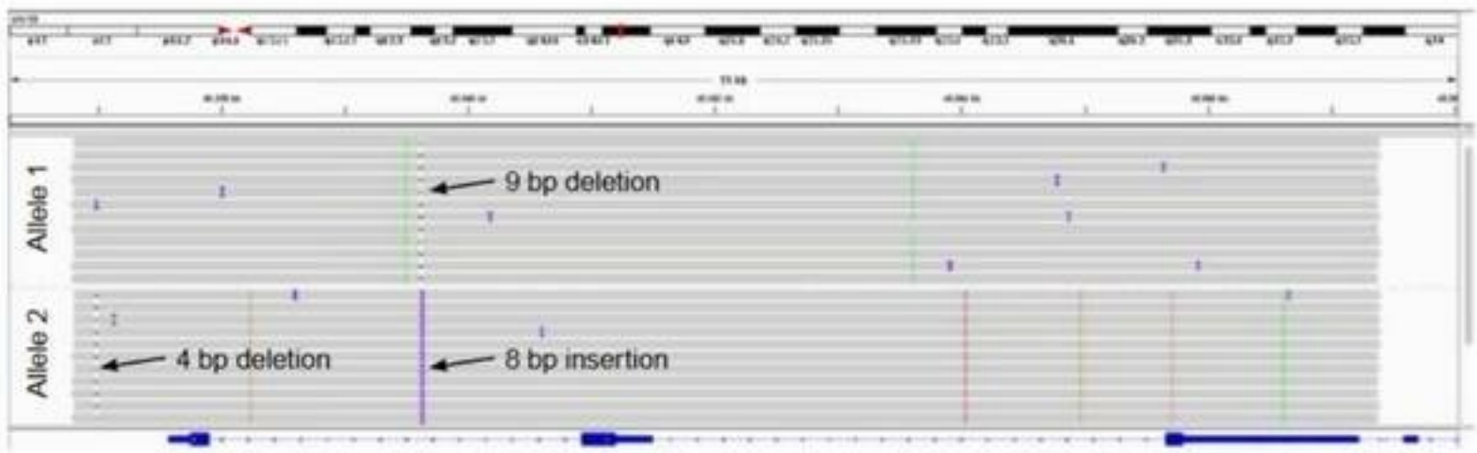
TARGETED SEQUENCING APPLICATIONS

Accurately discover and detect all variant types even in the hardest to reach regions of the genome

TARGETED SEQUENCING

Focus in on variation in even the most difficult regions of the genome

VARIANTS DETECTED IN PHASE ACROSS LONG AMPLICON



An Integrative Genomics Viewer image highlighting full-length reads and phased SNVs for a 9.6 kb gene covered by a single 10 kb amplicon. Sequencing results generated with Chemistry 3.0, SMRT Link v6.0 and a 20-hour movie collection time[8]. Data generated on the Sequel System.

SAMPLE PREPARATION RECOMMENDATIONS:

Start with high-quality, nucleic acids, as low as 250 ng for a 250 bp amplicon

Create SMRTbell templates from amplicons between 250 bp to 20 kb

Optimize throughput with flexible barcoding options:

Amplify PCR products using target-specific primers with incorporated barcodes

Add Barcoded Universal Primers into amplicons via a simple 2-step PCR process

Attach Barcoded Overhang Adapters during ligation without modifying existing primers

Multiplex up to 10,000 samples per SMRT Cell

Maximize output and turn-around-time with adjustable run parameters: For inserts <5 kb, recommend 10-hour movies, for inserts >5 kb, recommend 20-hour movies

Generate HiFi reads; Q20 single-molecule accuracy reads, up to 500,000 reads per SMRT Cell 1M, 4,000,000 reads per SMRT Cell 8M

Sequence to desired coverage based on project needs:

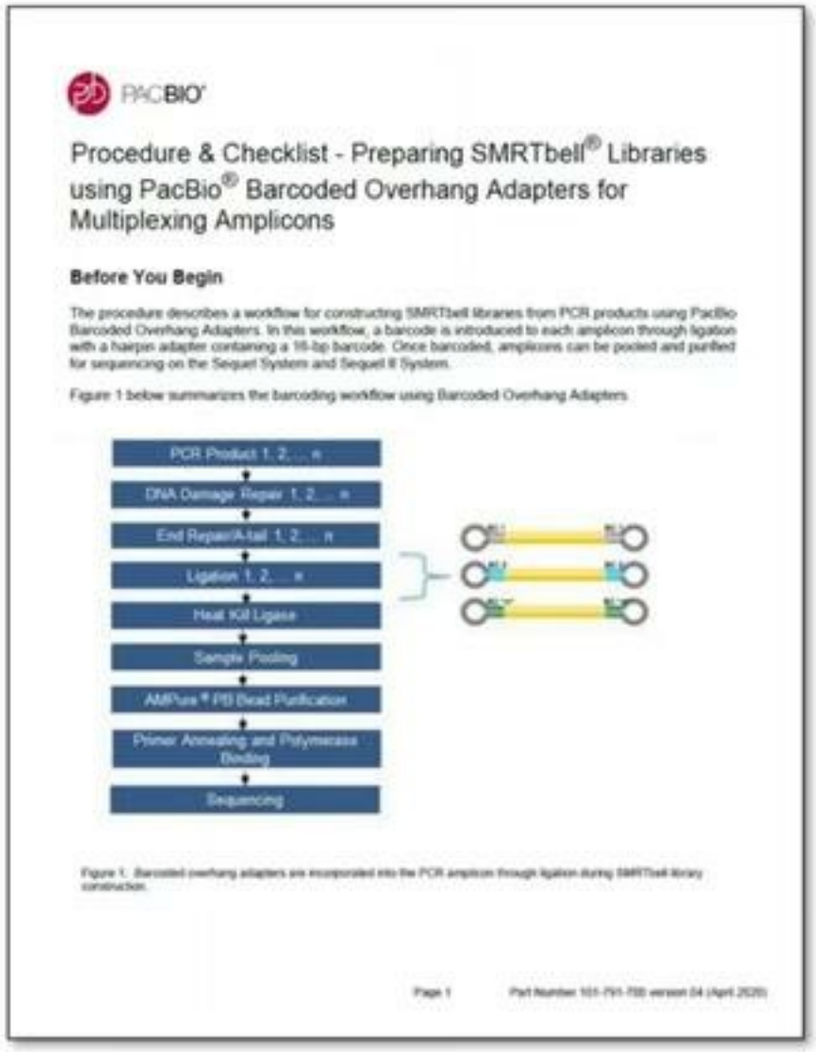
Target 30-fold coverage for variant detection

Increase coverage for minor variant detection (~6,000-fold coverage for 1% sensitivity)

PREPARING SMRTBELL LIBRARIES USING PACBIO BARCODED OVERHANG ADAPTERS FOR MULTIPLEX SMRT SEQUENCING

- Document (PN 101-791-700) describes a workflow for constructing SMRTbell libraries from PCR products using PacBio Barcoded Overhang Adapters.
- In this workflow, a barcode is introduced to each amplicon through ligation with a hairpin adapter containing a 16-bp barcode.
- Once barcoded, amplicons can be pooled and purified for sequencing on a single SMRT Cell
- Protocol document contains:

1. Best practices recommendations for generating high-quality PCR products for PacBio sequencing and input PCR amplicon QC
2. General recommendations for planning multiplexing experiments using Barcoded Overhang Adapters
3. Instructions for constructing barcoded SMRTbell libraries with input PCR amplicons using PacBio's SMRTbell Express Template Prep Kit 2.0 and Barcoded Overhang Adapter Kit 8A/8B



PACBIO

Procedure & Checklist - Preparing SMRTbell® Libraries using PacBio® Barcoded Overhang Adapters for Multiplexing Amplicons

Before You Begin

The procedure describes a workflow for constructing SMRTbell libraries from PCR products using PacBio Barcoded Overhang Adapters. In this workflow, a barcode is introduced to each amplicon through ligation with a hairpin adapter containing a 16-bp barcode. Once barcoded, amplicons can be pooled and purified for sequencing on the Sequel System and Sequel II System.

Figure 1 below summarizes the barcoding workflow using Barcoded Overhang Adapters

Figure 1. Barcoded overhang adapters are incorporated into the PCR amplicon through ligation during SMRTbell library construction.

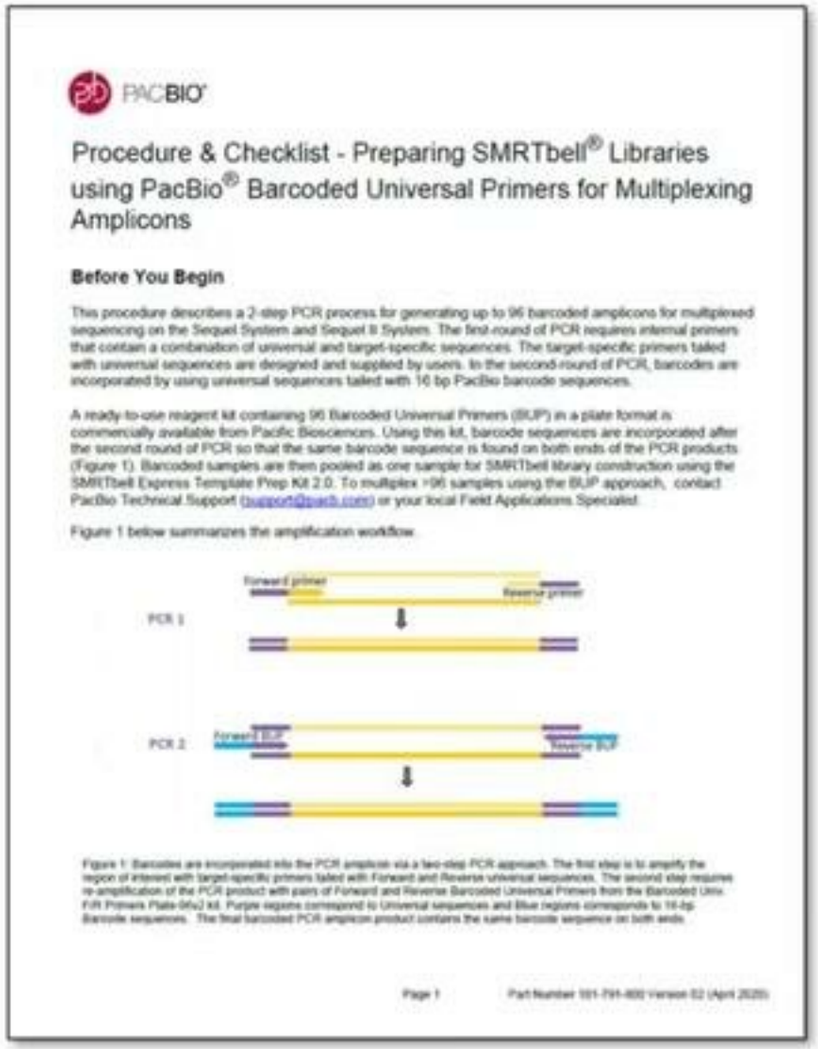
Page 1 Part Number 101-791-700 version 04 (April 2020)

https://www.pacb.com/support/documentation/

PREPARING SMRTBELL LIBRARIES USING PACBIO BARCODED UNIVERSAL PRIMERS FOR MULTIPLEX SMRT SEQUENCING

- Document (PN 101-791-800) describes a 2-step PCR method for generating up to 96 barcoded amplicons, with the first PCR step requiring internal primers containing a combination of universal and target-specific sequences
- In the second PCR step, barcodes are incorporated by using universal sequences tailed with 16-bp PacBio barcode sequences
- Once the PCR products are barcoded, they can be pooled into a single tube for SMRTbell library construction with the SMRTbell Express Template Prep Kit 2.0
- Protocol document contains:

1. Best practices recommendations for generating high-quality PCR products for PacBio sequencing
2. Instructions for generating barcoded amplicons using a 2-Step PCR method with PacBio's Barcoded Universal Primer Kit and performing equimolar pooling of barcoded amplicon products
3. Instructions for constructing SMRTbell libraries with pooled barcoded PCR amplicons using PacBio's SMRTbell Express Template Prep Kit 2.0



PACBIO

Procedure & Checklist - Preparing SMRTbell® Libraries using PacBio® Barcoded Universal Primers for Multiplexing Amplicons

Before You Begin

This procedure describes a 2-step PCR process for generating up to 96 barcoded amplicons for multiplexed sequencing on the Sequel System and Sequel II System. The first-round of PCR requires internal primers that contain a combination of universal and target-specific sequences. The target-specific primers tailed with universal sequences are designed and supplied by users. In the second-round of PCR, barcodes are incorporated by using universal sequences tailed with 16 bp PacBio barcode sequences.

A ready-to-use reagent kit containing 96 Barcoded Universal Primers (BUP) in a plate format is commercially available from Pacific Biosciences. Using this kit, barcode sequences are incorporated after the second round of PCR so that the same barcode sequence is found on both ends of the PCR products (Figure 1). Barcoded samples are then pooled as one sample for SMRTbell library construction using the SMRTbell Express Template Prep Kit 2.0. To multiplex >96 samples using the BUP approach, contact PacBio Technical Support (support@pacb.com) or your local Field Applications Specialist.

Figure 1 below summarizes the amplification workflow.

Forward primer
Reverse primer
PCR 1
PCR 2
Forward BUP
Reverse BUP

Figure 1: Barcodes are incorporated into the PCR amplicon via a two-step PCR approach. The first step is to amplify the region of interest with target-specific primers tailed with Forward and Reverse universal sequences. The second step requires re-amplification of the PCR product with pairs of Forward and Reverse Barcoded Universal Primers from the Barcoded Univ. F/R Primers Plate 96v2 kit. Purple regions correspond to Universal sequences and Blue regions corresponds to 16-bp Barcode sequences. The final barcoded PCR amplicon product contains the same barcode sequence on both ends.

Page 1 Part Number 101-791-800 Version 02 (April 2020)

https://www.pacb.com/support/documentation/

Microbial and Metagenomics Applications

MICROBIAL & METAGENOMICS APPLICATIONS

PACBIO

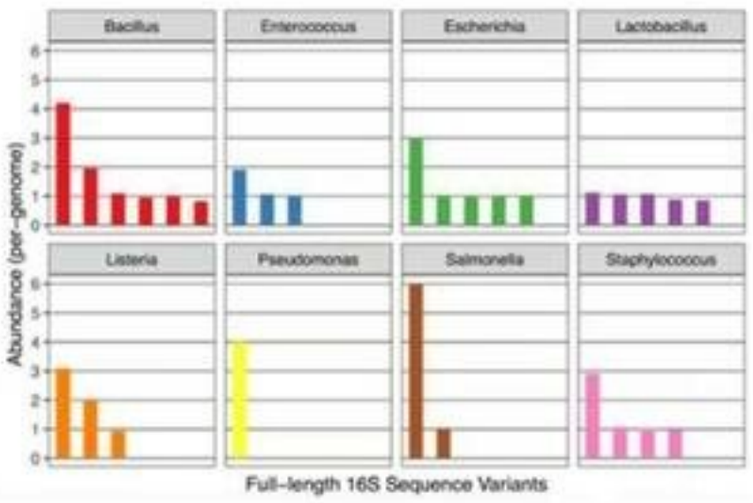
Long-read sequencing enables species-level resolution and drives functional insights with full-length 16S sequencing and 10 kb HiFi reads for metagenome profiling and assembly

WHOLE GENOME SEQUENCING | RNA SEQUENCING | TARGETED SEQUENCING | **COMPLEX POPULATIONS** | EPIGENETICS

COMPLEX POPULATIONS

Resolve closely related sequences within a heterogenous mixture

SMRT Sequencing generates the long, accurate, single-molecule reads you need to comprehensively characterize samples with complex variation



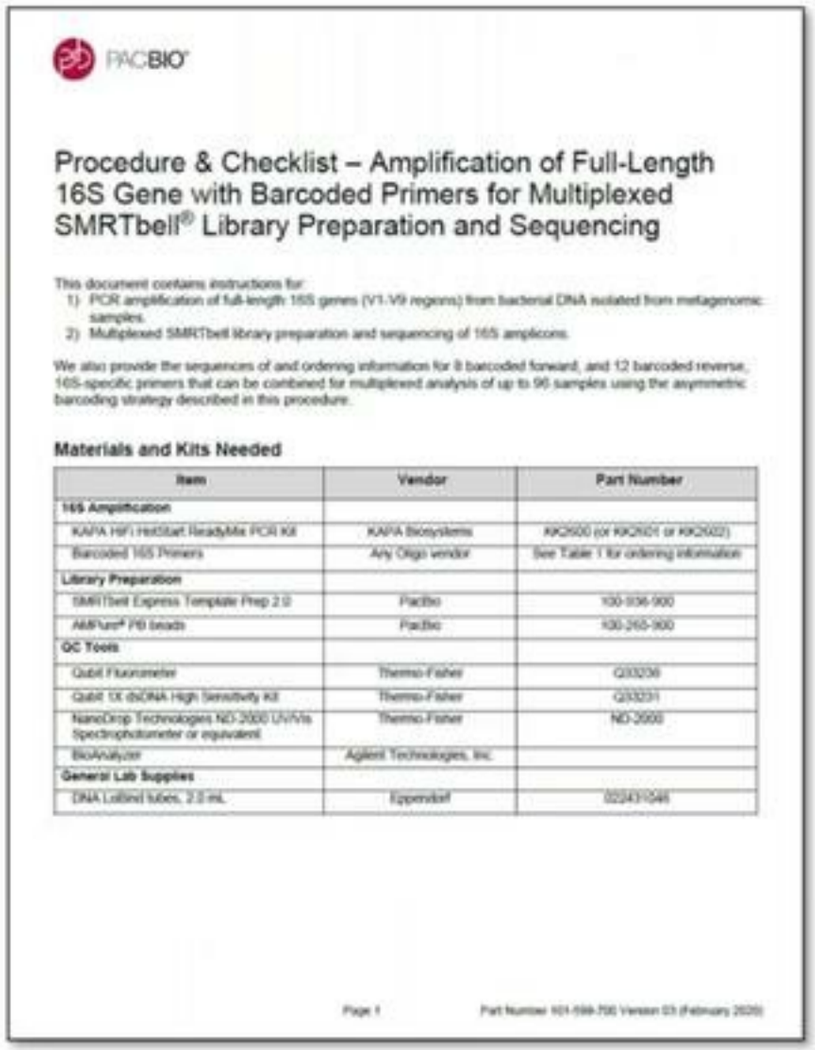
Frequencies of Full-Length 16S Variants Recovered from a Zymo Mock Community. Using DADA2 software, all copies of the 16S housekeeping gene can be recovered from microbial communities. 16S sequence variants from the same strain appear in the data in integer ratios that reflect their copy number in each genome. In some cases, this high-resolution information can be used to unambiguously identify not only the species but the strain present. *Nucleic Acids Res. 2019 47(18):e103.*

AMPLIFICATION OF FULL-LENGTH 16S GENE WITH BARCODED PRIMERS FOR MULTIPLEXED SMRTBELL LIBRARY PREPARATION AND SEQUENCING

- This document (PN 101-599-700) presents a workflow for amplifying full-length 16S genes from bacterial gDNA isolated from metagenomic samples and constructing multiplexed libraries using SMRTbell Express TPK 2.0 for sequencing on the Sequel and Sequel II Systems.

- Document also provides the sequences of and ordering information for 8 barcoded forward, and 12 barcoded reverse, 16S-specific primers that can be combined for multiplexed analysis of up to 96 samples using the asymmetric barcoding strategy described in this procedure.

- Protocol document contains:
  1. Recommendations for metagenomic DNA extraction QC and quantification
  2. Barcoded 16S Primer Sequences, ordering and storage Information
  3. Instructions for amplification of full-length 16S gene from bacterial gDNA extracted from metagenomic samples using barcoded primers in a single round of PCR
  4. Enzymatic steps for preparation of barcoded 16S SMRTbell libraries
  5. Sample setup guidance for preparing 16S SMRTbell libraries for sequencing on the Sequel and Sequel II Systems



PACBIO

Procedure & Checklist – Amplification of Full-Length 16S Gene with Barcoded Primers for Multiplexed SMRTbell® Library Preparation and Sequencing

This document contains instructions for:
1) PCR amplification of full-length 16S genes (V1-V9 regions) from bacterial DNA isolated from metagenomic samples.
2) Multiplexed SMRTbell library preparation and sequencing of 16S amplicons.

We also provide the sequences of and ordering information for 8 barcoded forward, and 12 barcoded reverse, 16S-specific primers that can be combined for multiplexed analysis of up to 96 samples using the asymmetric barcoding strategy described in this procedure.

Materials and Kits Needed

Item	Vendor	Part Number
16S Amplification		
KAPA HiFi HotStart ReadyMix PCR Kit	KAPA Biosystems	KK2500 (or KK2601 or KK2602)
Barcoded 16S Primers	Any Oligo vendor	See Table 1 for ordering information
Library Preparation		
SMRTbell Express Template Prep 2.0	PacBio	100-938-900
AMPure® PB beads	PacBio	100-265-900
QC Tools		
Qubit Fluorometer	Thermo Fisher	Q33226
Qubit 1X dsDNA High Sensitivity Kit	Thermo Fisher	Q33231
NanoDrop Technologies ND-2000 UV/Vis Spectrophotometer or equivalent	Thermo Fisher	ND-2000
BioAnalyzer	Agilent Technologies, Inc.	
General Lab Supplies		
DNA LoBind tubes, 2.0 mL	Eppendorf	022431048

Page 1 Part Number 101-599-700 Version 03 (February 2020)

https://www.pacb.com/support/documentation/

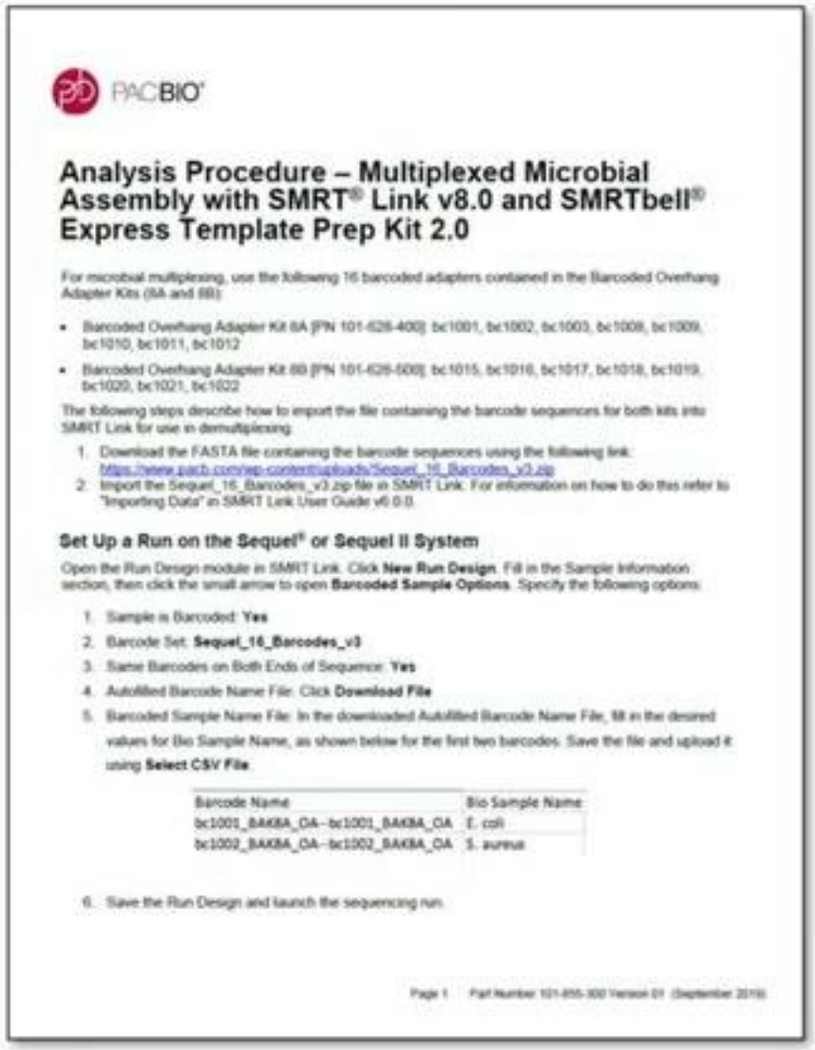
MICROBIAL WGS *DE NOVO* ASSEMBLY ANALYSIS APPLICATION IN SMRT LINK

PACBIO

A Dedicated Microbial Assembly Pipeline

- *New* : Chimera detection and filtering
- *New* : Plasmid sequence recovery and assembly
- *New* : Circular-aware polishing of contigs
- *New* : Circular rotation around the origin of replication
- *New* : Output annotated according to the NCBI guidelines

- Analysis document (PN 101-855-300) includes:
  1. How to import the FASTA file containing the barcode sequences for Barcoded Overhang Adapter Kit 8A/8B into SMRT Link
  2. How to set up a Run Design on the Sequel and Sequel II Systems to enable auto-demultiplexing of barcoded microbial sequencing data
  3. How to review the resulting demultiplexed microbial data sets using the SMRT Link Data Management module
  4. How to launch the SMRT Link Microbial Assembly analysis application on the microbial data sets
  5. Updated microbial assembly troubleshooting guidance



PACBIO

Analysis Procedure – Multiplexed Microbial Assembly with SMRT® Link v8.0 and SMRTbell® Express Template Prep Kit 2.0

For microbial multiplexing, use the following 16 barcoded adapters contained in the Barcoded Overhang Adapter Kits (8A and 8B):

- Barcoded Overhang Adapter Kit 8A [PN 101-628-400]: bc1001, bc1002, bc1003, bc1008, bc1009, bc1010, bc1011, bc1012
- Barcoded Overhang Adapter Kit 8B [PN 101-628-500]: bc1015, bc1016, bc1017, bc1018, bc1019, bc1020, bc1021, bc1022

The following steps describe how to import the file containing the barcode sequences for both kits into SMRT Link for use in demultiplexing.

1. Download the FASTA file containing the barcode sequences using the following link: https://www.pacb.com/wp-content/uploads/Sequel_16_Barcodes_v3.zip
2. Import the Sequel_16_Barcodes_v3.zip file in SMRT Link. For information on how to do this refer to "Importing Data" in SMRT Link User Guide v6.0.0.

Set Up a Run on the Sequel® or Sequel II System

Open the Run Design module in SMRT Link. Click **New Run Design**. Fill in the Sample Information section, then click the small arrow to open **Barcoded Sample Options**. Specify the following options:

1. Sample is Barcoded: **Yes**
2. Barcode Set: **Sequel_16_Barcodes_v3**
3. Same Barcodes on Both Ends of Sequence: **Yes**
4. AutoFilled Barcode Name File: Click **Download File**
5. Barcoded Sample Name File: In the downloaded AutoFilled Barcode Name File, fill in the desired values for Bio Sample Name, as shown below for the first two barcodes. Save the file and upload it using **Select CSV File**.

Barcode Name	Bio Sample Name
bc1001_BAK8A_OA--bc1001_BAK8A_OA	E. coli
bc1002_BAK8A_OA--bc1002_BAK8A_OA	S. aureus

6. Save the Run Design and launch the sequencing run.

Page 1 Part Number 101-855-300 Version 01 (September 2019)

https://www.pacb.com/support/documentation/

LONG-READ SEQUENCING PROVIDES HIGHER-RESOLUTION VIEWS OF BOTH METAGENOME MEMBERS AND FUNCTIONS

PACBIO

Full-Length 16S Sequencing Provides Species- and Strain-level Information

- FL 16S sequencing can provide strain-level identification of community members
- PacBio full-length 16S amplification & sequencing protocol available with recommended barcoded 16S primer sequences; *OR* use third-party all-in-one Shoreline Biome kits for DNA extraction, PCR amplification & data analysis

Long-Read Metagenomics Shotgun Profiling Reveals Intact Genes and Operons Without Assembly

- Circular consensus sequencing of 10 kb insert libraries delivers highly accurate HiFi (≥Q20 CCS) reads containing entire bacterial genes, operons, or gene clusters
- Community functions are revealed even without assembly
- Obtain high-confidence information even for low-abundance species

Metagenomics Shotgun Assembly From Long Reads Generates New References

- Resolve genomes of microbes that cannot be easily cultured
- Leverage epigenomic data to cluster contigs and plasmids from the same strain

16S SUB-REGION SEQUENCING SHOWS BIAS IN PROFILING TAXA


nature COMMUNICATIONS

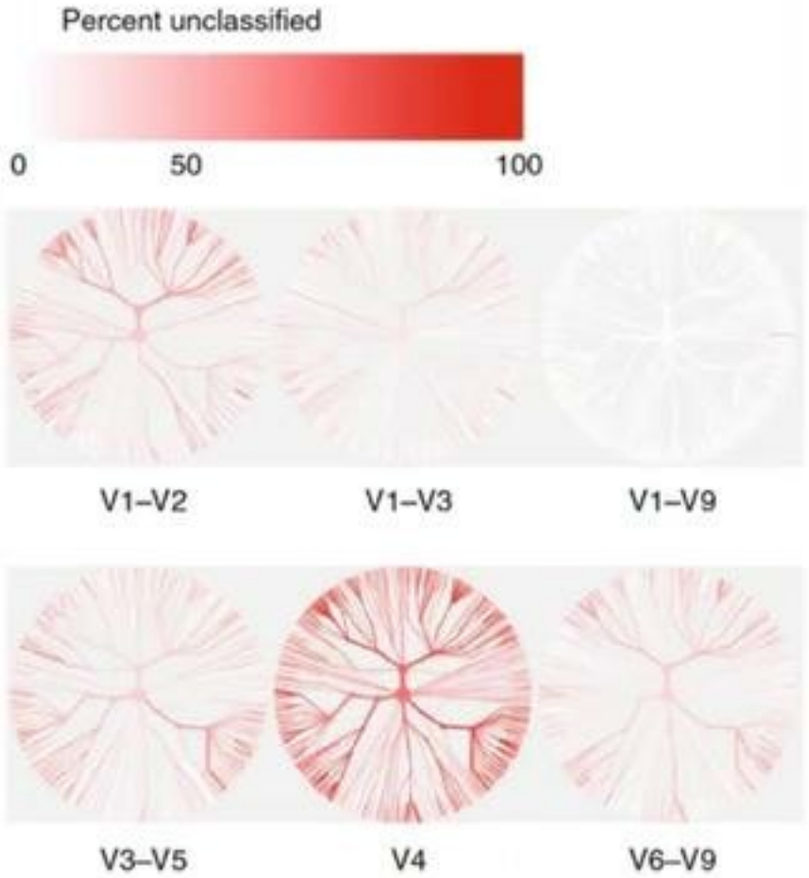
ARTICLE

OPEN

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

Jethro S. Johnson, Daniel J. Spakowicz, Bo-Young Hong, Lauren M. Petersen, Patrick Demkowicz, Lei Chen, Shana R. Leopold, Blake M. Hanson, Hanako O. Agresta, Mark Gerstein, Erica Sodergren & George M. Weinstock

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



- V4: Consistently poor performance
- V1–V2: poor for Proteobacteria
- V3–V5: poor for Actinobacteria

- V1–V3: good results for *Escherichia / Shigella*
- V3–V5: good results for *Klebsiella*,
- V6–V9: good results for *Clostridium* and *Staphylococcus*

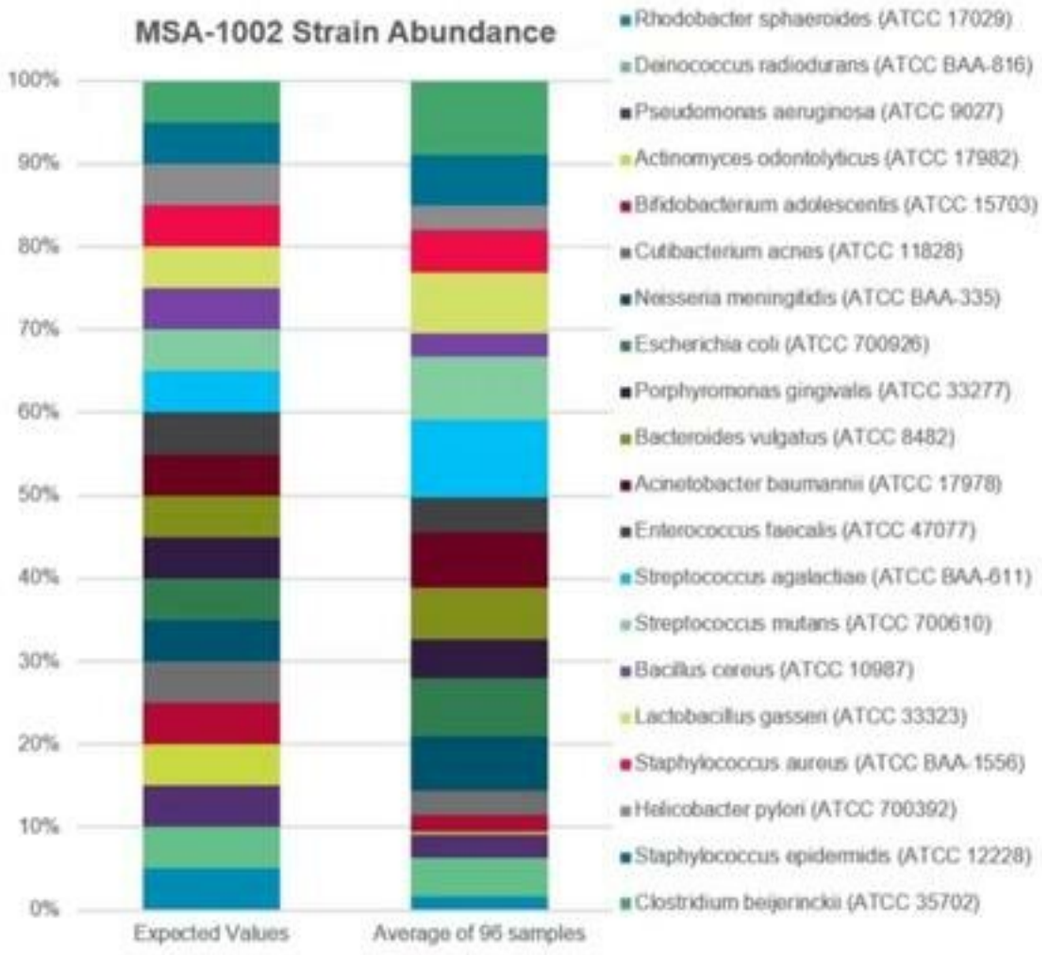
- **Full V1–V9 region: the only way to resolve ALL the clades that may be present in the human gut**

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

TAXONOMIC CLASSIFICATION RESULTS FOR 96-PLEX 16S LIBRARY SAMPLE (SEQUEL II SYSTEM)

PACBIO®

PacBio 16S Sequencing Faithfully Represents a Known Mock Community Sample



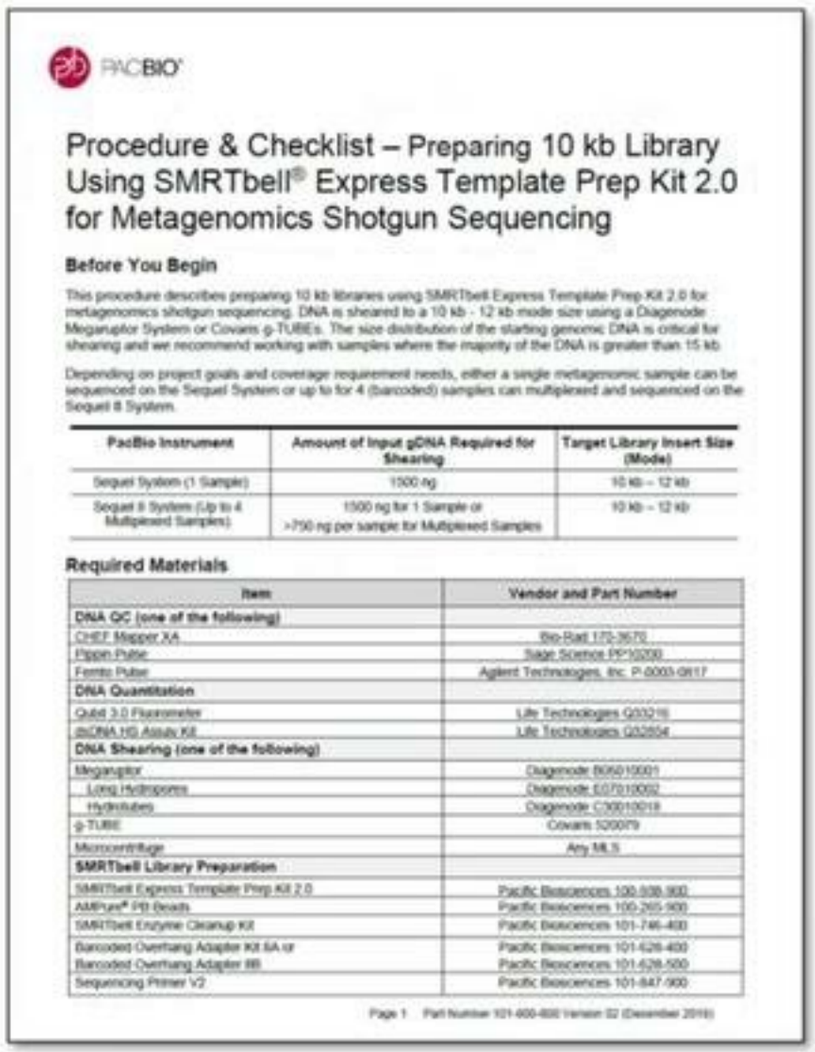
Full-length (V1-V9) 16S amplicon samples were pooled at 96-plex and sequenced on a <u>single</u> Sequel II SMRT Cell 8M

Download and explore this Sequel II System 16S data set

PROCEDURE & CHECKLIST

PREPARING 10 KB LIBRARY USING SMRTBELL EXPRESS TEMPLATE PREP KIT 2.0 FOR METAGENOMICS SHOTGUN SEQUENCING

- Workflow for preparing 10 kb libraries using SMRTbell Express Template Prep Kit 2.0 for metagenomics shotgun sequencing on the Sequel and Sequel II Systems
- Depending on project goals and coverage requirement needs, either a single metagenomic sample can be sequenced on the Sequel System or **up to for 4 (barcoded) samples can multiplexed** and sequenced on the **Sequel II System**.
- Protocol document contains:
  1. Recommendations for metagenomic DNA QC, shearing and quantification
  2. Enzymatic steps for preparation of non-barcoded and barcoded metagenomics shotgun SMRTbell libraries
  3. Instructions for size-selection of multiplexed metagenomics shotgun libraries using the AMPure PB Size Selection method
  4. Sample setup guidance for preparing metagenomics shotgun SMRTbell libraries for sequencing on the Sequel and Sequel II Systems



PACBIO

Procedure & Checklist – Preparing 10 kb Library Using SMRTbell® Express Template Prep Kit 2.0 for Metagenomics Shotgun Sequencing

Before You Begin

This procedure describes preparing 10 kb libraries using SMRTbell Express Template Prep Kit 2.0 for metagenomics shotgun sequencing. DNA is sheared to a 10 kb - 12 kb mode size using a Diagenode Megaruptor System or Covaris g-TUBEs. The size distribution of the starting genomic DNA is critical for shearing and we recommend working with samples where the majority of the DNA is greater than 15 kb.

Depending on project goals and coverage requirement needs, either a single metagenomic sample can be sequenced on the Sequel System or up to for 4 (barcoded) samples can multiplexed and sequenced on the Sequel II System.

PacBio Instrument	Amount of Input gDNA Required for Shearing	Target Library Insert Size (Mode)
Sequel System (1 Sample)	1000 ng	10 kb – 12 kb
Sequel II System (Up to 4 Multiplexed Samples)	1500 ng for 1 Sample or >750 ng per sample for Multiplexed Samples	10 kb – 12 kb

Required Materials

Item	Vendor and Part Number
DNA QC (one of the following)	
CHEF Mapper XA	Bio-Rad 170-3670
Pippin Pulse	Sage Science PP10200
Femto Pulse	Agilent Technologies, Inc. P-0003-0817
DNA Quantitation	
Qubit 3.0 Fluorometer	Life Technologies Q33216
dsDNA HS Assay Kit	Life Technologies Q32854
DNA Shearing (one of the following)	
Megaruptor	Diagenode B06010001
Long Hydropores	Diagenode E07010002
Hydrotubes	Diagenode C30010018
g-TUBE	Covaris 520079
Microcentrifuge	Any MLS
SMRTbell Library Preparation	
SMRTbell Express Template Prep Kit 2.0	Pacific Biosciences 100-938-900
AMPure® PB Beads	Pacific Biosciences 100-265-900
SMRTbell Enzyme Cleanup Kit	Pacific Biosciences 101-746-400
Barcoded Overhang Adapter Kit 8A or Barcoded Overhang Adapter 8B	Pacific Biosciences 101-628-400 Pacific Biosciences 101-628-500
Sequencing Primer V2	Pacific Biosciences 101-847-900

Page 1 Part Number 101-600-800 Version 02 (December 2019)

https://www.pacb.com/support/documentation/

HOW MANY METAGENOMICS SHOTGUN SAMPLES CAN BE MULTIPLEXED ON A SINGLE SEQUEL II SMRT CELL 8M?

PACBIO

The overall goals of the project will determine the needed coverage depth

Question 1: ***What is the estimated abundance of the rarest species you want to observe?***

Example: "I want to see species present at 1% abundance."

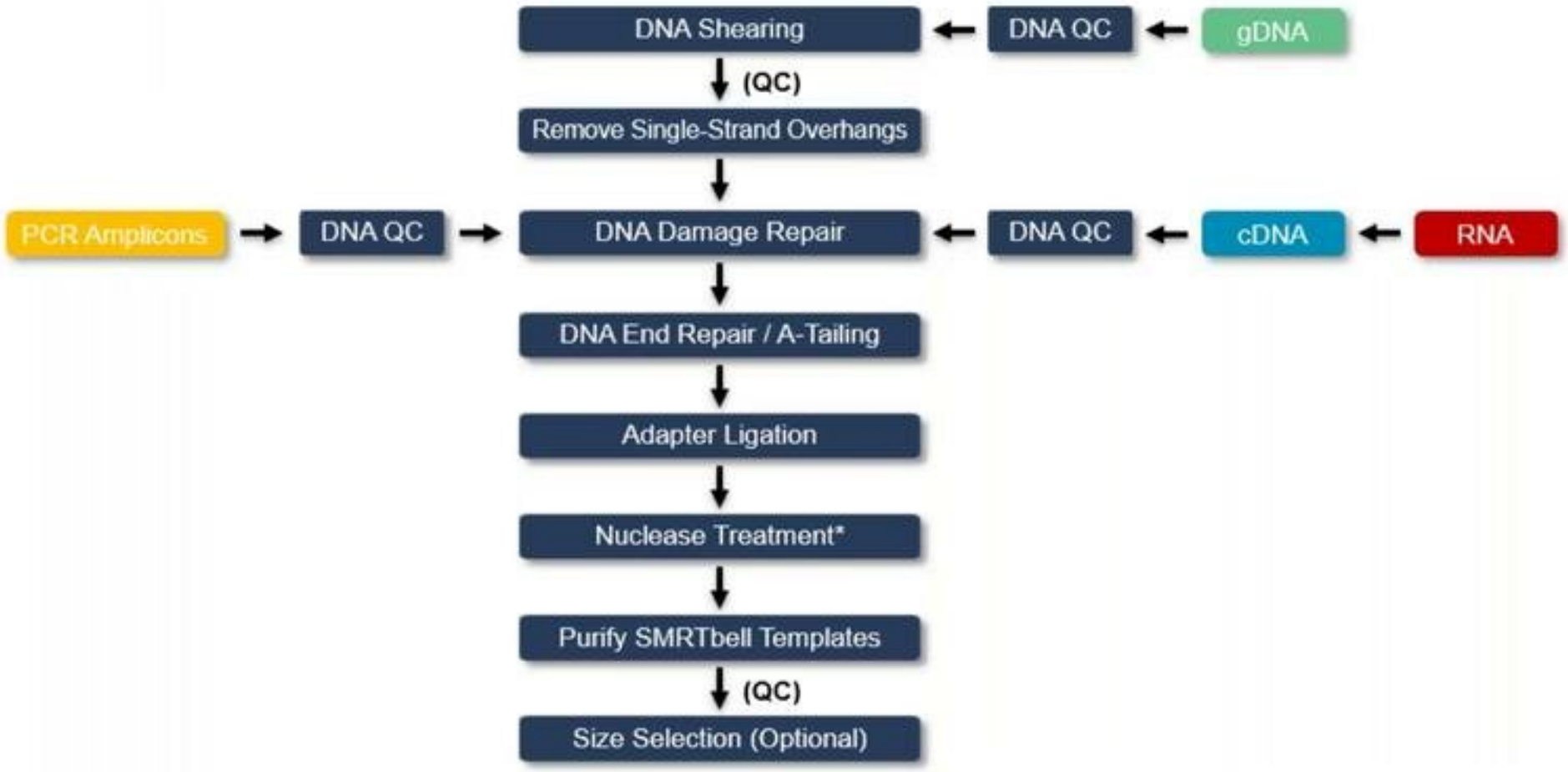
With 1 SMRT Cell 8M, you can expect ~2,400,000 x 0.01 = 24,000 HiFi (≥Q20) reads from a 1% abundant species with an 'average' genome size

Question 2: ***What is your goal?***

IN ORDER TO ACHIEVE...	...YOU NEED
Species detection	~100 reads
Comprehensive gene profiling / discovery*	5-Fold coverage; ~3,000 HiFi reads
Complete genome assembly*	20-Fold coverage; ~12,000 HiFi reads

* # Reads Needed = Coverage x 5 Mb Genome / 8.5 kb Median HiFi Read Length

PACBIO SEQENCING APPLICATIONS



WHOLE GENOME SEQUENCING

RNA SEQUENCING

TARGETED SEQUENCING

COMPLEX POPULATIONS

millennium science

PACBIO®

www.pacb.com

PACBIO® Sequencing services at the Australian Genome Research Facility

agrf Engineering Australian Genomics

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

BIOPLATFORMS AUSTRALIA

PACBIO®

1 PACBIO® Sequencing services at the Australian Genome Research Facility

2 Sequencing client prepared libraries

3 Other PacBio Services

4 Submitting to our service

5 Local Account Managers

PacBio-UQ@agrf.org.au

6 Your Operations Team

PACBIO® Sequencing services at the Australian Genome Research Facility

agrf
Empowering Australian Genomics

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

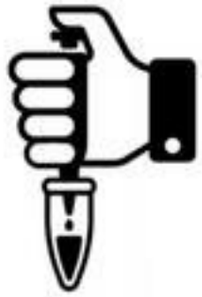
NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

BIOPLATFORMS AUSTRALIA

PACBIO®

Sequencing client prepared libraries

PACBIO CERTIFIED
SERVICE PROVIDER



We perform primer & polymerase binding

Library QC
Accurately QC HMW DNA using FEMTO *Pulse*

Sequence on the Sequel II
8x more data

Other PacBio Services

Library Prep

HMW DNA Extraction
Coming Soon

Bioinformatics

Submitting to our service

https://www.agrf.org.au/pacbio-uq

Click the

REQUEST A QUOTATION

Button

Submitting to our service

https://www.agrf.org.au/pacbio-uq

Click the

REQUEST A QUOTATION

Button

Local Account Managers

Dr Glenn D'Cunha

VIC, TAS & NZ

Anna Ikonomopoulos

Dr Leanne McGrath

SA & Internationals

Dr Oliver Distler

NSW / ACT

Saurabh Shrivastava

QLD / NT

Dr David Chandler

WA

PacBio-UQ@agrf.org.au

Your Operations Team

Dr David Hawkes

Trent Peters

Naga Kasinadhuni

Angelika Christ

Tim Bruxner

Stacey Anderson

PacBio-UQ@agrf.org.au

PacBio-UQ@agrf.org.au

https://www.agrf.org.au/pacbio-uq

millennium science®

PACBIO®

Thankyou for attending our workshop!!

Follow up

- Survey feedback
- Resources
- Recording of session
- Monthly update email

Events

- Telomere to Telomere conference: https://www.t2t-hprc-2020conference.com/

Virtual Event

PacBio Global Summit 2020

November 17, 2020–November 18, 2020