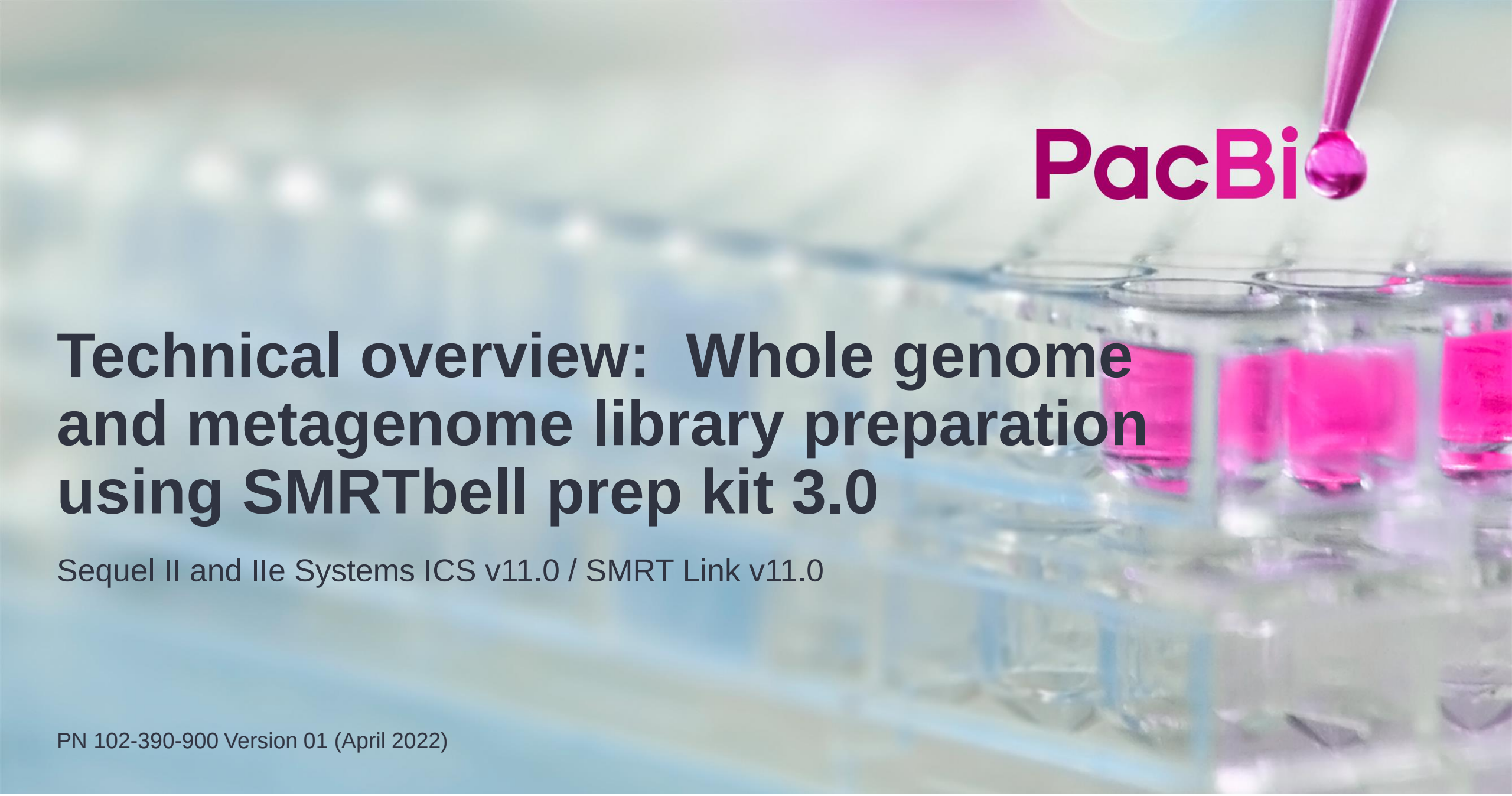


Technical overview: Whole genome and metagenome library preparation using SMRTbell prep kit 3.0

Sequel II and IIe Systems ICS v11.0 / SMRT Link v11.0

PN 102-390-900 Version 01 (April 2022)

Preparing whole genome and metagenome libraires using SMRTbell prep kit 3.0

Technical overview

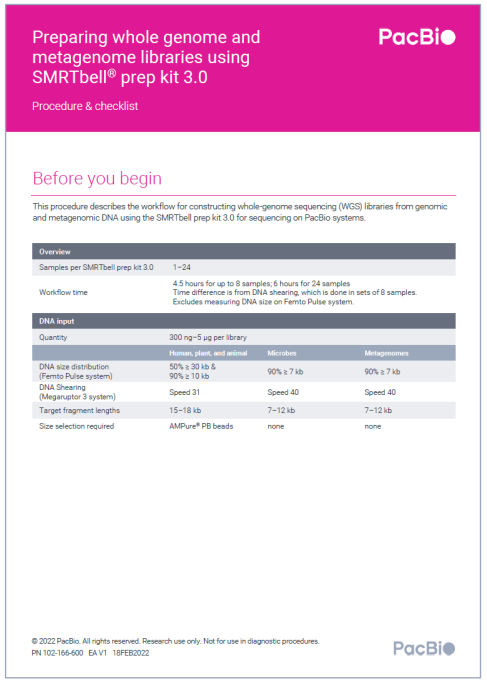
1. WGS library sample preparation workflow overview
2. WGS library sample preparation workflow details
3. WGS library sequencing preparation workflow overview
4. WGS data analysis recommendations for variant detection and *de novo* assembly applications
5. WGS library example sequencing performance data
6. Technical documentation & applications support resources
7. APPENDIX – Genomic DNA isolation & QC recommendations for PacBio WGS library sample preparation

Whole genome sequencing: How to get started

Application-specific Best practices guide → **Application-specific Procedure & checklist** → **Application-specific technical overviews** → **Library construction, sequencing & analysis**

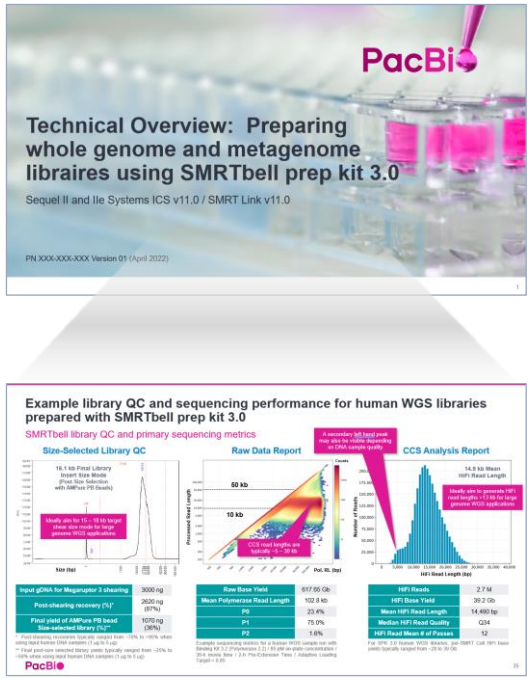
Application Briefs: WGS for *de novo* assembly – Best practices / Variant detection using WGS with HiFi Reads – Best practices

Summary overview of application-specific sample preparation and data analysis workflow recommendations



Procedure & Checklist – Preparing whole genome and metagenome libraires using SMRTbell prep kit 3.0 (102-166-600)

Technical documentation containing sample library construction and sequencing preparation protocol details



Technical overview: Whole genome and metagenome library preparation using SMRTbell prep kit 3.0 (102-390-900)

Technical overview presentations describe sample preparation details for constructing HiFi libraries for specific applications. Example sequencing performance data for a given application are also summarized.

Genomic DNA QC & shearing

≥1 µg per SMRT Cell 8M

Shear to 15 – 18 kb for large genomes

Shear to 10 – 12 kb for small genomes

Library construction (SMRTbell Prep Kit 3.0)

Multiplex WGS samples using SMRTbell barcoded adapter plate 3.0

Sequencing (Sequel II and IIe systems)

ABC* with Sequel II binding kit 3.2

15 hr or 30 hr movie collection time

Data analysis (SMRT Link or third-party)

Genome assembly

Variant detection

PacBio

* ABC = Anneal primer / Bind polymerase / Clean up bound complex

When is it appropriate to consider standard input, low DNA input or ultra-low DNA input workflows for whole genome sequencing applications?

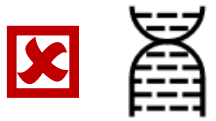
Sample & project considerations	Standard HiFi WGS (Large/small genomes & shotgun metagenomics)	Low DNA input WGS (2-plex)	Low DNA input WGS (Single sample)	Ultra-low DNA input WGS
Recommended Procedure & checklist	Preparing whole genome and metagenome libraires using SMRTbell prep kit 3.0 (102-166-600)	Preparing whole genome and metagenome libraires using SMRTbell prep kit 3.0 (102-166-600)	Preparing whole genome and metagenome libraires using SMRTbell prep kit 3.0 (102-166-600)	Preparing HiFi SMRTbell Libraries from Ultra-Low DNA Input (101-987-800)
Minimum DNA input	≥3 µg for a 3-Gb genome	300 ng for each genome	400 ng	5 ng
Amplification-based?	No	No	No	Yes
Genome size limit	N/A	600 Mb for each genome	1 Gb	500 Mb
Supported applications	*De novo* Assembly Human Variant Detection	*De novo* Assembly	*De novo* Assembly	*De novo* Assembly Human Variant Detection

Ultra-low DNA input: SUPPORTED APPLICATIONS

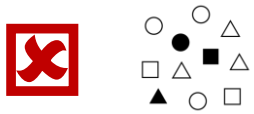
De novo assembly of insect/arthropod genomes (Up to 500 Mb)

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

Ultra-low DNA input: UNSUPPORTED APPLICATIONS



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



Metagenomics sequencing

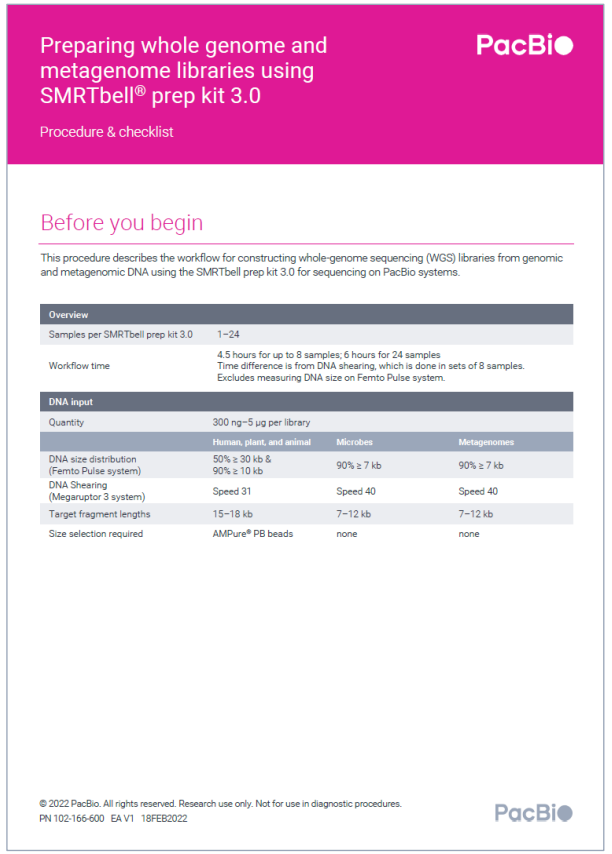
WGS library sample preparation workflow overview

WGS sample preparation procedure description

Procedure & Checklist – Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0 (102-166-600) describes a method for constructing SMRTbell libraries that are suitable for generating HiFi reads on the Sequel II and IIe Systems for WGS and metagenomic shotgun sequencing applications.

Procedure Highlights

- Uses SMRTbell Prep Kit 3.0 (102-182-70) and supports high-throughput processing using 500 ng – 5 µg of input genomic DNA amounts
  - We recommend starting with **≥1 µg of input DNA per SMRT Cell 8M** (or ~3 µg for up to a 3 Gb WGS sample to enable running 3 SMRT Cells 8M)
- Multiplexing of samples can be performed using SMRTbell barcoded adapter plate 3.0 (102-009-200)
- Recommend shearing high-quality gDNA using a Megaruptor 3 System (Diagenode)
  - 15 kb – 18 kb target insert size for large (plant / animal / human) genomes
  - 7 kb – 12 kb target insert size for small (microbial) genomes
  - 7 kb – 12 kb target insert size for shotgun metagenomic samples
- 4.5-hour workflow time to process up to 8 samples from shearing to size selection (6 hours for 24 samples)
  - Time difference is from DNA shearing, which can be performed in sets of 8 samples.
  - Excludes time needed for DNA sizing QC analysis using a Femto Pulse system.
- WGS SMRTbell libraries can be size-selected using AMPure PB Beads without the need for third-party equipment



Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0

Procedure & checklist

PacBio

Before you begin

This procedure describes the workflow for constructing whole-genome sequencing (WGS) libraries from genomic and metagenomic DNA using the SMRTbell prep kit 3.0 for sequencing on PacBio systems.

Overview			
Samples per SMRTbell prep kit 3.0	1–24		
Workflow time	4.5 hours for up to 8 samples; 6 hours for 24 samples Time difference is from DNA shearing, which is done in sets of 8 samples. Excludes measuring DNA size on Femto Pulse system.		
DNA input			
Quantity	300 ng–5 µg per library		
	Human, plant, and animal	Microbes	Metagenomes
DNA size distribution (Femto Pulse system)	50% ≥ 30 kb & 90% ≥ 10 kb	90% ≥ 7 kb	90% ≥ 7 kb
DNA Shearing (Megaruptor 3 system)	Speed 31	Speed 40	Speed 40
Target fragment lengths	15–18 kb	7–12 kb	7–12 kb
Size selection required	AMPure® PB beads	none	none

PN 102-166-600 EA V1 18FEB2022

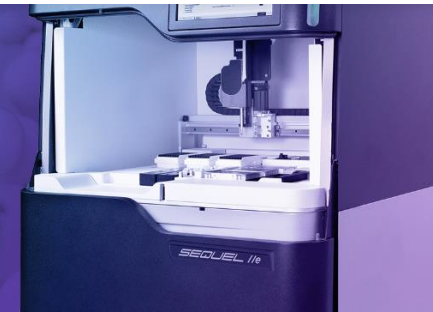
PacBio

PacBio Documentation (102-166-600)

APPLICATIONS

WHOLE GENOME SEQUENCING

De Novo assembly & variant detection
Microbial assembly
Shotgun metagenomics



WGS sample preparation & sequencing workflow overview

Workflow summary for constructing SMRTbell libraries suitable for sequencing on the Sequel II and IIe systems for WGS and metagenomic shotgun sequencing applications

Genomic DNA extraction, QC & shearing

- Perform DNA sample extraction using Circulomics Nanobind kits
- ≥1 µg of total input DNA per SMRT Cell 8M (for a single sample or across multiple samples when pooling)
- Large genome samples: Femto Pulse Genome Quality Number (GQN) at 30 kb ≥5.0
- Small genome samples: GQN at 7 kb ≥9.0
- Shear to target insert size with Megaruptor 3 system

SMRTbell library construction

- **Procedure & checklist – Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0** (102-166-600)
- Multiplex WGS samples using SMRTbell barcoded adapter plate 3.0 (102-009-200).
- Perform AMPure PB bead size selection for large genome WGS libraries or perform a standard cleanup using SMRTbell cleanup beads for small genome WGS libraries

Sequencing

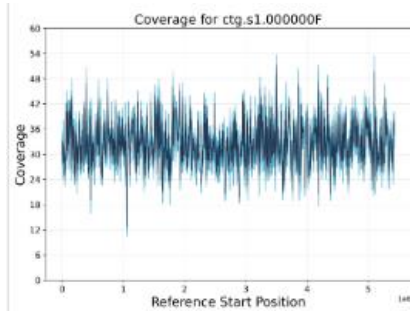
- Follow **SMRT Link Sample Setup** instructions for primer annealing, polymerase binding, complex cleanup and sample loading

HiFi Read

PacBio HiFi reads achieve 99.9% accuracy

Data analysis

- For *de novo* assembly, can use SMRT Link Genome Assembly or other third-party software
- For variant detection, can use DeepVariant for small variants <20 bp and SMRT Link PBSV for larger variants >20 bp



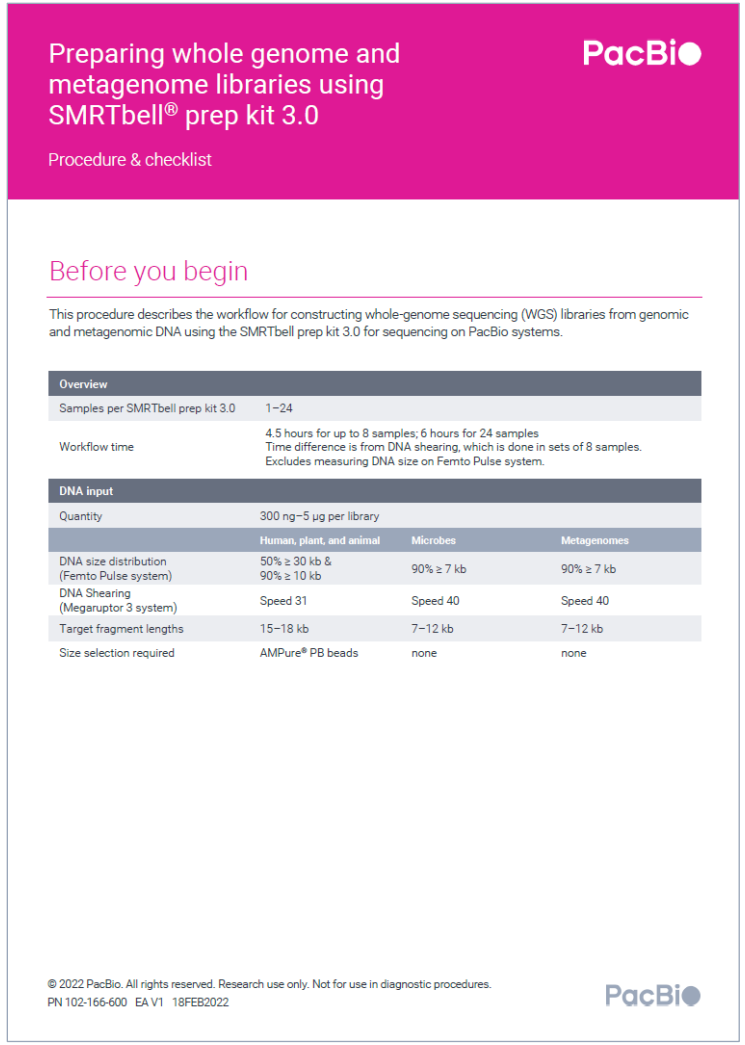
WGS library sample preparation workflow details

Procedure & checklist – Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0

Procedure & checklist 102-166-600 describes a method for constructing SMRTbell libraries using SMRTbell prep kit 3.0 that are suitable for generating high-accuracy long reads on the Sequel II and IIe systems for whole genome sequencing (*de novo* assembly, variant detection, microbial multiplexing) and shotgun metagenomic sequencing

Protocol Contents

1. Recommendations for gDNA quantification and sizing QC.
2. Recommendations for shearing gDNA to the desired target mode size using the Megaruptor 3 system (Diagenode).
3. Enzymatic steps for preparation of a WGS SMRTbell library using SMRTbell prep Kit 3.0 (102-182-700). (Instructions for preparing multiplexed samples using SMRTbell barcoded adapter plate 3.0 (102-009-200) are also provided.)
4. Instructions for size-selection of WGS SMRTbell libraries using AMPure PB bead size selection. (Size selection is not required for microbial WGS and metagenomic shotgun libraries where retention of shorter fragments is desired.)
5. Guidance for pooling barcoded WGS SMRTbell libraries for multiplexed sequencing on a single SMRT Cell.



Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0

PacBio

Procedure & checklist

Before you begin

This procedure describes the workflow for constructing whole-genome sequencing (WGS) libraries from genomic and metagenomic DNA using the SMRTbell prep kit 3.0 for sequencing on PacBio systems.

Overview			
Samples per SMRTbell prep kit 3.0	1–24		
Workflow time	4.5 hours for up to 8 samples; 6 hours for 24 samples Time difference is from DNA shearing, which is done in sets of 8 samples. Excludes measuring DNA size on Femto Pulse system.		
DNA input			
Quantity	300 ng–5 µg per library		
	Human, plant, and animal	**Microbes**	**Metagenomes**
DNA size distribution (Femto Pulse system)	50% ≥ 30 kb & 90% ≥ 10 kb	90% ≥ 7 kb	90% ≥ 7 kb
DNA Shearing (Megaruptor 3 system)	Speed 31	Speed 40	Speed 40
Target fragment lengths	15–18 kb	7–12 kb	7–12 kb
Size selection required	AMPure® PB beads	none	none

PN 102-166-600 EA V1 18FEB2022

PacBio

PacBio Documentation (102-166-600)

SPK 3.0 WGS library prep and sequencing workflow timing overview

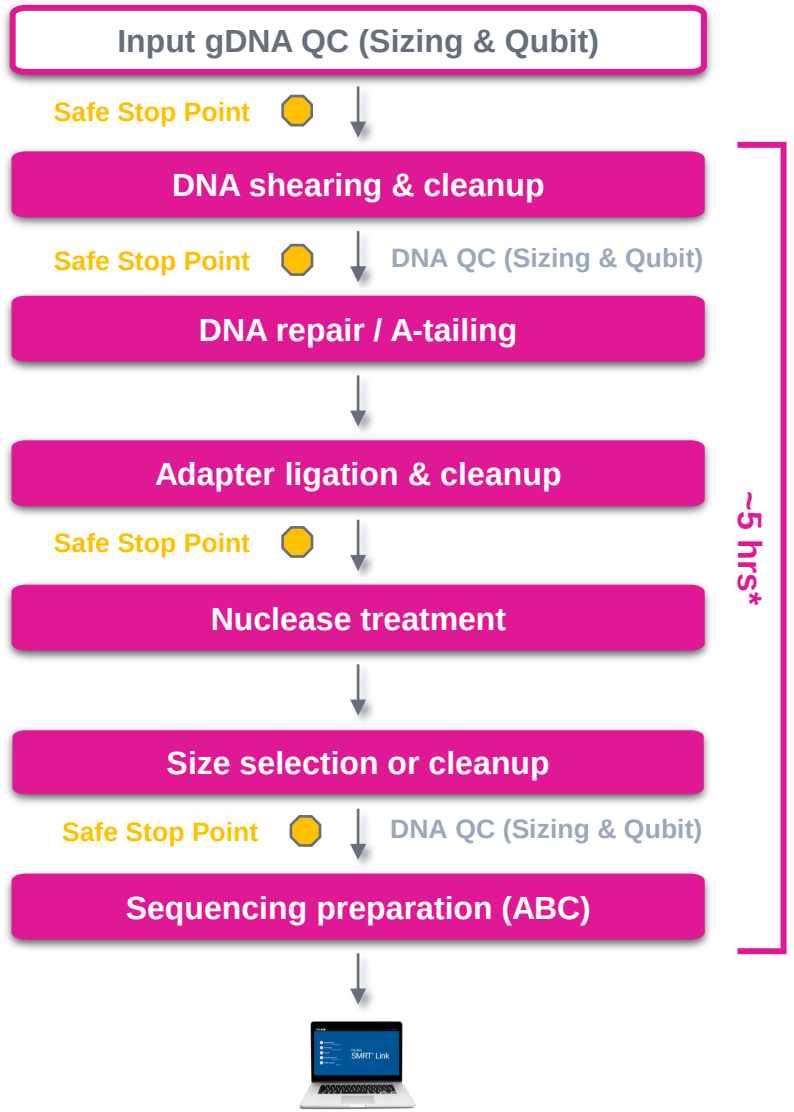
Go from DNA shearing to sequencing preparation in 1 day with SPK 3.0

Workflow step		Hands-on (min)	Walk-away (min)
DNA shearing*	DNA shearing (1 Megaruptor 3 cycle)	5	45
	1.0X SMRTbell bead cleanup	5	20
SMRTbell library construction*	DNA repair / a-tailing	5	35
	Adapter ligation (barcoded or non-barcoded adapter)	5	30
	1.0X SMRTbell bead cleanup	5	20
	Nuclease treatment	5	15
	AMPure PB bead size selection	10	30
Total		**40 min**	**3.3 hrs**

Workflow step		Hands-on (min)	Walk-away (min)
Sequencing preparation (ABC)	Primer annealing (Sequel II primer 3.2)	5	15
	Polymerase binding (Sequel II binding kit 3.2)	5	15
	Complex cleanup (1.2X SMRTbell cleanup beads)	5	20
Total		**15 min**	**0.83 hrs**

SPK 3.0 WGS Workflow

(102-166-600)



* **Excludes** additional time required to perform DNA sizing QC using Femto Pulse and DNA concentration QC

Input genomic DNA QC recommendations for WGS library construction using SMRTbell Prep Kit 3.0

- WGS library construction using SMRTbell prep kit 3.0 requires **high-quality, high-molecular weight genomic DNA***.
- Prior to library preparation, evaluate the **quantity** and **size distribution** of the input gDNA to determine whether it is suitable for the protocol.
- For each input gDNA sample:
  - ❑ Measure concentration and total mass of DNA with a **Qubit High Sensitivity dsDNA Assay** system (Thermo Fisher Scientific)
  - ❑ Measure DNA size distribution with a **Femto Pulse** system (Agilent)
  - ❑ Proceed with SMRTbell library construction if the **gDNA sample quality** is acceptable (see Table below)

Sample type	Input DNA metric	Requirement	Notes
All	**Per Library**	300 ng – 5 µg	• Starting with low DNA input amounts approaching ~300 ng may in some cases produce insufficient amounts of SMRTbell library to load at concentrations that optimize sequencing data yield. • For multiplexing applications, generally aim to use ≥300 ng of DNA input per sample, with a total mass ≥1 µg across all samples
All	**Per SMRT Cell 8M**	≥1 µg	• Start with ≥1 µg of total input DNA per SMRT Cell 8M (for a single sample or across multiple samples when pooling) to enable generation of sufficient library to load at concentrations that optimize sequencing data yield.
Large genome (Animal/plant/human)	**Longer than 30 kb**	≥50%	• Required to achieve target fragment lengths after DNA shearing. • For large genome samples, the Femto Pulse Genome Quality Number (GQN) at 30 kb should be ≥5.0. (Not applicable to microbial and metagenomic samples)
Large genome (Animal/plant/human)	**Longer than 10 kb**	≥90%	• Required for effective AMPure PB bead size selection. • For large genome samples, the GQN at 10 kb should be ≥9.0.
Small genome (microbial/metagenomic)	**Longer than 7 kb**	≥90%	• For microbial and metagenomic samples, the input DNA should be at least as large as the recommended insert lengths of 7–12 kb with a GQN at 7 kb ≥9.0. • Any degradation present should be due to shearing from the extraction process (e.g., bead beating) and not from poor sample handling or storage, or biochemical processes

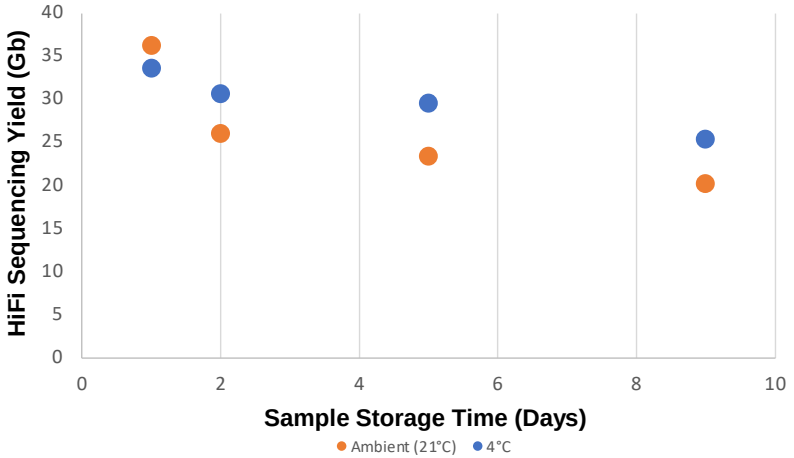
* See Technical note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

DNA extraction tech note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

Provides best practices for handling human whole blood samples* to generate optimal sequencing performance

Technical note 102-326-500 Discusses the effect of anticoagulant, sample storage time, storage conditions, and white blood cell count on the sequencing performance of DNA extracted using the Nanobind CBB Big DNA Kit (NB-900-001-01)

Nanobind CBB Big DNA Kit (NB-900-001-01) for isolating HMW DNA from cells, bacteria, & blood.



For optimal HiFi yield and read length performance, store human whole blood samples for **fewer than 2 days** at 4°C.

Stage	Variable	Best practice for PacBio HiFi sequencing
Before DNA extraction	Sample type	Human whole blood
	Anticoagulant	Potassium EDTA (K_2EDTA)
	Sample storage temperature	4 ± 3°C
	Sample storage time	≤ 2 days from collection to extraction
DNA extraction	Volume of whole blood	200 µL
	White blood cell (WBC) count	≥ 4 × 10^6 cells/mL for ≥ 3 µg of DNA
	DNA extraction method	Nanobind CBB Big DNA kit
After DNA extraction	DNA storage	Rest 1 day at ambient temperature, then store at 4 ± 3°C
	DNA size distribution	• 90% of DNA ≥ 10 kb (genomic quality number at 10 kb ≥ 9.0) • 50% of DNA ≥ 30 kb (genomic quality number at 30 kb ≥ 5.0)
	UV absorbance	• A260/280 nm ≥ 1.7 • A260/230 nm ≥ 1.5

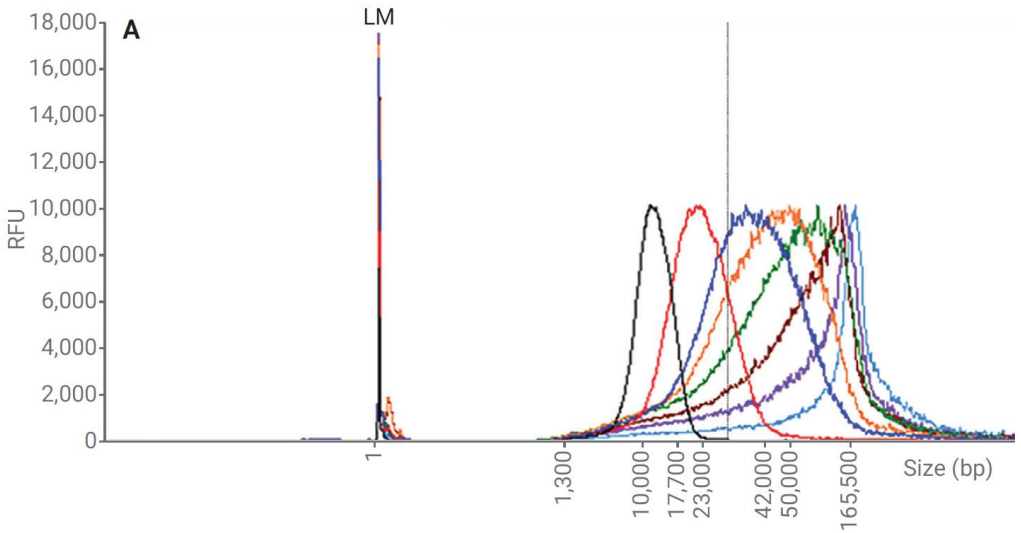
* For general DNA extraction recommendations see Technical note: Preparing DNA for PacBio HiFi sequencing – Extraction and quality control (102-193-651)

The Femto Pulse System is recommended for DNA sizing QC of genomic DNA for WGS applications

- Femto Pulse System (Agilent) is highly recommended for DNA sizing QC of input genomic DNA and SMRTbell libraries
  - Enables sizing of gDNA samples ranging from 1,300 bp to 165 kb
  - Requires <1 ng of sample DNA
  - Can analyze up to 12 samples in <1.5 hours
- The Femto Pulse system can be used in place of traditional pulse-field gel electrophoresis (PFGE) to quickly assess the initial integrity of genomic DNA, evaluate shears, determine appropriate size-selection thresholds, and conduct final QC before preparing libraries for SMRT Sequencing

Femto Pulse outputs quality metrics such as the **Genomic Quality Number (GQN)*** to quickly score the integrity of HMW gDNA

Ave. Smear Size (bp)	GQN Set at 30 kb
12,147	0
23,339	1.5
45,304	6.4
57,789	7.1
73,267	7.8
94,045	7.8
109,968	8.2
164,292	8.8



Femto Pulse System

Femto Pulse system offers a simplified QC workflow to generate SMRTbell libraries for WGS sequencing in **reduced time, and conserves sample** by using femtogram ranges of input DNA

* See Agilent Application Note – Quality Metrics for Nucleic Acids with the Agilent Fragment Analyzer and Femto Pulse Systems (5994-0521EN)

The Megaruptor 3 system is recommended for shearing genomic DNA For WGS applications

- Megaruptor 3 system (Diagenode) is highly recommended for DNA shearing*
  - Up to 8 samples can be sheared in parallel in ~45 minutes for high-throughput applications
  - Achieving the same size distribution across multiple samples provides more consistent sequencing performance
- Recommended library insert size distributions and Megaruptor 3 shear speed settings to use for different WGS applications are summarized on Page 7 in the procedure
  - Bring input gDNA to a final volume of **100 – 130 µL** with **Low TE buffer** [10 mM Tris-HCl (pH 8.0) + 0.1 mM EDTA] to target a DNA concentration of **3 – 39 ng/µL (ideal: 30 ng/µL)**
  - Perform shearing (**1 cycle**) using the conditions described in the table below

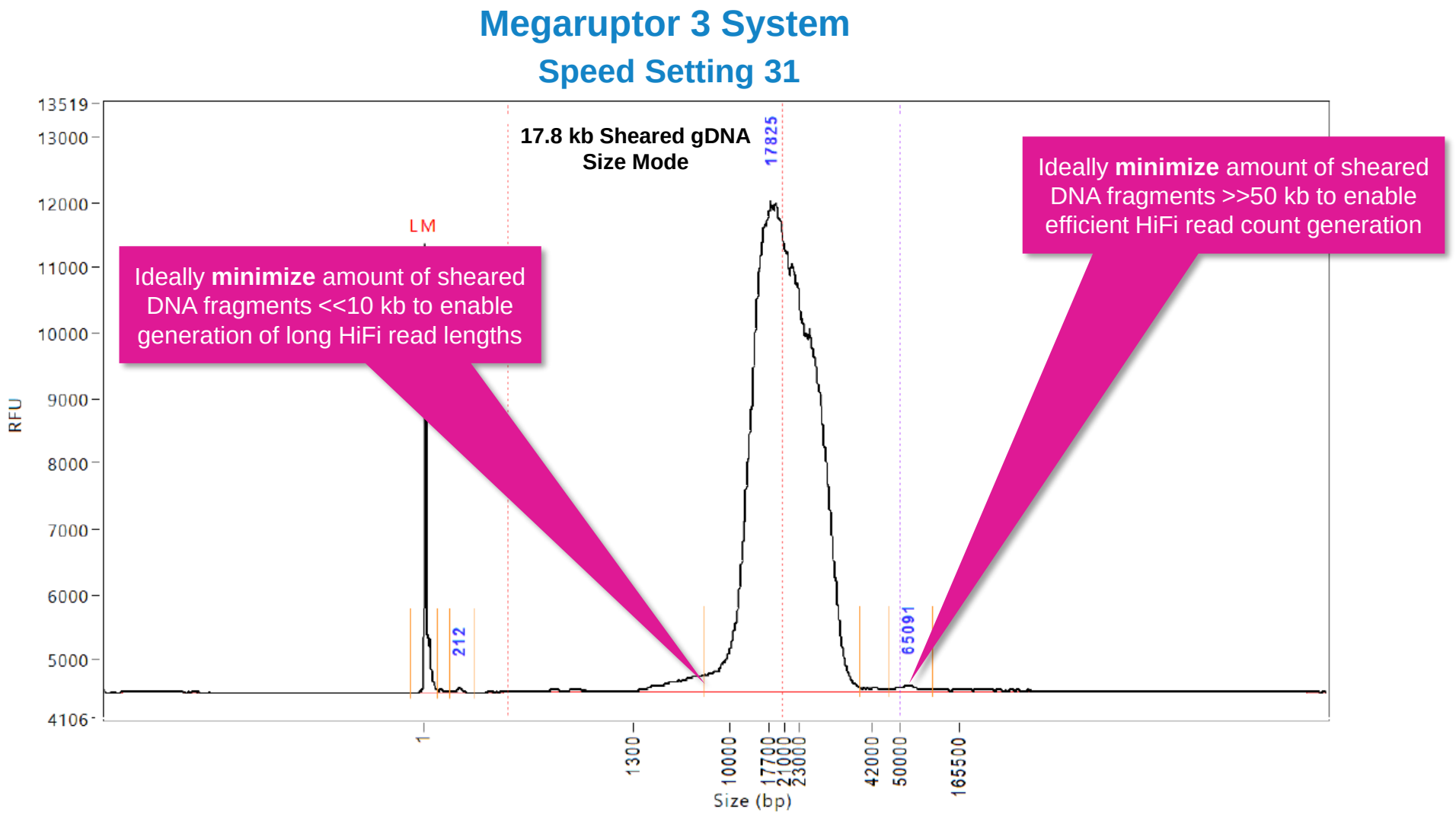
Megaruptor 3 System

Application	Recommended Library Insert Size (Mode)	Recommended Megaruptor 3 Shear Speed Setting
Animal / plant / human WGS	15 kb - 18 kb	31
Microbial WGS or shotgun metagenomics	7 kb - 12 kb	40

Because the response of individual gDNA samples can differ, **optimization of shearing conditions** may be needed to achieve the desired fragment distribution

*** Note:** The g-TUBE (Covaris) device generates a broader DNA fragment size-distribution compared to the Megaruptor 3 system. As a result, HiFi read quality and overall HiFi data yield may be reduced due to the residual presence of very large DNA fragments generated by g-TUBEs. For additional guidance, see Technical Note: Covaris g-TUBE DNA Shearing for SMRTbell Prep Kit 3.0 (102-326-501) or contact PacBio Technical Support or your local Field Applications Scientist.

Example Megaruptor 3 shearing results for a human genomic DNA sample

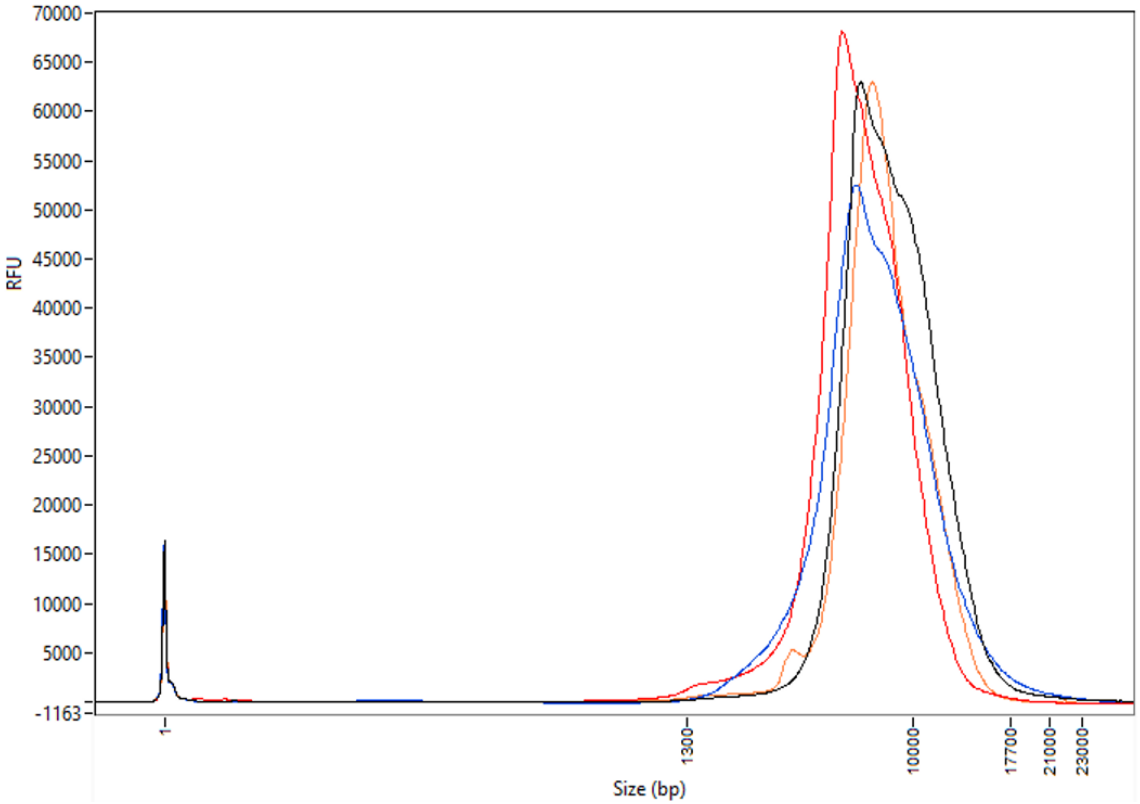


Femto Pulse DNA sizing QC analysis of a human gDNA sample sheared using a Megaruptor 3 with speed setting 31 (1-cycle shear). The fragment size distribution mode is 17.8 kb.

Example Megaruptor 3 shearing results for microbial genomic DNA samples

Megaruptor 3 System

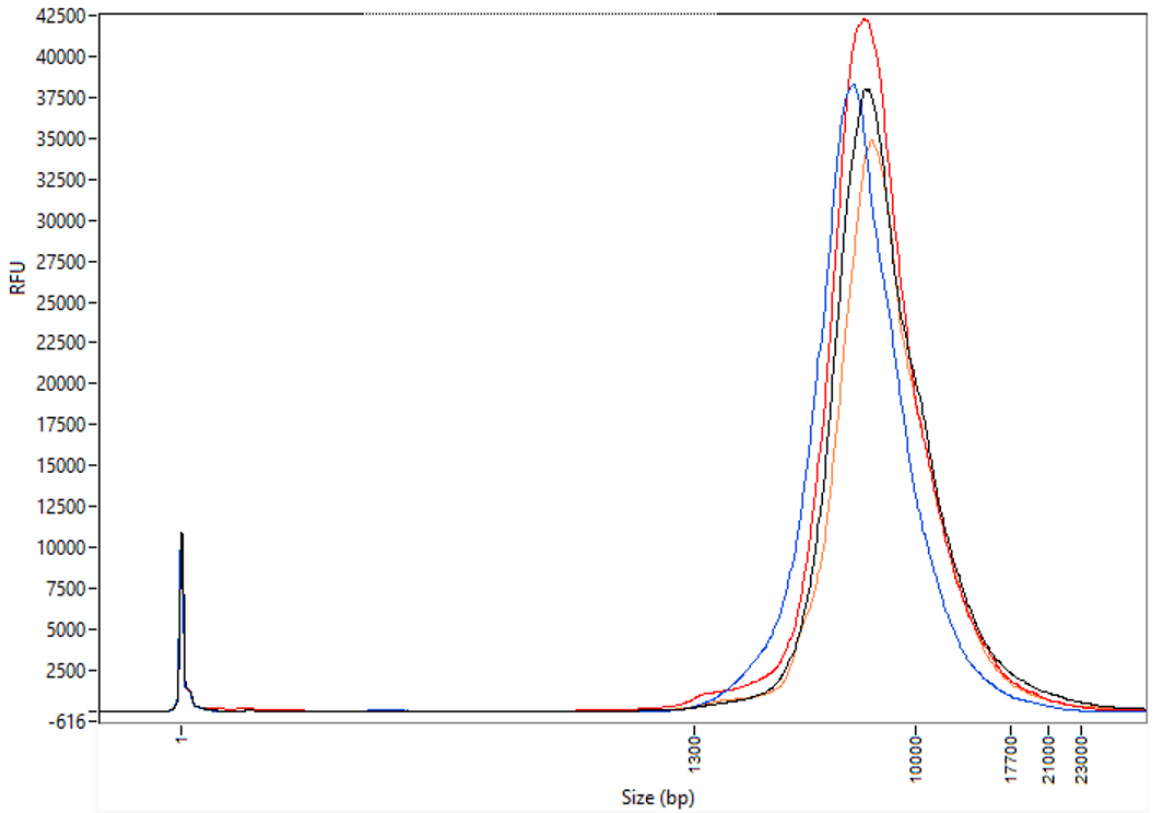
Speed Setting 40



Femto Pulse DNA sizing QC analyses of four different microbial gDNA samples sheared using a Megaruptor 3 System with speed setting 40 (1-cycle shear). The mean sheared DNA fragment size for all samples is ~7 kb – 10 kb.

g-TUBE

3287 x *g* [7000 RPM with Eppendorf MiniSpin Plus)



Femto Pulse DNA sizing QC analyses of four different microbial gDNA samples sheared using g-TUBES with a centrifugation speed of 3287 x *g*. The mean sheared DNA fragment size for all samples is ~7 kb – 10 kb.

SMRTbell barcoded adapter plate 3.0 is recommended for barcoding WGS samples

For Sequel II and IIe Systems, SMRTbell barcoded adapter plate 3.0 (102-009-200) is available for multiplexing up to 96 microbes per SMRT Cell 8M.

- Contains 96 barcoded adapters to support multiplexed SMRTbell library construction for up to 96 samples using SPK 3.0
- Can be used for Microbial Assembly and any other WGS or amplicon sequencing application that employs barcoded overhang adapters
- Each barcoded adapter contains a 5 bp padding sequence for more uniform ligation performance across different barcode sequences
- Each well on the plate contains a barcoded adapter with a unique 10-base pair PacBio barcode sequence
- Each barcoded adapter is present in only one well and supports a single reaction
- SMRT Link comes pre-installed with the following barcode set FASTA file containing SMRTbell barcoded adapter plate 3.0 barcode sequences*:
  `SMRTbell Barcoded Adapter Plate 3.0 (bc2001-bc2096)`

	1	2	3	4	5	6	7	8	9	10	11	12
A	BC2001	BC2009	BC2017	BC2025	BC2033	BC2041	BC2049	BC2057	BC2065	BC2073	BC2081	BC2089
B	BC2002	BC2010	BC2018	BC2026	BC2034	BC2042	BC2050	BC2058	BC2066	BC2074	BC2082	BC2090
C	BC2003	BC2011	BC2019	BC2027	BC2035	BC2043	BC2051	BC2059	BC2067	BC2075	BC2083	BC2091
D	BC2004	BC2012	BC2020	BC2028	BC2036	BC2044	BC2052	BC2060	BC2068	BC2076	BC2084	BC2092
E	BC2005	BC2013	BC2021	BC2029	BC2037	BC2045	BC2053	BC2061	BC2069	BC2077	BC2085	BC2093
F	BC2006	BC2014	BC2022	BC2030	BC2038	BC2046	BC2054	BC2062	BC2070	BC2078	BC2086	BC2094
G	BC2007	BC2015	BC2023	BC2031	BC2039	BC2047	BC2055	BC2063	BC2071	BC2079	BC2087	BC2095
H	BC2008	BC2016	BC2024	BC2032	BC2040	BC2048	BC2056	BC2064	BC2072	BC2080	BC2088	BC2096

Figure illustration of mapping between a specific well location and a unique PacBio barcode sequence on a 96-well plate in the SMRTbell barcoded adapter plate (102-009-200)

Reagent kit quantities support a **single use** of each of the 96 barcoded adapters in the plate for SMRTbell library preparations.

Plate Layout (Excel) [Link]

Barcode Sequences (FASTA) [Link]

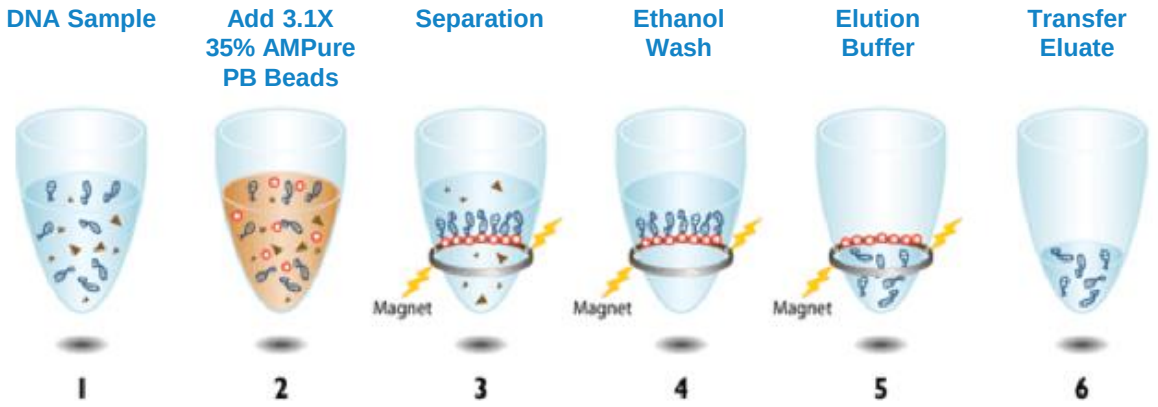
Product insert: SMRTbell barcoded adapter plate 3.0 (96 barcodes, 96 samples) [Link]

* If your SMRT Link version does not have the required barcode set FASTA file for a specific analysis application, please visit PacBio's Multiplexing website to download the appropriate FASTA file or contact PacBio Technical Support.

Size-selection of SMRTbell libraries with AMPure PB beads is recommended for WGS applications

- AMPure PB Beads are used as the default size selection method* to remove short DNA fragments (<5 kb) and enrich for the long fragments when preparing SMRTbell libraries for whole genome sequencing
- AMPure PB bead size selection of SMRTbell templates is performed follows:
  - Prepare a **35% dilution (v/v)** of the AMPure PB bead stock by adding 1.75 mL of resuspended AMPure PB Beads to 3.25 mL of Elution Buffer (EB). [35% AMPure PB beads solution can be stored at 4°C for 30 days.]
  - Add 3.1X v/v of resuspended, room-temperature 35% AMPure PB beads solution to each sample tube and incubate for 20 min at RT to allow beads to bind to DNA
  - Place sample tubes on a magnetic rack to immobilize AMPure PB beads; wash samples with 80% ethanol 2X; then elute samples in 15 µL of EB for 5 min at RT

AMPure PB Beads

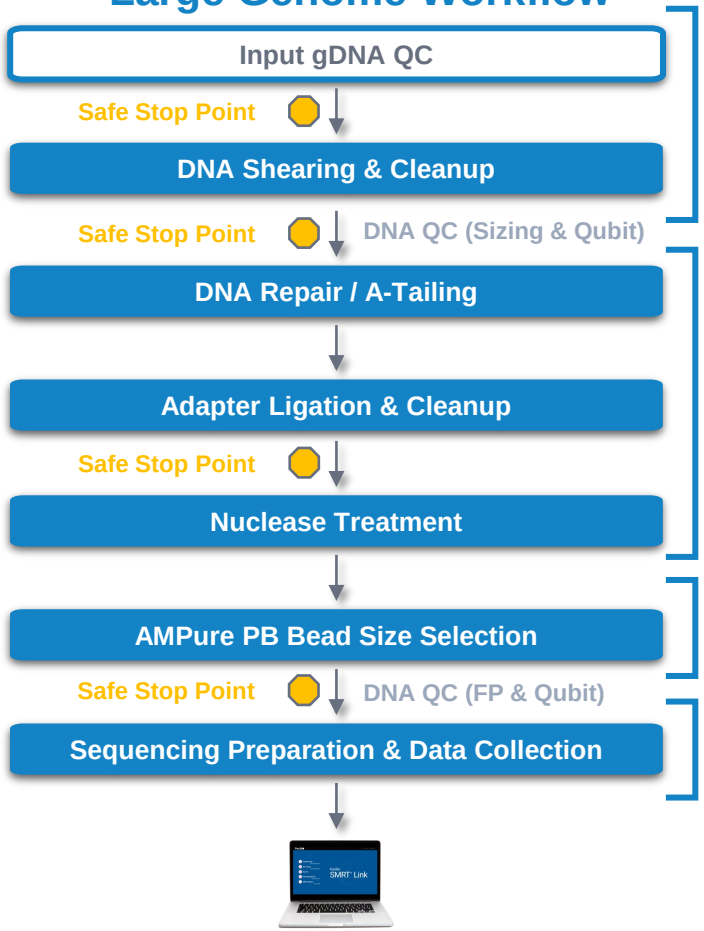


With high-quality WGS samples, AMPure PB bead size selection can recover sufficient SMRTbell library material to run **up to ~3 or more SMRT Cells 8M per 3 µg of starting input gDNA**

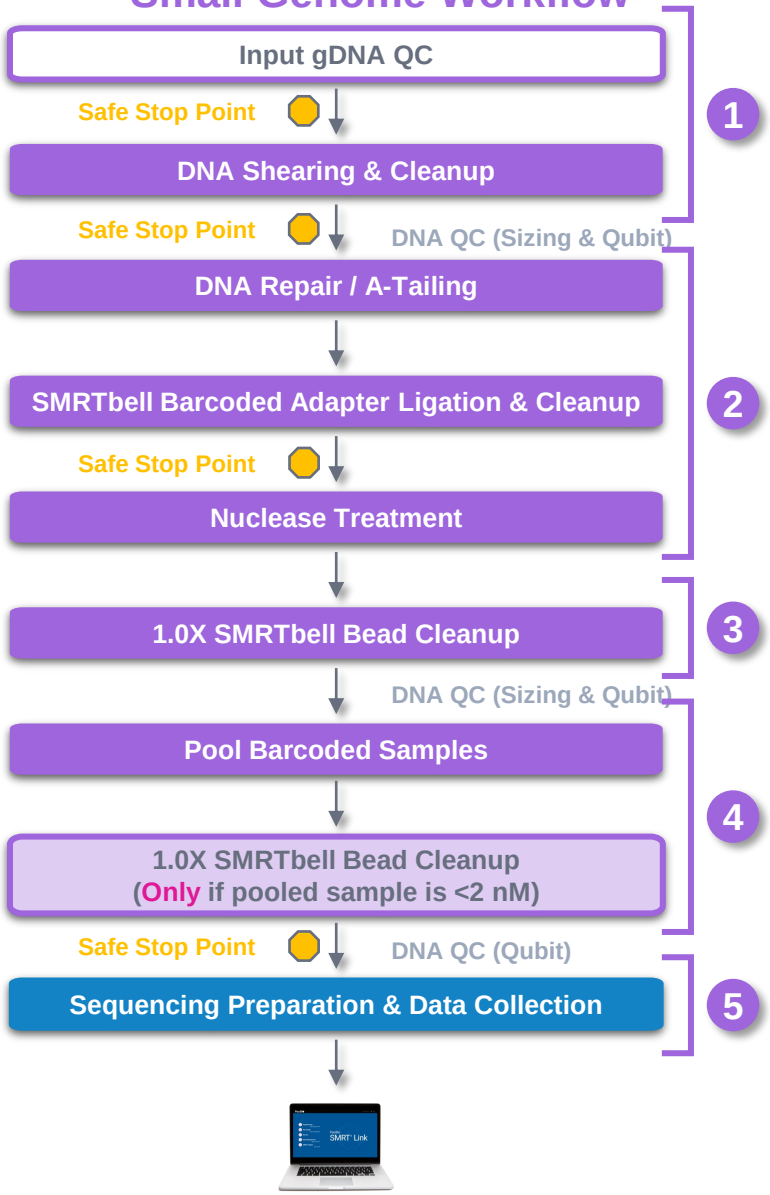
* **Note:** Although size-selection beads have many advantages, automated DNA size selection systems that utilize gel cassettes offer more flexibility in defining a size cutoff. Three automated DNA size selection tools that may **optionally** be employed for performing size selection on SMRTbell libraries for HiFi WGS applications include the PippinHT, BluePippin and SageELF systems from Sage Science. Note that use of these tools requires higher input DNA amounts (≥1.5 µg/SMRT Cell 8M). For more information, refer to Technical Note: Alternative size selection methods for SMRTbell prep kit 3.0 (TN103-110921), which provides detailed guidance for size selection of WGS libraries using automated DNA size selection tools or contact PacBio Technical Support or your local Field Applications Scientist.

Summary comparison of SPK 3.0 library sample preparation for large genome *vs*. small genome (microbial / shotgun metagenomic) WGS applications

Large Genome Workflow



Small Genome Workflow



Step	Action or metric	Large genome (Single sample)	Small genome (Multiplexed samples)
1	Input DNA quality	• 50% ≥ 30 kb • & 90% ≥ 10 kb	• 90% ≥ 7 kb
	Input DNA amount*	• Use ≥1 µg of total input DNA per SMRT Cell 8M	• Use ≥300 ng of input DNA per sample, with a total mass ≥1 µg across all samples
	Megaruptor system shearing	• Use speed setting 31	• Use Speed setting 40
2	Adapter ligation	• Use standard (non-barcoded) adapter included with SPK 3.0	• Use Barcoded adapter plate 3.0
3	Size selection	• Use AMPure PB bead size selection	• AMPure PB bead size selection is optional; otherwise perform a standard 1X cleanup with SMRTbell cleanup beads
4	Pooling	• N/A	• Perform equal mass pooling with barcoded samples
5	Sequencing prep & data collection	• 30 hr movie time	• 15 hr movie time

* Increase DNA input amounts to ≥1.5 µg per SMRT Cell 8M when using a gel-based automated size selection system.

How many shotgun metagenomic WGS samples can be multiplexed on a single SMRT Cell 8M?

The overall goals of your 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 ~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 HiFi 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

How many shotgun metagenomic WGS samples can be multiplexed on a single SMRT Cell 8M? (cont.)

Example calculation of estimated coverage levels achievable for rare species at different multiplex levels

	1 Sample / SMRT Cell 8M	2 Samples / SMRT Cell 8M	3 Samples / SMRT Cell 8M
Assignable HiFi (≥Q20) Reads per SMRT Cell 8M*	2.4 M	2.4 M	2.4 M
HiFi Reads per Sample	2.4 M	1.2 M	800,000
1% of Reads	24,000 → **assembly**	12,000 → **assembly**	8,000 → **profiling**
0.2% of Reads	4,800 → **profiling**	2,400 → **detection**	1,600 → **detection**

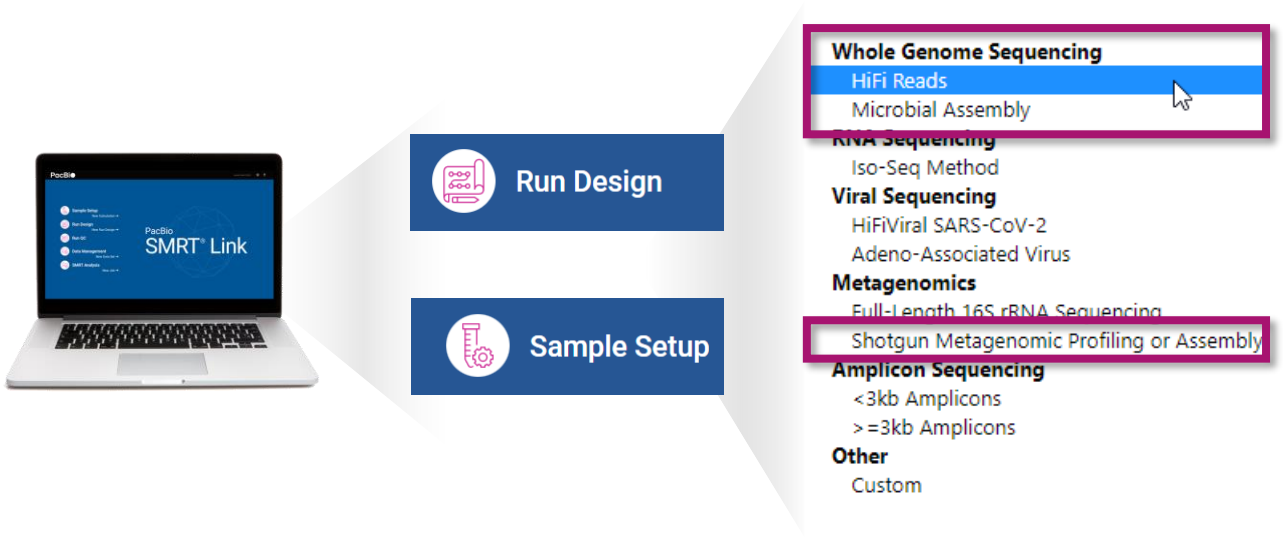
* Typically, ≥99.5% of HiFi reads have recoverable barcodes.

- The average HiFi read length for metagenomics samples is ~8.5 kb when following our recommended procedure with samples meeting the minimum DNA input quality requirements
- Choose your multiplex level depending on how many reads per rarest-OTU of interest you require for your metagenomic analysis plan

WGS library sequencing preparation workflow overview

Sample Setup & Run Design recommendations for WGS libraries

In SMRT Link Sample Setup & Run Design, select ‘HiFi Reads’, ‘Microbial Assembly’ or ‘Shotgun Metagenomic Profiling or Assembly’ for application type



- We recommend using Sequel II binding kit 3.2 & cleanup beads (102-333-300) to perform ABC (anneal primer / bind polymerase / clean up complex) with WGS samples
- Refer to Quick reference card – loading and pre-extension time recommendations for the Sequel II/IIe systems (101-769-100) for updates to ABC workflow for specific applications

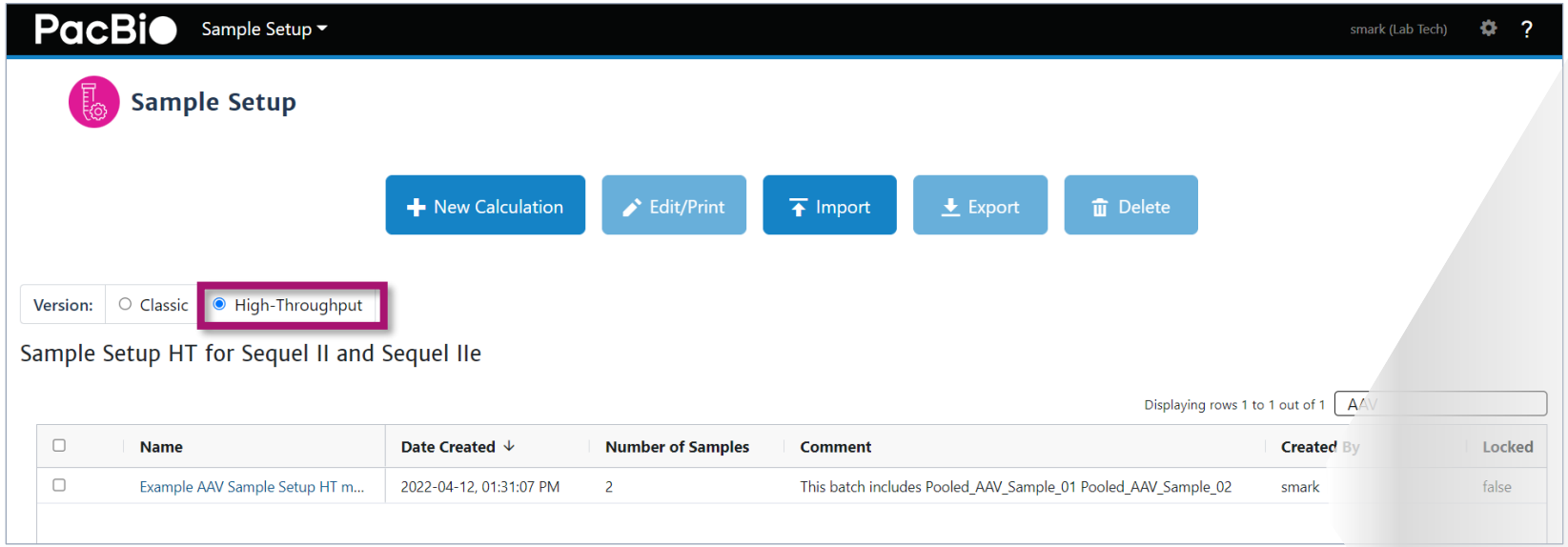
Sequel II binding kit 3.2 & cleanup beads (102-333-300) is recommended for preparing WGS samples for sequencing.

Sequel II binding kit 3.2 & cleanup beads (102-333-300) includes the following components:

- Sequencing primer 3.2
- Sequel II polymerase 2.2
- SMRTbell cleanup beads for complex cleanup
- DNA internal control 3.2 (defined 11 kb template bound to Polymerase 2.2)
- Supports ≥24 binding reactions, and up to 4 SMRT Cells 8M per binding reaction (96 cells total), depending on use case, sample size and concentration

Sample Setup guidance for WGS samples

Use SMRT Link Sample Setup High-Throughput (HT) mode and follow instructions to perform ABC (anneal primer / bind polymerase / clean up complex) using recommended settings for WGS samples



- Sample Setup High-Throughput mode provides a simplified, streamlined workflow to efficiently process either one sample or multiple samples with similar library properties (such as mean insert size and DNA concentration) in parallel
- You can also export the calculated values to a CSV file for laboratory automation

	Sample Group
Actions	Copy / Remove / Lock / Automate
Name	Example Human WGS Sample Setup
Comment	1. Human_Sample_01 2. Human_Sample_02
Application	HiFi Reads
Binding Kit	Sequel® II Binding Kit 3.2
Number of Samples	2 samples
SMRT Cells per Sample	3 cells
Available Volume per Sample	15 uL
Insert Size	18000 bp
Sample Concentration	60 ng/uL
Cleanup Anticipated Yield	75 %
Recommended Conc... on Plate	50-90 pM
Specify Concentration on Plate	85 pM
Minimum Pipetting Volume	1 uL
Warnings	

Example Sample Setup HT mode worksheet for a batch comprised of two human WGS samples.

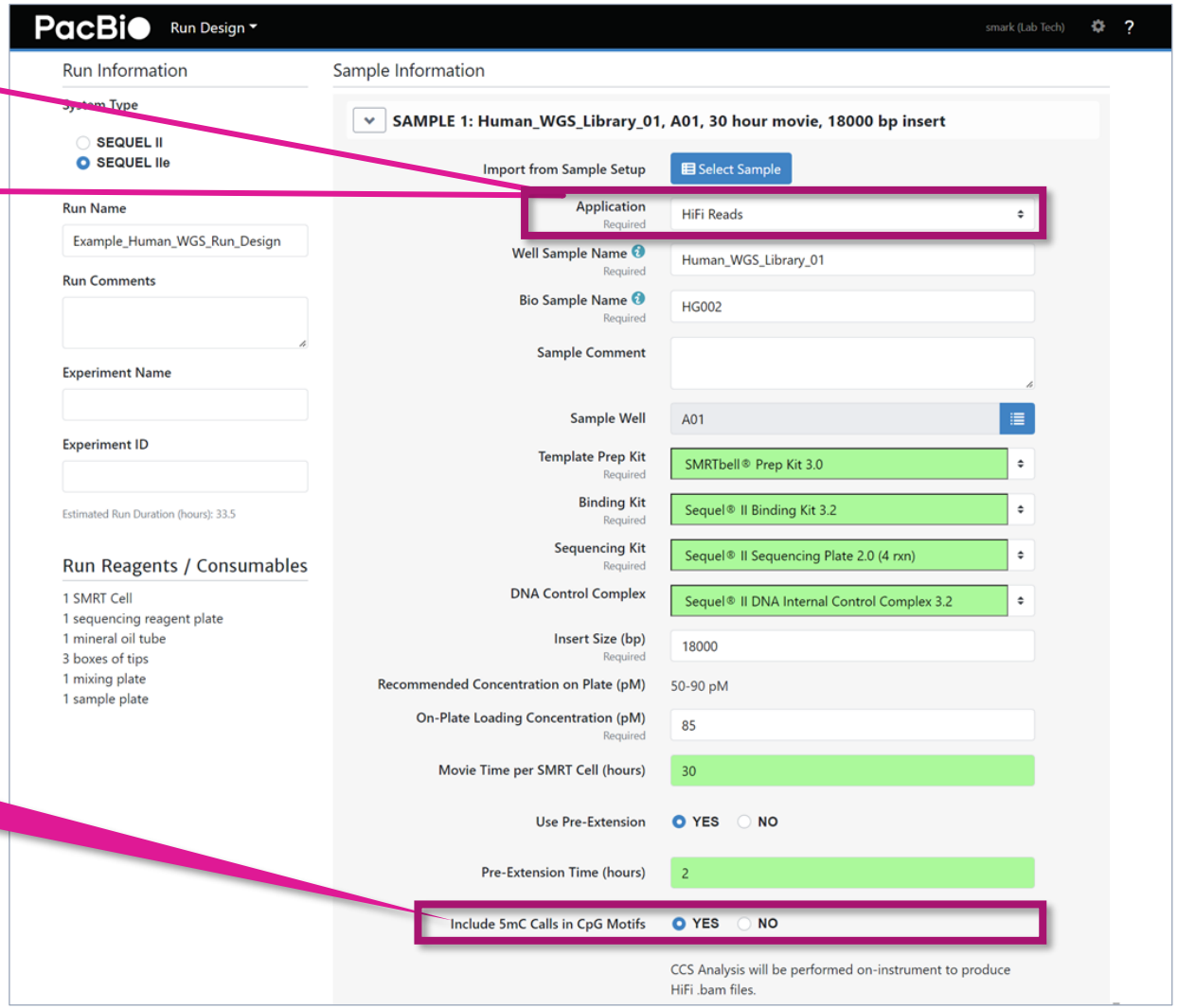
Binding Kit and **Cleanup Anticipated Yield** fields are auto-filled and highlighted in green after specifying application type

Run Design guidance for WGS samples

Follow SMRT Link Run Design instructions to set up a sequencing run using recommended settings for WGS samples

- Choose the appropriate **Application type** for your samples:
  - Select HiFi Reads if analyzing large genome WGS samples (default movie collection time = 30 hrs)
  - Select Microbial Assembly if analyzing small genome WGS samples (default movie collection time = 15 hrs)
  - Select Shotgun Metagenomic Profiling or Assembly if analyzing shotgun metagenomic WGS samples (default movie collection time = 30 hrs)

Include 5mC calls in CpG motifs is enabled by default if **HiFi Reads** or **Custom** is specified as the application type

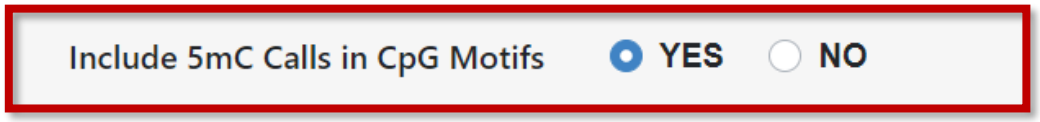


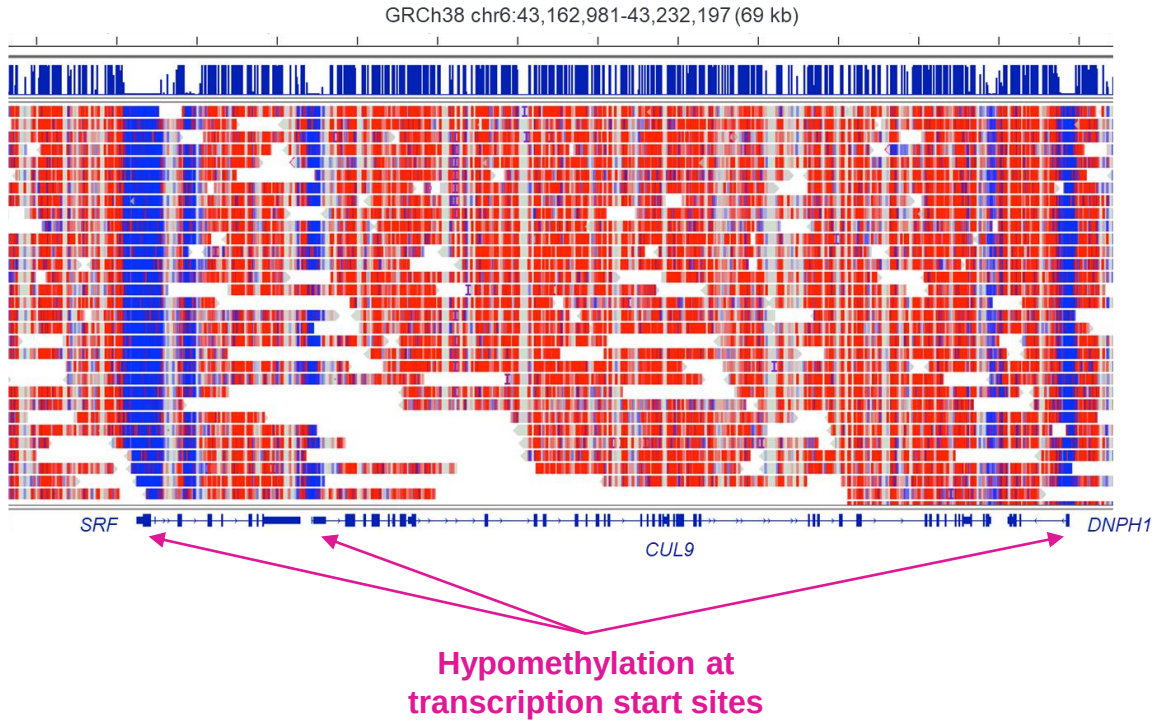
Example sample information entered into Run Design for sequencing a human WGS sample.

Include 5mC calls in CpG motifs option

If selected, kinetic signatures of cytosine bases in CpG motifs will be automatically analyzed to identify the presence of 5mC during on-instrument CCS (Sequel IIe system only) or during CCS analysis in SMRT Link

- **Default setting = YES** when specifying 'HiFi Reads' or 'Custom' application types
- 5mC detection is automatically performed on-instrument with the Sequel IIe system and in SMRT Link with the Sequel II system (data outputs are the **same** for both methods)
- 5mC calls are output in `hifi_reads.bam` as BAM standard MM and ML tags and can be easily visualized in IGV
- Processing and storage requirements are **minimal**:
  - File size increase is ~5%
  - On-instrument processing time for Sequel IIe systems is ~10 minutes
- Kinetics are not retained in the CCS analysis output by default, but they can **optionally** be retained as well.
- 5mC calls require a **CpG context and symmetrical methylation** (i.e., does not detect hemi-methylated sites)
- Though trained on human data, 5mC detection has been demonstrated to work on non-human data (e.g., plants (Maize)).
- 5mC consensus calling and other tools planned for a a future SMRT Link version.
  - For guidance on command line tool options for 5mC analysis, please contact your local PacBio support team or PacBio Technical Support



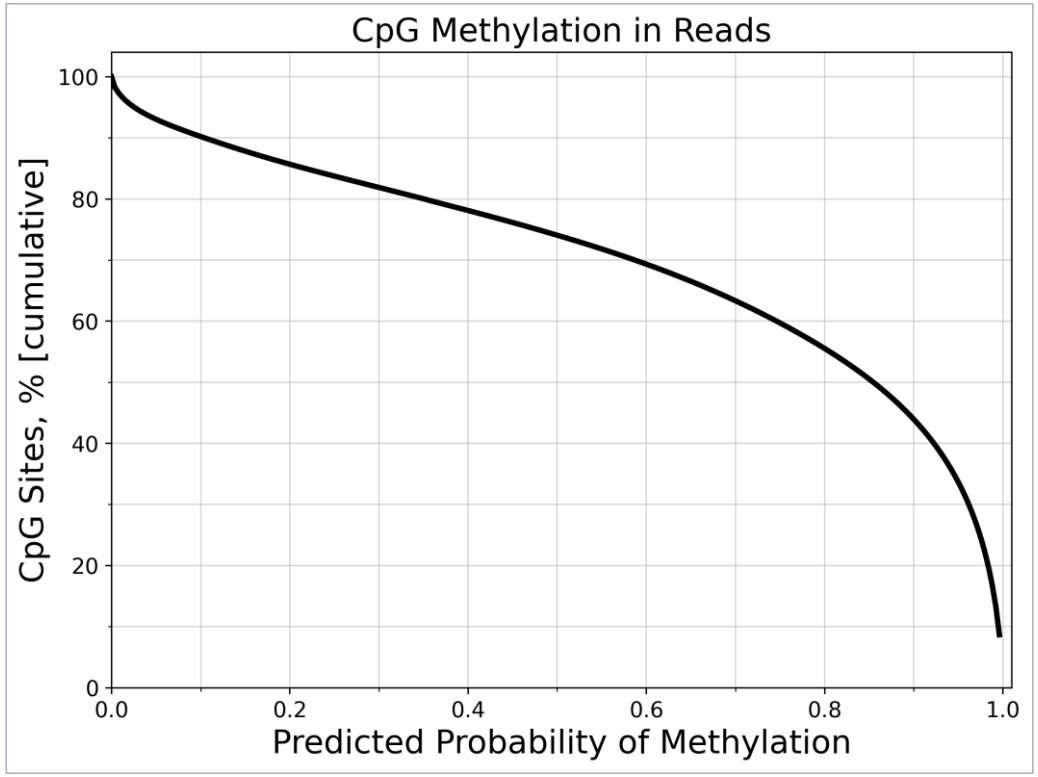


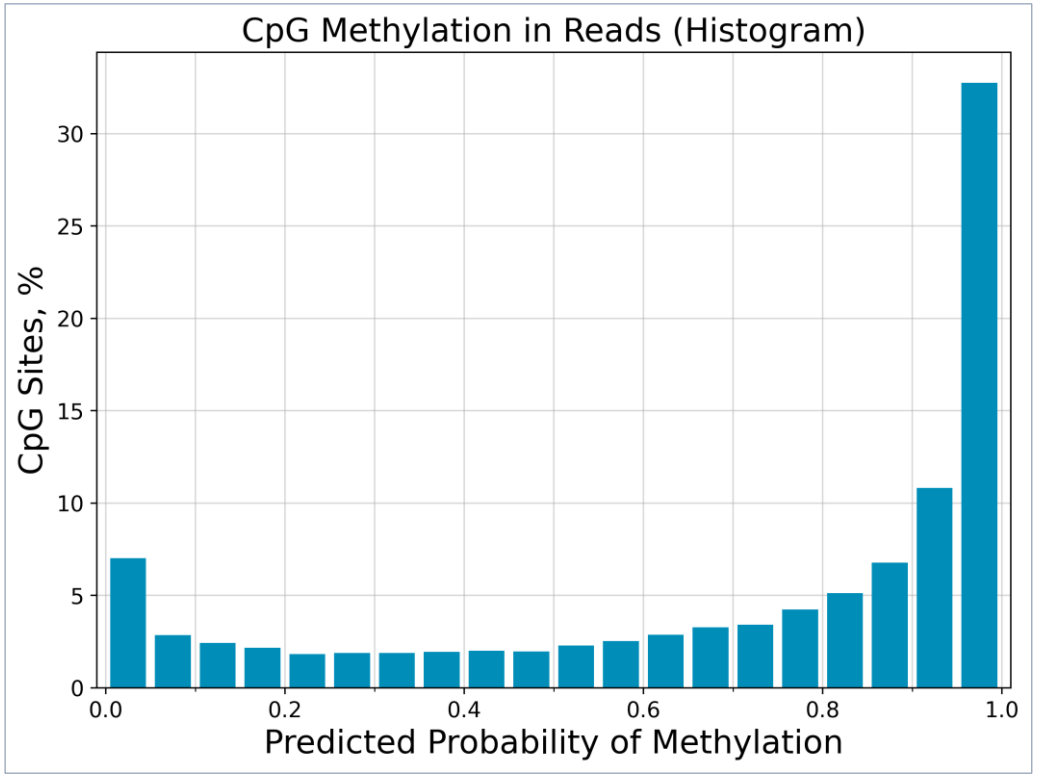
Example IGV plot demonstrating 5mC detection in HiFi reads for a human HG002 sample. Hypomethylation at active transcription start sites can be easily visualized (unpublished data).

https://downloads.pacbcloud.com/public/dataset/HG002-CpG-methylation-202202/

Include 5mC calls in CpG motifs option (cont.)

- The 5mC CpG detection utility generates the following reports:
  - **CpG Methylation in Reads:** Plots the cumulative percentage of CpG sites in the sample against the predicted probability of methylation. (Report appears in SMRT Link Run QC and Data Management)
  - **CpG Methylation in Reads (Histogram):** Histogram plot displaying the percentage of CpG sites in the sample versus the predicted probability of methylation (Report appears in Data Management only)

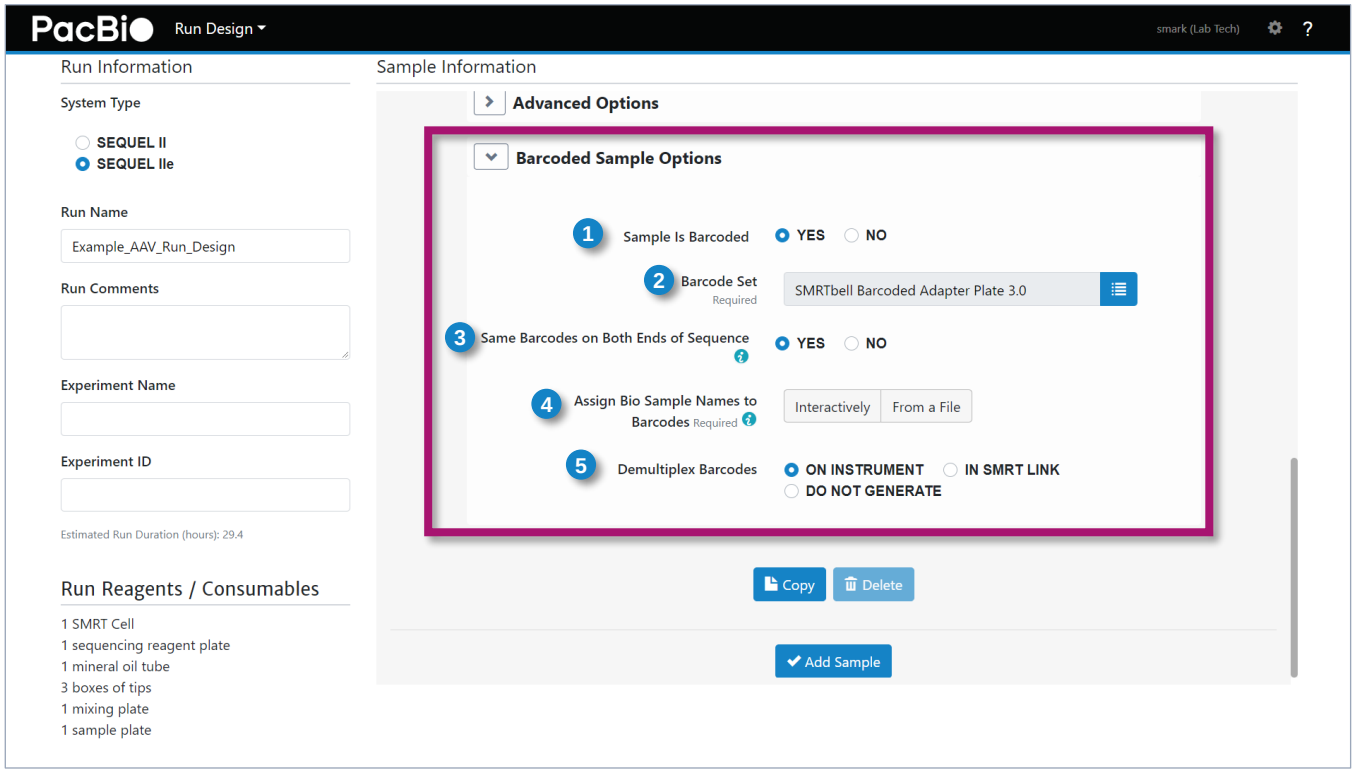




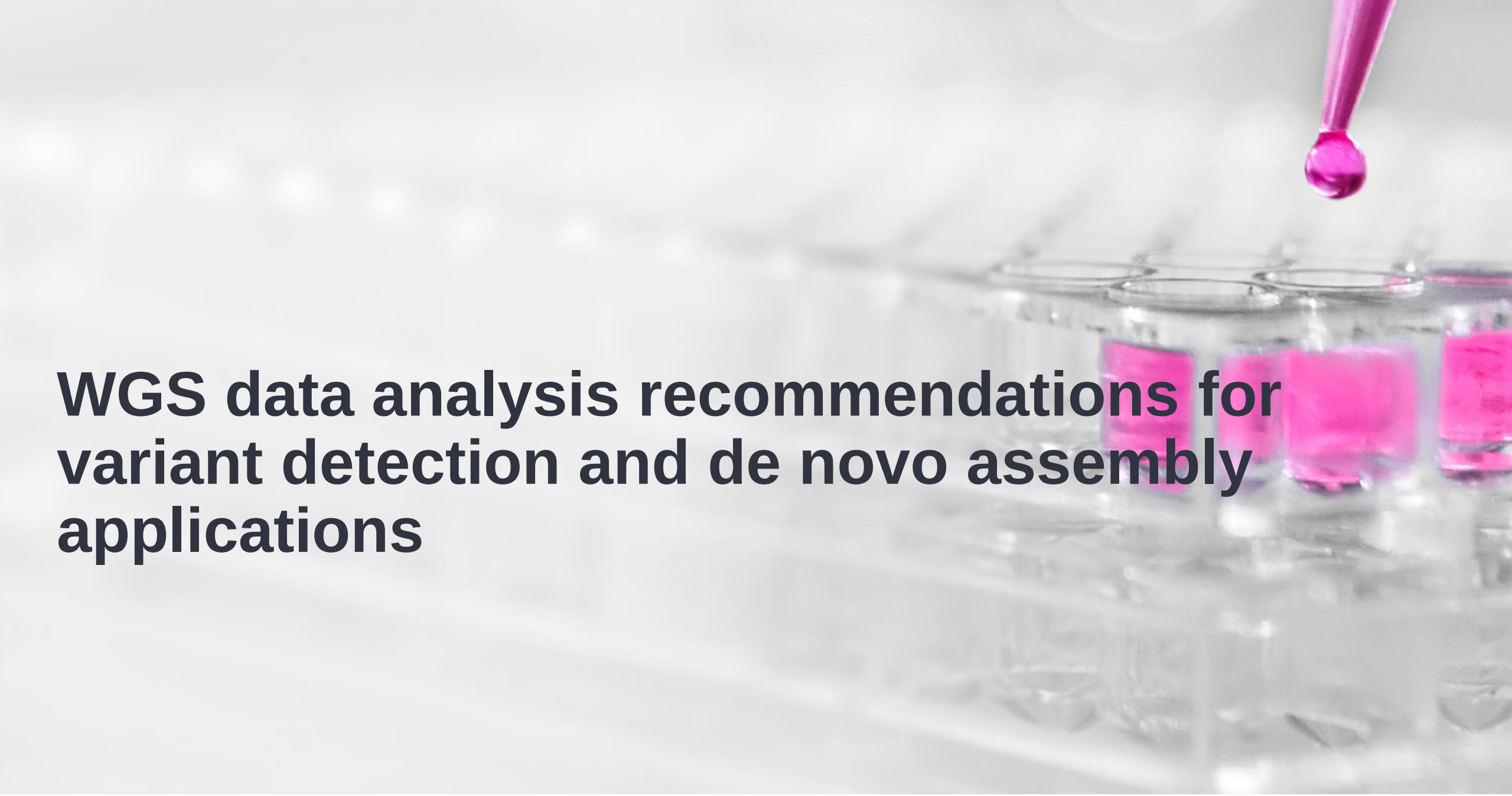
Run Design guidance for multiplexed WGS samples

Run Design setup procedure for automated barcode demultiplexing of WGS library samples barcoded with SMRTbell barcoded adapter plate 3.0

1. Sample is Barcoded: YES
2. Barcode Set: Select 'SMRTbell barcoded adapter plate 3.0 (bc2001-bc2096)' (or other custom barcode set if appropriate)
3. Same Barcodes on Both Ends of Sequence: YES
4. Assign a Biological Sample Name to each barcoded sample using one of two ways: From a (CSV) File or Interactively
5. Specify if barcode demultiplexing is to be performed on-instrument (Sequel IIe system only) or in SMRT Link. (Optionally specify Do Not Generate.)



Example barcoding information entered into Run Design for sequencing a multiplexed WGS sample.



WGS data analysis recommendations for variant detection and de novo assembly applications

HiFi WGS data analysis recommendations for large genomes

Using HiFi reads for *de novo* assembly analysis of large genomes

- Perform CCS analysis on-instrument using the Sequel IIe System or in SMRT Link to generate highly accurate and long single-molecule reads (HiFi reads)
- **10- to 15-fold HiFi read coverage per haplotype** is recommended for most *de novo* assembly projects

  → $Target\ HiFi\ Base\ Yield = [Haploid\ Genome\ Size\ (Gb)]x\ [Ploidy\ Level]x\ [Target\ HiFi\ Coverage\ per\ Haplotype]$

  E.g., for *de novo* assembly analysis of a 3 Gb diploid genome:

  Recommended Minimum Target HiFi Base Yield = 3 Gb x 2 x 10 = 60 Gb

- Output data in standard file formats, (BAM and FASTA/Q) for seamless integration with downstream analysis tools
- Can use SMRT Link Genome Assembly analysis application (powered by IPA) or other third-party software for *de novo* assembly analysis using HiFi reads:
  - Hifiasm
  - HiCanu
- Contact PacBio Technical Support (support@pacb.com) or your local Bioinformatics Field Applications Scientist for additional information about data analysis recommendations

HiFi WGS data analysis recommendations for large genomes (cont.)

Using HiFi reads for variant detection analysis of large genomes

- Perform CCS analysis on-instrument using the Sequel IIe System or in SMRT Link to generate highly accurate and long single-molecule reads (HiFi reads)
- **≥15-fold HiFi read coverage per sample** is sufficient for most human variant detection projects

  → $Target\ HiFi\ Base\ Yield = [Sample\ Haploid\ Genome\ Size\ (Gb)]\ x\ [Target\ Coverage\ per\ Sample]$

  E.g., For variant detection analysis of a human genome (3 Gb):

  Recommended Minimum Target HiFi Base Yield = 3 Gb x 15 = 45 Gb
- Output data in standard file formats, (BAM and FASTA/Q) for seamless integration with downstream analysis tools
- For detection of small variants (SNVs, InDels <50 bp):
  - Can use third-party software (e.g., Google DeepVariant)
- For detection of structural variants (SVs >50 bp):
  - Can use the Structural Variant Calling application in SMRT Analysis
- Contact PacBio Technical Support (support@pacb.com) or your local Bioinformatics Field Applications Scientist for additional information about data analysis recommendations

HiFi WGS data analysis recommendations small genomes (microbial multiplexing applications)

Using HiFi reads for *de novo* assembly and base modification detection analysis of microbial genomes

- Perform CCS analysis on-instrument using the Sequel IIe System or in SMRT Link to generate highly accurate and long single-molecule reads (HiFi reads)
- **15-fold HiFi read coverage per microbe** is recommended for most *de novo* assembly projects

  → $Target\ HiFi\ Base\ Yield = [Microbe\ Genome\ Size\ (Mb)]x\ [Target\ HiFi\ Coverage\ per\ Microbe]$

  E.g., for *de novo* assembly analysis of a 5 Mb microbial genome:

  Recommended Minimum Target HiFi Base Yield = 5 Mb x 15 = 75 Mb

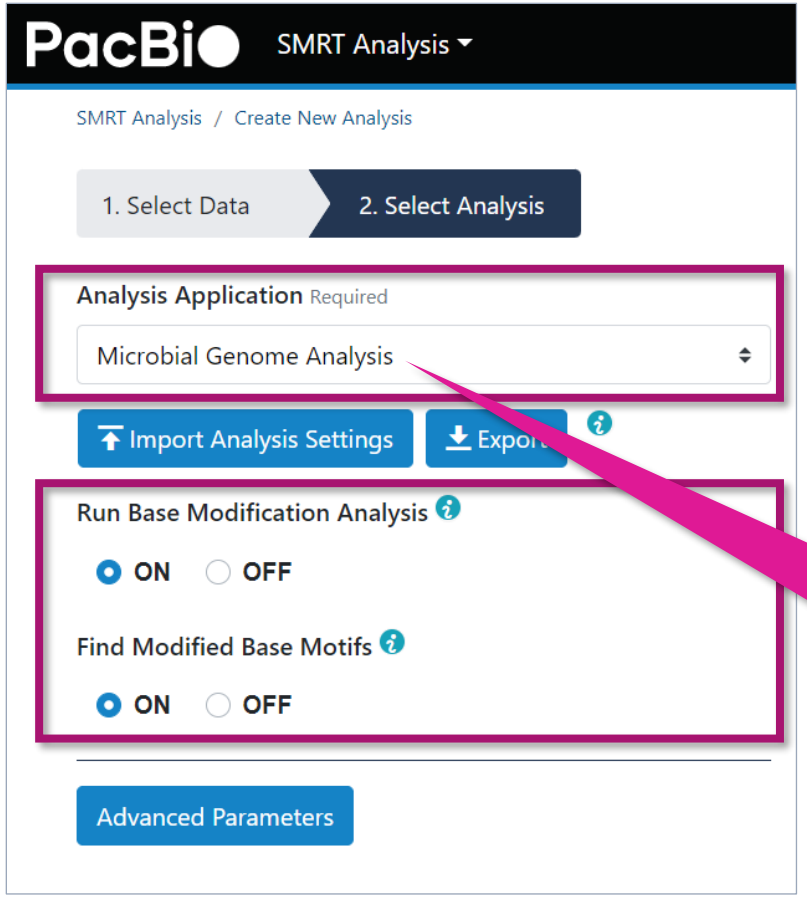
- Output data in standard file formats, (BAM and FASTA/Q) for seamless integration with downstream analysis tools
- Can use SMRT Link Microbial Genome analysis application for *de novo* assembly and base modification detection analysis using HiFi reads:
  - Easy to use (no requirement for laborious parameter input/optimization)
  - Enables fast and efficient microbial assembly results using HiFi reads (typical time to result is ~20-60 minutes* for analysis of a 96-plex microbial data set (up to 375 total sum of genome sizes))
  - Outputs complete, high-quality microbial genome assemblies (including chromosomes and plasmids)
- Contact PacBio Technical Support (support@pacb.com) or your local Bioinformatics Field Applications Scientist for additional information about data analysis recommendations

PacBio

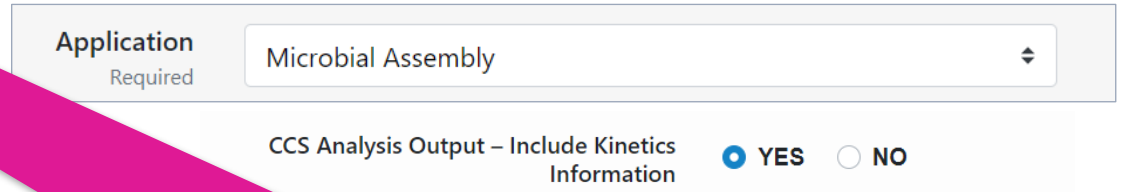
** Minimum Compute Requirements: Head Node - Cores: 32, RAM: 64 GB, 1 TB local tmp, 256 GB local db_datadir; Compute Nodes – Cores 64, RAM: 4GB per core, 1 TB local tmp, 256 GB local db_datadir*

HiFi WGS data analysis recommendations small genomes (microbial multiplexing applications) (cont.)

Use SMRT Link Microbial Genome Analysis application to perform microbial assembly and base modification detection using HiFi reads



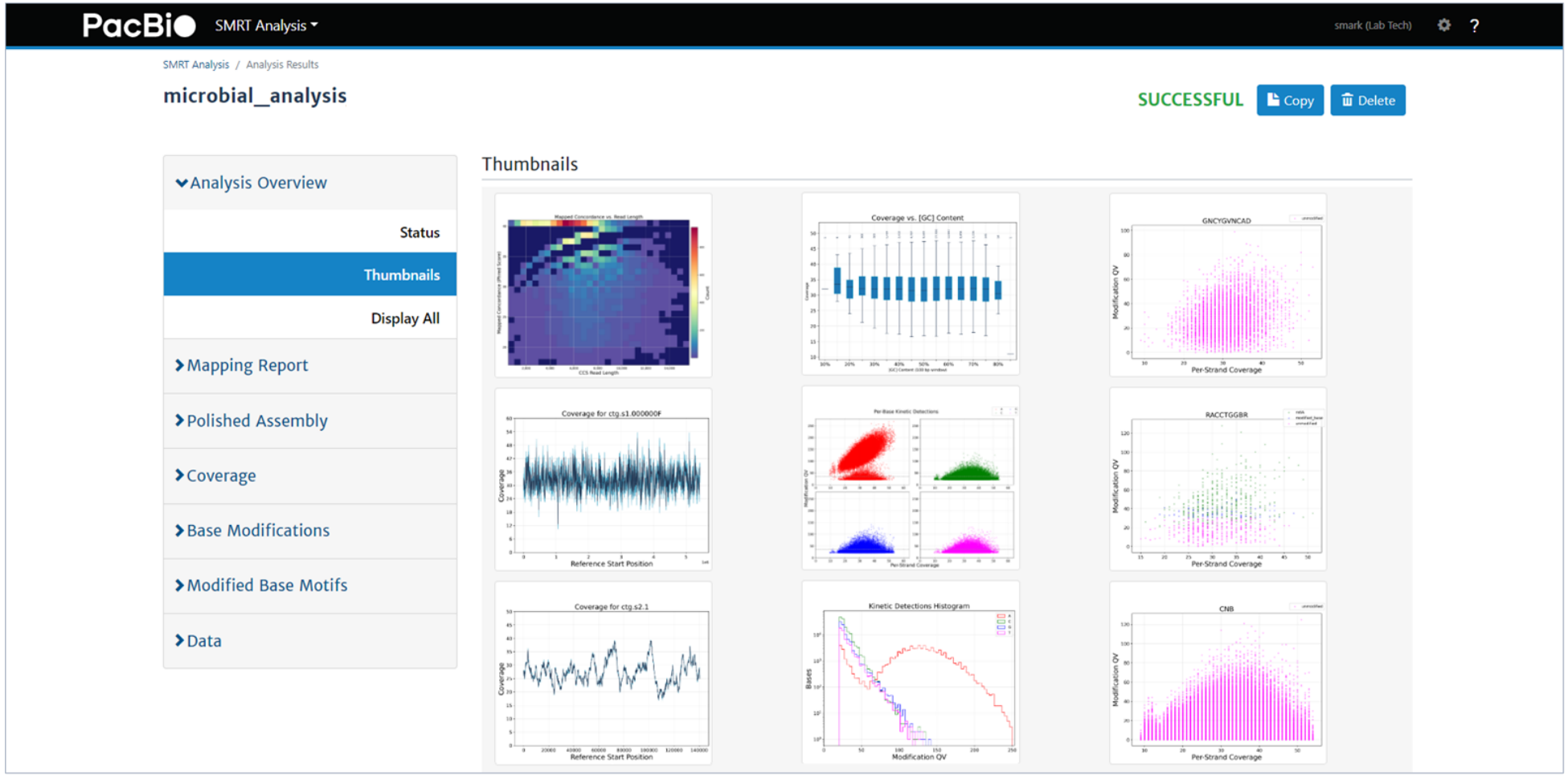
- Generate de novo assemblies of small **prokaryotic genomes between 1.9-10 Mb** and companion **plasmids between 2 – 220 kb**, and identify methylated bases and associated nucleotide motifs.
- Optionally include identification of **6mA and 4mC** modified bases and associated **DNA sequence motifs**. (This requires kinetic information.)
  - Unlike 5mC calling, microbial base modification detection is performed **off-instrument** (i.e., in SMRT Link only)
  - This requires a Run Design to specify that **kinetic information be retained** in the CCS analysis output
    - For the Microbial Assembly application type, Run Design **automatically** defaults to specifying **YES** for the '**CCS Analysis Output – Include Kinetics Information**' field



Note: This combines and **replaces** the Microbial Assembly and Base Modification Analysis applications in SMRT Link releases prior to v11.0.

HiFi WGS data analysis recommendations small genomes (microbial multiplexing applications) (cont.)

View SMRT Link microbial assembly results, detected base modifications and identified modified base motifs in a single analysis job report



HiFi WGS data analysis recommendations small genomes (shotgun metagenomic applications)

HiFi reads are compatible with third-party metagenomics data analysis tools

- Perform CCS analysis on-instrument using the Sequel IIe System or in SMRT Link to generate highly accurate and long single-molecule reads (HiFi reads)
- Output data in standard file formats, (BAM and FASTA/Q) for seamless integration with downstream analysis tools
- Use PacBio metagenomics tools available on GitHub for taxonomic classification and functional gene profiling using HiFi reads
- Can perform metagenomic shotgun assembly directly with HiFi reads using Hifiasm and evaluate & extract metagenome-assembled genomes using PacBio HiFi-MAG-Pipeline tool available on GitHub (see Portik *et al*.)
- Contact PacBio Technical Support (support@pacb.com) or your local Bioinformatics Field Applications Scientist for additional information about data analysis recommendations

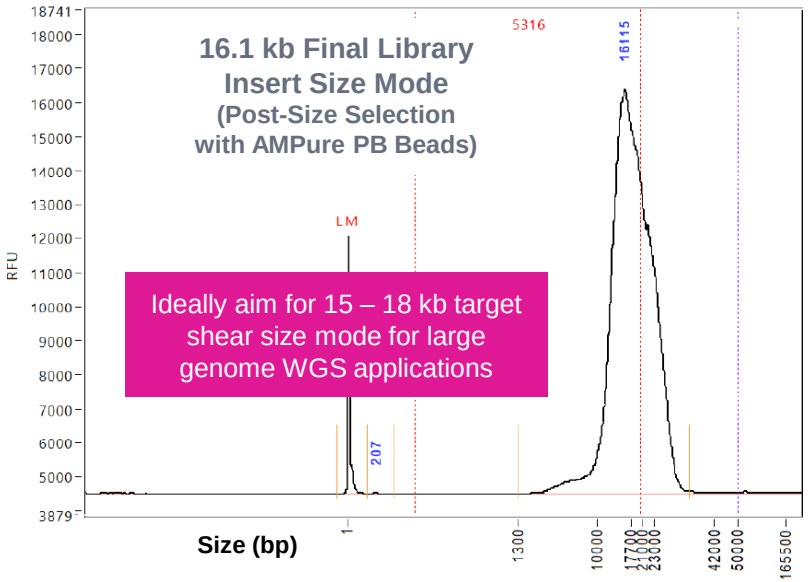
Portik D.M. et al. (2022) Evaluation of taxonomic profiling methods for long-read shotgun metagenomic sequencing datasets. BioRxiv Preprint doi: https://doi.org/10.1101/2022.01.31.478527

WGS library example sequencing performance data

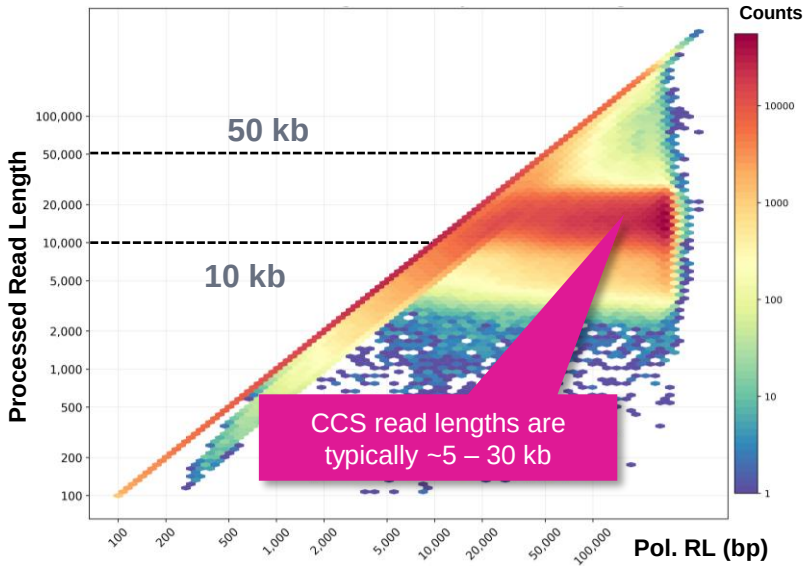
Example library QC and sequencing performance for human WGS libraries prepared with SMRTbell prep kit 3.0

SMRTbell library QC and primary sequencing metrics

Size-Selected Library QC

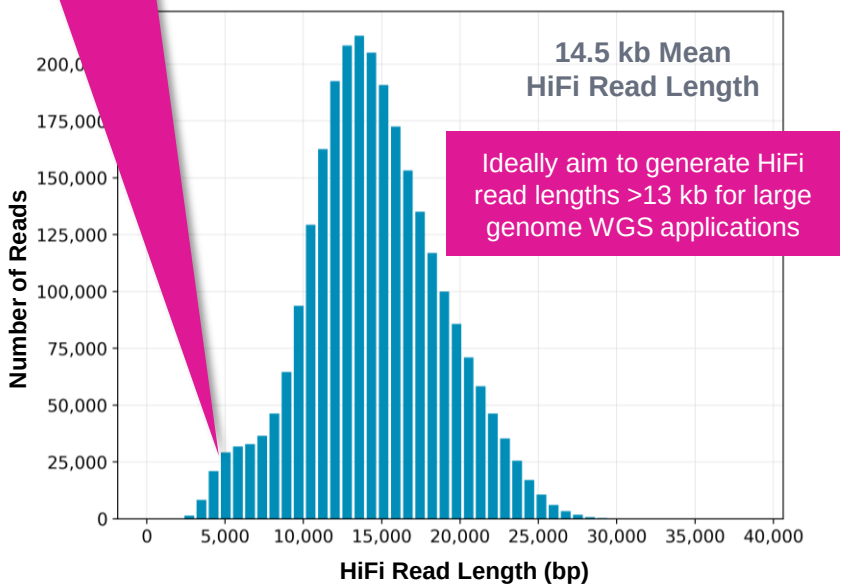


Raw Data Report



A secondary left hand peak may also be visible depending on DNA sample quality

CCS Analysis Report



Input gDNA for Megaruptor 3 shearing	3000 ng
Post-shearing recovery (%)*	2620 ng (87%)
Final yield of AMPure PB bead Size-selected library (%)**	1070 ng (36%)

* Post-shearing recoveries typically ranged from ~70% to >95% when using input human DNA samples (1 μg to 5 μg)

** Final post-size selected library yields typically ranged from ~25% to ~50% when using input human DNA samples (1 μg to 5 μg)

Raw Base Yield	617.65 Gb
Mean Polymerase Read Length	102.8 kb
P0	23.4%
P1	75.0%
P2	1.6%

Example sequencing metrics for a human WGS sample run with Binding Kit 3.2 (Polymerase 2.2) / 85 pM on-plate concentration / 30-h movie time / 2-h Pre-Extension Time / Adaptive Loading Target = 0.85

HiFi Reads	2.7 M
HiFi Base Yield	39.2 Gb
Mean HiFi Read Length	14,490 bp
Median HiFi Read Quality	Q34
HiFi Read Mean # of Passes	12

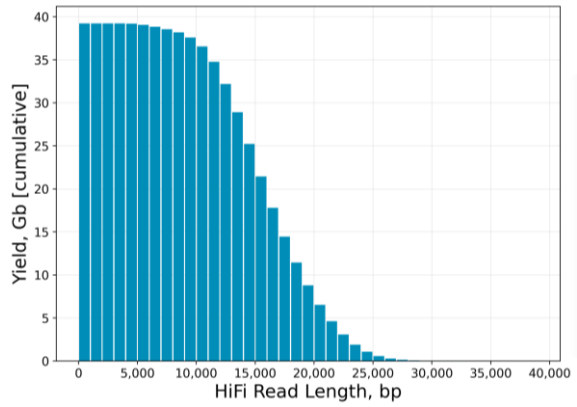
For SPK 3.0 human WGS libraries, per-SMRT Cell HiFi base yields typically ranged from ~28 to 39 Gb.

PacBio

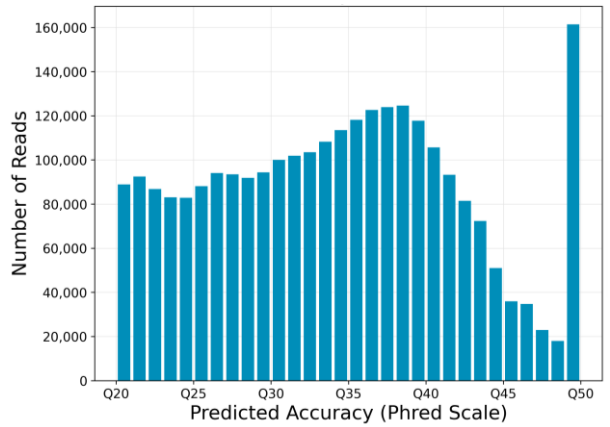
Example sequencing performance for human WGS libraries prepared with SMRTbell prep kit 3.0 (cont.)

Primary sequencing metrics (Cont.)

Yield by HiFi Read Length



Read Quality Distribution



HiFi Read Length Summary

Read Length (bp)	Reads	Reads (%)	Yield (bp)	Yield (%)
≥ 0	2,707,732	100	39,236,168,651	100
≥ 5,000	2,664,322	98	39,051,919,399	100
≥ 10,000	2,353,137	87	36,541,368,326	93
≥ 15,000	1,164,272	43	21,435,305,025	55
≥ 20,000	294,460	11	6,522,779,501	17
≥ 25,000	21,062	1	559,040,421	1
≥ 30,000	1,012	0	35,294,569	0
≥ 35,000	388	0	15,240,023	0
≥ 40,000	129	0	5,578,841	0

HiFi Read Quality Summary

Read Quality (Phred)	Reads	Reads (%)	Yield (bp)	Yield (%)
≥ Q20	2,707,732	100	39,236,168,651	100
≥ Q30	1,811,377	67	25,413,473,886	65
≥ Q40	679,582	25	8,150,599,400	21
≥ Q50	146,257	5	1,355,549,531	3

Example *de novo* assembly performance for human WGS libraries prepared with SMRTbell prep kit 3.0

HiFi WGS data sets generated with SPK 3.0 provide highly contiguous and highly accurate assemblies

HG002 Library ID	Contig_N50_Mbp
64009e_s10_cov30	35.4
64012e_s10_cov30	36.5
64015e_s10_cov30	36.7
64438e_s10_cov30	34.1
64441e_s10_cov30	33.6

HG002 Library ID	deNovo_asm_QV
64009e_s10_cov30	48.3
64012e_s10_cov30	48.3
64015e_s10_cov30	48.2
64438e_s10_cov30	48.3
64441e_s10_cov30	48.2

- Data were generated from five different human HG002 WGS libraries run on five different Sequel IIe system instruments
- Data were subsampled to 30-fold coverage and assembled using SMRT Link Genome Assembly analysis application

Example variant detection performance for human WGS libraries prepared with SMRTbell prep kit 3.0

HiFi WGS data sets generated with SPK 3.0 provide highly accurate variant calls

HG002 Library ID	INDEL.F1_Score
64009e_s10_cov30	0.995
64012e_s10_cov30	0.994
64015e_s10_cov30	0.994
64438e_s10_cov30	0.993
64441e_s10_cov30	0.994

HG002 Library ID	SNP.F1_Score
64009e_s10_cov30	0.999
64012e_s10_cov30	0.999
64015e_s10_cov30	0.999
64438e_s10_cov30	0.999
64441e_s10_cov30	0.999

- Data were generated from five different human HG002 WGS libraries run on five different Sequel IIe system instruments
- Data were subsampled to 30-fold coverage and analyzed with DeepVariant

F1 score = Harmonic mean of precision and recall

Example sequencing performance for a 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0

Sample preparation workflow

Experiment design

- 24 different microbes; each ligated independently to 4 different barcodes for 96-plex
- Selected microbes relevant to food safety and human health represent a range of genome sizes, GC content, and plasmid composition
- Total sum of genome sizes = 375 Mb

SMRTbell library construction

- 1 µg of input gDNA per microbe for shearing
- Target shear size = 7 kb – 10 kb
- 500 ng of sheared DNA into library prep
- Symmetrically barcoded samples using SMRTbell Barcoded Adapter Plate 3.0 Kit (102-009-200)
- No size-selection performed

Microbial genome assembly complexity*

Class I – Have few repeats except for the rDNA operon sized 5 to 7 kb

Class II - Class II genomes have many repeats, such as insertion sequence elements, but none greater than 7 kb.

Class III - Contain large, often phage-related, repeats >7 kb.

Microbial species	Genome size (bp)	GC content (%)	Microbial genome complexity	Barcode names
Acinetobacter baumannii AYE	3,960,239	39.35	Class 3	bc2001 / bc2025 / bc2049 / bc2073
Bacillus cereus 971	5,430,163	35.29	Class 1	bc2002 / bc2026 / bc2050 / bc2074
Bacillus subtilis W23	4,045,592	43.5	Class 1	bc2003 / bc2027 / bc2051 / bc2075
Burkholderia cepacia UCB 717	8,569,621	66.6	Class 3	bc2004 / bc2028 / bc2052 / bc2076
Burkholderia multivorans 249	7,008,277	66.68	Class 3	bc2005 / bc2029 / bc2053 / bc2077
Enterococcus faecalis OG1RF	2,739,503	37.75	Class 1	bc2006 / bc2030 / bc2054 / bc2078
Escherichia coli H10407	5,393,109	50.71	Class 1	bc2007 / bc2031 / bc2055 / bc2079
Escherichia coli K12 MG1655	4,642,522	50.79	Class 1	bc2008 / bc2032 / bc2056 / bc2080
Helicobacter pylori J99	1,645,141	39.19	Class 1	bc2009 / bc2033 / bc2057 / bc2081
Klebsiella pneumoniae BAA-2146	5,780,684	56.97	Class 2	bc2010 / bc2034 / bc2058 / bc2082
Listeria monocytogenes Li2	2,950,984	37.99	Class 1	bc2011 / bc2035 / bc2059 / bc2083
Listeria monocytogenes Li23	2,979,685	38.19	Class 1	bc2012 / bc2036 / bc2060 / bc2084
Methanocorpusculum labreanum Z	1,804,962	50.5	Class 1	bc2013 / bc2037 / bc2061 / bc2085
Neisseria meningitidis FAM18	2,194,814	51.62	Class 3	bc2014 / bc2038 / bc2062 / bc2086
Neisseria meningitidis Serogroup B	2,304,579	51.44	Class 1	bc2015 / bc2039 / bc2063 / bc2087
Rhodopseudomonas palustris CGA009	5,459,213	64.9	Class 3	bc2016 / bc2040 / bc2064 / bc2088
Salmonella enterica LT2	4,950,860	52.24	Class 1	bc2017 / bc2041 / bc2065 / bc2089
Salmonella enterica Ty2	4,791,947	52.05	Class 1	bc2018 / bc2042 / bc2066 / bc2090
Staphylococcus aureus Seattle 1945	2,806,348	32.86	—	bc2019 / bc2043 / bc2067 / bc2091
Staphylococcus aureus USA300_TCH1516	2,872,915	32.7	Class 1	bc2020 / bc2044 / bc2068 / bc2092
Streptococcus pyogenes Bruno	1,844,942	38.48	—	bc2021 / bc2045 / bc2069 / bc2093
Thermanaerovibrio acidaminovorans DSM6589	1,852,980	63.78	Class 1	bc2022 / bc2046 / bc2070 / bc2094
Treponema denticola A	2,842,721	37.87	—	bc2023 / bc2047 / bc2071 / bc2095
Vibrio parahaemolyticus EB101	5,146,979	45.33	Class 1	bc2024 / bc2048 / bc2072 / bc2096

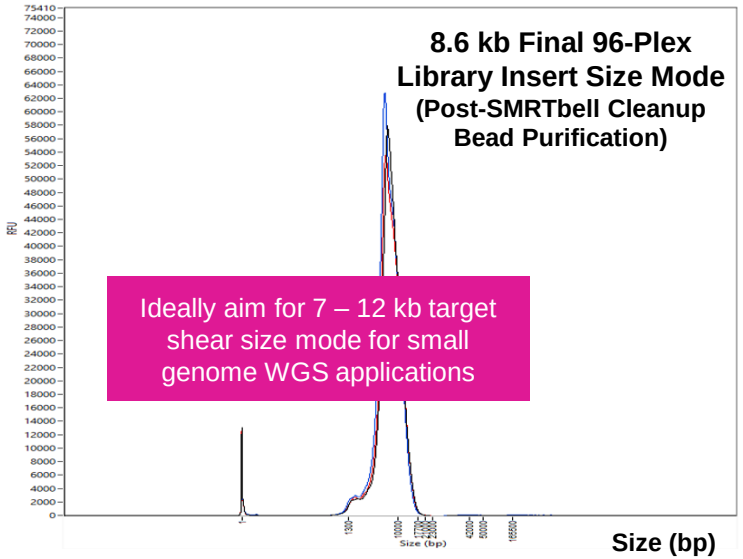
** Koren, S. et al. (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14, R101*

PacBio

Example sequencing performance for A 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0 (cont.)

Primary sequencing metrics

Size-Selected Library QC

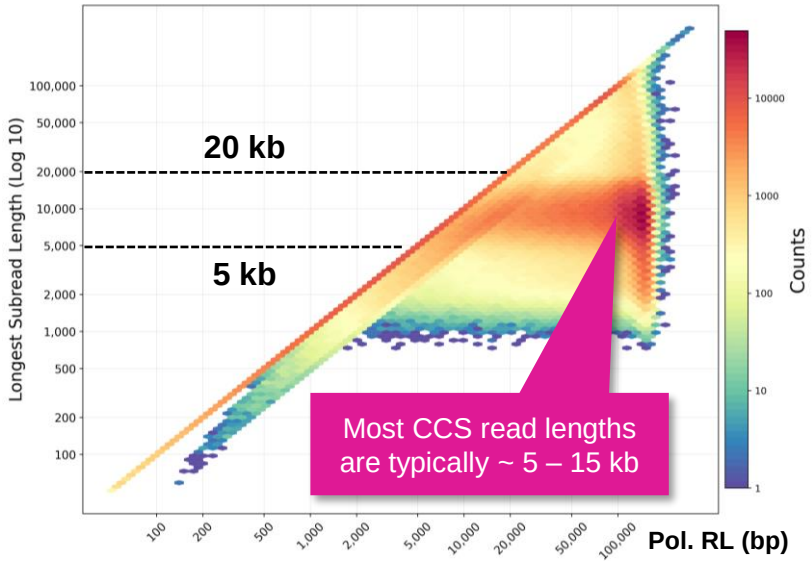


Input gDNA per microbe For Megaruptor 3 shearing	1 µg
Input sheared DNA per microbe For library construction	500 ng
Mean SMRTbell library construction yield per microbe before pooling (%)*	32%

* LC yields ranged from 17 – 52% across 96 microbes. Microbial libraries were barcoded with SMRTbell Barcoded Adapter Plate 3.0 Kit, independently purified with SMRTbell Cleanup Beads after nuclease treatment, and then pooled for sequencing on a single SMRT Cell 8M.

Raw Data Report

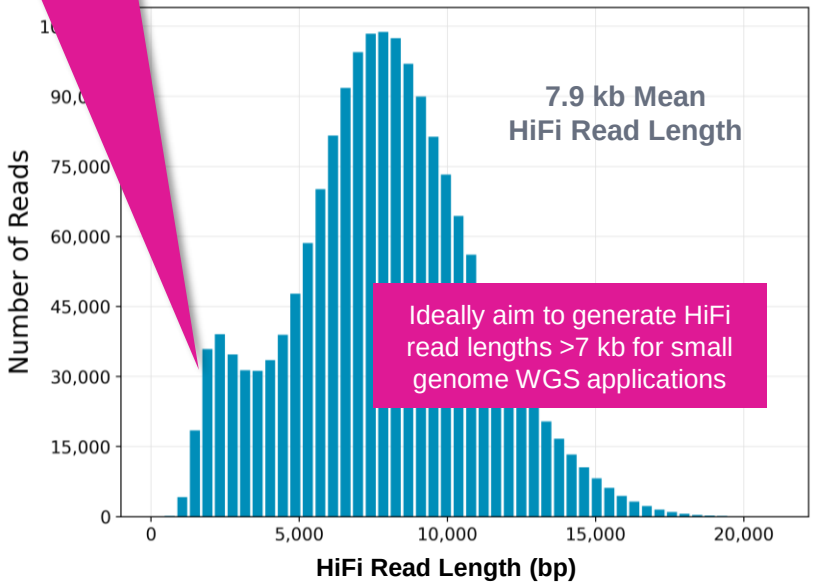
A secondary left hand peak may also be visible depending on DNA sample quality



Raw Base Yield	224.94 Gb
Mean Polymerase Read Length	79.2 kb
P0	63.3
P1	35.6
P2	1.1

90 pM on-plate concentration / 15-h movie time / 2-h Pre-Extension Time / Adaptive Loading Target = 0.85

CCS Analysis Report

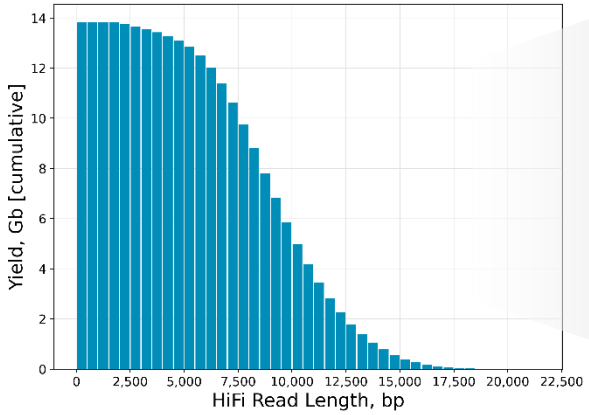


HiFi Reads	1.8 M
HiFi Base Yield	13.8 Gb
Mean HiFi Read Length	7,881 bp
Median HiFi Read Quality	Q38
HiFi Read Mean # of Passes	14

Example sequencing performance for A 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0 (cont.)

Primary sequencing metrics (Cont.)

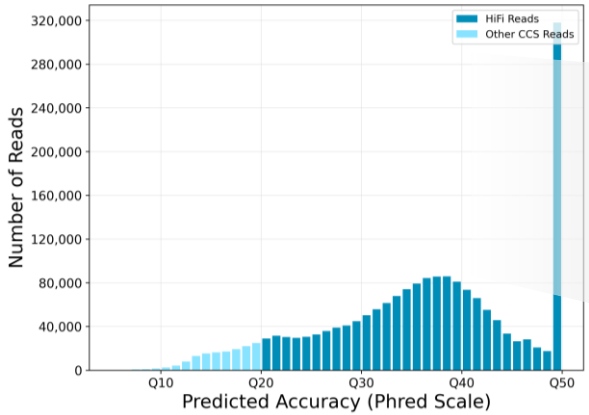
HiFi Read Length Distribution



HiFi Read Length Summary

Read Length (bp)	Reads	Reads (%)	Yield (bp)	Yield (%)
≥ 0	1,756,336	100	13,842,769,905	100
≥ 5,000	1,451,205	83	12,852,531,184	93
≥ 10,000	416,073	24	4,979,117,856	36
≥ 15,000	24,686	1	397,536,526	3
≥ 20,000	206	0	4,478,104	0
≥ 25,000	13	0	350,714	0
≥ 30,000	2	0	61,197	0
≥ 35,000	0	0	0	0
≥ 40,000	0	0	0	0

Read Quality Distribution



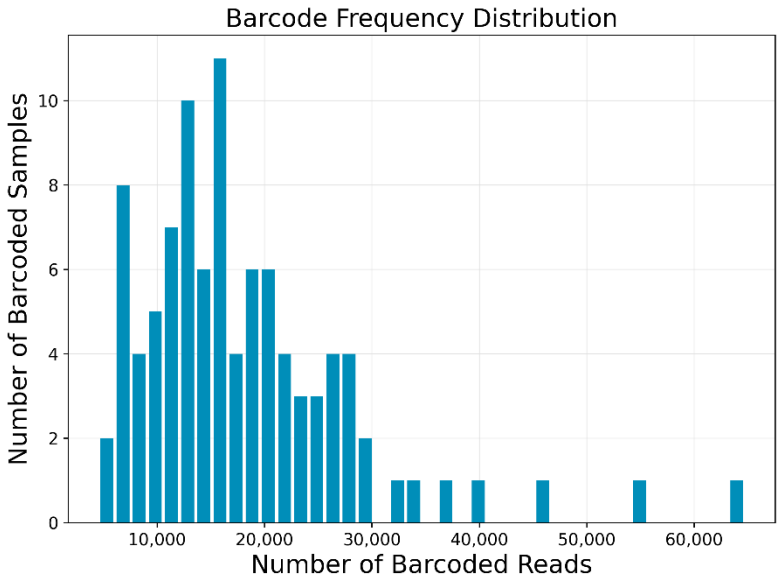
HiFi Read Quality Summary

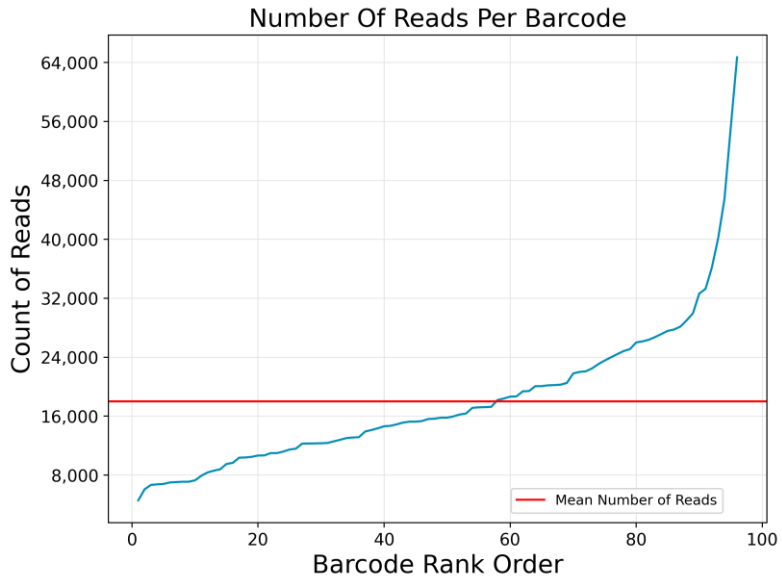
Read Quality (Phred)	Reads	Reads (%)	Yield (bp)	Yield (%)
≥ Q20	1,756,336	100	13,842,769,905	100
≥ Q30	1,411,586	80	10,691,739,197	77
≥ Q40	687,603	39	4,137,789,984	30
≥ Q50	302,228	17	1,376,190,958	10

Example sequencing performance for A 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0 (cont.)

Barcode demultiplexing results

Value	Analysis Metric
96	Unique Barcodes
1,731,704	Barcoded Reads
18,038	Mean Reads
64,709	Max. Reads
4,565	Min. Reads
7,856	Mean Read Length
24,632	Unbarcoded Reads
98.66%	Percent Bases in Barcoded Reads
98.59%	Percent Barcoded Reads

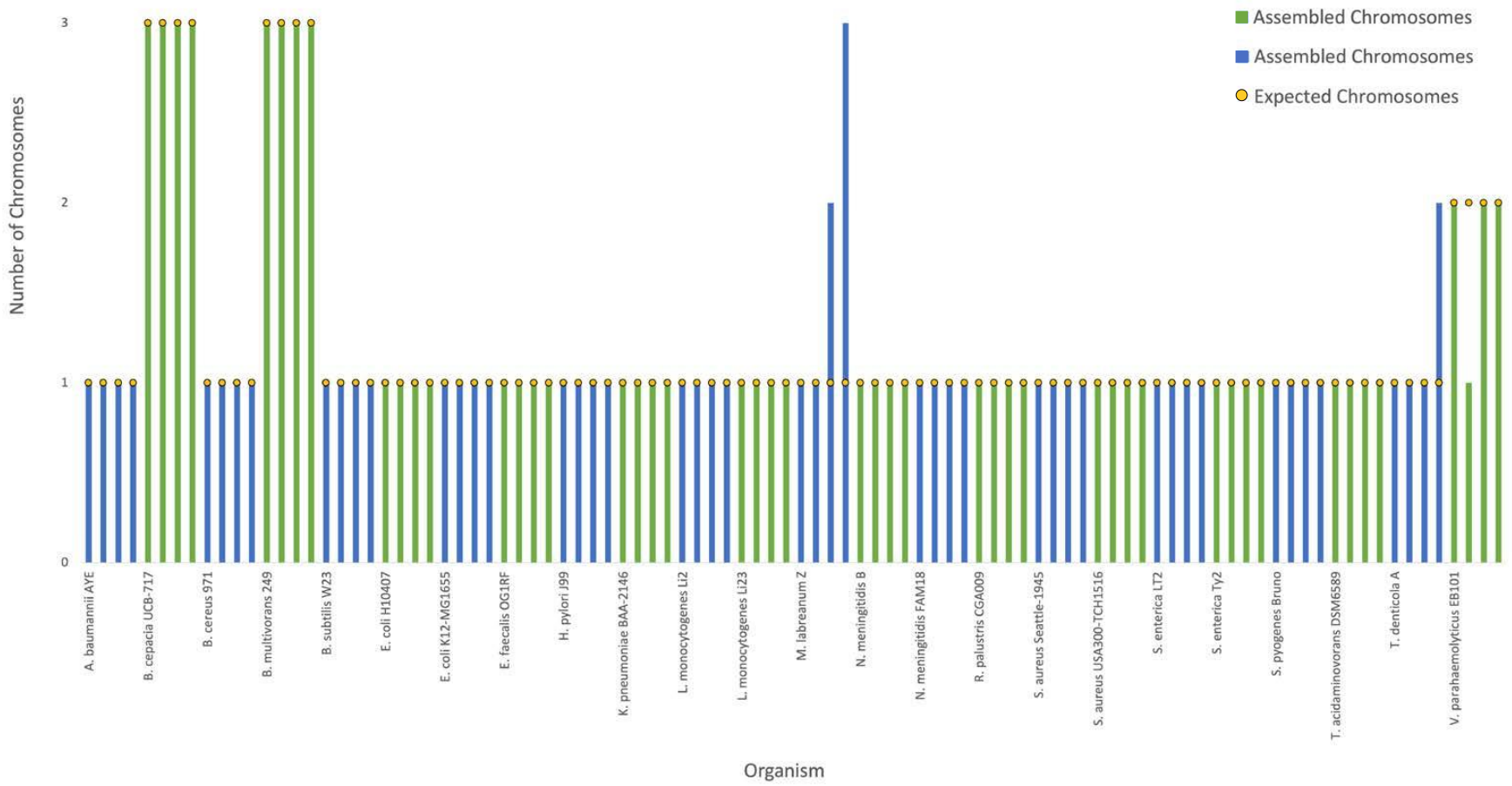




- **All 96 barcodes detected**
- Mean # of barcoded HiFi reads per microbe is ~18,000
- Mean HiFi base coverage per microbe is 36-fold (Range is 19- to 63-fold)

Example sequencing performance for A 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0 (cont.)

HiFi de *novo* assembly results – assembled chromosomes

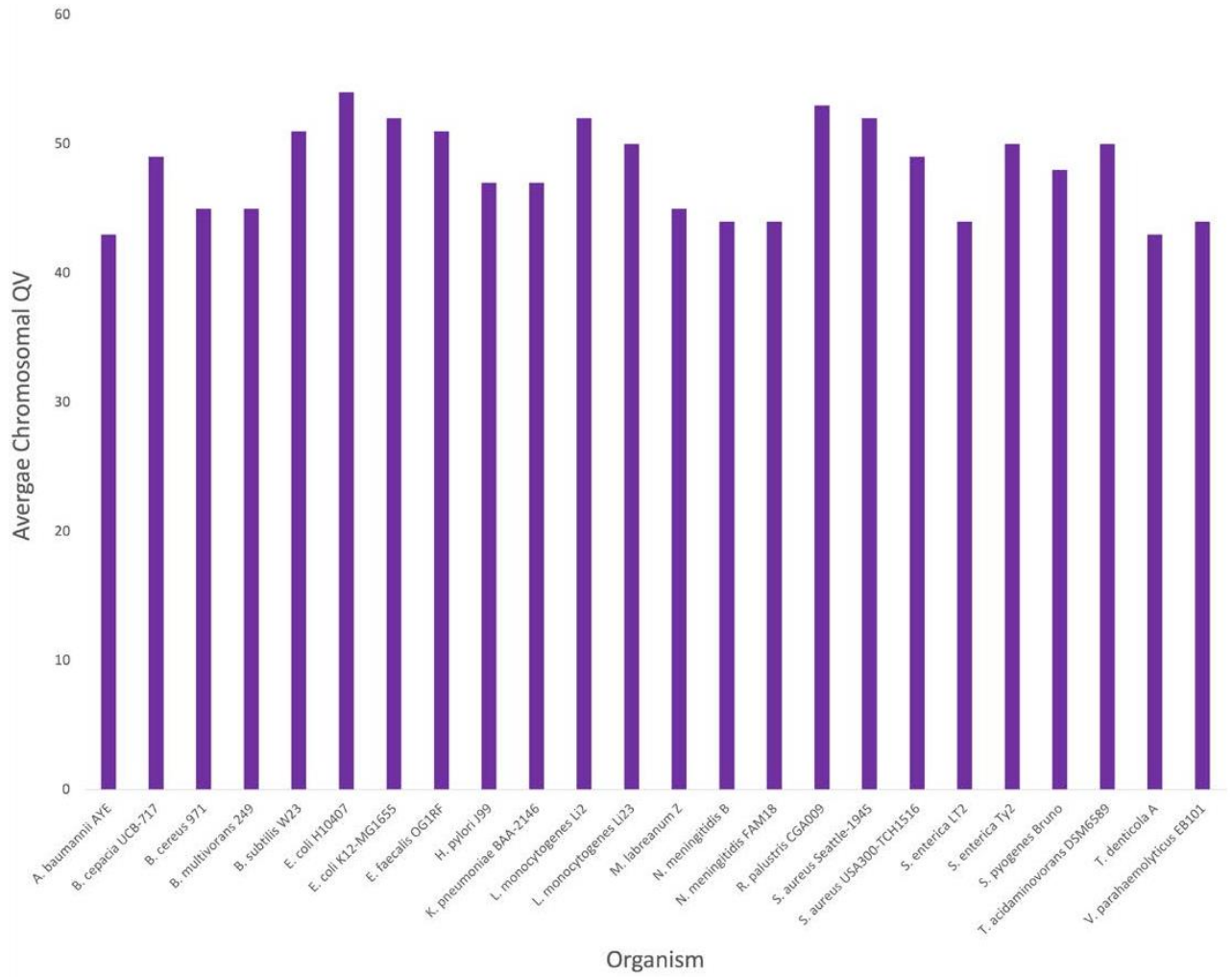


- Achieved **1 Contig / Chromosome** for 92 out of 96 assemblies
- **For all 96 microbes, most or all chromosomal assemblies were complete and of the expected sizes**

Microbial assembly statistics from a 96-plex pool of bacteria relevant to food safety and human health. These data were generated on the Sequel II system and assembled with the fully automated HiFi-based Microbial Assembly application in SMRT Link using the default parameters, without any manual curation. Download and explore the data yourself.

Example sequencing performance for A 96-plex microbial WGS library prepared with SMRTbell prep kit 3.0 (cont.)

HiFi *de novo* assembly results – representative assembly accuracies



- Accuracy of representative samples from a 96-plex microbial whole genome sequencing run on a Sequel II System
- With HiFi data and the Microbial Assembly application in SMRT Link, **genome assemblies are consistently >99.99% accurate**

Technical documentation & applications support resources

Technical resources for WGS library preparation, sequencing & data analysis

DNA extraction literature

- Circulomics Nanobind Application notes
- Circulomics Nanobind Kit handbooks
- Circulomics Nanobind Protocols
- Technical note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)
- Technical note: Preparing DNA for PacBio HiFi sequencing – Extraction and quality control (102-193-651)
- Technical note: Preparing samples for PacBio whole genome sequencing for *de novo* assembly – collection and storage (TN100-040518)

Sample preparation literature

- Application brief: Whole genome sequencing (WGS) for *de novo* assembly – Best practices (102-193-627)
- Application brief: Variant detection using whole genome sequencing with HiFi reads – Best practices (102-193-604)
- Overview – Sequel systems application options and sequencing recommendations (101-851-300)
- Procedure & checklist – Preparing whole genome and metagenome sequencing libraries using SMRTbell prep kit 3.0 (102-166-600)
- Quick reference card – Loading and pre-extension recommendations for the Sequel II and IIe systems (101-769-100)
- Technical note: Alternative size selection methods for SMRTbell prep kit 3.0 (TN103-110921)
- Technical note: Covaris g-TUBE DNA shearing for SMRTbell prep kit 3.0 (102-326-501)
- Technical overview – Whole genome and metagenome library preparation using SMRTbell prep kit 3.0 (102-390-900)

Technical resources for WGS library preparation, sequencing & data analysis (cont.)

Data analysis resources

- SMRT Link v11.0 user guide (102-278-200)
- SMRT Tools v11. 0 reference guide (102-278-500)
- Sequel II and IIe systems: Data files (102-144-100)

Example PacBio data sets

Whole genome sequencing application	Dataset	Data type	PacBio system
Assembly and variant detection	Homo sapiens – GIAB sample HG002	HiFi Reads	Sequel II System
Assembly	Oryza sativa – MH63	HiFi Reads	Sequel IIe System
Assembly (low-DNA input)	Anopheles gambiae – 2 plex	HiFi Reads	Sequel II System
Assembly (ultra-low DNA input)	Phlebotomus papatasi, Homo sapiens, Drosophila melanogaster	HiFi Reads	Sequel II System
Assembly	Food safety and infectious microbes – 96 plex	HiFi Reads	Sequel II System
5mC detection at CpG sites	Human HG002 CpG methylation status	HiFi Reads	Sequel IIe System

Technical resources for WGS library preparation, sequencing & data analysis (cont.)

Publications

- Nurk S. et al. (2022) The complete sequence of a human genome. Science. 376:44-53 doi: 10.1126/science.abj6987
- Noyes, D.N. et al. (2022) Familial long-read sequencing increases yield of de novo mutations. American journal of human genetics. 109:631-646. doi: https://doi.org/10.1016/j.ajhg.2022.02.014
- Bickhart, D.M. et al. (2022) Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. Nature biotechnology. doi: 10.1038/s41587-021-01130-z
- Lefoulon, E. et al. (2021) Greenhead (Tabanus nigrovittatus) Wolbachia and its microbiome: A preliminary study. Microbiol Spectr. 9(2):e0051721. doi: 10.1128/Spectrum.00517-21

Webinars

- PacBio Webinar (2022): Unlocking the genome with long-read sequencing in genetic disease research [Link]
- PacBio Webinar (2021): Methylation detection with PacBio HiFi sequencing [Link]
- PacBio Webinar (2021): HiFi sequencing: see what you've been missing [Link]
- PacBio Webinar (2021): Integrated rare disease using long-read genome sequencing [Link]
- PacBio Webinar (2021): Getting the most out of your breeding program with DNA and RNA sequencing [Link]
- PacBio Webinar (2021): Generation of lineage-resolved complete metagenome-assembled genomes in complex microbial communities [Link]

APPENDIX:
Genomic DNA isolation & QC recommendations for PacBio WGS library sample preparation

Sequel IIe System ICS v11.0 / SMRT Link v11.0

Genomic DNA isolation & QC recommendations for PacBio WGS sample preparation

Technical contents

1. Sample collection, preparation, and storage for SMRT sequencing
2. Genomic DNA extraction, QC and handling for SMRT sequencing
3. Example protocols and kit solutions for high-molecular weight genomic DNA isolation
4. Methods for evaluation of genomic DNA quality
5. Cleanup of genomic DNA and SMRTbell libraries
6. Storage and shipping of genomic DNA and SMRTbell libraries
7. DNA sample extraction literature resources

Sample collection, preparation, and storage for SMRT sequencing

Sample collection, preparation, and storage for SMRT sequencing whole genome sequencing projects

To obtain the highest quality genomic DNA, it is important to start with sample types compatible with high molecular weight (HMW) DNA extraction methods

PacBio Technical note: Sample prep (TN100-040518)

- Provides general guidance on biological sample collection, preparation, and storage across a range of commonly encountered sample types including Vertebrates, Invertebrates, Arthropods, Fungi and Plants

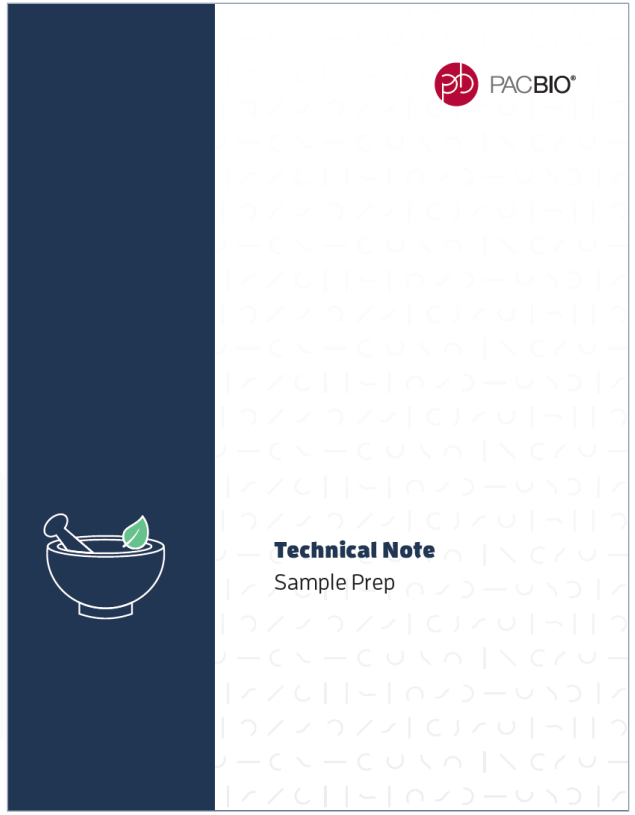
Invertebrates

When sampling from invertebrates, it is recommended to use a single individual. Some invertebrates have mucous membranes that inhibit the ability to obtain high-quality DNA. Please consider an extra cleanup step of the DNA if mucous-coated samples are used (see cleanup protocol here). It is also common to encounter invertebrate samples that are not easily separated from contaminants or do not have cell-dense tissues readily available. In these cases, sperm can be used as the input sample.

	Sample Type	Sample Storage
1	Cell-dense tissue (brain, kidney, muscle, etc.)	Fresh (within ~24 hours of collection and kept cold) or flash frozen with liquid nitrogen and stored at -80°C should be viable for several months.
2	Cell-culture	Room temperature cell-suspension or pellet is best entering the DNA isolation step. However, cryopreserved cells (slow frozen in cryoprotectant) should be viable at -80°C for several months.
3	Sperm	Fresh collected, room temperature sperm samples are best entering the DNA isolation step. However, customers have been successful with sperm frozen at -20°C.

Table 2 - Invertebrate sample types in order of preference

Note: It is NOT recommended to use liver tissue as input sample due to high abundance of enzymes that may degrade DNA.

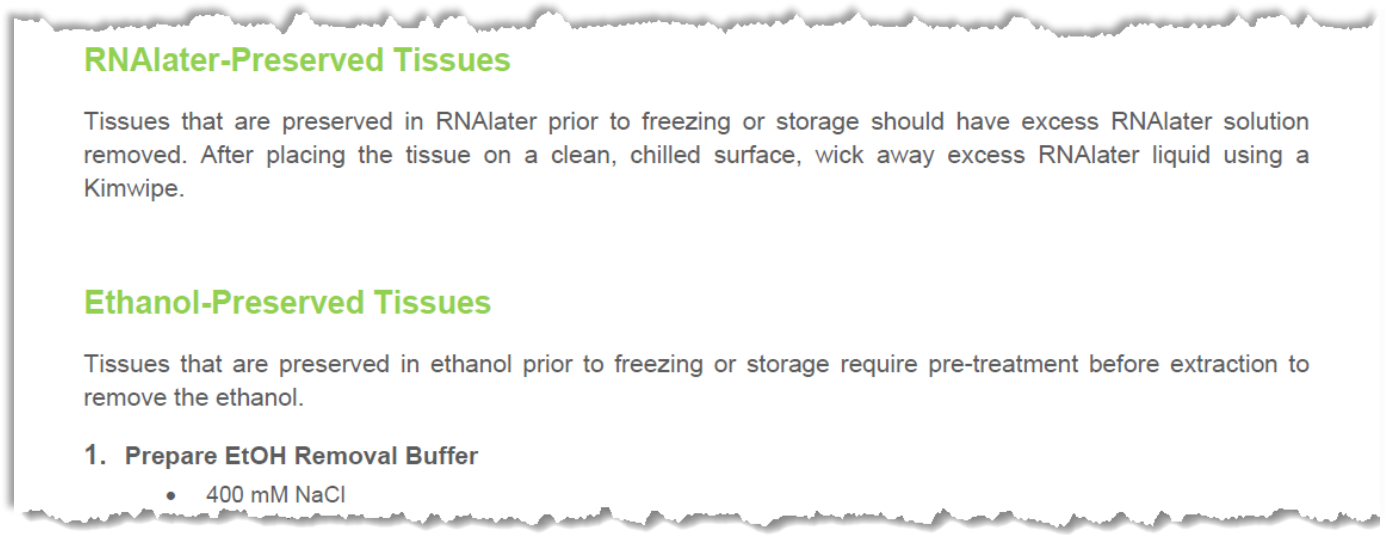


PacBio Technical note TN100-040518: Preparing samples for PacBio whole genome sequencing for *de novo* assembly – Collection and storage

Circulomics Nanobind kit handbooks

Nanobind Kit Handbooks contain information on HMW DNA isolation kit specifications, general tips, tissue preservation recommendations, extraction and sequencing performance, and troubleshooting tips.

- Nanobind CBB Big DNA Kit Handbook (HBK-CBB-001)
- Nanobind Plant Nuclei Big DNA Kit Handbook (HBK-PLT-001)
- Nanobind Tissue Big DNA Kit Handbook (HBK-TIS-001)



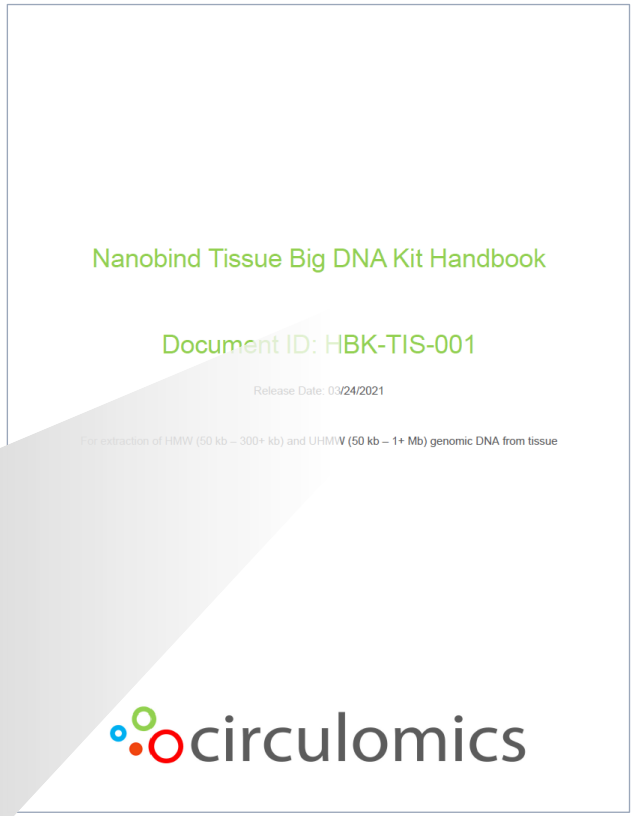
RNAlater-Preserved Tissues

Tissues that are preserved in RNAlater prior to freezing or storage should have excess RNAlater solution removed. After placing the tissue on a clean, chilled surface, wick away excess RNAlater liquid using a Kimwipe.

Ethanol-Preserved Tissues

Tissues that are preserved in ethanol prior to freezing or storage require pre-treatment before extraction to remove the ethanol.

1. **Prepare EtOH Removal Buffer**
   - 400 mM NaCl



Visit the Circulomics Nanobind Support website to find the latest resources for using Nanobind Kits for HMW DNA extraction and recommended tissue preservation methods.

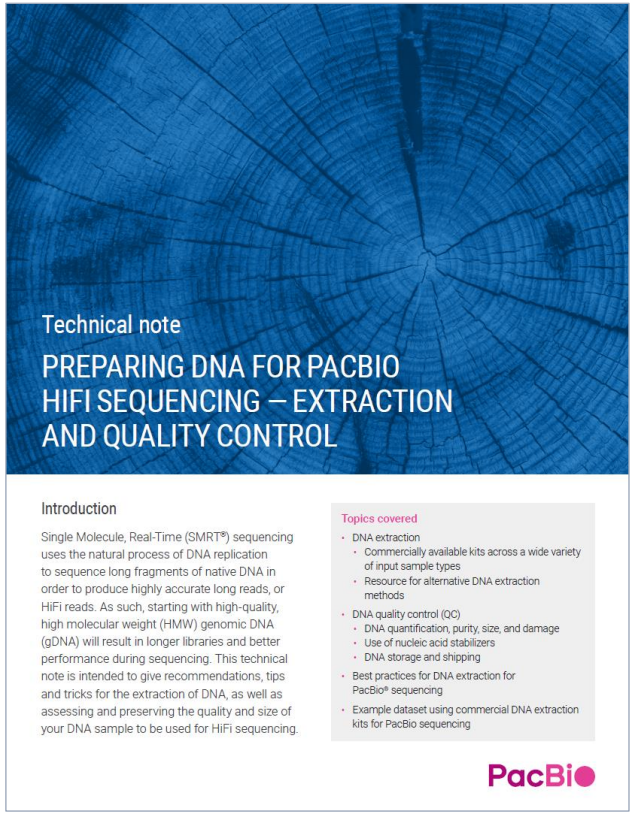
Genomic DNA extraction, QC and handling for SMRT sequencing

DNA extraction, QC and handling for SMRT sequencing whole genome sequencing projects

Starting with high-quality, high molecular weight (HMW) genomic DNA (gDNA) will result in longer libraries and better performance during sequencing

PacBio Technical note: DNA prep (102-193-651)

- Provides recommendations, tips and tricks for the extraction of genomic DNA, as well as assessing and preserving the quality and size of your DNA sample to be used for PacBio HiFi sequencing for *de novo* assembly
- Topics covered include:
  - DNA extraction
    - Commercially available kits across a wide variety of input sample types
    - Resource for alternative DNA extraction methods
  - DNA quality control (QC)
    - DNA quantification, purity, size, and damage
    - Use of nucleic acid stabilizers
    - DNA storage and shipping
  - General best practices for DNA extraction for PacBio sequencing
- This technical note also includes an example dataset for a California Redwood tree DNA sample that was isolated using the Nanobind Plant Nuclei Big DNA Kit (NB-900-801-01) from Circulomics



PacBio Technical Note: Preparing DNA for PacBio HiFi sequencing - Extraction and quality control (102-193-651)

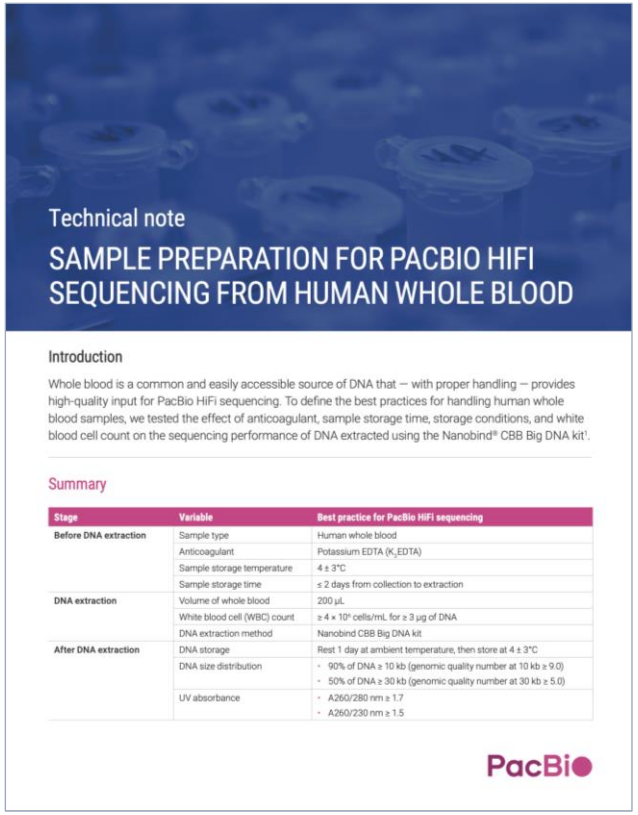
DNA extraction, QC and handling for SMRT sequencing whole genome sequencing projects (cont.)

Whole blood is a common and easily accessible source of DNA that – with proper handling – provides high-quality input for PacBio HiFi sequencing

PacBio Technical Note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

- To define the best practices for handling human whole blood samples, we tested the effect of anticoagulant, sample storage time, storage conditions, and white blood cell count on the sequencing performance of DNA extracted using the Nanobind CBB Big DNA Kit (NB-900-001-01).

Stage	Variable	Best practice for PacBio HiFi sequencing
Before DNA extraction	Sample type	Human whole blood
	Anticoagulant	Potassium EDTA (K_2EDTA)
	Sample storage temperature	4 ± 3°C
	Sample storage time	≤ 2 days from collection to extraction
DNA extraction	Volume of whole blood	200 µL
	White blood cell (WBC) count	≥ 4 × 10^6 cells/mL for ≥ 3 µg of DNA
	DNA extraction method	Nanobind CBB Big DNA kit
After DNA extraction	DNA storage	Rest 1 day at ambient temperature, then store at 4 ± 3°C
	DNA size distribution	• 90% of DNA ≥ 10 kb (genomic quality number at 10 kb ≥ 9.0) • 50% of DNA ≥ 30 kb (genomic quality number at 30 kb ≥ 5.0)
	UV absorbance	• A260/280 nm ≥ 1.7 • A260/230 nm ≥ 1.5



PacBio Technical Note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

DNA extraction, QC and handling for SMRT sequencing whole genome sequencing projects (cont.)

Circulomics offers innovative products for HMW DNA extraction to sequence common samples as well as more challenging samples such as animal tissue, insects, fungi, and plants

Circulomics Nanobind kit handbooks

- Nanobind Kit Handbooks contain information on HMW DNA isolation kit specifications, general tips, tissue preservation recommendations, extraction and sequencing performance, and troubleshooting tips.

Circulomics Nanobind protocols

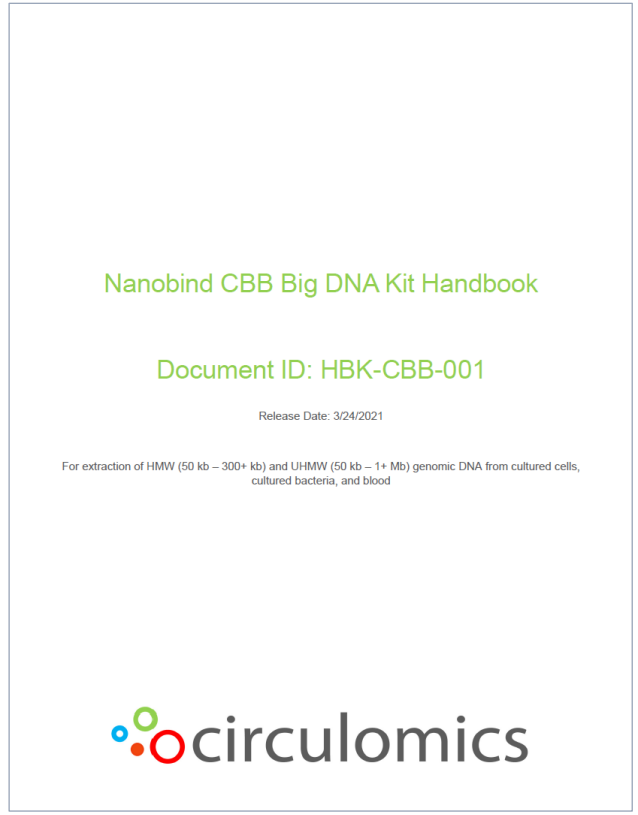
- Nanobind Protocols posted on the Circulomics website are always our most up to date versions and should take precedence over the Kit Handbooks. If an Application Note exists, that should be consulted first.

Circulomics Nanobind application notes

- Nanobind Application Notes provide detailed protocols for challenging or interesting samples with supporting sequencing data.

Circulomics Nanobind developmental protocols

- Developmental protocols exist for many sample types such as insect, worms, and fungal samples (contact Circulomics for details)



Visit the Circulomics Nanobind Support website to find the latest resources for using Nanobind Kits for HMW DNA extraction.

General recommendations for isolating high-molecular weight (HMW) genomic DNA

Before gDNA extraction:

- Use fresh or flash-frozen tissue
- Store flash-frozen tissue at -80°C and avoid freeze-thaw cycles
- Do not store blood samples longer than a few days at 4-8°C before DNA extraction
- Microbial gDNA Isolation:
  - ❑ Avoid culture incubation in complex or rich media
  - ❑ Harvesting from several replicate cultures rather than a single, high-density culture is preferred
  - ❑ Extraction of small culture volumes is preferred over large volumes to avoid accumulating high concentrations of potentially inhibiting secondary components

During gDNA extraction:

- Mechanically disrupt tissues using TissueRuptor (QIAGEN), Dounce homogenizer or liquid nitrogen grinding
- Inactivate nucleases and DNA binding proteins with a protease, such as proteinase K
- Remove all RNA with RNase A
- It is preferable to avoid oxidative agents such as phenol and/or chloroform if possible to minimize DNA damage
- Resuspend, or elute, DNA in a low salt buffer, such as 10 mM Tris-HCl pH 8.0-9.0 + 0.1 mM EDTA

General recommendations for isolating high-molecular weight (HMW) genomic DNA (cont.)

After gDNA extraction:

- Check concentration on both the NanoDrop and Qubit systems for concordance
- High-quality, pure DNA typically shows a A260/280 ratio ≥1.8 and a A260/230 ratio ≥2.0
- If gel purification is required, avoid using ethidium/UV based visualization methods. One alternative is to use SYBR Safe (Invitrogen) and visualize with blue light
- To help resuspend HWM DNA, pipette mix 1-10 times with a standard P200 pipette tip. Allow the DNA to rest overnight at 25°C.
- Overheating can introduce DNA damage. Inactivate DNase as recommended by the vendor kit. It is best to avoid heat inactivation when possible
- DNA storage conditions: 4°C (short-term); -20°C / -80°C (long-term)
- Ideally proceed to SMRTbell library preparation with freshly isolated DNA whenever possible

PacBio

Example protocols and kit solutions for high-molecular weight genomic DNA isolation

Example high-molecular weight genomic DNA isolation protocols and kit solutions

Cultured cells, cultured bacteria (gram negative / gram positive), and blood

Recommended kit product or protocol

Circulomics Nanobind CBB Big DNA Kit (NB-900-001-01)

- The Nanobind CBB Big DNA Kit is designed for rapid extraction of really big DNA from cultured cells, cultured bacteria (gram negative and gram positive), and blood (nucleated and non-nucleated).
- Includes protocols for HMW DNA (50 kb - 300+ kb) and UHMW DNA (50 kb - 1 Mb+) extraction.
- Typical yields are 5 µg - 100+ µg per extraction depending on input.
- A high-throughput version of this kit is available for automated extractions on Thermo Fisher KingFisher instruments. (Contact Circulomics for details.)

Other kit products or protocols*

QIAGEN Genomic-tip 20/100/500/G Kit [Link]

QIAGEN Gentra Puregene Kit [Link]

Geneaid DNA Isolation (Bacteria) Kit [Link]

Lucigen Masterpure Kit [Link]

JGI Bacterial Genomic DNA Isolation Protocol Using CTAB [Link]

NEB Monarch HMW DNA Extraction Kit [Link]

PacBio

* These products have not been extensively tested or validated by PacBio but are listed here as examples of other third-party kits or methods used by PacBio customers for isolating genomic DNA for SMRTbell library preparation.

Plant Tissue

Recommended kit product or protocol

Circulomics Nanobind Plant Nuclei Big DNA Kit (NB-900-801-01)

- The Nanobind Plant Nuclei Big DNA Kit is designed for rapid extraction of really big DNA from plant nuclei.
- First, nuclei are isolated from 1–5 g of plant tissue using one of the recommended nuclei isolation protocols. Then, HMW DNA is extracted from the nuclei using Nanobind disks. Each of the two purification steps (i.e., nuclei isolation + Nanobind extraction) removes different contaminants from the sample, resulting in clean, HMW DNA from even the most challenging plant species.
- For many plant types, big DNA (up to 300+ kb) can be obtained in <1 hour from the nuclei pellet stage.
- Typical yields are 5 µg - 20+ µg per extraction depending on input

Other kit products or protocols*

QIAGEN Genomic-tip 20/100/500/G Kit [Link]

Unsupported Protocol – Switchgrass (Panicum virgatum) DNA isolation [USDA] [Link]

Unsupported Protocol – DNA extraction of Chlamydomonas using CTAB [JGI] [Link]

QIAGEN User-Developed Protocol: Isolation of genomic DNA from plants and filamentous fungi using the QIAGEN Genomic-tip Kit [Link]

Modified QIAGEN Genomic-tip Protocol [King Abdullah University of Science and Technology] [Link]

* These products have not been extensively tested or validated by PacBio but are listed here as examples of other third-party kits or methods used by PacBio customers for isolating genomic DNA for SMRTbell library preparation.

Animal Tissue

Recommended kit product or protocol

Circulomics Nanobind Tissue Big DNA Kit (NB-900-801-01)

- The Nanobind Tissue Big DNA Kit is designed for rapid extraction of really big DNA from diverse animal tissues including mammalian, fish, avian, mollusk, and crustacean samples.
- Includes protocols for HMW DNA (50 kb - 300+ kb) and UHMW DNA (50 kb - 1+ Mb) extraction.
- Typical yields are 5 ug - 100+ ug per extraction depending on tissue type and input.
- Note: Insect, worm and fungal samples are not officially supported at this time. However, select insects, worms and fungi can be processed with supplemental buffers. Contact Circulomics for more details.

Other kit products or protocols*

QIAGEN Genomic-tip 20/100/500/G Kit [Link]

QIAGEN Gentra Puregene Kit [Link]

QIAGEN User-Developed Prototocol: Isolation of genomic DNA from mosquitoes or other insects using the QIAGEN Genomic-tip Kit [Link]

QIAGEN User-Developed Prototocol: Purification of archive-quality DNA from 10–20 mg fish tissue using the Gentra Puregene Tissue Kit or Gentra Puregene Mouse Tail Kit [Link]

Unsupported Protocol – Gentra Puregene Cell Kit (Qiagen) DNA Isolation [Univ. Washington] [Link]

Macherey-Nagel™ NucleoBond™ AXG 20/100/500 Gravity-flow Columns [Link]

PacBio

* These products have not been extensively tested or validated by PacBio but are listed here as examples of other third-party kits or methods used by PacBio customers for isolating genomic DNA for SMRTbell library preparation.

Yeast and Fungi

Recommended kit product or protocol

- Contact Circulomics for details on how to process select yeast and fungal samples with Nanobind Kits using supplemental buffers.

circulomics

Technical Support:
support@circulomics.com

Other kit products or protocols*

QIAGEN Genomic-tip 20/100/500/G Kit [Link]

QIAGEN Gentra Puregene Kit [Link]

QIAGEN DNeasy PowerLyzer PowerSoil Kit [Link]

GeneJET Plant Genomic DNA Purification Kit [Link]

Zymo Research Fungal/Bacterial DNA MidiPrep Kit [Link]

JGI Yeast and Fungal DNA Isolation Protocol [Link]

PacBio

* These products have not been extensively tested or validated by PacBio but are listed here as examples of other third-party kits or methods used by PacBio customers for isolating genomic DNA for SMRTbell library preparation.

Have challenging or interesting samples to process?

Please contact Circulomics for recommendations on how to process your challenging or interesting samples using Nanobind Kits.

circulomics
A PACBIO® COMPANY

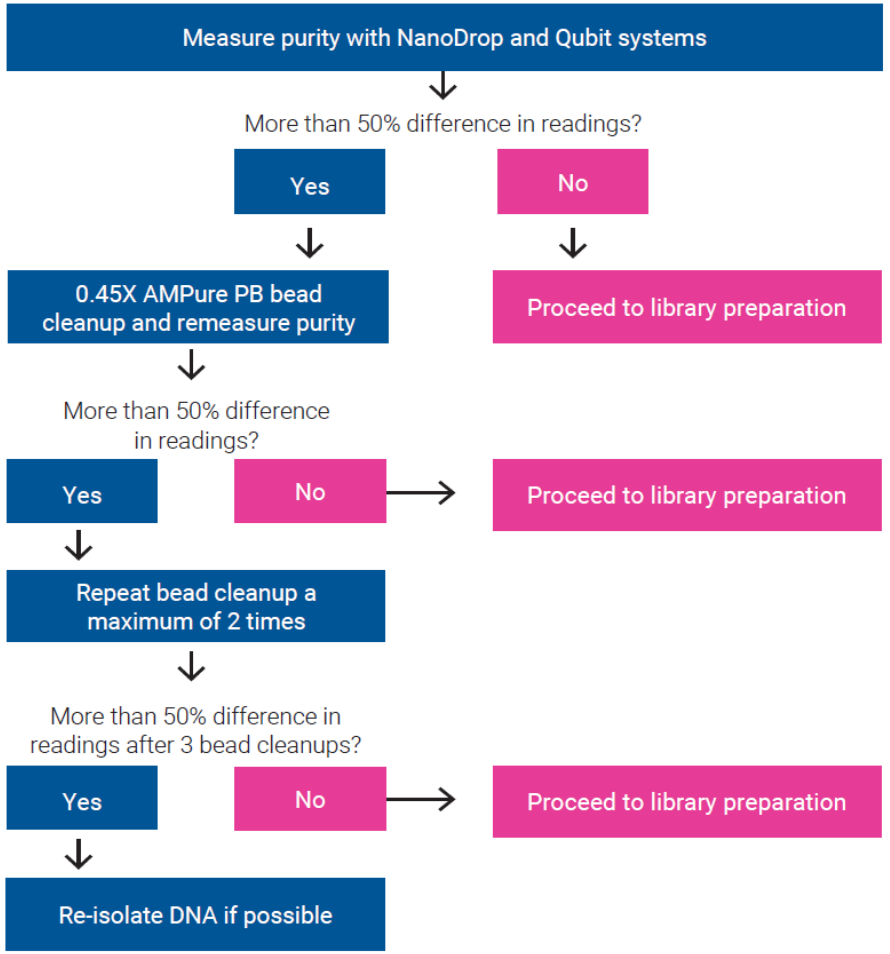
Technical Support:
support@circulomics.com

Cleanup of genomic DNA and SMRTbell libraries

General recommendations for cleanup of genomic DNA and SMRTbell libraries

AMPure PB beads can be used for cleanup of genomic DNA

- A quick and very effective check for sample purity is to compare the concentration readings between the NanoDrop Spectrophotometer and Qubit Fluorometer: High-quality DNA should show relative agreement in concentration measurements.
- If you observe a large difference in the concentration readings between the NanoDrop and Qubit systems, for example a difference of greater than or equal to 50% of the Qubit dsDNA assay reading, first check for RNA contamination using the Qubit RNA broad range assay.
- If there is no RNA contamination, then we recommend doing at least one to three rounds of AMPure PB bead purification until the concentrations are less than 50% different.
- If the agreement does not improve after three rounds of purification, try using either a commercial kit, isopropanol precipitation, or a new DNA extraction method to obtain a cleaner DNA sample.
- If there is RNA in the sample, then treat with RNase A followed by a round of AMPure PB bead purification.



Recommended cleanup process for isolated gDNA using 0.45X AMPure PB beads.* (1X AMPure PB beads may also be used.)

* See PacBio Technical Note: Preparing DNA for PacBio HiFi sequencing - Extraction and quality control (102-193-651)

Methods for evaluation of genomic DNA quality

Methods for evaluation of genomic DNA quality

Use recommended tools for evaluation of gDNA quality to generate optimal SMRT sequencing data quality

DNA Sizing QC

Use a Femto Pulse system or pulsed-field gel electrophoresis (PFGE) system for accurate DNA sizing QC of gDNA samples

- ☑ **High-quality, high-molecular weight gDNA** → Longer read lengths / higher data yields
- ☒ **Low-quality, degraded/damaged gDNA** → Shorter read lengths / lower data yields / lower library synthesis yields

DNA Purity QC

Use a NanoDrop instrument or other spectrophotometer device to determine DNA purity

- ☑ **High-quality, pure gDNA** → Longer read lengths / higher data yields
- ☒ **Low-quality, contaminated DNA** → Shorter read lengths / lower data yields / lower library synthesis yields

DNA Quantification QC

Use a Qubit fluorometric assay for accurate dsDNA quantitation

- ☑ **Accurate dsDNA quantitation** → Optimal library construction yields / higher data yields
- ☒ **Inaccurate dsDNA quantitation** → Lower library construction yields / lower data yields

PacBio

Genomic DNA sizing characterization

Femto Pulse System [Agilent Technologies]

Highly Recommended

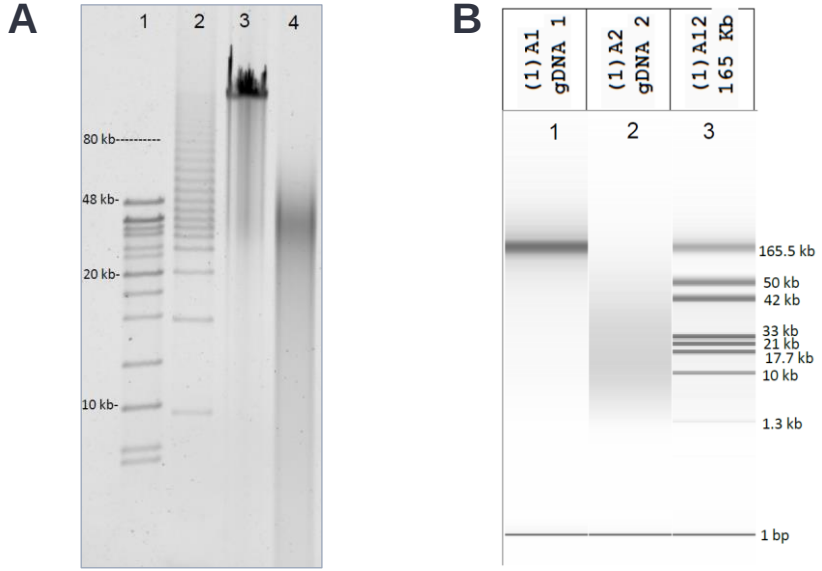
Resolves up to ~165 kb
Requires <1 ng of sample
<1.5-hour analysis time

PippinPulse System [Sage Science]

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

CHEF Mapper XA System [Bio-Rad]

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



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 genomic DNA.

Refer to the appropriate PacBio Procedure & checklist protocol documentation for recommended minimum input gDNA fragment size distributions for specific whole genome sequencing applications

DNA purity determination

- DNA purity can be determined by using a NanoDrop system [Thermo Fisher Scientific] or other spectrophotometer tool
- For ultrapure DNA, A260/280 ratio is typically between ~1.8 - 2.0 and A260/230 ratio is ≥2.0
- High UV absorbance values are *not* always a guarantee of optimal sequencing performance because not all inhibitors absorb at the wavelengths of 230, 260, and 280 nm.
- Conversely, low UV absorbance values are *not* always a guarantee that non-optimal sequencing performance will be obtained for a sample
  - PacBio has found that gDNA samples with A260/280 ratios ≥1.7 and A260/230 ratios ≥1.5 can still generate excellent HiFi sequencing performance.*
  - For WGS samples with absorbance ratios outside this range, we recommended performing a 1X AMPure PB bead cleanup to remove potential contaminants.

A260/A280 Ratio

- A low A260/A280 ratio may be the result of:
  - ❑ Protein
  - ❑ Phenol
  - ❑ 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

If A260/280 and A260/230 readings are out of the recommended ranges, perform one or more rounds of **AMPure PB bead purification** followed by re-assessment of the quantity and purity of the input DNA sample.***

A260/A230 Ratio

- A low A260/A230 ratio may be the result of:
  - ❑ Protein**
  - ❑ 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

* See PacBio Technical Note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

** See NEB Technical Note: A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers (2019)

*** See PacBio Technical Note: Preparing DNA for PacBio HiFi sequencing - Extraction and quality control (102-193-651)

DNA quantification

- Accurate measurement 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 system [Thermo Fisher Scientific] for DNA quantitation
  - Can use a Qubit dsDNA broad range (BR) assay kit for initial genomic DNA QC evaluation
  - Use a Qubit dsDNA high sensitivity (HS) assay kit for routine DNA quantitation during SMRTbell library construction and sequencing preparation using SMRT Link Sample Setup.
- 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, we recommend doing an AMPure PB bead purification step, 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 or more rounds of AMPure PB bead purification is not reduced, we recommend trying another cleanup approach before a re-measurement with both methods.*

* See PacBio Technical Note: Preparing DNA for PacBio HiFi sequencing - Extraction and quality control (102-193-651)

Storage and shipping of genomic DNA and SMRTbell libraries

Guidelines for storage and shipping of genomic DNA and SMRTbell libraries

High-molecular weight genomic DNA storage

- Very clean HMW gDNA can be stored at 4°C for weeks with no degradation. It can also be stored a few days at 25°C.
- HMW gDNA can also be frozen at -20/-80°C for extended storage. Avoid freeze/thaw cycles.

SMRTbell library 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 generally 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

* **Note:** Genomic DNA extracted with Circulomics Nanobind Kits can also be shipped in liquid form on wet ice.

DNA sample extraction literature resources

DNA sample extraction documentation & other literature

Technical notes

- Technical note: Preparing samples for PacBio whole genome sequencing for de novo assembly – Collection and storage (TN100-040518)
- Technical Note: Preparing DNA for PacBio HiFi sequencing – Extraction and quality control (102-193-651)
- Technical note: Sample preparation for PacBio HiFi sequencing from human whole blood (102-326-500)

Nanobind kit handbooks

- Nanobind CBB Big DNA Kit Handbook (HBK-CBB-001)
- Nanobind Plant Nuclei Big DNA Kit Handbook (HBK-PLT-001)
- Nanobind Tissue Big DNA Kit Handbook (HBK-TIS-001)

Nanobind protocols

- HMW DNA Extraction – Animal Tissue Protocol (Standard Dounce Homogenizer) (EXT-DHH-001)
- HMW DNA Extraction – Animal Tissue Protocol (Standard TissueRuptor) (EXT-TRH-001)
- HMW DNA Extraction – Cultured Cells Protocol (EXT-CLH-001)
- HMW DNA Extraction – Gram Negative Bacteria Protocol (EXT-GNH-001)
- HMW DNA Extraction – Gram Positive Bacteria Protocol (EXT-GPH-001)
- HMW DNA Extraction – Mammalian Whole Blood Protocol (200 µL) (EXT-BLH-001)
- HMW DNA Extraction – Nucleated Blood Protocol (EXT-NBH-001)
- HMW DNA Extraction – Plant Nuclei Protocol (EXT-PLH-001)
- Nuclei Isolation – Plant Tissue Protocol (TissueRuptor) (NUC-TRP-001)
- Nuclei Isolation – Plant Tissue Protocol (LN2) (NUC-LNP-001)

DNA sample extraction documentation & other literature (cont.)

Nanobind application notes

- cryoPREP Tissue Homogenization Application Note [Link]
- Human Breast Application Note [Link]
- Mammalian Brain Application [Link]
- Mammalian Liver Application [Link]
- Mammalian Spleen Application Note [Link]
- Fish Testis Application Note [Link]
- Snail Application Note [Link]
- Crab Application Note [Link]
- Fish Skeletal Muscle Application Note [Link]
- Aplysia Application Note [Link]

Have challenging or interesting samples to process?

Please contact Circulomics for recommendations on how to process your challenging or interesting samples using Nanobind Kits.

Nanobind automation scripts [contact Circulomics]

- Nanobind KF CBB Big DNA Kit Automated Reference Guide
- Automated HMW DNA Extraction – Cultured Cells
- Automated HMW DNA Extraction – 200 µL Mammalian Blood
- Automated HMW DNA Extraction – 1 mL Mammalian Blood
- Automated HMW DNA Extraction – Gram-Negative Bacteria
- Automated HMW DNA Extraction – Gram-Positive Bacteria
- Automated HMW DNA Extraction – Animal Tissue

circulomics
A PACBIO® COMPANY

Circulomics Technical Support: support@circulomics.com

DNA sample preparation online resource

Literature resource for sample collection and DNA extraction protocol references

The listing below is a collection of publications by the scientific community describing extraction protocols for high-molecular weight DNA followed by PacBio sequencing. When possible, the links point directly to the methods section (or supplementary information).

Animals Plants Fungi Protists

If you have protocols you would like to share, or have questions about DNA extraction for PacBio sequencing, p
contact ExtractDNA@pacb.com.

Animals

- Invertebrates
  - Panova2016 – DNA extraction protocols for whole-genome sequencing in marine organisms
  - *microinvertebrates*
    - Laumer2020 – protocols for diverse meiofauna species, including *C. elegans* & *Castrella truncata* (SMRT Leiden confere presentation)
  - *arthropods*
    - arachnids
      - Guerrero2019 – The Pacific Biosciences de novo assembled genome dataset from a parthenogenetic New Ze wild population of the longhorned tick, *Haemaphysalis longicornis* Neumann, 1901
      - Spider, *Stegodyphus dumicola*
      - p scab mite, *Psoroptes ovis*
      - *atophagoides pteronyssinus* supports identification of novel allergen
      - n for the pathogen vector *Ixodes ricinus* by ultra-long single
    - crustaceans
      - Liu2019 – Crab muscle extraction & sequencing (Circulomics application note)
      - Baldwin-Brown2018 – A New Standard for Crustacean Genomes: The Highly Contiguous, Annotated Genome Assembly of the Clam Shrimp *Eulimnadia texana* Reveals HOX Gene Order and Identifies the Sex Chromosome
      - Gonçalves Athanasio2016 – Optimisation of DNA extraction from the crustacean *Daphnia*

Methods

Sample collection

A female yellowbelly pufferfish (Fig. 2), reared in the fish breeding centre of Fujian Normal University in Fuzhou City of Fujian Province was used for genome sequencing and assembly. Fresh white muscle, eye, skin, gonad, gut, liver, kidney, blood, gall bladder and air bladder tissues were collected and quickly frozen in liquid nitrogen for one hour. White muscle tissues were used for DNA sequencing for genome assembly, while all tissues were used for transcriptome sequencing.

Fig. 2

A picture of the yellowbelly pufferfish used in the genome sequencing and assembly.

DNA and RNA sequencing

Genomic DNA from white muscle tissue was extracted using the

Plants

t genomic DNA for long-read sequencing of single molecules

ce Genome With Single-Molecule Sequencing Uncovers a Recent Burst of Genes

ssembly reveals the structure of the *Arabidopsis thaliana* Nd-1 genome and its gene set

sive, high-throughput plant genomic DNA extraction method suitable for

ng and

ect run

blolly pu

ce for the Fast-Growing Microalga *Picochlorum celeri*

Undaria pinnatifida: Chromosome-Level Assembly Using PacBio and Hi-C

ne suggests convergent functions of homeobox genes in algae and land plants (algae)

whole-genome sequencing in marine organisms (algae)

Fungi

- Faure2019 – Long-Read Genome Sequence of the Sugar Beet Rhizosphere Mycoparasite *Pythium oligandrum*
- Nagappan2018 – Improved nucleic acid extraction protocols for *Ganoderma boninense*, *G. miniatocinctum* and *G. tornatu*
- Schwessinger2017 – Extraction of high molecular weight DNA from fungal rust spores for long read sequencing
- Solomon2016 – Robust and effective methodologies for cryopreservation and DNA extraction from anaerobic gut fungi
- Sonnenberg2016 – A detailed analysis of the recombination landscape of the button mushroom *Agaricus bisporus var. bisporus*

www.ExtractDNAforPacBio.com

circulomics
A PACBIO® COMPANY

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 or Circulomics for more discussions around your particular species & sequencing project!

www.pacb.com