PACBIO®

Technical Note

DNA Prep

Preparing DNA for PacBio HiFi Sequencing – Extraction and Quality Control

Introduction

Single Molecule, Real-Time (SMRT®) Sequencing uses the natural process of DNA replication to sequence long fragments of native DNA in order to produce highly accurate long reads, or HiFi reads. As such, starting with high-quality, high molecular weight (HMW) genomic DNA (gDNA) will result in longer libraries and better performance during sequencing. This technical note is intended to give recommendations, tips and tricks for the extraction of DNA, as well as assessing and preserving the quality and size of your DNA sample to be used for HiFi sequencing.

Topics Covered

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

Best Practices for DNA Extraction for PacBio® Sequencing

Example Dataset Using Commercial DNA Extraction Kits for PacBio Sequencing

DNA Extraction

Simple and easy-to-use commercial DNA extraction kits are an excellent way to produce HiFi reads on the Sequel® and Sequel II Systems. The following are a list of kits that our customers have used in combination with the SMRTbell® Express Template Prep Kit 2.0 protocol for high yields and consistent sequencing results. Sample amounts and typical yields are based on each kit's manufacturer. Check with each manufacturer for the most up-to-date information and recommendations. DNA yields will vary by sample quality, type, and input amount.

Kit	Method	Sample Amount	Typical Yields
QIAGEN MagAttract HMW DNA Kit	Magnetic bead	• Blood: 200 mL • Bacterial cells: 2 x 10^9 • Tissue: up to 25 mg	• Blood: 4-8 µg • Gram-negative bacteria: up to 14 µg • Gram-positive bacteria: up to 3.5 µg • Tissue: 0.5-2.8 µg / mg tissue
QIAGEN PAXgene Blood DNA Kit	Precipitation	• Blood: 8.5 mL	• 150-500 µg depending on the number of nucleated cells
QIAGEN Gentra Puregene Kit	Precipitation	• Cells: 106 - 2.2*10^7 • Tissue: up to 100 mg	• 7 µg per 1 million cells • Tissue: 5-100 µg
QIAGEN Genomic-tip 20/G Kit	Anion exchange column	• Blood: 1 mL • Cells: 5 x 10^6 • Yeast: 1.5 x 10^9 • Bacteria: 4.5 x 10^9 • Tissue: up to 20 mg	• 1-20 µg
Circulomics Nanobind CBB Kit	Nanobind disc	• Blood: 200 mL • Cells: 1 x 10^6 • Bacteria: 5 x 10^8 - 5 x 10^9	• 5-34 µg depending on sample type and input amount

Kit	Method	Sample Amount	Typical Yields
Circulomics Nanobind Tissue Big DNA Kit	Nanobind disc	• Depends on tissue and preservation method; typically around 25 mg	• 5-100 µg
Circulomics Nanobind Plant Nuclei Big DNA Kit	Nanobind disc	• Up to 10 g	• 5-20 µg
Lucigen MasterPure Kit	Precipitation	• Cells: 1 x 10^6 • Blood: 200 mL • Bacteria: 3.5 x 10^6	• Cells: 3-12 µg • Blood: 3-9 µg • Bacteria: 1.3-1.6 µg
NEB Monarch Genomic DNA Purification Kit	Anion exchange column	• Blood: 100 mL • Bacteria: 2 x 10^9 • Cells: 5 x 10^6 • Tissue: 10 mg	• Blood: 2.5-4 µg • Gram-negative bacteria: 6-10 µg • Gram-positive bacteria: 6-9 µg • Mammalian cells: 7-9 µg • Tissue: 5-30 µg
Macherey-Nagel NucleoBond HMW DNA Kit	Anion exchange column	• Blood: 2 mL • Plant: 1.5 g of leaves • Bacteria: up to 100 mg • Cells: 10^7 • Animal tissue: up to 300 mg	• Dependent on sample type and input amount
QIAGEN DNeasy PowerMax Soil Kit	Bead-beating; Anion exchange	• Soil: up to 10 g	• Sample dependent
QIAGEN QIAamp PowerFecal DNA Kit	Bead-beating; Anion exchange	• Stool or biosolids: 250 mg	• Sample dependent

There are many alternative methods of DNA extraction not covered above. See www.extractdnaforpacbio.com as a resource database of customer protocols for DNA extraction for different organisms.

DNA QC

There are multiple aspects of DNA quality that can have an impact on the final sequencing result. Below are important guidelines and considerations to follow for the different aspects of DNA quality. This includes measuring DNA concentration, assessing DNA purity and molecular weight, and minimizing DNA damage.

Quantification

We recommend using a fluorometric method that specifically measures only double-stranded DNA (dsDNA). The most accurate and convenient way to do this is to use the Qubit® fluorometer (ThermoFisher Scientific) with either the Broad Range (BR) or High Sensitivity (HS) dsDNA assay kits. The NanoDrop® Spectrophotometer (ThermoFisher Scientific), commonly used for DNA quantification, is good for assessing DNA purity (see below) but measures all nucleic acid in solution and therefore frequently overestimates the true concentration of dsDNA in a sample.

If you are trying to quantify HMW DNA that is viscous, typically >100 kb in size on average, then we recommend diluting a small aliquot and pipette up and down 10-20 times to shear and better homogenize the DNA in solution. For even more accurate quantification, consider making 3 aliquots and take the mean of all three readings. The CV between the three readings should be less than 30%. If not, try additional mixing of the samples to homogenize the DNA, and repeat quantification.

Purity

DNA purity is the single largest factor affecting the success of your sequencing experiment. We recommend assessing DNA purity by measuring the absorbance at 230, 260, and 280 nm using the NanoDrop system. Pure DNA will have an A260/280 absorbance ratio of 1.8, and an A260/230 ratio of 2.0 or higher. A260/280 ratio below 1.8 indicates contaminating protein or phenol that is absorbing at 280 nm, or noise from a very low DNA concentration. High ratios (>1.8) may not indicate a problem, but the spectral profile should be reviewed for abnormalities. Common causes of a low A260/230 ratio are carryover contaminants such as carbohydrates (e.g. in mature plant tissue), phenol, residual guanidine or ethanol from a column-based kit, and residual glycogen from precipitation. A high A260/230 ratio doesn’t indicate carryover contaminants, rather it may indicate blanking with the wrong solution, not

PACBIO®

having the instrument well-calibrated, or a dirty pedestal. Again, the spectral profile should be reviewed for any abnormalities. In addition, absorbance measurements on the Nanodrop system are sensitive to pH, so ensure that the DNA is suspended in a proper buffer solution such as the PacBio Elution Buffer.

Another quick and very effective check for sample purity is to compare the concentration readings between the NanoDrop Spectrophotometer and Qubit Fluorometer. Despite the difference between how the systems estimate concentration, pure, high-quality DNA should show relative agreement in concentrations. 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 a standard 0.45X AMPure® PB bead purification until the concentrations are less than 50% different[1]. If the agreement does not improve after three rounds of purification, try using either a commercial kit, isopropanol precipitation, or new DNA extraction method to get cleaner DNA. If there is RNA in the sample, then treat with RNase A followed by a standard 0.45X AMPure PB bead purification.

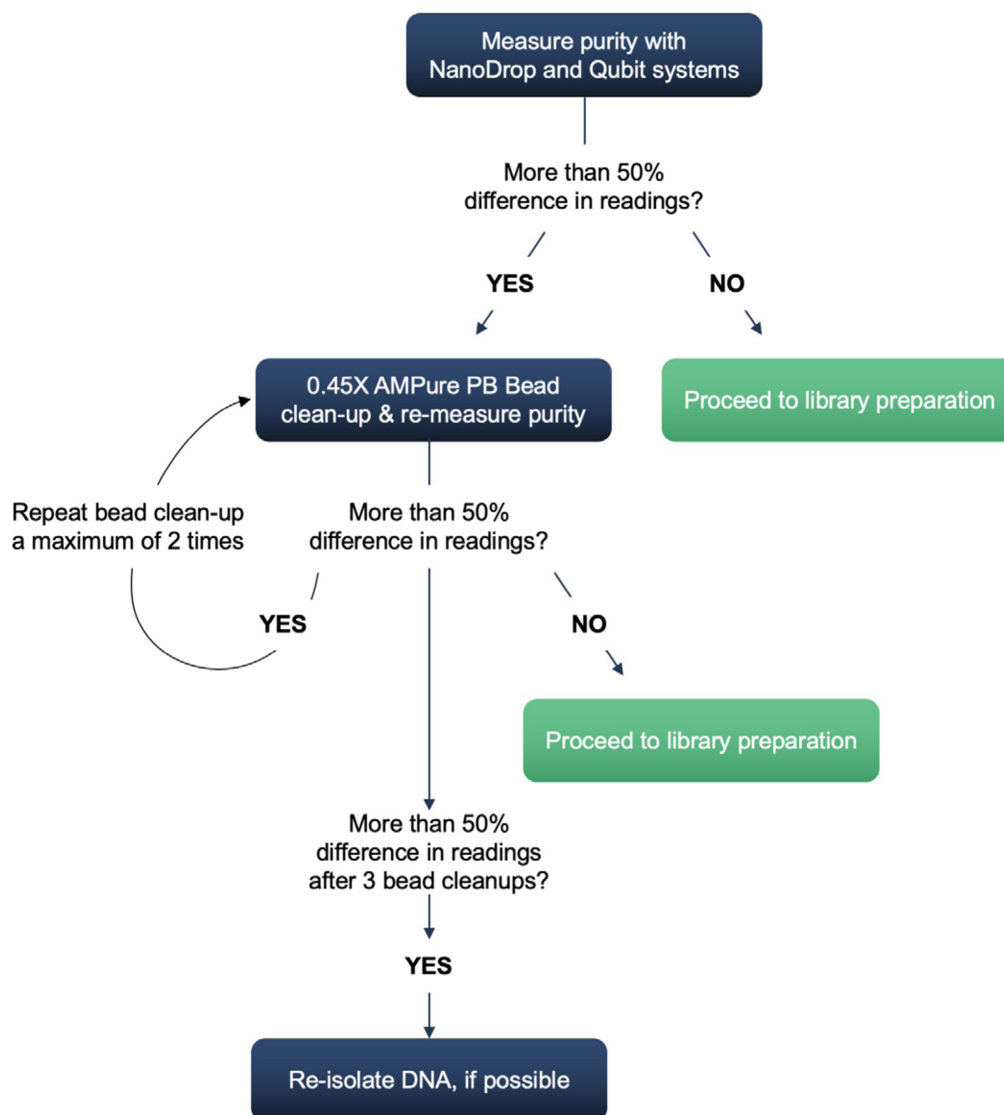

Figure 1. *Recommended cleanup process for isolated DNA.*

1 See the Concentrate DNA using AMPure® PB Beads section of the Preparing HiFi SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0 protocol for a 0.45X AMPure® PB cleanup procedure.

Size

Our new HiFi read-based applications on the Sequel II System have made gDNA molecular weight requirements less stringent than previous applications that relied on traditional long-read sequencing. Our current recommended HiFi library insert sizes for *de novo* assembly range from 10–25 kb. Despite the smaller insert sizes, we still recommend starting with gDNA >40 kb in size, on average, so that the size distribution of the library can be more narrowly focused with DNA shearing. In addition, the molecular weight of a sample is a good indicator of whether a sample has experienced significant damage. For example, degraded samples are much more likely to have significant DNA damage that can reduce read length and yield during sequencing, in addition to having overall lower insert sizes. Starting with a HMW sample will ensure that the sample is high quality and will sequence well. Please see individual library preparation protocols for application-specific gDNA size recommendations and the best tools to use to accurately measure the molecular weight.

In addition, customers have reported that allowing HMW DNA pellets to resuspend (in a low-salt buffer) at room temperature for one to three days improves the yield of DNA in solution and average molecular weight.

Also, precipitation and nanobind disks methods tend to yield higher molecular weight DNA than anion-exchange (column) and magnetic bead-based methods. This is primarily due to the latter methods having greater mechanical forces that shear the gDNA. Moreover, gentle handling of the gDNA, the use of wide-bore pipette tips, and avoiding vortexing will all help improve the overall molecular weight of the DNA.

DNA Damage

Nuclease Inactivation

We recommend using a DNA extraction method that includes proteinase K, or another type of protease, during lysis to inactivate native nucleases. Substituting a chaotropic agent for a protease, such as guanidinium thiocyanate, has resulted in samples with reduced sequencing performance. For tissue samples, it's important to flash-freeze the sample immediately upon extraction, or death, in order to prevent DNA degradation. Moreover, we also recommend cutting the tissue into small pieces, or grinding to a fine powder in liquid nitrogen, in order to allow proteinase K to reach the lysed cells.

Phenol/Chloroform

We also recommend avoiding extraction methods that use phenol/chloroform. These chemicals are strong oxidizers and can significantly damage the DNA if not properly used. For example, guanine is particularly sensitive to oxidation and exposure to phenol/chloroform can form 8-oxo-G bases. Despite the risks, we understand that in some circumstances phenol/chloroform may be required to remove certain compounds from organisms where DNA extraction is a challenge. If you are working with an organism for which the DNA extraction method requires the use of phenol/chloroform, then take care not to expose the DNA to these agents longer than necessary and consider using phase-lock tubes to better separate the aqueous and organic layers.

Hydrolysis of N-glycosyl Bonds

The most frequent type of DNA damage is the hydrolysis of the N-glycosyl bond, which creates an abasic site. These sites can then further degrade into nicks. Nicks in turn can make the DNA more susceptible to fragmentation. If not repaired, nicks will also adversely affect sequencing performance. Exposure to heat or changes in pH will accelerate the hydrolytic reactions. When working with HMW DNA it is best to minimize incubation times at high temperatures and always store the DNA in a buffered solution. Regardless, abasic sites and nicks will accumulate over time, even in lyophilized material, since it's practically impossible to rid the DNA of H_2O, the reactive species. For best possible sequencing results always use freshly isolated DNA.

Tissue and Nucleic Acid Stabilizers

In general, tissue preservatives such as ethanol, DMSO, or a commercial solution result in significantly lower quality DNA than using fresh or flash-frozen tissue. If using a preservative is necessary, we recommend storing the tissue at 4°C, as opposed to room temperature (~21°C). This should improve the molecular weight and quality of the DNA isolated from the tissue.

We have not tested commercial nucleic acid stabilizes (e.g. DNAgard and RNAlater) on sequencing performance, and therefore cannot support their use with DNA or RNA samples at this time.

DNA Storage and Shipping

Pure gDNA suspended in a buffered solution, such as PacBio Elution Buffer, can be safely stored at 4°C for several weeks to months. For long-term storage (months) we recommend freezing the DNA at -20°C or below and avoid freeze-thaw cycles. Freezing will induce some low-level of shearing from the ice-crystals.

When shipping, we recommend DNA and SMRTbell libraries be shipped frozen on dry ice. Keeping the DNA, or library, frozen helps prevent shearing which can occur from the jostling experienced during shipping.

Summary of Best Practices for DNA Extraction for PacBio HiFi Sequencing

- 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 week at 4-8°C before DNA extraction
- Grind tissue samples to a fine powder in liquid nitrogen
- Inactivate nucleases and DNA binding proteins with a protease, such as proteinase K
- Remove all RNA with RNase A
- Avoid guanidinium or guanidine thiocyanate (proteases are favored over chaotropic agents)
- Avoid oxidative agents such as phenol and/or chloroform to minimize DNA damage
- Resuspend, or elute, DNA in a low salt buffer, such as 10 mM Tris-HCL pH 8.5-9.0
- Check concentration on both the NanoDrop and Qubit systems for concordance
- Ensure DNA is pure with $A_{260/280}$ ratio at 1.8, and a $A_{260/230}$ ratio greater or equal to 2.0
- Proceed to SMRTbell library preparation with freshly isolated DNA

Example: California Redwood Sample using Circulomics Nanobind Plant Nuclei Big DNA Kit

The Circulomics Nanobind Plant Nuclei Big DNA Kit was used to isolate DNA from 10 g of fresh leaf tissue of *Sequoia sempervirens*. The DNA library was prepared following the protocol titled Preparing HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0 and a >17 kb fraction was size-selected on the BluePippin system (Sage Sciences). Sequencing was performed on the Sequel II System with v8.0 software and sequencing chemistry 2.0. Sequencing run time was 30 hours, with a 1-hour pre-extension prior to acquisition. CCS analysis was performed in SMRT® Link v8.0 with default parameters using a minimum 3 pass, Q20 filter. Shown is data from a single SMRT® Cell 8M loaded at 60 pM.

Sample	Total Bases (Gb)	Polymerase Read Length N50 (bp)	HiFi Reads	HiFi Sequencing Yield (bp)	HiFi Read Length Mean (bp)	HiFi Read Quality (median)
S. sempervirens	405.2	176,258	1,269,680	30,633,954,289	24,127	Q27

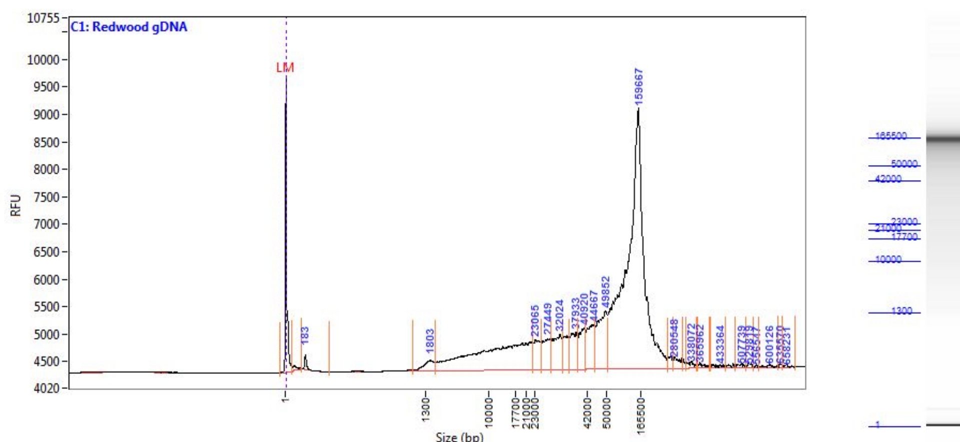

Figure 2. *Femto® Pulse system (Agilent) electropherogram of the S. sempervirens gDNA isolated with the Circulomics Nanobind Plant Nuclei Big DNA Kit. The majority of the gDNA is >100 kb in size.*

PACBIO®

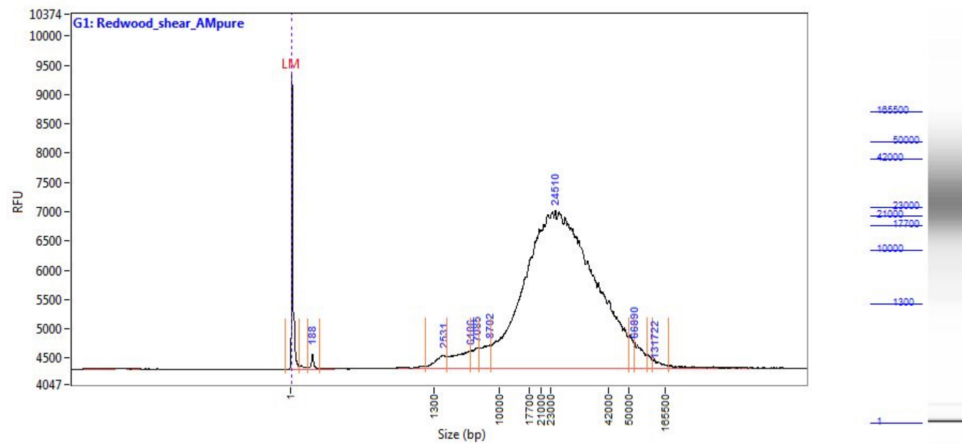

Figure 3. *Femto Pulse system (Agilent) electropherogram of the S. sempervirens gDNA sheared to a mode of 25 kb on the Megarupter 3 system (Diagenode).*

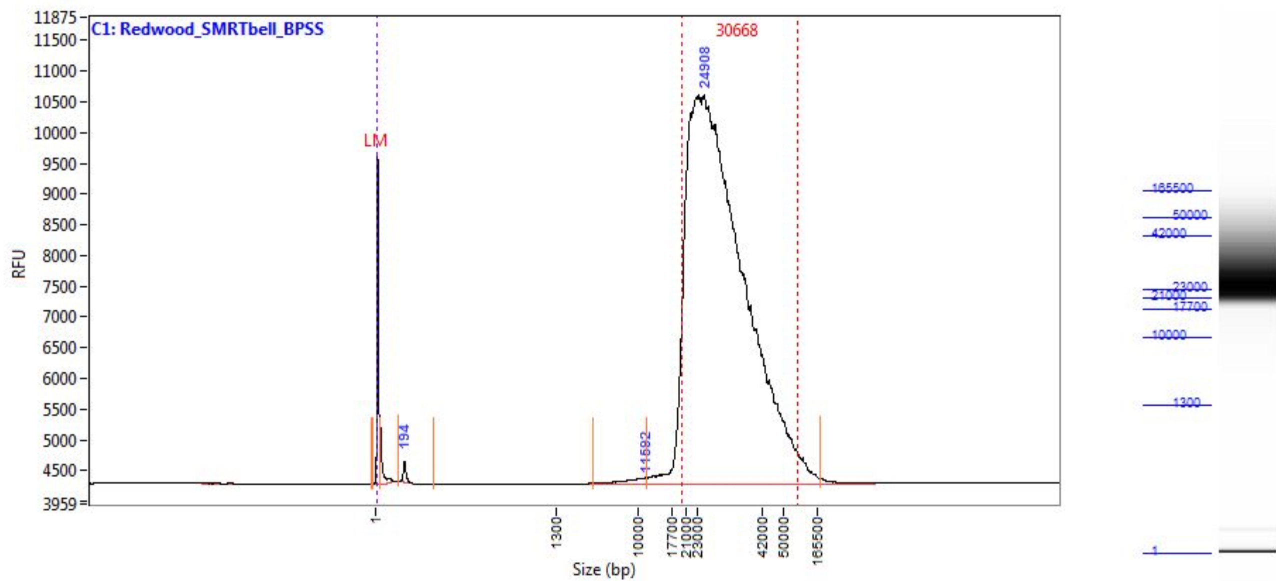

Figure 4. *Femto Pulse system (Agilent) electropherogram of the S. sempervirens size-selected SMRTbell library. Black and blue lines are the size-selected SMRTbell libraries using the Sage BluePippin system with a high-pass 17 kb cutoff. The red and orange lines are the pre-size-selected SMRTbell library.*

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at http://www.pacb.com/legal-and-trademarks/terms-and-conditions-of-sale/.
Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. Pacific Biosciences does not sell a kit for carrying out the overall No-Amp Targeted Sequencing method. Use of the No-Amp method may require rights to third-party owned intellectual property. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. Femto Pulse and Fragment Analyzer are trademarks of Agilent Technologies Inc. All other trademarks are the sole property of their respective owners.