Product note

FAST, ACCURATE DNA SIZING WITH THE AGILENT FEMTO PULSE SYSTEM FOR HIFI WGS

The Agilent Femto Pulse system is the recommended technology for the most accurate measurement of high molecular weight DNA for long-read sequencing on PacBio® HiFi systems. Accurately measuring the size distribution of your genomic DNA or SMRTbell® library provides more informative quality control (QC) checks and ensures the success of your long-read projects. The Femto Pulse system provides high-resolution genomic DNA sizing of 11 samples plus ladder through 165 kb in under 1.5 hours (Table 1 and Figure 2). The instrument automates analysis from three standard 96-well sample plates, and can process more than 190 samples, hands-free, in a 24-hour period. There are 3 points in the HiFi sequencing prep workflow where a QC check point is recommended (Figure 3).

Key benefits of the Agilent Femto Pulse system

- High-resolution DNA sizing through 165 kb in under 1.5 hours
- Ultra-sensitive analysis using only picograms of DNA
- Reduced project cost with accurate genomic DNA and SMRTbell library QC

Figure 1. **Agilent Femto Pulse system**

Size resolution	Up to 165 kb
Run time	<1.5 hours
Samples per run	11 + ladder
DNA input	500 pg
Recommended kit	Agilent FP-1002, Genomic DNA 165 kb kit

Table 1. Specifications for the Agilent Femto Pulse system for analyzing high molecular weight DNA.

Genomic DNA QC

Prior to preparing samples for HiFi sequencing, it is recommended to measure the size distribution of the genomic DNA using the Femto Pulse system, a parallel capillary electrophoresis instrument that uses a pulsed field power supply. Knowing the fraction of DNA that is high vs lower molecular weight informs performance expectations (e.g., yield and coverage) for the experiment under consideration. Technologies other than the Femto Pulse system may not as reliably predict long-read sequencing performance. The larger the fraction of high molecular weight DNA, the better the resulting DNA size distribution will be after mechanical shearing, leading to higher read lengths and sequencing coverage.

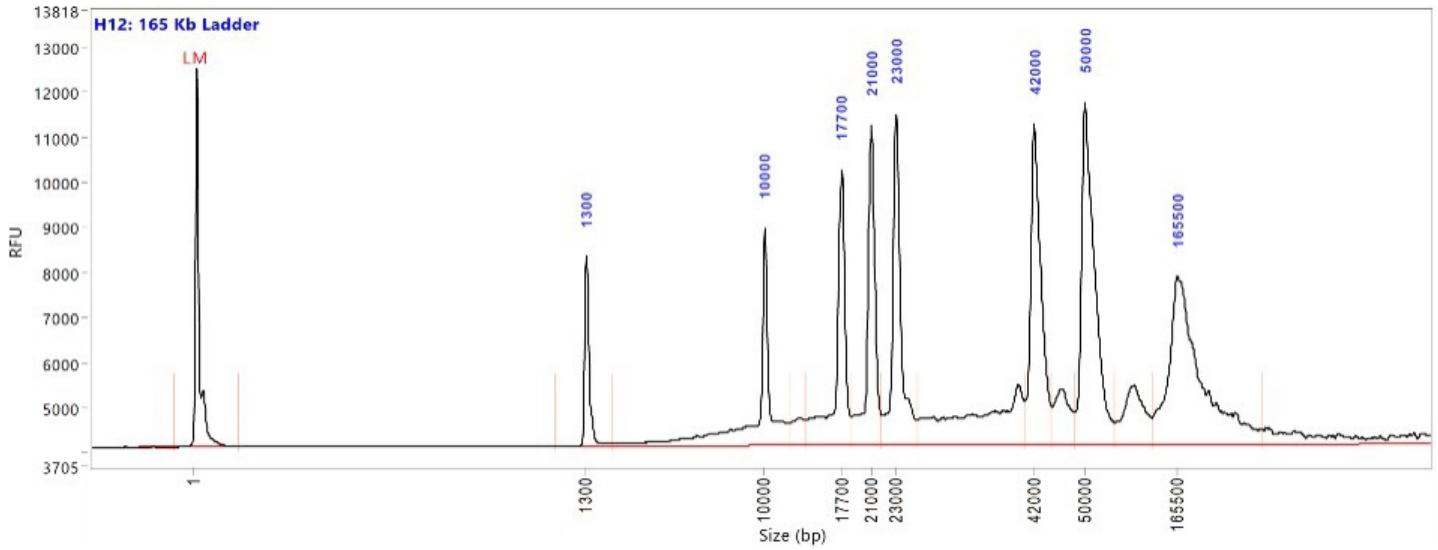


Figure 2. The FP 165 kb Ladder shows the ability of the Agilent Femto Pulse system to separate large DNA fragments up to 165 kb in 1.5 hours.

Agilent has developed a genomic quality number (GQN) metric for the Agilent ProSize data analysis software to score genomic DNA quality. The ProSize software calculates a GQN value from 0 to 10 based on the fraction of total measured concentration above a specified size threshold defined by the user.[1] The recommended threshold and acceptable GQN will depend on the application and sample prep workflow step. For whole genome sequencing, PacBio recommends that the starting genomic DNA has a GQN of at least 7.0 or higher at 10 kb (GQN_{10kb} ≥7.0) and a GQN of 5.0 or higher at 30 kb (GQN_{30kb} ≥5.0) for best results. Figure 4 shows an example of a human saliva sample analyzed on the Femto Pulse system using the Agilent Genomic DNA 165 kb kit. Because this sample is below the GQN thresholds at both 10 kb and 30 kb, treatment with the short read eliminator (SRE) kit would be recommended to deplete the lower molecular weight DNA fragments to improve sample quality prior to DNA shearing. Please refer to PacBio library prep protocols for additional information on genomic DNA quality, GQN guidelines, and size selection options.[2]

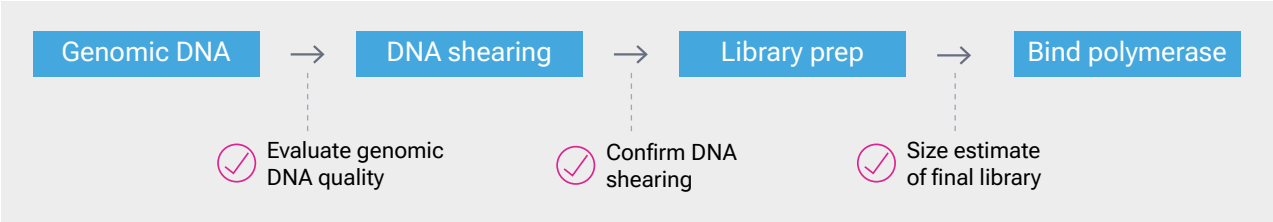
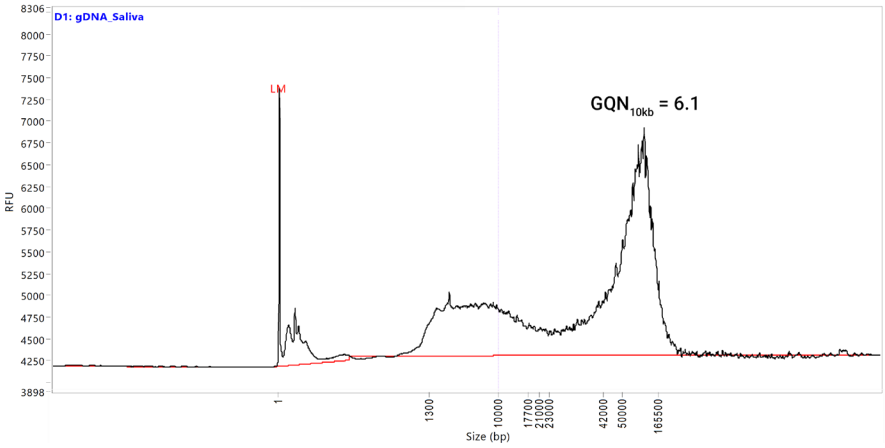


Figure 3. Overview of the HiFi sequencing workflow starting with genomic DNA through sequencing polymerase binding. The recommended QC checkpoints using the Agilent Femto Pulse system are shown in magenta.

A.



B.

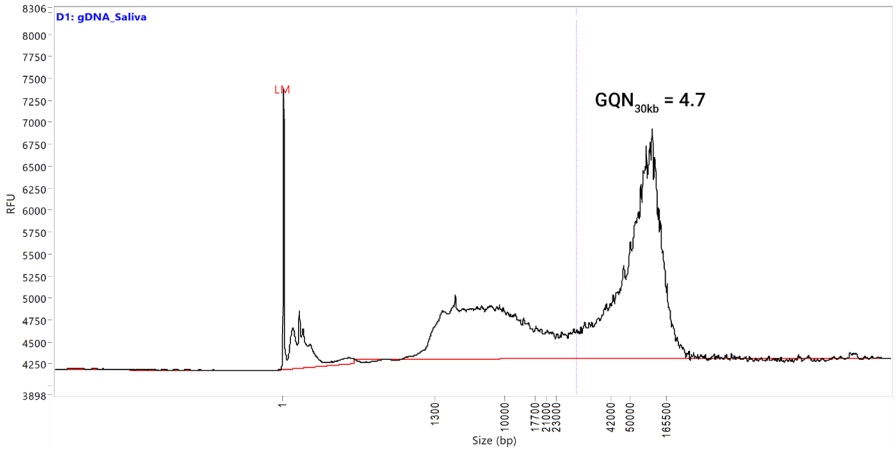


Figure 4. Example electropherograms from the Agilent ProSize software showing a genomic DNA human saliva sample measured on the Agilent Femto Pulse system. The GQN at 10 kb (A) and 30 kb (B) is 6.1 and 4.7, respectively. To increase HiFi read length and HiFi base yield, it is recommended to treat this sample with SRE prior to library prep to improve sample quality or to use a post-library size selection method such as the SageScience Pippin HT or Yourgene Health Lightbench systems to remove DNA less than 10 kb.

DNA shearing QC

It is important to verify that the genomic DNA has been sheared to the target fragment lengths (Table 2). DNA that is under sheared—meaning that DNA fragments are too large on average—will reduce sequencing yield and quality scores. Similarly, over shearing—meaning too many short DNA fragments on average—will also reduce sequencing yields because of shorter read lengths. The Femto Pulse system is the optimal technology to accurately determine whether your sample sheared appropriately. Technologies that use direct current may not resolve higher molecular weight DNA (e.g., >15 kb) and will not be predictive of sequencing performance. Examples of optimal DNA shearing profiles on the Femto Pulse system for whole genome sequencing are shown in Figure 5.

	Metric	Human, plant, and animal WGS	Microbial or metagenomic samples
Target fragment lengths	Mode (peak)	15–20 kb	7–20 kb
	Approx. range of sheared fragments	5 to 40 kb	5 to 40 kb

Table 2. Recommend DNA shearing size distributions for large and small genome shotgun sequencing. The mode is where the distribution should be centered and contain most of the fragments (area under the curve). The distribution of fragments should range from 5 to 40 kb. Microbial or metagenomic samples can have smaller mode sizes because these genomes tend to have less repetitive content and are haploid.

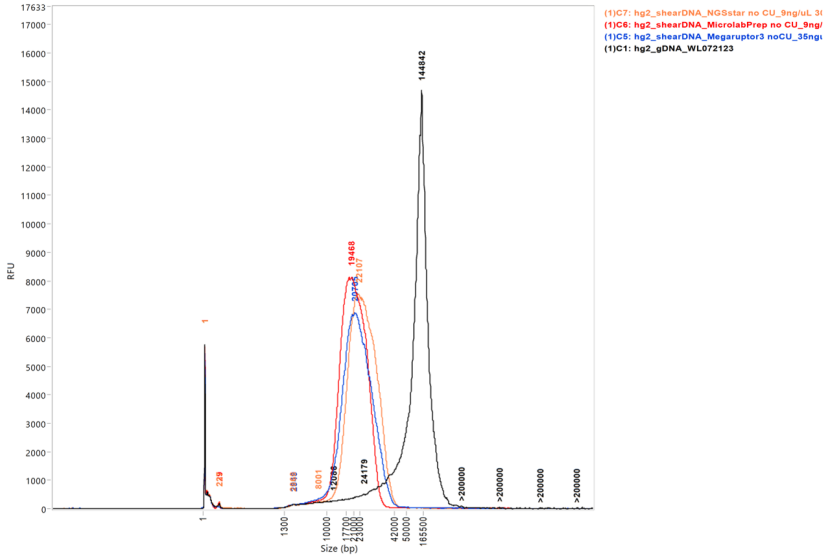


Figure 5. HG002 genomic DNA (black line with mode of ~144 kb) was sheared with multiple technologies and measured on the Agilent Femto Pulse system. DNA sheared with the Diagenode Megaruptor 3 is shown in blue, Hamilton Microlab Prep in red, and Hamilton NGS STAR MOA in orange.

Library prep QC

The final recommended QC step follows library prep. Accurately estimating the final size of the library is important for calculating the SMRT® Cell loading concentration. Inaccurate DNA sizing can lead to under loaded cells or unnecessary waste of library (by loading more than necessary). Performing a smear analysis in the ProSize software is recommended to estimate the average size of the SMRTbell library to use for polymerase binding and SMRT Cell loading (Figure 6).

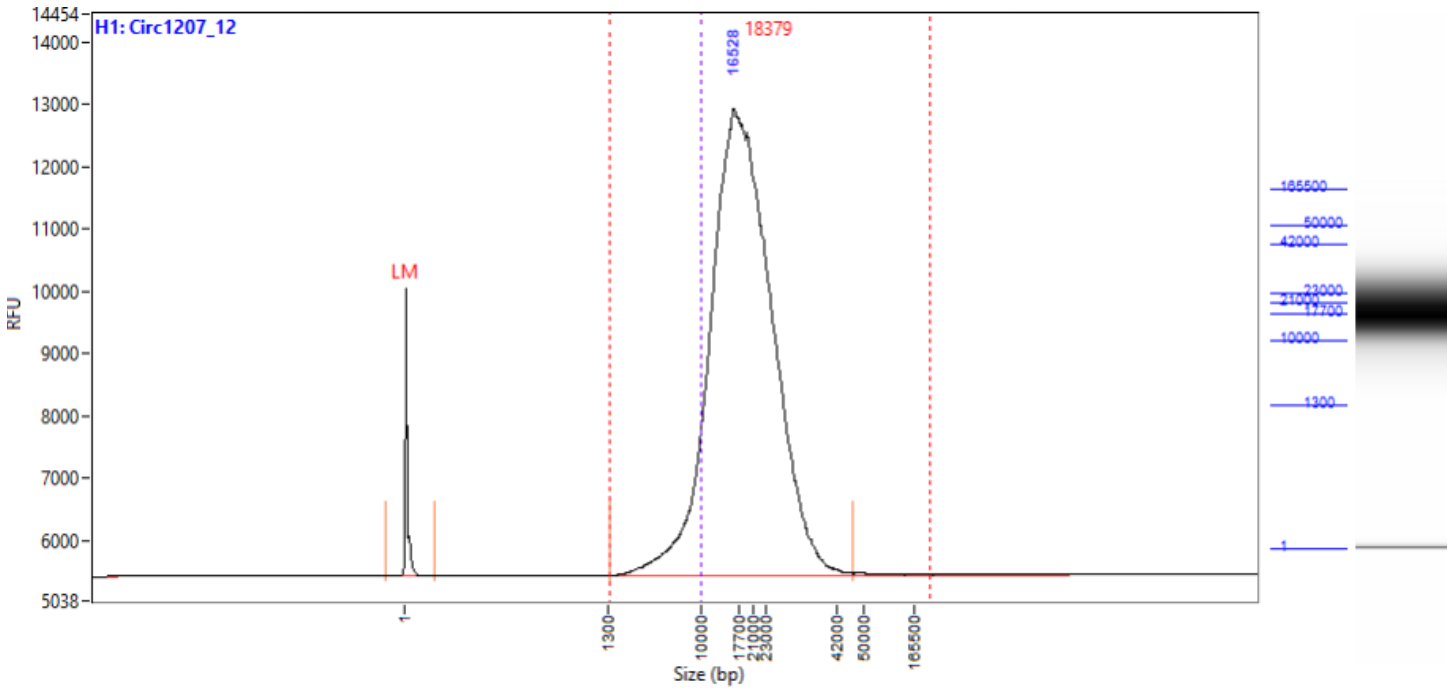


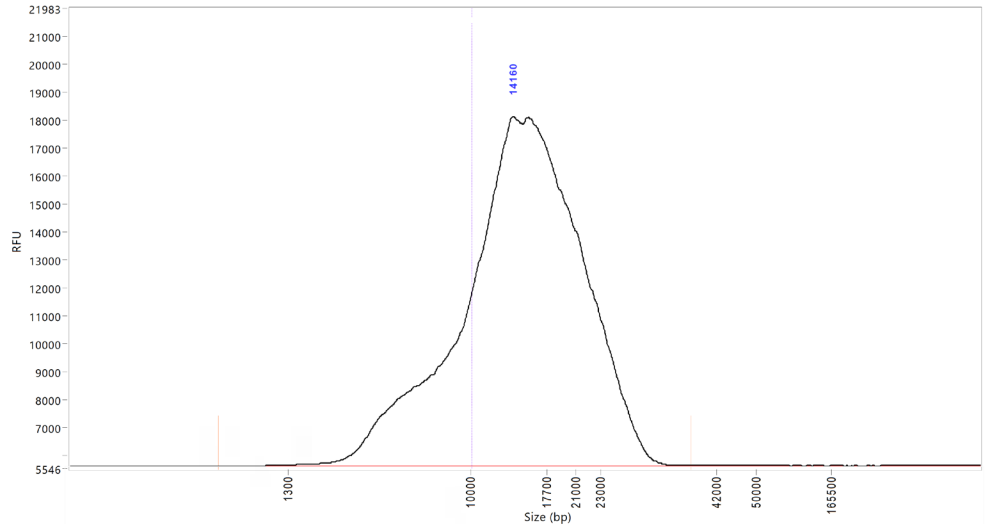
Figure 6. Electropherogram of the final SMRTbell library of the human whole blood sample shown in Figure 3. A smear analysis between 1,300 and 165,000 bp demonstrates that the average mass of the library is close to 18 kb with minimal fragments below 10 kb (blue dashed line).

Viewing by moles, not mass

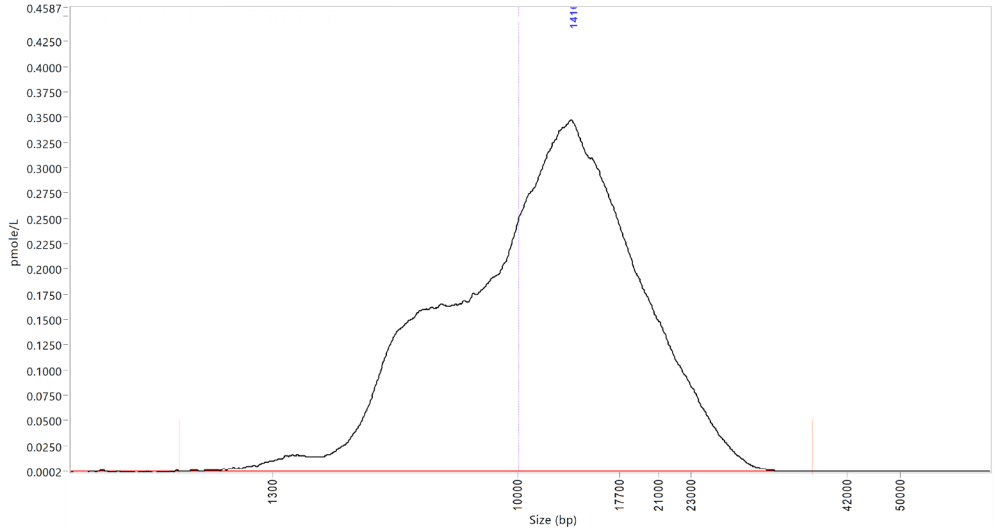
HiFi sequencing uses passive diffusion loading, meaning that the higher the molarity, the higher the likelihood the fragment will load into the SMRT Cell zero-mode waveguide (ZMW). Visualizing DNA by moles is therefore a better predictor of the final sequencing read length distribution. The Agilent Femto Pulse system provides a way to visualize the distribution of DNA fragments by converting the relative fluorescence unit (RFU) intensity to moles[3]. To visualize the electropherogram by moles in the ProSize software, go to the Options tab and ensure that the Display option "Scale to sample" is selected. Next, on the electropherogram, right-click on the Y-axis and select the nmol/L option[3].

Examples of this are shown in Figures 7 and 8, with two libraries analyzed on the Femto Pulse system and visualized by both mass (A) and moles (B). The corresponding tables for each figure demonstrate that the number of resulting HiFi sequencing reads at a given size range correlates with the molarity at that range. For instance, the sample in Figure 7 did not undergo treatment with the SRE kit, leaving more degraded genomic DNA under 10 kb throughout the library preparation. The high fraction of gDNA less than 10 kb in comparison to the fraction above 10 kb is reflected in the final sequencing data, with an increase in the number of shorter HiFi reads. In contrast, a sample that was treated with the SRE kit shows few fragments less than 10 kb (Figure 8). HiFi sequencing of this sample resulted in better overall read length and base yield. These samples demonstrate that the size distribution profile from the Femto Pulse when visualized in moles is a more accurate predictor of HiFi sequencing read distributions.

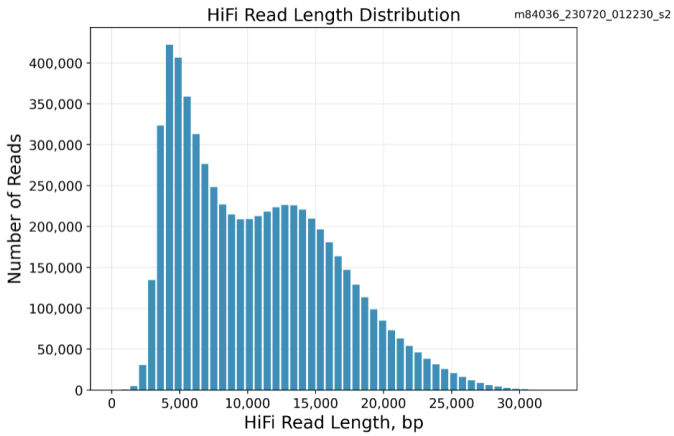
A.



B.



C.

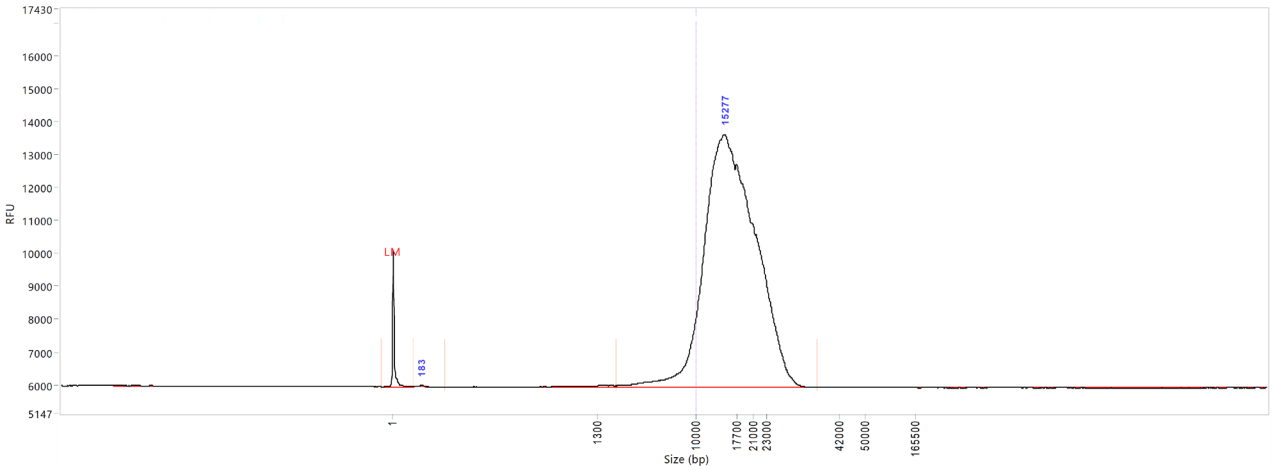


Analysis metric	Value
HiFi reads	6.4 M
HiFi base yield	68.72 Gb
HiFi read length (mean, bp)	9,915 bp
HiFi read Quality (median)	Q37

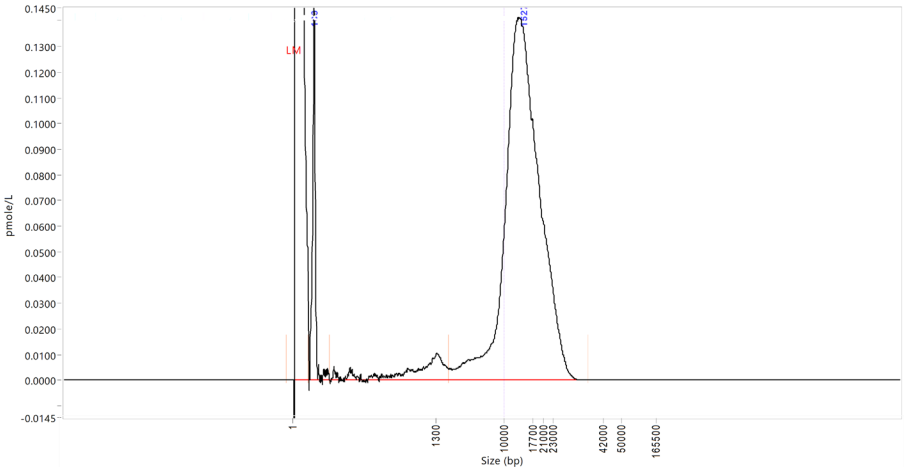
Table 3. Primary HiFi sequencing metrics for the sample shown in Figure 7. This sample did not go through size selection using the SRE kit, resulting in a high fraction of shorter DNA fragments present in the library. This reduces the read length and HiFi base yield.

Figure 7. Agilent Femto Pulse electropherograms of the same SMRTbell library plotted by A) RFU and B) pmole/L. Plotting the distribution by molarity shows that the fraction of lower molecular weight fragments is relatively higher than would be inferred from the RFU plot. This is further confirmed from the HiFi read length distribution plot of the same library C) after sequencing on the Revio™ system. This sample did not go through SRE treatment to deplete the genomic DNA <10 kb prior to library prep.

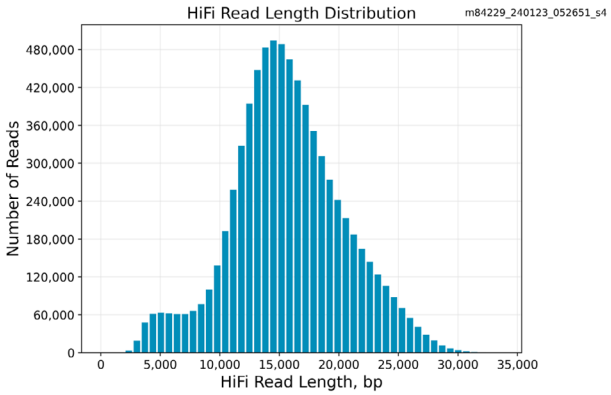
A.



B.



C.



Analysis metric	Value
HiFi reads	7.6 M
HiFi base yield	118.71 Gb
HiFi read length (mean, bp)	15,435 bp
HiFi read quality (median)	Q34

Table 4. Primary HiFi sequencing metrics for the sample in Figure 8. This sample went through size selection using the SRE kit, resulting in fewer short library fragments and higher HiFi read lengths and base yield.

Figure 8. Agilent Femto Pulse electropherograms of the same high-quality SMRTbell library plotted by A) RFU and B) pmole/L. This sample went through SRE treatment to deplete short genomic DNA fragments less than 10 kb prior to library prep. The depletion of short fragments from the sample is confirmed in the pmole/L plot and is reflected in the C) HiFi read length distribution plot from sequencing on the Revio system.

KEY REFERENCES

1. Agilent application note – Genomic DNA Sizing and Quality Control on the Agilent Femto Pulse System, Publication number 5994-0516. 2019.
2. PacBio procedure & checklist – Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0. PacBio documentation.
3. Agilent application note – Visualizing DNA for Long-Read Sequencing by Moles, Not Mass. Publication number 5994-7266. 2024.

Agilent Technologies. PR7001-2841

102-326-561 REV04 22AUG2024