Application Note

Genomics

Agilent

Trusted Answers

Quality Control in the PacBio HiFi Whole Genome Sequencing Library Preparation Workflow

Using the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems

Authors

Timothy Butler and Whitney Pike,
Agilent Technologies, Inc.

Aurelie Souppe, Gloria Diaz,
and Greg Young,
PacBio

Abstract

Accurate and reliable next-generation sequencing (NGS) results rely on starting with high-quality samples. In long-read sequencing workflows, such as the PacBio HiFi workflow, essential quality control (QC) steps include assessing input gDNA, intermediate steps throughout library preparation, and the final library. Electrophoretic analysis enables evaluation of sample size, purity, and integrity at each of these steps. Agilent offers several automated electrophoresis systems, including the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems, each with unique benefits for DNA QC. This application note demonstrates the capabilities and limitations of each of the automated electrophoresis systems within the PacBio HiFi whole genome sequencing (WGS) library preparation workflow, aiding in the selection of the appropriate QC system.

Introduction

Quality control (QC) in next-generation sequencing (NGS) workflows is vital for successful library preparation and sequencing. In the PacBio HiFi WGS sequencing workflow, electrophoretic analysis is recommended for assessing sample size and integrity at three key steps of library preparation, as outlined in Figure 1. First, QC of the input genomic DNA (gDNA) helps to ensure high-quality starting material. Next, QC after shearing confirms that the gDNA is of the correct size to proceed with library preparation. Finally, QC of the final library aids users in determining the size and molarity of the library for loading onto the sequencer.

Electrophoresis is a widely used technique in molecular biology laboratories, playing a key role in evaluating sample size, quantity, and quality. High molecular weight (HMW) DNA, such as gDNA and HiFi WGS libraries, can be analyzed using either constant or pulsed field electrophoresis. Constant or direct field gel electrophoresis applies current in a single direction and is commonly used for lower molecular weight DNA samples. In contrast, pulsed field gel electrophoresis (PFGE) alternates the direction of the electrical current moving through the gel medium, providing superior separation of HMW DNA, but often requires lengthy overnight runs. Modern systems have automated electrophoresis processes, allowing for easier set up, quicker run times, and objective data analysis. Agilent offers multiple automated electrophoresis systems for the analysis of nucleic acids, including the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems, each offering unique benefits as QC tools.

Designed for analyzing HMW DNA, the Femto Pulse system is the only automated pulsed-field capillary electrophoresis (PFCE) system currently available, capable of separating 12 samples in parallel in approximately 1.5 hours. Using the optimized pulsed-field separation methods of the Agilent Genomic DNA 165 kb kit (p/n FP-1002), the Femto Pulse provides accurate and reliable sizing of HMW DNA up to 165 kb, requiring only picogram amounts of DNA. This makes the Femto Pulse the optimal choice for analyzing the size and quality of DNA in the PacBio HiFi WGS workflows. Due to the large sizing range and high-resolution separations, the Femto Pulse can be used for any of the HiFi WGS library preparation QC steps.

The Fragment Analyzer and TapeStation systems are both automated constant field electrophoresis systems that offer high throughput and are compatible with various assays that can size DNA up to 60 kb. The Fragment Analyzer systems use parallel capillary electrophoresis (CE) to simultaneously analyze 12, 48, or 96 samples in 70 minutes or less.

The TapeStation systems offer highly automated, easy and fast analysis of one to 96 samples, in just one to two minutes per sample, using ready-to-use and proprietary Agilent ScreenTape technology.

The smaller sizing range of the Fragment Analyzer and TapeStation systems make them suitable for assessing the sheared gDNA and library QC steps in the HiFi WGS library preparation workflow. In contrast, the expansive sizing range of the Femto Pulse system allows for accurate sizing throughout the entire workflow and is the only automated electrophoresis instrument capable of accurately sizing samples larger than the HiFi WGS libraries, such as input HMW gDNA (Table 1). This application note describes several QC steps in the HiFi WGS workflow where systems in the Agilent automated electrophoresis portfolio can be applied (Figure 1). HMW gDNA, sheared DNA, and final HiFi WGS library samples were analyzed using the Femto Pulse, Fragment Analyzer, and TapeStation systems, and the average sizes measured by each system were compared.

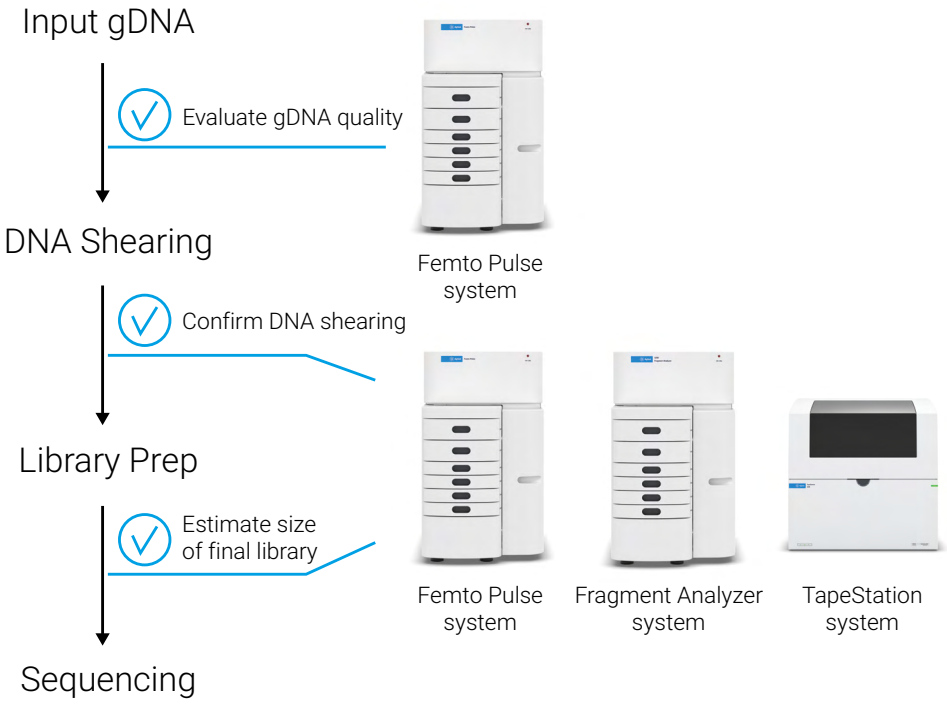

Figure 1. Overview of the PacBio HiFi WGS library preparation workflow, highlighting specific quality control steps that the Agilent automated electrophoresis systems, including the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems, can be used.

Table 1. Analytical specifications of the Agilent automated electrophoresis systems and their respective kits for high molecular weight DNA quality control.

System	Femto Pulse	Fragment Analyzer	TapeStation
Kit name	Genomic DNA 165 kb kit (p/n FP-1002-0275)	HS Large Fragment Analysis kit (p/n DNF-464-0500)	Genomic DNA ScreenTape (p/n 5067-5365) and reagents (p/n 5067-5366)
Sizing range	1.3 – 165 kb	75 bp – 48.5 kb	200 bp – 60 kb
Sizing accuracy	±15%	±15% at ≤ 15 kb ±25% at ≥ 15 kb	200 bp – 15 kb: ±15%
Sizing precision	20 %CV	N/A	200 bp – 15 kb: 15 %CV
Quantification accuracy	N/A	±25%	±20%
Quantification precision	25 %CV	20 %CV	15 %CV
Sample concentration range	Fragment: 0.3 – 30 pg/µL Smear: 5 – 500 pg/µL	Fragment: 5 – 600 pg/µL (optimal concentration 500 – 60 pg/µL) Smear: 50 pg/µL – 5 ng/µL (optimal concentration 1 ng/µL)	Quantitative range: 10 – 100 ng/µL
Electrophoresis method	Pulsed field	Constant field	Constant field

Experimental

Sample information

In this study, gDNA, sheared DNA, and HiFi SMRTbell library samples were used to illustrate the QC steps in the PacBio HiFi WGS library preparation process. Sheared DNA from *E. coli* and human (Hg002), along with the resulting libraries, were provided by PacBio. The *E. coli* library was sequenced using the PacBio Revio system, yielding a mean HiFi read length of 19.0 kb. The human library was sequenced using the PacBio Sequel II system, with a mean HiFi read length of 7.2 kb. Sheared samples were prepared from Human Mixed Genomic DNA (Promega, p/n G3041) using Covaris g-TUBEs (p/n 520079) to generate samples of various sizes.

Automated electrophoresis

All samples were analyzed using Agilent automated electrophoresis systems. Each sample was assessed on the Femto Pulse with the Genomic DNA 165 kb kit (FP-1002-0275) at a concentration of approximately 300 pg/µL. The samples were run on the 5200 Fragment Analyzer system using the Agilent HS Large Fragment Analysis kit (p/n DNF-464-0500) at a concentration of approximately 1 ng/µL. Additionally, the samples were analyzed on the 4200 TapeStation using the Agilent Genomic DNA ScreenTape (p/n 5067-5365) with Genomic DNA reagents (p/n 5067-5366) at a concentration of approximately 30 ng/µL.

Data analysis settings

Data acquired from the Agilent automated electrophoresis systems are automatically analyzed using the ProSize data analysis software for the Femto Pulse and Fragment Analyzer, or the TapeStation analysis software for the TapeStation. Both software programs provide digital gel images and electropherograms of each sample, as well as data tables with information such as sample size and concentration.

For accurate assessment of samples throughout the PacBio HiFi WGS library preparation process, a user-defined smear or region analysis can be applied in either software to obtain an average sample size. In the ProSize software, users can set upper and lower smear boundaries to include any region of the analysis window for average sizing (Figure 2). For best practice, this region should be set between the lower and upper markers (or the last ladder fragment for kits that do not use an upper marker), but it may be extended past this range if the sample smears out farther than the ladder. For example, the last fragment of the ladder used with the Femto Pulse gDNA kit is 165 kb and does not use an upper marker. The HS Large Fragment kit for the Fragment Analyzer uses a ladder with the last fragment at 48.5 kb. When the upper limit of the smear range is set at a larger size then 48.5 kb, the separation resolution is limited and the size of the smear is extrapolated by the ProSize data analysis software using a linear line of best fit generated by the ladder.

Similarly, the TapeStation software allows users to set defined regions for analysis. The gDNA ScreenTape assay uses a ladder that extends to 48.5 kb and does not include an upper marker. Smear analysis can start at the lower marker and extrapolate sizing up to 60 kb. Beyond 60 kb, the separation resolution is limited, and the TapeStation software will not include any portion of the smear that extends beyond this size range.

In this study, the smear range for gDNA analyzed on the Femto Pulse was set with the lower limit of 400 bp, while the upper limit varied to ensure inclusion of the entire smear. The sheared DNA and libraries were analyzed using a smear range from 400 bp to 165 kb with the Femto Pulse, and from 400 bp to 60 kb with the Fragment Analyzer and TapeStation, to provide a comparison of the systems.

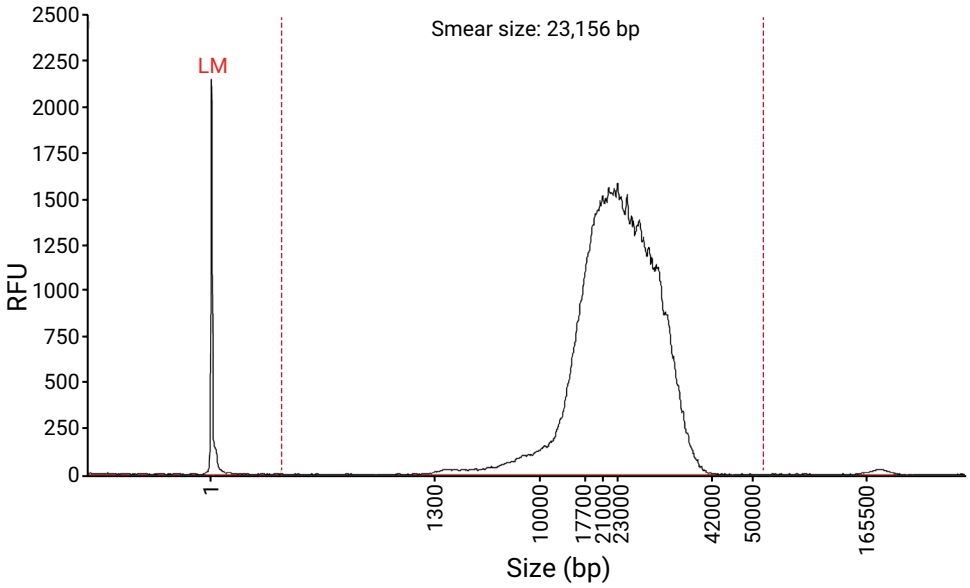

Figure 2. Representative example of a PacBio library analyzed on the Agilent Femto Pulse system, with the user-defined upper and lower bounds of the smear analysis (red dashed lines) encompassing the full range of the sample.

Results

QC of input genomic DNA

The first QC step in the PacBio library preparation workflow is to assess the quality of gDNA before shearing. Knowing the size and distribution of gDNA is important for assessing the sample's suitability for long-read sequencing. For instance, a high proportion of HMW gDNA will result in a better size distribution after shearing, enhancing read lengths and sequencing coverage.[1] Conversely, a sample with more degraded material will lead to shorter read lengths and reduced total sequencing yield.

The Femto Pulse is uniquely suited for this analysis, providing high-resolution separations and sizing of HMW gDNA through 165 kb. To demonstrate the performance of the Femto Pulse, multiple replicates of *E. coli* and human gDNA samples were assessed. On average, the Femto Pulse reported the smear size of *E. coli* gDNA at 124 kb and human gDNA at 146 kb (Figure 3), with excellent precision between replicates (2 %CV and 3 %CV, respectively). The Fragment Analyzer and the TapeStation systems were not used for assessing these large samples, as the smears extend beyond the sizing range of their analysis kits.

In addition to determining the size distribution of large HMW samples, PacBio recommends the Femto Pulse due to the Genomic Quality Number (GQN) metric generated in the ProSize software.[1,2] The GQN measures the percentage of the sample above a user-defined threshold and assigns a score from 0 to 10. A GQN score of zero indicates that none of the sample exceeds the user-defined threshold, while a score of 10 indicates that all of the sample is above the threshold.

The recommended threshold and acceptable GQN depends on the application and sample preparation workflow step. For example, PacBio recommends that the input gDNA should have a GQN of at least 7.0 or higher at 10 kb ($GQN_{10kb} \geq 7.0$) and a GQN of 5.0 or higher at 30 kb ($GQN_{30kb} \geq 5.0$) for whole genome sequencing.[1] In the examples shown in Figure 3, the *E. coli* sample displayed a GQN_{10kb} of 9.7 and a GQN_{30kb} of 9.0. The human gDNA had a GQN_{10kb} of 9.4 and GQN_{30kb} of 8.9.

The high GQN scores of each sample indicate that the samples are of sufficient quality to proceed with the shearing and sequencing workflow. The GQN score allows users to assess sample integrity, enabling a flexible empirical scoring strategy based on the application.

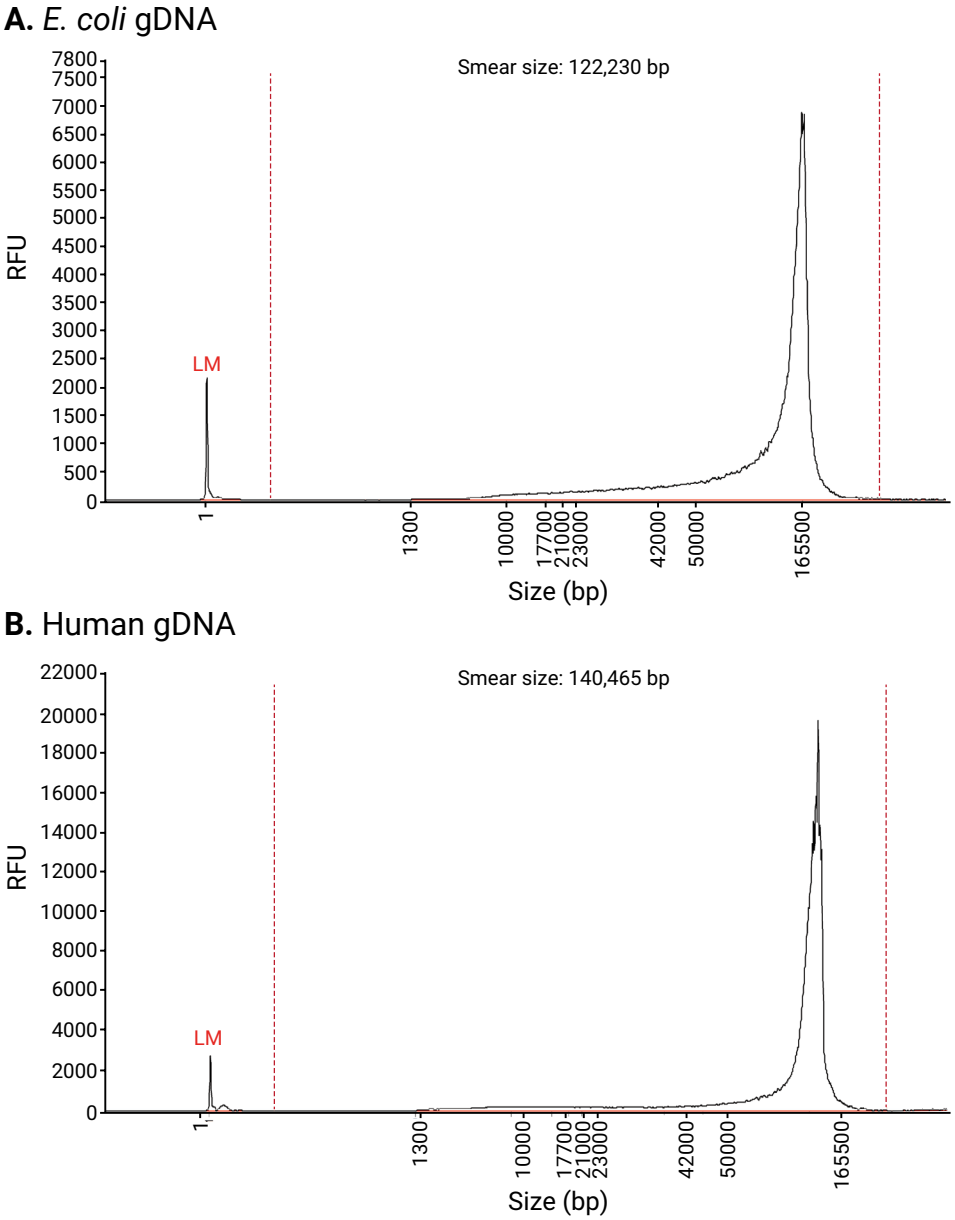

Figure 3. (A) *E. coli* and (B) human genomic DNA samples analyzed using the Agilent Femto Pulse system. Red lines indicate the smear analysis range used for assessment of the gDNA.

Sizing comparison of sheared DNA

The second QC step recommended in the PacBio HiFi WGS library preparation workflow occurs after shearing. At this stage, QC of sheared DNA confirms that the sample is within the correct size distribution to proceed with library preparation and subsequent sequencing. For PacBio HiFi WGS applications, the gDNA should be sheared to 10 to 25 kb.[3]

While the Femto Pulse is the recommended analysis tool for QC within the PacBio workflow, each of the Agilent automated electrophoresis systems has analysis kits covering this size range. As shown in Figure 4, multiple gDNA samples were sheared to various sizes between 10 and 20 kb and assessed on each system. Data analysis with the Femto Pulse used a smear range from 400 bp to 165 kb to encompass the entire range of the sample. The constant field systems used a range of 400 bp to 60 kb.

The sizes of the sheared samples were similar when compared between systems. For example, Shear 2 reported sizes close together across each instrument, with an average size of 14.1 kb using the Femto Pulse, 13.9 kb with the Fragment Analyzer, and 14.5 kb from the TapeStation. Other samples showed similar sizing, but with slight differences between instruments, such as Shear 3, with an average of 17.6 kb on the Femto Pulse, 14.6 kb on the Fragment Analyzer, and 15.7 kb on the TapeStation. Across each of the three systems, all samples showed excellent sizing precision of 11 %CV or less. These results demonstrate that any of the Agilent automated electrophoresis systems can be used interchangeably for QC of sheared gDNA in the PacBio HiFi WGS library preparation workflow.

QC of a sheared *E. coli* DNA for PacBio HiFi WGS library preparation

QC of sheared DNA within the PacBio library preparation protocol is primarily used to provide information regarding sample quality, such as the size distribution of the sample, prior to sequencing. When analyzed on the automated electrophoresis systems, sheared DNA typically has a larger average size than the final library. To demonstrate, a sheared *E. coli* DNA was assessed using each of the three automated electrophoresis systems. A similar distribution was observed across all instruments (Figure 5).

Smear analysis was used to determine the average size of the sheared sample with each automated electrophoresis system. The smear ranges used for this example were 400 to 165 kb for the Femto Pulse, and 400 to 60 kb for the Fragment Analyzer and TapeStation.

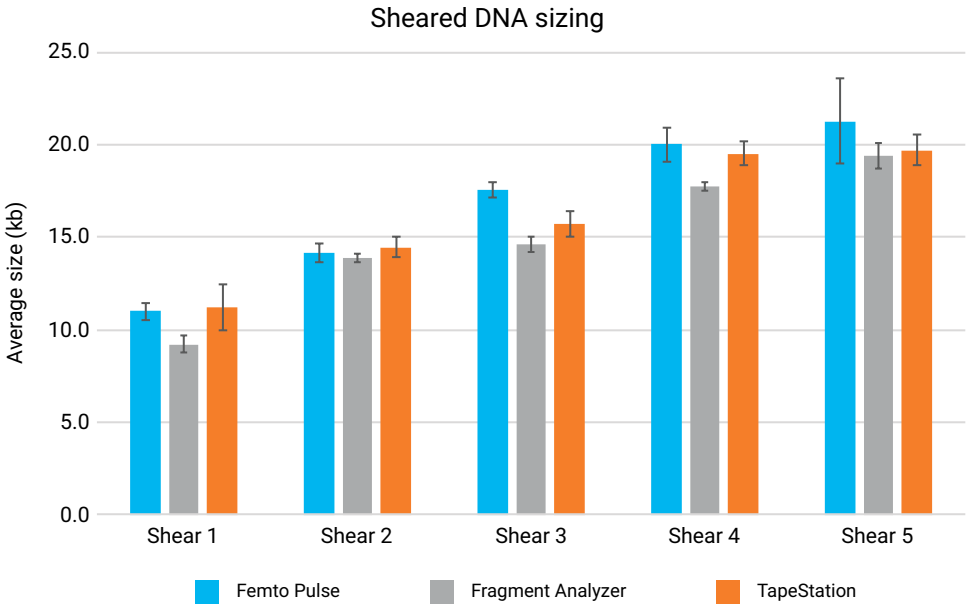

Figure 4. Average smear size for multiple sheared human genomic DNA samples analyzed using the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems. Error bars represent standard deviation. N > 3.

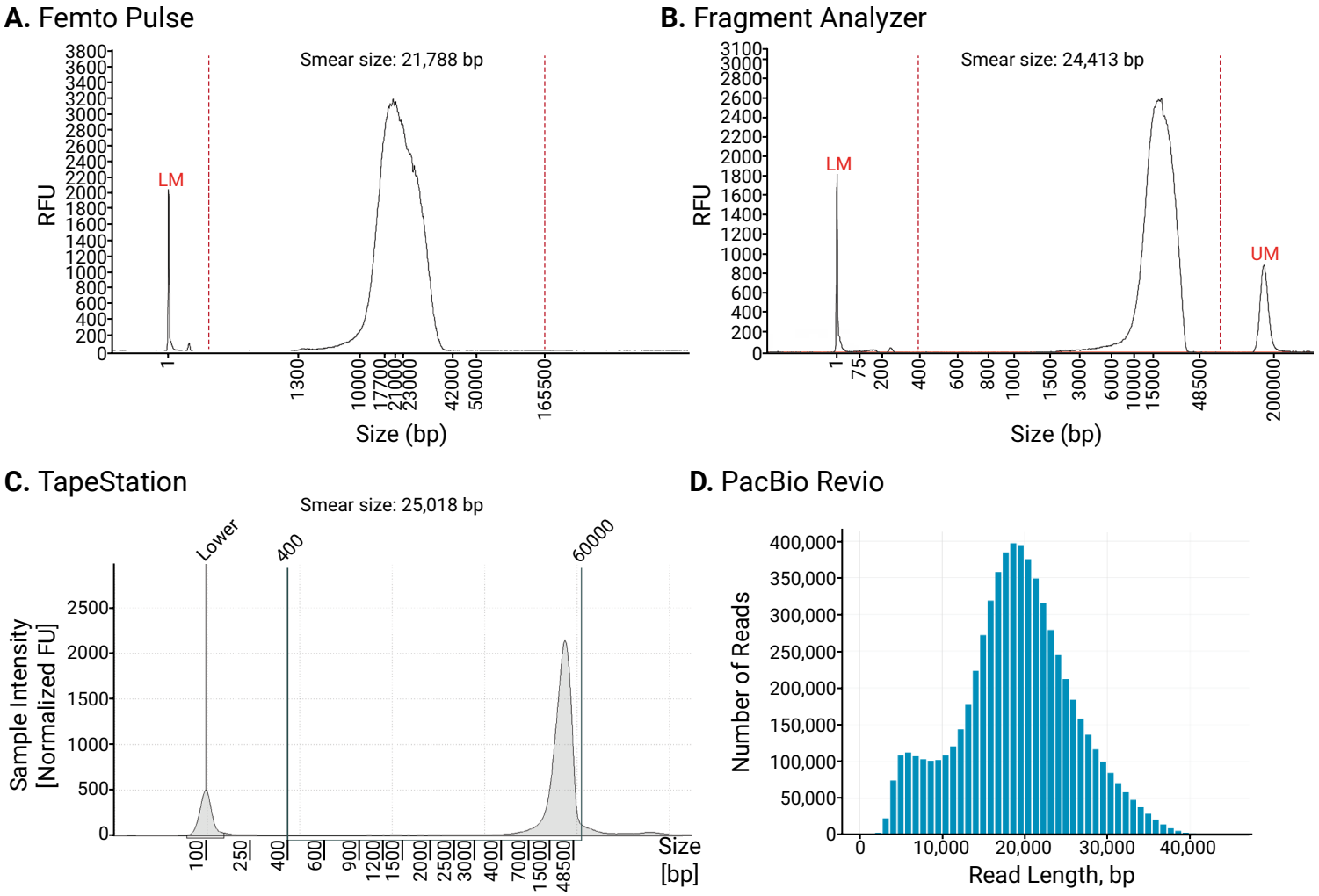

Figure 5. Electropherograms from assessment of sheared *E. coli* DNA on the (A) Agilent Femto Pulse system, (B) Agilent Fragment Analyzer system, and (C) Agilent TapeStation system. (D) PacBio HiFi WGS read length distribution plot from sequencing the final library on the Revio system.

The average smear size of multiple replicates of the sheared DNA reported by the Femto Pulse was 23.4 kb, while the Fragment Analyzer reported 27.1 kb, and the TapeStation reported 28.4 kb (Figure 6). Each system sized the sample similarly. Sizing with the Femto Pulse was closest to the sequencing results, which showed an average read length of 19.0 kb. Additionally, the Femto Pulse system showed the best sizing precision of the three instruments, at 3 %CV. By providing sizing closest to the sequencing results and excellent precision, the Femto Pulse proves to be the optimal system for this example.

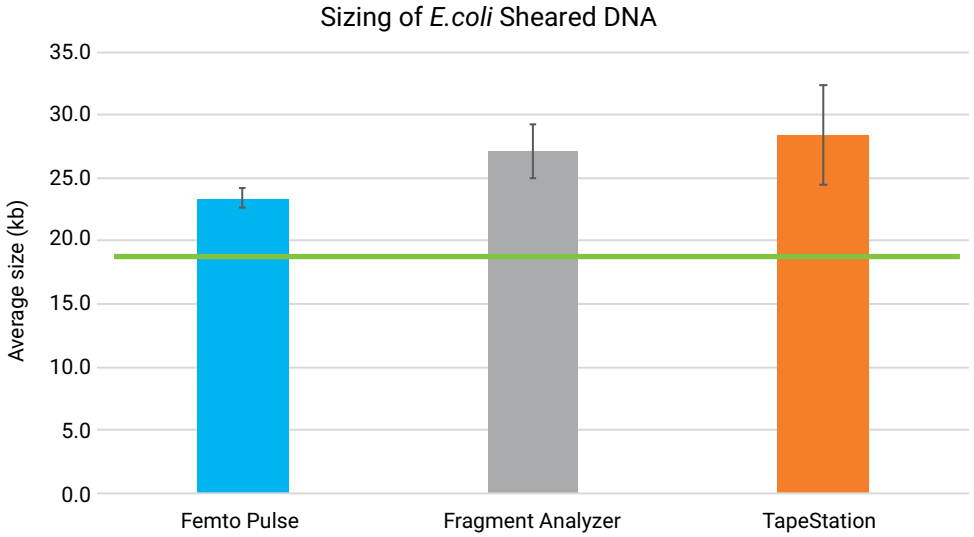

Figure 6. Average sequencing read length from PacBio HiFi WGS sequencing and corresponding sheared *E. coli* DNA sizes measured by the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems. Error bars represent standard deviation. N > 3. The green line indicates the average read length of the library sequenced on the PacBio Revio system.

QC of a human PacBio HiFi WGS library

The final QC step in the PacBio HiFi WGS library preparation workflow is to assess the quality and size of the final library. The human HiFi WGS library was assessed on each of the automated electrophoresis systems. As shown in the electropherograms in Figure 7, the library is visualized as a similarly shaped curve across all systems. Each system sized the sample close to the average read length of 7.2 kb. The average smear size was 7.7 kb using the Femto Pulse, 6.4 kb with the Fragment Analyzer, and 7.8 kb with the TapeStation (Figure 8). Additionally, all systems demonstrated high sizing precision, with 6 %CV or less. This example shows that any of the Agilent automated electrophoresis systems can provide reliable sizing analysis for QC of the library prior to sequencing.

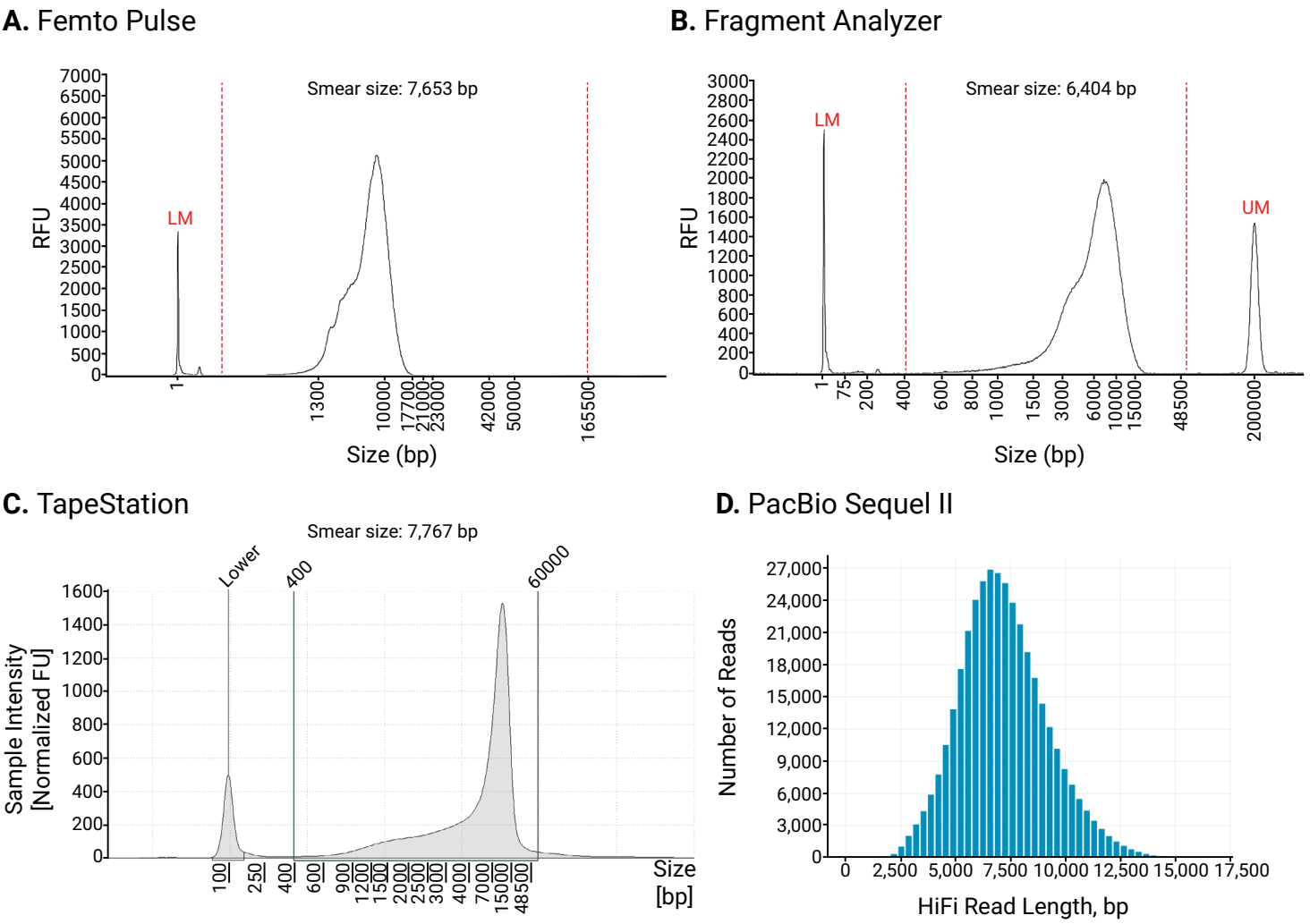

Figure 7. Electropherograms from assessment of human HiFi WGS library on the (A) the Agilent Femto Pulse, (B) the Agilent Fragment Analyzer, and (C) the Agilent TapeStation systems. (D) PacBio HiFi WGS read length distribution plot from sequencing on the Sequel II system.

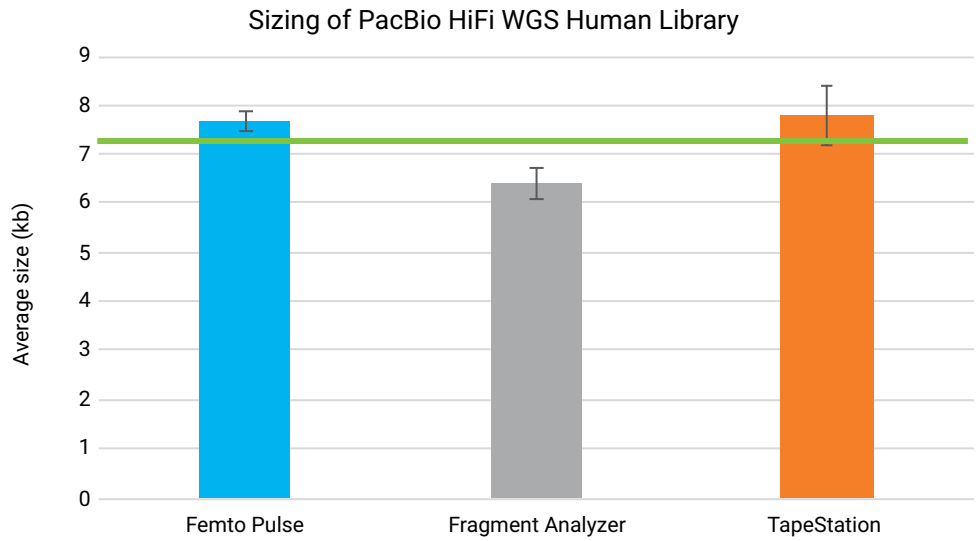

Figure 8. Average sizes for PacBio human DNA library measured by the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems. Error Bars represent standard deviation. N > 3. The green line indicates the average read length of the library sequenced on the PacBio Sequel II system.

Conclusion

QC in NGS library preparation is essential for successful sequencing. Throughout the PacBio HiFi WGS library preparation workflow, it is therefore recommended to assess the integrity and size of the DNA to help ensure a high-quality library for sequencing. This includes QC of the initial gDNA, confirmation of steps throughout library preparation (such as after shearing), and assessment of the final library before sequencing. In this application note, examples of *E. coli* and human gDNA, sheared samples, and PacBio HiFi WGS libraries were assessed using the Agilent automated electrophoresis systems, including the Agilent Femto Pulse, Fragment Analyzer, and TapeStation systems.

The Femto Pulse system is the optimal instrument for QC assessment throughout the PacBio long-read sequencing library preparation workflows. As demonstrated here, the Femto Pulse is the only automated electrophoresis system capable of sizing HMW input gDNA material for HiFi WGS, as well as assessing the sheared gDNA and the final library. The Fragment Analyzer and TapeStation systems are also suitable for analyzing smaller sheared gDNA and final libraries in the PacBio HiFi WGS workflow, making them desirable options for labs with higher throughput needs and routine sample analysis. Together, the Agilent automated electrophoresis systems provide solutions for QC throughout the long-read sequencing workflow.

References

1. Fast, Accurate DNA sizing with the Agilent Femto Pulse system for HiFi WGS, *PacBio Product Note*, publication number 102-326-561 REV04, **2024**.
2. Quality Metrics for Nucleic Acids with the Agilent Fragment Analyzer and Femto Pulse Systems, *Agilent Technologies Application Note*, publication number 5994-0521EN, **2019**.
3. Preparing DNA For PacBio HiFi Sequencing – Extraction and Quality Control, *PacBio Technical Note*, publication number 102-193-651 REV02, **2022**.

www.agilent.com/genomics/automated-electrophoresis

Published in the USA, July 1, 2025
5994-8448EN

Agilent
Trusted Answers