Application Note

Genomics

Agilent

Trusted Answers

Visualizing DNA for Long-Read Sequencing by Moles, Not Mass

Author

Whitney Pike
Agilent Technologies, Inc.

Abstract

Traditional measurements during NGS library preparation include the determination of the sample mass within specific size ranges. This analysis is easily performed with Agilent automated electrophoresis instruments, which provide visual results in the form of a digital gel image and electropherogram. The electropherogram displays the fluorescent signal as a graphical representation, with the size on the X-axis and relative fluorescence units (RFUs) on the Y-axis. The height of the fluorescent signal is therefore directly proportional to the mass of the sample at a given size. While this representation has been widely used for quality control of sheared gDNA and the final NGS library, examining the molarity of a sample may provide a better visual representation of the number of sequencing reads that can be produced by a sample, especially for long-read sequencing. High molecular weight samples were analyzed using the Agilent Femto Pulse system and the accompanying ProSize data analysis software. ProSize allows the user to visualize the electropherogram image as a product of either mass or molarity by switching the Y-axis from RFU to nmole/L. By visualizing the data in moles and using a smear analysis, the Femto Pulse can be used to determine the number of moles of sample found within different sizing brackets and provide a better prediction of long-read sequencing read lengths.

Introduction

Electrophoretic methods used in the analysis of nucleic acids often use a fluorescent intercalating dye, which binds to the nucleic acid in a regular pattern. For example, one molecule of ethidium bromide will be intercalated for every 2.5 base pairs of a sequence[1]. During electrophoresis, the sample is separated by size and visualized by fluorescence. Given the regular pattern of dye intercalation, the intensity of the resulting fluorescent signal correlates to the total mass present at any given size. This method of measuring sample distribution is customarily used to visualize NGS libraries prior to sequencing to assess length and concentration. However, NGS sequencing read counts are proportional to the molarity of each fragment length, and not the mass. Since long-read NGS libraries span a range of tens of thousands of base pairs in length, and it is known that the smaller fragments within this area are preferentially sequenced over the larger fragments[2], visualizing the sample with moles on the y-axis of an electropherogram instead of intensity may be a more accurate predictor of sequencing read length.

The Agilent automated electrophoresis instruments provide quick and easy assessment of nucleic acids, including genomic DNA (gDNA) and NGS libraries. For high molecular weight DNA and long-read sequencing libraries, the Femto Pulse system was specifically designed to automate pulsed field gel electrophoresis for samples through 165 kb, while also providing faster run times and increased sensitivity. With the Agilent ProSize data analysis software, the Femto Pulse presents a way to visualize large DNA samples in either moles or mass to better represent the data and predict sequencing results.

Methods

Samples

A one microgram aliquot of Lambda DNA (NEB p/n N3011S) was digested with *SacI* (NEB p/n R0156) at 37 °C for 15 minutes, followed by heat inactivation at 65 °C for 20 minutes.

Human gDNA was obtained from Zyagen (p/n HG-705) and diluted to 500 pg/µL for further analysis.

Human gDNA (Promega p/n G3041) was sheared using Covaris g-TUBEs (p/n 010145), following the manufacturer's protocol for 20 kb.

Agilent Femto Pulse system

All samples were analyzed with the Agilent Femto Pulse system and the ProSize data analysis software. *SacI* digest products were run with the Agilent 55 kb BAC kit (p/n FP-1003-0275), which is optimized for DNA fragments. Genomic DNA (gDNA) was run with the Agilent Genomic DNA 165 kb kit (p/n FP-1002-0275). To visualize the samples in moles, under the Option tab in ProSize, ensure that the Display option "Scale to sample" is selected. On the electropherogram, right click the Y-axis and select nmol/L.

Library preparation and sequencing

An NGS library was prepared from the gTUBE sheared DNA using the Oxford Nanopore Technologies MinION with the Ligation Sequencing kit (Oxford Nanopore Technologies, p/n SQK-LSK109) according to the manufacturer's specifications, using 500 ng DNA input. The library was sequenced using the Oxford Nanopore Technologies MinION Mk1B device equipped with MinION Flow Cells (version R9.4.1, p/n FLO-MIN106D).

Sequencing data analysis

Sequencing data was base called in real-time using the MinKNOW (v20.06.5) software. After filtering, "passed" reads were further processed using the NanoPack package, as previously described[2].

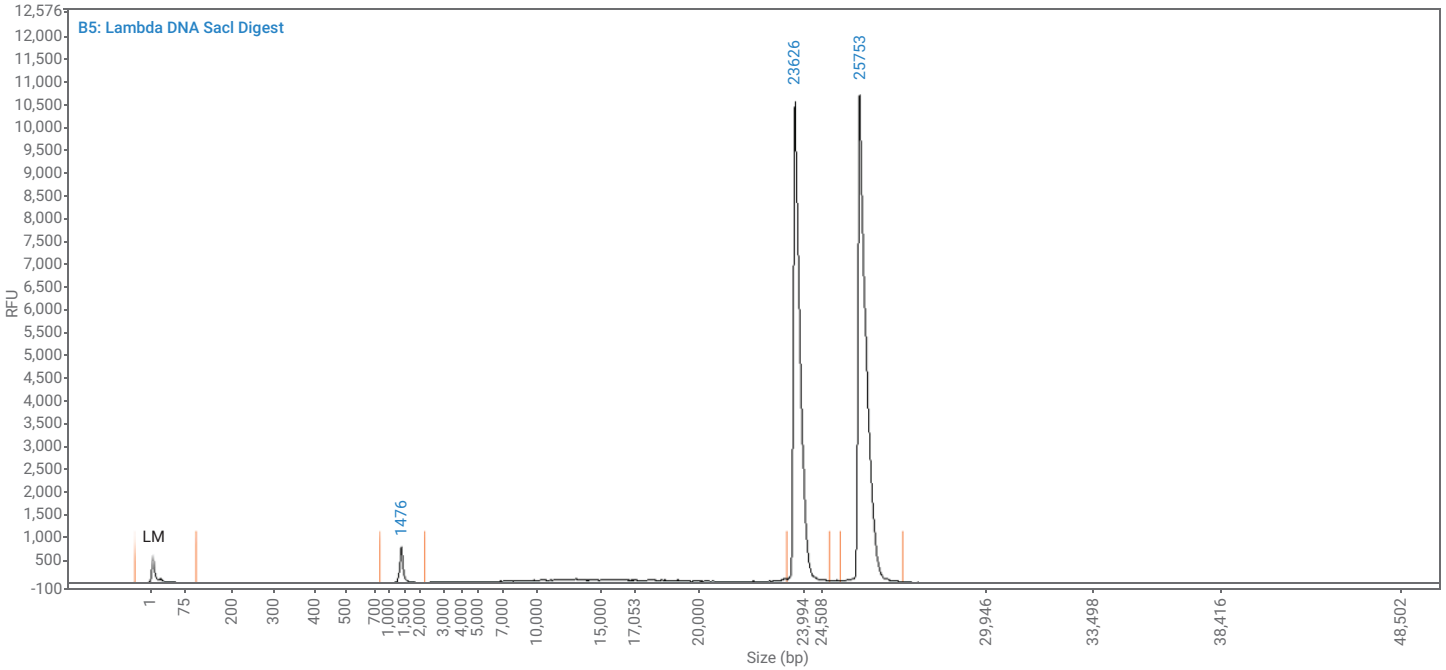
Results and discussion

Visualizing moles instead of mass using Agilent ProSize data analysis software

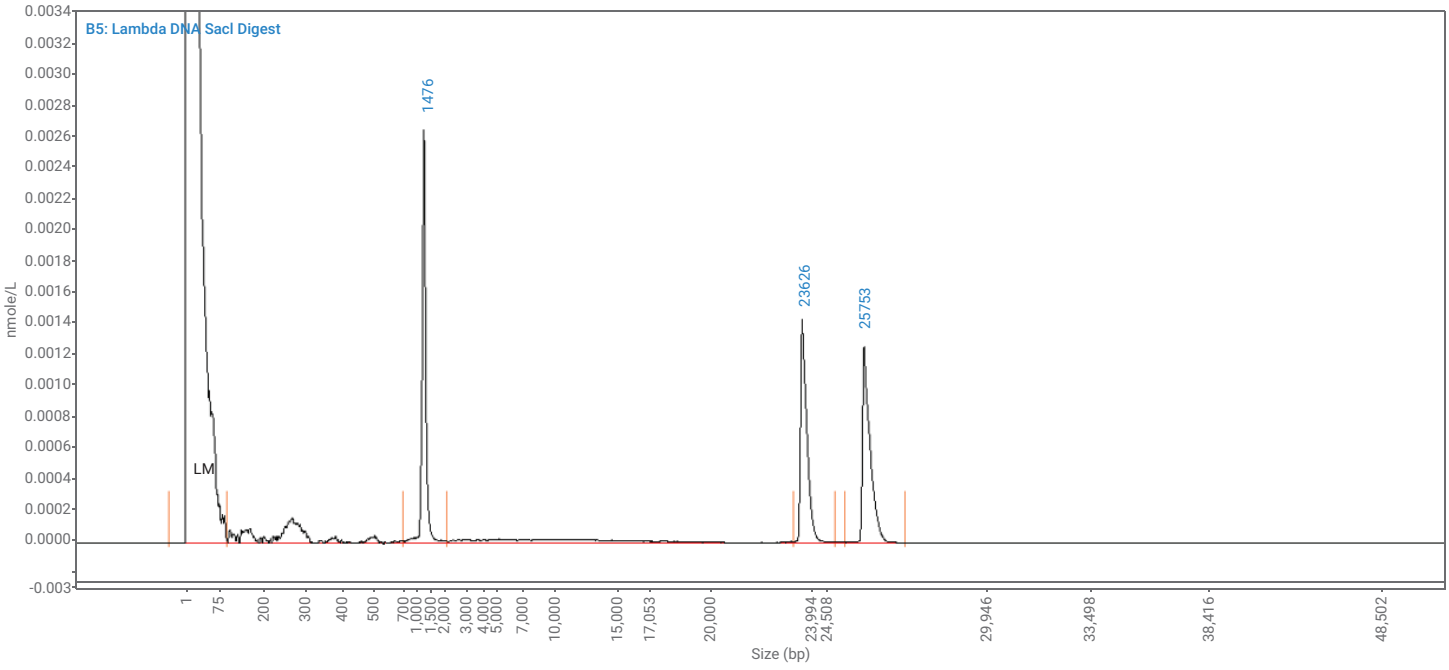
Samples assessed with the Femto Pulse system and corresponding ProSize data analysis software are automatically analyzed for size and concentration, with the results being displayed as a digital gel image, electropherogram, and peak table. The X-axis of the electropherogram displays the sizing scale, while the Y-axis, by default, displays relative fluorescence units (RFU). The intensity of the fluorescent signal is indicative of the mass, or concentration (ng/µL), of the sample. Users have the option to toggle the Y-axis units to moles (nmol/L) to visualize the sample in molarity instead of mass.

To demonstrate this concept, Lambda DNA was digested with *SacI* to produce three equimolar fragments of different lengths and analyzed on the Femto Pulse system, as shown in Figure 1. Despite having similar molarities, when visualized in terms of mass (RFUs) the electropherogram shows the 1,476 bp fragment at a much shorter peak height than the other two fragments (Figure 1A). Additionally, the reported concentration is much less for the first fragment (0.0273 ng/µL) compared to the other two fragments, which are each at ~ 0.4 ng/µL. When visualized by moles, all three peaks were similar in peak height (Figure 1B) and the calculated number of moles differed by no more than 10% between the peaks (Figure 1C).

A. Visualization by mass.



B. Visualization by moles.



C. Peak analysis.

Peak Number	Size (bp)	Concentration (ng/µL)	Molarity (nMol)
1	1,476	0.0273	0.0307
2	23,625	0.4230	0.0293
3	25,753	0.4930	0.0313

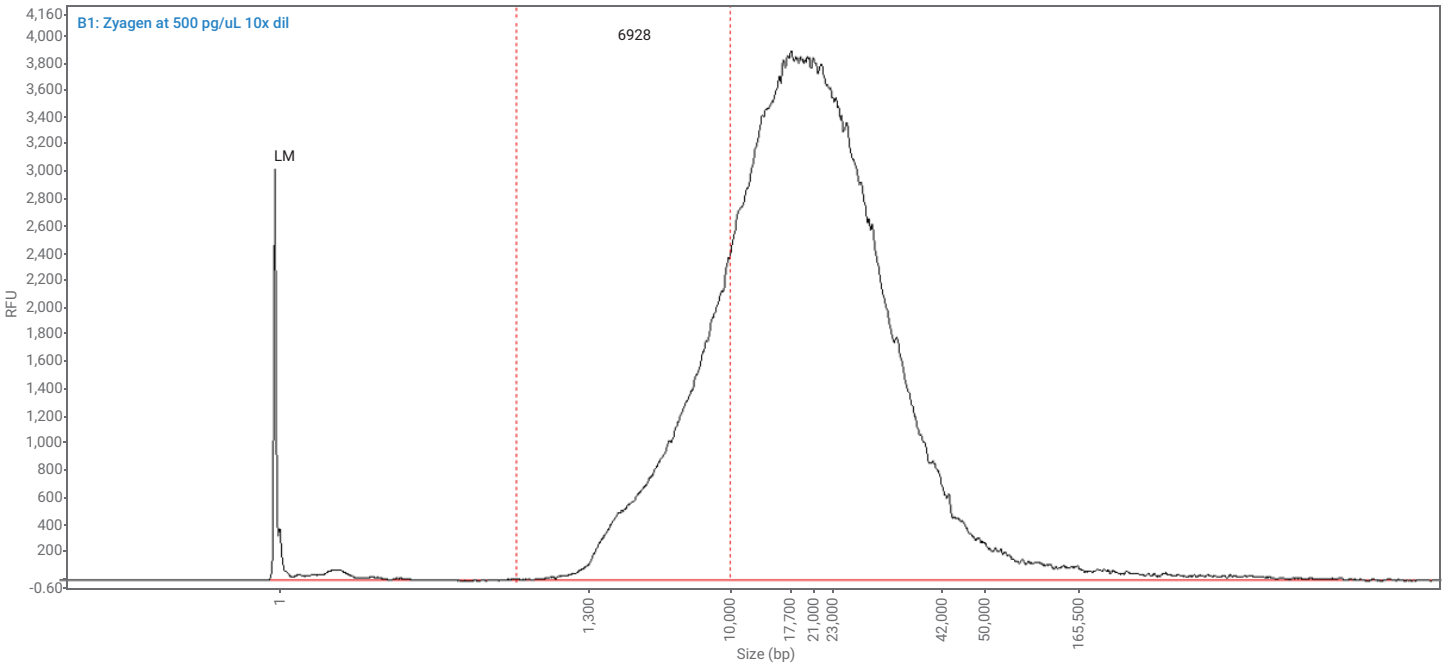
Figure 1. Digestion of Lambda DNA with *SacI* produces three equimolar fragments of different lengths, as shown in the zoomed-in electropherograms from analysis with the Agilent Femto Pulse system and the corresponding Agilent ProSize data analysis software. Right clicking on the Y-axis allows toggling between A) RFU (mass) and B) nmole/L (moles). C) The peak analysis table displays the reported size (bp), concentration, and molarity of each of the three fragments.

Smear analysis in moles instead of mass

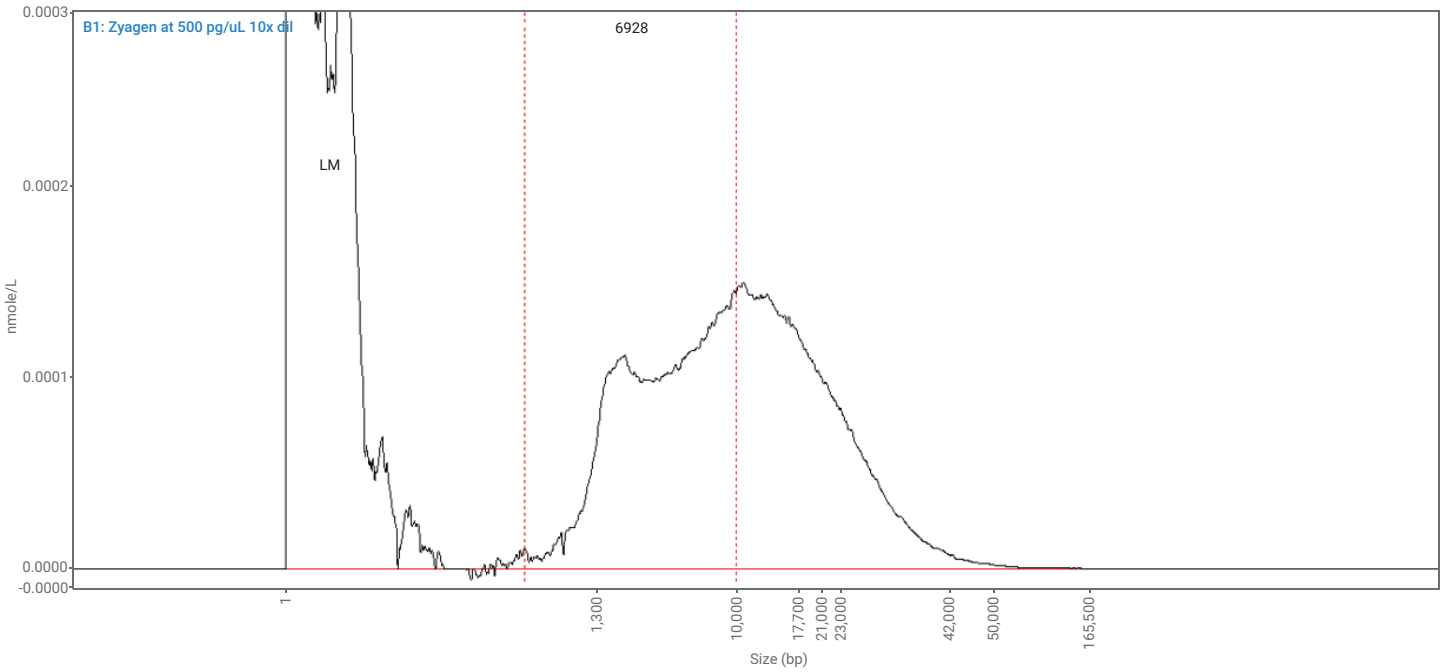
To demonstrate how the automated electrophoresis systems can aid in the assessment of smears, such as NGS libraries and high molecular weight DNA, Zyagen human gDNA was analyzed using the 165 kb gDNA kit on the Femto Pulse and evaluated using smear analysis. ProSize data analysis software allows for different smear regions to be set to assess specific portions of a sample. The number of nanograms and nanomoles is calculated for each smear region. In the example shown in Figure 2, two smear analysis regions were examined. The overall shape of the smear is altered when visualizing the electropherogram in terms of mass compared to moles. When visualized by mass, the smear is a symmetrical bell-shaped curve, with a slight bump on the left-hand side. However, when visualized in moles, this bump is enlarged compared to the rest of the smear.

For example, when the smear analysis is set from 1,000 to 10,000 bp, the mass of the sample below 10 kb is only about 20% (Figure 2A). However, the percent of moles in the sample, below 10 kb, is 66% (Figure 2B). With NGS sequencing, the number of reads at any given size is directly proportional to the number of molecules present at a given size. Thus, it can be inferred that if sequenced, most of the reads in this sample would be smaller than 10 kb. This analysis provides insight into whether additional size selection steps are necessary to achieve the desired sequencing read lengths.

A. Visualization by mass.



B. Visualization by moles.



C. Peak analysis.

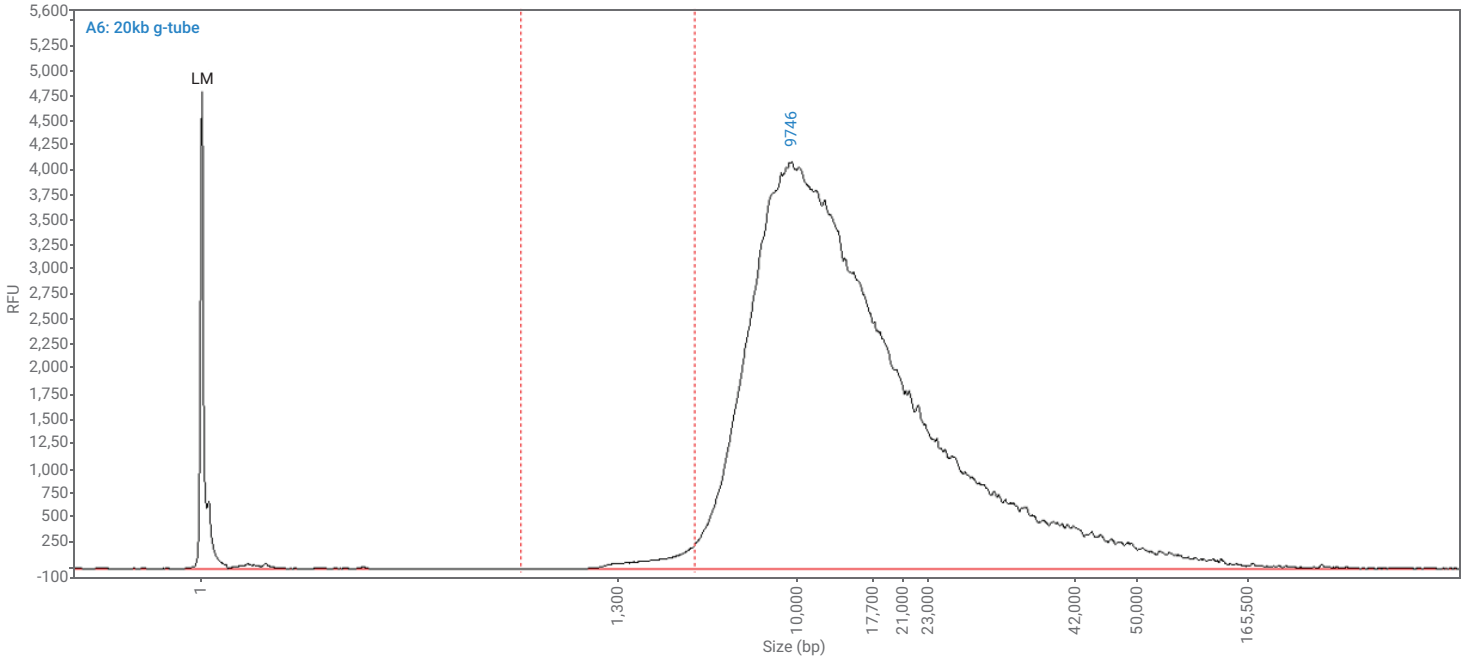
Smear Analysis Range (bp)	Percent of Mass	Percent of Moles
1,000 - 10,000	21.6%	66.0%

Figure 2. Zyagen human gDNA was analyzed on the Agilent Femto Pulse system with the Agilent Genomic DNA 165 kb kit. The resulting electropherogram can be visualized on the Agilent ProSize data analysis software with either A) mass or B) moles on the Y-axis. When assessed by moles it is evident that most of this sample is smaller than 10 kb, as reported in the smear analysis table (C).

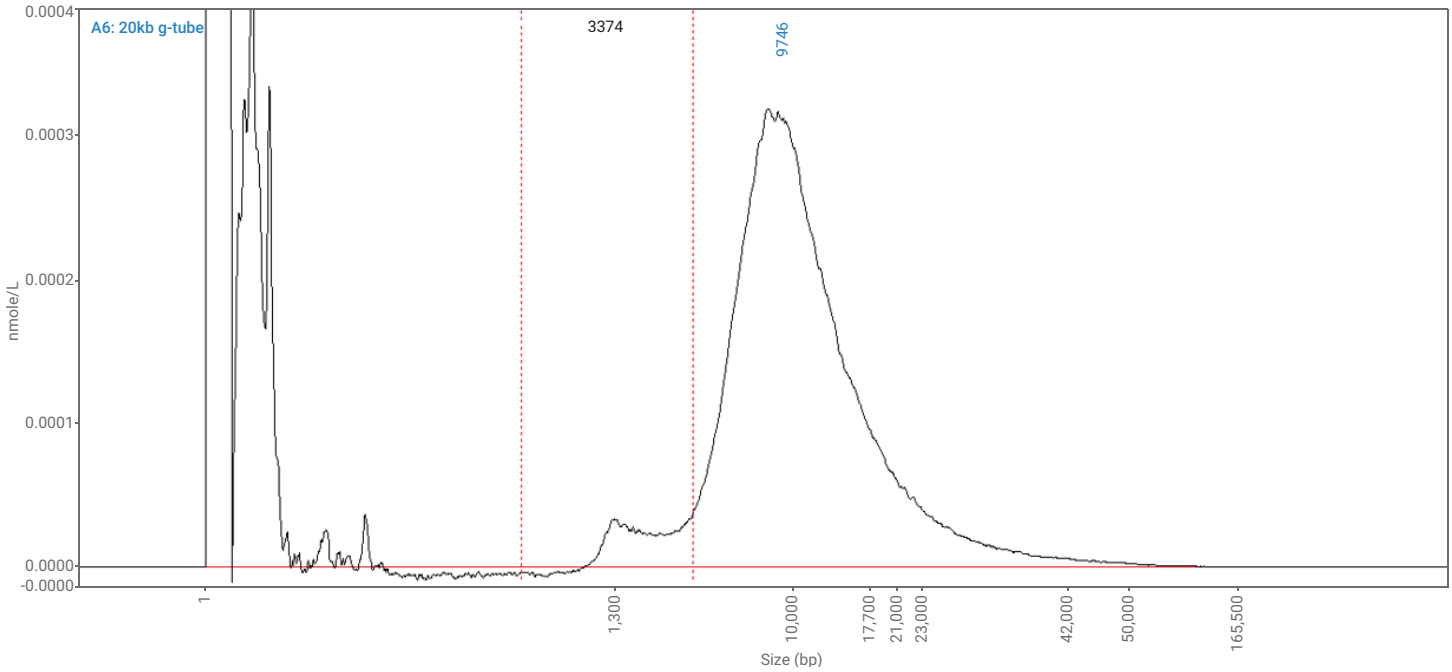
Molar visualization of NGS library corresponds with sequencing results

As demonstrated above, small-sized fragments within a sample can make up a significant number of the total moles present. However, this may not be clearly seen when visualizing the sample electropherogram by mass. Additionally, it is known that smaller-sized fragments within an NGS library are preferentially sequenced over the larger sizes. Long-read sequencing results can be maximized by loading only long fragments onto the sequencer, thereby eliminating any issues with this preferential sequencing of smaller fragments. However, proper visualization of the library is necessary to identify the presence of these small fragments and provide an indication of when additional size selection or cleanup steps are necessary in the library preparation workflow. It is thus important to obtain an accurate assessment of the proportions of the different areas of the library for successful sequencing results, which can be achieved by visualizing the electropherogram of the sample in moles, instead of mass, using ProSize. To demonstrate this, gDNA was sheared to 20 kb with a Covaris g-TUBE and analyzed on the Femto Pulse using the 165 kb gDNA kit. A smear analysis of the library from 1,000 to 5,000 bp indicated that this region encompassed 5% of the total mass of the library, but 18% of the total number of molecules in the library. The electropherogram of the library when visualized in mass (Figure 3A) shows a single smear with a peak maximum of 9,746 bp, and very little of the library within the smear analysis range (Figure 3C). However, when visualizing the electropherogram in moles (Figure 3B), a second peak at 1,759 bp is evident within the smear range, accounting for the 18% molarity total (Figure 3C). Upon sequencing this sample on an Oxford Nanopore MinION Mk1B device with MinION Flow Cells, the read length histogram showed two distinct peaks corresponding in size to the peaks seen when the electropherogram was set to nmoles on the y-axis (Figure 3D). Visualization of the data in both mass and moles provides valuable data for predicting long-read sequencing success.

A. Visualization by mass.



B. Visualization by moles.



C. Peak analysis.

Smear Analysis Range (bp)	Percent of Mass	Percent of Moles
1,000 - 5,000	5.0%	18.1%

D. Sequencing read lengths.

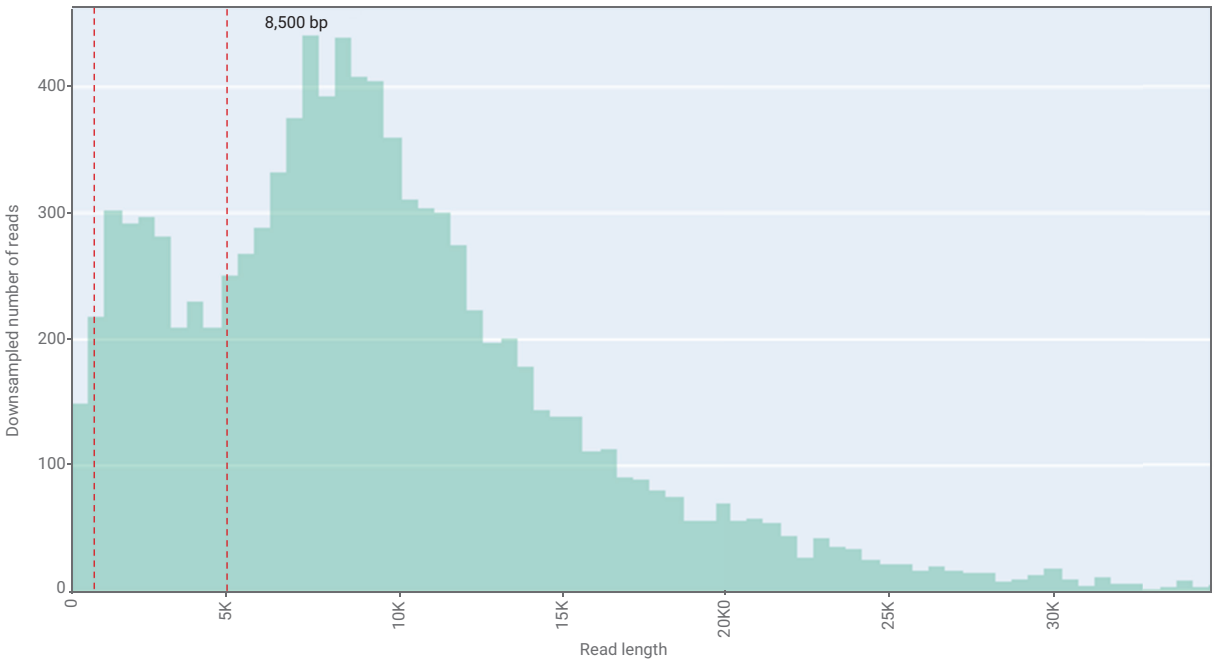


Figure 3. Genomic DNA was sheared to ~ 20 kb and analyzed on the Agilent Femto Pulse system with the Agilent Genomic DNA 165 kb kit prior to MinION sequencing. The resulting electropherogram can be visualized with the y-axis in A) mass or B) moles, with a smear range indicated by red lines at 1,000 to 5,000 bp. C) The percent of the mass and percent of moles within the smear range compared to the total library. D) The library was sequenced with the ONT MinION, and the number of reads of each read length shown in the histogram, which aligns more closely with the Femto Pulse electropherogram when visualized in moles.

Conclusion

The Agilent Femto Pulse system was designed to provide high sensitivity and resolution for accurate analysis of nucleic acids. When analyzing high molecular weight DNA samples by mass, the visual representations generated by different analytical software underrepresent the smaller fragments present in the sample. This corresponds with sequencing data that suggests that smaller fragments within a library are preferentially sequenced over larger fragments. The Agilent ProSize data analysis software allows users the option to visualize the electropherogram in moles, instead of mass. Visualization of the sample by moles using the Femto Pulse system can provide a clearer assessment of the length distribution, and better predict sequencing results.

References

1. Mandal, C.; Englander, S. W.; Kallenbach, N. R. Hydrogen-Deuterium Exchange Analysis of Ligand-Macromolecule Interactions: Ethidium-Deoxyribonucleic Acid System. *Biochemistry* **1980**, 19 (25), 5819–5825. https://doi.org/10.1021/bi00566a025.

2. Comparison of Agilent Femto Pulse System Sizing with Long-Read Sequencing Read Length. *Agilent Technologies application note*, publication number 5994-3078, **2021**.

www.agilent.com/genomics/automated-electrophoresis

PR7001-2428

Published in the USA, March 19, 2024
5994-7266EN

Agilent
Trusted Answers