Technical Note

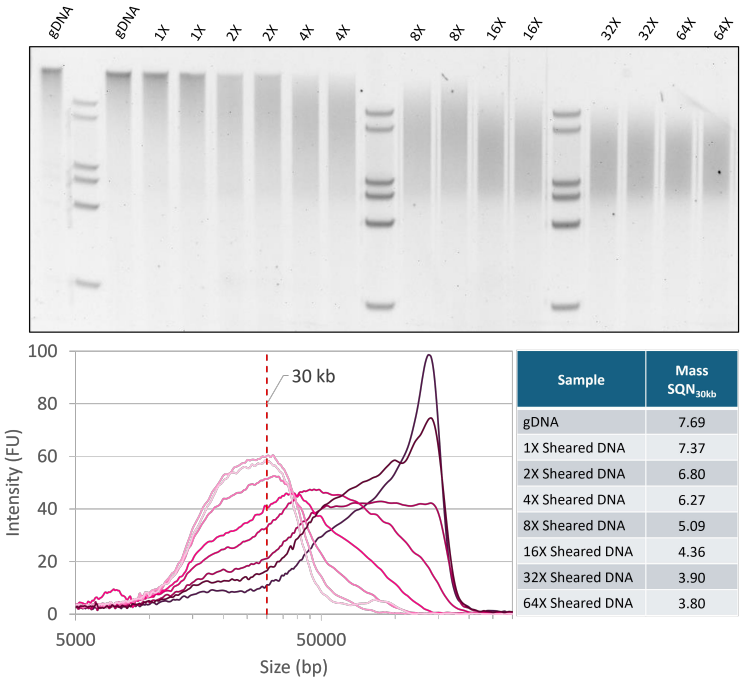
Quality Control and Size Selection for PacBio® Long-read Sequencing using the LightBench® Discover at HudsonAlpha's CoLab

Introduction

The integrity and sizing of input DNA can substantially impact the success of long-read sequencing, especially for whole-genome sequencing applications.

Options for quality control (QC) of nucleic acid samples and libraries in long-read sequencing workflows are limited, particularly due to the need to resolve DNA fragments larger than 20 kb. The LightBench® Discover addresses this challenge with integrated pulsed-field electrophoresis, providing fragment length analysis of DNA samples up to 150 kb. It also maintains functionality for gel-based size-selection that was introduced on the original LightBench® instrument. The LightBench® Discover thus provides a solution to complete both size selection and QC steps in long-read sequencing workflows.

The LightBench® Discover's assessment of HMW DNA quality was demonstrated at PacBio's Applications Lab using a single genomic DNA sample separated into different aliquots, and each subjected to varying levels of tip shearing via manual pipette mixing. For each tip shearing condition, samples were assessed in duplicate using the Large Fragment Analytics 3-75 kb protocol on the LightBench® Discover. Analysis with Ranger® Analytics software clearly shows a reduction in DNA fragment size with increased shearing. Correspondingly, the Sample Quality Number (Mass SQN_{30kb}) generated by the Ranger® Analytics application also declines as the degree of shearing increased (Figure 1). These results demonstrate that the LightBench® Discover can effectively evaluate the quality and degradation status of HMW gDNA samples – detecting the effect of a single pipette mix operation.



Sample	Mass SQN_{30kb}
gDNA	7.69
1X Sheared DNA	7.37
2X Sheared DNA	6.80
4X Sheared DNA	6.27
8X Sheared DNA	5.09
16X Sheared DNA	4.36
32X Sheared DNA	3.90
64X Sheared DNA	3.80

Figure 1 - Gel image captured by the LightBench® Discover during electrophoresis used to generate the electropherogram traces and Mass SQN at 30 kb.

Adoption at HudsonAlpha's CoLab

The CoLab at the HudsonAlpha Institute for Biotechnology[1] has integrated the LightBench® Discover into their PacBio whole genome sequencing (WGS) workflow, using it for size-selection of prepared libraries as well as sample QC at three key stages (Figure 2):

1. High molecular weight (HMW) DNA
2. Sheared DNA
3. Final HiFi DNA libraries

[1] The CoLab is a cooperative lab comprising the research teams of Dr. Alex Harkess and Dr. Josh Clevenger. Together, they use genomics to better understand plant reproduction and investigate potential crop improvements for sustainable agriculture.

From January to June 2025, the CoLab processed 1199 samples (1051 HMW analytics + 148 HMW size selection) through this workflow, using the LightBench® Discover for QC and size selection.

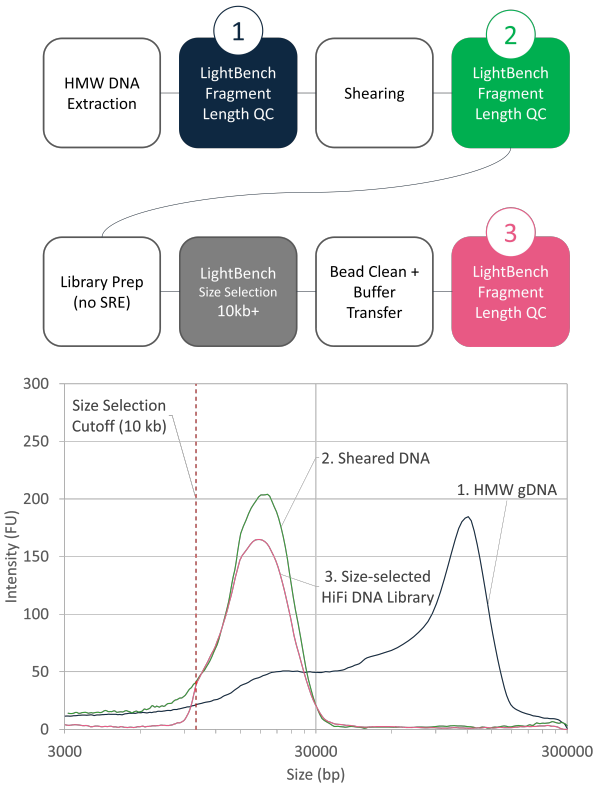
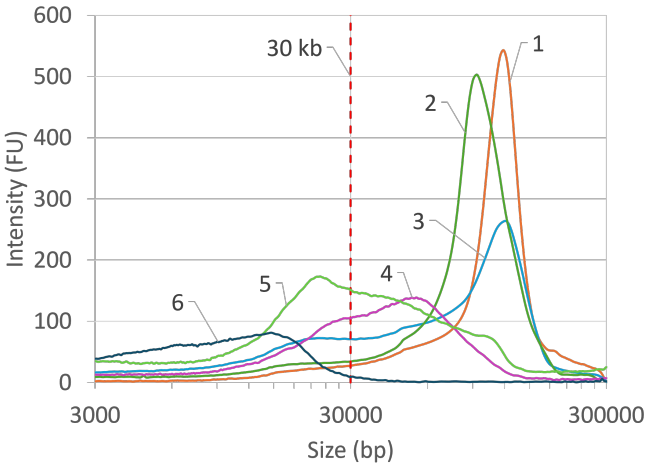


Figure 2 – Schematic of the integration of the LightBench® Discover in the CoLab's long-read sequencing workflow. Numbered positions indicate QC step performed by the LightBench® Discover (top). Electropherogram traces produced at each QC step (bottom). The vertical line indicates the location of the 10 kb size-selection cutoff.

Quality Control of HMW DNA

Following HMW DNA extraction, 1 µL of the sample is evaluated directly on the LightBench® Discover, without quantification or dilution. Representative electropherogram traces produced from four different plant species highlight variations in DNA quality, further reflected by the Mass SQN_{30kb} value (Figure 3). The fragment length distribution for a given sample informs whether to perform shearing prior to SMRTbell library preparation. In cases of very low-quality DNA, it may be decided not to proceed with library preparation at all, instead, returning to DNA extraction where possible.



Sample	Mass SQN_{30kb}
1	9.18
2	8.08
3	6.18
4	5.60
5	4.09
6	0.24

Figure 3 - Electropherogram traces generated by LightBench® Discover for four plant gDNA samples. SQN metric produced using the Ranger Analytics application indicates the proportion of DNA (by mass) >30 kb.

Quality Control of Sheared DNA

Following shearing, samples are evaluated to assess the efficacy of DNA shearing and the portion of fragments <10 kb (Figure 4).

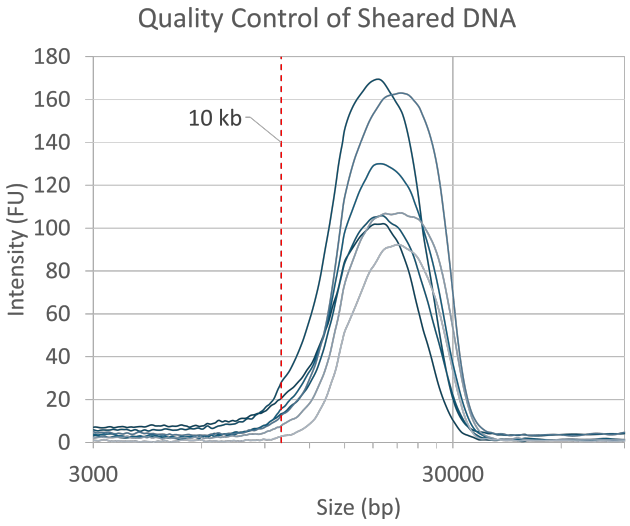


Figure 4 - Overlay of sheared DNA electropherograms with decreasing Mass SQN_{10kb} due to higher number of DNA fragments <10 kb.

If the desired fragment length distribution has not been achieved, the shearing is repeated and the sample reanalyzed on the LightBench® Discover prior to library preparation.

Size Selection and Final Quality Control of SMRTbell Libraries

Preferential sequencing of smaller DNA fragments can lower HiFi mean read length and data quality. Gel-based size selection on the LightBench® Discover was used to remove DNA fragments below 10 kb leading to increased mean read length from subsequent sequencing. PacBio's Small Read Eliminator (SRE) kit is not used in this workflow, being found unnecessary given the use of gel-based size-selection.

Fragment length analysis of the final size-selected SMRTbell library on the LightBench® Discover shows complete removal of small DNA fragments and a corresponding increase in the SQN value compared to the sheared DNA (Figure 5). The measured size of the size-selected library is then used to inform molar dilution of libraries prior to sequencing.

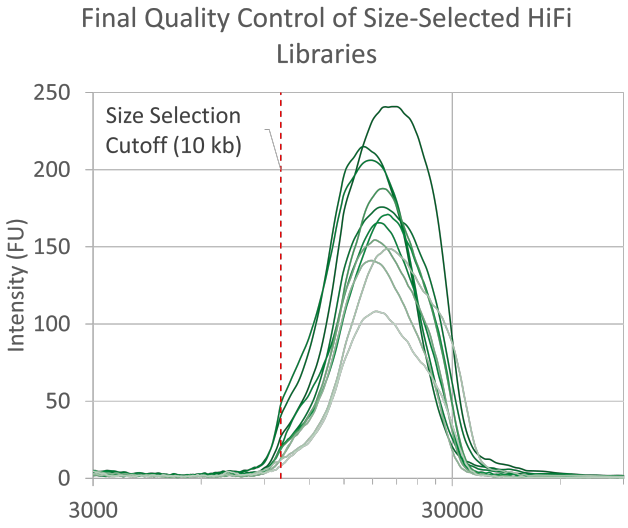


Figure 5 - Overlay of electropherograms traces of final QC of HiFi Libraries prior to sequencing.

In the comparative example below, the CoLab prepared two SMRTbell libraries from the same extracted gDNA – one of which underwent size selection using the LightBench® Discover. The size-selected library yielded a higher HiFi mean read length, resulting in more reads for fragments >10kb. Additionally, methylation calling data shows that size-selection has no impact on DNA methylation states (Figure 6).

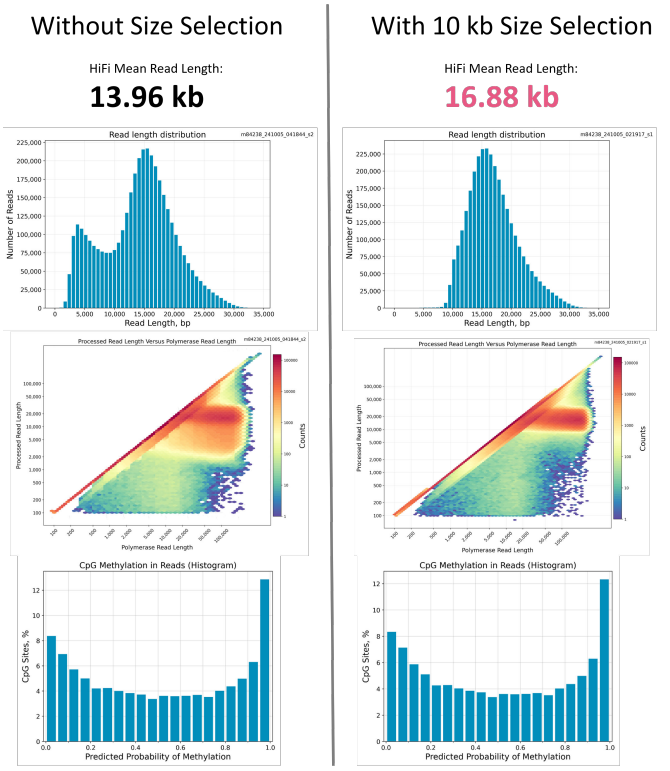


Figure 6 - Two SMRTbell libraries were generated from the same extracted HMW DNA and after having been processed using PacBio's SRE kit. One library was size selected on the LightBench Discover while the other was not. Mean HiFi Read Length was larger for the size-selected library. DNA methylation calls were unaffected by the size selection (Revio V1 chemistry).

The average size of 23 size-selected SMRTbell libraries was measured on the LightBench® Discover and compared to the corresponding HiFi mean read length, showing good correlation, and a ~10% discrepancy (~16.8 kb vs ~15.2 kb). This measured offset is consistent with the expectation that shorter library fragments will be converted to HiFi reads more efficiently than longer fragments during sequencing, decreasing the HiFi mean read length relative to the true average size of the input library (Figure 7). Sample-dependent variability in the magnitude of this bias contributed to differences in the correlation between average library size and HiFi mean read length.

PacBio

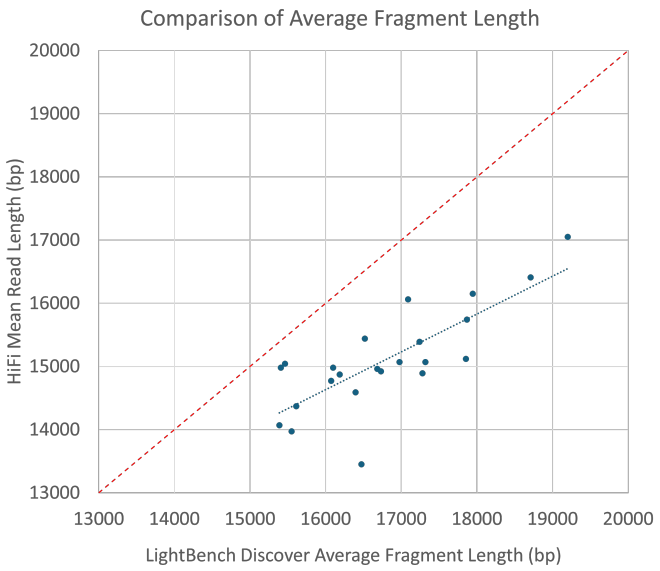


Figure 7 - Comparison of the average fragment length measured on the LightBench® Discover with the mean HiFi read length.

Conclusion

HudsonAlpha's CoLab has successfully implemented the LightBench® Discover in their PacBio WGS workflow, using it for precise gel-based size selection and fragment length quality control of all samples. Gel-based size selection on the LightBench® Discover consistently removes DNA fragments <10 kb, increasing mean HiFi read lengths. The fragment length distribution data informs real-time decisions at key stages in the workflow, including assessment of HMW DNA quality, confirmation of DNA shearing, and verification of the final DNA library size and quality. The LightBench® Discover offers a reliable solution for integrated size-selection and quality control in long-read sequencing applications.

Large Fragment Analytics Specifications

Input DNA Mass	10-100 ng
Functional Range (dependent on chosen protocol)	3 kb to 75 kb OR* 20 kb to 150 kb
Throughput	20 samples
Availability of initial low-resolution data	30 mins
Complete Run Time	120 mins
Sizing Accuracy	**Sample QC 3 kb to 75 kb protocol** 3 kb to 50 kb: 10% 50 kb to 75 kb: 20% **Sample QC 20 kb to 150 kb gDNA protocol** 20 kb to 150 kb: 20%

PROD0015 Rev 1

Required Equipment and Consumables

Equipment	Part Number
LightBench® Discover	CG-12500-05
LightBench® Discover Install Kit	CG-12491-30
Consumable Kit	
Large Fragment Analytics DNA Kit	CG-14500-20-050-24-DL
Includes:	
Dual Dye Loading Buffer - 2 kb marker	CG-14400-24-24
Large Fragment Analytics DNA Ladder	CG-14450-01
0.5% Large Fragment Analytics Cassette, 24 channel	CG-10600-20-050

Acknowledgements

Thanks to Alex Harkess, Josh Clevenger, Haley Hale, and Paige Phelps (CoLab, HudsonAlpha Institute for Biotechnology, Huntsville, AL.) for providing LightBench® Discover fragment analysis and Revio sequencing data. Thanks to Christine Lambert (PacBio Applications Lab, Menlo Park, CA) for generating the fragment length data using tip-sheared gDNA.

Discover More:

yourgenehealth.com

ranger@yourgenehealth.com

Yourgene Health

LightBench® Discover Resources:

Scan the QR code to learn more

Yourgene Health
Skelton House
Lloyd Street North
Manchester Science Park
Manchester
M15 6SH, UK