Application Note

Best practices: Sage Science size selection with Diagenode Megaruptor shearing for long-read sequencing library preparation

Christina MacNaughton[1], Madelyn Tolland[1], Kathryn Withers[1], Eric Veien, Wassim Lakhal[2], Francis Wilke[2], Sandy Carrillo[2], Matthieu Bossan[2], Céline Sabatel[2], Chris Boles[1]

[1] Sage Science, Inc., Beverly, MA, USA
[2] Diagenode, a Hologic Company, Liège, Belgium

Introduction

PacBio HiFi sequencing utilizes input DNA with a relatively tight size distribution centered within a 17-20kb size window. At many genomics labs, the workflow for generating such tightly sized input is to shear high molecular weight DNA (HMW, >30kb) using the Diagenode Megaruptor 3 followed by size selection on the Sage BluePippin or PippinHT. In this scheme, the Megaruptor shearing determines the average size and upper limit of the input DNA, and the size selection on the Sage instrument provides a sharp low molecular weight cutoff at around 15-17kb.

A Diagenode-Sage workflow may also prove useful for the recently developed "high duplex" sequencing at Oxford Nanopore, based on initial feedback from presenters at recent genomics meetings (Spring, 2023). In summary, there are some indications that a library size between 20-50kb may offer a good balance between fraction of duplex reads and overall sequence throughput for duplex sequencing.

For these reasons, Sage and Diagenode initiated the present study to determine best practice guidelines for using our two instrument systems in combination. We performed BluePippin and PippinHT size selections on Megaruptor 3-sheared samples ranging in average size from 11kb to 33kb. We used new Sage cassette definitions that are optimized for almost linear resolution between 9kb and 30kb, along with a new programming mode called "Range + T" that offers improved controls over the upper limit of the output size distribution. Our results demonstrate that successful long-read library sizing require careful coordination of Sage instrument settings with the average size of Megaruptor 3-sheared DNA inputs.

Materials and Methods:

High quality purified HMW human DNA was purchased from Promega.

At Sage, HMW DNA was sheared in 500ul aliquots at a concentration of 150 ug/ml using Megaruptor 3 "high concentration" protocols for one shearing cycle each using speeds of 29 and 31 and two cycles at speeds of 33, 36, and 40 (**Figure 1**).

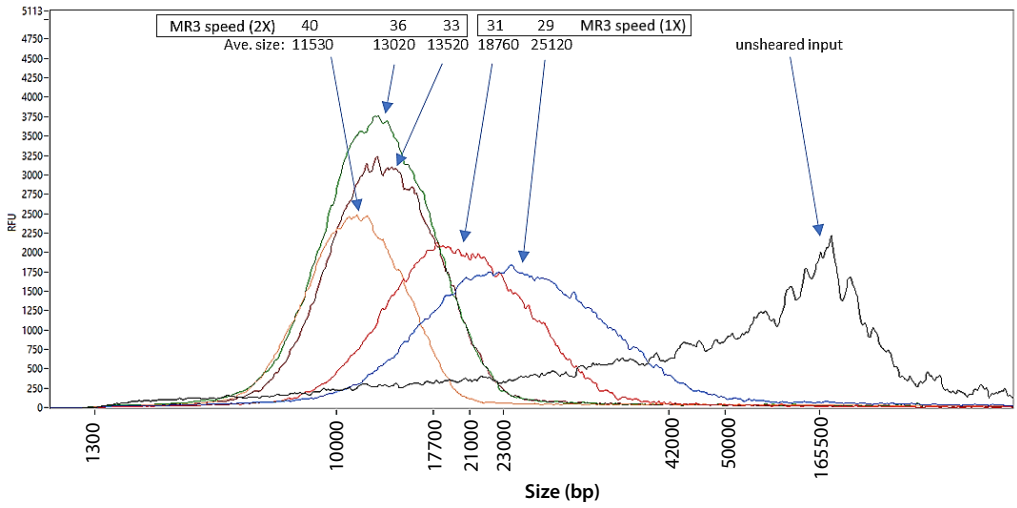

Figure 1. Megarupter DNA shears; one cycle with speeds of 1 X 29 and 31 and 2 X 33, 36 and 40. 500ul aliquots at 150 ug/ml.

In addition, other Megaruptor 3-sheared samples were generated at the Arizona Genomics Institute (AGI) using 100ul aliquots of Promega DNA at 150 ug/mL using one shearing cycle each at speeds of 29, 31, and 34. One additional sample was prepared by AGI using one shearing cycle at speed 28 followed by one additional shearing cycle at speed 29 (**Figure 2**).

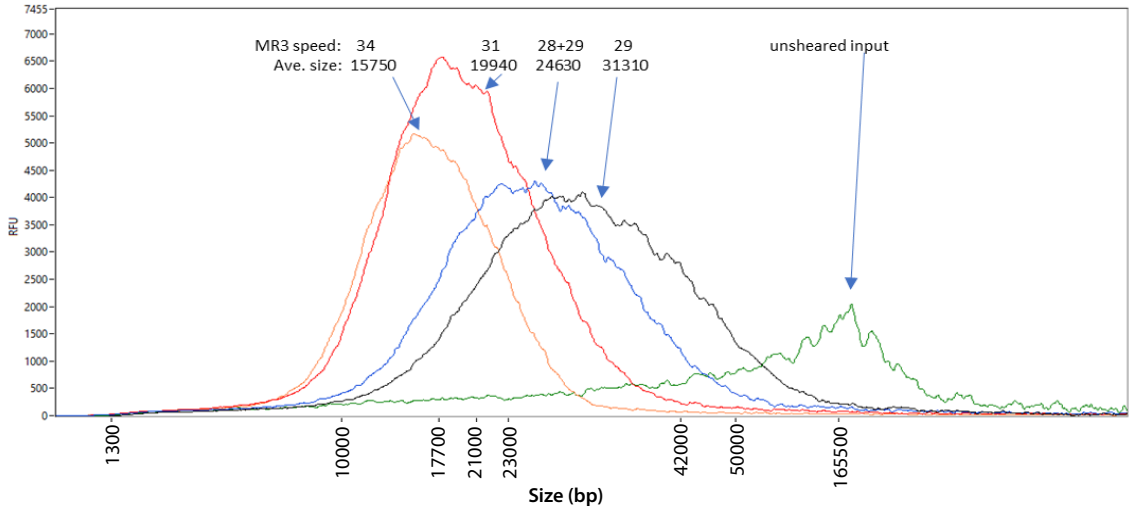

Figure 2. Megarupter DNA shears; one cycle with speeds of 29,31,and 34 and one two-cycle shear with speeds of 28 and 29. 100ul aliquots at 150 ug/ml.

Each of the sheared DNA samples was subjected to size selection on Sage BluePippin and PippinHT 0.75% agarose cassettes using new cassette definitions that have an extended range of linear resolution, allowing more accurate LMW cutoffs between 9kb and 30kb (**Figure 3**). In addition, we used a new Sage programming mode called "Range+T" (range plus time). In this mode, the user selects the DNA size at which product elution should start, and an elution time in minutes. This programming mode is designed to address the inaccuracy in terminating elution in Sage "Range" mode when the input samples that have a narrow size distribution.

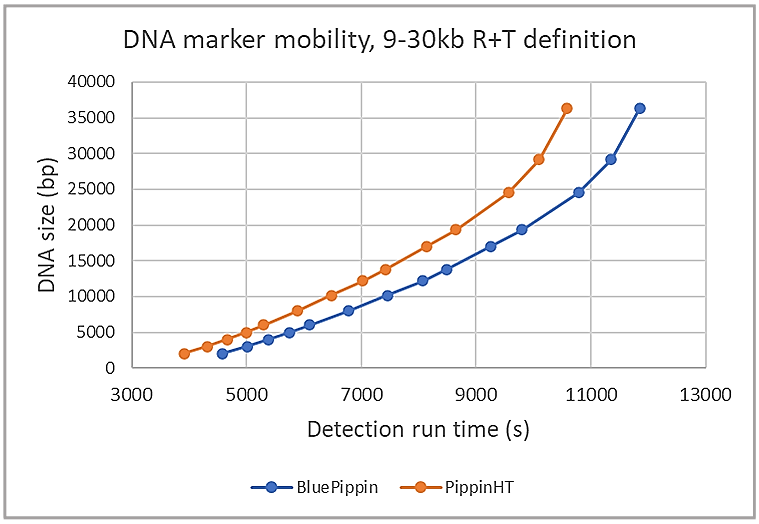

Figure 3. Linear range of resolution (9-30kb) on BluePippin and PippinHT 0.75% agarose gel cassettes with "Range+T" electrophoresis waveforms.

Size-selection products were quantified by Qubit HS assay (Life Technologies), diluted to 200 pg/ul in TE buffer (10mM Tris-HCl, 1mM EDTA pH 8) and analyzed on the Agilent FemtoPulse using the Genomic DNA 165kb kit. For Megaruptor 3 sheared inputs and the Sage size-selected samples, the average DNA size was determined using the NGS mode of Agilent ProSize software (version 4.0.2.7). To determine the LMW cutoff of the Sage size-selected samples in an objective manner, the FemtoPulse electropherogram data was exported to Microsoft Excel, and the DNA size at the leading edge of the size-selected product distribution at a height of 4.4% of the peak product RFU was used.

Results

Using a wide variety of Megauptor 3-sheared inputs with average sizes ranging from 11kb to 33kb, we performed size selection using LMW cutoff values chosen to be less than, equal to, and greater than the average size of the sheared inputs. The measured size of Sage product at the start of collection and the average size of the product were measured using the Agilent FemtoPulse. For both Sage instruments, the measured start of collection values were in good agreement with programmed start setting if the programmed start settings were smaller than the average size of the sheared input DNA (BluePippin data, **Figure 4**, left column); PippinHT data **Figure 5**, left column). When programmed start settings were larger than the average size of the sheared input the measured start of collection values were extremely variable and substantially lower than the programmed start values.

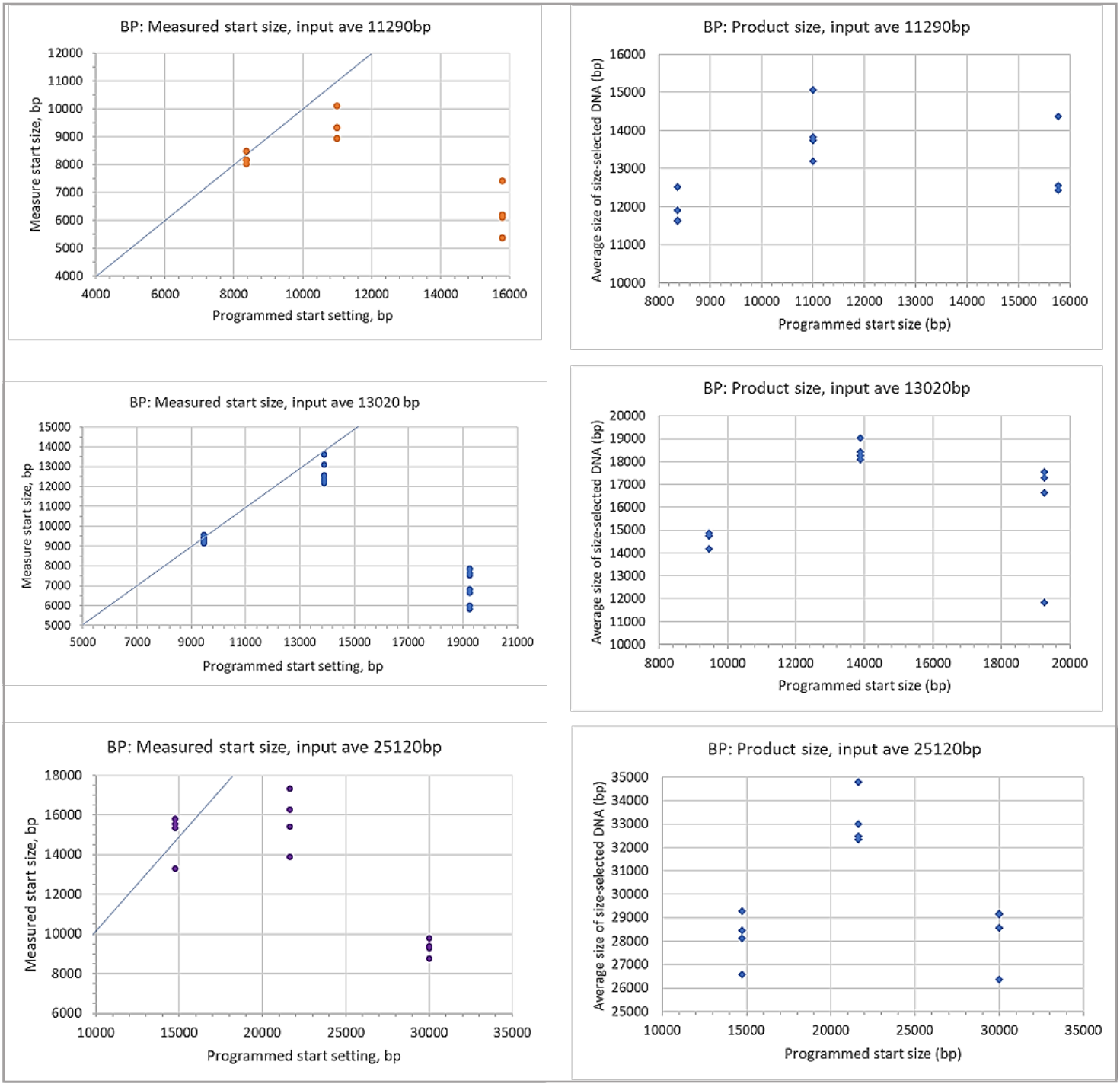

Figure 4. DNA size selection on the BluePippin. Start values are in the left column and average peak distribution sizes are in the right column.

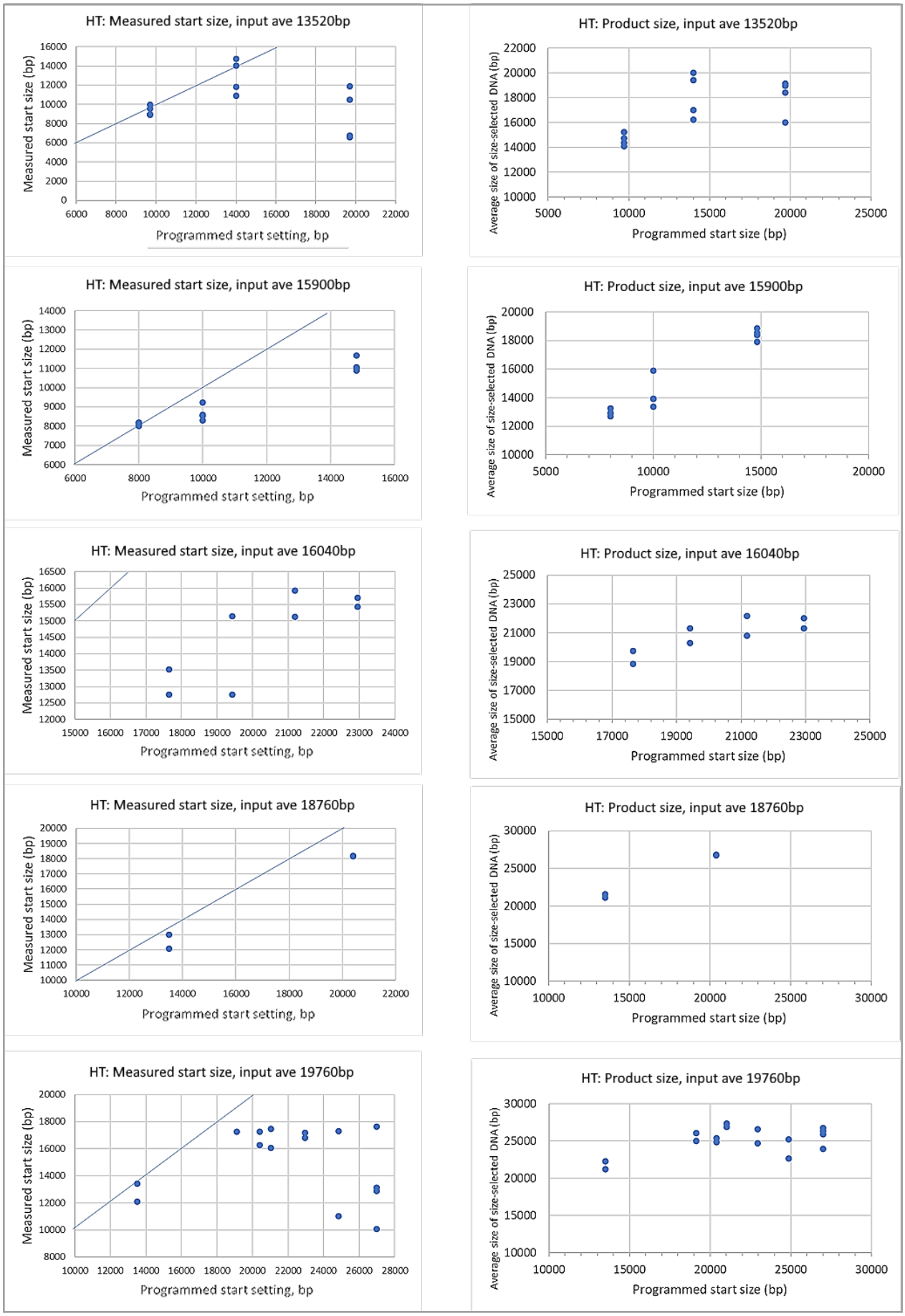

Figure 5. DNA size selection on the PippinHT. Start values are in the left column and average peak distribution sizes are in the right column.

When programmed start settings were equal to the average size of the sheared input, the measure start sizes were in reasonably good agreement with programmed values for smaller sheared inputs (see BluePippin data, **Figure 4**, for 11kb and 13kb inputs; PippinHT data, **Figure 5**, for 13kb input), but were less consistent for larger inputs.

To evaluate consistency of these trends with Megaruptor-sheared samples of different size, the percent error in measured start size was plotted as a function of the ratio of programmed start size to average input size (PippinHT data, **Figure 6, upper**; BluePippin data, **Figure 6, lower**). The data confirm that the trends in start value accuracy are similar across all sheared input sizes.

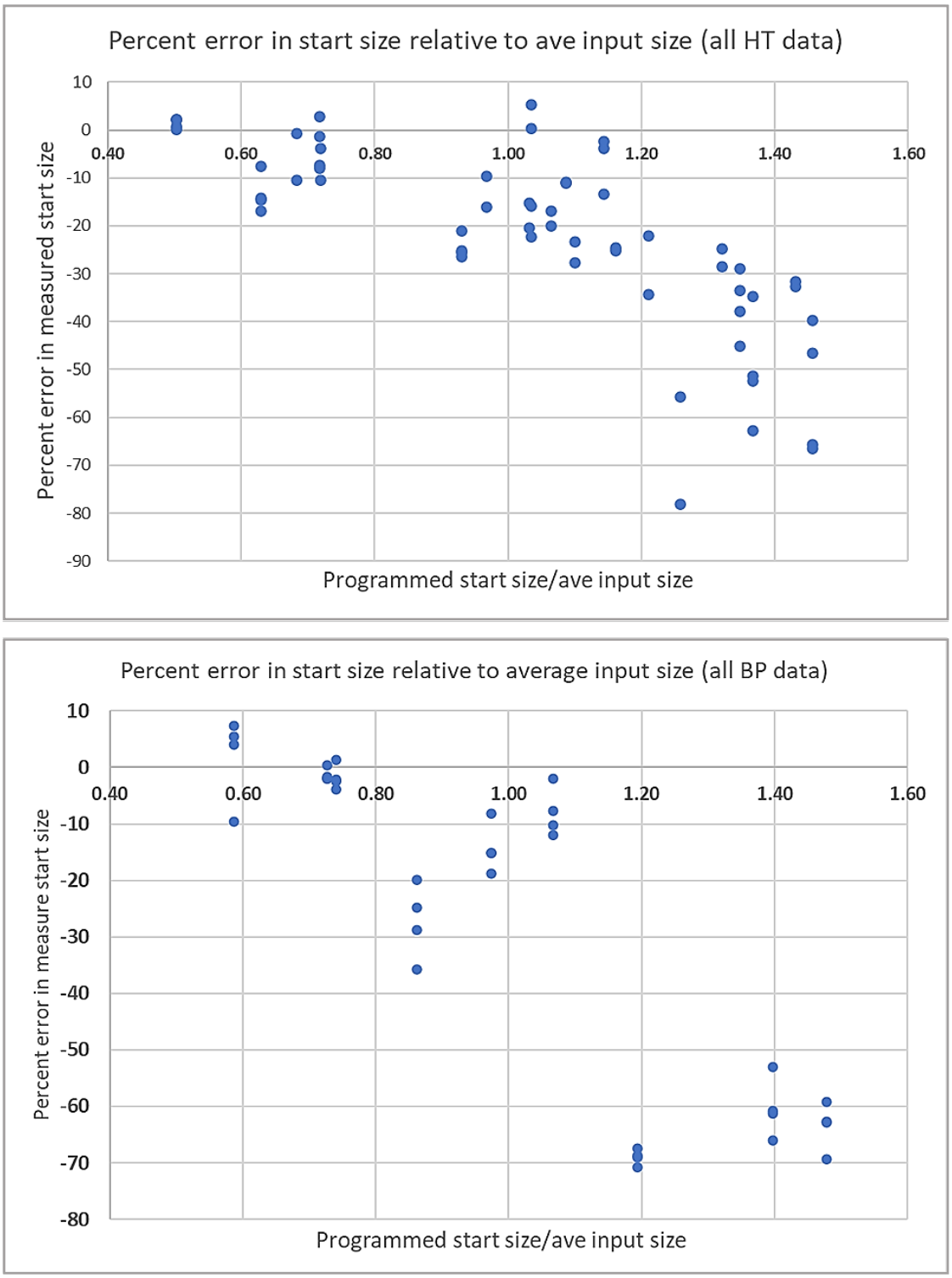

Figure 6. The percent error in measured start size, plotted as a function of the ratio of programmed start size to average input size (PippinHT upper, BluePippin lower)

To study the trends in accuracy for start values near the average input size, a larger, more detailed set of measurements was performed using two different 19kb sheared samples on the PippinHT (**Figure 7**). These data confirm the earlier findings and demonstrate consistent accuracy of the PippinHT size selections up to the average input size (at around -15% error).

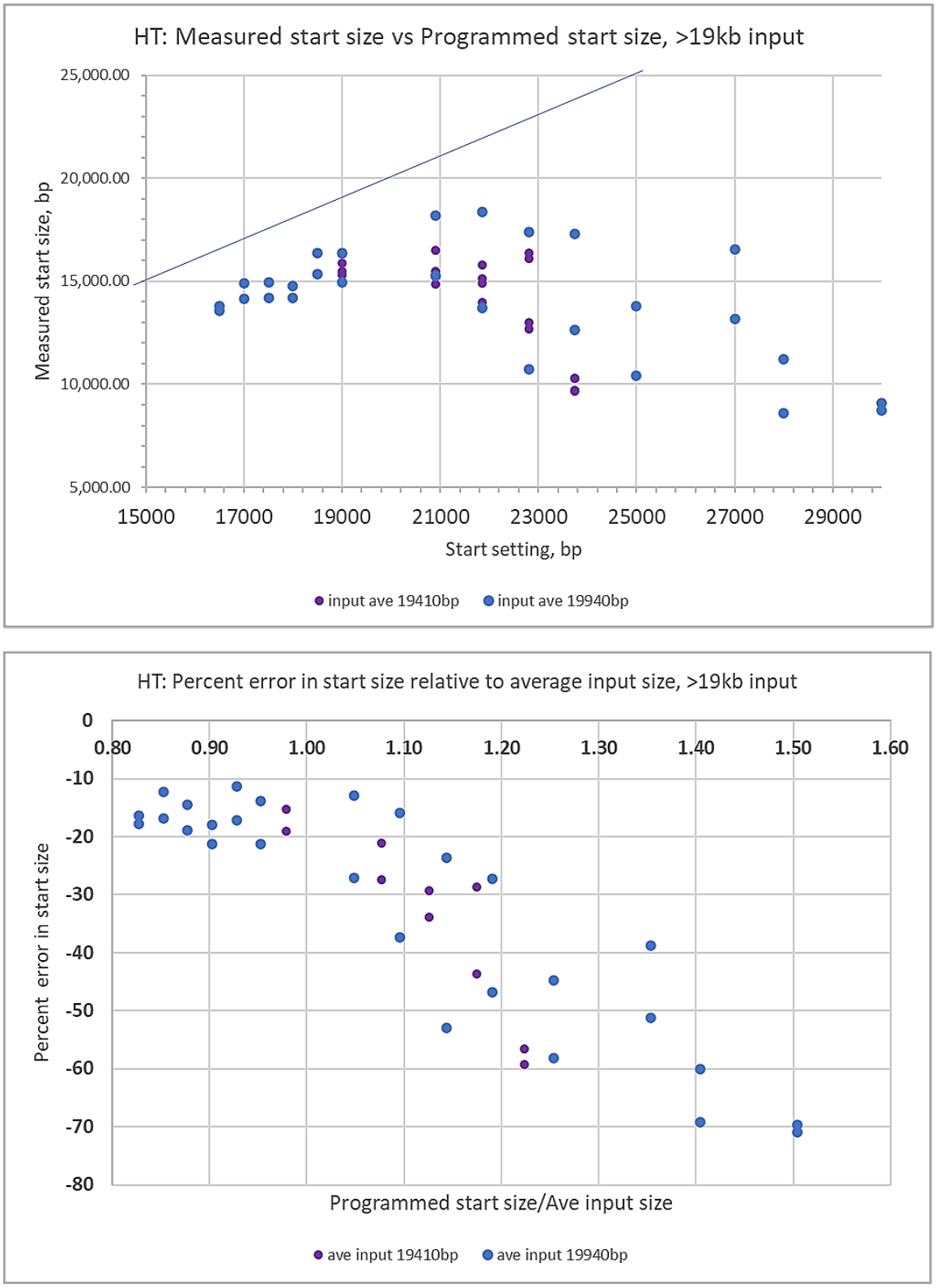

Figure 7. A detailed set of measurements was performed using two different 19kb sheared samples on the PippinHT shows consistent accuracy of the PippinHT size selections up to the average input size.

A similar trend was seen in the average size of Sage outputs. With programmed settings smaller than the average size of the sheared input sample, measured average size of the output increased with increasing programmed start values, but plateaued or decreased with programmed start values that were larger than the average input size.

To investigate possible causes for the inaccuracy of Sage size selection at setting greater than the input DNA size, we inspected FemtoPulse traces of size selected products across a variety of start size settings prepared from the same sheared input (**Figure 8 upper**, BluePippin, 11.3kb sheared input, **Figure 8 lower**, PippinHT, 20kb sheared input). In the PippinHT samples using a 20kb input, programmed start settings of 16.5kb, 18.5kb, and 19kb give products with sharp LMW cutoffs. However, at start setting of 21.5 and 22.8kb, the sharp LMW cutoff is lost and a pronounced LMW foot is present. At an even larger start setting of 30kb, this LMW foot enlarges to a pronounced bimodal shape that extends down to the lower limit of the input sample. A similar trend is seen using a much smaller 11.3kb average input (BluePippin, **Figure 8 upper**). Programmed start values of 8.4kb and 11kb give products with sharp LMW cutoffs, but a start value

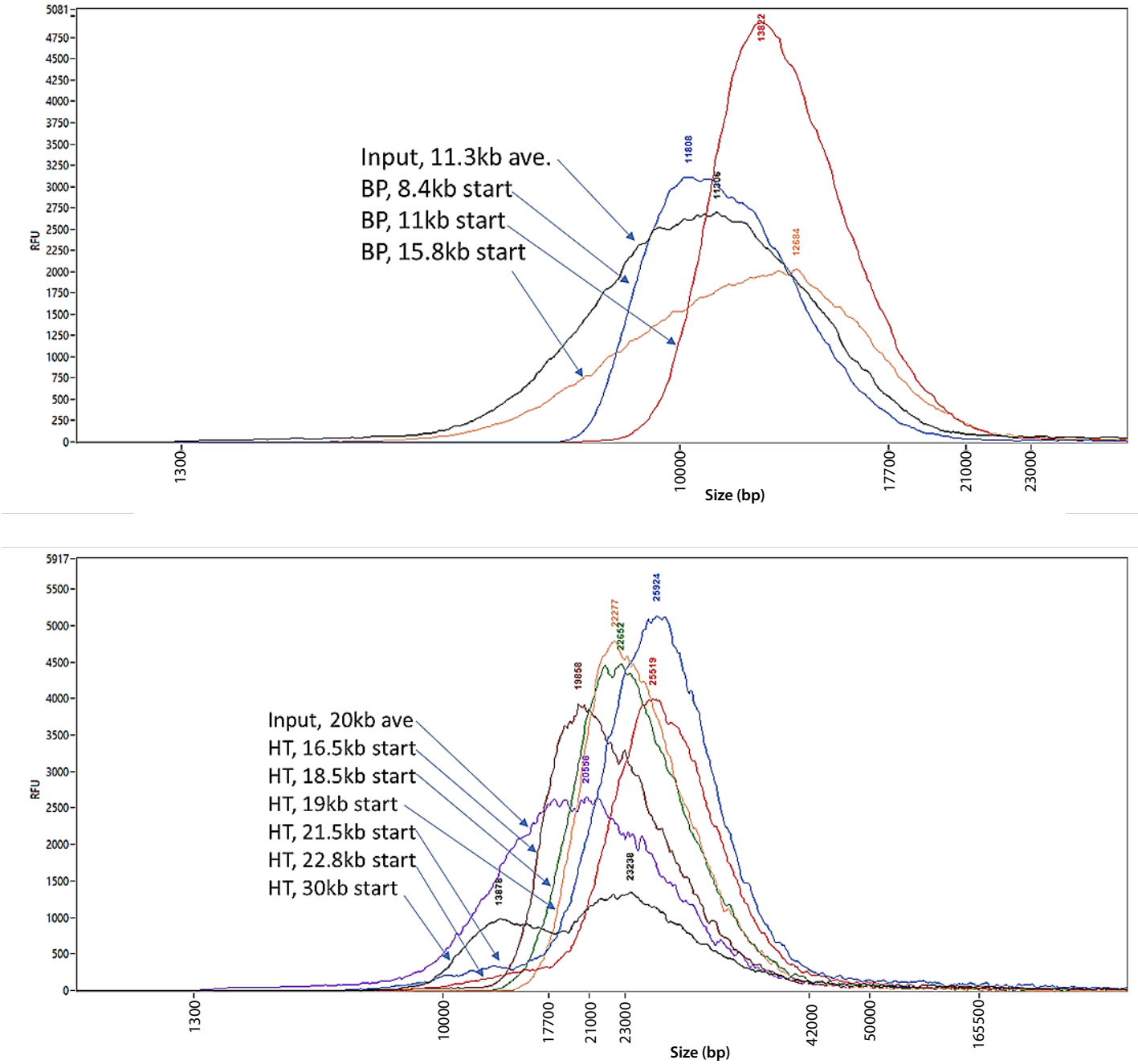

Figure 8. Agilent Femtopulse traces of starting size selection values before and after the average size of the input DNA; BluePippin (upper) and PippinHT (lower). Starting selection after the average sizes show a loss of LMW cutoff.

of 15.8kb gives a very broad product with a leading edge that extends down to the lower size limit of the input DNA. This band broadening phenomenon is similar to band shape distortions in capillary electrophoresis due to sample overloading termed "electromigration dispersion"(1,2,3). Band broadening in overloaded CE had been attributed to sample mobility alterations due to changes in the relative ionic contributions of the electrophoresis buffer ions and sample ions (DNA, in this case) from the front to back edges of the sample zone. Since these experiments used DNA inputs near the maximum possible inputs of the PippinHT and BluePippin cassettes (1.5 and 5 ug, respectively), it seems likely that the electromigration dispersion concept also applies to our size-selections on input samples with very narrow size distributions.

In all of the size selections shown so far (**Figures 4-8**), we used the new Sage Range+T programming mode, with an elution time of 30 minutes. This elution time is relatively long and was designed to capture most of the input DNA above the programmed low threshold size. To demonstrate the flexibility of the Range+T programming mode, we size selected two input samples with average sizes of 11.3kb and 33.6kb using programmed start values of 8kb and 17.5kb, respectively, and eluted for 7.5, 15, and 30 minutes. FemtoPulse traces of the outputs are shown in **Figure 9**. As expected, the width of the size-selection broadens with increased elution time. This programming feature gives the user additional control over the product distribution.

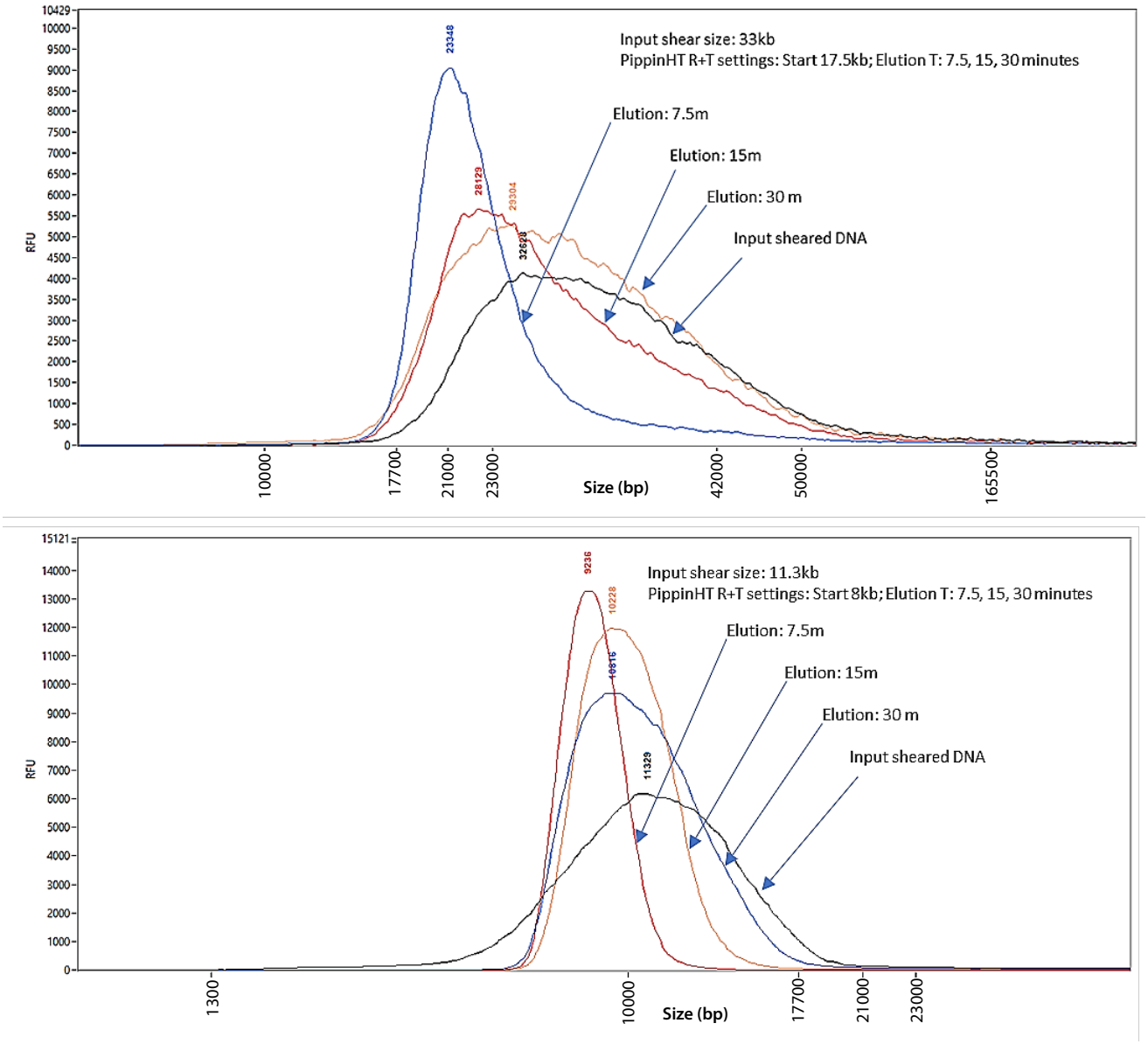

Figure 9. Aglilent Femtopulse traces of Range+T programming mode, with two input samples with average sizes of 33.6kb (upper) and 11.3kb (lower) using programmed start values of 17.5kb and 8kb, respectively, and eluted for 7.5, 15, and 30 minutes.

From the data presented in **Figures 4 and 5**, we compiled **Table 1** of Megaruptor 3 and Sage instrument settings that are useful for producing sheared genomic samples with average sizes ranging from 13kb to 29kb. Under the conditions listed, there is very little low molecular weight contamination below the Sage start of collection value. These settings should be used as approximate guidelines, since the Megaruptor 3 shearing depends on the size, volume, and concentration of the input DNA samples. In addition, as discussed above, the accuracy of the Sage size selection will depend on the ratio of the start setting to the average size of the sheared input sample.

Figure*	Final product size (ave)	Megaruptor 3 settings	Sage instrument	Size selection settings (all 30m elution)
5	~13kb	Speed 34, 100ul, 150ug/ml	HT	Start 8000bp
5	~15kb	Speed 33, 500ul, 150ug/ml	HT	Start 8500bp
5	~22kb	Speed 31, 500ul, 150ug/ml	HT	Start 13,500bp
5	~22kb	Speed 31, 100ul, 150ug/ml	HT	Start 13,500bp
4	~12kb	Speed 40, 500ul, 150ug/ml	BP	Start 8300bp
4	~14-15kb	Speed 33, 500ul, 150ug/ml	BP	Start 9300bp
4	~18-19kb	Speed 33, 500ul, 150ug/ml	BP	Start 13,000bp
4	~27-29kb	Speed 29, 500ul, 150ug/ml	BP	Start 15,000bp

Table 1. Megarupter 3 and Sage Instrument (HT = PippinHT and BP=BluePippin) recommended settings guidelines to produce a final product size with minimal LMW DNA below the start of collection. Data is compiled from Figures 4 and 5.

Conclusions

For long-read sequencing applications which require a relatively tight size distribution and a sharp LMW size cutoff, careful coordination of Megaruptor shearing and Sage size-selection conditions are important. In general terms, the Megaruptor shearing conditions should ideally be chosen to provide a sheared DNA input with an average size at, or slightly above, the desired library size. Sage Range+T programmed start size should be smaller than the average sheared input size. For these reasons, careful size QC of sheared samples is important for choice of the most useful Sage size-selection settings.

Choice of Sage start sizes greater than the average size of the sheared input sample will produce unpredictable results, usually yielding outputs with large amounts of DNA smaller than the programmed start size. Such outputs will compromise library insert lengths and read N50s.

The upper boundary of the size distribution will be determined by the Megaruptor shear conditions if long Sage Range+T elution times are used (i.e., 30 minutes or longer). Alternatively, Sage Range+T programming can provide significantly narrower size distributions by shortening the Range+T elution time.

Acknowledgments:

We thank Jayson Talag and Dario Copetti of the Arizona Genomics Institute for help in generating Megaruptor-sheared DNA samples.

References

1) Gas, B., Kenndler, E. 2002. Electrophoresis 23: 3817-3826.
2) Gas, B., Stedry, M., Kenndler. 1997. Electrophoresis 18: 2123-2133.
3) Poppe, H. 1992. Analytical Chemistry 64:1908-1919.