diagenode
Innovating Epigenetics Solutions

PACIFIC BIOSCIENCES®

APPLICATION NOTE

Optimal long DNA fragment generation prior to SMRTbell® library preparation

Several methods exist today for shearing genomic DNA, but few are reliable for generating DNA fragments >50 kb necessary for assembly of many microbial, human, plant and animal genomes *de novo*. Pacific Biosciences highly recommends the Megaruptor™ System for shearing genomic DNA to large fragments for long-read sequencing in the Sequel® System.

The Megaruptor uses mechanical shearing to fragment DNA. DNA in solution is pumped through the device's "hydropore" which has an array of uniform pores. The resulting turbulent flow stretches and breaks the DNA strands. The length of the resulting fragments is dependent on the fluid flow rate determined by the system's software and the size of the pores. Passage of the DNA molecules several times through the pores ensures that they will reach a minimum and uniform length as compared to a single pass. Additionally, the hydropore design ensures that DNA will not clog the pores, resulting in consistent and efficient shearing not realized by single orifice devices.

To demonstrate performance of the Megaruptor 2, a high molecular weight (>100 kb) human genomic DNA (NA12878) was diluted to concentration of 50 ng/µl, divided into seven 50 µl aliquots and sheared using the instrument's software settings of 10, 20, 30, 40, 50, 60 and 75 to target large DNA molecules. Sheared samples were recovered and analyzed using AATI FEMTO Pulse to determine fragment distribution (Figure 1).

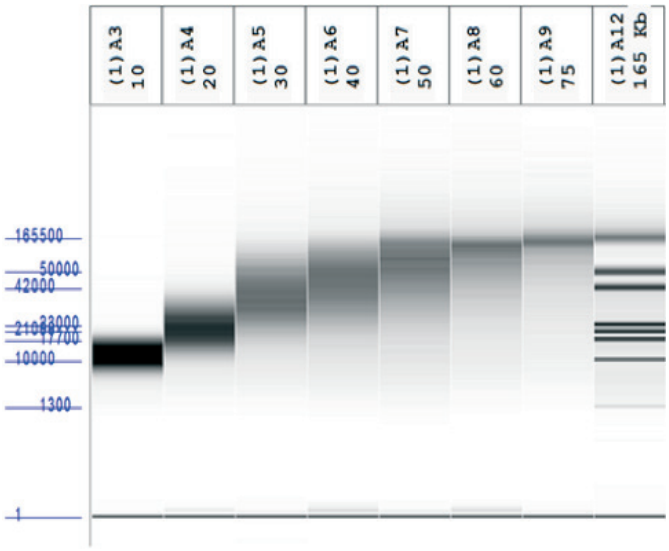

Figure 1: High molecular weight Human DNA (>100 kb) sheared using various shear settings and analyzed using the FEMTO Pulse. Higher flowrate (setting = 10) decreases fragment size while lower flowrate (setting = 75) increases fragment size.

High molecular weight human DNA >100 kb) responds well to varying settings for generating fragments from 10 kb to >70 kb. Its consistency and reliability is beneficial to laboratories performing *de novo* sequencing of different organisms using the Sequel System. In this example, software settings of 10 and 20 resulted in 10–30 kb fragments that is ideal for microbial genome assembly, while settings 30-75 yielded DNA fragments >30 kb,

www.diagenode.com Please contact us for more information
Europe Diagenode sa / LIEGE SCIENCE PARK // Rue du Bois Saint-Jean, 3 // 4102 Seraing (Ougrée) // Belgium // Phone: (+32) 4 364 20 50 // E-mail: info@diagenode.com
USA Diagenode Inc. / 400 Morris Avenue, Suite 101 // Denville, NJ 07834 // USA // Phone: (+1) 862 209-4680 // Mail: info.na@diagenode.com

ideal for whole genome assemblies of human, plants and animals.

To evaluate sequencing performance, sheared DNA from the 60-shear setting was constructed to a SMRTbell® library, size-selected using 30 kb cutoff in the Sage Science's BluePippin System and subsequently sequenced in the Sequel System using the Sequel Sequencing Kit v2.0. The average size of the final library was determined to be approximately 55 kb (Figure 2).

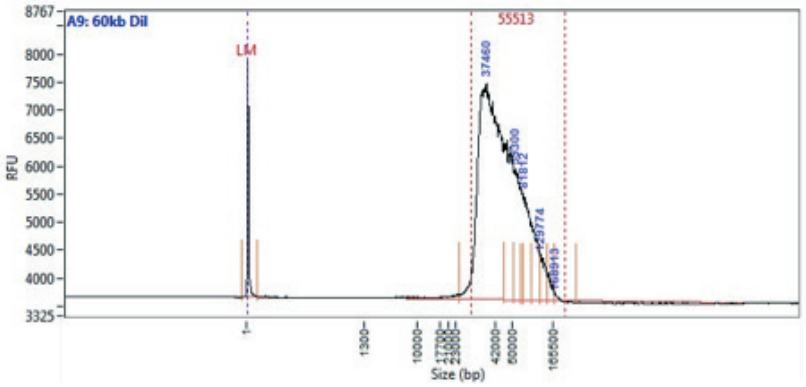

Figure 2: Smear analysis shows a size-selected SMRTbell library (30 kb cutoff) of approximately 55 kb.

Data from one SMRT® Cell 1M was mapped to Human reference hg19 using SMRT Analysis v4.0 and results are shown in Figure 3 and Table 1. The histograms represent mapped polymerase and subread length distributions with average read lengths of 15.4 kb and 13.6 kb, respectively. Table 1 summarizes key read length statistics including mapped N50 subread length (50% of useable bases are in reads greater than 22.6 kb) and maximum subread length of 66.2 kb.

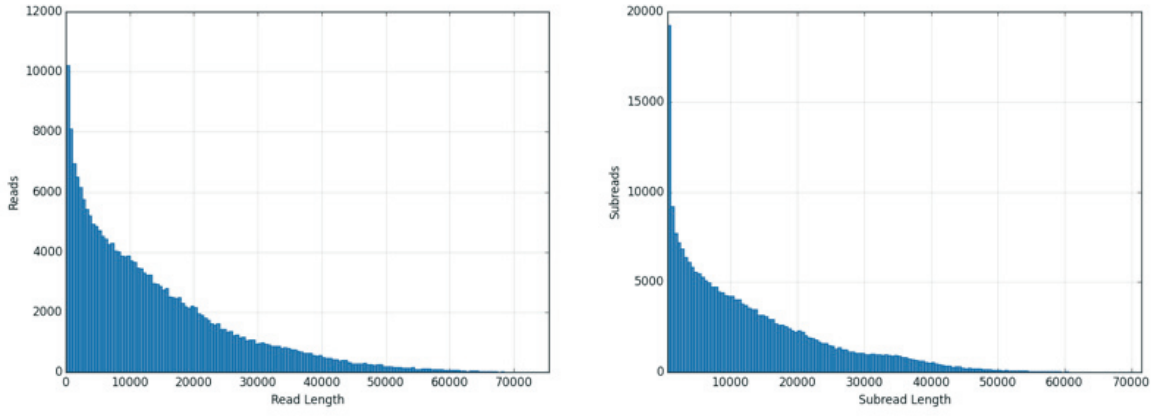

Figure 3: Polymerase and subread length statistics from a 55 kb Human SMRTbell library sequenced in the Sequel System using Sequel Chemistry v2.0 and analyzed using SMRT Analysis v4.0.

	Metrics
Number of Subreads (mapped)	306,455
Number of Subread Bases (mapped)	4,178,458,526
Subread Length Mean (mapped)	13,635
Subread Length N50 (mapped)	22,697
Subread Length 95% (mapped)	37,200
Subread Length Max (mapped)	66,147
Number of Polymerase Reads (mapped)	275,917
Polymerase Read Length Mean (mapped)	15,396
Polymerase Read N50 (mapped)	24,653
Polymerase Read Length 95% (mapped)	41,510
Polymerase Read Length Max (mapped)	72,537

***Table 1**: Read length statistics from a 55 kb human SMRTbell library mapped against human reference hg19. The data represented here is from one SMRT Cell 1M.*

The above performance can be easily achieved depending on the quality of the starting genomic DNA that is sheared to large fragment length. The Sequel System, combined with the Megaruptor system, delivers exceptional performance for generating long reads essential for comprehensive *de novo* assemblies.