NIPPON Genetics Co.,Ltd.

Application Note

Product Name : BluePippin (BLU0001)
Manufacturer : Sage Science
Application : Long-fragment library preparation for PacBio RSⅡ・P5-C3 chemistry

The following data were provided by the courtesy of our customer at Okinawa Institute of Advanced Sciences, Japan.

Background

P5-C3 chemistry substantially enhances activity retention of enzymes used in the sequencing reaction performed in PacBio RSⅡ sequencer (Pacific Biosciences), extending the read (polymerase read) length to an average of 8.5 kb, with the longest being 30 kb. However, the overall read lengths of inserts (subreads) become shorter when short libraries are present at a high frequency. In order to take advantage of P5-C3's ability to generate extremely long reads, it is essential to perform size selection for increasing the frequency of long libraries.
Lately, PacBio RSⅡ users are recommended to perform BluePippin size selection, and the combination of P5-C3 and BluePippin is offered as the standard protocol. In this study, we evaluated the effect of BluePippin size selection.

Experimental method

● Sample DNA

Origin : bacteria A
Purification method : Column purification using a kit

●Conditions for genomic DNA fragmentation

g-Tube (Covaris)

●Conditions for BluePippin size selection

Sample load : ≧ 4 kb → 15 µl / Lane (1.4 µg / Lane)
≧ 7 kb → 30 µl / Lane (2.7 µg / Lane)
Gel cassette : 0.75% DF Marker S1
Extraction condition : high-pass 4-10 kb vs2 (recover ≧4kb and ≧7kb fragments)

PacBio RSⅡ
(Pacific Biosciences)

●Conditions for Pippin Pulse electrophoresis

Electrophoresis buffer : 0.5 x TBE Buffer
Agarose : 0.75% Agarose S
Electrophoresis time : 15hr

●Comparison and evaluation methods

Library size : Fragment Analyzer (Advanced Analytical)
Read length : 20kb Library / 180 min. / 4 cell / sequenced with P5-C3

<Workflow>

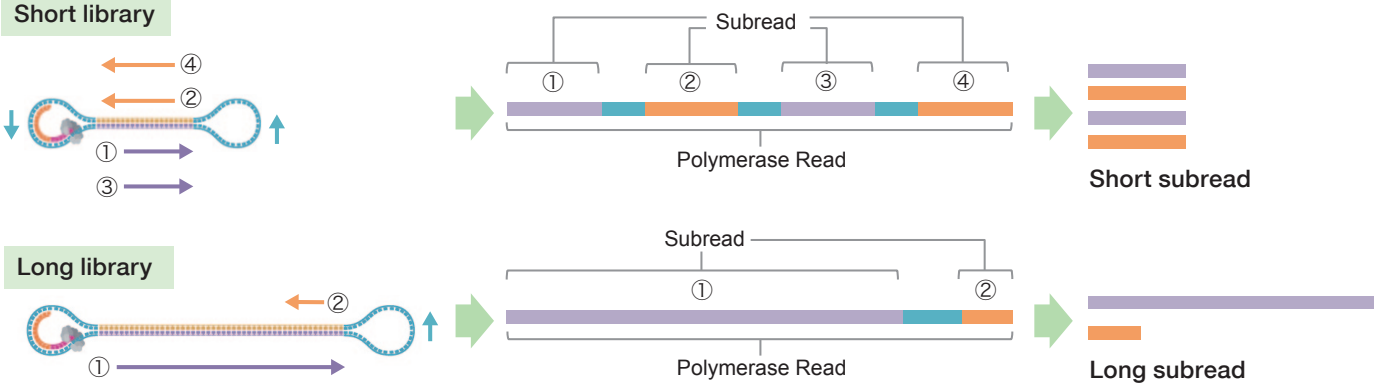
Genomic DNA extraction
↓
Fragmentation in g-Tube
↓
Library preparation
↓
BluePippin size selection, recovery of ≧7 kb fragments
↓
Purification, sequencing

BluePippin
Capable of performing pulsed-field electrophoresis.
Optimum for size selection of long size DNA fragments.

Results

BluePippin size selection

Short library



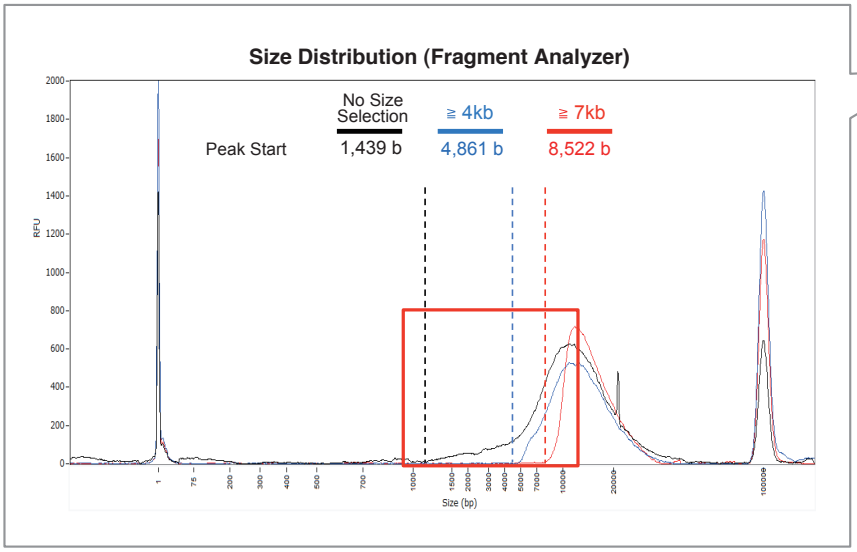
- Short library : inserts are read back and forth, yielding long polymerase reads but short subreads
- Long library : inserts are read back and forth at lower chances, yielding long subreads although the lengths of polymerase reads as a whole do not differ from that obtained from a short library

* In some cases, short libraries (around 2 kb) are intentionally prepared to obtain circular consensus sequencing (CCS) reads

It is essential to remove short libraries and increase the frequency of long libraries

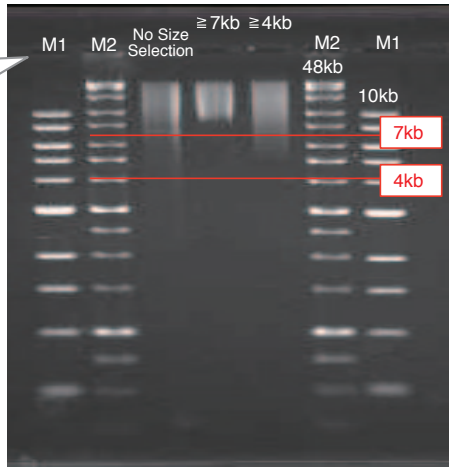
BluePippin

＜Results of size selection＞



Comparison of size distribution by Fragment Analyzer
In "No Size Selection", short libraries are present.
In "≧4kb", short libraries have been removed.
In "≧7kb", short libraries have been further removed.

PippinPulse



Condition：0.5 x TBE Buffer、0.75% Agarose S 、15hr
M1：Quick load 1 kb ladder
M2：M29 ladder

Results of electrophoresis
In "No Size Selection" and "≧4kb", more short libraries are detected compared to "≧7kb".
In "≧7kb", short libraries have been successfully removed.

Conclusion
Short libraries can be removed by BluePippin size selection

Results

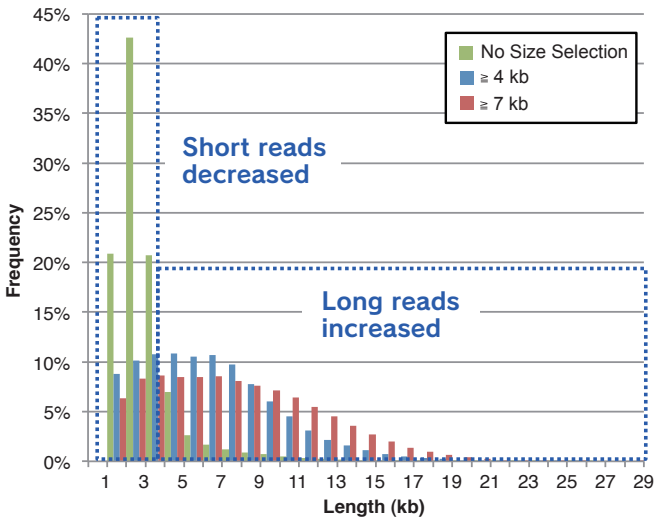
Effect of BluePippin size selection

Bacteria A genomic DNA
- Genome size: 2 Mb (GC 70%)
- 20kb Library / 180min / 4 cell / P5-C3

● Subreads from sequencing using P5-C3

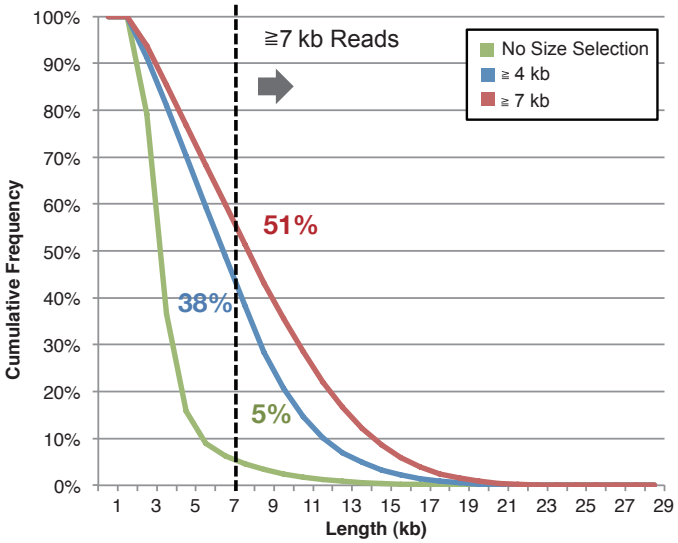
The average lengths in "≧4kb" and "≧7kb" were 2.6 and 3.3 times longer, respectively, compared to that in "No Size Selection".

Subreads	No Size Selection	≧4kb	≧7kb
# cells	4	4	4
# Reads	283,133	72,605	60,510
Total Bases (b)	988,682,713	386,942,346	403,633,307
Max Length (b)	25,888	28,574	27,851
Mean Length (b)	2,060	2.6-fold → 5,329	3.3-fold → 6,671
All Bases (b) /10,000reads	34,919,374	1.5-fold → 53,294,173	1.9-fold → 66,705,223
Coverage /10,000reads	17	25	32



● Frequency of reads

The frequency of short reads in "No Size Selection" (green) decreased after size selection
Consequently, the frequency of long reads increased in the size-selected "≧4kb" (red) and "≧7kb" (blue).



● Cumulative frequency diagram

The cumulative frequencies of ≧7kb reads were 5%, 38% and 51% in "No Size Selection", "≧4kb" and "≧7kb", respectively.

Conclusion

- The frequency of long reads and the average read length were increased by size selection.
- The amount of data per cell was increased 1.5-fold by size selection.

<Customer's comments>

BluePippin size selection could remove short libraries, yielding 2.6- and 3.3-fold increases in average read lengths (subreads) obtained from libraries size-selected for ≧4kb and ≧7kb, respectively, compared to that obtained from a library without size selection.
The frequencies of ≧7kb reads obtained from a library without size selection, with size selection for ≧4kb and with size selection for ≧7kb were 5%, 38% and 51%, respectively, demonstrating an increased frequency of long reads.
The amount of data per cell increased 1.5- and 1.9-folds in libraries size-selected for ≧4kb and ≧7kb, respectively, compared to that in a library without size selection.
These results demonstrate that BluePippin size selection, which is included in the latest standard protocol for PacBio RSⅡ, is actually very effective and is essential for obtaining long reads.

Nippon Genetics Co.,Ltd
http://www.n-genetics.com
TEL 03(3813)0961 FAX 03(3813)0962 info@genetics-n.co.jp