Taking Advantage of Long Read Lengths with Improved Library Preparation Methods

Joan Wilson, Kevin Travers, Kristi Spittle, Paul Peluso, David Rank
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Introduction

The recently released XL polymerase has read lengths that average ~5000 nt on the PacBio® *RS*. To take maximum advantage of these longer reads, new protocols are under development for the generation of 20 kb libraries. When combined with size selection to remove the shorter fragments that tend to load more favorably than long fragments, these libraries show greatly increased subread lengths.

Library Prep Workflow

Check integrity of gDNA on a field-inversion gel

↓

Shear gDNA to ~ 20 kb

↓

Prepare > 10 kb SMRTbell™ library

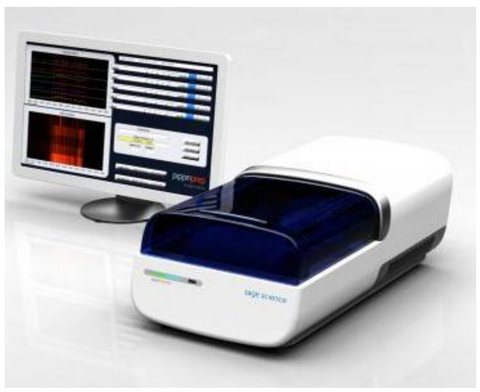
↓

Size-select using Blue Pippin™ System

↓

Concentrate by 0.45x AMPure® purification

Blue Pippin™ Size Selection



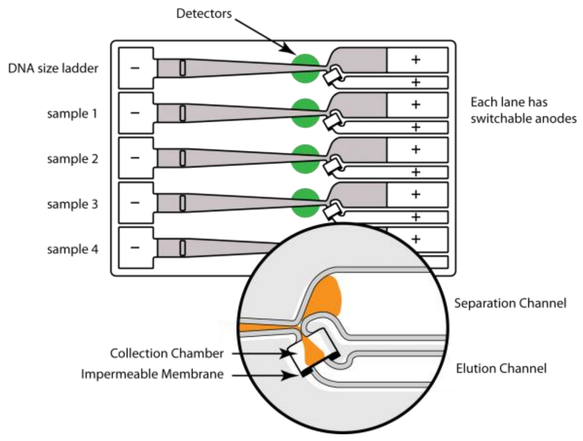


Figure 1. The Blue Pippin system provides preparative-scale separation and extraction of nucleic acids. It can be applied to the separation of large fragments of DNA, allowing the user to separate small DNA fragments from large. Each lane in the gel cassette is split halfway down the cassette. During electrophoresis, the fragments of interest are sent to the elution chamber (inset in lower image) by switching to the electrodes in the elution channel.

Size selection of large fragments

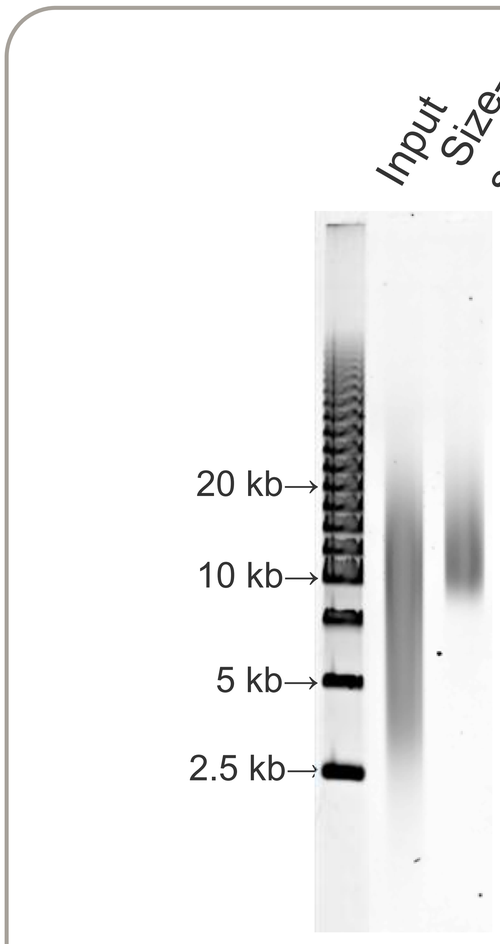


Figure 2. An example of size separation of larger DNA fragments. This image is from field inversion gel electrophoresis of both input DNA and size selected product. In this example, DNA was selected from ~7kb to 50kb.

References

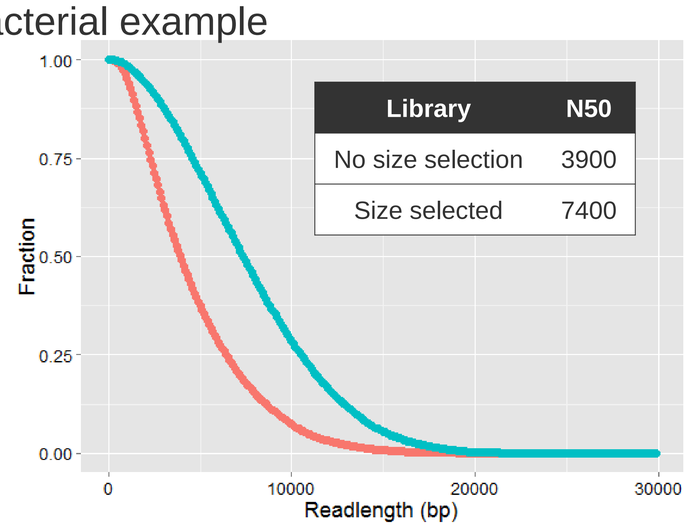
The protocol for preparation of >10kb SMRTbell libraries is available at:

http://www.smrtcommunity.com/Share/Protocol?id=a1q70000000HBDJAA4&strRecordTypeName=Protocol

For more information about the Blue Pippin™ system, see http://www.sagescience.com/

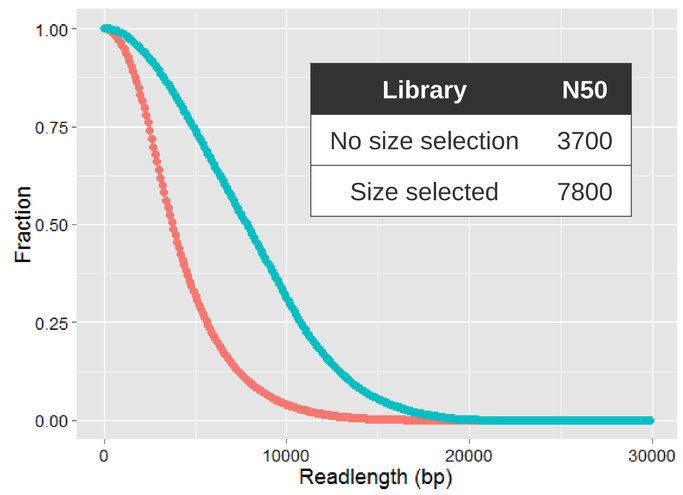
Size selection increases the number of long subreads

Bacterial example



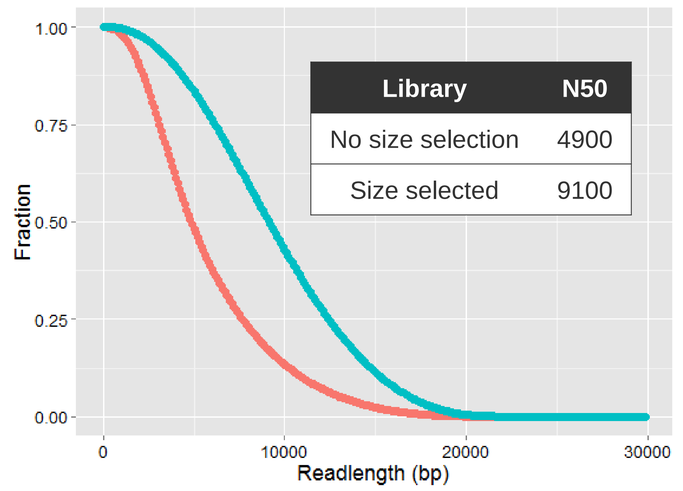
Library	N50
No size selection	3900
Size selected	7400

Fish example



Library	N50
No size selection	3700
Size selected	7800

Mammalian example



Library	N50
No size selection	4900
Size selected	9100

No size selection With size selection

Figure 3. Three examples illustrating the gains achieved with size selection. These plots represent the fraction of bases in reads longer than the value on the x-axis. We report the "N50" value – the readlength at which 50% of the bases are in reads longer than this value.

In all three cases, libraries of ~20 kb average size were prepared as described in the workflow above. Aliquots were sequenced both with and without size selection. In all cases, the size selection resulted in an approximately 2-fold increase in the number of bases in long reads.