Targeted Long Range Sequencing using AccuStart Long Range SuperMix

Subrata Panja PhD, Quantabio

Keywords: Long Range PCR, AccuStart Long Range SuperMix, SMRTbell® library, SMRT sequencing, Long Range Sequencing

ABSTRACT

Targeted genomic sequencing has proven to be very fast and cost effective in determining genetic variations for smaller gene panels and as a confirmatory method for genomic data generated by orthogonal methods. Recent advancements in Single Molecule, Real Time (SMRT) Sequencing and Sequel chemistry has enabled an easy and economical way to sequence multiple kilobases of the genome in a single read with high accuracy (~99%). Targeted sequencing with longer reads often depends on high quality long range PCR products. In this application note, we describe how the newly developed AccuStart Long Range SuperMix can revolutionize the generation of longer PCR products (1 to 24 kb) for longer, better and more reliable SMRT sequencing.

INTRODUCTION

Advancements in next-generation sequencing (NGS) has revolutionized the field of medical genetics. In spite of monumental success and several advantages, traditional NGS with short reads (150 bp) is limited in the ability to identify larger structural variants, mapping sequencing data into highly homologous and repetitive genomic regions, and phasing of alleles. These limitations significantly impact the diagnosis of several genetic disorders.

Sequencing with long reads (~5 to 20 kb) can overcome some of these challenges posed by short read sequencing. Several long range sequencing technologies are now available. Single Molecule Real Time (SMRT) sequencing by Pacific Biosciences and nanopore sequencing by Oxford Nanopore Technologies are at the forefront of the long range sequencing market. Long range sequencing can be used to sequence whole genomes, RNA, targeted regions of the genome and to find epigenetic signatures. In this application note we focus on long range PCR for use on PacBio® sequencing.

Targeted long range sequencing often starts with amplifying a large segment (several kilobases) of the target genome. Multiple amplicons of the same region from different samples undergo individual SMRTbell library preparation and then pooled together for SMRT sequencing (Figure 1).

Alternatively, amplicons of different regions of the same sample can be sequenced using the same procedure. The quality of the long range sequencing data vastly depends on the quality of the PCR amplicons. Non-specific PCR products or incomplete amplifications often occupies sequencing space and produce larger fractions of unusable data. This in turn impacts the quality of variant calling by increasing noise in the sequencing data; simultaneously increasing the overall sequencing cost. This problem can be addressed by using a high quality long range polymerase or master mix. AccuStart Long Range PCR SuperMix (4x concentration) from Quantabio is formulated to enable high quality, robust, reliable and reproducible long range PCR. The SuperMix can reliably amplify up to 24 kb with fast extension speeds (30–60 sec/kb). Also, the fidelity of this polymerase is 10 times higher than the Taq DNA polymerase and can be used for multiplex PCR for up to 6 targets up to 6 kb each.

For this work we collaborated with a large CAP-CLIA genetic testing company to amplify target genes and sequence the PCR amplicons using the PacBio Sequel system.

MATERIALS AND METHODS

Long Range PCR

NA12878 (Coriell Institute) DNA sample was amplified targeting several genes using proprietary primers and AccuStart Long Range SuperMix (Quantabio). The size of the PCR products for the target genes, CYP21A2, BRCA1, SMN1, GBA and mitogenome were 6.2 kb, 10 kb, 10.5 kb, 10 kb and 16.5 kb respectively. For all the PCR reactions, forward and reverse primers were added to the reaction at a final concentration of 0.33 µM (0.66 µM for mitogenome). For each reaction, 100 ng NA12878 DNA was amplified for 37 cycles (28 cycles for mitogenome) using Quantabio's PCR setup recommendation (Table 1 and 2).

Figure 1 **Schematic representation of targeted SMRT sequencing.** PacBio libraries were prepared by ligating barcoded SMRTBell adapters with the amplicons from the target genes. SMRTBell libraries from different samples were pooled together for sequencing (left side of the scheme). Alternatively, target genes can be amplified with primers containing Sequel barcode sequences. Multiple barcoded amplicons were polled together and SMRTBell libraries were prepared by an overhang adapter ligation using SMRTBell Express template Kit 2.0 (right side of the scheme).

Table 1 Guidelines for long range PCR set-up

Components	Volume for 1 Reaction (µl)	Final concentration
Nuclease-Free Water	To a final 50 µl reaction volume	NA
AccuStart Long Range SuperMix	12.5	1x
10 µM Primer Mix	1 to 4	0.2 to 0.8 µM
Template DNA	5 to 10	10 to 100 ng

Table 2 Guidelines for long range PCR set-up

Steps	Temperature	Time	Cycles
Initial Activation	95°C	3 min	1
Denaturation	92°C	30 s	25–40
Annealing	Approximately 5°C below Tm of primers	30 s	
Extension	68°C	30–60 s/kb	
Final Extension	68°C	10 min	1
Hold	10°C	infinite	1

Quality of each PCR product was determined by running 1–2 µl of the PCR product in a 1% agarose gel side by side with Quick-Load® Purple 1 kb Plus DNA Ladder (NEB).

NA12878 was also amplified targeting the mitogenome using PCR primers containing PacBio barcode sequences following the same procedure.

Library Preparation and Sequencing

PCR products were purified using 0.45x sparQ PureMag Beads (Quantabio) following standard protocol and were eluted in 15 µl of 10 mM Tris-HCl (pH 8.0) buffer. Purified PCR products were quantified using Qubit™ 1x dsDNA HS Assay Kit (Invitrogen, ThermoFisher Scientific). A total amount of 1.5 µg purified PCR products were used for the library preparation. DNA SMRTbell libraries were generated by ligating barcoded SMRTbell adapters (Sequel 384 barcodes v1) with the PCR products using Pacific Biosciences (PacBio) SMRTbell Barcoded Adapter Complete Prep Kit-96 (Part Number 100-465-800) and SMRTbell Damage Repair Kit (Part Number 100-465-900) following PacBio's published protocol.

Mitogenome PCR products that were amplified using barcoded primers were purified as described above and pooled together to achieve a total DNA amount of 1500 ng. SMRTbell libraries were prepared using PacBio's SMRTbell Express Template Prep Kit 2.0 (Part Number 101-685-400) following PacBio's published protocol.

All the libraries were sequenced on the Sequel (PacBio) instrument at the CCS (Circular Consensus Sequence) mode using Sequel Sequencing Kit 3.0 (4 rxn) (PacBio) and SMRT Cell 1M v3 LR (PacBio) flow cells. The libraries were loaded at a concentration of 4 to 8 pM and SMRT data were collected over 20 hours for individual flow cells. Data was analyzed using SMRTLink (V 8.0) and the BAM files were visualized using IGV (Broad Institute).

RESULTS

Quality of the PCR products

As stated above, in order to generate high quality SMRT sequencing data the quality of the PCR products needs to be high. PCR products of the target genes CYP21A2, BRCA1, SMN1 and GBA, produced extremely clean bands and expected migration patterns in a 1% agarose gel (Figure 2). Most importantly, there were no traces of smaller or non-specific PCR products. From a 30 µl PCR reaction, we obtained 2 to 4 µg of DNA after purification.

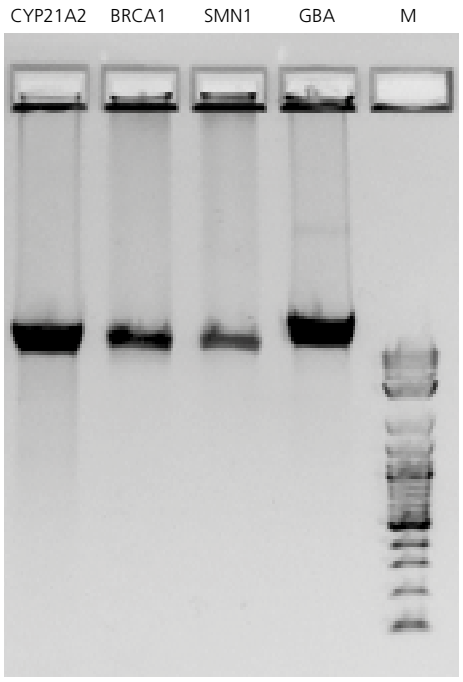

Figure 2 Long range PCR amplicons of different target genes were visualized in a 1% agarose gel.

Targeted SMRT sequencing of GBA

Targeted GBA templates were sequenced in Sequel in the CCS mode. CCS reads were then mapped into the hg19 reference genome. Mapped reads generated with the SMRT sequencing method fully covered the GBA regions of interest (10 kb) with 99% of the total reads mapped into the target region. For this experiment, 16 samples were sequenced for 20 hours at a loading concentration of 8 pM, resulting in >5000x coverage across 10 kb target regions of GBA (Figure 3).

All the homozygous and heterozygous SNPs are clearly visible from the data (Figure 3), which allowed excellent phasing performance. Another DNA sample from the Coriell Institute was used for the same experiment, clearly demonstrating that the sample is homozygous for a recombinant allele for GBA (data not shown).

Figure 3 IGV view of the targeted SMRT sequencing reads provide coverage for the entire 10 kb GBA gene.

Different methods of library preparation vs sequencing data quality

Human mitogenome (NA12878) was amplified with custom designed primers and with the same primers containing Sequel barcode sequences. PCR efficiency and the amplicon quality between two PCR methods were similar as evident from 1% agarose gel electrophoresis (Figure 4 A). As mentioned in the materials and method section, amplicons generated with primers containing adapter sequences went through a library preparation using PacBio's SMRTbell Express Template Prep Kit 2.0. 99% of the CCS data generated using both the methods were mapped into the target region. A side by side comparison of the bam files shows that all the homozygous SNPs were clearly visible in both cases and the variant calls were identical between the two methods (Figure 4 B). Although, the same number of samples were sequenced in Sequel at the same loading concentration, the second library prep method yielded higher amounts of data and the coverage across the mitogenome was 7-fold higher (3500x coverage in comparison to 500x).

Comparison between different polymerases

As quality of the PCR product is pivotal to achieve higher quality of SMRT sequencing data, we compared long range PCR efficiency of the GBA gene using different leading polymerases available in the market. Out of three polymerases tested, only one (Pol T) performed similar to AccuStart Long Range SuperMix (Figure 5). Performance of the other two polymerases were poor, resulting in lower yield. In the case of pol N, a second band is visible in the gel (Figure 5).

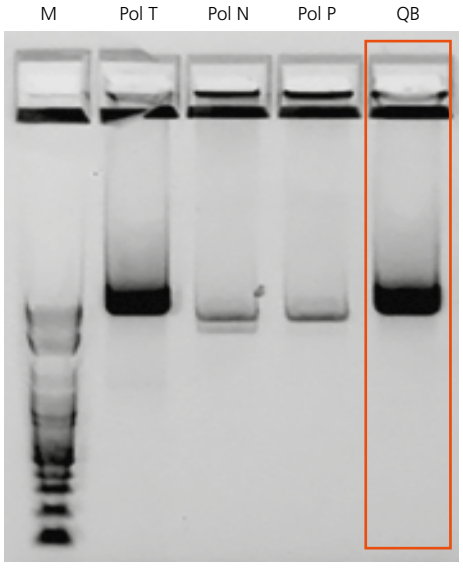

Figure 5 Amplification of GBA gene using different polymerases.

Long range PCR and DNA from different sources

In order to check the efficiency of long range PCR with AccuStart Long Range SuperMix, DNA from different sources (blood, buckle swab, oral rinse etc.) was utilized along with various extraction methods. The same amount of DNA was used to amplify the mitogenome using the protocol mentioned above. Figure 6 A shows that quality of the long range PCR was poor in many cases. To solve this problem, we purified the same set of DNA samples with 1x sparQ PureMag Beads and repeated the long range PCR with similar conditions. Figure 6 B shows that the quality of the long range PCR improves dramatically when bead purified DNA samples were used for the PCR.

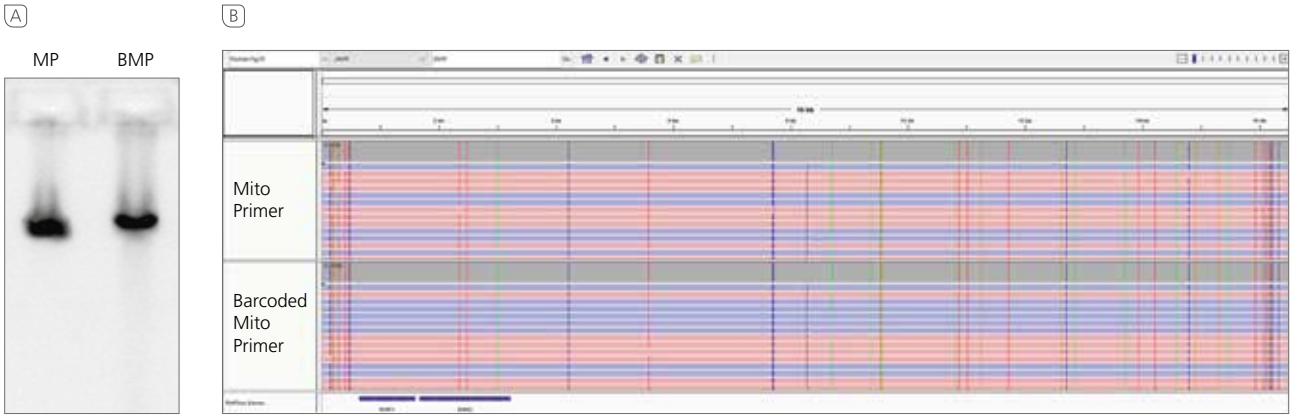

Figure 4 **A** Amplification of the mitogenome with the primers containing Sequel barcodes (BMP: barcoded mito primer) does not affect the quality of the long range PCR amplicon. **B** SMRTbell libraries prepared with SMRTbell Barcoded Adapter Complete Prep Kit-96 (upper panel) and SMRTbell Express Template Prep Kit 2.0 (lower panel) exhibited identical variant calling.

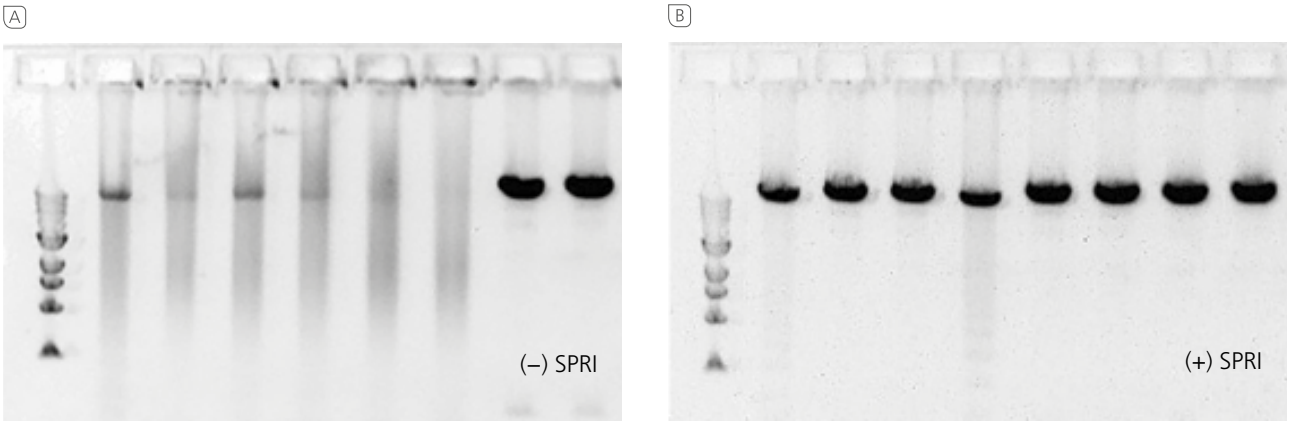

Figure 6 **DNA samples from different sources were amplified with mitogenome specific primers.** 1x SPRI bead cleanup improved the long range PCR performance.

DISCUSSION

In this application note, we highlight how targeted genomic sequencing can be achieved with longer reads by generating high quality long range PCR products using AccuStart Long Range SuperMix and sequencing those amplicons using SMRT technology developed by Pacific Biosciences. Two different PCR and library preparation methods are described where by the user can choose a method depending on their need and feasibility. Although, both the methods generated high quality data with full length coverages (>500 x for 24 samples), the libraries prepared with PacBio's SMRTbell Express Template Prep Kit 2.0 resulted in 6–7 times higher coverage than the libraries prepared with the SMRTbell Barcoded Adapter Complete Prep Kit-96. This may be due to the improved ligation efficiency of the over hanged adapters compared to the blunt end ligation method used in the first library prep method.

AccuStart Long Range SuperMix produced PCR amplicons that resulted in better performance in comparison to alternative long range polymerases. In addition to performance, the SuperMix comes packaged in a single tube, only requiring the addition of DNA template, making it very easy to use. For uniform performance across the sample types and extraction methods, a bead cleanup of the input DNA before PCR is highly recommended.

MK-AN-0003 REV 02 Long Range PCR 0121