PacBio® RS II sequencing chemistries provide read lengths beyond 20 kb with high consensus accuracy. The long read lengths of P4-C2 chemistry and demonstrated consensus accuracy of 99.999% are ideal for applications such as de novo assembly, targeted sequencing and isoform sequencing. The recently launched PS-C3 chemistry generates even longer reads with N50+ often >10,000 bp, making it the best choice for scaffolding and spanning structural rearrangements. With these chemistry advances, PacBio’s read length performance is now primarily determined by the SMRTBbell™ library itself.

Size selection of a high-quality, sheared 20 kb library using the BluePippin® System has been demonstrated to increase the N50 read length by as much as 5 kb with C3 chemistry. BluePippin size selection or a more stringent AMPure® PB selection cutoff can be used to recover longer fragments from degraded genomic material. The selection of chemistries, P4-C2 versus P5-C3, is highly dependent on the final size distribution of the SMRTBbell library and experimental goals. PacBio’s long read lengths also allow for the sequencing of full-length cDNA libraries at single-molecule resolution. However, longer transcripts are difficult to detect due to lower abundance, amplification bias, and preferential loading of smaller SMRTBbell constructs. Without size selection, most sequenced transcripts are 1-1.5 kb. Size selection dramatically increases the number of transcripts >1.5kb, and is essential for >3kb transcripts. * N50=X defined as half of the data in reads with length greater than X

**Introduction**

The new P5 polymerase and C3 Chemistry combined with Long read lengths allow sequencing of transcript isoforms from To capture full diversity of transcripts, additional coverage of ~25X of P5

**Large Genome Scaffolding with P5-C3**

- The new P5 polymerase and C3 Chemistry combined with 3-hr data collection are ideal for generating long reads for gap closing or scaffolding large genomes such as the highly repetitive Maize genome.
- Maximum long read benefits of P5-C3 can be achieved by constructing and sequencing 20 kb SMRTBbell library size-selected using the BluePippin™ system.

**Size Selecting Degraded Samples**

- When input DNA is already fragmented to the desired size or smaller, shearing is not necessary and may further reduce library insert size.
- For these samples, it is very important to remove shorter fragments that will be much less beneficial in assembly. The BluePippin system is the preferred method of size selection, if sample quantity is sufficient.
- Shown below, SMRTBbell libraries from partially degraded samples can be successfully prepared and sequenced in the PacBio RS II to generate long read lengths.

**Maximizing Long Reads in Iso-Seq Sequencing**

- Long read lengths allow sequencing of transcript isoforms from high-quality poly(A) RNA using PacBio’s Iso-Seq method.
- Full-length, intact transcripts are defined by the detection of both 5’ and 3’ PCR primers.
- To capture full diversity of transcripts, we recommend three size fractions for each cDNA sample: 1-2 kb, 2-3 kb, and 3-6 kb.
- Size selection can be performed by excision from agarose gels (traditional gel cuts), or with the BluePippin system.

**Conclusions**

Long continuous read lengths are essential for applications such as de novo assembly and IsoSeq Sequencing (Iso-Seq).

- The combination of P6 polymerase-C3 chemistry with a high quality BluePippin size-selected library has resulted in N50 subread length > 10.8 kb, enabling gap closure or contig scaffolding for complex genomes such as maize.
- Size selection using the BluePippin system greatly increases insert sequence lengths and assembly results from partially degraded gDNA, even with a low cutoff such as 4 kb.
- With Iso-Seq, size selection of transcripts allows the detection of isoforms up to 6 kb. With no size selection, the average transcript size is generally 1-1.5 kb. (See Poster Pi043)

**Acknowledgements**

The authors would like to thank our collaborators, Mark S. Mikl et al. from University of Illinois at Urbana-Champaign and Ellen Paxinos and Jenny Gu, Pacific Biosciences, for their contributions to this poster.

REFERENCE:
