Full-length cDNA Sequencing of Prokaryotic Transcriptome and Metatranscriptome Samples

Matthew Boitano, Ting Hon, Elizabeth Tseng, and Tyson A. Clark
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

Introduction

Next-generation sequencing has become a useful tool for studying transcriptomes. However, these methods typically rely on sequencing short fragments of cDNA, then attempting to assemble the pieces into full-length transcripts. Here, we describe a method that uses PacBio long reads to sequence full-length cDNAs from individual transcriptomes and metatranscriptome samples.

We have adapted the PacBio Iso-Seq™ protocol for use with prokaryotic samples by incorporating RNA polyadenylation and rRNA-depletion steps. In conjunction with SMRT® Sequencing, which has average readlengths of 10-15 kb, we are able to sequence entire transcripts, including polycistronic RNAs, in a single read.

Here, we show full-length bacterial transcriptomes with the ability to visualize transcription of operons. We also highlight the ability to detect full-length transcription of operons with alternative start and stop sites. In the area of metatranscriptomics, long reads reveal unambiguous gene sequences without the need for post-sequencing transcript assembly.

Sample Preparation Methods

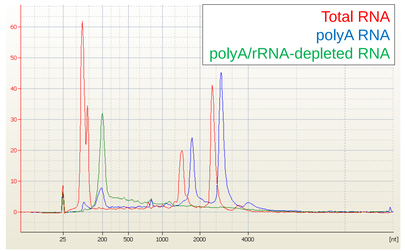

Figure 1. Bioanalyzer traces of *E.coli* total RNA, polyadentylated RNA and polyadentylated/rRNA-depleted RNA. PolyA-tail reaction has been optimized in order to add ~200 nucleotides. Polyadentylated/rRNA-depleted RNA shows good reduction in rRNA peaks and is the input for the cDNA synthesis reaction.

Sample Preparation Methods

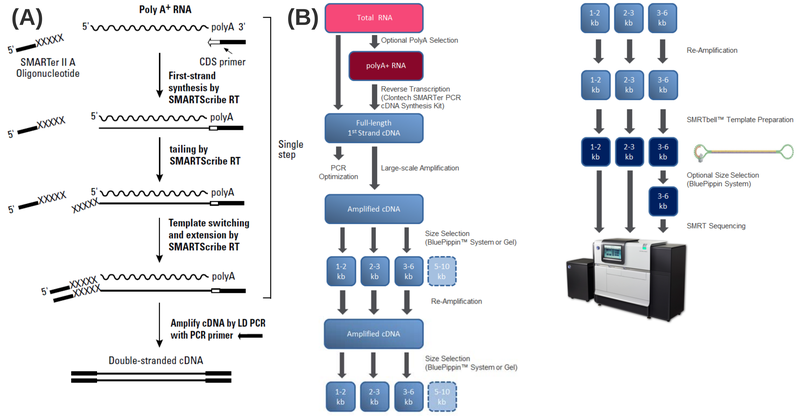

Figure 2. (A) Clontech® SMARTer ® PCR cDNA Synthesis Kit is used to generate double-stranded cDNA. (B) Double-stranded cDNA is then size-fractionated using the Sage BluePippin™ system to sizes of 1-2, 2-3, 3-6 and 5-10kb (if material is available at each size). This size-fractionated material is then used to make SMRTbell libraries. Alternatively, non-size selected material can also be used to generate SMRTbell libraries.

Affects of rRNA Depletion

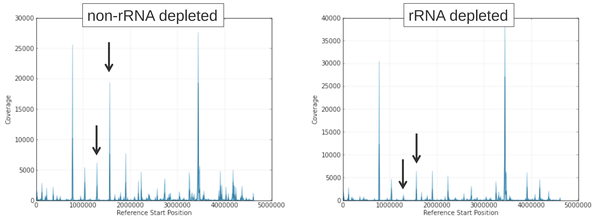

Figure 3. Sequence reads were mapped to the *E.coli* genome. Arrows show reduction in coverage of rRNAs when rRNA have been depleted. Shared peaks are most likely ribosomal associated genes.

Detection of poly-Cistronic and Full-Length Operon Transcripts

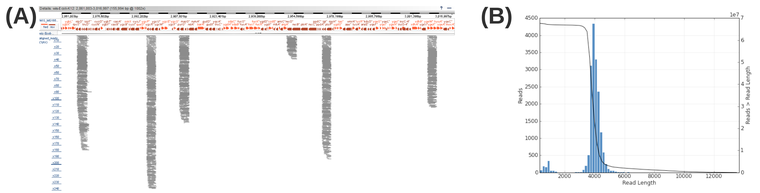

Figure 4. (A) Using long-read SMRT® Sequencing, poly-cistronic and full-length operon reads are easily obtained without the need for assembly of short fragments. Data shown are from 3-6kb size bin, which have an average insert size of 3,917bp (B).

Detection of Alternative Transcription Start/Stop Sites

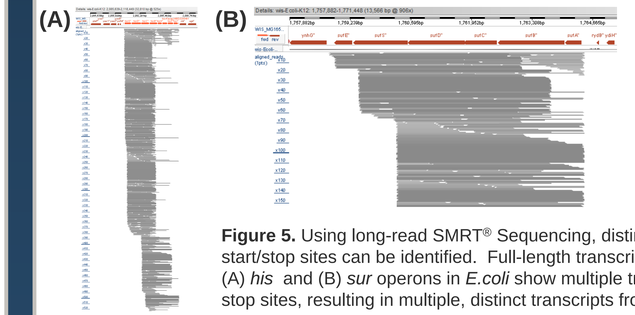

Figure 5. Using long-read SMRT® Sequencing, distinct transcription start/stop sites can be identified. Full-length transcript reads that map to the (A) *his* and (B) *sur* operons in *E.coli* show multiple transcription start and stop sites, resulting in multiple, distinct transcripts from the same operon.

Metatranscriptome Long-Read Sequencing

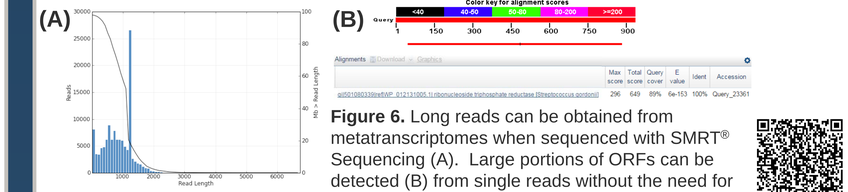

Figure 6. Long reads can be obtained from metatranscriptomes when sequenced with SMRT® Sequencing (A). Large portions of ORFs can be detected (B) from single reads without the need for assembling small fragments.