**Abstract**

PacBio SMRT® Sequencing has the unique ability to directly detect base modifications in addition to the nucleotide sequence of DNA. Because eukaryotes use base modifications to regulate gene expression, the absence or presence of epigenetic events relative to the location of genes is critical to elucidate the function of the modification. Therefore an integrated approach that combines multiple omic-scale assays is necessary to study complex organisms. Here, we present an integrated analysis of three sequencing experiments: 1) DNA sequencing, 2) base-modification detection, and 3) Iso-seq™ analysis. In Neurospora crassa, a filamentous fungus that has been used to make many landmark discoveries in biochemistry and genetics. We show that de novo assembly of a new strain yields complete assemblies of entire chromosomes, and additionally contains entire centromeric sequences. Base-modification analyses reveal candidate sites of increased interpulse duration (IPD) ratio, that may signify regions of 5mC, 5hmC, or 6mA base modifications. Iso-seq method provides full-length transcript evidence for comprehensive gene annotation, as well as context to the base-modifications in the newly assembled genome. Projects that integrate multiple genome-wide assays could become common practice for identifying genomic elements and understanding their function in new strains and organisms.

** Genome: de novo Assembly**

The *Neurospora crassa* genome:
- Seven chromosomes
- Approximately 40 Mb

We sequenced a new strain (T1) of *N. crassa*, which is an A mating type strain derived by DG Catchside from a cross between the Em a 5297 and Em A 5256 strains he obtained from Stirling Emerson in 1955.

**Sequencing statistics**

- Bases: 9 Gb (225 X)
- Bases per cell: 500 Mb
- Mean read length: 7.8 kb
- N50 read length: 10.4 kb
- Chemistry: C2

**Assembly statistics**

- Pre-assembly seed cutoff length: 15.9 kb
- Total assembled bases: 41.7 Mb
- Total contigs: 20
- N50 contig length: 5.3 Mb (limited by chrom. size)

Algorithm: HGAP.1 (Chin et al.)

Software: SMRT Analysis v2.0.1 (free, open-source)

We assembled four entire chromosomes into a single contig and the remaining three chromosomes were assembled with only 1-2 gaps remaining. By comparison, each “supercontig” annotated in the reference genome has between 38 – 89 gaps.

**Epigenome: Base Modifications**

PacBio SMRT sequencing has the unique ability to detect base modifications based on increased interpulse duration (IPD) between two base-incorporation events.

We can detect large regions of increased IPD. Above is an example of a 2 kb region on contig 60, where IPD is increased in both the forward and reverse directions.

**Integration**

A well assembled genome is an important backbone for annotating functional elements such as genes and epigenetic marks. The rich genomic context obtained by combining datasets and assays is greater than the sum of the individual parts.

**Transcriptome: Iso-seq™ Analysis**

We isolated and sequenced polyA+ RNA from the same T1 strain. A cDNA library was prepared and separated into three size fractions (1-2 kb, 2-3 kb, 3-6 kb) according to the Iso-seq protocol “Using Clontech cDNA Prep and BluePippin™ Size Selection”. Transcripts were obtained by applying the bioinformatics pipeline for consensus clustering (https://github.com/PacificBiosciences/cDNA_primer).

**Consensus clustering statistics**

- Identified 3,890 full-length cDNAs
- Aligned 68,244,024 bases
- 99.94% accuracy (0.057% error)
- 11,932 (0.017%) insertions
- 15,502 (0.022%) deletions
- 11,482 (0.016%) mismatches

**Transcript analysis statistics**

- 7,626 unique transcripts
- Found 4823 named/known transcripts
- Validated 2829 hypothetical transcripts that had no prior mRNA evidence
- Discovered 144 new transcripts

We also found many examples of antisense transcription. Here, a new transcript (PB.500.1) is antisense to gene NCU04664T.

**References**

- SMRT Analysis v 2.1.1 Software suite is free and open-source, and can be downloaded at http://pacbio.com.
- Iso-Seq™ sample preparation protocol can be downloaded at http://www.smrtcommunity.com/Share/Protocol?id=17000000979

---


1 Joint first authors, 2 Flinders University, Adelaide, Australia
2 Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025