Lawrence Berkeley National Laboratory
Recent Work

Title
Gap Closure with Roche/ 454 Reads

Permalink
https://escholarship.org/uc/item/9kh2h51f

Authors
Cheng, Jan-Fang
Tice, Hope N.
Malfatti, Stephanie
et al.

Publication Date
2008-09-08
# Abstract

As the Sanger sequencing in the de novo assemblies is being replaced by the next generation sequencing reads, finishing of remaining gaps in the genomes will face two major challenges. First, the technology will need to be fast enough to handle many more drafted genomes. Second, it will have to be a clone-free approach. We have been testing a method that utilizes a universal “bubble-tag” to perform extended amplification from contig ends and gap closure in a clone-free condition. The “bubble-tag” method was first described by Doug Smith (PCR Methods Appl. 2, 1992) to sequence lambda DNA. The advantage of this approach is that gap closure can be performed from all contig ends without the prior knowledge of contigs’ order and orientation. It has not been demonstrated, however, that this approach would work for the complex genomes. Here we describe the experimentation of this approach in closing gaps of the draft Kedonobacter racemifer genome. Genomic DNA was sheared and ligated to the bubble adapters. Primers derived from 96 contig ends were used with the universal bubble primer to amplify the gaps. We applied the AMPure beads to reduce the amount of small fragments. The remaining large amplified DNA appears to be suitable for sequencing. Different bead-to-DNA ratios were tested in order to generate long amplified templates. This approach enables the finishing of complex genomes in a clone-free process with the new sequencing platforms. More importantly, the uniformity of this approach is amenable for a massive, parallel finishing operation.

# Does it Work with Roche/454 Reads?

## Test Case - Gap Closure of Kedonobacter racemifer Genome

![Diagram of draft assembly contigs and gap closure](image)

Ninety six primers were selected from the draft genome for the last of primer walk with the bubble-PCR approach. Many of them have multiple matches to the genome, and not all primers were derived form the ends of contigs. Two-up of gDNA was sheared to 8Kb, end-polished, and ligated to a bubble adapter. About 0.9 ng of the ligated DNA was used in the each of the 96 PCR reactions. After AMPure bead cleaning, 2 ul of the PCR products were loaded on the gel (A), and the excluded DNA were also loaded in gel (B) for comparison. It appears that DNA fragments less than 500 bp were effectively removed by this process. We pooled 24 ul of the eluted product from each amplification and constructed a 454 Titanium library.

### Generation, Assembly, and Gap Filling of 454 reads

- **Sequence Depth Distribution in the Gap Filling Contig**
  - **A.** A brief overview of the experimental processes.
  - **B.** About 50% of the reads are shorter than 300bp. Based on the distribution of the sequence depth within the resulted contigs, we believe that the extremely short read length is caused by the nature of this library.
  - **C.** About 16% of the reads were trimmed by the bubble-tag sequence. In the cases of low sequence depth (5k to 40k reads), about 30% of the reads were not assembled into contigs. This percentage increases with an increasing amount of sequence coverage. The reads that were not assembled included short reads, repeats, too short, etc.
  - **D.** We are particularly interested in the gap closing or extension results generated by different amount of sequence coverage. Although the number of major contigs (200bp) increases with a higher sequence depth, the number of contigs that extend into the gaps reaches the highest number at 20k reads.
  - **E.** At 20k reads, 30 gaps were closed and 17 gaps gain new sequences.

## Alignment of the Gap Filling Contigs and the Draft Contigs

![Alignment diagram](image)

Shown here are a distant view (A) and two close-up views (B and C) of the ACT (Artemis Comparison Tool) display of the gap filling contigs and the draft assembly of the K. racemifer genome based on the MUMmer alignment. The contigs that extend or bridge the gaps in the draft genome is located at bottom of each graph.

### Conclusions

1. We have demonstrated that the bubble-tag PCR can be used to generate targeted templates for the finishing process. The sequencing and assembly can be performed by using the 454 platform.
2. This clone-free finishing process can be achieved by pooling a large number of DNA fragments generated from contig ends and poor quality regions. However careful quantification and equal mixing of DNA samples is important to minimize the representation bias.

### Acknowledgements

We would like to thank Linda Meincke and Cliff Han of LANL for providing the draft assembly and primer sequences, Eugene Gottman and Alex Copeland for suggestions of running the assemblies, and Dorothy Lang for running the MUMmer analysis.