# Development of High Throughput Process for Constructing 454 Titanium and Illumina Libraries.

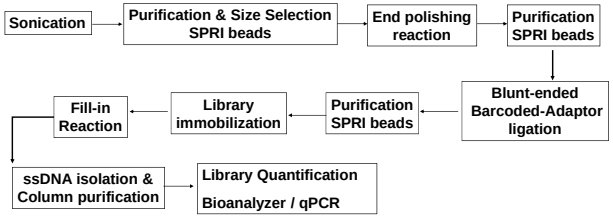
Shweta Deshpande[1], Chris Hack[1], Eric Tang[1], Stephanie Malfatti[2], Aren Ewing[1], Susan Lucas[2], and Jan-Fang Cheng[1]
1Lawrence Berkeley National Laboratory, 2 Lawrence Livermore National Laboratory

## Abstract

We have developed two processes with the Biomek FX robot to construct 454 titanium and Illumina libraries in order to meet the increasing library demands. All modifications in the library construction steps were made to enable the adaptation of the entire processes to work with the 96-well plate format. The key modifications include the shearing of DNA with Covaris E210 and the enzymatic reaction cleaning and fragment size selection with SPRI beads and magnetic plate holders. The construction of 96 Titanium libraries takes about 8 hours from sheared DNA to ssDNA recovery. The processing of 96 Illumina libraries takes less time than that of the Titanium library process. Although both processes still require manual transfer of plates from robot to other work stations such as thermocyclers, these robotic processes represent about 12- to 24-folds increase of library capacity comparing to the manual processes. To enable the sequencing of many libraries in parallel, we have also developed sets of molecular barcodes for both library types. The requirements for the 454 library barcodes include 10 bases, 40-60% GC, no consecutive same base, and no less than 3 bases difference between barcodes. We have used 96 of the resulted 270 barcodes to construct libraries and pool to test the ability of accurately assigning reads to the right samples. When allowing 1 base error occurred in the 10 base barcodes, we could assign 99.6% of the total reads and 100% of them were uniquely assigned. As for the Illumina barcodes, the requirements include 4 bases, balanced GC, and at least 2 bases difference between barcodes. We have begun to assess the ability to assign reads after pooling different number of libraries. We will discuss the progress and the challenges of these scale-up processes.

## Scaling Up of 454 Titanium Std and Illumina Library construction

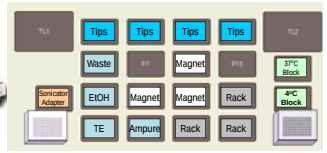
### Robotic Process Of Constructing 454 Titanium Libraries




A method to construct 96 454 shotgun libraries in parallel using a Beckman-Coulter BioMek FX robot to automate the repetitive pipetting steps is in development at the JGI. The goal of this project is to enable a single operator to be able to construct 96 454 shotgun libraries with or without barcodes in a single day with minimal ergonomic risk. All the modifications were made so that the entire process can be adapted to the 96-well plate format. One major change is replacing the gel size selection step with a SPRI bead size selection.

### Beckman-Coulter BioMek FX robot

Deck Layout



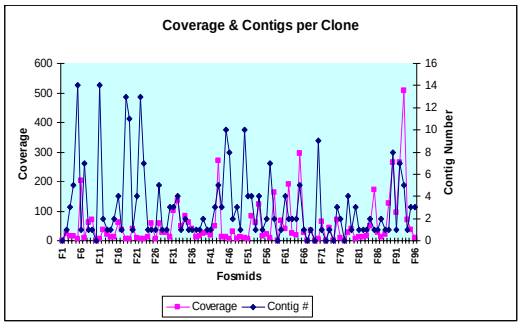

We were able to successfully develop and implement tests on lambda DNA for 454 production samples. The fosmid samples were subsequently pooled and sequenced on a 454 Titanium platform. We aim to convert this process to the parallel construction of barcoded 454 RAPID dsDNA libraries. It is our expectation that this will result in a more robust library construction process with higher yields of dsDNA libraries and greater sequencing efficiency.

## Pooling of Titanium Libraries with Molecular Barcodes

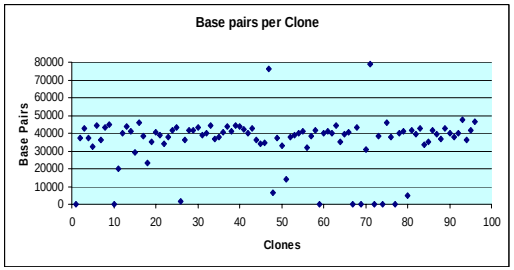
96 fosmid clones were selected from the Antarctic Marine Bacterioplankton Summer pool. Making 96 individual libraries and sequencing them separately was a challenge and would have generated huge amount of data. We designed the molecular barcode sequences for the Titanium libraries based on the requirements mentioned below. We have selected 96 MIDs to be used in creating libraries in a 96-well plate format. To test the ability of pooling projects and Newbler's ability in resolving these MIDs, we constructed 96 fosmid subclone libraries, each containing a unique MID. These libraries were pooled with an equal mass according to the Bioanalyzer reading.

**BARCODE SEQUENCE DESIGN REQUIREMENTS**
- Oligo length: 10 nucleotides (1,048,576 possible sequences)
- No consecutive same bases (78,732 sequences)
- 40-60% GC content (64,472 sequences)
- No more than 2 di-nt or tri-nt repeats (62,072 sequences)
- Adapter sequences differ by at least 4 bases (270 sequences)

## Coverage of 96 fosmid clones

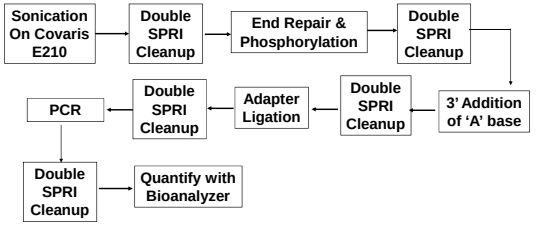



96 clones were barcoded and pooled. After sorting, the clones were assembled in to contigs ranging from 1 to 14 contigs. Of the 96 fosmids, 8 were not assembled due to some assembly problems, 38 were assembled into 1 contig with coverage ranging from 5X to 202X, 13 were assembled in 2 contigs with the coverage ranging from 9X to 191X, and 37 were assembled in 3 to 14 contigs. The average depth for the 88 fosmids is 57 and the standard deviation is 81, suggesting a wide range of sequence depth among these fosmids. There are rooms for improvement in library quantification and pooling.




The assembled contig lenth of the 96 fosmids is shown here. 80 fosmids were assembled to 30 to 48 Kb length. Two fosmids with total contig length of greater than 50 Kb may be resulted from mis-assembly. 14 fosmids received low sequence depth were only partially assembled.

## Constructing Pooled barcoded Illumina Library



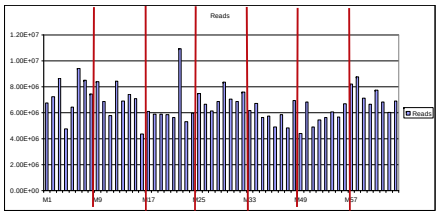

The Illumina library construction process above was constructed to be performed either robotically on the BioMek FX or manually. When constructing a barcoded library a barcoded adapter was used at the adapter ligation step instead of the standard Illumina adapter. Barcoded libraries can be created using the manual or robotic approach as well.

## Pooling of Illumina Libraries (Molecular Barcodes)

The samples selected for barcoding were derived from different recombinant inbred lines Illumina sequencing produces millions of reads per run, and barcoding libraries allows pooling of multiple samples to be run in the same lane. This approach has progressively become important as the total number of reads and raw bases per lane increases.

BARCODE SEQUENCE DESIGN REQUIREMENTS
- Oligo length: 4 nucleotides
- 50% GC content
- Adapter sequences differ by at least 2 bases

## Distribution of reads between 56 RILs




56 RILs libraries were individually barcoded and were pooled in a order of 8 libraries in one pool. Overall 7 pools were created. All 8 libraries were ran on a separate lane. Each pool generated 46 to 61 Million reads. The reads that sorted in to each RILs ranged from a low of 4.3 to 10 Million reads.

## Conclusions

1. We have been constructing Titanium Std libraries on a 96 well format using a modified version of the 454 protocol and have successfully programmed the BioMek robot to automate the construction of 96 libraries simultaneously with minimal ergonomic risk.

2. We have successfully designed 43 unique molecular barcodes that can be used to create quality Illumina libraries. This approach has allowed us to sequence multiple samples in the same run lane. The data presented here indicates that our method produces a relatively even number of reads for each unique barcode used.

3. We are currently working to optimize the efficiency, quality, and reproducibility of the libraries created by the Robot so that we can create Illumina and 454 RL libraries on it.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.