Title
Development of High Throughput Process for Constructing 454 Titanium Libraries

Authors
Deshpande, Shweta
Hack, Chris
Tang, Eric
et al.

Publication Date
2010-03-25

Development of High Throughput Process for Constructing 454 Titanium Libraries.

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

Abstract

We have developed a process with the Biomek FX robot to construct 454 titanium 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. Although this process 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. The requirements for the 454 library barcodes include 10 bases, 40-60% GC, no consecutive same base, and no less than 4 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. 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 this scale-up process.

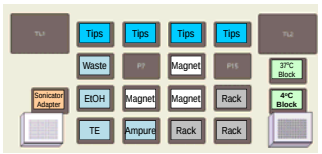
Scaling Up of 454 Titanium Std Library Construction

Robotic Process Of Constructing 454 Titanium Libraries

Sonication
↓
Purification & Size Selection SPRI beads
↓
End polishing reaction
↓
Purification SPRI beads
↓
Blunt-ended Barcoded-Adaptor ligation
↓
Purification SPRI beads
↓
Library immobilization
↓
Fill-in Reaction
↓
ssDNA isolation & Column purification
↓
Library Quantification
Bioanalyzer / qPCR

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.

Deck Layout of the BioMek Robot



We were able to successfully develop and implement tests on lambda DNA and two plates of 96 fosmid samples each. 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

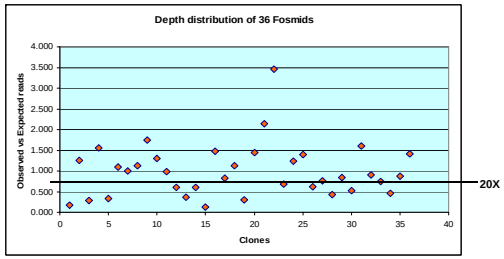
96 fosmid clones were selected from the Antarctic Marine Bacterioplankton Winter pools. Making 96 individual libraries and sequencing them separately was a challenge and would have generated huge amount of data. We designed the molecular barcode sequences for the Titanium libraries based on the requirements 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

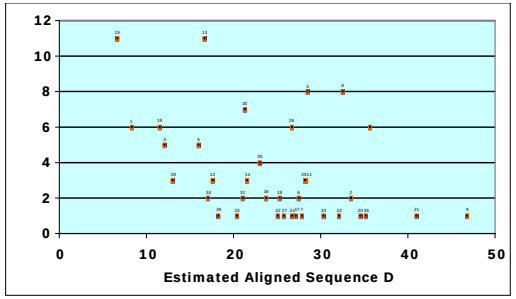
- Oligo length: 10 nucleotides (1,048,576 possible sequences)
- No consecutive same bases (78,732 sequences)
- 40-60% GC content (64,472 sequences)
- No more than 2 di-nt or tri-nt repeats (62,072 sequences)
- Adapter sequences differ by at least 4 bases (270 sequences)

Distribution of reads between 36 fosmid clones

To test this process fosmids were selected from a library derived from the endomycorrhizal fungus *Glomus intraradices*. 36 barcoded fosmid subclone libraries were sequenced in a quarter of a Titanium run. This run yielded 229,356 reads, containing 89.7 MB of sequence. If divided evenly between the fosmids, there would be 6,371 reads each; the actual read numbers sorted into these projects ranged from a low of 790 to a high of 22,083. Most of the projects have a less than 50% of read number deviation from the mean. We have also used the "number of allowed errors" parameter in the MID configuration file to test the ability of sorting these MIDs. 64 % of the clones got more than 20x coverage.



The Assembly of pooled 36 Fosmid clones

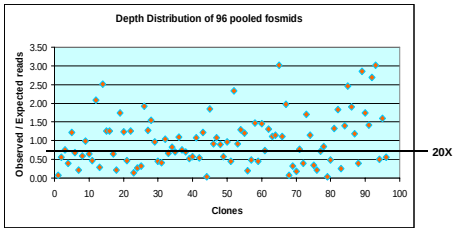


The Assembly of pooled 36 Fosmid clones

The assembly of these Titanium reads formed 13 complete fosmids (36.1%), 6 with 2 large contigs (16.7%), 12 with 3-6 large contigs (33.3%), and 5 with greater than 6 large contigs (13.9%). More than half of the fosmids (11 of 20) that received greater than 25-fold sequence depth formed single complete contigs. Local repeats within each fosmids do not seem to affect the assembly. So this dataset suggests that the pooled barcoded fosmid libraries with 454 sequencing seems to be a valid approach to sequence the *Glomus intraradices* genome.

Distribution of reads between 96 Fosmid clones

After the success of 36 barcoded fosmids clones, 96 barcoded fosmid clone libraries were pooled and sequenced in a quarter of a Titanium run. This run yielded 210,341 reads, containing 69.3 MB of sequence. If divided evenly between the fosmids, there would be 2191 reads; the actual read numbers sorted into these projects ranged from 93 to 6549 reads.



More detailed analysis of this data needs to be performed. From Glomus data, for example, if we had half a run then we would have gathered 200 Mb which means approximately 2 Mb per fosmid. Each fosmid size along with the vector is around 40- 43 kb which means most of the clones will get more than 20x coverage. For these 96 fosmid clones, a quarter Titanium run was performed so keeping the above as a guideline we would expect to get 20x coverage for most of the clone which is required for genome assembly. The above graph shows that 55 % of clones had more than 20x coverage.

Conclusions

1. We have been constructing Titanium Std libraries on a 96-well format using a modified version of the 454 protocol.
2. We have successfully programmed the BioMek robot to automate the construction of 96 libraries simultaneously with minimal ergonomic risk.
3. We have demonstrated that using a set of molecular barcodes to create and pool libraries for 454 sequencing is a valid strategy to sequence many target DNA and analyze separately. The data presented here shows that the assembly of glomus intaradices and the distribution of reads between 96 pooled Antarctic fosmids. 55 % of the clones received more than 20x coverage.

LLNL-POST-427520

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.