Title
Progress Towards a High Throughput 454 Titanium Library Construction

Authors
Cheng, J.-F.
Phung, W.
Peng, Z.
et al.

Publication Date
2009-01-30

Progress Towards a High Throughput 454 Titanium Library Construction

Jan-Fang Cheng, Wilson Phung, Ze Peng, Chris Hack, Harris Shapiro, JGI Production Sequencing Group, Chris Daum, and Susan Lucas

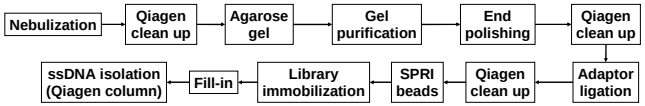
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Abstract

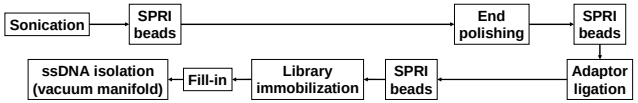
We have implemented the 454 Titanium sequencing as one of the production platforms at the Joint Genome Institute. The production tasks include de novo assemblies of many types of genomes ranging from bacteria to large eukaryotes, EST and metagenome sample sequencing. To construct a large number of high quality 454 Titanium standard (Std) and paired-end (PE) libraries on a routine basis has become a crucial step of this operation. In order to scale the Titanium library construction process, we have been developing new processes in three areas: (1) modify the 454 Titanium standard library construction process so it can be adapted to a 96-well plate format, (2) implement a set of MID and pooling process to sequence many targets simultaneously, and (3) develop a scalable process to create large insert paired-end libraries. Here we present the progress and the challenges of developing these scale-up processes.

Scaling Up of 454 Titanium Std Library Construction

single tube library construction



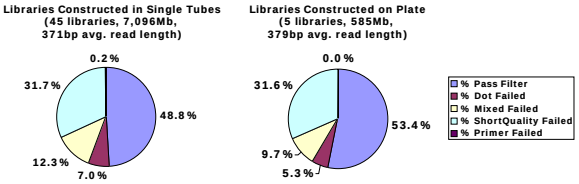
96-well plate library construction



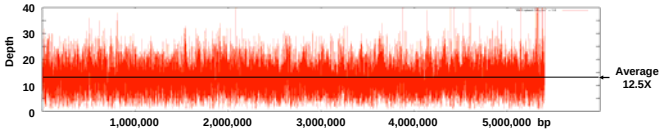
All the modifications were made so that the entire process can be adapted to the 96-well plate format. One major change is to replace the gel size selection step with a single SPRI bead size selection. We thought that it might result in an increase of short quality read failure.

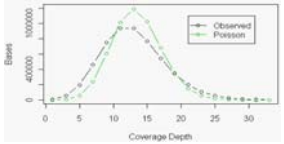
Quality Assessment of Libraries Constructed on 96-well Plate

(A) Read Quality



(B) Genome Coverage





We assessed the quality of the libraries constructed using a plate format by (A) comparing the read quality with libraries constructed in individual tubes, and (B) examining the randomness of the sequence coverage over the entire genome (Klebsiella variicola, 5.4Mb). All the parameters examined shows that the quality of the Titanium std libraries constructed on 96-well plates appears to be equivalent to those constructed in individual tubes.

Pooling of Titanium Libraries with Molecular Barcodes

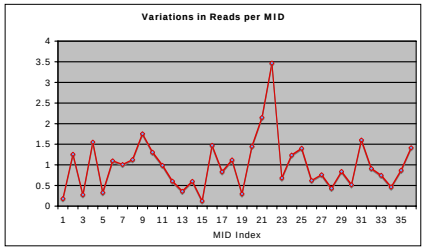
Rationale: The fosmids were selected from a library derived from the endomycorrhizal fungus *Glomus intraradices*. The whole genome shotgun sequences generated by the Sanger reads form thousands of contigs. One possibility is that the assembly is hindered by the occurrence of multiple copies of many nuclear genes, somewhat diverged in sequence. This 454 library pooling approach, if works, would provide a way to complete this genome.

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)

We have designed the molecular barcode sequences (AKA MID or multiple identifier) for the Titanium libraries based on the above requirements. We have selected 96 MID 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 36 fosmid subclone libraries, each containing a unique MID as shown in the right panel. These libraries were pooled with an equal mass according to the Bioanalyzer readings.

MID_index	MID_sequence
MID01	ACACTATGTC
MID02	CACACGTCTC
MID03	AGCATACTGT
MID04	GACAGTATAC
MID05	TACATGACGC
MID06	CGCATCTGCA
MID07	ATGACTGCGT
MID08	GCGATCTCTG
MID09	TCGAGCGTGA
MID10	GTGACACTGA
MID11	TGCTACGACG
MID12	CTCTACAGAG
MID13	ACTAGTACTC
MID14	CATACTCGCG
MID15	AGACTGATCA
MID16	CGACTCGAGA
MID17	ATCATCACTC
MID18	CTATACGCGT
MID19	GATACTACGA
MID20	TATAGCGACA
MID21	GCACTGAGAG
MID22	TCACTCGTCA
MID23	GTATGACAGA
MID24	TGATGCTGTA
MID25	AGTAGCTGAG
MID26	CGTCTCGTAT
MID27	ATACGCGCAC
MID28	CTACGTATGC
MID29	ACGAGTCTGT
MID30	CAGATAGAGT
MID31	AGAGTAGTGA
MID32	CGAGTCACAG
MID33	ATCGTCGCGA
MID34	CTCAGACTGC
MID35	ACTGCGACAG
MID36	CATGTCGAGC

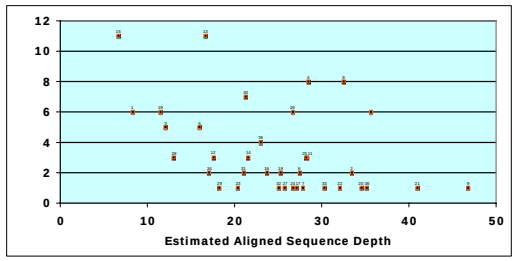


The pooled 36 barcoded fosmid subclone libraries were sequenced in a quarter of a Titanium run. This run yielded 229,356 reads, containing 89.7 MB of sequence. If divided evenly between the fosmids, there would be 6,371 reads each; the actual read numbers sorted into these projects ranged from a low of 790 (12% of the mean) to a high of 22,083 (~350% of the mean). 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.

We have also examined the ability of assigning these MIDs accurately by Newbler:

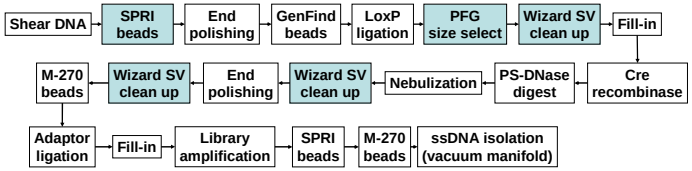
a) Allowing no error: 97.02% of the initial reads were assigned to a MID, the missing 3% could be due to a combination of error rate and contamination.
b) Allowing up to one error: 99.61% of the initial reads were assigned to a MID, all uniquely.
c) Allowed up to two errors: 99.80% of the initial reads were assigned to a MID, and 99.63% were uniquely assigned
d) Allowing up to three errors: 99.92% of the initial reads were assigned to a MID, but only 24.35% were assigned uniquely.

The Assembly of Pooled Fosmids



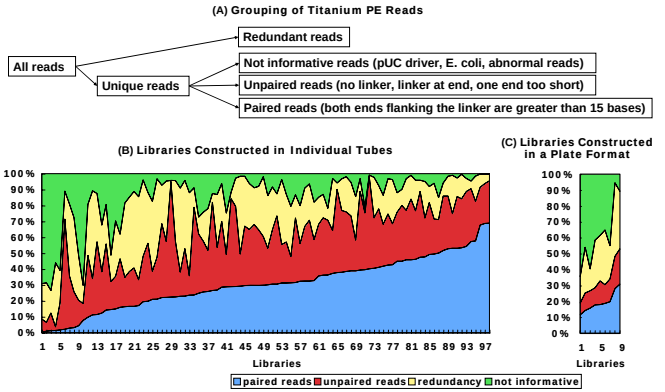
The assembly of these Titanium reads formed 13 complete fosmids (36.1%), 6 with 2 large contigs (16.7%), 12 with 3-6 large contigs (33.3%), and 5 with greater than 6 large contigs (13.9%). More than half of the fosmids (11 of 20) that received greater than 25-fold sequence depth formed single complete 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.

Developing a Scalable Method for Constructing Titanium Paired-end Libraries



This process is a modified version of the 454 Recombi Paired-end Library Construction Protocol. We can get the entire process in a 96-well format except the shearing of gDNA fragments to 20Kb at the beginning and the PFG size selection step. We continue to use pUC DNA as the driver so that we can measure the quality of the library at the end.

Quality Assessment of the Titanium Paired-end Libraries



The quality assessment process of the Titanium PE libraries is shown in (A). We grouped the reads in 4 categories. The redundant reads are defined as any two reads with greater than 95% of sequence match and no more than 2 bases difference in sequence length. We examined sequences of 98 Titanium PE libraries constructed using the 454 Recombi PE protocol (B). Of them, 9 contain less than 10% paired reads, 30 contain 10-30% paired reads, and 58 contain greater than 30% paired reads. The average insert size is about 17Kb. The chimeric rate based on two finished genomes is less than 5%. We have also constructed 9 PE libraries using the 96-well plate format. Their paired reads percentage ranges from 11.5 to 30.9% (C). We are in the process of optimizing this process.

Conclusions

1. **We have been constructing Titanium Std libraries on a 96-well format using a modified version of the 454 protocol. The sequence quality generated from these libraries are comparable to those constructed with individual tubes. We are currently programming a robot to perform this library construction tasks.**

2. **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 individual *Glomus intraradices* fosmids do not seem to be affected by the repetitive sequences in the genome. These repeats have been shown to hinder the whole genome shotgun assembly of this genome.**

3. **We have modified the 454 Recombi PE library construction process in the attempt to scale the 454 PE library production. Further improvement is underway to increase the quality of the libraries.**

Acknowledgements

We would like to thank Roche/454 Life Science for providing early access to the Titanium Recombi paired-end library construction reagents and protocol.

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.