Cloning and Sequencing with Trace Amount of DNA on Roche/454 and Illumina Platforms

[1]Mansi Chovatia, [2]Hope N. Tice, [1]Jeff Froula and [1]Jan-Fang Cheng

[1] Lawrence Berkeley National Laboratory, [2] Lawrence Livermore National Laboratory

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Abstract

As a User Facility, The US Department of Energy's Joint Genome Institute, in collaboration with scientists around the world, are able to generate DNA sequences from a diversity of organisms and environmental communities. Often times, the amount of genomic DNA provided for library construction is very limited. It is imperative to develop a protocol to minimize the amount of genomic DNA required for library construction for the second-generation of sequencing platforms. We have begun constructing Roche/454 Titanium and Illumina Paired End libraries with less than 1ug of genomic DNA by altering two key components from the standard operating protocol for library construction. The two key components that help minimize loss of genomic DNA are: shearing DNA via Covaris Adaptive Focused Acoustics™ (AFA) process instead of nebulization and utilizing Agencourt® AMPure® purification system to purify and select the size range of DNA fragments from contaminants and enzymes, with minimal loss of sample. This modified approach enables us to create Roche/454 Titanium libraries with less than 75ng of genomic DNA and Illumina libraries with only 1ng of genomic DNA as the starting material.

454 Titanium Library Creation

Initial DNA Mass: 1200ng, 600ng, 300ng, 150ng, 75ng
↓
DNA Fragmentation via Covaris
↓
Size Selection with SPRI Beads
↓
End-Polishing
↓
Adaptor Ligation
↓
Small Fragment Removal with SPRI Beads
↓
Library Immobilization
↓
Fill-In Reaction
↓
sstDNA Isolation
↓
8cycle Amplification
(of eluted sstDNA & beads from isolation)
↓
Size Selection with SPRI Beads
↓
Library Immobilization
↓
sstDNA Isolation

Note:

-Four different amount of starting genomic DNA of Actinobacillus succinogenes 130Z (2.3 Mb, 44.92%GC) were used to create the 454 Titanium libraries.

-The blue boxes to the left represent standard 454 Titanium library creation protocol.

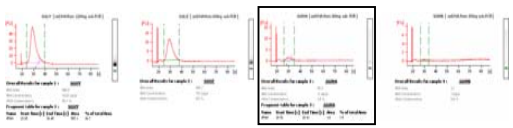
-The yellow boxes to the left represent the changes or additions made to the standard protocol optimized for minimal loss of sample.

-Three sets of libraries were constructed. The first set were created with no amplification, the second set were created with 8 cycles of PCR using the DNA on beads, and the third set were created with 8 cycles of PCR using the eluted DNA.

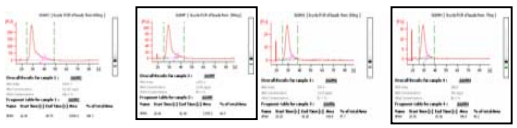
-The Bioanalyzer traces of these twelve 454 Titanium libraries that were eluted in 10µl of TE buffer, pH 8.0 are shown below.

sstDNA Bioanalyzer Traces for 454 Libraries

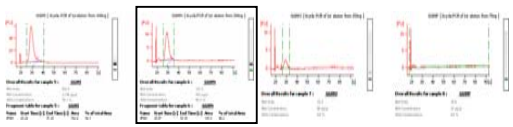
SET 1: Recovery of sstDNA without Amplification



SET 2: Recovery of sstDNA with Amplification of Beads from Isolation



SET 3: Recovery of sstDNA with Amplification of eluted Library from SET 1



We found that 300 ng seems to be the minimal amount of starting DNA for constructing libraries without amplification. With 8 cycles of PCR amplification, we can reduce the starting DNA to less than 75 ng based on the bioanalyzer readings. Only four of the twelve libraries, GGWA, GGWF, GGWH, and GGWN (outlined in black), were selected for sequencing and further quality analysis.

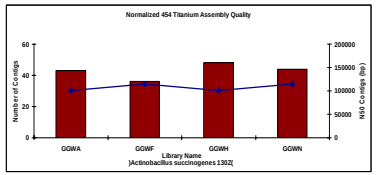
Quality of the 454 Libraries

A.

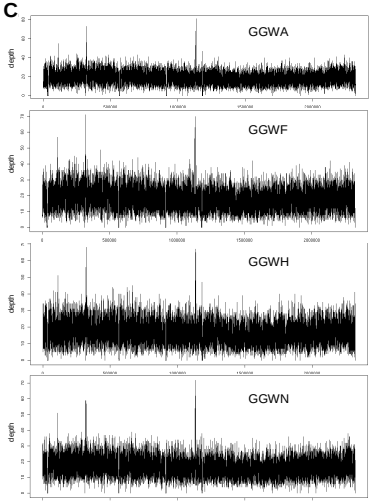
Summary Table of Sequences and Genome Mapping Results

Library Name	Sample Type	Number of reads used in the analysis	Average Size (bp)	Estimated Coverage	Aligned reads (%)	Gaps (%)
GGWA	300ng no PCR	118,908	390.46	19.99	98.57%	0.71%
GGWF	300ng w/ bead PCR	118,908	378.33	19.36	98.41%	0.72%
GGWH	75ng w/ bead PCR	118,908	345.7	17.66	97.75%	0.72%
GGWN	300ng w/ liquid PCR	118,908	343.81	17.6	98.57%	0.74%

B.



C



A. The quality of the sequence reads are very similar - We randomly selected the same number of sequence reads from each of the four constructed libraries for quality analysis. The reads were mapped to the reference genome using the Broad Aligner. We found that the percents of the total reads aligned to the genome and the percents of gaps are very similar among these libraries.

B. The quality of the assemblies of these four libraries are very similar - The selected sequences were assembled by Newbler. The resulted contig numbers range from 36 to 48. The N50 contig sizes range from 100.9 to 115.1 Kb.

C. The sequence coverages across the genome are very similar - We have found no significant coverage bias among the PCR amplified libraries (GGWF, GGWH, and GGWN). In fact, we have found a striking similarity of the sequence depth across the genome from the amplified as well as the non-amplified libraries. The locations of gaps and highly represented regions are very similar. The nature of the similarity in sequence coverage is under investigation.

Illumina Library Creation

Initial DNAMass: 1000ng, 100ng, 10ng, 1ng
↓
DNA Fragmentation via Covaris
↓
End-Repair and Phosphorylation
↓
Klenow-exo for dATP
↓
Adaptor Ligation
↓
Size Selection with SPRI Beads
↓
Library Immobilization
↓
Fill-In Reaction
↓
sstDNA Isolation
↓
12cycle Amplification
(of 1/30th of 1000ng and 1/3rd of 100ng, 10ng, and 1ng)

Note:

-The genomic DNA of Actinobacillus succinogenes 130Z (2.3 Mb, 44.92% GC) was used to create the Illumina libraries.

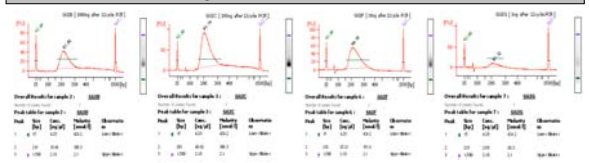
-The blue boxes to the left represent standard Illumina PE library creation protocol.

-The yellow boxes to the left represent the changes or additions made to the standard protocol optimized for minimal loss of sample.

-The amplified GGIB was eluted in 30µl of Qiagen Buffer EB and the remaining three, GGIC, GGIF, and GGIG, were all eluted in 15µl of Qiagen Buffer EB.

-All four of the Illumina libraries, GGIB, GGIC, GGIF, and GGIG, were selected for sequencing and further analysis.

dsDNA Bioanalyzer Traces for Illumina Libraries



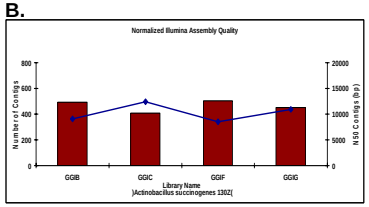
Assembly Quality of Illumina Libraries

A.

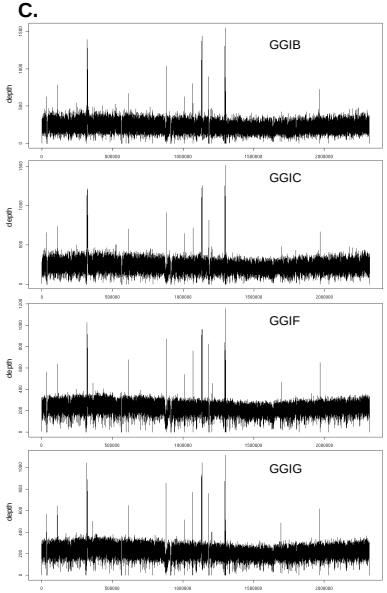
Summary Table of Sequences and Genome Mapping Results

Library Name	Sample Type	Number of reads used in the analysis	Average Size (bp)	Estimated Coverage	Aligned reads (%)	Gaps (%)
GGIB	1000ng	14,908,575	35	224.28	99.70%	1.08%
GGIC	100ng	14,908,575	35	223.58	99.39%	1.08%
GGIF	10ng	14,908,575	35	221.83	98.61%	1.08%
GGIG	1ng	14,908,575	35	216.21	96.11%	1.09%

B.



C.



A. The quality of the sequence reads are very similar - We randomly selected the same number of sequence reads from each of the four constructed libraries for quality analysis. The reads were mapped to the reference genome using the Broad Aligner. We found a small decrease of the percents of the total reads aligned to the genome when less DNA was used as the starting material. The percents of gaps are very similar among these libraries.

B. The quality of the assemblies of these four libraries are very similar - The selected sequences were assembled by Velvet. The resulted contig numbers range from 408 to 503. The N50 contig sizes range from 8.5 to 12.4 Kb.

C. The sequence coverages across the genome are very similar - We have found no significant increase of coverage bias when less amount of starting gDNA was used (GGIC, GGIF, and GGIG). We have also found a striking similarity of the sequence depth across the genome from all the Illumina libraries. The locations of gaps and highly represented regions are very similar. The nature of the similarity in sequence coverage is under investigation.

Conclusion

We have demonstrated the ability to create Roche/454 Titanium libraries and Illumina libraries with trace amount of genomic DNA. We can create Roche/454 Titanium libraries with as little as 300ng of genomic DNA, and with less than 75ng of genomic DNA after an 8cycle amplification of the sstDNA library. We can create Illumina libraries with only 1ng of genomic DNA as the starting material. The quality of the sequences and assemblies would not be affected by the amplification.

LLNL-POST-413430

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.