Modification of the GS LT Paired-end Library Protocol for Constructing Longer Insert Size Libraries

Ze Peng, Matthew Hamilton, Sara Ting, Hank Tu, Eugene Goltsman, Alla Lapidus, Susan Lucas, and Jan-Fang Cheng

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

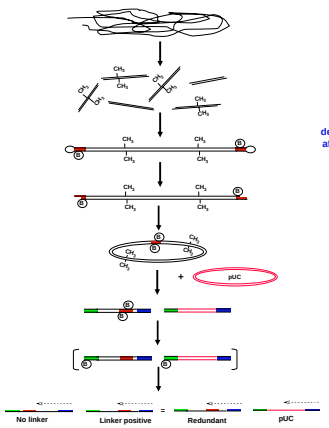
Abstraction

Paired-end library sequencing has been proven useful in scaffold construction during *de novo* assembly of genomic sequences. The ability of generating mate pairs with 8 Kb or greater insert sizes is especially important for genomes containing long repeats. While the current 454 GS LT Paired-end library preparation protocol can successfully construct libraries with 3 Kb insert size, it fails to generate longer insert sizes because the protocol is optimized to purify shorter fragments. We have made several changes in the protocol in order to increase the fragment length. These changes include the use of Promega SV column to increase the yield of large size DNA fragments, two gel purification steps to remove contaminated short fragments, and a large reaction volume in the circularization step to decrease the formation of chimeras. We have also made additional changes in the protocol to increase the overall quality of the libraries. The quality of the libraries are measured by a set of metrics which includes the amount of redundancy, the number of linker positive, linker negative, and half linker reads, as well as driver DNA contamination and read length distribution. We have also assessed the quality of the resulting mate pairs including levels of chimera, distribution of insert sizes, and genome coverage after the assemblies are completed. Our data indicated that all these changes have improved the quality of the longer insert size libraries.

454 LT Paired-end Library Construction Workflow and Quality Metrics

Workflow

1. Shear DNA
2. Purify the appropriate size fragments
3. Methylate DNA at EcoRI sites
4. Repair ends of DNA fragments
5. Ligate DNA to biotinylated hair-pin linker
6. Treat DNA with exonucleases
7. Purify and cut DNA with EcoRI enzyme
8. Determine DNA concentration
9. Circularize DNA fragments
10. Treat DNA with exonucleases
11. Add pUC DNA as driver
12. Nebulize DNA to average 300 bp in size
13. Repair ends of DNA fragments
14. Bind DNA to Streptavidin beads
15. Ligate LT adaptors to DNA
16. Remove nicks with polymerase
17. 15 cycles of PCR using biotinylated primers
18. Isolate DNA with Ampure beads
19. Bind DNA to Streptavidin beads
20. Denature ssDNA and neutralize



Quality Metrics

Quality determined after every run:
- Total reads
  - redundant reads
  - unique reads
    - pUC + E coli contamination
    - no linker reads
    - 1/2 linker reads
    - linker positive reads
      - ◊ read too short (<18)
      - ◊ sequence not found
      - ◊ repeats
      - ◊ chimera
      - ◊ true mate pairs
      - ◊ insert size distribution
      - ◊ genome coverage
      - ◊ sequence accuracy

Quality determined with ref genome or after bulk run assembly

GS Paired end LT Library Protocol Success in 3Kb Paired end LT Library Construction But failed in 8Kb Paired end LT Library Construction

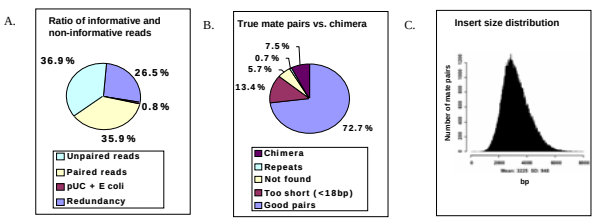
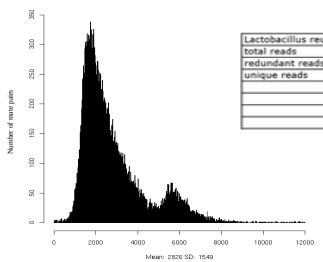


Fig.1 Shows Xylanimonas cellulosilytica DSM1589 3kb paired end LT library was constructed and 570K reads were generated from one 454 bulk run. 35.9% of the reads were paired reads (A), 72.7% of paired reads were match with reference genome (B), the average insert size is 3 kb (C).



Lactobacillus reuteri 100-23 (6KB) Bulk		
total reads	452226	100.0%
redundant reads	229474	50.7%
unique reads	222752	49.3%
pUC19 reads	20399	9.2%
no linker reads	135009	60.6%
1/2 linker reads	10696	4.8%
linker positive reads	56648	25.4%

Fig. 2, Using GS paired end LT library protocol was failed to construct 8kb paired end LT library of Lactibacillus reuteri 100-23. The 454 bulk run reads has 50.7% redundant (B) and major paired end reads are unexpected small insert size (A).

Modifying GS Paired end LT library Protocol to Improve 8 kb Paired end LT library Construction

1. Using Twice gel purification to remove small fragment contamination

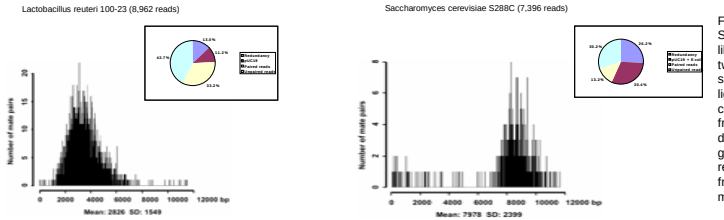


Fig.3: Shows the titration run data of S.cerevisiae 8kb paired end LT libraries constructed with or without twice gel purification (one after shearing, another after hairpin adaptor ligation) Without twice gel purification, contaminated unexpected small fragment has been selected become domain part of the library (A). Twice gel purification more completely remove unexpected small insert fragment, make 8 kb insert become major part of the library (B)

2. Using Tween-20 to Decrease pUC Driver DNA non specific binding to M-270 Streptavidin Beads

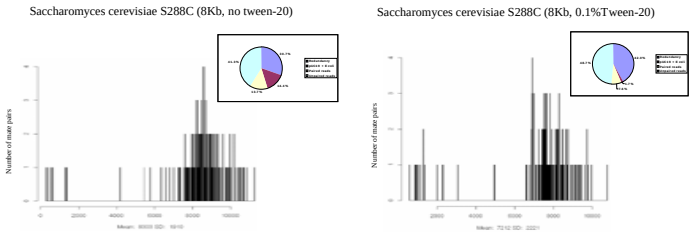


Fig.4 Shows the titration run data of S.cerevisiae 8kb paired end LT libraries constructed using the common condition of twice gel purification, 100ng/400ul 12C 24 hr circularization , and 15 cycle emPCR library amplify etc., except pre-treat M-270 beads with Tween-20 or not. There are more pUC 19 driver DNA no-specific binding to no-pretreated M-270 beads (A), Tween-20 can decrease this kind of non-specific binding. At the concentration of 0.1%, Tween-20 also decreased biotin labeled DNA binding to M-270 beads (B). Decrease the concentration of Tween-20 to 0.01% has no this effect. (data not show)

3. Increase DNA circularization volume to decrease Chimera Reads

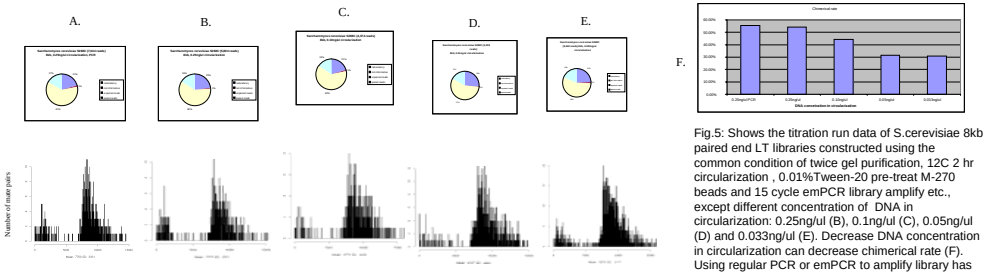


Fig.5: Shows the titration run data of S.cerevisiae 8kb paired end LT libraries constructed using the common condition of twice gel purification, 12C 2 hr circularization , 0.01%Tween-20 pre-treat M-270 beads and 15 cycle emPCR library amplify etc., except different concentration of DNA in circularization: 0.25ng/ul (B), 0.1ng/ul (C), 0.05ng/ul (D) and 0.033ng/ul (E). Decrease DNA concentration in circularization can decrease chimerical rate (F). Using regular PCR or emPCR to amplify library has little difference (A).

4. Using ATP dependant Plasmid-Save DNase to Further Decrease Chimera Reads

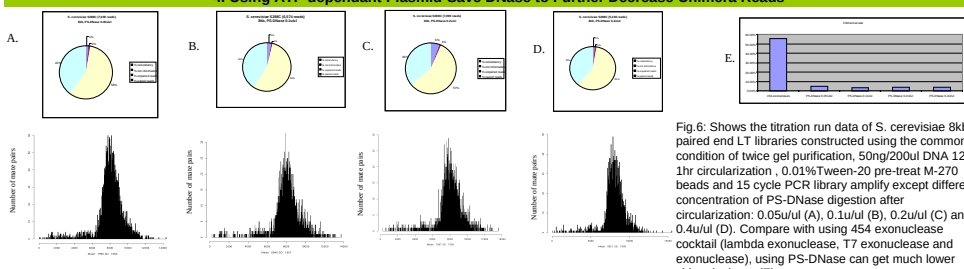


Fig.6: Shows the titration run data of S. cerevisiae 8kb paired end LT libraries constructed using the common condition of twice gel purification, 50ng/200ul DNA 12C 1hr circularization , 0.01%Tween-20 pre-treat M-270 beads and 15 cycle PCR library amplify except different concentration of PS-DNase digestion after circularization: 0.05u/ul (A), 0.1u/ul (B), 0.2u/ul (C) and 0.4u/ul (D). Compare with using 454 exonuclease cocktail (lambda exonuclease, T7 exonuclease and exonuclease), using PS-DNase can get much lower chimerical rate.(E)

Conclusion

Modify GS paired end LT library protocol can generate larger insert size (>8kb)library

Reference:

3K Long-Tag Paired End sequencing with the Genome Sequencer FLX System. NATURE METHODS, MAY 2008