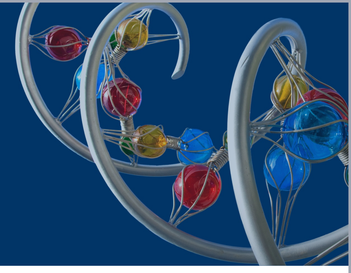
Comparison of Three Cre-LoxP Based Paired-End Library Construction Methods

Ze Peng*, Nandita Nath, Andrew Tritt, Shoudan Liang, James Han, Len Pennacchio and Feng Chen
US Department of Energy Joint Genome Institute, Walnut Creek, CA

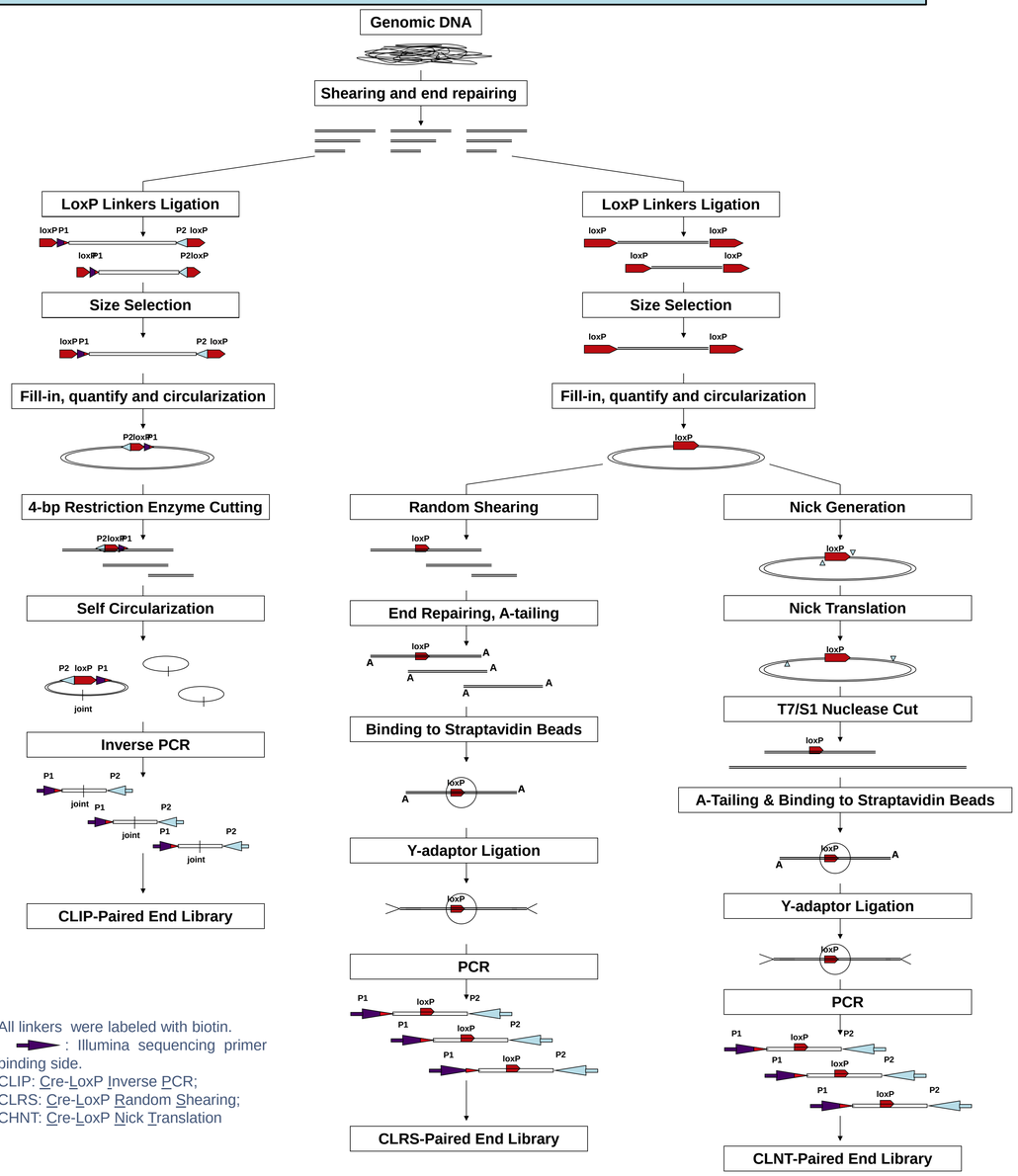
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



Abstract

Paired-end library sequencing has been proven useful in scaffold construction during *de novo* whole genome shotgun assembly. The ability of generating mate pairs with > 8 Kb insert sizes is especially important for genomes containing long repeats. To make mate paired libraries for next generation sequencing, DNA fragments need to be circularized to bring the ends together. There are several methods that can be used for DNA circulation, namely ligation, hybridization and Cre-LoxP recombination. With higher circularization efficiency with large insert DNA fragments, Cre-LoxP recombination method generally has been used for constructing >8 kb insert size paired-end libraries. Second fragmentation step is also crucial for maintaining high library complexity and uniform genome coverage. Here we will describe the following three fragmentation methods: restriction enzyme digestion, random shearing and nick translation. We will present the comparison results for these three methods. Our data showed that all three methods are able to generate paired-end libraries with greater than 20 kb insert. Advantages and disadvantages of these three methods will be discussed as well.

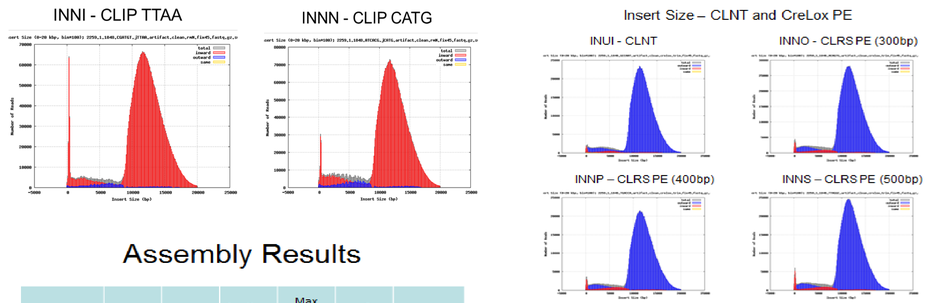
Three Cre-LoxP Paired-End library construction Procedure



All linkers were labeled with biotin.
: Illumina sequencing primer binding side.
CLIP: Cre-LoxP Inverse PCR;
CLRS: Cre-LoxP Random Shearing;
CHNT: Cre-LoxP Nick Translation

* E-mail: zpeng@lbl.gov

12 kb insert size libraries comparison



Assembly Results

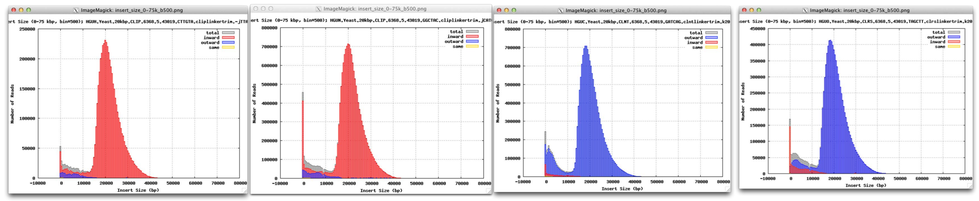
Library	# Scaffolds	# Contigs	Scaffold N50	Max Scaffold Length	Num Gaps (>50bp)	Total Bases
INUI - CLNT	217	511	122056	460049	266	12535209
INNI - CLIP	3036	7566	3749	22627	1422	9525652
INNN - CLIP	233	590	129544	662632	314	12895633
INNP – CLRS-3	266	635	120019	508484	332	12797136
INNS – CLRS-4	109	410	352030	992345	270	12275066
INNO – CLRS-5	105	366	441405	888174	229	12132645

- assemblies were done with ALLPATHS with 50x mate-pair coverage and 50x standard coverage (library HUSA)

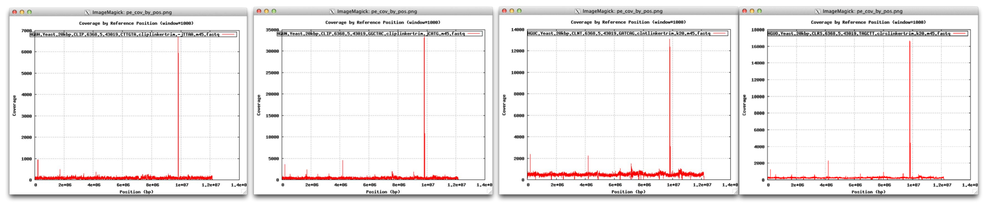
Yeast genomic DNA was used for constructing 12kb insert size CLIP, CLNT and CLRS paired-end libraries. CLIP library INNI was using MseI enzyme cut which generate 2bp overhang ends; CLIP library INNN was using NIaIII cut which generate 4bp overhang ends. CLRS libraries INNO, INNP and INNs were generated by random shearing of the insert to 300bp, 400bp and 500bp respectively. Although all libraries shown very similar insert size, libraries made from longer randomly sheared fragments (400-500bp) were the best in helping genome assembly.

20 kb insert size libraries comparison

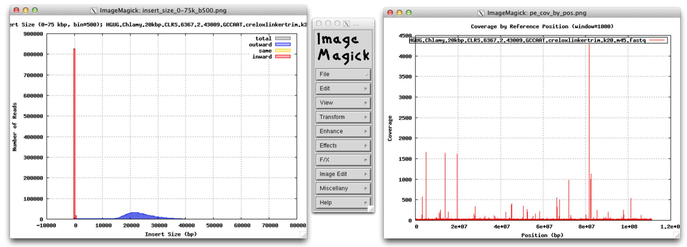
A



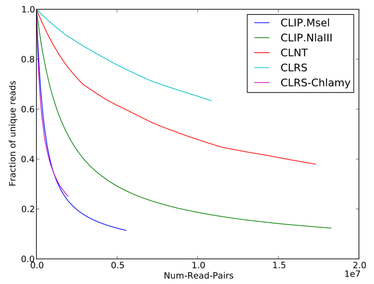
B



C



D



Yeast and Chlamy genomic DNA were used for constructing 20 kb CLIP, CLNT and CLRS paired-end libraries. CLIP library HGUH was made by using restriction enzyme MseI generating 2bp overhang ends; HGUN was made by using restriction enzyme NIaIII generating 4bp overhang ends; HGUC was made by CLNT method and HGUO is made by CLRS method. All three processes, CLIP, CLNT and CLRS made 20kb paired-end library from good quality DNA such as yeast gDNA (A). But from lower quality Chlamy DNA only CLRS made a library with a small right insert size peak along with big background peak (C) . Both CLIP and CLNT failed to make 20 kb libraries from Chlamy gDNA. Although CLIP, CLNT and CLRS libraries have very different unique reads (D), their genome coverage and cover pattern were very similar (B)

LBNL-6527E

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.