Fosmid Cre-LoxP Inverse PCR Paired-End (Fosmid CLIP-PE), A Novel Method for Constructing Fosmid Pair-End Library.

Ze Peng

7th Sequencing, Finishing, Analysis in the Future Meeting

Santa Fe, New Mexico

June 7, 2012

DOE Joint Genome Institute

Features and Applications of Fosmid Paired-end Library

- Features:

  Narrow insert size range (40±5 kb) selected by phage packaging

  High stability due to 1-2 copies in host cells.

- Applications:

  Genome sequence assemblies.

  Genome structural variation studies

Approaches of Generating Fosmid Paired-End Sequence

- Fosmid clone end sequencing by Sanger sequencer
  - Long reads but costly and laborious
- Fosmid Di-tag made by type III restriction enzymes cut
  - EcoP15I cutting generate short ends < 27bp

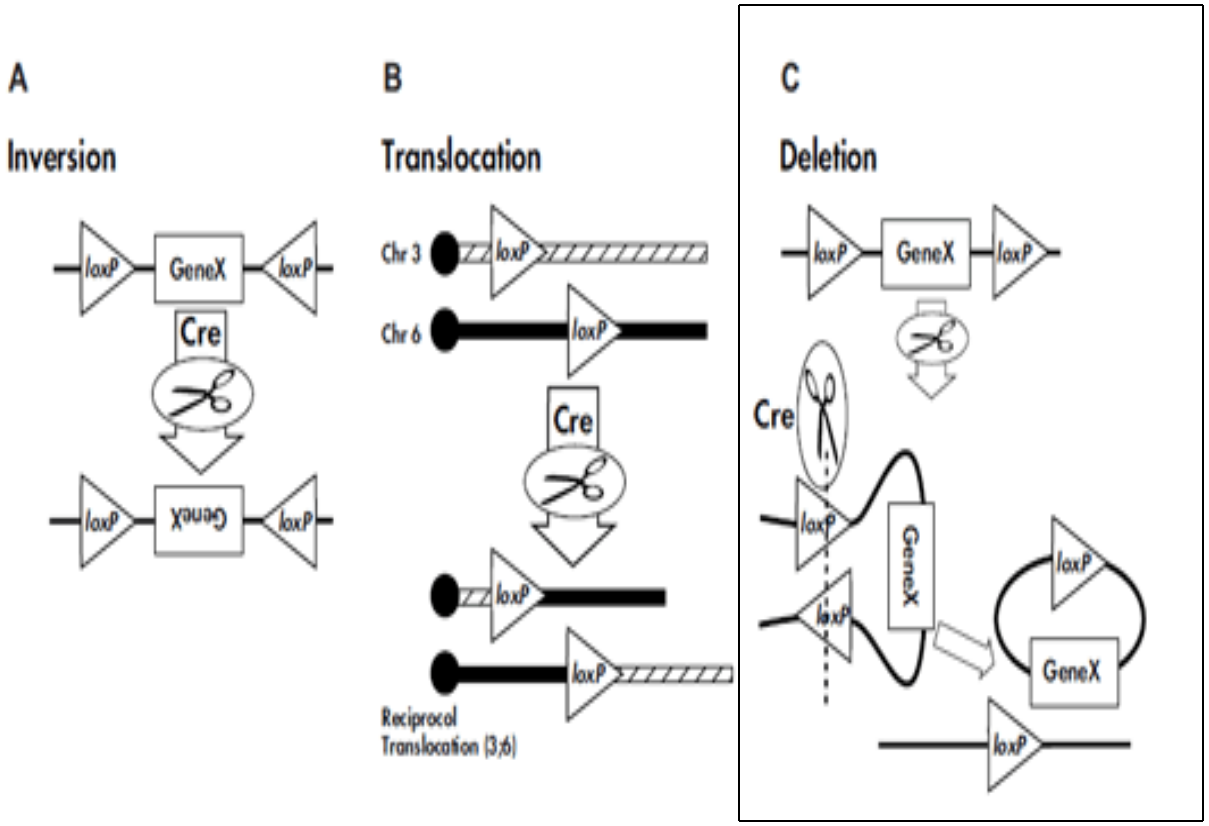
- Fosmid PE made by nick-translation (Broad)
  - Nb.BbvCI enzyme cut make nick site; blunt end ligation without known junction site
- Fosmid without some 4bp restriction enzyme cutting site (Lucigen)
  - Limited enzyme selection.

- Fosmid CLIP-PE
  - Vector deduction due to Cre-LoxP recombination make multiple enzyme digestion possible

Feature of Cre-LoxP System

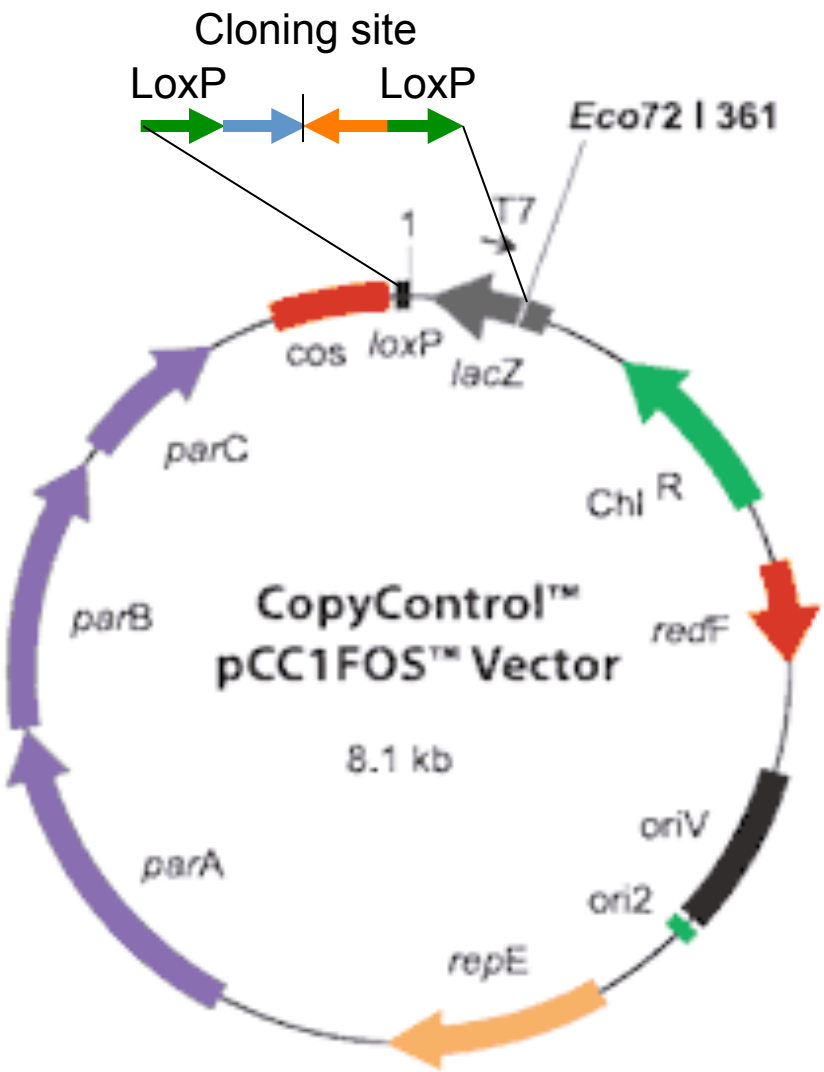
ATAACTTCGTATAGCATACATTATACGAAGTTAT

1. The 34-base loxP sequence. The core sequence (underlined) is flanked by two inverted repeats

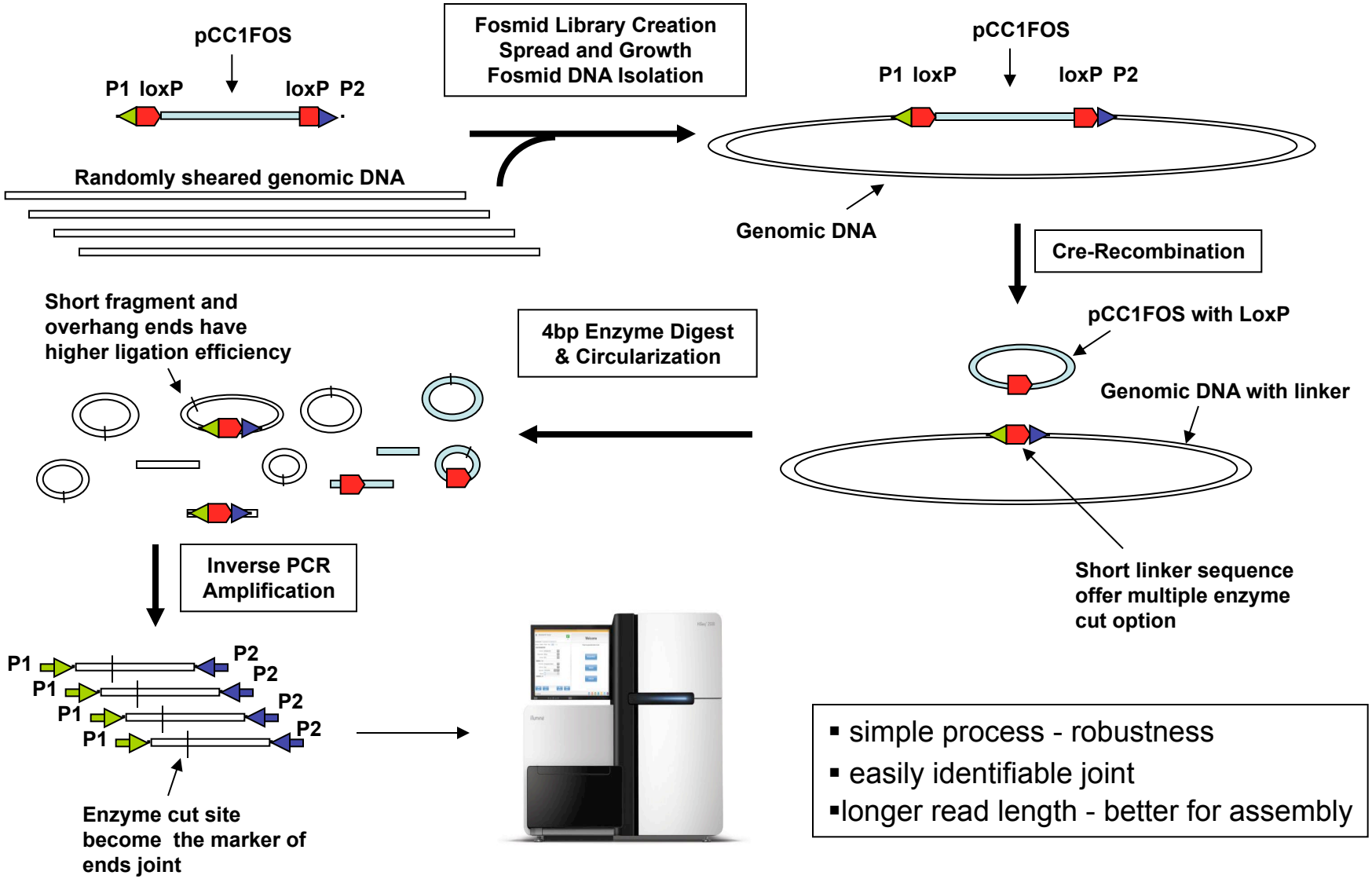


2. The outcome of a Cre-lox recombination is determined by the orientation and location of flanking loxP sites.

Modified pCC1Fos Vector



Illumina Fosmid CLIP-PE Workflow



- simple process - robustness
- easily identifiable joint
- longer read length - better for assembly

Candidate 4 bp Restriction Enzymes

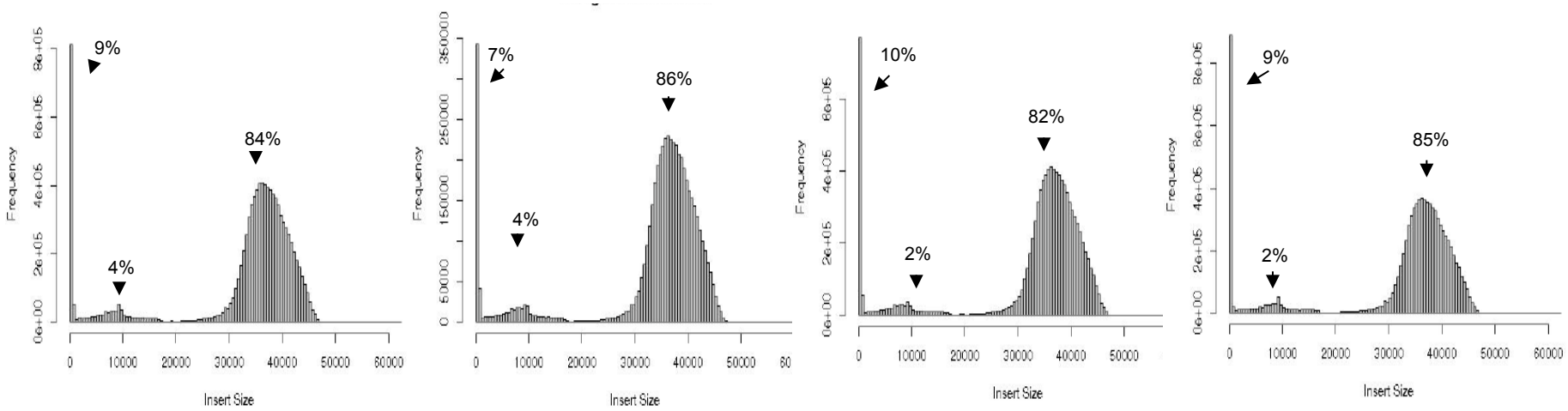
NlaIII	5′...CATG...3′ 3′...GTAC...5′
FatI	5′...CATG...3′ 3′...GTAC...5′
Tsp509I / MluCI	5′...AATT...3′ 3′...TTAA...5′
HpaII / MspI / AciI	5′...CCGG...3′ 3′...GGCC...5′
HhaI	5′...GCGC...3′ 3′...CGCG...5′
HinP1I	5′...GCGC...3′ 3′...CGCG...5′
TaqI	5′...TCGA...3′ 3′...AGCT...5′
CviAII	5′...CATG...3′ 3′...GTAC...5′
BfaI	5′...CTAG...3′ 3′...GATC...5′
CviQI	5′...GTAC...3′ 3′...CATG...5′
MseI	5′...TTAA...3′ 3′...AATT...5′
BstUI	5′...CGCG...3′ 3′...GCGC...5′
HaeIII / PhoI	5′...GGCC...3′ 3′...CCGG...5′
HpyCH4V	5′...TGCA...3′ 3′...ACGT...5′
RsaI	5′...GTAC...3′ 3′...CATG...5′

Principle of 4 bp Restriction Enzyme Selection

- Has no cutting site on linker sequence
- Generated ends has higher ligation efficiency:
  - 4 bp overhang > 2 bp overhang > blunt end
- Match GC content of the genome
- Features of the enzyme:
  - Reaction temperature, Heat inactive, working buffer etc.

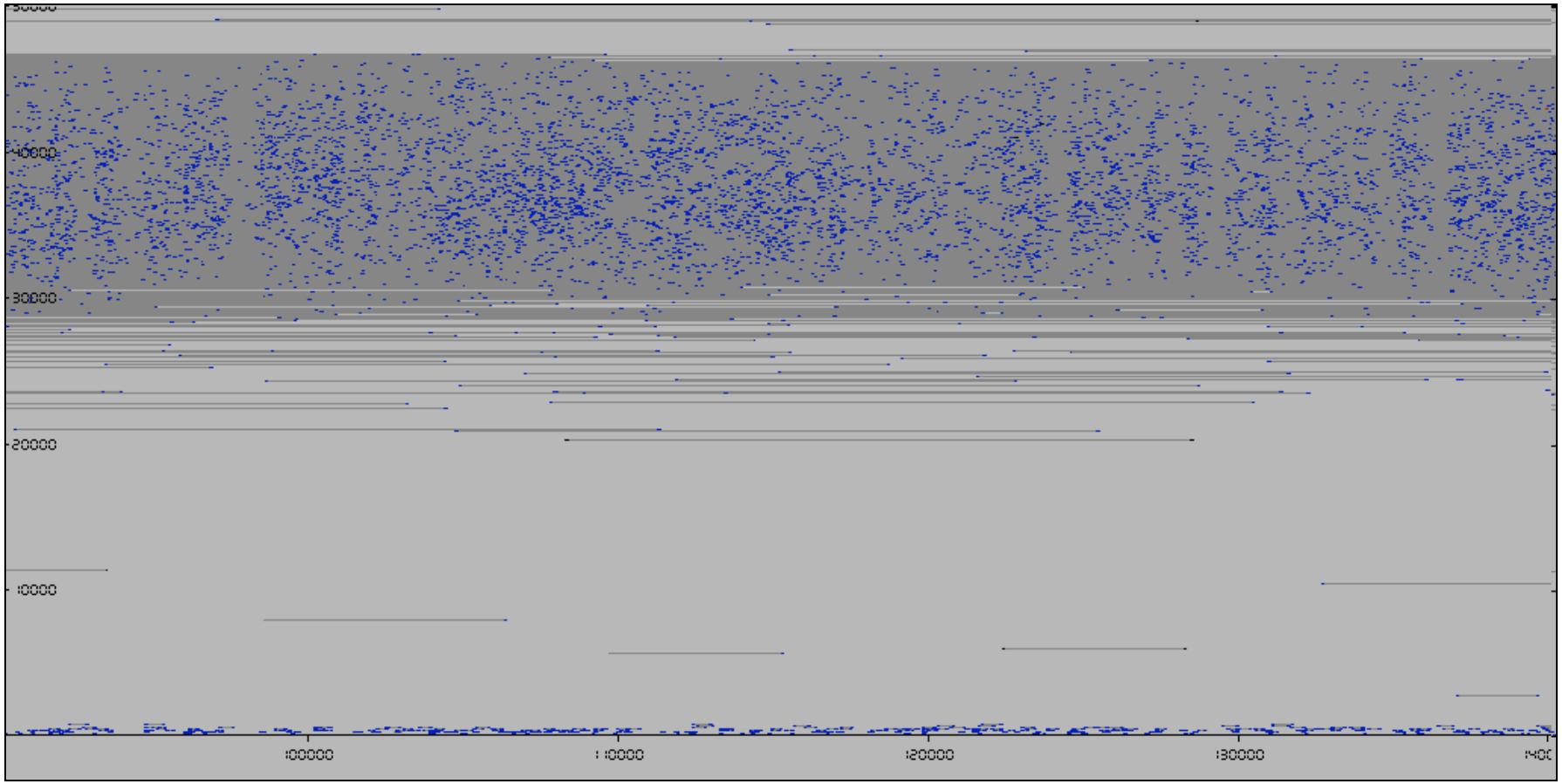
Results of Four S.cerevisiae S288c Genome Fosmid CLIP-PE Libraries

	HWFA	HWFC	HWFF	HWFG
Clone #	2,400,000	2,400,000	2,400,000	2,400,000
Total PE reads	42,266,926	20,413,370	38,530,578	50,100,726
# of Informative unique pairs	965,799	852,258	880,424	706,420
Median Insert size (kb)	36.37	36.48	36.41	36.27
Covered Bases	99.30%	99.30%	99.30%	99.30%
Median Clone Coverage	2,361	2,140	2,104	1,743



HWFA HWFC HWFF HWFG

HWFC Maps on S.cerevisiae Genome

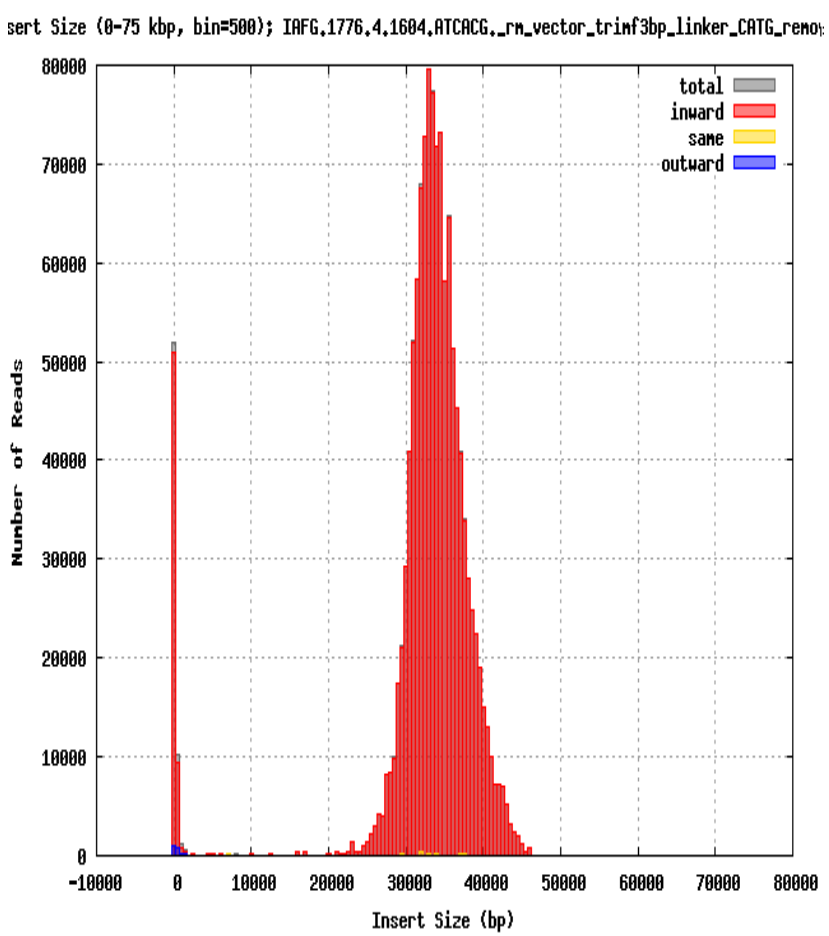


fosmid CLIP-PE libraries from 5 fungal genomes.

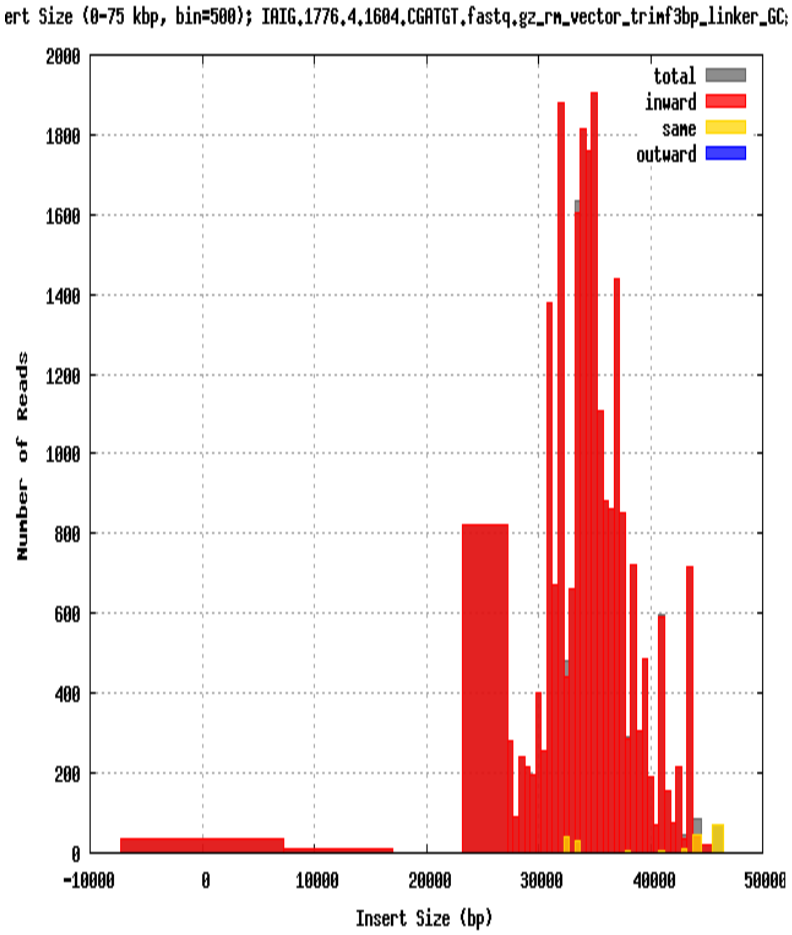
Genome	Lib. Name	CFU	Junction Tag	Insert size(kb)
Punctularia strigosozonata HHB-11173	**IAFG**	**2,340,000**	**CATG**	**31.505**
Punctularia strigosozonata HHB-11173	**IAIG**	**2,340,000**	**GCGC**	**34.922**
Coniophora puteana RWD-64-598 SS-2	**IAIH**	**2,880,000**	**GCGC**	**34.369**
Pichia membranifaciens NRRL Y-2026	**IAFI**	**120,000**	**CATG**	**31.249**
Auricularia delicata SS-5	**IAIN**	**780,000**	**GCGC**	**33.433**
Saccharomyces cerevisiae S228C	**IAIP**	**2,760,000**	**GCGC**	**31.673**

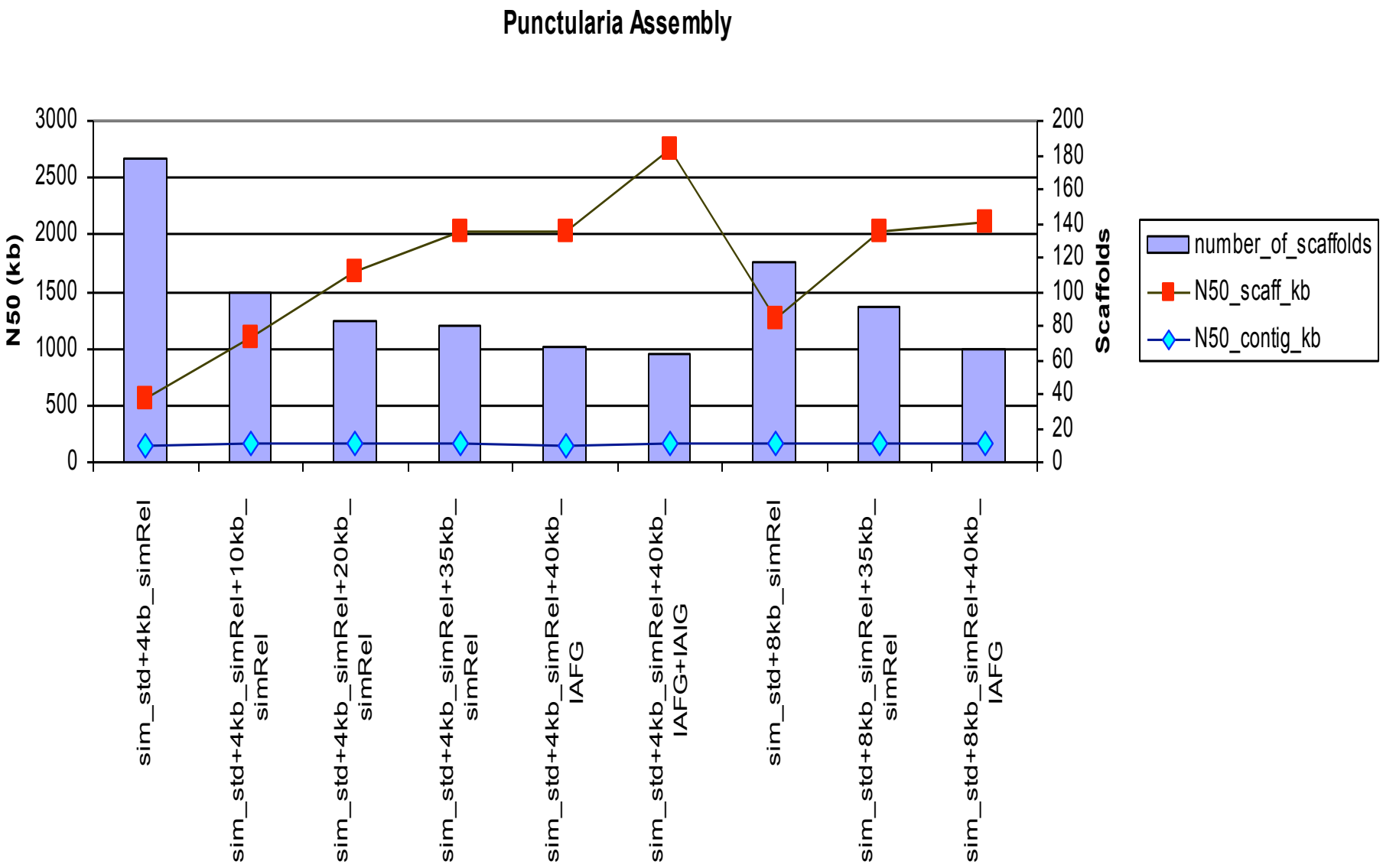
Punctularia strigosozonata HHB-11173 SS5 Fosmid Insert Size

IAFG: ▼CATG▲



IAIG: G▼CG▲C




Punctularia Assembly
N50 (kb)
Scaffolds
3000
2500
2000
1500
1000
500
0
200
180
160
140
120
100
80
60
40
20
0
number_of_scaffolds
N50_scaff_kb
N50_contig_kb
sim_std+4kb_simRel
sim_std+4kb_simRel+10kb_simRel
sim_std+4kb_simRel+20kb_simRel
sim_std+4kb_simRel+35kb_simRel
sim_std+4kb_simRel+40kb_IAFG
sim_std+4kb_simRel+40kb_IAFG+IAIG
sim_std+8kb_simRel
sim_std+8kb_simRel+35kb_simRel
sim_std+8kb_simRel+40kb_IAFG

Fosmid CLIP-PE Libraries from 6 Plant Genomes

Genome	Library name	CFU	Total clone	Junction Tag.	Insert size(kb)
Panicum hallii var filipes	IGYZ	15,000,000	30,750,000	CATG	36.005
Panicum hallii var filipes	IGZA	15,000,000		GCGC	36.534
Panicum hallii var filipes	IGZB	15,750,000		CATG	36.205
Panicum hallii var filipes	IGZC	15,750,000		GCGC	36.349
Capsella rubella Mte. Gargano (Italy)	IGZF	664.000	664.000	CATG	34.289
Capsella rubella Mte. Gargano (Italy)	IGZG	664.000		GCGC	20.65+/-19.08
Boechera stricta LTM	IGZH	928,000	928,000	CATG	26.18+/-17.06
Boechera stricta LTM	IGZI	928,000		GCGC	26.06+/-13.06
Brassica Rapa	INNA	3,168,000	8,608,000	CATG	31.56+/-13.36
Brassica Rapa	INNB	3,168,000		GCGC	31.46+/-9.73
Brassica Rapa	INNC	5,440,000		CATG	33.79+/-11.94
Brassica Rapa	INNF	5,440,000		GCGC	34.92+/-10.08
Thellungiella halophila	INIW	312,000	1,240,000	CATG	29.03+/-13.79
Thellungiella halophila	INIX	312,000		GCGC	26.88+/-14.45
Thellungiella halophila	INIY	928,000		CATG	30.04+/-14.16
Thellungiella halophila	INIZ	928,000		GCGC	28.61+/-14.30
Hordeum Vulgare	INIO	880,000	1,296,000	CATG	No reference
Hordeum Vulgare	INIP	880,000		GCGC	No reference
Hordeum Vulgare	INIS	416,000		CATG	No reference
Hordeum Vulgare	INIT	416,000		GCGC	No reference

Summery

- Fosmid CLIP-PE is a novel method making fosmid paired end libraries.
- Fosmid CLIP-PE help generate better genome assembly.

Acknowledgments

Library Sequencing:

Zhiying Zhao

Bioinformatics:

Jeff L. Froula

Kecia Duffy

Alicia Clum

Kurt LaButti

James Han

Supervisors:

Feng Chen

Jan-Fang Cheng

James Han

Alex C. Copeland

Len Pennacchio