Fosmid-pools and 40-kb mate-pairs

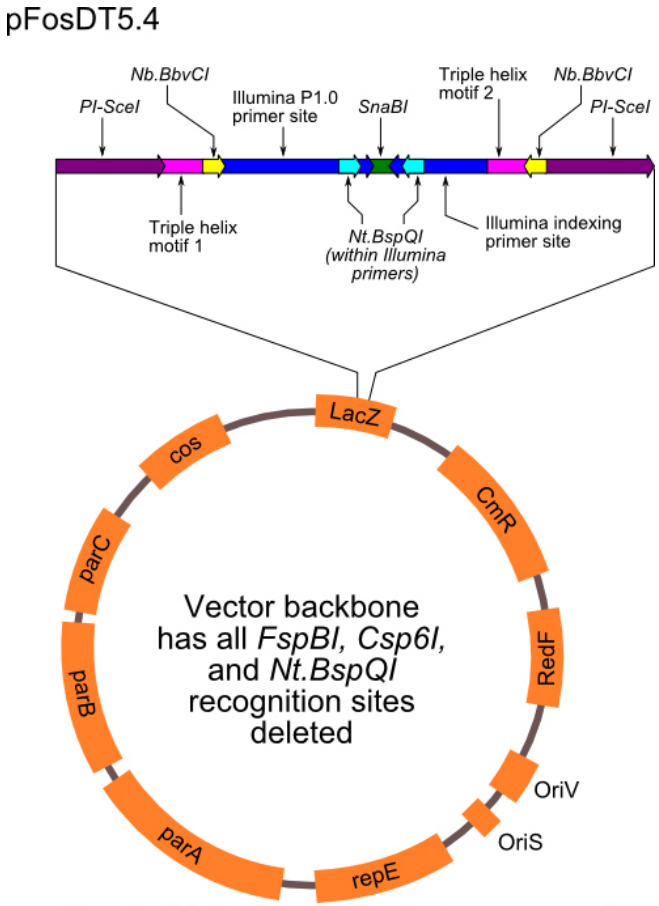
Children's Hospital Oakland Research Institute (CHORI)

BACPAC Resources Center

Strategies for Sequencing large genomes with high-repeat content

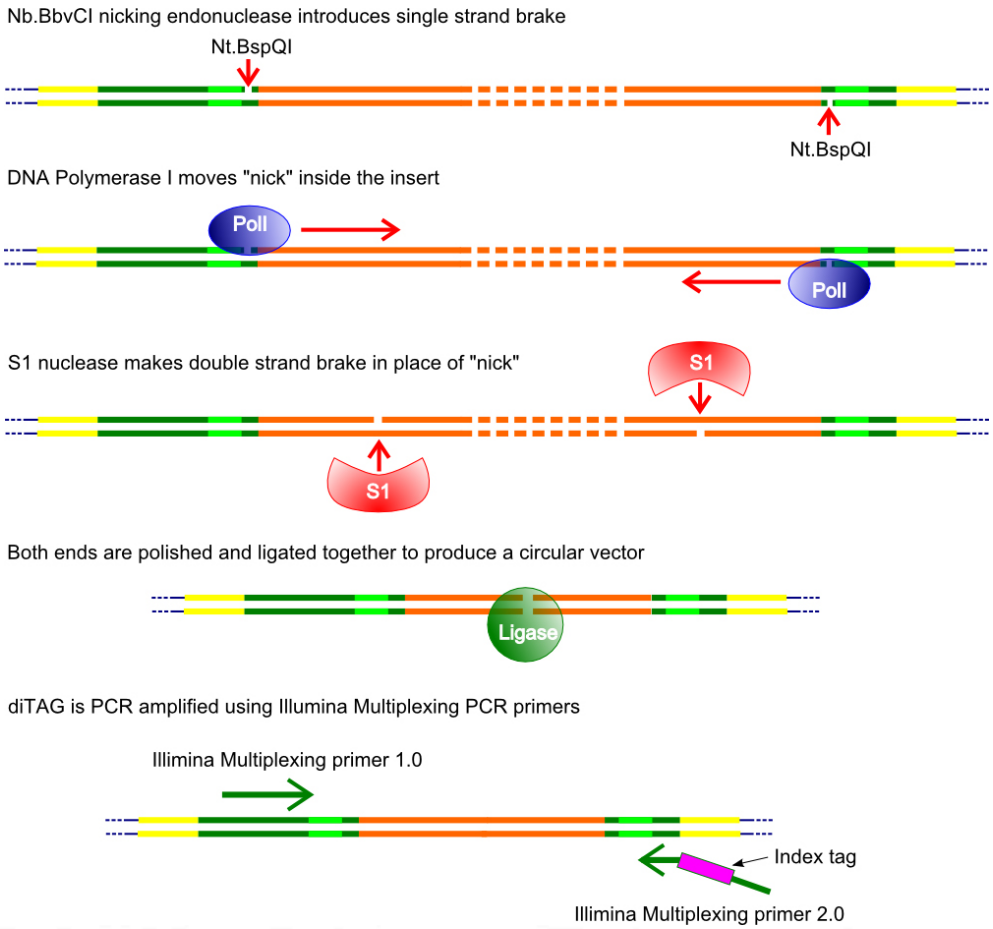
A. Scaffolding of contigs (produced by WGS or Pool Seq strategy) with an abundant source of long-range (~40 kb) mate-pairs.
B. Reduced complexity: NextGen sequencing from E.coli clone pools with combined inserts sizes far below genome size. Creates contigs limited to the size of clones.

Vector to create 40-kbp mate pair libraries ("diTags")



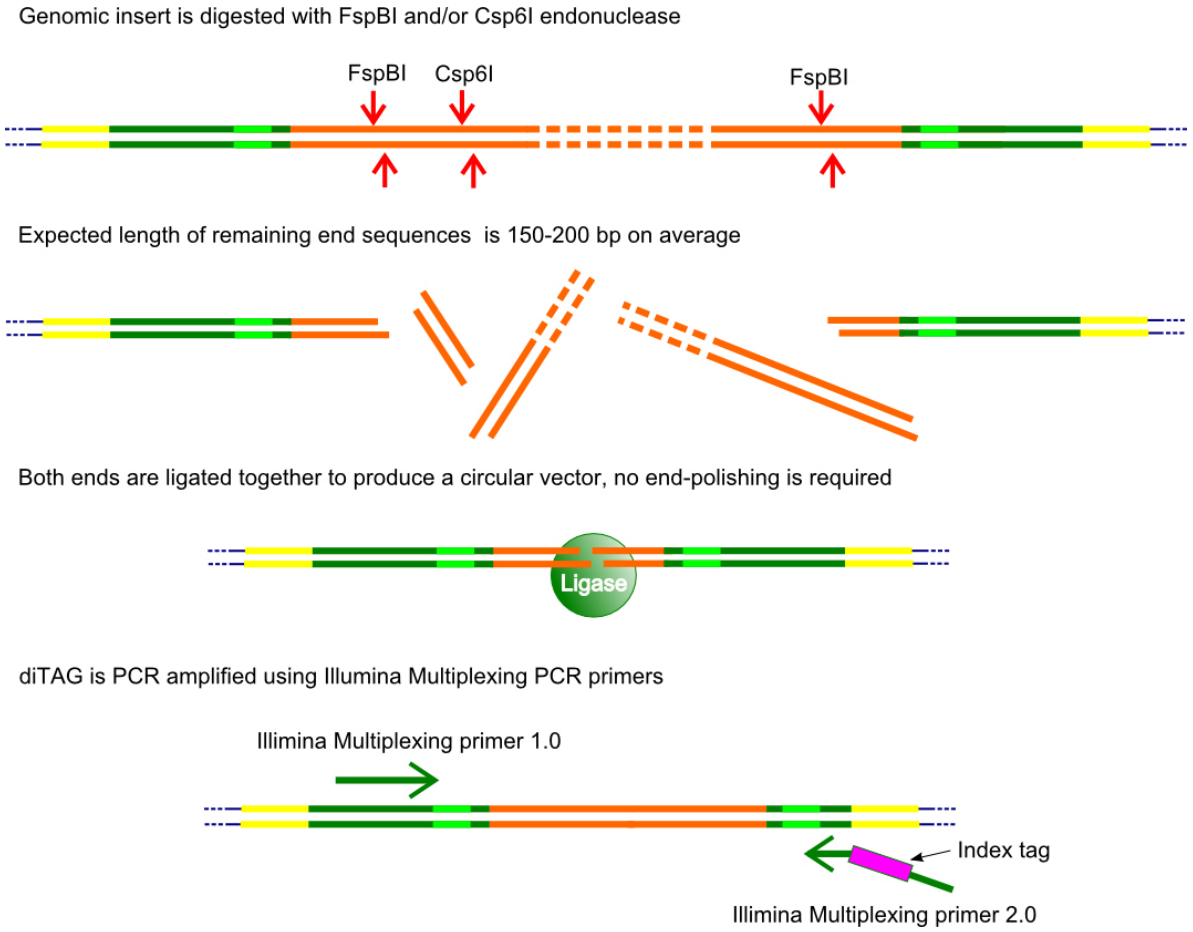
- Cloning of sheared genomic DNA in blunt-ended SnaBI site
- Includes options for creating mate-pairs by 5 strategies:
- Nicking, Nick translation, S1-nuclease (Nb.BbvCI or Nt.BspQI nicking sites)
- Fragment insert DNA with FspBI, Csp6I or both

Fragmenting fosmid inserts: nicking enzyme approach



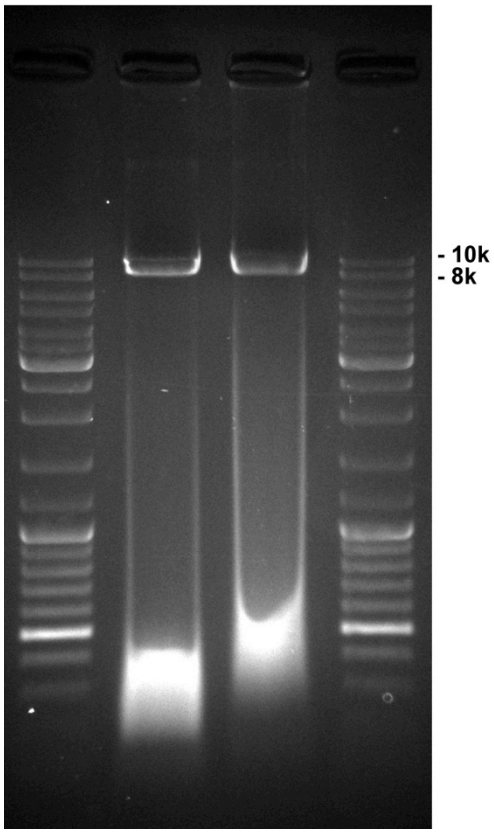
- Nicking of vector (and insert)
- Nick-translate from vector into insert
- Convert nick into break
- Isolate vector plus attached insert-ends
- Ligate
- PCT amplify
- Illumina Sequence

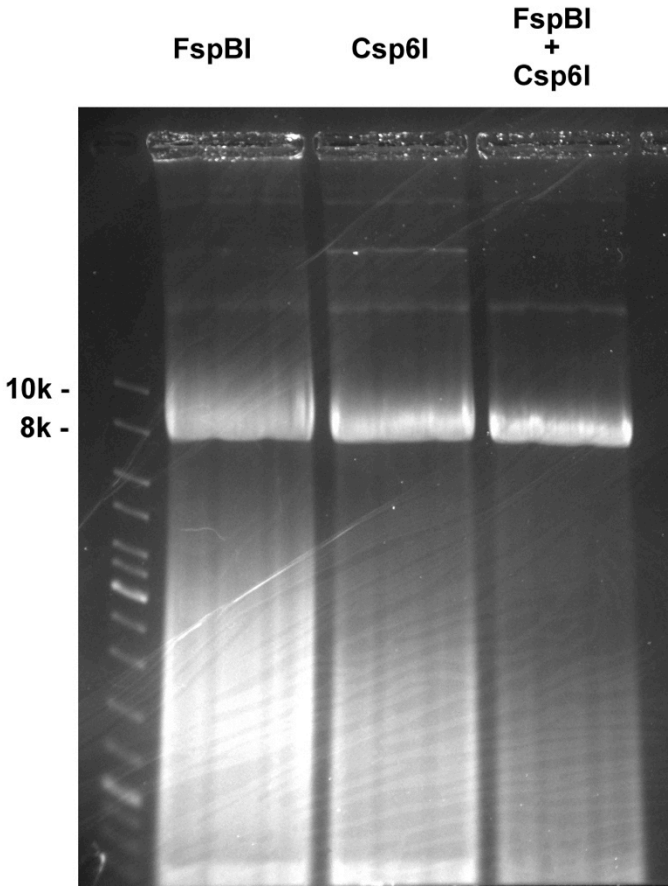
Fragmenting fosmid inserts: use of 4-cutters



- Digest with 4-cutter
- Isolate vector with attached insert-ends
- Ligate
- PCR-Amplify
- Illumina Sequence

Fragmenting fosmid inserts: use of FspBI, Csp6I or together



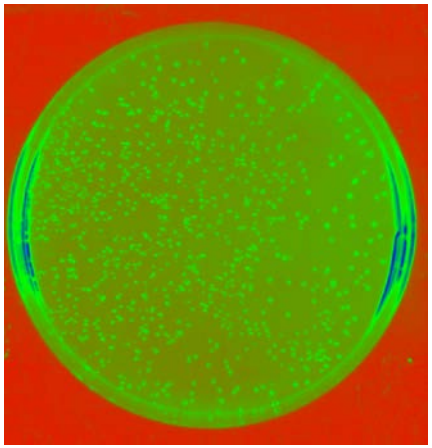


Reduced Complexity from E.coli clone pools

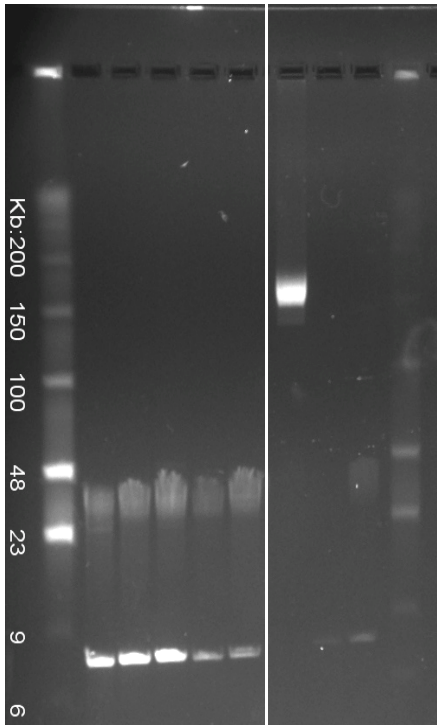
- Used by others to improve assemblies
- Used also to derive long-haplotype information
- Ideally: larger clones (BACs) are more desirable, more likely to span long “duplicons”
- Economically: much cheaper to create large fosmid libraries using sheared DNA (more random) with high genomic representation

Fosmid Pools

- DNA extraction
- Shearing (to average ~40 kbp)
- Size-purification (pulsed-field gel electrophoresis)
- Ligation to excess vector (dephosphorylated ends)
- Packaging (extracts from mcrA, B, C strains)
- Determine titer of the “particle library”
- Create E.coli colonies (e.g. 1,000/150 mm Petri dish)
- Isolate DNA from colonies and create Illumina seq libraries
- Quality Assessment: Quantitation & Q-PCR (TAQMAN assays) to determine E.coli & fosmid vector contamination



Fosmid colonies: Documented & counted prior to pooling



Fosmid pools: DNA digested (PI-SceI); Lane 6: Undigested pool

DNA Isolation from fosmid Pools

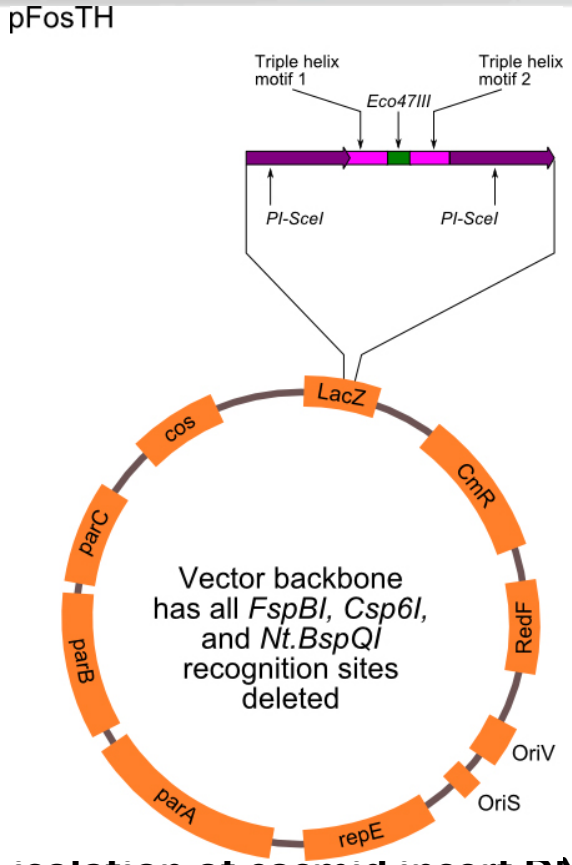
Criteria:

- Scalable?
- Representative?
- Artifact levels (e.g. chimeric reads?)
- Costs?

Options:

- Midi-prep columns: high quality/quantity/cost
- Small prep's followed by amplification (Rolling Circle Amplification, RCA)
- Crude bacterial lysates purified using biotin-labeled Triple-Helix probes (magnetic beads)
- Triple-helix followed RCA

Triple_Helix Purification of fosmid inserts*



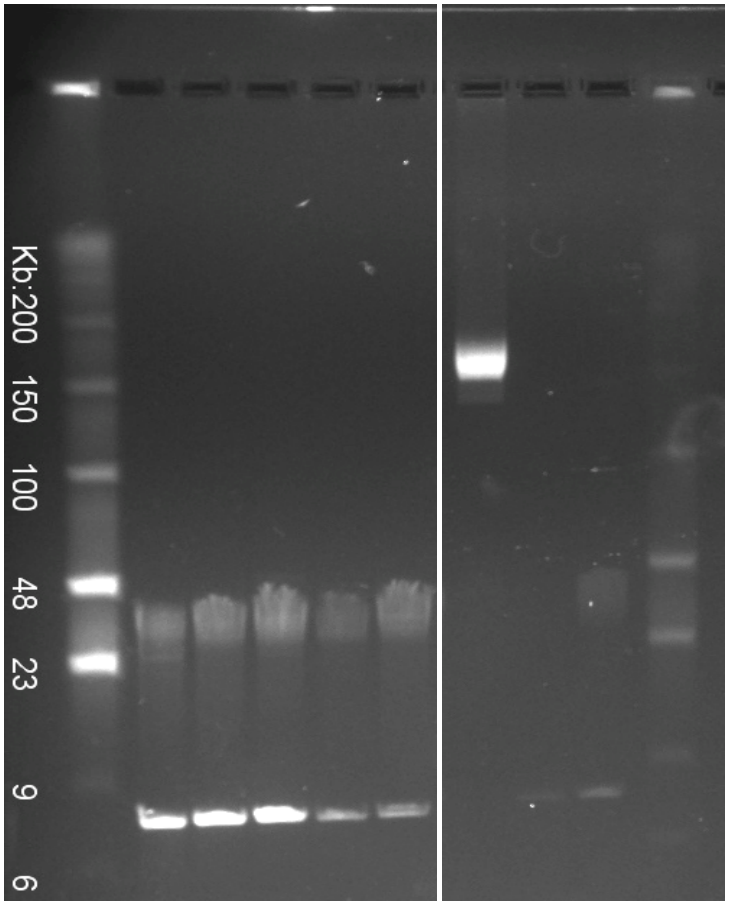
- Removal of vector & E.coli contamination
- Scalable
- Simple Purification

*** Rapid isolation of cosmid insert DNA by triple-helix-mediated affinity capture.**
Ji H, Smith LM, Guilfoyle RA. Genet Anal Tech Appl. 1994;11(2):43-7.

*** Rapid restriction mapping of cosmids by sequence-specific triple-helix-mediated affinity capture.** Ji H, Francisco T, Smith LM, Guilfoyle RA. Genomics. 1996 Jan 15;31(2):185-92.

Results

- Newest vector now in production: pFOSTH1.0
- Pools generated from midi-prep DNA, 12 pools sequenced
- Pools generated from mini-prep & RCA (enriched for E.coli mobile element: IS150)
- Pools now being generated with pFOSTH1.0



Fosmid pools in pFOSTH1.0 vector : PI-SceI -digested ; Lane 6: no digest

(l to r) Co-PD **Pieter de Jong**, **Ann Holtz-Morris**, **Maxim Koriabine**, **Boudewijn ten Hallers**

Acknowledgments

- Work at CHORI by Ann Holtz-Morris, Maxim Koriabine, Ph.D., and Boudewijn ten Hallers
- Loblolly DNA provided by Dana Nelson, Southern Institute of Forest Genetics
- Many collaborators within Loblolly Pine consortium
- PI: David Neale, UC Davis