

Lawrence Berkeley National Laboratory

Recent Work

Title

Strategies for Efficient Fosmid Sequencing Using 454 Sequencing Technology

Permalink

<https://escholarship.org/uc/item/35s20962>

Authors

Chen, Feng

Jett, Jamie

Smith, Douglas

et al.

Publication Date

2007-02-05

Strategies for Efficient Fosmid Sequencing Using 454 Sequencing Technology

Feng Chen, Jamie Jett, Douglas Smith, Edward Kirton, Andy Yuen, Jennifer Kuehl, Pilar Francino, Philip Hugenholtz, Eileen Dalin, and Paul Richardson

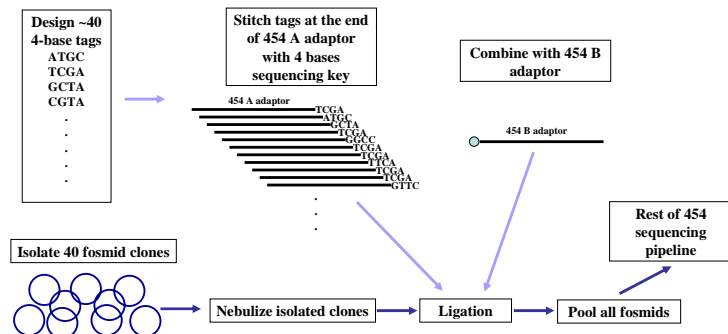
Introduction

The Department of Energy Joint Genome Institute (www.jgi.doe.gov) in Walnut Creek, CA is a high throughput DNA sequencing facility with a current throughput of approximately 3 billion Sanger base pairs per month. A major effort at JGI is the sequencing of microbial genomes of relevance to the DOE missions of carbon sequestration, bioremediation and energy production. The JGI Microbial Program and Community Sequencing Program together are responsible for the generation of sequencing data for over 400 microbial genomes. At the traditional Sanger sequencing side, JGI currently operates about 100 capillary instruments. JGI currently runs 2 Roche's GS20 instruments to supplement our traditional Sanger sequencing. Our current whole genome shotgun sequencing strategy is to sequence 3kb and/or 8kb shotgun libraries to a combined 4-8x draft coverage and to sequence fosmid ends to 1x sequence coverage with Sanger sequencing and to supplement that with 15-25x coverage with 454 sequencing platform depending on the sizes of the genomes.

JGI is also involved in an increasing number of metagenomic projects sequencing. Termite hindguts, sludge, gutless worm microbes, ventilator-associated pneumonia and terephthalate-degrading bioreactor are just a few examples for JGI metagenomics programs. Shotgun metagenomic sample sequencing and high throughput 16S rRNA sequencing and analysis are two major strategies JGI applies to metagenomics research.

Fosmid sequencing is crucial in many aspects of JGI sequencing and research programs. Our hybrid strategy of microbial genome sequencing utilizes 0.5 - 1x sequencing coverage of fosmid end sequences to build contig scaffolds. Remaining gaps after draft sequencing of traditional Sanger and 454 methods will be likely covered by fosmid clones. Obtaining the sequences of gap spanning fosmid clones is often used in the finishing stages of microbial genome sequencing. Fosmid clone sequencing also plays an important role in metagenomics research. Phylogenetically identified fosmid clones or fosmid clones with end sequences containing genes of interest can be sequenced to allow retrieval of large genomic fragments. Depending on the different purposes of downstream analyses, different fosmid clone sequencing strategy can be applied. Fosmid clones can be sequenced to 1) fully finished, 2) with ordered and oriented contigs but not finished, or 3) most of clone covered with contigs without order and orientation information. Cost and desired end products determine the strategy use.

Fosmid Sequencing Using Tagging Approach

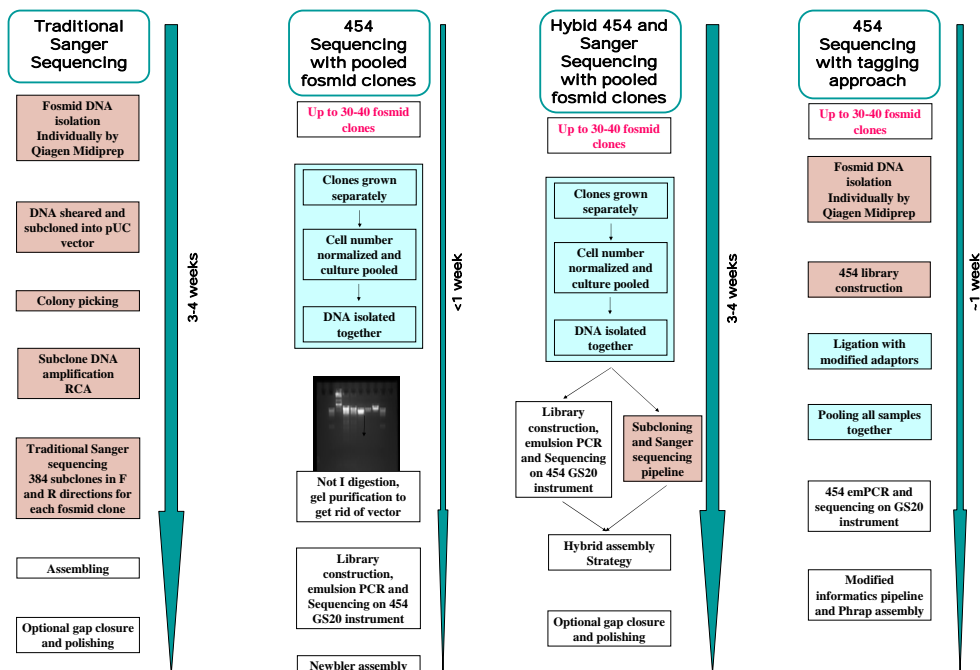


For each 454 read from the tagged fosmid shotgun library, the tag sequence (first four bases) was separated from the remaining sequence and the total error probability of the barcode was calculated. If the barcode sequence had an expected error rate of less than 5%, the unique read sequence was placed into the appropriate bin, otherwise the read was discarded. The reads in each bin belong to a particular fosmid and were assembled using Phrap assembler separately.

Discussion

Sanger only method is definitely the most time consuming and costly method. 454 only pooling method is most straightforward, least time consuming and least costly. It gives very good coverage and relatively good quality for the consensus sequences. The disadvantage is that this method does not provide assignment information, that is, the original fosmid clone for each assembled contig is not known. This method does not provide order and orientation information either. For certain applications, such as genome survey, this is a very good method to quickly generate enough sequencing information for downstream work. The assignment information as well as order and orientation information sometimes can be obtained when combined with known knowledge of the fosmid such as end sequences, or the genome such as gene structure information. Combining 454 pooling method with 2-3X of pooled Sanger sequencing provides best assembly results. The same pool of fosmid DNA goes to both 454 sequencing and Sanger sequencing. In our opinion, this is the most efficient way of sequencing large amount of fosmid clones. Without any finishing work, most of the sequencing information of the fosmid will be known. Tagging approach can be done in the same time frame as 454 only method but it can provide much more information of assigning contigs back to fosmid clones. Most of the contigs generated by Phrap assembler can be assigned back to original fosmids. However, it does not provide order and orientation information. This is very useful when the sequencing information of large genomic fragment around an interesting locus is needed.

Comparison of Methods



Results

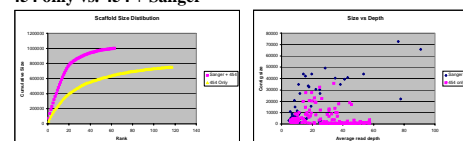
Coverage, assignability and order/orientation comparison:

The amount of 454 sequences assignable to fosmids was estimated by taking the average contig length, doubled to account for each fosmid end and divided by the average fosmid length.

The advantages of each method are highlighted in the table:

	Pooled 454 only	Pooled 454+Sanger subclones	Tagged and pooled 454
Average % bp covered	>99%	>99%	>99%
Average % bp assignable to fosmid	26.80%	99.86%	>99%
Average % bp ordered & oriented	26.80%	99.86%	28.45%
Difficulty level for the process	easy	difficult	moderate
Time needed for the process	~1 week	~4 weeks	~1 week
Cost	\$	\$\$	\$

454 only vs. 454 + Sanger



Acknowledgements

Production Department at JGI

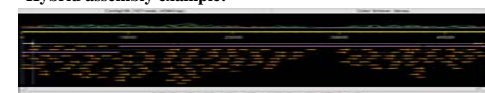
Hector Garcia Martin
Tanja Woyke
Vaishampayam Parag

Falk Warnecke
Darren Platt
Christopher Hack

Tagged and pooled fosmid sequencing method:

In pooled and tagged 454 sequencing, 99% of 454 reads contain high quality tag sequences. After binning and trimming of the tag sequences, assemblies looked very similar to the ones from pooled 454 only sequencing in terms of average number and length of contigs. Number of reads for each tag is not uniformly distributed even though we had a normalization step before we pooled all samples together. We will further optimize the normalization process.

Hybrid assembly example:

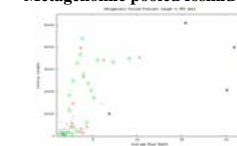


Graph provided by Falk Warnecke

Contigs can be scaffolded by known gene information:



Metagenomic pooled fosmid hybrid assembly:



Sequencing depth vs. contig size for different metagenomic samples in same pool of fosmid clones. Selected sludge fosmids actually overlap with each other.

Graph provided by Darren Platt