Fosmid transposition solves large sequence gaps

Chelsea Dunbar, Nicole Allen, Daniel Bessette, Tashi Lokyitsang, Charles Matthews, Tamrat Negash, Thu Nguyen, Michael FitzGerald
Finishing Laboratory Group, Broad Institute, Cambridge, MA, USA

Process Overview

Fosmid transposition is a workflow utilized by the Finishing Laboratory group to successfully capture large sequence gaps and rapidly provide a complete sequence for tiling path clones. This process can be executed when fosmids act as the only capturing clone in addition to sequencing large gaps. Since transposons insert themselves randomly throughout a target clone, thorough coverage can be achieved rapidly, often within two weeks. In our process, a single bacterial colony is isolated from a sample of the respective fosmid library glycerol. We electroporate the transposed fosmids in order to recover the 672 colonies required for sequencing. The transposed fosmid clones are sequenced bidirectionally from primer binding sites on the transposons. To improve the accuracy of our process we have implemented the end sequencing of several fosmid library colonies to confirm clone identity prior to transposition. In the future, TempliPhi amplification may replace our alkaline lysis DNA prep to improve time and cost efficiency of amplification and sequencing. This process currently yields 1344 reads of approximately 600bp each that have been used by the Computer Finishing group to successfully close many gaps in human, mouse, and magnaporthe genomes.

Fosmid Transposition Process

The following are three reasons a transposition order will be requested for a fosmid clone:

- To capture a large sequence gap existing between contigs
- To resolve misassembled regions of sequence
- To capture gaps in tiling path positions

The fosmid library clones ordered were constructed using the following vector:

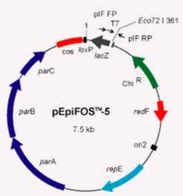
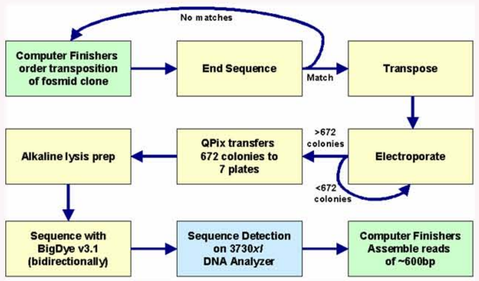


Figure 1. EpiFOS™-5 Fosmid Vector used for fosmid library construction (EPICENTER Biotechnologies; www.epibio.com).

Each transposition order of this library clone will enter the following pathway:



Order Placement

Once a sequence gap is determined to exist between contigs, fosmids identified to contain sequence spanning into both contigs are ordered for transposition.

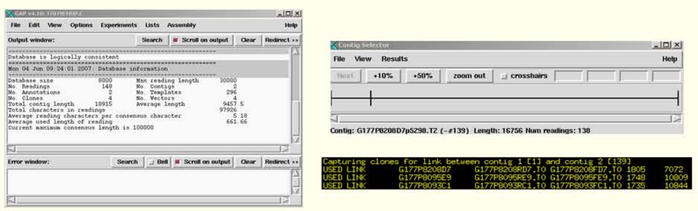


Figure 2. The screen shot above (left) displays the data on two contigs of a combined length of 18915bp. The position of the link between contigs is demonstrated in the Contig Selector window (top right). Three fosmid clones were identified to capture the link between the two contigs (bottom right). Each fosmid clone aligned with ~1800bp of contig 1 [1] and ~10kb of contig 2 [139]. Since our fosmids are ~40kb, a gap of ~30kb was identified. A fosmid transposition order was placed to sequence this gap.

End Sequencing

Twelve colonies are isolated from the fosmid library glycerol of the ordered clone. The purified DNA (extracted via alkaline lysis prep) is sequenced using T7 and Sp6 primers. Computer finishers select one clone whose end sequences are of the desired identity. This specific clone continues through the fosmid transposition process.

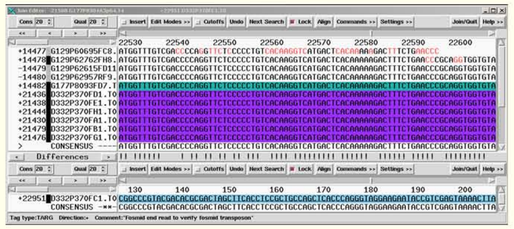


Figure 3. This screen shot demonstrates an example of end sequencing read alignments. The sequence tagged in teal is the parent sequence. The sequence tagged in blue does not match the parent sequence while all tagged in purple are matches. The clone representing the highest quality matching read is chosen.

Transposition

Tn5 is a transposon, a mobile DNA sequence, isolated from gram-negative bacteria. The Tn5 used in this process contains sequencing primer binding sites and a kanamycin resistance selection marker. Tn5 randomly inserts into our purified target fosmid DNA *in vitro* by Tn*5* transposase in a simple "cut and paste" step. Tn5 inserts with a high degree of randomness creating >10^6/µg insertion clones and therefore providing a representation of the fosmid sequence.

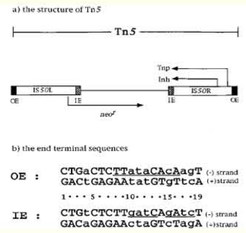


Figure 4. The EZ-Tn5™ Insertion Kit utilizes the above process (EPICENTER Biotechnologies; www.epibio.com). Tn5 containing a kanamycin resistance marker is used for our process.

Figure 5. The structure (a) and end sequences (b) of the Tn5 transposon.

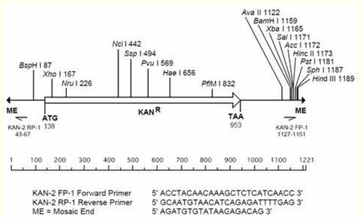


Figure 6. The layout and primer sequences for the EZ-Tn5™ <KAN-2> transposon (EPICENTER Biotechnologies; www.epibio.com).

Electroporation

The transposed DNA is transformed into DH10B electrocompetent cells via electroporation. This method uses a high-intensity electrical field to temporarily permeabilize the cell membrane, thus allowing large molecules such as our ~40kb fosmid DNA to enter the cell. Due to the high voltage used, our transposed DNA must be ethanol precipitated prior to electroporation to avoid "sparking".

BROAD INSTITUTE

http://www.broad.mit.edu

Sequencing of the Transposed Fosmid

The purified transposed fosmid DNA is sequenced bidirectionally using BigDye v3.1 and EZ:Tn5 kit sequencing primers in plate format and the sequences are detected on 3730*xl* DNA Analyzers. Each clone produces 1344 reads with an average NHGRI passing Q20 across all organisms of ~600. The Computer Finishing group then aligns the reads to attempt gap closure.

Figure 7. The screen shot above demonstrates successful closure of the gap shown in Fig. 2. The new total contig length confirms the presence of the large fosmid sequence data. The purple tag in the Contig Selector window represents the new transposed fosmid sequence data. The green tags represent the ends of the fosmid clone sequence.

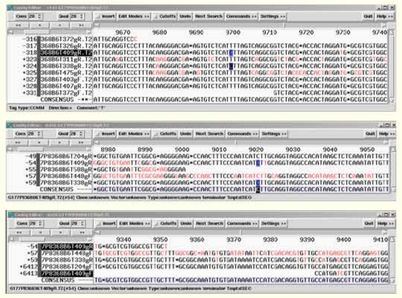


Figure 8. Misassemblies occur as demonstrated by the HQD (High Quality Discrepancy at a position between a base and the consensus sequence) positions marked in blue in the first window above. The sequences were realigned correctly as seen in the middle window. The forward and reverse sequence mates for one of the realigned reads (indicated in black) are shown aligned at the bottom.

Figure 9. This screen shot is an example of a misassembly that was unable to be correctly successfully. In this case another fosmid clone spanning the contig gap can be ordered for transposition.

Sample of Overall Process Results

- There are 5109 total mouse tiling path positions, 110 of which are held by transposed fosmid clones
- Transposed fosmids contribute 4.316Mb of sequence
- 2.24Mb of this sequence was contributed to the Broad mouse tiling path by transposed fosmids

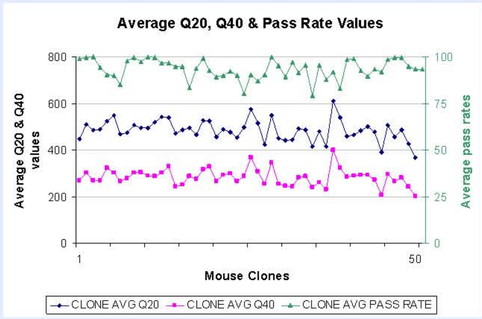


Figure 10. The Q20 and Q40 values for 65k reads are shown here grouped into data points representing 53 fosmid mouse clones. The average Q20 value here is 485, average Q40 value is 284, and average pass rate is 93%.

Future Direction

In the future, TempliPhi amplification may replace our alkaline lysis DNA prep to improve time and efficiency and drastically cut the costs of the amplification and sequencing steps.

References

Goryshin, I. Y. and Reznikoff, W. S. (1998) *J. Biol. Chem.* 273, 7367.
Jilk, R. A., York, D., and Reznikoff, W.S. (1996) *J. Bacteriology.* 178, 1671.