

New!

## An Improved CopyControl™ Fosmid Vector Maximizes End-Sequencing Results

Rob David, Lakeisha Tillery, Nadeem Tusneem, Erik Gustafson and Douglas R. Smith, Agencourt Bioscience Corporation

Agencourt Bioscience operates the largest commercial sequencing facility in the world. We produce approximately 20 million reads annually for commercial clients and also sequence genomes of biomedical interest as a member of the National Human Genome Research Institute (NHGRI) Network for Large Scale Sequencing. One of our current NHGRI-funded projects is aimed at sequencing a number of human genomes using fosmid paired ends to discover single nucleotide polymorphisms (SNPs) and, more significantly, structural variations. Fosmids are used exclusively as the sequencing template for this project due to their large 40 kb insert size capacity and characteristically tight range on insert size distribution.

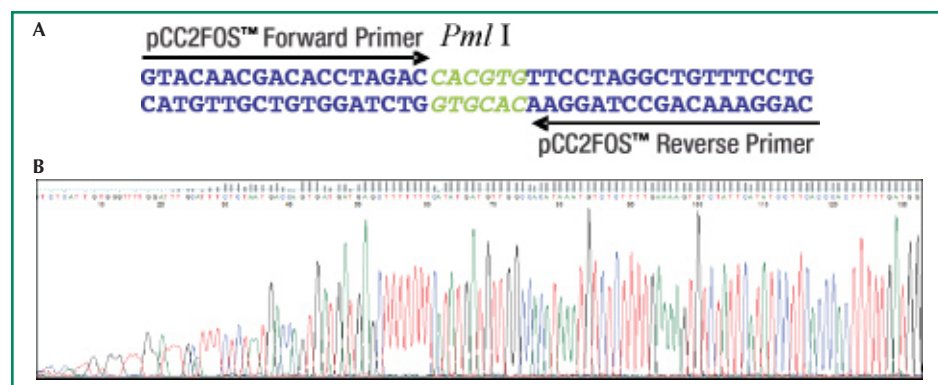
As part of these sequencing endeavors we pay close attention to technological developments that improve sequencing quality while reducing cost. One such development is EPICENTRE Biotechnologies' CopyControl™ Fosmid Library Production Kit. We use this kit to make our fosmid libraries because the kit is easy to use and provides a means to induce the single-copy fosmid vector to multiple copies per cell prior to DNA purification and sequencing. A second such development is our SprintPrep™ purification system where fosmid templates are easily prepared in an economical 384-well format. Two factors, however, diminish our sequencing efficiency.

First, in high throughput production (HTP), there is always some level of *E. coli* genomic DNA contamination present, even with the best DNA preparation methods. Although the contamination level is generally quite low, we noticed that when large numbers of thermal cycles were used for sequencing at low BigDye® Terminator concentrations, spurious products derived from the *E. coli* genome were produced, compromising the sequence quality. Secondly, the 3' termini of the forward and reverse sequencing primers that can be used with the original CopyControl pCC1FOS™ Vector are located 28 to 115 nucleotides from the cloning site, respectively. Utilization of these priming sites for large scale fosmid end-sequencing results in significant waste since up to ~70 megabases of vector-derived sequence must be removed prior to assembling a million reads of insert sequence.

To remedy these inefficiencies, we developed a new version of the CopyControl Fosmid Vector at Agencourt, pCC2FOS, which is being made available in EPICENTRE's new CopyControl HTP Fosmid Library Production Kit. Initially, we designed 5 sets of forward and reverse primers adjacent to the cloning site, each 1 base further upstream than the next. The best performing primers were 3 nucleotides from the cloning site, but their sequencing traces suggested there was a significant level of mispriming

occurring from low levels of contaminating *E. coli* DNA. To eliminate this problem, we needed to incorporate a 7-base sequence at the 3'-end of each primer that could not prime efficiently to the *E. coli* genome. The only 7-mer that doesn't occur at all in the genome (GCCTAGG) is not suitable for this purpose since it would result in primer-dimer formation. After evaluating a number of possibilities, we selected two of the lowest frequency 7-mers with 50% G+C content in the last 4 bases, and which did not have any primer-dimer issues. The two 7-mers are GTCTAGG and CTAGGAA, which occur 3 times and 7 times in the *E. coli* sequence, respectively. Finally, the base adjacent to each 7-mer was chosen such that there were no corresponding octamers with that sequence in the *E. coli* genome (FIG 1A).

Based on the improved sequencing performance we have observed with these primers on pCC2FOS (see FIG 1B) and a similarly modified high-copy-number plasmid vector (from approximately 3 million and 8 million attempted reads, respectively), we strongly recommend their use by other large-scale sequencing groups. We have therefore freely disclosed these primer sequences for research or commercial use.



**FIG 1. A.** The CopyControl™ pCC2FOS™ Vector differs from the pCC1FOS™ Vector by the insertion of a new primer cassette that eliminates wasteful vector-derived sequencing reads and minimizes the potential for priming on the *E. coli* genome. **B.** Typical sequencing results obtained with the pCC2FOS Forward Primer on a pCC2FOS clone at 1/48x BigDye dilution. Similar results were obtained with the pCC2FOS Reverse Primer (data not shown).

[www.EpiBio.com/ccfos.asp](http://www.EpiBio.com/ccfos.asp)

### CopyControl™ HTP Fosmid Library Production Kit

CCFOS059 1 Kit

For producing up to 10 complete and unbiased fosmid libraries.

Contents: CopyControl™ pCC2FOS™ Vector, End-Repair Enzyme Mix, End-Repair 10X Buffer, dNTP Mix, Fast-Link™ DNA Ligase, Fast-Link™ 10X Ligation Buffer, ATP Solution, GELase™ Gel-Digesting Preparation, GELase™ 50X Reaction Buffer, MaxPlax™ Lambda Packaging Extracts, Control DNA, Ligated Lambda Control DNA, TransforMax™ EPI300™-T1R Plating Strain, Control Lambda Plating Strain, and CopyControl™ Induction Solution. (Additional quantities of most of these reagents are also available separately.)