Random-Shear BAC Library Construction and Efficient Genome Gap Closing

Chengcang Wu, Sarah Vande Zande, Rebecca Hochstein, ***David A. Mead***, Ronald Godiska
Lucigen Corporation, Middleton, WI 53562

Abstract

Bacterial artificial chromosome (BAC) libraries are indispensable for physical mapping, positional cloning, genetic analysis, and sequencing of large genomes. BAC libraries have been created from many species, including *Arabidopsis, Drosophila*, rice, mouse, and human. A significant limitation of the current methods is the use of partial restriction digestion to generate genomic DNA fragments of 100-300 kb. The inherently skewed genome distribution of restriction sites causes at least 10-fold under- or over-representation of particular sequences, with some regions being entirely absent from the BAC libraries. Another drawback is the instability of inserts in current cloning vectors due to transcription and secondary structure formation. As a result, existing BAC libraries built with conventional methods and vectors are biased, and numerous gaps exist in all of the physical and sequencing maps of eukaryotic multi-cellulargenomes. To circumvent these problems we have successfully developed techniques to construct unbiased, randomly-sheared BAC libraries (>100 kb inserts). We have demonstrated that a single 5X random shear BAC library covers various genome regions uniformly and closes several gaps in the *Arabidopsis thaliana* genome. We believe it will be possible to finish the physical mapping and sequencing of *Drosophila, Arabidopsis,* rice, mouse, and human with this approach, closing all of the existing genomic gaps, including centromeres. We have also developed transcription-free BAC vectors. These vectors show much higher stability of inserts containing AT-rich sequences, direct and inverted repeats, and other deleterious DNAs. It is thus possible for the first time to construct unbiased BAC libraries to achieve complete closure of a large complex genome.

Background

Current BAC libraries are constructed from partial digestions of genomic DNA, cloned into a BAC vector. However, despite using multiple libraries, many gaps remain in all genomes studied (below). These gaps include but are not limited to repetitive DNA and centromeric regions.

Table 1. Gaps in Whole Genome Physical or Sequencing Maps

Species	Ref.	Genome Size (Mb)	# Libraries (coverage)	Contigs (chr. no.)	Genome Gaps
Plants					
Arabidopsis	1	125	Two (17x)	27 (5)	< 5%
Rice	2	430	Two (26x)	284 (12)	< 10%
Soybean	3	1,115	Three (10x)	2,905 (20)	~ 10%
Maize	4	2,500	Three (15x)	3,488 (10)	unknown
Animals					
Fruit Fly	5	97	One (14x)	9 (2)*	> 2%
Human	6	3,200	Five (15x)	246 (23)	~ 4%
Mouse	7	3,200	Two (33x)	296 (20)	~10%

References: [1]Mozo (1999) ; [2]Chen (2001); [3]Wu (2004); [4]www. genome.arizona.edu;[5]Hoskins (2000); [6]HGMC (2001); [7]Gregory(2002).
* *Drosophila* physical maps of chromosome 2, 3.

BAC Optimizied Electrocompetent Cells

Lucigen has long been a leader in providing electrocompetent cells of the highest cloning efficiency. We have now successfully developed the first-ever BAC-optimized electrocompetent cells. Manufactured with proprietary protocols, Lucigen's BAC Optimized *E. cloni* cells have the highest transformation efficiency available for BAC cloning (Figure 2).

BAC and Large Insert Cloning:Two- to Six-Fold More Recombininants with *E. cloni* BAC-Optimized Electrocompetent Cells

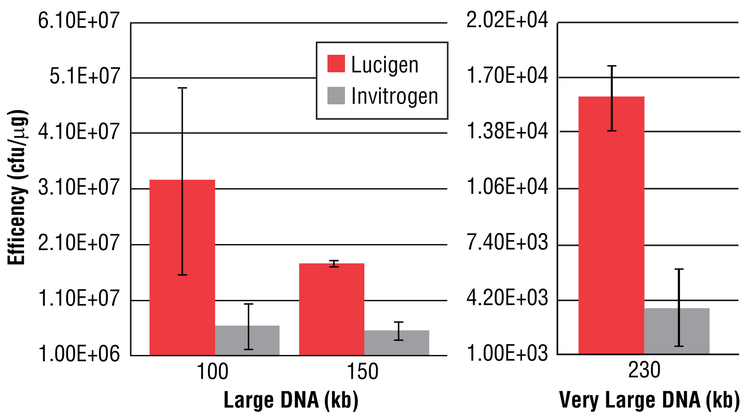

Figure 2. BAC clones were transformed into Lucigen's *E.cloni* BAC Optimized cells or competitors' strains. Values are relative to the transformation efficiency of large-insert BACs.

Minimizing Genomic Gaps

One major source of gaps in genomic libraries is non-random distribution of restriction sites in certain regions (e.g., centromeres), causing them to be absent or under-represented in the library. In addition, some regions are unstable in typical cloning vectors, resulting in their deletion.

Lucigen has developed improved cloning vectors and a method of randomly shearing genomic DNA to eliminate this cloning bias in BAC libraries.

Unbiased, low-background vectors

The new pSMART BAC v2.0 vector incorporates CloneSmart transcription-free technology to increase the stability of cloned inserts. In addition, a unique system selects against non-recombinant clones (Figure 1). However, unlike all other BAC vectors, this vector does NOT induce high-level expression of insert DNA, further increasing stability of recombinant clones.

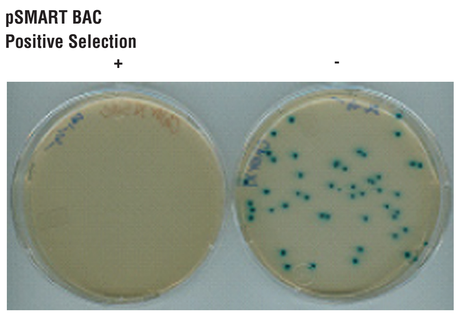

Figure 1. The pSMART BAC V2.0 vector was self-ligated, transformed into *E. cloni* cells, and plated on chloramphenicol plates with or without positive selection. Background clones are absent with positive selection.

Reduced bias in pSMART BAC vector

Tetrahymena genomic DNA (75% AT) is very difficult to clone. Fragments as small as 4-6 kb are often deleted when cloned into standard *E. coli* vectors. However, 8-20 kb inserts were stably cloned in the transcription-free pSMART BAC vector (Figure 3).

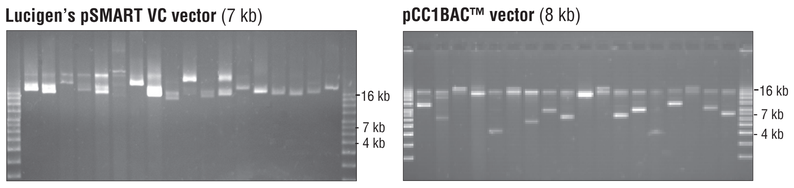

Figure 3. *Tetrahymena* genomic DNA was sheared to 6-20 kb, end-repaired, and cloned into the pSMART-BAC vector or a standard blue/white BAC vector. Uncut DNA from the pSMART BAC clones was >14 kb, indicating that inserts were > 8 kb (left). In contrast, many clones from the standard vector underwent substantial deletions, and were actually smaller than the parental vector (right).

>200 kb Partial-digest Libraries

To construct BAC libraries with inserts of >200 kb, Lucigen has optimized all aspects of library construction, including preparation of cells, vector, and HMW insert DNA. The result is complex libraries with majority insert sizes over 200 kb (Fig. 5).

Miniprep DNA was isolated from random clones, digested with Not I to excise the inserts, and fractionated on a pulse field gel (PFGE). The average insert size in this library was 212 kb (Fig 5).

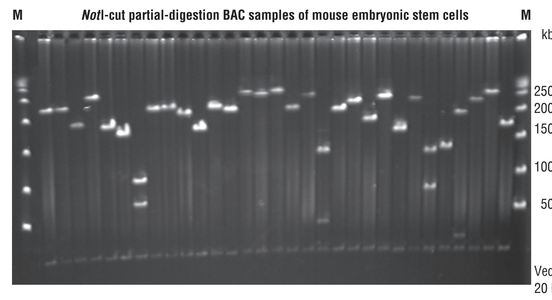

Figure 5. Mouse genomic DNA was partially digested with HindIII, size fractionated, and cloned into a BAC vector. DNA from random BAC clones was digested with NotI to release the inserts (average size = 212 kb). Lanes 1- 31, BAC clones; M, Lambda ladder size markers.

Random Shearing of Genomic DNA

Megabase regions of genomic DNA, such as centromeres, may completely lack recognition sites for common restriction enzymes (e.g., BamHI, EcoRI, HindIII; Figure 4, left).

Lucigen has developed methods to randomly shear genomic DNA into fragments of 100-400 kb. Significantly, the DNA from all genomic regions is sheared (Figure 4, right), which allows it to be cloned into BAC vectors.

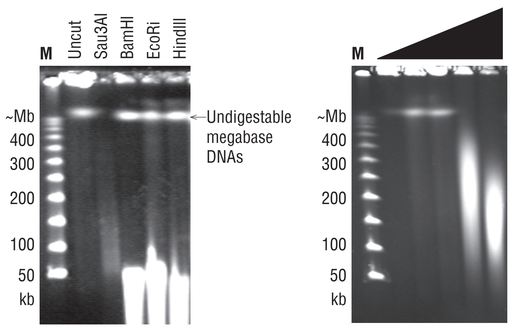

Figure 4. Mouse genomic DNA was over-digested by several restriction enzymes (Left panel) or fragmented by random shearing (Right panel). Lanes: 1, Uncut; 2, Sau3A; 3. BamHI; 4, HindIII; 5, EcoRI. Only Sau3A digested the band at ~1 Mb. In contrast, all the DNA was reduced to 50-500 kb, as the degree of random shearing was increased.

>100 Kb Random Shear Libraries

A random shear BAC library was constructed with potato genomic DNA and cloned into the pSMART BAC vector. The average insert size was > 100 kb. (Fig.6)

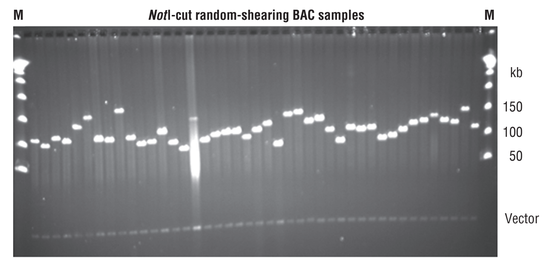

Figure 6. Genomic DNA was isolated from potato tissue, randomly sheared, size-selcted to >100 kb, and cloned into the pSMART BAC vector. DNA from minipreps was digested with *Not*I to excise inserts. The vector band is visible at 7 kb.

Unbiased Cloning in Random Shear Libraries

The "complete" BAC library of the *Arabidopsis* genome contains numerous regions that are under- or over-represented (Fig. 7, black bar graph). To show the unbiased distribution of clones in a random shear BAC library, *Arabidopsis* genomic DNA was randomly sheared, size-selected, and cloned into the pSMART BAC vector. A 5X coverage library was screened with overgo oligonucleotide probes specific for various regions of Chromosome 1. Significantly, clone coverage across all the probed regions, ***including the centromeric region***, were observed to a similar extent in the random shear library (Figure 7). In contrast, these regions show vastly different representation in the *Arabidopsis* genome project (15, 75, or <1 clone per 0.1 Mb, respectively; 17X coverage overall). Most importantly, we have been able to close existing centronomic gaps of this "finished" physical and sequence genomic map. The same probes also identified clones covering centromeric regions of other chromosomes.

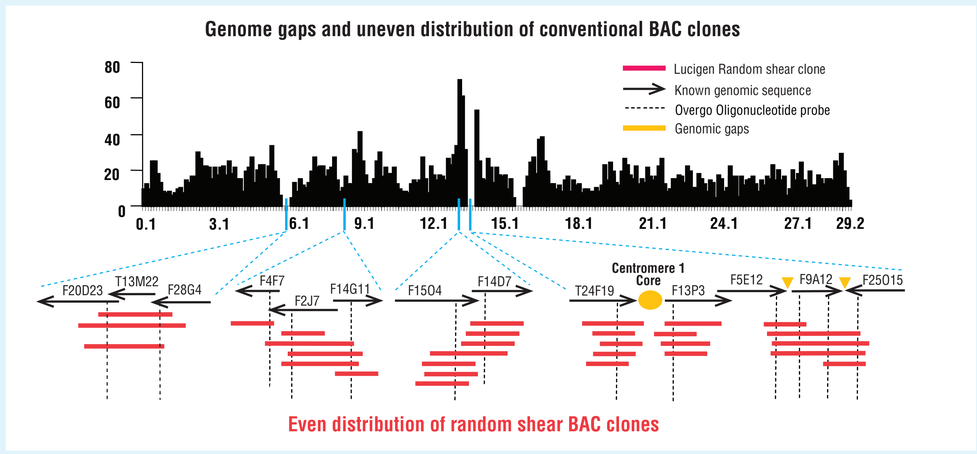

Figure 7. The distribution of BAC clones from Chromosome 1 of the *Arabidopsis* genome project is shown in the bar graph (Mozo, 1999). Overgo oligonucleotide probes were used to screen Lucigen's random shear library. The coverage of Lucigen clones is uniform over all regions tested. Several clone gaps were covered with this libary, including centromeric regions. Additional sequencing is underway.

Summary

- **Improved vector for BAC libraries.** A transcription-free vector provides more stable cloning with very low background.
- **BAC Optimized competent cells.** Electrocompetent cells prepared specifically for BAC cloning offer the highest possible transformation efficiency.
- **Random shearing for reduced bias.** Random shear libraries are a powerful tool for closing genomic gaps, including centromeric regions.
- **bSMART libraries.** Lucigen offers custom BAC library using standard partial restriction digestion or random shearing. These techniques provide unparalleled performance for obtaining large, unbiased BAC libraries.

Lucigen® Advanced Products for Molecular Biology

Lucigen Corporation 2120 W. Greenview Dr., Ste 9 Middleton, WI 53562 www.lucigen.com