

Random Shear BAC Libraries for Efficient Genomic Cloning

Ronald Godiska, Rosa Ye, Sarah Vande Zande, Rebecca Hochstein, David Mead, **Cheng-Cang Wu**

Lucigen Corporation, Middleton, WI 53562

Abstract

Bacterial artificial chromosome (BAC) libraries, BAC-based physical maps and whole genome sequencing maps have been under development for many genomes of the most important eukaryotic multi-cellular species, including *Arabidopsis*, *Drosophila*, rice, mouse, and human. However, DNA libraries built with conventional vectors and methods are biased, and genome gaps exist in all of the physical and sequencing maps. To optimize BAC cloning, physical mapping, and genome sequencing we have developed an optimized BAC cloning system. The CopyRight™ pSMART BAC 2.0 consists of the pSMART-BAC 2.0 vectors and the BAC-optimized Replicator™ 2.0 competent cells. This BAC vector lacks an indicator gene and associated promoter, has termination signals on either side of the insertion site, and simplifies BAC preparation by selectively using single or multi-copy number replicons. The vector shows much higher stability of inserts containing AT-rich sequences, direct and inverted repeats, and other deleterious DNAs, thus making it possible to construct unbiased BAC libraries. The CopyRight pSMART BAC 2.0 vector and the Replicator 2.0 cells also feature inducible amplification of copy number, increasing yields to as many as 50 copies per cell. The amplification is more robust than the similar CopyControl system and permits easy isolation of BAC DNA for sequencing, subcloning, or restriction mapping. To overcome the other gaps introduced by partial restriction digestion, we have successfully developed techniques to construct unbiased, randomly-sheared BAC libraries with large inserts (>100 kb). With these novel techniques and tools, we have cloned previously unclonable DNA and are working toward covering the entire genomes of many plants and animal species. We offer the Random Shear BAC Libraries as a custom service.

Background

Current BAC libraries are constructed from partial digestions of genomic DNA. 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					
<i>Arabidopsis</i>	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: *Mozo (1999); †Chen (2001); ‡Wu (2004); ‡www.genome.arizona.edu; †Hoskins (2000); †HGMC (2001); †Gregory(2002).

**Drosophila* physical maps of chromosome 2, 3.

BAC Optimized 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. coli* cells have the highest transformation efficiency available for BAC cloning (Figure 1).

Two- to Six-Fold More Recombinants with *E. coli* BAC-Optimized Electrocompetent Cells

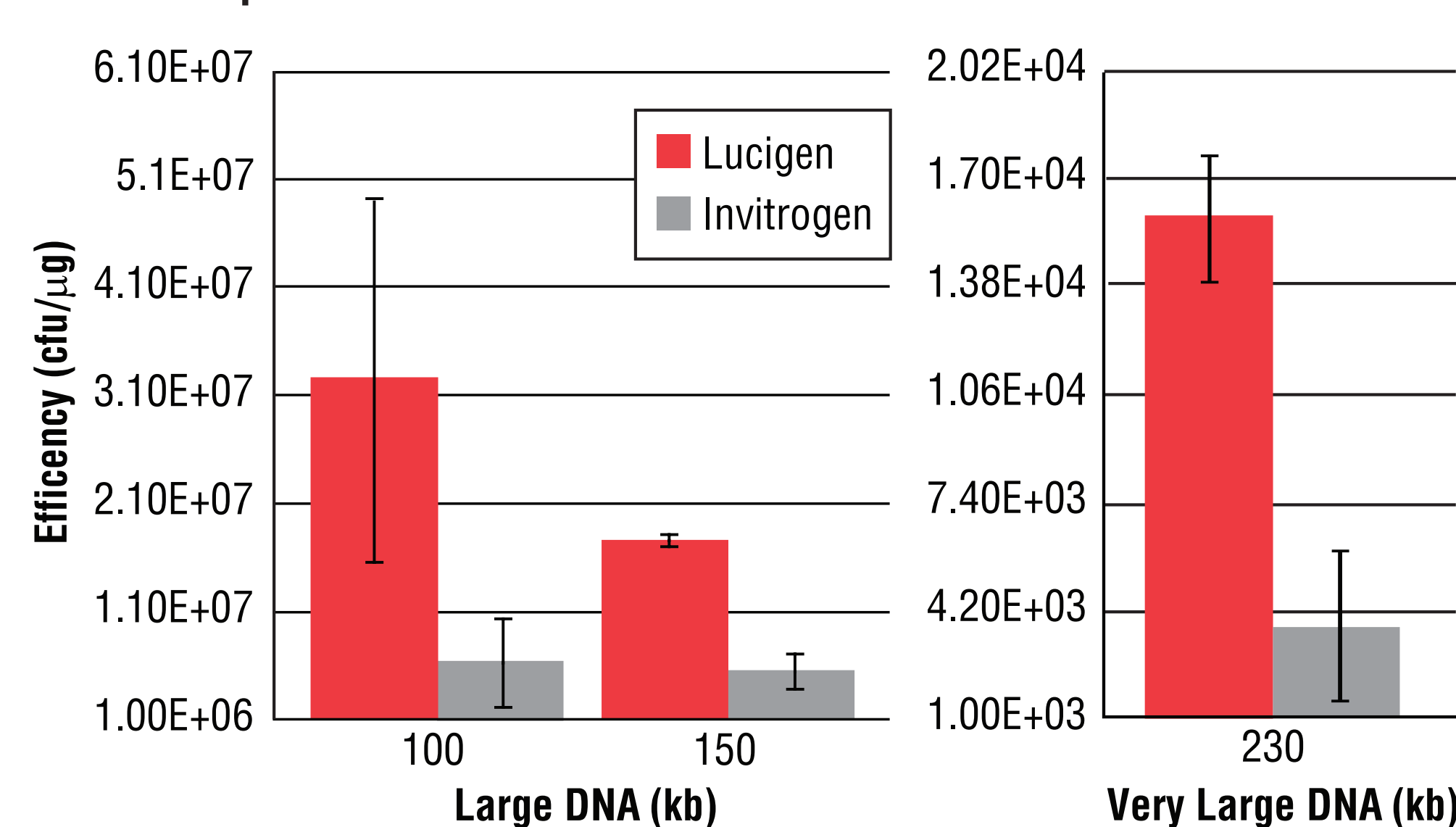


Figure 1. BAC clones were transformed into Lucigen's *E. coli* BAC Optimized cells or alternative strains.

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 2, 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 2, right), which allows it to be cloned into BAC vectors.

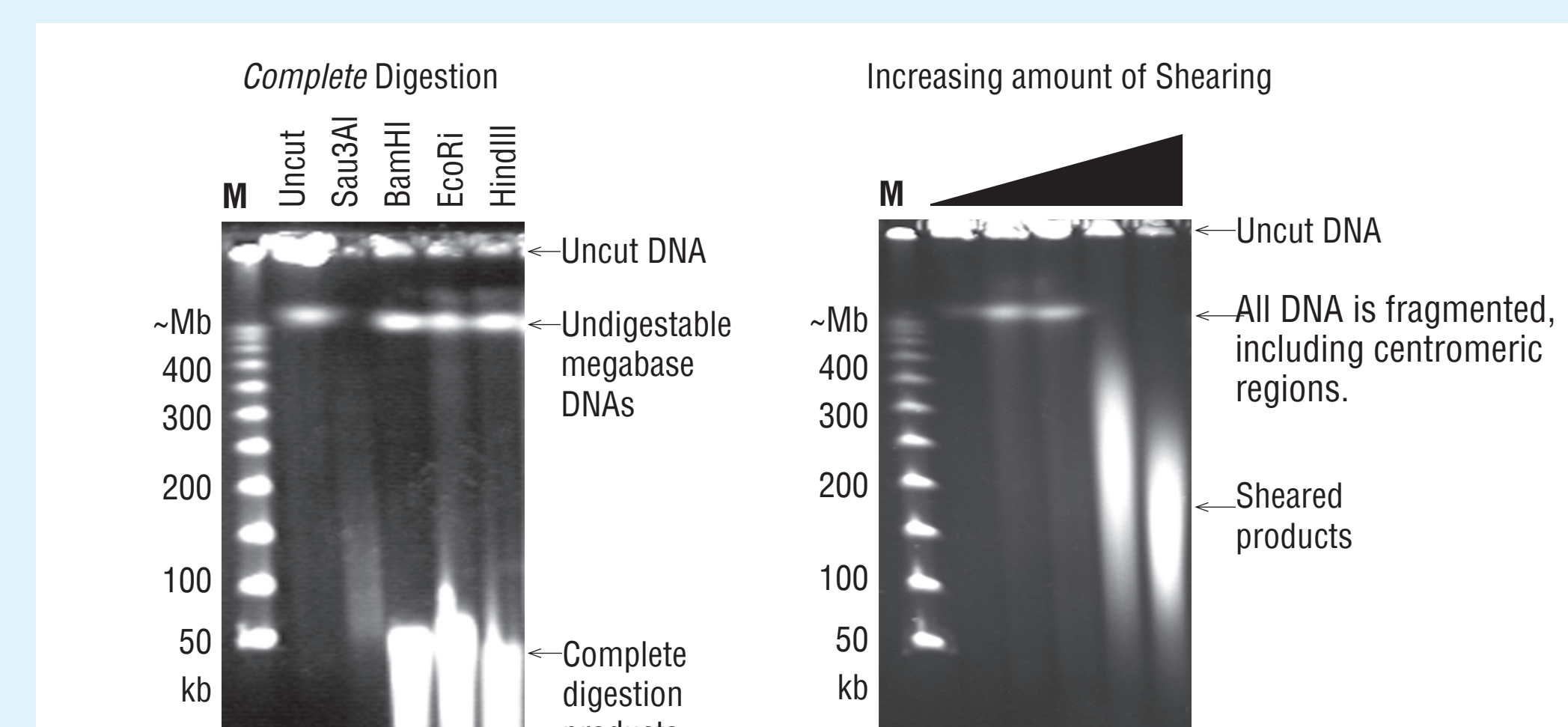


Figure 2. Mouse genomic DNA was over-digested by several restriction enzymes (Left panel) or fragmented by random shearing (Right panel). Lanes: 1, Uncut; 2, Sau3A; 3, BamHI; 4, HindIII; 5, EcoRI. Only Sau3A digested the band at ~1 Mb. In contrast, all the DNA was reduced to 50-500 kb random shearing.

Unbiased Cloning in Random Shear Libraries

The "complete" BAC library of the *Arabidopsis* genome contains numerous regions that are under- or over-represented (Fig. 4, 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, was similar in the random shear library (Figure 4). 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 centromeric gaps of this "finished" physical and sequence genomic map. The same probes also identified clones covering centromeric regions of other chromosomes.

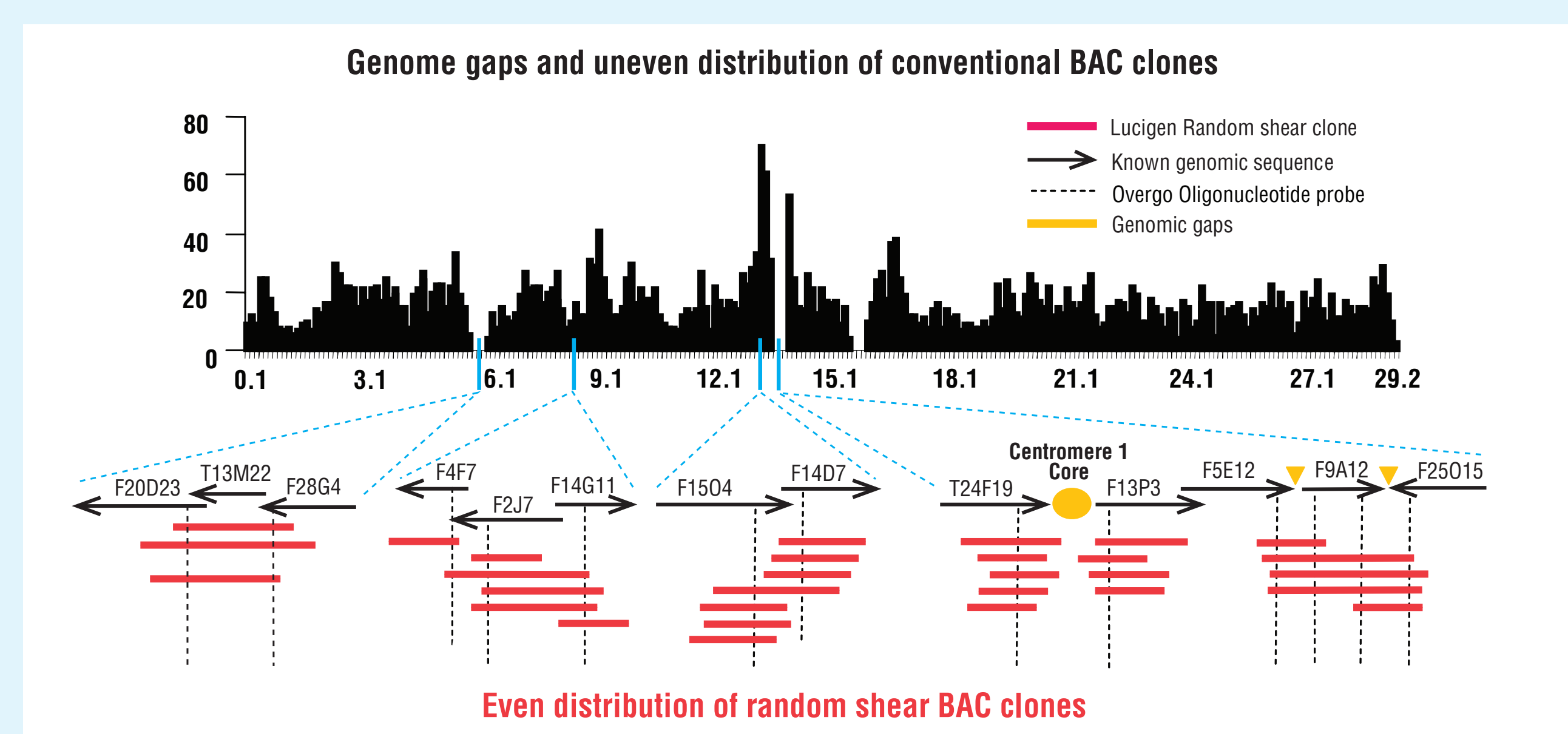


Figure 4. The distribution of BAC clones from Chromosome 1 of the *Arabidopsis* genome project is shown in the bar graph (Mozo, 1999). Overgo oligonucleotide probes were used to screen Lucigen's random shear library. The coverage of Lucigen clones is uniform over all regions tested. Several clone gaps were covered with this library, including centromeric regions.

>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.3)



Figure 3. Genomic DNA was isolated from potato tissue, randomly sheared, size-selected to >100 kb, and cloned into the pSMART BAC vector. DNA from minipreps was digested with *NotI* to excise inserts. The vector band is visible at 7 kb.

Lucigen

Advanced Products for Molecular Biology

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

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 5). However, unlike all other BAC vectors, this vector does NOT induce high-level expression of insert DNA, further increasing stability of recombinant clones.

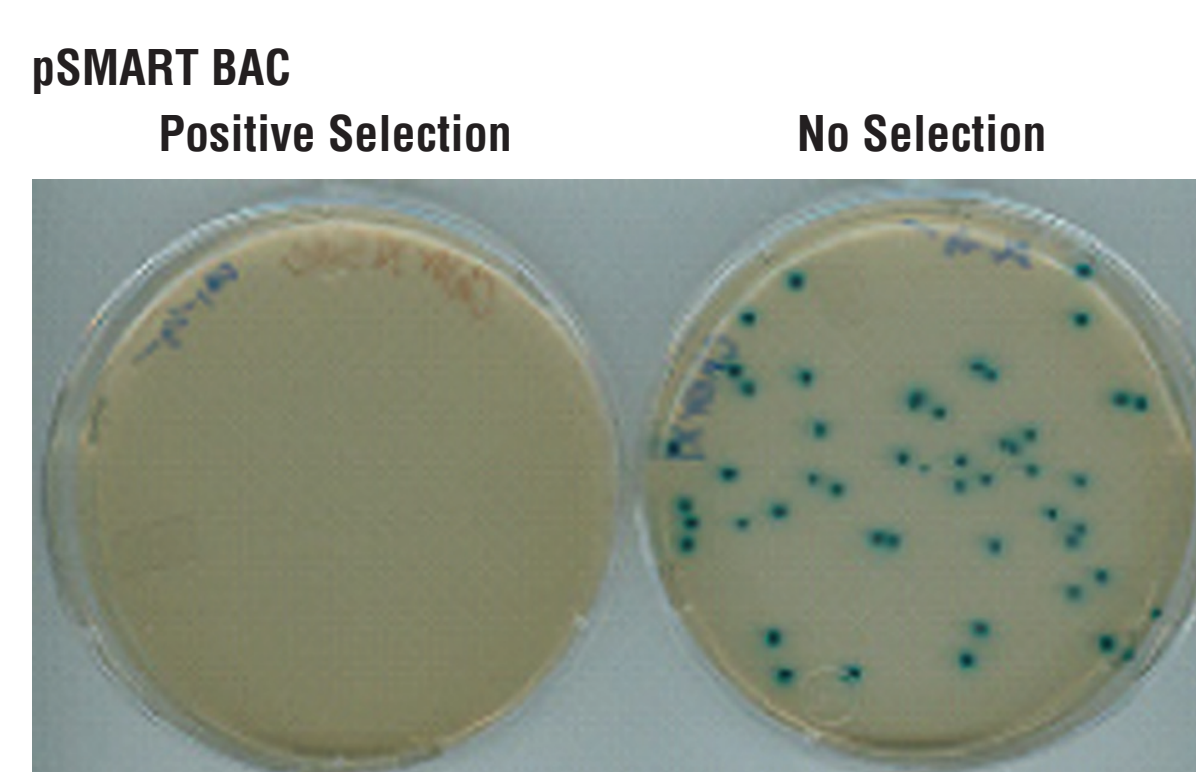


Figure 5. The pSMART BAC V2.0 vector was self-ligated, transformed into *E. coli* cells, and plated on chloramphenicol plates with or without positive selection. Background clones are absent with positive selection.

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 6).

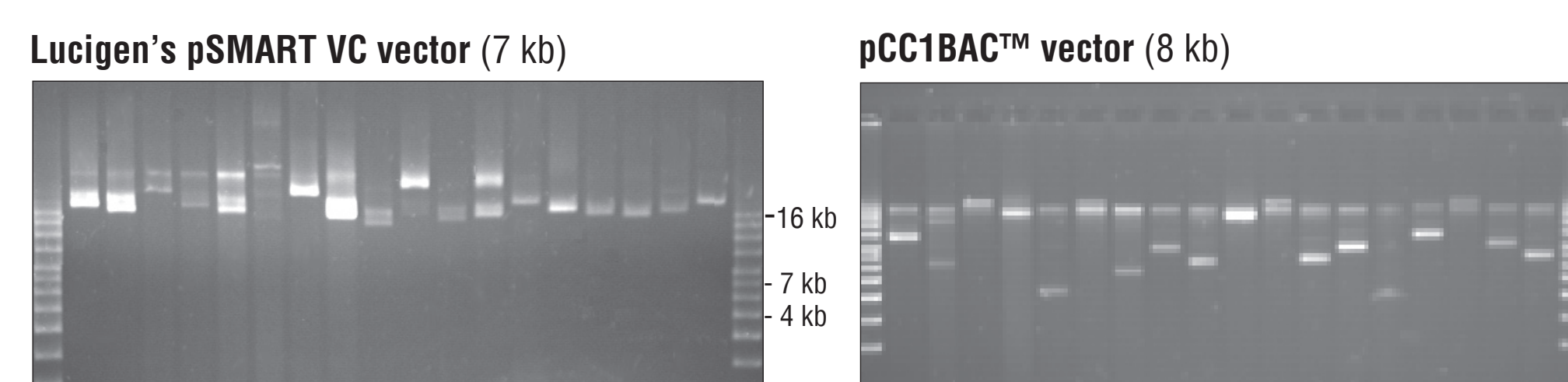


Figure 6. *Tetrahymena* genomic DNA was sheared to 8-20 kb, end-repaired, and cloned into the pSMART-BAC vector or a standard blue/white BAC vector. Uncut DNA from the pSMART BAC clones was >14 kb, indicating that inserts were > 8 kb (left). In contrast, many clones from the standard vector underwent 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 insert sizes over 200 kb (Fig. 7).

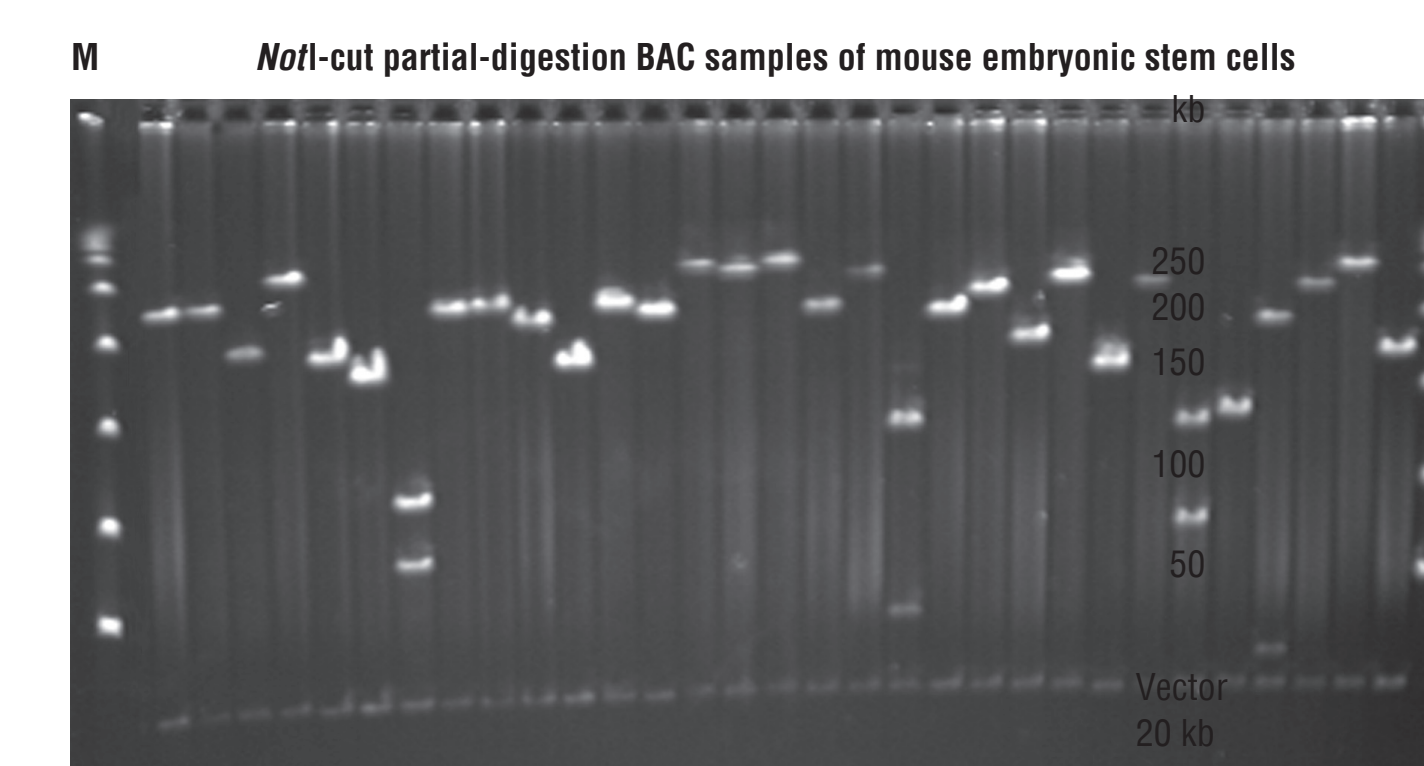


Figure 7. Mouse genomic DNA was partially digested with *HindIII*, size fractionated, and cloned into a BAC vector. DNA from random BAC clones was digested with *NotI* to release the inserts (average size = 212 kb). Lanes 1-31, BAC clones; M, Lambda ladder size markers.

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.
- We have successfully constructed random shear BAC libraries** from 8 plants and 5 animal species, including the important models: *Medicago truncatula*, *Xenopus Tropicalis*, and mouse. More importantly, we have constructed random shear BAC libraries of genomic DNA from crustacean species, which is impossible to digest by restriction enzymes.
- 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.