New Lucigen Custom Service:

Random Shear BAC Library Construction

The quality of genomic BAC libraries depends greatly on the cloning methods and vectors used. The bias introduced by partial restriction digestion of genomic DNA results in regions being highly under-represented, over-represented, and missing portions for all eukaryotic multi-cellular genomes studied, including *Arabidopsis, Drosophila*, rice, mouse, and human. This bias results in numerous clone gaps that can be impossible to close, even with multiple complementary libraries and up to 40X coverage, thus dramatically increasing the finishing costs.

To circumvent these problems, Lucigen has successfully developed unique Random Shear BAC Library construction techniques to efficiently construct unbiased, full coverage BAC libraries with large inserts (≥100 kb average). With these new techniques and tools, we have demonstrated that a Random Shear BAC Library of the *Arabidopsis* genome is unbiased. Importantly, with just 5X coverage we were able to close several clone gaps and fill existing centromeric gaps in the *Arabidopsis thaliana* genome. We have also developed transcription-free BAC cloning vectors to alleviate additional instability problems.

Random shearing of all genomic DNA, including undigestible DNA

Certain regions of genomic DNA, such as centromeres, highly repetitive sequences and telomeres, often contain megabase regions that completely lack recognition sites for common restriction enzymes (e.g., BamHI, EcoRI, and HindIII; Figure 1, upper panel). Lucigen has developed methods to randomly shear genomic DNA into fragments of 100-400 kb to successfully clone this "undigestible" DNA. Significantly, the DNA from all genomic regions is sheared (Figure 1, lower panel), which eliminates a major form of bias inherent to conventional methods.

>100 kb inserts in a Random Shear BAC Library of *Arabidopsis*

A Random Shear BAC Library of *Arabidopsis* genomic DNA was constructed and cloned into the transcription-free pSMART® BAC vector. The average insert size was > 100 kb (Figure 2).

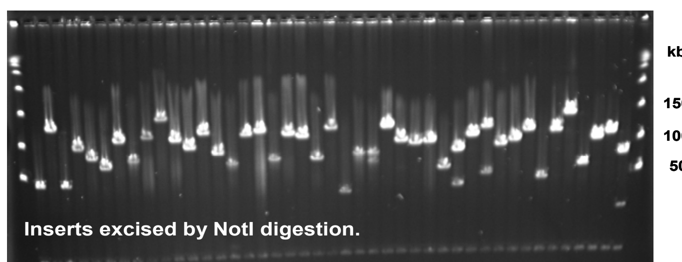

Figure 2. Genomic DNA was isolated from *Arabidopsis* tissue, randomly sheared, size-selected to >100 kb, and cloned into the pSMART® BAC vector. DNA from minipreps was digested with NotI to excise inserts. The vector band is visible at 7 kb.

Unbiased cloning in Random Shear Libraries

Numerous regions of the "finished" BAC library of the *Arabidopsis* genome are under-represented, over-represented, or completely missing (Figure 3, black bar graph, next page). 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 (overlapping oligonucleotide) probes specific for various regions of Chromosome 1. Significantly, clone coverage was uniform across all the probed regions, including the centromeric region, in the Random Shear Library (Figure 3, red bars, next page). 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 several existing centromeric gaps of this "finished" physical and sequence genomic map. The same probes also identified clones covering centromeric regions of other chromosomes. These results demonstrate that Random Shear BAC Libraries may offer the only way to finish existing genome sequencing projects.

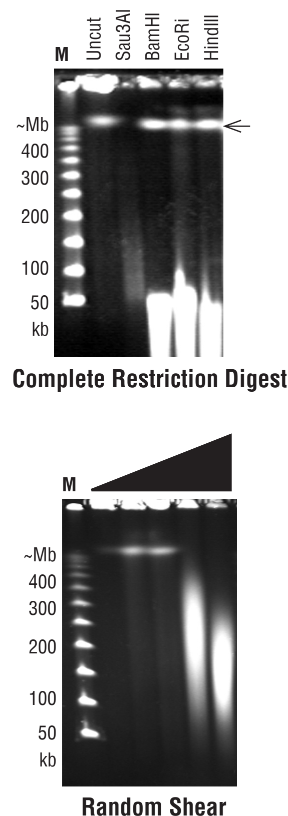

Figure 1. Mouse genomic DNA digested to completion by several restriction enzymes (upper panel) or fragmented by random shearing (lower panel). Lanes: M, Marker; U, Uncut; S, Sau3A; B. BamHI; H, HindIII; R, EcoRI. Only Sau3A digested the band at ~1 Mb. In contrast, all of the DNA was reduced to 100-400 kb as the degree of random shearing was increased. The arrow indicates megabase size, undigested DNA.

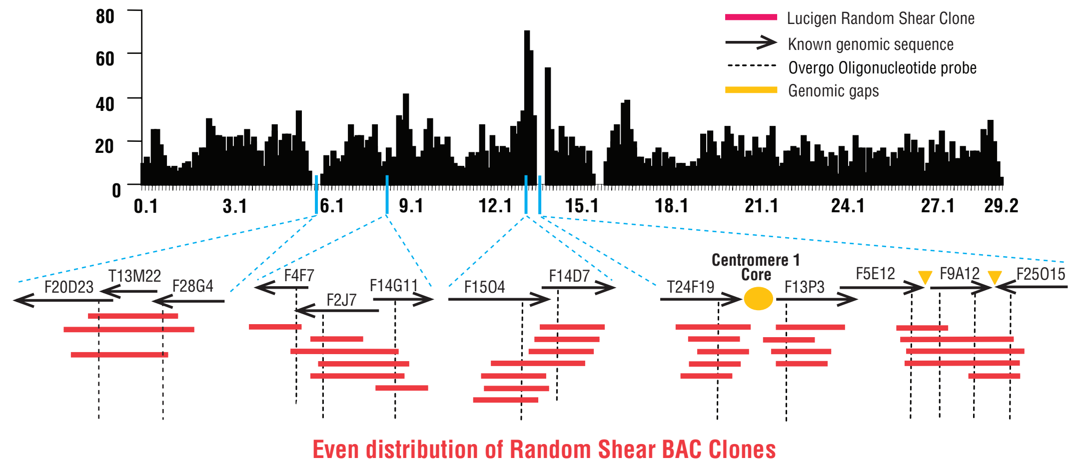

Figure 3. The distribution of BAC clones from Chromosome 1 of the *Arabidopsis* genome project as shown in the black bar graph (1). Overgo oligonucleotide probes were used to screen Lucigen's Random Shear BAC Library. The coverage of Lucigen clones is uniform over all regions tested (red bar graph). Several clone gaps were closed with this library, including centromeric regions. Additional sequencing is underway.

Random Shear BAC Libraries also are the ideal choice for initial library construction in new genome projects. Due to the unbiased coverage, only a single Random Shear Library is needed. More significantly, the finishing process is greatly simplified because of the reduced number of gaps. Random Shear BAC Libraries provide superior data with less time, effort, and expense (Table 1).

Table 1. Comparison of Random Shear and partial digestion BAC libraries

BAC library	Character	Number of Libraries Needed (coverage)	Cost of Library Construction ($US) *	Finishing cost ($US) **
Random Shear	Unbiased; No gaps	1 (10X)	Up to $30,000	Up to $1 million
Partial Digestion	Biased; gaps	≥2 (>20X)	≥ $60,000	≥$2 million

* Construction cost is based on one Random Shear BAC Library (10x) for a genome of ~500 Mb; and at least two complementary partial-digest BAC libraries (20x total), to minimize restriction site bias.

** Finishing cost includes BAC end sequencing, whole genome physical mapping, and integrating the physical map with about 1000 genetic makers.

The CopyRight® BAC Cloning System simplifies BAC preparation and sequencing

Lucigen's CopyRight BAC Cloning System consists of the pSMART® BAC vector and new Replicator™ v2.0 Electrocompetent Cells (2). The pSMART BAC vector contains both the single-copy origin of replication from the *E. coli* F-factor and the inducible high-copy *oriV* origin of replication. Replicator v2.0 Cells are an *E. coli* strain that contains the *trfA* gene for initiation of replication from *oriV*. The *trfA* gene is under tight control of the inducible *araBAD* promoter. Upon addition of the supplied Induction Solution, CopyRight BAC clones replicate to ~10-20 copies/cell and are stably maintained, providing high yields of high purity DNA. No stability problem is observed (Figure 4). This feature simplifies BAC DNA preparation and increases the success rate of BAC end sequencing to >95%.

Lucigen's BAC library construction team is led by Dr. Chengcang Wu, who has more than ten years of experience in large-insert DNA and BAC cloning at Pioneer Hi-Bred International, Inc.; Texas A&M University; and the National Institute of Sericultural and Entomological Sciences (NISES, now the National Institute of Agrobiological Sciences, NIAS), Japan. He has published peer-reviewed papers in this area and has co-authored several reviews about BAC cloning and related research (3-10).

The BAC team at Lucigen has constructed hundreds of custom BAC libraries for many satisfied clients. The novel Random Shear BAC Libraries are available only from Lucigen.

Contact Lucigen for a free quote on a Random Shear BAC or Fosmid library.

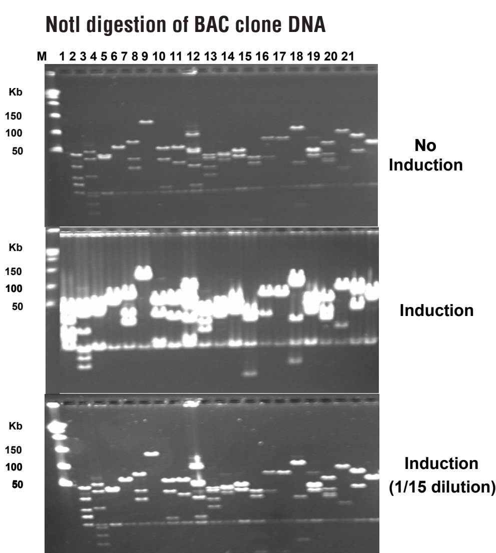

Figure 4. Amplification of large insert CopyRight BACs. BAC DNA was prepared according to a standard BAC mini-prep protocol, and 10 µl of each clone was digested with NotI. Insert sizes are 50 to ~200 kb.

M, lambda ladder DNA marker. Upper panel: BAC samples without induction; middle panel: BAC samples with induction; lower panel: BAC samples with induction, diluted 15-fold.

References:

1. Mozo, T., Dewar, K., Dunn, P., Ecker, J. R. and Fischer, S., et al. (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22**, 271.
2. Wild, J., Hradecna, Z., Szybalski, W. (2002). Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* **12,** 1434.
3. Wu, C. C., Asakawa, S., Shimizu, N., Kawasaki, S., and Yasukochi, Y. (1999). Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*. *Mol. Gen. Genet.* **261,** 698.
4. Wu, C. C., Santos, F. A., Nimmakayala, P., Springman, R., Meksem, K., Lightfoot, D., and Zhang, H. B. (2004). Construction of a complementary BAC library for soybean genome-wide physical mapping and whole genome sequencing. *Theor. Appl. Genet.* **109**, 1041.
5. Wu, C. C., Xu, Z. and Zhang, H. B. (2004). DNA libraries. In, *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, **Volume 3** (2nd Edition), pp385-425. Edited by Meyers R. A., Wiley-VCH Verlag GmbH: Weinheim, Germany.
6. Zhang, H. B. and Wu, C. C. (2001). BAC as tools for genome sequencing. *Plant Physiol. Biochem.* **39**, 1.
7. Ren, C., Sun, S., Lee, M. K., Xu, Z., Wu, C. C. and Zhang, H. B. (2005). Chapter 4. Genomic libraries and physical mapping (YACs and BACs). In, *Handbook of Plant Genome Mapping*, pp173-214. Edited by Meksem, K. and Gresshoff, P., Wiley-VCH Verlag GmbH: Weinheim, Germany.
8. Stupar, R. M., Beaubien, K. A., Jin, W., Song, J., Lee, M. K., Wu, C. C., Zhang, H. B., Han, B., and Jiang, J. (2006). Structural diversity and differential transcription of the patatin multicopy gene family during potato tuber development. *Genetics* **172,**1263.
9. Zhang, X., Scheuring, C., Tripathy, S., Xu, Z., Wu, C. C., Ko, A., Tian S. K., Arredond, F., Lee, M. K., Santos, A. F., Zhang, H. B., and Tyler, B. M. (2006). An integrated BAC and genome sequence physical map of *Phytophthora sojae. Molecular Plant-Microbe Interactions* **12,**1302.
10. Broggini, G. A. L., Le Cam, B., Parisi, L., Wu, C. C., Zhang, H. B., Gessler, C. and Patocchi, A. (2006). Construction of a contig of BAC clones spanning the region of the apple scab avirulence gene *AvrVg. Fungal Genetics and Biology*, (in press).