

## Research Article

# Construction and Characterization of a Bacterial Artificial Chromosome Library for the A-Genome of Cotton (*G. arboreum* L.)

Yan Hu, Yamin Lu, Dan Ma, Wangzhen Guo, and Tianzhen Zhang

National Key Laboratory of Crop Genetics & Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

Correspondence should be addressed to Tianzhen Zhang, cotton@njau.edu.cn

Received 16 March 2010; Revised 17 June 2010; Accepted 29 July 2010

Academic Editor: James Birchler

Copyright © 2011 Yan Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A bacterial artificial chromosome (BAC) library for the A-genome of cotton has been constructed from the leaves of *G. arboreum* L. cv. Jianglinzhongmian. It is used as elite A-genome germplasm resources in the present cotton breeding program and has been used to build a genetic reference map of cotton. The BAC library consists of 123,648 clones stored in 322 384-well plates. Statistical analysis of a set of 103 randomly selected BAC clones indicated that each clone has an average insert length of 100.2 kb per plasmid, with a range of 30 to 190 kb. Theoretically, this represents 7.2 haploid genome equivalents based on an A-genome size of 1697 Mb. The BAC library has been arranged in column pools and superpools allowing screening with various PCR-based markers. In the future, the A-genome cotton BAC library will serve as both a giant gene resource and a valuable tool for map-based gene isolation, physical mapping and comparative genome analysis.

## 1. Introduction

Cotton, one of the most important economical crops in the world, provides most natural textile fiber and some of oil resources for people. Cotton cultivars are diploids and tetraploids derived from natural hybridization of an African/Asian A-genome and an American D-genome species [1]. As one of the ancestors of tetraploids, the Asiatic cotton (*G. arboreum* L.) has been domesticated and cultivated for almost 2000 years in China since it was first introduced from India. It is still used as germplasm resources in the present cotton breeding program due to their early maturity, strong tolerance to stress, resistance to disease and insects, high fiber strength, and excellent plasticity. So decoding the A-genome will help for the digging of potential gene in Asian cotton, understanding the origin and evolution of allpolyploid, studying the genome structure of important agronomic trait and the relationship of gene structure and function, and elucidating the influence to the trait by the interaction between genes, which is important in crop improvement.

The bacterial artificial chromosome (BAC) cloning system has become an invaluable tool in many research areas of genome research because of its ability to stably maintain large DNA fragments and its ease of manipulation [2–4]. BAC libraries have been developed for a number of major crops including rice [5–7], soybean [8, 9], wheat [10–13], maize [14, 15], sorghum [16], and tomato [17]. For cotton, at least eight BAC/BIBAC libraries have been reported and made available to the public. These libraries were all made from different genotypes of AD-genome allotetraploid *Gossypium* species. Seven cotton libraries were made from upland cotton species, including Tamcot HQ95, Auburn 623 (<http://hbx7.tamu.edu/homelinks/bac.est/bac.htm>), TM-1 [18, 19], Maxxa [20], Suyun7235 [21], Zhongmiansuo12 [22] and 0-613-2R [23], and one from *G. barbadense* L. cv Pima90-53 [24]. Herein, we report the development of a deep-coverage BAC library for the diploid A-genome cotton species, which served as the maternal parent during polyploidization to generate the precursor of the commercially important allotetraploid species. *G. arboreum* L. cv. Jianglinzhongmian was selected for this aim.

It is a diploid species of A-genome, widely used in breeding programs because of its high level of resistance to Fusarium wilt disease and its superior fiber strength. These BAC and BIBAC libraries will provide resources essential for advanced genomics and genetics research of cotton.

## 2. Materials and Methods

**2.1. BAC Library Construction.** Cotton plants were grown in the dark for about 7 days. Etiolated cotyledons were used as the source of high-molecular-weight (HMW) DNA preparation. Nuclei were isolated, lysed and megabase DNA was purified as described by Yin et al. [25]. Agrose plugs preran to eliminate small fragments in the megabase genomic DNA. Then, megabase nuclear DNA was partially digested with *Hind*III to identify an appropriate partial digestion condition. Six plugs were prerun and used for large-scale digestion with optimal amount of *Hind*III. The DNA separation was performed in two stages. First, partially digested DNA was put into the center wells of a 1% agarose gel and separated by PFGE (6 V/cm, 50 s switch time, 18 hours run time, 12.5°C). The region of the compression zone containing DNA fragments in the size range from 150 to 450 kb was excised from the unstained gel and divided into three equal sections. For the second step, the three excised gel slices were embedded into a second gel and compressed by PFGE (6 V/cm, 3–5 s switch time, 16 hours run time, 12.5°C). The compressed DNA band was excised and recovered from the agarose gel slices using an Electroeluter model 422 (Biorad, CA, USA). Eluted DNA was ligated to vector pIndigoBAC-5 (*Hind*III-cloning Ready, Epicentre Technologies, Madison, WI, USA) and incubated under temperature-cycle conditions [26]. 2  $\mu$ L of ligation mixture was used to transform 18  $\mu$ L of ElectroMAX DH10B competent cells (Invitrogen, CA, USA) by electroporation at 17 kV/cm, 100  $\Omega$ , and 25  $\mu$ F. Clones were picked by hand into 384-well plates containing LB freezing media. Plates were incubated overnight, replicated, and then frozen at  $-80^{\circ}\text{C}$ .

**2.2. BAC Clones Characterization.** The BAC clones were picked from the library and inoculated into 3 mL 2  $\times$  TY medium containing chloramphenicol (12.5  $\mu\text{g}/\text{ml}$ ) and incubated at  $37^{\circ}\text{C}$  for 24 hours. BAC-DNA was miniprepared by the method of Sambrook [27] with some modifications. To estimate, insert size and determine distribution of clone size, a total of 103 BAC clones were selected at random throughout the library. The BAC-DNA was digested with 5U of *Not*I enzyme (3 hours at  $37^{\circ}\text{C}$ ). The digestion products were separated by PFGE (6 V/cm, 5–15 s switch time, 14 hours run time,  $12.5^{\circ}\text{C}$ ). The insert size was estimated by comparison with the midrange PFGE marker II (New England BioLabs, MA, USA).

Southern blots of size-separated BAC inserts were performed by standard protocols after UV nicking the DNA (Gene Linker, UVP, CA, USA). Total genomic cotton DNA digested by *Hind*III, and *Hae*III used as probe and labeled with DIG by standard random priming techniques (High

Prime DNA Labeling and Detection Starter Kit I, Roche, Mannheim, Germany).

Six BACs selected from the library at random for stability testing (with a volume of 5  $\mu\text{L}$ ) were cultured in 3 mL 2  $\times$  TY medium with antibiotic at  $37^{\circ}\text{C}$  for 24 hours. 5  $\mu\text{L}$  of this culture was used to inoculate a subsequent 3 mL 2  $\times$  TY. This procedure was continued for five cycles. Every 24 hours period was considered to represent about 20 generations [28]. DNA samples isolated from the 1st and 5th day cultures (0-generation cells and 100-generation cells) were digested with *Hind*III and ran on 0.8% agarose gel for 16 hours with 2.0 V/cm. The gel was stained with ethidium bromide for 30 minutes, destained in water for 30 minutes, and then photographed.

**2.3. BAC Library Pooling.** Under the pooling strategy by Yin et al. [25], the BAC library was arranged in two levels of pools (column pools; Super pools) allowing screening with various PCR-based markers (Figure 3). The strategy consists of a two-step approach. Firstly, for every 384-well plate, the clones (5  $\mu\text{L}$  cultures of each clone) in individual row A to P in the same column were combined into a pool containing 16 individual clones by the 12-channel transferpettor. Each column pool composed of 16 sequential 384-well plates in this way. In the second step, every column pool in the same row was mixed by the 8-channel transferpettor. So the entire BAC library of 123,648 clones was organized into 322 super pools, each consisting of 384 unique clones.

## 3. Results

As cotton leaves are particularly rich in polyphenols and polysaccharides, we modified the general library construction method to meet the needs of cotton BAC library construction. Modifications included using etiolated cotyledon as the DNA preparation source, addition of PVP40, and increasing the  $\beta$ -mercaptoethanol concentration in the extraction-washing buffer. The prepared DNA, consisting of about 1 Mb nearly free of protein and organelle DNA, was suitable for BAC library construction. HMW-DNA embedded in LMP agarose plugs was partially digested with the enzyme *Hind*III. The optimal concentration range of *Hind*III found to produce the maximum number of 150–450 Kb DNA fragments was 8–12 U/plug. The partially digested DNA was gel separated and size selected twice. The DNA fragments from the second size selection were electroeluted from the gel, and DNA concentration was at least 3 ng/ $\mu\text{L}$ . Total twelve separate ligation reactions gave rise to the BAC library of 123,648 individual clones stored in 322 384-well plates.

To analyze the distribution of insert size and estimate the average insert size in the BAC library, DNA samples were isolated from 103 clones. BAC-DNA was digested by *Not*I to release the insert and fractionated by PFGE (Figure 1(a)). Statistical analysis indicated the insert size of clones ranged from 30 to 190 kb, with an average size of 100.2 kb (Figure 1(c)). No clone was found without an insert. Based on an A-genome size of 1697 Mb [29], the coverage of

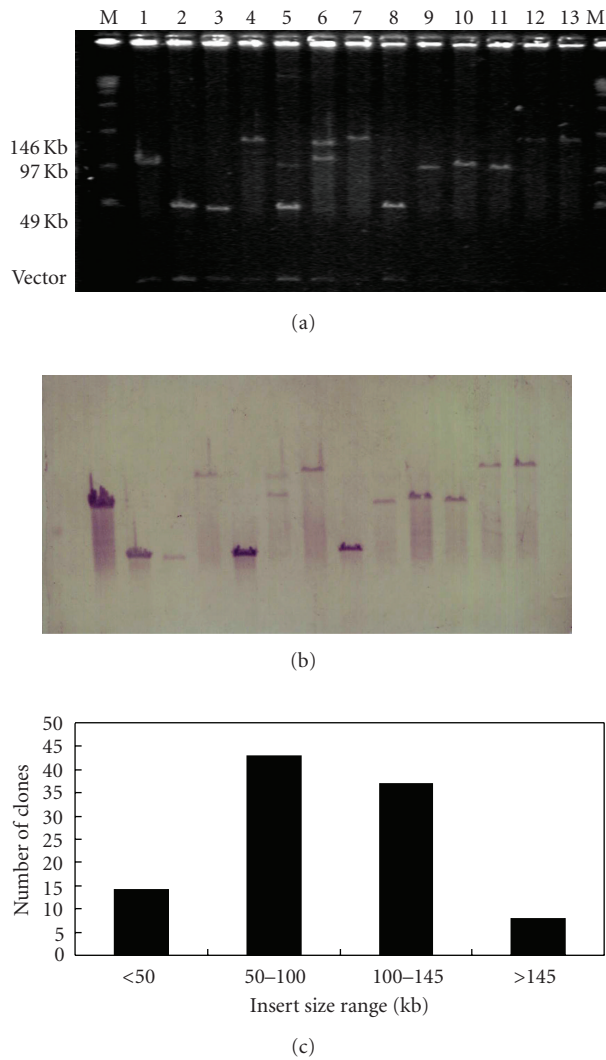


FIGURE 1: Analysis of the cotton BAC clones. (a) Ethidium bromide-stained agarose gel showing 13 BAC clones digested with *NotI* and separated by PFGE. The size marker is the midrange PFGE marker II and its sizes are indicated in the margin. (b) An autoradiograph of the same gel hybridized with total cotton genomic DNA. (c) Distribution of insert sizes based on the analysis of 103 clones randomly selected from the BAC library.

the library is approximately 7.2 haploid genome equivalents. This accounts for an over 99% probability of hitting a specific BAC clone containing any sequence in the genome. *NotI* is a GC-8 base cutter, while cotton genome is relatively AT-rich, so digestion with *NotI* should generate one or two insert bands plus a vector band (7.5 kb), which is consistent with our data. A Southern blot of the gel shown in Figure 1 probed with total cotton genomic DNA (Figure 1(b)) indicated that the source of cloned DNA originated from cotton.

To test the stability of BAC clones in *E. coli*, we analyzed the *HindIII* restriction patterns of six BAC clones in the 0 and 100 generations. No visible changes in fingerprints were seen between 0 and 100 generations (Figure 2), indicating the stability of the BACs.

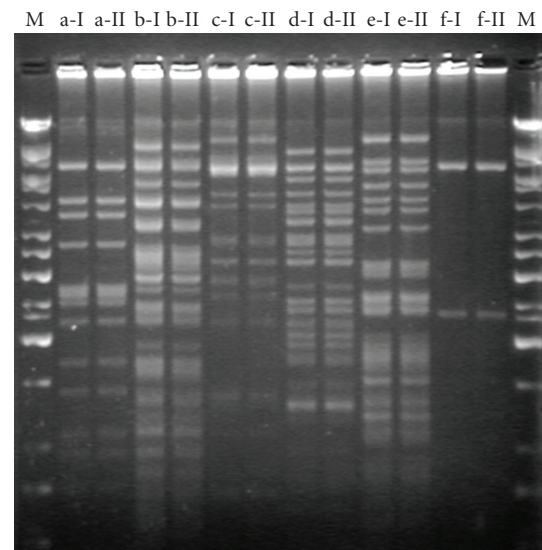


FIGURE 2: Stability study of BAC clones. DNA samples were isolated from 6 single BAC clones derived from 1st and 5th cultures (0-generation cells and 100-generation cells) and digested with *HindIII*. M 1 Kb ladder DNA marker+ $\lambda$ DNA/*HindIII*; a–f 6 single BAC clones; I 0-generation cells II, 100-generation cells.

#### 4. Discussion

Cotton has a relatively large genome with about 30%–60% of the genome consisting of highly repetitive sequences. Moreover, many homologous regions exist between the At and Dt genomes of the tetraploids. These characteristics hamper standard analysis of the cotton genome. Therefore, construction of a BAC library provides an important alternative tool for genomic research. Here, we described the development and characterization of a high-quality BAC library from the A-genome diploid species *G. arboreum* L. cv. Jianglinzhongmian, an elite cultivar. *G. arboreum* L. is an old cultivated cotton species that was planted extensively in China. To date, this species is still used in the tetraploid cotton-breeding program as an elite germplasm line, due to unique properties that include early maturity, environmental resistance, biotic resistance, high fiber strength, and excellent plasticity. Additionally, *G. arboreum* L. has historically been chosen as a model system to study fiber development. The A-genome produces spinnable fiber, whereas the D-genome alone is worthless in terms of fiber production [30]. Mei et al. [31] found among seven detected QTLs for six fiber-related traits, five were distributed among A-subgenome chromosomes, suggesting that the A-subgenome of the allotetraploid cotton contributes to the superior fiber production. Therefore, a BAC library established from A-genome species is useful in cotton genome studies.

Efficient library screening is crucial for all applications of the library. Screening can be performed either by hybridization on high-density filters or by PCR. Both methods are feasible, but the main advantage of hybridization is the ability to combine probes for screening the entire BAC library and identifying clones in a single experiment. PCR screening,

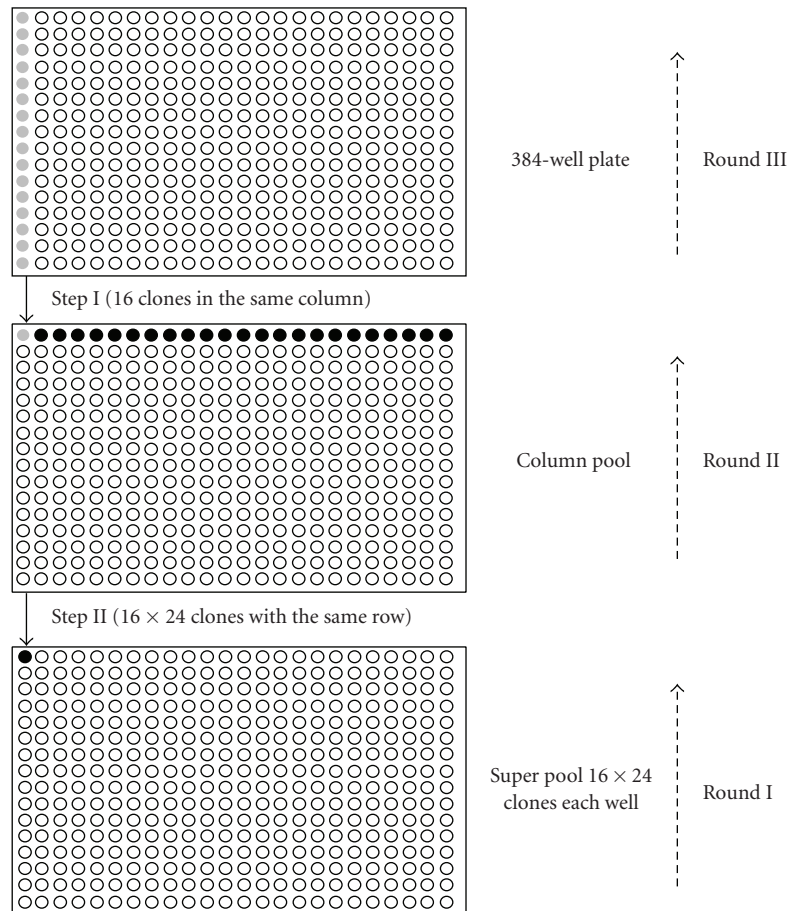


FIGURE 3: The scheme representations of BAC pooling and PCR-based screening strategy. “→” indicated the direction of BAC pooling. Step I, one column pool was mixed with each 384-well column and 16 clones in the same column (gray dot) were combined into a column pool. Step II, One super pool was mixed with each column pool in the same raw and the column pools in same raw (black dot) were mixed. The entire BAC library of 123, 648 clones on 322 384-well plates was organized into 322 super pools. “↑” indicated the direction of three rounds of PCR-based screening. Round I, screening against 322 Super pools. Round II, screening against the individual row of column pools. Round III, screening against the individual column of particular plate(s).

however, is much more reliable, faster, and efficient with higher specificity owing to effective avoidance of false positive clones identified by repeat sequences in probes by hybridization. Here, we described a three-step PCR screening procedure based on the BAC library pool system. The BAC pool strategy was sensitive enough to identify single positive clones among superpools containing 384 BAC clones. BAC clones cultured overnight served as PCR template directly, rather than using prepared BAC-DNA. This modification considerably simplifies the procedure and shortens the time required for library screening. With this change, the BAC library can be screened using SSRs, RAPD, and other PCR-based molecular tagging and will facilitate the development of whole-genome integrated physical/genetic maps of cotton.

In addition, the *G. arboreum* L. cv. Jianglinzhongmian BAC library will provide a valuable tool for a diverse range of other studies including comparative genomics, physical mapping, map-based cloning of gene(s) of QTL(s), and marker development based on BAC-end sequencing and

genotyping sequencing. Recently, we developed a BAC library from TM-1 [19], which is a tetraploid genetic standard line for upland cotton. In the future, the availability of these two cotton BAC libraries will allow us to perform further comparative studies between A-genome and AD-genome species and these comparisons will further reveal the evolution of cotton genome.

## Acknowledgments

This study is supported by grants from the Project of the Changjiang Scholars and Innovative Research Team in University (IRT0432) and the 111 Project (B08025), the Ministry of Education of China.

## References

- [1] J. F. Wendel and R. C. Cronn, “Polyploidy and the evolutionary history of cotton,” *Advances in Agronomy*, vol. 78, pp. 139–186, 2001.



- [2] H. Shizuya, B. Birren, U.-J. Kim et al., "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 18, pp. 8794–8797, 1992.
- [3] S.-S. Woo, J. Jiang, B. S. Gill, A. H. Paterson, and R. A. Wing, "Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*," *Nucleic Acids Research*, vol. 22, no. 23, pp. 4922–4931, 1994.
- [4] D. Peterson, J. Tomkins, D. Frisch, R. Wing, and A. Paterson, "Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide," *Journal of Agricultural Genomics*, vol. 5, pp. 1–100, 2000.
- [5] M. Nishimura, S. Nakamura, N. Hayashi et al., "Construction of a BAC library of the rice blast fungus *Magnaporthe grisea* and finding specific genome regions in which its transposons tend to cluster," *Bioscience, Biotechnology and Biochemistry*, vol. 62, no. 8, pp. 1515–1521, 1998.
- [6] G.-H. Jiang, W.-M. Wang, B. Xie, W.-X. Zhai, R.-L. Lu, and L.-H. Zhu, "Construction of a Bacterial Artificial Chromosome (BAC) contig encompassing the bacterial blight resistance gene *Xa4* locus in rice," *Acta Genetica Sinica*, vol. 28, no. 3, pp. 236–243, 2001.
- [7] Q. Tao, Y.-L. Chang, J. Wang et al., "Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis," *Genetics*, vol. 158, no. 4, pp. 1711–1724, 2001.
- [8] C.-C. Wu, P. Nimmakayala, F. A. Santos et al., "Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping," *Theoretical and Applied Genetics*, vol. 109, no. 5, pp. 1041–1050, 2004.
- [9] C. Wu, S. Sun, P. Nimmakayala et al., "A BAC- and BIBAC-based physical map of the soybean genome," *Genome Research*, vol. 14, no. 2, pp. 319–326, 2004.
- [10] D. Lijavetzky, G. Muzzi, T. Wicker, B. Keller, R. Wing, and J. Dubcovsky, "Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat," *Genome*, vol. 42, no. 6, pp. 1176–1182, 1999.
- [11] A. Cenci, N. Chantret, X. Kong et al., "Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*)," *Theoretical and Applied Genetics*, vol. 107, no. 5, pp. 931–939, 2003.
- [12] J. Janda, J. Bartoš, J. Šafař et al., "Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat," *Theoretical and Applied Genetics*, vol. 109, no. 7, pp. 1337–1345, 2004.
- [13] E. D. Akhunov, A. R. Akhunova, and J. Dvořák, "BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat," *Theoretical and Applied Genetics*, vol. 111, no. 8, pp. 1617–1622, 2005.
- [14] H. Fu and H. K. Dooner, "A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region," *Genome Research*, vol. 10, no. 6, pp. 866–873, 2000.
- [15] Y.-S. Yim, G. L. Davis, N. A. Duru et al., "Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization," *Plant Physiology*, vol. 130, no. 4, pp. 1686–1696, 2002.
- [16] S.-S. Woo, J. Jiang, B. S. Gill, A. H. Paterson, and R. A. Wing, "Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*," *Nucleic Acids Research*, vol. 22, no. 23, pp. 4922–4931, 1994.
- [17] C. M. Hamilton, A. Frary, Y. Xu, S. D. Tanksley, and H.-B. Zhang, "Construction of tomato genomic DNA libraries in a binary-BAC (BIBAC) vector," *Plant Journal*, vol. 18, no. 2, pp. 223–229, 1999.
- [18] J. M. Dong, R. J. Kohel, H. B. Zhang, and J. Yu, "Bacterial artificial chromosome(BAC) libraries constructed from the genetic standard of upland cottons," 1999, <http://algodon.tamu.edu/htdocs-cotton/tm1bac.html>.
- [19] Y. Hu, W.-Z. Guo, and T.-Z. Zhang, "Construction of a bacterial artificial chromosome library of TM-1, a standard line for genetics and genomics in upland cotton," *Journal of Integrative Plant Biology*, vol. 51, no. 1, pp. 107–112, 2009.
- [20] J. P. Tomkins, D. G. Peterson, T. J. Yang et al., "Development of genomic resources for cotton (*Gossypium hirsutum* L.): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development," *Molecular Breeding*, vol. 8, no. 3, pp. 255–261, 2001.
- [21] X. F. Wang, M. A. Jun, Z. Y. Ma, G. Y. Zhang, and Y. M. Zheng, "BAC library construction and characterization of Suyuan7235, a cotton germplasm with high fiber strength," *Cotton Science*, vol. 18, pp. 200–203, 2006.
- [22] Y. M. Zheng, X. F. Wang, G. Y. Zhang, X. H. Li, and Z. Y. Ma, "BAC library construction of Zhongmiansuo12 with highyield, highquality and disease resistance," *Journal of Agriculture University of Hebei*, vol. 27, pp. 17–20, 2004.
- [23] J.-M. Yin, W.-Z. Guo, and T.-Z. Zhang, "Construction and identification of bacterial artificial chromosome library for 0-613-2R in upland cotton," *Journal of Integrative Plant Biology*, vol. 48, no. 2, pp. 219–222, 2006.
- [24] X. F. Wang, J. Ma, W. S. Wang et al., "Construction and characterization of the first bacterial artificial chromosome library for the cotton species *Gossypium barbadense* L.," *Genome*, vol. 49, no. 11, pp. 1393–1398, 2006.
- [25] J. Yin, W. Guo, L. Yang, L. Liu, and T. Zhang, "Physical mapping of the *Rf1* fertility-restoring gene to a 100 kb region in cotton," *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1318–1325, 2006.
- [26] A. H. Lund, M. Duch, and F. S. Pedersen, "Increased cloning efficiency by temperature-cycle ligation," *Nucleic Acids Research*, vol. 24, no. 4, pp. 800–801, 1996.
- [27] J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, NY, USA, 2nd edition, 1989.
- [28] U.-J. Kim, H. Shizuya, P. J. De Jong, B. Birren, and M. I. Simon, "Stable propagation of cosmid sized human DNA inserts in an F factor based vector," *Nucleic Acids Research*, vol. 20, no. 5, pp. 1083–1085, 1992.
- [29] B. Hendrix and J. M. Stewart, "Estimation of the nuclear DNA content of *Gossypium* species," *Annals of Botany*, vol. 95, no. 5, pp. 789–797, 2005.
- [30] W. L. Applequist, R. Cronn, and J. F. Wendel, "Comparative development of fiber in wild and cultivated cotton," *Evolution and Development*, vol. 3, no. 1, pp. 3–17, 2001.
- [31] M. Mei, N. H. Syed, W. Gao et al., "Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*)," *Theoretical and Applied Genetics*, vol. 108, no. 2, pp. 280–291, 2004.