

Construction of a bacterial artificial chromosome library for *Gossypium herbaceum* var. *africanum*

GAO HaiYan^{1,2}, WANG XingFen², LIU Fang¹, PENG RenHai^{1,3}, ZHANG Yan², CHENG Hua³, MA ZhiYing^{2*} & WANG KunBo^{1*}

¹ State Key Laboratory of Cotton Biology/Cotton Research Institute, Chinese Academy of Agricultural Sciences, Anyang 455000, China;

² North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Key Laboratory for Crop Germplasm Resources of Hebei Province, Agricultural University of Hebei, Baoding 071001, China;

³ Anyang Institute of Technology, Anyang 455000, China

Received November 18, 2012; accepted March 14, 2013; published online May 31, 2013

Gossypium herbaceum var. *africanum* is the only wild cotton species within the cultivated diploid *G. herbaceum*. As the A sub-genome donor of tetraploid cotton, it is characterized by its resistance to insects, diseases, and other adversities. We have constructed the first bacterial artificial chromosome library (BAC) for *G. herbaceum* var. *africanum*. With high quality and broad coverage, this library includes 75000 clones, with an average insert size of about 115 kb and fewer than 4% empty clones. Our library is approximately five-fold the size of the A-genome (1667 Mb) and it provides 99.3% probability for isolating genes of interest or their sequences. Using nine SSR markers that are located on five different chromosomes and linked with resistance to *Verticillium* wilt, seven of nine could amplify the 40 superpools and got 1–14 hits. Because of its moderate wide coverage and relative large insert size, this library will be an important genomic resource for classifying and analyzing the evolution of cotton species, as well as for isolating disease-resistance genes and control elements.

wild cotton, *Gossypium herbaceum* var. *africanum*, bacterial artificial chromosome library, SSR marker, *Verticillium* wilt resistance

Citation: Gao H Y, Wang X F, Liu F, et al. Construction of a bacterial artificial chromosome library for *Gossypium herbaceum* var. *africanum*. Chin Sci Bull, 2013, 58: 3199–3201, doi: 10.1007/s11434-013-5864-5

The germplasm for cotton (*Gossypium*) includes both wild species and cultivated landraces. Although fiber yield is lower from wild species than from cultivars, the former shows better resistance to diseases and insects and has better tolerance to drought and low temperatures. Such plants also produce high-strength fibers and have cytoplasmic male sterility. Therefore, as a valuable genetic resource for breeding programs, genes from wild lines could be exploited to improve the fiber quality, yield, and resistance in commercial cultivars [1]. *Gossypium herbaceum* var. *africanum* is the only wild type for two diploid cultivated species (*G. herbaceum* and *G. arboreum*), and is the donor of the A sub-genome in tetraploid cotton. Although its fiber

yield is comparable to that of the cultivars and its tolerance to insects, diseases, and other adversities matches that of wild cotton, few reports have been made about it, and little research has focused on its potential bacterial artificial chromosome (BAC) library.

A BAC library is an important resources for genomics studies, such as genetic and physical mapping, genome sequencing, etc. [2]. laboratories in China led by both Zheng and Ma [3–5] have established an effective system for creating three cotton BAC libraries, and have found that “CCRI12” shows high yield and *Verticillium* wilt resistance, “Suyuan 7235” has greater fiber strength, and *G. barbadense* var. “Pima 90-53” exhibits good fiber quality and disease resistance. Yin et al. [6] and Hu [7] have reported BAC libraries for cytoplasmic male-sterile Line 0-613-2R

*Corresponding authors (email: wkbcrl@163.com; mzhy@hebau.edu.cn)

and Jianglingzhongmian. By screening its library with molecular markers associated with a fertility-restoring gene, both research groups have been able to construct a genetic map linked with that gene.

Nevertheless, a BAC library for *G. herbaceum* var. *africanum* has not yet been reported. We optimized our construction system for tetraploid cotton and established such a library for this wild cotton:

Nuclei DNA fragment concentrations of >700 kb. The main band was clear, and DNA was seldom degraded or mechanically broken (Figure 1).

The fragments of digested DNA were 100 or 300 kb long when 0.8 or 1.2 U of *Hind* III (Figure 2). Therefore, we selected 1.2 U for large-scale DNA digestion.

After digestion, the DNA was treated with two rounds of electrophoresis to remove the smaller (Figure 3) then estimated the concentration of DNA (Figure 4) to ligation and transformation. Examined of the insert size of BAC clones, the average insert size was 115 kb (Figure 5) and less than 4% of the clones were empty.

To validate the genome coverage and value of our BAC library, we selected nine SSR markers for screening 40 BAC super pools that were located on Chromosomes A_{1-a}05, A_{1-a}07, A_{1-a}08, A_{1-a}09 and A_{1-a}11. These markers are closely linked or even cover the resistance region [8]. We applied the screening method described by Wu et al. [9] and utilized the genomic DNA of "CCRI 8", *G. herbaceum* "Pima 90-53",

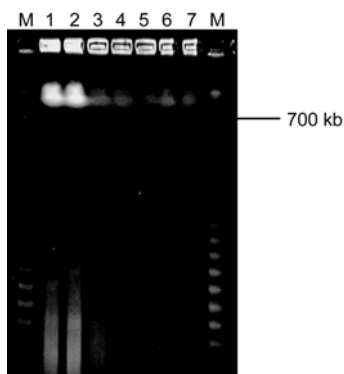


Figure 1 The quality of HMW DNA tested by PFGE.

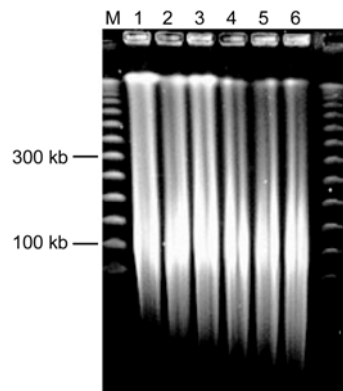


Figure 2 Determination of the optimized enzyme dosage for HMW DNA. M, PFGE marker; 1-6 indicated 0.4, 0.8, 1.2, 3.0, 4.0, and 5.0 U of *Hind* III, respectively.

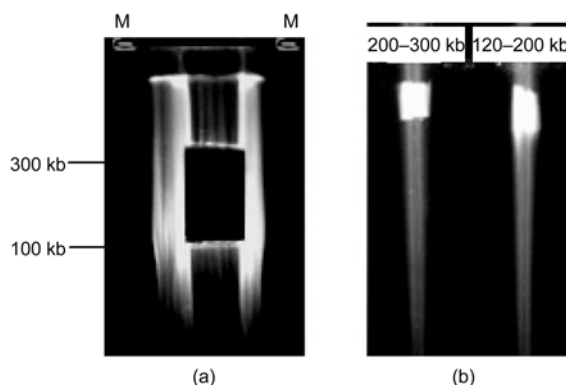


Figure 3 First (a) and second (b) size selections for partial digestion products of genomic DNA.

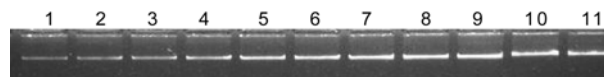


Figure 4 Detection of HMW DNA concentrations. Lanes 1-9, λ DNA with concentrations shown as nanograms per microliter; lanes 10 and 11, 120 and 200 kb of HMW DNA, respectively.

and *G. herbaceum* var. *africanum* as controls. Seven of nine could amplify the positive clones, and 1 to 14 hits were obtained (Table 1).

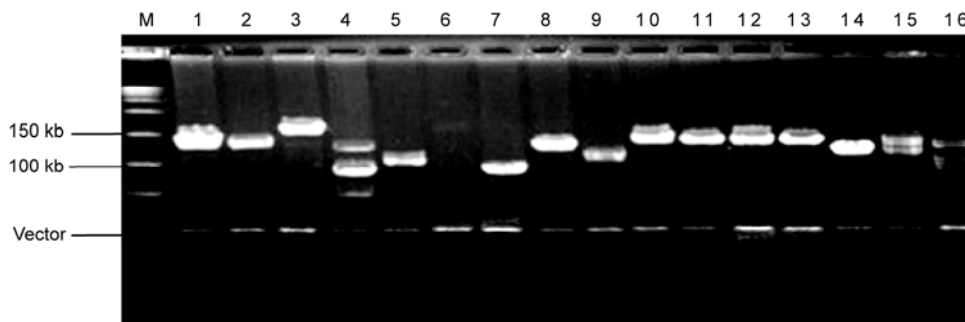


Figure 5 Insert sizes of BAC clones. Lanes 1-14, single clones; lanes 15 and 16, mixed BAC clones digested by *Not* I.

Table 1 Results of screening the BAC library with SSR markers linked to *Verticillium* wilt resistance

Number	SSR marker	Chro. No.	Size of amplification (bp)	Relative distance to the QTL (cM)	Number of hits from super pool
1	NAU5273	A _{1-a} 05	250	0.07	6
2	NAU3607	A _{1-a} 05	200	2.02	5
3	NAU2121	A _{1-a} 05	250	0	0
4	NAU1048	A _{1-a} 07	255	0.56	2
5	JESPR232	A _{1-a} 08	160	0.04	0
6	NAU3964	A _{1-a} 08	190	0.03	4
7	NAU3201	A _{1-a} 08	245	0.05	14
8	BNL3582	A _{1-a} 09	195	0.02	1
9	BNL3147	A _{1-a} 11	185	18.9	1

We report here the successful construction of the first BAC library for *Gossypium herbaceum* var. *africanum*, which is the A sub-genome donor for tetraploid cotton. In all, 75000 clones were found, with an average insert size of 115 kb and the empty clones are less than 4%. Seven SSR markers located on five different chromosomes were selected for screening 20% of this library, and positive clones were obtained for each chromosome. The BAC library was about five times larger than the size of the genome for *G. herbaceum* var. *africanum*, which is 1667 Mb [10]. Using the formula of $N=\ln(1-p)/\ln(1-f/G)$ [11], we determined that the probability was 99.3% for isolating some gene of interest. This BAC library will serve as an important genomics resource for investigating the classification, evolution, and relationships among cotton species. It can also be used as a foundation for further integrating of the physical chromosomes and for molecular genetics mapping.

We had previously digested the DNA with 3–4 U of *Hind* III when constructing a BAC library for tetraploid cotton [5]. However, because the genome for diploid cotton is smaller, we used 1.2 U of *Hind* III here. The period of electrophoresis also affected the efficiency of HWM DNA recovery when the desired size of DNA was selected through two rounds of PFGE. Higher efficiency was achieved with fragments concentrations of 100–200 kb. During our first

size selection, we changed the pulse time for PFGE from 50–50 s to 10–50 s. However, main bands were not as easily formed when 100–200 kb of DNA was used with 50–50 s, making it difficult to gather and recover during the second size selection. This was because the initial pulse, 50 s, was relatively long and we could not eliminate the smaller DNA that was intermingled.

This work was supported by the National Basic Research Program of China (2010CB126002). The authors are grateful to Priscilla Licht for critical reading of the manuscript.

- 1 Wang K B. *Cotton Sci*, 2007, 19: 354–361
- 2 Shizuya H, Birren B, Kim U J, et al. *Proc Natl Acad Sci USA*, 1992, 89: 8794–8797
- 3 Zheng Y M, Wang X F, Zhang G Y, et al. *J Agricult Univ Hebei*, 2004, 27: 17–20
- 4 Wang X F, Ma J, Ma Z Y, et al. *Cotton Sci*, 2006, 18: 200–203
- 5 Wang X F, Ma J, Zheng Y M, et al. *Genome*, 2006, 49: 1393–1398
- 6 Yin J M, Guo W Z, Zhang T Z. *Plant Biol*, 2006, 48: 219–222
- 7 Hu Y. The development and preliminary applications of genomic BAC libraries from *G. arboreum* and *G. hirsutum*. Doctoral Dissertation. Nanjing: Nanjing Agricultural University, 2008
- 8 Chang Y, Guo W Z, Li G Y, et al. *Plant Sci*, 2008, 174: 290–298
- 9 Wu Q, Cheng H, Liu F, et al. *Chin Sci Bull*, 2010, 55: 2099–2105
- 10 Hendrix B, Stewart J McD. *Ann Bot*, 2005, 95: 789–797
- 11 Clarke L, Carbon J. *Cell*, 1976, 9: 91–100

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.