A *Gossypium* BAC clone contains key repeat components distinguishing sub-genome of allotetraploid cottons

Yuling Liu¹,², Renhai Peng¹,³, Fang Liu¹, Xingxing Wang¹, Xinglei Cui¹, Zhongli Zhou¹, Chunying Wang¹, Xiaoyan Cai¹, Yuhong Wang¹, Zhongxu Lin²* and Kunbo Wang¹*

**Abstract**

**Background:** Dissecting genome organization is indispensable for further functional and applied studies. As genome sequences data shown, cotton genomes contain more than 60% repetitive sequences, so study on repetitive sequences composition, structure, and distribution is the key step to dissect cotton genome.

**Results:** In this study, a bacterial artificial chromosome (BAC) clone enriched in repetitive sequences, was discovered initiative by fluorescence in situ hybridization (FISH). FISHing with allotetraploidy cotton as target DNA, dispersed signals on most regions of all A sub-genome chromosomes, and only middle regions of all D sub-genome chromosomes were detected. Further FISHing with other cotton species bearing A or D genome as target DNA, specific signals were viewed. After BAC sequencing and bioinformational analysis, 129 repeat elements, size about 57,172 bp were found, accounting for more than 62% of the BAC sequence (91,238 bp). Among them, a type of long terminal repeat-retrotransposon (LTR-RT), LTR/Gypsy was the key element causing the specific FISH results. Using the fragments of BAC matching with the identified Gypsy-like LTR as probes, the BAC-57I23-like FISH signals were reappeared. Running BLASTN, the fragments had good match with all chromosomes of *G. arboreum* (A₂) genome and A sub-genome of *G. hirsutum* (AD₁), and had relatively inferior match with all chromosomes of D sub-genome of AD₁, but had little match with the chromosomes of *G. raimondii* (D₅) genome, which was consistent with the FISH results.

**Conclusion:** A repeats-enriched cytogenetic marker to identify A and D sub-genomes of *Gossypium* was discovered by FISH. Combined sequences analysis with FISH verification, the assembly quality of repetitive sequences in the allotetraploid cotton draft genome was assessed, and better chromosome belonging was verified. We also found the genomic distribution of the identified Gypsy-LTR-RT was similar to the distribution of heterochromatin. The expansion of this type of Gypsy-LTR-RT in heterochromatic regions may be one of the major reasons for the size gap between A and D genome. The findings showed here will help to understand the composition, structure, and evolution of cotton genome, and contribute to the further perfection of the draft genomes of cotton.

**Keywords:** *Gossypium*, BAC, Repetitive sequences, LTR-RT, FISH
The two progenitors of allotetraploidy cotton diverged 4–8 million years ago, and re-hybridized about 1–2 million years ago [6, 7]. There is enough time for sequence divergence, as well as subsequent genome stability. What’s more, there is a wide range in genome size across closely related diploid species (from 880 Mb to 2572 Mb per haploid nucleus) and well-established phylogeny in *Gossypium* [8]. So, cotton is also an excellent model system for studying polyploidization, genomic organization, and genome size variation. To dissect the genomic complexity in allotetraploidy cotton, extensive efforts have been performed. The ployploid parentage had been explained with the help of series of cytogenetic data combined with the observation derived from different studies. In early years, based on some classic cytogenetic and cytological studies, genome composition of the polyploids was investigated, which confirmed that the American allotetraploidy species are allopolyploids containing two resident genomes, an A-genome from Africa or Asia, and a D-genome similar to those found in the American diploids [9–11]. With the extensive application of FISH, more evidences that allotetraploidy cottons may be polyphyley have been obtained [12, 13].

It is believed that the proportion of protein-coding sequences is generally similar in different plant species [14], and repetitive DNA sequences are important factors in genome size variation [15–17]. Repetitive sequences can be classified into two categories: tandem repeats and transposable elements [18]. The former, which is usually found in specific genomic regions, such as centromeres or telomeres, has been extensively studied in different plant species [19–24]. Among the latter, retrotransposons replicating through a ‘copy and paste’ mechanism can result in the increase of the genome size to a great extent. Different methods had been used for analysis of repetitive DNA sequences, such as the low 

C_\text{G}\_\text{T} analysis [25, 26], bacterial artificial chromosome (BAC) end sequences analysis [27], full-length BAC sequence analysis [28, 29]. To date, the most powerful method to characterize the high copy fraction of a genome is next generation sequencing and subsequent bioinformatic analysis [30, 31]. Recently, the draft assemblies of cotton genomes have been reported. More than 60 % of repetitive DNA sequences in genomes were revealed [32–36]. So dissecting the repetitive DNA sequences of genome is helpful to further understand the composition, evolution, and function of the cotton genome.

Fluorescence in situ hybridization (FISH), which allows direct mapping of DNA sequences on chromosomes, has become the most important technique in plant molecular cytogenetics [37]. Unique distribution patterns of repetitive DNA sequences on chromosomes has been revealed by FISH [38, 39], which provided a wealth of information regarding the chromosomal location of repetitive DNA sequences and their evolution in polyploidy genomes.

Here we analyzed a repeats-rich BAC clone combining FISH verification with sequence analysis, and identified the key elements resulting in specific FISH signal patterns, that is, a type of long terminal repeat-retrotransposon (LTR-RT). Simultaneous FISH with different cotton species as target chromosomes provided visual cytogenetic evidences of the colonization and size variation of the genomes. Moreover, by integrating FISH results with the cotton draft genomes, we preliminarily assessed the assembly quality of the draft genome assemblies.

**Methods**

**Plant materials and BAC library**

The cultivated *Gossypium* species, *G. hirsutum* (AD\text{1}) (accession TM-1), *G. barbadense* (AD\text{2}) (cultivar Hai-7124), and *G. arboreum* (A\text{3}) (cultivar Shixiya-1) were planted at Institute of Cotton Research of Chinese Academy of Agricultural Sciences (CRI-CAAS) in Anyang City, Henan Province, China. The wild species *G. tomentosum* (AD\text{3}) (accession P0601211), *G. mustelinum* (AD\text{4}) (accession P0811704), *G. darwinii* (AD\text{5}) (accession AD\text{5}-7), *G. raimondii* (D\text{3}) (accession D\text{5}-2), and the artificial hexaploid cotton (*G. hirsutum* (AD\text{1}) x *G. stockii* (E\text{1})) are perennially growing in National Wild Cotton Nursery in Sanya city, Hainan Island, China. The BAC library of *G. herbaceum* var. *africanum* was constructed by Gao et al. [40].

**BAC clone screening**

During the screening of the 1\text{st} chromosome-specific BACs from the BAC library of *G. herbaceum* var. *africanum*, with SSR markers derived from a whole-genome marker map [41], the BAC clone 57123 enriched in repeats was found. The corresponding SSR marker Gh216, with primers (F\text{/R}): TCCACATTCCATGCACTACTC/CTAAAAACCTTATACATACAAAAATGAGC was used to screen the BAC library according to Cheng et al. [42] with a few modifications.

**BAC sequencing and repeats identification**

The screened BAC clone 57123 was sequenced and assembled by Shanghai Invitrogen Inc. Then BLASTN searches were performed using the BAC sequence as query, the draft genomes of cotton [33, 34, 36] as subjects respectively to detect the high copy repeats consisted in the BAC sequence. To further identify repeats types, online programs CENSOR (http://www.girinst.org/) [43], LTR-FINDER (http://tlife.fudan.edu.cn/ltr_finder/) [44] were used with the default parameters.

**Isolation of repeats**

The primers of the selected repeats, with better match to genome or higher score in CENSOR results,
were designed using NCBI primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Touchdown PCR was performed to obtain amplification products with the BAC-57I23 bacterium as template. The amplification procedure was as follows: firstly, 98 °C 5 min for pre-degeneration; then 98 °C for 11 s, 52 + 1 °C for 18 s, 68 °C for 2.5 min for 10 cycles; 98 °C for 11 s, 57 °C for 18 s, 68 °C for 2.5 min for 30 cycles with a final extension at 68 °C for 6 min.

DNA probes preparation
To visualize the distribution of the BAC-57I23 and its repeat elements, FISH was performed using BAC-DNA and repeat elements as probes respectively. BAC-DNA was isolated using Plasmid Miniprep Kit (Bionig) according to the handbook. The PCR products were purified using Wizard SV Gel and PCR Clean-up System (Promega). They were labeled with DIG-nick translation Mix, according to the instructions of the manufacturer (Roche, USA).

Chromosome preparation and FISH
Chromosome Preparation and the FISH procedure were conducted according to the previous protocols [45, 46]. The probes were detected with anti-digoxigenin-rhodamine (red) (Roche, USA). Images were captured using a CCD camera attached to a Zeiss Imager M1 microscope. Images were processed using Photoshop CS3.

Results
Discovery of the repeat-rich BAC clone 57I23
During the screening of the 1st chromosome-specific BACs from the BAC library of G. herbaceum var. africanaum, a genome-specific BAC clone 57I23 was obtained using SSR marker Gh216, which was genetically mapped to AD_chr.01 (A01) [47, 48]. FISHing with AD genome species as target DNA, the signals dispersed on all chromosomes except the terminal areas of A sub-genome, and only middle areas of all D sub-genome chromosomes (Fig. 1a-e). So the FISH with BAC-57I23 can distinguish A sub-genome from D sub-genome simultaneously. Further FISHing with diploid A and D species, high coverage signals on all chromosomes of A genome were found (Fig. 1g), but hardly any signal on chromosomes of D genome (Fig. 1h). When using the artificial hexaploid hybrid (G. hirsutum x G. stocksii) preparation as target chromosomes, the similar A and D sub-genome signal patterns were observed, and none signal on E sub-genome (Fig. 1f). More than 15 metaphase cells with clear chromosome spreads were chosen to analyze the distribution of the FISH signals along the chromosomes. Based on the signal pattern, we deduced that the BAC clone 57I23 enriched in some types of repetitive elements.

BAC sequencing and BLASTN analysis
To further understand the composition of BAC-57I23, BAC sequencing was performed by Shanghai Invitrogen Inc. Due to the existence of enriched repetitive sequences, three scaffolds with size of scaffold1-42,338 bp, scaffold2-26,803 bp, scaffold3-22,097 bp were obtained, respectively. By BLASTN using the BAC sequence as query and A2 draft genome (G. arboretum) [34] as subject sequence, we obtained ten DNA fragments (named after its sequence location in corresponding scaffold) from the BAC sequence, based on the more-than-80 % similarity and zero or approximate zero e-value. With the ten selected DNA fragments as query sequences, BLASTN
were performed against D5 (G. raimondii) and AD1 (G. hirsutum) draft genomes [33, 36] respectively. After comparing the distribution of the ten fragments in different cotton genomes, it was found that the copy number was the highest in A2 genome, but 10–25 times lower in D5 genome (Fig. 2), and with very bad match hits (data not shown), which maybe partially explain the FISH results in D genome species. We extracted the sequences of the ten fragments from the BAC sequence for the following analysis.

At the same time, taking into account the FISH results of BAC-57I23 in AD genome species, we compared the total repeated numbers of ten fragments in every chromosome of AD genome (Fig. 3). Result showed that the A sub-genome chromosomes had more than 10 times of repeats copy numbers than D sub-genome, and better consistency with the FISH results was viewed.

**Identification of repetitive sequences**

Based on CENSOR results, DNA transposon, LTR-RT, Non-LTR-RT, and other repetitive elements were identified from the BAC sequences, which account for more than 62 % of the assembled BAC sequence. Among them, LTR-RT was predominant, accounting for 88.11 % of total identified repetitive elements (55.21 to 62.66 %) (Fig. 4 and Table 1). The identified LTR-RTs were classified into LTR/Gypsy, LTR/Copia, LTR/BEL families. Especially, LTR/Gypsy accounted for more than 91 % of the total identified LTR-RTs. By combining the CENSOR with BLASTN analysis results, we selected 12 LTR-RTs with higher score value (Table 2), and extracted the corresponding sequences from the BAC sequences for FISH verification.

When running LTR-FINDER (version 1.05) using BAC sequence as query sequence, a 4118 bp full-length LTR-RT was identified in sequence region of scaffold1 (13558-17675). It belonged to the LTR/Copia family, and overlapped with Copia-80_ST-I identified by CENSOR.

By RepeatMasker (RepeatMasker version open-4.0.5) analysis, a 659 bp (sca2 (20662-21331)) Gypsy/DIRSI LTR element was identified, which had overlap region with sca2 (18785-21330) from the CENSOR results.
For further FISH verification, the partial above-mentioned fragments and LTR-RTs were PCR amplified and purified. Each purified DNA fragment had single band and expected size, which suited for the following work.

**Distribution of LTR-RTs in the cotton genomes**

The FISH analysis of somatic metaphase chromosomes showed differential distribution patterns for each LTR-RT subfamily. When using Gypsy-48 GR-I-like LTR-RTs as probes, BAC-57I23-like signals were reappeared (Fig. 5a, d-i). Using sca3 (5355-8188) as probes, the FISH signals only were observed on chromosomes of A sub-genome with lower coverage relative to BAC 57I23 (Fig. 5b), and no signal on *G. raimondii* chromosomes (Fig. 5c). Using sca1 (13558-17675), a 4118 bp-LTR/Copia element as probe, only a few dotty signals appeared (Fig. 5). But using sca2 (23904-25399), a Non-Gypsy-48 GR-I-like LTR-RT as probe, no signal appeared (Fig. 5k).

Pachytene chromosomes can display a differentiated pattern of heterochromatic and euchromatic regions [46, 49]. The pachytene-FISH results of *G. hirsutum* using fragment sca2 (18785-21330) as probe, which belonging to Gypsy-48 GR-I-like LTR-RT, showed high signal density throughout the partial pachytene chromosomes mainly following the distribution of heterochromatin, as white arrow shown (Fig. 5l).

**Discussion**

**Sub-genome-specific cytogenetic marker**

In early times, cotton chromosome identification was mainly based on the analysis of cytological characters, such as chromosomal relative lengths, arm ratios, and nuclear organization regions (NORs) in the mitotic or meiotic metaphase [50]. Because of the big number and small size of the chromosomes in cotton, the cytological identification of the chromosome has been hitherto limited. With the development of FISH, chromosome-specific FISH markers are effective tools for chromosome identification, analysis of genetic stocks, and physical mapping [13, 51 – 53]. BAC-57I23 displayed here can be used as a sub-genome specific FISH marker to identify A and D sub-genomes simultaneously in AD genome cotton species or allohexaploids containing A and D sub-genomes, due to the different FISH signal patterns on A and D sub-genome chromosomes. The discovery of BAC-57I23 provided a new FISH marker for identification of two or three sub-genomes at the same time, so the one-BAC FISH with 57I23 can take the place of GISH (genomic in situ hybridization) with two or three genomes DNA to achieve the identification of the sub-genomes.

**Assembly quality of repetitive sequences in allotetraploidie cotton draft genome**

Decoding cotton genomes is a foundation for understanding the functional and agronomic significance of polyploidy and genome size variation within the *Gossypium* genus. But high-quality assembly of allopolyploid plant genomes is a formidable task because of the large genomes and the existence of highly homeologous sub-

---

**Table 1** Summary of identified repeats in BAC sequence by CENSOR

| Repeat class          | Fragments | Length (bp) | Percentage accounting for BAC sequence |
|-----------------------|-----------|-------------|----------------------------------------|
| Integrated Virus      | 1         | 86          | 0.09 %                                 |
| DNA transposon        | 25        | 6086        | 6.67 %                                 |
| LTR Retrotransposon   | 94        | 50373       | 55.21 %                                |
| BEL                   | 1         | 65          | 0.07 %                                 |
| Copia                 | 4         | 4251        | 4.66 %                                 |
| Gypsy                 | 89        | 46057       | 50.48 %                                |
| Non-LTR Retrotransposon| 9       | 627         | 0.69 %                                 |
| Total                 | 129       | 57172       | 62.66 %                                |
genomes [36]. Mis-assemblies are common when draft genome sequences have been generated by de novo assembly of sequences obtained with NGS technologies [54]. It's possible that regions with repeated sequences might not be assembled successfully. FISH, allowing directly mapping of DNA sequences on chromosomes, has become an important technique in plant molecular cytogenetic research and can be used to guide draft genome assembly [37, 55, 56]. In this study, when blasting against the AD1-NAU draft genome using the identified repeats, the results had good consistency with the BAC-FISH results (Figs. 1 and 3). Based on this result, we can infer the assembly of the identified repetitive sequences in AD1-NAU draft genome has better matchup on their chromosome belonging.

| Name            | From     | To       | Class       | Name            | From     | To       | Class       | Name            |
|-----------------|----------|----------|-------------|-----------------|----------|----------|-------------|-----------------|
| scaffold1|size42338 | 4200     | 5326        | Gypsy-48_GR-I  | 1376     | 2498     | LTR/Gypsy   | sca1 (4200-5326) |
| scaffold1|size42338 | 13782    | 17200       | Copia-80_ST-I  | 194      | 3604     | LTR/Copia   | sca1 (13782-17200) |
| scaffold1|size42338 | 31153    | 32254       | Gypsy-48_GR-I  | 3205     | 4308     | LTR/Gypsy   | sca1 (31153-32254) |
| scaffold2|size26803 | 7498     | 8637        | Gypsy-48_GR-I  | 76       | 1221     | LTR/Gypsy   | sca2 (7498-8637) |
| scaffold2|size26803 | 23904    | 25399       | Gypsy-18 GR-I  | 1326     | 2899     | LTR/Gypsy   | sca2 (23904-25399) |
| scaffold3|size22097 | 10623    | 11420       | Gypsy-18 GR-I  | 2893     | 3697     | LTR/Gypsy   | sca3 (10623-11420) |
| scaffold3|size22097 | 17834    | 19556       | Gypsy-48_GR-I  | 2773     | 4562     | LTR/Gypsy   | sca3 (17834-19556) |
| scaffold3|size22097 | 20731    | 21832       | Gypsy-48 GR-I  | 3260     | 4363     | LTR/Gypsy   | sca3 (20731-21832) |
| scaffold2|size26803 | 13047    | 13565       | Copia-2 JC-I  | 4861     | 5382     | LTR/Copia   | sca2 (13047-13565) |
| scaffold2|size26803 | 19133    | 19572       | Gypsy-1 JC-I  | 276      | 715      | LTR/Gypsy   | sca2 (19133-19572) |
| scaffold2|size26803 | 18785    | 21330       | Gypsy-48 GR-I  | 50       | 2596     | LTR/Gypsy   | sca2 (18785-21330) |
| scaffold3|size22907 | 5533     | 6040        | Gypsy-1 JC-I  | 269      | 774      | LTR/Gypsy   | sca3 (5533-6040) |

Table 2 Selected LTR-RTs from CENSOR results

Fig. 5 FISH analysis of distribution of identified LTR-RTs in cotton genome. a, sca2 (18785-21330)- G. hirsutum; b, sca3 (5355-8188)- G. hirsutum; c, sca3 (5355-8188)- G. raimondii; d, sca1 (4200-5326)- G. hirsutum; e, sca2 (7498-8637)- G. hirsutum; f, sca3 (17834-19556)- G. hirsutum; g, sca3 (20731-21832)- G. hirsutum; h, sca1 (4200-5326)- G. arboreum; i, sca1 (4200-5326)- G. raimondii; j, scaffold1 (13558-17675)- G. hirsutum; k, sca2 (23904-25399)- G. hirsutum; l, sca2 (18785-21330)- G. hirsutum (pachytene). Bar = 5 μm

Genome size expansion and LTR-RTs
In diploid cottons, the A genome (1697 Mb) has nearly twice the size of the D genome (885 Mb) [1, 5]. The sequences analysis of cotton draft genome indicates that
the amount of sequence encompassing LTR-type retrotransposons increased from 348 Mb in G. raimondii to 1145 Mb in G. arboreum, whereas the protein-coding capacities of these two species remained largely unchanged [32, 34]. In this study, the significant difference of FISH signal patterns of the BAC-57123 between A and D genomes indicated that the BAC-57123 should have specific composition, which can partly explain the size gap between A and D genome (Fig. 1g, h). By sequence analysis, a type of Gypsy-like LTR-RTs was identified as the key element in the BAC. The genomic distribution of the identified Gypsy-LTR-RT was similar to the distribution of heterochromatin (Fig. 5I). The expansion of this type of Gypsy-LTR-RT in heterochromatic regions may be one of the major reasons for the size gap between A and D genome. Here we provided visualized evidence by FISH that the proliferation of a type of Gypsy-like LTR-RTs is one of the major reasons for genome size diversity between A and D, which further supported the former studies results [8, 57, 58].

The colonization of the genome

The previous studies showed that A-genome-specific dispersed repetitive sequences at the diploid level have colonized the D-genome at the polyploid level [38, 59]. Similarly, another study showed that a family of copia-like retrotransposable elements “horizontally” transferred across genomes following allopolyploid formation [60]. Page et al. discovered that approximately 900 kp of sequence in the polyploid genome have been converted from one genome to another in separate conversion events scattered across the genome by whole-genome re-sequencing [61]. Here, our results combined BAC-sequencing with FISH verification showed that a type of Gypsy-like LTR-RTs had high copies in G. arboreum (A2) genome, but none in the G. raimondii. (D5) genome (Fig. 5), however at the polyploidy level, obviously sequence expansion and colonization from A to D sub-genomes occurred, which dispersed on all D sub-genome chromosomes middle areas.

Conclusions

As an excellent system for studying genome evolution and polyploidization, cotton cytogenetic study is increasingly on the agenda. Combined sequences analysis with FISH verification, a new genome-specific cytogenetic marker for identification of sub-genome was discovered. The repetitive sequences assembly quality of allotetraploidy cotton draft genome was verified preliminarily, that is, the chromosome belonging of the repeats in AD1 draft genome has good consistency with the BAC-FISH results. A type of Gypsy-like LTR-RTs identified from the BAC-57123 can partially explain the size gap between A and D genome. During the process of polyploidization of cotton, “horizontally” transferred from the A sub-genome to D sub-genome The findings showed here will help to understand the composition, structure, and evolution of cotton genome, and also will contribute to the further perfection of the draft genomes of cotton, as well as provided the cytogenetic evidence for polyploidy formation.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

YL: Design of the study; Performed most of the experiments, manuscript writing. KW, RP, FL, ZL: Design of the study; Manuscript corrections; Supervision. XW, XC, CW, XC, YW, ZZ: Participated in the experiments, Manuscript corrections. All authors read and approved the final version of the manuscript.

Funding

This study was financially supported in part by a grant from the National Natural Science Foundation of China (No. 31471548), State Key Laboratory of Cotton Biology Open Fund (No.CB2014A07), National High Technology Research and Development Program (No.2013AA102601).

Author details

1State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Science, Anyang, Henan 455000, China.
2Anyang Institute of Agricultural University, Wuhan 430070, China. 3Anyang Institute of Technology, Anyang, Henan 455000, China.

Received: 18 February 2016 Accepted: 14 March 2016
Published online: 22 March 2016

References

1. Endrizzi JE, Turcotte EL, Kohel RJ. Genetics, cytogenetics, and evolution of Gossypium. Adv Genet. 1985;23:271–375.
2. Wendel JF, Schnabel A, Seelanan T. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). Proc Natl Acad Sci U S A. 1995;92(1):280–4.
3. Fryxell PA. A revised taxonomic interpretation of Gossypium L. (Malvaceae). Rheedea. 1992;2:108–65.
4. Stewart JM, Craven LA, Brubaker C, Wendel JF. Gossypium anapoides (Malvaceae), a New Species from Western Australia. Novon. 2015;23(4):447–51.
5. Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. Adv Agron. 2003;78:139–65.
6. Seelanan T, Schnabel A, Wendel JF. Congruence and consensus in the cotton tribe. Syst Bot. 1997;22:259–90.
7. Wendel JF, Albert VA. Phylogenetics of the cotton genus (Gossypium L): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Syst Bot. 1992;17:115–43.
8. Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res. 2006;16(10):1252–61.
9. Webber JM. Interspecific hybridization in Gossypium and the meiotic behavior of F1 plants. J Agric Res. 1953;51:1047–70.
10. Skovsted A. Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. J Genet. 1934;28:407–24.
11. Beasley JO. The origin of American tetraploid Gossypium species. Am Nat. 1940;74:285–6.
12. Liu SH, Wang KB, Song GL, Wang CY, Liu F, Li SH, et al. Primary investigation on GSH-NDR in cotton. Chinese Sci Bull. 2003;50(5):425–9.
13. Gani YM, Chen D, Liu F, Wang CY, Li SH, Zhang XD, et al. Individual chromosome assignment and chromosomal collinearity in Gossypium thurberi, G. hirsutum and D subgenome of G. barbadense revealed by BAC-FISH. Genes Genet Syst. 2011;86(3):165–74.
14. San Miguel P, Tikhonov A, Jin YK, Motchoukova N, Záhová D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. Science. 1996;274(5288):765–8.
15. Sober DE, Solitis PS, Bennett MD, Leitch U. Evolution of genome size in the angiosperms. Am J Bot. 2003;90(11):1596–603.
16. Bennetzen JL, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 2005;95(1):1127–32.
17. Lysak MA, Koch MA, Beaulieu JM, Meister A, Leitch U. The dynamic ups and downs of genome size evolution in Brassicaceae. Mol Biol Evol. 2009;26(1):158–98.
18. Kubis S, Schmidt T, Heslop-Harrison JS. Repetitive DNA elements as a major component of plant genomes. Ann Bot. 1998;82(Suppl):45–55.
19. Anaveesh EV, Phillips RL, Rines HW. Chromosome-specific molecular organization of maize (Zea mays L) centromeric regions. Proc Natl Acad Sci U S A. 1998;95(22):13073–7.
20. Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, et al. Genetic definition and sequence analysis of Arabidopsis centromeres. Science. 1999;286(5449):2468–74.
21. Kishii M, Nagaki K, Tsujimoto HA. Tandem repetitive sequence located in centromeric region of common wheat (Triticum aestivum) chromosomes. Chromosome Res. 2001;9(5):417–28.
22. Cheng ZK, Dong GF, Langdon TL, Ouyang S, Buehl CR, Gu MH, et al. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell. 2002;14(8):1691–704.
23. Kulkova O, Geurts R, Lamine M, Kim DJ, Cook DR, Leunissen J, et al. Satellite repeats in the functional centromere and pericentromeric heterochromatin of Medicago truncatula. Chromosoma. 2004;113(6):276–83.
24. Fajkus J, Sykorova E, Leitch AR. Telomeres in evolution and evolution of telomeres. Chromosome Res. 2005;13(5):469–79.
25. Ho SH, Leung FC. Isolation and characterization of repetitive DNA sequences from Panax ginseng. Mol Genet Genomics. 2002;266(6):951–61.
26. Liu WX, Thummasuwaw S, Seshag K, Chouvarine P, Peterson DG. Characterization of the genome of bald cypress. BMC Genomics. 2011;12:553.
27. Hong CP, Lee SJ, Park JY, Plaha P, Park YS, Lee YK, et al. Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. Mol Genet Genomics. 2004;271(6):709–16.
28. Choi HJ, Waminal NE, Park HM, Kim NH, Choi BS, Park M, et al. Major repeat components covering one-third of the genome (Panax ginseng CA. Meyer) genome and evidence for allotetraploidy. Plant J. 2014;77(6):306–19.
29. Tamura M, Hisatake Y, Motitsuka E, Watanabe A, Uchiyama K, Futamura N, et al. Analyses of random BAC clone sequences of Japanese cedar, Cryptomeria japonica. Tree Genet Genomes. 2015;11(3):50.
30. Macas J, Neumann P, Navratilova A. Repetitive DNA in the pea (Pisum sativum L.) genome comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. BMC Genomics. 2007;8:427.
31. Novak P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010;11:378. doi:10.1186/1471-2105-11-378.
32. Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, et al. The draft genome of a diploid cotton Gossypium raimondii. Nat Genet. 2012;44(10):1098–103.
33. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin DC, et al. Repeated polyploidization of Gossypium genomes and the evolution of spinny cotton fibres. Nature. 2012;492(7423):423–7.
34. Li FG, Fan GT, Wang KB, Sun FM, Yuan YL, Song GL, et al. Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet. 2014;46(6):567–72.
35. Li FG, Fan GT, Lu CR, Xiao GH, Zou CS, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30.
36. Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, Chen JD, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–7.
37. Jiang JM, Gill BS. Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome. 2006;49(9):1057–68.
38. Hanson RE, Zhao XP, Islam-Faridi MN, Paterson AH, Zwick MS, Crane CF, et al. Evolution of interspersed repetitive elements in Gossypium (Malvaceae). Am J Bot. 1998;85(10):1364–8.
39. Luo S, Mach J, Abramson B, Ramirez R, Schurr R, Barone P, et al. The cotton centromere contains a Ty3-gypsy-like LTR retroelement. PLoS ONE. 2012;7(4):e35261. doi:10.1371/journal.pone.0035261.
40. Gao HY, Wang HF, Liu F, Peng RH, Zhang Y, Cheng H, et al. Construction of the bacterial artificial chromosome library of G. barbadense var. africanum. Chinese Sci Bull. 2013;58(26):3199–201.