Toward Elucidating the Structure of Tetraploid Cotton Genome

GUO Wang-zhen

(*National Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, Nanjing* 210095, *China*)

Upland cotton has the highest yield, and accounts for >95% of world cotton production. Decoding upland cotton genomes will undoubtedly provide the ultimate reference and resource for structural, functional, and evolutionary studies of the species. Here, we employed GeneTrek and BAC tagging information approaches to predict the general composition and structure of the allotetraploid cotton genome. Further, based on our enhanced genetic map between *Gossypium barbadense* and *G. hirsutum*, we carried out the integration of 26 genetic and physical maps from *G. hirsutum* by chromosome-specific BACs and FISH technology and put a foundation for BAC-by-BAC sequencing from homologous chr. 12 and chr. 26. 142 BAC sequences from *G. hirsutum* cv. Maxxa were downloaded (www. ncbi. nlm. nih. gov) and confirmed. These BAC sequence analyses revealed that the tetraploid cotton genome contains over 70000 candidate genes with duplicated gene copies in homologous A and D sub-genome regions. Gene distribution is uneven, with gene-rich and gene-free regions of the genome. Twenty-one percent of the 142 BACs lacked genes. BAC gene density ranged from 0 to 33. 2 per 100 kb, whereas most gene islands contained only one gene with an average of 1. 5 genes per island. Retro-elements were found to be a major component, first an enriched LTR/gypsy and second LTR/copia. Most LTR retro-transposons were truncated and in nested structures. Based on our enhanced genetic map between *G. barbadense* and *G. hirsutum*, consisting of 2247 loci and covering 3514. 6 cM with an average inter-marker distance of 1. 5 cM, 166 polymorphic loci amplified with SSRs developed from 70 BAC clones were tagged on our backbone genetic map. Seventy-five percent (125/166) of the polymorphic loci were tagged on the D sub-genome. By comprehensively analyzing the molecular size of amplified products among tetraploid *G. hirsutum* cv. Maxxa, acc. TM-1, and *G. barbadense* cv. Hai 7124, and diploid *G. herbaceum* var. *africanum* and *G. raimondii*, 37 BACs, 12 from the A sub-genome and 25 from the D sub-genome, were further anchored to their corresponding sub-genome chromosomes. After a large amount of genes sequence comparison from different sub-genome BACs, the result showed that introns might have no contribution to different sub-genome size in *Gossypium*. Further, one set of chromosome-specific microsatellite-containing BACs and physical maps from *G. hirsutum* were identified by chromosome-specific BACs and FISH technology. 334 BACs, 146 for chr. 12 and 188 for chr. 26, were screened, and physical maps for the two homologous chromosomes were constructed.