

Whole genome sequencing of cotton—a new chapter in cotton genomics

CAO XiaoFeng

Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Received April 22, 2015; accepted April 27, 2015; published online April 29, 2015

Citation: Cao XF. Whole genome sequencing of cotton—a new chapter in cotton genomics. *Sci China Life Sci.* 2015, 58: 515–516, doi: 10.1007/s11427-015-4862-z

A Cotton Genome Consortium led by the Cotton Research Institute, Chinese Academy of Agricultural Sciences, with scientists from Peking University, Wuhan University, BGI and US Department of Agriculture Southern Plains Research Center has completed the genome sequence of the allotetraploid *G. hirsutum* (AtDt) by using the second-generation high-throughput sequencing technology assisted by 5-fold traditional BAC-to-BAC sequences. In this assembly 88.5% of the 2,173-Mb scaffolds which cover 89.6%–96.7% of the AtDt genome, are anchored and oriented to 26 pseudochromosomes. The *G. hirsutum* genome contains 76,493 protein-coding genes and transposable elements, which are mainly long terminal repeat (LTR) retrotransposons, accounted for 67.2% of the tetraploid cotton genome. The findings have been online published in *Nature Biotechnology* on April 20th entitled “Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution” [1]. Before publication of this tetraploid cotton genome, the same team has already sequenced and assembled two diploid ancestor genomes, *G. raimondii* (DD) [2] and *G. arboreum* (AA) [3].

Cotton (*Gossypium* spp.) is one of the most important industrial crops in the world. This genus constitutes at least five well-established tetraploid ($2n=4x=52$) and 46 diploid ($2n=2x=26$) species, which may have evolved from a common ancestor around 5–10 million years ago (MYA) that subsequently diversified to produce eight groups, including groups A–G and K. The tetraploid cotton was formed

around 1–2 MYA by hybridization between an A-genome ancestor and a D-genome ancestor, and coupled with chromosome doubling. Upland cotton (*G. hirsutum*, AtDt, $2n=4x=52$), native to Central America, is widely grown in over 80 countries and responsible for more than 90% of the world's cotton lint production in the world. Cotton fibers are the main source for textile industry and play an essential role in the global economy. Approximately 33 million hm^2 (5% of the world's arable land) is used for cotton planting, producing an annual global market value of textile mills of approximately 630.6 billion USD, mainly for cotton fiber, the natural and renewable materials for textile industry.

Apart from its economic value, cotton is also an excellent model system for studying polyploidization, cell elongation and cell wall biosynthesis. Cotton fibers are unusually long, single-celled epidermal seed trichomes, also known as cotton lint (about 3 cm final length and 20 μm width). Fiber length is a key factor determining cotton quality with the yield of cotton fiber depends mainly on the number of fiber cells of each ovule. Phytohormones, including ethylene and gibberellins (GA), are proposed to regulate cotton fiber initiation and elongation [4,5]. Comparative genomics analysis revealed that neither extremely high level of ethylene in D genome nor low level in A genome is beneficial to cotton fiber development, which suggest that fine-tuning of ethylene biosynthesis is necessary for normal cotton fiber cell elongation and maturation. These findings provided important clues for the study of allopolyploid formation in general and molecular regulatory mechanisms for cotton fiber development in particular. Publication of the tetraploid

email: xfciao@genetics.ac.cn

cotton genome opens a new chapter of cotton genomic studies. This cotton genome will facilitate the screening for new target genes important for fiber yield and quality. Also, it will be used as a template for genotyping, phenotyping, and genome-wide association studies to build more effective cotton breeding programs. It has been widely anticipated that, with the rapid accumulation of knowledge from cotton genomics, transcriptomics and proteomics studies, and genomics-assisted cotton breeding will gradually accelerate from theory to practice. Potentially, understanding the whole genome will enhance the yield and cotton lint quality.

- 1 Li FG, Fan GY, Lu CR, Xiao GH, Zou CS, Kohel RJ, Ma ZY, Shang HH, Ma XF, Wu JY, Liang XM, Huang G, Percy RG, Liu K, Yang WH, Chen WB, Du XM, Shi CC, Yuan YL, Ye WW, Liu X, Zhang XY, Liu WQ, Wei HL, Wei SJ, Huang GD, Zhang XL, Zhu SJ, Zhang H, Sun FM, Wang XF, Liang J, Wang JH, He Q, Huang LH, Wang J, Cui JJ, Song GL, Wang KB, Xu X, Yu JZ, Zhu YX, Yu SX. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum*

- TM-1) provides insights into genome evolution. *Nat Biotechnol*, 2015, doi: 10.1038/nbt.3208
- 2 Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, Yue Z, Cong L, Shang HH, Zhu SL, Zou CS, Li Q, Yuan YL, Lu CR, Wei HL, Gou CY, Zheng ZQ, Yin Y, Zhang YY, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu SX. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*, 2012, 44: 1098–1103
- 3 Li FG, Fan GY, Wang KB, Sun FM, Yuan YL, Song GL, Ma ZY, Li Q, Lu CR, Zou CS, Chen WB, Liang XM, Shang HH, Liu WQ, Shi CC, Xiao GH, Gou CY, Ye WW, Xu X, Zhang XY, Wei HL, Li ZF, Zhang GY, Wang JY, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu SX. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*, 2014, 46: 567–572
- 4 Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, Zhu YX. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell*, 2006, 18: 651–664
- 5 Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao L, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ, Zhou BL, Hu GJ, Guan XY, Chen ZJ, Wendel JF, Zhang TZ, Chen XY. Control of cotton fiber elongation by a homeodomain transcription factor GhHOX3. *Nat Commun*, 2014, doi: 10.1038/ncomms6519

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.