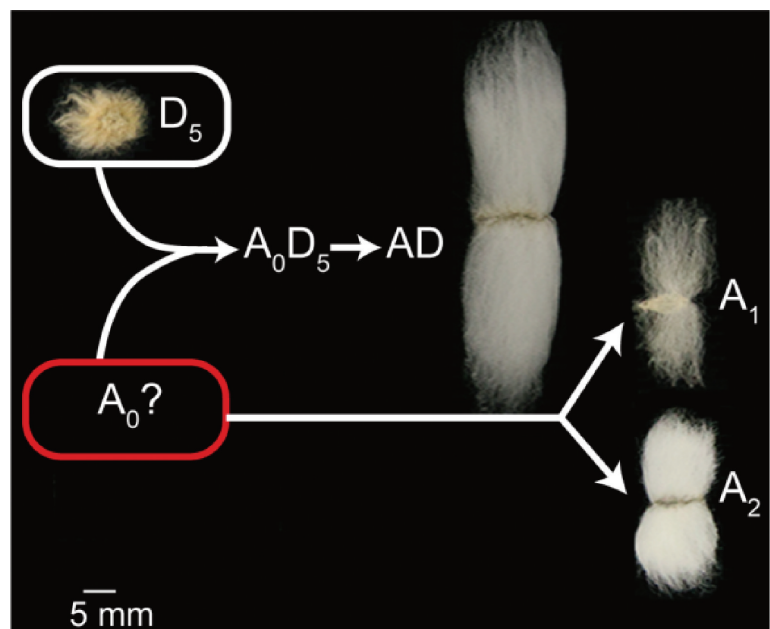


# Mystery surrounding the origin and evolution of cotton A-genome has been solved by new genomic data

With the support by the National Natural Science Foundation of China, the research team directed by Prof. Zhu YuXian (朱玉贤) at the Institute for Advanced Studies, Wuhan University, Wuhan, China, recently reported that genome sequence of *Gossypium herbaceum*, and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution, which was published in *Nature Genetics* (2020).

Cotton is one of the most important economic crops in the world. There are four human domesticated cotton cultivars, including diploid ( $A_1$ - and  $A_2$ -genome) and allotetraploid cotton [ $(AD)_1$ - and  $(AD)_2$ -genome]. Upland cotton, *G. hirsutum* [ $(AD)_1$ ], currently dominates the world's cotton commerce. AD-genome is formed by natural hybridization and chromosome doubling of D- and A-genome. There are continuing efforts and advances for Zhu' group in cotton D- and A-genomes since they published the first cotton  $D_5$ -genome in 2012 (*Nature Genetics*, 2012, 44: 1098–1103). Zhu' group has made great breakthroughs in the  $A_2$ -genome *G. arboreum* (*Nature Genetics*, 2014, 46: 567–572), widely-cultivated  $(AD)_1$ -genome *G. hirsutum* (*Nature Biotechnology*, 2015, 33: 524–530), and population variation maps of diploid cultivated cottons (*Nature Genetics*, 2018, 50: 796–802). Scholars recognized that its D-subgenome donor of (AD)-genome is *G. raimondii* ( $D_5$ -genome), but its A-subgenome donor has been in a state of debate for a long time and needs further investigation, especially from whole genome-scale phylogenetic analysis.

Upon assembling the first  $A_1$ -genome *G. herbaceum* and substantially improving the existing  $A_2$ -genome *G. arboreum* and  $(AD)_1$ -genome *G. hirsutum*, Zhu's group showed that all existing A-genomes may have originated from a common ancestor, referred to here as  $A_0$ , which was more phylogenetically related to  $A_1$  than  $A_2$ . Further, allotetraploid formation was shown to have preceded the speciation of  $A_1$  and  $A_2$ . Both A-genomes evolved independently, with no ancestor—progeny relationship. Analysis of their developed novel methods indicates that several long-terminal-repeat bursts that occurred from 5.7 million years ago (Ma) to less than 0.61 Ma contributed compellingly to A-genome size expansion and speciation. Abundant structural variations in genic regions changed the expression of many important genes, which may have led to fiber cell improvement in  $(AD)_1$ . These findings resolve existing controversial concepts surrounding A-genome origins and provide valuable genomic resources for cotton genetic improvement.



**Figure** The evolution of the allotetraploid cotton genome.