

Research & Reviews: Journal of Botanical Sciences

A Genetic Map Constructed from an Intraspecific Cross Between a Semi-Wild Line and a Cultivar in Upland Cotton

Wang Heng^{1,2†}, Wang Yuan-Yuan^{2,3†}, Zhou Zhong-li², Oluoch George^{2,4}, Kong Qing-Quan^{1,2}, Cai Xiao-Yan², Wang Xing-Xing², Zhang Zhen-Mei², Wang Kun-Bo^{2*} and Liu Fang^{2*}

¹Huazhong Agricultural University, Shizishan Street, Wuhan, 430070, Hubei, PR China

²State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, 455000, PR China

³Henan Institute of Science and Technology, Xinxiang, 453003, Henan, PR China

⁴Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

†These authors contributed equally to this work

Research Article

Received date: 01/04/2016

Accepted date: 20/06/2016

Published date: 22/06/2016

*For Correspondence

Wang Heng and Liu Fang, State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, 455000, PR China, Tel: +86-372-2525377

E-mail: 296130309@qq.com, liufcri@163.com

Keywords: Semi-wild cotton, Saline-alkaline tolerance, *G. hirsutum* races, SSRs, F₂ population, Genetic map.

ABSTRACT

Upland cotton (*Gossypium hirsutum*) accounts for more than 95% of the world cotton production but its genetic base has seriously narrowed. It is necessary to utilize desirable genes from its semi-wild lines which have enviable traits such as saline-alkaline tolerance. In this study, linkage groups were constructed using an F₂ population derived from an intraspecific cross between a semi-wild line (marie-galante 85) and a cultivar (CCRI-16) in the upland cotton (*G. hirsutum* L.). The map was based entirely on genome-wide simple sequence repeat markers. A total of 69 loci were mapped in 13 linkage groups covering a genome length of 615 cM with an average inter-locus distance of 8.9 cM. The average length of linkage group in this map is 47.3 cM, with an average number of five loci per linkage group. Based on the common primer SWU11887, one linkage group corresponded to chromosome 2. The results may lay a foundation for quantitative trait loci mapping which will facilitate the improvement of cultivars or varieties within upland cotton.

INTRODUCTION

Cotton (*Gossypium* spp.) is the world's leading natural fiber crop, which supplies approximately 35% of the total fiber used, and cotton also acts as an important oil and protein crop [1-3]. The genus *Gossypium* identified includes 53 species of two ploidy levels, of which 47 are diploid ($2n=2x=26$) and six are tetraploid ($2n=4x=52$) [4]. Moreover, there are four cultivated species: two diploids, *G. herbaceum* and *G. arboreum*, and two allotetraploids, *G. hirsutum* and *G. barbadense*. Among these species, upland cotton (*G. hirsutum*) accounts for over 95% of the world cotton production because of its wider adaptability across the globe. However, the narrow genetic background of cultivated upland cotton has resulted in inbreeding depression and reduced genetic variability. Moreover, different degrees of sterility and distinct segregation distortions, and propensity to preserve the parental haplotypes of interspecific hybrids are unusual features which hinder the progress of enriching the genomes of cultivated cotton species [5,6]. For the cultivated species of upland cotton, there are seven races (also named semi-wild cotton) from which the modern cultivated cotton (lines, cultivars or varieties) were directly domesticated, and hence they have the closest relationship. The races have a large indigenous range, collectively encompass a wealth of morphological forms, and span the wild-to-domesticated continuum [7,8], alleviating the limitation between interspecies and broadening the existing narrow genetic base.

Upland cotton (*G. hirsutum*) is widely distributed in Central and South America, the Caribbean, and even reaches distant islands in the Pacific (Solomon Islands, Marquesas). Hutchinson classified the diversity of upland cotton (*G. hirsutum*) into

seven geographical races^[9]: punctatum, latifolium, marie-galante, palmeri, morrilli, yucatanense and richmondi, which was also supported by the simple sequence repeats (SSR) study^[11]. The wide phenotypic and genetic diversity of the races were also confirmed in various ways^[10,11]. Meanwhile, the excellent phenotypes, together with the existence of genetic variation, attracted the researchers^[10,12].

The *G. hirsutum* race marie-galante occurs from El Salvador eastward through Costa Rica and Panama to the north of South America and the Caribbean. Previous reports suggested a few accessions of the race marie-galante were resistant or tolerant to Verticillium wilt and Fusarium wilt when screened. Meanwhile, marie-galante cottons survive in adverse conditions such as cold, drought and infertile growing areas^[13-15]. Zheng also found that the seeds of race marie-galante were the smallest among *G. hirsutum* races and uniformed in size, which made them close to cultivated Upland cotton (*G. hirsutum*)^[14]. Three accessions of marie-galante were identified to be resistant or tolerant to coastal salinity^[16]. Then with several methods, an accession 'marie-galante 85' was identified as saline-alkaline resistant stably in different areas^[17]. The trait of saline-alkaline resistance from marie-galante 85 can be used to improve cultivated upland cotton.

Genetic map can create a basic platform for the studies on marker assisted selection, gene/quantitative trait loci (QTL) cloning, genome sequence assembly, association mapping and evolutionary analysis. To date, several intra- and inter-specific genetic maps have been constructed using molecular markers in cotton. Moreover, most of the interspecific genetic maps were constructed using different types of populations between Upland cotton (*G. hirsutum*) and *G. barbadense*^[18-26]. Many reports were on intraspecific genetic maps, however, because of the distant relationships between the *Gossypium* species or very narrow genetic diversity within cultivated accessions in upland cotton (*G. hirsutum*), it may be an effective way to construct a genetic map from a semi-wild line with cultivars^[3,27-35].

Till now, there have been 19,074 publicly available SSRs (genomic SSRs and EST-SSRs), 3,541 RFLPs, 2,146 AFLPs and 1,018 single-nucleotide polymorphisms (SNPs) published on the Cotton Genome database (<http://www.cottongen.org/>), and most of them have been used on genetic maps construction and quantitative trait loci (QTL) detection. Among the different types of molecular markers, microsatellites or simple sequence repeats (SSRs) are popular for constructing genetic map, because they are specific and simple. However, few SSR markers have been used in cotton breeding. The development of more new polymorphic SSR markers and genetic map from the new SSRs is therefore of great importance in cotton breeding.

In this study, we used an F₂ population derived from the semi-wild line 'marie-galante 85' with a cultivar 'CCRI-16', to construct the genetic map using the SSR markers developed from *G. raimondii* sequences, designed by professor Zhang who came from Southwest University. The objective is to provide the basic foundation for a high density genetic map, the genetic dissection of important traits, the utilization of desirable genes from *G. hirsutum* races in breeding by marker-assisted selection (MAS) and map-based cloning^[36].

MATERIALS AND METHODS

Mapping Population

An F₂ population was developed from an intraspecific cross in upland cotton between the semi-wild line marie-galante 85 (as the male parent) and the cultivar CCRI-16 (as the female parent). Marie-galante 85, a line of semi-wild cotton or race of *G. hirsutum*, was identified to be saline-tolerant using different methods and in several areas during the last five years^[17]. The CCRI-16 is an outstanding cultivar with good characters, created by the Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR-CAAS). The parents were crossed in Hainan Island, China, during winter in 2014. In the same year, the F₁ were self-pollinated to produce F₂ seeds. Then 188 plants were randomly selected from the F₂ total plants for the construction of genetic map during summer in 2015 in Anyang, Henan, China. Only those plants that were mechanically damaged or showed unusual signs of insect or disease damaged were eliminated from consideration. Genomic DNA of parents, F₁ and 188 F₂ mapping individuals were extracted from young leaf tissue using a modified CTAB DNA extraction procedure. DNA solutions were quantified using NanoDrop 2000 (Thermo Co. Ltd.), diluted to working concentration (10 ng/μL) in 1 × TE buffer, and then stored at -20 °C until PCR amplification.

SSR Analysis

A total of 12,560 SSR primer pairs were first screened for polymorphism between the marie-galante 85 and CCRI-16. These primer sequences were designed from *G. raimondii* genome sequences, by professor Zhang from Southwest University, and were synthesized by Invitrogen Co. Ltd. (Shanghai, China). PCR amplification and product electrophoresis were conducted by TAKARA Bio Inc. TP 600 thermal cycler and silver stained according to the procedures described by Chen et al.^[36,37]. SSR data collection was done manually using 6% non-denaturing PAGE gel-based assays. Primer pairs with clear polymorphism were used to genotype the 188 individuals of the F₂ mapping population. According to the parents and F₁-based authentication, primers were identified as dominant or co-dominant. The data was recorded in A, B, H and AC, BD format. Multiple loci generated by single primer pair were named with the molecular weight following the primer name. The analysis of polymorphic rate was carried out using Microsoft Excel (Microsoft Corporation, Redmond, USA).

Genetic Map Construction

Linkage map was conducted using JoinMap 4.0. The Kosambi's mapping function was used to convert recombination frequencies into map distances^[38]. Log-of-odds (LOD) value of 3.0 was used to determine all linkage groups. Linkage groups were assigned to chromosomes based on common markers from previously reported maps^[39].

RESULT

SSR Marker Polymorphism

A total of 12,560 SSR primer pairs were used to screen polymorphic primer pairs between the two parents (marie-galante 85 and CCRI-16), 293 polymorphic primer pairs were obtained, with the polymorphism ratio of 2.3%. Then 91 polymorphic SSR primers, which produced clear and polymorphic bands, were used to genotype 188 individuals of the F₂ population and 97 marker loci were obtained, because six primer pairs amplified two stable bands with different molecular weights. These loci were therefore named for example as SWU12918a_163 and SWU12918b_300, which produced bands with about 163 bp and 300 bp respectively. Moreover, 68 loci of the total were codominant, accounting for 70.1% while 29.9% (29) were dominant.

Genetic Map Construction

JoinMap 4.0 mapping software being used to carry out linkage analysis of the 97 polymorphic loci and map distances in centimorgans (cM) being calculated using Kosambi's mapping function, 33 linkage groups were obtained, in which 13 had at least three loci each, accounting to 69 loci (**Figure 1**)^[38]. Another eight linkage groups had two loci each while the remaining twelve loci did not form any linkage group, and were classified as ungrouped. Of the 69 loci, 59 markers amplified single locus while five markers generated two loci each. The largest linkage group had eleven loci. Only linkage group two was possible to be linked to its corresponding chromosome (chromosome 2) based on Shang et al.^[39]. Out of the eleven mapped loci in linkage 2, seven were codominant, whereas four were dominant in which three received its alleles from *G. hirsutum* L var. CCRI-16 and one from *G. hirsutum* marie-galante 85.

Genetic Map Features

All the mapped loci spanned 615 cM with an average inter-loci distance of about 8.9 cM. The average length of the linkage groups in this map was 47.3 cM, with an average number of 5 loci per linkage group. Among the 13 linkage groups, the longest (linkage 2) was 125.8 cM while the shortest (linkage 8) was 11.8 cM.

DISCUSSION

Cultivated upland cotton has a very narrow genetic background, which results in low diversity of alleles controlling a target trait between two given parents. Therefore, far less polymorphic markers were mapped on the linkage groups derived from the intraspecific population, which have previously been reported^[3,27-29,31-33,35]. In this study, diversity of alleles were detected between the cultivated and the semi-wild upland cotton materials, *G. hirsutum* L var. CCRI-16 and *G. hirsutum* marie-galante 85, which had quite different characteristics in saline-alkaline resistance and many other agronomic traits, but had closer relationship without any reproductive isolation.

SSR markers were used in this study because they are robust, easy to generate and can be used to identify the heterozygotes among the individuals of mapping population as they are co-dominant markers and hence provide more genetic information while constructing genetic map. SSRs are highly abundant especially in eukaryotes, randomly distributed across the genome, and more often associated with non-repetitive DNA^[40]. Therefore, SSRs are highly polymorphic markers and can be employed for variety of species ranging from similar to diverse species. These are the reasons why most of the researchers considered them ideal markers in gene mapping studies^[41,42].

In the present study, the polymorphic markers accounted for 2.3% of all SSRs used, the actual polymorphic ratio is actually much lower. This low polymorphism ratio further confirmed that upland cotton has a very narrow genetic background. However, based on nearly the same primers by Shang et al.^[39], the polymorphic ratio here is higher than theirs (1.7%), indicating that polymorphism between semi-wild cotton is higher than cultivated upland cotton. Therefore, *G. hirsutum* races are one of the best choices for upland cotton variety improvement^[39]. However, several prior intraspecific upland cotton maps show relatively higher DNA marker polymorphic ratio than the present study, because of the very different primers and the varied parents. Zhang et al. constructed a RIL population genetic map with the 6.8% polymorphic primers between parents. Zhang et al.^[31] found 6.6% polymorphism in a three-parent composite population. Tang et al. found 5.8% polymorphism in a population composed of two cultivated upland cotton materials^[35].

In this study, 69 loci were mapped to linkage groups, and eleven loci were linked to chromosome 2 based on the common markers from the reported maps^[39]. The linkage groups generated in this study may lay a foundation for enriching the genetic map for further analysis. However, owing to the narrow genetic base in upland cotton, the differences among upland cotton genome sequences may be mainly due to single-base deletions, insertions and mutations. Using the genome sequence of upland cotton reported and developing polymorphism from SNPs could be more effective than SSRs in intraspecific population studies of upland cotton^[43,44].

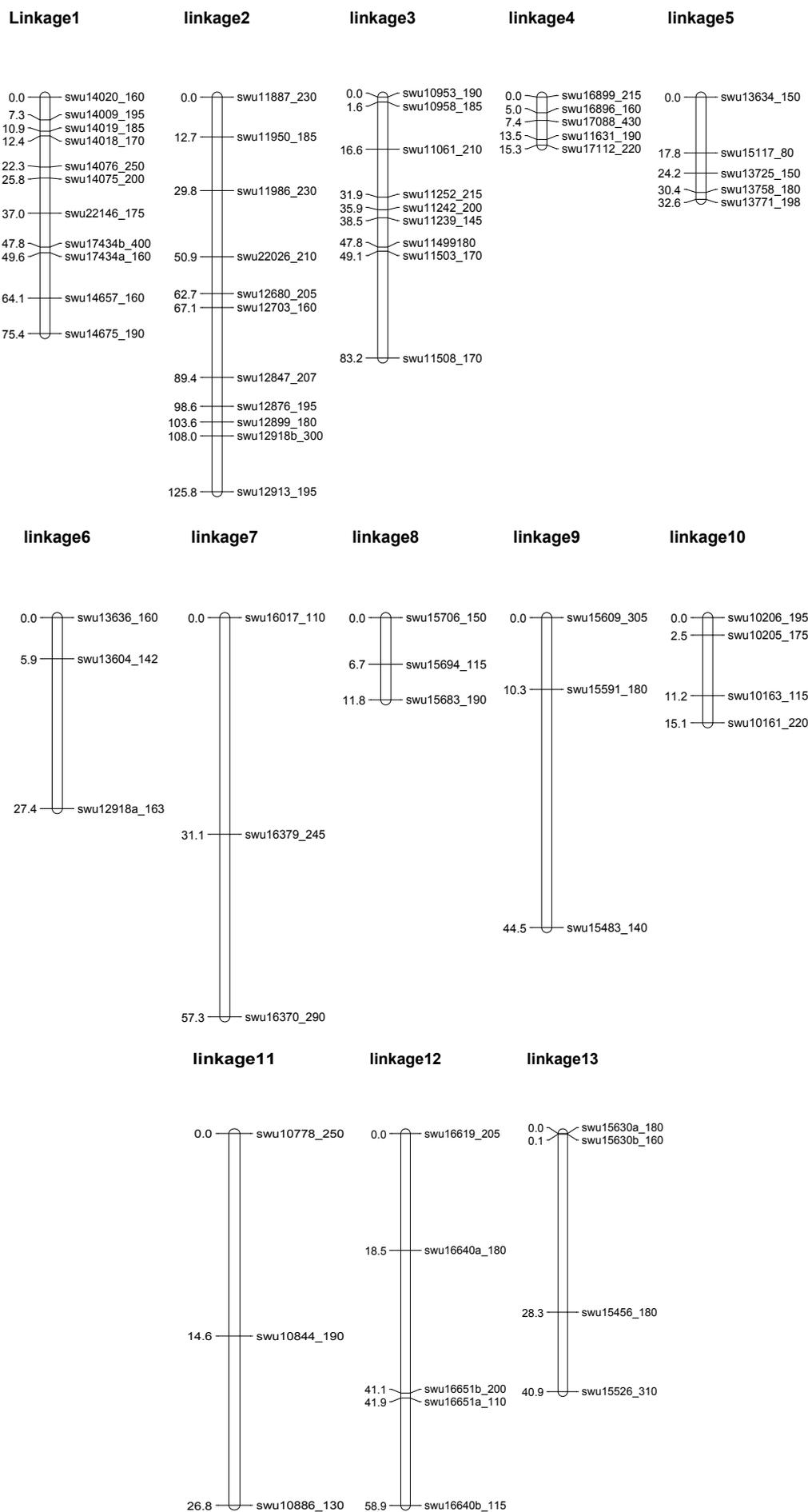


Figure 1. Locations of polymorphic loci from F₂ population of *G. hirsutum* and *G. hirsutum* 'marie-galante 85'

CONCLUSION

Soil salinization is such a serious problem in the whole world that the development of saline-alkali tolerant cotton is quite necessary. Based on SSRs, genetic map of the F_2 population derived from *G. hirsutum* (CCRI 16) × *G. hirsutum* marie-galante 85 was assembled. Totally, 13 linkage groups, covering a genome length of 615 cM with an average inter-locus distance of 8.9 cM, were constructed. The average length of linkage group was 47.3 cM, and the average loci number was five for each linkage group. The main results may lay a foundation for QTL mapping which will improve the saline-alkali tolerance of cultivated upland cotton.

ACKNOWLEDGEMENTS

This work was supported by grants from the national natural science fund project (No. 31530053).

REFERENCES

1. UADA-ERS, Cotton and wool: overview. 2013.
2. USDA-FAS, World agricultural production. 2013.
3. Zhang K, et al. Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding* 2012;29:335-348.
4. Grover CE, et al. Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genetic Resources and Crop Evolution*. 2015;62:103-114.
5. Jiang CX, et al. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution*. 2000;54:798-814.
6. Waghmare VN, et al. Genetic mapping of a cross between *Gossypium hirsutum* (cotton) and the Hawaiian endemic, *Gossypium tomentosum*. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. 2005;111:665-76.
7. Brubaker CL, et al. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome*. 1999;42:184-203.
8. Coppens d'Eeckenbrugge G and Lacape JM. Distribution and differentiation of wild, feral, and cultivated populations of perennial upland cotton (*Gossypium hirsutum* L.) in Mesoamerica and the Caribbean. *PLoS One*. 2014;9:e107458.
9. Hutchinson J. Intra-specific differentiation in *Gossypium hirsutum*. *Heredity*. 1951:169-193.
10. Abdurakhmonov IY, et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics*. 2008;92:478-487.
11. Lacape JM, et al. Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Molecular Breeding*. 2007;19:45-58.
12. Hu SA, et al. Studies of *G. hirsutum*. *Acta gossypii sinica* 6 (Suppl). 1994:15-18.
13. Shen D, et al. Identification and screening of Verticillium wilt resistant cotton. *Crop Germplasm Resources*. 1987:p 23-24.
14. Zheng S, et al. Evaluation of seed characters and Fusarium wilt-resistance in *G. hirsutum* races. *Seed*. 1994:p 6-12.
15. Lei J, et al. The difference of cold tolerance and disease resistance for semi-wild cotton in the northern of Xinjiang. *China Cotton* 2014;41:14-16.
16. Pan Z, et al. Salt Tolerance Identification of Cotton Germplasm Materials in Jiangsu Coastal Saline - alkali Soil. *Cotton Science*. 2013;35:11-14.
17. Zhou Z, et al. Primary report of the tolerance identification to saline on *Gossypium hirsutum* races. *China Cotton*. 2015;42:15-18.
18. Reinisch AJ, et al. A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics*. 1994;138:829-47.
19. Rong J, et al. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*, 2004;166:389-417.
20. Lacape JM, et al. A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. *Genome/National Research Council Canada=Genome/Conseil national de recherches Canada*. 2003;46:612-26.
21. Lin Z, et al. Construction of SRAP Genetic Map of Cotton. *Chinese Science Bulletin*, 2003;48:1676-1679.
22. Mei M, et al. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. 2004;108:280-291.
23. Han ZG, et al. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton.

- Molecular genetics and genomics: MGG. 2004;272:308-27.
24. Liang Z, et al. Toward allotetraploid cotton genome assembly: integration of a high-density molecular genetic linkage map with DNA sequence information. *BMC Genomics*. 2012;13.
 25. Yu H, et al. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One*. 2011;6:e17595.
 26. Yu JZ, et al. A High-Density Simple Sequence Repeat and Single Nucleotide Polymorphism Genetic Map of the Tetraploid Cotton Genome. *G3-Genes Genomes Genetics*. 2012;2:43-58.
 27. Shappley ZW, et al. Establishment of molecular markers and linkage groups in two F2 populations of Upland cotton. *Theor Appl Genet*. 1996;92:915-919.
 28. Qin H, et al. QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. *Theor Appl Genet*. 2008;117:883-894.
 29. Lin Z, et al. A high-density integrative linkage map for *Gossypium hirsutum*. *Euphytica*. 2009;166:35-45.
 30. Wu J, et al. Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of upland cotton. *Euphytica*. 2009;165:231-245.
 31. Zhang ZS, et al. Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding*, 2009;24:49-61.
 32. Sun FD, et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. *Molecular Breeding*. 2012;30:569-582.
 33. Ning Z, et al. Molecular tagging of QTLs for fiber quality and yield in the upland cotton cultivar Acala-Prema. *Euphytica*. 2014;195:143-156.
 34. Shao Q, et al. Identifying QTL for fiber quality traits with three upland cotton (*Gossypium hirsutum* L.) populations. *Euphytica*. 2014;198:43-58.
 35. Tang S, et al. Construction of genetic map and QTL analysis of fiber quality traits for Upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2015;201:195-213.
 36. Wang K, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*. 2012;44:1098-1103.
 37. Chen H, et al. A high-density SSR genetic map constructed from a F2 population of *Gossypium hirsutum* and *Gossypium darwinii*. *Gene*. 2015;574:273-286.
 38. Kosambi D. The Estimation of Map Distances from Recombination Values. *Annals of Human Genetics*. 1943;12:172-175.
 39. Shang L, et al. Identification of stable QTLs controlling fiber traits properties in multi-environment using recombinant inbred lines in Upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2015;205:877-888.
 40. Morgante M, et al. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet*. 2002;30:194-200.
 41. Hearne CM, et al. Microsatellites for linkage analysis of genetic traits. *Trends Genet*. 1992;8:288-294.
 42. Jarne P and Lagoda PJ. Microsatellites, from molecules to populations and back. *Trends Ecol Evol*. 1996;11:424-429.
 43. Zhang T, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33:531-537.
 44. Li F, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol*. 2015;33:524-530.