# SCIENTIFIC REPORTS

**OPEN**

# Chromosome structural variation of two cultivated tetraploid cottons and their ancestral diploid species based on a new high-density genetic map

Wen-wen Wang[1], Zhao-yun Tan[1], Ya-qiong Xu[1], Ai-ai Zhu[1], Yan Li[1], Jiang Yao[1], Rui Tian[1], Xiao-mei Fang[1], Xue-ying Liu[1], You-ming Tian[1], Zhong-hua Teng[1], Jian Zhang[1], Da-jun Liu[1], De-xin Liu[1], Hai-hong Shang[2], Fang Liu[2] & Zheng-sheng Zhang[1]

A high-resolution genetic map is a useful tool for assaying genomic structural variation and clarifying the evolution of polyploid cotton. A total of 36956 SSRs, including 11289 released in previous studies and 25567 which were newly developed based on the genome sequences of *G. arboreum* and *G. raimondii*, were utilized to construct a new genetic map. The new high-density genetic map includes 6009 loci and spanned 3863.97 cM with an average distance of 0.64 cM between consecutive markers. Four inversions (one between Chr08 and Chr24, one between Chr09 and Chr23 and two between Chr10 and Chr20) were identified by homology analysis. Comparative genomic analysis between genetic map and two diploid cottons showed that structural variations between the A genome and At subgenome are more extensive than between D genome and Dt subgenome. A total of 17 inversions, seven simple translocations and two reciprocal translocations were identified between genetic map and *G. raimondii*. Good colinearity was revealed between the corresponding chromosomes of tetraploid *G. hirsutum* and *G. barbadense* genomes, but a total of 16 inversions were detected between them. These results will accelerate the process of evolution analysis of *Gossipium* genus.

Cotton (*Gossypium* spp.) is one of the world's most important crops, with a wide variety of applications in the textile and paper industries, home fixtures, medical supplies, chemicals and oil[1]. The *Gossypium* genus is currently comprised of 52 species, including 45 diploids classified into 8 genome designations (A through G, and K) and 7 allotetraploids designated AD[2, 3]. Allotetraploid *Gossypium* are classic natural allotetraploids that derived from trans-oceanic dispersal of an A-genome ancestral African species resembling *G. herbaceum* (A$_1$) or *G. arboreum* (A$_2$), and its hybridization with a native American D-genome species resembling *G. raimondii*, followed by divergence from their common ancestor[4]. Although the two cultivated allotetraploids, *G. barbadense* and *G. hirsutum*, may share a common progenitor, the two species differ substantially in many ways.

In the past few years, several cotton genome sequences have been released, including diploids *G. arboreum* (A$_2$) and *G. raimondii* (D$_5$), and allotetraploids *G. hirsutum* (AD$_1$) and *G. barbadense* (AD$_2$)[5–11]. A series of structural variations have been detected among these genomes based on comparative analyses. *G. raimondii* and *G. arboreum* chromosomes 1, 4–6, and 9–13 were highly colinear, whereas large-scale rearrangements were observed on chromosomes 2 and 3 of *G. raimondii*, and deletions/insertions were observed on chromosomes 7 and 8 of *G. arboreum*[7]. Subsequently, 9 translocations and 28 inversions were identified by comparing the genome sequence of *G. hirsutum*[8] with *G. raimondii*[5]. In addition, with the release of the *G. barbadense* genome sequence[11], SNPs were identified by comparing the diploid progenitors (A$_1$, A$_2$, and D$_5$) and the tetraploid *G. hirsutum* (AD$_1$) with

[1]Engineering Research Center of South Upland Agriculture, Ministry of Education, Southwest University, Chongqing, 400716, China. [2]State Key Laboratory of Cotton Biology/Cotton Research Institute, Chinese Academy of Agricultural Sciences, Anyang, 455000, China. Wen-wen Wang, Zhao-yun Tan and Ya-qiong Xu contributed equally to this work. Correspondence and requests for materials should be addressed to F.L. (email: liufcri@163.com) or Z.-s.Z. (email: zhangzs@swu.edu.cn)

1

the At and Dt subgenomes of *G. barbadense* (AD$_2$)[11]. The distribution of SNPs showed that the genomic divergence between the two tetraploids (*G. barbadense* and *G. hirsutum*) was less than the divergence between the tetraploid and its diploid progenitors.

A high-resolution map is another way to study the diversity of genome structure and the morphological diversity of a genus and its biology. In recent years, many interspecific genetic maps have been developed from crosses between *G. hirsutum* and *G. Barbadense*[12–22]. Among these maps, only six possess more than 2000 markers[12, 15, 18, 19, 21, 22]. Rong *et al.*[12] reported several inversions and two reciprocal translocations between At and A chromosomes. Guo *et al.*[13] identified translocations between some chromosomes (Chr02 and Chr03, Chr04 and Chr05) based on homologous analysis, corroborating the result reported by Rong *et al.*[12]. Wang *et al.*[23] reported five inversions (Chr15-D02, LGA03-D07, Chr.12-D08, Chr.04-D09, Chr.10-D11) between a whole-genome DNA marker map and the genetic map released by Rong *et al.*[12]. Shi *et al.*[21] identified a few minor discrepancies in some marker orders along the D genome and the Dt or At subgenomes. Based on a high density genetic linkage map, Li *et al.*[22] reported that only a portion of Chr10 and Chr20 showed colinearity, whereas Chr07 and Chr16 had good homology but little colinearity. In order to support assembly of a precise genome of allotetraploid cotton, a sequence-based interspecific genetic map composed of 4,999,048 SNP loci was constructed, revealing some structural variations in the At genome and Dt genome, including 15 simple translocations reported for the first time[20].

It has become evident that higher-order genomic architectural features can lead to susceptibility to DNA rearrangements[24, 25]. Genomic rearrangements can cause Mendelian diseases or complex traits such as behaviors, or they can represent benign polymorphic changes. Revealing variation of genome architecture between the two tetraploid cultivated cottons will further deepen our understanding of the molecular basis of fiber formation. The release of genome sequences together with the high-density map is valuable for understanding genome structure and for exploring the genetic basis of important agronomic characters.

In this study, a new high-density genetic linkage map for tetraploid cotton was constructed based on SSR markers. This study aimed to (i) design SSR markers based on the sequence of two diploid species *G. arboreum*[7] and *G. raimondii*[6]; (ii) construct a new high-density genetic map; (iii) explain the relationship between genetic map and four sequenced species of cotton by aligning sequences of polymorphic loci to the genome of cotton; and (iv) reveal genome structural differences by comparing physical maps of two tetraploid cottons.

## Results

**Marker information.** In the present study, a total of 22567 new primers were designed according to the reference genomes of *G. raimondii*[6] (SWU10001-SWU22560) and *G. arboreum*[7] (CCRI00001-CCRI13007) (Supplementary Table 1–3). The 13007 primers from *G. arboreum*[7] were distributed over all 13 chromosomes, and the highest marker density (16.74/Mb) was on Chr13. For 12560 markers from *G. raimondii*[6], the number of SSR markers on the chromosome ranged from 852 (Chr13) to 1323 (Chr09). Chr12 (27.35/Mb) and Chr03 (25.22/Mb) had the higher marker densities. Dinucleotide and pentanucleotide repeats were the most common SSR type in both *G. arboreum*[7] and *G. raimondii*[6], followed by trinucleotide repeats (Supplementary Figure 1).

A total of 36956 markers, including 22567 newly developed primers and 11389 markers from cottongen (http://www.cottongen.org), were screened for polymorphic primers between CCRI35 and Pima S-7. Different series of primers showed different polymorphism ratios (Supplementary Table 4). For example, among 3926 NAU markers, 918 (23.4%) showed polymorphism, and generated 1028 loci, whereas 302 (9.5%) of 3177 HAU SSRs were polymorphic. For the newly developed markers in this study, 12560 SWU and 13007 CCRI SSRs generated 1432 and 1762 polymorphic SSRs with a polymorphism ratio of 11.40% and 13.60%, respectively. In total, 5394 primers showed clear polymorphism, for a polymorphism ratio of 14.60%, and generated 6022 polymorphic loci.

**Construction of a genetic map.** Based on the 6022 polymorphic loci, a genetic map with 6009 loci was constructed. The total length of this map was 3863.97 cM with an average of 0.64 cM between adjacent markers. The At genome contained 3049 loci with 1932.50 cM in total length and an average of 0.63 cM between adjacent markers, whereas the Dt genome contained 2960 loci with 1931.47 cM in total length and an average of 0.65 cM between adjacent markers. The chromosomes with the most loci were Chr05 (337 loci) and Chr19 (329 loci), whereas Chr04 (147 loci) had the fewest, with an average of 231 loci on each chromosome. The longest chromosome was Chr05 (236.42 cM), and the shortest was Chr04 (96.77 cM), with an average length of 148.62 cM (Supplementary Figure 2, Supplementary Table 5). In addition, among 6022 polymorphic loci, 307 (5.11%) loci showed segregation distortion (P < 0.05), including 104 on the At genome and 203 on the Dt genome (Supplementary Table 5). The largest number of distorted loci were on Chr02 (70), accounting for 34.48% of all distorted loci.

Regarding the 603 primers that segregate at multiple loci in the genetic map, most pairs of loci were mapped on homologous chromosomes (Supplementary Figure 3). All homologous chromosomes showed high colinearity. However, four inversions were identified in the present study: One inversion each was detected between homologous chromosomes Chr08 and Chr24, and Chr09 and Chr23, respectively. Two inversions were identified between Chr10 and Chr20. In addition, the two large well-known reciprocal translocations between Chr02 and Chr03 and between Chr04 and Chr05 were also identified.

**Colinearity between the genetic map and the physical maps of two tetraploid cultivated cottons.** To study the colinearity and genome variations, the physical maps of 26 chromosomes were constructed by blasting the polymorphic loci of the genetic map against the reference genome sequence of *G. hirsutum*[8] and *G. barbadense*[10]. Among the 6009 loci distributed on 26 chromosomes, 5442 (90.56%) and 5121 (85.22%) were anchored on the genome of *G. hirsutum*[8] and *G. barbadense*[10], respectively (Supplementary Figure 3). The average physical distance between adjacent markers is 0.32 Mb and 0.34 Mb, giving an overall density of 3 loci/Mb and 3 loci/Mb for *G. hirsutum*[8] and *G. barbadense*[10], respectively (Supplementary Table 5).

The relationship between the genetic map and two tetraploid genomes was explored on the basis of this genetic map and two physical maps of *G. hirsutum*[8] and *G. barbadense*[10]. All chromosomes showed high colinearity, except for a segment of Chr06 showing poor colinearity with corresponding chromosomes in *G. barbadense*[11] (Supplementary Figure 3). However, two inversions were identified between Chr05 (172.52–173.60 cM) and Gh_A05 (59.98–72.30 Mb), Gb_A05 (58.73–66.37 Mb); and between Chr08 (63.17–64.25 cM) and Gh_A08 (53.18–86.89 Mb), Gb_A08 (50.65–84.82 Mb) after comparing the genetic map and two physical maps of tetraploid cottons. In addition, 4884 of 6009 loci anchored on both *G. hirsutum*[8] and *G. barbadense*[10] were used for analyzing the genomic coline-arity of two cultivated tetraploid cottons. All chromosomes of *G. hirsutum* showed good colinearity with corresponding chromosomes of *G. barbadense* except for a segment of A06. Furthermore, at least 16 inversions were identified that differentiated the *G. hirsutum* and *G. barbadense* genomes (Supplementary Table 6).

### Comparison of the genetic map and two diploid genomes.

Sequences of 3868 and 4205 markers were homologous between the physical maps of *G. arboreum*[7] and *G. raimondii*[6], with alignment proportions of 64.37% and 69.97%, respectively. We found a high degree of colinearity between the genetic map and the *G. raimondii*[6] genome (Supplementary Figure 4), but poor colinearity between the genetic map and the *G. arboreum*[7] (Supplementary Figure 5). Therefore, only the *G. raimondii*[6] genome was used for further comparative genomic analysis. We identified 2 reciprocal translocations, 17 inversions and 7 simple translocations between this genetic map and the D genome after comparative analysis (Supplementary Table 7).

We were surprised to find that a segment of Chr02 and a segment of Chr03 of genetic map were anchored on an overlapping region (17.16 Mb − 31.31 Mb) of Gr_Chr05 (Supplementary Table 8). Therefore, the sequence of this overlapping region from *G. raimondii*[6] was used to anchor to *G. hirsutum*[8]. Further analysis showed that a total of 8 fragments (4 fragments from A02 and 4 from A03) are cross-distributed in this overlapping region of Gr_Chr05 (Supplementary Table 9).

## Discussion

Due to their high level of polymorphism and codominant inheritance, SSR (Simple Sequence Repeat) markers are a useful tool for a variety of applications including study of genetic variation, gene mapping, construction of link-age maps, and marker assisted selection. Although a number of SSRs for cotton have been developed from small segments of genomic libraries or EST sequences[26–28], these generally provided uneven and incomplete genome coverage. The release of cotton genome sequences allows us to mine SSRs at the genomic level, and analyze their distributions, putative functions, and evolution. Totals of 13007 and 12560 SSRs were designed on the basis of ref-erence sequences of *G. arboreum*[8] and *G. raimondii*[6], and the markers collectively covered approximately 99.92% and 99.77% of the physical length of the genomes, respectively. Among the 25567 SSR markers, 3138 showed polymorphism between CCRI35 and Pima S-7. Therefore, the newly developed SSR markers in the present study were applicable for map construction, fine mapping, map-based cloning and molecular marker assisted breeding. In addition, compared with publicly available markers (http://www.cottongen.org), the primers designed in the present study were endowed with precise information about their chromosomal affiliation and physical position. This new information provides us a convenient approach to unravel the relationship between diploid and tetra-ploid genomes or the homology between genetic and physical maps.

It is well known that species resembling *G. herbaceum* or *G. arboreum* and *G. raimondii*, respectively, were the putative contributors of the At and Dt subgenomes of tetraploid cotton[29]. These diploid genomes share a high level of synteny or colinearity with the corresponding subgenomes in tetraploid cotton[8–11]. In this study, comparative genomic analysis showed that colinearity between D and Dt genomes is much higher than that between A and At genomes. Possible reasons are that much more structural change exists between A and At genomes than between D and Dt genomes[8], or that much more misassemble exists in the genome sequence of *G. arboreum*[7] than *G. raimondii*[6].

Some rearrangements between the genetic map and *G. raimondii*[5] identified in this study were slightly dif-ferent from previous reports[20]. Two new inversions between Chr05 and Gr_Chr09 and between Chr08 and Gr_Chr04 were identified in this study. Moreover, the inversions were also found when we compared this genetic map with two physical maps of tetraploid cottons, which may be caused by the difference between parents and germplasm used for sequencing or a result of misassembles in the reference genome sequence (Supplementary Figure 3, Supplementary Table 7). Moreover, in addition to the rearrangements between Chr22 and Gr_Chr09, rearrangement between Chr04 and the same position of Gr_Chr09 was also identified between genetic map and *G. raimondii*[6] (Supplementary Table 7). Rearrangements between Chr02/Chr17 and Gr_Chr03, between Chr09/Chr23 and Gr_Chr06, between Chr11/Chr21 and Gr_Chr03 and between Chr13/Chr18 and Gr_Chr01 were also identified in the present study. Several researches considered the phenomenon as simple translocation. However, the existence of rearrangement in both At subgenome and Dt subgenome reveal that it may be due to misassembles of the *G. raimondii*[6] genome or the difference between *G. raimondii*[6] and its ancestor, rather than simple translocation between *G. raimondii* and the tetraploid genome reported by Zhang *et al.*[8] and Wang *et al.*[20].

Although *G. barbadense*[10] and *G. hirsutum*[8] are thought to share a common progenitor, intense directional selection by humans and the environment has led to different yield and/or quality characteristics between two tetraploid cottons. The release of genome sequences makes it easy to study genomic divergence between *G. bar-badense*[10] and *G. hirsutum*[8]. However, little literature reports the difference in genome structure between two cultivated tetraploid cottons[10, 11]. In the present study, comparative genomic analysis showed that the 26 chromo-somes of *G. hirsutum* showed good colinearity with corresponding *G. barbadense* chromosomes. However, some rearrangements were still detected between the corresponding chromosomes of the two tetraploid cottons. In the *G. barbadense*[10] genome, at least 16 large inversions (>1 Mb) were identified, compared with the genome of *G. hirsutum*[8] (Supplementary Table 6). It predicted that these structural variations occurred after polyploidization, perhaps in association with directional selection by humans and/or the environment. An important subject for

further investigation is whether these structural variations are linked to agronomic and/or quality traits that distinguish the two cultivated cotton species. Exploring these structural variations may facilitate understanding of genome evolution and gene/QTL cloning from the corresponding orthologous regions.

Two large reciprocal translocations between Chr02 and Chr03 and between Chr04 and Chr05 were identified when comparing either *G. arboreum*[7] or *G. raimondii*[6] with genetic map, consistent with the result of Brubaker *et al.*[30]. Furthermore, Chr02, Chr03, Chr04 and Chr05 showed good colinearity between two tetraploid cottons. Collectively, these findings indicated that chromosomal translocations between Chr02 and Chr03 and between Chr04 and Chr05 may occur after polyploidization and before divergence of the two tetraploid cottons.

It is worth noting that an overlapping region (16.77 Mb-31.31 Mb) on Gr_Chr05 was found after comparing Chr02/Chr03 with Gr_Chr05 and the segments of Chr02 and Chr03 anchored on Gr_Chr05 are cross-distributed in this overlapping region. We hypothesize that the reason for this phenomenon is the difference between Ga_Chr02 and Gr_Chr05 or chromosome rearrangement of the Ga_Chr02 before allopolyploidization. In other words, the fragments for A02 or A03 of tetraploid got together across multiple rearrangements (inversion or simple translocation) in this overlapping region before allopolyploidization. Then, this region was divided into two segments which were assembled on corresponding chromosomes after reciprocal translocations (Supplementary Figure 6).

Wang *et al.* proposed a new chromosome nomenclature for tetraploid cotton and established 13 homeologous chromosome pairs[31]. They treated Chr02/Chr14 and Chr03/Chr17 as homeologous chromosome pairs. After comparative analysis between genetic map and *G. raimondii*[6], we found that the proportion of segments on Chr02 corresponding to Gr_Chr03 is larger than that corresponding to Gr_Chr05, and the proportion of segment on Chr03 corresponding to Gr_Chr05 is larger than corresponding to Gr_Chr03. The same results had been reported by Wang *et al.*[20]. These findings support that the homologous chromosome of Chr02 is Chr17, whereas the homologous chromosome of Chr03 is Chr14, which had been applied in previous studies[19, 32]. Therefore, we consider that the current released chromosome names A02 and A03 should be interchanged.

## Materials and Methods

**Plant materials.**    *G. hirsutum* cv. CCRI35 as a recurrent parent was crossed with *G. barbadense* cv. Pima S-7 as a donor to establish a BC1 population. CCRI35, a high yielding and disease resistant cultivar, has been widely planted in China in the past decade[33]. Pima S-7 is a *G. barbadense* cultivar that is photoperiod insensitive and produces high-quality fiber in USA[34, 35]. Subsequently, the BC1 population [(CCRI35 × Pima S-7) × CCRI35] including 94 indivaduals was used as the mapping population.

**Development of SSRs.**    To construct a high-density genetic map, in addition to markers released by previous studies (http://www.cottongen.org), approximately 1000 primer pairs on each chromosome were designed from *G. raimondii*[6] and *G. arboreum*[7]. The reference genome sequences of *G. raimondii*[6] were retrieved from phytozome (https://phytozome.jgi.doe.gov/Graimondii_er) and *G. arboreum*[7] were retrieved from cottongen (http://www.cottongen.org). Simple sequence repeat (SSR) primers of *G. arboreum*[7] and *G. raimondii*[6] were designed using SSR locator 1[36]. The microsatellite motifs were searched by the following criteria: eighteen repeat units for mononucleotide (Mono) repeats, nine for dinucleotide (Di) repeats, six for trinucleotide (Tri) repeats, four for tetranucleotide (Tetra) repeats, three for pentanucleotide (Penta) repeats and three for hexanucleotide (Hexa) repeats. The major parameters for designing SSR primers were: (1) primer length from 18 to 27 bases; (2) PCR product size ranging from 100 to 200 bp; (3) melting temperature between 55 and 65 °C with 60 °C being the optimum annealing temperature; (4) GC content of 45–65% with an optimum of 50%. SSR primers developed form *G. raimondii* were named 'SWU', whereas primers designed with *G. arboreum* were named 'CCRI'. All primers were synthesized by Invitrogen Co. Ltd. (Shanghai, China).

**DNA extraction, PCR amplification, and electrophoresis.**    Total genomic DNA from fresh young leaves of the parents and 94 BC$_1$ individuals were extracted according to a modified CTAB method[37]. All SSR primer pairs, including 11289 released by previous studies and 25567 developed in this study, were screened for polymorphism between the mapping parents and those showing clear polymorphism were used to genotype the BC$_1$ population. PCR amplification and product tests were performed according to Zhang *et al.*[37]. Loci detected were named with the primer name. For multiple loci revealed by the same primer pair, an extra letter was added to the primer name, such as a/b/c, indicating the molecular size from smallest to largest.

**Map construction.**    The molecular map was generated using JoinMap 4.0 with the parameter that the logarithm of odds (LOD) threshold was ≥4.0[38]. Map distances in centiMorgans (cM) were calculated using the Kosambi mapping function[39]. The genetic linkage map was drawn with Mapchart 2.0[40]. In addition, each segregating marker was tested by a Chi-squared test to determine if it deviated significantly from the expected Mendelian segregation ratio.

**Colinearity and comparative genomic analysis among a genetic map and genome of cotton.**    The sequence of the *G. arboreum*[7], *G. raimondii*[6] and *G. barbadense*[10] and primers were obtained from cottongen (http://www.cottongen.org), while *G. hirsutum*[8] were retrieved from phytozome (https://phytozome.jgi.doe.gov/Ghirsutum_er). The sequences of polymorphic markers were carried out BLASTN to find corresponding chromosome and position on genome of *G. arboreum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*[6–8, 10]. The best hit was selected to analyze the relationship between the genetic map and physical map.

The relationship between genetic map and two diploid cottons was illustrated intuitively using the online drawing tool circos-0.67 (http://circos.ca/). In addition, markers anchored on both *G. hirsutum*[8] and *G. barbadense*[10] were used to analyze the relationship of two tetraploid cottons. The colinearity was illustrated intuitively using the online drawing tool D3 (https://github.com/d3/d3/wiki).

# References

1. Ensminger, M. E., Oldfield, J. E., & Heinemann, W. W. Excerpts with reference to cottonseed and cottonseed components. *Feeds. Nutr.* 252 (1990).
2. Gallagher, J. P., Grover, C. E., Rex, K., Moran, M. & Wendel, J. F. A New Species of Cotton from Wake Atoll, Gossypium stephensii (Malvaceae). *Systematic Botany* **42**, 115–123 (2017).
3. Grover, C. E. *et al.* Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. *Genet. Resour. Crop. Evol.* **62**, 103–114 (2015).
4. Senchina, D. S. *et al.* Rate variation among nuclear genes and the age of polyploidy in *Gossypium. Mol. Biol. Evol.* **20**, 633–643 (2003).
5. Wang, K. *et al.* The draft genome of a diploid cotton *Gossypium raimondii. Nat. Genet.* **44**, 1098–1103 (2012).
6. Paterson, A. H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).
7. Li, F. *et al.* Genome sequence of the cultivated cotton *Gossypium arboretum. Nat. Genet.* **46**, 567–572 (2014).
8. Zhang, T. *et al.* Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537 (2015).
9. Li, F. *et al.* Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **33**, 524–530 (2015).
10. Liu, X. *et al. Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci. Rep.* **5**, 14139 (2015).
11. Yuan, D. *et al.* The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* **5**, 17662 (2015).
12. Rong, J. K. *et al.* A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* **166**, 389–417 (2004).
13. Song, X., Wang, K., Guo, W., Zhang, J. & Zhang, T. A comparison of genetic maps constructed from haploid and BC1 mapping populations from the same crossing between *Gossypium hirsutum* L. and *Gossypium barbadense* L. *Genome* **48**, 378–390 (2005).
14. He, D. H. *et al.* QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense. Euphytica* **153**, 181–197 (2006).
15. Guo, W. *et al.* A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium. Genetics* **176**, 527–541 (2007).
16. Zhang, Y., Lin, Z., Xia, Q., Zhang, M. & Zhang, X. Characteristics and analysis of simple sequence repeats in the cotton genome based on a linkage map constructed from a BC1 population between *Gossypium hirsutum* and *G. barbadense. Genome* **51**, 534–546 (2008).
17. Lacape, J. M. *et al.* A new interspecific, *Gossypium hirsutum* x *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton. *Theor. Appl. Genet.* **119**, 281–292 (2009).
18. Yu, Y. *et al.* Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between *gossypium hirsutum* and *G. barbadense. BMC genomics* **12**, 15 (2011).
19. Yu, J. *et al.* A High-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploidcotton genome. *G3 (Bethesda)* **2**, 43–58 (2012).
20. Wang, S. *et al.* Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *Genome. Biol.* **16**, 108 (2015).
21. Shi, Y. *et al.* Constructing a high-density linkage map for *Gossypium hirsutum* x *Gossypium barbadense* and identifying QTLs for lint percentage. *J. Integr. Plant. Biol.* **57**, 450–467 (2015).
22. Li, X., Jin, X., Wang, H., Zhang, X. & Lin, Z. Structure, evolution, and comparative genomics of tetraploid cotton based on a high-density genetic linkage map. *DNA.Res.* **23**, 283–293 (2016).
23. Wang, Z. N. *et al.* A Whole-Genome DNA Marker Map for Cotton Based on the D-Genome Sequence of *Gossypium raimondii* L. *G3 (Bethesda)* **3**, 1759–1767 (2013).
24. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends. Genet.* **14**, 417–422 (1998).
25. Lupski, J. R. Genomic disorders ten years on. *Genomic. Med.* **1**, 1–11 (2009).
26. Zane, L., Bargelloni, L. & Patarnello, T. Strategies for microsatellite isolationa review. *Mol. Ecol.* **11**, 1–16 (2002).
27. Guo, Y. *et al.* BAC-derived SSR markers chromosome locations in cotton. *Euphytica* **161**, 361–370 (2007).
28. Xiao, J. *et al.* New SSR markers for use in cotton (*Gossypium spp.*) improvement. *J. Cotton. Sci.* **13**, 75–157 (2009).
29. Wendel, J. F., Brubaker, C. L. & Seelanan, T. The origin and evolution of *Gossypium. Physiol cotton* (pp. 1–18) (2010).
30. Brubaker, C. L., Paterson, A. H. & Wendel, A. J. F. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* **42**, 184–203 (1999).
31. Wang, K. *et al.* Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence *in situ* hybridization mapping. *Theor. Appl. Genet.* **113**, 73–80 (2006).
32. Fang, D. D. & Yu, A. J. Z. Addition of 455 Microsatellite Marker Loci to the High-Density *Gossypium hirsutum* TM-1 × *G. barbadense* 3–79 Genetic Map. *J Cotton Sci* **16**, 229–248 (2012).
33. Tan, Z. *et al.* Genetic map and QTL controlling fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica* **203**, 615–628 (2014).
34. Mei, M. *et al.* Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor. Appl. Genet.* **108**, 280–291 (2004).
35. Zhu, L. & Kuraparthy, V. Molecular Genetic Mapping of the Major Effect Photoperiod Response Locus in Pima Cotton (L.). *Crop. Sci.* **54**, 2492 (2014).
36. Da Maia, L. C. *et al.* SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. *Int. J. Plant. Genomics.* **2008**, 412696 (2008).
37. Zhang, Z.-S. *et al.* Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica* **144**, 91–99 (2005).
38. Van, O. J. & J. W. JoinMap 4.0: software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen (2006).
39. Kosambi, D. D. The Estimation Of Map Distances From Recombination Values. *Ann. Eugen.* **12**, 172–175 (1943).
40. Voorrips, R. E. M. C. MapChart2.2: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78 (2006).

## Acknowledgements

## Author Contributions

Z.S.Z. and F.L. conceived the experiment, designed SSR markers, analyzed the data and revised the manuscripts. W.W.W. contributed to genotyping, data analysis and writing. Z.Y.T. and Y.Q.X. assisted in DNA extraction and SSR genotyping. A.A.Z., Y.L., J.Y., R.T., X.M.F., X.Y.L. and Y.M.T. assisted in SSR genotyping. D.X.L., J.Z., D.J.L. and Z.H.T. made crosses and conducted field tests. H.H.S. contributed to designing SSR. All authors reviewed and approved this submission.

## Additional Information