Genome-Wide Mining, Characterization, and Development of Microsatellite Markers in *Gossypium* Species

Qiong Wang*, Lei Fang*, Jiedan Chen, Yan Hu, Zhanfeng Si, Sen Wang, Lijing Chang, Wangzhen Guo & Tianzhen Zhang

Although much research has been conducted to characterize microsatellites and develop markers, the distribution of microsatellites remains ambiguous and the use of microsatellite markers in genomic studies and marker-assisted selection is limited. To identify microsatellites for cotton research, we mined 100,290, 83,160, and 56,937 microsatellites with frequencies of 41.2, 49.1, and 74.8 microsatellites per Mb in the recently sequenced *Gossypium* species: *G. hirsutum*, *G. arboreum*, and *G. raimondii*, respectively. The distributions of microsatellites in their genomes were non-random and were positively and negatively correlated with genes and transposable elements, respectively. Of the 77,996 developed microsatellite markers, 65,498 were physically anchored to the 26 chromosomes of *G. hirsutum* with an average marker density of 34 markers per Mb. We confirmed 67,880 (87%) universal and 7,705 (9.9%) new genic microsatellite markers. The polymorphism was estimated in above three species by *in silico* PCR and validated with 505 markers in *G. hirsutum*. We further predicted 8,825 polymorphic microsatellite markers within *G. hirsutum* acc. TM-1 and *G. barbadense* cv. Hai7124. In our study, genome-wide mining and characterization of microsatellites, and marker development were very useful for the saturation of the allotetraploid genetic linkage map, genome evolution studies and comparative genome mapping.

Microsatellites, which consist of a variable number of tandem repeats, are known as simple sequence repeats (SSRs, defined as 1-6 bp) and are characterized by their high frequency, distribution, co-dominance, reproducibility, and high polymorphism. Among genetic marker systems such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-related amplified polymorphism (SRAP), and target region amplification polymorphism (TRAP), microsatellites have numerous uses, including linkage map, quantitative trait loci (QTL) mapping, marker-assisted selection, genetic diversity studies, and evolutionary studies.

Cotton (*Gossypium* SPP) is the most important fiber crop in the world and is also an important edible oil crop. The genus *Gossypium* has nine genome types: eight diploid genomes (A, B, C, D, E, F, G, and K; 2n = 2x = 26) and one tetraploid genome (AD; 2n = 4x = 52), based on observations of chromosome pairing. Two of the four cultivated allotetraploid species, *G. hirsutum* L. (AD), and *G. barbadense* L. (AD), account for 90% and 8% of world cotton production, respectively. Although *G. arboreum* (A2) and *G. raimondii* (D2) are regarded as the progenitors that led to the formation of the tetraploid cotton species 1-2 million years ago (MYA), the exact donor species that no longer exist.

State Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Hybrid R & D Engineering Center (the Ministry of Education), Nanjing Agricultural University, Nanjing 210095, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.Z. (email: cotton@njau.edu.cn)
Much research has been conducted to characterize microsatellites and develop markers from enriched genomic DNA\textsuperscript{11,12}, expressed sequence tags\textsuperscript{13–15}, sequences derived from bacterial artificial chromosomes\textsuperscript{16,17}, assembled contigs based on transcriptomic profile\textsuperscript{18} and sequenced genome\textsuperscript{19}. Current publicly available cotton simple sequence repeat (SSR) markers are described in the CottonDB (http://cottondb.org/) and Cotton Marker Database (CMD) (http://cottonmarker.org)\textsuperscript{20}. So far, 19,010 SSR markers have been described in the CottonDB. CottonGen supersedes CottonDB and the CMD with enhanced tools for easier data sharing, mining, visualization and data retrieval of cotton research data\textsuperscript{21}. Thanks to global efforts, high-resolution mapping in cotton has been published with segregating populations through interspecific crosses\textsuperscript{19,22–25}. However, the development of microsatellites in cotton remains limited, due to the paucity of DNA polymorphisms and the lack of complete genome sequences\textsuperscript{26–28}. Although a very-high-density whole genome marker map (WGMM) has been constructed for cotton based on the D genome, which has a total of 48,959 loci, about six times the number characterized in the most richly populated of the integrated maps published previously\textsuperscript{20,29}, the pattern of microsatellite distribution and the physical position or product of publicly available cotton SSRs in \textit{G. hirsutum} remained ambiguous.

**Figure 1.** Distribution of microsatellites identified in the assembled genomic sequences of \textit{G. hirsutum} (TM-1), \textit{G. arboreum} and \textit{G. raimondii}. (A) Frequency distribution of microsatellites with different motif lengths in \textit{Gossypium} species. (B) Distribution of microsatellites in genic and intergenic regions in \textit{G. hirsutum} (TM-1).
Recent developments in high throughput DNA sequencing technologies provide new opportunities to expedite molecular marker development. In this study, we conducted whole-genome microsatellite characterization and marker development in the sequenced genome of *G. hirsutum*, *G. raimondii*, and *G. arboreum*. We characterized and compared the frequency and distribution of motif length, type, and repeat number of microsatellites in assembled genomic sequences of these three species. In addition, we analyzed the genomic distribution of microsatellites, genes and transposable elements (TEs) in the assembled genome of *G. hirsutum*. Furthermore, we developed microsatellite markers from assembled genomic sequences, and evaluated polymorphism in different species. These markers may facilitate the advancement of many basic and applied genomic studies in cotton, including the development of high-resolution linkage maps, positional gene-cloning, and breeding in different cotton species.

### Results and discussion

**Frequency, distribution and characterization of microsatellite length, type and number.** The available 2,433 Mb, 1,694 Mb and 761 Mb genome sequences of *G. hirsutum*, *G. arboreum*, and *G. raimondii*, respectively, were searched for microsatellites with different types of desirable repeat motifs from mono- to hexanucleotide. A total of 100,290, 83,160 and 56,937 microsatellites were identified with an overall frequency of 41.2, 49.1, and 74.8 per Mb, or one every 24.3, 20.4 and 13.4 Kb, in above three assembled genomic sequences, respectively (Table 1). Meanwhile, a total of 50,443 and 41,119 microsatellites with perfect repeats were detected in the *A*, and *D*, respectively. These showed an overall density of 34.2 (*A*) and 49.5 (*D*) per Mb, or one every 29.3 and 20.2 Kb, respectively (Table 1). The determination of the genomic distribution of 100,290 microsatellites revealed 81,898 microsatellites mapped to the 26 chromosomes of *G. hirsutum* with an average density of 42.33 per Mb. In our study, physically mapped microsatellites showed a higher density of markers on the *D* (38,622, 49.87/Mb) than the *A* (43,276, 37.30/Mb), with maximums on A05 (4,248, 46.15/Mb) and D05 (3,553, 57.37/Mb), and minimums on A06 (3,204, 31.06/Mb) and D02 (3091, 45.94/Mb) (Table S1). Our results agree with a previous study which revealed a negative correlation between genome size and microsatellite density. Cucumber (367 Mb), wheat (3B, 1,000Mb), and maize (1,115 Mb) have microsatellite densities of 551.9 per Mb, 163 per Mb and 120 per Mb, respectively. Other microsatellite densities of 78.5 per Mb, 189.4 per Mb and 99.8 per Mb have been identified in *Arabidopsis*, *Brachypodium*, and *rice*, respectively.

The distributions of microsatellite length in the assembled genomic sequences of *G. hirsutum*, *G. arboreum* and *G. raimondii* showed that tri-, tetra-, penta- and hexanucleotide repeats accounted for very similar proportions, whereas mono- and dinucleotide repeats were relatively different in proportions among these genomes (Fig. 1A). Among the 100,290 microsatellites obtained in *G. hirsutum*, the hexanucleotide repeats were most abundant (39,506) with a proportion of 39.4%, followed by tri- (22,483, 22.4%), penta- (14,970, 14.9%), di- (12,445, 12.4%), tetra- (9,031, 9.0%), and mono-nucleotide (1,855, 1.8%) repeats (PNR) were found to be most abundant in the genome of *G. hirsutum*. The different results of these two criteria (minimum microsatellite length 18 vs 12) mostly due to longer reads used in genome assembling of *G. raimondii* such as Roche 454 data. Among these 100,290 microsatellites in *G. hirsutum*, 11,008 (11.0%) were detected in the genic regions.

| Species     | Microsatellites number | Genome length (Mb) | Frequency            |
|-------------|------------------------|--------------------|----------------------|
|             |                        |                    | Per Mb | One every (Kb) |
| *A*         | 50,443                 | 1477.1             | 34.2   | 29.3           |
| *D*         | 41,119                 | 831.0              | 49.5   | 20.2           |
| *G. hirsutum* | 100,290              | 2432.7             | 41.2   | 24.3           |
| *G. arboreum* | 83,160              | 1694.6             | 49.1   | 20.4           |
| *G. raimondii* | 56,937              | 761.4              | 74.8   | 13.4           |

Table 1. Overall frequency of microsatellites in *Gossypium* species. *A*, *D*; two subgenomes of allotetraploid cotton *G. hirsutum*. The different results of these two criteria (minimum microsatellite length 18 vs 12) mostly due to longer reads used in genome assembling of *G. raimondii* such as Roche 454 data. Among these 100,290 microsatellites in *G. hirsutum*, 11,008 (11.0%) were detected in the genic regions.
More intergenic microsatellites from mono- to hexanucleotide were identified than that in genic regions (Fig.1B).

The distributions of microsatellite type in the assembled genomic sequences of *G. hirsutum*, *G. arboreum*, and *G. raimondii* are presented in Fig. S2A and Table S2. Specifically, the major motifs were rich in A, AT, AAT/AAG, AAAT, AAAAT and AAAAAT, and the minor motifs were mostly riched in C/G (Fig. S2A, Table S2 and S3), which is consistent with previous reports on microsatellites from the assembled genomic sequences of *G. raimondii*\(^4\), *Cucumis sativus*\(^4\), and *Brassica napus*\(^4\). It was noted that the nucleotide composition characteristics of these A/T and C/G motifs corresponded well with their higher A/T than C/G content, such as 34.1% GC content in *G. hirsutum*\(^3\). However, AAT was the most common motif in allotetraploid *G. hirsutum*, whereas AT was most common in diploid *G. arboreum* and *G. raimondii*. More interestingly, we found significantly higher levels of AATCAG in At (2,590) and *G. arboreum* (2,646) compared to Dt (42) and *G. raimondii* (43), but much lower number of AACCCT motif were found in At (60) and *G. arboreum* (73) compared to Dt (332) and *G. raimondii* (421). In addition, the ACAGG repeat was only detected in At (283) and *G. arboreum* (224) (Table S2).

The distributions of microsatellite repeat numbers in the assembled genomic sequences of *G. hirsutum*, *G. arboreum* and *G. raimondii* revealed that microsatellite frequency decreased as the number of repeat units increased (Fig. S2B). Interestingly, as the motif repeat number increased, the microsatellite abundances decreased and the rate of change was slowest for dinucleotides, followed by trinucleotide repeats (Fig. 2). These finding are similar to the research on *Brassica*\(^4\). Although we found that the number of mono- and dinucleotide was dramatically reduced in *G. hirsutum*, high correlations were identified between these genomic sequences based on motif repeat numbers (Table S4).

In conclusion, almost all of the analyzed characteristics of microsatellite distribution in assembled genome of *G. hirsutum* and its two progenitors were highly similar, which suggests that the pattern of microsatellite distribution is conservative in *Gossypium* or retained after formation of allotetraploid cotton. This is understandable because allotetraploid cotton is reunited by *G. arboreum* and *G. raimondii* approximately 1-2 MYA ago\(^1\).

**Genomic distribution.** Based on the assembled genome of *G. hirsutum*, the genomic distributions of microsatellites, genes and transposable elements were investigated. We illustrated greater physical
densities in distal chromosomal regions than in the central regions; in agreement with a previous study that found a greater marker density in the distal gene-rich ends of the chromosomes. Specifically, the genomic distribution of microsatellites was positively correlated with genes and negatively correlated with TEs (Fig. 3, Table S5). For both At and Dt, the frequencies of microsatellites in the 1-Mb genomic intervals were significantly positively correlated with genes ($r = 0.78$ and $0.73$) and negatively correlated with TEs ($r = -0.34$ and $-0.14$). These results are similar to previous reports that microsatellites are associated with gene sequences in plants. It was interesting that the homoeologous chromosomes A05 and D05 exhibited the highest average frequency of microsatellites on the At and Dt, respectively.

### Development of genome-wide microsatellite markers

A total of 77,996 (83.2%), 63,263 (81.46%) and 44,388 (82.8%) identified microsatellite markers were designed from the flanking sequences of 93,736, 77,661, and 53,586 microsatellites from the genomic sequences of *G. hirsutum*, *G. arboreum* and *G. raimondii*, respectively. All of the 77,996 developed microsatellite markers were summarized with motif length, amplification length and polymorphism in Table S6. The remaining microsatellite markers failed to generate specific amplicons mostly due to the limited number of flanking sequences from each side of identified microsatellites. Similar observations have been reported in the mining of genome microsatellites in other crop plants.
The most useful microsatellites are those where the chromosome location is known on the assembled chromosomes. The physical location of 65,498 of the 77,996 microsatellite markers was revealed on the 26 chromosomes of *G. hirsutum*, with average marker density of 33.86 markers per Mb. The average marker density of 40.64 markers/Mb on the D_1 was larger than the 29.33 markers/Mb on A_1. The maximum density (47.07 markers/Mb) was found on D_05, followed by 44.35 markers/Mb on D_07, and 42.27 markers/Mb on D_09, and the minimum density (24.81 markers/Mb) were found on A_06 (Table 2). The physical positions of the developed markers across the 26 chromosomes can provide a high-density microsatellite map that contributes to genome-wide MAS in research of gene mapping, and to comparative genome mapping involving *G. hirsutum* and related crop plants. Only 10,116 (13%) identified microsatellite markers were available in the previous public database (www.cottongen.org), and the remaining 67,880 (87%) were universal.

Although several studies have been conducted to develop genic microsatellite markers from Expressed Sequence Tag (ESTs) transcripts of *G. arboreum* 14, *G. raimondii* 47,48, and *G. hirsutum* 48, the number of publicly available genic microsatellite markers in *Gossypium* was limited (http://www.cottongen.org). We have identified 10,449 (13.4%) genic microsatellite markers known as “functional markers” that have a high transferability across species. Of them, 7,705 were new genic markers by crosschecking within CottonGen and relative literatures. There were more intergenic microsatellites markers than genic markers as a result of 8.64% genic region in *G. hirsutum* genome, although microsatellites are preferentially associated with nonrepetitive DNA in plant genomes 34. Microsatellites in coding regions can regulate gene expression or function, and the mutation rate in coding sequences is lower than in noncoding sequences, therefore, the number of SSRs and polymorphisms is lower in coding regions 49. The main advantage of developing genic microsatellite markers is the possibility of finding associations between functional genes and phenotypes 50,51.

**In silico PCR analysis.** To avoid complicated errors in genotyping due to the polyploidy nature of *G. hirsutum*, all the developed genome-wide microsatellite markers were subjected to in silico PCR analysis based on the genome sequences of *G. hirsutum*, *G. raimondii* and *G. arboreum* (Table 3). As to 77,996 microsatellite markers designed in *G. hirsutum*, 0 (0.0%), 29,392 (37.7%), 20,911 (26.8%), 5,151 (6.6%), and 22,542 (28.9%) markers generated 0, 1, 2, 3, > 3 in silico PCR products from the de novo sequences of *G. hirsutum*, respectively (Table 3 and Table S6). Among all these microsatellite markers, an average of forty eight in silico PCR products were identified, because some element (TE) associated markers could generate tens of thousands of in silico products, as reported in *Brassica* 45. We also found 62,326 (79.9%) markers generating ≤ 10 in silico PCR products with an average of two alleles, and four alleles for 68,811 (88.2%) markers generating ≤ 50 in silico PCR products (Fig. S3).

Through in silico PCR analysis, a set of 20,911 (26.8%) high-quality double-locus microsatellite markers and 29,392 (37.7%) single-locus were established. In diploid species such as barley and rice, most microsatellite markers are single locus amplifying a maximum of two alleles. However, in allotetraploid *G. hirsutum*, most microsatellite markers are found in multiple loci and amplify multiple alleles from Chr.

| Chr. | Marker mapped | Density (Per Mb) | Chr. | Marker mapped | Density (Per Mb) |
|------|---------------|-----------------|------|---------------|-----------------|
| A01  | 2612          | 26.15           | D01  | 2452          | 39.90           |
| A02  | 2345          | 28.10           | D02  | 2491          | 37.02           |
| A03  | 2592          | 25.85           | D03  | 1849          | 39.60           |
| A04  | 1682          | 26.74           | D04  | 2038          | 39.61           |
| A05  | 3373          | 36.64           | D05  | 2915          | 47.07           |
| A06  | 2560          | 24.81           | D06  | 2430          | 37.79           |
| A07  | 2571          | 32.86           | D07  | 2453          | 44.35           |
| A08  | 2794          | 26.96           | D08  | 2606          | 39.55           |
| A09  | 2405          | 32.07           | D09  | 2166          | 42.47           |
| A10  | 2884          | 28.59           | D10  | 2438          | 38.47           |
| A11  | 3028          | 32.45           | D11  | 2712          | 41.04           |
| A12  | 2732          | 31.23           | D12  | 2493          | 42.18           |
| A13  | 2447          | 30.60           | D13  | 2430          | 40.14           |
| All  | 65498         | 33.86           |      |               |                 |

Table 2. Summary of chromosomal distribution and average density of microsatellite markers mapped on *G. hirsutum* chromosomes. Chr.: Chromosome A_1, D_1; two subgenomes of allotetraploid cotton *G. hirsutum*.
Table 3. Generated number (%) of in silico PCR products by genome-wide microsatellite markers in the sequenced genome of *G. hirsutum*, *G. raimondii* and *G. arboreum*.

| Markers from | in silico PCR in | Zero | One | Two | Three | > Three | Total |
|--------------|------------------|------|-----|-----|-------|---------|-------|
| *G. raimondii* | *G. arboresum* | 26647(60.0) | 12706(28.6) | 1326(3.0) | 523(1.2) | 3186(7.2) | 44388 (100) |
| *G. hirsutum* | 6532(14.7) | 14134(31.8) | 11553(26.0) | 2345(5.3) | 9824(22.1) | |
| *G. raimondii* | 42616(67.4) | 14973(23.7) | 1238(2.0) | 545(0.9) | 3891(6.2) | |
| *G. arboresum* | 0(0.0) | 37060(58.6) | 4543(7.2) | 1761(2.8) | 19899(31.5) | 63263 (100) |
| *G. hirsutum* | 6694(10.6) | 19880(31.4) | 14138(22.3) | 3061(4.8) | 19490(30.8) | |
| *G. raimondii* | 32868(42.1) | 30588(39.2) | 2876(3.7) | 1391(1.8) | 10273(13.2) | |
| *G. arboresum* | 25026(32.1) | 33406(42.8) | 3388(4.3) | 1501(1.9) | 14675(18.8) | 77996 (100) |
| *G. hirsutum* | 0(0.0) | 29392(37.7) | 20911(26.8) | 5151(6.6) | 22542(28.9) | |
Identification of microsatellites. Genome sequences were searched for perfect microsatellites using PRL5 script MicroSatellite (MISA, http://pgrc.ipk-gatersleben.de/misa/) with basic motifs from mono- to hexanucleotide\(^1\). Repeats with a minimum of 18, 9, 6, 5, 4, and 3 were defined for the mono- to hexanucleotide, respectively. Compound microsatellites were defined as ≥2 repeats interrupted by ≤100 bp, as previous report\(^2\).

Statistical analysis. Each chromosome was divided into 1-Mb for statistical analysis of microsatellites, genes, and TEs for the represent practical frequencies and average frequencies. Then, Excel statistical function CHISQ.TEST was used to calculate the significance level (\(P_{\text{corr}}\)) of these two frequencies of microsatellites as well as genes and TEs in 26 chromosomes.

Design of SSR primers. Primer pairs were designed from the flanking sequences of identified microsatellites using PRIMER3 software\(^3\), and two perl scripts, p3_in.pl and p3_out.pl served as interface modules between MISA and Primer3 with the primer designing parameters: 18–27 bp in length, 57–63 °C in melting temperature, 30–70% in GC content and 100–280 bp in product size. These two perl scripts were downloaded from MISA (http://pgrc.ipk-gatersleben.de/misa/). Primer3 was downloaded from http://www-genome.wi.mit.edu/genome_software/other/primer3.html. The p3_in.pl was used to create a primer3 input file which was submitted to Primer3. Then p3_out.pl was used to calculate and merge all information together.

In silico analysis of microsatellite polymorphisms. The primer-pair sequences of previously developed publicly available Gossypium SSR markers were downloaded from the CottonGen website (www.cottongen.org). In our study, all microsatellite markers were aligned to genomic sequences of G. hirsutum, G. arboreum and G. raimondii using an in silico PCR strategy with the following default parameters: 2bp mismatch, 1bp gap, 50bp margin and 50–1000 bp product size, as previous report\(^4\),\(^5\). The software (e-PCR-2.3.11) used for in silico PCR was downloaded from ftp://ftp.ncbi.nlm.nih.gov/pub/schuler/e-PCR/. And only one genome was used at a time.

Evaluation of polymorphisms in G. hirsutum and G. barbadense. Illumina reads of G. barbadense cv. Hai7124 (Biosample: SAMN03002317) were mapped to the genome of G. hirsutum acc. TM-1 using the Burrows-Wheeler Alignment tool (BWA ver 0.6.2)\(^6\), with mainly default parameters. SAMtools\(^7\) was used to call InDels ≥5 with mapped reads of a minimum mapping quality of 20. If InDels were located in the region of paired primers/microsatellite markers, we regarded the paired primer as a putative polymorphic primer between TM-1 and Hai7124.

Data Section. The detail information of 77,996 developed markers and in silico PCR amplification were deposited in Table S6. These markers will be publically available in http://mascotton.njau.edu.cn and CottonGen database.

References
1. Gupta, P. & Varshney, R. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163–185 (2000).
2. Powell, W., Machray, G. & Provan, J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1**, 215–222 (1996).
3. Agarwala, M., Shrivastava, N. & Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* **27**, 617–631 (2008).
4. Cavagnaro, P. et al. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* **11**, 569 (2010).
5. Zhang, H., Li, Y., Wang, B. & Chee, P. Recent advances in cotton genomics. *Int. J. Plant Genom.* **2008**, 742304 (2008).
6. Zhu, H. et al. Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon Ait*). *Theor. Appl. Genet.* **124**, 87–96 (2012).
7. Percival, A., Stewart, J. & Wendel, J. Taxonomy and germplasm resources. In: *Cotton: Origin, History, Technology, and Production* (Smith, C.W. & Cothren, J.T., eds) pp. 33–63. New York: John Wiley & Sons (1999).
8. Fryxell, P. *The natural history of the cotton tribe* (*Malvaceae, tribe Gossypieae*), pp. 245. College Station: Texas A & M University Press (1978).
9. Endrizzi, J., Turcotte, E. & Kohel R. Qualitative genetics, cytology, and cytogenetics. In: *Cotton, (Kohel, R.), & Lewis, C.F., eds*, pp.81–129, Madison: University of Wisconsin Press (1984).
10. Wendel, J. New World tetraploid cottons contain Old World cytoplasms. *Proc. Natl. Acad. Sci. USA* **86**, 4132–4136 (1989).
11. Hoffman, S. et al. Identification of 700 new microsatellite loci from cotton (*G. hirsutum* L.). *J. Cotton Sci.* **11**, 208–241 (2007).
12. Lacape, J. M., Dessaup, D., Rajab, M., Noyer, J. L. & Hau, B. Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol. Breeding* **19**, 45–58 (2007).
13. Edwards, K., Barker, J., Daly, A., Jones, C. & Karp, A. Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* **20**, 758–760 (1996).
14. Han, Z. et al. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor. Appl. Genet.* **112**, 430–439 (2006).
15. Park, Y. et al. Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol. Genet. Genomics* **274**, 428–441 (2005).
16. Frechtlowski, I. et al. Cotton genome mapping with new microsatellites from *Acala ’Maxxa’* BAC-ends. *Mol. Genet. Genomics* **275**, 479–491 (2006).
17. Guo, Y. et al. BAC-derived SSR markers chromosome locations in cotton. *Euphytica* **161**, 361–370 (2008).
18. Zhang, X. et al. Characterization of the global transcriptome for cotton (*Gossypium hirsutum* L) anther and development of SSR marker. *Gene* **551**, 206–213 (2014).
This work was financially supported in part by grants from the Major State Basic Research Development Program of China (973 Program, 2011CB109300), the National High Technology Research and
Development Program of China (863 Program) (2011AA10A102, 2012AA101108-04-04), National Key Technology Support Program (2015BAD02B00), the Priority Academic Program Development of Jiangsu Higher Education Institutions and the 111 project (B08025).

**Author Contributions**
T.Z.Z. conceived and designed the experiments. Q.W., L.F. and J.D.C. performed the experiments. Y.H., Z.F.S., S.W. and L.J.C. participated in part of materials and libraries preparation. Q.W. and L. F. analyzed the results. T.Z.Z., L.F. and W.Z.G. drafted the manuscript.

**Additional Information**

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, Q. et al. Genome-Wide Mining, Characterization, and Development of Microsatellite Markers in *Gossypium* Species. *Sci. Rep.* 5, 10638; doi: 10.1038/srep10638 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/