A Whole Genome DNA Marker Map for Cotton Based on the D-genome Sequence of *Gossypium raimondii* L.

Zining Wang¹, Dong Zhang¹, Xiyin Wang, Xu Tan, Hui Guo, Andrew H. Paterson²
Plant Genome Mapping Laboratory, University of Georgia, 111 Riverbend Road, Athens GA, 30605
A Whole Genome DNA Marker Map for Cotton Based on the D-genome Sequence of *Gossypium raimondii* L.

Key Words: quantitative trait loci, resistance gene analog, simple sequence repeat, restriction fragment length polymorphism, inversions

1 These authors contributed to this work equally.

2 Corresponding author: Plant Genome Mapping Laboratory, University of Georgia, 111 Riverbend Road, Rm 228, Athens GA 30605, Phone 1-706-583-0162, Fax 1-706-583-0160, Email paterson@plantbio.uga.edu

Abstract

We constructed a very high density whole genome marker map (WGMM) for cotton, using 18,597 DNA markers corresponding to 48,958 loci that were aligned to both a consensus genetic map and a reference genome sequence. The WGMM has a density of one locus per 15.6 kb, or an average of 1.3 loci per gene. The WGMM was anchored using colinear markers to a detailed genetic map, providing recombinational information. Mapped markers occurred at relatively higher physical densities in distal chromosomal regions, and lower physical densities in the central regions, with all 1 Mb bins having at least 9 markers. Hotspots for quantitative trait loci (QTL) and resistance gene analog (RGA) clusters were aligned to the map and DNA markers identified for targeting of these regions of high practical importance. Based on the cotton D genome reference sequence, the locations of chromosome structural rearrangements plotted on the map facilitate its translation to other *Gossypium* genome types. The WGMM is a versatile genetic map for marker assisted breeding, fine mapping and cloning of genes and QTLs, developing new genetic markers and maps, genome wide association mapping and genome evolution studies.
**Introduction**

Genetic mapping is an essential prerequisite for the activities of marker assisted selection, gene/QTL cloning, genome sequence assembly, association mapping and evolutionary studies (Duran *et al.* 2009). Genetic marker systems such as RFLP, SSR, SRAP, and others have been widely used in linkage and QTL mapping (Agarwal *et al.* 2008) in population sizes of up to a few hundred individuals, generally involving highly-divergent parents and in strong linkage disequilibrium such that a few hundred markers provided adequate information. Whole genome genotyping methods that are now being reduced to routine and cost-effective practice (Xie *et al.* 2010; Andolfatto *et al.* 2011; Elshire *et al.* 2011; Bus *et al.* 2012) open the door to investigations such as global trait mapping and association that require vastly higher DNA marker densities (Poland *et al.* 2011), and are likely to render many prior marker systems obsolete. However, the mapping information garnered from prior systems remains of value, with individual QTLs and particularly meta-analyses (Rong *et al.* 2007; Zhang *et al.* 2013) identifying genomic regions that might be searched at high resolution for causal genes.

High-contiguity ‘reference’ genome sequences provide a natural platform for unifying information from a range of sequence-tagged DNA marker systems, toward efficient application of new approaches to build upon collective knowledge of the biology of an organism. Most major crops now have a reference genome sequence, and some authors have projected that within a few years all of the ~200 widely-used domesticates will have such a resource (Paterson 2006). For the organism that we focus on herein, cotton (*Gossypium*), the smallest of 8 genome types (A-G, and K) and a suspected progenitor of cultivated polyploids, the D-genome of *Gossypium raimondii*, was recently sequenced (Paterson *et al.* 2012; Wang *et al.* 2012). A high degree of colinearity between various genomes in the Gossypium genus (Reinisch *et al.* 1994; Brubaker *et al.* 1999; Han *et al.* 2004; Rong *et al.* 2004; Desai *et al.* 2006; Becerra Lopez-Lavalle *et al.* 2011), suggests that the D reference genome will extrapolate well to most regions of most other Gossypium genomes.
In cotton, more than 30 genetic maps have been published, largely based on interspecific crosses between two species that are each domesticated but are treated by breeders as different gene pools, *G. hirsutum* × *G. barbadense* (Jiang et al. 1998; Zhang et al. 2002; Lacape et al. 2003; Nguyen et al. 2004; Rong et al. 2004; Guo et al. 2007; He et al. 2007; Yu et al. 2007; Lacape et al. 2009). The interspecific tetraploid genetic maps were valuable for finding new DNA markers at a time that primary genetic mapping was a high priority (Reinisch et al. 1994; Nguyen et al. 2004; Rong et al. 2004; Guo et al. 2007; Zhang et al. 2008; Yu et al. 2011).

However, conventional breeding programs seldom utilize such wide crosses (Ulloa and Meredith 2000; Ulloa et al. 2002) – and intraspecific crosses, largely within *G. hirsutum* (‘Upland’ cotton) suffered from a paucity of DNA polymorphism that often left much of the genome unmapped (Shappley et al. 1998a, b; Ulloa and Meredith 2000; Ulloa et al. 2002; Rong et al. 2005a; Shen et al. 2005, 2007; Wang et al. 2007; Chen et al. 2008; Xu et al. 2008; Zheng et al. 2008; Liu et al. 2011).

Increasing the DNA marker density of cotton genetic maps is urgent for marker assisted selection and genomic studies. This need has been widely recognized, and several efforts have interleaved existing maps based on subsets of shared DNA markers to form ‘consensus’ (Rong et al. 2005b) or ‘integrated’ maps (Yu et al. 2010; Reddy et al. 2011; Blenda et al. 2012) with as many as 8,254 loci (Blenda et al. 2012).

In this study, we constructed a whole genome marker map (WGMM) by integrating publicly-available sequence tagged DNA markers with the cotton D-genome sequence. Chromosomal affiliations were deduced based on DNA markers derived from diploid and tetraploid cotton genetic maps (Rong et al. 2004), noting known differences among the genomes and subgenomes that should be considered when using the cotton markers and maps. ‘Hotspots’ for QTLs previously-implicated in fiber development, and clusters of resistance-gene analogs (RGAs) identified in the genome sequence, were also aligned with the WGMM. The WGMM provides a foundational tool and resources for increasing knowledge of fundamental and
applied elements of cotton biology, via marker assisted breeding, fine mapping and cloning genes and QTLs, genome wide association mapping, and other studies.

Materials and methods

The materials used in this study include the D-genome sequence for *G. raimondii* (Paterson et al. 2012), a consensus genetic map and diploid D-genome genetic map (Rong et al. 2005a), tetraploid At and Dt maps (Rong et al. 2004; 2005a), a cotton QTL meta-analysis (Rong et al. 2007), and the Cotton Marker Database ([http://www.cottonmarker.org/](http://www.cottonmarker.org/)).

Construction of WGMM

The availability of sequence-tagged markers such as restriction fragment length polymorphism (RFLP) probes and simple sequence repeats (SSR) in the Cotton Marker Database ([http://www.cottonmarker.org/](http://www.cottonmarker.org/)) and a cotton consensus map (Rong et al. 2005a) provides alignable information to convert genetic positions (in centiMorgans, cM) of markers to physical positions (bp).

After marker sequences were prepared, Blastn (Altschul et al. 1990) was applied to anchor markers to the cotton D-genome pseudo molecules. Markers with alignments of E-value <= 1e-10 for RFLP/SSR sequences and <= 50 for SSR primers were assembled into loci. For RFLPs, the alignments with distance <= 5000bp were assembled into one RFLP locus. For SSR primers, one forward primer hit was combined with one reverse primer hit if the distance between the two hits was <= 1000bp.

Colinearity between genetic and physical positions was determined by ColinearScan 1.0.1 (Wang et al. 2006). The colinear markers aligned on the D-genome molecules were maintained as anchoring markers with their original genetic distance. The genetic distances of the non-colinear markers were estimated based on the genetic distance between the anchoring markers.
and the physical locations of the markers. Strikingly discrepant loci were removed, based on the order of markers in the original source. A QTL region was delineated by two flanking markers nearest to the likelihood peak that had alignment information.

**Identification of RGA clusters**

All cotton proteins were used to search for nucleotide binding site (NBS) domains (PF00931, NB-ARC) by a Hidden Markov Model method (Eddy 1998) implemented in hmmsearch version 3.0 with e-value cutoff = 1. To filter false positive hits, all identified NBS containing proteins were screened against the Pfam-A file (Bateman *et al.* 2004). NBS domains that overlap with other domains with lower E-value were considered false hits and abandoned. Likewise, the TIR domain (PF01582) was searched against all cotton proteins by hmmsearch with e-value cutoff = 1, and putatively false hits abandoned. To detect LRR motifs, predicted NBS encoding proteins were searched against 10 LRR families in the LRR clan (CL0022) with e-value cutoff = 1. All regions predicted as LRR motifs and not overlapping with other domains with lower e-value were inferred to be real LRR motifs. Coiled coil (CC) motifs were detected using NCOILS with default parameters. CCs at the N-terminus of NBS domains and not overlapping with other domains were considered to identify CC-NBS type genes. Only the RGA clusters (clusters containing only RGAs) were selected for this study.

**Dotplots**

To compare the genetic maps constructed for diploid D, tetraploid At and Dt, and a consensus map (Rong *et al.* 2004, 2005a), the markers in these maps were mapped onto the cotton pseudomolecules by running Blastn using their sequences against the constructed cotton assembly, at the criteria of e-value < 1e-10. Dotplots of genetic maps and the pseudomolecules were generated using an in-house perl program (similar to those available at http://chibba.agtec.uga.edu/duplication/).

**Results**

**Marker number and density on chromosomes**
The total of 20,096 sequence-tagged cotton DNA markers were first filtered for sequence/primer duplications. In total, 18,597 non-redundant markers (Table 1) were used to do Blastn searches against the D-genome pseudomolecules and 79,481 sequence alignments could be aligned to the D genome. After filtering clustered duplicate copies within 1 kb for SSR and 5 kb for RFLP, 48,546 loci were mapped on the 13 chromosomes of WGMM and 412 loci on the unassembled scaffolds (Table S1). The 18,597 markers used for the alignments have an average of 2.63 copies per marker sequence, ranging from 6493 (34.91%) markers with a single copy to 1195 markers with four copies (6.43%; Table 1). Among the 48,958 mapped loci, 42,794 (87.41%) are SSRs and 6164 (12.59%) are RFLPs. Loci from markers with five copies accounted for the largest share (45.81%) of the 48,958 loci, and the loci from four copy-markers had the smallest share (9.76%).

Marker density was closely correlated with the physical DNA content of chromosomes (r=0.81; p=0.0004), with size differences among the 13 chromosomes reflected by differences in locus number (Table S2). D12, with 34.9 Mb DNA, has the fewest loci mapped (2,386); D09 with 70.7 Mb has the most (5,370). Marker density along a chromosome ranges from 12.8 to 18.5 Kb per locus (D07, D10), with an average of 15.6 kb per locus. Relative to the 37,505 genes in the current D genome annotation, the WGMM provides 1.3 marker loci per gene.

**Marker distribution on chromosomes**

With one exception, the cotton chromosomes have relatively higher marker densities of 100-220 markers per 1 Mb in terminal chromosomal regions that are also gene-rich (Figure 1), and significantly lower marker densities (50-120 loci per Mb) in central regions that are gene-poor and repeat-rich (See the gene and repeat density heatmap in Figure 1). A few higher peaks appear in the middle regions of some chromosomes that are also gene rich.

In order to assess the coverage of this map, we determined the number of markers in each of the 34-69 1 Mb windows of D01-D13. The least-populated window (on D02) had 9 markers and a physical distance of 110 kb between markers, indicative of good coverage of the whole
genome by these markers. The WGMM covered 98.11% of the 747 Mb assembled portion of the cotton D genome, out of total 761.4 Mb (Paterson et al. 2012).

One unusual chromosome, D06, had high marker density consistent with the remaining chromosomes at one end, but unusually low density at the other end that was more consistent with the central regions of other chromosomes (Figure 1). The unusual chromosome arm (0-27 Mb) had lower gene and higher repeat densities that may have been the main reasons for the lower marker density, noting that most markers were SSRs developed from expressed sequence tags. We also searched the ribosomal DNA (rDNA) sequences. The rDNA (18S, 5S and 25S) was mainly distributed on 3 chromosomes, with 40 rDNA genes in a 249 kb region (20789590-21038157 bp) and a 544 kb region (32723102-33266822 bp) on D08, 45 rDNA genes in a 400 kb region (43290160-43692820 bp) on D09 and 111 rDNA genes in a 362 kb region (55503532-55865427 bp) on D01 (Figure 1). Another 33 rDNA genes were scattered, with 1-11 gene copies on D02, 03, 04, 05, 06, 07, 10, 12, and 13. The rDNA genes caused lower marker density in local regions that were too short to be labeled in the figure, and did not significantly influence marker and gene density.

**Marker alignment from consensus map to WGMM**

To integrate recombination patterns into the WGMM, it was aligned with a consensus genetic map of 2325 cM (Table S2) that was used in assembling the D genome pseudo molecules, and has been confirmed to be a good representation of At and Dt genetic maps (Table S1; Table S2; Rong et al. 2004). Altogether, 1725 markers from the consensus map could be mapped on the WGMM, with 86 to 220 per chromosome (Table S3). About 60% of the mapped markers showing colinearity between physical and genetic locations on the D genome were used as anchor markers on WGMM and their genetic distances transferred to the WGMM. Genetic locations of non-anchor markers on the WGMM were interpolated by using the genetic and physical spacings of flanking markers.

**Cotton fiber QTL hotspots**
From a meta-analysis of 432 cotton QTLs (Rong et al. 2007), we mapped to the WGMM 104 fiber QTLs that comprised 18 ‘hotspots’ containing QTLs for 3-7 fiber traits including fiber elongation, color, fineness, length, strength, uniformity, micronaire and short fiber content (Table 2). Chr.07 and Chr.23 each have the largest number of 16 QTLs. Table 2 lists the interval locations and the number of markers in the intervals on D-genome molecules.

**Cotton RGA hotspots**

A total of 63 RGA clusters were identified on 9 of the 13 D-genome chromosomes (Table S4). D07 has the largest number of clusters (21), while D04 has the smallest (1, with only 2 RGAs). The cluster positioned at 51,589,128 bp on D09 is the largest cluster with 7 RGAs. Because there is no other gene in the RGA clusters, some clusters with only two RGAs have no markers within the interval while the largest cluster has 5 markers. Cluster positions and their closest flanking markers are listed (Table S4) to facilitate the use of these markers in expedited searches for loci responsible for disease and/or pest resistance phenotypes.

**Comparative map alignments of Di, At, Dt to the D genome sequence**

To further enhance alignment of the WGMM to prior genetic information, we aligned the genetic maps of the diploid D (Rong et al. 2004), tetraploid At and Dt (Reinisch et al. 1994; Lacape et al. 2003; Rong et al. 2004), with the D-genome pseudo molecules. To avoid confusion, Di represents the D genome genetic map and D the D-genome sequence in the text, tables and figures hereafter. All chromosomes from each genetic map could be aligned to the WGMM with marker numbers per chromosome of 21-65 (Di), 20-93 (At), and 38-95 (Dt). The numbers of aligned markers on the corresponding and other chromosomes are listed (Table S3).

**Genome variations**

To study the colinearity and genome variations between the genetic maps and the D genome sequence, dotplots were performed between the Di, At and Dt genetic maps and the D genome pseudomolecules (Figure 2a-b). Here, we listed the large inversions (>1 Mb) different from
those found in Rong et al. (2004) based on the comparison of genetic maps. The D genome diploid genetic map had been used to construct the pseudomolecules, and therefore, the dotplot between them showed good consistency as expected (Figure 2a). The Dt genetic map also exhibited colinearity with 11 of the pseudo molecules, with Chr.15 having an inversion (Figure 2b, red oval) of 14.71 Mb accounting for 23.44% of D02 (Table 3). Four inversions (Figure 2b) relative to At chromosomes with sizes ranging from 3.33-6.27 Mb were identified on Chr.04, 10, 12 and LGA03, accounting for 5.36-9.95% of the corresponding chromosomes.

The At genomes displayed two reciprocal chromosome arm translocations (Rong et al. 2004) between Chr. 02/03 and between Chr. 04/05 (red arrows in Figure 2b) and their physical location information on the D-genome sequence was added here. The translocations broke the chromosomes near centromere regions, such as Chr.02/03 corresponding to pseudo molecule D03 at 24043418-29931543 bp (PAR0499-A1171) and Chr.03/02 to D05 at 23203279-34341901 bp (A1325-Unig26F10); Chr.04/05 to D09 at 46366195-39825818 bp (Unig27B06-Gate4DF07) and Chr.05/04 to D12 at 21985484-24460316 bp (Unig06C12-Coau2105).

Table 3 lists the flanking markers for the five inversions with their locations on both the genetic maps and the WGMM.

**Discussion**

The cotton whole genome marker map described herein has a total of 48,959 loci, about 6 times the number characterized in the most richly-populated of the integrated maps published previously (Blenda et al. 2012). Comparable to a linkage map of rice based on population sequencing with 15,795 SNPs (Xie et al. 2010), and a Brassica genetic map of 13,551 SRAP markers (Sun et al. 2007), this map may facilitate fine mapping, gene cloning, global association mapping of cotton genes and traits, and other genomic studies.

**Distribution of markers**
Genetic maps always face the problem of unevenly distributed markers and the resulting gaps. During meiosis, recombination does not happen evenly on the chromosomes. Further, marker sequences are not randomly dispersed, especially for sequence repeats such as SSRs. In centromeric and intergenic regions, more sequence repeats were found (Paterson et al. 2012). Previous genetic maps illustrated the uneven distribution of markers with many large gaps; on a recombinational scale, distal chromosomal regions tended to have lower marker density than the centromere regions (Rong et al. 2004). On the physical scale used in this study, the distal gene-rich ends of the chromosomes clearly have higher marker density, a natural outcome of the fact that many of the markers used are gene-derived. Not a single 1 Mb window in any chromosome had fewer than 9 markers (i.e. an average of one per 110 kb), and most had at least 50 markers, guaranteeing the availability of established DNA markers in any region of interest.

Alignment of QTL and RGA hotspots
Global genotyping is the trend for dissecting genes/QTLs controlling important phenotypes, and prior QTL information provides valuable evidence toward validation of statistically significant associations from genome-wide association studies. The cotton community has identified at least hundreds of QTLs for fiber related traits, plant architecture, disease resistance and stress tolerance [a meta-analysis now 6 years old already identified 432 QTLs (Rong et al. 2007)]. Here, we provide resources to quickly and efficiently target QTL ‘hotspots’ for fiber related or disease/pest related traits for rapid characterization, for example in searches for novel alleles. For example, genotyping of as few as 126 DNA markers would permit one to assess co-segregation at DNA markers flanking with 53% (159/300) of the resistance gene analogs in the cotton reference sequence.

Genome variations revealed by comparative analysis of maps
While the various cotton genomes have a high degree of colinearity, several rearrangements known from prior studies (Rong et al. 2004) and the five possible inversions identified here need to be accounted for in ‘translation’ of information from the D-genome based WGMM to
other cotton genomes. The At genome experienced two reciprocal chromosome arm translocations, between Chr.02 and 03 and between Chr.04 and 05, with break points near the centromere regions. These genome variations may facilitate the understanding of genome evolution and gene/QTL cloning from the related orthologous regions. Further to characterize the variations, the WGMM provides a foundation for and will facilitate the investigation of the relationships of phenotype and genotype of important agronomic traits, especially those controlled by A and At genomes. We noted that some of the newly identified and relatively smaller inversions could represent mis-assemblies of the reference genome sequence.

In conclusion, the consensus high density WGMM is a valuable resource with the potential for adding additional value as more information accumulates such as better knowledge of QTL hotspots for cotton fiber development; roles of different RGA clusters in conferring pest resistance genes and QTLs; global association studies of cotton; and/or genome structure and variation.

Acknowledgements

A.H.P. appreciates financial support from the U. S. National Science Foundation (DBI 0817707, IIP-0917856, IIP-1127755), USDA National Research Initiative (00-52100-9685, 02-35301-12045), Cotton, Inc, and the Pakistan-US Science and Technology Cooperation Program.

References


Guo, W., C. Cai, C. Wang, Z. Han, X. Song et al., 2007 A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in Gossypium. Genetics 176: 527-541.


Shappley Z, J. N. Jenkins, J. Zu, and J. C. J. McCarty, 1998b Quantita- D-subgenome bias of Xcm resistance genes in tetraploid gossypium tive trait loci associated with agronomic and fiber


Xie, W., Q. Feng, H. Yu, X. Huang, Q. Zhao et al., 2010 Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci U S A 107: 10578-10583.


Yu, Y., D. Yuan, S. Liang, X. Li, X. Wang et al., 2011 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.


Table 1 Numbers and percentages of markers and loci on the WGMM

<table>
<thead>
<tr>
<th>#Copy</th>
<th>Markers</th>
<th>%</th>
<th>Loci</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6,493</td>
<td>34.91</td>
<td>6,493</td>
<td>13.26</td>
</tr>
<tr>
<td>2</td>
<td>4,012</td>
<td>21.57</td>
<td>8,024</td>
<td>16.39</td>
</tr>
<tr>
<td>3</td>
<td>2,412</td>
<td>12.97</td>
<td>7,236</td>
<td>14.78</td>
</tr>
<tr>
<td>4</td>
<td>1,195</td>
<td>6.43</td>
<td>4,780</td>
<td>9.76</td>
</tr>
<tr>
<td>5</td>
<td>4,485</td>
<td>24.12</td>
<td>22,425</td>
<td>45.81</td>
</tr>
<tr>
<td>Total</td>
<td>18,597</td>
<td>100</td>
<td>48,958</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 2 QTLs from At and Dt tetraploid chromosomes with their consensus locations and markers on the D genome and WGMM

<table>
<thead>
<tr>
<th>Chr.</th>
<th>#QTL</th>
<th>D Chr.</th>
<th>Start bp</th>
<th>End bp</th>
<th>#Markers</th>
<th>QTL traits</th>
<th>QTL names</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5-6</td>
<td>D02</td>
<td>1390547</td>
<td>20704735</td>
<td>1344</td>
<td>EL, FF, FL, FS, FU, MIC</td>
<td>FL01.1(HVU2.5, HVuhm), EL01.1, FU01.1(HVU2.5), FS01.2, EL01.1, MIC01.1, FF01.2, EL01.2, FF15.1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>D03</td>
<td>3185870</td>
<td>35433985</td>
<td>1387</td>
<td>FL, FS, FU, SF</td>
<td>FL03.1(HVuhm, HVU2.5), FU03.1(HVuhm), SF03.1(SFCn), FS02.1, FS02.2</td>
</tr>
<tr>
<td>5</td>
<td>6-7</td>
<td>D05</td>
<td>4002919</td>
<td>49222777</td>
<td>2191</td>
<td>EL, FF, FL, FU, MIC</td>
<td>FL02.1, FF02.1, EL02.1(HVU2.5, HVuhm), FU02.1(HVU2.5, MIC02.1), FL02.2, FF02.2, MIC02.1, FF02.3</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>D07</td>
<td>3099961</td>
<td>5211929</td>
<td>2529</td>
<td>EL, FF, FL, MIC</td>
<td>FLA03.1(HVuhm, HVU2.5), FU03.1(HVuhm), MIC03.1, FF03.2</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>D07</td>
<td>13773747</td>
<td>19329003</td>
<td>485</td>
<td>EL, FF, FL, FU</td>
<td>FLA03.1(HVuhm, HVU2.5), FF03.2, FLA03.2, FF03.1, FU03.1, FLA03.2</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>D07</td>
<td>28919943</td>
<td>35433985</td>
<td>1387</td>
<td>FL, FS, FU, SF</td>
<td>FL03.1(HVuhm, HVU2.5), FU03.1(HVuhm), SF03.1(SFCn), FS03.1, FF03.2, SF03.2</td>
</tr>
<tr>
<td>8</td>
<td>4-6</td>
<td>D08</td>
<td>41534671</td>
<td>55710457</td>
<td>1450</td>
<td>EL, FF, FL, FU</td>
<td>FF03.1, FL03.1, FL03.1, FLA03.1, FF03.2, FF03.1, FLA03.2, FL03.1, SF03.1</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>D09</td>
<td>11863128</td>
<td>17291302</td>
<td>750</td>
<td>EL, FF, FS, FU, MIC</td>
<td>FF05.15, MIC05.1, FS05.2, FL05.05(FE), FF05.3</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>D01</td>
<td>17773028</td>
<td>45685292</td>
<td>1433</td>
<td>EL, FF, FL, MIC</td>
<td>FU16.1(FU), FL16.1, FF16.1, MIC16.1</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>D03</td>
<td>42011483</td>
<td>45719191</td>
<td>386</td>
<td>FC, FL, FU, MIC</td>
<td>FL17.1(HVU2.5, HVuhm), FF17.1(HVuhm), MIC17.1, FC17.1</td>
</tr>
<tr>
<td>22</td>
<td>4-5</td>
<td>D04</td>
<td>50443620</td>
<td>58823186</td>
<td>803</td>
<td>EL, FF, FL, FU</td>
<td>FD03.05, FD03.1, FL03.1(HVU2.5, HVuhm, FLU), FD03.1(HVuhm), EL03.1(FLONG)</td>
</tr>
<tr>
<td>LGD08</td>
<td>5-6</td>
<td>D05</td>
<td>39041081</td>
<td>62380583</td>
<td>1366</td>
<td>FF, FL, FS, FU, MIC</td>
<td>FD04.05, FD04.1, FL04.1(HVU2.5, HVuhm, MIC04.2), FF04.1, FS04.1, SS04.1</td>
</tr>
<tr>
<td>26</td>
<td>4-6</td>
<td>D06</td>
<td>37014370</td>
<td>47605556</td>
<td>1238</td>
<td>EL, FF, FL, FU</td>
<td>EL23.03, FF23.1, EL23.03(FE), EL23.1, FU23.1, HVU23.1, FS23.2</td>
</tr>
<tr>
<td>16</td>
<td>4-6</td>
<td>D07</td>
<td>17026813</td>
<td>48701291</td>
<td>1541</td>
<td>EL, FF, FL, FS</td>
<td>FF02.1, FL02.1(HVU2.5), FD02.1, FD02.1, FD02.1, FD02.1</td>
</tr>
<tr>
<td>LGD03</td>
<td>4-6</td>
<td>D08</td>
<td>37154657</td>
<td>55710457</td>
<td>1812</td>
<td>EL, FF, FS, FU, MIC</td>
<td>FF08.12, EL08.1, FF08.1, EL08.1, FD08.1, FL08.1, MIC08.1, FD08.1, FD08.1</td>
</tr>
<tr>
<td>23</td>
<td>9</td>
<td>D09</td>
<td>49420986</td>
<td>9480129</td>
<td>552</td>
<td>EL, FF, FS, FU, MIC</td>
<td>EL09.1, FF09.1, FL09.1, MIC09.1, FS09.1, FD09.1, EL09.1, FF09.1</td>
</tr>
<tr>
<td>23</td>
<td>5-7</td>
<td>D09</td>
<td>12049737</td>
<td>29728227</td>
<td>1773</td>
<td>EL, FF, FL, MIC, SF</td>
<td>FD09.3, MIC09.2, EL09.3, EL09.4, FD09.2, EL09.2, MIC09.2, FD09.2</td>
</tr>
<tr>
<td>LGD02</td>
<td>5</td>
<td>D11</td>
<td>2519220</td>
<td>3161683</td>
<td>57</td>
<td>EL, FF, FL, SF</td>
<td>EL20.1, FL20.1(HVU2.5, FLU), SF20.1, SF02.1, FL20.1, SF02.1</td>
</tr>
</tbody>
</table>

*EL: elongation; FC: color; FF: fineness; FL: length; FS: strength; FU: uniformity; MIC: micronaire; SF: short fiber content
Table 3 Genome variations identified via comparative mapping against D genome pseudomolecules

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Genetic interval (cM)</th>
<th>cM</th>
<th>Total cM</th>
<th>% of chr.</th>
<th>Start Marker</th>
<th>End Marker</th>
<th>Physical interval (Mb)</th>
<th>Mb</th>
<th>Chr. length (Mb)</th>
<th>% of Chr.</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr.15-D02</td>
<td>169.9-123.2</td>
<td>46.7</td>
<td>176.4</td>
<td>26.47</td>
<td>PAR08C07</td>
<td>PAR0935</td>
<td>17037231-2325639</td>
<td>14.71</td>
<td>62.75</td>
<td>23.44</td>
<td>Inversion</td>
</tr>
<tr>
<td>LGA03-D07</td>
<td>69.4-84.4</td>
<td>15</td>
<td>198</td>
<td>7.57</td>
<td>Gate4DB08</td>
<td>Gate1BF07</td>
<td>28319670-32633603</td>
<td>4.31</td>
<td>60.74</td>
<td>7.09</td>
<td>Inversion</td>
</tr>
<tr>
<td>Chr.12-D08</td>
<td>55.4-59.0</td>
<td>3.6</td>
<td>213.6</td>
<td>1.69</td>
<td>ESTS146</td>
<td>Coau2J13</td>
<td>39347694-43302432</td>
<td>4.00</td>
<td>58.26</td>
<td>6.87</td>
<td>Inversion</td>
</tr>
<tr>
<td>Chr.04-D09</td>
<td>166.7-181.8</td>
<td>15.1</td>
<td>186.1</td>
<td>8.11</td>
<td>PAR0372</td>
<td>Gate3DC07</td>
<td>66859322-70184607</td>
<td>3.33</td>
<td>62.13</td>
<td>5.36</td>
<td>Inversion</td>
</tr>
<tr>
<td>Chr.10-D11</td>
<td>111.2-85.2</td>
<td>26</td>
<td>182.7</td>
<td>14.23</td>
<td>BNL1161</td>
<td>pVNC163</td>
<td>16252771-22482522</td>
<td>6.23</td>
<td>62.59</td>
<td>9.95</td>
<td>Inversion</td>
</tr>
</tbody>
</table>
**Figure 1** DNA marker, gene, and repetitive DNA distributions on the 13 D-genome pseudomolecules and the WGMM. The x-axis indicates the lengths of the chromosomes in Mb and the y-axis shows the number of markers per 1 Mb window in a range of 0-300. The shaded fields in blue display the marker distribution in 1 Mb windows at 100 kb steps from 0 Mb to the end of each chromosome. Gene (red) and repetitive sequence (blue) density heatmaps for the chromosomes are above each marker distribution map.

**Figure 2** Comparison of the Di, At, and Dt genetic maps and their physical locations on the pseudo molecules of the D genome sequence of *G. raimondii*. The 13 D-genome pseudo molecules are plotted on all y-axes. A. Di01-Di13 indicates the 13 chromosomes on the D genome genetic map; B. Chr.01-Chr.26 indicates the 26 tetraploid chromosomes of cotton. Both the relative genetic and physical distances of the chromosomes on the plots are represented by the cells of different sizes according to the ratio of the chromosome lengths. Chromosome arm inversions (red ovals) and translocations (red arrows) are highlighted.