Genetic map construction and functional characterization of genes within the segregation distortion regions (SDRs) in the F$_{2:3}$ populations derived from wild cotton species of the D genome

KIRUNGU Joy Nyangasi[1†], MAGWANGA Richard Odongo[1,2†], SHIRAKU Margaret Linyerera[1], LU Pu[1],
MEHARI Teame Gereziher[1], XU Yuanchao[1], HOU Yuqing[1], AGONG Stephen Gaya[2], ZHOU Yun[3], CAI Xiaoyan[1*],
ZHOU Zhongli[1*] (iD), WANG Kunbo[1*] and LIU Fang[1,4*]

Abstract

Background: Segregation distortion (SD) is a common phenomenon among stable or segregating populations, and the principle behind it still puzzles many researchers. The F$_{2:3}$ progenies developed from the wild cotton species of the D genomes were used to investigate the possible plant transcription factors within the segregation distortion regions (SDRs). A consensus map was developed between two maps from the four D genomes, map A derived from F$_{2:3}$ progenies of *Gossypium klotzschianum* and *G. davidsonii* while Map B from *G. thurberi* and *G. trilobum* F$_{2:3}$ generations. In each map, 188 individual plants were used.

(Continued on next page)

* Correspondence: cxycri@163.com; zhonglizhou@163.com; wkbcri@163.com;
liufcri@163.com
†Joy Nyangasi KIRUNGU and Richard Odongo MAGWANGA contributed
equally to this work.
[1]State Key Laboratory of Cotton Biology /Institute of Cotton Research,
Chinese Academy of Agricultural Science, Anyang 455000, Henan, China
Full list of author information is available at the end of the article

(Continued from previous page)

Results: The consensus linkage map had 1 492 markers across the 13 linkage groups with a map size of 1 467.445 cM and an average marker distance of 1.037 0 cM. Chromosome $D_5$02 had the highest percentage of SD with 58.6%, followed by Chromosome $D_5$07 with 47.9%. Six thousand and thirty-eight genes were mined within the SDRs on chromosome $D_5$02 and $D_5$07 of the consensus map. Within chromosome $D_5$02 and $D_5$07, 2 308 and 3 730 genes were mined, respectively, and were found to belong to 1 117 gourp out of which 622 groups were common across the two chromosomes. Moreover, genes within the top 9 groups related to plant resistance genes (R genes), whereas 188 genes encoding protein kinase domain (PF00069) comprised the largest group. Further analysis of the dominant gene group revealed that 287 miRNAs were found to target various genes, such as the gra-miR398, gra-miR5207, miR164a, miR164b, miR164c among others, which have been found to target top-ranked stress-responsive transcription factors such as *NAC* genes. Moreover, some of the stress- responsive *cis*-regulatory elements were also detected. Furthermore, RNA profiling of the genes from the dominant family showed that higher numbers of genes were highly upregulated under salt and osmotic stress conditions, and also they were highly expressed at different stages of fiber development.

Conclusion: The results indicated the critical role of the SDRs in the evolution of the key regulatory genes in plants.

Keywords: Genetic map, Segregation distortion region, *Cis*-regulatory elements, Genes, miRNA

Background

Segregation distortion (SD) is described as a deviation from the expected Mendelian ratio within a segregating population due to various segregating distorters (Anhalt et al. 2008). Some of the factors that may lead to SDs include gametic and zygotic selections, non-homologous chromosome recombination, gene transfer, environmental agents, mapping population, marker types and genetic transmission (Mello et al. 1991). During the construction of genetic maps, it has been observed that some alleles in chromosomal regions skew from the normal Mendelian ratio. These alleles tend to cluster at segments of the chromosome, and these regions are referred to as the segregation distortion region (SDR) (Lu et al. 2002).

Research has shown that SD could bring errors in the marker order and map distances in the linkage map and thus reduce the accuracy of the maps (Yuan et al. 2019). However genes of significance have been mined within the SDRs, for instance, the gene for crown rot resistance in wheat was identified within the SDR (Bovill et al. 2006), while the gene responsible for stem rust tolerance, was detected in the SDR on chromosome 2B in wheat (Tsilo et al. 2008). Moreover, SD has been observed in a variety of populations of organisms including insects (Sandler and Golic 1985), plants (Yuan et al. 2019), and mammals (Kumari et al. 1992).

Higher frequencies of occurrence of the SDR have been found in populations developed through interspecific as compared with intraspecific crosses (Dai et al. 2017), for example in rice more SDRs were detected in the double haploid compared to the $F_{2:3}$ populations developed from the same intraspecific cross (Xu et al. 1997; Wu et al. 2010), thirty-six SDRs were detected on 20 chromosomes in recombinant inbred lines in

tetraploid cotton (Jamshed et al. 2016). Further evidence points out that the genes associated with zygotic and gametic selection could be responsible for SD (Manrique-Carpintero et al. 2016).

The use of molecular markers is preferred in the genotyping of populations because they are less influenced by phenotype and are significant in the study of SD (Zhang et al. 2013). The most used molecular marker in the analysis of SD is the simple sequence repeat (SSR); it has been widely used in the study of SD in the majority of plants and animals (Cheng et al. 2016; Wang et al. 2019). Several studies on SDs have been conducted in several plant species, including rice (Reflinur et al. 2014; Yang et al. 2014), maize (Lu et al. 2002; Wang et al. 2012), wheat (Kumar et al. 2007), barley (Liu et al. 2011), soybean (Liu et al. 2000), rapeseed (Yang et al. 2006), cotton (Wu et al. 2003; Amudha et al. 2012), and other plants. In the analysis of SD in the $F_{2:3}$ population of *Aegilops tauschii*, it was observed that some regions had skewed ratios towards particular alleles in the chromosomes (Fans et al. 1998).

The studies conducted in cotton showed that the majority of the SDs were mainly skewed towards the male parent rather than the female population, as was observed on chromosome 18 (Dai et al. 2017). However, in all the studies conducted to unravel the mystery of SDs in cotton, no experiment has been undertaken to explore the SDs in the $F_{2:3}$ population derived from the diploid wild cotton parental lines. The latest attempt to explore the SDs in the wild cotton progenitors involved a backcross population developed between *G. hirsutum* as the recurrent parent and *G. mustelinum* as the donor cultivar (Chandnani et al. 2017). And therefore, to explore the phenomena of the SDs in wild cotton progenitors, an interspecific population between *G. klotzschianum*

and *G. davidsonii*, and between *G. thurberi and G. trilobum* were developed. The four parental lines were primarily selected because of their diverse genetic traits and broader ecological niches. The four parental lines used in the construction of the genetic maps are known to have traits for resistance to bacterial blight (*G. davidsonii*) (Zhang et al. 2016), sucking pests such as aphids (*G. klotzschianum*) (Wei et al. 2017), Fusarium wilt, silver leaf whitefly and cotton bollworm resistance (*G. thurberi*) (Natwick 2006), Verticillium wilt (*G. trilobum*) (Dong et al. 2019). A total of 188 individuals were genotyped using SSR markers, primarily focusing on the exploitation of the genetic mechanism of the SD in severely distorted chromosome D_502 and chromosome D_507. The analysis of the SD from the genetic maps constructed from the diploid cotton of the D genome was conducted. The first map was then generated from two closely related parents, *G. klotzschianum* and *G. davidsonii* (Kirungu et al. 2018) and the second map developed from *G. thurberi and G. trilobum* (Li et al. 2018), in either of the maps, the $F_{2:3}$ population used, the genotypic data from the two maps were combined to generate the consensus map, and the consensus map was generated by using the two maps. The only available maps developed from the wild cotton species of the D genome. The focus was on chromosome D_502 and chromosome D_507 which showed severe distortions of markers from the two maps. Moreover, the marker segregation and genes within the SDRs were mined and analyzed. The genes mined within the SDR and understanding their roles will be significant in elucidating the role played by segregation distortion, and will help in improving the elite cultivated cotton germplasms with ever-shrinking genetic base and significantly lower adaptive mechanisms to various abiotic and biotic stress factors.

Materials & methods
Parental materials
The two genetic maps were generated from an interspecific population obtained from the four parental lines. The first genetic map (Map A) was constructed from the $F_{2:3}$ population derived from the self-pollinating F_1 population of *G. klotzschianum* (female parent) and *G. davidsonii* (male parent). Similarly, the second genetic map (Map B) was constructed from $F_{2:3}$ populations derived from *G. thurberi* (female parent) and *G. trilobum* (male parent). A total of 188 progenies were used as the mapping population. The $F_{2:3}$ progenies from the four parental lines were developed and grown in the wild cotton nurseries, managed by the Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS), located in Sanya, Hainan province, China. The development of the $F_{2:3}$ progenies followed a similar pattern as described by Magwanga et al. (2020) in the

development of the backcross progenies between *G. tomentosum* (donor male parental line) and *G. hirsutum* (recurrent female parental line).

Molecular markers genotyping
Total DNA was extracted from the $F_{2:3}$ progenies and their parental lines using the CTAB method (Zhang et al. 2000b). Polymerase chain reaction (PCR) was conducted. The amplified PCR products were electrophoresed on non-denaturing 10% polyacrylamide gel electrophoresis in the $1 \times$ TBE buffer, and the gels were then visualized after silver staining (Huang et al. 2018). The primers used were the SWU markers which were developed by Southwest University in China, hence the acronym SWU. In the construction of the genetic map A, a total of 12 560 SWU markers were screened of which 1 000 markers were found to be polymorphic. Out of the 1 000 polymorphic markers, 728 markers were mapped and generated 13 linkage groups, designated as chromosome D_501 to D_513. In the second genetic map, map B, 12 560 SWU markers were screened, of which 996 markers were polymorphic, and only 849 polymorphic markers were mapped onto the 13 linkage groups. For the construction of consensus map, 1 492 polymorphic markers were applied to generate the genetic map, after removing the duplicated markers. The details of the markers and their sequences are shown in Supplementary Table S1.

Linkage map construction and determination of the segregation distortion of molecular markers
Markers with less than 5% missing data were used in the mapping of the linkage groups in the three maps (Coulton et al. 2020). The Joinmap 4.0 mapping tool was applied with a recombination frequency of 0.40, and a LOD (logarithm of odds) score of 3.0, for any LOD above 2.5 is known to be above the noise level (Faleiro et al. 2003). The Kosambi mapping function was used to convert the recombination frequencies to map distances. The linkage groups were then constructed using Mapchart 2.3 software (Voorrips 2002). The consensus map was constructed by merging the two individual data sets. Maps were drawn using MapChart 2.2 (Voorrips 2002).

Segregation distortion analysis
Segregation distortion (SD) within the mapping population was determined when the genotypic ratios deviated significantly from the expected Mendelian expectation (Reflinur et al. 2014). A Chi-square (χ2) test was performed for each marker to assess whether it significantly deviated from Mendelian segregation ratios. The markers showing segregation distortion were indicated by asterisks. The level of distortion was determined as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.00$; ****, $P <$

0.001; *****, $P < 0.0005$; ******, $P < 0.0001$; *******, $P < 0.00005$; in which *******, $P < 0.00005$ denoted the highly distorted markers. The Chi-square test was used to calculate the distortion of each marker.

Annotation of genes at the segregation distortion regions (SDRs) and the analysis of phylogenetic tree

Sequences corresponding to the SSR markers were identified by BLASTN (Nucleotide basic local alignment search tool) to the cotton ESTs (Expressed sequence tag) with an $E \leq 1$ e-15 and were annotated using BLASTX (Translated nucleotide sequence searched against protein sequences) (NCBI, Bethesda, MD, USA). The four genotypes, *G. klotzschianum*, *G. davidsonii*, *G. thurberi* and *G. trilobum,* have not been sequenced, the D_5, *G. raimondii* was used as the reference genome. A similar method has been used to explore the genetic variation among the BC_2F_2 genotypes developed from *G. hirsutum* as the recurrent parent and *G. tomentosum* as the donor parent (Magwanga et al. 2018b). The mined genes within this SDR that belonged to the two most abundant subfamilies, the probable protein types and the serine/threonine-protein kinase were then analyzed for their properties and function. A phylogenetic tree was constructed and, the multiple sequence alignments of all the proteins were done by Clustal Omega, MEGA 7.0 software (Kumar et al. 2016). The neighboring method (NJ) was used with a bootstrap value of 1 000 replications, and other parameters were applied as per the default setup, as previously used in the analysis of the phylogenetic relationships of the LEA proteins in cotton (Magwanga et al. 2018b). Transcriptional response elements of genes for the two major subfamilies were predicted using an online tool, the PLACE database (http://www.dna.affrc. go.jp/PLACE/signalscan.html) (Higo et al. 1999). The genes targeted by miRNAs were predicted by searching 5′ and 3′ untranslated regions (UTRs) and the coding sequences (CDS) of all the genes for their complementary sequences for the cotton miRNAs using the psRNATarget server (http://plantgrn.noble.org/psRNATarget/function).

Gene Ontology (GO) annotation

Analysis of GO annotation was conducted using Blast2GO PRO software version 4.1.1 (https://www.blast2go.com). The GO annotations described the hierarchal roles of the genes and their products, it entailed three independent ontological terms, the molecular function (MF), biological process (BP), and cellular component (CC) (Langfelder and Horvath 2008; Magwanga et al. 2018c). The protein sequences of the dominant gene domains were obtained within the SDRs and subsequently analyzed through Blast2GO as previously applied in the analysis of the *LEA* genes in cotton (Magwanga et al. 2018b).

RNA and RT-qPCR validation of key genes harbored within the SDRs

Based on the previous work by our research team, *G. raimondii* (D5), *G. thurberi*, and *G. trilobum* were profiled under biotic stress conditions, in which the plants were exposed to *Verticillium dahliae* infection (Dong et al. 2019). The genes which were harbored within the SDR were also prominently expressed, and majorities were members of the Probable Protein Types and the Serine/Threonine-Protein Kinase. Moreover, the *de novo* sequencing of the *G. klotzschianum* and *G. davidsonii* revealed a similar pattern (the data yet to be published). The highly upregulated genes were further validated under abiotic stress conditions, in which the seedlings of *G. klotzschianum*, *G. davidsonii*, *G. thurberi,* and *G. trilobum* at the third-leaf stage were exposed to drought and salt stress by exposing the seedings to 15% Polyethylene glycol 6000 (PEG6000) and 250 mmol•L^{-1} NaCl, respectively. The leaf tissues were then harvested for RNA extraction at 0 h, 1 h, 3 h, 6 h and 12 h of post-stress exposure. RNA extraction, purification, and RT-qPCR analysis were carried out as described by Lu et al. (2018). Cotton *GrActin* was applied as the reference gene.

Results

Linkage map construction

The first map was developed from the $F_{2:3}$ population between *G. klotzschianum* and *G. davidsonii*, a total of 728 polymorphic markers were used. The total map length was 1 480.23 cM, with an average marker interval of 2.182 cM (Kirungu et al. 2018). This map was designated as map A. The second map, designated as map B, was derived by genotyping the $F_{2:3}$ population developed between *G. thurberi* and *G. trilobum*, and 849 polymorphic markers were used in the linkage map construction. The map size was 1 012.46 cM with an average marker distance of 1.193 cM. In both maps, it was observed that chromosome number two also annotated as D_502 had the least map size of 82.908 cM and 28.665 cM in map A and map B, respectively. Interestingly, in both maps, chromosome D_502 had a smaller map size but with the highest percentage of SD (Table 1). Similar results have been observed in other linkage maps in cotton (Yu et al. 2011; Li et al. 2016).

The consensus map was constructed by merging two data sets from the two genetic maps. A total of 1 492 markers were mapped onto the 13 linkage groups encompassing the 13 chromosomes, and only 85 markers remained unlinked. The diploid cotton species has 13 chromosomes, while the tetraploid cotton species has 52 chromosomes (Mendoza et al. 2013; Magwanga et al. 2018a). This work was based on the diploid cotton species of the D genome. The consensus map size was 1 467.445 cM with an average marker distance of 1.037

Table 1 Mapping statistics for the two individual maps and the consensus genetic maps of diploid cotton in the D Genome. Map A represents genotyping of *G. klotzschianum* and *G. davidsonii*, Map B represents genotyping of *G. thurberi* and *G triolobum* while Consens Map represents the combination of genotypic data of map A and B.

Chr	Marker numbers per chromosome			Average distance /cM			Map size /cM			Number of SD			Average SD /%		
	Map A	Map B	Consensus Map	Map A	Map B	Consensus Map	Map A	Map B	Consensus Map	Map A	Map B	Consens Map	Map A	Map B	Consensus Map
$D_5$01	89	60	143	1.304	1.713	0.788	116.045	102.761	112.698	3	12	13	3.371	20	9.091
$D_5$02	44	21	58	1.884	1.365	1.943	82.908	28.665	112.698	35	10	34	76.087	42.857	58.621
$D_5$03	45	56	94	2.59	1.136	1.259	116.528	63.601	118.325	8	5	7	17.778	8.929	7.447
$D_5$04	56	70	123	1.997	0.846	0.767	111.846	59.229	94.288	2	6	7	3.571	8.571	5.691
$D_5$05	49	89	136	2.361	1.04	0.856	115.671	92.563	116.432	17	8	25	34.694	8.989	18.382
$D_5$06	58	73	125	2.001	0.88	1.082	116.045	64.213	135.273	5	8	9	8.621	10.959	7.200
$D_5$07	86	60	142	1.446	1.15	0.809	124.358	69.003	114.899	35	23	68	40.698	38.333	47.887
$D_5$08	49	64	99	2.492	1.251	0.81	122.13	80.053	80.156	25	5	27	51.02	7.813	27.273
$D_5$09	69	93	141	1.697	1.038	0.944	117.06	96.559	133.117	12	10	15	16.216	10.753	10.638
$D_5$10	34	58	84	2.998	1.786	1.238	101.93	103.563	103.973	2	12	11	5.882	20.69	13.095
$D_5$11	63	82	140	1.806	0.788	1.007	113.801	64.604	140.985	5	23	25	7.937	28.049	17.857
$D_5$12	49	41	100	2.301	2.153	1.007	112.739	88.288	100.72	6	2	6	12.245	4.878	6.000
$D_5$13	37	82	107	3.491	1.212	0.971	129.164	99.356	103.881	4	11	7	10.526	13.415	6.542
Totals	728	849	1 492	2.182	1.193	1.037	1 480.23	1 012.458	1 467.445	159	135	254	22.2	15.783	18.133

cM. Even though the map size was relatively smaller than map A, the marker interval was low, which improved the precision of the consensus map. From the consensus map, we observed that Chromosome $D_5 02$ had the highest percentage of SD with 58.6%, followed by Chromosome $D_5 07$ with 47.9%. Chromosome $D_5 01$ had the highest number of markers (143), while Chromosome $D_5 02$ had the least number of markers (58) (Table 1). Most of the markers mapped on the consensus map were found to be contributed by map B rather than map A. A total of 797 markers from map B were mapped on the consensus map accounting for 53.4% while only 695 markers (46.6%) were from map A. The chromosome with the highest number of markers (143)was Chromosome $D_5 01$ while the chromosome with the least number of markers (58) was Chromosome $D_5 02$ (Fig. 1).
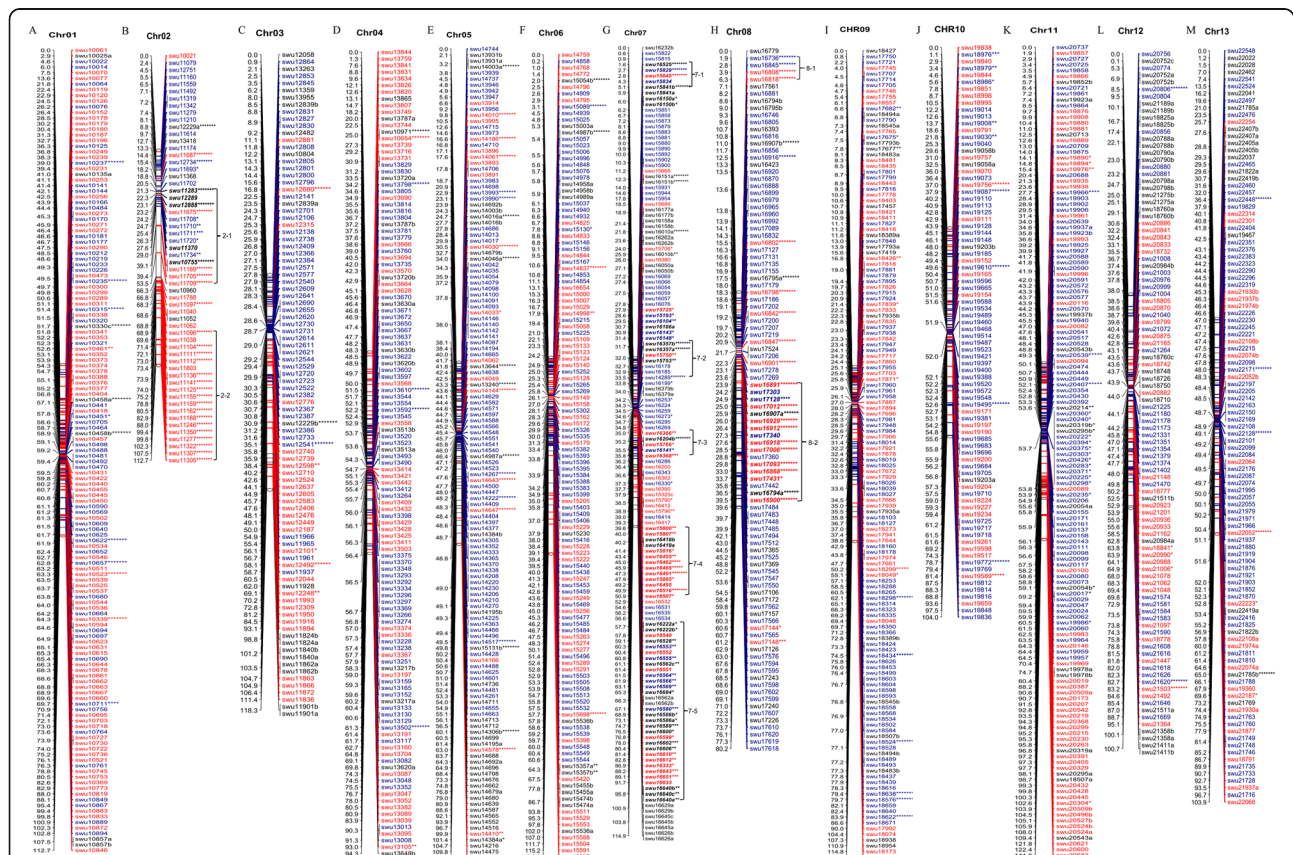
Segregation distortion (SD) analysis

In map A, out of the 728 markers mapped, 159 markers were distorted accounting for 22.2%, and the highest SD was observed in Chromosome $D_5 02$ with 76.1% followed by Chromosome $D_5 07$ with 40.7%. The SDRs were located on Chromosome $D_5 02$, $D_5 05$, $D_5 07$, and $D_5 08$. Chromosome $D_5 02$ had the largest SDR, while Chromosome $D_5 07$ had the highest number of SDR.

It was observed that the alleles in SDR were skewed towards a particular parental line, like that in Chromosome $D_5 02$ towards the female parent (G. klotzschianum), and in Chromosome $D_5 07$ towards the heterozygosity (Kirungu et al. 2018). In the second genetic map B, there was a slightly lower number of distorted markers, with only 135 accounting for 15.8%, and the highest two segregation distortions were observed in Chromosome $D_5 02$ and Chromosome $D_5 07$ with 42.8% and 38.3%, respectively (Table 1). Chromosomes that had the SDRs were $D_5 01$, $D_5 02$, $D_5 06$, $D_5 07$, $D_5 09$, $D_5 10$, and $D_5 11$. Moreover, the largest SDR was located on Chromosome $D_5 02$, while Chromosome $D_5 07$ had the highest number of SDR.

In the consensus map, the highest SDs were located on Chromosome $D_5 02$ and $D_5 07$, with distortion percentages of 58.6% and 47.9%, respectively. Similarly, the two chromosomes had the largest SDRs (Fig. 2). The

Fig. 1 Consensus genetic linkage map representing 13 linkage groups of the diploid cotton of D genome, developed from map A (G. klotzschianum and G. davidsonii) and map B (G. thurberi and G. trilobum). Markers in green font represent map B while markers in red font represent map A. The markers in black represent markers translocated from other chromosomes within the maps; the markers within the SDR are italicized and bold. **a-m**: represent the individual chromosomes, from chromosome 1 to chromosome 13

largest SDR was located on Chromosome D$_5$02–2 and was skewed toward the female parents while SDR located on Chromosome D$_5$02–1 was skewed towards the heterozygous. Chromosome D$_5$07 had the highest number of SDRs with a total of five SDRs, and all the SDR were skewed towards the heterozygotes except for the SDR located on Chromosome D$_5$07–1, which was skewed towards the female parents. The majority of the

SDRs were skewed towards the heterozygotes. Similar results were observed in the analysis of SDRs in tetraploid cotton, more specifically on Chromosome 18 (Dai et al. 2017), and rice (Wu et al. 2010), wheat (Fans et al. 1998). Based on the individual maps, the SDs were skewed towards the female compared with the male parent, the results obtained were in agreement with the study conducted on an interspecific F$_2$

Fig. 2 Segregation distortion region (SDR) in Chr D$_5$02 and D$_5$07 in Map A, Consensus map and Map B; Markers in green font represent map B while markers in red font represent map A. The markers in black represent markers translocated from other chromosomes within the maps; the markers within the SDR are italicized and made bold

population in which the segregated distorted markers were skewed towards the female parent (Li et al. 2007).

Annotation of genes at SDR

We conducted a BLAST search, and a total of 6 038 genes were mined within the SDR in Chromosome D_502 and D_507 (Supplementary Table S2), with 2 308 genes in Chromosome D_502 and 3 730 genes in Chromosome D_507. These genes were further divided into 1 117 groups were obtained. There are 622 groups which were shared between Chromosome D_502 and Chromosome D_507. The largest group consists of a total of 188 genes, among which the Pkinase (PF00069) was encoded; followed by a group with 132 genes, among which PF13855 (LRR_8, Leucine-rich repeat) is encoded. The third consists of 108 genes, among which PF07714 (Pkinase_Tyr, Protein tyrosine kinase) is encoded. The genes in the three main groups were highly correlated with abiotic stress response. The genes located within the largest 12 grous were analyzed. Out of the 12 groups, 9 were found containing 'members of the resistant genes (R group of genes), including that encoded Protein kinase, LRR_8, Protein tyrosine kinase , NB-ARC, LRRNT_2, Leucine-rich repeat N-terminal, PPR, PPR_2 family, Cytochromes P450 (CYPs), Myb-like DNA-binding, and RNA recognition motif (RRM, RBD, or RNP) (Tables 2 and 3).

Analysis of the physiochemical properties and structures of the genes obtained from the abundant/enriched domain mined within the SDR in chromosome D_502 and chromosome D_507

The abundant/enriched domain was the Protein kinase domain (PF00069). It has been widely studied; for instance, it was found to be the dominant domain in the

analysis of the genes conserved between the two upland cotton, *G. hirsutum* and its wild relative *G. tomentosum* (Magwanga et al. 2018b). We, therefore, explored the genes which belonged to this domain. The physiochemical properties of these genes showed significant variations; the molecular weight ranged between 10.351 kDa and 134.232 kDa, the charge was between – 27 and 39.5; Isoelectric point (pI) values were between 4.375 and 10.382; the GRAVY values ranged from – 0.721 to –0.251 while their protein lengths ranged from 611 aa to 12 310 aa (Supplementary Table S3). The GRAVY values were all below zero, indicating that these genes were mainly hydrophilic. The group that encoded Protein kinase domain contained 28 different subfamilies. The subfamily with the highest number of genes was Probable types with a total of 64 genes, which included members such as the Probable inactive receptor kinase (4 genes); Probable leucine-rich repeat receptor-like serine (21), Probable L-type lectin-domain containing receptor kinase (3 genes); Probable receptor-like protein kinase (25 genes) among others (Supplementary Table S4).

The two most abundant subfamilies, the probable protein types and the Serine/threonine-protein kinase were further analyzed, by looking into their classification based on the phylogenetic tree analysis. The genes were found to be grouped into five clades, with clade 2 being the majority (Fig. 3). The most interesting concept is that the members within clade 3 had a percentage bootstrap similarity value of 100%. The majority of these genes have previously been found to be highly correlated to biotic stress tolerance; for instance, 11 genes, i.e., *Gorai.007G335000, Gorai.002G039900, Gorai.002G040100, Gorai.002G041100, Gorai.002G041200, Gorai.002G041800, Gorai.002G042100, Gorai.002G047500, Gorai.002G047900, Gorai.007G182500* and *Gorai.007G334900,* are homologous to an *Arabidopsis*

Table 2 Characteristics of the genes found within the two common markers, SWU16562 and SWU16586, across maps A, B and the consensus map

Gene ID	Gene name	Description	Molecular weight /kDa	Charge	pI	GRAVY[*] value	Domain list	Domain
Gorai.007G355900	NA	NA	26.649	-12	4.563	−0.408	–	
Gorai.007G356000	At4g27220	Probable disease resistance protein At4g27220	252.737	−10	6.175	−0.127	PF00931	NB-ARC domain
Gorai.007G347200	LHP1	Chromo domain-containing protein LHP1	48.046	−13.5	4.855	−1.049	PF00385	Chromatin organization modifier
Gorai.007G347300	SIGB	RNA polymerase sigma factor sigB	64.627	15.5	9.115	−0.54	PF00140	Sigma-70 factor, region 1.2
Gorai.007G347400	NA	NA	16.507	17.5	9.897	−0.956	–	
Gorai.007G347500	NA	NA	47.568	14.5	9	−0.157	PF06219	Protein of unknown function (DUF1005)
Gorai.007G347600	TFCA	Tubulin-folding cofactor A	12.859	−5	4.781	−0.821	PF02970	Tubulin binding cofactor A
Gorai.007G347700	CYP89A2	Cytochrome P450 89A2	58.73	10	8.563	−0.074	PF00067	Cytochromes P450 (CYPs)
Gorai.007G347800	CYP89A2	Cytochrome P450 89A2	58.775	15.5	9.358	−0.151	PF00067	Cytochromes P450 (CYPs)

Notes: *Grand average of hydropathy.

Table 3 Distribution of genes of the 12 largest domains within Chromosome $D_5$02 and Chr $D_5$07 with the highest segregation distortion regions (SDRs) in the consensus map combing the genotypic data of the *G. klotzschianum* and *G. davidsonii* (Map A) and *G. thurberi* and *G. trilobum* (Map B)

PF number	Domain	$D_5$02 Gene number	$D_5$07 Gene number	Total genes per domain
PF00069	Protein kinase domain	71	117	188
PF13855	LRR_8; Leucine rich repeat	52	78	130
PF07714	Protein tyrosine kinase domain	55	53	108
PF00931	NB-ARC domain	15	82	97
PF08263	LRRNT_2; Leucine rich repeat N-terminal domain	31	58	89
PF00560	LRR_1; Leucine Rich Repeat	24	61	85
PF01535	Pentatricopeptide repeat (PPR)	26	51	77
PF13041	Pentatricopeptide repeat (PPR_2) repeat family	26	49	75
PF00067	Cytochromes P450 (CYPs)	29	32	61
PF00249	Myb-like DNA-binding domain	15	41	56
PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	25	29	54
PF13639	zf-RING_2; Ring finger domain	15	36	51
Total		**384**	**687**	**1 071**

gene, *At5g39020*, which has a functional role in leaf senescence during viral infection in *Arabidopsis* (Espinoza et al. 2007). Moreover, the remaining genes were homologous to an *Arabidopsis* gene, *At1g67000*, which plays a more significant role in salt stress pathways. It was also was found to be highly upregulated in the roots under salt stress conditions (Ma et al. 2006).
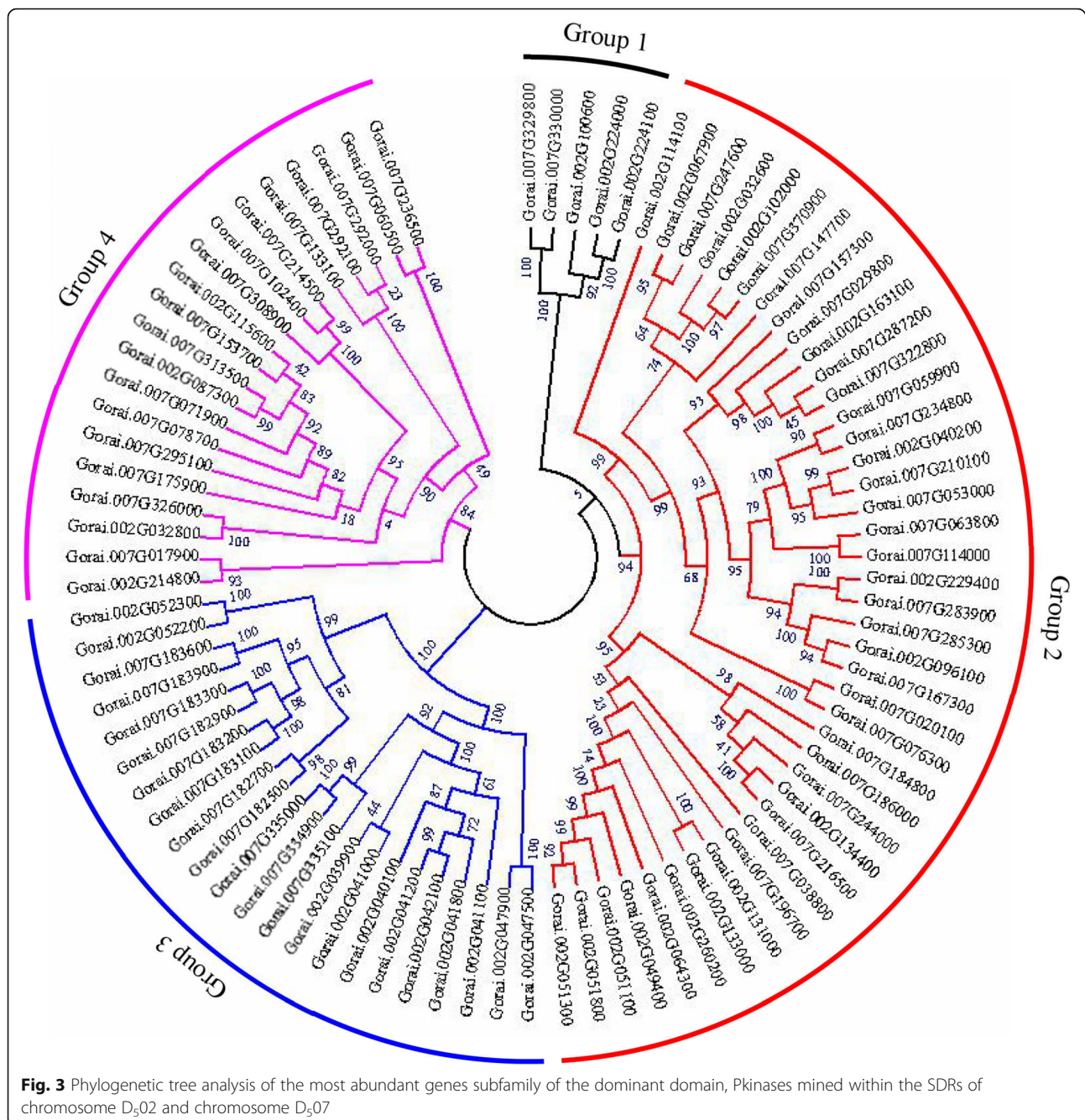
Cis-regulatory elements analyses of the major two subfamilies: the probable protein types and the serine/threonine-protein kinase

We examined the two major subfamilies to determine if there could be any of the regulatory elements related to either abiotic or biotic stress factors. *Cis*-regulatory elements are known to enhance the functions of the genes (Tümpel et al. 2006). In the analysis of the *cis*-elements, all the genes were found to be associated with either abiotic or biotic stress-responsive *cis*-regulatory elements; for instance, ARFAT with the sequence "TGTCTC" was found to be associated with 87 genes which function as ABA and auxin responsiveness. ABA is a plant phytohormone that is vital for plants' response towards stress (Trivedi et al. 2016). Other *cis*-regulatory elements predicted were CBFHV with a role in dehydration-responsive element / cold acclimation, DRECRTCOREAT functioning as activators that function in drought-, high-salt- and cold-responsive gene, lastly ABRELATERD1 with a function in early responsive to dehydration, AGMOTIFNTMYB2 induced by various stress such as wounding or elicitor treatment among others (Fig. 4 and Supplementary Table S5). The *cis*-regulatory elements detected such as ABRE have previously been found to associate with top-ranked plant stress-responsive transcription factors such as the NAC, MYB (Nakashima et al. 2009).

miRNA prediction for the major two subfamilies; the probable protein types and the serine/threonine-protein kinase

In the prediction analysis of the miRNA targeting the various genes obtained for the two major subfamilies, a total of 287 miRNAs were found to target 91 genes (Supplementary Table S5). The high number of miRNA targets detected for these genes showed that the genes obtained from the SDR on chromosome $D_5$02 and chromosome $D_5$07 have a significant role in various biological processes. The highest level of miRNA target was observed for the following genes: *Gorai.002G039900* (6 miRNAs), *Gorai.002G041100* (9 miRNAs), *Gorai.002G114100* (9 miRNAs), *Gorai.002G133000* (7 miRNAs), *Gorai.002G134400* (8 miRNAs), *Gorai.007G244000* (9 miRNAs), *Gorai.007G271300* (10 miRNAs) among the rest. The miRNAs and the genes association revealed that higher level of miRNA targets , for instance, a single gene was targeted by 2~10 miRNAs. Some of the miRNAs detected were gra-miR172a and gra-miR172b all found to target *Gorai.007G059900* which is a member of the serine/threonine-protein kinase. The SAPK2 mined within the SDR located on chromosome $D_5$07 has been found to have a function in fiber development in cotton (Abdurakhmonov et al. 2008). Moreover, miR398 has been extensively studied and found to have a role in enhancing abiotic stress tolerance in plants; for instance, gra-miR398 was found to be upregulated in plants exposed to water deficit conditions, and thus found to be responsible for enhancing tolerance towards oxidative stress, water deficit, salt stress, abscisic acid stress, ultraviolet stress, copper

KIRUNGU *et al. Journal of Cotton Research* (2020) 3:32

Page 10 of 19

Fig. 3 Phylogenetic tree analysis of the most abundant genes subfamily of the dominant domain, Pkinases mined within the SDRs of chromosome D₅02 and chromosome D₅07
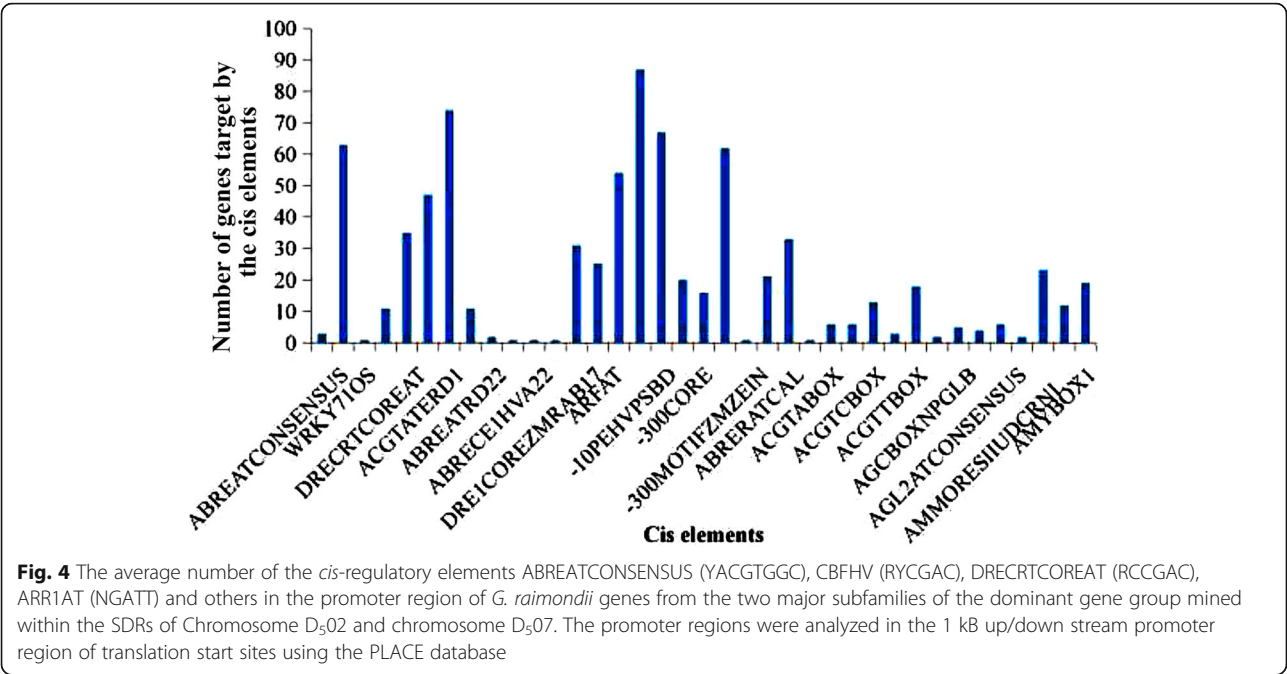
and phosphate deficiency, high sucrose and bacterial infection (Jia et al. 2009; Lu et al. 2010; Pashkovskii et al. 2010). The same miRNA was found to target *Gorai.007G335000*, a member of the probable receptor-like protein kinase mined within the SDR on chromosome D₅07.

GO annotation of the major two subfamilies; the probable protein types and the serine/threonine-protein kinase of the dominant gene groups

In the analysis of the GO terms, a total of 188 genes were found to have GO terms, in which a high number

of genes were found to be involved in biological process (BP), with functions such as regulation of the biological process, response to stimulus, single-organism process, metabolic process, and cellular process, in relation to cellular component (CC), four major functions were detected, namely the cell, cell part, membrane part, and membrane while in molecular functions (MF), and only two functions were observed, binding and catalytic activity (Fig. 5). Some unique results were found in some of the genes within the SDRs; for instance, *Gorai.002G14960* (BRASSINOSTEROID INSENSITIVE 1-
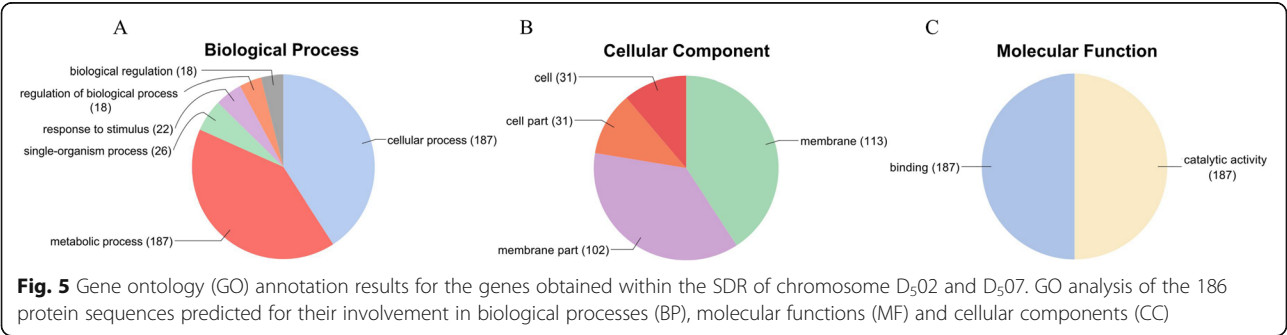
Fig. 4 The average number of the *cis*-regulatory elements ABREATCONSENSUS (YACGTGGC), CBFHV (RYCGAC), DRECRTCOREAT (RCCGAC), ARR1AT (NGATT) and others in the promoter region of *G. raimondii* genes from the two major subfamilies of the dominant gene group mined within the SDRs of Chromosome D$_5$02 and chromosome D$_5$07. The promoter regions were analyzed in the 1 kB up/down stream promoter region of translation start sites using the PLACE database

like) was found to have 20 GO functions, with 3 cellular component functions, namely endosome (C: GO: 0005768), plasma membrane (C: GO:0005886) and integral component of membrane (C: GO:0016021). Five molecular functions were: protein serine/threonine kinase activity (F: GO:0004674), steroid binding (F: GO: 0005496), ATP binding (F: GO:0005524), protein homo-dimerization activity (F: GO:0042803), and protein heterodimerization activity (F: GO:0046982). A very high number of biological processes were observed microtubule bundle formation (P: GO:0001578), protein phosphorylation (P: GO:0006468), skotomorphogenesis (P: GO:0009647), detection of brassinosteroid stimulus (P: GO:0009729), brassinosteroid mediated signaling pathway (P: GO:0009742), positive regulation of flower development (P: GO:0009911), response to UV-B (P: GO: 0010224), pollen exine formation (P: GO:0010584), leaf development (P: GO:0048366), anthers wall tapetum cell differentiation (P: GO:0048657), negative regulation of

cell death (P: GO:0060548) and regulation of seedling development (P: GO:1900140). Other genes harbored a range of GO functions from three to 10 different functions (Fig. 6 and Supplementary Table S6).

RNA sequence data analysis profiled under abiotic stress conditions and in different Fiber developmental stages

By the fact that the two major subfamilies were found to be targeted by stress-specific miRNAs and even found to be associated with some known *cis*-regulatory elements, we undertook to investigate if the genes would have any varying expression under drought, salt and even different stages of fiber development. Genes and their RNA sequences were then obtained from the *de novo* sequenced data. The raw data for the RNA sequencing were transformed into $\log_2$ form and used in the construction of the heat map. The RNA expression analysis showed that the genes were categorized into three groups, with members in group 1 exhibiting higher expression levels at
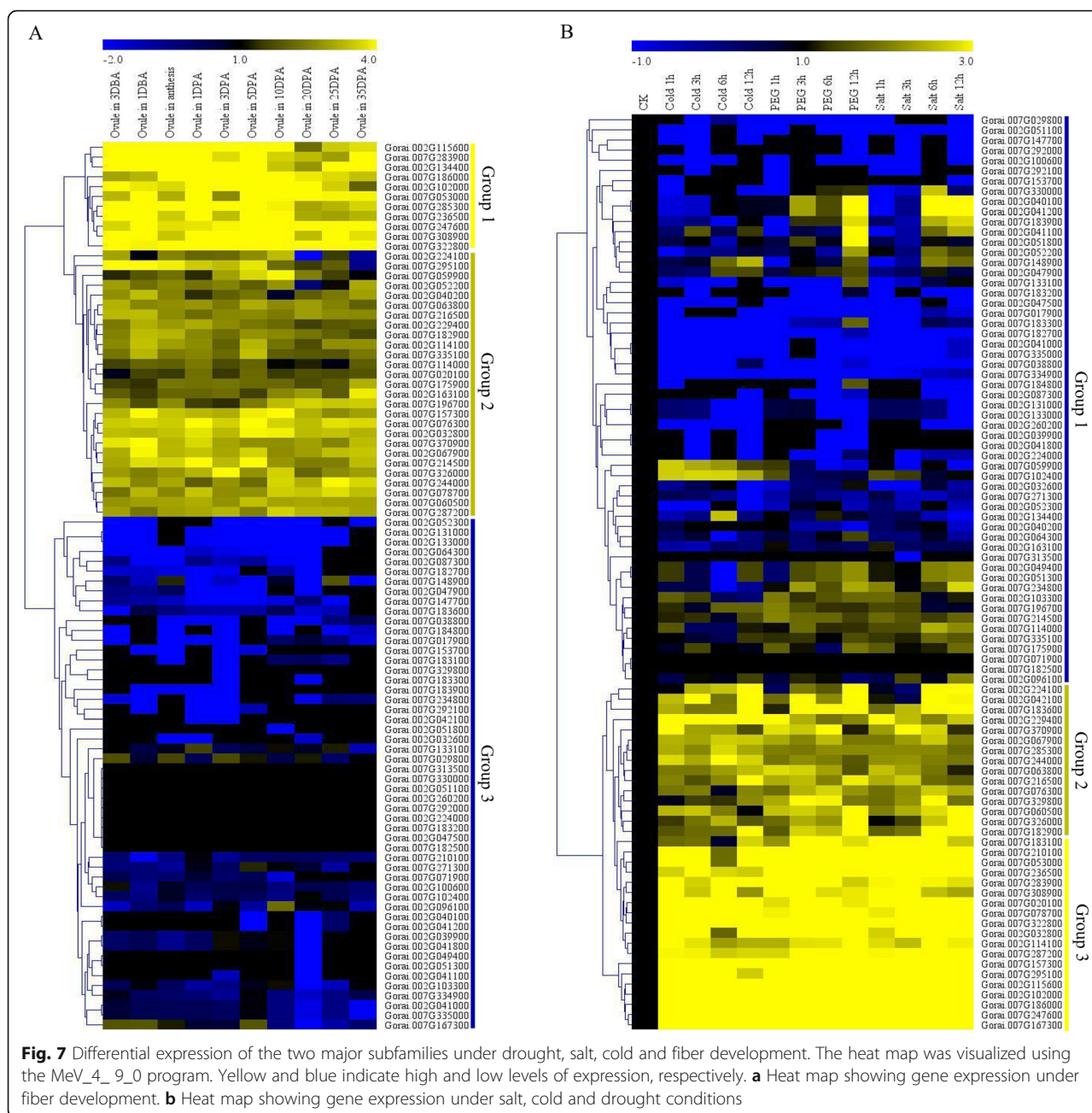

Fig. 5 Gene ontology (GO) annotation results for the genes obtained within the SDR of chromosome D$_5$02 and D$_5$07. GO analysis of the 186 protein sequences predicted for their involvement in biological processes (BP), molecular functions (MF) and cellular components (CC)

Fig. 6 Detailed gene ontology (GO) annotation, analysis of the significantly expressed genes within the SDRs of chromosome D$_5$02 and D$_5$07. GO analysis of the 186 protein sequences predicted for their involvement in biological processes (BP), molecular functions (MF) and cellular components (CC)

different fiber development stages (Fig. 7). The majority of the highly upregulated genes were obtained from the SDRs in chromosome D$_5$07, such as *Gorai.007G283900* (Serine/threonine-protein kinase Nek2), *Gorai.007G186000* (Probable inactive receptor kinase At1g48480), *Gorai.007G053000* (Serine/threonine-protein kinase SRK2I), *Gorai.007G285300* (Serine/threonine-protein kinase WNK1), *Gorai.007G235600* (Genome polyprotein), *Gorai.007G247600* (Serine/threonine-protein kinase ppk15) and *Gorai.007G308900*. It is interesting to note that the gene which was highly

upregulated in various stages of fiber development was also found to be targeted by gra-miR164a, and the same miRNA has been found to target the NAC transcription factor family (Xie et al. 2000). Moreover, mutant *Arabidopsis* lacking ath-miR164c was found to exhibit a slight defect in carpel fusion (Baker et al. 2005). In addition, miR164a,b,c has been found to have a regulatory role in the expression of CUP-SHAPED COTYLEDON1 (CUC1) and CUC2, which encode key transcriptional regulators involved in organ boundary specification (Huang et al. 2012). These previous findings showed that

Fig. 7 Differential expression of the two major subfamilies under drought, salt, cold and fiber development. The heat map was visualized using the MeV_4_ 9_0 program. Yellow and blue indicate high and low levels of expression, respectively. **a** Heat map showing gene expression under fiber development. **b** Heat map showing gene expression under salt, cold and drought conditions

the gene found to be targeted by miR164a/b/c could play an essential role in fiber development.

Under abiotic stress conditions, genes exhibited differential expression, with group 3 members exhibiting significantly higher expression under salt, cold and drought stresses. Some of the genes were highly expressed include *Gorai.007G167300* (Probable serine/threonine-protein kinase WNK11), *Gorai.007G247600* (Serine/threonine-protein kinase ppk15), *Gorai.007G186000* (Probable inactive receptor kinase At1g48480), *Gorai.002G102000* (Serine/threonine-protein kinase D6PKL2), *Gorai.002G115600* (Serine/threonine-protein

kinase CDL1), *Gorai.007G295100* (Serine/threonine-protein kinase CDL1), *Gorai.007G157300* (Serine/threonine-protein kinase MHK), *Gorai.007G287200* (Probable serine/threonine-protein kinase *At1g54610*), *Gorai.007G322800* (Probable serine/threonine-protein kinase *At1g09600*), *Gorai.007G078700* (Probable receptor-like protein kinase *At5g15080*) and *Gorai.007G020100* (Serine/threonine-protein kinase fray2). Among the highly expressed genes, *Gorai.007G167300* was targeted by gra-miR398. *Gorai.007G247600* was found to be targeted by gra-miR5207; miR398 is the first plant miRNA reported miRNA to be down-regulated by oxidative stresses. It has

been intensively studied and found to be important in the regulatory process of copper homeostasis, in response to abiotic stresses such as heavy metals toxicity, sucrose, and heat, in addition to biotic stresses through the down-regulation of the expression of Cu/Zn-superoxide dismutase (CSD) (Sunkar 2006; Lu et al. 2010; Pashkovskii et al. 2010). The result shows that the SDRs could be vital in the evolution of some of the key regulatory genes required for the survival of the plants.
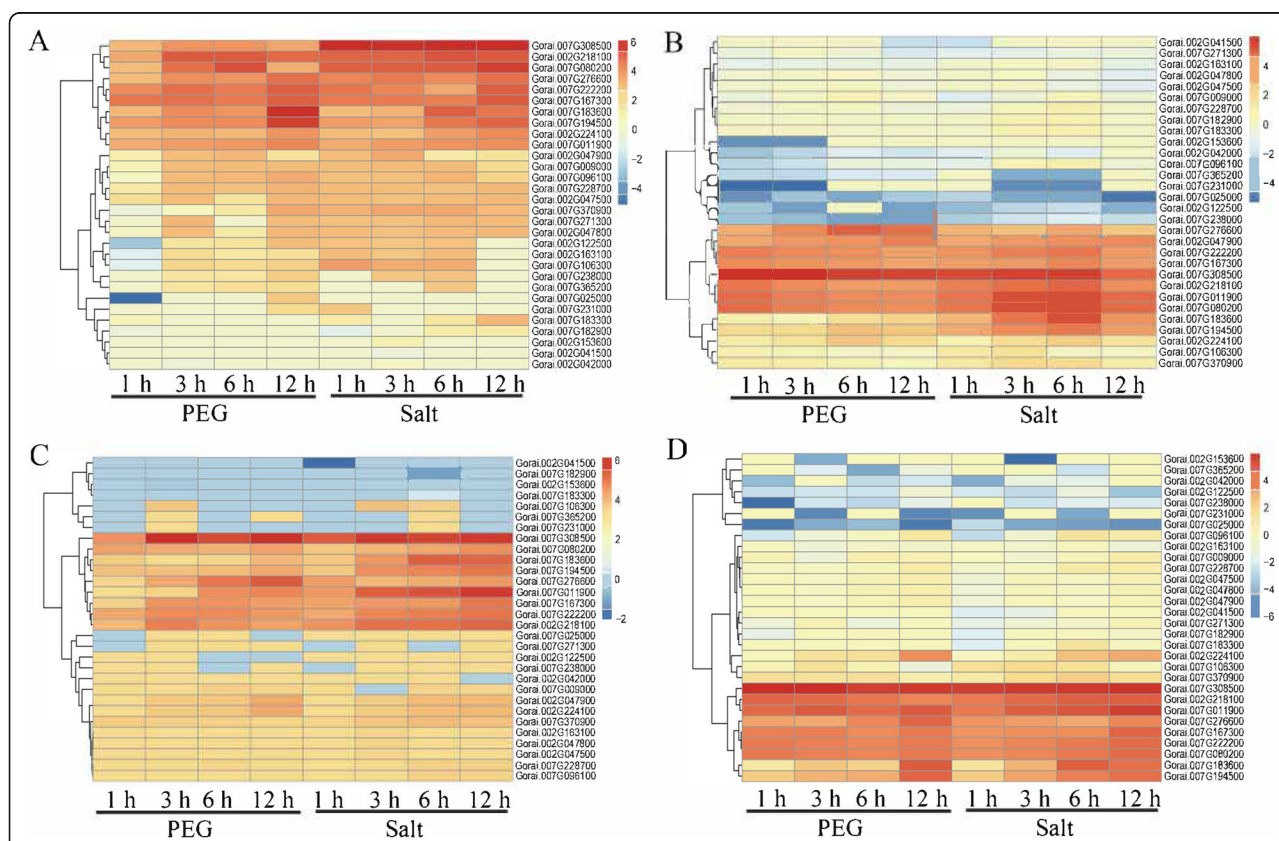
RT-qPCR validation of the selected genes within the SDRs of chromosome D$_5$02 and D$_5$07 under drought and salt stress conditions

Thirty genes were profiled on the leaf tissues of the four parental lines under drought and salt stress conditions. The genes exhibited three types of expressions across the four parental lines; however, more genes were found to be highly upregulated in the leaves of *G. klotzschianum* and *G. thurberi* compared with *G. davidsonii* and *G. trilobum* (Fig. 8a-d). The results obtained were in agreement to previous findings which have shown that

G. thurberi is more tolerant to both biotic stress conditions, more so to *Verticillium dahliae* which is a fungal pathogen causing Verticillium wilt, a terminal disease to various crops (Dong et al. 2019). Moreover, Cai et al. (2019) revealed that *G. thurberi* was highly tolerant to cold stress compared with *G.trilobum*. Furthermore, Kirungu et al. (2018) found that *G. klotzschianum* harbored more beneficial traits compared with *G. davidsonii*.

Discussion

Genetic maps have become significantly important in understanding markers, breeding, association genetics, map-assisted gene cloning, gene mining, and mapping of quantitative trait loci (QTLs) (Golestan Hashemi et al. 2015). In our study, we integrated two genetic maps from the D genome of the diploid cotton with a mapping size of 188 F$_{2:3}$ population. The first genetic map (Map A) was composed of a genetic cross between *G. klotzschianum* (female parent) and *G. davidsonii* (male parent) while the second genetic map (Map B) was developed from *G. thurberi* (female parent) and *G.*

Fig. 8 RT-qPCR validation of the selected genes within the SDRs of chromosome D$_5$02 and D$_5$07 under drought and salt stress conditions. The heat map was visualized using the MeV_4_ 9_0 program. Red and blue indicate high and low levels of expression, respectively, while white indicates none expressed genes. **a** Heat map showing gene expression in the leaf tissue of *G. klotzschianum*. **b** Heat map showing gene expression in the leaf tissue of *G. davidsonii*, **c** Heat map showing gene expression in the leaf tissue of *G. thurberi* and **d** Heat map showing gene expression in the leaf tissue of *G. trilobum*. Drought and salt stress was imposed by supplementing the Hoagland nutrient solution with 17% PEG and 250 mL of NaCl solution, respectively

trilobum (male parent). Map B had a higher number of markers linked and a smaller average distance as compared with map A. This map could play a fundamental role in the analysis of QTLs. In the construction of the consensus map, more markers were contributed by map B as compared with map A. Inconsistencies of marker order including the translocation or inversions between individual markers in consensus maps were observed especially on markers that were closely linked together in the SDR of Chromosome $D_5$02–2. Similar results were observed in the consensus map of flax seed (Cloutier et al. 2012).

The segregation distortion among the three maps ranged from 15.8% (map A) to 22.2% (map B). Segregation distorted markers have previously been studied in various plants (Takumi et al. 2013). The study of segregation distortion is significant because distorted markers may be linked to genes or traits of interest, these genes may be beneficial or lethal to organisms. Therefore, it's important to include the segregation distortion markers in the construction of genetic maps since the exclusion of such markers could cause bias of the data and result in the loss of significant genetic information. In our study, we examined the trend of segregation distortion within Chromosome $D_5$02 and $D_5$07. We observed that both chromosomes had higher segregation distorted markers. Chromosome $D_5$02 had the least mapped markers with a higher percentage of segregation distortion ranging from 42.9% to 76.1% in the three genetic maps. Similar results have been observed in cotton (Li et al. 2007; Khan et al. 2016; Shang et al. 2016). The two chromosomes also showed SDR which was skewed towards a specific allele. These SDRs may be due to pre- or post-zygotic selection and chromosome loss or rearrangements.

We observed that 29 genes were not disrupted by introns (intronless), intronless genes contain a single exon and do not contain introns from its beginning to the end neither in its UTR or CDS regions (Yan et al. 2016). The intronless genes are known to promote the efficiency of transcription initiation and elongation in spliced genes (Sakharkar et al. 2006). Their isoelectric point (pI) values ranged from both acidic to basic proteins. The pI values are known to affect the solubility of protein molecules; hence proteins are less soluble when the pH of the solution is at its isoelectric point (Dawes et al. 1994). All of the proteins were observed to have a GRAVY value less than zero, indicating that they were hydrophilic. Hydrophilic proteins have a high solubility, hence these proteins could be playing a significant role in desiccation tolerance (Hundertmark and Hincha 2008), and also aid in enzymatic activities involved in the biochemical processes.

The analysis of the genes mined within the SDR of chromosome $D_5$02 and $D_5$07 revealed that the dominant froup was the Pkinase gene family, with a Pfam number of PF00069. There were so many genes within this group, it was technically impossible to analyze all of them, and thus, we determined the dominant subfamily, and further analyzed two of them. The two major dominant subfamilies were the probable kinases and the serine/threonine kinases genes. These groups have been widely studied in both plants and animals (Jun et al. 2015). In the cotton plant, overexpression of *GbRLK*, a putative receptor-like kinase gene, has been found to confer tolerance to Verticillium wilt, a plant disease that is known to cause massive losses in cotton production (Jun et al. 2015).

Similarly, overexpression of the *GbRLK* gene isolated from *G. barbadense* has been found to confer drought and salt stress tolerance in transgenic *Arabidopsis* plants (Zhao et al. 2013). The detection of these genes within the SDRs demonstrates the significant role played by the SDRs in the evolution or synthesis of vital proteins with a profound role in enhancing tolerance levels of plants to various abiotic and biotic stress factors. The main genes found to be located within the SDR in the consensus map were the R genes. This group of genes is known to play an integral role in signaling during pathogen recognition, hence assist in the activation of plant defense mechanisms.

The R genes work in coordination with other groups to bring combinatorial variations in signal response specificity to pathogens. Moreover, the R genes are mainly associated with those encode proteins that identify specific pathogen effectors, known as avirulence proteins, which specific in terms of their activities. These genes are known to have a gene-to-gene interaction between an organism and its pathogens (Rouxel and Balesdent 2010). These genes were segregating within the SDR in synchrony intending to help in plant defense mechanisms, these mechanisms are involved in a series of enzymatic activities within the proteins. From the recent analysis, it has been observed that the proteins encoded by resistance genes (R genes) display modular domain structures and require several dynamic interactions between specific domains to perform their function (Wang et al. 2016), hence a very close interaction and coordination in terms of the activities of the genes located within the SDR. In a study conducted on determining significant QTLs for drought stress tolerance, the majority of the marker loci co-localized with known QTLs for blast tolerance or NBS-LRR disease resistance genes were located within the regions of significantly distortion levels (Dixit et al. 2014). Similar result was found on Bangladeshi rice landrace Capsule in relation to salt stress tolerance (Rahman et al. 2019). The four parental

lines used in the construction of the genetic map are known to contain traits for resistance to bacterial blight (*G. davidsonii*), sucking pests such as aphids (*G. klotzschianum*), Fusarium wilt, silver leaf whitefly and cotton bollworm resistance (*G. thurberi*), Verticillium wilt (*G. trilobum*). This explains the reasons for a large number of plant resistant genes (R genes) detected within the SDRs in chromosome D_502 and D_507.

The carrying out *insilico* analysis of the genes obtained within the SDRs, the *cis*-regulatory elements, miRNA and GO analysis showed that the R genes could play a significant role within the plant. Recent evidence indicates that plant miRNAs play a role in biotic and abiotic stress responses (Sunkar et al. 2007). In the analysis of the genes obtained within the SDRs, several miRNAs were found to target several genes; for instance, miR157a and miR157b were found to target a single gene *Gorai.007G063800*, a member of the serine/threonine-protein kinase. The same miRNA family was found to be the most abundant, followed by miR156, miR166, and miR168, with variation within each family in Pomegranate. This fruit has enormous importance in human health mainly because of its antioxidant properties, it does accumulate a high amount of anthocyanins in skin and arils (Saminathan et al. 2016). The antioxidant enzymes are important to plants in reducing the deleterious effects of reactive oxygen species (ROS). When plants are exposed to stresses, the production and elimination of the ROS process altered leading to excessive accumulation of ROS within the cell resulting in oxidative stress. The association of miR157 to induction of antioxidant enzymes, showed that these genes within the SDR are critical for plants.

The various *cis*-regulatory elements (CREs) targeting the genes within the SDRs, were found to perform a myriad of CREs with diverse functions. More specifically, it is geared towards enhancing plants tolerance to various environmental stresses; for instance, ABREATCONSENSUS targets not only the stress-responsive genes but also those involved in transportation such as the nitrate transporter (NRT) genes as evident in poplar plant (Aichi et al. 2006; Bai et al. 2013). The results obtained for the CREs were further augmented by GO annotation. The various genes obtained within the SDRs were found to play an integral in all the three GO functional annotations. In cellular component (CC), functions such as an integral component of membrane (GO: 0016021), cortical microtubule (GO: 0055028) among others were detected. The integrity of the cell membrane is important because the membrane is the communicating channel between intra and extracellular environments, and any damage to the cell membrane affects various biological processes such as osmosis, thus affecting cell water retention. The detection of these cellular component roles

showed that the genes found in the SDRs have a function in maintaining cell membrane stability, and therefore enhancing the delicate osmotic balance within the cell. Moreover, an integral component of the membrane was a function found to be unanimous with the *LEA* genes (Magwanga et al. 2018b).

Several gametophytic and zygotic barriers causing deviation of allele frequencies from Mendelian ratios have been reported in several plants such as rice (Wang et al. 2009). Therefore detection of SDRs in the two populations developed from the two wild parental lines is a common feature more so among the $F_{2:3}$ populations. It is assumed based on Mendelian law that there is an equal probability of transmission of alleles from either parent during sexual reproduction, but this has not been the case in several studies, being there tend to be phenomena referred to as the preferential transmission of alleles or genotypes known as segregation distortion (SD) (Nadeau 2017). The evolution of segregation distortion may have profound evolutionary implications. From previous studies the bulk pollen sequencing indicated a rapid evolution of segregation distortion (Corbett-Detig et al. 2019). SD has been described as powerful evolutionary tools that could lead to speciation (Liberman and Feldman 1982). SDR has been observed not only among the controlled population but also among the natural population (McLaughlin and Malik 2017). The results from the two maps and their consensus showed that SDs are a common feature in segregating population and could be used to mine genes of significance that could be introgressed into the already cultivated species.

Conclusions
The use of genetic map analysis has become increasingly significant in understanding markers-assisted selection, gene mining and cloning. However, intensive investigation of genes located within the SDR has not been widely studied. In our research we examined the only two interspecific maps developed in the D genome of the diploid cotton. We constructed a consensus map from the two genetic maps and noted that in all the three maps D_502 and D_507 had the highest of SD, and hence we mined the genes within the SDR of D_502 and D_507 to find out if there were genes of significance that could be segregating within this region. A total of 2 308 genes in D_502 and 3 730 genes in D_507 were mined within the SDR, these genes were divided into 1 117 groups of which 622 groups were shared between the two chromosomes. We further observed that the 12 largest domains had a significant role in the plant defense mechanism of which 9 belonged to the resistance genes (R group of genes), with 188 genes and a pfam number of PF00069. We analyzed for the properties of these genes, the largest subgroup encode the serine/threonine-

protein kinase. The genes that performed similar roles clustered together within the SDR. These genes have similar feature being hydrophilic. The study of these genes will provide an understanding of the significance of genes within the SDR and the role of the consensus map in mining these genes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-020-00072-2.

> **Additional file 1: Table S1**. Details of primers used in this research
>
> **Additional file 2: Table S2**. Details of primers used for the RT-qPCR validation of the 30 selected genes within the SDRs on chromosome $D_5$02 and $D_5$07
>
> **Additional file 3: Table S3**. Genes within the dominant group
>
> **Additional file 4: Table S4**. Genes mined within the SDR of chromosome $D_5$02 and $D_5$07
>
> **Additional file 5: Table S5.** Cis-regulatory promoter elements identified for the genes obtained within the SDRs
>
> **Additional file 6: Table S6.** miRNA targets prediction
>
> **Additional file 7: Table S7**. GO annotation for the genes obtained within the SDRs of chromosome $D_5$02 and $D_5$07.

Abbreviations

SDR: Segregation distortion region; GO: Gene ontology; NRT: Nitrate transporter; ROS: Reactive oxygen species; cM: centiMorgan; QTL: Quantitative trait loci; CRE: Cis- regulatory elements; PPR: Pentatricopeptide repeat; CYPs: Cytochromes P450; CC: Cellular component; MF: Molecular function; LEA: Late embryogenesis abundant proteins

Acknowledgements

We are indebted to the entire cotton biology research team for their immense support and technical assistance in the course of this research work.

Authors' contributions

Kirungu JN, Magwanga RO, Wang K, and Liu F conceptualized the concept; Kirungu JN, Magwanga RO, Wang K, Shiraku ML, Mehari TG, and Liu F performed data curation; Kirungu JN, Magwanga RO, Wang K, and Liu F performed formal analysis; Kirungu JN, Magwanga RO, Wang K, and Liu F are responsive to funding acquisition; Kirungu JN, Magwanga RO, Wang K, Liu F, Zhou Z, Pu L, Xu Y, Hou Y, Zhou Y, Cai X, Agong SG, Wang K and Liu F are responsive to plant resources; Kirungu JN, Magwanga RO, Wang K, Agong SG and Liu F alid the results; Kirungu JN and Magwanga RO wrote the original draft; Kirungu JN and Magwanga RO reviewed & edited the final manuscript. All authors approved the final manuscript.

Funding

This research program was financially sponsored by the National Key Research and Development Plan (2016YFD0100306) and the National Natural Science Foundation of China (31671745, 31530053).

Availability of data and materials

All files supporting the findings are included within the manuscripts as figures, tables, and supplementary files.

Ethics approval and consent to participate

No ethical nor consent to participate in this research was sought.

Consent for publication

No consent to publish the work was sort.

Competing interests

The authors declare no form of competing interest.

Author details

[1]State Key Laboratory of Cotton Biology /Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang 455000, Henan, China. [2]School of Biological, Physical, Mathematics and Actuarial Sciences (SBPMAS), Jaramogi Oginga Odinga University of Science and Technology (JOOUST), 210-40601, Bondo, Kenya. [3]School of Life Science, Henan University/State Key Laboratory of Cotton Biology/Henan Key Laboratory of Plant Stress Biology, Keifang 475004, Henan, China. [4]School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, Henan, China.

References

Abdurakhmonov IY, Devor EJ, Buriev ZT, et al. Small RNA regulation of ovule development in the cotton plant, G hirsutum L. BMC Plant Biol. 2008;8:93. https://doi.org/10.1186/1471-2229-8-93 .

Aichi M, Yoshihara S, Yamashita M, et al. Characterization of the nitrate-nitrite transporter of the major facilitator superfamily (the nrtP gene product) from the cyanobacterium Nostoc punctiforme strain ATCC 29133. Biosci Biotechnol Biochem. 2006;70:2682–9. https://doi.org/10.1271/bbb.60286 .

Amudha J, Balasubramani G, Malathi VG, et al. Segregation pattern of gene expression in cotton leaf curl virus-resistant transgenics. Arch Phytopathol Plant Protect. 2012;45:487–98. https://doi.org/10.1080/03235408.2011.587987 .

Anhalt UCM, Heslop-Harrison PJS, Byrne S, et al. Segregation distortion in Lolium: evidence for genetic effects. Theor Appl Genet. 2008;117:297–306. https://doi.org/10.1007/s00122-008-0774-7 .

Bai H, Euring D, Volmer K, et al. The nitrate transporter (NRT) gene family in poplar. PLoS One. 2013;8:e72126. https://doi.org/10.1371/journal.pone.0072126 .

Baker CC, Sieber P, Wellmer F, et al. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. Curr Biol. 2005;15:303–15. https://doi.org/10.1016/j.cub.2005.02.017 .

Bovill WD, Ma W, Ritter K, et al. Identification of novel QTL for resistance to crown rot in the doubled haploid wheat population "W21MMT70" x "Mendos". Plant Breed. 2006;125:538–43. https://doi.org/10.1111/j.1439-0523.2006.01251.x .

Cai X, Magwanga RO, Xu Y, et al. Comparative transcriptome, physiological and biochemical analyses reveal response mechanism mediated by CBF4 and ICE2 in enhancing cold stress tolerance in Gossypium thurberi. AoB Plants. 2019; 11:1–17. https://doi.org/10.1093/aobpla/plz045 .

Chandnani R, Wang B, Draye X, et al. Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species. Theor Appl Genet. 2017;130:2219–30. https://doi.org/10.1007/s00122-017-2952-y .

Cheng J, Zhao Z, Li B, et al. A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in capsicum. Sci Rep. 2016;6:1–12. https://doi.org/10.1038/srep18919 .

Cloutier S, Ragupathy R, Miranda E, et al. Integrated consensus genetic and physical maps of flax (Linum usitatissimum L.). Theor Appl Genet. 2012;125:1783–95. https://doi.org/10.1007/s00122-012-1953-0 .

Corbett-Detig R, Medina P, Frérot H, et al. Bulk pollen sequencing reveals rapid evolution of segregation distortion in the male germline of Arabidopsis hybrids. Evol Lett. 2019;3:93–103. https://doi.org/10.1002/evl3.96 .

Coulton A, Przewieslik-Allen AM, Burridge AJ, et al. Segregation distortion: utilizing simulated genotyping data to evaluate statistical methods. PLoS One. 2020;15(2):e0228951. https://doi.org/10.1371/journal.pone.0228951 .

Dai B, Guo H, Huang C, et al. Identification and characterization of segregation distortion loci on cotton chromosome 18. Front Plant Sci. 2017;7:2037. https://doi.org/10.3389/fpls.2016.02037 .

Dawes H, Boyes S, Keene J, et al. Protein instability of wines: influence of protein isolelectric point. Am J Enol Vitic. 1994;45:319–26.

Dixit S, Huang BE, Sta Cruz MT, et al. QTLs for tolerance of drought and breeding for tolerance of abiotic and biotic stress: an integrated approach. PLoS ONE. 2014;9(10):e109574. https://doi.org/10.1371/journal.pone.0109574 .

Dong Q, Magwanga R, Cai X, et al. RNA-sequencing, physiological and RNAi analyses provide insights into the response mechanism of the ABC-mediated resistance to Verticillium dahliae infection in cotton. Genes. 2019;10:110. https://doi.org/10.3390/genes10020110 .

Espinoza C, Medina C, Somerville S, et al. Senescence-associated genes induced during compatible viral interactions with grapevine and *Arabidopsis*. J Exp Bot. 2007;58:3197–212. https://doi.org/10.1093/jxb/erm165 .

Faleiro FG, Schuster I, Ragagnin VA, et al. Characterization of recombinant inbred lines and QTL mapping associated to the cycle and yield of common bean. Pesqui Agropecu Bras. 2003;38:1387–97.

Fans JD, Laddomada B, Gill BS. Molecular mapping of segregation distortion loci in *Aegilops tauschii*. Genetics. 1998;149:319–27.

Golestan Hashemi FS, Rafii MY, Ismail MR, et al. The genetic and molecular origin of natural variation for the fragrance trait in an elite Malaysian aromatic rice through quantitative trait loci mapping using SSR and gene-based markers. Gene. 2015;555:101–7. https://doi.org/10.1016/j.gene.2014.10.048 .

Higo K, Ugawa Y, Iwamoto M, et al. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 1999;27:297–300.

Huang L, Deng X, Li R, et al. A fast silver staining protocol enabling simple and efficient detection of SSR markers using a non-denaturing polyacrylamide gel. J Vis Exp. 2018; 134:e57192. https://doi.org/10.3791/57192 .

Huang T, Lopez-Giraldez F, Townsend JP, et al. RBE controls microRNA164 expression to effect floral organogenesis. Development. 2012;139:2161–9. https://doi.org/10.1242/dev.075069 .

Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics. 2008;9:118. https://doi.org/10.1186/1471-2164-9-118 .

Jamshed M, Jia F, Gong J, et al. Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. BMC Genomics. 2016;17:1–13. https://doi.org/10.1186/s12864-016-2560-2 .

Jia X, Wang WX, Ren L, et al. Differential and dynamic regulation of miR398 in response to ABA and salt stress in *Populus tremula* and *Arabidopsis thaliana*. Plant Mol Biol. 2009;71:51–9. https://doi.org/10.1007/s11103-009-9508-8 .

Jun Z, Zhang Z, Gao Y, et al. Overexpression of GbRLK, a putative receptor-like kinase gene, improved cotton tolerance to Verticillium wilt. Sci Rep. 2015;5: 15048. https://doi.org/10.1038/srep15048 .

Khan MKR, Chen H, Zhou Z, et al. Genome wide SSR high density genetic map construction from an interspecific cross of *Gossypium hirsutum* × *Gossypium tomentosum*. Front Plant Sci. 2016;7:436. https://doi.org/10.3389/fpls.2016.00436 .

Kirungu JN, Deng Y, Cai X, et al. Simple sequence repeat (SSR) genetic linkage map of D genome diploid cotton derived from an interspecific cross between *Gossypium davidsonii* and *Gossypium klotzschianum*. Int J Mol Sci. 2018;19:204. https://doi.org/10.3390/ijms19010204 .

Kumar S, Gill BS, Faris JD. Identification and characterization of segregation distortion loci along chromosome 5B in tetraploid wheat. Mol Gen Genomics. 2007;278:187–96. https://doi.org/10.1007/s00438-007-0248-7 .

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4. https://doi.org/10.1093/molbev/msw054 .

Kumari JR, Srikumari CR, Valenzuela CY. ABO segregation distortion in Visakhapatnam, India. Anthropol Anz. 1992;50:307–14.

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559 .

Li P, Kirungu JN, Lu H, et al. SSR-linkage map of interspecific populations derived from *Gossypium trilobum* and *Gossypium thurberi* and determination of genes harbored within the segregating distortion regions. PLoS One. 2018;13: e0207271. https://doi.org/10.1371/journal.pone.0207271 .

Li W, Lin Z, Zhang X. A novel segregation distortion in intraspecific population of asian cotton (*Gossypium arboretum* L.) detected by molecular markers. J Genet Genomics. 2007;34:634–40. https://doi.org/10.1016/S1673-8527(07)60072-1 .

Li X, Jin X, Wang H, et al. Structure, evolution, and comparative genomics of tetraploid cotton based on a high-density genetic linkage map. DNA Res. 2016;23:283–93. https://doi.org/10.1093/dnares/dsw016 .

Liberman U, Feldman MW. On the evolution of fluctuating segregation distortion. Theor Popul Biol. 1982;21:301–17. https://doi.org/10.1016/0040-5809(82)90020-X .

Liu F, Wu XL, Chen SY. Segregation distortion of molecular markers in recombinant inbred populations in soybean (*G. max*). Acta Genet Sin. 2000; 27(10):883–7 (in Chinese with English abstract).

Liu X, You J, Guo L, et al. Genetic analysis of segregation distortion of SSR markers in F_2 population of barley. J Agric Sci. 2011;3:172–7. https://doi.org/10.5539/jas.v3n2p172 .

Lu H, Romero-Severson J, Bernardo R. Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet. 2002;105:622–8. https://doi.org/10.1007/s00122-002-0970-9 .

Lu P, Magwanga RO, Lu H, et al. A novel G-protein-coupled receptors gene from upland cotton enhances salt stress tolerance in transgenic *Arabidopsis*. Genes (Basel). 2018;9(4):209. https://doi.org/10.3390/genes9040209 .

Lu Y, Feng Z, Bian L, et al. miR398 regulation in rice of the responses to abiotic and biotic stresses depends on CSD1 and CSD2 expression. Funct Plant Biol. 2010;38:44–53. https://doi.org/10.1071/FP10178 .

Ma S, Gong Q, Bohnert HJ. Dissecting salt stress pathways. J Exp Bot. 2006;57(5): 1097–107. https://doi.org/10.1093/jxb/erj098 .

Magwanga R, Lu P, Kirungu J, et al. GBS mapping and analysis of genes conserved between *Gossypium tomentosum* and *Gossypium hirsutum* cotton cultivars that respond to drought stress at the seedling stage of the BC_2F_2 generation. Int J Mol Sci. 2018a;19:1614. https://doi.org/10.3390/ijms19061614 .

Magwanga RO, Lu P, Kirungu JN, et al. Identification of QTLs and candidate genes for physiological traits associated with drought tolerance in cotton. J Cotton Res. 2020;3:3. https://doi.org/10.1186/s42397-020-0043-0 .

Magwanga RO, Lu P, Kirungu JN, et al. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. BMC Genet. 2018b;19:6. https://doi.org/10.1186/s12863-017-0596-1 .

Magwanga RO, Lu P, Kirungu JN, et al. Identification of cotton cyclin dependent kinase (CDK) genes and overexpression of *Gh_D12G2017 (CDKF4)* confer drought and salt stress tolerance in transgenic *Arabidopsis*. Int J Mol Sci. 2018c;19(9):2625.

Manrique-Carpintero NC, Coombs JJ, Veilleux RE, et al. Comparative analysis of regions with distorted segregation in three diploid populations of potato. G3: Genes Genomes Genetics. 2016;6:2617–28. https://doi.org/10.1534/g3.116.030031 .

McLaughlin RN, Malik HS. Genetic conflicts: the usual suspects and beyond. J Exp Biol. 2017;220:6–17. https://doi.org/10.1242/jeb.148148 .

Mello CC, Kramer JM, Stinchcomb D, et al. Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 1991;10(12):3959–70.

Mendoza CP, Ulloa M, Abdurakhmonov IY, et al. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. Botany. 2013;11:54. https://doi.org/10.1139/cjb-2012-0192 .

Nadeau JH. Do gametes woo? Evidence for their nonrandom union at fertilization. Genetics. 2017;207:369–87. https://doi.org/10.1534/genetics.117.300109 .

Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol. 2009; 149:88–95. https://doi.org/10.1104/pp.108.129791 .

Natwick E. Resistance to silverleaf whitefly, *Bemisia argentifolii* (Hem, Aleyrodidae), in *Gossypium thurberi*, a wild cotton species. J Appl Entomol. 2006. https://doi.org/10.1111/j.1439-0418.2006.01083.x .

Pashkovskii PP, Ryazanskii SS, Radyukina NL, et al. MIR398 and expression regulation of the cytoplasmic cu/Zn-superoxide dismutase gene in *Thellungiella halophila* plants under stress conditions. Russ J Plant Physiol. 2010;57:707–14. https://doi.org/10.1134/S1021443710050146 .

Rahman MA, Thomson MJ, De Ocampo M, et al. Assessing trait contribution and mapping novel QTL for salinity tolerance using the Bangladeshi rice landrace capsule. Rice. 2019;12:63. https://doi.org/10.1186/s12284-019-0319-5 .

Reflinur, Kim B, Jang SM, et al. Analysis of segregation distortion and its relationship to hybrid barriers in rice. Rice. 2014;7:3. https://doi.org/10.1186/s12284-014-0003-8 .

Rouxel T, Balesdent MH. Avirulence genes. In: Encyclopedia of life sciences. Chichester: John Wiley & Sons Ltd; 2010. https://doi.org/10.1002/9780470015902.a0021267 .

Sakharkar KR, Sakharkar MK, Culiat CT, et al. Functional and evolutionary analyses on expressed intronless genes in the mouse genome. FEBS Lett. 2006;580: 1472–8. https://doi.org/10.1016/j.febslet.2006.01.070 .

Saminathan T, Bodunrin A, Singh NV, et al. Genome-wide identification of microRNAs in pomegranate (*Punica granatum* L.) by high-throughput sequencing. BMC Plant Biol. 2016;16(1):122. https://doi.org/10.1186/s12870-016-0807-3 .

Sandler L, Golic K. Segregation distortion in drosophila. Trends Genet. 1985;1: 181–5. https://doi.org/10.1016/0168-9525(85)90074-5 .

Shang L, Wang Y, Wang X, et al. Genetic analysis and QTL detection on fiber traits using two recombinant inbred lines and their backcross populations in

upland cotton. G3 (Bethesda). 2016;6:2717–24. https://doi.org/10.1534/g3.116.031302 .

Sunkar R. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. Plant Cell Online. 2006;18:2051–65. https://doi.org/10.1105/tpc.106.041673 .

Sunkar R, Chinnusamy V, Zhu J, et al. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci. 2007;12:301–9. https://doi.org/10.1016/j.tplants.2007.05.001 .

Takumi S, Motomura Y, Iehisa JCM, et al. Segregation distortion caused by weak hybrid necrosis in recombinant inbred lines of common wheat. Genetica. 2013;141:463–70. https://doi.org/10.1007/s10709-013-9745-2 .

Trivedi DK, Gill SS, Tuteja N. Abscisic acid (ABA): biosynthesis, regulation, and role in abiotic stress tolerance. In: Tuteja N, Gill SS, editors. Abiotic stress response in plants. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2016. p. 315–26. https://doi.org/10.1002/9783527694570.ch15 .

Tsilo TJ, Jin Y, Anderson JA. Diagnostic microsatellite markers for the detection of stem rust resistance gene Sr36 in diverse genetic backgrounds of wheat. Crop Sci. 2008;48:253–61. https://doi.org/10.2135/cropsci2007.04.0204 .

Tümpel S, Cambronero F, Wiedemann LM, et al. Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (*Takifugu rubripes*). Proc Natl Acad Sci U S A. 2006;103:5419–24. https://doi.org/10.1073/pnas.0600993103 .

Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93:77–8. https://doi.org/10.1093/jhered/93.1.77 .

Wang G, He QQ, Xu ZK, et al. High segregation distortion in maize B73 x teosinte crosses. Genet Mol Res. 2012;11:693–706. https://doi.org/10.4238/2012.March.19.3 .

Wang S, Tan Y, Tan X, et al. Segregation distortion detected in six rice F_2 populations generated from reciprocal hybrids at three altitudes. Genet Res. 2009;91:345–53. https://doi.org/10.1017/S0016672309990176 .

Wang X, Yang B, Li K, et al. A conserved *Puccinia striiformis* protein interacts with wheat NPR1 and reduces induction of pathogenesis-related genes in response to pathogens. Mol Plant-Microbe Interact. 2016;29:977–89. https://doi.org/10.1094/MPMI-10-16-0207-R .

Wang X, Zhang Y, Qiao L, et al. Comparative analyses of simple sequence repeats (SSRs) in 23 mosquito species genomes : identification, characterization and distribution (*Diptera : Culicidae*). J Insect Sci. 2019:607–19. https://doi.org/10.1111/1744-7917.12577 .

Wei Y, Xu Y, Lu P, et al. Salt stress responsiveness of a wild cotton species (*Gossypium klotzschianum*) based on transcriptomic analysis. PLoS One. 2017;12(5):e0178313. https://doi.org/10.1371/journal.pone.0178313 .

Wu JH, Zhang XL, Luo XL, et al. Inheritance and segregation of transformants in cotton with two types of insect-resistant genes. Acta Genet Sin. 2003;30:631–6. (in Chinese with English abstract).

Wu YP, Ko PY, Lee WC, et al. Comparative analyses of linkage maps and segregation distortion of two F_2 populations derived from *japonica* crossed with *indica* rice. Hereditas. 2010;147:225–36. https://doi.org/10.1111/j.1601-5223.2010.02120.x .

Xie Q, Frugis G, Colgan D, et al. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 2000;14:3024–36. https://doi.org/10.1101/gad.852200 .

Xu Y, Zhu L, Xiao J, et al. Chromosomal regions associated with segregation distortion of molecular markers in F_2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet. 1997;253:535–45. https://doi.org/10.1007/s004380050355 .

Yan H, Dai X, Feng K, et al. IGDD: A database of intronless genes in dicots. BMC Bioinformatics. 2016;17:289. https://doi.org/10.1186/s12859-016-1148-9 .

Yang C, Wang Z, Yang X, et al. Segregation distortion affected by transgenes in early generations of rice crop-weed hybrid progeny: implications for assessing potential evolutionary impacts from transgene flow into wild relatives. J Syst Evol. 2014;52:466–76. https://doi.org/10.1111/jse.12078 .

Yang RC, Thiagarajah MR, Bansal VK, et al. Detecting and estimating segregation distortion and linkage between glufosinate tolerance and blackleg resistance in *Brassica napus* L. Euphytica. 2006;148:217–25. https://doi.org/10.1007/s10681-005-9003-5 .

Yu Y, Yuan D, Liang S, et al. Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. BMC Genomics. 2011;12:15. https://doi.org/10.1186/1471-2164-12-15 .

Yuan JZ, Peng N, Feng CJ, et al. Effect of marker segregation distortion on high density linkage map construction and QTL mapping in Soybean (*Glycine max* L.). Heredity. 2019;123:579–92. https://doi.org/10.1038/s41437-019-0238-7 .

Zhang F, Zhu G, Du L, et al. Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, *Gossypium davidsonii*. Sci Rep. 2016;6:20582. https://doi.org/10.1038/srep20582 .

Zhang J, Stewart JM. Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci. 2000;4:193–201.

Zhang Y, Wang L, Xin H, et al. Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol. 2013;13:1–12. https://doi.org/10.1186/1471-2229-13-141 .

Zhao J, Gao Y, Zhang Z, et al. A receptor-like kinase gene (*GbRLK*) from *Gossypium barbadense* enhances salinity and drought-stress tolerance in *Arabidopsis*. BMC Plant Biol. 2013;13:110. https://doi.org/10.1186/1471-2229-13-110 .