



Article

# Simple Sequence Repeat (SSR) Genetic Linkage Map of D Genome Diploid Cotton Derived from an Interspecific Cross between *Gossypium davidsonii* and *Gossypium klotzschianum*

Joy Nyangasi Kirungu <sup>1,†</sup>, Yanfeng Deng <sup>1,†</sup>, Xiaoyan Cai <sup>1</sup>, Richard Odongo Magwanga <sup>1,2,†</sup> , Zhongli Zhou <sup>1</sup>, Xingxing Wang <sup>1</sup>, Yuhong Wang <sup>1</sup>, Zhenmei Zhang <sup>1</sup>, Kunbo Wang <sup>1,\*</sup> and Fang Liu <sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Cotton Biology/Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China; joynk@cricaas.com.cn (J.N.K.); dengyf@cricaas.com.cn (Y.D.); caixy@cricaas.com.cn (X.C.); richard@cricaas.com.cn (R.O.M.); zhouzl@cricaas.com.cn (Z.Z.); wangxx@cricaas.com.cn (X.W.); wangyh@cricaas.com.cn (Y.W.); zhangzm@cricaas.com.cn (Z.Z.)

<sup>2</sup> School of Biological and Physical Sciences (SBPS), Main Campus, Jaramogi Oginga Odinga University of Science and Technology (JOUST), Main Campus, P.O. Box 210-40601 Bondo, Kenya

\* Correspondence: wkbcri@cricaas.com.cn (K.W.); liufang@caas.cn (F.L.)

† These authors contributed equally to this work.

Received: 6 December 2017; Accepted: 7 January 2018; Published: 11 January 2018

**Abstract:** The challenge in tetraploid cotton cultivars is the narrow genetic base and therefore, the bottleneck is how to obtain interspecific hybrids and introduce the germplasm directly from wild cotton to elite cultivars. Construction of genetic maps has provided insight into understanding the genome structure, interrelationships between organisms in relation to evolution, and discovery of genes that carry important agronomic traits in plants. In this study, we generated an interspecific hybrid between two wild diploid cottons, *Gossypium davidsonii* and *Gossypium klotzschianum*, and genotyped 188 F<sub>2:3</sub> populations in order to develop a genetic map. We screened 12,560 SWU Simple Sequence Repeat (SSR) primers and obtained 1000 polymorphic markers which accounted for only 8%. A total of 928 polymorphic primers were successfully scored and only 728 were effectively linked across the 13 chromosomes, but with an asymmetrical distribution. The map length was 1480.23 cM, with an average length of 2.182 cM between adjacent markers. A high percentage of the markers on the map developed, and for the physical map of *G. raimondii*, exhibited highly significant collinearity, with two types of duplication. High level of segregation distortion was observed. A total of 27 key genes were identified with diverse roles in plant hormone signaling, development, and defense reactions. The achievement of developing the F<sub>2:3</sub> population and its genetic map constructions may be a landmark in establishing a new tool for the genetic improvement of cultivars from wild plants in cotton. Our map had an increased recombination length compared to other maps developed from other D genome cotton species.

**Keywords:** wild cotton spp; genetic map; polymorphic primers; null alleles

## 1. Introduction

Wild cotton species have been well studied in terms of their variation in desired traits, such as chemical defenses [1], oil content [2], and fibre production [3,4]. With cotton being the backbone of the textile industries, its continued production with enhanced fibre quality is necessary for the sustainability of the sector [5,6]. The production of cotton is ever-declining due to abiotic and biotic stress: currently the loss in cotton production is estimated to stand at 30% as a result of stress effects [7].

This is projected to continue further due to available lands being highly polluted with increased salinity, and is estimated to cover more than 6% of the world's arable land [8]. In light of this, breeding of tolerant cotton genotypes is important for the existence of the textile industries. Elite cultivated cotton has a narrow genetic base with ever-declining hybrid vigor due to inbreeding [9]. For the continuation of cotton production, increased vigor and broadening of the genetic base is essential. The use of wild progenitors will help to break the bottleneck [10]. The direct use of wild cotton species in improving the elite tetraploid cotton, which has a small amounts of variation due to the specie's monophyletic origin and domestication compounded by intense selection, has proved futile through conventional breeding [4]. Wild cotton species have immense genetic diversity with great potential to improve many important agronomic traits, such as resistance to disease, fibre qualities, and abiotic stress tolerance [11–13]. The valuable agronomic traits in wild cotton progenitors can be exploited effectively to improve cultivated cotton cultivars and solve many problems associated with fibre quality, resistance to insects, pathogens, and tolerance to abiotic stress [11,14].

The majority of elite cotton (*Gossypium hirsutum* L.) varieties have been either bred from cultivars with common ancestry or are known to have related ancestors, but so far only limited increases in productivity have been obtained [15]. The use of D-genome cotton as a source of genetic diversity for gaining stress tolerance, fibre quality, and earlier maturity has been suggested [7,11,16]. *G. davidsonii* and *G. klotzschianum* are both wild diploid cotton species which are endemic to the Sonoran Desert of north-western Mexico [17]. The two cotton species are found in the subsection of *Integrifolia* within the genus *Gossypium* (Family Malvaceae) [17]. They are known to exhibit several characteristic traits such as salt tolerance, and resistance to sucking pests and bacterial blight [18,19].

Genetic linkage maps based on DNA markers are essential tools for plant molecular research in areas such as marker-assisted selection, quantitative trait loci (QTL) mapping, and map-based cloning. A number of DNA markers have been used in the development of genetic linkage maps for cotton [20–24]. The genetic maps have been used to map many beneficial agronomical genes and/or quantitative trait loci (QTLs). Several comparative maps using a common set of amplified fragment length polymorphisms (AFLP), restriction fragment length polymorphisms (RFLP), or SSR markers have been produced for cotton [25,26]. Interspecific linkage maps of diploid cottons have been constructed for the D genome (*G. raimondii* × *G. trilobum*), A genome (*G. arboretum* × *G. herbaceum*), [25,27] and G genome (*G. australe* × *G. nelsonii*) [28]. However, maps derived from wild progenitors are still lacking.

Marker-assisted selection (MAS) is the most recent innovation in molecular breeding for crop improvement [29]. MAS employs the technique of molecular markers in detecting or mining valuable traits in mapping various populations. The markers are developed based on DNA variation among individuals in a breeding population. In contrast to phenotypic selection, MAS is a direct form of genotype selection [30]. The application of markers derived from a specific genome provides the opportunity to detect the inherent traits of a given genotype [31]. SWU markers are recently developed SSR markers from the D genome of *Gossypium raimondii*. To date, no genetic map has been developed from mono-markers of the D genome in cotton. The application of mono-markers provides the best approach for examining genetic variability in a pure line developed from a given genome, as evidenced from SSR markers developed in barley [32]. So far, no comprehensive SSR-based interspecific linkage maps of  $F_{2,3}$  populations derived from two wild cotton generations of D genome in *Gossypium* with closer genetic relatedness have been reported. We managed to develop an  $F_{2,3}$  generation through the interspecific cross of two wild cotton species, *G. davidsonii* and *G. klotzschianum*, which we employed in developing the genetic map. This study aimed to genotype the  $F_{2,3}$  populations and develop fine linkage maps with broader applications in molecular studies, with the net effect of breeding for an enhanced or broader genetic base.

## 2. Results

### 2.1. Parental Polymorphism

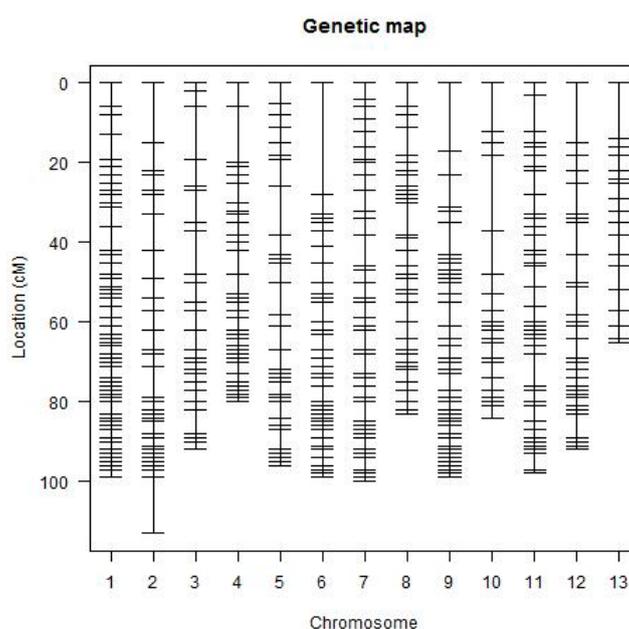
A total of 12,560 SWU SSRs markers derived from *G. raimondii* were used to screen the  $F_{2:3}$  interspecific populations formed using two wild cotton species of D genome: *G. davidsonii* (accession number PI 530809) and *G. klotzschianum* (accession number PI 499748). Among them, 1000 markers were found to be polymorphic, accounting for only 8% of the total markers used. A total of 728 loci were generated by 1000 polymorphic primer pairs. The number of polymorphic dominant loci identified was 200 (27.17%), while 528 (72.83%) non-dominant loci were identified (Table 1).

**Table 1.** Polymorphic rate of screened SWU primers.

Primer	No. of Markers	No. of Polymorphic Primers	Dominant Markers	Non-Dominant Markers	No. of Linked Markers	Polymorphic Rate
SWU	12,650	1000	200 (27.17%)	528 (72.83%)	728	8.00%

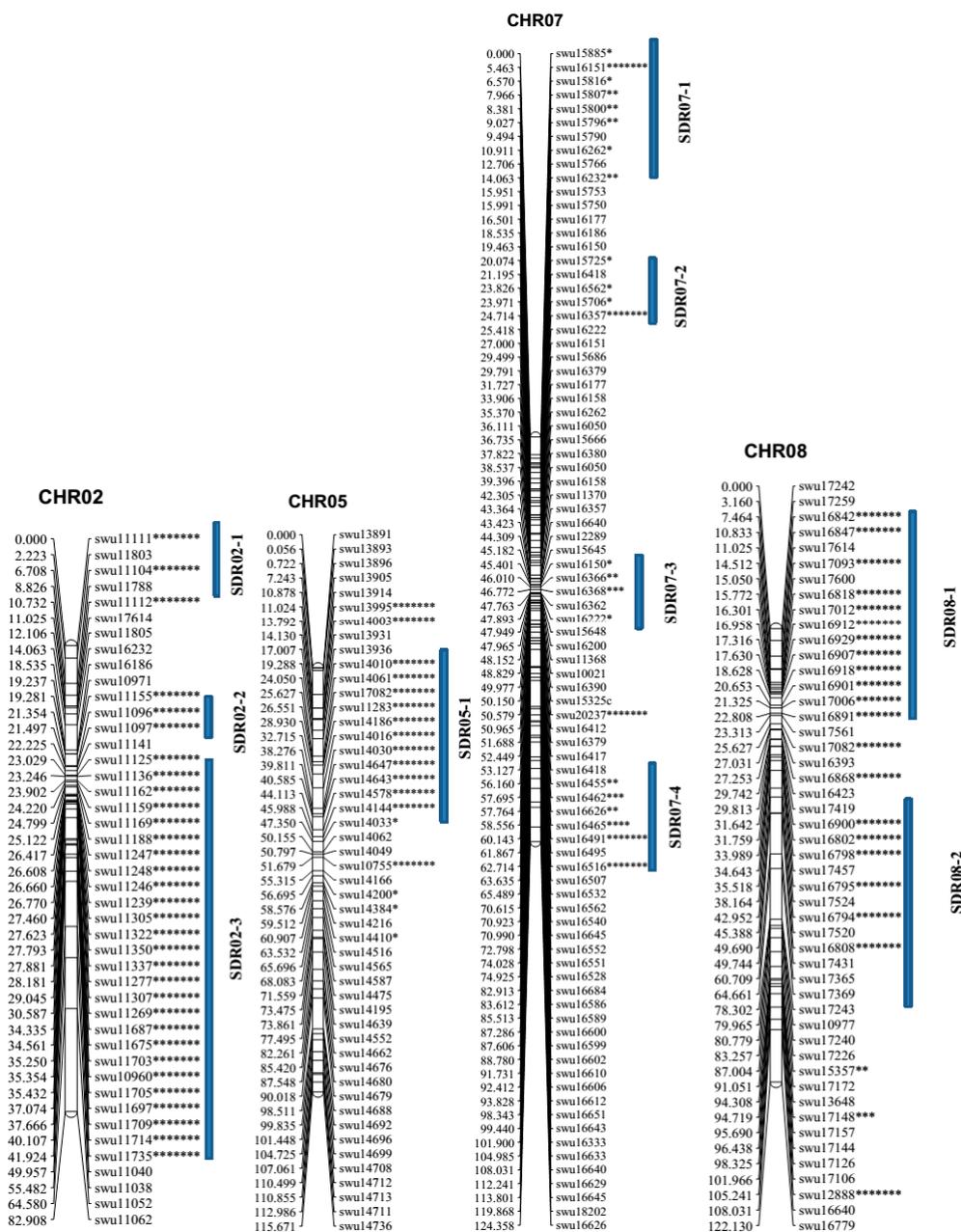
### 2.2. Linkage Analysis and Map Construction

A total of 728 segregating loci were assembled into 13 linkage groups (Figure 1). Two hundred (200) loci remained unlinked to any of the generated linkage groups (LGs). The total length of the map was 1480.225 cM. The average genetic distance between two loci was 2.011 cM, while the largest gap was 36.597 cM between adjacent loci on LG11 (Chr11), and the smallest gap was 0.002 cM on LG06 (Chr06). The largest linkage group consisted of 89 marker loci covering 116.045 cM in LG1 (Chr01), and the smallest was in linkage group LG10 (Chr10) consisting of 34 marker loci covering only 101.93 cM (Figure 1 and Table 2). All the markers were distributed evenly among the linkage groups without clustering of loci. The map obtained through mono-markers had similar attributes to earlier constructed D genome maps, and was composed of 763 loci and with map size of 1493.3 cM in 13 linkage groups [27]. This clearly shows a high level of success in the development of  $F_{2:3}$  populations from two closely related wild cotton genotypes, despite the difficulties faced in the development of the mapping population. The assigned chromosomes are provided in (Supplementary Figure S1).



**Figure 1.** Genetic map constructed using the  $F_{2:3}$  Population derived from the parental lines. The visual presentation was analysed by R software.

A total of 159 distorted loci were detected, which translates to (21.6%) of all the mapped loci. The loci showed significant deviation from the Mendelian ratios at  $p < 0.05$  to  $p < 0.00005$ . The highest level of distortion was observed in two linkage groups with 35 distorted markers on each, accounting for 22.01% of the entire linkage map (Table 3 and Figure 2). Several large clusters of segregation distortions were detected in LG02 (35), LG07 (35), LG08 (25), LG05 (17), and LG09 (12). LG02, LG05, LG07, and LG08 had dominant loci with preferential transmission of *Gossypium klotzschianum* alleles in the F<sub>2:3</sub> populations, while LG01, LG03, LG04, LG06, LG10, and LG12 showed preferential transmission of *Gossypium davidsonii* alleles in the F<sub>2:3</sub> populations (Table 3). In general, the alleles from *Gossypium klotzschianum* were highly preferentially transferred to the F<sub>2:3</sub> populations compared to *Gossypium davidsonii* alleles, as evidenced by levels of distortion segregation of 74 and 12, respectively (Table 3).



**Figure 2.** Linkage groups with the highest segregation distortion. Segregation distortion region (SDR); chromosome 2 has three regions, chromosome 5 has one region, chromosome 7 has four regions, and chromosome 8 has two regions. The marker positions are in centiMorgans (cM).

**Table 2.** Characteristics of the genetic map.

Chr.	Markers per Chromosome	SD	Average Distance (cM)	Average % SD	Map Size (cM)	Marker Distance Gaps (cM) per Chromosome						Ratio
						Smallest Gap (cM)	Largest Gap in (cM)	<5 cM	5–10 cM	10–20 cM	>20 cM	
Chr01	89	3	1.304	3.371	116.045	0.005	8.756	83	5	0	0	0.9326
Chr02	44	35	1.884	76.087	82.908	0.044	18.328	39	3	1	0	0.8864
Chr03	45	8	2.59	17.778	116.528	0.14	10.009	39	1	1	0	0.8667
Chr04	56	2	1.997	3.571	111.846	0.008	17.538	53	1	1	0	0.9464
Chr05	49	17	2.361	34.694	115.671	0.056	8.493	45	5	0	0	0.9184
Chr06	58	5	2.001	8.621	116.045	0.002	14.926	52	3	2	0	0.8966
Chr07	86	35	1.446	40.698	124.358	0.016	7.988	81	4	0	0	0.9419
Chr08	49	25	2.492	51.02	122.13	0.054	14.099	45	0	3	0	0.9184
Chr09	69	12	1.697	16.216	117.06	0.046	11.167	63	3	2	0	0.9130
chr10	34	2	2.998	5.882	101.93	0.117	12.472	25	7	1	0	0.7353
Chr11	63	5	1.806	7.937	113.801	0.042	36.597	60	1	0	1	0.9524
Chr12	49	6	2.301	12.245	112.739	0.065	9.588	43	5	0	0	0.8776
Chr13	37	4	3.491	10.526	129.164	0.137	25.3	28	7	0	1	0.7568
<b>Genetic map</b>	<b>728</b>	<b>159</b>	<b>2.182</b>	<b>22.20</b>	<b>1480.23</b>	<b>0.0563</b>	<b>15.020</b>	<b>656</b>	<b>45</b>	<b>11</b>	<b>2</b>	<b>0.9011</b>

Chr chromosome; SD segregation distortion; cM centiMorgan; % percentage; <less than; The ratio is the number of markers less than 5 divided by the number markers in the same linkage group.

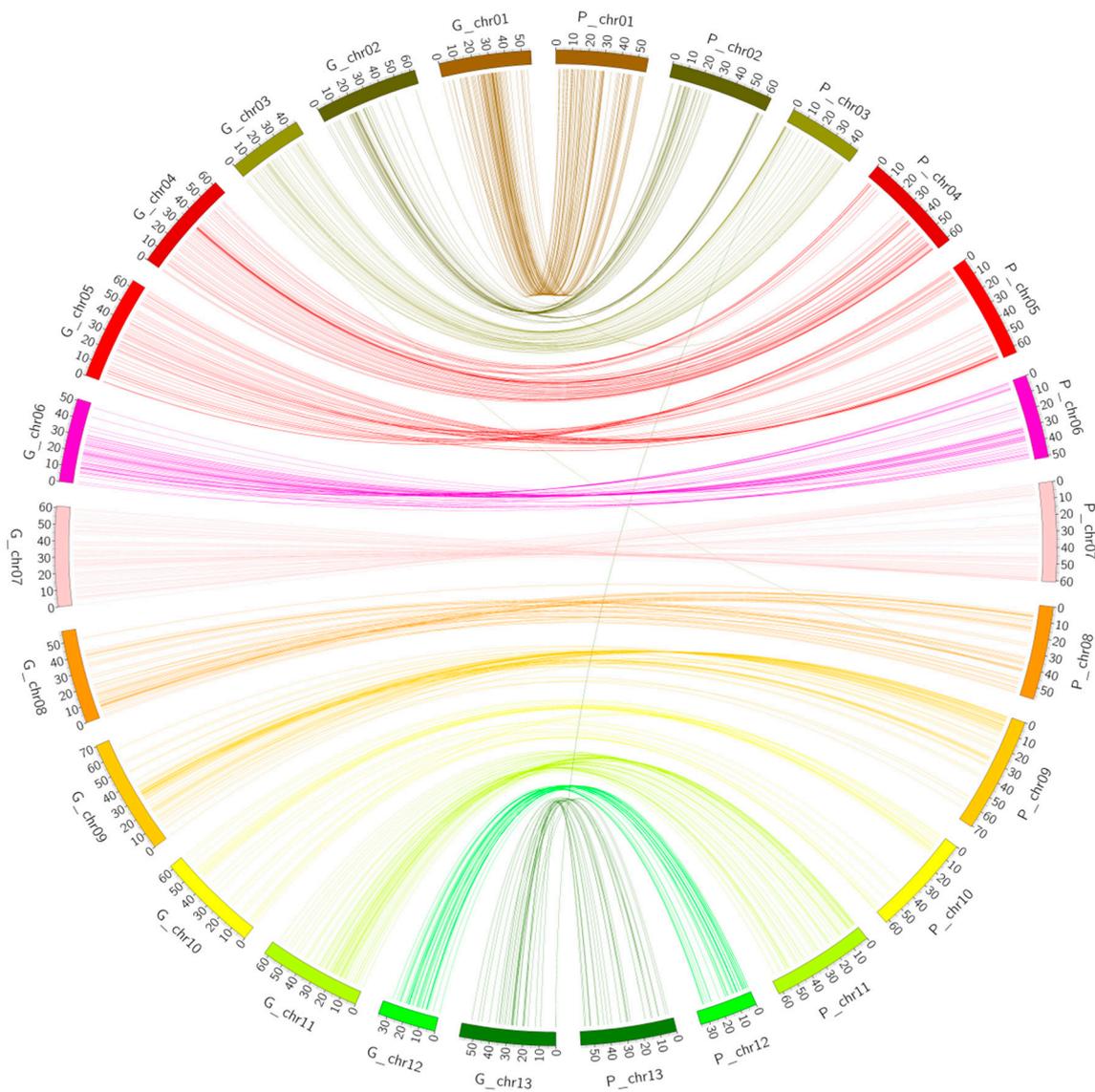
**Table 3.** Segregation distortion in the F<sub>2:3</sub> interspecific populations derived from two wild cottons of D genome.

Chromosome	Chi Square ( $\chi^2$ ) Values							Loci Number			
	<i>p</i> < 0.05	<i>p</i> < 0.01	<i>p</i> < 0.005	<i>p</i> < 0.001	<i>p</i> < 0.0005	<i>p</i> < 0.0001	<i>p</i> < 0.00005	<i>G. davidsonii</i>	<i>G. klotzschianum</i>	Heterozygote's	Totals
Chr01	0	1	0	0	0	0	2	2	0	1	3
Chr02	1	0	0	0	0	0	34	0	34	1	35
Chr03	2	2	1	0	0	0	3	1	0	7	8
Chr04	0	1	0	0	0	0	1	1	0	1	2
Chr05	3	1	0	0	0	0	13	0	11	6	17
Chr06	0	1	0	0	0	0	4	4	0	1	5
Chr07	11	10	5	1	1	0	7	1	5	29	35
Chr08	2	0	1	0	0	0	22	0	22	3	25
Chr09	5	5	1	0	1	0	0	0	0	12	12
Chr10	0	0	0	0	1	0	1	1	0	1	2
Chr11	3	1	0	1	0	0	0	0	0	5	5
Chr12	3	0	0	0	0	0	3	2	1	3	6
Chr13	2	0	0	0	0	0	2	0	1	3	4
<b>Total</b>	<b>32</b>	<b>22</b>	<b>8</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>92</b>	<b>12</b>	<b>74</b>	<b>73</b>	<b>159</b>

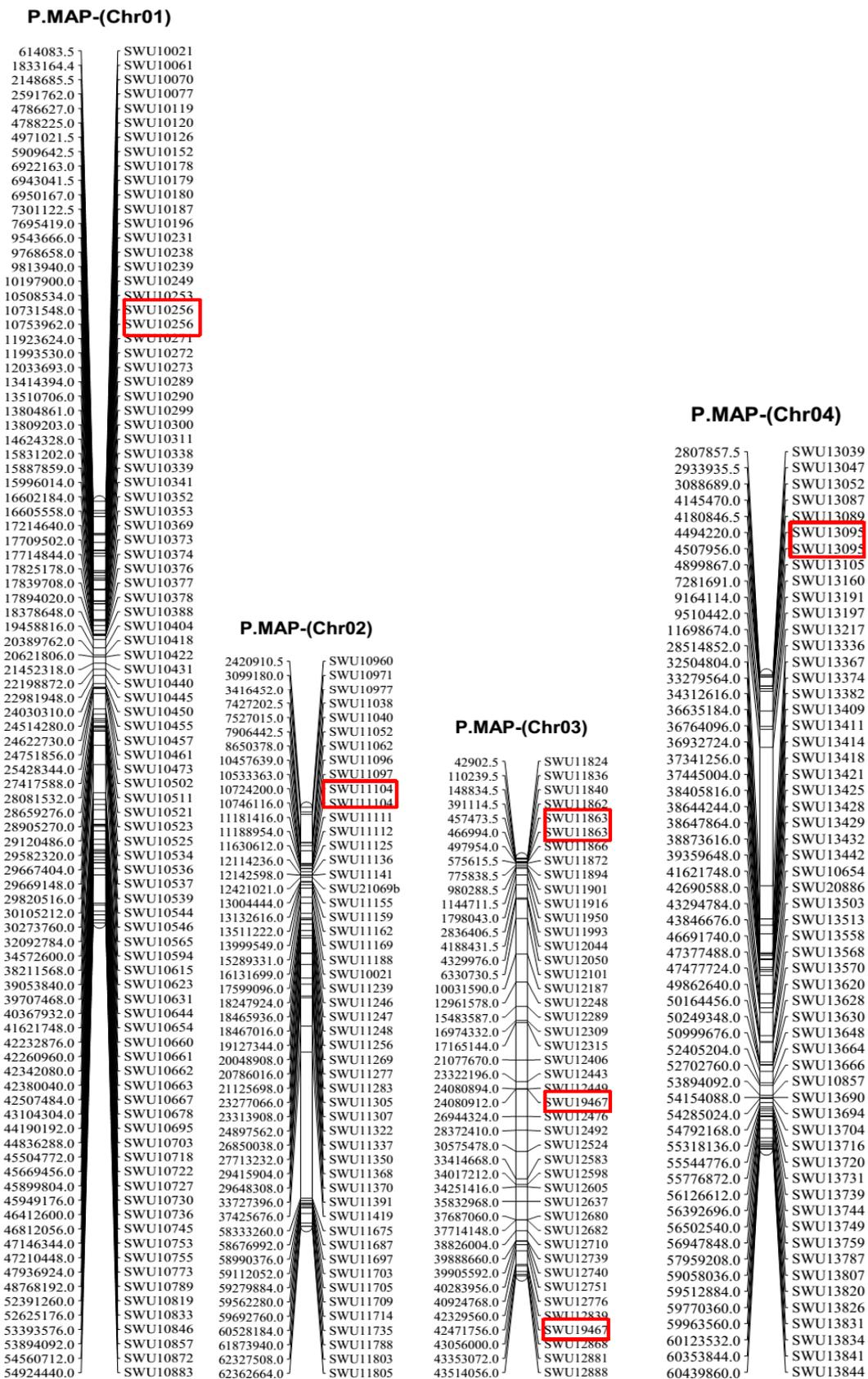
*p* < 0.05; *p* < 0.01; *p* < 0.005; *p* < 0.001; *p* < 0.0005; *p* < 0.0001; *p* < 0.00005 significant levels at 0.05, 0.01, 0.005, 0.001, 0.0005, 0.00001, and 0.00005 respectively.

### 2.3. Collinearity Analysis

Syntenic blocks were obtained by comparing the 728 markers' mapped positions on both genetic map and the physical map using D cotton genome as the reference. The results showed that most of the markers had good collinearity (Figure 3). However, duplication was observed; two types of duplication were noted: intra- and inter-chromosomal duplication. The majority of the duplication was observed for inter-chromosomal as opposed to intra-chromosomal duplication. A total of 54 markers showed inter-chromosomal duplication while only 11 had intra-chromosomal duplication (Figure 4 and Supplementary Figure S2). Similar observations have been reported in the A genome, in which inter-chromosomal duplicated loci were recorded to be twice the number of intra-chromosomal duplicated loci [33].



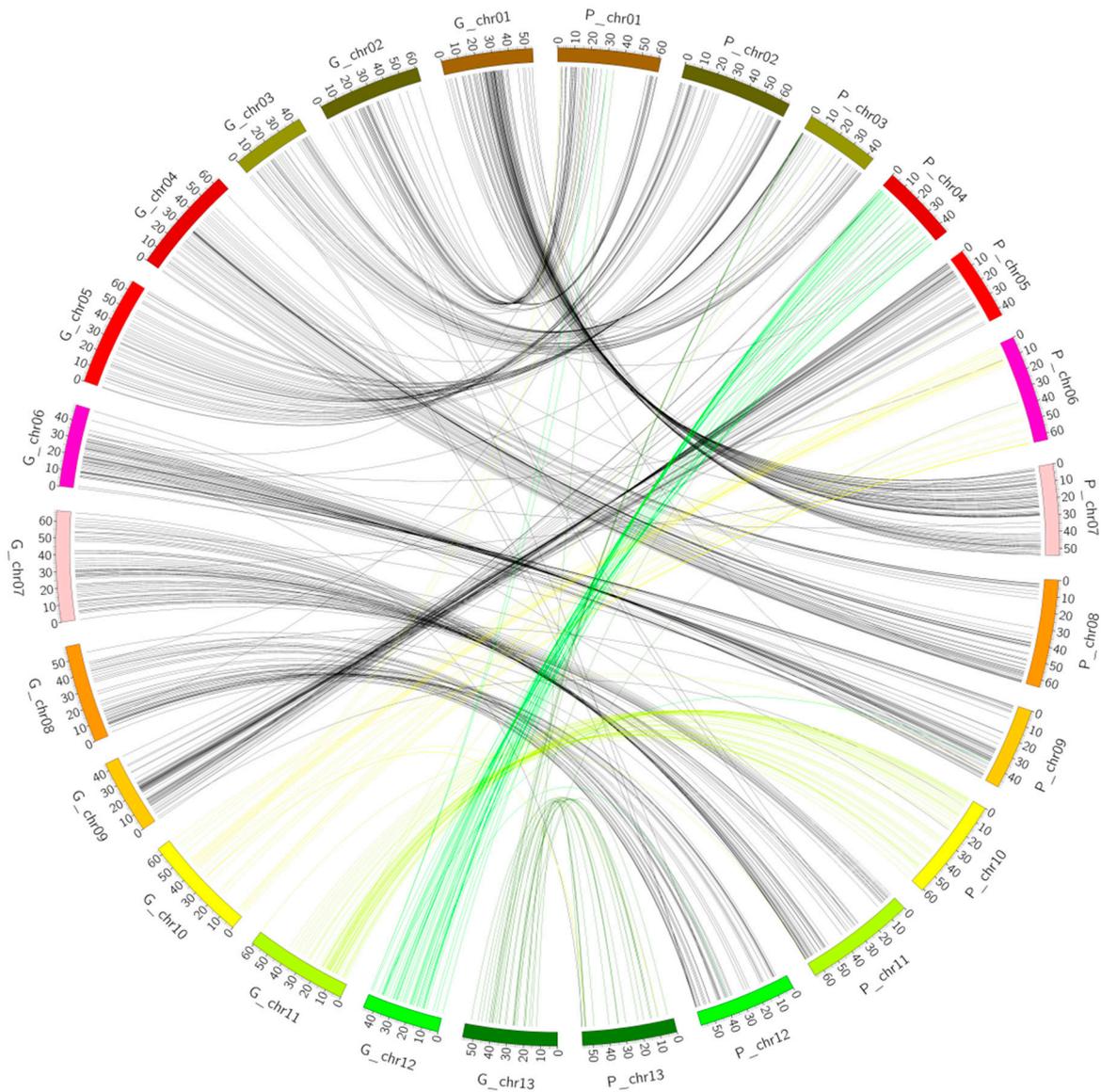
**Figure 3.** Collinearity between the genetic map and the physical map of diploid cotton.



**Figure 4.** A section of physical map (P.MAP) with intra-chromosomal duplication. The red boxes indicate the duplicated markers. The marker positions are in base pairs (Bp).

#### 2.4. Collinearity Analysis of the Genetic Map and the Physical Map of the (Dt) Sub-Genome

A total of 728 markers were blasted against the AD genome of *G. hirsutum* and 633 matches were obtained. Upon removal of the redundant markers, 599 markers were finally used in the collinearity analysis. The markers were used to analyze the collinearity blocks between the genetic map and the physical map of the Dt sub-genome of *G. hirsutum* (Figure 5 and Supplementary Table S1). The syntenic blocks showed good collinearity between the two maps.

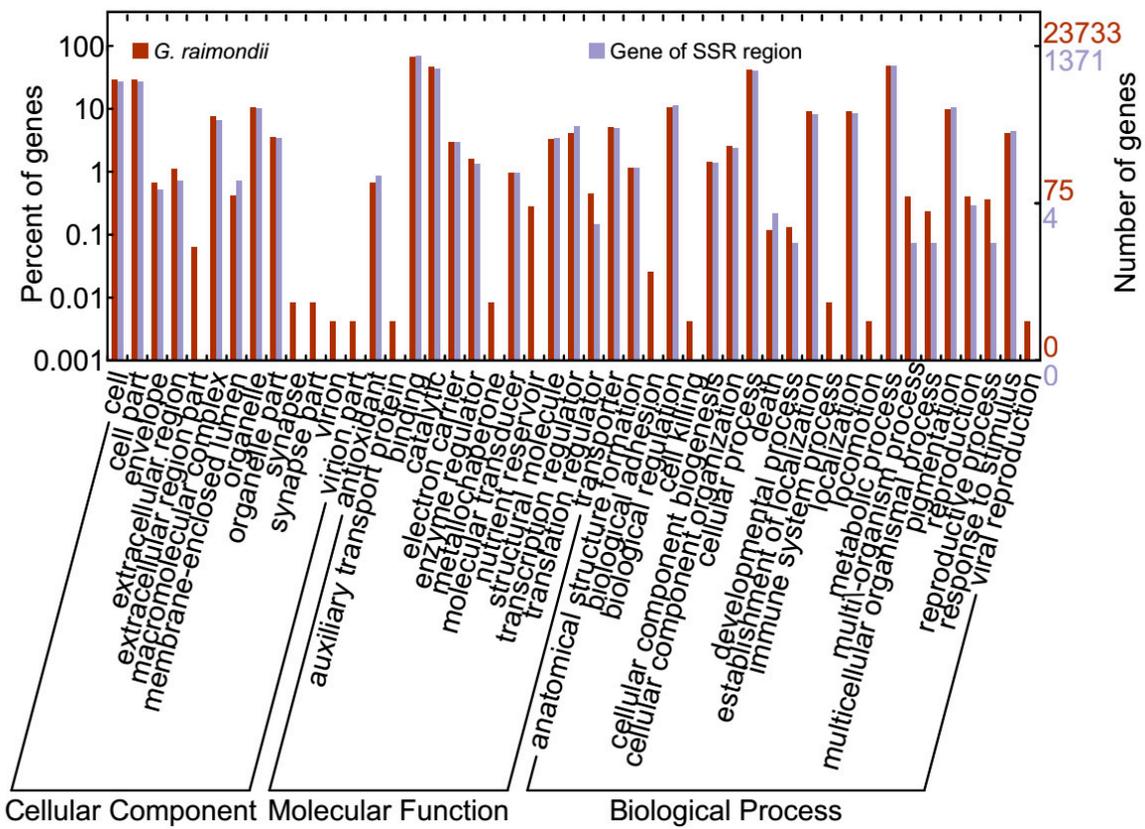


**Figure 5.** Collinearity between the genetic map and the physical map of the Dt sub-genome of *G. hirsutum*. The different line colors represent the various syntenic block regions between the chromosomes.

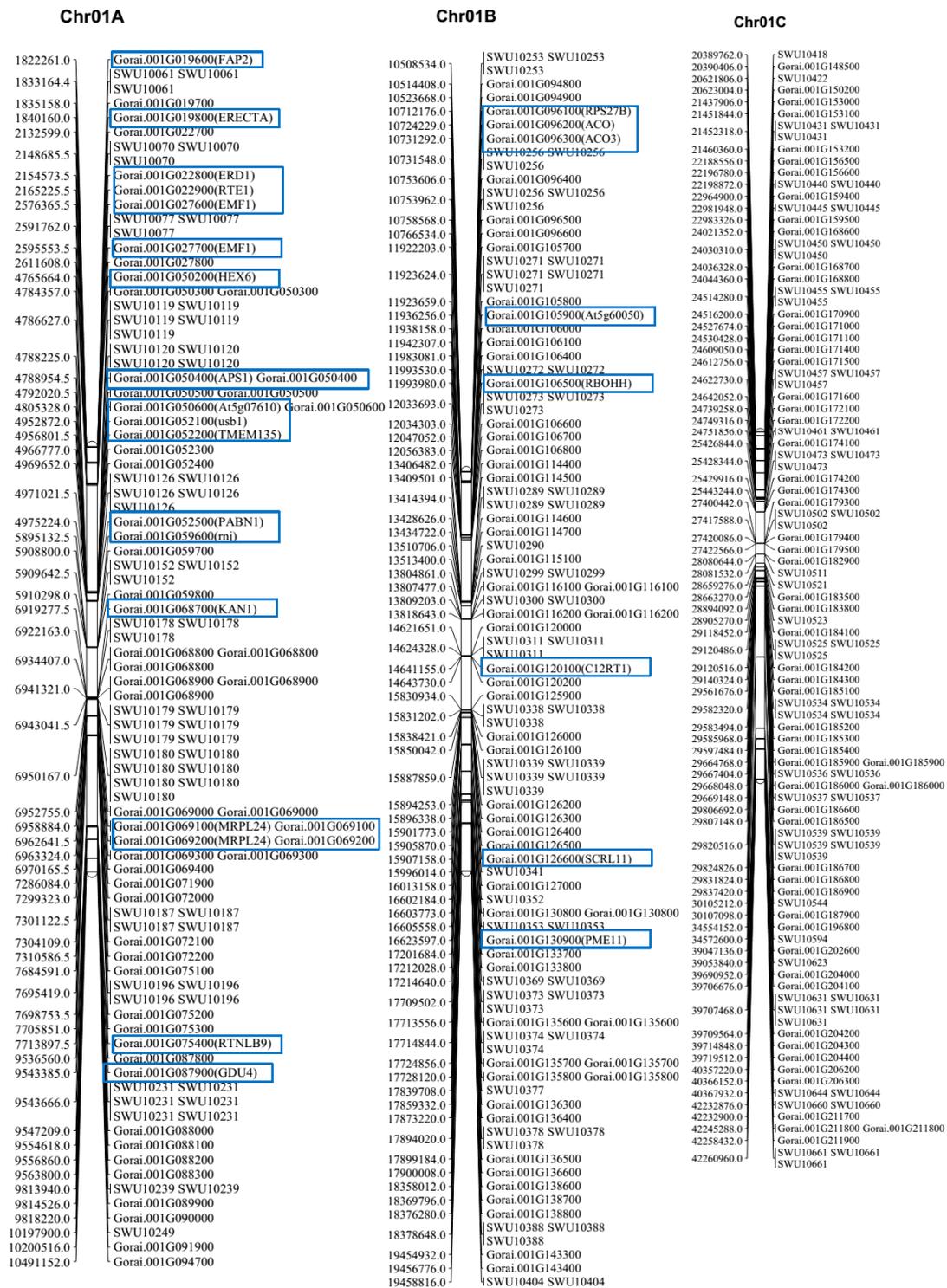
#### 2.5. Gene Mining, GO Functional Annotation, and Expression

A blast search of the regions 20 Kb up- and downstream of each SSR location was done and 2063 genes were obtained (Supplementary Table S2). The genes were distributed in all 13 chromosomes; the highest number of genes was found in chr09 with 269 genes, followed chr07 with 259 genes, while the least number of genes was detected in chr13 with only 58 genes. The rest of the chromosomes contained genes ranging from 100 to 200 in number (Supplementary Table S2). In the analysis

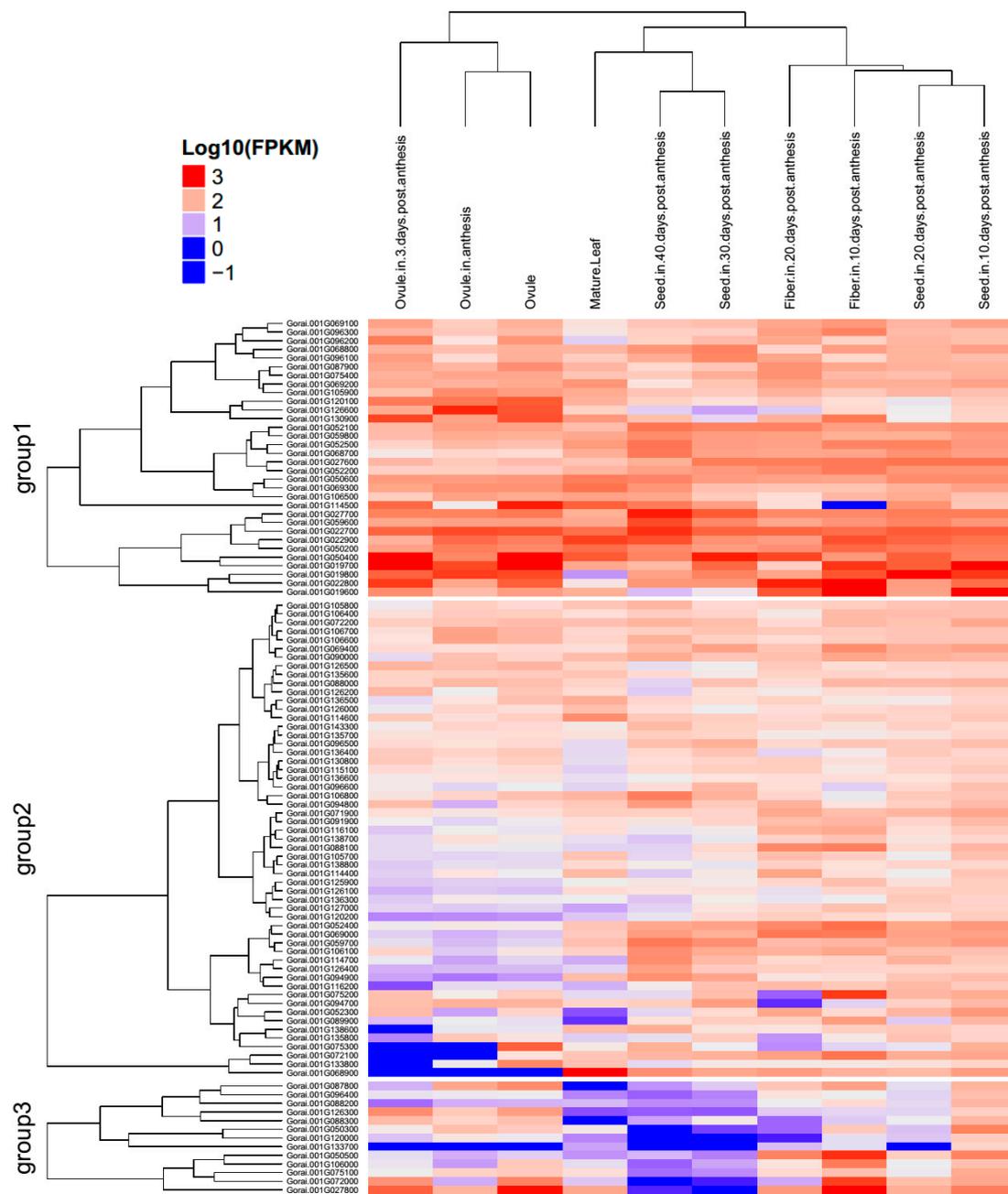
of the physiochemical properties of the mined genes, their molecular weights ranged from 4.272 to 318.34 kDa, their charge ranged from  $-159.5$  to  $68.5$ , the Isoelectric Point (PI) values ranged from 3.562 to 12.546, and the Grand Average of Hydropathy (GRAVY) values ranged from  $-1.982$  to  $1.148$ , which implies that the genes were both hydrophobic and hydrophilic in nature. In addition, we analysed the gene structures in order to determine the intron–exon interactions. Out of 2063 genes, only 274 were not disrupted, which implies that these genes were highly conserved. The highest intron disruption was detected in Gorai.009G087200, with 49 introns (Supplementary Table S3); similarly high levels of intron disruption have also been recorded for cyclin dependent kinase (CDK) genes [34]. We further analysed gene features in order to determine the various descriptions of the mined genes; several classes of genes of interest observed were stress- and fibre-related genes. The stress-related genes detected were: heat shock protein, abscisic acid receptor, aluminum-activated malate transporter, calcium-dependent protein kinase, E3 ubiquitin-protein ligase, ethylene-responsive transcription factor, expansin, Myb-like protein, and mitogen-activated protein kinase, among others. In addition, we carried out gene ontology (GO) annotation. The GO terms describe the genes in relation to cellular components (CC), molecular functions (MF), and biological processes. All GO terms were detected. In cellular components, functions such as cell, cell part, extracellular regions, and membrane enclosed lumens, among others, were detected for the genes mined. Similarly, 13 functions were observed under molecular function, while the highest numbers of genes were found to be involved in biological processes (Figure 6). Finally, we carried out RNA seq. expression in order to validate the functions annotated from gene ontology. The RNA sequences were downloaded from the cotton genome database (<https://cottonfgd.org/search/>). The genes showed differential expression: more than 70% of the genes were found to be up regulated at fibre developmental stages at 10 and 20 DPA; in seed development at 10, 20, 30, and 40 DPA; in mature leaves, in ovules in anthesis; and 3 days post-anthesis (Supplementary Table S4). We selected the top 100 highly up regulated genes in the various tissues and constructed a heatmap based on their respective expression to levels ( $\log_{10}$ ). The genes were categorized into three distinct groups. Group 1, had 32 genes that were significantly highly expressed, with expression levels of more than one. Among the 32 highly up regulated genes, *FAP2* (Fatty-acid-binding protein 2), with two GO functions: intramolecular lyase activity (GO:0016872-MF) and a cellular modified amino acid biosynthetic process (GO:0042398-CC); *ERECTA* (LRR receptor-like serine/threonine-protein kinase ERECTA) with seven GO functions such protein kinase activity (GO:0004672-MF), protein binding (GO:0005515-MF), ATP binding (GO:0005524-MF), protein phosphorylation (GO:0006468-MF), protein serine/threonine kinase activity (GO:0004674-MF), protein tyrosine kinase activity (GO:0004713-MF), and transferase activity, transferring phosphorus-containing groups (GO:0016772-MF); *EDR1* (Serine/threonine-protein kinase EDR1) with nine GO functions; *HEX6* (Hexose carrier protein HEX6) with one GO function; *RBOHH* (Putative respiratory burst oxidase homolog protein H) with 10 GO functions; *C12RT1* (Flavanone 7-O-glucoside-2''-O-beta-L-rhamnosyltransferase) with two GO functions; *SCRL11* (Putative defensin-like protein 244) with one GO function; and *PME11* (Putative pectinesterase 11) with three GO functions, could be the key genes responsible for fibre quality and other stress factor tolerance (Figure 7 and Supplementary Table S5). The second group, with 13 genes, was relatively down regulated in various tissues examined, while the third group, which was the majority with 55 genes, exhibited differential expression of both up- and down regulation. The expression profile is as illustrated in (Figure 8). As cotton is an important crop for fibre, seed development is critical for high fibre quality. The RNA seq. data provides clear indication that the genes mined within the SSR marker sequences have potential roles in both biotic stress and enhanced fibre quality.



**Figure 6.** Gene ontology (GO) annotation results of the mined genes from the SSR derived genetic map of the D genome. GO analysis of 2063 protein sequences predicted their involvement in biological processes (BP), molecular functions (MF), and cellular components (CC).



**Figure 7.** Physical map showing the exact locations of the identified key genes. The gene positions are in base pairs (bp). “SWU” are the SSR markers while the “Gorai” are the gene identities. The blue boxes indicate the key genes.



**Figure 8.** RNA sequence data analysis of 100 highly up regulated genes as per the RNA sequence. The heatmap was generated by  $\log_{10}$  of the expression values. Colour coding indicates expression as illustrated in the key. Y-axis is the relative expression ( $2^{-\Delta\Delta Ct}$ ).

### 3. Discussion

Simple sequence repeat (SSR) markers have gained significant use in plant genetics and breeding because of their multi-allelic nature, reproducibility, relative abundance, codominant inheritance, and good genome coverage [35]. SSRs are developed from genomic libraries, and can belong to either the transcribed region or the non-transcribed region of the genome [36,37].

In the present study, we explored the expressed sequence tag simple sequence repeat (EST-SSR) primers to access the nature of the  $F_{2:3}$  populations, which we developed from an interspecific cross between two wild cotton species of the D genome: *G. Klotzschianum* and *G. davidsonii*. Out of 12,500 SWU primers screened, only 1000 primers were found to be polymorphic, accounting for less than 10%

of all the primers used. The result showed that these markers had relatively low levels of polymorphism in the  $F_{2:3}$  populations used. The low level of polymorphism in these markers indicates that EST-SSR primers are less polymorphic. The results obtained are in agreement with previous findings in which 54% and 83.8% polymorphism levels were detected in EST-SSRs and gSSRs in rice, respectively [38]. Similar results have also been obtained in wheat, in which 53% of gSSRs were found to be polymorphic compared to 25% of EST-SSRs [39]. Despite low polymorphism rates among the EST-SSRs, significant deviation has been observed in a high percentage of eSSRs, showing a relatively high polymorphism rate; for example, a 93.5% polymorphism rate was observed in an intraspecific cross between two genotypes of diploid *Actinidia chinensis* [40]. High polymorphism rates of 66% in EST-SSRs have also been found between the parents of rye grass and tall fescue populations [41]. In cotton, higher polymorphism rates have been recorded in *G. hirsutum* and *G. barbadense* gSSR markers (49%, 56%) than in EST-SSRs (18.2%, 19.8%, 23.3%, 23.9%, and 26%) [35,42–44]. Despite the low polymorphism in EST-SSR markers, EST-SSR markers still remain the markers of choice in plant genome studies due to their close association with genes of known function, high levels of transferability, codominant inheritance, and low cost [45,46].

In allotetraploid cotton map constructions a high number of dominant marker loci ranging from 29% to 67.8% have been reported [25]. A total of 106 dominant restriction fragment length polymorphism (RFLP) loci, accounting for 38.5% of an interspecific A diploid map, have been reported [25]. In the present study of the D genome interspecific map construction, 280 (28%) markers were scored as dominant; among them 200 (27.17%) were mapped. The high number of dominant loci in diploid cotton could be explained by the presence of non-functional copies of genes as a result of genetic mutation. These copies of genes are referred to as null alleles [47]. The null alleles have caused perennial challenges in population genetics as a result of the introduction and application of each new techniques of molecular assay such as microsatellites, protein electrophoresis, serological typing, and RFLPs [48]. The A, B, and O blood grouping in humans is the best example of a locus with a null allele: in people with the blood groups AO and BO, the O allele is a null allele that produces no phenotype because it is masked by the presence of the A or B alleles, which are co-dominant to each other. In plants or animals, a microsatellite null allele is a type of allele which consistently fails to amplify during a polymerase chain reaction (PCR), and therefore is not detected in the process of genotyping individuals or populations [48]. A number of studies have recorded the detection of null alleles: for example, in bread wheat a 45% rate of polymorphism was recorded due to presence of null alleles [49]. Similarly, 10% of eSSR primers detected 7–14 null alleles in bread wheat [50].

There are three major explanations for the occurrence of null alleles. The first is due to mutation—either deletion or substitution in the primer site [51]—which causes poor primer annealing during the PCR amplification process. Another possible cause for the occurrence of null alleles is the differential amplification of size-variant alleles [52]: short length alleles amplify more efficiently than larger ones, such that only the smaller of two alleles might be detected in a heterozygous individual. The third option could be due to PCR failure as a result of inconsistent DNA template quality or low template quantity.

Non-normalized segregation ratios have been observed in cotton [25]. In some cases, the distortion percentage is as high as 80%. This is similar to what was observed in this study, in which the highest segregation distortion was 76.087% in Chr02. Similar results have been observed in previous publications, and this high frequency of segregation distortion in interspecific crosses is believed to occur as a result of species divergence [53]. The map generated in this study is more detailed than other maps which have been generated from the D genome. The map was developed from mono-markers and the map features are coherent to other maps produced from high quality methods such as the use of RLFPs. High levels of segregation distortion (SD) are not only limited to maps developed from interspecific crosses; even in intraspecific crosses, the frequency of distorted ratios is also relatively high, ranging from 44.1% to 52.49% [54].

SD is a problem often encountered in mapping populations. In this work, the map generated exhibited the lowest observed SDs in the following linkage groups; LG01 (3.371%), LG04 (3.571%), LG10 (5.882%), LG11 (7.937%), and LG06 (8.621%); medium SDs were detected in LG13 (10.526%), LG12 (12.245), and LG09 (16.216); while the following LDs showed highest level of SDs; LG05 (34.694%), LG07 (40.698%), and LG02 (76.087%). A number of variables could have led to this kind of distortion, such as genetic drift, or cytological attributes such as pollen tube competition, mutation of the SSR binding site, and redundant heterozygotes, which are possible causes of segregation distortion. Pollen death, hitchhiking, gametophyte selection, preferential fertilization, and zygotic selection all result in segregation distortion [55]. Seven hundred and twenty-eight (728) of the mapped markers in the genetic map were syntenic at the chromosomal level with their corresponding chromosomes in the D genome. Marker duplication is a common feature in most maps of diploid cotton. The two types of duplication are inter- and intra-chromosomal duplication. In this study, 94 markers were found to be duplicated. Out of the duplicated markers, the majority showed inter-chromosomal duplication as opposed to intra-chromosomal duplication. The results obtained correlate positively with a number of reports in which markers have been found to be duplicated in diploid cotton with inter-chromosomal duplication constituting the majority [25]. Similar findings have been reported in other plants; for instance in maize, only one intra-chromosomal duplication has been reported on chromosome 8 [56]. Helentjaris et al. [57] reported there being little evidence supporting the existence of extensive intra-chromosomal duplication in maize. More duplicated loci were shared between the linkage group pairs LG03-LG13 (4), LG06-LG08 (3), LG07-LG08 (5), LG09-LG12 (7), and LG10-LG13 (2). Also, LG07 shared duplicated loci with LG01, LG02, LG06, LG08, LG09, LG11, and LG12. Twelve intra-chromosomal duplications were observed in LG01, LG02, LG03, LG04, LG05, LG06, LG07, LG08, LG09, LG12, and LG13. The two forms of duplication, inter- and intra-chromosomal, have been explained by a number of mechanisms, and therefore the occurrence of conserved syntenic blocks common between two chromosomes of a diploid genome or two chromosomes of a polyploid genome is evidence of a paleopolyploid [58]. Therefore, in this study, a locus shared by two or more linkage groups was anticipated. The diploid genomes of cotton are paleopolyploids as proven by evidence through cytogenetic, biochemical, and genetic mapping [58].

Cotton production has been on the decline due to effects of biotic and abiotic stresses, which have been aggravated due to narrow genetic base of elite tetraploid cotton [59]. To solve this, the important agronomic traits of wild progenitors can be introgressed into the cultivated cotton cultivars [4]. Therefore, based on the genetic map developed from this study, a total of 2063 genes were mined in reference to *G. raimondii*. Of significance were the genes responsible for fibre development, and abiotic and biotic stress tolerance. In relation to stress factors, various drought-related stress genes were found, including eight NAC gene members: *NAC007* (NAC domain-containing protein 7), *NAC073* (NAC domain-containing protein 73), *NAC083* (NAC domain-containing protein 83), *NAC086* (NAC domain-containing protein 86), *NAC091* (NAC domain-containing protein 91), *NAM-B2* (NAC transcription factor *NAM-B2*), and *ONAC010* (NAC transcription factor *ONAC010*). The plant-specific NAC family has been shown to regulate several biological processes in wheat. NAC TFs are known to be involved in processes such as senescence and nutrient remobilization [60], as well as responses to abiotic and biotic stresses, ranging from biotic stresses such as stripe rust [61] to abiotic stresses including drought and salt [62,63]. The mitogen-activated protein kinase (MAPK) signaling cascades have been reported to play a significant role in plant environment adaptation, growth, and development [64]. Four members of MAPK genes were mined, such as *MPK4* (Mitogen-activated protein kinase 4), *MMK1* (Mitogen-activated protein kinase homolog *MMK1*), *NPK1* (Mitogen-activated protein kinase *NPK1*), and *YDA* (Mitogen-activated protein kinase *YODA*). The MAPKs have been linked to drought, cold, and salt stress signal-associated cascades, such as *AtMEKK1-AtMKK1/AtMKK2-AtMPK4* [65], and *MEKK1-MKK4/5-MPK3/6-WRKY22/WRKY29* which is actively involved in plant innate immunity [66], *AtMKK2-AtMPK10* MAPK has been found to be involved in the regulation of venation complexity by altering polar auxin transport efficiency [67].

Eighteen members of the serine/threonine-protein kinases were also found; these proteins have been found to have significant roles in plant cell growth, development, and oncogenesis [68]. Plant survival in ever-increasing environmental degradation has been associated with vital proteins known as ubiquitin carboxyl-terminal hydrolases. These proteins detect destructive molecules within the cell and either break or form compounds which are non-toxic to the plant. We identified three members of these proteins: *coq6* (Ubiquinone biosynthesis monooxygenase *coq6*, mitochondrial), *UBP15* (Ubiquitin carboxyl-terminal hydrolase 15), and *UBP8* (Ubiquitin carboxyl-terminal hydrolase 8). This group of proteins has been found to confer tolerance to drought and salt stress in *Brassica napus* [69].

In the RNA sequence expression of the first 100 highly upregulated genes, 32 genes were found to be significantly upregulated in the various tissues, for instance *FAP2* (Fatty-acid-binding protein 2) which is linked to heat stress tolerance in plants [70]. *ERECTA* (LRR receptor-like serine/threonine-protein kinase *ERECTA*) has been found to have roles in signaling and plant defense [71]. *FKBP42* (Peptidyl-prolyl cis-trans isomerase *FKBP42*) has been found to be induced by both abiotic and biotic stresses in maize [72], and by salt stress, heat shock, and cold shock in *Solanum tuberosum* [73]. *RBOHH* (Putative respiratory burst oxidase homolog protein H) has been found to have a crucial role in plant hormone signaling, development, and defense reactions [74]. The identification of these genes from the map constructed provides a platform for further use of the identified genes in improving the performance of elite tetraploid cotton. The D genome of cotton, has been found to be associated with various stable QTLs for physiological, morphological, and biochemical traits related to fibre qualities, and abiotic and biotic stress factors, when compared to the A genome as evident in the cotton QTL database (<http://www2.cottonqtl.org:8081/>).

## 4. Materials and Methods

### 4.1. Plant Materials

*G. klotzschianum* (accession number PI 499748) as the female parent was crossed with *G. davidsonii* (accession number PI 530809) as the male parent to establish an F<sub>1</sub> population. The F<sub>1</sub> was then selfed to generate F<sub>2,3</sub> progenies. *Gossypium davidsonii* has economically significant traits such as resistance to aphids, salinity, and bacterial blight [18], while *Gossypium klotzschianum* has traits for resistance to sucking pests [19]. The two parental materials and the F<sub>2,3</sub> lines were developed at the National Wild Cotton Nursery in Sanya, Hainan Island, China.

### 4.2. DNA Extraction, Quantification, and Qualification

The CTAB method for DNA extraction was used for the extraction of DNA for the entire population and the two parental lines. [26]. Fresh and young leaves were collected from each line, then immediately frozen in liquid nitrogen to prevent DNA degradation before extraction. Each sample was then ground in liquid nitrogen into a fine powder, and CTAB mixture was then added immediately. For each 100 mg of the homogenized tissue, 500 µL of CTAB Extraction Buffer was used. Upon addition of the buffer, the mixture was thoroughly vortexed. The homogenized mixture was then incubated in a 60 °C bath for 30 min. Following the incubation period, the mixture was centrifuged to homogenate for 5 min at 12,000 rpm. After centrifuging, the supernatant was transferred to a new tube. An equal volume of chloroform/isoamyl alcohol (24:1) was added then vortexed for 5 s. The sample was centrifuged for 1 min at 12,000 rpm to separate the phases. The upper phase was then pipetted into a new tube; the method was then repeated until the upper phase was clear. The upper clear aqueous phase was then transferred to a new tube. DNA was precipitated by adding 70% by volume ice-cold isopropanol and then incubated at –20 °C for 15 min. The precipitated DNA samples were then centrifuged at 12,000 × *g* for 10 min. The supernatant was then extracted without disturbing the pellet and subsequently washed with 500 µL of ice cold 70% by volume ethanol two times, and then finally with absolute alcohol. Then DNA pellets were later dissolved in 20 µL TE buffer (10 mM Tris, pH 8, 1 mM EDTA) [75]. DNA degradation and contamination were checked by

using 1% agarose gels, while the purity of the DNA extracted was evaluated by the use of a Nano Photometer<sup>®</sup> spectrophotometer (IMPLEN, Westlake Village, CA, USA). Ratios of 260 nm and 280 nm were the guidelines to assess the purity of the DNA. The DNA samples with ratios of ~1.8 were selected as pure [76]. For the use of SSR and for the PCR process, specific concentrations of DNA are necessary; we therefore determined the concentration of the DNA samples by using a Qubit<sup>®</sup> DNA Assay Kit in a Qubit<sup>®</sup> 2.0 Fluorimeter (Life Technologies, Carlsbad, CA, USA). The Qubit<sup>®</sup> dsDNA HS (High Sensitivity) Assay Kits make DNA quantitation easy and accurate. The manufacturer's protocol was followed. Upon mixing the reagents, 1 to 20  $\mu\text{L}$  was added to each DNA sample. The concentrations were read using the Qubit<sup>®</sup> Fluorometer. Only the DNA samples with concentration ranges of 10  $\text{pg}/\mu\text{L}$  to 100  $\text{ng}/\mu\text{L}$  were finally used [77].

#### 4.3. Screening of SWU Markers and Genotyping of $F_{2:3}$ Populations

The SWU markers used are expressed sequence tag-simple sequence repeat (EST-SSR) primers which were developed from *G. raimondii* by Southwest University, China, thus the acronym SWU. A total of 12,500 pairs of SSR primer were used to screen polymorphic loci between three parents, *G. davidsonii*, *Gossypium klotzschianum*, and their  $F_1$  generation [78]. In total, 1000 polymorphic loci were used to conduct genotype analysis of 188  $F_{2:3}$  populations. The polymerase chain reaction (PCR) profile was set as follows: denaturation at 94 °C for 2 min, 35 cycles of 30 s at 94 °C for denaturation, 30 s at 52 °C for annealing, 30 s at 72 °C for extension, and 5 min at 72 °C for final extension after the last cycle. The amplified PCR products were separated on 8% denaturing polyacrylamide gel and visualized by silver nitrate staining [79].

#### 4.4. Construction of the Linkage Maps

Linkage map analysis was conducted using Join Map 4.0 [69] with a recombination frequency of 0.40 and a LOD score of 2.5 for the  $F_{2:3}$  population. The Kosambi mapping function was used to convert the recombination frequencies to map distances. Linkage groups were assigned to chromosomes based on a marker sequence blast search, SWU markers being new sets of SSR markers and not previously annotated to any chromosome. The obtained linkage groups were drawn using Mapchart 2.2 Software [80] and R Software's mapping function [81].

#### 4.5. Collinearity Analysis

Based on the fine genetic linkage map constructed from the  $F_{2:3}$  population, the sequences corresponding to SWU markers were used to carry out collinearity analysis between genetic and physical maps of the D genome of *Gossypium raimondii*. A BLASTN Search with  $E \leq 1 \times 10^{-5}$ , identity  $\geq 80\%$ , and matched length  $\geq 200$  bp was applied (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The best hit for each marker was chosen and were illustrated intuitively using online drawing tools (<http://circos.ca/>).

#### 4.6. Gene Mining, GO Functional Annotation, and Expression Analysis

Regions 20 kb up- and downstream of each SSR location were screened in order to mine the genes from the physical map by the use of full sequences of the SWU markers used in construction of the genetic map. The 20 kb up- and downstream value was selected as the average distance of markers in the genetic map was 2.182 cM, and a similar method has also been adopted in mining several genes; for example, in the identification of the candidate gene for anthracnose disease resistance in *Lupinus angustifolius*, the largest gap ever used was 3.5 cM [82]. The genes mined were analysed for their gene features, protein characteristics, GO, and RNA expression profile using the cotton functional genome database (<https://cottonfgd.org/search/>), adopting *G. raimondii* as the reference genome. The RNA expression data obtained were then analysed and a heatmap was constructed using the R statistical software package [81].

## 5. Conclusions

In conclusion, wild cotton species harbor many important agronomic traits which can be used to improve the current cultivated cotton cultivar. Plant germplasm resources originate from a number of historical genetic events as a response to environmental stresses and selection; therefore, wild progenitors are important reservoirs of natural genetic variation that can be exploited to increase the genetic base of the elite cultivars. Our map is the first map developed from wild progenitors of the D genome; it covers a total map size of 1480.23 cM, with 728 mono-markers. We further identified a total of 2063 genes from the SSR regions of the physical map. Of significance were the stress-related genes, which exhibited differential expression in various tissues. The genes such as *NAC*, *Myb*, and ubiquitin-like genes, have been highly associated with abiotic and biotic stress factors in plants, and thus the detection of these genes provides valuable information for future exploration. Twenty-seven (27) key genes were identified with diverse roles such as plant hormone signaling, development, and defense reactions. The use of the two wild species, *G. davidsonii* and *G. klotzschianum*, in the construction of the first fine genetic map of two wild species in the D genome will provide insight for further genetic analysis. This is the first genetic map developed from the population of F<sub>2:3</sub>. The map covers a greater percentage of the whole D genome and therefore it forms a valuable tool for breeders in detecting QTLs related to salt stress and biotic stress tolerance, especially aphids, sucking pests, and the secondary metabolic product gossypol.

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/1422-0067/19/1/204/s1](http://www.mdpi.com/1422-0067/19/1/204/s1).

**Acknowledgments:** We are sincerely grateful to Liu Fang (Cotton Research Institute) for his valuable support, advice, and suggestions during the course of this research. To Wang Kunbo (Cotton research institute), thanks so much for the provision of the materials we needed during the research work and guidance. To the entire research team, friends, and any other person who contributed, we are indebted to you so much. This research was funded by the National Key Research and Development Plan (2016YFD0101401, 2016YFD0100203) and the National High Technology Research and Development Program (2013AA102601).

**Author Contributions:** Joy Nyangasi Kirungu, Yanfeng Deng, and Fang Liu conceived and designed the experiment; Joy Nyangasi Kirungu, Yanfeng Deng, and Richard Odongo Magwanga collected the data; Joy Nyangasi Kirungu and Richard Odongo Magwanga analysed the data and interpreted the results; Joy Nyangasi Kirungu and Richard Odongo Magwanga developed the manuscript; Fang Liu, Xiaoyan Cai, Zhongli Zhou, Kunbo Wang, and Fang Liu supervised the work and proofread the manuscript. All authors approved the manuscript.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## Abbreviations

LG	linkage group
EST-SSR	Express sequence tag-simple sequence repeat
gSSR	Genomic simple sequence repeat
GO	Gene ontology
BP	Biological process
MF	Molecular function
CC	Cellular components
GRAVY	Grand average of hydropathy
SWU	South west university
MPK	Mitogen-activated protein kinase

## References

1. Geber, M.A.; Griffen, L.R. Inheritance and Natural Selection on Functional Traits. *Int. J. Plant Sci.* **2003**, *164*, S21–S42. [[CrossRef](#)]
2. Huang, L.; Jiang, H.; Ren, X.; Chen, Y.; Xiao, Y.; Zhao, X.; Tang, M.; Huang, J.; Upadhyaya, H.D.; Liao, B. Abundant Microsatellite Diversity and Oil Content in Wild Arachis Species. *PLoS ONE* **2012**, *7*. [[CrossRef](#)] [[PubMed](#)]
3. Applequist, W.L.; Cronn, R.; Wendel, J.F. Comparative development of fiber in wild and cultivated cotton. *Evol. Dev.* **2001**, *3*, 3–17. [[CrossRef](#)] [[PubMed](#)]

4. Hulse-Kemp, A.M.; Ashrafi, H.; Zheng, X.; Wang, F.; Hoegenauer, K.A.; Maeda, A.B.V.; Yang, S.S.; Stoffel, K.; Matvienko, M.; Clemons, K.; et al. Development and bin mapping of gene-associated interspecific SNPs for cotton (*Gossypium hirsutum* L.) introgression breeding efforts. *BMC Genom.* **2014**, *15*, 945. [[CrossRef](#)] [[PubMed](#)]
5. Bengtsson, E. *Obtaining High Quality Textile Fibre from Industrial Hemp through Organic Cultivation by Elin Bengtsson*; SLU: Alnarp, Sweden, 2009.
6. Resta, B.; Gaiardelli, P.; Pinto, R.; Dotti, S. Enhancing environmental management in the textile sector: An Organisational-Life Cycle Assessment approach. *J. Clean. Prod.* **2016**, *135*, 620–632. [[CrossRef](#)]
7. Zhou, M.; Sun, G.; Sun, Z.; Tang, Y.; Wu, Y. Cotton proteomics for deciphering the mechanism of environment stress response and fiber development. *J. Proteom.* **2014**, *105*, 74–84. [[CrossRef](#)] [[PubMed](#)]
8. Diouf, L.; Pan, Z.; He, S.-P.; Gong, W.-F.; Jia, Y.H.; Magwanga, R.O.; Romy, K.R.E.; Rashid, H.; Kirungu, J.N.; Du, X. High-Density Linkage Map Construction and Mapping of Salt-Tolerant QTLs at Seedling Stage in Upland Cotton Using Genotyping by Sequencing (GBS). *Int. J. Mol. Sci.* **2017**, *18*, 2622. [[CrossRef](#)] [[PubMed](#)]
9. Islam, M.S.; Thyssen, G.N.; Jenkins, J.N.; Fang, D.D. Detection, Validation, and Application of Genotyping-by-Sequencing Based Single Nucleotide Polymorphisms in Upland Cotton. *Plant Genome* **2015**, *8*. [[CrossRef](#)]
10. Wegier, A.; Piñeyro-Nelson, A.; Alarcón, J.; Gálvez-Mariscal, A.; Álvarez-Buylla, E.R.; Piñero, D. Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (*Gossypium hirsutum*) at its centre of origin. *Mol. Ecol.* **2011**, *20*, 4182–4194. [[CrossRef](#)] [[PubMed](#)]
11. Chakravarthy, V.S.K.; Reddy, T.P.; Reddy, V.D.; Rao, K.V. Current status of genetic engineering in cotton (*Gossypium hirsutum* L.): An assessment. *Crit. Rev. Biotechnol.* **2012**, *8551*, 1–18. [[CrossRef](#)] [[PubMed](#)]
12. Tuberosa, R.; Forster, B.P.; Ellis, R.P.; Thomas, W.T.; Newton, A.C.; This, D.; El-Enein, R.A.; Bahri, M.H.; Ben Salem, M. The development and application of molecular markers for abiotic stress tolerance in barley. *J. Exp. Bot.* **2000**, *51*, 19–27. [[CrossRef](#)]
13. Khoury, C.K.; Castañeda-Alvarez, N.P.; Achicanoy, H.A.; Sosa, C.C.; Bernau, V.; Kassa, M.T.; Norton, S.L.; van der Maesen, L.J.G.; Upadhyaya, H.D.; Ramírez-Villegas, J.; et al. Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: Distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* **2015**, *184*, 259–270. [[CrossRef](#)]
14. Becerra Lopez-Lavalle, L.A.; Matheson, B.; Brubaker, C.L.; Van Deynze, A. A genetic map of an Australian wild *Gossypium* C genome and assignment of homoeologies with tetraploid cultivated cotton. *Genome* **2011**, *54*, 779–794. [[CrossRef](#)] [[PubMed](#)]
15. Rungis, D.; Llewellyn, D.; Dennis, E.S.; Lyon, B.R. Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Aust. J. Agric. Res.* **2005**, *56*, 301–307. [[CrossRef](#)]
16. Abdurakhmonov, I.Y.; Kohel, R.J.; Yu, J.Z.; Pepper, A.E.; Abdullaev, A.A.; Kushanov, F.N.; Salakhutdinov, I.B.; Buriev, Z.T.; Saha, S.; Scheffler, B.E.; et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* **2008**, *92*, 478–487. [[CrossRef](#)] [[PubMed](#)]
17. Kuester, A.P.; Nason, J.D. Microsatellite loci for *Gossypium davidsonii* (Malvaceae) and other D-Genome, Sonoran Desert endemic cotton species. *Am. J. Bot.* **2012**, *99*. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, F.; Zhu, G.; Du, L.; Shang, X.; Cheng, C.; Yang, B.; Hu, Y.; Cai, C.; Guo, W. Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, *Gossypium davidsonii*. *Sci. Rep.* **2016**, *6*, 20582. [[CrossRef](#)] [[PubMed](#)]
19. Wei, Y.; Xu, Y.; Lu, P.; Wang, X.; Li, Z.; Cai, X.; Zhou, Z.; Wang, Y.; Zhang, Z.; Lin, Z.; et al. Salt stress responsiveness of a wild cotton species (*Gossypium klotzschianum*) based on transcriptomic analysis. *PLoS ONE* **2017**, *12*. [[CrossRef](#)] [[PubMed](#)]
20. Han, Z.G.; Guo, W.Z.; Song, X.L.; Zhang, T.Z. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Mol. Genet. Genom.* **2004**, *272*, 308–327. [[CrossRef](#)] [[PubMed](#)]
21. Jiang, C.; Chee, P.W.; Draye, X.; Morrell, P.L.; Smith, C.W.; Paterson, A.H. Multilocus Interactions Restrict Gene Introgression in Interspecific Populations of Polyploid *Gossypium* (Cotton). *Evolution (N. Y.)* **2000**, *54*, 798–814. [[CrossRef](#)]

22. Zhang, Z.S.; Hu, M.C.; Zhang, J.; Liu, D.J.; Zheng, J.; Zhang, K.; Wang, W.; Wan, Q. Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol. Breed.* **2009**, *24*, 49–61. [[CrossRef](#)]
23. Jiang, C.; Wright, R.J.; El-Zik, K.M.; Paterson, A.H. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4419–4424. [[CrossRef](#)] [[PubMed](#)]
24. Reinisch, A.J.; Dong, J.M.; Brubaker, C.L.; Stelly, D.M.; Wendel, J.F.; Paterson, A.H. A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. *Genetics* **1994**, *138*, 829–847. [[PubMed](#)]
25. Desai, A.; Chee, P.W.; Rong, J.; May, O.L.; Paterson, A.H. Chromosome structural changes in diploid and tetraploid A genomes of *Gossypium*. *Genome* **2006**, *49*, 336–345. [[CrossRef](#)] [[PubMed](#)]
26. Cota-Sánchez, J.H.; Remarchuk, K.; Ubayasena, K. Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant Mol. Biol. Rep.* **2006**, *24*, 161–167. [[CrossRef](#)]
27. Rong, J.; Abbey, C.; Bowers, J.E.; Brubaker, C.L.; Chang, C.; Chee, P.W.; Delmonte, T.A.; Ding, X.; Garza, J.J.; Marler, B.S.; et al. A 3347-Locus Genetic Recombination Map of Sequence-Tagged Sites Reveals Features of Genome Organization, Transmission and Evolution of Cotton (*Gossypium*). *Genetics* **2004**, *166*, 389–417. [[CrossRef](#)] [[PubMed](#)]
28. Brubaker, C.L.; Brown, A.H. The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium* G genome. *Genome* **2003**, *46*, 774–791. [[CrossRef](#)] [[PubMed](#)]
29. Collard, B.C.Y.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B* **2008**, *363*, 557–572. [[CrossRef](#)] [[PubMed](#)]
30. He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.; Li, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*. [[CrossRef](#)] [[PubMed](#)]
31. Yu, H.; Xie, W.; Li, J.; Zhou, F.; Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* **2014**, *12*, 28–37. [[CrossRef](#)] [[PubMed](#)]
32. Li, H.; Chen, X.; Yang, Y.; Xu, J.; Gu, J.; Fu, J.; Qian, X.; Zhang, S.; Wu, J.; Liu, K. Development and genetic mapping of microsatellite markers from whole genome shotgun sequences in *Brassica oleracea*. *Mol. Breed.* **2011**, *28*, 585–596. [[CrossRef](#)]
33. Ma, X.X.; Zhou, B.L.; Lü, Y.H.; Guo, W.Z.; Zhang, T.Z. Simple sequence repeat genetic linkage maps of A-genome diploid cotton (*Gossypium arboreum*). *J. Integr. Plant Biol.* **2008**, *50*, 491–502. [[CrossRef](#)] [[PubMed](#)]
34. Wang, G. Genome-Wide Analysis of the Cyclin Family in *Arabidopsis* and Comparative Phylogenetic Analysis of Plant Cyclin-Like Proteins. *Plant Physiol.* **2004**, *135*, 1084–1099. [[CrossRef](#)] [[PubMed](#)]
35. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A.K. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2011**, *177*, 309–334. [[CrossRef](#)]
36. Gupta, P.K.; Rustgi, S. Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genom.* **2004**, *4*, 139–162. [[CrossRef](#)] [[PubMed](#)]
37. Luro, F.L.; Costantino, G.; Terol, J.; Argout, X.; Allario, T.; Wincker, P.; Talon, M.; Ollitrault, P.; Morillon, R. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genom.* **2008**, *9*, 287. [[CrossRef](#)] [[PubMed](#)]
38. Cho, Y.G.; Ishii, T.; Temnykh, S.; Chen, X.; Lipovich, L.; McCouch, S.R.; Park, W.D.; Ayres, N.; Cartinhour, S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **2000**, *100*, 713–722. [[CrossRef](#)]
39. Eujayl, I.; Sorrells, M.; Baum, M.; Wolters, P.; Powell, W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. In *Euphytica*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Volume 119, pp. 39–43.
40. Fraser, L.G.; Harvey, C.F.; Crowhurst, R.N.; De Silva, H.N. EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor. Appl. Genet.* **2004**, *108*, 1010–1016. [[CrossRef](#)] [[PubMed](#)]
41. Saha, M.C.; Mian, M.A.R.; Eujayl, I.; Zwonitzer, J.C.; Wang, L.; May, G.D. Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* **2004**, *109*, 783–791. [[CrossRef](#)] [[PubMed](#)]
42. Nguyen, T.B.; Giband, M.; Brottier, P.; Risterucci, A.M.; Lacape, J.M. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* **2004**, *109*, 167–175. [[CrossRef](#)] [[PubMed](#)]

43. Jenkins, J.N.; Brooks, T.; Bolek, Y.; El-zik, K.M. New Dinucleotide and Trinucleotide Microsatellite Marker Resources for Cotton Genome Research. *J. Cotton Sci.* **2001**, *5*, 103–113.
44. Bowers, J.E.; Abbey, C.; Anderson, S.; Chang, C.; Draye, X.; Hoppe, A.H.; Jessup, R.; Lemke, C.; Lenington, J.; Li, Z.; et al. A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* **2003**, *165*, 367–386. [[PubMed](#)]
45. Li, M.; Zhu, L.; Zhou, C.Y.; Lin, L.; Fan, Y.J.; Zhuang, Z.M. Development and characterization of EST-SSR markers from *Scapharca broughtonii* and their transferability in *Scapharca subcrenata* and *Tegillarca granosa*. *Molecules* **2012**, *17*, 10716–10723. [[CrossRef](#)] [[PubMed](#)]
46. Varshney, R.K.; Graner, A.; Sorrells, M.E. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol.* **2005**, *23*, 48–55. [[CrossRef](#)] [[PubMed](#)]
47. Wagner, A.P.; Creel, S.; Kalinowski, S.T. Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity* **2006**, *97*, 336–345. [[CrossRef](#)] [[PubMed](#)]
48. Dakin, E.E.; Avise, J.C. Microsatellite null alleles in parentage analysis. *Heredity* **2004**, *93*, 504–509. [[CrossRef](#)] [[PubMed](#)]
49. Gupta, P.K.; Rustgi, S.; Sharma, S.; Singh, R.; Kumar, N.; Balyan, H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genom.* **2003**, *270*, 315–323. [[CrossRef](#)] [[PubMed](#)]
50. Prasad, M.; Varshney, R.K.; Balyan, H.S.; Gupta, P.K.; Roy, J.K. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. *TAG Theor. Appl. Genet.* **2000**, *100*, 584–592. [[CrossRef](#)]
51. Callen, D.F.; Thompson, A.D.; Shen, Y.; Phillips, H.A.; Richards, R.I.; Mulley, J.C.; Sutherland, G.R. Incidence and origin of “null” alleles in the (AC)<sub>n</sub> microsatellite markers. *Am. J. Hum. Genet.* **1993**, *52*, 922–927. [[PubMed](#)]
52. La Rosa, R.; Angiolillo, A.; Guerrero, C.; Pellegrini, M.; Rallo, L.; Besnard, G.; Bervillé, A.; Martin, A.; Baldoni, L. A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. *Theor. Appl. Genet.* **2003**, *106*, 1273–1282. [[CrossRef](#)] [[PubMed](#)]
53. Paterson, A.H.; Lander, E.S.; Hewitt, J.D.; Peterson, S.; Lincoln, S.E.; Tanksley, S.D. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **1988**, *335*, 721–726. [[CrossRef](#)] [[PubMed](#)]
54. Shen, X.; Guo, W.; Zhu, X.; Yuan, Y.; Yu, J.Z.; Kohel, R.J.; Zhang, T. Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol. Breed.* **2005**, *15*, 169–181. [[CrossRef](#)]
55. Sibov, S.T.; Souza, C.L., Jr.; Garcia, A.A.F.; Silva, A.R.; Garcia, A.A.F.; Mangolin, C.A.; Benchimol, L.L.; Souza, A.P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. *Hereditas* **2003**, *139*, 107–115. [[CrossRef](#)] [[PubMed](#)]
56. Gaut, B.S. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Res.* **2001**, *11*, 55–66. [[CrossRef](#)] [[PubMed](#)]
57. Helentjaris, T.; Weber, D.; Wright, S. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* **1988**, *118*, 353–363. [[PubMed](#)]
58. Brubaker, C.L.; Paterson, A.H.; Wendel, J.F. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* **1999**, *42*, 184–203. [[CrossRef](#)]
59. Juturu, V.N.; Mekala, G.K.; Kirti, P.B. Current status of tissue culture and genetic transformation research in cotton (*Gossypium* spp.). *Plant Cell. Tissue Organ Cult.* **2015**, *120*, 813–839. [[CrossRef](#)]
60. Zhao, D.; Derkx, A.P.; Liu, D.C.; Buchner, P.; Hawkesford, M.J. Overexpression of a NAC transcription factor delays leaf senescence and increases grain nitrogen concentration in wheat. *Plant Biol.* **2015**, *17*, 904–913. [[CrossRef](#)] [[PubMed](#)]
61. Wang, F.; Lin, R.; Feng, J.; Chen, W.; Qiu, D.; Xu, S. TaNAC1 acts as a negative regulator of stripe rust resistance in wheat, enhances susceptibility to *Pseudomonas syringae*, and promotes lateral root development in transgenic *Arabidopsis thaliana*. *Front. Plant Sci.* **2015**, *6*. [[CrossRef](#)] [[PubMed](#)]
62. Xu, Z.; Gongbuzhaxi; Wang, C.; Xue, F.; Zhang, H.; Ji, W. Wheat NAC transcription factor TaNAC29 is involved in response to salt stress. *Plant Physiol. Biochem.* **2015**, *96*, 356–363. [[CrossRef](#)] [[PubMed](#)]

63. Huang, Q.; Wang, Y.; Li, B.; Chang, J.; Chen, M.; Li, K.; Yang, G.; He, G. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* **2015**, *15*. [[CrossRef](#)] [[PubMed](#)]
64. Xu, J.; Zhang, S. Mitogen-activated protein kinase cascades in signaling plant growth and development. *Trends Plant Sci.* **2015**, *20*, 56–64. [[CrossRef](#)] [[PubMed](#)]
65. Furuya, T.; Matsuoka, D.; Nanmori, T. Membrane rigidification functions upstream of the MEKK1-MKK2-MPK4 cascade during cold acclimation in Arabidopsis thaliana. *FEBS Lett.* **2014**, *588*, 2025–2030. [[CrossRef](#)] [[PubMed](#)]
66. Asai, T.; Tena, G.; Plotnikova, J.; Willmann, M.R.; Chiu, W.L.; Gomez-Gomez, L.; Boller, T.; Ausubel, F.M.; Sheen, J. Map kinase signalling cascade in Arabidopsis innate immunity. *Nature* **2002**, *415*, 977–983. [[CrossRef](#)] [[PubMed](#)]
67. Stanko, V.; Giuliani, C.; Retzer, K.; Djamei, A.; Wahl, V.; Wurzinger, B.; Wilson, C.; Heberle-Bors, E.; Teige, M.; Kragler, F. Timing is everything: Highly specific and transient expression of a map kinase determines auxin-induced leaf venation patterns in arabidopsis. *Mol. Plant* **2014**, *7*, 1637–1652. [[CrossRef](#)] [[PubMed](#)]
68. Rudrabhatla, P.; Reddy, M.M.; Rajasekharan, R. Genome-wide analysis and experimentation of plant serine/ threonine/tyrosine-specific protein kinases. *Plant Mol. Biol.* **2006**, *60*, 293–319. [[CrossRef](#)] [[PubMed](#)]
69. Chen, L.; Ren, F.; Zhong, H.; Jiang, W.; Li, X. Identification and expression analysis of genes in response to high-salinity and drought stresses in Brassica napus. **2010**, *42*, 154–164. [[CrossRef](#)] [[PubMed](#)]
70. Menard, G.; Martin Moreno, J.; Bryant, F.; Munoz-Azcarate, O.; Kelly, A.A.; Hassani-Pak, K.; Kurup, S.; Eastmond, P.J. Genome wide analysis of fatty acid desaturation and its response to temperature. *Plant Physiol.* **2017**, 1594–1605. [[CrossRef](#)] [[PubMed](#)]
71. Afzal, A.J.; Wood, A.J.; Lightfoot, D.A. Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. *Mol. Plant-Microbe Interact.* **2008**, *21*, 507–517. [[CrossRef](#)] [[PubMed](#)]
72. Marivet, J.; Frenedo, P.; Burkard, G. Effects of abiotic stresses on cyclophilin gene expression in maize and bean and sequence analysis of bean cyclophilin cDNA. *Plant Sci.* **1992**, *84*, 171–178. [[CrossRef](#)]
73. Godoy, A.V.; Lazzaro, A.S.; Casalongué, C.A.; San Segundo, B. Expression of a Solanum tuberosum cyclophilin gene is regulated by fungal infection and abiotic stress conditions. *Plant Sci.* **2000**, *152*, 123–134. [[CrossRef](#)]
74. Cheng, C.; Xu, X.; Gao, M.; Li, J.; Guo, C.; Song, J.; Wang, X. Genome-wide analysis of respiratory burst oxidase homologs in grape (*Vitis vinifera* L.). *Int. J. Mol. Sci.* **2013**, *14*, 24169–24186. [[CrossRef](#)] [[PubMed](#)]
75. Krizman, M.; Jakse, J.; Baricevic, D.; Javornik, B.; Prosek, M. Robust CTAB-activated charcoal protocol for plant DNA extraction. *Acta Agric. Slov.* **2006**, *87*, 427–433.
76. Wilfinger, W.W.; Mackey, K.; Chomczynski, P. 260/280 and 260/230 Ratios NanoDrop® ND-1000 and ND-8000 8-Sample Spectrophotometers. *BioTechniques* **1997**, *22*, 474–481. [[CrossRef](#)] [[PubMed](#)]
77. Carver, T.J.; Rutherford, K.M.; Berriman, M.; Rajandream, M.A.; Barrell, B.G.; Parkhill, J. ThermoFisher Scientific Qubit dsDNA HS Assay Kit. *Bioinformatics* **2016**, *21*, 3422–3423. [[CrossRef](#)] [[PubMed](#)]
78. Shang, L.; Wang, Y.; Cai, S.; Wang, X.; Li, Y.; Abduweli, A.; Hua, J. Partial Dominance, Overdominance, Epistasis and QTL by Environment Interactions Contribute to Heterosis in Two Upland Cotton Hybrids. *Genes Genomes Genet.* **2016**, *6*, 499–507. [[CrossRef](#)] [[PubMed](#)]
79. Sun, D.; Sun, J.; Jia, Y.; Ma, Z.; Du, X. Genetic Diversity of Colored Cotton Analyzed by Simple Sequence Repeat Markers. *Int. J. Plant Sci.* **2009**, *170*, 76–82. [[CrossRef](#)]
80. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [[CrossRef](#)] [[PubMed](#)]
81. Kuhnert, P.; Venables, W.N. *An Introduction to R: Software for Statistical Modelling & Computing*; Commonwealth Scientific and Industrial Research Organisation (CSIRO): Canberra, Australia, 2005; pp. 1–364.
82. You, M.; Boersma, J.G.; Buirchell, B.J.; Sweetingham, M.W.; Siddique, K.H.; Yang, H. A PCR-based molecular marker applicable for marker-assisted selection for anthracnose disease resistance in lupin breeding. *Cell. Mol. Biol. Lett.* **2005**, *10*, 123–134. [[PubMed](#)]

