Ditta A *et al.* (2024)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 52, Issue 3, Article number 13545
DOI:10.15835/nbha52313545
Research Article

AcademicPres

Notulae Botanicae Horti Agrobotanici Cluj-Napoca

Unveiling the genetic landscape: Exploring the SSR-based genetic architecture and amino acid dissection of *Gossypium barbadense* and *G. darwinii* genomes

Allah DITTA[1,2,3a], Zhongli ZHOU[1], Muhammad SHEHZAD[1b], Xiaoyan CAI[1], Yanchao XU[1], Yuqing HOU[1], Fang LIU[1,4]*, Kunbo WANG[1]*, Muhammad K.R. KHAN[2,3], Kotb A. ATTIA[5], Yaser M. HAFEZ[6], Arif A. MOHAMMED[5] Baohua WANG[7], Sajid FIAZ[8], Baohong ZHANG[9]

[1]*Chinese Academy of Agricultural Sciences, State Key Laboratory of Cotton Biology, Institute of Cotton Research, Anyang 455000, Henan, China; adbotanist@yahoo.com (A.D.); zhonglizhou@163.com (Z.Z.); mshehzad534@gmail.com (M.Z.); cxycri@163.com (X.C.); xuyanchao2016@163.com (Y.X.); houyuqing18@163.com (Y.H.); liufcri@163.com (F.L.) (*corresponding author); wkbcri@163.com (K.W.) (*corresponding author)*

[2]*Nuclear Institute for Agriculture and Biology (NIAB), P.O. Box 128, Jhang Road, Faisalabad 38000, Punjab, Pakistan; mkrkhan@gmail.com (M.K.R.K.)*

[3]*NIAB-C, Pakistan Institute of Applied and Engineering Sciences (PIEAS), Nilor Islamabad, Pakistan*

[4]*National Nanfan Research Institute of Chinese Academy of Agriculture Sciences, Sanya, China*

[5]*King Saud University, College of Science, Department of Biochemistry, P.O. Box 2455, Riyadh 11451, Saudi Arabia; kattia1.c@ksu.edu.sa(K.A.A.); moaahmed@ksu.edu.sa (A.A.M.)*

[6]*Kafr-el-sheikh University, Faculty of Agriculture, Plant Pathology and Biotechnology Laboratory, Agriculture Botany Department, EPCRS Excellence Center, 33516, Egypt; hafezyasser@gmail.com (Y.M.H.)*

[7]*Nantong University, School of Life Sciences, China; bhwangzm@126.com (B.W.)*

[8]*The University of Lahore, Institute of Molecular Biology and Biotechnology, 59540 Lahore, Pakistan; sajidfiaz50@yahoo.com (S.F.)*

[9]*East Carolina University, Department of Biology, Greenville, NC 27858, USA; zhangb@ecu.edu (B.Z.)*

[a,b]*These authors contributed equally to the work*

Abstract

Genetic maps highlight the genome organization and structure but also provide the chance of tagging superior traits for crop improvement through marker-assisted selection. Amino acids are building blocks of proteins and perform crucial function in regulating the signaling of molecules involved in the development and growth of plants. Plant architecture also have an impact on crop productivity. In order to select elite cultivars for breeding and identification of favorable alleles and their functional properties, a deep understanding of genetic architecture and development of genetic map is essential. In present investigation, an interspecific cross of *Gossypium barbadense* XH-18 × *G. darwinii* 5-7 was made to develop a genetic map utilizing single sequence repeat markers for the dissection of amino acids involved in genetic architecture of *G. barbadense* and *G. darwinii*. We measured chromosomal distribution of 20 amino acids across the whole genome of both species. The map consists of 613 markers spread across all 26 chromosomes, covering 2371.4 cM of cotton genome with an average inter-marker distance of 9.35 cM. The marker number anchored on the chromosomes varied from 5 to 76 with an average of 23.57 on each chromosome. The Dt sub-genome had more markers (83.03%)

Received: 04 Dec 2023. Received in revised form: 08 Apr 2024. Accepted: 03 Jun 2024. Published online: 20 Aug 2024.
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

than the At sub-genome (15.66%). Moreover, the longest chromosome was 143.387 cM, the shortest was 58.430 cM, and the average length was 91.207 cM. The Dt subgenome spans a greater genomic distance than the At subgenome. A sum of 21,035 genes were discovered, covering the complete genome of *G. barbadense*; *G. darwinii* and have been found to be involved in tRNA 3'-trailer cleavage, macromolecule modification, peptide deformylase activity, response to biotic stimulus and defense response. The minimum Glutamic acid (Glu), Histidine (His), and Lysine (Lys) were found on Chr.13 (0.00-17.74), Chr.02 (0.00-8.01), and Chr.06 (0.00-17.97), respectively found through chromosomal amino acid dissection. The genome-wide SSR interspecific genetic map of *G. barbadense* and *G. darwinii* is first of its kind, and studying chromosomal distribution of amino acids will set a landmark step to dissect the genome structure of *G. darwinii*.

Keywords: amino acid; fiber quality; genetic map; genetic variation; interspecific cross; wild cotton

Introduction

Cotton is known as white gold and considered a valuable agricultural commodity with diverse applications. It is primarily used for fabric production but also utilized in items like paper, household goods, and medical products. Additionally, cotton plays a vital role in feed and oil industries owing to nutrient-rich seeds, containing significant amounts of oil (28-33%) and protein (28-34%), respectively (Yu *et al*., 2012). Taxonomically, cotton is classified under the genus *Gossypium* and displays a wide range of phenotypic variations, encompassing over fifty distinct species (Wendel *et al*., 2009; Wendel anf Grover, 2015; Gallagher *et al*., 2017). Among these species, seven are tetraploid and forty-six are diploidin nature. The worldwide cultivation of cotton is mainly consisting of four distinct types, among two are diploids 2n = 2x = 26, and two are being allotetraploids 2n = 4x = 52. The two allotetraploid species known as *Gossypium barbadense* and *G. hirsutum* are the primary global cotton contributor (Wendel *et al*., 1992; Wang *et a*l., 2015b).

Approximately 90% of the world cotton production is attributed to *G. hirsutum*, known as upland cotton (Campbell *et al*., 2010) whereas, *G. barbadense* known as Pima cotton is esteemed for its exceptional quality of extra-long fibers. Pima cotton is cultivated in North-West South America and originated in modern day Egypt. It contributes approximately 8% to global lint yield (Wang *et al*., 2015a). The abundant genetic variations present in wild cotton have been extensively utilized to enhancing yield potential of cultivated cotton species. The DNA of *G. darwinii* Watt (AD 5), an allotetraploid cotton was solely found at Galapagos Islands. This species belongs to the *Karpas rafinesque* subgenus, which also encompasses *G. hirsutum*, *G. barbadense*, *G. tomentosum*, and *G. mustelium*. According to the classification of cotton germplasm resources, *G. darwinii* Watt and *G. hirsutum* belongs to the first germplasm bank and have the ability to directly crossbreed among each other. *G. darwinii*, a wild allotetraploid cotton displayed resilience for drought and nematodes, along with possessing fine fiber characteristics (Chen *et al*., 2015). Keeping in view, the limited genetic diversity of *G. hirsutum*, it is essential to expand its genetic basis for traits of fundamental importance (Wendel and Percy, 1990; Brubaker, Paterson and Wendel, 1999; Wendel, Brubaker and Seelanan, 2010; Wang *et al*., 2012). Recent research has revealed that *G. darwinii* containing AD5 genome shares an evolutionary relationship with *G. barbadense* whereas, divergent from the cultivated *G. hirsutum*. Notably, *G. darwinii* displays premium traits i.e., superior fiber quality, drought tolerance, and resistance to fusarium and verticillium wilt. These findings highlight the valuable attributes of *G. darwinii* and suggest its promising prospects for addressing specific objectives (Liu *et al*., 2016; Ditta *et al*., 2018).

Molecular markers are successfully applied because of their reliability in directly detecting allelic diversity and providing precise estimations of genetic distances. However, in cotton research the application of molecular markers as molecular mileposts has not been widespread compared to other crop species owing to

multiple reasons. Among these challenges, unavailability of a high-quality DNA extraction protocol for various molecular marker analyses is lacking. Furthermore, upland cotton possesses a limited genetic foundation, leading to a scarcity of polymorphism within the prominent varieties (Saha *et al*., 2003). Lastly, high level of homologous chromosome recombination and genetic redundancy, which hampers linkage analysis (Reinisch *et al*., 1994). It has been observed, some chromosomal regions contain recombination thereby, promoting wider genetic distances whereas, majority of chromosomal regions have low recombination rates owing to heterochromatin. The challenges mentioned above have been addressed through the development of abundant cotton-specific molecular markers i.e., SSRs simple sequence repeats (SSR). The construction of genetic map plays a crucial role in understanding the genomic basis for marker-assisted selection (MAS) and map-based cloning. It allows to harness the potential of *G. darwinii* in breeding programs, particularly for functional characterization of candidate genes associated with various traits of interest. To date, numerous high-density genetic molecular maps have been developed for various mapping populations using molecular markers (Chen *et al*., 2015; Khan *et al*., 2016). However, most of these maps are derived from hybrids between *G. hirsutum* and *G. barbadense* or hybrids between *G. hirsutum* and *G. darwinii* (Chen *et al*., 2016). However, there is lack of reported maps between *G. barbadense* and *G. darwinii.* Recently, a genetic map has been established between *G. barbadense* and *G. darwinii* using South Western University of China (SWU)-SSR markers derived from the D-genome. The SWU-SSRs, developed by the South Western University of China, have recently emerged as a novel tool for genetic map construction in tetraploid cotton, particularly in *G. barbadense* and *G. darwinii*. The choice of markers was primarily based on the fact that majority of fiber trait genes are linked to D-genome rather than A-genome (Jamshed *et al*., 2016).

For cotton molecular breeding, it is essential to conduct a comprehensive linkage mapping of the hybrid between *G. barbadense* and *G. darwinii*, focusing whole D-genome. The mapping will enable the efficient transfer of favorable genes responsible for strong fibers and drought resistance by studying the chromosomal distribution of gene/s. In present investigation, we successfully developed SSR-based linkage map derived from the D-genome of tetraploid cotton by utilizing the genetic resources of wild cotton. However, to further enhance the density of the genetic map, additional markers will be required which can be incorporated in subsequent experiments. The findings of present research will contribute to the differentiation of closely related tetraploid cotton genomes i.e., *G. barbadense* and *G. darwinii*. Furthermore, it will facilitate the identification of DNA based markers associated with fiber quality traits and drought tolerance in cotton.

Materials and Methods

The cultivar 'XH-18' (EG16002), a variant of *G. barbadense* L., is known for its exceptional stability and possesses unique characteristics such as extra-long fibers. On contrary, *G. darwinii 5-7* (EG16004), introduced from the exotic *G. darwinii* G. Watt., exhibits desirable traits such as strong fibers, tolerance to drought and salt, as well as nectariless. The strict self-pollination was employed to maintain parental material, developed at e National Wild Cotton Germplasm Nursery in Hainan, China.

Development of population and wild accessions

The hybridization involved crossing between *G. barbadense* var. 'XH-18' (as the female parent) with *G. darwinii* Watt 'darwinii 5-7' (as the male parent), to develop F_1 hybrid. The self-fertilization of the F_1 hybrid led to the development of 233 individuals in the F_2 generation. These F_2 plants were then planted in the field at China's Sanya Wild Cotton Germplasm Nursery. The F_2 population were carefully preserved for constructing a genetic map. Each autumn, the above-ground sections of the plants were pruned for effective preservation and maintenance.

DNA extraction and quantification

Young leaves from the parental plant materials, F_1 hybrids, and 233 F_2 individuals were collected and placed in an ice box for transportation to laboratory at Institute of Cotton Research (ICR), CAAS (Chinese Academy of Agricultural Sciences), located at Anyang, China. The samples were stored in an ultra-low temperature refrigerator at -80 °C to ensure freezing. The genomic DNA was extracted from the frozen leaves using Cetyltrimethylammonium bromide (CTAB) method (Zhang & Stewart, 2000)with minor adjustments. The quality of DNA was assessed by subjecting it to agarose gel electrophoresis with a concentration of 1% (w/v). The quantity of DNA was measured using Nanodrop, specifically by determining the absorbance ratio at 260/280 nm.

PCR mixture and program

Polymerase chain reaction (PCR) is a laboratory technique employed to amplify specific targeted DNA fragments in vitro, demonstrating its capability to amplify even minute quantities of DNA. The reaction volume of 10 µL was employed, comprising 5 µL of 2× Taq Master Mix containing buffer, dNTPs, and Taq DNA Polymerase, 2 µL of primers, 1 µL of DNA, and 2 µL of ddH_2O. The PCR amplification process was carried out utilizing the TAKARABio Inc. TP 600 Thermal Cycler, and the silver staining procedure was conducted following the protocol outlined by Zhang *et al.* (2002). The PCR amplification procedure began with an initial pre-denaturation step consisting of two cycles at 95 °C for 3 minutes followed by 30 cycles, each involving denaturation at 94 °C for 45 seconds, annealing at 57 °C for 36 seconds, and extension at 72 °C for 1 minute. Finally, a final extension step was performed at 72 °C for 5 minutes. To verify the reported amplicons were indeed amplified from genomic DNA and not due to primer artifacts, the control reaction excluded genomic DNA. In all PCR experiment conducted, no amplification products were detected in the absence of genomic DNA.

Gel reading and scoring of SSR markers

The genotyping process involved manual scoring of SSR markers based on gel analysis. Once polymorphic markers were selected, they were assessed on 233 plants from the F_2 generation. The primers used were classified as dominant, co-dominant, or recessive, based on the analysis of band patterns observed in both the F_1 hybrid and the parental plants (P1 and P2). For co-dominant markers, scoring was conducted following the Mendelian ratio of 1:2:1 for the F_2 population. The dominant markers were scored using the mendelian segregation ratio of 3:1. The SSR amplicons were denoted as "10" when a single upper band was present, "11" when two bands were observed, and "01" when a single lower band was detected. The presence of dominant loci was assigned the value "1," while the absence of dominant loci was assigned "0." Missing data, which included fuzzy or imprecise bands, were indicated by the "-" sign. In case of multi-allelic markers generating multiple distinct bands on a gel and each band representing a different molecular weight indicated segregation for multiple loci.

Genetic map construction

In this study, JoinMap 4.0 was utilized for map construction. Initially, the data codes were transformed from the "10" format to the following designations as required by JoinMap 4.0: "A" for Homozygous Female Parent, "B" for Homozygous Male Parent, "H" for Heterozygotes, "C" for cases where the B allele is dominant (representing genotypes similar to the male parent and heterozygotes), "D" for cases where the A allele is dominant (representing genotypes similar to the female parent and heterozygotes), and "-" for missing data. In this study, the female parent is *G. barbadense*, 'XH-18' (RG16004), and the male parent is *G. darwinii*, 'darwinii 5-7'. The co-dominant markers were designated using the ABH code. In this code, "A" represents the female parent, "B" represents the male parent, and "H" represents heterozygotes. In F_2 populations, individuals that displayed a band pattern similar to that of the female parent were denoted as "A," while individuals resembling the male parent were labeled as "B." If an individual exhibited both the band patterns of the female

and male parents, it was assigned the code "H." For dominant primers, the Code-DDB was employed in situations where the female parent exhibited a band indicating dominance. In such cases, F_2 progenies resembling the female parent were designated as "D," while those resembling the male parent (no band, 0) were denoted as "B." On the other hand, the Code-ACC was utilized when the male parent displayed dominance. In this scenario, progenies resembling the male parent were labeled as "C," whereas those resembling the female parent were given the code "A." The missing data were indicated by the "-" symbol.

Running JoinMap 4.0

In the case of dominant primers, the Code-DDB was used when the female parent displayed a dominant band. Consequently, F_2 progenies that resembled the female parent were assigned the code "D," while those resembling the male parent (lacking a band, coded as 0) were designated as "B." Alternatively, when the male parent exhibited dominance, the Code-ACC was applied. Progenies resembling the male parent were assigned the code "C," while those resembling the female parent were denoted as "A." Missing data were represented by the "-" symbol.

Segregation distortion

Interspecific crosses often exhibit distortion in the segregation of traits, leading to non-standard segregation ratios. To identify such deviations, the Chi-square ($\chi 2$) test was employed. The analysis involved selecting "Locus genotype Freq" and examining the data to assess skewness in the segregation ratios. Following the criteria established by Yu et al. (2011) a region was considered as a segregation distorted region (SDR) if it contained a minimum of three consecutive loci exhibiting significant distortion ($P<0.05$).

Genetic distances, locus order and mapping functions

The highest recombination rate 0.40 was observed which was utilized to convert recombination frequencies into map distances (centimorgan, cM) using the Kosambi mapping algorithm. In determining the relationship between two loci, a LOD score greater than 3.0 indicates a presumed linkage. For our analysis, we employed the independence LOD parameter, with a threshold value range from "Start" set at 3 to "End" set at 50, using a step of 1. This parameter allows the software to generate different trees based on the LOD values. At lower LOD values, most markers tend to group together, while at higher LOD values, markers tend to distribute into more distinct groups. A LOD score of 35 was chosen as the threshold to generate different linkage groups. Initially, we obtained 53 smaller linkage groups as a result of this process.

Chromosomal assignment

To consolidate the 53 small linkage groups into larger linkage groups representing all 26 chromosomes, a LOD threshold of ≥10 was utilized. However, some primers were not successfully assigned to any linkage group during this grouping process. Subsequently, a map was generated using Mapchart 2.2 software by inputting the loci and map distances obtained from JoinMap 4.0. The assignment of loci to specific chromosomes was based on information from previously published studies that had anchored SSR markers to respective chromosomes. In this study, SSR loci located on chromosomes 1 to 13 were assigned to the A sub-genome (At), while loci situated on chromosomes 14 to 26 were assigned to the D sub-genome (Dt) by following the established methodology outlined by Chen *et al*. (2015).

Gene identification and amino acid dissection

The SSR marker sequences were subjected to a homology search using the cotton functional genomic database (https://cottonfgd.org), resulting the identification of 21,035 genes shown in Supplementary Table S2. These genetic factors were found to be distributed across all 26 chromosomes. Furthermore, analysis of physiochemical properties of the identified genes was conducted in order to obtain diverse descriptions of the identified genes. Additionally, an amino acid dissection was performed to investigate the protein production

patterns and to characterize their roles in various gene functions related to fiber quality traits, as well as biotic and abiotic stress tolerance in tetraploid cotton.

Results

Parental screening

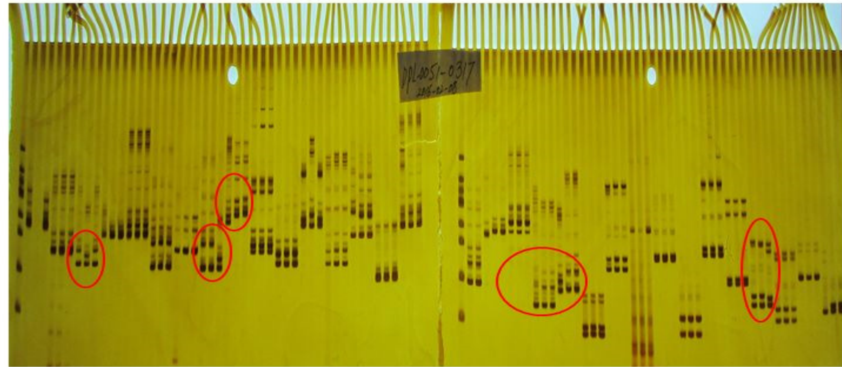
In present investigation, we utilized a total of 18,201 SSR primer pairs to investigate polymorphism between the parental lines, namely *G. darwinii* 5-7 (EG16002) and *G. barbadense* XH-18 (EG16004). The results of the parental screening process are described in Figure 1, presenting selected portion of the screening outcomes. Out of the total primers used, 13,577 (74.59%) were classified as EST-gSSRs, while 4,624 (25.40%) were categorized as EST-SSRs. Among the parents, a total of 2,648 SSRs exhibited polymorphism, with an average polymorphism rate of 14.54%. Within this set, 14.67% (1,993 SSRs) were identified as EST-gSSRs, while 655 SSRs were classified as EST-SSRs, showing a polymorphism rate of 14.16%. The BNL series markers displayed the highest polymorphism rate of 32.67%, followed by the DPL series with a polymorphism rate of 28.57%. Conversely, the NAU series exhibited the lowest polymorphism rate at 11.56%. The TMB and MON_CGR series of SSR markers demonstrated similar levels of polymorphism, with rates of 24.57% and 24.51%, respectively.

Among the EST-SSRs, the HAU produced highest polymorphism (16.04%) while MGHES series showed second highest polymorphism 15.54% and lowest polymorphism rate (11.56%) was observed in NAU series, respectively. In addition, total 550 polymorphic SWU SSR primers were applied during the research for construction of genetic map. The sequence of SWU SSR primers is provided in supplementary Table S1. This study found that the polymorphism rate of gSSR primers was higher compared to that of EST-SSR primers. The Table 1 and Figures 1 and 2 displayed the performance of SSR primers from various sources in terms of polymorphism between parents. The F_2 individuals were evaluated for genotyping and genetic map construction as shown in Figure 2.

Table 1. Genotyping of all SSRs used in the study

Primers Name	Primers Screened	Polymorphic Primers
SWU	12500	1740
HAU	2586	415
NAU	1954	226
TMB	350	86
MON_CGR	310	76
BNL	101	33
JESPR	150	27
MUSB	138	23
MGHES	84	14
DPL	28	8
Total	**18201**	**2648**

* All available primers were tested, and the polymorphism was detected as described above.



*G.darwanii*5-7, XH18 and XH21 screening with different DPL Primers

Figure 1. Screening of parents for polymorphism

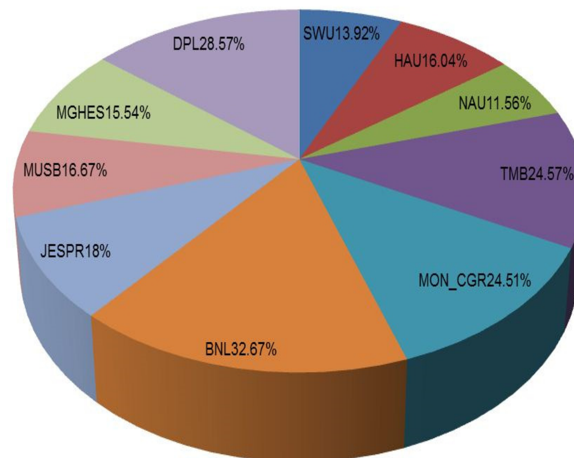


Figure 2. Percentage of polymorphism for the all SSR screened against the parents

Population genotyping

A set of 613 polymorphic SSR primer loci were employed to genotype 233 individuals from the F_2 population. The results of the SWU15345 primer can be seen in **Figure 3**, which depicts the PAGE analysis. The gel clearly displayed the presence of two distinct bands, indicating stable amplicons generated by this particular primer at different segregating pattern positions.

G. barbadense **×** ***G. darwinii*** **5-7 F_2 Evaluation**

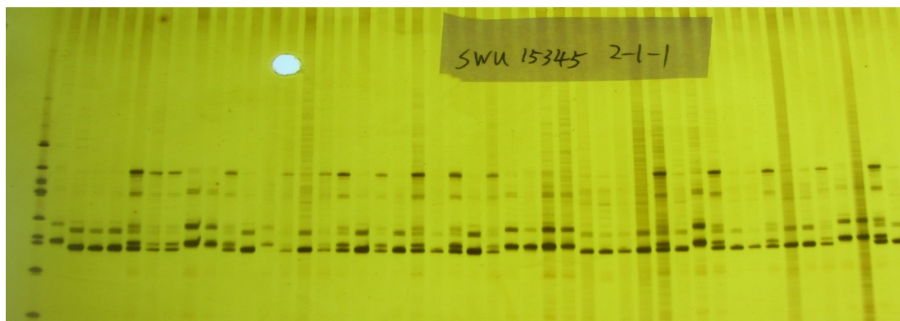


Figure 3. Electrophoresis pattern of F_2 population along with parent's *darwinii* 5-7 (*G. darwinii*) and *G. barbadense* (EG16004) by using gSSR marker SWU15345

Genetic map

In the present study, JoinMap 4.0 mapping software (Stam, 1993) was utilized to perform linkage analysis using data from 613 polymorphic loci. The LOD threshold was set at ≥ 10, and the maximum distance between markers was limited to ≤ 50 cM resulting a genetic linkage map comprising of 509 SSR primers was constructed, which successfully mapped a total of 613 SSR marker loci on the 26 chromosomes of cotton. However, BLAST analysis results shown 96 SSR loci out of the 613 markers were found to be distributed on genome whereas, remaining 517 loci were located in the Dt genome. The genetic map encompassed alleles derived from both *G. darwinii 5-7* and *Gossypium barbadense* XH-18 parental lines.

Genetic linkage map features

A total of 550 markers exhibited polymorphism between the parental lines 'XH-18' and 'Darwinii 5-7' and F_2 population. However, 41 loci could not be mapped due to significant segregation skewness and missing data, resulting their inability to form linkage groups. Consequently, the remaining 509 markers were successfully used to construct 26 linkage groups corresponding to the 26 chromosomes. The genetic map constructed using these 509 mapped markers had a total length of 2371.384 cM. The average inter-marker distance was found to be 15.755 cM for the A-genome and 2.949 cM for the D-genome. Detailed information regarding these results can be found in Table 2 and Figure 4. The constructed map demonstrated an average chromosome length of 91.207 cM, with an average of 118.96 loci per chromosome. To represent the A sub-genome, previously known as chromosome 1 (Chr.1), these chromosomes were labeled as Chr.1 (A1). Notably, the A and D sub-genomes exhibited variations for chromosomal length and the number of loci they contained. Among these chromosomes, Chr.16(D07) was identified as the longest chromosome in the D sub-genome, spanning 107.4 cM and containing of 76 loci 76. On contrary, Chr.2(A02) was the shortest, covering only 79.243 cM and consisting of 5 loci.

Chr.22(D04) exhibited the lowest number of loci, with only 21 markers mapped across a distance of 143.378 cM. Moreover, the largest inter-marker distance 46.794 cM was observed on the same chromosome. In contrast, Chr.24(D08) displayed the shortest inter-marker distance of 9.408 cM, indicating a higher density of loci on this chromosome. Among these chromosomes, Chr.3(A03) had the smallest inter-marker interval of 0.009 cM, indicating closely spaced markers. Similarily, Chr.3(A03) also exhibited the largest inter-loci gap of 49.678 cM, suggesting a significant distance between adjacent loci on this chromosome. The analysis identified a sum of 81 gaps larger than 10 cM, with 37 gaps present in the At sub-genome and 18 gaps in the Dt sub-genome (Table 3 and Figure 4). Similarily A sub-genome displayed a sum of 96 loci covering a genomic distance of 1145.771 cM, with an average inter-marker interval of 15.755 cM. The largest inter-marker distance observed in this sub-genome was 36.569 cM whereas, the smallest was 3.253 cM. Among the chromosomes in the A sub-genome, Chr.11(A11) was identified as the longest, spanning 69.766 cM, containing 13 loci. Similarly, Chr.06(A06) covered 82.183 cM and consisted of 12 loci.

In the At sub-genome, there was a sum of 37 gaps larger than 10 cM, with the largest gap observed on Chr.03 (A03) spanning over 49.417 cM (Table 3). The genomic distance covered by the 517 loci in the Dt sub-genome covered genomic distance of 1225.613 cM, with 2.949 cM of average interval between markers of. Among these markers, the largest inter-marker interval measures 19.281 cM whereas, smallest interval 0.103 cM. As mentioned previously, Chr.16(D07) displayed the highest recombination frequency and represented the longest chromosome meanwhile, Chr.24(D08) spans 92.439 cM and encompasses 57 marker loci. In the Dt sub-genome, there are 18 gaps larger than 10 cM, with the largest gap observed on Chr.22(D04), which spans a distance of 46.794 cM (Table 3).

The distribution of SSR markers between the At and Dt sub-genomes is not uniform, with a higher number of SSRs observed on the Dt sub-genome. Additionally, there is an uneven distribution of SSR markers among the chromosomes, with certain chromosomes exhibiting a greater abundance of SSR markers compared to others. Specifically, Chr.11(A11), Chr.16(D07), Chr.23(D09), Chr.24(D08), Chr.20(D10), and Chr.26 have a higher number of SSR markers. Furthermore, the distribution of SSR markers varies within each

chromosome. For instance, there was an even distribution of 9 SSR markers on Chr.01(A01), Chr.05(A05), 5 SSR markers on Chr.02(A02), Chr.03(A03), Chr.09(A09), and 7 SSR markers on Chr.07(A07), Chr.13(A13), while Chr.19(D05) and Chr.25(D06) contain 22 SSR markers. However, the distribution of SSR markers differs on the remaining chromosomes (Table 2, Figure S1).

Table 2. The main characteristics of the genomic map in this research

Chromosome	Number of Marker Loci	Recombination	Average distance between locations (cM)	The smallest gap distance (cM)	The most significant disparity (cM)	Number of gaps (>10cM)	Number of skewed/SD loci	Skewness percentage/standard deviation percentage
Chr.01(A01)	9	92.828	11.603	0.959	48.148	2	4	44.44
Chr.02(A02)	5	79.243	19.810	4.398	35.542	2	2	40.00
Chr.03(A03)	5	61.081	15.270	1.510	49.417	1	3	60.00
Chr.04(A04)	6	58.430	11.686	0.583	31.477	2	2	33.33
Chr.05(A05)	9	67.478	8.434	1.610	16.201	3	4	44.44
Chr.06(A06)	12	82.183	7.471	0.188	20.879	3	7	58.33
Chr.07(A07)	7	103.065	17.175	5.004	34.191	4	2	28.57
Chr.08(A08)	6	130.263	26.052	0.269	43.824	4	3	50.00
Chr.09(A09)	5	100.000	25.000	13.402	49.678	4	4	80.00
Chr.10(A10)	6	111.483	22.296	6.306	38.04	3	5	83.33
Chr.11(A11)	13	69.766	5.813	0.216	25.201	2	5	38.46
Chr.12(A12)	6	76.451	15.290	3.093	34.895	3	1	16.67
Chr.13(A13)	7	113.500	18.916	4.750	47.901	4	3	42.86
Subtotal	**96**	**1145.771**	**15.755**	**3.253**	**36.569**	**37**	**45**	**46.88**
Chr.15(D01)	34	100.000	3.125	0.084	14.004	3	13	38.235
Chr.14(D02)	35	83.742	2.701	0.212	10.432	1	18	51.429
Chr.17(D03)	32	83.191	2.683	0.077	10.813	1	13	40.625
Chr.22(D04)	21	143.387	7.169	0.171	46.794	3	8	38.095
Chr.19(D05)	22	94.917	4.519	0.324	15.368	3	13	59.091
Chr.25(D06)	22	91.044	4.335	0.143	19.865	2	13	59.091
Chr.16(D07)	76	107.372	1.431	0.013	40.003	1	27	35.526
Chr.24(D08)	57	92.439	1.650	0.009	9.408	0	17	29.825
Chr.23(D09)	61	83.644	1.394	0.050	13.362	1	27	44.262
Chr.20(D10)	52	93.364	1.830	0.010	10.385	1	18	34.615
Chr.21(D11)	32	82.137	2.649	0.071	15.431	1	17	53.125
Chr.26(D12)	44	99.000	2.301	0.047	30.749	1	15	34.091
Chr.18(D13)	29	71.376	2.549	0.125	14.045	1	13	44.828
Sub Total	**517**	**1225.613**	**2.949**	**0.103**	**19.281**	**19**	**212**	**41.006**

Distinctive features of warped segregation markers

The segregation distortion was observed in 257 marker loci, which deviated from the expected Mendelian ratios. These distorted loci were mapped at various cotton chromosomes, accounting 41.92% of the total number of mapped loci. Among these loci, 58 exhibited segregations towards the 'darwinii 5-7' allele, 34 showed segregation towards the 'XH-18' allele, while the remaining 165 loci displayed segregation towards the heterozygous allele (Figure 4). The proportion of distorted loci based on SSR analysis was found to be 17.829% (Table 3, Figure S1). However, the distribution of these distorted loci is not uniform across all 26 cotton chromosomes, with varying numbers of observed markers on each chromosome, ranging from 1 to 27 (Table 3). Notably, the Dt sub-genome contained 212 distorted loci, which is approximately 41.0% higher compared to the At sub-genome containing 45 distorted loci of 7.34%.

In this study, a sum of 51 SDRs were discovered in both At and Dt sub-genomes, with 10 SDRs located at At sub-genome and 41 at Dt sub-genome. Chromosomes Chr.16(D07), Chr.23(D09), Chr.14(D02), and Chr.21(D11) were found having highest number of distorted loci. Notably, the deformed loci tended to cluster on specific chromosomes or within the same SDRs i.e., Chr.14(D02), Chr.15(D01), Chr.16(D07), Chr.17(D03), Chr.18(D13), Chr.19(D05), Chr.20(D10), Chr.21(D11), Chr.23(D09), Chr.24(D08), Chr.25(D06), and Chr.26(D12) exhibited clustering of loci skewing toward the same allele (Figure 4). The

largest SDRs were identified as SDR4_15, SDR9_14, SDR18_19, SDR22_25, SDR23_16, SDR26_16, SDR30_24, SDR33_23, SDR37_20, SDR41_21, SDR45_26, and SDR50_18. The Chr.16(D07) and Chr.23(D09) had the highest number of distorted loci a sum of 27 whereas, Chr.12(A12) had the lowest distortion rate, with less than 20% of distortion. Chromosome Chr.12(A12) exhibited the lowest percentage of distorted loci at 16.67% whereas, Chr.10(A10) had the highest at 83.33% distortion. Notably, Chr.09(A09), Chr.03(A03), Chr.19(D05), and Chr.25(D06) displayed significantly higher distortion ratios of 80%, 60%, 59.091%, and 59.091%, respectively compared to other chromosomes. The overall segregation ratio across the entire genome was 43.94%, which was higher compared to other maps (Table 3).

Cottons undergo duplication, rearrangement, and translocation

Among the 613 SSR primer pairs used in current mapping study, 106 pairs (17.29%) amplified two loci, leading to the presence of duplicated loci that connected the 13 At/Dt chromosomes. Out of these 106 duplicated loci, 60% were found on the same chromosomewhereas, 93.39% were located on non-homologous chromosomes. These findings suggested that various rounds of duplication, including inter-chromosomal duplication, have occurred during the evolutionary process (Table 3). Furthermore, within the duplicated loci of the Dt sub-genome, five pairs showed two post-polyploidization translocations involving D1/D9, D3/D10, D4/D7, and D4/D18. The majority of duplicated loci were identified between the At and Dt chromosomes.

Table 3. Detail of duplicated marker loci and their chromosomal relationship

Sr. No.	Marker Loci	Chromosome	Marker Loci	Chromosome	Relationship
1	SWU10037	Chr.07(A07)	SWU10037	Chr.15(D01)	Chr.07-Chr.15
2	SWU10072	Chr.07(A07)	SWU10072	Chr.15(D01)	Chr.07-Chr.15
3	SWU10174	Chr.07(A07)	SWU10174	Chr.15(D01)	Chr.07-Chr.15
4	SWU10321	Chr.07(A07)	SWU10321	Chr.15(D01)	Chr.07-Chr.15
5	SWU10429	Chr.07(A07)	SWU10429	Chr.15(D01)	Chr.07-Chr.15
6	SWU10741	Chr.07(A07)	SWU10741	Chr.15(D01)	Chr.07-Chr.15
7	SWU10394	Chr.07(A07)	SWU10394	Chr.15(D01)	Chr.07-Chr.15
8	SWU10877	Chr.07(A07)	SWU10877	Chr.15(D01)	Chr.07-Chr.15
9	SWU10912	Chr.01(A01)	SWU10912	Chr.14(D02)	Chr.01-Chr.14
10	SWU10942	Chr.01(A01)	SWU10942	Chr.14(D02)	Chr.01-Chr.14
11	SWU11087	Chr.01(A01)	SWU11087	Chr.14(D02)	Chr.01-Chr.14
12	SWU11115	Chr.01(A01)	SWU11115	Chr.14(D02)	Chr.01-Chr.14
13	SWU11196	Chr.01(A01)	SWU11196	Chr.14(D02)	Chr.01-Chr.14
14	SWU11318	Chr.01(A01)	SWU11318	Chr.14(D02)	Chr.01-Chr.14
15	SWU11685	Chr.02(A02)	SWU11685	Chr.14(D02)	Chr.01-Chr.14
16	SWU11747	Chr.02(A02)	SWU11747	Chr.14(D02)	Chr.01-Chr.14
17	SWU11855	Chr.02(A02)	SWU11855	Chr.17(D03)	Chr.02-Chr17
18	SWU11989	Chr.02(A02)	SWU11989	Chr.17(D03)	Chr.02-Chr17
19	SWU12042	Chr.02(A02)	SWU12042	Chr.17(D03)	Chr.02-Chr17
20	SWU12187	Chr.02(A02)	SWU12187	Chr.17(D03)	Chr.02-Chr17
21	SWU12297	Chr.01(A01)	SWU12297	Chr.17(D03)	Chr.02-Chr17
22	SWU12343	Chr.01(A01)	SWU12343	Chr.17(D03)	Chr.02-Chr17
23	SWU12391	Chr.01(A01)	SWU12391	Chr.17(D03)	Chr.02-Chr17
24	SWU12958	Chr.01(A01)	SWU12958	Chr.17(D03)	Chr.02-Chr17
25	SWU13551	Chr.08(A08)	SWU13551	Chr.22(D04)	Chr.08-Chr.22
26	SWU13613	Chr.08(A08)	SWU13613	Chr.22(D04)	Chr.08-Chr.22
27	SWU13666	Chr.08(A08)	SWU13666	Chr.22(D04)	Chr.08-Chr.22
28	SWU13677	Chr.08(A08)	SWU13677	Chr.22(D04)	Chr.08-Chr.22
29	SWU13858	Chr.08(A08)	SWU13858	Chr.22(D04)	Chr.08-Chr.22
30	SWU13872	Chr.08(A08)	SWU13872	Chr.22(D04)	Chr.08-Chr.22

31	SWU13895	Chr.03(A03)	SWU13895	Chr.19(D05)	Chr.03-Chr.19
32	SWU14028	Chr.03(A03)	SWU14028	Chr.19(D05)	Chr.03-Chr.19
33	SWU14590	Chr.03(A03)	SWU14590	Chr.19(D05)	Chr.03-Chr.19
34	SWU14604	Chr.03(A03)	SWU14604	Chr.19(D05)	Chr.03-Chr.19
35	SWU14709	Chr.03(A03)	SWU14709	Chr.19(D05)	Chr.03-Chr.19
36	SWU14809	Chr.09(A09)	SWU14809	Chr.25(D06)	Chr.09-Chr.25
37	SWU14914	Scaff	SWU14914	Chr.25(D06)	Scaff-Chr.25
38	SWU15194	Chr.09(A09)	SWU15194	Chr.25(D06)	Chr.09-Chr.25
39	SWU15212	Chr.09(A09)	SWU15212	Chr.25(D06)	Chr.09-Chr.25
40	SWU15357	Chr.09(A09)	SWU15357	Chr.25(D06)	Chr.09-Chr.25
41	SWU15473	Chr.09(A09)	SWU15473	Chr.25(D06)	Chr.09-Chr.25
42	SWU15642	Chr.11(A11)	SWU15642	Chr.16(D07)	Chr.11-Chr.16
43	SWU15705	Chr.11(A11)	SWU15705	Chr.16(D07)	Chr.11-Chr.16
44	SWU15744	Chr.11(A11)	SWU15744	Chr.16(D07)	Chr.11-Chr.16
45	SWU15751	Chr.11(A11)	SWU15751	Chr.16(D07)	Chr.11-Chr.16
46	SWU15760	Chr.11(A11)	SWU15760	Chr.16(D07)	Chr.11-Chr.16
47	SWU15794	Chr.11(A11)	SWU15794	Chr.16(D07)	Chr.11-Chr.16
48	SWU15861	Chr.11(A11)	SWU15861	Chr.16(D07)	Chr.11-Chr.16
49	SWU15927	Chr.11(A11)	SWU15927	Chr.16(D07)	Chr.11-Chr.16
50	SWU16136	Chr.11(A11)	SWU16136	Chr.16(D07)	Chr.11-Chr.16
51	SWU16149	Chr.11(A11)	SWU16149	Chr.16(D07)	Chr.11-Chr.16
52	SWU16186	Chr.11(A11)	SWU16186	Chr.16(D07)	Chr.11-Chr.16
53	SWU16281	Chr.11(A11)	SWU16281	Chr.16(D07)	Chr.11-Chr.16
54	SWU16317	Chr.11(A11)	SWU16317	Chr.16(D07)	Chr.11-Chr.16
55	SWU16428	Scaff	SWU16428	Chr.16(D07)	Scaff-Chr.16
56	SWU16698	Chr.12(A12)	SWU16698	Chr.24(D08)	Chr.12-Chr.24
57	SWU16786	Chr.12(A12)	SWU16786	Chr.24(D08)	Chr.12-Chr.24
58	SWU16876	Chr.12(A12)	SWU16876	Chr.24(D08)	Chr.12-Chr.24
59	SWU17019	Chr.12(A12)	SWU17019	Chr.24(D08)	Chr.12-Chr.24
60	SWU17044	Chr.12(A12)	SWU17044	Chr.24(D08)	Chr.12-Chr.24
61	SWU17112	Scaff	SWU17112	Chr.24(D08)	Scaff-Chr.24
62	SWU17161	Chr.12(A12)	SWU17161	Chr.24(D08)	Chr.12-Chr.24
63	SWU17323	Chr.12(A12)	SWU17323	Chr.24(D08)	Chr.12-Chr.24
64	SWU17884	Chr.05(A05)	SWU17884	Chr.23(D09)	Chr.05-Chr.23
65	SWU18038	Chr.05(A05)	SWU18038	Chr.23(D09)	Chr.05-Chr.23
66	SWU18049	Chr.05(A05)	SWU18049	Chr.23(D09)	Chr.05-Chr.23
67	SWU18055	Chr.05(A05)	SWU18055	Chr.23(D09)	Chr.05-Chr.23
68	SWU18141	Chr.05(A05)	SWU18141	Chr.23(D09)	Chr.05-Chr.23
69	SWU18208	Chr.05(A05)	SWU18208	Chr.23(D09)	Chr.05-Chr.23
70	SWU18254	Chr.05(A05)	SWU18254	Chr.23(D09)	Chr.05-Chr.23
71	SWU18371	Chr.05(A05)	SWU18371	Chr.23(D09)	Chr.05-Chr.23
72	SWU18613	Chr.05(A05)	SWU18613	Chr.23(D09)	Chr.05-Chr.23
73	SWU18809	Chr.04(A04)	SWU18809	Chr.23(D09)	Chr.05-Chr.23
74	SWU18959	Chr.06(A06)	SWU18959	Chr.20(D10)	Chr.06-Chr.20
75	SWU19027	Chr.06(A06)	SWU19027	Chr.20(D10)	Chr.06-Chr.20
76	SWU19134	Chr.06(A06)	SWU19134	Chr.20(D10)	Chr.06-Chr.20
77	SWU19231	Chr.06(A06)	SWU19231	Chr.20(D10)	Chr.06-Chr.20
78	SWU19275	Chr.06(A06)	SWU19275	Chr.20(D10)	Chr.06-Chr.20
79	SWU19358	Chr.06(A06)	SWU19358	Chr.20(D10)	Chr.06-Chr.20
80	SWU19370	Chr.06(A06)	SWU19370	Chr.20(D10)	Chr.06-Chr.20
81	SWU19500	Chr.06(A06)	SWU19500	Chr.20(D10)	Chr.06-Chr.20

82	SWU19546	Chr.06(A06)	SWU19546	Chr.20(D10)	Chr.06-Chr.20
83	SWU19601	Chr.06(A06)	SWU19601	Chr.20(D10)	Chr.06-Chr.20
84	SWU19757	Chr.06(A06)	SWU19757	Chr.20(D10)	Chr.06-Chr.20
85	SWU19781	Chr.06(A06)	SWU19781	Chr.20(D10)	Chr.06-Chr.20
86	SWU19857	Chr.10(A10)	SWU19857	Chr.21(D11)	Chr.10-Chr.21
87	SWU20524	Chr.10(A10)	SWU20524	Chr.21(D11)	Chr.10-Chr.21
88	SWU20398	Chr.10(A10)	SWU20398	Chr.21(D11)	Chr.10-Chr.21
89	SWU20572	Chr.10(A10)	SWU20572	Chr.21(D11)	Chr.10-Chr.21
90	SWU20682	Chr.10(A10)	SWU20682	Chr.21(D11)	Chr.10-Chr.21
91	SWU20706	Chr.10(A10)	SWU20706	Chr.21(D11)	Chr.10-Chr.21
92	SWU20815	Chr.04(A04)	SWU20815	Chr.26(D12)	Chr.04-Chr.26
93	SWU20924	Chr.04(A04)	SWU20924	Chr.26(D12)	Chr.04-Chr.26
94	SWU21068	Chr.04(A04)	SWU21068	Chr.26(D12)	Chr.04-Chr.26
95	SWU21223	Chr.04(A04)	SWU21223	Chr.26(D12)	Chr.04-Chr.26
96	SWU21236	Scaff	SWU21236	Chr.26(D12)	Scaff-Chr.26
97	SWU21417	Chr.04(A04)	SWU21417	Chr.26(D12)	Chr.04-Chr.26
98	SWU21509	Chr.04(A04)	SWU21509	Chr.26(D12)	Chr.04-Chr.26
99	SWU21526	Chr.04(A04)	SWU21526	Chr.26(D12)	Chr.04-Chr.26
100	SWU21808	Chr.13(A13)	SWU21808	Chr.18(D13)	Chr.13-Chr.18
101	SWU21917	Chr.13(A13)	SWU21917	Chr.18(D13)	Chr.13-Chr.18
102	SWU21958	Chr.13(A13)	SWU21958	Chr.18(D13)	Chr.13-Chr.18
103	SWU22338	Chr.13(A13)	SWU22338	Chr.18(D13)	Chr.13-Chr.18
104	SWU22509	Chr.13(A13)	SWU22509	Chr.18(D13)	Chr.13-Chr.18
105	SWU22525	Chr.13(A13)	SWU22525	Chr.18(D13)	Chr.13-Chr.18
106	SWU22538	Chr.13(A13)	SWU22538	Chr.18(D13)	Chr.13-Chr.18

Gene identification and amino acid dissection

A total of 21,035 genes were identified employing BLASTx and Cotton Functional Genome database. The highest number of genes was found on Chr.19, accounting for 15.7% of the total whereas, lowest numbers were observed on Chr.13 and Chr.26 simultaneously, both representing only 0.2% (Figure 4). When examining the amino acid composition of the chromosomes, Chr.13 exhibited the lowest levels of Glutamic acid (Glu), Histidine (His), and Lysine (Lys) ranging from 0.00 to 17.74, followed by Chr.02 with levels ranging from 0.00 to 8.01, and Chr.06 with levels ranging from 0.00 to 17.97 (Figure 5). On contrary, Chr.19 showed the highest accumulation of amino acids, with a count of approximately 31,783.92 (Supplementary File 1 and 2). The enrichment analysis revealed variations in the percentage of genes involved in different processes across certain chromosomes, including Chr.02, Chr.05, Chr.10, Chr.14, Chr.15, Chr.17, Chr.19, Chr.20, and Chr.22. For instance, Chr.19 had a percentage of 0.01 for genes involved in tRNA 3'-trailer cleavage whereas, Chr.22 had a higher 0.38% of genes associated with defense response. Moreover, Chr.14 showed 0.20 % for genes involved in the response to biotic stimulus, and several other chromosomes had 0.01% for genes related to macromolecule modification and peptide deformylase activity (Figure 6).

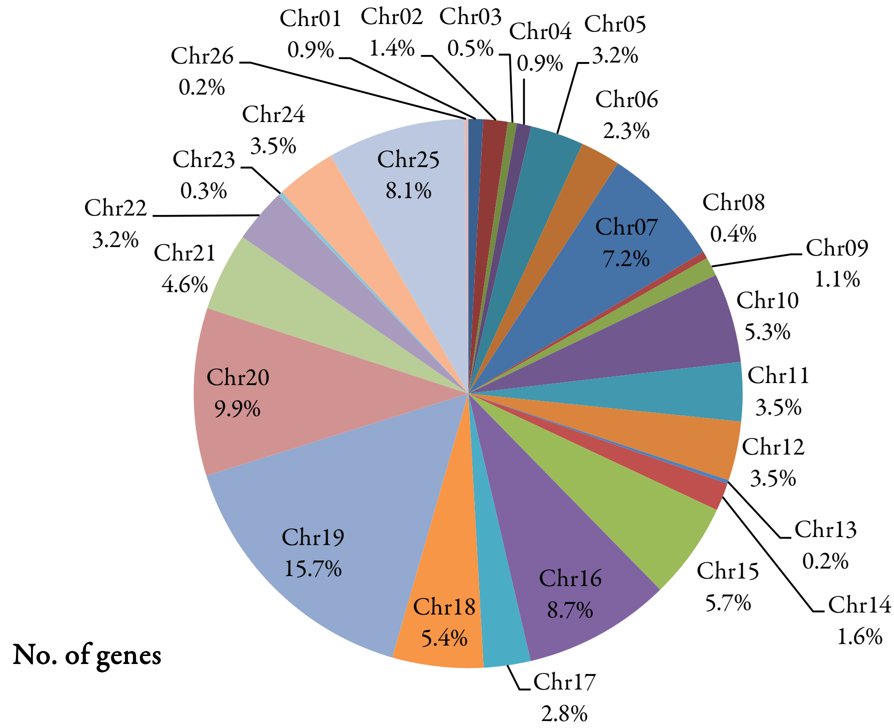


Figure 4. Percentage of chromosome wise gene distribution in the genetic map constructed between *G. barbadense* and *G. darwinii*

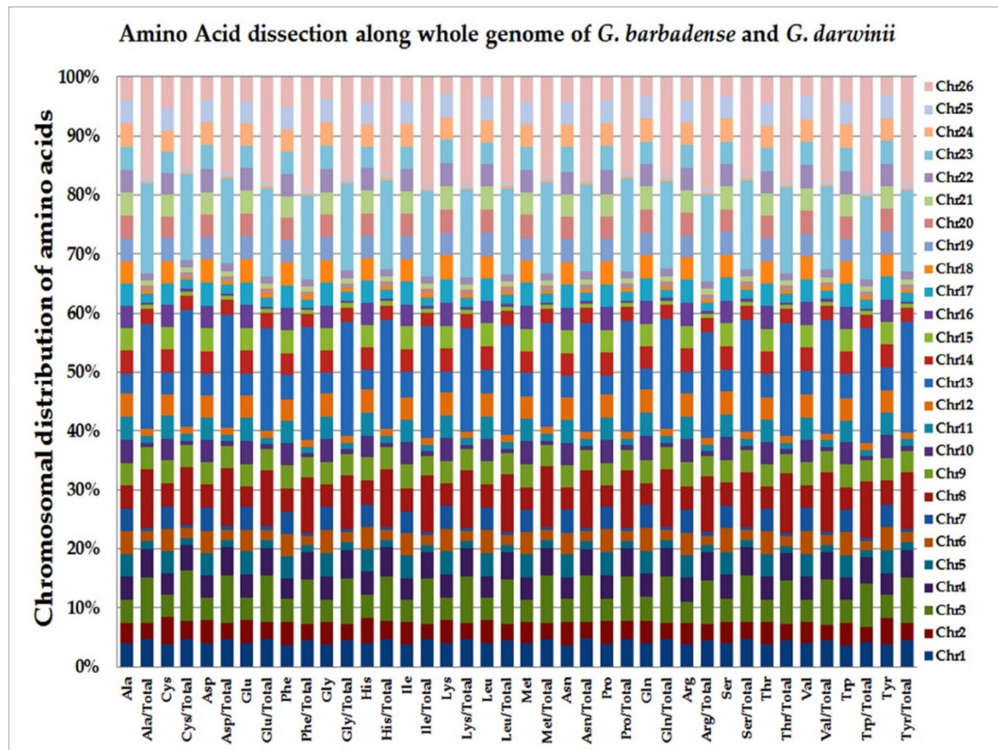


Figure 5. Dissection of amino acids average values with chromosomal distribution along whole genome of *G. barbadense* (AD2) and *G. darwinii* (AD5)

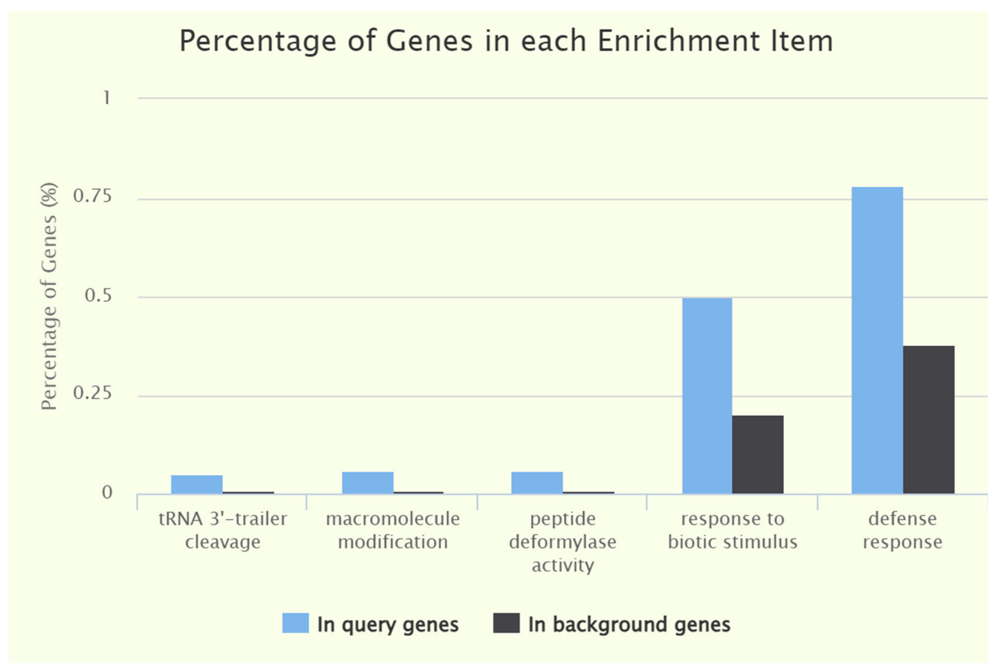


Figure 6. Gene ontology of 10248 genes distributed on chromosome 2, 5, 10, 14, 15, 17, 19, 20 and 22 through enrichment analysis at significance level 0.0001

Discussion

Choice of mapping population

To construct a reliable genetic map, it is crucial to develop an appropriate mapping population. Researchers have recently utilized various types and sizes of populations for developing genetic map further utilized for gene discovery. In addition, a wide range of markers and software packages have been employed for genetic analysis and genotyping of plants. Analyzing the factors that influence marker distances and recombination frequencies during crossing may provide insights into their impact on the quality, efficiency, and reliability of any genetic maps. Furthermore, these factors play a vital role in determining the success of the mapping process and the overall accuracy of results obtained from any genetic maps (Khan *et al*., 2016). Developing diverse maps using different populations derived from the same plant material has been shown in previous studies i.e., Paterson *et al*. (2000) resulting in variations in mapping efficiency and reliability. Therefore, it is crucial to carefully plan the following steps to ensure the development of an appropriate population.

Parents used for developing population

In present investigation, the male parent selected was the wild cotton *G. darwinii* ('*darwinii* 5-7'), known for its high tolerance to abiotic stress (such as drought and salt) and fine fiber characteristics with a short length however, the female parent, the tetraploid *G. barbadense* ('XH-18') was chosen based on its extra-long fiber traits.

These two cotton varieties were morphologically and genetically closely related however, exhibit significant divergence for their fiber characteristics. The primary objective of selecting *G. darwinii was* to incorporate beneficial alien genes into the cultivated *G. barbadense* genetic pool and to enhance the germplasm significance. Previous studies employing molecular assisted selection (MAS) have highlighted the limited

genetic diversity and narrow genetic base of cultivated cotton, resulting low levels of genetic variation which can be utilized for future breeding and crop improvement (Razzaq *et al*., 2021). Therefore, utilizing traits from wild genetic resources becomes crucial in developing comprehensive gene libraries for genomic research. A recent investigation by Chen *et al*. (2015) successfully constructed a dense genetic map between *G. hirsutum* and *G. darwinii*, serving as a significant advancement in comprehending the naturally occurring variation present in *G. darwinii*.

Therefore, it is imperative to develop molecular breeding tools for facilitating the effective utilization of wild genetic resources. Currently, there is insufficient evidence to establish a clear relationship between cultivated cotton species like *G. barbadense* and their wild relatives i.e., *G. darwinii*. Therefore, it is crucial to establish a genetic linkage map between *G. darwinii* and *G. barbadense*. To ensure the purity of both parents, a controlled selfing process was performed on '*darwinii* 5-7' for several generations, and *G. barbadense* was meticulously cultivated under close observation at the World Cotton Germplasm Nursery in Sanya, Hainan, China, before constructing F_2 population.

Development of mapping population

Several populations have been employed to investigate the genetics of cotton however, the F_2 population holds significant advantages due to its ease of development within a relatively short period of time (Bolek, 2016; Khan *et al*., 2016; Chandnani *et al*., 2018). The tetraploid nature of *G. darwinii,* a wild cotton species, makes it convenient to establish an F_2 population. In previous studies, researchers predominantly employed interspecific crosses between upland cotton and *G. barbadense* to generate high-density genetic linkage maps, primarily utilizing the BC_1 population. In this study, the F_2 population was selected due to its advantages over the BC_1 population. The F_2 population offers a greater number of segregation and dominant markers, which are available for both the maternal and paternal parents on the genetic map.

Populations' size

Theprevious studies on genetic mapping have employed population with size ranging from 50 to 250 samples. The choice of population size is determined by the population type and is used to calculate the recombination ratio. Additionally, population size plays a role in determining the density of the map by influencing the number of distinguishing loci. When establishing QTLs, the population size can be estimated based on the marker type and the type of population being studied. Ferreira *et al*. (2006) proposed that for the construction of precise genetic maps in plant species, a minimum population size of 200 individuals is required. According to their findings, larger population sizes significantly contribute for accuracy and highly precise mapping results. The careful consideration of population size is essential for the construction of a high-quality genetic map (Ronin *et al.*, 2015). In this study, a population comprising of 233 F_2 individuals was used to generate a genetic map.

Molecular markers selection

The scrutiny of molecular markers is imperial in research and specific objectives of the study. Molecular markers are the best choice in recent genomic research because they are neutral and not affected by the environment. In genomic mapping and MAS, SSRs are the most widely employed markers. Because SSRs are co-dominant markers, they can be used to identify heterozygotes in a mapping population, delivering more genetic information during construction of a genetic map. The SSRs are widespread, particularly in eukaryotes scattered at randomly across the genome. The current investigation involved examining almost all available SSRs (18201) for polymorphism between the two parents, resulting in a polymorphism rate of 14.55 percent.

Segregation distortion

Segregation distortion, a widespread phenomenon in plants, refers to the deviation of genotypic frequencies from the expected Mendelian segregation ratios. Accurately assessing these deviations is a complex phenomenon (Li *et al*., 2010), and segregation distortion is observed in both intra- and interspecific crosses, exerting an influence on species evolution (Taylor and Ingvarsson, 2003; Zhao *et al*., 2012). The use of morphological markers by Mangelsdorf and Jones (1926) led to the first report of segregation distortion in maize. Subsequently, Khan *et al*. (2016) and Chen *et al*. (2015) reported the occurrence of segregation distortion in F_2 populations obtained from crossing wild genetic sources with upland cotton. The segregation distortion is influenced by various factors, including pollen tube competition, pollen killer genes, chromosome translocation, selective abortion and fertilization (Zhu *et al*., 2007).

Segregation distortion poses a challenge in the construction of genetic maps and the identification of QTLs. In this study, a higher rate of segregation distortion was observed in 257 loci, accounting for 43.94 % of the 613 mapped loci, surpassing the percentage reported by Chen *et al*. (2015) and Khan *et al*. (2016). The high rate of segregation distortion observed in this study may be attributed to the use of *G. darwinii* and *G. tomentosum* as male parents and *G. darwinii* as female parent. Despite their morphological similarity, these species exhibit greater genetic divergence. As a consequence, an increased frequency of chromosomal rearrangements, translocations, and other structural modifications within the genome is expected, leading to notable segregation distortion. Recent scientific literature has suggested that a higher degree of genetic relatedness between parents enhances the probability of segregation distortion (Fu *et al*., 2020).

The number of loci exhibiting segregation distortion was found to be higher in the Dt sub-genome (212 loci) compared to the At sub-genome (45 loci). This observation aligns with the findings reported by Khan *et al*. (2016) observing greater number of distorted loci in the Dt sub-genome. Notably, a majority of the skewed loci in both sub-genomes exhibited bias towards heterozygous alleles. The segregation distortion regions (SDRs) accounted for 19.84% of all distorted loci, with a total of 51 SDRs distributed across the genome, predominantly in the Dt sub-genome. Interestingly, within each SDR, all distorted loci showed segregation in the same direction, confirming previous reports by Yu *et al*. (2011). According to Zhao *et al*. (2012), all large SDRs exhibited distortion toward the heterozygous allele.

Gene identification and amino acid dissection

The identified genes play diverse roles in various molecular and biological functions, including tRNA 3'-trailer cleavage, macromolecule modification, peptide deformylase activity, response to biotic stimulus, and defense response. Kirungu *et al*. (2019) and Ditta *et al*. (2018) have conducted comparable research to identify gene functions. Based these studies tRNA identification and candidate gene analysis, it was observed that green plants harbor a diverse set of tRNase Z enzymes, which are expected to reside in various subcellular compartments. Additionally, research indicated that numerous chloroplast tRNA genes in flowering plants encode incomplete CCA sequences (Fan *et al*., 2011). Proteins are known to have a significant impact on the tolerance of plants towards biotic and abiotic stress, as indicated by amino acid analysis. In the case of *Vicia faba*, the presence of exogenous proline inhibited stomatal opening, while other amino acids such as histidine, methionine, aspartic acid, glutamic acid, asparagine, and glutamine were found to increase it. The addition of exogenous proline resulted in a decrease in the permeability of the Vinca petal membrane to H^+ and OH^- ions. Similar findings were observed when arginine, asparagine, glutamine, alanine, and leucine were used (Rai and Kumari, 1983).

Conclusions

Understanding the genetic architecture of plants is essential for enhancing crop yield. By gaining comprehensive insights into the genetic architecture of cotton plants, we can identify favorable alleles and functional genes to breed superior cultivars. In this study, we constructed the first genome-wide SSR interspecific genetic map between *G. darwinii* and *G. barbadense* using D-genome sequenced primers SWU, named after South West University of China. This genetic map provides valuable information about the structure of *G. darwinii*'s genome, facilitating further research not only in cotton but also in other species. Furthermore, constructing backcross populations allows the collection of field data for agronomic traits and fiber characteristics which can be further utilized for estimation of QTLs, be fine-mapped and subsequently map-based cloning. By employing marker-assisted selection, chromosome substitution lines can be generated to improve the germplasm for better understanding of the organization of the *G. darwinii* genome.

Authors' Contributions

The tests were designed by K.W., A.D., and F.L. The tests were designed by A.D. and Z.Z., and the findings were analyzed by A.D. and Z.Z. Z.Z. gathered the materials, whereas A.D. performed the lab tests and contributed equally with M.S. All computational analyses were performed by A.D. The mapping tests were carried out by F.L., X.C., X.W., M.S., S.F., Y.X., and Y.H. M.K.R.K., B.W. K.A.A, A.A.M., A.M.A, Y.M.H., and B.Z. provided critical feedback and helped for the revision of the manuscript. The manuscript was written by A.D. S.F. and M.S., and it was corrected by K.W. All authors read and approved the final manuscript.

Ethical approval (for researches involving animals or humans)

Not applicable.

Acknowledgements

The authors extend their appreciation to the Researchers Supporting Project number (RSP-2024R369), King Saud University, Riyadh, Saudi Arabia. The State Key Laboratory of Cotton Biology, Institute of Cotton Research (ICR), Chinese Academy of Agricultural Sciences (CAAS), Anyang, China, provided a foundation for our research. We are forever grateful to the entire research team, friends, and anybody else who has ever contributed.

Funding

The National Natural Science Foundation of China (31621005, 31530053, 31671745) and the National key research and development plan (2016YFD0100306, 2016YFD0100203), (2021YFE0101200) provided financial support for this research program, as did the Pakistan Science Foundation (PSF/CRP/18thProtocol (07)), the Public Sector Development Programme under the Ministry of Planning, Development & Special Initiatives (PSDP Project No. 829), the International Foundation For Science, Sweden, and COMSTECH, Islamabad (IFS-C-6500-1).

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Bolek Y (2016). Insight in the utilization of marker assisted selection in cotton: A review. Molecular Plant Breeding 7(10): 1-17. *https://doi.org/10.5376/mpb.2016.07.0010*

Brubaker CL, Paterson AH, Wendel JF (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 203:184-203. *https://doi.org/10.1139/g98-118*

Campbell BT, Saha S, Percy R, Frelichowski J, Jenkins JN, Park W, … Podolnaya L (2010). Status of the global cotton germplasm resources. Crop Science 50(4):1161-1179. *https://doi.org/10.2135/cropsci2009.09.0551*

Chandnani R, Kim C, Guo H, Shehzad T, Wallace JG, He D, … Paterson AH (2018). Genetic analysis of *Gossypium* fiber quality traits in reciprocal advanced backcross populations. The Plant Genome 11(1):170057. *https://doi.org/10.3835/plantgenome2017.06.0057*

Chen H, Khan MKR, Zhou Z, Wang X, Cai X, Ilyas MK, … Wang K (2015). A high-density SSR genetic map constructed from a F2 population of *Gossypium hirsutum* and *Gossypium darwinii*. Gene 574(2):273-286. *https://doi.org/10.1016/j.gene.2015.08.022*

Ditta A, Zhou Z, Cai X, Shehzad M, Wang X, Okubazghi KW, … Liu F (2018). Genome-wide mining and characterization of SSR markers for gene mapping and gene diversity in *Gossypium barbadense* L. and *Gossypium darwinii* G. watt accessions. Agronomy 8(9):181. *https://doi.org/10.3390/agronomy8090181*

Gallagher JP, Grover, C. E., Rex, K., Moran, M., & Wendel, J. F. (2017). A new species of cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany, 42(1), 115-123. *https://doi.org/10.1600/036364417X694593*

Grover CE, Zhu X, Grupp KK, Jareczek JJ, Gallagher JP, Szadkowski E, … Wendel JF (2015). Molecular confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum* Wittmack. Genetic Resources and Crop Evolution 62:103-114. *https://doi.org/10.1007/s10722-014-0138-x*

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

Khan MK, Chen H, Zhou Z, Ilyas MK, Wang X, Cai X, … Wang K (2016). Genome wide SSR high density genetic map construction from an interspecific cross of *Gossypium hirsutum*× *Gossypium tomentosum*. Frontiers in Plant Science 7:436. *https://doi.org/10.3389/fpls.2016.00436*

Kirungu JN, Magwanga RO, Shiraku ML, Lu P, Mehari TG, Xu Y, … Liu F (2020). 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. Journal of Cotton Research 3:1-19. *https://doi.org/10.1186/s42397-020-00072-2*

Liu F, Zhou ZL, Wang CY, Wang YH, Cai XY, Wang XX, … Zhang ZS (2015). Collinearity analysis of allotetraploid *Gossypium tomentosum* and *Gossypium darwinii*. BioRxiv 031104. *https://doi.org/10.4238/gmr.15038391*

Reinisch AJ, Dong JM, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994). A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829-847. *https://doi.org/10.1093/genetics/138.3.829*

Ronin Y, Minkov D, Mester D, Akhunov E, Korol A (2015). Building ultra-dense genetic maps in the presence of genotyping errors and missing data. In: Advances in Wheat Genetics: From Genome to Field. *https://doi.org/10.1007/978-4-431-55675-6_14*

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Kantety RV (2003). Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355-364. *https://doi.org/10.1023/A:1023077209170*

Stam P (1993). Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant Journal 3:739-744. *https://doi.org/10.1111/j.1365-313X.1993.00739.x*

Wang H, Jin X, Zhang B, Shen C, Lin Z (2015a). Enrichment of an intraspecific genetic map of upland cotton by developing markers using parental RAD sequencing. DNA Research 22:147-160. *https://doi.org/10.1093/dnares/dsu047*

Wang S, Chen J, Zhang W, Hu Y, Chang L, Fang L, ... Zhang T (2015). Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. Genome Biology 16:1-18. *https://doi.org/10.1186/s13059-015-0678-1*

Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM (2009). Evolution and natural history of the *Cotton* genus. In: Genetics and Genomics of Cotton. *https://doi.org/10.1007/978-0-387-70810-2_1*

Wendel JF, Brubaker CL, Percival AE (1992). Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. American Journal of Botany 79(11):1291-1310. *https://doi.org/10.1002/j.1537-2197.1992.tb13734.x*

Wendel JF, Cronn RC (2003). Polyploidy and the evolutionary history of cotton. Advances in Agronomy. *https://doi.org/10.1016/S0065-2113%2802%2978004-8*

Wendel JF, Grover CE (2015). Taxonomy and evolution of the cotton genus, *Gossypium*. In: Cotton. *https://doi.org/10.2134/agronmonogr57.2013.0020*

Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, ... Jones DC (2012). Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. Euphytica 187:203-213. *https://doi.org/10.1007/s10681-012-0643-y*

Yu Y, Yuan D, Liang S, Li X, Wang X, Lin Z, Zhang X (2011). 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 12:1-14. *https://doi.org/10.1186/1471-2164-12-15*

Zhang J, Guo W, Zhang T (2002). Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L. x *Gossypium barbadense* L.) with a haploid population. Theoretical and Applied Genetics 105:1166-1174. *https://doi.org/10.1007/s00122-002-1100-4*

Zhang J, Stewart JMD (2000). Economical and rapid method for extracting cotton genomic DNA. Journal of Cotton Science 4(3):193-201.

The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.

License - Articles published in ***Notulae Botanicae Horti Agrobotanici Cluj-Napoca*** are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.
© Articles by the authors; Licensee UASVM and SHST, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.

Notes:

- Material disclaimer: The authors are fully responsible for their work and they hold sole responsibility for the articles published in the journal.
- Maps and affiliations: The publisher stay neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Responsibilities: The editors, editorial board and publisher do not assume any responsibility for the article's contents and for the authors' views expressed in their contributions. The statements and opinions published represent the views of the authors or persons to whom they are credited. Publication of research information does not constitute a recommendation or endorsement of products involved.