Araştırma Makalesi https://doi.org/10.46810/tdfd.995786 Research Article

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Genetic Diversity Analysis of Some Upland Cotton (*Gossypium hirsutum* L.) Genotypes Using SSR Markers

Sadettin ÇELİK[1*]

[1]Bingöl University, Genç Vocational School, Department of Forestry, Bingöl, Turkey
Sadettin ÇELİK ORCID No: 0000-0002-0588-1391

**Corresponding author: sadettincelik@bingol.edu.tr*

(Received: 15.09.2021, Accepted: 30.11.2021, Online Publication: 25.03.2022)

Keywords

Genetic diversity, MAS, PIC, SSR, Upland cotton

Abstract: Cotton plant is an important crop cultivated under biotic and abiotic stress conditions worldwide. The best way to avoid the harmful effects of chemicals used to combat these stresses is to develop tolerant or resistant varieties in plant breeding programs. In the present study, some of Upland cotton varieties were screened with 20 polymorphic SSR primers, and their population structure and genetic diversity analysis were examined. 17 SSR primers amplified 99 alleles with a 5.82 allele per locus. The mean PIC value of the markers was 0.312. The highest PIC value (0.491) belongs to the Nau3736 SSR marker while Bnl1611 and Bnl3449 markers had the lowest PIC value (0.105). The Genetic Distance (GD) values of the markers varied between 0.26 and 1.09. The highest GD values were between Sure Grow 96 and Carmen, Sealand-542 and Siokra ¼, and between Sphinx V and Stoneville-453 cultivars. As a result, the genetically distant cultivars (Acala maxxa, Carmen, Aleppo 40, Siokra ¼, and Tex) can be recommended to use as parents in Marker-assisted selection (MAS) technology to develop new cotton varieties which are resistant or tolerant to stress factors.

SSR Markörleri Kullanılarak Bazı Upland Pamuk *(Gossypium hirsutum* L.) Genotiplerinin Genetik Çeşitlilik Analizlerinin Yapılması

Anahtar Kelimeler

Genetik çeşitlilik, MAS, PIC, SSR, Upland pamuk

Öz: Pamuk bitkisi dünya genelinde biyotik ve abiyotik stres koşullarında yetiştirilen çok önemli bir tarla bitkisidir. Bu stres koşullarıyla mücadelede kimyasal ilaçlarının olumsuz etkilerinden kaçınmanın en etkili yolu, bitki ıslah metodlarıyla biyotik ve abiyotik stres faktörlerine dayanıklı/tolerant yeni bitki çeşitlerini geliştirmektir. Bu çalışmamızda bazı upland pamuk genotipleri 20 adet polimorfik SSR primerleriyle taranmış, populasyon yapıları ve genetik çeşitlilik analizleri yapılmıştır. 17 SSR primeri lokus başına 5.82 allel olmak üzere totalde 99 allel üretmiştir. Markörlerin ortalama PIC değerleri 0.312 olmuştur. En yüksek PIC değeri (0.491)'ni Nau3736 SSR marköründe elde edilirken, en düşük PIC değeri (0.105) Bnl1611, Bnl3449 markörlerinde elde edilmiştir. Markörlerin genetik mesafe değerleri 0.26 ile 1.09 arasında değişmiştir. En yüksek genetik mesafe Sure Grow 96 ve Carmen, Sealand-542 ve Siokra ¼ ile Sphinx V ve Stoneville-453 arasında ölçülmüştür. Sonuç olarak, genetik olarak birbirinden uzak olan çeşitler (Acala maxxa, Carmen, Aleppo 40, Siokra ¼, and Tex), stres faktörlerine karşı dirençli çeşit geliştirme teknolojisi olan Markör Destekli Seleksiyon (MAS) ıslah programına ebeveyn olarak kullanılabileceği güçlü bir şekilde önerilmektedir.

1. INTRODUCTION

Cotton (*Gossypium spp.*) is an important crop as a source of natural fiber and oil. It covers 35% of the fiber used worldwide. There are about 50 species of cotton, and 45 of these are known to be diploids (2n = 2x = 26), and 5 are allotetraploid (2n = 4x = 52) [1]. The most common diploid ones are *G. arboreum* L. and *G. herbaceum* L. with AA (2n = 2x = 26) genome group, and the tetraploid ones are *G. hirsutum* L. and *G. barbadense* L. with AADD (2n = 4x = 52) genome group [2].

Approximately there are 20 diseases, pests, and harmful stress factors that reduce the yield, quality, and restrict cotton cultivation worldwide. Particularly, more than 41% of the yield loss is due to biotic factors such as insects (15%), weeds (13%), and other harmful pathogens [3]. To

combat such stress factors, Cook. [4]. has proposed four possibilities: (i) breeding new tolerant or resistant varieties, (ii) developing a healthy root environment, (iii), increasing the quality of water, and (iv) protecting the plants from airborne threats. Developing tolerant/resistant genotypes against biotic and abiotic stress conditions are more environmentally safe compared to chemicals that pollute the atmosphere.

Corresponding to recent climate changes, the identification of desired alleles at QTL underlying tolerance/resistance to biotic and abiotic stresses is a primary breeding strategy for improving crop productivity and production under stress conditions [2;5]. The strategy involves germplasm screening, QTL mapping, and the development of DNA markers linked to QTL. Marker-assisted selection (MAS) in breeding programs has been applied to introduce the desired alleles into the genetic backgrounds of elite varieties [6].

Fundamental studies on quantitative traits associated with stress tolerance/resistance are necessarily required to apply marker-assisted selection to practical crop breeding. On the other hand, diversity studies may contribute to screening desired genotypes as parents for crossing but cannot directly contribute to breeding through marker-assisted selection. When a broad range of diverse germplasm and genome-wide DNA markers are used for the analysis, genome-wide association mapping using genotypic and phenotypic data is applicable to identify chromosomal regions involving QTL/genes conferring target traits [7; 8; 5; 9].

Genetic diversity comes from the allelic variation in the genome (insertion, deletions in DNA) and constitutes the basis of Marker-assisted breeding. Genetic diversity and the selection of parents play important roles in terms of the level of variation, and heterosis (hybrid vigor), hybrid strength, molecular breeding, and in obtaining cotton genotypes tolerant or resistant to diseases and pests, early with high efficiency. To achieve this, molecular markers that measure the genetic diversity at the DNA level must be employed [5; 9]. DNA markers can be classified as sequence-based (SNP, Single Nukliotit polymorphism), hybridization-based (RFLP, Restriction Fragment Lenght Polymorphism), and PCR-based markers (RAPD, Randomly Amplified Polymorphic DNA; AFLP, Amplified Fragment Lenght Polymorphism; SSR, Simple Sequences Repeats) [10]. Genetic markers are used for predicting the genetic diversity in wild and designed populations [11; 12; 13], Quantitative Trait Loci (QTL) and association mapping [7; 8; 14; 15; 16], pedigree analyses [17], heterotic group classification [18; 19], and for the protection of genotype rights [20].

The Simple Sequence Repeat (SSR) (Microsatellites or Short tandem repeat) molecular markers were used in the genetic diversity analyses because they have a high ability to show genetic differences between cotton genotypes, they are present in all eukaryotic cells, show uniform distribution throughout the genome, provide the opportunity to determine genetic diversity, are repeatable, allow working on low DNA samples amount, are cheap and co-dominant, and give reliable results. SSR markers are in 1-4 to 1-6 nucleotide length [21; 22; 23; 24; 25]. Bertini et al. [26] reported that they obtained 66 alleles in total, with an average of 2.13 alleles per locus in her genetic diversity with 53 cotton varieties using 31 SSR markers. Liu et al. [27] reported that they obtained 165 polymorphic DNA fragments in a study using SSR markers on Asian cotton (*Gossypium arboreum* L.), and the genetic similarities of the accessions were between 0.58 and 0.87. The development of new varieties against biotic and abiotic stress factors increased the quality, germination, emergence, seedling, growing, and yield of cotton [28; 29; 30; 31].

Some important parameters must be known about the individuals that will be used as parents in developing new varieties with the breeding method. Genetic differences and the degree of these differences among the genetic materials that will be used in the breeding program have critical importance in variety breeding. Islam et al. [32] reported that a high rate of stability was observed in yield in some regions because of the use of germplasm sources as breeding materials, where genetic diversity is very small. Therefore, genotypes to be used in plant breeding should be analyzed to determine their kinship degrees and those have close kinship degree is not recommended to use in plant breeding program.

The present study was conducted to determine the genetic diversity between upland cotton (*Gossypium hirsutum* L.) germplasm and examine population structure using 20 SSR markers.

2. MATERIALS AND METHOD

2.1. Plant Materials

A total of seventeen upland (*G. hirsutum* L.) cotton genotypes belong to the AD_1 genome group collected from different countries (The USA, Syria, Turkey, Australia, Albania) used for genetic diversity analysis, in Kahramanmaras Sutcu Imam University, Faculty of Agriculture, Department of Agricultural Biotechnology, in Kahramanmaraş City-Turkey.

2.2. DNA Extraction, and Visualization Genomic DNA

From the young leaves, 0.5 g amount picked up at the 3-4-leaf stage, from each genotype with sterile scissors. The samples were then washed with distilled water (dH_2O) and ethanol and placed in plastic tissue bags in thermally insulated containers with -80ºC dry ice during transport to the laboratory and were kept in the -80ºC freezer until DNA isolation. The Genomic DNA isolation protocol Centyltrimeyhtlaminiumbromide (CTAB) was developed by Doyle and Doyle. [33] was modified and used in this study.

2.3. SSR Amplification

Amplification with PCR and gel electrophoresis stage was performed according to Zhang and Stewart. [34]. The PCR protocol consisted of incubating at 94ºC for 5 min, then 34

cycles of at 94ºC for 1 min, at 60ºC for 1 min and 72ºC for 2 min; later at 72ºC for 7 min. A reaction volume of 15 µL was used for each PCR process. In PCR reaction mixture, there were 0.75 µL dNTP (Conc.10 mM), 1.5 µL 10X PCR buffer, 1 µL forward primer, 1 µL reverse primer, 0.5 µL Taq DNA polymerase enzyme (Conc.5 µL, 2 µL template DNA (Conc. 25ng/µL), 8.25 µL ddH_2O (double-distilled) components [35].

Table 1. Information about Simple sequence repeat (SSR) used in this study

Primer Name	Species, Germplasm	Chromosome	Repeat motif	Publication
BNL1079	*G. hirsutum* L, Deltapine 90	AD_chr.18	(CA)11 (GT)11	Mei et al. [36].
BNL1604	*G. hirsutum* L, Deltapine 90	AD_chr.7-16	(AG)25	Mei et al. [36].
BNL3449	*G. hirsutum* L, Deltapine 90	AD_chr.18	(AC)15, (TC)6T (AC)15G(CA)2	Yu et al. [37].
BNL3479	*G. hirsutum* L, Deltapine 90	AD_chr.18	(AC)15(TC)6T(AC)15 G(CA)2	Yu et al. [37].
BNL2571	*G. hirsutum* L, Deltapine 90	AD_chr. 13-18	(AG)13, (TC)13	Yu et al. [37].
BNL1611	*G. hirsutum* L, Deltapine 90	AD_chr.19	(AG)12	Yu et al. [37].
CIR 0141	*G. hirsutum* L Guazuncho-2	AD_chr.07-16	(TG)30	Nguyen et al.[38].
CIR0253	*G. hirsutum* L, Guazuncho-2	AD_chr.5-19-22	(TC)15, (N)8(A C)5(N)7(CA)8	Nguyen et al. [38].
CIR0099	*G. hirsutum* L, Guazuncho-2	AD_chr.18-25	(GT)8	Nguyen et al. [38].
NAU3736	*Gossypium raimondii*	AD_chr.01-15	(GTA)6	Guo et al. [39].
NAU2714	*Gossypium raimondii*	AD_chr.6-25	(TTA)7	Guo et al. [39].
NAU4024	*Gossypium raimondii*	AD_chr.14	(GTC)6	Guo et al. [39].
NAU2761	*Gossypium raimondii*	AD_chr.02-17	(TTAA)6	Guo et al. [39].
NAU2173	*G. hirsutum* L, Xuzhou 142	AD_chr.14	AAG (17)	Han et al. [40].
JESPR0065	*G. hirsutum* L, Tamcot Sphinx	AD_chr.04-12	(GAA)25	Reddy et al. [41].
JESPR0153	*G. hirsutum* L, Tamcot Sphinx	AD_chr.13,18,20	(CTA)18	Reddy et al. [41].
JESPR0114	*G. hirsutum* L, Tamcot Sphinx	AD_chr.09-23	(GT)12	Reddy et al. [41].
MGHES16	*G. hirsutum* L, Acala Maxxa	AD_chr.11-21	(CT)10, (TCT)4	Qureshi et al. [42].
MGHES31	*G. hirsutum* L, Acala Maxxa	AD_chr.12-26	(CAT)9	Qureshi et al. [42].
NAU2251	*G. hirsutum* L, Xuzhou 142	AD_chr.12-26	AGA (8)	Han et al. [40].

In table 1., it is indicated the SSR markers properties were obtained from the varieties of *G. hirsutum* L. and *Gossypium raimondii* cotton species. Bnl1079, Bnl1604, Bnl3449, Bnl3479, Bnl2571, Bnl1611 SSR markers were obtained from *G. hirsutum* L. species's Deltapine 90 variety; Cır0141, Cır0253, Cır0099 from Guazuncho-2; Nau2173, Xuzhou 142, Jespr0065, Jespr00654, JesprR00654 from Tamcot Sphinx; Mghes16 and Mghes31 were obtained from Acala Maxxa variety. Nau3736, Nau2714, Nau4024, Nau2761, Nau2173 SSR markers were obtained from the cotton species, *Gossypium Raimond*. Besides, in table 1, the references, repeat motifs, and chromosomes that SSR markers used in this study were also displayed [43].

2.4. Genetic diversity and phylogenetic analyses

Each SSR marker locus, amplified alleles were scored as either '1' (present) or '0' (absent). Each amplified allele of SSR was indicated with A, B, C and so on… letters. Analysis of genetic distance (GD) to obtain genetic dissimilarity matrix PopGENE 1.32 ver. software was used [44].

Genetic diversity was conducted for all SSR locus across germplasm individuals based on the number of alleles, polymorphic alleles numbers, percent of polymorphism, gene diversity [45], and the polymorphic information content (PCI) [46] was determined with PowrMarker 3.25 version program [47]. Genetic distances (GD) [48] were calculated, and the phylogenetic tree was built with the distance matrix using UPGMA (Unweighted pair group method with arithmetic) data panel provided in PowerMarker 3.25 [47]. The phylogenetic tree was created in MEGA-X 10.1.7 version. PIC values show the number of alleles at each locus and the distinctive features of the markers through the relative frequencies of alleles in the population [49]. PIC calculation was made based on the formula given below.

$$PIC=1-\Sigma(Pi)^2 \quad (1)$$

In equation 1, the P-value is the frequency of the i^{th} allele of the total Upland cotton genotypes subjected to the analysis [50].

2.5. Population Structure Analysis

The genetic structure of the subpopulation (*Q-matrix*) was analyzed with Bayesian Model-Based (MBB) analysis model provided in Structure 2.3.4 version software [51]. To calculate the Q matrix, the software was set up to run under 10,000 (Leng of burning periods)-100,000 (Number of reps after burning) Markov Chain Monte Carlo iteration after the burn-in and Number of the population (K) from 1 to 10 and number of iterations was 5. The average likelihood value L(K) was calculated for each K cross of all the runs. The number of populations is estimated by estimating ΔK [52]. The results were transferred to the web-based "Structure Harvester" (http://taylor0.biology.ucla.edu/structureHarvester/) program.

3. RESULTS and DISCUSSIONS

As a result of the amplification made using 20 polymorphic SSR primers, 99 alleles were produced. There was an average of 5.82 alleles per locus. Zhang et al. [53] and Lacape et al. [54] obtained 5.5 alleles per average locus that ranged between 2 and 26 per locus, which was similar to the results of the present study. Besides, Lacape et al. [54] also reported that the number of the alleles obtained per marker did not only stem from the diversity in germplasm, but also depended on the marker type, the fragment separation technique used, and the resolution.

As given in figure 1, after scanning with 20 SSR markers [43] the cotton varieties belong to *G. hirsutum* L. 17 SSR markers were found as polymorphic. Polymorphic information content (PIC) ranged from 0.0-0.5 with an average of 0.312; Each SSR marker amplified at least 2 loci and Jespr SSR produced the most alleles. Genetic diversity among SSR markers differed from 0.0 to 0.6 (Figure 1).

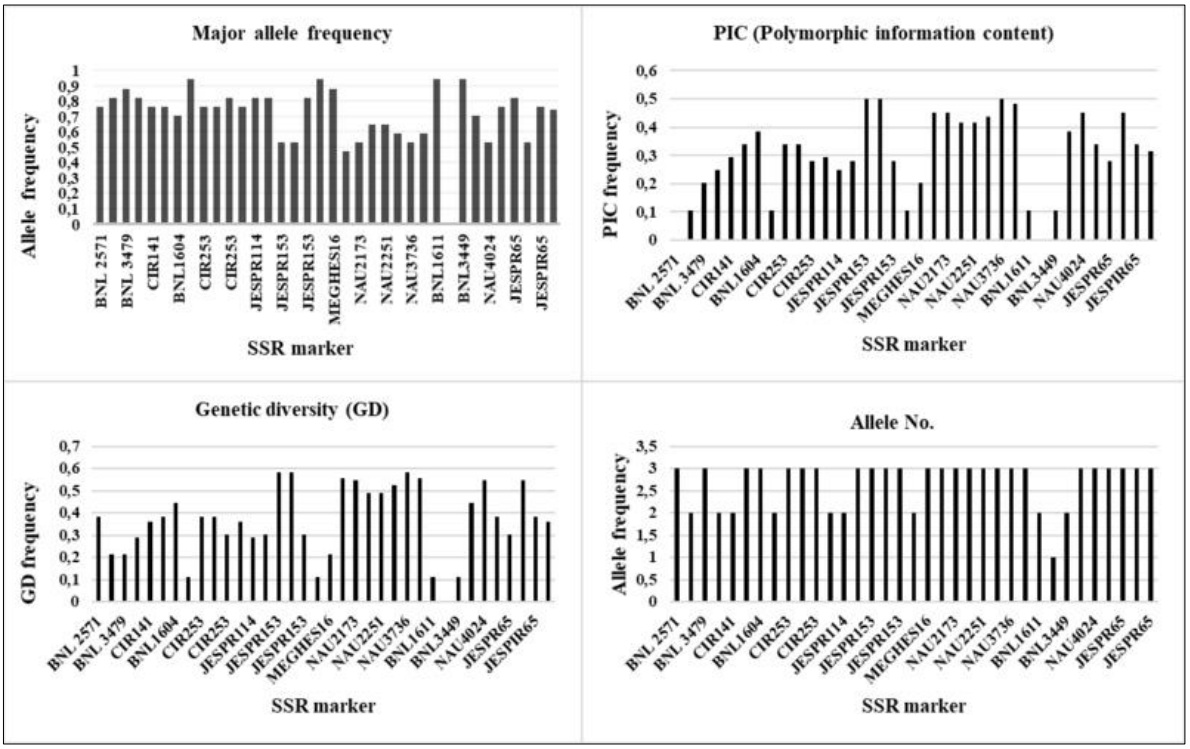

Figure 1. Major allele frequency (MAF), Polymorphic information content (PIC), Genetic Diversity (GD), and Number of Alleles per SSR locus

Table 2. Analysis results of SSR primers used in this study

Primer Name	Forward Primer (5'-3')	Reverse Primer (3'-5')	Alleles Band size (bp)	Allele no.	PIC
BNL1604	AGAGGGAGTAAAGATTTGGGG	TCCAGTTCTTTTTGCCTTGG	120-50	6	0.362
BNL3449	AAGCTGTGGCTATGATGCCT	AGAGCAAAAAACAATTACAAAAGC	180-150	2	0.105
BNL3479	AGTGGGTTGGACTTTCATGC	CACGGGCTTTTTTTTTTTCA	350-200	5	0.216
BNL2571	TCGCTATCGCTCTGAAATCA	ATGCCACGGAATTAGCAAAC	400-320	3	0.345
CIR 0141	CGCACAAGGAATAGAAG	ACCCAACATAAGGACTAAA	300-200	4	0.271
CIR0253	CCAACCAAGAAACCAG	GTAAGCATGGGCATTT	150-50	11	0.256
NAU3736	CATGTGCATTTCATCCTGTC	CCAAGTGAGAGGCATTTTCT	200-80	6	0.491
NAU2714	GCAGCCATTACAGAACATCA	TCATTGATCCATTGCTTCTG	300-200	6	0.384
NAU4024	ACAAGCATCTTCATGGACCT	AGAAGGATGATGCAAAGAGG	200-150	3	0.449
NAU2751	GACAAGTTTTTGGACCCACT	TTCATAGAGGGGTTTTGCAT	350-50	2	0.345
NAU2173	GCCAAATAGGTCACACACAA	AGCGAGAAGGAGACAGAAAA	350-320	6	0.450
JESPR0065	CCACCCAATTTAAGAAGAAATTG	GGTTAGTTGTATTAGGGTCGTTG	300-200	12	0.325
JESPR0153	GATTACCTTCATAGGCCACTG	GAAAACATGAGCATCCTGTG	300-100	12	0.390
JESPR0114	GATTTAAGGTCTTTGATCCG	CAAGGGTTAGTAGGTGTGTATAC	300-100	4	0.248
MGHES16	ACCCCAATACAACCCCATTT	GCAGAGAAAAGGGACAGAGG	400-50	5	0.154
NAU2251	TTCTCCAGTAACCAACAAAGG	AAAATATCATCCCCGTCAAA	400-200	9	0.422
BNL1611	CAATGAACAAAAAATGTAAGGG	TGGGCATTTAGCCATTTACC	100-50	3	0.105
				Mean= 5.82	Mean= 0.312

As indicated in table 2. the average Polymorphic Information Content (PIC) of the markers was 0.312. The PIC values of the markers changed between 0.105 and 0.491. The Nau3736 marker had the highest value with a PIC value of 0.491, and the Bnl1611 and Bnl3449 markers had the lowest value of 0.105. While the mean of alleles is 5.82, each of Jespr0065 and Jespr0153 SSR markers has amplified 12 loci (The highest alleles number) and both Bnl3449 and Nau2751 amplified 2 loci (2 alleles) markers. As a result of the amplification with 31 pairs of SSR primers, Bertini et al. [26] reported similar results with 2.13 alleles per average SSR locus, with PIC values ranging from 0.18-0.62 to 0-0.41 in 66 alleles. Seyoum et al. [35] propounded also close results such as PIC value ranged from 0.319 to 0.019 with 0.279 mean value, the number of alleles per locus ranged from 2 to 12 with 4.53 average.

Liu et al. [55] made amplification with 62 SSR primers and identified 139 alleles in 69 SSR locus as 2 alleles per locus, as well as 325 alleles as 5 alleles per locus; Lacape et al. [54] reported that they identified a total of 1128 alleles including an average of 5.61 alleles per locus because of 201 SSR markers using 47 wild *Gossypium* genotypes. The reason that Lacape et al. [54] had more alleles and higher average alleles per locus than us could be depended on their SSR amount and wild Cotton genotypes. Iqbal et al. [56] reported that they obtained 349 alleles with 50 primers, and detected polymorphism in 23 cotton genotypes using 49 primers. In these studies, it is considered that the reason for producing more alleles than in the present study is related to the excessive SSR primers used.

In figure 2, the Bnl1611 SSR marker produced 3 alleles with 100, 75, 50 bp between 100-50 bp fragment size. The Bnl1611marker amplified the target locus in all genotypes except Gloria and produced at least 2 maximum 3 alleles in genotypes except the Nazilli 84s. The reason for not being able to produce bands in the Gloria genotype may be due to pipetting errors or absences of template DNA.

The gel image taken under UV light after Bnl1611 running through 1% agarose was given in Figure 2. Scoring fragments were performed in all genotypes except the Gloria variety.

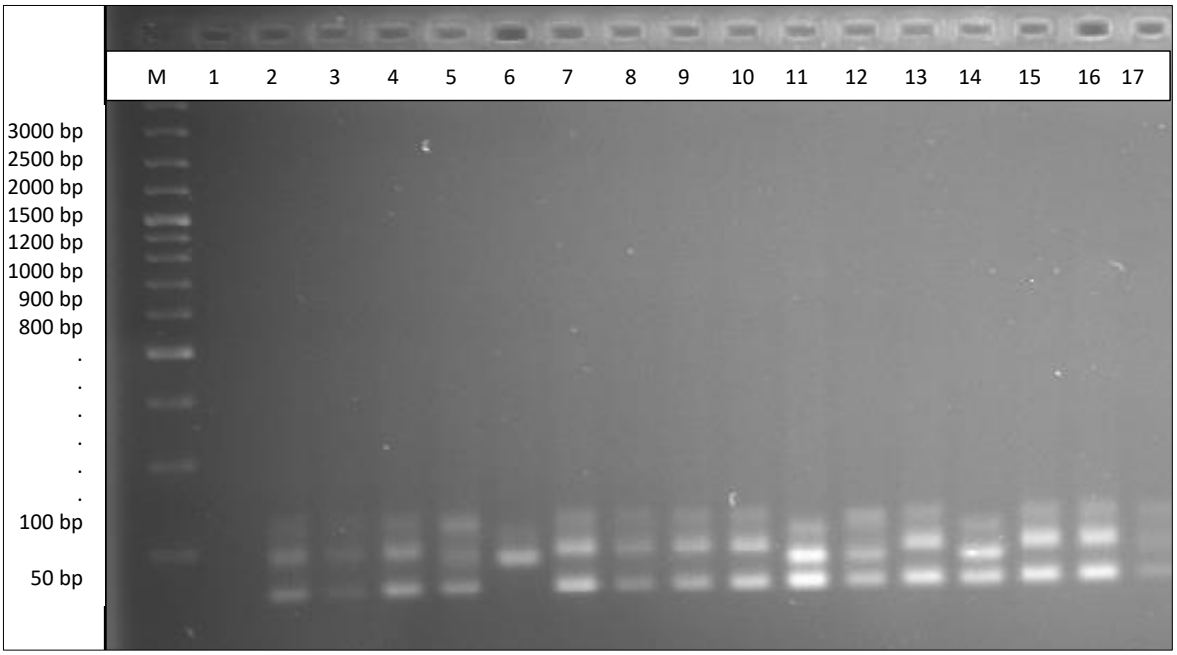

Figure 2. PCR gel image of BNL1611 (Locus) SSR primer. **M**=Marker, DNA ladder= Vivantis 10 bp (1% agarose 1 X TBE). **1**=Gloria, **2**=Aleppo-40, **3**=Carmen, **4**=Cukurova 1518,5=Albania-6172, **6**=Nazilli 84S, **7**=Siokra 1/4, **8**=Sphinx V, **9**=Stoneville-453, **10**=Sure grow 96, **11**=TamcotSP37, **12**=Tex, **13**=Sealand-542, **14**=Acala 1517-95, **15**=Acala Maxxa, **16**=TMS 108/2, **17**=Candia

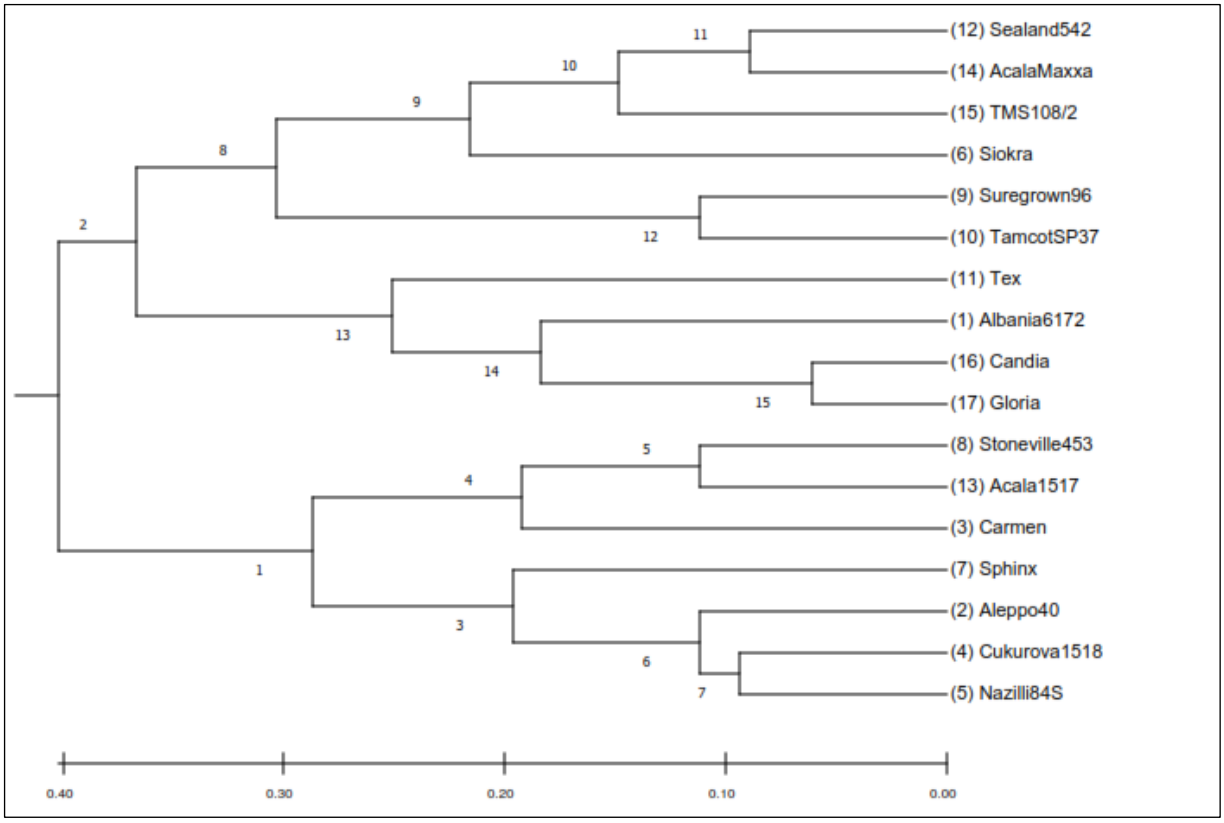

Figure 3. Dendrogram based on Nei's [45] genetic distance: Method = UPGMA of *G. hirsutum* L. genotypes based on 20 Genome-Wide SSR markers

The results of UPGMA (Unweighted pair group method with arithmetic) analysis were shown in Figure 3. The genotypes are divided into two main clusters. 1st main cluster consists of 7 genotypes (Nazilli 84S, Çukurova 1518, Aleppo 40, Sphinx, Carmen, Acala 1517, Stonville 453), the 2nd cluster consists of 10 genotypes (Sealand 542, Acala maxxa, TMS 108/2, Siokra, Sure grow 96, Tamcot SP 37, Tex, Albania 6172, Candia, Gloria). The 1st main cluster is divided into two subclusters and 2nd main cluster is also divided into two subclusters. Accordingly, 15 different subclusters emerged in total. In this respect, Sealand-542, Acala maxxa, TMS 108/2, Siokra ¼, Suregrow96, Tamcot, Tex, Albania-6172, Candia and Gloria were classified in a different cluster, in other words, they have pedigree relations, Stonville-453, Acala-1517, Carmen, Sphinx V, Aleppo-40, Çukurova 1518 and Nazilli 84S genotypes were classified in a different cluster. The first cluster was also divided into further subclusters, and Sealand-542, Acala maxxa, TMS 108/2, Siokra ¼, Sure grow 96, and Tamcot, were included in the first subcluster of this cluster while Albania-6172, Candia, Gloria, and Tex were included in the second subcluster.

Although Stoneville-453, Acala-1517, and Carmen varieties in Group 4 were collected from different geographical areas, there was a unity of origin among them. The Sphinx V, Aleppo-40, Çukurova 1518, and Nazilli 84S were classified in Group 3. Again, it was seen that the second subcluster of the second main cluster also may have common parents even though they were the genotypes obtained from different regions, as was the case in the first subcluster. Eminenur and Hancer. [57] conducted genetic diversity analysis with *G. hirsutum* L. genotypes (Flash, BA119, ST506, Tamcot Sp21, Tamcot 22, Tamcot 94, Tamcot Camd-eS and Tamcot Sp23, Sphinx v, Stn468), and *G. barbadance* L. (Giza 70) variety, and reported that Giza 70 was included in a separate cluster and the remaining genotypes were included in a separate cluster.

In Figure 3, some genotypes classified in the same cluster have some similar characteristics. This may be clarified that they come from a common pedigree. For instance, Sealand-542 which is an interspecies hybrid (*G. hirsutum* L X *G. barbadance* L.), and Acala maxxa were classified in the same cluster. Nazilli 84S, which is a hybrid of Carolina Queen X 153-F and Çukurova 1518 are also among the closest relatives. It was seen that Sure Grow 96 and Tamcot SP37 genotypes, which were clustered in Group 12 that originated from the USA, and Candia and Gloria, two Australian originating varieties, were in Group 15.

SSR markers were distributed on 20 different chromosomes, and produced a total of 99 alleles ranging from 2-12 with a 5.82 average. Tyagi et al. [58] reported similar results in that they produced 546 alleles in 141 loci in genetic diversity study with 381 cotton accessions using 120 Genome-wide SSR markers. Bardak and Bolek. [59] reported similar results with the present study as they obtained a total of 173 alleles, including 3.93 alleles per locus, as a result of their study with 39 SSRs and 5 ISSRs using for screening 25 cotton genotypes.

According to Nei's [45] pair-wise comparison on the genetic distance between genotypes, and according to the

results of the Genetic Distance matrix, the highest GD (1.09) was measured at the highest score between Carmen and Sure Grow 96, Sealand-542 and Siokra ¼, and Sphinx V and Stoneville-453 genotypes. The lowest GD (0.26) was measured between Gloria and Carmen, and Tex and Stoneville-453 genotypes. The GD between the genotypes varied between 0.26 and 1.09, in this respect, it was seen that the GD between Cukurova 1518 and Gloria, Nazilli 84S and Siokra ¼, Tamcot SP37 and Cukurova 1518, TMS 108/2 and Çukurova 1518, Sure Grow 96 and Tex, Acala Maxxa and Tamcot SP37 and Acala-1517 and TMS 108/2 were the same. Bardak and Bolek. [59] also reported close results to this study that the lowest GD (0.04) was detected between Siokra ¼ and Nazilli 84S, and the highest GD (0.58) was between Erşan 92 cotton variety of the AD genome and *G. sturtianum* of the C genome group *Nandewarense* variety.

In the present study, the Genetic Distance (GD) between upland cotton genotypes ranged from 0.26 to 1.09. The previous studies were reported Genetic distance between 0.06-0.34 [60], 0.06-0.38 [61]; 0.82-0.93 [56] and 0.19-0.36, [62] in upland cotton genotypes. These results are on the line with the previous genetic diversity studies.

Zhao et al. [63] have reported the highest genetic diversity in cotton from the USA followed by China; Chen and Du. [64] indicated the higher genetic diversity of introduced genotypes particularly obtained from China and USA countries than the domesticated genotypes.

The Structure analysis of population results revealed the highest population numbers took place at K:5 and this can be evaluated as cotton germplasm can be grouped into five subpopulations (figure 4). Five (5) different colors in Figure 4, and each color represents a population or genotype collected from the same geographical area.

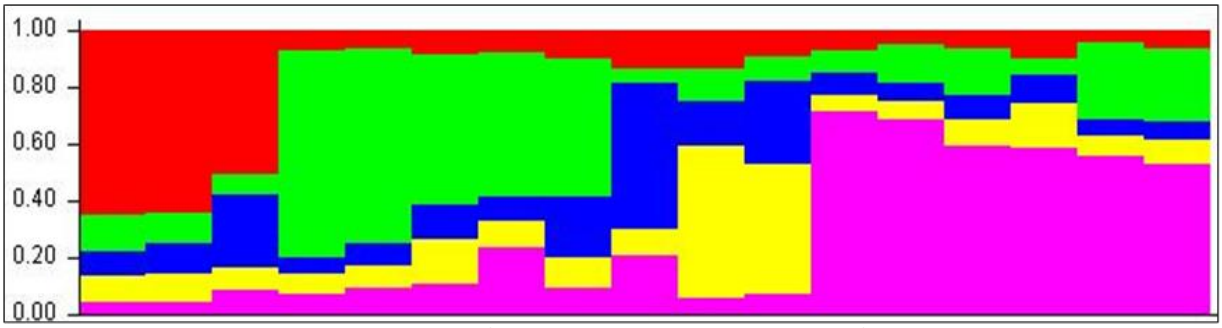

Figure 4. Plot Q-Matrix shows the genotypic data analysis of cotton genotypes in Structure 2.3.3 ver. software. Each colored subsection bar represents the cotton genotypes group origin

These genotypes were obtained from the United States (Acala maxa, Sure Grow 96, Stoneville-453, Tamcot SP 37, Teks, Sealand-542, Acala-1517, Siokra ¼, Sphinx V), Turkey (Çukurova 1518, TMS 108/2, Nazilli 84S), Albania (Albania-6172), Syria (Aleppo-40) and Australia (Carmen, Candia, Gloria). Abdurakhmonov et al. [65] reported that the germplasm materials they used were divided into three categories as native, Mexican, and African origin. Bardak et al. [66] reported that 48 cotton genotypes they used were divided into 3 different groups. Tyagi et al. [58] reported similar results as their 381 cotton genotypes were divided into 5 different groups. Genetically distant genotypes should be selected as parents as much as possible when breeding programs are carried out.

4. CONCLUSIONS

As it is already known, the success of breeding depends on the selection of the right parents, which is the first and most important step in the strength and success of this process. Genetic distance between parents in a breeding program brings the possibility of allelic diversity and improved phenotypic values as well as a higher chance of hybrid vigor.

Therefore, exploring the genetic diversity, in other words, determining the genetic distance between genotypes plays a key role in the development of a new variety.

In the study, most SSR primers produced polymorphic bands. Some of them didn't produce.

The cotton varieties with far distance from each other such as Acala maxxa, Carmen, Aleppo 40, Siokra ¼, and Tex, (onsidering the high yield and quality) are highly rCecommended to use as parents in Marker-assisted selection (MAS) to develop new varieties.

Acknowledgments

I am grateful to Associate Prof. Dr. Adem BARDAK for providing cotton germplasm materials, and Dr. Halil TEKEREK and Osman YİĞİT for the assistance on laboratory analysis in Kahramanmaraş Sutçu Imam University, Agricultural Biotechnology Lab.

REFERENCES

[1] Zhang J, Fang H, Zhou H, Sanogo S, Ma Z. Genetics, Breeding, and Marker-Assisted Selection for Verticillium Wilt Resistance in Cotton. *Crop Science*. 2014; 54(4):1289-1303.

[2] Hui-Fang BS. Development of Molecular Markers and Mapping of Quantitative Trait Locf for Resistance to Verticillium Wilt Disease Using Two Inbred Line Populations in Tetraploid Cotton. [P.H.D Thesis]: New Mexico State University Las Cruces, *New Mexico*; 2013.

[3] Pimentel D. Techniques for reducing pesticide use. Wiley, *Hoboken*; 1997.

[4]. Cook JR. Advances in plant health management in the twentieth century. *Annu Rev Phytopathol; 2000*: (38):95–116.

[5] Jia JZ. Molecular germplasm diagnostics and molecular marker assisted breeding. *Scientia Agricultura* Sinica. 1996; 29(4): 1-10.

[6] Varshney R, Hoisington D, Nayak S, Graner A. Molecular Plant Breeding: Methodology and Achievements. In: Gustafson J., Langridge P., Somers D. (eds) *Plant Genomics*. Methods in Molecular Biology™ (Methods and Protocols). *Humana Press*. 2009; 513.

[7] Young ND. A Cautiously Optimistic Vision for Marker-Assisted Breeding. *Molecular Breeding*. 1999;5: 505-510

[8] Kohel RJ, Yu J, Park YH, Lazo GR. Molecular Mapping and Characterization of Traits Controlling Fiber Quality in Cotton. *Euphytica*. 2001; 121: 163-172.

[9] Xie J, Cai Z, Liu XH, Li FH, Cao HL, Luan YC. Application of biotechnology on evaluation of genetic diversity of germplasm. *CROPS (Supplement*). 1998; 71-76.

[10] Meena KK, Sorty AM, Bitla UM, Choudhary K, Gupta P, Pareek A. Abiotic Stress Responses and Microbe-Mediated Mitigation in Plants: The Omics Strategies. Front. *Plant Scientist*. 2017; 8:172.

[11] Manjarrez-Sandoval P, Carter TE, Webb DM, Burton JW. Heterosis in Soybean and Its Prediction by Genetic Similarity Measures. *Crop Science*. 1997; 37(5):1443-1452.

[12] Tatineni V, Cantrell RG, Davis DD. Genetic Diversity in Elite Cotton Germplasm Determined by Morphological Characteristics and RAPDs. *Crop Science*. 1996; 36(1):186-192.

[13] Wendel JF, Brubaker CL, Percival AE. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *American Journal of Botany*. 1992; 79:1291-1310.

[14] Schuster EW, Kumar S, Sarma SE, Willers JL, Milliken GA. Infrastructure for Data-Driven Agriculture: Identifying Management Zones for Cotton Using Statistical Modeling and Machine Learning Techniques. In Emerging Technologies for A Smarter World (Cewıt). *8th International Conference & Expo*, 2013. Ieee. P.1-6.

[15] Lacape JM, Nguyen TB, Thibivilliers S, Courtois B, Bojinov BM, Cantrell RG, et al. A combined RFLP, SSR-AFLP map of tetraploide cotton based on a *G. hirsutum* x *G. barbadense* backcross population. *Genome*. 2003; 46: 612-626.

[16] Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly MD. Chromosomal Assignment of RFLP Linkage Groups Harboring Important QTLs on an Intraspecific Cotton (*Gossypium hirsutum* L.) *Joinmap. J. Hered*. 2005; 96: 132-144.

[17] Smith JSC, Chin ECL, Shu HO, Smith S, Wall SJ, Senior ML, et al. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea Mays* L.) comparisons with data from RFLPs and pedigree. *Theor Appl Genet*. 1997; 95:163-173.

[18] Dudley JW, Maroof MAS, Rufener GK. Molecular Markers and Grouping of Parents in Maize Breeding Programs. *Crop Science*. 1991; 31: 718-723.

[19] Senior ML, Murphy JP, Goodman MM, Stuber CW. Utility of SSRs for Determining Genetic Similarities and Relationships in Maize using an Agarose Gel System. *Crop Science*. 1998; 38: 1088-1098.

[20] Smith JSC, Smith OS. Fingerprinting crop varieties. Adv. *Agron*. 1992; 47:85-140.

[21] Chaters YM, Robertson A, Wilkinson MJ, Ramsay G. PCR analysis of oil seed rape cultivars (*Brassica napus* L. sp. oleifera) using 50-anchored simple sequence repeat (SSR) primers. *Theor. Appl. Genet*. 196; 92: 442-447.

[22] Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends in Plant Sciences*. 1996; 1:215-222.

[23] Abdalla AM, Reddy OUK, El-Zik K Man, d Pepper AE. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theoretical and Applied Genetics.* 2001; 102:222-229.

[24] Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE. A genetic bottleneck in the evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics*. 2001; 103:547-554.

[25] Lu HJ, Myers GO. Genetic relationships and discrimination of ten influential upland cotton varieties using RAPD markers. *Theoretical and Applied Genetics*. 2002; 105:325-331.

[26] Bertini CH, Schuster I, Sediyama T, Barros E, Moreira MA. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. *Genetics and Molecular Biology*. 2006; *29*(2): 321-329.

[27] Liu D, Guo X, Lin Z, Nie Y, Zhang X. Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. *Genetic Resources and Crop Evolution*. 2006; *53*(6): 1145-1152.

[28] Nachimuthu G, Webb AA. Closing the Biotic and Abiotic Stress-Mediated Yield Gap in Cotton by Improving Soil Management and Agronomic Practices. In: Senthil-Kumar M. (eds) Plant Tolerance to Individual and Concurrent Stresses. *Springer*. 2017.

[29] Liu HS, Li FM. Root respiration, photosynthesis and grain yield of two spring wheat in response to soil drying, *Plant Growth Regul*. 2005; 46, 233–240.

[30] Bange M. The impact of temperature extremes on cotton performance. *CSIRO Plant Industry*. 2004.

[31] Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA. Plant drought stress: effects, mechanisms and management. Agronomy for Sustainable Development, Springer Verlag/EDP *Sciences*/INRA. 2009; 29 (1): 185-212.

[32] Islam S, Haque MS, Emon RM, Islam MM, Begum SN. Molecular characterization of wheat (*Triticum aestivum* L.) genotypes through SSR markers. *Bangladesh Journal of Agricultural Research*. 2012; *37*(3): 389-398.

[33] Doyle and Doyle. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochem Bull*. 1997; 19: 11-5.
[34] Zhang J, Stewart JM. Economical and rapid method for extracting cotton genomic DNA. *J Cotton Sci*. 2000; *4*(3): 193-201.
[35] Seyoum M, Du X, He SP, Jia YH, Pan Z, Sun JL. Analysis of genetic diversity and population structure in upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. *Journal of genetics*. 2018; *97*(2): 513-522.
[36] Mei M, Syed N, Gao W, Thaxton P, Smith C, Stelly D, et al. Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). *Theoretical and applied genetics*. 2004; 108(2):280-291.
[37] Yu J, Yu S, Lu C, Wan W, Fan S, Song M, Zhang J. High-density linkage map of cultivated allotetraploid cotton based on SSR, TRAP, SRAP and AFLP markers. *Journal of Integrative Plant Biology*. 2007; *49*(5): 716-724.
[38] Nguyen TB, Gigel M, Brottier P, Risterucci AM, Lacape JM. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theoretical and applied genetics.* 2004; 109(1):167-75.
[39] Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B. A microsatellite-based, gene-rich linkage map reveals genom structure, function and evolution in Gossypium. *Genetics*. 2007; 176(1):527-541.
[40] Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T. Characteristics, development and mapping of Gossypium hirsutum derived EST-SSRs in allotetraploid cotton. *Theoretical and applied genetics.* 2006; 112(3):430-439.
[41] Reddy O, Pepper A, Abdurakhmonov I, Saha S, Jenkins J, Brooks T, et al. new dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *Journal of cotton science*. 2011; 5(2):103-113.
[42] Qureshi SN, Saha S, Kantety RV, Jenkins JN. EST-SSR: a new class of genetic markers in cotton. 2004.
[43] Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, et al. CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic acids research*. 2014; *42*(D1), D1229-D1236.
[44] Yeh, F.C., Yang, R. C., Boyle, T., Ye, Z. H., & Mao, J.X. (1999). POPGENE, version 1.32: the user friendly software for population genetic analysis.
[45] Nei M. Genetic distance between populations. *The American Naturalist*. 1972; *106*(949): 283-292.
[46] Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.1980*; 32: 314–331.
[47] Liu J, Muse SV. Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21:2128-2129.
[48] Nei M, Tajima FA, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol*. 1983; 19: 153–170.
[49] Pei Z, Gao JQ, Chen J, Wie Z, Li F, Luo L, et al. Genetic diversity of elite sweet sorghum genotypes assessed by SSR markers. *Biologia Plantarum*. 2010; 54 (4): 653-658.
[50] Weir BS. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. 2nd ed. Sunderland, MA, USA: *Sinauer Associates Inc*. 1996.
[51] Pritchard JK, Wena X, Falush D. Documentation for structure software: Version 2.3. Department of Human *Genetics, University of Chicago*.http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf; 2010.
[52] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol*. 2005; 14(8):2611-2620.
[53] Zhang Y, Wang XF, Li ZK, Zhang GY, Ma ZY. Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. *Genetics and Molecular research*. 2011; 10(3): 1462-1470.
[54] Lacape JM, Dessauw DM, Rajab JL, Noyer B, Hau B. Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol. Breed*. 2007; 19:45–58.
[55] Liu S, Cantrell RG, Mccarty JCJR, Stewart JMcD. Simple Sequence Repeat based assessment of genetic diversity in cotton race stock accessions. *Crop Sci*. 2000; 40:1459-1469.
[56] Iqbal MJ, Aziz N, Saeed NA, Zafar Y, Malik KA. Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theoretical and Applied Genetics*. 1997; *94*(1):139-144.
[57] Eminur E, Hançer T. Cotton (*Gossypium hirsutum* L.) Germination Analysis and Molecular Characterization of Genotypes in Constrained Irrigation Conditions. *Turkey Agricultural Research Journal*. 2016; *3*(2): 122-129.
[58] Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*. 2014; *127*(2): 283-295.
[59] Bardak A, Bolek Y. Genetic diversity of diploid and tetraploid cottons determined by SSR and ISSR markers. *Turk. J. Field Crops*. 2012; *17*(2): 139-144.
[60] Gutierrez OS, Basu S, Saha JN, Jenkins DB, Shoemaker CL, Cheatham JC, et al. Genetic distance among selected cotton genotypes and its relationship with F_2 performance. *Crop Sci*. 2002; 42:1841-1847.
[61] Zhang JF, Lu Y, Adragna H, Hughs E. Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. *Crop Sci*. 2005; 45:2363-2373.
[62] Khan AI, Fu YB, Khan IA. Genetic diversity of Pakistani cotton cultivars as revealed by simple sequence repeat markers. *Communications in Biometry and Crop Sci.* 2009; 4(1): 21-30.
[63] Zhao YL, Wang HM, Chen W, Li YH, Gong HY, Sang X, H et al. Genetic diversity and population structure of elite cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. *Plant Syst. Evol*. 2015; 301; 327–336.

[64] Chen G, Du XM. Genetic diversity of source germplasm of upland cotton in China as determined by SSR marker. *Acta Genet. Sinica*. 2006; 33: 1–10.
[65] Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Jenkins JN. Molecular diversity and association mapping of fiber quality traits in exotic G. hirsutum L. germplasm. *Genomics*. 2008; *92*(6): 478-487.
[66] Bardak A, Fidan Ms, Dağgeçen E, Tekerek H, Çelik S, Parlak D, Hayat K. Pamukta İlişkilendirme Haritalaması Yöntemiyle Gossypol ile İlişkili Markörlerin Belirlenmesi (Determination of Gossypol-Related Markers with the Association of Cotton Mapping Method). *Journal of agricultural and nature*. 2017; *20*: 236.