ASSESSING GENETIC DIVERSITY AND POPULATION STRUCTURE ANALYSIS IN UPLAND COTTON GERMPLASM THROUGH MICROSATELLITES

ÇELIK, S.

Department of Forestry, Genç Vocational School, Bingol University, Genç-Bingol 12500, Turkey
(e- mail: sadettincelik@bingol.edu.tr; phone: +90-535-070-6739; fax: +90-426-216-0029)

(Received 5th Mar 2022; accepted 23rd Aug 2022)

Abstract. Genetic diversity analyses were performed on 96 genotypes using 20 SSR markers to determine the genetic diversity and population structure of the genotypes of *Gossypium hirsutum* L. cotton at the molecular level. As a result of the bioinformatic analysis, 126 alleles ranging from 3 to 12 were produced for 20 SSR markers, with an average of 6.3 alleles per locus. In gene diversity analysis, the highest value was 0.4478 and the lowest value was 08.13, the mean GD was 0.2419. The PIC value ranged between 0.08 and 0.28, with a mean of 0.20. The genetic distance between the Upland cotton genotypes was calculated as the highest 0.973 and the lowest 0.156. According to the UPGMA method, the dendrogram obtained and the genotypes were divided into six main clusters. The optimum K number was K = 3 according to the analysis performed in the STRUCTURE v2.3.4 program examining the population structure. Accordingly, the genotypes used in the analysis were divided into three different groups. This information can be used in the development of mapping populations, in the selection of parents to be used in cultivar development breeding programs through crossbreeding, and in determining the polymorphic levels of SSR markers.
Keywords: A*llele, PIC, SSR, Gossypium hirsutum*

Introduction

Upland cotton (*Gossypium hirsutum* L.) is cultivated for its natural fibers and seeds. In addition to providing raw materials to many industries, it is also the main source of income for millions of people (FAO, 2021). There are approximately 50 cultivars of *Gossypium spp.* (Canbell et al., 2009) and 4 of them are cultivated. Two of the cultured ones are allotetraploid (2n = 4x = 52), while the other two are diploid (2n = 2x = 26). Most of the world cotton production is composed of allotetraploid cultivars, *G. hirsutum* L. and *G. bardabance* L. (Wendel et al., 1992). While the cultivation of Upland cotton, which has the highest adaptation, cultivation and spread in the world, is 90% of the world cotton production, the cultivation rate of *G. barbadense* L. is 8%, and the cultivation rate of *G. herbaceum* L. and *G. Arboretum* L. together is 2% (Abdurakhmonov et al., 2012; Zhang et al., 2008). Upland cotton (*G. hirsutum* L.) is superior in its high yield and wide adaptability, but poor in quality, disease and pest resistance. The second most cultivated cotton, *G. barbadense* L. (Sea-island), is resilient against diseases and pests, and high in quality, but has a low yield and limited adaptation abilities (Wang et al., 2008).

Very important agricultural parameters, such as high yield and wide adaptability of Upland cotton, have also improved other weaknesses, allowing biotic and abiotic disease and pest resistance breeding programs to focus on this cotton type (Seyoum et al., 2018). Underlying a successful and sustainable cultivar development breeding program is the wide cultivar of genetic resources from commercial cultivars, wild or foreign cultivars (McCarty et al., 2006) that breeders use when selecting parents

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

(Glaszmann et al., 2010). On the other hand, limited genetic variation of upland cotton genotypes limits variation development breeding programs (Paterson et al., 2004; Abdurakhmonov et al., 2012; Brown, 1983). Huang et al. (2002) put forward that genetic diversity is the basis of a successful breeding program, and limited genetic diversity has a negative effect on breeding programs to have cultivars that are resistant to biotic stress factors such as adaptation, diseases and pests, and abiotic stress factors such as drought, salinity and frost. Limited genetic variation causes the cotton cultivars to be prone to epidemic diseases and frail against climate change (Brown, 1983). Some studies have shown that *G. hirsutum* L. cotton cultivars shows the greatest genetic diversity among all other cultivars (Wendel et al., 1992), and the level of genetic diversity is higher than the other three cultivars (Wendel et al., 1992; Abdurakhmonov et al., 2012) but this has not been seen in current cultivars (Van Esbroeck et al., 1999).

Events such as the genetic drift in cotton variations, the founder effect, migration and gene flow, mate selection, and population bottleneck at the mutation, isolation (separation) and domestication phases (Iqbal et al., 2001), using low genetic breeding materials through crossbreeding in variation development breeding studies (May et al., 1995; Bowman et al., 1996; Wendel et al., 1992; Brubaker et al., 1999) cause changes in allele frequencies in the gene pool and consequently decrease the genetic diversity as well (Haw, 2013; Anonymous, 2021; Purves et al., 2003; California University, 2016). In addition, breeding studies conducted with close relatives cause decreases in genetic variation (Wendel et al., 1992; Esbroeck et al., 1999).

Choosing the right parents, the first step in developing resistant/tolerant cultivars to biotic and abiotic stress conditions through crossbreeding, increases the success of breeding programs. For this, parents should be chosen from those who are as far from each other as possible in terms of kinship. This is possible through genetic diversity analysis. Genetic diversity analyses are conducted using morphological and pedigree data (May et al., 1995; Van Esbroeck et al., 1999), biochemical markers (Wendel et al., 1992) and DNA-based markers (Yu et al., 2012). However, since morphological and pedigree data are affected by the environment (Huang et al., 2002), DNA-based markers in genetic diversity analyses are being used more frequently as they are not affected by the environment, can give safer results and are reproducible.

Molecular markers such as RFLP (Restriction fragment length polymorphism) (Ulloa et al., 2005), RAPD (Random amplified polymorphic DNA) (Mumtaz et al., 2010; Sapkal et al., 2011), AFLP (Amplified fragment length polymorphism) (Rana et al., 2005; Badigannavar et al., 2012), ISSR (Intersimple sequence repeats) (Noormohammadi et al., 2011) and SSRs (Simple sequence repeats) (Chen and Du, 2006; Abdurakhmonov et al., 2009; Cai et al., 2014; Tyagi et al., 2014; Zhao et al., 2014) are being used in genetic diversity analysis. However, SSR markers are used in cultivar fingerprinting, Association mapping, QTL mapping and Marker-assisted selection (MAS), especially in genetic diversity analyses, due to their advantages such as being multiallelic and highly polymorphic, their abundance in the plant genome and their ease of use (Reddy et al., 2001; Zhang et al., 2008). SSR markers can be effectively used in *Gossypium spp*. Germplasm resources to do molecular characterization (Liu et al., 2000; Lacape et al., 2007; Sun et al., 2009; Zhang et al., 2005; Reddy et al., 2001).

The aim of this study is to use 96 worldwide *Gossypium hirsutum* L. cotton cultivars germplasm genotypes to assess the genetic diversity, genetic distance, and population structure analyses.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

Materials and methods

Plant materials

A set of 96 genotypes of *Gossypium hirsutum* L., located at Agricultural Biotechnology germplasm, Faculty of Agricultural of Kahramanmaraş Sütçü İmam University (Turkey/Kahramanmaraş), collected from different locations worldwide, and which are known for their fiber quality, fiber yield, seed cotton yield, adaptation, number of bolls, important economic parameters such as earliness and tolerance to biotic and abiotic stress conditions were used in the study.

DNA extraction and SSR genotyping

Genomic DNA isolation was performed by taking 0.5 g leaf samples from true young leaves of cotton genotypes and using the Cetyl trimethyl ammonium bromide (CTAB) method developed by Zhang and Stewart (2000). Leaf samples were kept at -80 °C until isolation. DNA isolation was performed after physical grinding with liquid nitrogen, and the control of DNAs was tested by being executed in 1% agarose gel. To clean dirty DNA, Proteinase K enzyme was used, and they have been purified from histone proteins wrapped around the DNAs. Concentrations of DNA were determined by using the Nanodrop spectrophotometer (Thermo fisher).

Genetic diversity analyses of commercial cultivars of 96 *G. hirsutum* L. cotton were carried out using 24 Simple Sequence Repeat (SSR) markers distributed on 26 chromosomes of Upland cotton genome (AD genome). It can be said that in previous studies, of the 20 microsatellites that gave polymorphic bands, 7 belonged to BNL, 5 belonged to JESPR and 8 belonged to NAU marker types were used. The primer sequences of the SSR markers are available in the CottonGene (http://www.cottongene.org) and Cotton Marker (http://www.cottonmarker.org) databases.

PCR amplification, PCR solution and gel electrophoresis were performed according to the technique specified by Zhang and Stewart (2000). Reaction volume of 15 μL^{-1} was used for each PCR cycle. The PCR reaction mixture has 0.75 µL-1 dNTP (Conc.10 mM), 1.5 μL^{-1} 10X PCR buffer, 1 μL^{-1} forward primer, 1 μL^{-1} reverse primer, 0.5 μL^{-1} Taq DNA polymerase. PCR protocol consists of the following stages: Denaturation at 94 °C for 5 min, then at 94 °C with 34 cycles for 1 min, at 60 °C for 1 min, annealing at 72 °C for 2 min, and the extension at 72 °C for 7 min. (Conc.5 U/µL, 2 μL^{-1} template DNA (Conc. 25 ng/μL^{-1}), 8.25 µL ddH2O (double-distilled) components).

PCR products were horizontally electrophoresed with 1% agarose in 1X TBE (Tris-Borate-EDTA) buffer solution. The length of the tapes was measured using a 100-bp DNA ladder. After electrophoresis, the gels were kept in ethidium bromide solution for 20 min and visualized in a UV (Ultraviolet) device.

Data analysis

For alleles amplified by the SSR marker loci, if the bands were present, then they were scored as "1", and if not, as "0". Alleles amplified by each SSR marker locus were assigned letters such as A, B, and C. Parameters such as the number of alleles, genetic diversity (heterozygosity) (Nei, 1972) and Polymorphic information content (PIC; Botstein et al., 1980) were calculated using the PowerMarker ver. 3.25 program (Liu and Muse, 2005) with the scored alleles. Gene diversity (GD) of a locus is accepted as the response of expected heterozygosity (*He*). It shows the expected heterozygosity ratio of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

genotypes based on Hardy-Weinberg equilibrium as a measure of the genetic diversity of the population (Nei, 1973). The genetic distance in phylogenetics (Nei et al., 1983) was calculated based on the dissimilarity coefficient of the SSR markers between genotypes. A phylogenetic tree was created using the genetic distance coefficients matrix obtained from POPGENE 1.31 ver. (Yeh et al., 1999) program, and the clustering analysis for all the genotypes in the main and subgroups were conducted using the Unweighted Pair Group Method of Arithmetic Mean (UPGMA) technique. MEGA_11.0.10 ver. program was used in creating, interpreting and organizing the phylogenetic tree (Tamura et al., 2021). PIC value shows the number of alleles at each locus and the distinctive features of the markers through the relative frequencies of alleles in the population (Pei et al., 2010). PIC calculation was based on the formula given below:

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

The P-value is the frequency of the i^{th} allele of the 96 Upland cotton genotypes subject to the analyses (Weir, 1996).

Population structure

For clustering analysis of genotypes, a Bayesian model-based (Bayesian Model-Based (MBB) STRUCTURE ver. 2.3.4 program using co-dominant genotypic data was used (Pritchard et al., 2010). The subpopulation number (K) of the Upland cotton germplasm genotypes was obtained by calculating the ΔK (Evanno et al., 2005) value. The K value was adjusted between 2-10 and 5 repetitions were made for each K. The STRUCTURE program settings, Leng of Burning Periods and the iteration number of Markov Chain Monte Carlo (MCMC) was set as 10.000-100.000 (Chen et al., 2012; Zoric et al., 2012). Optimum K value was identified for the population. The Compressing results as a zip format were uploaded to the web-based "STRUCTURE HARVESTER" platform (Earl and Von Holdt, 2012).

Results and discussion

Phylogenetic analysis and genetic diversity

In genetic diversity analyses performed on 96 genotypes of Upland cotton, 24 SSR markers were used, 83.3% (20) of them produced polymorphic bands, 12.5% (3) monomorphic bands and 4.2% (1) did not produce any. 126 alleles were produced from 24 SSR markers in total: 6 from BNL0852, 8 from BNL1317, 12 from BNL1690, 4 from BNL1694, 11 from BNL3031, 5 from BNL3140, 6 from BNL3255, 6 from JESPR65, 6 from JESPR114, 6 from JESPR300, 9 from JESPR0092, 7 from JESPR0122, 3 from NAU0923, 9 from NAU2173, NAU2196, 5 from NAU4024, 6 from NAU1037, 6 from NAU1093, 4 from NAU1248, 3 and 4 from NAU2302 markers. For SSR markers that produced polymorphic bands, a mean of 6.3 alleles per locus was produced, while the marker producing the most alleles was BNL1690 (12), while the marker producing the least allele was NAU2196 (3) (*Table 1; Fig. 1*). Similar results were also produced in literature. 5.6 alleles (Lacape et al., 2007), 5.08 (Zhang et al., 2011), and 6.9 alleles (Moiana et al., 2012) were obtained per SSR locus. Liu et al. (2000a) amplified 62 loci in cotton using 56 polymorphic microsatellites and produced a total of 325 alleles, with a mean of 5 alleles per locus. Bertini et al. (2006), in their

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

genetic diversity analysis in cotton using 31 SSR primary farmers, reproduced 31 loci and duplicated BNL1964 and BNL3408 markers in two loci. They also obtained a total of 66 alleles, with an average of 2.13 alleles per microsatellite locus. Liu et al. (2000b) reproduced 2 loci with SSR primers. Gutiérrez et al. (2002) also reproduced 69 loci using 60 polymorphic SSRs and produced a total of 139 alleles, with a mean of 2 alleles per locus. Due to cotton's allotetraploid genomic structure, SSR primer pairs appear to reproduce multiple loci (Fang et al., 2013). Zhao et al. (2015) reported 2-5 alleles per SSR locus, with a mean of 2.26. Most of the SSR primers used in the study reproduced multiple marker loci polymorphically. In our study, the average allele amount per locus and the total number of alleles were higher than previous genetic diversity analyses (Bertini et al., 2006; Ai et al., 2017; Fang et al., 2013) performed on the same cotton cultivars. It is thought that this may be due to the use of SSR markers with broad-spectrum genetic diversity that show high polymorphism in genetic mapping.

In our study, the Polymorphic information content (PIC) value, which gives information about each SSR primer pair, ranged from 0.12 to 0.35, with a mean of 0.20 (*Table 1; Fig. 1*). As expected in the study, gene diversity (GD) values were higher than PIC values. The PIC values are always lower than the diversity (or *He*) values, and the proliferation of alleles and the increase in the flatness in allele frequencies (the low probability of individuals having similar heterozygous genotypes) bring the PIC value closer to the GD values. Luo et al. (2019) produced similar values in their study as well. These results are similar to the studies of Ai et al. (2017) who found 0.25 mean PIC, Fang et al. (2013) who found 0.29 mean PIC, Rungis et al. (2005) who found 0.37 mean PIC and Bertini et al., who found 0.40 mean PIC in 53 *G. hirsutum* L. cotton genotypes. Tiyagi et al. (2014) found mean PIC values of 0.17 for upland cotton and 0.16 PIC values for *G. barbadense* cotton cultivars in their genetic diversity analyses using *G. hirsutum* and *G. barbadense* cotton cultivars and obtained results similar to our study. A PIC value of "1" indicates that the marker is polymorphic, and a value of "0" indicates that it is a monomorphic marker. At the same time, PIC values of 0.5 and above have the potential to give more information about the marker, if they are between 0.5 and 0.25, they are medium, and if they are less than 0.25, they give very less information about the marker (Tiyagi et al., 2014). Polymorphic information content (PIC) is an analysis that measures the informativeness of the markers used in the analysis (Guo and Elston, 1999). The PIC value was first defined by Botstein et al. (1980) as a measure of the informative trait of a marker independent of the mode of inheritance of the associated trait. Other researchers also identified PIC as a measurement of polymorphism of SSR markers (Gupta and Vershey, 2000; Shete et al., 2000; Botstein et al., 1980). Liu et al. (2000b) obtained PIC values in cotton ranging between 0.05-0.082, with a mean of 0.31. Guang and Xiong-Ming (2006) obtained a mean of 3.6 alleles, between 2 and 8 per primer. PIC values ranged between 0.278-0.865 with a mean of 0.62. The fact that the PIC values in our study were lower than some studies in the literature may be due to genetic diversity analyses being conducted using cultured commercial cultivars. Iqbal et al. (1997) and Tatineni et al. (1996) concluded in their research that genetic diversity is low in molecular diversity studies conducted using the cultured cotton cultivars. The low biodiversity in cultivated Upland cotton is due to bottleneck events during the domestication phase (Brubaker and Wendel, 1994; Wendel and Cronn, 2003; May et al., 1995; Iqbal et al., 2001). However, developments in the technology of obtaining transgenic cotton also contribute to the decrease in the level of genetic diversity in other cotton growing countries, especially in the USA (Zhang et al., 2005).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

Table 1. *Polymorphic SSR marker properties used in this study*

No.	Primer name	Primer sequences	Repeat motify	Location	Allele no.	Gene diver (GD)	PIC
1	BNL0852	F: TGCTTTCAGCCAATGACTTG R: AACAATGCCCCCAATATTCA	(CA)13	Chr9.	6	0.3185	0.27
2	BNL1317	F: AAAAATCAGCCAAATTGGGA R: CGTCAACAATTGTCCCAAGA	(AG)14	Chr6.	8	0.2474	0.21
3	BNL1690	F: TTTGTCTTTCTGTTACCAAATGG R: CCAGGAAATTTGAGGTGGAA	(GA)10	Chr9	12	0.0813	0.08
4	BNL1694	F: CGTTTGTTTTCGTGTAACAGG R: TGGTGGATTCACATCCAAAG	(AG)19, (TC)19	Chr1.	4	0.3462	0.27
5	BNL3031	F: AGGCTGACCCTTTAAGGAGC R: AACCAACTTTTCCAACACCG	(AG)27	D09.	11	0.2547	0.22
6	BNL3140	F: CACCATTGTGGCAACTGAGT R: GGAAAAGGGAAAGCCATTGT	(GA)11	A09.	5	0.2960	0.24
7	BNL3255	F: GACAGTCAAACAGAACAGATATGC R: TTACACGACTTGTTCCCACG	(GC)6, AT(AC)14	A08.	6	0.4478	0.35
8	JESPR0065	F: CCACCCAATTTAAGAAGAAATTG R: GGTTAGTTGTATTAGGGTCGTTG	(GAA)25	A05.	6	0.3495	0.28
9	JESPR 114	F: GATTTAAGGTCTTTGATCCG R: CAAGGGTTAGTAGGTGTGTATAC	(GT)12	D10.	6	0.3117	0.26
10	JESPR 300	F: CGCATCACAAACCAAACAC R: CGGAAAATGATGATGATGAAGAAG	(CTT)5, (CAT)6	Chr8.	6	0.2306	0.20
11	JESPR0092	F: GGGACCTCTATTGAATAGCTGGAG R: CTCTTGGCATCATTAGTTCCTGG	(GAA)23	D12.	9	0.2324	0.20
12	JESPR0122	F: GCTGCTGGTTTTACTTGTTGG R: CTATGGTGGAGGAGCAACAAC	(CAT)5	Chr05.	7	0.1357	0.12
13	NAU0923	F: GGAATTCAAGGTTGAAGGAG R: CCTCTTCTTTGGCTCTGAAA	(TCTTTT)4	Chr6	3	0.3138	0.26
14	NAU2173	F: GCCAAATAGGTCACACACAA R: AGCGAGAAGGAGACAGAAAA	AAG (17)	NA	9	0.2054	0.17
15	NAU2196	F: TCAAGAAAACATGCCTGCTA R: CTATTTGCTCGTTGTTGACG	CAT (4)	Chr7.	3	0.3306	0.27
16	NAU4024	F: ACAAGCATCTTCATGGACCT R: AGAAGGATGATGCAAAGAGG	(GTC)6	Chr5	5	0.2345	0.20
17	NAU1037	F: CACCTTCACCTAACCATCAA R: GAAGAATTGCGAGAAGAGGA	(CTGCCA)3	-	6	0.1570	0.14
18	NAU1093	F: TGTGATGAAGAACCCTCTCA R: AAATGGCGTGCTTGAAATAC	(TA)14	-	6	0.1831	0.16
19	NAU1248	F: AATGTCAGCTGCCTATTTCC R: AAGACAGGCGATGTCATCTT	(TCTTCC)3	Chr5.	4	0.2281	0.20
20	NAU2302	F: CAAACCGTCAAATGAGACAA R: GCCTCTAAGGGTCCCTACTC	AT(12); GA (10)	Chr12	4	0.2419	0.20
Mean				-	6.3	0.2419	0.20
Total				-	126	-	-

In the current study, the genetic diversity varied between 0.08130 and 0.4478, with a mean of 0.2419 (*Table 1; Fig. 1*). While BNL3255, JESPR0065 and NAU2196 showed the highest genetic diversity values, BNL1690, JESPR0122 and NAU1037 markers showed the lowest genetic diversity values (*Table 1*). There are studies with similar results in the literature. Seyoum et al. (2017) obtained genetic diversity values ranging from 0.020 to 0.492, with a mean of 0.279. The low genetic diversity in genetic diversity studies conducted using the upland cotton may be due to the fact that the developed upland cotton cultivars were obtained using very few genotypes (Chen and Du, 2006; Du et al., 2007). Rungis et al. (2005) stated that the bottleneck events occurring in the first stage of modern upland cotton cultivar developments lead to a

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

significant decrease in the level of genetic diversity in upland cotton cultivars. Low genetic diversity studies were carried out on Upland cotton cultivars (Tiyagi et al., 2014; Campbell et al., 2009; Abdurakhmonov et al., 2008). Heterozygosity levels of marker data can be calculated by averaging the similarity frequencies of alternative alleles (Heterozygosity, *H*: 0.9 or 0.5 values indicate high heterozygosity, while *H*: 0.1 indicates low heterozygosity (Li et al., 2007).

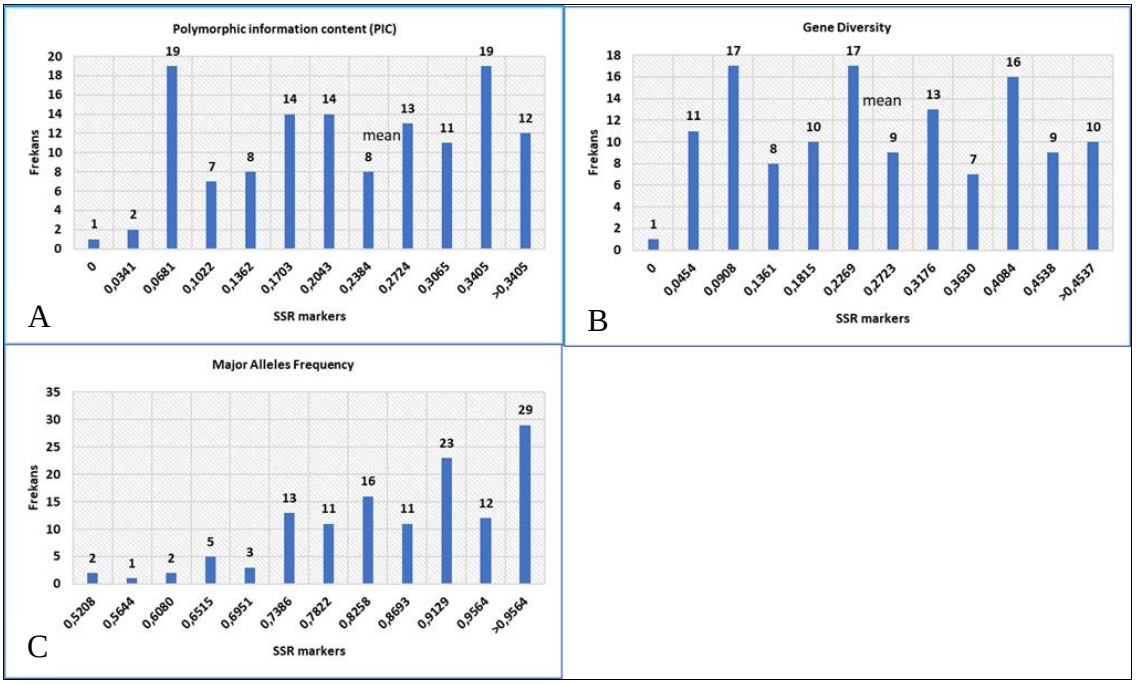

Figure 1. *(A) Polymorphic information content (PIC), (B) genetic diversity (GD), (C) major alleles frequency (MAF)*

The genetic distance between the Upland cotton genotype pairs was based on the pairwise matching technique of Nei (1972), and the genetic distance values between the Upland cotton genotypes varied between 0.973 and 0.015. The highest genetic distance was between 308 (Campo) and DPL-5614 and Israel-2 (0.9730). These were followed by 308 (Campo) and Stoneville 506 with 0.9322 genetic distance value, and the values between DP 5111, Nazilli 303 and 93 FF 01 and Nazilli 303 genotypes. The lowest genetic distance values were obtained between Delcerro and 4SP, CA-228; DPL50-DPL5409, DPL5614- DPL50 and DPL50- ERŞAN-92 genotypes. The low genetic distance between them is directly proportional to the degree of relation. Similar studies have been carried out in the literature. The similarity coefficient ranged between 0.407 and 0.767, with a mean value of 0.587 (Hancı and Gokçe., 2016). Nas et al. (2011) recorded a GD coefficient as low as 0.40 between the two cultivars and stated that this might be due to these cultivars being closely related. In addition, the genetic distance between heterogeneous genotypes (generally wild cultivars) should be higher than homogeneous (breeding genotypes) genotypes (Hinze et al., 2017). The genetic distance in the genetic diversity analyses performed in the F_2 cotton population varied between 0.06-0.34 (Gutierrez et al., 2002). The genetic distance between genotypes of the Upland cotton cultivars is higher than the F_2 segregation population from the same mother and father, due to the possibility of having different parents. The genetic

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

distance between *G. hirsutum* and *G. barbadense* varies between 42-54% (Kebede et al., 2007). Lacape et al. (2007) revealed that there is a higher genetic distance between *G. hirsutum* and *G. barbadense* (GD = 0.89-0.91). Analyses with highly polymorphic SSR markers showed a high difference between G. *hirsutum* and G. *tomentosum* (D:0.71-0.75) and between G. *barbadense* and G. *tomentosum* (D:0.80) (Lacape et al., 2007). Kebede et al. (2007) also stated that the genetic distance between genotypes belonging to cotton genome groups such as A and D varies between 29-42%. The genetic distance between those included in the AD tetraploid cotton group also varies between 0.80-0.88 (Liu et al., 2000).

Ninety-six upland cotton cultivars were divided into 6 main groups in phylogenetic tree analysis conducted in the MEGA_11.0.10 computer program using short tandem repeat markers (*Fig. 2*). The groups are as follows: Flash genotype in an independent group, Marcell leaf brown, israil 2, Paymaster 2379, Acala 552, DP 388, Dulcerro, 308 (Campo), 153 F, and 111 glandless genotypes in first group, TX0091-2, STV 474, PI 528420, STV-132, Acala Prema, Lockette genotypes in the second group, Okra 201, Nazilli 503, Mn nair 235-612, Acala 172, 93 FF01, 919 (leader) genotypes in the third group, Acala 1517-D, PI 528872 (SEALAND 3), Acala 1517-99, Nazilli 87, STV-506, DPL-529, Acala 3080, Is 10, Orgosta 644 genotypes in the fourth group, Tamcot Luxor and Sahel 1 genotypes in the fifth group, Cabu cs 2-1-8-3, Deltapine 5111, BA 525, STV 373, Deltapine 419, Celia, PI 529128 (Acala 1517 D) and Deltapine 50 genotypes in the sixth group. The remaining 63 genotypes formed subsets among themselves and clustered in a separate sixth group (*Fig. 2*). Similar studies were conducted in the literature. Ulah et al. (2012) stated that their dendrogram built using 19 Bt cotton genotypes was divided into three main groups and two independent groups, Fang et al. (2013) stated that 193 upland cotton genotypes formed 15 main groups and two independent groups, and Erdal (2018) stated that the genotypes were clustered in four main groups in dendrogram, and formed 8 subgroups.

Population structure analysis in MBB subpopulations

Using the STRUCTURE v 2.3.4 ver. (Pritchard et al., 2000) program, the structure of the cotton populations and the genetic relationship between 96 upland cotton (*G. Hirsutum* L.) cultivars obtained from different countries were tried to be revealed. The K value shows the number of the clusters obtained in cluster analysis using genotypic data. The most suitable K values was calculated via the log likelihood of the data and to find the value of optimum K (Number of subpopulations), the number of clusters was plotted against Delta K (△K), which indicates the highest peak point to K = 3. Two peaks occurred at K = 5 and K = 9. Since this provides information about these two different populations, it is important. According to Seyoum et al. (2018), the entire germplasm population was divided into three subpopulations in their clustering analysis conducted using the STRUCTURE program. Population structure analysis shows that the highest population number △K value is K = 3, and the whole population is divided into 3 subpopulations (*Figs. 2* and *3*).

Population structure analyzes are the guides for understanding genetic diversity and for association mapping studies and for planning the future. Since the presence of population structure data in Association mapping (AM) analyses may cause incorrect marker-trait matching, testing the population structure is the first step in establishing a correct marker-trait match and determining the gene/genes controlling the trait (Eltaher et al., 2018). While the number of subpopulations could not be determined in the Ln (K)

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

plot in the study (*Fig. 3a*), 3 groups (G1, G2, G3) were determined based on the optimum ΔK value (*Fig. 3b*). However, these three clusters happen to be mixed significantly (*Fig. 3c*). Each group is indicated by a color. Similar results were obtained in the phylogenetic tree (Neighbor-joining tree). Groups of different colors show that they were bred in different geographic locations (Tiyagi et al., 2014). Ai et al. (2017) has also reported similar results. Population structure and determining the relationship between Upland cotton genotypes are of great importance for the development, characterization and maintenance of genotypes. Results of our study are similar with the studies of Jia et al. (2014b) and Ai et al. (2017) who divided the genotypes into three groups as a result of STRUCTURE analyses using Upland cotton genotypes originated in China. Although Mei et al. (2013) divided Upland cotton genotypes into two subsets as a result of MBB structure analysis even though they used Upland cottons originated in China, Tiyagi et al. (2014) used upland cotton cultivars originating from the USA and the genotypes were divided into five main populations. It is thought that the difference here may be due to the cotton genotype type, the breeding stage and the geographical location of breeding.

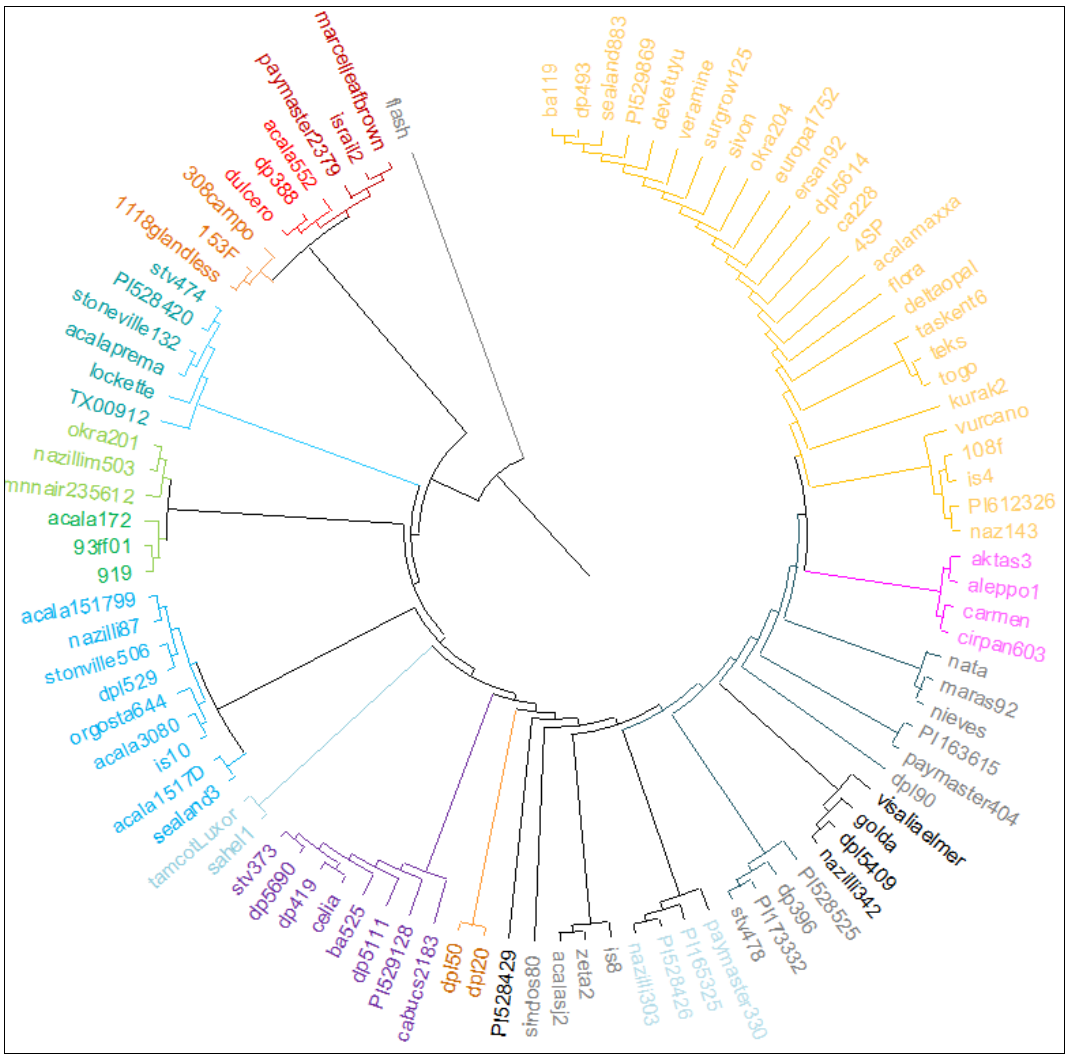

Figure 2. *Cluster analyses of 96 upland cotton genotypes using unweighted pair group of arithmetic mean (UPGMA) Basid Neil's 1972 in MEGA_11.0.10 (Tamura et al., 2021). Six major clusters were obtained.*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

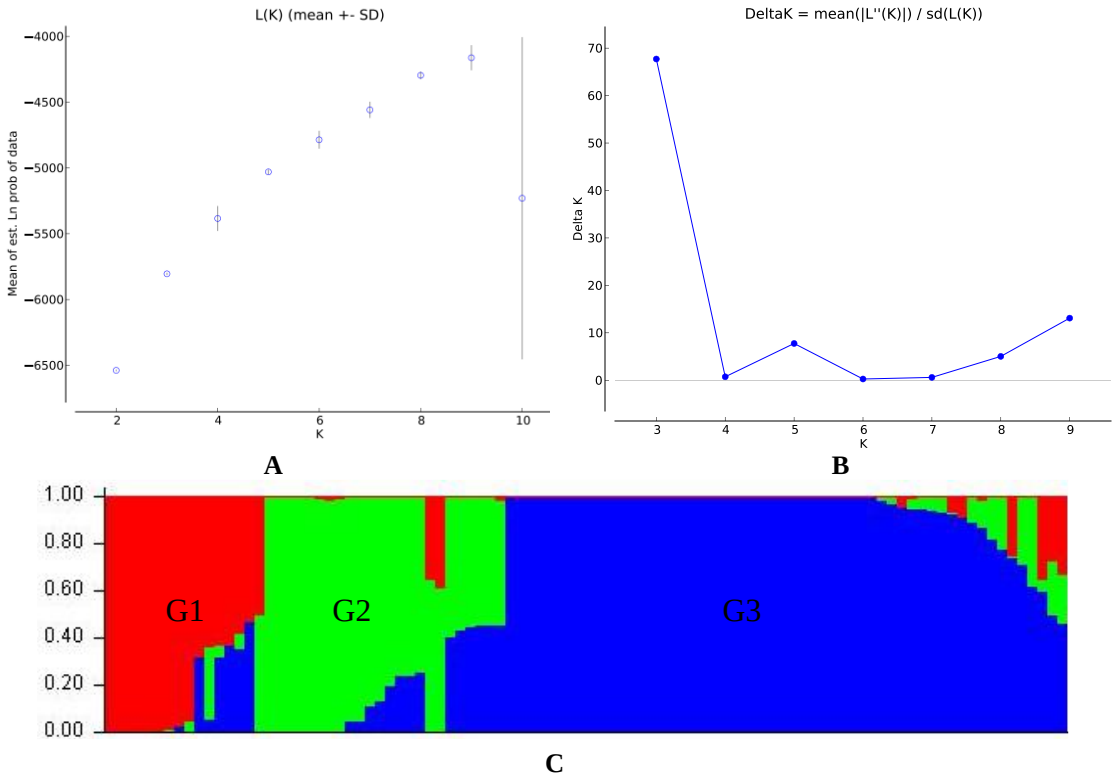

Figure 3. *(A) The plot of Ln (probability of data) analysis in the population structure. (B) Estimating the number of subpopulations DK value (⊿K) using Evanno et al. (2005) method. (C) The Q-plot indicates the Model-based Bayesian (MBB) structure clustering of a panel of 96 upland cotton genotypes. Each vertical bar represents a genotype. Each color represents the estimated member individuals of each K group*

Conclusions

Genetic diversity results from modification of the plant's genome to adapt to changing environmental conditions. As a result of such modifications, differences occur in the genomes of plants and Microsatellites, which can detect these differences, are used in genetic diversity analyses in cotton plants thanks to their high polymorphic capabilities. As expected, in the study, the PIC values were lower than gene diversity values. Polymorphic levels of SSR markers were found to be high and PIC values that provide information about the marker were calculated using the PowerMarker program. The genetic distance between the Upland cotton genotypes was similar to the literature. As a result of UPGMA analysis, genotypes were divided into 6 main clusters and a total of 5 large clusters were formed. The sixth major cluster covers the vast majority of cotton genotypes. In the analysis of the population structure, the clusters were divided into 3 groups depending on the geographical regions, but the genotypes were partially mixed with each other due to the use of the parents obtained from different geographic locations in breeding programs. The genetic diversity, which is generally low, is getting lower. Genetic diversity analyzes and population structure analyses enable benefitting from genetic variation in genomic studies. Considering the other economic characteristics of the genotypes that are genetically far from each other, the development of resistant cultivars against biotic and abiotic diseases and pests can be used as parents in breeding programs using marker-assisted selection technology. In

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

addition, SSR markers, whose polymorphic levels are determined, can enable the development of mapping populations for genetic mapping such as high-resolution linkage mapping and Association mapping.

Acknowledgements. The author is grateful to associate Prof. Dr. Adem Bardak for providing cotton germplasm materials, and Osman Yiğit for the assistance on laboratory analyses in Kahramanmaraş Sütçü Imam University, Agricultural Biotechnology Laboratory.

REFERENCES

[1] Abdurakhmonov, I., Kohel, R., Yu, J., Pepper, A., Abdullaev, A., Kushanov, F. et al. (2008): Molecular diversity and association mapping of fiber quality traits in exotic G. hirsutum L. germplasm. – Genomics 92: 478-487.

[2] Abdurakhmonov, I., Saha, S., Jenkins, J., Buriev, Z., Shermatov, S., Scheffler, B. et al. (2009): Linkage disequilibrium based association mapping of fiber quality traits in G. hirsutum L. variety germplasm. – Genetica 136: 401-417.

[3] Abdurakhmonov, I. Y., Buriev, Z. T., Shermatov, S. E., Abdullaev, A. A., Urmonov, K., Kushanov, F., Egamberdiev, S. S., Shapulatov, U., Abdukarimov, A., Saha, S., Jenkins, J. N., Kohel, R. J., Yu, J. Z., Pepper, A. E., Kumpatla, S. P., Ulloa, M. (2012): Genetic Diversity in *Gossypium* Genus. – In: Caliskan, M. (ed.) Genetic Diversity in Plants 7. InTech, London, pp. 313-338.

[4] Ai, X., Liang, Y., Wang, J., Zheng, J., Gong, Z., Guo, J., Qu, Y. (2017): Genetic diversity and structure of elite cotton germplasm (*Gossypium hirsutum* L.) using genome-wide SNP data. – Genetica 145(4): 409-416.

[5] Anonymous (2021): Genetic drift and the founder effect. – Evolution. http://www.pbs.org/wgbh/evolution/library/06/3/l_063_03.html.

[6] Badigannavar, A., Myers, G. O., Jones, D. C. 2012 Molecular diversity revealed by AFLP Markers in upland cotton genotypes. – J. Crop Improvement 26: 627-640.

[7] Bertini, C. H., Schuster, I., Sediyama, T., Barros, E. G. D., Moreira, M. A. (2006): Characterization and genetic diversity analysis of cotton cultivars using microsatellites. – Genetics and Molecular Biology 29: 321-329.

[8] Botstein, D., White, R. L., Skolnick, M., Davis, R. W. (1980): Construction of a genetic linkage map in man using restriction fragment length polymorphisms. – Am J Hum Genet 32: 314-331.

[9] Bowman, D. T., May, O. L., Calhoun, D. S. (1996): Genetic base of upland cotton cultivars released between 1970 and 1990. – Crop Sci 36: 577-581.

[10] Brown, W. L. (1983): Genetic diversity and genetic vulnerability: an appraisal. – Econ Bot 37: 4-12.

[11] Brubaker, C. L., Wendel, J. F. (1994): Re-evaluating the origin of domesticated cotton (Gossypium hirsutum: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). – Am J Bot 81: 1309-1326.

[12] Brubaker, C. L., Bourland, F. M., Wendel, J. F. (1999): The Origin and Domestication of Cotton. – In: Smith C. W., Cothren J. T. (eds.) Cotton: Origin, History, Technology, and Production. Wiley, New York, pp 3-32.

[13] Cai, C. P., Ye, W. X., Zhang, T. Z., Guo, W. Z. (2014): Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). – J. Integr. Plant Biol. 56: 51-62.

[14] Campbell, B. T., Williams, V. E., Park, W. (2009): Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. – Euphytica 169: 285-301.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

[15] Chen, G., Du, X. M. (2006): Genetic diversity of source germplasm of upland cotton in China as determined by SSR marker. – Acta Genet. Sinica 33: 1-10.
[16] Chen, X., Min, D., Yasir, T. A., Hu, Y. G. (2012): Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. – PLoS One 7: e44510.
[17] Du, X. M., Zhou, Z. L., Jia, Y. H., Liu, G. Q. (2007): Collection and conservation of cotton germplasm in China (English abstract). – CNKI Cotton Sci. 19: 346-353.
[18] Earl, D. A., vonHoldt, B. M. (2012): Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. – Conserv. Genet. Resour. 4: 359-361.
[19] Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M. et al. (2018): Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping-by-sequencing. – Front. Genet. 9: 76.
[20] Erdal, Ş. (2018): Determination of genetic distances of inbred maize lines. – Derim 35(1): 73-80.
[21] Esbroeck, G. V., Bowman, D. T., May, O. L., Calhoun, D. S. (1999): Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. – Crop Sci. 39: 323-328.
[22] Evanno, G., Regnaut, S., Goudet, J. (2005): Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. – Mol Ecol 14: 2611-2620.
[23] Fang, D. D., Hinze, L. L., Percy, R. G., Li, P., Deng, D., Thyssen, G. (2013): A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. – Euphytica 191(3): 391-401.
[24] FAO (2021): https://www.fao.org/land-water/databases-and-software/crop-information/cotton/en/ – Access: 12.10.2021.
[25] Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., Varshney, R. K. (2010): Accessing genetic diversity for crop improvement. – Curr. Opin. Plant Biol. 13: 167-173.
[26] Guang, C. H. E. N., Xiong-Ming, D. U. (2006): Genetic diversity of source germplasm of upland cotton in China as determined by SSR marker analysis. – Acta Genetica Sinica 33(8): 733-745.
[27] Guo, X., Elston, R. (1999): Linkage information content of polymorphic genetic markers. – Human Heredity 49(2): 112-118.
[28] Gupta, P. K., Varshney, R. K. (2000): The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. – Euphytica 113: 163-185.
[29] Gutierrez, O. A., Basu, S., Saha, S., Jenkins, J. N., Shoemaker, D. B., Cheatham, C. L., McCarty, J. C. (2002): Genetic distance among selected cotton genotypes and its relationship with F2 performance. – Crop Science 42: 1841-1847.
[30] Hanci, F., Gökçe, A. F. (2016): Molecular characterization of Turkish onion germplasm using SSR markers. – Czech Journal of Genetics and Plant Breeding 52(2): 71-76.
[31] Haw, J. (2013): Northern elephant seals: increasing population, decreasing biodiversity. – Scientific American. http://blogs.scientificamerican.com/expeditions/northern-elephant-seals increasing-population-decreasing-biodiversity/.
[32] Hinze, L. L., Hulse-Kemp, A. M., Wilson, I. W., Zhu, Q. H., Llewellyn, D. J., Taylor, J. M., Stelly, D. M. (2017): Diversity analysis of cotton (Gossypium hirsutum L.) germplasm using the CottonSNP63K Array. – BMC Plant Biology 17(1): 1-20.
[33] Huang, X., Börner, A., Röder, M., Ganal, M. (2002): Assessing genetic diversity of wheat (Triticum aestivum L.) germplasm using microsatellite markers. – Theoretical and Applied Genetics 105(5): 699-707.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

[34] Iqbal, M. J., Aziz, N., Saeed, N. A., Zafar, Y. (1997): Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. – Theoretical and Applied Genetic 94: 139-144.
[35] Iqbal, A. E., Reddy, O. U. K., El-Zik, K. M., Pepper, A. E. (2001): A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. – Theor Appl Genet. 103: 547-554.
[36] Jia, Y. H., Sun, J. L., Wang, X. W., Zhou, Z. L., Pan, Z. E., He, S. P. et al. (2014b): Molecular diversity and association analysis of drought and salt tolerance in *G. hirsutum* L. germplasm. – J. Integr. Agric. 30: 1845-1853.
[37] Kebede, H., Burow, G., Dani, R. G., Allen, R. D. (2007): A-genome cotton as a source of genetic variability for upland cotton (*Gossypium hirsutum*). – Genetic Resources and Crop Evolution 54(4): pp. 885-895.
[38] Lacape, J. M., Dessauw, D., Rajab, M., Noyer, J. L., Hau, B. (2007): Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. – Molecular Breeding 19(1): 45-58.
[39] Li, Y., Li, Y., Han, K., Wang, Z., Hou, W., Zeng, Y. et al. (2007): Estimation of multilocus linkage disequilibria in diploid populations with dominant markers. – Genetics 176: 1811-1821.
[40] Liu, K., Muse, S. V. (2005): PowerMarker: an integrated analysis environment for genetic marker analysis. – Bioinformatics 21: 2128-2129.
[41] Liu, S., Cantrell, R. G., McCarty-Jr, J. C., Stewart, J. Mc. D. (2000a): Simple sequence repeat based assessment of genetic diversity in cotton race stock accessions. – Crop Science 40: 1459-1469.
[42] Liu, S., Saha, S., Stelly, D., Burr, B., Cantrell, R. G. (2000b): Chromosomal assignment of microsatellite loci in cotton. – Journal of Heredity 91: 326-332.
[43] Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., Abdel-Haleem, H. (2019): Genetic diversity and population structure of a Camelina sativa spring panel. – Frontiers in Plant Science 10: 184.
[44] May, O. L., Bowman, D. T., Calhoun, D. S. (1995): Genetic diversity of U.S. upland cotton cultivars released between 1980 and 1990. – Crop Sci 35: 1570-1574.
[45] McCarty, J. C., Wu, J., Jenkins, J. N. (2006): Genetic diversity for agronomic and fiber traits in day-neutral accessions derived from primitive cotton germplasm. – Euphytica 148: 283-293.
[46] Mei, H., Zhu, X., Zhang, T. (2013): Favorable QTL alleles for yield and its components identified by association mapping in Chinese Upland cotton cultivars. – PLoS One 8: e82193.
[47] Moiana, L. D., Filho, P. S. V., Gonçalves-Vidigal, M. C., Carvalho, L. P. (2012): Genetic diversity and population structure of cotton (*Gossypium hirsutum* L. race latifolium H.) using microsatellite markers. – Afr. J. Biotechnol. 11: 11640-11647.
[48] Mumtaz, A. S., Naveed, M., Shinwari, Z. K. (2010): Assessment of genetic diversity and germination pattern in selected cotton genotypes of Pakistan. – Pak. J. Bot. 42: 3949-3956.
[49] Nas, M. N., Bolek, Y., Bardak, A. (2011): Genetic diversity and phylogenetic relationships of Prunus microcarpa CA Mey. subsp. tortusa analyzed by simple sequence repeats (SSRs). – Scientia Horticulturae 127(3): 220-227.
[50] Nei, M. (1972): Genetic distance between populations. – Am. Nat. 106 283-292.
[51] Nei, M. (1973): Analysis of gene diversity in subdivided populations. – Proc. Natl. Acad. Sci. USA 70: 3321-3323.
[52] Nei, M. (1987): Molecular Evolutionary Genetics. – Columbia University Press, New York.
[53] Nei, M., Tajima, F., Tateno, Y. (1983): Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. – J Mol Evol 19: 153-170.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

[54] Noormohammadi, Z., Jesvaghani, F. S., Sheidai, M., Farahani, F., Alishah, O. (2011): Inter simple sequence repeats (ISSR) and random amplified polymorphic DNA (RAPD) analyses of genetic diversity in Mehr cotton cultivar and its crossing progenies. – Afr. J. Biotechnol. 10: 11839-11847.
[55] Paterson, A., Boman, R., Brown, S., Chee, P., Gannaway, J., Gingle, A. et al. (2004): Reducing the genetic vulnerability of cotton. – Crop Sci. 44: 1900-1901.
[56] Pei, Z., Gao, J. Q., Chen, J., Wie, Z., Li, F., Luo, L., Shi, B., Ding, S., Sun, S. (2010): Genetic diversity of elite sweet sorghum genotypes assessed by SSR markers. – Biologia Plantarum 54(4): 653-658.
[57] Pritchard, J. K., Stephens, M., and Donnelly, P. (2000): Inference of population structure using multilocus genotype data. – Genetics 155: 945-959.
[58] Purves, W. K., Sadava, D., Orians, G. H., Heller, H. C. (2003): Genetic drift may cause large changes in small populations. – In: Purves, W. K. et al. (eds.) Life: The Science of Biology. Sinauer, Sunderland, MA, pp. 468-469.
[59] Rana, M. K., Singh, V. P., Bhat, K. V. (2005): Assessment of genetic diversity in upland cotton (Gossypium hirsutum L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. – Genet. Resour. Crop Evol. 52: 989-997.
[60] Reddy, O. U. K., Pepper, A. E., Abdurakhmonov, I., Saha, S., Jenkins, J. N., Brooks, T. D. (2001): New dinucleotide and trinucleotide microsatellite markers for cotton genomic research. – J. Cotton Sci. 5: 103-113.
[61] Rungis, D., Llewellyn, D., Dennis, E. S., Lyon, B. R. (2005): Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (Gossypium hirsutum L.) cultivars. – Australian Journal of Agricultural Research 56(3): 301-307.
[62] Sapkal, D. R., Sutar, S. R., Thakre, P. B., Patil, B. R., Paterson, A. H., Waghmare, V. N. (2011): Genetic diversity analysis of maintainer and restorer accessions in Upland cotton (*Gossypium hirsutum* L.). – J. Plant Biochem. Biotechnol. 20: 20-28.
[63] Seyoum, M., Du, X. M., He, S. P., Jia, Y. H., Pan, Z., Sun, J. L. (2017): Analysis of genetic diversity and population structure in upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. – Journal of Genetics 97(2): 513-522.
[64] Shete, S., Tiwari, H., Elston, R. C. (2000): On estimating the heterozygosity and polymorphism information content value. – Theor. Popul. Biol. 57: 265-271.
[65] Sun, D. L., Sun, J. L., Jia, Y. H., Ma, Z. Y., Du, X. M. (2009): Genetic diversity of colored cotton analyzed by simple sequence repeat markers. – Int. J. of Plant Sci. 170: 76-82.
[66] Tamura, K., Stecher, G., Kumar S. (2021): MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. – Molecular Biology and Evolution. https://doi.org/10.1093/molbev/msab120.
[67] Tatineni, V., Cantrell, R. G., Davis, D. D. (1996): Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. – Crop Science 36: 186-192.
[68] Tyagi, P., Gore, M. A., Bowman, D. T., Cambell, B. T., Udall, J. A., Kuraparthy, V. (2014): Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). – Theor Appl. Genet. 127: 283-295.
[69] Ullah, I., Iram, A., Iqbal, M. Z., Nawaz, M., Hasni, S. M., Jamil, S. (2012): Genetic diversity analysis of Bt cotton genotypes in Pakistan using simple sequence repeat markers. – Genetics and Molecular Research 11(1): 597-605.
[70] Ulloa, M., Saha, S., Jenkins, J. N., Meredith Jr, W. R., McCarty Jr, J. C., Stelly D. M. (2005): Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) join map. – J. Hered. 96: 132-144.
[71] University of California Museum of Paleontology (2016): Bottlenecks and founder effects. – Understanding Evolution. http://evolution.berkeley.edu/evolibrary/article/bottlenecks_01.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953

[72] Van Esbroeck, G. A., Bowman, D. T., May, O. L., Calhoun, D. S. (1999): Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. – Crop Sci 39: 323-328.
[73] Wang, H. M., Lin, Z. X., Zhang, X. L., Chen, W., Guo, X. P., Nie, Y. C., Li, Y. H. (2008): Mapping and quantitative trait loci analysis of Verticillium wilt resistance genes in cotton. – Journal of Integrative Plant Biology 50(2): 174-182.
[74] Weir, B. S. (1996): Genetic Data Analysis II: Methods for Discrete Population Genetic Data. 2nd ed. – Sinauer Associates Inc., Sunderland, MA.
[75] Wendel, J. F., Cronn, R. C. (2003): Polyploidy and the evolutionary history of cotton. – Adv Agron 78: 139-186.
[76] Wendel, J., Brubaker, C., Percival, A. (1992): Genetic diversity in Gossypium hirsutum and the origin of upland cotton. – Am J Bot 79: 1291-1310.
[77] Yeh, F. C., Yang, R. C., Boyle, T. (1999): POPGENE. Microsoft Windows Based Freeware for Population Genetic Analysis. Release 1.31. – University of Alberta, Edmonton.
[78] Yu, J. Z., Fang, D. D., Kohel, R. J., Ulloa, M., Hinze, L. L., Percy, R. G., Zhang, J., Chee, P., Scheffler, B. E., Jones, D. C. (2012): Development of a core set of SSR markers for the characterization of gossypium germplasm. – Euphytica 187: 203-213.
[79] Zhang, J., Stewart, J. M. (2000): Economical and rapid method for extracting cotton genomic DNA. – J Cotton Sci 4(3): 193-201.
[80] Zhang, J., Lu, Y., Cantrell, R. G., Hughs, E. (2005): Molecular markers diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. – Crop Sci 45: 1483-1490.
[81] Zhang H. B., Li Y., Wang B., Chee P. W. (2008): Recent advances in cotton genomics. – Int. J. Plant Genomics 20.
[82] Zhang, Y., Wang, X. F., Li, Z. K., Zhang, G. Y., Ma Z. Y. (2011): Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. – Genet. Mol. Res. 10: 1462-1470.
[83] Zhao, Y. L., Wang, H. M., Chen, W., Li, Y. H. (2014): Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. – PLoS One 9: e86308.
[84] Zhao, Y. L., Wang, H. M., Chen, W., Li, Y. H., Gong, H. Y., Sang, X. H. (2015): Genetic diversity and population structure of elite cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. – Plant Syst. Evol. 301: 327-336.
[85] Zoric, M., Dodig, D., Kobiljski, B., Quarrie, S., Barnes, J. (2012): Population structure in a wheat core collection and genomic loci associated with yield under contrasting environments. – Genetica 140: 259-275.

ELECTRONIC APPENDIX

This manuscript has an electronic appendix with the genotypes' properties.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2005_39393953