Pakistan Journal of Biochemistry and Biotechnology (PJBB) Volume 1, Issue 2 (2020), pp. 51-58

Molecular diversity of Pakistani upland cotton (*Gossypium hirsutum* L.) varieties based on microsatellite markers

Seema Aslam

Institute of Molecular Biology and Biotechnology,
Bahauddin Zakariya University, Multan, Pakistan

Huma Iqbal
Institute of Molecular Biology and Biotechnology,
Bahauddin Zakariya University, Multan, Pakistan

Mujahid Hussain
Institute of Molecular Biology and Biotechnology,
Bahauddin Zakariya University, Multan, Pakistan

Muhammad Baber
Institute of Molecular Biology and Biotechnology,
Bahauddin Zakariya University, Multan, Pakistan

Corresponding author: Muhammad Baber
E-mail: muhammadbabar@bzu.edu.pk

ABSTRACT:

For the management of germplasm and varietal characterization the understanding and estimation of genetic diversity plays a key role. In this study genetic diversity of thirty Pakistani Upland cotton (Gossypium hirsutum) varieties was determined by using thirteen SSR markers as these markers are highly polymorphic and efficient for comparative genome mapping. Total 48 band were produced by all thirteen markers; five markers were monomorphic and eight were polymorphic. Highest polymorphism was 68% that produced by NAU 2083 marker. To find the genetic relatedness between thirty studied varieties UPGMA was used for the construction of dendrogram. The similarity between studied thirty varieties ranged from 38.46% to 100%. Highest PIC value was 0.6484 shown by NAU 2083 and the mean PIC value was 0.2833 which predicted that very low genetic diversity is present among the cotton varieties under investigation.

Key words: Germplasm, Polymorphism, UPGMA, Dendrogram.

I. Introduction

Cotton has significant importance as a fiber crop in the world (Fang et al., 2017). It belongs to family *Malvaceae* and the genus *Gossypium* (Tigga, Patil, Edke, Roy, & Kumar, 2017). It nearby has 53 species, 45 diploid and almost 7 tetraploid (Shim, Mangat, & Angeles-Shim, 2018).The genus *Gossypium* cultivates in arid & semi-arid tropical regions. Four commercially important and cultivated species includes *Gossypium hirsutum L*. (upland cotton), *Gossypium barbadense L*. (Egyptian cotton), *Gossypium arboreum L.* (Asiatic cotton) and *Gossypium herbaceum L.* (Asiatic cotton)

(Renny-Byfield et al., 2016). *Gossypium* genus comprises on both diploid (2n=26) and allotetraploid (2n=4x=52) species. Diploid cotton genome categorized into eight types designated A to G &K. Most important tetraploid cotton specie is *Gossypium hirsutum L*. that carries AADD genome and refers as allotetraploid (2n=52). AADD genome results from polyploidization between A genome (*G.harbaceum*) and D genome (*G. raimondi*) (Jabran, Ul-Allah, Chauhan, & Bakhsh, 2019).

Almost 100 countries cultivate cotton in the world. Upland cotton is the most widely cultivated *Gossypium* specie around the globe, >90% cultivation than other species of the genus *Gossypium* mainly due to its high yield characteristic. Globally almost 70% cotton cultivation in Asia continent followed by Americas 20%, Africa almost 6%, and Europe <2%. Pakistan ranked at 4th in cotton production. It's a major commodity in the world economy (Wendel, Brubaker, Alvarez, Cronn, & Stewart, 2009). Cotton is a valuable crop; it has importance in different fields. Its seed is used to obtain oil that is used to get different products such as vegetable oil, soaps etc. Cotton has great importance in textile industry by providing raw material in order to make variety of products like variety of fabrics, garments, towel, papers etc. Cotton also used for biofuel production. There is a need to make genetic improvement to increase yield and quality by making them resistant to environmental stresses in order to fulfill the desired demands (Jabran et al., 2019).

Molecular markers are segments of the DNA that are found in the whole genome, it is used as a tag. These markers have significant importance in genome studies and genetic diversity analysis. Simple sequence repeats (SSR) markers are repeated sequence of nucleotides ranges 1-6 repeating units and also known as microsatellites. SSRs are highly polymorphic and co-dominant in nature, present in both coding and non-coding regions (Parekh et al., 2016). They exhibit variation among species in a population by providing information regarding allelic variation in genome. SSRs enables to detect polymorphism, as these are highly precise to their target and easily detect genomic variation. SSRs are desired markers that are used for genetic diversity analysis in cotton. Information that is obtained by using these markers is very useful to improve breeding strategies. (Sabev, Valkova, & Todorovska, 2020).

Genetic diversity analysis is important in different aspects. In order to introduce different desirable traits in plants to meet our demands there is a need to have complete information of genome. Genetic diversity analysis enables to find out the genes that leads to food limitation and reducing yield rate. SSRs helps to find out the genetic diversity among different varieties of *Gossypium hirsutum* L. (Ali et al., 2019). Objective of this study is to estimate the genetic diversity among 30 upland cotton varieties. If genetic diversity of different varieties is known we can use highly diverse variety in breading programs to obtain desirable trait and for many other purposes.

II. Materials and Methods

Germplasm of 30 Upland cotton varieties (Table 1) were collected from the Cotton Research Institute Multan, Pakistan. Young leaves samples were collected from the selected cotton varieties for the DNA extraction and packed in the zip lock plastic pack. These packs were placed into the ice box to protect the samples during transportation. In the laboratory these samples were stored at -80^{0}C until the DNA extraction. Wet lab experiments were carried out at the Genome Mapping Laboratory of the Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan.

Table 1: Selected upland cotton varieties utilized for diversity analysis

Serial No.	Varieties	Serial No.	Varieties
1	CIM-473	16	CIM-599
2	CIM-506	17	CIM-600
3	CIM-554	18	CIM-602
4	CIM-573	19	CIM-610
5	Niab Kiran	20	CIM-632
6	AGC-999	21	CIM-707
7	ALSEM-151	22	Cyto-178
8	BH-160	23	Cyto-179
9	BH-184	24	Cyto-608
10	Bt-121	25	FH-112
11	CIM-446	26	FH-118
12	CIM-482	27	FH-142
13	CIM-496	28	FH-901
14	CIM-534	29	FH-Lalazar A
15	CIM-598	30	1UB-13

Extraction of DNA and SSR marker analysis

DNA of selected 30 varieties of upland cotton was extracted from the young leaves following the CTAB method described by Doyle and Doyle (Doyle, 1990). 0.8% agarose gel was used to check the quantity and quality of extracted DNA. After electrophoresis, ultraviolet (UV) light (Trans Illuminator) was used to visualize DNA on the agarose gel. Spectrophotometer was also used to quantify the DNA and 3.5 ng/µL was the final working concentration utilized for 20-µL polymerase chain reactions (PCRs). Thirteen SSR markers (Table 2) were utilized for the analysis of 30 upland cotton varieties. These SSR markers we selected randomly and these markers covered most of the cotton genome. Different SSR primers sequences drawn from different sources NAU primers (Han et al., 2006); JESPR primers (Reddy et al., 2001); CIR primers (Nguyen, Giband, Brottier, Risterucci, & Lacape, 2004) and BNL primers from the Research Genetics Co. (Huntsville, AL, USA, http://www.resgen.com) were used in this study. SSRs are used due to their high polymorphism, abundance, wide genome distribution throughout the genome and co-dominance (Saha et al., 2003).

Polymerase chain reaction (PCR)

For diversity analysis PCR by was performed by utilizing gene-specific primers (Table 2). Total 20µL reaction volume was prepared adding 3.5 µL template DNA, 2.5 µL 10X PCR buffer, 2.5 µL MgCl2, 2.5 µL dNTPs, 1µL forward and 1µL reverse primer, 0.25 µL Taq polymerase, and 6.75 µL PCR water. Gene Amp® system (Applied Biosystems) was used to perform PCR. Total 35 cycles with PCR profile including initial denaturation at 95°C for 5 min, final denaturation for 1 min at 94°C, annealing at 55°C for 1min, extension at 72°C for 1 min, and final extension at 72°C for 10 min were performed. After the completion of PCR reaction products of PCR were separated on 2% agarose gel in the electrophoresis. One hundred base pair ladder was also loaded on the gel for the estimation of size PCR products of upland cotton.

Table 2: SSR primers used for diversity analysis in upland cotton varieties.

SrNo.	Name	5-3 sequence of primers	Length bases	Annealing temperature
1	NAU 2095 F	GGGACACAAACAAAACACAC	20	55ºC
	NAU 2095 R	GGAACTTGAGAACTTGAGG	20	
2	CIR 094 F	ATACCTCCTTTGGCATC	17	55ºC
	CIR 094 R	ATTCAGCAACTTCACACA	18	
3	BNL 3580 F	CTTGTTTACATTCCCTTCTTTATACC	26	55ºC
	BNL 3580 R	CAAAGGCGAACTCTTCCAAA	20	
4	NAU2083 F	AGAAGAGGTTGACGGTGAAG	20	55ºC
	NAU2083 R	TGAGTGAAGAACCTGCACAT	20	
5	NAU 1067 F	GATGCTTTTCTCACCCATTT	20	55ºC
	NAU 1067 R	ACCATAGCCAAAACAAGGAA	20	
6	BNL 3888 F	GCCCACTTGCCTCTTACAG	20	55ºC
	BNL 3888 R	AGCTTTTCCCCTTTCACCAT	20	
7	BNL 3971 F	CACATATTTTTGCCTCACGC	20	55ºC
	BNL 3971 R	TGTGGACCCAAAAAGGAAGA	20	
8	NAU 1072 F	ATCAGAGGTTTCGAACTCCA	20	55ºC
	NAU 1072 R	AATAGGATCCTGTGGCTGTG	20	
9	JESPR 101 F	CCAAGTCAAGGTGAGTTATATG	22	55ºC
	JESPR 101 R	GCTCTTTGTTACTGAAATGGG	21	
10	NAU 2265 F	CAATCACATTGATGCCAACT	20	55ºC
	NAU 2265 R	CGGTTAAGCTTCCAGACATT	20	
11	BNL 2443 F	TTTATTGGTCGGTCTTTGCC	20	55ºC
	BNL 2443 R	TTAGGGTGTTCTTTGGGCAC	20	
12	NAU 1248 F	AATGTCAGCTGCCTATTTCC	20	55ºC
	NAU 1248 R	AAGACAGGCGATGTCATCTT	20	
13	NAU 882 F	ATCATCCATTAGGCACCAAC	20	55ºC
	NAU 882 R	GAGGGGAAGAAGCAGCTAACA	20	

varieties. The electrophoresis bands on the gel were visualized under the UV rays. After confirmation of PCR products on the agarose gel electrophoresis PAGE (poly acrylamide gel electrophoresis) was performed.

Data Analysis

The bands visible on the gel were scored using the codes 1 indicating the presence of an allele, and 0 representing the absence of an allele. Only clear bands were scored and based on the presence or absence of bands polymorphism was calculated. A software power marker was used to calculate the Major Allele Frequency, Allele number, Gene Diversity and Polymorphism Content for all thirteen SSR markers that were used in this study for the amplification of 30 upland cotton varieties. Genetic distance among the studied 30 upland cotton varieties was also calculated by using power marker. Based on the genetic distances a dendrogram of these varieties was constructed by using unweighted pair group method with arithmetic averages (UPGMA).

III. Results

Genetic relationships between thirty cotton varieties were studied by using thirteen SSR markers. Only 8 SSR markers were polymorphic and 5 were monomorphic. NAU 2083 was highly polymorphic with 64% polymorphism. Polymorphism content (PIC), gene diversity, allele numbers and major allele frequency for each primer were determined. The mean numbers of effective alleles, PIC, major allele frequency and gene diversity were 2.7692, 0.7538, 0.2833 and 0.3168 respectively (Table 3). In genetic studies PIC has vast importance in the selection of markers as it calculates the capacity of marker to determine the polymorphism between the individuals of population. In other words, PIC value indicates the marker's quality and for co-dominant markers pic values range from to 1. Markers with PIC > 0.5 are premeditated to be immensely informative and markers having PIC values less than 0.25 are not informative and not recommended for the analysis of genetic diversity (Serrote, Reiniger, Silva, Rabaiolli, & Stefanel, 2020). In this study NAU 2083 showed highest PIC value 0.6484. In the present study low level of polymorphism was found which may be due to low level of genetic variability between cotton varieties as less number of SSR markers used, or insufficient genome coverage.

Table 3: Parameters of genetic diversity in 30 upland cotton varieties with 13 SSR markers.

Marker	Major Allele Frequency	Allele No.	Gene Diversity	PIC
NAU 2095	0.9333	2.0000	0.1244	0.1167
CIR 094	0.4333	6.0000	0.6844	0.6326
BNL 3580	0.6667	4.0000	0.5067	0.4625
NAU 2083	0.4000	5.0000	0.7000	0.6484
NAU 1067	1.0000	1.0000	0.0000	0.0000
BNL 3888	1.0000	1.0000	0.0000	0.0000
BNL 3971	1.0000	1.0000	0.0000	0.0000
NAU 1072	0.5667	3.0000	0.5844	0.5194
JESPR 101	0.5333	3.0000	0.5511	0.4561
NAU 2265	0.7667	4.0000	0.3822	0.3468
BNL 2443	1.0000	1.0000	0.0000	0.0000
NAU1248	1.0000	1.0000	0.0000	0.0000
NAU882	0.5000	4.0000	0.5844	0.4999
Mean	0.7538	2.7692	0.3168	0.2833

Dendrogram and cluster analysis

Dendrogram generated by UPGMA (Figure 1) was divided into three main clusters: 1,2 and 3. According to our study 38.46% to 100% similarity was observed among the studied 30 varieties of upland cotton revealing the moderated overall genetic diversity among these varieties. The genetic distance ranged from 0.000 to 0.6154.

Cluster-1 was divided into subclusters; sub-cluster A and sub-cluster B. Sub-cluster A was further divided into many small clusters. Total 22 verities were present in cluster-1 out of them 21 verities Cyto-178, 1UB-13, Cyto-179, CIM-446, BH-160, Bt-121, BH-184, CIM-573, CIM-554, CIM-482, ALSEM-151, CIM-598, CIM-534, CIM-506, CIM-473, CIM-496, Niab Kiran, AGC-999, AGC-999, FH-Lalazar A, FH-901 and CIM-602 were present in sub-cluster A and only 1 variety FH-142 was present in sub -cluster B. Cluster-2 was divided into two sub-clusters C and D. Sub-cluster C

constituted two varieties CIM-632 and CIM-610 with 100% similarity. Sub-cluster D also comprised 2 varieties CIM-600 and CIM-599 with 100% similarity. Cluster-3 was also divided into two sub-clusters; sub-cluster E and sub-cluster F. Sub-cluster E comprised two varieties FH-118 and FH-112 and sub-cluster F also comprised two varieties Cyto-179 and Cyto-608 with 93% similarity.

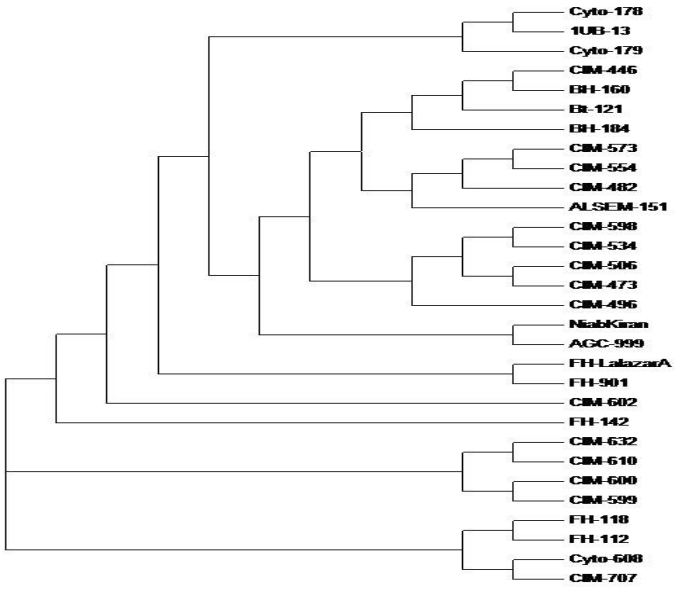

Figure 2: **UPGMA (Unweighted Pair Group Method with Arithmetic Mean) dendrogram showing the genetic relationships among 15 mango genotypes.**

IV. DISCUSSION

Estimation of genetic diversity between plant species gives essential information to develop a variety and to conserve crops (Singh, Mahenderakar, Jugran, Singh, & Srivastava, 2020). Molecular markers play key role in the estimation of genetic diversity and in crop improvement programs. Segments of genes containing useful traits are recognized by these markers (Kesawat & Kumar, 2009). Our research work was designed to estimate the informativeness of the SSR locus of thirty different varieties of Gossypium hirsutum L., and to determine the genetic distances between these varieties. Microsatellite markers are highly preferred for genetic fingerprinting of many plant species due to their highly polymorphic, codominant and reproducible nature (Vieira, Santini, Diniz, & Munhoz, 2016). These markers are used to detect the comparative allelic-variability because of their multi-allelic nature across a large range of germplasm (Kirungu et al., 2018).

In present study 13 microsatellites were utilized to study the genetic diversity among 30 upland cotton varieties. The frequency based genetic distance ranged from 0.000 to 0.6154 indicating genetic similarity from 38.46% to 100%. 5 markers were monomorphic and 8 were polymorphic producing 47 band. A moderate level of genetic diversity was calculated by these markers. NAU 2083 produced highest polymorphism of 64%. Based on SSR analysis a dendrogram comprising 3 main clusters (1, 2 and 3)

is formed having further subclusters. Cluster 1 has many subclusters comprising 22 varieties.

SSR markers always played significant role in the determination of genetic diversity in several plant species. SSR markers were used by (Guang & Xiong-Ming, 2006) to estimate genetic diversity among 43 upland cotton germplasm sources and to study ecological growing regions of china. 36 SSR primers were used that produced 130 gene alleles having 20% monomorphism and 80% polymorphism. Allele numbers ranged 2-8 per primer with mean of 3.6. PIC value ranged from 0.278 to 0.865, with mean of 0.62. Genotype diversity index (H') ranged from 0.451 and 2.039 with an average of 1.102. Genetic similarity coefficient of SSR markers ranged 0.409- 0.865 between germplasm sources with average of 0.610 representing high genetic diversity among studies source germplasm. At genomic level the diversity of base germplasm decreased from 2nd and 3rd breeding period in comparison with 1st period indicating gradual narrowing genetic background of cotton in China. 12 SSR markers were utilized by (Nazish et al., 2017) to estimate the genetic diversity among 15 mango varieties. Out of 12 ten SSR primers were polymorphic and 2 were monomorphic and total 181 band were produced from these primers. Dendrogram was constructed base of SSR analysis. Co-efficient of similarity between accessions ranged from 75% to 100% which indicated inbreeding among few parent cultivars and low genetic diversity.

V. Conclusion

SSR markers utilized in this study are significant to disclose genetic diversity between Upland cotton varieties. An understanding of genetic relatedness and differences between cultivars plays a significant role to select parent plants with desired genes. These genetic informations facilitate the breeding programs such that desired genes may be transferred from parents to progeny to develop a novel cultivar. cotton has narrow genetic bases so there is need to widen the genetic bases of cotton by selecting highly distant parent plants with desired genes. As genetic diversity can be stored in the plant genetic resources (PGR) form like, DNA library, gene bank and the biorepository which preserve genetic stuff for long time period present study will play an important role scientific literature and can be helpful in cotton breeding of these cotton varieties the varieties with high distance can be selected for further breading to develop a genetically improved variety. The development of large number of highly polymorphic microsatellite is required for high quality analysis of cotton crop for the performance of comparative mapping or recognition of quantitative trait loci.

References

Ali, I., Khan, N. U., Gul, S., Khan, S. U., Bibi, Z., Aslam, K., . . . Hussain, I. (2019). Genetic Diversity and Population Structure Analysis in Upland Cotton Germplasm. *INTERNATIONAL JOURNAL OF AGRICULTURE AND BIOLOGY, 22*(4), 669-676.

Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*, 13-15.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., . . . Zhou, B. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature Genetics, 49*(7), 1089.

Guang, C., & Xiong-Ming, D. (2006). Genetic diversity of source germplasm of upland cotton in China as determined by SSR marker analysis. *Acta Genetica Sinica, 33*(8), 733-745.

Han, Z., Wang, C., Song, X., Guo, W., Gou, J., Li, C., . . . Zhang, T. (2006). Characteristics, development and mapping of Gossypium hirsutum derived EST-SSRs in allotetraploid cotton. *Theoretical and Applied Genetics, 112*(3), 430-439.

Jabran, K., Ul-Allah, S., Chauhan, B. S., & Bakhsh, A. (2019). An introduction to global production trends and uses, history and evolution, and genetic and biotechnological improvements in cotton. *Cotton Production. 1st ed. Wiley*, 1-5.

Kesawat, M. S., & Kumar, B. D. (2009). Molecular markers: it's application in crop improvement. *Journal of Crop Science and Biotechnology, 12*(4), 169-181.

Kirungu, J. N., Deng, Y., Cai, X., Magwanga, R. O., Zhou, Z., Wang, X., . . . Liu, F. (2018). Simple sequence repeat (SSR) genetic linkage map of D genome diploid cotton derived from an interspecific cross between Gossypium davidsonii and Gossypium klotzschianum. *International journal of molecular sciences, 19*(1), 204.

Nazish, T., Shabbir, G., Ali, A., Sami-ul-Allah, S., Naeem, M., Javed, M., . . . Aslam, K. (2017). Molecular diversity of Pakistani mango (Mangifera indica L.) varieties based on microsatellite markers. *Genetics and Molecular Research, 16*(2).

Nguyen, T.-B., Giband, M., Brottier, P., Risterucci, A.-M., & Lacape, J.-M. (2004). Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theoretical and Applied Genetics, 109*(1), 167-175.

Parekh, M. J., Kumar, S., Zala, H. N., Fougat, R. S., Patel, C. B., Bosamia, T. C., . . . Parihar, A. (2016). Development and validation of novel fiber relevant dbEST–SSR markers and their utility in revealing genetic diversity in diploid cotton (Gossypium herbaceum and G. arboreum). *Industrial Crops and Products, 83*, 620-629.

Reddy, O. U. K., Pepper, A. E., Abdurakhmonov, I., Saha, S., Jenkins, J. N., Brooks, T., . . . El-Zik, K. M. (2001). New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research.

Renny-Byfield, S., Page, J. T., Udall, J. A., Sanders, W. S., Peterson, D. G., Arick, M. A., . . . Wendel, J. F. (2016). Independent domestication of two old world cotton species. *Genome biology and evolution, 8*(6), 1940-1947. doi: org./10.1093/gbe/evw/129

Sabev, P., Valkova, N., & Todorovska, E. G. (2020). Molecular markers and their application in cotton breeding: progress and future perspectives. *Bulgarian Journal of Agricultural Science, 26*(4), 816-828.

Saha, S., Karaca, M., Jenkins, J. N., Zipf, A. E., Reddy, O. U. K., & Kantety, R. V. (2003). Simple sequence repeats as useful resources to study transcribed genes of cotton. *Euphytica, 130*(3), 355-364.

Serrote, C. M. L., Reiniger, L. R. S., Silva, K. B., Rabaiolli, S., & Stefanel, C. M. (2020). Determining the Polymorphism Information Content of a molecular marker. *Gene, 726*, 144175. doi: 10.1016/j.gene.2019.144175

Shim, J., Mangat, P., & Angeles-Shim, R. (2018). Natural variation in wild Gossypium species as a tool to broaden the genetic base of cultivated cotton. *J. Plant Sci. Curr. Res, 2*(005).

Singh, R. B., Mahenderakar, M. D., Jugran, A. K., Singh, R. K., & Srivastava, R. K. (2020). Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene, 753*, 144800.

Tigga, A., Patil, S., Edke, V., Roy, U., & Kumar, A. (2017). Heterosis and inbreeding depression for seed cotton yield and yield attributing traits in intrahirsutum (G. hirsutum L. X G. hirsutum L.) hybrids of cotton. *International Journal of Current Microbiology and Applied Sciences, 6*(10), 2883-2887.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. d. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology, 39*(3), 312-328.

Wendel, J. F., Brubaker, C., Alvarez, I., Cronn, R., & Stewart, J. M. (2009). Evolution and natural history of the cotton genus. In P. A.H. (Ed.), *Genetics and genomics of cotton* (Vol. 3, pp. 3-22). : Springer, New York, NY.