Original Research Article

Evaluation of Genetic Diversity among CLCuD-Resistant Upland Cotton Varieties Using SSR Markers

I suggest revising the English.

.

ABSTRACT

Cotton leaf curl disease (CLCuD) is a dangerous viral disease that affects the productivity of cotton crops. Since early detection, this disease has spread worldwide and has caused serious damage to the production of cotton around the world particularly in Pakistan. In this study, we evaluated the genetic diversity of CLCuD-resistant upland cotton germplasm by using microsatellite markers. For this purpose, 25 pairs or microsatellite primers were screened and 9 were found to be polymorphic. These 9 primers were used for assessing the genetic diversity of 75 upland cotton varieties. Using these markers, a total of 46 loci were amplified with an average of 5.11 loci per primer. BNL0409 and BNL2835 produced the highest number of bands, 8 bands by each primer. Mean gene diversity was 0.651 with 0.91 (BNL2835) being the highest and 0.38 (JESPR0013) being the lowest. Polymorphism information content (PIC) and major allele frequency ranged between 0.91 to 0.36 (for BNL2835 and BNL0341) and 0.77 to 0.32 (for JESPR0013 and BNL0409) respectively. Mean PIC and major allele frequency were 0.62 and 0.49 respectively. BNL2835 showed the highest allele number of 23 and the mean allele number was 11. Principal component analysis (PCA) was performed and the PCA graph was plotted.These results showed that the 75 varieties, used in this study, using 9 microsatellite markers have an average level of genetic diversity. The findings of this study could be utilized for genetic analysis and the outcomes might be useful for cotton breeding programs.

Keywords: Genetic diversity, CLCuD, microsatellite, upland cotton, PIC, PCA

- Keywords must not be in the title.

1. INTRODUCTION

The agricultural sector is a major contributor to Pakistan's economy, accounting for 22.7% of gross domestic product (GDP) [1]. Among the various crops grown in Pakistan, cotton (Gossypium hirsutum L.) (Malvales: Malvaceae) is of particular importance, as the country's rural-industrial -has a significant impact on utilization In fact, the economic growth of this sector is highly dependent on this sector, which directly and indirectly affects employment [2]. These crops are susceptible to both biotic and abiotic degradation, and sustainable crop production is essential for the sustainability of Pakistan's economy.

The cotton crop faces challenges due to both biotic and abiotic factors. Fluctuating environmental conditions, such as increased temperatures and reduced rainfall, have a

considerable effect on drought occurrence, which has adverse effects on the quality and yield of cotton globally. Various pathogens contribute to diseases in the cotton cultivars, including Bacterial blight, Root rot, Verticillium and Fusarium wilts, and CLCuD. Among these, the Cotton crop is seriously threatened by CLCuV in the subcontinent [3]. Although Gossypium hirsutum, a species of cotton, exhibit some level of tolerance against CLCuV disease [4], it still remains susceptible to these ailments. In recent years, CLCuD has been consistently spotted in numerous countries in Africa and South Asia, particularly in Pakistan, northwestern regions of India, and more recently, in China [5].

CLCuD has cased substantial damage to the cotton productivity of China, India, and Pakistan leading to huge losses to the economy of the countries, out of which, Pakistan is among the top that faced serious losses due to this virus [6, 7]. The symptoms of this disease include thickened veins, curled leaves, and tiny enations formed on the leaves' lower side [3]. The symptoms are used for the characterization of this disease. The viruses responsible for causing this are termed CLCuD-associated Begomoviruses (CABs). These viruses are part of the genus Begomovirus in a family of viruses called Geminiviridae. The transmission of these viruses occurs through the vector white-fly called Bemesia tabaci [8]. There is a broader range of Geminiviridae. One of the special features of this family is that these viruses contain a circular, single-stranded DNA (ssDNA) which is contained inside icosahedral twin particles. These viruses have been divided into nine genera in accordance with their range of hosts, the medium of transmission, and genome organization [9]. The most widespread and diversified of them all is the genus Begomovirus [10]. Microsatellites are the marker preference for many scientists when it comes to genetic diversity and structure analyses, parentage analyses, and pairing system estimations when co-dominance is crucial. Microsatellites, which are sometimes referred to as SSRs or STRs, are commonly regarded as repetitive sequences of between one and six bases that are present within the nuclear and organelles' genetic material of eukaryotic organisms [11]. SSRs offer several advantages in genetic studies. They follow a co-dominant inheritance pattern, meaning that they can be amplified and detected in both heterozygous and homozygous loci. Their advantages include broad genome coverage, abundance, and reproducibility. Their detection is easier by polymerase chain reaction (PCR) technique and offers locus-specific information. These characteristics make SSRs valuable tools in various applications such as genetic mapping, population genetics, DNA fingerprinting, and molecular breeding programs.

In this study, we aimed to evaluate the genetic diversity among seventy-five upland cotton varieties using SSR markers.

- I suggest inserting biographical references from the last five years.

2. MATERIALS AND METHODS

.

This study was conducted at the Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University Multan from June 2022 to June 2023. In this study, genetic diversity analysis of cotton cultivars was done against whitefly resistance using simple

sequence repeat (SSR) markers A total of 75 CLCuD resistant cotton cultivars were used for evaluating genetic diversity. For this purpose, SSR markers were chosen. The names of the varieties used in this study are given in the table below. A total of 25 SSR markers were screened, out of which, 9 markers were found to be polymorphic (Table 1). These 9 markers were proceeded with for further study. DNA from fresh leaves was extracted using a modified CTAB extraction protocol [12]. The markers were used to amplify alleles of the 75 cotton varieties using PCR. PCR was performed using the standard procedure and the PCR products were subjected to agarose gel electrophoresis for bands visualization. The bands were visualized and recorded using GelDoc XR+ in combination with ImageLab software.

Bands data were recorded in the form of binary data matrix using excel. This binary data matrix was used for phylogenetic and statistical analysis. The data was imported in PowerMarker V 3.25 and different parameters were analyzed including major allele frequency, polymorphism information content (PIC), gene diversity and allele number [13]. Following that, genetic distance among the varieties was calculated and a neighbor-joining (NJ) phylogenetic tree was constructed. Principal component analysis (PCA) was performed using OriginPro and PC1 and PC2 were plotted.

Table 1: Cotton varieties utilized for genetic diversity assessment in current research work

Sr. No.	Variety Name	Sr. No.	Variety Name	Sr. No.	Variety Name
1	Barhi-M1	26	Cris-580	51	CIM-482
2	CIM-443	27	CIM-473	52	CIM-496
3	Cris-583	28	DPL New Cotton-33	53	Cris-587
4	Cris-599	29	SLH-13	54	CIM-446
5	Cris-613	30	FH-142	55	Sitar-008
6	SLH-119	31	S-14	56	SLH-317
7	VH-282	32	B-021	57	VH-305
8	VH-363	33	CIM-448	58	NIAB-2009
9	SI Okra1-23	34	TARZAN-2	59	GM-90
10	FH-152	35	NIAB-2010	60	IR-3701
11	Samaru-72	36	Cris-635	61	CA-325 IRABLT
12	AA-802	37	FVH-53	62	Cris-562
13	RH-112	38	CIM-499	63	CIM-1100
14	NIAB-112	39	Hari-Dost	64	SLH-12
15	FBS-30	40	SLS-90/2	65	SLH-06
16	CIM-506	41	LB-391	66	Cris-625
17	Malmal	42	VH-300	67	FH-113
18	Cris-541	43	Chilala-76/2	68	CIM-632
19	CEMB-33	44	SLS-B7/175	69	Barberton

20	LaOkra-541	45	TARZAN-1	70	FBS-37
21	FH-87	46	Gomal-105	71	MNH-886
22	Bt.CIM-602	47	DP-Acala-90	72	Cyto-179
23	Cris-601	48	Cris-628	73	IUB-222
24	BH-160	49	CIM-612	74	Cyto-124
25	S-32	50	MNH-329	75	CIM-554

Table 2: SSR primers' names and sequence used for genetic diversity assessment

Name	Annealing Temp. °C	Direction	Sequence 5' to 3'
JESPR0013	54.7	F	GCTCTCAAATTGGCCTGTGT
		R	GGTGGAGGCATTCCTGCTAAC
BNL2835	57.8	F	AAGATAATCGCCGGCTAGCT
		R	CCGCCAGTTTGCCATAATAT
JESPR0220	54.7	F	CGAGGAAGAAATGAGGTTGG
		R	CTAAGAACCAACATGTGAGACC
JESPR0274	57.8	F	GCCCACTCTTTCTTCAACAC
		R	TGATGTCATGTGCCTTGC
BNL0409	55.9	F	AGAAGTCGGACGTGGAAAGA
		R	GCCGTTTTCCTCAGTGATGT
BNL0387	54.7	F	GAAGGGGAATTTATAGCGGG
		R	AGAGACTCCCGCACTTGAAA
BNL0341	55.9	F	ACCTGGGGTACTTGTCCACA
		R	CCATCCCATTTGTGATACCC
BNL0243	54.7	F	GGGTTTTCTGGGTATTTATACAACA
		R	TCATCCACTTCAGCAGCATC
JESPR0042	57.8	F	CGTTGCCGTCTTCGACTCCTT
		R	GTGGGTGGCTAATATGTAGTAGTCG

I suggest that the methodology be more detailed.

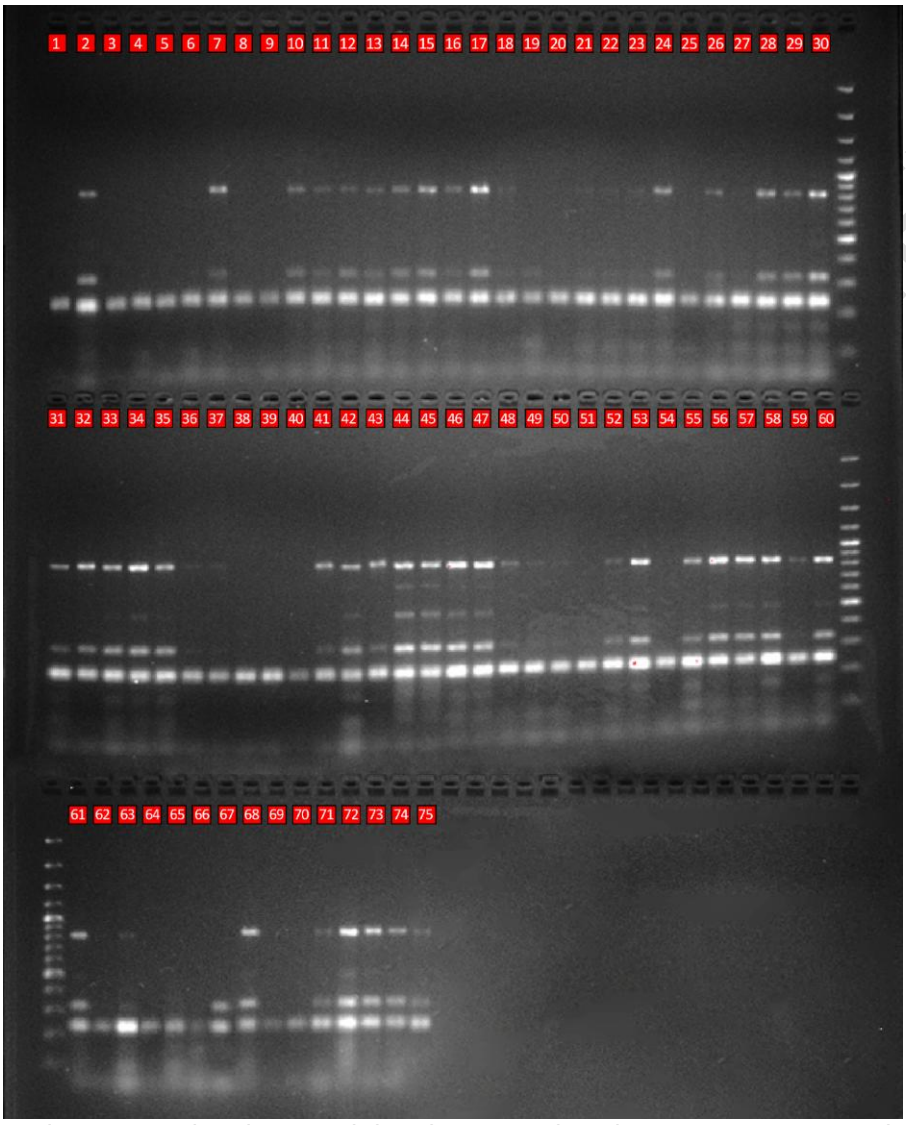
3. RESULTS AND DISCUSSION

To analyze the genetic diversity of upland cotton germplasm, in this research work, we used 9 sets of SSR primers on 75 cotton varieties. These 9 markers amplified 46 loci in selected cotton cultivars. The highest number of amplified by a single primer was 8 which was demonstrated by two markers, BNL2838 (Table 4.1) and BNL0409 respectively. The second

highest number of loci per marker was 7 which were amplified by BNL0243 and the lowest number of loci amplified was 3 by BNL0341 and JESPR0013 respectively.

Figure 1: Agarose gel visualization for BNL2835 on 75 samples

PowerMarker V3.25 was used to evaluate major allele frequency, number of alleles, gene diversity, and polymorphism information (PIC) from gel data. A binary data matrix of gel



results was used as input and the above-mentioned parameters were evaluated. Details of the amplified bands along with other analyzed parameters including major allele frequency, allele number, genetic diversity, and polymorphism information content (PIC) are provided in Table 3.

Table 3: Summary statistics of results including major allele frequency, allele number, genetic diversity, PIC, and number of unique bands.

Marker	Major Allele Frequency	Sample Size	Allele No.	Gene Diversity	PIC	Total Number of Unique Bands
BNL0409	0.3200	75	20	0.8469	0.8347	8
JESPR0220	0.5867	75	5	0.5607	0.4937	3
JESPR0274	0.6800	75	9	0.5077	0.4799	5
BNL2835	0.1467	75	23	0.9170	0.9113	8
BNL0341	0.6933	75	4	0.4405	0.3648	3
BNL0243	0.2933	75	17	0.8473	0.8328	7
JESPR0013	0.7733	75	6	0.3897	0.3749	3
JESPR0042	0.2800	75	11	0.8245	0.8043	4
BNL0387	0.6667	75	7	0.5312	0.5091	5
Mean	0.4933	75	11	0.6517	0.6228	5.1111

OriginPro was used to calculate to perform PCA. PC1 and PC2 had an eigenvalue of 5.098 and 3.535 and 10.85 and 7.52% of the variance. The scree plot and Scores plot of the PCA is provided in Figure 2 and 3 respectively. Comparatively higher eigenvalues of PC1 and PC2 represent the accuracy of PCA and represent their contribution in variance.

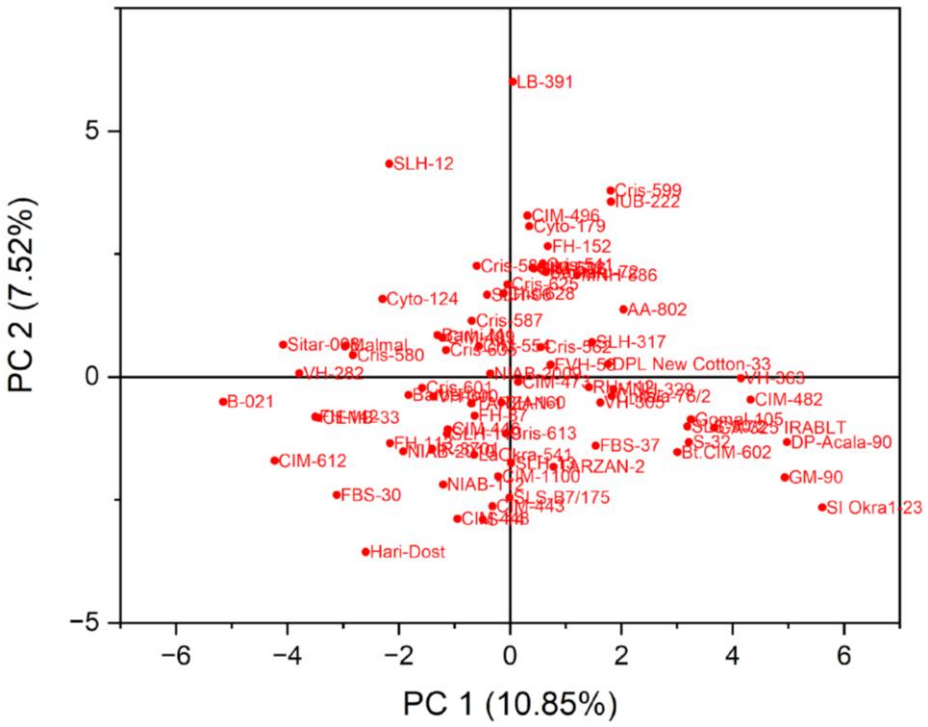


Figure 2: Score plot of PC1 and PC2 from PCA analysis

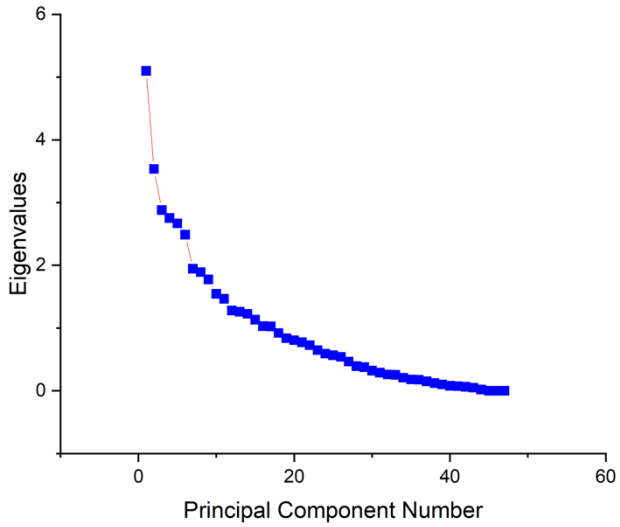


Figure 3: Scree plot of principal components of PCA

Based on the genetic distance matrix obtained, a neighbor-joining (NJ) phylogenetic tree was constructed. A circular graphical representation of this constructed phylogenetic tree is provided below in Figure 4..

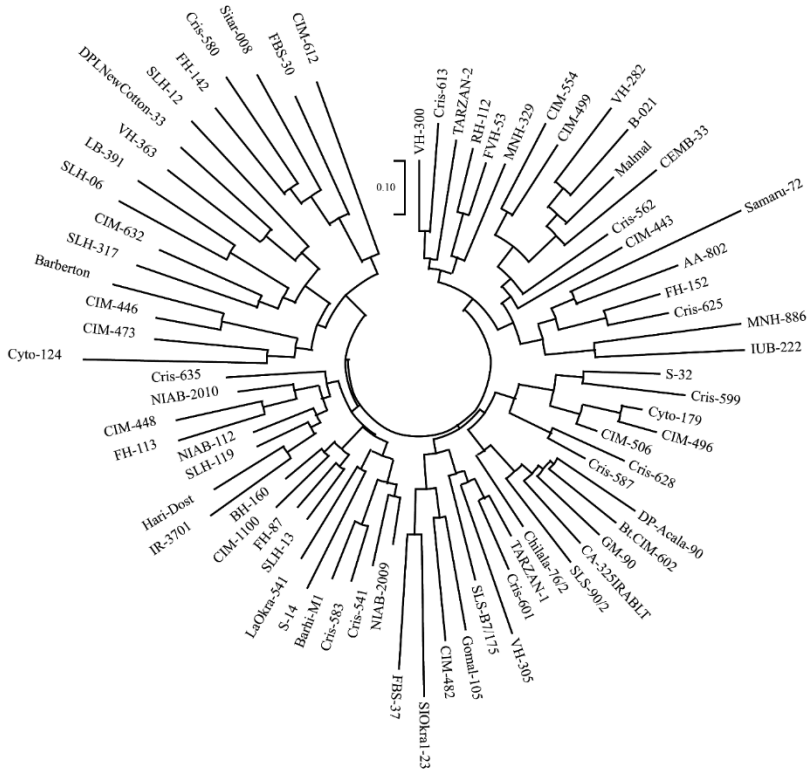


Figure 4.: Circular representation of result from the phylogenetic tree

This phylogenetic tree is divided into three major clusters A, B, and C, each one consisting of 20, 21, and 34 varieties respectively. Cluster A holds two sub-clusters and in this cluster, SAMARU-72 showed the most distant relationship with a genetic distance of 0.35. Following this two varieties, MNH-886 and IUB-222, showed the second-highest genetic distance of 0.28 in cluster A. In cluster B, SIOkra1-23 displayed the highest genetic distance of 0.30 followed by 0.29 for VH-305. Finally in cluster C, having 34 varieties, Cyto-124 showed the highest genetic distance of 0.34 in the first sub-cluster and 0.48 cumulative genetic distance.

Studies like this have been performed in the past which evaluated genetic diversity of cotton germplasm using SSR markers [14, 15]. In this study, we used higher number of varieties which gave more insight to the genetic diversity of upland cotton. On overage, our findings demonstrated higher gene diversity and PIC and lower major allele frequency than the findings of Ali, Khan [14] and Hussain, Farooq [15]. This suggests that the available upland cotton germplasm is more diverse that previous considerations.

Obtained results demonstrate that many varieties of the available cotton germplasm have considerable levels of genetic diversity and can be helpful in breeding programs to develop resistant varieties against CLCuV. The breeding programs can make use of the available germplasm to develop new better varieties which may not only show resistance to pasts and viral infections but may also have favorable characteristics like high yield and stress tolerance. Further genomic analysis, sequencing, transcriptomic and proteomic studies are required to validate these results.

Autors should explore the results further.

4. CONCLUSION

The average gene diversity and PIC values along with PCA analysis conclude higher levels of genetic diversity than what previous studies suggest. This can be helpful for developing improved varieties. Breeders can make use of this genetic diversity to develop new varieties with better morphological traits as well as improved resistance to diseases like CLCuD and other biotic and abiotic stresses. Cross-breeding of varieties with high yield with those showing resistance to CLCuD can produce new varieties with beneficial features from both. Further studies are required to expand the scope of current results.

REFERENCES

Please review the bibliographic references it is important to place them in the journal format.

1. Imtiaz Ahmad HMM, Naila Abbas Dar, Shuaib Malik. *Pakistan Economy Survey 2021-22*, F.D. Economic Adviser's Wing, Editor. 2022
2. Mayee C, Monga D, Dhillon S, Nehra P, and Pundhir P. *Cotton-wheat production system in South Asia: A success story*. 2008

3. Farooq A, Farooq J, Mahmood A, Shakeel A, Rehman KA, Batool A, et al. *An overview of cotton leaf curl virus disease (CLCuD) a serious threat to cotton productivity.* Aust J Crop Sci, 2011. **5**(13): p. 1823-1831
4. Zafar Y, Bashir A, Mansoor S, Saeed M, Asad S, Saeed N, et al. *Cotton leaf curl virus epidemic in Pakistan: virus characterization, diagnosis and development of virus resistant cotton through genetic engineering*. in *Proceedings of the Technical Seminar at the 56th Plenary Meeting of the International Cotton Advisory Committee, Washington DC*. 1997.
5. Cai JH, Xie K, Lin L, Qin B, Chen B, Meng J, et al. *Cotton leaf curl Multan virus newly reported to be associated with cotton leaf curl disease in China.* Plant Pathol, 2010. **59**(4): p. 794-795.doi: 10.1111/j.1365-3059.2010.02266.x
6. Mansoor S, Briddon RW, Zafar Y, and Stanley J. *Geminivirus disease complexes: an emerging threat.* Trends Plant Sci, 2003. **8**(3): p. 128-134.doi: 10.1016/S1360-1385(03)00007-4
7. Zubair M, Zaidi SS-e-A, Shakir S, Amin I, and Mansoor S. *An insight into Cotton leaf curl Multan betasatellite, the most important component of cotton leaf curl disease complex.* Viruses, 2017. **9**(10): p. 280.doi: 10.3390/v9100280
8. Briddon RW and Markham P. *Cotton leaf curl virus disease.* Virus Res, 2000. **71**(1-2): p. 151-159.doi: 10.1016/S0168-1702(00)00195-7
9. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, et al. *Capulavirus and Grablovirus: two new genera in the family Geminiviridae.* Arch Virol, 2017. **162**: p. 1819-1831.doi: 10.1007/s00705-017-3268-6
10. Fauquet C, Briddon R, Brown J, Moriones E, Stanley J, Zerbini M, et al. *Geminivirus strain demarcation and nomenclature.* Arch Virol, 2008. **153**: p. 783-821.doi: 10.1007/s00705-008-0037-6
11. Zane L, Bargelloni L, and Patarnello T. *Strategies for microsatellite isolation: a review.* Mol Ecol, 2002. **11**(1): p. 1-16.doi: 10.1046/j.0962-1083.2001.01418.x
12. Doyle JJ and Doyle JL. *A rapid DNA isolation procedure for small quantities of fresh leaf tissue.* Phytochem Bull, 1987
13. Liu K and Muse SV. *PowerMarker: an integrated analysis environment for genetic marker analysis.* Bioinform, 2005. **21**(9): p. 2128-2129.doi: 10.1093/bioinformatics/bti282
14. Ali I, Khan NU, Gul S, Khan SU, Bibi Z, Aslam K, et al. *Genetic diversity and population structure analysis in Upland cotton Germplasm.* Int J Agric Biol, 2019. **22**(4): p. 669-76.doi: 10.17957/IJAB/15.1114
15. Hussain SB, Farooq L, and Zubair M. *Assessment of Genetic Diversity for Cotton Leaf Curl Disease (CLCuD) and Qualitative Traits among Elite Cotton Cultivars.* Asian J Biotechnol Genet Engineering, 2022.doi: 10.1007/s00299-021-02669-6