*Available online at* www.ijpab.com

**ISSN: 2320 – 7051**
*Int. J. Pure App. Biosci.* **1 (3):** 81-93 (2013)
*Research Article*

**International Journal of Pure & Applied Bioscience**

# Genetic relationship among different species of cotton as revealed by SSR markers for fiber quality traits

**Kundan K. Mishra [1]*and R.S.Fougat [2]**
Department of Agricultural Biotechnology**,** B. A. College of Agriculture,
Anand Agricultural University, Anand – 388 110, Gujarat, India.
*Corresponding Author Email: kkmbt@yahoo.co.in

## ABSTRACT

*The objective of present investigation was to utilize and validate tetraploid species derived EST SSR markers linked to fiber quality traits to generate markers for mapping in diploid cotton species. Six lines of allotetraploid species viz, G. hirsutum (AD1) and seven and eleven lines of its putative diploid ancestors viz, G. arboreum (A2) and G. herbaceum (A1) respectively were used for comparative analysis. Of the total of 41 tetraploid-derived SSR primers, 85.36% gave amplification in both diploid species, indicating that flanking sequences around repeats are conserved in the diploid and tetraploid species of cotton. Average number of bands amplified by each primer was 15.57. The average observed mean heterozygosity was 0.60. DNA sequence information of 10 randomly selected SSR amplicons were used for phylogenetic analysis of Gossypium sps.which revealed that two diploid species of Gossypium (i.e.; G.herbaceum & G.arboreum), shares a close genetic similarity with each other. Tetraploid-derived microsatellites may be useful in comparative genetic mapping of both diploid Gossypium species, for evolutionary studies and marker assisted selection (MAS) for introgression of agriculturally important traits from exotic diploid and tetraploid germplasm sources. BLASTn analysis of randomly selected SSR sequence data showed significant similarity (70-97%) between diploid G. arboreum and G.herbaceum species and (10-27%) between diploid and tetraploids G. hirsutum species. BLASTx analysis of most of SSR markers revealed identity with hypothetical and uncharacterized plant protein.*

***Keywords***- *Cottons, Tetraploid, Diploid, Gossypium hirsutum, Gossypium herbaceum, Gossypium arboreum, SSR markers, BLASTn, BLASTx.*

## INTRODUCTION

The word 'cotton' refers to four species in the genus *Gossypium* (family *Malvaceae*), namely G.*hirsutum* L, G.*barbadense* L, G.*arboreum* L and G.*herbaceum* L. that were domesticated independently as sources of textile fiber. Globally, the *Gossypium* genus comprises of about 50 species (Fryxell, 1992).The genus *Gossypium* occurs naturally throughout tropical and subtropical regions, and includes 50 species split across two ploidy levels, 45 species are diploid (2n=2x=26) and five are allotetraploid (2n=4x=52) [1,2].(Percival and Wendel *et al.*, 1999; Wendal and Cronn, 2003). The haploid genome sizes are estimated to be 880-Mb for *Gossypium raimondii*, 1.75-Gb for *Gossypium arboreum* L, and ~2.5-Gb for *Gossypium hirsutum* L.(Hendrix and Stewart,2005). Among these approximate 50 species four are cultivated viz. *Gossypium arboreum* L. (n=x=13, A2) and *Gossypium herbaceum* L. (n=x=13 A1) are diploids while *Gossypium barbadense* L.[n=2x=26, (AD)2] and *Gossypium hirsutum* L.[n=2x=26, (AD)2] are teraploid Cotton is an important source of natural fiber, which plays an important role in global economy and as a raw material in the textile industry. Globally cotton crops is grown on 32 million hectares (mha) with

appoximately 71% of the production in developing countries. India, USA, China, Australia, Pakistan and Uzbekistan are main producers of cotton in the world. *Gossypium barbadense* L is grown on very less average area (3%) in Tamilnadu and Andhra Pradesh. *Gossypium herbaceum* L is limited to Gujarat and Karnataka. *Gossypium hirsutum* L and *Gossypium arboreum* L are grown in all the major cotton growing states of India.

Enhanced cotton resources are needed to facilitate the improvement of this important crop. Therefore, an important area of cotton genomics is germplasm characterization and utilization. Analysis of genetic diversity and relatedness between species and among genotypes is useful in plant breeding programs because it provides a tool for accurate organization of germplasm and efficient parental selection. Moreover exotic germplasm is an important source of genes which can be transferred in cultivated species through marker assisted selection. With advancement in molecular marker technology, marker-assisted selection (MAS) combined with conventional breeding has been one way in which fiber quality can be improved.

Molecular markers have been widely used in genetic analysis, breeding studies and investigation of genetic diversity and the relationship between cultivated species and their wild parents because they have several advantages as compared to morphological markers, including high polymorphism and independence from environmental effects and physiological stages of the plant. Simple sequence repeat (SSR) markers (microsatellite) are tandem repeated DNA motifs (1-6bp long) which may vary in the number of repeats at given locus, have been successfully employed in many genetic diversity studies (Liu *et al*., 2000; Gutierrez *et al.*,2002) and are useful for a variety of applications in plant genetics and breeding because of their reproducibility, multiallelic nature, codominant, relative abundance and good genomic coverage (Powel *et al*., 1996). SSR are easy to use and analyze (Morgante and Olivieri 1993). The variation among microsatellite is thought to be due to the slippage of DNA polymerase during replication or unequal crossing over, resulting in differences in the copy number of the actual nucleotide sequences (Yu *et al.*,1999). Polymorphism among individuals arises from changes in the number of the repeats. In other words, these markers meet most of the requirements for ideal markers in assessing gene flow. Tracking of microsatellite markers requires specifically designed primers for conserved flanking region of repeats and PCR amplification of this region. The availability and abundance of microsatellite markers throughout the cotton genome coupled with the fact that they are polymorphic, codominant and are based on polymerase chain reaction (PCR) make them particularly useful in genetic diversity studies of cotton (Reddy *et al.,* 2006). With in excess of 1000 microsatellite primers having already been isolated from cotton DNA genomic libraries (Nguyen *et al*.,2004) for molecular studies of the genetic diversity. However, more attention is needed to investigate and compare the genetic diversity of cotton plants cultivated in several countries of Asia with the specific objectives of estimating the informativeness of cotton microsatellite loci and selecting a set of microsatellite primers able to differentiate between the various cultivars studied and to estimate the genetic distance among various cotton cultivars in Asia.

## MATERIALS AND METHODS

### Plant genotypes

The genotypes of cotton were procured from Central Institute of Cotton Research Nagpur, Regional Research Station, AAU Anand, Main Cotton Research Station, Surat, Regional Cotton Research Station Viramgam and Regional Cotton Research Station, Bharuch. Total 24 cotton genotypes were used for study, out of which 11 genotypes were of ***G. herbaceum,*** 7 genotypes were of ***G. arboreum*** and 6 genotypes were of ***G. hirsutum***.

### DNA isolation and quantification

Fresh leaves (3 gm) of young plant genotypes were collected from polyhouse and crushed in the presence of liquid nitrogen. DNA extraction buffer (CTAB) containing 2.5 % polyvinyl pyrrolidone (PVP) was used to isolate DNA. Spectrophotometry was performed to determine DNA concentration by using

Nanodrop N.D.1000 (Software V.3.3.0, Thermo Scientific, USA) at absorbance ratio 260/280 nm and the quality of obtained DNA was checked on 0.8% agarose gel.

**Simple Sequence Repeat (SSR) analysis**

A total of 41 simple sequence repeat primers belonging to JESPR,NAU, CIR, BNL and Gh series were chosen randomly across the cotton genome and downloaded from Cotton Marker Database(CMD). The sequences for these SSRs are available at http://www.cottonmarker.org/projects/cm/. They were custom synthesized by MWG Biotech Pvt. Ltd, Bangalore, India.

**SSR amplification and electrophoresis**

An initial denaturing step of 4 min at 94°C was followed by 42 PCR cycles (denaturing at 94°C for 1 min, primer annealing at 40-60°C for, 1 min and primer extension at 72°C for 2 min). A final step of 7 min at 72°C was carried out for polishing the ends of PCR products. The PCR amplification cycles were carried out in a thermal cycler (T-Gradient, Bio-metra, BIORAD). Agarose gel of 2.5 % concentration was prepared in 1X TBE (2.5 g agarose in 100 ml 1X TBE and 2.5 µl Ethidium bromide from 10 mg/ml stock). PCR amplified products (9 µl and 1 µl 6X loading dye) were loaded into the wells. The 100 bp standard DNA ladder (1 µl) (marker) was also run along with the samples. The electrophoresis was conducted at 100 V current (constant) for 2.5 hrs. to separate the amplified bands. The separated bands were visualized under UV transilluminator (Biometra, Germany) and photographed using gel documentation system (Bio-rad, California).

**Sequencing of SSR amplicon**

Two genotypes one each from *G. herbaceum* (Gcot13) and *G.arboreum* (824) were randomly selected for cycle sequencing for five classes of primer (Amplified by BNL 3482, NAU 1369, NAU 1200, CIR 30 & Gh-G1) used for SSR markers amplicon analysis. Cycle sequencing was performed following the instructions supplied along with Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The reaction was carried out in a final reaction volume of 20 µl using 200 µl capacity thin walls PCR tube. The tubes containing the mixture were tapped gently, spun briefly and then were transferred to thermal cycler. The cycling protocol was designed for 29 cycles with the thermal ramp rate of 1ºC per second. After the reaction, the extension products were purified by Ethanol-EDTA protocol.

**Electrophoresis and Data Analysis**

Electrophoresis and data analysis was carried out on the ABI PRISM® 310 Genetic Analyzer (Perkin-Elmer, Applied Biosystems) using appropriate Module, Base caller, Dyeset/Primer and Matrix files. Sequencing Analyzer Software of Applied Biosystem, provided with Genetic analyzer was used to analyze the raw data obtained through genetic analyzer. By using Seq Scape Software, forward and reverse sequences of each representative sample of each SSR locus were assembled against most closely related reference sequence of respective gene and total length sequence was obtained.

**SSR amplicon sequence analysis.**

The SSR sequencing data of *Gossypium* sps were aligned using ClustalW 2.0 (Larkin et al. 2007) at their default alignment parameter and manually corrected by MEGA 4.0. Phylogenetic analysis was performed with MEGA 4.0 using the Neighbor-Joining method (Saitou et al, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein et al. 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). BLASTn analysis was done for alignment and comparative analysis of diploid cotton SSR marker DNA sequences with other tetraploid cotton DNA sequences along with other species DNA sequences in the NCBI database with standard search statistics (Reddy et al.2006). BLAST x analysis were used for analysis of predicted protein sequence derived from cotton

SSR DNA sequence amplified by respective primer to search the protein databases for similar proteins.

## RESULTS AND DISCUSSION

### SSR analysis for genetic diversity and polymorphism

In the present SSR analysis with 41 microsatellite markers only 35 markers gave the results. All thirty five markers produced 83 alleles. The average number of alleles per locus were found to be 2.47. A maximum of 5 alleles were recorded for marker BNL 4030, JESPR 208 and NAU 923 which were followed by BNL 2920, JESPR 307, NAU 4024 AND NAU 1531 which produced 4 alleles. BNL 1317, BNL 580, CIR30, NAU 1200 and NAU3393 produced three alleles each and BNL1531, CIR 246, CIR 307, JESPR65, NAU 1369, NAU2035, Gh-G1 and Gh-G3 produced 1alleles which was lowest in the present investigation. SSR markers amplified by primer CIR 30, NAU1200, NAU4024, NAU1531, BNL2920, JESPR208, JESPR127 and NAU923 gave high polymorphism and they are also abundant in both diploid (*G. herbaceum,G.arboreum*) genotypes therefore, they are best markers for marker assisted selection (MAS) in both diploid (*G. herbaceum and G.arboreum*) genotypes for fibre quality traits. The highest PIC value was recorded for NAU 923 (0.730489) and lowest for JESPR 29 (0.165289) whereas mean PIC value from all tested microsatellite was 0.465411. The average heterozygosity for 35 SSR markers was 0.60 (Banerjee *et al*, 2007).

Dendrogram Based on Nei's (1978) unbiased measures of genetic distance by UPGMA method formed two major clusters which grouped all 24 cotton genotypes. Cluster 1(B) which was smallest in all clusters included Gvhv 235. Second cluster formed two subgroups viz. A1 and A2. Cluster A1 included two genotypes Lintless DDK and 592. The A2 cluster again formed two sub group viz. A2A and A2B. A2A includes 11 genotypes viz. Gvhv133, G.cotH16, G.cotH10. G.cot20 , G.cot8, CINA329, CINA318, CINA344, Jawar Tapti , CINA333 and DLSA17. A2B includes 10 genotypes viz. G.cotH4 , G.cotH12, 824, G. cot21, Gvhv104, Gvhv505, G.cot13, G.cot13, V-797 and Gvhv544. Gvhv133, G.cotH16, G.cotH10, G.cot20, G.cot8, CINA329, CINA318, CINA344, Jawar Tapti , CINA333 and DLSA17 which leads us to assume that the A2A subgroup cluster included the varieties of similar genetic origin and which can be helpful for consideration of these varieties for the hybridization programme, while the A2B subgroup cluster contained G.cotH4 , G.cotH12, 824, G. cot21, Gvhv104, Gvhv505, G.cot13, G.cot13, V-797 and Gvhv544. The involvement of genotypes in the same cluster again indicated that the varieties with the same genetic composition were falling in the one cluster and this can help to conduct a hybridization programme with these genotypes. The similarity coefficients ranged from 0 to 0.71 for all accessions (Adawy et al, 2007). Thus, the relatedness of the cultivars studied was efficiently established through the use of SSR markers though with some differences in the positioning of some cultivars at various clusters. Dendrograms generated for SSR markers of cotton genotypes reflect relationships among most of the diploids (*G. herbaceum,G.arboreum*) and tetraploids (*G.hirsutum*) cultivars, depending upon their fiber quality traits.

### Primer wise Sequence Homology of SSR sequence between diploid A1 and A2 genome and the tetraploid AD1 Genome.

The SSR sequencing data (*Gossypium* sps.) were aligned using ClustalW 2.0 (Larkin *et al*, 2007) at their default alignment parameter and manually corrected by MEGA 4.0. There were considerable similarities and differences between SSR markers in tetraploid *G. hirsutum* and diploid species (Paterson *et al*,2004). These differences might be due to the genomic distance between the tetraploid and diploid species as well as mismatches during DNA amplification process. DNA sequence alignment for locus BNL 3482, CIR30, Gh-G1, NAU1200 and NAU1369 using a single representative amplicon from each diploid species were done. The SSR sequences from PCR products of the diploid species showed a high degree of identity for sequence with a sequence of the same primer product from the tetraploid cotton obtained from the cotton database. Amplicon variation was observed among sequences derived from both diploid species, which may be attributable to primer mismatch and or allelic variation due to heterozygosity (Jenkins *et al*,

2004). DNA sequence analysis of individual amplicon amplification products confirmed the presence of SSR repeats in both of the diploid species. A high degree of sequence conservation was detected among the flanking regions of diploid and tetraploid species, although single nucleotide differences and various pairwise insertion or deletion polymorphisms distinguished each of the species (Reddy *et al*, 2006).

**Phylogenetic analysis with MEGA 4.0 using Neighbor-Joining method.**

Phylogenetic analysis of *Gossypium* sps. reveals that two species of Gossypium (i.e.; *G.herbaceum* & *G.arboreum),* shares a close genetic similarity with each other. Further, *G.herbaceum* & *G.arboreum* forms three distinct phylogenetic clusters, which might be due to the independent cultivation events of this species. Overall; *G.hirsutum* shows a distant phylogenetic relationship with other two species of Gossypium, which was in accordance with the sequence pairwise mean percentage identity among *Gossypium* sps. ( *G.herbaceum* vs. *G.arboreum,* 72.00% & *G.herbaceum*/*G.arboreum* vs *G.hirsutum* 17.00%).

Phylogenetic analysis of SSR sequencing data (*Gossypium* sps.) was performed with MEGA 4.0 using the Neighbor-Joining method (Saitou et al, 1987). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein et al. 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option).

A dendrogram was generated from 10 amplicon DNA sequences of *G. herbaceum* and *G. arboreum* and DNA sequences of *G. hirsutum* from the cotton database for the same SSR primer . The combined data generated by flanking regions of 5 primer pairs distinguished three species at DNA sequence level. Three group of phylogenic cluster combinations were observed. First, three primer pairs (CIR30,Gh-G1 and NAU1200) produced closely related homologous sequences with less substitutions between *G. herbaceum and G. arboreum* indicating the conservation of this region between these two species. Second, two primer pairs (NAU1200 and NAU1369) exhibited homology between *G. herbaceum* and *G. arboreum* with little variation. Third, another set of three primer pairs (BNL3482 and NAU1369) also targeted a conserved region between *G. herbaceum* and *G. arboreum* with a range of nucleotide substitution. Fourth, there were three major groups of these DNA sequences among three species with a varying numbers of nucleotide substitutions revealing that SSR flanking regions contained a high number of informative substitutions and such variation proved to be particularly useful for distinguishing between very closely related species, such as *G. herbaceum* and *G arboreum.* Dendrograms of SSR amplicon sequence reflect relationships among most of the diploids (*G. herbaceum,G.arboreum*) and tetraploids (*G.hirsutum*) cultivars, depending upon their fiber quality traits.

Pairwise percentage identity profile of *Gossypium* sps. reveals that two species of Gossypium (i.e.; *G.herbaceum* & *G.arboreum),* shares a close genetic similarity with each other for all five primer pairs. Primer pair BNL3482 ,NAU 1200, CIR30, Gh-G1 and NAU 1369 shows 92%,89%, 85%, 58% and 52% similarity respectively between *G.herbaceum* (G.cot13) & *G.arboreum* (824) Further, *G.hirsutum* shows a distant phylogenetic relationship with other two diploid species of *Gossypium arboreum* and *G.herbaceum.* Primer pair BNL3482 ,NAU 1200, CIR30, Gh-G1 and NAU 1369 showed 19%, 27%, 10%, 15% and 10% similarity between *G.herbaceum* & *G.arboreum* (Reddy *et al*,2006).

**BLASTn Analysis:** BLASTn analysis were done for alignment and comparative analysis of diploid cotton SSR marker DNA sequences with other tetraploid cotton DNA sequences along with other species DNA sequences in the NCBI database with standard search ststistics. BLASTn alignment of diploid SSR amplicon sequence with all species shows significant match with tetraploid cotton species as well as other plants and animal species**.** Further BLASTn alignment with only Gossypium species secific alignment has

been done and % identity of sequenced amplicon of both *G herbaceum* & *G. arboreum* species were vary between 70-97% with Gossypium hirsutum. Gossypium species secific alignment clearly reveals that large portions of SSR flanking regions appear to be conserved among diploid and tetraploid genomes of cotton.

**BLASTx Analysis:** BLASTx analysis were used for analysis of predicted protein sequence derived from cotton SSR DNA sequence amplified by respective primer to search the protein databases for similar proteins. BLASTx analysis of most of SSR markers sequence amplified by primer BNL3482, Gh-G1 and NAU1369 reveals identity with hypothetical and uncharacterized plant proteins from different plant species (Arabidopsis, Glycine max, Hordeum vulgare, Oryza sativa Japonica and Zea mays etc). Two other primer CIR30 and NAU1200 SSR sequence donot show any similarity with any predicted, known, Unknown or hypothetical protein.

**Table-1 List of genotypes:**

| Sr. No. | *Gossypium herbaceum* | Sr. No. | *Gossypium arboreum* | Sr. No. | *Gossypium hirsutum* |
|---|---|---|---|---|---|
| 1. | Lintless DDK | 12. | 824 | 19. | G. cot H12 |
| 2. | Gvhv133 | 13. | CINA 329 | 20. | G. cot H4 |
| 3. | 592 | 14. | CINA318 | 21. | G. cot H16 |
| 4. | Gvhv235 | 15. | Jawahar Tapti | 22. | G. cot H10 |
| 5. | Gvhv544 | 16. | CINA344 | 23. | G cot 8 |
| 6. | V-797 | 17. | CINA333 | 24. | G. cot 20 |
| 7. | G. cot 13 | 18. | DLSA17 | ---- | ---- |
| 8. | G.cot 21 | --- | --- | --- | --- |
| 9. | G. cot 23 | --- | --- | --- | --- |

**Table-2 Simple Sequence Repeats (SSR) QTL Primers, amplicon size and repeat length.**

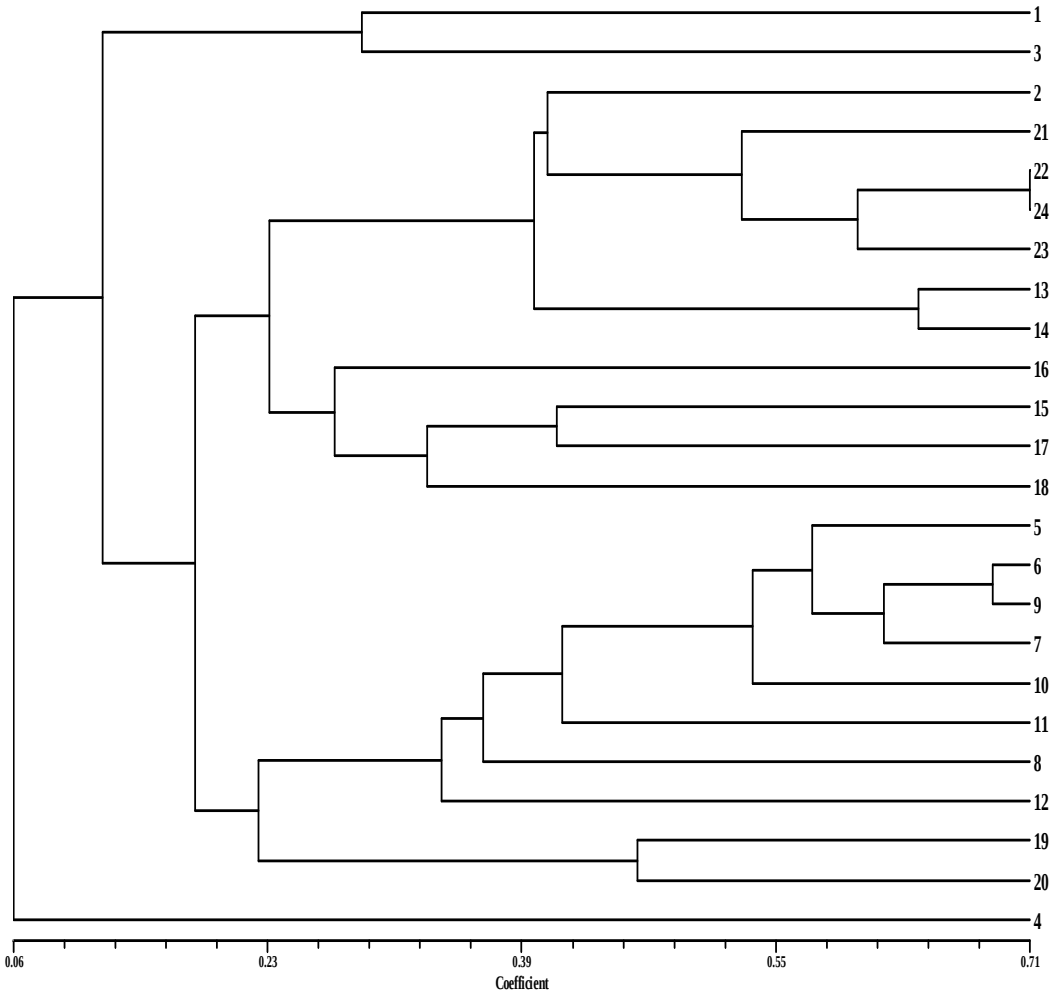
| Sl. No | Primer sequence | Primer | QTL | Amplicon size(bp) | SSR Linked Repeat length |
|---|---|---|---|---|---|
| 1. | TTTGGAGCCATTTACATGCA | **BNL1030 (F)** | QTL | 164-190 | (GT)16 (CA)13 |
| 2. | AAACCACTTCTGCATCTGGA | **BNL1030(R)** | QTL | 164-190 | (GT)16 (CA)13 |
| 3. | AAAAATCAGCCAAATTGGGA | **BNL1317 (F)** | QTL | 154-350 | (AG)14 |
| 4. | CGTCAACAATTGTCCCAAGA | **BNL1317(R)** | QTL | 154-350 | (AG)14 |
| 5. | AAAAACCCCTTTCCATCCAT | **BNL1414(F)** | QTL | 127-197 | (AG)16 |
| 6. | GGGTGTCCTTCCCAAAAATT | **BNL1414(R)** | QTL | 127-197 | (AG)16 |
| 7. | TAATAAAAGGGAAAGGAAAGAGTT | **BNL1231(F)** | QTL | 197 | (AG)15 |
| 8. | TATGGCTCTAGAATATTCCCTCG | **BNL1231(R)** | QTL | 197 | (AG)15 |
| 9. | CCTCCCTCACTTAAGGTGCA | **BNL4030(F)** | QTL | 156-679 | (GT)10 |
| 10. | ATGTTGTAAGGGTGCAAGGC | **BNL4030(R)** | QTL | 156-679 | (GT)10 |
| 11. | CACCATTGTGGCAACTGAGT | B**NL3140(F)** | QTL | - | (GA)11 |
| 12. | GGAAAAGGGAAAGCCATTGT | **BNL3140(R)** | QTL | - | (GA)11 |
| 13. | TCGATTTCCTTATTTGATTTCTG | **BNL3650(F)** | QTL | **76-110** | (TC)15+(TA)6 |
| 14. | AATTTGTCCAGATTCATTCTTCA | **BNL3650(R** | QTL | **76-110** | TC)15+(TA)6 |
| 15. | ATCCAAACCATTGCACCACT | **BNL3408(F)** | QTL | **110-138** | GT2AT(GT)12 |
| 16. | GTGTACGTTGAGAAGTCATCTGC | **BNL3408(F)** | QTL | **110-138** | GT2AT(GT)12 |
| 17. | CATGTTCTAATCATATATATATGTATAT ATATGTGT | **BNL3874(F)** | QTL | **195-328** | (AT)5+G+(TA)4(GT)1 0 |
| 18. | AAAATAACAAAAGCCATGGAATAA | **BNL3874(R)** | QTL | **195-328** | (AT)5+G+(TA)4(GT)1 0 |
| 19. | TTGAGGGCATCCAAATCCAT | **BNL3994(F)** | QTL | - | (CT)25+(GA)25 |
| 20. | CCTCCACCATACACGTGCTA | **BNL3994(R)** | QTL | - | (CT)25+(GA)25 |
| 21. | ATTTGCCCCAGGTTTTTTTT | **BNL3482(F)** | QTL | **230-340** | (AC)12 |
| 22. | GCAACACCTTTTCCTCCCTA | **BNL3482(R)** | QTL | **230-340** | (AC)12 |
| 23. | CTGCAACAAGAGCCTGTGTC | **BNL1531(F)** | QTL | **223** | (AG)14 |

| | | | | | |
|---|---|---|---|---|---|
| 24. | ATGGAGATTGGCTGAGATGG | **BNL1531(R)** | QTL | **223** | (AG)14 |
| 25. | TTCTTGCATTGAATAATACTGGC | **BNL2920(F)** | QTL | **154-340** | (AG)12 |
| 26. | CTTAATTCTAAAAATCAATAA | **BNL2920(R)** | QTL | **154-340** | (AG)12 |
| 27. | CTATGTTTGGCCTTGGCATT | **BNL580(F)** | QTL | **165-210** | (GT)+(GT)+(CT)14 |
| 28. | TAGTGACAGATATCCCCGGC | **BNL580(R)** | QTL | **165-210** | (GT)+(GT)+(CT)14 |
| 29. | ACCTGGGGTACTTGTCCACA | **BNL341(F)** | QTL | **80-125** | (GA)14 |
| 30. | CCATCCCATTTGTGATACCC | **BNL341(R)** | QTL | **80-125** | (GA)14 |
| 31. | TTAGGGTTTAGTTGAATGG | **CIR246(F)** | QTL | **230** | (TG)6 |
| 32. | ATGAACACACGCACG | **CIR246(R)** | QTL | **230** | (TG)6 |
| 33. | ACTAGCAGTGCGAATACA | **CIR45(F)** | QTL | **104-167** | (TG)7 |
| 34. | TGGTTAAGGGTTGGG | **CIR45(R)** | QTL | **104-167** | (TG)7 |
| 35. | TTTCCATCCTTTTGTGA | **CIR381(F)** | QTL | **-** | (AC)7 |
| 36 | AAGGAGAAGAACAAGCAA | **CIR381(R)** | QTL | **-** | (AC)7 |
| 37 | CAATATCTCACTTGGACCT | **CIR030(F)** | QTL | **117-356** | (C)8(TC)6(CA)8 |
| 38 | TGCTACACATCATAGTTGG | **CIR030(R)** | QTL | **117-356** | (C)8(TC)6(CA)8 |
| 39. | TGCATGATGAAGTTAGA | **CIR078(F)** | QTL | **295-336** | (GT)7 |
| 40. | ACATAAATCCCAAGAAC | **CIR078(R)** | QTL | **295-336** | (GT)7 |
| 41. | CTTCATCATAGTAGCGAGTT | **CIR182(F)** | QTL | **-** | (AC)10 |
| 42. | GAATCAAGCAGAGGATTT | **CIR182(R)** | QTL | **-** | (AC)10 |
| 43. | GACTTGAAAAGATTACACAC | **CIR307(F)** | QTL | **235** | (AC)18 |
| 44. | GAATTTGCTGGCTCT | **CIR307(R)** | QTL | **235** | (AC)18 |
| 45. | AACCACCAACCATTCA | **CIR070(F)** | QTL | **156-286** | (AC)8 |
| 46. | TGGGACTCGGTCATC | **CIR070(R)** | QTL | **156-286** | (AC)8 |
| 47. | CCACCCAATTTAAGAAGAAATTG | **JESPR65(F)** | QTL | **316** | (GAA)25 |
| 48. | GGTTAGTTGTATTAGGGTCGTTG | **JESPR65(R)** | QTL | **316** | (GAA)25 |
| 49. | CTTGGCCATGTATTCCTTCA | **JESPR307(F)** | QTL | **94-388** | (TGA)11 |
| 50. | GAAAGACACTAAGCTGAGGC | **JESPR307(R)** | QTL | **94-388** | (TGA)11 |
| 51 | CGCAACCAAACATATACTTCACAC | **JESPR208(F)** | QTL | **116-526** | (CT)15 |
| 52. | CCCTTTCCATCCATAGAACG | **JESPR208(R)** | QTL | **116-526** | (CT)15 |
| 53. | GATTTGGGTAACATTGGCTC | **JESPR127(F)** | QTL | **192-386** | (GA)9AA(GA)5 |
| 54. | CTGCAGTGTTGTGTTGGGTAGA | **JESPR127(R)** | QTL | **192-386** | (GA)9AA(GA)5 |
| 55. | CACCGTTTCCAAGTAAGATT | **JESPR29(F)** | QTL | **110-156** | (CTT)18 |
| 56. | GGTTAATCTTAGTTGAGGTC | **JESPR29(R)** | QTL | **110-156** | (CTT)18 |
| 57. | TTTTGCAGATGTTTGTAGGG | **NAU3260(F)** | QTL | **132-179** | (CT)4 4CA(CT)4 |
| 58. | TTTCTTCAACAGGGGCTAAG | **NAU3260(R)** | QTL | **132-179** | (CT)4 4CA(CT)4 |
| 59. | TTCGGGAAAGTTAGAGGAGA | **NAU1233(F)** | QTL | **-** | (AAT)6 |
| 60 | TCCTCAGAGCTCGGAATAGT | **NAU1233(R)** | QTL | **-** | (AAT)6 |
| 61. | CAACAGCAACAACCACAA | **NAU1200(F)** | QTL | **108-276** | (CAG)11 |
| 62 | CTGCCTCGAGGACAAATAGT | **NAU1200(R)** | QTL | **108-276** | (CAG)11 |
| 63. | CAGCCATCCCTCCTCTAATA | **NAU3393(F)** | QTL | **259-294** | (CTG)7 |
| 64 | GTCAGCAGCCATTCTAACCT | **NAU3393(R)** | QTL | **259-294** | (CTG)7 |
| 65 | ACAAGCATCTTCATGGACCT | **NAU4024(F)** | QTL | **102-569** | (GTC)6 |
| 66 | AGAAGGATGATGCAAAGAGG | **NAU4024(R)** | QTL | **102-569** | (GTC)6 |
| 67 | TTACCAGCAGCCAACACTAA | **NAU3654(F)** | QTL | **96-276** | (TGA)5 |
| 68 | TCCCCTTCAACATCTTCTTC | **NAU3654(R)** | QTL | **96-276** | (TGA)5 |
| 69 | TGGCAGAGATGAATGTAAGC | **NAU1369(F)** | QTL | **369** | (AGGCGG)3 |
| 70. | GGTAACGGATGGAAAATCAC | **NAU1369(R)** | QTL | **369** | (AGGCGG)3 |
| 71. | CGAGAAACTTCACTGGACCT | **NAU2035(F)** | QTL | **282** | (GATA)4 |
| 72 | GAAAAGGTAGGCTTGTTGGA | **NAU2035(R)** | QTL | **282** | (GATA)4 |
| 73 | GGAATTCAAGGTTGAAGGAG | **NAU923(F)** | QTL | **260-524** | (TCTTTT)4 |
| 74 | CCTCTTCTTTGGCTCTGAAA | **NAU923(R)** | QTL | **260-524** | (TCTTTT)4 |
| 75. | GTATCCGCCCACAAATAAAG | **NAU1043(F)** | QTL | **-** | (TTC)14 |
| 76. | GCATCGTGAGAGAAAGTGAA | **NAU1043(R)** | QTL | **-** | (TTC)14 |
| 77. | GCATGCTATCAATTGGGTTTGCATG | **Gh-G1(F)** | QTL | **-** | (CATG)3 |
| 78. | TTGGATTTCCGTTAGCTTCTTGAGC | **Gh-G1(R)** | QTL | **-** | (CATG)3 |
| 79. | GAATTGAAGCAAACCTCATTAATTACC | **Gh-G2(F)** | QTL | **117** | (CTT)12 |
| 80. | CTACCCTCATCTCATTCCAAAAAAC | **Gh-G2(R)** | QTL | **117** | (CTT)12 |
| 81. | CTCAACCTCCCACATGGC | **Gh-G3(F)** | QTL | **121** | (CAA)3 |
| 82. | GCTGCTCACAAGAGGGATG | **Gh-G3(R)** | QTL | **121** | (CAA)3 |

**Table-3 Result of multiple sequence alignment for individual primer amplicon sequence for diploid cotton (*G. arboreum and G.herbaceum)***

| Sample ID | Alignment Length | Alignment length (after alignment gap subtraction) | Conserved sites | Variable sites |
|---|---|---|---|---|
| BLN 3482 | 317 | 114 | 78 | 36 |
| CIR 030 | 114 | 111 | 81 | 30 |
| GhG1 | 543 | 89 | 57 | 32 |
| NAU 1200 | 556 | 111 | 81 | 30 |
| NAU 1369 | 341 | 91 | 47 | 44 |

**Fig.1: Dendrogram showing the genetic relationship between different cotton genotypes based on Nei's (1978) similarity, coefficients using UPGMA as the clustering method for the polymorphism data obtained at 35 microsatellite (SSR) loci.**




**Lintless DDK; 2.Gvhv133; 3.592; 4.Gvhv235; 5.Gvhv544; 6.V-797; 7.G. cot 13; 8.G.cot21; 9.G. cot 23; 10.Gvhv505; 11.Gvhv104; 12.824; 13.CINA 329; 14.CINA318; 15.Jawahar Tapti; 16.CINA344; 17.CINA333; 18.DLSA17; 19.G. cot H12; 20.G. cot H4; 21.G. cot H16; 22.G. cot H10; 23.G cot 8; 24.G. cot 20.**

**Fig .2: Frequency of PIC value of individual primer and their diversity for SSR marker**

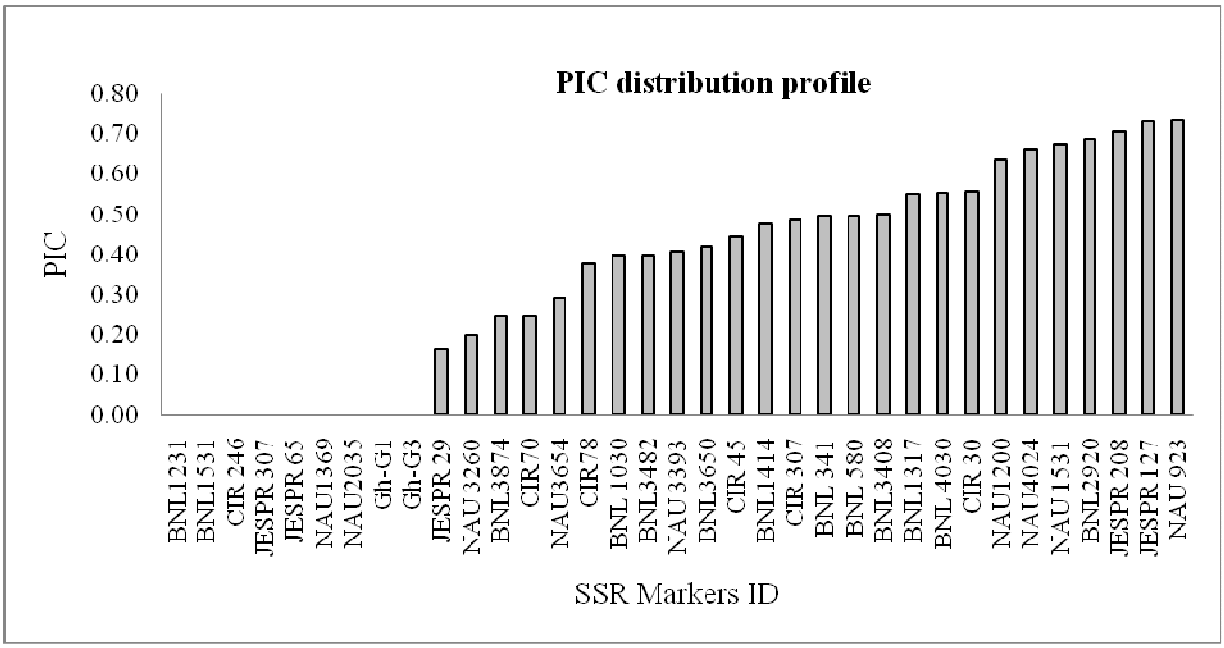
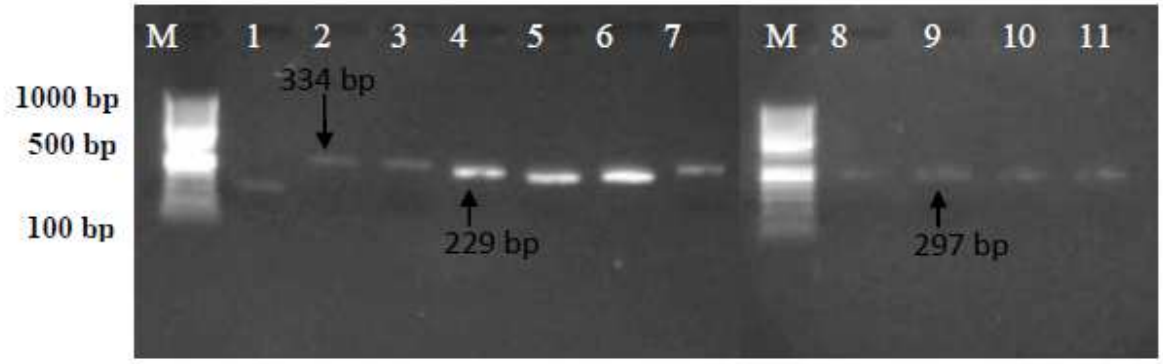



**Fig.3: Gel photograph SSR markers for cycle sequencing (Amplified by BNL 3482, NAU 1369 NAU 1200 CIR 30 Gh-G1).**




**1.Lintless DDK; 2. G. cot 13; 3. G. cot 13 ; 4. G. cot 13 ; 5. G. cot 13; 6. G cot 13;7. 824 ; 8. 824; 9. 824 10. 824; 11. 824**

**Fig.4: Sequence alignment of three species from SSR sequence amplified products using one teraploid derived SSR primer pair (BNL3482). DNA sequence for *G. hirsutum* was obtained from cotton database. DNA sequence identities across three species are indicated with different colour.**

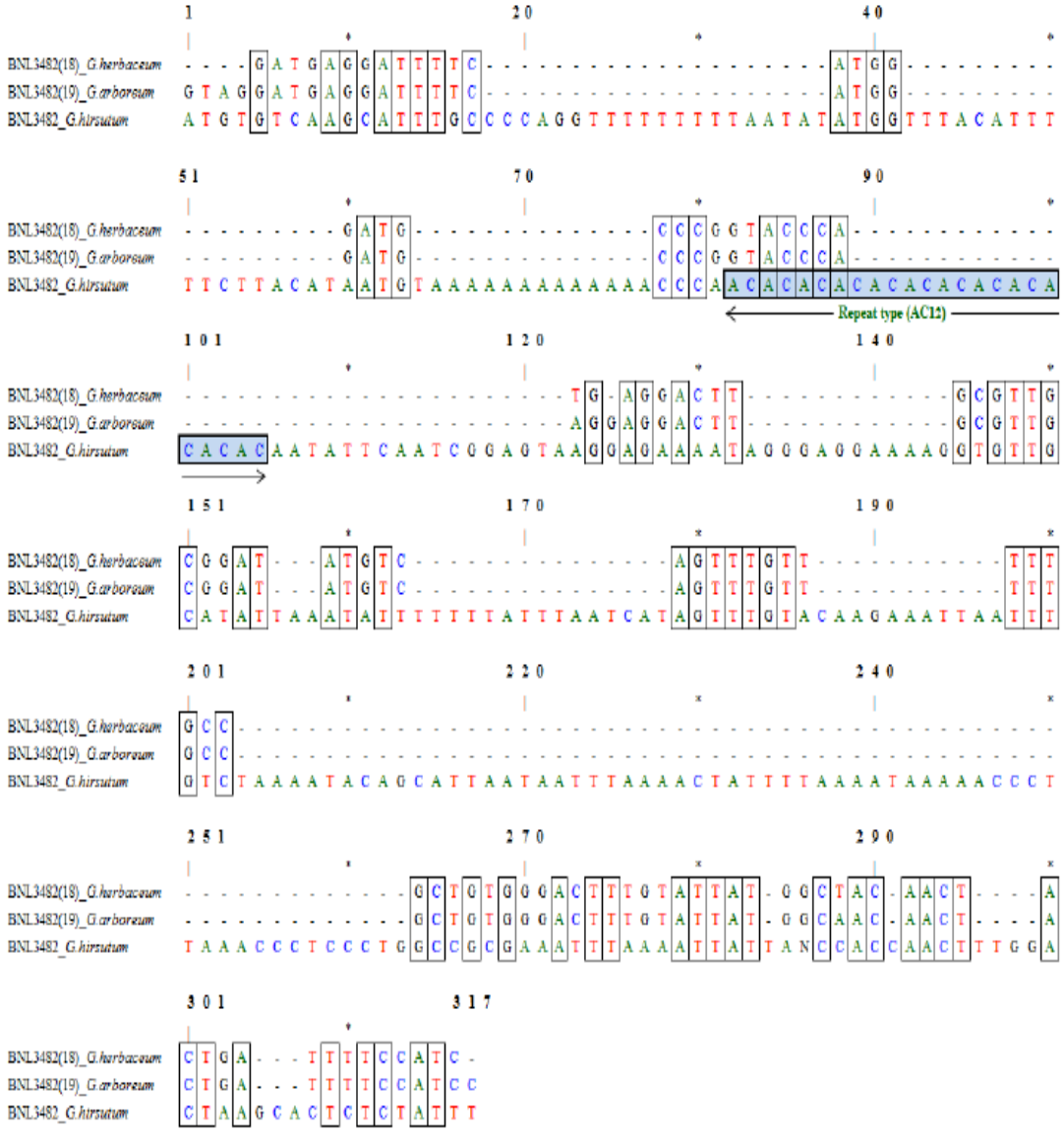
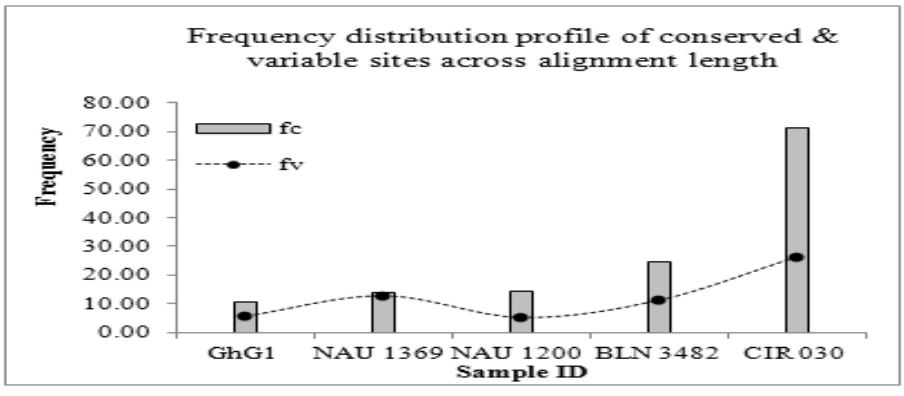



**Fig.5: Result of multiple sequence alignment for frequency distribution of conserved and variables sites across alignment length before gap subtraction.**



## CONCLUSION

From this comparative study between tetraploid and diploid cotton species it could be very well concluded that:

a. Tetraploid-derived EST SSR primers for fibre quality traits could be successfully used in diploid species of cotton for marker assisted breeding.

b. Large portions of SSR flanking regions appear to be conserved among diploid and tetraploid genomes of cotton.

c. SSR markers amplified by primer CIR 30, NAU1200, NAU4024, NAU1531, BNL2920, JESPR208, JESPR127 and NAU923 gave high polymorphism and they are also abundant in both diploid (*G. herbaceum,G.arboreum*) species therefore, they are best markers for marker assisted selection (MAS) in both diploid (*G. herbaceum and G.arboreum*) species for fibre quality traits.

d. Dendrograms generated for SSR markers of cotton genotypes reflect relationships among most of the diploids (*G. herbaceum, G.arboreum*) and tetraploids (*G.hirsutum*) cultivars, depending upon their fibre quality traits which can help to conduct hybridization programme with closely related genotypes.

e. Multiple Sequence alignment between SSR markers of genotypes through clustal W 2.0 reveals considerable similarities between SSR markers in tetraploids (*G. hirsutum)* and diploids *(G. herbaceum* and *G. arboreum)* species.

f. Phylogenetic and Pairwise percentage identity profile analysis of *Gossypium* species reveals that two species of Gossypium (i.e.; *G.herbaceum* & *G.arboreum),* shares very close genetic similarity with each other.

g. BLASTx analysis of most of EST SSR markers sequence amplified by primers reveals identity with hypothetical and uncharacterized plant proteins from different plant species.

## REFERENCES

1. Abdurakhmonov IY, Buriev ZT, Saha S, Pepper AE, et al. Microsatellite markers associated with lint percentage trait in cotton, Gossypium hirsutum. Euphytica, **156:** 141-156 (2007)
2. Adawy, S.S., Ebtissam, H.A. Hussein and A.E. Hanaiya 2006. Molecular characterization and genetic relationship among cotton genotypes 2 AFLP analysis. Arab J. Biotech. **9(3):** 478-492.
3. Bertini CHCM, Schuster I, Sediyama T, Barros EG. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet. Mol. Biol. 29: 321-329 (2006)
4. Brubaker, C. L., A. H. Paterson, and J. F. Wendel Comparative genetic mapping of allotetraploid cotton and its progenitors. Genome, **42:**184-203 (1999)
5. Candida H.C.de, Magalhaes Bertini, Ivan Schuster, Tocio Sediyama, Everaldo Goncalves de Barros and Maurilio Alves Moreira, Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genetics and Molecular Biology, **29 (2):** 321–329 (2006)
6. Chee PW, Rong J, Williams Coplin D, Schulze SR, Paterson AH. EST derived PCR-based markers for functional gene homologues in cotton. Genome, **47:** 449–462 (2004)
7. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang TZ, Guo WZ, Chen XY, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell RG, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Deynze AV, Zhu YX, Yu SX, Abdurakhmonov I, Katageri I, Kumar PA, Rahman M, Zafar Y, Yu JZ, Kohel RJ, Wendel JF, Paterson AH. Toward sequencing cotton (Gossypium) genomes. Plant Physiol **145:**1303–1310. doi:10.1104/pp.107.107672 (2007)
8. Dao-Hua He, Zhong-Xu Lin, Xian-long Zhang, Yi-Chun Nie, Xiao-Ping Guo, Chun-Da Feng and James McD. Stewart. Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. Euphytica, **144:**141-149 (2005)

9. Dongre A B, Bhandarkar Manoj and Banerjee Subha. Genetic diversity in tetraploid and diploid cotton (Gossypium spp) using ISSR and microsatellite DNA markers: Indian Journal of Biotechnology, **6:** 349-353 (2007)
10. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, **39:** 783-791 (1985)
11. Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol Genet Genomics, **275:** 479–491 (2006)
12. Guo WZ, Sang ZQ, Zhou BL, Zhang TZ. Genetic relationships of D-genome species based on two types of EST-SSR markers derived from G. arboreum and G. raimondii. Plant Science, **172(4):** 808-814 (2007)
13. Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B, Zhang TZ. A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in Gossypium. Genetics, **176:** 527–541. doi:10.1534/genetics.107.070375 (2007)
14. 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-SSR in allotetraploid cotton. Theor Appl Genet, **112:** 430–439 (2006)
15. Han ZG, Guo WZ, Song XL, Zhang TZ. Genetic mapping of EST derived microsatellites from the diploid Gossypium arboreum in allotetraploid cotton. Mol Genet Genomics. **272:** 308–327 (2004)
16. Imelfort, M. and Edwards, D. De novo sequencing of plant genomes using second-generation technologies. Brief Bioinform. **10(6):** 609-18 (2009)
17. Kalivas A., Xanthopoulos F, Kehagia O. and Tsaftaris A.S. Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers, Genetics and Molecular Research **10(1):** 208-217 (2011)
18. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG Clustal W and Clustal X version 2.0. Bioinformatics, **23:** 2947–2948 (2007)
19. Lascape JM, Nguyen TB, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, Burr B, Hau B. A combined RFLP-SSR, AFLP map of tetraploid cotton based on a Gossypium hirsutum x Gossypium barbadense backcross population. Genome, **46:** 612-626 (2003)
20. Lascape JM, Dessauw D, Rajab M, Noyer JL, Hau B. Microsatellite diversity in tetraploid Gossypium germplasm:assembling a highly informative genotyping set of cotton SSRs. Mol. Breed. **19(1):** 45-58 (2007)
21. Liu S, Cantrell RG, McCarty JC and Stewart JM. Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *Crop Sci.* **40:** 1459-1469 (2000)
22. Liu, S., S. Saha, D. Stelly, B. Burr, and R. G. Cantrell. Chromosomal assignment of microsatellite loci in cotton. *J. Hered*. **91:** 326-332 (2000)
23. M. Altaf-Khan, Samina N. Qureshi, Sukumar Saha, Johnie N. Jenkins, Curt L. Brubaker and Oumesh U. K. Reddy. Usefulness of SSR derived from Tetraploid Gossypium spp. for analyses of Diploid Gossypium spp, Journal of Crop Improvement,**16(1/2):** (#31/32) (2006)
24. Mei M., Syed N. H., Gao W., Thaxton P.M., Smith C. W., Stelly D. M., Chen Z. J. Genetic mapping and QTL analysis of fiber-related traits in cotton (Gossypium). Theor Appl Genet, **108:** 280-291 (2004)
25. Nguyen TB, Giband M, Brottier P, Risterucci AM, Lascape JM. Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor. Appl. Genet, **109:** 167- 175 (2004)

26. Park YH, Alabady MS et al, Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Gen Genomics **274:** 428–441 (2005)
27. Peng W Chee, Junkang Rong,Dawn Williams-Coplin,Stefan R. Schulze, and Andrew H. Paterson. EST derived PCR-based markers for functional gene homologues in cotton. Genome **47:** 449-462 (2004)
28. Percival A. E., J. F. Wendel, and J. M. Stewart. Taxonomy and germplasm resources. In: Smith, C. W., and J. T. Cothren (eds.), Cotton: origin, history, technology, and production. John Wiley & Sons, Inc., New York, pp. 33-64 (1999)
29. Powell W. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. **1:** 215–222 (1996)
30. Qureshi Samina N, Saha Sukumar, Kantely Ramesh v.Jenkins. J. N. EST-SSR, A New Class of Genetic Markers in Cotton. *The Journal of Cotton Science* **8:**112-123 (2004)
31. Rakshit Aruna, Rakshit S., singh J., Chopra S. K.. Balyan H. S., Gupta P.K and Bhatt Shripad R. Association of AFLP and SSR markers with agronomic and fiber quality traits in Gossypium hirsutum L. Journal of Genetics **89 (2):**( 2010)
32. Rohlf F.J. 1998. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1, User Guide. Exeter Software, New York.
33. Reddy Oumesh U. K. ,Khan M. Altaf, Quresh Samina Ni, Saha Sukumar, Johnie N. Jenkins, Brabaker Curt L, Usefulness of SSR Derived from Tetraploid Gossypium spp. for Analyses of Diploid Gossypium spp. *J. Crop Imp*., 1**6:**1-2, 1-20 (2006)
34. Sajid-ur-Rahman and Tanwir Ahmad Malik. Genetic Analysis of Fiber Traits in Cotton. *Int. J. Ag. & Bio.* ISSN Print: 1560-8530 (2008)
35. Saitou N & Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4:**406-425 (1987)
36. Shen X, Guo W, Zhu X, Yuan Y, Yu J, Kohel R, Zhang T 2005. Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol Breed 15:169–181.
37. Tamura K, Nei M & Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101:11030-11035 (2004)
38. Thomas W, Schlagenhauf E, Graner A, Close T, Keller B and Stein N. 454 sequencing put to the test using the complex genome of barley. BMC Genomics, **7:** 275 (2006)
39. Wangzhen, Caiping Cai, Changbiao, Zhiguo Han, Xianliang Song, Kai Wang, Xiaowei Niu, Cheng Wang, Keyu Lu, Ben Shi and Tianzhen Zhang. A Microsatellite-Based, Gene-Rich Linkage Map Reveals Genome Structure, Function and Evolution in Gossypium. Genetics **176:** 527-541 (2007)
40. Zheng-Sheng Zhang, Mei-Chun Hu, Jian Zhang, Da-Jun Liu, Jing Zheng, Ke Zhang, Wei Wang, Qun Wan. Construction of a comprehensive PCR-based marker linked map and QTL mapping for fiber quality traits in upland cotton (Gossypium hirsutum L.). Mol Breeding DOI 10.1007/s11032-009-92711 (2009)
41. Zhiguo Han, Changbio Wang, Xianliang Song, Wangzhen Guo, Jinying Gou, Chunhong Li, Xiaoya Chen and Tianzhen Zhang. Characteristics, development and mapping of Gossypium hirsutum derived EST-SSR in allotetraploid cotton. Theor Appl Genet 112: 430-439 (2006)
42. Z.-G. Han, W.Z. Guo, X.L. Song, T.Z. Zhang. Genetic mapping of EST derived microsatellite from the diploid Gossypium arboreum in allotetraploid cotton. Mol Gen Genomics. **272:** 308-327 (2004)
43. Zhang J, Lu Y, Cantrell RG, Hughs E. Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. Crop Sci **45:**1483–1490 (2005)