Pak. J. Agri. Sci., Vol. 53(2), 283-290; 2016
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
DOI: 10.21162/PAKJAS/16.2300
http://www.pakjas.com.pk

PHYLOGENETIC ANALYSIS OF COTTON SPECIES (DIPLOID GENOMES) USING SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) MARKERS

Tayyaba Shaheen[1,2], Yusuf Zafar [1] and Mehboob-ur-Rahman[1*]

[1]Plant Genomics and Molecular Breeding (PGMB) Labs., National Institute of Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan Atomic Energy Commission, Islamabad, Pakistan; [2]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan.
[*]Corresponding author's e-mail: mehboob_pbd@yahoo.com

Genus cotton has total 50 species, diploid species fall into 8 genomic groups (A–G, and K). To the extent of our knowledge, frequency of single nucleotide polymorphisms (SNPs) has not been calculated in genes of multiple genomes of the genus *Gossypium*. Here we present the frequency of SNPs in the transcribed regions of the multiple genes, and their utility in resolving phylogenies among 11 diploid species representing five of the eight diploid genomes of the genus Gossypium. We explored the expressed sequence tags (ESTs) data set of *G. arboreum* showing homology with genes encoding for mitochondrial small heat shock protein (MT-sHSP), histone H_2B_1, S-adenosyl methionine synthetase, cytochrome p450, actin-depolymerizing factor 2, C-terminal domain of helicases and histone H_2B_3 for designing primers. The resultant PCR products amplifying partial gene sequences were sequenced. In total, 61 SNPs were detected in cotton genomes which include 53 substitutions and 8 Indels in the total 1920 bp genome length. Phylogenetic analysis using this data revealed grouping of genomes comparable with previous studies. A_1 and A_2 genomes were most similar (98.0%) while D_9 and Oryza genomes were least similar (31.2%). *Oryza* was most distantly related with other genomes. In conclusion SNPs are potent markers to delineate cotton genomes according to their evolutionary positions.

Keywords*:* SNPs, genetic diversity, phylogeny, comparative genomics, *Gossypium.*

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most abundant source of variations in plant and animal genomes (Chen *et al.,* 1999; Ayeh, 2008; Duran *et al.,* 2009; Bansal *et al.,* 2010; Huang *et al.,* 2012; Huang *et al*., 2016) which are present in coding as well as non-coding regions of the genomes (Aerts *et al.,* 2002). Their distribution is random, and sometimes associated with morphological traits (Lindblad-Toh *et al.,* 2000; Gamazon *et al.,* 2010). SNPs are useful for characterizing allelic variation, QTL mapping, and implementing marker-assisted selection (MAS). For example, in *Arabidopsis*, more than one million non-redundant SNPs were identified which can be used in disequilibrium mapping studies (Clark *et al.,* 2007). Similarly, higher plants like barley (Kanazin *et al.,* 2002; Duran *et al.,* 2009), maize (Tenaillon *et al.,* 2001), soybean (Zhu *et al.,* 2003) and sugar beet (Schneider *et al.,* 2001) have been surveyed for discovering SNPs.

Genetic improvement of cotton fiber production and its properties will be enhanced by the availability of rapidly developing genetic resources and tools, including high-density genetic maps (Rong *et al.,* 2004; Lacape *et al.,* 2005). Very few studies have been reported on identification of SNPs in cotton because of huge genome size coupled with the polyploid nature of cultivated cotton which requires the distinction of allelic SNPs from paralogs. A total of 94 SNPs including 36 single-base changes (38.3%) and 58 indels (61.7%) were identified in 16 fiber gene fragments of *G. hirsutum* and *G. barbadense* by using a PCR based direct sequencing technique (Lu *et al.,* 2005). In the *FIFI* gene, regulating the fiber development in *G. barbadense*, three base substitutions were reported while comparing with the corresponding gene in *G. hirsutum* (Ahmad *et al.*, 2007). In another study, in six *R2R3-MYB* transcription factors influencing trichome length and density, one SNP per 77 bases were reported (An *et al.,* 2008). ESTs are a good source for SNPs detection in different plant species. A total of 10,000 SNPs and indels were reported among the ESTs sequences derived from *G. hirsutum* and its progenitor's species (Udall *et al.,* 2006).

In the present study, we present the phylogeny analysis of different diploid cotton genomes by comparing the sequences of the conserved coding regions of seven loci, which are least prone to mutations (Koornneef *et al.,* 2004), otherwise are difficult to study with conventional DNA marker system (Semagn *et al.,* 2006) because the SNPs are most effective to find the single base variations in any part of the genome as compared to other conventional markers (Ball *et al*., 2010). Therefore, a study was conducted to harness SNPs and Indels for studying phylogenetic relationships of cotton genomes and these sequences were also compared with the

corresponding sequences of the sequenced plant genomes. This study will lead to the conclusion how variations at single nucleotide level depict the phylogenetic backgrounds of genomes and what is their role in pace of evolution of different genes. Results obtained with other DNA markers like RAPD, SSRs, AFLP would be compared with the results obtained with SNPs. The present study would set a stage towards exploring the new horizons to scrutinize cotton genes involved in evolution of vital traits.

MATERIALS AND METHODS

***Plant material*:** Experimental material consisted of 11 diploid species (Table 1). Leaves of the cotton species were collected from CCRI Multan.

Table 1. Diploid cotton species (2n=2x-26) used in the present study.

Sr#	***Species name***	***Genome***	***Distribution***
1	*G. arboreum*	(A_2)	Old world
2	*G. herbaceum*	(A_1)	Africa
3	*G. robinsonii*	(C_2)	Australia
4	*G. sturtianum*	(C_1)	Australia
5	*G. aridum*	(D_4)	Mexico
6	*G. raimondii*	(D_5)	Peru
7	*G. gossypoides*	(D_6)	Mexico
8	*G. lobatum*	(D_7)	Mexico
9	*G. laxum*	(D_9)	Mexico
10	*G. stocksii*	(E_1)	Arabian peninsula
11	*G. nelsonii*	(G_3).	Australia

***Isolation of total genomic DNA*:** DNA was extracted from 10 individual plants of each genotype from all 11 species according to the method used by Iqbal *et al.* (1997). After RNase treatment the DNA concentration was measured by flouremeter DyNA QuantTM 200. DNA quality was checked by running 25 ng DNA on 0.8% agarose gel The DNA samples, which were not showing a discrete band, were rejected. The DNA was diluted in double distilled water to a concentration of 15 ng/µl for PCR analysis.

***Primer designing*:** Gene specific primers were designed based on conserved regions of ESTs showing homology with genes encoding for MT-sHSP, Histone H_2B_1, S-adenosyl methionine synthetase, Cytochrome p450, Actin-depolymerizing factor 2, C-terminal domain of helicases and Histone H_2B_3. Three ESTs for each gene were retrieved from model sequenced organisms (Arabidopsis, rice) to design primers. Primers were designed to amplify conserved partial gene sequences using primer 3 software on the basis of regions spanning the conserved regions. (Table 2) Polymerase chain reaction (PCR) was performed in a total volume of 20µl, using 2.5µl (15ng/µl) of cotton DNA, 10 x PCR buffer without $MgCl_2$ (10mM Tris-HCl, 50mM KCl, PH 8.3), 3mM $MgCl_2$, 0.1mM each of dATP, dGTP, dCTP and dTTP and 0.5 units of *Taq* DNA polymerase, 0.15 mM of each primer. *Taq* DNA polymerase together with 10 x PCR buffer, $MgCl_2$ and dNTPs were from MBI Fermentas. Polymerase chain reaction consisted of initial denaturation of 94°C for 5 min and 35 cycles of 94°C for 1 min, 50°C for 30 sec, 72°C extension for 1 min and final extension at 72°C for 10 min. PCR products were resolved on 1% agarose gel to check amplification.

***Sequencing of PCR product*:** Sequencing of PCR products was done on ABI automated DNA sequencer. Sequences were edited manually to get single read of whole sequence. Four runs of sequencing for each of the PCR product were done to avoid discrepancies in SNP detection. Those SNPs which were detected in all sequencing results were considered valid. Consensus sequence of each of the product was used for alignment and phylogenetic analysis using DNA star.

Table 2. Homology of ESTs used in study and Primer sequences.

EST	**Best Blast homology**	**Organism showing best homology**	**Primer sequence**
CON_005_03587	Histone H2B-3	*Arabidopsis thaliana*	5´GCTTAGCCAATTCACCAGGC 3´ 5´GCTCCAAGGTTGGTGAGAAG3´
CON_001_09243	NADPH-cytochrome P450 reductase	*Pisum sativum*	5´CACCCATTTAACCCTTCTCGC3´ 5´TGTATGTGTGTGGTGATGCC3´
CON_003_02307	C-terminal domain of helicases	*A. thaliana*	5´TGGCCTTATCTCCGTCACTC3´ 5´GAAGCGAAAGACCCTCGAAG3´
CON_002_01213	Mitochondrial small heat shock protein	*Solanum lycopersicum*	5´CTCTCCATCACCCTAAAG3´ 5´GTGGAACAGAACACACTC3´
CON_001_10630	S-adenosyl methionine synthetase	*Populus trichocarpa*	5´CCAATGTGATGAAGCTCC3´ 5´GGTGTACCTGAACCATTG3´
CON_008_04131	Histone H2B1	*Medicago truncatula*	5´AGAGAAGAAGCCTAAGGC3´ 5´TCACCAGGAAGTACAAGC3´
CON_006_03927	Actin-depolymerizing factor 2	*Populus trichocarpa*	5´CAACCGAAAGCTATGAGG3´ 5´TGTAGGAAGGAAGGAAGC3´

Sequences of already sequenced genomes *Arabidopsis thaliana, Carica papaya, Vitis vinifera, populus trichocarpa* and *Oryza sativa* were obtained from GenBank for all the seven loci. Maximum similarity of all cotton patial gene sequences sequences was searched with BLAST search tool in NCBI (Blastn).

***SNPs detection, similarity matrix and phylogenetic analysis*:** Seven gene sequences from each of eleven diploid species were used for SNP detection. The Genbank sequences of *A. thaliana, Carica papaya, Vitis vinifera, populus trichocarpa* and *Oryza sativa* were used as outgroup species. DNASTAR (DNASTAR Inc., Madison, WI, USA) and Clustal v were used for sequence alignment. Phylogenetic analyses were performed using Megalign DNASTAR. Similarity matrix and phylogenetic analysis was performed based on cumulative sequences of all seven partial gene sequences amplified from cotton and retrieved sequences from five sequenced genome.

***Chromosomal assignment*:** Positions of the genes were known on *Arabidopsis* chromosomes. This information was used following the comparative genomics of Arabidopsis and cotton (Rong *et al*., 2005) to find their position on cotton chromosomes.

RESULTS

***Comparison with other genomes*:** Sequences of the genes obtained from cotton (*G. arboreum*) were compared with reported corresponding gene sequences from other plant species (Table 2). Two sequences have shown maximum similarity with *A. thaliana*, two have shown with *populus*, while others with *Pisum sativum*, *Medicago truncatula*, *Solanum lycopersicum* and *Vitis vinifera* (Table 2).

***Types and distribution of SNPs within Gossypium diploid genomes*:** In total 61 SNPs were detected in 1920 bp (Total cumulative sequence length obtained by amplification of partial gene sequences in cotton) (out of these, 53 were substitutions and 8 were Indels. A total of 39 (73.5%) substitutions were transitions and 14 (26.5%) substitutions were transversions. Maximum SNPs (16) were detected in S-adenosyl methionine synthetase proteins gene and minimum (2) in gene encoding for C-terminal domain of helicases (Table 3).

***Assignment of the position of genes on Arabidopsis and cotton chromosomes*:** All genes included in the study are single copy genes. Position of these genes on *A. thaliana* chromosomes are given in Table 4. Position of these genes on *Gossypium* chromosomes according to comparative genomics of cotton and *Arabidopsis* by Rong *et al*. (2005) were determined (Table 4). Two of the genes were present near central region of chromosomes in Arabidopsis while four were present at distal ends. Mitochondrial small heat shock protein is a mitochondrial gene and position of Ubiquitin extension protein could not be determined.

***Similarity matrix*:** Similarity matrix studies for all partial gene sequences spanning total 1920bp revealed similarity among genomes ranging from 31.2% (D_6*Oryza) to 98.0%

Table 3. Position and nature of SNPs identified.

EST	Conserved region	Amplicons size	No. of SNPs/ Indels	SNP/ Indel location/ position	SNPs with A. thaliana
Mitochondrial small heat shock protein	138-440	300	13	152 T→C, 177 G→A, 178 –G 218 C→A, 246 –C, 332 A→C 336 T→C, 345 C→T, 346 –T 347 –C, 363 G→A, 361 T→C 394 A→G	101
S-adenosyl methionine synthetase	43-300	260	16	86 A→C, 89 G→A, 101 C→T 109 –G, 125 –C, 130 A→C 136 A→C, 149 A→C, 154 C→T 160 C→T, 186 C→T, 200 C→G 227 C→A, 228 A→C, 272 T→C 276 G→T	41
Histone H2B1	194-460	270	9	203 G→A, 230 A→G, 231 A→G 344 T→C, 382 –C, 388 G→A 410 C→T, 425 T→C, 458 G→A	50
Cytochrome p450	251-500	250	5	375 C→T, 376 A→T, 426 G→A 422 –C, 494 G→A	No similarity
Actin-depolymerizing factor 2	294-490	300	4	327 T→C, 348 T→C, 443 T→C 469 C→T	103
Histone H_2B_3	111-400	300	12	140 C→T, 155 C→T, 175 A→T 187 A→C, 189 T→C, 272 T→C 290 T→C, 338 T→C, 356 A→G 365 T→C, 383 G→A, 392 T→A	50
C-terminal domain of helicases	51-400	350	2	101 C→T, 120 C→T	50

Table 4. Location of genes on chromosomes of Arabidopsis and cotton.

Gene	Location on Arabidopsis	Expected location in cotton	Location on Arabidopsis chromosome
Histone H_2B_3	Chr. 2	Chr. 6	Distal end
cytochrome P450 reductase	Chr. 4	Chr. Not known	Near to centre
C-terminal domain of helicases	Chr. 1	Chr. 5 or 9	Distal end
S- adenosyl methionine synthetase	Chr. 1	Chr. 5 or 9	Distal end
Actin depolymerizing factor 2	Chr. 3	Chr 6	Distal end
Histone H_2B_1	Chr. 1	Chr. 5 or 9	Near to centre

Table 5. Similarity matrix for the cotton genotypes and plant species used in study.

	A1	A2	C1	C2	D4	D5	D6	D7	D9	G3	E1	*A. thaliana*	Populus	Papaya	Vitis	Rice
A1	100															
A2	98.0	100														
C1	77.8	75.7	100													
C2	75.5	74.0	92.4	100												
D4	73.9	70.0	64.2	76.3	100											
D5	76.0	77.4	69.3	78.8	68.2	100										
D6	71.6	73.3	80.4	55.4	72.0	86.9	100									
D7	62.1	64.5	61.8	65.9	76.3	75.3	74.9	100								
D9	65.6	63.3	55.9	56.7	77.1	73.2	76.0	83.2	100							
G3	82.6	84.5	77.4	75.7	65.1	63.8	66.5	67.5	67.8	100						
E1	73.3	76.6	73.9	72.5	72.0	68.1	77.1	69.7	57.1	66.4	100					
A. thaliana	57.0	53.2	48.2	47.3	55.4	51.3	49.2	46.6	42.6	46.0	42.0	100				
Populus	48.1	51.1	48.8	44.4	46.2	47.4	41.4	42.0	38.4	39.0	42.5	36.0	100			
Papaya	40.5	37.8	43.5	42.0	42.4	43.9	44.2	42.0	40.9	40.0	39.8	35.8	74.0	100		
Vitis	47.4	45.7	46.0	47.1	47.2	46.8	42.2	43.2	48.8	48.8	44.1	41.8	64.6	62.9	100	
Rice	37.3	34.3	36.0	36.2	35.9	38.4	31.2	34.1	37.6	39.2	34.5	32.5	49.2	58.4	53.4	100

(A_1*A_2). High degree of similarity was observed in C_2*C_1 (92.4%), D_9*D_7 (83.2%), and D_6*D_5 (86.9%). On average, C genome has shown similarity with A_1 genome 76.65% and with A_2 genome 74.85%. Average similarity of D genome with A_1 and A_2 was 69.84 and 69.7%, respectively. G genome has shown 84.5% similarity with A_2 genome and 82.6% similarity with A_1 genome. E genome has shown 73.3% similarity with A_1 and 76.6% similarity with A_2 genome.

Among the sequenced genomes Arabidopsis has shown average similarity with cotton genomes 50% while Populus, Papaya, Vitis and Oryza has shown 44.5, 41.6, 46.1, and 35.9%, respectively (Table 5).

***Phylogenetic assessment*:** Phylogenetic assessment revealed two large clusters A and B. Cluster A represent diploid genomes of cotton and further comprised of three subclusters (a_1, a_2, and a_3). Sub cluster a_1 included D genomes and E_1 genome also joined this cluster separately. Subcluster a_2 included A_1, A_2 and G_3 genomes. Subcluster a_3 included C_1 and C_2. *Arabidopsis thaliana* distantly combine to this main cluster. Second cluster B include Populus,

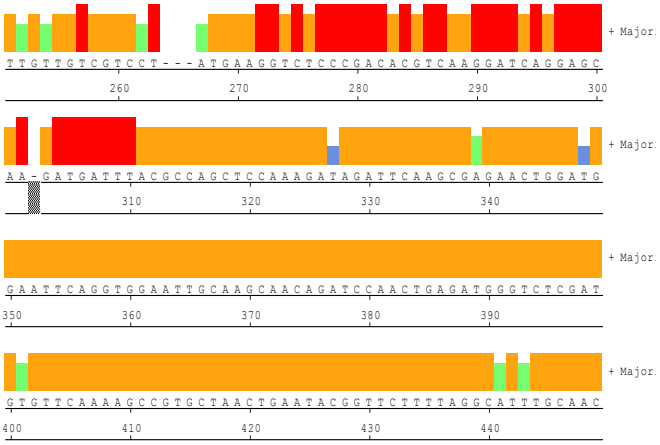

Figure 1. Alignment reports of cotton genomes and other sequenced genomes.

Similarity among sequences decrease from orange→red→green→light blue→blue. Only those SNPs were considered which were consistently present in all genomes.

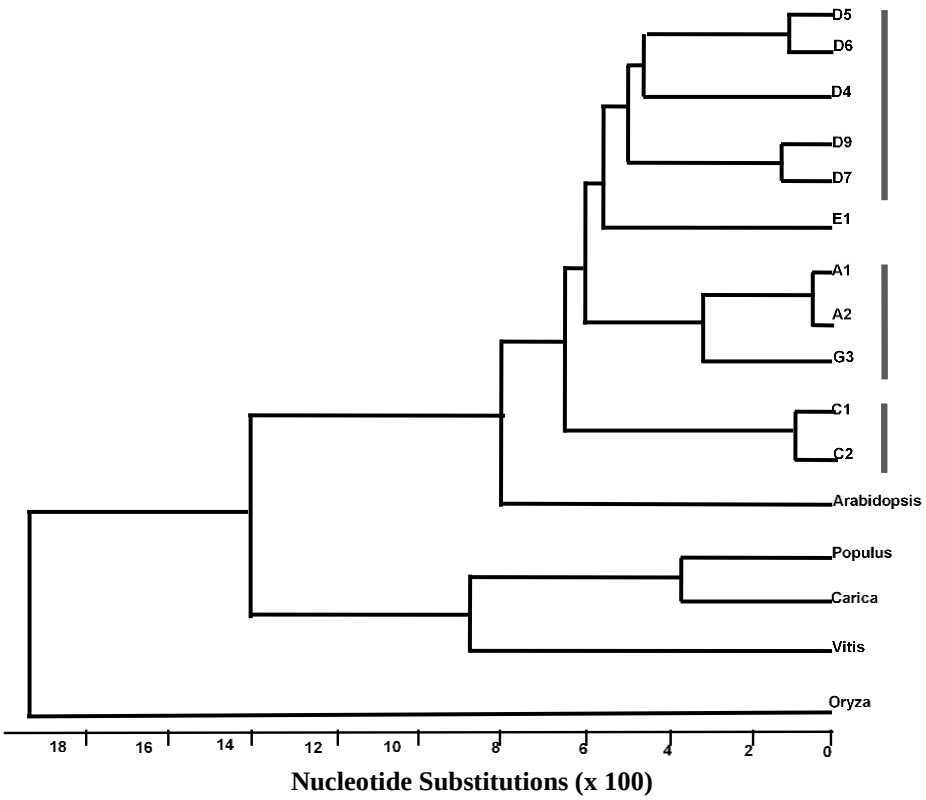

Figure 2. Phylogenetic tree of 11 diploid cotton species and five sequenced genome (Arabidopsis, Populus, Papaya, Vitis and Oryza) species on bases of SNPs studied.

Carica and Vitis while Populus and papaya are closely associated. Oryza is most distantly related with these clusters (Fig. 2).

DISCUSSION

We have harnessed SNPs for phylogenetic assessment of cotton genomes. This is an emerging technology which can facilitate the research area of evolutionary dynamics. SNPs being frequent in number can prove very authentic markers in phylogenetic studies (Batley and Edwards, 2007).

During this study of eleven diploid genome species, we observed 61 SNPs in conserved regions of seven ESTs showing homology with different genes (Table 2). Previously in a study SNPs were detected in tetraploid cotton by comparing the sequences of fiber related genes (Lu *et al.,* 2005). In another study the differential evolutionary dynamics of low copy number *Adh* genes in tetraploid cotton by comparison of sequences of genes were explained (Small and Wendel, 2000).

When sequences of the genes obtained from cotton (*G. arboreum*) in this study were compared with sequences from other plant species to see maximum similarity (Table 2). Two sequences shown maximum similarity with *A. thaliana*, two have shown with *populus*, while others with *Pisum sativum*, *Medicago truncatula*, *Solanum lycopersicum* and *Vitis vinifera*. Cotton is a member of family Malvaceae and its similarity with family Brassicaeae, Salicaceae, Fabaceae, Solanaceae and Vitaceae is indicating a wide evolutionary background of cotton.

Study of number of mutation in comparison with other organisms is vital to understand rate of evolution of genes like in a study by Rausher *et al*. (1999). Number of SNPs observed in two genes (MT-sHSP and Actin-depolymerizing factor 2) in comparison with *A. thaliana* were high (101 and 103) (Fig. 1) as compared to other genes which depict a high rate of evolution in these genes (Table 3) showing different pace of evolution for different genes. Mitochondrial genes are prone to somatic mutations which are not inherited to next generation which may be cause of high mutation rates in MT-sHSP (Fischel-Ghodsian *et al.,* 2004) while high rate of mutation in Actin-depolymerizing factor 2 gene is not known. A high degree of similarity between *Arabidopsis* and cotton (50%) among the outgroup species strengthen the concept given in earlier reports elucidating their origin from a common ancestor (Rong *et al.*, 2005). Positions of these genes were identified upon comparing with the *A. thaliana* chromosomes (Table 4). While according to position on chromosomes of Arabidopsis two genes were present near to centre and four were present at distal end (Table 4). Rate of mutations does not seem to be dependent on position of gene

on chromosome.

Nucleotide substitutions were more frequent than Indels. Indels can change the whole frame of codons translation (Liston *et al.,* 1995; Vage and Lingaas, 2008) and are usually not tolerated in coding regions. Among substitutions 73.5% were transitions and 26.5% were transversions which is congruent with the previous studies (Wang *et al.,* 1998; Schneider *et al.,* 2001). For example in ten STSs, most of SNPs were of transition type C↔T (A↔G) (52.7%) in Citrus (Novelli *et al.,* 2004). Such commonalties were found in human (66%, Zhang and Zhao, 2004), mouse (66%, Zhang and Zhao 2004), Drosophila (55%, Hoskins *et al.,* 2001), rice (61.8%, Feltus *et al.,* 2004) and Arabidopsis (52.8%, Jander *et al.,* 2002). Alterations of this type could be attributed to the actions of 5-methylcytosine deamination (Brookes, 1999; Hoskins *et al.,* 2001; Feltus *et al.,* 2004).

There are deep genetic backgrounds among diploid cotton germplasms for the genetic and breeding research. Within the diploid cotton species, there are many favorable characteristics such as fiber quality and yield traits, insect and pathogen resistance, tolerance to environmental stresses and so on, which may be utilized in development of superior cultivars (Abdalla *et al.,* 2001). So an understanding of the genetic and genomic relationships of cotton species is a critical step for further utilization of extant genetic diversity and genomic information (Wu *et al.,* 2007). Here we have utilized different genes to study relatedness of different cotton diploid species.

Tenaillon *et al*. (2001) suggested that the greater SNP rates can be correlated with the higher level of diversity. When comparing the different sequences among the cotton genomes we could verify that SNPs are present after every 31 bases in the cotton genomes which depict high diversity among diploid genomes of cotton at nucleotide level which is also evident with cytogenetic (Omran *et al.,* 2007), DNA marker studies with SSRs (Guo *et al.,* 2006; Wu *et al.,* 2007) and also with some gene sequences (Small and Wendel, 2000).

Even though the ETSs represent a very small portion of the cotton genome, the data generated in this work were congruent with the earlier phylogenetic studies using other markers in cotton such as SSRs, RFLPs and RAPDs (Khan *et al.,* 2000; Abdalla *et al.,* 2001; Guo *et al.,* 2006; Wu *et al.,* 2007). Species representing subgenomes of the same genome were grouped in one cluster consistent with the classification of genomes (Fig. 2). A-genome species were grouped together as well as C-genome species and D-genome species were closely associated. D_6 and D_5 are showing close relationship and D_7 and D_9 are also closely associated with them. G genome species grouped with A genome and E genome species was distantly related with other genomes which is according to the evolutionary relations of cotton genomes (Hawkins *et al.,* 2006).

Cotton and Arabidopsis have similarities at genomic level (Rong *et al.,* 2005) which has been elucidated in this study as well. While among the other genomes, Populus (family Salicaceae) and Papaya (family Caricaceae) are closely associated (84%) with each other and on average 44.5% and 41.6% with cotton genomes which both belongs to woody plant families and Vitis (Vitaceae) which is a climbing herb is 64.6% and 72.9% similar with these genomes, respectively, and on average 46.1% similar with cotton genomes (Table 5). Oryza is most genetically diverged from all the genomes showing an average 38.9% similarity and 35.8% similarity with cotton genomes. This low degree of similarity of Oryza can be explained on the basis of its being a monocot species which diverged from a common ancestor 150 mya (Chaw *et al.,* 2004). In a study a high degree of collinearity in gene order was observed in Arabidopsis, Papaya, Populus and Vitis which indicate the common ancestry of these genomes (Tang *et al.,* 2008).

While comparing A_1 genome and A_2 genome with the other genomes of the genus *Gossypium*, C genome has shown more affinity with A_1 genome as compared to A_2 genome. While other genomes including G and E genome has shown more affinity with the A_2 genome as compared to A_1 genome. While D-genome has shown approximately equal affiliation for both the genomes which reflect a different pattern of evolution between two A genome species. Among the genomes, D_7 genome has shown minimum similarity with A genome while G genome has shown maximum similarity with A-genome. These results indicate a high rate of evolution in A_1 genome as compared to A_2 genome with respect to G and E genome while a high rate of evolution in A_2 genome with respect to C genome species.

In conclusion SNPs are an effective tool for whole genome survey and are potent markers to survey conserved regions where other markers may not prove very effective. SNPs can be effectively utilized for phylogenetic studies as other DNA markers.

***Acknowledgements*:** Funds for the present study were provided by Higher Education Commission (HEC) Pakistan through Presidential Young Innovator (PYI) Program and funds to PhD student through indigenous PhD program.

REFERENCES

Abdalla, A.M., O.U.K, Reddy, K.M. El-Zik and A. E. Pepper. 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor. Appl. Genet. 102:222-229.

Aerts, J., Y. Wetzels, N. Cohen and J. Aerssens. 2002. Data mining of public SNP database for the selection of intragenic SNPs. Hum. Mutat. 20:162-173.

Ahmad, S., M. Ashraf, T. Zhang, N. Islam, T. Shaheen, M. Rahman. 2007. Identifying genetic variation in *Gossypium* L. based on single nucleotide polymorphism. Pak. J. Bot. 39:1245-1250.

An, C., S. Saha, J.N. Jenkins, D.P Ma, B.E. Scheffler, R.J. Kohel, J.Z. Yu and D.M. Stelly. 2008. Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping. Theor. Appl. Genet. 116:1015-1026.

Araujo, A.H., M.E. Fonseca, L.S. and Boiteux. 2007. Nucleotide diversity of a major carotenoid biosynthetic pathway gene in wild and cultivated *Solanum* (Section *Lycopersicon*) species. Brazil. J. Plant. Physiol. 19:233-237.

Ayeh, K.O. 2008. Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): Emerging molecular marker tools for improving agronomic traits in plant biotechnology. Afr. J. Biotech. 7:331-341.

Azhaguvel, P. and T. Komatsuda. 2007. A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. Ann. Bot. 100:1009-1015.

Ball, A.D., J. Stapley, D.A. Dawson, T.R. Birkhead, T. Burke and J . Slate. 2010. A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). BMC Genomics 11:218. doi: 10.1186/1471-2164-11-218.

Bansal, V. and O. Harismendy, R. Tewhey, S.S. Murray, N.J. Schork, E.J. Topol and K.A. Frazer. 2010. Accurate detection and genotyping of SNPs utilizing population sequencing data Published in Advance. Genome Res. 20:537-545.

Batley, J. and D. Edwards. 2007. SNP applications in plants. In: N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner and H.N. De Silva. Published in Association Mapping in Plants; pp.95-102.

Brookes, A.J. 1999. The essence of SNPs. Gene 234:177–186.

Chaw, S.M., C.C. Chang, H.L. Chen and W.H. Li. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 58:424-441.

Chen, X., L. Levine and P.Y. Kwok. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res. 9:492-498.

Clark, R.M., G. Schweikert, S. Ossowski, G. Zeller, C. Toomajian, P. Shinn, N. Warthmann, T.T. Hu, G. Fu, D.A. Hinds, H. Chen, K.A. Frazer, D.H. Huson, B. Scholkopf, M. Nordborg, G. Ratsch, J.R Ecker and D. Weigel. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Sci. 317:338-342.

Duran, C., N. Appleby, M. Vardy, M. Imelfort, D. Edwards and J. Batley. 2009. Single nucleotide polymorphism discovery in barley using auto SNPdb. Plant Biotechnol. J. **7**:326-333.

Feltus, F.A., J. Wan, S.R. Schulze, J.C. Estill, N. Jiang and A.H. Paterson. 2004. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res. 14:1812-1819.

Fischel-Ghodsian, N., R.D. Kopke and X. Ge. 2004. Mitochondrial dysfunction in hearing loss. Mitochondrion 4:675-694.

Gamazon, E.R., W. Zhang, A. Konkashbaev, S. Duan, E.O. Kistner, D.L. Nicolae, M.E. Dolanand and N.J. Cox. 2010. SCAN: SNP and copy number annotation. Bioinformatics 26:259-262.

Guo, W., W. Wang, B. Zhou and T. Zhang. 2006. Cross species transferability of *G. arboreum*- derived EST-SSRs in the diploid species of *Gossypium*. Theor. Appl. Genet. 112:1573-1581.

Hawkins, J.S., H. Kim, J. Nason, R.A. Wing and J.F. Wendel. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium.* Genome Res. 16:1252–1261.

Hoskins, R.A., A.C. Phan, M. Naeemuddin, F.A. Mapa, D.A. Ruddy, J.J. Ryan, L.M. Young, T. Wells, C. Kopczynski and M.C. Ellis. 2001. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. Genome Res. 11:1100–1113.

Huang, C., C. Chen, S.D. Mague, J.A. Blendy and L. Liu-Chen. 2012. A common single nucleotide polymorphism A118G of the μ opioid receptor alters its N-glycosylation and protein stability. Biochem. J. 441:379–386.

Huang, W., H. Zhang, Y. Hao, X. Xu, Y. Zhai, S. Wang, Y. Li, F. Ma, Y. Li, Z. Wang, Y. Zhang, X. Zhang, R. Liang, Z. Wei, Y. Cui, Y. Li, X. Yu, H. Ji, F. He, W. Xie and G.A. Zhou. 2016. Non-synonymous single nucleotide polymorphism in the HJURP gene associated with susceptibility to hepatocellular carcinoma among Chinese. PLoS One. 11(2): e0148618. doi: 10.1371/journal.pone.0148618.

Iqbal, M.J., N. Aziz, N.A. Saeed, Y. Zafar and K.A. Mailk. 1997. Genetic diversity of some elite cotton varieties by RAPD analysis. Theor. Appl. Genet. 94:139-144.

Jander, G., S.R. Norris, S.D. Rounsley, D.F. Bus, I.M. Levin and R.L. Last. 2002. Arabidopsis map-based cloning in the postgenome era. Plant Physiol. 129:440–450.

Kanazin, V., H. Talbert, D. See, P. DeCamp, E. Nevo and D. Blake. 2002. Discovery and assay of single nucleotide polymorphism in barley (*Hordeum vulgare*). Plant Mol. Biol. 48:529-537.

Khan, S.A., D. Hussain,, E. Askari, J. McD Stewart, K.A. Malik and Y. Zafar. 2000. Molecular phyllogeny of *Gossypium* species by DNA fingerprinting. Theor. Appl. Genet. 101:931-938.

Koornneef, M., C. Alonso-blanco and D. Vreugdenhil. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. Ann. Rev. Plant. Biol. 55:141–172.

Lacape, J.M., T.B. Nguyen, B. Courtois, J.L. Belot, M. Giband, J.P. Gourlot, G. Gawryziak, S. Roquesand and B. Hau. 2005. QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* _ *Gossypium barbadense* backcross generations. Crop. Sci. 45:123–140.

Lindblad-Toh, K., E. Winchester, M.J. Daly, D.G Wang, J.N. Hirschhorn, J.P. Laviolette, K. Ardlie , D.E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J.B. Fan, T. Gingeras, J. Warrington, N. Patil, T.J. Hudson and E.S Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat. Genet. 24:381-386.

Liston, P. and D.J. Briedis. 1995. Ribosomal frameshifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. J. Virol. 69:6742-6750.

Liu, A. and J.M. Burke. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321-330.

Lu, Y., J. Curtiss, J. Zhang, R.G. Percy and R.G. Cantrell. 2005. Discovery of single nucleotide polymorphisms in selected fiber genes in cultivated tetraploid cotton. National Cotton Council Beltwide Cotton Conference 946.

Novelli, V.M., M.A. Takita and M.A. Machado. 2004. Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus. Euphytica 138:227-237.

Omran, A., A. Asadollah and N. Saeid. 2007. Intragenomic diversity and geographical adaptability of diploid cotton species revealed by cytogenetic studies. Afr. J. Biotech. 6:1387-1392.

Rausher, M.D., R.E. Miller and P. Tiffin. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. 16:266-274.

Rong, J., J.E. Bowers, S.R. Schulze, V.N. Waghmare, C.J. Rogers, G.J. Pierce, H. Zhang, J.C. Estill and A.H. Paterson. 2005. Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. Genome Res. 15:1198-1210.

Rong, J.K., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X. Ding, J.J. Garza, B.S. Marler, C. Park, G.J. Pierce, K.M. Rainey, V. Rastogi, K. Schulze, N.L. Tronlinde, J.F. Wendel, T.A. Wilkins, R.A. Wing, R.J. Wright, X. Zhao, L. Zhu and A.H. Paterson. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389-417.

Schneider, K., B. Weisshaar, D.C. Borchardt and F. Salamini. 2001. SNPs frequency and allelic haplotypes structure of *Beta vulgaris* expressed genes. Mol. Breeding 8:63-74.

Semagn, K., A. Bjornstad and M.N. Ndjiondjop. 2006. An overview of molecular marker methods for plants. Afr. J. Biotech. 5:2540-2568.

Small, R.L. and J.F. Wendel. 2000. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in new world diploid cotton (*Gossypium* L., Malvaceae). Mol. Phylogenet. Evol. 16:73–84.

Tang, H., J.E. Bowers, X. Wang, R. Ming, M. Alam and A.H. Paterson. 2008. Synteny and colinearity in plant genomes. Sci. 320:486-488.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). Proc. Natl. Acad. Sci. 98:9161–9166.

Udall, J.A., J.M. Swanson, K. Haller, R.A. Rapp, M.E. Sparks, J. Hatfield , Y. Yu, Y. Wu, C. Dowd, A.B. Arpat, B.A. Sickler, T.A. Wilkins, J.Y. Guo, X.Y. Chen, J. Scheffler, E. Taliercio, R. Turley, H, McFadden, P. Payton, N. Klueva, R. Allen, D. Zhang, C. Haigler, C. Wilkerson, J. Suo , S.R. Schulze, M.L. Pierce, M. Essenberg, H. Kim, D.J. Llewellyn, E.S. Dennis, D. Kudrna, R. Wing, A.H. Paterson, C. Soderlund and J.F. Wendel. 2006. A global assembly of cotton ESTs. Genome Res. 16:441-450.

Vage, J. and F. Lingaas. 2008. Single nucleotide polymorphisms (SNPs) in coding regions of canine dopamine and serotonin-related genes. BMC Genet. 9:10. DOI: 10.1186/1471-2156-9-10.

Wang, D.G., J.B. Fan, , C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittman, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee and E.S. Lander. 1998. Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Sci. 280:1077-1082.

Wu, Y.X., M.K. Daud, L. Chen and S.J. Zhu. 2007. Phylogenetic diversity and relationship among *Gossypium* germplasm using SSRs markers. Plant. Syst. Evol. 268:199-208.

Zhang, F. and Z. Zhao. 2004. The influence of neighboring-nucleotide composition on single nucleotide polymorphisms (SNPs) in the mouse genome and its comparison with human SNPs. Genome 84:785-795.

Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tasell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young and P.B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134.