SHORT COMMUNICATION

Pros and cons of using genomic SSRs and EST-SSRs for resolving phylogeny of the genus *Gossypium*

Nabila Tabbasam · Yusuf Zafar ·
Mehboob-ur-Rahman

Abstract The genus *Gossypium* is comprised of 50 diverse cotton species representing eight different genomes (A through G and K), however, phylogenetic relationship using various DNA marker types such as RAPD and SSRs was determined on limited number of cotton species. In this report, we have demonstrated the application of genomic SSRs (gSSRs) and EST-SSRs, and after combining both the data sets, for resolving the phylogenies of 36 cotton species including seven races. Out of the 100 primer pairs surveyed (50 for gSSRs and 50 for EST-SSRs), 75 produced scorable amplification products in all species. Out of these, 73 were found to be polymorphic and amplified 135 alleles ranging from 1 to 5 alleles per SSR marker (average 2.87 alleles per marker). The gSSRs amplified higher number of alleles (72) compared to the EST-SSRs (63). In total 22 highly informative SSRs with PIC values ≥0.5 were identified. Genomic SSRs containing di-while EST-SSRs containing tri-nucleotide repeats exhibited high polymorphism compared to the other nucleotide repeats containing gSSRs/EST-SSRs. Number of tandem repeats and polymorphism were positively correlated. Neither the type of chromosome nor the location of the SSRs showed association with the polymorphism. *Gossypium herbaceum* var. *africanum* (Watt) Hutch. ex and Ghose and *Gossypium robinsonii* F. Muell. were found the most genetically diverse, while among races of *Gossypium hirsutum* L. "yucatanense" and *G. hirsutum* "punctatum" were found genetically diverse. Of the three data sets, clustering analysis based on EST-SSRs and combined data sets, revealed parallel results reported in earlier studies. This study further confirmed that *Gossypium darwinii* Watt has close relationship with *Gossypium barbadense* L. Moreover, *Gossypium raimondii* Ulbr. and *G. herbaceum/Gossypium arboreum* L. are close living relatives of the ancestor allotetraploid species. Our studies suggest that for resolving phylogenetic relationship among the various plant species EST-SSRs could be a better choice. This information can be instrumental in transferring novel alleles or loci from the wild species into the cultivated cotton species which would set a stage for cultivating genetically diverse cultivars—a way to achieve sustainable cotton production in changing climate.

Keywords *Gossypium* species · Genetic diversity · EST-SSRs · Genomic SSRs · Genome specificity and transferability · Phylogenetic relationship

Introduction

In total 50 species of the genus *Gossypium* (Fryxell et al. 1992) are representing eight different genomes A through G and K (Endrizzi et al. 1984; Stewart 1995), of these 45 are diploids ($2n = 26$) while five are allotetraploids ($2n = 52$). Out of these, four species, *G. arboreum*, *G. herbaceum*, *G. hirsutum* and *G. barbadense* are cultivated in different parts of the world. Allotetraploids evolved about 1.1–1.9 million years ago by hybridization of A-genome with D-genome [only sequenced genome (Paterson et al. 2012; Wang et al. 2012)], followed by duplication (Beasley 1940; Wendel et al. 1992). Among tetraploids, *G. hirsutum* is a widely distributed species comprising seven races (6 domesticated, "marie-galante", "punctatum", "richmondi", "morrillii", "palmeri", and

N. Tabbasam · Y. Zafar · Mehboob-ur-Rahman (✉)
Plant Genomics and Molecular Breeding Lab, National Institute for Biotechnology and Genetic Engineering (NIBGE), Jhang Road, PO Box 577, Faisalabad, Pakistan
e-mail: mehboob_pbd@yahoo.com

"latifolium") and one wild ("yucatanense") (Hutchinson 1951). First cultivated races are "punctatum" and "latifolium" (Lacape et al. 2007). Most of the cultivated *G. hirsutum* varieties are derived from "latifolium" (Hutchinson 1951). *G. hirsutum* "marie-galante" is thought to be hybrid of *G. hirsutum* and *G. barbadense* (Stephens 1967; Percival and Kohel 1990).

After the revolution in textile industry, *G. hirsutum* and *G. barbadense* replaced major cultivated area under diploid cultivated species especially in Asia (Iqbal et al. 2001; Rahman et al. 2008). *G. hirsutum* is predominantly cultivated on ∼90 % of the total cotton area (Zhang et al. 2005) whereas; limited area is under cultivation of *G. barbadense*, known for producing high quality lint fiber (Abdalla et al. 2001; Rahman et al. 2012). Like other cultivated crop species, all the cultivated tetraploids and diploids have a narrow genetic base within their respective gene pool (Liu et al. 2005; Rahman et al. 2008, 2011; Shaheen et al. 2010) because of developing cultivars from few genotypes (Rahman et al. 2002b, 2005a). This phenomenon has been demonstrated in multiple investigations (Liu et al. 2000; Gutierrez et al. 2002; Lacape et al. 2007; Rahman et al. 2002b). Second, limited use of wild cotton species for breeding new cultivars is another major reason for the narrow genetic base, a major cause of stagnation in cotton productivity worldwide (Rahman et al. 2002a; Zhang et al. 2005).

The genus *Gossypium* has sufficient genetic repository for many important traits like fiber strength, high fiber yield, high tolerance/immunity to viral disease. For example, A-genome species are immune to cotton leaf curl disease—causes substantial yield losses in Pakistan (Rahman et al. 2005b, 2008, 2011, 2012). It is, therefore, vital to assess the genetic diversity and phylogenetic relationship among the available cotton species as a mandatory requirement before utilizing the species in molecular breeding programs. Numerous methods including conventional and DNA-based conventional marker systems have been employed for estimating the extent of genetic diversity among the cotton species (Menzel 1954; Brubaker and Wendel 1994, 2000; Abdalla et al. 2001; Gutierrez et al. 2002; Guo et al. 2006; Lacape et al. 2007; Rahman et al. 2009; Kalivas et al. 2011). Earlier, RAPD markers were applied for working out the phylogeny of 31 cotton species (Khan et al. 2000). RAPDs are handicapped because of reproducibility concerns. After couple of years, SSRs were applied on 22 diploid cotton species, thus are not representing most of the important cotton species using SSRs (Guo et al. 2006). EST-SSRs (SSRs derived from coding sequences are called EST-SSRs, Saha et al. 2003) have more potential to identify changes in the genes accumulated during domestication (Wang et al. 2007). In contrary to EST-SSRs, genomic SSRs (gSSRs) are highly polymorphic, tend to be widely dispersed throughout the genome but have less transferability across other species (Peakall et al. 1998; Kuleung et al. 2004). In contrast, EST-SSRs are less informative (Decroocq et al. 2003) than the gSSRs because these are part of transcribed regions thus are more conserved (Cho et al. 2000; Thiel et al. 2003).

In the present studies, two types of markers: gSSRs and EST-SSRs were used (1) to find genetic diversity and phylogenetic relationship among both the diploid (wild, cultivated) and tetraploid (wild, cultivated) species/races (2) to compare the usefulness of these markers for calculating the genetic divergence and their phylogenetic relationship. The information generated will not only be helpful in validating the existing phylogenetic relationship about the genus *Gossypium* but can also be utilized for introgressing the novel traits from the wild accession to the cultivated species, an effective strategy to counter the negative impact of climate change.

Materials and methods

Plant material

Thirty-six *Gossypium* species; 24 diploid species representing seven genomes (3 A-, 10 D-, 3 E-, 1 C-, 3 G-, 3 B-, and 1 F-genome), 5 tetraploid species (AD-genome) and 7 races of *G. hirsutum* ("morilii", "palmeri", "marie-galante", "yucatanense", "punctatum", "latifolium" and "lanceolatum") have been explored in this study (Fig. 1). Leaf samples were taken from Central Cotton Research Institute (CCRI) Multan, Pakistan and the National Institute for Biotechnology and Genetic Engineering (NIBGE) Faisalabad, Pakistan. The genomic DNA of *G. hirsutum* "yucatanense", *G. capitis-viridis*, *G. longicalyx* Hutchinson & Lee and *G. australe* F. Muell. was provided by Prof J. McD. Stewart (University of Arkansas, Fayetteville, Arkansas, USA).

Extraction of DNA and SSR analysis

For genomic DNA isolation, young leaves from five plants of each was collected from the field and stored in liquid nitrogen. The samples were ground in liquid nitrogen by following a CATB method with little modification (Doyle and Doyle 1987). DNA concentration was measured using 200 Fluorometer DyNA Quant (Hoefer USA). Quality was checked by running 30 ng of genomic DNA of each species in 0.8 % agarose gel. We too validated the concentration of the genomic DNA.

Fig. 1 Location of countries on World map from which cotton species have been collected for current study **1**. *G. herbaceum* (A₁), Southern Africa & Arabian Peninsula **2**. *G.herbaceum* var. *africanum* (A₁), Africa **3**. *G. arboreum* (A₂), India, Pakistan **4**. *G. thurberi* (D₁), Mexico, Arizona **5**. *G. klotzschianum* (D₃₋ₖ), Galapagos Islands **6**. *G. herkensii* (D₂₋₂), Mexico **7**. *G. davidsonii* (D₃₋d), Mexico **8**. *G. aridum* (D₄), Mexico **9**. *G. raimondii* (D₅), Peru **10**. *G. gossypioides* (D₆), Mexico **11**. *G. lobatum* (D₇), Mexico **12**. *G. trilobum* (D₈), Mexico **13**. *G. laxum* (D₉), Mexico **14**. *G. hirsutum cv* NIBGE 115, (AD₁) Central & northern America, Caribbean, Pacific **15**. *G. tomentosum* (AD₃), Hawaii **16**. *G. mustilinum* (AD₄), Brazile **17**. *G. barbadense* (AD₂), Hawaii **18**. *G. darwinii* (AD₅), Galapagos **19**. *G. hirsutum* "marie-galante" (AD), Caribbean,Central America **20**. *G. hirsutum* "latifolium" (AD), Southern Mexico/Guatemala **21**. *G. hirsutum* "morrillii" (AD), Mexico **22**. *G. hirsutum* "palmeri" (AD), Mexico **23**. *G. hirsutum* "punctatum" (AD), Cameron **24**. *G. hirsutum* "yucatenense" (AD), Guadeloupe **25**. *G. lanceolatum* (AD), Mexican states of Oaxaca& Guerrero **26**. *G. anomalum* (B₁), Africa **27**. *G. barbosanum* (B₂), Africa **28**. *G. capitis-viridis* (B₃), Africa Cape Verde Island **29**. *G. somalense* (E₂), Africa/Arabia **30**. *G. stocksii* (E₁), Arabian penunsula **31**. *G. longicalyx* (F), *East Africa* **32**. *G. incanum* (E₄), Arabian Penunsula **33**. *G. robinsonii* (C₂), Australia **34**. *G. australe* (G₂), Australia **35**. *G. nelsonii* (G₃), Australia **36**. *G. bickii* (G₁), Australia

A total of 100 SSR primer pairs, 50 each of BNL (representing gSSRs) and MGHES (representing EST-SSRs derived from fiber tissues of *G. hirsutum*) series were used for the analysis. Sequences of these primers were downloaded from http://www.mainlab.clemson.edu/cmd/primer, and got synthesized from GeneLink, USA. Polymerase chain reaction (PCR) was carried out in gradient thermal cycler, (Eppendorf Germany). Total reaction mixture of 20 µl contained 1 × Fermentas *Taq* buffer, 2.5 mM MgCl₂, 0.25 mM dNTPs each, 0.15 µM primers, one unit *Taq* DNA Polymerase (Fermentas, USA) and 50 ng genomic DNA as a template. The programming of thermal cycler was adjusted for 5 min at 94 °C for 1 cycle and 35 cycles each of 94 °C for 30 s for denaturation, primer annealing (50–61 °C) for 30 s, 72 °C for 30 s for extension. The final extension temperature at 72 °C was granted for 10 min. The PCR products were fractionated in 4 % metaphore-agarose gels (Cambrex Corporation, USA). The gel was stained with ethidium bromide which was then visualized in UV (ultraviolet light) and scored manually.

Statistical analysis

The gels were scored by assigning '1' for the presence of amplified allele while we scored '0' for the absence of a fragment. The reactions were repeated twice to confirm the absence of fragment. Few loci could not be amplified in genotypes were scored as 'null' alleles. Bright fragments were considered for scoring. We did not consider 25 primers (16 gSSRs and 9 EST-SSRs) in the final analysis because of the poor amplification profile. Polymorphism information content (PIC) was calculated for determining

the diversity of each SSR locus (Anderson et al. 1993). Following formula was used for PIC values calculation.

$$PIC = 1 - \sum Pij^2$$

Pij is the frequency of the jth allele for the ith locus summed across all alleles for the locus.

With the help of scoring profile similarity matrix was calculate (Nei and Li 1979). These similarity coefficients were used to construct the phylogenetic tree using unweighted pair group method of arithmetic means (UP-GMA). We used the PAUP version 4.4 software. We developed three dendrogram using gSSRs, EST-SSRs and combined data sets. Amplification percentage was calculated by the formula described by Kuleung et al. (2004).

% age of amplification

= (no. of amplified markers × 100/total no. of markers)

Similarly, transferability of each SSR marker was calculated as the percentage of amplified products in all the cotton species. We too manually estimated the correlation between the repeat type and the polymorphism rate.

The frequency distribution of the alleles among all the cotton species was calculated manually on MS excel. In this regard, we divided it in 10 different categories, i.e., alleles with frequency of 0.1 or less, 0.19 or less, 0.29 or less, 0.39 or less, 0.49 or less, 0.59 or less, 0.69 or less, 0.79 or less, 0.89 or less 0.99 or less.

The association between the polymorphism rates with tandem repeats number was also calculated manually on the basis of the PIC values given in Table 1.

Results

Microsatellite polymorphism

A total of 36 *Gossypium* species/races were investigated using 75 SSR markers, amplified 87 loci. Out of these, 10 (MGHES-19, MGHES-33, MGHES-38A, MGHES-41, MGHES-60, BNL-448, BNL-1350, BNL-1878, BNL-3408 and BNL-3646) amplified two loci while one primer BNL-3955 amplified three loci. A total of 73 SSR primer pairs (97.70 %) were found polymorphic, while only two primers i.e., MGHES-13 and MGHES-17 were monomorphic. Out of the 73 polymorphic primers, 26 (34.66 %) were polymorphic because of amplifying null alleles in few species while 47 (62.66 %) were polymorphic because of amplifying alleles of different size. The range of amplified fragments was 70 bp for MGHES-12 to 700 bp for MGHES-07. A total of 135 alleles were amplified. Range of alleles detected on single locus was 1–5 with average

alleles (2.87) per locus. Maximum number of alleles (5) were amplified by each of the genomic SSRs (gSSRs), BNL-1878, BNL-2691, BNL-3955 and BNL-3985. Relatively, higher number of alleles were amplified in tetraploids (two alleles per locus) than the diploids (1.36 alleles per locus).

The frequency distribution of the 135 alleles is shown in Fig. 2. The frequencies of alleles ranged from 0.025 to 1 having average frequency 0.469 (Table 1). A total of 18 alleles (16.51 %) appeared with 0.10 or lower frequencies, whereas 29 (26.6 %) appeared with a frequency of 0.99 or higher frequencies. None allelic variations were observed in MGHES-13 and MGHES-17 loci and were amplified in all cotton species.

In this study, the average value of polymorphism information content (PIC) was 0.50, with the highest value 0.882 for MGHES-27 and lowest value 0.11 for BNL-3895 (Table 1). Genomic SSRs showed higher PIC value (average value 0.35) compared to the EST-SSRs (average 0.291). Diploids exhibited high (0.30) while tetraploids depicted low (0.21) average PIC values.

Genetic characterization

In the present study, 22 informative SSRs, 11 BNLs and 11 MGHES (Table 2) with PIC ≥ 0.5 were identified which can distinguish all the *Gossypium* species. All useful gSSRs (BNLs) contained di-nucleotide motif whereas 63.63 % of EST-SSRs (MGHES) contained tri and 27.27 % had di-, tetra- and penta-nucleotide motif. However, location and or position of SSRs on chromosome (either proximal to centromere or near distal end) has no effect on polymorphism information content (Table 2), while polymorphism rate was found positively correlated with tandem repeats number (Table 1).

Transferability of SSRs across *Gossypium* genomes and genome specificity.

Out of the 75 SSRs, 22 (29.33 %) produced amplicons in all the 36 species/races. We did not find any association between the repeat motif type with rate of transferability. Out of the total amplified fragments (75 SSRs × 36 species = 2,700), 44.16 % were found in more than one genome group. Four primers BNL-1350, BNL-3147, BNL-3065 and BNL-3558 produced the least number of species whereas MGHES-15, MGHES-17, MGHES-21, MGHES-26, MGHES-28, MGHES-30, MGHES-52, BNL-448, BNL-3672, BNL-3793 and BNL-3985 amplified fragments in most species. Among diploids, the species belonging to A-, B-, F-, and E-genomes showed high transferability rate while D-genome species exhibited low transferability rate (Table 3). We identified 15 genome- or species-specific primers (Table 4). None of the species belonging to A-, B-, C- and G-genome were amplified with BNL-3147,

Table 1 Markers with repeat motif type, number of alleles, frequencies and PIC value

Marker (loci)	Allele size range (bp)/loci	Repeat motif	No. of alleles	Allele freq	PIC value
MGHES 06	130–175	$(CCA)_7$	3	0.115, 0.769, 0.115	0.381657
MGHES 8	175–200	$(ATT)_{11}$	2	0.433, 0.566	0.491111
MGHES 11B	200–225	$(TTA)_3(CGG)_3$	2	0.769, 0.230	0.35503
MGHES 12	70–100	$(AC)_{10}$	2	0.304, 0.695	0.42344
MGHES 15	175–200	$(AAAC)_5$	2	0.523, 0.476	0.498866
MGHES 17	175–190	$(CAT)_6(TTC)_4$	2	0.341, 0.658	0.449732
MGHES 18	100–200	$(AT)_{13}$	2	0.333, 0.667	0.444444
MGHES 19(2)	110–250	$(AGA)_5(AGC)_3$	2	0.534, 0.465	0.497566
MGHES 20	200–225	$(CCA)_9$	3	0.333, 0.6, 0.066	0.524444
MGHES 21	200–230	$(GA)_{14}$	2	0.227, 0.772	0.35124
MGHES 22	175–250	$(AGA)_7(GAA)_3$	2	0.842, 0.157	0.265928
MGHES 27	225–280	$(TCT)_7(AAC)_4(TTC)_3$	3	0.375, 0.281, 0.343	0.881836
MGHES 30	175–210	$(CT)_{13}$	4	0.527, 0.416, 0.027, 0.027	0.546296
MGHES 33(2)	190–280	$(CT)_8(TC)_8$	2	0.25, 0.75	0.375
MGHES 34	200–250	$(CCA)_3(CAC)_4$	2	0.472, 0.528	0.498457
MGHES 36	160–210	$(CTT)_8(TCA)_4$	2	0.25, 0.75	0.375
MGHES 38A(2)	170–225	$(ACC)_5(TCT)_3$	2	0.448, 0.551	0.494649
MGHES 39	175–200	$(GCC)_6(TCT)_3$	2	0.702, 0.297	0.417823
MGHES 40	150–200	$(TTC)_6$	2	0.233, 0.767	0.357778
MGHES 41(2)	210–290	$(CAA)_8$	2	0.791, 0.208	0.329861
MGHES 42	170–240	$(AGA)_6$	2	0.5, 0.5	0.5
MGHES 44	200–290	$(GAA)_{10}(TCA)_3(CAT)_3$	3	0.178, 0.678, 0.142	0.487245
MGHES 51	190–210	$(ACAA)_5(TA)_4(AT)_5$	3	0.281, 0.468, 0.25	0.648672
MGHES 52	250–340	$(CAC)_5(AGG)_3(CCA)_3$	3	0.166, 0.694, 0.138	0.470679
MGHES 60(2)	125–200	$(AT)_{15}(TA)_5$	2	0.355, 0.644	0.458272
MGHES 63	160–200	$(TTTTA)_6$	3	0.057, 0.685, 0.257	0.460408
MGHES 76	200–230	$(GA)_{20}$	2	0.4, 0.6	0.48
BNL 448(2)	90–225	$(CT)_{13}$	4	0.125, 0.553, 0.196, 0.125	0.623724
BNL 827	250–275	$(CA)_{19}$	3	0.233, 0.6, 0.166	0.557778
BNL 1317	175–200	$(AG)_{14}$	3	0.086, 0.826, 0.086	0.302457
BNL 1350(2)	200–300	$(CA)_8(GA)_{16}$	3	0.222, 0.278, 0.379	0.623457
BNL 1878(2)	125–250	$(AG)_{14}$	5	0.025, 0.225, 0.5, 0.15, 0.1	0.66625
BNL 2691	230–260	$(GA)_{23}$	5	0.428, 0.5, 0.071, 0.75, 0.667	0.561224
BNL 3065	175–200	$(AG)_{21}$	3	0.108, 0.521, 0.369	0.579395
BNL 3103	150–200	$(GA)_{13}(TC)_{14}$	3	0.367, 0.408, 0.224	0.648063
BNL 3147	180–220	$(AG)_{11}$	3	0.189, 0.567, 0.243	0.582907
BNL 3255	225–320	$(GC)_6AT(AC)_{14}$	4	0.310, 0.206, 0.103, 0.379	0.706302
BNL 3408(2)	200–250	$(GT)_2AT(GT)_{12}$	4	0.0606, 0.469, 0.439, 0.030	0.581726
BNL 3558	200–250	$(AC)_{11}$	3	0.121, 0.609, 0.268	0.541344
BNL 3563	240–275	$(CA)_{13}(TA)_4$	3	0.041, 0.583, 0.375	0.517361
BNL 3627	160–200	$(TC)_{17}$	4	0.161, 0.483, 0.354, 0.233	0.613944
BNL 3646(2)	110–170	$(TC)_{14}$	4	0.054, 0.783, 0.054, 0.108	0.368152
BNL 3793	150–300	$(TG)_{15}$	3	0.225, 0.375, 0.4	0.63875
BNL 3888	200–250	$(TG)_{15}$	3	0.115, 0.576, 0.307	0.559172
BNL 3895	180–200	$(TG)_{10}$	2	0.058, 0.942	0.110727
BNL 3955(3)	170–350	$(CA)_{12}(GT)_{13}$	5	0.043, 0.108, 0.195, 0.152, 0.5	0.674858
BNL 3985	150–300	$(TC)_{23}$	5	0.086, 0.1521, 0.195, 0.5, 0.065	0.676749

The number in parenthesis in column one indicates the number of loci amplified

similarly none of the species of A- and AD-genome were amplified with MGHES-44 thus BNL-3147 can be used as A-, B-, C- and G-genome negative and MGHES-44 can be used as A- and D-genome negative. BNL-3888 did not amplify any of the D- and E-genome species which can be used as D- and E-genome negative. MGHES-16, BNL-
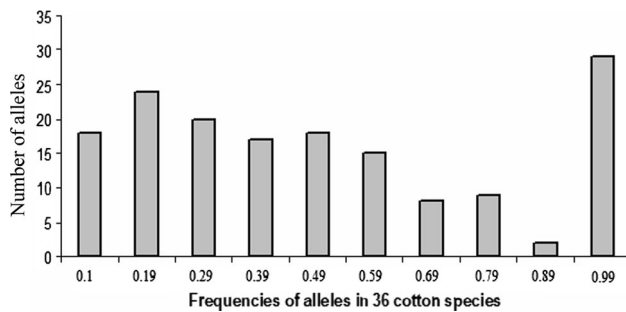
Fig. 2 Frequency distribution of 135 alleles in 36 cotton species/landraces

Table 2 Most informative SSRs with their position on chromosome

Marker	Chromosome	Position CM	PIC value
MGHES 06	D	62.7	0.582907
MGHES 19	A	31.6	0.581726
MGHES 20	D	31.5	0.524444
MGHES 21	D	144	0.63875
MGHES 27	A	0	0.881836
MGHES 30	A	151	0.546296
MGHES 36	D	65.6	0.648063
MGHES 40	D	159	0.541344
MGHES 51	D	135	0.648672
MGHES 52	A	168	0.579395
MGHES 63	A	81.4	0.557778
BNL 448	D	81.3	0.623724
BNL 1878	D	64.2	0.66625
BNL 2691	D	71	0.561224
BNL 3255	A	119	0.706302
BNL 3408	D	68.6	0.623457
BNL 3563	A	71.6	0.517361
BNL 3627	A	66.7	0.613944
BNL 3793	D	92.2	0.5
BNL 3895	A	64.5	0.110727
BNL 3955	D	167	0.674858
BNL 3985	D	15	0.676749

1053 and BNL-1359 did not amplify in *G. trilobum* (DC) Skovst. BNL-3482 did not amplify in *G. aridum* (Rose & Standl.) Skovst. Two EST-SSRs MGHES-20 and MGHES-21 could not amplify the genomic DNA of *G. darwinii* while MGHES-15 could not amplify in *G. aridum* and *G. trilobum* (Table 4). Thus, this set of primers can be utilized as species-specific primers.

Microsatellite performance among diploid (A and D) and tetraploid (AD) genome species

In the present investigation, 12 primer pairs (MGHES-12, MGHES-22, BNL-1878, BNL-2449, BNL-2634, BNL-

2691, BNL-3147, BNL-3103, BNL-3408, BNL-3793, BNL-3955 and BNL-3985) did not amplify A-genome species while, these primers produced some private alleles in D- and AD-genome species, indicating specificity of these primers for D-genome.

The sizes of many amplicons of tetraploids were different from diploids (A, D) (Fig. 3). The size of amplified fragments in all the A-, D- and AD-genome species was in the range of 101–700 bp. However, the percentage of fragments within 101–300 bp in AD-genome species was higher than that of A- and D-genome species. All the A-genome species produced relatively bigger fragments sizes between 301 and 400 bp (Fig. 3).

A- and D-genome species relationship with AD-genome species

Gossypium herbaceum showed 0.661 and 0.624 genetic similarity with *G. hirsutum* and *G. barbadense*, respectively. While *G. arboreum* was found 0.57 and 0.63, genetically similar with *G. hirsutum* and *G. barbadense*, respectively (Table 5). Among D-genome species, *G. raimondii* was more genetically close to *G. hirsutum* (0.642) and *G. barbadense* (0.667). Average coefficients of genetic similarity of diploid (A-, D-) cotton species (11 in number) with tetraploid (AD) species (five in number) were in the range of 0.56–0.64 (Table 5). *G. raimondii* showed the highest (0.652) while *G. lobatum* Gentry showed lowest (0.566) mean genetic similarity with the five tetraploid species.

Genetic similarity among diploid and tetraploid cotton species with EST and gSSRs

Gossypium arboreum (A_2) and *G. herbaceum* (A_1) showed the maximum genetic similarity (0.89) followed by *G. barbosanum* Phillips & Clement (B_2) and *G. capitis-viridis* (B_3). While the lowest genetic similarity coefficient (0.50) was observed between *G. herbaceum* var. *africanum* (A_1) and *G. robinsonii* (C_2). Among diploid species average genetic similarity coefficient was 0.67, *G. thurberi* Tod. (D_1) showed highest (0.71) while, *G. robinsonii* (C_2) showed lowest average similarity (0.62) to all other diploid species.

Genetic similarity coefficients between tetraploids species/races were in the range of 0.62–0.85 (average 0.73). Least genetic similarity (0.62) was found between *G. hirsutum* "yucatanense" and *G. mustilinum* Miers ex G. Watt while maximum genetic similarity (0.85) was found between *G. darwinii* and *G. barbadense* as well as between *G. tomentosum* Nutt. ex Seem. and *G. hirsutum* (0.85). The species *G. hirsutum* "punctatum" showed the highest genetic dissimilarity with the other tetraploids. Among the

Table 3 Transferability of G. hirsutum derivative SSRs in other Gossypium species/genomes

Genome	# of SSRs amplified each subgenome*	% age of SSRs amplified each subgenome	# of SSRs amplified partially**	# of null amplified
A	32	42.67	31	12
D	24	32.0	41	10
AD	45	60	0	30
B	40	53.34	24	11
C	24	32	44	7
E	19	25.34	43	13
G	28	37.33	25	22
F	58	77.33	0	17

Null amplified means that there was no amplification in any of the species of given genome

* Indicates that amplification in all species of a given genome

** Indicates there was no amplification in some species of the given genome

Table 4 Genome and species-specific amplification features of SSRs

Primers	Genomes							
	A	B	C	D	E	F	G	AD
MGHES-15	√	√	√	* 1	√	√	√	X
MGHES-16	√	×	×	* 3	√	√	√	√
MGHES-20	√	×	×	√	√	√	√	* 5
MGHES-21	√	√	×	√	* 4	×	√	* 5
MGHES-22	×	×	×	√	×	√	√	√
MGHES-44	×	√	√	√	√	√	√	×
BNL-1053	√	√	√	* 3	√	√	×	√
BNL-1359	×	√	×	* 3	√	√	√	×
BNL-2634	×	√	×	√	√	√	√	√
BNL-3066	√	√	√	√	√	×	√	×
BNL-3147	×	×	×	√	√	√	×	√
BNL-3279	√	√	×	√	* 4	×	√	×
BNL-3482	√	√	√	* 2	√	√	√	√
BNL-3599	√	√	×	√	×	√	√	√
BNL-3888	√	√	√	×	×	√	√	√

The check mark X indicate no amplification in any species belonging to this genome while * indicate partial amplification in the mentioned genomes

* 1 null in *G. aridum* and *G. trilobumi*

* 2 null in *G. aridum*

* 3 null in G. *trilobum*

* 4 null in *G. Incanum*

* 5 null in *G. Darwinii*

races, *G. hirsutum* "morrillii" Cook & Hubb. and *G. hirsutum* "punctatum" Schumach. showed close relatedness with *G. hirsutum* "palmeri" (Watt) Wouters and *G. hirsutum* "yucatanense", respectively.

Phylogenetic study of 36 cotton species with combined data of gSSRs and EST-SSRs

Average genetic similarity coefficient of 36 *Gossypium* species/races was 0.64, with a range of 0.49–0.89
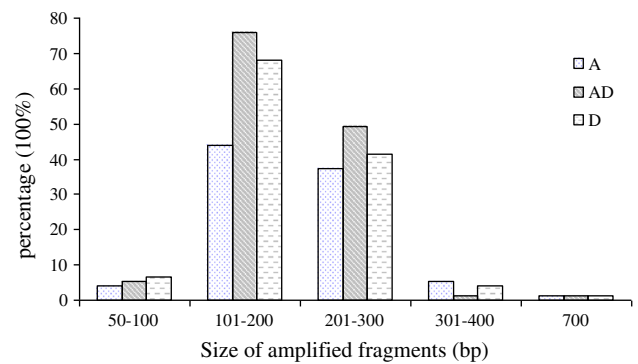
Fig. 3 Distribution of fragment sizes amplified by SSRs

(Table 6). A dendrogram (Fig. 4) was generated using these genetic similarity coefficients and it was found that species were grouped in three major clusters 'A', 'B' and 'C'. The major cluster 'A' consisted of two subclusters (a1, a2). All A-genome species were grouped in the subcluster 'a1' and all tetraploid species/races were grouped in 'a2' subcluster. Among allotetraploids, *G. barbadense* showed close relatedness with *G. darwinii*. The major cluster 'B' comprised of 10 species, two subclusters 'b1' and 'b2'. All B- and E-genome species (6 in number) were grouped in subcluster 'b1' while the subcluster 'b2' consisted of four species (Fig. 4).

A total of nine D-genome species constituted a major cluster 'C' containing two subclusters (C_1, C_2). In the subcluster 'C_1' *G. klotzschianum* Andersson D_2 with *G. davidsonii* Kellogg D_3 and *G. herkensii* Brandegee D_{2-2} with *G. aridum* D_3 formed sister clustering, respectively. *G. thurberi* D_1 was related to *G. klotzschianum* D_2 and *G. davidsonii* D_3 with genetic relatedness of 80 %. Similarly in subcluster 'C_2' *G. gossypioides* D_6 and *G. lobatum* D_7 formed a sister group relationship. The most divergent species of the dendrogram was *G. longicalyx* F_1 which was 62.15 % genetically related to all other species followed by *G. laxum* Phillips D_9 which was 64.94 % genetically related to the other species.

Table 5 Genetic similarity coefficient between tetraploid species and A-/D-genome species

	G. herbaceum A1	*G. arboreum* A2	*G. thurberi* D1	*G. klotzschianum* D2	*G. davidsonii* D3-d	*G. aridum* D4	*G. raimondii* D5	*G. gossypioides* D6	*G. lobatum* D7	*G. trilobum* D8	*G. laxum* D9
G. hirsutum 115 AD$_1$	0.661	0.570	0.582	0.630	0.618	0.618	0.667	0.545	0.552	0.594	0.600
G. barbedense AD$_2$	0.624	0.630	0.630	0.618	0.642	0.582	0.642	0.594	0.588	0.642	0.636
G. mustilinum AD$_4$	0.636	0.618	0.594	0.618	0.618	0.582	0.642	0.570	0.588	0.594	0.636
G. tomentosum AD$_3$	0.636	0.618	0.630	0.606	0.606	0.594	0.655	0.582	0.552	0.594	0.624
G. darwinii AD$_5$	0.648	0.667	0.618	0.606	0.630	0.606	0.655	0.570	0.552	0.655	0.600
Mean	0.641	0.621	0.611	0.616	0.622	0.596	0.652	0.572	0.566	0.616	0.619

Clustering of species with EST-SSRs

Cluster analysis based on EST-SSRs grouped the species into seven major clusters A through G (Fig. 5). The clustering of species with EST-SSRs (Fig. 5) is more close to the phylogenetic tree obtained from combined data set except few differences like grouping of *G. raimondii* with A-genome species and sister clustering between A- and D-genome species. Sister clustering between B- and E-genome species and also between C- and G-genome species was observed with EST-SSRs as well as with combined data set.

Clustering of species with gSSRs

Cluster analysis based on gSSRs grouped the species into six major clusters A through F (Fig. 6). Clustering of species with gSSRs deviated from both the EST-SSRs and combined data set. In gSSR-based dendrogram, E-genome species (*G. somalense* Gurke; E$_2$) was grouped with D-genome species. Similarly, *G. incanum* (Schwartz) Hillcoat (E$_4$) and *G. stocksii* Mast. (E$_1$) were grouped with *G. robinsonii* (C$_1$) and *G. longicalyx* (F$_1$), respectively. Also, the two races of *G. hirsutum* ("lanceolatum" and "latifolium") were grouped with D-genome species.

Discussion

Microsatellite polymorphism

Faint or failure of SSRs fragment amplification was expected because of the reason that SSR primer pairs were designed from the sequences derived from *G. hirsutum*. It is much likely that during evolution there was enough accumulation of mutations in annealing sites and or the loss of these loci in the diploid species which together may influence the annealing of these primers (Liu et al. 2000). In the present study, ∼97 % of the SSRs were polymorphic. Such commonalities have been reported while studying the genetic divergence among 31

Gossypium species using RAPDs (Khan et al. 2000) and 25 diploid *Gossypium* species using SSRs (Guo et al. 2006). Such a high allelic polymorphism rate among various species is the result of accumulation of mutations during evolution (Nei 2007). Average alleles per locus (2.87) were slightly higher than the previous reports (2 allele; Wu et al. 2007b). Similarly, more alleles were amplified in tetraploids than diploids, which is in agreement with earlier reports (Gutierrez et al. 2002; Kalivas et al. 2011). Multiple folds (30–36) increase in ploidy level of tetraploids (Paterson et al. 2012) is one of the possible explanations for amplifying more alleles. The propensity of the number of alleles is positively correlated with the repeat number (Lacape et al. 2007), ploidy level of the germplasm (Udall and Wendel 2006), number of genotypes surveyed (Lacape et al. 2007; Guo et al. 2007) and the accuracy of system used for resolving amplicons.

Polymorphic information content (PIC), an important parameter, helps in choosing SSRs for evaluating germplasm, gene tagging etc. (Peng and Lapitan 2005). In the present study, higher PIC value for gSSRs versus EST-SSRs suggesting that transcribed portions of the genome are conserved in the genomes (La Rota et al. 2005; Eujayl et al. 2002). In multiple reports, inconsistency in PIC values data have been reported (Kebede et al. 2007; Liu et al. 2000; Kalivas et al. 2011) which is attributed to the kind of germplasm explored, bottleneck in domestication (Thuillet et al. 2004; Vigouroux et al. 2005) and the kind of DNA markers (Liu et al. 2000; Gutierrez et al. 2002). Also, PIC values of the SSRs surveyed on diploid species were higher than the tetraploids. Most diploid species were wild except A-genome species. Wild species were not domesticated suggesting selection pressure for accumulating particular type of alleles was not applied is the reason for escalation in PIC values (Vigouroux et al. 2002; Qureshi et al. 2004).

Genetic characterization

The PIC values can guide us to select the most informative SSRs for calculating genetic divergence (Candida et al. 2006),

Table 6 Genetic similarity coefficients of 36 cotton species using two types of markers (EST-SSRs and gSSRs)

	G. herb	G. herb. afri	G. arbo	G. thur	G. klot	G. herk	G. davi	G. arid	G. raim	G. goss	G. loba	G. tril	G. laxu	G. hirs	G. tome	G. must	G. barb	G. darw	G. mari
G. herb	1																		
G. herb. afri	0.81	1																	
G. arbo	0.73	0.76	1																
G. thur	0.74	0.67	0.62	1															
G. klot	0.7	0.65	0.64	0.62	1														
G. herk	0.72	0.67	0.64	0.64	0.78	1													
G. davi	0.66	0.67	0.65	0.8	0.8	0.79	1												
G. arid	0.7	0.68	0.66	0.76	0.73	0.8	0.79	1											
G. raim	0.71	0.68	0.67	0.74	0.72	0.75	0.71	0.73	1										
G. goss	0.77	0.67	0.64	0.8	0.75	0.74	0.72	0.78	0.71	1									
G. loba	0.67	0.66	0.67	0.76	0.69	0.68	0.73	0.78	0.72	0.82	1								
G. tril	0.65	0.67	0.65	0.76	0.75	0.74	0.77	0.73	0.71	0.71	0.73	1							
G. laxu	0.73	0.62	0.63	0.77	0.69	0.7	0.65	0.69	0.68	0.75	0.67	0.67	1						
G. hirs	0.58	0.62	0.57	0.62	0.63	0.56	0.64	0.62	0.59	0.55	0.55	0.62	0.6	1					
G. tome	0.63	0.64	0.63	0.66	0.62	0.59	0.67	0.58	0.64	0.59	0.59	0.64	0.64	0.7	1				
G. must	0.59	0.56	0.62	0.64	0.62	0.58	0.62	0.58	0.59	0.57	0.59	0.64	0.64	0.7	0.76	1			
G. barb	0.63	0.61	0.62	0.64	0.61	0.55	0.61	0.59	0.59	0.58	0.55	0.65	0.62	0.85	0.77	0.79	1		
G. darw	0.62	0.65	0.67	0.65	0.61	0.58	0.63	0.61	0.65	0.57	0.55	0.65	0.6	0.76	0.85	0.77	0.85	1	
G. mari	0.58	0.61	0.58	0.58	0.58	0.6	0.57	0.58	0.56	0.57	0.55	0.65	0.56	0.71	0.65	0.65	0.71	0.73	1
	G. herb	G. herb. afri	G. arb	G. thur	G. klot	G. herk	G. davi	G. arid	G. raim	G. goss	G. loba	G. tril	G. laxu	G. hirs	G. tome	G. must	G. barb	G. darw	G. mari
G. lati	0.68	0.68	0.64	0.70	0.74	0.67	0.73	0.68	0.69	0.69	0.66	0.72	0.70	0.64	0.67	0.67	0.65	0.67	0.67
G. mori	0.62	0.67	0.62	0.63	0.67	0.61	0.66	0.64	0.61	0.61	0.59	0.65	0.65	0.74	0.64	0.64	0.67	0.67	0.65
G. palm	0.65	0.63	0.65	0.62	0.64	0.56	0.59	0.59	0.55	0.61	0.59	0.64	0.63	0.69	0.72	0.69	0.73	0.72	0.64
G. punc	0.63	0.66	0.63	0.65	0.63	0.58	0.63	0.63	0.58	0.58	0.58	0.63	0.65	0.71	0.67	0.64	0.70	0.70	0.71
G. yuca	0.61	0.64	0.62	0.62	0.64	0.61	0.64	0.68	0.63	0.57	0.55	0.64	0.58	0.70	0.64	0.62	0.65	0.73	0.75
G. lanc	0.65	0.65	0.64	0.66	0.69	0.70	0.68	0.69	0.64	0.69	0.68	0.69	0.72	0.59	0.61	0.65	0.61	0.61	0.61
G. anom	0.68	0.62	0.63	0.67	0.63	0.59	0.63	0.65	0.63	0.65	0.65	0.67	0.61	0.59	0.66	0.61	0.62	0.61	0.62
G. barb	0.67	0.70	0.63	0.62	0.61	0.62	0.63	0.61	0.64	0.61	0.59	0.63	0.59	0.56	0.61	0.61	0.56	0.60	0.62
G. capi	0.68	0.65	0.63	0.64	0.62	0.61	0.62	0.59	0.63	0.62	0.60	0.62	0.59	0.59	0.64	0.61	0.59	0.61	0.65
G. soma	0.61	0.50	0.61	0.68	0.66	0.64	0.68	0.68	0.64	0.66	0.62	0.72	0.62	0.58	0.58	0.63	0.62	0.64	0.59
G. stoc	0.60	0.57	0.61	0.62	0.60	0.57	0.61	0.61	0.61	0.59	0.59	0.61	0.57	0.57	0.61	0.61	0.58	0.62	0.61
G. long	0.70	0.62	0.58	0.65	0.65	0.67	0.64	0.65	0.59	0.70	0.65	0.66	0.67	0.50	0.59	0.62	0.55	0.50	0.53
G. inca	0.63	0.62	0.63	0.64	0.59	0.59	0.59	0.62	0.62	0.62	0.65	0.67	0.62	0.54	0.54	0.62	0.60	0.58	0.56
G. robi	0.56	0.50	0.52	0.64	0.55	0.54	0.61	0.57	0.55	0.59	0.61	0.58	0.55	0.52	0.52	0.55	0.52	0.53	0.54
G. aust	0.59	0.57	0.55	0.63	0.59	0.56	0.65	0.61	0.64	0.60	0.59	0.60	0.55	0.55	0.58	0.61	0.53	0.58	0.53
G. nels	0.56	0.55	0.55	0.62	0.54	0.53	0.56	0.59	0.61	0.60	0.62	0.53	0.59	0.56	0.58	0.58	0.56	0.57	0.52
G. bick	0.61	0.57	0.54	0.64	0.58	0.56	0.61	0.59	0.61	0.60	0.58	0.59	0.58	0.62	0.61	0.64	0.58	0.62	0.61

Table 6 continued

|	G. lati	G. mori	G. palm	G. punc	G. yuca	G. lanc	G. anom	G. barbo	G. capi	G. soma	G. stoc	G. long	G. inca	G. robi	G. aust	G. nels	G. bick
G. lati	1.00																
G. mori	0.75	1.00															
G. palm	0.68	0.77	1.00														
G. punc	0.66	0.76	0.76	1.00													
G. yuca	0.67	0.68	0.67	0.78	1.00												
G. lanc	0.72	0.63	0.64	0.65	0.60	1.00											
G. anom	0.68	0.62	0.61	0.62	0.65	0.70	1.00										
G. barbo	0.68	0.59	0.57	0.59	0.62	0.68	0.82	1.00									
G. capi	0.72	0.61	0.62	0.62	0.65	0.70	0.83	0.84	1.00								
G. soma	0.69	0.55	0.56	0.57	0.64	0.73	0.70	0.70	0.72	1.00							
G. stoc	0.65	0.56	0.58	0.55	0.62	0.67	0.73	0.69	0.78	0.70	1.00						
G. long	0.67	0.58	0.60	0.52	0.53	0.74	0.73	0.72	0.73	0.68	0.70	1.00					
G. inca	0.62	0.57	0.57	0.60	0.58	0.68	0.72	0.70	0.72	0.69	0.72	0.72	1.00				
G. robi	0.58	0.56	0.55	0.49	0.52	0.58	0.66	0.61	0.64	0.59	0.65	0.59	0.72	1.00			
G. aust	0.63	0.54	0.55	0.51	0.52	0.56	0.67	0.61	0.62	0.62	0.62	0.60	0.67	0.74	1.00		
G. nels	0.59	0.55	0.58	0.52	0.52	0.53	0.68	0.62	0.64	0.61	0.60	0.59	0.65	0.74	0.82	1.00	
G. bick	0.64	0.56	0.56	0.58	0.63	0.56	0.64	0.63	0.65	0.64	0.61	0.60	0.62	0.63	0.75	0.73	1.00

* Full names of species are given in Fig. 1

thus the number of SSRs can be reduced substantially (Candida et al. 2006) before initiating the genetic diversity and variety identification experiments (Macaulay et al. 2001; Masi et al. 2003; Jain et al. 2004). In this study, we proposed 22 (11 BNLs and 11 MGHES) out of the 75 SSRs—based on their high PIC values (PIC ≥ 5.0) and potential to amplify distinct DNA fragments for calculating the extent of genetic diversity among the 36 *Gossypium* species. Such findings were reported in multiple investigations including 39 SSRs for cotton genetics studies (Lacape et al. 2007) and 25 SSRs for *G. arboreum* accessions (Kantartzi et al. 2009). Large number of informative gSSRs had di-nucleotide repeats while larger portion of informative EST-SSRs had tri-nucleotide motif. Dominance of trimeric SSRs over the others is possibly due to the inhibition of non-trimeric SSRs in coding regions of genes for avoiding chances of frame shift mutations (La Rota et al. 2005). Another reason is that the high portion of trinucleotide repeats in coding regions may be due to the exertion of selection pressure for selecting particular single amino acid stretches (Morgante et al. 2002). Also the most informative SSRs contained ≥10 repeats which are in agreement with the previous studies (Vigouroux et al. 2002; Qureshi et al. 2004; Kantartzi et al. 2009). It was also observed that position of the SSR loci on the chromosome has no effect on the corresponding PIC values (Lacape et al. 2003, 2007).

In the present study, correlation was not observed between the rate of polymorphism and repeat motif type that is contradictory to the previous findings of Lacape et al. (2007). They found repeat motif type dependent polymorphism in cotton and showed that SSRs with GA repeat motif type exhibited higher PIC value with more number of alleles than SSRs with CA repeat motif while Thuillet et al. (2004) found SSRs with CA repeat motif type exhibited significantly fewer alleles than GA SSRs in wheat. This might be due to difference in nucleotide distribution in different genomes but still further investigations are required with higher number of markers to confirm whether polymorphism is repeat motif type dependent or not.

Performance of microsatellite between A, D and AD-genome species

The SSRs did not amplify distinctive fragments with genomic DNA of A-genome species but produced clear bands in the D- and AD-genome species were placed on D-subgenome of allotetraploid cotton (Lacape et al. 2003; Mei et al. 2004). Only (12) markers were D-genome specific reflecting substantial divergence of D-genome species from D-subgenome of allotetraploid cotton (Brubaker et al. 1999; Adams and Wendel 2004; Guo et al. 2006; Wu et al. 2007). Amplicon sizes (101–300 bp) of a number of SSRs were different in AD-genome species and their diploid ancestral species (A and D). Such commonalities have been
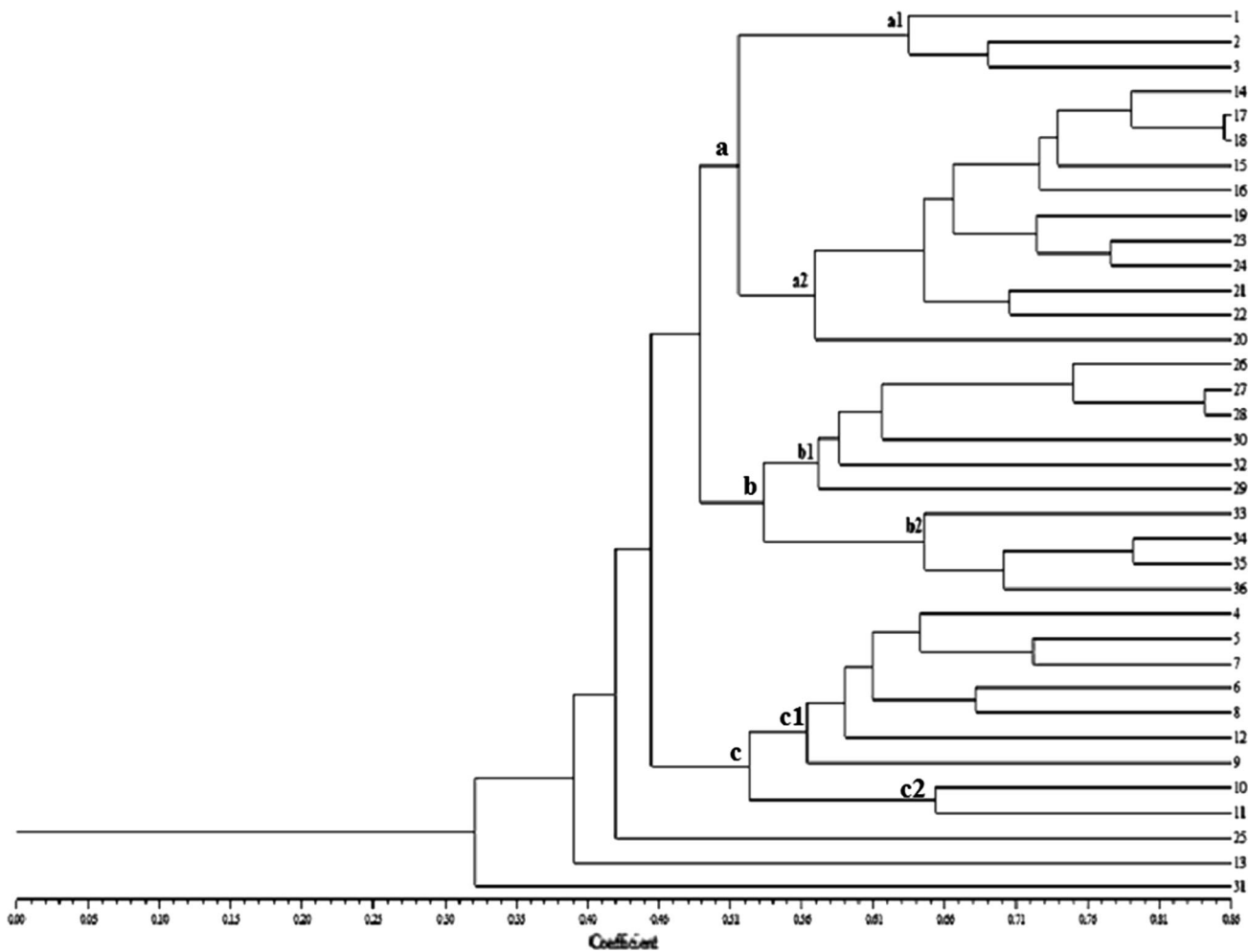
Fig. 4 Phylogenetic analysis of 36 cotton species with combined data set of ESTs and gSSRs: The code represents the species as *G. herbaceum* A$_1$ (1), *G. herbaceum* var. *africanum* A$_1$ (2), *G. arboreum* A$_2$ (3), *G. thurberi* D$_1$ (4), *G. klotzschianum* D$_2$ (5), *G. harknessii* D$_{2-2}$ (6), *G. davidsonii* D$_3$ (7), *G. aridum* D$_4$ (8), *G. raimondii* D$_5$ (9), *G. gossypioides* D$_6$ (10), *G. lobatum* D$_7$ (11), *G. trilobum* D$_8$ (12), *G. laxum* D$_9$ (13), *G. tomentosum* AD$_3$ (14), *G. hirsutum* AD$_1$ (15), *G. mustilinum* AD$_4$ (16), *G. barbadense* AD$_2$ (17), *G. darwinii* AD$_5$ (18), *G. hirsutum* "marie-galante" AD (19), *G. hirsutum* "latifolium" AD (20), *G. hirsutum* "morrillii" AD (21), *G. hirsutum* "palmeri" AD (22), *G. hirsutum* "punctatum" AD (23), *G. hirsutum* "yucatanense" AD (24), *G. hirsutum* "lanceolatum" AD (25), *G. anomalum* B$_1$ (26), *G. barbosanum* B$_3$ (27), *G. capitis-viridis* B$_4$ (28), *G. somalense* E$_2$ (29), *G. stocksii* E$_1$ (30), *G. longicalyx* F$_1$ (31), *G. incanum* E$_4$ (32), *G. robinsonii* C$_1$ (33), *G. australe* G$_2$ (34), *G. nelsonii* G$_3$ (35), *G. bickii* G$_1$ (36)

reported (Syed et al. 2001) which are due to type/number of repeat motif and flanking sequences (Buteler et al. 1999).

Amplified fragments size distribution of species containing AD- and D-genomes was dispersive while alleles amplified in A-genome species were of intense distribution (Fig. 3). Our outcomes are contradictory to Liu et al. (2006) who found dispersive distribution of fragment sizes in *G. arboreum* and relatively concentrated distribution in *G. hirsutum*.

Cross species amplification and genome specificity

Genome/species-specific SSRs can be useful in monitoring introgression of specific genomic portion of the donor species into the adaptive species (Guo et al. 2007), that can be instrumental in assigning species to unknown plants and in distinguishing cotton species. In this study 15 genome/species-specific SSRs were observed.

The transferability of SSRs derived from tetraploids to diploids indicates evolution of all genomes from one ancestor. We reported a high transferability rate in A-genome as compared to D-genome, indicating that D-subgenome in tetraploids deviated during polyploidization from their progenitor D-genome (Liu et al. 2006). Second, the higher transferability rate in A-genome species may be due to larger size of A-genome (Edwards et al. 1974; Reinisch et al. 1994).

In this study, gSSRs (BNLs) showed low transferability (37.28 %) and high polymorphism rate across the species versus high transferability (54.72 %) and low
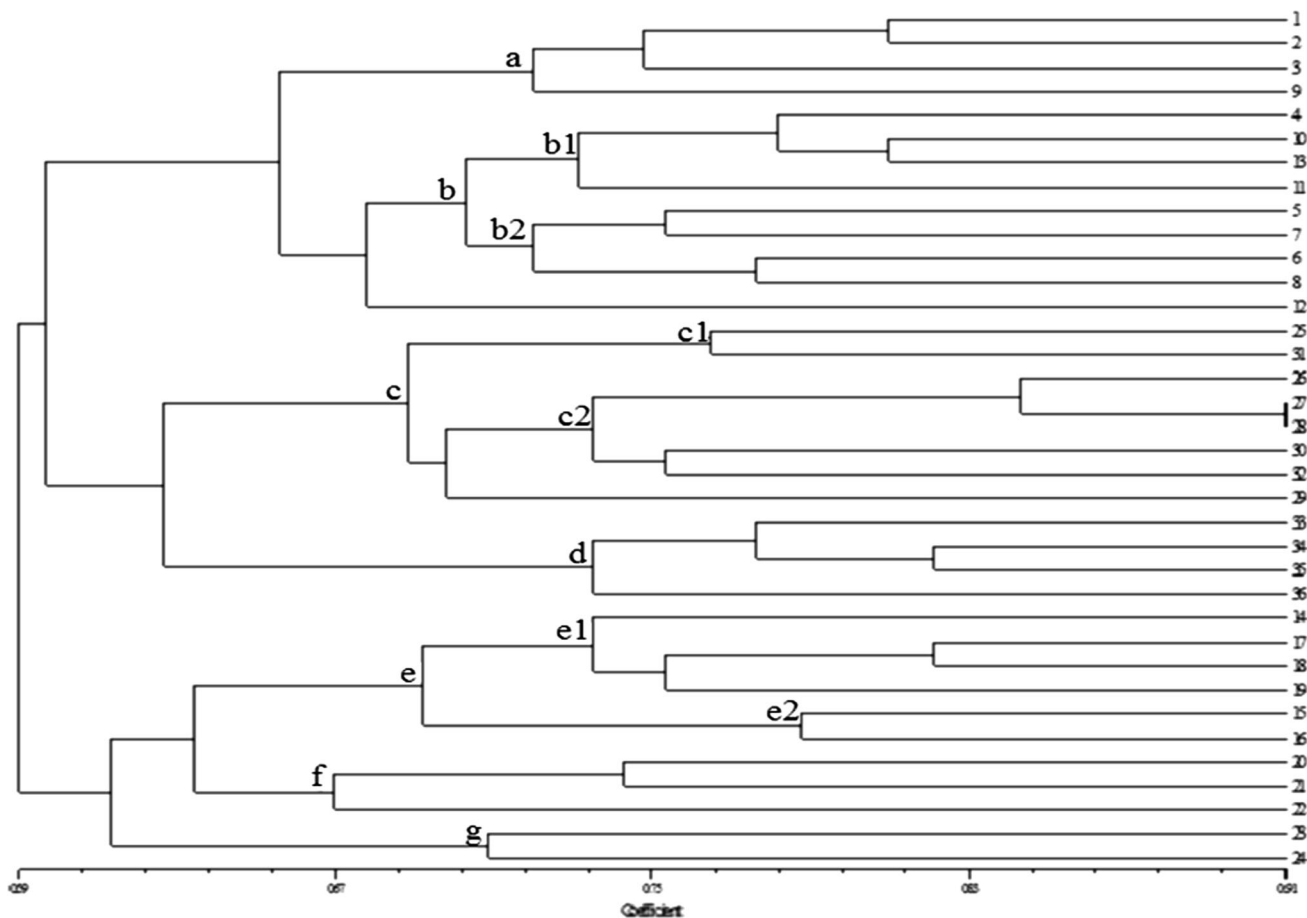
Fig. 5 Clustering of species with EST-SSRs The code represents the species as *G. herbaceum* A_1 (1), *G. herbaceum* var. *africanum* A_1 (2), *G. arboreum* A_2 (3), *G. thurberi* D_1 (4), *G. klotzschianum* D_2 (5), *G. harknessii* D_{2-2} (6), *G. davidsonii* D_3 (7), *G. aridum* D_4 (8), *G. raimondii* D_5 (9), *G. gossypioides* D_6 (10), *G. lobatum* D_7 (11), *G. trilobum* D_8 (12), *G. laxum* D_9 (13), *G. tomentosum* AD_3 (14), *G. hirsutum* AD_1 (15), *G. mustilinum* AD_4 (16), *G. barbadense* AD_2 (17), *G. darwinii* AD_5 (18), *G. hirsutum* "marie-galante" AD (19), *G. hirsutum* "latifolium" AD (20), *G. hirsutum* "morrillii" AD (21), *G. hirsutum* "palmeri" AD (22), *G. hirsutum* "punctatum" AD (23), *G. hirsutum* "yucatanense" AD (24), *G. hirsutum* "lanceolatum" AD (25), *G. anomalum* B_1 (26), *G. barbosanum* B_3 (27), *G. capitis-viridis* B_4 (28), *G. somalense* E_2 (29), *G. stocksii* E_1 (30), *G. longicalyx* F_1 (31), *G. incanum* E_4 (32), *G. robinsonii* C_1 (33), *G. australe* G_2 (34), *G. nelsonii* G_3 (35), *G. bickii* G_1 (36)

polymorphism rate exhibited by EST-SSRs, primarily because of their conserved nature (Cho et al. 2000; Thiel et al. 2003). The EST-SSRs derived from fiber tissues showed high level of transferability in diploid genomes, confirming the presence of fiber related genes in all the cotton genomes. Phenomenon of transferability has also been reported in other crop species (Kuleung et al. 2004; Saha et al. 2004).

Genetic relationship of tetraploid species with their wild relatives

Among A-genome species, *G. herbaceum* (A_1) was found relatively closer to *G. hirsutum* (AD_1) while *G. arboreum* (A_2) showed more closeness toward *G. barbadense*. Among D-genome species *G. raimondii* (D_5) was more similar to *G. hirsutum* (AD_1) (0.667). In another study, *G. herbaceum* was

found more genetically near to *G. hirsutum* (0.69) as compared to *G. arboreum* (0.52). It has also been observed that *G. raimondii* (D_5) and *G. gossypioides* (D_5) are genetically more comparable to *G. hirsutum* (AD_2) and *G. barbadense* (AD_2) (Kebede et al. 2007). In few cytogenetic studies, it was elucidated that *G. herbaceum* is more comparable to the ancestor species of tetraploid cotton than *G. arboreum* (Endrizzi et al. 1985; Wendel 1989; Percival and Kohel 1990).

Genetic diversity and phylogenetic relationship in the genus *Gossypium*

For evolutionary studies of cotton species, basic requirement is to workout their phylogeny (Khan et al. 2000; Abdalla et al. 2001; Paterson et al. 2002) and to estimate the extent of genetic diversity (Khan et al. 2000). For genetic diversity and phylogeny studies of cotton species
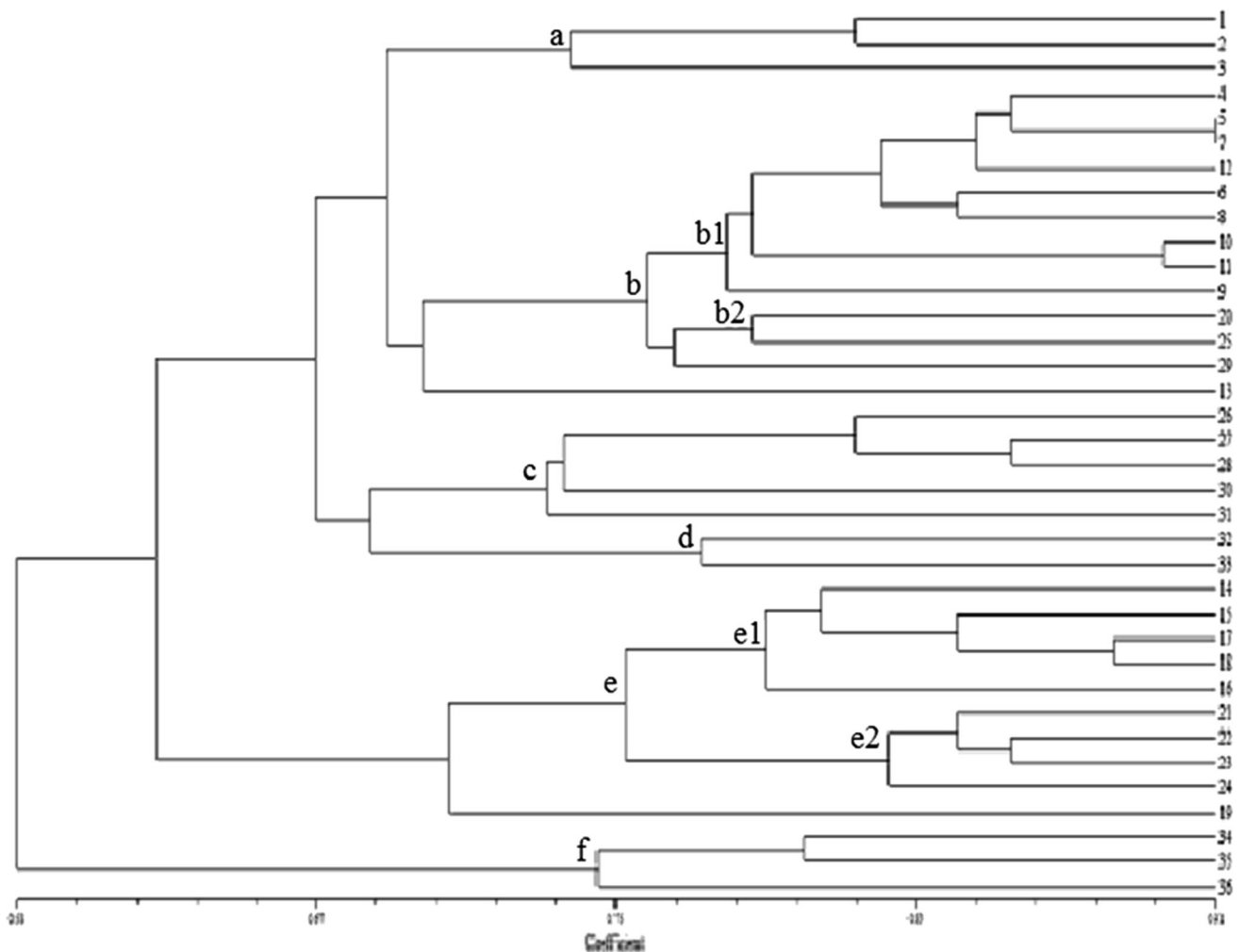
Fig. 6 Clustering of species with gSSRs The code represents the species as *G. herbaceum* A_1 (1), *G. herbaceum* var. *africanum* A_1 (2), *G. arboreum* A_2 (3), *G. thurberi* D_1 (4), *G. klotzschianum* D_2 (5), *G. harknessii* D_{2-2} (6), *G. davidsonii* D_3 (7), *G. aridum* D_4 (8), *G. raimondii* D_5 (9), *G. gossypioides* D_6(10), *G. lobatum* D_7 (11), *G. trilobum* D_8 (12), *G. laxum* D_9 (13), *G. tomentosum* AD_3 (14), *G. hirsutum* AD_1 (15), *G. mustilinum* AD_4 (16), *G. barbadense* AD_2 (17), *G. darwinii* AD_5 (18), *G. hirsutum* "marie-galante" AD (19), *G. hirsutum* "latifolium" AD (20), *G. hirsutum* "morrillii" AD (21), *G. hirsutum* "palmeri" AD (22), *G. hirsutum* "punctatum" AD (23), *G. hirsutum* "yucatanense" AD (24), *G. hirsutum* "lanceolatum" AD (25), *G. anomalum* B_1 (26), *G. barbosanum* B_3 (27), *G. capitis-viridis* B_4 (28), *G. somalense* E_2 (29), *G. stocksii* E_1 (30), *G. longicalyx* F_1 (31), *G. incanum* E_4 (32), *G. robinsonii* C_1 (33), *G. australe* G_2 (34), *G. nelsonii* G_3 (35), *G. bickii* G_1 (36)

various methods based on morphology, meiotic behavior, genetic and molecular techniques have been deployed. In this study, two types of SSR markers (EST-SSRs and genomic SSRs) have been utilized to study the phylogenetic relationship among cotton species. It is clear from this study that diploid species are genetically more diverse from each other as compared to the tetraploid germplasm. In this study, low to moderate level of genetic similarity among *Gossypium* species has been estimated. This report is consistent with the findings of Abdalla et al. (2001)—calculated relatively high estimates of genetic diversity among *Gossypium* species using AFLP marker system.

Among tetraploid germplasm lowest genetic similarity of *G. hirsutum* "punctatum" (ancient cultivated race: Brubaker and Wendel 1994; Lacape et al. 2007) with the other tetraploid germplasm reveals the existence of unique/useful alleles in this species. Such races could be a promising source for broadening the extent of genetic diversity within cultivated cotton. *G. hirsutum* "latifolium"—genetically more close to *G. hirsutum* (Lacape et al. 2007), would have least obstacles (Lubbers and Chee 2009) in attempting crosses. Within tetraploids, high genetic similarity estimates between *G. barbadense* and *G. darwinii* and *G. tomentosum* and *G. hirsutum* are in consistent with the earlier reports (Liu et al. 2000; Wendel and Cronn 2003; Lacape et al. 2007). Also, the variation in restriction sites in cpDNA and rDNA and in allozyme (14 enzyme systems) demonstrated more distinctiveness of *G. tomentosum* from *G. hirsutum* (0.82) than from

G. barbadense 0.65 (Dejoode and Wendel 1992) which is contradictory to our findings.

In cluster analysis, A-genome species made sister cluster with D-genome species using the data of both EST-SSRs and gSSRs while these made sister cluster with AD-genome species with combined data in a major cluster 'A'. Such commonalities have been found using RAPD markers (Khan et al. 2000), cpDNA, ITS and combined data set based phylogenies (Seelanan et al. 1997) and cpDNA restriction site based phylogeny (Wendel et al. 1992). The sister clustering of A-genome species with the AD-genome species strengthens the concept that A-genome is the cytoplasmic donor of AD-genome (Wendel 1989). It is likely that genomes of A-genome species have chromosome of larger size and more recurrence of repetitive sequences in their genome as compared to D-genome (Geever 1980), thus producing opportunities to amplify similar sequences (homology) among the genomes. Also, the rate of evolution of D-genome is faster than A-genome (Adams and Wendel 2004). All B-genome species showed close relationship with each other using EST-SSRs, gSSRs and combined data sets (Figs. 4, 5, 6) which is consistent with the previous reports (Wendel and Albert 1992), and the 'E' genome species were grouped with 'B' genome species. 'G' genome species also showed closeness with each other using three data sets. All 'G' genome species were grouped with 'C' genome species using EST-SSRs and combined data set while with gSSRs 'C' genome species were grouped with 'E' genome species whereas 'G' genome species grouped separately and appeared to be distantly related to all other genomic groups, illuminating that all 'G' genome species share a common ancestor (Fryxell et al. 1992; Liu et al. 2001).

Gossypium genomes, eight in number, comprise of four major lineages, spread in three continents. In Australia C-, G-, and K-genome species were found while in America D-genome species were present. While in Africa/Arabia first lineage of the A-, B-, and F-genome species, and second lineage of the E-genome species were found (Fryxell 1979; Fryxell et al. 1992). Clustering of (B and E) and (C and G) genome in one cluster with EST-SSRs and combined data is probably because of evolving from a common ancestry. A large number of genomic data showed consistency with the aforementioned taxonomy of the cotton species (Wu et al. 2007).

With three types of data sets, all D-genome species were grouped in one cluster except G. raimondii (D_5) that grouped with A-genome species using EST-SSR data. The grouping of G. raimondii D_5 with A-genome species with EST-SSR data set was not surprising as in several studies it has been found isolated from rest of the D-genome species (Fryxell 1979; Parks et al. 1975; Phillips 1966). It is disjunct geographically from rest of the subgenus. In D-genome species cluster (G. aridum D_4 and G. herkensii D_{2-2}),

(G. gossypioides D_6 and G. lobatum D_7) and (G. klotzschianum D_2 and G. davidsonii D_3) showed closeness with all three data sets. The position of remaining three D-genome species (G. thurberi D_1, G. trilobum D_8 and G. laxum D_9) remained unresolved as these species were grouped in different clusters with both the independent and combined data of EST and gSSRs while in few studies close genetic relationship was reported between G. thurberi D_1 and G. trilobum D_8 (Wu et al. 2007; Guo et al. 2007). The close relationship of G. klotzschianum D_2 and G. davidsonii D_3 is congruent with the earlier reports (Wu et al. 2007; Wendel and Albert 1992; Guo et al. 2008).

The position of G. longicalyx, only F-genome species, also remained unresolved in the present study. With the gSSR data, this species has shown close association with the E-genome species (G. stockii E1) while with EST-SSR data set it showed kinship to uncommon tetraploid species 'G. lanceolatum (AD)'. In earlier reports, high genetic similarity of G. longicalyx was reported with A-genome or allotetraploid derivatives (Wendel 1989). However, in majority of the earlier reports, relatively isolated position of G. longicalyx was reported (Phillips and Strickland 1966; Saunders 1961; Wu et al. 2007).

The phylogenetic tree of species with EST-SSR data is closely resembled with the tree obtained from the combined data set. On the basis of these findings it is, therefore, suggested that relatively limited number of EST-SSRs instead of using high number of markers can be instrumental in resolving phylogenies at species level. Moreover, this study confirms the usefulness of limited number EST-SSRs for fingerprinting of geographically isolated species (Gutierrez et al. 2002; Rahman et al. 2002b; Shaheen 2005) and determining the phylogeny of the species.

Conclusion

Genomic SSR markers are more informative for fingerprinting and for estimation of genetic diversity among the cotton species because of occurrence of more alleles in microsatellite regions. EST-SSRs are more powerful for determining the changes occurred in result of selection during domestication. Only few EST-SSR markers are sufficient for resolving phylogenetic relationship of cotton species instead of using large number of SSR markers. Number of repeats per locus showed positive correlation with the number of alleles amplified, allele size range and polymorphism information content. Repeat motif type and position of loci on the chromosome have no effect on polymorphism rate.

Tetraploid species amplified more alleles than the diploid species. The presence of 18 alleles (16.51 %) having 0.10 or lower frequency, is one of the indicators of mutations or introgressions of new alleles in germplasm pool. G.

mustelinum, *G. tomentosum*, *G. darwinii*, *G. hirsutum* "yucatanense" and *G. herbaceum* var. *africanum* were found genetically dissimilar which can utilized in breeding programs for broadening the genetic base of the cultivated cotton species, a way for achieving the sustainability in cotton production in the changing climate.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229

Adams KL, Wendel JF (2004) Exploring the genomic mysteries of polyploidy in cotton. Biol J Linn Soc 82:573–581

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage map. Genome 36:181–186

Beasley JO (1940) The origin of American tetraploid *Gossypium* species. Am Nat 74:285–286

Brubaker CL, Wendel JF (1994) Re-evaluating the origin of domesticated cotton (*Gossypium hirsutum*, Malvaceae) using nuclear restriction fragment length polymorphism (RFLPs). Am J Bot 81:1309–1326

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Buteler MI, Jarret RL, LaBonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. Theor Appl Genet 99:123–132

Candida H, Bertini DM, Schuster I, Sediyama T, Barros EGD, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol 29(2):321–329

Cho YG, Ishii T, Temnykh S, Chen X, Lopovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912–922

Dejoode DR, Wendel JF (1992) Genetic diversity and origin of the Hawaiian islands cotton, *Gossypium tomentosum*. Am J Bot 79:1311–1319

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bulletin 19:11–15

Edwards GA, Endrizzi AJE, Stein R (1974) Genome DNA content and chromosome organization in *Gossypium*. Chromosoma 47:309–326

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16:551–558

Endrizzi JE, Turcotte EL, Kohel RJ (1984) Qualitative genetics, cytology and cytogenetics. In: Kohel RJ, Lewis CF (eds) Madison. WI, Am Soc Agron, pp 81–129

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology and evolution of *Gossypium*. Advan in Genet 23:271–375

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsaltellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Fryxell PA (1979) The natural history of the cotton tribe. Texas A and M University Press, Texas

Fryxell PA, Craven LA, Stewart JM (1992) A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. Syst Bot 17:91–114

Geever RF (1980) The evolution of single-copy nucleotide sequences in the genome of *G. hirsutum*. PhD dissertation, University of Arizona, Tuscon, USA

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

Guo WZ, Sang ZQ, Zhou BL, Zhang TZ (2007) Genetic relationships of D-genome species based on two types of EST-SSR markers derived from *G. arboreum* and *G. raimondii* in *Gossypium*. Plant Sci 172:808–814

Guo Y, McCarty JC, Jenkins JN, Saha S (2008) QTLs for node of first fruiting branch in a cross of an upland cotton, *Gossypium hirsutum* L., cultivar with primitive accession Texas 701. Euphytica 163:113–122

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC (2002) Genetic distance among selected cotton genotypes and its relationship with F2 performance. Crop Sci 42:1841–1847

Hutchinson JB (1951) Intra-specific differentiation in *Gossypium hirsutum*. Heredity 5:169–193

Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. Theor Appl Genet 103:547–554

Jain S, Jain RK, McCouch S (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. Theor Appl Genet 109:965–977

Kalivas A, Xanthopoulos F, Kehagia O, Tsaftaris AS (2011) Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. Genet Mol Res 10(1):208–217

Kantartzi SK, Ulloa M, Sacks E, Stewart JMD (2009) Assessing genetic diversity in *Gossypium arboreum* L. cultivars using genomic and EST-derived microsatellites. Genetica 136:141–147

Kebede H, Burow G, Raviprakash G, Randy RA (2007) A-genome cotton as a source of genetic variability for Upland cotton (*Gossypium hirsutum*). Genet Resour Crop Ev 54:885–895

Khan SA, Hussain D, Askari E, Stewart JM, Malik KA, Zafar Y (2000) Molecular phylogeny of *Gossypium* species by DNA fingerprinting. Theor Appl Genet 101:931–938

Kuleung C, Baenziger PS, Dweikat I (2004) Transferability of SSR markers among wheat, rye and triticale. Theor Appl Genet 108:1147–1150

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice wheat and barley. BMC Genomics 6:23–35

Lacape JM, Nguyen TB, Thibivilliers S, Bojinnov TB, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium*

hirsutum × *Gossypium barbadense* backcross population. Genome 46:612–626

Lacape JM, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. Mol Breed 19:45–58

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7:R14

Liu S, Cantrell RG, McCarty JC, Stewart JM (2000) Simple sequence repeat based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459–1469

Liu S, Griffey CA, Shagai Maroof MA (2001) Identification of molecular markers associated with adult plant resistance to powdey mildew in common wheat cultivar Massey. Crop Sci 41:1268–1275

Liu D, Guo XP, Lin Z, Nie YC, Zhang X (2005) Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genet Resour Crop Ev 53:1145–1152

Liu DQ, Guo XP, Lin ZX, Nie YC, Zhang XL (2006) Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genet Resour Crop Ev 53:1145–1152

Lubbers EL, Chee PW (2009) The worldwide gene pool of *Gossypium hirsutum* and its improvement. In: Paterson AH (ed) Genetics and genomics of cotton, vol 3. Springer, New York, pp 23–52

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative ''genotyping set'' of barley SSRs. Theor Appl Genet 106:801–809

Masi P, Spagnoletti Zeuli PL, Donini P (2003) Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). Mol Breeding 11:303–313

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fibre related traits in cotton. Theor Appl Genet 108:280–291

Menzel MY (1954) A cytological method for genome analysis in *Gossypium*. Genetics 40:214–223

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Nei M (2007) The new mutation theory of phenotypic evolution. PNAS 104:12235–12242

Nei N, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Parks CR, Ezell WI, William DE, Dreyer DL (1975) The application of flavonoid distribution to taxonomic problems in the genus *Gossypium*. Bull Torrey Bot Club 102:350–361

Paterson A, Estill J, Rong J, Williams D, Marler B (2002) Toward a genetically anchored physical map of the cotton genomes. Cotton Sci 14((Suppl.)):31

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Rahman M, Rokhsar DS, Wang X, Schmutz J (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinable cotton fibers. Nature. doi:10.1038/nature11798

Peakall R, Gilmore S, Keys W, Morgante M, Rafaske A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeat (SSR) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Peng JH, Lapitan NLV (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomic 5:80–96

Percival AE, Kohel RJ (1990) Distribution, collection, and evaluation of *Gossypium*. Adv Agron 44:225–256

Phillips LL (1966) The cytology and Phylogenetic of the diploid species of *Gossypium*. Am J Bot 53(4):328–335

Phillips LL, Strickland MA (1966) The cytology of a hybrid between *Gossypium hirsutum* and *G. longicalyx*. Can J Genet Cytol 8:91–95

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8:112–123

Rahman M, Hussain D, Zafar Y (2002a) Estimation of genetic divergence among elite cotton (*Gossypium hirsutum* L.) cultivars/genotypes by DNA fingerprinting technology. Crop Sci 42:2137–2144

Rahman M, Malik TA, Aslam N, Asif M, Ahmad R, Khan IA, Zafar Y (2002b) Optimisation of PCR conditions to amplify microsatellite loci in cotton (*Gossypium hirsutum* L.). Int J Agric Biol 2(4):282–284

Rahman M, Asif M, Ullah I, Malik KA, Zafar Y (2005a) Overview of cotton genomic studies in Pakistan. In: Plant and Animal Genome Conference XIII. San Diego, CA, USA.

Rahman M, Hussain D, Malik TA, Zafar Y (2005b) Genetics of resistance to cotton leaf curl disease in *Gossypium hirsutum*. Plant Pathol 54:764–772

Rahanan M, Tabassum N, Ullah I, Asif M, Zafar Y (2008) Studying the extent of genetic diversity among *Gossypium arboreum* L. genotypes/cultivars using DNA fingerprinting. Genet Resour Crop Ev 55:331–339

Rahman M, Zafar Y, Paterson AH (2009) *Gossypium* DNA markers: types, numbers and uses. In: Paterson AH (ed) Genetics and genomics of cotton, vol 3. Springer, New York, pp 101–139

Rahman M, Asif M, Shaheen T, Tabbasam N, Zafar Y, Paterson AH (2011) Marker-assisted breeding in Higher Plants. In: Lichtfouse Eric (ed) Sustainable agriculture reviews 6; alternative farming systems, biotechnology, drought stress and ecological fertilization. Springer, Berlin, pp 39–76

Rahman M, Shaheen T, Tabbasam N, Iqbal MA, Ashraf M, Zafar Y, Paterson AH (2012) Genetic resources in cotton and their improvement. Agron Sustain Dev 32:419–432

Reinisch MJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Ramesh V, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Saha MC, Main MAR, Eujayl L, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Saunders JH (1961) The wild species of *Gossypium* and their evolutionary history. Oxford university press, london

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe. Syst Bot 259–290:22

Shaheen HL (2005) Global view of genetic diversity among cotton cultivars/genotypes by microsatellite analysis. MPhill thesis, Dept of Bot, University of Agriculture Faisalabad, Pakistan

Shaheen T, Zafar Y, Rahman M (2010) Detection of single nucleotide polymorphisms in the conserved ESTs regions of *Gossypium arboreum*. Electron J Biotechnol 13(5). doi:10.2225/vol13-issue5-fulltext-3

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16:491–501

Stephens SG (1967) Evolution under domestication of the new world cottons (*Gossypium* spp). Cienc Cult 19:118–134

Stewart, JMCD, (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: F.N. Constable GA (Editor), Challenging the future. Proceedings of the world cotton research conference-1, Brisbane Australia, CSIRO, Melbourne, pp. 313–327

Syed NH, Lee HS, Mei M, Thaxton P, Stelly DM, Chen ZJ (2001) Variability and evolution of microsatellite loci in cotton

(*Gossypium*) diploid and polyploidy genomes. Plant and Animal Genome IX Conference. Town and Country Hotel, San Diego, CA, January 13–17

Thiel T, Michalek V, Graner A (2003) Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Thuillet AC, Bataillon T, Sourdille P, David JL (2004) Factors affecting polymorphism at microsatellite loci in bread wheat (*Triticum aestivum* L. Thell): effects of mutation processes and physical distance from the centromere. Theor Appl Genet 108(2):368–377

Udall JA, Wendel JF (2006) Polyploidy and crop improvement. Plant Genome suppl Crop Sci 46:S3–S14

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith SC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19:1251–1260

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites. Genetics 169(3):1617–1630

Wang HY, Wei YuM, Yan ZH, Zheng YL (2007) EST-SSR DNA polymorphism in durum wheat (*Triticum durum* L.) collections. J Appl Genet 48:35–42

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Guo C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nature. doi:10.1038/ng.2371

Wendel JF (1989) New world tetraploid cottons contain old world cytoplasm. Proc Natl Acad Sci USA 86:4132–4136

Wendel JF, Albert VC (1992) Phylogenetics of the cotton genus (*Gossypium*): character- state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17:115–143

Wendel JF, Cronn R (2003) Polyploidy and the evolutionary history of cotton. Adv in Agro 78:139–186

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310

Wu YX, Daud MK, Chen L, Zhu SJ (2007) Phylogenetic diversity and relationship among *Gossypium* germplasm using SSRs markers. Plant Syst Evol 268:199–208

Zhang J, Lu Y, Cantrell RG, Hughs E (2005) Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the southwest USA. Crop Sci 45:1483–1490