Full Length Research Paper

Simple sequence repeats (SSR) and interspersed sequence repeats (ISSR) markers for genetic diversity analysis among selected genotypes of *Gossypium arboreum* race *'bengalense'*

Khushboo Sethi[1], Priyanka Siwach[1]* and Surender Kumar Verma[2]

[1]Deparment of Biotechnology, Chaudhary Devi Lal University, Sirsa, Haryana, India.
[2]Central Institute of Cotton Research, Regional Station, Sirsa, Haryana, India.

Genetic diversity among 65 selected genotypes of *Gossypium arboreum* race bengalense was explored using 62 simple sequence repeats (SSR) and 73 interspersed sequence repeats (ISSR) markers. The SSR primers produced a total of 170 alleles (all polymorphic), while ISSRs yielded 281 bands of which only 94.3% were polymorphic. Utility of various markers were evaluated by calculating different parameters like polymorphic information content (PIC), marker index (MI), and discriminative ability (D), on the basis of which 21 SSR and 53 ISSRs primers were found very efficient for genetic diversity analysis. ISSR outperformed the SSR for discriminative ability as it yielded higher number of banding patterns (ISSR-658, SSR-175), greater numbers of polymorphic bands/assay (ISSR-3.63 and SSR-2.7) and higher D values (ISSR-0.862 and SSR-0.442). Values of I (SSR-0.740 and ISSR-0.421) and He (SSR-0.433 and ISSR-0.262) indicated SSRs as more suitable for characterizing the species in terms of abundance and evenness of alleles. A slight difference was observed in terms of MI values of the SSR (1.20) and ISSR (MI-1.38), showing an edge for ISSR in detecting overall polymorphism among given genotypes. Phylogenetic analysis was carried out by SSR, ISSR as well as combined datasets of markers. The highest value of cophenetic correlation coefficient was obtained for ISSR (r=0.94), followed by combined datasets (r=0.91) and SSR markers (r=0.87).

Key words: Molecular markers, marker index, polymorphism information content, genetic diversity, *Gossypium arboreum,* discrimination coefficient.

INTRODUCTION

Cotton (*Gossypium* spp.) is one of the principal cash crops, providing most of the world's natural textile fiber.

The genus *Gossypium* (family *Malvaceae*) comprises nearly 45 diploid and 5 allotetraploid species. Spinnable

fibers are obtained only from four species; two allotetraploids or new world cotton (*Gossypium hirsutum* and *Gossypium barbedense*) and two diploids or Asiatic/old world cotton (*Gossypium herbaceum* and *Gossypium arboreum*).

India is the original home of domestication, diversification and development of Asiatic cultivated cottons. From 1500 BC to 1700 AD, India was recognized as the cradle of the cotton industry. The Indian monopoly in cotton muslins was broken up by the industrial revolution in England; new world cotton largely replaced the Asiatic cotton (Mohan et al., 2006). The major cause for this change was the unsuitability of diploid cotton fibers for mechanized spinning because of short length (<23 mm), high coarseness (>5.0 micronaire) and poor strength (<20 g/tex at 3.2 mm gauge) (Kulkarni et al., 2009). At present, tetraploid cotton (dominantly *G. hirsutum*) occupies a major fraction (>90%) of world cotton cultivation due to its suitability to mechanized harvesting and spinning. However, in marginal and drought-prone environments of Asia, diploid cottons are still popularly cultivated. This is because of certain inherent traits (which the tetraploids lack) like drought and salinity tolerance (Tahir et al., 2011); resistance to several pests including bollworms (Dhawan et al., 1991), aphids and leafhoppers (Nibouche et al., 2008); and diseases like rust, fungal (Wheeler et al., 1999) and viral (Akhtar et al., 2010).

Of the two diploid cultivated species, *G. arboreum* is more popular due to its suitability to a wider range of environments, and better fiber and plant features (Mohan et al., 2006). From its origin, dispersal and domestication of *G. arboreum* germplasm in different directions resulted in six races- indicum, burmanicum, cernuum, sinense, bengalense and soudanase. India is the only country where all six races are cultivated, the major share of which is contributed by 'bengalense' (cultivated commonly across central and North India).

G. arboreum germplasm constitutes an indispensable gene pool for modern cotton improvement programs. However, due to continuous selective breeding and selection during the last few decades, the germplasm is facing the constraints of narrow genetic base. Knowledge of genetic variation among *G. arboreum* germplasm is essential for future developments. Equally essential are the efficient tools which enable the detection of higher levels of genetic diversity (Ulloa et al., 2007). During the last two decades, various molecular markers have been extensively used for genetic diversity studies across species. *G. arboreum* germplasm has been explored with markers like randomly amplified polymorphic DNA (RAPD) (Deosarkar et al., 2010), interspersed sequence repeats (ISSR) (Bardak and Balek, 2012), simple sequence repeats (SSRs) (Noormohammadi et al., 2013a) etc.; and all studies report low polymorphism. Considering the edge of SSR and ISSR markers in cultivar fingerprinting and diversity studies, the present

study was planned to evaluate the utility of these two methods for assessing genetic diversity as well as phylogenetic analysis among elite genotypes of *G. arboreum* race 'bengalense'.

MATERIALS AND METHODS

Plant materials and DNA extraction

Seeds from 65 elite genotypes belonging to race 'bengalense' of *G. arboreum* (Table 1) were procured from the Central Institute of Cotton research (CICR), Regional Station, Sirsa, Haryana, India. The cotton plants were cultivated in two rows of 6 m length with 30 cm interplant distance in the experimental field of CICR, Sirsa, in a completely randomized design (CRD) with three replications. Fresh and young leaves of randomly selected single plants of each genotype were subjected to total genomic DNA extraction using the cetyltrimethylammonium bromide (CTAB) method (Saghai et al., 1984) with certain modifications. The quality and quantity of extracted DNA was examined by agarose gel (0.8%) electrophoresis and ultra violet (UV)-spectrophotometry, respectively.

SSR amplification

One hundred microsatellite primer pairs were obtained from Brookhaven National Laboratory (BNL), MGHES (M for Mississipi, GH- *G. hirsutum*, E- EST, S-SSR), CIR (CIRAD), JESPR (named after the names of Principal Investigators), Nanjing Agricultural University (NAU), and MUSS (M- Microsatellite, U- Last name of Principal Investigator, SS- Simple Sequences). Out of 100 primers, only 62 gave polymorphism and reproducible banding patterns and hence were selected for the present study (Table 2). The sequence information of these SSRs is available at http://www.cottonmarker.org.

Polymerase chain reaction (PCR) amplification was performed in a volume of 20 µl containing 2 µl of DNA (50 ng/µl), 0.5 µM of each primer (Sigma-Aldrich), 200 µM of dNTPs (Sigma-Aldrich), 0.5 U *Taq* polymerase (Sigma-Aldrich) and 1X PCR buffer (Sigma-Aldrich). Thirty five (35) cycles, each consisting of 1 min denaturation at 95°C, 2 min at annealing temperature (optimized separately for each primer pair, generally Tm-5°C) and 1 min polymerization at 72°C, were performed in a thermocycler (Bio-Rad, USA). The PCR products were separated by electrophoresis in a horizontal gel system at 100 V for 4 h in a 4% metaphor agarose gel. A 100 bp ladder (Thermo Scientific) was used for size determination of amplified products. Polymorphism was visualized by staining the gel with ethidium bromide, and it was photographed with the gel documentation system (Bio-Rad, USA).

ISSR amplification

One hundred ISSR primers were used for initial screening, out of which 73 primers gave informative banding patterns with good reproducibility. The selected 73 primers were 15-20-mers which included 54.7% di-nucleotide repeat motif, 31.5% tri-nucleotide repeat motif, 8.21 % tetra-nucleotide repeat motif and 5.47% penta-nucleotide repeat motif (Table S1). These were anchored at 5' end or 3' end by zero nucleotides or by one to three partially degenerated selective nucleotides.

PCR amplification was performed in a volume of 20 µl containing 2 µl of DNA (50 ng/µl), 0.4 µM of each primer (Sigma-Aldrich), 200 µM of dNTPs (Sigma-Aldrich), 0.5 U Taq polymerase (Sigma-

Table 1. The selected genotypes of *G. arboreum*-race 'bengalense'

Accession	Source	S/N	Accession	Source
CISA-6-187	Sirsa	34	DLSA-1006	Karnataka
CISA-6-123	Sirsa	35	LD-960	Punjab
CISA-6-209	Sirsa	36	LD-909	Punjab
CISA-6-214	Sirsa	37	FDK-124	Punjab
CISA-6-256	Sirsa	38	PAIG-8/1	Maharashtra
CISA-6-295	Sirsa	39	DAS-802	Karnataka
CISA-6-350	Sirsa	40	CCA-4	Tamilnadu
CISA-614	Sirsa	41	RAAS-931	Karnataka
CISA-6	Sirsa	42	GBaV-105	Gujarat
CISA-7	Sirsa	43	GBaV-120	Gujarat
CISA-8	Sirsa	44	ARBHA-0853	Karnataka
CISA-9	Sirsa	45	ARBAS-104	Karnataka
CISA-10	Sirsa	46	RAAS-36	Karnataka
CISA-294	Sirsa	47	RAAS-8	Karnataka
CISA-64	Sirsa	48	GAM-158	Gujarat
CISA-310	Sirsa	49	AKA-0106	Maharashtra
LD-327	Punjab	50	CINA-369	Maharashtra
LD-733	Punjab	51	CAN-1006	Maharashtra
ARBAS-105	Karnataka	52	HD-485	Hisar
TKA-9102/03	Tamilnadu	53	GAM-150	Gujarat
MDL-2617	Karnataka	54	JTAPTI-007	Madhya pradesh
GBaV-107	Gujarat	55	CCA-8	Tamilnadu
PA-532	Maharashtra	56	LD-694	Punjab
PA-686	Maharashtra	57	RG-8	Rajasthan
RG-526	Rajasthan	58	HD-123	Hisar
RG-540	Rajasthan	59	PA-255	Maharashtra
RG-541	Rajasthan	60	LD-987	Punjab
RG-514	Rajasthan	61	RG-579	Rajasthan
FDK-118	Punjab	62	LD-919	Punjab
TKA-9102	Tamilnadu	63	LD-936	Punjab
KWP-902	Madhya pradesh	64	LD-1010	Punjab
DLSA-17	Karnataka	65	RG-595	Rajasthan
DLSA-1005	Karnataka			

* All the 65 genotypes were collected from C.I.C.R, Regional station, Sirsa (Haryana), India, which in turn procured from respective source in India mentioned above.

Table 2. Different SSR and ISSR primers used for present study.

Marker type	Names
SSR (The sequence information of these SSRs is available at http://www.cottonmarker.org.)	**1**.BNL-530, **2**.BNL-686, **3**.BNL-852, **4**.BNL-1694, **5**.BNL-1679, **6**.BNL-3408, **7**.BNL-256, **8**.BNL-1030, **9**.BNL-834, **10**.BNL1317, **11**.BNL-1414, **12**.BNL-252, **13**.BNL-1053, **14**.BNL-3279, **15**.BNL-3649, **16**.BNL-2960, **17**.BNL-1707b, **18**.BNL-448, **19**.BNL-1434, **20**.BNL-3140, **21**.BNL-1231, **22**.MGHES-13, **23**.MGHES-14, **24**.MGHES-32, **25**.MGHES-58, **26**.MGHES-70, **27**.MGHES-7, **28**.MGHES-52, **29**.MGHES-50, **30**.MGHES-46, **31**.MGHES-45, **32**.NAU-2035, **33**.NAU-1047, **34**.NAU-1231, **35**.NAU-1068, **36**.NAU-3675, **37**.NAU-3519, **38**.NAU-2317, **39**.NAU-3008, **40**.NAU-1218, **41**.NAU-980, **42**.NAU-2083, **43**.NAU-862, **44**.NAU-3418, **45**.NAU-923, **46**.NAU-1233, **47**.NAU-3260, **48**.MUSS-563, **49**.MUSS422, **50**.MUSS-257, **51**.MUSS-300, **52**.MUSS-321, **53**.MUSS-88, **54**.MUSS-20, **55**.MUSS-49, **56**.MUSS-121, **57**.MUSS-439, **58**.JESPR-127, **59**.JESPR-65, **60**.JESPR-307, **61**.JESPR-297, **62**.CIR-070
ISSR	73 Primers (The sequence information along with their annealing temperature is given in Supplementary Table 1)

Aldrich) and 1X PCR buffer (Sigma-Aldrich). After a pre-denaturation step of 5 min at 95°C, amplification reactions were cycled forty times at 95°C for 1 min, at the annealing temperature (optimized separately for each primer pair, generally Tm-5°C) for 2 min and polymerization at 72°C for 1 min in a thermocycler (Bio-Rad, USA). The PCR products were visualized by running on 2% agarose gel, followed by staining with ethidium bromide. Finally, the gel was photographed as above.

Data analysis

Evaluating efficiency of different primers within each marker systems for diversity analysis

Within each marker system, the efficiency of each assay unit (that is, primer) was studied by: a) the number of scorable bands (NSB); b) the number of polymorphic bands (NPB); c) polymorphism information content (PIC); d) marker index (MI); e) the number of patterns (Tp); and, f) discrimination power (D). The formulas used for the above calculations are as follows:

The number of scorable bands (NSB) represents the average number of DNA fragments amplified/detected per genotype using a marker system. Of these, some loci (fragments or bands) may be polymorphic (NPB).
PIC for SSR markers was calculated according to Anderson et al. (1993). For ISSR markers, PIC of a band (PIC*i*) was calculated as follow: $\text{PIC}_i = 1 - \sum_j f_{ij}^{\,2}$, where *fij* is the frequency of the jth

pattern of the ith band (note that dominant markers have two patterns for a band as being present and absent). Then PIC of each ISSR primer was calculated as: $\text{PIC} = 1/n \sum_{i=1}^{n} \text{PIC}_i$, where

n is NPB for that primer.
The utility of a given marker system is a balance between the level of polymorphism detected and the extent to which an assay can identify multiple polymorphisms. Marker index is the product of PIC and effective multiplex ratio (EMR) (Powell et al., 1996). EMR is estimated as: EMR= NSB X ß, where ß is the fraction of polymorphic markers and is estimated after considering the polymorphic loci (np) and non-polymorphic loci (nnp) as ß = np / (np + nnp). Tp and D were calculated according to Tessier et al. (1999).

Comparison of two marker systems for diversity analysis

To compare the discriminating capacity of the SSR and ISSR markers, the following statistical calculations were performed manually according to Belaj et al. (2003): a) the number of assay units (U); b) the number of polymorphic bands (n$_p$); c) the number of monomorphic bands (n$_{np}$); d) the average number of polymorphic bands/assay unit (n$_p$/U); e) the number of Loci (L); f) number of loci/assay unit (n$_u$); g) the number of banding patterns (T$_p$); h) the average number of patterns/assay unit (I); i) average confusion probability (C); j) average discriminating power (D); and k), the average limit of discriminating power (D$_L$).
Several other genetic diversity parameters viz. effective number of allele (N$_e$), Shannons index (I) and expected heterozygosity (H$_e$) were determined using GenAlex 6.5.

Cluster analysis

For this analysis, each amplified band was treated in terms of binary code, based on the presence (1) and absence (0) of bands. To analyze data obtained from binary matrices, the NTSYS-pc ver

2.2 statistical package (Rohlf, 2000) was used. Three data sets were utilized, viz. SSR, ISSR and combined datasets of SSR and ISSR. The binary qualitative data matrices were then used to construct similarity matrices based on Jaccard similarity coefficients (Jaccard, 1908). The similarity matrices were then used to construct a dendrogram using the unweighted pair group method with arithmetic average (UPGMA). To compare SSR and ISSR based dendrograms, cophenetic matrices were derived from dendrograms using COPH (cophenetic values) program, and the goodness-of-fit of the clustering to the 2 data matrices was calculated by comparing the original similarity matrices with the cophenetic value matrices using the Mantel matrix correspondence test (Mantel, 1967) in the MXCOMP program. Similarly, a dendrogram was also constructed for combined dataset of SSR and ISSR markers.

RESULTS

SSR analysis

The 62 primers detected a total of 170 alleles (all polymorphic), and the number of alleles per locus varied from 2 to 6 with an average of 2.7 alleles per locus, in all 65 genotypes examined. The size of the alleles ranged from a minimum of 90 bp (loci JESPR-297) to maximum of 720 bp (loci MGHES-14). The PIC values ranged from 0.030 (MUSS-439, NAU-923, NAU-3675, BNL-1434, BNL-1694) to 0.809 (NAU-3008), giving an average of 0.38. Of the 62 SSR loci, 21 loci yielded a PIC value of ≥ 0.5 (Table 3) and produced quite distinct bands in the metaphor gels (Figure S1). These included 5 BNLs, 3 MGHES, 6 NAUs and 6 MUSS and 1 JESPR SSR loci, which were considered as highly informative markers (Table 3). Among these 21 loci, 52.1% had di-nucleotide motifs, 30.4% had tri-nucleotide motifs, while the remaining 17.5% had tetra/penta/hexa-nucleotide motifs.

Marker index (MI), considered to be an overall measure of the efficiency to detect polymorphism, was obtained in the range of 0.06-4.85 (average 1.20). The 21 informative primers, designated so on the basis of high PIC, also exhibited a high marker index value (more than 1.5). Primer NAU-3008 yielded highest MI value (4.85), which was obvious because it had the highest PIC and EMR.
The discriminating power (D) of a primer depends on the number of fragments it generates as well as the frequency of the banding patterns. In the present study, the maximum value of discrimination power (D) observed was 0.927 (NAU-3008) while the lowest was 0.44 (NAU-3675, BNL 1434, BNL 1694, NAU 923 and MUSS 439); with an overall mean value of 0.442 for the 62 SSR loci. The 21 above-mentioned informative primers pairs also exhibited high discrimination power (values of D more than 0.6) and thus these 21 primers were categorized as highly informative and discriminative primers (Table 3).

ISSR analysis

Across the 65 genotypes, the 73 ISSR primer pairs yielded a total of 281 reproducible bands, of which 265

Table 3. Description of 21 selected SSR markers for all the studied genotypes of *G. arboretum.*

Primer name	NSB	NPB	PIC	MI	Tp	D
BNL-834	3	3	0.51	1.53	4	0.66
BNL-252	3	3	0.55	1.65	5	0.77
BNL-2960	4	4	0.57	2.31	4	0.78
BNL-448	3	3	0.58	1.74	3	0.67
BNL-3140	4	4	0.61	2.45	4	0.7
MGHES-50	4	4	0.68	2.74	4	0.74
MGHES-46	3	3	0.59	1.77	3	0.69
MGHES-45	3	3	0.57	1.73	3	0.69
NAU-3519	3	3	0.51	1.55	3	0.60
NAU-2317	3	3	0.57	1.73	3	0.68
NAU-3008	6	6	0.80	4.85	9	0.92
NAU-1218	3	3	0.58	1.75	3	0.67
NAU-3418	4	4	0.64	2.56	6	0.82
NAU-3260	4	4	0.70	2.8	4	0.75
MUSS-257	3	3	0.57	1.71	3	0.66
MUSS-300	3	3	0.58	1.74	3	0.67
MUSS-321	3	3	0.55	1.67	3	0.65
MUSS-20	4	4	0.69	2.77	4	0.77
MUSS-49	3	3	0.57	1.73	3	0.72
MUSS-121	4	4	0.70	2.8	4	0.76
JESPR-65	4	4	0.70	2.8	4	0.79

NSB, Number of scorable bands; NPB, number of polymorphic bands (NPB); PIC, polymorphic information content; MI, marker index; Tp, number of banding patterns; D, discriminative ability.

(94.3%) were polymorphic. The number of loci (or bands) scored varied from 2 (ISSR-17, 19, 38, 47, 49, 59, 62, 69, 74, 81, 82, 84, 87, 96 and 103) to 15 (ISSR-18), with an average of 3.84 bands/loci per primer. The PIC values for 73 ISSR primer pairs ranged from 0.3 to 0.5, with an average of 0.38 per primer; and the highest PIC (0.5) was obtained for ISSR-82. Marker index (MI), calculated for each primer pair, was found in the range of 0.16 (ISSR-17, 19, 49, 81) to 5.28 (ISSR-18). The highest value of discrimination power was observed for primer ISSR-18 (0.998), while primer ISSR-17, 19, 49 and 81 yielded the lowest values of D (0.316).

On the basis of higher values of MI (more than 1), D (more than 0.8) and PIC (more than 0.3), 53 primer pairs were identified as very efficient for the present genetic diversity analysis (Table 4). Further, in addition to these 53 primer pairs, 7 more primer pairs viz. ISSR-15, 16, 59, 80, 87, 96 and 103 exhibited higher values of D (more than 0.8), though MI values were considerably low for some.

Comparison of marker systems

Performance of the two marker systems was compared based on two main aspects: The discriminating capacity (that is, efficiency of discrimination) between any two genotypes at random from the studied genotypes; and, the overall efficiency in detecting polymorphisms in all the studied genotypes.

Overall, SSR markers were more polymorphic (100% polymorphic bands) than ISSR (94.3% polymorphic bands), however, the number of polymorphic bands per assay unit was higher in ISSR (3.63) as compared to SSR (2.7). SSR markers are locus specific so only 1 loci was analyzed per assay, and 62 loci overall. ISSR primer pairs produced 281 bands, with each band considered as one locus, resulting in an average of 3.84 loci per assay unit. ISSR produced a higher number of banding patterns (658) than SSR (175) and so the average number of banding pattern per assay unit was also higher for ISSR (9.01) than for SSR (2.8).

The number of effective alleles (Na) in all 65 genotypes examined was higher in SSR (2.112) than in the ISSR assay (1.397), while the average discriminating capacity (D) was distinctly higher for ISSR (0.862), compared to SSR (0.442) (Figure 1). The average limits of discriminating powers (D_L) for both the markers were found to be very close to the actual value of the discriminating powers (D) of both.

A higher value for the Shannon index (I) was obtained for SSR (0.740), ISSR yielding a comparatively low value of I (0.421) (Figure 1). The average expected heterozygosity (He) values calculated for SSR and ISSR came

Table 4. Description of 53 ISSR markers for all the studied genotypes of *G. arboretum*.

Name	NSB	NPB	PIC	MI	Tp	D	Name	NSB	NPB	PIC	MI	Tp	D
ISSR1	3	3	0.45	1.35	7	0.89	ISSR-70	3	3	0.43	1.29	7	0.94
ISSR-2	4	4	0.38	1.52	12	0.96	ISSR-71	4	4	0.40	1.6	12	0.96
ISSR-3	4	4	0.31	1.24	10	0.94	ISSR-72	4	4	0.40	1.6	12	0.96
ISSR-4	3	3	0.34	1.02	6	0.92	ISSR-73	4	4	0.37	1.48	9	0.95
ISSR-5	7	7	0.32	2.24	17	0.90	ISSR-75	4	4	0.35	1.4	11	0.89
ISSR-6	3	3	0.43	1.29	7	0.93	ISSR-76	3	3	0.41	1.23	7	0.94
ISSR-7	4	4	0.33	1.32	11	0.95	ISSR-77	4	4	0.41	1.64	11	0.95
ISSR-10	4	4	0.42	1.68	14	0.97	ISSR-78	3	3	0.45	1.35	7	0.94
ISSR-18	15	13	0.41	5.28	33	0.99	ISSR-83	4	4	0.43	1.72	13	0.96
ISSR-27	5	5	0.30	1.5	14	0.98	ISSR-85	3	3	0.42	1.26	7	0.92
ISSR-28	3	3	0.38	1.14	5	0.94	ISSR-86	4	4	0.42	1.68	7	0.92
ISSR-29	8	8	0.37	2.96	26	0.97	ISSR-88	4	4	0.41	1.64	10	0.93
ISSR-31	12	10	0.38	3.78	32	0.95	ISSR-89	3	3	0.42	1.26	7	0.93
ISSR-34	5	5	0.31	1.55	13	0.96	ISSR-90	4	4	0.43	1.72	7	0.92
ISSR-35	9	9	0.30	2.7	27	0.87	ISSR-91	3	3	0.42	1.26	6	0.91
ISSR-36	4	4	0.39	1.56	12	0.93	ISSR-92	4	4	0.42	1.68	6	0.92
ISSR-40	9	8	0.32	2.53	24	0.93	ISSR-93	5	5	0.42	2.10	6	0.91
ISSR-45	3	3	0.34	1.02	7	0.94	ISSR-94	3	3	0.43	1.29	6	0.92
ISSR-50	4	4	0.30	1.2	12	0.98	ISSR-95	4	4	0.43	1.72	6	0.90
ISSR-60	4	4	0.34	1.36	12	0.97	ISSR-97	3	3	0.43	1.29	7	0.93
ISSR-61	4	4	0.41	1.64	13	0.95	ISSR-98	3	3	0.43	1.29	7	0.93
ISSR-63	4	4	0.37	1.48	11	0.97	ISSR-99	4	4	0.43	1.72	7	0.93
ISSR-64	4	4	0.39	1.56	10	0.92	ISSR-100	3	3	0.43	1.29	7	0.92
ISSR-65	4	4	0.30	1.2	10	0.96	ISSR-101	5	5	0.43	2.15	7	0.93
ISSR-66	3	3	0.40	1.2	7	0.94	ISSR-102	3	3	0.43	1.29	7	0.93
ISSR-67	5	5	0.40	2.0	20	0.98	ISSR-104	3	3	0.43	1.29	7	0.93
ISSR-68	4	4	0.40	1.6	13	0.96							

NSB, Number of scorable bands; NPB, number of polymorphic bands (NPB); PIC, polymorphic information content; MI, marker index; Tp, number of banding patterns; D, discriminative ability.
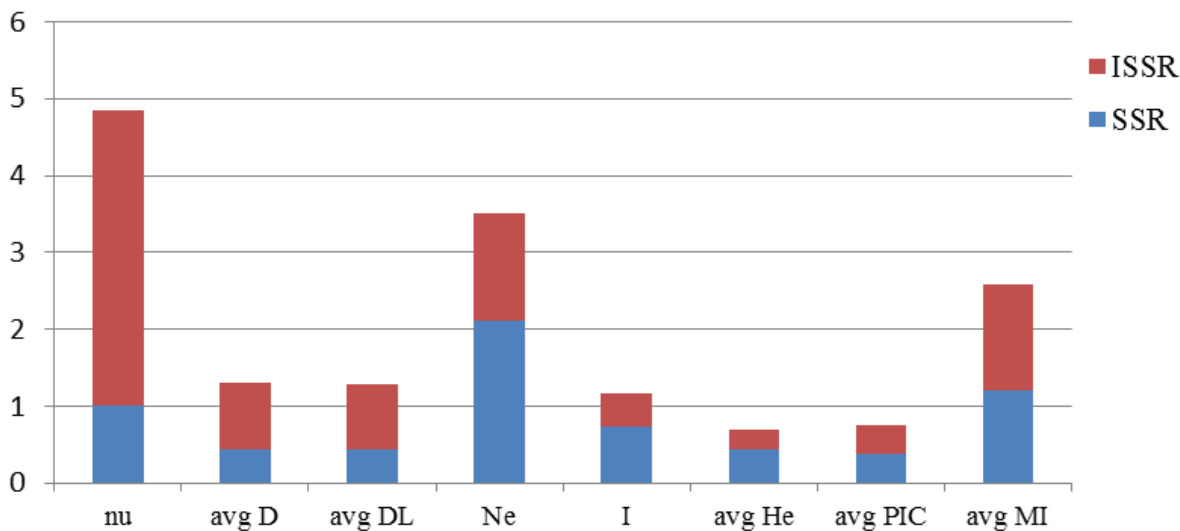

Figure 1. Comparative description of SSRs and ISSRs for parameters-number of loci/assay unit (nu), average discriminating power (avg D), average limit of discrimination power (avgDL), number of effective alleles (Ne), Shannon's index (I), average expected heterozygosity (avg He), average PIC (avg PIC), average MI (avg MI).
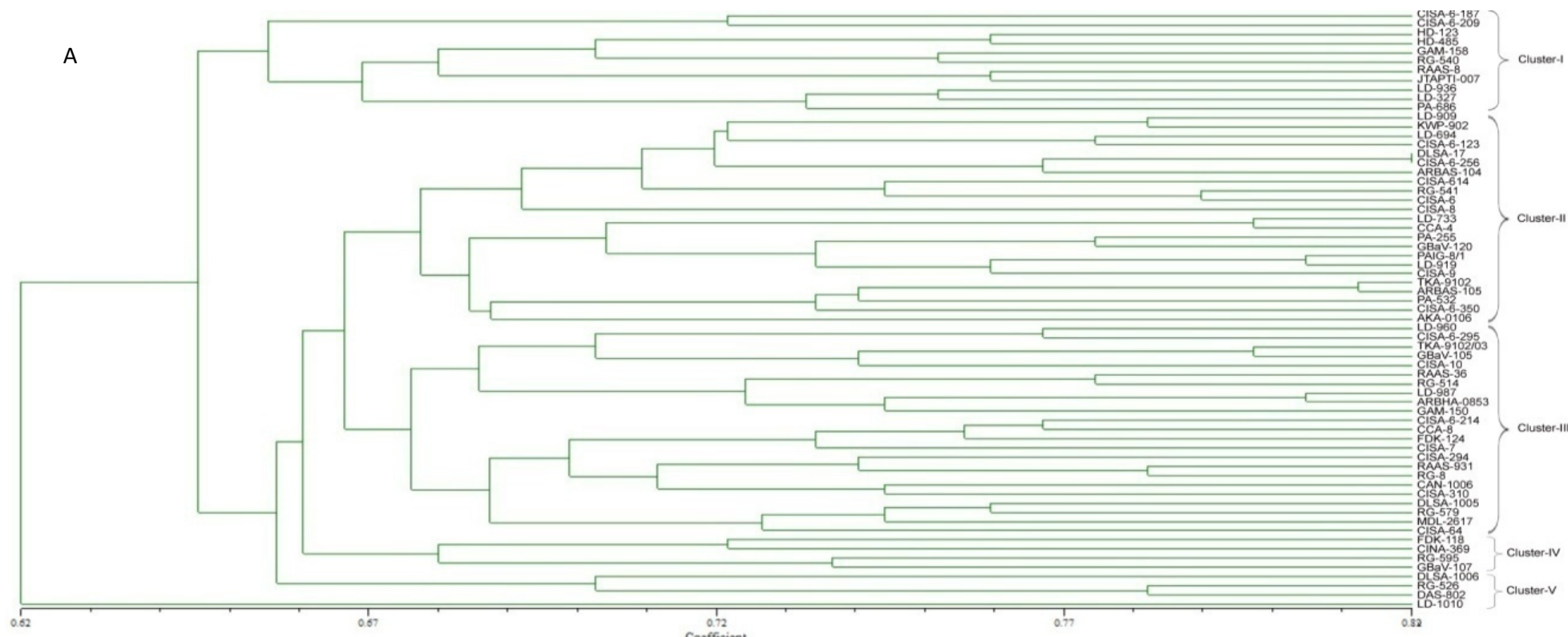
Figure 2. Dendrograms of 65 *G. arboreum* genotypes based on the similarity coefficient values calculated using: **(a)** SSRs, **(b)** ISSRs and **(c)** joined data set of markers.

out to be 0.433 and 0.264 respectively. The average PIC was found to be the same for both SSR and ISSR markers (0.38) in the studied genotypes, while the Average MI was slightly higher for ISSR compared to SSR markers.

Cluster analysis

A dendrogram obtained using the UPGMA method based on SSR, ISSR and SSR + ISSR data set (Figure 2) clearly distinguished all the

genotypes of the race 'bengalense' of *G. arboreum*. Genetic similarity coefficients were obtained in the range of 0.62-0.82 for the SSR marker, 0.56-0.86 for ISSR markers and 0.59-0.80 for the combined data of SSR and ISSR markers. Five main clusters were formed in all three dendrograms. Each cluster consists of a different number of genotypes with different genetic similarity coefficients.

In dendrograms based on SSR, ISSR and SSR + ISSR, the first cluster consists of 11, 14 and 9 genotypes, respectively, in which CISA-6- 187

have been found to be more distant than the other genotypes in all three dendrograms. The second cluster consists of 23, 21 and 21 genotypes, respectively, showing almost similar groupings of genotypes but with some differences in the similarity coefficient between different genotypes. For example, with SSR markers, DLSA-17 and CISA-6-256 exhibited a maximum similarity coefficient value of 0.82, while with ISSR the maximum value (0.86) was for CISA-6 and CISA-8. For combined datasets, a maximum similarity coefficient (0.785) within cluster 2 was obtained
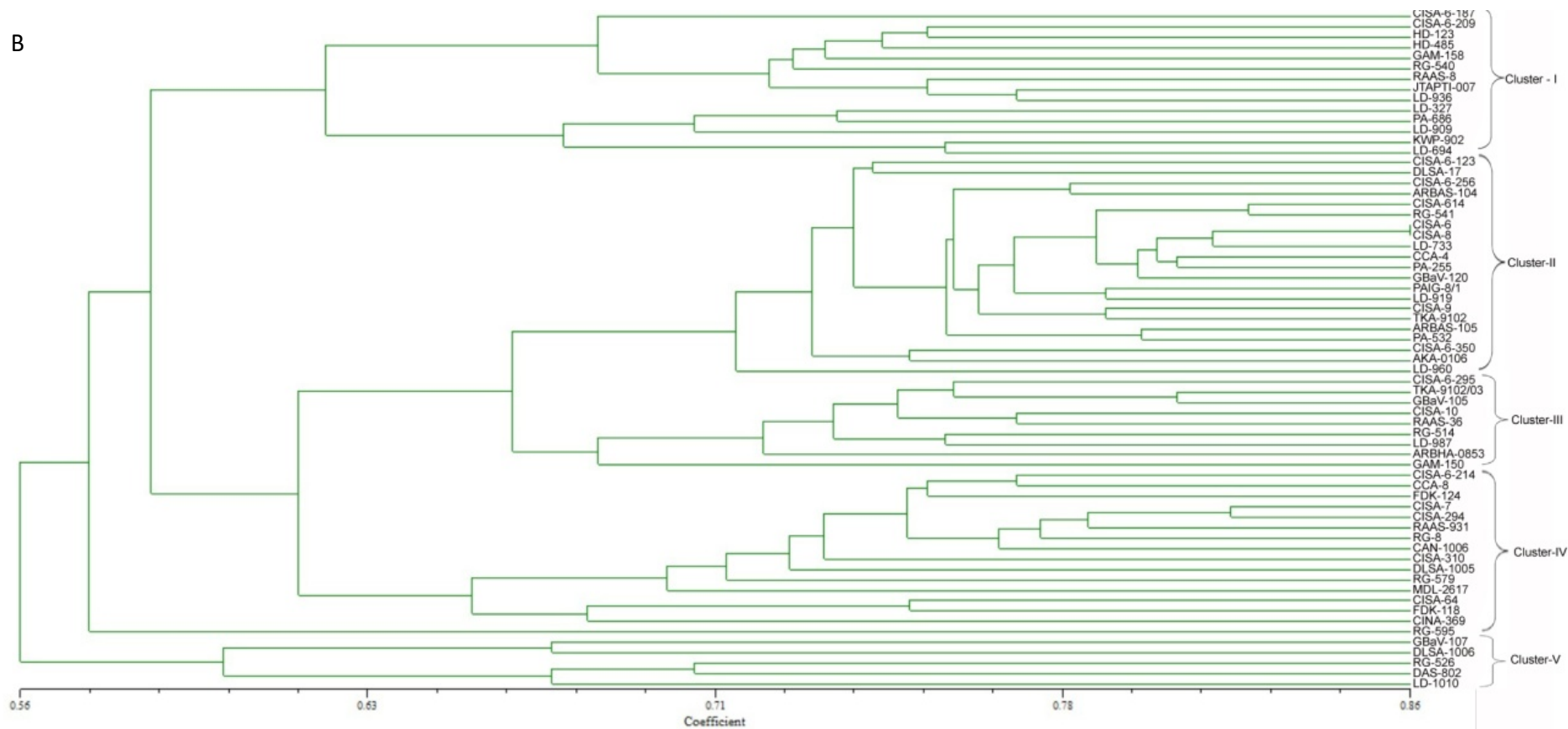
B

Figure 2. Contd.

for CISA-614 and RG-541. The third cluster of SSR-based dendrograms consisted of 23 genotypes, while in dendrograms based on ISSR and combined data, the third cluster consisted of nine genotypes each. Similar observations were made for cluster four, which consisted of 4, 16 and 19 genotypes in SSR, ISSR and SSR+ISSR based dendrograms, respectively; CISA-7 and CISA-294 were found to be closer than the rest of

the genotypes in the case of ISSR and SSR + ISSR dendrograms; but, in the case of SSR, these two genotypes were present in cluster 3. Cluster five consisted of almost similar number of genotypes, that is, 4, 5 and 7, in the three dendrograms formed. In this cluster, LD-1010 was found to be more distant than the rest of genotypes in all three of the dendrograms obtained.

Cophenetic correlation coefficients for individual techniques based on genetic similarity value matrices were obtained using the Mantel matrix correspondence test. High correlation coefficient values were obtained for ISSR markers (r = 0.94), for combined data set (SSR + ISSR) marker (r = 0.91) and for SSR markers (r = 0.87). All three dendrograms showed almost similar groupings with some differences in the genetic similarity

C

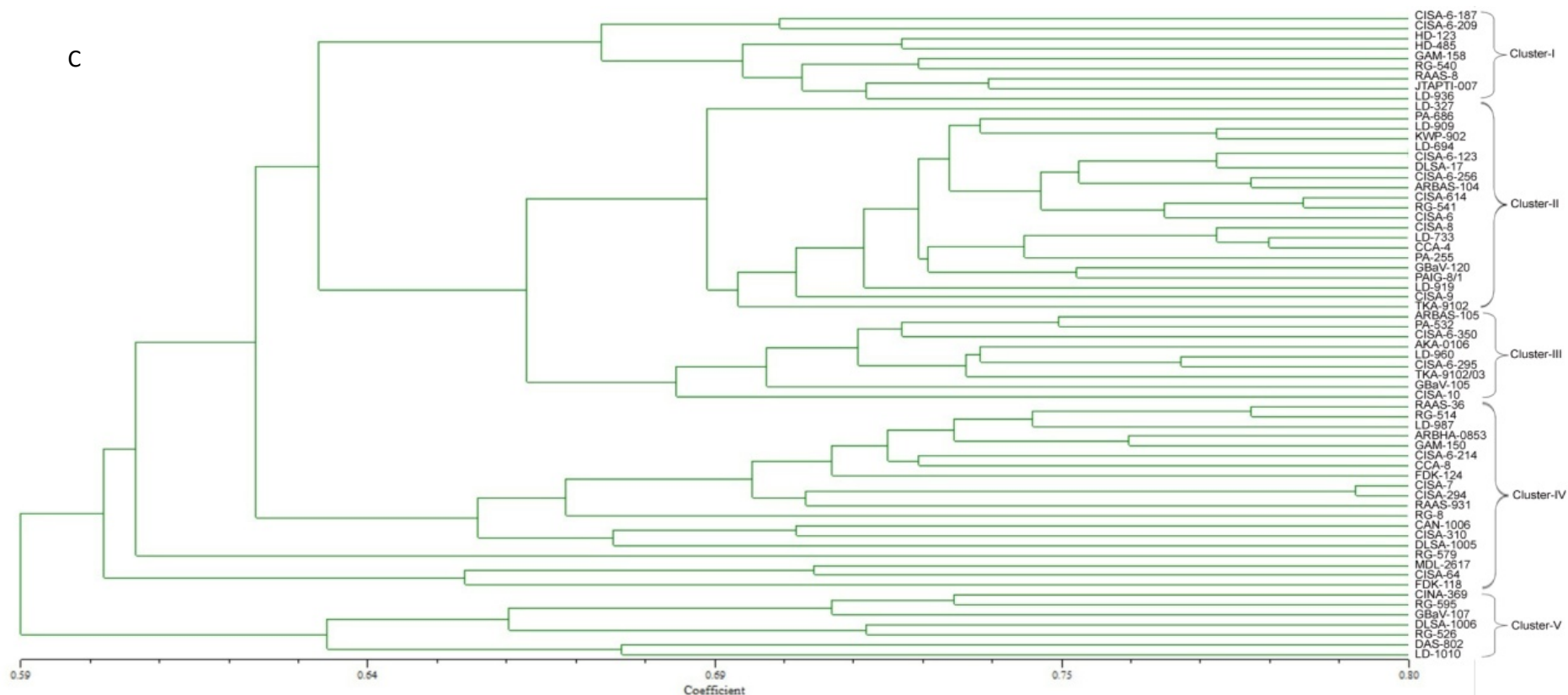Figure 2. Contd.

coefficient, as discussed above.

DISCUSSION

During the past few decades, molecular markers have been commonly used for assessing genetic diversity, which is the basis for the genetic improvement of any given species. The important criteria of selecting the right molecular marker depends on the specific application, presumed level of polymorphism, presence of sufficient technical facilities, time constraints and financial limitations (Kumar et al., 2009). Sometimes the combined use of two or more markers for the study of genetic diversity has been found to be better than respective individual markers (Anna Serra et al., 2007). In the past, a variety of molecular markers like RAPD, ISSR and SSR have been used for estimating the genetic diversity in *G. arboreum* (Dongre et al., 2011; Bardak and Bolek, 2012; Noormohammadi et al., 2013a). SSR are locus specific, co-dominant markers, and are considered ideal for fingerprinting; while ISSR are multi-locus, dominant markers, and have been found very efficient for diversity analysis. Therefore, the present study documents the comparative utility of these maker types for genetic diversity studies in accessions belonging to *G. arboreum* race bengalense.

Marker polymorphism

Both SSR and ISSR markers were found to reveal a similar level of polymorphism as revealed by the same average value of PIC (0.38) obtained for each. The average PIC value for SSR markers obtained during the present study was less than that obtained by Kantartzi et al. (2009) (average PIC 0.42) while genotyping various *G. arboreum* genotypes with SSR markers, though their highest PIC obtained was less (0.75) than that obtained during the present study (0.809). ISSRs are dominant markers and therefore a maximum PIC value of 0.50 can be expected for a given ISSR loci. During the present investigation, for one marker ISSR-82, this threshold was reached while values very close to the threshold were obtained for ISSR-1, ISSR-62, ISSR-69, and ISSR-78 (0.45, 0.45, 0.47 and 0.45, respectively). A PIC range of 0.00 to 0.5 with an average of 0.321 was also obtained previously in another study using ISSR markers for some tetraploid cotton (Noormohammadi et al., 2013b).

In addition to PIC, certain other parameters such as MI and D have been documented as very useful for evaluating the efficiency of molecular markers (Belaj et al., 2003; Myskow et al., 2010). The utility of any given marker is found in a balance between the level of polymorphism it can detect and its capacity to identify multiple polymorphisms (Powell et al., 1996). The MI is considered to be an overall measure of the efficiency of a marker to detect polymorphism, and is related to EMR value. Discriminating power is considered as a good estimator of the efficiency of a primer or locus. It depends not only on the number of patterns generated, but also on their relative frequency (Tessier et al., 1999). On the basis of these factors, a core set of 21 SSR primers (Table 3) were identified as highly informative markers with high PIC, very good discriminative power and MI. Likewise, 54 ISSR primer pairs could be identified on the basis of higher MI values. Multi locus marker systems like ISSR are expected to produce higher EMR and MI than single locus SSRs (Belaj et al., 2003). Markers with higher EMR and MI values are better for analysis of both interspecific and intraspecific genetic diversity (Singh et al., 2014). Several studies report such identification of a core set of highly polymorphic and discriminative markers to be very helpful for varietal identification and genetic diversity assessment (Masi et al., 2003; Jain et al., 2004; Kantartzi et al., 2009).

Comparative utility of marker system

The selection of a particular type of molecular marker is important and critically depends on the intended use (Gupta et al., 2002). The discriminative abilities of both marker systems were compared using certain selected parameters which have also been used earlier for such purposes in some studies (Mukherjee et al., 2013). The presence of rare bands/alleles can produce low frequency of patterns and result in lower D values. ISSR markers exhibited considerably higher number of banding patterns, more polymorphic bands/assays and higher discriminative powers compared to SSR during the present investigation. The similar edge of ISSR over SSR in terms of discriminative capability for a given set of genotypes has also been observed in certain other studies (Singh et al., 2014).

SSR markers are locus specific, multi-allelic and co-dominant in nature. These have been found to detect higher levels of polymorphism and so, generally, are the markers of choice in plant genetics and breeding (Kantartzi et al., 2009). ISSR are bi-allelic (hence supposed to be less informative) and are locus un-specific, but are more randomly distributed throughout genome than SSR (Kumar et al., 2009). This abundance of ISSR sometimes compensates for their bi-allelic nature and may make them very informative for a given germplasm (Vijayan, 2005). Further, the low development and running cost makes ISSR more suitable than SSR (Vijayan, 2005).

During the present study, SSR markers outperformed the ISSR in terms of Ne, I and He parameters. Ne represents the number of equally frequent alleles it would take to achieve a given level of gene diversity. The Shannon index (I) is a diversity index that is used to characterize species diversity and is an indicator of both the abundance and evenness of the species present. The reason for high heterozygosity in case of SSR markers is due to its co-dominant nature, which permits the detection of a high number of alleles per locus as these are multi-allelic as compared to ISSR markers, which are biallelic in nature (Belaj et al., 2003).

During the present study, the average PIC value for SSR was on the lower side (0.38) as SSR, being co-dominant, yielded PIC values in the range of 0 to 1.0. On the other hand, ISSR markers yielded a higher value of average PIC (0.38), while for dominant markers the range is 0 to 0.5. Further, ISSR also showed better utility in detecting multiple polymorphisms as revealed by high MI and high EMR (Table 4).

Phylogenetic relationships in examined germplasm

The present study has reported that both SSR and ISSR techniques, along with proper statistical tools, could be successfully applied to assess genetic diversity and perform phylogenetic analysis in *G. arboreum*. Although SSR and ISSR markers showed differences in detecting polymorphism and discriminating capacity, they showed similar groupings in dendrograms on the basis of similarity matrices. A high significant correlation coefficient was obtained for all the three dendrograms. The correlation coefficient between genetic similarity values depends not only on the kind of molecular technique and species examined, but also upon the range of discovered diversity. Noormohammadi et al. (2013b) and Sheidai et al. (2012) reported higher values (r=0.87-95) by using

different molecular techniques in cotton. High r values and identical topologies of dendrograms suggest that each method of molecular marker development, used independently, could be a reliable source of information about the relationships between analyzed germplasm (Myskow et al., 2010). In our study, ISSR and SSR+ISSR markers depicted better topology and high correlation coefficient than SSR markers.

In conclusion, although the average PIC is the same for both markers, there are certain parameters in which SSR exceeds like Ne, I and He, and in the rest of the parameters - MI, EMR, and D - ISSR was found better than SSR. So, a combination of both markers would be highly efficient in detecting genetic diversity and phylogenetic analysis between genotypes of race 'bengalense' of *G. arboreum*.

Conflict of interests

The authors have not declared any conflict of interest.

Abbreviations

ISSR, Interspersed sequence repeats; **SSR,** simple sequence repeats; **PIC,** polymorphic information content.

REFERENCES

Akhtar KP, Haidar S, Khan MKR, Ahmad M, Sarwar N (2010). Evaluation of *Gossypium* species for resistance to cotton leaf curl Burewela virus. Ann. Appl. Biol. 157:135-147.

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

Anna Serra I, Procaccini G, Carmela Intrieri M, Migliaccio M, Mazzuca S, Maria Innocenti A (2007). Comparison of ISSR and SSR markers for analysis of genetic diversity in sea grass (*Posidonia aceanica*). Mar. Ecol. Prog. Ser. 338:71-79.

Bardak A, Bolek Y (2012). Genetic diversity of diploid and tetraploid cottons by SSR and ISSR markers. Turk. J. Field Crops 17:139-144.

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I (2003). Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theor. Appl. Genet. 107:736-744.

Deosarkar VD, Gupta DB, Gaikwad DB (2010). Genetic diversity studies in intra-specific *desi* cotton (*G. arboreum*) through DNA marker. J. Cotton Res. Dev. 24:133-137.

Dhawan AK, Simwat GS, Sidhu AS (1991). Field reaction of some varieties of Asiatic Cotton (*Gossypium arboreum* L.) to sucking and boll worm pests. J. Res. Punjab Agric. Univ. 28: 57-62.

Dongre AB, Ramteke DM, Bhandarkar MR, Raut MP, Meshram KJ (2011). Genetic diversity analysis of *Gossypium arboreum* (diploid cotton) cultivars revealed by PCR based molecular markers. J. Cotton Res. Dev. 25:1-8.

Gupta PK, Varshney RK, Prasad M (2002). Molecular markers: principles and methodology. In: SM Jain, BS Ahloowalia, DS Brar, Eds. Molecular techniques in crop improvement. Kluwer Academic Publishers, Netherlands, pp. 9-54.

Jaccard P (1908). Nouvelles recherche sur la distribution florale. Bulletin de la societe vaudoise des sciences naturelles 44:223-270.

Jain S, Jain RK, McCouch S (2004). Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluoroscently labeled microsatellite markers. Theor. Appl. Genet.

109:965-977.

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

Kulkarni VN, Khadi BM, Manjula S, Lalitadas M, Deshapande A, Narayanan SS (2009). The Worldwide Gene Pools of *Gossypium arboreum* L. and *G. herbaceum* L. and their Improvement, In: AH Paterson ed, Genet. Genomics Cotton. Vol 3, pp. 69-100.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009). Potential of molecular markers in Plant Biotechnology. Plant Omics J. 2:141-162.

Mantel N (1967). The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209-220.

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

Mohan P, Mukewar PM, Singh VV, Singh P, Khadi BM, Amudha J, Deshpande VG (2006). Identification of sources of resistance to grey mildew disease (*Ramularia areola*) in diploid cotton (*Gossypium arboreum*). CICR Technical Bulletin No: 34, pp. 1-26.

Mukherjee A, Sikdar B, Ghosh B, Banerjee A, Ghosh E, Bhattacharya M, Roy SC (2013). RAPD and ISSR analysis of some economically important species, varieties, and cultivars of the genus *Allium* (Alliaceae). Turk J. Bot. 37:605-618.

Myskow B, Milczarski and Masoje P (2010). Comparison of RAPD, ISSR and SSR markers in assessing genetic diversity among rye (*Secale cereale* L.) inbred lines. Plant Breed. Seed Sci. 62:107-115.

Nibouche S, Brevault T, Klassou C, Dessauw D, Hau B (2008). Assessment of the resistance of cotton germplasm (*Gossypium* spp.) to aphids (Homoptera, Aphididae) and leafhoppers (Homoptera: Cicadellidae, Typhlocybinae): methodology and genetic variability. Plant Breed. 127:376-382.

Noormohammadi Z, Farahani YHA, Sheidai M, Baraki SG, Alishah O (2013b). Genetic diversity analysis in opal cotton hybrids based on SSR, ISSR and RAPD markers. Genet. Mol. Res. 12:256-269.

Noormohammadi Z, Taghavi E, Foroutan M, Sheidai M, Alishah O (2013a). Structure analysis of genetic diversity in tetraploid and diploid cotton genotypes. Int. J. Plant Animal Environ. Sci. 3:79-90.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey J, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2:225-238.

Rohlf FJ (2000). NTSYS-pc*:* Numerical Taxonomy System. Ver. 2.1. Exeter Software, Setauket, NY, USA. pp. 29-34.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. U.S.A 81:8014-8018.

Sheidai M, Noormohammadi Z, Shojaei-Jesvaghani F, Baraki SG, Farahani F, Alishah O (2012). SSR (Simple sequence repeat) and ISSR (Inter simple sequence repeat) analysis of genetic diversity in tissue regenerated plants of cotton. Afr. J. Biotechnol. 11(56):11894-11900.

Singh A, Dikshit HK, Jain N, Singh D, Yadav RN (2014). Efficiency of SSR, ISSR and RAPD markers in molecular characterization of mungbean and other vigna species. Ind. J. Biotechnol. 13:81-88.

Tahir MS, Khan NUI, Sajid Ur R (2011). Development of an interspecific hybrid (Triploid) by crossing *Gossypium hirsutum* and *G. arboreum*. Cytologia 76:193-199.

Tessier C, David J, This P, Boursiquot JM, Charrier A (1999). Optimisation of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. 98:171-177.

Ulloa M, Brubaker C, Chee P (2007). Cotton. In: Kole C (ed) Genome mapping & molecular breeding. vol 6: technical crops. Springer, New York, pp.1-49.

Vijayan K (2005). Inter simple sequence repeat (ISSR) polymorphism and its application in mulberry genome analysis. Int. J. Ind. Entomol. 10:79-86.

Wheeler TA, Gannaway JR, Keating K (1999). Identification of resistance to *Thielaviopsis basicola* in diploid cotton. Plant Dis. 83:831-833.

Table S1. List of 73 ISSR primers used in the present study, with sequence information and annealing temperature.

| S/N	ISSR name	Sequence	Annealing Temp. (°C)	S/N	ISSR name	Sequence	Annealing Temp. (°C)
1	ISSR-1	$(AGC)_5GA$	52	38	ISSR-69	$(AC)_8TA$	45
2	ISSR-2	$(AGC)_5GG$	55	39	ISSR-70	$(GT)_8TA$	49
3	ISSR-3	$(AGC)_5GT$	52	40	ISSR-71	$(GT)_8CG$	50
4	ISSR-4	$(AGC)_5GC$	55	41	ISSR-72	$(CAG)_5$	50
5	ISSR-5	$(CA)_7AT$	43	42	ISSR-73	$(CAA)_5$	55
6	ISSR-6	$(CA)_7AC$	44	43	ISSR-74	$(GATA)_4$	55
7	ISSR-7	$(CA)_7GT$	45	44	ISSR-75	$(TGGA)_4$	43
8	ISSR-10	$(CA)_7AA$	42	45	ISSR-76	$(CA)_7AG$	45
9	ISSR-15	$(GT)_7AT$	42	46	ISSR-77	$(ACTG)_4$	47
10	ISSR-16	$(GT)_7AC$	45	47	ISSR-78	$(GA)_8CG$	55
11	ISSR-17	$(GA)_9T$	49	48	ISSR-79	$CCAG(GT)_7$	50
12	ISSR-18	$(GA)_9A$	49	49	ISSR-80	$(GACAC)_4$	49
13	ISSR-19	$(GA)_9C$	51	50	ISSR-81	$(TG)_8TT$	49
14	ISSR-27	$(CA)_9GT$	52	51	ISSR-82	$(TGT)_5$	52
15	ISSR-28	$CAG(GA)_7$	47	52	ISSR-83	$(AGC)_5$	55
16	ISSR-29	$GCT(GA)_7$	47	53	ISSR-84	$(GAA)_5$	52
17	ISSR-31	$T(AG)_7$	45	54	ISSR-85	$GT(CAC)_7$	55
18	ISSR-34	$G(CA)_7$	42	55	ISSR-86	$CT(CAC)_7$	43
19	ISSR-35	$C(CA)_7$	42	56	ISSR-87	$CAG(CT)_8$	44
20	ISSR-36	$A(CA)_7$	45	57	ISSR-88	$CGT(CA)_8$	45
21	ISSR-38	$A(CT)_8$	45	58	ISSR-89	$AGG(CA)_8$	42
22	ISSR-40	$C(CT)_8$	47	59	ISSR-90	$(CAC)_5GT$	42
23	ISSR-45	$(TG)_7C$	42	60	ISSR-91	$(CAC)_5CT$	45
24	ISSR-47	$(GACA)_4$	45	61	ISSR-92	$(CAG)_5AT$	49
25	ISSR-49	$T(GA)_8$	47	62	ISSR-93	$(CAG)_5GT$	49
26	ISSR-50	$C(GA)_8$	47	63	ISSR-94	$TC(GACA)_4$	51
27	ISSR-58	$(CTC)_6$	46	64	ISSR-95	$G(TGGGG)_5$	52
28	ISSR-59	$(GGGTG)_3$	45	65	ISSR-96	$C(CAG)_5$	47
29	ISSR-60	$AGT(AG)_7$	47	66	ISSR-97	$G(CAG)_5$	47
30	ISSR-61	$GCG(GA)_7$	52	67	ISSR-98	$GT(GACA)_4$	45
31	ISSR-62	$AAG(GT)_7$	55	68	ISSR-99	$(GCTTC)_3$	42
32	ISSR-63	$CAC(TG)_7$	45	69	ISSR-100	$(AAG)_5$	42
33	ISSR-64	$AAG(CT)_7$	45	70	ISSR-101	$(AAG)_5GT$	47
34	ISSR-65	$(TC)_8A$	52	71	ISSR-102	$(AAG)_5GC$	45
35	ISSR-66	$(TC)_8G$	47	72	ISSR-103	$T(AAG)_5$	55
36	ISSR-67	$(AC)_8C$	54	73	ISSR-104	$G(AAG)_5$	55
37	ISSR-68	$(AC)_8CT$	55				
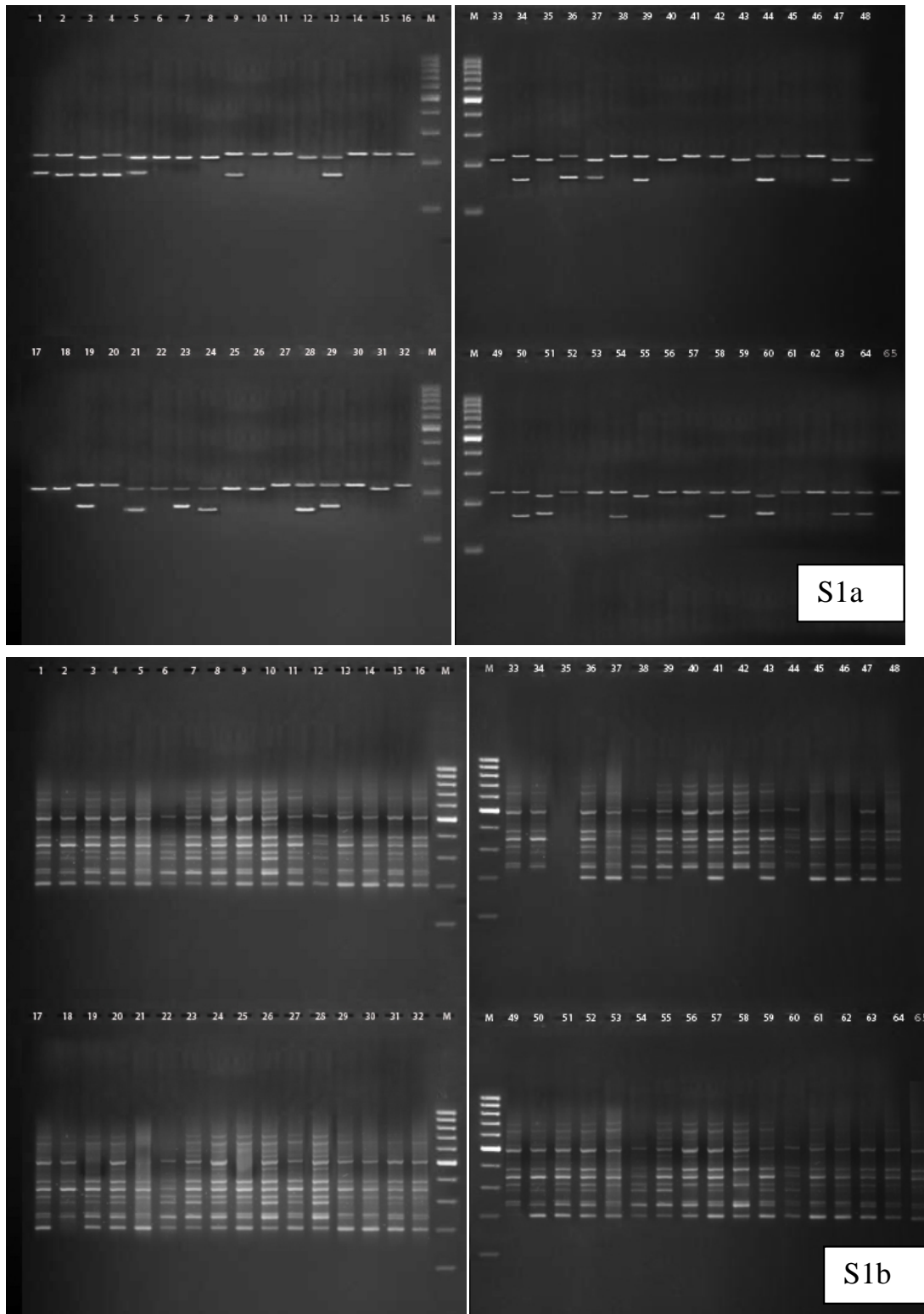
Figure S1. a) SSR profile with primer BNL2960, b) ISSR profile with primer ISSR 40, of the selected 65 genotypes (numbers are as per Table 1).