J. Cotton Res. Dev. **28** (2) 175-184 (July, 2014)

Unlocking the genetic potential for improvement of economic traits in large collection of cotton (*Gossypium* spp) germplasm

R. RAVIKESAVAN, S. VENKATESAN, S. SATHISH, P. KAVITHA AND N. M. BOOPATHI*
Center for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore – 641 003
**E-mail : nmboopathi@tnau.ac.in*

ABSTRACT : Totally 142 cotton germplasm accessions, including 91 *G. barbadense* and 51 *G. hirsutum* representing diverse origin were evaluated for morphological and molecular level. The data were used to estimate mean performance and heritability to choose parents for breeding programme based on *per se* performance and genetic divergence. All the investigated traits showed significant differences among the genotypes. D^2 analysis revealed that the *G. barbadense* and *G. hirsutum* accessions were grouped into 32 and 6 clusters, respectively indicating the existence of sufficient genetic diversity. Potential genotypes in *G. barbadense* and *G. hirsutum* were identified as parents based on high mean performance. Out of the 30 SSRs used in this study, the dendrogram classified *G. barbadense* accessions into 11 major clusters and *G. hirsutum* accessions into 7 clusters. The genotypes 85/2 and BARBADOS and ST-7A-OKRA and HH 6 had the least similarity percentage in *G. barbadense* and *G. hirsutum,* respectively, when compared with other genotypes and these could be useful for hybridization programme for transferring the desirable traits and to get more heterotic F_1s and better segregants in F_2 generation besides dissection of quantitative trait loci linked to traits of economic importance in cotton.

Key words : Genetic diversity, germplasm characterization, morphological markers, SSRs

Evolving cotton varieties or hybrids for economic traits such as patronizing quality fibres and increased production with sufficient degree of regional adaptability has assumed paramount importance because of the increasing internal demand and to encourage exports. To formulate such breeding strategy, it required a greater understanding of the genetic basis and variation of quantitative characters that decided fibre yield and quality.

Genetic diversity analysis played an important role, because hybrids between lines of high genetic diversity can display a great heterosis than those between closely related parents. Hence, for exploitation of heterosis and enlarging the variability in the subsequent segregating generations, an understanding of genetic diversity among parents is absolutely essential. Further, hybridization involving genetically diverse parents is known to provide an opportunity for pooling genes together to yield desirable transgressive segregants in advanced generation.

Genetic diversity is conventionally assessed by morphological traits. There are about 145 morphological markers identified in cultivated cotton and successfully applied in genetic diversity analysis during earlier days. However, such traits are affected by environment, phenology or developmental stages of the plant and the type of the plant material besides its difficulty in accumulating multiple markers in a single genotype. Further it requires testing in several replications and environments to establish the genetic contributions, which are highly influenced by several factors. Quicker methods of evaluating the width of genepool minimizing environmental influences are represented by the use of molecular markers. Thus the efficiency can be improved, if both morphological and molecular markers are employed in genetic diversity analysis (Boopathi *et al.*, 2011).

Among the molecular markers, simple sequence repeats (SSR) could be an ideal means to sample the genetic diversity and relationship of cotton resources at the genomic level (Saha *et al.*, 2003). Genetic diversity analysis using SSRs has been used as a successful tool in genotyping of the many plant species because of their reproducibility, multiallelic nature, co dominant inheritance, relative abundance and good genome coverage over other markers such as RFLPs, RAPDs or AFLPs. Most genetic studies in cotton, both pedigree and marker based, have found a narrow genetic base in upland cotton (Boopathi *et al.*, 2011). Hence, documenting the existing genetic diversity in the cotton germplasm

used in routine regional breeding programme is assuming paramount importance. In this context, an investigation was carried out with the following objective: To characterize *G. hirsutum* and *G. barbadense* germplasm lines using morphological, fibre quality traits, and SSR markers and identify superior parents with wide genetic base for the forthcoming molecular breeding programme.

MATERIALS AND METHODS

Plant materials : The experiments were conducted at the Cotton Breeding Station, Tamil Nadu Agricultural University, Coimbatore. One hundred and forty two cotton germplasm accessions, including 91 accessions belonging to *G. barbadense* and 51 to *G. hirsutum* of different geographical locations were evaluated under field conditions.

The seeds of each accession were raised in a randomized block design with two replications in 4.5 m rows. A spacing of 90 cm between rows and 45 cm between plants was adopted. All recommended agronomic practices and plant protection measures were followed. Five competitive plants in each accession in each replication were randomly selected and labeled for recording observations. The average values of the observations from these 5 plants represented the mean of that replication. Data were recorded on different yield paramenters such as days to first flowering, plant height, sympodia and bolls/plant, boll weight (g/tex), ginning outturn (%), lint index (g/100 seeds), seed index (g/100 seeds) and single plant yield (g). Besides, fibre quality characters such as 2.5 per cent span length (mm), fibre fineness (micronaire), bundle strength (g/tex) and uniformity ratio were also tested using in High Volume Instrument (Uster Model: HVI classic 900).

Statistical analysis : Statistical analysis was carried out using the Windostat statistical package. Heritability in broad sense was calculated and expressed in percentage. The data on yield and yield components obtained for 142 accessions were utilized for estimating the genetic diversity using D^2 analysis. Variance in respect of characters and the covariance between character pairs were calculated. A dendrogram was also constructed to depict the diversity of genotypes within and outside the clusters.

Genetic diversity using SSR markers : Attempts were also made to study the genetic diversity and phylogenetic relationship using SSR markers in 87 G. *barbadense* and 50 *G. hirsutum* genotypes. Totally 30 microsatellite primers that span the cotton genome were used for DNA amplification and the amplified products were resolved in 3 per cent MetaPhor agarose gel electrophoresis and the SSR bands were scored for their presence and absence. A dendrogram was constructed based on Jaccard's similarity coefficient with unweighted pair group method and arithmetic average (UPGMA) using the NTSYS-pc version 2.02 (Exeter Software, Newyork, USA). To measure the informativeness of the markers, the polymorphism information content for each SSR was calculated according to the formula: PIC = 1-Ópi^2, where pi is the frequency of the i^{th} allele.

RESULTS AND DISCUSSION

Mean performance and heritability : Genetic studies in cultivated cotton revealed very low level of genetic diversity (Lacape *et al.,* 2007). Pedigree analysis in cotton indicated a wide genetic base, but it conflicted with other type of distance estimates that revealed a narrow genetic base. Genetic variability analysis was carried out on 142 accessions at morphological and molecular level. Based on the *per se* performance, the accessions B 135, K5309-9155E, SIV 135-18, BarxLT, EC97631, EC111248, Sudan G-5-5, TADIA, SB-1085 and TCB 377 in *G. barbadense* were found to have high mean value for single plant yield, boll weight, bolls/plant, 2.5 per cent span length and bundle strength. On the other hand *G. hirsutum* accessions EC 95308, PEE DEE 01B, 6144, EMPIRE GLANDLES, SA72-62 and AR 9 were identified to have higher mean value for single plant yield, boll weight, bolls/plant, 2.5 per cent span length and bundle strength. Hence these accessions were recommended for selection and further utilization in crop improvement programme.

The heritability estimates ranged from 41.53 (days to first flowering) to 92.90 per cent

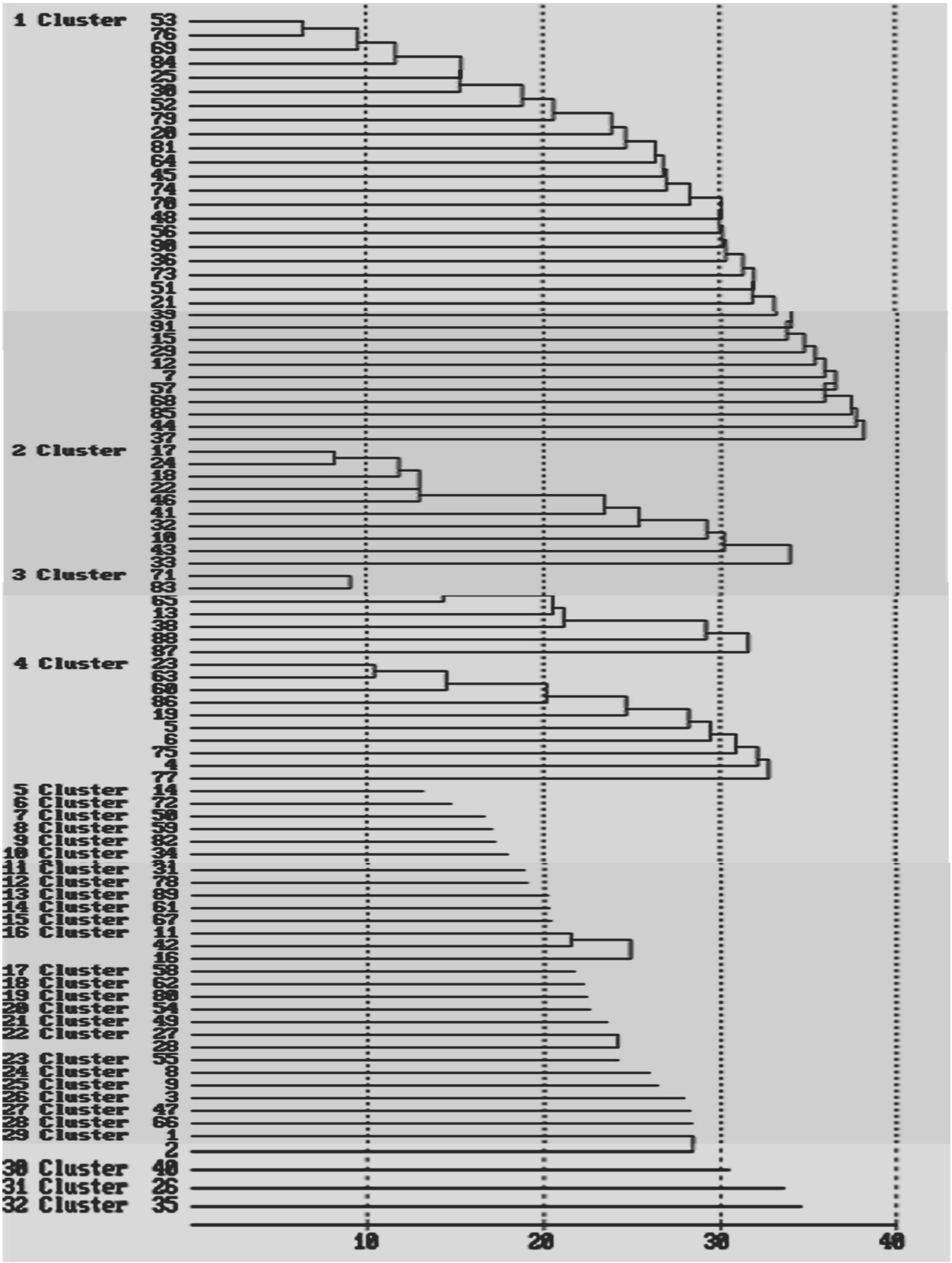

Fig. 1. Cluster analysis of *G. barbadense* accessions using morphological markers

(boll weight) in *G. barbadense*. In *G. hirsutum,* the heritability estimates ranged from 7.10 (2.5 % span length) to 92.49 per cent (single plant yield). High heritability coupled with low genetic advance was observed for ginning outturn and 2.5 per cent span length in both *G. barbadense* and *G. hirsutum*. This suggested that these characters were under the control of non additive gene action and heterosis breeding can be resorted for improving these characters.

Genetic diversity analysis using morphological markers : D^2 statistic helps in the selection of genetically divergent parents for their exploitation in the hybridization programme. For exploiting heterosis as a mean for increasing production, it is necessary to have parents with maximum genetic diversity. The more diverse the parents, there were more chances of pronounced heterotic effects and increased spectrum of variability in the segregating generations.

Ninety one accessions of *G. barbadense* of diverse origin were grouped into 32 clusters based on the morphological data (Fig. 1). It was observed that cluster I had the largest genotypes (32) followed by cluster II and IV with 10 genotypes in each cluster. Cluster III possessed with 7 genotypes, cluster XVI with 3 genotypes and clusters XXII and XXX had 2 genotypes and all the remaining 25 clusters were contained only one genotype each. The accessions in each cluster were genetically closer and the geographic origin was found to be different except in cluster XXII (had homogeneous accessions of same origin). Cluster I, II, III, IV, XVI and XXX were heterogeneous, consisting of genotypes from various origins.

In *G. hirsutum,* 51 accessions were classified into 6 clusters based on the morphological traits investigated in the study (Fig. 2). It was observed that cluster I was the largest with 39 accessions followed by cluster II and III with 6 and 3 genotypes, respectively. Remaining 3 clusters were grouped with one genotype each. The genotypes in each cluster were genetically similar. All clusters were heterogenous in their origin. The genotypes of same origin were arranged in different clusters. This further indicated that genetic diversity need not necessarily be related to geographical diversity.

Genetic diversity was the outcome of several factors including geographical diversity, genetic drift, natural selection forces and diverse environmental conditions within a country caused more diversity than geographical isolation. Due to domestication, heterogeneity, selection pressure in a given environment, considerable variability would have been created in these accessions. Therefore, the selection of parents for hybridization programme should be based on genetic diversity in relation to the character rather geographic diversity.

The results of this study indicated that the inter cluster distances were greater than intra cluster distances in both species, revealing considerable amount of genetic diversity among the accessions studied. Intra cluster distance was maximum for cluster I (Fig. 1). The relative divergence of each cluster from other cluster (inter cluster distance) indicated high order of divergence between cluster XVIII and XXIX (14.86) which included Egyptian 1 and PIMA 2 followed by cluster XVI and XXIX (14.59) which included 32/2R, SIA 7, 5746 U and PIMA 2. Whereas in *G. hirsutum* intra cluster distance was maximum for cluster I (8.03) followed by cluster II (6.87) and cluster III (6.62) (Fig. 2). The inter cluster distance was high between V and VI (18.44) which included B 61-1862 and EC 95308 followed by cluster IV and VI (16.70) which included TASHKENT 3 and EC 95308. Hybridization among the accessions between these clusters would produce successful heterotic hybrids and desirable segregants with high yield coupled with desirable fibre quality parameters.

The cluster mean values for different characters can also be considered to generate the highest possible variability in the yield components. Cluster XXV, XXX and X had high mean values for following characters *viz.,* plant height and sympodia/plant and showed low mean value for days to first flowering. The genotypes in these clusters can be used to improve those characters so as to obtain plants which would be early to flower and more sympodia/plant. For yield components *viz.,* single plant yield, ginning outturn, seed index, bolls/plant and boll weight, the clusters XIX, XX, XXXI, XVI and XVII had ideal cluster mean values. Hence, the accessions belonging to these clusters can be used to improve single plant yield. Cluster VI, VII, XIX

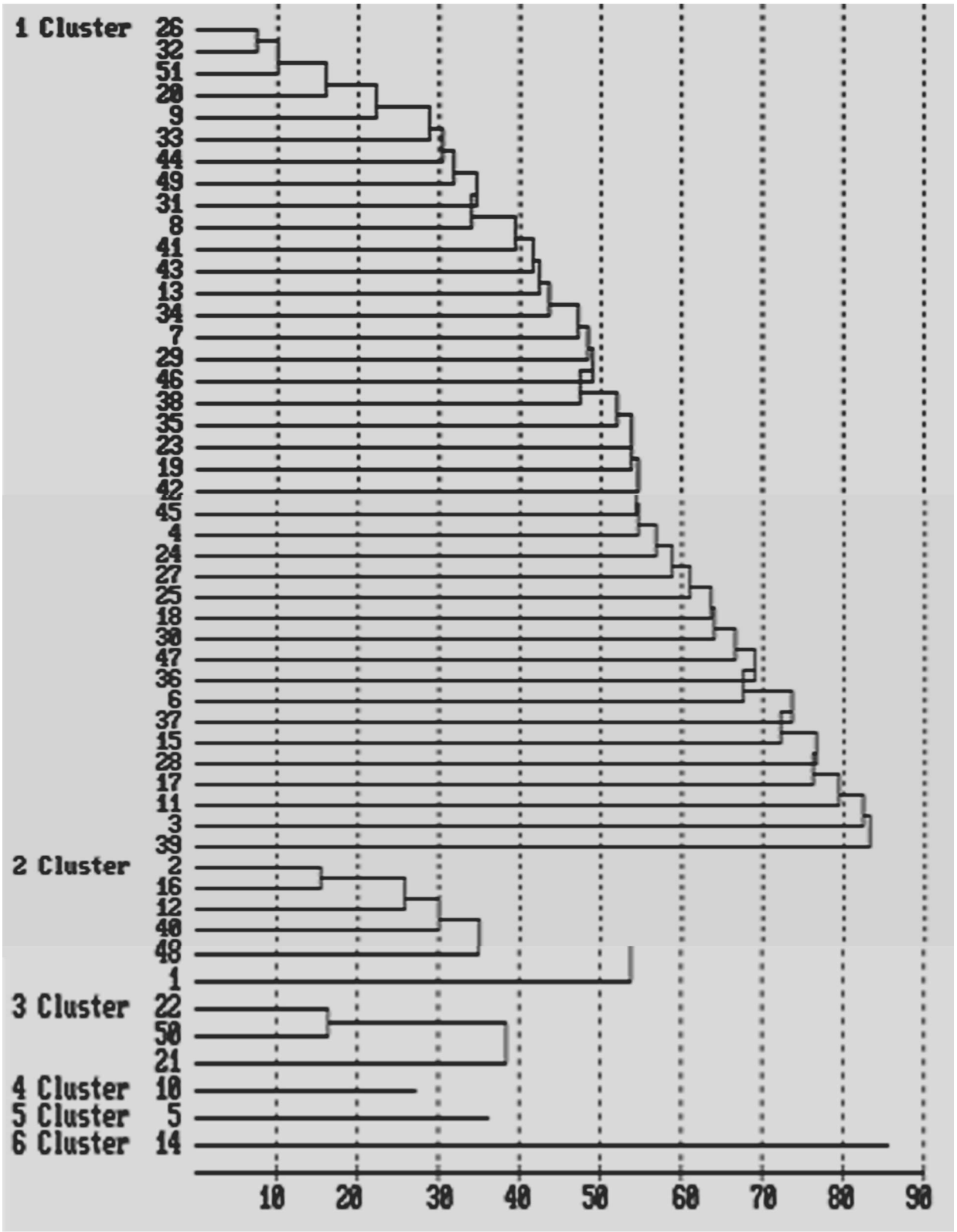

Fig. 2. Cluster analysis of *G. hirsutum* accessions using morphological markers

and XXVI showed highest mean value for quality traits namely uniformity ratio, micronaire, 2.5 per cent span length and bundle strength. The lines under these clusters can be used to improve fibre quality traits in *G. barbadense.* In *G. hirsutum*, cluster VI had lowest mean value for early flowering, boll weight and micronaire and highest value for sympodia/plant, lint index, ginning outturn and bundle strength. Cluster II had lowest mean values for plant height, single plant yield, lint index and ginning outturn. Cluster III had highest value for bolls/plant, single plant yield, uniformity ratio and micronaire value and it had lowest value for 2.5 per cent span length. Cluster IV, V, had highest mean value for plant height, boll weight, 2.5 per cent span length and seed index. Hence it is desirable to the breeders to choose the parents from these clusters based on his breeding objectives.

Genetic diversity analysis using molecular markers : The level of genetic diversity detected by molecular markers confers to them a high degree of precision and reliability in the framework of genetic studies, with a particular emphasis on the management of germplasm collection. Hence, 15 SSRs that generated 40 alleles were used to estimate the genetic diversity among 87 accessions of *G. barbadense;* whereas in case of *G. hirsutum* 13 SSRs were used, which generated 29 alleles among 50 accessions. The number of alleles revealed by each marker ranged from 2 to 4 with an average of 2.7 and 1 to 4 with an average of 2.4 in *G. barbadense* and *G. hirsutum,* respectively.

PIC revealed the amount of information that can be obtained from a particular primer. The PIC of an SSR marker provided an estimate of the discriminatory power of that SSR marker by taking into account not only the number of alleles that are detected but also the relative frequencies of those alleles (Smith *et al.*, 2000). In the present study, PIC value was highest for BNL3269 (0.65) and lowest for JESPR 127 (0.07) with a mean of 0.4 in *G. barbadense* and in *G. hirsutum* PIC value was highest for BNL 226, BNL 1047 and BNL 1672 (0.5) and low for BNL 3147 and JESPR 127(0.0) with the mean of 0.3. This is in close agreement with the results of Liu *et al.* (2000), Lacape *et al.*, (2007) and Candida *et al.*, (2006). Hence, the above said SSRs with higer PIC value have great significance in germplasm characterization.

The mean number of alleles per SSR locus was 2.7 and 2.4 for *G. barbadense* and *G. hirsutum*, respectively. This is similar to the findings of Wu *et al.*, (2007). However, it was lower than that reported by Liu *et al.*, (2000) and Lacape *et al.*, (2007) where the mean/locus was 5. This might be due to the fact that the lines used in the study came from breeding programmes and might therefore have a narrow genetic base (Candida *et al.*, 2006).

The highest similarity of 100 per cent was observed in the genotypes TCB 472/5 and TCB 472/4, 6002-1 and SUVIN, SUDAN G 5-3 and 16/2R, SBYF and B 125; while 76/3 and 19/61 had 97.7 per cent, 13/2 and SUDAN G 5-3, 16/2R and 13/2, 1085-6 and 24/2W, GIZA 1467 and EGYPTIAN 1, 85/2 and 17/3A had 97.8 per cent in *G. barbadense.* In *G. hirsutum,* the genotypes PEE DEE 9223 and EL 592, AHA 1-9-38 and K 3902, T 11 and K 3902 and AHA 1-9-38 and T 11 had 93 per cent similarity. Hence, intercrossing cannot yield good segregants between these genotypes as they have high similarity coefficient and low genetic diversity. Similar result was obtained by Hirut *et al.*, (2007). It was also stated that many of the accessions had common parents in their genetic backgrounds. This common parentage among the cultivars and lines may also be a reason for poor gene diversity.

The DNA banding pattern was scored for all the 137 cotton accessions. A data matrix was generated and analysed using SAHN clustering methods of the NTSYS-pc programme. A high level of polymorphism (100%) was observed among the 87 accessions of *G. barbadense* and the mean number of alleles/locus was 2.7. In case of *G. hirsutum* the level of polymorphism was only about 50 per cent which represented low level of polymorphism and the mean number of alleles/locus was 2.4. This is similar to the findings of Lacape *et al.*, (2007) and Candida *et al.*, (2006), who have mentioned that low level of polymorphism was common in *G. hirsutum.*

Based on the dendrogram, *G. barbadense* accessions have been grouped into 11 clusters (Fig. 3). These 11 clusters were identified at the 0.79 similarity coefficient whereas in case of *G. hirsutum* it has been grouped into 7 clusters (Fig. 4). These clusters were identified at the 0.74

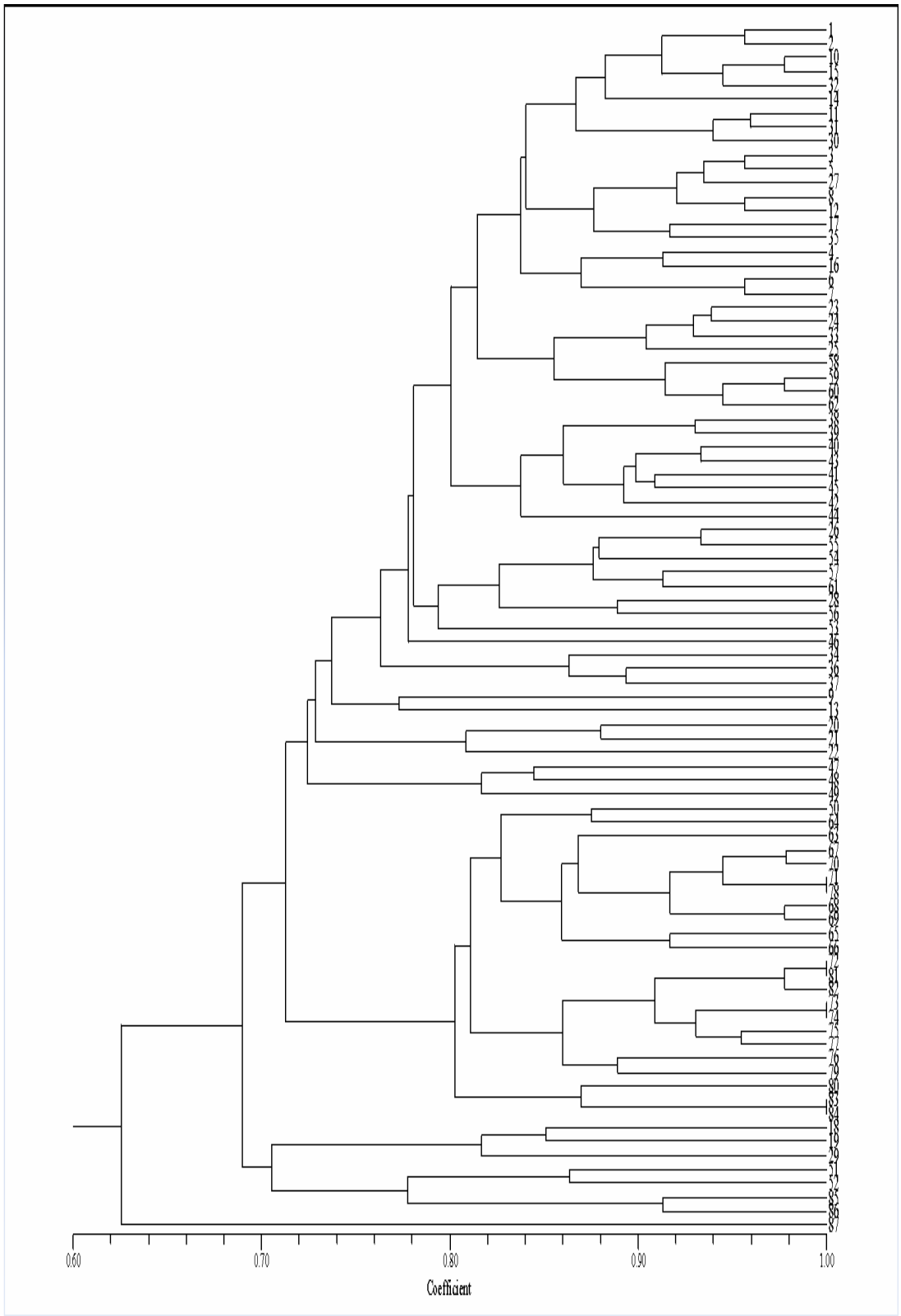

Fig. 3. Cluster analysis using SSR marker data among *G. barbadense* accessions

similarity coefficients. In *G. barbadense,* Cluster I comprised maximum number of genotypes (36) followed by cluster VIII (23). Cluster II had eight genotypes, cluster IV, cluster VI, cluster VII, cluster IX and cluster XI had 3 genotypes each. Cluster V and cluster X had 2 genotypes each and cluster III had one genotype. In *G. hirsutum,* cluster II comprised maximum number of genotypes (21) followed by cluster I (17). Cluster VI had 4 genotypes, cluster III had 3 genotypes and cluster V and VII had 2 genotypes each and Cluster IV had 1 genotype.

The formation of 11 clusters and 7 clusters in *G. barbadense* and *G. hirsutum,* respectively through hierarchial cluster analysis of SSR data revealed that the presence of genetic diversity at molecular level was high among the accessions and the SSR primers used in this study sorted the superior genotypes. Hence by using SSR technique a large set of informative data could be generated in less time than with morphological traits.

Comparative analysis of clustering of accessions based on phenotypic and molecular data : To determine the extent of diversity in *G. barbadense* and *G. hirsutum*, diversity analysis both at molecular and morphological level were done. A set of 13 morphometric traits were used to group varieties based on D^2 analysis. In this approach the varieties were grouped into 32 clusters and 6 clusters in *G. barbadense* and *G. hirsutum,* respectively, while it was 11 and 7 cluster in *G. barbadense* and *G. hirsutum,* respectively through molecular analysis. Some of the accessions *viz.,* 32/2R, 5746 U and SIA 7 in *G. barbadense* and HH 6, T 23, AP 18-2-1 and AR 8 in *G. hirsutum* fall under same cluster both in morphological and molecular analysis. Some of the lines *viz.,* B 125, SUDAN G 4-5 and Suvin in *G. barbadense* had high diversity as well as higher value for *per se* performances. So these accessions can be selected for the further breeding programmeme.

However molecular analysis revealed that the accessions 6002-1 and SUVIN, TCB 472/4 and TCB 472/5, SUDAN G 5-3 and 16/2R, B 125 and SBYF in cluster VIII in *G. barbadense* and the *G. hirsutum* genotypes *viz.,* EL 592 and PEE DEE 9223 in cluster I and K 3902, T 11 and AHA 1-9-38 in cluster II were focused as similar genotypes which have 100 per cent similarity. In contrast, cluster analysis using morphological data had placed these accessions in different clusters. The change in clustering pattern of accessions might be due to the presence of environmental impact on the plant, which does not precisely describe the genetic relationship that resulted in shuffling of some genotypes among the clusters. Morphological variation does not reliably reflect real genetic variation because of genotype environment interaction and the largely unknown genetic control of polygenically inherited morphological and agronomic traits. Most phenotypic traits, were controlled by several genes which may be highly influenced by the environment, and hence the molecular markers overcome this limitation as they have virtually no environment component and hence the grouping at the molecular level, based on the specific simple sequence repeat markers were highly reliable when compared to the phenotypic diversity. So the SSR could be an ideal mean for the identification of the genetic diversity and relationship of cotton resources at the genomic level.

The present study thus helped to select cotton accessions with wider genetic distances which could be used as parents for developing mapping populations to genetically dissect out quantitative trait loci linked to traits of agronomic importance in cotton besides their usefulness in regular breeding programme.

ACKNOWLEDGEMENT

This project was funded by Department of Biotechnology, Government of India, New Delhi.

REFERENCES

Boopathi, N. M., Thiyagu, K., Urbi, B., Santhoshkumar, M., Gopikrishnan, A., Aravind, S., Swapnashri, Gat and Ravikesavan, R. 2011. Marker assisted breeding as next generation strategy for genetic improvement of productivity and quality: Can it be realized in cotton? *Intl. J. Plant Genomics* 2011: Article ID 670104. doi:10.1155/2011/ 670104.

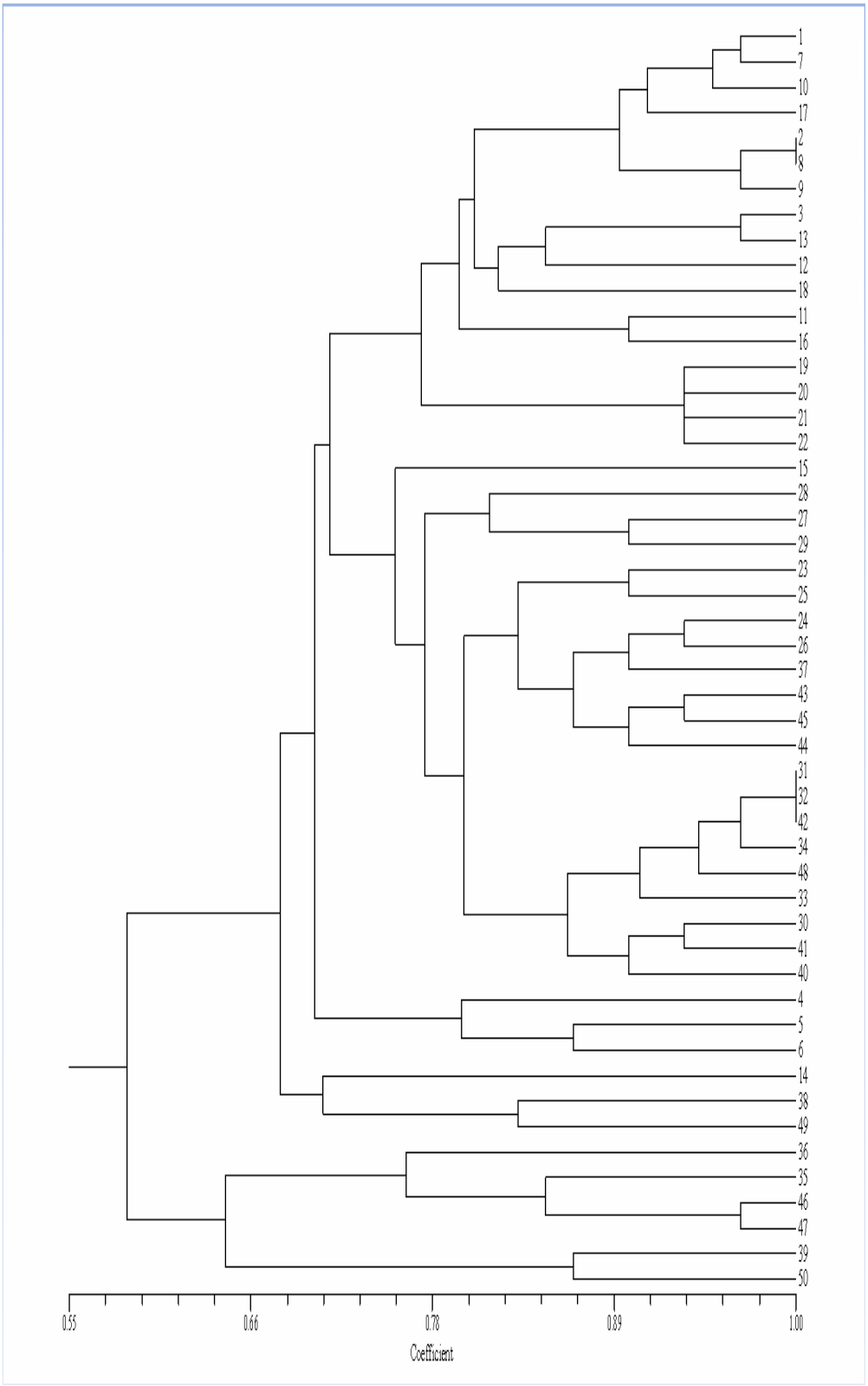

Fig. 4. Cluster analysis using SSR marker data among *G. hirsutum* accessions

Cândida, H.C., Bertini, de Magalhães, Schuster, Ivan, Sediyama, Tocio, Everaldo Gonçalves Barros, de and Moreira, Maurílio Alves 2006. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. *Genet. Mol. Bio.*, **29** : 321-29.

Hirut, K., Burow, G., Dani, Raviprakash G. and Allen, Randy D. 2007. A-genome cotton as a source of genetic variability for upland cotton (*Gossypium hirsutum*). *Genetic Res. Crop Evo.* **54**:885–95.

Lacape, J.M.A.E., Dessauw, D., Rajab, M., Noyer J. L. and Hau, A. E. B. 2007. Microsatellite diversity in tetraploid *Gossypium* germplasm: Assembling a highly informative genotyping set of cotton SSRs. *Mol. Breed.* **19** : 45–58

Liu, S., Cantrell, R. R., McCarty Jr., J. C. and Stewart, J. McD 2000. Simple sequance repeat based assesment of genetic diversity in cotton race stock accessions. *Crop Sci.,* **40** : 1459-69

Saha, S., Karaca, M., Jenkins, J.N., Zipf, A.E., Reddy, U.K. and Kantety, R.V. 2003. Simple sequence repeats as useful resources to study transcribed genes of cotton. *Euphytica*, **130**:355–64.

Smith, J.S.C., Kresovich, S., Hopkins, M.S., Mitchell, S.E., Dean, R.E., Woodman, W.L., Lee, M. and Porter, K. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Sci.,* **40** : 226-32.

Wu Y. X., Daud, M. K., Chen L., and Zhu, S.J. 2007. Phylogenetic diversity and relationship among *Gossypium* germplasm using SSR markers. *Pl. Syst. Evol.* **21** : 255-65.

Received for publication : December 31, 2012
Accepted for publication : August 9, 2013