
Genetic diversity analysis in G. barbadense accessions

GENETIC DIVERSITY ASSESSMENT OF G. BARBADENSE ACCESSIONS TO WIDEN COTTON (GOSSYPIUM SPP.) GENE POOL FOR IMPROVED FIBRE QUALITY

N. Manikanda Boopathi\(^1\)*, A. Gopikrishnan\(^1\), N. Jagadeesh Selvam\(^1\), R. Ravikesavan\(^2\), K. Iyanar\(^2\), S. Muthuraman\(^2\) and N. Saravanan\(^2\)

\(^1\) Department of Plant Molecular Biology and Biotechnology, Centre for Plant Molecular Biology and Biotechnology, \(^2\) Department of Cotton, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore-3

* Corresponding author: nmboopathi@tnau.ac.in; biotechboopathi@yahoo.com
ABSTRACT

Understanding the genetic diversity and breeding potential of *Gossypium* spp., accessions are vital for the genetic improvement of cotton fiber quality and productivity. A substantial variation of fiber quality traits is known to exist in *G. barbadense* germplasm. The objective of this study was to assess the genetic diversity and relationship among the *G. barbadense* accessions using simple sequence repeat (SSR) markers. Thirty *G. barbadense* accessions and five *G. hirsutum* cultivars were genotyped with 88 SSR markers that generated 151 alleles with an average of 1.72 alleles per locus. Polymorphism information content (PIC) value has an average of 0.39 with a range of 0.11–0.73. The genetic similarity (GS) coefficient was estimated and clustering analysis based on the GS grouped the 35 *Gossypium* accessions into two distinct clusters comprising *G. barbadense* and *G. hirsutum*. Grouping based on clustering analysis was in good agreement with available pedigree and genetic background information. Diverse pairs of *G. barbadense* accessions were identified which were polymorphic at many SSR loci and they can be used as parents for hybrid development to maximize the fiber productivity and quality and development of segregating populations to map genes controlling fiber quality in cotton.
**Key words:** Cluster analysis, Genetic distance, Genetic diversity, Genetic similarity co-efficient, Simple sequence repeats (SSRs)

**INTRODUCTION**

Success of any crop breeding program is based on the knowledge of and availability of genetic variability for efficient selection. Genetic similarity (or genetic distance) estimates among genotypes are helpful in selecting parental combinations for creating segregating populations so as to maintain genetic diversity in a breeding program and the classification of germplasm into heterotic groups for hybrid crop breeding (Ali et al., 2008). The search for and establishment of heterotic groups can be based on geographical origin, agronomical traits, pedigree data or on molecular marker data. Recently molecular markers found to be more useful and various types of molecular markers are available for genome analysis (Gingle et al., 2006). SSRs in particular have been reported to be very useful to analyze the structure of germplasm collections as these are abundant, co-dominant, multi-allelic, highly polymorphic and chromosome-specific and hence SSR markers have been extensively used in genetic diversity studies (Lacape et al., 2007).
Rising cotton productivity is increasingly relying on knowledge of countries’ capacity to develop improved cultivars. A better understanding of the available genetic diversity and the breeding potential of specific accessions is important for the choice of parents for use in breeding programs (Gingle et al., 2006). However, breeding programs are limited by the under-utilization of the available genetic diversity. Among the four species of *Gossypium* that produce seeds with spinnable fibres, *G. hirsutum* dominates (accounting for ~90% of) the world’s cotton fibre production. The second most cultivated species, *G. barbadense*, includes superior extra long, strong and fine cottons. However, compared with *G. hirsutum*, the marketing advantage of ‘high quality’ *G. barbadense* cottons is offset by their lower productivity and a narrower adaptability to harsh environments (Lacape et al., 2007). Since *G. hirsutum* and *G. barbadense* display complementary characteristics there is a possibility to transfer best genomic segments that governs fiber quality from *G. barbadense* to *G. hirsutum*. Several strategies were developed including advanced backcross QTL analysis (AB-QTL), which allow simultaneous identification and transferring of favourable QTL alleles from unadapted (e.g. related species, wild species, land races etc.,) to cultivated germplasm (Lacape et al., 2007). However, the central
dogma of molecular breeding involves the utilization of molecular marker fingerprints to improve selection efficiency in plant breeding programs and thus avoiding the introgression of already existing gene pool (Ali et al., 2008). Hence, it is important to characterize the available G. barbadense germplasm at molecular level and selecting the diverse accessions for the production of better cross combinations with G. hirsutum. The objective of the present study is analysing the genetic diversity exists in the G. barbadense and G. hirsutum accessions and selects the diverse parents for future breeding program which may widen the cotton gene pool for fiber quality traits.

MATERIALS AND METHODS

Plant materials

Thirty G. barbadense accessions and five G. hirsutum cultivars were selected to evaluate the genetic diversity and the relationship among them. G. barbadense accessions used in this study include B1, B2, B3, B4, B9, B10, B12, B16, B17, B21, B22, B23, B24, B26, B28, B33, B34, B35, B37, B41, B42, B45, B46, B48, B49, B50, B51, B52, B53 and B54. These are the accessions that have been selected as superior fiber quality G. barbadense lines and are being used in breeding program at TNAU, Coimbatore. In addition to these accessions, five G. hirsutum cultivars viz., MCU5, MCU12, MCU13,
KC2 and SVPR2 that are widely cultivated in Tamil Nadu were included as reference lines to this study. Seeds were obtained from germplasm bank maintained at TNAU, Coimbatore and plants were raised to collect the leaf samples.

**DNA extraction, PCR profile and gel electrophoresis**

DNA was isolated by essentially following the method described by Callahan and Mehta, (1991) and the DNA quality and quantity was tested by 0.8 % agarose gel electrophoresis. Cotton SSR primers were synthesized based on the published sequence information and the PCR reaction mix preparation and PCR conditions were performed as described by (Lacape et al. 2007). The 4% Metaphor agarose gel electrophoresis was done to separate the alleles generated by SSR primers and the gel was photographed using gel documentation system (Alpha Imager, USA).

**Genetic diversity estimation**

The amplified fragments of each SSR marker were scored as “1” and “0”, where “1” indicated the presence of a specific allele (band) and “0” indicated its absence. Polymorphism information content (PIC) of SSR markers was calculated using the formula developed by Anderson et al. (1993). A PIC value of each locus was calculated as:
PIC\(_j\) = 1 - \(\sum_{l=1}^{L} P_{lj}^2\),

where \(P_{lj}\) is the relative frequency of the \(l^{th}\) allele for the locus \(j\) and is summed across all the alleles (L) over all lines. PIC provides an estimate of the discriminatory power of a locus by taking into account, not only the number of alleles that are expressed, but also the relative frequencies of those alleles. PIC values may range from 0 (monomorphic) to 1 (very highly discriminative), with many alleles in equal frequencies.

Genetic diversity estimate related analyses were done using NTSYSpc ver.2.02i (Rohlf, 2000). Genetic similarities (GS) between pairs of accessions were measured by the DICE similarity coefficient based on the proportion of shared alleles with SIMQUAL module. Genetic distances between pairs of lines were estimated as GD or \(D = 1 - GS\). The clustering of accessions was done based on a similarity matrix using an unweighted pair group method with arithmetic average (UPGMA) algorithm following SAHN module. The clustering result was used to construct a dendrogram following TREE module (Ali et al., 2008).

**RESULTS AND DISCUSSION**

Understanding the organization of genetic diversity within agricultural species has long been a matter of interest to identify
genetic resources to be used as parents of breeding programs, to define heterotic groups for hybrid breeding, etc. Molecular markers have proven particularly helpful to address these issues. Until recently, analysis of molecular marker data has been approached by hierarchical structuring, using diverse algorithms such as UPGMA, Ward, and factorial analysis, principal component analysis, and principal coordinate analysis. This study was aimed to identify the genetic diversity among the *G. barbadense* accessions using UPGMA algorithm which was noticed as appropriate method for this kind of study (Ali *et al*., 2008)

**Allelic diversity at SSR loci**

Eighty-eight SSR markers that generated 151 alleles were used to estimate the genetic diversity among 35 *Gossypium* accessions. The number of alleles revealed by each marker ranged from 1-5 alleles with an average of 1.72 alleles per locus. Polymorphism information content (PIC) value for the SSR loci, a measure of gene diversity, has an average of 0.39 with a range of 0.11–0.73.

SSR markers have been shown higher levels of polymorphism or more alleles in cotton (Iqbal *et al*., 2001; Lacape *et al*., 2007) and highly informative SSR markers assembled in PCR-multiplex are publically available which can be applied for molecular genetic
diversity studies of large germplasm collections (Lacape et al., 2007). The mean number of alleles per SSR locus detected in this study was similar to that detected by Iqbal et al., (2001) but lower than that reported by Lacape et al., (2007). Use of wild species and land races by Lacape et al., (2007) might be one of the reasons for the high allelic diversity. Another possible reason for relatively low gene diversity in the present set of cotton accessions could be that many of these accessions used in this study were either developed from crosses involving parents that were selected from the early introduction of cotton entries or were direct selections from early introductions with narrow genetic base (Ali et al., 2008).

It was interesting to note that some of SSR markers such as CIR167, CIR253, CIR79, CIR398, BNL1440, BNL3556, CML60, CIR209 and CIR399 produced rare alleles. These rare alleles could be of particular interest as they were uniquely linked to some particular genotypes. Such alleles are important because they may be diagnostic for particular genotype or for particular regions of the genome of Gossypium spp.,

Cluster analysis and genetic diversity

Based on the 151 shared alleles, genetic similarity co-efficient was estimated for each pair of the 35 Gossypium accessions which
ranged from 0.38 to 1.00. Except for *G. barbadense* accessions B37 and B46, the dendrogram (Figure 1) clearly discriminated all the 35 *Gossypium* accessions. B37 and B46 were remained together with a genetic similarity co-efficient of 1.00. Both of these accessions may be developed from same parental lines or lines descended from early introductions.

**Figure 1.** Dendrogram of 30 *G. barbadense* and five *G. hirsutum* accessions revealed by cluster analysis of genetic similarity estimates generated by DICE coefficient based on 88 SSR markers.

The cluster analysis grouped the 35 *Gossypium* accessions into two main groups *viz.*, I and II (Figure 1). Group I included all the five *G. hirsutum* accessions and Group II included all the 30 *G. barbadense* accessions. Thus grouping based on clustering analysis
was in good agreement with available pedigree and genetic background information.

Group I was again sub-divided into Ia and Ib. All the three MCU lines, which were developed from at least one parent as common in their parentage, were clustered in Ia and KC2 and SVPR2 were grouped in IIb. Interestingly, KC2 and SVPR2 are commonly cultivated in rainfed regions and are well adapted to water limited environments.

Group II was sub-divided into IIa and IIb. Sub-group IIa comprised 29 *G. barbadense* accessions and sub-group IIb had only one *G. barbadense* accession, B53. However, both of these sub-groups had more than 80 per cent similarity. Group IIa included two *G. barbadense* accessions viz., B37 and B46 which were shown 100 per cent similarity in this study and thus these two lines might be evolved from common parents. In contrast, several lines were found to be clustered in two or more different groups instead of being grouped together (IIa in Figure 1). This separation is most likely due to selection for different traits.

In conclusion, the present study has revealed valuable information on the relationship among a group of *G. barbadense* and *G. hirsutum* accessions. This genetic relationship information will be
very helpful in future cotton breeding programs to improve fiber productivity and quality and maintain broad genetic diversity. This study has also identified pairs of cultivars which could be used as parents to create mapping populations that can be employed to map genes influencing fiber quality and productivity. Mapping population development is in progress at this University and they are in F1 stage.

**Acknowledgement:** This work is supported by Department of Biotechnology, Ministry of Science and Technology, Government of India under Program Support for Research and Development in Agricultural Biotechnology at TNAU.

**REFERENCES**


