Australian Journal of Crop Science AJCS
AJCS 9(2):143-152 (2015) ISSN:1835-2707

Genetic diversity and population structure of upland cotton Brazilian cultivars (*Gossypium hirsutum* L. raça *latifolium* H.) using SSR markers

Leonel Domingos Moiana[2], Pedro SoaresVidigal Filho*[1], Maria Celeste Gonçalves-Vidigal[1], Luiz Paulo de Carvalho[3]

[1]Universidade Estadual de Maringá, Pós-Graduação em Genética e Melhoramento, Av. Colombo, 5790, Bloco J45, CEP 87020-900, Maringá-PR-Brazil
[2]Instituto de Investigação Agrária de Moçambique, Av. das FPLM 2698, C.P. 2698, Maputo, Mozambique
[3]Embrapa Algodão. Rua Osvaldo Cruz, 1143, CEP 58428-095, Campina Grande–PB-Brazil

***Corresponding author: vidigalfilhop@gmail.com**

Abstract

To better understand the genetic diversity of the cultivated upland cotton (*Gossypium hirsutum L.*) and its structure at the molecular level, microsatellite markers were used. The objective of this study was to evaluate genetic diversity and population structure in tetraploid cotton (*Gossypium hirsutum* L. race *latifolium* H.). Twenty cultivars and inbred lines from Embrapa Cotton Breeding Program, Brazil, were analyzed. From a total of 33 microsatellite (SSR) markers, twenty seven markers revealed 91 polymorphic SSR alleles. Two sub-populations were identified applying different methods (The Bayesian analysis, Principal Coordinates Analysis and Neighbor Joining Tree). Most of the cultivars belongs to Embrapa Cotton Breeding Program were allocated in sub-population I. The FST index indicated moderate genetic variability among the studied cultivars. In general, Embrapa cotton cultivars were the most dissimilar to GIBANGA and IMA CD05-8221 cultivars. The dissimilarity index ranged from 0.13 to 0.73 and the lowest genetic divergence was observed between BRS PRECOCE and BRS 286 genotypes. Combination of Embrapa cotton cultivars, GIBANGA and IMA CD-05 8221 is recommended for obtaining superior segregation in order to improve yield.

Keywords: Upland cotton (*Gossypium hirsutum*), microsatellite markers, population's structure, cultivar characterization.
Abbreviations: AMOVA_Analysis of molecular variance; PCoA_Principal Coordinates Analysis; PIC_polymorphism information content; SSR_Simple Sequence Repeat.

Introduction

Cotton belongs to the genus *Gossypium*, which has extensive phenotypic diversity among the approximately 50 species representing this genus (Campbell et al., 2009). Generally, four species are cultivated worldwide. Two of these cultivated species are diploids, *G. herbaceum* and *G. arboreum* ($2n = 2x = 26$) and two are allotetraploids *G. hirsutum* L. and *G. barbandense* L. ($2n = 4x = 52$) (Giband et al., 2010; Fang et al., 2013; Tyagi et al., 2014).

The independent domestication process of these species was responsible for the fiber evolution in the diploid species separately from the allotetraploid from the new world (Stewart et al., 2010). Within species, *G. hirsutum* shows great phenotypic diversity. The level of genetic diversity within *G. hirsutum* has been found to be higher than the other three cultivated cotton species (Wendel et al., 1992; Abdurakhmonov et al., 2008). So far, studies have indicated that this diversity does not represent in the present cultivated germplasm of Upland cotton (Fang et al., 2013). The success of plant breeding depends on the genetic variability, which rises from the genetic relationship and the genetic diversity between and within plant groups (Rana and Bhat, 2005; Abdellatif and Soliman, 2013). Breeding efforts can be greatly improved through better parental selection for generating segregating populations. It is also helpful to identify heterotic groups, understand population structure, and distinguish core set of lines for genetic analysis studies (Tyagi et al., 2014).

Genetic diversity estimates have been made using pedigree and morphological data (Van Esbroeck et al., 1999), biochemical markers (Wendel et al., 1992), and DNA-based molecular markers (Yu et al., 2012). Therefore, the use of morphological markers has been limited in genetic diversity studies because they are not enough to cover the entire genome and can be influenced by the environment and developmental stage (Lukonge et al., 2007). Several studies have been done on genetic diversity of the *Gossypium* genus using biochemical markers (storage proteins, isoenzymes) (Wendel et al., 1992). Molecular markers, on the other hand, are more reliable and informative since they can directly measure allelic diversity and give robust estimates of genetic distances (Tyagi et al., 2014). More recently, research with amplified fragment length polymorphisms (AFLP)(Pillay and Myers, 1999; Abdalla et al., 2001; Lukonge et al., 2007; Zhang et al., 2008; Rakshit et al., 2010), random amplified polymorphic DNA (RAPD) (Khan et al., 2000; Zhang et al., 2008; Maleia et al., 2010), restriction fragment length polymorphism (RFLPs) (Brubaker and Wendel, 1994; Zhang et al., 2008), inter-simple sequence repeat (ISSR) (Liu and Wendel, 2001) and microsatellite markers (Zhang et al., 2008; Rakshit et al., 2010; Kalia et al., 2011; Moiana et al., 2012; Abdellatif and Soliman, 2013; Fang et al., 2013; Tyagi

et al., 2014; Zhao et al., 2014a, b) have been carried out in the past few years.

Researches on the characterization of cotton cultivars using SSRs have analyzed few cultivars and limited numbers of molecular markers (Bertini et al., 2006). The SSR markers combine the good proprieties such as high polymorphism's level, codominance. They can be developed by codicant and uncodificant regions of the plants (Collard et al., 2005; Agarwal et al., 2008; Lin et al., 2010; Kalia et al., 2011; Zhao et al., 2014a).

SSR markers are considered as ideal and user friendly tool once they are conducted through PCR, genetically defined, typically co-dominant, and uniformly dispersed in the plant genome (Kantartzi et al., 2013). They are derived from expressed sequences target (EST-SSR), which is transferable in more than one SSR genome and they are located in transcript regions of the genome (Turkoglu et al., 2010; Abdullatif and Soliman, 2013). Several researches had reports low or very low levels of diversity within upland cotton, in spite of the high polymorphism observed between the primitive and semi-primitive races (Abdurakhmonov et al., 2008).

In Brazil, the studies of genetic diversity and population structure of cotton utilizing molecular markers are few, despite, existence of some studies which evaluated the genetic diversity of the Brazilian cultivars (Bertini et al., 2006; Lacape et al., 2007; Alves et al., 2009; Moiana et al., 2012).

The current study was undertaken with the objective to evaluate the genetic diversity and population structure of the cotton cultivars to the Embrapa Cotton Breeding Program through the SSR markers. These will contribute to identify the best combinations which could be recommended for the Embrapa Cotton´s Breeding Program. Then, it could be helpful for breeders in choosing of diverse parents for the Embrapa future Cotton Breeding Program.

Results and Discussion

Genetic diversity

27 SSR markers detected 91 alleles which were subjected to statistical analysis. All molecular markers were polymorphic. The number of alleles range from 2 to 6 with a mean of 3.99 by locus (Table 3).

Liu et al. (2000) obtained a mean of five alleles by locus, ranging from 2 to 11, screening a collection of 97 cultivars and primitive species of *Gossypium* genus through the 62 SSR markers. Lacape et al. (2007) found mean of 5.6 alleles by locus (2 to 17) evaluating 47 accesses of *Gossypium* genus through 320 SSR markers. Moiana et al. (2012) assessed 35 cultivars and eight inbred lines through 15 SSR markers. They obtained a mean of 6.9 alleles by locus (between 4 and 10), whereas, Fang et al. (2013) evaluated the genetic diversity and linkage disequilibrium of upland cotton in some main cotton producer countries through SSR markers based on wide genome. They detected 523 polymorphic loci containing 1381 alleles, ranging from 2 to 8 by locus, with a mean of 2.64.

The maximum number of alleles was six which found in marker JESPR 152 (Table 3). Lacape et al. (2007) evaluated 47 accesses of *Gossypium* genus through SSR markers. They obtained a maximum of 17 alleles, which were amplified in the markers JESPR 152 and CIR413. Moiana et al. (2012) accessed 35 cultivars and eight inbred lines of cotton through 15 SSR markers, which were utilized in this study. They obtained a maximum of 10 alleles to BNL-1694 and JESPR 152, respectively.

Overall, the number of alleles observed per marker depends on the selection of markers, germplasm to be genotyped, as well as the platform used for resolution of amplified products (Lacape et al., 2007). Fewer alleles (3.0) per locus in upland cotton are similar to the trend observed in other self-pollinated crops (Tyagi et al., 2014). Our results showed that the mean of major allelic frequency ranges from 0.40 for marker BNL4035 for 0.95 to markers BNL1053, BNL 3257, BNL 3902, BNL 686 with a mean of 0.722 (Table 3). In relation to the genetic diversity, the locus with the major diversity between the cultivars was JESPR 152 with a value of 0.69, while markers BNL 1053, BNL 3257, BNL 3902, BNL 686 showed the minor genetic diversity (0.095) (Table 3).

The values of the PIC range from 0.090 to 0.580, with a mean of 0.361 (Table 3), indicating that informative potential was moderated (PIC between 0.25-0.5). Liu et al. (2000 evaluated a collection of 97 cultivars and primitive species of *Gossypium* genus through the 62 SSR markers. They found PICs of 0.05 to 0.82 (mean of 0.31), while, Lacape et al. (2007) evaluated 47 accesses of *Gossypium* through microsatellites and obtained 0.08 to 0.89 (mean of 0.55). Moiana et al. (2012) evaluated 435 cultivars and eight inbred lines of cotton through 15 SSR markers and obtained 0.37 to 0.77 with a mean of 0.65. Zhao et al. (2014a) evaluated a genetic diversity and population structure of elite cotton (*Gossypium hirsutum L.*) germplasm using SSR markers found a PIC mean of 0.30. Fang et al. (2013) evaluated the genetic diversity and the linkage disequilibrium of Upland cotton producer´s countries through the microsatellite markers based on genome wide study and obtained PIC values ranging from 0.0052 to 0.64 (with a mean of 0.29).
The results of this study indicated that cultivars from Embrapa had a moderate genetic diversity (Table 3).

Population structure and cluster analysis

The sub-division of the population were determined by software STRUCTURE 2.3.3 and Neighbor Joining Tree based on C.S. Chord genetic distance (Cavalli-Sforza and Edwards, 1967) showed a significant diversity between the sub-populations.

The Bayesian analysis (Pritchard et al., 2000) indicated that 20 cultivars were distributed in two distinct populations (Fig 1). For value K=2, 14 cultivars were allocated in sub-population 1 and other 6 cultivars were allocated in the sub-population 2. For value k=3, this situation was maintained (Fig 1).

The cultivar accessed was distributed in K=2 optimal value. This was determined by mean ln prob of data generated by Structure Harvester (Earl and vonHoldt, 2012). The genetic dissimilarity matrix and the Neighbor Joining Tree were performed with POWERMARKER 3.25 and MEGA 5.2, respectively, to estimate the phylogenetic relation between the sub-populations (Fig. 2).

Table 1. List of the cultivar assessed for genetic diversity analysis.

Number of the order	Origin	Cultivar	Pedigree
1	Embrapa	BRS PEROBA	Selection in ITA 90xDelta Opal
2	Embrapa	BRS 7H	Tamcot sp 37/IAC 17
3	Embrapa	ITA90	Selection in Deltapine 90
4	Embrapa	BRS 8H	CNPA 77-105*D3-79
5	Embrapa	BRS ARAÇÁ	CNPA 98-6399
6	Embrapa	BRS PRECOCE	C-100-7-81/PNH3
7	Embrapa	BRS SUCUPIRA	BRS 97-700
8	Embrapa	BRS 336	CHACO 520/BRS Itaúba/Delta Opal
9	Embrapa	BRS IPÊ	CNPA 97-2046
10	Embrapa	BRS 286	ITA 90 e BRS 7H
11	Embrapa	BRS CAMAÇARI	CNPA 97-1682
12	Embrapa	ITA96	Auburn 56/Deltapine
13	Embrapa	BRS 335	DP 4049/ ITA 96 /Delta Opal
14	Embrapa	BRS ANTARES	Selection in CNPA SRI_5
15	Embrapa	BRS 201	Selection in CNPA 96-12
16	Embrapa	BRS FACUAL	CNPA SRI_5 /Sicala 3-4 (94-171)
17	Embrapa	BRS PRECOCE1	Selection in GH-11-9-75
18	Embrapa	BRS CEDRO	CNPA 97-1067
19	Landrace	GIBANGA	-
20	IMAMT	IMA CD-05 8221	Selection in FM 910

Source: Bertini et al. (2005), Embrapa (2006), Embrapa (2011), Belot et al. (2012).

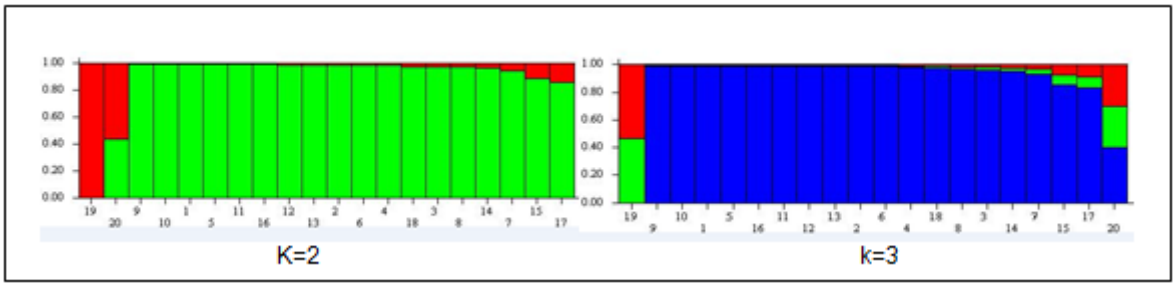

Fig 1. Q-plot showing clustering of 20 cotton cultivars based on analysis of genotypic data using STRUCTURE 2.3.3 for K=2 and K=3 values. Each accession is represented by a vertical bar (colors represent different sub-population). Sub-population 1 (1-BRS PEROBA, 2-BRS 7H, 3-ITA90, 4-BRS 8H, 5-BRS ARAÇÁ, 6-BRS PRECOCE, 8-BRS 336, 9-BRS IPÊ, 10-BRS 286, 11-BRS CAMAÇARI, 12-ITA96, 13-BRS 335, 16-BRS FACUAL and 18-BRS CEDRO) and subpopulation 2 (7- BRS SUCUPIRA ,14-BRS ANTARES, 15-BRS 201, 17-BRS PRECOCE 1, 19-GIBANGA and 20-IMA CD-05 8221).

Regarding to the C.S. Chord dissimilarity distance, the most divergent values were 0.73, which detached the following cultivars GIBANGA, BRS PEROBA and BRS 7H (Table 5). These cultivars can proportionate the heterotic effect when they are used as progenitors in the cotton breeding program, which involves hybridizations (Falconer, 1996).

Bertini et al. (2006) evaluated the genetic diversity of the cultivars through the microsatellite markers and obtained the dissimilarity index ranging from 0.0 to 0.41. In turn, Lacape et al. (2007) also utilized the SSR markers to evaluate the diversity within cultivars of *G. hirsutum*, races and species of cotton through the microsatellites observed that the dissimilarity index ranged from 0.12 to 0.94, meanwhile, Moiana et al. (2012) evaluated a diversity between 35 cultivars and eight inbred lines, utilizing SSR markers and found dissimilarity distances (0.06 to 0.90).

The major cotton cultivars had the same progenitors, and the most cultivars released in Brazil resulted from cultivars and inbred lines introduced from USA or from reselection of other cultivars (Bertini et al., 2005; 2006, Fang et al., 2013). According to Bertini et al. (2005) more than 30% of Brazilian cultivars of cotton were released between 1970 and 1990 in the USA and were obtained through selection among other cultivars. Cotton is considered as a self-pollinated plant with cross-pollinated rates. It is observed that the insects can participate in pollination and 12 of 30 cultivars utilized in research were obtained by reselection from other cultivars and lines. Some of them are similar to those used in this study.

Furthermore, according to Fang et al. (2013) seventeen cultivars released between 1899 and 1950 had 9 alleles and a mean of 0.53, although 27 cotton cultivars released from 1951 to 1980, contained 12 alleles and a mean of 0.44. In turn, 50 cultivars released since 1981-now had 12 alleles with an average of 0.24 (Bourland and Jones, 2010; Fang et al., 2013; Bowman et al., 2006).

These results show that the modern cultivars of USA are progenitors of many Brazilian cultivars (Bertini et al., 2006) and they gradually lost their genetic diversity in the last century. These cultivars were dominated by cultivars which were resistant to root-knot nematodes and Fusarium wilt. Despite that these cultivars are obsolete (Auburn 56, Auburn 634RNR), they have been used as a resistance source (Shen et al., 2006).

Also, the genitors were cultivars developed by the major Cotton Breeding Program of USA, as a Deltapine, Suregrow,

Table 2. Twenty seven primers SSR utilized in the reactions for amplification of the DNA fragments (PCRs)

Number of the order	Marker	Motif	Nucleotides sequences (5´- 3´)		Number of the order	Marker	Motif	Nucleotides sequences (5´- 3´)	
1	BNL1694	$(AG)_{19}$, $(TC)_{19}$	CGTTTGTTTTCGTGTAACAGG	Forward	15	JESPR152	$(GAA)_{50}$	GATGCACCAGATCCTTTTATTAG	Forward
	BNL1694		TGGTGGATTCACATCCAAAG	Reverse		JESPR152		GGTACATCGGAATCACAGTG	Reverse
2	BNL3408	$(GT)_2AT(GT)_{12}$	ATCCAAACCATTGCACCACT	Forward	16	CIR 081	$(AC)_7$	AAAGAACCCATGAGAAGA	Forward
	BNL3408		GTGTACGTTGAGAAGTCATCTGC	Reverse		CIR 081		GCTGTCTATGTTGGTGG	Reverse
3	BNL2495	$(AG)_{14}$, $(TC)_{14}$	ACCGCCATTACTGGACAAAG	Forward	17	BNL 1440	$(AG)_{15}$	CCGAAATATACTTGTCATCTAAACG	Forward
	BNL2495		AATGGAATTTGAACCCATGC	Reverse		BNL 1440		CCCCCGGACTAATTTTTCA	Reverse
4	BNL2572	$(GA)_{23}$	GTCCTATTACTAAAATTGTTAATTTAGCC	Forward	18	BNL 1414	$(AG)_{16}$	AAAAACCCCTTTCCATCCAT	Forward
	BNL2572		CGATGTTAAATCAATCAGGTCA	Reverse		BNL 1414		GGGTGTCCTTCCCAAAAATT	Reverse
5	BNL1053	$(AC)_{16}$	AGGGTCTGTCATGGTTGGAG	Forward	19	BNL 2921	$(AG)_{10}$	CGAGAGATTTTAAAGGGAAACA	Forward
	BNL1053		CATGCATGCGTACGTGTGTA	Reverse		BNL 2921		GGGAGTGGTCTGATGGAAAA	Reverse
6	BNL3649	$(TC)_{20}$	GCAAAAACGAGTTGACCCAT	Forward	20	BNL 1434	$(AG)_{13}$	AAATTCAAGAATCAAAAAACAACA	Forward
	BNL3649		CCTGGTTTTCAAGCCTGTTC	Reverse		BNL 1434		TTATGCCAAAGTATATGGAGTAACG	Reverse
7	BNL2544	$(AG)_{11}$	GCCGAAACTAAAACGTCCAA	Forward	21	BNL 2496b	$(GA)_{15}$	AATTGACGAAAGGTGGAACG	Forward
	BNL2544		TCCTTACTCACTAAGCAGCCG	Reverse		BNL 2496b		CAAGCGGTAATAGGAATGCA	Reverse
8	BNL2494	$(AG)_{14}$, $(TC)_{14}$	ACCGCCATTACTGGACAAAG	Forward	22	BNL 3590	$(CA)_{20}$	TCTTCCCTCTCTTTCTCTTTCG	Forward
	BNL2494		AATGGAATTTGAACCCATGC	Reverse		BNL 3590		ACACGGAAGACCAACCAAGT	Reverse
9	BNL3816	$(TG)_{15}$, $(TG)_5TA(TG)_{15}$	GTTAGCCACGTGTTAGTTCTATG	Forward	23	BNL 3257	$(AC)_{13}+(AT)_{10}$	CAATCTGGGATCAAAAAAACC	Forward
	BNL3816		ATCGATCACTTGCTGGTTCC	Reverse		BNL 3257		GGTGAAACATAGCGTGTTGC	Reverse
10	BNL3904	$(GT)_{11}$	ATGCATTAATGAGTCGATAGGC	Forward	24	BNL 1317	$(AG)_{14}$	AAAAATCAGCCAAATTGGGA	Forward
	BNL3904		GCACAAAGAAAACAAACTGTACG	Reverse		BNL 1317		CGTCAACAATTGTCCCAAGA	Reverse
11	BNL3998	$(AC)_{11}$	CGGCGAGAAGTTGAAAGATC	Forward	25	BNL 3902	$(GT)_{18}$	GAGTTTGGGGGCTGTGTATG	Forward

	BNL3998		TGTGCAAAAGTGGGTGGTAA	Reverse		BNL 3902		GGGGTGCTTATGTCAGACGT	Reverse
12	BNL4035	$(TG)_{10}$	TGCATCTGCATTTGGGATTA	Forward	26	BNL 686	$(GA)_{22}$	ATTTTTCCCTTGGTGGTCCT	Forward
	BNL4035		TAGCCAACCGTTACACACCA	Reverse		BNL 686		ACATGATAGAAATATAAACCAAACACG	Reverse
13	NAU864	$(TG)_{10}$	GGATTAATTAGCCCCCACAT	Forward	27	CM43	$(TC)_{20}$	GCGCAGATATTATTATCACAGC	Forward
	NAU864		TCTTTTTCAGCTTGGGTTCT	Reverse		CM43		TATATAAATTTGCATCAGTTGGC	Reverse
14	JESPR292	$(CTT)_7$	GCTTGCAATCTCCTACACC	Forward					
	JESPR292		GAATATGTTTCATAGAATGGC	Reverse					

Source: Cotton Marker Database (2013).

Table 3. Genetic diversity´s index estimates by 27 microsatellites markers assessed.

Marker	Major allele frequency	Allele number	Genetic diversity	PIC
BNL3408	0.550	5	0.620	0.571
BNL3816	0.650	5	0.530	0.489
BNL3904	0.700	2	0.420	0.332
BNL4035	0.400	3	0.655	0.580
BNL3649	0.600	5	0.595	0.561
BNL2544	0.750	4	0.410	0.379
BNL1053	0.950	2	0.095	0.090
NAU864	0.600	5	0.570	0.519
JESPR292	0.600	5	0.595	0.561
BNL1694	0.850	3	0.265	0.247
BNL2572	0.750	3	0.405	0.368
BNL3998	0.900	3	0.185	0.177
JESPR152	0.450	6	0.690	0.643
BNL2495	0.750	3	0.395	0.347
BNL2494	0.900	2	0.180	0.164
CIR-081	0.800	3	0.335	0.303
BNL-1440	0.550	4	0.595	0.531
BNL-1414	0.900	2	0.180	0.164
BNL-2921	0.450	4	0.625	0.551
BNL-1434	0.750	3	0.395	0.347
BNL-2496b	0.550	4	0.595	0.531
BNL-3590	0.650	4	0.510	0.452
BNL-3257	0.950	2	0.095	0.090
BNL-1317	0.800	3	0.335	0.303
BNL-3902	0.950	2	0.095	0.090
BNL-686	0.950	2	0.095	0.090
CM-43	0.800	2	0.320	0.269
Mean	0.722	3	0.399	0.361

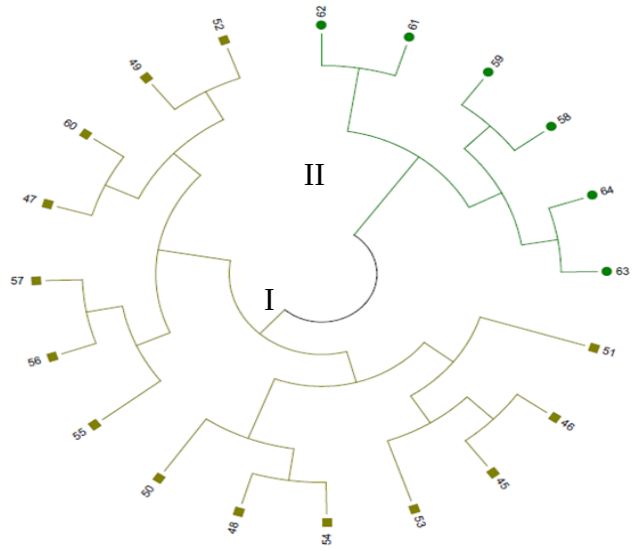

Fig 2. Distribution of 20 cultivars of cotton, according to Neighbor-Joining Tree based on C.S. Chord distance. Sub-population 1 (45-BRS PEROBA, 46-BRS 7H, 47-ITA90, 48-BRS 8H, 49-BRS ARAÇÁ, 50-BRS PRECOCE, 51- BRS SUCUPIRA, 52-BRS 336, 53-BRS IPÊ, 54-BRS 286, 55-BRS CAMAÇARI, 56-ITA96, 57-BRS 335, 60-BRS FACUAL,) and subpopulation 2 (58-BRS ANTARES, 59-BRS 201, 61-BRS PRECOCE 1, 62-BRS CEDRO, 63-GIBANGA and 64-IMA CD-05 8221).

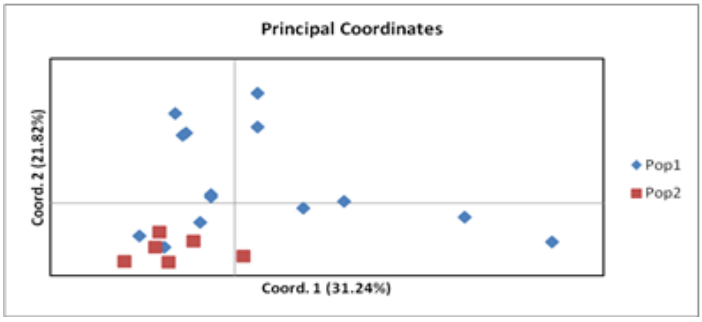

Fig 3. Principal Coordinates Analysis (PCoA1, 31.24% vs. PCoA2, 21.82%) of SSR data showing 53.06 % genetic diversity among 20 cotton cultivars. Sub-population 1 (BRS PEROBA, BRS 7H, ITA90, BRS 8H, BRS ARAÇÁ, BRS PRECOCE, BRS 336, BRS IPÊ, BRS 286, BRS CAMAÇARI, ITA96, BRS 335, BRS FACUAL and BRS CEDRO) and sub-population 2 (BRS SUCUPIRA , BRS ANTARES, BRS 201, BRS PRECOCE 1, GIBANGA and IMA CD-05 8221).

Paymaster, and Stoneville, and from the Cotton Breeding Program of Australia-CSIRO (Fibermax) (Fang et al., 2013). The Deltapine 50 and Deltapine 90 are the genitors of the major cultivars originated from Australia and USA. The great success of the cotton transformation was acquired utilizing mainly the cultivars Coker 312 and Coker 315 (Duncan, 2010). However, the eventual selection for cultivars adapted was developed from four basical categorical upland cotton cultivar (Acala, Delta, Plains, Eastern), which modernized the major part of the upland cotton in the world (Stewart et al., 2010). The cultivar and progenitor CNPA SRI_5 were obtained by the large population evolving several genitors in their pedigree (Bertini et al., 2006). Therefore, for the cotton breeding program in Brazil, the crosses between GIBANGA × BRS PEROBA, GIBANGA × BRS 7H, GIBANGA × BRS FACUAL, GIBANGA × IMA CD05-8221, IMA CD05-8221 × ITA90 should be the most recommended ones (Table 5). The cultivar GIBANGA, as a long growing season (bush tree), might be resistant to the main cotton disease and insect pests. The cultivar BRS PEROBA was obtained through genealogy selection method, applied in the population F_2, derived from the bi-parental cross between ITA 90 and Delta Opal (Embrapa, 2004). From this cross, the inbred line CNPA 98-7633 were selected. Its presents were major resistant to viruses, alternary (*Alternaria macrospora*), bacterial blight (*Xanthomonas axonopodis pv malvacearum*) and the *Fusarium* wilt and nematodes complex, when compared to ITA90. Regarding to other diseases such as ramulosis [*Colletotrichum gossypii* (Speg.) Cif.] that occur in Cerrados region, Brazil, this cultivar shows a moderate susceptibility, compared to ITA90 (Embrapa, 2006).

A cultivar BRS FACUAL was derived from the cross between CNPA SRI_5 and Sicala 34, which originated from the inbred line ITA 94-171 (Embrapa, 2006). This cultivar shows resistance to ramulosis, red virose and blue disease, bacterial blight, spot of *Stemphylium solani*, Areolate mildew (*Ramularia gossypii*), alternary (*Alternaria macrospora*). It is susceptible to root-knot nematode and *Fusarium* wilt and nematode complex (Embrapa, 2004, Embrapa, 2006).

The cuItivar ITA90 is derived from a composite of 13 plants selected from Deltapine Acala 90, which submitted to three cycles of mass selection for the viruses resistance (Embrapa, 2004). It is resistant to *Stemphylium solani* spots and is moderately resistant to alternary spot, nematodes and ramulose (Embrapa, 2004, Embrapa, 2006). The cultivar Sicala, derived from CSIRO Australia, was selected from cultivar DP 16 (Bertini et al., 2006). Furthermore, the cultivars ITA90, ITA96 were included in the phylogeny of the major cultivars utilized in this study and they have the Deltapine as a genitor (Embrapa, 2006).

Genetic diversity parameters

Genetic relationships between accessions were further studied using Principal Coordinates Analysis (PCoA) (Fig. 4). The first two axes of PCoA accounted for 53.06 % of the variation. These justify 53.06 % of genetic diversity in global population. These results were in accordance with Tyagi et al. (2014), which found 59.2% of genetic diversity, indicating a low level of genetic diversity in *G. hirsutum* germplasm with continuous variation between sub-populations.

The Analysis of Molecular Variance (AMOVA), the F_{st} (FWright) was equal to 0.108 reflected the existence of a moderate genetic variability (Balloux and Lugon-Moulin, 2002), while the level of genetic fixation in the sub-populations was 0.330 and the level of the genic fixation in the total population was 0.403 (Table 4). These results indicate a low heterozygosis in the sub-population, compared to global population, showing that the individuals were self-pollinated during the breeding process and had the grade of parentage and relationship (Wright, 1943; Menezes et al., 2014).

The variability was prominently conserved within the sub-populations (85%) compared to sub-populations (15%) (Table 4). Tyagi et al. (2014) evaluated the genetic diversity and population structure among the US upland cottons, and obtained conserved within the sub-populations (65.84 %)

Table 4. Analysis of molecular variance of two clusters (K=2) of 20 cotton cultivars detected by Structure analysis.

Source	DF	SS	MS	% of variation	Nm	F de Wright	Fis	Fit
Between Sub-population	1	34.2	34.2	15	2.06	0.108**	0.330**	0.403**
Within Sub-population	18	243	13.5	85				
Total	19	277.2						

DF=Degree freedom, SS=Sum Square, MS=Mean Square, SIG=Probability, Fis=Genic fixation in the subpopulation, Fit=Genic fixation in the global population, Nm=number of migrants ,**=significance by 1% de probability.

Table 5. Resume of matrix of Dissimilarity, C.S. Chord distance, obtained with 27 DNA markers, through SSR markers between 20 cotton cultivars.

Cultivars	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
45	0.00	0.30	0.30	0.33	0.40	0.33	0.40	0.37	0.27	0.37	0.47	0.43	0.53	0.50	0.47	0.37	0.40	0.40	0.73	0.57
46		0.00	0.47	0.43	0.33	0.33	0.37	0.33	0.37	0.40	0.47	0.50	0.53	0.50	0.40	0.50	0.50	0.50	0.73	0.50
47			0.00	0.30	0.33	0.30	0.43	0.37	0.33	0.37	0.43	0.37	0.43	0.43	0.47	0.37	0.40	0.47	0.70	0.60
48				0.00	0.23	0.20	0.33	0.33	0.27	0.13	0.27	0.33	0.30	0.33	0.33	0.40	0.33	0.30	0.67	0.43
49					0.00	0.27	0.37	0.23	0.30	0.20	0.20	0.27	0.30	0.30	0.30	0.33	0.30	0.37	0.63	0.43
50						0.00	0.27	0.33	0.23	0.13	0.27	0.30	0.30	0.37	0.37	0.40	0.33	0.30	0.63	0.40
51							0.00	0.33	0.30	0.27	0.37	0.37	0.37	0.37	0.30	0.50	0.37	0.47	0.60	0.33
52								0.00	0.33	0.30	0.33	0.37	0.37	0.30	0.27	0.40	0.37	0.40	0.70	0.43
53									0.00	0.23	0.33	0.27	0.37	0.40	0.43	0.37	0.40	0.37	0.67	0.50
54										0.00	0.13	0.23	0.20	0.27	0.27	0.33	0.27	0.23	0.63	0.40
55											0.00	0.20	0.23	0.30	0.37	0.33	0.37	0.33	0.67	0.47
56												0.00	0.17	0.27	0.40	0.30	0.43	0.40	0.67	0.50
57													0.00	0.17	0.30	0.37	0.27	0.30	0.60	0.40
58														0.00	0.17	0.37	0.23	0.33	0.60	0.33
59															0.00	0.43	0.23	0.33	0.60	0.27
60																0.00	0.33	0.37	0.67	0.53
61																	0.00	0.20	0.57	0.37
62																		0.00	0.67	0.43
63																			0.00	0.43
64																				0.00

Coeficient of CSchords dissimilarity for 20 genotypes (0,00 – 0,73) . 45-BRS PEROBA, 46-BRS 7H, 47-ITA90, 48-BRS 8H, 49-BRS ARAÇÁ, 50-BRS PRECOCE, 51-BRS SUCUPIRA, 52-BRS 336, 53-BRS IPÊ, 54-BRS 286, 55-BRS CAMAÇARI, 56-ITA96, 57-BRS 335, 58-BRS ANTARES, 59-BRS 201, 60-BRS FACUAL, 61-BRS PRECOCE, 62-BRS CEDRO, 63-GIBANGA e 64-IMA CD05-8221)

and among the sub-populations (31.4 %). The moderate level of variability among population can be explained by the high value of gene flow (Nm=2.06). These levels are inversely proportional differentiation among population (Fst) (Wright, 1943; Balloux and Lugon-Moulin, 2002). The low level of diversity observed by previous studies in upland cotton was the main bottle-neck, which occurs during the process of the domestication (Wendel and Crown, 2003). Recently, the development of the transgenic cultivars of cotton in the USA and other countries has contributed to the loss of the genetic diversity (Zhang et al., 2008).

Apart from the initial bottleneck encountered during domestication process, cotton breeding has frequently involved crossing and re-selections within small sets of breeding materials, which have led to the loss in genetic diversity (Tyagi et al., 2014).

The results of the genetic diversity of this study (Table 1), through the SSR markers are consistent with the breeding history and the pedigree of the major cultivars. According to Bertini et al. (2005) the low polymorphism observed in the Brazilian cultivars, is linked to the fact that the cotton breeding program has a low genetic base. Although, Fang et al. (2013) emphasized that the low level of polymorphism observed between the cotton cultivars was mainly due to bottleneck occurred during the domestication process. These facts illustrate the existence of a moderate differentiation among the cultivars accessed (Table 4). This study clearly shows that the cotton cultivars are losing their genetic diversity.

Materials and Methods

Plant materials

This study was conducted in the laboratory of molecular and biotechnology-NUPAGRI (Núcleo de Pesquisa Aplicada a Agricultura), Universidade Estadual de Maringá-PR, located at (latitude: 23 S25'31" and longitude: 51 W56'19"). Twenty cultivars, widely cultivated in Brazil, were selected for this study, 18 from the Embrapa Cotton Breeding Program, Campina Grande–PB, located at (latitude: 7 S09'50" and longitude: 35 W52'52"), one landrace from São Geraldo Farm, Porto Firme-MG, located at (latitude: 20 S40'22" and longitude: 43 W05'02") and one from Mato-Grossense Institute Cotton Breeding Program, Cuiabá-MT, located at (latitude 15 S35'46" and longitude 56 W05'48") (Embrapa, 2006, Embrapa, 2011, Belot et al., 2012) (Table1).

Genomic DNA extraction and quantification

Seeds of each cultivar and lines were grown in trials containing washed sand in the greenhouse, where they were kept until the harvesting of young leaves. Genomic DNA was extracted from fresh young leaves from four seedlings (Zhang and Stewart, 2000) of each cultivar (Table 1). After collection, leaves were immediately placed in eppendorf tubes and maintained in liquid nitrogen to preserve the DNA. DNA concentration was measured by a

Fluorometer Qubit™ and samples were diluted with buffer TE 0.1X to a concentration of 50 ng.μL^{-1} and stored at -20°C.

SSR genotyping

For the genetic divergence assay, 33 SSR markers were obtained from *Invitrogen-Induslab*. Only 27 most polymorphic markers with well-reproducible bands were selected. From these markers, 22 belong to BNLs, 1 NAU, 1 CIR, 2 JESPRs and 1 CM, showed high level of polymorphism (Table 2). Primer sequences for all SSR markers are publically available and were obtained from Cotton Marker Database (Cotton Marker Database, 2013). We preliminarily studied 3 genotypes (randomly) to identify SSR markers that gave reproducible amplification, the ideal mix and temperature of annealing could be confidently scored.

The verification of polymorphism were done with a thorough comparison of the products amplified from each genotype in the PCR reaction with AB device (Applied Biosystem Veritti™), utilizing the pair of the SSRs primers, following the protocol of Nguyen et al. (2004). Amplification reactions were conducted at a total volume of 20 μL which contained: 10 mM of buffer 10X (10 mM Tris HCl, pH 8.3, 50 mM KCl, 0.1% Triton X-100), 0.2 mM of dNTP, 1 U *Taq* DNA polymerase, 4 mM of each primers, $MgCl_2$ in a concentration compatible to the primer pair, 50 ng genomic DNA, and ultra-pure water.

The BNL primers amplification was performed using initial denaturation of 95°C for 12 min, followed by 30 cycles at 93°C for 1 min, annealing at 51 to 55°C for 2 min and extension at 72°C for 3 min. The last cycle was followed by an extension at 72°C for 7 min. The NAU and JESPR primers amplifications were performed using an initial denaturation of 94°C for 5 min, followed by 35 cycles at 94°C for 30 sec, annealing at 51 to 53°C for 1 min and an extension of 72°C for 1 min. The last cycle was followed by a final extension at 72°C for 8 min (Cotton Marker Database, 2013).

The DNA segments were fractionated in polyacrilamide gel (10%) containing 10 ml Acris/bis (29:1), 2.10 mL of glycerol , 14.47 mL of water, 2 mL TBE 10X, 22.50 μL of Temed and 450 μL of persulphate ammonium and a buffer TBE 1 X (Khantartzi et al., 2013). The fragments were visualized after the gel dyed with SYBER SAFE.

Statistical analysis

The software LAB IMAGE 1D, version 1.10 (Loccus Biotecnologia™) was used to genotype the samples. The major allelic frequency by locus, number of alleles by locus, mean of alleles by locus, mean of major allelic frequency and the genetic diversity by locus were obtained utilizing POWERMARKER 3.25 (Liu e Muse, 2005). The analysis was performed using a period of burn-in with 10.000, run length of 100.000 in the software STRUCTURE 2.3.3. The number of K was set from 2 to 20. A 20 independent structure runs were made for each K and the average likelihood value across 20 runs was calculated in the structure harvester (Pritchard et al., 2000; Earl and vonHoldt, 2012). We utilized the criterion proposed by Pritchard et al. (2000), where the mean ln of probability of the data "L(K)" was utilized to select the optimal value of K, as $L(K) = L(K)\text{-}L(K\text{-}1)$ minor than 50 and the optimal K was minor than 4 (Pritchard et al., 2000, Evanno, 2005).

In the population structure, the software GENALEX 6.3 was applied (Peakall and Smouse, 2006), generating the Principal Coordinates Analysis (PCoA) and the Analysis of Molecular Variance (AMOVA). The genetic distance between the accessions was determined through the C.S. Chords dissimilarity distance (Cavalli-Sforza and Edwards, 1967), which implemented by software POWERMARKER 3.25 (Liu e Muse, 2005), and the Neighbor-Joining Tree was generated using software MEGA 5.2 (Tamura et al., 2011).

Conclusion

A moderate variability was observed among the 20 analyzed cotton cultivars (*Gossypium hirsutum*). The cultivars from Embrapa Algodão (Embrapa Cotton Breeding Program), Brazil, showed more divergence, compared to cultivars GIBANGA and IMA CD05-8221. This revealed by the Probabilistic Method, Principal Coordinates Analysis and Neighbor-Joining Tree. For Cotton Breeding Programs in Brazil, the hybrid combinations between GIBANGA × BRS PEROBA, GIBANGA × BRS 7H, GIBANGA × BRS FACUAL, GIBANGA × IMA CD05-8221, IMA CD05-8221 × ITA90 could be used to obtain heterotic effect and transgressive segregation.

Acknowledgements

This study was made possible with technical and financial support from Ministery of Science and Technology of Mozambique and Embrapa-Algodão, Brazil. We would like to thank MCT-Mozambique, CAPES and CNPq-Brazil for scholarships, scholarship of productivity and financial support.

References

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

Abdellatif KF, Soliman YA (2013) Genetic relationships of cotton (*Gossypium barbadense* L.) genotypes as studied by morphological and molecular markers. Afr J Biotechnol. 12: 4736-4746.

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdukarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics. 92: 478–487.

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 27: 617–631.

Alves MF, Pereira FRA., Andrade AM, Menezes IPP, Hoffmann LV, Barroso PAV (2009) Marcadores moleculares polimórficos entre algodoeiros mocós e herbáceos. Rev Cienc Agron. 40: 406-411.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol. 11: 155–165.

Belot JL, Vilela PA, Salles AO (2012) Manual de boas práticas de manejo do algodoeiro no Mato Grosso. Cuiabá. Brasil, p 226.

Bertini CHCM, Schuster I, Sediyama T, Barros EG, Moreira MA (2005) Analysis of cotton genetic diversity

by microsatellites and pedigree. Crop Breed Appl Biotechnol. 5: 369-378.

Bertini CHCM, Schuster I, Sediyama T, Barros EG, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol. 2: 321-329.

Bourland FM, Jones DC (2010) Registration of 'UA48' cotton cultivar. J Plant Regist. 6: 15–18.

Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL (2006) Pedigrees of upland and pima cotton cultivars released between 1970 and 2005. Bulletin 1155. Mississippi Agricultural & Forestry Experiment Station Available in: http://msucares.com/pubs/bulletins/b1155.pdf. Accessed 14 of May of 2014.

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

Campbell BT, Williams VE, Park W (2009) Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. Euphytica. 169: 285.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet. 19: 233–257.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica. 142: 169–196.

Cotton Marker Database. Avalaible in: http://www.cottonmarker.org/. accessed on: 20 of January of 2013.

Duncan DR (2010) Cotton transformation. In: Zehr UB (ed.) Cotton, biotechnology in agriculture and forestry, Springer. 65: 65–77.

Earl DA, vonHold BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 4:359-361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14: 2611–2620.

Embrapa (2004) (Empresa Brasileira de Pesquisa Agropecuária-Centro Nacional de Pesquisa de Algodão). BRS Peroba-cultivar com resistência múltipla a doenças,1ª edição.

Embrapa (2006) (Empresa Brasileira de Pesquisa Agropecuária-Centro Nacional de Pesquisa de Algodão). Cultivares de algodoeiro herbáceo recomendadas para a região Meio-Norte do Brasil, p. 33.

Embrapa (2011) (Empresa Brasileira de Pesquisa Agropecuária-Centro Nacional de Pesquisa de Algodão. 8º Congresso Brasileiro de Algodão, Anais, p. 1907.

Fang DD, Hinze LL, Percy RG, Ping L, Deng D, Thyssen G (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. Euphytica. 10681-013-0886-2.

Falconer DS, Mackay TFC (1996). Introduction to quantitative genetics. England: Longman Group Limited, p. 464.Giband M, Dessauw D, Barroso PAV (2010). Cotton: Taxonomy, Origin and Domestication. In: Wakelyn PJ, Chaudhry MR (ed) Cotton: Technology for the 21st Century, International Cotton Advisory Committee. Washington, DC.

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

Kantartzi SK (2013) Microsatellites methods and protocols.Springer, New York Heidelberg Dordrecht, London. P. 377.

Kalia RK, Raí MK, Kalia S, Rohtas S, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica. 177: 309–334.

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.

Lin Z, Yuan D, Zhang X (2010) Mapped SSR markers unevenly distributed on the cotton chromosomes. Agriculture in China. 4: 257–264.

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

Liu KJ, Muse SV (2005) Powermarker: an integrated analysis environment for genetic marker analysis. Bioinformatics, 21: 2128–2129.

Liu B, Wendel JF (2001) Intersimple sequence repeat (ISSR) polymorphisms as a genetic marker system in cotton. Mol Ecol Notes. 1: 205-208.

Lukonge E, Herselman L, Labuschangne MT (2007) Genetic diversity of Tanzanian cotton (*Gossypium hirsutum L.*) revealed by AFLP analysis. Afr J Crop sci. 8: 773-776.

Maleia MP, Vidigal Filho PS, Gonçalves-Vidigal MC, Gonela A, Lacanallo GF, Moiana LD, Chamuene A, Sousa LL, Darben LM (2010) Genetic divergence among african and american cotton (*Gossypium hirsutum L.* race *latifolium* H.) cultivars inbred lines through random amplification of polymorphic DNA (RAPD) markers. Afr J Biotechnol. 9: 8539-8548.

Menezes IPP, Gaiotto FA, Hoffmann LV, Ciampi AY, Barroso PAV (2014) Genetic diversity and structure of natural populations of *Gossypium mustelinum*, a wild relative of cotton, in the basin of the De Contas River in Bahia, Brazil. Genet 142: 99–108.

Moiana LD, Vidigal Filho PS, Gonçalves-Vidigal MC, Lacanallo GF, Carvalho LP, Maleia MP, Pacheco CM, Ribeiro T, Neto HZ, Coimbra GK (2012) Genetic diversity and population structure of cotton (*Gossypium hirsutum* L. race *latifolium* H.) using microsatellite markers. Afr J Biotechnol.11: 11640-11647.

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers, Theor Appl Genet. 109: 167–175.

Peakall RS, Mousse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol. 6: 288-295.

Pillay M, Myers GO (1999) Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. Crop Sci. 39:1881–1886.

Prithchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Rakshit A, Rakshit S, Singh J, Chopra SK, Balyan HS, Gupta PK, Bhat SR (2010) Association of AFLP and SSR markers with agronomic and fiber quality traits in *Gossypium hirsutum* L. J Genet. 89:155–162.

Rana MK, Bhat KV (2005) RAPD marker for genetic diversity study among Indian cotton cultivars. Curr Sci. 88:1956-1961.

Shen X, Van Becelaere G, Kumar P, Davis RF, May OL, Chee P (2006) QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. Theor Appl Genet. 113:1539–1549.

Stewart JMCD, Oosterhuis D, Heitholt JJ, Mauney J (2010) Physiology of cotton. London, Springer, p. 563.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28: 2731-2739.

Turkoglu Z, Bilgener S, Ercisli S, Bakir M, Koc A, Akbulut M, Gercekcioglu R, Gunes M, Esitken A (2010) Simple sequence repeat-based assessment of genetic relationships among *Prunus* rootstocks. Genet Mol Res. 9: 2156-2165

Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet. 127:283–295.

Van Esbroeck GA, Bowman DT, May OL, Calhoun DS (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. Crop Sci. 39:323–328.

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

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

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78: 139-186.

Wright, S (1943) Isolation by distance. Genet. 28**:**114-138.

Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC (2012) Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. Euphytica. 187: 203–213.

Zhang J, Stewart JM (2000) Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci. 4: 193-201.

Zhang HB, Li YN, Wang BH, Chee PW (2008) Recent advances in cotton genomics. Int J Plant Genomics. P. 20.

Zhao Y, Wang H, Chen W, Li Y, Gong H, Sang X, Huo F, Zeng F (2014a) Genetic diversity and population structure of elite cotton (*Gossypium hirsutum L.*) germplasm revealed by SSR markers. Plant Syst Evol. 301(1):327-336.

Zhao Y, Wang H, Chen W, Li Y (2014b) Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. PLoS ONE. 9(1): e86308. doi:10.1371/journal.pone.0086308.