J. Crop Sci. Biotech. 2012 (June) 15 (2) : 93 ~ 99
DOI No. 10.1007/s12892-011-0120-5

RESEARCH ARTICLE

Molecular Diversity of Egyptian Cotton (*Gossypium barbadense* L.) and its Relation to Varietal Development

Kamal Fouad Abdellatif[1]*, Yehia Abdallah Khidr[1], Yasser Mohammad El-Mansy[2], Mohammad Mohammad El-Lawendey[2], Yasser Abdelraouf Soliman[2]

[1]*Plant Biotech. Dept., GEBRI, Sadat City, Minoufiya Univ., Egypt*
[2]*Cotton Research Institute, Agricultural Research Center, Giza, Egypt*

Received: December 16, 2011 / Revised: February 15, 2012 / Accepted: March 13, 2012
© Korean Society of Crop Science and Springer 2012

Abstract

Twenty-eight Egyptian cotton (*Gossypium barbadense* L.) genotypes (varieties and hybrids) were used for analysis of genetic diversity using DNA based markers (ISSR, SSR, and EST) and to study varietal development of cotton. The ISSR markers gave the highest percentage of polymorphic bands as well as polymorphic information content compared with the other molecular markers (i.e. EST and SSR markers). Using clustering analysis, no general clustering according to the pedigree history of the genotypes was observed. Using principal coordinate analysis (PCOORDA), cotton genotypes were separated by the first three principal coordinates (PC1, PC2, and PC3) accounting for 11.5, 8.6, and 7.2% of the total genetic variance, respectively. The cotton genotypes were distributed into three parts based on the first PC, each part containing a group of varieties having a common ancestor. 'Giza 12' variety was the common ancestor for the varieties included in the first part and 'Ashmouni' variety was the common ancestor for the varieties included in the second part, while both 'Sakha 3' and 'Sakha 4' varieties were common ancestors for the varieties included in the third part. The results of the PCOORDA also showed better resolution of the genetic diversity than cluster analysis especially in the illustration of the varietal development of cotton. That means that principal coordinate analysis can be strongly used either alone or in combination with cluster analysis to discuss both genetic diversity and varietal development in the cotton genotypes.

Key words: cluster analysis, Egyptian cotton, genetic diversity, molecular markers, PCOORDA

Introduction

Cotton is the world's most important fiber crop and the second most important oilseed crop. The primary product of the cotton plant has been the lint that covers the seeds within the seed pod, or boll. This lint has been utilized for thousands of years for clothing the people of ancient India, Asia, the Americas, and Africa. Cotton fabrics have been found in excavations at Mohenjo-Daro in India and in pre-Inca cultures in the Americas (Hutchinson et al. 1947). It is grown on every continent except Antarctica and in over 60 countries around the world. In many countries, cotton is one of the primary economic bases, providing employment and income for millions of people involved in its production, processing, and marketing (United Nations 2003).

Kamal Fouad Abdellatif (✉)
E-mail: kamal2004gr@yahoo.com

Cotton (*Gossypium* spp.) belongs to the genus Gossypium which contains about 50 species, 44 of which are diploid species (2n = 2x = 26) and six are allotetraploid (2n = 4x = 52). The diploid species comprise genomic groups A, B, C, D, E, F, G, and K and allotetraploid species are made up of two sub-genomic groups having an affinity with A and D genomes (Chen et al. 2007; Hussein et al. 2007; Stewart 1995). The cultivated cottons include *G. arboreum* L. and *G. herbaceum* L., both diploid species with an A genome native to southern Asia and Africa, and two allotetraploid species, *G. barbadense* L. and *G. hirsutum* L., with an AD genome from Central, North, and South America (Endrizzi et al. 1985). The cultivated cotton includes four species: two New World tetraploid, *G. hirsutum* L. and *G. barbadense* L., and two Old World diploid, *G. arboreum* L., and *G. herbaceum* L. Pima cotton or Egyptian cotton (*G. barbadense* L.) is grown for

its extra long, strong, and fine fiber (Hussein et al. 2007).

Genetic diversity of cotton plays an important role in sustainable development and food security, as it allows the cultivation of crops in the presence of various biotic and abiotic stresses. It is also important for the selection of parents that can be used in plant breeding programs. Characterizing genetic diversity and degree of association between and within varieties is the first step toward developing germplasm and crop cultivars. Successful crop improvement depends on genetic variability that arises from genetic diversity (Rana and Bhat 2004). A lack of genetic diversity may limit breeding progress and gain from selection. The information from genetic diversity is important when working to improve crop and develop new varieties (Cornelius and Sneller 2002).

Molecular markers have been used to measure genetic diversity and relationship within species and between their wild relatives in cotton. Polygenic morphological markers are influenced by the environment and are mostly quantitatively inherited (Lukonge et al. 2007). Among all molecular markers, Simple Sequence Repeats (SSR) also known as microsatellites, Inter-Simple Sequence Repeat (ISSR), and Expressed Sequence Tag (EST) have been involved in many genetic diversity studies. SSR are simple tandemly repeated di- to tetra-nucleotide sequence motifs flanked by unique sequences (McCouch et al. 2001). ISSR is a simple and informative genetic marker system in cotton for revealing both inter- and intraspecific variation (Liu and Wendel 2001). It uses primers that are complimentary to a single SSR and anchored at either the 5' or 3' end with a one- to three-base extension (Dongre et al. 2007; Preetha and Raveendren 2008). The SSR and ISSR markers are robust, reliable, quick, efficient, and reproducible with greater discriminative ability than the other techniques (Dongre et al. 2007; Preetha and Raveendren 2008). Expressed Sequence Tag (EST) is generated from data mining sequence information from the public databases. It offers an opportunity to identify simple sequence repeats (SSR) in ESTs by data mining. These sequences may provide an estimate of diversity in the expressed portion of the genome and may be useful for comparative mapping, for tagging important traits of interest, and for additional map-based cloning of important genes (Qureshi et al. 2004).

The present study focused on the analysis of genetic diversity among cotton genotypes in G. barbadense, using DNA-based markers (ISSR, SSR, and EST) and try to study the cotton varietal development and evolution.

Table 1. Pedigrees of the 28 cotton varieties used for molecular genetic diversity study

No.	Genotype	Pedigree*	Year of release
1	Giza 45	Giza 7 x Giza 28	1957
2	Giza 67	Giza 53 x Giza 30	1963
3	Giza 68	Menoufi x Giza 56	1963
4	Giza 69	Giza 51A x Giza 30	1966
5	Giza 70	Giza 59A x Giza 51B	1971
6	Giza 75	Giza 67 x Giza 69	1975
7	Giza 76	Menoufi x Pima S2	1980
8	Giza 77	Giza 70 x Giza 68	1982
9	Giza 80	Giza 66 x Giza 73	1981
10	Giza 81	Giza 67 x H10867/63	1983
11	Giza 83	Giza 72 x Giza 67	1990
12	Giza 84	Giza 68 x C.B.58	1986
13	Giza 86	Giza 75 x Giza 81	1995
14	Giza 87	Giza 77 x Giza 45-A	1998
15	Giza 88	Giza 77 x Giza 45-B	1997
16	Giza 89	Giza 75 x 6022	1997
17	Giza 90	Dendera x Giza 83	2000
18	Giza 92	Giza84 x (Giza74 x Giza 68)	2009
19	Giza89 x Giza86	Giza 89 x Giza 86	-
20	Giza77 x PimaS6	Giza 77 x Pima S6	-
21	Giza89 x Pima S6	Giza 89 x Pima S6	-
22	Ashmouni(Giza19)	Selected from Giza 2	1860
23	Dendera(Giza 31)	Selected from Giza 3	1951
24	Karnak (Giza29)	Maarad x Sakha 3	1939
25	Menoufi(Giza 36)	Wafeer x Sakha 3	1942
26	Pima S2	Amer.-Egy. Variety	-
27	Pima S6	Amer.-Egy. Variety	-
28	6022	-	-

* Pedigree and date of release information supported form Abdel-Salam (1999) and AlKelany (2010).

Materials and Methods

Plant material

Twenty-eight Egyptian cotton (*Gossypium barbadense* L.) genotypes (varieties and hybrids) were used for this study. Seed material was provided by the Cotton Research Institute, Sakha Research Station, Agricultural Research Center, Egypt (Table 1).

DNA isolation

Cotton seeds were grown in the greenhouse for 10 days, and leaves of seedlings were collected and grinded in liquid nitrogen using pestle and mortar. About 0.5 g of the grinded tissue was transferred in 1.5 mL sterilized Eppendorf tube. DNA isolation and purification was carried out using modified cetyl-tetramethyl ammonium bromide (CTAB) method (Dellaporta et al. 1983).

EST analysis

Five cotton specific EST primer pairs were used to perform the EST analysis (Table 2) according to the literature (Hussein et al. 2007; Xiao et al. 2009). The PCR amplification reactions were achieved in a 25-µL volume using 50 ng DNA containing 0.3 µM of each primer, 200 µM of dNTPs, 5 µL (1X) of *Taq* polymerase buffer, 2 mM $MgCl_2$, and 0.5 U *Taq* DNA polymerase. The reactions were carried out using Touchdown PCR program. The main program was performed for 7 cycles at 94ºC for 1 min, 56ºC for 1 min, decreasing 1ºC in every cycle, and 72ºC for 1 min, followed by 28 cycles at 94ºC for 1 min, 49ºC for 1 min, and 72ºC for 1 min. The previous programs were preceded by a denaturation step at 94ºC for 5 min and followed by an extension step at 72ºC for 5 min. PCR products were separated on 1.5% agarose gel electrophoresis.

SSR analysis

Six primer pairs specific for cotton microsatellite (SSR) were selected to carry out the SSR analysis (Table 2) according to the

Table 2. Primer names, sequences, number of amplified and polymorphic bands and the polymorphic information content (PIC) generated by EST, SSR, and ISSR markers in cotton genotypes

Primer name	Sequence (5' - 3')	No. of bands	Polymorphic bands	% Polymorphic	PIC
EST Primers		28	24	84.3	
NAU3442	F: TTTCCAAGACCAGTTCTTCC R: GACCAAAGGAGGTGCTCTTA	2	2	100	0.50
NAU3401	F: ATGCCGACGCTTTAAGTAAC R: CGATATGGGCATGTTTGATA	6	5	83.3	0.83
NAU3665	F: CAGCATGGAAATCCTAATCC R: TGAACTAGCTTGGCTGAATG	5	4	80	0.80
C3	F: CATCATGGTTTCCGTTTTT R: CCAGGATTGGTAAACCCGAT	3	2	66.7	0.45
E10	F: AAGAGTACAACCGCCACGG R: GAAAGGCCGAAGGATAGAGC	12	11	91.7	0.92
SSR Primers		52	44	85	
L11	F: AAAAACCCCTTTCCATCCAT R: GGTGTCCTTCCCAAAAA	9	6	66.7	0.81
M8	F: GGCATCTTACGGTGGAAATGAC R:GTTAGGTTTGGGGTGTTACATAC	10	9	90	0.90
M11	F: TGGACTAACCTAAACTTGACAC R: CCTATGTACATATGCTCTTC	5	4	80	0.80
C2-0109	F: GTGAAAACCCGCAAAG R: ATACCTAGTATTGCCCTTAT	9	8	89	0.89
C2-0119	F: GGTCCTTTTCGTCCTT R: GGTATAAATATAATGATGGT	6	6	100	0.83
SSR3	F: GCACTCGAAGGAATTAATTTT R: GAACAGTTGTTTCGTGTCGTA	13	11	84.6	0.91
ISSR Primers		47	43	90.2	
Cot 1	5'-(GA)7T-3'	4	3	75	0.75
Cot 2	5'-(GT)7CA-3'	7	7	100	0.86
Cot 3	5'-(AGC)5GA-3'	21	19	90.5	0.94
Cot 4	5'-(CA)7AT-3'	7	6	85.7	0.86
Cot 5	5'-(GT)7TA-3'	8	8	100	0.84

literature (Hussein et al. 2007; Xiao et al. 2009). The PCR amplification reactions were performed using 50 ng DNA at a 25-µL volume reaction containing 0.3 µM of each primer, 200 µM of dNTPs, 5 µL (1X) of *Taq* polymerase buffer, 1.5 mM $MgCl_2$, and 0.5 U *Taq* DNA polymerase. The SSR reactions were carried out using Touchdown PCR program. The main program was: 9 cycles at 94ºC for 1 min, 54ºC for 1 min, decreasing 1ºC in every cycle, and 72ºC for 1 min, followed by 28 cycles at 94ºC for 1 min, 45ºC for 1 min, and 72ºC for 1 min. The previous cycles were preceded by a denaturation step at 94ºC for 5 min and followed by an extension step at 72ºC for 5 min. PCR products were separated on 1.5% agarose gel electrophoresis.

ISSR analysis

Five SSR-anchored primers (ISSR primers) were selected in order to carry out the ISSR analysis (Table 2) according to the references (Dongre et al. 2004; Liu and Windel 2001). The 25-µL reaction volume contained 5 µL of reaction buffer (1X), 250 µM dNTPs, 2 mM $MgCl_2$, 0.35 µM of primer, 0.75 U *Taq* DNA polymerase, and 75 ng of genomic DNA. The PCR program included a denaturation step at 94ºC for 7 min, followed by 35 cycles at 94ºC for 1 min, 42ºC for 1 min, and 72ºC for 1 min, and a final extension step at 72ºC for 7 min. The PCR products were separated on 1.5% agarose gel electrophoresis.

Statistical analysis

All gels of the different molecular markers were scored as 0/1 for absence/presence of the bands, respectively. The total number of bands and the number of polymorphic bands were calculated as well as the polymorphic information content (PIC) which was calculated according to Anderson et al. (1993) using the following simplified formula:

$$PIC_i = 1- \Sigma p^2_{ij}$$

where p_{ij} is the frequency of the jth allele for marker ith summed across all alleles for the locus. Similarity coefficient matrices were calculated for all the markers (mixed together) using simple matching similarity algorithm (Sokal and Sneath 1963). Phylogenetic dendrogram was constructed using the UPGMA method (Unweighted Pair-Group Method with arithmetical algorithms Averages; Sneath and Sokal 1973). Principal coordinate analysis (PCOORDA) was also performed on the basis of the distance matrices using the standardized centered data in the NTSYS PC2.1 software (Rohlf 1998). The molecular data were standardized through NTSYS PC2.1 software, similarities matrices were calculated using SimInt option, decentered, and then the eignvactors and the eigenvalues were calculated using the ordination option in NTSYS PC program and then the three-dimensional diagram was obtained.

Results and Discussion

Polymorphic bands and PIC

Polymorphism analysis as detected by EST Twenty-eight bands were generated from all the EST primer pairs, 24 of them were polymorphic and representing of 85.7% of the total generated bands with an average 4.8 polymorphic bands per primer pair (Table 2). The total number of bands per primer ranged from two to 12 bands for the primer pairs NAU3442 and E10, respectively, with an average of 5.6 per primer pair. The primer pair NAU3442 gave the highest percentage of the polymorphic bands (100%) while the primer pair C3 produced the lowest percentage of polymorphic bands (66.7%; Table 2). The polymorphic information content of the EST primers ranged from 0.5 for the primer pair NAU3442 to 0.92 for the primer pair E10 (Table 2). These results are higher a double than that obtained by Hussein et al. 2007 who used 17 EST primer pairs to investigate genetic diversity among 11 cotton genotypes. They showed that the number of bands per primer pairs ranged from 2 to 5 with an average of 2.8, while the number of polymorphic bands varied from 1 to 4 and the average level of polymorphism was 70.2%. Likewise, Adawy et al. (2006) used 10 EST-SSR and seven EST primer pairs for genetic diversity estimation in 14 cotton genotypes. They found 118 bands, among which 76 were polymorphic (64.4%). The number of bands per primer pairs ranged from 1 to 11 with an average of 3.2, while, the number of polymorphic bands ranged from 1 to 5 with an average of 2 bands per primer pair.

Polymorphism analysis as detected by SSR Out of the 52 bands generated from the SSR primer pairs, 44 bands were polymorphic accounting for 84.6% of the total number of generated

bands with an average of 7.3 polymorphic bands per primer pair. The total number of bands generated from each primer pair was between five to 13 bands for primer pairs M11 and SSR3, respectively with an average of 8.7 bands per primer pair, while the polymorphic bands percentage ranged from 66.7% for the primer pair L11 to 100% for the primer pair C20119 (Table 2). The PIC of the SSR primer pairs ranged from 0.80 for the primer pair M11 to 0.91 for the primer pair SSR3 (Table 2). The average number of the polymorphic and the total number of bands per primer pair was higher than the results obtained by Kalivas et al. (2011), they analyzed 29 cultivars of *Gossypium hirsutum* and an interspecific hybrid (*G. hirsutum* x *G. barbadense*) using 12 pairs of SSR markers. They observed that two to four different alleles were amplified at each genomic locus, with a mean of 2.53 alleles per locus. Furthermore, Dongre et al. (2007) found that 17 out of the 25 microsatellite markers produced a total of 56 polymorphic bands, four markers were monomorphic and the remaining four produced non-scorable and non-reproducible bands. Moreover, Bertini et al. (2006) characterized 53 cultivars using 31 pairs of SSR primers; they obtained a total of 66 alleles with an average of 2.13 alleles per SSR locus. Similarly, Gutierrez et al. (2002) used 60 pairs of polymorphic primers amplifying 69 loci which resulted in 139 alleles with an average of two alleles per locus. However, Liu et al. (2000) used 56 polymorphic primer pairs to amplify 62 cotton loci and produced a total of 325 alleles with an average of 5 alleles per locus. Khan et al. (2009) employed 34 of 57 SSR primer pairs screened that displayed polymorphism and 122 (60%) of the 204 SSR bands detected by these polymorphic primer pairs were polymorphic across the cultivars. The number of polymorphic alleles detected per primer pair ranged from one to eight with an average of 3.6 alleles per primer pair. Buteler et al. (1999) claimed that the multi-locus amplification of the SSR is common in species with allopolyploid origin. Fisher and Bachman (1998) pointed out that SSR polymorphism is thought to be related to the number of repeats.

Polymorphism analysis as detected by ISSR The ISSR markers gave the highest percentage of polymorphic bands compared with the other molecular markers used in this study; whereas 91.4% of the total numbers of bands generated using the ISSR markers were polymorphic (43 out of 47 bands). The highest number of generated bands was obtained from the primer Cot3 (21 bands) and the lowest number of generated bands was obtained from the primer Cot1 (four bands; Table 2). The percentage of polymorphic bands for the ISSR primers ranged from 75% for the primer Cot1 to 100% for both Cot2 and Cot5 primers. The PIC of the ISSR primers ranged from 0.75 for the primer Cot1 to 0.94 for the primer Cot3 (Table 2).

In this study, it can be noticed that ISSR markers give higher polymorphic bands and polymorphic information content than both SSR and EST (Table 2). This may be due to the behavior of both SSR and EST markers whereas they are co-dominant markers and represent the repetitive DNA in the genome which is not expressed in most of the cases. On the other hand, ISSR is a dominant marker and represents the distance between two

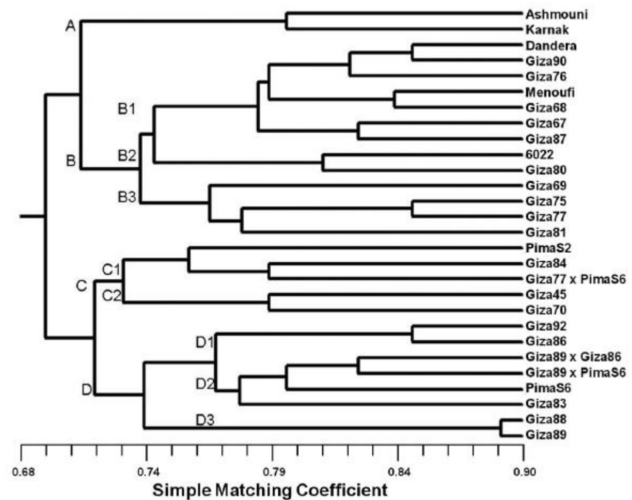

Fig. 1. Cluster analy5sis of 28 Egyptian cotton genotypes based on EST, SSR, and ISSR markers using the simple matching algorithm and the UPGMA clustering method.

microsatellites (SSR). The chance of having polymorphism in this distance between the microsatellites seems to be higher than in the SSRs and ESTs. The high incidence of detectable polymorphism through changes in repeat numbers is caused by an intramolecular mutation mechanism called DNA slippage (Gupta et al. 1996). The variation in the number of bands amplified by different primers influenced by variable factors such as primer structure, template quantity, and lower number of annealing sites in the genome (Kernodle et al. 1993).

Cluster analysis

According to the cluster analysis of the all-molecular data combined, all 28 genotypes used in this study were separated into four clusters using all the data generated from the three different molecular markers (ISSR, SSR, and EST markers; Fig. 1). Cluster number A was located at the uppermost part of the dendrogram and contained both ‘Ashmouni’ and ‘Karnak’ varieties. Cluster B was divided into three subclusters, subcluster B1 included ‘Dandera’, ‘Giza 90’, ‘Giza 76’, ‘Menoufi’, ‘Giza 68’, ‘Giza 67’, and ‘Giza 87’ varieties. Subcluster B2 included the strain ‘6022’ along with the variety ‘Giza 80’, where the subcluster B3 contained four varieties (‘Giza 69’, ‘Giza 75’, ‘Giza 77’, and ‘Giza 81’) which were clustered along with the variety ‘Giza 69’ (Fig. 1). The third cluster (cluster C) was divided into two subclusters (subclusters C1 and C2). ‘Pima S2’, ‘Giza 84’, and ‘Giza 77 X Pima S6’ genotypes formed subcluster C1, while ‘Giza 45’ and ‘Giza 70’ varieties formed subcluster C2 (Fig. 1). The last cluster (cluster D) was divided into three subclusters (D1, D2, and D3). Two varieties formed the first subcluster D1 (‘Giza 92’ and ‘Giza 86’) two another varieties formed the third subcluster D3 (‘Giza 88’ and ‘Giza 89’, Fig.1). The second subcluster (subcluster D2) contained the genotypes ‘Giza 89 X Giza 86’, ‘Giza 89 X Pima S6’, ‘Pima S6’, and ‘Giza 83’. In general, the two most related genotypes were ‘Giza 88’ and ‘Giza 89’. There was a common ancestor for these two varieties may explain why they are related to each other. This ancestor is the ‘Giza 36’ variety, which passed its genetic background through

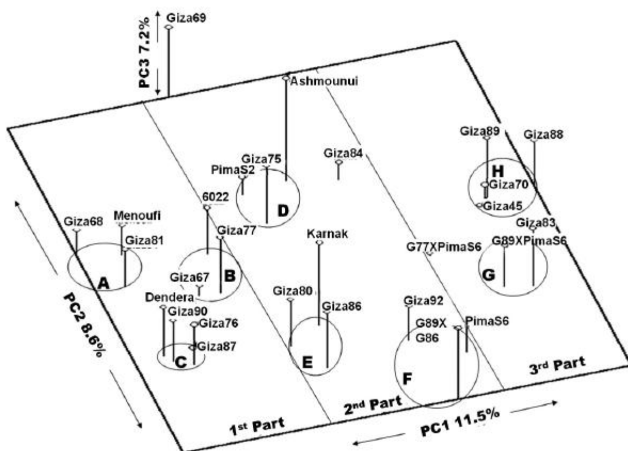

Fig. 2. Three-dimensional PCOORDA analysis for cotton genotypes obtained by molecular markers.

many generations of selections and may reflect its dominance habit. 'Menoufi' and 'Giza 68' varieties also were related each to other ('Menoufi' variety is the mother parent of Giza 68 variety according to the pedigree history, Table 1), then 'Dandera' and 'Giza 90' varieties ('Dandera' variety is the father parent of 'Giza 90' variety, Fig. 1). Unlike in the pedigree history, 'Giza 88' and 'Giza 87' varieties were clustered separately in different clusters although they generated from the same cross. These two varieties are different also in the morphological and technological traits, even in the color of the fiber. It can be said that no general clustering according to the pedigree history of the genotypes was observed (Fig. 1). In this regard, Zhang et al. (2005) found 'DP555BR' and 'DP449BR' shared cultivar 'DP5690' in their pedigree but they were grouped separately and they concluded that pedigree information or geographic origins of cultivars may not accurately reflect genetic relatedness among genotypes, whereas DNA markers could better reveal the genotypic relationships when there are sufficient markers and they are distributed across all chromosomes.

Principal Coordinate analysis (PCOORDA)

Using the principal coordinate analysis (PCOORDA), cotton genotypes were separated by the first three principal coordinates into seven groups (nominated A, B, C, D, E, F, G, and H, Fig. 2). These groups were distributed on the basis of the first principal coordinate in three parts according to their varietal development level. Nearly one third of the total genetic variance (27.3%) was represented by the first three principal coordinates PC1, PC2, and PC3, accounting for 11.5, 8.6, and 7.2%, respectively, which reflect the reliability of this analysis for both genetic diversity and varietal development studies.

The cotton genotypes were divided into three parts according to their representative level of PC1 (1st part, 2nd part, and 3rd part, Fig. 2). The 1st part represented a low level of PC1 and different levels of PC2. This part included three groups of Egyptian cotton genotypes according to both PC1 and PC2 (A, B, and C groups, Fig. 2). Group (A) contained three varieties 'Menoufi', 'Giza 68', and 'Giza 81'. The second group (B) also included three varieties, 'Giza 77', 'Giza 67', and the strain '6022', while the third group (C) contained the genotypes 'Dandera', 'Giza 90', 'Giza 76', and 'Giza 87'. It can be noted that 'Menoufi' variety is the mother parent of the variety 'Giza 68' in group A. The same can be also noticed for the 'Giza 90' variety, which was aggregated with its parent 'Dandera' in group C. These three groups are located in the first part of the first principal coordinate (PC1) and this may reflect the genetic base of development of these varieties. When we searched the ancestors of these varieties, we were surprised to find that all these varieties have a common ancestor selected from the father of the Egyptian cotton 'Ashmouni', called 'Giza 12'. This ('Giza 12') ancestor was distributed in Egypt as a variety for a period of time (from 1934 until 1944) before it was excluded for different reasons (Abdel-Salam 1999). Most of the varieties present in this part represent moderate levels of the PC3 except for the 'Giza 67' variety which represents a low level of PC3 and 'Giza 87' variety which represents zero level of that PC (Fig. 2).

The second part of the PC1 represented a moderate level of PC1 and different levels of PC2. Depending upon the PC2, it can be said that this part contained three groups (D, E, and F groups). The first group (group D) included the varieties 'Ashmouni', 'Pima S2', and 'Giza 75', while group E contained 'Karnak', 'Giza 80', and 'Giza 86' varieties. The last group (group F) included 'Giza 92', 'Pima S6', and 'Giza 89 X Giza 86' genotypes (Fig. 2). These three groups are located at the middle of the first PC and thus their varieties have a common ancestor (i.e. 'Ashmouni' variety). Of course, most of the Egyptian cotton varieties were developed from the 'Ashmouni' variety that originated and was distributed as a cotton cultivar in 1860 (Abdel-Salam 1999), but the members of these groups have a direct connection through the ancestors with the 'Ashmouni' variety. 'Giza 69' variety was located at the highest level of the PC2 in this part while 'Giza 84' variety was located near group D (Fig. 2). These two varieties were not grouped with any of the three above-mentioned groups. All members of this part have a comparable level of the third PC (moderate level) except for 'Giza 69', 'Karnak', and 'Ashmouni' which have a high level of PC3 (Fig. 2).

The third part of the PC1 included two groups (G and H groups, respectively). The genotypes 'Giza 83' variety and 'Giza 89 X Pima S6' hybrid formed Group G at a high level of the PC1 while 'Giza 89', 'Giza 88', 'Giza 70', and 'Giza 45' varieties formed group H at a very high level of PC1 and a moderate level of PC2 (Fig. 2). The members of this part have two common ancestors, 'Sakha 4' which is a common ancestor in 'Giza 83', 'Giza 88', and 'Giza 89'. At the same time, both 'Sakha 3' and 'Sakha 4' (the first variety was released in 1924 and the second was released in 1929 according to Abdel-Salam (1999) are common ancestors for 'Giza 70', 'Giza 45', 'Giza 88', and 'Giza 89' varieties. The genotype 'Giza 77 X Pima S6' represents a highly moderate level of PC1 and a moderate level of PC2 and was not grouped in any of the two groups of this part. All the members of this part represent a moderate level of PC3 except for the cross 'Giza 77 X Pima S6' which represents zero level of PC3 (Fig. 2).

Cluster analysis refers to a group of multivariate techniques

whose primary purpose is to group individuals or objects based on the characteristics they possess, so that individuals with similar descriptions are mathematically gathered into the same cluster" (Hair et al. 1995). The resulting clusters of individuals should then exhibit a high internal homogeneity (within cluster) and a high external heterogeneity (between clusters). On the other hand, principal coordinate analysis (PCOORDA) is a scaling or ordination method that starts with a matrix of similarities or dissimilarities between a set of individuals and aims to produce a low-dimensional graphical plot of the data in such a way that distances between points in the plot are close to original dissimilarities (Mohammadi and Prasanna 2003). Principal coordinate analysis is usually used to reduce the huge amount of data in groups and to obtain a new set of uncorrelated variables which are known as PCs. The first PC summarizes most of the variability present in the original data relative to all remaining PCs. The second PC explains most of the variability not summarized by the first PC and is uncorrelated with the first, and so on. Mohammadi and Prasanna (2003) reported that principal component analysis and principal coordinate analysis (the ordination methods) can be used in combination with cluster analysis for genetic diversity determination purposes, particularly when the first two or three PCs explain more than 25% of the genetic variation. In the present study, the first three principal coordinates proposed 27.3% of the total genetic variance for the molecular data. That means that principal coordinate analysis can be strongly used either alone or in combination with cluster analysis to discuss the genetic diversity in the cotton genotypes used. The results of the PCOORDA also showed better resolution of the genetic diversity than cluster analysis especially in the illustration of varietal development of cotton in this study. Melchinger (1993) reported that the ordination methods provided a faithful portrayal of the relationships between major groups of maize and barley lines compared to the cluster analysis.

References

Abdel-Salam ME. 1999. The Egyptian Cotton: Production, Quality and Marketing. Elkalema Press, 4th edn, Ahmed Barada St., Giza-Cairo, Egypt

Adawy SA, Ebtissam HA, El-Itriby AH. 2006. Molecular characterization and genetic relationships among cotton genotypes. Arab J. Biotech. 9: 477- 492

Alkelany MA. 2010. Genetical and Varietal Purity of Egyptian Cotton. Third Part: The Genetic Origins of Egyptian Cotton Varieties. Academic Press, Cairo

Anderson JA, Sorrels ME, Tanksley SD. 1993. RFLP analysis of genomic regions associated to preharvest sprouting in wheat. Crop Sci. 33: 453-459

Bertini CHCM, Schuster I, Sediyama T, Barros EG, Moreira MA. 2006. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet. Mol. Biol. 29: 321-329

Buteler MI, Jarret RL, La Bonte DR. 1999. Sequence characterization of microsatellite in diploid and polyploid *Ipomea*. Theor. Appl. Genet. 99: 123-132

Chen ZJ, Schelffler BE, Dennis E, Triplett BA, Zhang TZ et al. 2007. Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol. 145: 1303-1310

Cornelius BK, Sneller CH. 2002. Yield and molecular diversity of soybean lines derived from crosses between Northern and Southern elite parents. Crop Sci. 20: 187-190

Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA minipreparation version II. Plant Mol. Biol. Rep. 1: 19-21

Dongre AB, Bhandarkar M, Jee SB. 2007. Genetic diversity in tetraploid and diploid cotton using ISSR and SSR DNA markers. Ind. J. Biotechnol. 6: 349-353

Dongre A, Parkhi V, Gahukar S. 2004. Characterization of cotton germplasm by ISS, RAPD markers and agronomic values. Ind. J. Biotechnol. 3: 388-393

Endrizzi JE, Turcotte EL, Kohel RJ. 1985. Genetics, cytology, and evolution of Gossypium. Adv. Genet. 23: 271-375

Fisher D, Bachman K. 1998. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). Biotechniques 24: 796-802

Gupta PK, Balyan HS, Sharma PC, Ramesh B. 1996. Microsatellites in plants: A new class of molecular markers. Curr. Sci. 70: 45-54

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC. 2002. Genetic Distance among Selected Cotton Genotypes and Its Relationship with F2 Performance. Crop Sci. 42: 1841-1847

Hair JE, Anderson RE, Tatham RL, Black WC. 1995. Multivariate Data Analysis, MacMillan, New York

Hussein EHA, Osman MHA, Hussein MH, Adawy SS. 2007. Molecular characterization of cotton genotypes using PCR-based markers. J. Appl. Sci. Res. 10: 1156-1169

Hutchinson JB, Silow RA, Stephens SG. 1947. The Evolution of *Gossypium*. Oxford Univ. Press, London

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

Kernodle SP, Cannon RE, Scandalios JG. 1993. Concentration of primer and template qualitatively affects product in RAPD-PCR. Biotechniques 1: 362-364

Khan AI, Fu YB, Khan IA. 2009. Genetic diversity of Pakistani cotton cultivars as revealed by SSR markers. Inter. J. Fac. Agric. Biol. 4: 21-30

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

Liu S, Saha S, Stelly D, Burr B, Cantrell RG. 2000. Chromosomal assignment of microsatellite loci in cotton. J. Hered. 91: 326-332

Lukonge E, Herselman L, Labuschagne MT. 2007. Genetic diversity of Tanzanian cotton (*Gossypium hirsutum* L.) revealed by AFLP analysis. Afr. Crop Sci. Conf. Proc. Vol. 8. pp. 773-776

McCouch SR, Temnykh S, Lukashova A, Coburn J, DeClerck G, et al. 2001. Rice Genetics IV, In GS Khush, DS Brar, B

Hardy, eds, Science Publishers, Inc., New Delhi, pp. 117-135
Melchinger AE. 1993. Use of RFLP markers for analyses of genetic relationships among breeding materials and prediction of hybrid performance. In DR Buxton, Proceedings of the International Crop Science Congress, 1st, Ames, IA. July 1992. CSSA, Madison, pp. 621-628
Mohammadi SA, Prasanna BM. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci. 43: 1235-1248
Preetha S, Raveendren TS. 2008. Molecular marker technology in cotton. Biotechnol. Mol. Biol. Rev. 3: 32-45
Qureshi SN, Saha SR, Kantety V, Jenkins JN. 2004. EST-SSR: A new class of genetic markers in cotton. J. Cotton Sci. 8: 112-123
Rana MK, Bhat KV. 2004. A comparison of AFLP and RAPD markers for genetic diversity and cultivar identification in cotton. J. Plant Biochem. Biotechnol. 13: 19-24
Rohlf FJ. 1998. NTSYSpc. Numerical taxonomy and multivariate analysis system, version 2.02c. Exeter Software, New York
Sneath P, Sokal R. 1973. Numerical Taxonomy, vol. 12. Freeman, San Francisco, pp. 102-108
Sokal R, Sneath P. 1963. Principles of Numerical Taxonomy, WH Freeman, San Francisco
Stewart JM. 1995. Potential for Crop Improvement with Exotic Germplasm and Genetic Engineering. In GA Constable, Forrester, eds, Challenging the Future. Proc. World Cotton Res. Conf. 1 CSIRO, Melbourne, pp. 313-327
United Nations 2003. Cotton Facts. International Cotton Advisory Committee, Technical Paper No. 25 of the Common Fund for Commodities (M.R. Chaudhry and A. Guitchounts). New York, United Nations
Zhang J, Lu Y, Cantrell RG, Hughs E. 2005. Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. Crop Sci. 45: 1483-1490
Xiao J, Wu K, Fang DD, Stelly DM, Yu J, Cantrell RG. 2009. New SSR markers for use in cotton (*Gossypium* spp.) improvement. J. Cotton Sci. 13: 75-157