A multivariate procedure for the study of the behavior of current cultivars and future strains in Egyptian cotton under varying locations

Hamoud, H. M. E.; A. M. R. Abdel-Bary, and Y. A. M. Soliman
Cotton Research Institute, Agriculture Research Center , Egypt

ABSTRACT

Barbadense cotton cultivar grown in Egypt were derived from pedigree breeding system . All of the superior genotypes in this system were adapted under local condition . The objective of this study was to determine the important of genetic distance to cultivar. The choice of parent is often the most importance decision in a breeding program , little is known about the important of parental genetic distance to successful cotton cultivar development . Principle component and hierarchical (ward,s minimum variance) cluster were determine data from of five regional cotton variety tests . Seven current cultivars in delta region and seven expected future strains were used in this study . The two principal components accounted for each region no less than 70.4% . The PC1 has positive moderately correlation for most studied traits at most region . The genetic divergence among the genotypes based on Euclidean distances revealed that these entries trend to join in two groups according to breeding program especially in the region 1 which there are breeding program for extra and long staple varieties . The highly distantly-genetic was observed between Giza 87 and Giza 89 for all region except region 5. Giza 45 had unique position relative most of genotypes. Breeders need to increase variability by using a divergent and adaptability parents under local condition.

INTRODUCTION

Egyptian cotton cultivars grown in A.R.E were derived from a breeding programs depending on a pedigree method so, choice of parent in the beginning program is very important step for the success of such program. Knowledge about germplasm diversity

and genetic relationships (closely-related or distantly-related) among current cultivar and new promising strains could be an invaluable aid in improvement strategies. Diversity analysis of crop species is essential process of crop improvement programs , it serves to provide information about genetic diversity, as it facilitates the selection of potential parents for subsequent crossing and selection of progenies up to the final utilization of cultivars in production schemes. Accurate assessment of the levels and patterns of genetic diversity can be invaluable in crop breeding for diverse applications including (i) analysis of genetic variability in cultivars (Smith, 1984; Cox et al., 1986), (ii) identifying diverse parental combinations to create segregating progenies with maximum genetic variability for further selection (Barrett and Kidwell, 1998), and (iii) introgressing desirable genes from diverse germplasm into the available genetic base (Thompson *et al.*, 1998). An understanding of genetic relationships among inbred lines or pure lines can be particularly useful in planning crosses , in assigning lines to specific heterotic groups , and precise identification with respect to plant varietals protection (Hallaur and Mirand ,1988) .Analysis of genetic diversity in germplasm collection can facilitate reliable classification of accession , and identification of sub sets of core accessions with possible utility for specific breeding purposes.

The use of multivariate methods , more common in other discipline , is seen increasingly in plant breeding literature . In the analysis presented here , principal component and cluster analysis were used on agronomic trial data to give graphical presentation of relative cultivar performance , and to show interrelationships of cotton cultivars , based on agronomic performance and fiber quality measurements . The analysis also gives general overview of genetic variability among current Egyptian cotton cultivars and new lines across geographic regions of the cotton . It helps in the selection of genetically divergent parents for their exploitation in hybridization program , and measures the degree of diversification and determines proportion of each component character to the total divergence (Singh and Narayanan , 2000)

Materials and methods

Seven current cultivars , five new promising crosses and two advanced strains were grown and evaluated in five locations (environments) for yield and fiber properties . The 14 genotypes chosen represented at three groups .The first group, seven current cultivars used were Giza 87, Giza 88, Giza 70, Giza 45,Giza 85,Giza 89 and Giza 86. The second group, the cotton breeders   from two programs (Long and extra long staple) were each asked to supply 5 promising crosses that their research suggested had a good potential of replacing their current cultivars which were Giza 89 x Giza 86,Giza 89 x Pima s6 ,Giza 75 x Sea ,Giza 77 x Pima s6 and Giza 84 x   (Giza 74 x Giza 68). The third group, two advanced strains which chosen from the best strains inside breeding program ,Giza 85 x Giza 86 x Giza 89 , Giza (84 x (Giza 70 x Giza 51b)) x Pima S6 at the two breeding programs.

Data used in these analysis are from the 2005 regional cotton varieties tests at five environments (Locations) i.e: Kafr EL-Sheikh(Sakha Agriculture Research Station), Damitta, EL-Sharqia, EL-Behara and EL-Gharbia governorates. The experimental design was a randomized complete block with six replications for all five locations. Plot size was five rows, each 60 cm x 4 m. Seedlings were thinned to two plants for each hill at about 21 day after planting.

Data for yield and fiber properties were taken from a whole plot. Table 1 lists the abbreviations used for genotypes names in subsequent graphical figures and genotypes. The first column of table 2 lists the traits analyzed. The means of genotypes a cross locations and replications were used in the multivariate analysis .

Hierarchical clustering was then carried out on each data set using minimizes within-cluster sums of squares   a cross all partitions on the basis of Euclidean distances as outlined by Anderberg (1973) and developed by Hair *et al.* 1987. Principal component analysis was performed on the correlation matrix of the

Table 1 : genotypes in the five regional cotton trails , abbreviations used in the figures , and genotypes type

Current cultivars	Abbrev.	Breeding program	type	Recent genotypes	Abbrev.	Breeding program	genotypes type
Giza 87	G 1	Extra long staple	Commercial cultivar	Giza 84 x Giza 74 x Giza 68	G 8	Extra long staple	New promising strains
Giza 88	G 2	Extra long staple	Commercial cultivar	Giza 77 x Pima s6	G 9	Extra long staple	New promising strains
Giza 45	G 3	Extra long staple	Commercial cultivar	Giza84x(Giza70x Giza 51b) x Pima 62	G 10	Extra long staple	Advanced strains
Giza 70	G 4	Extra long staple	Commercial cultivar	Giza 89 xGiza86	G 11	Long staple	New promising strains
Giza 89	G 5	long staple	Commercial cultivar	Giza 89 x Pima s6	G 12	Long staple	New promising strains
Giza 86	G 6	long staple	Commercial cultivar	Giza 75 x Sea	G 13	Long staple	New promising strains
Giza 85	G 7	long staple	Commercial cultivar	Giza 89 x Giza 85 x Giza 86	G 14	Long staple	Advanced strains

traits of each trial (Johnson *et al.*, 1988) .All these computations were performed using SAS computer program.

Results and discussion

The relative magnitude of the coefficient of each trait relating it to the first two principal components from the component analysis can often provide an agronomic interpretation for each component axis table 2 . 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 uncorrelated with the first, and so on (Jolliffe, 1986). Because PCs are orthogonal and independent of each other, each PC revealed different properties of the original data and may be interpreted independently (Mohammadi and Prasanna , 2003) . As can be seen in this table that PC1 has moderately positive correlation with seed cotton yield , lint cotton yield at regions 2, 3 and 5, earliness index, and lint percentage at regions 1, 2 and 5 , micronaire and hair weight at all regions. Also moderately negative correlation with 2.5 % span length and yarn strength for all regions .

PC2 had moderately positive correlation with seed cotton yield and lint cotton yield at regions 1 and 4 and seed cotton yield in region 5 , and with stelometer and yarn strength at the regions 3 and 5 .This indicate that these traits are most moderately associated with directions of maximum variance in the data set . It also notable that PC1 had moderate correlation with most traits at most regions , which would indicate that to a certain PC1 is controlled by an over all increase in trait concentration .

The information of the percentage of variance indicated that the two principal components for each region account for no less than 70.4% and this agreed with Abd EL-Sayyed et al (2000) , while Van Esbrock et al (1999) found that six principal components accounted for 87% of the variation among cultivars. Raychaudhari *et al* (2000) found that first two principal components account for

Table 2 : Coefficients of the principal component of each agronomic traits in cotton by regions for the first two principal components and the variance of a principal component over all traits.

	Region 1		Region 2		Region 3		Region 4		Region 5	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Seed cotton yield (k/f)	0.294	0.479	0.328	−0.247	0.383	0.148	0.277	0.511	0.329	0.424
Lint cotton yield(k/f)	0.270	0.541	0.375	−0.179	0.415	0.158	0.300	0.491	0.373	0.299
Boll weight	0.194	0.371	−0.005	0.723	−0.282	0.002	0.073	0.179	−0.023	−0.231
Earliness index	0.355	0.003	0.354	−0.050	0.235	−0.427	−0.050	0.494	0.339	0.220
Lint %	0.342	0.067	0.400	0.146	0.334	0.139	0.296	0.209	0.365	0.094
2.5% span length	−0.354	0.067	−0.310	−0.384	−0.278	0.362	−0.371	0.287	−0.357	0.314
Micronire	0.352	0.276	0.336	−0.244	0.381	0.137	0.391	−0.185	0.363	−0.055
Stelometer	−0.234	0.349	0.071	0.365	0.044	0.530	−0.380	0.110	−0.159	0.551
Hair weight	0.359	−0.107	0.373	0.009	0.430	0.094	0.374	−0.146	0.349	−0.009
yern strength	−0.352	0.319	−0.332	−0.133	−0.150	0.556	−0.401	0.158	−0.309	0.461
Variance %	62.040	16.654	55.651	15.632	44.173	26.315	53.215	19.871	57.242	16.589
Total variance on first 2 PC	78.694		71.283		70.488		73.086		73.831	

Region 1 =Kafr EL-Shiekh , Region 2=Damitta, Region 3=EL-Sharqia , Region 4 = EL-Behara , Region 5= EL-Gharbia

over 90 % of the variance. The percent variance for PC1 account ranged from 44.17% in the region 3 to 62.040% in the region 1 and PC2 ranged from 15.632% in the region 2 to 26.315% in the region 3. also Brown (1991) reported that the variance of first two principal components (PC1 and PC2) ranged from 44.2% to 57.6% for all regions

Generally it is possible to include that the corresponding amount of variance in a two-dimensional plot of the components (Fig. 1). In Fig. 1, each genotype is plotted at its principal components score on each axis. The results revealed that there are relationship between each of principal components and cluster analysis which they gave a sensible and useful integration of the data, generally they appeared to be incomplete accordance. Principal components analysis is often used as a preprocessing step to clustering (Everitt 1993).

The results of the hierarchical cluster analysis in the form of dendrograms are presented in figure 2 and table 3 . As no definitive exit for determining of the number of population clusters , to illustrate relative genetic distance and genetic diversity within a given germplasm base . All of these entries were developed locally at Sakha Agricultural Research Station , Kafr EL-Shiekh and selection of these genotypes achieved throw its superior in trial B (mentaince experimental) in many environments for many years , so more than one of these entries adapted for more one location .

Also most of these entries as seen in table 3 are closely related so the genetic distance between them are small . However , for many crops , yield improvement have come from mating of closely-related genotypes (Van Esproeck, and Bowman, 1998) .Compared with other regions, show smaller genetic distance between genotypes and little variability within clusters, reflective of narrow genetic bases within these genotypes.

As seen in Fig. 1 and Fig. 2 region 1 , the genetic distance among the extra long staple genotypes were very small except for Giza 45 which is an occupied a basal position within all genotypes

Table 3 : Genetic distance for 14 extra and long staple genotypes

| Stages	Region 1			Region 2			Region 3			Region 4			Region 5			
	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Distance	Cluster 1	Cluster 2	Distanc	e
1	11	14	13.30	2	4	7.97	8	10	12.41	2	4	7.44	6	13	14.37	
2	8	10	19.04	6	12	11.94	1	4	20.02	5	7	13.7	14	12	15.95	
3	1	8	23.43	2	8	16.33	3	8	24.08	1	2	25.13	2	10	31.38	
4	1	4	27.53	2	9	26.61	7	12	30.09	5	14	29.83	2	8	41.71	
5	2	9	36.89	5	13	42.51	2	3	42.03	10	13	30.44	5	7	41.79	
6	1	2	49.07	1	2	43.08	1	9	42.73	8	11	32.73	3	6	44.34	
7	5	12	56.12	5	7	58.63	7	14	44.33	1	8	49.77	1	5	46.41	
8	11	13	67.91	1	3	65.39	5	7	64.33	9	10	52.37	2	11	48.32	
9	5	7	78.56	5	14	80.29	1	2	81.58	5	6	73.93	1	4	56.96	
10	1	3	80.92	1	10	101.35	5	11	93.83	1	3	93.27	2	9	58.16	
11	6	11	105.77	1	11	150.93	5	13	129.44	1	9	128.40	1	12	94.15	
12	5	6	184.02	5	6	202.4	5	6	180.20	5	12	161.47	2	3	110.74	
13	1	5	338.82	1	5	412.95	1	5	460.21	1	5	339.03	1	2	217.24	
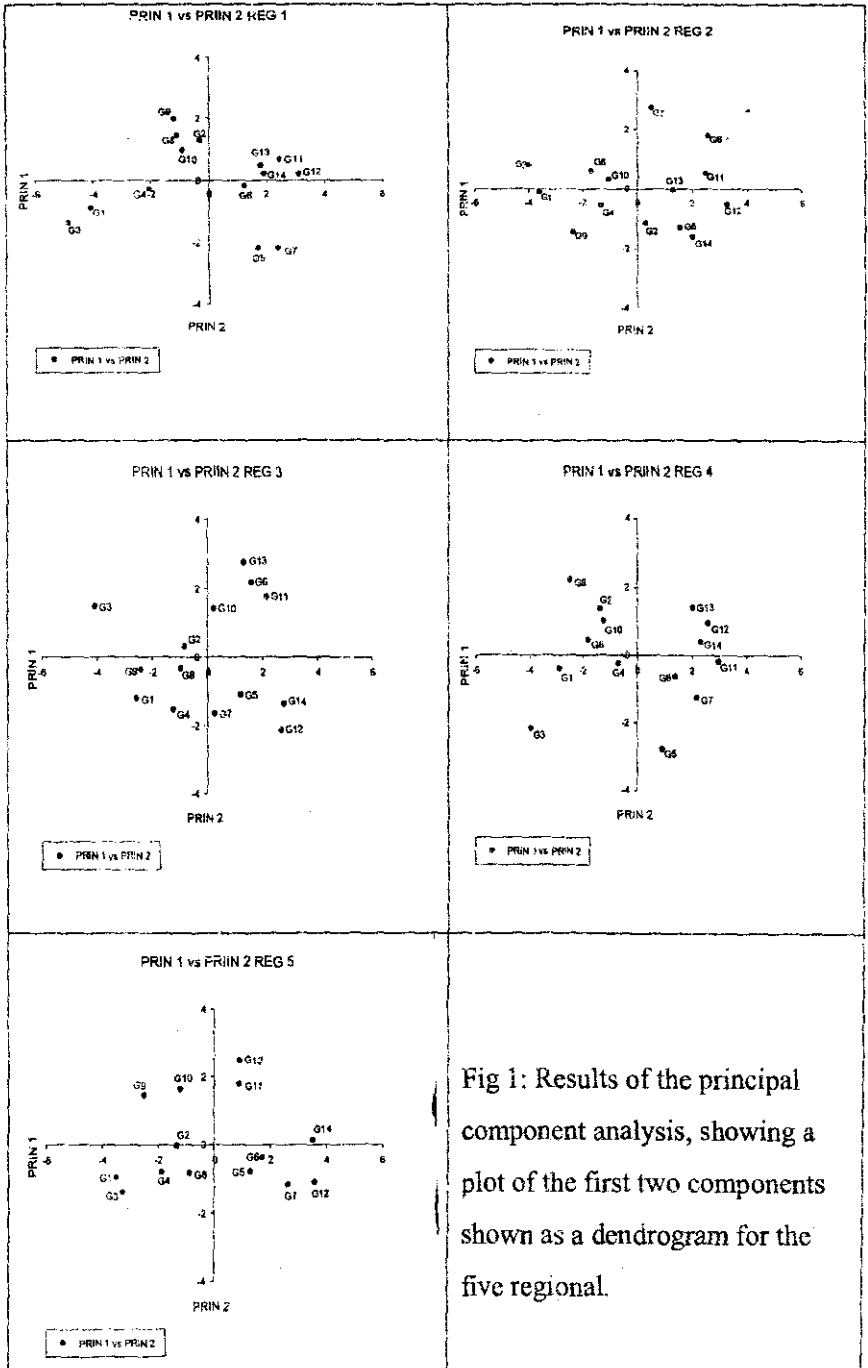
Fig 1: Results of the principal component analysis, showing a plot of the first two components shown as a dendrogram for the five regional.
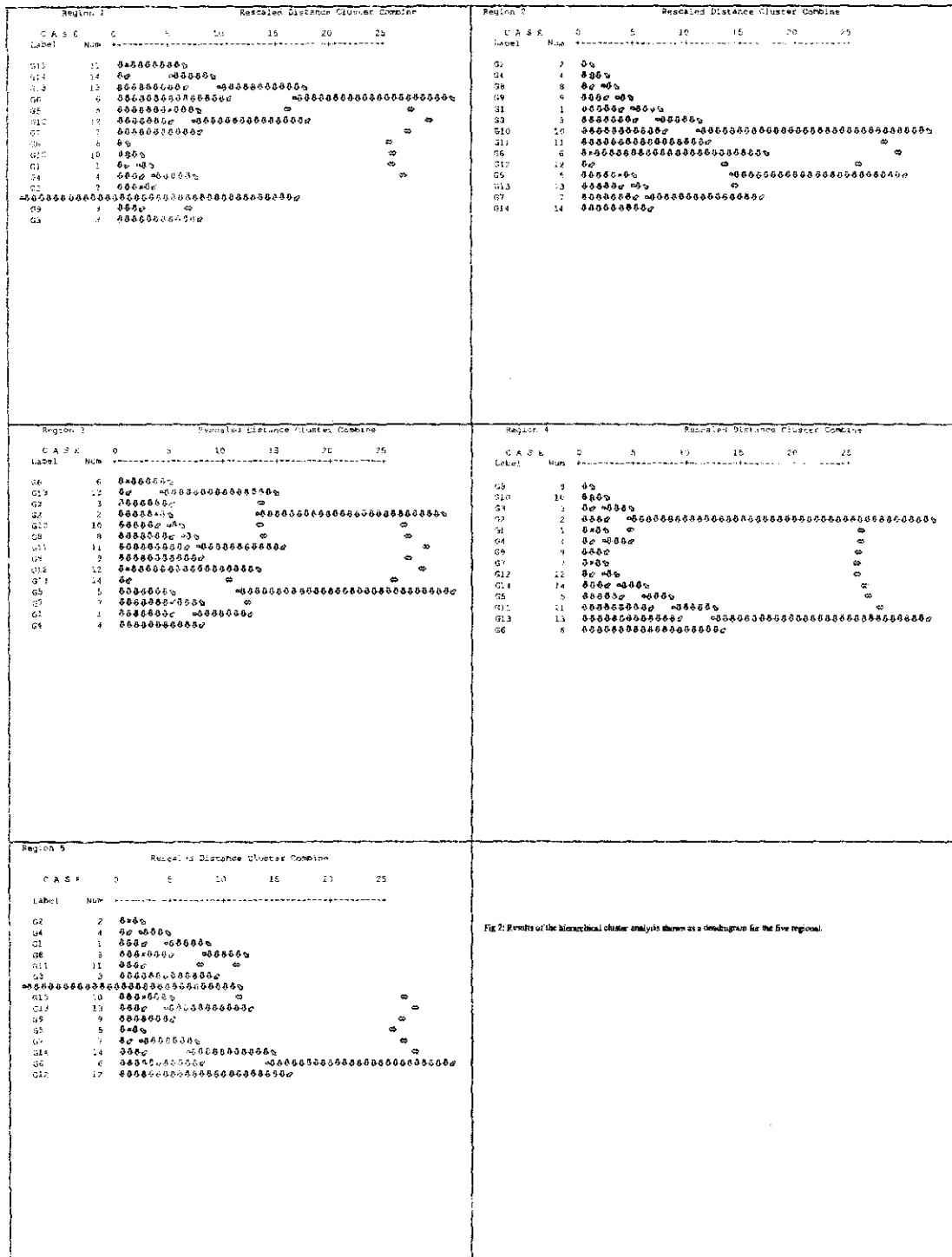
Fig 2: Results of the hierarchical cluster analysis shown as a dendrogram for the five regions.

in most regions, Giza 45 had a well-supported position at the base of the *G. barbadencse* cluster (Abdalla *et al* 2001). In this region most of long stable genotypes revealed more genetic distance and each of the two types grouped separated in two clusters and this due to that these genotypes were developed separately in two programs in this region . The closest genotypes were Giza 89 x Giza 86 with (Giza 85 x Giza 89) x Giza 86 while Giza 87 with Giza 89 had the greatest genetic distance.

The same trend was observed in region 2 except for the new promising cross Giza 89 x Giza 86 which behaved as extra long staple genotypes, while the smallest genetic distance for this group was between Giza 88 and each of Giza 70 and (Giza 84 x Giza 74) x Giza 68 and in the other group (long staple genotypes) was observed between Giza 89 x Giza 86 and Giza 89 x Pima s6, also the greatest genetic distance was noticed between Giza 87 and Giza 89. The region 3 and 5 revealed different behaved for genotypes which show interfere among most of genotypes except for Giza 89 x Pima s6, which seems to be outlier in the region 5, to which it is unadapted and both of Giza 89 x Pima s6 and Giza 86 show unique clusters in this region. Also showed the smallest genetic distance revealed between Giza 84 x   (Giza 74 x Giza 68) with (Giza 84 x (Giza 70 x Giza 51b)) x Pima 62 in region 3 and Giza 86 with Giza 75 x Sea in region 5, while the largest distance was noticed between Giza 87 with each of Giza 89 and Giza 88 in region 3 and 5 respectively.

While in the region 4 , most of the genotypes , especially extra long staple illustrated closely-related, while Giza 86 , Giza 75 x Sea , Giza 89 x Giza 86 and Giza 89 which seemed to be outlier compared with the other genotypes . This referred to most of this extra long staple were adaptation to this region, Generally all of these genotypes resulting from a pedigree method , and much of these genotypes have participate original ancestors.

The breeders should be increase the genetic base , which genetical diverse parents facilitate the creation of superior progeny .

In theory , mating of distantly-related parents will produce a greater number of transgressive segregates than mating of closely-related .Thus the recurrent mating of genotypes from different clusters in along time would be a step toward broadening the genetic base in breeding program to produce abundant of useful segregates .

Although that successful cultivars were most frequently developed from closely-related parents , with level of diversity similar to the average genetic relationship among regionally adapted cultivars from closely-related parents , indicated that their was sufficient variability or mechanisms to create variability to make breeding progress in a narrow germplasm base . The high frequency of closely-related parents in final cross of successful cultivars reflects the fact that new cultivars were for most part developed from high-yielding , closely-related , locally-adapted cultivars , (Van Esbroeck and bowman 1998)

REFERENCES

A bdalla, A.M., O.U.K. Reddy, K.M. EL-Zik, A.E. Pepper, 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet. 102:222-229.

Abd EL-Sayyed , S. M. , A. R. Abo-Arab and Y. M. EL-Mansy 2000. Genetical studies on off types of some Egyptian cottons . J. Agric. Sci. Mansoura Univ0., 25(11):6643-6657,2000

Anderberg ,M.R. 1973.Cluster analysis for applications academic press, New York

Barrett, B.A., and K.K. Kidwell. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 38:1261–1271.

Brown J. S. 1991. Principal component and cluster analyses of cotton cultivar variability a cross the U S cotton Belt . Crop Sci .31:915-922.

Cox, T.S., J.P. Murphy, and D.M. Rodgers. 1986. Changes in genetic diversity in the red winter wheat regions of the United States. Proc. Natl. Acad. Sci. (USA) 83:5583–5586.

Everitt, B.S. 1993. Cluster Analysis. John Wiley & Sons, New York, NY.

Hair, J. F. Jr., R. E. Anderson and R. L. Tatham. 1987. Multivariate data analysis with reading . Macnillan publ. Co., New York.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative genetics in maize breeding. 2nd edition, Iowa State University Press, Ames, IA.

Johanson, R.A., and D.W. Wichern. 1988. A pplicd multivariate staqtistical analysis. 2nd . ed Prentice-Hall, Englewood Cliffs, NJ.

Jolliffe, I.T. 1986. Principal component analysis. Springer-Verlag, Berlin.

Mohammadi S. A. and B. M. Prasanna.2003.   Analysis of genetic diversity in crop plants—Salient Statistical Tools and Considerations. Crop Sci 43:1235-1248.

Raychaudhuri, S., J. M. Stuart, and R. B. Altman 2000. Principal 38:1348–1355

SAS Institute . 1985 . SAS user's guide: Statistics. Version 5 ed. SAS Inst., Cary NC.

Singh P. and S.S Narayanan 2000. Biometrical Techniques In Plant Breeding. Klyani publishers, New Delhi, 2nd ed.

Smith, J.S.C. 1984. Genetic variability within U.S. hybrid maize: Multivariate analysis of isozyme data. Crop Sci. 24:1041–1046.

Thompson, J.A., R.L. Nelson, and L.O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci.   38:1348–1355.

Van Esbroeck G. A., D. T. Bowman 1998.Cotton germplasm diversity and its importance to cultivar development .the journal of Cotton Sci .2:121-129

Van Esbroeck G. A. , D. T. Bowman ,O. L. May and D.S. Calhoun. 1999. Genetic similarity Indices for Ancestral cultivars and their impact on genetic diversity estimates of modern cultivars .Crop Sci . 39.323-328

الملخص العربي

إجراء مقاييس التباينات المتعددة لدراسة سلوك الأصناف الحالية والسلالات المستقبلية في القطن المصري تحت مواقع مختلفة

هشام مسعد السيد حمود– عبد الناصر محمد رضوان عبدالبارى – ياسر عبدالرؤوف محمد سليمان

معهد بحوث القطن– مركز البحوث الزراعي

تم إجراء هذه الدراسة في موسم ٢٠٠٥ واستخدم فيها ١٤ تركيب وراثي عبارة عن ٧ أصناف تجارية و ٧ سلالات يتوقع أن تحل محـل الأصناف التجارية وتم تقيمها في خمس مواقع بهدف فهم طبيعة العلاقـة والمسافة الوراثية بين هذه التراكيب الوراثية وذلك لتحديد أنسبها كمصدر للتباين الوراثي وقد أظهرت النتائج أن تحليل مكونات التباين الرئيسية كانت أكبر من ٧٠,٤ % وكانت معظم الصفات هامة وذات تأثير معتدل للمكون الوراثي الأول وقد أظهرت دراسة التباعـد السـوراثي أن تلـك التراكيب الوراثية تميل إلى أن تتواجد في مجموعتين (مجموعة الأقطان طويلة وجه بحري ومجموعة الأقطان فائقة الطول) خاصة في المنطقـة الأولى حيث يتواجد فيها كلا البرنامجين، كما نلاحظ أن المسافة الوراثية بين التراكيب الوراثية لطبقة فائقة الطول اقل بالمقارنة بطبقة طويل وجه بحري وقد أظهرت أصناف وسلالات طبقة فائق الطول تقـارب شـديد خاصة في المنطقة ٤ مما يشير إلى أنها أكثر ملائمة لهذه المنطقـة فـي

766

J. Agric. Res. Kafer El-Sheikh Univ., 33(4) 2007

حين وجد تداخل في سلوك معظم التراكيب الوراثية في المنطقتين ٣و٥ .
وكانت أعلى مسافة وراثية بين الصنفين جيزة ٨٧ والصنف جيزة ٨٩ في
معظم المناطق. وقد اظهر الصنف جيزة ٤٥ موقع خـــاص فـــي معظـــم
المناطق. وعلى مربى القطن أن يسعى لزيادة التباين الوراثي من خلال
استخدام أباء متباعدة وراثيا ومتأقلمة تحت الظروف المصرية .