Principal component analysis as a reflector of combining abilities

Mehdi Ghaffari[1], Ebrahim Farrokhi[2]
[1]Agricultural and Natural Resources Research Station, Khoy, P.O.Box:383, Iran,
E-mail: mgh110@hotmail.com
[2]Seed and Plant Improvement Institute, Karaj, P.O.Box:31585-4119, Iran
E-mail: efarrokhi@yahoo.com

ABSTRACT

In this study Principal Component Analysis was used to reveal two dimensional structures based on general (GCA) and specific combining abilities (SCA) in a sunflower crossing program. GCA and SCA of new sunflower inbred lines accompanied by genetic variance components were estimated using Line x Tester analysis method. Six new restorer lines were crossed with four CMS lines as tester in 2006. 24 combinations were planted in randomized block design with four replications in Khoy Agricultural and Natural Resources Research Station in 2007. All traits except head diameter and plant height were under control of both additive and dominant effects. Seed yield was mainly under control of dominant effects. For the growth period, and in other traits, additive effects had a major influence. Overdominant effects for growth period and seed yield, complete dominance for plant height and partial dominance for other traits were evident. Principal component analysis revealed a close positive association between seed yield, oil yield and seed number, whereas there were negative associations between these traits and 1000 seed weight. Principal components located entries based on their whole performance in the biplots provided. Restorer line R21 and CMS line CMS356 had a strong effect on their corresponding hybrids. Ordination biplots provided simplicity of selection based on agronomic means instead of SCA data. R23 x CMS78 was a superior early matured hybrid with higher oil content and yield. We could use PCA based on agronomic data alone instead of combining abilities for determining real performance of entries considering multivariate traits.

Key words: combining ability sulfonylurea – principal component analysis – sunflower sulfonylurea – variance components.

INTRODUCTION

Sunflower is one of the most important oilseed crops of Iran but its planting area has been very variable and recently decreased to 30000ha. Breeding programs in seed and plant investigation institutes are carried out for producing new hybrids to replace open pollinated (OP) varieties. Iranian Hybrids have not been very successful to date but a new generation of F1 hybrids is now being produced. A basic research program in this field has been focused on producing parental lines and many CMS and restorers have been produced. In these programs, estimation of combining ability of lines to identify superior parents for hybridization is essential. Besides an estimation of combining abilities and identifying the type of gene action governing yield-related traits, this study was made to establish a two dimensional structure between agronomic performance and combining abilities. Predominant of dominance gene action was reported for plant height, head diameter, oil content, 100 seed weight and seed and oil yield (Gangappa et al., 1997). However, additive gene action for these traits has also been reported (Singh et al., 1989) Estimates of GCA and SCA (Bedov, 1985) indicated that additive effects were more important for oil content. Additive gene action has the greatest influence on flowering (Alvarez et al., 1992). A significant relationship between morphological differences between inbred lines and SCA effects was found for oil yield and seed number per plant (Luczkiewicz and Kaczamarek, 2004). These relationships could be used to evaluate lines directly based on agronomic values instead of on combining abilities. Biometrical methods such as principle component analysis allow us to recognize structures between genotypes based on multivariate traits. Tersac and et al. (1993) used PCA based on SCA to show a structure of sunflower populations by country of origin. De la Vega et al. (2001) used PCA for revealing two dimensional structures among genotypes and environments based on their interactions. They reported the effectiveness of PCA for demonstrating genotype x environment interactions. In this study we have investigated the efficiency of PCA in screening genotypes due to GCA and SCA of multivariate traits.

MATERIALS AND METHODS

This experiment was conducted in Khoy-Iran Agricultural and Natural Resources Research Station in 2007. The station is located in 38º 32' north latitude and 44º 58' east altitudes. Six new restorer lines were crossed with four cytoplasmic male sterile (CMS) lines in line x tester fashion to generate 24 single cross hybrids in first year and twenty four F1 hybrids planted in randomized block design with four replications in 2007. Each experimental plot consisted of 3 rows of 4 m in length with 60cm spacing between rows and 25cm within rows. Fertilizers were applied at the rate of 100:70:90 NPK kg/ha. Field practices were followed according to the regional sunflower planting handbook (Ghaffari, 2006). Data of measured traits for hybrids subjected to Line x Tester analysis (Kemptorne, 1957) to estimate general combining ability (GCA), specific combining ability (SCA), effects and their respective variances were collected. Principal Component Analysis (PCA) was used to arrange the entries in two dimensional biplots (Kroonenberg, 1997) based on agronomic performance and combining abilities.

RESULTS AND DISCUSSION

There were significant differences between crosses for all measured traits but differences between CMS lines were greater than those between restorers (Table 1).All traits except head diameter and plant height were under control of both additive and dominant effects. Seed yield was mainly under the control of dominant effects. For growth period, the dominant effects were more important than additives and in other traits additive effects had a major influence. Over dominant effects for growth period and seed yield, complete dominance for plant height and partial dominance for other traits were evident (Table 2).

Table 1. Analysis of variance for agronomic traits

Sources	D.F	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
Replication	3	44.9**	339.26	48.244**	314.2**	35628.6	1162745.4*	50.32**	182658.6
Crosses	23	32.4**	349.49**	5.445	154.7**	79052.1**	1155345.3**	18.12**	278046.6**
Lines	5	39.8**	550.8**	2.932	171.9**	34051.2	470879.6	33.48**	123211
Testers	3	68.2**	605.8*	5.653	610.33**	349693.9**	3627592.4**	58.64**	1043106**
Lines x Testers	15	22.8**	231.1	6.241	57.8*	39923.9*	889051.3**	4.89**	176646.7*
Error	69	6.6	160.9	5.373	28.7	20324.9	361540.3	7.039	86292.98
C.V.		2.42	6.86	12.47	7.42	21.73	19.29	5.87	20.80

* and ** significant at 0.05 and 0.01 level of probability respectively

Table 2. Variance components for agronomic traits

Variance	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
Additive	3.13*	34.72*	-0.19	33.33**	15194.86*	116018.49+	4.12**	40651.16*
SE	1.24	12.40	0.12	9.94	5568.92	58699.98	1.03	16712.78
Dominance	4.03**	17.53	0.22	7.28*	4899.76*	131877.70**	0	22588.43*
SE	0.99	10.47	0.29	2.55	1764.04	38865.08	0.26	7787.033934
Dominance rate	1.61	1.00	0.66*	0.66*	0.80	1.51	0.00	1.05

* and ** significant at 0.05 and 0.01 level of probability respectively

Restorers R50, R21 and R23 had the highest GCA for seed yield but none of them was significant (Table 3). R21, R23 and R56 had significant positive GCA for 1000 seed weight, seed number and oil content, respectively. This indicates that these restorers seemed to possess increasing alleles with additive effects for the mentioned traits. R26 and R50 had significant negative GCA for growth period indicating presence of alleles with additive effects for early maturity. Therefore, R26 seemed to possess additive alleles for dwarfness. Single branch restorer RG50 distinguished itself as being a good line for using in crossing programs because of having the desired GCA for seed yield and growth period.

Testers CMS52 and CMS148 had the highest GCA for seed yield but only CMS52 had significant GCA in the desired (positive) direction (Table 4). CMS 356, CMS52 and CMS78 lines had significant positive GCA for 1000 seed weight, seed number and oil content, respectively. CMS356 seemed to have alleles with additive effects for increasing the growth period, while CMS52 had alleles for decreasing it. It would seem that CMS52, because of having the desired GCA for seed yield and growth period, could be used as a valuable A-line in crossing programs.

Table 3. GCA of restorer lines for agronomic traits

Restorer	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
R19	0.07	-6.11	-0.32	-0.62	-50.91	-254.66	-1.28*	-158.63*
R21	-0.61	7.96*	0.58	-3.37	72.74*	114.51	-0.34	44.43
R23	2.51**	1.97	-0.43	5.41**	-28.18	82.43	1.14	82.07
R26	-1.55*	-7.72*	0.39	0.94	-31.40	-112.57	0.29	-41.77
R50	-1.49*	1.54	0.13	0.94	30.39	217.85	-1.82*	36.76
R56	1.07	2.37	-0.35	-3.31	7.38	-47.57	2.01**	39.20
SE	0.59	2.90	0.53	1.22	32.54	137.22	0.61	67.04

* and ** significant at 0.05 and 0.01 level of probability respectively

Table 4. GCA of CMS lines for agronomic traits

CMS	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
CMS78	-0.70	2.19	-0.35	-2.41	21.10	23.96	1.31*	54.65
CMS52	-1.32*	3.47	0.69	-2.05	87.16**	318.68*	0.60	154.93*
CMS148	-0.45	-7.46**	-0.33	-3.09	67.96	211.18	0.35	97.01
CMS356	2.47**	1.81	-0.01	7.53**	-176.21	-553.82**	-2.26	-306.57**
SE	0.46	2.24	0.41	0.95	25.20	106.29	0.47	51.93

* and ** significant at 0.05 and 0.01 level of probability respectively

Hybrids R23 x CMS78 and R56 x CMS356 had high SCA for both seed yield and growth period (Table 5). These hybrids had seed yield of 4105 and 3022 kg/ha, respectively, and growth period of 107 days (Table 6). Crossing of R21, R26, and R50 with CMS52 and R21 with CMS148 resulted in higher seed yield (over 3500 kg/ha) with short growth period (107 days) and makes them high yielding early maturing hybrids for summer cropping.

Table 5. SCA of crosses for agronomic traits

Restorer	CMS	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
R19	CMS78	-0.11	-1.51	1.18	1.13	-79.13	-308.12	-1.38	-172.54
R19	CMS52	-0.99	10.06*	-0.91	0.52	18.04	60.49	0.86	50.80
R19	CMS148	3.64**	1.31	0.53	1.94	59.81	406.32	-0.79	152.33
R19	CMS356	-2.53*	-9.89	-0.79	-3.56	1.28	-158.68	1.32	-31.98
R21	CMS78	-1.93	-3.96	-1.02	-7.25	-23.97	-403.96	1.46	-139.56
R21	CMS52	0.95	-2.46	1.92*	-1.60	149.51*	649.66*	-1.12	250.58*
R21	CMS148	0.57	2.09	-1.54	2.94	30.39	257.16	-0.46	109.74
R21	CMS356	0.41	4.30	0.64	5.94*	-155.94*	-502.85*	0.12	-222.15
R23	CMS78	-2.05*	-7.74	0.65	2.35	145.60*	881.47**	0.15	411.84**
R23	CMS52	-1.18	10.14*	-1.69	-0.88	-120.49*	-598.26*	0.74	-252.21
R23	CMS148	-2.05*	0.01	0.22	-0.22	-55.88	-222.43	0.11	-89.18
R23	CMS356	5.28**	-2.44	0.83	-1.22	30.76	-60.76	-1.00	-71.83
R26	CMS78	0.51	2.45	0.42	-0.56	-65.97	-350.21	0.80	-149.18
R26	CMS52	-0.11	2.32	-0.47	-0.54	44.58	235.07	1.07	148.78
R26	CMS148	0.26	-2.55	-0.35	-2.37	68.75	200.91	-0.35	71.64
R26	CMS356	-0.66	-2.25	0.41	3.50	-47.39	-85.76	-1.52	-72.63
R50	CMS78	2.45*	7.86	-1.07	1.94	-8.67	6.04	-0.48	-17.65
R50	CMS52	-0.43	-16.36**	0.54	-0.66	13.36	29.66	-0.40	3.84
R50	CMS148	-2.05*	-0.01	1.91*	-3.00	-37.44	-337.85	0.60	-132.12
R50	CMS356	0.03	8.49	-1.38	1.75	32.75	302.16	0.28	144.53
R56	CMS78	1.14	2.86	-0.16	2.44	32.12	174.80	-0.55	65.00
R56	CMS52	1.76	-3.74	0.62	3.21	-105.02	-376.60	-1.16	-203.87
R56	CMS148	-0.36	-0.89	-0.76	0.75	-65.63	-304.10	0.90	-114.49
R56	CMS356	-2.53*	1.74	0.30	-6.37	138.52*	505.91*	0.81	251.98*
SE		1.02	5.02	0.92	2.12	56.35	237.68	1.05	116.12

* and ** significant at 0.05 and 0.01 level of probability respectively

Ordination with PCA was used to determine if there is any structure related to agronomic performance, GCA and SCA regarding multivariate characters. According to combining abilities two principal components accounted for 62% of variability in the GCA and SCA of entries. Ordination in biplot was based on discrimination of entries by multivariate GCA and SCA for measured characters. Traits with higher weight in principal components could discriminate entries effectively to exert multivariate selection. Combining ability of data, seed number (SN), oil yield (OY) and seed yield (SY) had the highest weight in principal component 1, so this component could discriminate entries according to the traits. It can be seen that R23 x CMS78 with highest SCA for SY located further along in the positive direction of its vector (Fig. 1). Two other hybrids with high SCA located near the vector but with a different distance from the vector. The closer one (R21 x CMS52) has a higher SCA. On the other hand,

Table 6. Mean values of agronomic traits in the crosses

Restorer	CMS	Growth Period	Plant Height	Head Diameter	1000 Seed weight	Seed Number /head	Seed yield	Oil Content	Oil Yield
R19	CMS78	106.00	179.63	19.10	70.38	547.26	2,578.35	43.84	1,136.02
R19	CMS52	104.50	192.48	18.05	70.13	710.49	3,241.68	45.37	1,459.63
R19	CMS148	110.00	172.80	18.48	70.50	733.06	3,480.02	43.48	1,503.25
R19	CMS356	106.75	170.88	17.48	75.63	430.36	2,150.01	42.97	915.35
R21	CMS78	103.50	191.25	17.80	59.25	726.07	2,851.68	47.62	1,372.05
R21	CMS52	105.75	194.03	21.78	65.25	965.61	4,200.02	44.33	1,862.48
R21	CMS148	106.25	187.65	17.30	68.75	827.29	3,700.02	44.74	1,663.71
R21	CMS356	109.00	199.13	19.80	82.38	396.79	2,175.01	42.70	928.24
R23	CMS78	106.50	181.48	18.45	77.63	794.72	4,105.02	47.80	1,961.08
R23	CMS52	106.75	200.63	17.15	74.75	594.69	2,920.02	47.68	1,397.32
R23	CMS148	106.75	179.58	18.05	74.38	640.10	3,188.35	46.80	1,502.43
R23	CMS356	117.00	186.40	18.98	84.00	482.57	2,585.01	43.07	1,116.19
R26	CMS78	105.00	181.98	19.05	70.25	579.93	2,678.35	47.59	1,276.24
R26	CMS52	103.75	183.13	19.20	70.63	756.54	3,558.35	47.15	1,674.47
R26	CMS148	105.00	167.33	18.30	67.75	761.51	3,416.68	45.48	1,539.42
R26	CMS356	107.00	176.90	19.38	84.25	401.20	2,365.01	41.69	991.57
R50	CMS78	107.00	196.65	17.30	72.75	699.02	3,365.02	44.21	1,486.29
R50	CMS52	103.50	173.70	19.95	70.50	787.11	3,683.35	43.57	1,608.07
R50	CMS148	102.75	179.13	20.30	67.13	717.11	3,208.35	44.32	1,414.19
R50	CMS356	107.75	196.90	17.33	82.50	543.13	3,083.35	41.39	1,287.25
R56	CMS78	108.25	192.48	17.73	69.00	716.80	3,268.35	47.96	1,571.39
R56	CMS52	108.25	187.15	19.55	70.13	645.72	3,011.68	46.64	1,402.79
R56	CMS148	107.00	179.08	17.15	66.63	665.91	2,976.68	48.44	1,434.26
R56	CMS356	107.75	190.98	18.53	70.13	625.89	3,021.68	45.74	1,397.14
LSD5%		3.53	17.81	3.23	7.72	195.7	819.5	3.74	414.4

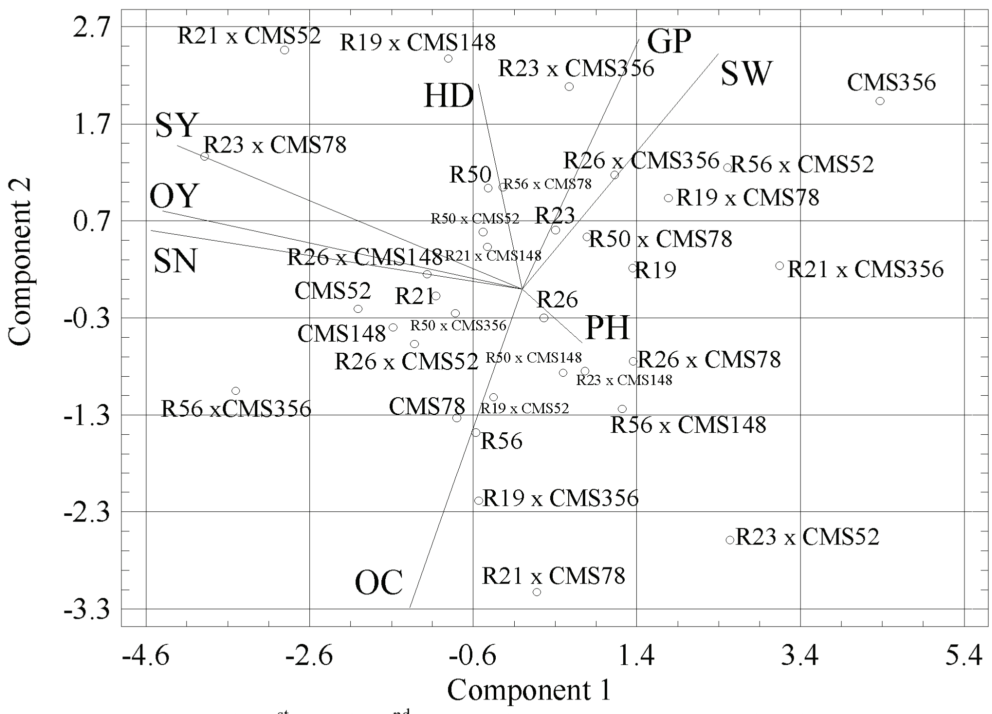

Fig. 1. Biplot of the 1st and the 2nd principal component for general and specific combining abilities.

the lowest SCA belongs to R23 x CMS52, and, logically, it is located further along in the inverse direction of its vector followed by the next high SCA hybrids R21 x CMS356 and R21 x CMS78. This statement is in accordance with de la Vega et al. (1997) who reported the discrimination of genotypes and environment based on their interactions. Acute angles for SY, OY and SN indicated positive associations between them. 1000 seed weight (SW), growth period (GP) and head diameter (HD) is closely associated in this respect. So selection should be made according to one of these traits, accompanied by selection according to the associated traits too, and this would allow making multivariate selection on breeding materials. We found that oil content (OC) has not been associated with SY and OY and that there is a strong negative association for SW with GP and HD because of the obtuse angle for their vectors. Among parental lines, CMS356 had the highest GCA for SW and was located in the same direction of its vector with a distance further away from the origin. R56 and CMS78 had higher SCA for OC and were located in the same direction of their vector. Genotypes with values close to the mean of entries were located near the origin. This was the situation for most parental lines.

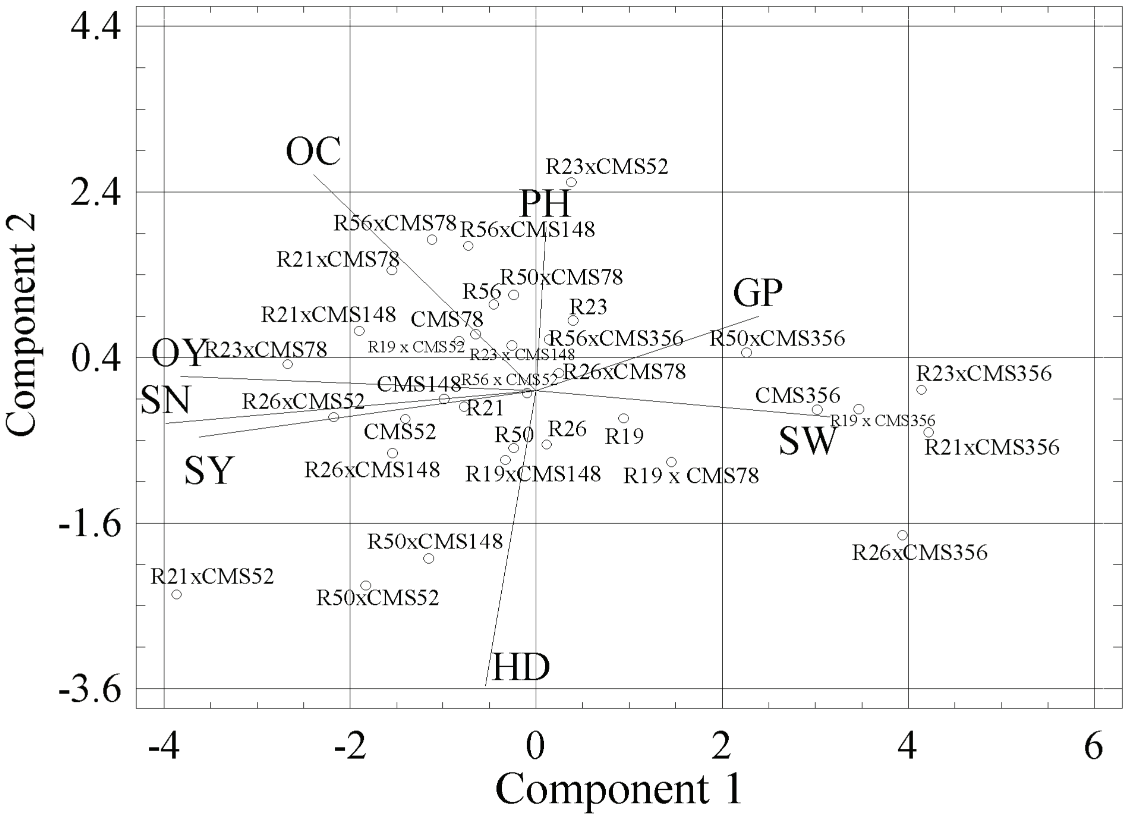

Fig. 2. Biplot of the 1st and the 2nd principal component for agronomic means

Traits with low weight in principal components could not discriminate entries efficiently .This is true for R56 x CMS148 due to its SCA for plant height (PH). Considering agronomic means of SN, OY and SY, these had the highest negative weight in principal component 1, whereas in principal component 2 positive weights of OC and PH and negative weight of HD were higher. A negative association for SN and SW is understood from biplot in Fig. 2. R21 x CMS52 and R23 x CMS78 were located further along in the positive direction of SY vector because of their high seed yields. Hybrids with a low seed yield located on the inverse side of biplot ordination of entries might be influenced by the presence of multivariate effects. For example, ordination of R26 x CMS52 with a lower seed yield than R21 x CMS 148 is not in agreement with the statements, and, in fact, it resulted from the effect of the GP. So, multivariate reactions could cause problems in the ordination of entries, which are felt by breeders considering multivariate selection, but PCA biplots generate equilibrium ordination due to different traits which could be used for precise selections. In this experiment, if it is desired to select an early mature hybrid with a higher oil yield and oil content using Fig. 2, R23 x CMS78 would be a good choice. Association and discrimination behavior of biplot in Fig. 2 is the same as that mentioned for Fig. 1 but the ordination of entries is slightly different. Except R21 x CMS356, all other combinations of R21 are on the left of the biplot in the same direction as that of their parents, except CMS356. It can be seen that R21 has a stronger positive effect on related hybrids than CMS356, and that agronomic means could discriminate genotypes more effectively than SCA data. Also, CMS 356 has a strong increasing effect on its crosses with all 6 restorer lines considering their SW and a decreasing effect on their SY, OY, SN and OC. R56 x CMS356 with higher SCA for SY has a SY close to the mean of entries and SCA was not able to show its real performance. So these biplots provide more useful information to the breeder, PCA based on agronomic data alone could be used for determining the real performance of entries considering multivariate traits.

REFERENCES

Alvarez D., P. Luduena, and E. Frutos. 1992. Variability and genetic advance in sunflower. p. 963-968. In: Proc. 13th Int. Sunflower Conf., Pisa, Italy.

Bedov, S. 1985. A study of combining ability for oil and protein contents in seed of different sunflower inbreds. p. 675-682. In: Proc. 11th Int. Sunflower Conference, Mar del Plata, Argentina.

De la Vega, A., S.C. Chapman, and A. J. Hall. 2001. Genotype by environment interaction and indirect selection for yield in sunflower I. Two-mode pattern analysis of oil yield and biomass yield across environments in Argentina. Field Crops Res. 27:17-38.

Gangappa, E., K.M. Channakrishnajiah, M. S. Harini, and S. Ramesh. 1997. Studies on combining ability in sunflower. Helia 20(27):73-84.

Ghaffari, M. 2006. Sunflower growing guide (In Persian). Agricultural and Natural Resources Research Center of west Azerbaijan-Iran.

Kemptorne, O. 1957. An introduction to genetic statistics. The IOWA State University Press.

Kroonenberg, P.M., 1997. Introduction to biplots for G x E tables. Research report No. 51. Center for Statistics. The University of Queensland, Brisbane, Australia.

Luczkiewicz, T., and Z. Kaczmarek. 2004. The influence of morphological and differences between sunflower inbred lines on their SCA effects for yield components. J. Appl. Genet. 45:175-182.

Sing, S.B., K.S. Labana, and D.S. Virk. 1989. Detection of epistatic, additive and dominance variation in sunflower. Indian J. Genet.,47:243-247.

Tersac, M., D. Vares, and P. Vincourt. 1993. Combining groups in cultivated sunflower populations and their relationships with country of origin. Theor. Appl. Genet. 87:603-608.