International Journal of Plant Breeding *©2007 Global Science Books*

Principal Component Analysis Based on Morphological Characters in Pea (*Pisum sativum* L.)

María Andrea Espósito[1*] • Leonardo Andrés Milanesi[1] • Eugenia Alejandra Martin[2] • Vanina Pamela Cravero[2] • Fernando Sebastian López Anido[1] • Enrique Luis Cointry[1]

[1] Genetic and Plant Breeding Department, Rosario National University (UNR), CC 14, Zavalla S2125ZAA, Argentina
[2] *CONICET* – Av. Rivadavia 1917, C1033AAJ, Buenos Aires, Argentina

Corresponding author: * **mesposi@unr.edu.ar**

Keywords: legumes, PCA, variability

Pea (*Pisum sativum* L.), a member of the *Leguminosae*, is an important cool season vegetable crop. It is a native species of Southwest Asia and was among the first crops cultivated by man. Wild field pea can still be found in Afghanistan, Iran and Ethiopia. It is also the second most important food legume worldwide after common bean. The increasing demand for protein-rich raw materials for animal feed or intermediary products for human nutrition have led to a greater interest in this crop as a protein source (Santalla *et al*. 2001). To establish a breeding program the yield components and characteristics of the plant must be evaluated since these have an affect on their harvest.

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components* (Chatfield and Collis 1980). The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The objectives of PCA are to discover or to reduce the dimensionality of the data set and to identify new meaningful underlying variables (Jollife 2002).

The mathematical technique used in PCA is called eigen analysis: We solve for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix (Harris 2001). Multivariate statistical methods have been successfully used to classify quantitative and qualitative variations in many crop species like pea (Amurrio *et al*. 1995; Rabbani *et al*. 1998; Berdahl *et al*. 1999; Chandran and Padya 2000; Cravero *et al*. 2002; Rosso and Pagano 2005; Bhargava *et al*. 2007).

An excellent way to see the results of PCA are through graphics Biplot (Gabriel 1971), specifically the connections between the ordinations of the rows (generally samples) and of the columns (generally markers) of the data table. The Biplot represent both ordinations simultaneously.

In the present study we carried out a PCA to determine which variable contributed more with its variance to total variance observed in the pea accessions. For this, fifteen characteristic were measured in two seasons (2005-2006): Length (LS) and width (WS) of stipule, leaflets (LL, WL), length of the internodes (LI), plant height (PH), number of nodes at the first flower (NF) and at the first pod (NV), and numbers of days to flowering (DF) were measured in the flowering period while the length and width of pod (LP, WP), numbers of pods (NP) and seeds per plot (NS), yield (Y) and grain diameter (GD) were measured at the dry seed stage. Four traits (plant height, size of pods and stipule, and diameter of grain) were recorded in cm, while the nodes at the first flower and at the first pod were counted with the average of three plants randomly selected in the centre of rows. The yield was estimated in g per plot at harvest. Seeds and pod per plot were counted. The days to flowering were estimated as days from sowing time to the day on which at least 50% of the plants in the plot had started to flower.

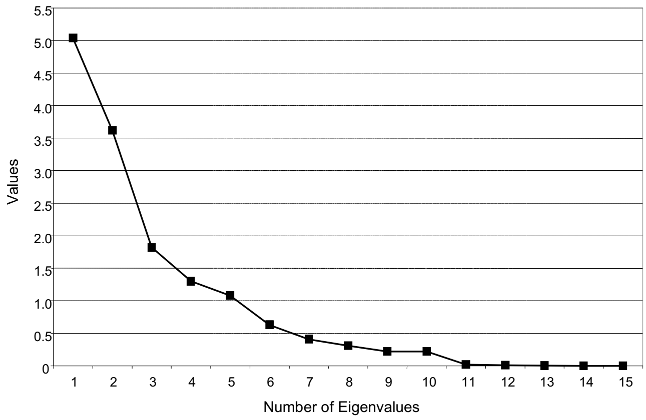

Fig. 1 Plot of Eigenvalues for 2005 season.

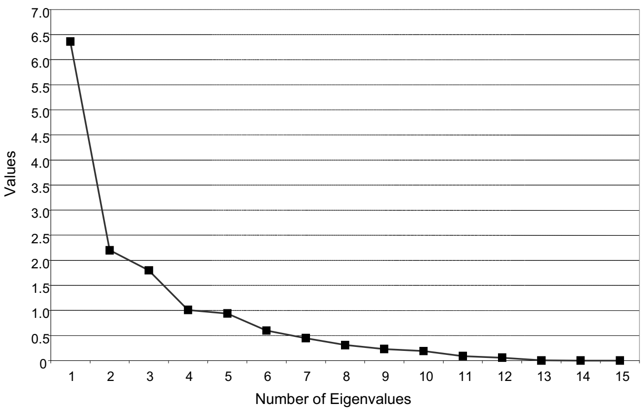

Fig. 2 Plot of Eigenvalues for 2006 season.

The PCA was carried out through InfoGen software (Balzarini and di Renzo 2003): four PCs (**Figs. 1, 2**) explained 81.0% of the total variation observed among the varieties for both seasons (**Tables 1A, 1B, 2A, 2B**). The

Received: 30 July, 2007. Accepted: 25 August, 2007.

Short Communication

Table 1A Eigenvalues obtained of the PCA for 2005 season.

Lambda	Value	%	% Accum
1	5.04	0.34	0.34
2	3.62	0.24	0.58
3	1.82	0.12	0.70
4	1.30	0.11	0.81
5	1.08	0.07	0.88
6	0.63	0.04	0.92
7	0.41	0.03	0.95
8	0.31	0.02	0.97
9	0.22	0.01	0.98
10	0.22	0.01	1.00
11	0.02	1.6E-03	1.00
12	0.01	6.9E-04	1.00
13	3.0E-03	2.0E-04	1.00
14	0.00	0.00	1.00
15	0.00	0.00	1.00

Table 1B Eigenvectors obtained of the PCA for 2005 season.

Variables	PC_1	PC_2	PC_3	PC_4
PH	0.25	0.32	-0.15	-0.02
LP	-0.20	-0.05	0.49	-0.36
WP	0.22	0.01	0.31	-0.34
Y	-0.01	0.46	-0.03	-0.27
NP	-0.11	0.40	-0.27	-0.25
NS	-0.23	0.32	-0.15	-0.29
NF	-0.27	0.28	-0.03	0.35
LI	0.34	0.26	-0.11	0.10
WS	0.19	0.31	0.32	0.22
LS	0.09	0.32	0.47	0.26
NV	-0.28	0.08	-0.28	0.38
WL	0.38	0.03	-0.20	-0.13
LL	0.39	-2.5E-03	-0.19	-0.15
GD	0.28	0.10	0.17	0.32
DF	-0.31	0.27	0.15	-0.10

Table 2A Eigenvalues obtained of the PCA for 2006 season.

Lambda	Value	%	% Accum
1	6.36	0.44	0.44
2	2.20	0.16	0.60
3	1.80	0.14	0.74
4	1.01	0.07	0.81
5	0.94	0.06	0.87
6	0.60	0.04	0.91
7	0.45	0.03	0.94
8	0.31	0.02	0.96
9	0.23	0.02	0.98
10	0.19	0.01	0.99
11	0.09	0.01	1.00
12	0.06	3.9E-03	1.00
13	3.5E-03	2.3E-04	1.00
14	0.00	0.00	1.00
15	0.00	0.00	1.00

Table 2B Eigenvectors obtained of the PCA for 2006 season.

Variables	PC_1	PC_2	PC_3	PC_4
PH	0.26	-0.18	0.29	0.07
LP	-0.13	0.09	0.59	-0.05
WP	0.14	0.19	0.20	-0.12
Y	0.36	0.01	0.14	0.17
NP	0.34	-0.14	0.12	0.23
NS	0.30	-0.19	0.19	0.33
NF	-0.06	-0.08	-0.59	0.36
LI	0.26	-0.28	-0.11	0.11
WS	0.14	0.56	0.01	0.04
LS	0.14	0.53	-0.03	0.22
NV	-0.29	0.04	0.03	0.40
WL	0.33	-0.10	-0.16	-0.37
LL	0.33	-0.11	-0.17	-0.37
GC	0.20	0.39	-0.20	-0.21
DF	-0.33	-0.14	0.07	-0.33

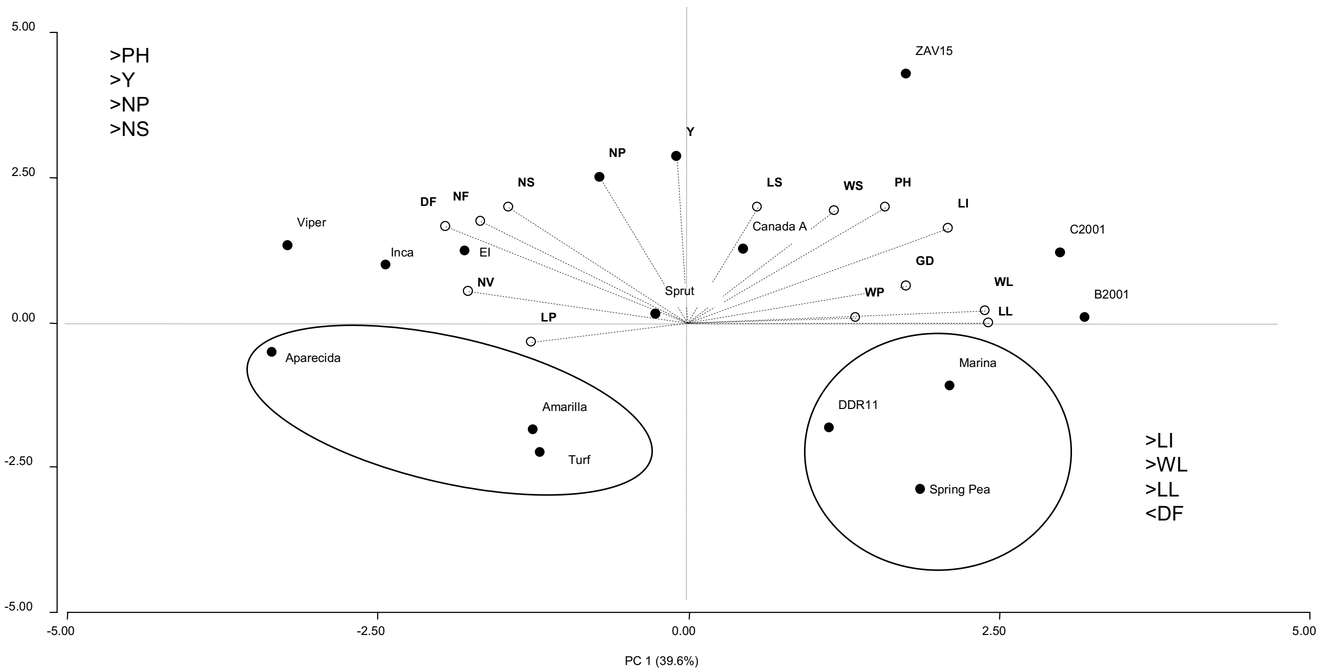

Fig. 3 Principal Component Analysis graphs. PC1 and PC2 are the two first principal components for 2005 season.

data was standardized previously at analysis. We considered only the first two PCs, which explained 67.7% (for 2005) and at 69.8% (for 2006) of the total variance.

According to their dispersion on the PCA diagram for 2005, we identified different groups of accessions (**Fig. 3**). Three varieties, Aparecida, Amarilla and Turf, were plotted on the same left portion of the PCA diagram, indicating that they are closely related genotypes, characterized by low yield, seeds per plot, plant height and length and width of leaflets by high days at flowering. We identified another group that consisted of Marina, DDR11 and Spring Pea on the lower right portion of the diagram, with low yield, seeds per plot, plant height and days at flowering by high length and width of leaflets and length of internodes.

For 2006, we also identified groups of accessions in the diagram of the PCA diagram (**Fig. 4**). Two accessions, Aparecida and Viper were plotted on the same left portion of this diagram, indicating that they are closely related genotypes, characterized by low yield and grain diameter. Also we identified another group that consisted of ZAV15

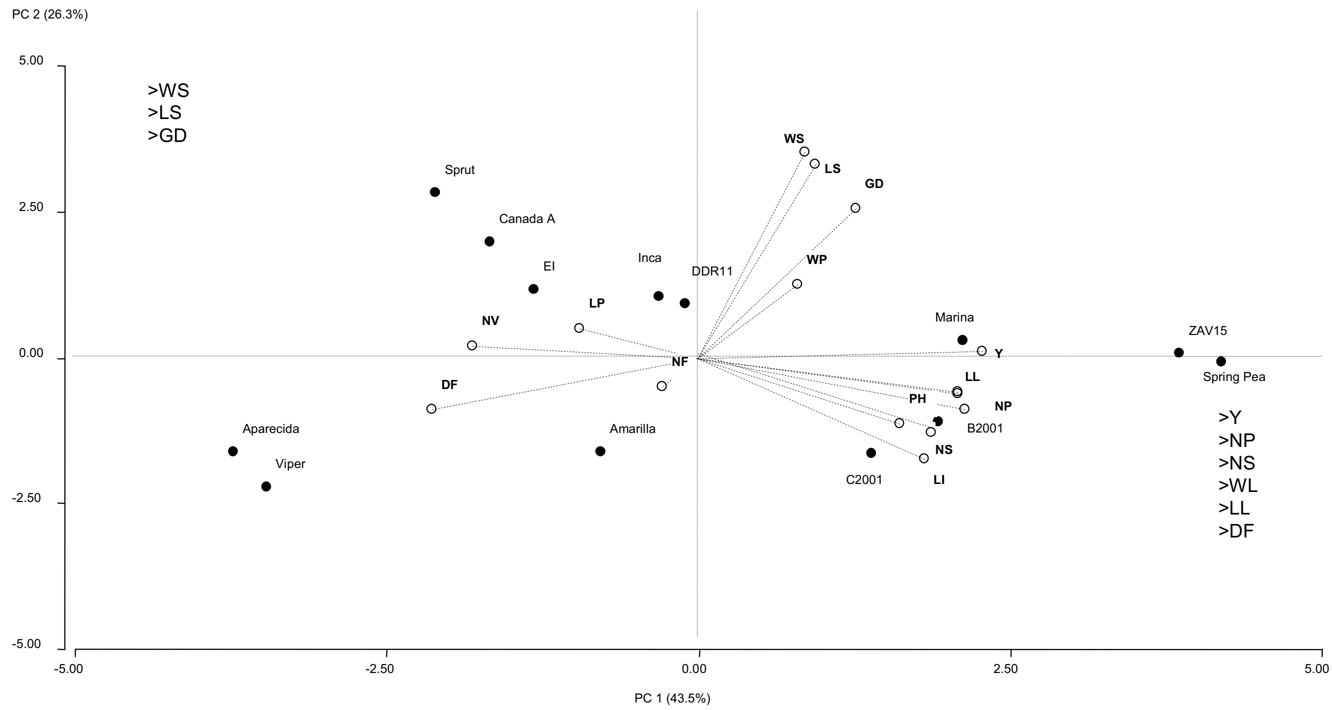

Fig. 4 Principal Component Analysis graphs. PC1 and PC2 are the two first principal components for 2006 season.

and Spring Pea with high yield and grain diameter. We can see that the accessions were grouped similarly in both seasons, except for any cultivars that changed their position in the figure for the different environments. This fact implies that the genotype × environmental interactions are present.

The summation of these results may provide information about the importance of some of the characters studied among all accessions analyzed because the fifteen characters previously identified (Espósito *et al.* 2007) as promising criteria for characterization of the material eleven were shown to be relevant. Our study provides a better insight regarding the relationship between various characters determining either timing of flowering or productivity, and provides information on the likely interest of a particular accession as a parent in initial crosses for breeding stable and high-yielding varieties.

ACKNOWLEDGEMENTS

Financial support for this research work was provided by PICT N° 08-14645 from the Agencia Nacional de Investigaciones Científicas y Técnicas (ANCYT).

REFERENCES

Amurrio J, de Ron A, Zeven AC (1995) Numerical taxonomy of Iberian pea landraces based on quantitative and qualitative characters. *Euphytica* **82**, 195-205

Balzarini M, di Renzo J (2003) Infogen: Software para análisis estadísticos de marcadores genéticos. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Córdoba

Berdahl J, Mayland HF, Asay KH, Jeerson PG (1999) Variation in agronomic and morphological traits among Russian wildrye accessions. *Crop Science* **39**, 1890-1895

Bhargava A, Shukla S, Rajan S, Ohri D (2007) Genetic diversity for morphological and quality traits in quinoa (*Chenopodium quinoa* Willd.) germplasm. Genetic *Resources and Crop Evolution* **54**, 167-173

Chandran K, Padya SM (2000) Morphological characterization of *Arachis* species of section Arachis. *Plant Genetic Resources Newsletter* **121**, 38-41

Chatfield C, Collis A (1980) *Introduction to Multivariate Analysis*, CRC Press, Boca Raton, 246 pp

Cravero VP, López Anido FS, Cointry EL (2002) Caracterización y selección de familias S1 de alcaucil a través de técnicas de análisis multivariado. *Horticultura Brasileira* **20**, 619-625

Espósito MA, Martin EA, Cravero VP, Cointry E (2007) Characterization of pea accessions by SRAPs markers. *Scientia Horticulturae* **113**, 329-335

Gabriel KR (1971) Biplot display of multivariate matrices with application to principal components analysis. *Biometrika* **58**, 453-467

Harris R (2001) *A Multivariate Analysis*, Laurence Girlbaum Associates, New Jersey, USA, 624 pp

Jolliffe I (2002) *Principal Component Analysis*. *Springer Series in Statistics* (2nd Edn), New York, USA, 487 pp

Rabbani MA, Iwabuchi A, Murakami Y, Suzuki T, Takayanagi K (1998) Phenotypic variation and the relationships among mustard (*Brassica juncea* L.) germplasm from Pakistan. *Euphytica* **101**, 357-366

Rosso B, Pagano E (2005) Evaluation of introduced and naturalized populations of red clover (*Trifolium pretense* L.) at Pergamino EEA- INTA, Argentina. *Genetic Resources and Crop Evolution* **52**, 507-511

Santalla M, Amurrio J, de Ron A (2001) Food and feed potential breeding value of green, dry and vegetal pea germplasm. *Canadian Journal of Plant Science* **81**, 601-610