UDC 575: 635.652
Original scientific paper

DIVERGENCE IN THE DRY BEAN COLLECTION BY PRINCIPAL COMPONENT ANALYSIS (PCA)

Mirjana VASIĆ, Jelica GVOZDANOVIĆ-VARGA
and Janko ČERVENSKI

Institute of Field and Vegetable Crops, Novi Sad, Serbia

Vasić M., J. Gvozdanović-Varga and J. Červenski (2008): *Divergence in the dry bean collection by principal component analysis (PCA).–* Genetika, Vol. 40, No. 1, 23 - 30.

We studied the divergence of our beans collection. The study included two qualitative traits, grain color and shape, and 13 quantitative traits, namely three components of plant height, five direct yield components and five chemical properties of grain. The principal component analysis (PCA) showed which of the traits were decisive in genotype differentiation. The principal components were formed based on the correlation matrix and shown through unrotated and rotated values of trait correlation with the main axes. The percentage contribution of particular principal components to total variability was shown, as was the accumulation of variability. The variability of the collection was interpreted based on the seven principal components, the first one describing genotype productivity, the second grain shape, the third grain size, the fourth genotype harvestability and the last three describing the chemical composition of grain.

Key words: divergence, qualitative and quantitative traits, PCA, *Phaseolus vulgaris*

Corresponding author: Mirjana VASIĆ, vasicka@ifvcns.ns.ac.yu, Institute of Field and Vegetable Crops, Maksima Gorkog 30, 21000 Novi Sad, Serbia

INTRODUCTION

Collections of bean genotypes have been established at the research centers dealing with bean breeding (ANTUNES *et al.*, 1981; CONTI, 1982; TRIFUNOVIĆ and MILOŠEVIĆ, 1982; VASIĆ, 1986; ESCRIBANO *et al.*, 1994). The collection was put together in such a way that it comprised genotypes with all grain colors of interest for the market (GENCEV and KIRJAKOV, 1994; KELLI *et al.*, 1998; KRSTEVA, 2000; VASIC, 2001). Simultaneously, such organization of the collection allows breeding for important market categories. The collection has been formed with the objective to study the variability of the genotypes in order to select those that are most suitable for breeding, both for hybridization and selection of lines from populations. When breeding for a particular set of growing conditions, it is highly important to know and use the local populations, since in them the relationships among yield components are balanced and in harmony with the effects of the specific climatic and edaphic factors. The principal component analysis (PCA), one of Multivariate Analysis methods, showed which of the traits were decisive in genotype differentiation (KOVACIC, 1994). PCA enables easier understanding of impacts and connections among different traits by finding and explaining them.

MATERIAL AND METHODS

We studied the divergence of our beans collection for a period of three years. They represent a part of the working collection of Institute of Field and Vegetable Crops in Novi Sad. The study included two qualitative traits, grain color and shape, and 13 quantitative traits, namely three components of plant height, five direct yield components and five chemical properties of grain.

For calculation each color and shape of grain were assigned a numerical code according to the idea of HUSSANI *et al.* (1977). The PCA was based on Pearson correlation matrix and Euclidean distances. Latent roots or Eigenvalues for all principal components were shown. The percentage contribution of particular principal components to total variability was shown, as was the accumulation of variability. The variability of the collection in was interpreted based on the seven principal components. Non-rotated and rotated values of Latent vectors (Component weights, Factor loadigins) were shown. The Warimax method (KAISER in WILKINS, 1990) was used for the rotation of principal components.

PCA were calculated by the statistical package SYSTAT, module STATS (WILKINSON, 1990).

RESULTS AND DISCUSSION

Number of Principal Components calculated from correlation matrix is 15 (Table 1). It is similar to number of observed traits. PCA concentrated variability in first principal components. Total variance explained with the first tree PC's more than

10%, in the the following three, to the seventh, is more than 5%. Variance explained with last three PC is irrelevant.

Table 1. Latent roots (Eigenvalues) and variability of non-rotated values of Principal Components

Principal components	Latent roots (Eigenvalues)	% of total variance explained	Summary communality in %
Y1	5,96	39,7	39,7
Y2	1,96	13,1	52,8
Y3	1,60	10,7	63,5
Y4	1,06	7,1	70,5
Y5	0,89	5,9	76,4
Y6	0,83	5,5	82,0
Y7	0,77	5,1	87,1
Y8	0,57	3,8	90,9
Y9	0,45	3,0	93,9
Y10	0,36	2,4	96,3
Y11	0,28	1,9	98,2
Y12	0,23	1,5	99,7
Y13	0,04	0,3	100,0
Y14	0,30	0,04	100,0
Y15	0,00	0,01	100,0

There is a certain number of criteria for selecting the number of principal components (PC) to be included in the future analysis, and mostly these are based on the height of Eigenvalues PC or needed Summary communality in percentage (KOVACIC, 1994). Values of Latent roots (Eigenvalues) fell lower than one after the fourth PC. With those four PCs circa 70 % of total variance has been accounted for, which is not enough to completely comprehend divergence in our dry bean collection (Table 1). Explanation of circa 90% of total variance would suffice our needs, which is why we have chosen to use seven principal components in our future research. Summary communality shows that with seven PCs 87,1% of variability is explained. Latent roots (Eigenvalues) are between 5,96 for the first and 0,77 for the seventh.

The first seven PC (principal components) are shown based on non-rotated values. Certain traits have a high coefficient of correlation with only one PC. Some have lower correlation level with more PC. (Table 2)

Table 2. Principal components analysis traits of common bean collection – non-rotate values

Traits	Principal components						
	Y1	Y2	Y3	Y4	Y5	Y6	Y7
Grain color	0,68	-0,23	0,23	-0,07	0,45	-0,21	-0,15
Grain shape	-0,31	0,25	0,59	0,40	-0,12	0,33	0,31
Total plant height	**0,77**	0,32	-0,08	0,42	0,19	-0,04	0,04
Height of the first pod	-0,38	0,48	-0,37	0,55	0,19	0,10	-0,05
Productive height	**0,89**	0,12	0,07	0,19	0,11	-0,05	0,06
N° of pods per plants	**0,88**	-0,27	0,05	0,07	-0,05	0,21	-0,12
N° of grains/plants	**0,95**	-0,13	-0,16	0,03	-0,10	0,19	0,06
N° of grains per pod	0,54	0,22	-0,55	-0,15	-0,15	-0,00	0,44
1000 grains mass	**-0,77**	-0,22	0,42	0,22	0,06	-0,03	-0,11
Grain mass per plant	0,70	-0,54	0,16	0,12	-0,03	0,21	-0,03
Proteins	0,35	0,64	0,01	-0,21	0,27	0,28	-0,42
Ashes	0,37	0,28	0,53	-0,36	0,04	0,33	0,22
Starch	-0,31	-0,56	-0,18	0,04	0,59	0,13	0,35
Fat	0,60	-0,26	0,13	0,33	-0,32	-0,20	-0,13
Cellulose	0,39	0,41	0,44	-0,06	0,11	-0,54	0,25
Latent roots	5,96	1,96	1,60	1,06	0,89	0,83	0,77
% of total variance explained	39,7	13,1	10,7	7,1	5,9	5,5	5,1
Summary communality	39,7	52,8	63,5	70,5	76,4	82,0	87,1

High percentage of more than half of observed traits in the first PC and low correlation with traits and next PC have facilitated the need for principal components rotation.

In most cases the number of principal components rotated depends on the number principal components chosen for the next analysis (SEAL, 1964). Here too all seven PCs have been rotated. In this way we had clearer distribution of particular traits in them, and traits were more closely tied to particular PComponents (Table 3.) Summary communality remained 87 % after rotation.

Table. 3. Principal components analysis traits of common bean collection - Rotated component loadings

Traits	Principal components						
	Y1	**Y2**	**Y3**	**Y4**	**Y5**	**Y6**	**Y7**
Grain color	0,63	-0,29	-0,06	-0,20	-0,23	0,46	0,24
Grain shape	-0,11	**0,88**	-0,28	0,11	0,05	0,05	-0,05
Total plant height	0,62	0,06	0,37	0,41	0,08	0,40	0,26
Height of the first pod	-0,30	0,12	0,01	**0,86**	-0,03	-0,11	0,10
Productive height	**0,71**	0,01	0,36	0,06	0,08	0,42	0,21
N° of pods per plants	**0,89**	-0,06	0,25	-0,20	0,05	-0,01	0,16
N° of grains per plants	**0,79**	-0,06	0,54	-0,13	0,07	0,01	0,14
N° of grains per pod	0,12	-0,09	**0,92**	0,03	0,04	0,07	-0,01
1000 grains mass	-0,36	0,24	**-0,78**	0,05	-013	-0,11	-0,26
Grain mass per plant	**0,87**	0,01	0,05	-0,28	-0,12	-0,08	-0,05
Proteins	0,05	0,01	0,12	0,11	0,21	0,10	**0,90**
Ashes	0,13	0,47	0,16	-0,49	0,03	0,24	0,47
Starch	-0,07	-0,06	-0,11	0,02	**-0,92**	-0,15	-0,23
Fat	0,72	-0,07	0,02	-0,01	0,33	0,14	-0,26
Cellulose	0,05	0,11	0,10	-0,11	0,18	**0,89**	0,07
Latent roots()	4,23	1,19	2,20	1,36	1,14	1,50	1,46
% of total variance explained	28,21	7,92	14,66	9,035	7,60	9,98	9,70
Summary communality	28,21	36,13	50,79	59.83	67,43	77,41	87,11

The first PC could be named component of productivity since it contains pronounced traits which determine the yield level. In the highest correlation with this PC are pod number, grain mass and grain number per plant. From hight components there is productive plant hight assisted by smaller portion of plant hight. Apart from that in lower correlation with this component is grain color and grain oil content. These traits have the largest participation in the divergence of the researched collection and carry the largest portion of its variability (Table 3). Using this principal component for genotype differentiation, we have distinguished between yielding genotypes with large number of pods and grain per plant, large productive height and high grain oil content. It is specific for the tested beans that black and white grain genotapes with the highest numeric values had the highest yield, eventhough they were included only due to grain color multivariate analysis. Such correlation is not necessarily expected in different samples and genotypes, as is neither connection of grain color in this manner with the first principal component.

The second PC could be named grain shape and it shows large variability for the tested genotypes. Grain shape is one of the most obvious grain quality indicators (KELLY *et al.*, 1998), grain market characteristics which sells bean as a species and foodstuff (ROSIĆ *et al.*, 1970). Selection route is based on grain shape (ČOROKALO and MILADINOVIĆ, 1980; ACQUAAH *et al.*, 1992; BROTHERS and KELLY,1993; VASIC *et al.*, 2001), which is why genetic collections are based on the presence of all grain shapes, with other traits different within one shape. The genotypes differentiated using it would take their place similarly within previously performed grain size distribution (KRUSTEVA, 1997).

The third PC comprises of direct grain-related yield components. It combines explanation of productivity and grain quality. The traits which are in the strongest correlation with it are grain number per pod and grain number per plant as important productivity factors, as well as 1,000 grains mass which is grain quality determining trait. Considering the fact that grain size is in negative correlation with this component, it could be used to distinguish between genotypes with high 1,000 grains mass and small number of grains per pod which would have small number of grains per plant in large percentage, and genotypes with small grain size with large number of grains per pod.

The fourth PC would best describe genotype harvest stability since it comprises of the highest portion of first pod height with plant hight influence. **In the following three principal components** seed chemical composition content influence is dominant. Correlation of starch component with the fifth principal component is high, as well as correlation of cellulose with the sixth and protein with the seventh PComponent. Since these components are independent, the portion of other traits in the last three principal components is low.

Some of these traits (mass of grain, number if grain per plant) are marked important in ACQUAAH *et al* (1992) in investigation associated with architecture and seed size in breeding genotypes of dry bean. Combining usage of the set seven PC (principal components) could yield a successful selection of genotypes suitable for donors of one or more important traits in breeding.

Received December 3rd, 2007
Accepted February 28th, 2008

REFERENCES

ACQUAAH, G., M.W. ADAMS and J.D. KELY (1992): A factor analysis of plant variables associated with architecture and seed size in dry bean, Euphytica, 60:171-177.

ANTUNES, I., M. TEIXEIRA, M.J.O. .ZIMMERMANN and J.G.C. COSTA (1981): Exploration of regional populations in common beans - concepts and procedures adopted at the national research centre for rice and beans- cnpaf/Brazil. BIC, 24-27.

BROTHERS, M.E. and J.D. KELLY (1993) Interrelationship of plant architecture and yield components in the pinto bean ideotype, Crop Sci., *33*: 1234 - 1238.

CONTI, L. (1982): Bean germplasm evaluation from the collection at Minoprio (Como, Italy) in view of a breeding program for the improvement of the protein content of the seed. Genet.Agr., (3-4): 375-392.

ESCRIBANO, M.R., A.M. DE RONA and J.M. AMURRIO (1994): Diversity in agronomical traits in common bean populations from North Western Spain, Euphytica, 76, 1-6.

GENCEV, D.D. and I.D. KIRJAKOV (1994): Obiknoven zrjal fasul (Phaseolus vulgaris L). Selekcionnite priznaci i tjahnata ocenka.. PablishSajSet - Agri OOD., 60 str.

HUSSAINI, S.H., M.GODMAN and D.H. TIMOTHY (1977): Multivariate analysis and the geographical distribution of the worl collection of Finger Millet. Crop.Sci. *17*: 257-263.

KELLI, J.D., J.M. KOLKMAN and K.SCHNEIDER (1998): Breeding for yield in dry bean (Phaseolus vulgaris L). Euphytica, *102*:343-356.

KOVAČIĆ, Z. (1994): Multivarijaciona analiza. Univerzitet u Beogradu, Ekonomski fakultet, 282str

KRUSTEVA, L.I. (2000): Green Bean Germplasm in an EX SITU Collection – Evaluasion, Identification,Conservation and Use. Tessis for Doctor of Agricultural Science degree. Plovdiv, Bulgaria.

ROSIĆ, K., M. POPOVIĆ and D.COROKALO (1970): Sortiment i selekcija pasulja (Sortiment and breeding in dry bean). Agronomski glasnik, Yugoslavia, (1): 9-19.

TRIFUNOVIC, M. and M. MILOŠEVIC (1982): Neke osobine domacih populacija pasulja. Savrem.polj., (7-8): 335-346.

VASIC, M. (1986): Osobine zrna nekih domacih populacija pasulja. Jug. Simp. "Intenzivna proizvodnja povrca za zdravu ishranu", Split, 119-125.

VASIĆ, M., J.GVOZDANOVIĆ-VARGA and J. ČERVENSKI (1997): The interdependence of morphological characters in Yugoslavian bean varieties (Phaseolus vulgaris L.), Acta Horticulturae, *462*: 235-239.

VASIĆ M., J. GVOZDANOVIĆ-VARGA and A.TAKAČ (2001): Selekcija pasulja (Dry bean breeading). Savremena poljoprivreda- Contemporary agriculture, Novi Sad, (1-2): 237-245.

WILKINSON, L. (1990): SYSTAT: The System for Statistics, 677 pp, Ewanston, IL: SYSTAT, Inc., USA.

DIVERGENTNOST U KOLEKCIJI PASULJA PRIKAZANA ANALIZOM GLAVNIH KOMPONENATA (PCA)

Mirjana VASIĆ, Jelica GVOZDANOVIĆ-VARGA i Janko ČERVENSKI

Institut za ratarstvo i povrtarstvo, Novi Sad, Srbija

I z v o d

Ispitivana je divergentnost kolekcije pasulja. Posmatrane su boja i oblik zrna pasulja kao kvalitativne i 13 kvantitativnih osobina (tri komponente visine biljke, pet direktnih komponenti prinosa i pet hemiskih osobina zrna). Analizom glavnih komponenata (PCA) utvrđeno je koje su od ispitivanih osobina odlučujuće u diferenciranju genotipova**.** Glavne komponente su formirane na osnovu korelacione matrice. Prikaz glavnih komponenti vršen je preko nerotiranih i preko rotiranih vrednosti korelacija osobina sa glavnim osama. Prikazan je procentualni udeo pojedinih glavnih komponenata u ukupnoj varijabilnosti, kao i kumulacija varijabilnosti. Za objašnjenje varijabilnost kolekcije dovoljno je bilo sedam glavnih komponenata. Prva glavna komponenta govori o produktivnosti genotipa, druga o obliku zrna, treća o krupnoći zrna, četvrta o pogodnosti genotipova za žetvu, a ostale tri o hemijskom sastavu zrna.

Primljeno 03. XII 2007.
Odobreno.28 II.2008.