Selection to high productivity and stink bugs resistance by multivariate data analyses in soybean

Felipe Bermudez[1*] (iD), José Baldin Pinheiro[2] (iD)

1. Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina – Centro de Pesquisa para a Agricultura Familiar – Chapecó (SC), Brazil.
2. Universidade de São Paulo – Escola Superior de Agricultura "Luiz de Queiroz" – Departamento de Genética – Piracicaba (SP), Brazil.

ABSTRACT: Stink bugs that affect soybeans are responsible for significant losses in seed production, quality and germination potential, in addition to hindering the mechanized harvest. To develop insect resistant materials, the breeder can compile a selection index by factor analysis. Therefore, the objective of this work was to validate the use of factor analysis, by means of its estimated gains, for the selection of highly productive and stink bugs resistant genotypes in two soybean segregating populations. For this, the phenotypic evaluation was performed in the generation $F_{2:3}$, in two distinct experiments, being the populations from the crosses between IAC-100 × PI 295952 and IAC-100 × PI 306712. The experiments were installed in an 18 × 9 alpha-lattice design, with three replicates for each population. Agronomic and resistance characters were evaluated. The factorial scores for each character were obtained for the creation of "supercharacters". These were designed to check if the selection in the new characters could provide satisfactory simultaneous gains in the original characters. Subsequently, the analysis of variance was performed for all factors, in both populations. The F test showed the presence of variability among genotypes, allowing the selection of superior genotypes. None of the factors selected progenies with all the characters favorably, and their use was not interesting for both populations. With this, complementary studies should be performed with other selection indices in these populations.

Key words: *Glycine max*, factor analysis, *Euschistus heros*, *Piezodorus guildinii*, *Nezara viridula*.

INTRODUCTION

Stink bugs that affect soybeans [*Glycine max* (L) Merrill] are responsible for significant loss in seed production, caused by grain or pod abortion, as well as leaf retention, hindering mechanized harvesting (Gazzoni 1998).

In Brazil, the most frequent stink bugs species are: *Nezara viridula* (L.), *Piezodorus guildinii* (Westwood) and *Euschistus heros* (F.) (Heteroptera: Pentatomidae) (Medeiros and Megier 2009). The increase in the use of insecticides to control this pest, besides causing environmental damage and raising the cost of crop production, has contributed to the selection of resistant insect populations.

There is an increase in the incidence of insect pest due to increased area and successive cultivation of crops, such as occurs in soybeans. In addition, according to Panizzi et al. (1986) and Lustosa et al. (1999), breeding directed to productivity of grains and to quality of plants or their derivatives can make them more susceptible to insects. Therefore, the development of resistant cultivars would reduce the use of insecticides, which would bring economic, ecological and social benefits (Ventura and Pinheiro 1999).

To develop insect resistant materials, the breeder can compile a selection index by factor analysis. This is a multivariate analysis technique that can be used to predict the selection gains, replacing the traditional method, or by complementing the simultaneous selection techniques based on selection indexes.

For Cruz and Carneiro (2006), factor analysis is a significant alternative because it structures and simplifies the original data so that a large number of variables comes to be represented by a smaller number, expressed by linear combinations of these original data, called factors. Traits grouped in one factor are intensely correlated with each other and weakly correlated with other factors.

These factors are extracted by principal component analysis, whose function is to simplify a set of *n* variables into factors with the ability of joining the maximum amount of original variation available, while remaining mutually independent (Cruz and Carneiro 2006).

The use of factor analysis to develop indexes that allow accurate selection of maize genotypes had been used in some studies (DoVale et al. 2011; Reis et al. 2017). However, there is an absence of information on the use of these factor analyses in more than one segregating population and for resistance to stink bug complex in soybean. Therefore, the objective of this work was to validate the use of factor analysis, by means of its estimated gains, for the selection of highly productive and resistant to stink bugs genotypes in two soybean segregating populations.

MATERIAL AND METHODS

Three parents were used to perform the crosses. One of them, the cultivar IAC-100, was used in the synthesis of both populations. This cultivar was developed by the Agronomic Institute of Campinas (IAC, from Portuguese *Instituto Agronômico de Campinas*), state of São Paulo, showing resistance to the stink bug complex by several mechanisms (Carrão-Panizzi and Kitamura 1995; Pinheiro et al. 2005). The other two parents were the exotic soybean genotypes PI 295952 and PI 306712.

The phenotypic evaluation was performed in the $F_{2:3}$ generation in two distinct experiments, the population derived from the cross IAC-100 × PI 295952 being considered "population 1", and the population derived from IAC-100 × PI 306712 considered "population 2". Both were composed of 160 individuals and their respective parents (IAC-100 and one of the exotic genotypes). These experiments were installed in December 2013, in a homogeneous area located at the Anhumas Experimental Station (22°17'43"S and 51°23'14"W), Department of Genetics, Luiz de Queiroz College of Agriculture (ESALQ/USP), located in the city of Piracicaba, SP. Interleaving the useful rows of each experiment, two rows of border were sown using the BRS133 susceptible cultivar to avoid damage in the plots during the evaluations and to approximate the environmental conditions of a commercial field. The experimental plot was represented by five plants derived from an F_2 plant, and the average of these plants was considered for the statistical analysis. The distance between plants was 50 cm on the useful lines and 1.5 m between the useful lines, totaling 3.75 m² per plot. The 18 × 9 alpha-lattice design was adopted, with three replicates.

No chemical insects' control was performed and, after flowering, the level of stink bugs infestation was evaluated using the beat cloth method for ten weeks, with a minimum of eight sample points per day of evaluation.

The evaluated agronomic characters were: number of days for flowering (NDF); number of days to maturity (NDM); plant height at maturity (PHM) in cm; lodging (L), evaluated at maturity by a scale of visual notes from 1 to 5 (1 corresponding to the erect plant and 5 to the plant fully lodged); agronomic value (AV), evaluated at maturity through a scale of visual notes from 1 to 5 (1 corresponding to the plant with no agronomic value and note 5 to the plant with excellent agronomic characteristics); and grain yield (GY), evaluated by the grains weight (g) of each plant.

Characteristics associated with insect resistance were also evaluated: grain filling period in days (GP); leaf retention (LR), evaluated at maturity by a scale of visual notes from 1 to 5 (1 attributed to the plant without leaf retention (normal senescence) and 5 to the plant with total leaf retention (green leaves and stems); weight of one hundred seeds in g (WHS), from a random sample after humidity standardization; and healthy seed weight in g (HSW), that is, without damage caused by stink bugs, evaluated using sieves after grains harvest and process, as proposed by Rocha et al. (2014).

Initially, for all evaluated characters, analyzes of variance were performed in each population, using the following model (Eq. 1):

$$Y_{ijk}=\mu+\alpha_i+r_j+b_{k(j)}+\varepsilon_{ijk} \tag{1}$$

where: Y_{ijk} is the value observed for the character in the i-th genotype, in the j-th repetition, in the k-th block; μ is the overall mean for the character; α_i is the effect of the i-th genotype (i = 1, 2, 3,..., 160), considered as having a random effect, where, $g\sim$NID $(0, \sigma^2{}_g)$; r_j is the effect of the j-th repetition (j = 1, 2, 3); $b_{k(j)}$ is the effect of the block within the repetition (k = 1, 2, 3,..., 9); ε_{ik} is the effect of the random error associated with the observation of order ijk, where $\varepsilon\sim$NID $(0, \sigma^2)$.

To perform a principal components analysis, it is assumed that X_{ij} is the standardized mean of the j-th character (j = 1, 2, ..., v) evaluated in the i-th genotype (i = 1, 2, ... g). The principal components technique consists of transforming the set of v characters ($X_{i1}, X_{i2}, ..., X_{iV}$) into a new set ($Y_{i1}, Y_{i2}, ..., Y_{iV}$), which are linear functions of the X_i's and independent of each other. Therefore, a principal component can be given by the following linear combination (Eq. 2):

$$Y_{i1} = a_1 x_{i1} + a_2 x_{i2} + ... +a_v x_{iv} \tag{2}$$

Among all components, the first presents greater variance than the second, and so on. Additionally, the covariance between each pair of components is zero. Each eigenvalue corresponds to an eigenvector with the same number of elements as the initial characters. The factor analysis model used (Eq. 3) was suggested by Cruz and Carneiro (2006):

$$X_j = I_{j1}F_1 + I_{j2}F_2 + ... + I_{jm} F_m + \varepsilon_j \tag{3}$$

where: X_j is the j-th character (j = 1, 2, ..., v); I_{jK} is the factorial load for the j-th variable associated with the k-th factor (k = 1, 2, ..., m); F_K is the k-th common factor; ε_j is the specific factor. The initial factorial load represents the correlation between the character j and the factor k, defined by Eq. 4:

$$I_{IJ}=\lambda_I V_{IJ}^2 \tag{4}$$

where: λ_I is the i-th eigenvalue greater than 1 obtained from the phenotypic correlation matrix; V_{IJ} is the j-th value of the i-th eigenvector.

The fraction of the variance of X_J explained by the factors is called *commonality* and is defined by Eq. 5:

$$C_J=I_{J1}^2+I_{J2}^2+...+I_{jm}^2 \tag{5}$$

The technique involves several stages with the establishment of the number of common factors to be used, the calculation of the initial loads of these factors and the relation of the factors, obtaining the final loads and allowing to define factors. The last step involves the estimation of factor scores.

Each eigenvalue equal or greater than the 1.00 corresponds to an eigenvector which consists of a number of values equal to the number of original characters. Therefore, a given factor will have individual loads for all characters, which explains the reason for the name "common factor", used in factor analysis theory.

The rotation method used was the *varimax*, since it was assumed that the common factors are orthogonal to each other (Cruz and Carneiro 2006). This is an important step in factor analysis to quantify the effect of each common factor on the expression of the characters. The final scores are obtained by Eq. 6 presented by Manly (1986):

$$F^*=(\Lambda`\Lambda)^{-1} \Lambda´X \tag{6}$$

where: F* is the vector of dimension 1 × m of factorial scores; Λ is the matrix of dimension p × m of the final factorial loads; X is the vector with size p × 1 of the characters of the k-th genotype.

With this, the factorial scores for each character were obtained for the creation of the "supercharacters", that were designed to check if the selection in these new characters could provide satisfactory simultaneous gains in the original characters.

The factors created by the sums of the multiplications of the factorial scores by the characters, when they presented negative values, were transformed by the sum of the constant 1 + [z], where: [z] is the magnitude of the lowest value.

The factors were submitted to analysis of variance by the same model used for the original characters. To obtain the predicted gains in factor selection, the direct selection of 25 superior progenies was adopted using Eq. 7:

$$GS = h^2(\bar{X}_s - \bar{X}_o) \tag{7}$$

where: h^2 is the heritability of the character; $\bar{X}_s$ e $\bar{X}_o$ are the means of the selected and original population, respectively.

Analyzes of variance were performed using the software Statistical Analysis System (SAS) version 9.2 (SAS 2007). Principal components analysis and factor analysis were performed using the Genes Program – Computational Application in Genetics and Statistics (Cruz 2013).

RESULTS AND DISCUSSION

All characters had a normal distribution and homogeneity of residues at 5% probability (data not shown).

The analyzes of variance for the two populations have evidenced the existence of variability for all the characters, allowing the selection of superior genotypes. The coefficients of variation were between 2.26 (NDM) and 45.56 (HSW) (data not shown). High CV values can be explained by stink bug behavior, as the distribution of these insects is not uniform in the field, resulting in some genotypes more damaged than others. High CV values were also found by Rocha et al. (2014).

To determine the factorial loads, held the principal components analysis, obtaining their eigenvalues and respective eigenvectors for each population. The eigenvalues that absorbed at least 80% of the total variation among the analyzed characters were considered, which also coincided with the selection of factors that considers the number of eigenvalues greater than unity. With this, four factors were obtained for population 1 (Table 1) and three for population 2 (Table 2).

Subsequently, factor analysis was performed for each population, and the final factorial loads were obtained after rotation by the *varimax* method. The communality is the measure of the efficiency of character representation by a common part, also involved in the other analyzed characters (Cruz and Carneiro 2006). According to Souza (1988), values of commonality

Table 1. Eigenvalues (λ_j), accumulated percentage of the variance [λ_j(%)] and eigenvectors obtained from the phenotypic correlation matrix among the characters the, evaluated in 162 F$_{2:3}$ soybean genotypes of population 1, Piracicaba-SP, Brazil, 2014.

λ_j	λ_j(%)	NDF	GP	NDM	PHM	L	LR	AV	GY	HSW	WHS
3.40	34.04	0.47	-0.44	0.26	0.28	0.34	-0.09	0.02	-0.22	-0.25	-0.44
2.07	54.73	0.01	-0.21	-0.10	0.48	0.27	-0.03	-0.25	0.56	0.51	0.10
1.61	70.81	0.11	-0.03	0.03	-0.22	-0.14	-0.50	0.65	0.26	0.33	-0.26
1.22	83.00	0.23	0.17	0.70	-0.11	-0.15	0.54	0.06	0.24	0.20	-0.04
0.59	88.88	-0.29	0.42	0.19	-0.25	0.75	-0.14	-0.07	0.00	0.05	-0.20
0.48	93.63	-0.15	-0.15	-0.46	-0.03	0.21	0.66	0.42	-0.03	0.10	-0.30
0.29	96.58	0.04	0.23	0.12	0.48	0.23	0.01	0.56	-0.10	-0.17	0.54
0.19	98.46	0.59	-0.14	-0.22	-0.52	0.30	0.05	-0.04	0.01	0.11	0.46
0.14	99.86	-0.51	-0.67	0.35	-0.23	0.10	-0.01	0.11	-0.02	-0.02	0.30
0.01	100	0.00	0.01	0.08	0.10	-0.04	-0.01	-0.05	-0.71	0.69	0.03

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

higher than 0.64 have been accepted as reasonable, which is equivalent to having a coefficient of correlation higher than 0.80 between the common factors and the character.

After obtaining the final factorial loads, the commonalities were higher than 0.64 for almost all characters, except for L. In the final factorial loads of population 1, it is evident that Factor 1 can be interpreted as the determinant of PHM, L and GP characters; Factor 2 expresses the characters GY and HSW; Factor 3 explains the AV and LR characters; while Factor 4 is the determinant of the NDM, NDF and WHS characters (Table 3). The factorial scores for each factor were obtained as follows (Eqs. 8–11):

$$F1 = 0{,}11NDF - 0{,}30GP - 0{,}13NDM + 0{,}37PHM + 0{,}31L - 0{,}11LR - 0{,}22AV + 0{,}03GY + 8{,}99.10^{-5}HSW - 0{,}08WHS \quad (8)$$

$$F2 = 0{,}01NDF + 0{,}02GP + 0{,}10NDM + 0{,}10PHM - 0{,}01L + 0{,}01LR + 0{,}08AV + 0{,}50GY + 0{,}49HSW + 0{,}05WHS \quad (9)$$

$$F3 = 0{,}04NDF - 0{,}12GP - 0{,}20NDM - 0{,}13PHM - 0{,}03L - 0{,}55LR + 0{,}45AV - 4{,}22.10^{-3}GY + 0{,}07HSW - 0{,}22WHS \quad (10)$$

$$F4 = 0{,}32NDF + 0{,}02GP + 0{,}60NDM - 0{,}10PHM - 0{,}06L + 0{,}28LR + 0{,}20AV + 0{,}09GY + 0{,}07HSW - 0{,}22WHS \quad (11)$$

Table 2. Eigenvalues (λ_j), accumulated percentage of the variance [$\lambda_j(\%)$] and eigenvectors obtained from the phenotypic correlation matrix among the characters the, evaluated in 162 $F_{2:3}$ soybean genotypes of population 2, Piracicaba-SP, Brazil, 2014.

λ_j	$\lambda_j(\%)$	NDF	GP	NDM	PHM	L	LR	AV	GY	HSW	WHS
5.75	57.53	-0.37	0.09	-0.40	-0.24	-0.18	-0.36	0.35	0.34	0.37	0.33
1.57	73.20	0.05	-0.40	-0.04	0.54	0.58	-0.16	-0.08	0.33	0.24	0.10
1.01	83.32	-0.15	0.78	0.03	0.06	0.39	0.17	-0.31	0.07	0.07	0.27
0.42	87.55	-0.02	-0.33	-0.14	-0.36	0.32	-0.01	-0.10	-0.45	-0.21	0.62
0.38	91.35	-0.24	-0.25	-0.07	-0.52	0.14	0.24	-0.52	0.23	0.28	-0.35
0.33	94.69	-0.08	-0.18	0.04	0.28	-0.46	0.63	-0.14	0.15	0.17	0.45
0.24	97.09	0.73	0.05	0.31	-0.31	-0.09	-0.19	-0.12	0.30	0.24	0.26
0.15	98.60	-0.01	-0.03	0.28	-0.27	0.36	0.48	0.66	0.23	-0.05	-0.03
0.10	99.63	0.47	0.10	-0.69	0.03	0.10	0.31	0.13	-0.23	0.30	-0.16
0.04	100	-0.14	-0.0018	0.40	0.05	0.02	-0.05	0.11	-0.55	0.71	-0.04

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

Table 3. Values of commonalities, initial and final factor loads obtained in the factors analysis from the mean values of the characters evaluated in 162 $F_{2:3}$ genotypes soybean of population 1, Piracicaba-SP, 2014.

Characters	Commonalities	Initial loads				Final loads			
		Factors				Factors			
		1	2	3	4	1	2	3	4
PHM	0.83	0.52	0.69	-0.28	-0.13	0.88	0.19	-0.19	-0.00
L	0.61	0.64	0.39	-0.18	-0.16	0.77	-0.07	-0.02	0.11
GP	0.80	-0.82	-0.31	-0.04	0.19	-0.82	0.12	-0.24	-0.24
GY	0.99	-0.41	0.80	0.33	0.27	0.03	0.99	-0.04	-0.07
HSW	0.98	-0.47	0.73	0.42	0.22	-0.06	0.98	0.06	-0.11
AV	0.82	0.05	-0.36	0.83	0.07	-0.38	0.05	0.75	0.33
LR	0.78	-0.17	-0.04	-0.63	0.60	-0.19	-0.03	-0.83	0.25
NDM	0.85	0.48	-0.15	0.04	0.77	0.02	-0.04	-0.17	0.91
NDF	0.84	0.87	0.01	0.14	0.26	0.53	-0.19	0.19	0.70
WHS	0.80	-0.81	0.14	-0.33	-0.05	-0.42	0.28	-0.46	-0.58

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

In population 2, all commonalities were higher than 0.64 for all characters (Table 4). The final factorial loads show that the determinant of the majority of characters is the Factor 1, while Factor 2 expresses the characters L and PHM, and Factor 3 explains only GP. The factor scores obtained for each factor are given by Eqs. 12–14:

$$F1 = -0,14NDF - 0,04GP - 0,16NDM + 0,09PHM + 0,15L - 0,17LR + 0,09AV + 0,24GY + 0,22HSW + 0,18WHS \quad (12)$$

$$F2 = 0,03NDF + 0,07GP + 0,05NDM + 0,41PHM + 0,57L + 0,03LR - 0,24AV + 0,19GY + 0,12HSW + 0,13WHS \quad (13)$$

$$F3 = -0,16NDF + 0,84GP + 0,02NDM - 0,15PHM + 0,14L + 0,19LR - 0,23AV - 0,03GY - 1,18.10^{-3}HSW + 0,22WHS \quad (14)$$

From these scores, the new "supercharacters" were estimated, here called "Factors". Subsequently, analysis of variance was performed for all Factors, in both populations (Table 5). The F test showed the presence of variability among the genotypes, allowing the selection of superior genotypes. The coefficients of variation were as expected for the populations with stink bug incidence.

Furthermore, the preferred directions for each character were stated, being desirable the increase of the characters NDF, AV, GY and HSW; and the decrease of GP, L, LR and WHS, for both populations. However, two characters had different preferential directions in each population, being desirable the nonmodification of NDM and the increase of PHM in population 1, while the decrease of them in population 2 is more interesting.

Considering the preferred directions mentioned above, neither factor for both populations fully met the requirements for efficient selection, with at most 7 of 10 directions being met for Factors 1 and 3 of population 1. Among the factors elaborated, the least efficient in directing the characteristics was Factor 3 of population 2, which attended only PHM.

The best genotypes in each population were selected based on the coefficients of weighting of traits from the scores obtained in each factor (Cruz and Carneiro 2006; DoVale et al. 2011). The direct selection in the Factors of population 1 can be compared to the Confidence Interval (CI) of original population means (Table 6). Some characters had the opposite signal as expected according to the factor scores in the formulas. Among these, NDF in Factors 2 and 3, GP in Factor 3, L and AV in Factor 2 and GY in Factor 3. In this population it would be interesting to increase NDF to select genotypes with higher juvenile period. However, there was the opposite effect on Factors 2 and 3, reducing the

Table 4. Values of commonalities, initial and final factor loads obtained in the factors analysis from the mean values of the characters evaluated in 162 $F_{2:3}$ genotypes soybean of population 2, Piracicaba-SP, 2014.

Characters	Commonalities	Initial loads			Final loads		
		Factors			Factors		
		1	2	3	1	2	3
HSW	0.90	0.90	0.30	0.07	0.95	-0.08	0.04
GY	0.84	0.82	0.42	0.07	0.92	0.05	-0.02
WHS	0.70	0.78	0.13	0.28	0.78	-0.08	0.29
AV	0.82	0.84	-0.10	-0.31	0.71	-0.55	-0.14
NDF	0.82	-0.89	0.06	-0.15	-0.80	0.33	-0.27
LR	0.79	-0.85	-0.20	0.17	-0.84	0.25	0.13
NDM	0.90	-0.95	-0.05	0.03	-0.88	0.34	-0.06
L	0.86	-0.42	0.73	0.39	-0.06	0.92	-0.01
PHM	0.78	-0.57	0.67	0.06	-0.24	0.79	-0.31
GP	0.92	0.21	-0.50	0.79	0.04	-0.14	0.95

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

characteristic that is already considered low. The increase in GP can be observed in Factor 3, which is undesirable as this is the period when plants are most susceptible to stink bug attack, while Factors 2 and 4 did not differ from the original population.

Almost all factors selected materials with reduction for NDM, except Factor 2, which was the only one to maintain the original population average, which is a desirable effect for this population. Factors 3 and 4 had the effect of reducing the PHM. This character has a high positive correlation with lodging (L), already reported by Gallon et al. (2016),

Table 5. Summary of analysis of variance, means, Coefficient of Variation in percentage [CV (%)] of factor scores for both populations evaluated in 162 $F_{2:3}$ soybean genotypes of population 1, Piracicaba-SP, Brazil, in 2014.

Source of variation	DF	Population 1				Population 2		
		Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3
Repetition	2	76.70**	4633.58**	36.05**	188.42**	1239.53**	1229.21**	2.27[ns]
Block(Rep)	24	32.15**	251.96**	4.60**	14.93**	64.30**	67.47**	15.95[ns]
Treatment	161	84.14**	97.82**	5.01**	24.10**	107.76**	67.93**	37.14**
Error	298	7.37	57.76	1.30	4.11	17.17	20.91	11.18
μ		-0.41	53.25	-33.81	86.86	-16.56	47.76	6.67
CV(%)		22.42	14.27	14.51	2.34	23.84	9.57	24.47

Degrees of freedom (DF); [ns]not significant, *significant at 5% and **significant at 1% of probability by F test.

Table 6. Predicted gains percentages estimates ($GS_{(\%)}$), selected population mean ($\bar{X}_s$) by direct selection in the factorial scores, and mean ($\bar{X}$) and confidence interval (CI) of original population for the characters evaluated in 162 $F_{2:3}$ of soybean genotypes in population 1, Piracicaba-SP, 2014.

	Factor		NDF	GP	NDM	PHM	L	LR
Pop. 1	1	$GS_{(\%)}$	2.18	-9.73	-0.51	41.27	9.67	-0.05
		$\bar{X}_s$	55.53	28.76	117.07	75.38	2.28	1.74
	2	$GS_{(\%)}$	-0.62	-0.16	-0.13	12.01	3.37	-0.49
		$\bar{X}_s$	53.61	32.53	117.62	58.44	2.03	1.70
	3	$GS_{(\%)}$	-1.81	1.86	-1.99	-14.46	-5.94	0.01
		$\bar{X}_s$	52.80	33.32	114.89	43.12	1.66	1.75
	4	$GS_{(\%)}$	2.45	-0.94	2.64	-8.55	-1.12	0.51
		$\bar{X}_s$	55.72	32.22	121.69	46.54	1.85	1.79
Pop. 1 (original)		$\bar{X}$	54.04	32.59	117.81	51.49	1.90	1.74
		CI	(53.63; 54.45)	(32.08; 33.10)	(117.20; 118.42)	(49.41; 53.57)	(1.82; 1.98)	(1.67; 1.82)
	Factor		AV	GY	HSW	WHS		Total
Pop. 1	1	$GS_{(\%)}$	-5.7	4.91	2.71	-2.58		42.17
		$\bar{X}_s$	2.83	40.54	36.23	11.81		
	2	$GS_{(\%)}$	-1.92	10.66	8.57	4.53		35.82
		$\bar{X}_s$	3.07	45.71	41.98	12.74		
	3	$GS_{(\%)}$	0.37	2.54	3.19	0.77		-15.46
		$\bar{X}_s$	3.22	38.42	36.70	12.25		
	4	$GS_{(\%)}$	3.01	-0.09	-0.04	-13.90		-16.03
		$\bar{X}_s$	3.39	36.06	33.54	10.33		
Pop. (original)		$\bar{X}$	3.20	36.14	33.58	12.15		
		CI	(3.13; 3.26)	(35.12; 37.16)	(32.64; 34.52)	(11.77; 12.53)		

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

and the selection gain to PHM had the same effect in L means. For LR, all factors maintained parental averages, but it would be interesting to reduce this character. According to Silva et al. (2013), LR may be caused by several biotic and abiotic factors, such as stink bug attack, water stress, nutritional imbalance, disease occurrence, and predisposition of cultivars. Therefore, reducing this feature would select materials that are tolerant of all these factors. Character AV was selected for its reduction in Factors 1 and 2, and kept the average in Factor 3, and its preferential direction would be to increase its values, since this character is measured considering its agricultural suitability as a whole. Both GY and HSW characters had their characters increased, regardless of previous character directions, except for Factor 4, which kept the original population average. This indicates that selection can increase the yield and weight of good seeds even by select pseudoresistance characters in an unfavorable direction, as in the WHS character that had all possible directions but did not change the HSW. Thus, it can be assumed that pseudoresistance mechanisms are not really efficient in resistance to stink bugs, because even selecting resistance characters contrary to the desired, there was an increase in the most important character, which is HSW.

Many effects were similar in population 2, such as reduced NDF and increased GP and WHS in Factors 1 and 3, maintenance of parental mean in NDM and WHS, increased PHM and L, and reduced AV in Factor 2 (Table 7).

Therefore, none of the factors selected progenies with all the characters in a favorable way, being not interesting for both populations.

Table 7. Predicted gains percentages estimates ($GS_{(\%)}$), selected population mean ($\bar{X}_S$) by direct selection in the factorial scores, and mean ($\bar{X}_S$) and confidence interval (CI) of original population for the characters evaluated in 162 $F_{2:3}$ of soybean genotypes in population 2, Piracicaba-SP, 2014.

	Factor		NDF	GP	NDM	PHM	L	LR
Pop. 2	1	$GS_{(\%)}$	-12.22	1.11	-11.70	-6.70	-4.40	-28.18
		$\bar{X}_S$	57.96	28.61	122.45	63.96	2.28	1.63
	2	$GS_{(\%)}$	2.15	-1.23	0.76	16.42	9.34	-4.16
		$\bar{X}_S$	68.39	26.98	141.55	84.52	2.95	2.49
	3	$GS_{(\%)}$	-10.79	5.20	-7.91	-14.64	-7.55	-13.26
		$\bar{X}_S$	59.00	31.45	128.26	56.90	2.13	2.16
Pop. 2 (original)		$\bar{X}$	66.83	27.84	140.39	69.92	2.49	2.64
		CI	(65.76; 67.90)	(27.41; 28.27)	(138.51; 142.26)	(68.03; 71.81)	(2.40; 2.58)	(2.50; 2.78)

	Factor		AV	GY	HSW	WHS	Total
Pop. 2	1	$GS_{(\%)}$	16.64	35.57	68.29	29.59	88.00
		$\bar{X}_S$	2.83	33.73	27.36	12.04	
	2	$GS_{(\%)}$	-2.63	14.97	12.44	1.06	49.12
		$\bar{X}_S$	2.14	27.32	16.79	8.59	
	3	$GS_{(\%)}$	10.86	13.27	34.64	23.53	33.35
		$\bar{X}_S$	2.62	26.79	20.99	11.31	
Pop. 2 (original)		$\bar{X}$	2.24	22.66	14.43	8.46	
		CI	(2.15; 2.33)	(21.21; 24.11)	(13.06; 15.81)	(8.01; 8.91)	

Number of days for flowering (NDF in days), grain filling period (GP in days), number of days to maturity (NDM in days), plant height at maturity (PHM in cm), lodging (L in notes), leaf retention (LR in notes), agronomic value (AV in notes), grain yield (GY in g), healthy seeds weight (HSW in g) and weight of one hundred seeds (WHS in g).

CONCLUSION

Factor analysis could not provide indexes for selection, in which all characters had an undesired response. Furthermore, selection in the "supercharacters" resulted in different select direction and undesired effects on the original variables, and were not ideal for the studied populations.

Therefore, complementary studies should be done with other selection indices in these populations, in order to establish an index in which all the characters are selected favorably and with satisfactory gains for the application in breeding programs.

ACKNOWLEDGMENTS

AUTHOR'S CONTRIBUTION

Conceptualization, Pinheiro J. B.; Methodology, Pinheiro J. B. and Bermudez F.; Investigation, Bermudez F.; Formal Analysis, Bermudez F.; Writing – Original Draft, Bermudez F.; Writing – Review and Editing, Pinheiro J. B.; Funding Acquisition, Pinheiro J. B.; Resources, Pinheiro J. B.; Supervision, Pinheiro J. B.

REFERENCES

Carrão-Panizzi, M. C. and Kitamura, K. (1995). Isoflavone content in Brazilian soybean cultivars. Japanese Journal of Breeding, 45, 295-300. https://doi.org/10.1270/jsbbs1951.45.295

Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. Acta Scientiarum Agronomy, 35, 271-276. https://doi.org/10.4025/actasciagron.v35i3.21251

Cruz, C. D. and Carneiro, P. C. S. (2006). Modelos biométricos aplicados ao melhoramento genético. 2 ed. Viçosa: Editora UFV.

DoVale, J. C., Fritsche-Neto, R. and Silva, P. S. L. (2011). Índice de seleção para cultivares de milho com dupla aptidão: minimilho e milho verde. Bragantia, 70, 781-787. https://doi.org/10.1590/S0006-87052011000400008

Gallon, M., Buzzello, G. L., Trezzi, M. M., Diesel, F., and Silva, H. L. (2016). Ação de herbicidas inibidores da PROTOX sobre o desenvolvimento, acamamento e produtividade da soja. Revista Brasileira de Herbicidas, 15, 232-240. https://doi.org/10.7824/rbh.v15i3.471

Gazzoni, D. L. (1998). Efeito de populações de percevejos na produtividade, qualidade da semente e características agronômicas da soja. Pesquisa Agropecuária Brasileira, 33, 1229-1237.

Lustosa, P. R., Zanuncio, J. C., Leite, G. L. D., and Picanço, M. (1999). Qualidade da semente e senescência de genótipos de soja sob dois níveis de infestação de percevejos (Pentatomidae). Pesquisa Agropecuária Brasileira, 34, 1347-1351. https://doi.org/10.1590/S0100-204X1999000800004

Manly, B. F. J. (1986). Multivariate Statistical Methods. London: Chapman & Hall.

Medeiros, L. and Megier, G. A. (2009). Ocorrência e desempenho de *Euschistus heros* (F.) (Heteroptera: Pentatomidae) em plantas hospedeiras alternativas no Rio Grande do Sul. Neotropical Entomology, 38, 459-463. https://doi.org/10.1590/S1519-566X2009000400003

Panizzi, A. R., Carrão-Panizzi, M. C., Bays, I. A., and Almeida, L. A. (1986). Danos por percevejos em genótipos de soja com semente pequena. Pesquisa Agropecuária Brasileira, 21, 571-577.

Pinheiro, J. B., Vello, N. A., Rossetto, C. J., and Zucchi, M. I. (2005). Potential of soybean genotypes as insect resistance sources. Crop Breeding and Applied Biotechnology, 5, 293-300.

Reis, G. G., Pereira, F. B., Granato, I. S. C., DoVale, J. C., and Fritsche-Neto, R. (2017). Tropical maize selection indexes genotypes for efficiency in use of nutrients: phosphorus. Revista Ceres, 64, 266-273. https://doi.org/10.1590/0034-737x201764030007

Rocha, F., Bermudez, F., Ferreira, M. C., Oliveira, K. C., and Pinheiro, J. B. (2014). Effective selection criteria for assessing the resistance of stink bugs complex in soybean. Crop Breeding and Applied Biotechnology, 14, 174-179. https://doi.org/10.1590/1984-70332014v14n3a26

[SAS] Statistical Analysis System. (2007). The SAS system for Windows, version 9.2. Cary: SAS Institute Inc. [Accessed 2020 Mar. 10]. Available at: https://www.sas.com/en_us/software/stat.html

Silva, J. P. G. F., Baldin, E. L. L., Souza, E. S., Canassa, V. F., and Lourenção, A.L. (2013). Characterization of antibiosis to the redbanded stink bug *Piezodorus guildinii* (Hemiptera: Pentatomidae) in soybean entries. Journal of Pest Science, 86, 649-657. https://doi.org/10.1007/s10340-013-0527-5

Souza, J. (1988). Métodos estatísticos nas ciências psicossociais: Análise de componentes principais e suas aplicações. Brasília: Thesaurus.

Ventura, M. U. and Pinheiro, J. B. (1999). Resistência a insetos. In: D. Destro, and R. Montalván (Org.). Melhoramento genético de plantas (p. 467-516). Londrina: UEL