

Methodological proposal for the characterization of accessions in Germplasm Banks using Generalized Procrustes Analysis applied to incomplete but connected trials

Propuesta metodológica para la caracterización de accesiones en Bancos de Germoplasma mediante Análisis de Procrustes Generalizado aplicado a ensayos incompletos pero conectados

Andrea Lina Lavalle ^{1*}, Raquel Defacio ², Mariano De Leo ³, Sergio Jorge Bramardi ⁴

Originales: *Recepción*: 03/09/2019 - *Aceptación*: 19/02/2021

ABSTRACT

Characterization of plant material conserved in germplasm banks allows the study and analysis of the genetic variability within a collection. When germplasm banks have a large number of accessions, field evaluation should be performed using assays with manageable accession subsets. Common checks connecting the different assays are required to compare these accession subsets. In this study, the Generalized Procrustes Analysis was proposed as a basis for obtaining a factorial plane where all individuals are projected. This technique is applied to genotypes common to all assays, iteratively generating scale factors and rotation matrices. Accessions only belonging to a given assay are considered supplementary elements. This proposal was illustrated using datasets of 54 maize accessions from the Pergamino Active Germplasm Bank of the Experimental Station at the Instituto Nacional de Tecnología Agropecuaria (INTA) in Argentina. The proposal achieved highly satisfactory results.

Keywords

three-way data • MANOVA • modeling biological process • plant genetic resources

-
- 1 Universidad Nacional del Comahue. Departamento de Estadística. Buenos Aires 1400. Neuquén. Argentina. * lavalleandrealina@gmail.com, andrea.lavalle@faea.uncoma.edu.ar.
 - 2 Estación Experimental Agropecuaria Pergamino. Instituto Nacional de Tecnología Agropecuaria (INTA).
 - 3 Universidad Nacional de General Sarmiento. Instituto de Ciencias.
 - 4 Universidad Nacional del Comahue. Departamento de Estadística. Centro de Investigaciones en Toxicología Ambiental y Agrobiotecnología del Comahue (CONICET - UNCo).

RESUMEN

La caracterización del material vegetal conservado en bancos de germoplasma permite el estudio y análisis de la variabilidad genética dentro de una colección. Cuando los bancos de germoplasma tienen una gran cantidad de entradas, la evaluación de campo debe realizarse utilizando ensayos en los cuales se evalúa un subconjunto de poblaciones manejable experimentalmente. Se requieren poblaciones testigo que conecten los diferentes ensayos para comparar estos subconjuntos de accesiones. En este estudio, se propuso utilizar el Análisis de Procrustes Generalizado como base para obtener un plano factorial donde se proyectan todos los individuos. Esta técnica se aplica a los genotipos que son comunes a todos los ensayos, para generar iterativamente factores de escala y matrices de rotación. Las accesiones que solo pertenecen a un ensayo dado se consideran elementos suplementarios. La propuesta se ilustró utilizando un conjunto de datos de 54 accesiones de maíz del Banco Activo de Germoplasma Pergamino de la Estación Experimental del Instituto Nacional de Tecnología Agropecuaria de Argentina, donde se obtienen resultados altamente satisfactorios.

Palabras clave

datos de tres vías • MANOVA • modelización de procesos biológicos • recursos fitogenéticos

INTRODUCTION

Germplasm banks play an important role in genetic resources conservation. Characterization of genetic resources is the process of describing accessions with respect to a particular set of characters (6). The evaluation process has several objectives: i) to measure the genetic variability of the studied group; ii) to establish the representativeness of the collection in relation to the genetic variability of species in a region or to intra-species genetic variability; and iii) to identify genes that can be used for research and practical purposes, such as biotic resistance (11). The evaluation of total conserved accessions is affected by limitations such as experimental plot areas, labor availability and logistics. These limitations call for trait evaluations on consecutive field assays, each of them with a subset of manageable number of accessions. Associated with the complete evaluation of a germplasm bank by agronomic and morphological descriptors, two considerations should be taken into account: a) to conduct consecutive assays for evaluating a large number of accessions, and b) to use multivariate statistical methods for analysis of set quantitative traits. Complete evaluation of accessions from a germplasm bank requires the inclusion of a set of checks or controls in all assays, performed over the seasons. These checks connect the assays and allow the comparison of the different accessions.

Taba *et al.* (1998) reported the results of field evaluations to develop a core subset of Caribbean maize (*Zea mays l.*) accessions from the maize germplasm bank of CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo). A total of 498 accessions were evaluated in two sets of 249 accessions each, and seven common checks. The authors present a combined analysis of these data sets. Firstly, they propose a mixed linear model including the following effects: assay, accession and their interaction, plus other design-associated factors. These effects were estimated and removed from the observed value of each entry for each trait, leaving only the genotype effect. Secondly, the adjusted trait means were used to cluster accessions into homogeneous groups through a multivariate analysis. Following this same approach, Reeb *et al.* (2007) developed a model to evaluate 145 maize accessions from the Active Germplasm Bank (BAP) of the Experimental Station at the Instituto Nacional de Tecnología Agropecuaria (EEA INTA) in Pergamino, Argentina. Both models have a high degree of unbalance and the interactions result in a large number of empty cells, making difficult the effects estimation. In addition, given that they are univariate models, they ignore relationships structure among variables within datasets.

On the other hand, data from every assay are arranged into matrices of accession by markers sharing, over time, a group of accessions characterized in all assays. This may be seen as a three-way data structure, where input matrices are incomplete but connected through common checks. The first way comprises the accessions, the second one, the characterization traits and the third one, the assays. Thus, a three-way data analysis should be a useful tool to deal with the multidimensional nature of the information.

Objective

To develop a multivariate methodology based on Generalized Procrustes Analysis (GPA) (8, 9, 25) to define similar accession groups in the context of the characterization assays, previously described.

MATERIALS AND METHODS

Plant material

Maize accessions were evaluated in three assays. Assay 1 was planted in Pergamino (Buenos Aires, Argentina) during the 2006/2007 season, and included 23 accessions of different racial forms: Cristalino Colorado (11); Cristalino Amarillo (1); Dentado Amarillo (6); and Dentado Blanco (8). Assay 2 was conducted in Ferré (Buenos Aires, Argentina) during the 2006/2007 season. This assay included 24 accessions of different racial forms: Cristalino Colorado (12); Avatí Morotí Ti (1); Dentado Amarillo (3); Dentado Blanco (8); and non-classifiable type (1). Seven common checks were included in both assays: four synthetic open pollinated (BS13p, Candelaria INTA, Payagua INTA and SP1234) and three accessions (ARZM14007 of Dentado Blanco, ARZM14023 of Dentado Amarillo and ARZM18004 of Cristalino Colorado). Assay 3 included all the 47 accessions of assays 1 and 2 and the seven common checks, and it was carried out in Ferré during the 2011/2012 season. Assay 3 was used as a reference for the comparative analysis of the methodology proposed. Random Complete Block Design with two replications was used for the three assays. Twelve morphological, agronomic and phenological traits were measured according to IBPGR descriptors (12): plant height (PL.HEIGHT-cm); ear height (E.HEIGHT-cm); kernel width (K.WIDTH-mm); kernel length (K.LENGTH-mm); number of kernel rows (N.ROW); ear diameter (E.DIAM-mm); 1000 kernel weight (W1000-gr); yield (YIELD-kg); prolificacy (PROL); root lodging (ROOT.L-%); days to anthesis (D.ANT) and days to silking (D.SILK) (3).

Principal Components Analysis on the concatenated table

Principal Component Analysis (PCA) adequately describes a set of n individuals and p variables through a small set of variables expressed as linear combinations of the original ones. In this way the information is optimally represented in a reduced dimensional space. In addition, the initial variables are usually correlated, whereas the new ones are not. This transformation facilitates data interpretation because it is possible to infer about linear relations between variables and similarities about individuals (5, 27). In this study, a single PCA matrix of 54 accessions by 12 variables was performed using the mean value of accessions over replications per assay and subsequent concatenation of the assays.

Principal Component Analysis on the matrix obtained after the estimation and elimination of model effects

A mixed linear model was adjusted for each variable according to Reeb *et al.* (2007) and Taba *et al.* (1998). This model firstly estimates the effect of each factor and then eliminates it from the value of each variable obtained for each accession:

$$y_{ikl} = \mu + G_i + E_k + B_{l(k)} + GE_{(ik)} + \mathcal{E}_{ikl}$$

where:

y_{ikl} = the observed value

μ = the general mean

G_i = the effect of the i -th accession $i=1, \dots, 54$

E_k = the effect of the k -th assay, $k=1,2$

$B_{l(k)}$ = the effect of the block-within-assay set, $l=1,2$

\mathcal{E}_{ikl} = the random error.

As a result, a matrix of 54 accessions by 12 adjusted variables was used to perform a PCA.

Principal Component Analysis on the reference assay

A matrix of individuals-by-variables was obtained using the average of the replications, based on data from the reference trial conducted in the 2011/2012 season. This matrix was used to perform the PCA.

Generalized Procrustes Analysis on incomplete but connected trials

An algorithm based on Generalized Procrustes Analysis (GPA) (8, 9, 25) was developed to generate a consensus configuration of the common checks from a multivariate approach. GPA can be applied to a set of individuals described by the same variables in different conditions. This technique searches for an optimal consensus configuration of the different individuals-by-variables data matrices. The consensus is obtained through a series of iterative algebraic steps that include translation, rotation, reflection and scaling of each individual configuration, optimizing a goodness-of-fit criterion. The latter relies on maintaining the relative distances among elements of the individual configurations and on minimizing the sums of squares between analogous points, *i.e.*, points that correspond to the same element under different configurations. After the initial standardization or the translations have been done and all the configurations have been transformed, an iteration is completed. The process is repeated until the change in the residual square sums between two consecutive steps is less than a particular value. The consensus configurations are obtained from the average of all individual transformed configurations. Once the iterative process of the GPA ends, the total variability can be partitioned in the form of a table of analysis of variance (ANOVA).

Some applications of this technique are related to measuring the consensus between the agronomic and molecular information (1, 3, 10, 18, 20, 22, 26). On the other hand, GPA can be applied in crop characterization and other agronomic objectives (4, 19, 23, 27).

The experimental situation under study can be analyzed using incomplete matrices which contain the information from q assays. In each k -th assay ($k=1 \dots q$), a set of $(n+n_k)$ accessions is assessed by p variables. Each set of $(n+n_k)$ accessions includes a set of n common checks measured under q conditions, and n_k other accessions, different in each k -th assay ($k=1 \dots q$). In matrix notation, information on the conditions is represented by q matrices X_k . They have both: n rows representing the individuals measured under all conditions (hereafter referred to as individuals in common), and n_k rows representing those individuals which correspond only to the k -th condition. The latter are unique to each particular condition and differ among conditions. Therefore, each X_k data matrix can be partitioned into two submatrices: X_k^C , of order $(n \times p)$ with the coordinates of the individuals in common in the k -th condition, and X_k^R , of order $(n_k \times p)$ with the coordinates of the individuals unique to the k -th condition (figure 1). A common space where to project all the set of individuals is required for analyzing the relationships between the $N = n + \sum_{k=1}^q n_k$ individuals. However, an assumption of the GPA is that all objects are measured under all conditions.

		q Conditions					
		X_1^C	X_2^C	...	X_k^C	...	X_q^C
N Individuals	X_1^R						
		X_2^R					
			...				
				X_k^R			
					...		
						X_q^R	
							X_q^R

Figure 1. Data structure. In each condition, the individuals corresponding to the k -th condition are represented by the matrix X_k^R , of order $(n_k \times p)$, while the individuals in common are represented by X_k^C , of order $(n \times p)$. The conditions are juxtaposed.

Figura 1. Estructura de los datos. En cada condición k los individuos pertenecientes a dicha condición se representan en la matriz X_k^R de orden $(n_k \times p)$ y los individuos comunes en la matriz X_k^C de orden $(n \times p)$. Las condiciones están yuxtapuestas.

Thus, the GPA-based algorithm must be applied on the n individuals in common, while the individuals unique to each condition are considered as supplementary elements (17). The n_k individuals in the k -th condition are centered, scaled and rotated through the same transformations applied for the individuals in common, but are excluded from the calculation of the parameters involved in these operations. The individuals in common act as pivots for the rest of the accessions, being the geometric reference for the successive rotations. Therefore, it is essential to select individuals in common that satisfy some requirements for obtaining reliable and accurate results.

The assumptions of the proposal are that individuals in common are clustered into groups showing significant differences between the considered traits and that these groups remain stable across conditions. These are verified by MANOVA, checking the absence of genotype-by-assay interaction and the presence of two or more significantly different groups. If these assumptions are verified, then the relationships between individuals in each condition can be analyzed through the set of individuals in common (figure 2). That is to say, individuals that belong only to the k -th condition X_k^R are compared with the individuals in common that correspond to this same condition, X_k^C . Then, using the GPA performed on the individuals in common, X_k^C is compared with $X_{k'}^C$ through the consensus configuration Y . Finally, individuals in common within the k' -th condition are compared with individuals corresponding only to this condition, $X_{k'}^R$.

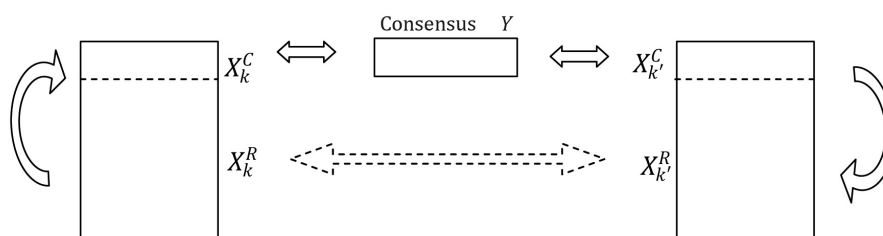


Figure 2: Flow scheme for comparing individuals in each condition through the consensus of individuals in common.

Figura 2: Esquema de flujo para comparar los individuos de cada condición a través del consenso de los individuos comunes.

In this multivariate solution, GPA is used as a basis for obtaining a factorial plane where all individuals are projected (7, 13, 14, 15, 17). Thus, individuals are grouped according to the similarity of their multivariate behavior (2).

A brief description of the steps involved in the methodological proposal is presented here:

- Fit a MANOVA model to the individuals in common, taking into account the main effects, the interactions and the design terms. The aim is to verify the assumptions of absence of genotype-by-assay interaction effect and presence of, at least, two significant different groups of the individuals in common (significant accession effect). The multivariate model is:

$$\mathbf{y}_{ikl} = \boldsymbol{\mu} + \mathbf{G}_i + \mathbf{E}_k + \mathbf{B}_{l(k)} + \mathbf{GE}_{(ik)} + \boldsymbol{\varepsilon}_{ikl} \quad (1)$$

where:

\mathbf{y}_{ikl} = the l -th multivariate replication of the p observable variables,

$\mathbf{y}_{ikl} = (y_{ikl1}, y_{ikl2}, \dots, y_{iklp})$;

$\boldsymbol{\mu}$ = the general mean;

\mathbf{G}_i = the effect of the i -th accession, $i=1, \dots, 7$;

\mathbf{E}_k = the effect of the k -th assay, $k=1, 2$;

$\mathbf{B}_{l(k)}$ = the block-within-trial set, $l=1, 2$;

$\boldsymbol{\varepsilon}_{ikl}$ = the random error.

- Generate the assay configurations X_k , $k=1\dots q$ by carrying out a PCA on each separate assay. Each X_k matrix is considered to be partitioned into X_k^C and X_k^R . The matrix X_k^C contains the coordinates of the individuals in common that belong to the k -th condition (controls). The matrix X_k^R contains the coordinates of those individuals belonging only to the k -th condition.

- Center each configuration X_k , $k=1\dots q$ on the gravity center of the individuals in common.
- Scale each matrix X_k , $k=1\dots q$, initialize the scale factors and calculate the initial residual sum of squares.

- Perform the rotation of the individuals in common. Compute the rotation matrix of each matrix X_k^C and use it to rotate the individuals only belonging to the X_k^R condition.

- Calculate the consensus configuration Y as the average of the configurations of the rotated individuals in common, and obtain a new residual sum of squares.

- If necessary, adjust the scaling factors and then recalculate the consensus configuration Y and the residual sums of squares.

- If the difference in the residual sum of squares between subsequent steps is greater than the set tolerance, go to step 5; else, concatenate into a single matrix X .

- Finally, perform a PCA on X to obtain the coordinates of the individuals.

The consensus configuration obtained was compared with those from the PCAs performed as described in (2.2), (2.3) and (2.4). For comparison, correlation coefficient between matrices of distances between individuals projected on the principal plane was calculated. These correlation coefficients were computed for each strategy. The proposed algorithm was implemented in an ad-hoc Matlab routine (16).

RESULTS AND DISCUSSION

Principal Components Analysis on the concatenated matrix

The principal plane obtained by applying the PCA on the concatenated matrix (2.2) explained 68.54% of total variability. The PCA plot (figure 3) shows a horizontal gradient from left to right of phenological traits, plant height, ear height, kernel width, kernel length and 1000 kernel weight, and in opposite direction: number of kernel rows. Yield and Prolificacy contributed to the vertical axis formation, establishing an upward gradient, whereas root lodging established a downward gradient.

The analysis of the individual's factorial plane (figure 4, page 41) shows that the accessions of Dentado Blanco are placed on the right end. These accessions include the plants of largest size and longest flowering cycle, with ARZM14066 and ARZM14090 standing out from the rest.

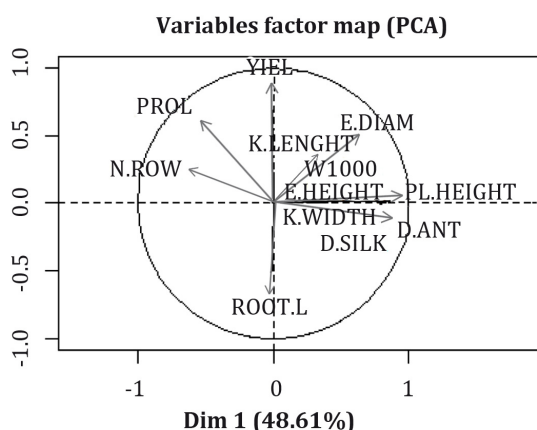


Figure 3. Projection of the variables on the correlation circle obtained performing a PCA after matrix concatenation.

Figura 3. Proyección de las variables en el círculo de correlación obtenido mediante un ACP sobre la matriz concatenada.

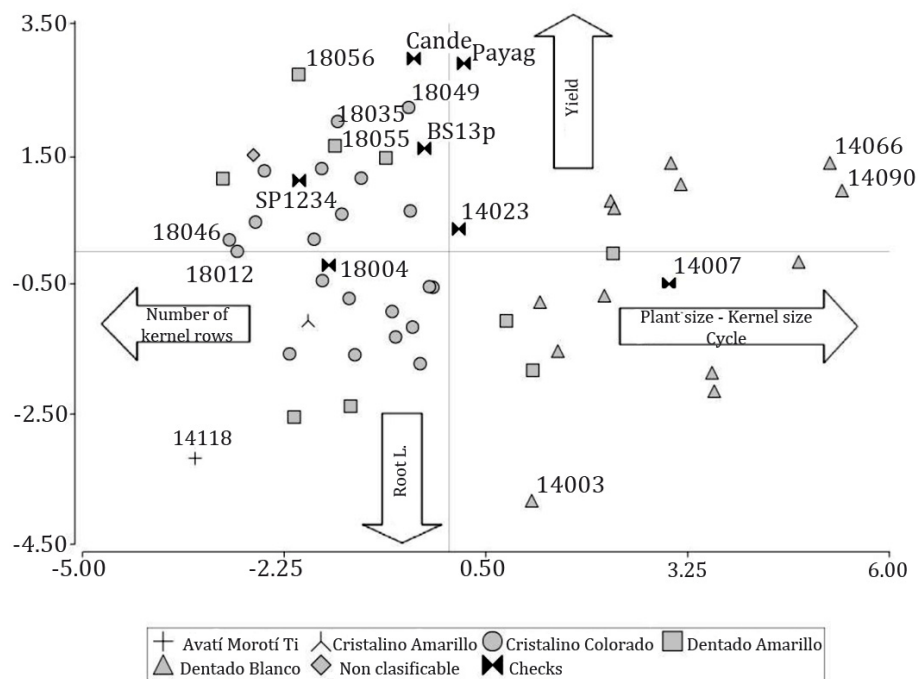


Figure 4. Configuration of individuals (maize landraces accessions) on the principal factorial plane obtained after matrix concatenation.

Figura 4. Configuración de los individuos (accesiones locales de maíz) sobre el plano principal obtenido luego de concatenar las tablas.

The accessions of Dentado Amarillo and Cristalino are placed on the left end. These accessions showed a larger number of kernel rows and a shorter flowering cycle than those of Dentado Blanco, particularly ARZM18046 and ARZM18012. The accessions having the highest yield values are situated on the upper end of the vertical axis. This group is composed by the synthetic varieties used as checks (Candelaria INTA, Payagua INTA, BS13p and SP1234), the accessions ARZM18055 and ARZM18056 of Dentado Amarillo and ARZM18035 and ARZM18049 of Cristalino Colorado. In contrast, the accessions ARZM14003 of Dentado Blanco and ARZM14418 of Avatí Morotí Ti are located at the lower end of the vertical axis, with the lowest yield and largest root lodging percentage.

The accessions of Dentado Amarillo are dispersed, showing the following distribution: the largest plants were grouped with the accessions of Dentado Blanco, those having high yield values were grouped with the synthetic varieties used as checks whereas those of low yield values are grouped with Avatí Morotí Ti. The accessions of Cristalino Colorado constitute an homogeneous group in terms of plant and kernel size and cycle duration, but they are separated into two subgroups in regard to high yield (upper left quadrant) and low yield values (lower left quadrant).

Principal Components Analysis on the matrix of adjusted means

The PCA principal plane obtained after the estimation and elimination of the model effects (2.3) explained 65.14% of the total variability. Traits behaved as in the previous analysis, except for number of kernel rows and prolificacy that contributed to a lesser extent to axis formation. In addition, ear diameter and yield contributed to form the vertical axis. Accessions were included in the same groups as in 3.1.

Principal Components Analysis on the reference trial

The PCA principal plane on data from the reference trial (2.4) explained 61.24% of total variability. Traits behaved as in previous analyses, except for number of kernel rows that contributed to the formation of the vertical axis. Accessions were grouped as before.

Generalized Procrustes Analysis on incomplete but connected trials

After verifying the assumptions, no genotype-by-environment interaction was detected and three groups of individuals in common revealed highly significant differences. Group 1: ARZM14007; Group 2: ARZM14023, ARZM18004, SP1234; and Group3: BS13p, Candelaria INTA, Payagua INTA. In step 2 of the presented methodological proposal, the replicates for each accession were averaged and a PCA was performed on each assay. Within the algorithm, and before the iterative process, both configurations were centered and scaled (Steps 3 and 4) and the initial residual sum of squares was calculated. Then, rotation and scaling were made until the convergence criterion was satisfied (Steps 5 to 8). The algorithm converged in two iterations (table 1).

Table 1. Residual sum of squares after rotation and scaling in two iterations.

Tabla 1. Sumas de cuadrados residuales luego de la rotación y el escalado en las dos iteraciones.

Iteration	SSR after rotation	SSR after scaling
0		0.316400
1	0.063110	0.063110
2	0.063108	0.063108

Figure 5 shows the configuration of individuals obtained by the methodology proposed in this research (Step 9). The accessions are grouped according to their race. Dentado Blanco landraces were separated from Dentado Amarillo and Cristalino along the first principal axis. The synthetic varieties are positioned at the upper end of the vertical axis, which is associated with yield. Accessions were grouped as in the other cases already described.

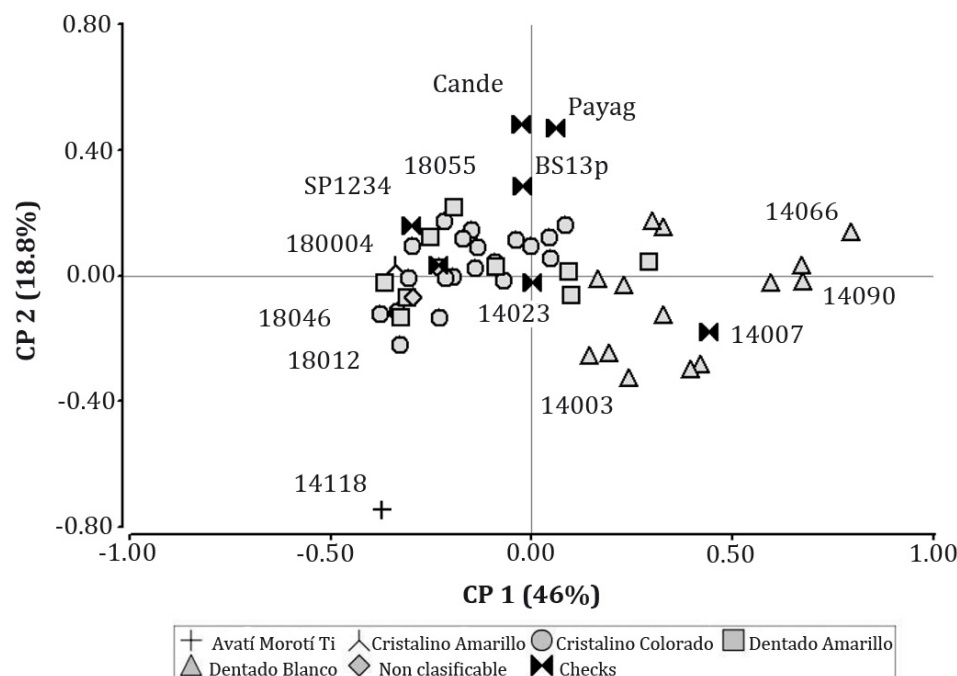


Figure 5. Principal Component Analysis plot of individuals belonging to trials 1 and 2 and consensus configuration of common checks, after applying the proposed methodology.

Figura 5. Representación en Componentes Principales de los individuos de los ensayos 1 y 2 y configuración consenso de los testigos comunes, luego de la aplicación de la propuesta.

Table 2 shows the correlation coefficients between matrices of distance among individuals obtained from the final configurations of the different strategies. The correlation coefficients indicate a high degree of agreement among them. The highest correlation coefficient value was observed between the presented methodological proposal and the strategy of model effects elimination. This was a very promising and encouraging result, since it may improve the way in which germplasm collections are characterized. So far, this goal has been performed by the elimination of model effects, a univariate methodological approach that requires a laborious process. This is due to the model being highly unbalanced, thus hindering the estimation of the effects. In this context the new proposal stands as a useful tool for data analysis in germplasm characterization and evaluation process, providing good results with easy implementation.

Table 2. Correlation coefficients between matrices of distance between individuals obtained from the configurations of the different strategies.

Tabla 2. Coeficientes de correlación entre matrices de distancia entre individuos de las configuraciones provenientes de cada estrategia.

	Concatenation	Elimination of effects	Reference trial	Proposal
Concatenation	1			
Elimination of effects	0.8788	1		
Reference trial	0.7724	0.8491	1	
Proposal	0.8518	0.9518	0.8202	1

The common checks were clustered into significant groups serving as gravity center for accessions that are similar to them. Figure 5 (page 42) shows that non-synthetic common checks were positioned close to the rest of the accessions belonging to their race. Common checks are required to be stable (without genotype-by-assay interaction) and to show a distinct behavior, so that they can serve as geometric reference for those accessions belonging only to a given assay

CONCLUSIONS

The objective of this study was to develop a multivariate methodology based on Generalized Procrustes Analysis (GPA) to define similar accession groups in the context of the characterization assays. The GPA is applied to a group of checks, common to all assays. The accessions that are only present in a given assay are considered as supplementary elements. Common checks must fulfill certain assumptions, *i.e.*, they are required not to interact with the evaluation assays and so to be stable across assays. The efficiency of the proposal was illustrated with seven common checks. After verifying the above-mentioned assumptions, the proposed technique was applied to obtain a factorial plane where all evaluated accessions are projected. This configuration was compared with those obtained from three strategies: concatenation of trials, estimation and elimination of the model effects and a reference assay. The correlations between matrices of distance among individuals reveal that the presented proposal provides similar results to those given by the currently used methodology. This traditional methodology is based on the estimation and elimination of effects, and has several disadvantages: a high degree of unbalance and many empty cells that complicate the estimates, in addition of being a univariate approach. These results support that the developed proposal is useful for identifying sets of accessions similar to those involved in germplasm evaluation trials, considering the multivariate structure of the data set.

REFERENCES

1. Bramardi, S.; Bernet, G.; Asíns, M.; Carbonell, E. 2005. Simultaneous agronomic and molecular characterization of genotypes via the generalised procrustes analysis: An application to cucumber. *Crop Sci.* 45: 1603-1609.
2. Bruno, C.; Balzarini, M. 2010. Ordenaciones de material genético a partir de información multidimensional. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 42(2): 183-200.
3. Defacio, R. 2009. Caracterización y evaluación de la variabilidad genética en poblaciones nativas de maíz (*Zea mays L.*) de la provincia de Buenos Aires en base a descriptores morfológicos y agronómicos. Tesis de Maestría en Genética Vegetal. Universidad Nacional de Rosario.
4. Del Médico, A.; Cabodevila, V.; Vitelleschi, M.; Pratta, G. 2020. Characterization of tomato generations according to a three-way data analysis. *Bragantia.* 79(1): 8-18.
5. Espósito, M. A.; Gatti, I.; Bermejo, C. J.; Cointry, E. L. 2020. Evaluation of a lentil collection (*Lens culinaris Medik*) using morphological traits and digital phenotyping. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 52(1): 1-13.
6. FAO. 2010. Segundo Informe sobre el estado de los Recursos Fitogenéticos para la alimentación y la agricultura en el mundo. Roma.
7. González, C.; Lavalle, A.; Bramardi, S.; Defacio, R. 2012. Programación en R para la obtención de una configuración consenso en ensayos multiambientales incompletos conectados. X Congreso Latinoamericano de Sociedades de Estadística. Córdoba.
8. Gower, J. 1975. Generalized Procrustes Analysis. *Psychometrika.* 40: 33-51.
9. Gower, J.; Dijksterhuis, G. 2004. *Procrustes Problems.* Oxford University Press.
10. Grandón, N.; Alarcón, Y.; Moreno, M.; Arolo, V.; Odorizzi, A.; Bruno, C.; Basigalup, D.; Gioco, J. Diversidad genética en genotipos de alfalfa (*Medicago sativa L.*) mediante marcadores SSR y características agronómicas. 2012. Ediciones INTA. <https://inta.gob.ar/documentos/diversidad-genetica-en-genotipos-de-alfalfa-medicago-sativa-l.-mediantemarcadores-ssr-y-caracteristicas-agronomicas>.
11. Hidalgo, R. 2003. Variabilidad genética y caracterización de especies vegetales. En Franco e Hidalgo (Eds): *Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos.* Boletín 8. IPGRI. Colombia.
12. IBPGR. 1991. Descriptors for maize. international maize and wheat improvement center, Mexico city/International board for plant genetic resources. Rome.
13. Lavalle, A.; Bramardi, S. 2009. Propuesta metodológica para la obtención de configuración consenso en ensayos multiambientales incompletos conectados. II Encuentro Iberoamericano de Biometría. Veracruz. México.
14. Lavalle, A.; Bramardi, S.; Defacio R. 2015a. Análisis Multivariado a tres vías aplicado a datos provenientes de ensayos con diseños desbalanceados. *Proceedings of V MACI 2015.* Asociación Argentina de Matemática Aplicada, Computacional e Industrial. Tandil. Argentina.
15. Lavalle, A.; Bramardi, S.; Defacio R. 2015b. Eficiencia de la metodología "Análisis de Procrustes Generalizado" para el tratamiento de datos a tres vías provenientes de diseños desbalanceados. I Congreso Argentino de Estadística. Buenos Aires. Argentina.
16. Lavalle, A.; Curia, L. 2016a. GPA for incomplete structures. Mathworks. File ID # 57300. <http://www.mathworks.com/matlabcentral/fileexchange/57300-gpa-for-incompletestructures>.
17. Lavalle, A.; Bramardi, S. 2016b. An algorithm based on generalized procrustes analysis to find the consensus of several configurations of individuals connected by common checks. *Advances and Applications in Statistics.* 49(1): 31-48.
18. Lorello, I.; García Lampasona, S.; Peralta, I. 2020. Genetic diversity of squash landraces (*Cucurbita maxima*) collected in Andean Valleys of Argentina. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 52(1): 293-313.
19. Matteucci, S.; Pla, L. 2010. Procrustes analysis as a tool for land management. *Ecological Indicators.* 10: 516-526.
20. Pocovi, M. I.; Collavino, N. G.; Gutiérrez, Á.; Taboada, G.; Castillo, V.; Delgado, R.; Mariotti, J. A. 2020. Molecular versus morphological markers to describe variability in sugar cane (*Saccharum officinarum*) for germplasm management and conservation. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 52(1): 40-60.
21. Reeb, P.; Bramardi, S.; Defacio, R. 2007. Estimación y tratamiento del efecto ambiente y sus interacciones en la caracterización de Recursos Genéticos en un Banco de Germoplasma. XI Conferencia Española y I Encuentro Iberoamericano de Biometría. Salamanca. España.
22. Ribotta, A.; López Colomba, E.; Bollati, G.; Strieker, G.; Carloni, E.; Griffa, S.; Quiroga, M.; Tommasino, E.; Grunberg, K. 2019. Agronomic and molecular characterization of Chlorisgayana cultivars and salinity response during germination and early vegetative growth. *Tropical Grasslands.* 7(1): 14-24.
23. Salles Pontes, D.; Silva Rosado, R.; Cruz, C.; Nascimento, M.; Cruz Oliveira, A.; Pensky, M. 2020. Traits election using Procrustes analysis for the study of genetic diversity in Conilon coffee. *Acta Scientiarum. Agronomy.* V. 42: e43195.

24. Taba, S.; Díaz, J.; Franco, J.; Crossa, J. 1998. Evaluation of Caribbean maize accession to develop a core subset. *Crop Science*. 38(5): 1378-1386.
25. ten Berge, J. 1977. Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*. 42: 267-276.
26. Vega, C.; Teich, I.; Acosta, M.; Lopez Lauenstein, D.; Verga, A.; Cosacov, A. 2020. Morphological and molecular characterization of a hybrid zone between *Prosopis alba* and *P. nigra* in the Chaco region of northwestern Argentina. *Silvae Genetica*. 69: 44-54.
27. Zuliani, R.; Bramardi, S.; Lavalle, A.; Defacio, R. 2012. Caracterización de poblaciones nativas de maíz mediante análisis de procrustes generalizado y análisis factorial múltiple. *Revista de la Facultad de Ciencias Agrarias*. Universidad Nacional de Cuyo. Mendoza. Argentina. 44(1): 49-64.

ACKNOWLEDGEMENTS

The authors wish to thanks Dra. Marta Zanelli, for the English text review and for her valuable technical advice. We are grateful to editors and the anonymous reviewers for the insightful comments that have improved this manuscript.

This work was supported by Secretaría de Ciencia y Técnica, Universidad Nacional del Comahue, and Instituto Nacional de Tecnología Agropecuaria, Argentina.