

# Assessment of phenotypic variability among EEA INTA Pergamino sunflower lines: Its relationship with the grain yield and oil content<sup>☆</sup>

Matías Domínguez<sup>1,2,\*</sup>, Stefanía Herrera<sup>3</sup> and Julio Horacio González<sup>1</sup>

<sup>1</sup> Instituto Nacional de Tecnología Agropecuaria, Estación Experimental Agropecuaria, Pergamino, Buenos Aires, Argentina

<sup>2</sup> Consejo Nacional de Investigaciones Científicas y Técnicas, Pergamino, Buenos Aires, Argentina

<sup>3</sup> Universidad Nacional del Noroeste de Buenos Aires, Pergamino, Buenos Aires, Argentina

Received 11 February 2021 – Accepted 11 May 2021

**Abstract** – The aims of the present study were to assess the phenotypic diversity among 221 sunflower accessions of INTA Pergamino Sunflower Breeding Program, to obtain discriminant functions that allow the classification of new accessions in similar groups and to evaluate the relationship between genetic distance pairwise accessions and hybrid performance for grain yield and oil content. We used 19 quantitative descriptors to evaluate phenotypic and morphological variability. Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were used to evaluate simultaneously all the variables and to describe phenotypic variation patterns of the germplasm. The distribution of germplasm in the dendrogram did not follow a clear pattern with regard to the predefined groups. This study revealed the variability observed among the lines that form the INTA Pergamino breeding program despite the highly selective forces applied to obtain inbred lines that produce superior hybrids for the Argentinean sunflower area. This work demonstrates the need for more in-depth study of genetic variability to be used as a predictor of heterosis in sunflower.

**Keywords:** sunflower / breeding / genetic variability / heterosis

**Résumé** – Évaluation de la variabilité phénotypique parmi les lignées de tournesol EEA INTA Pergamino : relation avec le rendement en grains et la teneur en huile. Les objectifs de la présente étude étaient d'évaluer la diversité phénotypique parmi 221 lignées de tournesol issues du programme de sélection de tournesol argentin à l'INTA Pergamino, afin d'identifier des fonctions discriminantes permettant de classer les nouvelles accessions dans des groupes similaires et d'évaluer la relation entre la distance génétique entre couples d'accessions et la performance des hybrides pour le rendement en grain et la teneur en huile. Nous avons utilisé 19 descripteurs quantitatifs pour évaluer la variabilité phénotypique et morphologique. Des analyses en Composantes Principales (ACP) et en Clusters Hiérarchiques (ACH) ont permis d'évaluer simultanément toutes les variables et de décrire les modèles de variation phénotypique du germplasm. La distribution du germplasm dans le dendrogramme n'a pas suivi un schéma clair par rapport aux groupes prédéfinis. Cette étude a démontré la variabilité observée parmi les lignées issues du programme de sélection de l'INTA Pergamino malgré les fortes pressions de sélection appliquées pour obtenir des parents d'hybrides performants et bien adaptés aux conditions argentines. Les résultats indiquent le besoin d'études plus approfondies de la variabilité génétique afin de l'utiliser pour prédire l'hétérosis chez le tournesol.

**Mots clés** : tournesol / sélection / variabilité génétique / hétérosis

<sup>☆</sup> Contribution to the Topical Issue "Sunflower / Tournesol".

\*Correspondence: [dominguez.matias@inta.gob.ar](mailto:dominguez.matias@inta.gob.ar)

## 1 Introduction

Sunflower (*Helianthus annuus*) is produced in more than fifty countries, and two thirds of its production is concentrated in Europe, including Ukraine, Russia and Turkey. Other major producing countries are Argentina, China, United States, and the South-Eastern parts of Africa. Sunflower oilseed came in fourth position on vegetable oils market in 2017/18 after palm, soybean, and rapeseed (Pilorgé, 2020).

Since 1931, Argentina has been working on sunflower breeding by exploiting the diversity of a broad range of international genetic resources in combination with introgressions of wild *Helianthus* species. At its Experimental Stations in Manfredi (Córdoba, Argentina) and Pergamino (Buenos Aires, Argentina), the Instituto Nacional de Tecnología Agropecuaria (INTA), has pioneered Argentinean sunflower breeding and has become one of the most prolific sunflower breeding groups in the country (Filippi *et al.*, 2020).

Multivariate analysis refers to a statistical technique that is widely used to analyze data which arise from more than one variable. Principal component analysis (PCA), hierarchical cluster analysis (HCA) and discriminant analysis (DA) are commonly employed multivariate techniques. PCA is one of the most useful methods employed in diversity studies. It reduces the dimensionality of the data while retaining most of the variation in the data set, and it is used to order genotypes according to their relationships while identifying the traits which explain most of the variability in the data set (Ringné, 2008; Dudhe *et al.*, 2019). HCA is a multivariate analysis widely used to assess relatedness and distance of any type of samples characterized by any type of descriptors. Therefore, it is used routinely to assess genetic diversity in germplasm collections (Peeters and Martinelli, 1989). DA maximizes between-group variability, minimizes within-group variability and provides a visual assessment of their relatedness. It also works as a predictive tool, enabling the classification of new accessions to previously characterized groups (Hernández *et al.*, 2019; Palacio *et al.*, 2020).

Knowledge of the amount and distribution of genetic diversity within sunflower breeding germplasm is essential to ensure the maintenance of the genetic variability of breeding pools, to make crop improvement more efficient through the directed accumulation of desirable alleles and to select the most diverse parental lines in crosses intended to generate breeding populations (Hladni *et al.*, 2017). For studies of genetic diversity in sunflower, morphological, physiological, biochemical, pedigree and molecular data have been used (Rama Subrahmanyam *et al.*, 2003; Dudhe *et al.*, 2019; Filippi *et al.*, 2015, 2020). Melchinger (1999) concluded that quantitative traits such as yield and heterotic response are expected to increase with parental genetic distance. Melchinger and Gumber (2015) defined a heterotic group as a group of related or unrelated genotypes from the same or different populations that exhibit similar combining ability when crossed with genotypes from other germplasm groups. According to Miller (1999) and Vear and Miller (1993) four heterotic groups are being utilized worldwide in sunflower breeding: a group of inbred female maintainer lines derived from the Russian open-pollinated varieties, a US restoration group, formed by crossing wild annual sunflower species with

lines of domesticated sunflowers, a grouping of Argentine germplasm and lastly a group of Romanian and South African female lines. The US restorer lines tend to be good sources for disease resistance and fertility restorer genes and the Argentine germplasm provides female parental lines with disease resistance and high grain yield. Also, French and Serbian inbred lines are important and widely used in sunflower seed industry.

The proposed heterotic groups in sunflower are less rigid than in other hybrid crops. The within group genetic distances appear to be greater and between group distances smaller than those observed in maize. Hongtrakul *et al.* (1997) also emphasized the need for further study of heterotic groups in sunflower, in their work on a group of public maintainer and restorer lines and using AFLP markers they found up to 4 subgroups within the restorer lines and 2 subgroups within the maintainer lines, suggesting that these could be defined as distinct heterotic groups. The heterotic groups are very useful for sunflower to catalog diversity and direct the process to introgress new alleles and create new heterotic groups (Cheres and Knapp, 1998; Cheres *et al.*, 2000; Lochner, 2011).

The aims of the present study were (i) to assess the diversity among a set of INTA Pergamino sunflower accessions using multivariate techniques, (ii) to compare the patterns of phenotypic variation obtained according to the different predefined groups of accessions, (iii) to obtain discriminant functions that allow the classification of new accessions and (iv) to evaluate the relationship between the genetic distance among accessions and the hybrid performance for grain yield and oil content.

## 2 Materials and methods

### 2.1 Sunflower accessions

A set of 221 inbred lines (maintainer and restorer) with desirable characteristics for subsequent use in hybrids and bred by the Sunflower Breeding Program of the Instituto Nacional de Tecnología Agropecuaria (INTA) in the Estación Experimental Agropecuaria (EEA) in Pergamino (Buenos Aires, Argentina) were included in this study. The only line that was included in this study and was not developed by INTA is the public line RHA 278. Five groups of accessions were previously determined according to the origin of the germplasm, the agronomic characteristics and if there are maintainer or restorer lines. The lines that form each group are shown in Supplementary Table 1. The characteristics of each group of lines and their origin are detailed in Table 1. For more information on the development of the lines released by INTA, their characteristics and selection strategies used, can be found in the work previously presented by González *et al.* (2015).

### 2.2 Field experiments

The germplasm characterization was conducted in the EEA INTA Pergamino (Buenos Aires, Argentina, 33° 56' S, 60° 33' W) during the 2014–2015, 2015–2016, 2016–2017, 2017–2018, 2018–2019 and 2019–2020 growing seasons (austral summer). All sunflower accessions were sown in three replicate plots per accession in fully randomized designs in each trial season. Each plot

**Table 1.** Characteristics and origin of each group of INTA Pergamino sunflower's lines.

Group	Characteristics	Origin
1	High content of oleic acid (maintainer lines)	<b>High content of oleic acid</b> lines were obtained from ND 01; VNNIMK 1646/ND 01; Sel HA 343; Luch/AO 01/03 and ND 01/BXC.
2	Maintainer lines with desirable traits such as grain yield, disease resistance and bird resistance	<b>Verticillium resistance</b> was originated from VNIIMK 1646; VNIIMK 1646/ND 01; SB/BZ BXC/KLM, BXC/KLM/HA 300, E/DXT; Local/Silvestre; P4 and HA 335/HA 412/ND 01/BZ. <b>Downy mildew resistance</b> was originated from E.75x MP 557/Negro Bellocq/KLM/HA 300; HA 335/HA412/71/538 /BXC; MP557/Negro Bellocq/KLM/HA 300/V112/E/BXC and MP557/Negro. From Bellocq/KLM/HA 300/V112/E/BXC were also obtained lines that were moderately resistant to <i>Verticillium</i> . <b>Sclerotinia resistance</b> was originated from MP557/Negro Bellocq/HA 89; ND 01; PGRK, V-13/BxC and AXB/BXC. From ND 01 was originated also high oleic lines and from AXB/BXC moderately <i>Verticillium</i> resistant lines. <b>Resistance to rust</b> was obtained from BXC and Pehuén/Caburé/Charata-SelRN160. <b>Resistance to birds</b> was originated from PGRK and sintética Os2.
3	Confectionary use lines	<b>Confectionery lines</b> were obtained from ND-NONOIL composite.
4	Restorer of fertility branched lines (oleic and linoleic)	<b>Restorer of fertility lines</b> were obtained from PRIII; Sel.105; Sel.105G; Sel.3; Sel.3107; Sel.34507; Sel.36506; Sel.379; Sel.70; Sel.731; Sel.860; Sel.871; Sel.90014 and Sel.9021.
5	Resistance to Imidazolinone	<b>Imidazolinone resistant lines</b> were obtained from IMI-Sun N.Dakota; B10/B83/HA425 and A83-3/Rfl. From Sel.2329 IMI-Sun N.Dakota and Sel.2330 IMI-Sun N.Dakota were originated <i>Verticillium</i> resistant lines.

consisted of one row 6 m long by 0.52 m wide, with a planting distance of 0.25 m, resulting in 24 plants per row. The dates of sowing in each season were in the first week of October.

A group of 70 A × R hybrids obtained from parental lines included in the collection of lines studied were evaluated in multi-environment trials (MET) in five locations of the Argentine sunflower region during the 2014–2015, 2015–2016, 2016–2017, 2017–2018 and 2018–2019 growing seasons (austral summer). The locations were Pergamino (33° 56' S, 60° 33' W), Bellocq (35° 54' S, 61° 32' W), San Carlos de Bolívar (36° 13' S, 61° 09' W), Coronel Suárez (37° 25' S, 61° 51' W) and Asturias (36° 10' S, 61° 52' W). The parental lines that form each hybrid included in the analysis are presented in Table 2. The A lines used to develop the hybrids were obtained by incorporating a unique source of cytoplasmic male sterility *Helianthus petiolaris* (CMS) PET1 gene (Leclercq, 1969) to the maintainer lines by backcrossing.

In each environment, a randomized incomplete block design with three replicates was used to test the 70 A × R hybrids. Environments were considered as random and hybrids as fixed. The trials were planted at 45 000 plants/ha. A plot size of 2 rows × 6 m and inter-row spacing of 0.70 m was used in all the trials. Planting took place within the normal sowing window (mid-October) at each location. Grain yield was determined by hand harvesting of 3.92 m<sup>2</sup> (both rows discarding border plants) and is presented at 11% moisture. Oil content was determined on a 10 g oven-dried achene sample Nuclear Magnetic Resonance (NMR) with an Oxford MQ5 equipment calibrated by solvent extraction (Grandlund and Zimmerman, 1975).

### 2.3 Quantitative descriptors

We measured phenotypic and morphological variability in four randomly selected plants from each row using 19 descriptors of the International Board for Plant Genetic Resources list (IBPGR, 1985) (Tab. 3).

DTF, PH, NL, SD, LL, LW, PL, CD, BL, NRF, RFW and RFL were measured in R5.5 (50% of the disk flowers have completed or are in anthesis) (Schneider and Miller, 1981). The % Oil was determined by Nuclear Magnetic Resonance (NMR) with an Oxford MQ5 equipment calibrated by solvent extraction (Grandlund and Zimmerman, 1975).

### 2.4 Statistical analysis

Descriptive and multivariate analyses were carried out to characterize the phenotypic diversity of sunflower accessions. The mean, range, standard deviation and coefficient of variation (CV) were obtained. Both statistical analyses were conducted using the statistical software R (R core Team, 2020).

PCA and HCA were used to simultaneously examine several variables for each accession and to describe phenotypic variation patterns in the germplasm characterized (Franco and Hidalgo, 2003). Principal component analysis was performed with standardized Euclidean distance obtained by Vegan package in R (Oksanen *et al.*, 2018) and depicted in a two-dimensional scatter plot using FactoMineR and factoextra R packages (Lê *et al.*, 2008; Kassambara and Mundt, 2020). For classification (HCA), the hierarchical agglomerative method with the average as fusion criteria was chosen using

**Table 2.** Parental lines of 70 hybrids tested in MET.

A-Line/ R-Line	IMI_05/04	IMI_13/03	IMI_13/04	Rf_00/03	Rf_00/04	Rf_00/07	Rf_00/08	Rf_00/09	Rf_00/10	Rf_00/11	Rf_00/13	Rf_00/17	Rf_00/18	Rf_15/03	Rf_17/01
AO_01/01		X					X							X	
AO_01/02		X									X				
AO_01/03	X					X								X	
AO_01/04	X		X			X						X		X	
AO_03/03			X			X					X				
AO_03/05							X								
AO_03/14			X			X					X				
AO_03/15					X	X					X				
AO_04/01		X					X								
AO_04/02			X					X			X		X		X
GP_02/04									X		X				
GP_02/05						X			X		X				
GP_03/02					X				X		X				
GP_06/04		X					X							X	X
GP_07/14													X		
SB_04/01			X			X								X	X
SB_04/02							X								

**Table 3.** Quantitative descriptors used.

Reference	Quantitative descriptor
DTF	Days to flowering from sowing
NAC	Number of achenes per capitulum
W100	Weight of 100 achenes (g)
Kernel (%)	Percentage of seed kernel
Oil (%)	% of oil content
AW	Achene width (mm)
AL	Achene length (mm)
AT	Achene thickness (mm)
PH	Plant height (cm)
NL	Number of leaves per plant
SD	Stem diameter (mm)
LL	Leave length (cm)
LW	Leave width (cm)
PL	Petiole length (cm)
CD	Capitulum diameter (central capitulum in branched lines) (cm)
BL	Bract length (cm)
NRF	Number of ray flowers
RFW	Ray flower width (mm)
RFL	Ray flower length (mm)

Euclidean distance derived from the principal components 1, 2 and 3 (PC1, PC2, PC3). The dendrogram was visualized and edited with *ape* and *factoextra* R packages (Paradis and Schliep, 2019; Kassambara and Mundt, 2020). The optimal number of clusters was determined using the *NbClust* R package (Charrad *et al.*, 2014).

Based on the groups obtained from the cluster analysis, a Linear Discriminant Analysis (LDA) was performed to define functions to discriminate between groups and to classify new accessions. Discriminant functions were obtained by splitting the original database in two subsets: the training data set (80%) and the testing data set (20%). Also, the variables that were highly correlated and provided redundant information were removed (AW, AT, LW and RFW). The testing data set was used to evaluate the accuracy of the LDA functions to classify new accessions into the groups. The LDA was performed using *MASS* R package (Venables and Ripley, 2002). The LD functions were graphed in a two-dimensional space using *ggplot2* R package (Wickham, 2016). The quality of the discriminant analysis was evaluated using the Wilks' Lambda statistic and was done using the basic R "manova" function.

The adjusted means for variables grain yield (GY) and Oil derived from multi-environment trials (MET) of the 70 A × R hybrids were regressed on Euclidean distance estimates obtained from the PC1, PC2 and PC3 to evaluate their association (Supplementary Table 2). The performance of these hybrids was also evaluated in relation to the group that was classified as the parental lines in the HCA. The relationship between the adjusted means of GY and Oil and the groups of classification of the parental lines were graphed in a violin plot using *ggplot2* R package (Wickham, 2016).

### 3 Results

Summary statistics for each quantitative descriptor are presented in (Tab. 4).

The PCA results showed that the first three components concentrate 43.7% of total variation. The first component (PC1) explained most of this variation (21.8%) and ordered the accessions according to a gradient of achene size, plant height, number of ray flowers, size of leaves (leave length and width) and weight of 100 achenes. In the opposite direction it ordered the accessions according to a gradient of oil content and kernel. As indicated in Figure 1, accessions with taller plants, wider capitulum, longer and wider achenes, longer petioles, numerous ray flowers and bigger leaves were placed in quadrants 1 (upper right) and 4 (lower right). Most of the confectionary lines were placed in these quadrants. Accessions with higher % of kernel and oil content were located in quadrants 2 (upper left) and 3 (lower left). The majority of restorer lines were in the quadrant 3.

The second component (PC2) explained 11.3% of variation and ordered the accessions according to the size of ray flowers, locating the accessions with the largest ray flowers in quadrants 1 and 2. In the same direction, it ordered the accessions with the most days to flowering, the highest number of leaves per plant and the number of achenes per capitulum. The accessions with the longer time to flowering, the largest number of achenes per capitulum and the number of leaves per plant were located in quadrants 1 and 2 (Fig. 1). All other components PC3 (10.6%), PC4 (9.8%) and PC5 (6.7%) explained less variation and did not differentiate between accessions. The biplot also showed that group 2 is the most diverse phenotypically and could discriminate between maintainer and restorer lines. Besides, it showed the positive associations between DTF and NAC, also positive associations were noticed between the variables of dimension of achenes and between the variables obtained on ray flowers. The vector of Oil showed an opposite direction to the vectors of the achene dimension variables.

The dendrogram resulting from HCA formed at a cut-off of 5.0 Euclidean distance showed three groups (S1, S2 and S3) and two un-grouped accessions (Fig. 2). The smallest group (S1) contained 23 accessions, the second (S2) has 45 and the third one (S3) 151 accessions. The main cluster was formed primarily by maintainer lines from group 1 and 2 and restorer lines from group 4. The second cluster was formed mainly by accessions derived from group 2 and in minor proportion by lines derived from group 5 and restorer lines. Most of the confectionary lines were concentrated in the smallest cluster. The HCA showed that group 2 was the predominant group in all clusters.

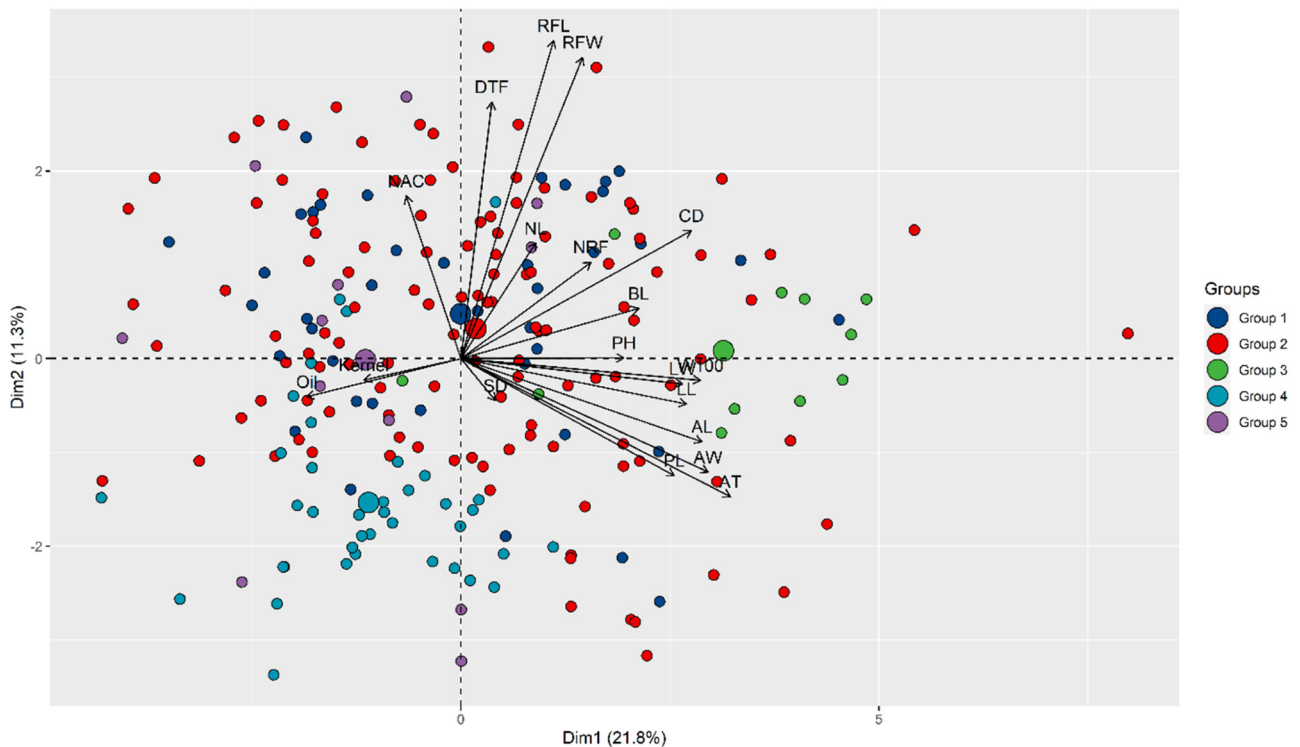
The LDA performed on the three clusters defined in the HCA showed that the discriminant functions obtained were significant in the Wilks test ( $p$ -value < 0.001) (Fig. 3). The two LDA functions obtained were:

$$LDA1 = 0.068(DTF) - 0.002(NAC) - 0.099(W100) + 0.065(Kernel) + 0.001(Oil) - 0.401(AL) - 0.012(PH) - 0.002(NL) + 0.051(SD) - 0.016(LL) - 0.076(PL) + 0.024(CD) - 0.034(BL) - 0.035(NRF) - 0.048(RFL),$$

$$LDA2 = -0.022(DTF) + 0.002(NAC) - 0.340(W100) + 0.059(Kernel) + 0.128(Oil) - 0.321(AL) + 0.005(PH) + 0.041(NL) - 0.011(SD) + 0.050(LL) + 0.005(PL) + 0.115(CD) - 0.164(BL) - 0.013(NRF) - 0.011(RFL).$$

**Table 4.** Summary statistics for quantitative descriptors included in this study.

Descriptor	Mean	SD	CV	Range
Days to flowering from sowing	71.83	7.90	11.00	57–99.5
Number of achenes per capitulum	581.59	197.83	34.01	135–1650
Weight of 100 achenes (g)	5.29	1.27	23.92	2.2–10
Percentage of seed kernel	73.18	5.35	7.31	29.6–84.6
% of oil content	46.53	5.42	11.64	29–57.5
Achene width (mm)	4.76	0.89	18.69	2.64–9.1
Achene length (mm)	9.57	1.26	13.19	5.44–13.26
Achene thickness (mm)	2.86	0.72	25.03	0.5–5.38
Plant height (cm)	131.75	24.17	18.35	71.25–185
Number of leaves per plant	26.98	6.67	24.73	15.33–72
Stem diameter (mm)	20.89	3.88	18.54	12–32.23
Leave length (cm)	18.79	3.45	18.35	9.63–28.63
Leave width (cm)	17.64	3.80	21.52	8.7–33.71
Petiole length (cm)	10.02	2.78	27.75	3.63–18.35
Capitulum diameter (cm)	13.57	3.41	25.12	6–24.25
Bract length (cm)	4.70	1.12	23.93	1.1–8.07
Number of ray flowers	42.94	8.91	20.76	10.67–72.5
Ray flower width (mm)	13.25	3.96	29.90	6–25.67
Ray flower length (mm)	57.62	13.81	29.89	23.4–89.75

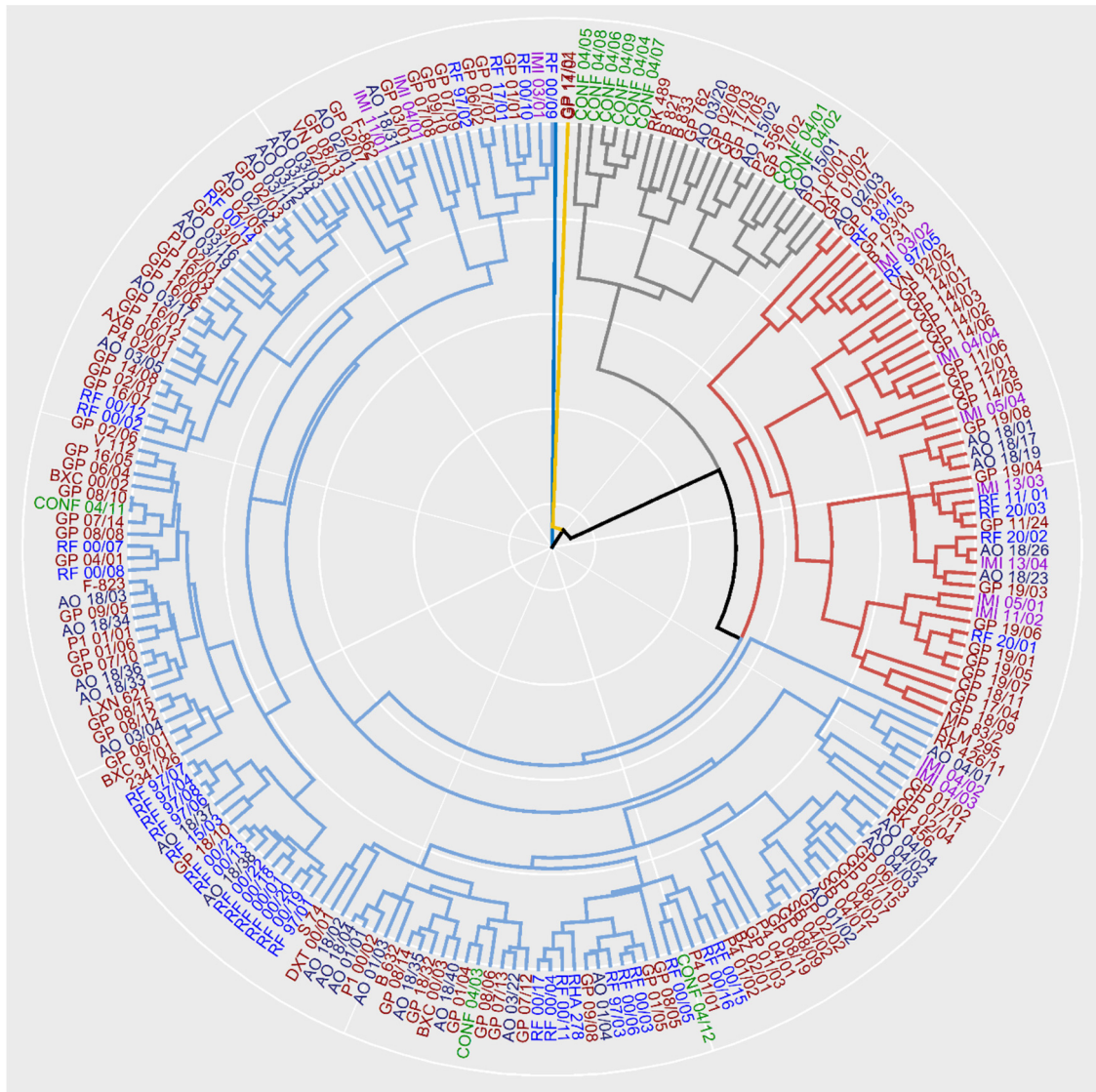


**Fig. 1.** Biplot of EEA INTA Pergamino sunflower lines.

The variables that differentiated best between accession groups were achene length and weight of 100 achenes, because they obtained the highest coefficients in both functions. The discriminant function LDA1 explained 78.8% of the variability while the function LDA2 accounted for the rest (Fig. 3).

The predictive capacity of the LDA functions obtained was evaluated through the testing data set by using a confusion matrix, showing an accuracy of 89%.

The associations between the adjusted means of GY and Oil and Euclidean distance were not significant. The simple



**Fig. 2.** Hierarchical cluster for EEA INTA Pergamino sunflower lines.

linear regression coefficient of Oil on Euclidean distance was 0.14 and the regression coefficient of GY on Euclidean distance was 16.27. Neither regression models explained the variability of the data, both multiple  $R^2$  were near to 0.

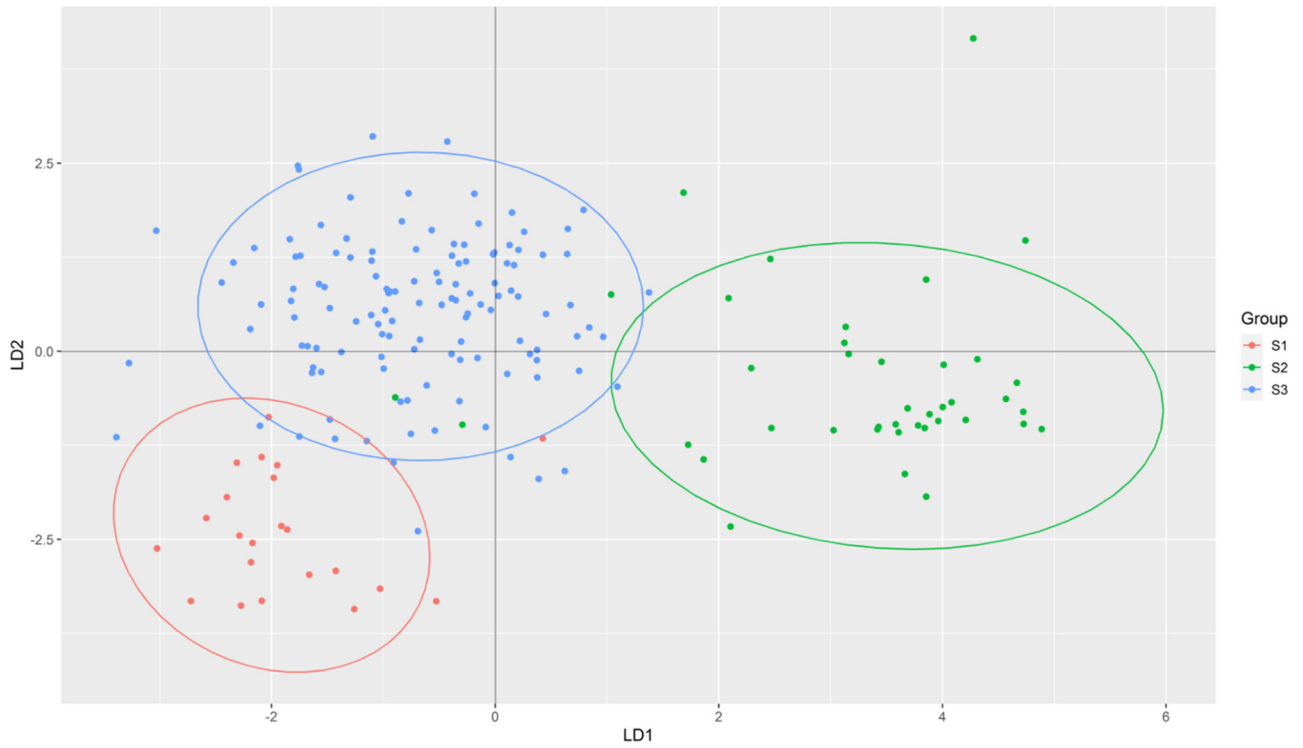
The adjusted means of the variable Oil for the 70 hybrids evaluated in MET showed no differences associated with the parental group classification obtained by HCA. The highest values were obtained by the crosses between lines belonging to the S3 group. On the other hand, for GY the values obtained for crosses between lines A belonging to group S2 and lines R belonging to group S3 were higher than the grain yields obtained between crosses of lines belonging to the same group. However, it can be seen that crosses of lines A of group S3 with lines R of group S2 showed the lowest yields (Fig. 4).

## 4 Discussion

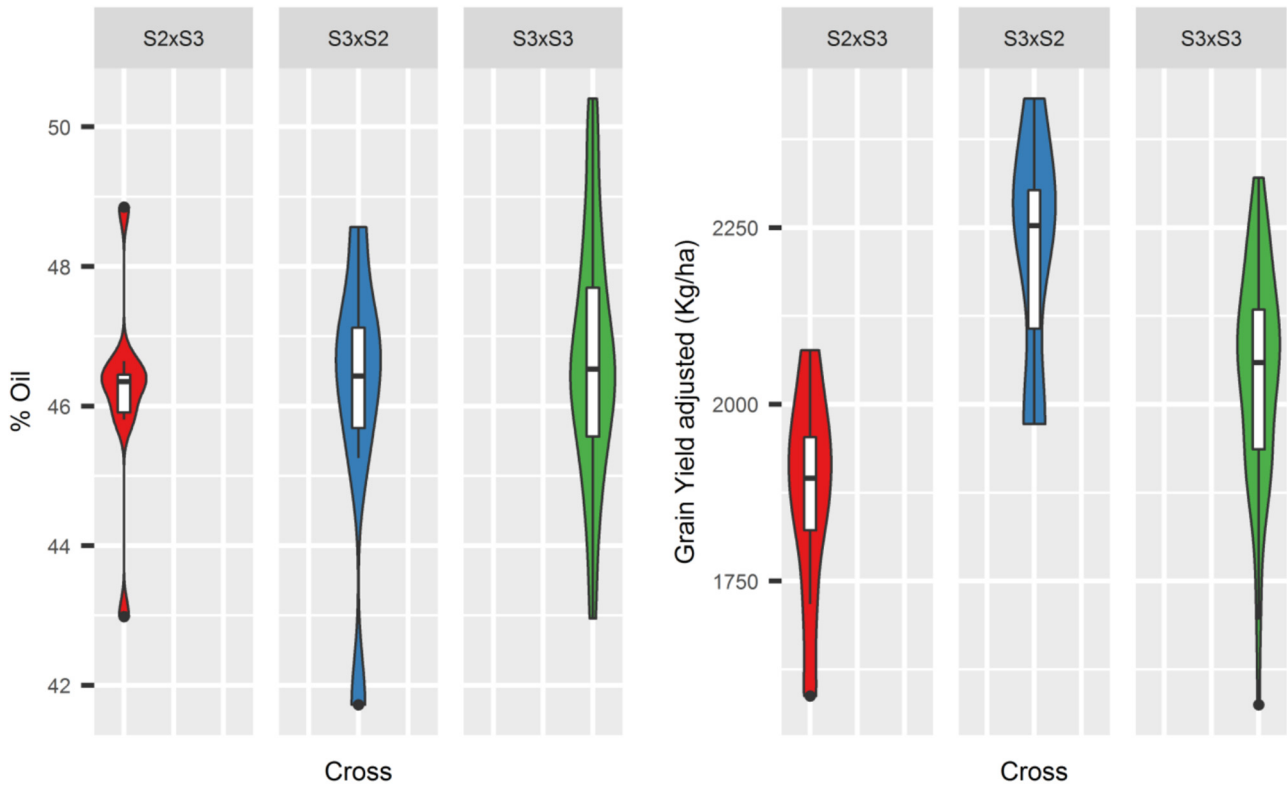
Germplasm collections are valuable resources for crop breeding. To exploit their potential, it is essential to characterize

them phenotypically and genotypically to understand the genetic diversity available in order to maximize the genetic gain in the breeding process.

PCA identified the traits which contributed most to explanation of the variability in the data. In this study, the variables related to achene size play an important role in the arrangement of accessions, with the greatest capacity to discriminate between them. Terzić *et al.* (2020) in a phenotypic study of the UGA-SAM1 population also reported achene size among the descriptors with the greatest power of genotype discrimination. Dudhe *et al.* (2019), in a diversity study on a germplasm bank collection, reported that oil content was one of the traits that contributed most to the variability of the data, whereas in our work the results were not similar. This difference can be explained by the limited range of variation of this trait in the set of lines selected for this study, since most of them were selected to increase oil content and they do not constitute a germplasm bank collection. DTF was one of the variables with greatest contribution to the variability explained by the PCs. Similarly, Ayaz *et al.* (2014) reported that DTF was



**Fig. 3.** Discriminant analysis scatter plot.



**Fig. 4.** Relationship between the adjusted means of oil and grain yield for the 70 hybrids evaluated in MET in relation to the classification cluster of the parental lines.



one of the variables with more incidence in the variability of the studied germplasm. In addition, the biplot clearly identified in opposite quadrants the restorer and maintainer lines, defined by their different characteristics in terms of oil content, flowering time and achene size. Similar results were expressed by Terzić *et al.* (2020) taking also into account the branching trait.

The PCA allowed us to interpret in a simple form the information collected through the different descriptors used, to identify groups of materials with certain phenotypic characteristics to direct future crosses, to identify the traits with less variability and to facilitate the search for specific characteristics in the materials available from the breeding program.

The distribution of germplasm in the dendrogram did not follow a clear pattern with regard to the predefined groups at the beginning of the work. The largest cluster included most of the restorer lines. Other population studies in sunflower germplasm collections reported the preponderance of this characteristic in cluster delimitation (Mandel *et al.*, 2011; Cadic *et al.*, 2013; Filippi *et al.*, 2015; Terzić *et al.*, 2020). This distinction between maintainer and restorer lines would be expected due to the breeding strategy of keeping these groups as separate heterotic pools to maximize heterosis in hybrid crosses (Fick and Miller, 1997). The close relationship between the restorer lines can be understood from their common ancestor, the RHA 274 line, which has the *Rfl* gene, one of the main sources of fertility restorer genes (Cadic *et al.*, 2013). In addition, the similarity between the restorer lines studied is justified by the fact that all of them are branched. The smallest group of accessions is composed of confectionary lines that are characterized by their achene size. These lines represent a small group in this study, their characteristics are different from the rest of the accessions due to the different objectives of confectionary sunflower breeding with respect to oil sunflower breeding. These results revealed the variability of the lines that form the INTA Pergamino sunflower breeding program despite the high selective forces applied to obtain inbred lines for superior hybrids production.

The discriminant analysis provided functions that allow us to classify the accessions in three groups that will allow a more efficient use of the genetic resources and the preservation of the genetic variability of the germplasm. In addition, it is a low-cost tool to make a better use of the phenotypic information generated in each growing season. In the same sense, Terzić *et al.* (2020) highlight the importance of using regularly collected phenotypic data in breeding programs. They propose to include as part of the germplasm selection criteria for breeding, traits related to leaf variation, seed coloration and certain flower traits, in order to preserve the variability of the populations used in breeding. Although the clusters defined by the HCA and used in the LDA differed from the groups mentioned at the beginning of the paper, this might be due to the fact that both analyses were performed using the information provided only by the morphological descriptors.

Developing hybrids is a costly long-term process; therefore, being able to predict heterosis is desirable. There are numerous papers in the literature on the relationship between genetic distance and heterosis or the response of traits such as yield and oil content. In sunflower, Cheres *et al.* (2000) found no satisfactory results in predicting grain yield as a function of genetic distance using molecular markers and

coefficients of coancestry. Similarly, Reif *et al.* (2013) reported no correlation between the genetic distance obtained through a genomic matrix of 572 AFLP and grain yield, oil yield and oil content for a set of A × R hybrids. In contrast, Darvishzadeh (2012) reported high correlations between distances obtained from morphological and molecular data and the yield of sunflower F1 hybrids. Hladni *et al.* (2018) in a study carried out on interspecific crosses of sunflower, positive associations were demonstrated between the genetic distance (obtained from a matrix of 37 SSR and 1 STS markers) and grain yield. The results obtained in our work indicate no relationship between the distances obtained and the performance of the hybrids we evaluated. In other crops such as chickpea, cotton and corn, low correlations have also been found between the distance of the parental genotypes and the hybrid performance (Ajmone Marsan *et al.*, 1998; Meredith and Brown, 1998; Sant *et al.*, 1999). In addition, Dias dos Santos *et al.* (2004) suggested that the inconsistent results obtained for the relationship between genetic distance and heterosis could be explained by the narrow genetic base of the germplasm used. Considering the genomic tools currently available, there is a clear need to deepen research on the relationship between genetic distance and heterosis in order to make the parental selection in breeding programs more efficient.

The higher oil content obtained by the hybrids whose parental lines belong to the S3 group can be explained by the additive action of the genes that control this characteristic. Similar results were obtained by Reif *et al.* (2013) evaluating a group of hybrids formed by crosses between and within heterotic groups (A × B, A × R and R × R), that did not present heterosis for oil content for any of the crosses. On the other hand, for grain yield, it was determined that hybrids obtained from the cross between lines of the group 2 and 3 had a greater grain yield. This tendency can be explained by the phenomenon of heterosis that is enhanced by the greater genetic distance between the lines of these groups. The fact that the reciprocal crosses (A lines of S2 × R lines of S3) showed a lower yield suggests the need for a further definition of the germplasm groups. The inclusion of molecular data in future studies on this set of lines and more extensive testing of the hybrids would probably allow a more precise differentiation of the groups.

In summary, this work points out the need for more in-depth study of genetic variability to be used as a predictor of heterosis. Future genotypic characterizations of the sunflower accessions of the EEA Pergamino germplasm are proposed to develop predictive models that allow a greater efficiency in the selection of parental lines for the production of superior hybrids exploiting heterosis. Also, the results obtained confirm the importance of INTA EEA Pergamino's sunflower breeding program in the development of improved, locally adapted and diverse materials as valuable resources for sunflower research.

## Supplementary Material

Supplementary Table 1.  
Supplementary Table 2.

The Supplementary Material is available at <http://www.ocl-journal.org/10.1051/oc/2021021/olm>.

**Acknowledgements.** We are grateful to the Instituto Nacional de Tecnología Agropecuaria for the financial support, to Juan Carlos Righetti and Gabriel Colacilli of the INTA Pergamino Sunflower Breeding Group. We are also grateful to Rumbidzai Matemba-Mutasa for the English revision, María Lujan Farace for the support in the statistical analysis and the anonymous reviewers for the valuable comments that contributed to the improvement of the manuscript.

## References

- Ajmone Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor Appl Genet* 96(2): 219–227. <https://doi.org/10.1007/s001220050730>.
- Ayaz U, Khan MF, Bashir S. 2014. Investigation of genetic divergence in local sunflower hybrids and inbred lines by applying morphological markers. *Int J Agron Agric Res* 5(2): 154–163.
- Cadic E, Coque M, Vear F, *et al.* 2013. Combined linkage and association mapping of flowering time in sunflower (*Helianthus annuus* L.). *Theor Appl Genet* 126(5): 1337–1356. <https://doi.org/10.1007/s00122-013-2056-2>.
- Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. Nbclust: An R package for determining the relevant number of clusters in a data set. *J Stat Softw* 61(6): 1–36. <https://doi.org/10.18637/jss.v061.i06>.
- Cheres MT, Knapp SJ. 1998. Ancestral origins and genetic diversity of cultivated sunflower: coancestry analysis of public germplasm. *Crop Sci* 38(6): 1476–1482. <https://doi.org/10.2135/cropsci1998.0011183x003800060012x>.
- Cheres MT, Miller JF, Crane JM, Knapp SJ. 2000. Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. *Theor Appl Genet* 100(6): 889–894. <https://doi.org/10.1007/s001220051366>.
- Darvishzadeh R. 2012. Phenotypic and molecular marker distance as a tool for prediction of heterosis and F1 performance in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. *Austral J Crop Sci* 6(4): 732–738.
- Dias dos Santos LA, de Toledo Picoli EA, Barros Rocha R, Alfenas AC. 2004. A priori choice of hybrid parents in plants. *Genet Mol Res: GMR* 3(3). [www.funpecrp.com.br](http://www.funpecrp.com.br).
- Dudhe MY, Mulpuri S, Meena HP, Ajjanavara RRG, Kodeboyina VS, Adala VR. 2019. Genetic variability, diversity and identification of trait-specific accessions from the conserved sunflower germplasm for exploitation in the breeding programme. *Agric Res*. <https://doi.org/10.1007/s40003-019-00406-w>.
- Fick G, Miller J. 1997. Sunflower breeding. In: Schneider AA, ed. Sunflower technology and production. Madison, Wisconsin, USA.
- Filippi CV, Merino GA, Montecchia JF, *et al.* 2020. Genetic diversity, population structure and linkage disequilibrium assessment among international sunflower breeding collections. *Genes* 11(3). <https://doi.org/10.3390/genes11030283>.
- Filippi CV, Aguirre N, Rivas JG, *et al.* 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol* 15(1). <https://doi.org/10.1186/s12870-014-0360-x>.
- Franco TL, Hidalgo R. 2003. Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos. *IBPGR* 89. [https://www.biodiversityinternational.org/fileadmin/\\_migrated/uploads/tx\\_news/Análisis\\_estadístico\\_de\\_datos\\_de\\_caracterización\\_morfológica\\_de\\_recursos\\_fitogenéticos\\_894.pdf](https://www.biodiversityinternational.org/fileadmin/_migrated/uploads/tx_news/Análisis_estadístico_de_datos_de_caracterización_morfológica_de_recursos_fitogenéticos_894.pdf).
- González J, Mancuso N, Alvarez D, Cordes D, Vázquez A. 2015. Contribution of the Argentine germplasm to the improvement of sunflower. *Helia* 38(62): 121–140. Institute of Field and Vegetable Crops. <https://doi.org/10.1515/helia-2014-0026>.
- Grandlund M, Zimmerman DC. 1975. Effects of drying conditions on oil content of sunflower (*Helianthus annuus* L.) seeds as determined by wide-line nuclear magnetic resonance (NMR). *Proc North Dakota Acad Sci* 27: 128–132.
- Hernández F, Presotto A, Poverene M, Mandel JR. 2019. Genetic diversity and population structure of wild sunflower (*Helianthus annuus* L.) in Argentina: reconstructing its invasion history. *J Hered* 110(6): 746–759. <https://doi.org/10.1093/jhered/esz047>.
- Hladni N, Terzić S, Mutavdžić B, Zoric M. 2017. Classification of confectionary sunflower genotypes based on morphological characters. *J Agric Sci* 155(10): 1594–1609. <https://doi.org/10.1017/S0021859617000739>.
- Hladni N, Zorić M, Terzić S, *et al.* 2018. Comparison of methods for the estimation of best parent heterosis among lines developed from interspecific sunflower germplasm. *Euphytica* 214(7). <https://doi.org/10.1007/s10681-018-2197-0>.
- Hongtrakul V, Huestis GM, Knapp SJ. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. *Theor Appl Genet* 95(3): 400–407. <https://doi.org/10.1007/s001220050576>.
- IBPGR - Int. Board Plant Genet. Resour. 1985. Sunflower descriptors. [https://www.biodiversityinternational.org/fileadmin/user\\_upload/Descriptors\\_sunflower.pdf](https://www.biodiversityinternational.org/fileadmin/user_upload/Descriptors_sunflower.pdf).
- Kassambara A, Mundt F. 2020. Factoextra: extract and visualize the results of multivariate data analyses. <https://cran.r-project.org/package=factoextra>.
- Lê S, Josse J, Husson F. 2008. FactoMineR: An R package for multivariate analysis. *J Stat Softw* 25(1): 18.
- Leclercq P. 1969. Une sterilité mâle cytoplasmique chez le tournesol. *Ann Amel Plantes* 19: 99–106. <https://ci.nii.ac.jp/naid/10005384804>.
- Lochner TC. 2011. Prediction of heterotic groups and hybrid performance in South African Sunflower (*Helianthus annuus* L.) germplasm using SSR analysis. (Issue November). South Africa: University of the Free State.
- Mandel JR, Dechaine JM, Marek LF, Burke JM. 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theor Appl Genet* 123(5): 693–704. <https://doi.org/10.1007/s00122-011-1619-3>.
- Melchinger AE. 1999. Genetic diversity and heterosis. In: Genetics and exploitation of heterosis in crops, pp. 99–118. <https://doi.org/10.2134/1999.geneticsandexploitation.c10>.
- Melchinger AE, Gumber RK. 2015. Overview of heterosis and heterotic groups in agronomic crops, Issue 25, pp. 29–44. <https://doi.org/10.2135/cssaspecpub25.c3>.
- Meredith WR, Brown JS. 1998. Heterosis and combining ability of cottons originating from different regions of the United States. *J Cotton Sci* 2(2): 77–84.
- Miller JF. 1999. Oilseeds and heterosis. In: Genetics and exploitation of heterosis in crops, pp. 399–404. <https://doi.org/10.2134/1999.geneticsandexploitation.c37>.
- Oksanen AJ, Blanchet FG, Friendly M, *et al.* 2018. Vegan: Community Ecology Package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists, Issue 25(1). <https://cran.r-project.org/package=veganER->

- Palacio F, Apodaca M, Crisci J. 2020. ANÁLISIS MULTIVARIADO PARA DATOS BIOLÓGICOS Teoría y su aplicación utilizando el lenguaje R. Vazquez Mazzini Editores.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.
- Peeters JP, Martinelli JA. 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theor Appl Genet* 78(1): 42–48. <https://doi.org/10.1007/BF00299751>.
- Pilorgé E. 2020. Sunflower in the global vegetable oil system: situation, specificities and perspectives. *OCL* 27(1): 34. <https://doi.org/10.1051/ocl/2020028>.
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. <https://www.r-project.org/>.
- Rama Subrahmanyam SVRM, Sudheer Kumar S, Ranganatha ARGR. 2003. Genetic divergence for seed parameters in sunflower (*Helianthus annuus* L.). *Helia* 26(38): 73–80. <https://doi.org/10.2298/hel0338073s>.
- Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V. 2013. Genomic prediction of sunflower hybrid performance. *Plant Breed* 132(1): 107–114. <https://doi.org/10.1111/pbr.12007>.
- Ringnér M. 2008. What is principal component analysis? *Nat Biotechnol* 26(3): 303–304. <https://doi.org/10.1038/nbt0308-303>.
- Sant VJ, Patankar AG, Sarode ND, *et al.* 1999. Potential of DNA markers in detecting divergence and in analysing heterosis in Indian elite chickpea cultivars. *Theor Appl Genet* 98(8): 1217–1225. <https://doi.org/10.1007/s001220051187>.
- Schneiter AA, Miller JF. 1981. Description of sunflower growth stages 1. *Crop Sci* 21(6): 901–903. <https://doi.org/10.2135/cropsci1981.0011183x002100060024x>.
- Terzić S, Zorić M, Seiler GJ. 2020. Qualitative traits in sunflower breeding: UGA-SAM1 phenotyping case study. *Crop Sci* 60(1): 303–319. <https://doi.org/10.1002/csc2.20059>.
- Vear F, Miller JF. 1993. Sunflower. In: Traditional crop breeding practices: an historical review to serve as a baseline for assessing the role of modern biotechnology, pp.95–111. <https://doi.org/10.2134/agronmonogr36.c17>.
- Venables WN, Ripley BD. 2002. Modern applied statistics with S, 4th ed. Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag. <https://ggplot2.tidyverse.org>.

**Cite this article as:** Domínguez M, Herrera S, González JH. 2021. Assessment of phenotypic variability among EEA INTA Pergamino sunflower lines: Its relationship with the grain yield and oil content. *OCL* 28: 33.