Legume Research, 42(6) 2019: 743-749
Print ISSN:0250-5371 / Online ISSN:0976-0571

AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com

# Construction and validation of core collections in *Pisum* sp. using different methodologies

**Espósito[1,2,3], María Andrea*, Gatti[1,4], Ileana*, Cointry[1,3] and Enrique**

Cátedra de Mejoramiento Vegetal y Producción de Semillas,
Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Santa Fe, Argentina.
Received: 30-08-2018 Accepted: 07-02-2019 DOI: 10.18805/LR-451

## ABSTRACT

Core collections contribute to a better utilization of accessions in breeding programs. Eighty-five accessions from a collection of *Pisum* germplasm were evaluated during 2015 and 2016. Phenotypic values of 12 morphological traits were measured and genotypic values (BLUP) were calculated. Molecular characterization was performed assaying a total of 15 SSR and 25 SRAP primer combinations. Four Cluster Analysis (phenotypic values, genotypic values, molecular markers and consensus) and four strategies to determine the number of accessions selected (constant, logarithmic, proportional and maximization strategies) were applied to construct 16 core collections. Validation of the core collections were performed through Mean difference percentage (MD), Variance difference percentage (VD), Coincidence rate of range (CR) Variable rate of coefficient of variation (VR), Shannon diversity index (SW) and the Taxonomy coverage (TC). The logarithmic strategy with data on genotypic values was the best strategy, while the least strategy was the proportional strategy with molecular marker data.

**Key words**: Genotypic values, Molecular characterization, Morphological characterization, *Pisum*, Validation of core collections.

## INTRODUCTION

Pea (*Pisum sativum* L.) is one of the most important food legumes in the world and constant efforts are made to improve its productivity and quality through breeding programs. Germplasm collections must provide genetic variants to the breeders to approach these challenges but in many cases, large size of such collections limits their study and utilization and lowers the quality of their management. Core collections were first defined as a limited set of accessions representing the genetic diversity of a crop species and its wild relatives with a minimum of repetitiveness (Frankel and Brown, 1984) and would offer a better starting material when searching for new traits. They have been widely utilized by the gene bank community, some of them representing the diversity in a complete genus, including wild species, whereas others represent only a small part of the known gene pool.

In practice, the construction of a core collection implies the selection of a limited number of accessions using any kind of information such as passport data and morphological (Hanci and Cebeci, 2018), molecular (Semwal *et al*., 2018) or biochemical characterizations (Jaggal *et al*., 2016).

Nevertheless, in breeding practice, the role of morphological and agronomical traits is irreplaceable, because introducing new variants in these kind of traits is the focal point when bearing a breeding program in mind (Peric *et al.*, 2014). In general, phenotypic values of quantitative traits are affected by genotypic-environmental interaction and cannot correctly reflect the genetic diversity of the initial germplasm resources, but genotypic values calculated in the absence of environmental effects would be a better estimator of such diversity. When accessions of an entire collection can be regarded as a random sample from the entire germplasm of the genus and can be considered stochastically independent, the Best Linear Unbiased Predictor (BLUP) has good predictive accuracy compared to other procedures to calculate genotypic values (Piepho *et al*., 2008) and will lead to a more accurate measure of the genetic distances among accessions and a more representative core collection (Hu *et al*., 2000).

The general procedure for creating a core collection comprehends the identification of germplasm collection to be represented (Van Hintum *et al*., 2003), the division of the collection into differentiated groups, the determination of the number of accessions that will be selected in each group and finally, the selection of the accessions.

Once constructed, the core collection must be validated to confirm to what extent it has achieved the original objectives of representing the variability, comparing

*Corresponding author's e-mail: esposito.maria@inta.gob.ar, ileana1111@gmail.com
[1]Cátedra de Mejoramiento Vegetal y Producción de Semillas, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario (UNR), Zavalla, Santa Fe, Argentina.
[2]Estación Experimental Agropecuaria- Instituto Nacional de Tecnología Agropecuaria-Oliveros, Oliveros, Santa Fe, Argentina.
[3]IICAR-CONICET Instituto de Investigaciones en Ciencias Agrarias de Rosario, Santa Fe, Argentina.
[4]CIUNR Consejo de Investigadores de La Universidad Nacional de Rosario.

it with the original collection. This comparison can be done using the traits used in the construction or those that were excluded, but for better results, both types of traits should be used (Van Hintum *et al*., 2003).

In this investigation, four strategies to determine the number of accessions selected from each group (constant, logarithmic, proportional and maximization strategies) and four kinds of data (phenotypic values, genotypic values, molecular markers and consensus) were applied to construct 16 core collections from a working collection of *Pisum* germplasm in the College of Agricultural Sciences, Rosario National University. These core collections were then validated against the original collection with the objective to select the one with the best representation of the existing variability.

## MATERIALS AND METHODS

**Morphological characterization:** Eighty-five accessions (Table 1) from the working collection of *Pisum* germplasm

**Table 1:** Common name, classification and source of 85 accessions studied.

| | | | |
|---|---|---|---|
| ***P. abyssinicum*** | | | |
| j54757 | USDA 15041 | j808 | USDA 639983 |
| ***P. sativum* subsp. *transcaucasicum*** | | ***P. sativum* subsp. *elatius*** | |
| j1448 | USDA 639982+ | Vir2524E27 | USDA 15008 |
| Virk2376 | USDA 639971 | Vir3115 | USDA 639973 |
| Virk4871 | USDA 639970 | JI 2629 | |
| ***P sativum* subsp. *sativum*** | | | |
| Aa38 | USDA 269862 | Mayflower | USDA 8810 |
| Accord | USDA 635187 | Miranda | USDA 600943 |
| Alaska WR | USDA 12620 | Moshong | USDA 116944 |
| AMA | commercial cultivar | NGB | USDA 101017 |
| APA | commercial cultivar | NN16 | commercial cultivar |
| Arveja Verde | USDA 109866 | Pai Wan Tou | JI 1922 |
| B1826a | USDA 3674 | Patani | USDA 123246 |
| B2001 | breeding line | Patani color | USDA 123246 |
| B22718 | USDA 343987 | Piscello Piccolo Provenz | commercial cultivar |
| B338[22] | USDA 113368 | Pleven2 | USDA 12723 |
| B4476 | USDA 119794 | Romack | USDA 613277 |
| C2001 | breeding line | Shandil | USDA 121976 |
| CanadaA | commercial cultivar | Spring Pea | commercial cultivar |
| CanadáB | commercial cultivar | Sugar Snap | USDA 642176 |
| Chinese Purple | USDA 137118 | Superscout | USDA 601010 |
| Cometodo | commercial cultivar | Thirty days | commercial cultivar |
| Cuarentinas | USDA 162557 | Turf | commercial cultivar |
| Dante | USDA 595591 | VaCo | commercial cultivar |
| DDR11 | commercial cultivar | Viper | commercial cultivar |
| Early Sweet | USDA 635190 | Wonder marrowfat | USDA206868 |
| Erick2 | commercial cultivar | Zavalla1 | breeding line |
| Ervilha T Flor R | USDA 117998 | Zavalla10 | breeding line |
| Explorer | commercial cultivar | Zavalla11 | breeding line |
| G2001 | Breeding line | Zavalla12 | breeding line |
| Granada | USDA 595579 | Zavalla14 | breeding line |
| Green Sugar | USDA 642175 | Zavalla15 | breeding line |
| Hoheinheimer PF | USDA180693 | Zavalla16 | breeding line |
| Ilca5041 | USDA 505082 | Zavalla17 | breeding line |
| Ilca5075 | USDA 505111 | Zavalla18 | breeding line |
| Ilca5115 | USDA 505144 | Zavalla19 | breeding line |
| Inca | commercial cultivar | Zavalla2 | breeding line |
| Keoma | Commercial material | Zavalla20 | breeding line |
| King Tut PurpPod | USDA 635206 | Zavalla21 | breeding line |
| Leo | USDA 595573 | Zavalla22 | breeding line |
| Marina | commercial cultivar | Zavalla26 | breeding line |
| Mattar | USDA 116843 | Zavalla4 | breeding line |
| Mattar color | USDA 116843 | Zavalla5 | breeding line |
| | | Zavalla9 | breeding line |

*JI: John Innes Centre, Norwich, UK. http://www.jic.ac.uk/GERMPLAS/pisum

+USDA: Plant Germplasm Introduction and Testing, USA. http://www.ars-grin.gov

belonging to cultivated species and subspecies were evaluated during 2015 and 2016 in the Experimental Field of the College of Agricultural Sciences, Rosario National University, located in Zavalla (33° 1′ S and 60° 53′ W) at 50 m above sea level. Experimental plots of 20 plants were sowed in a randomized complete block design with two replications in late June. Irrigation was used until flowering and pod setting stages.

The variables analyzed were plant height, first floral knot as the number of knots up to the first flower, first pod knot as the number of knots up to the first pod, pod length, pod width, days until 50% of plants flowering, days until first pod setting, days until harvest, number of pods per plot, number of grains per plot, grain diameter and plot yield.

Best linear unbiased prediction (BLUP) of the adjusted means or genotypic values for these morphological traits was calculated for each variable using META-R (Multi Environment Trial Analysis whit R for Windows.) Version 6.0 (Alvarado *et al*., 2016).

**Molecular characterization:** DNA was extracted using the commercial CTAB method (Smýkal *et al*., 2008). A total of 15 SSR primer combinations and 25 SRAP primer combinations were assayed on all accessions, the amplifications were carried out following the methodology described by Espósito *et al*. (2007) and fragments were scored for presence or absence in each sample.

**Construction of core collections:** Three cluster analysis were performed using the Average Linkage method (Sokal and Michener, 1958) with mean Euclidean distances for phenotypic, genotypic and molecular data (SSR and SRAP) respectively, and a fourth consensus cluster analysis using Gower distances (Gower, 1975) for phenotypic and molecular data were performed using the software InfoStat (Balzarini and Di Renzo, 2003).

To determine the number of accession to be included in the core collections, four strategies were used:

- constant strategy (C) where equal number of accessions from each group is selected;
- logarithmic strategy (L) where the number was calculated

$$\text{as NAC} = \frac{loga_i}{\sum_{i=1}^{ns} loga_i}\, nt$$

where: NAC is the number of accessions sampled per group; *ai* is the number of accessions of the *i*-th group; *nt* is the total number of accessions sampled and *ns* is the number of groups;

- Proportional strategy (P) where the number of accessions selected is a fix fraction of the number in each group (in this study the proportional factor was 2/7)
- Maximization strategy (M) implemented through a modified heuristic algorithm using PowerCore (v. 1.0) (Kim *et al*., 2007) software to select specific combinations of accessions that include complete coverage and retention.

**Validation of core collections:** The parameters utilized to validate the core collections were:

Mean difference percentage (MD) (Hu *et al*. 2000): MD = (St/m) x 100 where St is the number of traits with a significant difference ($\alpha = 0.05$) between the means of the whole collection and the core collection; m is the total number of traits. Significance of the difference was studied by a t test.

Variance difference percentage (VD) (Hu *et al*., 2000): VD = (St/m) x 100 where St is the number of traits with a significant difference (p = 0.05) between the variances of the whole collection and the core collection; m is the total number of traits. All comparisons between variances were carried out by F-test (Snedecor and Cochran, 1980).

Coincidence rate of range (CR) (Diwan *et al*., 1995):

$$\text{CR\%} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{Rc}{Re}\right)100$$

where *Rc* and *Re* represent the ranges of the trait in the core collection and the whole collection, respectively; m is the total number of traits.

Variable rate of coefficient of variation (VR) (Hu *et al*., 2000):

$$\text{VR\%} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{CVc}{CVe}\right)100$$

where *CVc* and *CVe* represent the coefficients of variation of the each trait in the core collection and the whole collection, respectively; *m* is the total number of traits.

Shannon diversity index (SW) (Shannon and Weaver, 1949):

$$\text{VR\%} = -\sum_{i=1}^{S} P_i \text{ x } \log_{10} p_i$$

where *Pi* is the proportion or relative ratio of the molecular marker in the population. Significance of the differences between S in the entire population and the CCs was stablished with a Hutcheson *t*-test for the Shannon diversity index (Hutcheson, 1970*)* where

$$t = \frac{\text{SWe - SWc}}{\sqrt{\text{S}^2\text{SWe} + \text{S}^2\text{SWc}}}$$

In the formula SW represents the Shannon diversity index for the entire collection (SWe) and each of the core collections (SWc). The bottom of the formula refers to the variance of each of the samples calculated as:

$$\text{S}^2\text{sw} = \frac{\sum \text{p}\,(Inp)^2 - (\Sigma p\, Inp)^2}{N} + \frac{\text{S} - 1}{2N^2}$$

where *S* is the total number of alleles, whilst *N* is the total abundance. The *p* is the proportion that each allele makes towards the total.

Another validation parameter calculated was the taxonomy coverage percentage (TC) taking into account the

number of species and subspecies included in the core collection from the number in the entire collection:

$$TC\ \% = \frac{NC}{Ne}100$$

where *Nc* is the number of species and subspecies in the core collection and Ne is the number of species and subspecies in the entire collection.

## RESULTS AND DISCUSSION

Results from the cluster analysis are presented in Fig 1 (phenotypic data), Fig 2 (genotypic values data), Fig 3 (SRAPs and SSR molecular markers data) and Fig 4 (consensus data). Cophenetic correlation coefficients were high for all analysis, ranging from 0.95 for the consensus analysis to 0.86 for the phenotypic data.

**Table 2:** Name and number of entries of the different Core collections obtained using different strategies and kinds of data.

| Strategy | Type of data | Name | N° |
|---|---|---|---|
| Constant (C) | Phenotypic values | CC1 | 16 |
| | Genotypic values | CC2 | 16 |
| | Molecular markers | CC3 | 13 |
| | consensus | CC4 | 15 |
| Proportional (P) | Phenotypic values | CC5 | 16 |
| | Genotypic values | CC6 | 16 |
| | Molecular markers | CC7 | 14 |
| | consensus | CC8 | 14 |
| Logarithmic (L) | Phenotypic values | CC9 | 13 |
| | Genotypic values | CC10 | 15 |
| | Molecular markers | CC11 | 14 |
| | consensus | CC12 | 15 |
| Maximization (M) | Phenotypic values | CC13 | 14 |
| | Genotypic values | CC14 | 14 |
| | Molecular markers | CC15 | 38 |
| | consensus | CC16 | 43 |

Ten groups were distinguished by truncating the dendrograms at the mean distance, and accessions closer to the centroid of each group were chosen for the strategies C, L and P. An optimal number of entries to represent the entire collection would be about 10 to 30% of the total number (Zhang *et al*. 2010), in this case from 9 to 25. The optimal number stablished was 16, which represents a 20% of the entire collection. This way, 12 core collections were obtained (Table 2) with a number of entries ranging from 13 to 16. Other 4 collections were obtained using the M strategy, being the number of entries in these cases defined by the program utilized (PowerCore) and ranging from 14 for phenotypic and genotypic values, to 38 for molecular data and 43 for the consensus analysis. In the present study, CC15 and CC16 obtained applying this strategy exceeded the optimal number of entries making their manageability difficult.

Results of the different parameters of validation for the 16 core collections analyzed are presented in Table 3. Criteria based on mean, variance, coefficient of variation or range have been used mainly to evaluate the quality of core collections based on continuous traits (Hu *et al*., 2000). A core collection is considered a good representative of the variation in the entire collection when MD and VD ≤ 20%, CR% ≥ 80% and VR% ≥ 100% (Hu *et al*., 2000).

As it can be seen, all core collections except CC3, CC4 and CC14 had MD lower than 20%, but for VD, only CC5, CC6, CC9, CC10, CC11, CC12, CC14, CC15 and CC16 matched this criteria, having the rest of them an elevated number of traits with significantly different variances when compared with the entire collection. For CR, CC1, CC2, CC3, CC4, CC10, CC13, CC14, CC15 and CC16 showed values superior to 80%; indicating that these core collections had high heterogeneity in each trait and high number of entries with extreme values.

**Table 3:** Validation parameters between the core collections and the entire collection.

| | MD % | VD% | CR% | VR% | SW | T Hutchenson | Taxonomy Coverage % |
|---|---|---|---|---|---|---|---|
| CC1 | 0 | 49.93 | 94.74 | 144.12 | 5.68 | 3.76*** | 100 |
| CC2 | 0 | 41.60 | 94.48 | 138.5 | 5.70 | 4.63*** | 100 |
| CC3 | 33.34 | 41.6 | 85.91 | 132.91 | 5.73 | 5.28*** | 100 |
| CC4 | 33.34 | 33.34 | 82.85 | 128.7 | 5.57 | 6.20*** | 100 |
| CC5 | 0 | 0 | 75.78 | 98.27 | 5.60 | 1.29ns | 75 |
| CC6 | 0 | 0 | 65.93 | 89.47 | 5.44 | 7.08*** | 75 |
| CC7 | 0 | 33.34 | 55.68 | 77.66 | 5.43 | 7.40*** | 50 |
| CC8 | 8.33 | 50 | 52.47 | 64.35 | 5.43 | 7.89*** | 50 |
| CC9 | 0 | 0 | 79.63 | 124.6 | 5.65 | 2.16** | 50 |
| CC10 | 0 | 8.33 | 87.73 | 127.85 | 5.64 | 1.71ns | 100 |
| CC11 | 8.33 | 0 | 67.59 | 92.06 | 5.48 | 5.60*** | 75 |
| CC12 | 8.33 | 16.66 | 65.86 | 88.01 | 5.52 | 3.33*** | 75 |
| CC13 | 0 | 41.6 | 96.01 | 144.34 | 5.65 | 2.11** | 75 |
| CC14 | 25 | 0 | 92.08 | 135.71 | 5.61 | 0.48ns | 75 |
| CC15 | 0 | 0 | 96.13 | 112.23 | 5.67 | 4.29*** | 100 |
| CC16 | 0 | 0 | 97.33 | 117.84 | 5.67 | 6.89*** | 100 |

* significant with $p < 0.05$, ** significant with $p < 0.01$

MD, mean difference percentage; VD, variance difference percentage; CR, Coincidence rate; VR, variable range, SW, Shannon diversity index (entire population SW = 5.61).

For VR, CC1, CC2, CC3, CC4, CC9, CC10, CC13, CC14, CC15 and CC16 showed values superior to 100% indicating that variation coefficients in each trait was higher than the ones in the entire collection and that this heterogeneity ensures a good representation of the variability.

The SW index reflexes the richness and the evenness of alleles (Balzarini *et al*. 2010) and has been used to validate core collections in *Pisum* before (Smýkal *et al*., 2008) because when there are no significant differences, the core collection can be considered a good representative of the entire collection. In this study, it was calculated using the molecular data and showed very similar values between the core collections CC5, CC10 and CC14 and the entire collection demonstrating that are good representatives of it; while the rest of the core collections are significantly different ($p < 0.05$) in the Hutcheson *t*-test using the Shannon diversity index.

Passport data or taxonomy are the two usually applied criteria when constructing core collections. In this study, the quantity of species and subspecies that remained present in the core collections was used as an index of

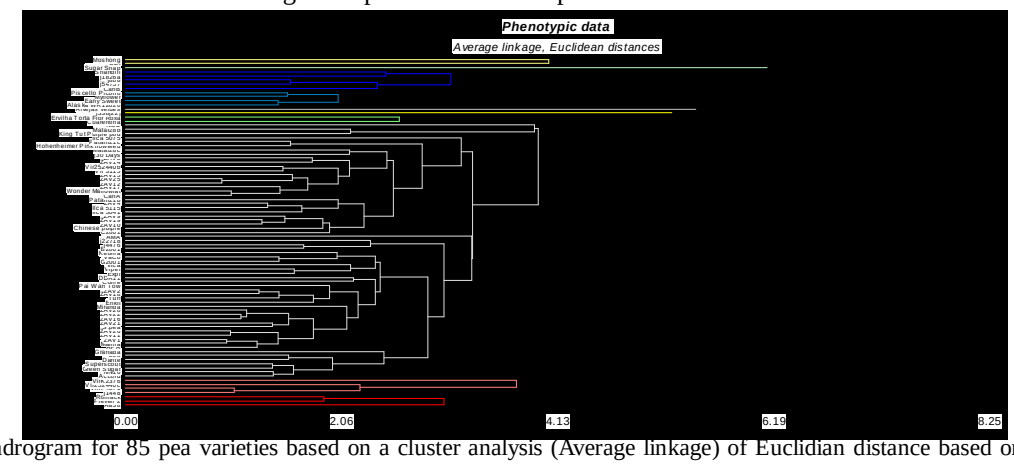


**Fig 1:** Dendrogram for 85 pea varieties based on a cluster analysis (Average linkage) of Euclidian distance based on phenotypic data.

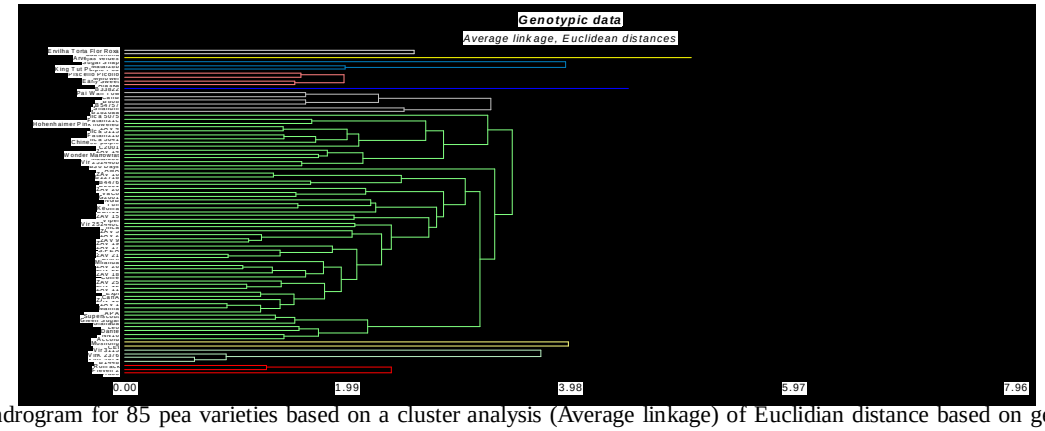


**Fig 2:** Dendrogram for 85 pea varieties based on a cluster analysis (Average linkage) of Euclidian distance based on genotypic values (BLUPs) data.

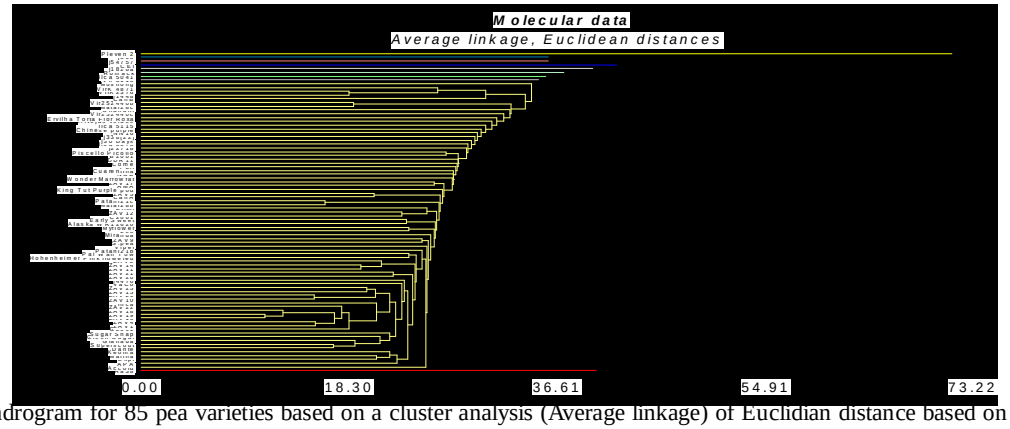


**Fig 3:** Dendrogram for 85 pea varieties based on a cluster analysis (Average linkage) of Euclidian distance based on molecular markers (SRAPs and SSR) data.

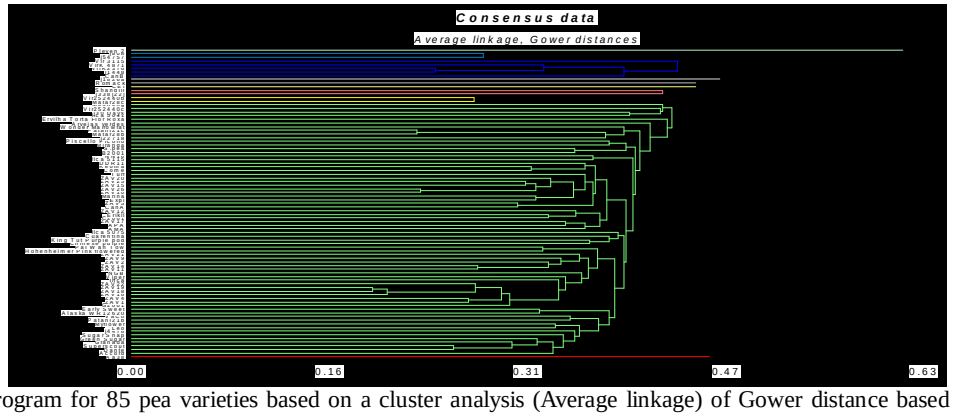


**Fig 4:** Dendrogram for 85 pea varieties based on a cluster analysis (Average linkage) of Gower distance based on phenotypic and molecular markers data.

representativeness to validate and qualify the core collections. The TC was maximum for CC1, CC2, CC3, CC4, CC10, CC15 and CC16, because all species and subspecies resulted represented in them. The lower TC were those of CC7, CC8 and CC9 where only 50% of the species and subspecies where represented.

In the past, the efficiency of the different strategies to construct core collections has been widely studied, being the L and P strategies the most appropriate. In the present study, considering all validation parameters together, the L strategy with genotypic values data was the best strategy (CC10) while the worst or the least strategy was P with molecular marker data (CC7).

## CONCLUSION

The best core collection to represent the working collection of pea resulted by combining the logarithmic strategy to select the number of entries with genotypic value data. This kind of collection provides useful information for breeding programs and can be used as a reference set for the evaluation of newly introduced accessions.

## REFERENCES

Alvarado, G., López, M., Vargas, M., Pacheco, A., Rodríguez, F., Burgueño, J., Crossa, J. (2016). META-R (Multi Environment Trial Analysis whit R for Windows.) Version 6.0 http://hdl.handle.net/11529/10201 International Maize and Wheat Improvement Center.

Balzarini, M., Di Renzo, J. (2003). Info-Gen: Software para análisis estadístico de datos genéticos. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Argentina.

Balzarini, M., Bruno, C., Peña, A., Teich, I., Di Renzo, J. (2010). *Estadística en Biotecnología. Aplicaciones en Info-Gen. Encuentro Grupo Editor. Córdoba, Argentina.*

Diwan, N., McIntosh, M.S., Bauchan, G.R. (1995). Methods of developing a core collection of annual Medicago species. *Theor Appl Genet.* **90:** 755-761.

Espósito, M.A., Martin, E.A., Cravero, V.P., López Anido, F.S., Cointry, E. (2007). Characterization of pea accessions by SRAP's markers. *Sci Hortic.* **113:**329-335.

Frankel, O.H., Brown, A.H.D. (1984). Plant genetic resources today: a critical appraisal. In: Crop Genetic Resources: Conservation & Evaluation. [Holden JHW, Williams JT (eds)]. George Alien & Unwin Ltd., London, pp 249-257.

Gower, J.C. (1975). Generalized Procrustes Analysis. *Physchometrika.* **40:** 33-51.

Hanci, F., Cebeci, E. (2018). Determination of morphological variability of different pisum genotypes using principal component analysis. *Legume Res.* DOI: 10.18805/LR-438. On line: 30/10/2018.

Hu, J.J., Zhu, J., Xu, H.H.M. (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor Appl Genet.* **101**(1-2): 264-268.

Hutcheson, K. (1970). A test for comparing diversities based on the Shannon formula. *J Theor Biol.* **29:** 151-154. https://doi.org/10.1016/0022-5193(70)90124-4.

Jaggal, L.G., Patil, B.R., Naik, P.M., Priya, K. (2016). Genetic diversity of selected accessions for seed protein among pigeonpea minicore collection. *Legume Res.* **39**(5): 704-708.

Kim, K.W., Chung, H.K., Cho, G.T., Ma, K.H., Chandrabalan, D., Gwag, J.G., Kim, T.S., Cho, E.G., Park, Y.J. (2007). PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics.* **23**(16): 2155-62.

Peric, V., Nikolić, A.,Babić, V., Sudarić, A.,Srebrić, M., Đorđević, V., Mladenović Drinić, S. (2014). *Genetika.* **46**(3): 839-854.

Piepho, H.P., Möhring, J., Melchinger, A.E., Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161:**209-228.

Semwal D.P., Ahlawat, S.P., Pradheep, K. (2018). Pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild spp. germplasm collection status, diversity distribution and trait-specific germplasm mapping using GIS tools in India. *Legume Res.* **41**(5): 656-662.

Shannon, C.E., Weaver, W. (1949). The mathematical theory of communication. The University of Illinois Press, Urbana, 117p.

Smýkal, P., Hýbl, M., Corander, J., Jarkovský, J., Flavell, A., Griga, M. (2008). Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. *Theor Appl Genet*. **117:** 413-424.

Snedecor, G.W., Cochran, W.G. (1980). Statistical Methods. 7ed, Iowa State University, Ames: 507 p.

Sokal, R.R., Michener, C.D. (1958). A Statistical Methods for Evaluating Systematic Relationships. *University of Kansas Science Bulletin*. **38:**1409-1438.

Van Hintum, T.J.L., Brown, A.H.D., Spillane, C., Hodgkin, T. (2003). Colecciones núcleo de recursos fitogenéticos. Boletín Técnico No. 3del IPGRI. Instituto Internacional de Recursos Fitogenéticos, Roma, Italia.

Zhang, J., Wang, Y., Zhang, X.Z., Li, T.H., Wang, K., Xu, X.F., Han, Z.H. (2010). Sampling strategy to develop a primary core collection of apple cultivars based on fruit traits. *Afr J Biotechnol*. **9:** 123-127.