PAB
Pesquisa Agropecuária Brasileira

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Genetics/ Original Article

Kohonen's self-organizing maps for the study of genetic dissimilarity among soybean cultivars and genotypes

Abstract – The objective of this work was to evaluate the genetic dissimilarity between soybean cultivars and genotypes for the selection of parents, as well as to propose a new method for using Kohonen's self-organizing maps (SOMs) and to test its efficiency through Anderson's discriminant analysis. The morphoagronomic descriptors of soybean cultivars and genotypes were evaluated. For data analysis, SOM-type artificial neural networks were used. The proposed method allowed the determination of the best network architecture in a nonsubjective way. Furthermore, at the beginning of training, it was possible to mitigate the randomness effect of the synaptic weights on the formed clusters. Six dissimilar clusters were formed; therefore, there is genetic dissimilarity between soybean cultivars and genotypes. Cultivars C25, C8, and C13 can be combined with C36, C31, C32, and C33 because they show good yield-related attributes and high dissimilarity. The proposed methodology is advantageous in comparison with the use of traditional SOMs, besides being efficient due to clustering consistency according to Anderson's discriminant analysis.

Index terms: *Glycine max*, artificial neural networks, multivariate analysis, plant breeding.

Mapas auto-organizáveis de Kohonen no estudo da dissimilaridade genética entre cultivares e genótipos de soja

Resumo – O objetivo deste trabalho foi avaliar a dissimilaridade genética entre cultivares e genótipos de soja para a seleção de genitores, bem como propor um novo método para a utilização de mapas auto-organizáveis de Kohonen (SOMs) e testar sua eficiência por meio da análise discriminante de Anderson. Foram avaliados os descritores morfoagronômicos de cultivares e genótipos de soja. Para análise dos dados, utilizaram-se redes neurais artificiais do tipo SOM. O método proposto permitiu a determinação da melhor arquitetura de rede de forma não subjetiva. Além disso, no início do treinamento, foi possível mitigar o efeito da aleatoriedade dos pesos sinápticos sobre os grupos formados. Foram formados seis grupos dissimilares; portanto, há dissimilaridade genética entre cultivares e genótipos de soja. As cultivares C25, C8 e C13 podem ser combinadas com as C36, C31, C32 e C33, por apresentarem bons atributos de produtividade e alta dissimilaridade. A metodologia proposta é vantajosa em comparação ao uso de SOMs tradicionais e se mostrou eficiente devido à consistência dos agrupamentos de acordo com a análise discriminante de Anderson.

Termos para indexação: *Glycine max*, redes neurais artificiais, análise multivariada, melhoramento genético vegetal.

Ludimila Geiciane de Sá(1),
Alcinei Mistico Azevedo(1 ✉),
Carlos Juliano Brant Albuquerque(1),
Nermy Ribeiro Valadares(1),
Orlando Gonçalves Brito(2),
Ana Clara Gonçalves Fernandes(1)
and Ignacio Aspiazú(3)

(1) Universidade Federal de Minas Gerais, Instituto de Ciências Agrárias, Avenida Universitária, nº 1.000, Bairro Universitário, CEP 39404-547 Montes Claros, MG, Brazil. E-mail: ludimilageiciane@hotmail.com, alcineimistico@hotmail.com, carlosjulianobrant@gmail.com, nermyrv@gmail.com, anaclaragoncalvesfernandes@gmail.com

(2) Universidade Federal de Lavras, Aquenta Sol, CEP 37200-900 Lavras, MG, Brazil. E-mail: orlandocefet@yahoo.com.br

(3) Universidade Estadual de Montes Claros, Campus Janaúba, Avenida Reinaldo Viana, nº 2.630, CEP 39440-000 Janaúba, MG, Brazil. E-mail: ignacio.aspiazu@unimontes.br

✉ Corresponding author

Received
October 11, 2021

Accepted
January 27, 2022

How to cite
SÁ, L.G. de; AZEVEDO, A.M.; ALBUQUERQUE, C.J.B.; VALADARES, N.R.; BRITO, O.G.; FERNANDES, A.C.G.; ASPIAZÚ, I. Kohonen's self-organizing maps for the study of genetic dissimilarity among soybean cultivars and genotypes. **Pesquisa Agropecuária Brasileira**, v.57, e02722, 2022. DOI: https://doi.org/10.1590/S1678-3921.pab2022.v57.02722

cc BY This is an open-access article distributed under the Creative Commons Attribution 4.0 International License

Introduction

Among plants of global economic importance, which are the targets of improvement programs, soybean [*Glycine max* (L.) Merrill] stands out for its worldwide production − 362.947 million tonnes and 127,842 million ha planted area −, in the 2020/2021 harvest (USDA, 2021). Efforts are focused on the increasing of soybean production through genetic improvement, to maintain Brazil as the world's largest producer of this crop.

However, the selection of superior individuals is not an easy task, and divergent parents with good performance per se should be used. Genetic diversity studies have been often carried out using traditional multivariate techniques, such as dendrograms (Arief et al., 2017), principal components (Hamawaki et al., 2012), and canonical variables (Vendruscolo et al., 2020). However, there is the possibility of carrying out these studies through computational intelligence, using artificial neural networks (ANNs) (Ferreira et al., 2018). The main advantages of ANNs are their nonparametric approach, tolerance to data loss, and the dispensability of detailed information on the modeled system, such as design and genealogies (Silva, et al., 2014; Azevedo et al., 2017).

Among the ANN techniques are the self-organizing maps (SOMs), developed by Teuvo Kohonen, in 1982 (Kohonen, 1982). SOMs are a type of artificial neural network trained by unsupervised competitive learning (Kohonen, 2001). Currently, SOMs are considered an essential tool in multivariate statistics, in the context of computational intelligence, as its algorithm is able to organize dimensionally complex data into groups according to their similarities (Kohonen, 1982).

SOMs have been effectively used to perform many tasks, which include genetic dissimilarity studies, among others. However, the network topology is usually selected in subjective way. Furthermore, at the beginning of the iterative process of SOM networks, synaptic weights are random, which can lead to different results for the same data set and network configuration. This can lead to discrediting this method which, therefore, requires the implementation of strategies to correct this problem.

The objective of this work was to evaluate the genetic dissimilarity between soybean cultivars and genotypes, in order to select parents, and propose a new method of using SOMs and test their efficiency through Anderson's discriminant analysis.

Materials and Methods

The experiment was carried out from February to July 2017, in the experimental area of the Instituto de Ciências Agrárias (ICA) of the Universidade Federal de Minas Gerais (UFMG), regional campus of Montes Claros county, in the state of Minas Gerais, Brazil. The experimental area is located between 16º51'00"S and 44º55'00"W, at 630 m altitude, and its soil is mostly classified as Cambisol (Santos et al., 2018). According to the Köppen-Geiger's classification, the climate in the region is Aw (wet tropical), with dry winter and rainy summer. During the experimental period, the annual means were 24.35°C for temperature and 264.90 mm rainfall.

Using a simple lattice design, 36 soybean cultivars and genotypes were evaluated (Table 1), with two replicates and 40 plants per plot, out of which 15 randomly selected plants were analyzed. The sowing lines were 5 m long with 0.5 m spacing apart.

In the phenotypic characterization process, 11 quantitative descriptors were analyzed in each plant, as follows: hypocotyl length (HL, mm), measured from the soil surface to the cotyledonary node, using a digital caliper, at the V_2 stage; hypocotyl diameter (HD, mm), using a digital caliper, at the V_2 stage; length of cotyledons 1 and 2 (LC1/LC2, mm) measured from the insertion of the cotyledon in the main stem to its end, using a digital caliper, at the V_2 stage; epicotyl length (EL, mm) measured from the cotyledonary node to the nodes of the unifoliolate leaves, using a digital caliper, at the V_3 stage; length of the petiole of the first trifoliate leaf (LPTL, mm) measured from the insertion of the petiole on the main stem to the insertion of the two lateral leaflets of the trifoliate leaf, using a digital caliper, at the V_3 stage; length of the central leaflet rachis of the first trifoliate leaf (LR, mm) measured from the junction of the two lateral leaflets to the insertion of the terminal leaflet, using a digital caliper, at the V_3 stage; plant height (PH, cm) obtained from the distance from ground level to the apical end of the plant, using a measuring tape, at the R_8 stage; height of insertion of the lowest pod (HILP, cm) obtained by the distance from the ground level to the first pod of the plant, using a measuring tape, at the R_8 stage; number of pods (NP) counted in each evaluated plant, for which only pods with seed were considered; seed weight (SW, g), using an analytical digital scale.

For data analysis, SOM-type artificial neural networks were used to study the genetic dissimilarity between cultivars and genotypes. The analysis was based on standardized data. Different network architectures were tested by varying the number of lines (1 to 5) and columns (1 to 5), totaling 24 configurations (excluding the combination with one row and one column).

Table 1. Trade names and codes of 36 soybean (*Glycine max*) cultivars and genotypes evaluated at the Instituto de Ciências Agrárias of Universidade Federal de Minas Gerais, Montes Claros, MG, Brazil.

Code(1)	Trade name	Maintainer
C1	97R21	DuPont Pioneer
C2	97R73	DuPont Pioneer
C3	98Y12	DuPont Pioneer
C4	98Y30	DuPont Pioneer
C5	99R03	DuPont Pioneer
C6	99R09	DuPont Pioneer
C7	AS 3610IPRO	Agroeste
C8	AS 3730IPRO	Agroeste
G9	BMXDESAFIO	Brasmax
G10	BMXPontaIPRO	Brasmax
C11	BMX Potência RR	Brasmax
C12	CD 2720IPRO	Coodetec
C13	CD 2728IPRO	Coodetec
C14	CD 2730IPRO	Coodetec
C15	CD 2737RR	Coodetec
C16	CD 2750IPRO	Coodetec
C17	CD 2817IPRO	Coodetec
G18	DM6563RSFIPRO	DONMARIO
C19	DS5916IPRO	Dow AgroSciences
C20	M5947IPRO	Monsoy
C21	M6210IPRO	Monsoy
C22	M6410IPRO	Monsoy
C23	M7110IPRO	Monsoy
C24	M7739IPRO	Monsoy
C25	M8210IPRO	Monsoy
C26	NA 5909 RG	Nidera
C27	NS 5959 IPRO	Nidera
G28	NS6906IPRO	Nidera
C29	NS 6909 IPRO	Nidera
C30	NS 7000 IPRO	Nidera
C31	NS 7209 IPRO	Nidera
C32	NS 7300 IPRO	Nidera
C33	NS 7338 IPRO	Nidera
C34	TMG 7062 IPRO	TMG
C35	RK6813 RR	GDM
C36	RK7814IPRO	Monsoy

(1)C, cultivar; G, genotype. Information on the cultivars was collected from the Registro Nacional de Cultivares (RNC), published by the Brazilian Ministry of Agriculture, Livestock and Food Supply (Ministério da Agricultura, Pecuária e Abastecimento, MAPA) (Brasil, 2021).

In order to select the best network architecture, 1,000 training sessions were performed for each combination and, for each combination, the average hit rate was estimated by Anderson's discriminant analysis and the smallest number of empty clusters. Subsequently, the best network architecture was selected (that with the highest average hit rate and the lowest number of empty clusters).

Anderson's discriminant analysis method was performed as described by Cruz et al. (2014), and the average hit rate was estimated by the relationship between the number of erroneous classifications and the total number of classifications.

After selecting the best network topology, 10,000 new training sessions were carried out and, subsequently, a dissimilarity matrix was built. To estimate the dissimilarity between cultivars and genotypes, the frequency at which the lines (two by two) were grouped into distinct neurons was estimated.

The unweighted pair group method (UPGMA) was applied to obtain a dendrogram, using the arithmetic averages from the dissimilarity matrix. The number of groups stablished in the dendrogram was based on the number of neurons in the best network topology. The consistency of the cluster was verified by Anderson's discriminant analysis.

The analyses were performed using the R software (R Core Team, 2016). For the use of SOM networks, we applied the RSNNS package (Bergmeir & Benitez, 2012). To obtain the dendrogram, the hclust function was employed and, for the representation of the dissimilarity matrix and the normalized means, the corrplot package was used (Wei & Simko, 2021).

Results and Discussion

The best network architecture was found using three rows and two columns (Figure 1). For this configuration, more than 99% of accuracy was found by Anderson's discriminant analysis, and 0% empty clusters.

To select the best network architecture, six neurons (three rows and two columns) can be used, therefore, the setting number of clusters equals 6. Kohonen (2001) emphasizes that the determination of the number of neurons and parameters of learning is an empirical process, based on the user's experience and trial and error methods. Several studies using SOMs defined their topology by trial or at random (Chaudhary et al., 2014). Therefore, the proposed method to find the best network architecture is very important, since, in the traditional method, for each time SOMs are used, they can present different results because the networks have random synaptic weights at the beginning of the training. In addition, it prevents the setting (number of rows and columns) from being subjectively selected.

After choosing the best network architecture, the dissimilarity matrix was obtained and graphically represented (Figure 2). The lighter colors indicate less distance between the genotypes, that is, they are more similar to each other, while darker colors indicate that the genotypes are more dissimilar (they are farther apart).

The most genotypes were observed as distant from each other, that means that they are dissimilar, indicating the presence of high genetic variability. Thus, the shorter the distance, the more similar the genetic or parental individuals. This was noticed among some materials from the same company, such as the C2, C3, C5, and C6 cultivars from DuPont Pioneer; C15, C16, and C17 cultivars from Coodetec; and C31, C32, and C33 cultivars from Nidera. Therefore, artificial crosses between these genotypes are not recommended.

The quantification of genetic dissimilarity existing between individuals generates information on the degree of similarity, or difference between genotypes, which allows of the formation of heterotic groups (by grouping methods) that are essential when choosing parents with good genetic complementarity (Lima & Peluzio, 2015). Hence, the evaluation of genetic diversity is essential in breeding programs, as it makes it possible the optimization of parental selection and, consequently, the prediction of the best hybrid combinations. However, in addition to being dissimilar, it is necessary that parents associate high means and variability in the characteristics that are being improved (Ferreira et al., 2018).

The dendrogram obtained by UPGMA from the dissimilarity matrix shows the formation of heterotic groups, which means that there is genetic diversity among the genotypes, and 0.95 cophenetic correlation coefficient was obtained, indicating a good fit of the dendrogram (Figure 3). This coefficient is used to assess the consistency of the clustering pattern (Ayed et al., 2016). For the analysis of the dendrogram, considering the pre-set number of six clusters, a cut was made at 60% of the distance.

Cluster I is composed of the cultivars C3, C6, C5, C2, C16, C15, and C17; cluster II is composed of C14, C25, C8, and C13; cluster III is composed of C1, C23, C12, G9, and C24; cluster IV is composed of C34, C4,

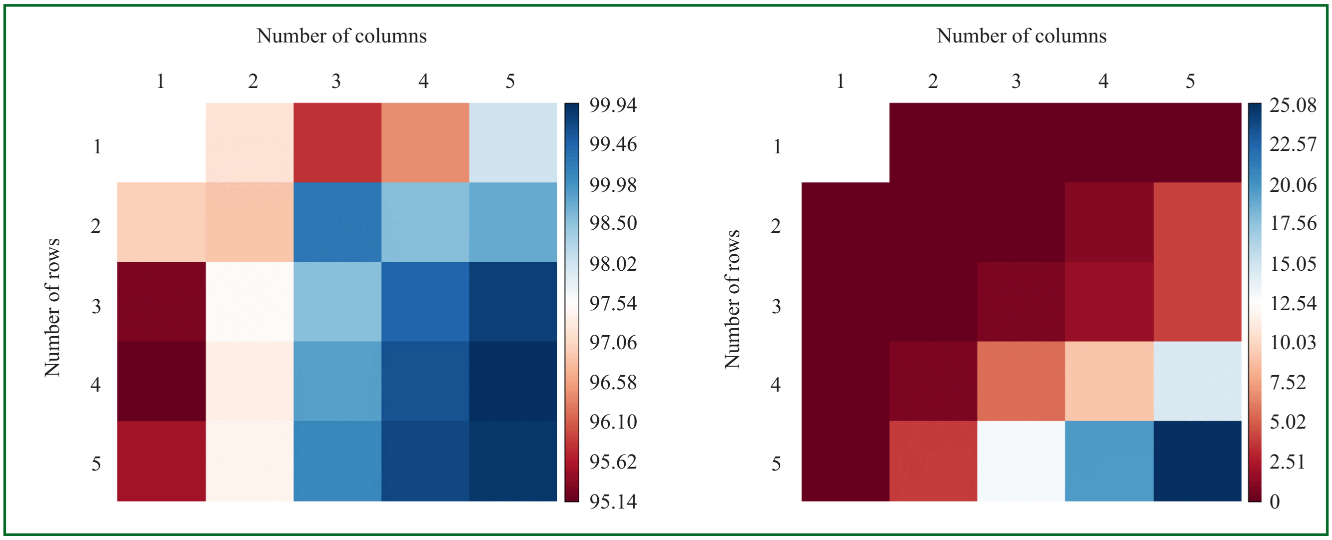

Figure 1. Average percentage of correct classifications by Anderson's discriminant analysis (A), and percentage of empty clusters (B), considering 24 neural network configurations by self-organizing maps (SOMs) in soybean (*Glycine max*) cultivars and genotypes cultivated in the municipality of Montes Claros, in the state of Minas Gerais, Brazil.

C36, C31, C32, and C33; cluster V is composed of C7, C21, C20, C26, C27, C19, and C29; and cluster VI is composed of the cultivars G28, C35, G10, C30, G11, G18, and C22. All clusters, except for the IV, showed a predominance of high values for hypocotyl length and cotyledon length one and two (Figure 3). High values for hypocotyl diameter were also observed. Groups IV, V, and VI showed lower values for epicotyl length, length of petiole of the first trifoliate leaf, rachis length, plant height and height of insertion of the lowest pod. In contrast, groups II, IV, and V showed higher values for weight and number of pods, which are characteristics that define the production.

According to Val et al. (2014), the measurement of agronomic characteristics of the crop, such as the height of insertion of the lowest pod, plant height, and number of pods is important for allowing the breeder

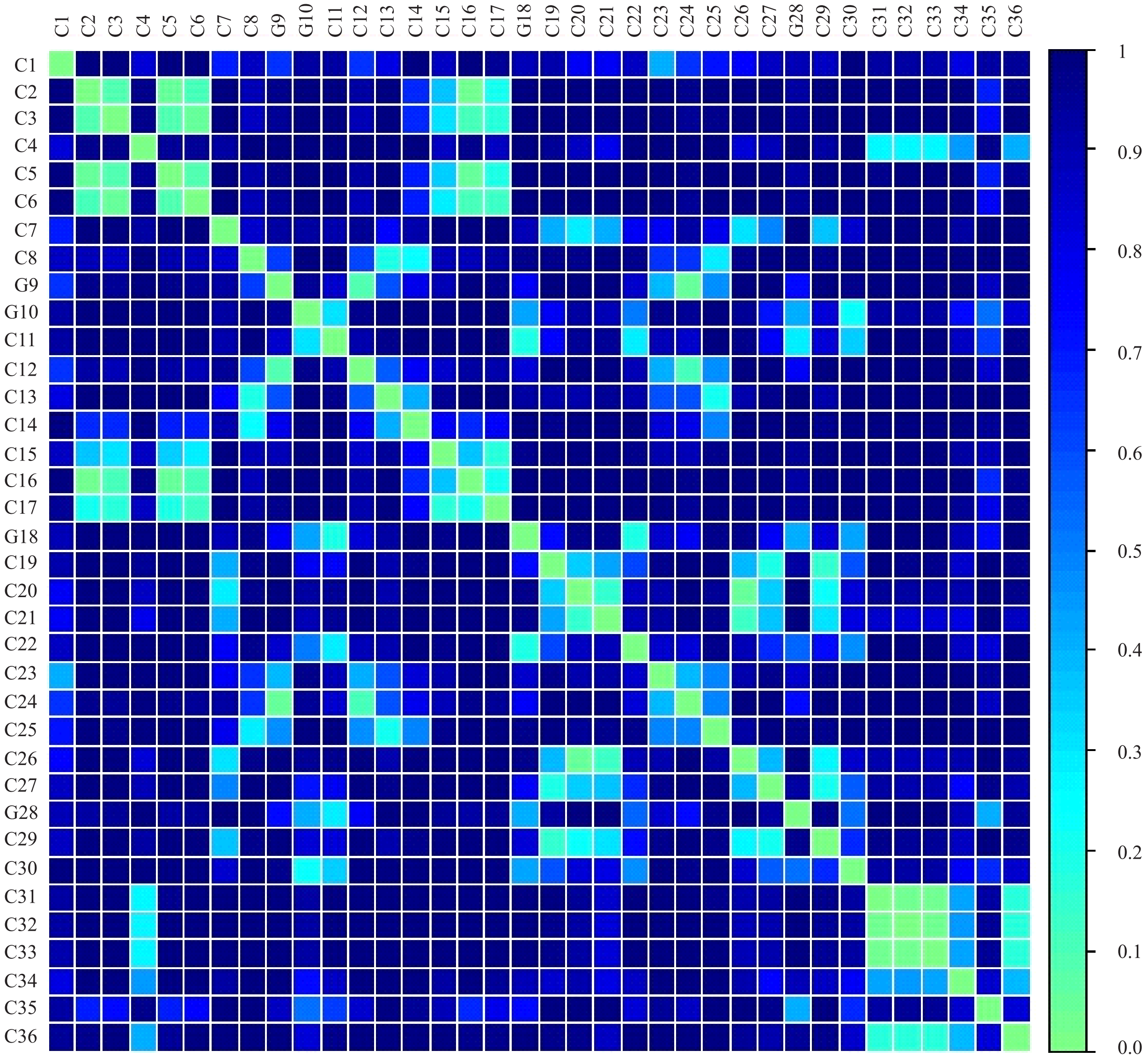

Figure 2. Dissimilarity matrix graphic representation obtained by Kohonen's self-organizing maps (SOMs) for 36 cultivars and genotypes (see Table 1) of soybean (*Glycine max*) cultivated in the municipality of Montes Claros, in the state of Minas Gerais, Brazil.

to identify and select the best genotypes for characters of great agronomic importance.

The distances between the six formed clusters showed the following values: the maximum inter-cluster distance (0.999) occurred between clusters I and V; cluster II was farther from the IV (0.994); cluster III was farther from the IV (0.977); cluster VI was farther from the II (D = 0.981); and the minimum inter-cluster distance (0.667) was observed between groups II and III (Table 2). Genotypes belonging to

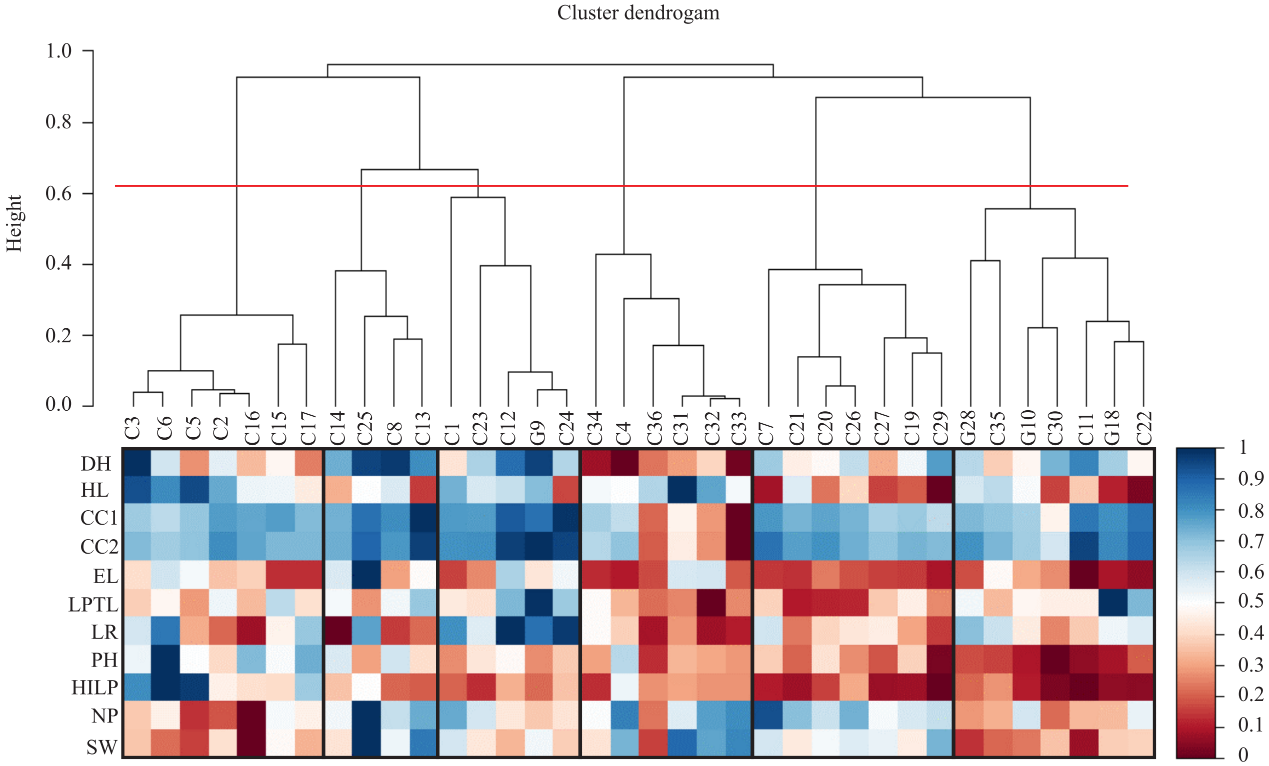

Figure 3. Dendrogram constructed by the UPGMA method from the complement coincidence matrix of 36 cultivars and genotypes (see Table 1) grouped by SOM-type networks, and representation of the normalized phenotypic means for soybean (*Glycine max*) cultivated in the municipality of Montes Claros, in the state of Minas Gerais, Brazil. Parameters: DH, hypocotyl diameter; HL, hypocotyl length; CC1/CC2, length of cotyledons 1 and 2; EL, epicotyl length; LPTL, length of the petiole of the first trifoliate leaf; LR, length of the rachis of the central leaflet of the first trifoliate leaf; PH, plant height; HILP, height of insertion of the lowest pod; NP, number of pods; and SW, seed weight.

Table 2. Average distances within (main diagonal) and between (off diagonal) clusters based on the dissimilarity matrix obtained by Kohonen's self-organizing map neural networks for soybean (*Glycine max*) cultivars and genotypes cultivated in the municipality of Montes Claros, in the state of Minas Gerais, Brazil.

Cluster	I	II	III	IV	V	VI
I	0.144	0.892	0.953	0.976	0.999	0.951
II	0.892	0.231	0.667	0.994	0.942	0.981
III	0.953	0.667	0.303	0.977	0.939	0.884
IV	0.976	0.994	0.977	0.219	0.914	0.943
V	0.999	0.942	0.939	0.914	0.256	0.870
VI	0.951	0.981	0.884	0.943	0.870	0.382

the most distant clusters can be used in hybridization programs, to obtain a wide spectrum of variation among segregants. The greater is the divergence between genotypes, the greater will be the heterosis of hybrids in a breeding program in the development of higher yielding varieties (Bekele et al., 2012). Within this context, cultivars from clusters II (C14, C25, C8, and C13) and IV (C34, C4, C36, C31, C32, and C33) can be selected as parents in hybridization programs, as they are genetically distant and have good yield-related attributes such as number and weight of pods (Figure 3).

The use of artificial neural networks as a clustering method is a promising path. Ferreira et al. (2018) concluded that SOM networks can provide more valuable results when compared to the traditional cluster analysis.

To verify the adequacy of the clusters obtained by the SOMs network method, we applied Anderson's discriminant analysis; the results provided by the method show that the accesses were 100% correctly classified in the clusters.

The use of Anderson's discriminant analysis is considered viable to verify the clustering consistency proposed by the technique of neural networks (Barbosa et al., 2011). In addition, Anderson's discriminant function proved to have a great potential and to be an additional tool to check the correct classification provided by the various methods of multivariate analysis (Sudré et al., 2006).

Conclusions

1. There is genetic dissimilarity between soybean cultivars and genotypes, and the cultivars M8210IPRO, AS 3730IPRO, and CD 2728IPRO, in the cluster II, can be combined with RK7814IPRO, NS 7209 IPRO, NS 7300 IPRO, and NS 7338 IPRO, in the cluster IV because they show good yield-related attributes.

2. The proposed methodology is advantageous (in comparison with the use of traditional SOM) for its efficiency, as it allows of clustering consistency in a non-subjective way, in accordance with Anderson's discriminant analysis and with the study of dissimilarity, without the influence of random synaptic weights at the beginning of training.

3. Self-organizing maps (SOMs) are efficient for the evaluation of genetic diversity of soybean cultivars for crop improvement programs.

Acknowledgments

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Finance Code 001); to Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for financial support.

References

ARIEF, V.N.; DELACY, I.H.; BASFORD, K.E.; DIETERS, M.J. Application of a dendrogram seriation algorithm to extract pattern from plant breeding data. **Euphytica**, v.213, art.85, 2017. DOI: https://doi.org/10.1007/s10681-017-1870-z.

AYED, R.B.; HASSEN, H.B.; ENNOURI, K.; REBAI, A. Genetic markers analyses and bioinformatic approaches to distinguish between olive tree (*Olea europaea* L.) cultivars. **Interdisciplinary Sciences: Computational Life Sciences**, v.8, p.366-373, 2016. DOI: https://doi.org/10.1007/s12539-016-0155-x.

AZEVEDO, A.M.; ANDRADE JÚNIOR, V.C.; SOUSA JÚNIOR, A.S.; SANTOS, A.A.; CRUZ, C.D.; PEREIRA, S.L.; OLIVEIRA, A.J.M. Eficiência da estimação da área foliar de couve por meio de redes neurais artificiais. **Horticultura Brasileira**, v.35, p.14-19, 2017. DOI: https://doi.org/10.1590/S0102-053620170103.

BARBOSA, C.D.; VIANA A.P.; QUINTAL, S.S.R.; PEREIRA, M.G. Artificial neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology**, v.11, p.224-231, 2011. DOI: https://doi.org/10.1590/S1984-70332011000300004.

BEKELE, A.; ALEMAW, G.; ZELEKE, H. Genetic divergence among soybean (*Glycine max* (L) Merrill) introductions in Ethiopia based on agronomic traits. **Journal of Biology, Agriculture and Healthcare**, v.2, p.6-12, 2012.

BERGMEIR, C.; BENITEZ, J.M. Neural networks in R using the Stuttgart neural network simulator: RSNNS. **Journal of Statistical Software**, v.46, p.1-26, 2012. DOI: https://doi.org/10.18637/jss.v046.i07.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Registro Nacional de Cultivares – RNC**. Available at: <http://sistemas.agricultura.gov.br/snpc/cultivarweb/cultivares_registradas.php>. Accessed on: Jan. 23 2022.

CHAUDHARY, V.; BHATIA, R.S.; AHLAWAT, A.K. A novel self-organizing map (SOM) learning algorithm with nearest and farthest neurons. **Alexandria Engineering Journal**, v.53, p.827-831, 2014. DOI: https://doi.org/10.1016/j.aej.2014.09.007.

CRUZ C.D.; CARNEIRO P.C.S.; REGAZZI A.J. **Modelos biométricos aplicados ao melhoramento genético**. 3.ed. Viçosa: UFV, 2014. v.2, 668p.

FERREIRA, F.; SCAPIM, C.A.; MALDONADO, C.; MORA, F. SSR-based genetic analysis of sweet corn inbred lines using artificial neural networks. **Crop Breeding and Applied Biotechnology**, v.18, p.309-313, 2018. DOI: https://doi.org/10.1590/1984-70332018v18n3n45.

HAMAWAKI, O.T.; SOUSA, L.B. de; ROMANATO, F.N.; NOGUEIRA, A.P.O.; SANTOS JÚNIOR, C.D.; POLIZEL, A.C. Genetic parameters and variability in soybean genotypes. **Comunicata Scientiae**, v.3, p.76-83, 2012.

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, v.43, p.59-69, 1982. DOI: https://doi.org/10.1007/BF00337288.

KOHONEN, T. **Self-organizing maps**. 3rd ed. Berlin: Springer, 2001. 501p. (Springer Series in Information Sciences, 30). DOI: https://doi.org/10.1007/978-3-642-56927-2.

LIMA, M.D. de; PELUZIO, J.M. Dissimilaridade genética em cultivares de soja com enfoque no perfil de ácidos graxos visando produzir biocombustível. **Agrária - Revista Brasileira de Ciências Agrárias**, v.10, p.256-261, 2015. DOI: https://doi.org/10.5039/agraria.v10i2a5333.

R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016. Available at: <https://www.R-project.org/>. Accessed on: Jan. 23 2022.

SANTOS, H.G. dos; JACOMINE, P.K.T.; ANJOS, L.H.C. dos; OLIVEIRA, V.Á. de; LUMBRERAS, J.F.; COELHO, M.R.; ALMEIDA, J.A. de; ARAÚJO FILHO, J.C. de; OLIVEIRA, J.B. de; CUNHA, T.J.F. **Sistema brasileiro de classificação de solos**. 5.ed. rev. e ampl. Brasília: Embrapa, 2018. 356p. Available at: <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1094003>. Accessed on: Jan. 23 2022.

SILVA, G.N.; TOMAZ, R.S.; SANT'ANNA, I. de C.; NASCIMENTO, M.; BHERING, L.L.; CRUZ, C.D. Neural networks for predicting breeding values and genetic gains. **Scientia Agricola**, v.71, p.494-498, 2014. DOI: https://doi.org/10.1590/0103-9016-2014-0057.

SUDRÉ, C.P.; CRUZ, C.D.; RODRIGUES, R.; RIVA, E.M.; AMARAL JÚNIOR, A.T. do; SILVA, D.J.H. da; PEREIRA, T.N.S. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, v.24, p.88-93, 2006.

USDA. United States Department of Agriculture. **Market and trade data**. 2021. Available at: <www.fas.usda.gov>. Accessed on: June 8 2021.

VAL, B.H.P.; FERREIRA JÚNIOR, J.A.; BIZARI, E.H.; DI MAURO, A.O.; UNÊDA TREVISOLI, S.H. Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. **Ciência & Tecnologia**, v.6, p.72-83, 2014.

VENDRUSCOLO, T.P.S.; SILVA, V.P. da; FELIPIN-AZEVEDO, R.; SILVA, R.S. da; CASTRILLON, M.A. de S.; CORRÊA, C.L.; TARDIN, F.D.; BARELLI, M.A.A. Genetic divergence in biomass sorghum genotypes through agronomic and physical-chemical characters. **Research, Society and Development**, v.9, e552997536, 2020. DOI: https://doi.org/10.33448/rsd-v9i9.7536.

WEI, T.; SIMKO, V. **R package 'corrplot'**: visualization of a correlation matrix. Version 0.92. 2021. Available at: <https://github.com/taiyun/corrplot>. Accessed on: Jan. 23 2022.