

Computational intelligence for studies on genetic diversity between genotypes of biomass sorghum

Abstract – The objective of this work was to evaluate the potential of computational intelligence and canonical variables for studies on the genetic diversity between biomass sorghum (*Sorghum bicolor*) genotypes. The experiments were carried out in the experimental field of Embrapa Milho e Sorgo, in the municipalities of Nova Porteirinha and Sete Lagoas, in the state of Minas Gerais, Brazil. The following traits were evaluated: days to flowering, plant height, fresh biomass yield, total dry biomass, and dry biomass yield. The study of genetic diversity was performed through the analysis of canonical variables. For the recognition of the organization pattern of genetic diversity, Kohonen's self-organizing map was used. The use of canonical variables and a self-organizing map were efficient for the study of genetic diversity. The application of computational intelligence using a self-organized map is promising and efficient for studies on the genetic diversity between biomass sorghum genotypes.

Index terms: *Sorghum bicolor*, canonical variables, Kohonen's self-organized map.

Inteligência computacional para estudos de diversidade genética entre genótipos de sorgo biomassa

Resumo – O objetivo deste trabalho foi avaliar o potencial da inteligência computacional e de variáveis canônicas para estudos de diversidade genética entre genótipos de sorgo (*Sorghum bicolor*) biomassa. Os experimentos foram realizados no campo experimental da Embrapa Milho e Sorgo, nos municípios de Nova Porteirinha e Sete Lagoas, no estado de Minas Gerais, Brasil. Foram avaliadas as seguintes características: dias para a floração, altura das plantas, produção de biomassa fresca, biomassa seca total e produção de biomassa seca. O estudo da diversidade genética foi realizado por meio da análise de variáveis canônicas. Para o reconhecimento do padrão de organização da diversidade genética, utilizou-se o mapa auto-organizável de Kohonen. O uso de variáveis canônicas e o mapa auto-organizável foram eficientes no estudo de diversidade genética. A utilização de inteligência computacional com mapa auto-organizável é promissora e eficiente para estudos de diversidade genética entre genótipos de sorgo biomassa.

Termos para indexação: *Sorghum bicolor*, variáveis canônicas, mapa auto-organizável de Kohonen.

Michele Jorge da Silva⁽¹⁾ ,
Antônio Carlos da Silva Júnior⁽¹⁾ ,
Cosme Damião Cruz⁽¹⁾ ,
Moysés Nascimento⁽²⁾ ,
Marciane da Silva Oliveira⁽¹⁾ ,
Robert Eugene Schaffert⁽³⁾  and
Rafael Augusto da Costa Parrella⁽³⁾ 

⁽¹⁾ Universidade Federal de Viçosa,
Departamento de Biologia Geral, Avenida
Peter Henry Rolfs, s/nº, Campus Universitário,
CEP 36570-000 Viçosa, MG, Brazil.
E-mail: michelejorgesilva@gmail.com,
antonio.silva.c.junior@gmail.com,
cdcruz@ufv.br,
marciane.silva.oliveira@gmail.com

⁽²⁾ Universidade Federal de Viçosa,
Departamento de Estatística, Avenida Peter
Henry Rolfs, s/nº, Campus Universitário,
CEP 36570-977 Viçosa, MG, Brazil.
E-mail: moysesnascim@ufv.br

⁽³⁾ Embrapa Milho e Sorgo, Rodovia MG-424,
Km 45, Caixa Postal 285, CEP 35701-970
Sete Lagoas, MG, Brazil.
E-mail: robert.schaffert@embrapa.br,
rafael.parrella@embrapa.br

 Corresponding author

Received
November 24, 2019

Accepted
October 28, 2020

How to cite
SILVA, M.J. da; SILVA JÚNIOR, A.C.
da; CRUZ, C.D.; NASCIMENTO, M.;
OLIVEIRA, M. da S.; SCHAFFERT, R.E.;
PARRELLA, R.A. da C. Computational
intelligence for studies on genetic diversity in
biomass sorghum. *Pesquisa Agropecuária
Brasileira*, v.55, e01723, 2020. DOI: [https://doi.
org/10.1590/S1678-3921.pab2020.v55.01723](https://doi.org/10.1590/S1678-3921.pab2020.v55.01723).

Introduction

Biomass sorghum [*Sorghum bicolor* (L.) Moench] is a crop with a high potential as a biomass source for bioenergy generation, both for obtaining second generation ethanol (Amaducci et al., 2000; Davila-Gomez et al., 2011; Silva et al., 2018, 2020) and for the energy cogeneration by direct biomass burning (Bennett & Anex, 2009; Zegada-Lizarazu & Monti, 2012).

For the generation of sorghum hybrids, three types of lines – called A, B, and R – are required (House, 1985). The A and B lines are isogenic and differ only for cytoplasm: A line has cytoplasm that confers the male sterility phenotype, when associated with recessive nuclear genes for restoration of fertility; and B line has normal cytoplasm, and the plant has the fertile male part, even with nuclear recessive alleles for fertility restoration (Smith & Frederiksen, 2000). Thus, the hybrid is obtained by crossing the A line (female) with an R line (restorative) that has dominant alleles for the fertility restoration gene.

Traditional genetic diversity studies, in which a group of different accessions are evaluated, are carried out with the purpose of identifying contrasting and higher potential parents that will form the base population, or hybrids (Cruz et al., 2011). However, the study of genetic diversity allows of other inferences, such as the possibility of recognizing inheritance patterns, especially in situations in which the cluster represents the proximity of hybrid combinations and their parents. In this situation, the clustering pattern can reflect the similarity of parents, and also important genetic phenomena, such as dominance and epistasis that produce similar phenotypes from different genotypic combinations (Ferrão et al., 2002). To determine how far one population or genotype is from another, biometric methods based on the quantification of heterosis, or on predictive processes of heterosis, are used. Predictive methods are usually quantified by measurements of dissimilarity (Cruz et al., 2012). Several methods may be used, including principal component analysis, canonical variables, and agglomerative methods (Guidoti et al., 2018). The choice of the method relies on the goal of the study, desired precision by the breeder, ease of analysis, and way the data are obtained (Cruz et al., 2012).

The computational intelligence approach aiming at the genetic improvement can be used to assist breeder

decision-making (Crain et al., 2018; Montesinos-López et al., 2018; Harfouche et al., 2019). From a database of well conducted experiments that characterize accessions in different environmental conditions, genotypes can be selected more efficiently by artificial neural networks (ANN), particularly to recognize the organization of this diversity in analyses based on self-organizing map (SOM) (Barbosa et al., 2011; Santos et al., 2019; Wolski & Kruk, 2020). SOMs are a type of ANN developed by Kohonen (1982) that are trained by using unsupervised learning to project high-dimensional, complex data into a two-dimensional grid (Kohonen, 1982, 2014; Augustijn & Zurita-Milla, 2013). In addition, SOMs have a nonlinear structure in which they can capture more complex data characteristics, either quantitative or qualitative, which is not always possible using traditional statistical techniques (Galvão et al., 1999; Santos et al., 2019).

Preliminary studies are performed from the population information (in the present study, regarding parents and hybrids) previously known in network training stage. Then, the efficiency of network discrimination with the adjusted structure for recognition purposes of organizational pattern is verified, allowing of the inferences on the phenotypic similarity from distinct genotypes.

The objective of this work was to evaluate the potential of computational intelligence and canonical variables for studies on the genetic diversity between biomass sorghum genotypes.

Materials and Methods

The experiments were carried out in the experimental field of Embrapa Milho e Sorgo, in the state of Minas Gerais (MG), Brazil, in the municipalities of Nova Porteirinha, northern of the state (15°48'S; 43°18'W), and Sete Lagoas (9°27'S; 44°15'W). Four experiments were conducted in different crop years and locations (Table 1), as also described by Silva et al. (2018).

In the crop year 2016/2017, 30 hybrids, the parent lines, and six non-mutant controls for the *bmr* gene (201636B005, 201636B006, 201636B008, 201636B019, 201636B004, and 'BRS 716') were evaluated in a triple lattice (7x7) design, at the experimental unit of Embrapa Milho e Sorgo, in Nova Porteirinha, MG. The sowing occurred on October 27, 2016. In the same crop year 2016/2017, 30 hybrids, the parent lines, and the same six non-mutant controls for the *bmr* gene,

were also evaluated in a triple lattice (7x7) design, at the experimental unit Embrapa Milho e Sorgo, in Sete Lagoas, MG. The sowing occurred on November 21, 2016.

In the 2017/2018 crop year, 40 hybrids, the parent lines, and two non-mutant controls for the *bmr* gene ('BRS 716' and N52K1009) were evaluated in an alpha-lattice (7x8) design, at Embrapa Milho e Sorgo, in the municipality of Nova Porteirinha, MG. The sowing occurred on October 28, 2017. In the same 2017/2018 crop year, 40 hybrids, the parent lines, and two non-mutant controls for the *bmr* gene ('BRS 716' and N52K1009) were evaluated in an alpha-lattice (7x8) design, at Embrapa Milho e Sorgo, in the municipality of Sete Lagoas, MG. The sowing occurred on October 26, 2017.

The plots of four experiments consisted of double 3 m rows, spaced at 0.70m. Plants were thinned two weeks after seedling emergence, maintaining a population of approximately 110,000 plants per ha. Fertilization was performed with 450 kg ha⁻¹ of the formulation 08-

28-16 (N-P₂O₅-K₂O) applied in the planting grooves, plus 200 kg ha⁻¹ urea as topdressing at 20–25 days after sowing. Weed control was performed by the application of atrazine and manual weeding. The other cultural treatments related to pests and disease control were carried out following the recommendations for the culture in each region (Coelho, 2015). Plants were harvested when they reached the physiological maturity of the grain.

The following traits were evaluated: days to flowering (FLOW, number of days), which consists of the days between sowing and the pollen liberation of 50% of the plants in the plot; plant height (PH, m), which is the mean height of the plants within the plot, measured from the soil surface to the top of the panicle; fresh biomass yield (FBY, Mg ha⁻¹), which was determined by weighing all plants of the useful area; total dry biomass (DB, %), which is the quantification of the dry matter content, by weighing samples of each treatment, and determining the green weight; green weight (GW). Samples were then stored

Table 1. Biomass sorghum (*Sorghum bicolor*) parents and hybrids used in the partial diallel in experiments I, II, III and IV⁽¹⁾.

Group II (R line)	Group I (A line)			
	BR007A bmr (1)	CMSXS156A bmr (2)	CMSXS157A bmr (3)	TX635Abmr(4)
Experiments I and II				
201635B004 bmr (1)	H ₍₁₋₁₎	H ₍₂₋₁₎	H ₍₃₋₁₎	-
201635B005 bmr (2)	H ₍₁₋₂₎	H ₍₂₋₂₎	H ₍₃₋₂₎	-
201635B006 bmr (3)	H ₍₁₋₃₎	H ₍₂₋₃₎	H ₍₃₋₃₎	-
201635B007 bmr (4)	H ₍₁₋₄₎	H ₍₂₋₄₎	H ₍₃₋₄₎	-
201635B008 bmr (5)	H ₍₁₋₅₎	H ₍₂₋₅₎	H ₍₃₋₅₎	-
201635B009 bmr (6)	H ₍₁₋₆₎	H ₍₂₋₆₎	H ₍₃₋₆₎	-
201635B010 bmr (7)	H ₍₁₋₇₎	H ₍₂₋₇₎	H ₍₃₋₇₎	-
201635B011 bmr (8)	H ₍₁₋₈₎	H ₍₂₋₈₎	H ₍₃₋₈₎	-
201635B012 bmr (9)	H ₍₁₋₉₎	H ₍₂₋₉₎	H ₍₃₋₉₎	-
201635B013 bmr (10)	H ₍₁₋₁₀₎	H ₍₂₋₁₀₎	H ₍₃₋₁₀₎	-
Experiments III and IV				
201635B004 bmr (1)	H ₍₁₋₁₎	H ₍₂₋₁₎	H ₍₃₋₁₎	H ₍₄₋₁₎
201635B005 bmr (2)	H ₍₁₋₂₎	H ₍₂₋₂₎	H ₍₃₋₂₎	H ₍₄₋₂₎
201635B006 bmr (3)	H ₍₁₋₃₎	H ₍₂₋₃₎	H ₍₃₋₃₎	H ₍₄₋₃₎
201635B007 bmr (4)	H ₍₁₋₄₎	H ₍₂₋₄₎	H ₍₃₋₄₎	H ₍₄₋₄₎
201635B008 bmr (5)	H ₍₁₋₅₎	H ₍₂₋₅₎	H ₍₃₋₅₎	H ₍₄₋₅₎
201635B009 bmr (6)	H ₍₁₋₆₎	H ₍₂₋₆₎	H ₍₃₋₆₎	H ₍₄₋₆₎
201635B010 bmr (7)	H ₍₁₋₇₎	H ₍₂₋₇₎	H ₍₃₋₇₎	H ₍₄₋₇₎
201635B011 bmr (8)	H ₍₁₋₈₎	H ₍₂₋₈₎	H ₍₃₋₈₎	H ₍₄₋₈₎
201635B012 bmr (9)	H ₍₁₋₉₎	H ₍₂₋₉₎	H ₍₃₋₉₎	H ₍₄₋₉₎
201635B013 bmr (10)	H ₍₁₋₁₀₎	H ₍₂₋₁₀₎	H ₍₃₋₁₀₎	H ₍₄₋₁₀₎

⁽¹⁾Values in parentheses refer to the 4 lines of group I, 10 lines of group II, and the 40 hybrids derived from the crosses between these lines for the experiments I and II, as well as to the 4 lines of group I, 10 lines of group II, and 40 hybrids from crosses between these lines for the experiments III and IV.

in paper bags and taken to a forced-air circulation oven (65°) for about 72 hours, to determine the dry weight (DW). Afterward, the following expression was used: $DB = (DW/GW) \times 100$. Also, dry biomass yield (DBY, Mg ha⁻¹) was measured by the expression: $DBY = DB \times FBY$.

The joint analysis of variance was performed to evaluate the traits, in Nova Porteirinha and Sete Lagoas, in the 2016/2017 and 2017/2018 crop years, for the common genotypes, considering the effects of treatment and environment as fixed, according to the following equation:

$$y_{ijk} = \mu + g_i + e_j + ge_{ij} + b/e_{jk} + \varepsilon_{ijk}$$

in which: y_{ijk} is the observed phenotypic value of the i^{th} genotype, in the k^{th} block, within the j^{th} environment; g_i is the effect of the i^{th} genotype; e_j is the effect of the j^{th} environment; ge_{ij} is the effect of the interaction of the i^{th} genotype with the j^{th} environment; b/e_{jk} effect of the k^{th} block within the j^{th} environment; and ε_{ijk} is the effect of the experimental error.

The study of genetic diversity was performed by graphic dispersion through canonical variables analysis, which is evidenced by the dispersion of scores in graphs, with the axes represented by the first two canonical variables. Thus, when determining the number of canonical variables that accumulate a minimum of 80% of the total available variance, we estimate the scores for each canonical variable that can be plotted on two or three dimensional graphs, and the

canonical variables are used as reference axes, in which graphical distances that represent similarities and dissimilarities between genotypes can be visualized (Piassi et al., 1995).

To recognize the organization of diversity, Kohonen's self-organizing maps (SOMs) were used. Different network architectures were tested by varying the number of rows (1 to 5) and columns (1 to 5). Kohonen (2001) points out that determining the number of neurons and learning parameters is an empirical process based on the researcher's experience, and trial and error methods. Then, to select the best network architecture, 2,000 trainings were done for each of the combinations. The defined topology was hextop, that is, with hexagonal neighborhood; the distance used to configure the artificial neural networks was the Euclidean distance. The software Matlab (Matlab, 2012) and Genes (Cruz, 2013) were used to perform the analysis.

Results and Discussion

In the four experiments, there were effects of genotypes (hybrids, parents, and controls) for most traits, in both crop years, except for dry biomass in 2017/2018, showing wide variability between genotypes (Table 2). Estimates of coefficients of variation (CV%) indicated an adequate experimental precision, according to other studies on biomass

Table 2. Summary of the joint analysis of variance for days to flowering (FLOW), plant height (PH), fresh biomass yield (FBY), total dry biomass (DB), and dry biomass yield (DBY) for the hybrids, parents, and controls of biomass sorghum (*Sorghum bicolor*), in the 2016/2017 and 2017/2018 crop years, in the municipalities of Nova Porteirinha and Sete Lagoas, in the state of Minas Gerais, Brazil.

Source of variation	Degree of freedom	Mean squares				
		FLOW	PH	FBY	DB	DBY
2016/2017 crop year						
Genotype (G)	43	3,085.41**	4.49**	4,241.95**	227.26**	304.36*
Environment (E)	1	3,354.09 ^{ns}	41.68**	36,859.01 ^{ns}	2,877.84 ^{ns}	3,940.75 ^{ns}
G x E	43	772.88**	0.32**	1,000.94**	1,000.19**	163.62**
Mean		131.68	4.10	100.74	31.96	36.30
Coefficient of variation (%)		3.86	5.14	6.88	8.09	6.54
2017/2018 crop year						
Genotype (G)	43	1,662.69**	5.15**	3,797.85**	147.89 ^{ns}	146.67**
Environment (E)	1	2,310.45 ^{ns}	3.18 ^{ns}	2,583.00 ^{ns}	2,498.49 ^{ns}	796.78 ^{ns}
G x E	43	466.79**	0.13**	1,085.58**	106.65**	80.37**
Mean		131.43	4.15	83.34	26.92	26.0
Coefficient of variation (%)		4.52	5.57	5.28	7.93	6.38

** , *Significant at 1% and 5% probability, by the F test. ^{ns}Nonsignificant.

sorghum (Silva et al., 2018; Oliveira et al., 2019). There was a significant effect for genotype x environment interaction (G x E) for all traits, in both crop years, indicating different behavior of genotypes at different environments. Thus, the genetic diversity between genotypes was studied separately for each environment, since the clustering pattern can be changed according to the environmental variation.

In Nova Porteirinha (first crop year, 2016/2017), the first two variables explained 92.94% of the total variation, according to values associated with the canonical variable (Table 3). When the first two canonical variables explain over 80% of the total variation, their use is satisfactory in the study of genetic diversity, by assessing the graphic dispersion of scores in relation to the canonical variables CV1 and CV2 (Cruz et al., 2012). In Sete Lagoas (2016/2017), the first two variables explained 97.45% of the total variation. In the second crop year (2017/2018), the first three variables explained 94.36% of total variation in Nova Porteirinha, and 94.71% of total variation in Sete Lagoas. Based on these results, the two-dimensional graphic dispersion using CV1 and CV2 was made (Figure 1).

In all the evaluated experiments, the A lines were observed as more divergent than the other lines (Figure 1). This can be verified because these lines are male-sterile, carrying the *bmr-6* allele, of low height and insensitive to photoperiod. In contrast, R lines

and hybrids were closer to each other, showing that R lines contribute to a higher concentration of favorable and dominant alleles for the formation of hybrids. In addition, the R lines carry the *bmr-6* allele, tall, with excellent biomass yield, and sensitive to photoperiod (Oliveira et al., 2019).

When transmitting their gametes, the genotypes with high concentration of dominant alleles leave more similar progeny than those with high concentration of recessive allele. In addition, dominant favorable alleles provide larger numbers of superior genotypes in the segregating generation, allowing of the identification of larger individuals with smaller population size. Another observation was that the controls evaluated in the 2016/2017 crop year were very close to each other, in comparison with the controls evaluated in the 2017/2018 crop year.

However, it is necessary to consider that the trait plant height, in sorghum, is controlled by four independently inherited *Dw* (Dwarf) genes, viz., *Dw1*, *Dw2*, *Dw3*, and *Dw4* (Hilley et al., 2016; Shukla et al., 2017). Plant height is determined by the interplay of the internode length and the number of nodes that the plant produces before flowering. The *Dw* genes have partial dominance for tallness, and their effects are additive in nature (Shukla et al., 2017). There may be also recessive epistasis in the manifestation of the combined additive effect of two or more genes for the other evaluated traits. That is, the observed phenotype of hybrids may not be explained and, in this case, the combined effect of these genes on a phenotype cannot be predicted from the sum of their individual effects (Leite et al., 2020).

For self-organizing maps, it was found that the network architecture using five columns and five rows for the four experiments was efficient (Figure 2). Several studies using the SOM network have also defined their topology either tentatively or randomly (Barbosa et al., 2011; Chaudhary et al., 2014; Gámez Albán et al., 2016; Santos et al., 2019). Therefore, the method for finding the best network architecture is very important because each time SOM networks are used, different results can be obtained, since the networks have random synaptic weights at the beginning of the training (Santos et al., 2019).

The lines were organized in different neurons, which are represented by the hexagons. Filling in the area of the hexagons indicates the concentration of

Table 3. Estimates of variance (eigenvalues) associated with canonical variables, related to five traits evaluated in biomass sorghum (*Sorghum bicolor*) in the municipalities Nova Porteirinha (NP) and Sete Lagoas (SL), in the state of Minas Gerais, Brazil.

Canonical variable accumulated (%)			
NP (2016/2017)		SL (2016/2017)	
CV1	79.45	CV1	90.00
CV2	92.94	CV2	97.45
CV3	98.93	CV3	98.84
CV4	99.75	CV4	99.47
CV5	100.0	CV5	100.0
NP (2017/2018)		SL (2017/2018)	
CV1	74.49	CV1	72.48
CV2	89.53	CV2	84.64
CV3	94.36	CV3	94.71
CV4	97.88	CV4	98.81
CV5	100.0	CV5	100.0

CV, coefficient of variation.

the lines contained in that neuron and, the higher the concentration of lines, the larger the filled area. Thus, the neurons that have the largest number of lines are represented by the complete filling of the hexagon (Figure 2). Neurons are composed of individuals who have some similarity to the neighboring class; the most divergent classes constitute the regions of the extremes, and the intermediate classes constituted the center of the map. Thus, genotypes contained in the most divergent classes show a greater dissimilarity and can be considered the most genetically distant. It is possible to verify which are

the genotypes in each neuron (Table 4), through the organization made by the Kohonen's maps. The topology used was able to organize the genotypes in groups closer to those obtained by the canonical variables (Figure 1). The SOM method proved to be efficient in the identification of similarity patterns and in the organization of the proximity of genotypes between groups, according to the grouping carried out, as shown by Santos et al. (2019), who used the SOM technique as an alternative method to evaluate the genetic diversity in plant breeding programs. However, in cases with higher numbers of neurons,

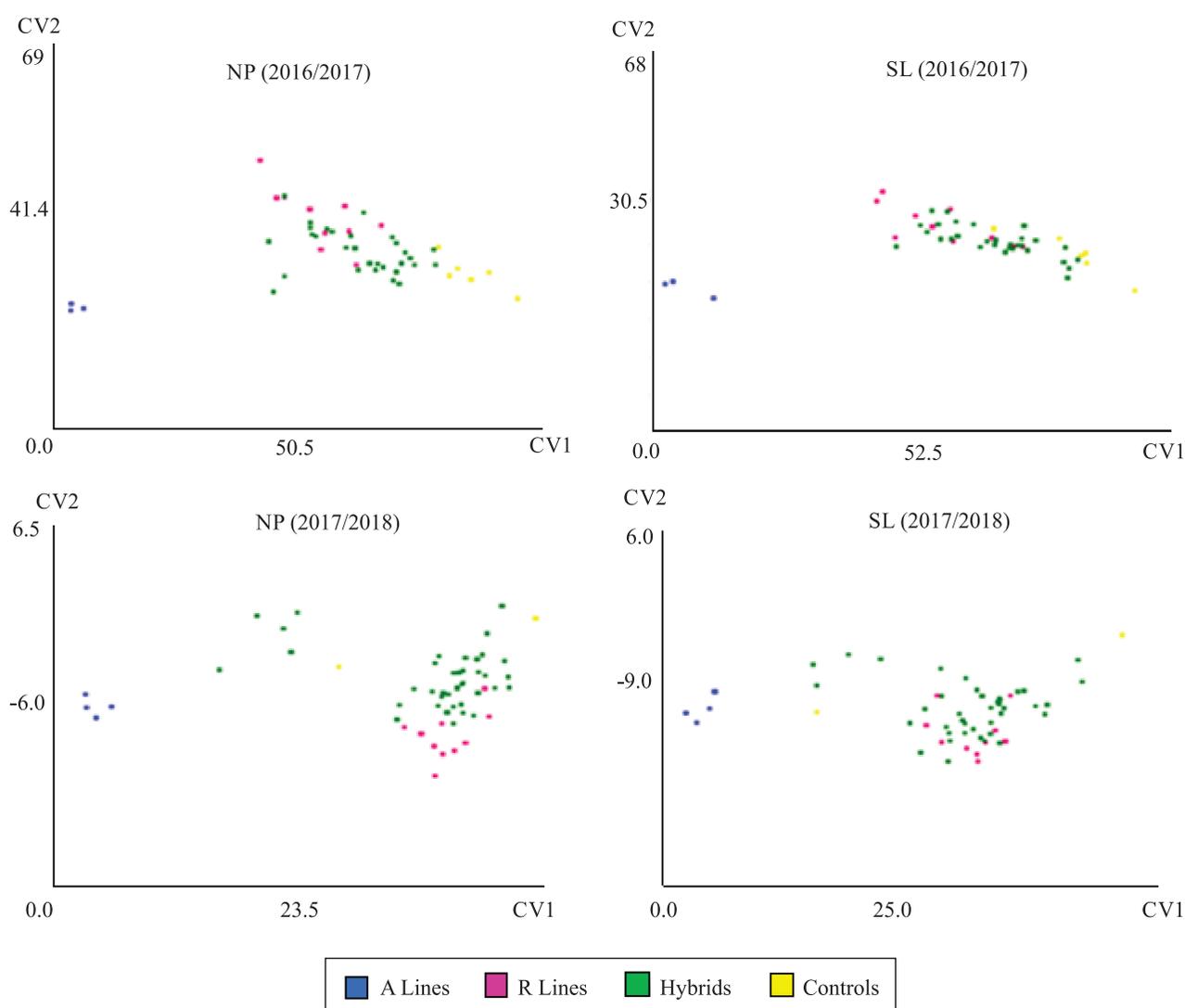


Figure 1. Graphical dispersion of biomass sorghum (*Sorghum bicolor*) lines based on the first two canonical variables (CV1 and CV2), which were obtained through the analysis of five traits evaluated in Nova Porteirinha (NP) and Sete Lagoas (SL), in the state of Minas Gerais, Brazil.

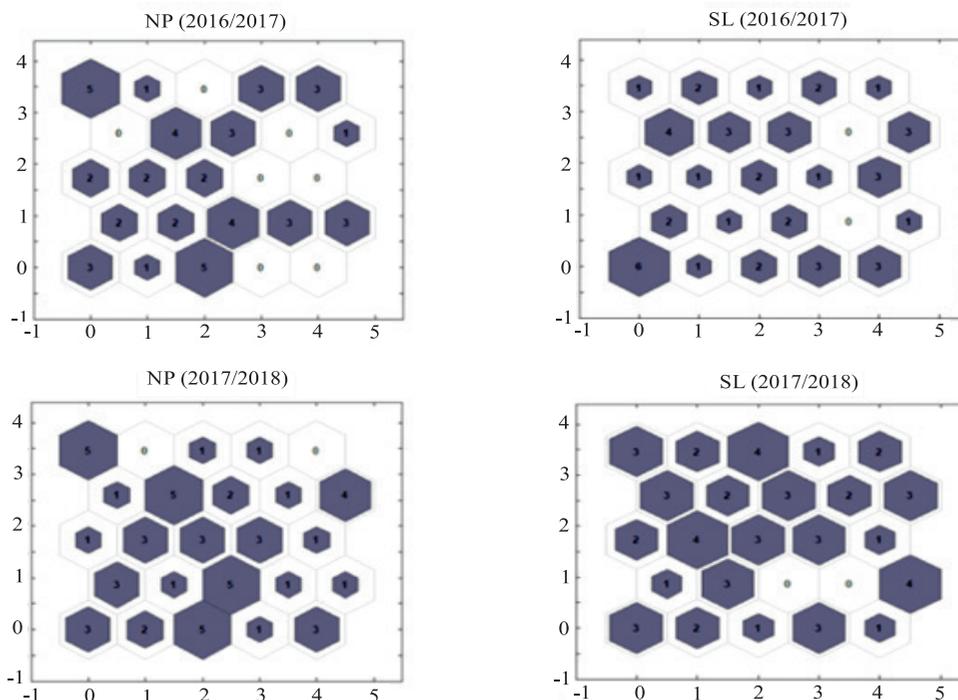


Figure 2. Arrangement of the clustering scenarios, using five columns and five rows for four experiments, according to the self-organizing map (SOM) for Nova Porteirinha (NP) and Sete Lagoas (SL), in the state of Minas Gerais, Brazil.

Table 4. Organization of genotypes in neurons through the self-organizing map (SOM) for Nova Porteirinha (NP) and Sete Lagoas (SL) in the 2016/2017 and 2017/2018 crop years, in the state of Minas Gerais, Brazil.

Neuron	2016/2017 crop year		2017/2018 crop year	
	Nova Porteirinha (NP)	Sete Lagoas (SL)	Nova Porteirinha (NP)	Sete Lagoas (SL)
1	44, 47, 48	7, 8, 27, 28, 31, 35	18, 23, 56	18, 19, 56
2	25	39	48, 51	17, 24
3	27, 31, 36, 37, 41	22, 36	22, 38, 39, 40, 41	20
4		6, 14, 30	34	16, 30, 49
5		11, 13, 21	20, 47, 49	47
6	46, 49	19, 41	19, 44, 54	21
7	16, 26	37	15	6, 28, 53
8	5, 14, 22, 38,	4, 32	8, 27, 29, 31, 50	
9	19, 29, 40		45	
10	1, 2, 3	12	30	1, 2, 3, 4
11	15, 45	24	36	32, 36
12	17, 18	42	17, 24, 26	40, 43, 52, 54
13	20, 35	9, 29	9, 21, 42	22, 23, 39
14		40	35, 46, 52	12, 33, 34
15		33, 34, 43	18	55
16		25, 38, 47	14	15, 37, 42
17	6, 9, 34, 39	16, 26, 49	5, 6, 7, 11, 13	5, 51
18	24, 32, 33	15, 20, 23	12, 25	29, 31, 41
19			16	7, 27
20	13	1, 2, 3	1, 2, 3, 4	10, 25, 45
21	7, 8, 10, 21, 23	44	10, 32, 37, 43, 53	38, 48, 50
22	28	18, 48		26, 46
23		17	33	8, 9, 13, 14
24	11, 30, 42	10	55	11
25	4	5		35, 44

there is a possibility of larger variation amounts in the allocation of genotypes, as affirmed by Kohonen (2014). Thus, the use of artificial neural networks as pattern recognition methods is a promising way, and the SOM network can provide more valuable results than traditional cluster analysis.

Conclusion

The application of computational intelligence using a self-organized map is promising and efficient for studies on the genetic diversity between biomass sorghum (*Sorghum bicolor*) genotypes.

Acknowledgments

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for scholarships and financial support; and to Banco Nacional de Desenvolvimento Econômico e Social (BNDES), for financial support.

References

- AMADUCCI, S.; AMADUCCI, M.T.; BENATI, R.; VENTURI, G. Crop yield and quality parameters of four annual fibre crops (hemp, kenaf, maize and sorghum) in the North of Italy. **Industrial Crops and Products**, v.11, p.179-186, 2000. DOI: [https://doi.org/10.1016/S0926-6690\(99\)00063-1](https://doi.org/10.1016/S0926-6690(99)00063-1).
- AUGUSTIJN, E.-W.; ZURITA-MILLA, R. Self-organizing maps as an approach to exploring spatiotemporal diffusion patterns. **International Journal of Health Geographics**, v.12, art.60, 2013. DOI: <https://doi.org/10.1186/1476-072X-12-60>.
- BARBOSA, C.D.; VIANA, A.P.; QUINTAL, R.S.S.; 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>.
- BENNETT, A.S.; ANEX, R.P. Production, transportation and milling costs of sweet sorghum as a feedstock for centralized bioethanol production in the upper Midwest. **Bioresource Technology**, v.100, p.1595-1607, 2009. DOI: <https://doi.org/10.1016/j.biortech.2008.09.023>.
- 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>.
- COELHO, A.M. Nutrição e adubação. In: RODRIGUES, J.A.S. (Ed). **Cultivo do sorgo**. Brasília: Embrapa, 2015. (Embrapa Milho e Sorgo. Sistemas de produção, 2). Disponível em: <https://www.spo.cnptia.embrapa.br/conteudo?p_p_id=conteudoportlet_WAR_sistemasdeproducaoof6_lgalceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&p_r_p_-76293187_sistemaProducaoId=8301&p_r_p_-996514994_topicoId=9204>. Acesso em: 30 jun. 2020.
- CRAIN, J.; MONDAL, S.; RUTKOSKI, J.; SINGH, R.P.; POLAND, J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. **Plant Genome**, v.11, art.170043, 2018. DOI: <https://doi.org/10.3835/plantgenome2017.05.0043>.
- CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v.35, p.271-276, 2013. DOI: <https://doi.org/10.4025/actasciagron.v35i3.21251>.
- CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620p.
- CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 4.ed. Viçosa: UFRV, 2012.
- DAVILA-GOMEZ, F.J.; CHUCK-HERNANDEZ, C.; PEREZ-CARRILLO, E.; ROONEY, W.L.; SERNA-SALDIVAR, S.O. Evaluation of bioethanol production from five different varieties of sweet and forage sorghums (*Sorghum bicolor* (L) Moench). **Industrial Crops and Products**, v.33, p.611-616, 2011. DOI: <https://doi.org/10.1016/j.indcrop.2010.12.022>.
- FERRÃO, M.A.G.; VIEIRA, C.; CRUZ, C.D.; CARDOSO, A.A. Divergência genética em feijoeiro em condições de inverno tropical. **Pesquisa Agropecuária Brasileira**, v.7, p.1089-1098, 2002. DOI: <https://doi.org/10.1590/S0100-204X2002000800006>.
- GALVÃO, C.O.; VALENÇA, M.J.S.; VIEIRA, V.P.P.B.; DINIZ, L.S.; LACERDA, E.G.M.; CARVALHO, A.C.P.L.F.; LUDERMIR, T.B. **Sistemas inteligentes: aplicações e recursos hídricos e ciências ambientais**. Porto Alegre: UFRGS/ABRH, 1999. 246p.
- GÁMEZ ALBÁN, H.M.; OREJUELA CABRERA, J.P.; SALAS ACHIPIZ, O.A.; BRAVOBASTIDAS, J.J. Aplicación de mapas de Kohonen para la priorización de zonas de mercado: una aproximación práctica. **Revista EIA**, v.13, p.157-169, 2016. DOI: <https://doi.org/10.24050/reia.v13i25.1024>.
- GUIDOTI, D.T.; GONELA, A.; VIDIGAL, M.C.G.; CONRADO, T.V.; ROMANI, I. Interrelationship between morphological, agronomic and molecular characteristics in the analysis of common bean genetic diversity. **Acta Scientiarum. Agronomy**, v.40, e33032, 2018. DOI: <https://doi.org/10.4025/actasciagron.v40i1.33032>.
- HARFOUCHE, A.L.; JACOBSON, D.A.; KAINER, D.; ROMERO, J.C.; HARFOUCHE, A.H.; SCARASCIA MUGNOZZA, G.; MOSHELION, M.; TUSKAN, G.A.; KEURENTJES, J.J.B.; ALTMAN, A. Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. **Trends in Biotechnology**, v.37, p.1217-1235, 2019. DOI: <https://doi.org/10.1016/j.tibtech.2019.05.007>.
- HILLEY, J.; TRUONG, S.; OLSON, S.; MORISHIGE, D.; MULLET, J. Identification of *Dw1*, a regulator of sorghum

- stem internode length. **PLoS ONE**, v.11, e0151271, 2016. DOI: <https://doi.org/10.1371/journal.pone.0151271>.
- HOUSE, L.R. **A guide to sorghum breeding**. 2nd ed. Patancheru: International Crops Research Institute for the Semi-Arid Tropics, 1985. 206 p.
- KOHONEN, T. **MATLAB implementations and applications of the self-organizing map**. Helsinki: Unigrafia Oy, 2014.
- 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. DOI: <https://doi.org/10.1007/978-3-642-56927-2>.
- LEITE, P.S. da S.; BOTELHO, T.T.; RIBEIRO, P.C. de O.; SCHAFFERT, R.E.; PARRELLA, R.A. da C.; NUNES, J.A.R. Intrapopulation recurrent selection in sweet sorghum for improving sugar yield. **Industrial Crops & Products**, v.143, art.111910, 2020. DOI: <https://doi.org/10.1016/j.indcrop.2019.111910>.
- MATLAB. version 7.10.0. Natick: The Math Works Inc., 2012. Software.
- MONTESINOS-LÓPEZ, A.; MONTESINOS-LÓPEZ, O.A.; GIANOLA, D.; CROSSA, J.; HERNÁNDEZ-SUÁREZ, C.M. Multi-environment genomic prediction of plant traits using deep learners with dense architecture. **G3: Genes, Genomes, Genetics**, v.8, p.3813-3828, 2018. DOI: <https://doi.org/10.1534/g3.118.200740>.
- OLIVEIRA, I.C.M.; MARÇAL, T. de S.; BERNARDINO, K. da C.; RIBEIRO, P.C. de O.; PARRELLA, R.A. da C.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; CARNEIRO, J.E. de S. Combining ability of biomass sorghum lines for agroindustrial characters and multitrail selection of photosensitive hybrids for energy cogeneration. **Crop Science**, v.59, p.1554-1566, 2019. DOI: <https://doi.org/10.2135/cropsci2018.11.0693>.
- PIASSI, M.; SILVA, M.A.; REGAZZI, A.J.; TORRES, R. de A.; SOARES, P.R.; TORRES JUNIOR, R.A. de A. Estudo da divergência genética entre oito grupos de aves de posturn por melo de técnicas de análise multivariada. **Revista Brasileira de Zootecnia**, v.24, p.715-727, 1995.
- SANTOS, I.G. dos.; CARNEIRO, V.Q.; SILVA JUNIOR, A.C. da; CRUZ, C.D.; SOARES, P.C. Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. **Acta Scientiarum. Agronomy**, v.41, e39803, 2019. DOI: <https://doi.org/10.4025/actasciagron.v41i1.39803>.
- SHUKLA, S.; FELDERHOFF, T.J.; SABALLOS, A.; VERMERRIS, W. The relationship between plant height and sugar accumulation in the stems of sweet sorghum (*Sorghum bicolor* (L.) Moench). **Field Crops Research**, v.203, p.181-191, 2017. DOI: <https://doi.org/10.1016/j.fcr.2016.12.004>.
- SILVA, M.J. da; CARNEIRO, P.C.S.; CARNEIRO, J.E. de S.; DAMASCENO, C.M.B.; PARRELLA, N.N.L.D.; PASTINA, M.M.; SIMEONE, M.L.F.; SCHAFFERT, R.E.; PARRELLA, R.A. da C. Evaluation of the potential of lines and hybrids of biomass sorghum. **Industrial Crops & Products**, v.125, p.379-385, 2018. DOI: <https://doi.org/10.1016/j.indcrop.2018.08.022>.
- SILVA, M.J. da; DAMASCENO, C.M.B.; CARNEIRO, J.E. de S.; PEREIRA, H.D.; CARNEIRO, P.C.S.; SCHAFFERT, R.E.; PARRELLA, R.A. da C. Combining ability of biomass sorghum in different crop years and sites for bioenergy generation. **Agronomy Journal**, v.112, p.1549-1563, 2020. DOI: <https://doi.org/10.1002/agj2.20123>.
- SMITH, C.W.; FREDERIKSEN, R.A. (Ed.). **Sorghum: origin, history, technology, and production**. New York: J. Wiley & Sons, 2000. 824p.
- WOLSKI, G.J.; KRUK, A. Determination of plant communities based on bryophytes: the combined use of Kohonen artificial neural network and indicator species analysis. **Ecological Indicators**, v.113, art.106160, 2020. DOI: <https://doi.org/10.1016/j.ecolind.2020.106160>.
- ZEGADA-LIZARAZU, W.; MONTI, A. Are we ready to cultivate sweet sorghum as a bioenergy feedstock? A review on field management practices. **Biomass and Bioenergy**, v.40, p.1-12, 2012. DOI: <https://doi.org/10.1016/j.biombioe.2012.01.048>.
-