Self-organizing maps in the study of genetic diversity among irrigated rice genotypes

Iara Gonçalves dos Santos[1*] , Vinícius Quintão Carneiro[1], Antônio Carlos da Silva Junior[1], Cosme Damião Cruz[1] and Plínio César Soares[2]

[1]Laboratório de Bioinformática, Universidade Federal de Viçosa, Av. PH Rolfs, s/n, 36570-000, Campus Universitário, Viçosa, Minas Gerais, Brazil. [2]Empresa de Pesquisa Agropecuária de Minas Gerais, Viçosa, Minas Gerais, Brazil. *Author for correspondence. E-mail: iara.santos@ufv.br

ABSTRACT. This study presents self-organizing maps (SOM) as an alternative method to evaluate genetic diversity in plant breeding programs. Twenty-five genotypes were evaluated in two environments for 11 phenotypic traits. The genotypes were clustered according to the SOM technique, with variable topology and numbers of neurons. In addition to the SOM analysis, unweighted pair group method with arithmetic mean clustering (UPGMA) was performed to observe the behavior of the clustering when submitted to these techniques and to evaluate their complementarities. Genotype ordering according to SOM was consistent with UPGMA results, evidenced by the basic structure of UPGMA groups being preserved in each group of the maps. Regarding genotype arrangement and the group neighbors, maps involving five neurons presented inferior organization efficiency compared to the six-map arrangements in both environments. It was observed that the organization pattern among the rice genotypes evaluated by the maps was complementary to the UPGMA approach, as observed in all scenarios. It can be concluded that self-organizing maps have the potential to be useful for genetic diversity studies in breeding programs.

Keywords: *Oriza sativa* L.; computational intelligence; clustering technique; SOM.

Introduction

Rice (*Oryza sativa* L.) is one of the most important crops in the world and is considered one of the main annual crops in Brazil. With the increase in population, demand has increased throughout the years, and it is estimated that by 2050 global rice production must increase from 60 to 110% to supply the population demand (Godfray et al., 2010; Tilman, Balzer, Hill, & Befort, 2011; Ray, Mueller, West, & Foley, 2013). However, this will only be possible as long as genetic variability is maintained. In Brazilian irrigated rice breeding programs, genotypic variability is restricted (Rabelo, Guimarães, Pinheiro, & Silva, 2015; Streck, Aguiar, Magalhães Júnior, Facchinello, & Oliveira, 2017); therefore, investigating the genetic diversity of rice genotypes is critical.

The rice cultivar indication process for commercial plantations is dynamic, and periodically new cultivars are recommended as substitutes for those less productive or with less commercial acceptance (Soares et al., 2008). New cultivar development is crucial to help increase food availability, and the success of breeding programs relies on the existence of genetic variability. Breeders have recommended the formation of a base population based on intercrossing the superior and genetically divergent cultivars. This is essential for the success of breeding programs (Cruz, Ferreira, & Pessoni, 2011).

One of the first steps in the formation of a base population is to guarantee genotypic variability through the morphological, physiological and molecular differences of genitors, generally expressed by a dissimilarity measure (Cruz, 2012). Genetic distance estimations are dependent on the data set available, as well as their phenotypic, genotypic, molecular or geographic features (Cruz et al., 2011).

Multivariate techniques such as the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Tocher method, Principal Components Analysis (PCA) and Canonical Variables (CV) have been used as alternatives to simultaneous comparisons of qualitative and quantitative traits, resulting in more precise distance estimates and more accurate genetic diversity predictions among genotypes (Barbosa, Viana, Quintal, & Pereira, 2011; Preisigke et al., 2015). Another promising approach for genetic diversity studies is

the self-organizing maps (SOM) method, which consists of a computational intelligence technique that allows for the visualization of similar patterns and data classification based on the distances between them (Kohonen, 2014).

SOM is a type of two-dimensional artificial neural network that organizes data from an unsupervised learning process and preserves notions of neighborhoods using Euclidean distance. The learning begins with the attribution of synaptic weights, and then a competition process is started in which each data sample is allocated to the neuron that best represents it. This neuron is called the "winner". Then the cooperation begins, in which the winning neuron determines the approximation of the other neurons in the order of proximity. Finally, the neurons that establish their neighborhood go to the adaptation phase, where there are weight adjustments. After all iterations, the map is organized in a topological structure that reflects the proximity of the elements under study.

SOM can present hexagonal topology, where each neuron has at most six direct neighbors, or quadratic topology with at most four direct neighbors. In addition, different arrangements are established that define the number of neurons available on the map. For example, a map with a two by three arrangement presents six neurons arranged in two columns and three rows. This technique is widely used in the various branches of science such as engineering (Akkiraju, Keskinocak, Murthy, & Wu, 2001), industry (Liukkonen, Laakso, & Hiltunen, 2013) and economics (Louis, Seret, & Baesens, 2013; Sarlin, 2013). Although self-organizing maps are widely used, this methodology is still relatively under-explored in plant breeding.

This work aims to test and present the self-organizing maps technique as an alternative method to evaluate the genetic diversity in plant breeding programs.

Material and methods

Twenty-five genotypes were evaluated (Table 1) from the irrigated rice breeding program of the Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), in partnership with Embrapa Arroz e Feijão, of which five were checks (Rio Grande, Ourominas, Seleta, Predileta, and Rubelita). Experiments were carried out in lowland soils under continuous flooding conditions in a randomized block design with three replications in the harvest of 2012/2013, in two municipalities of the state of Minas Gerais: Leopoldina (21º31'12"S, 42º38'43"W) and Lambari (21º58'32"S e 45º21'01"W). The experimental plots consisted of five-meter plant rows with 0.30 m row spacing, in a total plot area of 7.5 m^2 and with a useful area of 3.60 m^2. The plant density was 300 seeds m^{-2}.

Table 1. Rice genotypes origin and codification.

Code	Genotype	Origin	Code	Genotype	Origin
1	BRA 031001	Embrapa Arroz e Feijão	14	SELETA	Embrapa / EPAMIG
2	BRA 041099	Embrapa Arroz e Feijão	15	MGI 0517-25	Embrapa / EPAMIG
3	MGI 0712-1	Embrapa / EPAMIG	16	MGI 0902-8	Embrapa / EPAMIG
4	MGI 0901-5	Embrapa / EPAMIG	17	OUROMINAS	Embrapa Arroz e Feijão
5	MGI 0713-17	Embrapa / EPAMIG	18	CNAi 9091	Embrapa Arroz e Feijão
6	BRA 02691	Embrapa Arroz e Feijão	19	BRA 041230	Embrapa Arroz e Feijão
7	RUBELITA	Embrapa / EPAMIG	20	PREDILETA	Embrapa / EPAMIG
8	MGI 0902-7	Embrapa / EPAMIG	21	MGI 0909-15	Embrapa / EPAMIG
9	MGI 0607-1	Embrapa / EPAMIG	22	MGI 0717-18	Embrapa / EPAMIG
10	BRA 02706	Embrapa Arroz e Feijão	23	BRA 041236	Embrapa Arroz e Feijão
11	BRA 02708	Embrapa Arroz e Feijão	24	BRA 031018	Embrapa Arroz e Feijão
12	BRA 031006	Embrapa Arroz e Feijão	25	RIO GRANDE	Embrapa / EPAMIG
13	BRA 01330	Embrapa Arroz e Feijão			

The agronomic traits evaluated were grain yield (kg/ha), plant height (cm), flowering (days), 100 grain weight (g), grain size (length, width and thickness), length/width ratio, panicle length, number of full grains/panicle, full grains percentage and number of stems/m². All cultural practices were carried out as recommended for the culture (Borém & Nakano, 2015).

Joint analysis of variance was performed for each trait, considering the effects of genotypes to be fixed and environments to be random, according to Equation 1:

$$Y_{ijk} = \mu + B/E_{jk} + G_i + E_j + GE_{ij} + e_{ijk} \qquad (1)$$

For the genetic diversity study, dissimilarity matrices for each environment were obtained based on the average Euclidean distance. The clustering method using the conventional statistical approach was the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). The Mojena (1977) criterion was used to define the optimal number of dendrogram groups, adopting k = 1.25. To control cluster consistency and quality, the cophenetic correlation coefficient (CCC), the distortion between the dissimilarity matrix and the matrix obtained after dendrogram (graphical matrix), and the stress (adjustment precision obtained with the projection of dissimilarity matrix in the dendrogram) were obtained. The CCC was given by the correlation between the elements of the dissimilarity matrix and the elements from the matrix produced by the phenogram (cophenetic matrix) (Silva & Dias, 2013).

The genotypes were also clustered according to the technique of unsupervised learning machine of self-organizing maps. The replication averages for each genotype evaluated for all 11 variables in each assay were used as inputs for this approach. No outputs were stipulated a priori for each genotype, because this is an unsupervised technique. To evaluate the SOM consistency and the best configuration to be used when performing the clustering, eight scenarios were established that varied according to the number, the neurons conformation and the topology in use in the system. For each experiment, eight scenarios were tested in a total of 16 analyses.

Five or six neurons were initially adopted as possible centroids of the genotype groups to be formed. Because this technique is influenced both by the number of neurons adopted a priori, as well as the arrangement of these neurons, this study also evaluated whether there were any arrangements that allowed better visualization of the diversity among the genotypes. These arrangements varied with the number of neurons, and whether that amount was a prime number or not.

Thus, in scenarios for which a prime number of neurons was adopted, such as the arrangements containing five neurons, only one by five or the inverse could be used. In the condition of six neurons, a larger number of possibilities could be used; however, only the arrangements of two by three or the inverse were adopted. The decision was made not to use the other possible arrangements because they resembled the arrangements with five neurons.

Although Kohonen (2014) affirms that the hexagonal topology allows for better visualization of the general data structure, in addition to minimizing the errors, it is not yet known if there is more adequate topology that could be used in genetic diversity studies. Therefore, in this study we tested the grid or hexagonal topologies.

For SOM processing in the different scenarios, the standardized average Euclidean distance was used. For the iterative process, the number of 1000 iterations for each scenario was stipulated. The software Matlab (Matlab, 2010) and GENES (Cruz, 2012) were used to perform the analysis.

Results and discussion

The F test revealed significant effects for genotypes ($p < 0.05$) for the following traits: number of full grains per plant, full grains percentage, grain width and grain length/width ratio (Table 2). The coefficients of variation estimated for most traits were compatible with those obtained in other studies of rice (Silva, Silva, Guimarães, & Moura, 2011; Hosan, Sultana, Iftekharuddaula, Ahmed, & Mia, 2010; Streck et al., 2017), emphasizing the acceptable test quality.

The genotype x environment (G x E) interaction effect was significant for grain yield, panicle length, length, width, and grain thickness, in addition to the grain length / width ratio, indicating that the genotypes exhibit differential behavior in the evaluated environments for these traits. Therefore, the decision was made to study the genetic diversity among the genotypes separately for each environment, because the clustering pattern may vary according to environmental variation.

For the municipality of Leopoldina (Figure 1A), the UPGMA clustering presented CCC values of 0.7306, with a distortion of 2.53% and a stress of 15.90%. For Lambari (Figure 1B), the CCC, distortion and stress values were 0.7025, 4.02% and 20.07%, respectively, showing adequate adjustment values between the dissimilarity matrices and dendrograms. By means of the global criteria introduced by Mojena (1977), it was observed by reference to the last fusion level that at the 80% level of similarity, five different groups were formed for both dendrograms.

Table 2. Summary of joint analysis of variance for grain yield (GY), plant height (PH), panicle length (PL), number of filled grains per plant (NFGP), filled grain percentage (%FGP), number of stems per m² (NSM²), number of panicles per m² (NPM²), grain length (GL), grain width (GW), grain length / width ratio (GLWR), 100-grain weight (100GW) evaluated in 25 rice genotypes cultivated in two environments in the State of Minas Gerais, Brazil.

SV	DF	Mean squares					
		GY	PH	PL	NFGP	%FGP	NSM²
Genotype (G)	24	865773.6^{ns}	243.8^{ns}	12.8^{ns}	541.1^{**}	0.004^{**}	2234.9^{ns}
Environment (E)	1	1648713.8^{ns}	25846.4^{**}	836.6^{**}	1676.9^{ns}	0.0003^{ns}	5118288.3^{**}
G x E	24	1232619.7^{**}	279.7^{ns}	8.7^{**}	105.0^{ns}	0.0007^{ns}	2291.8^{ns}
Mean	--	4074.1	90.7	24.6	93.7	0.9	226.3
CV (%)	--	15.3	20.6	6.3	11.5	4.3	28.6
SV	DF	NPM²	GL	GW	GLWR	100GW	
Genotype (G)	24	26.6^{ns}	0.3^{ns}	0.05^{*}	0.01^{ns}	0.2^{*}	
Environment (E)	1	0.2^{ns}	16.7^{*}	0.2^{ns}	0.4^{*}	1.5^{**}	
G x E	24	29.0^{ns}	0.4^{**}	0.02^{**}	0.01^{**}	0.1^{**}	
Mean	--	37.5	6.9	2.1	1.7	3.3	
CV (%)	--	14.3	2.9	1.7	1.5	3.4	

[ns], [**] and [*]: Not significant, significant according to an F-test at the 0.05 and 0.01 probability level, respectively; SV: sources of variation CV: coefficient of variation; DF: degrees of freedom.
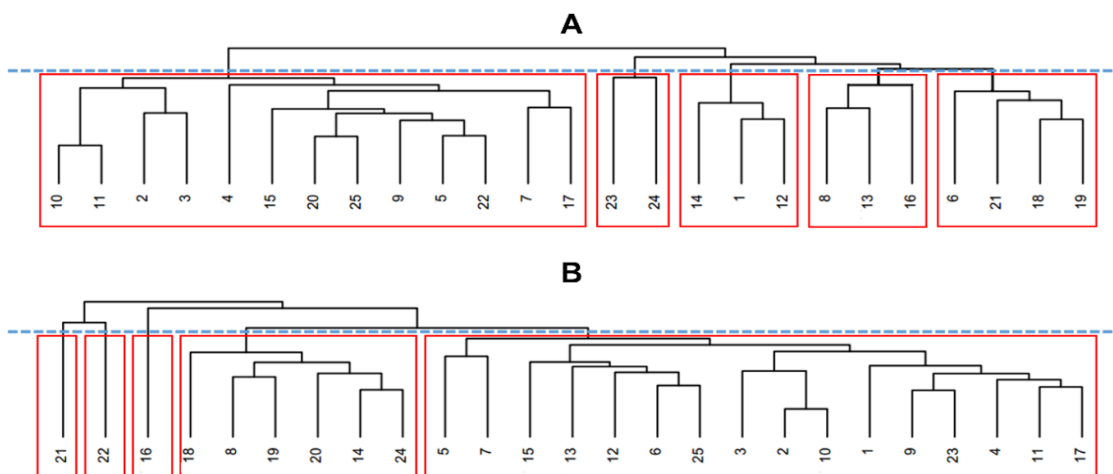
Figure 1. Dendogram obtained by dissimilatity matrix of Euclidian distance of 25 rice genotypes, using Unweighted Pair Group Method with Arithmetic Mean clustering (UPGMA) for the municipalities of Leopoldina (A) and Lambari (B).

Although the number of groups was the same, the clustering method gathered the genotypes differently in each environment, a result expected due to the significant GxE interaction for most evaluated traits. Despite this, the genotypes BRA 041099, MGI 0712-1, MGI 0901-5, MGI 0713-17, RUBELITA, MGI 0607-1, BRA 02706, BRA 02708, MGI 0517-25, OUROMINAS, and RIO GRANDE remained together in one unique group, independent of the environment (Figure 1A and B), reinforcing the idea of greater similarity among them.

In addition to cluster analysis under the conventional biometric approach, a spatial ordering of genotypes was also performed through self-organizing maps, which emphasizes the use of computational intelligence looking for an optimal solution (Kohonen, 1990). Smith and Ng (2003) affirm that it is difficult to quantify the efficiency of clustering made from SOMs, but they were able to generate clearly distinguishable groups.

Considering the experiment conducted in Leopoldina, it was verified that the maps with five neurons (scenarios one, two, three and four) showed genotypic organization with high agreement among each group (Table 3). Considering this configuration, the scenarios one, three and four used the same organization pattern for all genotypes. Only clusters one, two and three, belonging to scenario two, diverged in relation to the other scenarios with quadratic topologies. However, even if this variation existed, there was some agreement between these groups, because in all scenarios the genotype pairs six - 18 and eight - 13, for example, remained in the same group. The genotypes that diverged most in terms of classification in all the maps for this environment were the 16, 19, and 24 genotypes.

In Lambari, map results for the scenarios with five neurons were similar (Table 4). In this case, the genotypes 16, 21, and 22 were those that presented divergence regarding the allocated groups. In general, it has been observed that the maps with the one by five and five by one arrangements present similar classification results, mainly

because they are very simple and because the topology probably will not interfere in these configurations. In scenarios five, six, seven and eight, the map results presented low divergences among themselves, and in scenarios six, seven and eight, the genotypes were allocated to identical groups. Only groups two and three from scenario five differed from the other scenarios, with emphasis on genotypes 21 and 23, which had a distinct allocation pattern in other scenarios with five neurons. This result highlights the possibility of anomalies that occur due to this being an iterative process, and due to the genetic nature of these related and very similar genotypes.

Table 3. Clustering of 25 rice genotypes in the municipality of Leopoldina according self-organizing maps.

| Topology	Arrangement of clusters	Clusters	Genotypes	Topology	Arrangement of clusters	Clusters	Genotypes
Hexagonal	1 x 5	1	2, 3, 4, 5, 7, 9, 10, 11, 15, 17,	Grid	1 x 5	1	8, 13, 16, 19
		2	6, 16, 18, 19, 21			2	23, 24
		3	23			3	1, 12, 14
		4	8, 13, 24			4	6, 18, 21
		5	1, 12, 14			5	2, 3, 4, 5, 7, 9, 10, 11, 15, 17,
	5 x 1	1	2, 3, 4, 5, 7, 9, 10, 11, 15, 17,		5 x 1	1	2, 3, 4, 5, 7, 9, 10, 11, 15, 17,
		2	6, 16, 18, 19, 21			2	8, 13, 24
		3	23			3	6, 16, 18, 19, 21
		4	8, 13, 24			4	23
		5	1, 12, 14			5	1, 12, 14
	2 x 3	1	8, 13		2 x 3	1	2, 3, 10, 11
		2	23			2	4, 5, 7, 9, 15, 17, 20, 22, 25
		3	6, 16, 18, 19, 21			3	6, 8, 13, 16, 18, 19
		4	1, 12, 14			4	24
		5	2, 3, 4, 5, 7, 9, 10, 11, 15, 17,			5	23
		6	24			6	1, 12, 14, 21
	3 x 2	1	2, 3, 10, 11		3 x 2	1	6, 16, 18, 19, 21
		2	6, 16, 18, 19, 21			2	8, 13, 24
		3	8, 13, 24			3	23
		4	4, 5, 7, 9, 15, 17, 20, 22, 25			4	2, 3, 10, 11
		5	23			5	4, 5, 7, 9, 15, 17, 20, 22, 25
		6	1, 12, 14			6	1, 12, 14

Table 4. Clustering of 25 rice genotypes in the municipality of Lambari according self-organizing maps.

Topology	Arrangement of clusters	Clusters	Genotypes	Topology	Arrangement of clusters	Clusters	Genotypes
Hexagonal	1 x 5	1	16	Grid	1 x 5	1	16
		2	2, 3, 4, 5, 7, 10, 11, 17			2	2, 3, 4, 5, 7, 10, 11, 17, 23
		3	6, 12, 13, 15, 25			3	6, 12, 13, 15, 25
		4	1, 8, 9, 14, 18, 19, 20, 21, 23,			4	21
		5	22			5	1, 8, 9, 14, 18, 19, 20, 22,
	5 x 1	1	1, 8, 9, 14, 18, 19, 20, 24		5 x 1	1	1, 8, 9, 14, 18, 19, 20, 24
		2	22			2	22
		3	21			3	21
		4	2, 3, 4, 5, 7, 10, 11, 16, 17,			4	2, 3, 4, 5, 7, 10, 11, 16, 17,
		5	6, 12, 13, 15, 25			5	6, 12, 13, 15, 25
	2 x 3	1	16		2 x 3	1	21
		2	2, 3, 4, 5, 7, 10, 11, 17			2	1, 8, 9, 14, 18, 19, 20, 24
		3	1, 9, 18, 21, 23			3	2, 3, 4, 5, 7, 10, 11, 17, 23
		4	6, 12, 13, 15, 25			4	22
		5	22			5	16
		6	8, 14, 19, 20, 24			6	6, 12, 13, 15, 25
	3 x 2	1	22		3 x 2	1	1, 8, 9, 14, 18, 19, 20, 24
		2	1, 8, 9, 14, 18, 19, 20, 24			2	16
		3	16			3	2, 3, 4, 5, 7, 10, 11, 17, 23
		4	21			4	22
		5	2, 3, 4, 5, 7, 10, 11, 17, 23			5	21
		6	6, 12, 13, 15, 25			6	6, 12, 13, 15, 25

Although different clustering techniques provide different diversity views, an agreement is expected to exist among them. Therefore, the results obtained by the self-organizing maps were compared to those obtained by the conventional statistical approach, with a main goal to observe the clustering behavior associated with these techniques and evaluate their complementarities.

The genotypes were identified in each map according to the UPGMA clustering. Genotype ordering according the SOM technique was consistent with the hierarchical clustering results, because the basic structure of the

UPGMA groups was preserved in each group of the maps (Figures 2 and 3). Considering genotype arrangements and the group neighbors, maps involving five neurons presented inferior organization efficiency to the six-map arrangements in both environments, as can be observed in Figure 2 where the genotypes eight, 13 and 16 were separated into groups without neighborhoods in scenarios with five neurons, but remained in neighboring groups in all scenarios with six neurons.
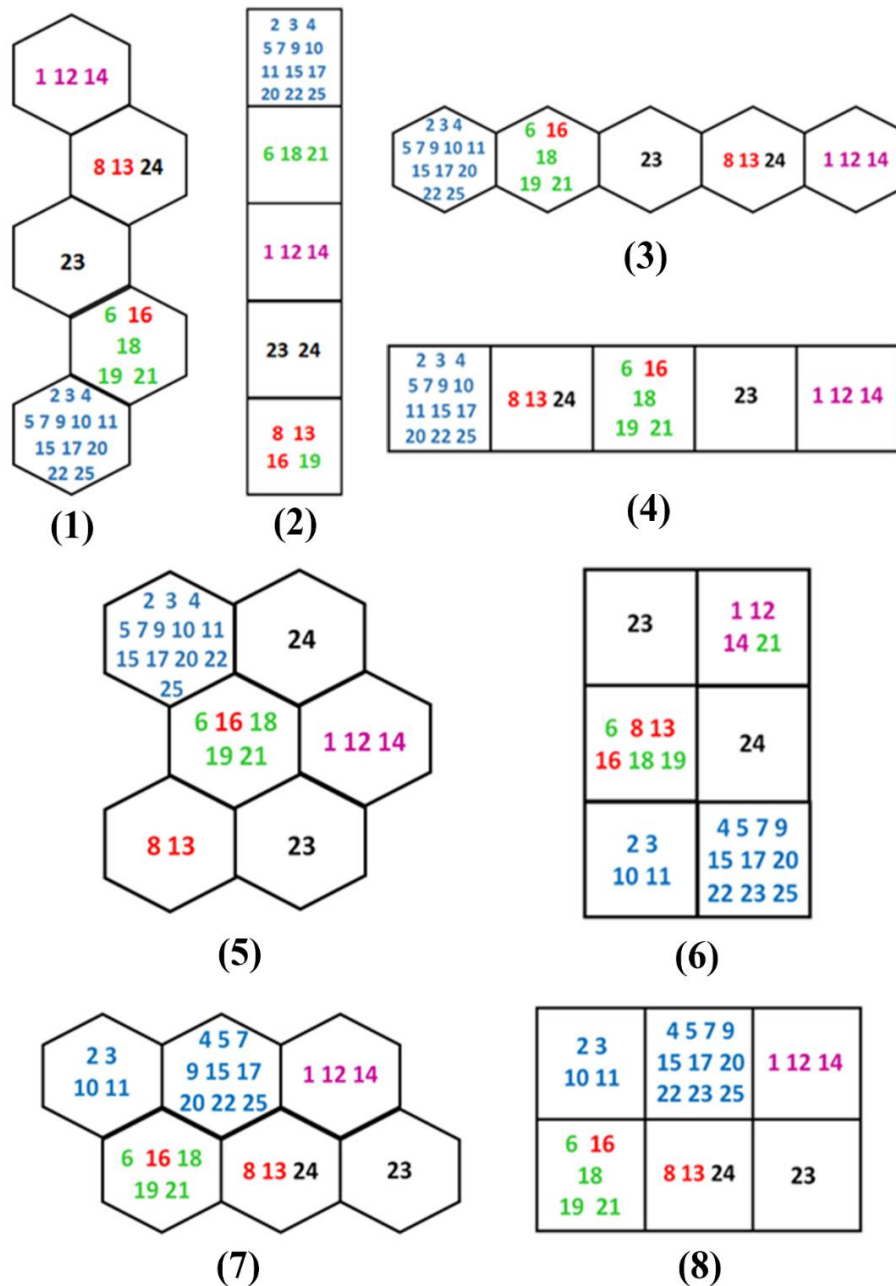
Figure 2. Arrangement of the clustering scenarios according to the SOM for Leopoldina. Genotypes belonging to the same group according to the UPGMA method were identified on maps using equal colors (i.e., numbers in blue, green, red, black, and purple represent a specific UPGMA cluster). Numbers in parentheses refers to their respective scenarios.

In scenarios with five neurons, each group has at most two direct neighbors, while in the six-neuron arrays the neighborhood is determined for up three groups; therefore, the technique is able to capture the proximity among groups and to organize the genotypes more efficiently in each one according to the grouping carried out. In addition, Figure 3 shows that genotypes one and nine of scenarios two, three, four, six and eight were distinctly allocated to the groups consisting of their peers according to UPGMA clustering. According to Kohonen (2014), hexagonal topology allows for better neuron arrangements; moreover, the simpler configurations like those used with five neurons may distort genotype organization, especially in cases where the material

studied is not easily distinguishable.
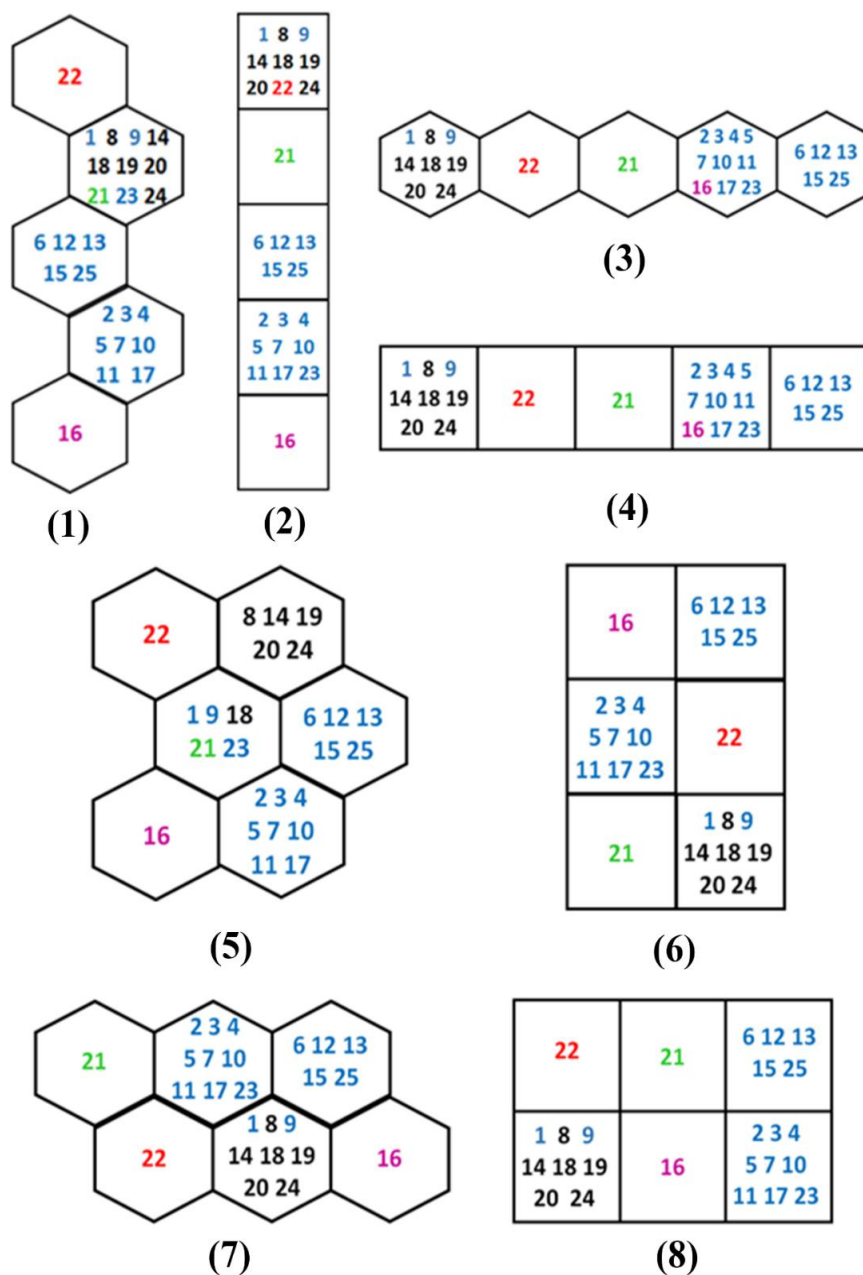
Figure 3. Arrangement of the clustering scenarios according to the SOM for Lambari. Genotypes belonging to the same group according to the UPGMA method were identified on maps using equal colors (i.e., numbers in blue, green, red, black, and purple represent a specific UPGMA cluster). Numbers in parentheses refers to their respective scenarios.

Having a fixed number of groups that depends on the number of neurons that are predefined when determining the SOM configuration can lead to some abnormalities, such as the separation of some large groups into two or more smaller subgroups and group union. In addition, it should be noted that the rice genotypes evaluated in these assays are in the final stage of breeding and are genetically close, which may lead to some difficulty in allocating these genotypes. However, in general, it is observed that the organization patterns among the rice genotypes evaluated by the maps is complementary to the UPGMA approach, as observed in all scenarios.

When evaluating the map arrangement in different scenarios, it was noted in Leopoldina that a three by two arrangement with hexagonal topology preserved the organization of UPGMA groups, a fact that was not observed in scenario five (two by three arrangement), where the genotypes 23 and 24 were allocated to distant groups (Figure 2). In Lambari, the three by two and two by three arrangements with the same topology were superior to the others because they permitted more connections among groups. These topologies were able to organize the genotypes into groups closer to those obtained by UPGMA clustering (Figure 3). In particular, genotypes of the

same group in the UPGMA remained in the same group or in linked groups in the SOM. However, in cases with higher numbers of neurons, there is a possibility of larger variation amounts in the allocation of genotypes, as affirmed by Kohonen (2014).

The SOM method has been shown to be an efficient way of identifying patterns of similarity, as shown by Mwasiagi (2011), who used the SOM technique to distinguish cotton genotypes. This author concluded that the method was efficient for separating thin wires from coarser ones, and the samples that were dispersed on the map would be outliers, implying irregularity of the material. Smith et al. (2003), studying the SOM efficiency for organizing web pages through navigation patterns, obtained a satisfactory result and concluded that the method can be easily incorporated; however, this needs to be developed for large scale applications. A similar conclusion was found by Fritzke (1994), who studied the map efficiency for supervised and unsupervised learning.

In addition to the strong agreement obtained by the SOM, this technique presented high complementarity to the stochastic approach. Thus, it is observed that the organization of genetic diversity through self-organizing maps is efficient, and the SOM technique has the potential to be useful for genetic diversity studies in breeding programs.

Conclusion

Self-organizing maps have the potential to be useful for genetic diversity studies in breeding programs.

References

Akkiraju, R., Keskinocak, P., Murthy, S., & Wu, F. (2001). An agent-based approach for scheduling multiple machines. *Applied Intelligence*, *14*(2), 135-144. DOI: 10.1023/A:1008363208898

Barbosa, C. D., Viana, A. P., Quintal, S. S. R., & Pereira, M. G. (2011). Artificial neural network analysis of genetic diversity in Carica papaya L. *Crop Breeding and Applied Biotechnology*, *11*(3), 224-231. DOI: 10.1590/S1984-70332011000300004

Borém, A., & Nakano, P. H. (2015). *Arroz: do plantio à colheita.* Viçosa, MG: Editora UFV.

Cruz, C. D., Ferreira, M. F., & Pessoni, L. A. (2011). *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco, MG: Editora Suprema.

Cruz, C. D. (2012). *Princípios de genética quantitativa*. Viçosa, MG: Editora UFV.

Fritzke, B. (1994). Growing cell structures – A self-organizing network for unsupervised and supervised learning. *Neural Networks*, *7*(9), 1441-1460. DOI: 10.1016/0893-6080(94)90091-4

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, *327*(5967), 812-818. DOI: 10.1126/science.1185383

Hosan, S. M., Sultana, N., Iftekharuddaula, K. M., Ahmed, M. N. U., & Mia, S. (2010). Genetic divergence in Landraces of Bangladesh rice (*Oryza sativa* L.). *The Agriculturists*, *8*(2), 28-34. DOI: 10.3329/agric.v8i2.7574

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464-1480. DOI: 10.1109/5.58325

Kohonen, T. (2014). *MATLAB implementations and applications of the self-organizing map*. Helsinki, FI: Unigrafia Oy.

Liukkonen, M., Laakso, I., & Hiltunen, Y. (2013). Advanced monitoring platform for industrial wastewater treatment: Multivariable approach using the self-organizing map. *Journal Environmental Modelling & Software*, *48*(1), 193-201. DOI: 10.1016/j.envsoft.2013.07.005

Louis, P., Seret, A., & Baesens, B. (2013). Financial efficiency and social impact of microfinance institutions using self-organizing maps. *World Development*, *46*(1), 197-210. DOI: 10.1016/j.worlddev.2013.02.006

Matlab. (2010). *Version 7.10.0* [Software]. Natick, MA: The Math Works Inc.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, *20*(4), 359-363. DOI: 10.1093/comjnl/20.4.359

Mwasiagi, J. I. (2011). Use of SOM to study cotton growing and spinning. In J. I. Mwasiagi (Ed.), *Self organizing maps – Applications and novel algorithm design* (p. 89-94). London, UK: IntechOpen. DOI: 10.5772/14106.

Preisigke, S. C., Neves, L. G., Araújo, K. L., Barbosa, N. R., Serafim, M. E., & Krause, W. (2015). Multivariate analysis for the detection of passiflora species resistant to collar rot. *Bioscience Journal*, *31*(6), 1700-1707. DOI: 10.14393/BJ-v31n6a2015-29300

Rabelo, H. O., Guimarães, J. F. R., Pinheiro, J. B., & Silva, E. F. (2015). Genetic base of Brazilian irrigated rice cultivars. *Crop Breeding and Applied Biotechnology*, *15*(3), 146-153. DOI: 10.1590/1984-70332015v15n3a26

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, *8*(6), e66428. DOI: 10.1371/journal.pone.0066428

Sarlin, P. (2013). Decomposin the global financial crisis: A Self-Organizing Time Map. *Patter Recognition Letters*, *34*(14), 1701-1709. DOI: 10.1016/j.patrec.2013.03.017

Silva, E. F., Silva, V. A. C., Guimarães, J. F. R., & Moura, R. R. (2011). Divergência fenotípica entre genótipos de arroz de terras altas. *Revista Brasileira de Ciências Agrárias*, *6*(2), 280-286. DOI: 10.5039/agraria.v6i2a1183

Silva, A. R., & Dias, C. T. S. (2013). A cophenetic correlation coeficiente for Tocher's method. *Pesquisa Agropecuária Brasileira*, *48*(6), 589-596. DOI: 10.1590/S0100-204X2013000600003

Smith, K. A., & Ng, A. (2003). Web page clustering using a self-organizing map of user navigation patterns. *Decision Support Systems*, *35*(2), 245-256. DOI: 10.1016/S0167-9236(02)00109-4

Soares, P. C., Soares, A. A., Cornélio, V. M. O., Reis, M. S., Morais, O. P., Cutrim, V. A., ... Silva, F. L. (2008). BRSMG Predileta: irrigated rice cultivar for lowlands in Minas Gerais. *Crop Breeding and Applied Biotechnology*, *8*(1): 251-254. DOI: 10.1590/1984-70332017v17n2c27

Streck, E. A., Aguiar, G. A., Magalhães Júnior, A. M., Facchinello, H. K., & Oliveira, A. C. (2017). Variabilidade fenotípica de genótipos de arroz irrigado via análise multivariada. *Revista Ciência Agronômica*, *48*(1), 101-109. DOI: 10.5935/1806-6690.20170011

Tilman, D., Balzer, C., Hill, J., & Befort, B. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260-20264. DOI: 10.1073/pnas.1116437108