ARTIGO

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais

V. 1 - N. 1 - OUT./ DEZ. 2022
https://revista.uemg.br/index.php/ambciencias

SELF-ORGANIZING MAPS AS TOOLS FOR UNDERSTANDING THE GENETIC VARIABILITY OF POPULATIONS

Marciane da Silva Oliveira, State University of Minas Gerais (UEMG), Carangola- MG, Brazil, e-mail: marciane.oliveira@uemg.br

Iara Gonçalves dos Santos, Federal University of Viçosa, Viçosa, Brazil

Cosme Damião Cruz, Federal University of Viçosa, Viçosa, Brazil

ABSTRACT

The maintenance of genetic diversity is fundamental to ensuring the population's viability in the mid- and long-term. Because evolutionary factors (e.g., genetic drift, selection, migration), as well as inbreeding can change the genetic structure of a population, it is important to understand how these factors act on populations and to infer about the maintenance of genetic diversity throughout generations. The Self-Organizing Maps (SOM) is an interesting approach to organizing genetic diversity and highlighting the effects caused by dispersive and systematic factors in populations. Briefly, the SOM algorithm maps the data, weighting similarities among inputs while keeping similar inputs close to each other in a topological map. The upside of this approach is that it organizes populations following biological principles. SOM have shown efficiency in organizing natural or breeding populations that are subject to processes that reduce variabilities, such as drift, inbreeding, and selection, and to processes that increase genetic variability, such as migration.

Keywords: genetic drift. selection. migration. inbreeding.

Conflitos de interesse: Os autores declaram não haver conflito de interesse.

Financiamento: CHAMADA 01/2021 PQ/UEMG, FAPEMIG.

Contribuição dos autores: MSO and IGS designed the project and performed the review. The work was curated by MSO. All authors wrote, read and approved the final manuscript.

Recebido: 23/05/2022 **Aprovado:** 19/07/2022

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

One of the most relevant requirements in genetic studies is the existence of phenotypic and genetic variability. The quantification of variability guides decisions in terms of selection from a breeding perspective, conservation, and management of natural resources. Once diversity is detected, researchers can manage to conserve the allelic variability in natural populations or cluster similar units and keep less similar individuals apart for breeding purposes (Cruz et al., 2020).

The choice of the right individuals to compose base populations determines the success of a breeding program (Oliveira et al., 2020a). If genetic diversity is properly quantified, the base populations will carry favorable alleles coming from different complementary individuals. The different approaches used to identify and quantify diversity are essential to make the decision process in a breeding program easier.

An additional question can be asked regarding the organization of genetic diversity. Given a set of accessions, one could recognize the existence of several clusters, for example, A, B, C, and D, and use this information in a breeding program. However, units within groups must have an organization that traditional approaches cannot access. Moreover, clusters can also present a very relevant pattern of organization or neighborhood. Thinking about different possibilities other than a linear organization (that can make an organization like ABCD, ABDC, and ADBC), non-linear patterns can also occur (Example: AB / CD, AC / BD, etc).

The ability to recognize clusters could be considerably expanded if, instead of using mathematical methods, the genetic diversity was mapped based on an artificial intelligence approach, unraveling the best organization of the individuals within clusters. The visualization of neighborhoods would allow ratifying or refuting hypotheses about gene flow, migration, origin, etc. The traditional approaches and basic genetic diversity studies are still fundamental, but adding information about the organization of individuals within clusters enriches the analysis and interpretation of the biological phenomena.

Self-organizing maps (SOM) are a class of neural networks that detect and organize similarities between input patterns. In other words, they constitute computational tools to solve pattern recognition, clustering, and data organization (Nascimento et al., 2018). The knowledge about the particularities and applicability of SOM techniques is desirable because it is a tool that generates information that adds to that information established in traditional approaches to studying genetic diversity. Furthermore, the extra knowledge SOM brings will help guide the decision-making process and optimize resources that guarantee a great use of genetic variability.

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Genetic structure of populations

Studies on population genetics are supported by different measures used to analyze data that provide a deep understanding of the natural and artificial genetic structure and dynamics of the population. Population genetics considers the laws of heredity, such as Mendel's laws and other principles of genetics, to understand the genetic structure of populations focused on allele and genotypic information. The genetic analyses of populations consider all individuals as part of the population. These studies provide basic information about allelic and genotypic frequencies and shed light on future generations by assuming that the probability of crosses between individuals corresponds to their genotypes in the population. Thus, the primary genetic descriptors of a population at a specific t time are estimated under certain environmental conditions. The analysis of these descriptors allows us to understand the genetic structure of populations under current conditions and its changes through environmental factors, mating systems, selection practices, and adaptive factors, among others.

In addition to the genetic structure, population genetic studies seek to understand the dynamics within and between populations and to discover and understand processes that can affect the frequency of alleles and genotypes such as migration, mutation, selection, and genetic drift. The genetic structure of a population can be studied from the measurement of its basic descriptors, which are the allelic frequencies and genotypic frequencies as well as their behavior over generations. These basic genetic descriptors provide a quantification of the genetic variation in a population.

Considering a hypothetical gene (A) with two allelic forms (A/a), the allelic frequencies are estimated by p (frequency of the reference allele) and q (frequency of the alternative allele), and the observed genotypic frequencies by D (frequency of the homozygote genotypes carrying two doses of the reference allele), H (frequency of heterozygotes), and R (frequency of homozygote genotypes carrying two doses of the alternative allele). Table 1 summarizes a genotyping analysis made in an agarose gel containing fragments generated by a codominant marker. Considering that the marker represents a gene A (locus A) with two allelic forms (A/a), the allelic frequencies are estimated by p and q, and the observed genotypic frequencies by D, H, and R, as follows.

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Table 1 - Genotypic and allelic frequencies of the studied population based on information from codominant molecular markers (adapted from Oliveira & Cruz, 2021)

Code	Correspondent genotype	Number of observed individuals	Genotypic frequencies	Genotypic frequencies
11	AA	10	D=N11/N	D=10/30=0,33
12	Aa	13	H=N12/N	H=13/30=0,44
22	Aa	7	R=N22/N	R=7/30=023
		30		1

The allelic frequencies are estimated as follows:

$$f_A=p=D+\frac{1}{2}(H)=0.33+\frac{1}{2}(0.44)=0.55$$

$$f_a=q=R+\frac{1}{2}(H)=0.23+\frac{1}{2}(0.44)=0.45$$

and

$$p+q=0.55+0.45=1$$

Whether from closely related populations or species, phylogenetic trees can be constructed from genetic distances quantified from allelic or genotypic frequencies obtained from different molecular markers (loci). Genetic and environmental factors can cause genetic distance or increase the similarity between populations. The degree of genetic diversity can be accessed by measures of heterozygosity, degree of fixation, or intergenotypic correlation.

Genetic differences between populations are generally slight when calculated from allelic frequency. However, suppose a single locus is analyzed. It is noteworthy that two populations would be genetically more distant when one population has allelic frequencies of p = 1 and q = 0 and the other population has p = 0 and q = 1, this situation represents what we call gene fixation. In other words, the first population has the A allele fixed while the second population has the alternative allele (a) fixed.

The basis of population genetics is the Hardy-Weinberg Equilibrium (HWE), which considers that equilibrium is reached in a sufficiently large population and the absence of mutation, selection, migration, and genetic drift after one generation of random mating. Under HWE, the genotypic frequencies are equivalent to the square of the sum of allelic frequencies. The successive generations of random mating do not change the allelic and genotypic frequencies under HWE (Oliveira & Cruz, 2021).

Although natural populations under genetic improvement do not meet these

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

assumptions, HWE is just a hypothetical condition. However, this predictive knowledge allows researchers to observe which deviations are occurring and propose which possible evolutionary factors are involved, thus recognizing the evolutionary dynamics of the species in certain regions. These deviations can be caused by the Wahlund effect, gene flow, mutations, non-random mating, selection, or genetic drift. In addition, quantifying the changes in allelic and genotypic frequencies through different generations in a population can improve the understanding of which evolutionary factors are present. For example, the occurrence of preferential mating only alters the genotypic frequencies relative to the genes specifically involved. In contrast, genetic drift is particularly effective and more quickly diagnosed in small populations, and its alterations cannot be predicted.

Deviations caused by selection, gene flow, and mutations of any kind often need significantly high values to be detected, which makes the Hardy-Weinberg proportion deviation test weak for evaluating changes in a population. An alternative is using SOM to understand the variations that evolutionary factors cause in populations. Oliveira et al. (2021) verified that SOM could organize populations under selection, genetic drift, migration, or divergent selection in a topological grid. The organization can be explained by the particularities of each evolutionary factor. Furthermore, it was verified in this study that the SOM was efficient in detecting these alterations using 100 SNPs.

However, when a higher number of genetic markers are evaluated, the linkage disequilibrium (LD) must be considered (Flint-Garcia et al., 2003). LD is the non-random allelic association of different loci in the gametes. LD is important to elucidate the genetic and evolutionary phenomena that occur over generations in populations or species (Santos et al., 2020). Genetic markers in LD with traits of interest are useful in marker-assisted selection and also in genetic models since they can increase the accuracy of the genetic findings.

According to Cruz et al. (2020), the presence of LD may indicate that despite the population being panmictic, a stratification may persist. A preferential gene pool remains, derived from ancestors, which subdivides the population, which will gradually homogenize. This homogenization will depend upon the recombination rate between the two loci, which may be extremely slow, especially between closely linked loci, but not only in this condition. Linkage equilibrium, when verified, will indicate that the population has already gone through successive cycles of random mating and is free from evolutionary forces.

If LD is detected, despite random matings being known, it is hypothesized that some evolutionary factors must have acted and constitute disturbing elements concerning the Hardy-Weinberg model. Even if selective processes cover only one generation, they can also generate new gametic disequilibrium, which can persist for many generations between closely linked loci. Therefore, the great interest in evaluating LD relies on the fact that it

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

indicates events of allele introduction and changes in their frequencies in the population's past (Cruz et al., 2020). Several methodologies for estimating LD have been extensively described. More explanations and examples of LD can be found in reference (Santos et al., 2020; Cruz et al., 2020).

How genetic diversity studies are carried out using conventional approaches?

Genetic diversity is defined as any measure that quantifies the magnitude of genetic variability within a population and is considered an essential tool for evaluating the conservation and management of species. However, to understand genetic diversity, it is necessary to access it. Genetic markers are used to assess genetic diversity at an allelic level.

There are dominant and co-dominant molecular markers, and the access to the genotypic information is specific according to the type of markers. Dominant markers do not show sensitivity to differentiate the heterozygous from the dominant homozygous genotype. The two genotypes are then considered a single class, whereas the "recessive" homozygous genotype (aa) is identified by the absence of the band in the gel (null phenotype). Thus, with dominant markers, it is possible to study the diversity within the population or between individuals. The populations evaluated, in most cases, are not hierarchically structured, and the samples are random representations of the population.

The analysis of the genetic diversity between populations is performed using estimates of genetic dissimilarities obtained from multivariate techniques by different similarity indices. After that, techniques for grouping or making graphical projections of dissimilarity measures can be adopted. Although Cruz et al. (2020) bring a series of similarity coefficients described in the literature, these authors also recommend that dissimilarity indices be used for cluster analysis. They also indicate the most recommended formulas to convert molecular or phenotypic information into similarity or dissimilarity. The choice of which similarity index should be used is up to the researcher since each has specific characteristics.

On the other hand, codominant markers can differentiate the dominant homozygote from the heterozygote genotype, providing extra genetic information. In this case, dominant homozygote genotypes are coded with 11, heterozygotes with 12, and the recessive homozygote with 22. If the marker is multiallelic, numbers corresponding to the number of each allele are added. Once the genotypes of individuals in the population are known, the allelic frequencies can be calculated. However, if dominant markers are used, it is only possible to calculate the allelic frequencies if the sampled population is in Hardy-Weinberg equilibrium.

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Once the codominant markers are coded with 11, 12, or 22, it is possible to obtain the dissimilarity matrix and apply specific techniques for clustering genotypes (Cruz et al., 2020). Thus, one can study the genetic diversity within the population. However, the information obtained by these markers can also be used to estimate the genetic distance between populations. Cruz et al. (2020) present several approaches that estimate genetic variation and determine the level of population structure. Those approaches have several applications at the individual, intrapopulation, and interpopulation levels.

From molecular data, it is possible to analyze the genetic descriptors within and between populations through traditional techniques in population genetics, such as the H statistics of Nei (Perez, 2020) and Hedrick's genotypic measure (Cruz et al., 2020). Nei's statistics are based on Heterozygosity (H) and can be applied at different hierarchical levels, using the GST metric to measure diversity (Cruz et al., 2020). Another important diversity measure is the measure of Hedrick. This measure only considers the diversity prediction between populations and genotypic frequencies. Hedrick's measure is different from Nei's statistics because the latter uses genotypic and allelic frequencies (Cruz et al., 2020). Regardless of the method, the determination of the genetic distance between and within a population is fundamental to guiding breeding decisions and understanding the behavior of natural populations.

What is and why use self-organizing maps?

Options other than traditional approaches are becoming common to predict and organize the genetic variability within and between populations, such as Artificial Neural Networks (ANNs). ANNs are models inspired by the neural network of living beings, that can be interpreted as a non-parametric prototype of the human brain, with the ability to recognize patterns and generate learning from input vectors (Moura et al., 2015). They resemble the functioning of biological neurons through the implementation of algorithms and mathematical functions (Kitani, 2013). The ANN approach has the advantage of not requiring many parameters and supports the loss of some data. Several types of ANNs can be used in germplasm bank management, such as the self-organizing map algorithm (SOM) (Cruz et al., 2020).

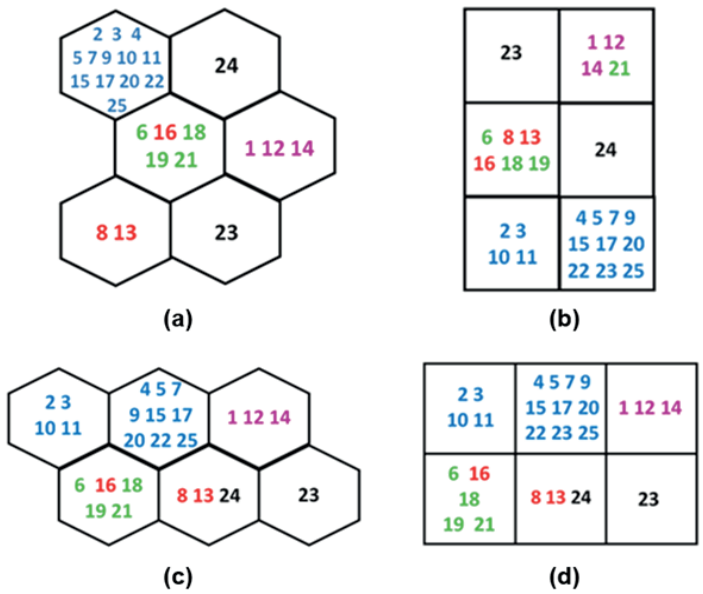
Self-organizing maps (SOM) were first introduced by Teuvo Kohonen around 1981. The technique is related to vector quantization, used extensively in digital signal processing and transmission (Kohonen, 2016). It is a non-linear way to map inputs through the construction of models, in which similar models get associated with nodes that are closer to the grid, whereas less similar models will be farther away from the grid. SOM networks can also be thought of as an approach to reducing dimension, starting from a multi-dimensional space to a lower-dimensional (usually two-dimensional) plot while preserving the

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

original topology of the data (Spanoghe et al., 2020).

SOM detect similarities between input patterns through a process of competition and preserves notions of distance by the adoption of an activation function, such as the Euclidian distance (Nascimento et al., 2018). The learning process begins with the attribution of synaptic weights to the different neurons, then a competitive process begins in which the winning neuron is determined. Next, the cooperation process begins with the winning neuron determining the approximation of other neurons. Once the neighborhood is established, the adaptation phase occurs, where the synaptic weights are adjusted. After all iterations, the map is organized in a topological structure that reflects the proximity between the accessions under study. SOM can present a hexagonal topology, where each neuron has at most six direct neighbors, or a quadratic topology with at most four direct neighbors (Figure 1). Different arrangements can also be established to 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.

Figure 1 - Self-organizing maps with hexagonal (a and c) or quadratic (b and d) topology. Genotypes (represented by number within each hexagon or square) belonging to the same group according to the UPGMA method were identified using equal colors on maps. Adapted (Santos et al., 2019).



In other words, SOM search for an organizational pattern in the dataset with optimal efficiency. Since it is an ANN unsupervised approach, SOM classifies the data with a minimum of distortion, meaning that this technique is a powerful tool for information extraction. It can be used to capture and organize genetic patterns of natural or breeding populations. The most exciting thing about using SOM rather than stochastic models to transform raw genetic data into diverse distance measures is to avoid the loss of information regar-

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

ding the real dataset. The main advantages (Spanoghe et al., 2020) of using SOM are:

i. to avoid an arbitrary choice of a distance measure that can be questionable sometimes;

ii. stochastic approaches emphasize means and covariances while there is no guarantee that the directions of maximum variance mean good discrimination;

iii. SOM are not affected by the presence of outliers, missing data, and disjointed data matrix (that may generate wrong clusters and lead to misinterpretations).

SOM also make easier the visualization of patterns established through different generations in populations under some genetic diversity effects. They can organize subpopulations by the level of similarity between them on a map to facilitate observation (Ibrahim et al., 2016). This neural algorithm follows a different methodology from the traditional groupings applied to population genetics (Nascimento et al., 2018). Therefore, the best way to describe the action of SOMs is to simulate populations under Hardy-Weinberg equilibrium in which the genetic parameters are known. When SOM is applied to the divergent selection, it allows predicting the direction and magnitude of the frequency changes, making comparisons easier to be interpreted in successive generations.

Practical application of SOM

SOM have been explored under different contexts, however, applications in biology and breeding are still limited. Some of the main applications of such an approach are next presented. SOM were used to map genetic drift, selection, migration, and inbreeding effects over generations from the allelic and genotypic frequencies in a simulation study (Oliveira et al., 2021). The SOM showed to be efficient in organizing the populations since it was possible to recognize the expected pattern of each tested genetic effect. The authors highlighted SOM are efficient in organizing populations that are subject to a process that reduces variability, such as drift, inbreeding, and selection, and to processes that increase genetic variability, such as migration.

However, the authors used a constant index between generations under divergent selection (± 0.25). A divergent selection of 0.25 is considered a high value, as mentioned by the authors. High values can be found in populations under artificial selection. Thus, a critical perspective opens since the SOM algorithm can still be used to map selection effects over generations, considering lower adaptation rate values, better representing the selection effects in natural populations (Mathieson & Mcvean, 2013; Kim et al., 2017).

Santos et al. (2020) used phenotypic and molecular markers as inputs to the SOM analysis to assess the genetic diversity of alfalfa populations. The bi-dimensional maps revealed the presence of genetic diversity. The authors selected alfalfa accessions based on the SOM to compose a base population to be used at the beginning of the alfalfa breeding program in the tropics.

Spanoghe et al. (2020) applied the SOM method to explore the population structure and

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

genetic relationships among a panel of potato lines, belonging to both different market segments and spatio-temporal groups, using 21 polymorphic microsatellite markers. The authors showed that the SOM was suitable for classifying varieties into the main detected groups and visualizing inter-group genetic dissimilarities. The SOM also provided additional information on intra-group diversity.

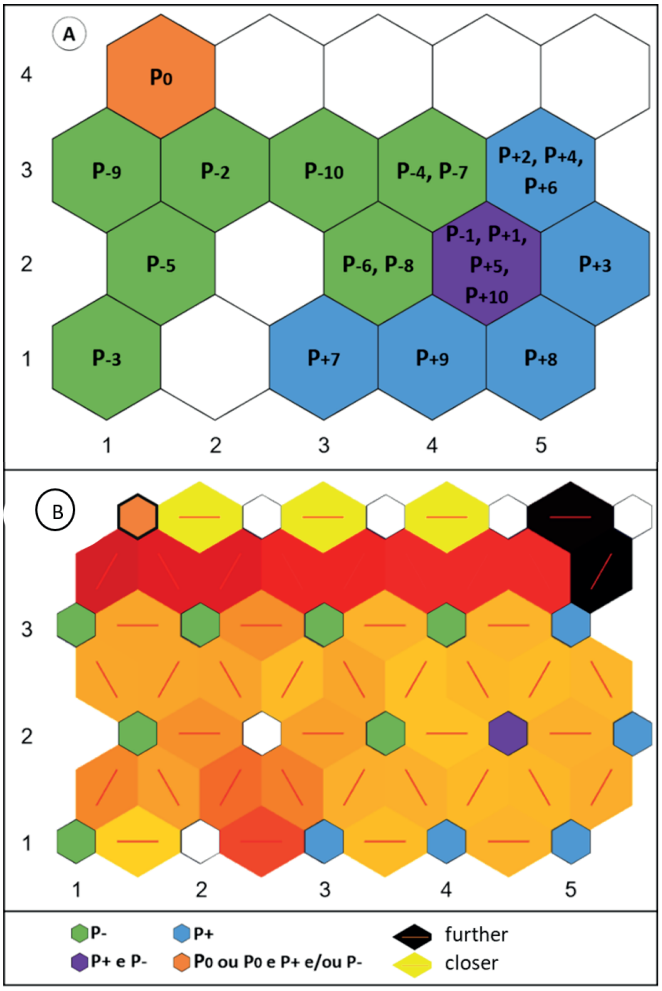
SOM were also used as an alternative method to evaluate genetic diversity in rice breeding programs. Twenty-five rice genotypes were evaluated in two environments for 11 phenotypic traits. The genotypes were organized according to SOM and to the Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering. Genotype ordering according to SOM was consistent with UPGMA results, since the basic structure of UPGMA groups was preserved in each group of the maps. The authors observed that the organization pattern among the rice genotypes evaluated by the maps was complementary to the UPGMA approach.

To verify the genetic diversity in populations simulated under divergent selection, Fonseca et al. (2020), Oliveira et al. (2020) and Fonseca (2021) demonstrated the efficiency of SOM. The authors were able to detect patterns of diversity for both higher selection index (W1 = ±0.02) and indices of W2= ±0.02 and W3= ±0.002 (Figure 1). Fonseca (2021) also performed the Tocher clustering to compare with clusters from Nei's GST statistics and Hedrick's genotypic measure and stated that all the techniques clustered subpopulations with greater genetic diversity, although there was a difficulty in clustering less diverse subpopulations.

Another important statement highlighted by Fonseca (2021) is that traditional techniques form groups with the characteristic of the distance between clusters being higher than the distance of populations within clusters. Although clustering is carried out with traditional approaches, there are limitations to representing the closeness between clusters and the way subpopulations are organized. These limitations do not mean that traditional techniques are not worthy. Traditional techniques are suitable for clustering rather than mapping. SOM, on the other hand, forms a topological segmentation of subpopulations, and this organization allowed for a more visual relationship to be traced about the patterns of genetic diversity (Figure 2).

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Figure 2 - A. Self-organizing map for the selection index W2= ± 0,02 using a 4 x 5 arrangement and a hexagonal topology (20 neurons). B. Heatmap of the 20 neurons to the selection index W2= ± 0,2. Base-population (P0); Subpopulations under a positive effect of selection W1= (P+1, P+2...P+10); Subpopulations under the negative effect of selection W1= (P-1, P-2...P-10) (Fonseca, 2021).



Vidon et al. (2021) performed a study that aimed to understand how conventional statistical techniques and SOM perceive the effects of migration among populations. The simulations of the base population, donor, and recipient populations, considered the Hardy-Weinberg equilibrium. Each population had 1000 individuals and was genotyped with 100 codominant biallelic markers with allelic frequencies varying from 0 to 1 ($0 \leq p \leq 1$). Simulation of the populations was performed in Portal Genes, as well as the process of migration. Briefly, migration consisted of m migrants coming from the donor population for g generations forming subpopulations. Subpopulations were then analyzed using Nei, Hedrick, and Tocher clustering techniques. SOMs were used to perform the organization of the same subpopulations. The mapping used the basic descriptors of populations (q, D, and H, which are the allele frequencies of the A2 allele, from the homozygous genotype to the A1 allele and the heterozygotes) as inputs. By analyzing different values of m (50, 100, and 300) over ten generations, the authors found out that SOM is capable of percei-

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

ving genetic variations.

Nei and Hedrick's measures, when grouped by Tocher, were not able to perceive the variability between generations for m = 50 and m = 100, forming a single cluster with all subpopulations. For m=300, the Tocher clustered the subpopulation in four groups, placing the subpopulations of generations 4 to 10 in group one, generations 1 to 3 in group two, the recipient population in group three, and the donor population in group four. SOM organized populations 1 and 2 (donor and recipient) in distant neurons, and subpopulations were placed in intermediate neurons. The first generations were close to the neuron with the recipient, and as more generations and more migrants, these subpopulations were approaching the donor population (Figure 3). This confirms that migration narrows the differences among populations. This behavior was observed for the different values of m. The smaller the number of migrants, the less distant the outputs on the map, which is evidenced by the clearer color between neurons for m = 50 and a darker color for m=300 (Figure 3). Thus, this work also confirms that SOM is more suitable for perceiving genetic variability, bringing additional information by organizing populations on a map.

SOM were also used to organize populations under different evolutionary factors by Oliveira et al. (2021). The results for selection and migration corroborate the results presented by references (Fonseca et al., 2020; Oliveira et al., 2020b; Vidon et al., 2021; Fonseca, 2021), respectively. In addition, Oliveira et al. (2020c) also evaluated populations under genetic drift and inbreeding (Figure 4). Populations subjected to genetic drift for ten generations were organized by SOM according to the generations. For the first generation, populations became more centralized and closer, and as generations progressed, populations were organized more distantly and less centralized (Figure 4A). This organization represented what happens with genetic variability among sampled populations. In this situation, the genetic variability among populations under genetic drift increases. The organization of inbreeding populations by SOM also reflected what occurs with genetic variability (Figure 4B). Because the genetic variability between inbreeding populations decreases over generations and thus, these were allocated to the same neuron.

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

ARTIGO

Figure 3 - Self-organizing map of 50 migrants from the donor population (D) to the recipient population (R) over 10 generations (m=50 and g=10) in a 4 x 3 arrangement and a hexagonal topology (A). Self-organized map of 300 migrants from the donor population (D) to the recipient population (R) over 10 generations (m=300 and g=10) in a 4 x 3 arrangement and a hexagonal topology (B). The subpopulations were composed of migrants plus R (R+m). The first generation is represented y (1), the second generation is the sum of (1) plus m migrants, and so on. Adapted (Vidon et al., 2021).

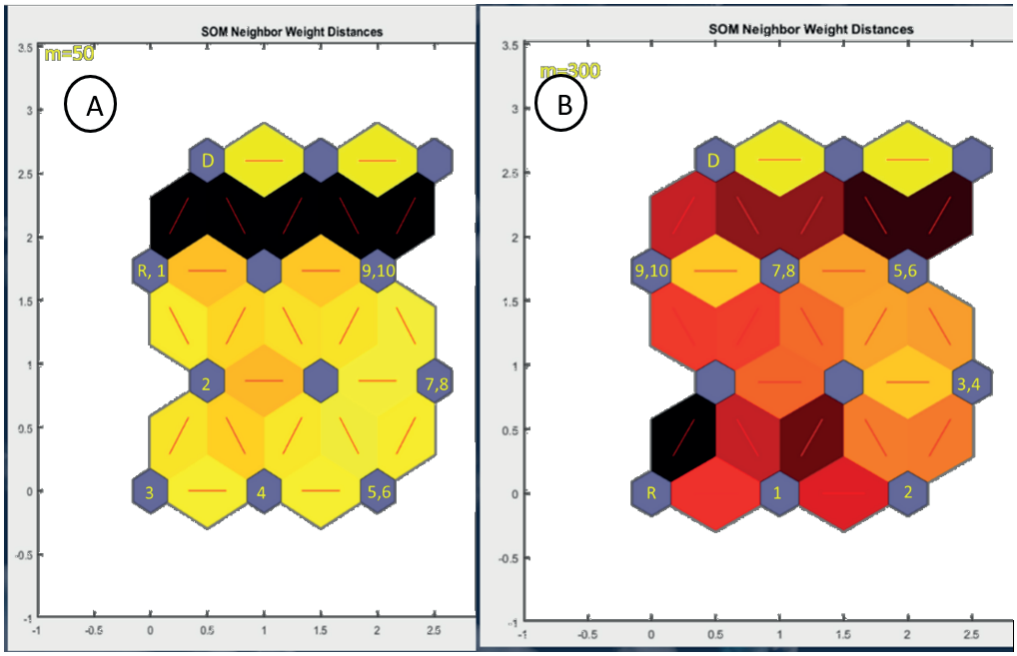
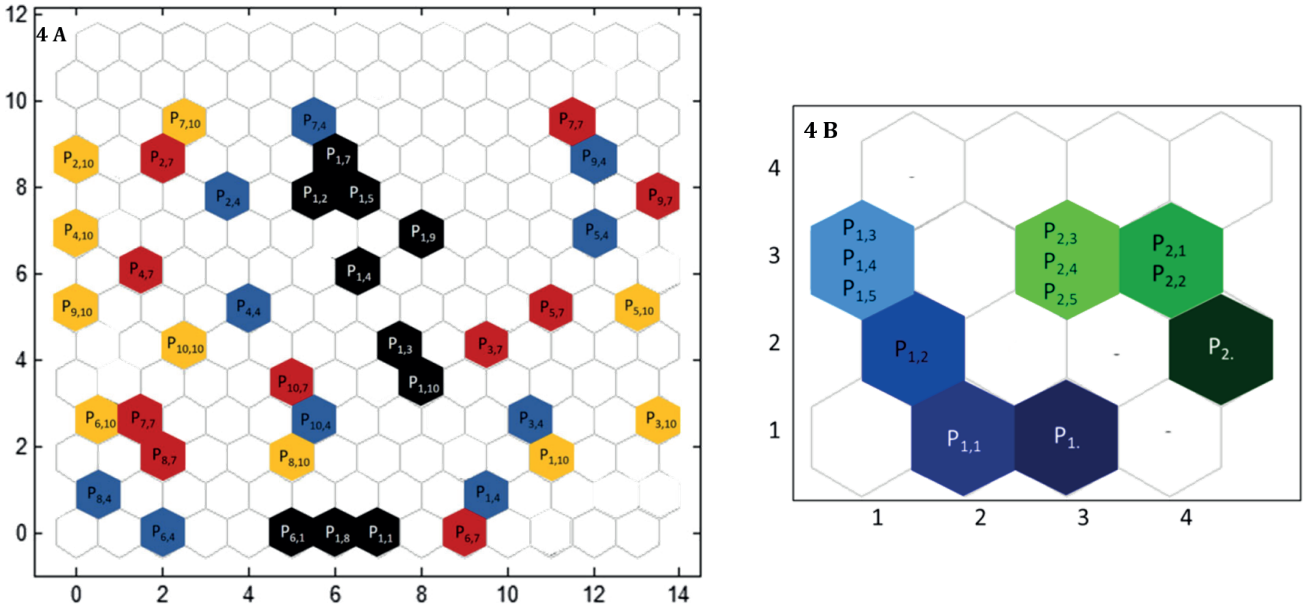


Figure 4 - A. Self-organizing map for the genetic drift effect. Pi,j denotes the population i (i = 1, 4, 7 and 10) in the generation j (j = 1, 2,..., 10). The base population is represented in black and their derivated subpopulations are represented in blue (P,4), red (P,7) and yellow (P,10); B. Self-organizing map of the populations submitted to self-pollination to generate the inbreeding effect. P1 and P2 are base populations that have gone through five generations of self-pollination, each. Neurons represented in blue show the organization of populations derived from P1; and those in green, show the series of self-pollinations from P2 (Oliveira et al., 2020c).



Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Thus, it can be seen that self-organizing maps are more suitable to perceive genetic variability and bring additional information by organizing populations in a map, in addition to recognizing the pattern of variation in genetic variability generated by evolutionary factors.

The application of computational intelligence using a self-organized map is promising and efficient for studies on the genetic diversity between biomass sorghum genotypes (Silva et al., 2020). SOM were also efficient in classifying Egyptian wheat varieties (Ibrahim et al., 2016), as well as in analyzing the genetic diversity of papaya (Carica papaya L.) accessions (Barbosa et al., 2011). Furthermore, these authors stated the maps are promising due to their performance compared to other clustering techniques (STRUCTURE and FST) for organizing genetic diversity (Peña-Malavera et al., 2014). Also, SOMs are superior to the traditional methods of organization of the dissimilarity and genetic divergence in germplasm banks (Moura et al., 2015).

The current review leads to the conclusion that SOMs are a promising technique for the organization of populations regardless of the type of molecular markers (Peña-Malavera et al., 2014) and phenotypic markers (Barbosa et al., 2011; Ibrahim et al., 2016; Silva et al., 2020). Furthermore, co-dominant markers can be used for estimating allelic and genotypic frequencies to be used as input in SOMs in real or simulated populations (Fonseca et al., 2020; Oliveira et al., 2020; Vidon et al., 2021; Fonseca, 2021). The efficiency of SOM in the organization of the genetic diversity of populations is unquestionable. However, each new population requires a unique investigation. The literature and researcher's experience have to be considered to determine the number of neurons and learning parameters since SOM is an empirical process (Kohonen, 2001).

Acknowledgments

We thank the Federal University of Viçosa (UFV) and the State University of Minas Gerais (UEMG) for the opportunity of our research work. We are especially grateful to the student Vicente Paulo Gomes Fonseca, Letícia Rodrigues Vidon and Maria Eduarda Farias Pigaiani for their efforts in developing their research.

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Referências Bibliográficas

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, 1(3), 224-231. https://doi.org/10.1590/S1984-70332011000300004

Cruz, C. D., Ferreira, F. M., Pessoni, L. A. (2020). Biometria aplicada ao estudo da diversidade genética (2nd ed.). Suprema.

Flint-Garcia, A. S, Thornsberry, J. M, Buckler, I. V. (2003). Structure of Linkage Disequilibrium in Plants. Annual Review of Plant Biology, 54(1), 357-374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Fonseca, P. G. F. (2021). Diferenciação genética em populações simuladas sob seleção divergente utilizando mapas auto-organizáveis [Completion of Course Work, University of Minas Gerais State].

Fonseca, V. P. G., Oliveira, M. S., Cruz, C. D. (2020). Padrão da diversidade gerada pela seleção divergente pelos mapas auto-organizáveis de Kohonen (SOM). [Conference presentation]. XI International Symposium on Genetics and Breeding, Viçosa, UFV. https://ee111266-e722-45e5-a03b-fee372-f1afb5.filesusr.com/ugd/7f9ade_c1f6b8d6dd644016829ed4619273da1f.pdf

Ibrahim, O. M., Tawfik, E. M. M., Badr, A., Wali, A. M. (2016). Evaluating the Performance of 16 Egyptian Wheat Varieties Using Self-Organizing Map (SOM) and Cluster Analysis. Journal of Applied Sciences,16(2), 47-53. https://doi.org/10.3923 / jas.2016.47.53

Kim, B. Y., Huber, C. D., Lohmueller, K. E. (2017). Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics, 206(1), 345-361. https://doi.org/10.1534 / genetics.116.197145

Kitani, E. C. (2013). Mapeamento e visualização de dados em alta dimensão com mapas auto-organizados [Doctoral dissertation, Polytechnic School, University of São Paulo]. https://doi.org/10.11606/T.3.2013.tde-11072014-114804

Kohonen, T. (2016). Essentials of the self-organizing map. Neural Networks, 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.

Kohonen, T. (2001). Self-Organizing Maps (3nd ed.). Springer. https://doi.org/10.1007/978-3-642--56927-2.

Mathieson, L., Mcvean, G. (2013). Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics, 193(3), 973-984. https://doi.org/10.1534 / genética.112.147611

Moura, M. C. C. L., Azevedo, A. M., Silva, D. J. H., Cruz, C. D. (2015). Potencialidades das redes neurais artificiais na avaliação de recursos genéticos em bancos de germoplasma. Revista RG News: Sociedade Brasileira de Recursos Genéticos. 1(1), 14-19. https://doi.org/10.1590/0102-7786355000009

Nascimento, M., Nascimento, A. C. C., Cruz, C. D. (2018). SOM - Mapas Auto-Organizáveis de Kohonen, In C.D. Cruz, M. Nascimento (Eds.), Inteligência Computacional Aplicada ao Mellhoramento Genético. Editora UFV.

Oliveira, M. S., Cruz, C. D. (2021). Genética de populações com o aplicativo GPOP. Brazil Publishing. https://doi.org/10.31012/9786550163563

Ambciências
Revista Brasileira de Tecnologia, Educação e Ciências Ambientais
https://revista.uemg.br/index.php/ambciencias

Oliveira, G. l., Souza, A. P., Oliveira, F. A., Zucchi, M. I., Souza, L. M., Moura, L. M. (2020a). Genetic structure and molecular diversity of Brazilian grapevine germplasm: Management and use in breeding programs. Plos One, 15(10): e0240665. https://doi.org/10.1371/journal.pone.0240665

Oliveira, M. S., Fonseca, V. P. G., Cruz, C. D. (2020b). Distância Genética de Nei e Hedrick e mapa auto organizáveis de Kohonen na percepção dos efeitos da seleção divergente. [Conference presentation]. XI International Symposium on Genetics and Breeding; Viçosa: UFV; https://ee111266-e722-45e5-a03b-fee372f1afb5.filesusr.com/ugd/7f9ade_c1f6b8d6dd644016829ed4619273da1f.pdf

Oliveira, M. S., Santos, I. G., Cruz, C. D. (2020c). Self-organizing maps: a powerful tool for capturing genetic diversity patterns of populations. Euphytica, 216(3), 1–9. https://doi.org/10.1007/S10681-020-2569-0

Peña-Malavera, A., BRUNO, C., FERNANDEZ, E., BALZARINI, M. (2014). Comparison of algorithms to infer genetic population structure from unlinked molecular markers. Statistical applications in genetics and molecular biology, 13(4), 39-402. https://doi.org/10.1515 / sagmb-2013-0006

Perez, C. C. M. (2008). Measures of genetic differentiation in simulate populations under inbreeding and divergent selection [Mester dissertation, Viçosa Federal University]. http://locus.ufv.br/handle/123456789/2371

Santos, I. G., Carneiro, V. Q., Silva Júnior, A. C., Cruz, C. D. (2019). Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. Acta Sci., 41, e39803. https://doi.org/10.4025/actasciagron.v41i1.39803

Santos, I. G., Rocha, J. R. A. S. C., Vigna, B. B. Z., Cruz, C. D., Ferreira, R. P., Basigalup, D. H., Marchini, R. M. S. (2020). Exploring the diversity of alfalfa within Brazil for tropical production. Euphytica, 216(5), 1–15. https://doi.org/10.1007/S10681-020-02606-W

Silva, M. J., Silva Júnior, A. C. S., Cruz, C. D., Nascimento, M., Oliveira, M. S., Schaffert, R. E., Parrella, R. A. C. (2020). Computational intelligence for studies on genetic diversity between genotypes of biomass sorghum. Pesq. agropec. bras., 55, e01723, https://doi.org/10.1590/S1678-3921.pab2020.v55.01723

Spanoghe M. C., Marique T., Rivière J., Moulin M., Dekuijper C., Nirsha A., Bonnave M., Lanterbecq D. (2020). Genetic patterns recognition in crop species using self-organizing map: the example of the highly heterozygous autotetraploid potato (Solanum tuberosum L.). Genetic Resources of Crop Evolution, 67, 947–966. https://doi.org/10.1007/s10722-020-00894-8.

Vidon, L. R, Pigaiani, M. E. F, Oliveira, M. S. (2021, November 24-26). Mapas auto-organizáveis na percepção da Diferenciação genética causada por migração. [Conference presentation]. 23ª Seminário de Pesquisa e Extensão of the University of Minas Gerais State, Carangola: UEMG.