

## DISSIMILARIDADE GENÉTICA DE HÍBRIDOS DE TOMATEIRO UTILIZANDO APRENDIZADO DE MÁQUINA NÃO-SUPERVISIONADO

Edesio Rodrigues de Souza Junior<sup>1</sup>, Lucas Medeiros Pereira<sup>2</sup>, Orlando Ribeiro de Oliveira<sup>1</sup>, Vinicius Augusto Pereira<sup>1</sup>, Gabriel Mascarenhas Maciel<sup>1</sup>, Ana Carolina Silva Siquieroli<sup>1</sup>

<sup>1</sup> Universidade Federal de Uberlândia, Monte Carmelo, MG (edesio.junior@ufu.br); <sup>2</sup> Universidade Federal de Uberlândia, Uberlândia, MG

**RESUMO:** O tomateiro (*Solanum lycopersicum* L.) é uma hortaliça de grande importância, onde os programas de melhoramento genético buscam obter híbridos superiores. Técnicas de análise de agrupamento são importantes ferramentas para ajudar na tomada de decisão em relação à dissimilaridade genética. No tomateiro, novas técnicas de agrupamento vêm sendo utilizadas, como é o caso do método de agrupamento K-means e o mapa auto-organizável de Kohonen (SOM). Logo, o objetivo deste trabalho foi avaliar a capacidade de formação de grupos dos métodos de agrupamento K-means e SOM em híbridos de tomateiro provenientes do banco de germoplasma da UFU. O experimento foi realizado na Estação Experimental de Hortaliças (EEH-UFU) da UFU campus Monte Carmelo. Foram utilizados 11 híbridos de tomateiro pertencentes ao banco de germoplasma da UFU. Foi utilizado o delineamento em blocos casualizados com três repetições. Foram avaliadas as características de DL, DT, FS, EP, NL, Brix, NFP e PMF. Os dados foram submetidos à análise de variância, pelo teste F ( $p \leq 0,05$ ). Para o K-means foi utilizado  $k=4$ . Já para o SOM, foi definido arranjo 3x3, com raio igual a um e topologia hexagonal. Após a obtenção dos resultados, foi possível observar que os dois métodos foram capazes de identificar a dissimilaridade genética entre os híbridos. O método SOM apresentou maior capacidade de formação de grupos, onde foram formados seis grupos distintos. A utilização de métodos de agrupamentos com classificação não-supervisionada são ferramentas importantes na identificação da dissimilaridade genética em programas de melhoramento genético de tomateiro.

**Palavras-chave:** Melhoramento genético, clusterização, algoritmos de agrupamento.

### 1 INTRODUÇÃO

O tomateiro (*Solanum lycopersicum* L.) é uma hortaliça que apresenta grande importância de cultivo mundial, sendo responsável pela grande demanda de mão-de-obra, gerando impacto econômico e social nas regiões de cultivo (Carvalho *et al.*, 2014).

Devido à importância da cultura, os programas de melhoramento genético de tomateiro buscam a obtenção de híbridos que apresentem resiliência contra os estresses bióticos e abióticos. Os melhoristas analisam os híbridos desenvolvidos para obterem materiais produtivos, plantas mais compactas para otimização de espaço e resistência contra pragas, doenças e outros estresses.

Para auxiliar na tomada de decisão sobre o desempenho dos híbridos, as análises de agrupamento são utilizadas como ferramentas para análise da dissimilaridade genética dos híbridos analisados. No tomateiro, diferentes técnicas de agrupamento podem ser utilizadas, como os agrupamentos hierárquicos (Pereira *et al.*, 2024). Entretanto, novas técnicas de agrupamento vêm sendo testadas e utilizadas no tomateiro, como é o caso da análise de agrupamento pelo método K-means e pelo mapa auto-organizável de Kohonen (SOM) (Pereira *et al.*, 2024).

O método K-means é uma técnica de classificação não-supervisionada de fácil implementação e baixo custo computacional. A técnica utiliza a definição prévia do número de grupos e posterior realocação interativa dos indivíduos em função da melhor posição dos centroides (Fachieri; Suaide, 2025). Já o SOM, utiliza o princípio das redes neurais artificiais, reduzindo a dimensionalidade dos dados, facilitando a visualização de padrões. Diferentemente do K-means, no SOM não é necessário definir o número de grupos do agrupamento, é necessário apenas definir a topologia do mapa e este se auto-organiza em função dos dados (Cruz; Nascimento, 2018).

Portanto, o objetivo deste trabalho foi avaliar a dissimilaridade genética de híbridos de tomateiro pertencentes ao banco de germoplasma da UFU com bases nas técnicas de agrupamento K-means e SOM.

## 2 MATERIAL E MÉTODOS

O experimento foi conduzido na Estação Experimental de Hortaliças da Universidade Federal de Uberlândia, campus Monte Carmelo (18°42'43,19" S; 47°29'55,8" O; 873 m de altitude) no período de fevereiro a junho de 2025. Foram avaliados 11 híbridos pertencentes ao Banco de Germoplasma da UFU campus Monte Carmelo (UFU-TOMd1, UFU-TOMd2, UFU-

TOMd3, UFU-TOMd4, UFU-TOMd5, UFU-TOMd6, UFU-TOMd7, UFU-TOMd8, UFU-TOMd9, UFU-TOMd10 e UFU-TOMd11), distribuídos em três repetições no delineamento em blocos casualizados.

Para a análise da dissimilaridade dos híbridos, foi avaliada os diâmetros longitudinal (DL) e transversal (DT) dos frutos, em centímetros (cm). O formato de fruto (FF) foi obtido pela razão entre o DL e DT. A espessura da polpa (EP) em cm, foi obtida após o corte transversal do fruto. O número de lóculos (NL) foi obtido pela contagem dos lóculos visualizados após o corte transversal do fruto. O teor de sólidos solúveis (BRIX) em °Brix, foi obtido pela medida com o auxílio de refratômetro digital com graduação de 0,2°Brix. O número de frutos por planta (NFP) foi obtida pela contagem do número de frutos comerciais. O peso médio de fruto (PMF) foi obtido pela razão entre a massa de frutos e o número de frutos.

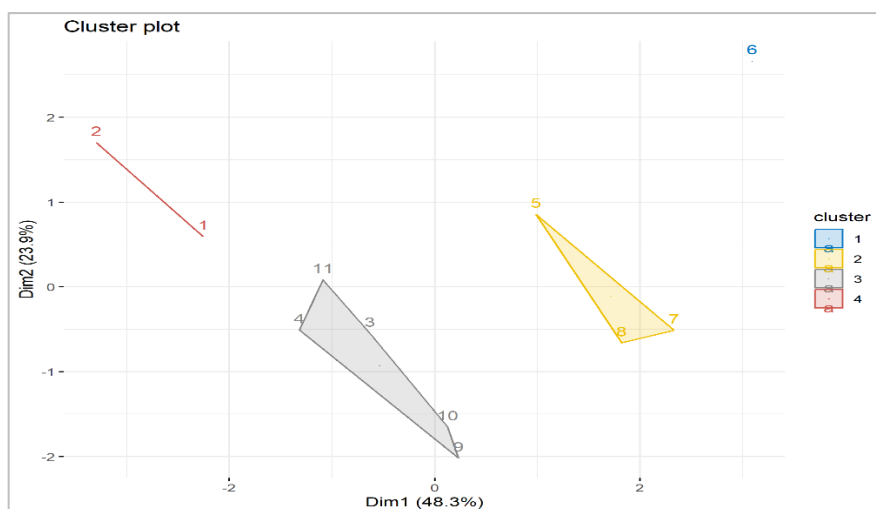
Para as análises de dissimilaridade genética, o método K-means foi utilizado com a definição de  $k=4$ , sendo este número de grupos definido com base no método Silhouette (Rousseeuw, 1987). Para o método SOM, foi definido o arranjo 3x3 com base na melhor relação entre os erros topológico e quantização, com raio igual a um e topologia hexagonal.

As análises foram realizadas no software R v. 4.2.1 (R Core Team, 2025). Para a utilização do método K-means foi utilizado o pacote factoextra (Kassambara, 2021). Para gerar o SOM, foi utilizado o pacote Kohonen (Wehrens; Kruisselbrink, 2018).

### 3 RESULTADOS E DISCUSSÃO

A obtenção do agrupamento pelo método K-means permitiu a formação de quatro grupos distintos (Figura 1).

Figura 1 – Agrupamento K-means com  $k=4$ .



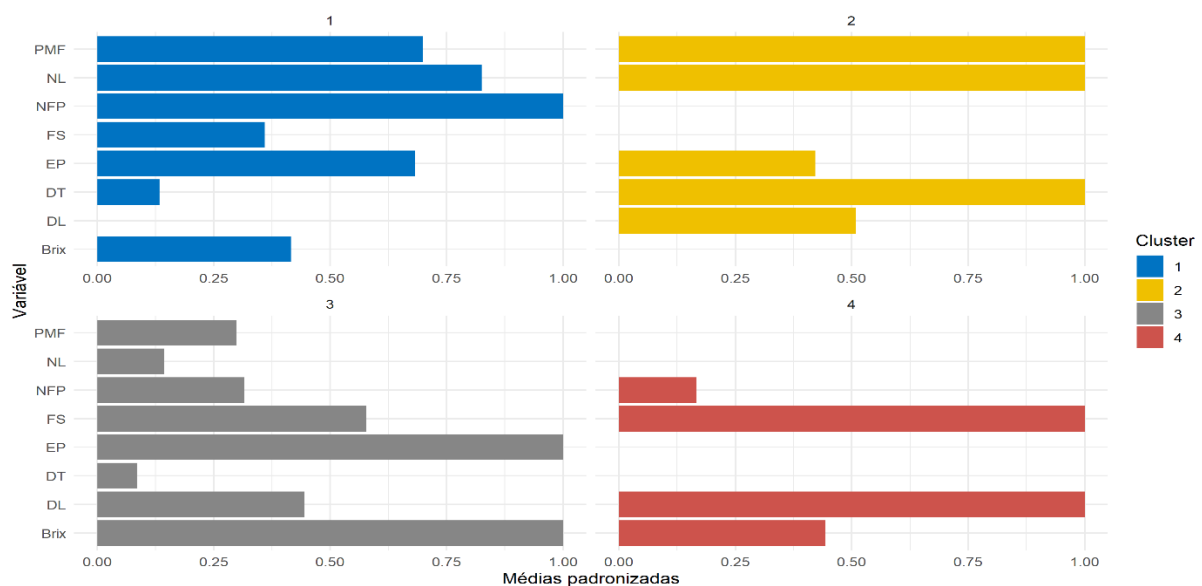
Nota: 1: UFU-TOMd1, 2: UFU-TOMd2, 3: UFU-TOMd3, 4: UFU-TOMd4, 5: UFU-TOMd5, 6: UFU-TOMd6, 7: UFU-TOMd7, 8: UFU-TOMd8, 9: UFU-TOMd9, 10: UFU-TOMd10 e 11: UFU-TOMd11.

Fonte: elaborado pelo próprio autor.

O primeiro grupo foi formado pelos híbridos UFU-TOMd1 e UFU-TOMd2; o segundo grupo por UFU-TOMd3, UFU-TOMd4, UFU-TOMd9, UFU-TOMd10 e UFU-TOMd11; o terceiro por UFU-TOMd5, UFU-TOMd7 e UFU-TOMd8 e o quarto grupo pelo híbrido UFU-TOMd6.

Foi possível observar também quais variáveis mais influenciaram em cada grupo formado (Figura 2).

Figura 2 – Importância das variáveis nos quatro clusters formados em função das médias de diâmetro longitudinal em cm (DL); diâmetro transversal em cm (DT); relação entre DL e DT (FF); espessura de polpa em cm (EP); teor de sólidos solúveis em °Brix (Brix); número de frutos por planta (NFP) e peso médio de fruto em gramas (PMF).



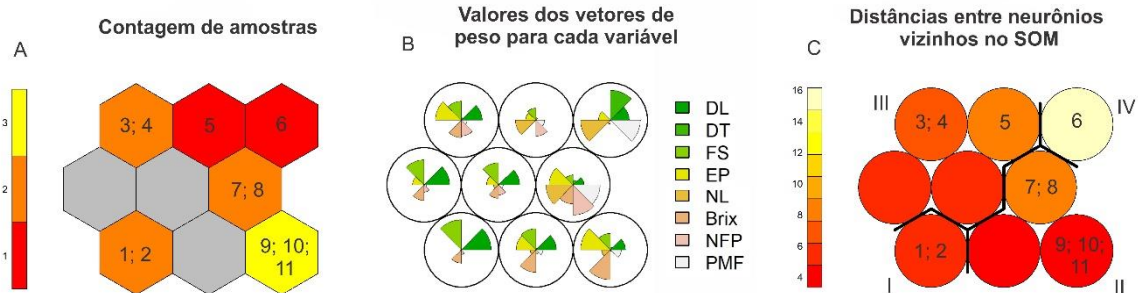
Nota: 1: UFU-TOMd1, 2: UFU-TOMd2, 3: UFU-TOMd3, 4: UFU-TOMd4, 5: UFU-TOMd5, 6: UFU-TOMd6, 7: UFU-TOMd7, 8: UFU-TOMd8, 9: UFU-TOMd9, 10: UFU-TOMd10 e 11: UFU-TOMd11.

Fonte: elaborado pelo próprio autor.

O método K-means possui fácil implementação e facilidade na interpretação dos resultados. Entretanto, a presença de outliers no conjunto de dados pode comprometer a formação dos grupos devido ao ruído provocado nos centróides (Lopes; Gosling, 2021).

Utilizando o método de agrupamento SOM foi possível a obtenção de seis grupos distintos (Figura 3).

Figura 3 - (A): Classificação dos tratamentos em função do número de neurônios. (B): Representação dos neurônios e a magnitude de influência das variáveis. (C): Distâncias dos vizinhos mais próximos e clusterização dos neurônios com base no método UPGMA em função das distâncias. Diâmetro longitudinal em cm (DL); diâmetro transversal em cm (DT); relação entre DL e DT (FS); espessura de polpa em cm (EP); teor de sólidos solúveis em °Brix (Brix); número de frutos por planta (NFP) e peso médio de fruto em gramas (PMF).



Nota: 1: UFU-TOMd1, 2: UFU-TOMd2, 3: UFU-TOMd3, 4: UFU-TOMd4, 5: UFU-TOMd5, 6: UFU-TOMd6, 7: UFU-TOMd7, 8: UFU-TOMd8, 9: UFU-TOMd9, 10: UFU-TOMd10 e 11: UFU-TOMd11.

Fonte: elaborado pelo próprio autor.

A utilização do método SOM permitiu a formação de número maior de grupos (Figura 3A). O primeiro grupo (linha I e coluna I) foi formado pelos híbridos UFU-TOMd1 e UFU-TOMd2; o segundo grupo (linha I e coluna III) por UFU-TOMd9, UFU-TOMd10 e UFU-TOMd11; o terceiro grupo (linha II e coluna III) por UFU-TOMd7 e UFU-TOMd8; o quarto grupo (linha III e coluna I) por UFU-TOMd3 e UFU-TOMd4; o quinto grupo (linha III e coluna II) por UFU-TOMd5 e o último grupo (linha III e coluna III) pelo híbrido UFU-TOMd6.

Foi possível observar que as variáveis avaliadas influenciaram os neurônios de diferentes formas (Figura 3B). O terceiro grupo apresentou um maior peso dos vetores em relação ao neurônio, com um destaque maior para EP, NL, Brix, NFP e PMF. Já o primeiro grupo apresentou maior vetor de peso para as variáveis DL e FS.

A análise de proximidade dos neurônios (Figura 3C) foi possível observar que pelo método UPGMA houve a classificação dos neurônios em novos quatro grupos, onde o grupo IV, formado pelo híbrido UFU-TOMd6 foi o mais distante em relação aos demais.

Em relação ao método de agrupamento K-means, não houve coerência entre o método SOM, onde no primeiro método houve a formação de quatro grupos e no segundo método, a formação de seis grupos. A utilização do SOM apresenta vantagens, onde o método apresenta uma maior robustez e capacidade de manipulação e processamento de dados (Kohonen, 2014). A técnica vem sendo utilizada como uma ferramenta eficiente para análise da dissimilaridade genética em programas de melhoramento genético (Santos *et al.*, 2019).

## 4 CONCLUSÕES

As duas técnicas de análise de agrupamentos foram capazes de identificar a dissimilaridade genética presente entre os híbridos de tomateiro. O SOM se destacou por apresentar uma maior capacidade de formação de grupos em relação ao K-means. A utilização do SOM pode auxiliar a tomada de decisão dos melhoristas em programas de melhoramento genético do tomateiro.

## REFERÊNCIAS

CARVALHO, C. R. F. *et al.* Viabilidade econômica e de risco da produção de tomate no município de Cambuci/RJ, Brasil. **Ciência Rural**, v. 44, n. 12, p. 2293–2299, 2014.

CRUZ, C. D. and NASCIMENTO, M. **Inteligência computacional aplicada ao melhoramento genético**. 1th ed. Viçosa: UFV, 2018.

FACHIERI, D. N.; SUAIDE, A. A. P. Introdução a métodos de aprendizado de máquina não supervisionados através de um experimento simples de medidas de densidades de sólidos. **Revista Brasileira de Ensino de Física**, v. 47, e20240443, 2025.

KASSAMBARA, Alboukadel. factextra: **Extract and Visualize the Results of Multivariate Data Analyses**. Versão 1.0.7. [S.l.]: CRAN, 2021. Disponível em: <<https://CRAN.R-project.org/package=factextra>>. Acesso em: 2 jul. 2025.

KOHONEN, T. **MATLAB Implementations and Applications of the Self-Organizing Map**. Unigraphia Oy: Helsinki, Finland, 2014.

LOPES, H. E. G.; GOSLING, M. DE S. Cluster Analysis in Practice: Dealing with Outliers in Managerial Research. **Revista de Administração Contemporânea**, v. 25, n. 1, e200081, 2021.

PEREIRA, L. M. *et al.* Introgression of the self-pruning gene into dwarf tomatoes to obtain salad-type determinate growth lines. **Plants**, v. 13, n. 11, p. 1522, 2024.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53-65, 1987.

SANTOS, I. G. *et al.* Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. **Acta Scientiarum. Agronomy**, v. 41, e39803, 2019.

WEHRENS, R.; KRUISSELBRINK, J. *Flexible Self-Organising Maps in kohonen 3.0*. **Journal of Statistical Software**, v. 87, n. 7, p. 1-18, 2018.