Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1807-8621
https://doi.org/10.4025/actasciagron.v48i1.75392

Functional genetic diversity in popcorn lines (*Zea mays* L. var. Everta) focusing on productivity through SSR-EST markers

Lavínia Santana Ladeira Gomes*, Marcelo Vivas, Alex Souza Rodrigues, Lilia Marques Gravina, Ramon de Moraes, Rafael Nunes de Almeida, Helaine Christine Cancela Ramos and Antônio Teixeira do Amaral Júnior

Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego, 2000, Parque Califórnia, 28013-602, Campos dos Goytacazes, Rio de Janeiro, Brazil. *Author for correspondence. E-mail: laviniasladeiragomes@gmail.com

ABSTRACT. Popcorn is a distinct type of maize with a high market demand. In Brazil, there is a shortage of national cultivars that meet production and marketing needs, so seeds are imported, hindering the growth of this sector. Nevertheless, its cultivation is important and can be enhanced with investments in genetic improvement aimed at optimizing characteristics such as productivity. Molecular markers support this type of research: They allow the genetic characterization of lines and inferences about genetic variability. Microsatellites such as simple-sequence repeats (SSRs) are widely used in genetic studies due to their high reproducibility and polymorphism. SSR-expressed sequence tag (EST) markers are microsatellites related to transcribed sequences. Functional variability is linked to the maximization of genetic gains, making the study of these regions valuable. In this context, the present study used SSR-EST markers to characterize the productivity of 47 popcorn lines from the UENF Germplasm Bank, selected for their high performance in previous studies. Young leaves of the lines were collected for DNA extraction, which was used for polymerase chain reaction (PCR) with 30 preselected primers. The amplified material was subjected to capillary electrophoresis for identification and separation of alleles. The data were converted into a dissimilarity matrix, which allowed for the estimation of diversity metrics such as the number of alleles, the polymorphic information content, observed heterozygosity, expected heterozygosity, and the inbreeding coefficient. A total of 94 distinct alleles were detected, with an average of 3.13 alleles per locus, indicating that, although the lines present high levels of allelic fixation and inbreeding, they retain moderate genetic variability for population configuration. A dendrogram created using unweighted pair group method using arithmetic mean revealed an optimal number of six populations. This information provides indications of specific genotypes for crosses that can maximize heterosis in hybridizations to support the improvement program.

Keywords: microsatellites; dissimilarity; heterosis.

Received on January 24, 2025.
Accepted on June 4, 2025.

Introduction

Maize (*Zea mays*) is one of the most widely cultivated cereals in the world. It is highly versatile and capable of being processed into several by-products (Wang & Hu, 2021). There are several distinct types, including sweet maize, dent maize, white maize, and others, which are classified into heterotic groups or subspecies (Mascarenhas, 2019). Among the various types of maize produced, popcorn maize (*Zea mays* L. var. Everta) stands out due to its specific grain composition, which differs from other varieties (Kaur et al., 2021).

Popcorn kernels are generally small, rich in starch, and have a hard pericarp, which contributes to one of their primary qualities: expansion capacity (EC). The EC enables popcorn kernels to transform into actual popcorn, through the heating of the endocarp and the rupture of the pericarp under high temperature and pressure. This phenomenon results in the formation of light flakes, known as popcorn, which is consumed as a snack throughout the world (Zulkadir & İdikut, 2021).

Popcorn maize has a relatively limited breeding history compared with other types of maize (e.g., dent maize). Moreover, one of the crop's main traits, productivity, has a negative genetic correlation with the EC, making it more challenging to improve (Cabral et al., 2016). Furthermore, compared with common maize, popcorn has a lower grain yield, is less vigorous, is more susceptible to diseases and pests, is more prone to lodging, and has a less developed root system. These factors have a direct influence on production rates (Coan et al., 2019). In addition, there are various biotic and abiotic factors that affect the crop and impact production

(Guimarães et al., 2025). Overall, popcorn varieties are genetically limited and have suboptimal agronomic potential compared with common maize (Parsons et al., 2020). In this context, it is necessary to develop cultivars with good performance for key traits to meet the domestic market demand and to reduce the high dependence on seed imports (Almeida et al., 2024).

There has been increased public and private investment in the genetic improvement of popcorn in Brazil to reduce seed imports and to develop varieties adapted to different regions of the country (Guimarães et al., 2018). These improvement programs use diverse methodologies, including classical approaches based on predicting the genetic value of individuals from their phenotypes, and biotechnological methods, such as molecular markers, which allow precise and early genomic analyses without environmental interference, enhancing genetic gains and the efficiency of the programs (Galvão et al., 2015; Peixoto et al., 2022). Studies on popcorn markers have used the common maize genome as a reference (Muzdalifah et al., 2020). For this reason, reference maize genomes deposited in databases such as the National Center for Biotechnology Information (NCBI) and Maize Genetics and Genomics Database (MaizeGDB), sequenced using high-coverage technologies, are of great value, as they can serve as parameters for popcorn genome studies (Pinto et al., 2023).

Among the available molecular markers, the simple sequence repeat (SSR) is commonly used in maize genetic studies because of its high abundance in the genome, ease of analysis, and accuracy. SSR markers are tandemly repeated DNA sequences that are multiallelic, highly polymorphic, cost-effective, highly reproducible, and efficient (Galvão et al., 2015; Muzdalifah et al., 2020). The use of SSR markers in popcorn genomes allows inferences at genetic and genomic levels. Genomic variations are those that occur anywhere in the genome, while genetic variations occur exclusively in coding regions, referred to as transcribed sequences. These have the additional advantage of assessing the functionality of the products (Galvão et al., 2015). SSR gene markers are known as SSR-expressed sequence tags (ESTs). They have been described in the literature as being associated with productivity. Although they are reported to be less polymorphic than genomic markers, they are considered to be superior in terms of functional diversity and transferability (Kara, 2023). Polymorphisms can modify the final product of genetic expression, influence the regulation of genetic expression of other compounds, or lead to changes in metabolic cascades, potentially modulating traits of agronomic interest (Galvão et al., 2015).

This study aimed to investigate the genomes of popcorn maize lines by detecting functional genetic variations through SSR-EST markers. The genomes were previously selected from other stages of crop improvement within the State University of the North Fluminense Darcy Ribeiro (UENF) Popcorn Genetic Improvement Program; they all showed good yields. The results of this study will assist in identifying genetically contrasting groups for the definition of crossing blocks. The application of this technique could reduce the number of required crosses by targeting the most promising genitors (Ferreira Junior et al., 2020).

Material and methods

Plant material

Forty-seven lines from the UENF Germplasm Bank were selected based on their superior performance in key agronomic traits relevant to popcorn cultivation (Table 1). The earlier field selection stages were conducted by the Laboratory of Agricultural Engineering (LEAG).

Table 1. Identification of each line and the number of lines according to the genealogy/progeny and institution in which they were originated or developed.

Number of lines	Genotypes	Genealogy/ progeny	Development institution/origin
7	L381, L383, L385, L386, L390, L391, L394	SAM	South America/USA
2	L653, L656	ARZM 13 050	Argentina (CIMMYT)
8	L213, L206, L222, L209, L204, L207, L216, L217	IAC 125	Campinas, São Paulo State (Brazil)
1	L443	BOYA 462	Colombia (CIMMYT)
2	L592, L594	RS 20	IPAGRO, Porto Alegre, Rio Grande do Sul State (Brazil)
1	L480	SE 013	Maringá, Paraná State (Brazil)
1	L625	PA 091	Maringá, Paraná State (Brazil)
2	L501, L503	PA 170 ROXO	Paraguay (CIMMYT)
1	L274	PARA 172	Paraguay (CIMMYT)
3	L61, L63, L69	Angela	Embrapa, Sete Lagoas, Minas Gerais State (Brazil)
2	L353, L357	PR 023	UEM, Maringá, Paraná State (Brazil)

4	L683, L689, L690, L691	UENF 14	UENF, Rio de Janeiro State (Brazil)
5	L325, L326, L328, L330, L332	Barão de Viçosa	UFV, Viçosa, Minas Gerais State (Brazil)
3	L76, L77, L80	Viçosa-Viçosa	UFV, Viçosa, Minas Gerais State (Brazil)
2	L55, L55*	Beija-Flor	UFV, Viçosa, Minas Gerais State (Brazil)
3	L291, L292, L294	URUG298	Uruguay (CIMMYT)

*It indicates that there was variation in grain color for lineage 55 during the seed multiplication phase that preceded this work. CIMMYT, Centro Internacional para el Mejoramiento de Maíz y Trigo; IPAGRO, Instituto de Pesquisas e Desenvolvimento Agrícola e Ambiental; UEM, Universidade Estadual de Maringá; UFV, Universidade Federal de Viçosa; UENF, Universidade Estadual do Norte Fluminense Darcy Ribeiro.

DNA extraction

To evaluate genetic diversity, seeds from the selected lines were planted in plastic trays containing 96 tubes, with three replications per genotype and two seeds per tube. The seeds were irrigated and maintained in a greenhouse until the V2 development stage. At that point, young, healthy, and active leaves were collected, immediately wrapped in aluminum foil, and stored at a low temperature to ensure the integrity of the DNA for future extraction. The plant material was taken to the DNA Markers Sector/LMGV at UENF. Then, the cetyltrimethylammonium bromide (CTAB) protocol was performed as described by Doyle and Doyle (1987), with minor modifications.

DNA quantification

The extracted DNA was stained with a GelRed + Blue Juice solution (1:1) and subjected to electrophoresis in 1% metaphor agarose gel in 1× Tris, sodium acetate, and ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (pH 8.0) under constant voltage (80 V) and current (0.20 A). Subsequently, the gel was exposed to ultraviolet light to visualize the DNA bands and an image was taken using a ultraviolet (UV)-sensitive camera. The GelQuant software provided an estimate of the DNA concentration and assessed the quality and integrity of the material. Then, the DNA samples were diluted to 5 ng μL^{-1} for the subsequent experiments.

Primer acquisition

A literature search was conducted to identify polymorphic SSR-EST primers and information associated with productivity for genotyping of the selected lines. This selection involved the use of NCBI and MaizeGDB, which provide developed and mapped sequences for *Z. mays*. Thirty primers were selected to obtain adequate genome samples (Table 2).

Table 2. List of popcorn gene-associated microsatellite loci used for the analysis of the target lines.

Primer ID	Locus	Chromosome	Ta (ºC)	Primer sequence (forward and reverse)
1	UENF7	1	58	GACAGCAAGGGATGGACAGT TCAAATCAATGGCTGCTCAA
2	UENF9	1	55	GCGACCACAAGGAGAAGAAG TGCATGACACCAGTGAAACA
3	UENF74	9	55	GCGTCATCAGCCTAGCAGTT CAGGCCATGATGAGATCAGA
4	BNLG1331	1	58	CGCCTGCCAAGGTACATCAC GCACTCACGCGCACAGGTCA
5	pbf1	2	64	ACAAGGCCATCATGAAGAGCAGTA CACGGTCTGGCACACTAACCTTAT
6	fdx3	1	63	GTGGGTTTTGTCTGTAGGGTGGTA TCCATCCACTCGACTTAAGAGTCC
7	bip2	4	60	AGCAAGCAGTTCGAAACAAGGAT GACACCAGCACCACTTGAACG
8	umc1031	4	60	ACGTGGACAACCAGTCTATCAACA TTGGGTTCATACCTCCTAGGAACA
9	umc1152	10	60	CCGAAGATAACCAAACAATAATAGTAGG ACTGTACGCCTCCCCTTCTC
10	umc2292	5	62	AGCAGAAGAGGACAAACCAGATTC ACTTCCGGCATGTCTTGTGTTT
11	umc1252	2	63	CTTCTGCATCATCATCATCGTCTT GCGTCGGAGAAGTACATCAAGTTT
12	umc1082	1	63	CCGACCATGCATAAGGTCTAGG GCCTGCATAGAGAGGTGGTATGAT
13	umc1448	2	63	ATCCTCTCATCTTTAGGTCCACCG

14	umc1389	5	63	CATATACAGTCTCTTCTGGCTGCTCA AAAACACAACGCTGGACATCAAC
15	umc1221	5	63	GGTCGTTTTGCTTAGCCCATTTTA AAACAGGCACAAAGCATGGATAG
16	umc1155	5	63	GCAACAGCAACTGGCAACAG CCTGAGGGTGATTTGTCTGTCTCT
17	umc1157	8	63	TCTTTTATTGTGCCCGTTGAGATT AACTCGCTATCGAAAAACCACAAG
18	umc1176	10	63	TCGGATTTTAGCTGAGCTTGTACC ATGAGTTCATGACAGAGCGCTACC
19	umc1415	8	63	GAGTTTGTTCGTTTGTGTGTGGAG GTGAGATATATCCCCGCCTTCC
20	umc2343	9	63	AGACTTCCTGAAGCTCGGTCCTA TCATCTTCCCCACAAATTTTCATT
21	umc1130	8	63	GACTGACAACTCAGATTTCACCCA TTGGGACTCATTACTTCCGGACT
22	umc1604	2	63	GCTAGGGGAAAGCTCGTACTATGG ACAGCCACCACCACTCCATC
23	umc1135	3	63	AGATCTTCGAGAGCGTGGAGAAC CGCTAGCTTAGCTCCATCGTTTTA
24	umc1587	5	63	TTTTTAACCTCACGAGCATCGTCT AGGTGCAGTTCATAGACTTCCTGG
25	umc2401	8	63	ATGCGTCTTTCACAAAGCATTACA TTTTCTTCTCCTTCCCTCACCTG
26	umc2246	2	63	GCACCTGATGCAGTAGGGAGTC AGGCTCCAGCTCTAGGGGAGT
27	umc1867	9	63	GTGAACTGTGTAGCGTGGAGTTGT ATAAGCTCGTTGATCTCCTCCTCC
28	umc1506	10	63	TGGTCTTCTTCGCCGCATTAT AAAAGAAACATGTTCAGTCGAGCG
29	umc1450	7	63	ATAAAGGTTGGCAAAACGTAGCCT ACTTTATTTAGCCCACGTCACTCG
30	umc1948	1	56	AGTATGACACGGGATTTTGCTGTT ATCTATCTGGTTTCACGATCTCGC TGTTGAAATAATGGAACACCTCCC

The "Chromosome" column indicates the chromosome on which each locus is located. Ta (°C) specifies the annealing temperature used for each primer. The forward sequences are presented 5′→3′, and the reverse sequences are presented as 3′→5′.

Polymerase chain reaction (PCR)

Each PCR was performed in a final volume of 12 µL, containing 2.0 µL of DNA (5 ng μL^{-1}), 1.0 µL of 25 nM $MgCl_2$ (Thermo Scientific, Fermentas, USA), 1.2 µL of 2 mM dNTPs (Sigma-Aldrich, USA), 1.0 µL of the forward and reverse primers (Sigma-Aldrich), 0.2 µL of Taq DNA polymerase (5 U μL^{-1}; Thermo Scientific, Fermentas), 10× buffer (NH_4) 1:2 reaction (Thermo Scientific, Fermentas), and 5.4 µL of ultrapure water. The Eppendorf Applied Biosystems/Veriti 96 Well thermocycler was used for PCR. The thermal cycling program was initial denaturation at 94°C for 5 minutes; 35 cycles of denaturation at 94°C for 1 minute, annealing at the optimal temperature for each primer (Ta in Table 2) for 1 minute, extension at 72°C for 2 minutes; and a final extension at 72°C for 7 minutes.

Capillary electrophoresis protocol

The PCR products were diluted in a specific reagent and distributed into 96-well plates for capillary electrophoresis using a Fragment Analyzer machine, along with markers containing fragments ranging from 1 to 500 base pairs (bp). In the sample plate, a 250-bp DNA ladder (Invitrogen, USA) was added to well H12 to determine the size of the amplified fragments. Capillary electrophoresis was conducted by applying an electric current of 7.5 kW for approximately 140 minutes.

Data analysis: diversity metrics

The DNA that was obtained from electrophoretic separation was used for allelic coding to identify distinct alleles within each locus. This information was converted into a numerical matrix following the procedure described by Ramos et al. (2014). This matrix served as the basis for estimating genetic distance through dissimilarity using the unweighted index (UI), with the support of the Genes software (version 2018.23).

The matrix was also utilized to estimate diversity metrics with the assistance of the Powermarker 3.5 software (Liu & Muse, 2005). The following diversity were calculated: the number of alleles per locus (Na), expected heterozygosity (He), observed heterozygosity (Ho), the polymorphic information content (PIC), and the inbreeding coefficient (F) (Rodrigues et al., 2023).

He was determined using the following expression:

$$HE = 1 - \sum_{i=1}^{k} x_i^2$$

where X_i is the frequency of allele i and K is the total number of alleles.

Ho was estimated using the formula:

$$HO = \frac{\sum_{j=1, j\neq 1}^{a} nij}{n}$$

where Nij is the number of heterozygous individuals at each locus and N is the total number of individuals analyzed.

The PIC was calculated using the formula:

$$PIC = 1 - \sum_{i=1}^{a} pi^2 - \sum_{i,j=1}^{a} . \sum_{(i\neq j)}^{a} pi^2 pj$$

where $\Sigma i\ pi^2$ is the informative value of the primer and $\Sigma\Sigma\ pi^2\ pj^2$ is the frequency of allele p in primer j.

The fixation index, which estimates the average F, was determined using the formula:

$$F = 1 - \frac{HE}{HO}$$

where H_O is the frequency of homozygous individuals and H_E is the frequency of heterozygous individuals.

Data analysis: Clustering method

The genetic dissimilarity matrix was exported to the STRUCTURE software (version 2.3.4) to determine the optimal number of populations (K) among the 47 popcorn genotypes using a Bayesian model. A total of 100,000 iterations were performed for each K value, ranging from 1 to 12, ensuring a broad coverage of possibilities. The method proposed by Evanno et al. (2005) was applied to select the optimal K based on the calculation of delta K (ΔK), which analyzes the variation of the logarithmic probability LnP(D) across different K values. The Structure Harvester tool was employed to identify the ideal K, avoiding both underestimation and overestimation of clusters.

After defining K, the simulation with the lowest LnP(D) value among the repetitions was selected to ensure data consistency. Finally, the dissimilarity matrix was imported into R, where cluster analyses were performed using the unweighted pair group method with arithmetic mean (UPGMA). The optimal cutoff point for clustering was established based on the K derived from the STRUCTURE analysis.

Results and discussion

Diversity metrics

Based on the literature search, 30 SSR-EST markers were found to be polymorphic for the studied germplasm, indicating that the selection was effective in capturing the genetic variability of the germplasm. Table 3 shows the genetic diversity metrics for the lines.

Table 3. Data from the microsatellite loci analysis performed using the Genes program for 47 popcorn genotypes, showing estimates for the number of alleles (Na), expected heterozygosity (He), observed heterozygosity (Ho), the polymorphism information content (PIC), and the inbreeding coefficient (F).

Locus	Na	He	Ho	PIC	F
UENF7	3	0.434	0	0.38	1
UENF9	4	0.627	0.106	0.567	0.834
UENF 74	4	0.555	0.085	0.473	0.85
BNLG1331	4	0.549	0.043	0.506	0.923
pbf1	5	0.764	0.064	0.726	0.918
fdx3	4	0.543	0	0.479	1

bip2	4	0.556	0.149	0.507	0.737
umc1031	6	0.786	0.022	0.754	0.973
umc1152	3	0.475	0.026	0.385	0.946
umc2292	4	0.55	0.021	0.504	0.962
umc1252	3	0.365	0.319	0.327	0.136
umc1082	3	0.564	0.022	0.475	0.962
umc1448	4	0.636	0.043	0.587	0.934
umc1389	4	0.594	0.07	0.526	0.885
umc1221	2	0.471	0.022	0.36	0.955
umc1155	2	0.481	0.022	0.365	0.956
umc1157	2	0.498	0	0.374	1
umc1176	2	0.485	0	0.367	1
umc1415	2	0.32	0	0.269	1
umc2343	6	0.78	0.098	0.746	0.878
umc1130	2	0.315	0	0.265	1
umc1604	2	0.398	0.548	0.319	-0.364
umc1135	2	0.165	0	0.152	1
umc1587	2	0.482	0	0.366	1
umc2401	2	0.42	0.022	0.332	0.948
umc2246	2	0.317	0.07	0.267	0.785
umc1867	3	0.64	0	0.568	1
umc1506	4	0.484	0.054	0.444	0.891
umc1450	2	0.462	0	0.355	1
umc1948	2	0.144	0.031	0.134	0.789
Average	3.133	0.495	0.061	0.429	0.863

The genetic variability of the 47 popcorn lines, subjected to successive cycles of self-fertilization, was assessed based on several genetic metrics to understand the potential of this sample group. Na correlates positively with the polymorphism rate of the evaluated loci, as highly polymorphic loci exhibit a greater number of allelic variants in the population (Gregolin et al., 2024). The higher the number of unique alleles in a strain or population, the lower the gene flow, and consequently, that strain or population tends to be genetically more distant from others within the same species (Azevedo et al., 2019).

The 47 lines had an Na of 94, distributed among 30 loci, with an average of 3.13 alleles per locus. SSR-EST markers typically exhibit lower polymorphism compared with genomic markers due to the greater conservation of DNA sequences in coding regions (Ferreira Junior et al., 2020). Additionally, self-fertilization tends to reduce the Na in a sample group, as it increases homozygosity and promotes the loss of rare alleles over generations. This process decreases genetic variability by concentrating alleles in specific lines (Eisele et al., 2021). Thus, a lower Na was expected, but the result was considerably higher than what has been reported in other studies using the same type of markers and germplasm (Ferreira Junior et al., 2020). The allelic variability per locus ranged from 2 to 6 alleles, suggesting that some genomic regions are more fixed, while others still show greater polymorphism. The presence of loci with up to 6 alleles is significant, as it indicates that, despite conditions favoring lower diversity, sources of genetic variability persist in the analyzed germplasm. This suggests multiple genetic origins in the formation of the sample group and effective management within the breeding program (Haliloglu et al., 2020).

Heterozygosity is another critical metric for diversity analysis, as it assesses whether a population is in Hardy-Weinberg equilibrium and helps detect factors influencing its genetic structure. He reflects the probability that two randomly chosen alleles in a population are distinct, while Ho is the actual proportion of heterozygous loci in the sample group. For the evaluated lines, the average He was higher than the average Ho, indicating a larger number of homozygous alleles in the sample group and a deficit of heterozygotes relative to expectations under Hardy-Weinberg equilibrium (Gregolin et al., 2024; Tan, 2020). This phenomenon is expected in self-fertilized populations because self-fertilization reduces heterozygosity over generations due to allelic fixation (Rahimi et al., 2023). Thus, the difference between He and Ho reflects inbreeding, characteristic of small or self-fertilized populations (Oliveira et al., 2022).

The PIC is crucial for evaluating the discriminatory power of molecular markers, that is, their ability to detect genetic variation based on genetic relationships among individuals. It estimates the degree of polymorphism for each locus, classified as highly informative (>0.5), moderately informative (0.25-0.5), or non-informative (<0.25) (Botstein et al., 1980; Rodrigues et al., 2023). In this study, the PIC ranged from 0.754 (locus umc1948) to 0.134 (locus UENF9), with an overall average of 0.429. This average is considered moderate

compared to related studies (Galvão et al., 2015; Ferreira Junior et al., 2020; unpublicated, 2023). Because the PIC depends on Na and relative allele frequencies, an identical or divergent PIC for the same microsatellite locus may arise across studies (unpublicated, 2023). Loci with a higher PIC are more informative and valuable for selection, enabling better discrimination between lines. However, loci with a lower PIC have reduced discriminatory power, potentially limiting selection progress (Gregolin et al., 2024).

F quantifies the degree of homozygosity in a population. In breeding programs aimed at developing lines, F infers homozygosity levels and verifies line purity (Galvão et al., 2015). Highly inbred lines are desirable for fixing favorable traits and creating pure lines as bases for hybrid crosses. Crossing inbred lines maximizes heterosis, resulting in hybrid vigor: The descendants from the first generation outperform the parental lines (Rajan et al., 2023). This occurs due to increased heterozygosity in hybrids, masking recessive deleterious alleles fixed during inbreeding. Greater genetic distance between parents enhances hybrid vigor. Thus, generating inbred lines is critical for maximizing crop performance, alongside molecular characterization of genotypes. F ranges from -1 to +1, where values near zero indicate random crossing; negative values denote excess heterozygosity; and positive values reflect high inbreeding (Ferreira Junior et al., 2020). The observed F ranged from -0.364 to 1.000, indicating medium to high homozygosity among the accessions across the investigated loci (Azevedo et al., 2019). The average F was 0.863, indicating a high degree of inbreeding in the sample group. This is attributed to cycles of self-fertilization (Eisele et al., 2021). Despite the high inbreeding, loci with greater polymorphism—supported by the moderate PIC—suggest that there is still genetic variability in the genome of the lines.

In summary, the genetic metrics indicate that, although there is a high degree of inbreeding and reduced heterozygosity among the 47 popcorn lines, there are still loci with significant genetic diversity. The strategic use of crosses between genetically divergent lines can help restore genetic variability and improve genetic progress in subsequent generations. Therefore, the assessment of genetic variability is critical to the success of genetic improvement programs, enabling the development of hybrids with superior traits and greater adaptability to different environmental conditions.

Cluster analysis

The distance matrix based on genetic dissimilarity can be converted into a dendrogram using the UPGMA. In this process, the genetic distances between populations or individuals are grouped hierarchically. The consistency of the clustering method was evaluated using the cophenetic correlation coefficient (CCC), which assesses how well a dendrogram faithfully represents the original genetic distances from the dissimilarity matrix (Vittorazzi et al., 2018). The dendrogram in Figure 1, generated from the UI, has a CCC of 0.88. Thus, the dendrogram has a high correlation with the dissimilarity matrix—specifically, 88% of the original genetic variations are well represented by the hierarchical structure of the dendrogram (Oliveira et al., 2022; Vianna et al., 2019).

The dendrogram shows six distinct groups, as inferred by Bayesian analysis (Gregolin et al., 2024). This separation indicates significant genetic differentiation among the 47 popcorn lines, confirming the descriptive analysis of the loci. The formation of groups is related to the extent to which the genotypes share the analyzed genomic regions (Oliveira et al., 2022).

Group I consists of line L653, descended from the ARZM 13 050 population, developed at the Centro Internacional para el Mejoramiento de Maíz y Trigo (CIMMYT) (International Maize and Wheat Improvement Center) in Argentina. Group II comprises line L274, which has its genealogy traced back to the PARA 172 population, originating from the CIMMYT in Paraguay (Saluci et al., 2020). The fact that these groups are each composed of a single line, forming isolated groups, indicates that these lines exhibit greater genetic dissimilarity compared with the other evaluated lines.

Group III is the largest group, representing 77.5% of the evaluated sample. This group includes lines from diverse origins and reflects greater genetic diversity, as it contains lines derived from different ancestries, originating from various geographic locations, and developed by different institutions. The lines in this group include seven originating from the North American landrace— the South American variety (SAM) population, which were developed in the United States from the combination of North and South American germplasms, resulting in hybrid genetic material. These lines are closely grouped in the dendrogram, suggesting that they share common alleles. Additionally, five lines originating from the Barão de Viçosa population, developed at the Federal University of Viçosa (UFV) in Brazil, are also closely grouped, showing a genetic relationship among them due to their common origin. Two other lines, belonging to the Viçosa-Viçosa population, also

developed at UFV, are similarly grouped. The group also includes seven lines originating from the IAC 125 population, generated from the intervarietal hybrid IAC 125, developed by the Agronomic Institute of Campinas (IAC) in São Paulo State, Brazil, which are very close in the dendrogram, reflecting greater genetic homogeneity among them. Other lines include two from the UENF 14 population, developed by the State University of the North Fluminense Darcy Ribeiro (UENF) through cycles of intrapopulation recurrent selection, as well as two lines from the RS 20 population, developed at the Agricultural Research Institute (IPAGRO) in Rio Grande do Sul State, Brazil. Other lines come from the State University of Maringá (UEM) Paraná State, Brazil, such as the PR 023 population, and from the CIMMYT in Uruguay.

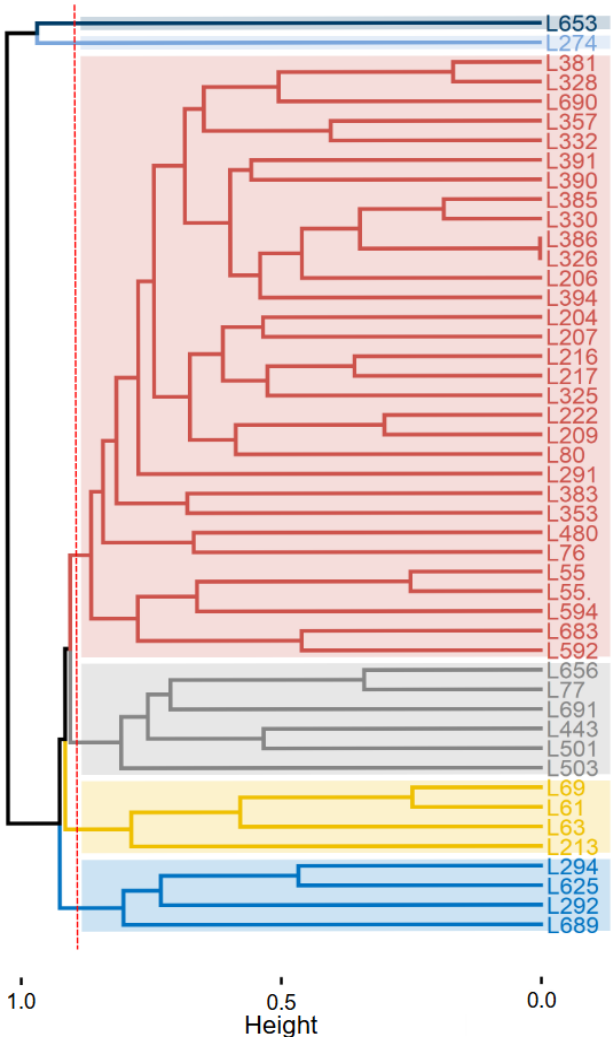

Figure 1. The unweighted pair group method with arithmetic mean dendrogram based on the analysis of 47 popcorn genotypes and 30 loci related to grain yield characteristics, considering the genetic distance calculated using the unweighted index (Cruz, 2013) (cophenetic correlation coefficient = 0.88). After determining the optimal number of populations based on population analysis in the Structure program (K = 6), the dendrogram was cut at 0.8.

Group IV includes line L656, descended from the ARZM 13 050 population, also developed at CIMMYT in Argentina. It also includes line L77, originating from the Viçosa-Viçosa population, developed at UFV; line L691, from the UENF-14 population; line L443, derived from the parental BOYA 462, developed at the CIMMYT in Colombia; and lines L501 and L503, originating from the PA 170 ROXO population, developed at the CIMMYT in Paraguay. The genetic similarity among them suggests that, despite their distinct origins, they share characteristics that result in genetic proximity within the group.

Group V consists of four lines, three of which are derived from Embrapa/Brazil and originate from the Angela variety: L69, L61, and L63. Additionally, the group includes line L213, also derived from the intervarietal hybrid IAC 125, developed by the Agronomic Institute of Campinas in São Paulo State, Brazil.

Finally, Group VI is also composed of four lines, including L294 and L292, descended from URUG 298, originating from CIMMYT in Uruguay; line L625, derived from the parental PA 091 and developed by IPAGRO

in Rio Grande do Sul State, Brazil; and the line L689, originating from the UENF-14 population, developed at UENF in Rio de Janeiro State, Brazil.

The fact that lines from the same genealogy are not separated into the same groups may reflect the number of markers or which markers were used for evaluation. However, it is also important to consider that the high number of closely grouped genotypes may be related to the reduced genetic base of popcorn maize (Almeida et al., 2024). Nevertheless, the analysis of the genetic groups revealed significant diversity among the popcorn maize lines, with different geographic and institutional origins, reflecting the complexity of the genetic improvement process.

The interaction between lines from different regions and countries shows how different selection contexts can result in genetic materials with common or differentiated characteristics. Lines originating from similar selection processes demonstrate genetic proximity, while lines from different geographic origins and improvement processes tend to exhibit genetic variations that influence their phenotypic characteristics and performance. The genetic diversity observed in the analyzed groups suggests that, over time, breeding programs have adapted these lines to local needs and conditions, resulting in a valuable genetic base for future advances in popcorn maize cultivation.

The genetic variability observed in the groups can be highly beneficial for the Popcorn Genetic Improvement Program at UENF. The presence of lines with different origins and selection histories provides a broad range of genetic material for crosses. Thus, by combining genetically distant lines, it is possible to enhance heterosis effects, creating more productive hybrids. Therefore, the genetic diversity found in this study constitutes a solid foundation for generating strong hybrids, particularly based on molecular data, which can infer the prediction of hybrid performance (Yu et al., 2024).

Conclusion

Analysis of genetic diversity metrics and clustering provided an accurate understanding of the genetic structure of the 47 popcorn maize lines evaluated based on the 30 SSR-EST markers. This allowed for the identification of patterns of similarity and dissimilarity among the lines, highlighting the existing genetic variability between them. The observed genetic diversity is crucial for genetic improvement, because it offers opportunities to exploit heterosis, with the aim of developing new cultivars with high agronomic performance.

Data availability

Dataset available on request from the authors.

References

Almeida, R. N., Vivas, M., Souza, Y. P., Saluci, J. C. G., Vasconcelos, L. C., & Amaral Júnior, A. T. (2024). Proposal of a super trait for increased grain yield, popping expansion, and ear rot resistance in popcorn. *Crop Science*, *64*(5), 2789-2800. https://doi.org/10.1002/csc2.21308

Azevedo, C. D. O., Rodrigues, R., & Sudré, C. P. (2019). Microsatellites for detecting inconsistencies in *Capsicum* cultivars registration in Brazilian database: more than meets the eye. *Horticultura Brasileira*, *37*(3), 285-293. https://doi.org/10.1590/S0102-053620190306

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314-331.

Cabral, P. D. S., Amaral Júnior, A. T., Freitas, I. L. J., Ribeiro, R. M., & Silva, T. R. C. (2016). Relação de causa e efeito de caracteres quantitativos sobre a capacidade de expansão do grão em milho-pipoca. *Revista Ciência Agronômica*, *47*(1), 108-117. https://doi.org/10.5935/1806-6690.20160013

Coan, M. M. D., Pinto, R. J. B., Kuki, M. C., Amaral Júnior, A. T., Figueiredo, A. S. T., Scapim, C. A., & Warburton, M. (2019). Inheritance study for popping expansion in popcorn vs. flint corn genotypes. *Agronomy Journal*, *111(5)*, 2174-2183. https://doi.org/10.2134/agronj2019.04.0295

Cruz, C. D. (2013). GENES: Software para análise de dados em estatística experimental e em genética quantitativa. *Acta Scientiarum. Agronomy*, *35*(3), 271-276. https://doi.org/10.4025/actasciagron.v35i3.21251

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11-15.

Eisele, T. G., Lazzari, D., Silva, T. A., Pinto, R. J. B., Matsuzaki, R. A., Maioli, M. F. S. D., Alves, A. V., & Amaral Júnior, A. T. (2021). Combining ability and genetic divergence among tropical maize inbred lines using SSR markers. *Acta Scientiarum. Agronomy*, *43*(1), 1-12. https://doi.org/10.4025/actasciagron.v43i1.53317

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ferreira Junior, J. A., Goncalves, G. M. B., Crevelari, J. A., Souza, Y. P., Goncalves, V. M. L., Viana, F. N., Vettorazzi, J. C. F., & Pereira, M. G. (2020). Genetic progress in 20 years of reciprocal recurrent selection in maize. *Functional Plant Breeding Journal*, *2*(1), 13-23. http://dx.doi.org/10.35418/2526-4117/v2n1a2

Galvão, K. S. C., Ramos, H. C. C., Santos, P. H. A. D., Entringer, G. C., Vettorazzi, J. C. F., & Pereira, M. G. (2015). Functional molecular markers (EST-SSR) in the full-sib reciprocal recurrent selection program of maize (*Zea mays* L.). *Genetics and Molecular Research*, *14*(3), 7344-7355. https://doi.org/10.4238/2015.July.3.10

Gregolin, F. S., Vieira, H. D., Santa-Catarina, R., & Viana, A. P. (2024). Molecular approach in elite parents of papaya with a focus on seed physiological quality. *Revista Brasileira de Fruticultura*, *46*, 1-18. https://doi.org/10.1590/0100-29452024637

Guimarães, A. G., Amaral Júnior, A. T., Almeida Filho, J. E., Pena, G. F., Vittorazzi, C., & Pereira, M. G. (2018). Population structure and impact of recurrent selection on popcorn using EST-SSR markers. *Acta Scientiarum. Agronomy*, *40*(1), 1-10. https://doi.org/10.4025/actasciagron.v40i1.35218

Guimarães, L. J. M., Santana, D. P., & Vasconcellos, J. H. (2025). *Milho-Pipoca*. https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/milho/producao/sistemas-diferenciais-de-cultivo/milho-pipoca

Haliloglu, K., Turkoglu, A., Tan, M., & Poczai, P. (2022). SSR-based molecular identification and population structure analysis for forage pea (*Pisum sativum* var. *arvense* L.) landraces. *Genes*, *13*(6), 1-16. https://doi.org/10.3390/genes13061086

Kara, D., & Orhan, E. (2023). Tolerance evaluation and genetic relationship analysis among some economically important chestnut cultivars in Türkiye using drought-associated SSR and EST-SSR markers. *Scientific Reports*, *13*(1), 1-12. https://doi.org/10.1038/s41598-023-47951-7

Liu, K., & Muse, S. V. (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128-2129. https://doi.org/10.1093/bioinformatics/bti282

Kaur, S., Rakshit, S., Choudhary, M., Das, A. K., & Kumar, R. R. (2021). Meta-analysis of QTLs associated with popping traits in maize (*Zea mays* L.). *PLoS ONE*, *16*(8), 1-14. https://doi.org/10.1371/journal.pone.0256389

Mascarenhas, K. (2019). *Conheça as principais variedades de milho usadas na alimentação humana e animal*. Fapemig. https://ufla.br/noticias/pesquisa/12714-conheca-as-principais-variedades-de-milho-usados-na-alimentacao-humana-e-animal.

Muzdalifah, I., Sunarti, S., Andayani, N. N., & Pabendon, M. B. (2020). Genetic characterization of popcorn (*Zea mays* everta) genotyping based on SSR (simple sequence repeats) markers. *IOP Conference Series: Earth and Environmental Science*, *484*, 1-10. https://doi.org/10.1088/1755-1315/484/1/012013

Oliveira, J. A. V. S., Santos, E. A., Viana, A. P., Walter, F. H. B., & Ribeiro, R. M. (2022). Genetic-molecular characterization in guava full-sib progeny. *Bragantia*, *81*, 1-16. https://doi.org/10.1590/1678-4499.20210267

Parsons, L., Ren, Y., Yobi, A., Hurst, P., Angelovici, R., Rodriguez, O., & Holding, D. R. (2020). Production and selection of quality protein popcorn hybrids using a novel ranking system and combining ability estimates. *Frontiers in Plant Science*, *11*(698), 1-20. https://doi.org/10.3389/fpls.2020.00698

Peixoto, L. A., Bhering, L. L., & Cruz, C. D. (2022). *Seleção genômica aplicada ao melhoramento genético*. Editora UFV.

Pinto, V. B., Vidigal, P. M. P., Dal-Bianco, M., Almeida-Silva, F., Venancio, T. M., & Viana, J. M. S. (2023). Transcriptome-based strategies for identifying aluminum tolerance genes in popcorn (*Zea mays* L. var. everta). *Scientific Reports*, *13*, 1-9. https://doi.org/10.1038/s41598-023-46810-9

Rahimi, A., Kahrizi, D., Mirmoayedi, A., Zarei, L., & Jamali, S. (2023). Genetic characterizations of the Iranian honey bee (*Apis mellifera meda Skorikov* 1929) populations using the microsatellite DNA markers. *Biochemical Genetics*, *61*, 2293-2317. https://doi.org/10.1007/s10528-023-10368-y

Rajan, N., Debnath, S., Perveen, K., Khan, F., Pandey, B., Srivastava, A., Khanam, M. N., Subramaniyan, V., Kumarasamy, V., Paul, P. J., & Lal, M. (2023). Optimizing hybrid vigor: A comprehensive analysis of

genetic distance and heterosis in eggplant landraces. *Frontiers in Plant Science*, *14*, 1-13. https://doi.org/10.3389/fpls.2023.1238870

Ramos, H. C. C., Pereira, M. G., Pereira, T. N. S., Barros, G. B. A., & Ferreguetti, G. A. (2014). Genetic-molecular characterization of backcross generations for sexual conversion in papaya (*Carica papaya* L.). *Genetics and Molecular Research*, *13*(4), 10367-10381. https://doi.org/10.4238/2014.December.4.32

Rodrigues, A. S., Ramos, H. C. C., Miranda, D. P., Pirovani, A. A. V., Santa-Catarina, R., Duarte, R. P., Braga, C. S., Bohry, D., & Pereira, M. G. (2023). Intra-genotypic variability in elite parent lines of papaya. *Anais da Academia Brasileira de Ciências*, *95*(1), 1-11. https://doi.org/10.1590/0001-3765202320220468

Saluci, J. C. G., Vivas, M., Almeida, R. N., Dutra, I. P., Carlos, M. C., Amaral Junior, A. T., & Scapim, C. A. (2020). Potential of popcorn S4 lines for resistance to Southern Corn Leaf Blight. *Functional Plant Breeding Journal*, *2*(2), 1-9. https://doi.org/10.35418/2526-4117/v2n2a6

Tan, Y.-D. (2020). Recombination disequilibrium in ideal and natural populations. *Genomics*, *112*(6), 3943-3950. https://doi.org/10.1016/j.ygeno.2020.06.034

Vianna, L. S., Pereira, T. N. S., Santos, E. A., Viana, A. P., Pereira, M. G., Ramos, H. C. C., & Rossi, A. A. B. (2019). ISSR and SSR Markers for determining genetic relationships among three wild species of passiflora. *Genetics and Molecular Research*, *18*(1), 1-10. https://doi.org/10.4238/gmr18040

Vittorazzi, C., Júnior, A. T. Amaral., Guimarães, A. G., Silva, F.H.L., Pena, G. F., Daher, R. F., Gerhardt, I. F. S., Oliveira, G. H. F., Santos, P. H. A. D., Souza, Y. P., Kamphorst, S. H., & Lima, V. J. (2018). Evaluation of genetic variability to form heterotic groups in popcorn. *Genetics and Molecular Research*, *17*(3), 1-17. https://doi.org/10.4238/gmr18083

Yu, G., Li, F., Wang, X., Zhang, Y., Zhou, K., Yang, W., Guan, X., Zhang, X., Xu, C., & Xu, Y. (2024). Enhancing across-population genomic prediction for maize hybrids. *Plants*, *13*(21), 1-13. https://doi.org/10.3390/plants13213105

Wang, J., & Hu, X. (2021). Research on corn production efficiency and influencing factors of typical farms: Based on data from 12 corn-producing countries from 2012 to 2019. *PLoS ONE*, *16*(7), 1-17. https://doi.org/10.1371/journal.pone.0254423

Zulkadir, G., & İdikut, L. (2021). Determination of popping traits and grain quality of landraces popcorn populations. *Journal of Food Science and Technology*, *58*, 1302-1312. https://doi.org/10.1007/s13197-020-04639-4