

SNP genotyping for fast and consistent clustering of maize inbred lines into heterotic groups

Lander Santos de Oliveira^{1*}, Ivan Schuster¹, Evandro Novaes² and Welison Andrade Pereira²

Crop Breeding and Applied Biotechnology
21(1): e367121110, 2021
Brazilian Society of Plant Breeding.
Printed in Brazil
<http://dx.doi.org/10.1590/1984-70332021v21n1a10>

Abstract: *Advances in genotyping technologies have transformed the way breeding programs manage their genetic resources. The identification of single nucleotide polymorphisms (SNPs) can improve understanding of the genetic diversity of maize (*Zea mays*) inbred lines and their classification into heterotic groups, which is useful in determining certain crosses to obtain hybrids with higher yield performance. The genetic diversity of 293 inbred lines was investigated with 5252 SNPs with minor allele frequency (MAF)>5%. There was an average of 525 SNPs per chromosome. Polymorphism information content (PIC) averaged 0.297. The unweighted pair group method with arithmetic mean analysis (UPGMA) and principal component analysis (PCA) based on the genetic distance matrix revealed four similar clusters and high cophenetic correlation coefficients (0.953 and 0.863, respectively). The results showed consistency between genetic distance-based grouping and the heterotic groups previously established using pedigree and topcross information for the inbred lines studied.*

Keywords: *genetic diversity, single nucleotide polymorphism, *Zea mays**


INTRODUCTION

Maize (*Zea mays*) is one of the main crops worldwide, widely used both for human and animal consumption. Brazil is the third largest maize producer in the world, with estimated production of 100 million metric tons in the 2018/19 crop year (CONAB 2020). To meet growing demand for maize worldwide, maize breeding programs have developed high-yielding inbred lines adapted to different environments, which are used as parents in hybrid production (Smith et al. 2017). The dramatic increase in the number of inbred lines produced by these programs has made evaluation of the phenotypic performance of all possible hybrid combinations impractical. Thus, classification of inbred lines into heterotic groups has had to be performed in a different way, such as through molecular markers, which has become routine practice in maize breeding programs (Andorf et al. 2019).

Heterotic groups are defined as sets of related genotypes from the same or different populations, based on combining abilities (Reif et al. 2005). Genotypes from the same group show similar combining ability and, when crossed with genotypes from another group, exhibit heterosis. Therefore, high yielding hybrids can be developed through crosses between inbred lines from different heterotic groups (Souza Júnior 2011).

***Corresponding author:**

E-mail: lander_sto@hotmail.com

 ORCID: 0000-0001-7508-4243

Received: 29 February 2021

Accepted: 21 March 2021

Published: 10 May 2021

¹ LongPing High-Tech, Departamento de Pesquisa e Desenvolvimento, 14140-000, Cravinhos, SP, Brazil

² Universidade Federal de Lavras, Departamento de Biologia, 37200-000, Lavras, MG, Brazil

Furthermore, genetic diversity studies are commonly carried out in maize breeding programs in an effort to classify lines into heterotic groups (Wu et al. 2016, Leng et al. 2019, Silva et al. 2020) because parents with high *per se* performance and genetic divergence from each other are important requirements for manifestation of heterosis (Prasad and Singh 1986). Classification of inbreds into heterotic groups using information on genetic diversity facilitates directed crosses between inbreds from contrasting groups and reduces the number of hybrid crosses made in a breeding program, thus increasing the efficiency of the program and leading to accelerated genetic gains from selection (Reif et al. 2005). Thus, systematized knowledge of maize genetic resources has become necessary to better evaluate their diversity.

Genetic diversity studies to classify inbred lines into groups can be performed using morphological characteristics and analysis of combining ability based on diallel and line \times tester information. These designs involve field trials, which provide the actual performance information with regards to *per se* performance and combining ability, and hence heterotic responses. However, they are costly and require large fields, hand pollination, and detasseling labor. Consequently, the number of hybrids evaluated in such genetic studies is usually restricted (Fernandes et al. 2015, Wu et al. 2016, Kulka et al. 2018). Genotypes can also be allocated into groups based on their genealogy. Although this method is simple, it requires detailed pedigree information, which is not always available (Lee et al. 2007, Adu et al. 2019b, Leng et al. 2019). Therefore, the use of molecular markers has become the best method to make inferences regarding genetic diversity among genotypes (Muhammad et al. 2017, Scherlosky et al. 2018, Adu et al. 2019a, Silva et al. 2020), without the need for making numerous crosses and evaluating hybrids in the field (Andorf et al. 2019). Some molecular markers are highly polymorphic and independent of environmental effects and plant physiological stages. These qualities are advantageous for selecting more divergent parents that will give rise to populations with high variability and adaptability to environments (Govindaraj et al. 2015, Nadeem et al. 2018).

Single nucleotide polymorphisms (SNPs) are the most abundant source of variation in genomes, showing dense coverage across genomes compared to other types of molecular markers. This increases the likelihood of some SNPs being associated with genes of interest, which can contribute to improved accuracy of evaluation of genetic diversity in breeding programs. New-generation sequencing technologies have offered high-throughput and reduced costs, with automated platforms appropriate to breeding programs requirements, leading to wide use of SNPs in studies (Rasheed et al. 2017, Guo et al. 2019).

Molecular information allows estimation of genetic similarity among individuals in terms of identity by descent (IBD) or identity by state (IBS) alleles (Messmer et al. 1993). The IBD between a pair of individuals is the probability that an allele in a given locus of a genotype and an allele from the same locus of another genotype are copies from a common ancestor (Cox et al. 1985). In contrast, IBS is the genotypic similarity of alleles alike in "state", that is, indistinguishable by their effects and ancestry. The estimates of genetic similarity based on molecular markers reveal the proportion of IBS alleles, regardless of whether their identity is caused by IBD or IBS alleles (Messmer et al. 1993).

The objective of this study was to examine the accuracy of the methods for clustering of maize inbred lines using genetic dissimilarity information obtained from SNP data compared to the heterotic groups previously established with pedigree and combining ability information. The results of this study may be useful to maize breeders interested in using molecular markers for classifying inbred lines into heterotic groups for planning crosses to accelerate genetic gains from selection.

MATERIAL AND METHODS

This study comprised 293 maize inbred lines developed by the breeding program of LongPing High-Tech (LPHT), Brazil. The lines are important to the company's breeding program and belong to four heterotic groups of tropical and temperate genetic backgrounds (Table 1), previously defined based on pedigree and topcross information (LPHT proprietary information not disclosed). Most of the inbred lines are doubled haploid, and a few were developed by selfing and are at least S_{10} .

This study was conducted in the biotechnology laboratory of LPHT in Cravinhos, SP, Brazil. The DNA of each maize line was extracted from leaf samples according to the Fast ID Genomic DNA Extraction Kit protocol (Fast ID NA, Inc., Fairfield, IA, USA). DNA quantity and quality were checked by fluorimetry using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) before genotyping. The SNP genotyping was performed using the MaizeSNP50 BeadChip according to the Infinium HD Assay Ultra Protocol Guide (Illumina, Inc., San Diego, CA, USA).

Table 1. Comparison of the number of inbred lines in previous heterotic group classification to the number of lines found in the clusters formed in the dendrogram and PCA

Original heterotic group			Dendrogram			PCA		
Heter- otic group	Genetic back- ground	Num- ber of lines	Dendro- gram cluster	Num- ber of lines	Lines from other groups	PCA clus- ter	Num- ber of lines	Lines from other groups
G3	Temperate	4	Yellow	11	4 from G3; 6 from G1; 1 from G4	Yel- low	17	4 from G3; 11 from G1; 2 from G4
G1	Tropical	56	Green	52	50 from G1; 1 from G2; 1 from G4	Green	47	50 from G1; 1 from G2; 1 from G4
G4	Tropical	154	Violet	151	151 from G4	Violet	150	151 from G4
G2	Tropical	79	Red	79	78 from G2; 1 from G4	Red	79	78 from G2; 1 from G4

A total of 6231 SNPs were available for this study. SNP markers with minor allele frequency (MAF) less than 5% and with more than 5% missing data did not pass quality control and were not used in the analysis. Therefore, the diversity analyses were carried out with a total of 5252 polymorphic SNPs.

Genetic information on SNP markers was estimated by MAF and polymorphism information content (PIC) parameters using PowerMarker version 3.25 (Liu and Muse 2005). The physical distribution of SNPs across the maize chromosomes was determined using the web-based software PhenoGram (Center for System Genomics, Pennsylvania State University; <http://visualization.ritchielab.psu.edu/>).

Genetic distances were estimated by the complement of identity by state (1 – IBS) with TASSEL 5.0 software (Bradbury et al. 2007). Based on the distance matrix, the maize lines were clustered by UPGMA analysis and principal component analysis (PCA) using R 3.5.1 (R Core Team 2018) with the *ggplot2* version 3.1.0 (Wickham 2016) and *ape* version 5.3 (Paradis and Schliep 2018) packages, respectively. PCA plots were generated with DataWarrior 5.0.0 (Sander et al. 2015). The cophenetic correlation coefficient was calculated, and Mantel's test was performed to check cluster analysis fitness to the genetic distance matrix using the *stats* version 3.5.1 (R Core Team 2018) and *ade4* version 1.7-13 (Dray and Dufour 2007) packages.

RESULTS AND DISCUSSION

The SNPs were distributed across the maize genome. On average, there were 525 SNPs in each chromosome, ranging from 327 in chromosome 10 to 926 in chromosome 1 (Figure 1A). This distribution of SNPs throughout the genome was similar to other studies in maize using this type of marker (Li et al. 2018, Guo et al. 2019, Leng et al. 2019, Silva et al. 2020). Unlike primary molecular marker systems, high-density marker genotyping allows simultaneous analysis of markers widely distributed throughout the genome. SNP markers provide higher genomic coverage than other available markers, such as RFLP, AFLP, and SSR, among others (Xu et al. 2017, Scherlosky et al. 2018).

The magnitude of informativeness of the marker depends on its degree of polymorphism, which is reflected in the genetic diversity among the genotypes under study (Chesnokov and Artemyeva 2015). In this study, PIC ranged from 0.092 to 0.375, with a mean of 0.297, whereas MAF ranged from 0.051 to 0.5, with a mean of 0.284 (Figure 1B), indicating that the 5252 SNPs used across the genomes of 293 inbred lines were highly informative. Liu et al. (2015) and Li et al. (2018) found similar PIC range values when studying genetic properties of Chinese maize germplasm using SNP marker data. Wu et al. (2016) found similar values of PIC in analysis of tropical and temperate maize inbred lines from CIMMYT. PIC values ranging from 0.25 to 0.5 indicate that multiallelic markers are moderately informative (Botstein et al. 1980). Considering the biallelic nature of SNPs, in which the maximum value of PIC is 0.375, we can consider PIC values in the higher quartile, such as those found in our study, highly informative. Based on these criteria, 65.6% of the markers in this study were highly informative (Figure 1B). Regarding MAF, which is used to quantify the degree of genetic differentiation in the population (Li et al. 2018), the average value in this study was higher than that found by Li et al. (2018) and Liu et al. (2015). Higher MAF are usually preferred in order to increase the average allelic differentiation (Xu et al. 2017).

Based on the 5252 polymorphic SNPs, a genetic distance matrix was built among all pairs of inbred line, ranging from 0 to 0.491, from most closely related to most distant, respectively, with an average of 0.375. Figure 1C shows the genetic distance frequencies among all pair of lines. The distance range was greater than the ranges estimated by Ertiro

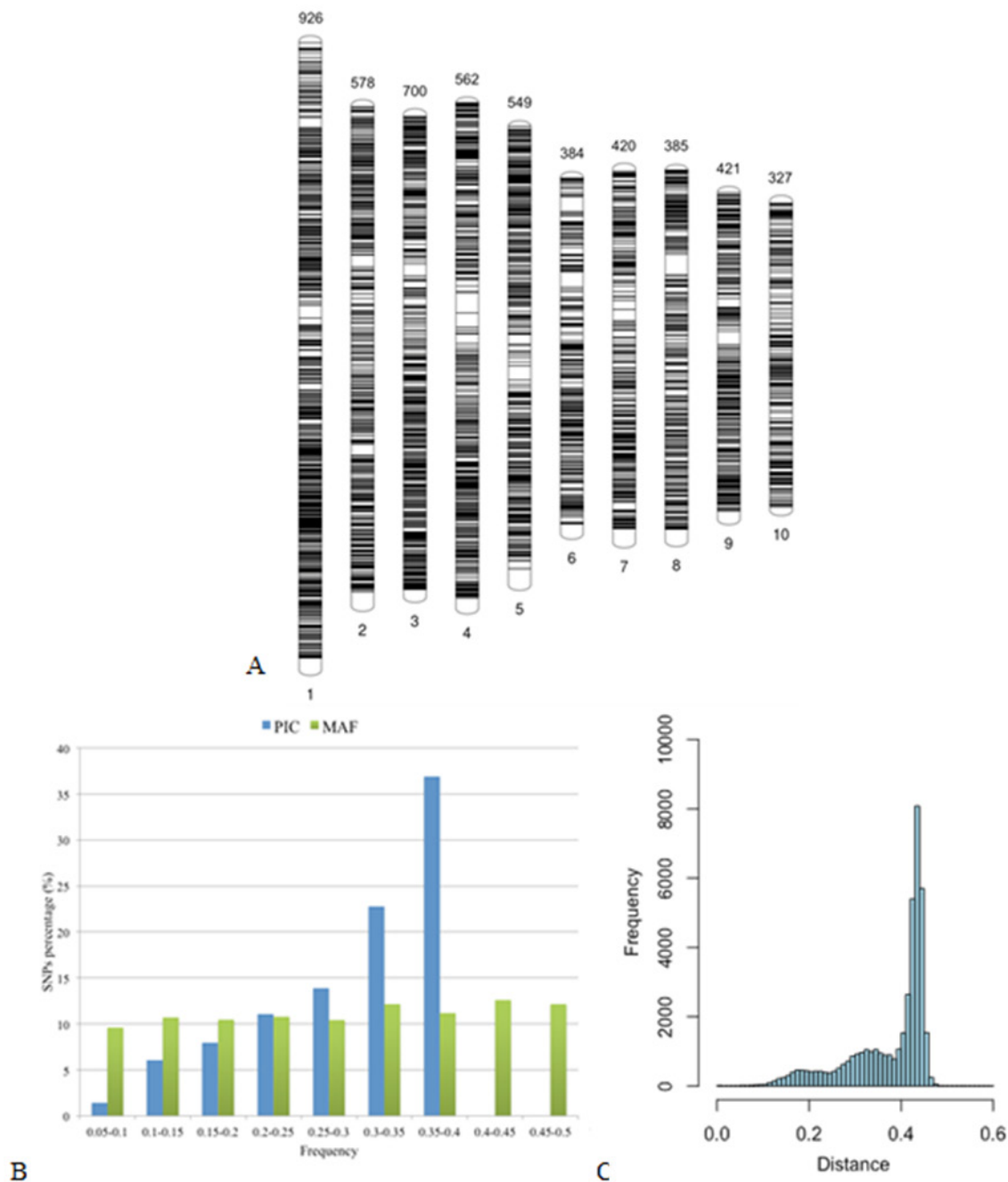


Figure 1. Physical distribution of the 5252 SNPs across the maize genome (A). Distribution of polymorphism information content (PIC) and minor allele frequency (MAF) for 5252 SNPs in a population of 293 maize inbred lines (B). Frequency of genetic distance values among pairs of inbred lines (C).

et al. (2017) and Silva et al. (2020), in which distances of tropical maize lines based on IBS genetic similarity ranged from 0.011 to 0.346 (average of 0.313) and 0.003 to 0.253 (average of 0.193), respectively. It is important to highlight that the lower average distances found by Ertiro et al. (2017) may be related to the substantially larger number of SNPs used in their study (220.878) and number of genotypes similar to our study. Silva et al. (2020) analyzed more lines (1.041)

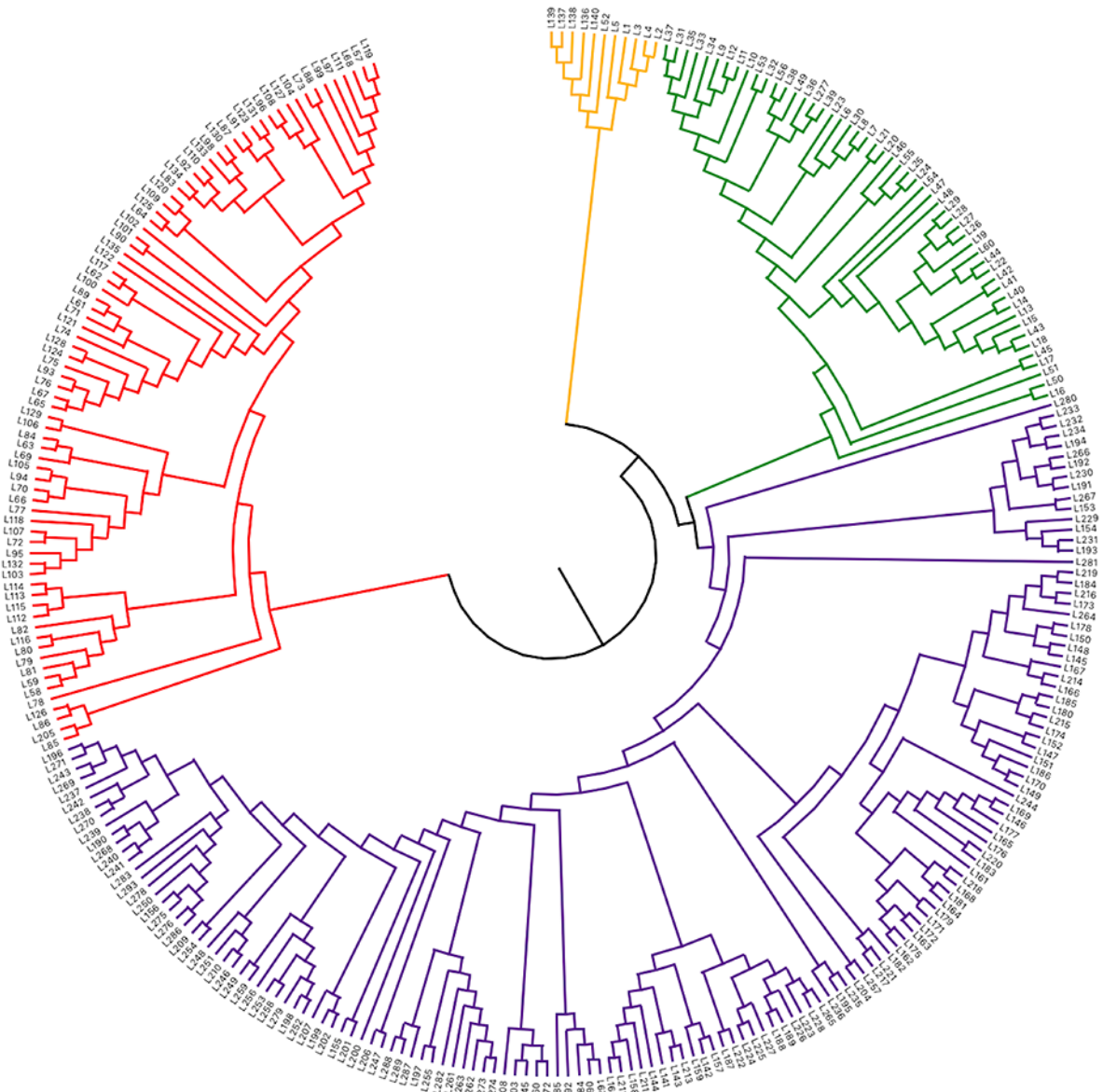


Figure 2. Dendrogram based on molecular data displaying four clusters and their respective number of inbred lines (in parentheses): yellow (11), green (52), violet (151), and red (79). Compared to previously established heterotic groups based on pedigree and top-cross information, 10 of the individuals had a classification different from what was expected.

in their report. Among the 42.778 estimated distances, the smallest were obtained between seven pairs of lines: L61 × L89, L57 × L119, L98 × L130, L124 × L128, L154 × L193, L154 × L231, and L173 × 216. These materials are highly related considering their pedigree (LPHT internal information). In contrast, the greatest distance was obtained between lines L136 × L159, previously classified as temperate and tropical, respectively. All the other pairs of lines with the smallest genetic distances agreed to the tropical groups. The average of the distance estimates in our study were higher than the averages in the studies cited (Ertiro et al. 2017, Silva et al. 2020), indicating that there is still genetic variability among the lines, even though we evaluated elite maize genotypes originating from breeding programs, which in theory could have led to narrowing of variability (Scherlosky et al. 2018).

The UPGMA cluster analysis based on the genetic distance matrix formed 4 distinct clusters, as shown in Figure 2. The dendrogram clusters were separated so that the number of lines arranged together was closest to the number of lines of the previously known heterotic groups. The significant cophenetic correlation coefficient ($r = 0.953$; $P < .0001$; 10,000 permutations) indicated that the cluster analysis well fit the genetic distance matrix on which it was based, according to Mohammadi and Prasanna (2003). Beckett et al. (2017) also affirm that an accurate dendrogram is important to help breeders classify their germplasm. Fernandes et al. (2015) and Nikolić et al. (2015) found positive cophenetic values when making inferences regarding genetic diversity using microsatellites ($r = 0.59$ and $r = 0.80$, respectively). In our study using SNP data, the cophenetic coefficient was higher ($r = 0.953$).

The population of this study was previously organized into four heterotic groups (G1, G2, G3, and G4) based on their genealogy and breeding history information (data not shown). The dendrogram revealed four distinct clusters (yellow, green, violet, and red), with most of the inbred lines grouped as expected. However, not all the clusters coincided with the known heterotic groups (Table 1). Although the genealogy of the inbred lines was not disclosed, the lines were obtained from breeding populations resulting from crossing tropical and temperate germplasm (LPHT confidential information), and they must therefore have a substantial amount of temperate germplasm in their genetic composition. For example, the yellow cluster had a total of 11 inbred lines, 4 of which consisted of G3 individuals, while 7 other lines were from other groups (6 from G1 and 1 from G4). Since the G3 group contains all temperate maize lines, the seven

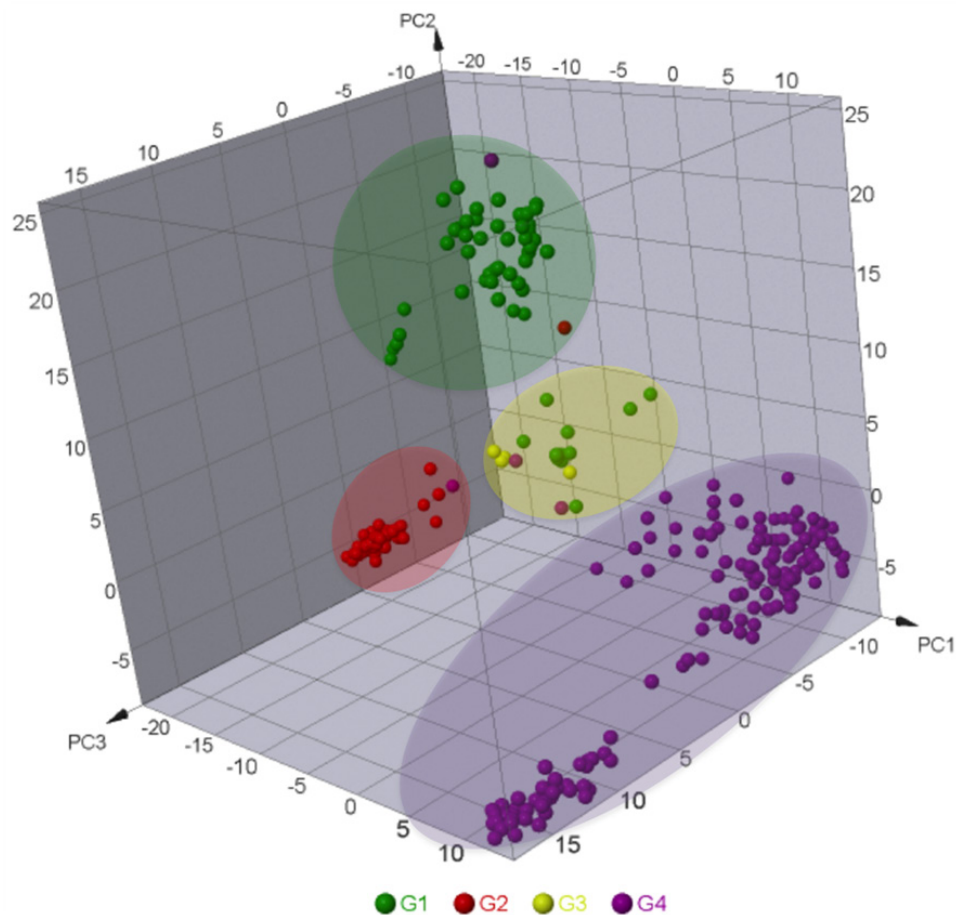


Figure 3. Principal component analysis (PCA) based on genetic distance using the first three PCs. Colored dots refer to heterotic groups previously classified by pedigree and topcross information (G1, G2, G3, G4). The shaded areas highlighted (yellow, green, violet, red) refer to the clusters formed from molecular data. Each cluster contains most of the inbred lines of its respective heterotic group, except for 17 individuals that had classifications different from what was expected. The yellow cluster contains 17 inbred lines; the green, 47; the violet, 150; and the red, 79.

inbred lines assigned to the yellow cluster are likely to have a considerable temperate genetic background. The green cluster consisted of 52 lines, which are mostly from the G1 group, but 2 are from other groups (1 from G2 and 1 from G4). The violet cluster comprised 151 genotypes, including most of the lines from the G4 group, and did not show any lines from other groups allocated to it. Finally, the red cluster had 79 inbred lines, mostly from the G2 group, along with one genotype from the G4 group.

In summary, out of the 293 maize inbred lines analyzed in the present study, 10 (3.4%) received a classification different from the previous heterotic group classification. This shows the importance of associating molecular and conventional breeding for a more accurate genetic diversity analysis (Wu et al. 2016). The inconsistencies found in classification of genotypes using marker data may be due to errors in pedigree information, genetic drift during the process of inbred line development (Nikolić et al. 2015), or labelling errors during storage of the lines.

The PCA plot was built with the first three principal components (PCs) and displayed four clusters (Figure 3). The cophenetic correlation value ($r = 0.863$; $P < .0001$; 10,000 permutations) between the PCA distance matrix and the genetic distance matrix suggests that this clustering method was also reliable, since the lines were clustered in a manner that was consistent with their known pedigree and breeding history information. The first three PCs explained 30.5%, 14.6%, and 7.7%, respectively, of the total variations among the inbred lines. Out of the 293 inbred lines, 17 (5.8%) were classified differently from the previous heterotic group categorization based on pedigree and topcross information. According to this grouping method, we also found more inbred lines (12) that are likely to have a temperate genetic background (Table 1).

The UPGMA dendrogram and PCA grouping of inbred lines was consistent with the known heterotic groups and pedigrees, with very few exceptions. Only 10 inbred lines were not grouped as expected in the dendrogram, and only 17 in the PCA plot. The UPGMA is a hierarchical clustering method that groups lines interactively, starting from the most similar lines (Sokal and Michener 1958). PCA is based on orthogonal (independent) linear combinations (principal components) that extract most of the variation in the genetic distances among lines (Ringnér 2008). Therefore, it is expected that some differences will appear in the results of the methods. However, the fact that different methods can achieve equivalent general conclusions increases the robustness of our inferences. In fact, all 10 lines that were not grouped as expected in the dendrogram are included in the 17 lines not grouped as expected in the PCA plot.

The main goal of this study was to show that grouping maize lines using high-density molecular marker data is a precise methodology, without the need for performing vast combining ability experiments in the field to achieve this classification. The results of the present study support the findings of other genetic diversity studies using SNPs as an important tool for classifying inbred lines into genetically related groups and directing hybrid crosses between individuals of different groups to obtain higher yield performance (Richard et al. 2016, Dari et al. 2018, Silva et al. 2020). In addition, SNPs are useful for quickly determining the genetic relationship of new inbred lines by genotyping and including them in a new cluster analysis (Beckett et al. 2017), which can benefit breeding programs. However, it should be noted that the information provided by genetic distances does not ensure good hybrids, since the parents of hybrids need to show genetic complementarity (Hablak 2019).

Classifying maize lines using SNPs showed high accuracy in accordance with the classification previously performed with pedigree and breeding information. Although most of the genotypes in this study were classified into the four heterotic groups previously established by the company's breeding program, subgroups within each of the groups could be visualized, indicating the abundant genetic variability existing in maize germplasm. In addition, the genotyping data generated in this study can be used in further models of hybrid prediction, allowing more effective identification of hybrids, thus improving breeding efficiency.

CONCLUSION

The results of this study indicated that clustering methods based on genetic diversity estimates using SNP markers offer reliable classification of maize inbred lines into heterotic groups, confirmed by the consistency found between these methods and the methods using pedigree and breeding information. The use of these highly dense markers as a complement to the breeding program provided more detailed information on the heterotic groups identified and allowed enhanced exploitation of genetic variability within the inbred lines.

ACKNOWLEDGMENTS

This study was funded by the company LongPing High-Tech, Brazil.

REFERENCES

- Adu GB, Awuku FJ, Amegbor IK, Haruna A, Manigben KA and Aboyadana PA (2019a) Genetic characterization and population structure of maize populations using SSR markers. **Annals of Agricultural Sciences** **64**: 47-54.
- Adu GB, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ and Gedil M (2019b) Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. **PLoS ONE** **14**: 1-12.
- Andorf C, Beavis WD, Hufford M, Smith S, Suza WP, Wang K, Woodhouse M, Yu J and Lübberstedt T (2019) Technological advances in maize breeding: past, present and future. **Theoretical and Applied Genetics** **132**: 817-849.
- Beckett TJ, Morales AJ, Koehler KL and Rocheford TR (2017) Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. **PLoS ONE** **12**: 1-23.
- Botstein D, White RL, Skolnick, M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics** **32**: 314-331.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics** **23**: 2633-2635.
- Chesnokov YV and Artemyeva AM (2015) Evaluation of the measure of polymorphism information of genetic diversity. **Agricultural Biology** **50**: 571-578.
- CONAB - Companhia nacional de abastecimento (2020) **Série histórica das safras**. Available at: <https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras>. Accessed on February 29, 2020.
- Cox TS, Lookhart GL, Walker DE, Harrell LG, Albers LD and Rodgers DM (1985) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electrophoretic patterns. **Crop Science** **25**: 1058-1063.
- Dari S, MacRobert J, Minnaar-Ontong A and Labuschagne MT (2018) SNP-based genetic diversity among few-branched-1 (Fbr1) maize lines and its relationship with heterosis, combining ability and grain yield of testcross hybrids. **Maydica** **63**: 1-14.
- Dray S and Dufour A (2007) The ade4 Package: Implementing the duality diagram for ecologists. **Journal of Statistical Software** **22**: 1-20.
- Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, Wegary D, Azmach G, Ogugo V, Keno T, Abebe B, Chibsa T and Menkir A (2017) Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. **BMC Genomics** **18**: 1-11.
- Fernandes EH, Schuster I, Scapim CA, Vieira ESN and Coan MMD (2015) Genetic diversity in elite inbred lines of maize and its association with heterosis. **Genetics and Molecular Research** **14**: 6509-6517.
- Govindaraj M, Vetriventhan M and Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. **Genetics Research International** **2015**: 1-14.
- Guo Z, Wang H, Tao J, Ren Y, Xu C, Wu K, Zou C, Zhang J and Xu Y (2019) Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. **Molecular Breeding** **39**: 1-12.
- Hablak S (2019) The concept of allelic and nonallelic mechanism of heterosis. **Biochemistry and Molecular Biology** **3**: 7782.
- Kulka V, Silva TA, Contreras-Soto RI, Maldonado C, Mora F and Scapim CA (2018) Diallel analysis and genetic differentiation of tropical and temperate maize inbred lines. **Crop Breeding and Applied Biotechnology** **18**: 31-38.
- Lee EA, Ash MJ and Good B (2007) Re-examining the relationship between degree of relatedness, genetic effects, and heterosis in Maize. **Crop Science** **47**: 629-635.
- Leng Y, Lv C, Li L, Xiang Y, Xia C, Wei R, Rong T and Lan H (2019) Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). **Molecular Breeding** **39**: 1-19.
- Li T, Qu J, Wang Y, Chang L, He K, Guo D, Zhang X, Xu S and Xue J (2018) Genetic characterization of inbred lines from Shaan A and B groups for identifying loci associated with maize grain yield. **BMC Genetics** **19**: 1-12.
- Liu C, Hao Z, Zhang D, Xie C, Li M, Zhang X, Yong H, Zhang S, Weng J and Li X (2015) Genetic properties of 240 maize inbred lines and identity-by-descent segments revealed by high-density SNP markers. **Molecular Breeding** **35**: 1-12.
- Liu K and Muse SV. (2005) PowerMaker: An integrated analysis environment for genetic marker analysis. **Bioinformatics** **21**: 2128-2129.
- Messmer MM, Melchinger AE, Herrmann RG and Boppenmaier J (1993) Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. **Crop Science** **33**: 944-950.
- Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants - Salient statistical tools. **Crop Science** **43**: 1235-1248.
- Muhammad RW, Qayyum A, Ahmad MQ, Hamza A, Yousaf M, Ahmad B, Younas M, Malik W, Liaqat S and Noor E (2017) Characterization of maize genotypes for genetic diversity on the basis of inter simple sequence repeats. **Genetics and Molecular Research** **16**: 1-9.

- Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G and Baloch FS (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology and Biotechnological Equipment** **32**: 261-285.
- Nikolić A, Ignjatović-Mićić D, Kovačević D, Čamdžija Z, Filipović M and Drinić SM (2015) Genetic diversity of maize inbred lines as inferred from SSR markers. **Genetika** **47**: 489-498.
- Paradis E and Schliep K (2018) Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. **Bioinformatics** **35**: 526-528.
- Prasad SK and Singh TP (1986) Heterosis in relation to genetic divergence in maize (*Zea mays* L.). **Euphytica** **35**: 919-924.
- R Core Team (2018) **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Available at <<http://www.R-project.org/>>. Accessed on February 29, 2020.
- Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK and He Z (2017) Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. **Molecular Plant** **10**: 1047-1064.
- Reif JC, Hailauer AR and Melchinger AE (2005) Heterosis and heterotic patterns in maize. **Maydica** **50**: 215-223.
- Richard C, Osiru DS, Mwala MS and Lubberstedt T (2016) Genetic diversity and heterotic grouping of the core set of southern African and temperate maize (*Zea mays* L) inbred lines using SNP markers. **Maydica** **61**: 1-9.
- Ringnér M (2008) What is principal component analysis? **Nature Biotechnology** **26**: 303-304.
- Sander T, Freyss J, Von KM and Rufener C (2015) DataWarrior: An open-source program for chemistry aware data visualization and analysis. **Journal of Chemical Information and Modeling** **55**: 460-473.
- Scherlosky A, Marchioro SV, Assis FF, Braccini AL and Schuster I (2018) Genetic variability of Brazilian wheat germplasm obtained by high-density SNP genotyping. **Crop Breeding and Applied Biotechnology** **18**: 399-408.
- Silva KJ, Guimarães CT, Guilhen JHS, Guimarães PEO, Parentoni SN, Trindade RS, Oliveira AA, Bernardino KC, Pinto MO, Dias KOG, Bernardes CO, Dias LAS, Guimarães LJM and Pastina MM (2020) High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. **Crop Science** **2020**: 1-9.
- Smith JS, Gardner CAC and Costich DE (2017) Ensuring the genetic diversity of maize and its wild relatives. In Watson D (ed) **Achieving sustainable cultivation of maize. Volume 1: From improved varieties to local applications**. Burleigh Dodds Science Publishing Limited, England, p. 3-50.
- Sokal R and Michener C (1958) A statistical method for evaluating systematic relationships. **University of Kansas Science Bulletin** **38**: 1409-1438.
- Souza Júnior CL (2011) Cultivar development of allogamous crops. **Crop Breeding and Applied Biotechnology** **11**: 8-15.
- Wickham H (2016) **ggplot2: Elegant graphics for data analysis**. Springer, New York, 276p.
- Wu Y, San Vicente F, Huang K, Dhlwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X and Babu R (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. **Theoretical and Applied Genetics** **129**: 753-765.
- Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P and Zou C (2017) Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. **Molecular Breeding** **37**: 1-12.