

SNP-Based Evaluation of Genetic Diversity and Purity in Elite hybrid Maize breeding in Ethiopia

Azalech Oli

Addis Ababa Science and Technology University

Solomon Benor

`solomon.benor@aastu.edu.et`

Addis Ababa Science and Technology University

Gizachew Haile

Addis Ababa Science and Technology University

Berhanu Tadesse

International Maize and Wheat Improvement Center

Manje Gowda

International Maize and Wheat Improvement Center

Research Article

Keywords: Parental line, Genetic purity, Genetic diversity, Genetic distance

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11306331/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

In maize breeding, cultivar release, and multiplication of the seed, verifying the genetic diversity and uniformity of inbred lines and the resulting commercial hybrid is a crucial technique of quality control. Two thousand two hundred and ninety-four single-nucleotide polymorphism (SNP) markers were used to assess 87 genotypes. Bi-allelic SNP data was processed using R software to assess genetic purity and genotype variation. Polymorphism information content (PIC) values greater than 0.2 were observed in 89% of highly polymorphic SNPs. The parental lines genetic distances varied from 0.004 to 0.548, with CML395/CML202 seed producer5 (SP5) and KIT21XW32XKIT seed producer1 (SP1) showing the largest distance (0.548) and CML495 and CML202 the smallest (0.004). Heterozygosity above 5% was found in 64% of lines, indicating poor purity and the need for additional inbreeding. The lines were divided into three distinct genetic clusters by cluster analysis, which may improve breeding efficiency and provide advantageous alleles to maize breeding programs. Out of the 56 parental lines (36%) that passed the parent offspring test. However, 64% of inbred lines purity % less than 95% which were not pure need to be purified by further inbreeding. The lines were separated into three categories, which might offer useful alleles for breeding maize. This study provides molecular quality control for preserving genetically pure parental lines by evaluating and confirming the genetic purity of inbred lines using genome-wide SNP markers. In order to identify genetic variation, contamination, or residual heterogeneity that could impact hybrid seed quality and breeding efficiency, it also compares lines from various seed producers.

Introduction

Maize (*Zea mays* L.) was cultivated around 9,000 years ago in the Mesoamerican and southern Mexico region (Matsuoka et al. 2002; Piperno et al. 2009). It belongs to the Gramineae family, grown worldwide and the third most significant crop of cereals. Its high genetic yield potential has earned it the designation of “queen of cereals. For many people in developing countries, it is a key source of food and nutritional security (Shiferaw et al. 2011), and it also serves as animals feed and raw ingredients for industrial products. Maize occupies a prominent position in the global economy (Omprakash et al. 2017; Reddy et al. 2021; Tippannavar et al. 2019).

The development of many hybrids and genetic advancement has grown on a global level (Bocianowski et al. 2021). Hybrid breeding is widely used to enhance crop yield by exploiting the phenomenon of heterosis. Maize genetic purity is can easily be affected by female lines' self-pollination during the production of hybrid maize seeds. One of the primary causes of the contamination is incomplete problem, which increases inbreeding levels and reduces the seed's genetic purity and physiological vigor, all of which may ultimately decrease crop yield (Salgado et al. 2006). To address this challenge, quality control (QC) analysis of the genetic purity of parental lines with commercial hybrids (F₁) is a vital element in maize breeding and seed systems (Ertiro et al. 2015).

In hybrid maize production, preserving genetic purity and confirming the purity of F₁ seed lots are essential for achieving the best agronomic performance while maintaining breeding programs'

intellectual property rights (Josia et al. 2021; Semagn et al. 2012). The authentic progeny of the commercial hybrids (F1) from the inbred lines without self-pollination can be tested using various parent offspring testing methods. This includes: molecular markers, biochemical markers and field evaluation (grow out test (GOT)). Grow-out tests use a set of morphological descriptors in the field, which is affected by environmental factors, slow and labor intensive. The biochemical marker involves analysis of the protein profile and, provides rapid and accurate results but with limited polymorphism and availability of the molecular markers. Molecular markers detect variations at the DNA level. Some of the features of molecular makers include high polymorphism, repeatability independent from environmental factors, abundance throughout the genome, known locations and observable at all phases of plant developmental (Dube et al. 2023).

Purity testing using DNA markers is increasingly prevalent in many agricultural crops and horticultural hybrids because it ensures the genuineness of seed before it is distributed to farmers (Pallavi et al. 2010). Genetic polymorphism of maize has been studied using several methods, like random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR), amplified fragment length polymorphism (AFLP) and single nucleotide polymorphism (SNP) (Abakemal et al. 2015; Shinde et al. 2021). Among these, SSRs and SNPs in particular are highly polymorphic, repeatable, environment-independent, expressed across developmental stages, and frequently associated with significant characteristics (Gowda et al. 2017). SNP markers are now commonly used in genetic studies due to their high prevalence throughout the genome, co-dominant inheritance pattern, and high accuracy of prediction potentials (Dube et al. 2023).

When heterozygosity is less than 5%, parental maize lines are accepted as genetically pure. According to (Gowda et al. 2017; Josia et al. 2021), lines with heterozygosity beyond this level are either genetically contaminated or inadequately fixed. Maize may retain up to 10% residual heterozygosity in inbred materials due to its high cross-pollination nature (Nepolean et al. 2013). For commercial hybrid seed production, inbred lines with a genetic purity of greater than 95% are typically suitable and below this level they are either not pure or not fixed. Such genetic contamination results in the reduction of a seed's physiological condition and genetic purity. This results in the reduction of hybrid yield performance. Therefore, genetic diversity evaluation, parental- offspring testing, and genetic purity assessment are crucial quality control methods for successful hybrid breeding programs (Daniel et al. 2012). Despite extensive genetic diversity studies in maize inbred lines, previous research has focused primarily on genetic diversity studies on broad germplasm collections (Abakemal et al. 2015; Adu et al. 2024; Adu et al. 2019; Ayesiga et al. 2023; Boakyewaa et al. 2019; Dube et al. 2023; Ertiro et al. 2017; Josia et al. 2021). On the other hand, there are few studies that combine genome-wide diversity assessment and residual heterozygosity evaluation in the collections of elite tropical maize inbred lines that are directly associated with commercial hybrid breeding programs. The novelty of this study lies in the integration of well-defined tropical maize breeding populations that serves as the parental foundation for commercially produced F1 hybrids in Ethiopia, but not in the use of well-established analytical techniques. Therefore, the present study aimed to evaluate SNP based genetic diversity, residual heterozygosity, fixation status, genetic purity and parental verifications in 56 tropical maize inbred lines and 31 F1 hybrids of maize.

Materials and Methods

Plant Material

A total of 87 genotype including 56 parental line and 31 certified commercial maize hybrids (F1) were collected from different Ethiopian seed suppliers. Of which 21 of them were three-way cross hybrids (3WCHs) and 10 of them were one- way cross hybrids (Supplementary Table S1) were used in this study.

DNA Extraction and Sequencing

The maize leaves were collected from 2 to 3 week old seedlings for DNA extraction using the high throughput mini-prep Cetyl Trimethyl Ammonium Bromide (CTAB) modified protocol by CIMMYT (CIMMYT 2005) and sent to Intertek Laboratory in Australia for mid density DArTseq genotyping. The genotyping technique was carried out using a procedure by (Kilian et al. 2012). In short, leaf samples were used to extract genomic DNA, which was then purified and digested at specific target sites using restriction enzymes. The resultant fragments were then ligated with short DNA adapters that included unique sequence tags. The genomic areas containing the target SNP markers were then selected using a unique complexity-reduction and amplification approach. To produce enough DNA for SNP detection, the amplified fragments were put through PCR. To ascertain whether SNP alleles were present in each sample, the resultant products were hybridized with particular probes, imaged, and assessed with the DArT platform. Genotype calling and data quality checking were performed using the methods designated in (Kilian et al. 2012).

Data Analyses

The DArT platform generated 3,305 SNP markers through its analytical pipeline. To ensure reliability, marker quality control was done in R software version 4.5.1 (R core team, 2025). All 2294 polymorphic SNPs with outstanding quality were used after removing for monomorphic SNPs having less than 2% minor allele frequency. and 80% call rate. Polymorphism information content, fixation index (F_{IS}), homogeneity, expected heterozygosity (H_e), observed heterozygosity (H_o), and other genetic diversity and purity indicators were assessed using R version 4.5.1 (R core team 2025). A distance-based clustering technique was used to examine the genetic similarity between parental lines and F_1 hybrid maize, which was created using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). The Jaccard coefficient was used to generate pairwise dissimilarity matrices and a dendrogram (Josia et al. 2021).

Results

Diversity and Inbreeding of Maize Genotype

Depending on the marker-based analysis, the values for minor allele frequency (MAF) varied from 0.011 to 0.5, with a mean value of 0.203. The observed heterozygosity (H_o) and varied from 0 to 0.924, with a mean value of 0.263 (Fig. 1A & B). Moreover, PIC different from 0.043 to 0.637, with a mean value of 0.397, while Inbreeding coefficient (F_{IS}) varied from - 0.86 to 1, with a mean value of 0.34. However, expected heterozygosity (H_e) for the majority of SNPs (61%) ranged between 0.40001 and 0.6500 and (Fig. 1C). The PIC value higher than 0.20 were identified in nearly 89% of the markers utilized in this investigation. In contrast, PIC value of 0.25 or less was observed in 11% of markers used for the study (Fig. 1D).

The inbreeding coefficient has a mean of 0.34 and ranges between - 0.86 and 1. Negative F_{IS} values were found in about 24% of the SNPs. F_{IS} values larger than 0.2 were found in around 75% of the SNPs (Fig. 1E). The range of gene diversity across genotypes was 0.042 to 0.644, with an average of 0.399. Eighty nine percent (89%) of the expected heterozygosity and polymorphic information lies greater than 0.20001 (Fig. 1C&D), while the expected heterozygosity gene diversity between genotypes varied from 0.042 to 0.644, a mean of 0.399 (Fig. 1C), (Supplementary Table S2)

Results of the genetic diversity parameter estimates using 2294 SNPs were shown in (Supplementary Table S3). Each chromosome had a different number of SNPs distribution; chromosome 10 had 176, while chromosome 5 had 293. Gene diversity values for hybrid maize ranged from 0.356 to 0.399, with a mean of 0.379. The observed heterozygosity of the SNP loci for the hybrid ranged from 3% to 4%. However, the ancestral line's gene variation and PIC value varied from 0.36 to 0.40 across the ten chromosomes, with a mean of 0.38. In contrast, the parental lines PIC and the value of expected heterozygosity, ranged from 0.38 and 0.41, having a mean of 0.40. The inbreeding coefficient (F_{IS}) for the commercial hybrid and parental line (inbred lines) ranged from - 0.08 to -0.14 and 0.52 to 0.61 with a mean of -0.09 and 0.59 respectively.

Genetic Purity of Maize

Parental line and commercial hybrid genetic purity values ranged from 57 to 96% and 41 to 85%, respectively (Supplementary Table S4). In terms of percent heterozygosity, 64% of parental lines had heterozygosity of > 5% and 36% of parental lines had heterozygosity less than 5% (Supplementary Table S5).

Parental Verification

Single crosses (P1XP2) and three-way crosses resulting from a single cross (SC) with a third parent (SCxP3) were also assessed for genetic purity level (Supplementary Table S6). The single-cross hybrids' genetic purity ranged from 85% to 100%; BH-5211 exhibited a 100% genetic purity. Majority (87.5%) of single crosses have high genetic identity values ranging from 92% to 100%. Relatively lower, genetic purity was observed in the three-way cross hybrids. The single cross hybrids such as AMH-850, AMH-854, BH-546, BH-5211 and BHQ-545 showed effective transfer of parental alleles into the final hybrid,

preserving very high identity levels (98%). Sixty-four percent (64%) of the commercial hybrid exhibited genetic purity of greater than 95%. However, few hybrids, like the AMH853 and BH-661, showed genetic purity levels of 80% and 83%, respectively.

Comparison of Maize Inbred Lines across Seed Producers

Assessment of genetic purity between similar hybrid genotypes produced from different seed sources were also carried out. The result indicated that, two inbred lines demonstrated consistent results across producers. However, depending on the seed producer, the genetic purity of the paternal line BKL001 varied significantly, with a minimum and maximum value of 68% and 96%, respectively. (Supplementary Table S7).

Genetic Relationship between 56 Parental Lines

The genetic relationships of the parental lines were assessed using distance-based cluster analyses. The 56 parental lines were clustered into three distinct clusters based on Jaccard's similarity coefficient (Fig. 2). The mean genetic distance between the parental lines was 0.278, with the maximum and minimum genetic distances of 0.548 and 0.004, respectively. The parental lines CML395/CML202 from seed producer (SP2) and KIT21XW32XKIT from (SP1) had the largest genetic distance (0.548). The inbred lines CML495 (SP2) and CML202 from (SP2 & SP4) had the smallest genetic distance (0.004). The genotypes were moderately and highly heterogeneous, with sixty-seven percent (67%) of the genetic distance values falling between 0.3001 and 0.5485 (Fig. 2). Cluster III had fifty-four parental lines, whereas Cluster I and Cluster II each had one parental line (Fig. 3). The clusters analysis well grouped the parental lines into distinct categories (Fig. 3).

Discussion

Evaluating the genetic purity of maize is important for maintaining their genetic uniqueness and assuring their quality by creating necessary corrections during the seed development process. Moreover, it is important to assess a seed lot's authenticity and genetic similarity. Genetic purity assessment helps to determine the presence of selfed seeds, admixtures, or contaminants that could lower hybrid vigor and to confirm successful hybridization between the designated parental inbred lines. In addition, it helps maintain outstanding seed quality, predictable field performance, and varietal distinctiveness. According to the study of (Ertiro et al. 2017; Ertiro et al. 2015; Gowda et al. 2017; Semagn et al. 2012), a heterozygosity rate of 5% or less is the standard level of genetic purity in inbred lines. Studies suggested that (Gowda et al. 2017; Semagn et al. 2012), any inbred lines in the samples with a level of heterozygous loci between 5 and 15% may need to be isolated via ear-to-row selection. On the other hand, a line with more than 15% heterozygous loci is probably contaminated with unrelated genetic material and needs to be either discarded or thoroughly reselected for the original genotype. When heterozygosity was higher than 5% (Daniel et al. 2012), Parental lines are also expected to maintain all the

genetic characters that the breeder selected. The present study thirty six percent (36%) of the tested parental lines were genetic purity are considered as highly stable and appropriate. These lines assist the effectiveness of quality control systems in maize breeding programs and help produce genetically homogeneous and high performing hybrids. However, 64% of parental lines have not yet attained the level of homozygosity predicted for elite breeding materials, as indicated by the finding that the majority of maize inbred lines showed genetic purity below 95%, as indicated genetic contamination and inadequate maintenance of inbred lines during seed development. Thus, before being utilized as parents in the generation of hybrid seeds, these lines could need additional purification and stringent quality control. On the other hand, 77% of single cross hybrid and 71% of three-way cross hybrid varieties exhibited acceptable pedigree conformance with genetic profiles that matched their expected parental combinations. All of these outcomes indicate to the necessity of improving parental line purity through ongoing inbreeding and regular SNP-based quality control in order to guarantee the production of genetically uniform and superior hybrid seed, even though hybrid identity was mostly maintained.

High genetic diversity is a sign of successful hybridization and heterozygosity, which have a strong connection with heterosis and enhanced agronomic performance. On the other hand, low diversity could indicate inadequate seed production, contamination, or self-pollination. Thus, a strong basis for quality control in seed production, variety protection, and the creation of enhanced maize cultivars is provided by efficient genetic diversity assessment. High-throughput genotyping provides an efficient and reliable technique for assessing genetic variation among maize genotypes. These markers will be especially beneficial in maize breeding programs, where finding markers associated with certain traits is crucial to enhancing selection efficiency.

The mean of PIC and expected heterozygosity (H_e) value in this study was 0.4, suggesting that the markers are suitable for identifying genetic variation in the genotype of maize. The current study has similar findings to those reported by (Josia et al. 2021). Several reports indicated higher values compared with the current study (Abakemal et al. 2015; Adu et al. 2024; Ayesiga et al. 2023; Dao et al. 2014; Kondwakwenda et al. 2020). and lower values also reported by (Mushtaq et al. 2016). Particularly, a total of 89% of the SNP markers indicated PIC values greater than 0.2. The result shows a high genetic diversity within the seed lot. This greater diversity confirms and indicates that successful hybridization between various parental lines has taken place. Therefore, the maize population exhibits higher hybrid vigor and dependable seed production methods with little genetic narrowing or loss of diversity. Additionally, this high proportion of informative markers suggests that these SNPs are appropriate for distinguishing between hybrids, allowing reliable distinct genotype identification, off-type detection, and genetic purity confirmation, all of which are essential for breeding programs and quality control.

The SNP markers showed a broad range of genetic diversity, as indicated by MAF values which ranged from 0.01 and 0.5, with a mean value of 0.20. The inclusion of markers with relatively low MAF guaranteed the representation of uncommon alleles, offering a deeper understanding of genetic diversity, even when self-pollination produces fixed or low-frequency alleles in maize populations. The observed mean MAF was higher than in (Ayesiga et al. 2023; Dao et al. 2014), but less than the study by

(Adu et al. 2024; Adu et al. 2019), due to variations in population structure and germplasm background. Importantly, rare alleles enable the discovery of understated distinctions and off-types, while markers with intermediate MAF contribute the most to heterozygosity and are very helpful for differentiating between hybrid genotypes. Additionally, rare alleles allow the discovery of small variations and off-types, while markers with intermediate MAF contribute the most to heterozygosity and are very helpful for distinguishing between hybrid genotypes.

The majorities of the parental lines developed a significant degree of homozygosity and were fixed for a particular allele. The mean of observed heterozygosity (H_o) and inbreeding coefficient was 0.26 and 0.34, respectively. This study's observed heterozygosity was greater than that of (Gowda et al. 2017; Josia et al. 2021), and its F_{IS} was higher than reported by (Kondwakwenda et al. 2020). The parental lines in the current study displayed a larger inbreeding coefficient, suggesting a higher degree of homozygosity, when compared to the report by (Labate et al. 2003). On the contrary, as the hybrids had a low mean F_{IS} value, indicating an abundance of heterozygotes. These findings support considerable heterozygosity in the hybrids and significant inbreeding in the parental lines.

The genetic similarity of parental lines must be established to facilitate efficient exploitation of heterosis in hybrid breeding. Two percent (2%) of lines' genetic distance is greater than 0.50. This indicated important for developing heterotic groups. While 34% of parental lines had the lowest genetic distance, those lines suggest that sister lines with shared pedigrees across most generations. The genetic distance between 0.06 to 0.65 values was reported by (Boakyewaa et al. 2019).

Based on their genetic similarities, parental lines were separated into three main clusters by the UPGMA clustering analysis. Previous studies on maize diversity showed a similar clustering structure in which parental lines were separated into distinct clusters due to their genetic links, pedigree background, and breeding history. The present study's observed clustering suggests that the assessed parental lines have significant genetic diversity, which is advantageous for choosing divergent parents for hybrid breeding programs (Josia et al. 2021). This clustering approach identified a substantial genetic variation between the parental lines. The largest sub-cluster of parental lines represents a genetically homogeneous group, indicating that a substantial proportion of the population shares similar genetic traits. Due to selective breeding within a limited number of genes or shared heritage, this could imply that these lines are closely related. In order to improve particular maize qualities, the smaller groups might reflect underrepresented genetic backgrounds or unique sources of genetic diversity. The consistency of parental lines across production systems allows for the identification of genetically redundant and very different parental lines, as well as the detection of potential contamination or genetic drift during seed processing. The results improve the effectiveness and reliability of hybrid maize breeding and seed production by offering valuable information for heterotic group assignment, parental line selection, quality control, and molecular certification.

Implications for Molecular Quality Control in Maize Hybrid Breeding

The present study covers the application of SNP markers beyond traditional evaluations of genetic diversity and genetic purity by integrating these analyses with parental verification to confirm commercial maize hybrids. Previous studies have successfully employed SNP markers for assessing genetic diversity and genetic purity in maize germplasm (Semagn et al. 2012; Qu et al. 2022) and, in some cases, for evaluating parent–offspring relationships within breeding populations (Josia et al. 2021). However, there hasn't been any thorough confirmation of the authenticity of commercial maize hybrids using a medium-density SNP panel to verify that they inherited genes from their reported male and female paternal inbred lines.

The genetic integrity of the parental lines and the related commercial hybrids is evaluated in this study. Also, it assesses the consistency of seed management techniques by comparing genetically identical inbred lines maintained by various seed companies. This study offers a thorough molecular quality evaluation of the commercial maize seed production system by connecting parental genetic purity with hybrid genetic purity and confirming the genetic linkage between parents and their producing hybrids. Medium-density SNP markers can support regular quality control and seed certification procedures, authenticate parent-hybrid connections, and efficiently identify genetic variations across seed sources.

The integration of these complementary analyses provides a complete framework for molecular quality control in commercial maize breeding and seed production. The findings support more precise hybrid authentication, early detection of pedigree mistakes, pollen contamination, seed admixture, and labeling errors during hybrid development and seed multiplication. Furthermore, this method strengthens molecular-based seed quality assurance and breeding program management by confirming the genetic identity, purity, and authenticity of hybrid seed before release to farmers.

Over all this study provides an integrated molecular quality assessment of commercial tropical maize hybrids and their parental inbred lines using a medium-density SNP panel. This study provided new evidence for improving molecular quality assurance and seed certification systems in commercial maize seed production by showing significant variation in parental genetic purity among seed producers and an unexpected variation between parental purity and hybrid purity.

Conclusion

These findings supported the importance of SNP markers in determining the genetic purity of parental lines and commercial hybrid maize (F1). The SNP markers were also verification of hybrid maize genetic purity. In order to identify any contamination that might have occurred during inbred line development, routine genetic quality assessment was carried out in the present investigation. The genetic purity of parental lines was indicated as highly pure to low purity. This analysis showed significant variations among seed sources, indicating heterogeneity in inbred line seed quality. The values of the genetic purity and diversity assessment ranged from high to low levels. When the inbred lines CML 202 and KIT/SNSYN(N3/TUX))-C1-F1-##-21-2-3-1-1-3-##*3 purity level had equal value confirmations for seed produced by different seed producers, only 36% of the samples met the well accepted requirements from

56 parental lines samples. When we analyzed parental verification of single cross and three-way cross hybrids, 75% and 71% accepted the purity percentage respectively. So, low genetic purity was seen in a number of parental lines, those lines indicating that pollen contamination and improper isolation occurred during the development of the hybrid. However, the majority of the same commercial hybrid variety (F1) and some parental lines produced by different seed producers have different purity level, which suggests need for quality improvement in hybrid development programs. To ensure the development of genetically pure and more superior inbred line maize seed for hybrid production, preservation of improved inbred line, strict field isolation, and improved quality control techniques are required for hybrid development programs in Ethiopia.

Abbreviations

AARC: Ambo Agricultural Research Center

BNMRC: Bako National Maize Research Center

MARC: Melkasa Agricultural Research Center

OSE: Oromia Seed Enterprise

ESE: Ethiopia Seed Enterprise

SP1= Seed producer1, SP2= Seed producer2SP17= Seed producer 17

Declarations

Conflict of Interest

The authors declare that they have no conflicting financial or private interests that might have influenced the study's findings. The authors declare that they have no competing financial or personal interests that could have influenced the results of this study.

Author Contribution

Manuscript Drafting: Azalech Oli
Conceptualization: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda
Data curation: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda
Formal Analysis: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda
Funding Acquisition: Berhanu Tadesse, Manje Gowda
Investigation: Azalech Oli, Solomon Benor, Gizachew Haile
Methodology: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda
Resources: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda
Software: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje

GowdaSupervision: Solomon Benor and Gizachew Haile Validation: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje GowdaVisualization: Azalech Oli, Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda , Yoseph Beyene, Manigben Kulai AmuduReview and Editing: Solomon Benor, Gizachew Haile, Berhanu Tadesse, Manje Gowda

Acknowledgments

The authors are grateful to the International Maize and Wheat Improvement Center (CIMMYT)

Nairobi, Kenya, Ethiopian Institute of Agricultural Research, Addis Ababa Science and Technology University, Institution support seed system in Ethiopia and Ethiopia Institution of Agricultural Research.

Data Availability

SNP data were generated, analysed and included in this report.

References

1. Abakemal D, Hussein S, Derera J, Semagn K (2015) Genetic purity and patterns of relationships among tropical highland adapted quality protein and normal maize inbred lines using microsatellite markers. *Euphytica* 204:49–61. <https://doi.org/10.1007/s10681-014-1332-9>
2. Adu GB, Awuku FJ, Garcia-Oliveira AL, Amegbor IK, Nelimor C, Nboyine J, Karikari B, Atosona B, Manigben KA, Aboydana PA (2024) DArTseq-based SNP markers reveal high genetic diversity among early generation fall armyworm tolerant maize inbred lines. *PLoS ONE* 19(4):e0294863. <https://doi.org/10.1371/journal.pone.0294863>
3. Adu GB, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ, Gedil M (2019) Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE* 14:1–12. <https://doi.org/10.1371/journal.pone.0214810>
4. Ayesiga SB, Rubaihayo P, Oloka BM, Dramadri IO, Edema R, Sserumaga JP (2023) Genetic variation among tropical maize inbred lines from NARS and CGIAR breeding programs. *Plant Mol Biology Report* 41(2):209–217. <https://doi.org/10.1007/s11105-022-01358-2>
5. Boakyewaa Adu G, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ, Gedil M (2019) Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE* 14(4):e0214810. <https://doi.org/10.1371/journal.pone.0214810>
6. Bocianowski J, Nowosad K, Wróbel B, Szulc P (2021) Identification of associations between SSR markers and quantitative traits of maize (*Zea mays* L). *Agronomy* 11(1):182. <https://doi.org/10.3390/agronomy11010182>
7. Daniel IO, Adetumbi JA, Oyelakin OO, Olakojo SA, Ajala MO, Onagbesan SO (2012) Application of SSR markers for genetic purity analysis of parental inbred lines and some commercial hybrid maize (*Zea*

- mays L). *Am J Experimental Agric* 2(4):597–606. <https://doi.org/10.9734/AJEA/2012/1579>
8. Dao A, Sanou J, Mitchell SE, Gracen V, Danquah EY (2014) Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. *BMC Genet* 15:1–14. <https://doi.org/10.1186/s12863-014-0127-2>
 9. Dube SP, Sibiya J, Kutu F (2023) Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers. *Sci Rep* 13:17851. <https://doi.org/10.1038/s41598-023-44961-3>
 10. Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, Wolde L, Tadesse B, Crossa J, Beyene Y, 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* 18:777
 11. Ertiro BT, Ogugo V, Worku M, Das B, Olsen M, Labuschagne M, Semagn K (2015) Comparison of kompetitive allele specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genomics* 16:908. <https://doi.org/10.1186/s12864-015-2180-2>
 12. Gowda M, Worku M, Nair SK, Palacios-Rojas N, Huestis G, Prasanna BM (2017) Quality assurance/quality control (QA/QC) in maize breeding and seed production: Theory and practice. International Maize and Wheat Improvement Center (CIMMYT)
 13. Josia C, Mashingaidze K, Amelework AB, Kondwakwenda A, Musvosvi C, Sibiya J (2021) SNP-based assessment of genetic purity and diversity in maize hybrid breeding. *PLoS ONE* 16(3):e0249505. <https://doi.org/10.1371/journal.pone.0249505>
 14. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. In: Kantety T (ed) *Data production and analysis in population genomics: Methods and protocols*, vol 888. Humana, Totowa, NJ, pp 67–89. https://doi.org/10.1007/978-1-61779-870-2_5.Mooney
 15. Kondwakwenda A, Sibiya J, Amelework AB, Zengeni R (2020) Diversity analysis of provitamin A maize inbred lines using single nucleotide polymorphism markers. *Acta Agriculturae Scand Sect B–Soil Plant Sci* 70(4):265–271. <https://doi.org/10.1080/0906471020201718198>
 16. Labate JA, Lamkey KR, Mitchell SE, Kresovich S, Sullivan H, Smith JS (2003) Molecular and historical aspects of corn-belt dent diversity. *Crop Sci* 43(1):80–91. <https://doi.org/10.2135/cropsci2003.8000>
 17. Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*. 99(9): 6080–6084
 18. Mushtaq M, Bhat MA, Bhat JA, Mukhtar S, Shah AA (2016) Comparative analysis of genetic diversity of maize inbred lines from Kashmir Valley using agro-morphological and SSR markers. *SABRAO J Breed Genet* 48(4):518–527
 19. Nepolean T, Singh I, Hossain F, Pandey N, Gupta HS (2013) Molecular characterization and assessment of genetic diversity of inbred lines showing variability for drought tolerance in maize. *J*

- Plant Biochem Biotechnol 22(1):71–79. <https://doi.org/10.1007/s13562-012-0112-7>
20. Omprakash R, Rajeshsing C (2017) Character association and path coefficient studies for grain yield and related yield attributes among maize inbreds. *Int J Adv Biol Res* 7(1):91–95
 21. Pallavi HM, Ramegowda K, Vishwanath K, Shadakshari YG (2010) Evaluation and confirmation of hybridity on the basis of seed proteins and isozymes in sunflower hybrids. *Res J Agric Sci* 1(3):189–192. <https://doi.org/10.2298/HEL1154059P>
 22. Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Science*. 106(13):5019–5024
 23. Qu J, Chassaigne-Ricciulli AA, Fu F, Yu H, Dreher K, Nair SK, Gowda M, Beyene Y, Makumbi D, Dhliwayo T, San Vicente F, Olsen M, Prasanna BM, Li W, Zhang X (2022) Low-density reference fingerprinting SNP dataset of CIMMYT maize lines for quality control and genetic diversity analyses. *Plants* 11(22):3092
 24. R Core Team (2025) R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
 25. Reddy PR, Talekar SC, Naidu GK, Harlapur SI (2021) Estimation of genetic variability, correlation, and path coefficient analysis in the early segregating generation of maize (*Zea mays* L.). *J Farm Sci* 34(4):357–360. <https://doi.org/10.61475/jfm.v34i04.152>
 26. Salgado KC, Vieira MG, Von Pinho ÉV, Guimarães CT, Von Pinho RG, Souza LV (2006) Genetic purity certificate in seeds of hybrid maize using molecular markers. *Revista Brasileira de Sementes* 28(1):169–175. <https://doi.org/10.1590/S0101-31222006000100024>
 27. Semagn K, Beyene Y, Makumbi D, Mugo S, Prasanna BM, Magorokosho C, Atlin G (2012) Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. *Theor Appl Genet* 125:1487–1501. <https://doi.org/10.1007/s00122-012-1928-1>
 28. Shiferaw B, Prasanna BM, Hellin J, Bänziger M (2011) Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Sec* 3:307–327
 29. Shinde NA, Bharose AA, Sarode DK, Swathi RS, Pimpale PA, Shinde SS (2021) Assessment of hybrid purity in maize (*Zea mays* L.) using RAPD and SSR markers. *Pharma Innov J* 10(4):870–874
 30. Tippannavar PS, Talekar SC, Mallapur CP, Kachapur RM, Salakinkop SR, Harlapur SI (2019) An outbreak of fall armyworm in the Indian subcontinent: A new invasive pest on maize. *Maydica* 64(M4):1–10

Figures

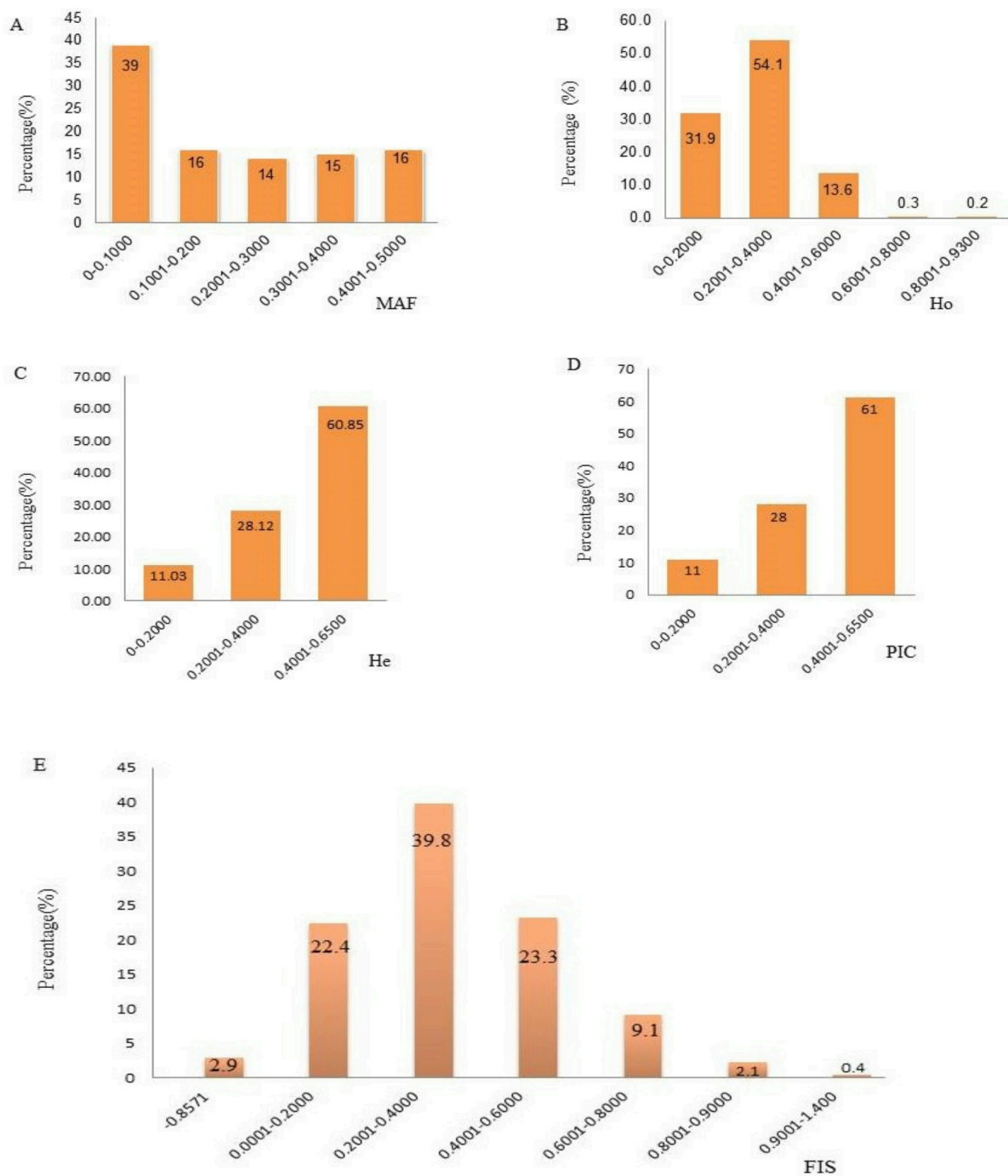


Figure 1

Summary of 2294 SNP markers frequency of 87 maize genotype A) Minor Allele Frequency (MAF), B) Observed Heterozygosity (Ho) C) Expected Heterozygosity (He) D) PIC and E) Inbreeding coefficient (F_{IS}).

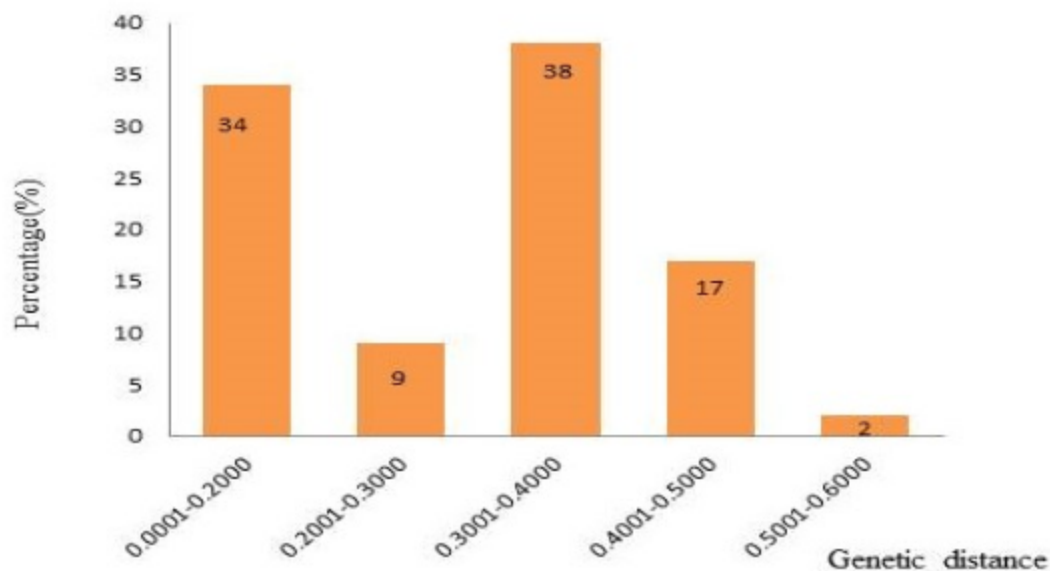


Figure 2

Summary of genetic distance of 56 maize lines using 2294 SNPs.

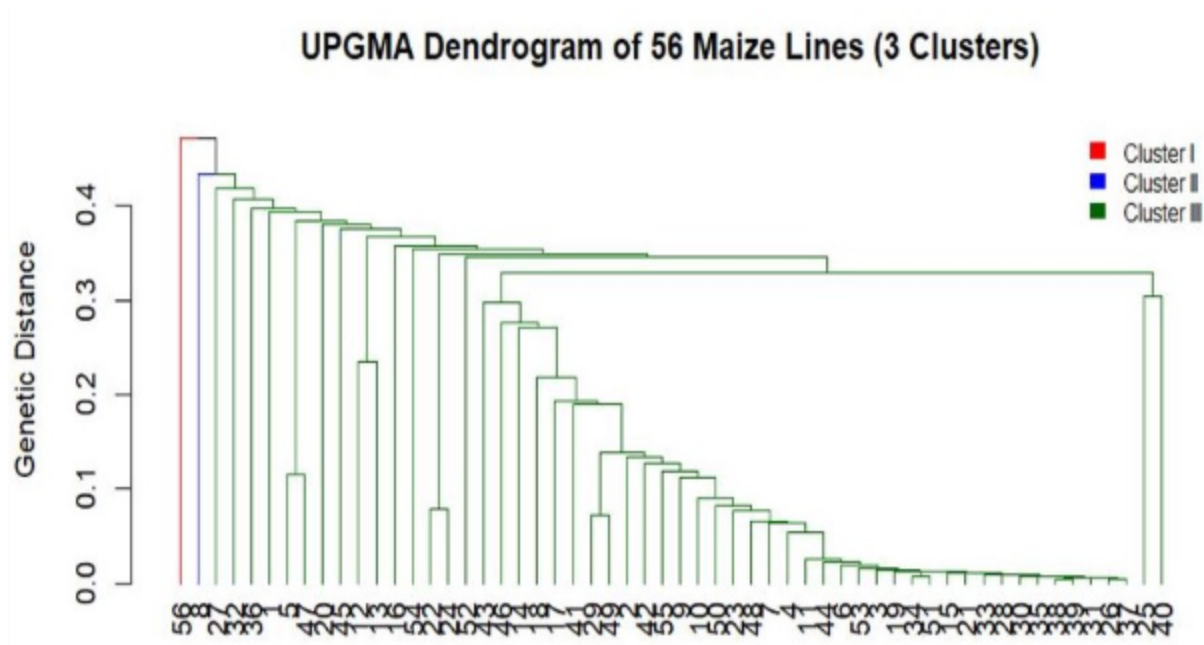


Figure 3

UPGMA dendrogram based on the genetic relationship between 56 parental maize lines based on 2294 SNPs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [2ndSuplimentaryTables.xlsx](#)
- [SupplementaryDataTables.docx](#)