PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.4011

HORIZON
e-Publishing Group
HePG

REVIEW ARTICLE

Genomic insights into maize: Advanced techniques for analysing diversity and enhancing crop traits

S Sneha[1], K R V Sathya Sheela[2*], R Ravikesavan[3], T Selvakumar[4] & V Babu Rajendera Prasad[5]

[1]Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai 625 104, India

[2]Department of Genetics and Plant Breeding, Maize Research Station, Vagarai 624 613, India

[3]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, India

[4]Department of Agronomy, Maize Research Station, Vagarai 624 613, India

[5]Department of Crop Physiology, Tamil Nadu Agricultural University, Coimbatore 641 003, India

*Email: sathyakrv@yahoo.co.in

OPEN ACCESS

ARTICLE HISTORY
Received: 04 July 2024
Accepted: 18 September 2024
Available online
Version 1.0 : 05 December 2024

Check for updates

Additional information
Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE
Sneha S, Sheela KRVS, Ravikesavan R, Selvakumar T, Prasad VBR. Genomic insights into maize: Advanced techniques for analyzing diversity and enhancing crop traits. Plant Science Today (Early Access). https:/doi.org/10.14719/pst.4011

Abstract

Maize is the third important staple food crop grown globally. The demand for maize production has increased significantly due to its multiple uses, including food, feed and various industrial applications. As a result, the area under maize cultivation is expanding, driven by its lucrative market price. Being a highly adaptive crop, the development of high-yielding hybrids better suited to climate change will help bridge the gap between demand and supply. Maize is an allogamous crop, exhibiting greater genetic diversity compared to autogamous crops. Therefore, intensified exploration of maize genetic diversity and effective utilization of germplasm will enhance the maize breeding programs. However, the domestication of maize has led to a decline in genetic diversity and the loss of valuable alleles. Human selection has significantly altered the morphology of maize from its wild ancestor. CIMMYT currently maintains around 28000 maize accessions, including landraces and wild relatives. Genetic diversity can be analysed using D^2 statistics and clustering methods, employing morphological, molecular, quantitative and qualitative data. Careful consideration is needed when selecting appropriate methods and software for such analyses based on available data. In recent years, SSR markers and SNPs have gained popularity for diversity analysis. Studying genetic diversity in maize is crucial for identifying novel traits and the introgression of these traits into new hybrids using advanced technology requires further attention.

Keywords

genetic distance; traditional tools; markers; cluster analysis; utilization; future breeding

Introduction

Maize (*Zea mays* L.), a staple food crop worldwide, belongs to the family Gramineae and has a chromosome number of 2n=20. It is the third most important cereal crop globally, after rice and wheat, contributing to food security in tropical and subtropical regions. Maize is highly adaptable to diverse agroclimatic zones (1). It is cultivated on 193.7 million ha worldwide, producing 1147.7 million tonnes with an average productivity of 5.75 tonnes per ha (2). In India, during 2021-2022, maize was grown on 9.95 million ha, yielding 33.72 million tonnes with a productivity of 33.87 quintals per ha (3). In 2022-23, India exported 3453680.58 MT of maize, valued at Rs. 8987.13 crores/1116.17 USD

million (4). Due to its immense genetic yield potential, maize is often referred to as the 'Queen of Cereals'. Maize serves as a vital source of food, feed, forage, oil and biofuel and its demand continues to grow with the rising global population (5). To meet this increasing demand, developing new hybrids by harnessing the genetic diversity of the crop is crucial.

Genetic diversity refers to the heritable variation within and between populations of organisms. Domestication often leads to a reduction in this diversity, primarily due to the decrease in population size. According to a report, genetic diversity is the level of variability of heritable traits in a population of a given species, while genetic variation is defined as the differences in morphological characteristics such as plant height, flower position, flower colour, etc as well as physiological characteristics, biochemical traits and DNA sequence (6). Since genetic diversity may not be expressed phenotypically, it encompasses the total genetic differences present among individuals, genotypes, strains, clones or populations within a species (7). It was described that genetic diversity as the presence of differences in alleles and genotypes, their resulting phenotypes and the overall sum of the genome (8).

Genetic diversity in maize

The domestication of maize originated from Balsas teosinte (*Zea mays* subsp. *parviglumis*) in Mexico around 9000 years ago (9). The primary aim of domestication was to enhance yield and simplify the harvesting process (10). Due to the distinct morphological differences between the maize ear and its ancestral teosinte, taxonomists classified them as separate species. The teosinte ear can be distinguished from the maize ear by 4 key characteristics. First, maize kernels are more easily harvested and digested compared to teosinte kernels, which are encased in tough outer glumes. Second, maize ears can contain hundreds of kernels arranged in multiple rows, while teosinte ears hold only 5-12 kernels arranged in 2 ranks-one on each side. Third, maize ears have a non-shattering rachis, while teosinte ears have a shattering one; the non-shattering trait was selected during the domestication of cereals to facilitate easy kernel collection. Fourth, each cupulate fruit case of a maize ear contains paired mature spikelets, potentially doubling the number of kernels, whereas in teosinte, each fruit case has a single mature spikelet, resulting in one kernel per spikelet (11). Cultivated maize produces fertile offspring when crossed with *Teosinte-parviglumis*. Domestication led to an increase in ear size while reducing the number of ears per plant. This domestication syndrome allowed maize to adapt to farmers' cultivation and harvesting practices (9). However, during this process, some heritable variation in ear architecture and morphology was lost. For example, maize plants developed fewer or no branches and produced larger ears. Traits selected by early farmers based on their needs showed a strong correlation (12). The alleles and different allelic combinations present in maize germplasm, which contribute to traits like increased yield, tolerance to biotic stress, disease resistance and nutritional quality, were not lost but instead distributed across various populations (13).

Since the domestication of maize, it has experienced a notable reduction in genetic diversity. Isozyme data indicates that maize has lost approximately 25 % of the gene diversity present in Balsas teosinte. Similarly, nucleotide data estimates suggest a 30 % loss of gene diversity in maize compared to teosinte, a figure closely aligned with isozyme-based estimates (12). The maize HapMap_v2 dataset, which includes genome-wide sequence data, was used to analyze diversity through 55 million SNPs across landraces, 75 inbred lines and teosinte. This analysis revealed that landraces have retained 83 % of the teosinte genes. It also showed that human selection during domestication had a greater impact on genetic diversity than subsequent modern breeding efforts (14). One contributing factor to the loss of genetic diversity is rapid urbanization and extreme, unpredictable climatic events such as drought, heat and flooding (13). However, certain landraces still maintain some additive genetic variance, which is valuable for future crop improvement. Additive genetic variance is particularly important in selection-based crop improvement because it is heritable, fixable and responds well to selection (15). Modern maize breeding continues to benefit from the rich genetic diversity offered by maize landraces (12). The domestication difference in maize is depicted in Fig. 1.

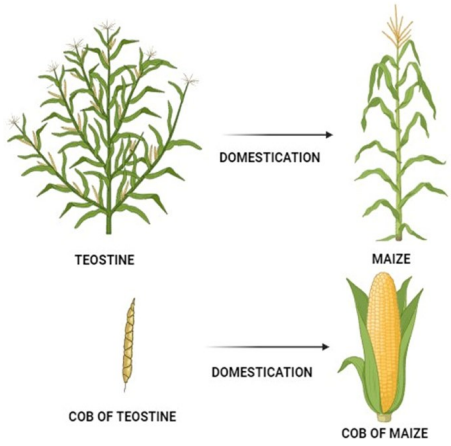

Fig. 1. Changes in the architecture of teosinte through the years of the domestication process which finally evolved as a most common food crop.

In a specific region of Sikkim, India, farmers continue to conserve and cultivate a unique collection of maize landraces known as Sikkim Primitive. These landraces are characterized by popcorn-like kernels, a lack of apical dominance, multiple ears per stalk (ranging from 5 to 9), long drooping tassels and uniform ear size. They form a crucial foundation of genetic diversity and serve as a reservoir of traits essential for future climate-resilient breeding. One such landrace, Murli Makai, known for its green leaves even after maturity, is particularly well-suited for animal fodder (13). The "stay-green" (SG) trait allows maize plants to retain green leaves and continue photosynthesis for an extended period after flowering, especially under water-stressed conditions. This trait extends the grain maturation phase and often results in higher yields compared to non-SG maize varieties. The prolonged greenness of SG maize leaves makes them an ideal source of fresh green feed for animals (16). The introduction of wild germplasm, especially from species closely related to cultivated maize, has contributed significantly to the genetic variability observed in modern maize. The exchange of genetic material between maize and

teosinte is commonly observed in fields where both are grown, particularly in regions of Mexico and Guatemala (17). Genetic diversity analysis plays a critical role in classifying genotypes into heterotic groups and assessing the variability present in genetic resources (18).

Basics of genetic diversity analysis

The individuals selected for genetic diversity analysis may include inbred lines (IL), pure lines, clones or populations. Data collected for this analysis can encompass qualitative or quantitative traits, isozymes and DNA markers. The fundamental steps in genetic diversity analysis, whether using traditional or modern approaches involve:

I. Estimating the genetic distance or similarity between pairs of entities and

II. using these estimates to group the entities accordingly (7).

Genetic distance

Genetic distance is defined as “the extent of gene differences between populations or species, measured by a numerical quantity" (19). Similarly, it was described that genetic distance as a “quantitative measure of genetic difference, whether at the sequence level or allele frequency level, calculated between individuals, populations, or species" (20). Genetic distance forms the foundation for subsequent analysis in population genetics and can be estimated at various levels, either among individuals or populations. A summary of different data types and their corresponding methods for calculating genetic distance is provided in Table 1.

PIC

The polymorphic information content (PIC) is an estimate of the degree of polymorphism at a marker locus (7). It is calculated using the following formula:

$$PIC = \sum_{i=1}^{n} Pi^2 \qquad \text{(Eqn. 1)}$$

Where Pi represents the frequency of the i^{th} allele (18). PIC is also referred to as expected heterozygosity (H_e). To determine the average PIC score for a marker system, data from all loci evaluated for the system can be combined. Research indicates that the average PIC score (H_{av} = average heterozygosity) for all polymorphic markers in a system can be calculated as:

$$H_{av} = \sum H_{ei}/np \qquad \text{(Eqn. 2)}$$

Where np is the number of polymorphic loci in the population and H_{ei} is the expected heterozygosity or PIC value of the i^{th} marker locus (7).

Table 1. A consolidated summary for the elucidation of genetics distance for various traits depends on different data obtained from morphological, molecular and population basis.

Types of data	Genetic distance	References
For morphological data	Euclidean or straight line measure $D_{ij} = \left[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots (x_p - y_p)^2\right]^{1/2}$	(7)
	Gower distance $d_G X_i, X_j = \frac{\sum_{c=1}^{m} W_{ijc}\, d_{ijc}}{\sum_{c=1}^{m} W_{ijc}}$	(94)
For molecular data	Roger’s distance $D_{ij} = Constant\left(\sum \lvert X_{ai} - X_{aj}\rvert^r\right)^{1/r}$	(7)
	Nei & Li’s coefficient $GD_{NL} = 1 - \left[\frac{2(N_{11})}{(2N_{11} + N_{10} + N_{01})}\right]$	(95)
	Jaccard’s coefficient $GD_J = 1 - \left[\frac{(N_{11})}{(N_{11} + N_{10} + N_{01})}\right]$	(96)
	Simple matching coefficient $GD_{SM} = 1 - \left[\frac{(N_{11} + N_{00})}{(2N_{11} + N_{10} + N_{01} + N_{00})}\right]$	(97)
For populations	Total gene diversity (H) $H = 1 - 1/m \sum_{i=1}^{m} \sum_{j=1}^{n_j} p_{ij}^2$	(7)
	Nei ‘s geometric distance	(7)
	Modified roger’s distance $GD_{MR} = \sqrt{\left[\frac{(N_{10} + N_{01})}{2N}\right]}$	(98)

Grouping

Grouping is crucial for reflecting evolutionary patterns, providing either a graphical or textual description of the population being studied. The method used to analyse data for multiple traits is known as multivariate analysis (7). This approach is widely employed for examining variation patterns and genetic relationships.

Cluster analysis

Pairwise distance matrices are used as the input for clustering in GD-based clustering algorithms, which produce a graphical representation in the form of a tree or dendrogram (7). Distance-based clustering techniques are the most commonly applied, with model-based (Bayesian) clustering also used to estimate genetic relationships among individuals (21). These clustering techniques are divided into 2 types: hierarchical and non-hierarchical. Hierarchical clustering is more frequently applied in studies of genetic diversity in plant species. Among the hierarchical methods, the unweighted pair group method with arithmetic mean (UPGMA) and Ward's method are widely used (22). Ward's method is often considered more suitable for clustering based on molecular data, as it avoids the "chaining" effects frequently observed with UPGMA (23). Mahalanobis distance (D^2) developed by P.C. Mahalanobis in 1928 and later suggested by C.R. Rao in 1952, is commonly used for clustering quantitative traits to determine genetic diversity in crops. The cluster diagram is illustrated using the square roots of the mean intra- and inter-cluster D^2value, with the degree of germplasm diversification assessed by the distance between clusters (15). This methodology is applicable to both quantitative characters and gene frequency data (19). For example, a study constructed a dendrogram using the neighbour joining algorithm, based on 45000 SNP markers, which grouped 222 inbred lines into 3 major clusters (24).

PCA

Principal component analysis (PCA) is used to reduce the number of variables and group genotypes. PCA can generate 2D or 3D scatter plots of the entities in a study (7), providing key insights by simplifying data related to complex traits (25). While reducing the dimensionality of the data, PCA retains most of the variation within the dataset (26). By combining comprehensive variables that represent various traits referred to as PC scores with genome wide association studies (GWAS), it is possible to avoid overlooking important genomic regions that might be missed when analysing individual traits (27). Clustering methods often use principal component scores as input to group the data entries (28).

PCoA

Principal coordinate analysis (PCoA), also known as metric multidimensional scaling, analyses data represented as either a similarity or dissimilarity matrix (29). This technique produces a graphical representation in a reduced number of dimensions using the input similarity or dissimilarity matrix. The distances between any 2 points on this plot reflect how different the entities originally were from one another. PCoA is advantageous over PCA when dealing with a large amount of missing data or when the number of traits exceeds the number of genotypes in the study (7). Another benefit of PCoA is its ability to handle both quantitative and qualitative data, offering greater flexibility in variable types. However, nonlinear PCA can also be used for qualitative data when optimal scaling is applied to convert qualitative data into quantitative form (30).

Conventional methods for genetic diversity analysis

Historically, genetic diversity was analysed based on phenotypic characters through visualization methods. The diversity analysis of morphological data encompasses both quantitative and qualitative traits. For the phenotypic characterization of maize, the international union for the protection of new varieties of plants (UPOV) has provided scoring scales for numerous morphological and physiological characteristics (31). Additionally, descriptors from the International Maize and Wheat Improvement Centre (CIMMYT) and the International Board for Plant Genetic Resources (IBPGR, 1991) have been utilized to estimate variation in morphological traits (21). Significant variability has been observed in various phenotypic traits of maize landraces, including plant height, 100 grain weight, ear height, ear diameter, number of kernel rows per ear, protein content and shelling % (32). Other traits such as days to silking, days to anthesis, rachis diameter and yield per plot have also shown considerable variability among landraces. These morphological traits contribute to diversity analysis and the variation within them can be assessed using analysis of variance (21). The recorded morphological data can be effectively utilized by decomposing them into principal components for further analysis (33). The phenotypic characterization indicates a significant degree of variability in quantitative traits within the Saharan maize germplasm. Consequently, clustering populations based on comparable phenotypic features may facilitate genetic enhancement. Inbred lines exhibit genetic diversity that supports phenotypic variation, thus enabling the breeding and selection of stress-tolerant maize hybrids (34). The phenotypical characterization for qualitative traits is easy as it is a visual scoring of the presence or absence of character (35). Phenotypic traits are useful in the preliminary evaluation of genetic diversity providing practical and essential information needed for genetic resource characterization. Morphological analysis for genetic diversity assessment presents many limitations, such as low polymorphism, less information and the influence of environment on phenotypic expression (21). On behalf of these limitations and for fine analysis, markers stepped into diversity analysis.

Modern methods for genetic diversity analysis

Molecular markers are independent of environmental influences, making them effective for leveraging genetic differences and expediting breeding efforts. In the 1990s, Restriction Fragment Length Polymorphism (RFLP) was the most widely used molecular marker for assessing genetic diversity. Subsequently, PCR-based markers, such as Simple Sequence Repeats (SSRs) and Amplified Fragment Length Polymorphism (AFLPs), gained prominence in these studies. RFLPs became less favored due to their labor-intensive and time-consuming nature (36). RFLP markers require a significant amount of DNA, typically between 5 to 10 µg, whereas AFLP markers need only 100 to 500 ng and SSR

markers require about 100 ng. Furthermore, RFLP markers exhibit a moderate degree of polymorphism, AFLP markers show a higher degree of polymorphism, while SSR markers display the highest degree of polymorphism (7). Additionally, AFLP markers are a quick and reliable method that demonstrates a strong correlation with the results obtained from the RFLP system. Genetic diversity assessed through these markers can be analyzed using Analysis of Molecular Variance (AMOVA) (32).

While determining the genetic distance, use of SSR markers instead of RFLP markers proved to be both valuable and user-friendly (35). Utilizing a greater number of SSR markers can yield more reliable results and is frequently employed in the analysis of maize genetic diversity (13). Due to the high degree of polymorphism present in SSR markers, they can effectively identify heterotic groups within genotypes. It was concluded from their experiments that the microsatellite loci umc1066, bnlg1805, phi116 and phi065 exhibited higher polymorphic information content (PIC) values in genetic analysis, highlighting the effectiveness and potential of SSR markers (37). Recently, there has been significant use of SSR markers in varietal protection, providing a means to uniquely identify specific genotypes based on their distinct allelic patterns. Research employing a larger number of markers may be beneficial in identifying genotypes with enduring, promising outcomes (38).

Relying solely on morphological characterization of germplasm or landraces can be ineffective due to environmental influences and limited reproducibility (32). However, phenotypic characterization remains important as it offers preliminary insights into genetic resources. Combining morphological data with marker data can lead to more comprehensive results.

Advanced genomic techniques for diversity analysis

Single nucleotide polymorphisms (SNPs)

Among all molecular markers, Single Nucleotide Polymorphisms (SNPs) provide the highest density of coverage across genomes, making them the most prevalent source of genetic variation (39). As a third-generation marker, SNPs offer significant advantages, particularly in high-throughput genomic selection (40). They have been applied in the fingerprinting of common corn, increasing the likelihood that some SNPs will be linked to genes of interest, thereby enhancing the accuracy of breeding programs in assessing genetic variation. The widespread use of SNPs in research can be attributed to the high throughput and low cost of next-generation sequencing technology as well as automated platforms that meet the requirements of breeding programs (41). In a study, 9642 SNP markers were utilized to genotype 439 maize inbred lines for the analysis of molecular diversity (42). These inbreds were categorized into 10 clusters. Notably, clusters 6 and 2, which accounted for 2.05 % of the total inbreds, included a higher number of inbreds derived from the '2009 TZE OR1 DT STR' population, characterized by orange endosperm kernels, drought tolerance and resistance to Striga.

Genotyping-by-sequencing (GBS)

In recent years, significant advancements in sequencing technologies have enhanced our understanding of the maize genome. Genotyping-by-sequencing (GBS) has emerged as a cost-effective approach for genotyping large numbers of samples, yielding a higher number of SNPs compared to traditional SNP arrays. It was genotyped 2240 maize lines from 8 tropical maize populations using a one-enzyme-based GBS method (43). This approach generated a total of 995120 raw SNPs for each individual. The average % of missing SNP genotypes were 39.05 % for bi-parental populations and 57.72 % for association panels. These high missing data rates can largely be attributed to the multiplex methodology and the inadequate coverage of the sequencing method used. Therefore, it is crucial to develop better imputation pipelines for GBS data. To address this issue, 2 imputation pipelines from the TASSEL software were employed.

Cost-effectiveness is one of the primary advantages of GBS. While GBS enables genome-wide SNP identification, the low sequencing coverage presents challenges, including high genotyping error rates, elevated missing data rates, reduced accuracy in calling heterozygous SNPs and potentially lower marker density. To mitigate these issues, several enhanced GBS techniques have been developed. For instance, Tunable Genotyping-by-Sequencing amplifies and sequences double-digested fragments to reduce the missing rate and improve genotyping accuracy, particularly for heterozygous loci (44).

Genome-wide association studies (GWAS)

Genome-wide association study (GWAS) is an effective method for identifying the genetic factors underlying complex traits. It involves genotyping a large number of accessions using numerous SNPs and analysing the relationships between these SNPs and agronomic traits. GWAS has become a valuable tool for identifying beneficial alleles linked to desirable traits in plant species, utilizing both phenotypic and genotypic variations to uncover advantageous genetic variations (45). The primary goal of GWAS is to establish relationships between target traits and genotypes, serving as a potent alternative to traditional QTL mapping by accurately identifying the genetic loci underlying traits at a high resolution. Several statistical approaches for GWAS have been employed, utilizing software such as Tassel version 5.0. GWAS analysis can be conducted after collecting both genotypic and phenotypic data to enhance mapping power. It is regarded as a flexible genetic tool for improving complex quantitative traits (46). GWAS is a reliable and effective method for tracing complex phenotypic features back to their underlying genetics. In 2008, researchers conducted the first genome-wide association study (GWAS) in maize using 8590 loci in 553 elite inbred lines to identify loci with a primary effect on the oleic acid content in seeds. The fad2-associated marker MZA10924 identified in this study can be used to predict oleic acid levels in plants, enabling the enhancement of 18:1 fatty acid content from 2 5% to 40 % (47). Since then, the application of GWAS for various phenotypes has rapidly advanced, culminating in the publication of the reference maize genome, known as "B73" (48). The steps involved in genome wide association studies (GWAS) are illustrated in Fig. 2, while Fig. 3 depicts the methodologies associated with GWAS analysis for a specific desirable trait.

Fig. 2. The steps involved in the Genome Wide Association Studies (GWAS) for the analysis of specific genes and also for the associated candidate genes.

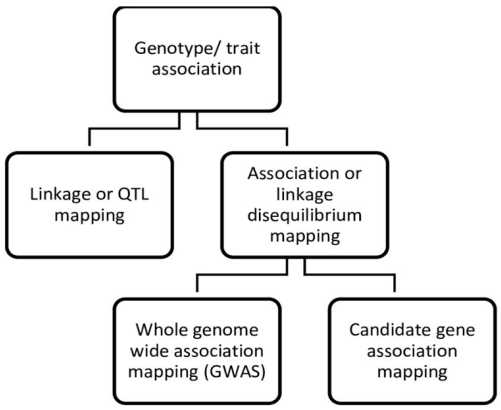

Fig. 3. Methodologies associated with the GWAS analysis for a specific desirable trait.

Applications of GWAS in dissecting complex traits in maize

The effective use of GWAS underscores its significant relevance in examining various traits affected by drought, salt and temperature stresses. In maize, GWAS investigations have identified new gene candidates responsible for abiotic stress, providing valuable insights for maize breeding and the development of climate-resilient varieties (49). A study established the association between genetic variation in ZmDREB2 and the level of drought tolerance by performing GWAS on 368 maize varieties (50). Additionally, cysteine-rich polycomb-like (CPP) proteins were discovered in maize, with the gene ZmCPP contributing to a better understanding of maize growth and development under various abiotic stress conditions due to its expression level variations under specific stresses such as cold, salt, heat and drought (51). It was utilized a metabolome-based GWAS in maize kernels to explore the entire biochemical landscape and to identify both general and specific patterns (52), thereby investigating another significant layer influencing genetic variation in maize. The GWAS data currently available have been fully leveraged by these molecular-level association studies, enhancing our understanding of the intrinsic functional genome that underlies trait variation.

Genomic selection (GS)

Genomic selection (GS) is an innovative strategy used to predict the performance of genetically distinct individuals through analysis of historical phenotypic data and genome-wide molecular markers (53). For GS to be successful, a training population (TP) must be established. This can be achieved by accurately phenotyping a collection of maize lines for the desired traits and then genotyping these lines using highly discriminating molecular markers. Genomic selection can significantly accelerate grain production. Research conducted by the International Maize and Wheat Improvement Center (CIMMYT) in Sub-Saharan Africa, India and Mexico has demonstrated that genomic selection can reduce the breeding interval by at least half compared to traditional methods and can generate lines that significantly outperform others in terms of grain yield (54). From a statistical perspective, the core concept of GS is to model all markers simultaneously in order to estimate genomic estimated breeding values (GEBV) (55). Unlike QTL mapping and association studies, GS utilizes every molecular marker for genomic-enabled prediction (GP) of the candidates' performance in making selections. Genomic-enabled prediction is a method used to forecast the genetic potential or breeding value of individuals based on their genome-wide DNA marker data. Thus, the primary objective of GS is to predict genetic and/or breeding values. To determine the GEBVs of individuals in a testing population (TST) that have been genotyped but not phenotyped, GS integrates molecular and phenotypic data from a training population (TRN) (54).

Software in statistical analysis

R program: The R programming language is widely used in the 21st century for various statistical analysis. It was utilized R software for conducting Analysis of Variance (ANOVA) (34). For correlation analysis, the Pearson correlation coefficient is employed, using the "corrupt" package in R. The "adegenet" package is utilized to estimate various genetic parameters, such as Polymorphic Information Content (PIC), major allele frequency, number of alleles, heterozygosity and gene diversity. Additionally, the "ggp-ubr" package in R is used for Principal Component Analysis (PCA). It was employed the "ade4" package in R to estimate Principal Coordinate Analysis (PCoA) using a genetic distance matrix for analyzing landraces (56). Allelic richness (Ar) and genetic diversity (He) for landraces can also be calculated using ad hoc scripts in the R program. The "cluster" package in R is utilized for clustering genotypes, offering enhanced analytical capabilities. Overall, R is a versatile tool that can be applied to various analyses in genetic diversity, as demonstrated through its applications in ANOVA, correlation analysis, PCA, PCoA, estimation of genetic parameters, allelic richness, genetic diversity and clustering.

SAS: The PROC GLM (General Linear Models) program provided by SAS can be used to perform a variety of analyses, including regression, analysis of variance (ANOVA), analysis of covariance (ANCOVA), multivariate analysis of variance (MANOVA) and partial correlation (32).

PHYLIP: The phylogenetic tree was constructed using the neighbor-joining algorithm implemented in the PHYLIP computer program (56).

GenAlex: It was utilized GenAlex for the analysis of molecular diversity (AMOVA) (21). To determine genetic parameters such as the number of alleles (N), number of different alleles (Na), number of effective alleles (Ne), expected heterozygosity (He), observed heterozygosity (Ho) and Shannon diversity index (I) among populations and SSR loci, they used GenAlex software version 6.3 (57).

SPSS 22.0: The correlation coefficient matrix, analysis of variance table, principal component load matrix and eigenvector for each index parameter can be generated using SPSS version 22.0.

STRUCTURE v2.3.4 sofware31: It was developed STRUCTURE, a freely available population analysis tool (58). STRUCTURE utilizes a Bayesian iterative method to classify samples into groups based on similar patterns of variation, analyzing the differences in the distribution of genetic variants among populations. The software identifies populations and assigns individuals to the population that best fits the observed variation patterns. Therefore, STRUCTURE can effectively be used to obtain insights into population structure (59).

SeeD: SeeD is one of many initiatives and online platforms dedicated to exploring genetic diversity and molecular data related to maize. Developed by CIMMYT, seeds of discovery (SeeD) represent an approach aimed at uncovering answers within the realm of genetic resources. In light of the declining agricultural resource base, genetic resources hold significant and largely untapped potential for improving yields, likely ranking just behind advancements in agronomic practices (60). A list of various software and their functions for data collection is presented in Table 2.

Utilization of genetic diversity

Yield improvement

Crop improvement breeding programs are primarily driven by the genetic variability and heritability of various traits inherent in the germplasm (57). Yield is not an independent trait; it is heavily influenced by multiple factors such as plant architecture, environmental conditions and nutrient availability. Therefore, improving yield requires consideration of all the components that directly or indirectly impact yield traits. Upright plant architecture, for example, supports a higher plant density per unit area, ultimately leading to increased maize yields (17). It was concluded that the upright architecture allele (UPA2), which reduces leaf angle, was present in teosinte and maize's wild ancestor but was lost during domestication (61). Reducing the average number of tassel branches from 13.0 to 0.5 branches at a density of 76600 plants/ha has been shown to increase grain yield by 6 q/ha (17). Maize with a high ear-to-plant height ratio is prone to lodging. The optimal plant architecture for yield improvement involves maintaining a balanced ear height to plant height ratio (62).

Abiotic stress tolerance

In recent years, climate change has significantly impacted rising temperatures during both day and night, along with other abiotic factors that negatively affect maize, resulting in a drastic reduction in crop productivity (61). Crops with greater genetic diversity within their populations are better equipped to adapt to sudden environmental changes. The genetic variability among lines in stressful conditions facilitates the selection of parent plants that will produce heterozygous and adaptive progeny (63). In crop breeding programs, selecting genetically distinct genotypes can help develop new, highly heterotic hybrids that are adaptable under various ecological and stress conditions (64).

In India, the primary climatic factor affecting maize productivity has been identified as dry spells, particularly during the flowering period (65). To enhance the resilience of maize production in heat-stressed conditions, there is a pressing need to effectively target the development of heat-tolerant varieties (66). The public-private partnership known as heat stress tolerant maize (HTMA) for Asia aims to assist South Asian smallholder farmers and underprivileged communities impacted by climate change and severe weather (67). The National Maize Research Program (NMRP) collaborated with the International Maize and Wheat Improvement Center (CIMMYT) to develop heat-tolerant maize hybrids (HTMHs), which were subsequently distributed across Nepal with the support of commercial seed companies. Rampur Hybrid-8 and Rampur Hybrid-10 are examples of heat -tolerant single cross hybrids (68). The anthocyanin coloration in root, stem and leaf tissues serves as a morphological marker that confers resistance to various abiotic stresses. The allele responsible for this anthocyanin coloration was introgressed from teosinte into cultivated maize. The red and hairy sheath trait of teosinte is advantageous in cooler climates. Additionally, anthocyanin accumulates in response to wounds and pathogens (17). It was studied 33 landraces from sub-

Table 2. Software and their function in data collection.

S.no	Software's	Function	Developer	References
1	Flapjack	A multiplatform program called Flapjack makes use of single nucleotide polymorphism (SNP) data to allow for the graphical display of both basic and complex genotypes and haplotypes. Flapjack makes it possible to sort samples based on phenotype and to depict QTLs.	The James Hutton Institute and CIMMYT collaborated to develop Flapjack.	(60)
2	Curly Whirly	Another multi-platform program called CurlyWhirly was created to facilitate interactive multi-variate analysis visualization. It focuses specifically on the results of principal coordinate analysis, principal components analysis, canonical analysis, and multidimensional scaling. Interpretation and point selection are made easier when results are integrated with additional categorical data.	In partnership with CIMMYT, the James Hutton Institute created CurlyWhirly.	(99)
3	KDSmart	Using a tablet or smartphone, KDSmart is an Android platform digital app for capturing phenotypic data in the field, greenhouse, or lab.	Diversity Arrays Technology (DArT) and CIMMYT worked together to produce KDSmart	(100)

Saharan Africa to identify genes resistant or tolerant to drought stress, heat stress and combined drought-heat stress. The landrace *GH-3505* demonstrated greater drought tolerance, while *GH-4859* and *TZm-1353* exhibited tolerance to abiotic stresses along with desirable secondary traits. Drought, heat and combined heat-drought tolerance are largely regulated by different genes (70). Heat stress adversely affects seed set by influencing silk and pollen receptivity, making these traits crucial for selection in heat stress tolerance (71). Landraces exhibiting adaptive traits such as early flowering and small stature were identified as potential sources of genetic diversity for developing varieties resistant to abiotic stresses, with the Burkinabe gene pool proving to be particularly unique (69). Compared to previous research, there was increased genetic diversity in cold tolerance and most traits exhibited relatively high heritability. It was identified 187 significant single-nucleotide polymorphisms (SNPs) associated with emergence and early growth-related phenotypes, which were incorporated into 159 quantitative trait loci (QTL) (72).

Biotic resistance

A major contributing factor to yield loss in maize is biotic stress, often caused by diseases or insect pests. By examining landrace varieties, it may be possible to successfully introduce resistant genes into contemporary cultivars (73). For instance, the solution for corn blight involved the introduction of blight-resistant alleles from a wild relative of Mexican maize, *Tripsacum dactyloides* L., into commercial corn lines (74). Rootworms are another destructive pest affecting corn in the United States and resistance genes from gama grass (*T. dactyloides* L.) have been deployed into cultivated corn to combat this issue. To address grey leaf spot, researchers explored untapped allelic diversity in the teosinte NIL population, where *Zea parviglumis* exhibited resistance to both grey leaf spot and southern corn leaf blight (75). A recent analysis of a teosinte-derived maize population using QTL mapping revealed a novel QTL on chromosome 5 and four minor QTLs on chromosomes 1, 3, 4 and 8 that provide resistance against sheath blight and banded leaf (76). QTL mapping is a widely used method for identifying genomic regions associated with specific quantitative traits or disease resistance in plants. The tetraploid genetic background of *Zea perennis* may harbor potential resistance genes that could function in modern maize, potentially explaining its higher resistance to viruses (77).

It was also discovered that *Tripsacum*-derived maize lines are a source of resistance genes against the maize weevil. *Z. parviglumis* was found to possess resistance genes located on chromosomes 1 and 3, specifically targeting the maize red flour beetle (78). The genetic variation among hybrids of maize resistant to the larger grain borer presents an opportunity to develop variants resistant to storage pests. To create maize cultivars suitable for the region, two landraces (ZM 4236 and ZM 7114) and one open-pollinated cultivar (Pool 16) were selected from Zambia for their desirable agro-morphological traits and resistance to fall armyworm (FAW) (79).

After nearly 14 generations of backcrossing *Tripsacum* to maize through recurrent selection, backcrossing and selfing, it was assessed the western corn rootworm tolerance of *Tripsacum*-introgressed maize lines (80). However, only a few of the lines exhibited tolerance to the western corn rootworm. To maintain stable yields in the face of future climatic fluctuations and pest outbreaks, research into new sources of resistance with beneficial alleles and the development of multiple resistant lines should be continuous processes.

Striga resistance

Striga parasitizes maize to obtain photosynthetic assimilates from the plant, leading to stunted growth and reduced yield of the host. In cases of high infestation, crop failures are often complete (81). The strigolactones secreted by maize roots are responsible for triggering *Striga* seed germination (82). To develop maize genotypes that produce low levels of strigolactones while exhibiting favorable agronomic traits, genetic variation is essential (83). A substantial reservoir of genetic variation is necessary to create multiple mechanisms that contribute to *Striga* resistance and tolerance (18). These mechanisms include reduced germination stimulant activities, physical root barriers and highly reactive responses following attachment as well as the ability to ignore *Striga* toxins (84). The development of stable and long-lasting resistance to *Striga* from wild maize (*Zea diploperennis* L.) and African landraces has been a key focus of the International Institute of Tropical Agriculture's maize improvement program (MIP-IITA). *T. dactyloides*, one of maize's wild relatives, demonstrates incompatibility with parasitic *Striga*, inhibiting haustoria development. This incompatibility is attributed to signals produced by *T. dactyloides* that prevent haustorial development, necessitating modern breeding techniques for the transfer of resistance. It was investigated the genetic diversity of 41 inbred lines using SSR and AFLP marker systems (85), emphasizing the substantial diversity among these lines that can be utilized in genetic improvement programs, particularly when maize germplasm with a higher level of *Striga* resistance is limited. Utilizing *Z. diploperennis*, a wild relative of maize, it was created maize inbred lines and identified multiple genes and alleles from the wild relatives that confer resistance against *Striga hermonthica* (86). These findings could serve as valuable resources for maize improvement. Increased grain yield and enhanced durability of *Striga* resistance were demonstrated by TZdEI 352, which was derived from the cross between TZEW Pop DT STR and *Z. diploperennis* (87). Research indicated that additive gene action primarily confers resistance to *Striga* species (88).

Nutritional enhancement

Since maize germplasm is a rich source of genetic diversity, it is possible to exploit alleles associated with nutrient levels in both grain and fodder (89). *Teosinte* carries numerous alleles that can enhance the oil, protein, starch and other nutrient content of kernels in contemporary maize varieties or hybrids. Throughout evolution, the genome of maize's progenitor (*Teosinte*) underwent segmental duplication, resulting in divergent copies of the regulatory opaque-2 gene. Many genes have been identified for nutrient enhancement, including semi-dominant floury mutants (fl-1, fl-2 and fl-3) and recessive opaque mutants (o1, o2, o5, o9-11, o13, o16 and o17) (90). These novel alleles can also serve as unique sources of variation for additional QTL and molecular research.

Traditional yellow maize exhibits considerable natural variation in its carotenoids, particularly lutein and zeaxanthin, but it lacks provitamin A activity (89). Osorno and it was studied the genetic relationships among 10 populations using agronomic and quality trait data, leading to a preliminary characterization of early-maturing germplasm, the measurement of their genetic effects and the clustering of these populations into genetically diverse groups (90). A selection of 24 distinct maize inbred lines (ILs) was made based on the total carotenoid content of their kernels, utilizing 36 microsatellite/SSR markers (91). Natural accumulation of provitamin A carotenoids, such as α-carotene, β-carotene and β-cryptoxanthin, occurs in yellow maize, serving as substrates for the synthesis of lutein and zeaxanthin. To create elite tropical inbred lines with varied carotenoid composition and content, it is beneficial to continuously infuse different temperate and tropical exotic maize germplasm as pre-breeding lines, which can be integrated into nutrient enrichment breeding programs. This process is followed by iterative visual selection for desirable kernel properties and agronomic traits during the early stages of inbreeding. It was found significant genetic variation for total carotenoids in Indian and exotic (HarvestPlus) tropical maize lines, indicating potential for genetic enhancement of this crucial attribute (92). The initial step in breeding for provitamin A-enriched maize germplasm is to determine the genetic diversity in kernel carotenoid concentration within elite germplasm (93).

Conclusion

Exploring the diversity among landraces and wild germplasm has been recommended for maize improvement. Continuous evaluation of germplasm should be conducted to uncover resistant genes against various pests. Genetic diversity assessed through molecular markers is generally more reliable than that evaluated through morphological and biochemical markers; however, morphological markers are often used as a preliminary step in diversity analysis. Among the various statistical methods, principal component analysis (PCA) is predominantly used for quantitative traits, whereas principal coordinate analysis (PCoA) is more suitable for molecular data. Clustering based on D^2 statistics and UPGMA is widely employed for grouping. For precise analysis of genetic diversity, software such as R and SAS are utilized. The deployment of identified genes for yield improvement, nutritional enhancement and resistance to biotic and abiotic stresses contributes to the development of resilient hybrids. Furthermore, the vast diversity in germplasm reveals novel traits such as stay-green types, anthocyanin pigments, low strigolactone secretion and provitamin A, which can enhance crop improvement. Knowledge of divergence and genetic variability between cultivars and landraces, along with the identification of novel genes using advanced molecular techniques and their deployment for maize improvement, will pave the way for successful breeding in the future.

Acknowledgements

The authors acknowledge Tamil Nadu Agricultural University for providing technical support.

Authors' contributions

SS and KRVS conceptualized the draft of this review. SS collected the literature and wrote manuscript. KRVS suggested the ideas for writing and finalized the manuscript. All authors helped in editing and summarizing the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None.

References

1. World Maize Scenario - ICAR-Indian Institute of Maize Research [Internet]. Available from: World Maize Scenario – ICAR-Indian Institute of Maize Research.
2. FAO. Biannual Report on Global Food Markets [Internet]. 2020. Available from: https://doi.org/10.4060/ca9509en
3. Indiastat [Internet]. 2021. Available from: www-indiastat-com.
4. APEDA [Internet]. 2023. Available from: Maize (apeda.gov.in)
5. Soliman ER, El-Shazly HH, Börner A, Badr A. Genetic diversity of a global collection of maize genetic resources in relation to their subspecies assignments, geographic origin and drought tolerance. Breeding Science. 2021;71(3):313-25. https://doi.org/10.1270/jsbbs.20142
6. Begna T. Role and economic importance of crop genetic diversity in food security. International Journal of Agricultural Science and Food Technology. 2021;7(1):164-69.https://doi.org/10.17352/2455-815X.000104
7. Singh BD, Singh AK. Marker-assisted plant breeding: principles and practices. Springer. 2015. https://doi.org/10.1007/978-81-322-2316-0
8. Rao R, Hodgkin T. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell Tissue and Organ Culture. 2002;68:1-19. https://doi.org/10.1023/A:1013359015812
9. Wills DM, Whipple CJ, Takuno S, Kursel LE, et al. From many, one: genetic control of prolificacy during maize domestication. PLoS Genetics. 2013;9(6):e1003604. https://doi.org/10.1371/journal.pgen.1003604
10. Doebley J. The genetics of maize evolution. Annu Rev Genet. 2004;38:37-59. https://doi.org/10.1146/annurev.genet.38.072902.092425
11. Chen Z, Hu K, Yin Y, Tang D, Ni J, et al. Identification of a major QTL and genome-wide epistatic interactions for single vs. paired spikelets in a maize-teosinte F2 population. Molecular Breeding. 2022;42(2):9. https://doi.org/10.1007/s11032-022-01276-x
12. Yang CJ, Samayoa LF, Bradbury PJ, Olukolu BA, Xue W, et al. The genetic architecture of teosinte catalyzed and constrained maize domestication. Proceedings of the National Academy of Sciences. 2019;116(12):5643-52. https://doi.org/10.1073/pnas.1820997116
13. Prasanna B. Diversity in global maize germplasm: characterization and utilization. Journal of Biosciences. 2012;37(5):843-55. https://doi.org/10.1007/s12038-012-9227-1
14. Flint-Garcia SA. Genetics and consequences of crop domestication. J Agric Food Chem [Internet]. 2013 Sep 4 [cited 2024 Sep 12];61(35):8267

-76. Available from: https://pubs.acs.org/doi/10.1021/jf305511d

15. Phundan S, Narayanan S. Biometrical techniques in plant breeding. Kalyani Publishers; 1993.

16. Kamal NM, Gorafi YSA, Abdelrahman M, Abdellatef E, Tsujimoto H. Stay-green trait: A prospective approach for yield potential and drought and heat stress adaptation in globally important cereals. IJMS [Internet]. 2019 Nov 20 [cited 2024 Sep 12];20(23):5837. Available from: https://www.mdpi.com/1422-0067/20/23/5837 https://doi.org/10.3390/ijms20235837

17. Adhikari S, Joshi A, Kumar A, Singh NK. Diversification of maize (*Zea mays* L.) through teosinte (*Zea mays* subsp. *parviglumis* Iltis & Doebley) allelic. Genetic Resources and Crop Evolution. 2021;68 (7):2983-95. https://doi.org/10.1007/s10722-021-01170-z

18. Shayanowako A, Shimelis H, Laing M, Mwadzingeni L. Genetic diversity of maize genotypes with variable resistance to *Striga asiatica* based on SSR markers. Cereal Research Communications. 2018;46:668-78. https://doi.org/10.1556/0806.46.2018.044

19. Nei M. Molecular evolutionary genetics. Columbia University Press; 1987. https://doi.org/10.7312/nei-92038

20. Beaumont MA, Ibrahim KM, Boursot P, Bruford MW. Measuring genetic distance. In A. Karp, P. G. Isaac, D. S. Ingram, Editors. Molecular Tools for Screening Biodiversity. Springer Netherlands; 1998:315-25. https://doi.org/10.1007/978-94-009-0019-6_58

21. Belalia N, Lupini A, Djemel A, Morsli A, Mauceri A, Lotti C, et al. Analysis of genetic diversity and population structure in Saharan maize (*Zea mays* L.) populations using phenotypic traits and SSR markers. Genetic Resources and Crop Evolution; 2019;66:243-57. https://doi.org/10.1007/s10722-018-0709-3

22. Odong T, Van Heerwaarden J, Jansen J, van Hintum TJ, et al. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? Theoretical and Applied Genetics. 2011;123:195-205.https://doi.org/10.1007/s00122-011-1576-x

23. Ward JH. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association [Internet]. 1963 Mar [cited 2024 Feb 14];58(301):236-44. https://doi.org/10.2307/2282967

24. Gasura E, Nyandoro B, Mabasa S, Setimela PS, Kyalo M, Yao N. Breeding strategy for resistance to *Striga asiatica* (L.) Kuntze based on genetic diversity and population structure of tropical maize (*Zea mays* L.) lines. Genetic Resources and Crop Evolution. 2022;1-10.https://doi.org/10.21203/rs.3.rs-336459/v1

25. Ringnér M. What is principal component analysis? Nature Biotechnology; 2008;26(3):303-04. https://doi.org/10.1038/nbt0308-303

26. Jolliffe IT. Principal component analysis for special types of data. Springer; 2002.

27. Zhang W, Gao X, Shi X, Zhu B, Wang Z, Gao H, et al. PCA-based multiple-trait GWAS analysis: a powerful model for exploring pleiotropy. Animals. 2018; 8(12):239. https://doi.org/10.3390/ani8120239

28. Jadhav R, Mehtre S, Patil D, Gite V. Multivariate analysis using D2 and principal component analysis in mung bean [*Vigna radiata* (L.) Wilczek] for study of genetic diversity. Legume Research- An International Journal. 2023;46(1):10-17. https://doi.org/10.36953/ECJ.15472494

29. Gower JC. Principal coordinates analysis. Wiley StatsRef: Statistics Reference Online. 2014;1-7. https://doi.org/10.1002/9781118445112.stat05670.pub2

30. Katayama H, Mori Y, Kuroda M. Variable selection in nonlinear principal component analysis. In: IntechOpen eBooks [Internet]. 2022. Available from: https://doi.org/10.5772/intechopen.103758

31. Singh DP, Marker S. Comparative evaluation of maize (*Zea mays* L.) genotypes based on distinctness, uniformity and stability (DUS) testing using physiological and morphological characters. JANS [Internet]. 2016 Jun 1 [cited 2024 Feb 14];8(2):652-57. https://doi.org/10.31018/jans.v8i2.853

32. Yousuf N, Dar SA, Shikari AB, Dar ZA, Lone AA, Sofi PA, et al. Assessment of genetic diversity at molecular and morphological levels of temperate maize landraces collected from diverse ecological niches in Kashmir. Indian Journal of Genetics and Plant Breeding. 2021;81(04):557-65. https://doi.org/10.31742/IJGPB.81.4.8

33. Al-Naggar A, Soliman A, Hussien M, Mohamed A. Genetic diversity of maize inbred lines based on morphological traits and its association with heterosis. SABRAO Journal of Breeding and Genetics. 2022;54(3). https://doi.org/10.54910/sabrao2022.54.3.11

34. Dube SP, Sibiya J, Kutu F. Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers. Scientific Reports. 2023;13(1):17851. https://doi.org/10.1038/s41598-023-44961-3

35. Ranatunga M, Meenakshisundaram P, Arumugachamy S, et al. Genetic diversity analysis of maize (*Zea mays* L.) inbreds determined with morphometric traits and simple sequence repeat markers. Maydica. 2009;54(1):113.

36. Leal A, Mangolin C, do Amaral Júnior A, Gonçalves L, Scapim C, et al. Efficiency of RAPD versus SSR markers for determining genetic diversity among popcorn lines. Genet Mol Res. 2010;9(1):9-18. https://doi.org/10.4238/vol9-1gmr692

37. Nashath M, Mubarak A, Kumara A. Nashath MN, et al. Exploring genetic diversity of maize (ZEA MAYS l.) accessions in Sri Lanka by employing SSR markers on photosynthetic, canopy architectural and grain yield traits.2023;19(1): 188-203. https://doi.org/10.4038/jas.v19i1.10195

38. Singh D, Nitesh Kushwaha S, Parveen R, Mohanty TA, Suman SK, Kumar R, et al. Assessment of quality protein maize (QPM) inbreds for genetic diversity using morphological characters and simple sequence repeats markers. Electronic Journal of Plant Breeding. 2023;14(3):1276-84. https://doi.org/10.37992/2023.1403.145

39. Oliveira LSD, Schuster I, Novaes E, Pereira WA. SNP genotyping for fast and consistent clustering of maize inbred lines into heterotic groups. Crop Breed Appl Biotechnol [Internet]. 2021 [cited 2024 May 14]; 21 (1):e367121110. https://doi.org/10.1590/1984-70332021v21n1a10

40. Zebire DA. Applications of molecular markers in genetic diversity studies of maize. Nig J Biotechnol [Internet]. 2020 Aug 28 [cited 2024 Feb 8];37(1):101-08. https://doi.org/10.4314/njb.v37i1.11

41. Guo Z, Wang H, Tao J, Ren Y, Xu C, Wu K, et al. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Mol Breeding [Internet]. 2019 Mar [cited 2024 May 14];39(3):37. https://doi.org/10.1007/s11032-019-0940-4

42. Badu-Apraku B, Garcia-Oliveira AL, Petroli CD, Hearne S, et al. Genetic diversity and population structure of early and extra-early maturing maize germplasm adapted to sub-Saharan Africa. BMC Plant Biol [Internet]. 2021 Feb 17 [cited 2024 Sep 12];21(1):96. Available from: https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-021-02829-6 https://doi.org/10.1186/s12870-021-02829-6

43. Wang N, Yuan Y, Wang H, Yu D, Liu Y, Zhang A, et al. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. Sci Rep [Internet]. 2020 Oct 1 [cited 2024 May 10];10(1):16308. https://doi.org/10.1038/s41598-020-73321-8

44. Ott A, Liu S, Schnable JC, Yeh CT 'Eddy,' Wang KS, Schnable PS. tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Research [Internet]. 2017 Dec 1 [cited 2024 May 10];45(21):e178-e178. https://doi.org/10.1093/nar/gkx853

45. Rabieyan E, Bihamta MR, Moghaddam ME, Alipour H, et al. Analysis of genetic diversity and genome-wide association study for drought tolerance related traits in Iranian bread wheat. BMC Plant Biol [Internet]. 2023 Sep 15 [cited 2024 Sep 12];23(1):431. Available from: https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-023-04416-3. https://doi.org/10.1186/s12870-023-04416-3

46. Shikha K, Madhumal Thayil V, Shahi J, Zaidi P, Seetharam K, Nair SK, et al. Genomic-regions associated with cold stress tolerance in

Asia-adapted tropical maize germplasm. Scientific Reports. 2023;13(1):6297. https://doi.org/10.1038/s41598-023-33250-8

47. Beló A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, et al. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol Genet Genomics [Internet]. 2008 Jan [cited 2024 May 10];279(1):1-10. https://doi.org/10.1007/s00438-007-0289-y

48. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: Complexity, diversity and dynamics. Science [Internet]. 2009 Nov 20 [cited 2024 May 10];326(5956):1112-15. Available from: The B73 Maize Genome: Complexity, Diversity and Dynamics | Science. https://doi.org/10.1126/science.1178534

49. Farooqi MQU, Nawaz G, Wani SH, Choudhary JR, Rana M, et al. Recent developments in multi-omics and breeding strategies for abiotic stress tolerance in maize (*Zea mays* L.). Front Plant Sci [Internet]. 2022 Sep 23 [cited 2024 Feb 29];13:965878. https://doi.org/10.3389/fpls.2022.965878

50. Liu S, Wang X, Wang H, Xin H, Yang X, Yan J, et al. Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. Springer NM, editor. PLoS Genet [Internet]. 2013 Sep 26 [cited 2024 Feb 29];9 (9):e1003790. https://doi.org/10.1371/journal.pgen.1003790

51. Song X, Zhang Y, Wu F, Zhang L. Genome-wide analysis of the maize (*Zea mays* L.) CPP-like gene family and expression profiling under abiotic stress. Genet Mol Res. 2016;15(3). https://doi.org/10.4238/gmr.15038023

52. Wen W, Li D, Li X, Gao Y, Li W, Li H, et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nature Communications [Internet]. 2014 Mar 17;5(1):3438. https://doi.org/10.1038/ncomms4438

53. Heslot N, Yang HP, Sorrells ME, Jannink JL. Genomic selection in plant breeding: a comparison of models. Crop Science. 2012;52 (1):146-60. https://doi.org/10.2135/cropsci2011.06.0297

54. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, et al. Genomic selection in plant breeding: Methods, models and perspectives. Trends in Plant Science [Internet]. 2017 Nov [cited 2024 May 13];22(11):961-75.https://doi.org/10.1016/j.tplants.2017.08.011

55. Marinho CD, Coelho IF, Peixoto MA, Carvalho Junior GAD, Resende Junior MFRD. Genomic selection as a tool for maize cultivars development. RBMS [Internet]. 2022 Apr 29 [cited 2024 May 13];21.https://doi.org/10.18512/rbms2022v21e1285

56. Diaw Y, Tollon-Cordet C, Charcosset A, Nicolas SD, Madur D, et al. Genetic diversity of maize landraces from the South-West of France. PLoS One. 2021; 16(2):e0238334. https://doi.org/10.1371/journal.pone.0238334

57. Roy PR, Haque MA, Ferdausi A, Al Bari MA. Genetic variability, correlation and path co-efficients analyses of selected maize (*Zea mays* L.) genotypes. Fundamental and Applied Agriculture. 2018;3 (1):382-89. https://doi.org/10.5455/faa.284502

58. Pritchard JK, Wen X, Falush D. Documentation for structure software: Version 2.3. University of Chicago, Chicago, IL. 2010;1-37.

59. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, et al. An overview of STRUCTURE: applications, parameter settings and supporting software. Frontiers in Genetics [Internet]. 2013;4. https://doi.org/10.3389/fgene.2013.00098

60. Tools: Flapjack - high-throughput genotype visualizer » Seeds of Discovery [Internet]. Available from: Tools: Flapjack – high-throughput genotype visualizer » Seeds of Discovery

61. Zinn KE, Tunc-Ozdemir M, Harper JF. Temperature stress and plant sexual reproduction: uncovering the weakest links. Journal of Experimental Botany; 2010;61(7):1959-68. https://doi.org/10.1093/jxb/erq053

62. Wang Q, Liao Z, Zhu C, Gou X, Liu Y, Xie W, et al. Teosinte confers specific alleles and yield potential to maize improvement. Theor Appl Genet [Internet]. 2022 Oct [cited 2024 Feb 14];135(10):3545-62. https://doi.org/10.1007/s00122-022-04199-5

63. Patil N, Shanthakumar G, Biradar B. Genetic diversity analysis of Maize (*Zea mays* L.) inbred lines under drought stress using grain yield and yield related traits. Journal of Crop and Weed. 2021;17 (3):192-98. https://doi.org/10.22271/09746315.2021.v17.i3.1509

64. SathyaSheela KRV, Selvakumar T, Narayanan SL, Radhajeyalakshmi R. Analyzing the variability in maize (*Zea mays* L.) genotypes using principal component analysis under varied ecosystems. IJECC [Internet]. 2023 Nov 3 [cited 2024 Jun 3];13(11):2172-80. https://doi.org/10.9734/ijecc/2023/v13i113379

65. Dubey L, Prasanna B, Ramesh B. Analysis of drought tolerant and susceptible maize genotypes using SSR markers tagging candidate genes and consensus QTLs for drought tolerance. Indian Journal of Genetics and Plant Breeding. 2009;69(04):344-51.

66. Tesfaye K, Zaidi P, Gbegbelegbe S, Boeber C, Rahut DB, et al. Climate change impacts and potential benefits of heat-tolerant maize in South Asia. Theoretical and Applied Climatology. 2017130:959-70. https://doi.org/10.1007/s00704-016-1931-6

67. Heat Stress Tolerant Maize for Asia (HTMA) [Internet]. Available from: Heat Stress Tolerant Maize for Asia (HTMA) – CIMMYT

68. Kulkarni AP, Tripathi MP, Gautam D, Koirala KB, Kandel M, et al. Impact of adoption of heat-stress tolerant maize hybrid on yield and profitability: Evidence from Terai region of Nepal. Front Sustain Food Syst [Internet]. 2023 Apr 3 [cited 2024 May 26];7:1101717. https://doi.org/10.3389/fsufs.2023.1101717

69. Nelimor C, Badu-Apraku B, Tetteh AY, N'guetta AS. Assessment of genetic diversity for drought, heat and combined drought and heat stress tolerance in early maturing maize landraces. Plants. 2019;8(11):518. https://doi.org/10.3390/plants8110518

70. Cairns JE, Crossa J, Zaidi P, Grudloyma P, Sanchez C, Araus JL, et al. Identification of drought, heat and combined drought and heat tolerant donors in maize. Crop Science. 2013;53(4):1335-46. https://doi.org/10.2135/cropsci2012.09.0545

71. Alam MA, Seetharam K, Zaidi PH, Dinesh A, Vinayan MT, Nath UK. Dissecting heat stress tolerance in tropical maize (*Zea mays* L.). Field Crops Research. 2017;204:110-19. https://doi.org/10.1016/j.fcr.2017.01.006

72. Yi Q, Álvarez-Iglesias L, Malvar RA, Romay M, Revilla P. A worldwide maize panel revealed new genetic variation for cold tolerance. Theoretical and Applied Genetics. 2021;134:1083-94. https://doi.org/10.1007/s00122-020-03753-3

73. Wani SH, Samantara K, Razzaq A, Kakani G, Kumar P. Back to the wild: mining maize (*Zea mays* L.) disease resistance using advanced breeding tools. Molecular Biology Reports. 2022;49 (6):5787-803. https://doi.org/10.1007/s11033-021-06815-x

74. Maxted N, Kell S. Establishment of a global network for the *in situ* conservation of crop wild relatives: Status and needs. Thematic Background Study [Internet]. 2009 [cited 2024]. Available from: https://www.fao.org/4/i1500e/i1500e18a.pdf

75. Lennon JR, Krakowsky M, Goodman M, Flint-Garcia S, Balint-Kurti PJ. Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. Crop Science. 2016;56(1):209-18. https://doi.org/10.2135/cropsci2014.07.0468

76. Adhikari S, Joshi A, Kumar A, Singh NK, Jaiswal JP, Jeena AS, et al. Identification of QTL for banded leaf and sheath blight in teosinte-derived maize population. Agricultural Research. 2022;11(2):155-63. https://doi.org/10.1007/s40003-021-00567-7

77. de Lange ES, Balmer D, Mauch-Mani B, Turlings TC. Insect and pathogen attack and resistance in maize and its wild ancestors. The Teosintes. 2014. https://doi.org/10.1111/nph.13005

78. Joshi A, Adhikari S, Singh NK. Mapping genomic regions for red flour beetle (*Tribolium castaneum* (Herbst)) resistance in teosinte (*Zea mays* L. subsp. *parviglumis* HH Iltis & Doebley) derived maize backcross inbred line population. Genetic Resources and Crop

Evolution. 2021;68:1529-44. https://doi.org/10.1007/s10722-020-01083-3

79. Kasoma C, Shimelis H, Laing MD, Shayanowako AI, Mathew I. Revealing the genetic diversity of maize (*Zea mays* L.) populations by phenotypic traits and DArTseq markers for variable resistance to fall armyworm. Genetic Resources and Crop Evolution. 2021;68:243-59. https://doi.org/10.1007/s10722-020-00982-9

80. Prischmann D, Dashiell K, Schneider D, Eubanks M. Evaluating Tripsacum-introgressed maize germplasm after infestation with western corn rootworms (Coleoptera: Chrysomelidae). Journal of Applied Entomology. 2009;133(1):10-20. https://doi.org/10.1111/j.1439-0418.2008.01311.x

81. Amusan IO, Rich PJ, Menkir A, Housley T, Ejeta G. Resistance to *Striga hermonthica* in a maize inbred line derived from *Zea diploperennis*. New Phytologist. 2008;178(1):157-66. https://doi.org/10.1111/j.1469-8137.2007.02355.x

82. Adewale SA, Badu-Apraku B, Akinwale RO, Paterne AA, Gedil M, et al. Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. BMC Plant Biology. 2020;20:1-16. https://doi.org/10.1186/s12870-020-02360-0

83. Karaya H, Njoroge K, Mugo S, Ariga E, Kanampiu F, Nderitu J. Determination of levels of Striga germination stimulants for maize gene bank accessions and elite inbred lines. International Journal of Plant Production. 2012;6(2):209-23. https://doi.org/10.22069/IJPP.2012.776

84. Ejeta G. The striga scourge in africa: a growing pandemic. Integrating New Technologies for Striga Control. 2007;3-16.

85. Menkir A, Kling J, Badu-Apraku B, Ingelbrecht I. Molecular marker-based genetic diversity assessment of *Striga*-resistant maize inbred lines. Theoretical and Applied Genetics. 2005;110:1145-53. https://doi.org/10.1007/s00122-005-1946-3

86. Shaibu AS, Badu-Apraku B, Ayo-Vaughan MA. Enhancing drought tolerance and *Striga hermonthica* resistance in maize using newly derived inbred lines from the wild maize relative, *Zea diploperennis*. Agronomy. 2021;11(1):177. https://doi.org/10.3390/agronomy11010177

87. Akaogu I, Badu-Apraku B, Adetimirin V, Vroh-bi I, Oyekunle M, Akinwale R. Genetic diversity assessment of extra-early maturing yellow maize inbreds and hybrid performance in *Striga*-infested and *Striga*-free environments. The Journal of Agricultural Science. 2013;151(4):519-37. https://doi.org/10.1017/S0021859612000652

88. Pfunye A, Rwafa R, Mabasa S, Gasura E. Genome-wide association studies for *Striga asiatica* resistance in tropical maize. International Journal of Genomics. 2021. https://doi.org/10.1155/2021/9979146

89. Langyan S, Bhardwaj R, Kumari J, Jacob SR, Bisht IS, Pandravada SR, et al. Nutritional diversity in native germplasm of maize collected from three different fragile ecosystems of India. Frontiers in Nutrition. 2022.;9:812599. https://doi.org/10.3389/fnut.2022.812599

90. Osorno J, Carena M. Creating groups of maize genetic diversity for grain quality: Implications for breeding. Maydica. 2008;53(2):131.

91. Sivaranjani R, Santha I, Pandey N, Vishwakarma AK, Nepolean T, Hossain F. Microsatellite-based genetic diversity in selected exotic and indigenous maize (*Zea mays* L.) inbred lines differing in total kernel carotenoids. Indian Journal of Genetics and Plant Breeding. 2014;74(01):34-41. https://doi.org/10.5958/j.0975-6906.74.1.005

92. Tiwari A, Prasanna B, Hossain F, Guruprasad K. Analysis of genetic variability for kernel carotenoid concentration in selected maize inbred lines. Indian Journal of Genetics and Plant Breeding. 2012;72(01):1-6.

93. Sivaranjani R, Prasanna BM, Hossain F, Santha IM. Genetic variability for total carotenoid concentration in selected tropical maize (*Zea mays*) inbred lines. Indian J Agri Sci [Internet]. 2013 Apr. 9 [cited 2024];83(4). Available from: https://epubs.icar.org.in/index.php/IJAgS/article/view/28851.

94. Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;857-71. https://doi.org/10.2307/2528823

95. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences. 1979;76(10):5269-73.https://doi.org/10.1073/pnas.76.10.5269

96. Jaccard P. Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Na. 1908;44:223-70.

97. Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin [Internet]. 1958;38:1409-38. Available from: A statistical method for evaluating systematic relationships | Semantic Scholar

98. Wright S. Evolution and the genetics of populations, volume 4: variability within and among natural populations. Vol. 4. University of Chicago press; 1984.

99. Tools: CurlyWhirly - multi-dimensional data visualizer» Seeds of Discovery [Internet]. Available from: https://seedsofdiscovery.org/catalogue/curlywhirly-multi-dimensional-data-visualizer

100. Tools: kdsmart-hand-held-data-capture-application» Seeds of Discovery. [Internet]. Available from: Tools: kdsmart-hand-held-data-capture-application » Seeds of Discovery