



## OPEN ACCESS

## EDITED BY

Patricio Hinrichsen,  
Agricultural Research Institute (Chile), Chile

## REVIEWED BY

Dusan Stanisavljevic,  
of Field And Vegetable Crops Novi Sad  
(IFVCNS), Serbia  
Hongjun Yong,  
Chinese Academy of Agricultural Sciences  
(CAAS), China

## \*CORRESPONDENCE

Thomas Lübberstedt  
✉ thomasl@iastate.edu

RECEIVED 20 May 2023

ACCEPTED 10 July 2023

PUBLISHED 27 July 2023

## CITATION

Ledesma A, Ribeiro FAS, Uberti A,  
Edwards J, Hearne S, Frei U and  
Lübberstedt T (2023) Molecular  
characterization of doubled haploid  
lines derived from different cycles  
of the Iowa Stiff Stalk Synthetic  
(BSSS) maize population.  
*Front. Plant Sci.* 14:1226072.  
doi: 10.3389/fpls.2023.1226072

## COPYRIGHT

© 2023 Ledesma, Ribeiro, Uberti, Edwards,  
Hearne, Frei and Lübberstedt. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Molecular characterization of doubled haploid lines derived from different cycles of the Iowa Stiff Stalk Synthetic (BSSS) maize population

Alejandro Ledesma<sup>1</sup>, Fernando Augusto Sales Ribeiro<sup>1</sup>,  
Alison Uberti<sup>1</sup>, Jode Edwards<sup>2</sup>, Sarah Hearne<sup>3</sup>, Ursula Frei<sup>1</sup>  
and Thomas Lübberstedt<sup>1\*</sup>

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, United States, <sup>2</sup>USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA, United States, <sup>3</sup>International Maize and Wheat Improvement Center (CIMMYT), El Batán, Texcoco, Mexico

Molecular characterization of a given set of maize germplasm could be useful for understanding the use of the assembled germplasm for further improvement in a breeding program, such as analyzing genetic diversity, selecting a parental line, assigning heterotic groups, creating a core set of germplasm and/or performing association analysis for traits of interest. In this study, we used single nucleotide polymorphism (SNP) markers to assess the genetic variability in a set of doubled haploid (DH) lines derived from the unselected Iowa Stiff Stalk Synthetic (BSSS) maize population, denoted as C0 (BSSS(R)C0), the seventeenth cycle of reciprocal recurrent selection in BSSS (BSSS(R)C17), denoted as C17 and the cross between BSSS(R)C0 and BSSS(R)C17 denoted as C0/C17. With the aim to explore if we have potentially lost diversity from C0 to C17 derived DH lines and observe whether useful genetic variation in C0 was left behind during the selection process since C0 could be a reservoir of genetic diversity that could be untapped using DH technology. Additionally, we quantify the contribution of the BSSS progenitors in each set of DH lines. The molecular characterization analysis confirmed the apparent separation and the loss of genetic variability from C0 to C17 through the recurrent selection process. Which was observed by the degree of differentiation between the C0\_DHL versus C17\_DHL groups by Wright's F-statistics (FST). Similarly for the population structure based on principal component analysis (PCA) revealed a clear separation among groups of DH lines. Some of the progenitors had a higher genetic contribution in C0 compared with C0/C17 and C17 derived DH lines. Although genetic drift can explain most of the genetic structure genome-wide, phenotypic data provide evidence that selection has altered favorable allele frequencies in the BSSS maize population through the reciprocal recurrent selection program.

## KEYWORDS

zea mays L., diversity, genetic resources, homozygous lines, genetic diversity

## Introduction

The maize Iowa Stiff Stalk Synthetic (BSSS) population has undergone recurrent selection since 1939. This population was developed by intermating 16 inbred lines selected for superior stalk quality (Sprague and Jenkins, 1943). The C0 base population was subjected to multiple cycles of recurrent selection. Currently, C19 is available. The BSSS maize population has been under recurrent selection for increased grain yield, low grain moisture at harvest and increased resistance to root and stalk lodging. Phenotypic and genotypic changes have been observed in this population (Messmer et al., 1991; Labate et al., 1999; Hagdorn et al., 2003; Edwards, 2011; Gerke et al., 2015), suggesting loss of genetic variability from C0 to more advanced cycles of selection. As noted by Ledesma (2020) when they evaluated the level of phenotypic diversity and identified significant SNPs by GWAS in different cycles of recurrent selection of the BSSS population based on doubled haploid (DH) lines. Alleles present in a heterogeneous population of heterozygous individuals can be fixed in homozygous and homogenous DH lines and part of the genetic diversity in a population can be harnessed for breeding by production of DH lines (Böhm et al., 2017). However, the success of this approach relies on the choice of promising populations and extensive characterization of the produced DH lines (Böhm et al., 2017). The combination of DH technology with high-throughput genotyping drives progress in major maize breeding programs today (Andorf et al., 2019) and has been applied in this study to understand the evolution and genotypic composition of different cycles of BSSS maize population.

Molecular markers like single nucleotide polymorphisms (SNPs) have proven to be valuable for the characterization of maize germplasm and their application becoming more feasible over the past two decades due to the availability of new, high density and affordable genotyping technologies (Lu et al., 2009). Useful measures of the quality of genetic markers' polymorphisms are the expected heterozygosity ( $H_{exp}$ ). Expected heterozygosity ( $H_e$ ) is defined as the probability that any two alleles at a single locus, chosen randomly from the population, are different from each other (Nei and Roychoudhury, 1974; Nei, 1978).

The genetic relationship based on genetic distance was first defined by Nei (1973) as the difference between two samples that can be described by allelic variation, meaning that genotypes with many similar genes have a smaller genetic distance between them. The degree of genetic differentiation using the fixation index ( $F_{ST}$ ) is a standard measure for the degree of genetic differentiation among subpopulations (Wright, 1951). The  $F_{ST}$  provides important insights into the evolutionary processes that influence the structure of genetic variation within and among populations (Holsinger and Weir, 2009). The  $F_{ST}$  estimates can identify regions of the genome that have been targeted for selection (Beckett et al., 2017; Guo et al., 2021; Wijayasekara and Ali, 2021). The comparison of  $F_{ST}$  from different genome regions can provide insights into populations demographic history (Holsinger and Weir, 2009).

Characterizing and understanding the genetic diversity and relationships of lines within a breeding program is essential for germplasm improvement (Andorf et al., 2019). Molecular markers

have been used to estimate the relative strengths of evolutionary forces: mutation, natural selection, migration and genetic drift (Ouborg et al., 1999) and a possible loss of genetic diversity in specific populations, including BSSS (Gerke et al., 2015). Gerke et al. (2015), when evaluating different cycles of selection in a recurrent selection program, found that the populations steadily decreased in genetic diversity within populations and increased in genetic differentiation between populations mainly due to genetic drift and selection. According to the same authors, the C0 population has drifted away from the BSSS founders, despite the absence of intentional selection during the creation and maintenance of C0. In our study, we used different methods proposed as genetic diversity and differentiation measures using genotypic information. Additionally, we used developed DH lines instead of individual heterozygous plants representing the different cycles of the BSSS population.

Population structure is referred to as any form of relatedness among subgroups within the overall sample, including ancestry differences or cryptic relatedness (Sul et al., 2018). Population structure analysis involves grouping of individuals into subpopulations based on shared genetic variants and can be assessed through principal component analysis (PCA). PCA can identify differences in ancestry among populations and individuals, regardless of the historical patterns underlying population structure (Price et al., 2006; Zhu and Yu, 2009), since PCA clusters individuals based on the number of markers that are identical by state among them. Based on this grouping and relationship information among individuals, plant breeders can direct crosses, avoiding the mating closely related individuals and providing a reduction in inbreeding in their breeding programs. For instance, kinship coefficients have been used to estimate the genetic relationships within populations and to estimate the genetic contribution of a set of parents to its descendants (Yang et al., 2011; Ertiro et al., 2017; Wegary et al., 2019). Therefore, the estimation of kinship coefficients represents a way to utilize breeding resources more efficiently (Beckett et al., 2017).

An identity by descent (IBD) segment refers to DNA segments descended from common ancestors and could be useful to estimate the genetic relationships in a population. IBD occurs when identical alleles are inherited from a common ancestor and constitutes a measure of the degree of relationship between individuals (Wright, 1922). The estimation of the degree of the relationship depends on the description of an ancestral population, which by definition, is assumed to be the base from where past ancestry is no longer accounted (Wright, 1922). With the advent of high-throughput genotyping technologies, IBD segments can be estimated at a molecular scale. The identification of shared segments in the genome and haplotype information has been used for a range of purposes, including the quantification of inbreeding (Keller et al., 2011), identification of patterns of inheritance (Kirin et al., 2010), genotype imputation and haplotype inference (Browning and Browning, 2007), genetic characterization and diversity analysis (Nelson et al., 2008), the genetic contribution of a set of founder lines in commercial maize breeding programs (Coffman et al., 2019), and to improve the accuracy of genome-wide association analysis (GWAS; Maldonado et al., 2019) and genomic prediction (Won et al., 2020).

In this study of the BSSS, we propose to determine how much of the genomic variation in C0 has been lost during the selection process. C0 may be a reservoir of untapped favorable genetic diversity for previously unselected traits. Developing DH lines from earlier cycles of selection could be an alternative approach to conventional breeding for introduction of diversity into related elite lines. Genetic heterogeneity and high genetic load present in C0 could be overcome by production of DH lines (Böhm et al., 2017) to unlock genetic diversity. Diversity may have been lost not only due to selection can also be attributed to genetic drift or genetic hitchhiking effects, since no new genetic material was intentionally introduced into the BSSS population.

Our overall question in this and a companion paper Ledesma (2020) was whether potentially useful genetic diversity is available in earlier cycles of selection in the recurrent selection process, which may be more accessible sources of alleles compared with founding non-adapted landraces and other such genetic resources. Here, we used SNP markers to i) estimate and compare the genetic diversity within different subsets of DH lines derived from the BSSS maize population after different cycles of selection, ii) determine, if genetic diversity was lost from C0 to C17, iii) assess the genetic relationships and genetic divergence within and among the cycles of selection, and iv) perform a haplotype analysis based on IBD segments to quantify the contribution of the progenitors to each set of DH lines.

## Materials and methods

### Breeding populations

Three synthetic populations BSSS, BSSS(R)C17, and BSSS/BSSS (R)C17 representing different cycles of selection in the reciprocal recurrent selection program with BSSS, and the Iowa Corn Borer Synthetic number 1 (BSCB1) were used to develop DH lines. The synthetic BSSS corresponds to the unselected base population (C0) formed by intermating 16 inbred lines selected for above average stalk quality in 1934 (Sprague, 1946). The C0 seed used came from subsequent cycles of seed multiplication in C0 for maintenance over time. The BSSS(R)C17 (C17) population corresponds to the seventeenth cycle of reciprocal recurrent selection with BSCB1 (Penny and Eberhart, 1971; Lamkey, 1992; Keeratinijakal and Lamkey, 1993; Edwards, 2011). Finally, BSSS/BSSS(R)17 was created by crossing plants from BSSS with plants in BSSS(R)C17 and intermating to create the BSSS/BSSS(R)C17 population (C0/C17). We also included in this study 14 (A3G-3-3-1-3, CI 540, I-159, IL12E, Oh 3167B, Os 420, Tr 9-1-1-6, WD 456, I224, LE 23, 461, Hy, AH83, CI 187-2) of the 16 known progenitors of the BSSS, plus the two parents (Fe and B2) of the F1B1 line. That is, a total of 16 progenitors were included in the study. Seed from the progenitor lines CI 617 and F1B1 were not available.

### DH line development

Randomly selected individuals within each population were pollinated with a maternal haploid inducer BHI301 (Almeida et al.,

2020) in an isolation field to generate the haploid seed. Seed produced from these plants was screened and kernels expressing the *R-nj* marker gene in the endosperm, but not in the embryo, were classified as haploid kernels. The haploid seed was germinated in plug trays in the Department of Agronomy greenhouse. Once seedlings developed 2-3 leaves, a colchicine treatment was applied following the protocol used by the DH Facility at ISU (Vanous et al., 2017). Two days after the colchicine treatment, haploid seedlings were transplanted in the field at the Agricultural Engineering and Agronomy Research Farm, Boone, IA. At flowering stage, putative DH<sub>0</sub> plants shedding pollen were self-pollinated to produce DH<sub>1</sub> seed. Seed multiplication was performed during subsequent generations and lines were screened for uniformity and discarded if they were segregating or variable. In total, 132 DH lines from BSSS(R)C0 (C0\_DHL), 185 DH lines from BSSS(R)17 (C17\_DHL), and 170 DH lines from BSSS(R)C0/BSSS(R)17 (C0/C17\_DHL) were obtained. The DH lines were developed by the DH Facility at ISU (<http://www.plantbreeding.iastate.edu/DHF/DHF.htm>).

### Genotyping and quality control

Genomic DNA was extracted from DH line seedlings established in a greenhouse. Leaf tissue samples from three plants per DH line were collected at the 3-4 leaf developmental stage, and DNA extraction was done using the standard International Maize and Wheat Improvement Center (CIMMYT) laboratory protocol (Warburton, 2005). Genotyping was carried out using the Diversity Arrays Technology sequencing (DArT-seq) method (Kilian et al., 2012) provided by the Genetic Analysis Service for Agriculture (SAGA) laboratory at CIMMYT. DArT-seq is a high-throughput, robust, reproducible, and cost-effective genotyping technology based on genome complexity reduction using a combination of tailored restriction enzymes, followed by multiplexed sequencing of resulting libraries to simultaneously assay thousands of markers across the genome (Sansaloni et al., 2011). Across the samples assessed a total of 51,418 SNP markers were generated, of these 32,929 SNP markers were successfully aligned to the B73 RefGen\_v4 (Jiao et al., 2017). Monomorphic and multi-allelic markers were removed. Un-imputed data without filtering for minor allele frequency (MAF) were used for further analyses.

The inbred line B73 was used as technical control and was repeated in seven separate plates to verify assay reproducibility. The resulting SNP core set was 24,885 SNP markers corresponding to 487 DH lines (132 C0\_DHLs, 170 C0/C17\_DHLs, 185 C17\_DHLs) and 15 progenitors). The progenitor CI 187-2 was omitted because of heterozygosity greater than 8.8% (not expected in inbred lines) and was removed from further analyses. After this point, only 15 progenitors with low heterozygosity were used in the study.

### Genotypic data analysis

Minor allele frequency analysis for each locus across the genotypes was calculated using the 24,885 SNP markers with the function 'Geno summary' analysis tool in the software TASSEL

v.5.2.64 (Bradbury et al., 2007). The expected heterozygosity ( $H_{exp}$ ) was calculated to quantify the genetic variation in the maize lines sampled. The expected heterozygosity is defined as the probability that two alleles randomly chosen from the test sample are different (Nei, 1978). The expected heterozygosity was calculated using the R package “Poppr” (Kamvar et al., 2014), with the following formula:  $H_{exp} = \left(\frac{n}{n-1}\right) \left(1 - \sum_{i=1}^k p_i^2\right)$ , where  $p$  is the allele frequency at a given locus, which goes from  $i$  to  $k$ , and  $n$  is the number of observed alleles for each locus (Nei, 1978).

The computation of dissimilarity coefficients or Euclidean genetic distance (Gower and Legendre, 1986) between DH lines and progenitor groups was performed with the 24,885 SNP markers using the R package “Poppr” (Kamvar et al., 2014). The genetic distances were calculated based on the average genetic distance of all lines within each other group. Cluster analyses were performed to subdivide the three sets of DH lines and the progenitor group into genetic subgroups using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Finally, dendrograms were constructed based on genetic distances using the visualization software Interactive Tree of Life (iTOL; Letunic and Bork, 2019).

To assess the degree of genetic differentiation between the groups of DH lines and the progenitors, we used the Wright’s  $F_{ST}$  statistics ( $F_{ST}$ ) on a per locus basis using the methodology described by Weir and Cockerham (1984), which accounts for unequal population sizes and sampling variances since heterozygous loci are weighted by the number of alleles observed in each population. The R package “hierfstat” (Goudet, 2005) was used to obtain estimates of  $F_{ST}$ . The  $F_{ST}$  values can range from zero to one, where high  $F_{ST}$  values showed a considerable difference in the allele frequency among two populations.

The pairwise relative kinship for all 487 DH lines and the 15 progenitors was estimated based on the 24,885 SNP markers using the software TASSEL v.5.2.64 (Bradbury et al., 2007) using the centered\_IBS method (Endelman and Jannink, 2012). The relative kinship reflects the approximate degree of identity between two given individuals over the average probability of identity between two random individuals (Yu et al., 2006). The pairwise relative kinship was used to measure the genetic resemblance among individuals. A relative kinship close to zero indicates no relationship, and values close to one indicate a close relationship. Marker-based kinship coefficients show the relationship among lines based on genotypic information and rely on the marker allele frequencies in the reference population, which in practice is not known (Wang, 2014). However, 15 of the 16 progenitors of BSSS are known. These estimates commonly use the sample of genotyped individuals as the reference population, resulting in estimates that two homologous genes within or between individuals are shared by descent (Wang, 2014). Marker-based estimation of kinship coefficients can result in negative values. Wang (2014) states that the kinship coefficient’s negative values could be interpreted as a lower probability that two homologous alleles are shared by descent compared with the probability that two alleles are taken at random from the reference population.

The 487 DH lines and the 15 progenitors were known to belong to the four subpopulations BSSS(R)C0, BSSS(R)C17, BSSS(R)C0/C17 and the progenitor groups, respectively. To examine the overall

population structure across all lines, we performed a principal component analysis (PCA). PCA analysis allows the classification of individuals into genetically similar groups. PCA relies on reducing dimensionality by using principal components to maximize genetic variability (Price et al., 2006). Each principal component will account for a percentage of the total genetic variance by grouping the individuals into clusters with similar genetic information. After reducing dimensionality, a linear regression model was fitted to each of the axes of variation, and the residuals were extracted to compute associations (Price et al., 2006). PCA avoids any prior information about individual ancestries, the population of origin, and assumptions about the data, handling genome-wide data for thousands of individuals (Paschou et al., 2007). PCA was performed using the software GAPIT v.3 (Lipka et al., 2012). Bayesian Information Criterion (BIC; Schwarz, 1978) was used to identify the optimal number of principal components by selecting the lowest BIC model. The principal component results were used to display the first two principal components in R software (R Core Team, 2021).

The average linkage disequilibrium (LD) decay among SNP markers for each chromosome was determined in each group of DH lines using the squared Pearson correlation coefficient ( $r^2$ ) among alleles at two loci, for all possible combinations of alleles, and then weighting them according to the allele frequency. P-values were determined by a two-sided Fisher’s Exact test (Bradbury et al., 2007). The option “Full Matrix LD” on TASSEL v.5.2.64 was used to calculate LD for every combination of sites in the alignment (Bradbury et al., 2007). The resulting data were imported into R (R Core Team, 2021) to create LD decay plots and fit a smooth line using Hill and Weir expectations of  $r^2$  among adjacent sites (Hill and Weir, 1988).

To quantify the progenitors genetic contributions to the different sets of DH lines, we used high-resolution detection of identity by descent (IBD) segments. An IBD segment refers to DNA segments descended from common ancestors. IBD occurs when identical alleles are inherited from a common ancestor and could be used to estimate the genetic contribution. Estimation of IBD segments with genotypic data allows the quantification of the proportion of the covered genome descended from each progenitor. For the genetic contribution and the average LD decay among SNP marker analysis, a different filtering process of the genotypic data was conducted to have the most reliable SNP markers and ensure genotype concordance. From the 32,929 SNP markers successfully called within the B73 RefGen\_v4 (Jiao et al., 2017). SNP markers with missing information rate above 10%, duplicated and monomorphic markers were removed in TASSEL v.5.2.64 (Bradbury et al., 2007). Genotypes were phased and imputed by using Beagle v.5.1 (Browning et al., 2018). Physical distance for each marker was converted to genetic distance using a dense 0.2 cM resolution map (Ogut et al., 2015), with on average 1385.6 kb per cM. After completing filtering and quality control, the genotypic data file contained 10,344 SNP markers for each of the 502 genotypes (487 DH lines and 15 progenitors) covering 2102.7 Mb (1,517.5 cM) of the genome and with one marker per 203.2 kb on average. The SNP markers not included in an IBD segment were referred to as non-IBD markers, while those within the IBD segment were labeled with the progenitor sharing the segment. The proportion of the genome descended from a progenitor was calculated by dividing the total number of SNP markers classified

as IBD by the total number of polymorphic SNP markers. Regions in the genome (IBD segments) that have been inherited from the progenitor were identified with the identity by descent linkage disequilibrium (IBDL) program v.3.38 (Han and Abney, 2011; Han and Abney, 2013). The IBDL program uses a probabilistic approach with a hidden Markov model to estimate IBD segments in pairs of individuals. The IBDL program further expresses the emission probability conditioned on the true genotype of  $n$  previous loci to account for linkage disequilibrium (Han and Abney, 2011). IBD segments were constrained for each pair of individuals to have a minimum length of 350 kb, have more than 10 SNP markers and SNP markers with an IBD probability above 70%. These parameters force the segment to be a long IBD segment, avoiding segments formed by an occasional genotyping error or missing genotype occurring in otherwise-unbroken segments that could underestimate IBD segments for each pair of individuals (McQuillan et al., 2008).

## Results

The initial number of SNP markers in the DArT-seq data set was 51,418. A total of 32,929 SNP markers were successfully called within the B73 RefGen\_v4 (Jiao et al., 2017). After removing monomorphic and multi-allelic markers, the final SNP marker data set included 24,885 SNPs distributed across the ten chromosomes. The SNP density varied among chromosomes ranged from 3,976 to 1,688 markers on chromosome 1 and 10, respectively (Table 1). Heterozygosity varied from 1.2% on chromosomes 2 and 7 to 1.6% on chromosome 9, with a mean value of 1.3% across the ten chromosomes. We found heterozygous loci among the DHLs which ranged from 0.40 (C17\_DHL045) to 2.24% (C0/C17\_DHL146; Supplemental Table S1).

TABLE 1 Genotypic data summary for the 24,885 SNP markers and the entire panel of DH lines derived from different BSSS selection cycles.

Chromosome number	Number of SNP markers	Heterozygosity rate (%)
1	3,976	1.3
2	2,978	1.3
3	2,721	1.2
4	2,453	1.5
5	2,798	1.3
6	1,929	1.4
7	2,203	1.2
8	2,105	1.5
9	2,034	1.6
10	1,688	1.4
Genome-wide	24,885	1.3

## Molecular characterization analysis

The 24,885 SNP markers were polymorphic with a MAF greater than zero (Figure 1). The average MAF was 0.19, 0.16, 0.13 and 0.07 in the progenitor, C0\_DHL, C0/C17\_DHL and C17\_DHL groups, respectively (Table 2). The highest expected heterozygosity was in the progenitor's group ( $H_{exp} = 0.28$ ), followed by the C0\_DHL group with  $H_{exp} = 0.21$  (Table 2). The lowest expected heterozygosity value was observed in the C17\_DHL group as expected. In comparison, the group C0/C17\_DHL had an expected heterozygosity value of  $H_{exp} = 0.19$ . The MAF and expected heterozygosity values all ranked populations in the same order. Higher values in progenitor and C0\_DHL group were expected, which represents higher allelic variation in relation to the C0/C17\_DHLs and C17\_DHL groups. These values in the C0/C17\_DHL group (F1 cross) were according with the expectation and were predictable values since we knew the parent populations (C0\_DHL and C17\_DHL) values.

## Genetic differentiation analysis

The greatest genetic distance was observed between the progenitor group and the C17\_DHL group (0.18) and the smallest genetic distance was observed between C17\_DHL and C0/C17\_DHL groups (0.11; Table 3). The UPGMA method separated the different groups of DH lines and the progenitor group (Figures 2–4). We observed that the grouping of lines and progenitors followed their origin. That is, lines and progenitors within groups were more related than among groups. In addition, we found high genetic diversity among the DH lines (C0\_DHL, C0/C17\_DHL and C17\_DHL) and progenitors of each group.

The lowest  $F_{ST}$  among the DH lines was observed between the progenitors and the C0\_DHL group (0.15). The highest value was observed between progenitors and C17\_DHL (0.50; Table 3). Manhattan plots showed the genetic differentiation among the different comparisons performed between the progenitors and the different groups of DH lines across the ten chromosomes, with similar patterns across chromosomes (Figures 5, 6).  $F_{ST}$  values of 1 and closer to 1 were observed between the progenitor group and the C17\_DHL group across the genome as expected, demonstrating a considerable differentiation.

In relation to the pairwise relative kinship distribution for the entire set of 487 maize DH lines and 15 progenitors, 53.2% of the kinship coefficient was equal to 0 (Figure 7). Whereas, 46.0% of the entire panel ranged between 0 and 0.4, and only 0.8% were greater than 0.5. Thus, most lines were either not or only distantly related to each other.

Based on PCA, DH lines developed from BSSS can be divided into three subgroups (Figure 8). The first two principal components explained 12.5% of the total SNP variation in the entire panel. Based on discriminant analysis of principal components (DAPC), we observed a clear grouping of the DH lines into the C0\_DHL, C17\_DHL and C0/C17\_DHL. The progenitor lines were grouped within the C0\_DHL cluster, as expected, since the combination of

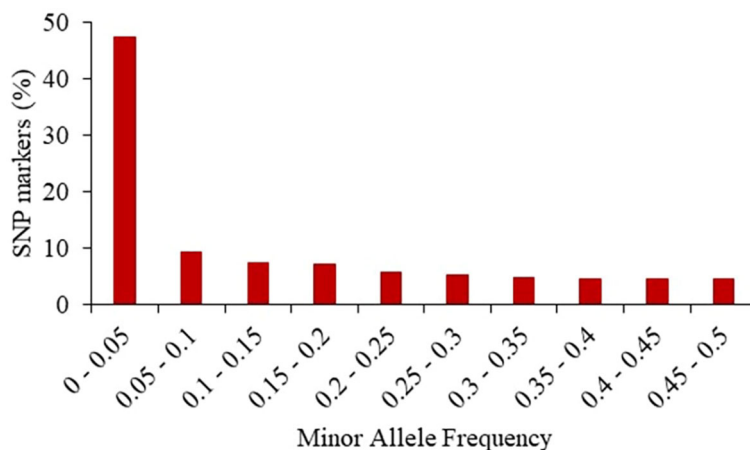


FIGURE 1

Frequency distribution of minor alleles in the entire panel of 487 BSSS DH lines and the 15 progenitors based on 24,885 SNP markers.

these 16 progenitor lines originated this population. The C0/C17\_DHL group were scattered over a wider range, similar to the C0\_DHL group.

The LD decay was variable across the ten chromosomes and different genetic regions within chromosomes in each group (Figure 9). The C17\_DHL group showed the longest LD decay distances ranging from 1,229 to 2,709 kb on chromosomes 3 and 1, respectively. In contrast, the C0/C17\_DHL group displayed the shortest LD decay distances (384 kb on chromosome 5 to 1,024 kb on chromosome 3). For C0\_DHL, the LD decay varied from 486 kb to 1,322 kb for chromosomes 7 and 3, respectively.

For the progenitors' genetic contribution to each set of DH lines, a total of 10,344 polymorphic SNP markers distributed across the whole genome were used to estimate IBD segments among the 15 progenitors and 487 DH lines (Supplemental Table S2, Figure 10). In general, the progenitor A3G-3-3-1-3 had a low

genetic contribution to the different sets of DH lines with 0.91, 0.87 and 0.63% in the C0\_DHL, C0/C17\_DHL and C17\_DHL groups, respectively. In comparison, the progenitor WD 456 had a high genetic contribution to the different sets of DH lines with 5.76, 4.90 and 4.14% in the C0\_DHL, C0/C17\_DHL and C17\_DHL line groups, respectively. The progenitors CI 540 and Os 420 had a similar contribution to the different groups of DH lines. In general, the 15 progenitors evaluated had a higher genetic contribution to C0\_DHLs, ranging from 0.91 to 5.87% for individual progenitors, compared with C0/C17\_DHL (0.87 to 4.90%) and C17 (0.63 to 4.62%). The progenitor with the highest genetic contribution in C0 (Oh 3167B with 5.87%) had a lower contribution in C0/C17\_DHL and C17 with 4.78 and 3.71%, respectively. On average, progenitor lines had 60.1% of the genome classified as identical by descent within C0\_DHLs, 50.0% within the C0/C17\_DHL and 41.6% within C17. The remaining 39.9, 50.0, and 58.4% in C0, C0/C17\_DHL and

TABLE 2 Average Minor Allele Frequency (MAF) and expected heterozygosity ( $H_{exp}$ ) within each group of DH lines and progenitors.

Group	Genotypes	Average MAF	$H_{exp}$
Progenitors	15	0.19 ± 0.001	0.28 ± 0.001
C0_DHL	132	0.16 ± 0.001	0.21 ± 0.001
C0/C17_DHL	170	0.13 ± 0.001	0.19 ± 0.001
C17_DHL	185	0.07 ± 0.001	0.09 ± 0.001

TABLE 3 Pairwise genetic distance and degree of genetic differentiation ( $F_{ST}$ ) between different groups of DH lines and the progenitors of the BSSS maize population.

Group	Progenitors	C0_DHL	C0/C17_DHL	C17_DHL
Progenitors		0.168	0.170	0.175
C0_DHL	0.148		0.141	0.147
C0/C17_DHL	0.220	0.092		0.108
C17_DHL	0.496	0.340	0.131	

Below diagonal: pairwise  $F_{ST}$  estimates between different groups of DH lines and the progenitors, above diagonal: genetic distances.

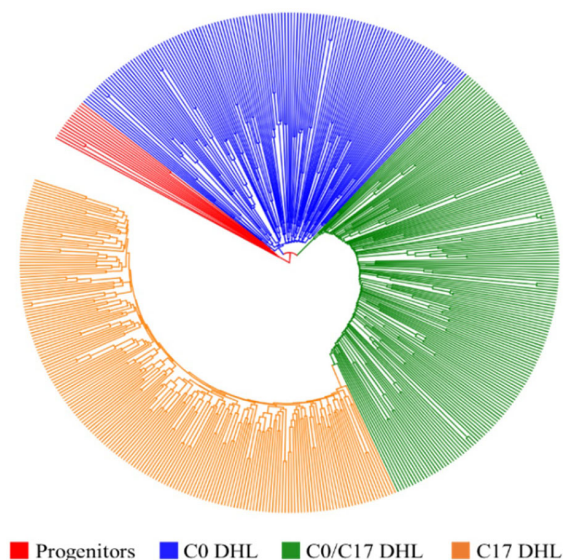


FIGURE 2

Dendrogram constructed from Euclidean genetic distance based on the UPGMA tree method for a panel of 15 progenitors and 495 DH lines derived from BSSS maize population.

C17, respectively are referred to as non-IBD markers. Those SNP markers were not included within the IBD segments between DH line groups and the progenitors.

## Discussion

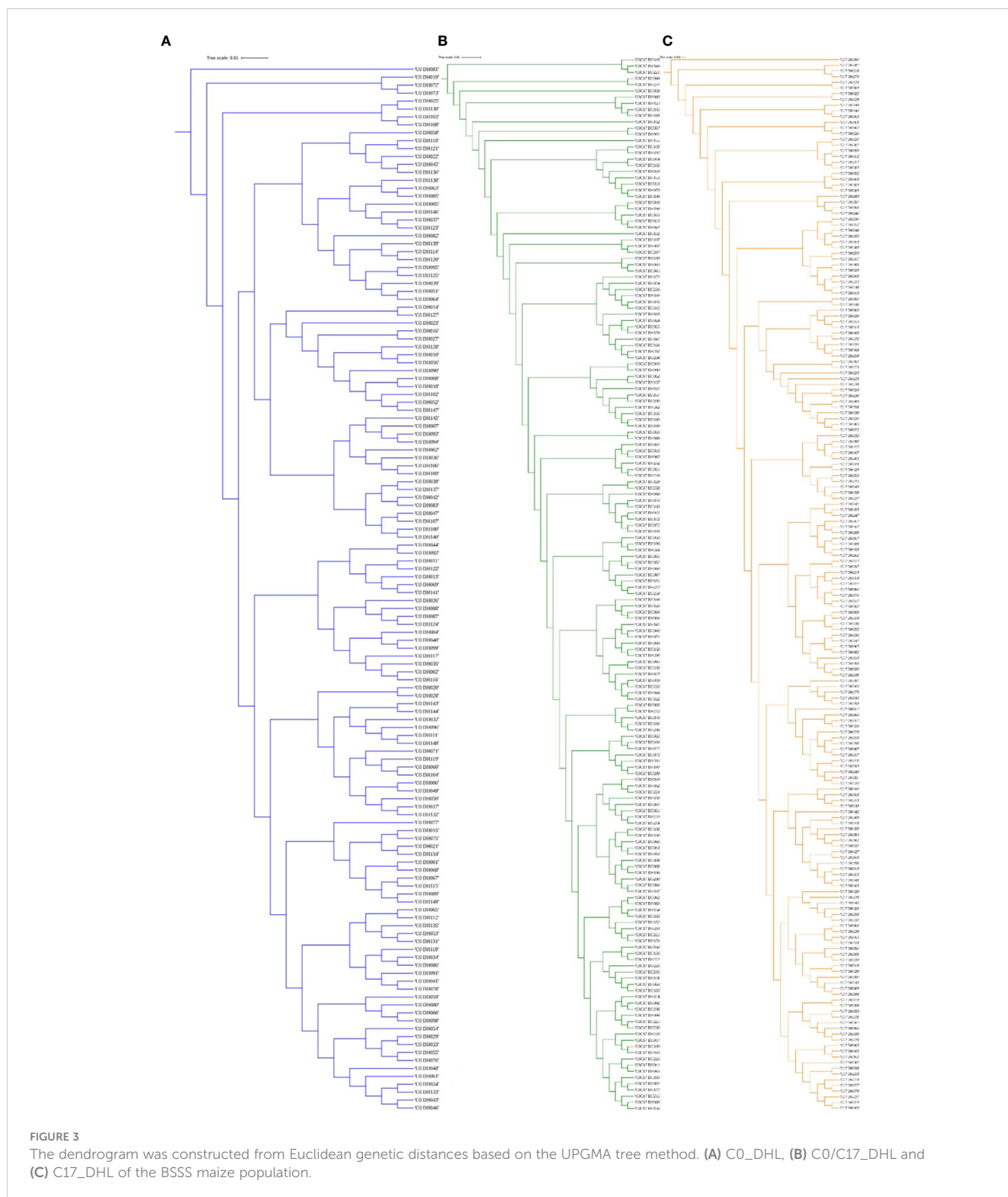
Molecular markers, including SNP markers have been used in many crops including maize for characterizing and quantifying genetic diversity of a given set germplasm for further improvement in a breeding program. The analysis of genetic variation among genetic materials is important to plant breeders, as it contributes to create a core set of germplasm, selecting parental lines, assigning heterotic groups, performs association analysis and prediction potential genetic gains for traits of interest. SNP markers, due to their abundance of availability of sophisticated, rapid, and affordable high-throughput detection systems, have become the principal resource for characterizing and quantifying genetic differences within and among species.

In the present study, the final SNP marker data set included 24,885 SNPs distributed across the ten chromosomes and 502 genotypes corresponding to DH lines derived from different cycles of recurrent selection (132 C0\_DHL, 185 C17\_DHL, and 170 C0/C17\_DHL) plus 15 progenitors of the BSSS maize population. The rationale of using un-imputed data without filtering for MAF was that the BSSS maize population came from 16 founder genotypes. For some SNP markers, an allele was provided by only one founder. The expected frequency would in such a case be ~6.2%. If genetic drift occurred, the actual frequency in C0 can be even lower. C0 seed used in this research came from subsequent cycles of seed multiplication for maintenance, increasing the chance of genetic drift to occur.

## Changes in genetic diversity in different subsets of DH lines

When dividing the number of SNP with heterozygous loci by the total number of SNPs, we observed that our DH lines presented a very low rate of heterozygous loci (less than 3%). Therefore, our DH lines attained an appreciable level of homozygosity, and the DH technology was efficient to fix the loci without requiring further generations of purification. Higher MAF and expected heterozygosity values of the progenitor and C0\_DHL groups (Table 2) were expected due to the large number of alleles that occurred in a few progenitor lines and were lost over recurrent selection cycles (Hagdorn et al., 2003). Additional recombination occurred because of population maintenance. Unfortunately, we do not have adequate records indicating how the seed has been maintained since 1939 when the population was created. Conversely, when comparing the C0\_DHL and C17\_DHL groups, we found a reduction in MAF and expected heterozygosity. The reduction in MAF among these groups was expected due to the recurrent selection process and genetic drift.

The high expected heterozygosity values found in the C0\_DHL group were an indication for the presence of more rare alleles in C0. This could be an important source for new functional alleles of desirable traits, which have been lost during multiple generations of recurrent selection. Potential reduction in genetic diversity in advanced cycles were consistent with previous studies of the BSSS maize population in different cycles of the recurrent selection program (Messmer et al., 1991; Labate et al., 1997; Hagdorn et al., 2003; Hinze et al., 2005), where genome-wide genetic diversity has decreased across cycles of selection. Gerke et al. (2015) found a clear separation, when analyzing the progenitors and individuals from different cycles in the BSSS population. As this was a closed selection process, the substantial increase in genetic distance from

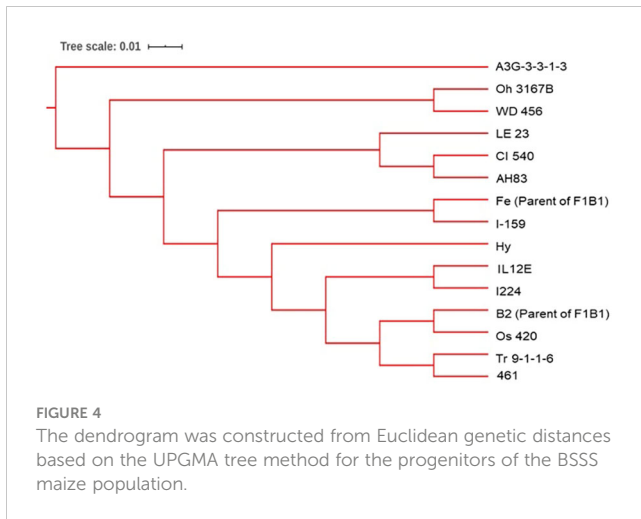


C0\_DHL to C17\_DHL could only arise from genetic differentiation due to selection and genetic drift (Gerke et al., 2015).

Improvement of plant characteristics like flag leaf angle, anthesis-silking interval, plant height, tassel branch number, total number of leaves and grain yield has been observed when advancing cycles in the BSSS recurrent selection program (Brekke et al., 2011;

Edwards, 2011). These changes suggest fixation of favorable alleles during the recurrent selection program. Thus, exploring BSSS cycles using DH technology may reveal useful genetic diversity for plant characteristics left behind in the recurrent selection process and could be an important resource to help drive future genetic gains in maize breeding program.





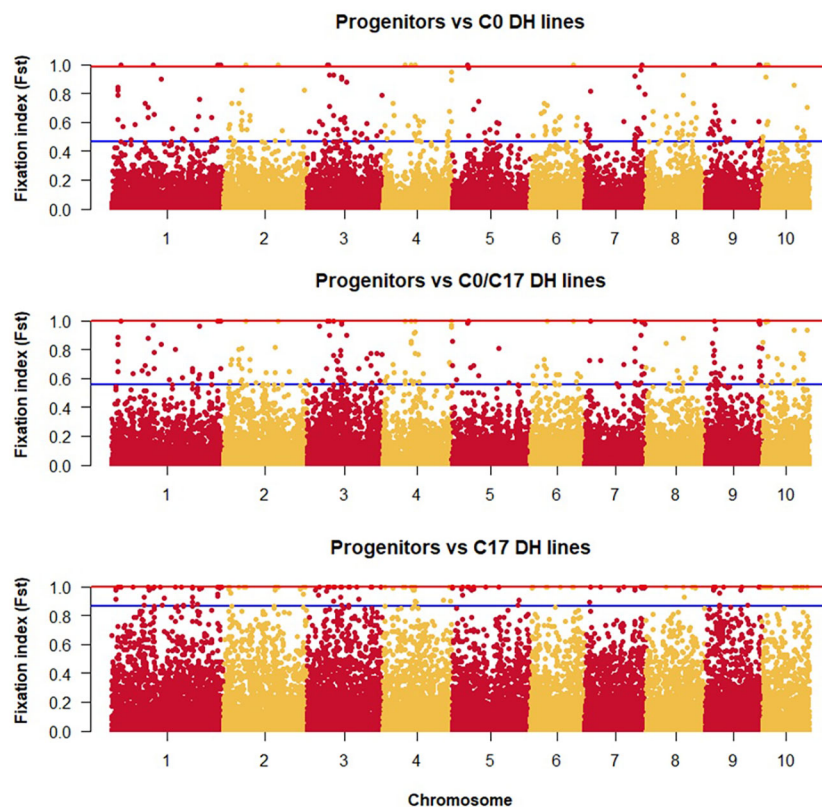
## Genetic relationship and divergence within and among cycles of selection

The Wright's  $F_{ST}$  statistics ( $F_{ST}$ ) used to measure population substructure and the overall genetic divergence among the different groups showed that the degree of differentiation was higher between the progenitor inbred lines and the C17\_DHL group compared to C0\_DHL and C0/C17\_DHL groups as

expected since the two groups share fewer alleles. Lower  $F_{ST}$  values indicate limited differentiation between groups of DH lines. When we compare the  $F_{ST}$  values of C0\_DHL versus C17\_DHL, we observe a clear genetic differentiation among these two groups. These results can be confirmed with the wider genetic distance found among them, reflecting the uniqueness of most lines within these groups. Similar results were found by Gerke et al. (2015) when evaluating the progenitors and samples from different cycles of the BSSS maize population (C0, C4, C8, C12 and C16), indicating a clear differentiation between the founder lines and the population at C16 caused by the loss of different alleles within BSSS maize population. Gerke et al. (2015) conducted extensive simulations using BSSS founder haplotypes to gauge the roles of selection and drift among the cycles of selection and the results showed that most of the reduction in diversity observed among cycles can be attributed to genetic drift alone.

Population structure based on principal component analysis (PCA) is used to reveal genetic divergence among populations (Price et al., 2006). In this study, the results suggest a clear separation into three significant subgroups among all the BSSS DH lines and the progenitors. Also, we observed that the C0/C17\_DHL group was scattered over a wide range, similar to C0\_DHL, indicating a broader genetic divergence among these DH lines than for C17\_DHL.

Kinship coefficients are defined by pedigree and can be estimated based on molecular information. Thus, it is possible to



**FIGURE 5**

Genetic differentiation compares the progenitor group and the different groups of DH lines across chromosomes (x-axis) with the  $F_{ST}$  value (y-axis). Dots between the red and the blue lines represent the highest 1% of the  $F_{ST}$  values.

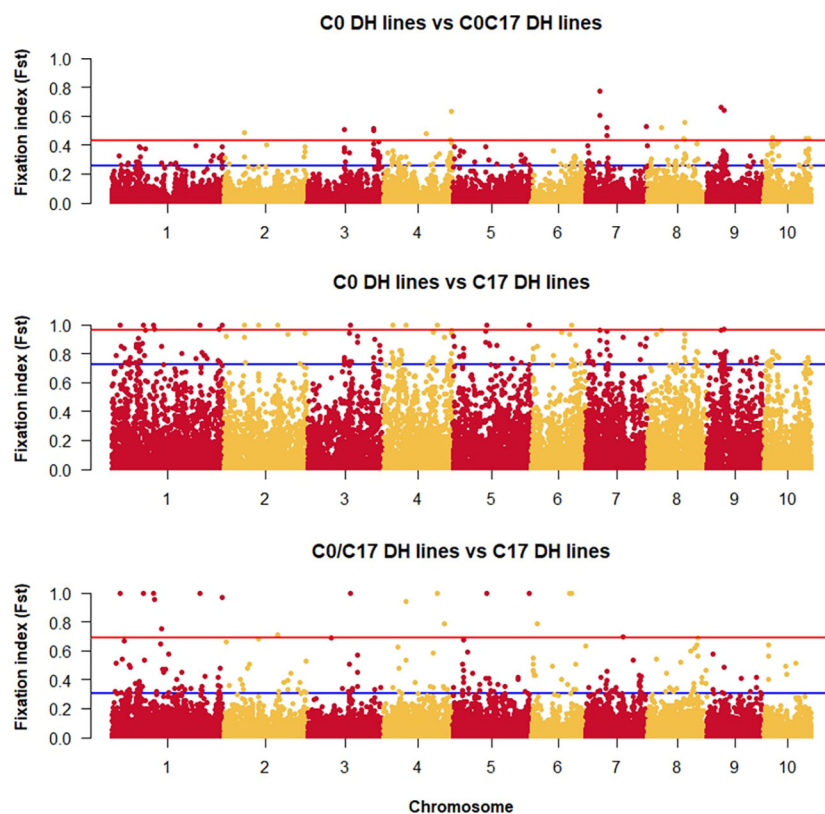


FIGURE 6

Genetic differentiation compares the different groups of DH lines across chromosomes (x-axis) with the  $F_{ST}$  value (y-axis). Dots between the red and the blue lines represent the highest 1% of the  $F_{ST}$  values.

find hidden relationships. We found that most of the DH lines in the entire panel were distantly related to each other. Therefore, this shows us a low relationship between DH lines of the C17\_DHL and C0\_DHL. The estimation of the degree of the relationship depends on the description of an ancestral population, which by definition, is assumed to be the base from where the past ancestry is no longer accounted (Wright, 1922). Thus, the lower the number of generations separating the ancestral with the current population, the higher the kinship coefficient among individuals because of a reduced number of possible recombination events (Wang, 2014). Low or negative relative

kinship coefficients among pairs of DH lines were found in the C0/C17\_DHL group reflecting the uniqueness of most lines.

## Linkage disequilibrium in BSSS DH lines

Linkage disequilibrium (LD) refers to the non-random co-segregation of alleles at two loci. Recombination events shuffle genetic material during meiosis among homologous chromosomes and cause LD to decay with increasing distance. Multiple factors are

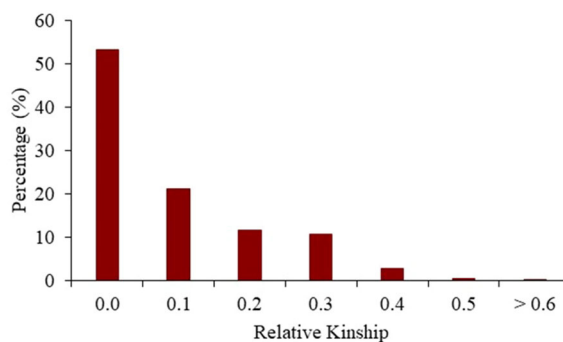


FIGURE 7

Distribution of pairwise relative kinship for 487 maize DH lines and 15 progenitor lines of the BSSS maize population calculated using 24,885 SNP markers.

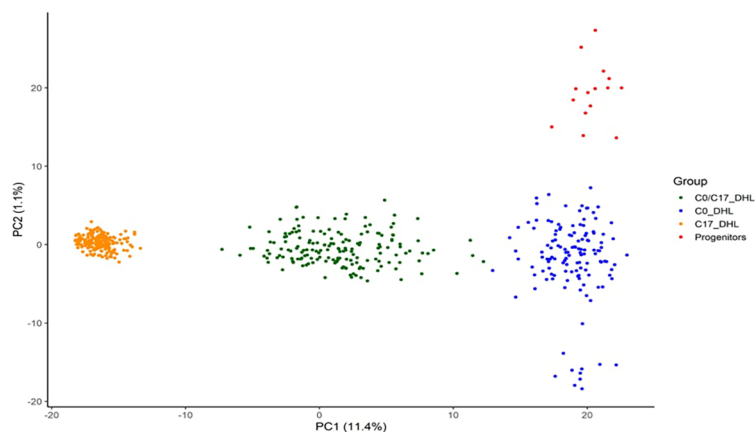


FIGURE 8

Scatter plot of the discriminant analysis of principal components based on 487 DH lines and 15 progenitors of the BSSS maize population. The dots represent each of the DH lines within their respective population. The axes represent the first two discriminant functions, respectively.

affecting LD in crops. Generally, LD decays faster in cross-pollinated crops, diverse populations, but also, different genes and genomic regions in the same crop can exhibit different rates of LD decay. It is expected in maize, for genome regions to decay at distances around 1 kb for exotic landraces, as described by (Romay et al., 2013). In the Ames panel subset corresponding to 384 lines (Pace et al., 2015) the LD decay rate was similar across chromosomes with an average distance of 10 kb throughout the genome. In this study, the LD decay distance among lines of the C17\_DHL group was larger compared among lines of the C0\_DHL and C0/C17\_DHL groups. The longer LD decay distances in C17\_DHL was consistent with the lower average MAF and expected heterozygosity results, as the rate of effective recombination declines over selection cycles due to the occurrence of bottlenecks or due to fixation for favorable alleles over time. The 17 cycles of recurrent selection did lead to a lower genetic diversity in the C17\_DHL group, and LD decays more rapidly in pools of lines with higher genetic diversity (Romay et al., 2013; Wu et al., 2016). The distance over which LD persists determines the number and density of markers, and

experimental design needed to perform an association analysis (Flint-Garcia et al., 2003). This was actually applied when generating the IBM Syn10 ultra-high-density map to precisely map a quantitative trait locus (Liu et al., 2015) at a higher genetic resolution than the IBM Syn4 map (Hu et al., 2016). In contrast, additional cycles of recurrent selection in the BSSS maize population increased homozygosity and LD decay distances due to selection and drift. Consistent with Gerke et al. (2015) genome-wide expected heterozygosity decreases steadily across cycles of selection. The loss of heterozygosity indicates the loss of different alleles within BSSS maize population.

### Progenitor genetic contributions to different subsets of DH lines

On average, the mean genetic contribution of the BSSS progenitor lines estimated using high-resolution detection of IBD segments changed in the different groups of DH lines. The

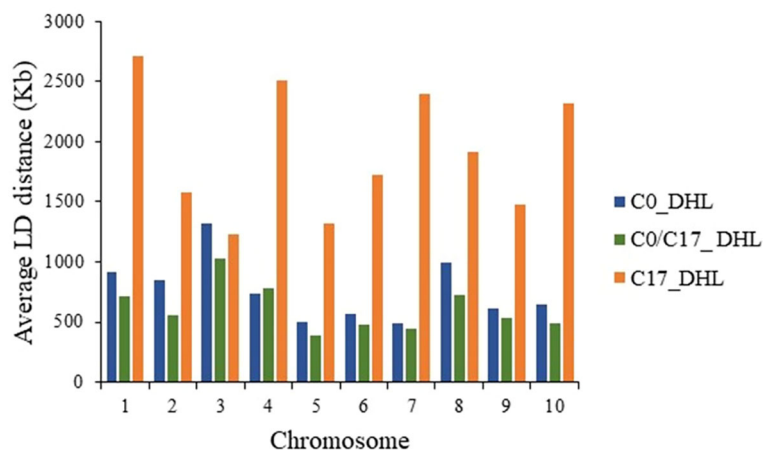


FIGURE 9

Linkage Disequilibrium (LD) decay distance per chromosome in the different groups of DH lines.

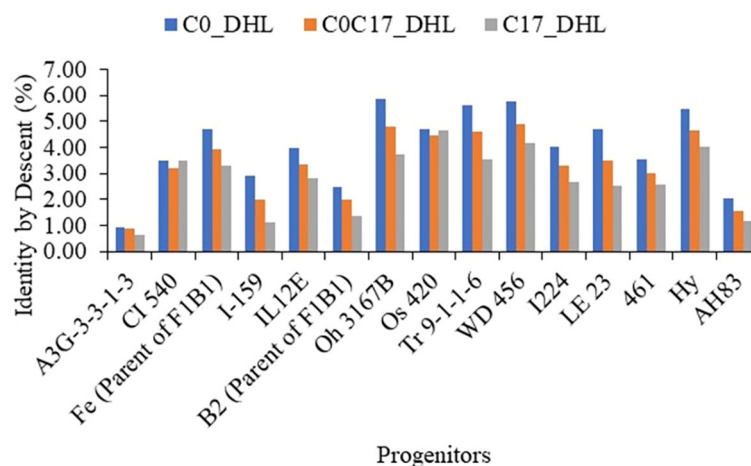


FIGURE 10

The genome's proportion classified as IBD among the BSSS progenitors inbred lines for each group of DH lines evaluated (C0\_DHL, C0/C17\_DHL and C17\_DHL) identified with marker-based dissimilarity values.

progenitors had the highest genetic contribution in the C0\_DHL group, due to their use in obtaining the population, and the lowest contribution in the C17\_DHL group in relation to the other groups. This suggests that relationships caused by more recent ancestry had the most significant contribution in the IBD segments among individuals. Additionally, 17 cycles of recurrent selection have changed the allele frequencies in the C17\_DHL group, because only individuals with superior performance for the selected traits contributed alleles to the next generation.

In the identification of regions in the genome inherited from the progenitors, we found prevalence of small to medium sized segments, where 50.4% of the segments were between 2.4 to 4.1 Mb. And, 28.2% of the segments ranged from 4.1 to 8.1 Mb inherited from the progenitor inbred lines. The number of segments decreased with increased length segments. IBD segment sizes from the progenitors changed across groups of DH lines. We found that some progenitors showed longer IBD segments in the C0\_DHL group and others longer in the C17\_DHL. Large, preserved regions in the genome could be associated with selection processes, resulting in long DNA segments inherited as a block from the parents. Therefore, under positive selection favoring a phenotype, a slight increase in LD surrounding the favored alleles will be produced. In these cases, the length of the IBD segment surrounding the alleles subject to selection will increase, experiencing less recombination at the population level (Albrechtsen et al., 2010). Albrechtsen et al. (2010) states that a reduced recombination rate in the genome, leading to significant LD, could be explained as a function of the effective population size. These could partially be explained by an increase in random genetic drift because of the population size, which will increase the length of DNA that will be shared among individuals in the population similar to what could happen in the C17\_DHL with the 17 cycles of the recurrent selection process. The detection of long IBD segments in populations could be used as evidence for strong and recent selection processes because these segments have not suffered from recombination. However, many recombination's could have occurred because of subsequent cycles of seed multiplication and

population maintenance. Unfortunately, we do not have adequate records indicating how the seed has been maintained since 1939 when the population was created. In cases where alleles within long IBD segments are in linkage disequilibrium, specifically in repulsion phase, unfavorable alleles will persist in the population, inducing the hitch-hiking effect and reducing the genetic diversity (Hospital and Chevalet, 1993). This hitch-hiking will increase genetic drift and significantly decrease the effective population size (Smith and Haigh, 1974). More studies should necessarily be done to confirm the possibility of the hitch-hiking effect having an effect in this population. Conversely, the restricted population size of both from founding (16 lines) and from continued population maintenance, may have provided the maintenance of long IBD segments. IBD segments shared between different groups of DH lines and the 15 progenitor lines will allow the estimation of genetic diversity and progenitor genetic contributions to new released lines.

In this study, we measured the genetic diversity among different sets of DH lines derived from the BSSS maize population and our results confirmed the separation from BSSS(R)C0 to BSSS(R)17 through the recurrent selection process. The selection process and the effective population size applied to the BSSS maize population have reduced the genetic variability. Consistent with previous studies (Messmer et al., 1991; Labate et al., 1997; Hagdorn et al., 2003; Hinze et al., 2005). Although genetic drift can explain most of the genetic structure genome-wide, phenotypic data provide evidence that selection has altered favorable allele frequencies in the BSSS maize population. We also found that the greatest genetic distance and  $F_{ST}$  observed between the progenitors group and the C17\_DHL group demonstrated a clear genetic differentiation among groups caused by the loss of different alleles during the recurrent selection program in the BSSS maize population, reflecting the uniqueness of most lines within these groups of DH lines. Thus, these DH lines can be evaluated in replicated trials, and genomic selection can be applied for the estimation of the breeding value for each DH line. Additionally, DH lines derived from the BSSS maize population could be ideal for association mapping due to

the low population structure. Thus, we could identify genes or regions in the genome associated with a particular trait. Using genome-based data and DH technology was a powerful tool for access to the genetic diversity available in C0\_DHL or C0/C17\_DHL groups, which would be beneficial to incorporate in BSSS(R)17 to broaden its genetic variation while minimizing yield or other penalties. Thus, the results of this research will also help maize breeders to explore useful genetic variation for further improvement in a breeding program.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Iowa State University DataShare, accession 22893878. DOI: <https://doi.org/10.25380/iastate.22893878.v1>.

## Author contributions

TL, AL, JE, UF and SH conceived and designed the experiments. AL analyzed the genotypic data and conducted the molecular characterization. AL and FA conducted the IBD analysis. AL, AU and TL wrote the manuscript, with contributions from all the other authors. All authors contributed to the article and submitted and approved the submitted section.

## Funding

Funding for this work was provided by USDA's National Institute of Food and Agriculture (NIFA) Project, No. IOW04314, IOW01018, and IOW05510; and NIFA award 2018-51181-28419. Funding for this work was also provided by the R.F. Baker Center

## References

- Albrechtsen, A., Moltke, I., and Nielsen, R. (2010). Natural selection and the distribution of identity-by-descent in the human genome. *Genetics* 186, 295–308. doi: 10.1534/genetics.110.113977
- Almeida, V. C., Trentin, H. U., Frei, U. K., and Lübberstedt, T. (2020). Genomic prediction of maternal haploid induction rate in maize. *Plant Genome* 13, e20014. doi: 10.1002/tpg2.20014
- Andorf, C., Beavis, W. D., Hufford, M., Smith, S., Suza, W. P., Wang, K., et al. (2019). Technological advances in maize breeding: past, present and future. *Theor. Appl. Genet.* 132, 817–849. doi: 10.1007/s00122-019-03306-3
- Beckett, T. J., Morales, A. J., Koehler, K. L., and Rocheford, T. R. (2017). Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. *PLoS One* 12, 1–23. doi: 10.1371/journal.pone.0189277
- Böhm, J., Schipprack, W., Utz, H. F., and Melchinger, A. E. (2017). Tapping the genetic diversity of landraces in allogamous crops with doubled haploid lines: a case study from European flint maize. *Theor. Appl. Genet.* 130, 861–873. doi: 10.1007/s00122-017-2856-x
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *J. Bioinform.* 23 (19), 2633–2635. doi: 10.1093/bioinformatics/btm308
- Brekke, B., Edwards, J., and Knapp, A. (2011). Selection and adaptation to high plant density in the Iowa Stiff Stalk Synthetic maize (*Zea mays* L.) population. *Crop Sci.* 51, 1965–1972. doi: 10.2135/cropsci2010.09.0563
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015
- Coffman, S. M., Hufford, M. B., Andorf, C. M., and Lübberstedt, T. (2019). Haplotype structure in commercial maize breeding programs in relation to key founder lines. *Theor. Appl. Genet.* 133, 547–561. doi: 10.1007/s00122-019-03486-y
- Edwards, J. (2011). Changes in plant morphology in response to recurrent selection in the Iowa Stiff Stalk Synthetic maize population. *Crop Sci.* 51, 2352–2361. doi: 10.2135/cropsci2010.09.0564
- Endelman, J. B., and Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *Genes|Genomes|Genetics* 2, 1405–1413. doi: 10.1534/g3.112.004259
- Ertiro, B. T., Semagn, K., Das, B., Olsen, M., Labuschagne, M., Worku, M., et al. (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 Genom.* 18, 1–11. doi: 10.1186/s12864-017-4173-9
- Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. B.IV (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

for Plant Breeding, Plant Sciences Institute, and K.J. Frey Chair in Agronomy at Iowa State University

## Acknowledgments

Alejandro Ledesma Miramontes acknowledges the National Council for Science and Technology (CONACYT), International Maize and Wheat Improvement Center (CIMMYT) and the National Institute for Agricultural, Livestock, and Forestry Research (INIFAP) for the scholarship 2016 for Ph.D. studies.

## Conflict of interest

The authors declare that they have no conflicts of interest. Availability of data and material data transparency

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1226072/full#supplementary-material>

- Gerke, J. P., Edwards, J. W., Guill, K. E., Ross-Ibarra, J., and McMullen, M. D. (2015). The genomic impacts of drift and selection for hybrid performance in maize. *Genetics* 201, 1201–1211. doi: 10.1534/genetics.115.182410
- Goudet, J. (2005). HIERFSTAT, a Package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 2, 184–186. doi: 10.1111/j.1471-8278
- Gower, J. C., and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.* 3, 5–48. doi: 10.1007/BF01896809
- Guo, R., Chen, J., Petrol, C. D., Pacheco, A., Zhang, X., San Vicente, F., et al. (2021). The genetic structure of CIMMYT and US inbreds and its implications for tropical maize breeding. *Crop Sci.* 61, 1666–1681. doi: 10.1002/csc2.20394
- Hagdorn, S., Lamkey, K. R., Frisch, M., Guimaraes, P. E., and Melchinger, A. E. (2003). Molecular genetic diversity among progenitors and derived elite lines of BSSS and BSCB1 maize populations. *Crop Sci.* 43, 474–482. doi: 10.2135/cropsci2003.4740
- Han, L., and Abney, M. (2011). Identity by descent estimation with dense genome-wide genotype data. *Genet. Epidemiol.* 2335, 557–567. doi: 10.1002/gepi.20606
- Han, L., and Abney, M. (2013). Using identity by descent estimation with dense genotype data to detect positive selection. *Eur. J. Hum. Genet.* 21, 205–211. doi: 10.1038/ejhg.2012.148
- Hill, W. G., and Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33, 54–78. doi: 10.1016/0040-5809(88)90004-4
- Hinze, L. L., Kresovich, S., Nason, J. D., and Lamkey, K. R. (2005). Population genetic diversity in a maize reciprocal recurrent selection program. *Crop Sci.* 45, 2435–2442. doi: 10.2135/cropsci2004.0662
- Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat. Rev. Genet.* 10, 639–650. doi: 10.1038/nrg2611
- Hospital, F., and Chevalat, C. (1993). Effects of population size and linkage on optimal selection intensity. *Theor. Appl. Genet.* 86, 775–780. doi: 10.1007/BF00222669
- Hu, S., Lübberstedt, T., Zhao, G., and Lee, M. (2016). QTL mapping of low-temperature germination ability in the maize IBM Syn4 RIL population. *PLoS One* 11, 1–11. doi: 10.1371/journal.pone.0152795
- Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* 546, 524–527. doi: 10.1038/nature22971
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2013, 1–14. doi: 10.7717/peerj.281
- Keeratinijakal, V., and Lamkey, K. R. (1993). Responses to reciprocal recurrent selection in BSSS and BSCB1 maize populations. *Crop Sci.* 33, 73–77. doi: 10.2135/cropsci1993.0011183X003300010012xh
- Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249. doi: 10.1534/genetics.111.130922
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). “Diversity arrays technology: a generic genome profiling technology on open platforms.” In: Pompanon, F., Bonin, A. (eds) Data Production and Analysis in Population Genomics. *Methods Mol. Biol.* (Hertfordshire, UK: Humana Press). 67–89. doi: 10.1007/978-1-61779-870-2\_5
- Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., Mckeigue, P. M., and Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5, 1–7. doi: 10.1371/journal.pone.0013996
- Labate, J. A., Lamkey, R., Lee, M., and Woodman, W. L. (1997). Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. *Crop Sci.* 37, 416–423. doi: 10.2135/cropsci1997.0011183X003700020018x
- Labate, J. A., Lamkey, K. R., Lee, M., and Woodman, W. L. (1999). Temporal changes in allele frequencies in two reciprocally selected maize populations. *Theor. Appl. Genet.* 99, 1166–1178. doi: 10.1007/s001220051321
- Lamkey, K. (1992). Fifty years of recurrent selection in the Iowa stiff stalk synthetic maize population. *Maydica* 37, 19–28.
- Ledesma, A. (2020). Molecular and phenotypic characterization of doubled haploid lines derived from different cycles of the Iowa Stiff Stalk Synthetic maize population [Dissertation/PhD thesis]. [Ames, IA]: Iowa State University.
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, 256–259. doi: 10.1093/nar/gkz239
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *J. Bioinform.* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, H., Niu, Y., Gonzalez-Portilla, P. J., Zhou, H., Wang, L., Zuo, T., et al. (2015). An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. *BMC Genom.* 16, 1–16. doi: 10.1186/s12864-015-2242-5
- Lu, Y., Yan, J., Guimarães, C. T., Taba, S., Hao, Z., Gao, S., et al. (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor. Appl. Genet.* 120, 93–115. doi: 10.1007/s00122-009-1162-7
- Maldonado, C., Mora, F., Scapim, C. A., and Coan, M. (2019). Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: hapLA4. *PLoS One* 14, e0212925. doi: 10.1371/journal.pone.0212925
- McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007
- Messmer, M. M., Melchinger, A. E., Lee, M., Woodman, W. L., Lee, E. A., and Lamkey, K. R. (1991). Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. *Theor. Appl. Genet.* 83, 97–107. doi: 10.1007/BF00229231
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U.S.A.* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590. doi: 10.1093/genetics/89.3.583
- Nei, M., and Roychoudhury, A. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* 76, 379–390. doi: 10.1093/genetics/76.2.379
- Nelson, P. T., Coles, N. D., Holland, J. B., Bubeck, D. M., Smith, S., and Goodman, M. M. (2008). Molecular characterization of maize inbreds with expired U.S. plant variety protection. *Crop Sci.* 48, 1673–1685. doi: 10.2135/cropsci2008.02.0092
- Ogut, F., Bian, Y., Bradbury, P. J., and Holland, J. B. (2015). Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity* 114, 552–563. doi: 10.1038/hdy.2014.123
- Ouborg, N. J., Piquot, Y., and Van Groenendael, J. M. (1999). Population genetics, molecular markers and the study of dispersal in plants. *J. Ecol.* 87, 551–568. doi: 10.1046/j.1365-2745.1999.00389.x
- Pace, J., Gardner, C., Romay, C., Ganapathysubramanian, B., and Lübberstedt, T. (2015). Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). *BMC Genom.* 16, 1–12. doi: 10.1186/s12864-015-1226-9
- Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., et al. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet.* 3, 1672–1686. doi: 10.1371/journal.pgen.0030160
- Penny, L. T., and Eberhart, S. A. (1971). Twenty years of reciprocal recurrent selection with two synthetic varieties of maize (*Zea mays* L.). *Crop Sci.* 11, 900–903. doi: 10.2135/cropsci1971.0011183X001100060041x
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- R Core Team (2021). “R: A language and environment for statistical computing.” (R Foundation for Statistical Computing).
- Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14, 1–18. doi: 10.1186/gb-2013-14-6-r55
- Sansaloni, C., Petrol, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., et al. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* 5, 1–2. doi: 10.1186/1753-6561-5-s7-p54
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136
- Smith, M. J., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672308009579
- Sprague, G. F. (1946). Early testing of inbred lines of corn. *J. Am. Soc. Agron.* 38, 108–117. doi: 10.2134/agronj1946.00021962003800020002x
- Sprague, G. F., and Jenkins, M. T. (1943). A comparison of synthetic varieties, multiple crosses, and double crosses in corn. *J. Agron.* 35, 137–147. doi: 10.2134/agronj1943.00021962003500020007x
- Sul, J. H., Martin, L. S., and Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genet.* 14, 1–22. doi: 10.1371/journal.pgen.1007309
- Vanous, K., Vanous, A., Frei, U. K., and Lübberstedt, T. (2017). Generation of maize (*Zea mays*) doubled haploids via traditional methods. *Curr. Protoc.* 2, 147–157. doi: 10.1002/cppb.20050
- Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *J. Evol. Biol.* 27, 518–530. doi: 10.1111/jeb.12315
- Warburton, M. (2005). Laboratory Protocols: CIMMYT applied molecular genetics laboratory, 3rd ed. (Mexico, D.F: CIMMYT).
- Wegary, D., Teklewold, A., Prasanna, B. M., Ertiro, B. T., Alachiotis, N., Negera, D., et al. (2019). Molecular diversity and selective sweeps in maize inbred lines adapted to African highlands. *Sci. Rep.* 9, 1–15. doi: 10.1038/s41598-019-49861-z
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641
- Wijayasekara, D., and Ali, A. (2021). Evolutionary study of maize dwarf mosaic virus using nearly complete genome sequences acquired by next-generation sequencing. *Sci. Rep.* 11, 1–14. doi: 10.1038/s41598-021-98299-9

- Won, S., Park, J. E., Son, J. H., Lee, S. H., Park, B. H., Park, M., et al. (2020). Genomic prediction accuracy using haplotypes defined by size and hierarchical clustering based on linkage disequilibrium. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00134
- Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.* 56, 330–338.
- Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.* 15, 323–354.
- Wu, Y., San Vicente, F., Huang, K., Dhliwayo, T., Costich, D. E., Semagn, K., et al. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor. Appl. Genet.* 129, 753–765. doi: 10.1007/s00122-016-2664-8
- Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B. M., Li, L., et al. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526. doi: 10.1007/s11032-010-9500-7
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhu, C., and Yu, J. (2009). Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182, 875–888. doi: 10.1534/genetics.108.098863