Original Paper Open Access

# Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study

**Akula Dinesh[1], Ayyanagouda Patil[1]*, PH Zaidi[2], PH Kuchanur[1], MT Vinayan[2], K Seetharam[2]**

[1]Department of Genetics and plant breeding, University of agricultural sciences, Raichur, Karnataka, India
[2]Department of molecular biology and Agril. Biotechnology, University of agricultural sciences, Raichur, Karnataka, India
[3]International Maize and Wheat Improvement Center (CIMMYT)- Asia c/o ICRISAT, Patancheru, 502324, India
*Corresponding author: E-mail: ampatil123@gmail.com

## Abstract

Rising temperatures has led to reduced maize yields in tropical and sub tropical countries. This provides the necessity for identifying the diverse inbred lines that can produce high yielding hybrids under high temperature regimes. With this view, the present study was conducted to analyse the extent of genetic diversity and population structure among 64 CIMMYT maize inbred lines using SNP markers derived from GBS (Genotyping by sequencing) along with characterization of haplotype blocks and linkage disequilibrium. The average polymorphic information content (0.37) and gene diversity was very high (0.5) with mean kinship coefficients of 0.28 and genetic distance more than 0.4 between pair of two inbred lines. Clustering analysis based on ward's method and euclidian distance showed presence of three sub groups. The population structure analysis using principle components showed three sub population. The average physical distance between pairs of markers was 27.7 kb with linkage disequilibrium (LD) estimation ($r^{\backslash}$) of 0.36 across all chromosomes., with rapid LD decay of 6.34 kb at $r^2$ = 0.2. Haplotype analysis with 75,664 SNPs under confidence interval model revealed 616 halpotype blocks across all chromosomes with highest number of blocks on chromosome 5. The results clearly indicate the uniqueness of the majority of the inbred lines, which can contribute to new alleles in breeding programme for heat tolerance.

**Keywords**: linkage disequilibrium, haplotype blocks, genetic diversity, population structure, *Zea mays*

## Introduction

Maize is one of the important food crops grown across the world for its yield potential, nutritive value and economical importance. Globally maize production was 967 mt (metric ton) in 2013 - 14, cultivated in an area of 177 mn hectare with productivity of 5.5 mt hectare$^{-1}$ (Anon, 2014). In India, maize has a pride of place in food grain scenario of the country, which contributes more than 8 % to the national food basket with productivity of 2.5 mt hectare$^{-1}$ which is very less compared to global productivity. Despite of lower productivity in India, there is an increased demand for maize production for its multiple uses like poultry feed, production of starch for textile, pharmaceutical, cosmetic industries, high quality corn oil, protein, alcoholic beverages etc. The maize production is mainly hampered by biotic and abiotic stress in addition to failure in exploitation of heterosis. Gain from heterosis can be achieved only when highly diverse inbred lines selected for successive breeding program.

Maize is a model genetic organism with immense genetic diversity. Although it was first domesticated in Mexico, maize landraces are widely found across the continents (Prasanna, 2012) due to presence of high diversity. Information about the genetic diversity and population structure in breeding material is of fundamental importance for crop improvement (Van Inghelandt, 2010) and exploitation of diversity helps breeding new genotypes or selection for desirable genotypes for specific environment/stress condition. Genetic diversity analysis among maize lines based on morphological and molecular data were reported by previous studies (Hartings et al, 2008; Lu et al, 2009; Thirunavukkarasu et al, 2013). The presence of phenotypic variation in germplasm is the primary requisite for identifying genotypes with specific characters. However, there is more chance of getting error in diversity analysis because of environmental influence (Smith and Smith, 1992).

On contrary DNA based markers are highly heritable and not influenced by environmnet, ideal for diversity and population structure analysis (Beyene et al, 2006; Cholastova and Knotova, 2012). The genetic diversity can be assessed based on genetic distance and kinship coefficient values along with Linkage disequlibrium (LD). Marker based characterization of tropical and subtropical maize representing CIMMYT germplasm were carried out using 79 SSR in 155 inbred lines (Xia et al, 2004), 32 RFLP markers in 219 lines (Warburton et al, 2005) and 25 SSR markers in maize landraces, OPVs, and inbred lines (Warburton et al, 2008) for diverse purposes. SSR markers were

 RECEIVED 5/31/2016

proved good source for estimation of genetic diversity and population structure, but they were expensive, less automated and does not represent genome wide coverage in comparison to single nucleotide markers (SNPs). Moreover, very large number of SNP markers were available in maize and many of which represent functional markers. For this reason, SNP markers were highly used in maize improvement, including genetic diversity, population structure and haplotype block estimation. Molecular characterization of CIMMYT (394), Chinese (282), and Brazilian lines (94) using 1,034 SNP markers (Lu et al, 2009) and 450 maize inbred lines using 1,065 SNP markers (Semagn et al, 2012) was carried out. The study of Linkage disequilibrium (LD) and haplotype blocks present in germplasm will assist understanding the genetic nature of inheritance of genes. LD refers to non random association between two markers (Gupta et al, 2005). The germplasm with high LD shows less diverse compared to low LD. The rate at which, the LD breakdown known has LD decay. The germplam with rapid LD decay shows high recombination and are more diverse. The average LD decay in tropical maize germplasm (5 – 10 kb) was two to ten times faster than that in the temperate germplasm, which represents the diverse nature of tropical maize (Yan et al, 2009). The haplotype block pattern which represents the magnitude of recombination across the genome was studied by Thirunavukkarasu et al (2013) in tropical and sub tropical maize lines using 29,619 SNPs. The haplotype blocks for particular trait can be used in marker assisted selection. Earlier report on estimation of genetic diversity, population structure and haplotype block patterns among CIMMYT lines used low SNP markers. In this regard, the present study utilizes large number of GBS derived SNP markers for molecular characterization of CIMMYT maize inbred lines for documentation of LD decay, haplotype pattern, genetic diversity and population structure at genome level.

Moreover, the study of genetic diversity and population structure along with linkage disequilibrium and haplotype block patterns help in identification of diverse parental genotypes, distinct allele frequency which can be useful in breeding hybrids. LD estimates and haplotype blocks further help in understanding the With this background knowledge, the present investigation was carried out to study the genomic characteristics in defining genetic diversity along with population structure in inbreds selected for heat tolerance studies.

## Materials and Methods

### *Plant material*

A panel of sixty four maize inbred lines representing tropical maize germplasm developed by the International Maize and Wheat Improvement Center (CIMMYT) were genotyped at Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA using GBS v2.7 version (Elshire et al, 2011). A total of 9,55,690 SNP markers were generated. The physical coordinates of GBS data were derived from AGPv2.

### *SNP marker characteristics, Genetic diversity and Population structure*

The genetic diversity among the genotypes was analysed using 1,629 unlinked high quality SNPs with zero missing data. The Polymorphism information content (PIC), major allelic frequency, heterozygosity and gene diversity (He) which assess the genetic diversity at each locus was analysed using Powermarker V3.25 (Liu and Muse, 2005). For calculating Polymorphic information content (PIC = 1-∑Pij 2), the relative frequency of the jth allele for the ith locus was summed across all the alleles for the locus over all lines. The kinship or relatedness between the individuals was assessed by kinship matrix and genetic distance using modified Euclidean distance implemented in TASSEL 4.1v. Population structure among 64 inbred lines was estimated with 1629 unlinked SNP markers using SVS 7.7v and GenABEL package in R programme. The linkage disequilibrium was estimated as squared allele frequency correlations ($r^2$) between pairs of SNP markers according to Weir (2008). The Pairwise LD patterns between SNPs explained by $r^2$ was investigated using TASSEL 4.1v (Bradbury et al, 2007). The Pairs of loci were considered to be in significant LD if P was < 0.01. To determine the extent of LD decay across genome and among chromosome, a high quality 75,664 SNPs with CR ≥ 0.94 and MAF ≥ 0.1 having pairwise $r^2$ values and physical distances among these SNPs (Remington et al, 2001) was extracted from SVS v 7.7.8 and viuaslised using R. The haplotypes blocks were analyzed using Haploview 4.2v (Barrett et al, 2005) using three models confidence interval (CI), four gamete rule (FGR), and solid spine of LD (SS).

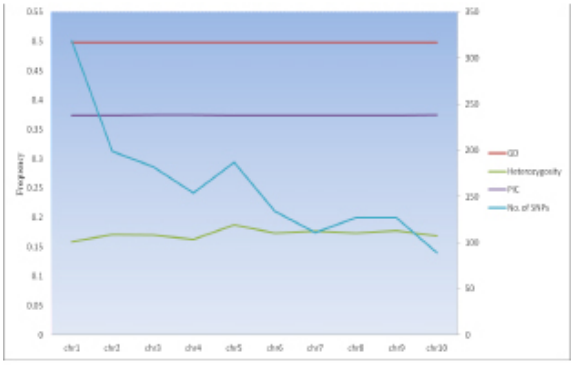

Figure 1 - SNP markers representing gene diversity (GD), polymorphic information content (PIC), No. of SNP per chromosome and heterozygosity values.Y1-axis: frequency, X-axis: chromosome number (increasing from left to right) and Y2-axis: number of SNPs.

## Results

### *Genetic diversity among inbred lines using SNP markers*

Each SNP marker exhibited 2 alleles per locus with mean heterozygosity of 0.17. The mean genetic diversity calculated at each locus for allelic Polymorphism Information Content (PIC) value was 0.374 with minimum PIC value of 0.372 and maximum of 0.375 across the inbred lines. The inbred lines exhibited gene diversity of 0.5 with minimum 0.49 and maximum 0.51 (Figure 1). The genetic diversity assessment based ward method and Euclidian distance using molecular marker showed presence of three clusters. The sixty four genotype were grouped into three distinct sub population set G1, G2, and G3 having 11, 17 and 36 genotypes respectively (Supplementary Figure 1). Out of 4,096 pairwise comparison, relative kinship coefficients between pairs of samples varied from 0 to 1.99 with an average of 0.28 (Figure 2A). The estimates of kinship/relatedness between individuals showed that 80 % of pairwise comparisons have value less than 0.3 and 96.5 % of pairwise kinship estimate have value less than 0.5 advocating the existence of considerable diversity in the panel. The kinship value of close to zero indicates no relation between the individuals which indicates high diversity. The genetic distance measured between pair of individuals based on modified Euclidean method showed more than 96% of pair wise comparisons have genetic distance value more than 0.4 indicating sufficient diversity in selected panel of inbred lines (Figure 2B).

### *Linkage disequilibrium (LD), Haplotype and Linkage disequilibrium (LD) decay characterization in tropical maize*

For LD estimation a total of 75,664 SNPs were used for, the average physical distance between pairs of markers showed 27.7 kb with LD values as revealed by mean $r^2$ of 0.36 across all chromosomes. The average $r^2$ value between chromosome ranged from 0.34 to 0.38, slightly higher on chromosomes 8 and 4 with $r^2$ 0.38 compared with chromosome 7 (0.34). Across entire genome, 17,112 pairwise SNPs were considered to be under high LD ($r^2 \geq 0.8$), most of which were present on chromosome 1 (16.3%) and least on chromosome 10 (6.5%). The variation in chromosome one and ten regarding number of SNPs in LD (Supplementary Figure 2). Chromosome

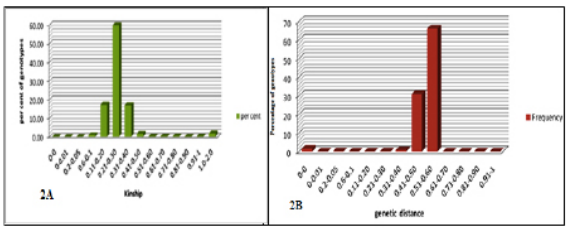


Figure 2 - Pairwise comparison among sixty four genotypes for kinship coefficient (2A) and genetic relatedness (2B).

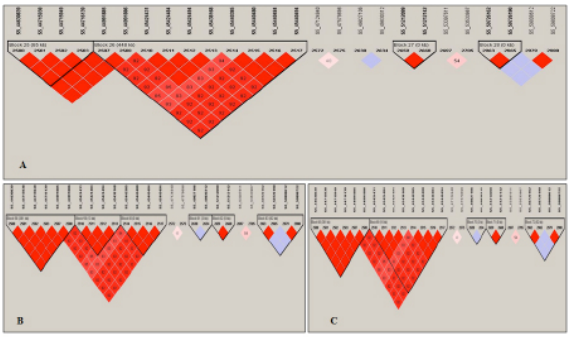


Figure 3 - Haplotype blocks ranging from 44.6 Mb to 59.8 Mb on chromosome five under three models A) Confidence intervals model, B) Four gamete rule and C) Solid spine of LD model viewed in Haploview. Inverted triangle presents haplotype block.

1 and 5 has minimum of 24.5 kb physical distance between pairs of markers and chromosome 4 reported to have maximum physical distance of 33.8 kb. The mean physical distance between pair of markers on chromosome six, seven and eight was 28.6 Kb. The haplotype patterns and LD visualization were analysed using 75,666 SNPs with Haploview 4.2v software with default window size under three models: 1) Confidence Intervals (CI), 2) Four Gamete Rule (FGR), and 3) Solid Spine (SS) model. The haplotype blocks varied according to chromosome and with changing model. As expected minimum number of haplotype blocks were identified in CI model compared to FGR and SS model. Under CI model, 616 halpotype blocks were formed across all the chromosomes. The chromosome five showed maximum of eighty four blocks with maximum average block length of 448 kb spanning 10 SNPs from 44.9 Mb to 45.4 Mb of physical distance. Similarly, FGR and SS models reported 1,172 and 1,270 haplotype blocks across genome with highest LD blocks of 167 and 182 on chromosome five respectively (Figure 3). The CI model showed minimum number of 39 blocks on chromosome six and SS model also reported the same trend of minimum blocks on chromosome six with 90 haplotype blocks. FG model reported 82 blocks on chromosome nine considered lowest among all chromosomes. Despite of minimum number of blocks, chromosome nine showed second largest haplotype block with 388 Kb length. The chromosome one reported second largest in having 77, 157 and 178 haplotype blocks from CI, FGR and SS models respectively. The chromosome wise number of haplotype blocks are presented in Table 1. Under CI model, chromosome seven showed smallest haplotype block lesser than 1 Kb.

Estimation of LD decay is very important in association mapping which depicts the minimum number of markers required to cover genome efficiently for mapping of traits. Association mapping with slow LD decay leads to low resolution map with less markers (coarse mapping). In contrast, fast LD decay uses ge-

Table 1 - Chromosome wise number of haplotype block under three distinct model 1) Confidence interval (CI), 2) Four Gamete Rule (FGR) and 3) Solid Spine model (SS).

| Chromosome | Model | | |
|---|---|---|---|
| | Confidence interval | Four gamete rule | Solid spine |
| 1 | 77 | 157 | 178 |
| 2 | 76 | 137 | 142 |
| 3 | 72 | 141 | 147 |
| 4 | 54 | 92 | 101 |
| 5 | 84 | 167 | 182 |
| 6 | 39 | 84 | 90 |
| 7 | 47 | 93 | 96 |
| 8 | 67 | 117 | 132 |
| 9 | 45 | 82 | 92 |
| 10 | 55 | 102 | 110 |

nome wide markers to tag gene of interest, yields high resolution (fine mapping). Genome-wide LD decay at MAF ≥ 1%, was 6.34 kb at $r^2$ = 0.2 and 18.27 kb at $r^2$ = 0.1. The visual display of genome wide LD decay and chromosome wise LD decay are presented in Supplementary Figure 3. Chromosome-wise LD analyses showed that the slowest LD decay was observed on chromosome 8 (45.75 kb, $r^2$ = 0.1), and fastest decay was found on chromosome 2 (12.2 kb, $r^2$ = 0.1). The chromosome wise LD decay is presented in Table 2. Similarly, slower decay at $r^2$ = 0.2 was found on chromosome 8 and fast decay on chromosome 2.

***Analysis of population structure in a sub-set of 64 tropical maize inbred lines***

The population structure among the 64 inbred lines was estimated using principle components estimated from SVS 7.7v software and GenABEL package of R software. Genetic structure analysis using both the programs revealed presence of three sub population. (Figure 4). Among 64 inbred lines 5% of individuals set to group 1 (G1), 8% to group 2 (G2) and 87% to group 3 (G3) (Supplementary Table 1). The inbred lines from G1 consist of early maturing lines which were mainly composed of CIMMYT-Asia lines and possessing distinct characteristic like susceptible to leaf firing, tassel blast and senescence. The sub population G2 consist of CLQ lines with characteristic feature like late maturing and resistance to tassel blast. Most of the CML lines and line with CML has one of the parent were grouped in group G3, which also consist of DTPW, POOL, POP, CLQ and CA lines. Most of lines in G3 were medium maturity group with dwarf nature and high yielding.

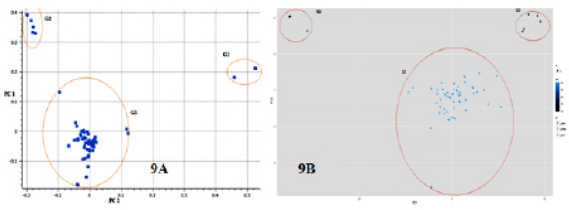


Figure 4 - Effect of Mancozeb treatment on Disease development at different developmental stages
Effect of mancozeb treatment on disease developmental parameters i.e. disease index (DI), disease severity (DS), rate of infection (R-Value) and area under disease progress curve (A-Value) at different disease developmental stages i.e. knee height stage, silking stage and tasseling stage.

## Discussion

In the present investigation, genetic diversity among sixty four genotypes was assessed using 1629 unlinked randomly distributed high quality SNPs with zero per cent missing data derived from GBS pipeline. The extent of genetic diversity in a population assessed by using polymorphic information content (PIC), gene diversity, kinship coefficient values and genetic distance between individuals (Semagn et al, 2012; Zhang et al, 2012; Thirunavukkarasu et al, 2013). The mean polymorphic information content (PIC) value of 0.37 was in accordance with Thirunavukkarasu et al (2013). The mean gene diversity among lines was 0.5 which is referred as maximum for biallelic markers (Van Inghelandt et al, 2010). Similar results were also reported by Dao et al (2014) among INERA maize inbred lines and Roy choudhury et al (2014) in rice using 36 SNPs which is in highly accordance with our findings. The high gene diversity may be due to the selection of high polymorphic SNP markers from large dataset. The average kinship coefficient among pair of individual was 0.28, which is less than the value of 0.37 that was reported by Semagn et al (2012) among 450 lines using 1065 SNP markers. A vast majority (96%) of pair wise comparisions have genetic distance value more than 0.4 indicating presence of diversity which is in agreement with the reports of Semegn et al (2012) and slightly higher than reports of Dao et al (2014). This may be due to difference in population used for study. The mean genetic distance reported in the present study of 0.49 which is in accordance with Wen et al (2011) who reported mean genetic diastance 0.54 between the 498 maize genotypes with 1,041 SNP markers. There was high genetic distance and low kinship coefficients among most pairs of lines, clearly indicating the uniqueness of the majority of the inbred lines in these maize breeding programs and which has potential to contribute new alleles in breeding program. In previous studies, many scientists opine that the selection of parental pairs based on genetic dissimilarity would be a good starting point to identify potential heterotic combinations (Dao et al, 2014; Semagn et al, 2012; Zhang et al, 2012).

Principal component analysis (PCA) has been proposed as an alternative to STRUCTURE software for studying population structure among genotypes (Patterson et al, 2006; Dao et al, 2014; Thirunavukkarasu et al, 2013). In the present investigation population structure was analysed using principle components derived from SVS 7.7v software and GenABEL package of R. The results from both the software were consistent and stratified population into three sub groups. The stratification of the population was in accordance with the pedigree information, as most

of the lines with similar pedigree tended to cluster into the same group. Our results were in consistence with reports of Warburton et al (2005), Semagn et al (2012), and Wen et al (2011) who also reported clustering of tropical maize lines based on pedigree information. The group 1 consist of early maturing genotypes contrary to group 2 which consist of late maturity. Our study thus provides information for developing new hybrids with different maturity dates by performing selective crosses between and within maturity groups that can cope up with high temperature.

*Linkage disequilibrium, haplotype and Linkage disequilibrium decay*

We characterized the genome wide and chromosome wise LD and LD decay in our set of inbred lines. Approximately, across genome 22.4% of SNPs with high LD ($r^2 > 0.8$) were scattered throughout the genome. The high LD regions were mostly interspersed with low LD regions, indicative of maize genome complexity and the random nature of recombination events across the genome (Tenaillon et al, 2001; Rafalski and Morgante, 2004). The mean $r^2$ value is higher on chromosome eight and four, similar report was found by Thirunavukkarasu et al (2013). This might be due to less recombinant events (Zhu et al, 2008) and directional selection to some specific traits (Ackay and Powell, 2007). More than sixty per cent of SNP pair across genome exhibit LD at $r^2 > 0.1$, which indicates presence of diversity in panel. Earlier authors Ching et al (2002) and Liu et al (2003), reported extent of LD detected by SNPs or SSRs was higher in narrow germplasm than in diverse germplasm. In the present study, LD decay declined rapidly with increased distance between markers. Similar trend was reported by Doa et al (2014), Thirunavukkarasu et al (2013), and Wang et al (2012). The decrease of the LD decay with increasing genetic distance indicates that, the portion of LD is conserved with linkage and proportional to recombination (Stich et al, 2005). In the present investigation average LD decay (18.7 Kb) is slightly higher than that observed in 632 (Yan et al, 2009) and 447 (Lu et al, 2011) maize inbred lines. However, LD decay reported in present investigation reveals presence of high genetic diversity. The reason for increase in LD distance is due to low sample size (64) used in the current investigation (Yan et al, 2009). There was difference in chromosome wise LD decay which is similar to the reports of Yan et al (2009). This may be due to the great variation in recombination rates along the chromosomes, including a low recombination rate in centromeric regions and a high recombination rate within genic regions due to retrotransposon insertions (Dooner and He, 2008). The haplotype blocks were indicative of magnitude of recombination across the genome. In the present study, haplotype blocks across genome assessed using three popular models, Confident Interval (CI), Four Gamete Rule (FGR) and Solid Spine (SS). The genome-wide SNP genotyping revealed a total of 616 haplotype blocks varying in size from <1 Kb to 448 Kb for CI model. FGR and SS model revealed 1,172 and 1,270 haplotype blocks respectively. The CI model identified fewer and shorter haplotype blocks than FGR and SS models. These results were in agreement with Thirunavukkarasu et al (2013) who reported FGR and SS models have more haplotype blocks. The change in haplotype blocks with model mainly due to change in high confidence interval bound cut-off values and algorithm. All these three models showed more haplotype blocks on chromosome 5 which may indicate fixation of alleles (Pfaffelhube et al, 2008).

Table 2 - LD decay estimated with 75,664 SNPs with MAF ≥ 0.1 across chromosomes and chromosome wise.

| chromosome | LD decay | |
|---|---|---|
| | $R^2 = 0.1$ | $R^2 = 0.2$ |
| 1 | 23.11 | 8.02 |
| 2 | 12.42 | 4.31 |
| 3 | 14.02 | 4.87 |
| 4 | 28.83 | 10 |
| 5 | 13.46 | 4.66 |
| 6 | 13.03 | 4.52 |
| 7 | 16.73 | 5.8 |
| 8 | 45.75 | 15.88 |
| 9 | 23.18 | 8.04 |
| 10 | 17.25 | 5.98 |
| Across genome | 18.27 | 6.34 |

In conclusion, the present study clearly revealed significant molecular diversity among the maize inbred lines selected for heat tolerance study. There was high genetic distance and low kinship coefficients among most pairs of lines along with rapid LD decay, clearly indicating the uniqueness of the majority of the inbred lines which can contribute new alleles for heat tolerance traits in breeding programme. The haplotype blocks length across all chromosomes was relatively smaller, revealing high occurrence of recombinations and allelic diversity in maize inbred lines. The population stratification, resulted most of the lines in accordance with the pedigree and breeding programme. The selected markers can be used for diversity assessment in tropical germplasm for low cost genotyping. The results from this study will be useful to breeders in selecting best parental combinations to exploit heterosis for varied demands.

## References

Ackay I, Powell W, 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 122: 57–63

Anonymous, 2014. India maize summit, pp. 1-27

Barrett JC, Fry B, Maller J, Daly MJ 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265

Beyene Y, Botha A, Alexander AM, 2006. Genetic diversity among traditional Ethiopian highland maize accessions assessed by simple sequence

repeat (SSR) markers. Genet Res Crop Evol 53: 1579-1588
Botstein D, White RL, Skolnick M, Davis RW 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331
Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E, 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19): 2633-2635
Ching A, Caldwell KS, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski MA, 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite Maize inbred lines. BMC Genetics 3: 19
Cholastova T, Knotova D, 2012. Using morhological and microsatellite (SSR) Markers to assess the genetic diversity in Alfalfa (*Medicago sativa* L.). Int J Biol Food Vet Agric Eng 6(9): 146-152
Dao A, Sanou J, Mitchell SE, Gracen V, Danquah EY, 2014. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. BMC Genetics 15: 127
Dooner HK, He L, 2008. Maize genome structure variation: Interplay between retrotransposon polymorphisms and genic recombination. Plant Cell 20: 249-258
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE, 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS One 6(5): e19379
Gupta P, Rustgi S, Kulwal P, 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57: 461-485
Hartings H, Berardo N, Mazzinelli GF, Valoti P, Verderio A, Motto M, 2008. Assessment of genetic diversity and relationships among maize (*Zea mays* L.) italian landraces by morphological traits and AFLP profiling. Theor Appl Genet 117: 831–842
Liu K, Muse SV, 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129
Liu KJ, Goodman M, Muse S, Smith JS, Buckler ES, Doebley J, 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165: 2117-2128
Lu Y, Yan J, Guimaraes GT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C et al, 2009. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120: 93-115
Patterson NJ, Price AL, Reich D, 2006. Population structure and eigen analysis. PLos Genetics 2: e190
Pfaffelhuber P, Lehnert A, Stephan W, 2008. Linkage disequilibrium under genetic hitchhiking in finite populations. Genetics 179: 527-537
Prasanna BM, 2012. Maize production in a changing climate: Impacts, adaptation, and mitigation strategies. Adv Agron 114: 1-65
Rafalski A, Morgante M, 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20(2): 103–111
Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. The National Academy of Science 98: 11479-11484
Roy Choudhury RD, Singh N, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Ahmad A, Singh NK, Singh R, 2014. Analysis of genetic diversity and population structure of rice germplasm from north-eastern region of India and development of a core germplasm Set. PLoS ONE 9(11): e113094
Semagn K, Magorokosho C, Bindiganavile SV, Makumbi MD, Beyene Y, Mugo S, Prasanna BM, Warburton ML, 2012. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BioMed Genomics 13: 113
Smith OS, Smith JSC, 1992. Measurement of genetic diversity among maize hybrids; A comparison of isozymic, RFLP, pedigree, and heterosis data. Maydica 37: 53–60
Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111: 723–730
Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS, 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). Proc Natl Acad Sci USA 98(16): 9161–9166
Thirunavukkarasu N, Hossain F, Shiriga K, Mittal S, Arora K, Rathore A, Mohan S, Shah T, Sharma R, Namratha PM et al, 2013. Unraveling the genetic architecture of subtropical maize (*Zea mays* L.) lines to assess their utility in breeding programs. BMC Genomics 14: 877
Van Inghelandt D, Albrecht EM, Claude L, Benjamin S, 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120: 1289-1299
Warburton ML, Reif JC, Frisch M, Bohn M, Bedoya C, Xia XC, Crossa J, Franco J, Hoisington D, Pixley K, Taba S, Melchinger AE, 2008. Genetic diversity in CIMMYT nontemperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. Crop Sci 48: 617-624
Warburton ML, Ribaut JM, Franco J, Crossa J, Dubreuil P, Betrán FJ, 2005. Genetic characteriza-

tion of 218 elite CIMMYT inbred maize lines using RFLP markers. Euphytica 142: 97–106

Weir BS, 2008. Linkage disequilibrium and association mapping. Annu Rev Genomics Hum Genet 9: 129–142

Wen W, Araus JL, Shah T, Cairns J, Mahuku G, Bänziger M, Torres JL, Sánchez C, Yan J, 2011. Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci 51: 1-13

Xia XC, Reif JC, Hoisington DA, Melchinger AE, Frisch M, Warburton ML, 2004. Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: I. Lowland tropical maize. Crop Sci 44: 2230-2237

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch JH, 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4(12): e8451

Zhang Q, Wu C, Li Y, Zhang C, 2012. Association analysis of important agronomical traits of maize inbred lines with SSRs. Aust J crop sci 6(6): 1131-1138

Zhu C, Gore M, Buckler ES, Yu J, 2008. Status and prospects of association mapping in plants. The Plant Genome 1: 5-20