Molecular Diversity Study in Tropical Maize inbred lines Using Microsatellites Markers

NITIN KUMAR[1, 2], RAJENDRA KUMAR KHANNA[1], BRIJ MOHAN MEENA[3], [4]RANBIR SINGH, BHUPENDER KUMAR*[2]
[1]Department of Basic and Applied Sciences, Vivekananda Global University, Sector 36, Sisyawas, NRI Road, Jagatpura, Jaipur, Rajasthan-303012, India.
[2]ICAR- Indian Institute of Maize Research, PAU Campus, Ludhiana, Punjab-141004, India.
[3]Department of Zoology, Rajgarh P.G. College, Alwar, Rajasthan-301408, India.
[4]Devision of Water Technology Centre, Indian Agriculture Research Institute, New Delhi-110012, India.

Abstract:

Background:

Maize Tropical (*Zea mays L.*, 2n=20) used in India has high complexity because of the diverse lines of this area. In the study, study, the genetic diversity and population structure decay of 352 selected tropical inbred lines collected from the breeding plan of India. Characterized using the maize lines with 70 Microsatellites.

Results:

With respect to molecular diversity, six sub grouped were identified. Markers were amplified and shown high polymorphism. With increase K value, the tropical group showed vertical structure with division into further groups All these SSR markers were select to detect high polymorphism. Primers ware showed in various bands showing different types of alleles.

Conclusion:

For the panel most of widely used tropical inbred lines in India, this work representatively not only illustrates the foundations and evolutions trend of maize inbred panel as a experimental reference for the improvement of tropical maize, but also provide Population structure were investigated using three complementary analysis methods such as the NTSYS and STRUCTURE, statistical analysis based on Microsatellites markers data..

Key words: Tropical group, Genetic Diversity, Maize, Inbred, Cluster Analysis, SSR marker

Background:

Maize (Zea mays L.), which is one of the important crop in the India and worldwide, and exceeds rice and wheat in production, plays an essential role in global food security (USDA). With a highly demand for maize, raising its production is an urgent challenge today. It is an open-pollinated species with a complex genome (Schnable *et al*., 2009), there is astounding genetic diversity (GD) in the maize genome and it is designed a considerable factor in heterosis (Gore *et al*., 2009). From the 1930s to the current, three stages in maize inbred selfing history have been defined according to the source of parents: (a) Inbred lines directly derived from tropical during the 1930s–1950s; (b) Inbred lines derived from crosses among artificial selected inbred lines during the 1950s–1980s and (c) Modern commercial breeding programs have brought about more than six-fold greater grain yields than those in previous decades (Wu *et al*.,2014). Inbred lines developed from cultivating elite inbred lines for commercial use (Lee., 2007). In heterosis, a hybrid offspring shows superior performance to the parents. At the genetic level, it is contributed by variation in the presence of genes or novel beneficial alleles and gene expression modification (Whitt *et al*., 2002). Anyhow, owing to the number of valuable loci targeted during artificial selection, the GD of maize has gradually narrowed during the breeding process (Springer *et al*., 2007). Theoretically, to maximize heterosis in maize, two inbred lines separated by a large genetic distance are selected as parents for commercial hybrids. Thus, identification of heterotic groups and patterns is the foundation of hybrid breeding. Previously, breeders assigned different inbred lines to specific groups using pedigrees, morphological traits and testcrosses. However, it is difficult to accredit groups definitely by relying only on depend information, particularly for lines with similar phylogenic background and complex sources. Currently, SNPs are

widely used because of efficient cost and high throughput (Yan *et al*., 2009). For this reason, molecular markers are widely used for this purpose. Application of molecular markers has undergone three main stages, from tens of restriction fragment length polymorphism markers (RFLPs) to hundreds of SSRs to millions of SNPs. During the last 30 years, many studies have focused on worldwide maize germplasm diversity. An early study used RFLPs to assign 148 inbred lines to two main groups, Iowa Stiff Stalk Synthetic (BSSS) and Lancaster, represented by B73 and Mo17, respectively (Mumm *et al*., 1994). In a large-scale analysis, 96 SSRs were used to genotype 964 individuals representing almost the entire set of ~350 races native to the Americas. The entire panel was divided into four main clusters: highland Mexican, northern United States (U.S.), tropical lowland, and Andean races (Liu *et al*., 2003). The analysis showed that the southwestern U.S. was a transition area between Mexico and the northern U.S. Then, another study genotyped 94 SSRs to separate 260 inbred lines into four groups: Tropical or Semitropical (TS), Stiff Stalk (SS), non-Stiff Stalk (NSS), and a mixed group. TS group showed the highest GD (Vigouroux *et al*., 2008). More recently, a new and low-cost SNP-genotyping technology named genotyping by sequencing (GBS) was used to genotype 2,815 maize inbred accessions in the U.S. seed bank. The results showed that the international germplasm pools were different from those commonly used in North America (Lu *et al*., 2009). In a later study, using the Golden Gate SNP chip, 770 inbred lines from CIMMYT, China and Brazil were clustered into eight groups, covering temperate and tropical germplasm and a special Indian group Sappington (SPT) (Romay *et al*., 2013). Maize was introduced into India approximately 600 years ago. Despite the narrow genetic background, there are two reasons for the complexity of Indian germplasm: (1) Chinese breeding programs have always integrated landraces and introduced germplasm and there is no complete record of this process (2) Different germplasm was introduced into different areas and in different periods, and new germplasm is introduced continuously and (Li Y *et al*., 2002). For these reasons, systematic study of Chinese germplasm resources is important, particularly in the era of high-throughput SNP genotyping. Indian inbred lines are classified into four to six heterotic groups generally including Reid, SPT, and Lancaster as major groups according to the results of molecular characterization and pedigree. Using 70 SSRs and 1,034 SNPs, 187 and 282 Indian inbred lines were grouped into six same clusters: PA, PB, Luda Red Cob (LRC), BSSS, Lancaster and SPT almost at the same time, respectively (Xie C *et al*., 2007). Recently, using the 56 k MaizeSNP50 Bead Chip, 1,015 SNPs were screened to classify 367 inbred lines into five subgroups containing P and Tem-tropic I in addition to the three major groups. Anyhow, given that tropical germplasm and landraces for commercial hybrids are widely used in Southwest China because of a variety of breeding goals that contribute to a large different series of inbred lines comparing to other regions, thus these divisions of heterotic groups above are not sufficient for practical relevance. With the development of high-throughput SNP genotyping, genome-wide association studies (GWAS) are widely used for gene mapping of complex quantitative traits. However, besides population structure, which is considered an essential factor resulting in false-positive estimation, linkage disequilibrium (LD) also determines the resolution of GWAS results by the decay distance corresponding to the marker coverage (Remington *et al*., 2001). More rapid LD decay corresponds to shorter LD decay distance, requiring high marker density to cover functional loci and achieve high resolution for association mapping. In maize, the LD decay distance varies from less than 1 kb in landraces to more than 100 kb in commercial inbred lines (Yu J., 2006). Current genome marker coverage is much higher than the previous coverage and allows fine estimation of LD decay distance in the genome-wide profile. Moreover, different populations and different segments of chromosomes always show varying LD. For example, in a study using a 1536-SNP array to analyze LD decay distance in 632 lines from temperate, tropical, and subtropical breeding programs, the LD decay distance among 10 chromosomes ranged from 1 to 10 kb (Yan J *et al*., 2009). In a comparison of temperate, tropical, and subtropical lines, the LD decay distance in temperate lines was much higher than that of tropical and subtropical lines, owing to lower diversity and less rare alleles (Lu *et al*., 2011). Recently, up to 681,257 SNPs were used for GWAS in 384 inbred lines from the Ames panel. The LD decay distance of that was close to 10 kb (Pace *et al*., 2015). It implies that different degrees of panel diversity, marker density, and methods contribute to the variation in LD decay distance estimates. However, previous studies have focused on population structure and GWAS in global germplasm, but for Southwest China, the complex ecological conditions in maize production contribute to the different germplasms used in this region so that pedigree of some local and new cultivated inbred lines are not clear. Besides, for the excellent resistance to the low soil fertility, severe disease caused by high

temperature, wet, and less sunlight during maize growing period, the widespread use of tropical and semi-tropical lines for hybrids is the greatest characteristic in this region. As a consequence, these aspects contribute to the complexity in the construction of heterotic groups and heterotic pattern. For the present study, 362 diverse inbred lines widely used in Southwest China were collected for genotyping with the MaizeSNP50 BeadChip. Systematic characterization of germplasm including hererotic group clustering, GD comparison, and LD estimation was performed for the purpose of characterization of breeding resources and simplification of heterosis models in India.

Material and method:

Plant material

352, tropical maize genotype including exotic accessions inbred lines and breeding lines from the public sectors were used for genetic diversity. These genotypes were taken from the germplasm collection maintained at IIMR, New Delhi, CIMMYT, AICRP and NBPGR. 70 Microsatellites markers were taken for polymorphism in genes.

SSR Markers and Genotyping

Genomic DNA was extracted from approximately 200 mg fresh leaf tissue using the cetyl tri-methyl ammonium bromide (CTAB) method (Saghai-Maroof *et al.*, 1984). 70 Microsatellites primers, which were distributed evenly over the six maize chromosomes, were selected based on the information available in the Maize-GDB database (http://archive.maizegdb.org). PCR amplifications were carried out in 20 µl reaction volumes containing 20-30ng of2µL template DNA, 1µM each of 3.0µL primer, 5x 2.25 Taq-buffer, 0.05µL of 5 units µL-1 Taq DNA polymerase, 2.5µl dNTPs of 1µL, 25mM $Mgcl_2$ of 0.60 µl and dH_2O 10µL. PCR protocols consisted of 32 cycles of 94 for 45s, an annealing temperature at either 45, 50, 55 or 60°C depending on the individual SSR primers for 45 s, and 72°C for 60s, and a final extension step of 72°C for 10 min. PCR products were analyzed by 3.5% Metaphor gel electrophoresis and visualized by blue dye staining with gel-doc.

Genetic Diversity Analysis

For each SSR locus, polymorphic bands were scored as one or in respected to presence or absence of the bands at the same mobility, respectively. Gene diversity (PIC) ware calculated for each marker using the formula: $PIC=1-\sum fi2$, where fi is the allele frequency for the i-th locus summed across all alleles for that locus. (Liu *et al*., 2005) used the program Power-Marker v3.25 to calculate allele number, allele frequency, and genetic diversity of each locus as mentions.

Results:

70 randomly selected polymorphic SSRs marker was amplified in 352 genotypes of maize, which include tropical inbred line collection (which were collected from different states of India). From 0.63 to 0.83 the PIC value for these 70 polymorphic SSRs markers with respect to 352 genotypes. The average of polymorphism information content for the markers was 0.57±0.19. Maximum PIC value was noticed for umc1332 (0.69) followed by Phi 049 (0.37) and umc1232 (0.83), while the lowest PIC value (0.01) was found in bnlg1092 (0.33), umc2208, and Phi 92. Majority of the SSRs (60%) showed >0.50 PIC value. There were only four SSRs which showed PIC value in ranges 0.63 to 0.83 (Table.1) The highest PIC value was observed in individual markers umc2303 (0.83) indicated that this marker had shown high allelic variation among all. However, Lowest PIC value indicated the low allelic variation. The groups wise highest PIC value ware observed 0.67 in bnlg252, while; lowest PIC value was found in-group Phi033 markers with value of 0.54. Highly polymorphism of band in the present study was reported in primer umc1232, *umc1332, umc2303* (Fig. 4).

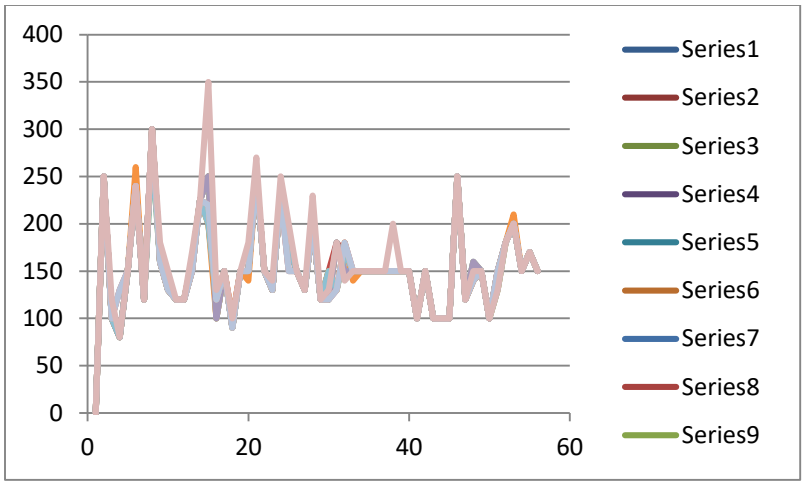


Fig. 1 Series 1 – 9 chromosome number and PIC value

Table: 1. SSR Markers identified which have shown polymorphism (PIC>0.50) among all tropical maize lines

S.No	Markers	Forward Sequence	Reserves Sequence	Chromosome No	PIC Values
1.	umc1232	GGAATTACCACAACAAACTAAACTTGG	AGGCTCTAGCTACCTGGCTACGTT	4	0.69
2.	umc1332	CCTCTTGCTTCCTCGTCATGTACT	AAGGAGCTGGAACATAAAACACCA	5	0.83
4	umc2303	AGAAGAAGGTGGAGGTCCAAGACT	CTGGTATCTGATCAGGGTGCG	5	0.83

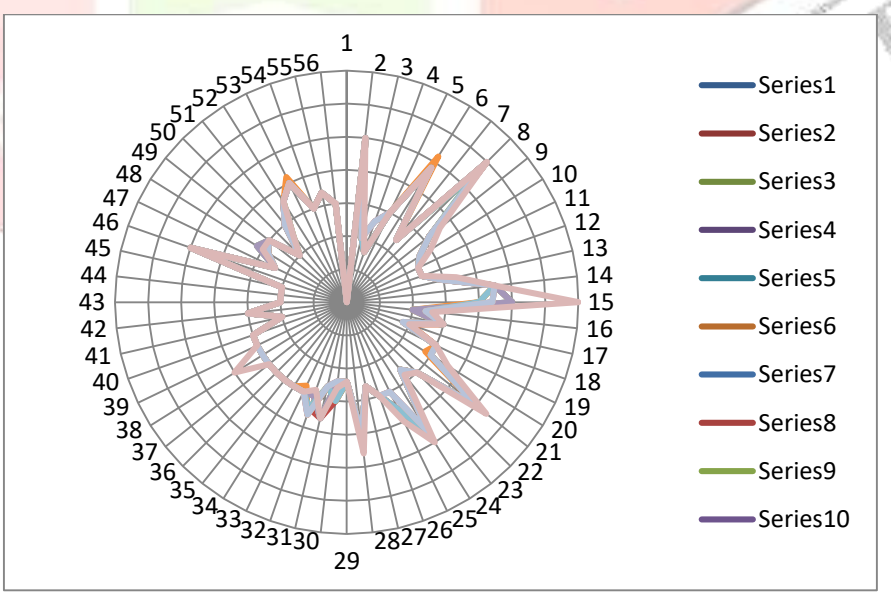


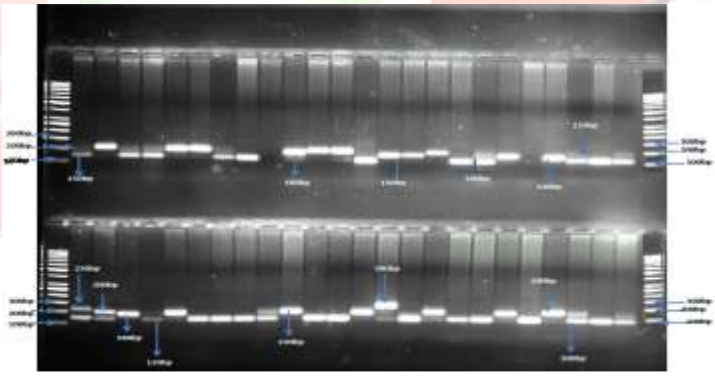
Figure: 2. SSR markers with high PIC value (> 0.97) sky blue color represents markers number and color brown genotypes.

Based on the maximum probability, 352 inbred lines were assigned into 6 subpopulations (Red, Green, Blue, yellow, Purple and Sky blue) (Fig.2). Red color included 64 tropical maize inbred lines closed to yellow color of 57 maize lines in genetic backgrounds. Green color had 24 inbred lines related to green color in genetic background. Blue color had

included 22 inbred lines. Purple color had included 52 tropical maize inbred lines. Sky blue color had 38 inbred lines at six genetic backgrounds, (Table 1) which described as population structure subgroup as described by (Xie *et al*. (2008)
The similarity coefficient among 288 maize inbred were lines ranged from 0.1 to 0.97 with average showing 0.40. While the similarity coefficient was 0.62, cluster analysis of 352 inbred lines cleared grouped into six subpopulations also clustering were consistent with their assignments using STRUCTURE software (Fig.2). Moreover, showed six subpopulations.

Figure: 3-. Genetic diversity of 352 maize inbred lines

70 SSR markers were analysed in the 352 maize inbred lines into six major groups. Inbred lines in Red were mainly distributed in between 2 to 6 cluster group, green distributed in between 1 to 5 group, blue distributed in between 4 to 5 group, yellow distributed in between 3 to 6 group, purple distributed in between 2 to 3 group, sky blue distributed in between 1 to 4 group. Most individuals within Red and sky blue subpopulations were grouped more closely. As inferred analysis by STRUCTURE software. The green color II group indicates higher genetic diversity subpopulations.



(a)

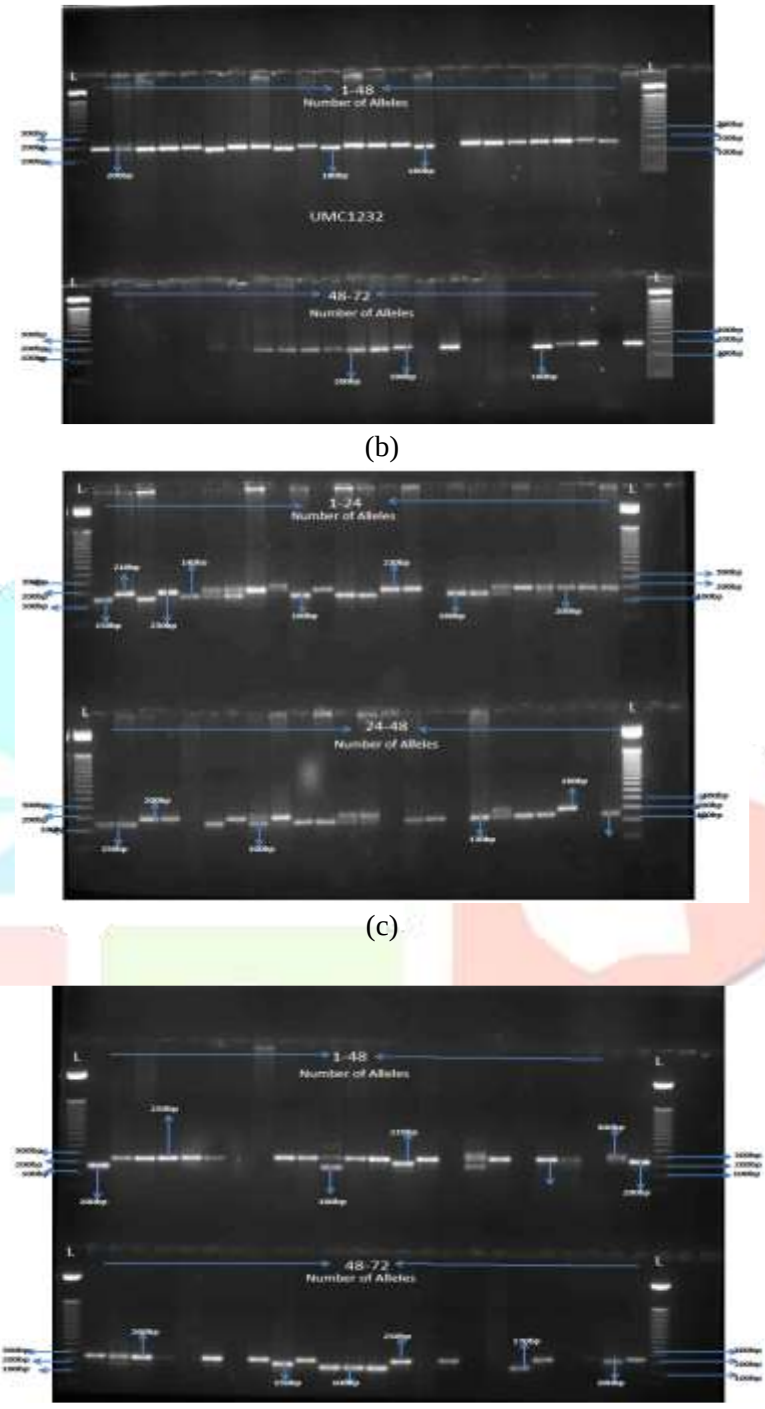


Figure: 4. SSR markers profiling for umc1232, (c) umc1332 and (d) umc2303 have been depicted in gels with (primers 100 ladder L).

Discussion and Conclusion:

Using the maizeGDP, we performed the genetic diversity and population structure distance analysis of 352 important inbred lines selected from the breeding programme of public sector in India. We identified one three four and six subgroups according the bio-informatics study and field experiment by the structure software which NTSYS and STRUCTURE , Markers SSR due to their co-dominance, locus specificity and abundance, have been extensively used to assess genetic diversity in tropical maize inbred lines (Sarcevic *et al*., 2008). In the present study, all the 352 tested maize

inbred lines were derived from public sector inbred selected lines. The average polymorphic information content (PIC) value was 0.01, which was lower than that in highest inbred lines with PIC over 0.97 (Wang *et al*., 2008; Xie *et al*., 2008). In our study, alleles were obtained at the genome level Chromosome 1 showed the lowest allele number 38 genotypes and the highest alleles is 57. Therefore, we have thus determined that there is a higher-level genetic diversity in the 352 tropical maize-selected lines, which has the potential to enhance the genetic diversity of Indian and foreign maize breeding materials. Nevertheless, for the 288 tested maize inbred -selected lines, the pedigree information was not in accordance with their clustering. Therefore, it is of significant importance to understand population structure and genetic diversity among inbred lines is for maize improvement. It means that highest frequencies of polymorphism shown the susceptibility and survival in adverse condition.

Acknowledgements:

Authors are indebted to the Indian Council of Agricultural Research and ICAR-Indian Institute of Maize Research, New Delhi and Vivekananda Global University, Jaipur for financial support under in fellowship. Authors are also thankful to Director, IIMR and Project Director, Department of Genetic Plant Breeding and for grant of permission to conduct the present study.

Reference:

1. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., Allard, R. W.: Ribosomal DNA spacer length polymorphisms in barley mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad*. Sci., 81, 8014-8018 (1984).
2. United States Department of Agriculture (USDA) Foreign Agricultural Service. http://www.fas.usda.gov/data/world-agricultural-production. Accessed 9 Apr 2015.
3. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA. The B73 maize genome: complexity, diversity, and dynamics. *Science.* (2009), 326(5956),1112–5.
4. Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J. A first-generation haplotype map of maize. *Science.* (2009), 326 (5956), 1115–7.
5. Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, Li Y. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theor Appl Genet*. (2014) 127(3), 621–31.
6. Lee EA, Tollenaar M. Physiological Basis of Successful Breeding Strategies for Maize Grain *Yield. Crop Sci*. (2007), 47 (Supplement_3), S-202.
7. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci* U S A. (2002) 99(20), 12959–62.
8. Springer NM, Stupar RM. Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Res*. (2007)17(3), 264–75.
9. Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y. High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol Breed*. (2009)25(3), 441–51.
10. Mumm RH, Dudley J. A classification of 148 US maize inbreds: I. Cluster analysis based on RFLPs. Crop Sci. (1994)34(4), 842–51.
11. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics*. (2003)165(4),2117–28.
12. Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez J, Doebley J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am J Bot*. (2008) 95(10),1240–53.
13. Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor Appl Genet*. (2009)120(1),93–115.
14. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. (2013)14(6),R55.

15. Li Y, Du J, Wang T, Shi Y, Song Y, Jia J. Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. *Maydica*. 2002;47(2):93–102.
16. Teng WT, Cao JS, Chen YH, Liu XH, Jing XQ, Zhang FJ, Li JS. Analysis of Maize Heterotic Groups and Patterns during Past ecade in China [J]. *Sci Agric Sin*. (2004)12, 003. in Chinese.
17. Wang R, Yu Y, Zhao J, Shi Y, Song Y, Wang T, Li Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor Appl Genet*. (2008) 117(7), 1141–53.
18. Xie C, Zhang S, Li M, Li X, Hao Z, Bai L, Zhang D, Liang Y. Inferring genome ancestry and estimating molecular relatedness among 187 Chinese maize inbred lines. *J Genet Genomics*. (2007) 34(8),738–48.
19. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci* U S A. (2001)98(20), 11479–84.
20. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol*. (2006)17(2),155–60.
21. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One.* (2009)4(12), 8451.
22. Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Prakash AB, et al. Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. *PLoS One*. (2011)6(9), 24861.
23. Pace J, Gardner C, Romay C, Ganapathysubramanian B, Lubberstedt T. Genome-wide association analysis of seedling root development in maize (Zea mays L.). *BMC Genomics.* (2015)16,47.
24. Xie, C., Warburton, M., Li, M., Li, X., Xiao, M., Hao, Z., Zhao, Q., Zhang, S.: An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. *Mol Breeding*., (2008) 21, 407-418.
25. Wang, R., Yu, Y., Zhao, J., Shi, Y., Song, Y., Wang, T., Li, Y.: Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor Appl Genet*., (2008), 117, 1141-1153.
26. Sarcevic, H., Pejic, I., Baric, M., Kozumplik, V.: Originality of M3S maize population and changes in allele frequencies revealed by SSR markers after two cycles of selfed progeny recurrent selection. *Euphytica.*, (2008), 161, 97-105.