Australian Journal of Crop Science AJCS
AJCS 5(13):1790-1795 (2011) ISSN:1835-2707

Evaluation of the genetic diversity and genome-wide linkage disequilibrium of Chinese maize inbred lines

Ming Wang[1], Xiaobo Zhang[1], Jiuran Zhao[2], Wei Song[2], Yonglian Zheng[1*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, 430070, Wuhan, China
[2]Maize Research Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, P.R. China

***corresponding author: yonglianzheng@gmail.com**

Abstract

Investigation of the genetic diversity, population structure, and linkage disequilibrium of maize will assist in the selection of parental lines for enhanced efficiency of maize breeding. In the present research, we investigated genetic diversity, linkage disequilibrium, and population structure among 173 inbred lines that were commercially important and/or were parental lines used for breeding in China. Using the model-based Bayesian clustering analysis, these lines could be assigned to four subgroups, Lan, P, Reid, and TSPT, most of which were in agreement with the pedigree of information. Using 78 SSR markers, the genetic diversity was determined to be an average of 8.1 alleles per locus (range 2 to 17). The average values of PIC (Polymorphic information content) and gene diversity were 0.667 and 0.704, respectively. A total of 547 pairs of SSRs on the same chromosome demonstrated significant linkage disequilibrium (LD) at the 0.01 level. Our results suggested that this population may be suitable for future marker-phenotype association analysis. In addition, the proportions of the observed LD were higher than that those of expected LD, indicating that we be careful for a high risk of spurious association, which could be generated by population stratification when using association mapping. The genetic diversity, population structure, and LD analysis in this study provided provide a basis for future association mapping for genes and quantitative trait loci.

Keywords: population structure, linkage disequilibrium, maize, gene diversity.
Abbreviations: *SSR*, Simple Sequence Repeat, *LD*, linkage disequilibrium, *TSPT*, Tangsipingtou.

Introduction

Maize (*Zea mays L.*) is one of the most economically important crops. Maize is also considered as a model organism for communicating the benefits of genetics and breeding because of its diversity of striking phenotypes. Moreover, maize has been the subject of extensive genetic research, including the construction of linkage maps (Burr et al., 1988), quantitative trait locus mapping (Edwards et al., 1987; Austin et al., 2001), molecular evolution (Henry and Damerval, 1997; Ching et al., 2002), developmental genetics (Poethig 1988; Fowler and Freeling, 1996), and physiological genetics (Crosbie et al., 1978). Conventional linkage mapping and the recently developed linkage disequilibrium-based association mapping have been employed to construct genetic maps (Xu et al., 2009). In comparison with conventional linkage mapping, which dissects complex traits, QTL mapping based on linkage disequilibrium (LD) (also called association mapping) is a promising approach to mine more alleles for genetic variation, and to bridge the gap between phenotype variation and genetic factors (Yu and Buckler, 2006). Furthermore, association mapping offers more power and higher resolution for QTL mapping, in addition to being less laborious and time consuming (Flint-Garcia et al., 2005). Association mapping was first used in the plant kingdom in 2001 (Thornsberry et al., 2001) and has been widely applied in plant species (Thornsberry et al., 2001; Zhao et al., 2007; Myles, et al., 2009; Yan et al., 2011). Association mapping is defined as non-random association of alleles at different loci, usually involved in linkage disequilibrium (LD). LD has become an important aspect in studies on the population structure in animals and plants, and in human evolution (Agrama and Eizenga, 2008). Differences in the extent of LD are dramatically affected by the breeding behavior of a species (Flint-Garcia et al., 2003). Generally, LD in self-pollinated species extends over a much longer distance than in cross-pollinated crops, because self-crossing reduces opportunities for recombination (Yan et al., 2011). For instance, in outcross species, such as grapevine, LD decays within 300bp (Lijavetzky et al., 2007), but in inbred systems, like rice (*Oryza sativa*), LD can be dramatically extended up to 120kb (Huang et al., 2010). However, LD is also dependent on the target population structure. The distance of LD decay is approximately 100 kb for commercial elite inbred lines, but can be as low as 2000 bp for diverse maize inbred lines, and 1000 bp for maize landraces (Yan et al., 2011). Other factors, such as genetic drift, natural and artificial selection, and admixture of different populations, could increase or decrease the extent of LD. A number of molecular techniques have been developed to investigate genetic diversity and population structure in plants. Among them, simple sequence repeats (SSRs) are the most useful for distinguishing closely related germplasms (Zhang et al., 2008). SSRs possess numerous advantages, such as high levels of polymorphism and even distribution across the

genome. They also provide codominant, accurate and reproducible data. Other types of molecular markers, such as AFLPs and SNPs, have been used to investigate both genetic diversity and population structure in several crops (Hyten et al., 2008; D'hoop et al., 2010; Zhang et al., 2010). Several studies have investigated population structure and genetic diversity with SSR markers in maize (Wang et al., 2008; Xie et al., 2007, 2008). The objectives of this study were to: (1) investigate genetic diversity; (2) assess the population structure; and (3) determine the genome-wide patterns of linkage disequilibrium.

Results

Population structure of inbred lines

To assess the genetic structure of the association mapping population, an admixed model-based approach using the program STRUCTURE subdivided the lines into appropriate subgroups. The results are shown in Fig.1. The number of subgroups, K, (equal to the number of populations) was four. The model-based subgroups are largely consistent with the known pedigree of accessions if K=4. The four subgroups are Lancaster, Reid, Tangsipingtou (TSPT), and P (Fig.1). The Reid subgroup is the largest subgroup, with 33 inbred lines. Many of this population were derived from foreign germplasms. For example, Tie7922 was selected from the American maize hybrid 3382, and Shen5003 was derived from the American maize hybrid 3147. In turn, some inbreds originated from Shen5003, such as 3189, Dan9046, and Liao3053. The subgroup TSPT, which included 29 inbred lines, was the second largest subgroup. Many members of this subgroup were derived from Huangzaosi, which is one of the founders in Chinese maize breeding programs. The Lancaster group included the Mo17 pedigree and the Zi330 pedigree. Longkang11, 4F1, and Ji1037, derived from Mo17, were also classified into the Lancaster group. In addition, some Ludahonggu germplasms, such as Dan340, were also assigned to the Lancaster group. Lastly, the smallest subgroup, P, comprising 18 inbreds, contains members that were also mostly derived from foreign germplasms. There is another subgroup that had <0.7 membership in each of the other four subgroups and was assigned as a mixed subgroup (Table S1).

Profile of SSR diversity

Seventy-eight pairs of SSR loci distributed randomly across the whole genome were employed to survey 173 diverse maize inbred lines. A total of 634 alleles were obtained from the 78 SSR loci, with an average of 8.1 alleles per locus (range, 2 to 17). The average value of PIC was 0.667 (range, 0.171 to 0.891). The gene diversity averaged of 0.704 (range, 0.188–0.899) (Table 1). In addition, the values of statistical parameters are not equal among the four subgroups. The number of alleles and allele per locus in the Reid subgroup were the largest, at 428 and 5.487, respectively, and smallest in the P group at 257 and 3.259, respectively. Their values were positively correlated with the size of the subgroups. The gene diversity in group P was lower than in the other three groups. This trend was also observed for the PIC value. Among 634 alleles, 22.082% were subgroup-specific. The number of subgroup-specific alleles in the TSPT group was much higher than that in P. The TSPT group also had more line specific alleles compared with that of the P group. This suggests that the TSPT group embodies more diverse germplasm and that the genetic background of the P subgroup is narrow.

Linkage disequilibrium

Among the 78 SSRs across the whole genome, LD was assessed in all the inbred lines and in each of the subgroups using TASSEL, with the permutation of Fisher's exact test (100,000 permutations). The results are shown in Table 3. The 78 SSR loci are distributed on 10 maize chromosomes, covering 72% of the entire genome. The result showed that 72 pairs of loci on the same chromosome were in LD at a significance level of 0.01. The 116 and one intrachromosomal pairs were linked at $r^2 > 0.1$ and $r^2 > 0.01$, respectively. Furthermore, 547 pairs of loci at different chromosomal locations showed significant LD in these lines at the 0.01 level. One was linked at $r^2 > 0.01$.

Discussion

Chinese maize germplasm

Chinese maize germplasms have distinct historical and geographical characteristics, contributing to population stratification. On the one hand, the history of the introduction of foreign germplasms on a large scale is short, although maize was first introduced into China in approximately 1530 (Li, 1998). Over the last 50 years, a large number of foreign maize germplasms, such as 3147, 78599, and 3382, were introduced into China and hybridized with our landraces. These materials thus formed two classes of foreign populations, termed the P and Reid groups. On the other hand, because of great differentiation of the climate in China, Chinese maize germplasms have suffered from selection pressure under different natural conditions. The Ludahonggu group is a good example; it originated from a landrace, Ludahonggu, in Luda, in Liaoning Province of Northeast China. The climate of Luda was advantageous, being rainy with little sunshine, but warm and humid. This area also has a high disease prevalence and a monsoon climate in the summer.Thus, the Chinese maize germplasm is more complicated than others, such as those from Europe and America, which are divided into only two groups each, including flint and dent germplasm groups in Europe, and Iowa stiff stalk (SS) and non-stiff stalk (NSS) germplasm groups in the United States (Stich et al., 2005).

Population structure

Evaluation of the population structure of Chinese accessions is essential for conservation, management, and utilization of these genetic resources. Many factors, such natural history and breeding system, can influence the population structure. It is difficult for breeders to distinguish which germplasm should be assigned to the corresponding subgroup because of the genetic complication of Chinese maize germplasms. However, Chinese maize germplasms embody four to six subgroups according to both the pedigree information and combining ability. In the current research, Chinese maize germplasms could be divided into four subgroups, Reid, Lancaster, TSPT and P. In a similar report, Wang et al. (2008) analyzed 288 inbred lines and subdivided them into four subgroups according to major empirical germplasm origins in China; i.e., Lancaster, Reid, SPT and P. However, Xie et al. (2007, 2008) detected six subpopulations: BSSS, PA, PB, Lancaster, Ludahonggu (LRC), and TSPT, from among 187 commonly used Chinese maize inbred lines, which represented the genetic diversity among public, commercial, and historically important lines for maize breeding. LRC was assigned to the TSPT group in Wang's research (Wang et al. 2008); however, the LRC group was not assigned to one of the four subgroups in the current study. For

Table1. Summary statistics for all lines and each subgroup.

Subgroup	Over all	Lan	P	Reid	TSPT
Sample size	173	24	17	32	29
Alleles	634	407	257	428	426
Allele/locus	8.128	5.218	3.295	5.487	5.462
Gene diversity	0.704	0.611	0.453	0.647	0.624
PIC	0.667	0.566	0.404	0.602	0.582
Subgroup-specific alleles	140	35	17	36	52
sub-group alleles(%)	22.082	8.6	6.615	8.411	12.207
sub-group alleles/line		1.458	1	1.125	1.793

Lan, P, Reid and TSPT represented Lancaster, P, Reid and Tangsipingtou groups respectively.

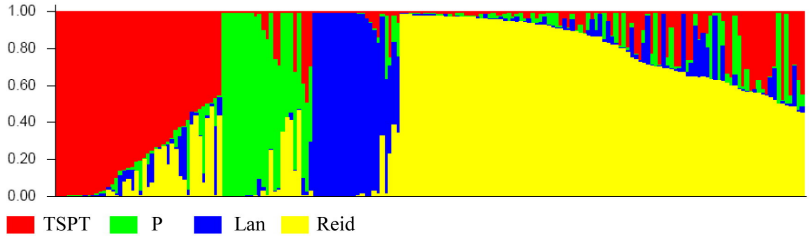

Fig 1. population structure of the 173 inbred lines based on the 78 SSRs. Population structure of the 173 inbred lines based on the 78 SSRs. Bar plot of the genetic composition of 173 individual lines based on 78 linked SSRs generated by STRUCTURE 2.2 using the admixture model. Groups for each inbred line are represented by colors. Each column represents an inbred's genotype and is partitioned into segments, the length of which represents the estimated genetic fraction of every line from each of the four inferred subpopulation.

xample, Dan598 originating from LRC, is classified into the P group, but Dan340 from LRC was assigned into the Lan subgroup. In addition, BSSS and PA were assigned into the Reid group in both Wang's study and ours. To some extent, this might reflect the differences in the target population and the use of different markers could give rise to dissimilar results.

Genetic diversity

Attaining maximum genetic diversity for maize germplasms is critical for mining the resolution of association mapping. A higher genetic diversity means more extensive history of recombination and more alleles. In the present study, the value of SSR allele diversity and PIC found in the 173 maize lines (the value of genetic diversity was 0.704, the PIC value was 0.667) was much lower than other studies involving inbred lines, representing most publicly available lines from the United States, Europe, Canada, South Africa, and Thailand (Liu et al. 2003 genetic diversity value 0.818). However, the genetic diversity is close to that of Stich et al. (2005) (0.68), Matsuoka et al. (2002) (0.62), and Wang et al. (2008). Meanwhile, the genetic diversity value is greater than that of a study involving 187 maize inbred lines (an average of 4.14 alleles per locus, Xie et al., 2008). The level of diversity correlated with the population underlying the study and involved markers (Wang et al., 2008). To some extent, the higher genetic diversity demonstrated that the target population was broader, which can be explained both by temporal and geographical variation trends closely related to agriculture systems and breeding program (Liu et al., 2003). In addition, the level of genetic diversity using more dinucleotide type of lines, as opposed to the many closely related lines in our larger population.

SSRs was higher than that calculated using fewer dinucleotide types (Vigouroux et al., 2002).

Linkage disequilibrium

Our results suggested that approximately 25% of SSR markers embodied significant LD. This result was higher than that of Remington (2001), but was much lower than those of Wang et al. (2008), Stich et al. (2005), and Liu et al. (2003), partly because of their higher density of SSR pairs. Nevertheless, decay of LD is affected by many factors, such as genetic drift, natural and artificial selection, mating system, and admixture of populations (Yu and Buckler, 2006). LD generated by linkage would be attractive and available for association mapping. However, LD generated by population structure and genetic drift would give rise to type I errors in genotype-phenotype associations (Rafalski, 2010). In the present study, the percentage of significant pairwise LD was lower in the subgroups than for all accessions (Table 3). The small size of the population could contribute to LD. Thus, the permutation of random samples of the same size equal to that of the subgroup was chosen. The proportion of observed LD was higher than that of expected LD. This suggests that the population structure does contribute to LD within the subgroups. The actual percentage in the TSPT subgroup is much higher than expected, which implies population stratification or linkage effects among the TSPT population. This result conflicts with that of Wang et al. (2008), who suggested that population structure, relatedness, and genetic drift did not strongly influence the LD of SSR loci in each subgroup. This could be the result of different sampling manner, where they avoided closely related

Table 2. Summary of LD using 85 pairs of SSR loci in 173 inbred lines, on the same chromosome and on the inter chromosome.

	Chromosome Coverage (%)		Numbers of detected LD			
			r^2>0.01	r^2>0.1	D'>0.1	P<0.01
Intra Chromosome	Chr1	92	17	0	29	11
	Chr2	73	2	0	4	3
	Chr3	70	24	0	51	13
	Chr4	60	22	0	41	14
	Chr5	82	15	0	23	11
	Chr6	91	7	0	12	3
	Chr7	64	9	0	20	7
	Chr8	74	6	0	14	2
	Chr9	43	11	1	23	6
	Chr10	69	3	0	16	2
	Total	72	116	1	233	72
Inter Chromosome	Total	72	900	0	896	547
Genome wide	Total	72	1016	1	1129	619

Table 3. Percentage of SSR pairs in LD at a P=0.01 level.

Population	Overall	Reid	TSPT	Lan	P
No of lines	173	32	29	24	17
Observed % in LD	24.74	4.70	10.12	4.23	1.45
Expected % in LD		4.25	3.91	3.28	1.27

Prospects in Marker-trait selection

To date, some studies have used association mapping and identified QTLs or genes corresponding to complex traits in plants (Atwell et al., 2010; Huang et al., 2010; Lu et al., 2010). The costs of the technology of next-generation sequencing, such as Illumina's Genome Analyzer, Applied Biosystems' SOLiD, and Roche's 454, have been decreasing (Myles et al. 2002), such that genome-wide association studies, where sufficient markers are genotyped across the genome such that functional alleles will likely be in LD with at least one of the genotyped markers, will become feasible. In maize, genotyping microarrays have been the first choice for genome-wide association, making it possible to explain the natural phenotypic variation. Recently, the Infinium MaizeSNP50 BeadChip was developed in collaboration with leading maize researchers, and provides uniform genomic coverage and the highest quality content available for identifying desirable traits in corn samples. In the current study, 78 SSR markers were genotyped as background markers to gauge the genetic diversity and population structure of this maize panel. Future research will focus on whole genome-wide association studies of abiotic and biotic tolerance traits, such as waterlogging and resistance to head smut and Corn sheath blight. Successful performance in association mapping would rely on detecting LD between markers and functional alleles corresponding to phenotypic variation. Our results suggest that the degree of LD supports genome wide association mapping. However, our results also indicate that a high risk of spurious association could be generated by population stratification when using association mapping.

Materials and methods

Plant materials

A set of inbred lines was used, comprising 163 maize accessions from China and 10 accessions from the United States. The collection of inbred lines was chosen to represent the genetic diversity and common utilization among current germplasms. The pedigrees and/or origins of all lines used in this study are shown in Supplementary Table S2.

DNA preparation and SSR genotyping

Young leaves from six random 30-day seedlings per accession were bulked together and ground in liquid nitrogen. DNA was manually isolated from leaves using a modified CTAB method (Saghai-Maroof et al., 1984). Seventy-eight SSR primer pairs (Supplementary Table S2), which covered the whole maize genome, were randomly selected to analyze the population structure, genetic diversity, and linkage disequilibrium. Primer sequences motifs of SSRs were obtained from MaizeGDB (http://www.maizegdb.org). PCR reactions for SSR analyses were performed as follows: 35 cycles of 94°C for 30 s, 54–60°C for 30 s (modified when needed), and 72°C for 1 min. Electrophoresis was performed with 6% denaturing polyacrylamide gels. Silver staining of the gels followed the procedures given in Lia et al. (2007).

Population structure and genetic diversity

To investigate genetic structure, Bayesian clustering was performed on multi-locus SSR data using the software package STRUCTURE (Pritchard et al., 2004) in its revised version 2.1 (Falush et al., 2003). The admixture model and independent allele frequency were utilized to explore the dataset with prior population information. Five runs of STRUCTURE were done for each number of populations (K) (set from 1 to 10). Burn-in time and replication number were both set to 100,000 in each run. The maximum likelihood ratio was used to assign the accessions to clusters, and the cut-off probability for assignment to a cluster was 0.70. Allele richness, gene diversity, PIC from both subgroups and subgroup-specific alleles were estimated using PowerMarker V3.25 (Liu and Muse, 2005).

Evaluation of linkage disequilibrium (both genome-wide and intrachromosome)

LD was evaluated for each pair of SSR loci using TASSEL (http://www.maize-genetics.net/bioinformatics/tasselindex.htm) for D′ and r2, which represent the LD measures modified for loci that were used (Hedrick, 1987; Weir, 1996). Significance (P value) of D′ for each SSR pair was determined by 100,000

permutations if $P < 0.01$. For each SSR locus, the rare alleles (i.e., those present in less than 1% of the panels) were combined into one allelic class, as described by Maccaferri (2005).

Acknowledgments

The author is grateful to Prof. Zhenhua Wang from Northeast Agriculture University and Prof. Guoqing Tan from Jilin Academy of Agricultural Sciences for supplying seeds for this study. This research was supported by the 111 Project of China (B07041) and the Beijing Agricultural Innovative Platform-Beijing Natural Science Fund Program (D08070500690802).

References

Agrama HA, Eizenga GC (2008) Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of Oryza sativa and its wild relatives. Euphytica 160: 339–355.

Ardlie K, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3: 299–309.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631.

Austin DF, Lee M, Veldboom LR (2001) Genetic mapping in maize with hybrid progeny across testers and generations: plant the height and flowering. Theor Appl Genet 102: 163–176.

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172: 1165–1177.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, SFlint-Garcia (2009) The genetic architecture of maize flowering time. Science 325: 714-718.

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize, Genetics 118: 519–526.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3: 1–14.

Crosbie TM, Mock JJ, Pearce R (1978) Inheritance of photosynthesis in a diallel among eight maize inbred lines from Iowa Stiff Stalk Synthetic. Euphytica 27: 657–664.

D'hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RG, Eck HJ, Eeuwijk FA (2010) Population structure and linkage disequilibrium unraveled in tetraploid potato Theor Appl Genet 121: 1151-1170.

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116: 113–125.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J 44: 1054–1064.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374.

Fowler JE, Freeling M (1996) Genetic analysis of mutations that alter cell fates in maize leaves: dominant Liguleless mutations. Developmental Genetics 18: 198–222.

Henry A, Damerval C (1997) High rates of polymorphism and recombination at the Opaque-2 locus in cultivated maize. Mol Genet Genomics 256: 147-157.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945-952.

Huang XH, Wei XH, Sang T, Zhao Q, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, Li M, Fan DL, Guo YL, Wang AH, Wang L, Deng LW, Li WJ, Lu YQ, Weng QJ, Liu KY, Huang T, Zhou TY, Jing YF, Li W, Lin Z, Buckler E S, Qian Q, Zhang QF, Li JY, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42: 961-967.

Li Y (1998) Development and germplasm base of maize hybrids in China. Maydica 43:259–269.

Lia VV, Confalonieri VA, Poggio L (2007) B chromosome polymorphism in maize landraces: adaptive vs demographic hypothesis of clinal variation. Genetics 177: 895-904.

Lijavetzky D, Cabezas JA, Ibanez A, Rodriguez V, Martinez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (Vitis vinifera L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8: 424.

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.

Liu KJ, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21: 2128–2129.

Lu YL, Zhang SH, Shah T, Xie CX, Hao ZF, Li XH, Farkhari M, Ribaut JM, Cao MJ, Rong TZ, Xu YB (2010) Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. P Natl Acad Sci USA 107: 19585-19590.

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289.

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in Zea—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 1:436–450.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21: 2194-2202.

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456: 720-723.

Poethig RS (1988) Heterochronic mutations affecting shoot development in maize. Genetics 119: 959-973.

Pritchard JK, Wen W (2004) Documentation for structure software: Version 2. http://pritch-bsduchicagoedu/ software/ readme_structure2_1pdf.

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13: 174-180.

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. P Natl Acad Sci USA: 11479-11484.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW, (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. P Natl Acad Sci USA 81: 8014-8019.

Shi MM (2001) Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clin Chem 47: 164-172.

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.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28: 286-289.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beaviset WD, Stephen J, Smith C, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19: 1251-1260.

Wang RH, Yu YT, Zhao JR, Shi YS, Song YC, Wang TY, Li Y (2008) Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. Theor Appl Genet 117:1141-1153.

Xie CX, Zhang SH, Li MS, Li XH, Hao ZF, Bai L, Zhang DG, Liang YH (2007) Inferring genome ancestry and estimating molecular relatedness among 187 Chinese inbred lines. J Genet Genomics 34:738–748.

Xie CX, Warburton M, Li MS, Li XH, Xiao MJ, Hao ZF, Zhao Q, Zhang SH (2008) An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. Mol Breed 21:407–418.

Xu YB, Skinner DJ, Wu HX, Palacios-Rojas N, Araus JL, Yan JB, Gao SB, Warburton ML, Crouch JH (2009) Advances in maize genomics and their value for enhancing genetic gains from breeding. Int J Plant Genomics 957602.

Yan JB, Warburton ML, Crouch J (2011) Association mapping for enhancing maize (Zea mays L.) genetic improvement. Crop Sci 51: 433-449.

Yan JB, Yang XH, Shah T, Sánchez-Villeda H, Li JS, Warburton M, Zhou Y, Crouch JH, Xu YB, (2009) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breeding 3: 441-451.

Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17: 155-160.

Zhang HL, Zhang DL, Wang MX, Sun JL, Qi YW, Li JJ, Wei XH, Qiu ZG Tang SX, Li ZC (2010) A core collection and mini core collection of Oryza sativa L. in China. Theor Appl Genet 122: 49-61.

Zhang XY, Blair MW, Wang SM (2008) Genetic diversity of Chinese common bean (Phaseolus vulgaris L.) landraces assessed with simple sequence repeat markers. Theor Appl Genet 117: 629-640.

Zhao KY, Aranzana MJ, Kim S, Lister C, Shindo C, Tang CL, Toomajian C, Zheng HG, Dean C, Marjoram P, Ordborg M (2007) An Arabidopsis example of association mapping in structured samples. Plos Genet. 3e4.