INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
14–043/2014/16–5–851–861
http://www.fspublishers.org

Full Length Article

Genetic Diversity and Linkage Disequilibrium Estimation among the Maize Breeding Germplasm for Association Mapping

Yang Liu, Jinjie Guo, Dongmei Zhang, Yongfeng Zhao, Liying Zhu, Yaqun Huang and Jingtang Chen*
Hebei Agricultural University Hebei sub-center for National Maize Improvement Center, Key Laboratory Constructed by Ministry of Education and Hebei Province, Key Laboratory for Crop Germplasm Resources of Hebei, North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Baoding, 007100, P.R. China
*For Correspondence: chenjingtang@126.com

Abstract

Analyzing the genetic basis and linkage disequilibrium (LD) of maize inbred lines is important for maize breeding and marker-trait association. In this study, a total of 201 SSR markers were used to assay the genetic diversity, population structure, and LD of a maize association mapping population consisting of 290 inbred lines, which mainly represented temperate Chinese and US germplasm. Results of genetic diversity analysis showed that this population presented relatively abundant genetic variation and a high level of gene diversity. According to population structure analysis, breeding lines could be clustered into 5 sub-groups, which corresponded well with known pedigree records. Compared with the other 4 sub-groups, the Lvdahonggu (LRC) sub-group showed higher genetic diversity. LD evaluation results showed significant LD levels in pair-wise SSR markers, with up to 40.6% of pairs linked with chromosomes, ranging from 38.1 to 80.1%. The results of the present studies will provide useful information to perform genome-wide association study to improve the efficiency of maize breeding in maize growing areas represented by the panel. © 2014 Friends Science Publishers

Keywords: Maize (*Zea mays* L.); Breeding germplasm; Genetic diversity; Linkage disequilibrium; Simple sequence repeat

Introduction

Maize (*Zea mays* L.) is one of the most agronomically important cereal crops worldwide, in addition to its roles in genetic mechanism of quantitative traits studies as a model plant species (Schnable *et al*., 2009; Haley, 2011). Maize is a cross-pollinated species which exhibits high levels of recombination and a low level of linkage disequilibrium (Vigouroux *et al*., 2002b; Flint-Garcia *et al*., 2005). Consequently, even though relatively strong selection was placed on maize during the course of domestication and breeding, there is a logical expectation that many alleles have been retained in elite germplasm, and that positive alleles have been enriched during artificial selection, allowing contributions to the increasing of grain yield (Varshney *et al*., 2005). Molecular marker-assisted selection may play a crucial role in allele mining to introduce new beneficial alleles into elite germplasm. The essential prerequisite of marker-assisted selection in crop breeding is to explore the locus of useful alleles by feasible strategy. Genetic diversity analyses and association mapping are important in marker-assisted germplasm resources evaluation and are of quite practical significance in molecular breeding (Zhu *et al*., 2008; Van Inghelandt *et al*., 2011).

In the past several decades, pedigree information, combining ability and phenotype data was used to study the genetic diversity of maize landraces, inbred lines and wild relatives (Liu *et al*., 2010; Li *et al*., 2002a). However, these data could describe genetic structure unreliably due to environmental factors. With the development of DNA markers, the evaluation of germplasm at the molecular level has become practical in related research fields (Li *et al*., 2002b; Yang *et al*., 2010; Truntzler *et al*., 2012). Currently, simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) are two main classes of DNA markers used in genetic diversity, population structure and LD level analyses of maize accessions. Because of the high level of polymorphism, reliability, reproducibility, discrimination and economical efficiency (Smith *et al*., 1997), SSR markers have played an important role in assessing genetic diversity and population structure among crop germplasm resources (Yu *et al*., 2007; Liu *et al*., 2012; Nantoumé *et al*., 2013). Recently, with the advance of next-generation sequencing technologies, SNP markers have also been applied in maize genotyping on a large scale (Yan *et al*., 2009; Yang *et al*., 2010). Due to a biallelic character, individual SNPs have lower information content than SSRs. Therefore, many more SNPs are required to obtain as much information as from SSRs. Good experimental results

To cite this paper: Liu, Y., J. Guo, D. Zhang, Y. Zhao, L. Zhu, Y. Huang and J. Chen, 2014. Genetic diversity and linkage disequilibrium estimation among the maize breeding germplasm for association mapping. *Int. J. Agric. Biol.,* 16: 851–861

could also be achieved by using SSRs in the estimation of genetic diversity, population structure and LD level on a small scale without the upfront investment of SNPs (Van Inghelandt *et al*., 2010).

Genome-wide association study (GWAS) is an available strategy to explore the potential use of new alleles for trait improvement (Huang *et al*., 2011; Li *et al*., 2012; Xue *et al*., 2013). The high diversity and rapid LD decay make maize become an ideal species for GWAS (Li *et al*., 2012). However, the complexity of the maize population structure increased the false positive rate in association mapping. For an extreme example, due to the influence of population structure, the non-reproducibility of associations between flowering time greatly influenced by population structure and polymorphisms in the *dwarf 8* gene in three panels was not available (Camus-Kulandaivelu *et al*., 2006). On the other hand, the results of Huang *et al*. (2010) indicated that LD leads to dramatically different resolutions of association study at different genomic regions. Population structure of crop accessions and the extent of LD across the genome play a predominant accessory role in association mapping, and determining the precision and accuracy of an association study. Therefore, the analysis of population structure and assessment of LD are necessary for a successful association analyses.

Previous genetic characterization studies of core collections have allowed the elaboration of population structure and diversity levels in these populations (Yu *et al*., 2007; Wang *et al*., 2008). In populations developed by breeders representing elite lines, genetic diversity and population structure are not always widely known. Therefore, the objectives of this study were to (1) analyze the genetic diversity in an association mapping panel containing Chinese and US inbred breeding lines; (2) assess the population structure, and the level and the distribution of LD between SSR loci of our association mapping population.

Materials and Methods

Plant Materials

A previous study in this laboratory was based on a set of 290 maize inbred lines representing elite China and North American germplasm, including 220 elite Chinese lines (the core collections and their derivatives in China), and 70 elite US lines (36 ex-PVP lines with expired plant variety protection act certificates and 34 public US lines with publically available, non-PVP lines) which used for introducing highly dynamic genetic changes into the maize genome by modern breeding (Jiao *et al*., 2012). These lines represent an extensive collection of the most advanced publically available maize breeding germplasm (Table 1).

SSR Genotyping

Genomic DNA was extracted with a modified CTAB procedure (Saghai-Maroof *et al*., 1984) and diluted to 50 ng/μL. Two hundred and one SSR primer pairs were chosen from the Maize GDB database (http://www.maizegdb.org.), chosen to give even coverage of the ten maize chromosomes. Polymerase chain reaction (PCR) was carried out in a 10 μL reaction volume containing 25 ng genomic DNA, 10 mM dNTP, 1 x *Taq* polymerase buffer, 1 unit of *Taq* polymerase, and 5 pM of SSR primer. The thermal cycling program was 94ºC for 5 min; 95ºC for 1 min; 60ºC for 1 min, and 72ºC for 2 min, which was repeated a total of 34 cycles; followed by a final extension step of 5 min at 72ºC. pBR322/Msp I fragment was employed as molecular weight standard during electrophoresis in polyacrylamide gels for visualization of results.

Statistical Analysis

Genetic Diversity Analysis

The allele frequency, allele numbers per locus, gene diversity, and polymorphism information content (PIC) of the maize accessions were calculated using PowerMarker V3.25 (Liu and Muse, 2005) with the Roger (1972) parameter. PIC value is descried according to (Botstein *et al*., 1980), and was estimated as follows:

$$PIC_i = 1 - \Sigma p_i^2 - \Sigma\Sigma 2 p_i^2 p_j^2$$

Where pi and pj are the frequencies of the i^{th} and j^{th} alleles, respectively.

Average allele numbers, gene diversity and PIC of each subgroup were also estimated. Because sample size has an effect on allele number and gene diversity, a random sampling strategy was utilized to compare the difference between gene diversity and allele number in different sub-populations to create the same sized sample in all sub-populations (Liu *et al*., 2012). Random sampling was repeated 150 times. The analysis of population-differentiation statistic (*Fst*) was calculated using Arlequin version 3.1 (Excoffier and Lischer, 2010).

Population Structure Analysis

The population structure of the association mapping panel was investigated with the STRUCTURE 2.3.4 package (Pritchard *et al*., 2000a, b).

The number of sub-populations (*k*) ranged from 1 to 15, and each *k* value was run 3 times with a burn-in period of 500,000 and 500,000 iterations (Lu *et al*., 2009). The number of subgroups was determined by ΔK (Evanno *et al*., 2005). Lines with probabilities of membership greater than 50% were placed into the related groups, while those with membership probabilities lower than 50% were allotted to a "mixed" group (Liu *et al*., 2012).

Table 1: List of the 290 inbred lines used in this study

ID	Germplasm identity	Pedigree/Origin	ID	Germplasm identity	Pedigree/Origin
1	Mo17	Public_US	69	Lo415	America derived
2	R31	America derived	70	D20	America derived
3	698-3	America Hybrid	71	Lo5-6	unknown
4	qiong51	unknown	72	ys06	unknown
5	757	unknown	73	L473	unknown
6	DH138	America derived	74	qun3	Synthetic
7	R08	American Hybrid P78641	75	qun4	Synthetic
8	78599	American Hybrid 78599	76	18-599	America Hybrid 78599
9	807D	unknown	77	697	America derived
10	802	unknown	78	K10	(Chang3×Shen5003)×Chang3
11	song1145	American Hybrid 78599	79	hai014	unknown
12	E600	Synthetic	80	chang3	Improved from Landrace
13	E200	Synthetic	81	ji444	A619×HuangZaoSi
14	L061F	Limagrain inbred lilne	82	huangyesi3	(YeJiHong×HuangZaoSi) × HuangZaoSi
15	DM101B	unknown	83	zaC546	Variant plants from C103
16	D88	Synthetic	84	ji63	(127-32×Tie84)×(Wei24×Wei20)
17	DM07	America derived	85	luyuan92	Yuanqi123×1137
18	SC24-1	unknown	86	ye8112	American Hybrid 3382
19	D1139	America derived	87	zheng32	American Hybrid 3382
20	W222	HuoBai×L1029	88	C8605-2	Tie7922×Shen5003
21	Y223	American Hybrid	89	7922	American Hybrid 3382
22	C521	unknown	90	ye832	American Single-cross
23	LM-2	Limagrain hybrid	91	chang72	Improved from Landrace
24	B97	BSCB1(R)C9	92	wu109	unknown
25	B100	Public_US	93	lv28	LvDaHongGu
26	B98	Public_US	94	yan812	Improved from U8112
27	B95	Public_US	95	qi205	(VeiAi141×ZhongXi017)×Population 70
28	SC-9	unknown	96	DH854	America derived
29	KP3130	Korea Hybrid	97	DH857	America derived
30	L-1	France inbred line	98	DH856	America derived
31	W172	France inbred line	99	DH864	America derived
32	zong3	Synthetic	100	DH869	America derived
33	W238	unknown	101	DH881	America derived
34	D33A	America derived	102	DH883	America derived
35	G966	America Hybrid	103	DH886	America derived
36	G668	America Hybrid	104	DH1049	America derived
37	G967	America Hybrid	105	DH1051	America derived
38	G969	America Hybrid	106	3H2	(WeiDa202×Zi330)×H84
39	G968	America Hybrid	107	yu87-1	American Hybrid
40	18	unknown	108	nan21-3	Jugoslavian Hybrid
41	1614	unknown	209	B37	Public_US
42	1610	unknown	210	Pa91	Public_US
43	zheng58	Variant plants from Ye478	111	ye52106	(Ye1075×Ye106)×AiJin525
44	changD	America derived	112	ye8001	5003×8112
45	M14	Public_US	113	suwan1611	Suwan2
46	M101	Public_US	114	dan599	American Hybrid 78599
47	2005-4	unknown	115	D892	unknown
48	chen322	unknown	116	1205A	Public_US
49	H601	Synthetic	117	su75	7BU9×511
50	H588	Synthetic	118	R136	America derived
51	changK	unknown	119	200B	Zi330×187-2
52	huangchanga	unknown	120	3489a	America derived
53	953	America derived	121	Z31B	America derived
54	huangchangb	unknown	122	N68a	Public_US
55	w499	America Hybrid	123	619	America derived
56	SC30-1	unknown	124	shen137	American Hybrid 6JK611
57	468-3	unknown	125	3335	America derived
58	XF117	America derived	126	Beck	America derived
59	XF223	America derived	127	98F1	unknown
60	XF134	America derived	128	Maxa	America derived
61	M131-5	Public_US	129	N42	unknown
62	XOP2	Public_US	130	fangyin	unknown
63	R1656	unknown	131	GY3	unknown
64	jian1496b	unknown	132	P167	America Hybrid
65	Max	unknown	133	5032	America derived
66	CN104	America derived	134	zheng22	Dan340×E28
67	La2-4	America derived	135	DF32	unknown
68	T24	unknown	136	FAP1360A	Europe inbred line

Table 1: Continued

137	7903E	America derived	196	zi330	OH43×KeLi67
138	dan340	Lv9×Wide Pod Corn	197	Q1261	Improved from K12
139	chong72	America derived	198	H99	Public_US
140	M101[6]	America Hybrid	199	F7	Ex-PVP
141	dan9046	Shen5003×Tie7922	200	danhuang02	Synthetic of 10 Lv lines
142	P25	America derived	201	xun971	unknown
143	SS99	America derived	202	dan598	Improved from OH43H3
144	BM	America derived	203	7884	America Hybrid Ci7×L289
145	d140	Unknown	204	W64a	Public_US
146	W9706	Public_US	205	A619	Public_US
147	R25	America derived	206	A554	Public_US
148	R150	Unknown	207	Ms71	Public_US
149	R98	America derived	208	B76	Public_US
150	qi319	America derived	209	B37	Public_US
151	20762	Unknown	210	Pa91	Public_US
152	20837	Unknown	211	A679	Public_US
153	20564	Unknown	212	Co109	Public_US
154	XF77	America derived	213	Sg1533	Public_US
155	Los-6	Unknown	214	Va102	Public_US
156	XF27	America derived	215	W182bn	Public_US
157	huotanghuang	HTB42 ×Hai1917×Mo17Ht	253	PHP55	Ex-PVP
158	SC11-1	America derived	254	PHR62	Ex-PVP
159	SC14	America derived	255	PHT22	Ex-PVP
160	811A	S1147×1112	256	PHW20	Ex-PVP
161	806A	1688×HuangC	257	2FACC	Ex-PVP
162	9058	American Hybrid6JK×8085Tai	258	NX928	TangSiPingTou
163	PH6WC	Pineer Hybrid	259	NX926	TangSiPingTou
164	PH4CV	Pineer Hybrid	260	AHU1	xiuqing73-1
165	hai9-21	American Hybrid	261	xun248	TangSiPingTou
166	DH40	unknown	262	AHU24	101314
167	928	Tangsipingtou	263	AHU2	522(white seed)
168	926	Tangsipingtou	264	AHU3	T-Hz4
169	A801	Dan9042×(Dan9046×MoHuang9)	265	AHU4	2105(big38)
170	8982	America derived	266	AHU5	(5237×shan811-1)×5237
171	DF20	unknown	267	AHU6	212×97-1
172	DF27	unknown	268	AHU7	zheng58×92-8
173	DF43	unknown	269	AHU8	73-1×212
174	DF24	unknown	270	AHU9	hengdan11
175	7236	unknown	271	AHU10	sanbei8
176	433-7	unknown	272	AHU11	(P53×ETO)×P53
177	shen977	America derived	273	AHU12	yongyan4×35-1
178	niu2-1	Improved from Landrace	274	AHU13	sanbei8×yong4-1
179	68139	America derived	275	AHU14	liangyu88
180	68202	America derived	276	AHU15	AP13
181	17564	America derived	277	AHU16	xingK36×zheng58
182	huangC	(HuangXiao162×Zi330×O2)×Tuxpenno	278	AHU17	178×zheng58)
183	P007	America Hybrid	279	H21	HuangZaoSi×H84
184	M7	Public_US	280	AHU18	HOO4-2
185	9702	unknown	281	R548	Ex-PVP
186	9710	unknown	282	AHU19	Zh907039
187	y9961	unknown	283	chang7-2	TangSiPingTou
188	7026B	America derived	284	AHU20	heyu2
189	196	unknown	285	huangzaosi	TangSiPingTou
190	LD61	unknown	286	AHU21	nongdan118
191	W344	unknown	287	AHU22	nongxi5678
192	68122	unknown	288	AHU23	yongyan4
193	M22	Public_US	289	B394	Public_US
194	shan89	unknown	290	B73	Public_US
195	ji853	(HuangZaoSi×Zi330)×Zi330			

Neighbor-joining trees and principal component analysis (PCA) were also used to infer population structure of the association mapping panels. Neighbor-joining tree was conducted based on Nei's genetic distance via PowerMarker V3.25 (Liu and Muse, 2005). The rooted neighbor-joining tree was drawn with software MEGA V3.1 (Kumar *et al*., 2004). PCA was carried out by using the NTSYSpc2.1 (Rohlf, 2002).

Linkage Disequilibrium Estimation

The linkage disequilibrium (LD) measurement (D') was calculated between all pairs of SSR markers in the study;

Table2: Summary of genetic diversity among 201 SSRs

SSR locus	Allele	Gene diversity	PIC	SSR locus	Allele	Gene diversity	PIC
umc1177	7	0.7910	0.7657	phi078	6	0.7083	0.6543
umc1354	3	0.2930	0.2521	umc2170	4	0.6168	0.5675
umc2224	6	0.7503	0.7071	umc1063	5	0.7463	0.7039
umc1071	6	0.6856	0.6412	umc2059	5	0.5742	0.5231
umc1397	5	0.5170	0.4072	umc1545	10	0.8168	0.7918
dupssr2	6	0.7647	0.7269	fhi057	7	0.4934	0.4678
phi109275	6	0.7846	0.7500	umc1016	10	0.8695	0.8561
umc1169	4	0.7058	0.6500	bnlg1380	9	0.8228	0.8004
bnlg1627	11	0.8577	0.8427	umc2142	7	0.7498	0.7080
umc1166	9	0.7707	0.7385	umc1865	4	0.6234	0.5496
umc1431	3	0.5643	0.4986	umc1888	3	0.5259	0.4136
umc2237	3	0.5556	0.4856	umc1710	5	0.7303	0.6838
umc1144	7	0.3628	0.3494	umc1782	7	0.7713	0.7413
umc2025	5	0.6765	0.6138	phi328175	6	0.7601	0.7196
phi011	4	0.7441	0.6964	umc1103	6	0.6522	0.6194
umc2227	6	0.7409	0.7070	umc2197	7	0.7205	0.6745
umc1003	6	0.5779	0.5025	umc1760	8	0.8558	0.8385
phi083	5	0.6681	0.6078	umc1786	7	0.7808	0.7454
phi96100	6	0.8024	0.7727	umc1034	16	0.8337	0.8161
umc2184	7	0.6039	0.5363	umc2147	5	0.7587	0.7184
phi046	3	0.5182	0.4043	phi014	5	0.5831	0.5161
umc1386	7	0.6865	0.6381	umc1562	14	0.8878	0.8775
phi03225	3	0.5632	0.4739	umc1960	16	0.7814	0.7548
umc1057	2	0.4869	0.3684	umc1997	11	0.7789	0.7504
umc1062	3	0.6473	0.5745	umc1724	7	0.7014	0.6483
bnlg1957	6	0.6717	0.6138	umc1268	9	0.7561	0.7149
umc1949	5	0.6677	0.6184	umc1933	10	0.7838	0.7519
umc2155	10	0.8195	0.7973	phi233376	11	0.7663	0.7297
umc1164	3	0.5691	0.4737	umc1957	9	0.7611	0.7202
umc1659	4	0.4019	0.3716	umc2393	6	0.5786	0.5241
umc2265	5	0.6465	0.6029	umc2093	5	0.6452	0.5785
bnlg1257	7	0.8331	0.8117	bnlg430	9	0.8023	0.7744
dupssr5	13	0.7469	0.7128	phi027	6	0.5945	0.5317
umc1395	3	0.4151	0.3552	umc1771	7	0.6571	0.5971
umc1504	3	0.6175	0.5475	umc2342	9	0.7024	0.6522
phi213984	7	0.7186	0.6840	umc1078	13	0.8629	0.8495
umc1622	6	0.5927	0.5579	umc1794	4	0.3656	0.3453
phi120	5	0.6020	0.5584	umc1366	6	0.7282	0.6868
umc2243	4	0.5623	0.4893	umc1942	8	0.4141	0.3962
bnlg1025	10	0.7977	0.7723	umc1505	3	0.5287	0.4208
bnlg1064	8	0.7805	0.7511	umc1380	6	0.6327	0.5757
umc1534	7	0.7190	0.6735	umc1293	5	0.7428	0.6986
umc1371	3	0.5588	0.4708	umc2053	3	0.5183	0.4067
umc1180	3	0.5847	0.4988	umc1319	7	0.5816	0.4943
bnlg589	8	0.7399	0.6988	phi059	8	0.7448	0.7053
umc2287	6	0.8108	0.7839	bnlg1716	7	0.6218	0.5898
umc2286	6	0.8118	0.7844	umc2163	10	0.7845	0.7553
umc1776	6	0.3342	0.3225	umc2043	5	0.6948	0.6368
bnlg1556	10	0.8478	0.8292	umc1061	4	0.5920	0.5072
umc2082	10	0.8329	0.8121	umc1993	7	0.6338	0.5710
umc1755	6	0.7165	0.6749	umc2351	3	0.4927	0.4007
umc1464	5	0.4872	0.4290	phi024	5	0.7057	0.6565
phi053	7	0.7606	0.7241	umc2036	2	0.4497	0.3486
umc1259	5	0.7184	0.6658	umc2115	13	0.8541	0.8374
umc1136	8	0.7946	0.7642	umc1935	8	0.7523	0.7232
umc1185	6	0.6664	0.6082	umc2373	9	0.7692	0.7361
umc1012	14	0.7373	0.7143	umc1990	7	0.6423	0.5788
bnlg1175	18	0.8935	0.8849	umc1171	4	0.5920	0.5185
phi072	6	0.7330	0.6971	umc1019	11	0.8340	0.8170
umc2101	5	0.6962	0.6501	bnlg1237	15	0.6053	0.5856
umc2277	8	0.7194	0.6770	phi048	6	0.7382	0.6968
umc1017	10	0.7496	0.7140	umc2136	15	0.8636	0.8496
umc1757	7	0.7660	0.7302	fdx2	11	0.8146	0.7918
umc1824	15	0.8393	0.8197	mmc0241	8	0.6867	0.6372

Table 2: Continued

phi047	3	0.6014	0.5207	umc1687	10	0.6495	0.6245
umc1244	7	0.6947	0.6570	mmc0282	12	0.8735	0.8605
umc1124	4	0.6389	0.5770	umc1248	17	0.7435	0.7219
umc1173	9	0.7461	0.7050	umc1178	12	0.8289	0.8091
dupssr34	13	0.8028	0.7783	umc1853	7	0.6489	0.6033
umc2166	9	0.8169	0.7921	nc010	5	0.6254	0.5493
umc2391	4	0.4371	0.3754	umc2386	11	0.7779	0.7444
umc2365	8	0.6994	0.6691	umc1462	7	0.7467	0.7079
umc1821	4	0.6899	0.6354	bnlg391	10	0.7743	0.7421
phi093	4	0.5561	0.4566	umc2309	9	0.6577	0.5919
bnlg469	6	0.7244	0.6769	umc1018	9	0.7930	0.7621
umc2228	3	0.5427	0.4399	umc1069	10	0.7277	0.6958
umc1590	15	0.8669	0.8541	umc1231	9	0.7212	0.6777
phi076	4	0.6270	0.5493	umc1125	10	0.7515	0.7155
umc2256	5	0.6393	0.5900	umc1489	5	0.6789	0.6122
bnlg1755	15	0.8978	0.8888	umv1492	11	0.7616	0.7261
bnlg1047	9	0.7813	0.7582	umc1536	8	0.7988	0.7694
bnlg1098	12	0.8727	0.8596	umc1429	7	0.6177	0.5546
phi104127	3	0.4854	0.3710	phi299852	12	0.8106	0.7835
umc1896	7	0.7574	0.7298	umc1335	12	0.7759	0.7581
umc1042	12	0.8054	0.7840	umc2084	11	0.7787	0.7455
umc1506	8	0.7486	0.7136	umc1147	9	0.4965	0.4657
umc2007	13	0.8532	0.8390	bnlg240	8	0.7985	0.7763
bnlg1940	12	0.8419	0.8234	bnlg1523	12	0.8392	0.8211
phi065	4	0.6230	0.5605	bnlg2291	13	0.8175	0.7964
umc1999	5	0.6262	0.5592	bnlg1380	9	0.8228	0.8004
phi072	5	0.7002	0.6630	bnlg339	10	0.8335	0.8131
umc1432	6	0.3401	0.3260	umc1760	8	0.8558	0.8385
umc2160	8	0.8039	0.7758	bnlg1237	15	0.6053	0.5856
umc1936	5	0.6010	0.5543	bnlg2305	17	0.8541	0.8380
bnlg439	9	0.8115	0.7895	umc2209	4	0.4556	0.4149
umc1705	15	0.8684	0.8548	bnlg2235	31	0.9163	0.9114
umc1979	4	0.6849	0.6268	bnlg1863	18	0.9218	0.9163
phi389203	2	0.3680	0.3003	phi115	6	0.6365	0.5684
umc1020	7	0.7372	0.6905	bnlg1823	16	0.9038	0.8958
umc1250	4	0.6202	0.5714	umc1033	15	0.8657	0.8567
phi034	3	0.6598	0.5855				

it is commonly used to evaluate the level of LD for each two multi-allelic loci, and was described by Flint-Garcia *et al*. (2003) and was estimated as follows:

$$|D'| = (D_{ab})^2/\min(\pi_A\pi_b, \pi_a\pi_B) \text{ for } D_{ab}<0;$$
$$|D'| = (D_{ab})^2/\min(\pi_A\pi_b, \pi_a\pi_B) \text{ for } D_{ab}>0$$

Where D_{ab} is the difference between the observed and expected haplotype frequencies, π_A, π_B, π_a, and π_b are the frequencies of the A, B, a and b alleles, respectively.

The software package TASSEL 2.1 (Bradbury *et al*., 2007) was used to estimate the statistical significance (*P*-value) and depict the extent of LD. The statistical significance value (*P*) less than 0.01was chosen to represent pair-wise polymorphic loci in the state of LD.

Results

Description of Genetic Diversity

The results of the genetic diversity calculations using 201 SSR markers showed that the breeding germplasm contains a high average PIC value and gene diversity (Table 2). The mean number of alleles per locus was 7.7 for a total of 1,556 alleles from the 201 SSR loci. The number of alleles varied from 2 to 31 per locus over the total population assayed. The gene diversity ranged from 0.2930 to 0.9218, with an average of 0.6978. The mean PIC value was 0.6545 varying from 0.2521 to 0.9163 with peak values appearing between 0.6676 and 0.7230 (Fig. 1). Pair-wise comparisons of the population-differentiation statistic (*Fst*) among subgroups revealed the highest level of differentiation between P group and Tangsipingtou (TSPT) group (*Fst* = 0.28), while the lowest level of differentiation was between Lancaster and LRC (*Fst* = 0.07). Among the 1,556 SSR alleles in the panel, most were present at low frequencies, as approximately 53.53% had the alleles were present at frequencies of less than 0.05 (Fig. 2).

Analysis of Population Structure

The STRUCTURE model-based method indicated that the true *k* value as decided by maximum ΔK calculated from

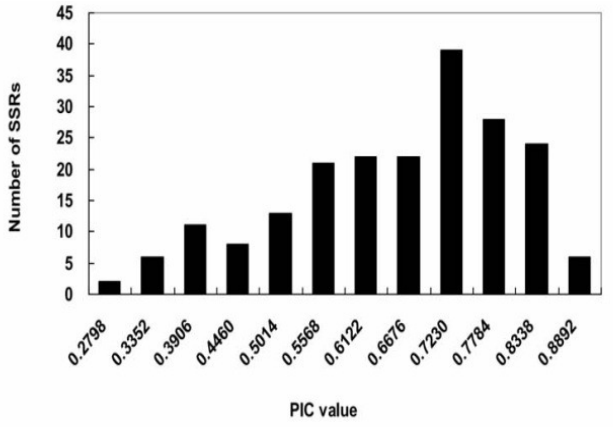

Fig. 1: Histogram of PIC distribution

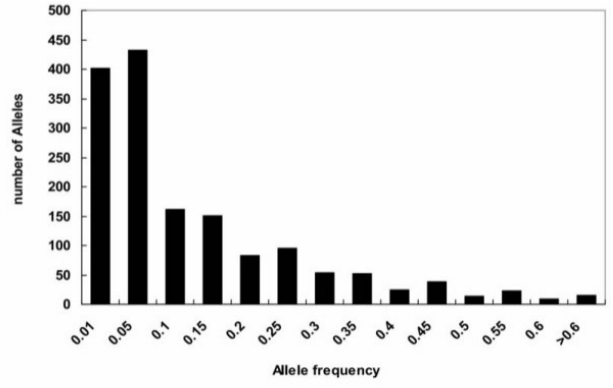

Fig. 2: Histogram of allele frequency distribution

LnP (D) was *k*=5 (Fig. 3a); thus the 290 maize accessions were divided into five subgroups. On the basis of known pedigree information and breeding history of these lines, the five sub-groups were denoted as Reid, Lancaster, Tangsipingtou (TSPT), P group and Lvdahonggu (LRC) (Fig. 3b). Representative lines qi 319, 78599, and Shen137 all belonged to the P group, which consisted of 41 inbred lines; Huangzao4 belonged to the TSPT group, which consisted of 17 inbred lines; lv28 and dan340 belonged to LRC, which consisted of 38 inbred lines; Mo17 and Huotanghuang belonged to the Lancaster group, which consisted of 57 inbred lines; and B73, ye478 and zheng58 belonged to Reid group, which consisted of 64 inbred lines (Table 3). Among the five sub-groups, TSPT and LRC were selected from landraces in China, while P group, Lancaster and Reid all had American pedigree. Additionally, for the reason that any lines with a probability <0.5 could not be clearly divided into any one of the five sub-groups, 73 inbred lines (25.2%) were classified into a "mixed" group.

A neighbor-joining tree based on Nei's genetic distance was constructed for the sake of further insight into the phylogenetic relationships of the 290 maize accessions. The resulting neighbor-joining tree had six divergent groups. Based on the pedigree information, there were three groups

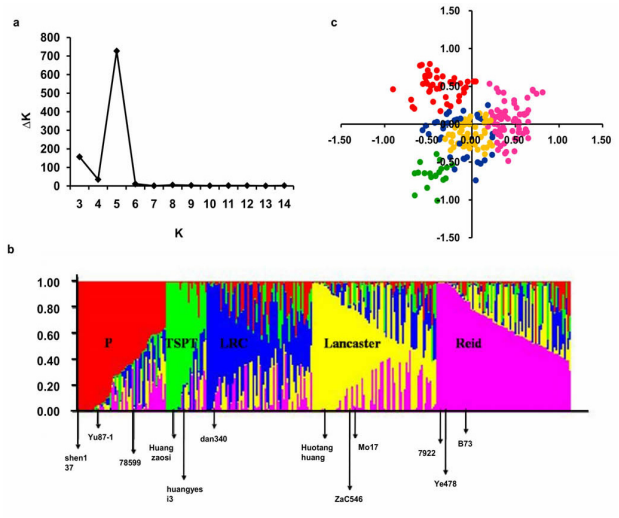

Fig. 3: Population structure of 290 maize inbred lines

a. The true number of population was deduced by delta K value. b. Population structure analysis of 290 inbred lines by STRUCTURE. The inbred lines that the arrows represent which in the figure is the representative lines for each subgroup. c. Principal component analysis of 290 inbred lines. The five subgroups identified from the population structure analysis figure are color-coded in b-c

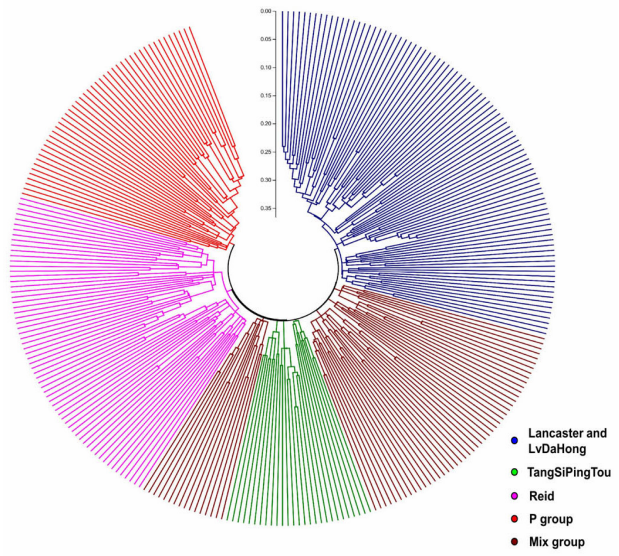

Fig. 4: Dendrogram of 290 maize inbred lines based on the Nei's genetic distance

corresponding to P, Reid and TSPT clusters, which exhibited a consistence with the population structure analysis. However, the Lancaster and LRC sub-groups were not separated in the phylogenetic tree; this agrees with the low *Fst* differentiation value seen between these sub-populations, and indicates that they are not well separated on a genetic level. In addition, 63 inbred lines were not clustered into each one of the four groups, were assigned to two additional "mixed" groups (Fig. 4).

Table 3: Summary of population structure analysis among 290 inbred lines

Group	Sample size	Percent (%)	Representative lines
P group	41	14.1	qi319,78599,yu87-1,shen137
TSPT	17	5.9	Huangzao4,Huangyesi3,xun926,xun928
LRC	38	13.1	lv28,dan340,danhuang02
Lancaster	57	19.7	Mo17,Huotanghuang,zaC546
Reid	64	22.1	zheng58,7922,ye478,B73,B37,zheng32
Mixed group	73	25.2	

Table 4: Summary statistics of average number of allele and gene diversity for model-based groups

Group	Sample size	Average Allele No. (mean ±SD)	Average gene diversity (mean ±SD)
LRC	38	5.58(4.56±0.11)[a]a[b]	0.654(0.631±0.008) a
Lancaster	57	5.67(4.42±0.09) b	0.647(0.622±0.009) b
Ried	64	5.36(3.97±0.09) c	0.596(0.568±0.014) c
P group	41	4.94(4.03±0.30) d	0.565(0.566±0.039) c
TSPT	17	3.59e	0.464d

a Figures in the brackets refers to the results based on random sampling with fixed sample size of 17
b Refers to the significance difference at 0.05 level.

Table5: Percentage of linked pair-wise SSR loci in significant LD and mean D' value

chr.	Percentage of LD locus pairs(%)	mean of D'
1	72.8	0.26
2	38.1	0.24
3	40.7	0.23
4	52.7	0.27
5	75.9	0.28
6	71.6	0.24
7	78.4	0.25
8	80.1	0.32
9	57.8	0.24
10	60.1	0.24
overall	40.6	0.22

Principal component analysis (PCA) indicates that the first and second principle coordinate (PC) contribution rates were 14.3% and 13.7%, respectively. These results are very similar to the structure analysis and clustering analysis. PCA of the entire set of 290 tested lines exhibited a relatively clear separation into five groups. Reid, P group, TSPT and Lancaster groups were densely distributed on the plot, while the inbred lines of LRC group were more dispersed on the plot of these components, and viewed from this angle (Fig. 3c). The three groups (PB, Reid and TSPT) were well divided into three distinct groups. However, this PCA did not clearly separate the Lancaster and LRC groups, in agreement with the cluster analysis and low *Fst* value between them. A large number of inbred lines in the LRC group were scattered around Lancaster group and even inside the Lancaster group.

Evaluation of Genetic Diversity within Subgroups

Further analysis to assess the genetic diversity within each sub-group revealed that the Lancaster and LRC sub-groups presented higher genetic diversity, with an average number of alleles per locus of 5.67 and 5.58 and the average gene diversity of 0.647 and 0.654, respectively (Table 4). The next were the Reid and P groups, which had average allele numbers of 5.36 and 4.94 per locus, and average gene diversity of 0.596 and 0.565, respectively. The TSPT subgroup was the least diverse with an average allele number per locus of 3.59 and gene diversity of 0.464.Random sampling of 150 times demonstrated a similar tendency of genetic diversity parameters for each sub-population (Table 4).

Analysis of Linkage Disequilibrium

Linkage disequilibrium (LD) was investigated at a whole genome level. The results showed that, no matter pair-wise SSRs were on the same chromosome or not, there existed a certain extent of LD (Table 5). In the 290 inbred lines, the percentage of loci pairs in significant LD was 40.6%, and mean D' value was 0.22. The proportion of pair-wise loci in significant LD and the mean D' value for each pair of loci on the same chromosomes is presented in Table 5.

Discussion

The present study uncovered an abundant level of allelic variation of SSR markers within an association mapping

population. Compared with the studies performed by Yu *et al*. (2007) and Wu *et al*. (2010), we obtained higher levels of gene diversity. However, these results were much lower than that reported by Liu *et al*. (2003) using di-nucleotide repeat (and more variable) SSR markers, and Liu *et al*. (2012), who characterized a larger and more diverse panel. Many of the inbred lines used in the present study were breeding lines, in which genetic diversity may have been reduced during the course of ongoing artificial selection (Yamasaki *et al*., 2005).

Genetic diversity within each subgroup was not even distributed. The random sampling results showed that the LRC was the subgroup with the highest diversity, which had a good agreement with the previous studies (Wang *et al*., 2008; Liu *et al*., 2012). Most early cluster studies also showed that the LRC derived lines were assigned to different subgroups (Teng *et al*., 2004; Yu *et al*., 2007). This implies that the LRC inbred lines contains complex germplasm resources, leading to the highest genetic diversity. On the contrary, the TSPT sub-group was the least diverse sub-group. The most productive inbred line of the TSPT sub-group is "Huangzaosi", with high general combining ability and excellent agronomic characters. Because most inbred lines in the TSPT sub-group are derived from Huangzaosi, the genetic diversity is lower than the other sub-groups.

Understanding the genetic structure of maize is of great importance for parent selection in breeding and reducing false association in association mapping. In the present study, population structure analysis and PCA of genotypes lead to similar results as past studies (Lu *et al*., 2009; Liu *et al*., 2012) and agreed with pedigree records. The 290 maize inbred lines could be divided into 5 subgroups including Reid, Lancaster, TSPT, P group, and LRC. A traditional classification of maize inbred lines includes Lancaster and Reid (from the U.S.A), TSPT and LRC (in China). The P group was developed from the Pioneer hybrid P78599 and related hybrids and has a short breeding history. According to the agronomic contribution and molecular characterization results, Wang *et al*. (2004) classified P78599 and related lines into an independent pool, named as "P group" or the Temperate-Tropical group (Teng *et al*., 2004). Nevertheless, the result of the present clustering analysis based on Nei's genetic distance placed Lancaster and LRC into the same group. Pair-wise comparisons of the population-differentiation statistic (*Fst*) among subgroups also showed that the population differentiation between Lancaster and LRC was the least (*Fst*=0.07) compared with that between any two other subgroups. Some past studies also place Lancaster and LRC into the same group (Wang *et al*. 2008; Yu *et al*., 2007). Although pedigree information places Lancaster and LRC into two independent sub-groups, this is not supported by marker studies. The analysis of population structure based on a model method and cluster analysis based on genetic distance were not in complete accordance for assigning inbred lines to clusters. These results suggested that the model clustering was superior to genetic distance clustering and it not only can clearly discriminate the genetic relationships among groups, but also can dissect the genetic basis among inbred lines.

Seventy three of the lines had a membership < 0.50 for any of the five clusters, and were assigned to a mixed ancestry. Of thes，23 lines belonged to American derivatives, 21 lines belonged to China derivatives, 11 lines belonged to US public breeding lines, 2 lines belonged to Ex-PVP lines, and 16 lines had unknown pedigree information, most of which were new germplasm. Developed from crosses between the different clusters in breeding practice, these improved new elite lines were not likely to be divided into the currently known breeding groups. Through Q-matrices output by STRUCTURE, we could see the complicated genetic composition of these improved resources, and make use of them in a breeding program, which is one advantage of STRUCTURE compared with the other kinds of cluster analysis software. The results confirm that molecular markers allow a better classification of genotypes than pedigree information do, especially those inbred lines with unknown pedigree information (Van Inghelandt *et al*., 2010).

Genome-wide scans with SSR loci revealed that 40.6% of the SSR marker pairs presented significant LD. The LD in the investigated association mapping population was higher than previously reported by Remington *et al*. (2001), which was most likely attributable to the higher SSR marker density used in the present study. However, it is lower than the LD reported value by Wang *et al*. (2008), who studied a smaller and less diverse population than used in our study. Diversity of germplasm and sample size are main two factors influencing LD level. Previous studies indicated that the more abundant the diversity of germplasm, the lower the level of LD, and the greater the number of sample, the lower the level of LD (Liu *et al*. 2003; Yan *et al*. 2009).

Linkage disequilibrium (LD) generating and influence factors including recombination, mutation, genetic drift, relatedness, selection, population structure and linkage (Flint-Garcia *et al*., 2003; Gupta *et al*., 2005). LD which generated by linkage is beneficial to the genome-wide association mapping, On the contrary, LD which generated by population structure and genetic drift will result in spurious genotype-phenotype association (Wang *et al*., 2008). In our analysis, we estimated the percentage of SSR marker pairs in significant LD and the mean value of D' for each chromosome. There was a high proportion of pair-wise SSR marker in significant LD on the same chromosomes (Table 5), therefore, linkage was deemed to be an important force which generated LD in the association mapping population. The results suggested that the population could be used in further genome-wide association study.

In this study, we screened 290 maize inbred lines including core collections (Li *et al*., 2005) as well as derivatives of breeding lines from public programs in China and the U.S.A, and parents and derivatives of important commercial hybrids. These 290 lines possess a great depth of allelic diversity. This diversity may be valuable for germplasm conservation and pre-breeding practice. Therefore, the germplasm resources with abundant diversity are urgently needed. Similarly, many rare alleles were captured in the association mapping population. Finally, we have defined the population structure and the extent of LD in this panel, which indicate that it may be useful for genome-wide association study to seek alleles having application value in maize production.

Acknowledgement

The authors thank Prof. Jingrui Dai and Jinsheng Lai for plant materials support. We are also grateful to Dr. Ruihui Wan for suggestions. This study was supported by the National High Technology Research and Development Program of China (863 Program No. 2011AA10A103-3), Research Fund for New Teachers of the Doctoral Program of Higher Education of China (20131302120001) and Modern Agricultural Technology System Maize Innovation Team of Hebei Province.

References

Botstein, D., R.L. White, M. Skolnick and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.*, 32: 314

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler, 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23: 2633–2635

Camus-Kulandaivelu, L., J.B. Veyrieras, D. Madur, V. Combes, M. Fourmann, S. Barraud, P. Dubreuil, B. Gouesnard, D. Manicacci and A. Charcosset, 2006. Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. *Genetics,* 172: 2449–2463

Evanno, G., S. Regnaut and J. Goudet, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol*., 14: 2611–2620

Excoffier, L. and H.E. Lischer, 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour*., 10: 564–567

Flint-Garcia, S.A., A.C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman and E.S. Buckler, 2005. Maize association population: a high–resolution platform for quantitative trait locus dissection. *Plant J.,* 44: 1054–1064

Flint-Garcia, S.A., J.M. Thornsberry and B. IV, 2003. Structure of Linkage Disequilibrium in Plants. *Annu. Rev. Plant Biol.,* 54: 357–374

Gupta, P.K., S. Rustgi and P.L. Kulwal, 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol*., 57: 461–485

Haley, C., 2011. A cornucopia of maize genes. *Nat. Genet*., 43: 87–88

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu and Z. Zhang, 2010. Genome–wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.,* 42: 961–967

Huang, X., Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, W. Li, Y. Guo, L. Deng and C. Zhu, 2011. Genome–wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet*., 44: 32–39

Jiao, Y., H. Zhao, L. Ren, W. Song, B. Zeng, J. Guo, B. Wang, Z. Liu, J. Chen and W. Li, 2012. Genome–wide genetic changes during modern breeding of maize. *Nat. Genet.,* 44: 812–815

Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*., 5: 150–163

Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, J. Wang, Y. Han, Y. Chai, T. Guo and N. Yang, 2012. Genome–wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet*., 45: 43–52

Li, M., S. Zhang and X. Li, 2002a. Study on heterotic groups among maize inbred lines based on SCA. *Sic. Agric. Sin.,* 35: 600–605

Li, Y., J. Du, T. Wang, Y. Shi, Y. Song and J. Jia, 2002b. Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. *Maydica*, 47: 93–102

Li, Y., Y. Shi, Y. Cao and T. Wang, 2005. Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. *Genet. Resour. Crop Evol.,* 51: 845–852

Liu, K. and S.V. Muse, 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21: 2128–2129

Liu, K., M. Goodman, S. Muse, J.S. Smith, E. Buckler and J. Doebley, 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics*, 165: 2117–2128

Liu, Z.Z., X. Wu, H.I. Liu, Y.X. Li, Q.C. Li, F.G. Wang, Y.S. Shi, Y.C. Song, W.B. Song, J.R. Zhao, J.S. Lai, Y. Li and T.Y. Wang, 2012. Genetic diversity and population structure of important Chinese maize inbred lines revealed by 40 core simple sequence repeats (SSRs). *Sci. Agric. Sin.*, 45: 2107–2138

Liu, Z.Z., Y.C. Song, Y.S. Shi, Y.L. Cai, W.D. Cheng, L.Q. Qin, Y. Li and T.Y. Wang, 2010. Racial classification and characterization of maize land races in China. *Sci. Agric. Sin.,* 43: 899–910

Lu, Y., J. Yan, C.T. Guimaraes, S. Taba, Z. Hao, S. Gao, S. Chen, J. Li, S. Zhang and B.S. Vivek, 2009. Molecular characterization of global maize breeding germplasm based on genome–wide single nucleotide polymorphisms. *Theor. Appl. Genet.,* 120: 93–115

Nantoumé, A., S. Andersen and B. Jensen, Genetic differentiation of watermelon landrace types in Mali revealed by microsatellite (SSR) markers. *Genet. Resour. Crop Evol*., 60: 2129–2141

Pritchard, J.K., M. Stephens and P. Donnelly, 2000a. Inference of population structure using multilocus genotype data. *Genetics,* 155: 945–959

Pritchard, J.K., M. Stephens, N.A. Rosenberg and P. Donnelly, 2000b. Association mapping in structured populations. *Amer. J. Hum. Genet.,* 67: 170–181

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman and E.S. Buckler, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci.,* 98: 11479–11484

Roger, J., 1972. *Measures of Genetic Similarity and Genetic Distance Studies in Genetics*. VII Univ Taxas, 7213: 145–153

Rohlf, F., 2002. *NTSYSpc*: *Numerical Taxonomy System, ver. 2.1*. Exeter Publishing, Ltd [Links], Setauket, New York, USA

Saghai-Maroof, M., K. Soliman, R.A. Jorgensen and R. Allard, 1984. Ribosomal DNA spacer–length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci*., 81: 8014–8018

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton and T.A. Graves, 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326: 1112–1115

Smith, J., E. Chin, H. Shu, O. Smith, S. Wall, M. Senior, S. Mitchell, S. Kresovich and J. Ziegle, 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theor. Appl. Genet.,* 95: 163–173

Teng W.T., Q.S. Cao, Y.H. Chen, X.H. Liu, X.Q. Jing, F.J. Zhang and J.S. Li, 2004. Analysis of maize heterotic groups and patterns during past decade in China. *Sic. Agric. Sin.,* 37: 1804–1811

Van Inghelandt, D., A.E. Melchinger, C. Lebreton and B. Stich, 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.,* 120: 1289–1299

Van Inghelandt, D., J.C. Reif, B.S. Dhillon, P. Flament and A.E. Melchinger, 2011. Extent and genome–wide distribution of linkage disequilibrium in commercial maize germplasm. *Theor. Appl. Genet*., 123: 11–20

Varshney, R.K., A. Graner and M.E. Sorrells, 2005. Genomics–assisted breeding for crop improvement. *Trends Plant Sci.,* 10: 621–630

Vigouroux, Y., J.S. Jaqueth, Y. Matsuoka, O.S. Smith, W.D. Beavis, J.S.C. Smith and J. Doebley, 2002a. Rate and pattern of mutation at microsatellite loci in maize. *Mol. Biol. Evol*., 19: 1251–1260

Vigouroux, Y., M. McMullen, C. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka and J. Doebley, 2002b. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci.*, 99: 9650–9655

Wang, R., Y. Yu, J. Zhao, Y. Shi, Y. Song, T. Wang and Y. Li, 2008. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor. Appl. Genet*., 117: 1141–1153

Wang, Y.D., M.X. Duan and J.F. Xing, 2004. Current Situation and Prospect of Using P Maize Group to Improve the Heteros is Utility and to Development New Germplasm. *J. Maize Sci.*, 12: 10–12,15

Wu, C.L., Q.Q. Zhing, B.X. Dong, S.F. Li and C.Q. Zhang, 2010. Analysis of genetic structure and relationships of maize inbred lines in China. *Acta. Agron. Sin.*, 36: 1820–1831

Xue, Y., M.L. Warburton, M. Sawkins, X. Zhang, T. Setter, Y. Xu, P. Grudloyma, J. Gethi, J.-M. Ribaut and W. Li, 2013. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor. Appl. Genet*., 126: 2587–2596

Yamasaki, M., M.I. Tenaillon, I.V. Bi, S.G. Schroeder, H. Sanchez–Villeda, J.F. Doebley, B.S. Gaut and M.D. McMullen, 2005. A large–scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell Online*, 17: 2859–2872

Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen and J. Crouch, 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*, 4: e8451

Yang, X., J. Yan, T. Shah, M.L. Warburton, Q. Li, L. Li, Y. Gao, Y. Chai, Z. Fu and Y. Zhou, 2010. Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor. Appl. Genet*., 121: 417–431

Yu, Y., R. Wang, Y. Shi, Y. Song, T. Wang and Y. Li, 2007. Genetic diversity and structure of the core collection for maize inbred lines in China. *Maydica*, 52: 181–194

Zhu, C., M. Gore, E.S. Buckler and J. Yu, 2008. Status and prospects of association mapping in plants. *Plant Genome*, 1: 5–20

(Received 17 January 2014; Accepted 12 March 2014)