Theor Appl Genet (2008) 117:1141–1153
DOI 10.1007/s00122-008-0852-x

ORIGINAL PAPER

Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China

Ronghuan Wang · Yongtao Yu · Jiuran Zhao · Yunsu Shi · Yanchun Song · Tianyu Wang · Yu Li

Received: 2 February 2008 / Accepted: 16 July 2008 / Published online: 12 August 2008

Abstract Understanding genetic diversity, population structure, and the level and distribution of linkage disequilibrium (LD) in target populations are of great importance and the prerequisite for association mapping. In the present study, 145 genome-wide SSR markers were used to assess the genetic diversity, population structure, and LD of a set of 95 maize inbred lines which represented the Chinese maize inbred lines. Results showed that the population included a diverse genetic variation. A model-based population structure analysis subdivided the inbred lines into four subgroups that correspond to the four major empirical germplasm origins in China, i.e., Lancaster, Reid, Tangsipingtou and P. Among all of the inbred lines, 65.3% were assigned into the corresponding subgroups; others were assigned into a "mixed" subgroup. LD was significant at a 0.01 level between 63.89% of the SSR pairs in the entire sample and with a range of 18.75–40.28% in the subgroups. Among factors influencing LD, linkage was the major cause for LD of SSR loci. The results suggested that the population may be used in the detection of genome-wide SSR marker–phenotype association.

Introduction

Maize is one of the most important crops in the world, as well as in China. Identification of genetic regions that are responsible for agronomically important traits is of fundamental significance for maize improvement. In the past several decades, linkage mapping has been extensively utilized in genetic dissection of simple or complex traits in maize. Recently, association mapping, which has several advantages over traditional linkage analysis (Kraakman et al. 2004; Flint-Garcia et al. 2005), has been proved to be an effective approach to connecting structural genomics and phenomics (Thornsberry et al. 2001).

Association mapping has been used extensively in human genetics (Corder et al. 1994; Templeton 1995). It was first introduced into plant genetics in 2001 (Thornsberry et al. 2001) mainly due to little information on the population structure and linkage disequilibrium (LD) pattern in plants (Flint-Garcia et al. 2003). Since many important crops have a long and complex domestication and breeding history, together with the limited gene flow in most wild plants, many crop populations exist as complex population structures (Sharbel et al. 2000; Flint-

---

R. Wang and Y. Yu contributed equally to the work.

Communicated by J. Yu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0852-x) contains supplementary material, which is available to authorized users.

R. Wang · Y. Yu · Y. Shi · Y. Song · T. Wang (✉) · Y. Li (✉)
Institute of Crop Science, Chinese Academy of Agricultural Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), 100081 Beijing, People's Republic of China
e-mail: wangtianyu@263.net; yuli@mail.caas.net.cn

R. Wang
e-mail: ronghuanwang@126.com

J. Zhao
Maize Research Center, Beijing Academy of Agricultural and Forestry Sciences, 100097 Beijing, People's Republic of China
e-mail: jiuran@263.net

Garcia et al. 2003). When performing association analysis based on these populations without considering the effects of population structure, spurious association between genotype and phenotype variation may be detected because of the unequal allele frequency distribution between subgroups (Knowler et al. 1988). This has also been verified in maize (Andersen et al. 2005; Camus-Kulandaivelu et al. 2006; Wilson et al. 2004). Recently, with the development of statistics, independent markers that are distributed through whole genome are successfully used to detect population structures (Pritchard et al. 2000a, b). The resolution of association mapping depends on the extent and distribution of LD across the genome within a given population (Remington et al. 2001). LD is generally dependent on the history of the population, but other factors such as population structure, selection, mutation, relatedness, and genetic drift would also cause LD. However, among all of these factors, LD caused by linkage is the most significant importance for association mapping (Stich et al. 2005).

In our previous study, a mini core set of maize inbred lines (94 accessions) was defined to represent the genetic diversity of Chinese maize inbred lines (Li et al. 2004; Yu et al. 2007). Together with B73, a total of 95 inbred lines were used as the association mapping population for further research. However, little information of genetic diversity, population structure, and LD is known for the association mapping panel until now.

The objectives of our research were to (1) assess the genetic diversity of our association mapping population; (2) investigate the population structure among the inbred lines; (3) detect the extent and distribution of LD between SSR loci pairs.

Materials and methods

Plant materials

In our previous study, 288 maize inbred lines including 242 inbred lines of the core collection were established in China (Li et al. 2004) and some elite lines used in recent years in breeding programs were genotyped for genetic diversity at 49 simple sequence repeat (SSR) loci. These markers which were publicly available (http://www.maizegdb.org) covered the maize genome. With the help of SSR fingerprinting, a mini core set of 94 inbred lines representing 87% of the SSR allelic diversity of the 288 inbred lines was defined (Yu et al. 2007). They are genetically diverse but mainly of Chinese origin. These 94 inbred lines together with B73 constituted the association mapping population used in the present study. The 95 inbred lines and their pedigrees or sources are listed in Table 1.

Table 1 List of the 95 inbred lines used in this study

ID	Inbred line	Pedigree/source
1	Ji63	(127-32 × Tie84) × (Wei24 × Wei20)
2	Aijin525	Wuzhiai × Golden Queen
3	C103	Noah Hershey
4	Zi330	Oh43 × Keli67
5	Tangsipingtou	Landrace Tangsipingtou
6	Huobai	Landrace Huojiabaimaya
7	Mo17	C103 × 187-2
8	Lu28	Landrace Ludahonggu
9	Weifeng322	W59E × Fengke
10	Hua160	Hybrid “Huadong 2”
11	Huangzaosi	Landrace Tangsipingtou
12	E28	(Lu9kuan × A619$^{\mathrm{Ht}}$) × Lu9kuan
13	Ye8112	A foreign hybrid “3382”
14	5003	American hybrid “3147”
15	Ye478	U8112 × 5003
16	Dan340	Baigulu9 × *Z. mays* - tunicata
17	Huangyesi3	(Huangzaosi × Yejihong) × Huangzaosi
18	Huotanghuang17	Huotangbai42 × Hai1917 × Mo17$^{\mathrm{Ht}}$
19	Zong3	Selected from a synthetic
20	Zheng58	Selected from inbred Ye478
21	Luyuan92	Yuanqi123 × 1137
22	Qi319	Selected from Pioneer hybrid “78599”
23	K12	Huangzaosi × Huaichun
24	Ziyu3	Selected from hybrid “Luziyu 3”
25	Jingnuo2	Unknown
26	Jitian15	Selected from a sweet corn hybrid
27	D729	Selected from synthetic “D Tuan”
28	Tai184	Selected from population “Hunxuan 1”
29	XZ19	Jin0-14 × Xin335
30	13A/O2	Unknown
31	Fu96	Nanchong5 × Ai13-31
32	Liao7794	7922 × 8112/(32 × 5003)
33	Ji846	Mo17 × Ji63
34	Ji880	Selected from inbred ZheB77A
35	Jiu03	Unknown
36	Fu842	Dian11 × Yi210/Fu746
37	Lu65	Beijin14 × 330
38	Yan103	Selected from inbred Yuanwu02
39	Qi318	Selected from Pioneer hybrid “78599”
40	48-2	Selected from a synthetic
41	FR218	Introduced from America
42	DaMO	Unknown
43	Yi67	Unknown
44	91huang5	Dian11A × Lu9
45	Liao5110	7922 × 5003
46	Han102	Landrace Henanduosui
47	De811	[B68 × [B73$^{\mathrm{Ht}}$ × (C103 × Mp3204)Sel.]]
48	A632	(Mt42 × B14) × B14(3)

Table 1 continued

ID	Inbred line	Pedigree/source
49	Lo1125	Pioneer ADA (PR3374)
50	K36	Selected from “Ku6”
51	Zun90110	Selected from Pioneer hybrid “78599”
52	P138	Selected from Pioneer hybrid “78599”
53	H21	Huangzaosi × H84
54	488	8112 × 5003
55	X178	Selected from North American single-cross hybrid “P”
56	Chang7-2	Huangzaosi × Wei95
57	DH65232	DH6327 × 5003
58	Shen137	Selected from Pioneer hybrid “6JK111”
59	H2	Selected from an American hybrid
60	Jinsui54	Zi330
61	GB	Selected from landrace Tianjinbaiyumi
62	X.L9010-3/O2	5105 × opaque2
63	Longkang15	RC103 × Chang3
64	Wanxi23	Va35 × B73
65	Guan17-1	Guan73 × Mo17
66	Yue267-1-1	5003 × Kanghandalihuang
67	Yue89E4-2	Wu102 × Huangzaosi
68	Yue20-3	Selected from hybrid “Yedan 13”
69	Wu202	Wu403 × Bup29
70	CML67	Introduced from Mexico
71	Liao2204	Selected from an American hybrid
72	HuangC	(Huangxiao162 × Zi330/O2) × Tuxpeno
73	CN165	Selected from an American hybrid
74	Shen135	Selected from Pioneer hybrid “78599”
75	Yuanwu02	Wu105 × Duo229
76	444	A619 × Huangzaosi
77	Zao49	Tuxpeno × Zao22
78	Chaoxianbai	Selected from landrace Chaoyangbaiyumi
79	H205	Selected from a foreign hybrid
80	87-20	Unknown
81	Yan172	Selected from landrace Puchengaiyumi
82	Zhonger/O2	Selected from hybrid “Zhongdan 2”
83	Feng273	Selected from hybrid “Chunza 1”
84	P39/su	Introduced from America
85	92huang7	W153/C103
86	785	330 × Huangzaosi
87	92huang40	853 × Mo17
88	Chihuang32	Lu9 × Ci7
89	Xing83	144 × 147
90	Longkang1	Liao1311(Dian11 × Dafeng22)
91	Lo1067	Pioneer 3780 × Lo87602
92	HR962	Selected from inbred Huangzaosi
93	55113-3-3-5	ZPDC551B
94	Cheng18	Dingshangyumi × (Gong70 × 60–22)
95	B73	Lowa Stiff Stalk Synthetic (BS13C5)

SSR genotyping

Genomic DNA was extracted from the leaves of 1-month-old maize seedlings according to the CTAB procedure (Saghai-Maroof et al. 1984). A total of 145 SSRs loci, randomly distributed across the genome, were used to genotype the mini core set of inbred lines. Among them, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide SSRs accounted for 30.34, 37.24, 24.13, 5.52 and 2.76%, respectively. Most of SSR repeat motifs and sequences were obtained from MaizeGDB (http://www.maizegdb.org).

TP-M13 method was performed in our analysis, in which three primers, i.e., a sequence-specific forward primer, a sequence-specific reverse primer with M13 tail (5′-CACGACGTTGTAAAACGAC-3′) at the 5′ end, and a universal fluorescent-labeled M13 primer, were used (Schuelke 2000). The fluorescent dyes used in the analysis included FAM, VIC, PET and NED. PCR products were size separated on an ABI Prism 3700 DNA Sequencer (Perkin Elmer Biotechnologies, Foster City, CA, USA) and were classified into specific alleles by GeneTyper 2.1 software (Perkin Elmer/Applied Biosystems).

Genetic diversity analysis

PowerMarker V3.25 (Liu and Muse 2005) was used to calculate the summary statistics including allele number, allele frequency, gene diversity (or expected heterozygosity) and PIC. In addition, line-specific alleles and rare alleles (with frequency <5%) were also calculated. Gene diversity (D) was defined as the probability of two randomly chosen alleles from the population and calculated at each locus as.

$$\hat{D}_l = (1 - \sum_{u=1}^{k} \tilde{p}_{lu}^2)/(1 + \frac{1+f}{n})$$

Where p_{lu} is the frequency of the uth allele, f is the inbreeding coefficient, n is the sample size. Polymorphism information content (PIC) (Botstein et al. 1980) was estimated as.

$$\widehat{PIC}_l = 1 - \sum_{u=1}^{k} \tilde{p}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2\tilde{p}_{lu}^2 \tilde{p}_{lv}^2$$

Where p_{lu} is the frequency of the uth allele, and p_{lv} is the frequency of the vth allele.

In addition, the relationship of the total number of alleles with the sample size was also investigated. Resampling was repeated 1,000 times and the results were averaged for each size of sample.

Population structure analysis

To evaluate the population structure of the association mapping population, software package STRUCTURE 2.1 (Pritchard et al. 2000a, b) was employed to subdivide inbred lines into genetic subgroups. The number of subgroups (*K*) was set from 2 to 10. For each *K*, three runs were performed separately. And the burn-in length and iterations were all set to 500,000. Lines with membership probabilities ≥ 0.8 were assigned to the corresponding subgroups and lines with membership probabilities <0.8 were assigned to a "mixed" subgroup.

Allele number, gene diversity, PIC of each subgroup and subgroup-specific alleles were calculated. A re-sampling strategy was also used to obtain genetic diversity of each subgroup. The same number of samples equal to the number of samples of the smallest subgroup was selected randomly from the larger subgroups. The procedure was repeated 1,000 times and the results were averaged.

Linkage disequilibrium estimation

A permutation version of Fish's exact test in PowerMarker V3.25 (Liu and Muse 2005) was used to calculate the extent of LD (r^2) between SSR pairs at $P = 0.01$ in the entire population and each subgroups. As sample size affects the statistic power of LD test, we used a re-sampling strategy to obtain comparable estimate. That is, random samples in each subgroup with the same size were drawn from the entire samples, and the expected percentage of SSR pairs in significant LD was calculated. The procedure was repeated 50 times and the resulting estimates were averaged.

In addition, the ratio of the percentage of linked to unlinked SSR pairs in significant LD was also calculated. SSR loci which were located on the same chromosome were defined as linked loci, and SSR loci located on different chromosomes were defined as unlinked loci.

Results

Genetic diversity

A total of 145 SSR loci, randomly distributed across the genome, were used to evaluate the genetic diversity of the mini core set of inbred lines. All of the 145 SSR loci were polymorphic across the 95 inbred lines and a total of 1,365 alleles were detected (Table 2 and Supplementary material). The average number of alleles per locus was 9.4, ranging from 2 to 38. The average genetic diversity was 0.6831 with a range of 0.2921–0.9489. In addition, the average PIC value was 0.6439 with a range of 0.2555–0.9465.

Of the 145 SSR loci, 44 were dinucleotide SSRs and the others were longer-repeat SSRs. The results showed that the allele number, the gene diversity, and the PIC were not equal among different types of SSR loci (Table 3). Dinucleotide SSRs had more alleles, higher gene diversity, and higher PIC than other longer-repeat SSR loci.

Among the 1,365 alleles, 320 private alleles (23.44%) were found only in one of the 95 inbred lines. Frequencies of most alleles were low, and rare alleles with frequency of less than 5% accounted for 55.75% (Fig. 1). In order to clarify the relationship of the allele number with the sample size, a re-sampling strategy was used to select different number of samples (2–95) from the 95 inbred lines for 1,000 times. Then, alleles numbers of 1,000 times for a given sample size were averaged. The results showed that 65 random samples could capture 90% of the total alleles in the entire sample (Fig. 2).

Population structure

In order to understand the genetic structure of the association mapping population, a model-based approach in the STRUCTURE software was used to subdivide each inbred line to the corresponding subgroup. STRUCTURE software was run for the number of fixed subgroups *K* from 2 to 10, and three runs were performed for each *K*. As the STRUCTURE software overestimates the number of subgroups for inbred lines (Pritchard and Wen 2004), and it is difficult to choose the "correct" *K* from the Ln probability of data, Ln $P(D)$ (Fig. 3). Thus, the results were compared with the known pedigree of the inbred lines for each run of different *K*. The results showed that when $K = 4$, the model-based subgroups were largely consistent with known pedigrees of the inbred lines. The four subgroups corresponded to the four major germplasm origins in China, i.e., Lancaster, Reid, Tangsipingtou (TSPT) and P (Fig. 4).

The Lancaster subgroup was the largest subgroup and included 30 inbred lines. The lines closely related to the Mo17 pedigree and the Zi330 pedigree. In addition, some lines derived from a landrace "Ludahonggu" were also designed into this subgroup. The next subgroup TSPT had 13 inbred lines, which were mainly derived from Huangzaosi, one of founder parents in maize breeding of China. The Reid subgroup included 12 inbred lines. The P subgroup was smallest with only 7 inbred lines, in which Shen137 was derived from a Pioneer hybrid "6JK111" and the other 6 lines were all derived from "78599". Additionally, 33 inbred lines that had <0.8 membership in each of the four subgroups and had a mixture of two or more subgroups were assigned to a mixed subgroup (Table 4).

Table 2 Summary statistics for the 145 SSR loci used in the present study

SSR locus	Allele	Gene diversity	PIC	SSR locus	Allele	Gene diversity	PIC
phi056	7	0.7703	0.7391	phi308090	9	0.6305	0.5916
umc1169	9	0.7083	0.6671	bnlg1189	11	0.7612	0.7363
phi011	8	0.7394	0.6937	bnlg1784	10	0.8196	0.7978
umc1185	13	0.7362	0.6978	umc1101	14	0.7817	0.7566
umc2105	12	0.6907	0.6754	umc1940	6	0.6762	0.6199
phi073	5	0.6335	0.5631	umc1109	6	0.4574	0.4335
umc1266	8	0.3661	0.3563	nc130	6	0.6656	0.6061
umc1008	14	0.7840	0.7565	umc2115	6	0.8063	0.7781
phi072	7	0.6872	0.6557	phi109188	7	0.6304	0.5793
nc004	12	0.6223	0.5897	umc1447	7	0.6767	0.6195
phi079	4	0.3167	0.3002	umc1705	26	0.9187	0.9134
umc1574	9	0.7511	0.7127	phi331888	5	0.5844	0.5390
phi076	3	0.6587	0.5845	umc1332	10	0.7941	0.7638
umc1050	7	0.7296	0.6820	umc1860	7	0.7486	0.7033
phi008	3	0.4669	0.3765	phi333597	5	0.6309	0.5770
phi085	7	0.6560	0.6023	phi087	7	0.6900	0.6422
phi075	5	0.6290	0.5611	bnlg2305	14	0.5643	0.5413
phi031	4	0.5534	0.4975	phi048	11	0.8070	0.7849
phi078	4	0.5215	0.4071	umc2143	7	0.6892	0.6424
umc1545	11	0.8734	0.8605	umc1153	5	0.6587	0.5997
phi112	14	0.7692	0.7416	bnlg1043	21	0.9231	0.9179
phi114	10	0.7251	0.6948	bnlg161	27	0.9201	0.9148
phi034	7	0.6912	0.6470	umc1143	7	0.668	0.6185
phi115	6	0.5783	0.4959	bnlg249	18	0.7796	0.7612
umc1741	15	0.8291	0.8121	umc1517	4	0.4866	0.4548
phi014	5	0.5720	0.4959	phi389203	14	0.7218	0.7044
umc1202	11	0.8444	0.8280	bnlg1702	12	0.7520	0.7269
phi015	6	0.7208	0.6754	bnlg2249	15	0.7412	0.7208
phi080	7	0.7454	0.7163	umc2170	10	0.6688	0.6203
phi033	8	0.4925	0.4666	phi123	6	0.6144	0.5491
phi065	4	0.6619	0.5972	phi299852	11	0.7599	0.7210
phi032	7	0.6865	0.6330	umc2160	21	0.8528	0.8420
umc2345	5	0.5085	0.4496	umc1301	4	0.4627	0.3701
umc1179	2	0.4317	0.3385	umc1408	7	0.8172	0.7932
umc2016	5	0.5409	0.4658	phi328175	5	0.7253	0.6657
phi062	5	0.4957	0.4059	phi069	3	0.6394	0.5674
phi084	6	0.6359	0.5667	phi116	6	0.6885	0.6373
bnlg1112	27	0.9201	0.9156	phi420701	6	0.7059	0.6710
bnlg1179	13	0.8704	0.8574	umc1359	12	0.7257	0.7016
bnlg1007	11	0.8120	0.7946	umc1139	6	0.6417	0.5861
bnlg439	19	0.8436	0.8282	umc1304	4	0.5297	0.4202
umc1144	6	0.2921	0.2816	phi100175	5	0.6160	0.5602
umc1297	17	0.8923	0.8829	phi121	9	0.5570	0.5417
umc1122	8	0.6791	0.6390	umc1121	6	0.6494	0.5933
umc1147	7	0.3274	0.3179	umc2212	6	0.4724	0.4378
phi423298	8	0.4818	0.4497	umc1161	6	0.6845	0.6449
umc2047	4	0.6780	0.6098	phi233376	5	0.7132	0.6625
phi308707	9	0.7713	0.7385	bnlg1272	20	0.8808	0.8733

Table 2 continued

SSR locus	Allele	Gene diversity	PIC	SSR locus	Allele	Gene diversity	PIC
phi064	9	0.8093	0.7878	umc1279	5	0.4556	0.3994
phi96100	8	0.7839	0.7528	bnlg1520	12	0.7882	0.7615
bnlg125	38	0.9402	0.9375	umc2084	10	0.8426	0.8232
umc1542	11	0.7278	0.6916	umc1170	12	0.8386	0.8201
umc1845	22	0.8555	0.8474	umc1634	7	0.6975	0.6646
umc2007	13	0.7870	0.7620	umc1492	7	0.6014	0.5754
nc131	9	0.8189	0.7972	phi108411	8	0.5327	0.5161
nc133	3	0.2963	0.2555	umc1231	13	0.6351	0.5922
bnlg1138	12	0.8317	0.8109	phi448880	6	0.5144	0.4853
bnlg1831	13	0.8070	0.7862	bnlg1129	10	0.8657	0.8509
bnlg1940	16	0.8619	0.8495	umc1675	6	0.7122	0.6648
phi090	5	0.4575	0.4111	umc1277	4	0.5393	0.4382
phi127	6	0.7676	0.7319	phi041	10	0.7646	0.7297
phi427434	6	0.7464	0.7056	phi059	5	0.5215	0.4411
phi101049	9	0.7643	0.7288	phi96342	3	0.3815	0.3487
phi453121	9	0.6966	0.6509	umc1432	11	0.4868	0.4667
phi374118	10	0.7778	0.7487	phi063	7	0.6985	0.6485
phi053	8	0.7123	0.6665	phi050	6	0.5789	0.5194
phi102228	8	0.6687	0.6301	umc1367	8	0.5948	0.5575
umc1489	10	0.6645	0.6118	umc2163	19	0.8854	0.8759
phi046	3	0.6156	0.5340	umc1506	8	0.7433	0.7046
bnlg1754	35	0.9489	0.9465	umc1196	4	0.7110	0.6566
umc1136	9	0.7947	0.7665	bnlg1450	27	0.9338	0.9300
phi213984	6	0.5650	0.5146	phi071	3	0.4689	0.3883
bnlg490	17	0.7834	0.7573				

Table 3 Summary statistics for different types of SSR loci

SSR repeat motif	No. of SSR	No. of allele	Gene diversity	PIC
Dinucleotide	44	15.5227	0.7755	0.7551
Trinucleotide	54	6.8148	0.6333	0.5861
Tetranucleotide	35	7.0286	0.6661	0.6170
Pentanucleotide	8	5.3750	0.5758	0.5349
Hexanucleotide	4	6.2500	0.7016	0.6533
Average		9.4138	0.6831	0.6439

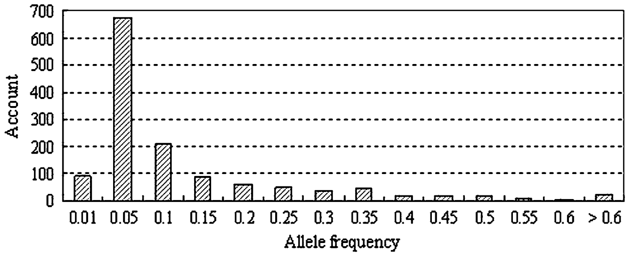

Fig. 1 Distribution of allele frequencies for the 1,365 alleles detected in the study

Genetic diversity of subgroups

The genetic diversity for each subgroup was assessed (Table 5). The Lancaster subgroup was the most diverse subgroup, with a total of 924 alleles, 6.37 alleles per locus, and gene diversity of 0.65. The next was the TSPT subgroup, which had 649 alleles totally, 4.48 alleles per locus, and gene diversity of 0.58. The Reid subgroup was less diverse than the TSPT subgroup. In addition, among all of the alleles, 35.09% were subgroup-specific. The Lancaster subgroup had more subgroup-specific alleles (256 or 27.71%) and 8.53% were line-specific. This also indicated that the Lancaster subgroup included higher genetic variation.

To understand the effect of subgroup size, a re-sampling strategy was also performed to evaluate the subgroup diversity. Since the P subgroup only had 7 inbreds, we randomly selected 7 inbreds from other three subgroups to calculate the total number of alleles and the gene diversity. The resulting data from 1,000 repetitions were averaged to assess the genetic diversity for each subgroup. The results showed that the trend of genetic diversity for each subgroup remained the same although the same number of

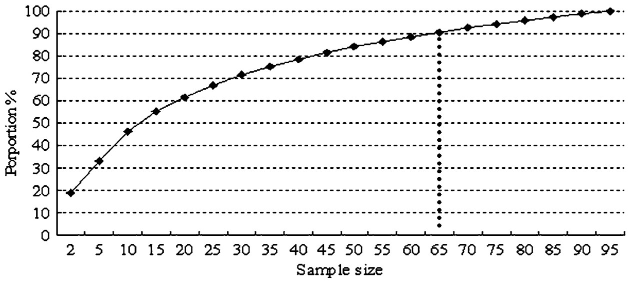

Fig. 2 Plot of the expected number of alleles in samples of different sizes

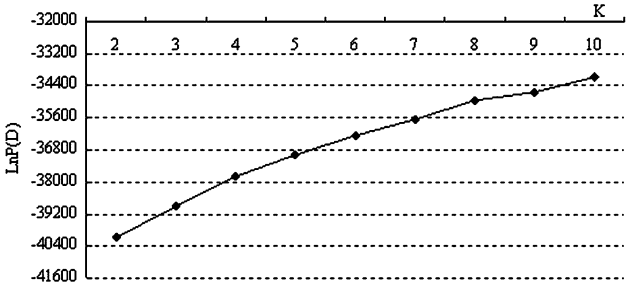

Fig. 3 Plot of the Ln probability of data, Ln $P(D)$, averaged over the replicates

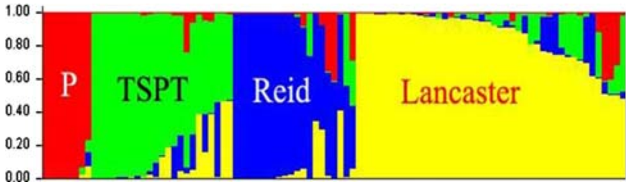

Fig. 4 Population structure of 95 inbreds based on 145 SSRs

inbreds was randomly selected from other subgroups except the P subgroup (Figs. 5, 6).

Linkage disequilibrium

Linkage disequilibrium (LD) among SSRs was investigated in the entire set of inbred lines and in each of the subgroups. In the 95 inbred lines, LD was significant at a 0.01 level between 63.89% of the SSR pairs, but the proportion within each of the subgroups was less (Table 6). Furthermore, the percentage of SSR pairs in LD in the Lancaster subgroup was much higher than the other subgroups, and the P subgroup was the lowest. Because the statistical ability of LD depends on the sample size, a re-sampling strategy was adopted to calculate the expected proportion of pair-wise SSR loci in significant LD. The results showed that when we selected random samples of the same size as in each subgroup from the entire set of inbred lines, the observed and expected proportion of significant pair-wise LD was almost equal. This indicated that sample size substantially contributed to the higher percentage of pair-wise SSR loci in LD in the entire sample than in each subgroup, but population structure and relatedness did not remarkably affect the LD in the subgroups.

In order to investigate the relationship of linkage and LD, we estimated the percentage of linked SSR loci pairs in significant LD in the entire set of inbred lines and in each subgroup. In the 95 inbred lines, 83.33% linked pair-wise SSR loci were in significant LD at the 0.01 level on average. For each model-based subgroup, most of the linked SSR loci were in significant LD though the percentage varied among chromosomes (Table 7). Overall, linkage was the main factor resulting in the pair-wise SSR loci with significant LD in the entire sample and each subgroup.

Discussion

Genetic diversity of the mini core inbred lines

Choice of germplasm is one of the key factors determining the resolution of association mapping. In order to detect more alleles, germplasm selected should include all the genetic variation of a specific species theoretically because diverse germplasm include more extensive recombination in the history and allow a high level of resolution. The species for which a core collection has been established, the core would be the idea material for association mapping (Whitt and Buckler 2003). We have constructed the core collection for maize germplasm preserved in Chinese National Genebank which included 951 landraces and 242 inbred lines (Li et al. 2004). Later, these 242 inbred lines and 46 elite lines used in recent years in Chinese breeding programs were genotyped for genetic diversity at 49 SSR loci, and a mini core set of 94 inbred lines representing 87% of the SSR allelic diversity of the 288 inbred lines was defined (Yu et al. 2007). During the definition, some lines of great agronomical importance were included and the mini core panel selected represented the maximum number of alleles of the 288 inbred lines. These 94 inbred lines together with B73 constituted the association mapping population for further analysis.

In the present study, 145 SSR loci, randomly distributed across the genome, were used to detect the genetic diversity of the population. A total of 1,365 alleles with an average of 9.4 alleles per locus were detected in the entire population, and the average gene diversity and PIC was 0.6831 and 0.6439, respectively. The genetic diversity was much higher than that of Xie et al. (2007, 2008) (PIC was

Table 4 Membership of inbred lines corresponding to each subgroup

Subgroup	Line	Membership of inbred lines corresponding to each subgroup			
		Lancaster	TSPT	P	Reid
Lancaster	91huang5	0.994	0.002	0.002	0.002
	Zi330	0.993	0.003	0.002	0.001
	Tai184	0.993	0.001	0.002	0.004
	92huang40	0.992	0.003	0.003	0.002
	Lu65	0.990	0.005	0.002	0.002
	Jinsui54	0.989	0.007	0.003	0.002
	Fu96	0.988	0.007	0.002	0.003
	48-2	0.988	0.004	0.003	0.006
	Lu28	0.987	0.003	0.007	0.003
	Zong3	0.986	0.004	0.003	0.006
	Xing83	0.981	0.003	0.004	0.012
	Guan17-1	0.978	0.004	0.012	0.007
	Cheng18	0.977	0.003	0.001	0.019
	D729	0.969	0.026	0.003	0.002
	Ji63	0.965	0.003	0.002	0.031
	Mo17	0.965	0.005	0.027	0.003
	Ji880	0.965	0.024	0.004	0.006
	Dan340	0.959	0.004	0.004	0.033
	Huotanghuang17	0.958	0.031	0.004	0.007
	Weifeng322	0.952	0.031	0.006	0.011
	Ji846	0.945	0.008	0.004	0.043
	Luyuan92	0.944	0.034	0.009	0.012
	Feng273	0.924	0.072	0.001	0.003
	Yue20-3	0.912	0.078	0.006	0.005
	XZ19	0.911	0.037	0.007	0.044
	Aijin525	0.910	0.019	0.056	0.014
	Fu842	0.886	0.109	0.002	0.002
	Zhonger/O2	0.873	0.013	0.005	0.109
	Longkang1	0.810	0.184	0.004	0.001
	Chihuang32	0.807	0.165	0.006	0.021
TSPT	Yue89E4-2	0.001	0.997	0.001	0.001
	HR962	0.001	0.997	0.001	0.001
	Huangyesi3	0.002	0.996	0.001	0.001
	Yuanwu02	0.001	0.996	0.002	0.001
	Huangzaosi	0.005	0.990	0.003	0.002
	Yan103	0.004	0.987	0.004	0.005
	Zao49	0.011	0.984	0.003	0.002
	Yan172	0.007	0.980	0.002	0.011
	444	0.010	0.968	0.018	0.004
	GB	0.022	0.942	0.011	0.025
	Chang7-2	0.009	0.902	0.002	0.087
	Huobai	0.128	0.856	0.003	0.012
	Tangsipingtou	0.189	0.802	0.006	0.003
Reid	Ye8112	0.001	0.001	0.001	0.997
	Ye478	0.001	0.001	0.001	0.997
	Liao7794	0.001	0.001	0.001	0.997

Table 4 continued

Subgroup	Line	Membership of inbred lines corresponding to each subgroup			
		Lancaster	TSPT	P	Reid
	Liao5110	0.001	0.001	0.001	0.996
	De811	0.002	0.001	0.002	0.995
	5003	0.001	0.002	0.003	0.993
	488	0.001	0.001	0.006	0.991
	B73	0.008	0.001	0.003	0.987
	FR218	0.015	0.003	0.001	0.980
	A632	0.024	0.003	0.003	0.970
	Longkang15	0.048	0.013	0.011	0.928
	Lo1125	0.063	0.003	0.084	0.850
P	Zun90110	0.001	0.001	0.998	0.001
	Shen137	0.001	0.001	0.998	0.001
	Qi318	0.001	0.001	0.997	0.001
	P138	0.001	0.001	0.997	0.001
	X178	0.001	0.001	0.997	0.001
	Shen135	0.002	0.001	0.994	0.002
	Qi319	0.025	0.039	0.934	0.002
Mixed	P39/su	0.768	0.020	0.003	0.210
	Ziyu3	0.763	0.006	0.021	0.211
	13A/O2	0.745	0.003	0.004	0.249
	Han102	0.745	0.202	0.008	0.044
	Jitian15	0.733	0.251	0.003	0.012
	DaMO	0.731	0.261	0.006	0.002
	92huang7	0.709	0.268	0.017	0.006
	E28	0.699	0.019	0.003	0.279
	Lo1067	0.617	0.002	0.002	0.379
	55113-3-3-5	0.605	0.271	0.115	0.009
	Yi67	0.513	0.007	0.409	0.071
	CN165	0.507	0.085	0.406	0.002
	87-20	0.503	0.170	0.319	0.008
	Hua160	0.484	0.463	0.015	0.038
	H21	0.002	0.785	0.016	0.196
	Wu202	0.047	0.722	0.018	0.214
	CML67	0.029	0.690	0.242	0.039
	H205	0.057	0.670	0.064	0.209
	Jingnuo2	0.386	0.602	0.008	0.003
	Chaoxianbai	0.164	0.561	0.047	0.229
	Jiu03	0.388	0.551	0.054	0.006
	Yue267-1-1	0.011	0.539	0.002	0.449
	C103	0.465	0.529	0.004	0.002
	K12	0.470	0.519	0.008	0.003
	H2	0.073	0.075	0.765	0.087
	HuangC	0.004	0.257	0.004	0.734
	785	0.346	0.003	0.001	0.649
	Zheng58	0.292	0.001	0.076	0.630
	Liao2204	0.019	0.003	0.356	0.619

Table 4 continued

Subgroup	Line	Membership of inbred lines corresponding to each subgroup			
		Lancaster	TSPT	P	Reid
	K36	0.003	0.002	0.413	0.582
	Wan23	0.413	0.004	0.007	0.576
	DH65232	0.010	0.439	0.003	0.548
	X.L9010-3/O2	0.062	0.276	0.288	0.375

Table 5 Summary statistics for each subgroup

	Overall	Subgroup			
		Lancaster	Reid	TSPT	P
Sample size	95	30	12	13	7
Alleles	1365	924	578	649	413
Alleles/locus	9.41	6.37	3.99	4.48	2.85
Gene diversity	0.68	0.65	0.56	0.58	0.46
PIC	0.64	0.60	0.51	0.54	0.41
Subgroup-specific alleles	479	256	87	90	46
Subgroup-specific alleles (%)	35.09	27.71	15.05	15.57	7.96
Subgroup-specific alleles/line (%)	7.73	8.53	7.25	6.92	6.57

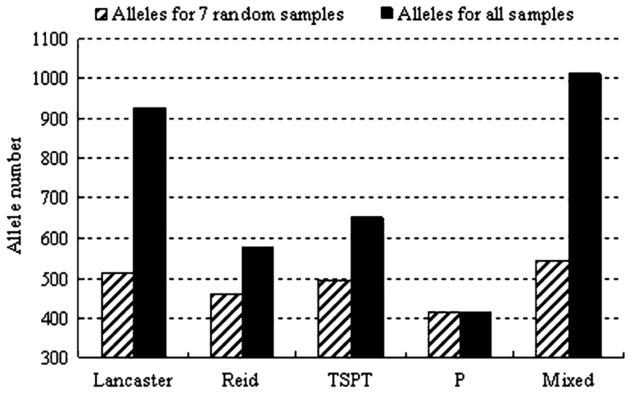

Fig. 5 Comparison of number of alleles for all samples and 7 random samples in each subgroup

0.615), and close to that of Stich et al. (2005) (genetic diversity was 0.68) and Matsuoka et al. (2002) (gene diversity was 0.62), but much lower than that of Liu et al. (2003) (gene diversity was 0.82). The main reason for the difference was the germplasm under study and the SSRs used. The higher genetic diversity detected by Liu et al. (2003) was mainly due to the broad range of germplasm and more dinucleotide type of SSRs. Mutation rate of dinucleotide type of SSR was much higher than that of other types (Vigouroux et al. 2002), which was confirmed in the present study.

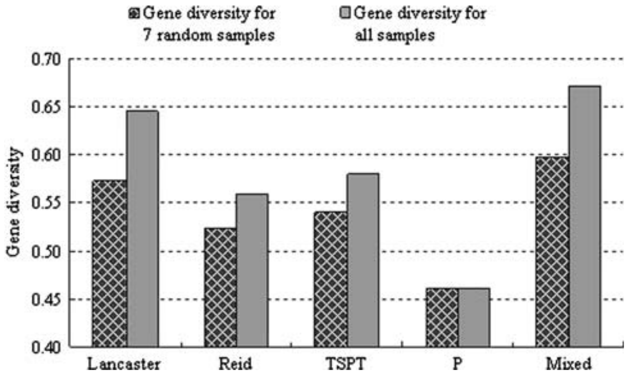

Fig. 6 Comparison of gene diversity for all samples and 7 random samples in each subgroup

Table 6 Percentage of SSR pairs in LD at a $P < 0.01$ level

Population	Inbred no.	Observed % in LD	Expected % in LD
Overall	95	63.89	
Lancaster	30	40.28	45.01
TSPT	13	34.72	33.89
Reid	12	29.86	33.36
P	7	18.75	21.95

Population structure of the mini core inbred lines

Chinese maize inbred lines often have complex genetic background; therefore, understanding population structure and relationships among inbred lines is of significant importance for maize improvement and association analysis. In the present study, 145 SSRs that covered the entire maize genome were selected to analyze the population structure of the 95 inbreds. We selected ≥ 0.8 membership as the subgroup subdivision criterion and the analysis showed that when $K = 4$, the model-based subgroups were consistent with known pedigrees of the inbred lines, and the subgroups were consistent with the four major empirical germplasm origins, i.e., Lancaster, Reid, TSPT and P subgroup. Among all of the inbred lines, 65.3% were assigned into the corresponding subgroups. Lancaster,

Table 7 Percentage of linked SSR loci pairs in significant LD in overall set and different subgroups

Chromosome	Overall	Subgroup			
		Lancaster	TSPT	Reid	P
Chr.1	90.00	88.89	100	100	80
Chr.2	81.82	90.00	87.50	100	50
Chr.3	75.00	75.00	75.00	100	50
Chr.4	100	100	100	100	100
Chr.5	88.89	100	85.71	100	100
Chr.6	77.78	100	87.50	71.43	100
Chr.7	83.33	100	80.00	66.67	100
Chr.8	81.82	100	100	87.50	66.67
Chr.9	84.62	85.71	100	33.34	100
Chr.10	77.78	66.67	100	100	100

TSPT, Reid and P subgroup accounted for 31.6, 13.7, 12.6 and 7.4% of the entire population, respectively.

The results in this study showed that the derivatives of Zi330, Mo17 and Ludahonggu had high genetic similarity, and thus were classified into the Lancaster subgroup. The Reid germplasm were introduced from the USA during the period from the 1950s to the 1970s. Chinese breeders developed a lot of inbred lines from these germplasm. For example, Ye8112 was selected from the maize hybrid "8112", 5003 was selected from the maize hybrid "3147", and a series of inbred lines such as Ye478, 488, Liao7794 and Liao5110 were derived from 5003. These lines were subdivided into the Reid subgroup in our analysis. Since the late 1980s, some Pioneer hybrids have been introduced into China; therefore, a new group defined as "P" was generated (Wang et al. 2004). "78599", one of the most important hybrids among them, was used widely in selecting inbred lines. Up to now, more than 100 hybrids have been released by using "78599"-derived inbreds. In our population, a few lines were selected from "78599".

Genetic structure of Chinese maize inbred lines was documented in a few previous studies. The consistent opinion was that Chinese maize inbred lines could be classified into 4–6 subgroups, most corresponding to the heterotic groups established according to the pedigree information and combining ability (Peng et al. 1998; Wang et al. 1998, 1999; Yuan et al. 2001; Xie et al. 2007, 2008). Recently, Xie et al. (2007, 2008) analyzed 187 commonly used Chinese maize inbred lines, representing the genetic diversity among public, commercial and historically important lines for maize breeding, and detected six subpopulations, that is, BSSS, PA, PB, Lancaster, Ludahonggu (LRC) and TSPT. But when only three clusters were allowed, the clusters were associated with geographic origins, i.e., A (PA, BSSS, Lancaster), B (PB) and D (LRC, TSPT). Interestingly, Ludahonggu is a landrace originally grown in Luda, Liaoning Province of Northeast China, and was probably introduced from the USA in the 1920s (Li et al. 2002). Previously, the derivatives of Ludahonggu were regarded as a dependent group called "Ludahonggu group" (Li et al. 2002; Peng et al. 1998; Xie et al. 2007, 2008). Other reports, on the other hand, suggested that Dan340, a typical Ludahonggu-derived inbred line, could be classified into Zi330 group (Sun et al. 1999; Yuan et al. 2001; Li et al. 2003; Teng et al. 2004; Zheng et al. 2006). Our results also supported this classification. In addition, the integration of PA and BSSS identified by Xie et al. (2008) was also accepted by breeders and researchers in China, since they contained the Reid germplasm origin. Although some discrepancies of the results existed among different researches, which probably resulted from the difference of materials and SSRs (type and number) used, the general profile of genetic structure of Chinese maize inbred lines was largely consistent.

In addition, the remaining 33 inbred lines which had a membership <0.8 with none of the four subgroups were classified into a mixed subgroup, and accounted for 34.7% of the total inbred lines. Among the 33 inbreds, 9 lines were selected from foreign hybrids, 3 lines from Chinese hybrids, and 2 lines from Chinese landraces. This also indicated that the mini core set of inbred lines came from wide origins and contained extensive genetic variation. On the other hand, population structure analysis could help us understand the genetic composition of lines, especially for those with unknown pedigree information, such as 87-20, Yi67, DaMo, CML67 and H205. Unexpectedly, an important inbred line from the US, C103 had 52.9% similarity with the TSPT germplasm and 46.5% similarity with the Lancaster germplasm. This needs to be investigated further although it does not suggest that C103 originated from TSPT, a Chinese landrace.

Linkage disequilibrium and the forces causing LD

In the present study, 63.89% of the SSR pairs exhibited significant LD; however, in each model-based subgroup the percentage of SSR pairs in LD was much lower with a range of 18.72–40.28%. The result was considerably higher than that of Remington et al. (2001), possibly due to the higher density of SSRs used in our study. However, it was lower than that of Stich et al. (2005) and comparable to the results reported by Liu et al. (2003), which was in accordance with the previous studies that LD level detected by SNPs or SSRs would be higher in narrow germplasm than in diverse germplasm (Ching et al. 2002; Liu et al. 2003).

LD observed in a population is the result of interplay of many factors including linkage, population structure, relatedness, selection, mutation and genetic drift (Huttley et al. 1999; Flint-Garcia et al. 2003; Rafalski and Morgante 2004; Gupta et al. 2005). Forces generating and conserving LD in a population were paid more attention to in recent years, and have been demonstrated by experimental data (Stich et al. 2005, 2006) and computer simulations (Stich et al. 2007). LD generated by linkage is considerably useful for genome-wide association mapping. But LD generated by population structure and genetic drift would result in spurious marker-trait associations. As for LD generated by selection, mutation and relatedness, the influence just depends on the population under consideration. Additionally, since Vigouroux et al. (2002) suggested that mutation rate of different types of SSRs in maize was very low, the influence of mutation on LD of SSR loci could be neglected. In our analysis, different number of random samples equal to the number of inbred lines for each subgroup was selected from the entire samples. The results showed that the expected percentage of SSR loci in significant LD was almost the same to that of in each subgroup. This indicated that the population structure, the relatedness, and the genetic drift did not strongly influence the LD of SSR loci in each subgroup. As the high percentage of linked SSRs in significant LD in the entire sample and each subgroup, linkage was assumed to be the major force that generated LD in both the entire sample and each subgroup.

In the present study, a mini core set of maize inbred lines consisting of 95 inbreds for association mapping has been constructed. Diversity analysis by using 145 SSR loci which covered the entire maize genome showed that the population was representative for Chinese maize inbred lines and included diverse genetic variation. Population structure analysis showed four subgroups existed in the population. Though many factors contributed to the LD between SSR loci, linkage was the major force generating and conserving LD of SSR loci. The results suggested that the population may be used in the detection of genome-wide SSR marker–phenotype association.

Acknowledgments The authors thank Dr. Chenyang Hao for technical support, Prof. Jizeng Jia and Xueyong Zhang for advice. We are also grateful to Zhizhai Liu and Ronghua Guo for their kind assistance and suggestions. The research was partly supported by the projects of the Ministry of Science and Technology (2006CB101700), the National Natural Science Foundation (30571133), the Ministry of Agriculture (nb07-070401) and the Beijing Agricultural Innovative Platform - Beijing Natural Science Fund Program (Yzpt02-06).

References

Andersen JR, Schrag T, Melchinger AE, Zein I, Lübberstedt T (2005) Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). Theor Appl Genet 111:206–217

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

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

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

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180–184

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

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

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Huttley GA, Smith MW, Carrington M, Brien SJ (1999) A scan for linkage disequilibrium across the human genome. Genetics 152:1711–1722

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) $Gm^{3;5,13,14}$ and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43:520–526

Kraakman AT, Niks RE, van den Berg PMMM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Li Y, Du J, Wang TY, Shi YS, Song YC, Jia JZ (2002) Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. Maydica 47:93–101

Li XH, Yuan LX, Li XH, Zhang SH, Li MS, Li WH (2003) Heterotic grouping of 70 maize inbred lines by SSR markers. Sci Agric Sin 36:622–627

Li Y, Shi YS, Cao YS, Wang TY (2004) 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 KJ, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21(9):2128–2129

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

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

Peng ZB, Liu XZ, Fu JH, Li LC, Huang CL (1998) Heterotic groups and patterns of maize inbred lines established by RAPDs. Acta Agron Sin 24:711–717

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software. The University of Chicago Press, Chicago

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20:103–111

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Saghai-Maroof MA, Soliman KL, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nat Biotech 18:233–234

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Mol Ecol 9:2109–2118

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

Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman J, Sørensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR marker. Mol Breed 17:217–226

Stich B, Melchinger AE, Piepho HP, Hamrit S, Schipprack W, Maurer HP, Reif JC (2007) Potential caused of linkage disequilibrium in a European maize breeding program investigated with computer simulations. Theor Appl Genet 115:529–536

Sun SM, Li SM, Zheng ZR, Guo BT, Sun ZL, Wang B, Zhang CL, Weng ML (1999) Evaluation of the genetic relationships among basic maize inbred lines in Shangdong. Acta Agron Sin 25:727–732

Templeton AR (1995) A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction-endonuclease mapping or DNA-sequencing. 5. Analysis of case-control sampling designs—Alzheimer's-disease and the apoprotein-E locus. Genetics 140:403–409

Teng WT, Cao JS, Chen YH, Liu XH, Jing XQ, Zhang FJ, Li JS (2004) Analysis of maize heterotic groups and patterns during past decade in China. Sci Agric Sin 37:1804–1811

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ESIV (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 YB, Wang ZH, Lu LX, Wang YP, Zhang X, Tian ZY (1998) Studies on maize germplasm base, division of heterosis groups and utilizing models of heterosis in China. J Maize Sci 6(1):9–13

Wang YB, Wang ZH, Wang YP, Zhang X, Lu LX, Tian ZY (1999) The improvement of principal maize germplasm and utilization of heterosis model in China. J Maize Sci 7(1):1–8

Wang YD, Duan MX, Xing JF, Wang JD, Zhang CY, Guo JL, Zhao JR, Chen SJ (2004) Current situation and prospect of using P maize group to improve the heterosis utility and to development new germplasm. J Maize Sci 12(2):10–12

Whitt SR, Buckler ESIV (2003) Using natural allelic diversity to evaluate gene function. Methods Mol Biol 236:123–139

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ESIV (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

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

Yu YT, Wang RH, Shi YS, Song YC, Wang TY, Li Y (2007) Genetic diversity and structure of the core collection for maize lines in China. Maydica 52:181–194

Yuan LX, Fu JH, Zhang SH, Liu XZ, Peng ZB, Li XH, Warburton M, Khairallah M (2001) Heterotic grouping of maize inbred lines using RFLP and SSR markers. Acta Agron Sin 27:149–156

Zheng SY, Wang SC, Liu DZ (2006) Heterotic grouping of maize inbred lines classified by using SSR markers. J Maize Sci 14(5):26–29