植物遗传资源学报 2022，23（1）：290-300
Journal of Plant Genetic Resources
DOI：10.13430/j.cnki.jpgr.20210712001

玉米低密度育种芯片开发及在种质资源评价中的应用

郭子锋[1]，王山荭[1]，刘　蓓[2]，李文学[1]，王红武[1,2]
（[1]中国农业科学院作物科学研究所，北京 100081；[2]中玉金标记生物（北京）技术股份有限公司，北京 102206）

摘要：SNP 基因分型芯片是分子育种的重要工具，高密度 SNP 芯片往往存在标记冗余、价格高、目标性不强等问题，是分子育种走向常规化、规模化的主要限制因素之一。本研究介绍了一款新开发的低密度育种芯片，并就芯片在种质资源评价中的价值进行了分析。首先，对 37 份玉米自交系进行 10× 重测序，获得了 18.2 Mb 的 SNP 标记，从中挑选 2080 个 SNPs；再从已开发 55 K 芯片中挑选 3390 个缺失率低、多态性高、标记类型为高多态分辨率的标记；最后从 HapMap3 中挑选 586 个标记，设计的育种芯片共包含 6056 个 SNPs，采用靶向测序基因型检测（GBTS，genotyping by target sequencing）技术对标记进行检测。通过自然群体、双亲群体和多亲本重组自交系（MAGIC，multiparent advanced generation inter-cross）群体验证表明，育种芯片检测到的原始设计位点数为 4773~5967 个，自然群体中最小等位基因频率（MAF，minor allele frequency）>0.4 和多态性信息含量（PIC，polymorphism information content）>0.4 的标记比例分别为 57.6% 和 88.6%，MAGIC 群体平均捕获率为 70.6%。用该芯片对 226 份玉米种质资源进行评价，主成分分析可以将其划分为温带和热带两大类群，UPGMA 聚类分析进一步将其划分为 6 个已知类群，分别是瑞德、兰卡斯特、PB、旅大红骨、四平头和热带类群，利用 Structure 软件进行群体结构分析，没有出现最佳 *K* 值，但热带材料都独立成群；类群内和类群间的遗传距离平均值分别为 0.394 和 0.471，其中 PB 群内的遗传距离最小（0.316），热带类群内的遗传距离最大（0.424）；类群间，瑞德与热带之间的遗传距离最大（0.493）；类群间的遗传分化系数（F_{ST}）表明，PB 类群与其他类群间的 F_{ST} 均较大。

关键词：玉米；SNP；育种芯片；种质资源

Development of Maize Low Density Breeding Chip and Its Application in Germplasm Resources Evaluation

GUO Zi-feng[1], WANG Shan-hong[1], LIU Bei[2], LI Wen-xue[1], WANG Hong-wu[1,2]
([1]*Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081*;
[2]*China Golden Marker Biotechnology Co., Ltd, Beijing 102206*)

Abstract: Single nucleotide polymorphism (SNP) genotyping chip is an important tool for molecular breeding. High density SNP chip often has some problems such as marker redundancy, high price, and poor target, limiting its use in the normalization and scale of molecular breeding. Here we developed a low density breeding chip. The panel of 6056 markers were assembled using three resources, consisting of: (1) 2080 key SNPs of 18.2 Mb SNPs which were identified from 10× sequencing of 37 maize inbred lines; (2) 3390 qualified markers with low missing rate, high polymorphism and Conversion Type being as Poly High Resolution within 55 K SNP array; (3) 586 markers that were selected from HapMap3. Genotyping by target sequencing (GBTS) technology was used to detect the markers. Through the verification of natural population, bi-parent population and multiparent advanced generation inter-cross (MAGIC) population, the original design markers were detected ranged from 4773-5967, and 57.6% and 88.6% markers were found by applying minor allele frequency (MAF)>

收稿日期：2021-07-12　**修回日期：**2021-08-24　**网络出版日期：**2021-09-02
URL：http://doi.org/10.13430/j.cnki.jpgr.20210712001
第一作者研究方向为玉米分子育种，E-mail：guozifeng@caas.cn
通信作者：王红武，研究方向为玉米高产育种，E-mail：wanghongwu@caas.cn
基金项目：国家重点研发计划课题（2017YFD0101201）；中央级公益性科研院所基本科研业务费专项（Y2020PT20）
Foundation projects: the National Key Research and Development Program of China (2017YFD0101201), Central Public-interest Scientific Institution Basal Research Fund (Y2020PT20)

0.4 and polymorphic information content (PIC)>0.4 in natural population, respectively. The average capture rate of the breeding chip was 70.6% in MAGIC population. We evaluated 226 inbred lines with this breeding chip. Within this collection two groups (temperate *vs.* tropical) had been classified by principal component analysis (PCA), and six known groups (Reid, Lancaster, PB, LRC, SPT and tropical) were proposed using cluster analysis. Structure analysis has not revealed the best *K* value. The mean genetic distance within and among groups were 0.394 and 0.471, respectively. The genetic distance within PB group was the smallest (0.316), and the genetic distance within tropical group was the largest (0.424). The largest genetic distance (0.493) was observed between group Reid and tropical. The genetic differentiation coefficient (F_{ST}) among the groups indicated that the F_{ST} of PB group was larger than that of other groups.

Key words: maize; SNP; breeding chip; germplasm resource

分子育种技术的大规模推广应用依赖于分子标记成本的降低[1]。相比其他遗传标记，单核苷酸多态性（SNP，single nucleotide polymorphisms）标记在基因组中含量丰富、稳定性强、鉴定效率高[2-3]。在玉米基因组中，每 44~75 bp 就会出现 1 个 SNP[3-5]。此外，SNP 一般为二等位基因标记，易于估计群体中等位基因频率。随着分子生物学的发展，不同通量的 SNP 基因分型平台得到了广泛的应用，对于少数 SNP，主要的技术平台是竞争性等位基因特异性 PCR（KASP，kompetitive allele specific PCR）和 Taqman 技术，而高密度 SNP 标记的鉴定主要通过全基因组重测序、简化基因组测序（GBS，genotyping by sequencing）、固相芯片和靶向测序基因型检测（GBTS，genotyping by target sequencing）（液相芯片）。全基因组重测序或者简化基因组测序可以获得大量的 SNP，适合用于群体遗传研究，但需要配备相应的服务器并建立生物信息学分析平台。固相芯片的理论基础是杂交测序，具有特定的物理载体，目前的两种主要平台是 Affymetrix Axiom 和 Illumina Infinium，分别使用显微光蚀刻和微珠技术。利用固相芯片已经在水稻[6]、小麦[7-8]、大豆[9]等作物上开发了一系列芯片。液相芯片因具有灵活性、广适性、高效性等特点，在动植物中已经开发了 50 余套标记集[10]。目前，玉米中已经开发了多款芯片[11-13]，在种质资源鉴定和分子标记辅助选择（MAS，marker-assisted selection）中发挥了一定的作用。然而，高密度的芯片在育种中往往存在标记冗余、价格高、目标性不强等问题，是分子育种走向常规化、规模化的主要限制因素之一，开发低密度的育种芯片，不但可以降低成本，还可以提高检测效率。因此，本研究根据遗传多样性丰富的 37 份玉米自交系的基因组重测序数据，并结合 55 K 芯片中的优良 SNP 标记，开发出一款高质量、高性价比，适合在育种中应用的低密度育种芯片，并利用该育种芯片对国内外常见的玉米自交系进行了评价。

1　材料与方法

1.1　试验材料

试验材料共包括 2 个部分，第 1 部分材料包含 37 份自交系，进行 10× 重测序，用于芯片的标记开发，材料名称详见表 1；第 2 部分材料，包含 1 个自然群体、1 个双单倍体（DH，doubled haploid）群体和 1 个多亲本重组自交系（MAGIC，multiparent advanced generation inter-cross）群体，用于芯片的验证。其中自然群体包含了 226 份玉米自交系，材料名称详见表 2，编号 1~10 引自美国，编号 11~147 为国内常用自交系，包括 B73、铁 7922、丹 598 等，编号 148~226 引自国际玉米小麦改良中心（CIMMYT，The International Maize and Wheat Improvement Center），包含 36 份 CML 自交系和 43 份 CML 衍生系。DH 群体包含了 171 份自交系，亲本分别为 B73 和 CXS161，其中 CXS161 是掖 478 与旱 23 的回交改良系，系谱为（掖 478 × 旱 23）BC_2F_8。MAGIC 群体是 16 份亲本（其中 8 份热带材料，分别为 CML290、CML411、CML426、CML432、CML496、TR0401、TR0505 和 DTMA238，8 份温带材料，分别为 Mo17、Lx9801、旱 21、吉 419、铁 7922、81565、郑 58 和昌 7-2）通过两两杂交，再经过自交产生的，目前已经自交了 7 代，共包含 297 份自交系，编号分别为 M1-M297，加上亲本共 313 份。DH 群体的亲本 B73 和 CX161 进行了两次生物学重复。

表 1 37 份重测序自交系清单

Table 1 List of 37 re-sequenced inbred lines

名称 Name	类群 Groups	Clean_Reads	Clean_Base	Q20（%）	Q30（%）	GC 含量（%） GC content
1538	瑞德	92835438	27807565006	96.79	91.71	46.18
CR1HT	兰卡斯特	79968771	23955500672	96.46	91.26	46.31
LH220Ht	瑞德	89945383	26948949996	96.73	91.68	46.32
NL001	瑞德	70279504	21055094098	96.82	91.87	45.88
PHW30	瑞德	95098271	28490853030	97.05	92.28	46.60
PHW43	PB	85754934	25690524148	97.17	92.56	45.90
GR1HT	兰卡斯特	83409266	24987564928	97.28	92.73	46.12
8107	旅大红骨	75474550	22608503290	96.88	91.89	46.33
B73	瑞德	73039343	21882696630	97.17	92.47	45.96
B84	瑞德	78723531	23585538274	96.52	91.32	46.10
K12	四平头	83874495	25128388128	96.61	91.45	46.04
PHN47	四平头	81362735	24377004684	97.30	92.80	45.98
PI143	瑞德	73441386	22003113246	96.85	91.90	46.36
PI36	四平头	82246831	24637540616	97.39	93.06	46.44
Q1261	四平头	79552113	23826969660	96.89	92.14	45.88
独 321 Du 321	瑞德	89899708	26929202788	97.11	92.34	46.41
多黄 29 Duohuang 29	PB	88818134	26608395508	95.68	89.58	46.41
吉 046 Ji 046	瑞德	76795925	23007439138	96.81	91.81	45.95
吉 A-034 Ji A-034	PB	83735850	25087741192	97.25	92.71	46.26
金黄 55 Jinhuang 55	PB	75055688	22486943616	97.07	92.34	46.13
辽 526 Liao 526	旅大红骨	79676323	23870076556	96.71	91.55	46.39
辽白 371 Liaobai 371	热带	80924668	24242259934	96.70	91.68	46.19
四 273 Si 273	旅大红骨	75788362	22703522746	96.77	91.66	45.95
四 -279 Si-279	四平头	85631695	25654237216	96.94	92.07	46.05
绥系 605 Suixi 605	瑞德	78939259	23646359204	96.51	91.14	46.25
早 8-3 Zao 8-3	unknown	83335153	24964472608	96.44	91.04	46.22
中黄 204 R2040 Zhonghuang 204 R2040	四平头	80546170	24131219900	96.20	90.40	46.28
综 31 Zong31	旅大红骨	76465200	22908665376	96.75	91.63	46.01
CXS62	四平头	75557001	22634887126	96.78	91.76	45.90
CXS109	四平头	72883759	21836184634	97.07	92.29	45.94
CXS161	瑞德	73317868	21964107868	96.87	91.86	45.89
TR0509	热带	71905300	21531347726	96.45	91.06	46.12
DTMA165	热带	71973704	21557206180	96.82	91.73	45.94
Mo17	兰卡斯特	80658999	24162633284	96.74	91.58	45.94
444	unknown	82168214	24615130404	96.20	90.75	46.26
早 21 Han 21	unknown	81881214	24532176054	97.00	92.15	46.23
黄早四 Huangzaosi	四平头	72449175	21698233168	97.06	92.30	45.99

表 2　自然群体包含的玉米种质

Table 2　Maize germplasms in the natural population

编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups
1	807	温带	瑞德	28	8107	温带	旅大红骨	55	CN962	温带	四平头	82	PI36	温带	四平头
2	78004	温带	瑞德	29	89-1	温带	四平头	56	CP619F	热带	热带	83	R09	温带	PB
3	B47	温带	瑞德	30	7537-1	温带	瑞德	57	CWF	温带	瑞德	84	S37	温带	旅大红骨
4	LH146Ht	温带	瑞德	31	7595-2	温带	瑞德	58	CWM	热带	热带	85	S7913	温带	瑞德
5	LH190	温带	瑞德	32	2002F22	温带	四平头	59	D185	温带	兰卡斯特	86	铁 9010	温带	旅大红骨
6	PHG47	温带	兰卡斯特	33	200B	温带	四平头	60	D387	温带	兰卡斯特	87	Va35	温带	兰卡斯特
7	PHJ89	温带	兰卡斯特	34	3H-2	温带	四平头	61	DH34	温带	旅大红骨	88	本 M130	温带	兰卡斯特
8	ICI193	温带	瑞德	35	5022（B）	温带	瑞德	62	东 46	温带	PB	89	川 29	温带	PB
9	OS602	温带	瑞德	36	53 选 3	温带	兰卡斯特	63	D 黄 212	温带	四平头	90	川 321	温带	瑞德
10	IRF310	温带	兰卡斯特	37	698-1	温带	瑞德	64	H10	温带	瑞德	91	丹 1324	温带	兰卡斯特
11	32	温带	瑞德	38	698-3	温带	瑞德	65	H152	温带	四平头	92	丹 3130	温带	PB
12	141	温带	PB	39	706 辐	温带	兰卡斯特	66	H201	温带	瑞德	93	丹 598	温带	四平头
13	196	温带	四平头	40	888-9	温带	兰卡斯特	67	H3	温带	瑞德	94	丹 988	温带	PB
14	412	温带	兰卡斯特	41	B234	温带	瑞德	68	HZ85	温带	瑞德	95	东 156	温带	四平头
15	485	温带	兰卡斯特	42	B73	温带	瑞德	69	J001	温带	兰卡斯特	96	冬 10	温带	四平头
16	495	温带	兰卡斯特	43	B84	温带	瑞德	70	K10	温带	兰卡斯特	97	关花	温带	四平头
17	501	温带	PB	44	BJ005	温带	瑞德	71	M0113	温带	瑞德	98	海 014	温带	四平头
18	西 502	温带	四平头	45	ВИР44	温带	旅大红骨	72	MBNA	温带	瑞德	99	海 9-21	温带	PB
19	653	温带	瑞德	46	C416	温带	兰卡斯特	73	N528-1（1284）	温带	兰卡斯特	100	黄 C	温带	瑞德
20	803	温带	旅大红骨	47	C649	温带	兰卡斯特	74	P138	温带	PB	101	获唐黄	温带	兰卡斯特
21	812	温带	瑞德	48	C8605-2	温带	瑞德	75	PA276	温带	瑞德	102	获唐黄 17	温带	兰卡斯特
22	835	温带	四平头	49	CA112	温带	旅大红骨	76	PA281	温带	瑞德	103	吉 412	温带	兰卡斯特
23	3189	温带	瑞德	50	CA156	温带	瑞德	77	PA311	温带	瑞德	104	吉 A-034	温带	PB
24	5213	温带	四平头	51	CA339	温带	旅大红骨	78	PH6WC	温带	瑞德	105	冀 35	温带	四平头
25	5311	温带	瑞德	52	CA375	温带	旅大红骨	79	PHN47	温带	四平头	106	金黄 63	温带	四平头
26	6103	温带	瑞德	53	CAL70	温带	旅大红骨	80	PI10	温带	瑞德	107	金黄 73	温带	旅大红骨
27	6523	温带	瑞德	54	CN165	热带	热带	81	PI143	温带	四平头	108	金黄 76	温带	四平头

表 2（续）

编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups	编号 ID	名称 Name	类型 Type	类群 Groups
109	金黄 96B	温带	PB	139	CXS77	温带	四平头	169	CML326	热带	热带	199	IMAS130A	热带	热带
110	京 7 黄	温带	四平头	140	CXS137	温带	瑞德	170	CML330	热带	热带	200	TR0395	热带	热带
111	辽 138	温带	旅大红骨	141	CXS181	温带	瑞德	171	CML360	热带	热带	201	TR0396	温带	PB
112	辽 184	温带	旅大红骨	142	XZY357-1	温带	旅大红骨	172	CML361	热带	热带	202	TR0412	热带	热带
113	辽 2345	温带	瑞德	143	XZY454-1	温带	PB	173	CML408	热带	热带	203	TR0415	热带	热带
114	辽 3053	温带	瑞德	144	XZY342-1	温带	PB	174	CML411	热带	热带	204	TR0418	热带	热带
115	辽 5114	温带	瑞德	145	Q-9-188	温带	PB	175	CML415	热带	热带	205	DTMA144A	热带	热带
116	南 23-32	温带	旅大红骨	146	Y-9-141	热带	热带	176	CML423	热带	热带	206	TR0423	热带	热带
117	南五	温带	兰卡斯特	147	C5 RIL 3	热带	热带	177	CML426	热带	热带	207	622016	温带	四平头
118	品 1P6Co	温带	旅大红骨	148	CML31	热带	热带	178	CML430	热带	热带	208	DTMA11A	热带	热带
119	齐 205	温带	兰卡斯特	149	CML68	热带	热带	179	CML454	热带	热带	209	CKL05018	热带	热带
120	齐 206	温带	兰卡斯特	150	CML85	热带	热带	180	CML465	热带	热带	210	CLA161	热带	热带
121	沈 136	温带	PB	151	CML94	热带	热带	181	CML468	热带	热带	211	CL-G1839A	热带	热带
122	沈 3336	温带	瑞德	152	CML99	热带	热带	182	CML470	热带	热带	212	CLQRCYQ14	热带	热带
123	双 M9	温带	四平头	153	CML103	热带	热带	183	CML493	热带	热带	213	CLYN231	热带	热带
124	四 -279	温带	四平头	154	CML118	热带	热带	184	TR0143	热带	热带	214	TR0465	热带	热带
125	四 F1	温带	兰卡斯特	155	CML122	热带	热带	185	TR0400	热带	热带	215	CY9169	温带	旅大红骨
126	四川地方种质	温带	四平头	156	CML127	热带	热带	186	TR0401	热带	热带	216	CZL068	热带	热带
127	特 70	温带	兰卡斯特	157	CML130	热带	热带	187	TR0403	热带	热带	217	DO620Y	温带	旅大红骨
128	铁 7922	温带	瑞德	158	CML133	热带	热带	188	TR0390	热带	热带	218	TR0467	热带	热带
129	文黄 31413	温带	四平头	159	CML154	热带	热带	189	TR0404	热带	热带	219	DTMA269A	热带	热带
130	系 14	温带	兰卡斯特	160	CML166	热带	热带	190	TR0406	热带	热带	220	DTMA241	热带	热带
131	豫自 87-1	温带	PB	161	CML171	热带	热带	191	DTMA25A	热带	热带	221	DTMA231	热带	热带
132	早 8-3	温带	瑞德	162	CML191	热带	热带	192	TR0448	热带	热带	222	DTMA233A	热带	热带
133	郑 28	温带	瑞德	163	CML282	热带	热带	193	TR0452	热带	热带	223	TR0473	热带	热带
134	CXS34	温带	PB	164	CML287	热带	热带	194	TR0453	热带	热带	224	TR0479	热带	热带
135	CXS36	温带	PB	165	CML292	热带	热带	195	TR0456	热带	热带	225	TR0480	热带	热带
136	CXS41	温带	PB	166	CML304	热带	热带	196	TR0457	热带	热带	226	H-16	热带	热带
137	CXS46	温带	PB	167	CML323	热带	热带	197	TR0461	热带	热带				
138	CXS54	温带	PB	168	CML325	热带	热带	198	TR0248A	热带	热带				

1.2　芯片设计与开发

芯片原始设计 6056 个 SNP 标记（图 1，详见 http://doi.org/10.13430/j.cnki.jpgr.20210712001，附表 1），SNP 标记共有 3 个来源：（1）对 37 份玉米自交系进行 10× 重测序，共获得 2963.4 Mb 的 Clean Reads，与 B73 RefGen_v4 参考基因组（ftp://ftp.ensemblgenomes.org/pub/plants/release-40/fasta/zea_mays/dna/Zea_mays.AGPv4.dna.toplevel.fa.gz）比对后检测到 18.2 Mb 的 SNPs，过滤后的 SNP 数为 2.8 Mb，根据最小等位基因频率（MAF，Minor allele frequency）和标记的分布挑选了高质量的 SNP 共 2080 个；（2）根据已有 8 盘共 3072 个样品的 55 K 芯片数据，从中挑选缺失率（Missing rate）<5.0%、MAF>0.35、标记类型为高多态分辨率的标记，考虑性状相关标记和标记在染色体上的均匀分布，共挑选了 3390 个标记，其中包含与抗穗腐病、耐渍和耐低磷相关的标记 393 个[14-16]；（3）最后从 HapMap3 中挑选 586 个标记，用于填补前两个步骤中标记间隔比较大的区间。标记的检测采用 GBTS 技术，在中玉金标记生物（北京）技术股份有限公司进行检测。GBTS 主要包括以下 5 个步骤[1]：A 利用超声波将基因组 DNA 进行片段化并加上测序接头；B 将带有生物素标记的 RNA 探针与已经带有接头序列的 DNA 片段结合；C 链霉亲和素包裹的磁珠与上一步的双链复合物相结合（探针过量）；D 清洗得到目标区域的 DNA，目的是去除非特异性杂交，提高捕获效率；E 对洗脱下来的 DNA 产物进行 PCR 扩增，构建 Illumina 测序文库。测序数据下机后使用 GATK 的 HC 模型进行变异检测，用 vcftools 进行标记过滤。

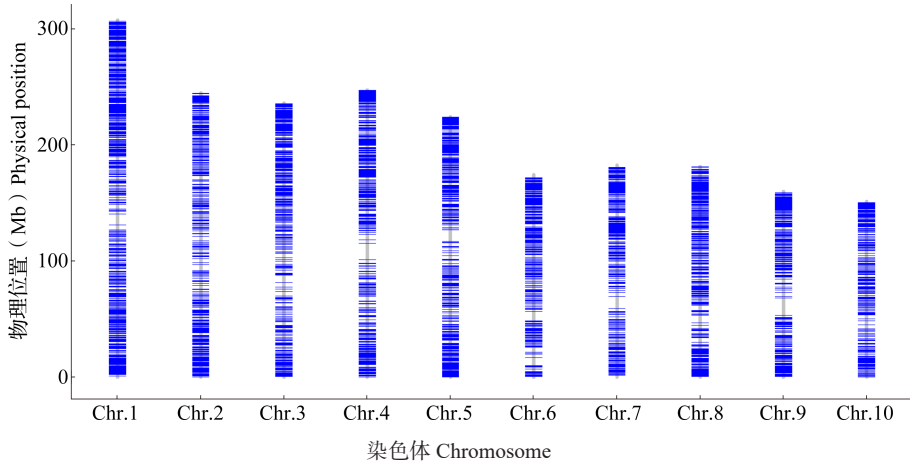

图 1　SNP 标记在玉米 10 条染色体上的分布

Fig.1　Distribution of single nucleotide polymorphism（SNP）on 10 chromosomes of maize

1.3　数据分析

对自然群体检测的标记，在 Excel 2019 中计算缺失率、MAF 和多态性信息含量（PIC，Polymorphic information content），其中 PIC 根据 Botstein 等[17]的描述进行计算，公式如下：

$$PIC = 1 - \left(\sum_{i=1}^{n} P_i^2\right) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2P_i^2P_j^2$$

其中，P_i 和 P_j 是第 i 个和第 j 个等位基因的频率。系统发育树在 TASSEL5.0 软件中构建，聚类方法采用非加权组平均法（UPGMA，unweighted pair-group method with arithmetic mean），聚类结果以 *.tree 格式导出后在 FigTreev1.4.3 软件中进一步作图优化。群体结构划分采用 Structure 软件，亚群数 K 设置为 2~10，Length of burn-in period 和 Number of MCMC Reps after Burn-in 都设置为 10000。最佳 K 值估算参照 Evanno 等[18]的方法。在 TASSEL5.0 中计算两两自交系之间的遗传距离矩阵，用不同类群间自交系之间的平均值代表类群之间的遗传距离。类群间的遗传分化系数（F_{ST}，Genetic differentiation coefficient）根据如下公式在 Excel 2019 中计算。

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

其中，H_T 表示群体总的杂合性，H_S 表示亚群的平均杂合性，F_{ST} 越大，表示亲缘关系越远；F_{ST} 越小，表示亲缘关系越近。

2　结果与分析

2.1　标记统计分析

将 226 个玉米自交系构成的自然群体和 171

个 DH 系的靶向测序数据分别比对到玉米 B73 RefGen_v4 参考基因组上，使用 GATK 软件进行变异检测。自交系群体比对获得 SNP 标记 358135 个，过滤后剩余 283343 个，以个体 GQ10 miss0.2 maf 0.05 过滤后获得 SNP 标记 35589 个，其中 32217 个 SNPs 在原始设计位点上或者上下游 100 bp 内；以个体 GQ20 miss0.1 maf 0.05 过滤获得 24521 个 SNPs，标记均在目标位点或上下游 150 bp 内，其中原始设计位点 4773 个。从 SNP 在基因组中的分布情况看，外显子中 SNP 分布最多，占 25%，其次为基因上游，占 19%（图 2A）；每个目标区段平均检测到 5.1 个 SNP（图 2B）。DH 群体使用相同的软件及参数进行分析，在个体符合 GQ10 miss0.2 的过滤条件下，上述 24521 个 SNPs 在 DH 家系中有 22192 个被检测到，比例为 90.5%。对 MAGIC 群体及亲本的 313 份材料目标区域测序后进行比对，GATK 检测获得 SNP 标记 199154 个，过滤后剩余 145646 个，以 GQ10 miss0.2 maf 0.05 过滤后获得 55586 个 SNPs，其中 5967 个 SNPs 是最初设计的原始标记，以 GQ10 miss0.1 maf 0.05 过滤后获得 50130 个 SNPs，其中 5767 个 SNPs 是最初设计的原始位点。其中，自然群体和 DH 群体检测的 SNP 与 MAGIC 群体检测的 SNP 相比，共同检测到的原始设计位点数为 4560 个。

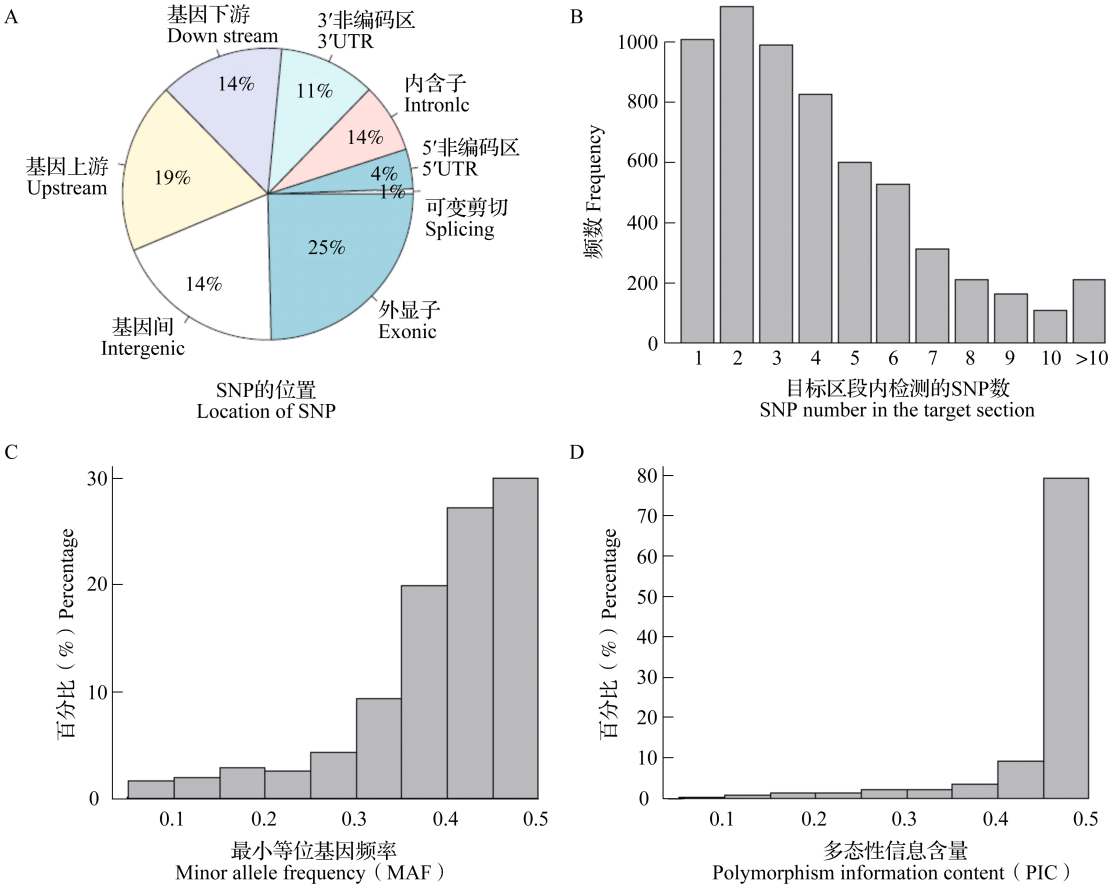

图 2 SNP 在基因组上的位置及标记的表征

Fig.2 Location of SNP and marker characterization

DH 群体 2 个亲本 B73 和 CXS161 生物学重复的一致性均为 99.2%，不考虑缺失基因型时一致性均为 99.9%。以 MAGIC 群体检测为例，平均比对率为 99.1%，捕获效率在 56.0%~77.5% 之间，平均值为 70.6%。目标区域测序深度在 99.3~391.1×，平均测序深度为 184.8×，表 3 为随机选择 10 个家系所得的捕获效率和目标区段测序深度。原始设计位点具有较高的多态性，以自然群体为例，MAF 和 PIC 的平均值分别为 0.39 和 0.46，其中 MAF>0.4 和 PIC>0.4 的标记比例分别为 57.6% 和 88.6%，而 MAF<0.2 和 PIC<0.2 的比例分别为 6.7% 和 2.2%，标记具有较高的多态性（图 2C、D）。

表 3　10 个 MAGIC 群体家系捕获效率与目标区段测序深度

Table 3　Target ratio and sequencing depth of target section for 10 multiparent advanced generation inter-Cross（MAGIC）lines

样本 Sample	干净读长（Mb）Clean reads	比对读长（Mb）Mapped reads	目标读长（Mb）Target reads	比对率（%）Mapping rate	捕获效率（%）Target ratio	目标区段测序深度（×）Sequencing depth of target section
M078	2.22	2.20	1.66	99.23	75.28	152.81
M083	2.37	2.35	1.63	99.11	69.35	143.79
M118	2.93	2.90	2.15	98.92	74.22	191.32
M128	2.78	2.75	2.06	98.99	74.91	180.91
M169	3.49	3.46	2.49	98.93	71.95	216.22
M185	4.50	4.45	3.23	98.94	72.63	281.86
M219	3.53	3.50	2.44	99.18	69.84	221.22
M226	3.41	3.39	2.30	99.26	67.95	209.23
M231	3.03	3.01	2.18	99.26	72.51	197.54
M287	3.88	3.85	2.53	99.20	65.56	232.76

2.2　类群划分

主成分分析结果显示，当主成分（PC，Principal component）设置为 2 时，自然群体可以明显地划分为 2 个类群，分别是热带类群和温带类群，热带类群不同自交系的前两个主成分比较聚集，而温带类群则比较分散（图 3A）。UPGMA 聚类分析可以将 226 个自交系分为 6 个类群，依据骨干自交系，这 6 个类群分别是热带、瑞德、兰卡斯特、PB、四平头和旅大红骨（图 3B）。热带材料共有 80 份，包括 36 份 CML 自交系以及 CML 衍生的自交系，一些含有热带血缘的自交系如 CN165 也被划分为热带类群；瑞德类群包含了 49 份自交系，骨干自交系 B73、B84、黄 C 及先玉 335 的母本系 PH6WC 都被划分为瑞德类群，本次类群划分与之前我们用 55 K 芯片和 20 K 系列芯片划分的类群略有不同[1,12]，区别是将 PA、Iodent 类群统一为瑞德，如铁 7922、辽 3053、辽 5114 等传统上属于 PA 群，由于 PA 群与瑞德的关系较近，本次我们将其归为瑞德类群；兰卡斯特、PB、四平头和旅大红骨则分别包含了 28、22、29 和 18 份自交系（表 2），与之前的划分结果一致[1,12]，总体上，对 226 份种质资源的评价结果与已知的中国玉米材料类群划分是一致的[19-20]。

用 Structure 软件将 *K* 值设置为 2~10，与 Lu 等[19]的研究结果一样，没有出现最佳 *K* 值（图 3C），当 *K*=2 时，自然群体可以被明显地划分为热带和温带两个类群，当 *K*=3 时，热带材料被单独划分为一个类群，而温带材料则被划分为国内种质和国外种质两个类群。以 *K*=6 时的划分结果为例，群体结构与聚类结果一致（图 3D，详见 http://doi.org/10.13430/j.cnki.jpgr.20210712001，附表 2）。

2.3　遗传距离和遗传分化系数

利用低密度育种芯片计算不同类群间的遗传距离，结果见表 4，类群间遗传距离大于类群内遗传距离，其平均值分别为 0.471 和 0.394。类群内遗传距离范围为 0.316~0.424，其中 PB 类群内的遗传距离最小，热带类群内的遗传距离最大，证明热带材料具有更丰富的遗传多样性。类群间遗传距离范围为 0.456~0.493，兰卡斯特与瑞德和旅大红骨之间的遗传距离均最小，瑞德与热带类群之间的遗传距离最大，热带类群与其他几个类群之间具有较大的遗传距离，此外，PB 与四平头之间具有较大的遗传距离，说明 PB 与四平头之间可能存在较强的杂种优势。利用新开发的低密度育种芯片计算的遗传距离高于 55 K 芯片和 20 K 芯片计算的结果[1,12]。

类群间的 F_{ST} 范围为 0.069~0.149（表 4），热带类群与旅大红骨之间的 F_{ST} 最小（0.069），四平头与 PB 之间的 F_{ST} 最大（0.149），这与它们之间的遗传距离大小一致，进一步证明这两个类群之间可能存在较强的杂种优势；总体上，PB 类群与其他类群间的 F_{ST} 均较大。国内类群四平头与旅大红骨之间的 F_{ST} 为 0.070，低于旅大红骨与瑞德和兰卡斯特之间的 F_{ST}，其值分别为 0.078 和 0.076，瑞德与兰卡

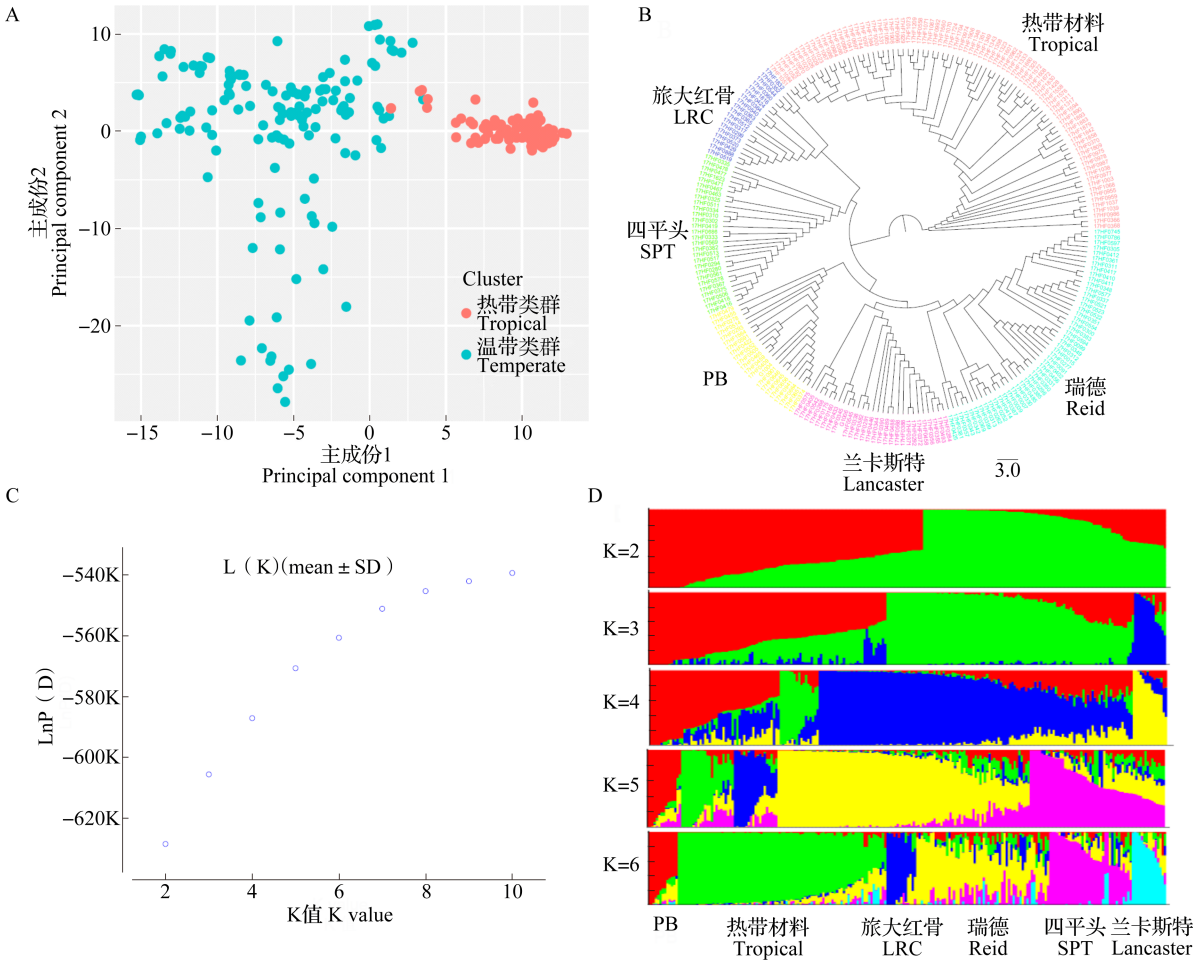

图 3 自然群体主成分分析（A）、邻接进化树（B）、LnP（D）值随 *K* 值变化曲线（C）与群体结构（D）

Fig.3 Principal coordinate analysis（A）, adjacent phylogenetic tree（B）, the LnP（D）statistic for each given *K*（C）, and population structure（D）for the natural population

表 4 不同类群遗传距离及遗传分化系数（F_{ST}）

Table 4 Genetic distances and pairwise F_{ST} comparisons among different groups

类群 Group	瑞德 Reid	兰卡斯特 Lancaster	PB	四平头 SPT	旅大红骨 LRC	热带类群 Tropical
瑞德 Reid	0.406	0.075	0.129	0.088	0.078	0.094
兰卡斯特 Lancaster	0.456	0.389	0.145	0.090	0.076	0.099
PB	0.459	0.468	0.316	0.149	0.136	0.137
四平头 SPT	0.471	0.466	0.482	0.405	0.070	0.072
旅大红骨 LRC	0.464	0.456	0.472	0.459	0.422	0.069
热带类群 Tropical	0.493	0.489	0.483	0.472	0.475	0.424

对角线和对角线下方分别表示类群内和类群间遗传距离，对角线上方表示遗传分化系数

The diagonal cell and lower diagonal represent the genetic distance as measured by pairwise differences within and among groups respectively, the upper diagonal represent pairwise F_{ST} comparisons

斯特之间的 F_{ST} 为 0.075，同样低于其与四平头和旅大红骨之间的 F_{ST}，这反映出国内外种质的遗传分化规律，国外种质与国内种质之间较易产生杂种优势。

3 讨论

利用育种芯片进行选择，一般能够缩短世代间隔，提高选择的准确度，加快遗传进展。中国育种

芯片研究始于 1997-1998 年，尽管起步晚，但是发展很快，在玉米、水稻、小麦、大豆等主要农作物上开发了不同密度的育种芯片或者微阵列。由于玉米基因组大、结构复杂、重组率高、材料类型丰富，开发高质量、高密度 SNP 阵列存在很多障碍[21]。目前，玉米中开发的芯片包括基于 Affymetrix 平台的 600 K 芯片[11]、55 K 芯片[12]和 6H-60K 芯片[13]，基于 Illumina 平台的 50 K 芯片[22]和 3 K 芯片[23]，基于 GoldenGate 平台的 1536 微阵列[24]和 3072 微阵列[25]，基于 GenoBaits 的 20 K 系列液相芯片[1]等。这些芯片的开发为玉米遗传研究提供了重要的支撑。传统的固相芯片的理论基础是 DNA 双螺旋的互补配对。虽然固相芯片鉴定的准确度比较高，但存在的问题是价格高、鉴定周期受样品量的限制。以我们开发的 55 K 芯片为例，定制规格为 384 × 55 K，5 个样本的生物学重复一致率在 97.1%~98.9%，缺失率平均值为 1.83%[12]，虽然略低于目前开发的低密度育种芯片，但相比其他芯片已经具有较高的质量。然而，在实际应用的时候，往往需要几个科研机构或者育种公司凑足 384 个样本才能上机，基因型鉴定的周期从 15 d 延长到平均 2 个月甚至更长，大大限制了其在分子标记辅助选择中的应用。我们开发的 20 K 系列液相芯片解决了 55 K 芯片的上述限制[1]，20 K 系列液相芯片适合用于种质资源鉴定、遗传图谱构建和全基因组选择等。利用 20 K 系列液相芯片进行种质资源鉴定和类群划分时，我们发现 20 K 与 10 K、5 K、1 K 标记的划分结果完全一致，说明玉米育种中类群的划分使用更低密度的芯片即可，而降低密度可以进一步降低成本。例如，当标记密度降低到原来的 1/2 时，成本降低为原来的 2/3，而当标记密度降低为原来的 1/4 时，成本约为原来的 1/2[1]，因此，有必要开发精准高效的低密度育种芯片。我们利用新开发的低密度育种芯片对国内外常用的自交系进行群体结构和杂种优势群的划分，当 PC=2 时，首先将 226 份自交系划分为温带与热带 2 个类群（图 3），其中来自于 CIMMYT 的自交系大多属于热带材料，部分含有热带血缘的材料如 CN165 和 CWM（中糯父 S9）也被划分为热带类群，这与前人的研究结果和育种实际相一致[12,26]。进一步利用 UPGMA 聚类分析将 226 份自交系划分为 6 个类群，除热带类群外，温带类群包括瑞德、兰卡斯特、PB、四平头和旅大红骨 5 个亚群，这与中国玉米主要类群相一致[27-31]，与应用高密度芯片相比，低密度育种芯片能够简化类群划分，如应用 55 K 芯片将 OS602 划分为 Iodent 类群，将 653、BJ005 和 CA156 等划分为 PA 类群[12,15]，此次统一划分为瑞德类群。因此，新开发的低密度育种芯片能够在降低成本的同时优化玉米类群划分。

参考文献

[1] Guo Z, Wang H, Tao J, Ren Y, Xu C, Wu K, Zou C, Zhang J, Xu Y. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Molecular Breeding, 2019, 39: 37

[2] Brumfield R T, Beerli P, Nickerson D A, Edwards S V. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution, 2003, 18: 249-256

[3] Bevan M W, Uauy C. Genomics reveals new landscapes for crop improvement. Genome Biology, 2013, 14: 1-11

[4] Gore M A, Chia J M, Elshire R J, Sun Q, Ersoz E S, Hurwitz B L, Peiffer J A, McMullen M D, Grills G S, Ross-Ibarra J, Ware D H, Buckler E S. A first-generation haplotype map of maize. Science, 2009, 326: 1115-1117

[5] Varshney R K, Nayak S N, May G D, Jackson S A. Next generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology, 2009, 27: 522-530

[6] Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, Tang X, Zhou F, Wang X, Zhang Q. A high-density SNP genotyping array for rice biology and molecular breeding. Molecular Plant, 2014, 7: 541-553

[7] Winfield M O, Allen A M, Burridge A J, Barker G L A, Benbow H R, Wilkinson P A, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King L, Bentley A R, Edwards K J. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnology Journal, 2016, 14 (5): 1195-1206

[8] Sun C, Dong Z, Zhao L, Ren Y, Zhang N, Chen F. The wheat 660 K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnology Journal, 2020, 18 (6): 1354-1360

[9] Lee Y G, Jeong N, Kim J H, Lee K, Kim K H, Pirani A, Ha B K, Kang S T, Park B S, Moon J K, Kim N, Jeong S C. Development, validation and genetic analysis of a large soybean SNP genotyping array. The Plant Journal, 2015, 81: 625-636

[10] 徐云碧，杨泉女，郑洪建，许彦芬，桑志勤，郭子锋，彭海，张丛，蓝昊发，王蕴波，吴坤生，陶家军，张嘉楠 . 靶向测序基因型检测（GBTS）技术及其应用 . 中国农业科学，2020，53（15）：2983-3004
Xu Y B, Yang Q N, Zheng H J, Xu Y F, Sang Z Q, Guo Z F, Peng H, Zhang C, Lan H F, Wang Y B, Wu K S, Tao J J, Zhang J N. Genotyping by target sequencing (GBTS) and its applications. Scientia Agricultura Sinica, 2020, 53 (15): 2983-3004

[11] Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom T M, Fries R, Pausch H, Bertani C, Davassi A, Mayer K F, Schön C C. A powerful tool for genome analysis in maize: development and evaluation of

the high density 600 k SNP genotyping array. BMC Genomics, 2014, 15: 823

[12] Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P, Zou C. Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. Molecular Breeding, 2017, 37: 20

[13] Tian H, Yang Y, Yi H, Xu L, He H, Fan Y, Wang L, Ge J, Liu Y, Wang F, Zhao J. New resources for genetic studies in maize (*Zea mays* L.): a genome-wide Maize6H-60K single nucleotide polymorphism array and its application. The Plant Journal, 2021, 105: 1113-1122

[14] Guo Z, Zou C, Liu X, Wang S, Li W X, Jeffers D, Fan X, Xu M, Xu Y. Complex genetic system involved in Fusarium ear rot resistance in maize as revealed by GWAS, bulked sample analysis, and genomic prediction. Plant Disease, 2020, 104: 1725-1735

[15] Xu C, Zhang H, Sun J, Guo Z, Zou C, Li W X, Xie C, Huang C, Xu R, Liao H, Wang J, Xu X, Wang S, Xu Y. Genome-wide association study dissets yield components associated with low-phosphorus stress tolerance in maize. Theoretical and Applied Genetics, 2018, 131: 1-16

[16] Guo Z, Zhou S, Wang S, Li W X, Du H, Xu Y. Identification of major QTL for waterlogging tolerance in maize using genome-wide association study and bulked sample analysis. Journal of Applied Genetics, 2021, 62: 405-418

[17] Botstein D, White R L, Skolnick M, Davis R W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics, 1980, 32: 314-331

[18] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 2005, 14: 2611-2620

[19] Lu Y, Yan J, Guimarães C T, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek B S, Magorokosho C, Mugo S, Makumbi D, Parentoni S N, Shah T, Rong T, Crouch J H, Xu Y. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theoretical and Applied Genetics, 2009, 120: 93-115

[20] 赵久然,李春辉,宋伟,王元东,张如养,王继东,王凤格,田红丽,王蕊 . 基于 SNP 芯片揭示中国玉米育种种质的遗传多样性与群体遗传结构 . 中国农业科学, 2018, 51(4): 626-634
Zhao J R, Li C H, Song W, Wang Y D, Zhang R Y, Wang J D, Wang F G, Tian H L, Wang R. Genetic diversity and population structure of important Chinese maize breeding germplasm revealed by SNP-chips. Scientia Agricultura Sinica, 2018, 51(4), 626-634

[21] Romay M, Millard M J, Glaubitz J C, Peiffer J A, Swarts K L, Casstevens T M, Elshire R J, Acharya C B, Mitchell S E, Flint-Garcia S A, McMullen M D, Holland J B, Buckler E S, Gardner C A. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology, 2013, 14: R55

[22] Ganal M W, Durstewitz G, Polley A, Bérard A, Buckler E S, Charcosset A, Clarke J D, Graner E M, Joets J, Le Paslier M C, McMullen M D, Montalent P, Rose M, Schön C C, Sun Q, Walter H, Martin O C, Falque M. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE, 2011, 6(12): e28334

[23] Rousselle Y, Jones E, Charcosset A, Moreau P, Robbins K, Stich B, Knaak C, Flament P, Karaman Z, Martinant J P, Fourneau M, Taillardat A, Romestant M, Tabel C, Bertran J, Ranc N, Lespinasse D, Blanchard P, Kahler A, Chen J, Kahler J, Dobrin S, Warner T, Ferris R, Smith S. Study on essential derivation in maize: III. Selection and evaluation of a panel of single nucleotide polymorphism loci for use in European and North American germplasm, Crop Science, 2015, 55: 1170-1180

[24] Yan J, Yang X, Shah T, Sánchez H, Li J, Warburton M, Zhou Y, Crouch J H, Xu Y. High-throughput SNP genotyping with the Golden Gate assay in maize. Molecular Breeding, 2010, 25: 441-451

[25] Tian H, Wang F, Zhao J, Yi H, Wang L, Wang R, Yang Y, Song W. Development of maize SNP3072 array, for DNA finger printing identification of Chinese maize varieties. Molecular Breeding, 2015, 35: 136

[26] Wu X, Li Y, Shi Y, Song Y, Wang T, Huang Y, Li Y. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. Theoretical and Applied Genetics, 2014, 127: 621-631

[27] Xie C, Zhang S, Li M, Li X, Hao Z, Bai L, Zhang D, Liang Y. Inferring genome ancestry and estimating molecular relatedness among 187 Chinese maize inbred lines. Journal of Genetics and Genomics, 2007, 34(8): 738-748

[28] Wang R, Yu Y, Zhao J, Shi Y, Song Y, Wang T, Li Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. Theoretical and Applied Genetics, 2008, 117: 1141-1153

[29] 黎裕,王天宇 . 我国玉米育种种质基础与骨干亲本的形成 . 玉米科学, 2010, 18(5): 1-8
Li Y, Wang T Y. Germplasm base of maize breeding in China and formation of foundation parents. Maize Science, 2010, 18(5): 1-8

[30] Yang X, Yan J, Shah T, Warburton M L, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J. Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theoretical and Applied Genetics, 2010, 121: 417-431

[31] 滕文涛,曹靖生,陈彦惠,刘向辉,景希强,张发军,李建生 . 十年来中国玉米杂种优势群及其模式变化的分析 . 中国农业科学, 2004, 37(12): 1804-1811
Teng W T, Cao J S, Chen Y H, Liu X H, Jing X Q, Zhang F J, Li J S. Analysis of maize heterotic groups and patterns during past decade in China. Scientia Agricultura Sinica, 2004, 37(12): 1804-1811