

温热带玉米株高的全基因组关联分析

郭子锋, 徐孝洁, 刘小刚, 王山荭

(中国农业科学院作物科学研究所, 北京 100081)

摘要: 以365份遗传背景丰富的玉米自交系为材料, 采用自主研发的55K芯片, 在海南三亚、河南新乡和北京顺义3个环境下对玉米自交系株高进行全基因组关联分析(Genome-wide association study, GWAS)。结果表明, 株高在3个环境下变异非常丰富, 有从低纬度到高纬度逐渐增加的趋势, 株高在基因型、环境、基因型与环境互作方面均达极显著差异水平。全基因组关联分析表明, 不同环境下均检测到与株高显著关联的位点, 在第1、3、5、6、7、8号染色体上均发现了显著关联的SNP。新乡和顺义环境下, 在bin5.05区段发现了显著关联的SNP; 三亚和顺义环境下, 在bin6.05区段发现了显著关联的SNP。对关联分析显著的SNP位点比对到B73参考基因组上, 在maizeGDB网站上得到22个最可能的候选基因, 为探索玉米株高遗传机制提供参考。

关键词: 玉米; 株高; 全基因组关联分析

中图分类号: S513.035.3

文献标识码: A

Genome-Wide Association Study of Plant Height in Temperate and Tropical Maize

GUO Zi-feng, XU Xiao-jie, LIU Xiao-gang, WANG Shan-hong

(Institute of Crop Sciences, Chinese Academy of Agricultural Science, Beijing 100081, China)

Abstract: Genome-wide association study(GWAS) was conducted on inbred maize lines with 365 inbred maize lines with rich genetic background. The results showed that plant height was highly variable in three environments, with a trend of increasing gradually from low latitude to high latitude. Plant height was significantly different in genotype, environment, genotype, and environment interaction. Genome-wide association analysis showed that significant sites associated with plant height were detected under different environments, and significant SNPS were found on chromosomes 1, 3, 5, 6, 7 and 8. Under the environment of Xinxiang and Shunyi, significant SNP was found in bin5.05, while under the environment of Sanya and Shunyi, significant SNP was found in bin6.05. The significant SNP loci of association analysis were compared to the B73 reference genome, and the 22 most likely candidate genes were obtained on the maize GDB website, which provided an important reference for exploring the genetic mechanism of maize plant height.

Key words: Maize; Plant height; Genome-wide association study

株高是影响玉米抗倒伏和生态适应性的重要农艺性状。株高是一种复杂的数量性状, 受多基因控制, 具有较高的遗传力和杂种优势^[1]。不同研究者利用不同的作图群体和作图方法对玉米的株高进行了解析。利用连锁分析已经定位了很多与株高相关

的区间, 从1~10号染色体上均检测到与株高相关的位点, 但各个位点的效应值均较小^[2]。张志明等^[3]用玉米自交系组合R15×掖478的F₂群体构建连锁图谱, 并利用146对SSR引物检测到8个与株高相关的QTL, 分别位于第2、3、4、5、8号染色体上。郑德波等^[4]以玉米两个F₂群体为作图群体, 用复合区间作图法对株高进行QTL定位, 在7号染色体上定位到2个主效QTL。何文昭等^[5]利用郑58和HD568及其组配得到的220个重组自交系, 在3个环境下, 对株高和穗位高的遗传特性进行解析, 发现株高受环境影响显著, 广义遗传力较大。全基因组关联分析(Genome-wide association study, GWAS)利用覆盖整

录用日期: 2019-03-20

基金项目: 国家重点研发计划项目(2016YFD0101201, 2017YFD0101201)、中国农业科学院农业科技创新工程行动(CAAS-XTCX2016009)

作者简介: 郭子锋(1986-), 助理研究员, 主要从事玉米遗传育种研究。E-mail: jzyyn@163.com

个基因组的高密度 SNP 进行基因型鉴定,在植物复杂数量性状的剖析中发挥了重要作用。玉米具有极端的遗传和表型多样性,与其他物种相比,具有更快速的连锁不平衡(linkage disequilibrium, LD)衰变,适合进行 GWAS 分析。利用 GWAS 已经定位了一些与玉米株高相关的 SNP。刘坤等^[6]利用 284 份关联群体和 56 万个 SNP 标记进行全基因组关联分析,在第 4、5、8 号染色体上检测到株高相关的 SNP。李凯等^[7]以 360 份具有广泛遗传变异的玉米自交系为材料,在 4 个环境下利用 44 569 个 SNP 标记对玉米株高进行全基因组关联分析,在第 5 号染色体上检测到 6 个与株高显著关联的 SNP 位点,解释表型变异的 14.26%。本文利用自主研发的适应温热带玉米自交系的 55K 芯片^[8],采用来源广泛的国内外玉米自交系,在不同纬度地区进行表型鉴定,对玉米株高的遗传性状进行解析。

1 材料与方法

1.1 供试材料与表型鉴定

关联群体包含 365 份玉米自交系,其中,温带材料 229 份,包含黄早四、Mo17、综 31、B73、PHW86、掖 478 和齐 319;热带材料 136 份,包含 CIMMYT 的杂种优势群 A(CML9)和杂种优势群 B(CML21),材料来源于中国、美国和国际玉米小麦改良中心(CIMMYT)。材料于 2014 年种植在海南三亚,2016 年种植在河南新乡,2017 年种植在北京顺义。随机区组排列,两次重复,种植行长 4 m,株距 25 cm,每行 17 株。在玉米自交系株高稳定后,随机选取 10 株,采用塔尺进行测量,测量从根部到雄穗顶端的长度。

1.2 DNA 提取与基因型鉴定

每个自交系取 8 粒种子在育苗盘中进行发苗,在 3 叶 1 心期从不同植株取 3 片叶进行混合,采用 CTAB 法高通量提取基因组 DNA,以 0.5%的琼脂糖凝胶电泳和 nanodrop 对 DNA 的质量和浓度进行检测。检测合格的 DNA 用自主研发的 55K 芯片进行基因型鉴定^[8]。

1.3 表型数据分析

利用 Excel 对数据进行初步整理,利用 SAS4.0 软件对表型数据进行统计分析,分析不同环境下株高的均值、方差、标准差、偏度和峰度等。利用 QTL IciMapping 软件进行单因素多环境下的联合方差分析。

1.4 全基因组关联分析

群体结构估算和亲缘关系计算参考之前的研究^[9]。结合株高表型数据和 55K 芯片数据,采用 TASSEL5.0 中的混合线性模型(MLM)进行全基因组关联分析。55K 芯片共含有 55 230 个 SNP 标记,在进行全基因组关联分析之前,对 SNP 进行筛选,SNP 的筛选标准为 MAF>0.05,缺失率<20%,符合要求的 SNP 数量为 39 797 个 SNP。

关联分析具体模型如下:

$$y = X\beta + Zu + \varepsilon$$

式中, y 表示观测值向量; β 是指固定效应向量;包含遗传标记和群体结构因素; u 为随机效应向量; X 和 Z 为已知的设计矩阵; ε 为随机剩余效应向量。

对关联分析显著的 SNP,基于 LD 衰减距离,利用 maizeGDB 数据库,在 200kb 窗口查找候选基因。

2 结果与分析

2.1 表型数据分析

表 1 不同环境下玉米自交系株高的表型数据统计分析

Table 1 Statistical analysis of phenotypic data of maize inbred lines in different environments

环境 Environment	范围(cm) Range	均值(cm) Mean	方差 Variance	标准差 Standard deviation	偏度 Skewness	峰度 Kurtosis	相关系数 Correlation coefficient		
							三亚 Sanya	新乡 Xinxiang	顺义 Shunyi
三亚	75.2 ~ 211.6	147.1	395.703	19.892	0.047	0.337	1.000		
新乡	113.4 ~ 288.5	203.9	853.477	29.214	0.073	-0.025	0.630**	1.000	
顺义	111.1 ~ 305.2	208.6	1 116.302	33.411	0.131	-0.017	0.550**	0.750**	1.000

注:**表示 0.01 差异显著水平。

Note: ** indicated a significant level of 0.01 difference.

365 份玉米自交系在三亚、新乡和顺义的表型统计见表 1。株高在 3 个环境下的变异非常丰富,在三亚的均值为 147.1 cm,分布在 75.2 ~ 211.6 cm;在

新乡的均值为 203.9 cm,分布在 113.4 ~ 288.5 cm;在顺义的均值为 208.6 cm,分布在 111.1 ~ 305.2 cm。株高有从低纬度(三亚)到高纬度(顺义)逐渐增加的

趋势, 方差和标准差也有增加的趋势, 说明株高随着纬度的增加变异越来越丰富。相关分析表明, 株高在3个环境下呈极显著相关性, 其中新乡和顺义之间的相关性最高, 相关系数为0.754; 三亚和顺义之

间的相关性最低, 为0.55, 但也达极显著相关性。利用D统计量进行正态分布检验, 图1结果表明, 株高在3个环境下均呈正态分布, 适合进行全基因组关联分析。

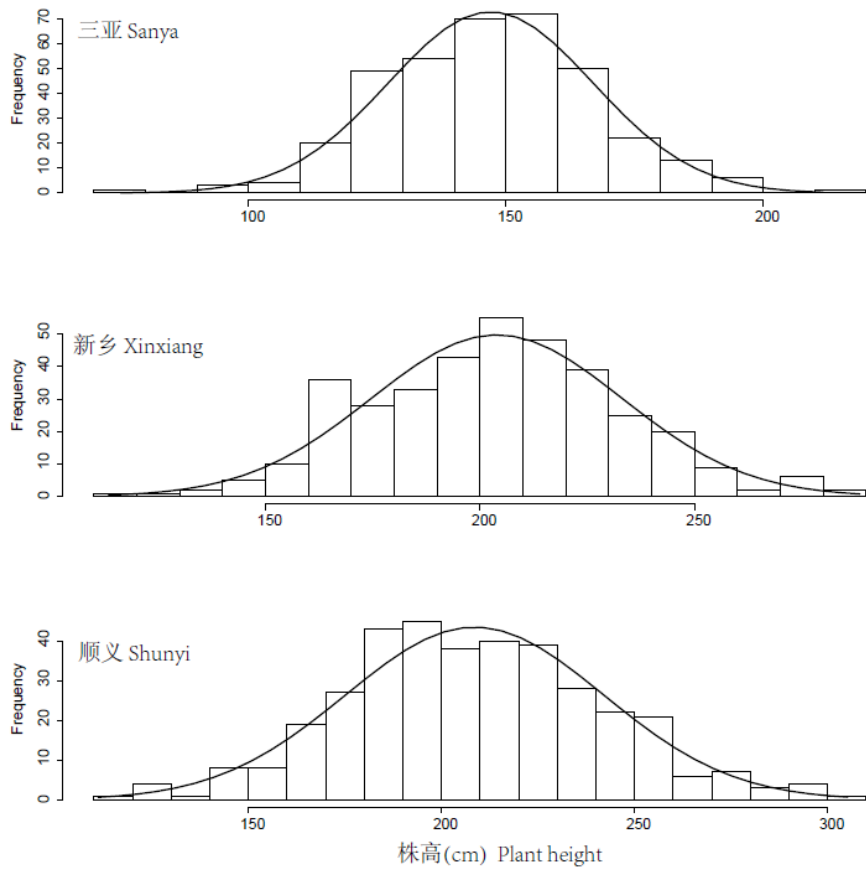


图1 365份玉米自交系株高的分布直方图
Fig.1 Distribution histogram of plant height of 365 maize inbred lines

2.2 方差分析

利用QTL IciMapping 软件对不同环境下株高进行方差分析, 检测环境、基因型、环境与基因型互作对玉米株高的影响。从表2可以看出, 3个环境下, 株高在基因型之间都表现出极显著的差异, 不同环境之间株高也达到了极显著的差异, 环境与基因型

的互作也达极显著差异, 说明株高受基因型、环境、环境与基因型的互作影响非常大。株高在三亚、新乡和顺义的广义遗传力分别为0.91、0.97和0.81, 联合遗传力为0.62, 株高的遗传力相对较高, 说明株高受遗传影响较大。

表2 3个环境下365份玉米自交系株高方差分析
Table 2 Variance analysis of plant height of 365 maize inbred lines in three environments

环境 Environment	来源 Source	自由度 Degrees of freedom	总方差 Total variance	平均方差 Mean variance	F值 F value	P值 P value
三 亚	区 组	1	12 827.0	12 827.0	1 197.8	0.0
	基因型	364	288 863.1	793.6	74.1	0.0
	广义遗传力	0.91				
新 乡	区 组	1	9 72.9	972.9	11.6	0.0
	基因型	364	623 037.8	1 711.6	20.4	0.0
	广义遗传力	0.97				

续表2 Continued 2

环境 Environment	来源 Source	自由度 Degrees of freedom	总方差 Total variance	平均方差 Mean variance	F值 F value	P值 P value
顺 义	区 组	1	524.2	524.2	2.3	0.1
	基因型	364	809 835.7	2224.8	9.6	0.0
	广义遗传力	0.81				
联 合	区组/环境	3	14 320.8	4 773.6	44.0	0.0
	基因型	364	1 284 611.5	3 529.2	32.5	0.0
	环 境	2	1 718 391.9	601.1	7 914.7	0.0
	基因型环境互作	728	437 572.8	601.1	5.5	0.0
	广义遗传力	0.62				

2.3 全基因组关联分析

图2为玉米株高在3个不同环境下进行全基因组关联分析的曼哈顿图,阈值设置为 1.0×10^{-4} 。在1、3、5、6、7、8号染色体上均发现了显著关联的SNP。不同环境下检测到显著SNP的数目不同,海南环境下在3、6号染色体上检测到2个和1个显著SNP,新乡环境下仅在5号染色体上检测到3个显著的SNP,顺义环境下检测到显著SNP位点最多,但比

较分散。不同环境下检测到一些位于共同区段的SNP,新乡和顺义环境下,都检测到位于bin5.05区段的显著SNP;海南和顺义环境下,都检测到位于bin6.05区段显著的SNP。不同环境下虽然都检测到一些与株高显著的SNP,但这些SNP的效应值均较小,其中,效应值最大的SNP标记是AX-90971140 ($R^2=0.076$),这说明控制株高的基因为微效多基因。

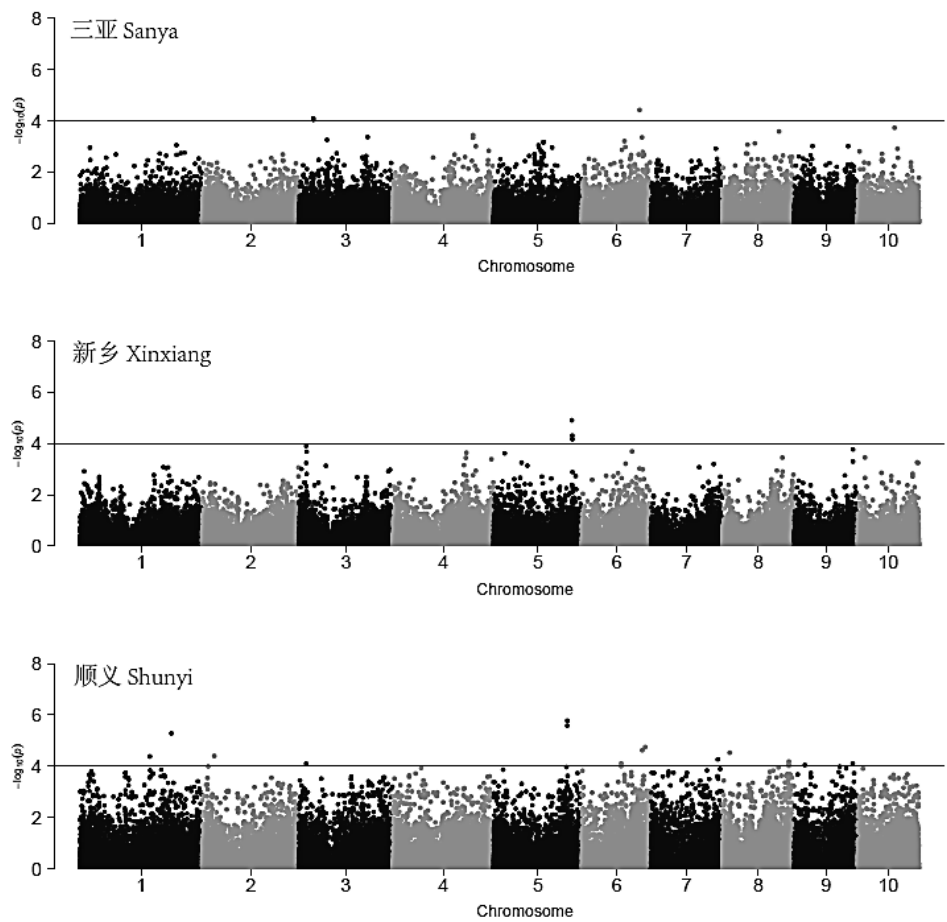


图2 3个环境下株高全基因组关联分析的曼哈顿图

Fig.2 Manhattan plot of Genome-wide association study of plant height in three environments

2.4 候选基因分析

对关联分析显著的SNP位点对到参考基因组上,在maizeGDB网站上得到22个最可能的候选基因,其中,8个有功能注释,在5、8号染色体上的候选基因最多,在新乡和顺义不同环境下都在5号染色体的相近区域发现了候选基因(表3)。其中,*GRMZM2G091119*编码一个导入蛋白亚基(importin subunit alpha-1a),*GRMZM2G061910*基因编码含有1个LEA_2结构域蛋白(LEA_2 domain-containing protein),*GRMZM2G168744*编码1个含有S1-P1结构域

的RNA核酸酶结构蛋白(S1-P1_nuclease domain-containing protein),*GRMZM2G160685*编码琥珀色脱氢酶9(succinate dehydrogenase9),*GRMZM2G061969*编码磷脂酶D家族蛋白(phospholipase D family protein),*GRMZM5G898314*编码组蛋白去乙酰化酶hdt3(histone deacetylase HDT3-like),*GRMZM2G326734*编码磷脂酶D Z(phospholipase D Z),*GRMZM2G152373*编码热休克蛋白结合蛋白(heat shock protein binding protein)。

表3 全基因组关联分析显著SNP位点及最可能候选基因

Table 3 Significant SNP loci by genome-wide association study and most likely candidate genes

环境	SNP标记	染色体	物理位置	P值	R ²	候选基因
Environment	SNP marker	Chromosome	Physical location	P value		Candidate gene
三 亚	AX-86326097	6	141 134 963	3.81E-05	0.022 1	<i>GRMZM5G875997</i>
	AX-90803248	3	35 698 699	7.98E-05	0.026 0	<i>GRMZM2G091119</i>
	AX-90803262	3	35 622 843	9.19E-05	0.063 4	<i>GRMZM2G105136</i>
新 乡	AX-86249507	5	192 342 240	1.24E-05	0.053 6	<i>GRMZM2G702141</i>
	AX-90971140	5	193 321 441	5.01E-05	0.076 0	<i>AC209208.3_FG005</i>
	AX-90971192	5	193 479 893	6.91E-05	0.064 6	<i>GRMZM2G061910</i>
顺 义	AX-123948031	5	181 197 390	1.699 2E-06	0.055 8	<i>AC205703.4_FG005</i>
	AX-90635223	5	181 193 868	2.628 5E-06	0.054 6	<i>AC205703.4_FG005</i>
	AX-90708770	1	225 032 393	5.251 0E-06	0.053 5	<i>GRMZM2G168744</i>
	AX-91018054	6	154 881 953	1.802 1E-05	0.049 0	<i>GRMZM2G160685</i>
	AX-86262068	6	147 881 469	2.389 6E-05	0.048 5	<i>GRMZM2G062452</i>
	AX-90595642	8	17 392 400	2.998 1E-05	0.047 5	<i>GRMZM2G061969</i>
	AX-90738324	2	29 060 630	3.991 9E-05	0.036 5	<i>GRMZM2G059365</i>
	AX-90587366	1	171 781 205	4.195 2E-05	0.037 8	<i>GRMZM2G072939</i>
	AX-86279252	7	164 317 839	5.433 9E-05	0.035 7	<i>GRMZM2G045287</i>
	AX-86320497	8	162 713 009	6.634 2E-05	0.035 2	<i>GRMZM5G898314</i>
	AX-91001484	6	96 053 535	0.000 07819	0.034 9	<i>GRMZM2G326734</i>
	AX-86268128	3	18 157 650	7.940 4E-05	0.045 2	<i>GRMZM5G874256</i>
	AX-91804978	9	144 079 299	8.016 9E-05	0.035 2	<i>GRMZM2G159824</i>
	AX-90627509	8	162 510 906	8.539 4E-05	0.034 8	<i>GRMZM2G001088</i>
	AX-86281499	9	26 543 917	9.051 4E-05	0.034 6	<i>GRMZM2G152373</i>
	AX-90594524	8	162 518 453	9.758 4E-05	0.034 4	<i>GRMZM2G001088</i>

3 结论与讨论

关联分析又称为连锁不平衡作图(LD mapping)或关联作图(association mapping),是一种以连锁不平衡为基础,鉴定某一群体内目标性状与遗传标记或候选基因关系的分析方法^[10]。根据扫描范围,关联分析可分为候选基因途径和全基因组途径两种,前者基于生理生化途径中重要的功能基因、QTL研究定位区域所包含的基因和近缘物种研究中效应较大的同源基因,一般涉及候选基因的功能预测;后者

基于覆盖全基因组的分子标记,通过对表型变异进行全基因组扫描来实现,寻找与变异可能关联的标记位点^[11]。关联分析的作图精度取决于群体的大小、目标等位基因效应的大小、所使用的标记的密度、标记与目标等位基因之间的LD衰减、表型和基因分型数据中的错误以及期望的结果^[12]。由于受种植成本和管理限制,群体大小需要控制在合理的范围。对于数量性状,等位基因效应一般较小,需要较高密度的标记才能检测较小效应的等位基因,而利用优良自交系或者使用SNP单倍型代替单个SNP

可以显著减少GWAS所需的标记数量^[13],因此选择合适的种质对于关联分析的成功至关重要。玉米上首次应用关联分析研究了玉米 *Dwarf8* 基因与玉米开花期和植株高度之间的关系^[14]。目前,已有多达数十种玉米相关性状应用关联分析方法进行了研究。本研究利用新开发的55K芯片,其中包含了一些热带特有的等位基因,在不同纬度进行株高表型鉴定,除在bin5.05定位到一些与前人研究结果一致的SNP位点^[15,16],还发现了一些新的影响株高的SNP位点,这些新发现的SNP多数集中在3、6号染色体上,在染色体bin6.05区段检测到在两个环境下都显著的SNP位点。株高是由多基因控制的复杂性状,每个位点的效应值较小,单个位点的选择效果非常有限,利用全基因组选择是未来研究的方向之一^[17]。

参考文献:

- [1] Yu S B, Li J X, Xu C G, et al. Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice [J]. Theoretical and Applied Genetics, 2002, 104, 619–625.
- [2] Yan J B, Tang H, Huang Y Q, et al. Dynamic QTL analysis for plant height in different developing stages in maize[J]. Chinese Science Bulletin, 2008, 48: 1959–1964.
- [3] 张志明,赵茂俊,荣廷昭,等. 玉米SSR连锁图谱构建与株高及穗位高QTL定位[J]. 作物学报, 2007, 33(2): 341–344.
Zhang Z M, Zhao M J, Rong T Z, et al. SSR linkage map construction and QTL identification for plant height and ear height in maize (*Zea mays* L.)[J]. Acta Agronomica Sinica, 2007, 33(2): 341–344. (in Chinese)
- [4] 郑德波,杨小红,李建生,等. 基于SNP标记的玉米株高及穗位高QTL定位[J]. 作物学报, 2013, 39(3): 549–556.
Zheng D B, Yang X H, Li J S, et al. QTL identification for plant height and ear height based on SNP mapping in maize(*Zea mays* L.) [J]. Acta Agronomica Sinica, 2013, 39(3): 549–556. (in Chinese)
- [5] 何文昭,王红武,胡小娇,等. 玉米株高和穗位高在不同环境下的数量遗传分析[J]. 作物杂志, 2017(3): 13–18.
He W Z, Wang H W, Hu X J, et al. Quantitative genetic research of plant height and ear height in maize under different environments [J]. Crops, 2017(3): 13–18. (in Chinese)
- [6] 刘 坤,张雪海,孙高阳,等. 玉米株型相关性状的全基因组关联分析[J]. 中国农业科学, 2018, 51(5): 821–834.
Liu K, Zhang X H, Sun G Y, et al. Genom-wide association studies of plant type traits in maize[J]. Scientia Agricultura Sinica, 2018, 51 (5): 821–834. (in Chinese)
- [7] 李 凯,张晓祥,管中荣,等. 玉米株高和穗位高的全基因组关联分析[J]. 玉米科学, 2017, 25(6): 1–7.
Li K, Zhang X X, Guan Z R, et al. Genome-wide association analysis of plant height and ear height in maize[J]. Journal of Maize Sciences, 2017, 25(6): 1–7. (in Chinese)
- [8] Xu C, Ren Y H, Jian Y Q, et al. Development of a maize 55K SNP array with improved genome coverage for molecular breeding[J]. Molecular Breeding, 2017, 37: 20.
- [9] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study[J]. Molecular Ecology, 2005, 15(8): 2611–2620.
- [10] Flint-Garcia S A, Thornsberry J M, Buckler E S. Structure of linkage disequilibrium in plant[J]. Annual Review of Plant Biology, 2003, 54: 357–374.
- [11] 于海霞,肖 静,田纪春,等. 关联分析及其在植物中的应用[J]. 基因组学与应用生物学, 2009, 28(1): 187–194.
Yu H X, Xiao J, Tian J C, et al. Association analysis and its application in plants genetics[J]. Genomics and Applied Biology, 2009, 28(1): 187–194. (in Chinese)
- [12] Gordon D, Finch S J. Factors affecting statistical power in the detection of genetic association[J]. The Journal of Clinical Investigation, 2005, 115: 1408–1418.
- [13] Yan J B, Warburton M L. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement[J]. Crop Science, 2011, 5: 433–449.
- [14] Thornsberry J M, Goodman M M, Doebley J, et al. Dwarf8 polymorphisms associate with variation in flowering time[J]. Nature Genetics, 2001, 28: 286–289.
- [15] Tang J H, Teng W T, Yan J B, et al. Genetic dissection of plant height by molecular markers using a population of recombinant inbred lines in maize[J]. Euphytica, 2006, 155(1–2): 117–124.
- [16] Weng J F, Xie C X, Hao Z F, et al. Genome-Wide association study identifies candidate genes that affect plant height in Chinese elite maize(*Zea mays* L.) inbred lines[J]. PLoS ONE, 2011, 6(12): 29229.
- [17] Liu X, Wang H, Wang H, et al. Factors affecting genomic selection revealed by empirical evidence in maize[J]. Crops, 2018, 6(4): 341–352.

(责任编辑:朴红梅)