

# 玉米穗轴粗和出籽率全基因组预测分析

马娟, 曹言勇, 朱卫红

(河南省农业科学院粮食作物研究所, 郑州 450002)

**摘要:** 穗轴粗和出籽率均是典型的数量性状, 在不同程度上影响玉米产量。全基因组选择整合全基因组关联分析 (GWAS, genome-wide association study) 的先验信息是提高性状预测准确性的有效方法。本研究利用 309 份玉米自交系穗轴粗和出籽率表型和基因分型测序技术获得的基因型数据, 研究基因组最佳线性无偏预测 (GBLUP, genomic best linear unbiased prediction)、贝叶斯 A (Bayes A) 和再生核希尔伯特空间 (RKHS, reproducing kernel Hilbert space) 模型对 2 种 GWAS 方法即固定和随机模型交替概率统一 (FarmCPU, fixed and random model circulating probability unification) 和压缩混合线性模型 (CMLM, compressed mixed linear model) 衍生的不同数量标记集、随机选择标记集和所有标记对预测准确性的影响。对于 2 个性状 FarmCPU 和 CMLM 衍生标记集, 3 个预测模型间的预测准确性差异较小, 差值变异范围介于 0~0.03。对于随机标记集, 相比其他 2 个模型的预测准确性, RKHS 对穗轴粗可提高 3.57%~15.91%, 而 3 个预测模型对出籽率具有相似的预测效果。除了 50 和 100 个标记, 3 个模型利用 CMLM 衍生标记对 2 个性状的预测效果均优于 FarmCPU。相比随机标记集, 穗轴粗 GWAS 衍生标记的预测准确性可提高 15.52%~88.37%; 出籽率利用衍生标记可提高 1~5.89 倍。所有衍生标记集的预测准确性均高于所有标记。这些结果均表明, 全基因组选择整合 GWAS 衍生标记有利于提高穗轴粗和出籽率的预测准确性。

**关键词:** 全基因组关联分析衍生标记; 全基因组预测; 基因组最佳线性无偏预测; 再生核希尔伯特空间; 贝叶斯 A

## Genome-Wide Prediction Analysis for Ear Cob Diameter and Kernel Ratio in Maize

MA Juan, CAO Yan-yong, ZHU Wei-hong

(Institute of Cereal Crops, Henan Academy of Agricultural Sciences, Zhengzhou 450002)

**Abstract:** Maize ear cob diameter (CD) and kernel ratio (KR) are controlled by multiple quantitative loci and both traits associate with the yield production. The genomic selection in conjugation with information identified by genome-wide association study (GWAS) is an effective method to improve the prediction accuracy. By taking use of the phenotypic datasets of CD and KR and the genotypic data derived from genotyping-by-sequencing in 309 maize inbred lines, here we investigated the genomic prediction accuracy using three GS models (genomic best linear unbiased prediction, GBLUP, Bayes A, reproducing kernel Hilbert space, RKHS) and different marker subsets (GWAS-derived markers: fixed and random model circulating probability unification, FarmCPU, compressed mixed linear model, CMLM, randomly selected markers, and all markers). By taking use of FarmCPU-and CMLM-derived markers at both traits, only slight difference (0-0.03) on the prediction accuracy using three prediction models was observed. For random markers, GS using RKHS model represented higher prediction accuracy of CD (3.57%-15.91%) if compared to two other models, whereas no difference for KR was detected. Except for 50 and 100 markers, the prediction accuracy of CMLM-derived marker using three models were higher than that of FarmCPU-derived markers. Compared to random markers,

收稿日期: 2021-04-05 修回日期: 2021-06-03 网络出版日期: 2021-06-05

URL: <http://doi.org/10.13430/j.cnki.jpgr.20210405002>

第一作者研究方向为数据分析, E-mail: majuanjuan85@126.com

基金项目: 河南省科技攻关项目 (192102110008, 212102110279); 河南省农业科学院优秀青年基金 (2020YQ04)

**Foundation projects:** Science and Technology Project of Henan Province (192102110008, 212102110279), Science-Technology Foundation for Outstanding Young Scientists of Henan Academy of Agricultural Sciences (2020YQ04)

GWAS-derived markers were able to increase the prediction accuracy (15.52%-88.37% for CD, 1-5.89 fold for KR). The prediction accuracy by deployment of GWAS-derived marker subsets was higher than that of all markers. Collectively, these results indicated that genomic selection using GWAS-derived markers could improve the prediction accuracy of CD and KR in maize.

**Key words:** markers derived from genome-wide association study; genome-wide prediction; genomic best linear unbiased prediction; reproducing kernel Hilbert space; Bayes A

玉米 (*Zea mays* L.) 穗轴粗和出籽率是典型的数量性状, 主要受微效多基因控制。目前, 利用全基因组关联分析 (GWAS, genome-wide association study) 和传统定位方法已鉴定到大量控制出籽率和穗轴粗的关键位点<sup>[1-4]</sup>。分子标记辅助选择可以对遗传效应较大的位点进行有效选择, 已成功用于对玉米丝黑穗病抗性、支链淀粉和赖氨酸含量等性状的主效位点聚合改良<sup>[5-6]</sup>。但该方法依赖于定位结果的准确性和位点遗传效应的高低, 限制了其在育种中的应用。全基因组选择 (GS, genomic selection) 能够对所有遗传变异和遗传效应进行检测和估计, 因此可以有效地利用微效位点。相比分子标记辅助选择, GS 可以提高玉米散粉吐丝间隔、散粉期和吐丝期的预测准确性<sup>[7]</sup>。在产量和秸秆产量的遗传进度上, GS 选育的玉米株系比分子标记轮回选择的高 14%~50%<sup>[8]</sup>。而且, GS 选择株系的产量测交表现比传统系谱选择方法选择的提高 7.3%<sup>[8]</sup>。在相同的选择增益下, GS 比传统表型选择方法降低 32% 的花费<sup>[9]</sup>。

预测模型和标记数量是影响 GS 预测准确性的重要因素。基因组最佳线性无偏预测 (GBLUP, genomic best linear unbiased prediction)、岭回归最佳线性无偏预测 (RR-BLUP, ridge regression BLUP)、贝叶斯方法 (Bayes)、半参数模型如再生核希尔伯特空间 (RKHS, reproducing kernel Hilbert space) 和非参数模型神经网络和支持向量机等是目前常用的预测模型<sup>[10-14]</sup>。GBLUP 和 Bayes 均基于混合线性模型<sup>[10, 12]</sup>。前者假定随机效应方差相等, 多利用 van Raden<sup>[15]</sup> 提出的基因组关系构建 G 矩阵计算基因组估计育种值。Bayes 模型中标记效应方差具有不同的先验分布假设<sup>[10, 12]</sup>。其中, Bayes A 模型假定每个标记都有效且服从正态分布, 但每个标记方差不同, 标记效应方差服从标准的逆卡方分布<sup>[10]</sup>。RKHS 是 Gianola 等<sup>[11]</sup> 提出的一种结合混合模型和核函数拟合标记间上位性效应的半参数模型。在此基础上, de los Campos 等<sup>[16-17]</sup> 提出了 RKHS 单核模型和多核模型。不同模型对标记效应

分布的假设不同导致了不同模型间预测准确性的差异。由于性状遗传结构复杂多样以及不同群体类型的限制, 目前还没有一种模型能够适用于所有情况<sup>[18]</sup>。多数研究表明, 分子标记个数的增加可以提高预测的准确性<sup>[19-21]</sup>。相比随机选择的标记, 利用 GWAS 衍生标记可以有效地提高性状的预测准确性<sup>[22-24]</sup>, 但这些研究均是基于一种 GWAS 方法的结果。不同 GWAS 方法衍生标记对预测准确性的研究, 目前尚缺少报导。在玉米 GS 研究中, 预测也多针对产量、穗行数、株型、抗旱和抗病等性状<sup>[12, 22-25]</sup>, 对穗轴粗和出籽率缺乏研究。因此, 本研究以 309 份玉米自交系穗轴粗和出籽率的表型和简化基因组测序获得的基因型数据为研究对象, 利用 GBLUP、Bayes A 和 RKHS 模型对 2 种 GWAS 方法即固定和随机模型交替概率统一 (FarmCPU, fixed and random model circulating probability unification) 和压缩混合线性模型 (CMLM, compressed mixed linear model) 衍生的不同数量标记集、随机选择标记集和所有标记进行预测, 研究 GS 整合 GWAS 衍生标记用于预测穗轴粗和出籽率的效果。

## 1 材料与方法

### 1.1 试验材料

本研究利用的 309 份自交系主要包括黄淮海骨干自交系、中国核心种质和美国保护期过期的种质等<sup>[1-2]</sup>。2017 年, 该群体种植在河南周口郸城、河南新乡原阳、河南商丘虞城和海南三亚。2019 年, 该群体种植在原阳。采用随机区组试验设计, 2 行区, 3 个重复。每个小区, 选择授粉较好的 3 个果穗, 测量穗轴粗、单穗重和单穗粒重, 计算出籽率。利用 GBS 简化测序技术获取 309 份材料的基因型, 具体实施方法见文献<sup>[1-2]</sup>。以最小等位基因频率大于 0.05, 杂合率和缺失率均小于 0.10 为筛选标准, 共获得 58129 个单核苷酸多态性位点 (SNP, single nucleotide polymorphisms) 用于关联分析和基因组预测分析。利用 R 语言将低频等位基因型编码

为2,高频等位基因型编码为0,杂合基因型编码为1。同时,根据已知基因型的基因型频率,对缺失的SNP进行随机插补。利用QTL IciMapping v4.0软件的AOV功能<sup>[26]</sup>对不同环境表型计算最佳线性无偏估计值(BLUE, best linear unbiased estimate)作为本研究穗轴粗和出籽率的表型。

## 1.2 试验方法

利用R语言GAPIT包的FarmCPU和CMLM方法计算标记和性状的关联性即 $-\log_{10}(P)$ 。2个模型的群体结构Q值和亲缘系数K值计算方法同文献<sup>[1-2]</sup>。将 $-\log_{10}(P)$ 从大到小排序,选择50、100、500、1000、2000、5000、8000和10000个SNP用于穗轴粗和出籽率基因组预测分析。同时,随机选择等数量的标记和所有标记(58129)用于预测分析。利用GBLUP、Bayes A和RKHS模型进行预测。RKHS采用的是多核模型,其中宽带参数设置为1/5M、1/M和5/M, M为欧氏距离的均方<sup>[27]</sup>。3个模型均利用R语言BGLR包计算,迭代的次数均为12000,预烧为3000次,其他参数按默认设置<sup>[27]</sup>。采用5倍交叉验证方案,将研究群体分成训练集和验证集,重复100次。利用验证集100次预测的表型值和BLUE值间相关系数均值作为GS的预测准确性( $r_{MG}$ )。A相比B预测准确性提高百分比的计算公式为 $\frac{(r_{MGA} - r_{MGB})}{r_{MGB}} \times 100\%$ 。利

用R语言stats包的aov分别对不同预测模型间、不同衍生标记集间、不同随机选择标记集以及不同标记选择方案间准确性进行方差分析。利用R语言agricolae包的duncan.test进行 $P < 0.05$ 水平多重比较检验。

## 2 结果与分析

### 2.1 不同GS模型预测准确性的比较

利用GBLUP、Bayes A和RKHS模型比较CMLM和FarmCPU衍生的不同标记集、随机选择标记集和所有标记的预测准确性(图1、图2)。对于所有标记集,穗轴粗的预测准确性均高于出籽率。这可能是因为穗轴粗的遗传力(0.77)高于出籽率(0.51)<sup>[1-2]</sup>。对于CMLM衍生的8个标记集,3个预测模型对穗轴粗的预测准确性基本相同,差值变异范围介于0~0.01(图1A)。对于穗轴粗FarmCPU衍生的标记,3个预测模型具有相似的预测效果,差值变异范围介于0~0.02(图1C)。对于出籽率衍生标记集,3个模型间准确性差异较小,差

值变异范围介于0~0.03(图1B、D)。穗轴粗不同衍生标记间准确性的差值介于0.11~0.17,低于出籽率不同衍生标记间准确性的差值(0.22~0.29)。对于穗轴粗随机选择的标记集,RKHS相比GBLUP和Bayes A可提高3.57%~15.91%(图1E)。对于出籽率,除了50个随机选择标记外,3个模型间准确性差异不显著(图1F)。

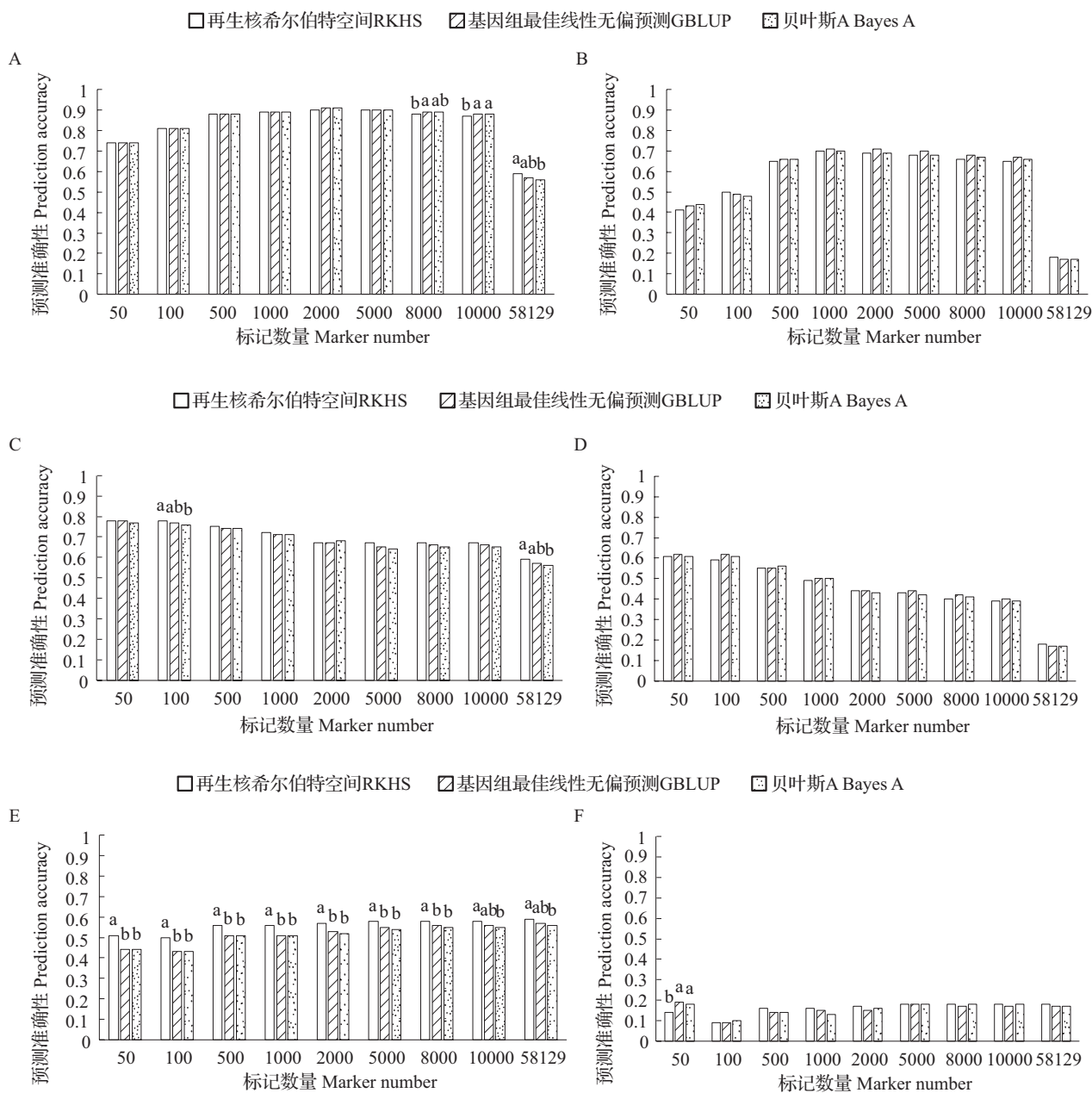
### 2.2 不同GWAS方法衍生标记的预测准确性比较

在穗轴粗和出籽率的3个模型中,FarmCPU衍生标记和CMLM衍生标记间的准确性差异达到显著水平( $t$ 测验, $P < 0.05$ ,详见<http://doi.org/10.13430/j.cnki.jpgr.20200405002>,附图1)。FarmCPU衍生的50个标记对穗轴粗的预测效果高于CMLM模型,可提高4.05%~5.41%(图2、详见<http://doi.org/10.13430/j.cnki.jpgr.20200405002>,附图1)。相比FarmCPU,3个模型利用100~10000个CMLM衍生标记对穗轴粗的预测准确性可提高3.85%~40.63%。FarmCPU衍生的50个和100个标记对出籽率的预测效果优于CMLM,可提高18%~48.78%。对于出籽率其他标记集,3个模型CMLM衍生标记的预测效果均明显高于FarmCPU,可提高17.86%~69.23%(图2、详见<http://doi.org/10.13430/j.cnki.jpgr.20200405002>,附图1)。2个性状不同数量衍生标记间的预测准确性存在显著差异(图2)。3个模型利用CMLM衍生标记集的预测准确性均随着标记个数的增加表现为先增后降趋势,穗轴粗的预测准确性在2000个衍生标记处达到峰值,为0.90~0.91,而出籽率在1000个衍生标记处准确性最高,为0.70~0.71(图2)。3个模型利用FarmCPU衍生标记对穗轴粗和出籽率的预测准确性均在50个标记最高,分别为0.77~0.78和0.61~0.62,然后随着标记个数的增加逐渐降低(图2)。

### 2.3 GWAS衍生标记和随机选择标记预测准确性的比较

在3个模型中,相比随机选择标记集,利用FarmCPU和CMLM衍生标记集可显著提高穗轴粗和出籽率的预测准确性(图2)。相比随机标记集的预测准确性,穗轴粗GWAS衍生标记可提高15.52%~88.37%,而出籽率利用衍生标记可提高1~5.89倍。所有GWAS方法衍生标记集的预测能力均显著高于所有标记(图2)。这些研究结果均表明,全基因组预测模型中拟合GWAS衍生标记有利于提高穗轴粗和出籽率的预测准确性。





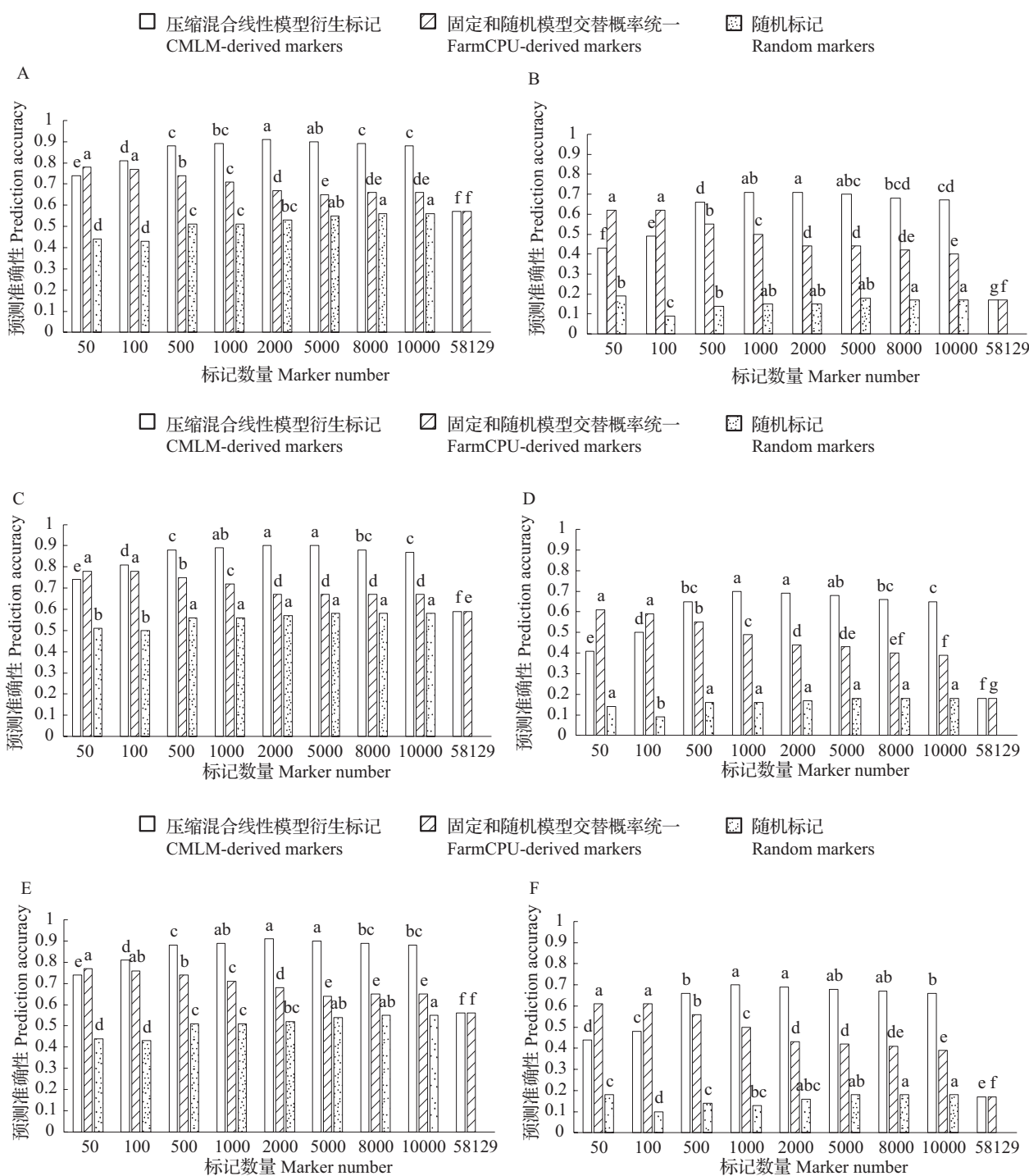
A、C、E: 分别为穗轴粗 CMLM 衍生标记、FarmCPU 衍生标记和随机选择标记; B、D、F: 分别为出籽率 CMLM 衍生标记、FarmCPU 衍生标记和随机标记。不同字母表示预测模型准确性均值在  $P < 0.05$  水平存在差异

A, C, E: represent CMLM-derived markers, FarmCPU-derived markers and random markers for ear cob diameter, respectively.

B, D, F: represent CMLM-derived markers, FarmCPU-derived markers and random markers for kernel ratio, respectively.

Different letters represent mean accuracies among prediction models are different at  $P < 0.05$  level

图 1 穗轴粗和出籽率 3 个预测模型的预测准确性  
Fig.1 Prediction accuracy of three prediction models for ear cob diameter and kernel ratio



A、C、E：分别表示穗轴粗 GBLUP、RKHS 和 Bayes A 模型；B、D、F：分别表示出籽率 GBLUP、RKHS 和 Bayes A 模型；  
不同字母表示不同标记集间准确性均值在  $P < 0.05$  水平存在差异

A, C, E: represent GBLUP, RKHS and Bayes A models for ear cob diameter, respectively.

B, D, F: represent GBLUP, RKHS and Bayes A models for kernel ratio, respectively.

Different letters represent mean accuracies between different marker numbers are different at  $P < 0.05$  level

图2 穗轴粗和出籽率不同标记数量的预测准确性

Fig.2 Prediction accuracy of different marker numbers for ear cob diameter and kernel ratio

### 3 讨论

GWAS 是全基因组范围鉴定性状遗传变异的一种定位方法。而 GS 是通过已知基因型和表型信息,建立模型进行预测,从而对研究材料进行有效预测和选择的一种分子育种方法。虽然 GWAS 和 GS 用来解析遗传性状的不同方面,但两者具有相同的数据结构即基因型和表型。因此, GWAS 挖掘的先验信息可以整合到 GS 模型中,用来对复杂数量性状进行遗传改良。

本研究将不同 GWAS 方法挖掘的先验信息,整合到 GBLUP、Bayes A 和 RKHS 模型中,对穗轴粗和出籽率进行了预测研究。多数衍生标记集中,3 个模型表现出相似的基因组预测能力。在其他作物、研究群体和性状中,3 个模型表现出不同程度的差别。RKHS 对小麦抽穗期和产量的预测准确性高于 Bayes A<sup>[28]</sup>。在玉米育种群体中,相比 GBLUP, RKHS 对于干旱环境产量和散粉丝间隔的预测略有优势<sup>[29]</sup>。对于玉米高油相关性状, RKHS 的基因组预测能力低于 Bayes A,且两者的预测能力均低于 RR-BLUP 模型<sup>[30]</sup>。理论研究表明, GBLUP 和 RR-BLUP 是等价的<sup>[31]</sup>。在玉米 F<sub>2:3</sub> 群体中, RKHS 对产量和株高的预测准确性优于 Bayes A 和 GBLUP<sup>[12]</sup>。大多数情况下,相比其他 2 个模型, RKHS 利用随机选择标记和所有标记表现出预测优势。RKHS 能够有效地捕捉杂交群体中的非加性效应,因此可以提高预测的准确性<sup>[11]</sup>。本研究中,309 份自交系简化测序数据存在一定程度的杂合性。因此,即使是随机选择的标记, RKHS 仍具有较好的预测能力。当利用 GWAS 衍生标记时, RKHS 的优势并不明显。这说明,性状的遗传结构对预测结果的影响较大。不同衍生标记预测准确性的差异远大于不同模型间的差异也证明如此。

在出籽率和穗轴粗显著位点检测功效上, FarmCPU 优于 CMLM<sup>[1-2]</sup>。对于衍生标记的预测准确性,2 种方法随着标记个数的变化表现出较大差异。2 个性状中, FarmCPU 均利用 50 个衍生标记获得最高的预测准确性,且高于 CMLM 方法。对于 CMLM 衍生标记,穗轴粗和出籽率的预测准确性均随着标记的数量增加呈现先增后降趋势。除了出籽率 50 个和 100 个衍生标记外,500~10000 个 CMLM 衍生标记的准确性高于 FarmCPU。CMLM 方法将研究个体进行聚类分组,利用小组间的亲

缘关系即 K 矩阵代替个体间的 K 矩阵<sup>[32]</sup>。而 FarmCPU 通过交替利用一个固定效应模型和一个随机效应模型来解决亲缘关系、群体结构和标记效应间的混杂问题,其具体原理为利用固定效应模型检测所有遗传标记,利用可能关联位点作为协变量控制假阳性,随机效应模型中采用 bin 方法根据标记 P 值和位置信息预测优化可能关联位点<sup>[33]</sup>。2 种 GWAS 方法利用不同的策略来提高位点检测的统计功效和解决模型中的混杂问题,是可能导致其衍生标记预测准确性的差别的原因。即使多数标记集中 FarmCPU 的预测准确性要低于 CMLM 方法,但其对出籽率和穗轴粗也表现出较高的预测能力,最高可达 0.62 和 0.90,可以满足预测的要求。

在 3 个模型中,相比随机选择的标记集,利用 FarmCPU 和 CMLM 衍生的标记集可以显著提高穗轴粗和出籽率的预测准确性。而且所有 GWAS 模型衍生标记集的预测效果均明显高于所有标记。这些结果均表明,全基因组选择模型中拟合 GWAS 衍生标记有利于提高性状的预测准确性。这与多数研究结果一致<sup>[18, 22-24]</sup>。由于 GWAS 衍生标记的个数远少于所有标记,因此减少了标记间的共线性问题和模型估计育种值的复杂性<sup>[18]</sup>,从而提高了预测的准确性。另一个解释是 GWAS 衍生标记可能与目标性状相关。基因组预测准确性会随着衍生标记的排序和个数的变化而产生差异。本研究 3 个预测模型中,穗轴粗和出籽率分别利用 2000 个和 1000 个 CMLM 衍生标记获得最高的预测准确性;而 FarmCPU 方法中,仅需要 50 个衍生标记即可获得最高的预测效果。刘小刚<sup>[34]</sup>研究发现,在玉米自然群体中,性状关联标记数为 1000~2000 时,预测准确性达到最高,而双亲群体中对应的标记数为 100~200。因此在利用 GS 进行预测时,有必要利用 GWAS 或连锁分析方法对性状的遗传结构进行解析,确定最佳的标记信息,来提高基因组估计育种值的准确性。对于遗传力较低的性状,预测准确性即使增加 1% 也会大大提高遗传进度<sup>[35]</sup>。因此, GWAS 和 GS 的整合是改良复杂性状的有效方法,可以用来预测表型,提高育种效率。

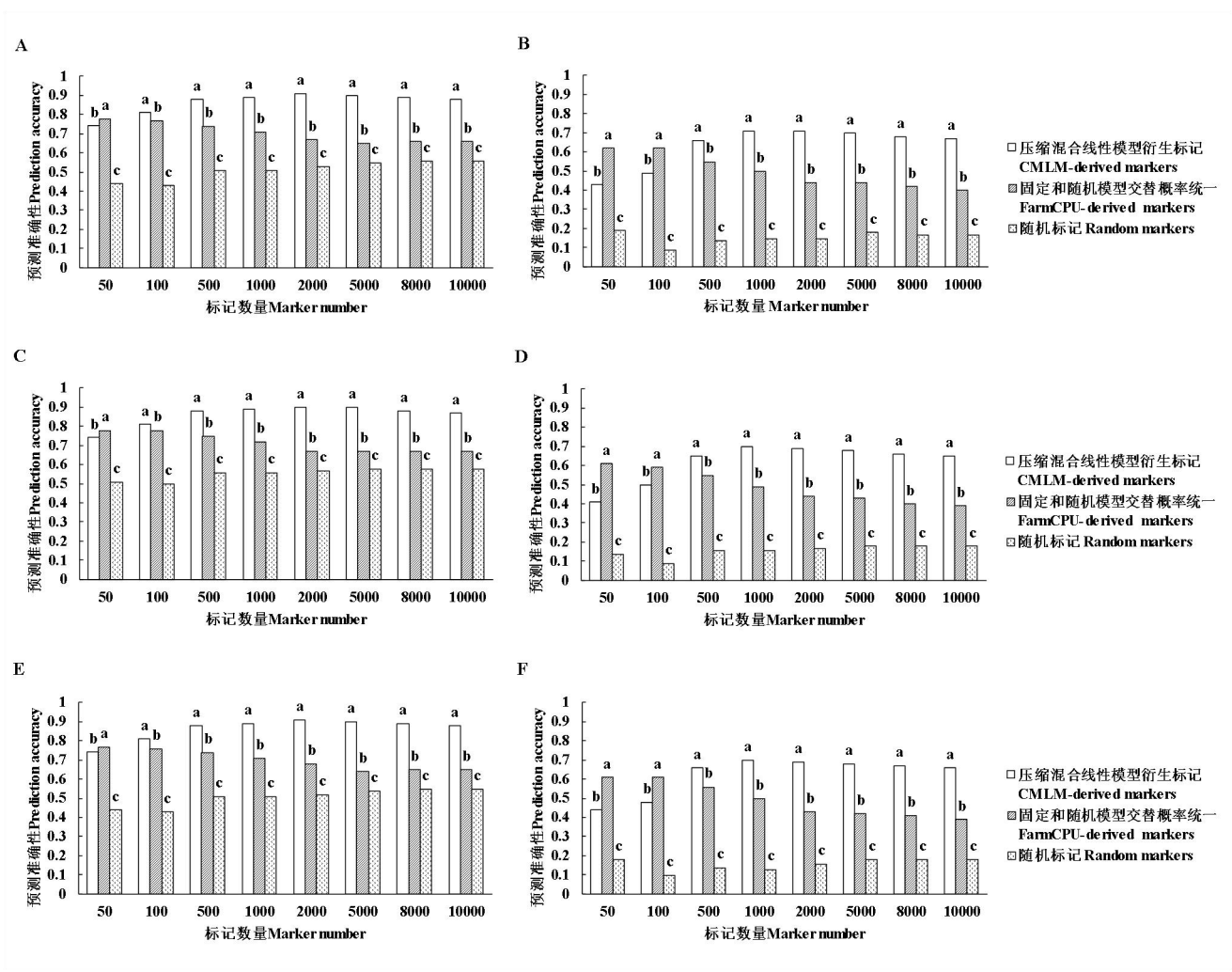
#### 参考文献

- [1] 马娟,曹言勇,李会勇. 玉米穗轴粗全基因组关联分析. 作物学报, 2021, 47(7): 1228-1238

- Ma J, Cao Y Y, Li H Y. Genome-wide association study of ear cob diameter in maize. *Acta Agronomica Sinica*, 2021, 47(7): 1228-1238
- [2] 马娟, 王利锋, 曹言勇, 李会勇. 玉米出籽率全基因组关联分析. *植物遗传资源学报*, 2021, 22(2): 448-454
- Ma J, Wang L F, Cao Y Y, Li H Y. Genome-wide association studies for kernel ratio in maize. *Journal of Plant Genetic Resources*, 2021, 22(2): 448-454
- [3] Su C, Wang W, Gong S, Zuo J, Li S, Xu S. High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Frontiers in Plant Science*, 2017, 8: 706-719
- [4] 赵小强, 任斌, 彭云玲, 徐明霞, 方鹏, 庄泽龙, 张金文, 曾文静, 高巧红, 丁永福, 陈奋奇. 8 种水旱环境下 2 个玉米群体穗部性状 QTL 间的上位性及环境互作效应分析. *作物学报*, 2019, 45(6): 856-871
- Zhao X Q, Ren B, Peng Y L, Xu M X, Fang P, Zhuang Z L, Zhang J W, Zeng W J, Gao Q H, Ding Y F, Chen F Q. Epistatic and QTL  $\times$  environment interaction effects for ear related traits in two maize (*Zea mays*) populations under eight watering environments. *Acta Agronomica Sinica*, 2019, 45(6): 856-871
- [5] 邸宏, 宫程旭, 孙培元, 刘显君, 王振华. 分子标记辅助选择改良玉米自交系昌 7-2 的丝黑穗病抗性. *玉米科学*, 2021, 29(1): 20-25
- Di H, Gong C X, Sun P Y, Liu X J, Wang Z H. Improving head smut resistance in maize inbred line Chang7-2 by marker assisted selection. *Journal of Maize Sciences*, 2021, 29(1): 20-25
- [6] Yang L, Wang W, Yang W, Wang M. Marker-assisted selection for pyramiding the waxy and opaque-16 genes in maize using cross and backcross schemes. *Molecular Breeding*, 2013, 31(4): 767-775
- [7] Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna B, Banziger M, Crossa J. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science*, 2015, 55(1): 154-163
- [8] Massman J M, Jung H J G, Bernardo R. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Science*, 2013, 53(1): 58-66
- [9] Beyene Y, Gowda M, Olsen M, Robbins K R, Pérez-Rodríguez P, Alvarado G, Dreher K, Gao S Y, Mugo S, Prasanna B M, Crossa J. Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. *Frontiers in Plant Science*, 2019, 10: 1502
- [10] Meuwissen T H E, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157: 1819-1829
- [11] Gianola D, van Kaam J B. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, 2008, 178(4): 2289-2303
- [12] Liu X, Wang H, Wang H, Guo Z, Xu X, Liu J, Wang S, Li W X, Zou C, Prasanna B M, Olsen M S, Huang C, Xu Y. Factors affecting genomic selection revealed by empirical evidence in maize. *The Crop Journal*, 2018, 6(4): 341-352
- [13] Rahaman M M, Ahsan M A, Chen M. Data-mining techniques for image-based plant phenotypic traits identification and classification. *Science Reports*, 2019, 9(1): 19526
- [14] Li G L, Dong Y, Zhao Y S, Tian X K, Würschum T, Xue J Q, Chen S J, Reif J C, Xu S T, Liu W X. Genome-wide prediction in a hybrid maize population adapted to Northwest China. *The Crop Journal*, 2020, 8(5): 830-842
- [15] van Raden P M. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 2008, 91(11): 4414-4423
- [16] de los Campos G, Gianola D, Rosa G J M. Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. *Journal of Animal Science*, 2009, 87(6): 1883-1887
- [17] de los Campos G, Gianola D, Rosa G J M, Weigel K A, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, 2010, 92: 295-308
- [18] Ali M, Zhang Y, Rasheed A, Wang J K, Zhang L Y. Genomic prediction for grain yield and yield-related traits in Chinese winter wheat. *International Journal of Molecular Sciences*, 2020, 21(4): 1342-1359
- [19] Bernardo R, Yu J M. Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, 2007, 47(3): 1082-1090
- [20] Calus M P, Veerkamp R F. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics*, 2007, 124(6): 362-368
- [21] Solberg T R, Sonesson A K, Woolliams J A, Meuwissen T H. Genomic selection using different marker types and densities. *Journal of Animal Science*, 2008, 86(10): 2447-2454
- [22] Liu X G, Wang H W, Hu X J, Li K, Liu Z F, Wu Y J, Huang C L. Improving genomic selection with quantitative trait loci and nonadditive effects revealed by empirical evidence in maize. *Frontiers in Plant Science*, 2019, 10: 1129
- [23] Yuan Y, Cairns J E, Babu R, Gowda M, Makumbi D, Magorokosho C, Zhang A, Liu Y, Wang N, Hao Z, San V F, Olsen M S, Prasanna B M, Lu Y, Zhang X. Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize. *Frontiers in Plant Science*, 2019, 9: 1919
- [24] An Y, Chen L, Li Y X, Li C, Shi Y, Zhang D, Li Y, Wang T. Genome-wide association studies and whole-genome prediction reveal the genetic architecture of KRN in maize. *BMC Plant Biology*, 2020, 20(1): 490
- [25] Sitonik C, Suresh L M, Beyene Y, Olsen M S, Makumbi D, Oliver K, Das B, Bright J M, Mugo S, Crossa J, Tarekegne A, Prasanna B M, Gowda M. Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theoretical and Applied Genetics*, 2019, 132(8): 2381-2399
- [26] Meng L, Li H, Zhang L, Wang J. QTL IciMapping: integrated

- software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal*, 2015, 3: 269-283
- [ 27 ] Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 2014, 198 ( 2 ): 483-495
- [ 28 ] Pérez-Rodríguez P, Gianola D, González-Camacho J M, Crossa J, Manès Y, Dreisigacker S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *Genes, Genomes, Genetics*, 2012, 2 ( 12 ): 1595-1605
- [ 29 ] Crossa J, Beyene Y, Kassa S, Pérez P, Hickey J M, Chen C, de los Campos G, Burgueño J, Windhausen V S, Buckler E, Jannink J L, Lopez Cruz M A, Babu R. Genomic prediction in maize breeding populations with genotyping-by-sequencing. *Genes, Genomes, Genetics*, 2013, 3 ( 11 ): 1903-1926
- [ 30 ] Hao Y F, Wang H W, Yang X H, Zhang H W, He C, Li D D, Li H H, Wang G Y, Wang J H, Fu J J. Genomic prediction using existing historical data contributing to selection in biparental populations: A study of kernel oil in maize. *The Plant Genome*, 2019, 12 ( 1 ): 180025
- [ 31 ] Hayes B J, Visscher P M, Goddard M E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetical Research*, 2009, 91 ( 1 ): 47-60
- [ 32 ] Zhang Z, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, Gore M A, Bradbury P J, Yu J, Arnett D K, Ordovas J M, Buckler E S. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 2010, 42 ( 4 ): 355-360
- [ 33 ] Liu X, Huang M, Fan B, Buckler E S, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, 2016, 12 ( 2 ): e1005767
- [ 34 ] 刘小刚. 玉米产量相关性状的全基因组选择. 北京: 中国农业科学院, 2018  
Liu X G. Genomic selection for yield-related traits in maize. Beijing: Chinese Academy of Agricultural Sciences, 2018
- [ 35 ] Rice B, Lipka A E. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *The Plant Genome*, 2019, 12 ( 1 ): 180052-180065





A、C、E：分别表示穗轴粗 GBLUP、RKHS 和 Bayes A 模型；B、D、F：分别表示出籽率 GBLUP、RKHS 和 Bayes A 模型。不同字母表示 CMLM 衍生标记、FarmCPU 衍生标记和随机选择标记间准确性均值在  $P < 0.05$  水平存在差异

A, C, E: represent GBLUP, RKHS, and Bayes A models for ear cob diameter, respectively. B, D, F: represent GBLUP, RKHS, and Bayes A models for kernel ratio, respectively. Different letters represent mean accuracies among CMLM-derived markers, FarmCPU-derived markers, and random markers are different at  $P < 0.05$  level

附图 1 穗轴粗和出籽率不同标记选择方案的预测准确性

Fig. S1 Prediction accuracy of different marker selection strategies for ear cob diameter and kernel ratio