

利用 ML-GWAS 解析玉米籽粒含水量相关性状的遗传基础

胡芳林¹, 赵永锋¹, 朱彦辉², 李真¹, 贾晓艳¹, 郭晋杰¹, 祝丽英¹

(¹河北农业大学农学院/国家玉米改良中心河北分中心/河北省作物种质资源重点实验室, 保定 071001; ²石家庄市农林科学研究院, 石家庄 050800)

摘要: 玉米成熟期籽粒含水量 (KMC, kernel moisture content) 是影响玉米机械化粒收的关键因素, 利用多位点全基因组关联分析 (ML-GWAS, multi-locus genome-wide association study) 挖掘与籽粒含水量相关性状显著关联的遗传位点, 解析其遗传基础, 可为适时收玉米品种的选育和遗传改良提供参考。本研究以 205 份玉米自交系为材料, 在 4 个环境下测定成熟期籽粒、苞叶与穗轴的含水量, 使用 mrMLM、FASTmrMLM、FASTmrEMMA、PLARmEB、PKWmEB 和 ISIS EM-BLASSO 6 种多位点关联分析方法结合分布于全基因组的 76492 个单核苷酸多态性位点 (SNP, single nucleotide polymorphism) 进行关联分析, 挖掘其候选基因并进行基因注释。表型分析显示, 205 份材料的籽粒含水量相关性状的变异系数在 10.47%~33.90%, 广义遗传率在 67.39%~81.24%。6 种 ML-GWAS 方法共检测到 26 个、15 个和 23 个 SNP 位点分别与籽粒含水量、苞叶含水量和穗轴含水量显著关联。其中 3 种及以上方法共同检测到 14 个显著关联 SNP 位点, 表型贡献率在 1.13%~17.21%。6 种方法中 pLARmEB 检测到的显著位点最多, FASTmrEMMA 检测到的最少。基于 3 种及以上方法共同检测到且平均 PVE≥5% 的显著 SNP 位点 Chr1_9232728、Chr4_176596174、Chr8_57716249 与 Chr5_191021635 上下游 220kb 共挖掘到 17 个候选基因, 主要富集到细胞解剖实体、新陈代谢过程和细胞进程与催化活性, 推测这些基因通过调节细胞代谢与催化活性最终影响玉米籽粒、苞叶和穗轴的含水量。

关键词: 玉米; 关联分析; 籽粒含水量; 苞叶含水量; 穗轴含水量

Utilizing ML-GWAS to Elucidate the Genetic Basis of Kernel Moisture Content Traits in Maize

HU Fanglin¹, ZHAO Yongfeng¹, ZHU Yanhui², LI Zhen¹, JIA Xiaoyan¹, GUO Jinjie¹, ZHU Liying¹

(¹College of Agronomy, Hebei Agricultural University/Hebei Sub-center of National Maize Improvement Center/ Hebei Key Laboratory of Crop

Germplasm Resources, Baoding 071001; ²Shijiazhuang Academy of Agriculture and Forestry Sciences, Shijiazhuang 050041)

Abstract: Kernel moisture content (KMC) at maturity is a crucial factor affecting the mechanical harvesting of maize. By utilizing multi-locus genome-wide association study (ML-GWAS), this study aims to explore significant genetic loci associated with KMC and provide insights into the genetic basis, which can contribute to the breeding and genetic improvement of maize varieties for appropriate harvesting. In this research, 205 maize inbred lines were used to measure the moisture content of kernels, husks, and cobs at maturity in four different environments. Six ML-GWAS methods, including mrMLM, FASTmrMLM, FASTmrEMMA, PLARmEB, PKWmEB, and ISIS EM-BLASSO, were applied in combination with 76,492 single nucleotide polymorphisms (SNPs) distributed throughout the genome for association analysis. Candidate genes were identified and gene annotation was performed. Phenotypic analysis showed that the coefficient of variation for KMC-related traits ranged from 10.47% to 33.90%, with a broad-sense heritability ranging from 67.39% to 81.24%. A total of 26, 15, and 23 SNP loci were significantly associated with kernel moisture content, husk moisture content, and cob moisture content, respectively. Among them, 14 significant SNP loci were identified by three or more methods, with a phenotypic contribution ranging from 1.13% to 17.21%. pLARmEB detected the most significant loci, while FASTmrEMMA detected the fewest. Based on the detection by three or more methods with an average PVE (Proportion of Variance

第一作者研究方向为玉米遗传育种, E-mail: lin248163264@foxmail.com

通信作者: 祝丽英, 研究方向为玉米遗传育种, E-mail: zhuliying73@163.com

基金项目: 河北省重点研发计划项目 (21326325D)

Foundation project: The Key Research and Development Project of Hebei Province (21326325D)

Explained) of $\geq 5\%$, significant Single Nucleotide Polymorphism (SNP) loci Chr1_9232728, Chr4_176596174, Chr8_57716249, and Chr5_191021635 were identified. Around these loci, within a 220kb upstream and downstream range, a total of 17 candidate genes were excavated. These genes were predominantly enriched in cellular anatomical entities, metabolic processes, and cellular processes with catalytic activity. It is hypothesized that these genes may influence the moisture content of maize kernels, husks, and cobs by modulating cellular metabolism and catalytic activity.

Key words: Maize; Association analysis; Kernel moisture content; Husk moisture content; Cob moisture content

玉米是我国种植面积最大的作物^[1], 随着劳动力的减少, 玉米生产全程机械化成为农业发展的必然趋势, 然而我国玉米生产中普遍存在收获期籽粒含水量较高, 机收籽粒易造成籽粒破碎, 严重制约了玉米籽粒机械化收获^[2]。因此, 选育成熟期籽粒含水量低的玉米品种已成为玉米育种工作的重点目标。玉米籽粒含水量是多基因控制的数量性状, 其遗传基础复杂。研究发现玉米的苞叶性状、穗部性状均会影响成熟期玉米籽粒的含水量^[3,4]。姚凌翔利用 257 份玉米自交系为材料, 基于 4.7 万个 SNP 的基因型信息, 使用混合线性模型 (MLM, mixed linear model), 探索与玉米果穗脱水相关的性状, 定位到 2 个与苞叶层数有关的显著位点 SNP7441 和 SNP25777^[5]。李露露等使用 242 份重组自交系, 测定玉米成熟后籽粒含水量以及苞叶数、苞叶长等性状, 共检测到 15 个相关 QTL^[6]。吕晨曦利用 290 份玉米自交系, 基于 13697 个高质量 SNP 标记, 使用广义线性模型 (GLM, generalized linear model) 和 MLM 模型, 定位到 9 个与成熟期籽粒含水量相关的 SNP 位点, 筛选到 4 个候选基因^[7]。党冬冬使用 522 份玉米自交系, 关联分析定位到与玉米籽粒成熟期含水量显著关联的 3 个 SNP 位点, 筛选到 7 个有功能注释的候选基因^[8]。Li 等以低含水量的 Si287 和高含水量的 JiA512 构建的 120 个双单倍体 (DH, doubled haploid) 为群体, 利用全基因组复合区间作图 (GCIM, genome-wide composite interval mapping) 和多位点随机 SNP 效应混合线性模型 (mrMLM, multi-locus random-snp-effect mixed linear model) 分别检测到 10 个 QTL 和 27 个 SNP 位点与籽粒含水量有关^[9]。

单位点关联分析显著性的判断大多采用较为严格的矫正方法, 从而导致一些效应较小的位点被淘汰。针对这一问题而开发的 ML-GWAS 分析方法, 它的优势是先用一个比 Bonferroni 略宽松的标准筛选 SNP, 然后用多位点遗传模型分析, 不需要多次校正^[10], 已被证明对数量性状核苷酸 (QTNs, quantitative trait nucleotides) 的检测具有更高的功效和准确性。Cui 等使用 6 种 ML-GWAS 方法 (mrMLM、FASTmrMLM、FASTmrEMMA、pLARmEB、pKWMEB、ISISEM-BLASSO) 对 478 份水稻种子萌发期耐盐性状进行分析, 共检测到 371 个 QTNs, 其中 56 个至少被 3 种方法同时定位到^[11]。Zhang 等利用 4 种 ML-GWAS 方法 (mrMLM、FASTmrEMMA、ISIS EMBLASSO、pLARmEB) 鉴定到 423 个与玉米抗倒相关性状的 QTNs, 并报道了该方法的可靠性和互补性^[12]。然而, 目前利用多种 ML-GWAS 方法解析玉米籽粒含水量相关性状遗传基础的研究还鲜见报道。

本研究以 205 份玉米自交系为关联群体, 使用 6 种 ML-GWAS 方法对籽粒、穗轴及苞叶含水量进行关联分析, 检测与其显著关联的 SNP 位点, 筛选低含水量的基因型, 进一步挖掘其候选基因, 为分子标记辅

助选育适机械化粒收玉米品种提供参考。

1 材料与方法

1.1 试验材料

本试验使用 205 份玉米自交系。于 2017 年和 2018 年分别在河北保定（115.47°E 38.87°N）和河北辛集（115.29°E 37.83°N）种植，田间试验均采用完全随机区组试验设计，2 行区，行长 4m，行距 0.6m，种植密度为 67500 株/hm²，2 次重复。保定播种时间为 5 月上旬，辛集播种时间为 6 月中旬。田间管理与当地大田生产管理一致。

1.2 表型测定与分析

选取行内整齐一致的单株，吐丝前和散粉前分别对雌穗和雄穗进行套袋隔离。待一个自交系有 10 株以上吐丝和散粉后挑选长势一致的植株自交授粉。调查每个自交系的成熟期，待其籽粒黑层形成时，每个自交系选取 5 个长势一致的果穗。首先剥苞叶，测定苞叶鲜重；然后脱粒，取果穗中部 100 粒称重记为籽粒鲜重；再测定穗轴鲜重；将称重后的苞叶、籽粒和穗轴在 105°C 下烘 30min 杀青，然后 80°C 下烘干至恒重，测定籽粒、苞叶和穗轴的干重^[13]。

使用 Microsoft Excel 2021 对两年两点 4 个环境下玉米籽粒、苞叶和穗轴的鲜重和干重数据进行整理并计算籽粒含水量（KMC, kernel moisture content）、苞叶含水量（HMC, husk moisture content）和穗轴含水量（CMC, cob moisture content）。计算公式如下：含水量(%)=(鲜重-干重)/鲜重×100%^[1]。

使用 IBM SPSS Statistics 26 对 4 个环境下的含水量数据进行描述统计和方差分析。使用 R 语言 (<https://cran.r-project.org/>) 平台下的 GGally 软件包计算相关系数并绘图^[14]。参照 Knapp 等的方法计算广义遗传率^[15]，计算公式如下： $H_B^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/r)$ 其中 σ_g^2 为遗传方差， σ_{ge}^2 为基因型与环境互作方差， σ_e^2 为随机误差方差，n 为环境数，r 为每个试验环境中的重复次数。使用 lme4 软件包中的混合线性模型进行最佳线性无偏估计(BLUE, best linear unbiased estimation)值的计算^[16]。

1.3 基因型测序与分析

利用 CTAB 法提取 205 份玉米自交系叶片 DNA，采用 Illumina 标准流程进行双末端 100 bp 测序，将测序所得有效数据对比到 B73RefGen_v4 基因组上^[17]。使用 plink v1.9 软件对基因型数据质控，去除缺失率>0.2 和次等位基因频率 (MAF, minor allele frequency) <0.05 的位点后共获得 76492 个 SNP 标记。过滤后的基因型数据使用 beagle_5.4 软件进行填补，生成用于 GWAS 分析的基因型文件^[18]。

1.4 全基因组关联分析

使用 R 语言平台下的 mrMLM.GUI_V4.02 软件包对 76492 个 SNPs 和各性状的 BLUE 值进行 GWAS 分析，使用软件包内置的 mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB 和 ISIS EM-BLASSO 6

种方法，采用 Q+K 模型，亲缘关系（K）由软件计算，群体结构 Q 矩阵由 admixture 软件计算，显著性阈值使用软件默认值 LOD=3，其他设置均保持默认^[19]。

1.5 候选基因挖掘

使用 plink v1.9 软件对填补后的基因型文件进行连锁不平衡(LD, linkage disequilibrium)分析，当 $r^2=0.2$ 时的衰减距离作为候选基因的筛选范围^[20]。在 maizeGDB (<https://www.maizegdb.org/>)网站的 B73RefGen_v4 参考基因组中查找候选基因。使用 NCBI(<https://www.ncbi.nlm.nih.gov/>)、GO(<http://www.geneontology.org/>)数据库网站查询相关基因的信息。根据玉米 B73 粽粒动态发育转录组数据分析粽粒含水量相关基因的表达情况^[21]。

2 结果与分析

2.1 粽粒含水量相关性状的描述统计分析

对 4 个环境下粽粒、苞叶和穗轴含水量数据进行描述统计分析（表 1）。结果显示 205 份自交系的 3 个含水量相关性状存在广泛变异，变异系数在 10.47%~33.90%。3 个性状中，苞叶含水量的变异系数最大。除 2017 辛集环境外，其它环境下的 3 个含水量性状相关数据的偏度和峰度的绝对值小于等于 1，表明这些性状在多数环境下服从正态分布。

表 1 粽粒、苞叶与穗轴含水量描述统计分析

Table 1 Descriptive statistical analysis of moisture content of kernel, husk and cob

性状 Trait	环境 Environment	范围 Range	均值±标准差 Mean±SD	偏度 Skewness	峰度 Kurtosis	变异系数(%) <i>CV</i>
粽粒含水量 (%) KMC	17BD	10.83~39.69	26.86±7.04	-0.22	-0.74	26.21
	17XJ	17.63~48.57	38.01±4.80	-1.14	2.96	12.62
	18BD	13.01~41.18	25.57±6.39	0.22	-0.56	24.98
	18XJ	15.80~46.99	35.40±5.66	-0.87	0.97	15.98
	BLUE	15.29~44.15	31.28±4.85	-0.26	0.29	15.51
苞叶含水量 (%) HMC	17BD	13.94~63.52	38.04±12.82	0.21	-1.00	33.70
	17XJ	20.71~72.58	45.35±12.85	-0.02	-1.11	28.34
	18BD	12.21~67.81	35.31±11.97	0.10	-0.69	33.90
	18XJ	16.32~68.18	41.27±12.70	0.23	-0.66	30.76
	BLUE	16.78~67.11	39.66±10.39	0.07	-0.64	26.20
穗轴含水量 (%) CMC	17BD	16.59~85.40	58.42±14.64	-0.78	0.05	25.06
	17XJ	37.26~82.14	66.00±7.23	-0.69	1.53	10.95
	18BD	15.57~78.28	54.03±13.60	-0.67	0.11	25.18
	18XJ	41.05~81.28	65.39±6.85	-0.70	0.85	10.47
	BLUE	28.57~87.32	60.87±8.46	-0.63	1.07	13.90

17BD: 2017 年保定; 17XJ: 2017 年辛集; 18BD: 2018 年保定; 18XJ: 2018 年辛集; BLUE: 最佳线性无偏估计; 下同

17BD: 2017 Baoding; 17XJ: 2017 Xinji; 18BD: 2018 Baoding; 18XJ: 2018 Xinji; KMC: kernel moisture content; HMC: husk moisture content; CMC: cob

moisture content; BLUE: best linear unbiased estimation; The same as below

2.2 粒粒含水量相关性状的方差分析

对 4 个环境下籽粒含水量、苞叶含水量和穗轴含水量的数据分别进行联合方差分析（表 2），3 个含水量性状基因型间差异、环境间差异及基因型与环境互作均达到极显著水平，籽粒含水量、苞叶含水量和穗轴含水量的广义遗传率分别为 78.06%、81.24% 和 67.39%，说明这 3 个含水量性状同时受遗传因素和环境影响，遗传因素的影响相对更大。

表 2 粒粒、苞叶和穗轴含水量的联合方差分析

Table 2 Joint analysis of variance for kernel, husk and cob moisture content

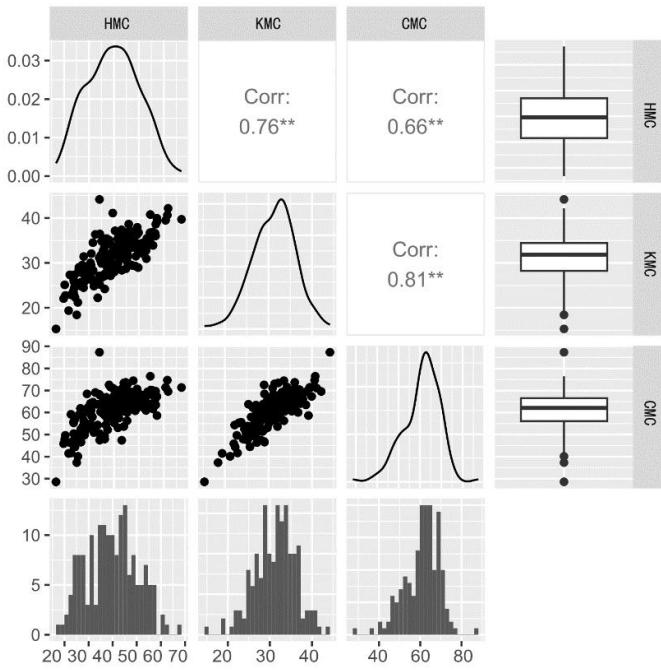
性状 Trait	F 值 F-value				广义遗传率 (%) Broad-sense heritability
	区组		环境	基因型	
	Block	Environment	Genotype	Genotype×Environment	
籽粒含水量 KMC	2.09	465.46**	6.10**	1.52**	78.06
苞叶含水量 HMC	0.70	60.99**	8.06**	1.71**	81.24
穗轴含水量 CMC	0.02	136.62**	5.77**	2.00**	67.39

*和** 分别表示在 0.05 和 0.01 水平上差异显著，下同

* and ** indicate significant differences at the levels of 0.05 and 0.01, respectively, The same as below.

2.3 粒粒含水量相关性状的相关分析

对籽粒、苞叶和穗轴含水量的 BLUE 值进行相关分析（图 1），可见这 3 个性状两两之间均呈极显著正相关，其中籽粒含水量与穗轴含水量间的相关系数最大，为 0.81。表明这 3 个含水量性状相互影响，成熟期穗轴含水量与籽粒含水量的相关最为密切。



籽粒、苞叶和穗轴含水量箱线图、频率分布直方图与曲线图、性状间散点图。频率分布曲线图纵坐标为数据所占比例，横坐标为含水量；散点图横纵坐标为相对应性状的含水量；频率分布直方图纵坐标为数量，横坐标为含水量；箱线图纵坐标为含水量

The box plots, frequency distribution histograms, curve graphs and scatter plot for KMC, HMC, and CMC are as follows: The y-axis of the frequency distribution curve graph represents the proportion of the data, while the x-axis represents the water content. In the scatter plots, both the x-axis and y-axis correspond to the water content of the respective traits. The y-axis of the frequency distribution histogram represents the quantity, and the x-axis represents the water content. In the box plot, the y-axis represents the water content

图 1 粟粒、苞叶和穗轴含水量 BLUE 值的相关分析

Fig.1 Correlation analysis of BLUE values of kernel, husk and cob moisture content

2.4 粟粒含水量相关性状的全基因组关联分析

使用 mrMLM.GUI 软件包中的 6 种方法对籽粒含水量、苞叶含水量和穗轴含水量的 BLUE 值进行多位点 GWAS 分析，检测到与籽粒含水量、苞叶含水量、穗轴含水量显著关联的 SNP 位点分别为 26 个、15 个和 23 个（图 2），分布在第 1~10 染色体上。6 种 ML-GWAS 方法 FASTmrEMMA、FASTmrMLM、SIS EM-BLASSO、mrMLM、pKWMEB 和 pLARmEB 分别检测到 8、22、22、17、23 和 24 个显著 SNP 位点。不同方法检测到的显著 SNP 位点数量有较大差异，其中 FASTmrEMMA 检测到的显著位点最少，pLARmEB 检测到的显著位点最多（图 3）。

3 种及以上方法共同检测到的 SNP 位点有 14 个（表 3），其中 6 个与籽粒含水量显著关联，表型贡献率为 1.65%~13.09%；2 个与苞叶含水量显著关联，表型贡献率为 2.21%~9.40%；6 个与穗轴含水量显著关联，表型贡献率为 1.13%~17.21%。14 个显著 SNP 位点中，分布在 1、10 染色体上各 3 个，2、4 和 6 染色体各 2 个，5 和 8 染色体上各 1 个。

对不同性状定位结果对比发现籽粒含水量与苞叶含水量有 2 个共位点 Chr1_253390359 和

Chr10_2640900, 其中 Chr10_2640900 被 3 种方法同时检测到与籽粒含水量显著关联, 被 4 种方法同时检测到与苞叶含水量显著关联。籽粒含水量与穗轴含水量有 3 个共位点 Chr2_110828978、Chr2_213305770 和 Chr6_108012658。

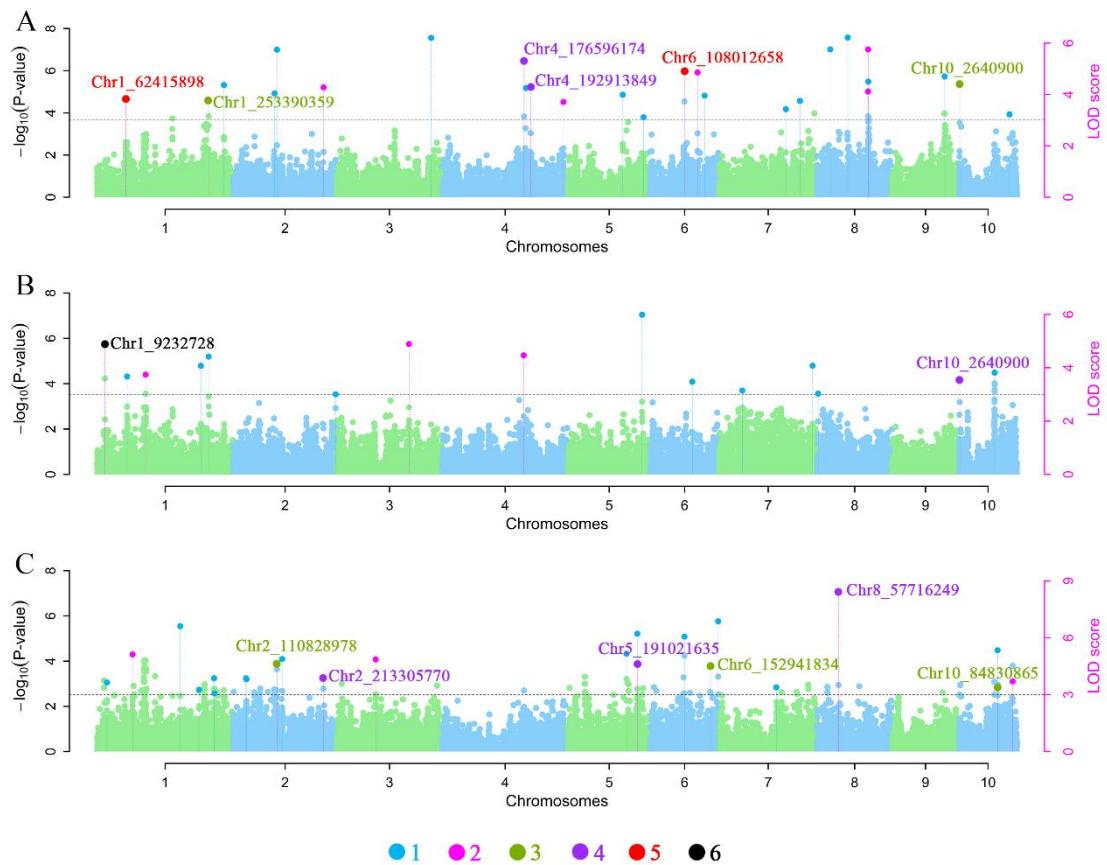


图 2 粒、苞叶和穗轴含水量曼哈顿图

Fig. 2 Manhattan plot of kernel, husk and cob moisture content traits

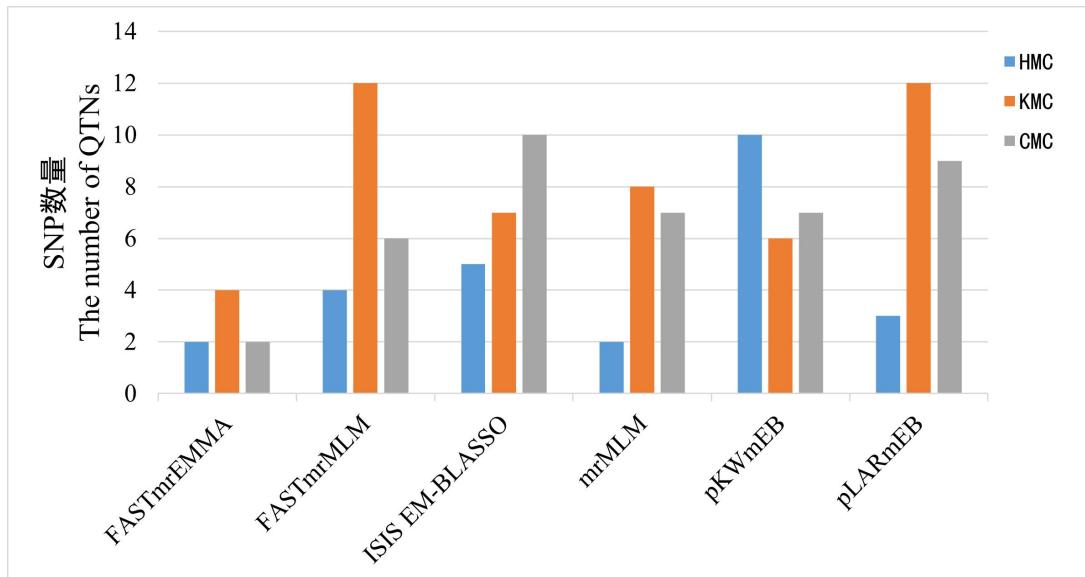


图 3 不同 ML-GWAS 方法检测到的显著位点数量

Fig. 3 Number of significant loci detected by different ML-GWAS method

表 3 3 种及以上方法共同检测到的显著 SNP

Table 3 Significant SNPs co-detected by three and more methods

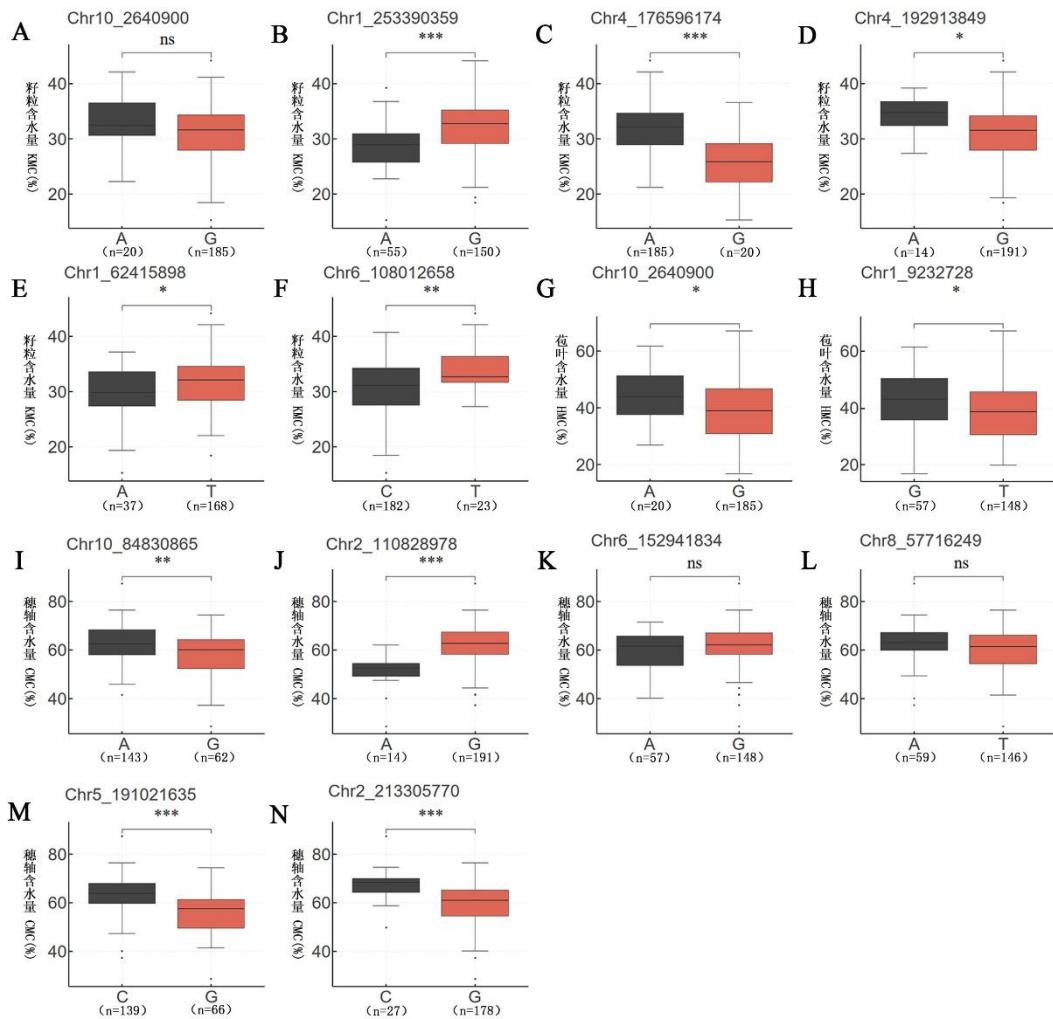
性状 Trait	关联位点 SNP	Bin	LOD	表型贡献率 (%) Phenotypic variation explained	ML-GWAS 方法 ML-GWAS Method
籽粒含水量 KMC	Chr10_2640900	10.01	3.87~4.70	2.35~4.09	1,2,5
	Chr1_253390359	1.08	3.30~6.50	3.55~5.88	2,4,5
	Chr4_176596174	4.07	3.33~8.80	3.37~13.09	1,2,3,6
	Chr4_192913849	4.08	3.34~5.91	2.02~6.25	1,2,4,5
	Chr1_62415898	1.04	3.33~9.18	1.65~4.79	1,2,3,4,5
	Chr6_108012658	6.04	3.92~5.50	3.01~6.92	1,2,3,4,5
苞叶含水量 HMC	Chr10_2640900	10.01	3.03~4.48	2.21~5.54	2,3,5,6
	Chr1_9232728	1.01	3.06~7.09	3.52~9.40	1,2,3,4,5,6
穗轴含水量 CMC	Chr10_84830865	10.03	3.33~4.54	1.13~4.14	2,4,5
	Chr2_110828978	2.05	3.69~6.80	2.30~6.41	3,5,6
	Chr6_152941834	6.05	3.03~5.95	1.90~4.86	1,2,5
	Chr8_57716249	8.03	6.64~9.46	3.58~17.21	1,2,5,6
	Chr5_191021635	5.05	4.11~5.27	1.22~11.42	1,4,5,6
	Chr2_213305770	2.08	3.70~8.15	2.49~5.34	1,2,3,5

ML-GWAS 方法: 1: mrMLM; 2: FASTmrMLM; 3: ISIS EM-BLASSO; 4: FASTmrEMMA; 5: pLARmEB; 6: pKWMnEB

ML-GWAS methods: 1: mrMLM; 2: FASTmrMLM; 3: ISIS EM-BLASSO; 4: FASTmrEMMA; 5: pLARmEB; 6: pKWMnEB

2.5 等位基因变异数分析

对 3 种及以上方法共同检测到的 14 个显著 SNP 位点进行等位基因分析（图 4）。按照位点基因型将表型数据分为两类进行 T 检验，结果显示除籽粒含水量相关的位点 Chr10_2640900 与穗轴含水量相关的位点 Chr6_152941834、Chr8_57716249 外，其余 11 个 SNP 位点的两类基因型的表型数据间差异均达显著或极显著水平。各位点含水量低的基因型即为该位点的有利等位基因。



ns 表示在 0.05 水平不显著; *, ** 和 *** 分别表示在 0.05, 0.01 和 0.001 水平上差异显著。

ns indicate not significant differences at the level of 0.05; *, ** and *** indicate significant differences at the levels of 0.05, 0.01 and 0.001, respectively

图 4 显著 SNP 不同等位基因的表型差异分析

Fig.4 Analysis of phenotypic difference in different alleles of significant SNPs

2.6 穗粒含水量相关性状的候选基因挖掘

本研究中，筛选同时由 3 种及以上方法共同检测到且平均 PVE \geq 5% 的显著 SNP 位点^[22]，用于挖掘玉米穗粒含水量相关性状的候选基因。根据 $r^2=0.2$ 时的衰减距离确定基因的筛选范围为上下游 220kb。通过综合 SNP 信息和基因注释信息，共鉴定出 17 个候选基因（表 4）。其中苞叶含水量相关位点筛选出 7 个候选基因，包含编码锌指蛋白、60S 核糖体蛋白、ABC 转运蛋白 MRP4、相互作用蛋白 RPM1、谷氨酸-tRNA 连接酶、SBT1.2 亚基蛋白酶和转录因子 MYB2。穗粒含水量相关位点筛选出 5 个候选基因，有丝氨酸蛋白、锌指结构域超家族蛋白、钙调神经磷酸酶、 β -1,2-N-乙酰葡萄糖胺基转移酶 II 和 DNA 复制许可因子 MCM5。穗轴含水量的两个位点共筛选出 5 个候选基因，包括呼吸氧化酶同源蛋白、未知功能蛋白（DUF640）、酪蛋白激酶家族蛋白、细胞壁相关受体激酶 5 和转录因子 Myb4。

GO 分类结果显示这 17 个基因中有 6 个主要富集在细胞解剖实体，4 个富集在新陈代谢过程，4 个富集在细胞进程和催化活性（图 5 A），这些基因产物通过调节细胞代谢与催化活性最终影响玉米籽粒、苞叶和穗轴的含水量。

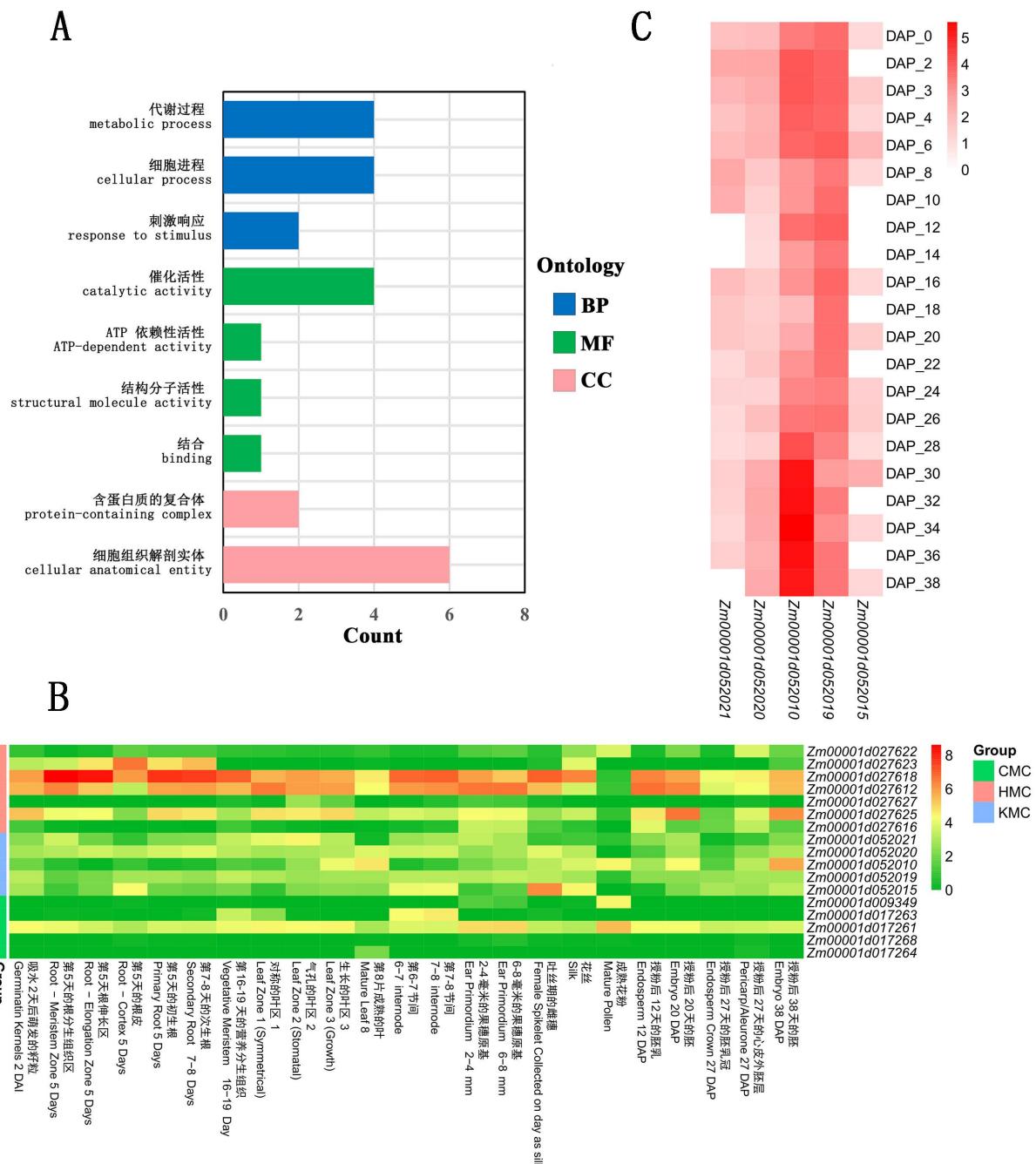
结合玉米自交系 B73 基因表达数据分析，发现这 17 个候选基因在玉米 23 个组织中的表达量存在较大差异（图 5B）。穗轴含水量相关基因（除 *Zm00001d017261* 外）表达量普遍较低，籽粒含水量相关基因在不同部位的表达也存在差异，苞叶含水量相关的两个基因 *Zm00001d027618* 和 *Zm00001d027612* 在各个组织表达最活跃。

进一步分析 5 个籽粒含水量候选基因（*Zm00001d052010*、*Zm00001d052019*、*Zm00001d052021*、*Zm00001d052020* 和 *Zm00001d052015*）在籽粒发育不同阶段的表达情况（图 5 C），结果表明，这 5 个候选基因在授粉后籽粒发育的 0~38 天内表达存在显著差异。特别是 *Zm00001d052010* 在籽粒发育的 28~38 天表达量显著增加，而 *Zm00001d052019* 在籽粒发育的 0~20 天相对较高。相比之下，*Zm00001d052021*、*Zm00001d052020* 和 *Zm00001d052015* 在 0~38 天期间表达量较低，且变化不大。

表 4 3 种及以上方法共定位 SNP 挖掘的候选基因

Table 4 Candidate genes identified from Significant SNPs co-detected by three and more methods

性状	SNP 标记	平均表型贡献率	候选基因	基因注释
Trait	SNP Marker	Average PVE%	Candidate gene	Gene annotation
苞叶含水量	Chr1_9232728	7.32	<i>Zm00001d027622</i>	锌指 (C3HC4 型 RING 指) 家族蛋白
HMC				
			<i>Zm00001d027623</i>	MYB2 转录因子
			<i>Zm00001d027618</i>	60S 核糖体蛋白 L21
			<i>Zm00001d027612</i>	谷氨酸-tRNA 连接酶
			<i>Zm00001d027627</i>	SBT1.2 亚基蛋白酶
			<i>Zm00001d027625</i>	ABC 转运蛋白 C 家族 MRP4
			<i>Zm00001d027616</i>	RPM1 相互作用蛋白 13
籽粒含水量	Chr4_176596174	7.24	<i>Zm00001d052021</i>	DNA 复制许可因子 MCM5
KMC				
			<i>Zm00001d052020</i>	β-1,2-N-乙酰葡糖胺基转移酶 II
			<i>Zm00001d052010</i>	钙调神经磷酸酶 B 亚基
			<i>Zm00001d052019</i>	假定的 RING 锌指结构域超家族蛋白
			<i>Zm00001d052015</i>	丝氨酸蛋白
穗轴含水量	Chr8_57716249	9.25	<i>Zm00001d009349</i>	假定的呼吸爆发氧化酶同源蛋白 H
CMC				
穗轴含水量	Chr5_191021635	6.21	<i>Zm00001d017263</i>	假定的未知功能蛋白 (DUF640)
CMC				
			<i>Zm00001d017261</i>	假定的酪蛋白激酶家族蛋白
			<i>Zm00001d017268</i>	Myb 相关蛋白质 Myb4
			<i>Zm00001d017264</i>	细胞壁相关受体激酶 5



A: 候选基因富集分类结果, BP: 生物过程 Biological process, MF: 分子功能 Molecular function; CC, 细胞组分 Cellular component; B: 17个候选基因在玉米 23 个组织中的表达量;C: 5个候选基因在籽粒发育 0-38 天的表达量

A: Candidate gene enrichment classification results, BP: Biological process; MF: Molecular function; CC: Cellular component. B: The expression levels of 17 candidate genes across 23 tissues in maize. C: The Expression of five candidate genes on days 0-38 of seed development

图 5 候选基因的富集分析和动态表达分析

Fig.5 Enrichment and dynamic expression analysis of candidate genes

3 讨论

3.1 GWAS 定位结果的比较

玉米籽粒含水量是多基因控制的数量性状，其遗传基础复杂。本研究定位结果显示，籽粒、苞叶和穗轴含水量分别检测到 26、15 和 23 个显著位点。其中，籽粒含水量位点 Chr10_2640900 与李文强关于籽粒含水量的研究中定位到的授粉后第 52 天籽粒含水量相关位点 Chr10_3167875 位于同一 bin 区域；Chr1_253390359 与授粉后第 46 天籽粒含水量相关位点 Chr1_228555720 位于同一 bin 区域^[23]。苞叶含水量相关的位点 Chr1_9232728 与张磊关于籽粒脱水速率研究中第五周含水量 QTL 位点 qFwc1-1 位于同一 bin 区域^[24]。穗轴含水量相关的位点 Chr5_191021635 与刘显君等关于玉米成熟期脱水速率的研究中挖掘的籽粒含水量 QTL 位点 qKdr-5-1 区段重叠^[25]。

另外，苞叶性状、穗轴性状及叶片性状均会对成熟时籽粒含水量产生影响。本研究中相关分析表明籽粒含水量与苞叶含水量、穗轴含水量之间存在显著相关性。通过关联分析，发现籽粒含水量与苞叶含水量有 2 个共位点 Chr1_253390359 和 Chr10_2640900，籽粒含水量与穗轴含水量有 3 个共位点 Chr2_110828978、Chr2_213305770 和 Chr6_108012658，进一步从遗传基础上说明籽粒含水量与苞叶含水量和穗轴含水量存在显著相关性，这与 zhang 等关于籽粒和穗轴含水量研究的结果一致^[1]。因此，为实现成熟期低籽粒含水量的育种目标，研究苞叶和穗轴性状对籽粒含水量的影响显得尤为重要。

3.2 不同 ML-GWAS 方法的比较

An 等人在玉米穗行数研究中同时采用了单位点和多位点 GWAS 方法进行关联分析，发现多位点模型检测到的显著位点数量多于单位点模型。对于多基因控制的数量性状，挖掘更多位点对全基因组选择育种的效率有显著提升^[26]。Zhang 等^[27]、Wen 等^[28]、Liu 等^[29]对多种 GWAS 模型的研究也表明，多位点模型在检测复杂数量性状时能提供较高的检测效率并保持准确性。

本研究中，发现不同 ML-GWAS 方法对玉米籽粒含水量相关性状的检测效力存在差异。pLARmEB 检测到的位点最多，而 FASTmrEMMA 检测到的位点数量最少。在 6 种 GWAS 方法中，mrMLM 方法提高了检测小效应基因位点的能力，但可能会出现过拟合问题^[30]。FASTmrMLM 通过实施最小角回归算法解决了 mrMLM 的过拟合问题^[31]。FASTmrEMMA 采用近似算法来加速计算，可能遗漏一些真正的 QTNs^[28]。pLARmEB 结合了多基因位点分析和经验贝叶斯方法，提高了罕见变异与性状关联的检测能力^[27]。pKWmEB 结合了多性状分析、亲缘关系权重和经验贝叶斯方法，提高了对复杂性状相关基因位点的检测能力，由其是在处理大规模数据集时^[32]。ISISEM-BLASSO 结合了工具变量和稀疏贝叶斯回归技术，对复杂性状相关基因位点的检测能力得到加强^[33]。结合前人研究发现，对于不同性状，不同 GWAS 方法挖掘的位点数量存在

较大差异。如 Xu 等关于玉米淀粉糊化特性的研究中发现 FASTmrEMMA 检测到的位点最多, GEMMA 检测到的位点最少。Zhang 等对玉米抗倒性状 GWAS 研究中发现 ISISEM-BLASSO 方法检测到的显著位点最多^[12]。结合前人定位结果, 发现与含水量相关的位点大多集中在 1 号^[25]、5 号^[7]和 6 号染色体^[26], QTL 定位结果和 GML、MLM 等单位点 GWAS 模型很少检测到 9 号、10 号染色体上与成熟期含水量相关的位点。本研究使用 6 种 ML-GWAS 方法检测到含水量相关基因在 10 条染色体上均有分布^[34], 其中 1 号染色体上显著位点数量最多, 为 13 个, 9 号、10 号染色体上分别有 1 和 6 个显著位点。这表明多位点关联分析方法更适合挖掘微效多基因。对于这 3 个含水量性状 pLARmEB 方法检测到更多位点, 可以为后续开展分子标记辅助育种以及全基因组选择提供更多参考位点。

3.3 候选基因分析

为确保挖掘位点的可靠性, 本研究筛选了至少被 3 种多位点 GWAS 方法共同检测到的显著位点, 并根据表型贡献率 $\geq 5\%$ 进一步筛选显著位点后, 查找影响玉米籽粒、穗轴和苞叶含水量的基因。在苞叶含水量显著位点 Chr_9232728 附近发现编码锌指家族蛋白的 *Zm00001d027622* 基因。前人研究表明该家族成员作为 RNA 结合蛋白, 在 mRNA 加工过程中具有调节功能, 并参与非生物和生物应激, 调节植物的生长发育^[35]。*Zm00001d027623* 编码的 MYB2 转录因子参与植物发育、次生代谢、激素信号转导、抗病性和非生物胁迫^[36,37]。在籽粒含水量显著位点 Chr4_176596174 附近, 发现与 E3 泛素连接酶功能相关的 *Zm00001d052019* 锌指结构域家族蛋白。这些蛋白在植物生长发育以及对干旱、盐、温度、活性氧和有害金属等非生物胁迫的响应中发挥重要作用^[38]。在穗轴含水量相关的显著位点 Chr8_57716249 附近, 发现了编码呼吸爆发氧化酶同源基因的 *Zm00001d009349*, 基因表达后产生的活性氧是胁迫反应的重要调节剂^[39]。在 Chr5_191021635 附近的位点与影响细胞壁相关受体激酶的 *Zm00001d017264* 相关, 这些激酶在植物细胞伸长和对生物和非生物胁迫的反应中起重要作用^[40]。尽管这些基因在调节植物发育中起着关键作用, 但它们是否影响玉米籽粒发育及其具体机理尚不清楚。结合前人关于籽粒发育的动态转录组数据, 我们对 5 个籽粒含水量的候选基因表达量进行了分析。结果显示, 在授粉后籽粒发育的 0~38 天内, 这些基因在不同时期的表达量各有差异。其中, *Zm00001d052010* 在授粉后籽粒发育的第 28 天表达量开始增加, 并持续至第 38 天, 这表明 *Zm00001d052010* 可能是影响玉米籽粒后期发育过程中脱水的关键基因。

参考文献

- [1] Zhang M Y, Xiangchen C Y, Yan J Q, Chengxu Y J, Liu H, Zou C Y, Pan G T, Shen Y, Ma L L. Genome-Wide Association Study Reveals the Genetic Basis of Kernel and Cob Moisture Changes in Maize at Physiological Maturity Stage. *Plants*, 2022, 11(15): 1989
- [2] Qian Y L, Zhang X Q, Wang L F, Chen J, Chen B R, Lv G H, Wu Z C, Guo J, Wang J, Qi Y C, Li T C, Zhang W, Ruan L, Zuo X L. Detection of QTLs controlling fast kernel dehydration in maize (*Zea mays* L.). *Genetics and molecular research*, 2016, DOI:10.4238/gmr.15038151
- [3] Jiang S Q, Zhang H B, Ni P Z, Yu S, Dong H X, Zhang A, Cao H Y, Zhang L J, Ruan Y Y, Cui Z H. Genome-Wide Association Study Dissects the Genetic Architecture of Maize Husk Tightness. *Frontiers in Plant Science*, 2020, DOI:10.3389/fpls.2020.00861

- [4] Xu C C, Zhang P, Wang Y Y, Luo N, Tian B J, Liu X W, Wang P, Huang S B. Grain yield and grain moisture associations with leaf, stem and root characteristics in maize. *Journal of Integrative Agriculture*, 2022, 21(7): 1941-1951
- [5] 姚凌翔. 玉米果穗脱水相关性状的全基因组关联分析. 四川: 四川农业大学, 2021
Yao L X. Genome-wide association analysis of dehydration-related traits in maize ear spikes. Sichuan: Sichuan Agricultural University, 2021
- [6] 李露露, 张磊, 崔敏, 邓德祥, 刘欢欢, 任姣姣, 李志鹏, 印志同. 玉米籽粒含水量相关性状 QTL 定位. 江苏农业科学, 2019, 47(9): 93-96
Li L L, Zhang L, Cui M, Deng D X, Liu H H, Ren J J, Li Z P, Yin Z T. QTL mapping of grain moisture content related traits of maize. Jiangsu Agricultural Sciences, 2019, 47(9): 93-96
- [7] 吕晨曦. 玉米籽粒含水量及相关性状的全基因组关联分析. 四川: 四川农业大学, 2019
Lv C X. Genome Wide Association Study of Grain Water Content and Related Traits in Maize. Sichuan: Sichuan Agricultural University, 2019,
- [8] 党冬冬. 利用 GWAS 解析玉米生理成熟期籽粒脱水性状的遗传基础. 沈阳: 沈阳农业大学, 2019
Dang D D. Analysis of Genetic Basis of Maize Kernel Dehydration Traits During Physiological Maturity Using GWAS. Shenyang: Shenyang Agricultural University, 2019
- [9] Li S F, Zhang C X, Lu M, Yang D G, Qian Y L, Yue Y H, Zhang Z J, Jin F X, Wang M, Liu X Y, Liu W G, Li X H. QTL mapping and GWAS for field kernel water content and kernel dehydration rate before physiological maturity in maize. *Scientific Reports*, 2020, 10(1): 13114
- [10] Li S F, Zhang C X, Yang D G, Lu M, Qian Y L, Jin F X, Liu X Y, Wang Y, Liu W G, Li X H. Detection of QTNs for kernel moisture concentration and kernel dehydration rate before physiological maturity in maize using multi-locus GWAS. *Scientific Reports*, 2021, 11(1): 1764
- [11] Cui Y R, Zhang F, Zhou Y L. The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice. *Frontiers in Plant Science*, 2018, DOI:10.3389/fpls.2018.01464
- [12] Zhang Y L, Liu P, Zhang X X, Zheng Q, Chen M, Ge F, Li Z L, Sun W T, Guan Z R, Liang T H, Zheng Y, Tan X L, Zou C Y, Peng H W, Pan G T, Shen Y. Multi-Locus Genome-Wide Association Study Reveals the Genetic Architecture of Stalk Lodging Resistance-Related Traits in Maize. *Frontiers in Plant Science*, 2018, DOI:10.3389/fpls.2018.00611
- [13] 李真, 刘文童, 杨硕, 郭晋杰, 赵永锋, 黄亚群, 陈景堂, 祝丽英. 成熟期玉米自交系籽粒含水量的快速测定方法. 种子, 2020, 39(9): 156-160
Li Z, Liu W T, Yang S, Guo J J, Zhao Y F, Huang Y Q, Chen J T, Zhu L Y. Rapid Determination of Grain Water Content in Maize Inbred Lines at Mature Stage. *Seed*, 2020, 39(9): 156-160
- [14] 马娟, 刘京宝, 朱卫红, 黄璐, 宇婷, 乔江方. 玉米农艺性状配合力全基因组关联分析和预测. 核农学报, 2023, 37(5): 944-956
Ma J, Liu J B, Zhu W H, Huang L, Ning T, Qiao J F. Genome-Wide Association Study and Prediction for Combining Ability of Maize Agronomic Traits. *JOURNAL OF NUCLEAR AGRICULTURAL SCIENCES*, 2023, 37(5): 944-956
- [15] Knapp S J, Stroup W W, Ross W M. Exact Confidence Intervals for Heritability on a Progeny Mean Basis1. *Crop Science*, 1985, 25(1): 7347
- [16] 李婷, 王亚鹏, 董远, 郭瑞士, 李冬梅, 唐雅伶, 张兴华, 薛吉全, 徐淑兔. 基于杂交群体解析玉米籽粒大小相关性状及其配合力的分子遗传机制. 作物学报, 2022, 48(10): 2451-2462
Li T, Wang Y P, Dong Y, Guo R S, Li D M, Tang Y L, Zhang X H, Xue J Q, Xu S T. Dissecting the genetic basis of kernel size related traits and their combining ability based on a hybrid population in maize. *Acta Agronomica Sinica*, 2022, 48(10): 2451-2462
- [17] Allen G C, Flores-Vergara M A, Krasynanski S, Kumar S, Thompson W F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols*, 2006, 1(5): 2320-2325
- [18] 李周帅, 董远, 李婷, 冯志前, 段迎新, 杨明羨, 徐淑兔, 张兴华, 薛吉全. 基于杂交种群体的玉米产量及其配合力的全基因组关联分析. 中国农业科学, 2022, 55(9): 1695-1710
Li Z S, Dong Y, Li T, Feng Z Q, Duan Y X, Yang M X, Xu S T, Zhang X H, Xue J Q. Genome-Wide Association Analysis of Yield and Combining Ability Based on Maize Hybrid Population. *Scientia Agricultura Sinica*, 2022, 55(9): 1695-1710
- [19] Zhang Y M, Jia Z Y, Dunwell J M. Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Frontiers in Plant Science*, 2019, DOI:10.3389/fpls.2019.00100
- [20] 渠建洲, 冯文豪, 张兴华, 徐淑兔, 薛吉全. 基于全基因组关联分析解析玉米籽粒大小的遗传结构. 作物学报, 2022, 48(2): 304-319.
Qu J Z, Feng W H, Zhang X H, Xu S T, Xue J Q. Dissecting the genetic architecture of maize kernel size based on genome-wide association study. *Acta Agronomica Sinica*, 2022, 48(2): 304-319
- [21] Chen J, Zeng B, Zhang M, Xie S J, Wang G K, Hauck A, Lai J S. Dynamic Transcriptome Landscape of Maize Embryo and Endosperm Development. *Plant*

Physiology, 2014, 166(1): 252-264

- [22] 李婷, 董远, 张君, 冯志前, 王亚鹏, 郝引川, 张兴华, 薛吉全, 徐淑兔. 玉米杂交种穗部性状的全基因组关联分析. 中国农业科学, 2022, 55(13): 2485-2499
Li T, Dong Y, Zhang J, Feng Z Q, Wang Y P, Hao Y C, Zhang X H, Xue J Q, Xu S T. Genome-Wide Association Study of Ear Related Traits in Maize Hybrids. *Scientia Agricultura Sinica*, 2022, 55(13): 2485-2499
- [23] 李文强. 玉米籽粒含水量及其动态变化的遗传结构解析. 武汉: 华中农业大学, 2017
Li W Q. Dissecting the Genetic Architecture of Maize Grain Moisture and the Dynamic Change. Wuhan: Huazhong Agricultural University, 2017
- [24] 张磊. 玉米籽粒脱水速率 QTL 的初步定位. 扬州: 扬州大学, 2016
Zhang L. The QTL Analysis of Kernel Dehydration Rate in Maize. Yangzhou: Yangzhou University, 2016
- [25] 刘显君, 王振华, 王霞, 李庭锋, 张林. 玉米籽粒生理成熟后自然脱水速率 QTL 的初步定位. 作物学报, 2010, 36(1): 47-52
Liu X J, Wang Z H, Wang X, Li T F, Zhang L. Primary Mapping of QTL for Dehydration Rate of Maize Kernel after Physiological Maturing, 2010, 36(1): 47-52
- [26] An Y X, Chen L, Li Y X, Li C H, Shi Y, Zhang D F, Li Y, Wang T Y. Genome-wide association studies and whole-genome prediction reveal the genetic architecture of KRN in maize. *BMC plant biology*, 2020, 20(1): 490
- [27] Zhang J, Feng J Y, Ni Y L, Wen Y J, Niu Y, Tamba C L, Yue C, Song Q, Zhang Y M. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity*, 2017, 118(6): 517-524
- [28] Wen Y J, Zhang H W, Ni Y L, Huang B, Zhang J, Feng J Y, Wang S B, Dunwell J M, Zhang Y M, Wu R L. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics*, 2018, 19(4): 700-712
- [29] Liu X L, Huang M, Fan B, Buckler E S, Zhang Z W. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genetics*, 2016, DOI:10.1371/journal.pgen.1005767
- [30] Wang S B, Feng J Y, Ren W L, Huang B, Zhou L, Wen Y J, Zhang J, Dunwell J M, Xu S, Zhang Y M. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports*, 2016, 6(1): 19444.
- [31] Tamba C L, Zhang Y M. A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*, 2018: 341784.
- [32] Ren W L, Wen Y J, Dunwell J M, Zhang Y M. pKWMEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity*, 2018, 120(3): 208-218.
- [33] Tamba C L, Ni Y L, Zhang Y M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLOS Computational Biology*, 2017, 13(1): e1005357.
- [34] Zhou G F, Zhu Q L, Mao Y X, Chen G Q, Xue L, Lu H H, Shi M L, Zhang Z L, Song X D, Zhang H M, Hao D R. Multi-Locus Genome-Wide Association Study and Genomic Selection of Kernel Moisture Content at the Harvest Stage in Maize. *Frontiers in Plant Science*, 2021, DOI:10.3389/fpls.2021.697688
- [35] Peng X J, Zhao Y, Cao J G, Zhang W, Jiang H Y, Li X Y, Ma Q, Zhu S W, Cheng B J. CCCH-Type Zinc Finger Family in Maize: Genome-Wide Identification, Classification and Expression Profiling under Abscisic Acid and Drought Treatments. *PLOS ONE*, 2012, DOI: 10.1371/journal.pone.0040120
- [36] Katiyar A, Smita S, Lenka S K, Rajwanshi R, Chinnusamy V, Bansal K C. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics*, 2012, 13(1): 544
- [37] Pratyusha D S, Sarada D V L. MYB transcription factors-master regulators of phenylpropanoid biosynthesis and diverse developmental and stress responses. *Plant Cell Reports*, 2022, 41(12): 2245-2260
- [38] Han G L, Qiao Z Q, Li Y X, Yang Z R, Wang C F, Zhang Y Y, Liu L L, Wang B S. RING Zinc Finger Proteins in Plant Abiotic Stress Tolerance. *Frontiers in Plant Science*, 2022, DOI: 10.3389/fpls.2022.877011
- [39] Liu D, Li Y Y, Zhou Z C, Xiang X H, Liu X, Wang J, Hu Z R, Xiang S P, Li W, Xiao Q Z, Wang Y Y, Hu R S, Zhao Q. Tobacco transcription factor bHLH123 improves salt tolerance by activating NADPH oxidase NtRbohE expression. *Plant Physiology*, 2021, 186(3): 1706-1720
- [40] Liu X T, Wang Z C, Tian Y, Zhang S Y, Li D D, Dong W Q, Zhang C Q, Zhang Z. Characterization of wall-associated kinase/wall-associated kinase-like (WAK/WAKL) family in rose (*Rosa chinensis*) reveals the role of RcWAK4 in Botrytis resistance. *BMC plant biology*, 2021, 21(1): 526