

王钰文, 钟 鑫, 范晓庆, 等. 基于 96K SNP 芯片的谷子抽穗期性状全基因组关联分析[J]. 江苏农业科学, 2025, 53(6): 33–47.  
doi:10.15889/j.issn.1002-1302.2025.06.004

# 基于 96K SNP 芯片的谷子抽穗期性状全基因组关联分析

王钰文<sup>1</sup>, 钟 鑫<sup>1</sup>, 范晓庆<sup>1</sup>, 刘颖慧<sup>1</sup>, 赵治海<sup>2</sup>, 范光宇<sup>2</sup>, 王小明<sup>2</sup>

(1. 河北北方学院农林科技学院, 河北张家口 075000; 2. 张家口市农业科学院谷子研究所, 河北张家口 075000)

**摘要:**抽穗期是衡量农作物重要性的一个关键指标,它直接影响作物对地区和季节的适应能力。了解不同谷子品种的适宜种植区域,对于农业生产和实践具有极其重要的意义。为揭示控制抽穗期性状的遗传机制,为开展理想性状标记辅助选择育种奠定基础,调查了 230 个样本的抽穗期叶姿、刚毛长度、刚毛颜色、护颖颜色、花药颜色、茎秆长度、茎秆粗细、伸长节间数 8 个抽穗期性状并进行基因组重测序,拟以水稻抽穗期相关性状为研究对象,利用 R 软件中 GWASpoly 方法对水稻抽穗前后的相关性状进行多层次的相关性研究,利用基于 1-随机和正双(种群结构+血缘)的杂交 Q+K(种群结构+血缘)方法,分别构建分位数-分位数(QQ-Plot)和曼哈顿图。筛选与表型相关性显著的单核苷酸多态性位点(SNPs),在 10 kb 区间找到与表型相关的基因,并对这些 SNPs 进行功能注释。明确了每个性状与其具有显著关联性的染色体位置以及关联度,与抽穗期叶姿、刚毛长度、刚毛颜色、护颖颜色、花药颜色、茎秆长度、茎秆粗细、伸长节间数关联的 SNP 的关键位点在 9 条染色体上均有分布。

**关键词:**谷子;全基因组关联分析;SNP;抽穗期性状;候选基因

**中图分类号:**S188;S515.01 **文献标志码:**A **文章编号:**1002-1302(2025)06-0033-15

谷子(*Setaria italica*)是一种起源于我国的多年生草本植物,小米是谷子去壳后的产物。谷子是我国最主要的农作物之一,也是我国陆地生态系统的核心组成部分<sup>[1]</sup>,其抗旱节水能力极强,有望成为我国应对未来全球气候变化的战略性后备作物。笔者所在课题组前期已经完成了谷子 SSR 简单重复序列分子标记的检测<sup>[2-3]</sup>,为开展谷子分子育种研究打下了坚实的基础。

近年来,研究人员在基因组层面上开展了一系列对谷子的研究工作,包括利用重测序技术对重要农艺性状进行全基因组关联分析(GWAS)<sup>[4-5]</sup>;利用杂交分离群体对重要性状进行数量性状基因座(QTL)作图<sup>[6-8]</sup>,以及抗逆基因的克隆与功能验证<sup>[9-10]</sup>等。谷子是短日喜温的农作物,对光照和温度敏感,具有显著的地域特征,是开展谷子生态适宜度评估的基础。通过对谷子的生态适应性评估,发掘出适应性广的谷子种质,对于培育跨地域适应性广的谷子新品种有重大的实际应用价值。抽穗期是反映作物生态适应能力的一个重要指标,还可以反映作物的生长状况,光照、温度等环境因素对其有显著影响<sup>[11-12]</sup>。Gupta 等对 184 个谷子品种的产量性状进行了研究,并利用 50 对 SSR 分子标记对其进行遗传连锁,并将其定位在第 5 条染色体上<sup>[13]</sup>。Jia 等对 916 种谷子品种进行重测序,与 47 个性状进行了关联分析<sup>[4]</sup>。谷子的全基因组测序工作<sup>[14-15]</sup>已全部完成,但其在全基因组范围内的相

收稿日期:2024-04-08

基金项目:国家现代农业产业技术体系建设专项(编号:CARs-06-14.5-A7);河北省重点研发计划(编号:21326302D);河北北方学院校级课题;河北宣化谷子科技小院项目。

作者简介:王钰文(1998—),男,山东潍坊人,硕士研究生,研究方向为作物栽培,E-mail:929660014@qq.com;共同第一作者:钟 鑫(1997—),女,河北承德人,硕士研究生,研究方向为分子植物育种,E-mail:1815708558@qq.com。

通信作者:刘颖慧,博士,教授,研究方向为作物栽培。E-mail:leely519@126.com。

pepper line PI201234 (*Capsicum annuum* L.) [J]. Molecular Breeding, 2016, 36(6): 66.

[31] 雷 阳, 成 妍, 乔 宁, 等. 辣椒苗期抗感疫病比较转录组学分析[J]. 华北农学报, 2019, 34(3): 194–202.

[32] 雷 阳, 成 妍, 乔 宁, 等. 辣椒苗期抗感黄瓜花叶病毒病比较转录组学分析[J]. 中国蔬菜, 2021(7): 24–31.

[33] 王红娟, 唐荣莉, 蒋晓英, 等. 辣椒不同发育时期果实的转录组分析[J]. 西南农业学报, 2022, 35(3): 517–525.

[34] 周 欣. 辣椒素合成结构基因及 MYB 转录因子的克隆和功能验证[D]. 广州: 华南农业大学, 2016: 44–49.

[35] 张 威, 张丽平, 吴 丹, 等. 黄灯笼辣椒 MYB4 转录因子的克隆与序列分析[J]. 分子植物育种, 2023, 21(6): 1801–1810.

关研究<sup>[4,13]</sup>鲜有报道。本研究拟在前期研究工作的基础上,通过对谷子抽穗前后 8 个主要特征(抽穗期叶姿、刚毛长度、刚毛颜色、护颖颜色、花药颜色、茎秆长度、茎秆粗度和伸长节间数)进行重测序,开展 SNP 标记与 8 个抽穗期性状的全基因组关联分析,获取关联的 SNP 标记。通过对关联区域的候选基因进行功能分析,找出与谷子抽穗期相关的关键基因,为进一步精细定位和育种奠定基础。

## 1 材料与方法

### 1.1 试验材料和种植方法

本研究所用的 230 份谷子材料由张家口市农业科学院提供,于 2023 年 5 月中旬播种于张家口市农业科学院试验田,采用田间试验,每份谷子材料种植 2 行,行长 2 m,行距 50 cm,株距 3~5 cm。按当地常规管理,每份材料于抽穗期测量田间性状。

### 1.2 基本试验策略

对样本进行核 DNA 提取,并确保其质量达到试验标准。为评估 DNA 的质量,采用多种检测技术。首先,使用紫外-可见分光光度计(NanoDrop ND-2000,赛默飞世尔科技公司)设备对 DNA 进行定量和定性分析,以确保 DNA 的浓度和纯度符合试验要求。此外,还通过凝胶电泳技术来评估 DNA 的完整性,确保其分子量分布和条带清晰度满足试验标准。在 DNA 质量检测合格后,按照芯片检测的标准操作流程,对样本 DNA 进行后续处理,包括 DNA 的扩增、片段化、沉淀、重悬等步骤,为芯片杂交做好准备。在芯片杂交阶段,将 DNA 与芯片上的探针进行结合,以检测特定的遗传变异。杂交完成后,进行洗涤和染色步骤,去除未特异性结合的 DNA,并使信号更加清晰,便于后续的信号检测和分析。

### 1.3 原始谷子芯片数据处理

首先,使用 GeneTitan 系统扫描获得的原始数据被导入到 Affymetrix Power Tools 软件中。该软件用于对原始数据进行聚类分析和基因分型处理。处理完成后, Affymetrix Power Tools 将数据导出为 PLINK 格式,以便于后续的数据分析。将 PLINK 格式的数据通过 PLINK 程序转换为 SNP 的 VCF 格式。VCF(variant call format)是一种用于存储遗传变异信息的标准文件格式,它包含了关于 SNP 和其他遗传变异的详细信息。将数据转换为 VCF 格式后,可以对这些数据进行更深入的分析,包括关联研究、群体遗传学分析等。

### 1.4 芯片 SNP 数据的转化

对以上总的 SNP VCF 文件进行筛选,按不同时期有表型的样本数,获得抽穗期性状 SNP 的 VCF 文件(使用 VCFtools 处理)。

### 1.5 样品群体结构分析

用 ADMIXTURE 软件进行主成分分析,得到样品的群体聚类情况。

### 1.6 连锁不平衡(LD)分析

使用连锁不平衡系数( $r^2$ )评估染色体上 SNP 点之间的遗传关联强度,当  $r^2$  为 0 时,表明群体处于遗传平衡或缺乏关联链接,而  $r^2$  值越高,意味着 SNP 之间的空间关系越紧密。通常,若  $r^2 \geq 0.8$ ,会被认定为存在连锁不平衡现象。通过计算  $r^2$  下降至原来的一半时所对应的 SNP 间距,定义为连锁不平衡的衰减距离,这一步骤借助 VCFtools 软件进行定量分析。随后采用非线性回归模型,精确地描绘  $r^2$  随 SNP 物理距离增长而逐渐减弱的动态趋势图。

### 1.7 芯片数据的 GWAS 分析和候选基因的预测

谷子特异定制的 SNP 芯片包含 82 738 个独特的遗传标记,经过严格筛选,最终在抽穗阶段锁定了一组普遍存在于 95% 样本中的变异位点,数量达到了 54 975 个。接下来,运用先进的 GWAS 策略,借助 R 语言包 GWASpoly 的深度分析流程,对各个生长阶段的多种性状进行细致的关联研究。在显性模型(1-dom)和关联强度 2 种分析模式下,采用了混合的群体结构+亲缘关系(Q+K)方法,生成分位数-分位数(QQ-Plot)图表和具有洞察力的曼哈顿图。对于那些表现出显著表型关联的 SNP,在它们所在基因区域的上下游 10 kb 碱基范围内进行深入挖掘。然后,对这些 SNP 位点进行详尽的功能注释,以揭示它们可能的生物学意义和影响。

## 2 结果与分析

### 2.1 230 份谷子品种群体结构分析

ADMIXTURE 是一种用于遗传数据聚类的软件,它通过计算不同 K 值(假设的祖先群体数量)下的交叉验证误差来评估模型的拟合度。在聚类分析过程中,K 值的趋势通常表现为先下降后上升,其中 K 值开始上升的点被认为是遗传背景最相似的祖先类群的最佳数量。根据图 1 所示的趋势,本试验中的最佳 K 值大约为 35,这意味着样本可以被划分为大约 35 个遗传背景相似的祖先类群。

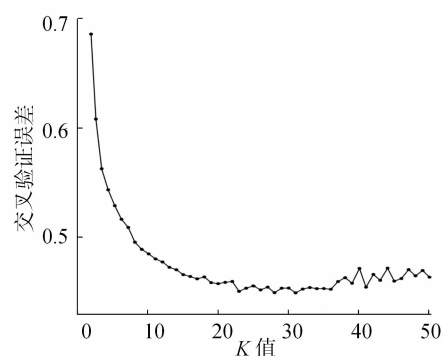


图1 230 份谷子群体结构分析

2.2 LD 分析

对谷子基因组的连锁不平衡进行深入分析,以确定 LD 衰减距离。通过分析发现,谷子基因组的 LD 衰减距离大约为 9 Mb(图 2),较短的 LD 衰减距离意味着在谷子基因组中,遗传变异之间的连锁关系可以在较短的物理距离内迅速衰减,这对于后续的遗传关联分析具有显著的优势。在进行关联分析时,研究者通常要确定候选基因区域,以寻找与特定性状相关的遗传变异。较短的 LD 衰减距离有助于缩小这些候选区域的范围,从而减少后续分析的工作量。

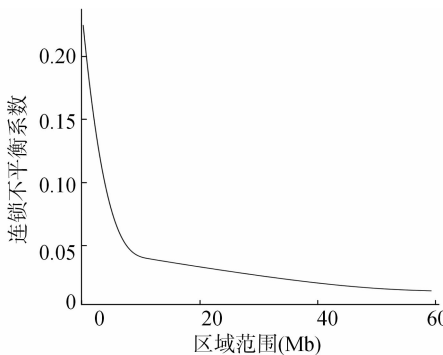


图2 LD 衰减分布

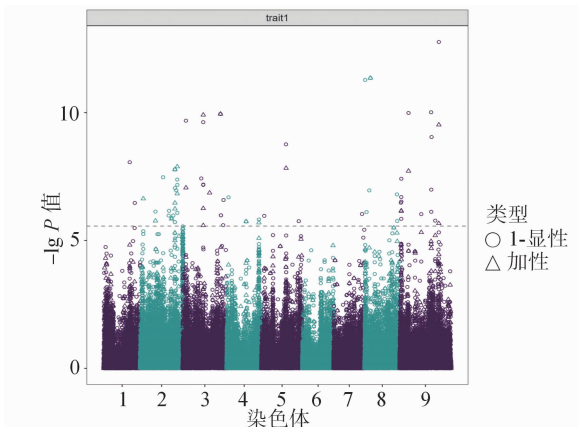
2.3 谷子抽穗期质量性状的全基因组关联分析

质量性状表现为非连续性的特征转变,这些性状的差异不是渐进的,而是呈现出明显的断裂式变化。这类性状通常由 1 个或少数几个基因决定。在谷物的 8 个抽穗期特性中,质量性状涵盖了抽穗期叶姿、刚毛颜色、护颖色泽以及花药颜色。研究发现,有 SNP 位点与谷物抽穗期的这些质量性状显著相关。抽穗期叶姿的显著 SNP 分布在全部 9 条染色体上,其中与 8 号染色体上的 SNP 位点(8\_6603655)关联最为紧密(表 1、图 3、图 4)。和谷子刚毛颜色相关的显著型 SNP 在 9 条染色体上均有分布,其中和 4 号染色体(4\_7395553)的关联度最

高(表 2、图 5、图 6)。和谷子护颖颜色相关的显著型 SNP 在 9 条染色体上均有分布,和 4 号染色体(4\_7395553)的关联度最高(图 7、图 8 和表 3)。和谷子花药颜色相关的显著型 SNP 在 9 条染色体上均有分布,和 6 号染色体(6\_34378428)的关联度最高(图 9、图 10 和表 4)。

表 1 谷子抽穗期叶姿相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
8_6603655	8	6 603 655	T	C	11.350 754 530
3_43379637	3	43 379 637	T	C	9.946 054 443
9_44980329	9	44 980 329	C	T	9.523 890 412
2_42694544	2	42 694 544	T	C	7.889 239 967
5_28258269	5	28 258 269	G	A	7.827 629 571
4_22349318	4	22 349 318	C	T	5.739 543 485
1_29854195	1	29 854 195	G	A	4.826 200 647
6_35260273	6	35 260 273	C	T	4.801 922 799
7_1609360	7	1 609 360	C	T	4.222 082 582



显著差异的 SNP 位点被标记在虚线以上( $P<0.01$ ),表示这些位点与谷子抽穗期叶姿性状具有显著关联。下图同图3 谷子抽穗期叶姿性状的曼哈顿图

2.4 谷子抽穗期数量性状的全基因组关联分析

连续变异的特征被称为数量性状,这种性状在群体内各个体间呈现连续变化,并且由多对等位基因共同作用,同时易受环境条件的影响。在谷物的抽穗期性状中,有 4 个数量性状值得注意,分别是刚毛长度、茎秆长度、茎秆粗度和伸长节间数。和谷子刚毛长度相关的显著型 SNP 在 9 条染色体上均有分布,和 1 号染色体(1\_38241023)的关联度最高(图 11、图 12 和表 5)。研究发现,存在若干 SNP 位点与这些茎秆特性显著相关。关于茎秆长度,显著的 SNP 位点遍布 9 条染色体,其中 4 号染色体上的 1 个位点(4\_38139078)显示出最强关联(图 13、图

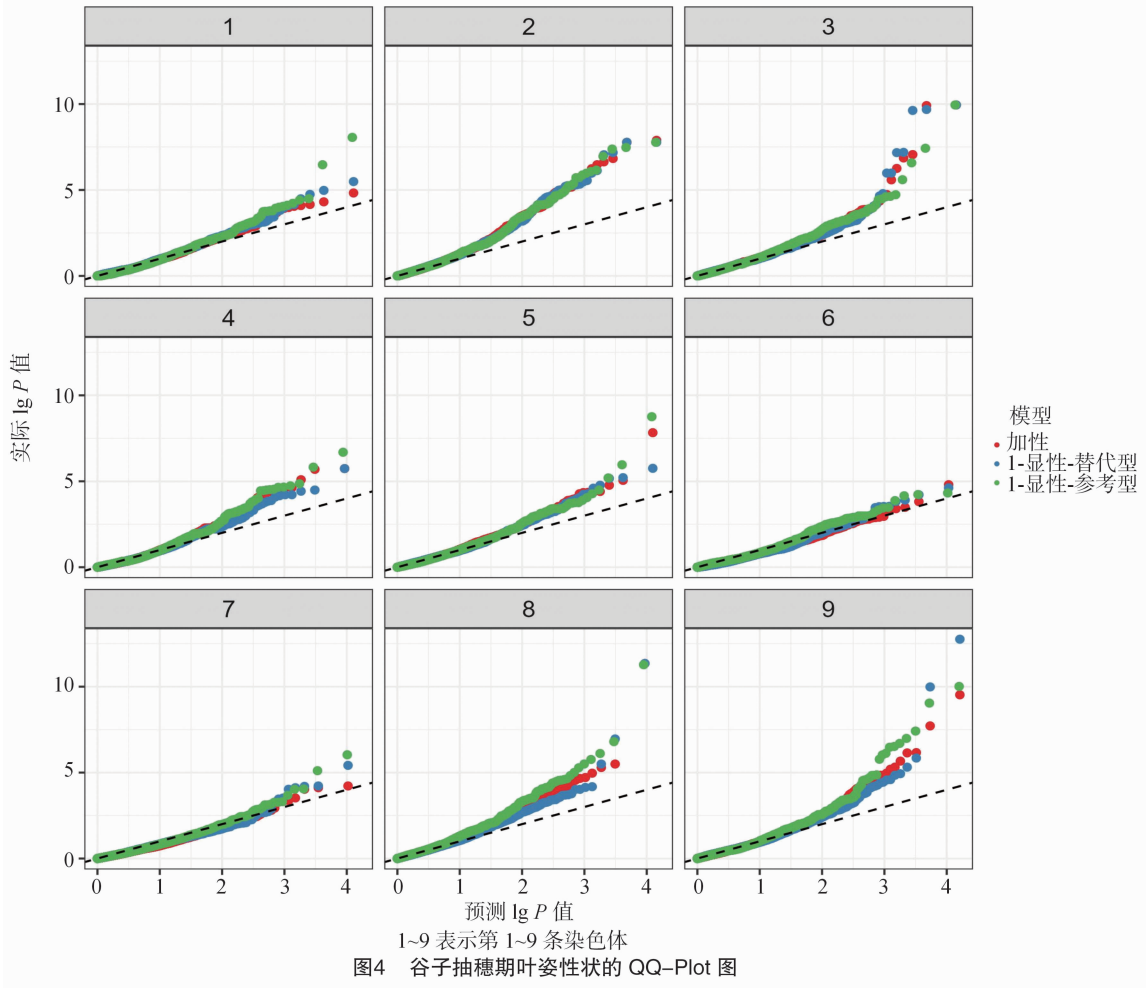
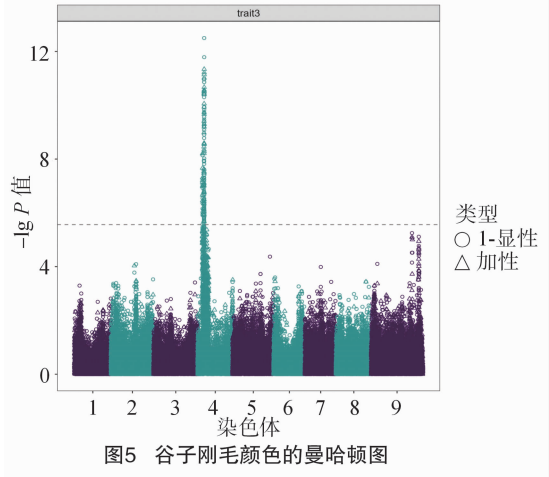


表 2 谷子刚毛颜色相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
4_7395553	4	7 395 553	T	C	11.341 040 490
9_46698524	9	46 698 524	A	G	5.101 290 918
2_26813558	2	26 813 558	A	G	4.033 885 677
6_1504470	6	1 504 470	T	C	3.450 080 812
8_34320023	8	34 320 023	C	T	3.436 112 980
5_26621146	5	26 621 146	G	A	3.087 423 931
7_15495205	7	15 495 205	G	A	2.991 603 042
3_47437967	3	47 437 967	C	T	2.959 784 793
1_6977058	1	6 977 058	C	T	2.624 831 724

14 和表 6)。对于茎秆粗度,同样有显著的 SNP 位点存在于所有 9 条染色体,而 3 号染色体上的 1 个位点(3\_167397)关联性最强(图 15、图 16 和表 7)。至于伸长节间数,也有显著的 SNP 位点在 9 条染色体上被识别,其中 3 号染色体的相同位点(3\_167397)关联度最高(图 17、图 18 和表 8)。



2.5 候选基因的预测

在筛选出的一系列基因中,外显子区域的变异基因总数达到了 173 个。其中,与关键发育阶段(抽穗期)叶姿相关的基因共有 20 个,它们分布在不同的染色体上,具体来说,这些基因分布在第 2、3、4、5、9 号染色体(表 9)。与刚毛长度相关的基因

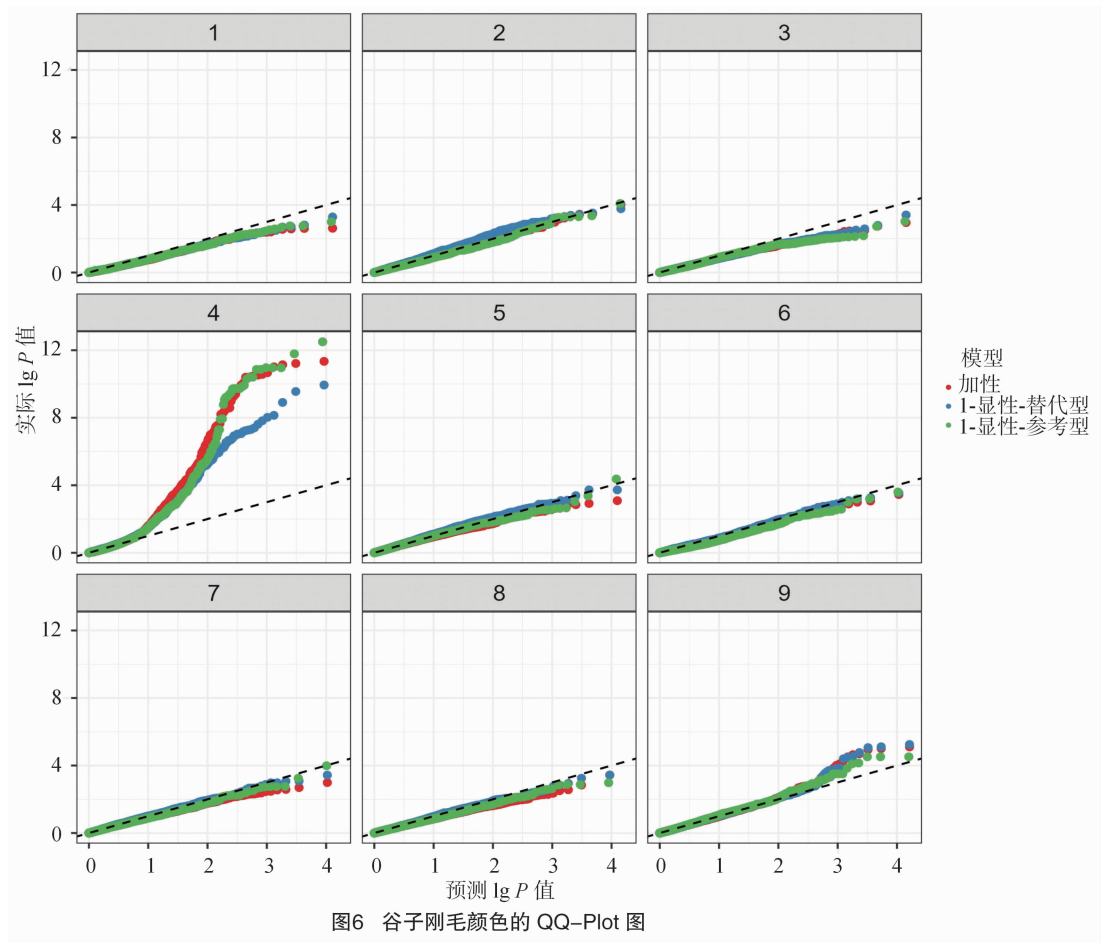


图6 谷子刚毛颜色的 QQ-Plot 图

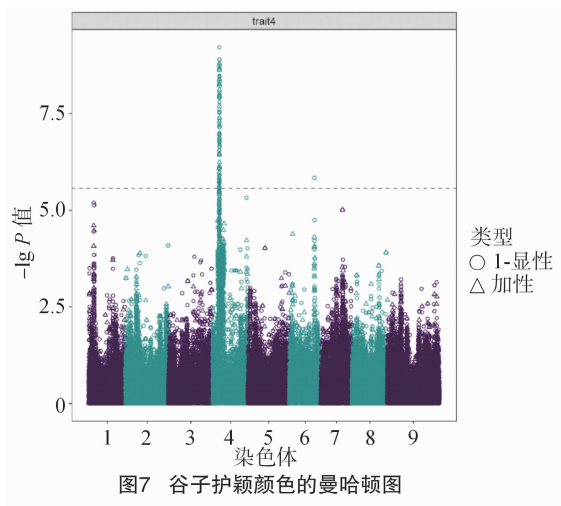


图7 谷子护颖颜色的曼哈顿图

数量为 19 个,它们主要集中在 1 号和 5 号染色体(表 10)。与刚毛颜色相关的候选基因最多,多达 54 个,集中在 4 号染色体(表 11)。同样,护颖颜色的关联基因也有 30 个,全部聚集在 4 号染色体(表 12)。对于花药颜色的变化,发现了 46 个可能的候选基因,分布在 6 号、7 号和 9 号染色体上,但以 6 号染色体上的基因最多(表 13)。在植物茎秆粗度

的遗传调控中,有 3 个显著的候选基因分布在 2 号和 3 号染色体上(表 14)。最后,仅发现 1 个与伸长节间隔数相关的候选基因,它位于 3 号染色体上(表 15)。茎秆长度的候选基因没有成功预测。

### 3 讨论与结论

在过去的科研工作中,水稻和小麦的抽穗关键阶段相关基因已陆续被定位并克隆<sup>[16-19]</sup>。至于高粱,关于抽穗期的基因研究也已经进行了数量性状位点(QTL)分析<sup>[20]</sup>。本研究利用谷子全基因组关联分析策略,利用 96K SNP 芯片对选定的样本进行基因型鉴定。这一过程通常包括挑选适合目标的谷子种质资源,然后扫描芯片以评估标记之间的 LD,同时对种质的表型数据进行验证,最后将所有数据整合,进行全基因组关联分析。与之相反,候选基因关联分析则较为有限,主要聚焦于候选基因的序列分析和多态性检测,随后通过 Q + K 模型配合表型数据,对候选基因进行关联分析。深入研究复杂性状的全基因组关联分析(GWAS)不仅揭示了遗传机制的复杂性,还为后续的基因定位提供了有

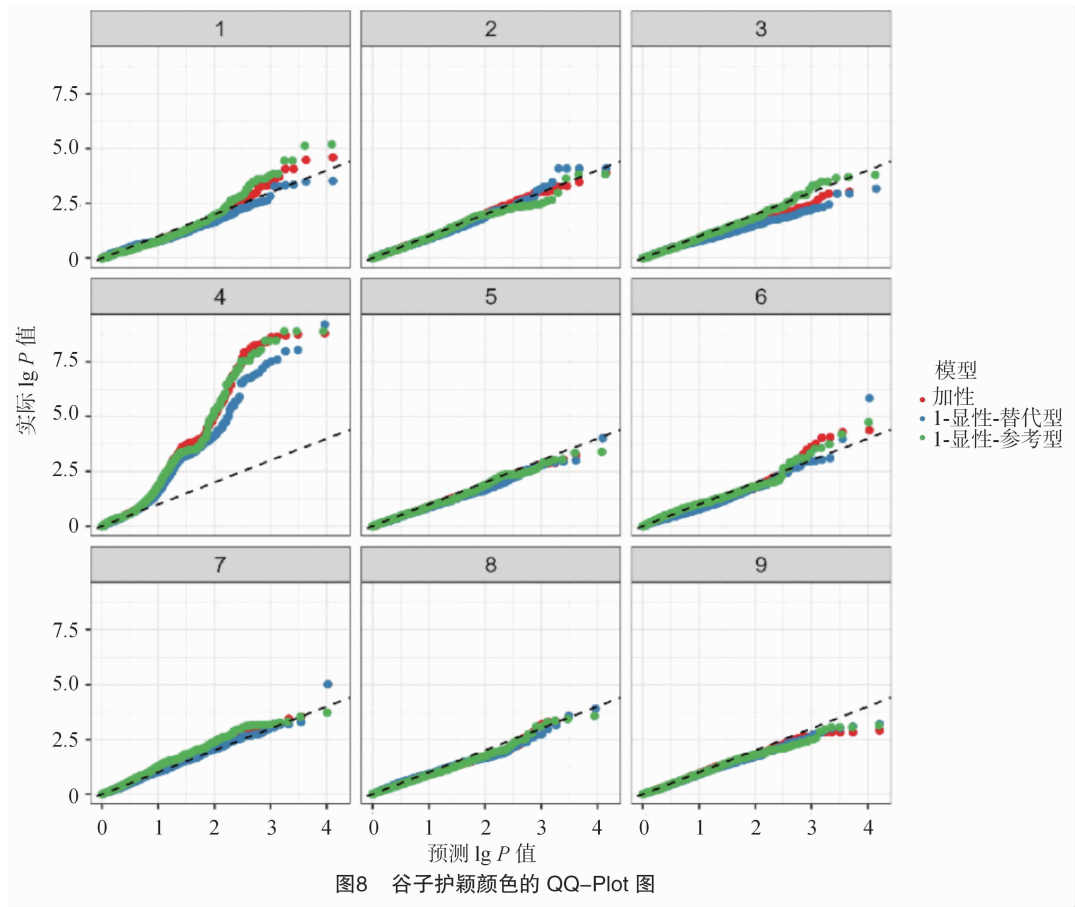
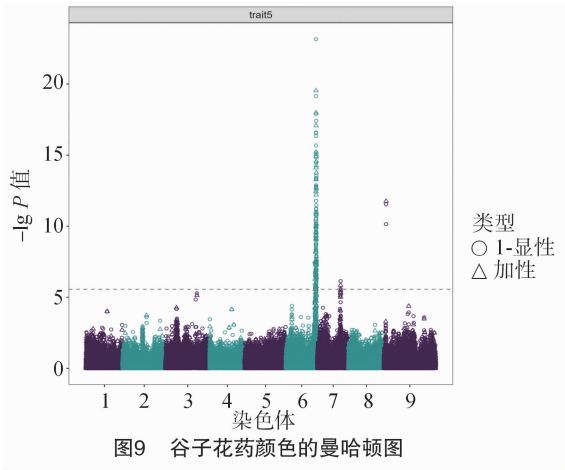


表 3 谷子护颖颜色相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
4_7395553	4	7 395 553	T	C	8.802 841 933
7_24732920	7	24 732 920	C	T	5.009 659 622
1_5480917	1	5 480 917	G	C	4.602 622 268
6_3697836	6	3 697 836	G	T	4.380 919 998
5_19302539	5	19 302 539	G	A	4.019 083 213
8_38727855	8	38 727 855	T	C	3.902 537 239
2_16242655	2	16 242 655	A	G	3.886 674 420
3_21710578	3	21 710 578	A	G	3.167 969 597
9_15237910	9	15 237 910	C	T	2.914 907 598

价值的线索。单倍型作为染色体上多个等位基因共享的单元,蕴含着等位基因间连锁关系的珍贵信息。传统 GWAS 通过逐一检测 SNP 来寻找关联,而单倍型 GWAS (hGWAS) 则将单倍型区块作为整体纳入分析,以发现与性状有显著关联的特定区域<sup>[21]</sup>。当前,作物表型评估主要依赖人工手段,耗时且易受主观因素影响。随着高通量表型平台的兴起,未来有望实现表型数据的高效且精确获取,



从而充分挖掘 GWAS 的潜力,突破传统对疾病抗性和产量等少数性状的关注,扩展到对数十个甚至更多难以人工衡量的作物特性,如抗逆性、生长速度等,从而揭示大量显著关联位点,进一步挖掘隐藏的基因资源<sup>[22-23]</sup>。本试验找出了与谷子抽穗期 8 个性状具有高度关联的 SNP 位点,发现在 9 条染色体上均有分布,在 4 号染色体上数量最多。与谷子

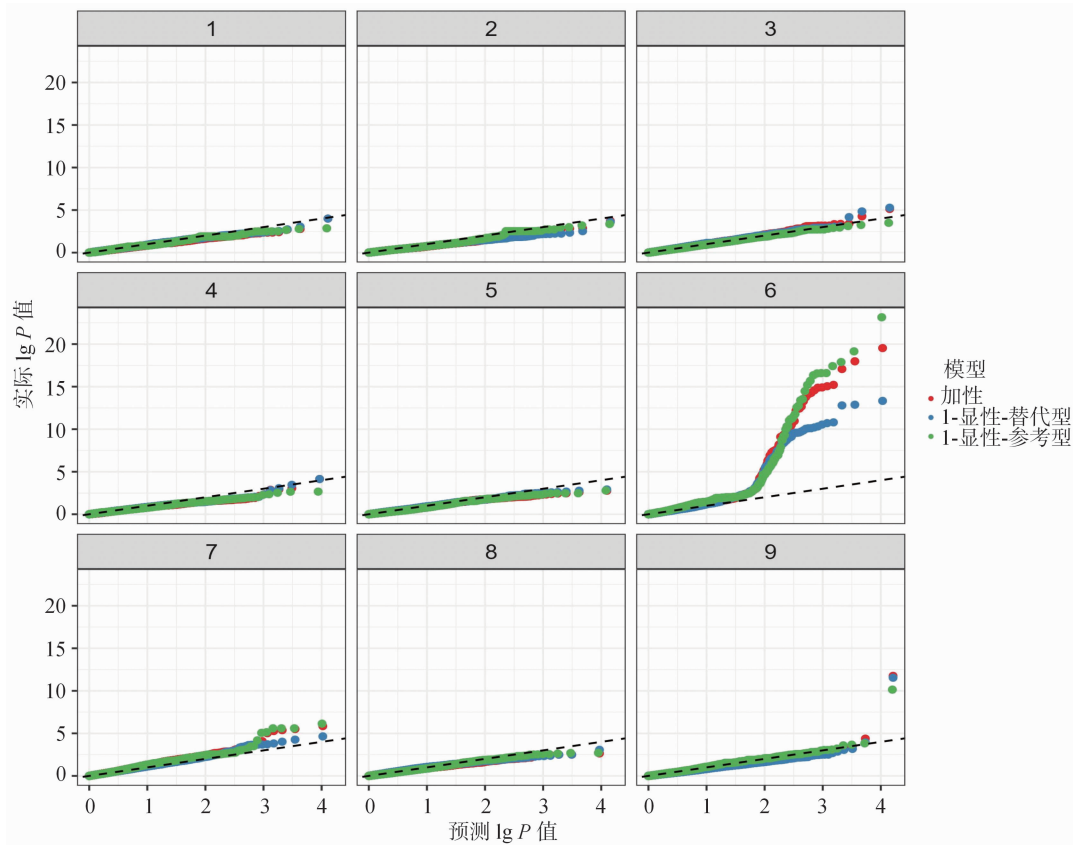


图10 谷子花药颜色的 QQ-Plot 图

表 4 谷子花药颜色相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
6_34378428	6	34 378 428	C	G	19.534 747 140
9_2235798	9	2 235 798	C	T	11.757 592 890
7_26573539	7	26 573 539	G	A	5.852 841 833
3_35865069	3	35 865 069	T	C	5.131 720 082
4_25078258	4	25 078 258	T	C	4.143 290 764
2_27215645	2	27 215 645	C	T	3.721 695 442
1_8028426	1	8 028 426	A	C	2.776 759 817
5_39489042	5	3 948 042	T	C	2.762 724 853
8_30242405	8	30 242 405	G	A	2.642 383 627

抽穗期叶姿关联性最强的 SNP 位点位于 8 号染色体,与谷子刚毛长度关联性最强的 SNP 位点位于 1 号染色体,但与谷子刚毛颜色关联的 SNP 位点在 4 号染色体上最多,与谷子护颖颜色关联性最强的 SNP 位点位于 4 号染色体,与谷子花药颜色关联性最强的 SNP 位点位于 6 号染色体,与谷子茎秆长度关联性最强的 SNP 位点位于 4 号染色体,但与谷子茎秆粗度关联的 SNP 位点在 3 号染色体上最多,与

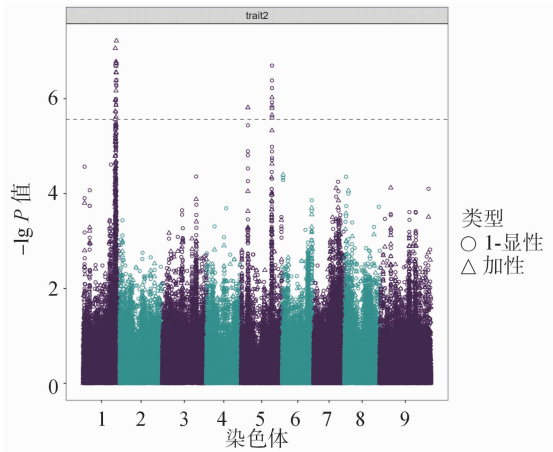


图11 谷子刚毛长度的曼哈顿图

谷子伸长节节间数关联性最强的 SNP 位点位于 4 号染色体。

水稻中控制成熟期性状的基因多,涉及包括植物激素、泛素-蛋白酶体通路、MAPK 信号、表观修饰、G-蛋白信号等分子路径<sup>[22-26]</sup>。本研究中,推测与谷子成熟期性状关联的候选基因功能主要包括细胞构成、跨膜运输、氧化还原酶、DNA 整合、离子结合蛋白、蛋白激酶等。

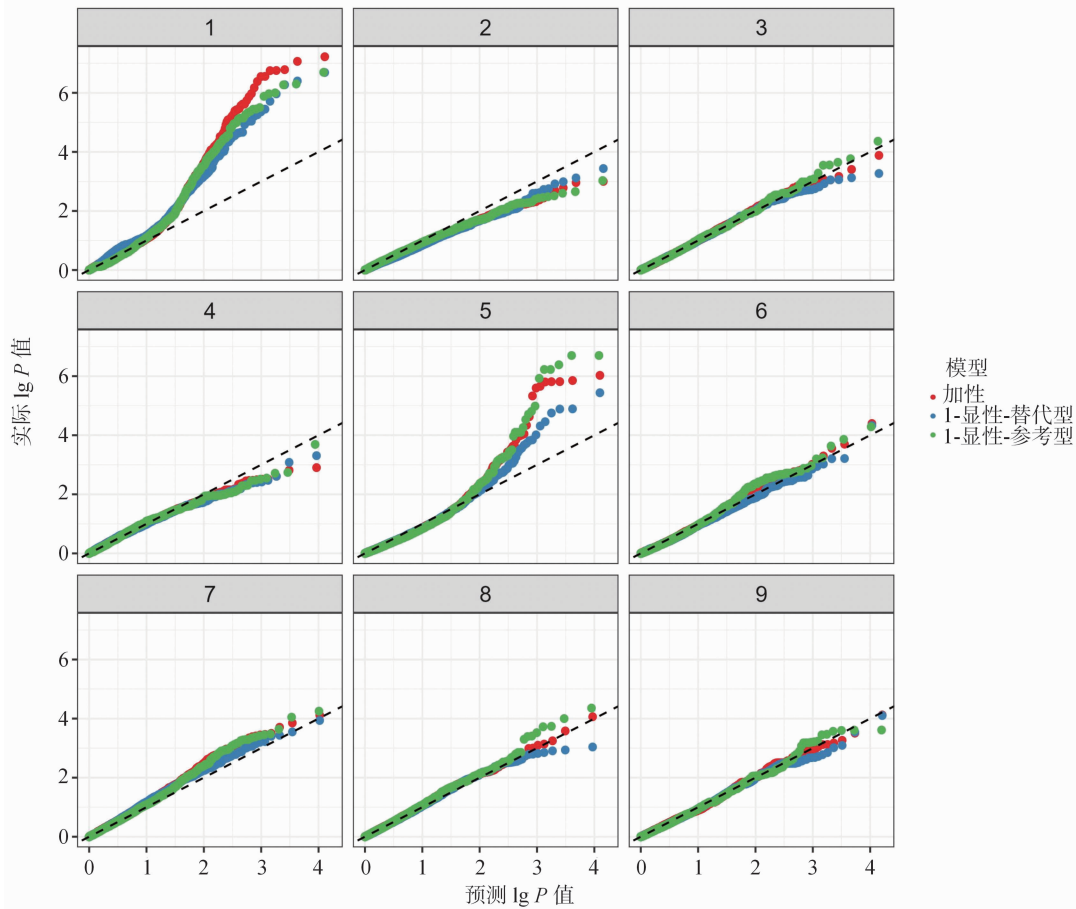


图12 谷子刚毛长度的 QQ-Plot 图

表 5 谷子刚毛长度相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
1_38241023	1	38 241 023	T	C	7.22 1315 299
5_35484548	5	35 484 548	A	G	6.027 084 896
6_1035814	6	1 035 814	G	A	4.396 465 644
9_12721349	9	12 721 349	T	G	4.122 496 713
7_26209186	7	26 209 186	C	T	4.116 026 782
8_4452279	8	4 452 279	A	G	4.062 157 945
3_38971556	3	38 971 556	A	G	3.881 579 749
4_20409866	4	20 409 866	A	G	2.90 3882 723
2_3206660	2	3 206 660	A	G	2.783 058 644

参考文献:

[1] Yang X Y, Wan Z W, Perry L, et al. Early millet use in northern China[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(10): 3726 – 3730.

[2] 贾冠清, 刁现民. 谷子[*Setaria italica* (L.) P. Beauv.] 作为功能基因组研究模式植物的发展现状及趋势[J]. 生命科学, 2017, 29(3): 292 – 301.

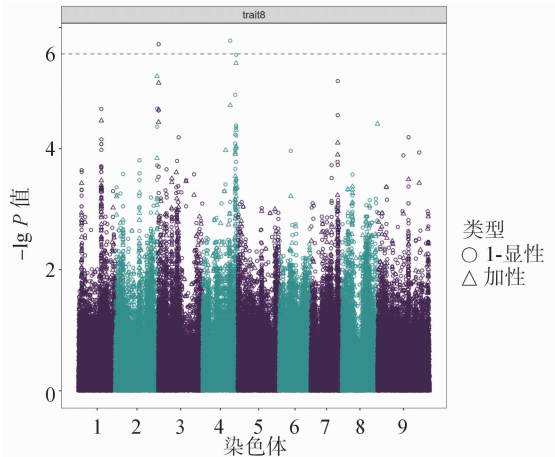


图13 谷子茎秆长度的曼哈顿图

[3] Jia X P, Zhang Z B, Liu Y H, et al. Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. Beauv.][J]. Theoretical and Applied Genetics, 2009, 118(4): 821 – 829.

[4] Jia G, Huang X, Zhi H, et al. A haplotype map of genomic variations and genome – wide association studies of agronomic traits in foxtail millet(*Setaria italica*) [J]. Nature Genetics, 2013, 45(8): 957 – 961.

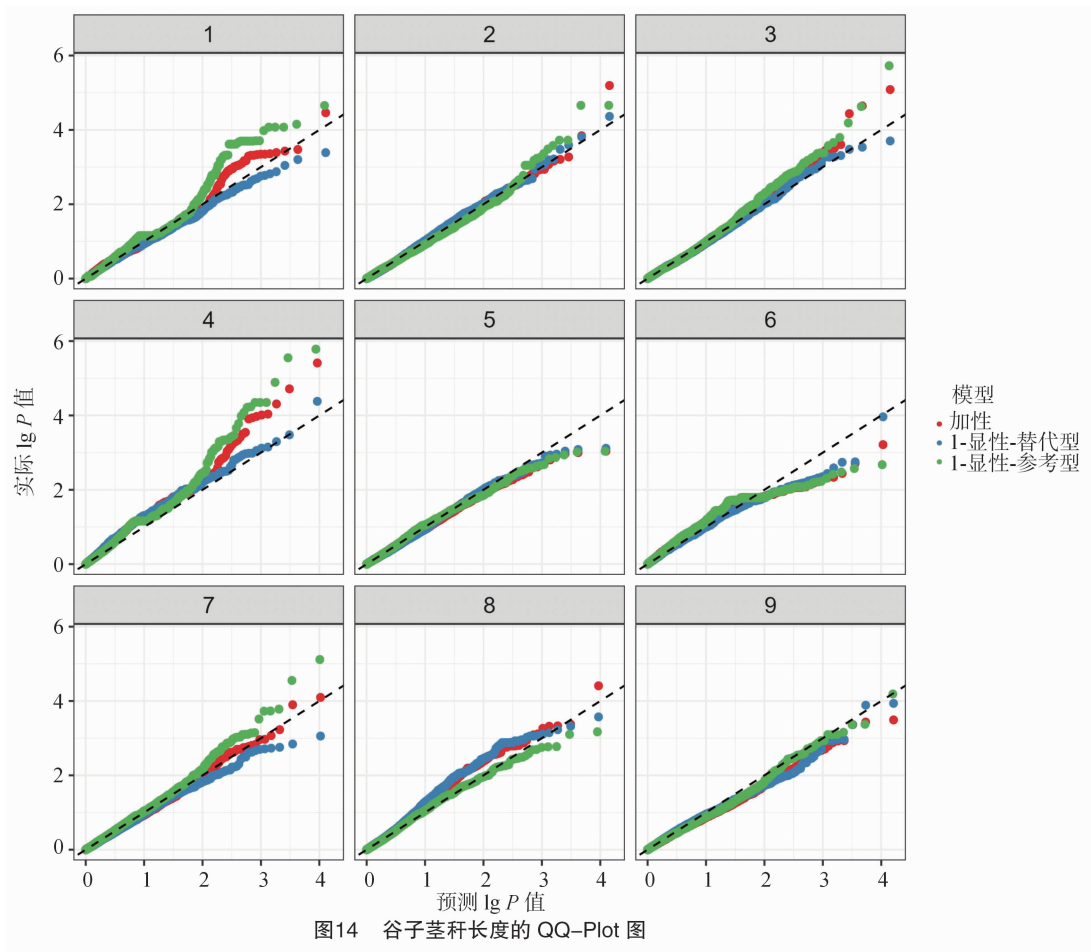


图14 谷子茎秆长度的 QQ-Plot 图

表 6 谷子茎秆长度相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
4_38139078	4	38 139 078	G	T	5.411 861 069
2_47595592	2	47 595 592	A	G	5.1940 773 680
3_235117	3	235 117	A	G	5.084 938 747
1_25952988	1	25 952 988	G	T	4.460 839 510
8_40348504	8	40 348 504	G	A	4.408 632 199
7_30813042	7	30 813 042	A	G	4.096 009 581
9_35310093	9	35 310 093	C	A	3.490 919 024
6_13108538	6	13 108 538	A	G	3.213 856 816
5_6053946	5	6 053 946	G	A	3.038 462 534

[5] 贾小平, 张 博, 董志平, 等. 海南短日照条件下谷子穗部性状的全基因组关联分析[J]. 河南农业科学, 2018, 47(9): 33–40.

[6] Fang X M, Dong K J, Wang X Q, et al. A high density genetic map and QTL for agronomic and yield traits in foxtail millet [*Setaria italica* (L.) P. Beauv.][J]. BMC Genomics, 2016, 17: 336.

[7] Wang J, Wang Z L, Du X F, et al. A high-density genetic map and QTL analysis of agronomic traits in foxtail millet [*Setaria italica* (L.) P. Beauv.] using RAD-seq[J]. PLoS One, 2017, 12(6): e0179717.

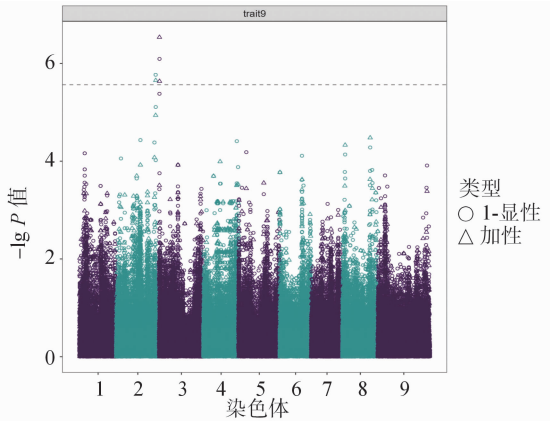


图15 谷子茎秆粗度的曼哈顿图

[8] Zhang K, Fan G Y, Zhang X X, et al. Identification of QTLs for 14 agronomically important traits in *Setaria italica* based on SNPs generated from high-throughput sequencing[J]. G3, 2017, 7(5): 1587–1594.

[9] 元香梅, 禾 璐, 张凯烨, 等. 谷子 XTH 基因家族与抗旱相关基因的分析[J]. 山西农业大学学报(自然科学版), 2017, 37(1): 1–6.

[10] 程 璐, 张 彬, 张耀元, 等. 谷子 MYB 转录因子家族与抗旱关系的研究[J]. 山西农业大学学报(自然科学版), 2016, 36(12): 846–849, 867.

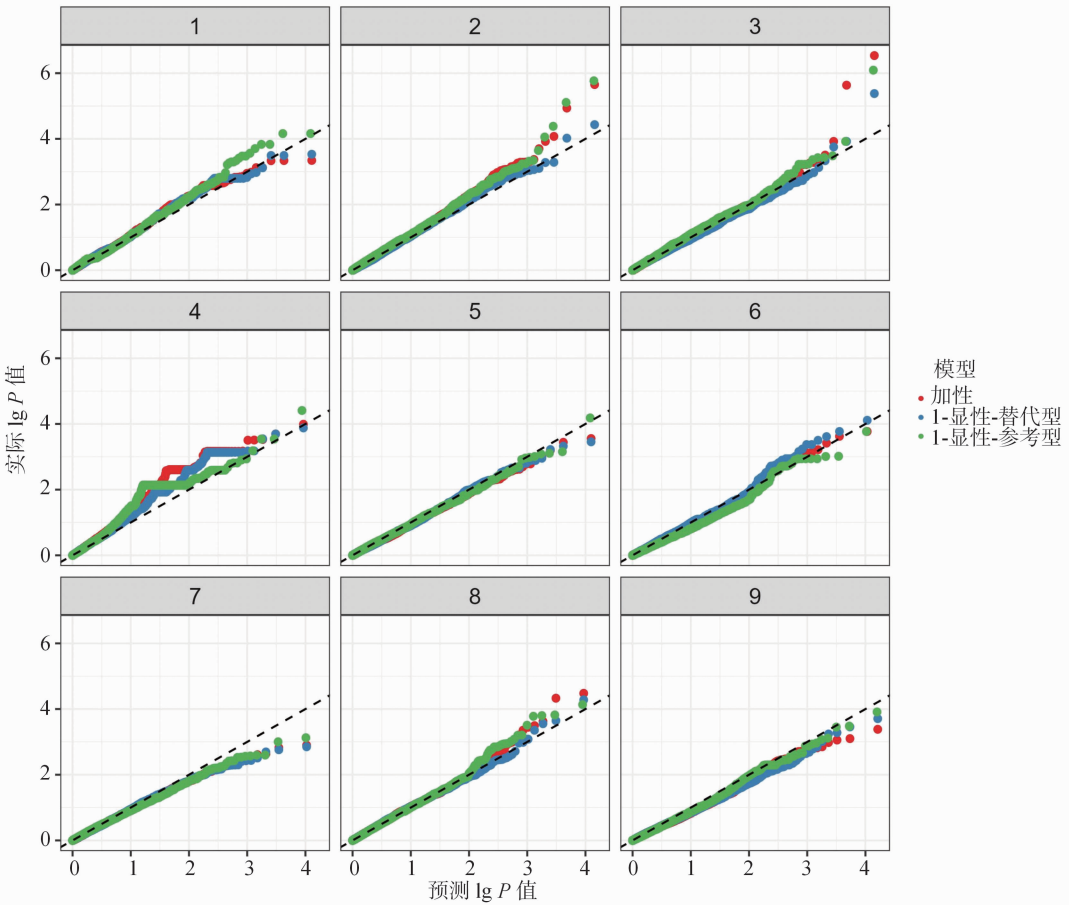


图16 谷子茎秆粗度的 QQ-Plot 图

表 7 谷子茎秆粗度相关关键 SNP 记录

标记	染色体	位置 (bp)	参考 碱基	替代 碱基	关联度
3_167397	3	167 397	C	G	6.535 909 593
2_45076519	2	45 076 519	C	G	5.652 968 264
8_31572478	8	31 572 478	G	T	4.479 039 661
4_19002203	4	19 002 203	G	T	3.990 426 427
6_73447	6	73 447	G	A	3.770 995 599
5_28780439	5	28 780 439	A	G	3.555 235 813
9_56085166	9	56 085 166	C	T	3.383 538 843
1_4724041	1	4 724 041	C	A	3.343 419 668
7_30815797	7	30 815 797	A	G	2.898 473 493

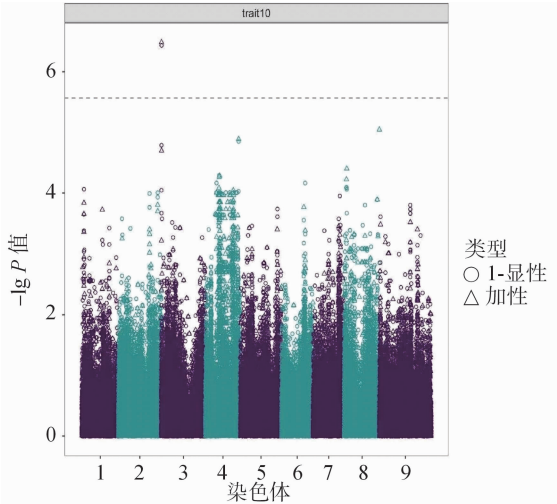


图17 谷子伸长节节间数的曼哈顿图

[11] 贾小平,李剑峰,全建章,等. 不同光周期条件下谷子农艺性状的光周期敏感性评价[J]. 植物遗传资源学报,2018,19(5): 919-924.

[12] 普晓英,曾亚文,申时全,等. 低速磷胁迫对云南地方稻种核心种质抽穗期的影响[J]. 植物遗传资源学报,2004,5(3):271-275.

[13] Gupta S, Kumari K, Muthamilarasan M, et al. Population structure and association mapping of yield contributing agronomic traits in foxtail millet[J]. Plant Cell Reports,2014,33(6):881-893.

[14] Bennetzen J L, Schmutz J, Wang H, et al. Reference genome sequence of the model plant *Setaria* [J]. Nature Biotechnology, 2012,30(6):555-561.

[15] Zhang G Y, Liu X, Quan Z W, et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential [J]. Nature Biotechnology,2012,30(6):549-554.

[16] 曾 莲,张 华,柳 絮,等. 水稻抽穗期基因 *Hd7m* 的定位及遗传分析[J]. 山东农业科学,2015,47(1):10-13.

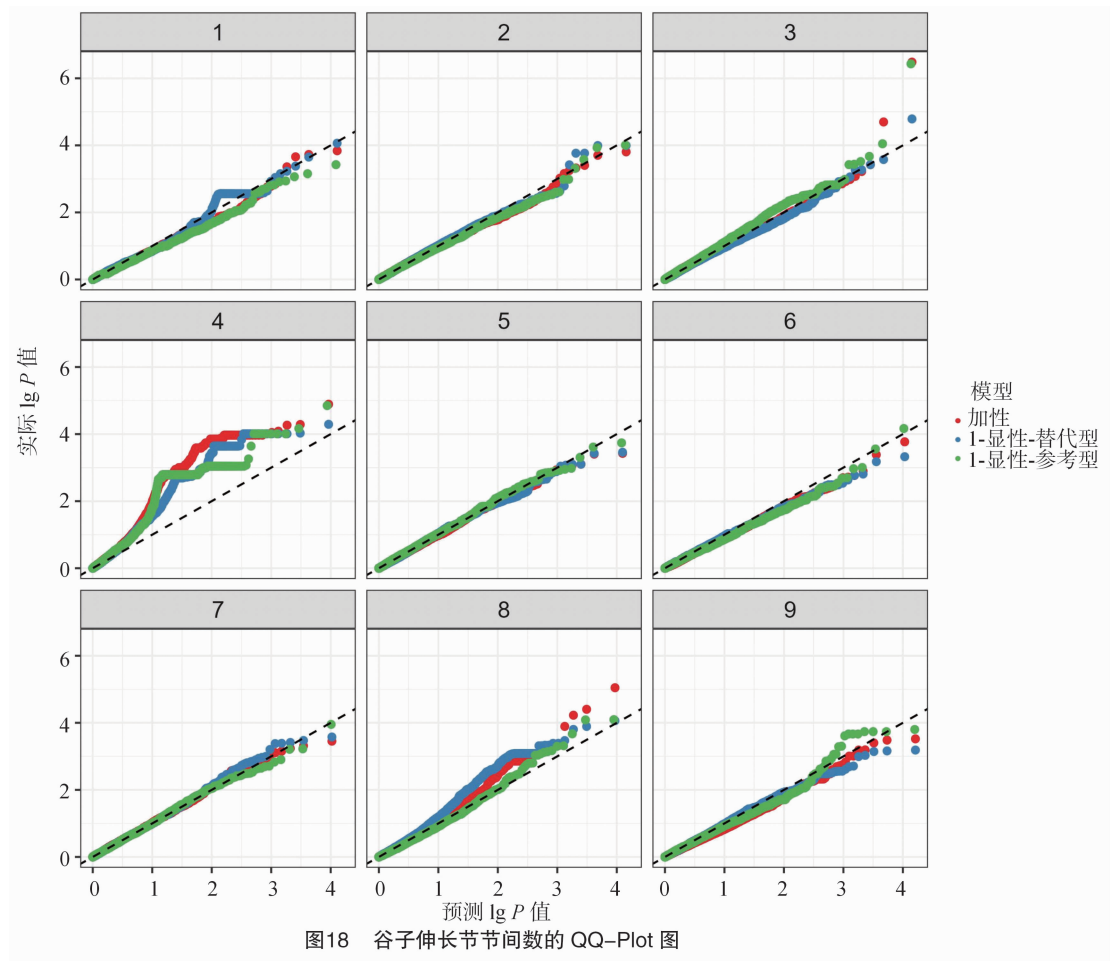


图18 谷子伸长节节间数的 QQ-Plot 图

表 8 谷子伸长节节间数相关关键 SNP 记录

标记	染色体	位置(bp)	参考碱基	替代碱基	关联度
3_167397	3	167 397	C	G	6.482 894 796
8_40348504	8	40 348 504	G	A	5.046 427 542
4_38139078	4	38 139 078	G	T	4.890 536 701
1_2411978	1	2 411 978	A	C	3.838 600 662
2_44977400	2	44 977 400	C	T	3.807 075 879
6_25453849	6	25 453 849	G	C	3.769 628 341
9_35248220	9	35 248 220	C	T	3.521 592 360
7_29582620	7	29 582 620	G	T	3.453 672 011
5_10111842	5	10 111 842	C	T	3.425 540 784

[17]李允振,黄永禄,谢旭阳,等. 水稻抽穗期基因 *EHD8* 的遗传分析及精细定位[J]. 中国科技论文,2017,12(12):1336–1340.

[18]葛永申. 水稻高代回交导入系群体产量相关性状的 QTL 定位及抽穗期基因 *DTH6* 的鉴定[D]. 扬州:扬州大学,2017.

[19]刘国祥. 小麦抽穗期基因 *TaHdM605* 精细作图[D]. 北京:中国农业科学院,2015.

[20]王 燕. 高粱抽穗期基因 *Ma1* 和 *Ma3* 的分子进化及抽穗期的 QTL 分析[D]. 北京:中国农业大学,2015.

[21]谭力治,赵毅强. 全基因组关联分析中混合模型的原理、优化与应用[J]. 中国农业科学,2023,56(9):1617–1632.

[22]任生林,吴才文,经艳芬,等. 全基因组关联分析在作物中的研究进展[J]. 分子植物育种,2024,22(11):3594–3602.

[23]杨 彬,周嘉润,沈玉婷,等. 水稻种子活力性状全基因组关联分析研究进展[J]. 江苏农业科学,2023,51(3):16–21.

[24]Duan P G, Rao Y C, Zeng D L, et al. SMALL GRAIN 1, which encodes a mitogen – activated protein kinase 4, influences grain size in rice [J]. The Plant Journal, 2014, 77 (4) : 547 – 557.

[25]Hu J, Wang Y X, Fang Y X, et al. A rare allele of *GS2* enhances grain size and grain yield in rice [J]. Molecular Plant, 2015, 8 (10):1455–1465.

[26]Li J, Chu H W, Zhang Y H, et al. The rice *HGW* gene encodes a ubiquitin – associated (UBA) domain protein that regulates heading date and grain weight[J]. PLoS One, 2012, 7(3):e34231.

表 9 谷子抽穗期叶姿关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置 (bp)	参考碱基替代碱基		关联度	数据库注释情况
2_3413115	2	3 413 115	G	A	6.634 901 884	钙腔蛋白
2_36097250	2	36 097 250	T	C	5.842 489 507	呼吸爆发氧化酶
2_39937209	2	39 937 209	C	A	7.774 644 443	26S 蛋白酶体调控亚基 T1
2_39987073	2	39 987 073	C	T	6.467 986 803	后期促进复合体亚基 4
2_40030828	2	40 030 828	T	C	6.235 823 424	转运蛋白颗粒复合体亚基 2
2_42647784	2	42 647 784	C	T	6.830 812 010	转录因子 MYB,植物
2_42694544	2	42 694 544	T	C	7.889 239 967	四肽重复蛋白
3_3503201	3	3 503 201	C	T	7.060 199 775	新薄荷醇脱氢酶
3_23490601	3	23 490 601	C	T	9.909 676 656	拟南芥组氨酸激酶 2/3/4(细胞分裂素受体)
3_23491117	3	23 491 117	G	A	6.253 978 581	拟南芥组氨酸激酶 2/3/4(细胞分裂素受体)
3_23512541	3	23 512 541	A	G	5.588 733 753	血红素合成酶
3_43379637	3	43 379 637	T	C	9.946 054 443	F-box 蛋白和富亮氨酸重复蛋白 2/20
4_22349318	4	22 349 318	C	T	5.739 543 485	二氨基羟磷酸核糖氨基嘧啶脱氢酶/ 5-氨基-6-(5-磷酸核糖氨基)尿嘧啶还原酶
4_37187322	4	37 187 322	G	T	5.693 982 810	含脱氢酶结构域的蛋白
5_28258269	5	28 258 269	G	A	7.827 629 571	1,3 半乳糖基转移酶
9_1551542	9	1 551 542	G	A	6.145 756 776	GDP-甘露糖运输车
9_1563933	9	1 563 933	G	A	6.163 803 736	E3 泛素蛋白连接酶
9_9929075	9	9 929 075	T	G	7.715 855 075	鼠李糖半乳糖 I 鼠李糖基转移酶
9_44936346	9	44 936 346	G	A	5.667 119 895	DNA 退火解旋酶和内切酶 ZRANB3
9_44980329	9	44 980 329	C	T	9.523 890 412	拉布 GDP 解离抑制剂

表 10 谷子刚毛长度关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置 (bp)	参考碱基	替代碱基	关联度	数据库注释情况
1_36904322	1	36 904 322	T	C	5.737 220 088	丙酮酸脱氢酶 E1 组分亚基 β
1_36927015	1	36 927 015	A	G	5.972 552 534	核糖体 RNA 加工蛋白
1_36940228	1	36 940 228	C	T	7.064 231 455	纤毛和鞭毛相关蛋白
1_37734436	1	37 734 436	T	C	5.864 089 135	蛋白质 LSM12
1_37748213	1	37 748 213	G	A	6.382 815 500	蛋白酶类 9
1_37751602	1	37 751 602	C	T	6.169 694 380	蛋白酶类 9
1_37752372	1	37 752 372	A	G	6.749 776 790	CCR4-NOT 转录复合体亚基 1
1_37764490	1	37 764 490	G	A	6.556 799 811	丝氨酸蛋白酶抑制剂
1_37769264	1	37 769 264	T	C	5.619 142 308	转录因子 III B 90 ku 亚基
1_37772937	1	37 772 937	G	A	5.617 997 990	转录因子 III B 90 ku 亚基
1_38241023	1	38 241 023	T	C	7.221 315 299	伪响应调节器 5
1_38258488	1	38 258 488	A	G	6.548 928 549	伪响应调节器 5
5_7681034	5	7 681 034	G	A	5.812 636 511	类固醇 3-氧化酶
5_7683014	5	7 683 014	T	G	5.814 572 841	类固醇 3-氧化酶
5_35432787	5	35 432 787	G	T	5.851 796 984	26S 蛋白酶体调控亚基 N9
5_35433569	5	35 433 569	A	G	5.803 960 404	26S 蛋白酶体调控亚基 N9
5_35438502	5	35 438 502	C	T	5.597 748 435	跨膜 9 超家族成员 2/4
5_35455595	5	35 455 595	G	A	5.656 413 903	ADP-核糖基化因子 1/2
5_35484548	5	35 484 548	A	G	6.027 084 896	DNA 拓扑异构酶 II

表 11 谷子刚毛颜色关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置(bp)	参考碱基	替代碱基	关联度	数据库注释情况
4_5504406	4	5 504 406	A	C	8.201 680 302	真核生物翻译起始因子 2C
4_5536398	4	5 536 398	T	C	7.666 349 395	$\beta$ -葡萄糖苷酶
4_5541868	4	5 541 868	G	C	7.146 836 152	丙二酸半醛脱氢酶(乙酰化)/甲基丙二酸半醛脱氢酶
4_5542955	4	5 542 955	T	G	7.306 038 923	丙二酸半醛脱氢酶(乙酰化)/甲基丙二酸半醛脱氢酶
4_5545280	4	5 545 280	C	G	6.660 819 107	丙二酸半醛脱氢酶(乙酰化)/甲基丙二酸半醛脱氢酶
4_5546856	4	5 546 856	G	A	7.004 611 793	丙二酸半醛脱氢酶(乙酰化)/甲基丙二酸半醛脱氢酶
4_5571438	4	5 571 438	T	G	6.455 023 881	翻译起始因子 3 亚基 M
4_5577361	4	5 577 361	A	G	7.003 388 333	DNA 修复蛋白 RadA/Sms
4_5577544	4	5 577 544	C	A	6.987 856 538	DNA 修复蛋白 RadA/Sms
4_6899797	4	6 899 797	G	T	6.173 861 441	泛素 B
4_7031399	4	7 031 399	G	C	7.543 531 432	多肽-二磷酸低聚糖-蛋白糖基转移酶亚基 DDOST/WBP1
4_7064794	4	7 064 794	G	A	5.695 955 283	ATP 结合盒, G 亚族(WHITE), 成员 2
4_7069247	4	7 069 247	C	T	6.776 957 696	乙醇胺磷酸转移酶
4_7077615	4	7 077 615	T	G	6.333 325 728	E3 泛素蛋白连接酶 SIAH1
4_7089111	4	7 089 111	C	T	7.460 531 694	乙醇胺磷酸转移酶
4_7100149	4	7 100 149	A	G	6.371 646 437	LRR 受体样丝氨酸/苏氨酸蛋白激酶
4_7101715	4	7 101 715	T	C	6.795 965 841	赖氨酸组氨酸转运蛋白
4_7106869	4	7 106 869	G	A	6.428 365 447	Ase1/PRC1/MAP65 家族蛋白
4_7118623	4	7 118 623	T	C	6.004 865 850	LRR 受体样丝氨酸/苏氨酸蛋白激酶
4_7124254	4	7 124 254	A	G	7.629 831 730	泛素 C
4_7148347	4	7 148 347	C	A	6.117 025 495	过氧化物酶
4_7149710	4	7 149 710	A	G	5.951 193 996	过氧化物酶
4_7152150	4	7 152 150	T	G	8.979 430 421	甘油醛 3-磷酸脱氢酶(磷酸化)
4_7164916	4	7 164 916	A	G	7.014 921 255	邻氨基甲酸酯合成酶组分 II
4_7188853	4	7 188 853	C	T	7.110 915 888	谷氨酰胺酰-tRNA 合成酶
4_7198273	4	7 198 273	A	G	9.410 187 288	跨膜抗坏血酸依赖还原酶
4_7215041	4	7 215 041	A	T	10.439 166 430	抗病毒解旋酶 SKI2
4_7222372	4	7 222 372	A	G	8.210 834 905	抗病毒解旋酶 SKI2
4_7236888	4	7 236 888	T	C	8.348 424 500	抗病毒解旋酶 SKI2
4_7286633	4	7 286 633	T	C	9.939 099 219	生长素反应蛋白
4_7297172	4	7 297 172	C	T	5.924 921 372	单链 DNA 结合蛋白 WHY2
4_7309945	4	7 309 945	T	G	8.593 316 050	t 复合体蛋白 1 亚基
4_7321520	4	7 321 520	T	C	10.526 955 740	ADP-核糖基化因子 1/2
4_7322957	4	7 322 957	G	A	9.740 637 651	细胞膜蛋白
4_7324751	4	7 324 751	C	T	8.566 362 000	细胞膜蛋白
4_7328327	4	7 328 327	T	C	8.282 989 933	ESCRT-I 复合体亚基 VPS37
4_7340914	4	7 340 914	T	G	10.058 711 550	ADP-核糖基化因子 1/2
4_7341983	4	7 341 983	T	C	6.493 776 203	ADP-核糖基化因子 1/2
4_7344062	4	7 344 062	A	G	7.504 195 643	细胞膜蛋白
4_7356777	4	7 356 777	A	G	5.671 299 177	细胞周期蛋白 K
4_7357094	4	7 357 094	A	G	10.493 265 830	细胞周期蛋白 K
4_7366617	4	7 366 617	A	G	6.753 178 860	翻译因子 GUF1, 线粒体
4_7370286	4	7 370 286	T	C	9.665 803 830	翻译因子 GUF1, 线粒体
4_7373702	4	7 373 702	T	C	10.568 543 980	翻译因子 GUF1, 线粒体
4_7377207	4	7 377 207	T	C	8.548 831 168	核酸酶 S1
4_7378625	4	7 378 625	T	G	10.402 264 300	核酸酶 S1
4_7387053	4	7 387 053	C	T	8.566 362 000	DNA 解旋酶 MCM9
4_7390404	4	7 390 404	G	A	11.141 066 530	DNA 解旋酶 MCM9
4_7390482	4	7 390 482	A	C	10.671 606 820	DNA 解旋酶 MCM9
4_7392347	4	7 392 347	A	T	9.195 221 557	COP9 信号小体复合物亚基 3
4_7395553	4	7 395 553	T	C	11.341 040 490	苏氨酰氨甲酰腺苷
4_7399114	4	7 399 114	C	G	10.406 364 090	苏氨酰氨甲酰腺苷
4_7400384	4	7 400 384	G	T	11.216 171 840	DNA 解旋酶 MCM9
4_7431011	4	7 431 011	T	C	9.294 874 437	聚腺苷化因子亚基 2

表 12 谷子护颖颜色关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置(bp)	参考碱基	替代碱基	关联度	数据库注释情况
4_7069247	4	7 069 247	C	T	5.841 572 929	乙醇胺磷酸转移酶
4_7074178	4	7 074 178	G	A	5.747 179 307	酰基甘油脂肪酶
4_7148347	4	7 148 347	C	A	5.598 360 174	过氧化物酶
4_7149710	4	7 149 710	A	G	5.757 447 809	过氧化物酶
4_7188853	4	7 188 853	C	T	5.650 145 474	谷氨酰胺 tRNA 合成酶
4_7198273	4	7 198 273	A	G	7.935 170 612	跨膜抗坏血酸依赖还原酶
4_7215041	4	7 215 041	A	T	8.751 723 165	抗病毒解旋酶 SKI2
4_7222372	4	7 222 372	A	G	6.983 887 526	抗病毒解旋酶 SKI2
4_7236888	4	7 236 888	T	C	6.431 830 255	抗病毒解旋酶 SKI2
4_7309945	4	7 309 945	T	G	6.139 541 301	20S 蛋白酶体亚基 $\beta$ 2
4_7321520	4	7 321 520	T	C	8.618 308 244	E3 泛素蛋白连接酶 ATL6/9/15/31/42/55
4_7322957	4	7 322 957	G	A	7.699 729 737	苏氨酸脱水酶
4_7324751	4	7 324 751	C	T	7.187 178 972	细胞膜蛋白
4_7328327	4	7 328 327	T	C	8.118 983 193	ESCRT - I 复合体亚基 VPS37
4_7340914	4	7 340 914	T	G	8.651 470 641	ADP - 核糖基化因子 1/2
4_7341983	4	7 341 983	T	C	8.694 630 171	ADP - 核糖基化因子 1/2
4_7344062	4	7 344 062	A	G	6.816 057 182	细胞膜蛋白
4_7357094	4	7 357 094	A	G	8.342 824 580	细胞周期蛋白 K
4_7366617	4	7 366 617	A	G	6.878 530 345	翻译因子 GUF1, 线粒体
4_7370286	4	7 370 286	T	C	7.304 366 972	翻译因子 GUF1, 线粒体
4_7373702	4	7 373 702	T	C	6.881 558 188	翻译因子 GUF1, 线粒体
4_7377207	4	7 377 207	T	C	6.464 965 104	核酸酶 S1
4_7378625	4	7 378 625	T	G	7.949 762 438	核酸酶 S1
4_7387053	4	7 387 053	C	T	7.187 178 972	DNA 解旋酶 MCM9
4_7390404	4	7 390 404	G	A	8.215 286 996	DNA 解旋酶 MCM9
4_7390482	4	7 390 482	A	C	7.987 813 113	DNA 解旋酶 MCM9
4_7392347	4	7 392 347	A	T	6.239 046 817	COP9 信号小体复合物亚基 3
4_7395553	4	7 395 553	T	C	8.802 841 933	N6 - 氨基苏氨酸腺苷合酶
4_7399114	4	7 399 114	C	G	7.381 317 194	N6 - 氨基苏氨酸腺苷合酶
4_7400384	4	7 400 384	G	T	8.259 193 949	DNA 解旋酶 MCM9

表 13 谷子花药颜色关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置(bp)	参考碱基	替代碱基	关联度	数据库注释情况
6_33324856	6	33 324 856	A	C	7.465 401 812	钙结合蛋白
6_33332369	6	33 332 369	T	C	7.527 322 928	长链酰基辅酶 a 合成酶
6_33337310	6	33 337 310	A	G	6.274 853 161	长链酰基辅酶 a 合成酶
6_33345652	6	33 345 652	T	C	5.936 733 873	乙酰谷氨酸激酶
6_33994198	6	33 994 198	G	A	8.174 464 835	I 类果糖二磷酸醛缩酶
6_34177699	6	34 177 699	C	T	7.536 977 258	锌指蛋白 IDD3/4/5/6/7/8/9/10/11/12/13
6_34183654	6	34 183 654	C	T	9.988 802 927	溶质载体家族 15( 多肽/组氨酸转运蛋白), 成员 3/4
6_34187945	6	34 187 945	C	T	9.957 015 822	ATP 结合盒, B 亚族( MDR/TAP), 成员 1
6_34192624	6	34 192 624	T	C	9.542 619 665	ATP 结合盒, B 亚族( MDR/TAP), 成员 1
6_34193647	6	34 193 647	G	T	6.509 513 023	ATP 结合盒, B 亚族( MDR/TAP), 成员 1
6_34193934	6	34 193 934	G	A	10.515 598 040	ATP 结合盒, B 亚族( MDR/TAP), 成员 1
6_34197840	6	34 197 840	T	C	8.260 534 819	锌指蛋白 IDD3/4/5/6/7/8/9/10/11/12/13

表 13(续)

标记	染色体	位置(bp)	参考碱基	替代碱基	关联度	数据库注释情况
6_34295674	6	34 295 674	C	T	12.434 073 060	氯通道 7
6_34314885	6	34 314 885	A	G	7.0563 648 640	过氧化物酶
6_34318301	6	34 318 301	A	G	12.424 566 360	氯通道 7
6_34335007	6	34 335 007	C	T	9.311 742 523	乙醇胺磷酸胞苷酰基转移酶
6_34338113	6	34 338 113	T	C	9.140 362 664	MAP 激酶底物 1
6_34378428	6	34 378 428	C	G	19.534 747 140	多药耐药蛋白,MATE 家族
6_34380043	6	34 380 043	C	T	17.979 729 030	多药耐药蛋白,MATE 家族
6_34385609	6	34 385 609	C	T	17.069 963 690	溶血磷脂酰基转移酶
6_34401163	6	34 401 163	G	A	9.373 659 546	多药耐药蛋白,MATE 家族
6_34404693	6	34 404 693	A	G	9.107 007 509	GPI 锚定转氨酶亚基 T
6_34413170	6	34 413 170	G	A	13.281 626 090	溶质载体家族 25(线粒体磷酸转运蛋白),成员 23/24/25/41
6_34415325	6	34 415 325	T	C	6.938 090 649	ATP 依赖性 RNA 解旋酶 DDX3X
6_34418102	6	34 418 102	C	T	15.048 269 610	ATP 依赖性 RNA 解旋酶 DDX3X
6_34427379	6	34 427 379	A	G	15.208 395 960	甲氧苄啶氨基肽酶
6_34428312	6	34 428 312	C	A	6.9113 166 960	甲氧苄啶氨基肽酶
6_34430322	6	34 430 322	T	C	14.875 277 820	大亚基核糖体蛋白 L38e
6_34476417	6	34 476 417	T	C	14.908 275 410	小亚基核糖体蛋白
6_34503085	6	34 503 085	C	T	7.420 983 500	多位点特异 tRNA 甲基转移酶
6_34507823	6	34 507 823	A	G	9.584 373 409	多位点特异 tRNA 甲基转移酶
6_34512966	6	34 512 966	T	C	7.396 036 958	LA 相关蛋白 7
6_34550336	6	34 550 336	G	A	14.566 802 040	转运蛋白 SEC24
6_34578973	6	34 578 973	G	A	13.764 801 540	机械敏感离子通道蛋白 1/2/3
6_34587229	6	34 587 229	C	T	10.166 209 260	3-酮脂酰 CoA 硫解酶合酶
6_34605459	6	34 605 459	G	A	12.665 921 070	丝氨酸-苏氨酸激酶受体相关蛋白
6_34613893	6	34 613 893	G	A	12.204 072 720	热休克蛋白 90 ku $\beta$
6_34629991	6	34 629 991	C	T	14.263 591 340	真核生物翻译起始因子 2C
6_34644290	6	34 644 290	A	T	6.444 938 724	真核生物翻译起始因子 2C
6_34663112	6	34 663 112	T	C	7.847 221 480	蛋白磷酸酶 PTC2/3
6_34675926	6	34 675 926	A	G	5.763 200 576	DNA(胞嘧啶-5)-甲基转移酶
6_34735070	6	34 735 070	T	C	7.551 506 022	核糖体组装蛋白 RRB1
6_34749844	6	34 749 844	C	T	7.3001 948 460	16S rRNA(胞嘧啶 967- $\epsilon$ 5)-甲基转移酶
6_34763325	6	34 763 325	G	A	7.672 234 924	通过与 t-SNAREs 的相互作用进行囊泡运输
7_26573539	7	26 573 539	G	A	5.8528 418 330	Rho GDP 解离抑制剂
9_2235798	9	2 235 798	C	T	11.757 592 890	沃纳综合征 ATP 依赖性解旋酶

表 14 谷子茎秆粗度关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置 (bp)	参考碱基	替代碱基	关联度	数据库注释情况
2_45076519	2	45 076 519	C	G	5.652 968 264	热休克蛋白 90 ku $\beta$
3_167397	3	167 397 000	C	G	6.535 909 593	RNA 聚合酶相关蛋白 LEO1
3_235117	3	235 117 000	A	G	5.634 794 093	过氧化物酶

表 15 谷子伸长节间数关联区域检测候选基因位置及其在数据库的注释情况

标记	染色体	位置 (bp)	参考碱基	替代碱基	关联度	数据库注释情况
3_167397	3	167 397	C	G	6.482 894 796	RNA 聚合酶相关蛋白 LEO1