

小麦籽粒相关性状全基因组关联分析

张志辉, 程宇坤, 任毅, 王睿, 董一帆, 耿洪伟

(新疆农业大学农学院/新疆农业大学优质专用麦类作物工程技术研究中心, 乌鲁木齐 830052)

摘要: 小麦籽粒大小和形态是决定产量的主要因素之一, 挖掘籽粒大小和形态性状的关联位点, 筛选相关候选基因对于提高小麦产量具有重要意义。本研究以国内外具有代表性的 300 份冬小麦自然群体为研究材料, 对千粒重、粒长、粒宽、粒厚、籽粒长宽比、籽粒面积、籽粒周长、籽粒形状和籽粒饱满度等 9 个籽粒性状进行表型鉴定, 利用小麦 90 K SNP 芯片进行基因型采集, 通过混合线性模型 (MLM+Q+K) 的方法对籽粒大小和形态性状进行全基因组关联分析。研究结果表明, 小麦籽粒大小和形态性状表现出丰富的表型变异, 变异系数范围 3.80%~26.06%, 广义遗传力在 56.25%~91.89% 之间。通过 GWAS 检测出 66 个与籽粒大小和形态相关的稳定关联位点 ($P \leq 0.001$), 分布在除 3D、4D、5D 外的 18 条染色体上, 可解释 3.74%~14.34% 的表型变异。检测到 37 个与两个及以上籽粒性状关联的一因多效位点, 其中 3B 染色体的 BS00022512_51 标记同时与 4 个籽粒性状 (粒长、粒宽、粒厚和籽粒长宽比) 关联, 具有最大的表型贡献率 (7.06%~14.34%), 6D 染色体的 wsnp_Ex_c4480_8055475 标记同时与除粒厚、籽粒形状和籽粒饱满度以外 6 个籽粒性状关联, 表型贡献率为 3.81%~8.25%。将 BS00022512_51 和 wsnp_Ex_c4480_8055475 标记进行单倍型分析, 发现位于 6D 染色体上的 wsnp_Ex_c4480_8055475 位点存在 GC-Hap1、AT-Hap2 和 AC-Hap3 三种单倍型, 单倍型 GC-Hap1 为籽粒较大的高千粒重单倍型。三种单倍型的分布频率分别为 65.58%、32.25% 和 2.17%, 单倍型 GC-Hap1 在中国四个冬麦区品种 (系) 中被大量选育。对 37 个一因多效位点进行发掘, 筛选到 9 个籽粒大小和形态性状相关的候选基因。

关键词: 小麦; 籽粒大小和形态; 单倍型; GWAS; 候选基因

Genome-wide Association Analysis of Wheat Grain related Traits

ZHANG Zhi-hui, CHENG Yu-kun, REN Yi, WANG Rui, Dong Yi-fan, GENG Hong-wei

(College of Agronomy/Special High Quality Triticeae Crops Engineering and Technology Research Center, Xinjiang Agricultural University Urumqi 830052)

Abstract: Grain size and morphology is one of the main factors that determine the wheat yield. It is of great significance to mine the associated loci of grain size and morphology related traits and screen related candidate genes for improving wheat yield. In this study, 300 representative natural populations of winter wheat were used to identify the phenotype at nine grain traits, including 1000-grain weight, grain length, grain width, grain thickness, grain length-width ratio, grain area, grain perimeter, grain shape and grain plumpness. This collection was genotyped by 90K SNP chip, followed by genome-wide association study using mixed linear model (MLM+Q+K). The results showed that these traits showed rich phenotypic variation, with the coefficient of variation ranged from 3.80%~26.06% and the broad heritability from 56.25% to 91.89%. Sixty-six loci ($P \leq 0.001$) were detected by GWAS on 18 chromosomes except 3D, 4D and 5D, explaining 3.74%~14.34% of phenotypic variation at each locus. Thirty-seven pleiotropic loci were found associating with two or more grain traits. Especially, the BS00022512_51 marker on chromosome 3B was simultaneously correlated with 4 grain traits (grain length, grain width, grain thickness and grain length-width ratio), with the largest apparent contribution rate (7.06%~14.34%). The wsnp_Ex_c4480_8055475 markers on chromosome 6D were simultaneously associated with six grain traits except grain thickness, grain shape and grain plumpness, and the phenotypic contribution rate was 3.81%~8.25%. Haplotype analysis of BS00022512_51 and wsnp_Ex_c4480_8055475 markers showed that there were three haplotypes of GC-Hap1,

¹收稿日期: 修回日期: 网络出版日期:

URL:

第一作者研究方向为小麦遗传育种, E-mail: zhangzhihui0105@163.com

通信作者: 耿洪伟, 研究方向为小麦遗传育种, E-mail: hw-geng@163.com

基金项目: 新疆自治区重大科技专项 (2022A03004-3)

Foundation Project: Xinjiang Autonomous Region Major Science and Technology Project (2022A03004-3)

AT-Hap2 and AC-Hap3 in wsnp_Ex_c4480_8055475 locs on chromosome 6D, and Haplotype GC-Hap1 is a haplotype with large grains and high 1000-grain weight. The three haplotypes accounted for 65.58% and 32.25% and 2.17% in population, respectively. Haplotype GC-Hap1 was predominant in four winter wheat regions in China. Nine candidate genes were proposed based on 37 significantly-associated multiple effect SNPs.

Key word: Wheat; Grain size and morphology; Haplotype; GWAS; Candidate genes

小麦 (*Triticum aestivum* L.) 是全球分布最广的粮食作物之一, 以蛋白质含量高、口感好、品质优良称著^[1]。随着世界经济的发展和人口的激增, 到 2050 年小麦产量需提高约 60% 方能满足人们对粮食的需求^[2]。籽粒性状能直接影响小麦的产量和品质, 也是普通小麦在驯化过程中受到选择的基本性状之一^[3]。小麦产量构成三要素中的粒重常用千粒重来表示, 主要由籽粒大小 (粒长、粒宽、粒厚) 等基本要素来反映^[4]。不同小麦品种籽粒大小和粒重存在显著遗传变异, 籽粒大小可导致产量的增加, 同时也影响着幼苗活力^[5]。籽粒的长宽厚不仅直接影响小麦产量, 而且通过影响籽粒形态间接影响着小麦品质, 大粒、球形、质地均匀的籽粒深受农户和厂商的青睐, 往往拥有更高的市场价值^[6-7]。因此, 研究籽粒大小和形态相关性状对于培育高产优质小麦新品种具有重要意义。

标记辅助选择 (MAS) 能从分子水平上快速准确的分析个体的遗传组成, 其有效性取决于可用基因的数目和标记-性状的关联程度^[8]。与传统育种相比, MAS 可以直接通过基因选择加快育种效率, 不受环境条件影响。籽粒大小和形态相关性状是复杂的数量性状, 主要受遗传因素影响, 同时也受到环境的制约^[9]。以往对小麦籽粒大小和形态性状的研究多通过连锁分析检测相关 QTL, 如周小鸿等^[10]以西藏半野生小麦 Q1028 和郑麦 9023 组成的 RIL 群体为研究材料, 对 3 个年度环境中 5 个籽粒性状进行 QTL 定位, 发现 13 个粒长、6 个粒宽、2 个长宽比、3 个粒厚和 9 个籽粒大小的 QTL, 可解释 4.64%~14.69% 的表型变异。邹林翰等^[11]以周 8425B×小偃 81 衍生的含有 102 个家系的 RIL 群体为研究材料, 对 3 个环境下的 4 个籽粒性状进行 QTL 定位, 共检测出 22 个 QTL 位点, 可解释 4.77%~19.95% 的表型变异, 同时结合自然群体材料鉴定出两个与千粒重相关的单倍型优势类群。近年来, 全基因组关联分析因不针对特定性状, 省去了 QTL 作图群体构建时的费时费力, 而且通过多个世代的周期重组提供了广泛的变异, 还对于复杂性状的捕捉具有更高的分辨能力^[12-13], 已被应用于许多作物的重要位点及基因的挖掘研究中^[14-16]。张芳等^[5]利用 50 K SNP 芯片对 121 份在新疆种植的小麦品种 (系) 的千粒重、粒长、粒宽、籽粒周长、籽粒面积和籽粒长宽比进行了 GWAS, 在两个及以上环境中重复检测到 29 个 SNP 位点, 这些位点分布在 1A、1B、1D、2A、3B、5A、5D、6B、6D、7B 和 7D 染色体上, 可解释 9.30%~22.70% 的表型变异。Wang 等^[17]以 198 份小麦品种 (系) 为研究材料, 利用 35 K SNP 芯片对 6 个环境下的 5 个籽粒性状进行全基因组关联分析, 共鉴定出 146 个标记-性状关联 (MTA) 位点, 可解释表型变异的 5.91%~15.02%, 在发现的位点中有 12 个 MTA 在多种环境中被鉴定到。Zhao 等^[18]以 211 份节节麦为研究材料, 利用 55 K SNP 芯片对连续两年 5 个环境下的 4 个籽粒

性状进行全基因组关联分析，基于 BLUE 值的关联分析共鉴定出 139 个关联位点，可解释粒长、粒宽、籽粒周长、籽粒长宽比的表型变异率为 5.86%~17.71%、6.83%~8.42%、5.99%~17.71%和 5.80%~19.61%，并对 26 个稳定位点利用双尾 t 检验进行验证，鉴定出 6 个与籽粒相关的候选基因。

虽然已经报道了一些小麦籽粒大小和形态相关的 QTL，但鉴于小麦基因组的规模庞大（约 17G）及复杂性，对籽粒大小和形态相关性状的分子调控机制的研究还是非常有限的，还需要进一步挖掘其调控基因及进行标记鉴定。本研究以国内外 300 份小麦品种（系）为研究材料，对小麦籽粒大小和形态相关性状进行全基因组关联分析，挖掘影响小麦籽粒大小和形态的稳定 QTL 位点，将挖掘到与多性状显著关联的稳定主效 QTL 位点进行二次挖掘，寻找对籽粒性状有利的单倍型，并分析影响籽粒性状的单倍型在不同生态区的分布规律，同时根据基因注释筛选潜在候选基因，以期小麦籽粒大小和形态的遗传改良和标记辅助选择提供理论依据。

1 材料与方法

1.1 试验材料和田间试验

用于关联分析的试验材料为 300 份冬小麦品种（系）组成的自然群体。该群体包含 65 份国外引进品种（系），121 份黄淮冬麦区的品种（系），51 份北部冬麦区的品种（系），41 份长江中下游冬麦区品种（系）和 22 份西南冬麦区品种（系），均具有较好的代表性。上述试验材料由中国农业科学院作物科学研究所小麦品质课题组夏先春研究员惠赠，所有材料均于 2018-2019 和 2019-2020 连续两个年度种植于新疆农业科学院玛纳斯综合试验站（N 44° 17′，E 86° 15′，海拔 400 m），两个环境分别记为 E1 和 E2。田间试验采用随机区组设计，设置 3 次重复，3 行种植，行长 2.0 m，行间距 20.0 cm，人工点种，田间管理遵循当地小麦生产基本田间管理方式。

1.2 表型测定

小麦成熟时随机选择 30 个代表性穗进行人工脱粒，使用万深 SC-G 自动考种分析仪检测千粒重（TKW）、粒长（GL）、粒宽（GW）、籽粒长宽比（LWR）、籽粒面积（GA）和籽粒周长（GC），每份材料选择不少于 500 粒；使用数显游标卡尺（精读为 0.02 mm）测量粒厚（GT）。

参照《小麦种质资源描述规范和数据标准》^[19]对籽粒形状（KS）按照籽粒清选后，随机抽取约 100 粒，采用目测法观察。根据观察结果并参照粒型模式图，确定种质籽粒的形状，籽粒形状评价参数为：1：长型；2：卵型；3：椭圆型；4：圆型。籽粒饱满度（GP）按照籽粒清选后，随机选取 100 粒左右，采用目测法观察。根据观察结果及下列说明，确定种质籽粒的饱满度。1：不饱满（籽粒未被胚乳充满，种皮有明显凹陷，籽粒瘪瘦），2：中等（籽粒基本被胚乳充满，种皮略有凹陷），3：饱满（籽粒完全被胚乳充满，种皮无凹陷）。

1.3 表型数据分析

采用 Excel 2019、IBM SPSS Statistics 20 和 Origin 2021 软件对千粒重、粒长、粒宽、粒厚、籽粒面积、籽粒周长、籽粒长宽比、籽粒形状和籽粒饱满度 9 个小麦籽粒大小和形态相关性状进行统计分析和相关性分析。采用 QTL IciMapping v4.1 软件中的 ANOVA 功能进行遗传力的估算。

1.4 基因型测定及连锁不平衡分析

本研究采用小麦 90 K SNP 芯片对 300 份冬小麦品种（系）进行基因型测定，90 K SNP 芯片由 Illumina 公司开发，包含 81587 个 SNP 标记。课题组前期完成 300 份品种（系）的数据质量控制，将杂合基因型作为缺失数据、过滤掉缺失率大于 20% 的标记，并去除最小等位基因频率（MAF, minor allele frequency）小于 5% 的标记，最后保留 16710 个高质量 SNP 标记用于进行全基因组关联分析。以 r^2 作为衡量多态性位点连锁不平衡（LD, linkage disequilibrium）参数，运用 Tassel v5.0 软件进行 LD 估算，将第 95 百分位处的 r^2 值与 LD 衰减曲线的交叉点作为 LD 衰减距离，染色体组平均 LD 衰减距离用于判断关联分析中显著关联 SNP 标记间的物理位置能否划分为同一位点。

1.5 群体结构

群体结构分析采用 Power Maker V3.25 软件计算多态性信息量（PIC, polymorphic information content, $PIC=1-\sum P_{ij}^2$, P_{ij} 表示第 i 个位点的第 j 个等位变异出现的频率）^[20]。从筛选过的标记中，选取 2000 个在 21 条染色体上均匀分布的 SNP 标记，通过 Structure V2.3.4 软件对进行群体结构分析，亚群数目 $K=2\sim 12$ ，每个 K 值重复 5 次，通过 Structure Harvester 在线分析软件获得最佳 K 值下的亚群数目。使用 Tassel v5.0 软件^[21]进行主成分分析，将 PCA 分析结果前三个主成分基于 Python v3.8 的 Axes3D 程序包完成数据可视化。

1.6 全基因组关联分析

基于筛选获得的高质量 SNP 标记，对 300 份小麦品种（系）的籽粒大小和形态相关性状在每个环境下的均值及 BLUE 值进行 GWAS。为了避免亲缘关系和群体结构造成假阳性，采用 TASSEL v5.0 软件中的混合线性模型（MLM, mixed linear model）进行关联分析。以 $-\log_{10}(p)=3$ 作为 SNP 标记与目标性状关联结果的显著性阈值，将至少在两个环境中检测到的位点视为稳定位点。GWAS 结果基于 R v4.2.0 中的 CMplot 程序包绘制，曼哈顿图（Manhattan plot）展示标记的分布情况， $Q-Q$ 图（Quantile-quantile plot）用于评估关联分析结果的准确性。

1.7 单倍型分析及候选基因筛选

利用 Haploview v4.2 软件对所有环境中检测到表型贡献率大，且同时与 4 个及以上籽粒性状显著关联的 SNP 位点进行单倍型分析，通过在线网站 HILOT (<https://hiplot-academic.com/>) 对籽粒表型值进行小提琴图绘制，剔除 SNP 杂合的材料。将在两个及以上环境中检测到与多个籽粒性状关联的 SNP 序列为探针，在小麦基因组数据库（IWGSC, CS RefSeq v2.1, <https://www.wheatgenome.org/>)、NCBI (<http://www.ncbi.n>

lm.nih.gov/) 及 ENA (<https://www.ebi.ac.uk/ena>) 等数据库中检索比对, 确定候选基因并对其进行功能注释。

2 结果与分析

2.1 小麦籽粒大小和形态性状表型数据

通过对小麦籽粒大小和形态性状的表型分析 (表 1), 显示 300 份小麦品种 (系) 籽粒大小和形态性状在不同环境中呈连续分布。各性状在不同环境中均具有较大的表型变异, 变异系数范围为 3.80%~26.06% 之间, 广义遗传力适中, 处于 56.25%~91.89% 之间, 表明这些性状的表达受到遗传因素和环境因素的双重影响。峰度值中, E2 环境中籽粒面积的峰度值最大, 说明该环境中籽粒面积性状正态分布相对陡峭, 存在部分极端值。偏度值中, 籽粒性状的偏度的绝对值都处于 1 以下, 除粒宽的 BLUE 值 (-0.76) 略微左偏外, 其余性状在偏度上均具有较好的表现。相关性分析中发现 (图 1), 除了粒长、籽粒面积、籽粒周长与籽粒饱满度, 籽粒长宽比、籽粒形状与籽粒面积无显著相关关系外, 其余性状之间均呈显著 ($P \leq 0.01$) 相关关系。其中, 粒长和籽粒周长正相关系数最大 ($r=0.95$), 籽粒长宽比和籽粒形状负相关系数最大 ($r=0.70$)。以上结果表明, 本研究中 9 个籽粒性状是多基因控制的数量性状, 符合正态分布, 适合用于全基因组关联分析。

表 1 小麦籽粒大小和形态性状表型数据分析和遗传力

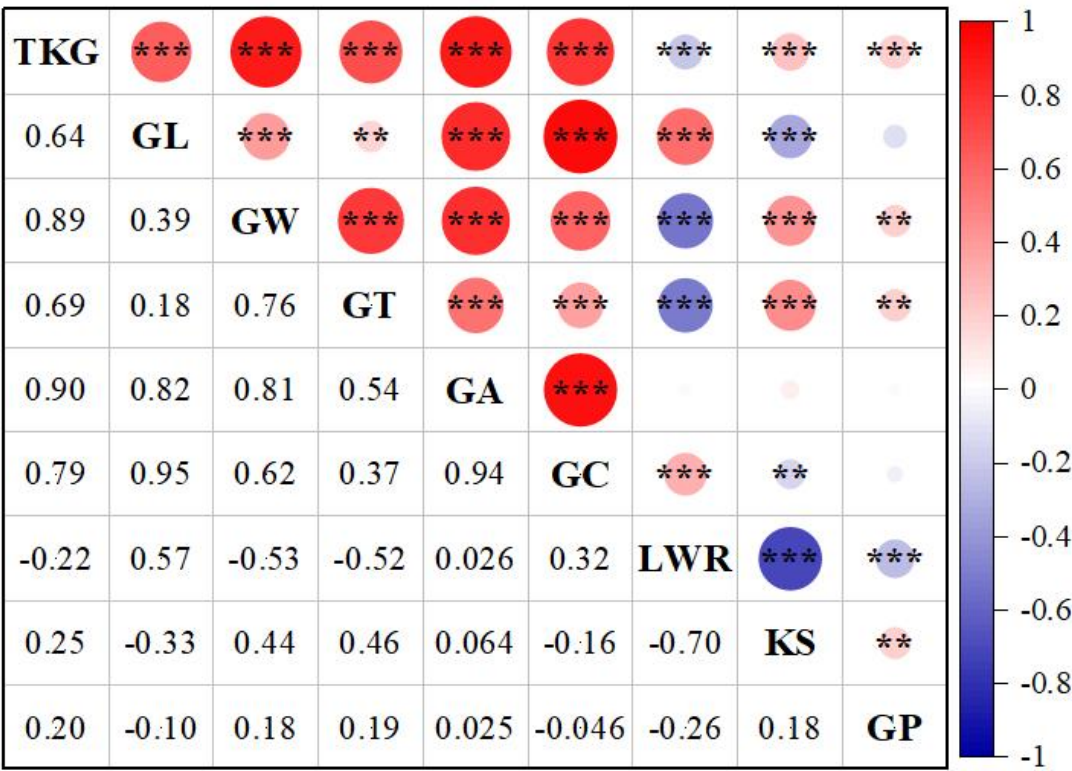
Table 1 Phenotypic data analysis and heritability of grain size and morphological traits in wheat

性状	环境	变异范围	均值±标准差	偏度	峰度	变异系数	遗传力
Trait	Environment	Range	Mean±SD	Ske.	Kur.	CV(%)	H ² (%)
千粒重	E1	24.71—53.84	42.4±4.58	-0.38	0.40	10.81	84.75
	E2	26.54—57.49	43.64±4.54	-0.46	0.73	10.41	
	BLUE	25.58—53.11	42.97±4.27	-0.60	0.73	9.95	
粒长	E1	5.60—7.52	6.63±0.29	0.07	0.38	4.39	91.98
	E2	5.23—7.11	6.3±0.29	-0.03	0.69	4.62	
	BLUE	5.47—7.20	6.47±0.28	-0.02	0.50	4.35	
粒宽	E1	2.90—3.79	3.47±0.16	-0.64	0.39	4.54	86.40
	E2	2.82—3.76	3.36±0.15	-0.56	0.77	4.45	
	BLUE	2.86—3.71	3.42±0.14	-0.76	0.83	4.21	
粒厚	E1	2.62—3.68	3.13±0.16	-0.20	0.59	5.12	75.69
	E2	2.64—3.68	3.12±0.16	-0.12	0.86	5.09	
	BLUE	2.65—3.52	3.12±0.14	-0.36	0.37	4.50	
籽粒面积	E1	12.53—21.14	17.89±1.29	-0.35	0.67	7.23	84.74
	E2	11.51—23.49	16.41±1.35	0.56	3.89	8.20	
	BLUE	12.14—21.38	17.15±1.22	-0.17	1.49	7.12	
籽粒周长	E1	14.73—19.44	17.55±0.68	-0.19	0.83	3.86	87.55
	E2	14.25—20.36	17.04±0.73	0.08	2.16	4.28	
	BLUE	14.55—19.33	17.29±0.66	-0.23	1.29	3.80	
籽粒长宽比	E1	1.68—2.35	1.94±0.1	0.40	0.89	5.08	90.64
	E2	1.69—2.31	1.9±0.09	0.58	1.30	4.81	
	BLUE	1.68—2.33	1.92±0.09	0.50	1.38	4.73	

性状 Trait	环境 Environment	变异范围 Range	均值±标准差 Mean±SD	偏度 Ske.	峰度 Kur.	变异系数 CV(%)	遗传力 H ² (%)
籽粒形状	E1	1.00—4.00	2.36±0.54	-0.36	0.28	22.83	61.94
KS	E2	1.00—4.00	2.33±0.61	0.01	0.00	26.06	
	BLUE	1.00—3.63	2.34±0.47	-0.34	0.46	20.03	
籽粒饱满度	E1	1.00—3.00	2.15±0.52	-0.46	-0.17	24.17	56.25
GP	E2	1.00—3.00	2.26±0.57	-0.39	-0.52	25.06	
	BLUE	0.97—3.07	2.21±0.43	-0.51	-0.08	19.42	

TKG: 千粒重; GL: 粒长; GW: 粒宽; GT: 粒厚; GA: 籽粒面积; GC: 籽粒周长; LWR: 籽粒长宽比; KS: 籽粒形状; GP: 籽粒饱满度; E1: 2018-2019 年度玛纳斯环境点; E2:2019-2020 年度玛纳斯环境点; BLUE: 最佳线性无偏估计。下同

TKG: values for 1000-grain weight; GL: grain length; GW: grain width; GT: grain thickness; GA: grain area; GC: grain circumference; LWR: grain length-width ratio, KS: kernel shape; GP: grain plumpness; E1: Manas 2018-2019 environmental point; E2: Manas 2019-2020 environmental point; BLUE: Best Linear Unbiased Estimate. The same as below



*表示 $P \leq 0.05$ 水平上相关显著; **表示 $P \leq 0.01$ 水平上相关显著; ***表示 $P \leq 0.001$ 水平上相关显著。下同

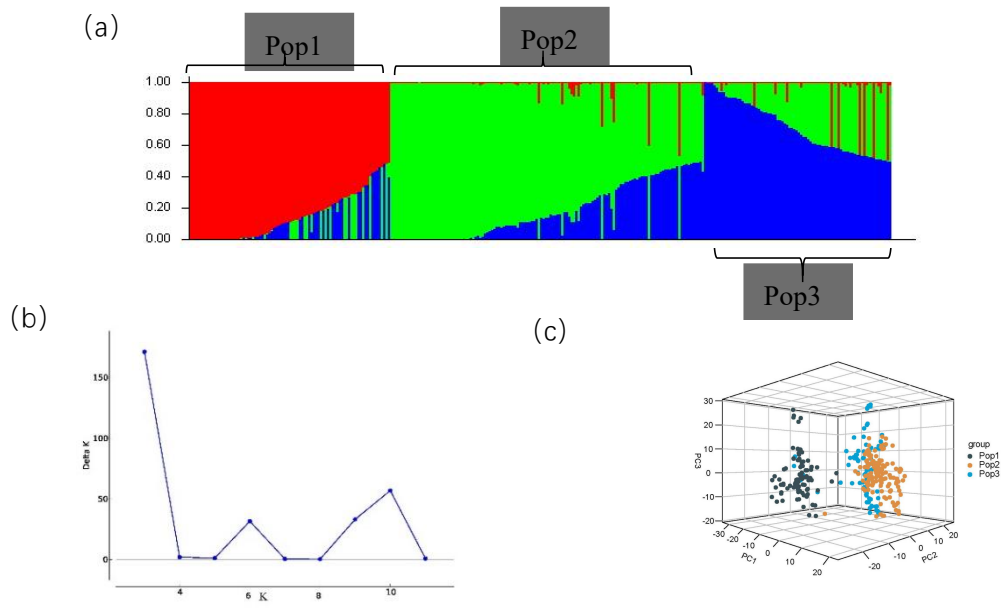
*: Significant at $P \leq 0.05$; **: Significant at $P \leq 0.01$; ***: Significant at $P \leq 0.001$. The same as below

图 1 小麦籽粒大小和形态性状相关性分析

Fig.1 Correlation analysis of grain size and morphological characters in wheat

2.2 连锁不平衡及群体结构

对 300 份小麦品种（系）的连锁不平衡分析得到，A、B 和 D 亚基因组及全基因组的 LD 衰减距离为 5、5、8 和 8 Mb。依据全基因组水平将物理图谱上小于 8 Mb 区间内的位点划分为同一位点，每个位点选择重复检测到或者显著性最高的标记作为代表性标记。采用 Structure v2.3.4 软件进行群体结构分析，由（图 2-b）可知， ΔK 在 $K=3$ 时最大，表明该群体适宜亚群数为 3。同时对供试材料进行 PCA 分析，结果显示（图 2-c）供试材料同样分为 3 个亚群。因此，本研究以 $K=3$ 生成的群体结构 Q （固定影响因子）+亲缘关系 K （随机影响因子）的 MLM 模型进行标记-性状关联分析。



(a) : 群体结构示意图； (b) : 群体结构分析 ΔK 随 K 值变化； (c) : 主成分分析。
(a) : schematic of population structure; (b) : population structure analysis ΔK as a function of K value; (c) : principal component analysis.

图 2 群体结构及主成分分析图

Fig.2 Population structure and principal component analysis diagram

2.3 标记-性状关联分析

采用混合线性模型（MLM, mixed linear model）对两个环境及 BLUE 值中的 9 个小麦籽粒大小和形态性状进行标记-性状关联分析。图 3 左侧为曼哈顿图（Manhattan plot），以 $-\log_{10}(p)=3$ 为阈值，高于阈值时，认为该 SNP 位点与该性状显著相关。右侧为 $Q-Q$ 图（Quantile-quantile plot），表示实际观测值与期望值的拟合程度。将至少在两个及以上环境中检测到的 SNP 位点视为稳定位点，通过附表 1 可知，关联分析共检测到 66 个稳定 SNP 位点。这些位点分布在除 3D、4D、5D 外的 18 条染色体上，可解释 3.74%~14.34%的表型变异。在所有环境（包含 BLUE 值环境）中均检测到的关联位点有 16 个，包含 4 个粒长、4 个粒宽、2 个粒厚、1 籽粒面积和 5 个籽粒长宽比的关联位点。本研究在 1B 染色体上检测到籽粒大小和形态的关联位点最多（12 个），包含 3 个粒长、1 个粒厚、1 个籽粒周长、6 个籽粒长宽比和 1 个籽粒饱满度位点。1B 染色体可能对籽粒大小和形态性状有较大影响，可作为研究籽粒性状的重要染色体。值得关注的是，在本研究中发现有 12 个位点疑似新位点，这些位点分别影响着千粒重、粒长、粒宽、粒厚和籽粒长宽比。

通过表 2 可知，共检测出 37 个与多个籽粒大小和形态性状显著关联的稳定位点。这些位点分布在 1A、1B、1D、2A、2B、2D、3A、3B、4A、4B、5B、6A、6B、6D、7A、7B 和 7D 染色体上，可解释 3.74%~14.33% 的表型变异。其中，发现 4 个与 4 个及以上籽粒性状显著关联的位点。在 E1 和 BLUE 值环境中的 6B 染色体的（188.19-191.99 Mb）处及 E2 和 BLUE 值环境中的 7B 染色体（23.44-26.24 Mb）处检测到控制 4 个籽粒性状的多效位点。在所有环境中检测到控制 4 个及以上性状的多效位点有：3B 染色体的 687.76 Mb 处控制粒长、粒宽、粒厚和籽粒长宽比的位点及 6D 染色体的 461.92-464.94 Mb 处控制千粒重、粒长、粒宽、籽粒面积、籽粒周长和籽粒长宽比的位点。上述发现的稳定且同时与 4 个及以上籽粒性状关联的位点可靠，可以作为分子标记辅助育种和 QTL 精细定位的重点。

2.4 单倍型分析

将所有环境中挖掘到的 BS00022512_51 和 wsnp_Ex_c4480_8055475 位点进行单倍型分析，这两个位点均具有较大的表型贡献率（7.06%~14.34%，3.81%~8.25%），且同时与 4 个及以上籽粒性状关联。如图 4-b 所示，位于 6D 染色体的 wsnp_Ex_c4480_8055475 标记处的单倍域中存在 3 个单倍型（GC-Hap1、AT-Hap2 和 AC-Hap3）。显著性分析（图 4-a）发现，具 GC-Hap1 单倍型的品种（系）在千粒重（ $43.38 \text{ g} \pm 4.19 \text{ g}$ ）、粒宽（ $3.44 \text{ mm} \pm 0.14 \text{ mm}$ ）、籽粒面积（ $17.26 \text{ mm}^2 \pm 1.19 \text{ mm}^2$ ）和籽粒周长（ $17.33 \text{ mm} \pm 0.63 \text{ mm}$ ）上显著（ $P \leq 0.05$ ）高于具 AT-Hap2 单倍型的品种（系）的千粒重（ $41.77 \text{ g} \pm 4.21 \text{ g}$ ）、粒宽（ $3.37 \text{ mm} \pm 0.14 \text{ mm}$ ）、籽粒面积（ $16.81 \text{ mm}^2 \pm 1.17 \text{ mm}^2$ ）和籽粒周长（ $17.17 \text{ mm} \pm 0.67 \text{ mm}$ ）；具 GC-Hap1 单倍型的品种（系）的籽粒长宽比（ 1.90 ± 0.09 ）显著（ $P \leq 0.05$ ）低于具 AT-Hap2 单倍型的品种（系）的籽粒长宽比（ 1.94 ± 0.10 ）；具 AC-Hap3 单倍型的品种（系）的籽粒性状在 $P \leq 0.05$ 水平上与具 GC-Hap1 和 AT-Hap2 单倍型的品种（系）的籽粒性状均无显著差异，整体均值介于 GC-Hap1 和 AT-Hap2 单倍型的品种（系）之间。具 AT-Hap2 单倍型的品种（系）虽然在籽粒长宽比性状方面显著（ $P \leq 0.05$ ）高于具 GC-Hap1 单倍型的品种（系），但两个单倍型的品种（系）间粒长不存在显著差异，粒宽存在显著差异，其效应主要来源于具 GC-Hap1 单倍型的品种（系）的粒宽，因此三个单倍型中单倍型 GC-Hap1 为籽粒较大的高千粒重单倍型。

从单倍型整体分布频率看，具 GC-Hap1 单倍型的品种（系）占 65.58%，具 AT-Hap2 单倍型的品种（系）占 32.25%，具 AC-Hap3 单倍型的品种（系）占 2.17%，具 GC-Hap1 单倍型的品种（系）分布频率明显大于具 AT-Hap2 和 AC-Hap3 单倍型的品种（系）。从单倍型在不同麦区分布频率看，具 GC-Hap1 单倍型的品种（系）分布频率表现为西南冬麦区（90.00%）>长江中下游冬麦区（89.74%）>黄淮冬麦区（69.73%）>北部冬麦区（69.57%）>国外品种（32.26%）；具 AT-Hap2 单倍型的品种（系）分布频率表现为国外品种（67.74%）>黄淮冬麦区（27.52%）>北部冬麦区（26.09%）>西南冬麦区（10.00%）>长江中下游冬麦区（7.69%）；具 AC-Hap3 单倍型的品种（系）仅在北部冬麦区、黄淮冬麦区和长江中下游冬麦区分布，且在各麦区分布频率不超过 5%（图 4-c）。麦区分布结果表明，三种单倍型在麦区分布频率上存在较大差异。

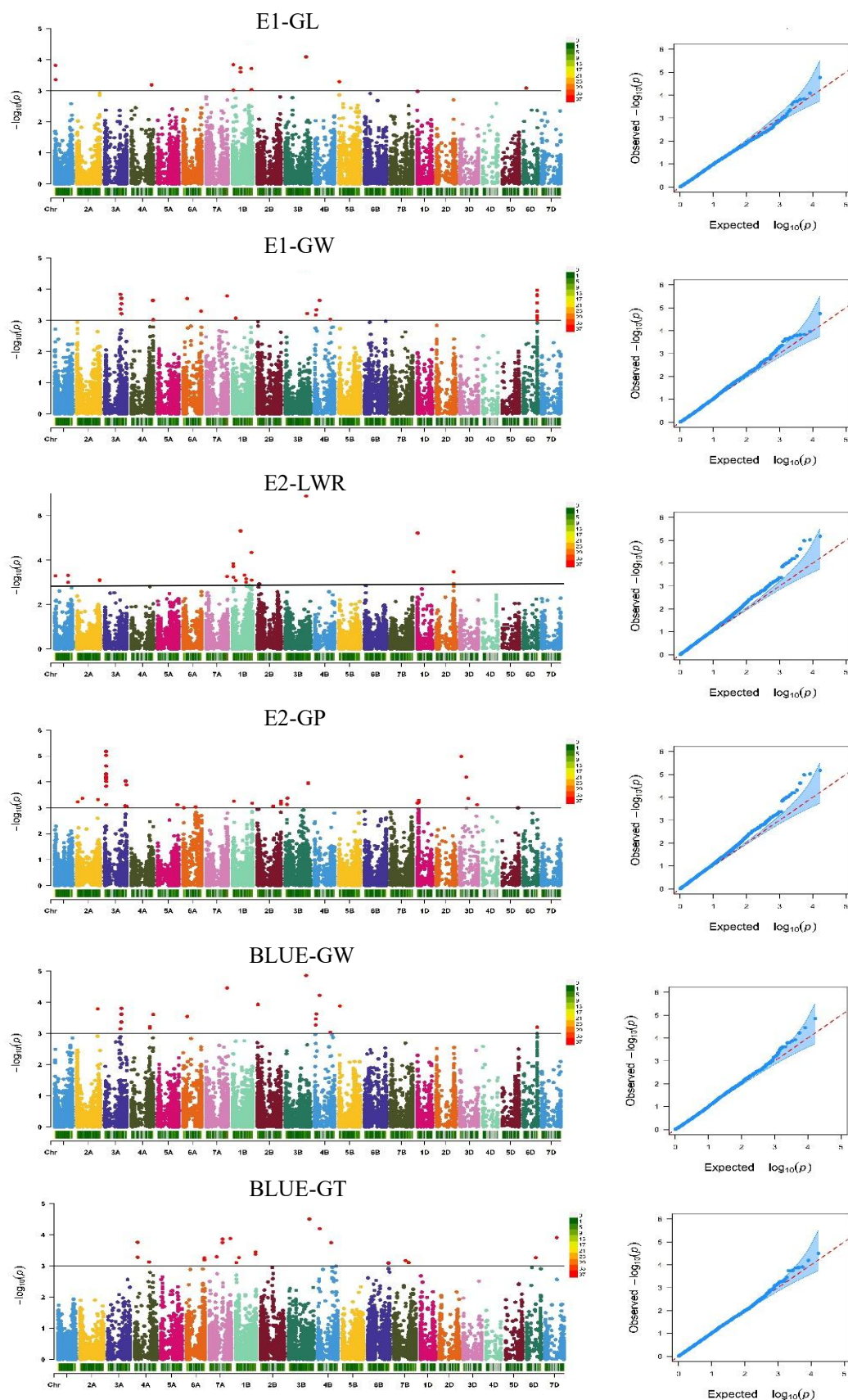


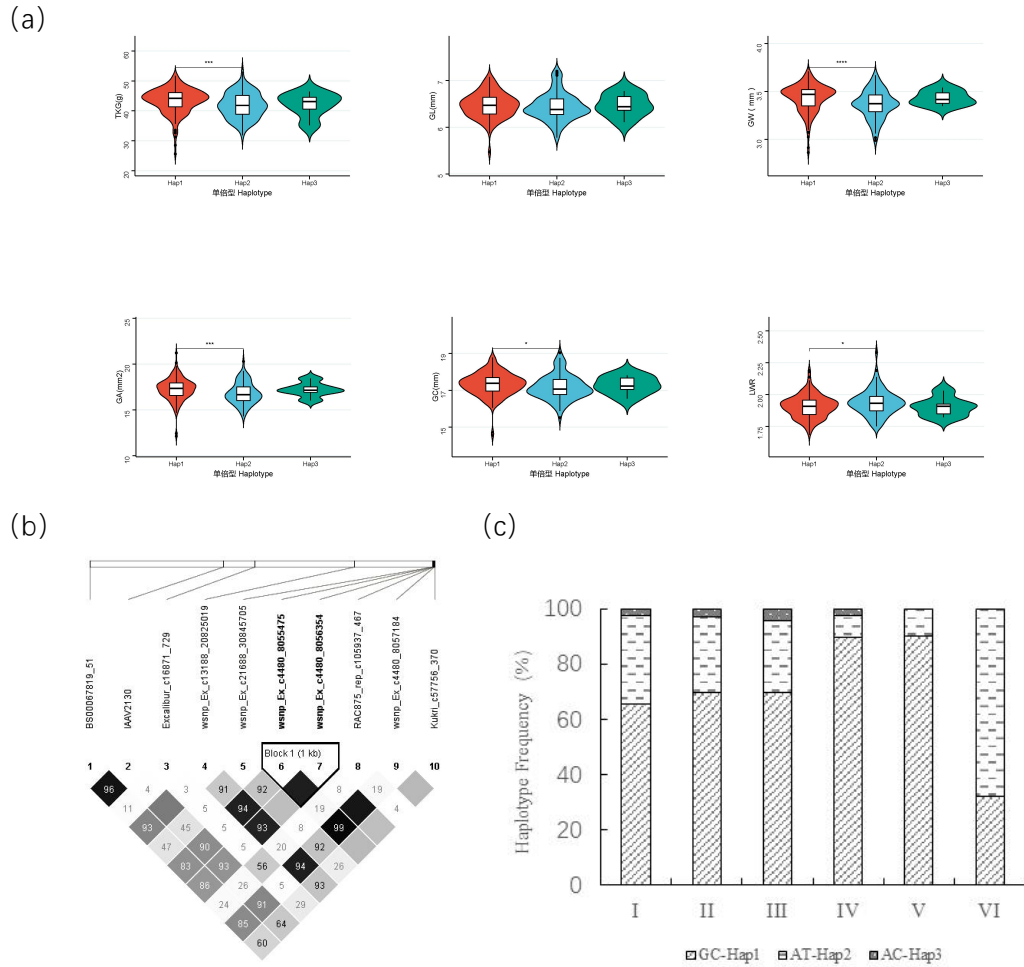
图3 部分性状曼哈顿图及Q-Q图展示

Fig.3 Manhattan and Q-Q plot display of partial traits

表 2 同时与多个籽粒性状显著关联的位点信息

Table 2 Locus information significantly associated with multiple grain traits at the same time

性状	位点	染色体	物理位置	<i>P</i> 值	表型贡献率	环境
Trait	Marker	Chr.	Position (Mb)	P value	<i>R</i> ² (%)	Environment
GL/LWR	Tdurum_contig60037_441	1A	23.97—28.62	1.54E-04—6.92E-04	4.15—5.65	E1/E2/BLUE
TKG/GW	Ex_c15768_799	1A	517.48	1.18E-04—9.59E-04	3.74—5.1	E2/BLUE
GT/KS	wsnp_Ex_c5780_10153638	1B	20.18—26.19	3.73E-04—7.88E-04	4.12—4.6	E1/E2/BLUE
GL/GT/LWR	BS00033332_51	1B	38.83—40.19	1.45E-04—9.56E-04	3.93—5.38	E1/E2/BLUE
GW/GT/LWR	GENE-0403_301	1B	109.73	2.72E-05—8.54E-04	3.89—6.24	E1/E2/BLUE
GL/LWR	Kukri_c19641_753	1B	275.67	4.65E-06—3.44E-04	4.36—7.36	E1/E2/BLUE
GL/GC/LWR	Excalibur_c10111_127	1B	637.62—642.68	1.70E-05—9.30E-04	3.8—6.5	E1/E2/BLUE
GT/GP	wsnp_Ex_c38849_46284348	1B	661.63—668.25	3.59E-04—7.70E-04	3.95—4.64	E2/BLUE
LWR/GP	IAAV426	1D	16.79	4.77E-06—6.51E-04	7.74—13.07	E1/E2/BLUE
GL/LWR	RAC875_c259_1339	2A	759.83—760.56	4.69E-04—9.67E-04	4.12—5.79	E2/BLUE
TKG/GW/GA	BS00099658_51	2B	24.91	3.59E-06—7.97E-04	5.15—9.11	E1/E2/BLUE
LWR/KS	BS00081406_51	2B	140.85—142.45	1.11E-04—9.23E-04	3.83—5.21	E1/BLUE
LWR/KS	Excalibur_c42512_584	2D	577.11—580.19	3.34E-04—8.25E-04	3.94—4.55	E2/BLUE
TKG/GP	Kukri_c14029_117	3A	53.34—55.09	6.68E-06—8.93E-04	3.76—7.19	E1/E2
GW/KS	IAAV8990	3A	375.82	7.11E-04—7.87E-04	4.03—4.43	E2/BLUE
TKG/GP	BS00097939_51	3A	729.58—732.04	1.29E-04—8.84E-04	4.65—5.84	E1/E2
GL/GC	Excalibur_c18641_1849	3B	230.34—235.87	1.12E-04—8.32E-04	3.8—5.19	E2/BLUE
GL/GW/GT/LWR	BS00022512_51	3B	687.76	4.05E-09—8.17E-05	7.06—14.34	E1/E2/BLUE
TKG/GW	Tdurum_contig31852_251	4A	734—734.07	1.24E-04—9.62E-04	5.07—6.76	E1/E2/BLUE
TKG/GW	IAAV971	4B	37.69—40.75	3.45E-04—9.04E-04	3.82—4.44	E1/E2/BLUE
TKG/GW/GT	Excalibur_c56787_95	4B	59.21	2.40E-04—8.59E-04	4.11—5.88	E1/E2/BLUE
GW/GT	Excalibur_c52517_464	4B	167.35	6.01E-05—6.48E-04	4.01—5.59	E1/E2/BLUE
LWR/GT	wsnp_Ex_c25373_34639805	4B	480.59—482.65	3.51E-04—6.10E-04	4.12—5.53	E1/E2/BLUE
TKG/GW	GENE-2771_327	4B	526.73—526.93	6.66E-04—9.37E-04	3.74—4.26	E1/E2/BLUE
GL/GA/GC	wsnp_Ex_c831_1625061	5B	10.53	2.51E-04—8.32E-04	4.97—5.83	E1/BLUE
TKG/GW/GA	BobWhite_c14575_323	6A	129.74	1.65E-04—6.06E-04	4.11—4.97	E1/BLUE
GW/GT/KS	RFL_Contig6053_3082	6A	594.99—597.79	4.47E-04—8.11E-04	3.94—4.31	E1/BLUE
GA/GC/LWR/KS	Jagger_c555_287	6B	188.19—191.99	3.75E-04—9.48E-04	3.81—5.08	E1/BLUE
TKG/KS	Kukri_c50603_164	6B	451.93—456.76	3.41E-04—7.98E-04	4.05—4.75	E1/E2
GL/GA/GC	Kukri_c65194_111	6D	98.13	1.56E-04—9.22E-04	3.89—5.04	E1/BLUE
TKG/GL/GW/GA/GC/LWR	wsnp_Ex_c4480_8055475	6D	461.92—464.94	1.07E-04—9.95E-04	3.81—8.25	E1/E2/BLUE
TKG/GT	BobWhite_c5396_296	7A	442.27—443.96	1.38E-04—9.95E-04	3.77—5.18	E1/E2/BLUE
TKG/GC	Kukri_c24408_743	7A	670.82	3.21E-04—9.55E-04	3.85—4.67	E1/BLUE
GW/GT/LWR	Tdurum_contig29240_206	7A	702.4	3.50E-05—6.28E-04	4.02—5.97	E1/E2/BLUE
GL/GA/GC/GP	CAP7_c1383_548	7B	23.44—26.24	1.16E-04—9.62E-04	3.74—5.17	E2/BLUE
TKG/GA/GP	wsnp_Ku_c665_1371121	7B	58.25—58.73	5.23E-04—8.78E-04	3.95—5.18	E2/BLUE
LWR/KS	Excalibur_c32138_394	7D	633.2—633.6	1.45E-04—7.64E-04	4.1—6.29	E1/BLUE



a: 千粒重 (TKG)、粒长(GL)、粒宽(GW)、籽粒面积(GA)、籽粒周长(GC)、籽粒长宽比(LWR)的小提琴图。b: 6D 染色体的 wsnp_Ex_c4480_8055475 标记关联的连锁区域。c: wsnp_Ex_c4480_8055475 标记处 3 种单倍型在国内外 300 份小麦品种 (系) 的分布频率及 3 种单倍型在中国 4 个冬麦区和国外品种的分布频率。I: wsnp_Ex_c4480_8055475 标记处 3 种单倍型整体的分布频率; II: 黄淮冬麦区; III: 北部冬麦区; IV: 长江中下游冬麦区; V: 西南冬麦区; VI: 国外品种。

a: violin diagram of 1000-grain weight, grain length, grain width, grain area, grain perimeter and grain length-width ratio. b: Linked regions associated with wsnp_Ex_c4480_8055475 markers on chromosome 6D. c: The distribution frequencies of three haplotypes in 300 wheat varieties (lines) at wsnp_Ex_c4480_8055475 markers, and the distribution frequencies of three haplotypes in four winter wheat regions in China and foreign varieties. I: The overall distribution frequency of three haplotypes at wsnp_Ex_c4480_8055475 markers; II: Huang-Huai winter wheat region; III: northern winter wheat region; IV: middle and lower Yangtze River winter wheat region; V: southwest winter wheat region; VI: foreign varieties.

图 4 单倍型分析

Fig. 4 Haplotype analysis

2.5 候选基因筛选

结合普通小麦中国春基因组数据库 (IWGSC, <https://www.wheatgenome.org/>), 本研究对与多个籽粒大小和形态性状显著关联 SNP 标记进行候选基因筛选, 发现了 9 个可能的候选基因 (表 3)。在 1B 染色体上存在两个候选基因 *TraesCS1B03G0270700* 和 *TraesCS1B03G1104100*, 分别编码 MYB 转录因子 (MYB transcription factor) 和激酶 (Kinase)。前者参与植物次生代谢途径, 后者主要在细胞的信号传导及其复杂的

生命活动中起作用。2B 染色体的 *TraesCS2B03G0104000* 编码的细胞色素 P450 (Cytochrome P450) 在植物体内起生物合成的作用, 具有解毒抗逆等功能。位于 1D 染色体的 *TraesCS1D03G0071600* 和 3A 染色体的 *TraesCS3A03G0182600* 编码富含半胱氨酸的受体样激酶 (Cysteine-rich receptor-like kinase), 主要在植物生长发育和应激反应中起关键作用。3B 染色体的 *TraesCS3B03G1107600* 编码五肽重复序列的蛋白质 (Pentatricopeptide repeat-containing protein), 主要参与植物的逆境胁迫和生长发育。位于 6B 染色体的 *TraesCS6B03G0448200* 编码泛素 (Ubiquitin), 泛素介导着靶蛋白的降解, 泛素途径影响着植物种子大小测定^[22]。位于 6D 染色体的 *TraesCS6D03G0893200* 编码 RING/U-box 超家族蛋白 (RING/U-box superfamily protein), 参与植物中多种激素信号的转导; 位于 7A 染色体的 *TraesCS7A03G1260500* 编码 F-box 家族蛋白 (F-box family protein), 该家族蛋白影响着植物的诸多方面, 如: 对生长发育、信号转导和生物及非生物逆境胁迫等方面起到重要作用。

表 3 筛选获得候选基因信息
Table 3 Screening to obtain candidate gene information

性状 Trait	SNP 标记 SNP marker	染色体 Chr.	物理位置 Physical position (Mb)	候选基因 Gene ID	功能注释 Functional annotation
GW/GT/LWR	GENE-0403_301	1B	115.97	<i>TraesCS1B03G0270700</i>	MYB 转录因子
GL/GC/LWR	Excalibur_c10111_127	1B	644.63	<i>TraesCS1B03G1104100</i>	激酶, 推定
LWR/GP	IAAV426	1D	18.15	<i>TraesCS1D03G0071600</i>	富含半胱氨酸的受体激酶样蛋白
TKG/GW/GA	BS00099658_51	2B	27.54	<i>TraesCS2B03G0104000</i>	细胞色素 P450
TKG/GP	Kukri_c14029_117	3A	53.51	<i>TraesCS3A03G0182600</i>	富含半胱氨酸的受体样激酶
GL/GW/GT/LWR	BS00022512_51	3B	703.84	<i>TraesCS3B03G1107600</i>	含有五肽重复序列的蛋白质
GA/GC/LWR/KS	Jagger_c555_287	6B	198.25	<i>TraesCS6B03G0448200</i>	泛素, 推定
TKG/GL/GW/GA/GC/LWR	wsnp_Ex_c4480_8055475	6D	484.02	<i>TraesCS6D03G0893200</i>	RING/U-box 超家族蛋白, 推定
GW/GT/LWR	Tdurum_contig29240_206	7A	707.89	<i>TraesCS7A03G1260500</i>	F-box 家族蛋白

3 讨论

3.1 籽粒性状的表型观察

小麦籽粒大小和形态是影响籽粒产量、磨粉品质和市场价值的重要因素, 其遗传解析对于我国粮食安全具有重要意义^[23]。本研究中, 小麦籽粒大小和形态相关性状整体呈正态分布, 遗传力在 56.25%~91.89% 之间, 所有目标性状均具有较大的表型变异, 适合用于全基因组关联分析。千粒重作为产量的主要构成因素之一, 受到籽粒长度、宽度、厚度和籽粒面积等因素的共同影响^[5]。本研究表明除籽粒长宽比外, 其余性状均与千粒重呈极显著 ($P \leq 0.001$) 正相关关系, 籽粒面积与千粒重的相关性最高 ($r=0.90$)。籽粒面积受到籽粒长宽的共同影响, 这二者虽然都与千粒重呈极显著正相关。粒宽与千粒重的相关性明显大于粒长与千粒重的相关性。同时粒宽与千粒重的相关性也明显大于粒厚与千粒重的相关性, 这表明目前小麦产量的增加受到粒宽的影响更大, 这些结果在前人的研究中也有证明^[24]。同时, 本研究中的粒形整体呈卵型偏椭圆型, 其原因可能与育种家对小麦籽粒宽度偏好选择存在极大的关系。Gegas 等^[25]分析了大多数已知小麦属种的籽粒大小和形态性状, 发现小麦的籽粒从一种细长的原始籽粒进化成一种更宽、更短的椭圆型现

代籽粒。Marshall 等^[26]对小麦籽粒用几何模型分析,发现不同大小和形态的籽粒出粉率存在一定差异,长型小麦籽粒出粉率最低,椭圆型小麦籽粒出粉率最高。这些结果表明了粒宽对产量和磨粉品质的影响相较于其他籽粒大小性状更大,籽粒的宽度相较于长度和厚度是更有效的育种性状。

3.2 当前 GWAS 结果与前人研究比较

GWAS 作为检测基因特征和表型变异的有力工具,在植物遗传育种中发挥越来越多的作用^[27]。小麦籽粒大小和形态相关性状属于微效多基因控制的数量性状,其遗传解析往往受到基因型和环境互作的影响,因此在不同环境中检测到稳定的标记-性状关联位点往往具有更高的参考价值^[28]。本研究以 $-\log_{10}(p)=3$ 为阈值,在两个及以上环境中共检测到 66 个稳定的标记-性状关联位点,可解释 3.74%~14.34%的表型变异。其中,1B 染色体定位到的位点最多(12 个),这 12 位点分别影响着粒长、粒厚、籽粒周长、籽粒长宽比和籽粒饱满度。严勇亮等^[29]的研究也发现 1B 染色体对小麦籽粒性状调控具有重要影响,在今后对小麦籽粒性状的研究中可优先选择 1B 染色体的位点。将本研究结果和前人研究进行比较,发现一些相邻或相近的位点和基因。本研究于 4A 染色体定位到与粒宽显著关联的标记 Tdurum_contig31852_251 与 Yue 等^[30]鉴定的粒重基因 6-*SFT-A2* 与相距约 4.9 Mb,推测该位点控制籽粒长度的效应可能来源于该基因。Zhao 等^[18]于 1D 染色体定位到控制籽粒长宽比位点 AX-108835501 与本研究定位到的籽粒长宽比 IAAV426 位点相距仅 0.14 Mb,在本研究中能够解释较高表型变异(10.84%~13.07%),该位点可能存在调控小麦籽粒长宽比的基因。同时,本研究于 2B 染色体的 24.91 Mb 处检测到控制籽粒面积的 BS00099658_51 位点。Daba 等^[31]在该染色体 16.90 Mb 处检测到控制粒宽的标记 S2B_16897947,邹林翰等^[11]在该染色体 1.60-10.17 Mb 处检测到控制粒厚的 QTL, Malik 等^[32]在该位点下游 16.59 Mb 处检测到了粒长位点,这些位点与本研究位点均不足 20 Mb,推测该染色体区段可能是控制小麦籽粒大小和形态性状的重要候选区段。本研究定位于 4B 染色体短臂的 Excalibur_c56787_95 标记同时控制千粒重、粒宽和粒厚,前人研究也证明了 4B 染色体上的 QTL 对千粒重和籽粒大小有多效性影响^[4,33]。同时位于 6A 染色体 129.74 Mb 处的控制千粒重、粒宽和籽粒面积的 BobWhite_c14575_323 标记与 Daba 等^[31]定位的粒长和邹林翰等^[11]定位到控制粒厚的标记位置相距 3 Mb 范围内,该位点处还未发现籽粒性状相关的基因,推测这些性状可能为同一基因调控。定位于 6D 染色体与 6 个籽粒性状关联的 wsnp_Ex_c4480_8055475 标记与 Li 等^[34]定位到的千粒重标记 AX_109481324 和 Zhao 等^[18]定位的籽粒长宽比标记 AX-95156239 相距不足 2 Mb,该位点可能是控制小麦籽粒性状的主效 QTL。综上,本研究发现的 SNP 位点与已知小麦籽粒相关性状的 QTL 位点或基因吻合,表明本研究的全基因组关联分析具有较高的可信度。同时发现与 6 个籽粒性状关联的 wsnp_Ex_c4480_8055475 标记和 1D 染色体定位的表型贡献率大的 IAAV426 标记,可作为 QTL 精细定位和标记开发的首选。

虽然前人利用不同群体已定位到较多的小麦籽粒性状 QTL,但小麦基因组庞大而复杂,因此克隆的基因数目有限^[35]。本研究以国内外 300 份小麦品种(系)为研究材料,发现了 12 个籽粒大小和形态相关性状

的新位点。定位于 1B (Kukri_c19641_753)、3B (BS00022512_51), 4B (wsnp_Ex_c25373_34639805) 和 7A (Tdurum_contig29240_206) 的标记是影响籽粒性状的一因多效标记。其中, 3B 染色体上的标记 BS00022512_51 在本研究中解释最大 7.06%~14.34% 的表型变异, 这些定位到的一因多效位点可能涉及小麦籽粒大小和形态发育过程中组成性表达基因。下一步研究中, 可优先选择以 3B 染色体为主的多效性标记进行图位克隆和功能验证, 以期进一步推进小麦籽粒大小和形态遗传机制的研究。

3.3 籽粒性状有利单倍型鉴定及分布

近年来, 全基因组关联分析已经成为分析复杂性状的首选方法^[14]。然而, 用双等位基因 SNP 位点描述性状和位点之间的复杂关系的分辨率是有限的, 基于连锁不平衡构建单倍型进行相关分析是克服这一局限的有效途径^[36]。单倍型将两个或多个 SNP 等位基因作为一个遗传单位, 通过关联分析结合单倍型分析的方法, 这极大提高了关联分析的可靠性。本研究对表型贡献力大且控制 4 个及以上籽粒性状的 2 个关联位点进行单倍型分析, 发现位于 6D 染色体的 wsnp_Ex_c4480_8055475 标记存在 3 个单倍型。具单倍型 GC-Hap1 和 AT-Hap2 的品种(系)在 5 个籽粒性状上差异均达到了显著水平($P \leq 0.05$), 进一步分析发现单倍型 GC-Hap1 为籽粒较大的高千粒重单倍型。在作物育种中, 优异等位变异或单倍型总会受到育种家的正向选择, 使其在育成品种中出现的频率提高^[37]。本研究同样也发现与籽粒性状好、高千粒重紧密相关的 GC-Hap1 单倍型 (65.58%) 中的品种(系)的频率远高于 AT-Hap2 单倍型 (32.25%) 和 AC-Hap3 单倍型 (2.17%) 中的品种(系)的频率。这表明单倍型 GC-Hap1 受到了育种家的正向选择, 但同时也表明单倍型 GC-Hap1 对小麦籽粒性状及粒重的遗传改良具有较大的提升空间。进一步对不同麦区分布频率分析 (图 4-c), 籽粒性状上与高千粒重紧密相关的具 GC-Hap1 单倍型的品种(系)在不同麦区分布频率存在较大差异。具 GC-Hap1 单倍型的品种(系)在中国育成小麦品种(系)中被大量选育, 在本研究中的 4 个中国冬麦区分布频率均达到了 69% 以上, 而在国外品种中的分布频率仅有 32.26%。中国育成小麦品种(系)整体分布频率大于国外引进品种(系)可能由于生态条件和选择压力的原因, 对小麦的千粒重等籽粒性状产生了积极的影响, 而国外小麦育种对于千粒重的选择不强, 对于产量提升的需求也不大, 其产量的提升更依赖于穗数和穗粒数^[38]。国内四个冬麦区中, 长江中下游冬麦区和西南冬麦区在具 GC-Hap1 单倍型的品种(系)中分布频率最高, 达到了 85% 以上, 原因可能与两地的气候条件密切相关, 病害频发和生育后期高温等不利因素是影响两地产量提升的主要原因, 所以育种家在品种选育上更加重视籽粒灌浆程度和籽粒饱满度的选择, 在一定程度上小麦籽粒灌浆充实的籽粒一般抗逆抗病性好, 千粒重高^[39]。本研究进行单倍型分析的标记可依据该标记的位置及候选基因筛选的结果, 进一步通过连锁分析验证标记的实用性, 并深入解析候选基因的作用机制。同时, 在今后的小麦育种工作中, 可以根据不同麦区的实际情况选择适宜的单倍型的品种进行遗传改良。

3.4 候选基因分析

在作物育种中，已经验证了一些与籽粒大小和形态相关的候选基因。例如，细胞色素 P450 通过影响小麦种皮数量的多少影响籽粒大小，其沉默可导致小麦籽粒大小下降 11%^[40]；RING/U-box 超家族蛋白在泛素通道起作用，而泛素途径在水稻和拟南芥中影响着籽粒大小已经得到了证明^[22,41]。本研究对与多性状显著关联的 SNP 位点进行候选基因分析（表 3），发现了 9 个可能的候选基因，本研究中发现的候选基因 *TraesCS6D03G0893200* 和 *TraesCS2B03G0104000* 在前人研究中已经证实。其次，位于 1B 染色体检测到两个候选基因，*TraesCS1B03G0270700* 编码着 MYB 转录因子，MYB 家族参与植物逆境胁迫响应和植物生长发育的许多重要阶段，朱治等^[37]对小麦 *TaMYB5-3B* 基因研究发现，该基因与株高、穗下节长和千粒重显著相关。*TraesCS1B03G1104100* 编码着激酶，在激酶中最大族群是蛋白质激酶，蛋白质激酶在细胞中的信号转导和生命活动中起到了重要作用，关于激酶在小麦籽粒性状方面的作用还有待进一步研究。位于 3B 染色体上拥有表型贡献力最大标记处发现的基因 *TraesCS3B03G1107600*，该基因编码含有五肽重复序列的蛋白质（PPR），该家族蛋白质在植物细胞中主要对植物生长发育和响应逆境胁迫等方面起到重要作用。LI 等^[42]在对花生种子大小和重量的研究中发现，在鉴定的候选基因中，有 1/2 都编码五肽重复序列。本研究筛选的候选基因编码着不同蛋白，在调控小麦籽粒大小和形态方面起到了直接或间接作用，需要进一步深入解析，为小麦籽粒大小和形态改良奠定基础。

4 结论

本研究利用小麦 90 K SNP 芯片，对国内外 300 份小麦品种（系）的籽粒大小和形态相关性状进行全基因组关联分析。在两个及以上环境中共检测到 66 个稳定显著的 SNP 标记，与前人研究对比后发现了 12 个新位点，同时发现 37 个一因多效位点。对所有环境中表型贡献率大且同时与 4 个及以上性状的位点进行单倍型分析，于 6D 染色体的 *wsnp_Ex_c4480_8055475* 标记处发现了籽粒较大的高千粒重单倍型 GC-Hap1；对发现的与多性状显著关联的 SNP 进行候选基因筛选后，鉴定出 9 个候选基因。

参考文献:

- [1] Pask A J D, Reynolds M P. Breeding for yield potential has increased deep soil water extraction capacity in irrigated wheat. *Crop Science*, 2013, 53(5): 2090-2104
- [2] Shiferaw B, Smale M, Braun H J, Duveiller E, Reynolds M, Muricho G. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, 2013, 5(3): 291-317
- [3] Tekeu H, Ngonkeu E L M, Bélanger S, Djocgoué P F, Abed A, Torkamaneh D, Boyle B, Tsimi P M, Tadesse W, Jean M, Belzile F. GWAS identifies an ortholog of the rice D11 gene as a candidate gene for grain size in an international collection of hexaploid wheat. *Scientific Reports*, 2021, 11(1): 19483
- [4] Gegas V C, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan J H, Snape J W. A genetic framework for grain size and shape variation in wheat. *The Plant Cell*, 2010, 22(4): 1046-1056
- [5] 张芳, 任毅, 曹俊梅, 李法计, 夏先春, 耿洪伟. 基于SNP标记的小麦籽粒性状全基因组关联分析. *中国农业科学*, 2021, 54(10): 2053-2064
Zhang F, Ren Y, Cao J M, Li F J, Xia X C, Geng H W. Genome-wide association analysis of wheat grain size related traits based on SNP markers. *Scientia Agricultura Sinica*, 2021, 54(10): 2053-2064
- [6] Wu W, Li C J, Ma B L, Shah F, Liu Y, Liao Y C. Genetic progress in wheat yield and associated traits in China since 1945 and future prospects. *Euphytica*, 2014, 196(2): 155-168
- [7] Cornish P S, So H B, McWilliam J R. Effects of soil bulk density and water regimen on root growth and uptake of phosphorus by ryegrass. *Australian Journal of Agricultural Research*, 1984, 35(5): 631-644
- [8] Rasheed A, Wen W E, Gao F M, Zhai S N, Jin H, Liu J D, Guo Q, Zhang Y J, Dreisigacker S, Xia X C, He Z H. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theoretical and Applied Genetics*, 2016, 129(10): 1843-1860
- [9] Edae E A, Byrne P F, Haley S D, Lopes M S, Reynolds M P. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theoretical and Applied Genetics*, 2014, 127(4): 791-807
- [10] 周小鸿, 马建, 罗伟, 蒋云峰, 孙敏, 杨宇杰, 江千涛, 刘亚西, 陈国跃, 魏育明, 郑有良, 兰秀锦. 西藏半野生小麦粒型性状的QTL定位. *麦类作物学报*, 2016, 36(01): 27-35
Zhou X H, Ma J, Luo W, Jiang Y F, Sun M, Yang Y J, Jiang Q T, Liu Y X, Chen G Y, Wei Y M, Zheng Y L, Lan X J. QTL mapping for kernel morphology traits of tibetan Semi-wild wheat. *Journal of Triticeae Crops*, 2016, 36(01): 27-35
- [11] 邹林翰, 周新颖, 张泽源, 蔚睿, 袁梦, 宋晓朋, 简俊涛, 张传量, 韩德俊, 宋全昊. 小麦周8425B×小偃81重组自交系群体千粒重相关性状的QTL定位及单倍型分析. *中国农业科学*, 2022, 55(18): 3473-3483
Zou L H, Zhou X Y, Zhang Z Y, Yu R, Yuan M, Song X P, Jian J T, Zhang C L, Han D J, Song Q H. QTL Mapping of Thousand-Grain-Weight and Its Related Traits in Zhou 8425B × Xiaoyan 81 Population and Haplotype Analysis. *Scientia Agricultura Sinica*, 2022, 55(18): 3473-3483
- [12] Cortes L T, Zhang Z W, Yu J M. Status and prospects of genome-wide association studies in plants. *Plant Genome*, 2021, 14(1): e20077
- [13] Gupta P K, Kulwal P L, Jaiswal V. Chapter Two - Association mapping in plants in the post-GWAS genomics era. *Advances in Genetics: Academic Press*, 2019: 75-154
- [14] Muhammad A, Hu W C, Li Z Y, Li J G, Xie G S, Wang J B, Wang L Q. Appraising the genetic architecture of kernel traits in hexaploid wheat using GWAS. *International Journal of Molecular Sciences*, 2020, 21(16): 5649
- [15] Huang X H, Zhao Y, Wei X H, Li C Y, Wang A H, Zhao Q, Li W J, Guo Y L, Deng L W, Zhu C R, Fan D L, Lu Y Q, Weng Q J, Liu K Y, Zhou T Y, Jing Y F, Si L Z, Dong G J, Huang T, Lu T T, Feng Q, Qian Q, Li J Y, Han B. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*, 2012, 44(1): 32-39
- [16] 张爱民, 阳文龙, 方红曼, 吕慧颖, 邓向东, 葛毅强, 魏珣, 杨维才. 作物种质资源研究态势分析. *植物遗传资源学报*, 2018, 19(03): 377-382.
Zhang A M, YANG W L, FANG H M, LV H Y, DENG X D, GE Y Q, WEI X, YANG W C. Current status of research on crop genetic germplasms. *Journal of Plant Genetic Resources*, 2018, 19(03): 377-382.
- [17] Wang P, Tian T, Ma J F, Liu Y, Zhang P P, Chen T, Shahinnia F, Yang D L. Genome-wide association study of kernel traits using a 35K SNP array in bread wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 2022, 13: 905660
- [18] Zhao X P, Lv L L, Li J H, Ma F F, Bai S L, Zhou Y, Zhang D L, Li S P, Song C P. Genome-wide association study of grain shapes in *Aegilops tauschii*. *Euphytica*, 2021, 217(7): 144
- [19] 李立会, 李秀全, 杨欣明. 小麦种质资源描述规范和数据标准. 北京: 中国农业出版社, 2006: 33-81

- Li L H, Li X Q, Yang X M. Descriptors and data standard for wheat (*Triticum aestivum* L.). Beijing: China Agricultural Press, 2006: 33-81
- [20] Wang S X, Zhu Y L, Zhang D X, Shao H, Liu P, Hu J B, Zhang H, Zhang H P, Chang C, Lu J, Xia X C, Sun G L, Ma C X. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *Plos One*, 2017, 12(11): e188662
- [21] Bradbury P J, Zhang Z W, Kroon D E, Casstevens T M, Ramdoss Y, Buckler E S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 2007, 23(19): 2633-2635
- [22] Li N, Li Y H. Ubiquitin-mediated control of seed size in plants. *Frontiers in Plant Science*, 2014, 5: 332
- [23] Abdipour M, Ebrahimi M, Izadi-Darbandi A, Mastrangelo A M, Najafian G, Arshad Y, Mirniyam G. Association between grain size and shape and quality traits, and path analysis of thousand grain weight in iranian bread wheat landraces from different geographic regions. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 2016, 44(1): 228-236
- [24] Simmonds J, Scott P, Brinton J, Mestre T C, Bush M, Del Blanco A, Dubcovsky J, Uauy C. A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theoretical and Applied Genetics*, 2016, 129(6): 1099-1112
- [25] Gegas V C, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan J H, Snape J W. A genetic framework for grain size and shape variation in wheat. *Plant Cell*, 2010, 22(4): 1046-1056
- [26] Marshall D R, Mares D J, Moss H J, Ellison F W. Effects of grain shape and size on milling yields in wheat. II. Experimental studies. *Crop and Pasture Science*, 1986, 37(4): 331-342
- [27] 李嘉琦, 逢洪波, 解元坤, 李玥莹, 陈强. 基于农艺植物粒型相关性状探讨GWAS分析方法研究进展. *福建农林大学学报: 自然科学版*, 2020, 49(02): 153-158
- Li J Q, Pang H P, Jie Y K, Li Y Y, Chen Q. Advances on GWAS analysis methods based on studies on grain size related traits for agronomic plants. *Journal of Fujian Agriculture and Forestry University : Natural Science Edition*, 2020, 49(02): 153-158
- [28] Gao Y T, Xu X R, Jin J J, Duan S N, Zhen W C, Xie C J, Ma J. Dissecting the genetic basis of grain morphology traits in Chinese wheat by genome wide association study. *Euphytica*, 2021, 217(4): 56
- [29] 严勇亮, 张恒, 张金波, 时晓磊, 耿洪伟, 肖菁, 路子峰, 倪中福, 丛花. 春小麦主要籽粒性状的全基因组关联分析. *麦类作物学报*, 2022, 42(10): 1182-1191
- Yan Y L, Zhang H, Zhang J B, Shi X L, Geng H W, Xiao J, Lu Z F, Ni Z F, Cong H. Genome-wide association study of grain traits of spring wheat. *Journal of Triticeae Crops*, 2022, 42(10): 1182-1191
- [30] Yue A Q, Li A, Mao X G, Chang X P, Li R Z, Jing R L. Identification and development of a functional marker from 6-SFT-A2 associated with grain weight in wheat. *Molecular Breeding*, 2015, 35(2): 63
- [31] Daba S D, Tyagi P, Brown-Guedira G, Mohammadi M. Genome-wide association studies to identify loci and candidate genes controlling kernel weight and length in a historical united states wheat population. *Frontiers in Plant Science*, 2018, 9: 1045
- [32] Malik P, Kumar J, Sharma S, Meher P K, Balyan H S, Gupta P K, Sharma S. GWAS for main effects and epistatic interactions for grain morphology traits in wheat. *Physiology and Molecular Biology of Plants*, 2022, 28(3): 651-668
- [33] Sun X Y, Wu K, Zhao Y, Kong F M, Han G H, Jiang H M, Huang X J, Li R J, Wang H G, Li S S. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica*, 2009, 165(3): 615-624.
- [34] Li Y Y, Tang J Q, Liu W L, Yan W Y, Sun Y, Che J Y, Tian C, Zhang H J, Yu L H. The genetic architecture of grain yield in spring wheat based on genome-wide association study. *Frontiers in Genetics*, 2021, 12: 728472
- [35] Yang L, Zhao D H, Meng Z L, Xu K J, Yan J, Xia X C, Cao S H, Tian Y B, He Z H, Zhang Y. QTL mapping for grain yield-related traits in bread wheat via SNP-based selective genotyping. *Theoretical and Applied Genetics*, 2020, 133(3): 857-872
- [36] Qian L W, Hickey L T, Stahl A, Werner C R, Hayes B, Snowdon R J, Voss-Fels K P. Exploring and harnessing haplotype diversity to improve yield stability in crops. *Frontiers in Plant Science*, 2017, 8: 1534
- [37] 朱治, 李龙, 李超男, 毛新国, 郝晨阳, 朱婷, 王景一, 常建忠, 景蕊莲. 小麦转录因子TaMYB5-3B与株高和千粒重相关. *作物学报*, 2022, <https://kns.cnki.net/kcms/detail/11.1809.S.20220826.1455.002.html>: 1-12
- Zhu Z, Li L, Li C N, Mao X G, Hao C Y, Zhu T, Wang J Y, Chang J Z, Jing R L. Transcription factor TaMYB5-3B is associated with plant height and 1000-grain weight in wheat. *Acta Agronomica Sinica*, 2022, <https://kns.cnki.net/kcms/detail/11.1809.S.20220826.1455.002.html>: 1-12
- [38] 何中虎, 夏先春, 陈新民, 庄巧生. 中国小麦育种进展与展望. *作物学报*, 2011, 37(02): 202-215

He Z H, Xia X C, Chen X M, Zhuang Q S. Progress of wheat breeding in China and the future perspective. *Acta Agronomica Sinica*, 2011, 37(02): 202-215

[39] 周阳. 中国冬小麦产量潜力及重要农艺性状的遗传改良. 西北农林科技大学, 2005

Zhou Y. Genetic Improvement in Yield and Its Associated Traits of Winter Wheat in China. Northwest A&F University, 2005

[40] Ma M, Wang Q, Li Z J, Cheng H H, Li Z J, Liu X L, Song W N, Appels R, Zhao H X. Expression of TaCYP78A3, a gene encoding cytochrome P450 CYP78A3 protein in wheat (*Triticum aestivum* L.), affects seed size. *The Plant Journal*, 2015, 83(2): 312-325

[41] Yee D, Goring D R. The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *Journal of Experimental Botany*, 2009, 60(4): 1109-1121

[42] Li Z F, Zhang X G, Zhao K K, Zhao K, Qu C X, Gao G Q, Gong F P, Ma X L, Yin D M. Comprehensive transcriptome analyses reveal candidate genes for variation in seed size/weight during peanut (*arachis hypogaea* L.) domestication. *Frontiers in Plant Science*, 2021, 12: 666483