

引用:朱沂昕,李重阳,王俊,等.大豆耐储藏性相关位点的GWAS定位及候选基因的单倍型分析[J].聊城大学学报(自然科学版),2025,38(6):997-1005.

Citation:ZHU Yixin, LI Chongyang, WANG Jun, LI Minmin, GENG Yating, GAO Huawei, LIU Like. Dissection of genetic characteristics of storability via GWAS and haplotype analysis of candidate genes in soybean[J]. Journal of Liaocheng University (Natural Science Edition), 2025, 38(6): 997-1005.

文章编号 1672-6634(2025)06-0997-09

DOI 10.19728/j.issn1672-6634.2025030004

大豆耐储藏性相关位点的GWAS定位及 候选基因的单倍型分析

朱沂昕¹,李重阳¹,王俊²,李敏敏³,耿雅婷¹,高华伟^{4,5},刘立科¹

(1. 聊城大学 农业与生物学院,山东 聊城 252059;2. 神农实验室,河南 郑州 450002;3. 聊城市农业科学院,山东 聊城 252000;
4. 中国农业科学院作物科学研究所,北京 100081;5. 中国农业科学院南繁研究所,海南 三亚 572024)

摘要 大豆种子的耐储藏性是决定储藏种子发芽势和发芽率的重要因素,对农业生产和食品加工具有重要意义。通过测定879份大豆种质自然条件下储藏3.5年种子的发芽势和发芽率,利用全基因组关联研究(genome-wide association study, GWAS),检测到58个显著性SNP位点,其中1个SNP位点与发芽势显著相关,57个SNP位点与发芽率显著相关。这些显著相关的SNP位点或孤立或成簇分布在10条不同的染色体上,将位于第2、9和10号染色体成簇分布的SNP位点确认为发芽率相关的3个QTL。其定位区间与5个已知的QTL位点区间重叠并包含2个候选基因 *Glyma.02g170500* 和 *Glyma.09g182900*。根据重测序SNP数据对这2个基因进行单倍型分析,结果表明 *Glyma.02g17050* 基因的11个标签SNP主要位于非翻译区,可以划分为4个block,其中3个block的不同单倍型间发芽势和发芽率差异显著; *Glyma.09g182900* 具有38个标签SNP,形成一个大的block,共有840个单倍型,因每种单倍型所含的种质数量太少而未能进行单倍型分析。在候选基因 *Glyma.02g17050* 的不同block中同时具有高发芽势和高发芽率的优良单倍型拥有黄淮地区来源的种质多于其他地区的种质,地方种质多于栽培种质。本研究结果将为后续培育耐储藏性大豆品种和解析大豆耐储藏遗传机理提供理论基础和技术支持。

关键词 大豆;耐储藏性;发芽率;发芽势;GWAS;单倍型分析

中图分类号 S565.1

文献标志码 A

开放科学(资源服务)标识码(OSID)



本文电子版文后含有增强出版内容。

Dissection of genetic characteristics of storability via GWAS and haplotype analysis of candidate genes in soybean

ZHU Yixin¹, LI Chongyang¹, WANG Jun², LI Minmin³,
GENG Yating¹, GAO Huawei^{4,5}, LIU Like¹

(1. College of Agriculture and Biology, Liaocheng University, Liaocheng 252059, China; 2. The Shennong Laboratory, Zhengzhou 450002, China; 3. Liaocheng Academy of Agricultural Sciences, Liaocheng 252000, China; 4. Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China; 5. National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572024, China)

收稿日期:2025-03-06

基金项目:国家重点研发计划项目(2016YFD0100201-09);农业农村部农业科技重大项目;聊城市“科技副总”协同创新项目(2024XT06)资助

通信作者:刘立科,男,汉族,博士,教授,博士生导师,研究方向:作物遗传育种,E-mail:liulike@lcu.edu.cn.

Abstract The storability of soybean seeds is a crucial factor in determining the germination potential (GP) and germination rate (GR) of stored seeds, which are of great significance for agricultural production and food processing. In this study, the GP and GR of 879 soybean germplasms stored under natural conditions for 3.5 years were measured. Through a genome-wide association study (GWAS), 58 significantly associated SNPs were found, among which one SNP was associated with GP and 57 SNPs were related to GR. These significant SNPs were distributed either independently or in clusters on ten different chromosomes. The SNP clusters on chromosomes 2, 9, and 10 were identified as three QTLs related to GR. Five known QTLs that overlapped with these three QTLs were considered co-located QTLs, and two genes, namely *Glyma.02g170500* and *Glyma.09g182900*, were identified as candidate genes. The results of haplotype analysis showed that 11 tag SNPs located in the non-coding region of *Glyma.02g170500* could be divided into 4 blocks, with GP and GR differing significantly between different haplotypes in all blocks except the block 2. A total of 38 tag SNPs were found for *Glyma.09g182900*, forming 840 haplotypes. Due to the small number of accessions for each haplotype, further analysis was not conducted. The haplotype with significantly higher GP and GR than the other haplotypes contained more accessions from the Huang-Huai region than from other regions, and more landraces than improved cultivars. The results of this study will provide a theoretical basis and technical support for the subsequent breeding of soybean varieties with high storability and the elucidation of the genetic mechanism of soybean seed storability.

Keywords soybean; storability; germination rate; germination potential; GWAS; haplotype analysis

0 引言

大豆[*Glycine max* (L.) Merr]是世界主要油料作物,也是我国重要的粮食和经济作物。大豆种子中的蛋白质约占 40%,脂肪约占 20%,在储藏过程中极易因环境中温度和湿度的改变发生变质,使种子活力下降。在大豆生产中,表现为出苗率降低、生长势下降、抗逆能力变差,进而造成减产和经济损失;同时也是大豆种质资源保存方面一直面临的问题^[1]。因此,相较于其他作物,大豆种子的耐储藏性较差,常被视作短命种子^[2]。种子活力是一种包含种子寿命、幼苗生长速度和幼苗发育过程中抗逆性的复杂性状,受到环境因素与复杂的生理、遗传和分子机制的综合影响^[3]。在自然条件下,随着储藏时间的延长,种子老化现象严重。许多生物化学和植物生理学的变化,如酶活性的丧失,膜完整性的改变和遗传性的变化,都会导致种子活力下降^[4]。目前研究种子老化的方法主要是自然老化和人工老化,前者是将种子保存在自然环境中,后者是将种子置于人为模拟的较极端环境中,加速种子老化进程^[5]。由于自然老化所需时间较长,因此在研究中通常采用高温高湿的人工老化方式加速种子老化^[6]。

大豆的耐储藏性在不同种质间具有较大的差异,不同大豆种质在相同储藏条件下的老化程度不同,对不利储藏条件表现出的抗性水平也不同^[7-9]。大豆种子的耐储藏性属于复杂的数量性状,受多基因控制^[7],且存在着显著的基因型与环境的互作^[10]。Jiamtae 等人通过对 4℃条件下储藏 8 年的 109 个重组自交系和 4℃条件下储藏 1 个月的 90 个重组自交系进行不同温度下的自然老化 6 个月和人工老化处理,定位了 5 个与种子生活力(seed viability)和种子活力(seed vigor)性状相关的 QTLs,分别位于第 2、6 和 8 号染色体上^[11]。Dargahi 等对来自于不耐储藏的菜用大豆和耐储藏的地方品种的杂交 F_{2:3} 和 F_{2:4} 株系进行自然老化和人工老化处理,利用复合区间作图法定位到 3 个耐储藏相关 QTL 位点^[10]。战宇航等人通过 112 份重组自交系群体和 343,907 个 SNP(single nucleotide polymorphism)标记定位到与大豆种子耐储藏性相关性状的 31 个 QTLs,共覆盖大豆的 10 条染色体^[12]。Wu 等通过全基因组关联分析(genome-wide association study, GWAS)对 168 份大豆种质进行储藏耐受性评价,利用筛选出的 23156 个高质量 SNP 定位到 14 个与大豆耐储藏性相关的位点,并结合表达模式分析鉴定出两个可能参与种子耐储藏性的候选基因(*Glyma.03g058300*, *Glyma.16g074600*)^[1]。总体而言,对于大豆种子耐储藏性有关基因的研究虽然较多,但不同

研究所定位到的 QTL 位点的一致性并不高,所用材料也主要集中于双亲组合来源的重组自交系或少量种质,因此,应用更多的种质资源,更加全面地解析大豆的耐储藏性非常重要。

因此,本研究通过测定自然老化 3.5 年的大豆种质的发芽势和发芽率指标对 879 份种质的耐储藏性进行评价,并利用 1,493,355 个高质量 SNP 标记对发芽势和发芽率性状进行 GWAS 分析,定位大豆种子耐储藏性相关的 QTLs。同时,对耐储藏性相关 QTLs 的候选基因进行单倍型分析,鉴定优异单倍型,为研究大豆种子耐储藏性的遗传机理奠定基础,同时也为大豆种子耐储藏性育种改良提供基因资源。

1 材料与方法

1.1 实验材料

本研究所用的 879 份大豆种质由中国农业科学院作物科学研究所邱丽娟研究员提供,于 2017 年夏季播种于山东省聊城市东昌府区道口铺试验田。这些大豆种质包括 399 份地方种质、367 份栽培种质和 113 份未知类型种质,地理来源于东北地区(221 份)、黄淮地区(329 份)、南方地区(62 份)、美国(61 份)和未知来源地区(206 份),具体种质类型的信息见 Li 等^[13]。

1.2 自然老化处理

2017 年收获的大豆种子在自然条件下保存,至 2021 年春季取出,储藏时长约为 3.5 年。

1.3 种质的耐储藏性评价

以自然老化种子的发芽势(Germination Potential, GP)和发芽率(germination rate, GR)作为衡量种质的耐储藏性指标,测定方法参考《粮油检验发芽试验》进行^[14]。对于种子量足够的大豆种质,挑选 30 粒饱满且完整的种子用于单次实验,并进行 3 次重复;种子量不充足的种质进行 2 次或 1 次实验。测试种子经 2% 次氯酸钠溶液杀菌后,使用无菌水清洗干净,然后用于实验。将两层滤纸置于培养皿中,用无菌水湿润滤纸,作为发芽床。将处理好的测试种子放入培养皿中,然后置于 25 °C 的人工气候培养箱中进行发芽实验,期间保持滤纸湿润。每天观察并记录种子的发芽情况,在第 4 天统计种子发芽势,第 7 天统计发芽率。

$$\eta_{\text{发芽势}}(\text{GP}/\%) = (M_{\text{第4天正常发芽种子数}} / M_{\text{供试种子总数}}) \times 100\%;$$

$$\eta_{\text{发芽率}}(\text{GR}/\%) = (M_{\text{第7天正常发芽种子数}} / M_{\text{供试种子总数}}) \times 100\%。$$

1.4 全基因组关联研究定位耐储藏性相关 SNP 位点

使用 EMMAX 软件中的混合线性模型(mixed linear model, MLM)对 879 份大豆种质的发芽势和发芽率数据进行 GWAS 分析。在分析过程中剔除等位基因频率小于 0.05 和缺失数据大于 0.2 的 SNP 位点。采用多重假设检验校正方法 Bonferroni 法对阈值进行调整。显著性 SNP 的阈值设置为 0.001/SNP,总数将检测到超过阈值的关联位点视为与性状显著关联的 SNP 位点,并绘制曼哈顿图(Manhattan Plot)和 Quantile-Quantile 图(Q-Q Plot)。

1.5 耐储藏相关 QTL 的确定及其候选基因和共定位 QTL 的筛选

在 Soybase 网站(www.soybase.org)将与种子活力相关的显著 SNP 位点,映射到大豆参考基因组“Williams 82. a2. v1”物理图谱和连锁图谱上。为了降低随机误差造成的定位错误,在 1 Mb 区间内至少有 3 个连续的显著 SNP 位点存在的区间,才被认定为是与耐储藏性相关的 QTL。根据参考基因组注释信息,将与耐储藏相关 QTL 定位区间有重叠的已知 QTL 位点确定为共定位 QTL,将耐储藏相关 QTL 定位区间内的基因确定为候选基因。

1.6 耐储藏相关 QTL 位点候选基因的单倍型分析

通过 vcftools 和 bcftools 软件对 879 个大豆种质资源重测序结果 vcf 数据文件进行处理,根据目标基因在染色体上的定位区间(参考基因组为大豆品种 Williams 82. a2. v1),获取基因内部 SNP 变异信息^[15]。利用 TASSEL 5.0 软件中基于 LD 的 KNNi 方法对 879 份种质的 SNP 数据缺失值进行填充。利用 R 程序包 geneHap 软件进行单倍型鉴定^[16],以 $R^2 > 0.8$ 的阈值确定每个 block 内的单倍型序列及频率(去除杂合及缺失位点),并将单倍型序列与表型数据进行关联分析,确定目标基因内每个 block 内的单倍型序列与表型之间是否存在显著差异。使用 Exon-Intro Graphic Maker 绘制基因结构。

1.7 数据分析与统计

利用 Excel 2016 整理数据,使用 SAS 9. 4(SAS Institute Inc. , Cary, North Carolina)进行所有的统计分析。

2 结果

2.1 自然老化处理 879 份大豆种质的发芽势和发芽率分布

对自然条件下保存 3.5 年的 879 份大豆种质进行发芽实验,获得 GP 和 GR 数据。结果表明,GP 和 GR 的数值范围分别为 0% ~56. 67%和 0% ~90. 00%,均值分别为 1. 18%和 3. 65%(图 1)。在自然条件下储藏 3.5 年后,约 80%的大豆种质完全丧失萌发能力,剩余 20%的大豆种质具有不同的 GP 和 GR。GP 的范围在 4. 00% ~56. 67%,只有 2 份大豆种质的 GP > 50%,即大豆(17D12901)和 PI438342(17D24710);GR 的范围在 1. 43% ~90. 00%之间,有 3 份种质的 GR ≥ 80%,即小黑豆(17D22211),PSB313(17D14105)和 PI438342(17D24710)。

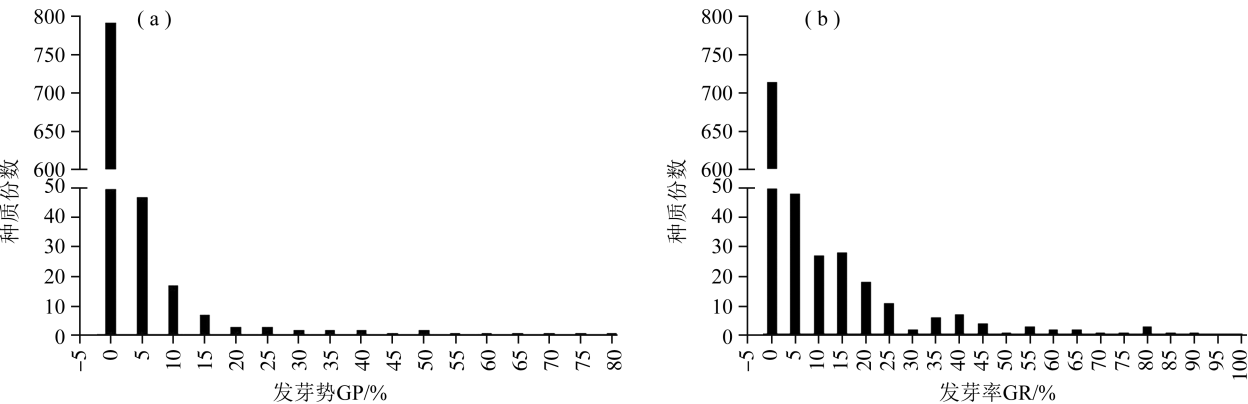


图 1 自然老化 3.5 年后 879 份大豆种质 GP(a)和 GR(b)的分布

Fig. 1 GP (a) and GR (b) distribution patterns of 879 soybean germplasms after 3.5 years of natural aging

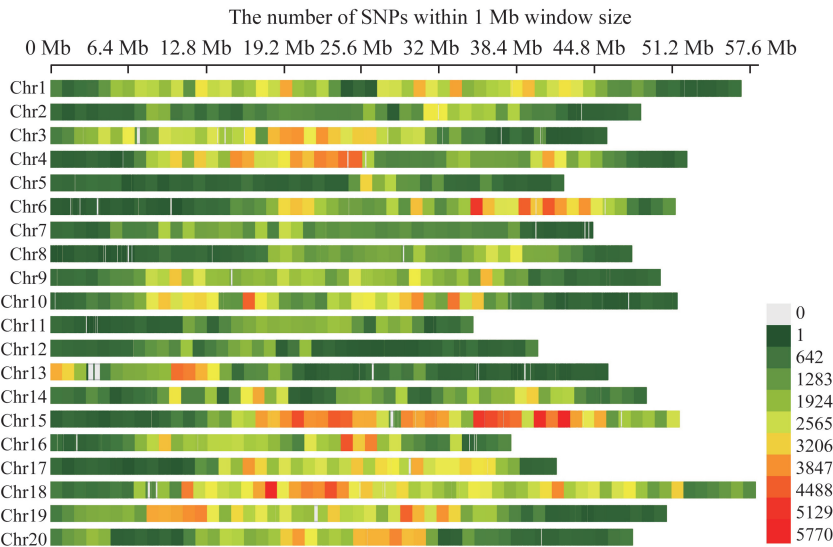


图 2 发芽势和发芽率 GWAS 分析用 SNP 标记在不同染色体上的分布

Fig. 2 Distribution of SNPs on 20 chromosomes of soybean genome for GP and GR GWAS analysis

2.2 全基因组关联研究定位发芽势和发芽率相关 QTL 位点

在去除等位基因频率小于 0.05 和缺失数据大于 0.2 的位点之后,在大豆基因组中共筛选出 1 493 355 个 SNP 位点,覆盖大约 955.38 Mb,平均密度是 1 563 个 SNP / Mb(图 2)。在 $-\log_{10}(p \text{ 值})$ 为 9.17(红线)时,有 1 个 SNP 与 GP 显著相关,57 个 SNP 与 GR 显著相关。这些 SNP 单独或成簇分布在第 1、2、3、4、7、

9、10、19 和 20 号染色体上,如图 3(a, b)所示。与 GP 相关的显著性 SNP 位点只有 1 个,附近没有其他显著的 SNP 位点作为支撑,因此该位点并没有被认定为是与 GP 相关的可靠 QTL 如图 3(a)所示;与 GR 相关的 3 个 QTL 分别被定位到第 2、9 和 10 号染色体上,如图 3(b)所示。其中 $qGR1$ 与 3 个已知的 QTL 和基因 *Glyma. 02g170500* 共定位, $qGR2$ 与 1 个已知的 QTL 和基因 *Glyma. 09g182900* 共定位, $qGR3$ 与 1 个已知的 QTL 共定位(表 1)。

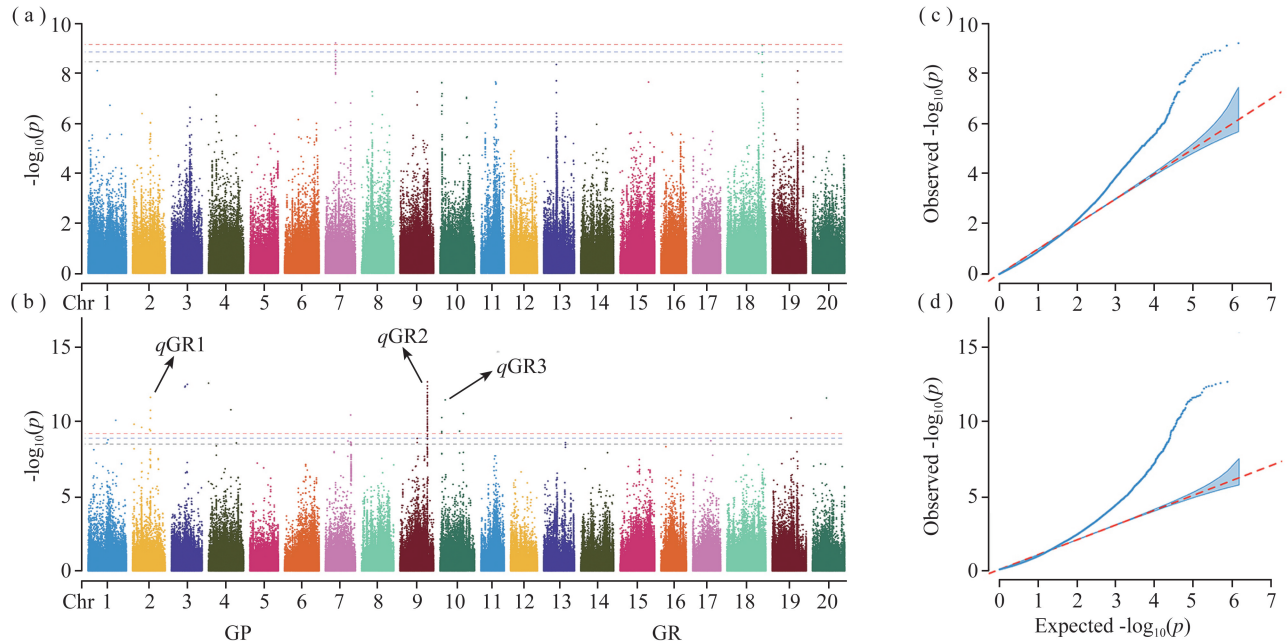


图 3 GP 和 GR 的 Manhattan plots(a, b)和 Q-Q plots(c, d)。在 Manhattan plots 中,红色、蓝色和灰色水平线分别代表不同的全基因组显著性阈值: $-\log_{10}[\text{阈值 } 1(0.001/\text{总 SNP})]=9.17$ 、 $-\log_{10}[\text{阈值 } 2(0.002/\text{总 SNP})]=8.87$ 和 $-\log_{10}[\text{阈值 } 3(0.005/\text{总 SNP})]=8.48$ 。Q-Q plots 的阴影区域表示 95% 的置信区间

Fig. 3 Manhattan plots (a, b) and Q-Q plots (c, d) of GP and GR, utilizing EMMAX software. In the Manhattan plots, the red, blue, and gray horizontal lines represent different genome-wide significance thresholds: $-\log_{10}[\text{Threshold } 1(0.001/\text{total SNP})]=9.17$, $-\log_{10}[\text{Threshold } 2(0.002/\text{total SNP})]$, and $-\log_{10}[\text{Threshold } 3(0.005/\text{total SNP})]=8.48$, respectively. The shaded regions of the Q-Q plots indicate a 95% confidence interval

表 1 利用 EMMAX 定位到的 GR 相关 QTL
Tab. 1 GR related QTLs identified by EMMAX

染色体	QTL	起始位点 (bp)	终止位点 (bp)	显著 SNP 数量	候选基因与共定位 QTLs
2	$qGR1$	25 484 690	26 602 170	9	<i>Glyma. 02g170500</i>
					Seed linolenic 7-3 ^[17]
					Seed isoflavone 11-1 ^[18]
					Seed tocopherol, total 2-3 ^[19]
9	$qGR2$	40 797 191	40 927 437	33	<i>Glyma. 09g182900</i>
10	$qGR3$	2 562 823	2 568 701	3	Seed oil 43-22 ^[20]
					Seed oil 24-2 ^[21]

2.3 发芽率相关 QTL 位点候选基因的单倍型分析

2.3.1 $qGR1$ 位点候选基因 *Glyma. 02g170500* 的单倍型分析。根据参考基因组注释信息, *Glyma. 02g170500* (Chr02:26593977..26611436, forward) 位于 2 号染色体上, 全长为 17 371 bp, 由 5' 非翻译区、23 个外显子、23 个内含子和 3' 非翻译区构成, 如图 4(a)。 *Glyma. 02G170500* 的编码区全长 2 970 bp, 编码蛋白有 989 个氨基酸, 含有肌球蛋白和驱动蛋白的马达结构域(myosin and kinesin motor domain)。分析 879 个大豆种质资源 SNP 数据, 在 *Glyma. 02g170500* 基因区域内共存在 460 个 SNPs。经过连锁不平衡分析,

得到 11 个标签 SNPs,其中前 4 个 SNPs 位于 5'非翻译区,第 5 个 SNP 是位于编码区的一个同义突变位点,其余后面 6 个 SNPs 均位于内含子区,如图 4(a)。根据连锁不平衡分析,这 11 个标签 SNPs 被划分为 4 个 blocks,如图 4(b)。在 block 1、block 3 和 block 4 内,不同单倍型间的 GP 和 GR 显著性差异,而 block2 的不同单倍型间 GP 和 GR 差异不显著,因此后续不再分析(图 S1)。

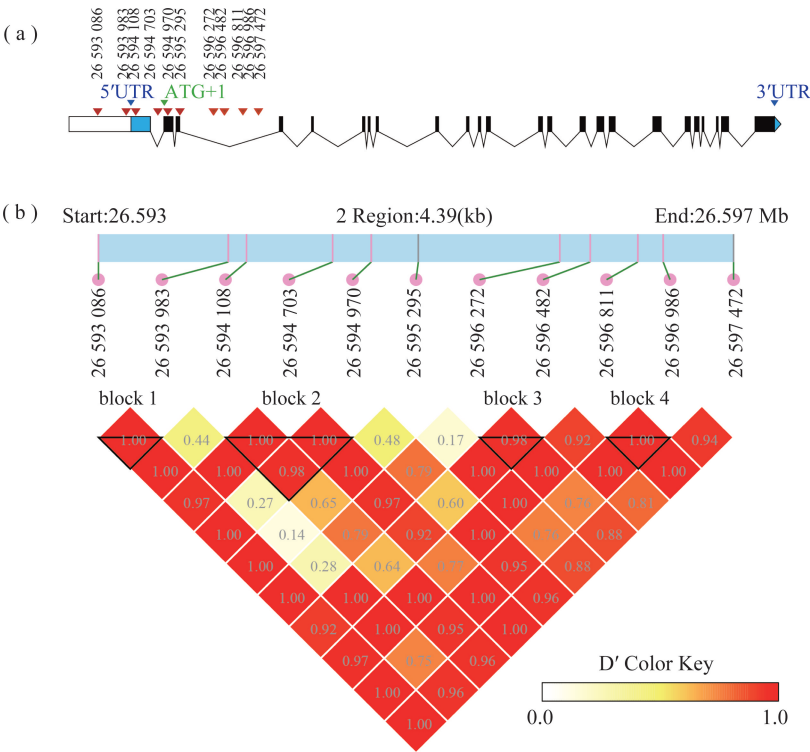


图 4 *Glyma.02g170500* 的基因结构(a)与连锁不平衡分析(b)

Fig. 4 Gene structure of *Glyma.02g170500* (a) and linkage disequilibrium analysis of *Glyma.02g170500* (b)

在 block 1 中共有 3 种单倍型,其不同单倍型依次包含 416、368 和 94 份种质。单倍型 Hap001 的 GP 和 GR 显著高于其他两个单倍型,其包含来自黄淮地区的种质最多(48.56%),而其他两个单倍型中黄淮地区种质分别占 28.26%和 24.47%。Hap001 中地方种质占 60.82%,而其他两个单倍型中地方种质分别占 30.98%和 32.98%,如图 S2(a)。block 3 共有 4 种单倍型,依次包含有 443、353、82 和 1 份种质。因 Hap004 仅有 1 份种质,不具有统计学意义,因此不再进行后续的分析。同 block 1 类似的是,block 3 的 Hap001 也具有显著高于其他两种单倍型的 GP 和 GR,其含有的黄淮地区的种质同样占最大的比例(46.28%),地方种质占多数为 58.69%,而其他两个单倍型中黄淮地区的种质和地方种质所占比例均不足 1/3,如图 S2(b)。block 4 共有 3 种单倍型,同上述两个 block 中的单倍型类似,Hap001 的 GP 显著高于其他两个单倍型,但 Hap001 的 GR 只显著高于 Hap002。在 block 4 的 Hap001 中,同样来源于黄淮地区的种质占最大的比例(49.30%),其他两个单倍型中黄淮种质都不到 1/3;地方种质在 Hap001 中占 61.63%,而在 Hap002 和 Hap003 中分别为 29.30%和 44.91%,如图 S2(c)。整体上看,在 block1、block3 和 block4 中,具有高 GP 和 GR 的单倍型中来源于黄淮地区的种质均占多数、地方种质占大多数,表明不同大豆生态区的大豆在驯化和改良过程中虽然优良单倍型存在一定程度上的选择但并未被固定。

2.3.2 *qGR2* 位点候选基因 *Glyma.09g182900* 的单倍型分析。根据 879 份大豆种质资源 SNP 数据,在 *Glyma.09g182900* 基因区域内共存在 165 个 SNPs,通过连锁不平衡分析,得到 Gm09_40805636 等 38 个标签 SNPs,被划分为 1 个总的 block(图 S3)。由于包含的 SNPs 数量较多且存在多个 SNP 数据缺失,该 block 共形成了 840 个单倍型,只有 Hap001 和 Hap002 所含种质数量超过 10 个,且其 GP、GR 差异不显著;其他单倍型种质数量较少,不具有统计学意义,故不再进行后续分析(图 S4)。

3 讨论

本研究通过 GWAS 鉴定出 3 个与 GR 相关的 QTL,但未能鉴定出与 GP 相关的 QTL,其原因可能在于在自然老化种子的发芽试验过程中,因种子活力不足,常出现早期未发芽但后期发芽的现象或已经发芽的种子因子叶腐烂而产生发芽后期失败的现象,造成 GP 和 GR 统计数据一致性较差,进而导致关联分析结果的差异。另外,在其他作物上也发现这两个指标之间有不一致的现象,例如在水稻中,发芽势和发芽率并不一致,特别是低发芽势种子,往往会发生烂种等现象导致实际发芽率降低^[22];不同储藏条件下,虽然两者的变化趋势相同,但发芽势的变化幅度大于发芽率,也表明这两个指标间存在差异^[23]。虽然如此,GP 和 GR 仍是两个紧密相关的两个指标,在本研究中其相关系数为 0.76($p < 0.001$)达到极显著水平。虽然未定位到与 GP 相关的 QTL,但在 GR 相关 QTL 位点候选基因的单倍型分析过程,多个不同单倍型间 GP 和 GR 差异显著性是一致的,这也从另一个方面表明这两个指标是紧密相关的。

根据 Soybase 网站注释,这 3 个 GR 相关 QTL 定位区间与 5 个已报道的其他性状 QTL 的定位区间重叠,但与已报道的种子耐储藏性 QTL 无重叠,5 个已知的位点分别为 *seed linolenic* 7-3^[17]、*seed isoflavone* 11-1^[18]、*seed tocopherol total* 2-3^[19]、*seed oil* 43-22^[20]和 *seed oil* 24-2^[21]。此外,在定位区间内,还筛选到 2 个与 GR 相关 QTL 的候选基因。大豆异黄酮(isoflavone)具有抗氧化作用,其含量随储藏时间和温度的增加而降低^[24]。生育酚(tocopherol)是一种小分子抗氧化剂,通过降低活性氧水平和抑制脂质过氧化作用,对大豆油的储藏稳定性起关键作用^[25]。异黄酮与亚麻酸(linolenic)呈显著正相关,与油脂含量呈显著负相关,说明油脂积累和脂肪酸组成可能通过影响异黄酮浓度影响大豆种子活力^[26]。在 25 °C/75%相对湿度条件下,种子含油量与储藏寿命呈高度负相关^[27]。这可能是由于油脂中多不饱和脂肪酸易氧化,在储藏条件和储藏时间的影响下,容易发生过氧化而产生一些有毒物质,从而导致种子活力下降甚至丧失。这 3 个 GR 相关 QTL 的 2 个候选基因分别是:*Glyma. 02g170500* 和 *Glyma. 09g182900*,它们分别编码微管马达蛋白和 RNA 结合蛋白。这些候选基因的产物均未直接参与上述共定位性状的代谢途径,那么这些候选基因是通过与其他基因互作来影响上述性状还是候选基因仅与这些共定位性状 QTL 连锁遗传,尚需进一步的研究证实。

本研究中,*qGR1* 位点的候选基因编码肌球蛋白和驱动蛋白的马达结构域,而肌球蛋白和驱动蛋白广泛地参与了细胞生长过程中的细胞骨架的动态和组织调整^[28]。再考虑到在此候选基因的不同连锁区块内不同单倍型间 GP 和 GR 差异显著,因此该基因很可能就是该 QTL 的候选基因。RNA 结合蛋白以其 RNA 识别基序(RRM)结合到靶转录物中的特定二级结构或序列上,并在多种过程中发挥关键作用,包括选择性剪接、mRNA 核输出、mRNA 稳定性和翻译。脱落酸(ABA)调节的 RNA 结合蛋白 1(ARP 1)突变体的过表达和敲低株系都会导致 ABA、高盐和脱水胁迫下种子发芽延迟,这表明 ARP 1 转录物水平的精确调控是正常种子发芽的必要条件^[29]。*qGR2* 位点候选基因也编码 RNA 结合蛋白,可能具有与 ARP1 相似的功能。但此候选基因中因为单倍型过多而导致每个单倍型所具有的种质数量太少,无法对单倍型间的 GP 和 GR 的差异显著性进行分析,尚需通过扩大群体样本和提高 SNP 数据质量进行进一步的分析。当然,上述两个候选基因是否真正地与大豆种子耐储藏相关,尚需要通过进一步的实验验证。

在 *qGR2* 候选基因 *Glyma. 02g170500* 的高 GP 和 GR 的单倍型中,均是来源于黄淮地区的种质占大多数,这可能是由于黄淮地区的大豆种质能够更好地适应同属黄淮地区的聊城当地的自然环境,因而其种子质量好于其他地区来源的种质。当然也有另外一种可能,即黄淮地区的种质本身具有比其他地区种质更好的耐储藏性。要解决此问题,尚需进一步的研究证实。从种质类型上看,具有高 GP 和 GR 的单倍型中,地方种质均占大多数,表明在地方种质中存在着较多的耐逆优良等位基因。这种现象在作物改良和驯化过程中屡见不鲜,比如地方种质对疫病的抗性明显好于栽培种质^[30];地方种质的瘪荚率也低于栽培种质^[31];大豆包囊线虫抗性位点 *rhg1-b*、*rhg1-a* 和 *Rhg4* 仅存在于 PI88788 和北京小黑豆两个地方种质中,而在栽培种质中不存在^[32]。

基因编码区域突变导致的基因功能丧失通常会引起显著的表型变化,而基因调控区域的突变往往仅改

变基因转录水平和表达模式,对生长的影响相对较小,但仍是性状转变的重要遗传基础,仍是作物驯化和改良过程中的重要选择位点^[33]。比如,在水稻中影响分蘖夹角的 *TIG1* 基因,该基因启动子区域在籼稻中的突变导致转录水平降低,进而造成分蘖由野生水稻的倾斜生长变为直立生长^[34]; *qSH1* 基因启动子调控区域中的一个 SNP 位点导致粳稻中离层缺失而丧失了落粒性^[35]。除了在启动子区域内的突变会影响基因的表达,5'UTR 的碱基突变^[36]和内含子区域的突变也同样会影响基因的表达^[37]。在本研究中, *qGR1* 位点候选基因内的 SNP 位点主要位于 5'UTR 和内含子等非编码区,表明该候选基因对于大豆种子的耐储藏性的影响可能主要是通过调控转录和转录后的基因表达水平而非通过改变基因的功能来实现的,但其具体的分子机理尚待进一步的研究。

4 结论

本研究利用自然老化 3.5 年的 879 份种质资源的发芽势和发芽率数据,通过全基因组关联分析定位到了 3 个发芽率相关的 QTL,并根据定位区间找到 2 个 QTL 的候选基因 *Glyma. 02g170500* 和 *Glyma. 09g182900*。经过单倍型分析,发现 *Glyma. 02g170500* 基因的 11 个标签 SNPs 主要位于非编码区,形成 4 个连锁 block,其中 3 个 block 的不同单倍型间 GP 和 GR 差异显著。在候选基因 *Glyma. 02g170500* 的高 GP 和 GR 单倍型中来源于黄淮地区的种质占近 50%,而在其他单倍型中占比不到 1/3;地方种质占比约为栽培种质的两倍,而在其他单倍型中两者占比差别不大或栽培种质占据优势。本研究得到的结果将为后续大豆耐储藏性遗传基础解析及耐储藏大豆品种的培育奠定基础。

参 考 文 献

- [1] WU X, WANG Y, XIE J, et al. Identification of candidate genes for soybean storability via GWAS and WGCNA approaches[J]. *Agronomy*, 2024, 14, 2457.
- [2] 唐志华, 马继凤. 大豆种子活力研究进展[J]. *作物研究*, 2007, 21(5): 625-628.
- [3] ZINSMEISTER J, LEPRINCE O, BUITINK J. Molecular and environmental factors regulating seed longevity[J]. *Biochemical Journal*, 2020, 477(2): 305-323.
- [4] PARRISH D J, LEOPOLD A C. On the Mechanism of aging in soybean seeds[J]. *Plant Physiology*, 1978, 61: 365-368.
- [5] 范玲娟. 三种人工老化方法对大豆种子活力和生理生化影响[D]. 太原: 山西大学, 2007.
- [6] MARLI M G F, FRANCISCO C K, PATRICIA P E, et al. Standardization of the accelerated aging test for evaluation of wheat seed vigor[J]. *Journal of Seed Science*, 2024, 46: 1-10.
- [7] SHELAR V R, SSHAUKHR S, NIKAM A S. Soybean seed quality during storage: a review[J]. *Agricultural Reviews*, 2008, 29(2): 125-131.
- [8] 高华伟, 满强, 武猛, 等. 耐储藏大豆种质资源的筛选[J]. *大豆科学*, 2014, 33(1): 6-12.
- [9] CHANDRA S, TALUKDAR A, TAAK Y, et al. Seed longevity studies in wild type, cultivated and inter-specific recombinant inbred lines (RILs) of soybean [*Glycine max* (L.) Merr.] [J]. *Genetic Resources and Crop Evolution*, 2021, 69(1): 399-409.
- [10] DARGAHI H, TANYA P, SRINIVES P. Mapping of the genomic regions controlling seed storability in soybean (*Glycine max* L.). [J]. *Journal of Genetics*, 2014, 93: 365-370.
- [11] JIANTAE P, Hashiguchi M, Lelapiyarnit I, et al. QTL analysis for soybean (*Glycine max* L. Merr.) seed storability in high-temperature storage conditions[J]. *Euphytica*, 2022, 218: 169.
- [12] 战宇航, 吴旭, 赵雪, 等. 大豆种子耐储藏性相关 QTL 分析及候选基因挖掘[J]. *大豆科学*, 2023, 42(4): 406-415.
- [13] LI Y H, QIN C, WANG L, et al. Genome-wide signatures of the geographic expansion and breeding of soybean[J]. *Science China Life Science*, 2023, 66(2): 350-365.
- [14] 河南工业大学, 广东国家粮食质量监测中心, 中央储备粮天津东丽直属库粮油检验 发芽试验: GB/T 5520-2011[S]. 中华人民共和国国家质量监督检验检疫总局 (中国国家标准化管理委员会), 2011-11-01.
- [15] SCHMUTZ J, CANNON S B, SCHLUETER J, et al. Genome sequence of the palaeopolyploid soybean[J]. *Nature*, 2010, 463(7278): 178-183.
- [16] ZHANG R, JIA G, DIAO X. geneHapR: an R package for gene haplotypic statistics and visualization[J]. *BMC Bioinformatics*, 2023, 24(1): 199.

- [17] BACHLAVA E, DEWEY R E, BURTON J W, et al. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations[J]. *Crop Science*, 2009, 49(2): 433-442.
- [18] HAN Y, Teng W, Wang Y, et al. Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents[J]. *Plant Breeding*, 2015, 134(3): 300-309.
- [19] LI H, WANG Y, HAN Y, et al. Mapping quantitative trait loci (QTLs) underlying seed vitamin E content in soybean with main, epistatic and QTL $\times$ environment effects[J]. *Plant Breeding*, 2016, 135(2): 208-214.
- [20] MAO T, Jiang Z, Han Y, et al. Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments[J]. *Plant Breeding*, 2013, 132(6): 630-641.
- [21] QI Z M, WU Q, HAN X, et al. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes[J]. *Euphytica*, 2011, 179(3): 499-514.
- [22] 贾军, 石广跃, 谢红梅. 杂交稻种子发芽率与发芽势不一致的原因及其对策[J]. *杂交水稻*, 2005, 20(6): 33-34.
- [23] 彭毛, 张欣, 左文杰. 储藏方式对稻谷发芽势和发芽率的影响[J]. *粮油食品科技*, 2017, 25(5): 62-65.
- [24] KIM S H, JUNG W S, AHN J K, et al. Analysis of isoflavone concentration and composition in soybean [*Glycine max* (L.)] seeds between the cropping year and storage for 3 years[J]. *European Food Research and Technology*, 2004, 220(2): 207-214.
- [25] CHU D, ZHANG Z, HU Y, et al. Genome-wide scan for oil quality reveals a coregulation mechanism of tocopherols and fatty acids in soybean seeds[J]. *Plant Communications*, 2023, 4: 100598.
- [26] ZHANG J, GE Y, HAN F, et al. Isoflavone content of soybean cultivars from maturity group 0 to VI grown in Northern and Southern China[J]. *Journal of the American Oil Chemists Society*, 2014, 91(6): 1019-1028.
- [27] ŠIMIĆ B, VRATARIĆ M, SUDARIĆ A, et al. Effect of storage longevity under different storage conditions on seed vigor and oil content in maize, soybean and sunflower[J]. *Poljoprivreda*, 2005: 1-6.
- [28] NEBENFUHR A, DIXIT R. Kinesins and Myosins: Molecular motors that coordinate cellular functions in plants[J]. *Annual Review Plant Biology*, 2018, 69: 329-361.
- [29] JUNG H J, KIM M K, KANG H. An ABA-regulated putative RNA-binding protein affects seed germination of *Arabidopsis* under ABA or abiotic stress conditions[J]. *Journal of Plant Physiology*, 2013, 170(2): 179-184.
- [30] ZHU X, WANG W, GAO H, et al. Evaluation of soybean germplasms for resistance to stay-green syndrome[J]. *Scientific Reports*, 2025, 15(1): 8164.
- [31] 耿雅婷, 王俊, 高华伟, 等. 大豆瘪荚率的遗传解析及其遗传位点的关联分析定位[J]. *西南林业大学学报(自然科学)(网络首发)*, 2025: 1-15.
- [32] HYTEN D L, SONG Q, ZHU Y, et al. Impacts of genetic bottlenecks on soybean genome diversity[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(45): 16666-16671.
- [33] SWINNEN G, GOOSSENS A, PAUWELS L. Lessons from Domestication: Targeting cis-regulatory elements for crop improvement[J]. *Trends Plant Science*, 2019, 24(11): 1065.
- [34] ZHANG W, TAN L, SUN H, et al. Natural Variations at *TIG1* encoding a TCP transcription factor contribute to plant architecture domestication in rice[J]. *Molecular Plant*, 2019, 12(8): 1075-1089.
- [35] KONISHI S, IZAWA T, LIN S Y, et al. An SNP caused loss of seed shattering during rice domestication[J]. *Science*, 2006, 312(5778): 1392-1396.
- [36] 徐志民, 鄢梦雨, 兰海燕. 5'UTR 在植物基因表达调控中的功能研究进展[J]. *植物生理学报*, 2024, 60(1): 9-19.
- [37] BACK G, WALTHER D. Identification of cis-regulatory motifs in first introns and the prediction of intron-mediated enhancement of gene expression in *Arabidopsis thaliana*[J]. *BMC Genomics*, 2021, 22(1): 390.

(责任编辑:刘庆松)

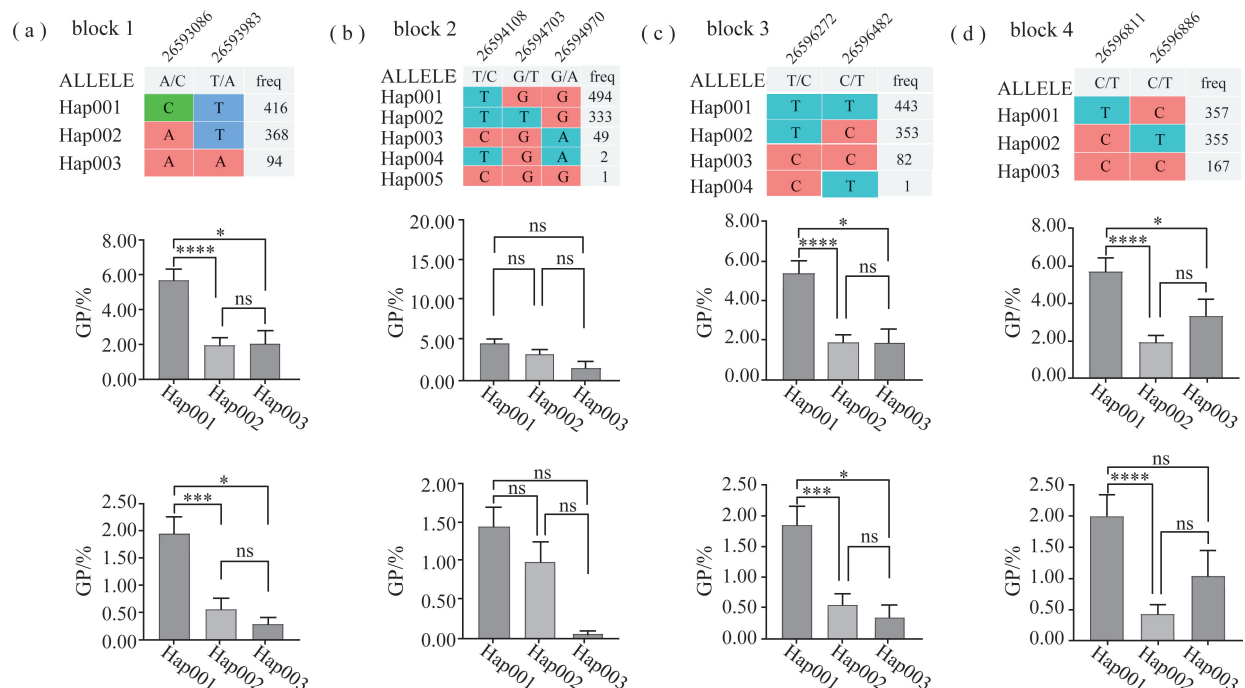
支撑材料

Supporting Information

大豆耐储藏性相关位点的 GWAS 定位及
候选基因的单倍型分析Dissection of genetic characteristics of storability
via GWAS and haplotype analysis of
candidate genes in soybean朱沂昕¹,李重阳¹,王俊²,李敏敏³,耿雅婷¹,高华伟^{4,5},刘立科¹ZHU Yixin¹,LI Chongyang¹,WANG Jun²,LI Minmin³,GENG Yating¹,GAO Huawei^{4,5}, LIU Like¹

(1. 聊城大学 农业与生物学院,山东 聊城 252059;2. 神农实验室,河南 郑州 450002;3. 聊城市农业科学院,山东 聊城 252000;
4. 中国农业科学院作物科学研究所,北京 100081;5. 中国农业科学院南繁研究所,海南 三亚 572024)

(1. College of Agriculture and Biology, Liaocheng University, Liaocheng 252059, China; 2. The Shennong Laboratory, Zhengzhou 450002, China; 3. Liaocheng Academy of Agricultural Sciences, Liaocheng 252000, China; 4. Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China; 5. National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572024, China)

图 S1 *Glyma.02g170500* 的不同连锁区内各单倍型间 GP、GR 的显著性差异分析Fig. S1 Analysis of the significant differences in GP and GR among the haplotypes within different linkage blocks of *Glyma.02g170500*

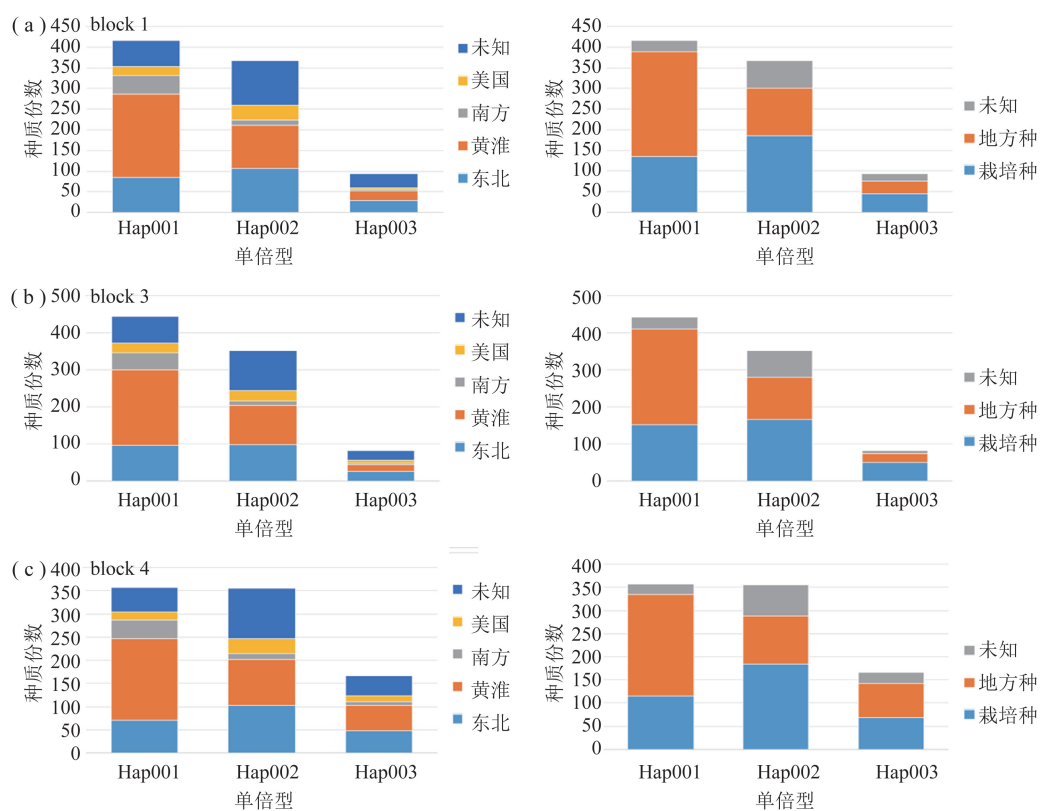


图 S2 连锁区块内不同单倍型中种质来源和种质类型的分布 (左为种质的地理来源, 右为种质的类型)

Fig. S2 The distribution of germplasm sources and germplasm types among different haplotypes within the linkage block (The geographical sources of germplasm are on the left, and the types of germplasm are on the right)

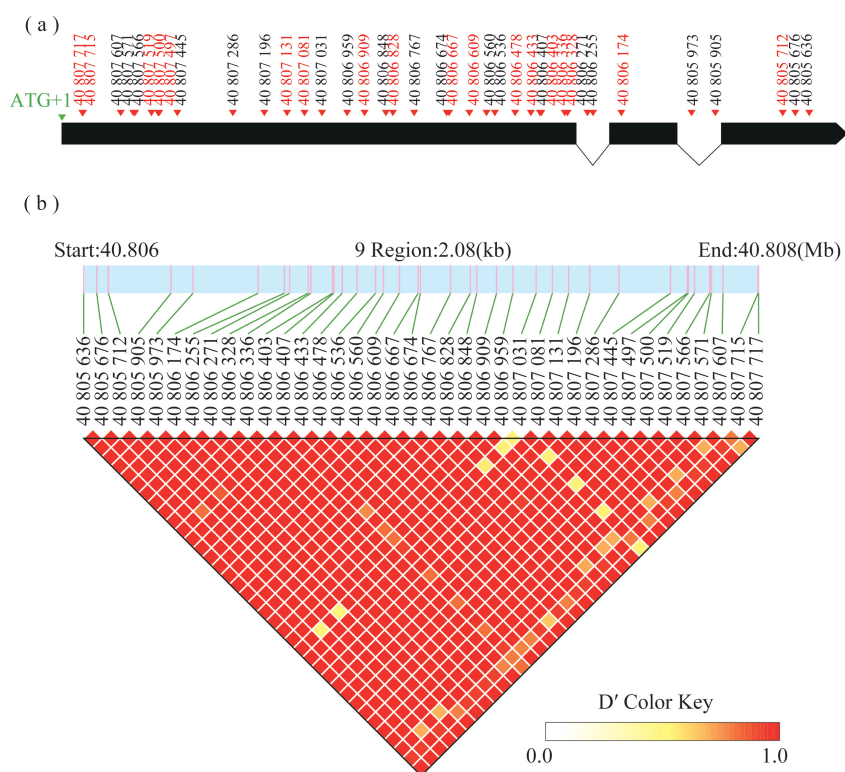


图 S3 *Glyma.09g182900* 的基因结构(a)与连锁不平衡分析(b)

Fig. S3 Gene structure of *Glyma.09g182900* (a) and linkage disequilibrium analysis of *Glyma.09g182900* (b)

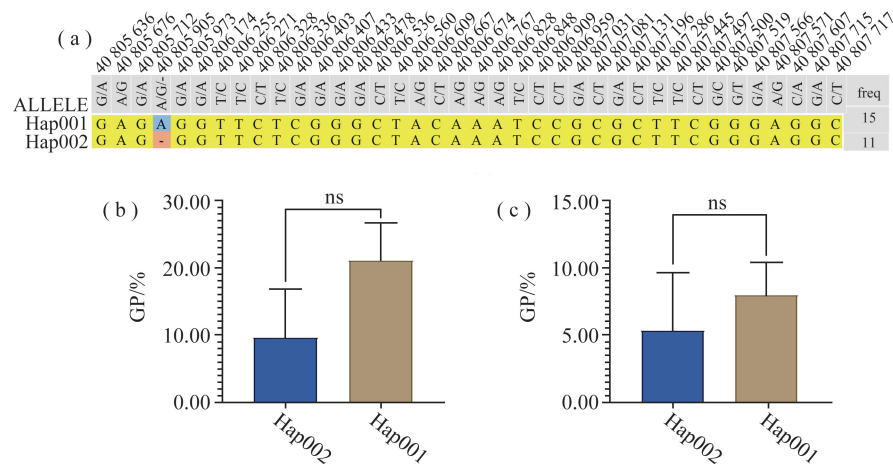


图 S4 *Glyma.09g182900* 的单倍型之间 GP、GR 的显著性差异分析

Fig. S4 Analysis of the significant differences in GP and GR among haplotypes of *Glyma.09g182900*

附表 1 以 $-\log_{10}(p \text{ 值})$ 为 9.17 筛选的发芽势和发芽率显著性相关 SNP

Supplementary Tab. 1 SNPs significantly associated with germination potential and germination rate at a threshold of $-\log_{10}(p \text{ value}) = 9.17$

SNP	染色体	物理位置/bp	$-\log_{10}(p)$	性状
Gm07_15195343	7	15 195 343	9.23	GP
Gm01_41129348	1	41 129 348	10.68	GR
Gm02_21306440	2	21 306 441	10.25	GR
Gm02_13458459	2	13 458 459	9.59	GR
Gm02_25484690	2	25 484 690	9.46	GR
Gm02_26602138	2	26 602 138	11.60	GR
Gm02_26602139	2	26 602 139	11.60	GR
Gm02_26602146	2	26 602 146	10.74	GR
Gm02_26602159	2	26 602 159	10.21	GR
Gm02_26602169	2	26 602 169	9.37	GR
Gm02_26602170	2	26 602 170	9.37	GR
Gm03_20068982	3	20 068 982	12.29	GR
Gm03_20479461	3	20 479 461	12.35	GR
Gm03_23609635	3	23 609 635	12.47	GR
Gm04_46669	4	46 669	12.55	GR
Gm04_33344050	4	33 344 050	10.77	GR
Gm07_37719748	7	37 719 748	10.42	GR
Gm09_40797191	9	40 797 191	10.02	GR
Gm09_40797217	9	40 797 217	10.88	GR
Gm09_40800473	9	40 800 473	9.89	GR
Gm09_40800850	9	40 800 850	9.66	GR
Gm09_40803213	9	40 803 213	10.14	GR
Gm09_40803374	9	40 803 374	9.38	GR
Gm09_40803720	9	40 803 720	9.42	GR
Gm09_40803848	9	40 803 848	10.14	GR
Gm09_40804753	9	40 804 753	10.65	GR

(续)附表 1 以 $-\log_{10}(p)$ 值为 9.17 筛选的发芽势和发芽率显著性相关 SNP

SNP	染色体	物理位置/bp	$-\log_{10}(p)$	性状
Gm09_40805189	9	40 805 189	9.61	GR
Gm09_40806433	9	40 806 433	11.58	GR
Gm09_40806478	9	40 806 478	10.14	GR
Gm09_40807081	9	40 807 081	9.36	GR
Gm09_40807607	9	40 807 607	10.29	GR
Gm09_40835564	9	40 835 564	9.63	GR
Gm09_40835901	9	40 835 901	9.80	GR
Gm09_40836119	9	40 836 119	11.69	GR
Gm09_40836507	9	40 836 507	11.43	GR
Gm09_40836701	9	40 836 701	12.37	GR
Gm09_40836767	9	40 836 767	11.67	GR
Gm09_40837173	9	40 837 173	11.74	GR
Gm09_40837681	9	40 837 681	11.33	GR
Gm09_40838579	9	40 838 579	12.18	GR
Gm09_40838762	9	40 838 762	11.11	GR
Gm09_40839298	9	40 839 298	11.31	GR
Gm09_40839497	9	40 839 497	11.17	GR
Gm09_40839771	9	40 839 771	11.92	GR
Gm09_40839798	9	40 839 798	16.01	GR
Gm09_40839836	9	40 839 836	12.64	GR
Gm09_40841705	9	40 841 705	11.28	GR
Gm09_40847711	9	40 847 711	10.50	GR
Gm09_40847867	9	40 847 867	10.72	GR
Gm09_40927437	9	40 927 437	9.53	GR
Gm10_2562818	10	2 562 823	9.23	GR
Gm10_2567817	10	2 567 817	9.32	GR
Gm10_2568701	10	2 568 701	10.25	GR
Gm10_7795638	10	7 795 638	11.43	GR
Gm10_29145752	10	29 145 752	9.34	GR
Gm10_34825680	10	34 825 680	10.52	GR
Gm19_28097352	19	28 097 352	10.22	GR
Gm20_20807791	20	20 807 791	11.56	GR