



大豆 *GmSWEET39* 基因在高蛋白种质 HPCD5 中的定位、功能分析与育种应用

张英虎, 薛松, 于文青, 杨红燕, 沈会权, 徐肖, 梁志浩, 程怡幡

(江苏沿海地区农业科学研究所, 江苏 盐城 224002)

摘要: 为探究高蛋白大豆新种质 HPCD5 及其籽粒蛋白质含量主效位点在品种选育中的应用价值, 本研究通过 F_2 分离群体连锁定位, 结合转录组与基因序列分析, 挖掘控制籽粒蛋白质含量的主效位点, 进一步构建重组自交系群体并开展表型与基因型检测, 明确主效位点的育种利用效率。结果表明: 在大豆 15 号染色体 2.4 ~ 5.7 Mb 的区间内检测到同时控制籽粒蛋白质、油脂和百粒重性状的位点, 该区间包含控制大豆蛋白质、油脂和百粒重性状的基因 *GmSWEET39* (*GmSWEET10a*, Glyma. 15G049200)。序列与表达分析显示, *GmSWEET39* 基因在 HPCD5 和南农菜豆 5 号中存在差异, 同时南农菜豆 5 号开花后 30 d 籽粒中该基因表达量显著高于 HPCD5, 表明 *GmSWEET39* 基因是 HPCD5 中控制籽粒蛋白质含量等性状的关键基因; 以 HPCD5 为亲本, 与南农 99-6、武德、徐豆 14 和中黄 39 配置杂交组合, 构建包含 334 份家系的重组自交系群体, 通过两年表型鉴定与基因型检测联合分析, 在 *GmSWEET39* 基因附近检测到多个与蛋白、油脂和百粒重关联的 SNP 标记, 其中基因内 SNP 标记 TZX-94 对三者的表型解释率分别为 26.9%、29.9% 和 7.8%, 明确 HPCD5 中 *GmSWEET39* 基因对大豆蛋白、油脂和百粒重具有选择效果。同时开发了 2 个 *GmSWEET39* 基因的 InDel 标记, 丰富了该基因的分子标记类型。本研究明确了 HPCD5 中来自野生大豆的控制籽粒蛋白质等性状的遗传位点, 拓展了 *GmSWEET39* 基因等位变异资源, 验证了其在多遗传背景下的育种应用潜力, 创制出多个高蛋白大豆新种质, 为高蛋白大豆品种的分子育种提供了重要基因资源与技术支持。

关键词: 大豆; 高蛋白; *GmSWEET39*; 标记辅助选择; 育种应用

Localization, Functional Analysis and Breeding Application of *GmSWEET39* Gene in High Protein Soybean Germplasm HPCD5

ZHANG Yinghu, XUE Song, YU Wenqing, YANG Hongyan, SHEN Huiquan, XU Xiao, LIANG Zhihao, CHENG Yifan

(Jiangsu Coastal Area Institute of Agricultural Sciences, Yancheng 224002, Jiangsu, China)

Abstract: To explore the application value of the high-protein soybean new germplasm HPCD5 and its major quantitative trait locus (QTL) for seed protein content in variety breeding, this study performed linkage mapping using an F_2 segregating population, combined with transcriptome and gene sequence analyses, to identify the major QTL controlling seed protein content. Furthermore, a recombinant inbred line (RIL) population was constructed, followed by phenotypic and genotypic characterization to clarify the breeding utilization efficiency of the major QTL. The results indicated that a QTL simultaneously regulating seed protein content, oil content, and 100-seed weight was detected in the interval of 2.4-5.7 Mb on soybean chromosome 15, which contains the gene *GmSWEET39* (*GmSWEET10a*, Glyma. 15G049200) responsible for these three traits. Sequence and expression analyses revealed distinct variations in *GmSWEET39* between HPCD5 and Nannong Caidou 5, and the expression level of this gene in the seeds of Nannong Caidou 5 at 30 days after flowering was significantly higher than that in HPCD5, demonstrating that *GmSWEET39* is a key gene controlling seed protein content and other related traits in HPCD5. Using HPCD5 as the parent, hybrid combinations were developed with soybean varieties including Nannong 99-6, Wude, Xudou 14, and Zhonghuang 39, and an RIL population consisting of 334 families was established. Combined analysis of two-year phenotypic evaluation and genotypic detection of the RIL population identified multiple SNP markers associated with protein content, oil content, and 100-seed weight in the vicinity of *GmSWEET39*. Among these markers, the intragenic SNP marker TZX-94 explained 26.9%, 29.9%, and 7.8% of the phenotypic variation for the three traits, respectively, confirming the stable selective effect of the *GmSWEET39* haplotype from HPCD5 on soybean protein content, oil content, and 100-seed weight. Additionally, two InDel markers for *GmSWEET39* were developed, enriching the molecular marker types of this gene. This study identified the major genetic locus derived from wild soybean in HPCD5 that controls seed protein content and other traits, expanded the allelic variation resources of *GmSWEET39*, verified its breeding application potential under multiple genetic backgrounds, and created several new high-protein soybean germplasms. These results provide important genetic resources and technical support for the molecular breeding of high-protein soybean varieties.

Keywords: soybean; high protein content; *GmSWEET39*; marker-assisted selection; breeding application

收稿日期: 2026-02-01

基金项目: 国家重点研发计划 (2024YFD1201400); 国家自然科学基金 (31701452); 盐城市科技项目 (YCBK2024055, YCBK2025070)。

第一作者: 张英虎, 男, 博士, 研究员, 主要从事大豆遗传育种研究。E-mail: zyhnjau008@163.com。

大豆是我国重要的油脂和蛋白质来源,其产量和品质对国民经济发展和人民生活至关重要。栽培大豆一般含有 40% 的蛋白质和 20% 的油脂含量,提高籽粒蛋白质含量和油脂含量是大豆品质育种的主要目标。大豆籽粒蛋白质和油脂含量是受遗传和环境协同控制的数量性状,国内外研究者利用连锁分析、关联分析等方法,检测到多个控制大豆籽粒蛋白质和油脂含量的 QTL/基因^[1-4],这些数量性状位点(QTL)广泛分布于大豆所有染色体上,然而多数位点的遗传效应值较低,致使其在大豆品质改良育种实践中应用较为有限。特别是野生大豆中的优良等位变异在栽培大豆中的利用则更少。

多个研究者在大豆 15 号染色体 4.0 Mb 区域范围内检测到蛋白质和油脂的 QTL^[5-7],通过分子生物学和群体遗传学等手段表明该位点的控制基因为 *GmSWEET39*(*GmSWEET10a*, Glyma. 15G049200),其功能区的 CC(DT15-3875101)插入或缺失是影响 *GmSWEET39* 基因功能的关键变异,该变异在大豆驯化过程中受到选择,95% 以上野生大豆中为含有 CC 片段的高蛋白等位变异类型,而 80% 以上栽培大豆中为 CC 片段缺少的高油脂等位变异类型^[8-10],但能否利用野生大豆中 *GmSWEET39* 高蛋白等位变异来提高栽培大豆中籽粒蛋白质含量的研究鲜有报道。

南农菜豆 5 号是南京农业大学选育的夏大豆品种,该品种籽粒大、蛋白质含量高^[11],利用野生大豆 ZYD4219 为供体亲本,南农菜豆 5 号为轮回亲本,通过回交 3 代自交 7 代(BC₃F₇)创制出高蛋白大豆新种质 HPCD5,其籽粒蛋白质含量为 50.9%,显著高于论文亲本南农菜豆 5 号的 46.2%。2008—2018 年黄淮海南部育成品种蛋白质为 41.77%^[12],新创制的 HPCD5 蛋白质含量显著高于黄淮海南部育成品种的平均值,表明该种质在蛋白质的利用上具有较高的应用价值。前期通过分析 HPCD5 与徐豆 14、中黄 39 杂交组合 F_{2:3} 和 F_{2:4} 分离世代家系蛋白质含量,发现利用 HPCD5 可以显著提高了栽培大豆籽粒蛋白质含量^[13],但 HPCD5 中来自野生大豆控制籽粒蛋白质含量的关键位点及其育种利用效率尚未明确。本研究以南农菜豆 5 号和 HPCD5 杂交后代为材料,鉴定 HPCD5 中来自野生大豆的控制籽粒蛋白质等性状的遗传位点,以 HPCD5 与南农 99-6、武德、徐豆 14 和中黄 39 的杂交组合后代重组自交系(Recombinant Inbred Lines, RIL)群体为材料,旨在明确鉴定位点对大豆蛋白质含量等性状的

选择效果,为大豆高蛋白育种提供理论依据和基因资源。

1 材料与方法

1.1 试验材料

南农菜豆 5 号,南农菜豆 5 号高蛋白回交系(HPCD5),南农菜豆 5 号和 HPCD5 杂交产生的 F₂ 群体(包含 209 份家系),HPCD5 与南农 99-6、武德、徐豆 14 和中黄 39 的杂交组合后代重组自交系群体(分别包含 100, 70, 99 和 65 份家系),其中南农菜豆 5 号、HPCD5、南农 99-6、武德、徐豆 14 和中黄 39 等亲本材料均由南京农业大学国家大豆改良中心提供。

1.2 试验设计

试验材料种植在江苏沿海地区农业科学研究所南洋试验场(33°25'N, 120°11'E),大豆生长时间为 6 月至 10 月,播种前结合旋耕施入 300 kg·hm⁻² 复合肥(N:P:K = 15:15:15)作基肥,开花初期施 750 kg·hm⁻² 尿素作追肥。

检测 HPCD5 中籽粒蛋白质含量等性状关键位点。利用 HPCD5 和南农菜豆 5 号杂交产生的分离后代检测 HPCD5 中籽粒蛋白质含量等性状关键位点,2017 年种植 F₂ 单株并测定基因型,2018 年种植 F_{2:3} 株行,行长 2 m,行距 0.5 m,株行收获。

分析 HPCD5 和南农菜豆 5 号籽粒蛋白质含量等候选基因序列和表达情况。测定 HPCD5 和南农菜豆 5 号籽粒蛋白质含量等候选基因序列,发掘基因序列差异,测定 HPCD5 和南农菜豆 5 号籽粒不同发育阶段关键基因的表达情况,发掘基因表达差异。

利用 *GmSWEET39* 基因提高栽培大豆蛋白质含量的育种应用研究。2021 年和 2023 年种植 HPCD5 与南农 99-6、武德、徐豆 14 和中黄 39 的杂交组合后代重组自交系群体,随机区组设计,行长 2 m,行距 0.5 m,3 次重复,株行收获。

1.3 表型测定

大豆收获后,随机选择 100 粒种子测定其百粒重,利用 FOSS 近红外 Infratec™ 1241 测定大豆籽粒蛋白质含量和油脂含量。

1.4 重测序与基因分型

利用 CTAB 法提取全部试验材料的叶片基因组 DNA,利用 Illumina HiSeq 高通量测序仪对南农菜豆 5 号和 HPCD5 进行重测序,平均每个样品测序数据量约为 30 倍基因组覆盖深度,覆盖度在 95% 以上,对南农菜豆 5 号和 HPCD5 杂交后代 F₂ 进行简化基

因组测序,因部分样品 DNA 建库失败,共获得 178 份家系基因型,平均每个样品测序数据量约为 0.5 倍基因组覆盖深度,覆盖度为 10% 左右,使用 GATK4.0 进行 SNP 和 InDel 分析,由上海凌恩生物科技有限公司进行检测。利用基于靶向测序的高通量基因型检测液相芯片“大豆 40K-GBTS”鉴定 HPCD5 与南农 99-6、武德、徐豆 14 和中黄 39 的杂交组合后代重组自交系群体家系的基因型,由河北石家庄博瑞迪生物技术有限公司进行检测^[14]。

1.5 转录组分析

使用北京天根生化科技有限公司 RNAsimple 总 RNA 提取试剂盒提取 HPCD5 和南农菜豆 5 号开花后 20, 30, 40 和 50 d 种子总 RNA, 利用以大豆 mRNA 为模板,以随机寡核苷酸为引物,用逆转录酶合成第一条 cDNA 链,然后用 DNA 聚合酶 I 以 dNTPs 为原料合成第二条 cDNA 链,使用核酸纯化试剂盒(AMPure XP beads)筛选片段大小,PCR 富集测序样本,纯化 PCR 产物,制备测序文库。用 Illumina HiSeq 2500 高通量测序平台进行转录组测序。利用 perl 软件计算 *GmSWEET39* 基因不同时期的 FPKM 值,由上海凌恩生物科技有限公司进行检测。

1.6 基因扩增和 InDel 标记开发

从 Phytozome13 网站中获取 *GmSWEET39* 基因的序列,设计扩增引物 F1-R1 和 F2-R2(表 1)。以 HPCD5 和南农菜豆 5 号基因组 DNA 进行 PCR 扩增,试剂盒进行胶回收,并进行测序,利用 Primer

Premier 5.0 软件设计 *GmSWEET39* 基因内 InDel 标记,由上海凌恩生物科技有限公司进行检测。

表 1 *GmSWEET39* 基因引物序列
Table 1 Primer sequence of *GmSWEET39*

测序区域 Sequencing region	引物名称 Primers	引物序列(5'-3') Primer sequence(5'-3')
F1-R1	F1	GTATGATCTGGCGGCCGTAA
	R1	AACCTCTCACCTTGTGTTGCAGGC
F2-R2	F2	GCACCATCTGAACTAGCCCC
	R2	CCTATACGCTATTTCGCGGCC

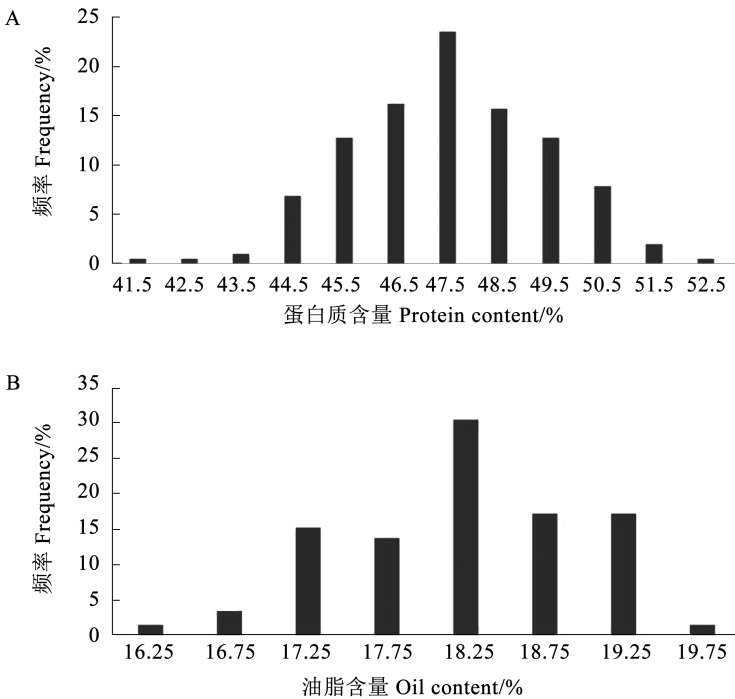
1.7 数据分析

利用 Excel 2021 进行统计分析,使用 MstMap 软件(<http://mstmap.org/>)构建南农菜豆 5 号和 HPCD5 杂交产生的 F₂ 群体的图谱,利用 TASSEL5^[15]中 GLM(General Linear Model)进行标记和性状关联分析。

2 结果与分析

2.1 高蛋白大豆种质 HPCD5 籽粒性状位点鉴定

2.1.1 HPCD5 和南农菜豆 5 号组合后代籽粒蛋白质、油脂和百粒重表型分布 测定 HPCD5 和南农菜豆 5 号杂交 F_{2:3} 代家系蛋白质、油脂和百粒重表型,因部分材料种子量不足,仅获得 204 份家系的蛋白质和油脂表型,杂交组合后代、蛋白质、油脂和百粒重平均值分别为 47.5%、18.2% 和 23.6 g,变幅分别为 41.6%~52.2%、16.2%~19.8% 和 17.2~28.7 g(图 1)。



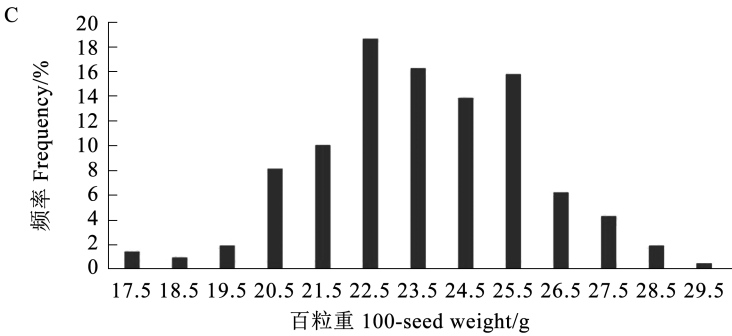


图1 HPCD5 和南农菜豆 5 号杂交后代蛋白质 (A)、油脂 (B) 和百粒重 (C) 表型分布
Fig.1 Phenotypic distribution of protein (A) , oil (B) and 100-seed weight (C) in hybrid offspring of HPCD5 and Nannongcaidou 5

2.1.2 亲本品种 HPCD5 和南农菜豆 5 号间序列差异分析 利用重测序技术对南农菜豆 5 号和 HPCD5 进行 DNA 序列测定,分别检测到 3 432 093 318和31 253 450 477 bp 序列(约 30 ×),将测定的序列与参考基因组 Williams 82 比对,分别检测到 2 593 812 和 2 659 154 个 SNP,378 103 和 393 030 个 InDel,比较南农菜豆 5 号和 HPCD5 两材料序列间的差异,检测到 2 895 个 SNP 和 361 个 InDel,这些位点主要分布在全基因组 11 个区间范围内(表 2)。

表 2 HPCD5 和南农菜豆 5 号差异序列区间
Table 2 Differential sequence intervals between HPCD5 and Nannongcaidou 5

染色体 Chromosome	基因数 Gene number	开始位置 Start position/bp	结束位置 Final position/bp	区间长度 Interval length/bp
Glyma. 02	69	42742465	44037854	1295389
Glyma. 04	193	48793077	52384837	3591760
Glyma. 06	273	3565056	13355060	9790004
Glyma. 08	24	2780328	10973479	8193151
Glyma. 09	203	293205	3515655	3222450
Glyma. 10	190	44413074	51532858	7119784
Glyma. 12	30	13372	511003	497631
Glyma. 14	20	6123866	7074900	951034
Glyma. 15	225	2402697	5747315	3344618
Glyma. 17	24	11034761	13442292	2407531
Glyma. 19	84	4309	1570574	1566265
总计 Total	1335			41979617

2.1.3 HPCD5 中控制大豆籽粒性状的关键位点检测 利用简化基因组测序技术测定 HPCD5 和南农菜豆 5 号杂交后代 F₂家系基因型,成功根据 178 份家系鉴定出 1 586 个有效 SNP 标记,联合基因型和表型数据,在 15 号染色体 2.4 ~5.7 Mb 鉴定到 1 个同时控制大豆蛋白质、油脂和百粒重的位点(图 2),表型解释率分别为 21.5%、16.7%和 17.7%。该位点是 HPCD5 中基于重测序检测的 11 个片段之一,推测该位点是 HPCD5 中控制大豆籽粒性状的关键位点。

2.2 高蛋白大豆种质 HPCD5 籽粒性状候选基因分析

前人研究表明在大豆 15 号染色体 2.4 ~5.7 Mb 区间内检测到大豆百粒重、蛋白质和油脂含量的 QTL,同时 *GmSWEET39*(Glyma. 15G049200) 为其控制基因^[8-10]。为明确该基因是否为 HPCD5 中控制大豆百粒重、蛋白质和油脂含量的基因,对两各材料间该基因进行序列分析,发现南农菜豆 5 号和 HPCD5 在 *GmSWEET39* 基因启动子、编码区和下游区域存在多个位点差异,其中包含 20 个 InDel 差异

和 52 个 SNP 差异。4 个 InDel 差异如图 3 所示,其中 Chr15_3875101 位于编码区,其序列变化会导致氨基酸变化,被认为是该基因的主要功能区域,Chr15_3876981 位于启动子区域,可能调控该基因的表达情况。HPCD5 中 *GmSWEET39* 基因在 Chr15_3875101 序列与野生大豆基因型一致^[7],推测

GmSWEET39 是 HPCD5 中控制大豆百粒重、蛋白质和油脂含量的基因。本研究在内含子和启动子上游区域检测到两个序列差异较大的位点 Chr15_3875550 和 Chr15_3884603,其差异序列长度分别为 74 和 27 bp,这两个位点是否有功能需要进一步研究,但该位点可以用于设计 InDel 标记。

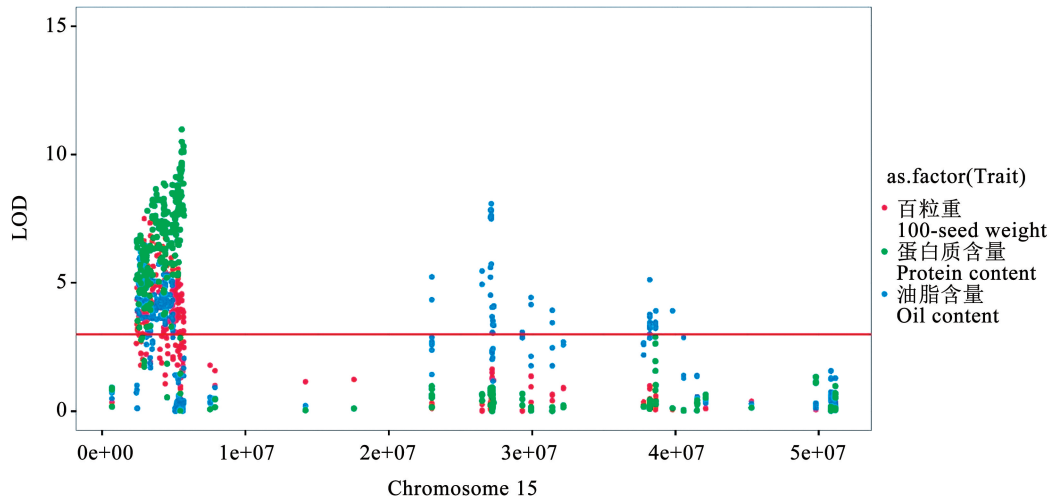


图2 HPCD5 和南农菜豆 5 号杂交 F₂ 蛋白质、油脂和百粒重定位结果

Fig. 2 Mapping results of protein, oil and 100-seed weight of F₂ for the cross of HPCD5 and Nannongcaidou 5

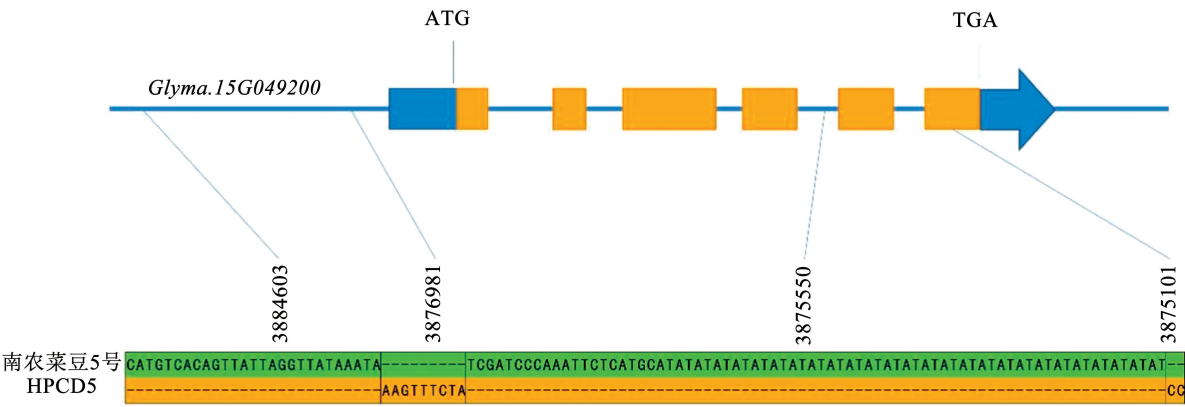


图3 HPCD5 和南农菜豆 5 号中 *GmSWEET39* 基因序列差异分析

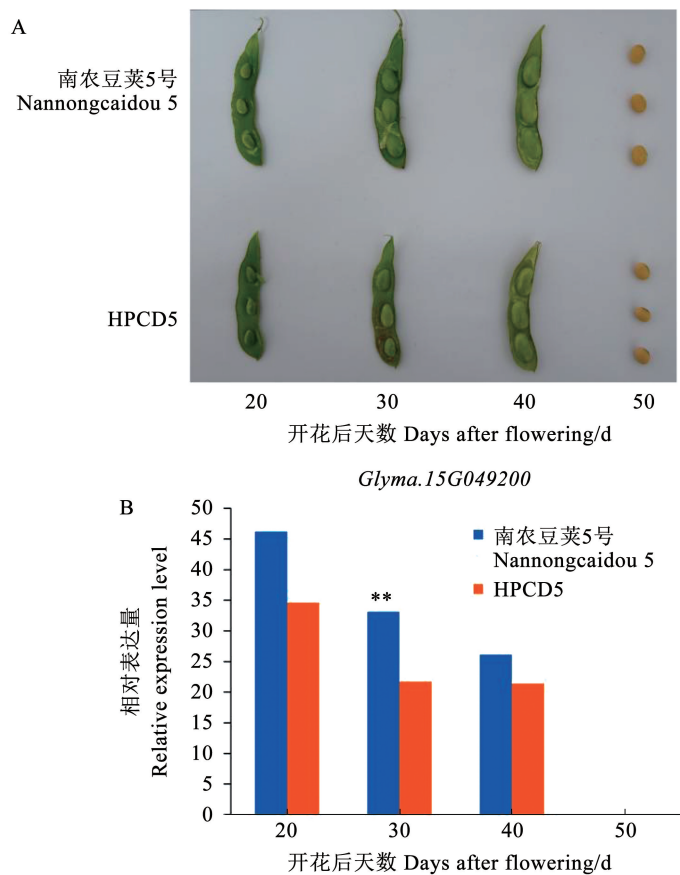
Fig. 3 Analysis of *GmSWEET39* gene sequence differences between HPCD5 and Nannongcaidou 5

利用转录组测序数据分析 *GmSWEET39* 在籽粒不同发育时期基因表达,发现该基因随着籽粒的成熟表达量呈逐渐降低,开花后 50 d 几乎不表达,在南农菜豆 5 号中的表达量高于 HPCD5,在开花后 30 d 时差异显著,表明 *GmSWEET39* 是 HPCD5 百粒重、蛋白质和油脂含量较亲本变化的关键基因(图 4)。

2.3 高蛋白大豆种质 HPCD5 的育种应用

配置 HPCD5 和南农 99-6、武德、徐豆 14 和中黄 39 的杂交组合,经过单粒传法,分别形成了 100,70,99 和 65 份家系,共包含了 334 份杂交后代,对杂交

后代进行籽粒蛋白质、油脂和百粒重的测定,同时利用芯片技术检测了 15 号染色体上的分子标记,结果表明 *GmSWEET39* 基因附近多个 SNP 与蛋白、油脂和百粒重关联,两年平均的表型解释率分别为 26.9%、29.9% 和 7.8%,表明该位点对蛋白和油脂含量的选择效率较高(图 5)。对 4 个杂交组合后代进行分析,结果表明 *GmSWEET39* 基因中 HPCD5 类型家系在 4 个杂交组合后代中均表现出高蛋白、低油脂和低百粒重,表明该位点在不同遗传背景下均具有选择效果(表 3)。

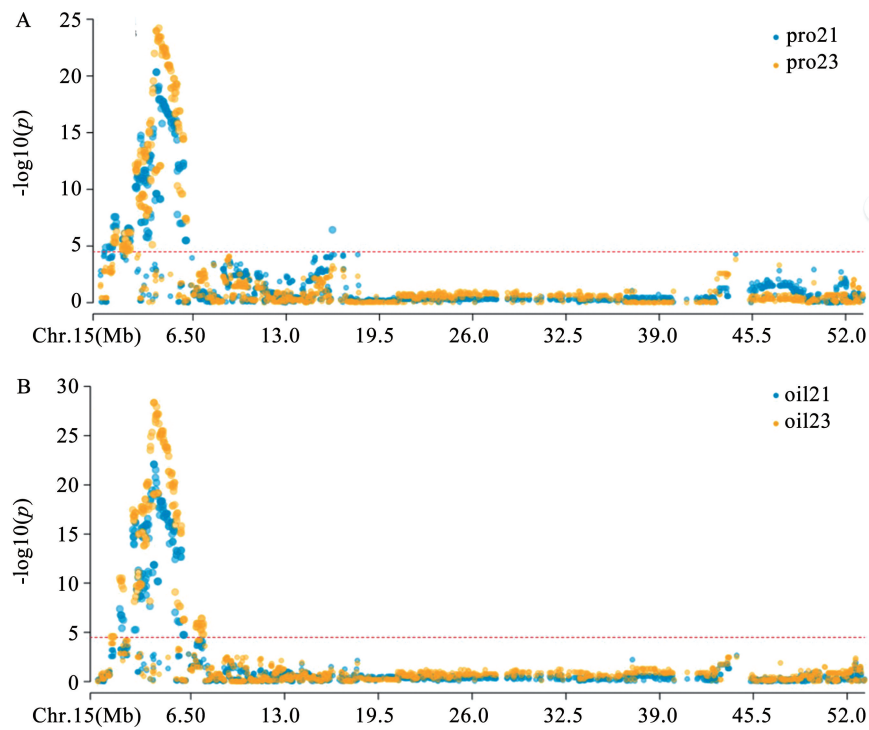


注:A. 南农菜豆 5 号和 HPCD5 在开花后不同时期的籽粒形态; B. 南农菜豆 5 号和 HPCD5 在开花后不同时期籽粒 *GmSWEET39* 基因表达量直方图。DAF. 开花后天数。* * 表示 $P < 0.01$ 水平存在显著差异。

Note: A. The seed morphology of Nannongcaidou 5 and HPCD5 at different stages after flowering; B. The histogram of *GmSWEET39* gene expression levels of Nannongcaidou 5 and HPCD5 at different stages after seed flowering. DAF. Days after flowering. * * means significant difference at $P < 0.01$ level.

图 4 Nannongcaidou 5 和 HPCD5 在开花后不同时期籽粒 *GmSWEET39* 基因表达情况

Fig. 4 Expression of *GmSWEET39* gene in seeds of Nannongcaidou 5 and HPCD5 at different stages after flowering



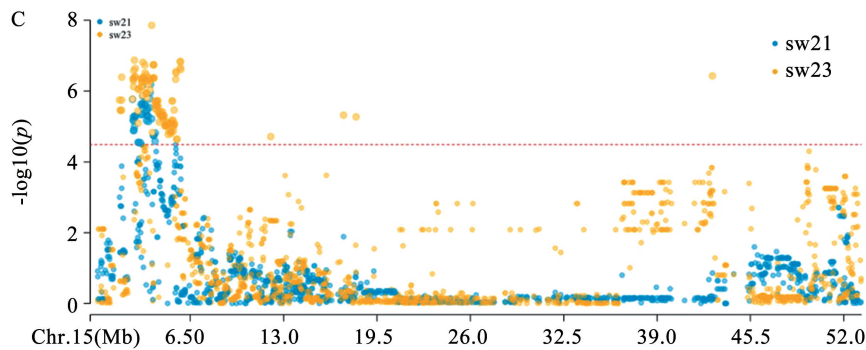


图5 *GmSWEET39* 基因附近 SNP 对杂交后代家系中蛋白质 (A)、
油脂 (B) 和百粒重 (C) 的选择效果图

Fig.5 Selection effect of SNPs near *GmSWEET39* gene on protein (A),
oil (B), and 100 seed weight (C) in hybrid offspring families

表3 杂交组合后代家系在 *GmSWEET39* 基因内 SNP (TZX-94) 不同等位变异下蛋白质、油脂和百粒重的表型值

Table 3 Phenotypic values of protein, lipid, and 100-seed weight under different allelic variations
of SNP (TZX-94) in *GmSWEET39* gene in hybrid offspring families

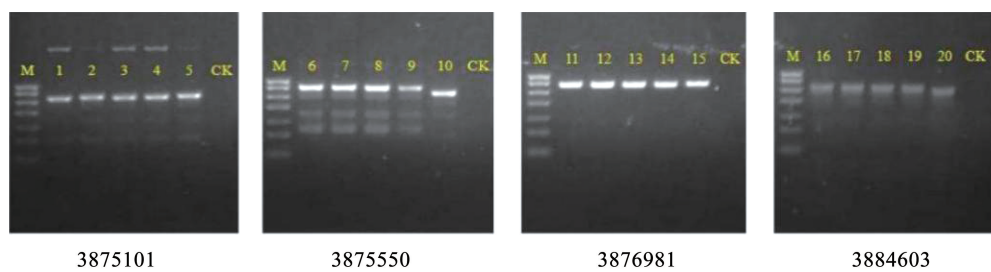
组合 Cross	类型 Type	家系数 Offspring number	蛋白含量 Protein content/%		油脂含量 Oil content/%		百粒重 100-seed weight/g	
			2021	2023	2021	2023	2021	2023
HPCD5 × 南农 99-6	HPCD5	38	48.8	48.0	18.1	18.4	21.3	21.5
HPCD5 × Nannong 99-6	南农 99-6	57	45.8	45.3	19.6	19.8	23.6	23.9
武德 × HPCD5	HPCD5	27	49.4	48.6	18.4	18.9	22.6	22.6
Wude × HPCD5	武德	40	46.9	45.6	19.6	20.2	25.3	25.0
徐豆 14 × HPCD5	HPCD5	46	46.7	47.0	18.8	18.9	21.0	23.5
Xudou 14 × HPCD5	徐豆 14	50	45.3	45.0	19.5	20.0	22.3	24.5
中黄 39 × HPCD5	HPCD5	28	47.9	47.4	18.7	18.9	22.6	24.4
Zhonghuang 39 × HPCD5	中黄 39	33	45.3	44.9	19.8	20.3	23.4	25.3

为拓展 *GmSWEET39* 利用范围,对 HPCD5 和南农菜豆 5 号中在 *GmSWEET39* 基因内和附近位置的 InDel 设计分子标记(表 4),利用分子标记检测中黄 39、南农 99-6 和徐豆 14 等品种在 *GmSWEET39* 基因上的等位变异类型,结果表明 3 875 550 和 3 884 603 位置可以通过琼脂糖电泳进行区分(图 6)。

表4 *GmSWEET39* 基因差异位点开发的 InDel 标记

Table 4 InDel markers for*GmSWEET39* gene differential sites

测序区域 Sequencing region	引物名称 Primers	引物序列 Primer sequence(5'-3')
3875101	3875101-F	GTATGATCTGGCGGCCGTAA
	3875101-R	AACTCTCACCTTGTTGCAGGC
3875550	3875550-F	GCACCATCTGAAGTAGCCCC
	3875550-R	CCTATACGTATTTCGGGGCC
3876981	3876981-F	TGGGGGATGTGATGAGCAAC
	3876981-R	CTGCTTCCGAGGGGATGTTT
3884603	3884603-F	TCTCTGACAAGTGACTTGACA
	3884603-R	AAAGGAGAGTGGATATTGCCT



注:从左到右材料分别为中黄 39、南农 99-6、徐豆 14、南农菜豆 5 号和 HPCD5。
Note: The materials from left to rightwere Zhonghuang 39, Nannong 99-6, Xudou 14, Nannongcaidou 5, and HPCD5, respectively.

图 6 InDel 标记的电泳图
Fig. 6 Electrophoresis image of InDel markers

3 讨论

选育高蛋白大豆品种是大豆品质育种的重要目标之一,野生大豆是提高育成品种蛋白质含量的优异基因源^[16-17]。本研究利用野生大豆 ZYD4219 为供体亲本,南农菜豆 5 号为轮回亲本,成功创造出蛋白质含量显著高于南农菜豆 5 号的高蛋白新种质 HPCD5,联合定位分析、基因序列分析和表达分析,明确位于 15 号染色体的 *GmSWEET39* (Glyma.15G049200)基因为来自野生大豆的控制蛋白质含量的关键基因。前人研究表明,*GmSWEET39* 基因受到选择,高蛋白等位变异类型主要存在在野生大豆中,在现代栽培品种中频率很低^[10],本研究分析了中黄 39、徐豆 14 和南农 99-6 等现代育成品种,发现育成品种中均不含有 *GmSWEET39* 高蛋白等位变异类型,表明利用野生大豆中含有的 *GmSWEET39* 高蛋白等位变异可以提高栽培大豆中籽粒蛋白质含量。

利用 *GmSWEET39* 提高育成品种蛋白质含量的同时,显著降低了油脂和百粒重,表明 *GmSWEET39* 基因具有多效性。从野生大豆到栽培大豆,大豆经历了油脂和百粒重增加,蛋白质下降的过程,当前也检测到多个同时增加大豆油脂和百粒重的基因^[18]。但在育成品种中,多个研究表明大豆油脂和百粒重呈负相关或相关性不显著,而蛋白质含量和百粒重呈正相关^[13, 19-20],因此在利用 *GmSWEET39* 高蛋白等位变异时需要协调其他高百粒重位点,本研究利用的高蛋白新种质 HPCD5 在具备高蛋白优异性状的同时具备百粒重不低的特点,在育种利用的 4 个组合中均创制出 *GmSWEET39* 基因为高蛋白类型,籽粒蛋白质高于 48%,百粒重高于 25 g 的大豆新种质。下一步拟检测 HPCD5 中提高百粒重的

位点,提高新种质 HPCD5 的利用效率。
液相芯片是一种基于二代测序新开发的基因分型技术^[14],在植物分子标记辅助选择方面已有多个报道^[21-22],本研究利用大豆上开发的液相芯片^[23],检测了 *GmSWEET39* 所在染色体的分子标记,发现 *GmSWEET39* 基因附近的 SNP 标记对蛋白质等性状均有选择效果,特别是该基因内分子标记 TZX-94 对大豆蛋白质等性状选择效果显著,表明利用液相芯片技术可以用来进行分子标记辅助选择。InDel 标记具有低成本和易于检测等优点,其对检测的设备要求低,为更好地对 *GmSWEET39* 基因进行选择,对该基因两个大片段 indel 开发了 InDel 标记。
综上所述,本研究发现了来自野生大豆的 *GmSWEET39* 基因新的等位变异,发现该基因内 SNP 标记在实际育种中对大豆蛋白、油脂和百粒重有较强的选择效果,创制出多个高蛋白新种质。

4 结论

本研究利用连锁分析对创制的高蛋白新种质 HPCD5 中控制籽粒蛋白质含量等性状的遗传位点进行检测,在 15 号染色体 2.4 ~ 5.7 Mb 鉴定到 1 个同时控制大豆蛋白质、油脂和百粒重的位点。结合基因组序列分析、基因表达等分析,明确该区间内 *GmSWEET39* 基因为控制大豆蛋白质、油脂和百粒重的关键基因。配组 HPCD5 和育成品种的杂交组合,发现 HPCD5 可以作为高蛋白新种质进行大豆高蛋白育种,同时来自野生大豆的 *GmSWEET39* 高蛋白等位变异用来提高栽培大豆蛋白质含量的选择效率较高。本研究创制的高蛋白新种质和发掘的 *GmSWEET39* 高蛋白等位变异为大豆高蛋白育种提高了种质和基因资源,同时为大豆高蛋白分子育种提供了实例。

参考文献

[1] Tian H, Yin Y, Li X, et al. Identification of *HSSPI* as a regulator of soybean protein content through QTL analysis and Soy-SPCC network[J]. *Plant Biotechnology Journal*, 2025, 23(7): 2673-2688.

[2] Zhang Y, He J, Meng S, et al. Identifying QTL-allele system of seed protein content in Chinese soybean landraces for population differentiation studies and optimal cross predictions [J]. *Euphytica*, 2018, 214(9): 157.

[3] Zhang Y, He J, Wang H, et al. Detecting the QTL-allele system of seed oil traits using multi-locus genome-wide association analysis for population characterization and optimal cross prediction in soybean[J]. *Frontiers in Plant Science*, 2018, 9: 1793.

[4] Hu Y, Liu Y, Wei J J, et al. Regulation of seed traits in soybean [J]. *aBIOTECH*, 2023, 4(4): 372-385.

[5] Kim M, Schultz S, Nelson R L, et al. Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15[J]. *Crop Science*, 2016, 56(1): 219-225.

[6] Patil G, Mian R, Vuong T, et al. Molecular mapping and genomics of soybean seed protein: A review and perspective for the future[J]. *Theoretical and Applied Genetics*, 2017, 130(10): 1975-1991.

[7] Yang H, Wang W, He Q, et al. Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean [J]. *Theoretical and Applied Genetics*, 2019, 132(10): 2793-2807.

[8] Miao L, Yang S, Zhang K, et al. Natural variation and selection in *GmSWEET39* affect soybean seed oil content [J]. *New Phytologist*, 2020, 225(4): 1651-1666.

[9] Wang S, Liu S, Wang J, et al. Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication[J]. *National Science Review*, 2020, 7(11): 1776-1786.

[10] Zhang H, Goettel W, Song Q, et al. Selection of *GmSWEET39* for oil and protein improvement in soybean[J]. *PLoS Genetics*, 2020, 16(11): e1009114.

[11] 王玉红, 赵正金, 凌良振, 等. 南农菜豆5号特征特性及栽培技术[J]. *江苏农业科学*, 2003, 31(5): 42-42, 101.

Wang Y H, Zhao Z J, Ling L Z, et al. Characteristics and cultivation techniques of Nannong Caidou 5 [J]. *Jiangsu Agricultural Sciences*, 2003, 31(5): 42-42, 101.

[12] 王大刚, 陈圣男, 李杰坤, 等. 2008—2018年黄淮南部审定大豆品种主要性状分析[J]. *大豆科学*, 2019, 38(5): 671-680.

Wang D G, Chen S N, Li J K, et al. Analysis on the main traits of soybean varieties approved in southern Huang-Huai during 2008-2018[J]. *Soybean Science*, 2019, 38(5): 671-680.

[13] 杨红燕, 沈会权, 胡哲, 等. 黄淮海地区大豆品种蛋白质、油脂含量和百粒质量的表型分析及籽粒蛋白质含量的改良[J]. *江苏农业科学*, 2021, 49(22): 84-87.

Yang H Y, Shen H Q, Hu Z, et al. Phenotypic analysis of protein, oil content and 100-grain weight of soybean varieties in Huanghuaihai area and improvement of grain protein content[J]. *Jiangsu Agricultural Sciences*, 2021, 49(22): 84-87.

[14] 徐云碧, 杨泉友, 郑洪建, 等. 靶向测序基因型检测(GBTS)技术及其应用[J]. *中国农业科学*, 2020, 53(15): 2983-3004.

Xu Y B, Yang Q N, Zheng H J, et al. Genotyping by target sequencing (GBTS) and its applications[J]. *Scientia Agricultura Sinica*, 2020, 53(15): 2983-3004.

[15] Bradbury P J, Zhang Z, Kroon D E, et al. TASSEL: Software for association mapping of complex traits in diverse samples [J]. *Bioinformatics*, 2007, 23(19): 2633-2635.

[16] 刘森, 毕影东, 刘建新, 等. 黑龙江省不同积温区野生大豆品质分析与优异资源筛选[J]. *大豆科学*, 2025, 44(3): 79-85.

Liu M, Bi Y D, Liu J X, et al. Quality analysis and superior resource screening of wild soybean resources in different accumulated temperature zones of Heilongjiang Province [J]. *Soybean Science*, 2025, 44(3): 79-85.

[17] 高倩, 冯燕, 杨雅华, 等. 基于全基因组关联分析挖掘野生大豆蛋白含量QTL[J]. *华北农学报*, 2021, 36(S1): 23-30.

Gao Q, Feng Y, Yang Y H, et al. Detecting QTL underlying wild soybean protein content through genome wide association study [J]. *Acta Agriculturae Boreali-Sinica*, 2021, 36(S1): 23-30.

[18] Duan Z, Li Q, Wang H, et al. Genetic regulatory networks of soybean seed size, oil and protein contents[J]. *Frontiers in Plant Science*, 2023, 14: 1160418.

[19] 王大刚, 胡晨, 胡国玉, 等. 大豆品系在黄淮南部主产区产量及品质性状分析[J]. *中国农学通报*, 2014, 30(18): 104-111.

Wang D G, Hu C, Hu G Y, et al. Analysis on yield and quality traits of soybean lines in planting region of southern Huanghuai [J]. *Chinese Agricultural Science Bulletin*, 2014, 30(18): 104-111.

[20] 徐瑶, 冷苏凤, 张玉明, 等. 1982—2021年江苏省审定大豆品种主要农艺性状、产量、品质及抗性演变分析[J]. *中国油料作物学报*, 2022, 44(4): 780-789.

Xu Y, Leng S F, Zhang Y M, et al. Evolution analysis of main agronomic traits, yield, quality and resistance of soybean varieties released in Jiangsu Province from 1982 to 2021 [J]. *Chinese Journal of Oil Crop Sciences*, 2022, 44(4): 780-789.

[21] 薛振宇, 张康康, 张元元, 等. 优质抗旱小麦种质的筛选及功能基因检测[J]. *中国农业科技导报*, 2025, 27(1): 35-49.

Xue Z Y, Zhang K K, Zhang Y Y, et al. Screening and functional gene detection of high-quality and drought-resistant wheat germplasms[J]. *Journal of Agricultural Science and Technology*, 2025, 27(1): 35-49.

[22] 裴法敬, 张文轩, 张晓, 等. 长粒香型的超短生育期水稻新品系创制[J]. *作物学报*, 2024, 50(7): 1684-1698.

Pei F J, Zhang W X, Zhang X, et al. Developing new rice lines with ultrashort-duration, long-grain, and fragrance [J]. *Acta Agronomica Sinica*, 2024, 50(7): 1684-1698.

[23] Liu Y, Liu S, Zhang Z, et al. GenoBaits Soy40K: A highly flexible and low-cost SNP array for soybean studies[J]. *Science China Life Sciences*, 2022, 65(9): 1898-1901.