

大豆蛋白含量新位点 *qPRO-19-1* 的定位

武阳春^{1,2}, 郭兵福³, 谷勇哲², 栾晓燕⁴, 邱红梅⁵, 刘鑫磊⁴, 李海燕¹, 邱丽娟^{1,2}

(¹ 吉林农业大学生命科学学院, 长春 130000; ² 中国农业科学院作物科学研究所 / 农作物基因资源与遗传改良国家重大科学工程 /

农业部北京大豆生物学重点实验室, 北京 100081; ³ 江西省农业科学院作物研究所, 南昌 330200;

⁴ 黑龙江省农业科学院大豆研究所, 哈尔滨 150086; ⁵ 吉林省农业科学院大豆研究所, 长春 130033)

摘要: 大豆是重要的粮食作物和经济作物, 其籽粒蛋白约为 40%, 是植物蛋白的重要来源之一。国产大豆主要用于食用, 提高大豆蛋白含量是主要的育种目标。因此, 发掘大豆蛋白含量相关基因, 对开发分子标记并培育高蛋白食用大豆具有重要意义。本研究以低蛋白大豆品种中黄 35 为母本, 以源自日本的高蛋白大豆十胜长叶为父本, 构建了重组自交系 (RIL, recombination inbred lines) 群体。利用集群分离分析法 (BSA, bulked segregant analysis) 在 3 条染色体筛选出 9 个与蛋白含量相关的 SSR 标记, 其中位于 19 号染色体的 QTL 尚未见报道。进一步利用完备区间作图法 (ICIM-ADD) 分析 RIL 群体 $F_{2:15}$ 和 $F_{2:16}$, 在 19 号染色体重复定位了 1 个蛋白质含量相关 QTL *qPRO-19-1*, 位于分子标记 SSR_19_38 和 SSR_19_59 之间, LOD 值分别为 3.43 和 3.98, 贡献率分别为 7.81% 和 14.87%, 高蛋白等位基因来自于高蛋白亲本十胜长叶。*qPRO-19-1* 的定位区间长度为 385 kb, 共有注释基因 36 个。本研究定位了蛋白质含量相关的新位点 *qPRO-19-1*, 为大豆高蛋白基因的图位克隆及分子标记育种奠定了基础。

关键词: 大豆; 重组自交系; 蛋白质含量; QTL

Mapping of a New Quantitative Locus *qPRO-19-1* Associated with Seed Crude Protein Content in Soybean (*Glycine max* L.)

WU Yang-chun^{1,2}, GUO Bing-fu³, GU Yong-zhe², LUAN Xiao-yan⁴,

QIU Hong-mei⁵, LIU Xin-lei⁴, LI Hai-yan¹, QIU Li-juan^{1,2}

(¹ College of Life Sciences, Jilin Agriculture University, Changchun 130000; ² National Key Facility for Crop Gene Resources and Genetic Improvement/Key Laboratory of Soybean Biology in Beijing (MOA)/Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081; ³ Crops Research Institute of Jiangxi Academy of Agricultural Sciences, Nanchang 330200;

⁴ Soybean Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086;

⁵ Soybean Research Institute of Jilin Academy of Agricultural Sciences, Changchun 130033)

Abstract: Soybean is an economically-important food crop that contains about 40% of grain protein. The soybeans harvested from China are mainly used for food, and breeding for higher protein content is one of the main targets in soybean. Exploration of the genes related to soybean protein content and development of molecular markers are thus of great significance. In this study, a recombinant inbred line population was used by crossing a low protein soybean variety ‘ZH35’ (female parent) with the high protein accession ‘Tokachi nagaha’ (male parent) that was introduced from Japan. With the Bulk Segregant Analysis (BSA) analysis, nine SSR markers that associated to protein content were identified on three chromosomes, including chromosome 19 where no QTL was previously reported. This QTL *qPRO-19-1* was allocated using an inclusive complete interval mapping method (ICIM-ADD) to the genetic interval of the molecular markers SSR_19_38 and SSR_19_59. This elite allele, which was derived from ‘Tokachi Nagaha’, could explain 7.81% and 14.87% of phenotypic variations in $F_{2:15}$ and $F_{2:16}$ with the LOD values of 3.43 and 3.98, respectively. The physical interval of *qPRO-19-*

收稿日期: 2020-02-26 修回日期: 2020-04-22 网络出版日期: 2020-07-24

URL: <http://doi.org/10.13430/j.cnki.jpgr.20200226001>

第一作者研究方向为大豆优异基因挖掘, E-mail: 447901573@qq.com

通信作者: 邱丽娟, 研究方向为大豆基因资源挖掘与利用, E-mail: qiulijuan@caas.cn

基金项目: 国家自然科学基金 (31960408); 中国农业科学院科技创新工程

Foundation projects: National Natural Science Foundation of China (31960408), The Agricultural Science and Technology Innovation Program

I expanded 385 kb that harbored 36 annotated genes. Collectively, the genetic mapping of the new QTL *qPRO-19-1* laid the foundation for map-based cloning gene and molecular marker assisted breeding for high protein soybean varieties.

Key words: soybean; RIL; protein content; QTL

大豆是提供可食用蛋白质和油脂的主要来源之一^[1]。随着我国人民生活水平的提高,国内市场对高蛋白大豆的需求逐年提高。我国是大豆的原产国,具有丰富的种质资源,且大豆籽粒中蛋白质含量变异十分广泛,高蛋白种质的蛋白质含量可高达 50% 以上,低蛋白种质的蛋白质含量为 30% 左右^[2]。如何将蛋白质含量提高到 43% 甚至是 45% 以上是近年来我国大豆育种的主要目标。因此,挖掘大豆高蛋白含量相关基因,有助于认识其调控网络,对定向培育高蛋白大豆品种具有重要意义。

大豆蛋白含量是受多基因控制的数量性状,它既受遗传因素的影响,又受环境因素的影响^[3],因此,大豆蛋白表型选择难度很大。随着遗传标记的发展及功能基因组的发展,为构建遗传图谱并从多基因控制的性状中挖掘出基因创造了条件。自 Botstein 等^[4]首次将 DNA 限制性片段长度多态性(RFLP)应用在人类的连锁图谱上,人们相继利用 RAPD^[5]、AFLP^[6]、SSR^[7]等分子标记和不同群体(F_2 、RIL、BC)构建出许多高密度的遗传连锁图谱。从 Keim 等^[8]首次利用 RFLP 分子标记定位出与大豆种子硬实性状有关的 QTL 定位后,越来越多的学者相继定位出与大豆数量性状相关的 QTL。截止到 2020 年 1 月, Soybase 数据库(<https://www.soybase.org>)公布的与大豆蛋白质含量相关的 QTL 共 248 个,这些 QTL 分布在 20 条染色体上,其中 20 号染色体定位到的蛋白质含量 QTL 最多,共 25 个 QTL 位点,可解释的遗传贡献率范围为 0.07%~65%。其次是 6 号染色体和 9 号染色体,两条染色体上蛋白质含量 QTL 定位数量均为 17 个,可解释的遗传贡献率分别为 0.05%~27.6% 和 0.07%~24.4%。虽然定位的大豆蛋白质含量 QTL 很多,但能够稳定表达的 QTL 非常有限。例如, Liang 等^[9]利用 1 个由 474 个株系组成的 F_{13} 重组自交系群体,定位到 6 个与蛋白质含量相关的 QTL,分别在 B2(Chr.14)、C2(Chr.06)、G(Chr.18)、H(Chr.12)连锁群上,其中两年重复定位的仅有 C2(Chr.06)连锁群。Lu 等^[10]在 2 年 3 个地点对 $F_{2,9}$ 重组自交系群体蛋白质含量进行分析,定位了 10 个与蛋白质含量相关的 QTL,分布在 A1(Chr.05)、C2(Chr.06)、M(Chr.07)、A2(Chr.08)、K(Chr.09)、B1(Chr.11)、H(Chr.12)、G(Chr.18)、L(Chr.19)、I(Chr.20)连锁群,其中在相

同地点连续两年定位的 QTL 只有 A2(Chr.08)连锁群,且定位区间仅部分重叠,可解释的表型变异率为 4.5%~18.2%。Wang 等^[11]利用 2 个重组自交系,将蛋白质含量 QTL 定位在 D1a(Chr.01)、A1(Chr.05)、B2(Chr.14)、I(Chr.20)连锁群,其中多个地点都定位的是 I(Chr.20)连锁群,可解释的遗传变异率为 20%~42%。Wang 等^[12]用 SD024-59 $\times$ A02-3811002018 群体定位 11 个大豆蛋白质含量 QTL,用 SD02911 $\times$ SD00-1501 群体进行验证,有 5 个蛋白质含量 QTL 位点与定位群体相同,分布在 B1(Chr.11)、D1b(Chr.02)、D2(Chr.17)、K(Chr.09)连锁群,其中,仅 D2(Chr.17)连锁群的 2 个蛋白质含量 QTL 位点,与定位群体的蛋白质含量 QTL 位点完全重叠。Shan 等^[13]在 5 个不同地点对 154 个重组自交系对蛋白质含量进行 QTL 定位,用复合区间作图法定位到 20 个与蛋白质含量相关的 QTL,在 C2(Chr.06)连锁群上重复定位 3 次。用多重区间作图法定位到 9 个与蛋白质含量相关的 QTL,在 A1(Chr.05)和 C2(Chr.06)重复定位 2 次;利用两种方法均将蛋白质含量 QTL 定位在 C2(Chr.06)连锁群上。Qi 等^[14]在 2 年 3 个地点利用重组自交系对蛋白质含量进行 QTL 分析,除 M(Chr.07)、A2(Chr.08)、D2(Chr.17)3 个连锁群外,均定位到与蛋白质含量相关的 QTL。其中, *qPRO20-1* 位点重复定位 4 次,可解释表型变异范围为 4.4%~9.95%。*qPRO7-5* 位点重复定位 3 次,可解释的表型变异范围为 7.2%~14.5%。这些结果表明,大豆蛋白质含量 QTL 定位受环境及遗传因素影响较大,挖掘具有重演性、稳定性的大豆蛋白含量 QTL 可促进分子标记辅助大豆高蛋白育种。

本研究以低蛋白中黄 35 为母本,以我国育种中利用较多的高蛋白十胜长叶^[15]为父本,构建了重组自交系(Recombinant Inbred Line)群体,结合 SSR 标记基因型数据和群体间蛋白质含量的表型数据,采用 QTL IciMapping4.1 完备区间作图法(ICIM-ADD),定位到 1 个与蛋白质含量相关的 QTL 新位点,为大豆高蛋白基因克隆及分子标记育种奠定基础。

1 材料与方法

1.1 试验材料

本试验以低蛋白品种中黄 35 为母本,以日本引进高蛋白种质十胜长叶为父本,通过杂交获得 F_2 群

体。利用单粒传法在北京、海南 2 个地点进行加代繁殖,形成了一个稳定的由 199 个家系组成的 RIL 群体。2018 年和 2019 年将亲本及 RIL 群体 $F_{2:15}$ 、 $F_{2:16}$ 播种在中国农业科学院作物研究所北京顺义试验基地。亲本及 RIL 群体的株系分别种植 1 行,行长 3 m,行距 0.45 m,株距 0.1 m。四周均设保护行,田间管理同常规生产。并在成熟后每行选择间距相等、长势一致的 5 个单株进行收获。

1.2 蛋白质含量的测定

选取形态完整的大豆籽粒,利用德国 Bruker 公司生产的傅立叶变换近红外光谱仪测定蛋白质含量,每份样品测量 3 次;数据分析利用 OPUS 软件在建立好的大豆蛋白质干基模型中分析光谱信息,取 3 次的平均值为样品的最终蛋白质含量。每个株行测量 5 个单株的蛋白质含量,以 5 株平均值代表该株系的蛋白质含量。

1.3 极端混池的构建

根据 BSA 法^[16],结合中黄 35 和十胜长叶 $F_{2:15}$ 群体蛋白质含量及群体株系数(199 个株系),参考蛋白质含量频数分布直方图(图 1),以蛋白质含量高于 44% 和低于 40% 为界,筛选出占群体总数 5% 的单株数各 10 株构建低蛋白(LP)和高蛋白(HP)基因池(表 1)。

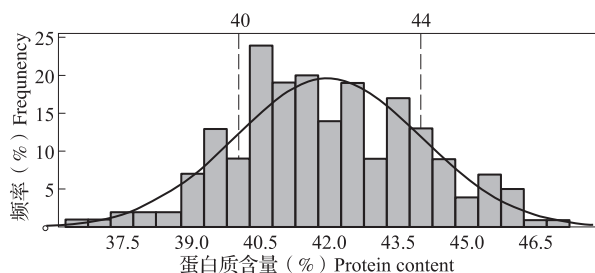


图 1 $F_{2:15}$ 群体蛋白质含量分布及构建基因池筛选标准

Fig.1 $F_{2:15}$ population protein content distribution and construction of gene pools screening criteria

表 1 构建低蛋白和高蛋白基因池各株系蛋白质含量

Table 1 Construct low protein and high protein gene pools for protein content of each strain

株系编号 Name	低蛋白 LP	株系编号 Name	高蛋白 HP
1	39.53	11	44.01
2	36.78	12	44.33
3	37.69	13	44.40
4	38.92	14	44.61
5	38.99	15	44.75
6	39.16	16	45.43
7	39.36	17	45.43
8	39.55	18	45.91
9	39.66	19	46.48
10	39.71	20	47.19
平均值 Mean	38.94	平均值 (Mean)	45.25

LP: Low protein, HP: High protein. The same as below

1.4 基因组 DNA 提取

以 CTAB 法制备大豆 DNA。取新鲜 50~100 mg 叶片置于 2.0 mL 装有钢珠的离心管中,液氮冷冻后放入打样机中打碎,加入 800 μ L CTAB,于 65 $^{\circ}$ C 水浴锅中水浴 1 h,每 10 min 摇匀 1 次;加入 800 μ L 酚氯仿,混匀,12000 rpm 离心 10 min;将上清转移至 1.5 mL 离心管中,加入 600 μ L 异丙醇,12000 rpm 离心 10 min;弃上清液,加入 100% 无水乙醇,12000 rpm 离心 10 min;弃 100% 无水乙醇,加入 75% 无水乙醇,12000 rpm 离心 10 min。弃 75% 无水乙醇后晾干。加入 200 μ L ddH₂O 溶解 DNA。用 1% 琼脂糖凝胶电泳检测 DNA 质量,并用紫外分光光度计检测 DNA 浓度,−20 $^{\circ}$ C 保存。

1.5 SSR 标记选择及基因型鉴定

利用本实验室保存的分布在大豆全基因组上的 872 对 SSR 引物^[17]在亲本间进行多态性筛选,筛选出具有条带清晰、多态性高的 SSR 引物。并构建高蛋白及低蛋白 2 个极端池,与亲本进行对比,筛选出连锁标记。

PCR 反应体系为 20 μ L,其中 DNA 模板 5 μ L (20 ng/ μ L), 10 $\times$ Easy Taq Buffer 2 μ L, 2.5 mmol/L dNTPs 1.5 μ L, Easy Taq 酶 (5U/ μ L) 0.3 μ L, 2 μ mol/L SSR 引物(正反向引物混合)3 μ L, ddH₂O 8.2 μ L。PCR 反应程序为 95 $^{\circ}$ C 预变性 5 min, 95 $^{\circ}$ C 变性 30 s, 55 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 30 s, 循环 35 次,最后 72 $^{\circ}$ C 延伸 5 min, 4 $^{\circ}$ C 保存。PCR 产物中加入 6 $\times$ loading buffer 5 μ L, 95 $^{\circ}$ C 变性 3 min, 变性后迅速放入冰水混合物中直至冷却。

利用 6% 变性聚丙烯酰胺电泳鉴定基因型,将与母本中黄 35 带型一致的基因型记为 A,将与父本十胜长叶带型一致的基因型记为 B,杂合带型记为 H,缺失或模糊不清的带型记为“−”。

1.6 数据统计分析

利用 Microsoft Excel 2013 和 SPSS19.0 分析群体最大值、最小值、平均数、标准差、偏度、峰度、变异系数和相关系数,利用 Minitab16 绘制蛋白质含量分布图,利用 R 语言对定位区间两端标记基因型与蛋白质含量进行差异性分析。

1.7 QTL 定位

利用 QTL IciMapping4.1 软件中的 Map 程序进行遗传图谱构建,利用 BIP 功能中的完备区间作图法对大豆蛋白质含量进行定位。

2 结果与分析

2.1 亲本及 $F_{2:15}$ 和 $F_{2:16}$ 群体的蛋白质含量分布

2 个亲本的蛋白质含量差异很大,但每个亲本蛋白质含量在 2018 年和 2019 年之间变化很小,其中,中黄 35 两年蛋白质含量平均数分别为 40.31% 和 39.25%,变异幅度为 1.06%;十胜长叶两年蛋白质含量平均数分别为 46.21% 和 44.47%,变异幅度为 1.74%。中黄 35 和十胜长叶之间蛋白质含量存在显著差异,并且在两年间蛋白质含量变化稳定,表明中黄 35 和十胜长叶分别为稳定的低蛋白和高蛋白品种,适合作为大豆蛋白质含量 QTL 定位群体的亲本。这 2 个亲本衍生的 RIL 群体 199 个株系的蛋白质含量变异很大(表 2)。从表中可以看出, $F_{2:15}$ 群

表 2 两年两代亲本及重组自交系群体蛋白质含量统计分析

Table 2 Statistical analysis of protein content in two generations of parents recombinant inbred lines

年份 / 环境 Year/Environment	中黄 35 ZH35	十胜长叶 SSCY	重组自交系 Recombinant inbred lines			
	平均数 $\pm$ 标准差	平均数 $\pm$ 标准差				
	(%) Mean $\pm$ SD	(%) Mean $\pm$ SD	平均数 $\pm$ 标准差 (%) Mean $\pm$ SD	变异系数 (%) CV	峰度 Kurtosis	偏度 Skewness
2018/ 北京	40.31 $\pm$ 0.53	46.21 $\pm$ 0.61	41.94 $\pm$ 2.02	4.81	-0.26	0.09
2019/ 北京	39.25 $\pm$ 0.81	44.47 $\pm$ 1.01	40.36 $\pm$ 2.26	6.49	-0.41	-0.27

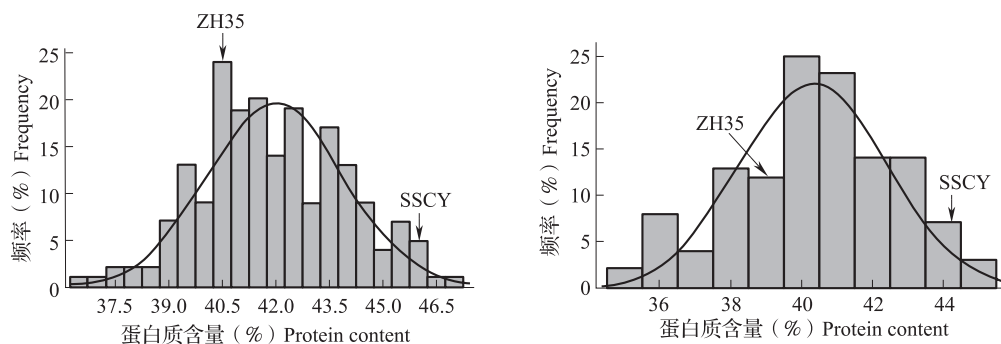


图 2 $F_{2:15}$ 家系 (左)、 $F_{2:16}$ 家系 (右) 蛋白质含量频率分布

Fig.2 Frequency distribution of protein content in $F_{2:15}$ (left), $F_{2:16}$ (right) strains

2.2 用 BSA 法筛选多态性 SSR 标记

利用 872 对 SSR 引物扩增亲本中黄 35 和十胜长叶的基因组 DNA,共筛选出在亲本间具有多态性的 SSR 位点 316 个,占筛选 SSR 位点总数的 36.2% (图 3)。用筛选出的 SSR 标记分析蛋白高低“基因池”,鉴定出 9 个 SSR 标记存在差异,包括 3 号染色体 2 个(表 3),相距为 1.2 Mb;18 号染色体 4 个,相距为 3 Mb;19 号染色体 3 个(图 4),相距为 1.2 Mb。其中,3 号染色体 (SSR_03_75 和 SSR_03_48) 位于

体蛋白质含量的变异幅度为 36.63%~47.18%,平均数为 41.94%,变异系数为 4.81%; $F_{2:16}$ 群体蛋白质含量的变异幅度为 34.78%~44.87%,平均数为 40.36%,变异系数为 6.49%。2 个世代均存在超亲分离现象,且随着时间和环境的变化超亲分离比有所不同,超亲品系为选育高蛋白材料提供了基础。利用 KS- 检验,对中黄 35 和十胜长叶 2 个世代群体蛋白质含量分布进行正态分布检验,中黄 35 和十胜长野 $F_{2:15}$ 和 $F_{2:16}$ 群体, P 值分别为 0.787、0.848,两代群体 P 值均大于 0.05,表明两代群体蛋白质含量分布呈正态分布(图 2),可用于蛋白质含量的 QTL 定位。

2 个世代蛋白质含量相关性系数为 0.393,呈极显著正相关,这表明,尽管两代群体蛋白质含量变异幅度存在差异,但群体整体变化趋势一致。

前人 (Sat_304 和 Satt549) 已定位区间内^[18];18 号染色体标记 (SSR_18_77 和 SSR_18_48) 与前人 (AW734137 和 Satt610) 的定位区间有部分重叠^[19],区间长度为 440 kb;19 号染色体标记 (SSR_19_27 和 SSR_19_98) 区间未见报道。因此,进一步在 19 号染色体连锁标记 (SSR_19_27 和 SSR_19_98) 附近选择 43 个已报道的 SSR 标记^[17],并在亲本间进行多态性标记筛选,鉴定出 25 个多态性 SSR 标记,多态性比例为 58.1%。

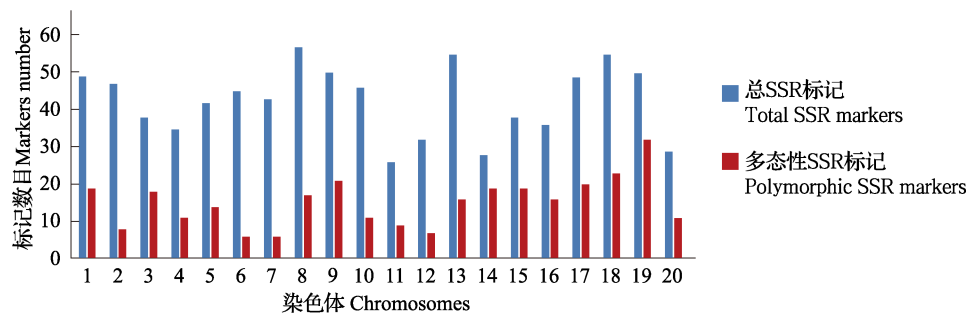
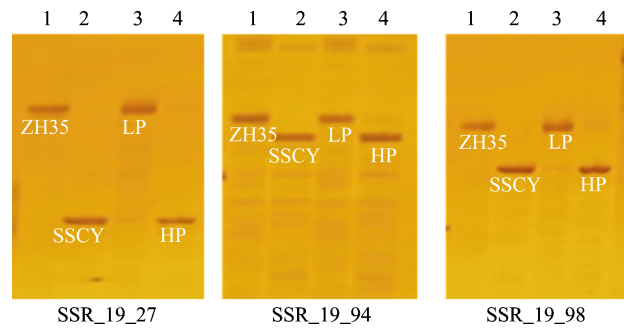


图 3 大豆 20 条染色体筛选 SSR 标记总数及多态性标记数
Fig.3 Screening SSR markers and polymorphism markers on 20 chromosomes

表 3 BSA 法筛选到的蛋白质含量相关标记
Table 3 Protein content related markers screened by BSA method

标记 Markers	染色体 Chr.	物理位置 Location	亲本基因型 Parental genotype		混池基因型 Pool genotype		参考文献 References
			中黄 35 ZH35	十胜长叶 SSCY	低池 LP	高池 HP	
SSR_03_75	3	37631393	A	B	H	B	[18]
SSR_03_48	3	38834802	A	B	H	B	[18]
SSR_18_77	18	3162724	A	B	H	B	[19]
SSR_18_95	18	5348281	A	B	H	B	[19]
SSR_18_16	18	5702492	A	B	H	B	[19]
SSR_18_48	18	6169553	A	B	H	B	[19]
SSR_19_27	19	47167467	A	B	A	B	-
SSR_19_94	19	48338071	A	B	A	B	-
SSR_19_98	19	48389366	A	B	A	B	-

-: 未见报道
-: No reference was found



1: 中黄 35; 2: 十胜长叶; 3: 低蛋白家系池; 4: 高蛋白家系池
1: ZH35, 2: Tokachi nagaha, 3: Low protein strains pool, 4: High protein strains pool

图 4 3 个 SSR 标记在 19 号染色体亲本和后代 DNA 混合池间的多态性
Fig.4 Polymorphism of three SSR markers between parent and progeny DNA pools on chromosome 19

2.3 大豆蛋白质含量 QTL 分析

从 19 号染色体蛋白质含量连锁标记 (SSR_19_27 和 SSR_19_98) 附近用亲本筛选 25 个多态性 SSR 标记中, 根据物理位置选择均匀分布的 22 个 SSR 标记分析 RIL 群体基因型, 结合蛋白质含量表型数据定位蛋白质含量相关 QTL。用 Icimapping4.1map 程序构建遗传连锁图谱, Bip 程序设定 QTL 的显著水平为 0.05, 进行 1000 次模拟运算获得 QTL 显著的 LOD 阈值, 利用完备区间作图法定位蛋白质含量 QTL。结果表明, $F_{2:15}$ 、 $F_{2:16}$ 两代数据的阈值分别为 2.5 和 1.8, 均将蛋白质含量 QTL 定位在分子标记 SSR_19_38 和 SSR_19_59 之间 (图 5), 命名为 *qPRO-19-1*, 区间长度为 384 kb, 包括 36 个基因。其中, $F_{2:15}$ 群体 LOD 值为 3.43 (表 4), 可解释的表型变异率为 7.81%, 加性效应为 -0.56。 $F_{2:16}$ 群体 LOD 值为 3.98, 可解释的表型变异率为 14.87%, 加性效应为 -0.84。两代蛋白质含量 QTL *qPRO-19-1* 加性效应均为负值, 表明定位高蛋白的基因来自于高蛋白亲本十胜长叶。

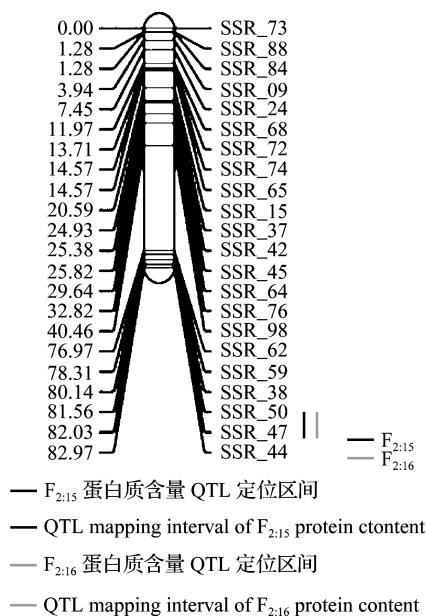


图 5 中黄 35 和十胜长叶 RIL 群体 $F_{2:15}$ 和 $F_{2:16}$ 在大豆 19 号染色体定位蛋白质含量 QTL *qPRO-19-1*

Fig.5 The protein content of QTL *qPRO-19-1* in soybean chromosome 19 was determined by $F_{2:15}$ and $F_{2:16}$ of ZH35 and SSCY RIL populations

表 4 $F_{2:15}$ 和 $F_{2:16}$ 群体定位蛋白质含量 QTL 的位置和参数

Table4 Chromosome locations and parameters associated with the quantitative loci (QTL) of protein contents in the $F_{2:15}$ and $F_{2:16}$

群体 Population	染色体 Chr.	标记区间 Marker interval	LOD	表型变异率 (%) PVE	加性效应 Additive effective
$F_{2:15}$	19	SSR_19_38-SSR_19_59	3.43	7.81	-0.56
$F_{2:16}$	19	SSR_19_38-SSR_19_59	3.98	14.87	-0.84

2.4 *qPRO-19-1* 位点两端标记基因型与蛋白质含量的差异性分析

统计 $F_{2:15}$ 和 $F_{2:16}$ 群体在 *qPRO-19-1* 的定位标记 SSR_19_38 和 SSR_19_59 基因型以及蛋白质含量。 $F_{2:15}$ 群体中, 在 SSR_19_38 位点, 具有中黄 35 基因型 (A) 的后代有 100 个, 蛋白质含量变化范围为 36.63%~46.10%, 平均数为 41.38%; 具有十胜长叶基因型 (B) 的后代有 83 个, 蛋白质含量变化范围为 38.97%~46.48%, 平均数为 42.54%。在 SSR_19_59 位点, 具有中黄 35 基因型 (A) 的后代有 106 个, 蛋白质含量变化范围为 36.63%~46.10%, 平均数为 41.44%; 具有十胜长叶基因型 (B) 的后代有 82 个, 蛋白质含量变化范围为 38.97%~46.48%, 平均数为 42.47%。 $F_{2:16}$ 群体中, 在 SSR_19_38 位点, 具有中黄 35 基因型 (A) 的后代有 64 个, 蛋

白质含量变化范围为 35.78%~44.59%, 平均数为 39.55%; 具有十胜长叶基因型 (B) 的后代有 49 个, 蛋白质含量变化范围为 35.37%~44.87%, 平均数为 41.36%。在 SSR_19_59 位点, 具有中黄 35 基因型 (A) 的后代有 66 个, 蛋白质含量变化范围为 34.78%~44.63%, 平均数为 39.66%; 具有十胜长叶基因型 (B) 的后代有 47 个, 蛋白质含量变化范围为 35.37%~44.87%, 平均数为 41.29%。用 R 语言对两端标记基因型 (A) 和基因型 (B) 蛋白质含量进行差异性分析。 $F_{2:15}$ 群体中 (图 6), 在 SSR_19_38 位点, P 值为 0.00018, 在 SSR_19_59 位点, P 值为 0.0013。 $F_{2:16}$ 群体中 (图 7), 在 SSR_19_38 位点, P 值为 1×10^{-5} , 在 SSR_19_59 位点, P 值为 5.6×10^{-5} 。2 个世代 2 个位点的 2 种基因型的蛋白质含量均达极显著差异。

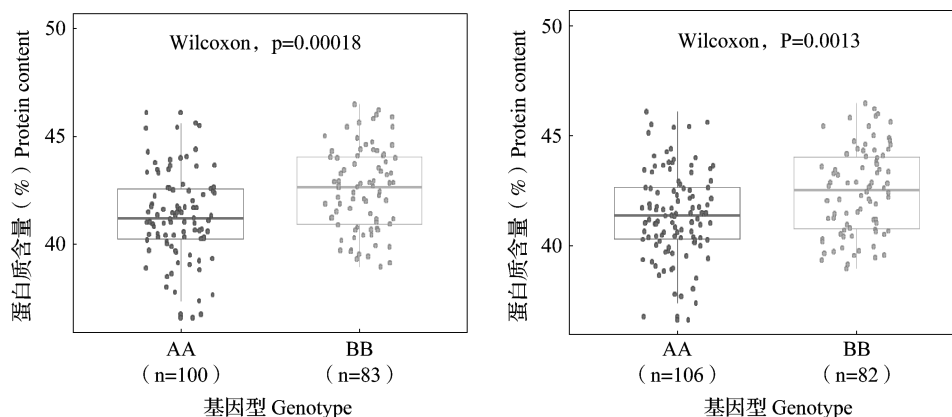


图 6 $F_{2:15}$ 群体 *qPRO-19-1* 两侧标记 SSR_19_38 (左)、SSR_19_59 (右) 不同基因型的蛋白质含量分布及差异性显著测验
Fig.6 The protein content distribution and significant differences of SSR_19_38 (left) and SSR_19_59 (right) markers on both sides of *qPRO-19-1* in $F_{2:15}$ population were tested

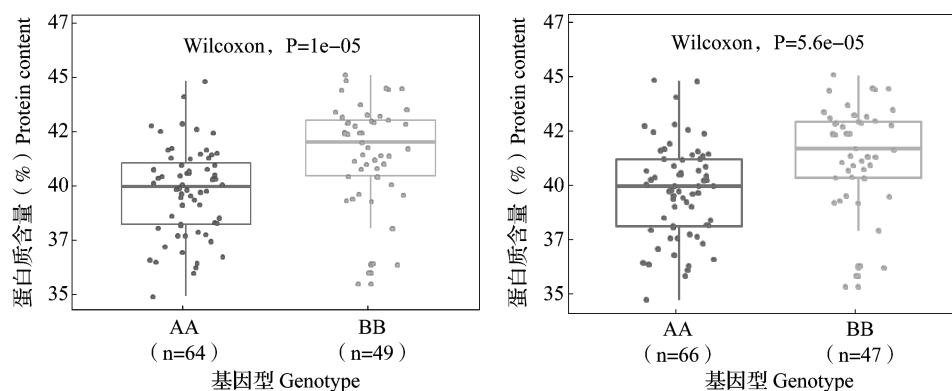


图 7 $F_{2:16}$ 群体 *qPRO-19-1* 两侧标记 SSR_19_38 (左)、SSR_19_59 (右) 不同基因型的蛋白质含量分布及差异性显著测验
Fig.7 The protein content distribution and significant differences of SSR_19_38 (left) and SSR_19_59 (right) markers on both sides of *qPRO-19-1* in $F_{2:16}$ population were tested

2.5 定位区间候选基因的分析

通过 Soybase 网站查询定位区间内的 36 个基因,其中 31 个基因有功能注释(表 5)。根据基因功能注释进行检索,推测有 3 个基因可能与种子蛋白质含量相关,其中 *Gene-18* 基因参与半胱氨酸生物合成, *Gene-25* 基因是半胱氨酸型肽酶, *Gene-34* 基因具有转运氨基酸的功能。

3 讨论

3.1 大豆优异种质的筛选

大豆蛋白质含量为数量性状,遗传机制比较复杂,易受温度、光照和水分等外界环境条件影响^[20]。随着中国大豆优势产业区划的逐步形成与完善以及大豆加工业专用化和综合化利用的需要,种质资源鉴定和优异品种的选育推广越来越重要^[21]。选育高蛋白大豆品种也越来越受到许多学者的重视。在本研究 RIL 群体中有些家系的蛋白存在超高亲现

象,其中 $F_{2:15}$ 中有 2 个家系, $F_{2:16}$ 中有 3 个家系;在超过 43% 和 45% 的高蛋白标准家系中, $F_{2:15}$ 分别有 58 个和 16 个, $F_{2:16}$ 分别有 14 个和 0 个。在 2 个世代中,蛋白质含量均高于 43% 的家系共 7 个,其中与高蛋白含量亲本十胜长叶基因型一致的为 6 个。这些蛋白含量稳定且具有高蛋白基因型的优异家系,为选育高蛋白品种奠定了材料基础。

3.2 QTL *qPRO-19-1* 是控制大豆蛋白质含量 QTL 新位点且稳定表达

截止到 2020 年 1 月,在大豆 Soybase 数据库 (<https://www.Soybase.org>) 中,19 号染色体上已定位的 QTL 位点共有 10 个,分别为 Seed protein 1-7 (35672961-36154603)、Seed protein 16-2 (48220448-50545228)、Seed protein 2-2 (45014383-46509345)、Seed protein 33-4 (2059590-7127430)、Seed protein 34-10 (14048834-33865153)、Seed protein 36-31 (45182458-46227880)、Seed protein 36-32 (40319984-

表 5 Soybase 网站定位区间内基因功能注释信息

Table5 The Soybase website locates information gene annotation in the interval

基因 Gene	功能注释 Functional annotation	基因 Gene	功能注释 Functional annotation
Gene-1	细胞内蛋白质运输; 核苷酸结合; ATP 结合	Gene-18	辅酶生物合成过程; 核苷酸代谢过程; 半胱氨酸的生物合成
Gene-2	生物过程中分子功能	Gene-19	翻译后蛋白质修饰; 微管细胞骨架组织; 转录调控
Gene-3	钾离子跨膜转运	Gene-20	根系发育的调控; 纤毛虫介导的细胞间运输
Gene-4	转移除氨基酸酰基以外的其他酰基	Gene-21	参与多维细胞生长的植物型细胞壁修饰; 植物型细胞壁组织
Gene-5	生物过程中分子功能	Gene-22	生物过程中分子功能
Gene-6	蛋白质翻译的他 RNA 氨酰化; 脯氨酰 tRNA 氨酰化	Gene-23	淀粉生物合成过程; 核苷酸转移酶活性
Gene-8	碳水化合物代谢过程	Gene-25	蛋白水解; 半胱氨酸型肽酶
Gene-9	负调控转录	Gene-26	蛋白质泛素化; 脂质储存; 糖介导的信号通路
Gene-10	碳水化合物代谢过程; 氨基酸转运; 硝酸盐运输	Gene-27	RNA 结合; 核酸结合
Gene-11	调控转录; 促进根生长	Gene-28	细胞外区域生物过程
Gene-12	调控转录; 信号转导; 花青素代谢过程的正调控	Gene-29	生物过程中分子功能
Gene-13	调节有丝分裂细胞周期; 蛋白质运输	Gene-31	核苷酸生物合成过程; 染色体组织调节
Gene-14	转录调控; 特异性序列 DNA 结合转录因子	Gene-33	翻译的负调控; 信号转导
Gene-15	生物过程中分子功能	Gene-34	转运氨基酸; N 端蛋白豆蔻酰化
Gene-16	羧酸酯酶活性	Gene-35	脱落酸运输; 与钙离子结合
Gene-17	硝酸盐运输; 水解酶活性		

45182458)、Seed protein 41-8 (386234-843081)、Seed protein 8-1 (42119600-42311222)、Seed protein 30-7 (40154846-40637071), 这些 QTL 区间互不重叠。其中, Orf 等^[22]和 Tajuddin 等^[23]定位的蛋白质含量 QTL (Seed protein 8-1 和 Seed protein 30-7) 物理位置较近, 物理区间大约相差 1.4 Mb。Diers 等^[24]定位的蛋白质含量 QTL (Seed protein 1-7) 物理区间较小, 大约为 482 kb。Lu 等^[25]在 19 号染色体上定位到蛋白质含量 QTL (Seed protein 34-10) 物理区间最大, 大约为 19.8 Mb。其中, Mao 等^[26]在 19 号染色体上定位到 2 个蛋白质含量 QTL (Seed protein 36-31 和 Seed Protein 36-32), 区间大小分别为 1.1 Mb 和 4.8 Mb。对本研究定位的高蛋白 QTL *qPRO19-1* 与 Mansur 等^[27]定位蛋白质含量的 QTL (Seed protein 2-2) 区间最近, 但相距 783 kb, 是尚未见报道的新位点。

数量性状易受环境条件的影响, 可在不同地点、不同年份多次定位到的 QTL, 说明该 QTL 具有重演性强、稳定性高、利用价值高等特点。截止到 2020 年 1 月, Soybase 数据库 (<https://www.soybase.org>) 公布的与大豆蛋白质含量相关的 QTL 超过 200 个, 但具有重演性的 QTL 很少^[28-30]。本研究利用中黄 35 和十胜长叶 RIL 群体进行定位, 虽然 $F_{2:15}$ 和 $F_{2:16}$ 的群体规模不同, 分别为 199 个和 120 个株系, 但均将蛋白质含量相关的 QTL 区间定位在

相同位置, 说明该蛋白质含量相关的 QTL 为稳定的 QTL 区间, 具有研究和利用潜力, 也为高蛋白育种的分子标记选择提供了选择技术。

3.3 大豆蛋白质 QTL 的精细定位鲜见报道

目前大豆蛋白 QTL 定位研究 (<https://www.Soybase.org>) 较多, 但关于精细定位和图位克隆基因的报道较少。Yang 等^[31]利用染色体片段置换系群体, 将蛋白质含量 QTL 物理区间精细定位在 329 kb 内, 并预测 *Glyma.15g049200* 基因可能与大豆蛋白含量相关。Lestari 等^[32]通过比较与大豆蛋白质含量相关的重复基因组区域, 选择了 2 个蛋白质含量定位区间, 其中 10 号染色体共有 35 个基因, 20 号染色体中共有 19 个基因。本研究利用 BSA 法, 将大豆蛋白质含量 QTL 定位在 1.2 Mb 区间, 通过定位区间加密分子标记, 将定位区间缩小至 384 kb。在 *qPRO-19-1* 精细定位区间内有 31 个假定基因, 根据功能推测, *Gene-18*、*Gene-25* 和 *Gene-34* 这 3 个基因与大豆种子蛋白含量关系比较密切。对 3 个基因的蛋白家族进行查询, 其中 *Gene-18* 基因是亲环素 -ABH 家族基因, 与蛋白质折叠相关; *Gene-25* 基因是 1 个木瓜蛋白酶蛋白家族基因, 在植物中可以调节种子中的贮藏蛋白含量^[33]; *Gene-34* 基因是 1 个 Cupin 蛋白家族基因, 在植物中能够催化种子贮藏蛋白含量。同时, 对定位区间内候选基因种子发育不同时期表达量进行查询 (图 8), *Gene-*

Gene-6、*Gene-25*、*Gene-27*、*Gene-28* 和 *Gene-34* 基因在种子中表达量较高, 推测基因 *Gene-25* 和 *Gene-34* 可能是候选基因, 但有待进一步的研究验证。

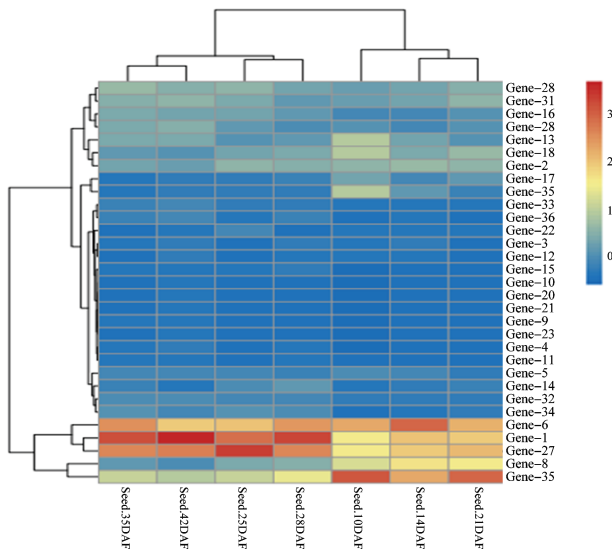


图8 Soybase 网站候选区间内基因种子不同发育时期表达量
Fig.8 The expression of seeds at different developmental stages in the candidate interval of Soybase website

参考文献

- [1] Leamy L J, Zhang H Y, Li C B, Chen C Y, Song B H. A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genomics*, 2017, 18(1): 1-18
- [2] 王文真, 刘兴媛, 曹永生, 张明. 中国大豆种质资源的蛋白质含量研究. *作物品种资源*, 1998, 17(1): 35-36
Wang W Z, Liu X Y, Cao Y S, Zhang M. Study on protein content of soybean germplasm resources in China. *Crop Germplasm Resources*, 1998, 17(1): 35-36
- [3] Patil G, Vuong T D, Kale S, Valliyodan B, Deshmukh R, Zhu C, Wu X, Bai Y, Yungbluth D, Lu F, Kumparla S, Shannon J G, Varshney R K, Nguyen H T. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal*, 2018, 16(11): 1939-1953
- [4] Bostein D R, White R L, Skolnick M, Davis R W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 1980, 32(3): 314-331
- [5] 邱丽娟, Randaiil N, Lilao V. 利用 RAPD 标记鉴定大豆种质. *作物学报*, 1997, 4(4): 408-417
Qiu L J, Randaiil N, Lilao V. Evaluation of soybean germplasm with random amplification polymorphic DNA (RAPD) Markers. *Acta Agronomica Sinica*, 1997, 4(4): 408-417
- [6] Young W P, Schupp J M, Keim P. DNA methylation and AFLP marker distribution in the soybean genome. *Theoretical and Applied Genetics*, 1999, 99(5): 785-792
- [7] Mello M B, Milene M, Imaculada Z M, Quecini V, Pinheiro J B. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. *Pesquisa Agropecuária Brasileira*, 2010, 45(3): 276-283
- [8] Keim P, Diers B W, Shoemaker R C. Genetic analysis of soybean hard seediness with molecular markers. *Theoretical and Applied Genetics*, 1990, 79(4): 465-469
- [9] Liang H Z, Yu Y L, Wang S F, Lian Y, Wang T F, Wei Y L, Gong P T, Liu X Y, Fang X J, Zhang M C. QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). *Agricultural Sciences in China*, 2010, 9(8): 1108-1116
- [10] Lu W G, Wen Z X, Li H C, Yuan D H, Li J Y, Zhang H, Zhong W H, Cui S Y, Du W J. Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *Theoretical and Applied Genetics*, 2013, 126(2): 425-433
- [11] Wang J, Chen P, Wang D, Shannon G, Zeng A, Orazaly M, Wu C. Identification and mapping of stable QTL for protein content in soybean seeds. *Molecular Breeding*, 2015, 35(3): 92-102
- [12] Wang X Z, Jiang G L, Green M, Scott R A, Song Q J, Hyten D L, Cregan P B. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Molecular Genetics and Genomics*, 2014, 289(5): 935-949
- [13] Shan D P, Liu C Y, Jiang H W, Dong X H, Chen Q S, Hu G H. QTL analysis of soybean protein content using two methods in 5 different environments. *Chinese Journal of Oil Crop Sciences*, 2011, 33(1): 9-14
- [14] Qi Z M, Hou M, Han X, Liu C Y, Jiang H W, Xin D W, Hu G H, Chen Q S. Identification of quantitative trait loci (QTLs) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTLs under multiple environments. *Plant Breeding*, 2014, 133(4): 499-507
- [15] 郭娟娟, 常汝镇, 章建新, 张巨松, 关荣霞, 邱丽娟. 日本大豆种质十胜长叶对我国大豆育成品种的遗传贡献分析. *大豆科学*, 2007, 26(6): 807-812
Guo J J, Chang R Z, Zhang J X, Zhang J S, Guan R X, Qiu L J. Contribution of Japanese soybean germplasm Tokachi-nagaha to Chinese soybean cultivars. *Soybean Science*, 2007, 26(6): 807-812
- [16] Michelmore R W, Paran I, Kesseli R V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America*, 1991, 88(21): 9828-9832
- [17] Song Q J, Jia G F, Zhu Y L, Grant D, Nelson R T, Hwang E, Hyten D L, Cregan P B. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Science*, 2010, 50(5): 1950-1960
- [18] Kabelka E A, Diers B W, Fehr W R, Leroy A R, Baianu I C, You T F, Neece D J, Nelson R L. Putative alleles for increased yield from soybean plant introductions. *Crop science*, 2004, 44(3): 784-791

- [19] Panthee D R, Pantalone V R, West D R, Saxton A. Quantitative trait loci for seed protein and oil concentration and seed size in soybean. *Crop Science*, 2005, 45 (5): 2015-2022
- [20] 韩雪, 齐慧冬, 齐照明, 陈庆山. 基于 GGE 双标图的大豆蛋白质多环境的稳定性分析. *农业科学*, 2018, 8 (7): 801-808
Han X, Qi H D, Qi Z M, Chen Q S. Analysis of soybean protein content for environment stability based on GGE biplot. *Journal of Agricultural Science*, 2018, 8 (7): 801-808
- [21] 李为喜, 朱志华, 刘三才, 刘方, 张晓芳, 李燕, 王述民. 中国大豆 (*Glycine max*) 品种及种质资源主要品质状况分析. *植物遗传资源学报*, 2004, 5 (2): 185-192
Li W X, Zhu Z H, Liu S C, Liu F, Zhang X F, Li Y, Wang S M. Quality characters of Chinese soybean (*Glycine max*) varieties and germplasm resources. *Journey of Plant Genetic Resources*, 2004, 5 (2): 185-192
- [22] Orf J H, Chase K, Jarvik T, Mansur L M, Cregan P B, Adler F R, Lark K G. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Science*, 1999, 39 (6): 1642-1651
- [23] Tenku T, Satoshi W, Ryoichi M, Kyuya H, Sumio K. Application of near infrared transmittance spectroscopy to the estimation of protein and lipid contents in single seeds of soybean recombinant inbred lines for quantitative trait loci analysis. *Journal of Near Infrared Spectroscopy*, 2002, 10 (1): 315-348
- [24] Diers B W, Keim P, Fehr W R, Shoemaker R C. RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, 1992, 83 (5): 608-612
- [25] Lu W G, Wen Z X, Li H C, Yuan D H, Li J Y, Zhang H, Huang Z W, Cui S Y, Du W J. Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *Theoretical and Applied Genetics*, 2013, 126 (2): 425-433
- [26] Mao T T, Jiang Z F, Han Y P, Teng W L, Zhao X, Li W B. Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding*, 2013, 132 (6): 630-641
- [27] Mansur L, Orf J, Chase K, Jarvik T, Cregan P B, Lark K. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Science*, 1996, 36 (5): 1327-1336
- [28] 仕相林, 孙亚男, 王家麟, 刘春燕, 陈庆山, 胡国华. 大豆叶片性状 QTL 的定位及 Meta 分析. *作物学报*, 2012, 38 (2): 256-263.
Shi X L, Sun Y N, Wang J L, Liu C Y, Chen Q S, Hu G H. Mapping and Meta-analysis of QTLs for leaf traits in soybean. *Acta Agronomica Sinica*, 2012, 38 (2): 256-263
- [29] 刘春燕, 齐照明, 韩冬伟, 单大鹏, 蒋洪蔚, 陈庆山, 胡国华. 大豆产量相关性状的多年多点 QTL 分析. *东北农业大学学报*, 2010, 41 (11): 1-9
Liu C Y, Qi Z M, Han D W, Shan D P, Jiang H W, Chen Q S, Hu G H. QTL analysis of yield components on soybean under different environment. *Journal of Northeast Agricultural University*, 2010, 41 (11): 1-9
- [30] 范冬梅, 孙殿君, 马占洲, 刘春燕, 杨振, 曾庆力, 辛大伟, 蒋洪蔚, 邱鹏程, 陈庆山, 胡国华. 多种环境下大豆单株粒重 QTL 的定位与互作分析. *作物学报*, 2013, 39 (6): 1021-1029
Fan D M, Sun D J, Ma Z Z, Liu C Y, Yang Z, Zeng Q L, Xin D W, Jiang H W, Qiu P C, Chen Q S, Hu G H. QTL mapping and interaction analysis of seed weight per plant in soybean among different environments. *Acta Agronomica Sinica*, 2013, 39 (6): 1021-1029
- [31] Yang H Y, Wang W B, He Q Y, Xiang S H, Tian D, Zhao T J, Gai J Y. Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. *Theoretical and Applied Genetics*, 2019, 132 (6): 2793-2807
- [32] Lestari P, Van K, Lee J, Kang Y J, Lee S H. Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. *Frontiers in Plant Science*, 2013, 2 (4): 1-8
- [33] 张焕, 仇忠凯, 阎新, 欧阳解秀, 李绍波, 王鑫. 水稻种子特异表达基因 *OsEnS38* 的克隆与表达. *湖南农业大学学报: 自然科学版*, 2018, 44 (3): 278-283
Zhang H, Chou Z K, Yan X, Ouyang J X, Li S B, Wang X. Cloning and expression analysis of a seed-specific gene *OsEnS38* in rice. *Journal of Human Agricultural University: Natural Sciences*, 2018, 44 (3): 278-283