

大豆蛋白质含量相关位点 *qPRO-19-1* 的精细定位

刘亭萱^{1,2}, 郭兵福³, 栾晓燕⁴, 王 俊¹, 邱丽娟²

(¹ 长江大学农学院, 湖北荆州 434025; ² 中国农业科学院作物科学研究所, 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)

摘要: 大豆是人类膳食蛋白的主要来源之一, 提高大豆蛋白含量是主要的育种目标。因此, 挖掘调控大豆蛋白质含量的关键基因, 对高蛋白大豆品种的定向选育有重要意义。本研究以中黄 35 和十胜长叶为亲本, 构建了 $F_{2:16}$ 与 $F_{2:17}$ 重组自交系 (RIL, Recombination Inbred Lines) 群体, 以前期定位的蛋白质含量新位点 *qPRO-19-1* 为基础, 对区间内基因测序, 发现 *Glyma.19g223300* 在亲本间有 1 个 InDel 变异并开发标记 IN-1, 同时从区间内 22 个 SSR 标记中筛选出 5 个多态性标记, 通过整合基因型和蛋白质含量表型数据, 将定位区间 *qPRO-19-1* 从 384 kb 缩小到 68.03kb, 包括注释基因 4 个, 其中 *Glyma.19g221800* 和 *Glyma.19g222000* 两个基因在种子不同发育时期存在极显著的表达差异。研究结果为该大豆蛋白含量相关基因的图位克隆及分子标记辅助育种提供了参考。

关键词: 大豆; 蛋白含量; 精细定位

Fine Mapping of the Protein-Related Locus *qPRO-19-1* in Soybean

LIU Ting-xuan^{1,2}, GUO Bing-fu³, LUAN Xiao-yan⁴, WANG Jun¹, QIU Li-Juan²

(¹ College of Agriculture, Yangtze University, Hubei Jingzhou 434025; ² 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)

Abstract: Soybean is one of the main sources supplying human dietary protein, and increasing its protein content is one of major targets in soybean breeding. Identification of the genes modulating the protein content is of great significance in breeding of soybean varieties with higher protein content. In this study, the $F_{2:16}$ and $F_{2:17}$ recombinant inbred lines (RIL) populations derived from Zhonghuang 35 and Tokachi nagaha were applied to delimit the previously-mapped protein content locus *qPRO-19-1*. By sequencing the candidate genes in the interval, an InDel marker in *Glyma.19g223300* between parents and five polymorphic markers from 22 SSR markers were developed. The physical interval harboring *qPRO-19-1* was further delimited from 384 kb to 68.03 kb, including four annotated candidate genes, of which *Glyma.19g221800* and *Glyma.19g222000* showed extremely significant difference on transcript level at different stages of seed development. These results provided a basis for future map-based cloning and molecular marker-assisted breeding of this protein content-related gene in soybean.

Key words: soybean; protein content; fine mapping

大豆是一种富含优质植物蛋白的经济作物, 所提供的蛋白质约占全世界蛋白质消费总量的 68%^[1]。大豆籽粒蛋白富含人体必需氨基酸, 氨基酸的组成

基本接近人体的需要, 属于营养价值较高的优质蛋白^[2]。近几年, 国内高蛋白大豆消费急剧增加, 供需不平衡, 市场对培育高蛋白大豆品种提出了更高

收稿日期: 2022-05-19 修回日期: 2022-06-27 网络出版日期: 2022-08-08

URL: <https://doi.org/10.13430/j.cnki.jpgr.20220519001>

第一作者研究方向为大豆优异基因挖掘, E-mail: 18834408825@163.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: qielujuan@caas.cn

王 俊, 研究方向为大豆分子生物学与遗传育种, E-mail: wangiagri@yangtzeu.edu.cn

基金项目: 国家自然科学基金 (31960408)

Foundation project: National Natural Science Foundation of China (31960408)

的要求^[3]。因此,挖掘大豆高蛋白相关基因,解析蛋白含量的遗传机制,对大豆高蛋白品种的培育具有重要的实践意义。

在许多大豆育种项目中,对籽粒蛋白含量进行了广泛的研究^[4-6]。大豆蛋白含量是一个典型的数量性状,受遗传和环境因素的共同影响,遗传规律极其复杂^[7],相关基因定位研究一直以来是大豆分子生物学的研究热点。截至 2022 年, Soybase (<https://www.soybase.org>) 数据库公布了 241 个与大豆蛋白质含量相关的 QTLs, 在 20 条染色体上均有分布。Csanadi 等^[8]利用 F₂ 群体在 5 个不同的环境下,筛选到 4 个与蛋白含量有关的标记,分别位于第 1、6、7 和 9 号染色体上。Warrington 等^[9]利用高产晚熟品种与高蛋白中熟品种构建 RIL 群体,检测到 4 个与蛋白含量相关的 QTL, 其中 20 号染色体上的位点可解释 55% 的表型变异。Pandurangan 等^[10]利用重组自交系群体对蛋白质含量进行 QTL 定位,将蛋白质含量 QTL 定位在 4 条染色体上,其中,在 6 号染色体上区间最小为 122 kb,在 20 号染色体上区间最大为 22 Mb。Zhang 等^[11]利用重组自交系群体鉴定到 1 个与蛋白质含量相关的 QTL 位点,位于第 14 号染色体,解释了 12.4% 的表型变异。陈庆山等^[12]利用美国半矮秆大豆 Charleston 与东农 594 构建的重组自交系,检测到 5 个与蛋白含量相关的 QTL 位点,分别位于第 8、10、19 和 20 号染色体,解释的表型变异在 5.16%~14.29% 之间。经报道,已定位的大豆蛋白含量相关的 QTL 位点在第 19 号染色体上报道较少,其中 Li 等^[13]通过连锁分析结合关联分析,在多个环境下鉴定蛋白含量相关的 QTL,发现第 19 号染色体上有两个位点。Jun 等^[14]利用 LD 进行关联定位,鉴定大豆种子蛋白 QTL,在第 19 号染色体上检测到两个位点。Lu 等^[15]在第 19 号染色体上定位到 1 个蛋白质含量 QTL,物理区间大约为 19.8 Mb。目前,与大豆高蛋白相关的基因只有少数被克隆,比如 Fliege 等^[16]发现 *cqSeedprotein-003* QTL 的高或低种子蛋白含量等位基因由 *Glyma.20G85100* 基因编码的 CCT 结构域蛋白中插入的转座子引起,该基因可增加种子大约 2% 的蛋白质含量。Wang 等^[17]发现一对 SWEET 同源基因 *GmSWEET10a* 和 *GmSWEET10b* 在大豆驯化过程中经历了逐步选择,可以同时改变大豆种子大小、含油量和蛋白质含量。

本研究以前期定位在第 19 号染色体的蛋白

质含量 QTL 新位点 *qPRO-19-1* 为基础^[18],利用中黄 35 和十胜长叶构建的 F_{2:16} 与 F_{2:17} RIL 群体,将 *qPRO-19-1* 的定位区间由 384kb 缩小到 68.03 kb,区间内有 4 个注释基因。该研究为大豆高蛋白相关基因克隆、功能研究及分子标记育种提供借鉴与参考。

1 材料与方法

1.1 试验群体

本研究利用低蛋白质品种中黄 35 和高蛋白质品种十胜长叶配置组合,从 F₂ 通过单粒传法构建 RILs 群体。2020 年,在海南三亚种植 F_{2:16} 群体,包含 193 个株系;2021 年,在北京昌平种植 F_{2:17} 群体,包含 173 个株系。亲本及 RIL 群体的株系分别种植 1 行,行长 2 m,行距 0.45 m,株距 0.1 m。成熟时每行收获 5 个单株。

1.2 表型调查

挑选形态完整籽粒饱满的种子样品,用德国 Bruker 公司生产的傅立叶变换近红外光谱仪测定大豆蛋白质含量,每个单株种子扫描 3 次,平均值为该单株的蛋白质含量,每个株系蛋白质含量是 5 个单株蛋白质含量的平均值。利用 OPUS 软件中大豆蛋白质干基模型分析光谱获得数据。同时用 SPSS 26 软件对亲本及 RIL 群体表型数据进行描述性统计,包括性状在环境下的变异系数、平均值、偏度、峰度和正态分布检验等。

1.3 DNA 提取及 SSR 标记的选择

用经典的 CTAB 法^[19]提取亲本以及 F_{2:16}、F_{2:17} RIL 群体的 DNA。选择初定位区间 *qPRO-19-1* 内的 22 对 SSR 标记(表 1),用于筛选亲本之间的多态性标记。PCR 总反应体系为 25 μ L,包括 2 \times Rapid Taq Master Mix 12.5 μ L、ddH₂O 9.5 μ L、引物(10 μ mol) 1.0 μ L 和 DNA(70 ng/ μ L) 1.0 μ L。反应程序为 95 $^{\circ}$ C 5 min, 95 $^{\circ}$ C 15 s, 60 $^{\circ}$ C 15 s, 72 $^{\circ}$ C 15 s, 34 个循环, 72 $^{\circ}$ C 5 min, 4 $^{\circ}$ C 保存。

1.4 Indel 标记开发与鉴定

前期在第 19 号染色体定位到 1 个蛋白质含量新位点 *qPRO-19-1*,对区间内基因进行扩增。用 Phytozome (<https://phytozome-next.jgi.doe.gov/>) 在线网站获得基因序列,用 Primer3 软件设计特异性引物扩增基因,将合格扩增产物送至中国农业科学院作物科学研究所重大工程楼一楼测序公司。

表 1 鉴定 *qPRO-19-1* 所用 SSR 标记Table 1 Identification of SSR markers for *qPRO-19-1*

引物名称 Primer name	正向引物 Forward primer (5' -3')	反向引物 Reverse primer (5' -3')
SSR_19_1438	TCGATTATGCCATGTTGCAC	CAATTGTACCCACGTCTTGTTT
SSR_19_1439	GGATTTTTTCGATGTCGCAGT	GCAAAACACGTCGTTAATTAGG
SSR_19_1440	GGCTCAGAAATTTGGAGAAGG	TAGTGCATTAGGCTCCCACC
SSR_19_1441	CGCTACCCTGCTTCTTCACT	AAATTTTTCCTGCCGTGTTG
SSR_19_1442	TCCTTCAAACATTGAAGCAAA	AATTTAGTCCGCTTTAATATGTTTCT
SSR_19_1443	ATATGGATACCCCTCCCGAC	AACTCTTTTGGCTTCTCGATT
SSR_19_1444	AAATAGAGCACAAGTGTGCGCA	AATAAAAGGTGCCCATGCAG
SSR_19_1445	TGAACAAAAATAAAATGCAATCAA	AAAATAACCTCTCATCAACCATCA
SSR_19_1446	TGATGGTTGATGAGAGGTTATTTT	TGGGAGAATCAAATTACAATAATACA
SSR_19_1447	CACAATATAATTGAGAGACACTTTCAT	CCCAAGTTTTTCATTGTCTCAA
SSR_19_1448	TCGATGTGGACGACAGATT	TCTGCAGCACAACTCATCA
SSR_19_1449	GGCACGCTTTTATCCATGTT	AGATCTAAGAGCGCCACCAA
SSR_19_1450	GCTCGTCCGTACACTTCACA	CAACAAACCAGAAACGGTCC
SSR_19_1451	TCCACAAAAGTCTACATATTGAAAAA	GACGCAATAATTTGAGCGTGT
SSR_19_1452	TGACACATTCTGAAACGGATG	GTAGCATTAAATTAAGGCAAAAGA
SSR_19_1453	TCCGTAAAAGCTGCATGTTG	ACCCCTTCAAGCTCACAGAA
SSR_19_1454	CAGAGGCGGACTTAGAATGG	TTCGTTGACTCGGTATTCCA
SSR_19_1455	TAAGTCTCATAGGTCGGCGT	TGGAGTACGAGGAACGTAATAGG
SSR_19_1456	TATGGCCCCGAAAATAACGAA	CGCATATGACAAGGAAGCAA
SSR_19_1457	GACGTTAATGCCGAGTTTT	TTTTGATTATACCAAACCTTAACAAA
SSR_19_1458	TGCAAGGTGGTCATTTGAAG	TCTCACGAAACCAACAGTCAA
SSR_19_1459	TCCAACCCTAATCTGTCCTGTT	GAGAAGGTTTTGCTACGCCA

基因测序结果表明,基因 *Glyma.19g223300* 在亲本间检测到 1 个 InDel 变异位点,根据变异位点设计包含 InDel 的标记 IN-1 的引物。PCR 总反应体系为 25 μ L,包括 2 \times Rapid Taq Master Mix 12.5 μ L、ddH₂O 9.5 μ L、上下游引物 (10 μ mol) 各 1 μ L 和 DNA (70 ng/ μ L) 1 μ L。反应程序:95 $^{\circ}$ C 3 min, 95 $^{\circ}$ C 15 s, 58 $^{\circ}$ C 15 s, 72 $^{\circ}$ C 15 s, 35 个循环;72 $^{\circ}$ C 5 min, 4 $^{\circ}$ C 保存。用 1.5% 琼脂糖凝胶电泳检测标记 IN-1 的扩增产物。

1.5 QTL 定位及候选基因预测

用 QTL ICIMapping4.1 软件对蛋白质含量进行 QTL 定位,采用的方法是完备区间作图法 (ICIM-ADD),经 1000 次排列确定 LOD 阈值,当实际得出的 LOD 值大于 LOD 阈值时,就认为存在 1 个 QTL。根据定位区间物理位置查询 SoyBase (<https://www.soybase.org/>) 获得基因注释。

1.6 荧光定量 PCR 分析

实时荧光定量 PCR 分子实验采用 PerfectStart® Green qPCR SuperMix (+Dye I) 试剂盒,用 ABI PRISM 7300 荧光定量仪器进行 qRT-PCR 扩增。

PCR 总反应体系为 20 μ L,包括 PerfectStart® Green qPCR SuperMix (+Dye I) 10 μ L、ddH₂O 7.2 μ L、上下游引物 (10 μ mol) 各 0.4 μ L 和 cDNA 2 μ L。qRT-PCR 反应程序如下:95 $^{\circ}$ C 30 s, 94 $^{\circ}$ C 5 s, 58 $^{\circ}$ C 15 s, 72 $^{\circ}$ C 10 s, 40 个循环;72 $^{\circ}$ C 31 s, 4 $^{\circ}$ C 保存。使用 *GmActin* 作为内参基因 (F: 5-CGGTGGTTCTATCTTGGCATC-3; R: 5-GTCTTTTCGCTTCAATAACCCTA-3), 计算候选基因相对表达水平。

1.7 数据统计分析

利用 Microsoft Excel 2016 和 SPSS 26 分析 RIL 群体蛋白含量最大值、最小值、平均数、标准差、偏度、峰度和变异系数等,利用 GraphPad 软件分析定位区间内 InDel 标记 IN-1 各个基因型对应的蛋白质含量分布。以中黄 35 和十胜长叶生殖时期的 R3 (始荚期)、R4 (盛荚期)、R5 (始粒期)、R6 (鼓粒期) 和 R8 (成熟期) 为材料,使用 *GmActin* 作为内参基因 (F: 5-CGGTGGTTCTATCTTGGCATC-3; R: 5-GTCTTTTCGCTTCAATAACCCTA-3), 用候选基因相对定量 ($2^{-\Delta\Delta CT}$) 方法计算基因的相对表达量。

2 结果与分析

2.1 亲本及 RIL 群体蛋白质含量分布

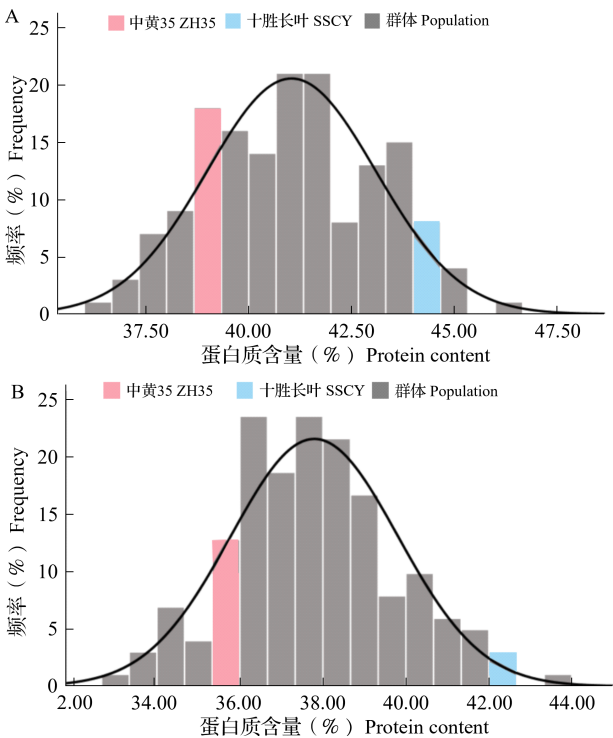
在 2020 年,中黄 35 平均蛋白质含量为 38.86%,十胜长叶平均蛋白质含量为 44.70%;在 2021 年,中黄 35 平均蛋白质含量为 35.59%,十胜长叶的蛋白质含量为 42.70%。亲本之间蛋白质含量差异较大,十胜长叶的蛋白含量高于中黄 35(表 2)。F_{2:16} 群体蛋白质含量的变异范围在 36.37%~46.64% 之间,

平均值为 41.04%,变异系数为 5.00%;F_{2:17} 群体蛋白质含量的变异范围在 33.25%~43.82% 之间,平均值为 37.79%,变异系数为 5.37%。其中,F_{2:16} 和 F_{2:17} 群体分别有 193 个和 173 个株系,群体蛋白质含量均呈连续性变异(图 1)。用 R 语言对 F_{2:16} 和 F_{2:17} RIL 群体蛋白质含量进行相关性分析,相关性系数为 0.27,*P*<0.01,有极显著相关性(图 2)。由表 2 可知,中黄 35 和十胜长叶蛋白质含量差异较大,2 个环境下十胜长叶的蛋白质含量显著高于中黄 35。

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 generation parents and recombinant inbred line populations

年份 / 环境 Year/Environment	中黄 35	十胜长叶	重组自交系			
	平均数 ± 标准差 (%)	平均数 ± 标准差 (%)	Recombinant inbred lines			
	ZH35 Mean ± SD	SSCY Mean ± SD	平均数 ± 标准差 (%) Mean ± SD	变异系数 (%) CV	峰度 Kurtosis	偏度 Skewness
2020 海南 2020 Hainan	38.86 ± 0.91	44.70 ± 0.56	41.04 ± 0.16	5.00	-0.627	0.119
2021 北京 2021 Beijing	35.59 ± 0.40	42.07 ± 0.51	37.79 ± 0.15	5.37	-0.092	0.243

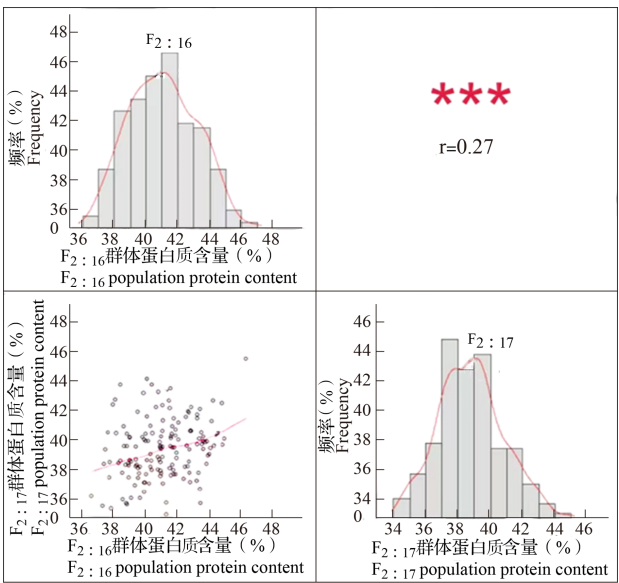
ZH35: 中黄 35; SSCY: 十胜长叶; 下同
ZH35: Zhonghuang 35; SSCY: Tokachi nagaha; the same as below



A: F_{2:16} 群体蛋白质含量分布图; B: F_{2:17} 群体蛋白质含量分布图
A: Distribution of protein content of F_{2:16} population;
B: Distribution of protein content of F_{2:17} population

图 1 ZH35/SSCY 重组自交系及亲本的蛋白质含量表型分布特征

Fig.1 Distribution of protein content phenotype of ZH35/ZH13 RIL population and parents



*** 表示 *P*<0.01 水平上差异显著
*** means significant at the *P*<0.01 level
图 2 F_{2:16} 与 F_{2:17} RIL 群体蛋白质含量相关性分析
Fig.2 Correlation analysis of protein content between F_{2:16} and F_{2:17} RIL populations

2.2 Indel 标记开发与鉴定

武阳春等^[18]利用中黄 35 和十胜长叶杂交构建 RIL 群体在第 19 号染色体上定位到 1 个蛋白质含量相关的新位点(*qPRO-19-1*),区间内共

有 36 个基因。对区间内基因进行扩增,发现基因 *Glyma.19g223300* 存在 1 个 InDel 变异位点,中黄 35 相对于十胜长叶增加了 16 个碱基,中黄 35 PCR 产物长度为 169 bp,十胜长叶长度为 153 bp。用 Phytozome (<https://phytozome-next.jgi.doe.gov/>) 下载含 InDel 变异位点的序列,用 Primer3 软件设计 InDel 标记 IN-1 的特异性引物(表 3)。此外,用 1.5% 琼脂糖凝胶电泳鉴定标记 IN-1 在中黄 35 与十胜长叶之间的特异性(图 3),可以满足定位需求。

表 3 InDel 标记引物序列及亲本 PCR 产物长度

Table3 InDel marker primer sequence and length of parental PCR product

引物名称 Primer name	正向引物 Forward primer (5' -3')	反向引物 Reverse primer (5' -3')	片段长度 (bp) Product size	
			中黄 35 ZH35	十胜长叶 SSCY
IN-1	ACTTAAGGAGTTATAGATAATTTC	AAAGTGGTCTATAAAAAATCTGCTCT	169	153

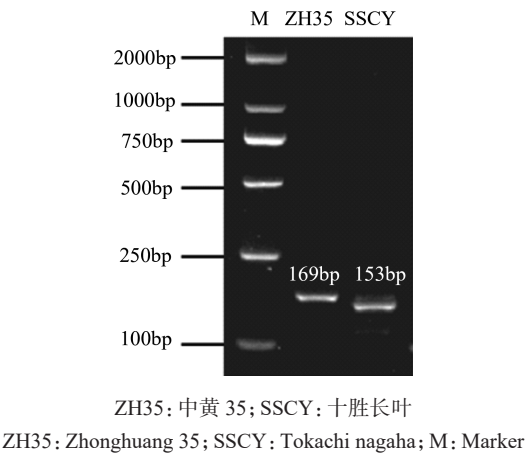


图 3 InDel 标记 IN-1 鉴定 ZH35 和 SSCY 胶图
Fig.3 InDel marker IN-1 to identify ZH35 and SSCY glue maps

2.3 大豆蛋白质含量相关 QTL 的精细定位

前期定位的蛋白质含量新位点 *qPRO-19-1*, 介于标记 BARCSOYSSR_19_1438 和 BARCSOYSSR_19_1459 之间, 区间大小为 384 kb。为进一步分析新位点 *qPRO-19-1*, 选择区间内开发的 Indel 标记 IN-1 和多态性 SSR 标记进行定位。用开发的标记 IN-1 鉴定 $F_{2:16}$ 与 $F_{2:17}$ 群体基因型, 同时将 *qPRO-19-1* 区间内的 22 对 SSR 标记分析亲本间多态性, 共检测到 8 个多态性 SSR 标记, 分别是 SSR_19_1442、SSR_19_1443、SSR_19_1447、SSR_19_1449、SSR_19_1450、SSR_19_1451、SSR_19_1455 和 SSR_19_1459。从 8 个

多态性 SSR 标记中选择物理位置分布均匀且条带清晰的 5 个 SSR 标记鉴定 $F_{2:16}$ 与 $F_{2:17}$ 群体基因型。SSR 标记和 InDel 标记的基因型数据结合蛋白质含量表型数据, 用 ICIMapping4.1 软件定位大豆蛋白质含量相关 QTL, 获得了 1 个稳定的蛋白质含量 QTL 位点(图 4), *qPRO-19-1* 区间大小为 68.03 kb, 位于标记 SSR_19_1443 到 SSR_19_1447 之间, 区间内包含 4 个基因。其中, $F_{2:16}$ 群体 LOD 值为 2.13(表 4), 可解释的表型变异率为 6.23%, 加性效应为 -0.53。 $F_{2:17}$ 群体 LOD 值为 3.96, 可解释的表型变异率为 10.45%, 加性效应为 -0.66。

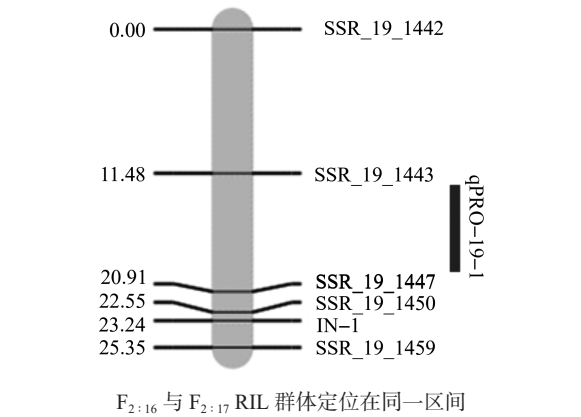


图 4 RIL 群体在 19 号染色体定位蛋白质含量 QTL
Fig.4 RIL populations map protein content QTLs on chromosome 19

表 4 完备区间作图法定位 $F_{2:16}$ 和 $F_{2:17}$ 群体蛋白质含量 QTL 的染色体位置和参数

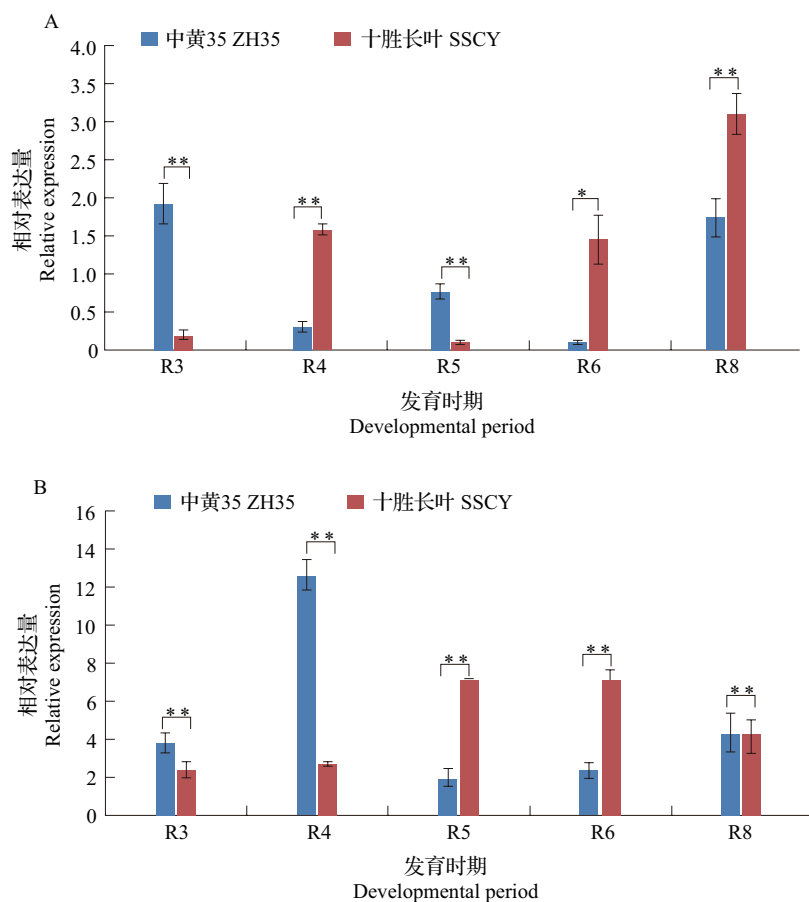
Table4 Chromosome locations and parameters of QTL for protein content of ICIM model $F_{2:16}$ and $F_{2:17}$ populations

群体 Population	左侧标记 Left marker	右侧标记 Right marker	LOD 值 LOD value	表型变异率 (%) PVE	加性效应 Additive effect
$F_{2:16}$	SSR_19_1443	SSR_19_1447	2.13	6.23	-0.53
$F_{2:17}$	SSR_19_1443	SSR_19_1447	3.96	10.45	-0.66

2.4 大豆蛋白质定位区间候选基因分析

为进一步确定候选基因,用 qRT-PCR 技术检测定位区间内基因相对表达量,分析其在中黄 35 和十胜长叶大豆种子生长发育时期鲜重 R3 (27~42 mg)、R4 (95.1~114.9 mg)、R5 (200~250 mg)、R6 (250.1~300 mg) 和 R8 (>350 mg) 5 个时期的表达差异 (图 5)。在亲本种子各个发育时期均未检测到 *Glyma.19g221900* 和 *Glyma.19g222100* 表达;虽然 *Glyma.19g221800* 在亲本种子各个发育时期有

表达,但相对表达量都比较低,其中十胜长叶在 R3 和 R5 非连续低表达,分析可能有两个原因,第一,样品来自于同一株系,但对蛋白质的积累能力不完全一致,第二,大豆蛋白质含量是一个复杂的数量性状,受种植环境影响较大。基因 *Glyma.19g222000* 在中黄 35 和十胜长叶种子的 R3、R4、R6 和 R8 之间的表达量都呈极显著差异,其中中黄 35 在 R3 和 R4 持续高表达,可能在这个时期种子蛋白质开始积累。



A 和 B 分别为基因 *Glyma.19g221800* 和 *Glyma.19g222000* 的相对表达量,黑色方差条表示平均值的标准误差,

* 代表在 $P \leq 0.05$ 水平具有显著性差异, ** 代表在 $P \leq 0.01$ 水平具有显著性差异

A and B are the relative expression levels of the genes *Gene1* and *Gene3*, respectively. The black variance bar represents the standard error (SEM) of the mean value., * correlation is significant at the $P \leq 0.05$ level, ** Correlation is significant at the $P \leq 0.01$ level

图 5 定位区间内候选基因在种子不同发育时期的相对表达量

Fig.5 Relative expression levels of candidate genes in the mapping interval at different developmental stages of seeds

3 讨论

3.1 大豆蛋白质含量的影响因素

大豆蛋白质属于复杂的数量性状,受多基因控制,同时种植环境也对蛋白表型有很大影响,比如光照强度^[20]、种植密度^[21]、地理纬度^[22]、大气 CO₂

浓度^[23]和播期^[24]等,这些环境因素对大豆蛋白表型的精准鉴定有一定影响。在环境和基因的交互作用下,高蛋白大豆育种更具有挑战性^[25]。据北京气象局统计,2021 年 (6-8 月) 平均降水量多达 627.4 mm,分别较常年同期和近十年同期多 70% 和 60%,为近 20 年降雨量最多,其中汛期 63.8% 的

降雨集中于7月,8月降水量虽趋于减少,但仍比常年同期偏多20%,同时降雨过程偶尔伴有冰雹或6级以上短时大风天气。鉴定2021年RIL群体的表型时,大豆籽粒蛋白质含量随着降雨量增多而降低,这种现象在王树宇等^[26]的研究中被提及,研究结果表明降雨量与蛋白质含量极显著负相关,不利于籽粒干物质的积累。在本研究中,由于两年种植地生态区域有较大的差异,亲本及RIL群体蛋白质含量有一定的差异,但 $F_{2:16}$ 和 $F_{2:17}$ RIL群体蛋白质含量相关性系数为0.27, $P < 0.01$,具有极显著相关性。

3.2 大豆高蛋白质 QTL 定位新位点

据 Soybase 数据库 (<https://www.soybase.org>) 统计,与大豆蛋白质含量相关的 QTL 有 241 个,第 19 号染色体上的蛋白含量 QTL 位点数目较少且都相互独立。其中, Lu 等^[15]定位的蛋白质含量 QTL 最大,区间大小为 19.80 Mb,可解释的表型变异率为 5.9%。Diers 等^[27]定位的蛋白质含量 QTL 最小,区间大小为 482 kb,可解释的表型变异率为 5.90%。Tajuddin 等^[28]在两个环境中检测到 1 个蛋白质含量 QTL 位点,大小为 2.0 Mb。Mao 等^[29]在两个不同的环境下,定位到两个蛋白质含量 QTL 位点,区间大小相差约 3.7 Mb。Jun 等^[14]利用关联分析对大豆蛋白质含量进行定位,在第 19 号染色体鉴定到 1 个 QTL,可解释的表型变异率 16%,与 Diers 等^[27]报道的位点一致。本研究将蛋白质含量 QTL (*qPRO19-1*) 位点由 384 kb 缩小至 68.03 kb,是目前第 19 号染色体定位区间最小的 QTL,小于 Mansur 等^[30]和 Jun 等^[14]报道的 QTL 区间大小,同时也是 1 个新的蛋白质含量 QTL。

3.3 候选基因预测

根据 SoyBase (<https://www.soybase.org/>) 获得 4 个基因的功能注释(表 5),基因 *Glyma.19g221800* 是几丁质酶家族蛋白,几丁质酶有广泛的生物学作用,参与共生固氮作用,调控植物生长发育,并且在植物保护方面有重要应用价值^[31]。其中,几丁质酶可以使固氮菌的结瘤因子失活降解,并且氮有利于大豆蛋白质含量的增加^[32-36]。基因 *Glyma.19g221900* 参与生长素激活的信号通路,调节植物根的生长发育。基因 *Glyma.19g222000* 属于 Hairy/E (spl) 家族的转录抑制子,参与调控光敏色素转录过程,对植物的生长发育有非常重要的作用。基因 *Glyma.19g222100* 属于 SNF7 蛋白质家族,主要参与细胞内蛋白质的运输,可能会影响

大豆蛋白质含量。推测基因 *Glyma.19g221800* 和 *Glyma.19g222100* 可能与大豆蛋白质含量相关,但是候选基因的选择还有待于下一步的鉴定和功能验证。

表 5 Soybase 网站目标区间内基因功能注释信息

Table 5 Gene function annotation information in the positioning interval on the soybase website

基因 Gene	功能注释 Feature notes
<i>Glyma.19g221800</i>	几丁质酶代谢过程、碳水化合物代谢过程、氨基酸转运、硝酸盐运输
<i>Glyma.19g221900</i>	调控转录、调节根生长发育、蛋白结合
<i>Glyma.19g222000</i>	调控转录、光敏色素相互作用因子
<i>Glyma.19g222100</i>	细胞内蛋白质运输

4 结论

利用中黄 35 和十胜长叶构建的 $F_{2:16}$ 和 $F_{2:17}$ RIL 群体进行定位。用开发的 InDel 标记 IN-1 和区间内多态性 SSR 标记将大豆蛋白质含量 QTL 位点 *qPRO-19-1* 由 384 kb 缩小到 68.03 kb,位于分子标记 SSR_19_1443 到 SSR_19_1447 之间,包括 4 个基因,在 $F_{2:16}$ 和 $F_{2:17}$ RIL 群体分别解释了 6.23% 和 10.45% 的遗传变异,加性效应均为负值,说明定位高蛋白位点来自于十胜长叶。本研究结果为大豆高蛋白质的基因克隆及分子标记辅助育种创造了条件。

参考文献

- [1] Zhang Y H, Liu M F, He J B, Wang Y F, Xing G N, Li Y, Yang S P, Zhao T J, Gai J Y. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) merr.]. Theoretical and Applied Genetics, 2015, 128(6): 1061-1072
- [2] 王军,张海生,李方舟,古晓红,杨婷婷. 大豆蛋白质的开发与利用. 农产品加工, 2020, 5: 11-13
Wang J, Zhang H S, Li F Z, Gu X H, Yang T T. Development and utilization of soybean protein. Farm Products Processing, 2020, 5: 11-13
- [3] 李琼,刘强,杨青春,舒文涛,李金花,常世豪,张东辉,张保亮,张来成,耿臻. 大豆高蛋白基因分子标记辅助育种的应用. 山西农业科学, 2020, 48(8): 1192-1197
Li Q, Liu Q, Yang Q C, Su W T, Li J H, Chang S H, Zhang D H, Zhang B L, Zhang L C, Geng Z. Application of molecular marker assisted breeding of soybean high protein gene. Journal of Shanxi Agricultural Sciences, 2020, 48(8): 1192-1197
- [4] Helms T C, Orf J H. Protein, oil, and yield of soybean lines selected for increased protein. Crop Science, 1998, 38(3): 707-711
- [5] Cober E R, Voldeng H D. Developing high-protein, high-yield

- soybean populations and lines. *Crop Science*, 2000, 40(1): 39-42
- [6] Panthee D R, Pantalone V R, West D R, Saxton A M, Sams C E. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Science*, 2005, 45(5): 2015-2022
- [7] 魏荷, 王金社, 卢为国. 大豆籽粒蛋白质含量分子遗传研究进展. *中国油料作物学报*, 2015, 37(3): 394-400
- Wei H, Wang J S, Lu W G. Molecular genetic advances in soybean seed protein. *Chinese Journal of Oil Crop Sciences*, 2015, 37(3): 394-400
- [8] Csanadi G, Vollmann J, Stift G, Jelle T. Seed quality QTLs identified in a molecular map of early maturing soybean. *Theoretical and Applied Genetics*, 2001, 103(6): 912-919
- [9] Warrington C V, Abdel-Haleem H, Hyten D L, Cregan P B, Orf J H, Killam A S, Bajjalieh N, Li Z, Boerma H R. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theoretical and Applied Genetics*, 2015, 128(5): 839-850
- [10] Pandurangan S, Pajak A, Molnar S J, Cober E R, Dhaubhadel S, Hernández-Sebastià C, Kaiser W M, Nelson R L, Huber S C, Marsolais F. Relationship between asparagine metabolism and protein concentration in soybean seed. *Journal of Experimental Botany*, 2012, 63(8): 3173-3184
- [11] Zhang W K, Wang Y J, Luo G Z, Zhang J S, He C Y, Wu X L, Gai J Y, Chen S Y. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theoretical and Applied Genetics*, 2004, 108(6): 1131-1139
- [12] 陈庆山, 张忠臣, 刘春燕, 王伟权, 李文滨. 应用 Charleston × 东农 594 重组自交系群体构建 SSR 大豆遗传图谱. *中国农业科学*, 2005, 38(7): 1312-1316
- Chen Q S, Zhang Z C, Liu C Y, Wang W Q, Li W B. Construction and analysis of soybean genetic map using recombinant inbred line of Charleston × Dongnong594. *Scientia Agricultura Sinica*, 2005, 38(7): 1312-1316
- [13] Li X Y, Wang P, Zhang K X, Liu S L, Qi Z Y, Fang Y L, Wang Y, Tian X C, Song J, Wang J J, Yang C, Sun X, Tian Z X, Li W X, Ning H L. Fine mapping QTL and mining genes for protein content in soybean by the combination of linkage and association analysis. *Theoretical and Applied Genetics*, 2021, 134(4): 1-28
- [14] Jun T H, Van K J, Kim M Y, Lee S H, Walker D R. Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica*, 2008, 162(2): 179-191
- [15] Lu W, Wen Z, Li H, 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
- [16] Fliege C, Ward R A, Vogel P, Nguyen H, Quach T, Guo M, Vianna J P J, Santos L B D, Specht J E, Clemente T E. Fine mapping and cloning of the major seed protein QTL on soybean chromosome 20. *The Plant Journal*, 2022, 110(1): 114-128
- [17] Wang S D, Liu S L, Wang J, Yokosho K, Zhou B, Yu Y C, Liu Z, Frommer W B, Ma J F, Chen L Q, Guan Y F, Shou H X, Tian Z X. Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *National Science Review*, 2020, 7(11): 1776-1786
- [18] 武阳春, 郭兵福, 谷勇哲, 栾晓燕, 邱红梅, 刘鑫磊, 李海燕, 邱丽娟. 大豆蛋白含量新位点 *qpro-19-1* 的定位. *植物遗传资源学报*, 2021, 22(1): 139-148
- Wu Y C, Guo B F, Gu Y Z, Luan X Y, Qiu H M, Liu X L, Li H Y, Qiu L J. Mapping of a new quantitative locus *qPRO-19-1* associated with seed crude protein content in soybean (*Glycine max* L.). *Journal of Plant Genetic Resources*, 2021, 22(1): 139-148
- [19] Doyle J, Doyle J L. Isolation of plant DNA from fresh tissue. *Focus*, 1990, 12(13): 39-40
- [20] 胡国华, 宁海龙, 王寒冬, 王继安, 张大勇, 李文滨. 光照强度对大豆产量及品质的影响 I. 全生育期光照强度变化对大豆脂肪和蛋白质含量的影响. *中国油料作物学报*, 2004, 26(2): 87-89
- Hu G H, Ning H L, Wang H D, Wang J A, Zhang D Y, Li W B. Effect of photo-intensity on quality and yield of soybeans I. Effect of light-intensity on oil content and protein content of soybeans in the whole growth period. *Chinese Journal of Oil Crop Sciences*, 2004, 26(2): 87-89
- [21] 宋旭, 李志刚, 马日亮. 不同栽培密度对大豆产量、脂肪和蛋白质含量的影响. *内蒙古农业科技*, 2009, 2: 29-30
- Song X, Li Z G, Ma R L. Effect of different planting density on yield, fat and protein content of soybean. *Journal of Northern Agriculture*, 2009, 2: 29-30
- [22] 闫晓艳, 邱强, 刘凤珍, 石一鸣, 唐晓博. 不同类型大豆品种鼓粒至成熟期脂肪和蛋白质相对积累规律初探. *吉林农业大学学报*, 2006, 28(5): 473-477
- Yan X Y, Qiu Q, Liu F Z, Shi Y M, Tang X B. An initial study on relative accumulative rule of oil and protein contained in different soybean varieties from beginning seed to maturity. *Journal of Jilin Agricultural University*, 2006, 28(5): 473-477
- [23] 蒋跃林, 张仕定, 岳伟, 姚玉刚, 张庆国. 大气 CO₂ 浓度增加对大豆籽粒品质的影响. *中国粮油学报*, 2005(5): 89-92
- Jiang Y L, Zhang S D, Yue W, Yao Y G, Zhang Q G. Effects of free CO₂ enrichment on grain quality of soybean. *Journal of the Chinese Cereals and Oils Association*, 2005(5): 89-92
- [24] 杜佳兴, 刘婧琦, 盖志佳, 蔡丽君, 顿国强, 张敬涛. 播期对高蛋白大豆籽粒品质及产量的影响. *中国农学通报*, 2019, 35(35): 31-34
- Du J X, Liu J Q, Gai Z J, Cai L J, Dun G Q, Zhang J T. Sowing dates affect seed quality and yield of high protein soybean. *Chinese Agricultural Science Bulletin*, 2019, 35(35): 31-34
- [25] 王鹏, 侯思宇, 武艳杏, 李贵全. 基于农艺性状的大豆种质资源多样性分析及评价. *山西农业科学*, 2021, 49(9): 1025-1030
- Wang P, Hou S Y, Wu Y X, Li G Q. Diversity analysis and comprehensive evaluation of soybean germplasm resources based on agronomic traits. *Journal of Shanxi Agricultural Sciences*, 2021, 49(9): 1025-1030

(下转 1736 页)

- Breeding, 2017, 15 (2): 763-773
- [23] 张桂权. 基于 SSSL 文库的水稻设计育种平台. 遗传, 2019, 41 (8): 754-760
Zhang G Q. The platform of breeding by design based on the SSSL library in rice. Hereditas, 2019, 41 (8): 754-760
- [24] 杨梯丰. 基于 SSSL 的水稻耐冷 QTL 的定位和聚合效应. 广州: 华南农业大学, 2016
Yang T F. Identification and pyramiding of QTLs for cold tolerance by use of single segment substitution lines in rice (*Oryza sativa* L.). Guangzhou: South China Agricultural University, 2016
- [25] Li F P, Gao Y H, Wu B Q, Cai Q P, Zhan P L, Yang W F, Shi W X, Li X H, Yang Z F, Tan Q Y, Luan X, Zhang G Q, Wang S K. High-Quality de novo genome assembly of Huajingxian 74, a receptor parent of single segment substitution lines. Rice Science, 2021, 28 (2): 109-113
- [26] Eshed Y, Zamir D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics, 1995, 141 (3): 1147-1162
- [27] Wissuwa M, Wegner J, Ae N, Yano M. Substitution mapping of Pup1: A major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. Theoretical and Applied Genetics, 2002, 105 (6-7): 890-897
- [28] 杨梯丰, 曾瑞珍, 朱海涛, 陈岚, 张泽民, 丁效华, 李文涛, 张桂权. 水稻粒长基因 GS3 在聚合育种中的效应. 分子植物育种, 2010, 8 (1): 59-66
Yang T F, Zeng R Z, Zhu H T, Chen L, Zhang Z M, Ding X H, Li W T, Zhang G Q. Effect of grain length gene GS3 in pyramiding breeding of rice. Molecular Plant Breeding, 2010, 8 (1): 59-66
- [29] 赵芳明, 张桂权, 曾瑞珍, 杨正林, 凌英华, 桑贤春, 何光华. 利用单片段代换系研究水稻产量相关性状 QTL 加性及上位性效应. 作物学报, 2012, 38 (11): 2007-2014
Zhao F M, Zhang G Q, Zeng R Z, Yang Z L, Ling Y H, Sang X C, He G H. Epistatic and additive effects of QTLs for yield-related traits using single segment substitution lines of rice (*Oryza sativa* L.). Acta Agronomica Sinica, 2012, 38 (11): 2007-2014
- [30] Shirasawa S, Endo T, Nakagomi K, Yamaguchi M, Nishio T. Delimitation of a QTL region controlling cold tolerance at booting stage of a cultivar, 'Lijiangxintuanheigu', in rice (*Oryza sativa* L.). Theoretical and Applied Genetics, 2012, 124 (5): 937-946
- [31] Cui D, Xu C Y, Tang C F, Yang C G, Yu T Q, A X X, Cao G L, Xu F R, Zhang J G, Han L Z. Genetic structure and association mapping of cold tolerance in improved *japonica* rice germplasm at the booting stage. Euphytica, 2013, 193: 369-382
- [32] Fujino K, Obara M, Shimizu T, Koyanagi Kanako O, Ikegaya T. Genome-wide association mapping focusing on a rice population derived from rice breeding programs in a region. Breeding Science, 2015, 65: 403-410
- [33] Ulzii Bat B, Ohta H, Fukushima A, Shirasawa S, Kitashiba H, Nishio T. Examination of candidates for the gene of cold tolerance at the booting stage in a delimited QTL region in rice cultivar 'Lijiangxintuanheigu'. Euphytica, 2016, 211: 331-341
- [34] Members of the Complex Trait Consortium. The nature and identification of quantitative trait loci: A community's view. Nature Reviews Genetics, 2003, 4 (11): 911-916
- [35] Tanksley S D. Mapping polygenes. Annual Review of Genetics, 1993, 27: 205-233

(上接 1725 页)

- [26] 王树宇, 付连舜, 刘德恒, 董友魁. 积温、降水量和日照对大豆主要农艺性状和品质的影响. 农业开发与装备, 2013 (5): 42-43
Wang S Y, Fu L S, Liu D H, Dong Y K. Effects of accumulated temperature, precipitation and sunshine on main agronomic characters and quality of soybean. Agricultural Development and Equipments, 2013 (5): 42-43
- [27] 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
- [28] Tajuddin T, Watanabe S, Masuda R, Harada K, Kawano S. 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 (4): 315-325
- [29] 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
- [30] Mansur L M, Orf J H, Chase K, Jarvik T, Cregan P B, Lark K G. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Science, 1996, 36 (5): 1327-1336
- [31] 欧阳石文, 赵开军, 冯兰香, 谢丙炎. 植物几丁质酶的研究进展. 生物工程进展, 2001 (4): 30-34
Ouyang S W, Zhao K J, Feng L X, Xie B Y. Advances in research of plant chitinases. China Biotechnology, 2001 (4): 30-34
- [32] Minic Z, Leproust-Lecoester L, Laporte J, Kouchkovsky Y D, Brown S C. Proteins isolated from lucerne roots by affinity chromatography with sugars analogous to nod factor moieties. The Biochemical Journal, 2000, 345 (2): 255-262
- [33] 高凤菊. 影响大豆蛋白质形成和积累的因素研究现状. 大豆科技, 2014 (3): 14-19
Gao F J. Progress on factors influencing protein formation and accumulation in soybean. Soybean Science and Technology, 2014 (3): 14-19
- [34] 孙智, 郝秀梅. 大豆籽粒脂肪及蛋白质含量与营养条件的关系. 内蒙古农业科技, 2004 (6): 30-31
Sun Z, Hao X M. The relationship between soybean kernel fat and protein content and nutritional conditions. Journal of Northern Agriculture, 2004 (6): 30-31
- [35] 薛红. 增施氮、磷、钾和有机肥对大豆产量、品质的影响及经济效益分析. 安徽农学通报, 2009, 15 (7): 109-110
Xue H. Effects of nitrogen, phosphorus, potassium and organic fertilizers on soybean yield and quality and economic benefit analysis. Anhui Agricultural Science Bulletin, 2009, 15 (7): 109-110
- [36] 张洪刚, 周琴, 何小红, 张国正. 播期、密度和肥料对菜用大豆南农 9610 产量和品质的影响. 江苏农业学报, 2008, 24 (5): 662-667
Zhang H G, Zhou Q, He X H, Zhang G Z. Effects of sowing date planting density and N, P and K fertilizer on yield and quality of vegetable soybean. Jiangsu Journal of Agricultural Sciences, 2008, 24 (5): 662-667