



大豆水溶性蛋白质的全基因组关联分析

沈甲诚, 张小利, 黄建丽, 邓肃霜, 郭娜, 赵晋铭, 邢邯

(南京农业大学 农学院/国家大豆改良中心/农业部大豆生物学与遗传育种重点实验室/作物遗传与种质创新国家重点实验室, 江苏 南京 210095)

摘要:为进一步解析中国大豆种质水溶性蛋白的遗传基础,为大豆高水溶性蛋白质的分子标记辅助选择育种及品质改良提供理论依据,本研究以 224 份大豆种质为试验材料,于 2017 和 2018 年对大豆水溶性蛋白质含量进行测定,利用 1 514 个高质量的 SNP 标记分别对 2017、2018 年水溶性蛋白质含量及两年均值进行全基因组关联分析,共检测到 18 个显著关联的 SNP 标记,这些 SNP 标记涉及 16 个位点,有 8 个位点至少被检测到 2 次,其余 8 个位点仅被检测到 1 次,表明其受环境因素影响较大。16 个位点中有 7 个尚未见报道,分别位于 8、11、13、14 和 15 号染色体上,是新发现的控制大豆水溶性蛋白的位点。对表型变异解释率较高且稳定关联的 2 个位点 *qWSPC7* 和 *qWSPC8-1* 候选区间内的基因进行预测,共获得 25 个候选基因,其中有 7 个基因 (*Glyma.07g195000*、*Glyma.08g103100*、*Glyma.08g108900*、*Glyma.08g105100*、*Glyma.08g107800*、*Glyma.08g107700* 和 *Glyma.08G115800*) 在大豆籽粒、根或根瘤中具有较高的表达水平。这些基因可作为水溶性蛋白质的候选基因,可能具有调控大豆水溶性蛋白质的功能。

关键词:大豆; 水溶性蛋白质; 全基因组关联分析; 候选基因挖掘

Genome Wide Association Analysis on Water-Soluble Protein in Soybean

SHEN Jia-cheng, ZHANG Xiao-li, HUANG Jian-li, DENG Su-shuang, GUO Na, ZHAO Jin-ming, XING Han

(College of Agriculture, Nanjing Agricultural University/National Center for Soybean Improvement/ Key Laboratory of Biology and Genetics and Breeding for Soybean, Ministry of Agriculture/ State Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing 210095, China)

Abstract: In order to further analyze the genetic basis of water-soluble protein in Chinese soybean collection, this study investigated the water-soluble protein content phenotype of 224 soybean collection of 2017 and 2018 years. Genome-wide association study (GWAS) of WSPC was performed in 2017, 2018 and the average across of the two years with 1 514 high-quality SNP markers. A total of 18 significantly associated SNPs were detected. These SNPs involved 16 sites, with 8 sites were stably detected in at least two times, and the remaining 8 loci were detected only in a single environment, indicating that it was greatly affected by environmental factors. Seven of the 16 loci have not been reported, located on chromosomes 8, 11, 13, 14, and 15 respectively. They are newly discovered loci that control the water-soluble protein of soybean. Prediction of genes in the two candidate regions of *qWSPC7* and *qWSPC8-1* with high phenotypic variation interpretation rates and stable associations obtained a total of 25 candidate genes, of which 7 genes (*Glyma.07g195000*, *Glyma.08g103100*, *Glyma.08g108900*, *Glyma.08g105100*, *Glyma.08g107800*, *Glyma.08g107700* and *Glyma.08G115800*) had higher expression levels in soybean seed, roots or nodules. These genes can be used as candidate genes for water-soluble proteins and may have the function of regulating soybean water-soluble proteins. This study provides theoretical basis for molecular marker-assisted selection breeding and quality improvement of soybean high water-soluble proteins.

Keywords: Soybean [*Glycine max* (L.) Merr.]; Water-soluble protein; Genome-wide association analysis; Mining for candidate gene

大豆蛋白是一种优质的植物蛋白,富含人体营养所必需的多种氨基酸,常被用作动物蛋白的主要替代品^[1]。大豆水溶性蛋白(water-soluble protein content, WSPC)是大豆蛋白的重要组成部分,具有凝

收稿日期:2019-11-27

基金项目:国家重点研发计划(2018YFD0100800, 2017YFD0101500);中央高校基本科研业务费专项资金(KYT201801);现代农业产业技术体系建设专项(CARS-004-PS10);江苏省重点研发计划(BE2019376-2);长江学者和创新团队发展计划(PCSIRT_17R55);仲英作物种业创新中心;江苏省现代作物生产协同创新中心。

第一作者简介:沈甲诚(1993-),男,硕士,主要从事大豆遗传育种研究。E-mail:2017101159@njau.edu.cn。

通讯作者:赵晋铭(1978-),男,博士,副教授,主要从事大豆遗传育种研究。E-mail:jzm3000@126.com;

邢邯(1963-),男,博士,教授,主要从事大豆育种与大豆分子遗传研究。E-mail:hanx@njau.edu.cn。

胶性、乳化性、发泡性等加工特性,其含量对大豆食品的加工和利用具有重要作用^[2]。前人研究表明,大豆水溶性蛋白质含量具有广泛的遗传变异,一般为10%~45%^[3,4],表明大豆水溶性蛋白具有较大的遗传改良潜力。提高大豆水溶性蛋白质含量对于大豆蛋白食品加工及品质的改良具有十分重要的意义。

大豆水溶性蛋白质含量属于多基因控制的数量性状,且易受环境因素的影响^[3]。利用双亲作图群体对该性状 QTL 定位研究的报道较少, Lu 等^[4]利用水溶性蛋白质含量较高的品种 ZDD09454 和水溶性蛋白质含量较低的品种豫豆 12 杂交衍生的 212 份重组自交系群体在 2 年 3 点 6 个环境下共检测到 11 个与大豆水溶性蛋白质相关的 QTL,表型变异解释率范围为 4.5%~18.2%,其中位于 8 号染色体上的 2 个 QTL(*qsp8-4* 和 *qsp8-5*) 在多个环境下被稳定检测到,表型变异解释率的范围为 8.1%~18.2%。陈强等^[5]以冀豆 12 和黑豆(ZDD03651)杂交衍生的 188 份重组自交系群体为试验材料,采用双尾法在 3 个不同选择压力下对大豆水溶性蛋白质进行 QTL 定位,共检测到 23 个 QTL,其中在 10% 选择压力下稳定检测到 2 个与水溶性蛋白质相关的 QTL,表型变异解释率为 7.98%~9.78%。

QTL 定位方法的发展加快了数量性状相关位点的发掘,但是传统 QTL 定位往往需要构建作图群体,耗时较长且定位精度不高^[6-7],而基于自然群体的关联分析具有花费时间短,定位精度高,能够同时检测多个等位基因的优点^[8]。已被广泛应用于玉米^[9-10]、水稻^[11-12]、大豆^[13-14]等作物数量性状遗传基础的解析当中。目前,高通量测序技术发展以及大豆参考基因组序列的公布^[15]极大地推动了大豆基因组学的发展,关联分析已经成为大豆数量性状遗传解析及相关基因挖掘的有效工具。Zhang 等^[3]利用包含 192 份大豆种质的自然群体基于 1 536 个高质量 SNP 标记对 4 个环境下的大豆水溶性蛋白质进行关联分析,共检测到 8 个与水溶性蛋白质显著关联的 SNP 标记,表型变异解释率的范围为 8.5%~15.1%,其中位于 8 号染色体上的 1 个谷蛋白 B2 类基因与显著关联的 SNP 标记 BARC-041663-08059 距离较近,可能与大豆水溶性蛋白质相关。Zhang 等^[16]利用包含 219 份大豆种质的自然群体和 152 份重组自交系群体,基于关联分析和连锁分析共检测到 16 个与水溶性蛋白质相关的 QTL,

分别位于除大豆 2、3、4、5、6、9 和 14 号的染色体上,表型变异解释率为 3.0%~19.3%,其中 *Glyma.13g123500* 和 *Glyma.13g194400* 2 个基因在籽粒中特异性表达,推测这 2 个基因可能为大豆水溶性蛋白质的候选基因。

关联分析中要求所用自然群体尽可能拥有广泛的遗传基础^[17],本研究 224 份大豆种质材料是基于中国 23 587 份大豆种质资源筛选构建的,能够代表资源材料 63.5% 以上的遗传多样性,具有丰富的遗传变异^[18],目前该群体已被成功用于成熟籽粒硬度、脂肪酸合成和疫霉根腐病等大豆多个性状的关联分析^[19-21]。遗传变异丰富的自然群体含有足够丰富的等位基因,能够提高关联分析的检测效率,其检测结果更为可靠^[22]。尽管已有水溶性蛋白质相关 QTL 的研究报道,但水溶性蛋白质属于复杂的数量性状,其遗传基础尚未全面解析。本研究以 224 份遗传多样性丰富的大豆种质材料为研究对象,借助用于分析比较栽培大豆和野生大豆遗传多样性以及与脂质合成降解通路相关的 1 514 个 SNP 标记对水溶性蛋白质含量进行关联分析,并进一步预测与水溶性蛋白质相关的重要候选基因,为大豆高水溶性蛋白质的分子标记辅助选择育种及品质改良提供理论依据。

1 材料与方法

1.1 材料

224 份供试材料分别来源于中国的 4 个大豆生态区,其中 39 份来源于东北大豆生态区,37 份来源于北方大豆生态区,48 份来源于黄淮海大豆生态区,100 份来源于南方大豆生态区。供试材料由中国农业科学院邱丽娟研究员提供。

1.2 试验设计

分别于 2017 和 2018 年将各大豆品种种植在南京农业大学江浦试验站(32°12'N,118°37'E)。每个材料种植 1 行,行长 4 m,行距 0.5 m,株距 0.1 m,采用随机区组设计,3 次重复。于完熟期收获种子,37℃下烘干至恒重后测定大豆水溶性蛋白质含量。

1.3 方法

1.3.1 表型测定 参照 Zhu 等^[7]和陈强等^[5]方法进行大豆水溶性蛋白质含量的提取和测定。

标准曲线的绘制:以不同浓度(0,0.2,0.4,0.6,0.8 和 1.0 mg·mL⁻¹)的牛血清蛋白标准溶液绘制标准曲线。

水溶性蛋白质的提取和测定:将烘干后的种子于研磨机中充分研磨,100 目过筛,称取 0.1 g 粉末于 2 mL 离心管中,加入 1 mL 蒸馏水,充分振荡混匀,置于恒温振荡器(20 ℃、280 r·min⁻¹)中振荡 2 h,20 ℃ 12 000 r·min⁻¹ 离心 10 min,将上清液转移至 50 mL 容量瓶中,再向沉淀中加入 1 mL 蒸馏水,如上重复浸提 3 次,定容至 50 mL,制成水溶性蛋白质待测液。使用凯基生物公司的 Bradford 蛋白含量检测试剂盒,取 50 μL 待测液置于离心管中,加入 2 mL 考马斯亮蓝溶液,混匀后室温下静置 2 min,利用酶标仪在 595 nm 波长下比色,记录吸光值。将吸光值带入标准曲线公式,计算出蛋白质含量,每份材料 3 次重复,取平均值即为大豆水溶性蛋白质含量。利用 SPSS 20.0 软件对大豆水溶性蛋白质含量进行描述性统计分析。采用 SAS 9.3 软件对表型数据进行方差分析,并计算遗传率,公式为:

$$h^2 = \frac{\sigma_g^2}{\left(\sigma_g^2 + \frac{\sigma_{ge}^2}{n} + \frac{\sigma_e^2}{rn}\right)}$$

其中, σ_g^2 表示基因型方差, σ_{ge}^2 代表基因型 × 环境互作方差, σ_e^2 代表误差方差, r 代表重复数, n 代表环境个数。

1.3.2 基因分型、LD 分析和群体结构分析 前期 Huang 等^[12]利用 1 654 个 SNP 标记对 224 份大豆种质材料进行了基因分型。这些 SNP 标记整合了与野生大豆和栽培大豆遗传多样性相关的 554 个 SNP^[23] 和大豆籽粒脂肪酸构成研究的 1 536 个 SNP^[21]。为提高关联分析功效,本研究利用 TASSEL v5.2 软件,以最小等位基因频率(minor allele frequency, $MAF > 0.05$)和杂合率(heterozygosity, $H < 0.15$)为标准,对标记进行进一步质量筛选,得到 1 514 个高质量 SNP 标记用于关联分析。

参考 Jing 等^[19]方法,通过 TASSEL v5.2 计算 SNP 标记间的 LD 程度 r^2 ,当 $r^2 = 0.04$ 时,该群体在 20 条染色体上的平均 LD 衰减距离是 544.01 kb。

利用 STRUCTURE 2.2 软件估算 224 份大豆种质材料的群体结构。将群体内存在的亚群数目 K 值设定为 1~6,迭代次数设置为 10 000,蒙特卡罗迭代(marko chain montecarlo, MCMC)次数设置为 100 000,设置 K 值重复运行 4 次,取 $\text{Ln}P(K)$ 值最大时对应的亚群即为种质所属亚群,也就是 Q 值矩阵。本研究对关联群体的群体结构分析显示当 $K = 2$ 时, ΔK 最大,该群体可分为两个亚群^[19]。

1.3.3 全基因组关联分析 利用 TASSEL v5.2 软件中的 MLM($Q + K$)模型对该群体水溶性蛋白质含量进行关联分析,将标记与目标性状显著关联的阈值设为 $-\log_{10}P \geq 2.0$ 。

1.3.4 候选基因预测 在关联分析结果的基础上,在显著关联且表型变异解释率较高的位点所在候选区间内($LD \pm 544.01$ kb),基于 Phytozome 数据库(<https://phytozome.org/>)和 Soybase 数据库(<https://www.soybase.org/>)中基因的功能注释预测与水溶性蛋白质相关的候选基因。在 Soybase 数据库中的 RNA-seq 数据库下载候选基因的表达数据,利用 R 软件绘制候选基因组织表达热图。

2 结果与分析

2.1 大豆水溶性蛋白质含量的表型变异

2017 和 2018 年分别对关联群体的水溶性蛋白质含量的测定和统计分析结果表明:该关联群体 2 年水溶性蛋白质含量表型变异幅度分别为 7.47%~32.60% 和 10.00%~34.60%,变异系数分别为 20.18% 和 17.23%,表明水溶性蛋白质含量在 2 年中存在较大的表型变异。方差分析表明 WSPC 在基因型、环境以及基因型与环境互作中均达到极显著水平($P \leq 0.01$)。该群体水溶性蛋白质含量遗传率为 77.14%,表明其主要受遗传因素影响(表 1)。该群体水溶性蛋白质含量呈连续性变异,符合数量性状的遗传特征,适合用于进一步的关联分析研究(图 1)。

表 1 大豆水溶性蛋白质含量的统计分析

Table 1 Statistical analysis of water-soluble protein content in soybean

年份	均值	标准差	最小值	最大值	变异系数	显著性			遗传率
Year	Mean	SD	Min.	Max.	CV/%	<i>G</i>	<i>E</i>	<i>G</i> × <i>E</i>	<i>h</i> ² /%
2017	18.83	3.80	7.47	32.60	20.18	**	**	**	77.14
2018	21.41	3.69	10.00	34.60	17.23				

G 代表基因型效应; *E* 代表环境效应; *G* × *E* 代表基因型与环境互作效应; **: 在 $P \leq 0.01$ 水平有极显著变异。
G represents the genotype effect; *E* represents the environment; *G* × *E* represents the interaction effect of genotype and environment; **: There is significant variation at $P \leq 0.01$ level.

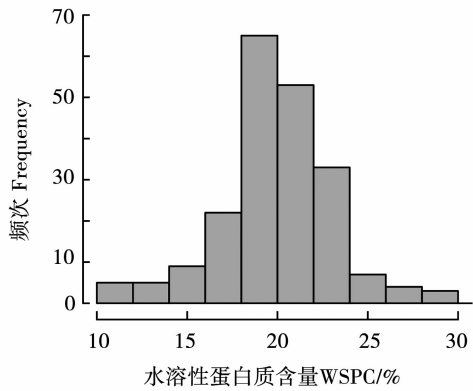
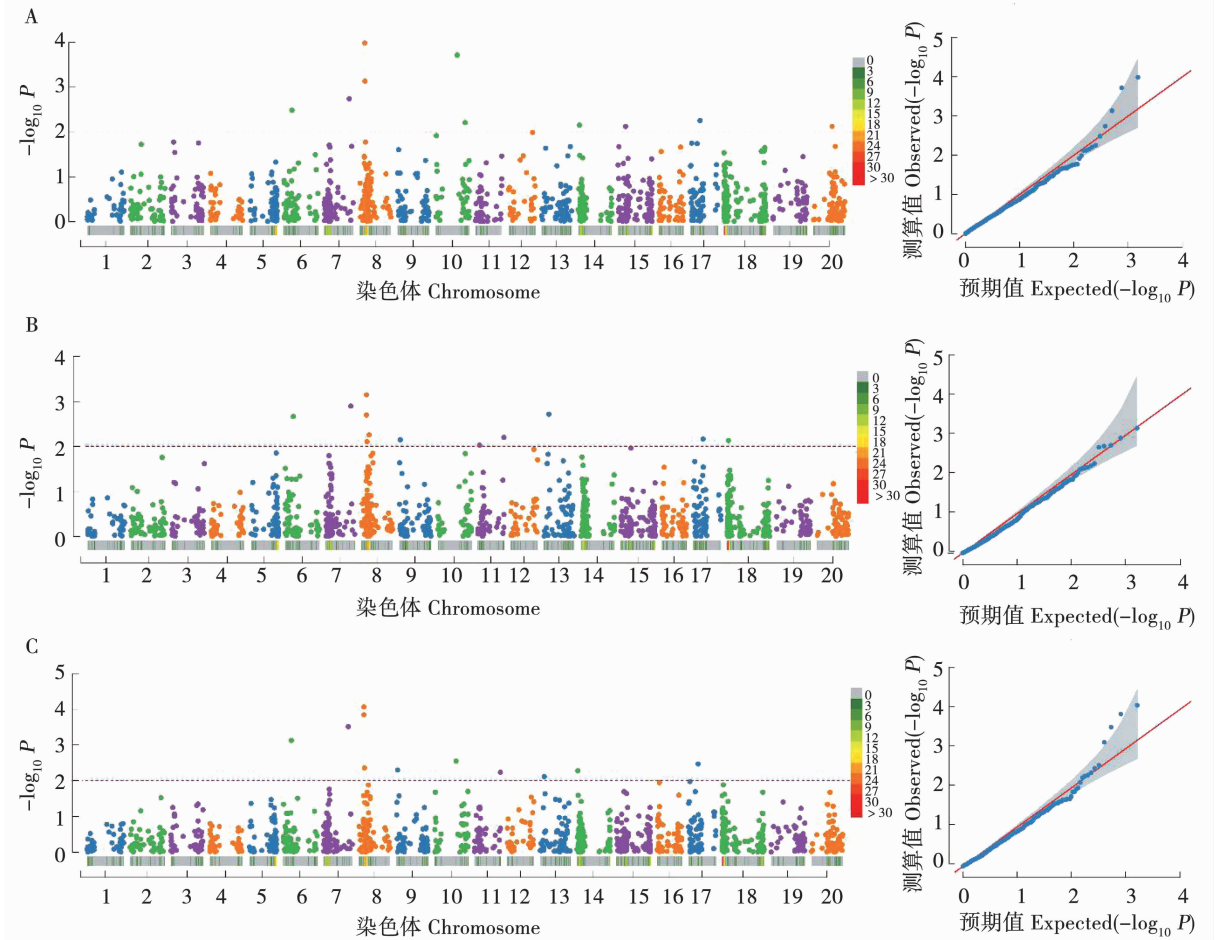


图1 大豆水溶性蛋白质含量的分布图
Fig.1 The distribution of WSPC in soybean

2.2 大豆水溶性蛋白质的 GWAS 分析

对关联群体 2017、2018 年水溶性蛋白含量及 2 年均值的关联分析共检测到 18 个 SNP 标记与大

豆水溶性蛋白质含量显著关联(图 2),表型变异贡献率的范围为 3.24% ~ 8.96% (表 2)。通过 LD block 分析发现 18 个 SNP 标记对应于 16 个 QTL 位点,分别位于 6、7、8、9、10、11、13、14、15、17、18 和 20 号染色体上(表 2)。在所检测到的位点中,*qWSPC6*、*qWSPC7*、*qWSPC8-1*、*qWSPC9*、*qWSPC10-1*、*qWSPC11-1*、*qWSPC14* 和 *qWSPC17* 8 个位点至少被检测到 2 次,表明这些位点受环境因素影响较小。其余 8 个位点仅被检测到 1 次,表明其受环境因素影响较大。其中 *qWSPC7* 和 *qWSPC8-1* 2 个位点表型变异解释率较高,其范围为 4.61% ~ 8.03%,在 2017、2018 年和 2 年均值下均被检测到,表明这 2 个位点附近可能存在调控大豆水溶性蛋白质的关键基因。



A:2017 年; B:2018 年; C:两年均值。
A: 2017; B: 2018; C: Average.

图2 大豆水溶性蛋白质全基因组关联分析的曼哈顿图和 Q-Q 图

Fig.2 Manhattan plots and Q-Q plots of genome wide association for WSPC in soybean seed

表 2 与大豆水溶性蛋白相关的位点
Table 2 Locus associated with WSPC in soybean

位点 Locus	SNP 标记 SNP marker	染色体 Chromosome	物理位置 Position	环境 Environment	显著性 -log ₁₀ <i>P</i>	贡献率 <i>R</i> ² /%	被检测 到次数 Number of detections	相关 QTL Related QTL
<i>qWSPC6</i>	Q-06-0079992	6	12470088	2017/2018/均值	2.48 ~ 3.13	4.26 ~ 5.59	3	Seed protein 34-2
<i>qWSPC7</i>	Q-07-0217819	7	36518586	2017/2018/均值	2.74 ~ 3.51	6.28 ~ 8.03	3	Seed protein 41-9
<i>qWSPC8-1</i>	Q-08-0054375	8	7785180	2017/2018/均值	2.69 ~ 4.06	4.61 ~ 7.66	3	cqSeed protein-013,cqSeed protein-016, Seed protein 30-4,Seed protein 34-4, Seed protein 34-5,GqWSPC8
	Q-08-0055285	8	7887402	2017/2018/均值	3.13 ~ 3.84	5.57 ~ 7.17		
	Q-08-0061616	8	8751104	2018/均值	2.11 ~ 2.36	5.57 ~ 7.17		
<i>qWSPC8-2</i>	Q-08-0087375	8	11647122	2018	2.26	3.70	1	/
<i>qWSPC9</i>	BARC-016397-02579	9	1935225	2018/均值	2.15 ~ 2.3	4.94 ~ 5.44	2	Seed protein 36-27,Seed protein 40-3, Seed protein 41-7
<i>qWSPC10-1</i>	BARC-018019-02497	10	32502722	2017/均值	2.55 ~ 3.72	6.04 ~ 8.96	2	Seed protein 36-40
<i>qWSPC10-2</i>	BARC-031677-07213	10	44421744	2017	2.21	5.23	1	Seed protein 27-5
<i>qWSPC11-1</i>	BARC-025703-04999	11	34311816	2018/均值	2.2 ~ 2.23	5.13 ~ 5.31	2	/
<i>qWSPC11-2</i>	Q-11-0017683	11	2674088	2018	2.03	3.24	1	/
<i>qWSPC13-1</i>	BARC-014535-01572	13	15673257	均值	2.11	4.99	1	/
<i>qWSPC13-2</i>	Q-13-0055305	13	14630992	2018	2.72	6.08	1	/
<i>qWSPC14</i>	Q-14-0016532	14	2058660	2017/均值	2.15 ~ 2.28	3.60 ~ 3.82	2	/
<i>qWSPC15</i>	Q-15-0079298	15	11284739	2017	2.12	3.51	1	/
<i>qWSPC17</i>	Q-17-0112995	17	14794222	2017/2018/均值	2.16 ~ 2.47	4.83 ~ 5.60	3	Seed protein 26-2,Seed protein 36-17, Seed acidic fraction 1-1,Seed conglycinin 1-1, Seed Cys 3-5,Seed Met 3-5
<i>qWSPC18</i>	BARC-027452-06569	18	1789574	2018	2.13	4.90	1	Seed protein 26-12
<i>qWSPC20</i>	Q-20-0136783	20	29062110	2017	2.12	3.51	1	cqSeed protein-003,Seed protein 1-1

2.3 候选基因预测

Phytozome 和 Soybase 数据库的预测结果显示:在 *qWSPC7* 和 *qWSPC8-1* 2 个位点候选区间内共获得 25 个与水溶性蛋白质相关的候选基因,主要涉及蛋白质或氨基酸的合成、运输及代谢等生物学过程(表 3)。

根据 Soybase 数据库中 RNA-seq 数据下载到 25 个候选基因在大豆 14 个组织中的表达数据。其中 7 个基因在大豆籽粒、根及根瘤中具有较高的表达。*Glyma. 08G103100* 基因在大豆籽粒中特异性表达,

特别是在种子发育后期表达量较高;*Glyma. 08G108900* 和 *Glyma. 08G105100* 基因在大豆 14 个组织中都具有较高的表达水平,特别是在根及根瘤中表达量最高;*Glyma. 07G195000*、*Glyma. 08G107800* 和 *Glyma. 08G115800* 3 个基因在大豆籽粒、根和根瘤等组织中具有相对较高的表达量;*Glyma. 08G107700* 基因在大豆籽粒及根瘤中表达(图 3)。

这 7 个基因在大豆籽粒、根或根瘤中表达量较高,可作为大豆水溶性蛋白质的候选基因,可能具有调控大豆水溶性蛋白质合成的功能。

表 3 大豆水溶性蛋白候选基因预测及功能注释

Table 3 Functional annotationfor candidate genes of WSPC in soybean

位点 QTL	染色体 Chromosome	候选区间 Candidate interval		候选基因 Candidate gene	功能注释 Function annotation
		起始 Start/bp	终止 End/bp		
qWSPC7	7	35974576	37062596	Glyma. 07G192200	STRUBBELIG-受体家族 3 STRUBBELIG-receptor family 3
					类丝氨酸羧肽酶 31 Serine carboxypeptidase-like 31
				Glyma. 07G192700	酵母氧化酶装配体同源物 1 Homolog of yeast oxidase assembly 1
				Glyma. 07G193700	网格蛋白轻链蛋白 Clathrin light chain protein
					类贩运蛋白粒子复杂亚基蛋白 Trafficking protein particle complex subunit-like protein
				Glyma. 07G195000	蛋白激酶超家族蛋白 Protein kinase superfamily protein
				Glyma. 07G197200	含有 DNAJ 热激 N – 末端结构域的蛋白 DNAJ heat shock N-terminal domain-containing protein
					线粒体底物载体蛋白家族 Mitochondrial substrate carrier family protein
				Glyma. 07G199200	包含 MAC/Perforin 结构域的蛋白 MAC/Perforin domain-containing protein
				Glyma. 07G199800	蛋白激酶超家族蛋白 Protein kinase superfamily protein
qWSPC8-1	8	7241170	9295114	Glyma. 08G099500	类 IAP – 蛋白 1 IAP-like protein 1
					半乳糖苷转移酶家族蛋白 Galactosyltransferase family protein
				Glyma. 08G100300	多药耐药相关蛋白 Multidrug resistance-associated protein 3
				Glyma. 08G101500	赖氨酸 – 组氨酸转运蛋白 Lysine histidine transporter 1
					ARM 重复超家族蛋白 ARM repeat superfamily protein
				Glyma. 08G101900	脲酶 Urease
				Glyma. 08G103000	核转运因子 2B Nuclear transport factor 2B
					FKBP 型肽基脯氨酰顺反异构酶家族蛋白 FKBP-type peptidyl-prolyl cis-trans isomerase family
				Glyma. 08G103100	富亮氨酸重复蛋白激酶家族蛋白 Leucine-rich repeat protein kinase family protein
				Glyma. 08G107700	天冬氨酸激酶 – 高丝氨酸脱氢酶 ii Aspartate kinase-homoserine dehydrogenase ii
					丝氨酸羟甲基转移酶 4 Serine hydroxymethyltransferase 4
				Glyma. 08G107800	种子贮藏 2S 清蛋白超家族蛋白 Seed storage 2S albumin superfamily protein
				Glyma. 08G108900	
				Glyma. 08G112300	

续表 3

位点	染色体	候选区间 Candidate interval		候选基因	功能注释
QTL	Chromosome	起始 Start/bp	终止 End/bp	Candidate gene	Function annotation
				<i>Glyma. 08G112600</i>	FAD 结合小檗碱家族蛋白 FAD-binding Berberine family protein
				<i>Glyma. 08G113400</i>	氨基酸通透酶家族蛋白 Amino acid permease family protein
				<i>Glyma. 08G115800</i>	通用调节因子 9 General regulatory factor 9

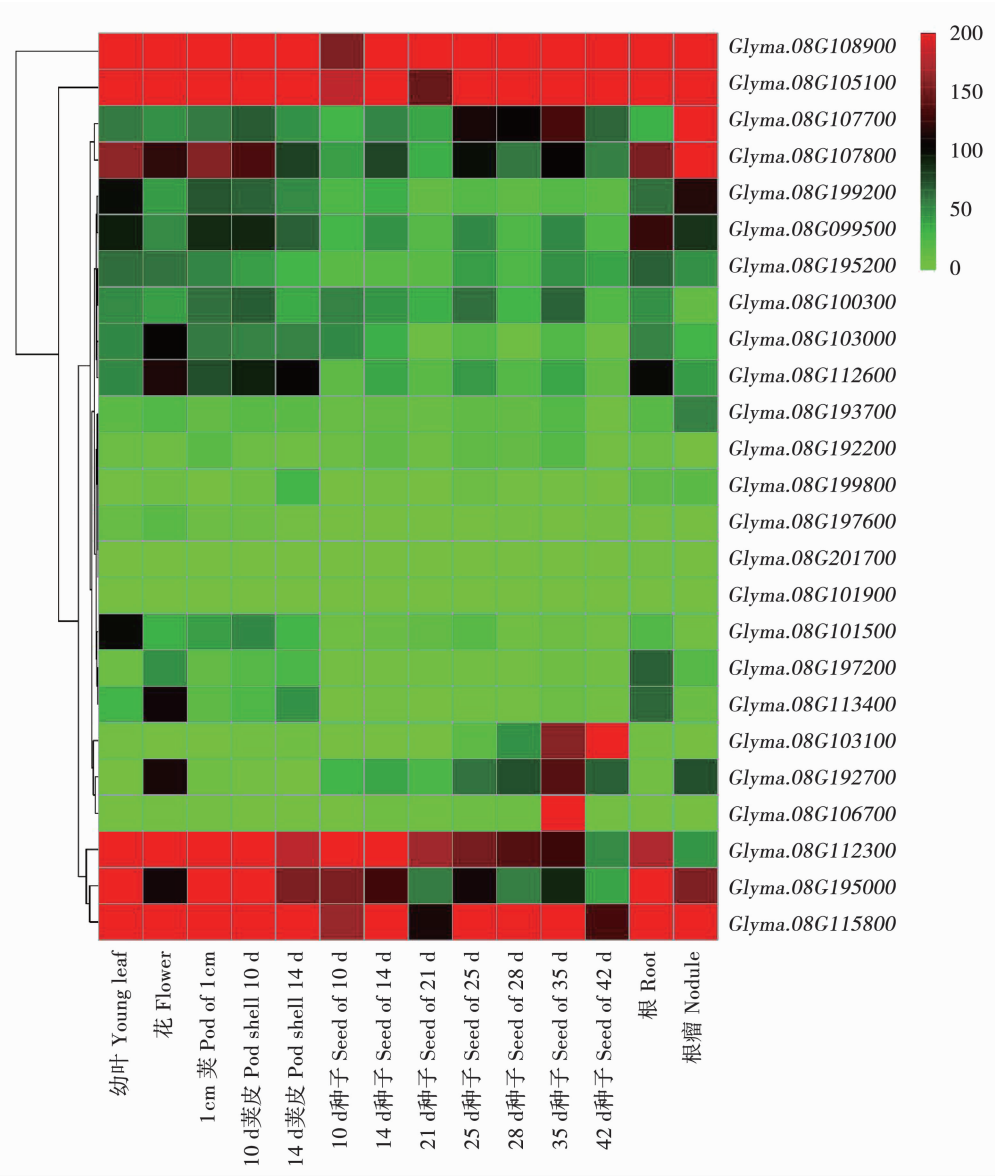


图 3 候选基因组织表达热图

Fig. 3 Heatmap profiles of the candidate genes in tissues

3 讨 论

大豆水溶性蛋白是影响大豆蛋白品质的重要因素。大豆蛋白中水溶性蛋白质含量的高低会影响大豆蛋白的分离,进而影响大豆蛋白食品的加工。溶解度较高的蛋白质可以获得良好的乳化和

发泡等功能特性,在食品加工和利用中具有重要作用^[4]。在大豆蛋白食品加工过程中可以通过一些物理和化学的方法提高大豆蛋白的溶解度^[24-26]。但是在提高大豆蛋白的溶解度的同时增加了大豆蛋白食品生产过程的成本,是一种不可持续的方法。从大豆种质资源中挖掘与大豆水溶性蛋白相

关的功能标记,通过标记辅助选择育种方法培育具有高水溶性蛋白质含量的大豆品种是一种经济有效的手段。

目前与大豆水溶性蛋白质相关的标记仅有少数报道,本研究通过全基因关联分析的方法共检测到18个与水溶性蛋白质显著相关的SNP标记,涉及16个位点。其中,8个位点(*qWSPC6*、*qWSPC7*、*qWSPC8-1*、*qWSPC9*、*qWSPC10-1*、*qWSPC11-1*、*qWSPC14*和*qWSPC17*)至少被检测到2次,表明这些位点受环境影响较小,其余8个位点仅被检测1次,说明其受环境影响较大。本研究检测到的16个位点中,9个位点位于前人已定位的大豆水溶性蛋白质、总蛋白及氨基酸等相关性状的QTL区段内。*qWSPC6*与Lu等^[4]检测到的水溶性蛋白QTL位点*qsp8-4*和*qsp8-5*重叠;*qWSPC7*与Jun等^[27]已定位到的蛋白质QTL一致;*qWSPC8-1*与Zhang等^[3,16,28-29]检测到的水溶性蛋白关联位点重叠;*qWSPC9*位于Mao等^[30-33]定位的3个蛋白质QTL内;10号染色体上的2个位点*qWSPC10-1*和*qWSPC10-2*与Pathan等^[28,34]定位的蛋白质QTL基本一致。*qWSPC17*与Wang等^[35-36]报道的6个蛋白质或氨基酸相关的QTL重叠。*qWSPC18*与Reinprecht等^[34]研究结果一致。*qWSPC20*位于已定位到2个蛋白质QTL内^[37-38]。本研究检测到的大豆水溶性蛋白质相关的位点大多与已报道的大豆水溶性蛋白质、总蛋白质和氨基酸相关的QTL基本一致,这些位点所在的基因组区域将成为挖掘大豆水溶性蛋白质编码基因的重点区域。此外,本研究中有7个水溶性蛋白质的位点被首次检测,产生原因不排除定位群体、环境以及测量方法的不同,有待进一步研究。

大豆蛋白质主要在籽粒发育阶段积累,而水溶性蛋白质是大豆蛋白的重要组成部分^[39-40]。大豆蛋白质合成过程中需要大量的氮(N),在多数情况下,氮素主要源于根瘤的固氮作用^[41]。大豆蛋白质含量随着根瘤固氮量的降低而减少^[42],生物固氮量的降低可影响大豆籽粒蛋白含量的积累^[43]。本研究根据RNA-seq数据库观察25个候选基因的表达模式,发现有7个基因(*Glyma.07g195000*、*Glyma.08g103100*、*Glyma.08g108900*、*Glyma.08g105100*、*Glyma.08g107800*、*Glyma.08g107700*、*Glyma.08G115800*)在大豆籽粒、根及根瘤中具有较高的表达水平,为大豆水溶性蛋白质的重要候选基因,这些基因是否参与大豆籽粒水溶性蛋白质的调控尚有待于进一步功能验证。

4 结 论

本研究利用全基因组关联分析共检测到18个

SNP标记与大豆水溶性蛋白质显著关联,涉及16个QTL位点,其中9个QTL与已报道的QTL重叠。对表型变异解释率较高的2个位点*qWSPC7*和*qWSPC8-1*候选区间内共预测获得25个候选基因,其中有7个基因在大豆籽粒、根或根瘤中高表达。这些基因可作为大豆水溶性蛋白质的候选基因,可能具有调控大豆水溶性蛋白质的功能。

参考文献

[1] Singh P, Kumar R, Sabapathy S N, et al. Functional and edible uses of soy protein products[J]. Comprehensive Reviews in Food Science and Food Safety, 2008, 7(1): 14-28.

[2] Yu X, Yuan F, Fu X, et al. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds[J]. Food Chemistry, 2015, 196:776-782.

[3] Zhang D, Kan G, Hu Z, et al. Use of single nucleotide polymorphisms and haplotypes to identify genomic regions associated with protein content and water-soluble protein content in soybean [J]. Theoretical and Applied Genetics, 2014, 127(9): 1905-1915.

[4] Lu W, Wen Z, Li H, et al. Identification of the quantitative trait loci (QTL) underlying water-soluble protein content in soybean [J]. Theoretical and Applied Genetics, 2013, 126(2):425-433.

[5] 陈强,雷雅坤,闫龙,等. 利用双尾法定位大豆水溶性蛋白QTL[J]. 华北农学报, 2014, 29(4):80-86. (Cheng Q,Lei Y K, Yan L,et al. QTL mapping of soluble protein content by two tail analysis[J]. Acta Agriculturae Boreali-Sinica,2014, 29(4): 80-86.)

[6] Rafalski J A. Association genetics in crop improvement [J]. Current Opinion in Plant Biology, 2010, 13(2):174-180.

[7] Zhu C, Gore M, Buckler E S, et al. Status and prospects of association mapping in plants[J]. Plant Genome, 2008, 1(1):5-20.

[8] Yu J, Buckler E S. Genetic association mapping and genome organization of maize [J]. Current Opinion in Biotechnology, 2006, 17(2):155-160.

[9] Li H, Peng Z, Yang X, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels [J]. Nature Genetics, 2013, 45(1): 43.

[10] Wen W, Li D, Li X, et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights[J]. Nature Communications, 2014, 5: 3438.

[11] Dong H, Zhao H, Xie W, et al. A novel tiller angle gene, *TAC3*, together with *TAC1* and *D2* largely determine the natural variation of tiller angle in rice cultivars [J]. PLoS Genetics, 2016, 12(11): e1006412.

[12] Huang X, Kurata N, Wang Z X, et al. A map of rice genome variation reveals the origin of cultivated rice[J]. Nature, 2012, 490(7421): 497.

[13] Fang C, Ma Y, Wu S, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean[J]. Genome Biology, 2017, 18(1): 161.

[14] Zhang J, Song Q, Cregan P B, et al. Genome-wide association

- study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*) [J]. Theoretical and Applied Genetics, 2016, 129(1): 117-130.
- [15] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 463 (7278): 178.
- [16] Zhang D, Lyu H, Chu S, et al. The genetic architecture of water-soluble protein content and its genetic relationship to total protein content in soybean[J]. Scientific Reports, 2017, 7(1): 5053.
- [17] Jiang P, Zhang P P, Zhang X, et al. Genetic diversity and association analysis for solvent retention capacity in the accessions derived from soft wheat Ningmai 9 [J]. International Journal of Genomics, 2017(9):2413150.
- [18] 邱丽娟, 李英慧, 关荣霞, 等. 大豆核心种质和微核心种质的构建、验证与研究进展[J]. 作物学报, 2009, 35(4): 571-579. (Qiu L J, Li Y H, Guan R X, et al. Establishment, representative testing and research progress of soybean core collection and mini core collection[J]. Acta Agronomica Sinica, 2009, 35(4): 571-579.)
- [19] Jing H, Na G, Li Y, et al. Phenotypic evaluation and genetic dissection of resistance to *Phytophthora sojae* in the Chinese soybean mini core collection [J]. BMC Genetics, 2016, 17 (1):85.
- [20] Zhang X, Zhao J, Bu Y, et al. Genome-wide association studies of soybean seed hardness in the Chinese mini core collection[J]. Plant Molecular Biology Reporter, 2018, 36(4): 605-617.
- [21] Li Y, Reif J C, Ma Y, et al. Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean[J]. BMC Genomics, 2015, 16(1): 841.
- [22] Jiang P, Zhang P P, Zhang X, et al. Genetic diversity and association analysis for solvent retention capacity in the accessions derived from soft wheat Ningmai 9 [J]. International Journal of Genomics, 2017(9): 2413150.
- [23] Li D, Sun M, Han Y, et al. Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*) [J]. Euphytica, 2010, 172(1):49-57.
- [24] Malhotra A, Coupland J N. The effect of surfactants on the solubility, zeta potential, and viscosity of soy protein isolates[J]. Food Hydrocoll, 2004, 18(1):101-108.
- [25] Ortiz S E M, Wagner J R. Hydrolysates of native and modified soy protein isolates: Structural characteristics, solubility and foaming properties [J]. Food Research International, 2002, 35 (6): 511-518.
- [26] Walsh D J, Cleary D, McCarthy E, et al. Modification of the nitrogen solubility properties of soy protein isolate following proteolysis and transglutaminase cross-linking[J]. Food Research International, 2003, 36(7): 677-683.
- [27] Jun T H, Van K, Kim M Y, et al. Association analysis using SSR markers to find QTL for seed protein content in soybean [J]. Euphytica, 2008, 162(2): 179-191.
- [28] Pathan S M, Vuong T, Clark K, et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean[J]. Crop Science, 2013, 53 (3):765-774.
- [29] Tajuddin T, Watanabe S, Yamanaka N, et al. Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines [J]. Breeding Science, 2003, 53(2): 133-140.
- [30] 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.
- [31] Qi Z, Hou M, Han X, et al. 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[J]. Plant Breeding, 2014, 133(4): 499-507.
- [32] Jun T H, Van K, Kim M Y, et al. Association analysis using SSR markers to find QTL for seed protein content in soybean [J]. Euphytica, 2008, 162(2): 179-191.
- [33] Panthee D R, Kwanyuen P, Sams C E, et al. Quantitative trait loci for β -conglycinin (7S) and glycinin (11S) fractions of soybean storage protein[J]. Journal of the American Oil Chemists' Society, 2004, 81(11): 1005-1012.
- [34] Reinprecht Y, Poysa V W, Yu K, et al. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm [J]. Genome, 2006, 49(12): 1510-1527.
- [35] Wang X, Jiang G L, Song Q, et al. Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean [J]. Euphytica, 2015, 201 (2): 293-305.
- [36] Nichols D M, Glover K D, Carlson S R, et al. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits[J]. Crop Science, 2006, 46(2): 834-839.
- [37] Diers B W, Keim P, Fehr W R, et al. RFLP analysis of soybean seed protein and oil content [J]. Theoretical and Applied Genetics, 1992, 83(5): 608-612.
- [38] Cheng L, Yuan H Y, Ren R, et al. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in *Glycine max* [J]. Frontiers in Plant Science, 2016, 7: 515.
- [39] Sale P W G, Campbell L C. Changes in physical characteristics and composition of soybean seed during crop development [J]. Field Crops Research, 1980, 3: 147-155.
- [40] Hill J E, Breidenbach R W. Proteins of soybean seeds: II. Accumulation of the major protein components during seed development and maturation[J]. Plant Physiology, 1974, 53(5): 747-751.
- [41] Matheny T A, Hunt P G. Effects of irrigation on accumulation of soil and symbiotically fixed n by soybean grown on a Norfolk Loamy sand [J]. Agronomy Journal, 1983, 75(5): 719-722.
- [42] Vollmann J, Fritz C N, Wagentristl H, et al. Environmental and genetic variation of soybean seed protein content under Central European growing conditions[J]. Journal of the Science of Food and Agriculture, 2000, 80(9): 1300-1306.
- [43] Tamagno S, Adee E A, Ciampitti I A. Effects of nitrogen in soybean seed quality definition during seed-filling period [J]. Kansas Agricultural Experiment Station Research Reports, 2018, 4(7): 8.