

利用关联分析方法挖掘自然群体中大豆油分和蛋白质含量相关 SSR 标记

于志远¹,王伟威¹,魏 峡¹,陈庆山²,赵贵兴¹,齐照明²,李 杰²,刘丽君¹

(1. 黑龙江省农业科学院 大豆研究所,黑龙江 哈尔滨 150086; 2. 东北农业大学 大豆研究所,黑龙江 哈尔滨 150030)

摘 要:为研究自然群体中大豆品种的油分和蛋白质含量变化,筛选出相关标记的优异等位变异,以 327 份东北主推品种和优异亲本材料构成的自然群体为试验材料。以分布于大豆 20 条染色体的 186 对 SSR (simple sequence repeat) 分子标记检测所有试验材料的基因型,利用 STRCTURE 2. 3. 4 软件分析群体结构,将试验材料分为 7 个亚群。利用 TASSEL 3. 0 软件的混合线性模型的方法对大豆自然群体的油分和蛋白质含量进行关联分析。在极显著水平 ($P < 0. 01$) 且贡献率大于 1% 情况下,共检测到 33 个关联位点。与油分含量极显著关联位点 12 个,解释率为 2. 19% ~ 10. 05%;与蛋白质含量极显著关联位点 11 个,解释率为 2. 65% ~ 9. 08%;与油分和蛋白质含量同时关联位点 6 个,分别为 Satt005、Satt117、Satt565、Satt469、Satt594 和 Satt546,其中 Satt594 解释率最高,油分为 10. 05%,蛋白质为 9. 08%。
关键词:大豆;油分含量;蛋白含量;关联分析
中图分类号:S565. 1 **文献标识码:**A **DOI:**10. 11861/j. issn. 1000-9841. 2015. 06. 0977

Exploiting SSR Loci Related with Soybean Oil and Protein Content by Using Association Analysis in Natural Population

YU Zhi-yuan¹, WANG Wei-wei¹, WEI Lai¹, CHEN Qing-shan², ZHAO Gui-xing¹, QI Zhao-ming², LI Jie², LIU Li-jun¹
(1. Soybean Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China; 2. Soybean Research Institute of Northeast Agricultural University, Harbin 150030, China)

Abstract: In order to study the oil and protein content in resource population of soybean, and to select trait related markers, 327 soybean breeding-used cultivars from Northeast China were used as experimental materials. 186 pairs of SSR (simple sequence repeat) markers, distributed on 20 chromosomes of soybean, were selected to detect genotypes of all testing materials, Software STRCTURE 2. 3. 4 was used to analysis population structure and experimental material were divided into 7 subgroups. The MLM (Mixed Linear Model) process of software TASSEL 3. 0 was used to associate oil and protein contents trait data with SSR marker data. 33 associated SSR loci were found highly significant ($P < 0. 01$) related with contribution rate greater than 1%. Among them, 12 loci were significantly associated with oil content, explaining 2. 7%-10. 08% of variation, 11 loci were significantly associated with protein content, explaining 2. 4%-9. 08% variation. There are 6 identical loci associated with both oil and protein content, which are Satt005, Satt117, Satt565, Satt469, Satt594 and Satt546. Explanation rate of Satt594 is the highest, which is 10. 05% for oil content variation and 9. 08% for protein content variation. This study provides information for molecular assisted breeding of soybean with high oil and protein content.
Keywords: Soybean; Oil content; Protein content; Association analysis

大豆蛋白质和油分含量作为重要的育种性状,影响大豆加工品质。挖掘相关基因位点能够为分子辅助育种提供帮助。目前,研究数量性状的方法主要包括连锁分析和关联分析,连锁分析以亲本杂交建立遗传群体为材料,亲本材料的遗传差异是决定 QTL 定位精确性的基础,定位基因存在一定的局限性。关联分析是一种定位和挖掘数量性状位点的新方法^[1-3],以自然群体为材料,以连锁不平衡为基础,将目标性状变异与基因多样性进行分析,可直接鉴定出与表型变异相关基因位点或标记性位点对目标性状有贡献率的优异等位变异。Jun 等^[4]用 150 对 SSR 分子标记与美国农业部收集的 96 份大豆试验材料的油分和蛋白质含量做关联分析,发现分别位于 J 和 M 连锁群的 Satt431 和 Satt551 与

蛋白质含量相关。Shi 等^[5]用 65 对 SSR 分子标记和 105 份亚洲和美国的大豆实验材料的油分和蛋白质含量做关联分析,发现 13 个标记与油分含量相关,19 个标记与蛋白含量相关。齐照明等^[6]用 32 对 SSR 分子标记与中国东北地区 的 96 份大豆材料的油分和蛋白含量做关联分析,发现 12 个标记位点与油分含量相关,10 个位点与蛋白含量相关。
东北是中国优质大豆的主要产区,本研究选用中国东北地区主要育种材料 327 份构成自然群体,利用关联分析方法,挖掘油分和蛋白优异等位变异,定位油分和蛋白含量相关的 SSR 标记位点,为高油高蛋白育种提供分子基础,为分子辅助选择育种提供一定的理论实践基础。

收稿日期:2015-03-06
基金项目:黑龙江省农业科技创新工程(2D028);国家现代农业产业技术体系(CARS04-PS05);哈尔滨市科技局科技创新人才项目(2013RFQYJ016);黑龙江省归国基金项目(LC2012C37)。
第一作者简介:于志远(1986-),男,硕士,主要从事大豆遗传育种研究。E-mail:ddyuzhiyuan@163.com。
通讯作者:刘丽君(1958-),女,研究员,主要从事作物遗传育种研究。E-mail:nkyssbd@126.com。

1 材料与方法

1.1 材料

选取东北地区主要育种材料 327 份,于 2012 年 5 月中旬播种于黑龙江省农业科学院民主园区。采用随机区组设计,行长 4 m,株距 5 cm,3 次重复。定期田间除草和病虫害防治。为去除边际效应,秋季选取中间 1 m 大豆进行收获,测定油分和蛋白质含量。

1.2 SSR 分子标记分析

在田间幼苗期采集大豆叶片 100 mg,利用 AxyPrep 试剂盒提取 DNA,材料 DNA 浓度范围均在 100 ~ 200 ng·μL⁻¹。根据 SoyBase 网站(<http://soybase.org/>)提供的 SSR 引物信息,由上海生工合成大豆引物 1 015 对,随机选取 10 份大豆材料对 1 015 对引物进行筛选,选择条带清晰且多态性较高的 186 对引物用于群体扩增实验。

采用 20 μL 的 PCR 反应体系:MgCl 1.5 μL, Buffer 2.0 μL, dNTP 0.3 μL, Taq 酶 3.0 μL,引物 3.0 μL,模板 DNA 3.0 μL,ddH₂O 10 μL。PCR 反应在 ABI Veriti PCR 仪中进行,条件为 95℃ 预变性 5 min,94℃ 变性 30 s;47℃ 复性 30 s;72℃ 延伸 30 s;35 个循环,最后 72℃ 延伸 5 min,4℃ 永久保存。然后用 6% 的聚丙烯酰胺胶电泳分离,电泳条件为 120 W 恒功率,电泳时间为 90 min。

1.3 蛋白质含量及油分含量测定

种子蛋白质含量的测定采用国标 GB2905-82 方法。种子油分含量的测定采用国标 GB2906-82 方法。

1.4 关联分析与优异等位变异

利用 STRCTURE 2.3.4^[7] 软件对材料进行群体结构分析,由于材料个数多,假定群体 K 是 1 ~ 13,不作数迭代从 20 000 ~ 200 000,每次增加 10 000,重复 5 次,确定获得稳定可靠的结果。然后利用 TASSEL 3.0^[8] 软件,Q 值作为协变量,用混合线性模型 (Mixed Linear Mode, MLM) 进行 SSR 分子标记和蛋白质含量、油分含量的关联分析,利用文自翔^[9] 提出的位点极显著性标准 $P < 0.01$ 且贡献率大于 1% 对关联位点进行判定,定位与蛋白质含量和油分含量相关联的 SSR 分子标记位点。

2 结果与分析

2.1 大豆油分和蛋白含量分析

大豆自然群体油分和蛋白质含量在实验重复间无显著性差异,蛋白质含量最大值 48.95%,最小值 34.57%,蛋白质含量 34.57% ~ 37.57% 的材料 50 个,蛋白质含量 37.57% ~ 40.57% 的材料 183 个,蛋白质含量 40.57% ~ 43.57% 的材料 69 个,蛋白质含量 43.57% ~ 46.57% 的材料 14 个,蛋白质含量 46.57% ~ 48.95% 的材料 1 个。油分含量最大值 23.43%,最小值 14.53%,其中油分含量 14.53% ~ 16.53% 的材料 2 个,油分含量 16.53% ~ 17.53% 的材料 17 个,油分含量 17.53% ~ 18.53% 的材料 113 个,油分含量 19.53% ~ 20.53% 的材料 15 个,油分含量 20.53% ~ 23.43% 的材料 72 个。蛋白质和油分含量变异基本符合正态分布。

表 1 大豆自然群体中油分和蛋白质含量的分析统计
Table 1 Descriptive statistics of soybean oil and protein content

性状 Trait	最大值 Maximum/%	最小值 Minimum/%	平均值 Mean/%	标准差 STDEV	变异系数 Variation	极差 Range
蛋白质 Protein	48.95	34.57	39.63	1.24	0.0003	14.38
油分 Oil	23.43	14.53	20.69	2.10	0.1014	8.90

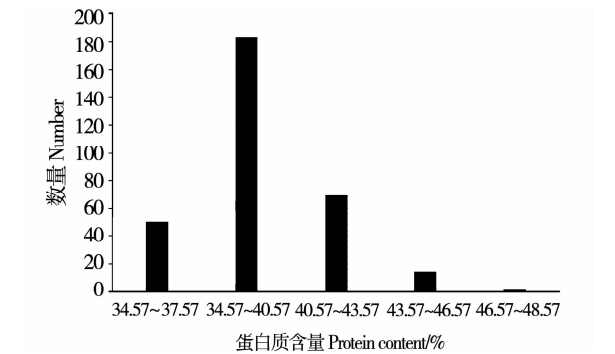


图 1 大豆蛋白质含量分布频率

Fig. 1 Distribution frequency of protein content

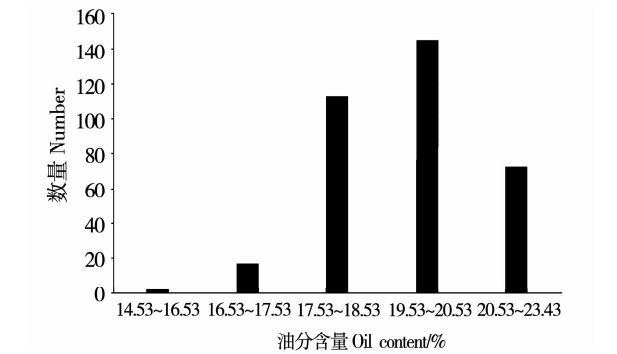


图 2 大豆油分含量分布频率

Fig. 2 Distribution frequency of oil content

2.2 群体结构分析

用 STRCTURE 2.3.4 软件对实验材料群体结构分析,本研究中似然值随着 K 值的增大而增大,因此选用 ΔK 值的变化为判定标准来确定群体的亚群数。当 $K=7$ 时, ΔK 有最大值(图 3),推断群体可分为 7 个亚群(图 4)。

2.3 连锁不平衡

用 Tassel 3.0 软件绘制 186 对 SSR 分子标记间连锁不平衡分布图,群体中标记的连锁不平衡程度能够代表大豆基因组中连锁不平衡的分布和程度。由图 5 可知,连锁不平衡在各个连锁群中均有分布,其中连锁群 A2、G 和 I 中连锁不平衡强度较大($D > 0.8$),共线连锁群主要分布在 I 和 G 连锁群,非共线

的连锁群主要分布在 G 和 F 连锁群中。在 $P < 0.01$ 情况下,标记间的连锁不平衡组合大幅减少。

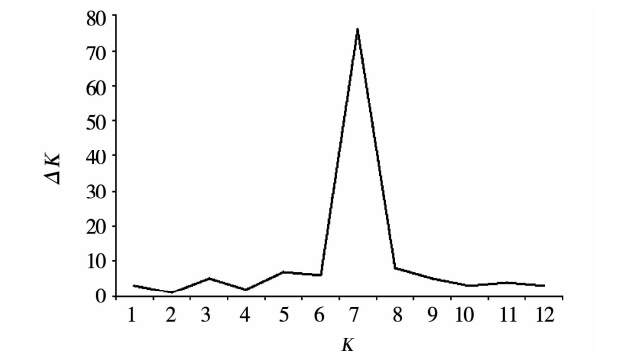
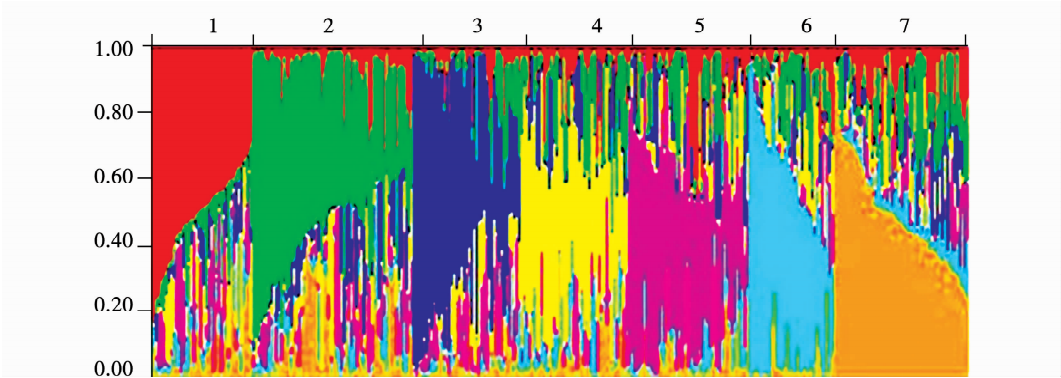


图 3 ΔK 值随着 K 变化图
Fig.3 Changes of ΔK according to K-values



不同颜色表示不同亚群;颜色多少表示占该亚群的概率。
Different colors represent different subgroups; Areas of different colors represent probability of the subgroups。

图 4 基于 186 个 SSR 分子标记对材料的群体结构分析

Fig.4 Population structure analysis based on 186 SSR maker data

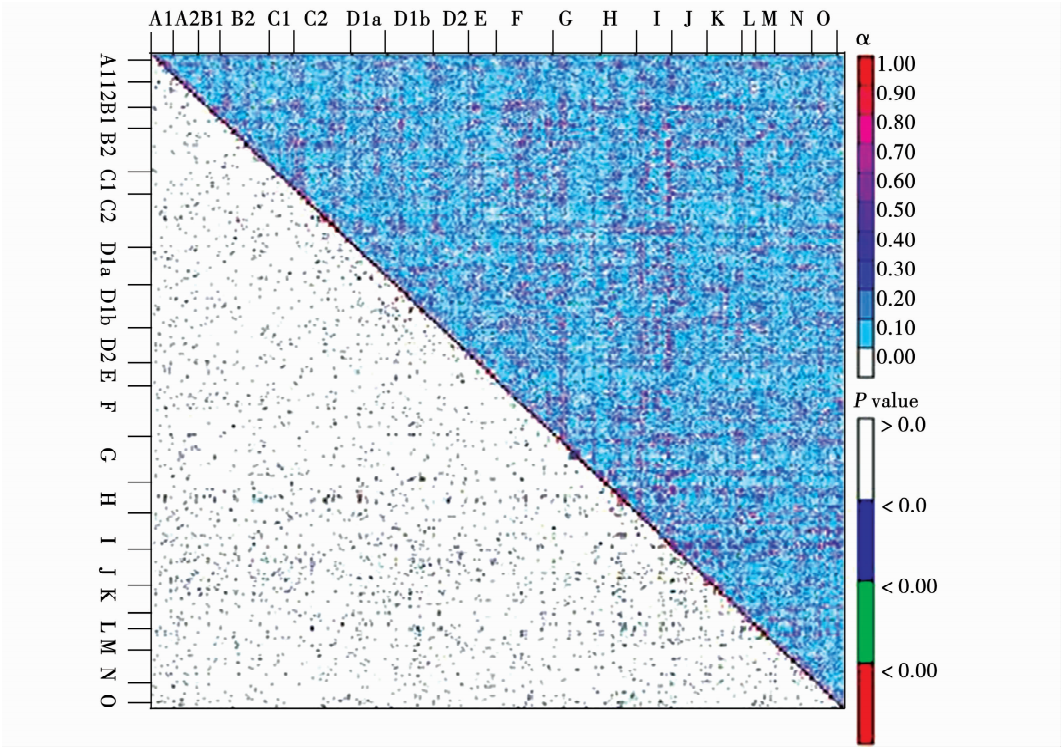


图 5 大豆 20 个连锁群 SSR 位点间的连锁不平衡分布

Fig.5 Distribution of linkage disequilibrium among SSR loci on 20 soybean linkage groups

2.4 关联分析与标记位点

用 186 个 SSR 标记对 327 份自然群体大豆的油分和蛋白质含量进行关联分析,在 12 个位点检测到与油分含量达到极显著 ($P < 0.01$) 的关联标记,分别是 C2 连锁群 Satt557、Satt640, D1a 连锁群的 Sat_345、Sat_346, D1b 连锁群 Satt005、Satt546、Satt506、Satt274, E 连锁的 Satt117, F 连锁群的 Satt656, H 连锁群的 Satt469, G 连锁群的 Satt594, 其中 Satt594 贡献率最高,为 10.05% (表 2)。在 11 个位点检测到与蛋白质含量达到极显著 ($P < 0.01$) 的关联标记,分别是 A1 连锁群的 Satt300, C2 连锁群 Satt640, D1b 连锁群 Satt005、Satt041、Satt537、Satt546, E 连锁群 Satt117, F 连锁群 Satt656、Satt252, G 连锁群的 Satt594, H 连锁群 Satt469, 其中 Satt594 贡献率达到 9.08%。油分和蛋白质含量同时关联到位点是 Satt005、Satt117、Satt565、Satt469、Satt594 和 Satt546 (表 3)。

表 2 与油分含量显著相关位点信息

Table 2 Marker loci associated with oil content				
名称 Name	连锁群 LG	解释率 Contribution/% R ²	等位基因个数 Alleles number	材料个数 Materials number
Sat_345	D1a	2.19	5	254
Satt656	F	2.36	3	316
Satt557	C2	2.73	3	296
Satt506	D1b	2.80	3	277
Satt274	D1b	2.82	4	266
Sat_346	D1a	2.91	4	234
Satt640	C2	3.12	5	263
Satt117	E	3.58	4	242
Satt005	D1b	3.57	4	254
Satt546	D1b	4.19	3	291
Satt469	H	8.85	6	268
Satt594	G	10.05	5	302

表 3 与蛋白质含量显著相关位点信息

Table 3 Marker loci associated with protein content				
名称 Name	连锁群 LG	解释率 Contribution /% R ²	等位基因个数 Alleles number	材料个数 Materials number
Satt005	D1b	2.65	4	254
Satt252	F	2.12	4	291
Satt656	F	2.18	3	317
Satt041	D1b	2.20	3	269
Satt300	A1	2.40	4	296
Satt640	C2	2.64	5	263
Satt537	D1b	3.51	5	278
Satt117	E	3.85	4	242
Satt546	D1b	4.40	3	291
Satt469	H	8.46	6	268
Satt594	G	9.08	5	302

3 讨论

群体结构遗传分析基于 Hardy-Weinberg 定律,

对数据数学聚类分析。群体结构能够影响群体的连锁不平衡,容易出现关联结果的假阳性^[10]。本研究选用东北地区育种材料 327 份,对试验材料群体结构分析,表明材料群体结构划分与地理生态分类有关,齐照明等^[6]对东北 92 份大豆主栽材料的群体结构划分,表明群体的划分与地理生态分类有关,张军等^[11]对中国东北大豆育成品种的群体机构分析表明,群体的划分与地理分群和遗传分类有关。由于实验材料间差异,群体划分也会有差异,但是东北地区大豆自然群体群体结构划分主要和地理生态分类有关。

关联分析方法已广泛应用于大豆蛋白质和油分的研究中。本研究结果的 5 个油分相关的位点与前人的研究相同,分别是 Satt557、Satt005、Satt274、Satt546 和 Satt117,其中 Satt557 与张军等^[11]定位的大豆油分含量 QTL 相同,其中 Satt274 与 Panthee 等^[12]和 Qi 等^[13]定位的大豆油分含量 QTL 相同, Satt005 与 Rossi 等^[14]定位的大豆油分含量 QTL 相同, Satt546 与 Kim 等^[15]定位的大豆油分含量 QTL 相同, Satt117 与 Chen 等^[16]定位的大豆油分含量 QTL 相同。蛋白质关联分析研究中,发现有 3 个相关位点与前人的研究一致,分别是 Satt117、Satt469 和 Satt537,其中 Satt117 和 Satt537 与 Chen 等^[16]定位的蛋白质含量 QTL 相同。Satt469 和 Lu 等^[17]定位的大豆蛋白质含量 QTL 相同。Satt252 是 Panthee 等^[12]研究与蛋氨酸和半胱氨酸含量相关的 QTL,本研究得出与大豆蛋白质含量相关。前人的研究较多采用重组自交系,本研究选用在自然群体,获得相同的结果,表明本研究获得关联位点真实可靠。

研究中发现 7 个位点与油分和蛋白质存在共同关联,分别是 Satt005、Satt117、Satt469、Satt546、Satt594、Satt640 和 Satt656,表明这些位点的关联基因极易受到环境和地理位置变化,影响品质形成。在 Williams82 遗传图谱中查找油分和蛋白质含量相关位点,与油分关联位点中有 4 个位于 D1b 连锁群中,与蛋白质关联位点中同样有 4 个位于 D1b 连锁群中。在蛋白质和油分中具有最高贡献率的引物是 Satt594,解释率分别为 10.05% 和 9.08%。在 Williams82 参考基因组中查找 Satt594 位点,其上游 3 cM 内找到 9 个基因,下游 3 cM 内找到 15 个基因。上游找到多个未知功能的蛋白,可能与油分的代谢合成有关。

许多学者已经利用重组自交群体在大豆蛋白质和油分含量的关联分析中定位到很多关联位点,由于使用特定的杂交群体材料,难以实现标记位点的广泛应用。本实验以自然群体为试验材料,打破了遗传群体中关联位点只适用于试验群体内的局限性,所关联出的位点可以应用于东北地区大豆各品种中。对后续的关联基因获取以及性状关联位

点开发和性状的分子标记选择提供了重要信息,为分子标记辅助育种奠定基础。

参考文献

[1] Corder E H, Saunders A M, Risch N J, et al. Protective effect of polio protein E type 2 allele for late on set Alzheimer disease [J]. Nature Genetics,1994,7:180-184.

[2] Chapman A, Pantalone V R, Ustun A, et al. Quantitative trait loci for agronomic and seed quality traits in an F₂ and F_{4,6} soybean population[J]. Euphytica, 2003, 129: 387-393.

[3] Meksem K, Njiti V N, Banz W J, et al. Genomic regions that underlie soybean seed isoflavone content[J]. Journal of Biomedicine and Biotechnology, 2001, 1:38-45.

[4] Jun T, van K, Kim M Y, et al. Association analysis using SSR markers to find QTL for seed protein content in soybean [J]. Euphytica, 2008,162:179-191.

[5] Shi A, Chen P, Zhang B, et al. Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States [J]. Plant Breeding, 2010, 3 (129): 250-256.

[6] 齐照明,侯萌,韩雪,等. 东北地区大豆主栽品种油分蛋白含量的关联分析 [J]. 中国油料作物学报, 2014,36(2):168-174. (Qi Z M, Hou M, Han X, et al. Correlation analysis in northeast soybean cultivars oil protein content[J]. Chinese Journal of Oil Crop Science, 2014,36(2): 168-174.)

[7] Department of Human Genetics University of Chicago. /http://pritch.bsd.uchicago.edu/structure.html/2010-2-2.

[8] Edward Buckler Lab. http://www.maizegenetics.net/bioinformatics/ta-ssel,2011-10-22.

[9] 文自翔,赵团结,郑永战,等. 中国栽培和野生大豆农艺品质性状与 SSR 标记的关联分析:1. 群体结构及关联标记[J]. 作物学报, 2008, 34(7): 1169-1178. (Wen Z X, Zhao T J, Zheng

Y Z, et al. Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: I. Population structure and associated markers[J]. Acta Agronomica Sinica, 2008, 34(7): 1169-1178.)

[10] Reinprecht Y, Poysa V, 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.

[11] 张军,赵团结,盖钧镒. 中国东北大豆育成品种遗传多样性和群体遗传结构分析 [J]. 作物学报, 2008, 34(9): 1529-1536. (Zhang J, Zhao T J, Gai J Y. Genetic diversity and genetic structure of soybean cultivar population released in Northeast China [J]. Acta Agronomica Sinica, 2008, 34(9): 1529-1536.

[12] Panthee D R, Pantalone V R, West D R, et al. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean[J]. Crop Science, 2005, 445(5):2015-2022.

[13] Qi Z M, Wu Q, Han X, et al. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes[J]. Euphytica, 2011, 179:499-514.

[14] Rossi M E, Orf J, Liu L J, et al. Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian × Chinese crosses [J]. Theoretical and Applied Genetics, 2013, 126(7):1809-1823.

[15] Kim H K, Kim Y C, Kim S T, et al. Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean[J]. Life Science, 2010, 20(8): 1186-1192.

[16] Chen Q S, Zhang Z C, Liu C Y, et al. QTL analysis of major agronomic traits in soybean [J]. Agricultural Sciences in China, 2007 (4):399-405.

[17] 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, 2012, 126(2). Doi: 10. 1007/s00122-012-1990-8.

(上接第 976 页)

[11] Niu C F, Wei W, Zhou Q Y, et al. Wheat WRKY genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic *Arabidopsis* plants [J]. Plant, Cell & Environment, 2012, 35(6):1156-1170.

[12] Berri S, Abbruscato P, Faivre-Rampant O, et al. Characterization of WRKY co-regulatory networks in rice and *Arabidopsis*[J]. BMC Plant Biology,2009, 9(1):120.

[13] Wang Q, Wang M, Zhang X, et al. WRKY gene family evolution in *Arabidopsis thaliana*[J]. Genetica,2011, 139(8):973-983.

[14] van Aken O, Zhang B, Law S, et al. *AtWRKY40* and *AtWRKY63* modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins [J]. Plant Physiology, 2013, 162(1):254-271.

[15] Shultz J L, Kurunam D, Shopinski K, et al. The Soybean Genome Database (SoyGD): A browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max* [J]. Nucleic Acids Research, 2006, 34(Suppl 1):D758-D765.

[16] Yin G, Xu H, Xiao S, et al. The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups [J]. BMC Plant Biology,2013, 13(1):148.

[17] Zhou Q Y, Tian A G, Zou H F, et al. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *Gm-*

WRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants [J]. Plant Biotechnology Journal,2008, 6(5):486-503.

[18] Luo X, Bai X, Sun X, et al. Expression of wild soybean WRKY20 in *Arabidopsis* enhances drought tolerance and regulates ABA signalling [J]. Journal of Experimental Botany, 2013, 8 (64):2155-2169.

[19] Zhang L, Wang X, Bi Y, et al. Isolation and functional analysis of transcription factor GmWRKY57B from soybean [J]. Chinese Science Bulletin,2008, 53(22):3538-3545.

[20] Hu Y, Dong Q, Yu D. *Arabidopsis* WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen *Pseudomonas syringae*[J]. Plant Science,2012, 185:288-297.

[21] Mukhtar M S, Nishimura M T, Dangl J. NPR1 in plant defense: It's not over'til it's turned over [J]. Cell, 2009, 137(5): 804-806.

[22] Fountain J C, Raruang Y, Luo M, et al. Potential roles of WRKY transcription factors in regulating host defense responses during *Aspergillus flavus* infection of immature maize kernels [J]. Physiological and Molecular Plant Pathology,2015, 89:31-40.

[23] Tripathi P, Rabara R C, Lin J, et al. *GmWRKY53*, a water-and salt-inducible soybean gene for rapid dissection of regulatory elements in BY-2 cell culture [J]. Plant Signaling & Behavior,2013, 8(5): e24097-e24102.