DOI: 10.3724/SP.J.1006.2021.04249

利用动态转录组学挖掘大豆百粒重候选基因

曾 健 徐先超 徐昱斐 王秀成 于海燕 冯贝贝 邢光南*

南京农业大学大豆研究所 / 国家大豆改良中心 / 农业农村部大豆生物学与遗传育种重点实验室(综合) / 作物遗传与种质创新国家重点实验室/江苏省现代作物生产协同创新中心, 江苏南京 210095

摘 要: 百粒重是影响大豆产量的重要农艺性状, 揭示其分子基础发掘关键候选基因对大豆改良具有重要意义。本研究通过对 12 个大豆品种籽粒发育 3 个时期共 36 个样本的转录组数据进行加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA), 得到 20 个基因共表达模块, 与百粒重及 4 个粒形性状关联后发现 green 模块与表型最为相关, 之后根据 Gene Significance (GS)值和 Eigengene Connectivity (*kME*)值筛选出 13 个 green 模块内的核心基因(hub gene); 然后对 2 组百粒重存在极显著差异的大豆品种的籽粒发育 3 个时期分别进行基因差异表达分析发现大豆在籽粒发育前中期可能通过 MAPK 信号通路调节百粒重大小; 之后对其进行 SNPs/InDels 挖掘并根据 Gene Ontology (GO)注释发现 green 模块内的 *Glyma.14G043900* 和 *Glyma.15G217400* 由于 SNP 变异造成同义以及非同义突变, 且存在调控基因表达相关的 GO Terms 以及锌指结构域, 表明它们可能通过调控 hub 基因和差异表达基因调控大豆百粒重及粒形性状。 *Glyma.15G217400* 位于已报道的 4 个百粒重 QTL 中, 而 *Glyma.14G043900* 位于已报道的一个籽粒蛋白含量及一个油脂含量 QTL 中。通过比对大豆公共数据库发现这 2 个基因的百粒重增效等位变异受到人工选择, 其频率从野生大豆到地方品种再到育成品种的过程中逐渐升高。这些结果为进一步发掘大豆百粒重候选基因及其表达调控机制提供了新思路。

关键词: 大豆; 百粒重; 转录组学; WGCNA; 候选基因

Utilization of dynamic transcriptomics analysis for candidate gene mining of 100-seed weight in soybean

ZENG Jian, XU Xian-Chao, XU Yu-Fei, WANG Xiu-Cheng, YU Hai-Yan, FENG Bei-Bei, and XING Guang-Nan*

Soybean Research Institute, Nanjing Agricultural University / National Center for Soybean Improvement / Key Laboratory for Biology and Genetic Improvement of Soybean (General), Ministry of Agriculture and Rural Affairs / National Key Laboratory for Crop Genetic and Germplasm Enhancement / Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing 210095, Jiangsu, China

Abstract: 100-seed weight of soybean is an important agronomic trait that affects yield, and it is of great significance to reveal its molecular basis and discover key candidate genes for soybean improvement breeding. In this study, weighted gene co-expression network analysis (WGCNA) was performed on the transcriptome data of 36 samples from 12 soybean varieties at three stages of seed development, and 20 gene co-expression modules were obtained. After correlating with 100-seed weight and four-seed shape traits, the green module was found to be most correlated with the phenotypes. Then 13 hub genes of green module were screened

本研究由国家重点研发计划项目(2016YFD0100201), 国家自然科学基金项目(31571694), 中央高校基本科研业务费专项资金(KYT201801), 长江学者和创新团队发展计划(PCSIRT_17R55), 高等学校学科创新引智基地项目 111 (B08025), 国家现代农业产业技术体系(大豆)建设专项(CARS-04)和江苏省优势学科建设工程专项和江苏省 JCIC-MCP 项目资助。

This study was supported by the National Key Research and Development Program of China (2016YFD0100201), the National Natural Science Foundation of China (31571694), the Fundamental Research Funds for Central Universities (KYT201801), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT_17R55), the Project of Intellectual Base for Discipline Innovation in Colleges and Universities (B08025), the China Agriculture Research System (CARS-04), the Jiangsu Higher Education PAPD Program, and the Jiangsu JCIC-MCP Program.

* 通信作者(Corresponding author): 邢光南, E-mail: xinggn@njau.edu.cn

第一作者联系方式: E-mail: 2019101129@njau.edu.cn

Received (收稿日期): 2020-11-20; Accepted (接受日期): 2021-03-19; Published online (网络出版日期): 2021-04-01.

URL: <https://kns.cnki.net/kcms/detail/11.1809.S.20210401.0925.002.html>

based on the Gene Significance (GS) and Eigengene Connectivity (*kME*) value. Gene differential expression of two groups of soybean varieties with extremely significant differences in 100-seed weight showed that the MAPK signaling pathway in the early and mid-term of seed development might regulate the 100-seed weight in soybean. According to SNPs/InDels calling and Gene Ontology (GO) annotation, *Glyma.14G043900* and *Glyma.15G217400* in the green module caused synonymous and non-synonymous coding mutations due to SNP mutations, and there were GO Terms and zinc finger domains related to gene expression regulation. These results suggested that they might regulate the 100-seed weight and seed shape of soybeans by regulating the hub gene and differentially expressed genes. Furthermore, *Glyma.15G217400* was located in four reported QTLs of 100-seed weight, while *Glyma.14G043900* was located in a reported seed protein content QTL and an oil content QTL. Compared with soybean public database, the increasing 100-seed weight alleles of the two genes were artificially selected and their frequency was gradually increased from wild accessions to landraces, resulting in the improved cultivars. These results provide new ideas for further discovering 100-seed weight candidate gene in soybean and its expression regulation mechanism.

Keywords: soybean; 100-seed weight; transcriptomics; WGCNA; candidate gene

百粒重是影响大豆产量的重要性状, 遗传育种家常利用连锁定位和关联定位的方法从基因组水平解析大豆百粒重的遗传机制, 如 Lu 等^[1]利用栽培大豆和野生大豆构建的重组自交系群体进行全基因组重测序及数量性状位点(quantitative trait locus, QTL)定位, 从野生大豆中克隆了 1 个控制大豆百粒重的基因 *PP2C-1*; Li 等^[2]通过对大豆籽粒百粒重、粒长和粒宽的全基因组关联分析(genome-wide association study, GWAS)发现了 1 个新的控制籽粒大小的位点 *SW9-1*。由于大豆百粒重性状具有复杂的遗传基础, 由多位点控制, 且在不同种植条件下表型变异大, 使得挖掘百粒重相关的基因比较困难。

随着测序技术的发展与测序成本的降低, 越来越多的基因组学、转录组学以及代谢组学方法被用来进行复杂生物学问题的研究, 其中以转录过程作为研究对象是最先发展且应用最广的组学技术, 如物种鉴定、种质资源挖掘、遗传多样性分析以及植物系统分类等方面^[3]。然而, 随着越来越多的 RNA 测序(RNA sequencing, RNA-seq)数据被公布, 如何处理如此海量的数据, 挖掘出有用的信息成为了研究人员新的研究目标。RNA-seq 最常用于分析 2 组样本间的差异表达基因(differential expression gene, DEG), 但是传统的成对比较不能有效反映所有样本的整体动态特性^[4]。面对这个问题, 网络研究的方法脱颖而出, 如基因调控网络、蛋白调控网络和代谢网络等^[5]。在基因调控网络中, 应用比较广泛的是加权基因共表达网络(weighted gene co-expression network analysis, WGCNA)^[5]。加权基因共表达网络分析可以有效地利用大量转录组数据将全基因组的基因分类为不同的基因共表达模块, 然后进一步研究共表达模块与性状之间的相关性。该方法特别适合研究处于不同发育阶段的多个样品^[6-7], 是研究

多样本、大数据的高效方法。例如, Yang 等^[8]利用 5 个时期的大豆转录组数据构建加权基因共表达网络, 成功鉴定出一批与大豆油脂合成相关的基因。Lu 等^[9]利用栽培及野生大豆共 40 个转录组数据进行 WGCNA, 鉴定出了 1 个限制赤霉素合成的基因 *GmGA20OX*, 它调控大豆籽粒重量及大小。

通过 RNA-seq 数据不仅能检测基因表达量, 还可以检测基因的等位变异。等位变异常被用做分子标记研究复杂生物学问题。例如传统的 GWAS 基于全基因组重测序技术获取数以百万计位点的等位变异作为分子标记, 将其与表型关联从而鉴定与表型相关的基因座。GWAS 是研究复杂表型遗传基础的有效手段, 但其总成本十分昂贵^[10]。因而之后出现了简化测序技术, 如基因组测序之前用限制酶处理以及外显子组测序等^[11]。RNA-seq 也是降低基因组复杂性的一种经济有效的方法^[11]。虽然基于 RNA-seq 的 SNPs/InDels 挖掘只能检测来自转录区域的变异信息, 但是大多数的来自转录区域的 SNPs/InDels 才是直接影响功能的变异^[10-11]。并且基于 RNA-seq 挖掘特定区域(如转录区域)的 SNPs/InDels 能提供非常高深度的变异信息, 还能检测到高表达基因中低频的变异^[10-11]。

随着大豆泛基因组公布^[12], 为大豆育种提供了许多新的生物信息学资源。本研究利用泛基因组中公布的籽粒发育转录组数据, 结合品种间百粒重存在极显著差异的 12 个大豆品种的百粒重以及粒形数据, 通过转录组数据从 WGCNA、基因差异表达分析以及 SNPs/InDels 挖掘 3 个方面分析这些品种不同时期的基因表达差异及其序列差异, 探索大豆百粒重基因表达调控机制, 以期挖掘出大豆百粒重特异相关模块、候选基因及其等位变异, 为进一步研究大豆百粒重及其他重要农艺性状分子机制提供新的思路。

1 材料与方法

1.1 供试材料及表型数据获取

供试大豆材料为晋豆 23、邯豆 5 号、克山 1 号、齐黄 34、丰地黄、铜山天鹅蛋、莒选 23、Amsoy、东农 50、十胜长叶、豫豆 22 和铁丰 18, 以上材料由中国农业科学院作物科学研究所提供。试验于 2018 年 6 月 15 日和 2019 年 6 月 17 日在安徽省马鞍山市当涂县南京农业大学皖江基地(118°37'22"E, 31°33'4"N)进行, 行播, 每个品种 2 行, 行距 50 cm, 行长 2 m, 试验地四周种 3 行保护行。常规田间管理, 这 12 份品种在当涂县夏播都能正常成熟, 且熟期差异不太大, 2 年分别为 83~101 d 和 94~112 d。

大豆品种成熟后, 每个品种随机选 10 个单株, 单株收获, 脱粒, 分单株测量籽粒性状。采用籽粒图像采集和计算机软件处理的方法^[13]对籽粒粒形进行测定并计数粒数。丁琦等^[13]使用良田高拍仪在 LED 背光板上对大豆籽粒图像进行采集, 基于分水岭和统计矩开发了测量大豆籽粒粒形的软件, 可获得粒数、粒长、粒宽、粒周长和粒面积等数据。用电子天平称其重量, 精确到千分之一, 用公式百粒重=粒重/粒数×100 计算出百粒重。运用 Excel 2010 对获取的数据进行整理, 用 SAS 9.4 软件对 2 年数据进行联合方差分析和 Duncan's 多重比较。

1.2 转录组数据的获取与加权基因共表达网络分析

大豆不同时期籽粒组织转录组数据来自 Biological Project Library (BioProject, [https://bigd.](https://bigd.big.ac.cn/bioproject/)

[big.ac.cn/bioproject/](https://bigd.big.ac.cn/bioproject/)), 获取码为 PRJCA002030^[12]。种子开始结荚记为第 1 天, 转录组测序取样时间分别为在大豆籽粒发育的前期(G, 种子发育第 4 周前)、中期(H, 种子发育第 6 周)和后期(I, 种子发育第 8 周)进行取样, 共 3 个不同时期。大豆参考基因组和注释基因组 Williams 82 v2.0 都来自 SoyBase (<https://www.soybase.org/>)。选择东农 50、Amsoy、莒选 23、十胜长叶、豫豆 22、铁丰 18、丰地黄、铜山天鹅蛋、齐黄 34、邯豆 5 号、晋豆 23 和克山 1 号共 12 个品种的籽粒发育 3 个时期转录组数据, 共 36 个样本进行加权基因共表达网络构建。首先使用 HISAT2 v2.0.5 将 reads 比对到参考基因组, 然后用 StringTie v2.1.1 定量转录本^[14], 最后用 R v4.0.2 中的 DESeq2 包^[15]的 varianceStabilizingTransformation 函数对 36 个样本进行标准化后得到基因表达矩阵, 去除在所有样本中表达量均为 0 的基因。使用 R v4.0.2 中 WGCNA 包的 biweight mid-correlation 计算基因相关性, pickSoftThreshold 函数计算软阈值, 确定阈值 β 为 5 (图 1), 并且模块内基因数最少为 30, 一步法构建 Signed Networks^[16]。根据模块内基因与性状相关性指标 Gene Significance (GS)值以及模块内基因与模块相关性指标 Eigengene Connectivity (kME) 值筛选核心基因 (hub gene), 拓扑重叠性 (weight) 值代表模块内两基因间相关性。使用 R 语言中的 pheatmap 包进行基因表达量可视化。GO 注释 v2.0 来自 SoyBase, clusterProfiler 包^[17]计算 Gene Ontology (GO)富集以及结果可视化($p < 0.01$, $q < 0.05$)。Cytoscape v3.7.1^[18]可视化共表达网络。

图 1 软阈值 β 的确定

Fig. 1 Determination of soft threshold β

A 图纵坐标是无尺度网络模型指数, 红线代表 $R^2=0.85$ 。B 图纵坐标每一个软阈值对应的平均连接度。图 A 和图 B 的横坐标均代表软阈值 β 。

The ordinate represents the index of scale free network model in figure A, the red line represents that R^2 is equal to 0.85. The ordinate represents the average link degree of each soft threshold in figure B. The abscissa in both figures A and B represent the soft threshold β .

1.3 基因差异表达分析

根据 Duncan’s 多重比较结果, 选择百粒重存在极显著差异的 2 组大豆品种, 分别对籽粒发育 3 个时期进行基因差异表达分析, 其中 2 年平均百粒重小于 12 g 的小粒品种组包括东农 50、Amsoy 和莒选 23, 2 年平均百粒重大于 18 g 的大粒品种组包括邯豆 5 号、晋豆 23 和克山 1 号。经 HISAT2、StringTie 和 DESeq2 处理得到 DEGs, $P_{adj}<0.05$, $|\log_2(\text{Fold Change})|>1$, Fold Change 代表基因差异表达倍数。使用 R 中的 clusterProfiler 包计算 Kyoto Encyclopedia of Genes and Genomes (KEGG)富集及结果可视化。

1.4 SNPs/InDels 挖掘

选择百粒重存在极显著差异的小粒品种东农

50、Amsoy、莒选 23 与大粒品种邯豆 5 号、晋豆 23、克山 1 号进行 SNPs/InDels 挖掘。参考 Engström 等^[19]的建议, 首先使用 STAR v2.5.3a^[20]将 reads 比对到参考基因组 Williams 82 v2.0 上。SAMtools v1.6^[21]软件构建索引, Picard Toolkit v2.22.0 (<http://broadinstitute.github.io/picard/>)对序列排序、标记并去除重复。再使用 Genome Analysis Toolkit (GATK, v4.1.0.0)进行 SNPs/InDels 检测^[22], 保留满足表 1 中所有条件的等位变异, 同时去除杂合等位变异, 然后使用 vcftools v0.1.15^[23]去除含缺失的等位变异, 从而获得高质量的 SNPs/InDels。最后使用 snpEff^[24]进行 SNPs/InDels 注释。使用 Circos v0.69-9^[25]绘制 SNPs/InDels 以及 DEGs 染色体分布图。

表 1 SNPs/InDels 过滤标准
Table 1 SNPs/InDels filter standards

SNPs 过滤标准 SNPs filter standards	InDels 过滤标准 InDels filter standards
QualByDepth (QD) > 2.0	QualByDepth (QD) > 2.0
RMSMappingQuality (MQ) > 40.0	—
FisherStrand (FS) < 60.0	FisherStrand (FS) < 200.0
StrandOddsRatio (SOR) < 3.0	StrandOddsRatio (SOR) < 10.0
MappingQualityRankSumTest (MQRankSum) > -12.5	MappingQualityRankSumTest (MQRankSum) > -12.5
ReadPosRankSumTest (ReadPosRankSum) > -8.0	ReadPosRankSumTest (ReadPosRankSum) > -8.0

满足以上所有条件的等位变异被保留。Alleles satisfying all the above conditions are retained.

1.5 驯化与选择

从 SoyBase 下载 Davoud 等^[26]发布的 1007 份大豆(包含育成品种 895 份和地方品种 112 份)的重测序等位变异数据, 以及 45 份野生豆重测序数据 (https://sobase.org/data/public/Glycine_max/Wm82.gnm2.div.G787), 分析 green 模块的 *Glyma.14G043900* 和 *Glyma.15G217400* 内等位变异频率从野生大豆到地方品种再到育成品种过程中的变化。

2 结果与分析

2.1 大豆品种间百粒重的变异

根据 12 个大豆品种百粒重的箱型图(图 2)发现, 东农 50、Amsoy、莒选 23、十胜长叶、豫豆 22 和铁丰 18 为百粒重较小的品种; 丰地黄、铜山天鹅蛋、齐黄 34、邯豆 5 号、晋豆 23 和克山 1 号为百粒重较大的品种。通过 Duncan’s 多重比较发现, 当 $P=0.01$ 时, 东农 50、Amsoy、莒选 23 和邯豆 5 号、晋豆 23、克山 1 号的百粒重间存在极显著差异, 可分成小籽粒和大籽粒 2 组大豆品种。

图 2 12 个大豆品种百粒重的箱型图
Fig. 2 Box diagram of 100-seed weight of 12 soybean varieties
横坐标代表 12 个大豆品种, 纵坐标表示百粒重。一个点代表一次观察值。品种间不同大写字母说明百粒重基于 Duncan 氏新复极差法在 $P=0.01$ 水平的差异显著性。
The abscissa represents 12 soybean varieties, and the ordinate represents the 100-seed weight. One point represents one observation. Soybean varieties with different uppercase letters are significantly different in 100-seed weight at $P=0.01$ by Duncan’s new multiple-range test.

2.2 大豆籽粒发育的加权基因共表达网络分析
基于 12 个大豆品种 3 个籽粒发育时期共 36 个

样本的转录组数据进行表达量分析，获得 50,595 个基因的标准化表达矩阵。构建加权基因共表达网络后获得 20 个模块(图 3)，模块内基因数量为 33 (lightyellow)到 13,680 (grey)个，其中 grey 模块代表未分配到任何其他模块的基因。该模块内可能包含大量调控植物基本生命活动所必需的管家基因，不随着时间和品种变化，因而未能分配到其他共表达

模块。
通过百粒重及粒形相关性状与模块的关联发现，green 模块与 5 个籽粒性状均有显著关联($P<0.05$) (图 4-A)。根据聚类热图，green 模块内 2904 个基因可以分为 2 类，当第一类的基因在大豆发育过程中高表达时，另一类别则显示出较低的表达，反之亦然(图 4-B)，这可能暗示 green 模块内基因的正负调节关系。

图 3 模块内基因数目柱状图
Fig. 3 Histogram of the number of genes in modules
横轴代表模块，纵轴代表模块内基因数目。
The abscissa represents the module and the ordinate represents the number of genes in the module.

图 4 模块与性状关联热图(A)及 green 模块内 2904 个基因的表达量热图(B)
Fig. 4 Heat map of module-trait relationships (A) and gene expression count of 2904 genes in green module (B)
A 图中横轴表示百粒重及粒形性状，纵轴表示每一个模块的特征向量；红色的格子代表性状与模块具有正相关性，蓝色的格子代表性状与模块具有负相关性。B: green 模块内 2904 个基因在 12 个大豆品种 3 个时期表达量行标准化后绘制。G、H 和 I 分别为籽粒发育前期、中期和后期。
The horizontal axis represents 100-seed weight and seed shape traits, and each column represents the co-expression module in the figure A; red color of each box represents the positive correlation between module and trait, blue color of each box represents the negative relationships between module and trait. B: 2904 genes in the green module are mapped after normalization of the expression levels by row of 12 soybean varieties in three stages. G, H, and I represent the early stage, the middle stage, and the later stage of seed development, respectively.

GO 富集结果显示, green 模块内显著富集到与发育相关的生物功能, 如脱落酸介导的信号通路(GO:0002679)、器官形态发生(GO:0010089)以及发育性生长(GO:0007346)等。并且发现 green 模块内显著富集的基因大多与基因沉默、转录和转录调控(包

括依赖 DNA 的转录调控、mRNA 加工、DNA 结合、胞质 mRNA 加工体和 RNA 聚合酶 II 转录辅因子活性等)以及细胞蛋白的降解修饰等相关(图 5), 这可能暗示这些 GO Terms 在大豆籽粒生长发育过程中发挥了重要作用。

图 5 green 模块内 2904 个基因的 GO 富集
Fig. 5 GO enrichment of 2904 genes in green module
横坐标(每列)代表富集基因占输入基因比例, 纵坐标(每行)代表 GO Terms。点的大小代表富集的基因个数, 点的颜色代表经费舍尔精确测验校正后的 P_{adjust} 值。
The abscissa (each column) represents the proportion of enriched genes to input genes, and the ordinate (each row) corresponds to GO Terms. The size of the dot represents the number of enriched genes, and the color of the dot represents the corrected P_{adjust} value by the Fisher exact test.

2.3 不同百粒重大豆品种组间籽粒发育时期的基因差异表达

通过对 2 组百粒重存在极显著差异的大豆品种分 3 个时期进行基因差异表达分析发现, 差异表达基因集中在籽粒发育前期和中期, 籽粒发育后期基因差异表达数目明显减少。其中前期 G 有 533 个基因差异表达, 相比大百粒重组品种, 小百粒重组品种中 442 个基因上调, 91 个基因下调; 中期 H 有 709 个基因差异表达, 小百粒重组中 534 个基因上调, 175 个基因下调; 后期 I 有 53 个基因差异表达, 小百粒重组中 21 个基因上调, 32 个基因下调(图 6-A)。3

个时期中上调基因共 961 个, 下调基因共 284 个, 上调基因没有显著富集到任何 KEGG 通路, 而下调基因仅显著富集到植物 MAPK 信号通路(图 6-B)。

2.4 与大豆百粒重相关的核心基因和差异表达基因

根据基因与 5 个性状的 GS 值(大于 0.2)和与 green 模块的 kME 值(大于 0.8), 共筛选到 green 模块内的 13 个 hub 基因(图 7-A)。同时, green 模块内存在 22 个 DEGs, 通过聚类热图发现, 这 22 个 DEGs 在百粒重存在极显著差异的 2 组大豆品种中的表达模式分为 2 种。第 1 种包含 7 个 DEGs, 其在小粒品

种籽粒发育中期显著上调表达(图 7-B)。第 2 种包含 15 个 DEGs, 它们在大粒品种籽粒发育的前中期显著上调表达(图 7-B)。其中 *Glyma.16G103400* 同时属于 hub 基因和 DEGs。

2.5 不同百粒重大豆品种组间 SNPs/InDels 挖掘

为进一步揭示 green 模块内基因表达调控机制, 用百粒重差异极显著的 2 组大豆品种进行 SNPs/

InDels 挖掘, 共检测到 52,395 个 SNPs 和 6713 个 InDels, 且发现 DEGs 与 SNPs/InDels 大致位于染色体相同区域(图 8)。其中 SNPs 主要分布于基因外显子区域, 但大多造成同义突变, InDels 主要分布于基因上下游区域(表 2)。在大粒和小粒 2 组大豆品种的组间共发现 10 个非同义突变 SNPs 位于 green 模块的 10 个基因内(表 3)。

图 6 百粒重存在极显著差异的 2 组大豆品种籽粒发育不同时期基因的表达差异(A)及小粒品种下调基因的 MAPK 信号通路富集图(B)
Fig. 6 Gene expression differences in different stages of seed development between two panels of soybean varieties with extremely significant differences in 100-seed weight (A) and KEGG pathway enrichment map of down-regulated genes in small 100-seed weight varieties (B)

G、H 和 I 分别代表籽粒发育前期、中期和后期。Down-regulated genes 代表小粒品种中下调的基因, Up-regulated genes 代表小粒品种中上调的基因。

G, H, and I represent the early stage, the middle stage, and the later stage of seed development, respectively. Down-regulated genes represent down-regulated genes in small 100-seed weight varieties, and up-regulated genes represent up-regulated genes in small 100-seed weight varieties.

图 7 green 模块内 hub 基因 GS 值和 kME 值的柱形图(A)及不同百粒重大豆品种组间差异表达基因的表达量聚类热图(B)
Fig. 7 Histogram of GS and kME value of hub gene in green module (A) and cluster heat map of differentially expressed genes among soybean varieties with different 100-seed weight (B)

100-SW: 百粒重; SA: 籽粒面积; SL: 籽粒长度; SP: 籽粒周长; SW: 籽粒宽度。G、H 和 I 分别代表籽粒发育前期、中期和后期。Large 和 Small 分别代表 $P = 0.01$ 差异极显著的大粒大豆和小粒大豆品种组。图 B 中加粗的基因代表第 1 种表达模式的基因。

100-SW: 100-seed weight; SA: seed area; SL: seed length; SP: seed perimeter; SW: seed width. G, H, and I represent the early stage, the middle stage, and the later stage of seed development, respectively. Large and Small represent large soybean varieties panel at $P = 0.01$, respectively. Gene in bold represents the first expression pattern in figure B.

图 8 SNPs/InDels 及 DEGs 在大豆染色体的分布图
Fig. 8 Distribution map of SNPs/InDels and DEGs on soybean chromosomes
A: SNPs 密度; B: InDels 密度; C: $\log_2$ (Fold Change) of DEGs, 内圈为在小粒品种组中显著上调基因, 外圈为下调基因。
A: SNPs density; B: InDels density; C: $\log_2$ (Fold Change) of DEGs. The inner circle represents significantly up-regulated genes in small 100-seed weight varieties, and the outer circle represents significantly down-regulated genes in small 100-seed weight varieties.

表 2 SNPs/InDels 分布区域
Table 2 SNPs/InDels distributed region

变异位置 Variation region	SNPs		InDels	
	数目 Count	比例 Percentage (%)	数目 Count	比例 Percentage (%)
DOWNSTREAM	41,636	23.67	5788	26.22
EXON	72,950	41.48	3028	13.72
INTERGENIC	2100	1.19	424	1.92
INTRON	8576	4.88	1625	7.36
SPLICE_SITE_ACCEPTOR	31	0.02	11	0.05
SPLICE_SITE_DONOR	31	0.02	12	0.05
SPLICE_SITE_REGION	334	0.19	69	0.31
TRANSCRIPT	—	—	1	0.01
UPSTREAM	29,780	16.93	4128	18.70
UTR_3_PRIME	14,683	8.35	4400	19.94
UTR_5_PRIME	5749	3.27	2586	11.72

加粗代表变异位置中 SNP 或 InDel 最大的数目以及其所占百分比。
Values in bold represent the largest count and percent of SNP or InDel in variation region.

表 3 Green 模块内在大粒和小粒品种组间存在非同义 SNP 突变的 10 个基因
Table 3 Ten genes with non-synonymous_coding SNP mutants between the large and small 100-seed weight varieties panels in green module

基因 Gene	染色体 Chromosome	位置 Position	小粒品种组等位变异 Allele of small 100-SW	大粒品种组等位变异 Allele of large 100-SW
<i>Glyma.02G139500</i>	Gm02	14480344	G	A
<i>Glyma.02G140100</i>	Gm02	14531600	T	C
<i>Glyma.06G179600</i>	Gm06	15246959	A	G
<i>Glyma.07G018200</i>	Gm07	1456246	A	G
<i>Glyma.12G057100</i>	Gm12	4156207	T	C
<i>Glyma.14G043900</i>	Gm14	3335050	T	C
<i>Glyma.15G217400</i>	Gm15	35930038	A	T
<i>Glyma.17G071400</i>	Gm17	5580413	A	G
<i>Glyma.17G076400</i>	Gm17	5974992	C	T
<i>Glyma.18G237000</i>	Gm18	52573168	T	C

加粗基因代表候选基因及其相关变异信息。100-SW 代表百粒重。
Values in bold represent candidate genes and related variation information. 100-SW represents 100-seed weight.

根据 GO 注释发现, 其中 *Glyma.14G043900* 与转录调控相关, *Glyma.15G217400* 与 3'UTR 介导的 mRNA 稳定调控相关(表 4), 并且 *Glyma.14G043900* 和 *Glyma.15G217400* 内均存在锌指结构域。*Glyma.14G043900* 内非同义突变发生在锌指结构域外部, 可能会影响该基因的转录, 而 *Glyma.15G217400* 内非同义突变发生在锌指结构域内部, 可能对该基因功能造成影响。基于基因共表达网络关系发现, green 模块内的 13 个 hub 基因和 22 个 DEGs 与这 2 个基因存在共表达关系(图 9)。其中, 13 个 hub 基因及在大粒品种籽粒发育前中期显著上调表达的 15 个 DEGs 与这 2 个基因间 weight 大于 0.1, 而在小粒品种籽粒发育中期显著上调表达的 7 个 DEGs 与这 2 个基因的 weight 值小于 0.1 但大于 0.01。

基于 *Glyma.14G043900* 和 *Glyma.15G217400* 的 GO 注释, 及其在 2 组大豆品种间的非同义突变, 及其与 hub 基因和 DEGs 的共表达关系, 推测其可能是控制大豆百粒重的候选基因。

2.6 候选基因的驯化和人工选择

通过比对公共数据库的 1052 份大豆 SNP 变异信息发现, green 模块中的 *Glyma.14G043900* 内 Gm14:3335050 和 *Glyma.15G217400* 内 Gm15:35930038 存在非同义 SNP 变异(表 3), 其中 Gm14:3335050 位点的百粒重增效等位变异(C)频率从野生大豆 42%到地方品种 57%再到育成品种 67%的过程中明显上升(图 10)。Gm15:35930038 位点的百粒重增效等位变异(T)频率从野生大豆 30%到地方品种 38%再到育成品种 45%的过程中也持续上升。

表 4 候选基因 *Glyma.14G043900* 和 *Glyma.15G217400* 的 GO 注释
Table 4 GO annotations of candidate genes *Glyma.14G043900* and *Glyma.15G217400*

基因 Gene	GO 编号 GO ID	GO 类型 GO type	注释描述 Annotation description
<i>Glyma.14G043900</i>	GO:0006355	生物过程 Biological process	以 DNA 为模板的转录调控 Regulation of transcription, DNA-templated
	GO:0008270	分子功能 Molecular function	锌离子结合 Zinc ion binding
	GO:0061158	生物过程 Biological process	3'端非翻译区介导的 mRNA 失稳 3'-UTR-mediated mRNA destabilization
<i>Glyma.15G217400</i>	GO:0005829	细胞组分 Cellular component	胞质溶胶 Cytosol
	GO:0003730	分子功能 Molecular function	mRNA 3'端非翻译区结合 mRNA 3'-UTR binding
	GO:0046872	分子功能 Molecular function	金属离子结合 Metal ion binding

图 9 green 模块内候选基因、hub 基因以及差异表达基因的共表达网络

Fig. 9 Co-expression network of candidate genes, hub genes, and differentially expressed genes in the green module

红色点代表候选基因, 黄色点代表 hub 基因, 绿色点代表 DEGs, 蓝色点代表既是 DEGs 同时又是 hub 基因。线的颜色深浅代表基因间的 weight 值大小。黑色的线代表基因间的 weight 大于 0.1, 灰色线代表基因间的 weight 小于 0.1 但大于 0.01。

Red dots represent candidate genes, yellow dots represent hub genes, green dots represent DEGs, and blue dot represents both DEG and hub gene. The color of the line represents the weight value between genes. Black line represents the weight between genes greater than 0.1, and gray line represents the weight less than 0.1 but greater than 0.01.

图 10 不同大豆类型中 *Glyma.14G043900* 和 *Glyma.15G217400* 非同义 SNP 突变等位变异的频率变化

Fig. 10 Allele percentage line chart of two non-synonymous coding SNP mutant of *Glyma.14G043900* and *Glyma.15G217400* in different soybean types

圆点代表百粒重的增效等位变异, 而三角形代表百粒重的减效等位变异。

The dots represent the positive effect alleles of 100-seed weight, and the triangles represent the negative effect alleles of 100-seed weight.

并且 *Glyma.14G043900* 内 Gm14:3335660 还存在同义突变, 其对应的百粒重增效等位变异(T)频率从野生大豆 45%增加到地方品种 100%和育成品种 99%。*Glyma.15G217400* 内 Gm15:35914738、Gm15:35914891 及 Gm15:35920369 三个位点也存在同义突变, 其中 Gm15:35914738 信息在野生大豆中未查询到, 但其增效等位变异(C)频率从地方品种的 10%增

加到育成品种 20%, Gm15:35914891 位点的百粒重增效等位变异(T)频率从野生大豆 10%增加到地方品种 14%和育成品种 19%, Gm15:35920369 位点的百粒重增效等位变异(T)频率从野生大豆 10%增加到地方品种 15%和育成品种 22%。从野生大豆到地方品种再到育成品种的非同义和同义 SNP 等位变异频率变化都说明这 2 个基因受到了明显的驯化和人工选

择, 特别是 *Glyma.14G043900*。

3 讨论

3.1 大豆百粒重及籽粒性状的加权基因共表达网络分析

通过构建加权基因共表达网络获得了 20 个基因共表达模块, 将百粒重及粒形性状与共表达模块关联, 发现 green 模块与这 5 个性状都显著相关 ($P < 0.05$)。由于 green 模块内基因的共表达且与 5 个籽粒性状密切相关, 因此 green 模块内显著富集的 GO Terms 可能与粒重及粒形密切相关。在这些 GO Terms 中, 脱落酸信号通路 (GO:0002679) 介导种子成熟过程, 成熟后期休眠诱导取决于强脱落酸信号, 且在拟南芥中脱落酸信号通路还通过 HAIKU 途径调节胚乳发育来控制种子大小^[27-28]。此外, 几种转录调节因子已被确定为植物种子大小控制的重要因素, 包括转录因子、转录共激活因子和参与染色质修饰的调控因子^[27], 如 APETALA2 型转录因子 *SMALL ORGAN SIZE1 (SMOS1)* 调控水稻籽粒大小^[27,29]; WRKY 家族转录因子 *WRKY15a* 的表达水平与野生大豆和栽培大豆的种子大小变化相关^[27,30]。加权基因共表达网络模块中, 代表基因和性状相关性的 GS 值以及基因和模块相关性的 *kME* 值可以有效筛选 hub 基因^[8]。本研究根据 GS 值和 *kME* 值确定了与 5 个籽粒性状高度相关的 13 个 hub 基因, 它们大多与转录、调控、蛋白质结合、氧化还原酶活性和生长素激活的信号通路等 GO Terms 相关, 并且已有研究表明这些 GO Terms 在调控种子大小过程起到重要作用^[27,31-32]。由于材料的遗传背景不同以及种植条件等因素, 影响大豆百粒重的基因 *PP2C-1*^[1]和 *GmGA20OX*^[9]未分类到 green 模块。

3.2 不同百粒重大豆品种组间基因差异表达分析

3 个不同熟期大豆品种的干物质积累都呈 S 型曲线变化, 籽粒形成期干物质始终是增长的, 都是在前期增长速率较快, 大部分干物质是在发育初期所积累^[33]。本研究选取种子发育阶段的 3 个不同时期包含了影响大豆百粒重的关键时期, 百粒重存在极显著差异的大豆品种组间籽粒的差异表达基因主要集中在籽粒发育前中期, 说明前中期是大豆籽粒发育积累营养物质的主要时期。在 KEGG 富集中, 小百粒重品种组中的上调基因没有富集到显著的 KEGG 通路, 下调基因仅在植物中 MAPK 信号通路显著富集。有研究表明 MAPK 信号通路在拟南芥和

水稻种子形成过程对种子大小中起重要作用^[27,31]。

例如在水稻中, *OsMKKK10*、*OsMKK4* 和 *OsMAPK6* 充当级联调节谷物大小^[34]。KEGG 富集结果表明, MAPK 信号通路可能也在大豆籽粒发育中调控籽粒大小。

3.3 大豆百粒重的候选基因

本研究基于 2 组百粒重存在极显著差异的大豆品种发现, 10 个 green 模块中存在非同义突变的基因。根据 GO 注释, 10 个基因中, *Glyma.14G043900* 存在依赖 DNA 的转录调控 GO Term (GO:0006355), 说明该基因发生的非同义突变可能造成转录调控功能受到影响; *Glyma.15G217400* 存在 3'UTR 介导使 mRNA 不稳定的 GO Term (GO:0061158), 说明该基因可能与 UTR 介导的基因表达调控过程相关^[35]。且 *Glyma.15G217400* 位于已经报道的控制粒重的 4 个 QTL 中 (SoyBase 编号为 *Seed weight 2-3*^[36]、*Seed weight 6-1*^[37]、*Seed weight 32-2*^[38]和 *Seed weight 49-11*^[39]), 而 *Glyma.14G043900* 虽然没有位于粒重及粒形性状 QTL 中, 但它位于一个籽粒油脂含量 QTL (SoyBase 编号为 *Seed oil 45-6*^[40])以及蛋白含量 QTL (SoyBase 编号为 *Seed protein 45-1*^[40])中。在 green 模块的共表达网络中, *Glyma.14G043900* 和 *Glyma.15G217400* 与 13 个 hub 基因及 green 模块中第 2 种表达模式的 15 个 DEGs (即在大粒品种组中显著上调表达的 DEGs)存在强共表达, 这暗示了它们之间的表达调控关系。并且根据已公布的 1052 份大豆 SNP 变异数据发现, *Glyma.14G043900* 和 *Glyma.15G217400* 对应的非同义突变位点 Gm14:3335050 和 Gm15:35930038 以及同义突变位点 Gm14:3335660、Gm15:35914891 和 Gm15:35920369 中, 百粒重增效等位变异从野生大豆到地方品种, 再到育成品种的过程中整体呈现上升趋势, 可能暗示这 2 个位点的等位变异频率受到了人为的驯化和人工改良选择。百粒重是影响大豆产量的重要性状, 本研究发掘出的百粒重候选基因可用于预测大豆百粒重, SNP 可用于开发 KASP 标记, 从而为选育适合不同用途的大豆品种如小粒豆芽专用品种和大粒菜用大豆品种提供依据与标记。

4 结论

利用 WGCNA、基因差异表达分析及 SNPs/InDels 挖掘, 共发现 36 个基因可能参与调控大豆百粒重。其中 *Glyma.14G043900* 和 *Glyma.15G217400*

在大粒和小粒品种组间存在同义及非同义突变, 可能通过调控 hub 基因以及 DEGs 调控大豆百粒重及粒形相关性状, 且这 2 个基因受到了驯化和人工改良选择, 因而可能是大豆百粒重的重要候选基因。本研究结果可为后续解析大豆百粒重发育及调控机制提供重要的试验数据和理论基础。

References

- [1] Lu X, Xiong Q, Cheng T, Li Q T, Liu X L, Bi Y D, Li W, Zhang W K, Ma B, Lai Y C, Du W G, Man W Q, Chen S Y, Zhang J S. A *PP2C-I* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol Plant*, 2017, 10: 670–684.
- [2] Li J, Zhao J, Li Y, Gao Y, Hua S, Nadeem M, Sun G, Zhang W, Hou J, Wang X, Qiu L. Identification of a novel seed size associated locus *SW9-1* in soybean. *Crop J*, 2019, 7: 548–559.
- [3] 蒋费涛, 王书平, 祁俊生, 熊春霞. 转录组学技术及其在植物系统学上的研究进展. 现代盐化工, 2020, 47(4): 14–17.
Jiang F T, Wang S P, Qi J S, Xiong C X. Research progress of transcriptional technology and its advances in plant phylogeny. *Mod Salt Chem Ind*, 2020, 47(4): 14–17 (in Chinese with English abstract).
- [4] Rory S, Marta G, James H. RNA sequencing: the teenage years. *Nat Rev Genet*, 2019, 20: 631–656.
- [5] 杨宇昕, 桑志勤, 许诚, 代文双, 邹彬. 利用 WGCNA 进行玉米花期基因共表达模块鉴定. 作物学报, 2019, 45: 161–174.
Yang Y X, Sang Z Q, Xu C, Dai W S, Zou C. Identification of co-expression modules for maize flowering stage genes using WGCNA. *Acta Agron Sin*, 2019, 45: 161–174 (in Chinese with English abstract).
- [6] Hollender C A, Kang C, Darwish O, Geretz A, Matthews B F, Slovin J, Alkharouf N, Liu Z. Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiol*, 2014, 165: 1062–1075.
- [7] Greenham K, Guadagno C R, Gehan M A, Mockler T C, Weinig C, Ewers B E, McClung C R. Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in *Brassica rapa*. *eLife*, 2017, 6: e29655.
- [8] Yang S N, Miao L, He J B, Zhang K, Li Y, Gai J Y. Dynamic transcriptome changes related to oil accumulation in developing soybean seeds. *Int J Mol Sci*, 2019, 20: 2202.
- [9] Lu X, Li Q T, Xiong Q, Li W, Bi Y D, Lai Y C, Liu X L, Man W Q, Zhang W K, Ma B, Chen S Y, Zhang J S. The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J*, 2016, 86: 530–544.
- [10] Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, Monnin D, Filali A E, Carareto C M, Vieira C, Picard F, Kremer N, Vavre F, Sagot M F, Lacroix V. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res*, 2016, 44: e148.
- [11] Rogier O, Chateigner A, Amanzougarene S, Lesage-Descauses M C, Balzergue S, Brunaud V, Caius J, Soubigou-Taconnat L, Jorge V, Segura V. Accuracy of RNAseq based SNP discovery and genotyping in *Populus nigra*. *BMC Genomics*, 2018, 19: 909.
- [12] Liu Y C, Du H L, Li P C, Shen Y T, Peng H, Liu S L, Zhou G A, Zhang H K, Liu Z, Shi M, Huang X H, Li Y, Zhang M, Wang Z, Zhu B G, Han B, Liang C Z, Tian Z X. Pan-genome of wild and cultivated soybeans. *Cell*, 2020, 182: 162–176.
- [13] 丁琦, 徐伟, 李蒙, 王秀成, 卢伟, 盖钧镒, 王玲, 邢光南. 基于分水岭和统计矩的大豆籽粒形态参数测量方法. 大豆科学, 2019, 38: 960–967.
Ding Q, Xu W, Li M, Wang X C, Lu W, Gai J Y, Wang L, Xing G N. Measurement method of soybean seed morphological parameters based on watershed and statistical moment. *Soybean Sci*, 2019, 38: 960–967 (in Chinese with English abstract).
- [14] Pertea M, Kim D, Pertea G M, Leek J T, Salzberg S L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*, 2016, 11: 1650–1667.
- [15] Michael I L, Wolfgang H, Simon A. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 2014, 15: 550.
- [16] Peter L, Steve H. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 2008, 9: 1–32.
- [17] Yu G C, Wang L G, Han Y Y, He Q Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, 2012, 16: 284–287.
- [18] Paul S, Andrew M, Owen O, Nitin S B, Jonathan T W, Daniel R, Nada A, Benno S, Trey I. Cytoscape: a software environment for integrated. models of biomolecular interaction networks. *Genome Res*, 2003, 13: 2498–2504.
- [19] Engström P G, Steijger T, Sipos B, Grant G R, Kahles A, Rätsch G, Goldman N, Hubbard T J, Harrow J, Guigó R, Bertone P. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods*, 2013, 10: 1185–1191.
- [20] Alexander D, Carrie A D, Felix S, Jorg D, Chris Z, Sonali J, Philippe B, Mark C, Thomas R G. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 2013, 29: 15–21.
- [21] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 2011, 27: 2987–2993.
- [22] Heldenbrand J R, Baheti S, Bockol M A, Drucker T M, Hart S N, Hudson M E, Iyer R K, Kalmbach M T, Kendig K I, Klee E W, Mattson N R, Wieben E D, Wierper M, Wildman D E, Mainzer L S. Recommendations for performance optimizations when using GATK3.8 and GATK4. *BMC Bioinf*, 2019, 20: 722.
- [23] Petr D, Adam A, Goncalo A, Cornelis A A, Eric B, Mark A D, Robert E H, Gerton L, Gabor T M, Stephen T S, Gilean M, Richard D. The variant call format and VCFtools. *Bioinformatics*, 2011, 27: 2156–2158.
- [24] Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X Y, Ruden D M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 2012, 6: 80–92.
- [25] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones S J, Marra M A. Circos: an information aesthetic for comparative genomics. *Genome Res*, 2009, 19: 1639–1645.
- [26] Torkamaneh D, Laroche J, Valliyodan B, O'Donoghue L, Belzile F. Soybean (*Glycine max*) Haplotype Map (GmHapMap): a

- universal resource for soybean translational and functional genomics. *Plant Biotechnol J*, 2020, 19: 324–334.
- [27] Li N, Xu R, Li Y. Molecular networks of seed size control in plants. *Annu Rev Plant Biol*, 2019, 70: 435–463.
- [28] Cheng Z J, Zhao X Y, Shao X X, Wang F, Zhou C, Liu Y G, Zhang Y, Zhang X S. Absciscic acid regulates early seed development in *Arabidopsis* by ABI5-mediated transcription of *SHORT HYPOCOTYL UNDER BLUE1*. *Plant Cell*, 2014, 26: 1053–1068.
- [29] Koichiro A, Tokunori H, Kanna S I, Miyako U T, Hidemi K, Makoto M. A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol*, 2014, 55: 897–912.
- [30] Gu Y Z, Li W, Jiang H W, Wang Y, Gao H H, Liu M, Chen Q S, Lai Y C, He C Y. Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *J Exp Bot*, 2017, 68: 2717–2729.
- [31] Li N, Li Y H. Signaling pathways of seed size control in plants. *Curr Opin Plant Biol*, 2016, 33: 23–32.
- [32] Hans W, Ljudmilla B, Ulrich W. Molecular physiology of legume seed development. *Annu Rev Plant Biol*, 2005, 56: 253–279.
- [33] 倪资园, 蔡丽, 谢皓, 陈学珍. 大豆籽粒发育过程中营养物质的积累动态分析. 北京农学院学报, 2011, 26(2): 1–3.
- Ni Z Y, Cai L, Xie H, Chen X Z. Analysis of dynamic accumulation on nutrients during maturing process in soybean seed. *J Beijing Agric Coll*, 2011, 26(2): 1–3 (in Chinese with English abstract).
- [34] Xu R, Duan P G, Yu H Y, Zhou Z K, Zhang B L, Wang R C, Li J, Zhang G Z, Zhuang S S, Lyu J, Li N, Chai T Y, Tian Z X, Yao S G, Li Y H. Control of grain size and weight by the OsMKKK10-OsMKK4-OsMAPK6 signaling pathway in rice. *Mol Plant*, 2018, 11: 860–873.
- [35] Srivastava A K, Lu Y M, Zinta G, Lang Z B, Zhu J K. UTR-dependent control of gene expression in plants. *Trends Plant Sci*, 2018, 23: 248–259.
- [36] Mian M A, Bailey M A, Tamulonis J P, Shipe E R, Carter T E, Parrott W A, Ashley D A, Hussey R S, Boerma H R. Molecular markers associated with seed weight in two soybean populations. *Theor Appl Genet*, 1996, 93: 1011–1016.
- [37] 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 Sci*, 1999, 39: 1642–1651.
- [38] Li W X, Zheng D H. QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). *J Crop Sci Biotechnol*, 2008, 11: 171–176.
- [39] Teng W, Han Y, Du Y, Sun D, Zhang Z, Qiu L, Sun G, Li W. QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). *Heredity*, 2009, 102: 372–380.
- [40] Akond M, Liu S M, Boney M, Kantartzi St K, Meksem K, Bel-laloui N, Lightfoot D A, Kassem M A. Identification of quantitative trait loci (QTL) underlying protein, oil, and five major fatty acids' contents in soybean. *Am J Plant Sci*, 2014, 5: 158–167.