

张玉梅, 萧涵, 蓝新隆, 等. 基于重测序的南方高蛋白大豆品种福豆 234 的遗传变异 [J]. 福建农业学报, 2024, 39 (6): 652–661.

ZHANG Y M, XIAO H, LAN X L, et al. Genetic Variations in Southern High-protein Soybean Fudou 234 by Re-sequencing [J]. *Fujian Journal of Agricultural Sciences*, 2024, 39 (6): 652–661.

基于重测序的南方高蛋白大豆品种 福豆 234 的遗传变异

张玉梅, 萧涵, 蓝新隆, 夏春英, 林国强*

(福建省农业科学院作物研究所/福建省特色旱作物品种选育工程技术研究中心/

福建省蔬菜遗传育种重点实验室, 福建 福州 350013)

摘要:【目的】揭示南方高蛋白大豆遗传变异提供理论参考。【方法】通过高通量测序技术对南方高蛋白大豆品种福豆 234 进行全基因组重测序。【结果】共获得 64 757 037 条 Clean reads, 测序深度为 17×, 基因组覆盖度分别达 98.08% (1×) 和 96.25% (5×)。共鉴定出 1 478 393 个单核苷酸多态性 (Single nucleotide polymorphism, SNP) 位点和 356 739 个小片段插入缺失 (Small insertion-deletion, Small InDel) 位点。其中共鉴定出 14 323 个非同义 SNP 突变的基因, 4 186 个 Small InDel 的突变基因位于编码区 (Coding sequence, CDS)。通过 COG (Clusters of orthologous groups of proteins) 分析发现信号传导机制、转录、碳水化合物转输和代谢等和 KEGG (Kyoto encyclopedia of genes and genomes) 分析发现碳代谢、淀粉和蔗糖代谢、氨基酸生物合成、植物激素信号传导、内质网蛋白质加工等通路与福豆 234 遗传变异相关。此外, 本研究以大豆籽粒蛋白质含量数量性状基因座 (Quantitative trait locus, QTL) 的两个主要区段内的候选基因进行分析, 发现有 65 个基因发生 SNP 或 Small InDel 水平的变异, 其中, SNP 变异类型达 10 种, Small InDel 变异类型达 7 种。【结论】本研究初步揭示南方高蛋白大豆的遗传变异规律, 为高蛋白大豆品种选育和分子标记的开发提供数据支持和理论依据。

关键词: 福豆 234; 全基因组重测序; 单核苷酸多态性; 小片段插入缺失; 变异

中图分类号: S565.1

文献标志码: A

文章编号: 1008-0384 (2024) 06-0652-10

Genetic Variations in Southern High-protein Soybean Fudou 234 by Re-sequencing

ZHANG Yumei, XIAO Han, LAN Xinlong, XIA Chunying, LIN Guoqiang*

(Crop Research Institute, Fujian Academy of Agricultural Sciences/Fujian Engineering Research Center for Characteristic Dry Crop Varieties Breeding/Fujian Key Laboratory of Vegetable Genetics and Breeding Fuzhou, Fuzhou, Fujian 350013, China)

Abstract: 【Objective】Genetic variations in the southern high-protein soybeans were revealed by re-sequencing the whole genome. 【Method】High-throughput sequencing was conducted on the whole genome of the southern high-protein soybean Fudou 234 to detect genetic variations. 【Result】In the 64 757 037 clean reads, the sequencing depth was 17× with genome coverage of 98.08% (1×) and 96.25% (5×). There were 1 478 393 single nucleotide polymorphisms (SNPs) and 356 739 small insertion-deletions (Small InDels) identified. Among them, 14 323 non-synonymous SNP mutations and 4 186 Small InDel mutated genes were found in the coding sequence (CDS). Analysis by COG (Clusters of orthologous groups of proteins) revealed that signal transduction mechanisms, transcription, carbohydrate translocation and metabolism and KEGG (Kyoto encyclopedia of genes and genomes) analyses revealed that the pathways of carbon metabolism, starch and sucrose metabolism, amino acid biosynthesis, phytohormone signalling, and endoplasmic reticulum protein processing were associated with genetic variation in Fudou 234. In addition, by studying the candidate genes in two major segments of soybean kernel protein

收稿日期: 2024-05-17 修回日期: 2024-06-11

作者简介: 张玉梅 (1980—), 女, 博士, 副研究员, 主要从事大豆育种与栽培研究, E-mail: 2407270791@qq.com

* 通信作者: 林国强 (1965—), 男, 研究员, 主要从事大豆育种与栽培研究, E-mail: 542057656@qq.com

基金项目: 福建省科技厅区域发展项目 (2023N3001); 福建省科技计划公益类专项 (2022R1031006); 福建省财政专项 (CXTD2021011-2)

quantitative trait locus (QTL), 10 SNP and 7 Small InDel type variations were discovered in 65 genes. 【Conclusion】 The genetic variations in the southern high-protein soybeans deviated from the regular varieties were unveiled to provide new venue for breeding and developing molecular markers in studying soybeans.

Key words: Fudou 234; whole genome re-sequencing; single nucleotide polymorphism; small InDel; variation

0 引言

【研究意义】大豆 (*Glycine max* L. Merr.) 是全球重要的粮食作物之一, 为食用和饲用提供植物蛋白和油脂来源^[1]。随着消费者对大豆需求的增加, 育种工作者有望挖掘出大豆优异性状的品种, 如高油、高蛋白、抗逆性强、早熟品种等。作为我国传统的经济作物, 大豆种质资源保护和利用势在必行。【前人研究进展】近年来, 全基因组重测序在分子育种和基因组研究遗传多样性中发展迅速, 通过与基因组参考物比对, 重测序个体可以识别 DNA 序列的遗传变异信息, 能够有效提升对作物基因组的认识和选择育种提供新见解, 因此可以获得更全面的基因组信息, 应用于作物改良^[2-3]。为发掘大豆优异种质特性、筛选和鉴定影响大豆产量、品质和抗逆性等性状的代谢通路和候选基因提供良好的平台^[4]。目前, 基于全基因组重测序的大豆优异性状研究成为热点。Yang 等^[5]对 250 份大豆进行全基因组重测序, 发现 *GSTT1*、*GL3*、*GSTL3*、*CKX3*、*CYP85A2* 等基因参与调控大豆产量和品质性状。Liu 等^[6]对鲜食大豆进行全基因组重测序技术, 发现高质量的 SNP 大多位于基因间区和内含子, 位于编码区较少; 同时鉴定出 355.55 K InDel, 其中 4856 个 InDel 位于外显子区域。蔗糖合酶基因 (*SoyZH13_18G042801*、*SoyZH13_18G042802*、*SoyZH13_18G042803* 和 *SoyZH13_20G077905*)、蔗糖磷酸合酶基因 (*SPS3*) 和糖转运基因 (*SoyZH13_04G095700*、*SoyZH13_09G175600*、*SWEET6b* 和 *SWEET16*) 作为鲜食大豆甜度和籽粒大小等重要性状相关的候选基因。Lee 等^[7]基于大豆早花突变体全基因组重测序鉴定了 332 821 个 SNP 和 65 178 个 InDel, 其中 *GI*、*AGL18*、*TOC1* 和 *ELF3* 等基因影响 *GmFT2a* 的表达, 导致花期提前。Maldonado dos Santos 等^[8]对 28 份巴西大豆通过全基因组重测序鉴定出 541 762 个 SNP、98 922 个 InDel 和 1 093 个 (Copy number variation, CNV), 此外还发现共有 327 个基因的 668 个等位变异与适应热带气候有关。Jiang 等^[9]对 427 个大豆重组自交系进行全基因组重测序, 在 2 号染色体 RSC11K 区域鉴定出 429 个 SNP 和 142 个 InDel 特异位点, 同时发现 34 个与大豆抗花叶病毒的相关候选基因。Yuan 等^[10]对大豆近等基因系进行全基因组重测序, 鉴定

出 55 853 个 SNPs 和 3 188 个 InDel, 多态性位点主要富集在 13 号染色体, 还发现候选基因 *Glyma.13g25950*、*Glyma.13g25970* 和 *Glyma.13g26380* 与大豆抗花叶病毒有关。【本研究切入点】目前, 基于全基因组重测序分析多数聚焦于大豆含油量、甜度、产量和抗逆性等性状, 而在大豆高蛋白性状上却鲜有研究。因此, 聚焦于全基因组水平上剖析大豆高蛋白性状的遗传多样性及遗传变异具有重要意义。【拟解决的关键问题】本研究旨在对南方高蛋白大豆品种福豆 234 进行全基因组重测序, 对其 SNP 和 Small InDel 进行深度挖掘和分析, 通过功能富集分析对变异基因进行注释。以期为高蛋白大豆遗传多样性、种质资源鉴定、遗传改良及亲本选配提供理论依据。

1 材料与方 法

1.1 植物材料

大豆品种福豆 234 (国审豆 2007024) 由福建省农业科学院作物研究所自主选育, 为高蛋白品种, 蛋白质含量和脂肪含量分别达 46.33% 和 17.68%, 适应性广, 丰产性和稳产性好^[11]。将种子种于蛭石中, 待长出三出复叶时, 采集三出复叶中间叶片, 液氮冻存后于 -80 °C 保存备用。

1.2 大豆叶片 DNA 提取、文库构建和质量控制

采用改良十六烷基三甲基溴化铵 (Hexadecyl trimethyl ammonium bromide, CTAB) 法提取大豆叶片基因组 DNA^[12]。利用 Nanodrop 2000 检测 DNA 样品的纯度、浓度和完整性, 并用琼脂糖凝胶电泳检测 DNA 质量。随后将提取的高质量 DNA 送样北京百迈克生物科技有限公司进行全基因组重测序, 在片段大小选择和扩增之前, 通过添加腺嘌呤核苷酸 (A) 和测序适配器来修复 DNA, 以清除碎片 DNA 的 3' 端的间隙, 然后通过 1% 琼脂糖凝胶中电泳进一步选择修复的 DNA 片段^[13]。通过 PCR 扩增产物构建测序文库, 利用 Illumina HiSeqTM 2500 测序平台对合格质量控制的文库进行测序^[14]。通过去除测序适配器并过滤低质量 reads, 对 Raw reads 进行处理以获得 Clean reads。通过 BWA (Burrow-Wheeler Aligner) 软件^[15]将 Clean reads 映射到参考基因组序列 Williams 82 (*Glyma.Wm82.a2v1*, <http://soybase.org>)。利用 SAMtools 软件^[16]将对齐后的 sam 文件

转换成 bam 文件, Picard (<http://sourceforge.net/projects/picard>.) 用于删除标记 PCR 重复项, 以此对结果进行分析。

1.3 SNP 和 Small InDel 的注释

基于样本与参考基因组的比对结果, 利用 GATK 软件^[17]的 HaplotypeCaller 算法检测 SNP 和 Small InDel。为降低 SNP 和 Small InDel 检测的错误率, 利用其中的 Variant Filtration 算法进行过滤, 过滤参数设置为: QUAL < 30.0、QD < 2.0、FS> 60.0、MQ<40 及cluster-Size-2-cluster-Window-Size5^[18]。最后使用 SnpEff 软件^[19]对 SNP 和 Small InDel 进行注释。

1.4 变异基因的筛选与功能注释分析

基于大豆参考基因组的基础, 统计非同义突变的 SNP 位点和编码区 (CDS) 发生 InDel 变异的基因数量。基因功能通过序列比对, 基于蛋白质直系同源簇 (Clusters of Orthologous Groups of Proteins, COG) 和京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 数据库进行注释。

2 结果与分析

2.1 测序数据质控评估

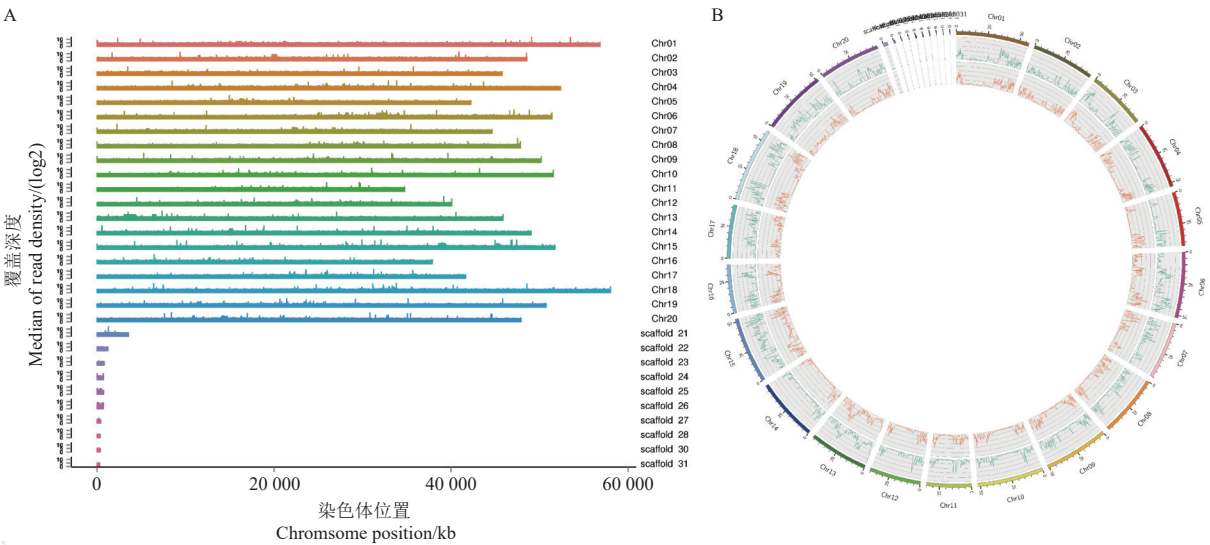
福豆 234 全基因组重测序数据通过合格质量控制, 共获得 64 799 941 条 Raw reads。Clean reads 为剔除 Raw reads 中包含适配器的读数、ployp-N 的读数和低质量读数, 共得到 64 757 037 条 Clean reads。经过严格的数据质控和过滤后, Q30 质控评分均值为 91.82%, GC 含量为 35.90%。以大豆参考基因组

Williams 82 (Glyma.Wm82.a2v1, <http://soybase.org>) 信息为参考, 福豆 234 所得序列在大豆基因组中的定位比达 98.66%, 双端定位比达 91.34%。测序覆盖深度达到 17×, 在 1×的覆盖深度下, 基因组上的覆盖度为 98.08%; 在 5×的覆盖深度下, 测序深度序列在基因组上的覆盖度达 96.25%。福豆 234 染色体上的读数覆盖分布如图 1A 所示, 读数覆盖更均匀地分布在染色体上。以上结果表明福豆 234 数据质量较高, 序列能够在染色体上均匀分布, 测序结果可信度高, 可进行后续分析。变异检测结果表明, 福豆 234 中共获得 1 478 393 个 SNP 和 356 739 个 Small InDel (图 1B)。

2.2 SNP 检测与注释

重测序读数映射到大豆参考基因组 Williams 82 (Glyma.Wm82.a2v1,<http://soybase.org>), 采用 GATK 软件进行对比, 删除冗余 SNP, 福豆 234 共计 1 478 393 个高质量 SNP, 其中, 转换类型的 SNP (961 071 个) 多于颠换类型的 SNP (517 322 个), 占比分别为 65.01% 和 34.99%。转换/颠换比率 (Ti/Tv) 的平均值为 1.85; 纯合性的 SNP 较杂合性的 SNP (233 652 个) 多, 为 1 244 741 个, 占比为 84.20%。总体杂合率为 15.80%。

福豆 234 基因组的 SNP 注释采用 SnpEff 软件进行。如图 2 所示, 福豆 234 的 SNP 变异主要分布在基因间区, 共鉴定出 673 309 个 (47.42%)。其次为基因上游区域 (5000bp 以内)、基因下游区域 (5000bp 以内) 和内含子, 分别为 312 129



A: 福豆 234 染色体覆盖深度分布; B: 福豆 234 各类型变异在染色体的分布。

A: distribution of chromosomes coverage depth of Fudou 234; B: the distribution of the variations of Fudou 234 on the chromosomes.

图 1 福豆 234 测序数据分析

Fig. 1 Analysis of sequencing data of Fudou 234

(21.98%)、248175 (17.48%) 和54082 个(10.88%)。此外，编码区域内的 58618 个 SNP 中，获得了 25318 个同义编码突变，占 43.19%，以及 32313 个非同义编码突变，占 55.12%。

2.3 Small InDel 检测与注释

福豆 234 重测序读数被映射到 Williams 82 参考基因组，去除重复的 Small InDel，共鉴定出 356739 个 Small InDel。其中，4965 个 Small InDel 变异从 CDS 区域鉴定得到。其中，插入、缺失、纯合突变和杂合突变从 CDS 区域鉴定得到，分别为 2486、2479、3605、1360 个。福豆 234 Small Indel 长度在编码区域和全基因组的分布比较如图 3 所示。+1、-1、+3、-3 类型变异在编码区域占较大比例。在全基因组范围内存在较多的+1、-1 类型变异，而+3、-3 类型变异较编码区域相对占比较小。

福豆 234 的 Small InDel 注释统计如图 4 所示。共鉴定出 111649 个 Small InDel 在全基因组中主要集中于基因间区 (31.73%)。其次位于基因上游区域 (5000 bp 以内) (29.23%)、基因下游区域 (5000 bp 以内) (21.52%) 和内含子 (14.05%)，分别为102833、75716、49430 个。共鉴定出 2880 个位于编码区域的 Small InDel 主要集中于移码突变 (45.67%)。其次为密码子的删除和插入、密码子插入和密码子删除。分别

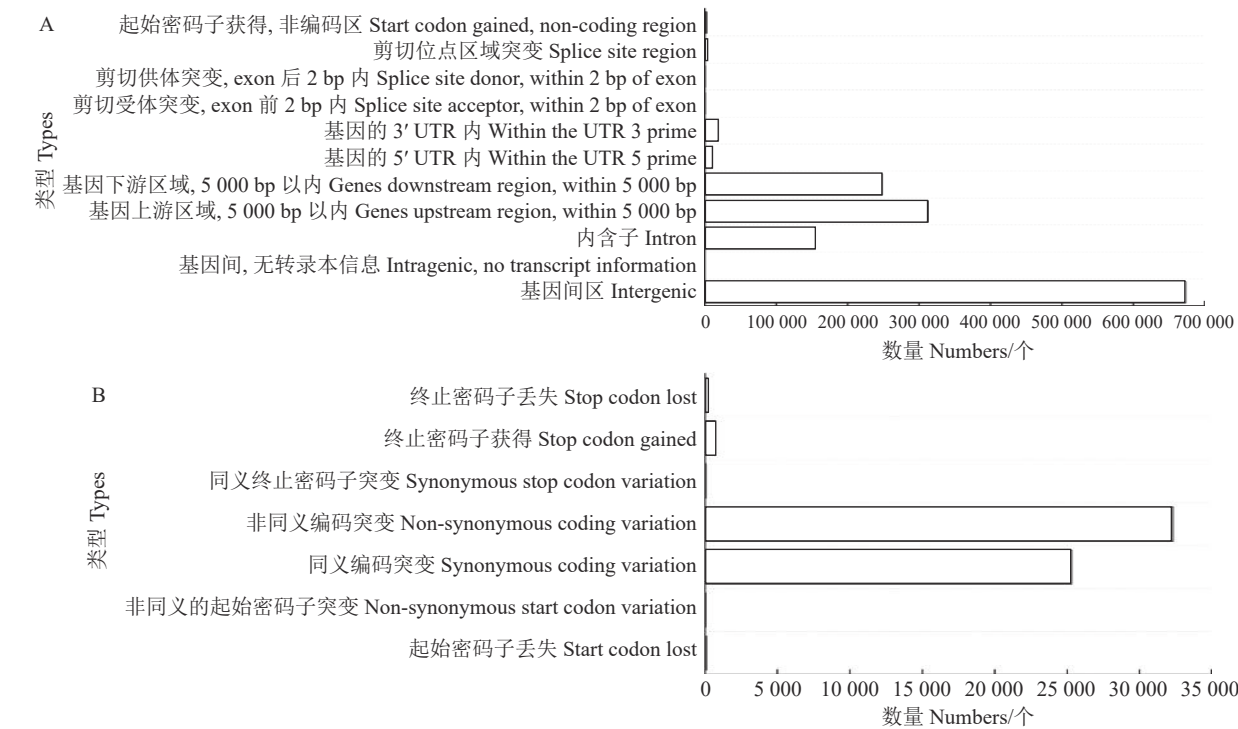
为1341 (21.27%)、724 (11.48%) 和617 个(9.78%)。

2.4 遗传变异分析

基于大豆参考基因组 Williams 82 的基础上，福豆 234 基因组一共鉴定出 18509 个突变基因。其中，共鉴定出 14323 个非同义 SNP 突变的基因，占 77.38%；共鉴定出 4186 个 Small Indel 的突变基因位于 CDS 区域，占 22.62%。

将福豆 234 基于 DNA 水平变异进行 COG 和 KEGG 数据库进行注释。基于 COG 注释结果 (图 5)，信号传导机制 (684 个)，转录 (559 个)，碳水化合物转输和代谢 (555 个) 等在福豆 234 存在大量的变异基因。根据福豆 234 基因组重测序进行 KEGG 注释如图 6 所示。变异基因主要注释到 KEGG 代谢通路涉及碳代谢 (ko 01200) 156 个、淀粉和蔗糖代谢 (ko 00500) 147 个，氨基酸生物合成 (ko 01230) 142 个，植物激素信号传导 (ko 04075) 137 个，内质网蛋白质加工 (ko 04141) 110 个。

前人研究表明 20 号染色体 (I 连锁群，29.8-31.6 Mbp) 和 15 号染色体 (E 连锁群，38.1-39.7 Mbp) 在自然环境和复杂的遗传种群中定位到大豆籽粒蛋白质含量 QTL^[20-21]。Paril 等^[22] 对两个主要的大豆蛋白质含量 QTL (Chr.20 和 15) 以及区域之间的片段复制区间预测出 86 个候选基因。本研究对福豆 234 基于全基因组测序基础上，根据之前报道的候选基因



A: 福豆 234 的 SNP 变异分布全基因组注释结果; B: 福豆 234 的 SNP 变异分布 CDS 注释结果。
A: genome-whole annotation statistics of SNP variant distribution of Fudou 234; B: CDS annotation statistics of SNP variant distribution of Fudou 234.

图 2 福豆 234 的 SNP 注释结果
Fig. 2 SNP annotations of Fudou 234

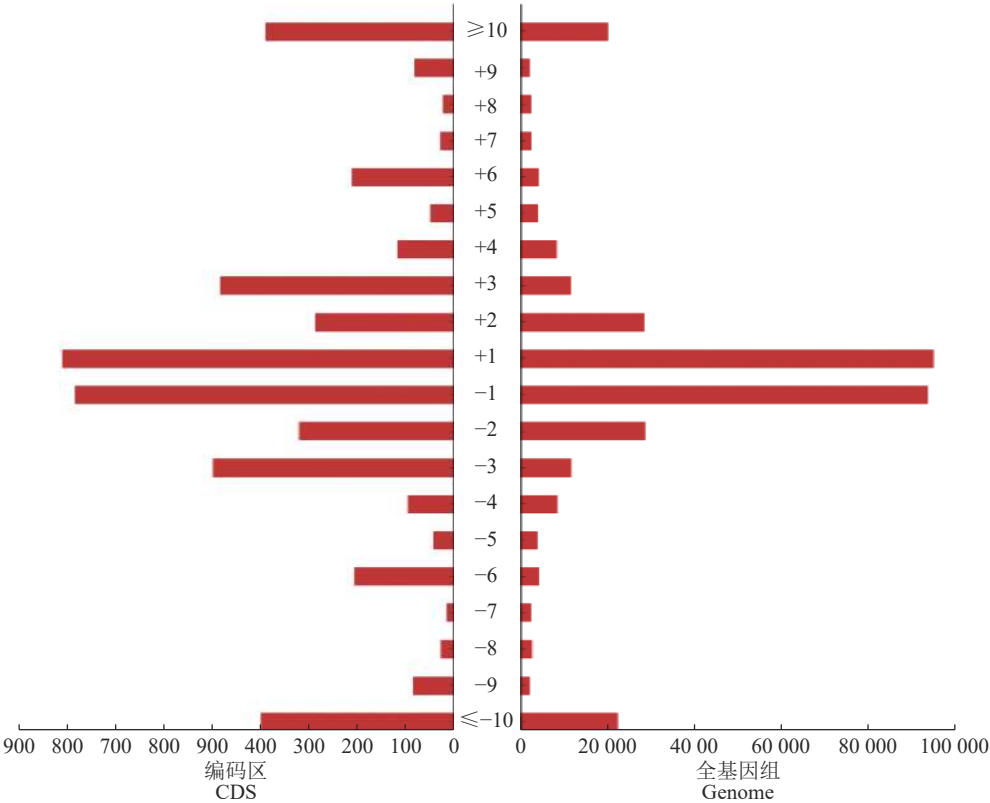
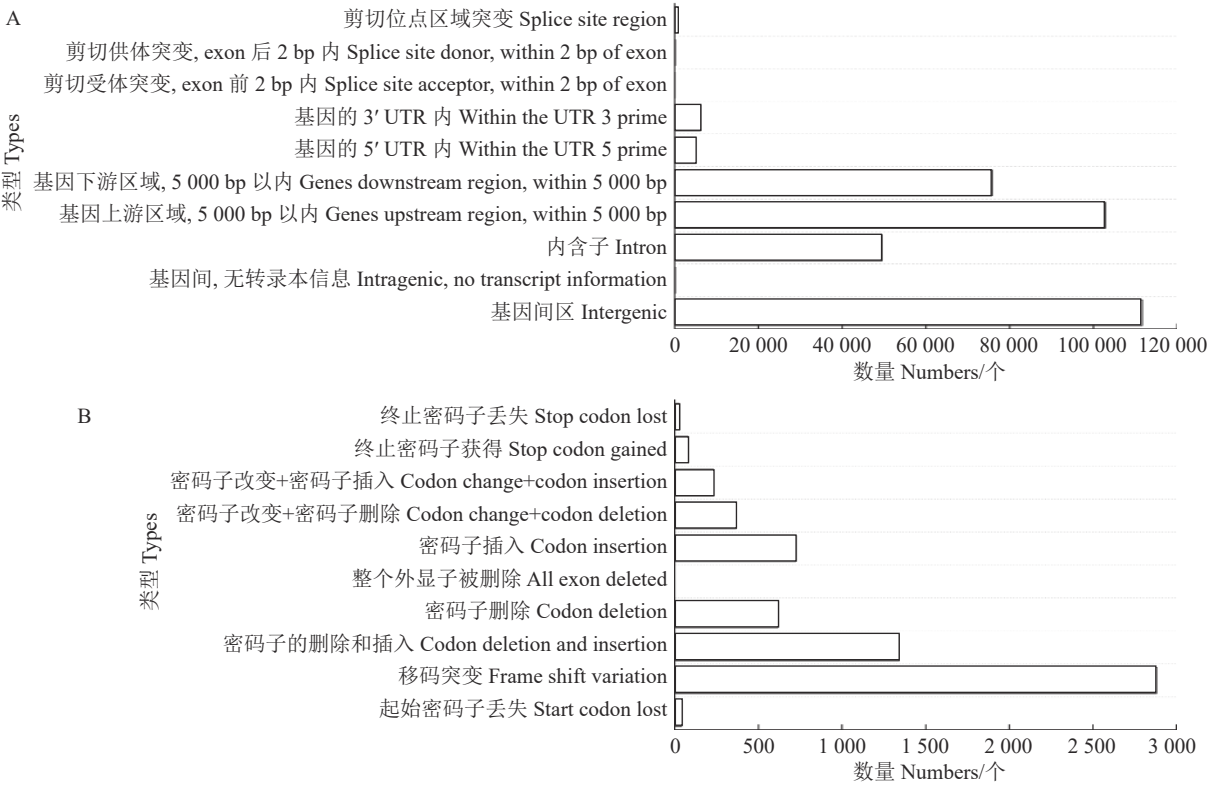


图 3 福豆 234 的 CDS 区和全基因组 Small Indel 长度分布

Fig. 3 CDS region and distribution of Small Indel lengths in whole genome of Fudou 234



A: 福豆 234 的 Small Indel 变异分布全基因组注释结果; B: 福豆 234 的 Small Indel 变异分布 CDS 注释结果。

A: genome-whole annotation statistics of Small Indel variant distribution of Fudou 234; B: CDS annotation statistics of Small Indel variant distribution of Fudou 234.

图 4 福豆 234 的 Small Indel 注释结果

Fig. 4 Small Indel annotations of Fudou 234

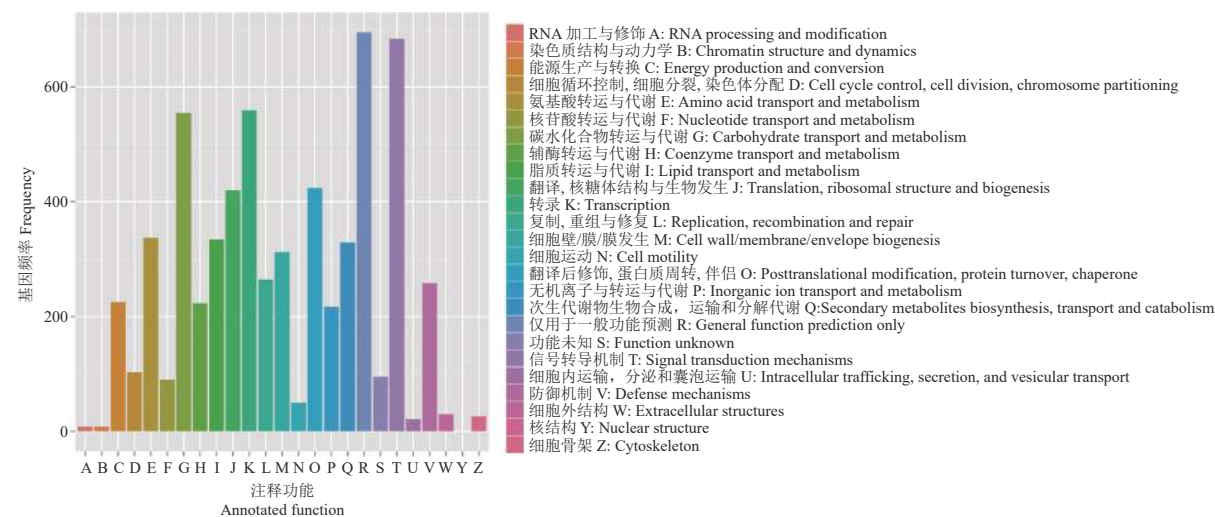


图 5 福豆 234 的遗传变异基因 COG 富集

Fig. 5 COG enrichment of genetic variants in Fudou 234

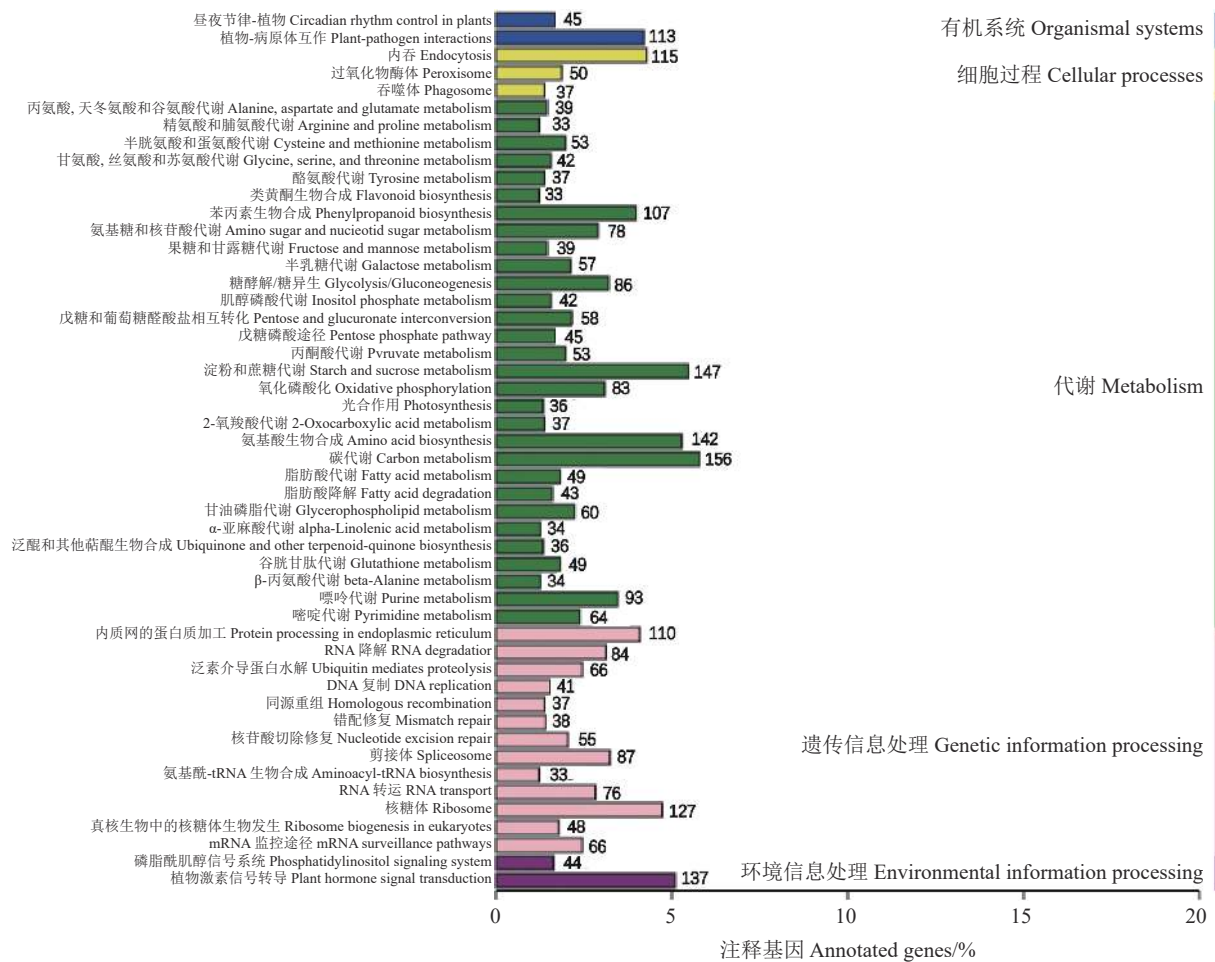


图 6 福豆 234 的遗传变异基因 KEGG 富集

Fig. 6 KEGG enrichment of genetic variants in Fudou 234

进行分析。由表 1 所示，共鉴定出 65 个候选基因存在 SNP 或 Small InDel 变异。其中 SNP 变异类型达 10 种，占比为 15.38%，主要集中于基因上游区域和基因下游区域。其次是内含子区域和编码区内同义突变，其他变异类型较少，无密码子插入和移码突

变，CDS 区域内非同义突变有 16 个位点，涉及 10 个基因；Small InDel 变异类型相对 SNP 变异类型而言相对较少，仅 7 种，占比为 10.77%。主要集中在基因上游区域和基因下游区域，其次是内含子区域，其他变异类型较少。

表 1 福豆 234 中与大豆蛋白质含量主要 QTL 区段候选基因 SNP 和 Small InDel 变异

Table 1 SNP and Small InDel variations of candidate genes in main QTL segments related to soybean protein content in Fudou 234

基因名称 Gene ID	变异类型 Variation type		基因名称 Gene ID	变异类型 Variation type		基因名称 Gene ID	变异类型 Variation type	
	SNP	Small InDel		SNP	Small InDel		SNP	Small InDel
Glyma.20G082300	2	1	Glyma.20G086900	/	2	Glyma.10G133900	1	1
Glyma.20G082700	/	1	Glyma.20G087000	6	/	Glyma.10G134100	1,7,2	7
Glyma.20G082800	7	/	Glyma.20G087200	1	/	Glyma.10G134400	1	1,7
Glyma.20G082900	2	/	Glyma.15G048600	2	2	Glyma.10G134500	1,7	1
Glyma.20G083100	1,7,5,8,6,2	1	Glyma.15G048700	1, 7	2	Glyma.10G136100	1	/
Glyma.20G083200	1,2	1,2	Glyma.15G048800	1, 7, 5,3, 2	1,11,7,3,2	Glyma.10G136300	2,1	2
Glyma.20G083300	1,7, 5,3,2	1,2	Glyma.15G048900	1, 4, 6,2	1,7,2	Glyma.10G136400	2	2
Glyma.20G083500	1,7,4,2	1,7,2	Glyma.15G049000	1,10,6,7,3,2	1,4,11,3,7	Glyma.10G136600	7	/
Glyma.20G083600	1,10,5,6,3,4,2	4,2	Glyma.15G049500	3,5	/	Glyma.10G136800	1,7,2	1,2,3
Glyma.20G083800	4,1	1	Glyma.15G049600	3,7,6,1	1,2,7	Glyma.08G182200	/	1
Glyma.20G084000	1,4,8,7,2	1,2	Glyma.15G049700	2,3,6,10,4,1	2,1	Glyma.08G182300	/	1,7
Glyma.20G084100	1,10,5,3,4,2	1,2,4	Glyma.15G049800	2,5,8,7,1	2,7,1	Glyma.08G182400	/	2
Glyma.20G084200	1,5	1	Glyma.15G049900	2,6,7,9,1	1,11,7,3,4	Glyma.08G182500	/	7
Glyma.20G084500	1,5,2	/	Glyma.15G050100	2,7	2	Glyma.08G182700	/	1,2,7,9
Glyma.20G084900	1	/	Glyma.15G050200	5,1	2,1	Glyma.08G182900	/	1,7
Glyma.20G085000	1	2	Glyma.15G050300	3	1	Glyma.08G183000	/	1,7,3,2
Glyma.20G085100	/	2	Glyma.15G050500	1,2,7	1,7	Glyma.08G183400	/	1
Glyma.20G085300	/	1	Glyma.15G050600	1,2,7	1	Glyma.08G183500	/	7,3,2
Glyma.20G085700	2	2	Glyma.10G132200	1,7,5,6,2	1,2	Glyma.08G183600	/	2
Glyma.20G085900	1	2	Glyma.10G132700	2,6	2,7	Glyma.08G183900	/	7,1
Glyma.20G086100	/	1	Glyma.10G132800	3,1	/	Glyma.08G184100	/	2,7,1
Glyma.20G086800	2	2	Glyma.10G133700	2	/			

11种变异类型中，1~11分别代表变异位点发生在基因上游区域、基因下游区域、基因的3'UTR内、基因的5'UTR内、编码区内同义突变、编码区内非同义突变、内含子、剪切位点区域、剪切受体突变、非编码区内起始密码子获得、密码子插入以及移码突变。

1~11 indicate mutation sites in 11 variation types occurred upstream, downstream, UTR 3 prime, UTR 5 prime, synonymous coding, non-synonymous coding, intron, splice site region, splice site donor, start gained, codon insertion, and frame shift, respectively.

3 讨论

大豆是食用和饲用植物蛋白的主要来源，粒用大豆含有 38%~42% 的蛋白质^[22]。随着全球消费者对大豆蛋白质需求日益增加，这一比例还在逐年上升^[23]。因此，通过遗传改良选育高蛋白大豆品种是育种工作者的目标之一。大豆作为豆科植物的一个典型，自花授粉特性对保持基因组同质性和减少基因组变异产生强烈影响。由于大豆基因组的复杂

性，挖掘高蛋白含量的优异特性也具有一定的局限性。自大豆种质资源全基因组测序的完成，大豆基因组遗传图谱揭示加快了新一代大豆分子育种进程^[24-25]。本研究通过全基因组重测序技术，比较于 Williams 82 参考基因组，在高蛋白大豆福豆 234 共获得 1478393 个 SNP 和 356739 个 Small InDel，变异产生的突变基因有 18509 个。变异基因 COG 注释主要发生在信号传导机制类、转录类、碳水化合物运输和代谢类等途径，而在 KEGG 注释主要发生在碳

代谢、淀粉和蔗糖代谢、氨基酸生物合成、植物激素信号传导以及内质网蛋白加工等途径。大豆籽粒中含有 40% 蛋白质，为脂肪（20%）的两倍，虽然蛋白质单位生物能低于脂肪，但在物质分配和能量转运占据优势^[26]。籽粒蛋白质积累与蔗糖合成酶活性呈正比，其活性影响蔗糖卸载量进而影响碳代谢^[27]。

本研究根据已报道的与大豆籽粒蛋白质含量相关的 QTL 通过共线性分析预测到的 86 个候选基因，对福豆 234 的变异基因进行 SNP 和 Small InDel 分析，与参考基因组 Williams 82 相比，共有 65 个候选基因发生 SNP 或 Small InDel 变异，其中有 10 个候选基因编码区内至少一个位点发生非同义突变。王嘉等^[28]对大豆籽粒蛋白质全基因组重测序及 QTL 定位。共获得 2453344 个 SNP 和 476953 个 Small InDel，变异产生的突变基因共 38896 个，14491 个基因存在 2 种或 2 种以上类型的突变。其中，候选基因 *Gy5* (*Glyma.13G123500*)、*ABI3* (*Glyma.18G176100*) 和 *CAR1* (*Glyma.03G163500*) 与大豆储藏蛋白合成相关。此外，QTL 结果表明，与大豆籽粒蛋白质含量相关的 6 个 QTL 位点分别分布在 C2、E、I 和 M4 连锁群上。此外，Fliege 等^[29]研究发现，位于 20 号染色体的 *Glyma.20G085100* 基因影响大豆籽粒蛋白质含量，在低蛋白大豆导入该基因有助于提高 2% 的蛋白质含量。糖作为信号分子参与植物生长发育，其中糖转运蛋白是植物体运输途径的关键组成部分，包括单糖转运蛋白、蔗糖转运蛋白和糖最终输出转运蛋白（Sugar will eventually be exported transporter, SWEET）^[30-31]。蔗糖在成熟叶片叶肉细胞中合成，SWEET 蛋白为韧皮部转载之前介导蔗糖从叶肉细胞流出到质外体的核心参与者^[32]。另一方面，SWEET 蛋白与其他蛋白质相互协调，调节“源-流-库”关系^[33]。大豆籽粒由成熟胚组成，胚胎发育时期，微小的胚胎快速生长并从发育中种子的胚乳中获取大量糖分^[34]。之前的研究表明 *GmSWEET24* (*Glyma.08G183500*) 在籽粒发育时期表达量较高^[35-36]。*WAT1* (Walls are thin 1) 是一种细胞内生长素转运蛋白，对木质部前体中生长素积累有直接作用^[37]。Pal 等^[38]研究发现参与植物激素信号传导的 *WAT1* 对油菜种子重量和产量具有显著影响。Liu 等^[39]研究表明 3 个编码 *WAT1* 相关蛋白不仅与苧麻纤维产量呈高度相关，同时发现与氮代谢相关正向选择基因共计 22 个，证实与参与粗蛋白含量和营养生长等驯化性状。本研究结果表明，候选基因 *Glyma.15G049700*、*Glyma.15G049800* 和 *Glyma.15G049900* 具有 *WAT1* 的功能，可能是促进福豆 234 高产和高蛋白特征的关键。

4 结论

大豆籽粒蛋白质含量是典型的数量性状，一般是由主基因控制，还受到微效多基因和环境因素的影响，不同大豆蛋白质含量基因间存在独立或互作的关系^[40]。本研究对高蛋白大豆品种福豆 234 全基因组重测序数据分析，获得 1478393 个 SNP，356739 个 Small InDel。通过 COG 分析发现信号传导机制、转录、碳水化合物转输和代谢等和 KEGG 分析发现碳代谢、淀粉和蔗糖代谢、氨基酸生物合成、植物激素信号传导、内质网蛋白质加工等通路与福豆 234 遗传变异相关。由变异产生的突变基因达 18509 个，从先前报道过的 15 号染色体和 20 号染色体区段内的基因进行变异分析，65 个基因发生 SNP 或 Small InDel 水平的变异，发现 SNP 变异类型占主体，而 Small InDel 变异类型相对较少。本研究初步阐明了福豆 234 全基因组突变类型，为大豆高蛋白品种的选育和优异基因挖掘奠定基础，可进一步为基因研究和分子标记辅助育种等提供大量的分子标记资源，为研究该性状的遗传调控提供方向。

参考文献：

- [1] CAO P, ZHAO Y, WU F J, et al. Multi-omics techniques for soybean molecular breeding [J]. *International Journal of Molecular Sciences*, 2022, 23(9): 4994.
- [2] XU X Y, BAI G H. Whole-genome resequencing: Changing the paradigms of SNP detection, molecular mapping and gene discovery [J]. *Molecular Breeding*, 2015, 35(1): 33.
- [3] HUANG X H, FENG Q, QIAN Q, et al. High-throughput genotyping by whole-genome resequencing [J]. *Genome Research*, 2009, 19(6): 1068-1076.
- [4] PETEREIT J, MARSH J I, BAYER P E, et al. Genetic and genomic resources for soybean breeding research [J]. *Plants*, 2022, 11(9): 1181.
- [5] YANG C M, YAN J, JIANG S Q, et al. Resequencing 250 soybean accessions: New insights into genes associated with agronomic traits and genetic networks[J]. *Genomics, Proteomics & Bioinformatics*, 2022, 20(1): 29-41.
- [6] LIU N, NIU Y C, ZHANG G W, et al. Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean [J]. *Horticulture Research*, 2022, 9: uhab052.
- [7] LEE K J, KIM D S, KIM J B, et al. Identification of candidate genes for an early-maturing soybean mutant by genome resequencing analysis [J]. *Molecular Genetics and Genomics*, 2016, 291(4): 1561-1571.
- [8] MALDONADO DOS SANTOS J V, VALLIYODAN B, JOSHI T, et al. Evaluation of genetic variation among Brazilian soybean cultivars

- through genome resequencing [J]. *BMC Genomics*, 2016, 17: 110.
- [9] JIANG H, JIA H Y, HAO X S, et al. Mapping Locus R_{SC11K} and predicting candidate gene resistant to Soybean mosaic virus strain SC11 through linkage analysis combined with genome resequencing of the parents in soybean [J]. *Genomics*, 2022, 114(4): 110387.
- [10] YUAN Y, YANG Y Q, SHEN Y C, et al. Mapping and functional analysis of candidate genes involved in resistance to soybean (*Glycine max*) mosaic virus strain SC3 [J]. *Plant Breeding*, 2020, 139(3): 618–625.
- [11] 林国强, 张轼, 滕振勇, 等. 高蛋白大豆福豆 234 的选育及高产农艺措施数学模型 [J]. *福建农业学报*, 2005, 20 (2): 69–73.
- LIN G Q, ZHANG S, TENG Z Y, et al. Breeding and mathematical model of agronomic measures for high yield and protein content soybean variety Fudou 234 [J]. *Fujian Journal of Agricultural Sciences*, 2005, 20(2): 69–73. (in Chinese)
- [12] ABOUL-MAATY N A F, ORABY H A S. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method [J]. *Bulletin of the National Research Centre*, 2019, 43(1): 25.
- [13] 张彦威, 李伟, 张礼凤, 等. 基于重测序的大豆新品种齐黄 34 的全基因组变异挖掘 [J]. *中国油料作物学报*, 2016, 38 (2): 150–158.
- ZHANG Y W, LI W, ZHANG L F, et al. Genome-wide variations of soybean cultivar Qihuang 34 by whole genome re-sequencing [J]. *Chinese Journal of Oil Crop Sciences*, 2016, 38(2): 150–158. (in Chinese)
- [14] 郭丹丹, 袁凤杰, 郁晓敏. 基于重测序的籽粒型和鲜食型大豆的全基因组变异分析 [J]. *分子植物育种*, 2019, 17 (22): 7306–7312.
- GUO D D, YUAN F J, YU X M. Genome-wide variation analysis of grain and vegetable soybeans based on re-sequencing [J]. *Molecular Plant Breeding*, 2019, 17(22): 7306–7312. (in Chinese)
- [15] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754–1760.
- [16] LI H, HANDSAKER B, WYSOKER A, et al. The Sequence Alignment/Map format and SAMtools [J]. *Bioinformatics*, 2009, 25(16): 2078–2079.
- [17] MCKENNA A, HANNA M, BANKS E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome Research*, 2010, 20(9): 1297–1303.
- [18] 沈丽丽, 曹斌斌, 杨光耀, 等. 毛竹及其 2 种竿型变异类型的全基因组重测序分析 [J]. *基因组学与应用生物学*, 2023, 42 (6): 581–592.
- SHEN L L, CAO B B, YANG G Y, et al. Whole genome resequencing analysis of moso bamboo (*Phyllostachys edulis*) and its two culm variants [J]. *Genomics and Applied Biology*, 2023, 42(6): 581–592. (in Chinese)
- [19] CINGOLANI P, PLATTS A, WANG L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3 [J]. *Fly*, 2012, 6(2): 80–92.
- [20] 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 [J]. *中国油料作物学报*, 2015, 37 (3): 394–410.
- WEI H, WANG J S, LU W G. Molecular genetic advances in soybean seed protein [J]. *Chinese Journal of Oil Crop Sciences*, 2015, 37(3): 394–410. (in Chinese)
- [21] BANDILLO N, JARQUIN D, SONG Q J, et al. A population structure and genome-wide association analysis on the USDA soybean germplasm collection [J]. *The Plant Genome*, 2015, 8(3): eplantgenome2015.04.0024.
- [22] PATIL G, MIAN R, VUONG T, et al. Molecular mapping and genomics of soybean seed protein: A review and perspective for the future [J]. *Theoretical and Applied Genetics*, 2017, 130(10): 1975–1991.
- [23] VAN K, MCHALE L K. Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed [J]. *International Journal of Molecular Sciences*, 2017, 18(6): 1180.
- [24] LAM H M, XU X, LIU X, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection [J]. *Nature Genetics*, 2010, 42: 1053–1059.
- [25] SCHMUTZ J, CANNON S B, SCHLUETER J, et al. Genome sequence of the palaeopolyploid soybean [J]. *Nature*, 2010, 463: 178–183.
- [26] 郭茜茜. 大豆子粒蛋白质积累与碳代谢关系的研究 [D]. 哈尔滨: 东北农业大学, 2010: 16–41.
- GUO Q Q. Study on the relationship between protein accumulation and carbon metabolism in soybean seeds [D]. Harbin: Northeast Agricultural University, 2010: 16–41. (in Chinese)
- [27] PATIL G, VUONG T D, KALE S, et al. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping [J]. *Plant Biotechnology Journal*, 2018, 16(11): 1939–1953.
- [28] 王嘉, 曾召琼, 梁建秋, 等. 基于全基因组重测序的大豆分子标记开发及籽粒蛋白质含量 QTL 定位 [J]. *中国农业科学*, 2019, 52 (16): 2743–2757.
- WANG J, ZENG Z Q, LIANG J Q, et al. Development new molecular markers for quantitative trait locus (QTL) analysis of the seed protein content based on whole genome re-sequencing in soybean [J]. *Scientia Agricultura Sinica*, 2019, 52(16): 2743–2757. (in Chinese)
- [29] FLIEGE C E, WARD R A, VOGEL P, et al. Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20 [J]. *The Plant Journal: for Cell and Molecular Biology*, 2022, 110(1): 114–128.
- [30] MA Q J, SUN M H, LU J, et al. Transcription factor AREB2 is involved in soluble sugar accumulation by activating sugar transporter and amylase genes [J]. *Plant Physiology*, 2017, 174(4): 2348–2362.
- [31] 张计育, 王刚, 王涛, 等. SWEET 蛋白在植物生长发育中的功能作用研究进展 [J]. *植物资源与环境学报*, 2023, 32(5): 1–15.
- ZHANG J Y, WANG G, WANG T, et al. Research progress on functional roles of SWEET proteins in plant growth and development [J]. *Journal of Plant Resources and Environment*, 2023, 32(5): 1–15. (in Chinese) Development [J]. *Journal of Plant Resources*

- and Environment, 2023, 32(5): 1-15. (in Chinese)
- [32] CHEN L Q, QU X Q, HOU B H, et al. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport [J]. *Science*, 2012, 335(6065): 207-211.
- [33] TAKAHASHI F, SATO-NARA K, KOBAYASHI K, et al. Sugar-induced adventitious roots in *Arabidopsis* seedlings [J]. *Journal of Plant Research*, 2003, 116(2): 83-91.
- [34] WANG S D, YOKOSHO K, GUO R Z, et al. The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo [J]. *Plant Physiology*, 2019, 180(4): 2133-2141.
- [35] 柯博洋, 李文龙, 张彩英. 大豆 SWEET 基因在荚粒发育过程中与逆境胁迫下的表达 [J]. 中国农业科技导报, 2023, 25 (8): 33-52.
- KE B Y, LI W L, ZHANG C Y. Expressions of SWEET genes during pod and seed developments and under different stress conditions in soybean [J]. *Journal of Agricultural Science and Technology*, 2023, 25(8): 33-52. (in Chinese)
- [36] PATIL G, VALLIYODAN B, DESHMUKH R, et al. Soybean (*Glycine max*) SWEET gene family: Insights through comparative genomics, transcriptome profiling and whole genome re-sequencing analysis [J]. *BMC Genomics*, 2015, 16(1): 520.
- [37] RANOGA P, DENANCÉ N, VANHOLME R, et al. Walls are thin 1 (WAT1), an *Arabidopsis* homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers [J]. *The Plant Journal: for Cell and Molecular Biology*, 2010, 63(3): 469-483.
- [38] PAL L, SANDHU S K, BHATIA D, et al. Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus* subsp. *napus*) [J]. *Physiology and Molecular Biology of Plants*, 2021, 27(9): 1933-1951.
- [39] LIU C, ZENG L B, ZHU S Y, et al. Draft genome analysis provides insights into the fiber yield, crude protein biosynthesis, and vegetative growth of domesticated ramie (*Boehmeria nivea* L. Gaud) [J]. *DNA Research: an International Journal for Rapid Publication of Reports on Genes and Genomes*, 2018, 25(2): 173-181.
- [40] 刘顺湖, 周瑞宝, 盖钧镒. 大豆蛋白质有关性状遗传的分离分析 [J]. 作物学报, 2009, 35 (11): 1958-1966.
- LIU S H, ZHOU R B, GAI J Y. Segregation analysis for inheritance of protein related traits in soybean [*Glycine max* (L.) merr.] [J]. *Acta Agronomica Sinica*, 2009, 35(11): 1958-1966. (in Chinese)

(责任编辑：于洪杰)