

基于 BSA-seq 技术定位绿豆种皮颜色基因

黄琬婷, 王 茜, 张泽燕, 朱慧珺, 闫虎斌, 张耀文

(山西农业大学农学院, 太原 030031)

摘要: 种皮颜色是作物重要的驯化性状和形态标记, 绿豆种皮颜色与黄酮类化合物含量有关。挖掘绿豆种皮颜色相关基因有助于开发新品种, 提高绿豆种皮的应用价值。本研究以冀绿 9 号(黑色种皮)和资源 330(黄色种皮)为亲本构建 F₂ 分离群体, 采用 BSA-seq 方法进行初定位。结果表明, SNPs 和 InDels 关联区域交集位于 4 号染色体共 3.26 Mb 的区段, 包含 324 个基因, 其中非同义突变基因共 49 个, 移码突变基因 15 个。进一步开发 KASP 分子标记进行精细定位, 筛选出高质量 KASP 引物 11 对。最终将绿豆种皮色位点定位于 4 号染色体上的 KASP330-KASP421 之间, 物理距离为 16,302,330~18,013,421 bp(1.71 Mb)。结合转录组数据分析和 qRT-PCR 表达分析共筛选出 6 个候选基因, 其中基因 *LOC106758748* 基因注释为 MYB90, 其功能为参与调控黄酮类化合物生物合成, 故可能是调控绿豆种皮颜色的关键基因。本研究结果可为绿豆种皮色相关基因的克隆和在育种中的利用奠定基础。

关键词: 绿豆; 种皮颜色; 基因定位; BSA-seq

Mapping of A Seed Coat Color Related Gene by BSA-seq in Mung Bean (*Vigna radiata* L.)

HUANG Wan-ting, WANG Qian, ZHANG Ze-yan, ZHU Hui-jun, YAN Hu-bin, ZHANG Yao-wen

(College of Agronomy, Shanxi Agricultural University, Taiyuan 030031)

Abstract: Seed coat color is an important agronomic trait that associates with crop domestication and serves as morphological marker. In mung bean, the seed coat color was related to the content of flavonoids. Cloning and application of seed coat color-related genes becomes of interest in development of new mung bean varieties with improved nutritional properties. In this study, the varieties 'Jilv 9' (black seed coat) and 'Ziyuan 330' (yellow seed coat) were used as parents to generate an F₂ segregating population. The BSA-seq approach was applied for preliminary mapping of the genes underlying the seed coat color. The association analysis using integrated SNPs and InDels suggested an interval of 3.26 Mb harboring 324 predicted genes, of which 49 genes were found with non-synonymous mutations and 15 genes were detected with frameshift mutation. By further use of 11 high-quality KASP markers in fine mapping, the candidate interval was finally delimited between KASP330 and KASP421 in the physical interval of 16,302,330–18,013,421 bp (1.71 Mb) on chromosome 4. The transcriptome data analysis and qRT-PCR expression analysis suggested six differently-expressed candidate genes, of which the *LOC106758748* was annotated as a transcription factor MYB90 that was reported with a function in the flavonoid biosynthesis and served as key candidate gene regulating the seed coat color in mung bean. This study laid a solid foundation for further cloning of mung bean seed coat color related gene(s) and the utilization in breeding.

收稿日期: 2022-10-25 修回日期: 2022-11-15 网络出版日期:
URL:

第一作者研究方向为绿豆分子遗传育种, E-mail: hwt503362871@163.com

通信作者: 张耀文, 研究方向为食用豆育种及栽培技术研究, E-mail: zyw8118571@126.com

基金项目: 山西省基础研究计划项目(20210302124504); 山西农业大学博士科研启动项目(2021BQ43); 山西省博士毕业生、博士后研究人员来晋工作奖励资金科研项目(SXBYKY2021050); 山西省农业科学院作物科学研究所博士基金计划(ZB1101)

Foundation project: Shanxi Basic Research Program Project (20210302124504), the Ph.D. of Shanxi Agricultural University Scientific Research Start-up Project (2021BQ43), the Shanxi Province Doctoral Graduates, Postdoctoral Researchers Come to Shanxi Province to Work Incentive Fund Scientific Research Project (SXBYKY2021050), the Shanxi Academy of Agricultural Sciences Institute of Crop Science Doctoral Fund Program (ZB1101)

Keywords: mung bean; the color of seed coat; mapping; BSA-seq

种皮由珠被（母本）发育而来，指被覆于种子周围的皮，可以保护种子的内部不受伤害。种皮中含有丰富的化合物，如黄酮、蛋白质和多肽等。研究表明，不同颜色的种皮中黄酮类化合物含量不同^[1]。黄酮类化合物是指 2 个苯环通过 3 个碳原子相互连接而成的一系列化合物的总称，其中包括黄酮醇、花青素和原花青素等，在作物种皮中广泛分布^[2]。作为亚洲最重要的豆类之一，绿豆种皮也富含黄酮类物质，并被广泛应用于功能食品和保健药物研发^[3]。深入研究绿豆种皮颜色的调控机理，可为今后功能性品种选育和开发利用奠定理论基础。

迄今为止，国内外研究人员对作物种皮颜色均已开展了较为频繁的研究。赵钰涵等^[4]利用二代 SNP 芯片结合 BSA 方法对花生黑种皮基因进行了定位，在 10 号染色体 70~109 Mb 区间内筛选到一个与黑种皮性状紧密连锁的 SSR 标记 pTsaSSR107.16；Herniter 等^[5]研究表明一个或多个 *MYB113* 的变异可能影响黑色素的合成，并认为 *Vigun05g039500* 可控制豇豆种皮的黑色素合成；García-Fernández 等^[6]认为白色菜豆种皮与非白色菜豆种皮分离比为 1:1，由单基因控制；而棕色种皮与黑色种皮分离比为 3:1，由两个上位性相互作用的基因控制，并进一步将白色种皮基因定位在 7 号染色体，黑色种皮基因则分别定位在 6 号和 8 号染色体。Zhang 等^[7]则发现花生红种皮受一单隐性基因 *AhRt2* 控制，并应用 BSA-seq、功能注释、表达质谱和序列变异等分析确定 *Arahy.IK60LM* 为花生红种皮的候选基因。

随着高通量全基因组测序技术的发展序列的释放，基于深度测序的批量分离分析(bulked segant analysis, BSA)从基因组中检测 SNP 和 INDEL 可快速发掘候选基因，并已在大豆^[8]、向日葵^[9]、油菜^[10]等物种中广泛应用，也被用于绿豆抗病虫基因的挖掘^[11-12]。作为小宗作物，虽然全基因组序列已经释放^[13]，但相关的分子遗传学及基因组学研究相对滞后。本研究利用冀绿 9 号（黑色种皮）和资源 330（黄色种皮）杂交的 F₂ 分离群体，构建了不同种皮色基因池，并通过 BSA-seq 全基因组重测序结果，初步筛选出调控绿豆种皮颜色的候选区域，在候选区域内设计 KASP 分子标记进行精细定位，并结合不同时期种皮转录组分析结果、GO、KEGG 富集分析，进一步筛选出调控绿豆种皮颜色的候选基因。研究结果对加快分子标记辅助选择育种及进一步阐明绿豆种皮颜色调控机制具有重要意义。

1 材料与方法

1.1 材料及表型观测

冀绿 9 号（河北）种皮黑色，结荚一致，不炸荚，抗倒性强；资源 330 种皮黄色，山西沁县农家品种。2019 年 6 月利用冀绿 9 号和资源 330 配置组合，同年 8 月底收获杂交 F₁ 代并统计表型数据。2020 年 5 月将亲本和杂交荚同时种植，收获 F₂ 脱粒后晒干备用，统计表型数据。2021 年 5 月将亲本及 F₂ 分离群体 527 株（后期 23 株花落未结荚），种植在山西省晋中市山西农业大学东阳试验基地，小区长 5 m，宽 2 m，4 行种植，每行 20 株，共种植 10 个小区，种子成熟后统计表型数据，进行卡方检验。

1.2 试验方法

1.2.1 混池测序 选取后代极端黑色种皮和黄色种皮绿豆各 30 株，取相对应植株顶端幼嫩叶片 1-2 片，称取 1-2 g，采用改良的 CTAB 法^[14]提 DNA，制备 DNA 3 μg，样品浓度 20 ng/μL。将黑色种皮 30 份 DNA 和

黄色种皮 30 份 DNA 分别等量混合，构建 2 个后代极端混池，送北京百迈克生物科技有限公司质检合格后用超声破碎将 DNA 随机打断成 350 bp 的片段，并经末端修复、加测序接头、纯化、PCR 扩增等，完成测序文库构建。文库质检合格后通过 Illumina HiSeq 上机测序。

1.2.2 测序数据分析 为了保证信息分析质量，原始测序序列依次通过以下步骤的筛选^[15]。(1)对原始序列过滤，去除带接头（adapter）的 reads。(2)若一条 reads 上 N（未能确定出具体的碱基类型）的比例大于 10%，过滤掉该 Pair end reads。(3)去除低质量 reads（质量值 $Q \leq 10$ 的碱基数占整条 read 的 50% 以上）。质量合格后，将二代高通量测序得到的短序列与参考基因组比对，通过比对定位 Clean Reads 在参考基因组上的位置，统计各样品的测序深度、基因组覆盖度等信息，并做变异检测。

1.2.3 SNP 和 InDel 的检测和分析 SNP 检测主要使用 GATK^[16]软件工具包实现。根据 Clean Reads 在绿豆参考基因组的定位结果，使用 Picard 去重复（mark duplicates）、GATK 局部重比对（local realignment）和碱基质量值校正（base recalibration）等预处理，以保证检测得到的 SNP 准确性，再使用 GATK 检测、过滤并得到最终的 SNP 位点集。运用 SnpEff 得到变异位点在基因组发生的区域；InDel 检测根据样品的 Clean Reads 在参考基因组上的定位结果，检测样品与参考基因组之间是否存在小片段的插入与缺失（Small InDel:1-5 bp）。

1.2.4 SNPs 和 InDels 关联分析 过滤后，得到高质量 SNPs 和 InDels 的位点。应用欧式距离法(euclidean distance, ED)^[17]和混池间的基因型频率差异做关联分析（简称 SNP-index）^[17-18]。ED 方法通过寻找不同颜色混池间存在显著差异标记，并以此评估与颜色性状关联区域。SNP-index 是通过混池间的基因型频率存在的，显著差异标记进行关联分析，用 $\Delta(\text{SNP-index})$ 值统计。标记 SNP 与性状关联度越强， $\Delta(\text{SNP-index})$ 越接近于 1。

1.2.5 KASP 分子标记的开发 基于冀绿 9 号与资源 330 的 BSA-seq 全基因组重测序数据，在初定位的候选区域内，选择非同义突变的 SNP 位点上下游 150 bp 序列，通过 DNAMAN 和 Primer 3. (<https://bioinfo.ut.ee/primer3-0.4.0/>) 设计引物，在每个 SNP 位点设计出两个等位基因特异的正向引物 (Allele-specific forward primer 1 and 2) 和一个通用反向引物 (Common reverse primer)，由生工生物工程股份有限公司合成引物。对引物进行 AQP 基因分型试验，反应体系如下：第一步：95℃ 10 min；第二步：95℃ 20 s，61-55℃ 40 s，每个循环降 0.8℃，共进行 10 个循环；第三步 95℃ 20 s，55℃ 40 s，共进行 35 个循环。反应结束后，将 PCR 产物放入荧光定量 PCR 仪中进行荧光数据读取，挑选出分型清晰的 KASP 引物。

1.2.6 候选区域筛选与功能注释 关联分析结果合并取交集，用 BLAST^[19]通过多个数据库 (NR、Swiss-Prot、GO、KEGG、COG^[20-23]) 深度注释候选区间内的编码基因，快速筛选候选基因。

1.2.7 荧光定量 PCR 取开花后第 10 天、第 15 天和第 20 天 3 个时期亲本的种皮，使用 RNAprep Pure Plant Plus Kit 法（多糖多酚植物总 RNA 提取试剂盒）提取 RNA，具体步骤按其说明书进行操作。反转录后进行 qRT-PCR 实验，采用二步法进行反应：第一步预变性 95℃ 30 s；第二步 PCR 反应 95℃ 5 s，60℃ 34 s，40 个循环，内参为 Actin3 (Vradi03g00210)，3 个生物学重复，3 次技术重复^[24]。

2 结果与分析

2.1 种皮颜色的遗传分析

表型分析表明，两个亲本种皮颜色有显著差异，冀绿 9 号种皮颜色为黑色，资源 330 种皮颜色为黄色。由于种皮是由珠被发育而来的，进行杂交获得 F₁ 种子种皮颜色为母本种皮颜色，F₂ 代种子种皮颜色为黑色，未发生分离。因此，相对于黄色种皮，黑色种皮是完全显性性状。F₃ 代种子种皮颜色发生分离，出现除亲本之外的中间型绿色种皮。卡方检验结果表明，F₃ 代种子种皮颜色黑色、绿色与黄色的分离比例符合 12:3:1(表 1)。由此推测绿豆的种皮颜色可能由两对基因控制，同时两对基因间可能存在显性上位作用。

表 1 绿豆种皮的遗传分析

Table 1 Genetic analysis of mung bean seed coat

种子世代	总株数	黑粒株数	绿粒株数	黄粒株数	分离比例	²	P-value
Generation	Total plants	Black	Green	Yellow	Expected ratio		
F ₁	25	20	0	0			
F ₂	98	98	0	0			
F ₃	504	372	97	35	12:3:1	2.007	0.05

2.2 基于 BSA-seq 技术的绿豆种皮颜色基因初定位

2.2.1 测序质量评估 对原始测序数据处理后最终得到 67.67 Gbp 高质量的 Clean Reads 数据。其中亲本冀绿 9 号和资源 330 的 Clean Reads 共为 104,954,882，GC 含量分别为 34.31%和 34.74%；黑色种皮和黄色种子代混池的 Clean Reads 共为 122,830,167，GC 含量分别为 33.75%和 33.76%，测序数据质量较高（Q20≥97.94%，Q30≥93.76%）（表 2）。

运用二代高通量 BWA 软件,比对定位 Clean Reads 在绿豆参考基因组上的位置,平均比对效率为 98.17%（表 2）。统计参考基因组上不同染色体区域的覆盖度和测序深度分布情况，覆盖深度在 23×~32×之间，1×覆盖率百分比超过 96.48%，5×覆盖率百分比超过 94.83%，10×覆盖率百分比超过 90.44%，基因组平均覆盖深度约为 28.00×，基因组覆盖度约为 97.53%（至少有一个碱基覆盖）。覆盖均匀，变异检测准确，测序随机性好，有利于下一步的检测和筛选。

表 2 原始测序数据与参考基因组比对结果统计分析

Table2 Statistical analysis of the comparison between raw sequencing data and reference genome

样本	过滤后序列	过滤后碱基数	Q20 含量	Q30 含量	GC 含量(%)	平均测序深度	1×覆盖率	5×覆盖率	10×覆盖率
Simple	数	Clean bases	(%)	(%)	Clean_GC_rate	度	(%)	(%)	(%)
	Clean reads		Q20	Q30		Ave_depth	Cov_ratio1X	Cov_ratio5X	Cov_ratio10X
J9	51,754,834	15,363,374,502	97.94	93.76	34.31	26	97.23	96.03	93.05
330	53,200,048	15,773,244,276	98.04	94.00	34.74	23	96.48	94.83	90.44
Bpool	60,865,703	18,099,744,344	98.01	97.64	33.75	31	98.19	97.26	95.47
Ypool	61,964,464	18,434,322,572	97.99	93.86	33.76	32	98.2	97.22	95.49

J9: 冀绿 9 号, 330: 资源 330; Bpool: 黑色种皮子代极端混池, Ypool: 黄色种皮子代极端混池。

J9: Jilv 9, 330: Ziyuan 330; Bpool: the black offspring mixing pool, Ypool: the yellow offspring mixing pool.

2.2.2 SNP 和 InDel 检测和注释 SNP 检测结果显示，亲本之间共获得 854,835 个 SNP 位点，其中非同义

突变的 SNP 18,506 个，混池之间获得的 SNP 数量为 251,864 个；InDel 检测亲本之间共获得 181,477 个 Small InDels（插入或缺失 1-5 bp）；混池之间共获得 55,563 个 Small InDels。亲本与混池之间共同有 SNP 数量为 564,545 个；共同有 InDel 数量为 117,598 个（图 1）。大部分 SNP 和 InDel 分布在基因间、基因上游区域、基因下游区域。

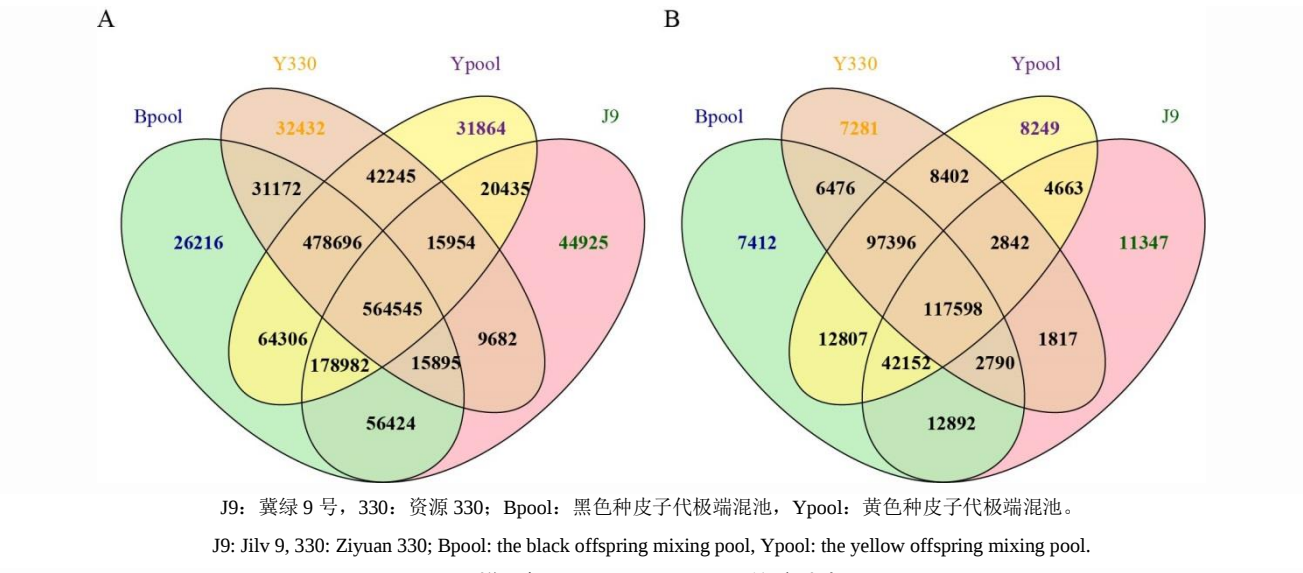
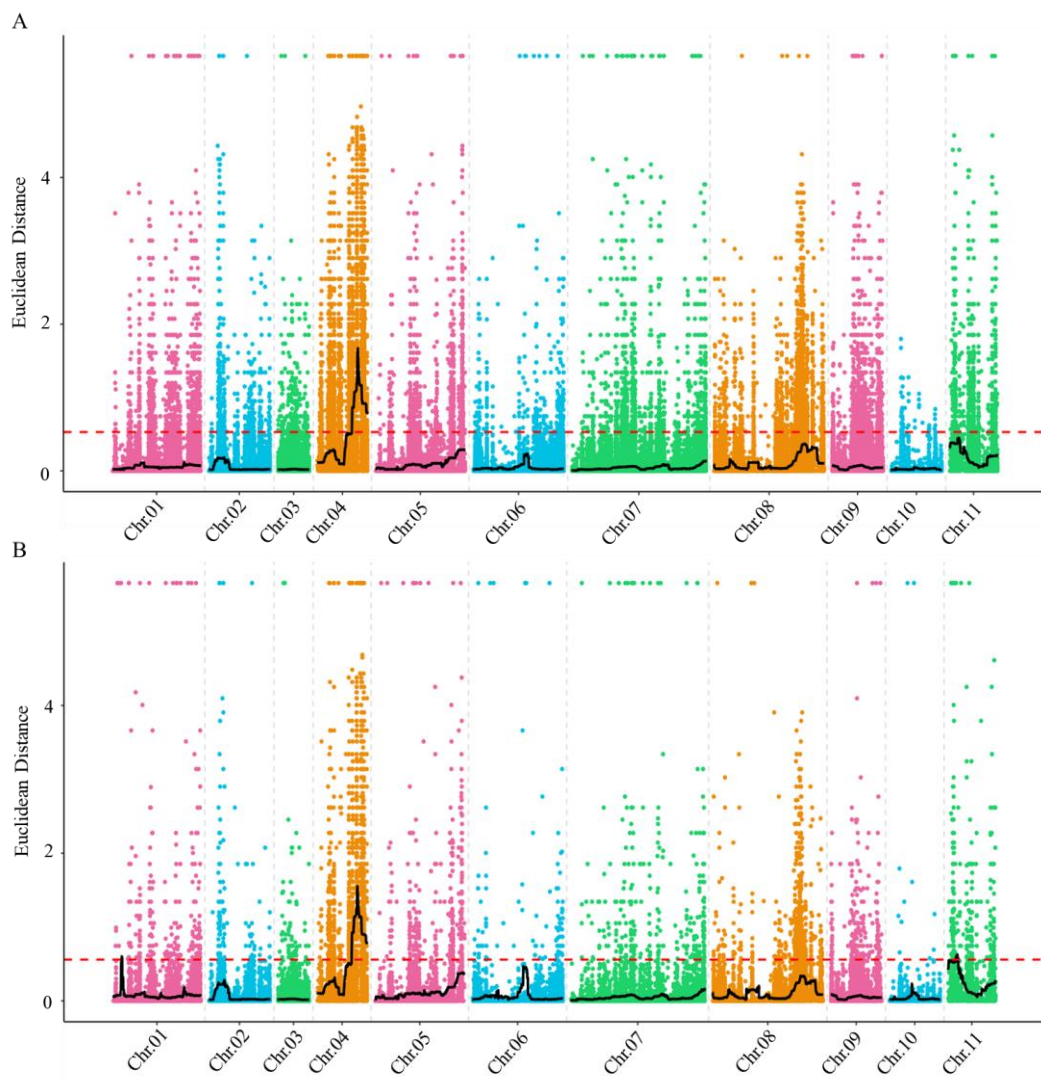


图 1 样品间 SNP(A)和 InDel(B)的统计韦恩图

Fig. 1 Statistical Venn plots of SNP(A) and InDel(B) between samples

2.2.3 基于关联分析的绿豆种皮颜色基因预测 过滤后得到高质量 SNP 位点 537,469 个。ED 算法进行关联分析，取所有位点拟合值的 median+3SD 作为分析的关联阈值，计算得出关联阈值 0.53，共得到 1 个区域，位于绿豆 4 号染色体长臂，总长度为 6.44 Mb，共包含 640 个基因，其中具有非同义突变位点的基因共 136 个（图 2）；SNP-index 算法关联阈值（0.6）共得到 2 个区域，均位于 4 号染色体长臂，总长度为 3.35 Mb，共包含 335 个基因，其中非同义突变位点的基因共 53 个。将 2 种分析方法的结果取交集，在 4 号染色体上得到 2 个定位区间，分别为 15,380,000~18,640,000 bp 和 19,000,000~19,090,000 bp，总长度为 3.35 Mb，共有 335 个基因。

对过滤后得到高质量 InDel 位点 116,462 个。ED 算法进行关联分析，关联阈值为 0.56，共得到 3 个区域，分别在 1 号、4 号、11 号染色体，总长度为 7.53 Mb，共包含 740 个基因，其中具有移码突变位点的基因共 42 个（图 2）；SNP-index 算法关联阈值（0.6）共得到 1 个区域，位于 4 号染色体，总长度为 3.33 Mb，共包含 328 个基因。将上述区域取交集得到一个区域：15,320,000~18,650,000 bp，总长度为 3.33 Mb，共包含 328 个基因，其中具有移码突变位点的基因共 16 个。对 SNPs 和 InDels 关联分析结果取交集，得到 4 号染色体 15,380,000~18,640,000 bp 的区域，总长度为 3.26 Mb，共有 324 个基因，包含非同义突变基因 49 个，移码突变基因 15 个。



横坐标为染色体名称，彩色的点代表每个 SNP/InDel 位点的 ED 值，黑色的线为拟合后的 ED 值，红色的虚线代表显著性关联阈值，ED 值越高，代表该点关联效果越好。

The abscissa: Chromosome ID; Each coloured dot represents an ED-based linkage value of an SNP/InDel site. Black line represents ED value after fitting. Red dashed line represents linkage threshold. A larger ED value indicates a stronger linkage of the SNP site to target trait.

图 2 SNP-ED(A)和 InDel-ED(B)关联值在染色体上的分布

Fig. 2 Distribution of SNP-ED(A) and InDel-ED(B) based linkage value on chromosomes

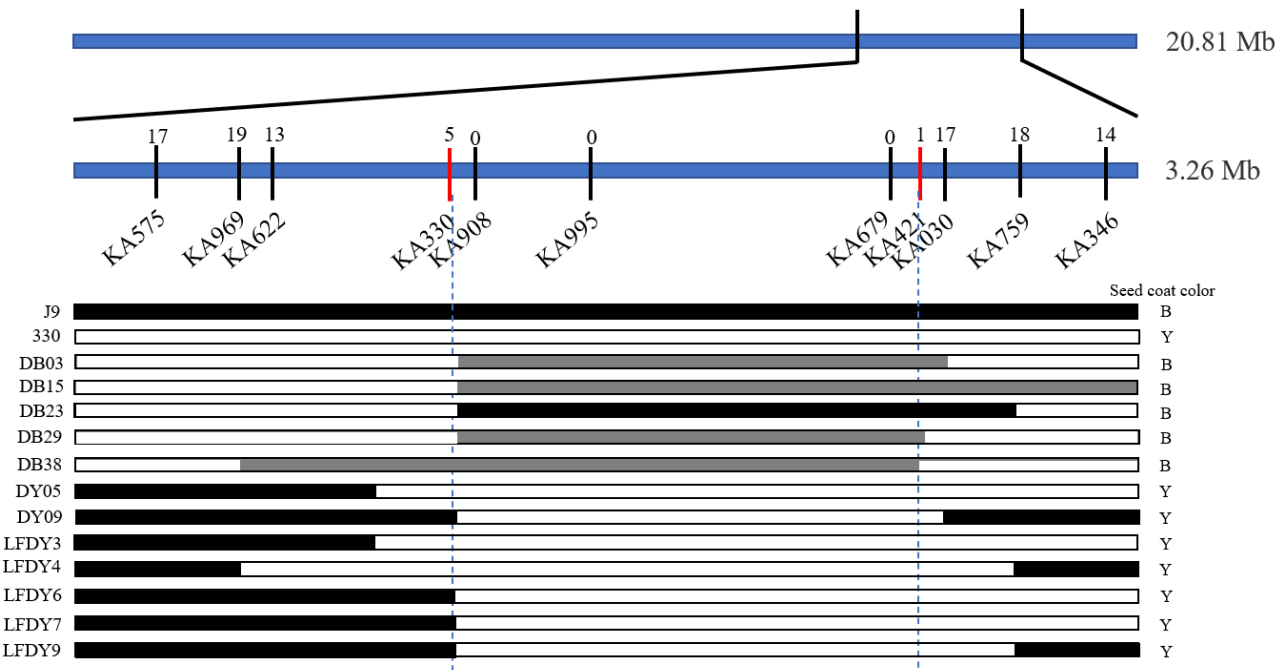
2.3 基于 KASP 标记的绿豆种皮色位点的精细定位

根据 BSA-seq 的测序结果，在初定位到的 3.26 Mb 候选区间内设计出 47 对 KASP 引物。应用 AQP 基因分型技术，筛选出高质量有明显分型趋势的 KASP 引物 11 对（表 3）。利用子代群体中表型为黑色种皮植株 372 株，表型为黄色种皮植株 35 株进行基因分型，其中标记 KASP575 和 KASP330 共检测到 17 株交换单株和 5 株交换单株，标记 KASP346 和 KASP421 共检测到 14 株交换单株和 1 株交换单株，故将候选区域逐步缩小至标记 KASP330 和 KASP421 之间，物理距离为 16,302,330-18,013,421 bp，长度为 1.71 Mb（图 3），共包含非同义突变基因 22 个，移码突变基因 2 个。

表 3 精细定位所用 KASP 分子标记

Table3 KASP Polymorphic molecular markers for fine mapping

位置	引物名	正向引物	反向引物
Position	称	Forward sequence (5'-3')	Reverse sequence (5'-3')
	Primer name		
1562557	KASP57	GAAGGTGACCAAGTTCATGCTCCAATAGTGTGCACTGGCTCC	TGAACTCAAAATATGTATGAAAGCAT
5	5	GAAGGTGCGAGTCAACGGATTCCAATAGTGTGCACTGGCTCT	
1588496	KASP96	GAAGGTGACCAAGTTCATGCTGGTGGGTTGGACTAGGTTCG	CACTAAGTGACCAACCCGACTCATT
9	9	GAAGGTGCGAGTCAACGGATTTGGTGGGTTGGACTAGGTTC	
1598162	KASP62	GAAGGTGACCAAGTTCATGCTTTGTTTCAAGCTCAAGAACTCTAGAA	TGCATGTGTTATTTGGGTTACCTTT
2	2	AT	
		GAAGGTGCGAGTCAACGGATTTGTTTCAAGCTCAAGAACTCTAGAA	
		AC	
1630233	KASP33	GAAGGTGACCAAGTTCATGCTCGTATAGACTGGCGATGGAAACG	CGATTCCGCGAACTGATGAAGAAATA
0	0	GAAGGTGCGAGTCAACGGATTCTGATAGACTGGCGATGGAAACT	
1638190	KASP90	GAAGGTGACCAAGTTCATGCTGCCACCCACTACTTCGAGT	TGAATCATCGGCTTCTGAAA
8	8	GAAGGTGCGAGTCAACGGATTGCCACCCACTACTTCGAGA	
1688299	KASP99	GAAGGTGACCAAGTTCATGCTTGGTAGTCCTCCGAACGGTTTA	CCAACCTTCGACCATTATTAGTGC
5	5	GAAGGTGCGAGTCAACGGATTGGTAGTCCTCCGAACGGTTTG	
1786767	KASP67	GAAGGTGACCAAGTTCATGCTCCTAAAGAGAGAAATGAGCGTCAA	GTATGTCGAACGTCCTCCTTGTTTC
9	9	GAAGGTGCGAGTCAACGGATTCTAAAGAGAGAAATGAGCGTCAG	
1801342	KASP42	GAAGGTGACCAAGTTCATGCTGCTCAAAACCCACGAGACTTAAT	GTTATTAATCCATTACAAATGTTTAATTT
1	1	GAAGGTGCGAGTCAACGGATTGCTCAAAACCCACGAGACTTAAG	GC
1801603	KASP03	GAAGGTGACCAAGTTCATGCTGACGGAAGCAGTTATAAAAGGGC	CCATGTAATCTCATCCACAGTGA
0	0	GAAGGTGCGAGTCAACGGATTGACGGAAGCAGTTATAAAAGGGG	
1826675	KASP75	GAAGGTGACCAAGTTCATGCTCTTAGCCCATCGGGAACCG	GACAATAGAGAATCAAGCCATGCGTAA
9	9	GAAGGTGCGAGTCAACGGATTCTTAGCCCATCGGGAACCC	
1855234	KASP34	GAAGGTGACCAAGTTCATGCTCATCCCTTTCAACTTAACCCTGATC	TCCCTCTTGTCATGCCTTCTTTA
6	6	GAAGGTGCGAGTCAACGGATTATCCCTTTCAACTTAACCCTGATG	



黑色、白色和灰色格分别表示冀绿 9 号、资源 330、杂合体基因型。B 为黑色种皮的交换单株，Y 为黄色种皮的交换单株。

The phenotypes and genotypes of the swapped individual plants used for fine mapping are shown in black, white and grey grids for Jilv 9, Ziyuan 330 and heterozygote genotypes. B is the exchange single plant with black seed coat, Y is the exchange single plant with yellow seed coat.

图 3 用于精细定位的交换单株的表型和基因型

Fig.3 Phenotypes and genotypes of exchanged individual plants used for fine mapping

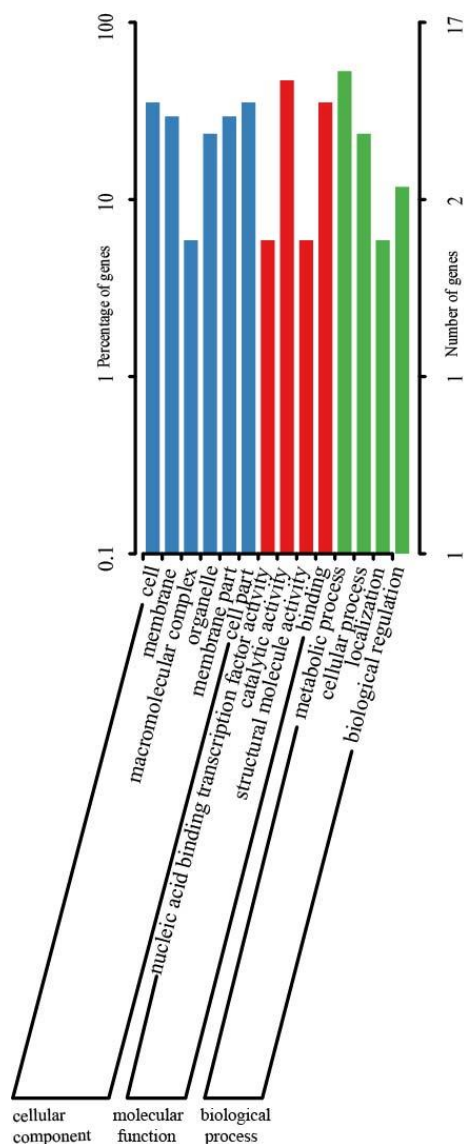
2.4 精细定位的候选区域筛选和功能注释

对精细定位区间内非同义突变和移码突变的 24 个基因进行功能注释（表 4）。其中 *LOC106759622*、*LOC106758992*、*LOC106759100*、*LOC106758437*、*LOC106759555*、*LOC106758197* 和 *LOC106758809* 在所有的基因注释文库里的功能未知，推测这 7 个基因可能为新基因或假基因。应用 BLAST 软件对候选区间内剩余的 17 个基因进行多个数据库（NR、Swiss-Prot、GO、KEGG、COG）的深度注释。其中 GO 注释聚类的基因，分别从细胞定位、分子功能、生物学过程 3 个方向进行分类注释，聚类基因共有 17 个（图 4）。

表 4 候选基因功能注释

Table4 Functional annotation of candidate genes

基因 ID	基因功能注释	
Gene ID	Gene function annotation	
LOC106758464	乙烯响应转录因子 ERF118	Ethylene-responsive transcription factor ERF118
LOC106758122	三螺旋转录因子 ASIL2	Trihelix transcription factor ASIL2
LOC106759554	叶绿体转录终止因子 MTEF1	transcription termination factor MTEF1, chloroplastic
LOC106758456	ALP1 蛋白	Protein ALP1
LOC106759353	含有 ELMO 结构域的蛋白质 A 亚型 X3	ELMO domain-containing protein A isoform X3
LOC106758052	枯草杆菌蛋白酶 Glyma18g48580	Subtilisin-like protease Glyma18g48580
LOC106758196	PRA1 家族蛋白 F4 类同种型 X1	PRA1 family protein F4-like isoform X1
LOC106758571	2-脱氢-3-脱氧磷酸酯醛缩酶	2-dehydro-3-deoxyphosphooctonate aldolase
LOC106759622	未知功能蛋白	uncharacterized protein
LOC106758992	未知功能蛋白	uncharacterized protein
LOC106759100	未知功能蛋白	uncharacterized protein
LOC106759605	半胱氨酸蛋白酶 ATG4 亚型 X1	Cysteine protease ATG4 isoform X1
LOC106758960	MDIS1 相互作用的受体激酶 2	MDIS1-interacting receptor like kinase 2
LOC106758788	枯草杆菌蛋白酶 Glyma18g48580	Subtilisin-like protease Glyma18g48580
LOC106758437	未知功能蛋白	uncharacterized protein
LOC106759555	未知功能蛋白	uncharacterized protein
LOC106758748	转录因子 MYB90	transcription factor MYB90
LOC106758197	未知功能蛋白	uncharacterized protein
LOC106758726	枯草杆菌蛋白酶 Glyma18g48580	Subtilisin-like protease Glyma18g48580
LOC106758701	含有五肽重复序列蛋白 At1g71460	Pentatricopeptide repeat-containing protein At1g71460
LOC106758809	未知功能蛋白	uncharacterized protein
LOC106759331	核苷	Nucleolin
LOC106758447	富含亮氨酸重复受体类蛋白激酶 At1g35710	Leucine-rich repeat receptor-like protein kinase At1g35710
LOC106758452	受体类蛋白 30	Receptor like protein 30



横坐标为 GO 各分类内容，纵坐标左边为基因数目所占百分比，右边为基因数目。此图展示的是关联区域内所有基因背景下 GO 各二级功能的基因分类情况。

X-axis: GO terms and classification; Y-axis: Percentage of total annotated genes (Left) and number of genes annotated to the term (Right). This figure shows the enrichment of genes in candidate regions to Level 2 GO terms.

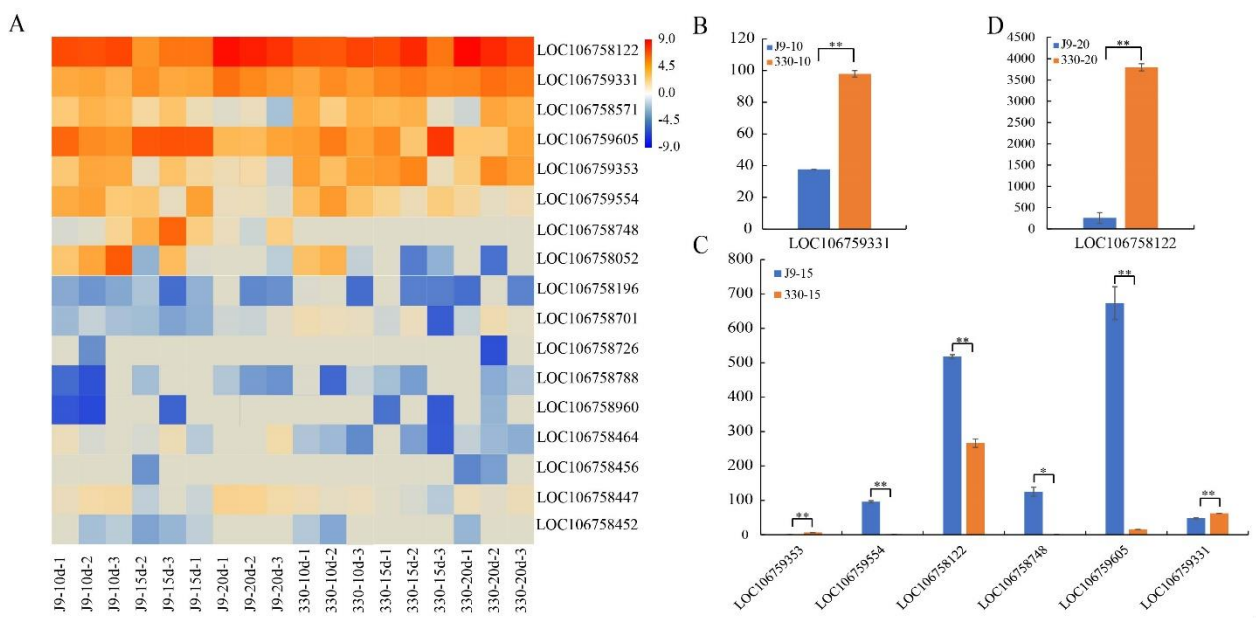
图 4 GO 基因富集分析

Fig. 4 GO enrichment of genes in candidate regions

2.5 调控绿豆种皮的关键基因预测

对候选区间内有功能注释的 17 个候选基因，进行荧光定量 PCR 验证。结合两个材料开花后种皮第 10 天、第 15 天和第 20 天三个时期转录组数据，筛选开花后同一时期两种材料的差异表达基因。GFOLD 绝对值大于 2 的基因被视为差异表达基因，根据上述基因的表达量（FPKM, fragments per kilobase of exon per million mapped）使用 R 包绘制热图（图 5），其中 *LOC106759331* 在第 10 天的冀绿 9 号和资源 330 种皮中表达差异显著，*LOC106759353*、*LOC106758701*、*LOC106758122*、*LOC106758748*、*LOC106759605* 和 *LOC106759331* 在第 15 天种皮中表达差异显著，*LOC106758122* 在第 20 天的种皮中表达差异显著。以上 3 组基因可能参与绿豆种皮颜色的调控。设计相关基因定量引物，使用 qRT-PCR 验证热图中差异表达基因。

在开花后第 10 天种皮样本中， *LOC106759331* 在资源 330 中的表达量远高于冀绿 9 号；在开花后第 15 天种皮样本中， *LOC106759353*、*LOC106759554*、*LOC106758122*、*LOC106758748*、*LOC106759605* 和 *LOC106759331* 在冀绿 9 号中的表达量高于资源 330；在开花后第 20 天种皮样本中 *LOC106758122* 在资源 330 中表达量较高，与第 15 天表达趋势相反；以上 6 个基因可能为调控绿豆种皮的关键基因。



J9: 冀绿 9 号; 330: 资源 330; 图 A: 候选基因热图。图 B: 开花后第 10 天, 冀绿 9 号和资源 330 种皮中差异基因表达分析; 图 C: 开花后第 15 天, 冀绿 9 号和资源 330 种皮中差异基因表达分析; 图 D: 开花后第 20 天, 冀绿 9 号和资源 330 种皮中差异基因表达分析。使用学生 *t* 测验进行显著性分析, **P*<0.05, ***P*<0.01。

J9: Jilv 9; 330: Ziyuan 330; (A) Heat map of candidate genes. (B) Differential gene expression analysis of Jilv 9 and Ziyuan 330 seed coat on the 10th day after flowering. (C) Differential gene expression analysis of Jilv 9 and Ziyuan 330 seed coat on the 15th day after flowering. (D) Differential gene expression analysis of Jilv 9 and Ziyuan 330 seed coat on the 20th day after flowering. Student *t* test was used for significance analysis, **P*<0.05, ***P*<0.01.

图 5 差异基因表达分析

Fig.5 Differential gene expression analysis

3 讨论

BSA-seq 是一种针对目标性状, 选择两个亲本和表型极端差异的子代构建混池进行全基因组重测序, 检测混池间 DNA 差异片段即为候选区间的先进技术^[7]。相比较传统的测序方式, 该技术可直接利用两亲本间具有多态性的 SNPs 进行定位, 缩小目标区域速度快, 准确率高, 广泛应用于挖掘候选基因的研究中。曾维英等^[8]应用大豆高抗豆卷叶螟材料赶泰-2-2 和高感豆卷叶螟材料皖 82-178 进行杂交构建 F₂ 代分离群体, 将 BSA-seq 与转录组测序相结合, 挖掘出 12 个控制大豆抗豆卷叶螟性状相关的候选基因; Rani 等^[11]利用两对杂交组合 F₂ 代群体中 144 个多态 SSR 标记进行 BSA 检测, 定位到绿豆 MYMIV 抗性基因位于六号染色体的两个 SSR 标记 GMAC7L 和 Satt322 之间的 77.115 kb 区间内; 程须珍等^[12]利用 BSA 检测获得一个共显性标记, 并初步认为此标记与 TC1966 的抗豆象基因位点紧密连锁。本研究选取的两个绿豆品种经多年培育, 稳定生长, 种皮颜色差异大, 适用于 BSA-seq 准确挑取极端个体, 快速定位到候选区域。

近年来, 种皮颜色的相关基因已有大量报道, 多数基因与黄酮类化合物密切相关。小麦中控制籽粒颜

色的基因 *Myb10* 定位在染色体 3A/3B/3D 长臂上, 该基因通过调节花青素的积累来控制种皮颜色^[25]; 水稻中 *Rc* 是控制种皮颜色的关键基因, 位于第 7 染色体上 R1807 和 R1582 之间, 是原花青素生物合成的转录因子, 参与类黄酮生物合成^[26]; Zhang 等^[27]利用 F₂ 分离群体鉴定到 *Arahy.IK60LM* 可能为花生红种皮的候选基因, 该基因编码花青素还原酶, 花青素作为黄酮类化合物, 具有很强的抗氧化能力和重要的营养价值, 花青素含量和组成是决定花生种皮颜色的重要因素; Chu 等^[28]通过 AG118 (黑色种皮) 和京农 6 号 (红色种皮) 小豆的 RNA-seq、qRT-PCR 分析, 鉴定出 *VaSDC1* 基因为调控黑色和红色种皮颜色的关键基因, *VaSDC1* 基因编码 R2R3-MYB 转录因子, 可以激活种皮中类黄酮代谢途径的结构基因, 导致小豆中红色和黑色种皮之间的差异。廉雪等^[1]研究发现绿豆种皮颜色与黄酮类化合物含量有关, 黑绿豆中黄酮类化合物含量高于绿绿豆和黄绿豆。

在 20 世纪中后期, 国外学者认为绿豆种皮颜色受一对基因控制^[29], 也有学者认为种皮颜色由两对基因控制^[30]。21 世纪初, 国内学者王丽侠等^[31]通过对黄绿豆(黄)×夹杆括角(绿)、明光绿豆(绿)×翁豆(黄)两对杂交组合进行三代验证, 认为绿色种皮对黄色种皮由单基因控制, 绿色为显性。贾龙等^[32]在研究绿豆相关农艺性状基因 QTL 定位时, 定位到绿豆种皮色基因位于 4 号染色体 17,486,086~18,014,973 bp, 但仅检测出单个 QTL, 贡献率达到 61%。本研究通过遗传分析结果推测绿豆种皮颜色可能受两对基因控制, 可能位于 4 号和 11 号染色体上 (图 2B)。由于 11 号染色体上 InDel-ED 关联值较低, 故本研究主要对 4 号染色体进行精细定位, 利用 BSA-seq 技术与 KASP 分子标记将控制种皮颜色的相关基因定位在 4 号染色体 16,302,330~18,013,421 bp, 共有 24 个候选基因, 与贾龙等定位到的候选区域接近。进一步结合 GO 数据库对候选基因的富集分析和转录组数据分析发现 6 个与种皮颜色相关的基因, 分别是 *LOC106758122*、*LOC106759605*、*LOC106759353*、*LOC106759554*、*LOC106759331* 和 *LOC106758748*, 其中 *LOC106758122* 的基因注释为三螺旋转录因子 ASIL2, 三螺旋转录因子主要参与光反应应答促进光合作用积累营养物质、参与生物与非生物逆境胁迫、干旱和冷胁迫^[33]; *LOC106759605* 基因注释为半胱氨酸蛋白酶 ATG4, 该基因在叶片衰老中高度表达^[34]; *LOC106759353* 基因注释为含有 ELMO 结构域的蛋白质 A 亚型 X3; *LOC106759554* 基因注释为叶绿体转录终止因子 MTEF1; *LOC106759331* 基因注释为核苷; *LOC106758748* 的基因注释为 MYB90, MYB90 为 R2R3-MYB 家族成员, 该家族参与黄酮类化合物生物合成, 花青素生物合成途径、黄酮醇合成途径和其他植物色素沉着^[35], 此家族成员已有报道参与颜色基因的调控, 故推测 *LOC106758748* 为该绿豆种皮色位点的候选基因。本研究存在的问题在于所定位到的候选区间大, 交换单株数量少, 候选基因数量多。下一步计划扩大群体数量, 进一步缩小定位区间, 针对定位区间中候选基因的鉴定以及功能验证的研究将在后续的实验中进行分析 and 验证。

4 结论

通过 BSA-Seq 全基因组重测序, 将绿豆种皮色候选基因定位到 4 号染色体 1.71 Mb 的区间。区间内共注释到 24 个基因, 其中非同义突变基因 22 个, 移码突变基因 2 个。结合基因的同源注释和转录组数据分析, *LOC106758122*、*LOC106759605*、*LOC106759353*、*LOC106759554*、*LOC106759331* 和 *LOC106758748* 可能为绿豆种皮颜色相关的候选基因, 这些基因可能在调控种皮颜色过程中起着重要作用, 为后期构建绿豆种皮色高密度遗传图谱奠定了基础。

参考文献

- [1] 廉雪, 张耀文. 不同种皮色绿豆黄酮类含量的初步研究. 山西农业科学, 2020, 48 (01): 55–57.
Lian X, Zhang Y W. Preliminary study on flavonoids content of mung bean in different seed coat color. Journal of Shanxi Agricultural Sciences, 2020, 48 (01): 55–57.
- [2] Shen S L, Tang Y S, Zhang C, Yin N W, Mao Y Y, Sun F J, Chen S, Hu R, Liu X Q, Shang G X, Liu L Z, Lu K, Li J N, Qu C. Metabolite profiling and transcriptome analysis provide insight into seed coat color in *Brassica juncea*. International Journal of Molecular Sciences, 2021, 22 (13): 7215.
- [3] Cantero G, Campanella C, Mateos S, Cortés F. Topoisomerase II inhibition and high yield of endoreduplication induced by the flavonoids luteolin and quercetin[J]. Mutagenesis, 2006, 21(05): 321-325.
- [4] 赵钰涵, 周希萌, 赵慧玲, 付春, 夏晗, 厉广辉, 赵术珍, 马长乐, 王兴军, 赵传志. 利用 SNP 芯片结合 BSA 的方法定位花生黑色种皮颜色基因. 分子植物育种, 2021, 19 (09): 2977–2984.
Zhao Y H, Zhou X M, Zhao H L, Fu C, Xia H, Li G H, Zhao S Z, Ma C L, Wang X J, Zhao C Z. Mapping of gene controlling black testa of peanut using SNP Array and BSA analysis. Molecular Plant Breeding, 2021, 19 (09): 2977–2984.
- [5] Herniter I A, Muñoz-Amatrián M, Lo S, Guo Y N, Close T J. Identification of candidate genes controlling black seed coat and pod tip color in cowpea (*Vigna unguiculata* (L.) Walp). Genes Genomes Genetics, 2018, 8 (10): 3347–3355.
- [6] García-Fernández C, Campa A, Ferreira J J. Dissecting the genetic control of seed coat color in a RIL population of common bean (*Phaseolus vulgaris* L.). Theoretical and Applied Genetics, 2021, 134 (11): 3687–3698.
- [7] Zhang K, Yuan M, Xia H, He L, Ma J, Wang M, Zhao H, Hou L, Zhao S, Li P, Tian R, Pan J, Li G, Thudi M, Ma C, Wang X, Zhao C. BSA-seq and genetic mapping reveals AhRt2 as a candidate gene responsible for red testa of peanut. Theor Appl Genet, 2022, 135 (05): 1529-1540.
- [8] 曾维英, 赖振光, 孙祖东, 杨守臻, 陈怀珠, 唐向民. 基于 BSA-Seq 和 RNA-Seq 方法鉴定大豆抗豆卷叶螟候选基因. 作物学报, 2021, 47 (08): 1460–1471.
Zeng W Y, Lai Z S, Sun Z D, Yang S Z, Chen H Z, Tang X M. Identification of the candidate genes of soybean resistance to bean pyralid (*Lamprosema indicata* Fabricius) by BSA-Seq and RNA-Seq. Acta Agronomica Sinica, 2021, 47 (08): 1460–1471.
- [9] Liu S, Wang P, Liu Y, Wang P. Identification of candidate gene for resistance to broomrape (*Orobanche cumana*) in sunflower by BSA-seq. Oil Crop Science, 2020, 5 (02): 47–51.
- [10] 张尧锋, 张冬青, 余华胜, 林宝刚, 华水金, 丁厚栋, 傅鹰. 基于极端混合池(BSA)全基因组重测序的甘蓝型油菜有限花序基因定位. 中国农业科学, 2018, 51 (16): 3029–3039.
Zhang Y F, Zhang D Q, Yu H S, Lin B G, Hua S J, Ding H D, Fu Y. Location and mapping of the determinate growth habit of *Brassica napus* by bulked segregant analysis (BSA) using whole genome re-sequencing. Scientia Agricultura Sinica, 2018, 51 (16): 3029–3039.
- [11] Rani A, Kumar V, Gill B S, Rath P, Shukla S, Singh R K, Husain S M. Linkage mapping of mungbean yellow mosaic India virus (MYMIV) resistance gene in soybean. Breed Science, 2017, 67 (02): 95–100.
- [12] 程须珍, 王素华, 吴绍宇, 周吉红, 王述民, 杨又迪. 绿豆抗豆象基因 PCR 标记的构建与应用. 中国农业科学, 2005, 38 (08): 6.
Cheng X Z, Wang S H, Wu S Y, Zhou J H, Wang S M, Yang Y D. Tagging and utilization of bruchid resistance gene using PCR markers in mungbean. Scientia Agricultura Sinica, 2005, 38 (08): 6.
- [13] Kang Y J, Kim S K, Kim M Y, Lestari P, Kim K H, Ha B K, Jun T H, Hwang W J, Lee T, Lee J, Shim S, Yoon M Y, Jang Y E, Han K S, Taeprayoon P, Yoon N, Somta P, Tanya P, Kim K S, Gwag J G, Moon J K, Lee Y H, Park B S, Bombarely A, Doyle J J, Jackson S A, Schafleitner R, Srinives P, Varshney R K, Lee S H. Genome sequence of mungbean and insights into evolution within *Vigna* species. Nat Commun, 2014, 5: 5443.
- [14] Porebski S, Bailey L G, Baum B R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter, 1997, 15 (01): 8–15.
- [15] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 2009, 25 (14): 1754–1760.
- [16] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M A. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 2010, 20 (09): 1297–1303.
- [17] Hill J T, Demarest B L, Bisgrove B W, Gorski B, Su Y C, Yost H J. MMAPP: mutation mapping analysis pipeline for pooled RNA-seq. Genome Research, 2013, 23 (04): 687–697.
- [18] Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, Saitoh H, Mitsuka C, Utsushi H, Uemura A, Kanzaki E, Kosugi S, Yoshida K, Cano L, Kamoun S, Terauchi R. MutMap+: genetic mapping and mutant identification without crossing in rice. The Public Library of Science, 2013, 8 (07): e68529.
- [19] Altschul S F, Madden T L, Schäffer A A, Zhang J, Zhang Z, Miller W, Lipman D J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 1997, 25 (17): 3389–3402.
- [20] 邓泱泱, 荔建琦, 吴松峰, 朱云平, 陈耀文, 贺福初. nr 数据库分析及其本地化. 计算机工程, 2006, 32(05): 71–74.

- Deng Y Y, Li J Q, Wu S F, Zhu Y P, Chen Y W, He F C. Integrated nr database in protein annotation system and its localization. *Computer Engineering*, 2006, 32 (05): 71–74.
- [21] Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S S, Eppig J T. Gene ontology: tool for the unification of biology. *Nature genetics*, 2000, 25 (01): 25–29.
- [22] Tatusov R L, Galperin M Y, Natale D A, Koonin E V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 2000, 28 (01): 33–36.
- [23] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Research*, 2004, 32 (suppl_1): D277–D280.
- [24] 李佳伟, 马钰聪, 杨鑫雷, 王梅, 崔顺立, 侯名语, 刘立峰, 胡梦蝶, 蒋晓霞, 穆国俊. 花生种皮色素合成相关通路的转录组-代谢组学联合分析. *植物遗传资源学报*, 2022, 23 (01): 240–254.
- Li J W, Ma Y C, Yang X L, Wang M, Cui S L, Hou M Y, Liu L F, Hu M D, Jiang X X, Mu G J. Transcriptomics-metabolomics combined analysis highlight the mechanism of testa pigment formation in peanut (*Arachis hypogaea* L.). *Journal of Plant Genetic Resources*, 2022, 23 (01): 240–254.
- [25] Groos C, Gay G, Perretant M R, Gervais L, Bernard M, Dedryver F, Charmet G. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a whitexred grain bread-wheat cross. *Theoretical and Applied Genetics*, 2002, 104 (1): 39–47.
- [26] Megan T S, Michael J T, Bernard E P, Susan M. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell*, 2006 (02), 18: 283–294.
- [27] Zhang K, Yuan M, Xia H, He L, Ma J, Wang M, Zhao H, Hou L, Zhao S, Li P, Tian R, Pan J, Li G, Thudi M, Ma C, Wang X, Zhao C. BSA-seq and genetic mapping reveals *AhRt2* as a candidate gene responsible for red testa of peanut. *Theoretical and Applied Genetics*, 2022, 135 (05): 1529–1540.
- [28] Chu L W, Zhao P, Wang K L, Zhao B, Li Y S, Yang K, Wan P. *VaSDC1* is involved in modulation of flavonoid metabolic pathways in black and red seed coats in adzuki bean (*Vigna angularis* L.). *Frontiers in Plant Science*, 2021, 12: 679892.
- [29] Manguiat P H, Lantican R, Carpena A, Catedral I. Inheritance of seed color, size, and luster in mungbean (*Vigna radiata* Wilczek). *Philippine Journal of Crop Science*, 1978, 3 (04): 238–241.
- [30] Bhadra S K, Akhter M I, Quashem A. Genetics of seed lustre and joint inheritance of seed-coat colour and seed lustre in mungbean (*Vigna radiata* (L.) Wilczek). *Bangladesh Journal of Botany (Bangladesh)*, 1991, 20: 61–64.
- [31] 王丽侠, 程须珍, 王素华, 刘岩. 绿豆几个表型性状的遗传特性. *作物学报*, 2013, 39 (07): 1172–1178.
- Wang L X, Cheng X Z, Wang S H, Liu Y. Inheritance of several traits in mungbean (*Vigna radiata*). *Acta Agronomica Sinica*, 2013, 39 (07): 1172–1178.
- [32] 贾龙, 王洁, 王素华, 陈红霖, 程须珍, 杜吉到, 王丽侠. 绿豆 11 个农艺性状相关基因的 QTL 定位. *植物遗传资源学报*, 2022, 23 (02): 442–449.
- Jia L, Wang J, Wang S H, Chen H L, Cheng X Z, Du J D, Wang L X. Identification of QTL associating with eleven agronomic traits in mungbean (*Vigna radiata* L.). *Journal of Plant Genetic Resources*, 2022, 23 (02): 442–449.
- [33] 于冰, 陈孟迪, 王宇光. 植物三螺旋 Trihelix 转录因子家族与环境相互作用的研究进展[J]. *植物遗传资源学报*, 2019, 20 (05): 1134–1140.
- Yu B, Chen M D, Wang Y G. Advances of plant trihelix transcription factor family interacting with environmental factors. *International Journal of Molecular Sciences*, 2019, 20 (05): 1134–1140.
- [34] Frank S, Hollmann J, Mulisch M, Matros A, Carrión CC, Mock HP, Hensel G, Krupinska K. Barley cysteine protease PAP14 plays a role in degradation of chloroplast proteins. *Journal of Experimental Botany*. 2019 18;70(21):6057–6069.
- [35] Ke Y J, Zheng Q D, Yao Y H, Ou Y, Chen J Y, Wang M J, Lai H P, Yan L, Liu Z J, Ai Y. Genome-wide identification of the MYB gene family in *Cymbidium ensifolium* and its expression analysis in different flower colors. *International Journal of Molecular Sciences*, 2021, 22 (24): 13245.