



# 野生大豆 × 栽培大豆 RIL 群体高密度遗传图谱构建及 SNP 偏分离分析

刘德泉<sup>1</sup>, 聂波涛<sup>1,2</sup>, 邱红梅<sup>1</sup>, 陈亮<sup>1</sup>, 陈健<sup>1</sup>, 崔正果<sup>1</sup>, 姬文秀<sup>2</sup>, 王跃强<sup>1</sup>

(1. 吉林省农业科学院 大豆研究所, 吉林 长春 130033; 2. 延边大学 农学院, 吉林 延吉 133002)

**摘要:** 通过研究野生大豆与栽培种大豆群体构建过程中产生的偏分离现象, 发掘偏分离区间 (Segregation Distortion Region, SDR) 和候选基因, 有助于探究偏分离在大豆中的产生机制。应用地方品种“一千粒”和野生品种“长岭野生豆”配置杂交组合, 获得 F<sub>2</sub> 代重组自交系 (Recombinant Inbred Lines, RIL) 株系 200 株, 利用 SLAF-seq 进行测序分析, 构建高密度遗传图谱, 获得该群体可靠的 4 564 个 SNP 标记。偏分离分析发现, 648 个标记发生偏分离 ( $P < 0.05$ ), 占总标记的 14.20%。获得 22 个 SDR, 分布在 9 个不同的染色体上。在 SDR 区间内共发现 8 个重度偏分离热点区域 (Extreme Segregation Distortion, ESDR), 分布在 5 个不同的染色体上, 其中 3 个 ESDR 偏向父本野生型, 5 个 ESDR 偏向母本栽培型。利用基因功能注释及全基因组重测序数据, 结合 ESDR 区域, 影响胚胎发育 (*Glyma. 01G051400*) 及雌配子体发育 (*Glyma. 16G072700*) 的基因分别被认为是 ESDR1-1 和 ESDR16-1 的候选基因。本研究结果为今后偏分离基因定位提供可靠依据, 并为阐明偏分离现象奠定基础。

**关键词:** 栽培大豆; 野生大豆; 偏分离; SNP; 重组自交系

## High Density Genetic Map Construction and SNP Segregation Distortion Analysis in RIL Population of Wild Soybean × Cultivated Soybean

LIU Dequan<sup>1</sup>, NIE Botao<sup>1,2</sup>, QIU Hongmei<sup>1</sup>, CHEN Liang<sup>1</sup>, CHEN Jian<sup>1</sup>, CUI Zhengguo<sup>1</sup>, JI Wenxiu<sup>2</sup>, WANG Yueqiang<sup>1</sup>

(1. Soybean Research Institute, Jilin Academy of Agricultural Sciences, Changchun 130033, China; 2. College of Agriculture, Yanbian University, Yanji 133002, China)

**Abstract:** This study examined the segregation distortion during the process of constructing wild soybean and cultivated soybean population, to explore the segregation distortion regions (SDRs) and candidate genes, and to shed some light on understanding the mechanism of segregation distortion in soybean. The wild variety ‘Changling wild soybean’ was used as the male parent and the landrace variety ‘Yiqianli’ was used as the female parent to develop a hybrid group, resulting in 200 recombinant inbred lines (RIL). SLAF-seq was used for sequencing analysis and constructing a high-density genetic map. 4 564 SNP markers were obtained and reliably identified for this population. Through segregation distortion analysis, 648 markers were found to have genetic distortion ( $P < 0.05$ ), accounting for 14.20% of the total markers. 22 SDRs were found, which were distributed across 9 different chromosomes. In SDRs, 8 extreme SDRs regions (ESDRs) were found, distributed on 5 different chromosomes, of which 3 ESDRs were biased towards the wild type of the male parent, and 5 ESDRs were biased towards the cultured type of the female parent. Using gene function annotation and genome-wide resequencing data, combined with the ESDR region, the affecting embryonic development (*Glyma. 01G051400*) and female gametophyte development (*Glyma. 16G072700*) genes were identified as candidate genes in ESDR1-1 and ESDR16-1, respectively. This study provides a reliable basis for locating the segregation distortion genes in the future, and lays the foundation for elucidating the segregation distortion.

**Keywords:** cultivated soybean; wild soybean; partial segregation; SNP; recombinant inbred line

偏分离 (Segregation Distortion, SD) 是指后代分离群体中基因型频率偏离孟德尔理论的分离比现象<sup>[1]</sup>。偏分离在大多数动植物杂交过程中均有发现, 被认为是一种重要的生物进化动力, 其通过影响基因型频率进而影响群体的进化方向<sup>[2]</sup>。研究发现, 大部分偏分离现象是由偏分离热点区域 (Segregation Distortion Region, SDR)、遗传搭车效应、配子体基因及环境因素等造成的<sup>[3-5]</sup>。偏分离的出现频率在不同双亲遗传背景、不同群体类型及

不同杂交组合中存在差异<sup>[6]</sup>。种间群体的偏分离比例要高于种内群体, 重组自交系 (Recombinant Inbred Lines, RIL) 群体的比例要高于 F<sub>2</sub> 群体和回交群体<sup>[7]</sup>。此外, RIL 群体中存在可稳定遗传的 SDR, 这些 SDR 可能与造成偏分离的特定基因相关<sup>[8]</sup>。

在大豆中, 分离群体的偏分离现象也普遍存在, 特别是在栽培种大豆 × 野生大豆的群体中<sup>[9]</sup>。生物学证据支持栽培大豆是从野生大豆驯化而来,

收稿日期: 2023-06-13

基金项目: 吉林省农业科技创新工程 (CXGC2022RCG005, CXGC2022RCY008)。

第一作者: 刘德泉 (1988—), 男, 博士研究生, 主要从事大豆资源与遗传育种研究。E-mail: liudequan5566@live.com;

聂波涛 (1998—), 男, 硕士研究生, 主要从事大豆遗传育种研究。E-mail: 2021010600@ybu.edu.cn。

通讯作者: 王跃强 (1968—), 男, 博士, 研究员, 主要从事大豆资源与遗传育种研究。E-mail: 825164942@qq.com。

驯化过程使大豆逐渐失去野生祖先种的部分生理形态和遗传特性,能够满足人类需要的遗传变异被不断积累和强化,最终形成高产优质的现代栽培种<sup>[10]</sup>。通过对偏分离位点的研究不但能进一步解析大豆驯化进程,而且有利于挖掘可能的关键基因。目前,对于导致大豆偏分离现象的基因挖掘还处在定位分子标记和物理区间<sup>[11-12]</sup>,尚鲜有功能基因被克隆。

过往研究使用 SSR、AFLP、RFLP 和 RAPD 等多种分子标记构建大豆遗传图谱,同时发现了众多偏分离位点<sup>[13-15]</sup>。但由于先前的群体分子标记密度过低,所定位的区间都不够精确。随着大豆全基因组测序信息的公布,利用重测序紧密连锁的单核苷酸多态性(SNP)分子标记,能够构建高密度遗传图谱。本研究即利用 SNP 标记构建“一千粒”(地方品种)×“长岭野生豆”的 200 个 F<sub>6</sub> 代 RIL 株系的高密度遗传图谱,发现并更精准地确认了遗传群体中的 SDR 区间和重度偏分离区间(Extreme Segregation Distortion,ESDR),同时利用重测序技术探究 ESDR 区间内可能的候选基因。该研究旨在为偏分离基因的克隆打下基础,同时也为阐明大豆偏分离规律及大豆驯化过程提供可靠参考。

1 材料与方法

1.1 群体构建

大豆(*Glycine max*)材料来自吉林省农业科学院大豆研究所,以“一千粒”(地方品种)作为母本,“长岭野生豆”作为父本配置杂交组合,随后采用单粒传法(single seed descent)连续自花授粉,至 F<sub>6</sub> 代无选择。F<sub>3</sub> 群体于 2015 年种植于吉林省农业科学院公主岭试验地盆栽场,以单株盆栽繁殖获得 F<sub>6</sub> 代 RIL 株系 200 株。

1.2 遗传图谱构建

使用改良的十六烷基三甲基溴化铵(cetyltrimethylammonium bromide,CTAB)方法<sup>[16]</sup>提取 RIL 群体各单株及两父母本叶组织总 DNA。由北京百迈客生物科技有限公司自主研发的 SLAF-Seq (Specific-Locus Amplified Fragment Sequencing)<sup>[17]</sup>进行测序分析,获得该群体可靠的 SNP 标记。采用 R 软件构建遗传图谱。

1.3 基因频率分析

基因频率(gene frequency)是指在该群体中,某个基因占全部等位基因数的比率。“一千粒”的基因频率计算方法为  $X_{gr} = (X_1 + 0.5 \times X_3) / (X_1 + X_2 + X_3)$ ,其中,X<sub>1</sub>为“一千粒”的基因型个体数,X<sub>2</sub>为“长岭野生豆”的基因型个体数,X<sub>3</sub>为杂合基因型个体数。

1.4 偏分离分析

将亲本基因频率在群体中分布按孟德尔分离比(1:1)进行卡方检测。位点卡方检验计算公式  $\chi^2 = [(X_a - X_e)^2 + (X_b - X_e)^2] / X_e$ ,其中 X<sub>a</sub>为标记“一千粒”基因观测值,X<sub>b</sub>为标记“长岭野生豆”基因观测值,X<sub>e</sub>为标记位点按孟德尔分离比的期望值。 $df$ (自由度)=1,选择显著水平  $\alpha = 0.05$  或  $\alpha = 0.01$ ,求得  $P < 0.05$  的  $\chi^2$  临界值为 3.841, $P < 0.01$  的  $\chi^2$  临界值为 6.635,以此判断 SNP 标记位点的偏分离显著情况。本研究将存在大于等于 6 个以上相邻的  $P < 0.05$  的偏分离 SNP 且偏分离方向一致的染色体区域,定义为 1 个 SDR<sup>[18]</sup>;6 个以上相邻的  $P < 0.01$  的同侧偏分离 SNP,定义为 1 个 ESDR。

1.5 候选基因筛选

依据群体连锁不平衡(Linkage Disequilibrium, LD)衰减距离,截取 ESDR 区间上下游 200 kb 区域作为候选区间,参照 Phytozome 13 (<https://phytozome.jgi.doe.gov>)数据库 G. max Wm82. a4. v1 大豆基因组序列,筛选出 ESDR 区间基因。利用高通量重测序技术,进一步标记“一千粒”和“长岭野生豆”存在的单核苷酸变异 SNV(Single-Nucleotide Variant)和 Indel 的序列差异,再根据大豆参考基因组(G. max Wm82. a4. v1)的注释及 RNA-Seq 数据,进一步选择偏分离相关基因作为候选基因。使用Tbtools 构建候选基因在不同组织表达量的 heatmap 图<sup>[19]</sup>。

2 结果与分析

2.1 基因型及基因频率在 RIL 群体的分布

SLAF-seq 测序分析获得该群体 4 564 个可靠的 SNP 标记,构建获得 20 个连锁群,总图距为 3 803.54 cM,平均间距为 0.83 cM(图 1、表 1)。统计分析 RIL 群体中基因型的分布,发现来自栽培种“一千粒”的基因型 AA 占 46.79%,来自“长岭野生豆”的基因型 BB 为 47.68%,杂合基因型占 2.83%,缺失基因型占 2.71%。亲本基因型在群体中的分离比为 0.98:1。不同染色体中基因型分离比例不同,其中 7 条染色体母本的遗传贡献率高于父本,13 条染色体父本的遗传贡献率高于母本。17 条染色体的亲本基因型分离比为 1/1.19 ~ 1/0.81。1 号和 12 号染色体基因型分离比为 1/1.37 和 1/1.42,其更多偏向父本基因型;而 16 号染色体基因型分离比为 1/0.71,其更多偏向母本基因型。在 200 个株系中,每个单株 20.40% ~ 75.69% 的基因型来源于母本,平均为 48.34%;22.71% ~ 76.09% 的基因型来源于父本,平均为 49.09%;0.04% ~ 15.87% 的杂合基因型,平均为 2.87%。基因频率在群体各单株的比例为 1:3.51 ~ 1:0.31,平均为 1:1.12。

表 1 RIL 群体 SNP 基因型分布  
Table 1 Distribution of SNP genotype in RIL population

染色体 Chromosome	SNP 数目 Number of SNP	AA 基因型数目 Number of AA genotype	BB 基因型数目 Number of BB genotype	AB 基因型数目 Number of AB genotype	缺失数 Missing number	基因型分离比例 Segregation ratio of genotype
1	186	14348	19624	1956	1272	1/1.37
2	300	29343	30315	266	76	1/1.03
3	327	29855	33184	1668	693	1/1.11
4	300	30614	27230	1020	1136	1/0.89
5	164	14826	17638	154	182	1/1.19
6	461	45958	45389	509	344	1/0.99
7	217	20185	21015	1486	714	1/1.04
8	178	16456	16831	1808	505	1/1.02
9	268	25722	26993	608	277	1/1.05
10	96	7193	8520	954	2533	1/1.18
11	75	7124	5764	829	1283	1/0.81
12	122	8575	12206	1802	1817	1/1.42
13	130	10952	11092	1193	2763	1/1.01
14	374	34770	36933	1749	1348	1/1.06
15	223	19058	22162	1933	1447	1/1.16
16	166	16592	11712	1679	3217	1/0.71
17	115	8928	10210	1228	2634	1/1.14
18	475	48752	43742	2283	223	1/0.90
19	216	21917	20195	773	315	1/0.92
20	171	15903	14423	1935	1939	1/0.90

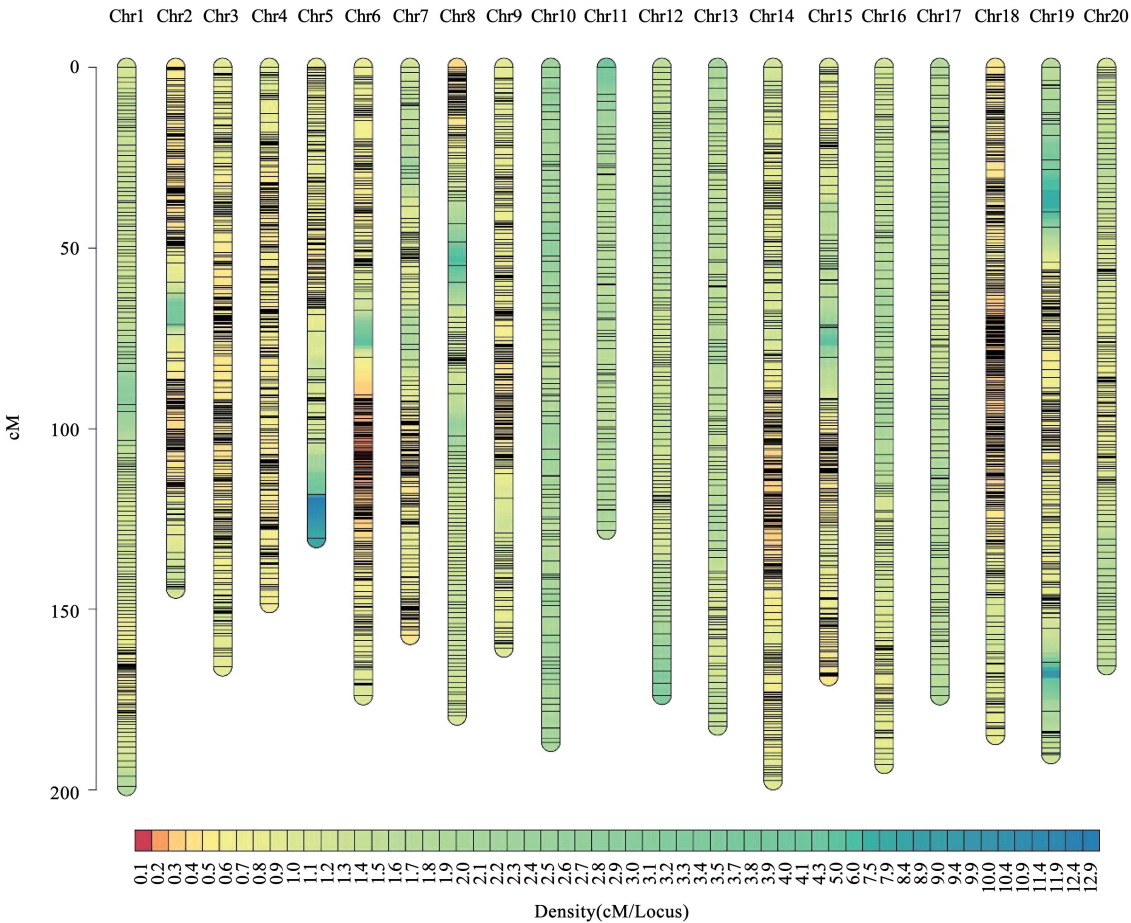


图 1 染色体标记分布图  
Fig. 1 Linkage group marker distribution map

2.2 偏分离标记的统计分析

对 4 564 个 SNP 标记进行卡方检验,  $P < 0.05$  水平上的偏分离标记 648 个, 频率为 14.20%。其中偏向母本栽培品种“一千粒”的 SNP 有 231 个, 占偏分离 SNP 标记的 35.65%, 偏向父本“长岭野生豆”的 SNP 有 417 个, 占偏分离 SNP 标记的 64.35%。  $P < 0.01$  水平上的偏分离标记 218 个, 频率为 4.78%。通过表 2 可知, 20 个遗传连锁群 SNP 标记偏分离的频率可划分为 4 个等级, 其中第 1、8、12 和 16 号染色体上所占偏分离标记比例超过

50%, 分别为 68.28%、82.58%、59.84% 和 68.67%; 第 5、10、11 和 13 号染色体上存在 22.92% ~ 36.00% 的标记发生偏分离; 第 9、15、18 和 19 号染色体存在 3.37% ~ 6.94% 的标记发生偏分离; 第 2、3、4、6、7、14、17 和 20 号染色体只存在 1 ~ 2 个标记偏分离或没有标记发生偏分离。分析发现各染色体 SNP 的偏分离方向也有所不同, 在偏分离发生频率高于 22.92% 的染色体中, 第 1、5、10 和 12 号染色体偏分离 SNP 全部偏向父本野生型, 而第 11 和 16 号染色体偏分离 SNP 全部偏向母本栽培型。

表 2 20 条染色体上总 SNPs 及偏分离 SNPs

Table 2 Summary of SNPs and segregation distortion SNPs in 20 chromosomes

染色体	SNP 数目	偏分离 SNP 数目	占各染色体 SNP 比例	偏向母本数目	偏向父本数目
Chromosome	Number of SNP	Number of segregation distortion SNP	Percentage in total SNPs in relevant chromosome/%	Number of deviated towards female	Number of deviated towards male
1	186	127	68.28	0	127
2	300	2	0.67	1	1
3	327	1	0.31	0	1
4	300	2	0.67	0	2
5	164	47	28.66	0	47
6	461	0	0.00	0	0
7	217	0	0.00	0	0
8	178	147	82.58	60	87
9	268	13	4.85	1	12
10	96	22	22.92	0	22
11	75	27	36.00	27	0
12	122	73	59.84	0	73
13	130	30	23.08	18	12
14	374	0	0.00	0	0
15	223	12	5.38	0	12
16	166	114	68.67	114	0
17	115	0	0.00	0	0
18	475	16	3.37	5	11
19	216	15	6.94	5	10
20	171	0	0.00	0	0

2.3 偏分离热点区域

通过对偏分离 SNP 在染色体上的分布进行分析发现, 偏分离 SNP 在染色体呈现连续、同方向分布。本研究共发现 22 个 SDR, 单个 SDR 内的偏分离 SNP 数为 6 ~ 61, SDR 区间的物理距离跨度在 0.3 ~ 14 239.1 kb。这些 SDR 分布在 9 个不同的染色体上(表 3)。其中 15 个 SDR 偏向父本野生型, 7 个 SDR 偏向母本栽培型。第 1、5、10、12 和 15 号

染色体上的 SDR 都偏向父本野生型, 第 11、13 和 16 号染色体上的 SDR 都偏向母本栽培型, 偏向不同方向 SDR 只存在于第 8 号染色体上。  $P < 0.01$  偏分离 SNP, 6 个以上相邻同方向 ESDR 有 8 个, 分布在 5 个不同的染色体上, 其中 3 个 ESDR 偏向父本野生型, 5 个 ESDR 偏向母本栽培型(表 4)。单个 ESDR 内的偏分离 SNP 数为 6 ~ 25, ESDR 区间的物理距离跨度在 61.6 ~ 14 335.3 kb。

表 3 SDRs 在遗传连锁图上的分布

Table 3 Distribution of SDRs in the genetic linkage map

偏分离热点区域 Segregation distortion region	染色体 Chromosome	SNP 区间 SNP interval	偏分离 SNP 数目 Number of segregation distortion SNP	SDR 跨度 Span of SDR/bp	偏向 Direction of deviated
SDR1-1	1	Marker566 ~ Marker606	16	5413514 ~ 5473048	male
SDR1-2	1	Marker633 ~ Marker807	41	5640704 ~ 6475886	male
SDR1-3	1	Marker842 ~ Marker1002	29	6512646 ~ 7555877	male
SDR1-4	1	Marker1031 ~ Marker1302	24	7629410 ~ 9039176	male
SDR5-1	5	Marker23327 ~ Marker22220	16	1341085 ~ 1899517	male
SDR5-2	5	Marker22231 ~ Marker22363	31	2045565 ~ 3603742	male
SDR8-1	8	Marker35750 ~ Marker35815	24	549501 ~ 1464847	male
SDR8-2	8	Marker35835 ~ Marker35875	15	1704044 ~ 2192638	male
SDR8-3	8	Marker35896 ~ Marker36010	23	2441551 ~ 4135502	male
SDR8-4	8	Marker36448 ~ Marker36706	49	10131835 ~ 13275563	female
SDR8-5	8	Marker39897 ~ Marker39938	15	46628828 ~ 47208047	male
SDR10-1	10	Marker51141 ~ Marker51261	10	43024558 ~ 44461708	male
SDR11-1	11	Marker51999 ~ Marker52118	23	1857347 ~ 3750776	female
SDR12-1	12	Marker56964 ~ Marker56970	7	8630401 ~ 8630725	male
SDR12-2	12	Marker57049 ~ Marker57177	32	9124003 ~ 10053876	male
SDR12-3	12	Marker57204 ~ Marker59187	21	10380979 ~ 24620069	male
SDR13-1	13	Marker64412 ~ Marker64425	6	44916876 ~ 45041080	female
SDR13-2	13	Marker64439 ~ Marker64474	7	45205279 ~ 45841137	female
SDR15-1	15	Marker72238 ~ Marker72725	13	5308659 ~ 11297235	male
SDR16-1	16	Marker793 ~ Marker79899	31	6407275 ~ 8399500	female
SDR16-2	16	Marker57072 ~ Marker80283	9	9394156 ~ 14567738	female
SDR16-3	16	Marker80907 ~ Marker82613	61	19260738 ~ 32432522	female

表 4 ESDRs 在遗传连锁图上的分布

Table 4 ESDRs in the genetic linkage map

重度偏分离热点区域 Segregation distortion region	染色体 Chromosome	SNP 区间 SNP interval	重度偏分离 SNP 数目 Number of extreme segregation distortion SNP	ESDR 跨度 Span of ESDR/bp	偏向 Direction of deviated
ESDR1-1	1	Marker668 ~ Marker746	14	5947964 ~ 6170533	male
ESDR8-1	8	Marker36531 ~ Marker36722	25	11111046 ~ 13512901	female
ESDR11-1	11	Marker52047 ~ Marker52091	10	2679691 ~ 3111014	female
ESDR12-1	12	Marker57133 ~ Marker57148	9	9665030 ~ 9726634	male
ESDR12-2	12	Marker7286 ~ Marker59187	21	10284762 ~ 24620069	male
ESDR16-1	16	Marker793 ~ Marker79585	6	6407275 ~ 7550604	female
ESDR16-2	16	Marker7170 ~ Marker80956	8	8767933 ~ 18521571	female
ESDR16-3	16	Marker80939 ~ Marker80588	7	18883055 ~ 22194619	female

### 2.4 探究偏分离候选基因

以ESDR区间上下游200 kb作为候选区间,参考G. max Wm82. a4. v1,区间内共401个基因。分析“一千粒”和“长岭野生豆”全基因组重测序数据,发现12 572个SNV和1 086个Indel在ESDR区间内存在差异,363个候选基因所在区域包含1个及以上SNV或Indel的序列差异。SNV位点在外显子有381个非同义突变和290个同义突变,5'-UTR和3'-UTR分别为163和203个,671和757个在上游和下游的突变。Indel在外显子有12个非移码突变和17个移码突变,5'-UTR和3'-UTR分别为27和25个,98和101个在上游和下游的突变(表5)。进一步通过基因注释筛选出8个基因编码的蛋白参与配子体形成和胚胎发育过程。其中4个基因(*Glyma. 01G051400*、*Glyma. 08G152100*、*Glyma. 12G107100*、*Glyma. 16G072700*)在外显子区域存在非同义SNV,6个基因(*Glyma. 01G051400*、*Glyma. 01G049900*、*Glyma. 08G161000*、*Glyma. 08G162000*、*Glyma. 12G107100*、*Glyma. 16G073000*)在启动子区域(5'-UTR上游1 kb)存在1个以上SNV差异。*Glyma. 01G051400*和*Glyma. 16G072700*基因存在非移码删除Indel;*Glyma. 01G051400*、*Glyma. 01G049900*、*Glyma. 08G162000*和*Glyma. 12G107100*启动子区域

存在Indel位点。此外,*Glyma. 16G072700*基因在mRNA拼接(splicing)位点存在1个Indel差异(表6)。这8个基因的Heatmap图揭示了其在大豆不同组织和不同发育阶段基因的表达情况(图2)。3个基因(*Glyma. 01G051400*、*Glyma. 08G161000*和*Glyma. 16G073000*)在配子体发育期的大豆花蕾中高表达,*Glyma. 01G051400*和*Glyma. 08G161000*也在种子胚发育早期高表达。

进一步分析候选基因的编码序列,编码胚胎缺陷蛋白(*emb1579*)的*Glyma. 01G051400*基因在启动子区域存在12个SNV突变和1个Indel突变,外显子存在9个SNV突变和1个Indel缺失突变。其中外显子5个非同义突变SNV和1个21 bp非移码缺失Indel(表7、图3A)。通过对比该基因所翻译的蛋白质,发现“长岭野生豆”相较于“Williams 82”和“一千粒”存在7个氨基酸的缺失(图3B)。*Glyma. 16G072700*基因编码大豆MYB98转录因子蛋白,该基因在外显子区域存在2个非同义SNV突变和1个非移码缺失,内含子区域存在1处mRNA拼接(splicing)位点差异。比对蛋白质序列发现该基因在野生豆编码蛋白质不同位置中存在59个氨基酸的增加及4个氨基酸的缺失(图4)。

表5 ESDRs内候选基因SNV与Indel突变的分类汇总

Table 5 Categorized summary of SNV and Indel mutations of candidate genes in ESDR region

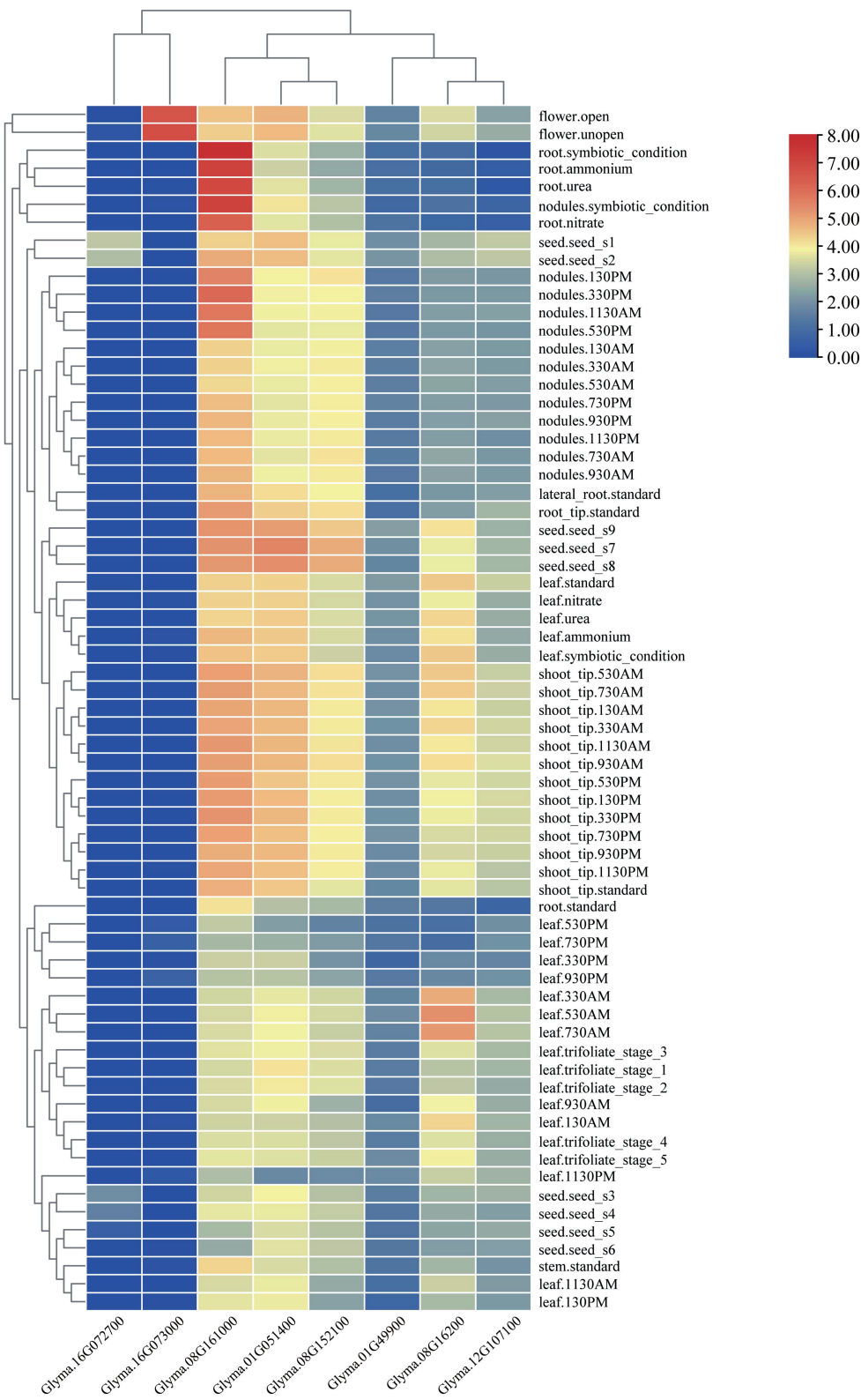
类别 Category	SNV 数量 Number of SNVs	类别 Category	Indel 数量 Number of Indels
downstream	757	downstream	101
Exonic_nonsynonymous_SNV	381	Exonic_frameshift_deletion	9
Exonic_stopgain_SNV	14	Exonic_frameshift_insertion	8
Exonic_stoploss_SNV	3	Exonic_nonframeshift_deletion	9
Exonic_synonymous_SNV	290	Exonic_nonframeshift_insertion	3
intergenic	8079	Exonic_stopgain_SNV	1
intronic	1935	intergenic	553
splicing	4	intronic	244
upstream	671	splicing	2
Upstream;downstream	71	upstream	98
3'-UTR	203	Upstream;downstream	6
5'-UTR	163	3'-UTR	25
共计 Total	12571	5'-UTR	27
		共计 Total	1086

表 6 ESDRs 内 SNV/InDel 突变的候选基因

Table 6 Candidate genes for SNV/InDel mutations in ESDRs

序号 No.	候选基因 Candidate gene	染色体 Chromosome	起始位置 Start	终止位置 End	SNV/InDel 数量 Number of SNV/InDel	SNV 类别 SNV category	InDel 类别 InDel category	同源基因 Homologs	文献 Reference
1	<i>Glyma. 01G049900</i>	Chr01	5851509	5853133	3/1	上游(2);下游(1)	上游(1)	<i>EMB93_SOLDATI0</i>	[20]
2	<i>Glyma. 01G051400</i>	Chr01	6073560	6083832	42/11	3'-UTR(1);上游(12);内含子(16);外显子同义 SNV(4);外显子非同义 SNV(5);下游(5)	上游(1);3'-UTR(1);外显子非移码缺失(1);内含子(5);下游(3)	<i>emb1579</i>	[21]
3	<i>Glyma. 08G152100</i>	Chr08	11762747	11770527	36/2	3'-UTR(1);内含子(20);外显子同义 SNV(1);外显子非同义 SNV(1);下游(13)	内含子(1);下游 3'-UTR(1)	<i>GFA2</i>	[22]
4	<i>Glyma. 08G161000</i>	Chr08	12544043	12551016	8/0	上游(1);内含子(5);下游(1);5'-UTR(1)		<i>emb2742</i>	[23]
5	<i>Glyma. 08G162000</i>	Chr08	12672098	12674964	2/2	上游(1);外显子同义 SNV(1)	5'-UTR(1);上游(1)	<i>EMB1586_ISE1</i>	[24]
6	<i>Glyma. 12G107100</i>	Chr12	9801389	9804042	9/1	上游(5);内含子(1);外显子非同义 SNV(1);下游(2)	上游(1)	<i>emb1923</i>	[25]
7	<i>Glyma. 16G072700</i>	Chr16	7275757	7279176	9/3	3'-UTR(1);内含子(5);外显子非同义 SNV(2);下游(1)	下游(1);外显子非移码缺失(1);剪切部位(1)	<i>AtMYB98、MYB98</i>	[26]
8	<i>Glyma. 16G073000</i>	Chr16	7303752	7307985	10/0	上游(5);外显子同义 SNV(2);下游(3)		<i>ATMYB21、ATMYB3、MYB21</i>	[27]

注:“( )”内数字表示对应 SNV 和 InDel 类别的数量。  
Note: Number in brackets represents the number of corresponding SNV and InDel categories.



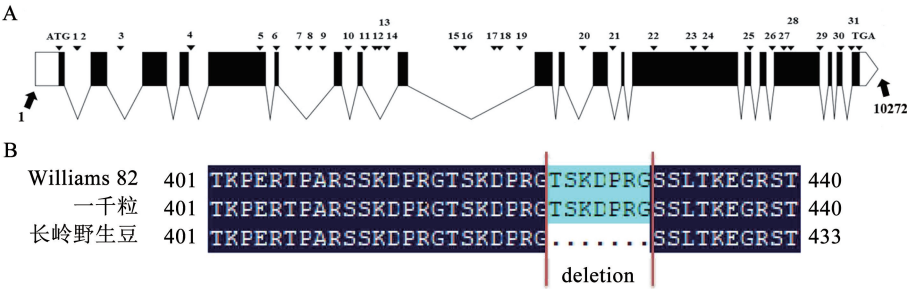
注:FPKM 值取 lg2。

Note:Normalized FPKM values are depicted on the lg2 scale.

图 2 ESDRs 内候选基因在不同组织和发育阶段相对表达的可视化热图  
Fig.2 Heatmap visualization of relative expressions in different tissues and developmental stages, for candidate genes in ESDRs

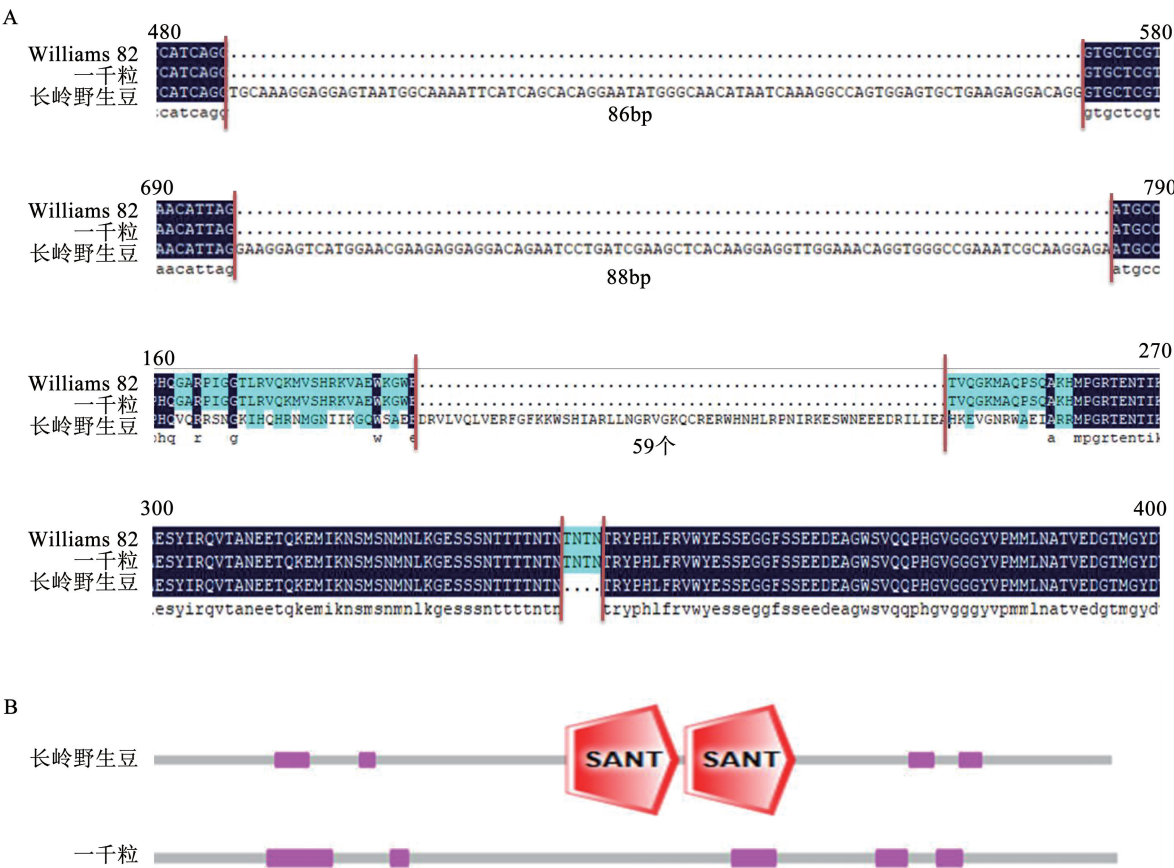
表 7 *Glyma. 01G051400* 的 SNV/InDel 汇总  
Table 7 Summary of SNV/InDel of *Glyma. 01G051400*

序号 No.	物理位置 Position/bp	Williams 82	一千粒 Yiqianli	长岭野生豆 Changling wild soybean
1	504	C	C	T
2	516	T	C	T
3	1038	T	T	C
4	1897	A	A	T
5	2739	GGATCCCCGTGCTACTTCAAAA	GGATCCCCGTGCTACTTCAAAA	G
6	2939	T	T	G
7	3206	C	C	A
8	3339	T	T	C
9	3504	A	G	A
10	3817	AT	AT	A
11	4010	A	A	T
12	4143	TC	TC	T
13	4191	G	G	A
14	4286	G	G	A
15	5138	CCTA	CCTA	C
16	5215	AT	AT	A
17	5594	C	C	CAC TT
18	5663	C	C	T
19	5905	G	G	C
20	6674	A	G	A
21	7039	G	G	T
22	7540	G	G	T
23	8023	A	A	G
24	8166	A	G	A
25	8713	A	A	G
26	8984	T	A	T
27	9125	A	T	A
28	9210	A	A	G
29	9565	A	A	G
30	9820	T	T	C
31	9949	T	T	A



注: A. *Glyma. 01G051400* 的 SNV/InDel 的物理位置; B. Williams 82、“一千粒”和“长岭野生豆”在 423 ~ 430 个氨基酸区间内的差异。图上黑框为外显子,折线为内含子,左侧白色方块为 5'-UTR,右侧白色三角为 3'-UTR,黑色倒三角表示此位置发生突变,红线内是其缺失位置。  
Note: A. Physical location of SNV/InDel of *Glyma. 01G051400*; B. The differences among Williams 82, Yiqianli, and Changling wild soybean in 423-430 amino acid region. The black box represents the exon, the line represents the intron, the white square on the left represents 5'-UTR, and the white triangle on the right represents 3'-UTR. The black inverted triangle indicates a mutation at this location. Deletion occurs between red lines.

图 3 *Glyma. 01G051400* 在 3 个品种中的基因序列比对结果  
Fig. 3 Comparative results of gene sequences of *Glyma. 01G051400* in three varieties



注:A. 在 mRNA 和氨基酸上的缺失位置;B. 蛋白质结构示意图。  
Note: A. Deletion on mRNA and amino acid; B. Schematic diagram of protein structure.

图 4 *Glyma. 16G072700* 在 3 个品种的基因序列比对结果  
Fig. 4 Comparative results of gene sequences of *Glyma. 16G072700* in three varieties

3 讨论

本研究利用野生大豆“长岭野生豆”与地方品种“一千粒”构建的 200 个株系 F<sub>6</sub> 代 RIL 群体,采用 SLAF-seq 测序分析技术,获得包含 4 564 个 SNP 标记的高密度遗传图谱。亲本基因型在群体中的分离比为 0.98:1,接近 1:1 的理论比例,双亲的遗传贡献率基本一致。但由于偏分离区间的存在,不同的染色体之间存在不同程度的偏分离。标记偏分离分析发现,648 个 SNP 位点表现出偏分离,占总标记的 14.20%。高于 Hisano 等<sup>[28]</sup> 通过 STR 和 SSR 标记检测 1 个由 94 个株系组成的 F<sub>8</sub> 栽培种重组自交系群体的 5.1% ( $P<0.05$ );同时高于 Song 等<sup>[29]</sup> 通过 SNP 位点检测 5 600 个 F<sub>5</sub> 单株,其来自 40 个栽培大豆品种构建的多个 RIL 群体,总偏分离为 3.75% ( $P<0.01$ );高于 Liu 等<sup>[30]</sup> 通过 282 SSR 标记检测 1 个由 97 个株系组成的 F<sub>8</sub> 野生种与栽培种大豆

RIL 群体的 9.6% ( $P<0.05$ );高于 Oki 等<sup>[31]</sup> 通过 163 个 SNP 和 236 个 SSR 标记检测 1 个由 202 个株系组成的 F<sub>7</sub> 野生种与栽培种大豆 RIL 群体的 9.3% ( $P<0.05$ );低于吴晓雷等<sup>[14]</sup> 利用 2 个栽培品种通过 4 种分子标记方式构建出 24 个连锁群、792 个遗传标记组成的大豆高密度连锁图谱,偏分离率为 21.73% ( $P<0.05$ );先前的报道中,在半野生大豆与栽培种大豆所构建的分离群体中也检测到了较高的偏分离现象<sup>[15]</sup>。

在检测到的 SDR 中,12 个偏向野生种父本,遵循父本基因引入规律<sup>[32]</sup>,但依然有 5 个 SDR 区间偏向母本。在 17 个偏分离 SDR 中进一步获得了重度偏分离 ( $P<0.01$ ) 的区间 8 个。值得注意的是,多数在同一条染色体上的 SDR 都偏向同一方向,但 8 号染色体上的 SDR8-4 偏向母本,区别于该染色体上其他 SDR 区间偏向方向。这一结果进一步说明了大豆偏分离现象产生原因的复杂性。与先前研

究对比,如 Zhang 等<sup>[33]</sup>在 2013 年两个 F<sub>2</sub> 群体;Oki 等<sup>[31]</sup>2017 年 202 株 RILs(F<sub>7</sub>) 群体;Hisano 等<sup>[28]</sup>在 2007 年的 98 株 RILs(F<sub>8</sub>)栽培与半野生杂交后代群体;Ortega 等<sup>[34]</sup>2011 年中 *Rps8/rps8* 偏分离区间等相比,本研究并未发现 SDR 区域彼此之间有重叠区间或临近位点。这一结果可能由于可参考的文献目前相对较少,并且在其他作物上不同群体偏分离区间也较少发生重叠。本研究获得的 SDR 还需要通过构建特定偏分离区间的回交群体,进一步加以确认。

本研究发现,编码胚胎缺陷基因 *emb1579* (*Glyma.01G051400*)是 ESDR1-1 区间内的候选基因。基于大豆 G. max Wm82. a4. v1 比对“Williams82”“一千粒”和“长岭野生豆”基因组序列差异,发现野生大豆相较于两个栽培品种编码的蛋白序列存在 7 个氨基酸的缺失(图 3B)。已有研究证实 *emb1579* 基因缺失影响胚胎中的细胞扩增和扭曲,造成拟南芥胚胎发育缺陷,还会造成植物开花延迟<sup>[21]</sup>。该氨基酸序列的差异可能造成胚胎发育的差异,导致不同等位基因植株间后代存活数量不同,最终导致偏分离现象的出现。在 ESDR16-1 区间内的 *Glyma.16G072700* 编码 MYB98 转录因子,该基因在雌性配子体发育过程中协同控制细胞内特定特征的发育,影响花粉管引导和助细胞分化<sup>[26]</sup>。“长岭野生豆”中该基因的核苷酸序列相较于两个栽培型大豆品种存在 mRNA 剪切位点差异和 1 处非移码片段缺失(图 4),特别是野生型相较于栽培型在该基因所编码蛋白中增加了两个 SANT 蛋白质结构域<sup>[35]</sup>。该蛋白质结构域的改变可能造成不同等位基因个体间花粉管发育差异,导致偏分离现象的发生。尽管如此,依然不能排除偏分离所在区域其它基因作为候选基因的可能性。因此,需要进一步通过构建回交群体,缩小候选区间,从而更准确地确定候选基因。

### 4 结论

相较之前低密度分子标记对于偏分离 SDR 的研究,本研究借助高密度遗传图谱的构建,更精准地筛选出栽培与野生大豆 RIL 群体 SDR 及 ESDR 区间,同时挖掘了可能的候选基因。该研究为今后偏分离基因定位提供了可靠依据,也有助于解析偏分离在大豆驯化过程中的重要作用及其产生机制。

### 参考文献

[1] LYTTLE T W. Segregation distorter [J]. Annual Review of Genetics, 1991, 25(1): 511-557.

[2] TAYLOR D R, INGVARSSON P K. Common features of segregation distortion in plants and animals [J]. An International Journal of Genetics and Evolution, 2003, 117(1): 27-35.

[3] SIBOV S T, LOPES DE SOUZA J R C, GARCIA A A F, et al. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation [J]. Hereditas, 2003, 139(2): 96-106.

[4] MATSUSHITA S, ISEKI T, FUKUTA Y, et al. Characterization of segregation distortion on chromosome 3 induced in wide hybridization between indica and japonica type rice varieties [J]. Euphytica, 2003, 134(1): 27-32.

[5] WANG G, HE Q Q, XU Z K, et al. High segregation distortion in maize B73 × teosinte crosses [J]. Genetics and Molecular Research, 2012, 11(1): 693-706.

[6] 王哲. 植物杂交后代中基因偏分离的产生原因及其进化意义 [J]. 遗传, 2016, 38(9): 801-810. (WANG Z. Distorted segregation in plant hybrids and its implication for evolution [J]. Hereditas, 2016, 38(9): 801-810.)

[7] XU Y, ZHU L, XIAO J, et al. Chromosomal regions associated with segregation distortion of molecular markers in F<sub>2</sub>, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.) [J]. Molecular Genetics and Genomics, 1997, 253(5): 535-545.

[8] 严建兵, 汤华, 黄益勤, 等. 玉米 F<sub>2</sub> 群体分子标记偏分离的遗传分析 [J]. 遗传学报, 2003, 30(10): 913-918. (YAN J B, TANG H, HUANG Y Q, et al. Genetic analysis of segregation distortion of molecular markers in maize F<sub>2</sub> population [J]. Journal of Genetics and Genomics, 2003, 30(10): 913-918.)

[9] 张明科, 张鲁刚, 巩振辉, 等. 大白菜分子连锁图谱的构建与分析 [J]. 分子植物育种, 2009, 7(1): 82-88. (ZHANG M K, ZHANG L G, GONG Z H, et al. Construction and analysis of a molecular genetic map of Chinese cabbage [J]. Molecular Plant Breeding, 2009, 7(1): 82-88.)

[10] WANG X, CHEN L, MA J. Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication [J]. Genome Biology, 2019, 20(1): 1-15.

[11] KOPISCH-OBUCH F J, DIERS B W. Segregation at the SCN resistance locus *rhg1* in soybean is distorted by an association between the resistance allele and reduced field emergence [J]. Theoretical and Applied Genetics, 2006, 112: 199-207.

[12] BAUMBACH J, ROGERS J P, SLATTERY R A, et al. Segregation distortion in a region containing a male-sterility, female-sterility locus in soybean [J]. Plant Science, 2012, 195: 151-156.

[13] 刘峰, 吴晓雷, 陈受宜. 大豆分子标记在 RIL 群体中的偏分离

分析[J]. 遗传学报, 2000(10): 883-887. (LIU F, WU X L, CHEN S Y. Segregation distortion of molecular markers in recombinant inbred populations in soybean (*G. max*) [J]. Journal of Genetics and Genomics, 2000(10): 883-887. )

[14] 吴晓雷, 贺超英, 王永军, 等. 大豆遗传图谱的构建和分析[J]. 遗传学报, 2001(11): 1051-1061. (WU X L, HE C Y, WANG Y J, et al. Construction and analysis of a genetic linkage map of soybean [J]. Journal of Genetics and Genomics, 2001(11): 1051-1061. )

[15] 张德水, 陈受宜, 惠东威, 等. 栽培大豆与半野生大豆杂种 F<sub>2</sub> 群体中 RFLP 标记的偏分离及其形成原因的分析[J]. 遗传学报, 1997(4): 362-367. (ZHANG D S, CHEN S Y, HUI D W, et al. Segregation distortion of RFLP markers in F<sub>2</sub> population of cultivated/semi-wild soybean and the causal analysis[J]. Journal of Genetics and Genomics, 1997(4): 362-367. )

[16] HAMWIEH A, XU D. Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans [J]. Breeding Science, 2008, 58(4): 355-359.

[17] SUN X, LIU D, ZHANG X, et al. SLAF-seq: An efficient method of large-scale *De Novo* SNP discovery and genotyping using high-throughput sequencing[J]. PLoS One, 2013, 8(3): e58700.

[18] 付忠军, 官玲, 张丕辉, 等. 玉米 RIL 群体 SNP 偏分离分析[J]. 农业生物技术学报, 2019, 27(9): 1542-1549. (FU Z J, GUAN L, ZHANG P H, et al. Segregation distortion analysis on maize (*Zea mays*) RIL population by SNP [J]. Journal of Agricultural Biotechnology, 2019, 27(9): 1542-1549. )

[19] CHEN C, CHEN H, ZHANG Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data[J]. Molecular Plant, 2020, 13(8): 1194-1202.

[20] ERRAMPALLI D, PATTON D, CASTLE L, et al. Embryonic lethals and T-DNA insertional mutagenesis in *Arabidopsis* [J]. The Plant Cell, 1991, 3(2): 149-157.

[21] ZHANG Y, LI Z, CHEN N, et al. Phase separation of *Arabidopsis* EMB1579 controls transcription, mRNA splicing, and development[J]. PLoS Biology, 2020, 18(7): e3000782.

[22] LU H, ROMERO-SEVERSON J, BERNARDO R. Chromosomal regions associated with segregation distortion in maize [J]. Theoretical and Applied Genetics, 2002, 105: 622-628.

[23] HICKL D, SCHEURING D, MÖHLMANN T. CTP Synthase 2 from *Arabidopsis thaliana* is required for complete embryo development[J]. Frontiers in Plant Science, 2021, 12: 652434.

[24] KIM I, HEMPEL F D, SHA K, et al. Identification of a developmental transition in plasmodesmatal function during embryogenesis in *Arabidopsis thaliana* [J]. Development, 2002, 129: 1261-1272.

[25] MEINKE D W. Genome-wide identification of *EMBRYODEFECTIVE* (*EMB*) genes required for growth and development in *Arabidopsis* [J]. New Phytologist, 2020, 226(2): 306-325.

[26] KASAHARA R D, PORTEREIKO M F, SANDAKLIE-NIKOLOVA L, et al. *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis* [J]. The Plant Cell, 2005, 17(11): 2981-2992.

[27] MANDAOKAR A, THINES B, SHIN B, et al. Transcriptional regulators of stamen development in *Arabidopsis* identified by transcriptional profiling [J]. The Plant Journal, 2006, 46(6): 984-1008.

[28] HISANO H, SATO S, ISOBE S, et al. Characterization of the soybean genome using EST-derived microsatellite markers [J]. DNA Research, 2007, 14(6): 271-281.

[29] SONG Q J, YAN L, QUIGLEY C, et al. Genetic characterization of the soybean nested association mapping population [J]. The Plant Genome, 2017, 10(2): 1-14.

[30] LIU B, FUJITA T, YAN Z H, et al. QTL mapping of domestication-related traits in soybean (*Glycine max*) [J]. Annals of Botany, 2007, 100(5): 1027-1038.

[31] OKI N, KAGA A, SHIMIZU T, et al. QTL mapping of antixenosis resistance to common cutworm (*Spodoptera litura* Fabricius) in wild soybean (*Glycine soja*) [J]. PLoS One, 2017, 12(12): e0189440.

[32] 袁洋, 敖和军, 周仲华, 等. 水稻杂种偏分离位点 *SegD8* 的精细定位[J]. 中国水稻科学, 2023, 37(1): 37-44. (YUAN Y, AO H J, ZHOU Z H, et al. Fine-mapping of *SegD8* loci for rice hybrid segregation distortion [J]. Chinese Journal of Rice Science, 2023, 37(1): 37-44. )

[33] ZHANG Y, ZHAO Q, YANG C, et al. Recombination frequencies between cultivated soybean (*Glycine max*) and its wild relative *Glycine soja* based on molecular marker analysis [J]. African Journal of Biotechnology, 2013, 12(22): 3522-3527.

[34] ORTEGA M A, DORRANCE A E. Microsporogenesis of *Rps 8/tps 8* heterozygous soybean lines [J]. Euphytica, 2011, 181: 77-88.

[35] GRÜNE T, BRZESKI J, EBERHARTER A, et al. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI [J]. Molecular Cell, 2003, 12(2): 449-460.