

大豆基因组 *GmSRK* 及其同源基因的生物信息学分析

钱 雪,孙晓丽,端木慧子,成舒飞,才 华,纪 巍,赵超越,朱延明

(东北农业大学 农业生物功能基因重点实验室,黑龙江 哈尔滨 150030)

摘 要:以野生大豆 *GsSRK* 氨基酸序列为检索序列,在 Phytozome 数据库中查找大豆基因组中其同源基因 *GmSRK*。利用 Phytozome 数据库以及 MEGA5.0,ClustalX 等软件挖掘大豆基因组 *GmSRK* 同源基因,最终确定同源性较高的 2 个基因 *Glyma12g21640* 和 *Glyma08g06550*。对 *GmSRK*、*Glyma12g21640* 和 *Glyma08g06550* 这 3 个基因进行基因结构,染色体定位,蛋白理化性质、一级、二级结构分析,结果表明三者均含有 6 个内含子,7 个外显子;定位于 3 条不同染色体上;都具有 B-lectin, S_locus_glycop, PAN_AP 以及 S_TKc 四个结构域,属于 G-type 外源凝集素类受体蛋白激酶。进化树分析发现 *GmSRK* 与 *Glyma12g21640* 亲缘关系最近且与四季豆、苜蓿中同源蛋白聚为一支, *Glyma08g06550* 与拟南芥 RLKs 亲缘关系较近。

关键词:大豆; *GmSRK*; 类受体蛋白激酶; 生物信息学

中图分类号: S565.1

文献标识码: A

DOI: 10.11861/j.issn.1000-9841.2014.04.0497

Bioinformatics Analysis of *GmSRK* Homologous Genes in Soybean Genome

QIAN Xue, SUN Xiao-li, DUANMU Hui-zi, CHENG Shu-fei, CAI Hua, JI Wei, ZHAO Chao-yue, ZHU Yan-ming

(Key Laboratory of Agricultural Biological Functional Genes, Northeast Agricultural University, Harbin 150030, China)

Abstract: We identified the *Glycine max GmSRK* gene by using BLASTP with the amino acid sequence of *Glycine soja GsSRK* gene. By using the phytozome database and some softwares such as MEGA5.0 and ClustalX, we confirmed the *GmSRK* homologous genes, *Glyma12g21640* and *Glyma08g06550*. Bioinformatics analysis indicated that the three genes were located in different chromosome, all of them had 6 introns and 7 exons, and belonged to the G-type lectin receptor-like kinase family, which was characterized by four conserved domains: B-lectin, S_locus_glycop, PAN_AP and S_TKc. Analysis of phylogenetic relationship showed that *GmSRK* and *Glyma12g21640* had the closest relationship, and grouped into the same cluster with homologous proteins of *Phaseolus vulgaris*, *Medicago truncatula*. *Glyma08g06550* had close relationship with *Arabidopsis thaliana* RLKs.

Key words: Soybean; SRK; Receptor-like kinases; Bioinformatics

大豆 (*Glycine max*) 是一种重要的粮食作物和经济作物。随着生物信息学的不断发展以及大豆基因组测序^[1]的完成和序列释放,人们可以在基因组水平上对大豆中编码重要功能的基因进行挖掘、比较和功能推断,为功能基因组学以及大豆品种改良奠定坚实的理论基础。

植物长期暴露于外界自然环境中,经常会受到各种生物和非生物环境因子的影响。植物细胞具有感受刺激的能力,并能通过细胞膜进行跨膜转导,最终调整其转录组与蛋白质组,从而对胞外信号做出相应的生理生化反应^[2]。在这些信号转导途径中,植物类受体蛋白激酶 (receptor-like kinases, RLKs) 担当着一个极为重要的角色。与动物体中的受体蛋白激酶结构相似^[3],植物类受体蛋白激酶也由 3 部分构成,即胞外受体结合区、跨膜区和胞内激酶区^[4]。根据胞外区的多样性,RLKs 可以分为不

同的亚家族,其中植物凝集素类受体蛋白激酶 (lectin-receptor-like kinases, LecRLKs) 是植物 RLKs 中的一大类亚家族^[5],在植物生理反应过程中起重要作用,如 *At3g53810* 编码的 SGC LecRLK 蛋白与植株花粉正常发育有关^[6]; *pid-2* 基因水稻超表达植株获得了稻瘟病抗性^[7]; *PsLecRLK* 基因是一种盐胁迫抗性基因^[8]; *MtLecRK1*; 1, *MtLecRK7*; 1, *MtLecRK7*; 2, *MtLecRK7*; 3 与蒺藜苜蓿共生固氮有关^[9],但是仍有许多基因功能未知,需要不断探索。

本研究根据实验室前期 Sun 等^[10]克隆得到的具有耐盐特性的野生大豆 (*Glycine soja*) 外源凝集素类受体蛋白激酶基因 *GsSRK*,对大豆基因组中其同源基因进行挖掘,运用生物信息学方法进行分析,以期进一步揭示此类基因的结构与功能,为后续研究提供理论依据。

收稿日期: 2014-01-06

基金项目: 转基因生物新品种培育重大专项 (2011ZX08004-002); 国家自然科学基金 (31171578); 黑龙江省高校科技创新团队建设计划 (2011TD055)。

第一作者简介: 钱雪 (1989-), 女, 硕士, 主要从事植物基因工程与分子生物学研究。E-mail: snow_qx@163.com。

通讯作者: 朱延明 (1955-), 男, 教授, 博士生导师, 从事植物基因工程与分子生物学研究。E-mail: ymzhu2001@neau.edu.cn。

1 材料与方法

1.1 序列来源

大豆全基因数据,拟南芥、苜蓿、四季豆、水稻、玉米、可可树、毛果杨、甜橙 *SRK* 同源蛋白的氨基酸序列均来源于 Phytozome 数据库 (<http://www.phytozome.net>),序列号见表 1。

1.2 *GmSRK* 及其同源基因的查询与确定

以实验室前期获得的野生大豆基因 *GsSRK* 氨基酸序列为检索序列在大豆数据库 (Phytozome) 中进行 BLASTP 筛选同源基因 *GmSRK*,根据序列同源性和结构域进行进一步筛选,选取同源性较高的基因进行生物信息学分析。

1.3 *GmSRK* 及其同源基因结构分析和染色体定位

利用 GSDS^[11] 在线工具 (<http://gsds.cbi.pku.edu.cn/>) 制作基因外显子-内含子结构示意图。通过 MapInspect 程序显示基因在染色体上的位置。

1.4 *GmSRK* 蛋白的理化性质、一级、二级结构分析

利用 ProtParam (<http://expasy.org/tools/prot-param.html>) 在线分析氨基酸理化性质;使用在线工具 SignalIP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) 进行信号肽的预测;利用 TMHMM (<http://www.ebs.dtu.dk/services/TMHMM/>) 进行跨膜结构域预测;利用 (<http://www.expasy.ch/egibin/protscal.pl>) 网站的 ProtScale 工具进行亲疏水性分析;采用 PSORT (<http://psort.ims.u-tokyo.ac.jp/>

[form.html](http://psort.ims.u-tokyo.ac.jp/form.html)) 软件进行亚细胞定位预测;利用在线的蛋白修饰位点预测程序 PROSCAN (<http://npsa-pbil.ibcp.fr/>) 进行蛋白修饰位点预测;利用 SMART (<http://smart.embl-heidelberg.de/>) 进行蛋白质功能结构域分析;利用 predator (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_predator.html) 进行蛋白质二级结构预测。

1.5 *GmSRK* 蛋白系统发生分析

利用 ClustalX 对大豆及其它物种 *SRK* 同源蛋白进行多重序列比对,运用 MEGA5.0 中的邻接法 (Neighbor-Joining, NJ) 构建系统发育树,并且用 Bootstrap (校正次数为 1 000) 进行校正,生成最终的系统进化树。

2 结果与分析

2.1 *GmSRK* 及其同源基因的查询与确定

以实验室前期获得的野生大豆外源凝集素类受体蛋白激酶基因 *GsSRK* 的氨基酸序列为检索序列,在大豆基因组数据库 Phytozome 中进行 BLASTP,得到其同源基因 *Glyma06g39930* (命名为 *GmSRK*),二者氨基酸序列同源性为 99.6%。运用 SMART 在线工具对 *GmSRK* 蛋白的保守结构域进行分析发现 (图 1),*GmSRK* 蛋白包括 B-lectin (33 ~ 149 aa), S_locus_glyco (187 ~ 288 aa), PAN_AP (318 ~ 384 aa) 3 个胞外结构域,跨膜结构域 (434 ~ 456 aa) 以及胞内蛋白激酶区域 S_TKc (513 ~ 787 aa),表明 *GmSRK* 为 G-型外源凝集素类受体蛋白激酶。

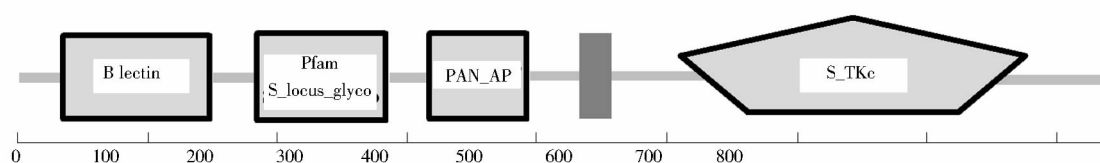


图 1 SMART 软件分析 *GmSRK* 蛋白结构

Fig. 1 Analysis of *GmSRK* protein structure by SMART

在 phytozome 大豆基因组注释数据库 (Phytozome V9.1) 中搜索到 51 个与 *GmSRK* 蛋白具有相同结构域的基因,SMART 分析发现,其中 9 个基因在 PAN_AP 和 S_TKc 结构域之间没有跨膜结构,因此将其剔除^[12],对其余 43 个基因进行聚类分析,发现 *Glyma12g21640*、*Glyma08g06550* 与 *GmSRK* 的亲缘关系最近,序列同源性分别达到 75.7% 和 44.0%。其余 40 个蛋白与 *GmSRK* 同源性较低 (均小于 42%),亲缘关系较远 (不在同一分支上)。这

与 Sun^[10] 等提出的 *GsSRK* 也许是植物 RLK 家族中的一类新成员一致,因此仅对相似性较高的 3 个蛋白即 *Glyma06g39930* (*GmSRK*)、*Glyma12g21640*、*Glyma08g06550* 进行生物信息学分析。

2.2 *GmSRK* 及其同源基因结构分析和染色体定位

利用 GSDS 在线工具对 *GmSRK* 及其同源基因的结构进行分析 (图 2),结果表明 3 个基因均含有 6 个内含子和 7 个外显子,具有一定的保守性。

表 1 不同物种中 SRK 同源基因
Table 1 SRK homologous genes in different species

Phytozome 序列登录号 Phytozome accession number	物种 Species
Glyma06g39930.1	大豆 <i>Glycine max</i>
Glyma12g21640.2	大豆 <i>Glycine max</i>
Glyma08g06550.1	大豆 <i>Glycine max</i>
AT1G11300.1	拟南芥 <i>Arabidopsis thaliana</i>
AT4G21380.1	拟南芥 <i>Arabidopsis thaliana</i>
AT1G65790.1	拟南芥 <i>Arabidopsis thaliana</i>
AT1G11340.1	拟南芥 <i>Arabidopsis thaliana</i>
AT1G65800.1	拟南芥 <i>Arabidopsis thaliana</i>
Medtr8g061110.1	苜蓿 <i>Medicago truncatula</i>
Phvul.011G149500.1	四季豆 <i>Phaseolus vulgaris</i>
GRMZM2G322728.T01	玉米 <i>Zea mays</i>
LOC_Os01g57480.1	水稻 <i>Oryza sativa</i>
Thecc1EG046868.t1	可可树 <i>Theobroma cacao</i>
Potri.005G014800.1	毛果杨 <i>Populus trichocarpa</i>
orange1.1g003846.m	甜橙 <i>Citrus sinensis</i>

图 2 大豆 *GmSRK* 及其同源基因内含子外显子结构

Fig.2 The gene structure of *GmSRK* homologous genes in soybean

根据 Phytozome 数据库中提供的基因与染色体相关位置的信息,通过 MapInspect 软件将每条基因显示在染色体的相应位置上, *GmSRK*、*Glyma08g06550*、*Glyma12g21640* 分别位于第 6、8、12 号染色体上(图 3)。

2.3 *GmSRK* 蛋白的理化性质分析

野生大豆 *GsSRK* 在大豆中的同源基因 *GmSRK* 编码 834 个氨基酸,分子量为 93.35 kDa,理论等电点 pI 为 5.85;总共包括 13 008 个原子,分子式为 $C_{4132}H_{6448}N_{1130}O_{1259}S_{39}$;在组成 *GmSRK* 蛋白的 20 种氨基酸中,亮氨酸(Leu)所占的比例最高,为 11.4%,组氨酸(His)所占的比例最低,为 1.4%;不稳定指数为 39.73,说明此蛋白为稳定蛋白;脂溶指数为 87.10;总平均疏水性值(GRAVY)为 -0.220。大豆中其另外 2 个同源性较高的蛋白理化性质信息如表 2 所示。由表 2 可知,随着氨基酸数目的增加,

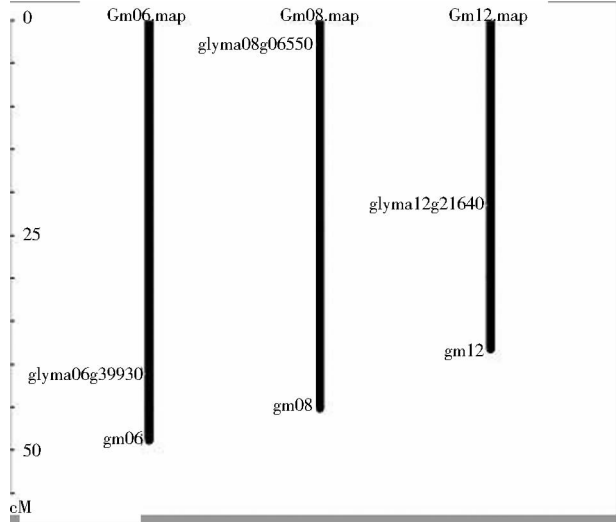


图 3 大豆 *GmSRK* 及其同源基因的染色体定位
Fig.3 The location of *GmSRK* homologous genes in soybean chromosomes

SRK 蛋白的相对分子质量也增加。3 个同源蛋白的理化性质具有一定的相似性:理论等电点均小于

7.0,都是酸性氨基酸^[13];GRAVY 均为负值;脂溶指数均小于 100。

表 2 GmSRK 及其同源蛋白理化性质信息

Table 2 Physicochemical properties of GmSRK homologous proteins

基因序列号 Gene ID	氨基酸数目 No. of aa	等电点 pI	相对分子质量 Molecular weight/kDa	脂溶指数 Aliphatic index	不稳定指数 Instability index	总平均疏水值 GRAVY
Glyma06g39930	834	5.85	93.35	87.10	39.73	-0.220
Glyma08g06550	838	6.92	94.61	79.75	38.30	-0.268
Glyma12g21640	845	6.63	95.38	87.23	41.84	-0.253

2.4 GmSRK 蛋白的一级、二级结构分析

GmSRK 蛋白 N 端含有信号肽序列,剪切位点位于第 23~24 位氨基酸,表明成熟肽始于第 23 位氨基酸。TMHMM2.0 预测显示,该蛋白在 434~456 位氨基酸之间形成一个跨膜区。亲疏水性预测发现在 430 位氨基酸附近有一个明显的疏水区域,这正是前面预测的跨膜结构所在区域。亚细胞定位分析发现该蛋白定位于质膜上。蛋白合成之后一般需要经过信号肽的剪切和翻译后修饰才能成为成熟的蛋白质,使用 PROSCAN 进行蛋白质修饰位点预测发现该蛋白含有 16 个糖基化位点,2 个 cAMP 和 cGMP 依赖的蛋白激酶磷酸化位点,18 个蛋白激酶 C 磷酸化位点,12 个酪蛋白激酶 II 磷酸化位点,11 个豆蔻酰化位点,1 个酰胺化位点,1 个丝苏氨酸蛋白磷酸化活性位点。二级结构分析表明

该蛋白由 α -螺旋、 β -折叠和无规则卷曲组成,主要组成部分为无规则卷曲,约占 60.43%。

2.5 GmSRK 蛋白系统进化树分析

根据 GmSRK 的氨基酸序列在 Phytozome 数据库中检索其他物种中的同源蛋白,获得 9 个物种的 15 条氨基酸序列(表 1),利用 MEGA5.0 构建进化树(图 4)。由图 4 可知,大豆中 GmSRK (Glyma06g39930)与 Glyma12g21640 亲缘关系最近,与四季豆、苜蓿、可可树、甜橙、毛果杨和水稻中的同源蛋白聚为一支;Glyma08g06550 与玉米、拟南芥的 RLKs 聚为一支,这与氨基酸同源性比对结果一致;GmSRK 与 Glyma12g21640 同源性较高,与 Glyma08g06550 同源性较低。GmSRK 与拟南芥中的所有 RLKs 亲缘关系都比较远。

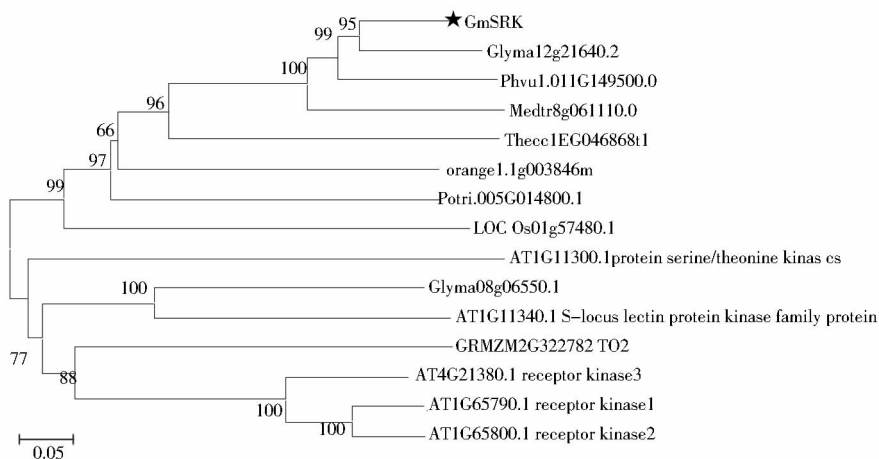


图 4 系统进化树分析

Fig. 4 Phylogenetic analysis

3 讨论

RLKs 是植物中最大的一类受体蛋白,其中 Lec-RLKs 是 RLKs 的一类亚家族。LecRLKs 根据其胞外结构域,可以分为 G-type、C-type 和 L-type 三类^[14]。G-type 又是 SRK (S-locus receptor kinase) 的一类^[15],已有研究证明 SRK 与自交不亲和性有

关^[16]。Sun 等^[10]克隆获得野生大豆 G-型外源凝集素类受体蛋白激酶基因 *GsSRK*,并将其在拟南芥中超表达进行功能研究,发现超表达和野生型植株在正常条件下无任何差异,当受到盐胁迫时,超表达植株在不同的生长阶段都比野生型更加耐盐,这说明 *GsSRK* 是一个在植物耐盐性方面起重要作用的蛋白激酶^[10]。

本研究采用生物信息学手段对大豆基因组中

GsSRK 的同源基因进行挖掘,确定其同源基因 *Glyma06g39930* (*GmSRK*), SMART 分析发现其与 *GsSRK* 具有相同的结构域,推测二者功能相似,可能在植物耐盐方面起作用。然后根据同源比对和保守结构域分析,检索到 51 个与 *GmSRK* 同源的大豆基因,其中 9 个不含有跨膜结构,对其余的 43 个基因进行初步的聚类分析发现, *GmSRK* 只与 *Glyma12g21640*、*Glyma08g06550* 基因亲缘关系较近,同源性较高,与大部分基因亲缘关系较远,同源性较低,推测原因是大豆基因组在进化过程中至少发生了两次复制,引起了整个基因组的高度重复,两次复制后紧接着出现了基因多样化和基因丢失,大量的染色体发生重排^[1]。对同源性较高的 3 个基因进行一系列的生物信息学分析,发现三者定位于 3 条不同的染色体上,具有相同的内含子外显子数目,部分理化性质具有一定的相似性。系统进化树分析发现,整个进化树分为两大支,每一支中都含有单子叶和双子叶植物,由此可推测此类基因在单子叶植物分化之前就已经存在。其中大豆中 3 个基因的进化关系并不是最近的,说明在进化过程中基因的功能可能发生了分化。*GmSRK* (*Glyma06g39930*) 与 *Glyma12g21640* 聚为一支,与豆科植物四季豆、苜蓿的同源蛋白关系较近,与拟南芥的各类 RLKs 关系较远,说明 *GmSRK* 可能与拟南芥 RLKs 执行不同的功能,是植物 RLK 家族中具有新功能的一类基因。

大豆 *Glyma06g39930* (*GmSRK*) 与耐盐相关基因 *GsSRK* 同源性高达 99.6%, 推测 *GmSRK* 也是耐盐相关基因,但是其具体功能以及机制并不清楚,需要相关实验进行进一步验证研究。

参考文献

- [1] Jeremy S, Steven B C, Jessica S, et al. Genome sequence of the palaeopolyploid soybean[J]. *Nature*, 2010, 463: 178-183.
- [2] 闫锋, 祝传书, 庞保平. 植物类受体蛋白激酶的研究进展[J]. *西北植物学报*, 2009, 29(4): 851-858. (Yan F, Zhu C S, Pang B P. Advance in research of plant receptor-like protein kinases[J]. *Acta Botanica Boreale-Occidentalia Sinica*, 2009, 29(4): 851-858.)
- [3] Stevan R H, Miller W T. Receptor tyrosine kinases: mechanisms of activation and signaling[J]. *Current Opinion in Cell Biology*, 2007, 19(2): 117-123.
- [4] 张兆沛, 王志伟, 张惠蓉. 植物类受体蛋白激酶研究概况[J]. *山西农业科学*, 2009, 37(8): 75-78. (Zhang Z P, Wang Z W, Zhang H R. Research advance in receptor-like protein kinases of plant[J]. *Journal of Shanxi Agricultural Sciences*, 2009, 37(8): 75-78.)
- [5] Herve C, Dabos P, Galaud J P, et al. Characterization of an *Arabidopsis thaliana* gene that defines a new class of putative plant receptor kinases with an extracellular lectin-like domain[J]. *Journal of Molecular Biology*, 1996, 258(5): 778-788.
- [6] Wan J R, Ami P, Melanie M, et al. A lectin receptor-like kinase is required for pollen development in *Arabidopsis*[J]. *Plant Molecular Biology*, 2008, 67(5): 469-482.
- [7] Chen X W, Li S G, Xu J C, et al. Identification of two blast resistance genes in a rice variety, Digu[J]. *Journal of Phytopathology*, 2004, 152(2): 77-85.
- [8] Amita J, Hung Q D, Neha V, et al. Pea lectin receptor-like kinase promotes high salinity stress tolerance in bacteria and expresses in response to stress in planta[J]. *Glycoconjugate Journal*, 2009, 27(1): 133-150.
- [9] Navarro-Gochicoa M T, Camut S, Timmers A C, et al. Characterization of four lectin-like receptor kinases expressed in roots of *Medicago truncatula*. Structure, location, regulation of expression, and potential role in the symbiosis with *Sinorhizobium meliloti*[J]. *Plant Physiology*, 2003, 133(4): 1893-1910.
- [10] Sun X L, Yu Q Y, Tang L L. *GsSRK*, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress[J]. *Journal of Plant Physiology*, 2013, 170(5): 505-515.
- [11] 郭安源, 朱其慧, 陈新, 等. GSDS: 基因结构显示系统[J]. 遗传, 2007, 29(8): 1023-1026. (Guan A Y, Zhu Q H, Chen X, et al. GSDS: a gene structure display server[J]. *Hereditas*, 2007, 29(8): 1023-1026.)
- [12] 邓克勤. 植物凝集素类受体蛋白激酶 *LecRK-b2* 的生物信息学分析及基因功能研究[D]. 长沙: 湖南大学, 2009. (Deng K Q. Bioinformatic analysis and function investigation of *Arabidopsis* lectin receptor like kinase *lecrk-b2*[D]. Changsha: Hunan University, 2009.)
- [13] 张春华, 上官凌飞, 俞明亮, 等. 桃 NAC 基因家族生物信息学分析[J]. *江苏农业学报*, 2012, 28(2): 406-414. (Zhang C H, Shanguan L F, Yu M L, et al. Bioinformatics analysis of NAC gene family in peach[J]. *Jiangsu Journal of Agricultural Sciences*, 2012, 28(2): 406-414.)
- [14] Bouwmeester K, Govers F. *Arabidopsis* L-type lectin receptor kinases: phylogeny, classification, and expression profiles[J]. *Journal of Experimental Botany*, 2009, 60(15): 4383-4396.
- [15] 王育, 叶佳卓, 常洪平. 植物凝集素类受体激酶的研究进展[J]. *生命科学研究*, 2012, 16(3): 266-271. (Wang Y, Ye J Z, Chang H P. Progresses on lectin receptor like kinase[J]. *Life Science Research*, 2012, 16(3): 266-271.)
- [16] Takasaki T, Hatakeyama K, Suzuki G, et al. The S receptor kinase determines self-incompatibility in *Brassicastigma*[J]. *Nature*, 2000, 403(6772): 913-916.