

于妍,姜威,陈庆山,等.大豆赖氨酸合成关键酶基因的克隆与定位分析[J].江苏农业科学,2018,46(23):41-45.

doi:10.15889/j.issn.1002-1302.2018.23.010

大豆赖氨酸合成关键酶基因的克隆与定位分析

于妍¹,姜威²,陈庆山²,邱红梅³,管宇⁴,蒋洪蔚²,李灿东⁵,唐敬仙¹

(1. 长春科技学院,吉林长春 130600; 2. 东北农业大学农学院,黑龙江哈尔滨 150030; 3. 吉林省农业科学院大豆研究所,吉林长春 130033;

4. 黑河出入境检验检疫局综合技术中心,黑龙江黑河 164300; 5. 黑龙江省农业科学院佳木斯分院,黑龙江佳木斯 154007)

摘要:为了从大豆中克隆 *DAPD*、*DAPE* 和 *DHDPR* 3 个大豆赖氨酸合成关键酶基因,并完成其电子定位和结构分析,借助已构建的本地化表达序列标签(expressed sequence tag,简称 EST)数据库,与不同物种的对应基因序列进行比对、拼接,同时利用已构建的大豆整合图谱,完成这 3 个关键酶基因的电子克隆及电子定位。结果表明,RT-PCR 克隆、序列分析的结果与预测序列基本一致。通过电子定位,得到这 3 个基因分别在 J、L 和 N 3 个连锁群上。通过对基因的结构分析,得到 *DAPD*、*DAPE*、*DHDPR* 3 个基因的 cDNA 长度分别为 1 467、1 080、1 035 bp, gDNA 长度分别为 4 662、3 728、3 158 bp,外显子数量分别为 8、9、8 个,内含子数量分别为 7、8、7 个。研究结果为其他氨基酸合成关键酶基因的克隆提供了参考依据,也为后续基因功能研究以及分子辅助育种打下了良好基础。

关键词:大豆;赖氨酸;合成酶基因;电子克隆;电子定位

中图分类号: S565.103 **文献标志码:** A **文章编号:** 1002-1302(2018)23-0041-04

大豆(*Glycine max*)是重要的油料作物、经济作物,植物性蛋白质的重要来源之一便是大豆蛋白质,其营养价值较高。大豆中人体必需的氨基酸含量丰富,其中赖氨酸含量远远高于谷物。由于赖氨酸能够促进人体充分吸收和利用其他营养物质,一旦缺乏将直接或间接影响机体的生长和发育,所以进一步提高大豆中的赖氨酸含量十分必要。在植物赖氨酸合成途径中,二氢吡啶、二羧二氢吡啶、二羧酸先后在 2,3-二氢吡啶二羧酸还原酶(DHDPR)、二氨基庚二酸异构酶(DAPE)、*N*- α -酰基-L,L-二氨基庚二酸脱酰酶(ADPD)的催化下生成赖氨酸^[1]。表达序列标签(expressed sequence tag,简称 EST)已经成为人类克隆不同物种特异性表达基因以及寻找未知基因的重要手段^[2-3]。截至目前,已有多项研究通过电子克隆手段获得了许多人类功能基因^[4-6]。但是由于目标物种序列信息有限,所以在植物基因克隆研究中,大部分试验方法仍局限于原有的、常规的思路。但是,随着植物 EST 数据库的逐步完善,其信息量快速增加,借助生物信息学技术来电子克隆基因势必成为基因克隆的一个方便快捷、有效的途径。另外,在分子标记研究领域,目前的研究多是将标记定位在连锁群上的,真正的基因定位鲜见报道。如果以基因自身的序列为依据,进而获得连锁群的定位,这样可以实现基因更加准确精细的定位。

本研究利用电子克隆手段,获得了大豆赖氨酸合成 3 个关键酶基因的全长 cDNA 序列,同时以已构建的大豆基因组遗传图和物理图的整合图谱为基础,完成这些基因的定位,并

通过 cDNA 和 gDNA 序列的比对,获得所有基因的结构信息,从而为后续基因功能的研究及分子辅助育种打下良好基础。

1 材料与方法

1.1 序列信息下载及生物软件

本研究从美国国立生物技术信息中心(National Center for Biotechnology Information,简称 NCBI)网站上分别下载了不同外源物种赖氨酸合成 3 个关键酶基因信息(表 1),并用 phytozome 网站(<http://www.phytozome.net/soybean.php>)进行在线比对。从 phytozome 网站(<http://soybase.org/SequenceIntro.php>)下载获得标记图谱 BARCSOYSSR Potential SSR Database。从 NCBI 的 FTP 网站(<ftp://ftp.ncbi.nih.gov/blast/executables/release/2.2.16/>)下载用于本地比对的 blast2.2.16 软件包。利用 DNAMan 软件进行 EST 片段比对,并筛选序列进行拼接。

表 1 外源物种赖氨酸合成关键酶基因信息

名称	登录号	外源物种名称	序列类型
<i>DAPD</i>	AY122998	拟南芥(<i>Arabidopsis thaliana</i>)	核苷酸
<i>DAPE</i>	ABA99594	水稻(<i>Oryza sativa</i>)	氨基酸
<i>DHDPR</i>	ABF94935	水稻(<i>O. sativa</i>)	氨基酸

1.2 试验植物

先将大豆品种东农 42 的种子在自来水中浸泡过夜,再将吸胀后的种子种在花盆中,于光照培养箱中培养,待长出 3 出复叶后,将幼嫩的叶片取下用作试验材料。

1.3 电子克隆

本研究依据不同物种间存在相同基因核苷酸序列同源性较高的特点,借助已构建的本地化大豆 EST 数据库,将已获得的不同物种的赖氨酸合成关键酶基因序列利用 Blast 软件与其进行比对,从比对出的数条 EST 序列中选取多条同源性高,且能完全覆盖整个基因 cDNA 片段的 EST 序列才可进行

收稿日期:2017-08-18

基金项目:黑龙江省自然科学基金重点项目(编号:ZD201213);教育部新世纪优秀人才支持计划(编号:NECT-1207-01)。

作者简介:于妍(1984—),女,辽宁沈阳人,硕士,讲师,研究方向为大豆遗传育种。E-mail:35789362@qq.com。

通信作者:唐敬仙,硕士,副教授,研究方向为农林经济管理。

E-mail:29364685@qq.com。

后续拼接工作。利用 DNAMan 5.0 软件的 Multiple sequence alignment 功能将目的基因与其同源序列进行联配分析,之后运用 Sequence Assembly 工具对所选 EST 序列进行进一步拼接组装完成电子延伸,进而获得全长基因序列。

表 2 赖氨酸合成关键酶基因引物信息

基因名称	上游引物	下游引物
DAPD	5'-ATGGCAGGCTCAAACATTCTTT-3'	5'-TCAAAGCCCCTCAAAAAACC-3'
DAPE	5'-ATGGCCATAACCGCCACC-3'	5'-TTAAAGGGGCAAAGATCCGTAAT-3'
DHDPR	5'-ATGGCAACATCTCTTCTGAAACTC-3'	5'-TCATCGCATATTACCCTCTCG-3'

1.5 大豆叶片总 RNA 提取及第一链 cDNA 的合成

依据 TRIzol 说明书提供的方法进行大豆叶片总 RNA 的提取。参照 M-MLV 使用说明书反转录合成 cDNA 的第一链,直接用作模板用于 PCR 扩增。

1.6 基因的克隆及测序

以东农 42 大豆品种 cDNA 作为模板,利用已设计的大豆 DAPD、DAPE 和 DHDPR 3 个酶基因特异引物,借助 PCR 扩增全长基因片段。反应条件如下:94 ℃ 5 min;94 ℃ 35 s,60 ℃ 35 s,72 ℃ 2 min,28 个循环;72 ℃ 7 min。待反应结束后,分别取 10 μL PCR 产物于 10 g/L 琼脂糖凝胶中进行电泳检测。回收目的片段并连接到 pGEM T-easy 载体上,转化大肠杆菌(Escherichia coli) DH5α 感受态细胞,在氨苄青霉素(Amp)抗性 X-gal/IPTG LB 培养基平板上将白色克隆挑出,并送交生物工程(上海)股份有限公司进行测序。

1.7 基因的电子定位

将已获得的 3 个大豆赖氨酸合成关键酶基因序列与本地数据库进行比对,将获得的序列与从 phytozome 网站下载的数据库再一次进行比对分析,由此便可准确地将基因定位在遗传图谱上。结合 Song 等已构建的图谱^[7],进而确定这些基因定位的连锁群以及距基因较近的两侧标记信息。

1.8 基因结构分析

根据比对分析结果,进一步确定基因的内含子与外显子数量,从而获得基因的精细结构。

1.4 引物设计

以拼接预测的基因序列为模版,利用 DNAMan 软件设计引物,引物序列见表 2。引物委托生工生物工程(上海)股份有限公司合成。

2 结果与分析

2.1 大豆赖氨酸合成关键酶基因的电子克隆

2.1.1 基因的电子预测 本试验将获得的不同外源基因序列与已构建的大豆本地化 EST 数据库进行比对,从比对结果中选取多条同源性高的序列,并进行拼接,进而获得全长基因,从而预测出目的基因的全长序列。经比对、拼接,大豆赖氨酸合成 3 个关键酶基因的全长序列预测信息见表 3。

表 3 大豆赖氨酸合成关键酶基因的预测信息

名称	基因长度 (bp)	氨基酸长度 (个)
DAPD	1 467	489
DAPE	1 080	360
DHDPR	1 035	345

2.1.2 基因的 RT-PCR 鉴定与测序结果分析 以大豆品种东农 42 叶片总 RNA 反转录的 cDNA 为模板,通过 RT-PCR 扩增,获得大豆 DAPD、DAPE 和 DHDPR 3 个关键酶基因 PCR 产物,经 10 g/L 琼脂糖凝胶电泳检测,结果显示泳道均出现单一条带,DAPD、DAPE、DHDPR 3 个关键酶基因大小分别约为 1 460、1 080、1 030 bp,全部与预测的基因长度相符(图 1)。将 3 个 PCR 产物分别与 pGEM-T easy 载体连接,进而获得对应的重组质粒。最后将重组质粒进行 PCR 扩增,经电泳检测,证实结果均与预测基因长度相符。

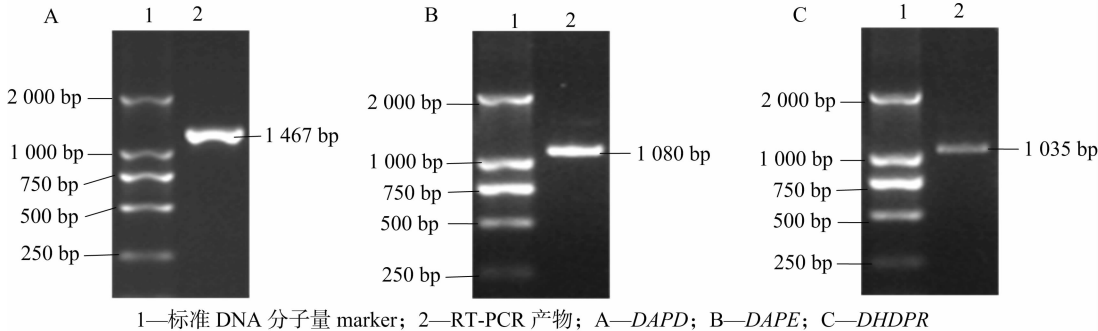


图1 大豆 DAPD、DAPE 和 DHDPR 基因的 RT-PCR 结果

借助 DNAMan 软件,将大豆赖氨酸合成 DAPD、DAPE 和 DHDPR 3 个关键酶基因重组质粒的测序结果与之前的预测结果进行比对分析,结果显示,在核苷酸水平上,所获得的 3 个基因 cDNA 序列与预测序列的同源性分别为 99.66%、99.17%、95.94%。利用 EBI→InterProScan 分别对 DAPD、DAPE 和 DHDPR 3 个酶基因编码的蛋白保守结构域进行预测。由表 4 可以看出,DAPD、DAPE、DHDPR 3 个关键酶基因含有的保守结构域分别具有 2,6-二氨基庚二酸脱羧酶、二

氨基庚二酸异构酶和 2,3-二氢吡啶二羧酸还原酶功能。通过以上分析,进一步验证了通过电子克隆获得的这 3 个全长基因序列是正确的。

2.2 大豆赖氨酸合成关键酶基因的电子定位与结构分析

2.2.1 基因的物理图定位 借助大豆基因组数据库,将已获得的基因序列片段与之进行比对,进而完成物理图定位分析。如果与基因组进行比对时出现多个片段同时匹配的情况,是由于该基因具有内含子。表 5 中显示了基因以及与基因组匹

表 4 大豆赖氨酸合成关键酶基因编码蛋白保守结构域信息

基因名称	氨基酸位置	保守结构域名称
<i>DAPD</i>	126 ~ 352	吡哆醛脱羧基酶(Orn_Arg_deC_N)
	71 ~ 486	2,6 - 二氨基庚二酸脱羧酶(Lys A;diaminopimelate decarboxylase)
<i>DAPE</i>	77 ~ 199,230 ~ 352	DAP 异构酶(DAP_epimerse)
	75 ~ 359	二氨基庚二酸异构酶(diaminopimelate epimerase)
<i>DHDPR</i>	69 ~ 344	2,3 - 二氢吡啶二羧还原酶(dihydropidicollinate reductase)

表 5 大豆赖氨酸合成关键酶基因的物理图定位

基因名称	基因长度 (bp)	基因匹配区段	基因组匹配区段	<i>E</i> 值	平均匹配率 (%)	基因组编号
<i>DAPD</i>	1 467	1 ~ 1 467	314 760 ~ 319 421	$0 \sim 8 \times 10^{-40}$	99.38	Super 40
<i>DAPE</i>	1 080	1 ~ 1 080	2 825 711 ~ 2 829 438	$1 \times 10^{-148} \sim 8 \times 10^{-26}$	99.78	Super 62
<i>DHDPR</i>	1 035	1 ~ 1 035	1 521 652 ~ 1 524 809	$1 \times 10^{-132} \sim 3 \times 10^{-31}$	99.75	Super 154

配区段的信息,此外还可以看出,*E* 值均趋近于 0,平均匹配率均在 99% 以上。这些数据足以说明,这些基因与基因组序列的匹配是真实且可信的。

2.2.2 基因的遗传图定位 根据赖氨酸合成 3 个关键酶基因与基因组匹配区段,同时参考标记图谱 BARCSOYSSR Potential SSR Database,*DAPD*、*DAPE* 和 *DHDPR* 3 个基因分别定位在 L、N、J 3 个连锁群上,获得了基因的两侧标记以及定位区间。由表 6、图 2 可以看出,*DAPD* 基因被定位在 L 连锁群 Super40 的 314 760 ~ 319 421 之间,且 BARC - 014655 -

01607 - BARC - 028787 - 06014 是分别位于此区间两侧最近的标记,距离是 1.1 cM;*DAPE* 基因被定位在 N 连锁群 Super62 的 2 825 711 ~ 2 829 438 之间,且 BARC - 024329 - 04849 - BARC - 028205 - 05792 是分别位于此区间两侧最近的标记,距离是 0.6 cM;*DHDPR* 基因被定位在 J 连锁群 Super154 的 1 521 652 ~ 1 524 809 之间,且 BARC - 022453 - 04332 - BARC - 014459 - 01381 是分别位于此区间两侧最近的标记,且定位于连锁群的同一位置,定位距离是 3 个基因中最近的。

表 6 大豆赖氨酸合成关键酶基因遗传图基因定位

名称	基因组编号	位置	附近标记	定位区间 (cM)	连锁群
<i>DAPD</i>	Super 40	314 760 ~ 319 421	BARC - 014655 - 01607 - BARC - 028787 - 06014	89.6 ~ 90.7	L
<i>DAPE</i>	Super 62	2 825 711 ~ 2 829 438	BARC - 024329 - 04849 - BARC - 028205 - 05792	64.4 ~ 65.0	N
<i>DHDPR</i>	Super 154	1 521 652 ~ 1 524 809	BARC - 022453 - 04332 - BARC - 014459 - 01381	26.3	J

2.2.3 基因结构分析 通过对大豆赖氨酸合成 3 个关键酶基因进行 cDNA 和 gDNA 序列的比对,进一步确定各个基因长度以及外显子、内含子的数量。表 7 结果显示,*DAPD*、*DAPE*、*DHDPR* 3 个酶基因 cDNA 长度分别为 1 467、1 080、1 035 bp;gDNA 的长度分别为 4 662、3 728、3 158 bp;外显子数量分别为 8、9、8 个;内含子的数量分别为 7、8、7 个。

表 7 赖氨酸合成关键酶基因结构比对信息

名称	cDNA 基因 长度(bp)	gDNA 基因 长度(bp)	外显子 数量(个)	内含子数量 (个)
<i>DAPD</i>	1 467	4 662	8	7
<i>DAPE</i>	1 080	3 728	9	8
<i>DHDPR</i>	1 035	3 158	8	7

3 讨论与结论

3.1 大豆赖氨酸合成关键酶基因的电子克隆

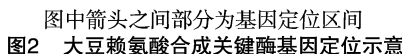
目前,电子克隆方法已经被广泛应用于很多领域,例如在人^[8]、甘蓝^[9]和牛^[10]的研究中均有报道。但大部分研究仅限于利用外源物种核苷酸或氨基酸序列与 GenBank 网站上提供的 EST 数据库进行比对,且由于 EST 数据库中的信息包含多个物种,所以造成信息量过大,因而比对起来必然存在以下问题:(1)时效性差;(2)会将许多不需要的物种信息同时比对出来,所以不仅须要比对多次,另外还须对结果进行细致筛选,为序列拼接带来不必要的麻烦。本研究在成功构建大

豆 EST 数据库的基础上进行本地比对,这样针对性更强,结果更准确。由此可见,以本地化 EST 数据库为基础进行基因电子克隆,既准确又快捷。

3.2 大豆赖氨酸合成关键酶基因的电子定位

目前,针对大豆脂肪酸含量^[11]、种子油分和蛋白^[12]、大豆花叶病毒抗性^[13]及胞囊线虫抗性^[14]等性状的基因定位均有报道,但关于基因方面的研究并不多见。本研究借助已有的大豆基因信息,构建了本地化 EST 数据库。同时利用生物学软件,完成了基因准确、快速的定位。而传统的分子标记,需要一定的群体以及大量人力、物力,既耗时又费力。基因电子定位是借助生物信息学相关软件并建立在测序基础之上的另一种用于基因定位的手段,其获得的基因两侧标记与基因位点之间的物理距离非常近,可为今后基于基因分子标记的开发奠定良好基础,从而真正完成基因准确精细的定位。

由本试验结果还可以看出,通过基因的电子定位,不仅能够更加快速、准确地相互找到对应标记,同时还能获得相邻序列的信息,这样就可以挖掘更多未知的功能性基因,进而完成 QTL 到 QTG 更深入的转换,同时也很好地将挖掘基因与遗传作图二者密切地联系在一起。除此之外,电子定位还可将基因 cDNA 序列和 gDNA 序列进行比对,同时对精细的基因结构,即内含子与外显子信息进行分析,为后续进一步研究基因提供了必要条件。目前基于基因序列的研究大部分仍然局限于 cDNA 水平,却忽视了基于 gDNA 水平上基因结构的分析,



王其鹏,王丽乔,张巍巍,等. 洋葱 EST-SSR 标记在大葱中的通用性及多态性分析[J]. 江苏农业科学,2018,46(23):45-49.
doi:10.15889/j.issn.1002-1302.2018.23.011

洋葱 EST-SSR 标记在大葱中的通用性及多态性分析

王其鹏¹, 王丽乔², 张巍巍³, 袁瑞江², 冯大领³

(1. 河北农业大学现代科技学院, 河北保定 071000; 2. 石家庄市农林科学研究院蔬菜研究所, 河北石家庄 050021;
3. 河北农业大学生命科学学院, 河北保定 071000)

摘要:以 31 个大葱自交系为材料,用 14 个洋葱 EST-SSR 标记,对各个自交系的 PCR 扩增产物进行分析。结果表明,其中 13 个 EST-SSR 标记在大葱不同自交系上具有通用性,1 个标记(ACM024)在 31 个大葱自交系上没有扩增产物,可用于大葱与洋葱的区分;13 对引物在 31 个自交系中共检测到等位位点 28 个,平均每对引物可以检测到 2.15 个;通过扩增产物的多态性分析,10 个引物具有多态性;通过聚类分析获得了各个自交系间的遗传距离及各自交系间的亲缘关系。在相似性系数 0.72 水平上,自选材料 1、日本品种山木一本与其他类型关系较远。各自交系间的遗传距离在 0~2.11 之间,其中遗传距离最小为 0(2013 年山东田选寒青巨葱与大梧桐象青叶 1 号;河北无极鸡腿葱与赞皇选日本型大葱;横滨大葱与自选材料 6),遗传距离最大值为 2.11(自选材料 1 与山木一本)。该研究为大葱品种鉴定、遗传多样性分析、遗传图谱构建等奠定了基础。

关键词:大葱;洋葱;EST-SSR 标记;通用性;多态性

中图分类号:S633.103.2 **文献标志码:**A **文章编号:**1002-1302(2018)23-0045-05

葱 (*Allium fistulosum* L.) 是百合科 (Liliaceae) 葱属

(*Allium* L.) 多年生草本植物。在我国有 3 个变种,即普通大葱 (*A. fistulosum* L. var. *giganteum* Makion)、分葱 (*A. fistulosum* L. var. *caespitosum* Makino) 和楼葱 (*A. fistulosum* L. var. *viviparum* Makino), 各个变种内还有不同的栽培类型。在北方地区种植的章丘大葱、北京高脚白、河北隆尧大葱、山东莱芜鸡腿葱等均为普通大葱的栽培类型^[1]。大葱在我国栽培历史悠久,不仅作为重要的蔬菜和调味料作物,其食用及药用价值也不断被人们发掘^[2-3],因此大葱的价值越来越被人们所重视。但长期以来,大葱育种工作发展缓慢,大葱作物的品质和产量潜力尚未被充分挖掘。

收稿日期:2017-08-12

基金项目:河北省科技支撑计划(编号:14226306D-6);河北省科技厅农业关键共性技术攻关专项(编号:16226304D-3);河北农业大学大学生创新创业训练计划(编号:cxzr2014009)。

作者简介:王其鹏(1994—),男,山东德州人,主要从事植物分子生物学及基因组学的研究。E-mail:1261870339@qq.com。

通信作者:袁瑞江,研究员,主要从事蔬菜育种和栽培等研究工作, E-mail:yrj208@126.com;冯大领,博士,副教授,主要从事植物遗传育种及利用等工作, E-mail:bjdalingfeng@163.com。

本研究对这方面内容进行了有效的补充。

参考文献:

- [1] Viola R E. The central enzymes of the aspartate family of amino acid biosynthesis[J]. Accounts of Chemical Research, 2001, 34(5): 339-349.
- [2] Huminiecki L, Bicknell R. In silico cloning of novel endothelial-specific genes[J]. Genome Research, 2000, 10(11): 1796-1806.
- [3] Wiemann S, Weil B, Wellenreuther R, et al. Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs[J]. Genome Research, 2001, 11(3): 422-435.
- [4] Schultz J, Doerks T, Ponting C P, et al. More than 1 000 putative new human signalling proteins revealed by EST data mining[J]. Nature Genetics, 2000, 25(2): 201-204.
- [5] 张德礼, 丁培国, 凌伦奖, 等. 人类新基因 *C17orf32* 的电子克隆和编码区序列 RT-PCR 验证[J]. 生物化学与生物物理进展, 2002, 29(4): 543-549.
- [6] 唐 军, 邓武军, 李仕新. 人类 PGAM 基因的电子克隆及生物信息学分析[J]. 仲恺农业工程学院学报, 2012, 25(2): 8-12.
- [7] Song Q J, Marek L F, Shoemaker R C, et al. A new integrated genetic linkage map of the soybean[J]. Theoretical and Applied Genetics, 2004, 109(1): 122-128.
- [8] 王 杰, 薛 鹏, 程 睿, 等. 人类新基因 ACBP5cDNA 的克隆及其同源物在小鼠/鸡胚发育过程中的表达[J]. 中国生物化学与分子生物学报, 2005, 21(2): 163-169.
- [9] 郭 慧, 金司阳, 刘 寒, 等. 甘蓝 AP2/ERF 转录因子的克隆和生物信息学分析[J]. 中国药师, 2017, 20(1): 6-10.
- [10] 靳聪飞, 梁婷玉, 刘新峰, 等. 牛 *GUCY1A3* 和 *SFXN1* 基因的克隆及生物信息学分析[J]. 中国畜牧兽医, 2017, 44(2): 357-364.
- [11] 邹 筱, 韩粉霞, 陈明阳, 等. 大豆脂肪酸主要组分含量 QTL 定位[J]. 作物学报, 2014, 40(9): 1595-1603.
- [12] 闫海波, 王 艳, 赵 琳, 等. 大豆蛋白和油分含量的 QTL 分析[J]. 大豆科学, 2016, 35(2): 228-233.
- [13] 韩英鹏, 赵 雪, 李修平, 等. 大豆种质对花叶病毒病和疫霉根腐病抗病性的 SSR 标记辅助鉴定[J]. 大豆科学, 2014, 32(1): 27-30.
- [14] 张闻博, 蒋洪蔚, 李仙东, 等. 基于元分析的大豆胞囊线虫抗性 QTL 的整合[J]. 中国油料作物学报, 2010, 32(1): 104-109.