

大豆脱氢抗坏血酸还原酶基因的电子克隆及进化分析

李晓晓,李蕊,李雅轩,蔡民华,胡英考

(首都师范大学生命科学院,北京 100037)

摘要 利用生物信息学和实验验证的技术路线,以拟南芥脱氢抗坏血酸还原酶基因 cDNA 序列为信息探针,对大豆 EST 数据库进行同源搜索,经同源性比对和序列组装,获得全长为 955bp 的大豆脱氢抗坏血酸还原酶基因的 cDNA 序列(GenBank 登录号为 DQ006810),经 RT-PCR 扩增,分子克隆和序列分析验证,结果表明与电子克隆完全一致。该 cDNA 序列具有完整的开放阅读框架(ORF,36bp—815bp),推测编码蛋白为 259 个氨基酸。通过与几种植物脱氢抗坏血酸还原酶蛋白序列的比较,发现该基因具有较高的保守性。结果表明根据物种间同源基因序列,对跨物种间 EST 数据库进行同源检索筛选,拼接,是新基因克隆的一条有效途径。

关键词 大豆;电子克隆;RT-PCR;脱氢抗坏血酸还原酶(DHAR);EST

中图分类号 S 565.1 **文献标识码** A **文章编号** 1000-9841(2007)01-0045-06

IN SILICO CLONING AND EVOLUTION ANALYSIS OF DEHYDROASCORBATE REDUCTASE cDNA FROM GLYCINE MAX

LI Xiao-xiao, LI Rui, LI Ya-xuan, CAI Min-hua, HU Ying-kao

(College of Life Science, Capital Normal University, Beijing 100037)

Abstract Using bioinformatics strategy and validating by experiment, a novel dehydroascorbate reductase gene(DHAR) of *Glycine max* was cloned and identified which was being blasted by search of *Glycine max* EST database with homologous gene cDNA of *Arabidopsis thaliana*. This sequence was confirmed by RT-PCR, molecular cloning and sequencing. The full length *Glycine max* DHAR gene cDNA of 955bp(GenBank accession, DQ006810) encoded dehydroascorbate reductase (DHAR) of 259 amino acid and contained a complete ORF of 777bp. Compared with several plants, we found the gene was almost conserved. The results revealed that it was a convenient technique for cloning novel gene by searching EST database with homologous gene of model plants such as the *Arabidopsis thaliana*.

Key words *Glycine max*; In silico cloning; Dehydroascorbate reductase; RT-PCR; EST

随着基因定位(连锁图谱,物理图谱,转录图谱)和人类基因组测序及生物信息技术的迅猛发展,特

别是人类基因组工作框架图和精细图谱及其初步分析结果的先后公布^[1,2],表达序列标签(EST)成为人

收稿日期:2006-06-28

基金项目:北京市教委科技发展计划项目基金资助(200310028112)

作者简介:李晓晓(1980-),女,硕士研究生,主要从事植物转基因方面的研究。E-mail:lix8012@163.com

通讯作者:胡英考副教授。E-mail:yingkaohu@sohu.com

类寻找未知功能的新基因,以及克隆不同时空差异表达基因的重要标志物^[3,6]。基于 EST 的电子克隆(in silico cloning)策略已成为克隆新基因的主要方法,它的主要原理是利用日益发展的生物信息学技术,借助电子计算机的巨大运算能力,通过对 EST 序列的组装和拼接,进一步利用 RT-PCR 的方法快速克隆功能基因,以此获得目的基因 cDNA 全长序列。迄今为止已有许多科研人员和实验室利用电子克隆的方法得到了很多人的功能基因^[7,8]。由于受到序列资料丰富度的限制,在植物领域基因克隆仍以常规的实验方法为主。但随着 EST 数据库的不断完善,利用生物信息学的方法进行某些植物基因的电子克隆已经成为可能^[9]。

脱氢抗坏血酸还原酶(DHAR)是抗坏血酸代谢途径的关键酶。它的主要功能是使抗坏血酸的二次氧化产物脱氢抗坏血酸在未自发分解为 2,3-二酮古洛糖酸之前,在谷胱甘肽参与下将其还原为抗坏血酸^[10]。因而脱氢抗坏血酸还原酶(DHAR)可使植物循环利用脱氢抗坏血酸,减少植物中抗坏血酸含量的损失。

本文采用电子克隆的方法,以拟南芥脱氢抗坏血酸还原酶(DHAR) cDNA 序列(GenBank, AY085616)为信息探针,BLAST 检索 GenBank 中的大豆 EST 数据库,拼接有部分同源的 EST 序列,获得大豆 DHAR 基因的 cDNA 序列,并进一步采用 RT-PCR 验证,克隆了大豆 DHAR 基因。在此基础上进一步利用生物信息学的方法分析该 cDNA 序列的基因特征,并和其他物种的该基因进行同源性比较分析,为进一步通过基因工程的方法利用该基因提高植物中抗坏血酸的含量打下基础。

1 材料和方法

1.1 电子克隆

电子克隆所用到的生物信息学数据库和生物软件:美国国家生物信息中心(National Center for Biotechnology Information, NCBI)提供并维护的 BLAST 数据库中的 nr 和 EST 数据库。网址是: <http://www.ncbi.nlm.nih.gov/BLAST/>。利用 BioEdit 软件进行 EST 序列片段的拼接。拼接得到的序列利用 ORF Finder(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)软件预测其可能的开放阅读框架。用 DNAMAN 软件进行不同物种间氨基

酸序列比对,分析其同源性进而绘制系统发育树。

1.2 实验克隆

1.2.1 实验材料、菌株和质粒 大豆种子于光照培养室培养一周左右,取其幼嫩子叶作为实验材料;宿主菌 *E. coli* DH5 α 由本实验室提供;质粒 pGEM-T Easy 载体为 Promega 公司产品。

1.2.2 工具酶和试剂 总 RNA 抽提所用的 Trizol 试剂购自 Invitrogen 公司;Taq DNA 聚合酶, DNA 凝胶回收试剂盒为 TaKaRa 公司产品;T₄ DNA 连接酶, SpeI, 反转录试剂盒为 Promega 公司产品; DNA 分子质量标准 D016-2 为鼎国公司的产品。

1.2.3 大豆叶片总 RNA 的提取及 cDNA 第一条链的合成 采用 Trizol 法提取大豆叶片的总 RNA。实验方法为:取大豆幼嫩叶片于液氮中研磨后加入 1mL Trizol,冰浴 10min;然后加入 350 μ L 氯仿,剧烈震荡 15s,12000rpm,4 $^{\circ}$ C 离心 10min;将上清移至无菌且无 RNase 的离心管中,加入等体积异丙醇混匀,-20 $^{\circ}$ C 沉淀 30min;12000rpm,4 $^{\circ}$ C 离心 10min 回收 RNA;弃上清,加入 1 mL 75%乙醇,12000rpm,4 $^{\circ}$ C 离心 5min;稍干后用无 RNase 的无菌水溶解沉淀。cDNA 的合成按反转录试剂盒说明步骤进行。

1.2.4 引物设计及合成 以电子克隆所得到的完整 cDNA ORF(编码 259 个氨基酸残基)序列为依据,设计合成特异引物。P1: CTA TCTAGA CCATGG CCACTGTGAGAGTTCAAG; P2: GAC ACTAGT TTAGCCTTCCACTTTAGGAC。特异引物由上海生工生物公司合成。

1.2.5 PCR 扩增大豆 DHAR 基因片段 用所设计的特异引物,以供试大豆 cDNA 为模板扩增特异片段。PCR 程序为:94 $^{\circ}$ C 预变性 5min,94 $^{\circ}$ C 变性 1min,55 $^{\circ}$ C 复性 1min,72 $^{\circ}$ C 延伸 1min,共 35 个循环;72 $^{\circ}$ C 延伸 10min。反应完毕后取 8 μ L 于 10g, L⁻¹进行琼脂糖凝胶电泳检测。

1.2.6 重组测序质粒的构建、克隆和 DNA 序列测定 PCR 产物经割胶回收后与载体 pGEM-T Easy(摩尔比 3 : 1)连接,4 $^{\circ}$ C 过夜。连接产物转化到大肠杆菌中,在 X-gal/IPTG Amp 平板上挑选白色克隆,由 TaKaRa 公司完成 DNA 序列测定。

2 结果和分析

2.1 大豆 DHAR 基因 cDNA 的电子克隆与序列分析

以拟南芥抗坏血酸还原酶(GenBank 登录号为:AY085616)的 cDNA 序列为信息探针,在 <http://www.ncbi.nlm.nih.gov/BLAST/> 站点进行大豆 EST 数据库检索,得到与之部分同源的大豆 EST 群,从中选取一条 EST(BG726302)作为种子序列再次 BLAST 检索大豆 EST 数据库,从得到的与种子序列同源性较高或有部分重叠的 EST 序列中选取同源性最高的四条 EST 序列,它们在 GenBank 的登录号分别为:BG652359, BG157481, AW831123, BG726302(图 1)。将这四条 EST 序列拼接组装得到全长 955bp 的大豆 DHAR 基因 cDNA 序列。该 cDNA 序列经 ORF Finder 软件分析后发现,该序列仅有一个完整的开放阅读框架(ORF),起始密码子在 36 位,终止密码子在 815 位,共编码 259 个氨基酸。

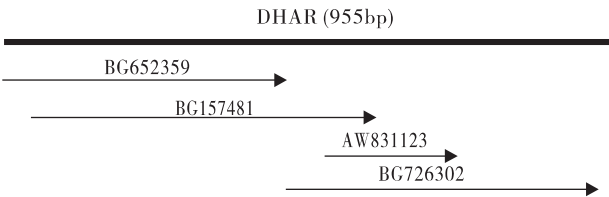
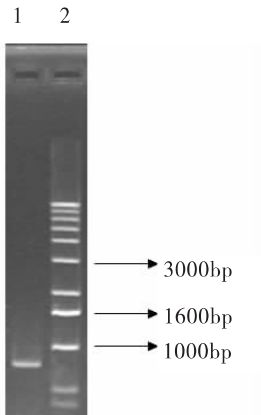


图 1 大豆 DHAR 基因的 cDNA 序列拼接示意图
Fig.1 The joint map of *Glycine max* DHAR cDNA sequence

2.2 RT-PCR 及酶切鉴定结果

PCR 结果经 10g/L 琼脂糖凝胶电泳检测,电泳结果出现单一条带,约 780bp,与预计大小相符(图 2)。将末端带 A 的 PCR 产物,插入 pGEM-T Easy 载体的多克隆位点的 *LacZ* 基因中,获得的重组质粒命名为 pGEM-T Easy-DHAR。该重组质粒经 *Spe* I 双酶切后,电泳显示的片段与预计的相符(图 3)。



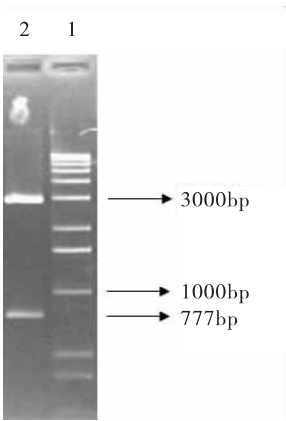
1 RT-PCR 产物 2 DNA 标准分子量(1kb)
1 RT-PCR product 2 DNA molecular marker(1kb)
图 2 大豆 DHAR cDNA RT-PCR 产物的琼脂糖电泳分析鉴定
Fig.2 Identification of the RT-PCR product of 777 bp ORF of *Glycine max* DHAR cDNA by agarose gel electrophoresis

2.3 测序结果和分析

重组质粒 pGEM-T Easy-DHAR 的测序结果与电子克隆序列经比对后发现二者是完全一致的。大豆 DHAR cDNA 的完整序列(图 4)及其编码蛋白的氨基酸序列的 GenBank 登录号为: DQ006810 和 AAY85184。

2.4 不同物种间蛋白同源性比较分析

将大豆 DHAR 基因与拟南芥(*Arabidopsis thaliana*, BAD43518),截形苜蓿(*Medicago truncatula*, AAY85185),百脉根(*Lotus corniculatus*,



2 DNA 标准分子量(1kb) 1 Spe I 酶切产物
1 DNA molecular marker(1kb) 2 Spe I digest product
图 3 重组质粒 pGEM-T Easy-DHAR Spe I 酶切产物电泳分析鉴定
Fig.3 Identification of the *Spe* I digest product of recombination plasmid by agarose gel electrophoresis

AAY52461), 花椰菜(*Brassica oleracea*, BAD14935),西红柿(*Lycopersicon esculentum*, AAY47049),水稻(*Oryza sativa*, BAD38160)等植物的该基因进行同源性比较发现在氨基酸水平上有很高的一致性,分别为 66.28%,79.92%,81.37%,67.05%,63.06%,60.66%(图 5)。

从利用 DNAMAN 软件构建的几种植物的 DHAR 系统发育树(图 6)上可以看出,同为豆科的百脉根、大豆和截形苜蓿同源性最高;拟南芥和花椰菜均为十字花科形成了同一分支,与大豆的同源性

略低;茄科的西红柿和禾本科的水稻虽单列为一支,但与大豆比较也显示了一定的同源性。从系统发育树上可以看出该基因在亲缘关系近的物种间同源性很高,即使亲缘关系较远的物种间其同源性也是较

高的。这说明 DHAR 基因在不同物种间还是具有较强的保守性,它们所对应的蛋白属于同一个蛋白家族。系统发育树基本上代表了它们在经典分类上的地位。

GGCACAAACGCAACATCATCATCAGCTTCAGCA

```

36  ATGTCCACTGTGAGACTTCAAGTTTCGGCGTGTGTGTTTTCTGCCACCGTCAACAACCTTTGCCTCCGCCAAAACGCCGTCGTTTCCTTC
    M S T V R V Q V S A C V F S A T V N N L C L R Q N A S V S F

126 AGAAAGAAGAAGCTTCTCAGATTGGTCTCAATGTCTTCGCTTCTCCTCCCAACCTTCGAAATGCCGTTAAAGCTTCGCTCACCACA
    R K K K L L R L V S M S S V P P S Q P F E I A V K A S V T T

216 CCCAACAGGCTCGGCGACTGCCCTTTTGGCAAAGGCTGTGTCTGACACTGGAGGAAAAACATCTACCTTATGACCCAACTTGGTGGAT
    P N R L G D C P F C Q R V L L T L E E K H L P Y D P K L V D

306 TTGACCAACAAGCCAGAATGCTTCCTTAAAGTCAATCCTGATGGTAAAGTTCCTGTCATAAAGTTTGATGAGAAGTGGGTTCTGATTCA
    L T N K P E W F L K Y N P D G K V P V I K F D E K W V P D S

396 GATATTATTACTCAAAACATTGGAAGAGAAGTATCCCAGCCCCCATTTGTTGACCCACCTGAAAAGGCCACGGCGGGTCAAAGATCTTT
    D I I T Q T L E E K Y P S P P L L T P P E K A T A G S K I F

486 TCAACATTTATTGCTTTTCTCAAGAGCAAGCATCCCAATGATGGAACAGACAAGCATTACTTACTGAACTGAGCTCTTTCACTGATTAC
    S T F I G F L K S K D P N D G T E Q A L L S E L S S F S D Y

576 ATCAAGGAAAATGCTCTTTCATCAACGGTACTGAGATATCTGCAGCTGACCTATCACTTGGACCAAAGCTATACCATTTAGAGATTGCT
    I K F N G P F I N G S E I S A A D L S L G P K L Y H L E I A

666 TTGGGGCATTATAAGAAATGGACAGTGCCTGATTCACTTACCTCTTTGAAGTCTTATATGAAGGCCATTTTCTCGAGGGAATCGTTGCTT
    L G H Y K K W T V P D S L T S L K S Y M K A I F S R E S F V

756 AAAACAAGTGCACAACCACAGGATGTCATTGAAGCTTGGCGTCTCTAAAGTGAAGGCTAATATTCAGTATCAGCAGTATCATTTCTAT
    K T S A Q P Q D V I E G W R P K V E G *

816 GAGACTCTGAGGACTATACCCCAACAAATAGCAATCCATCAAGAGTTCTAGACCTATATCGTTTTGTATTATATAACCTTATTTATTTAT

876 TCACATATGCACATATGGAATA
  
```

图4 大豆 DHAR 基因的 cDNA 序列及其推导的氨基酸序列(星号表示终止密码子)

Fig. 4 The nucleotide acid sequence and deduced amino acid sequence for DHAR of *Glycine max*
stop codon is indicated by asterisk(*)

3 讨论

脱氢抗坏血酸还原酶(DHAR)在谷胱甘肽(GSH)参与下能将脱氢抗坏血酸(DHA)还原为抗坏血酸(ASA),从而增加了植物细胞内维生素C的含量,对保护植物细胞内物质抵制氧化损伤起重要作用,同时也能增强植物耐炎热、寒冷和盐碱环境的能力^[11]。脱氢抗坏血酸还原酶基因首先由日本科学家从菠菜细胞中发现,随后又在其它植物中发现了该物质。利用转基因技术向大肠杆菌中植入各种基因并对其产生的酶一一进行鉴定,最终在大肠杆菌中发现了制造脱氢抗坏血酸还原酶的基因^[10]。利用EST数据库进行电子克隆是基因克隆的新途径,与传统的基因克隆相比,具有快捷、成本低、针对性强等特点,但也受到EST数据库数量和质量的限

制^[12],因此,只在EST资料非常丰富的模式物种(如人,鼠)中应用较多。在实际应用中,应该注意作为探针基因的物种应与目的基因的物种亲缘关系不宜太远^[13],否则可能会给基因的结构分析带来困难。但是生物中有些关键基因,大多具有一定的保守性,它们在不同物种间往往具有高度的同源性。因此可利用这些资源,以模式植物(如拟南芥)已知基因的EST为起点,进行跨物种间EST数据库同源检索,充分利用已有的资源,加速目标物种同源基因的克隆^[14]。在进行同源比较分析基因功能时,由于DNA的编码存在冗余性(第三个碱基的简并性),在DNA水平上需要有40%以上的一致性,才可能具有潜在的生物学意义,而蛋白质水平之间的25%同源性就可提示其功能的相似性,所以使用蛋白序列进行后续分析更能发现生物学意义^[15]。

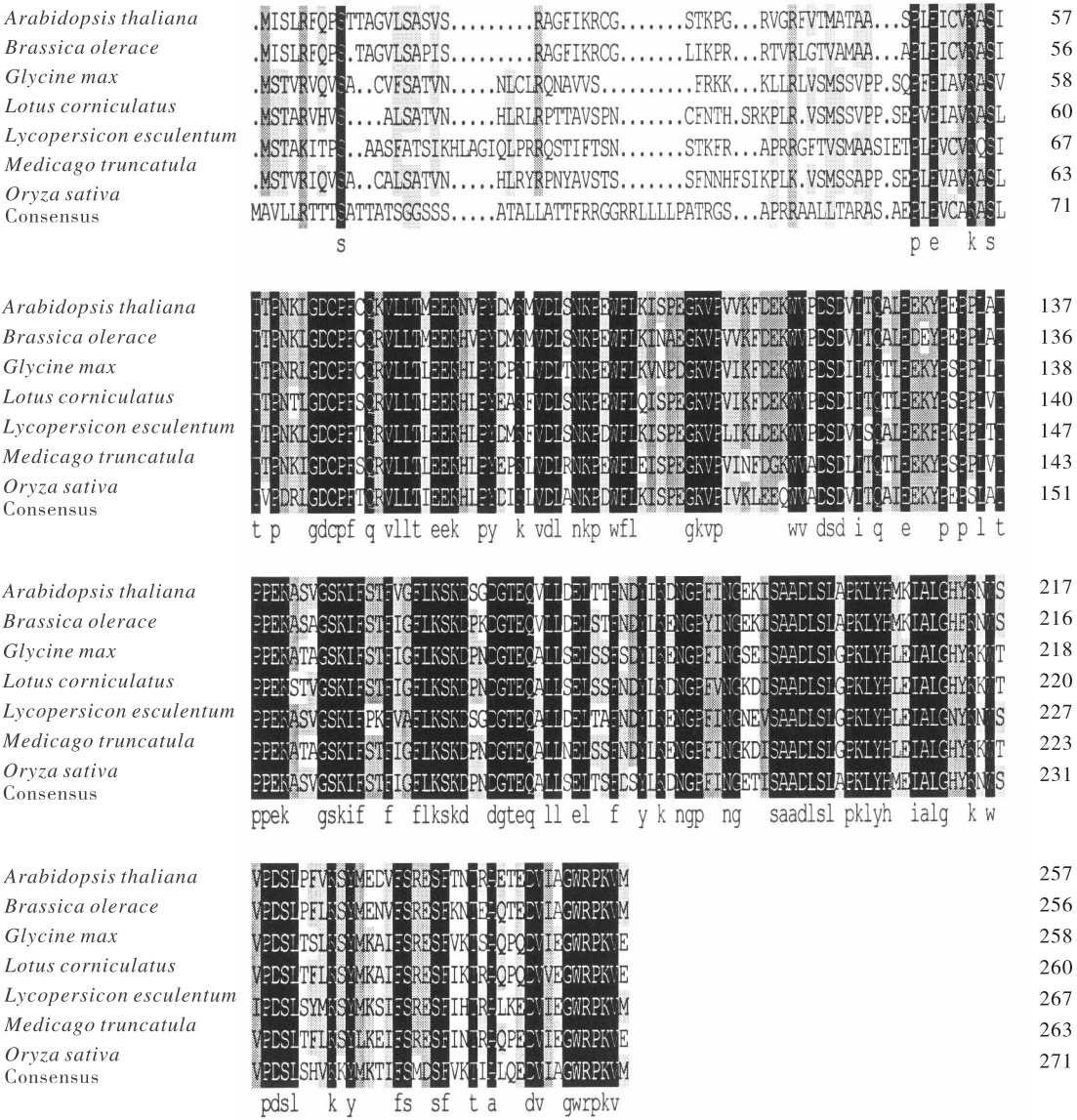


图 5 大豆 DHAR 基因蛋白序列与其他物种同源蛋白序列的比较

Fig. 5 An Alignment of *Glycine max* DHAR and homologous protein from other species *Arabidopsis thaliana* (BAD43518), *Brassica oleracea* (BAD14935), *Lotus corniculatus* (AAY52461), *Lycopersicon esculentum* (AAY47049), *Medicago truncatula* (AAY85185), *Oryza sativa* (BAD38160)

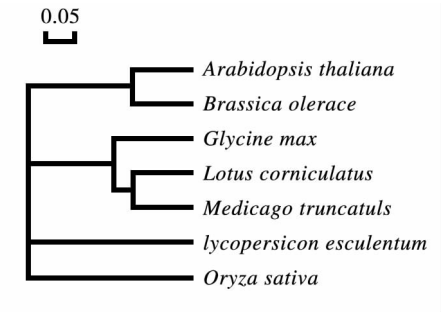


图 6 植物 DHAR 的系统发育树 (图上线段表示进化距离)
Fig. 6 Phylogenetic tree of plant DHAR (The length means evolution distance)

本文利用物种间同源基因的保守特性,以拟南芥脱氢抗坏血酸还原酶基因的 cDNA 序列为模板,对大豆 EST 数据库进行同源检索筛选、拼接及 RT-PCR 验证,成功获得了大豆脱氢抗坏血酸还原酶基因完整的 cDNA 序列。表明利用不同物种间的关键基因的高度保守性,采用生物信息学技术对目标物种的 EST 数据库进行筛选和拼接,完全可能快速克隆不同物种间的同源基因,从而为人们进行不同物种间功能基因的克隆提供了一条有效途径。

参 考 文 献

- [1] Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome[J]. Nature, 2001, 409 (6822):860—921.
- [2] Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome[J]. Science, 2001, 291 (5507):1304—1351.
- [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] Blumberg H, Conklin D, Xu W F, et al. Interleukin 20: discovery, receptor indentification, and role in epidermal function[J]. Cell, 2001, 104 (1): 9—19.
- [5] Huminiecki L, Bicknell R. In silico cloning of novel endothelial specific genes[J]. Gemone Research, 2000, 10 (11): 1796—1806.
- [6] Schultz J, Doerk T, Ponting C P, et al. More than 1 000 putative new human signaling proteins revealed by EST data mining[J]. Nat Genet, 2000, 25(2): 201—204.
- [7] OBrien K P, Tapia Paez I, Stahle Backdahl M, et al. Characterization of five novel human genes in llq13—q22 region[J]. Biochemistry Biophsiology Research Commum, 2000, 273: 90—94.
- [8] Zhang D L, Sun X J, Ling L J, et al. Molecular cloning, characterization,chromosomal assignment, genomic organization and verification of SFRS12(SRrp508), a novel member of human SR protein superfamily and a human homolog of rat SRrp86[J]. Acta Genetica Sinica, 2002, 29(5):377—383.
- [9] Tang X R, Wu H, Jia M, et al. Isolation and expression analysis of cDNA encoding a dsRNA binding protein homologue OsRBP of rice[J]. Plant Physiology and Molecular Biology, 2002, 28:41—45.
- [10] Chen Z, Young T E, Ling J, et al. Increase vitamin C content of plant through enhanced ascorbate recycling[J]. Proc Natl Acad Sci USA, 2003, 100:3525—3530.
- [11] 靳月华,陶大利,郝占庆,等. 环境胁迫和抗坏血酸的氧化还原状态[J]. 植物学报, 2003,45(7):795—801.
- [12] Prigent C,Gill R, Trower M, et al. In silico cloning of a new protein kianse, Aik2, related to Drosophila aurora using the new tool: EST Blast[J]. In Silico Biology, 1999, 1(2):123—128.
- [13] Huang J, Wang J F, Zhang H S, et al. In silico cloning of glucose—6—phosphate dehydrogenase cDNA from rice(*Oryza sativa* L.)[J]. Acta Genetica Sinica, 2002, 29(11):1012—1016.
- [14] 陈大福,牛宝龙,翁宏飙,等. 蜜蜂(*Apis mellifera*)腺苷酸转移载体基因 cDNA 的电子克隆[J]. 中国农业科学,2004,37(6):923—927.
- [15] 张成岗,贺福初. 生物信息学方法与实践[M]. 北京:科学出版社,2002. 63—93.

《大豆科学》2003~2005 年的综合评价

根据《中国学术期刊综合评价数据库(CAJCED)》2003~2005 年 6000 种左右统计刊源 524 万余条中国期刊引文数据及 CNKI“中国期刊网”中心网站的全文下载记录统计,《大豆科学》2003~2005 年度引文计量指标及 Web 下载统计指标如下:

2003~2005 年《大豆科学》的计量指标统计表

| 年份 | 统计期刊源 | | 被引频次 | 影响因子 | 引文即年指标 | 他引总引比 | 被引半衰期 | 载文量 | Web 下载率 |
|------|---------|------|------|--------|--------|--------|-------|-----|---------|
| | 引文条数(篇) | 期刊总数 | | | | | | | |
| 2003 | 140 万 | 5584 | 539 | 0.5899 | 0.1364 | 0.7069 | 6.8 | 66 | 15.9859 |
| 2004 | 170 万 | 5941 | 591 | 0.644 | 0.015 | 0.8037 | 7.1 | 6.5 | 34.5 |
| 2005 | 214 万 | 6182 | 718 | 0.771 | 0.015 | 0.87 | 7.5 | 68 | 23.1 |

另据《中国学术期刊综合评价数据库(CAJCED)》2005 年析出的 214 万条中国期刊引文数据及 CNKI“中国期刊网”中心网站 2005 年 1~12 月全文下载记录(1.5 亿篇次)的大样本数据分析,《大豆科学》的 5 年影响因子为 0.815。

薛津
《大豆科学》编辑部