

大豆种子不同发育时期全长均一化 cDNA 文库的构建

李 晨, 闫晓红, 周新安, 沙爱华, 单志慧, 周 蓉, 魏文辉

(中国农业科学院油料作物研究所基因组学与分子生物学研究室/农业部油料作物生物学重点开放实验室, 武汉 430062)

摘要: 【目的】为获得大豆种子发育相关基因, 并为大豆基因组资源提供材料。【方法】采用 SMART (switching mechanism at 5' end of RNA transcript) 与 DSN (duplex-specific nuclease) 均一化相结合的技术, 构建了全长均一化 cDNA 文库, 采用涂平板测定和 PCR 快速鉴定的技术分析文库质量, 用 3730 测序仪对文库克隆进行测序。【结果】构建了大豆种子不同发育时期全长均一化 cDNA 文库, 最大限度地获得了大豆种子不同发育阶段表达的基因序列, 原始文库的库容为 6.0×10^5 cfu/mL, 重组率接近 100%, 插入片段大小在 0.6—2.0 kb 之间, 平均长度超过 1.0 kb。经过大规模的质粒提取和测序, 共得到了 36 656 条高质量的 EST 序列, 序列的平均读长在 600 bp 以上。【结论】经过 EST 序列拼接分析, 整个文库有着很高的非冗余性。文库的代表性和重组片段的完整性均达到了分离筛选目的基因的建库要求。

关键词: 大豆; cDNA 文库; 均一化技术; 表达序列标签

Construction of a Normalized Full-Length cDNA Library of Soybean Seed at Different Developmental Stages

LI Chen, YAN Xiao-hong, ZHOU Xin-an, SHA Ai-hua, SHAN Zhi-hui, ZHOU Rong, WEI Wen-hui

(Department of Genomics and Molecular Biology, Institute of Oil Crops, Chinese Academy of Agricultural Sciences/Key Laboratory for Biological Sciences of Oil Crops, Ministry of Agriculture, Wuhan 430062)

Abstract: 【Objective】 The objective of the study is to obtain full-length genes related to seed development and to provide resources for soybean genomics. 【Method】 A normalized cDNA library enriched in full-length sequences was constructed using DSN (duplex-specific nuclease)-normalization method combined with SMARTTM(switching mechanism at 5' end of RNA transcript) technique. 【Result】 The titer of unamplified cDNA library was about 6.0×10^5 cfu/mL. The average cDNA inserts was more than 1.0 kb with a recombination rate of nearly 100%. After large-scale plasmid extraction and sequencing, 36 656 EST sequences were generated from the soybean seed cDNA library and the average length was more than 600 bp. All the EST sequences were assembled and 27 982 unigenes were gotten, which indicated that the cDNA library was a non-redundant library. 【Conclusion】 These results indicate that the normalized full-length cDNA library has been established successfully, which is convenient for further studying the molecular mechanism and gene cloning of seed development.

Key words: soybean; cDNA library; normalization; EST

0 引言

【研究意义】大豆起源于中国, 是一种重要的粮油作物, 各种营养成分包括碳水化合物、油脂和蛋白质, 超过种子干重的 90%。自 1923 年以来, 利用常

规育种方法, 中国已培育出了 1 000 余个大豆品种^[1], 在促进大豆生产和保障消费需求方面发挥了重要作用。随着分子生物学的发展, 人们逐渐认识到影响大豆产量及营养成分的关键因素在于不同基因相互作用的结果。在花的发育、荚、籽粒形成过程中, 储存在

收稿日期: 2009-06-30; 接受日期: 2009-09-23

基金项目: 国家自然科学基金面上项目 (30671312)、中央级公益性科研院所基本科研业务费专项资金 (20065049)

作者简介: 李 晨, 博士研究生。Tel/Fax: 027-86817881。通信作者魏文辉, 副研究员, 博士。Tel/Fax: 027-86817881; E-mail: whwei@oilcrops.cn

植株体内的营养物质开始在籽粒中积累, 籽粒达到生理成熟期时也达到了最大干重^[2], 一系列基因的表达控制着种子的发育。据统计, 大豆基因组约有 6 万余个基因^[3]。基因以一定时序性表达, 影响着复杂的代谢途径, 并涉及众多基因的协同作用^[4]。对于种子发育相关基因的研究, 特别是了解控制胚和胚乳相关基因的作用机理, 可以使人们在分子水平上解析种子发育和营养物质合成与积累的分子机理, 有助于更有效地开展种子产量和品质改良的基因工程。构建含有种子发育相关基因的 cDNA 文库并进行 EST 测序, 对于相关基因的克隆及研究其调节网络有着重要意义。【前人研究进展】对于胚发育相关基因的研究主要集中在拟南芥等模式生物中。已鉴定和分离了一些影响胚胎形成的基因, 如 *GNOM*^[5]、*FACKE*^[6]、*LEC2*^[7]。其中 *FACKE* 参与脂类的生物合成, 影响胚细胞的分裂与排列方式。目前, 大豆种子发育相关基因的研究还处于起步阶段, 利用全长 cDNA 文库克隆分析相关基因 EST 已取得了一些成绩。Umezawa 等^[8]测序并分析了大豆 40 000 个 cDNA 克隆, Shoemaker 等^[9]通过对大量 EST 的分析, 建立了一系列重叠群, 并根据重叠群的表达模式, 对相同功能的基因进行了鉴定和分析。全长 cDNA 文库已经成功应用于基因的克隆和表达分析, 这些研究包括: 拟南芥、水稻、小麦和玉米等^[10-15]。【本研究切入点】通过大规模的 EST 测序分析, 可以最终获得基因组中编码序列区的核苷酸全序列图。并通过已知功能基因的相似性分析推测每条 EST 所代表基因的功能。因此构建含有大豆种子不同发育时期的全长 cDNA 文库, 是一种克隆相关基因、研究其功能及代谢途径和网络的好方法。据统计, 1999 年 6 月, 大豆 EST 数量在 GenBank 中仅排在第 17 位, 远远落后于水稻、拟南芥、番茄和玉米等植物 EST 的研究 (<http://www.ncbi.nlm.nih.gov/dbEST>)。然而, 2009 年 5 月, 大豆 EST 数量已经达到 1 386 000 条, 在植物中仅次于水稻和拟南芥, 但由于组织特异性基因和特殊时期才表达的基因的存在, 部分特殊发育时期表达的基因常未能包含在一般的 cDNA 文库中, 因此有必要构建含有大豆种子发育相关基因的全长 cDNA 文库。【拟解决的关键问题】以高油脂含量的大豆栽培种为材料, 通过对种子发育不同时期 RNA 的提取, 构建了包含大豆种子不同发育时期特异性表达基因的 cDNA 文库, 通过对文库 EST 的进一步分析和筛选, 预计可以得到一批 (100—200 条) 与大豆产量、蛋白含量、脂肪酸含量、养分高效利用等重要经济性状密

切相关的基因序列。

1 材料与方法

1.1 材料

所用材料为大豆品种“中豆 32”, 由中国农业科学院油料作物研究所大豆研究室提供。该品种是以“湘春豆 10 号”作母本, 与父本“铁丰 18”杂交后, 经系谱法选择育成的大豆品种 (审定编号: 国审豆 2006017)。在开花后 10、15、20、25、30、35、40、45、50 d (DAF, day after flowering) 的 9 个时间点, 分别取胚作为试验材料。将第 1 天开的花用线做好标记, 再在不同的发育时间点取材。

1.2 方法

1.2.1 总 RNA 的提取 取 0.1 g 左右不同时期的大豆胚, 液氮速冻, 研钵研磨后加入 1 mL Trizol 试剂, 严格按照 Trizol 说明书步骤进行操作。然后通过紫外分析仪检测所提 RNA 的浓度、纯度, 甲醛变性琼脂糖凝胶电泳检测 RNA 完整性。

1.2.2 全长 cDNA 的合成 将大豆 9 个时期样品提取的总 RNA 等量混合后, 用 Clontech 公司的 Creator SMART cDNA Construction Kit (Cat.No.634903) 合成一、二链。其中利用均一化试剂盒 Trimmer-Director Kit (Evrogen, Cat.No.NK002) 中 CDSIII/3M adaptor 替换 CDSIII/3'primer, 方便进一步的均一化操作。采用 LD-PCR 对反转录得到的 cDNA 进行扩增, PCR 试剂使用 Clontech 公司的 Advantage 2 PCR Kit (Cat.No.639207), 95℃ 7 s, 66℃ 20 s, 18 个循环。反应结束后, 取 5 μL 产物于 1.1% 含 EB 的胶上电泳检测 cDNA 合成结果。

1.2.3 全长 cDNA 的均一化 使用 QIAquick PCR Purification Kit (Qiagen Cat.No.28104) 将扩增得到的双链 cDNA 纯化, 将 cDNA 溶于 15 μL 无菌水中, 取 1 μL 测定浓度, 用无菌水调节浓度至 100 ng·μL⁻¹ 左右。参照 Trimmer-Director Kit (Evrogen, Cat.No.NK002) 的方法, 利用 DSN (duplex-specific nuclease) 处理 cDNA, 取稀释的 cDNA 进行不同循环数的 PCR 扩增, 分别取样电泳确定最佳循环数, 再进行第 2 次 PCR 扩增, 条件为 95℃ 15 s, 66℃ 20 s, 72℃ 3 min, 12 个循环, 接着 64℃ 15 s, 72℃ 3 min, 电泳检测均一化效果, 并将产物回收。

1.2.4 均一化全长 cDNA 文库构建 取均一化 cDNA, 经 *Sfi* 酶切后, 电泳检测, 将酶切产物的 1—3 kb 区域切下, 并将产物回收后与 pDNR-LIB 载体连接, 将连

接产物加到 0.25 μm Millipore 纯化膜上脱盐纯化 1 h, 电击转化到大肠杆菌 DH10B 中。菌液涂于氯霉素平板上。

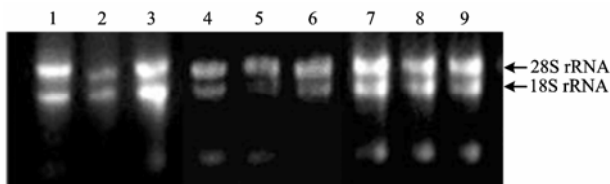
1.2.5 菌落鉴定 取 2 μL 菌液涂于氯霉素平板上, 计算平板上的库容数, 得出库容为克隆数 \times 稀释倍数 \times 连接产物体积/转化用连接物体积。笔者采用蓝白斑筛选剔出假阳性克隆, 随机挑取 12 个白斑克隆进行菌落 PCR 鉴定, 电泳检测插入片段大小及小片段比率。

1.2.6 EST 测序及分析 该文库在送检合格后, 进行了大规模质粒提取和测序, 采用 3730 测序仪对文库克隆进行 5'测序。对所有的序列去低质量、去载体、拼接, 并且根据所有 contig 和 singlets 的基本情况统计。该工作由北京六合华大基因科技股份有限公司完成。

2 结果

2.1 总 RNA 的提取

以大豆 9 个时期胚为材料, 提取总 RNA, 电泳效果较好(图 1), 用紫外分光光度计测定 OD 值(表), $\text{OD}_{260/280}$ 在 1.9—2.2, 符合 RNA 比值范围, $\text{OD}_{260/230}$ 在 2.0 左右, 表明 RNA 比较纯, 满足建库要求。



1: 10 天; 2: 15 天; 3: 20 天; 4: 25 天; 5: 30 天; 6: 35 天; 7: 40 天; 8: 45 天; 9: 50 天
1: 10 day; 2: 15 day; 3: 20 day; 4: 25 day; 5: 30 day; 6: 35 day; 7: 40 day; 8: 45 day; 9: 50 day

图 1 大豆种子发育 9 个时期总 RNA 电泳图

Fig. 1 Gel analysis of total RNA from soybean seed at 9 different developmental stages

2.2 全长 cDNA 的合成

经反转录得到的 cDNA 通过 LD-PCR, 电泳显示 cDNA 中含有若干条亮带(图 2), 代表高丰度基因, 而且不同大小和丰度的 mRNA 都得到了有效的反转录和扩增。

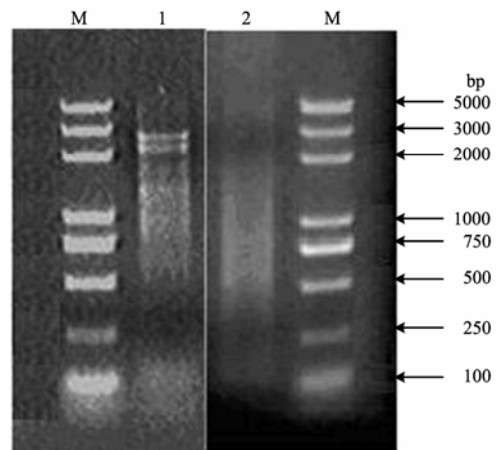
2.3 均一化效果检测

经过 DSN 处理后, 电泳呈现一条均匀的弥散条带, 均一化前的亮带已基本消失, 同时弥散条带至

表 大豆种子 9 个发育时期总 RNA 测定 OD 值及浓度

Table Total RNA purity and quantity at 9 different developmental stages

不同时期 Different stages	光密度比值 $\text{OD}_{260/280}$	光密度比值 $\text{OD}_{260/230}$	RNA 浓度 RNA concentrations ($\text{ng}\cdot\mu\text{L}^{-1}$)
10 day	2.01	2.22	1421
15 day	1.98	2.12	385
20 day	2.03	2.15	1580
25 day	1.95	1.88	1203
30 day	1.96	1.83	1103
35 day	2.02	2.30	1506
40 day	1.93	2.10	1305
45 day	1.97	2.28	1368
50 day	2.05	2.28	1420



1: 均一化前 cDNA; 2: 均一化后 cDNA; M: DL2000 plus marker
1: Non-normalized cDNA; 2: Normalized amplified cDNA; M: DL2000 plus marker

图 2 反转录 cDNA 均一化前后电泳图

Fig. 2 Agarose gel electrophoresis of non-normalized and normalized amplified SMART-prepared cDNA

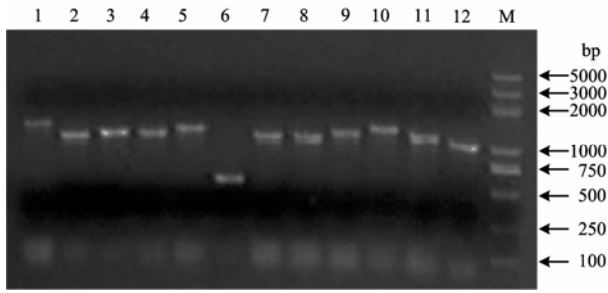
3 000 bp 以上, 长片段 cDNA 扩增后亮度较之前有明显增加(图 2), 说明均一化使原本丰度较低的 cDNA 相对含量得到了增加, 而高丰度 cDNA 含量明显下降。

2.4 文库库容及插入片段大小的分析

文库库容为 6.0×10^5 cfu, 重组率约 100%, 插入片段大小在 0.6—2.0 kb(图 3), 平均大小超过 1.0 kb。

2.5 测序结果分析

采用 3730 测序仪对构建的 cDNA 文库的插入片段进行了大规模的测序, 共得到了 36 656 条高质量的



1—12: 插入片段; M: DL2000 plus marker
1-12: Inserts; M: DL2000 plus marker

图 3 文库插入片段大小检测

Fig. 3 Size of cDNA library inserts

EST 序列, 序列的平均读长在 600 bp 以上。去除载体等序列后得到高质量 EST, 各长度 EST 统计结果见图 4。500—600 bp 读长的 EST 有 28 706 条, 600 bp 以上的 525 条, 超过 500 bp 读长的 EST 共 29 231 条, 达到 79.7%, 只有 2.7% 的 EST 读长少于 200 bp, 而读长少于 100 bp 的 EST 被认为没有参考价值被去除掉。所有 EST 序列经过拼接得到 27 982 条 unigene, unigene 的丰度达到了 66%, 显示整个文库有着很高的非冗余性。依据文库序列与公共数据库的 BLAST 比对结果, 统计得出不同 unigene 比对上基因的起始位置 (图 5), 因为测序采用 5' 测序, 如果是从该基因的起始密码子处开始比对上, 则认为该 unigene 是全长。一位密码子代表起始密码子, 共有 11 696 个 unigene 与公共数据库中基因的起始密码子比对上, 占全部比对的 25 354 个 unigene 的 46.1%, 这一结果表明文库的全长率达到了 46.1%, 文库中全长 cDNA 序列所占比例较高。

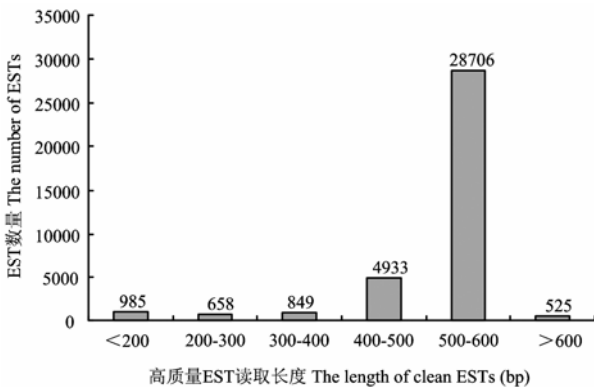


图 4 cDNA 文库高质量 EST 测序读取长度

Fig. 4 The length of clean EST

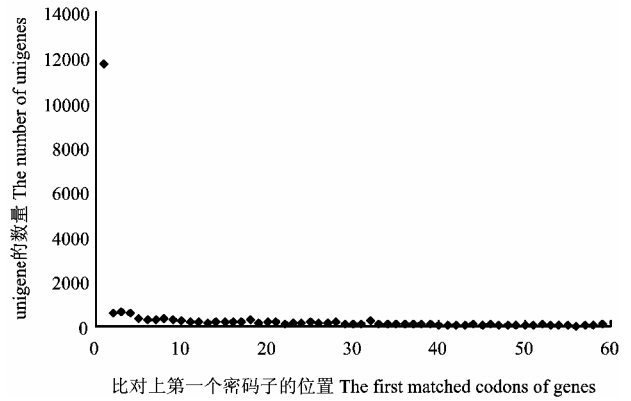


图 5 cDNA 文库的全长率

Fig. 5 The rate of full-length cDNA in the library

3 讨论

构建全长均一化 cDNA 文库的步骤大致包括细胞总 RNA 的提取、cDNA 第 1 链的合成、cDNA 第 2 链的合成、全长 cDNA 的均一化、双链 cDNA 的克隆等。在这些步骤中, RNA 质量是影响 cDNA 文库的关键因素, 因此必须要保证 RNA 纯度高且无降解。本试验采用 Trizol 方法提取的总 RNA 无蛋白和其它杂质污染, 琼脂糖凝胶电泳显示 28S rRNA 和 18S rRNA 的条带清晰, 且亮度比值为 2 : 1 以上, 无扩散现象, 说明其纯度和完整性高, 完全符合建库的要求。

构建全长 cDNA 文库的方法较多, 前人也作了大量的研究^[16-19]。Clontech (美国) 公司推出的 SMART 技术, 是目前应用最广泛的构建全长 cDNA 文库的方法之一^[20]。以前均一化文库时利用羟磷灰石等^[21], 不适合全长 cDNA 的均一化。目前针对全长 cDNA 的复性式均一化技术主要有磁珠的方法^[22]和 DSN 酶均一化的方法^[23], 前者步骤较为繁琐。

本试验利用 DSN 酶均一化的方法, DSN 酶是近年来发现的一种热稳定核酸酶^[24], 可以选择性降解双链 DNA 和 DNA-RNA 杂交体中的 DNA。经 DSN 酶处理后, 双链 cDNA 中代表高丰度基因的亮带消失, 呈现出一条均匀的弥散条带, 说明利用 DSN 酶均一化的方法可以有效地使不同丰度基因的拷贝数趋于一致, 相对提高低丰度基因的拷贝数, 克服了基因转录水平上的差异给文库筛选和分析带来的障碍, 这样更容易检测出时空特异性表达的低拷贝基因, 大大增加了克隆低丰度基因的机率, 在 EST 测序过程中可以避免大量相同基因的干扰。

一个全长 cDNA 文库的质量主要从以下 3 方面评价: (1) 库容量。原始文库的库容为 6.0×10^5 cfu, 保证了文库的完整性与覆盖度; (2) 插入片断的大小。该文库插入片段的平均长度超过 1 000 bp, 表明序列的完整性很高; (3) 均一化效果。通过均一化前后电泳对比可以看出均一化后高丰度基因明显降低。通过大规模的 EST 测序, 所有 EST 序列经过拼接得到 unigene, 丰度达到 66% (结果未发表), 显示整个文库有着很高的非冗余性, 说明均一化文库很少有重复的克隆, 均一化效果很好。

同时, 作为优化的文库系统及大规模测序结果可以应用于进一步的芯片制作。分析不同基因在种子发育过程中的表达, 从中筛选出种子发育特异性表达的基因, 进一步分析其功能, 并最终研究种子发育相关基因调控网络, 将为育种工作奠定理论基础。

大豆 ($2n=40$) 基因组较大, 约 1 115 Mb, 它可能起源于复杂的基因组范围的扩增事件^[25-26]。尽管其基因组比较复杂, 但其基因组草图序列已经发布 (<http://www.phytozome.net/soybean>), 这为大豆基因甚至更广泛领域的植物科学研究都带来了便利。

除了完整的基因组序列, 基因的转录产物对于分子生物学的研究也至关重要。现在公共数据库中, 尽管大豆不同组织、器官表达基因的 EST 均已非常丰富 (<http://www.ncbi.nlm.nih.gov/dbEST>), 但专门报道构建其不同发育时期种子的混合全长均一化 cDNA 文库还是第 1 次。因此, 通过笔者所构建的全长 cDNA 文库不仅可以应用于目的基因的克隆, 同时对大豆基因组数据的充实也起到了一定的补充作用。

4 结 论

本研究构建了高质量的大豆种子发育不同阶段全长 cDNA 文库, 同时进行了质粒提取并采用 3730 测序仪对构建的文库进行了大规模的测序, 为研究种子发育相关基因及调控网络等后续工作奠定基础。共得到了 36 656 条高质量的 EST 序列, 序列的平均读长在 600 bp 以上。

References

[1] 邱丽娟, 王昌陵, 周国安, 陈受宜, 常汝镇. 大豆分子育种研究进展. 中国农业科学, 2007, 40(11): 2418-2436.
Qiu L J, Wang C L, Zhou G A, Chen S Y, Chang R Z. Soybean molecular breeding. *Scientia Agricultura Sinica*, 2007, 40(11): 2418-2436. (in Chinese)

[2] 刘晓冰, 金 剑, 王光华, 杨恕平, 张秋英, 李艳华. 蛋白质含量不同的大豆籽粒形成过程中内源激素活性的变化. 大豆科学, 2000, 19(3): 238-242.
Liu X B, Jin J, Wang G H, Yang S P, Zhang Q Y, Li Y H. Endogenous hormone activities during seed development of soybean genotypes differing in protein content. *Soybean Science*, 2000, 19(3): 238-242. (in Chinese)

[3] He C Y, Tian A G, Zhang J S, Zhang Z Y, Gai J Y, Chen S Y. Isolate and characterization of a full length resistance gene homolog from soybean. *Theoretical and Applied Genetics*, 2003, 106: 786-793.

[4] Coen E S. The role of homeotic genes in flower development and evolution. *Annual Review of Plant Physiology and Plant Molecular Biology*, 1991, 42: 241-279.

[5] Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld J U, Salchert K, Koncz C, Jürgens G. A conserved domain of the *Arabidopsis* GNOM protein mediates subunit interaction and cyclophilin 5 binding. *The Plant Cell*, 2000, 12: 343-356.

[6] Schrick K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangl J, Schmidt J, Jürgens G. FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. *Genes and Development*, 2000, 14: 1471-1484.

[7] Stone S L, Kwong L W, Yee K M, Pelletier J, Lepiniec L, Fischer R L, Goldberg R B, Harada J J. *LEAFY COTYLEDON2* encodes a B3 domain transcription factor that induces embryo development. *Proceedings of the National Academy of Sciences of the USA*, 2001, 98(20): 11806-11811.

[8] Taishi U, Tetsuya S, Yasushi T, Atsushi T, Motoaki S, Atsushi I, Kenji A, Atsushi K, Takuhiro Y, Keiichi M, Mie K, Daisuke T, Kyonoshin M, Kazuo N, Akiko E, Saho M, Selina A, Kyoko Y, Kyuya H, Yasutaka T, Masaki H, Shusei S, Toyooki A, Masao I, Hideyuki F, Masayoshi T, Mitsuru O, Takuro S, Ryo A, Yoshiyuki S, Kazuko Y S, Kazuo S. Sequencing and analysis of approximately 40000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Research*, 2008, 15: 333-346.

[9] Shoemaker R, Keim P, Vodkin L, Retzel E, Clifton S W, Waterston R, Smoller D, Coryell V, Khanna A, Erpelding J, Gai X W, Brendel V, Christina R S, Shoop E G, Vielweber C J, Schmatz M, Pape D, Bowers Y, Theising B, Martin J, Dante M, Wylie T, Granger C. A compilation of soybean ESTs: generation and analysis. *Genome*, 2002, 45: 329-338.

[10] Jia J P, Fu J J, Zheng J, Zhou X, Huai J L, Wang J H, Wang M, Zhang Y, Chen X P, Zhang J P, Zhao J F, Su Z, Lü Y P, Wang G Y. Annotation and expression profile analysis of 2073 full-length cDNAs from

- stress-induced maize (*Zea mays* L.) seedlings. *The Plant Journal*, 2006, 48: 710-727.
- [11] Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li C J, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y. Collection, mapping, and annotation of over 28000 cDNA clones from *japonica* rice. *Science*, 2003, 301: 376-379.
- [12] Qi J L, Zhang W J, Liu S H, Wang H, Sun D Y, Xu G H, Shi M W, Liu Z, Zhang M S, Zhang H M, Yang Y H. Expression analysis of light-regulated genes isolated from a full-length-enriched cDNA library of *Onosma paniculatum* cell cultures. *Journal of Plant Physiology*, 2008, 165: 1474-1482.
- [13] Oghihara Y, Mochida K, Kawaura K, Murai K, Seki M, Kamiya A, Shinozaki K, Carninci P, Hayashizaki Y, Shin I T, Kohara Y, Yamazaki Y. Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. *Genes and Genetic Systems*, 2004, 79: 227-232.
- [14] Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K. Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science*, 2002, 296: 141-145.
- [15] Haas B J, Volfovsky N, Town C D, Troukhan M, Alexandrov N, Feldmann K A, Flavell R B, White O, Salzberg S L. Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biology*, 2002, 3(6): 0029. 1-0029. 12.
- [16] Chen L, Cao L X, Zhou L H, Jing Y D, Chen Z Z, Deng C, Shen Y, Chen L B. Trehalose as a good candidate for enriching full-length cDNAs in cDNA library construction. *Journal of Biotechnology*, 2007, 127: 402-407.
- [17] Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y. High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Research*, 1997, 4: 61-66.
- [18] Suzuki Y, Sugano S. Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods in Molecular Biology*, 2003, 221: 73-91.
- [19] Kato S, Ohtoko K, Ohtake H, Kimura T. Vector-capping: A simple method for preparing a high-quality full-length cDNA library. *DNA Research*, 2005, 12: 53-62.
- [20] Wellenreuther R, Schupp I, The German cDNA Consortium, Poustka A, Wiemann S. SMART amplification combined with cDNA size fractionation in order to obtain larger full-length clones. *BMC Genomics*, 2004, 5: 36.
- [21] Patanjali S R, Parimoo S, Weissman S M. Construction of a uniform-abundance (normalized) cDNA library. *Proceedings of the National Academy of Sciences of the USA*, 1991, 88: 1943-1947.
- [22] Carninci P, Shibata Y, Hayatsu N, Sugahara Y, Shibata K, Itoh M, Konno H, Okazaki Y, Muramatsu M, Hayashizaki Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Research*, 2000, 10: 1617-1630.
- [23] Zhulidov P A, Bogdanova E A, Shcheglov A S, Shagina I A, Vagner L L, Khazpekv G L, Kozhemiako V V, Lukyanov S A, Shagin D A. A method for the preparation of normalized cDNA libraries enriched with full-length sequences. *Russian Journal of Bioorganic Chemistry*, 2005, 31(2): 170-177.
- [24] Shagin D A, Rebrikov D V, Kozhemyako V B, Altshuler I M, Shcheglov A S, Zhulidov P A, Bogdanova E A, Staroverov D B, Rasskazov V A, Lukyanov A. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Research*, 2002, 12: 1935-1942.
- [25] Shoemaker R C, Schlueter J, Doyle J J. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology*, 2006, 9: 104-109.
- [26] Nelson R T, Shoemaker R. Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. *BMC Genomics*, 2006, 7: 204.