

• 药材与资源 •

千里光全长 cDNA 文库的构建及分析

平军娇, 张 珍, 蔡振锋, 汤贤春, 钱 刚*

遵义医学院 细胞生物学与遗传学教研室, 贵州 遵义 563099

摘 要: **目的** 构建千里光全长 cDNA 文库, 以期研究千里光的功能基因组学信息, 为克隆药理学性状相关的功能基因提供数据资源。**方法** Trizol 法提取千里光叶片总 RNA, 通过 SMART (switching mechanism at 5' end of RNA transcript) 构建全长 cDNA 文库, 随机挑取 600 个单克隆测序分析文库滴度、全长率及冗余率, 得到的 EST 序列进行 Blast 分析 (NR、NT、Swiss-Prot、KEGG) 及 COG 功能分类。**结果** 文库的库容为 4.3×10^6 cfu/mL, 插入片段大小平均 1.7 kb, 文库重组率 96.35%, 全长率 58.24%, 冗余率 10.88%; 获得 524 条全长 EST 序列, 含有 467 条独立基因 (unigenes), 其中 5 条序列与千里光次生代谢产物的合成、运输与代谢有关。**结论** 经检测, SMART 技术成功构建了千里光全长 cDNA 文库, 该文库可用于千里光功能基因组鉴定、新基因筛选及次生代谢产物生物合成的表达调控研究。

关键词: 千里光; 全长 cDNA 文库; SMART; EST; 基因组学

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2012)03-0557-05

Construction of a full-length cDNA library for *Senecio scandens*

PING Jun-jiao, ZHANG Zhen, CAI Zhen-feng, TANG Xian-chun, QIAN Gang

Department of Cell Biology and Genetics, Zunyi Medical College, Zunyi 563099, China

Abstract: Objective In the present study, our information from *Senecio scandens* full-length cDNA clones will serve as a useful resource for elucidating functional genes and will also aid a precise annotation of genomics in Compositae plants. **Methods** The total RNA was extracted from *S. scandens* using Trizol method. SMART (switching mechanism at 5' end of RNA transcript) was applied to constructing the full-length cDNA library. Titer of the library, full-length ratio, and redundancy rate for 600 monoclonal randomly selected sequencing library were evaluated by PCR amplification. NCBI and COG database was used to compare those sequences. **Results** Parameters of the the quality of cDNA library were as follows: the capacity of the library (4.3×10^6 cfu/mL), the average size of the inserted fragment (1.7 kb), the recombination rate (96.35%), the full-length rate (58.24%), and the redundancy rate (10.88%). EST sequences for 524 full-length were obtained in this study, involving 467 unigenes, among which five sequences associated with synthesis, transport, and metabolism of *S. scandens* secondary metabolites. **Conclusion** SMART technology could be effectively applied to constructing the full-length cDNA library of *S. scandens*. Our results indicate that the full-length cDNA library could be further applied to regulation of expression on identification of functional genome, screening of novel genes, and biosynthesis of metabolism genes in *S. scandens*.

Key words: *Senecio scandens* Buch.-Ham. ex D. Don; full-length cDNA library; SMART; EST; genomics

千里光 *Senecio scandens* Buch.-Ham. ex D. Don 为菊科 (Compositae) 千里光属多年生草本植物^[1], 叶片组织有清热解毒、杀虫、明目、凉血、生肌、驱风除湿等功效^[2]。其主要具有抗菌、抗钩端螺旋体、抗滴虫及抗氧化等药理活性, 对各种炎症性疾病及细菌性感染具有一定疗效^[3-5], 是一种临床疗效

显著、不良反应少的广谱抗菌中药^[6]。

cDNA 文库的构建已成为当前分子生物学研究和基因工程操作的基础以及研究功能基因组学的重要手段, 迄今, 拟南芥^[7]、水稻^[8]、小麦^[9]、北柴胡^[10]等重要模式植物的全长 cDNA 文库已成功构建并用于后期蛋白质表达及功能分析。近年来, 中药功能基

收稿日期: 2011-11-20

基金项目: 贵州省优秀科技教育人才省长基金项目 (黔省专合字 2008-61 号)

作者简介: 平军娇 (1987—), 女, 河北邯郸人, 在读硕士研究生, 研究方向为药用植物分子遗传学。

Tel: (0852)8609692 Fax: (0852)8609535 E-mail: pingjunjiao@126.com

*通讯作者 钱 刚 Tel: (0852)8609692 E-mail: qiangang69@yahoo.com.cn

因的研究,特别是药用植物次生代谢产物的生物合成基因调控研究已成为中药现代化研究的热点。

目前,何首乌、甘草、冬虫夏草、水母雪莲和青蒿等中药的 cDNA 文库已经成功构建,其中水母雪莲和青蒿为菊科中药,已有研究发现其次生代谢产物种类丰富。目前对于菊科中药有效成分的生物合成途径及基因调控的基础研究相对缺乏。千里光是一种具有良好开发前景的抗菌菊科中药,但国内外关于千里光的研究主要集中在化学成分、药理作用、分析方法、毒理、临床应用等方面,对于千里光功能基因组学方面的研究未见报道。本研究以千里光为材料,采用文库构建试剂盒 Creator SMART cDNA Library Constuction Kit 构建高质量的千里光叶片全长 cDNA 文库并进行 EST 序列分析,为进一步探讨千里光的分子遗传机制、功能基因组学、生物合成途径及次生产物代谢调节机制提供基础。

1 材料

千里光 *Senecio scandens* Buch.-Ham. ex D. Don 系本研究组采自贵州省遵义地区野生种,由遵义医学院细胞生物学与遗传学教研室钱刚副教授鉴定。取千里光幼苗的新鲜叶片置于液氮中保存备用。

总 RNA 提取试剂 Trizol 购自宝生物工程(大连)公司;文库构建试剂盒 Creator SMART cDNA Library Constuction Kit 及 Advantage 2 PCR Kit 购自 Clontech 公司;DH10B 感受态细胞购自 Invitrogen 公司。

2 方法

2.1 总 RNA 的提取及检测

将 0.1 g 千里光叶片组织用液氮研磨后加入 1 mL Trizol,静置 5 min 后离心,吸取上清,加氯仿剧烈震荡 15 s,待充分乳化后离心,吸取上清加入等体积异丙醇,充分混匀后静置 10 min,离心,在试管底部可见白色沉淀,75%乙醇清洗 2 次,最后用少量 DEPC 处理水溶解沉淀。采用分光光度计和 1%琼脂糖凝胶电泳检测其浓度、纯度和完整性。

2.2 cDNA 的合成

采用 Clontech 的 Creator SMARTTM cDNA 合成试剂盒合成一链和二链,引物 CDSIII/3 primer 改为 CDS-3M adapter (5'-AAGCAGTGGTATCAACG-CAGAGTGGCCGAGGCGGCC (T)₂₀VN-3', N=A、C、G、T; V=A、G、C)。一链 cDNA 合成体系总体积为 10 μ L,其中包括总 RNA 0.96 μ g,寡聚核苷酸 (SMART Oligonucleotide) 1 μ L, CDS-3M PCR 引物 1 μ L,一链合成缓冲液 (5 $\times$ First Strand

Buffer) 2 μ L,二硫苏糖醇 (DTT) (20 mmol/L) 1 μ L, dNTP Mix (10 mmol/L) 1 μ L 和反转录酶 PowerScriptTM Reverse Transcriptase 1 μ L,混匀后 42 $^{\circ}$ C 孵育 1 h,冰上终止第 1 链反应。

采用 LD-PCR 法以 cDNA 第 1 链为模板扩增获得双链 cDNA,PCR 扩增总体积为 100 μ L,其中包括一链 cDNA 2 μ L,去离子水 80 μ L,PCR 缓冲液 (10 $\times$ Advantage[®] 2 PCR 缓冲液) 10 μ L, 50 $\times$ dNTP Mix 2 μ L,PCR 引物 4 μ L 和聚合酶 (50 $\times$ Advantage 2 Polymerase) 2 μ L。PCR 反应条件为: 95 $^{\circ}$ C、1 min; 95 $^{\circ}$ C、7 s, 66 $^{\circ}$ C、20 s, 72 $^{\circ}$ C、4 min, 19 个循环。产物经 1%琼脂糖凝胶电泳检测。

2.3 双链 cDNA 纯化及短链 cDNA 去除

采用 QIAquick PCR Purification Kit (QIAGEN Cat.No.28104)进行双链 cDNA 纯化,终体积用 80 μ L ddH₂O 溶解。纯化的 cDNA 用 *Sfi* I Enzyme 酶切 (50 $^{\circ}$ C 水浴, 2 h),酶切产物采用 1%琼脂糖凝胶电泳检测。酶切产物通过试剂盒 QIAquick PCR Purification Kit 回收大片段 cDNA (1~3 kb),并检测。

2.4 cDNA 与载体的连接转化及菌落 PCR 鉴定

100 ng 的 cDNA 与 pDNR-LIB 质粒连接,连接产物经 0.25 μ mol/L Millipore 纯化膜上脱盐纯化 1 h,电击法转化到 50 μ L 大肠杆菌 DH10B 中,涂平板,37 $^{\circ}$ C 培养过夜,计算平板上的库容数及文库滴度;从文库中随机挑取 30 个克隆进行菌落 PCR,产物经 1%琼脂糖凝胶电泳检测。

2.5 单克隆 EST 测序及分析

从平板上随机挑选 600 个单克隆接种于 96 孔板上,采用 T7 引物从 cDNA 5' 方向测序 (华大基因科技股份有限公司)。序列通过 Cross_match 软件、Phrep 软件去除载体序列、核糖体及小于 100 bp 序列后,聚类拼接,得到所测千里光 EST 的相应 Unigene 数据,采用 Getorf 软件分析其开放阅读框并分析文库的全长率,冗余率。得到的 EST 序列进行 Blast 分析 (NR、NT、Swiss-Prot、KEGG)。用 BLASTP 将 Unigene 数据与 COG (Cluster of Orthologous Groups of Protein) 数据库 (<http://www.ncbi.nlm.nih.gov/COG>) 进行比对分析,获得千里光表达基因的功能注释及其 COG 功能分类。

3 结果

3.1 总 RNA 的分析

总 RNA 提取后,检测得出总 RNA 的质量浓度为 321 ng/ μ L, $A_{260}/A_{280}=1.98$, $A_{260}/A_{230}=2.09$,表

明总 RNA 纯度良好; 1%琼脂糖凝胶电泳结果表明总 RNA 完整性良好 (图 1), 可以用来建库。

3.2 cDNA 合成结果与分析

采用 LD-PCR 法扩增 cDNA 第一链以获得双链 cDNA, 检测结果显示 (图 2), 双链 cDNA 的长度在 250~5 000 bp 分布及含有中高丰度基因带, 表明 mRNA 得到有效的反转录和扩增。

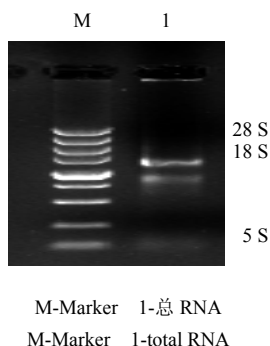


图 1 总 RNA 电泳结果

Fig. 1 Electrophoresis of total RNA

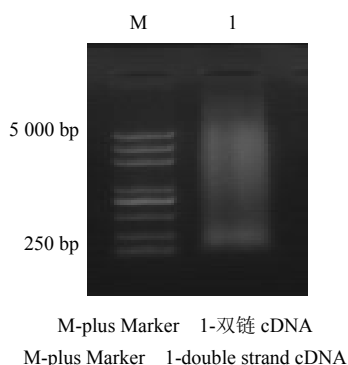


图 2 双链 cDNA 电泳结果

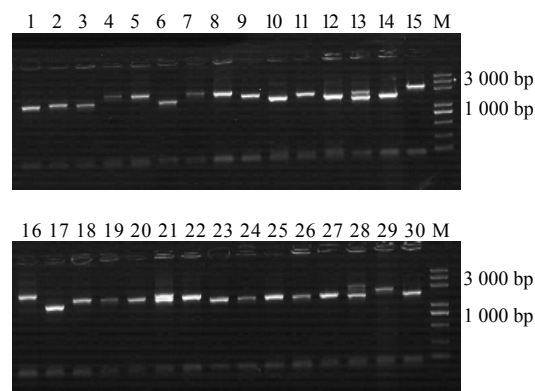
Fig. 2 Electrophoresis of double strand cDNA

3.3 cDNA 文库及 PCR 鉴定结果分析

经过计算分析文库滴度为 1.3×10^6 cfu/mL, 文库的库容大于 4.3×10^6 cfu/mL。如图 3 所示, 随机选取 30 个克隆的插入片段在 1.0~3.0 kb, 平均大小约为 1.7 kb, 满足全长 cDNA 文库构建要求。

3.4 单克隆 EST 测序结果及分析

从原始文库中随机挑选 600 个单克隆进行 5'端 EST 测序, 结果显示, 文库重组率高达 96.35%, 去除低质量的序列后得到 524 条高质量 ESTs 序列, 其平均长度为 1.7 kb。含有 467 条独立基因 (unigenes), 其中包括 43 个重叠群 (contigs) 和 424 个单拷贝基因 (singlets), 独立基因片段长度及其 ORF 长度如图 4 所示, 文库的冗余率为 10.88%, 全长率为 58.24%, 独立基因的大小分布见表 1。NR、



M-Marker 1~30-单克隆 PCR 结果
M-Marker 1—30-results of monoclonal PCR

图 3 全长 cDNA 文库插入片段大小鉴定

Fig. 3 Identification of inserted fragment length in full-length cDNA library

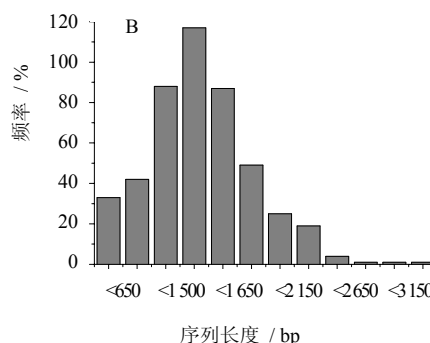
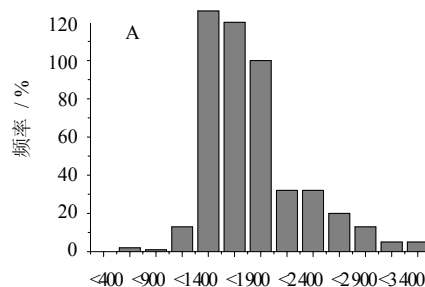


图 4 独立基因序列长度 (A) 及其 ORF 长度 (B)

Fig. 4 Sequence length of unigene (A) and its ORF (B)

NT、KEGG、Swiss-Prot COG 数据库 ESTs Blast 比对, EST 数目分别为 460、427、459、392、242, 通过计算得到 272 条全长序列。467 条千里光独立基因在 COG 数据库中有 242 个独立基因可获得生物学功能注释 (图 5)。

4 讨论

明确中药材药用品质成为制约中药现代化进程

表 1 独立基因的数目及比例
Table 1 Number of ESTs in unigene

簇大小 / 个	独立基因数目	独立基因比例 / %
1	424	90.79
2	34	7.28
3	6	1.28
4~5	2	0.43
6~10	1	0.21
11~20	0	0.00
21~50	0	0.00
51~100	0	0.00
>100	0	0.00

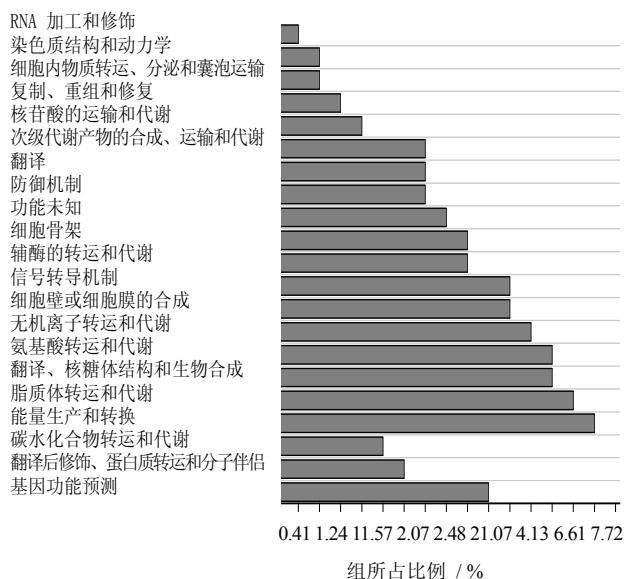


图 5 千里光表达基因的 COG 功能分类
Fig. 5 Functional classification of *S. scandens* expression genes

的瓶颈之一，药用植物活性成分及其功能基因研究成为我国药用生物资源可持续利用发展的一项最重要的基础研究^[11-12]。本研究得到的 cDNA 文库代表着千里光正在表达的有功能的基因，得到的 ESTs 序列不含内含子序列和调控序列，与氨基酸序列具有对应性，为深入了解菊科药用植物功能基因组序列信息提供基础资料，也为分离千里光的次生代谢基因、揭示药用品质的分子遗传机制奠定基础。cDNA 文库种类很多，经典 cDNA 文库存在克隆片段短等缺点，而全长 cDNA 文库能提供完整的 mRNA 信息^[13]，且只需通过 1 次阳性筛选，即可获得基因的全长序列，可最大程度地缩短获取全长基因所花费的时间^[14]。对于基因组庞大，近期内尚不能进行全基因组测序的中药千里光来说更是进行功

能基因组研究的有效途径。本研究结果表明该文库可用于基因鉴定、启动子分析和功能基因组研究，为快速、高效获得完整的基因全序列信息并进行相关功能的研究提供有效途径。

cDNA 文库的质量对于大规模 EST 序列测定至关重要，评价全长 cDNA 文库质量主要有两方面：即文库的代表性和插入序列的完整性，文库的代表性是指 cDNA 文库中所含的 mRNA 种类的完整性，即要有足够的克隆数使能够含有低丰度表达的 mRNA，可用文库滴度来衡量。一般来说，文库滴度能达到 10^5 cfu/mL 以上即为有效文库^[15]，过低则文库含有的组织内表达基因的全面性和代表性就会受到质疑，而且筛选到低丰度基因的概率也会小很多。而 LD-PCR 法合成的二链 cDNA，相比于普通的引物法增大了低丰度基因的表达率，因此本研究采用 LD-PCR 法构建了千里光的 cDNA 文库，文库的滴度为 1.3×10^6 cfu/mL，高于理论标准值，从中筛选低丰度目的基因在理论上具有可行性。完整性通常用文库的全长率和插入片段大小来衡量，cDNA 文库的构建方法主要有 Oligo-capping 法^[16]，CAPture 法^[17]，SMART 法^[18]，Cap-Select 法^[19]，Vector-Capping 法^[20-21]以及 Cap-jumping 法^[22]等，通常都是以 oligo (dT) 为引物，利用 mRNA 3'端的 poly (A) 尾巴，反转录合成 cDNA 的第一链。该方法的缺点在于若 mRNA 较长或者具有复杂二级结构，则转录得到的 mRNA 序列往往会存在 5'端的缺失情况，造成完整性降低。本研究采用 Clontech 公司的 SMART 技术^[23]能够克服这一不足，可避免因 mRNA 在分离纯化时的多步骤所造成 mRNA 的降解作用，可以有效保证 cDNA 的完整性，而且所获得的 cDNA 全长比例也较高^[24]。相对于普通 cDNA 文库，全长 cDNA 文库的优点主要在于提高全长 cDNA 的比例。本研究为了评价全长 cDNA 在文库中所占的比例，随机挑取 600 个单克隆进行完整性分析，通过序列比对与分析得到文库的全长率为 58.24%，插入片段大小平均约为 1.7 kb，保证了 cDNA 片段的完整性，满足了全长基因筛选的要求。

尽管全基因组测序是获知功能基因组信息的有效途径，但花费巨大，并需要辅以强有力的计算机软件乃至国际间的合作。鉴于表达基因只占整个基因组序列的 2%，因此，源于 mRNA 表达水平研究是进行功能基因筛选快捷、经济、有效手段^[25]。本研究利用构建千里光全长 cDNA 文库，并通过基因

组数据库已知序列信息比较, 获得了 524 条高质量 ESTs 序列, 242 条序列与 GenBank 上公布的序列有较高的同源性, 通过 COG 功能注释分类发现有 31 条序列参与了翻译后修饰, 蛋白质转移, 分子伴侣的功能, 有 28 条序列参与了碳水化合物运输和代谢, 有 17 条序列参与了能量产生和转换。千里光的药理学作用主要集中于千里光的次生代谢产物, 本课题组的结果表明, 2.07% 的序列参与了次生代谢产物的合成、运输和代谢, 这为研究千里光药理学性状功能基因奠定了基础; 而且, 288 条未知功能基因序列信息也为进一步探讨新的功能基因提供了序列信息。

参考文献

- [1] 陈录新, 马鸿雁, 张勉, 等. 千里光化学成分研究 [J]. 中国中药杂志, 2006, 31(22): 1872-1875.
- [2] 杨秀东, 冯乾坤, 李婷婷, 等. 麻叶千里光化学成分及药理作用研究 [J]. 长春中医药大学学报, 2006, 22(3): 70-71.
- [3] 张卫明, 赵伯涛, 马世宏, 等. 千里光在皮肤洗剂中的应用 [J]. 中国野生植物资源, 1995, 2: 481.
- [4] 张文平, 张瑞琪, 张文书, 等. 含千里光血清体外抗菌作用的研究 [J]. 江西医学检验, 2004, 22(6): 537-538.
- [5] 张文平, 张文书, 曾雪英. 千里光与甲氧苄氨嘧啶联用抗菌作用的研究 [J]. 时珍国医国药, 2006, 17(6): 944-945.
- [6] 全国中草药汇编编写组编. 全国中草药汇编 (上册) [M]. 北京: 人民卫生出版社, 1975.
- [7] Seki M, Caminci P, Nishiyama Y, *et al.* High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper [J]. *Plant J*, 1998, 15(5): 707-720.
- [8] Kikuchi S, Satoh K, Nagata T, *et al.* Collection, mapping, and annotation of over 28, 000 cDNA clones from japonica rice [J]. *Science*, 2003, 301(5631): 376-379.
- [9] Yasunari O, Keiichi M, Kanako K, *et al.* Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags [J]. *Genes Genet Syst*, 2004, 79(4): 227-332.
- [10] 隋春, 战晴晴, 魏建和, 等. 北柴胡皂苷生物合成途径关键酶 IPPI 的全长 cDNA 克隆及其序列分析 [J]. 中草药, 2010, 41(7): 1178-1184.
- [11] 吴春颖, 宋经元, 陈士林. 表达序列标签在药用植物研究中的应用 [J]. 中草药, 2008, 39(5): 778-782.
- [12] 王伟, 朱平, 程克棣. 药用植物基因组及 ESTs 研究 [J]. 中国生物工程杂志, 2004, 24(1): 1-5.
- [13] Suzuki Y, Yoshitomo N, Maruy A K. Construction and characterization of a full length-enriched and a 5'-end enriched library [J]. *Gene*, 1997, 200(1/2): 149-156.
- [14] Zhao D X, Song S H, Wang Q, *et al.* Discovery of immunerelated genes in Chinese mitten crab (*Eriocheir sinensis*) by expressed sequence tag analysis of haemocytes [J]. *Aquaculture*, 2009, 287: 297-303.
- [15] 李太武, 相建海, 刘瑞玉. 中国对虾 cDNA 文库的构建 [J]. 动物学报, 1998, 44(2): 237-238.
- [16] Maruyama K, Sugano S. Oligo-capping: A simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides [J]. *Gene*, 1994, 138: 171-174.
- [17] Edery I, Chu L L, Sonenberg N, *et al.* An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture) [J]. *Mol Cell Biol*, 1995, 15(6): 3363-3371.
- [18] Du X J, Wang J X, Liu N, *et al.* Identification and molecular characterization of a peritrophin-like protein from fleshy prawn (*Fenneropenaeus chinensis*) [J]. *Mol Immunol*, 2006, 43: 1633-1644.
- [19] Schmidt W M, Mueller M W. CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs [J]. *Nucl Acid Res*, 1999, 27(21): 31-39.
- [20] 韩慧霞, 王鲁, 陆洪光. 全长 cDNA 文库构建技术 Vector-Capping 法及其应用进展 [J]. 医学综述, 2010, 16(22): 3378-3380.
- [21] Kato S, Ohtake H, Kimura T. Vector-capping: a simple method for preparing a high-quality full-length cDNA library [J]. *DNA Res*, 2005, 12(1): 53-62.
- [22] Efimov V A, Chakhmakheva O G, Archdeacon J, *et al.* Detection of the 5'-cap structure of messenger RNAs with the use of the cap-jumping approach [J]. *Nucl Acid Res*, 2001, 29(22): 4751-4759.
- [23] Liu C Q, Lu T F, Feng B G, *et al.* Construction of cDNA library and preliminary analysis of expressed sequence tags from Siberian tiger [J]. *Int J Biol Sci*, 2010, 6(6): 584-589.
- [24] 徐峰, 武晓丽, 邱德文. 利用磁珠构建稻瘟菌 cDNA 文库 [J]. 微生物学报, 2008, 48(6): 806-810.
- [25] Schuler G D. Pieces of the puzzle: expressed sequence tags and the catalog of human gene [J]. *J Mol Med*, 1997, 75: 694-698.