

DOI: 10.3724/SP.J.1005.2010.00974

# 下一代测序 cDNA 样品制备及优化技术探讨

秦巧平, 张岚岚, 李南羿, 崔永一, 徐凯

浙江农林大学农业与食品科学学院, 临安 311300

**摘要:** 下一代测序技术目前已经应用于微生物、人类、动物、植物等的基因组分析。样品制备是开展大规模测序的必要前提和测序成功的根本保证。对大规模测序造成干扰的主要因素有: polyA 干扰测序信号及高丰度基因对低丰度基因的掩盖等。文章以堇菜(*Viola verecunda* A.Gray)叶片为试材, 提取总 RNA, 合成双链 cDNA, 利用 DSN 核酸酶对双链 cDNA 进行均一化处理, 并对双链 cDNA polyA 进行了切除, 将处理后的 cDNA 进行了 TA 克隆, 挑取 100 个克隆随机测序。结果表明, 未处理的 cDNA 样本测序有 15 个克隆由于 polyA 的存在而影响了附近碱基的正确阅读, 独立克隆只有 62 个, 而处理后的 cDNA 样本经测序未发现 polyA, 独立克隆有 94 个。序列分析发现, 经过处理的 cDNA 样本随机测序有两个克隆是经 MALDI-TOF 检测在样本中有蛋白质峰, 而基因克隆一直没有分离到的序列。以处理后的 cDNA 样本为模板扩增 2 个已知表达丰度差异较大的基因显示, 处理后这 2 个基因的 PCR 扩增产量差异明显减小。这些结果表明 polyA 切除、DSN 核酸酶处理的 cDNA 样本完全满足大规模测序、寻找新基因的要求。

**关键词:** cDNA; 下一代测序; 堇菜

## Optimizing of cDNA preparation for next generation sequencing

QIN Qiao-Ping, ZHANG Lan-Lan, LI Nan-Yi, CUI Yong-Yi, XU Kai

School of Agriculture and Food Science, Zhejiang Agriculture and Forestry University, Lin'an 311300, China

**Abstract:** Next generation sequencing has already been used for genomic analysis of microorganism, human being, animals, and plants. Sample preparation is prerequisite and most important for large-scale sequencing. There are two major interferences for large-scale sequencing, polyA and abundant genes' concealment for rare genes. In order to solve these problems, we used total RNA extracted from violaceae leaves to produce double stranded cDNA. DSN nuclease was used to treat the ds cDNA prior to removing the polyA. Randomly sequencing 100 clones of the treated cDNA showed that there were 94 independent clones in the treated sample, and the sequences did not contained polyA. However, only 62 independent clones were found in the untreated sample, and 15 of the sequencing files were affected by polyA. By randomly sequencing of the treated cDNA, we also found two clones encoded two interested genes. We failed to isolate these genes although the protein mass peaks of them had been found in the MALDI-TOF trace. Furthermore, we designed primers from two known genes with different expression abundances. The PCR yields were approaching similar using the treated cDNAs as templates. These results showed that, removal of the polyA and enrichment of rare genes with DSN can meet the requirements of large-scale sequencing and discovery of new genes.

**Keywords:** cDNA; next generation sequencing; *Viola verecunda* A.Gray

收稿日期: 2009-12-11; 修回日期: 2010-01-21

基金项目: 国家留学基金项目(编号: 2007106650), 国家自然科学基金项目(编号: 30771762), 浙江农林大学青年教师创新团队项目(编号: 2009RC02), 国家科技支撑计划子课题(编号: 2008BAD92B08-6), 林业科技支撑计划子课题(编号: 2006BAD03A15)和浙江省自然科学基金项目(编号: Y305222)资助

作者简介: 秦巧平(1975-), 女, 副研究员, 研究方向: 园艺植物分子生物学。E-mail: qinqp@zafu.edu.cn

通讯作者: 徐凯(1965-), 男, 教授, 研究方向: 园艺植物栽培生理。Tel: 0571-63741169; E-mail: xukai1965@sohu.com

获得未知基因组全序列、寻找新基因、识别突变和保守区域、探明重要代谢途径组件等是全基因组测序的主要目的,也是对基因功能开展研究的必要前提。下一代测序技术(Next generation sequencing)是近几年出现的一项革命性测序技术,利用其高通量高覆盖度的优势,可以在极短的时间内获取数百万甚至数十亿的序列数据信息。GS FLX 系统是由罗氏 454 生命科学公司推出的目前应用最广的超高通量基因组测序系统,已经应用于微生物、人类、动物、植物等的基因组分析<sup>[1~3]</sup>。

基因组 DNA、PCR 产物、BAC 文库、cDNA 等,其中以 cDNA 测序最为普遍。cDNA 是以 mRNA 为原料经体外反转录合成互补的 DNA, cDNA 分析是研究功能基因组学的基本手段之一。20 世纪末 21 世纪初,许多研究机构兴起构建 cDNA 文库,以便于保存资源和筛选目的基因,但是近几年科研人员发现,构建 cDNA 文库不仅费时费力,由于特定 cDNA 仅代表某一发育阶段表达出来的遗传信息,并不包含一个生物的完整遗传信息,且构建过程有多步 PCR 扩增、载体包装等,这些对基因定量表达分析完全不适合。罗氏公司推出的 GS FLX 系统无需构建 cDNA 文库,只需在双链 cDNA 两端连接特定的接头即可进行高通量测序。我们与公司测序部联系,他们认为目前测序技术日趋完善,而有的项目(如未知物种表达谱构建、寻找新基因)不成功的原因主要在于样品制备方面,即制备样品时未能去除 polyA,造成其对测序信号的严重干扰,另外,高丰度表达基因对低丰度表达基因的掩盖是造成寻找新基因失败或获得许多无意义大量重复序列的主要原因。

去除 polyA 的方法有 3 种:一种是用 RNase 酶消化反转录后的第一链 cDNA,然后进行第二链 cDNA 合成<sup>[4]</sup>;另一种是通过限制性内切酶将合成的双链 cDNA 中的 polyA 切除<sup>[5]</sup>;第三种方法是利用 PCR 对 polyA 结构进行突变<sup>[6]</sup>。但是,第一种方法较复杂,需要优化的参数较多,第三种方法由于增加了 PCR 步骤从而使克隆的准确性降低。第二种方法目前是 Roche 主要推荐的方法。本研究采用第二种方法并加以改进。均一化处理是近几年科研人员为了寻找低丰度表达基因而发展的一项技术,报道文献较少<sup>[4]</sup>。本研究以制备高质量、均一化 cDNA 用于大规模测序为目的,以堇菜(*Viola verecunda*

A.Gray)叶片为试材,提取总 RNA,合成双链 cDNA,利用 DSN 核酸酶对所合成的 cDNA 进行了均一化处理,并对逆转录合成的双链 cDNA polyA 进行了切除,将均一化后的 cDNA 进行了 TA 克隆,进行随机测序,结果表明,处理后的 cDNA 文库完全满足下一代测序系统的要求,在随机测序的 100 个克隆中,已经发现两个我们正在寻找的新基因。目前该样本已经委托罗氏公司开展大规模测序。

## 1 材料和方法

### 1.1 材料

实验材料取自培养在无菌培养室内的堇菜叶片,采摘后放入液氮中速冻后保存于-80℃超低温冰箱。

### 1.2 方法

#### 1.2.1 总 RNA 提取

采用 SDS 法提取总 RNA。3~5 g 叶片研磨后加入到 5 mL 匀浆液(150 mmol/L Tris (pH 7.5)、2% SDS、50 mmol/L EDTA)中,依次加入 0.25 体积乙醇、0.11 体积乙酸、1 体积氯仿/异戊醇(24:1),每一步均充分混匀,12 000 r/min 离心 30 min;取上清液用等体积酚/氯仿抽提一次;取上清加入 1/10 体积 3 mol/L 醋酸钠、1 体积异丙醇,-80℃放置 15 min,12 000 r/min 离心 30 min,沉淀溶于 5 mL 水中;加入等体积 4 mol/L LiCl,4℃过夜,12 000 r/min 离心 30 min,80%乙醇洗涤沉淀 1 次,沉淀风干后溶于 500  $\mu$ L 水中。总 RNA 质量用分光光度计和凝胶电泳进行检测。

#### 1.2.2 cDNA 合成

取 1  $\mu$ g 总 RNA 用 Evrogen MINT cDNA synthesis kit(Evrogen JSC, Russia)进行逆转录。为了去除 polyA 对大规模测序的影响,对试剂盒中的反转录引物 oligod(T)<sub>20</sub> 进行改造,加入 *Gsu* 酶切位点: AAGCAGTGGTATCAACGCAGAGTCTGGAGTTTT TTTTTTTTTTTT TTTTVN(下划线表示 *Gsu* 酶切位点)。以合成的 cDNA 第一链为模板,通过 PCR 反应扩增得到双链 cDNA。反应结束后,取 5  $\mu$ L PCR 产物,1.5%琼脂糖凝胶电泳检测 cDNA 质量。

#### 1.2.3 cDNA 的均一化处理

将通过 Evrogen MINT cDNA synthesis kit 合成的双链 cDNA 经 QIAquick PCR Purification

Kit(QIAGEN, Germany)进行纯化后,用 98%乙醇沉淀回收双链 cDNA,定量分析后参考 Evrogen Trimmer cDNA Normalization kit(Evrogen JSC, Russia)试验手册进行参数优化,实验方法略作改进。DSN 稀释为 1/1、1/2、1/4、1/8,取 12  $\mu$ L cDNA,加入 4  $\times$  Hybridization buffer,在 PCR 仪上 98  $^{\circ}$ C 杂交 2 min,之后在 68  $^{\circ}$ C 进行再杂交 5 h,加入 5  $\mu$ L 预热的 DSN master buffer 及未稀释和稀释过的 DSN 核酸酶,68  $^{\circ}$ C 均一化处理 15~20 min,加入 10  $\mu$ L 5 mmol/L EDTA 终止反应。处理后的 cDNA 保存于-20  $^{\circ}$ C。

#### 1.2.4 均一化后 cDNA 的扩增

以 1  $\mu$ L DSN 处理后的 cDNA 作模板,加入 40.5  $\mu$ L ddH<sub>2</sub>O、5  $\mu$ L 10  $\times$  PCR Buffer、1  $\mu$ L 50  $\times$  dNTP Mix、1.5  $\mu$ L Evrogen PCR primer M1、1  $\mu$ L 50  $\times$  Polymerase Mix, 95  $^{\circ}$ C 7 s、66  $^{\circ}$ C 20 s、72  $^{\circ}$ C 4 min,分别进行 7、9、11、13 个循环,1.5%琼脂糖凝胶电泳检测确定最佳循环数。以扩增最佳的第 1 次 PCR 产物 2  $\mu$ L 为底物进行第 2 次 PCR 扩增,加入 80  $\mu$ L ddH<sub>2</sub>O、10  $\mu$ L 10  $\times$  PCR Buffer、2  $\mu$ L 50  $\times$  dNTP Mix、4  $\mu$ L Evrogen PCR primer M2、2  $\mu$ L 50  $\times$  Polymerase Mix, 95  $^{\circ}$ C 7 s、66  $^{\circ}$ C 20 s、95  $^{\circ}$ C 7 s、64  $^{\circ}$ C 20 s、72  $^{\circ}$ C 4 min,扩增 12 个循环,取 5  $\mu$ L 进行 1.5%琼脂糖凝胶电泳检测。扩增后的双链 cDNA 用 *Gsu* 进行单酶切,酶切产物直接用于质量检测。

#### 1.2.5 均一化 cDNA 的检测

分别取 1  $\mu$ L 未均一化和均一化后的 cDNA,与 pGEMT Easy 载体连接,转化 TOP10 感受态细胞,分别随机挑取 100 个克隆,提取、纯化质粒,以 M13+为引物进行测序。基因序列分析采用 BLAST、CLUSTLX 等进行。同时,以已知的高丰度表达基因和低丰度表达基因设计引物,分别以未均一化和均一化后的 cDNA 为模板进行 PCR 扩增。

#### 1.2.6 样品的 MALDI-TOF 检测

叶片用 50% (v/v) ACN(含 0.05% TFA)研磨,离心去除残留物,上清液冷冻干燥。干燥物用 0.05% TFA 溶解,脱盐,用 C18 Ziptip (Millipore)进行样品浓缩。样品与 CHCA 等量混合,点于 2 $\times$ 96 平板,空气中干燥后用氮气横扫以去除表面附着杂质。MALDI-TOF mass 分析用 Voyager-DE STR mass spectrometer (Ap-

plied Biosystems)进行,参数为:加速电压为 20 kV,延时 165 ns。Data Explorer 4.0 软件用于数据分析及处理。

## 2 结果与分析

### 2.1 全长 cDNA 的合成

1.5%琼脂糖凝胶电泳检测显示,莴菜叶片双链 cDNA 的长度分布于 500~4 000 bp,中间有若干较亮条带,表明不同大小和丰度的 mRNA 都得到了有效的反转录及扩增,在 100 bp(引物二聚体、多聚体)附近无亮带,表明引物聚合物较少(图 1)。

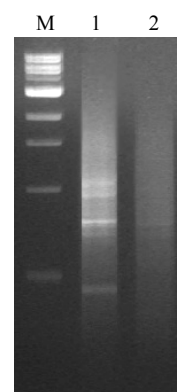


图 1 莴菜双链 cDNA 经 1.5%琼脂糖凝胶电泳检测结果  
M: 1 kb DNA ladder; 1: 未均一化的双链 cDNA; 2: 均一化后的双链 cDNA。

### 2.2 均一化效果检测

将双链 cDNA 进行 DSN 处理,反应完成后进行第 1 次 PCR 和第 2 次 PCR 扩增,扩增结果显示,经 DSN 处理 25 min 后,双链 cDNA 较长片段明显减少,而 500~1 000 bp 左右的片段明显增加,处理 20 min 可以显著改善这一现象,处理 15 min 双链 cDNA 分布情况基本与未处理相同,与对照相比,代表高丰度基因的亮带基本消失(图 1)。

以 2 个已知表达丰度的基因(A1, A2)设计引物进行 PCR 扩增,反应体系中加入等量 cDNA 模板。结果显示,扩增 20 个循环后,以未处理双链 cDNA 为模板进行扩增,这两个基因扩增后的 A1 PCR 产量明显大于 A2;而以处理双链 cDNA 为模板进行 PCR 扩增,这两个基因的 PCR 产量基本一致。这表明,处理后 cDNA 中高丰度表达基因明显降低,低丰度表达基因已经得到富集(图 2)。

### 2.3 随机测序分析

将 DSN 处理后的双链 cDNA 进行 *Gsu*I 酶切,以

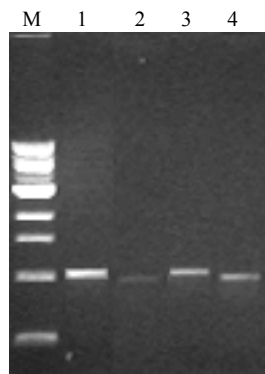


图2 均一化后双链 cDNA 的 PCR 扩增检测

M: 1 kb DNA ladder; 1: 高丰度表达基因以未处理双链 cDNA 为模板; 2: 低丰度表达基因以未处理双链 cDNA 为模板; 3: 高丰度表达基因以处理后双链 cDNA 为模板; 4: 低丰度表达基因以处理后双链 cDNA 为模板。

未处理 cDNA 为对照, 酶切后分别取 1  $\mu$ L cDNA 与 pGEMT Easy 载体连接, 转化 TOP10 感受态细胞, 分别随机挑取 100 个克隆, 提取、纯化质粒, 以 M13+ 为引物进行测序。结果显示, 对照 100 个随机克隆中独立克隆只有 62 个独立序列, 测序发现 15 个克隆由于 polyA 的存在而影响了附近碱基的正确阅读, 而处理后的 cDNA 样本测序未发现 polyA, 独立克隆有 94 个。

序列分析发现经过处理的 cDNA 样本随机测序有两个克隆是经 MALDI-TOF 检测在样本中有蛋白质峰, 而基因克隆一直没有分离到的序列(图 3)。

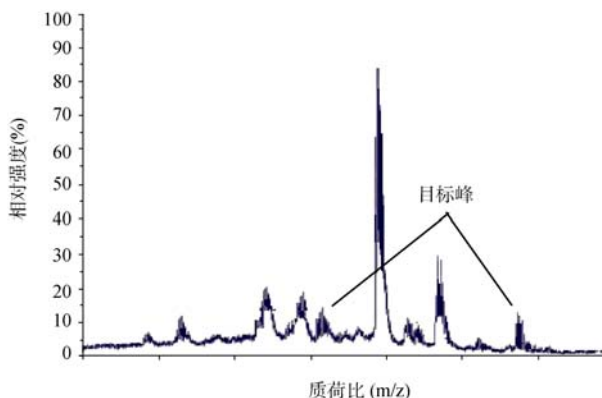


图3 样品 MALDI-TOF MS 分析

### 3 讨论

本研究以为准备大规模测序样品为目的, 开展了高质量双链 cDNA 制备及优化技术探讨。我们利用 Evrogen JSC 公司提供的技术方案进行了董菜 cDNA 合成及均一化处理, 结果显示, 经本实验合

成及优化的双链 cDNA 完全可以满足 GS FLX 系统测序要求。此外, 在过去很长时间的实验中我们发现, 由于两个目的蛋白质含量偏低(图 3), HPLC 难以进行分离纯化, 不能满足蛋白质测序要求, 因而一直没有获得这两个蛋白质的准确序列。而在本实验中, 经过均一化后, 随机测序 100 个克隆中已经发现两个目的基因编码目标蛋白, 这表明该 cDNA 样品完全可以满足我们寻找新基因的需要。我们之前用 Clontech 公司的 SMART cDNA 合成试剂盒合成双链 cDNA 中引物多聚体等非特异性片段较多, 而用 MINT 合成的 cDNA 则质量较高, 且 cDNA 弥散程度较 SMART 合成的大, 这为寻找更多基因提供了优势。同时, 本研究也为其他植物开展大规模测序前的样品制备提供了借鉴。

### 参考文献(References):

- [1] Tangphatsornruang S, Somta P, Uthapaisanwong P, Chanprasert J, Sangsakru D, Seehalak W, Sommanas W, Tragoonrungs S, Srinives P. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biol*, 2009, 9(1): 137.
- [2] Steuernagel B, Taudien S, Gundlach H, Seidel M, Ariyadasa R, Schulte D, Petzold A, Felder M, Graner A, Scholz U, Mayer KF, Platzer M, Stein N. *De novo* 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. *BMC Genomics*, 2009, 10(1): 547.
- [3] Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*, 2009, 10(1): 399.
- [4] Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acid Res*, 2004, 32(3): e37.
- [5] Shibata Y, Carninci P, Sato K, Hayatsu N, Shiraki T, Ishii Y, Arakawa T, Hara A, Ohsato N, Izawa M, Aizawa K, Itoh M, Shibata K, Shinagawa A, Kawai J, Ota Y, Kikuchi S, Kishimoto N, Muramatsu M, Hayashizaki Y. Removal of polyA tails from full-length cDNA libraries for high-efficiency sequencing. *Biotechniques*, 2001, 31(5): 1042, 1044, 1048–1049.
- [6] Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME, Robinson GE. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science*, 2007, 318(5749): 441–444.

## •综合信息•

## 2011 年部分生物、农林类学术期刊联合征订表

刊物名称	邮发 代号	刊 期	年价 (元)	期刊网址	编辑部 E-mail
癌变·畸·突变	80-285	双月刊	60	www.egh.net.cn	cemsctm@stu.edu.cn
动物学研究	64-20	双月刊	150	www.zoores.ac.cn	zoores@mail.kiz.ac.cn
动物学杂志	2-422	双月刊	360	http://dwzzz.ioz.ac.cn	journal@ioz.ac.cn
分子植物育种	84-23	双月刊	240	www.molplantbreed.org	mpb@hibio.org
国际遗传学杂志	14-55	双月刊	90	www.cma.org.cn	genetics@ems.hrbmu.edu.cn
激光生物学报	42-194	双月刊	150	www.jgswxb.net	jgswxb@hunnu.edu.cn
菌物学报	2-499	双月刊	480	http://journals.im.ac.cn/jwxtcn	jwxt@im.ac.cn
昆虫知识	2-151	双月刊	360	www.ent-bull.com.cn	entom@ioz.ac.cn
林业科学	82-6	月 刊	300	www.linyekexue.net	linyhx@forestry.ac.cn
农业生物技术学报	2-367	双月刊	240	www.jabiotech.org.cn/	nsjxb@cau.edu.cn
人类学学报	2-384	季 刊	100	www.ivpp.ac.cn	acta@ivpp.ac.cn
生命科学	4-628	月 刊	480	www.lifescience.net.cn	cbbs@sibs.ac.cn
生命科学研究	42-172	双月刊	108	http://smky.chinajournal.net.cn	life@hunnu.edu.cn
生物工程学报	82-13	月 刊	780	http://journals.im.ac.cn/cjbcn	cjb@im.ac.cn
生物化学与生物物理进展	2-816	月 刊	720	www.pibb.ac.cn	prog@sun5.ibp.ac.cn
生物技术通报	18-92	月 刊	300		biotech@mail.caas.net.cn
生物技术通讯	82-196	双月刊	150	http://swtx.chinajournal.net.cn	swtx@263.net
微生物学报	2-504	月 刊	660	http://journals.im.ac.cn/actamicrocn/	actamicro@im.ac.cn
微生物学通报	2-817	月 刊	576	http://journals.im.ac.cn/wwxtbcn	tongbao@im.ac.cn
武汉植物学研究	38-103	双月刊	180	http://whzwxxyj.cn	editor@rose.whiob.ac.cn
畜牧兽医学报	82-453	月 刊	360	www.xmsyxb.com	xmsyxb@263.net
遗传	2-810	月 刊	600	www.chinagene.cn	yczz@genetics.ac.cn
遗传学报	2-819	月 刊	600	www.jgenetgenomics.org	jgg@genetics.ac.cn
云南植物研究	64-11	双月刊	150	http://journal.kib.ac.cn	bianji@mail.kib.ac.cn
植物遗传资源学报	82-643	双月刊	120	www.zwyczy.cn	zwyczyxb2003@163.com
植物学报	2-967	双月刊	480	www.chinbullbotany.com	cbb@ibcas.ac.cn
中国实验动物学报	2-748	双月刊	120	www.calas.org.cn	A67761337@126.com
中国生态农业学报	82-973	双月刊	210	www.ecoagri.ac.cn	editor@sjziam.ac.cn
中国生物工程杂志	82-673	月 刊	960	www.biotech.ac.cn	biotech@mail.las.ac.cn
中国水产科学	18-250	双月刊	180	www.fishscichina.com	zgscckx@cafs.ac.cn
中国水稻科学	32-94	双月刊	120	www.ricesci.cn	cjrs@263.net
作物学报	82-336	月 刊	600	www.chinacrops.org/zwxb	xbzw@chinajournal.net.cn