

# 几种全长 cDNA 文库构建方法比较

毛新国, 景蕊莲, 孔秀英, 赵光耀, 贾继增

(中国农业科学院作物科学研究所, 农作物基因资源与基因改良国家  
重大科学工程, 农业部作物种质资源与生物技术重点开放实验室, 北京 100081)

**摘要:** 全长 cDNA 文库是高效、大规模获得基因全序列信息的一条有效途径, 尤其是对基因组庞大、近期内尚不能进行全基因组测序的生物来说, 是一条开展功能基因组研究的重要途径。文章对几种全长 cDNA 文库的构建方法进行了概述, 对各种方法的原理及优缺点做了分析、比较, 并结合实验室的结果, 重点介绍了 Cap-trapper 法在小麦全长 cDNA 文库中的应用及文库中全长基因比例判定方法。

**关键词:** 全长 cDNA 文库; Cap-trapper 法; Capture 法; Oligo-capping 法; Cap-jumping 法; SMART 法

中图分类号: Q75

文献标识码: A

文章编号: 0253-9772(2006)07-0865-09

## Comparison of Methods to Construct a Full-Length cDNA Library

MAO Xin-Guo, JING Rui-Lian, KONG Xiu-Ying, ZHAO Guang-Yao, JIA Ji-Zeng

(National Key Facility for Crop Gene Resources and Genetic Improvement, Key Laboratory of Crop Germplasm & Biotechnology / the Ministry of Agriculture, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

**Abstract:** The use of full-length cDNA libraries is an effective tool to obtain complete gene information in a high-efficiency, high-throughput manner, especially in organisms with huge genomes that are not amenable to whole genome sequencing. In this review, we outlined several methods of full-length cDNA library construction and compared their advantages and disadvantages based on their respective principles. Drawing on our own experience, we described the Cap-trapper method in detail, with an emphasis on its application in wheat full-length cDNA library construction as well as the determination of the ratio of full-length cDNA in a library.

**Key words:** Full-length cDNA library; Cap-trapper method; Capture method; Oligo-capping method; Cap-jumping method; SMART method

测序技术的不断完善和计算机技术的快速发展, 使得一大批模式生物和人类基因组全序列测定得以提前完成, 也随之产生了大量核酸序列数据信息, 使从基因组水平了解生物的生长、发育、分化以及各种生理和病理过程成为可能。由于生物基因组的复杂性, 不同生物基因组在结构、组成以及大小上的多样性, 以及人们对基因转录及转录后的拼接、加

工、修饰方式的机制缺乏深入了解, 所以完全依靠生物信息学的方法来预测基因及其功能尚存在一定困难。此外, 采用生物信息学的方法预测基因的功能经常会出现错误<sup>[1]</sup>。一般来讲, 基因预测软件对外显子边界预测的准确率约为 80%, 如果一个基因含有 5 个外显子, 那么基因预测的准确率仅有 33%, 且外显子的数量越多, 准确率也就越低。就目前的

收稿日期: 2005-07-05; 修回日期: 2005-08-16

基金项目: 国家重点基础研究发展规划(973 计划)项目(编号: 2004CB1172)[Supported by Key Project of Chinese National Programs for Fundamental Research and Development (973 program)(No. 2004CB1172)]

作者简介: 毛新国(1971—), 男, 博士, 研究方向: 小麦功能基因组学。E-mail: maoxg2001@sina.com

通讯作者: 贾继增(1945—), 男, 研究员, 研究方向: 植物遗传学。Tel: 010-62186623; E-mail: jzjia@mail.caas.net.cn

基因预测软件而言,如 Genescan<sup>[2]</sup>、Genemark.HMM<sup>[3]</sup>、GlimmerM<sup>[4]</sup>等,它们只能对基因的开放阅读框(ORF)进行预测,无法对基因两端的非编码区(UTR)进行预测,而许多基因的非编码区与基因的功能非常密切,所以使用基因预测软件分析得到的结果,其准确性、可信度及其利用价值一直倍受争议。就研究现状而言,要真正实现从基因组水平了解生命活动的机理还有很长的路要走。

全长 cDNA 文库的出现在客观上促进了人们对基因结构和功能的认识。所谓全长 cDNA 是指那些不仅包含完整的阅读框架,还拥有 5' 和 3' 端非编码区的 cDNA。全长 cDNA 文库的优点非常明显,其绝大多数克隆是全长的,这不仅能大大提高基因测序和生物信息学分析的进程,还利于后期蛋白质表达及功能分析<sup>[5]</sup>。此外,全长 cDNA 文库是高效、大规模获得基因序列信息的一条有效途径,尤其是对基因组庞大,近期内尚不能进行全基因组测序的生物来说更是进行功能基因组研究的一条重要途径。即使在模式生物,如拟南芥、水稻、美丽线虫等全基因组测序完成后,分子生物学家们仍旧构建了相应生物的全长 cDNA 文库,并进行了大规模的全长 cDNA 测序,以深入了解基因的功能。然而,构建高质量全长 cDNA 文库仍旧不是一件非常容易的事情。首先,获取完整的 mRNA 比较困难;其次,全长 cDNA 也不易得到;第三,有效区分全长和截短的基因非常棘手;第四,由于基因序列长短不同,克隆生长快慢也不一致,很容易导致小片段基因的富集。而基于普通 cDNA 文库中的基因片段来获取全长基因的方法(如 RACE 技术)在整个基因组范围内大规模、高通量地发掘新基因是难以想象的。

构建全长 cDNA 文库已经成为科学家们关注的焦点,许多重要的模式生物全长 cDNA 文库已经构建,如:拟南芥(*A. thaliana*)<sup>[6]</sup>,小鼠(*M. musculus*)<sup>[7]</sup>,果蝇(*D. melanogaster*)<sup>[8]</sup>,水稻(*O. sativa*)<sup>[9]</sup>等等,产生了大量有价值的数据,由此科学家们也取得了很多全新的研究成果,极大地促进了功能基因组学研究。

## 1 几种全长 cDNA 文库构建方法

有关全长 cDNA 文库构建的报道最早出现于 1994 年<sup>[10]</sup>。目前,在全长 cDNA 文库构建方面已经取得了很大的进展,全长 cDNA 文库的构建方法也

不断涌现,其中主要有:CAPture 法<sup>[11]</sup>、Oligo-capping 法<sup>[12,13]</sup>、SMART 法<sup>[14]</sup>、Cap-jumping 法<sup>[15]</sup>以及 Cap-trapper 法<sup>[5,6,16,17]</sup>等。所有的这些方法都着眼于真核生物 mRNA 5' 端的帽子结构,但又各具独到之处。总的来说,每种方法都存在一定的缺陷,它们有的涉及 PCR 扩增,极易改变文库中克隆的代表性,并影响难扩增基因的克隆;有的以质粒为载体,不利于大片段基因的克隆;有的实验流程长,步骤繁琐。

### 1.1 CAPture 法

CAPture 法(mRNA Cap Retention Procedure)最早由 Edery I<sup>[11]</sup>提出,它充分利用真核生物 mRNA 的帽子结构和帽子结合蛋白(转录起始因子 eIF-4e)相互作用的动力学原理来捕获全长 cDNA。首先,在反转录酶的作用下将 mRNA 转录为 cDNA,形成 cDNA/mRNA 双链复合体;接着,用 RNaseA 对 cDNA/mRNA 双链分子进行酶切。如果反转录不彻底,cDNA 没有延伸到 mRNA 的帽子结构部位,那么靠近 mRNA 5' 端的 mRNA 将以单链形式存在,这种情况下,RNaseA 就能将这类 mRNA 的帽子结构切除掉,因此这类 cDNA/mRNA 双链复合体也就不再携带帽子结构。然后,利用帽子结合蛋白结合、分离全长 cDNA(帽子结合蛋白结合在 eIF-4e 葡聚糖抗体上)(图 1);最后,根据帽子结合蛋白与 m<sup>7</sup>GDP 结合的动力学原理,将帽子结合蛋白上结合的 mRNA 帽子替换下来,从而达到洗脱全长 cDNA 的目的。

从原理上分析,这种方法非常巧妙,但在实际操作过程中却遇到一些的问题。首先,该方法采用 RNaseA 来酶切 cDNA/mRNA 复合体,以除去短截的(没有完全转录)复合体中 mRNA 5' 端的帽子结构。RNaseA 具有碱基偏爱性,它能够有效切割富含嘧啶的单链 RNA,而对嘌呤含量较高的 mRNA 消化效率却很低,甚至不能切割。一般来说,mRNA 的 5' 端 G+C 含量普遍较高,尤其在 5' 端非编码区,这种现象更为明显,也正是由于这种原因,致使短截 cDNA 掺入到全长 cDNA 中,导致全长 cDNA 在文库中的比例下降。其次,经过 RNaseA 消化后的小片段 RNA 依旧存在于反应体系中,尤其是切割后形成的小分子帽子结构,它会干扰帽子结合蛋白对全长 cDNA 的捕获。第三,它利用 m<sup>7</sup>GDP 来取代被结合的 mRNA 帽子,无法完全避免短截 cDNA 的掺入。第四,这种方法对 mRNA 的需求量特别大(100 μg

左右),实际操作也相当繁琐。由于以上缺陷,致使这种方法构建的文库中全长 cDNA 的比例不是很高(60%~70%)。

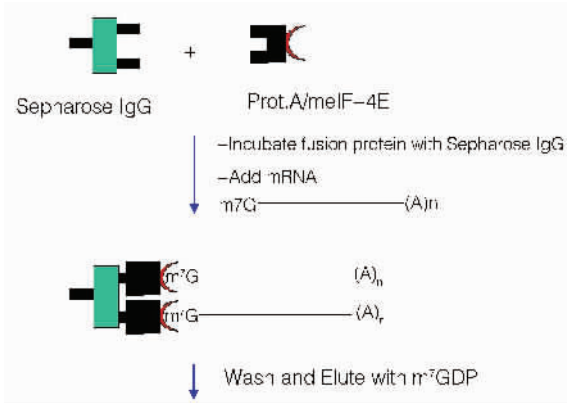


图 1 mRNA 帽子置换原理(参考文献[11])

Fig.1 Principle of “cap” replacement(Reference[11])

### 1.2 Oligo-capping 法

Oligo-capping 法 (Oligo-capping method) 最早由 Suzuki Y 等<sup>[12]</sup> 提出,其实验流程如图 2 所示。首先,利用细菌碱性磷酸酶 (Bacterial alkaline phosphatase BAP) 水解 5' 端不完整的 mRNA 的 5' 磷酸基团,防止短截的 mRNA 在后续反应中与寡聚核糖核

酸连接;接着,用烟草酸焦磷酸酶 (Tobacco acid pyrophosphatase TAP) 除去 mRNA 5' 端的帽子结构,使 mRNA 5' 端帽子结构处的磷酸基团暴露出来;然后,用 T4 RNA 连接酶在 mRNA 的 5' 端连上一段寡聚核糖核酸,作为引发第二链 cDNA 合成的引物结合位点,最后经反转录,PCR 扩增、酶切、连接,建成目的文库。

Oligo-capping 法构建 cDNA 文库涉及多种酶促反应,酶的效率将直接影响文库的最终质量。实验中所用的 T4 RNA 连接酶对文库的构建至关重要,通常 RNA 连接酶的连接效率没有 DNA 连接酶高。此外,该方法涉及 PCR 扩增,而 PCR 反应常常会由于模板 DNA 的长度不同,G+C 碱基的含量高低不一致以及模板自身的二级结构等因素的影响,导致扩增产物之间比例失调,甚至有些基因由于自身结构比较复杂,无法用 PCR 的方法进行扩增。以上不利因素都会影响稀有表达基因的发掘。

为了提高全长 cDNA 在文库中的比例,Clepet C 等<sup>[18]</sup> 对 Oligo-capping 法做了一些改进,用 T4 DNA 连接酶取代 T4 RNA 连接酶,在一定程度上提高了寡核苷酸 mRNA 的连接效率。

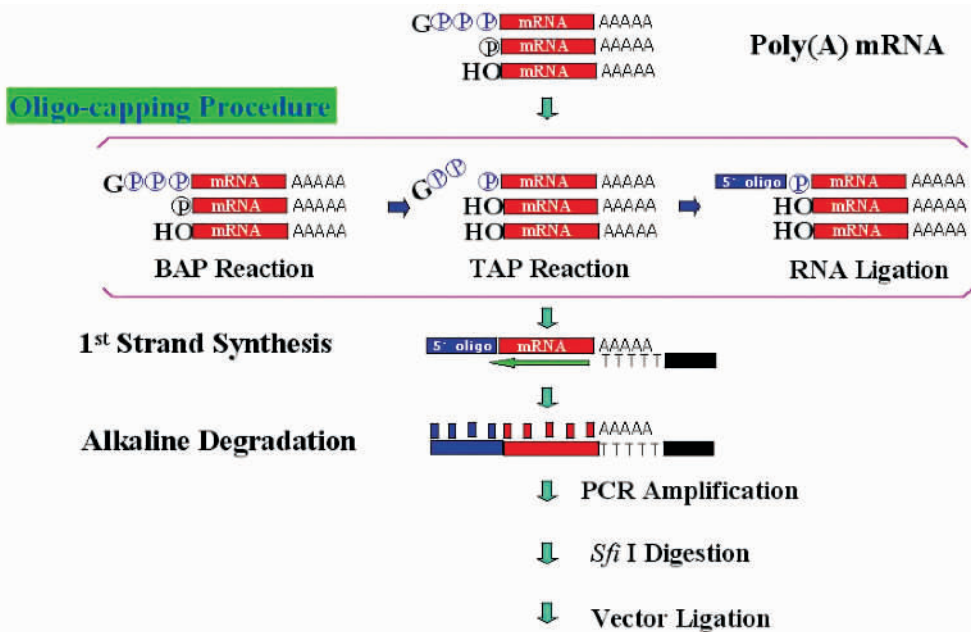


图 2 Oligo-capping 法原理(参考文献[19])

Fig.2 Schematic diagram of “Oligo-capping” method(Reference [19])

### 1.3 SMART 法

SMART 法 (Switching Mechanism At 5' end of RNA Transcript method/ SMART method)<sup>[14]</sup> 充分利用了反转录酶 Powerscript™ RT 和限制性内切酶 *Sfi* I 的特性 (流程如图 3)。该方法中全长 cDNA 的获得是借助 Powerscript™ RT 的末端转移酶活性来实现的。当反转录达到 mRNA 的 5' 端时, Powerscript™ RT 就能够在双链核酸的 3' 端添加几个脱氧胞嘧啶 (dC), 而对于非全长 cDNA, 由于反转录延伸没有达到 mRNA 的 5' 端, Powerscript™ RT 不能在其不完整的 3' 末端加上 dC。在 cDNA 第二链合成时, 3' 端携带 oligo(dG) 的第二链引物也就不能与短截的 ss-cDNA 结合, 因而这类 cDNA 不能合成互补链, 最终得到的 dsDNA 都是全长的。此外, 在实验设计时, 在 cDNA 第一、二链引物的 5' 端引入了 *Sfi* I

(A) 和 *Sfi* I (B) 位点, 因此只需对目的 cDNA 进行单酶切 (*Sfi* I 酶切), 即可实现对其定向克隆。

SMART 法设计很巧妙, 但也存在一定的缺陷。由于相当多的 cDNA 内部存在寡聚 dC, 而这种方法在第二链 cDNA 合成时复性温度不是很高, 有些存在于基因内部的寡聚 dC 有机会与末端携带几个 dG 的第二链引物退火, 从而在 cDNA 的内部引发第二链 cDNA 的合成, 导致文库中短截 cDNA 的比例升高。此外, 由于 cDNA 的 5' 端 G+C 含量较高, 所以这类事件发生的几率也大大提高, (大规模测序的结果也证明了这一点)。国内有不少实验室采用这种方法构建 cDNA 文库, 但全长 cDNA 的比例都不是很高。最近 Wellenreuther 等将琼脂糖电泳分级引入 SMART 法中来构建全长 cDNA 文库, 以提高文库中全长 cDNA 的比例<sup>[20]</sup>。

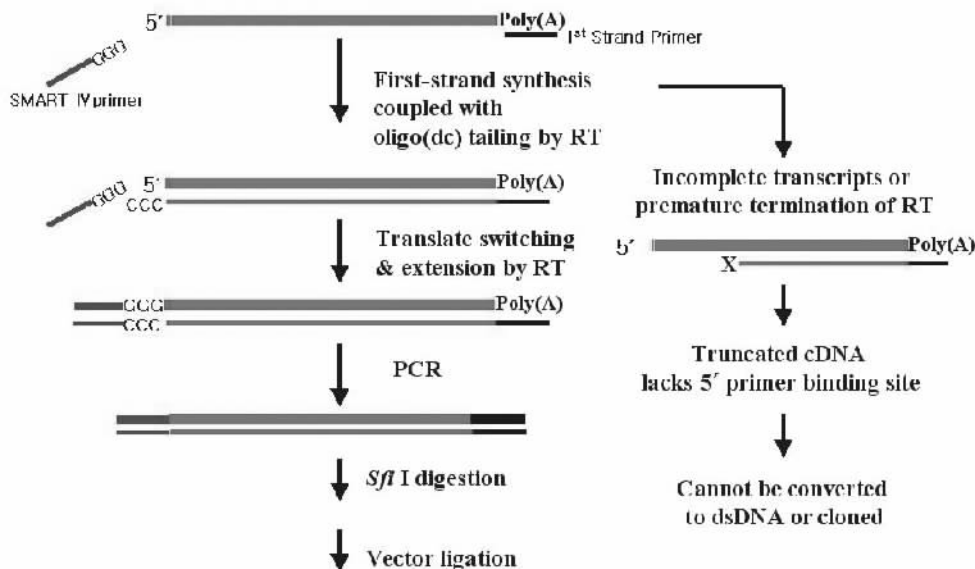


图 3 SMART 法构建全长 cDNA 文库流程 (参考文献 [21])

Fig. 3 Flowchart of full-length cDNA construction with SMART method (Reference [21])

### 1.4 Cap-trapper 法

Cap-trapper 法 (Cap-trapper method) 最早由 Carninci 等<sup>[5]</sup> 提出, 经不断改进, 现已日臻完善<sup>[6,16,17]</sup>。这种方法在全长 cDNA 获得方面采取了以下措施。首先, 在全长 cDNA 合成时, 为了得到尽可能多的全长的 cDNA, 向反应体系中加入海藻糖和山梨(糖)醇, 这两种物质都能提高反转录酶的热稳定性, 因此可以大幅度提高反应温度, 减少 mRNA 的二级结构对反转录所带来的负面影响, 同

时也加大了得到所有全长 cDNA 的可能性。第二, 全长 cDNA 的获得。利用高碘酸钠的氧化特性, 在低温、避光条件下特异氧化 cDNA/mRNA 复合体中 mRNA 5' 和 3' 端末位核糖上的两个相邻的羟基 (2-OH 和 3-OH)。经  $\text{NaIO}_4$  作用后, mRNA 两端的邻二醇基团被氧化成二醛基团 (图 4), 后者在一定条件下能够与生物素结合, 而生物素化的 cDNA/mRNA 复合体可被链霉亲和素包被的磁珠来分离出来。第三, 为防止短截 cDNA 的掺入, 采用 RNase I 对双链

复合体进行酶切, RNase I 可以消化以单链状态存在的 mRNA, 而且没有碱基特异性。第四, 第二链 cDNA 的合成。第二链引物结合位点的引入可采用两种方法: 一种是通过末端转移酶在单链 cDNA 的 3' 端加上一段 poly(G), 另一种是在利用 DNA 连接酶在 cDNA 的 3' 端加上一段寡核苷酸。

我们在利用 Cap-trapper 方构建小麦全长 cDNA 文库时, 采用了末端转移法来引入第二链引物结合位点。通过严格控制末端转移的反应时间, 将 poly(G) 的长度有效地控制在 9~13 个碱基, 将 poly(G) 对测序的影响降低到最低限度, 大大提高了测序的成功率。此外, 为了简化实验操作, 在第一、二链引物中引入了 II 类限制性内切酶 *Bsa* I 的识别位点,

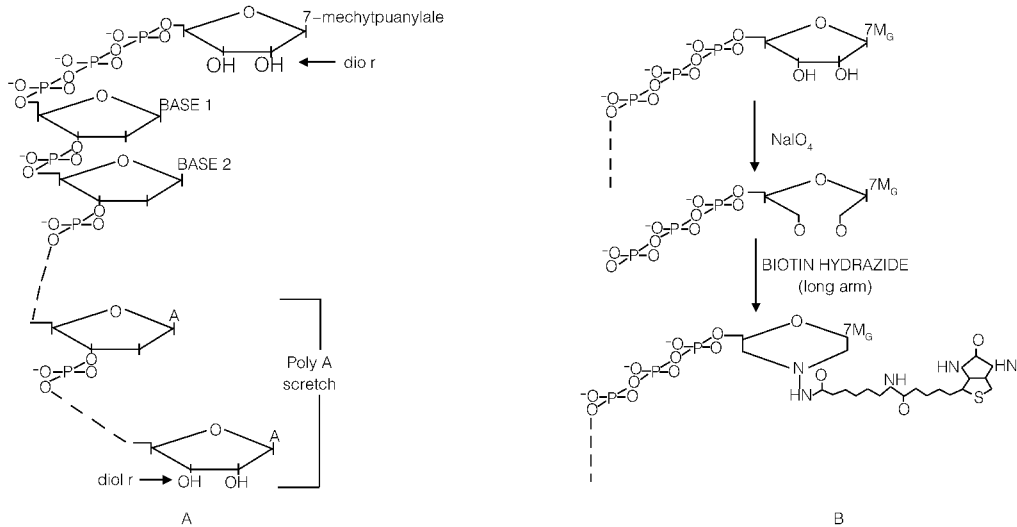


图 4  $\text{NaIO}_4$  氧化和生物素标记示意图(参考文献[6])

A: mRNA 3' 和 5' 端的结构; B:  $\text{NaIO}_4$  氧化和生物素标记

Fig. 4 Principle of  $\text{NaIO}_4$  oxidation and biotinylation (Reference [6])

A: Structure of mRNA 3' and 5' ends; B:  $\text{NaIO}_4$  oxidation and biotinylation.

## 1.5 Cap-jumping 法

Cap-jumping 法 (Cap-jumping method)<sup>[16]</sup> 在某些方面与 Oligo-capping 法和 Cap-trapper 法类似 (图 5), 之所以说它类似于 cap-trapper 法, 是因为它也是利用高碘酸钠来氧化 mRNA 5' 和 3' 端核糖上的邻二醇基团, 使之变为二醛基团。在 cap-trapper 法中, 二醛基团与生物素的氨基结合, 而在 Cap-jumping 法中, 用乙二胺取代了生物素。该方法与 Oligo-capping 法类似的地方在于它也在 mRNA 的 5' 端加上一段寡核糖核酸或 3' 端携带一个核糖集团的脱氧寡核糖核酸 (oligonucleotide template ex-

在 *Bsa* I 切割序列位点处分别引入 *Eco*R I 和 *Xho* I 的酶切位点, 这样通过用 *Bsa* I 单酶切就可以在 cDNA 的两端分别产生 *Eco*R I 和 *Xho* I 的黏性末端, 从而实现 cDNA 的定向克隆。采用这种技术, 很好地解决了双链 cDNA 经 *Eco*R I 和 *Xho* I 酶切可能导致克隆丢失和实现 cDNA 定向克隆这两个难题<sup>[22]</sup>。利用这种方法, 我们分别构建了小麦不同倍性材料的全长 cDNA 文库, 获得了近 10 000 条小麦全长 cDNA。结果分析表明, 全长 cDNA 的比例在 90% 左右<sup>[22]</sup>。

这种方法的最大缺点就是实验流程长, 技术要求高, 需要用同位素做平行实验, 操作繁琐。

tender, OTE), 但添加寡核糖核酸的方式不同, 它不需要除去 mRNA 5' 端的帽子结构, 而是直接用化学方法进行连接。经  $\text{NaIO}_4$  氧化后的 mRNA 在一定条件下与乙二胺作用成为氨基化的 mRNA, 这种 mRNA 在特定条件下可以和 3' 端 OTE 连接, 成为 5' 端携带一段核糖核酸或脱氧核糖核酸的 mRNA。而对于缺少帽子结构的 mRNA, 其 5' 端核糖集团上仅有一个羟基, 不能被  $\text{NaIO}_4$  氧化, 所以不能被胺反, 因而也就不能与氧化后的 OTE 连接。

经加工的 mRNA 作为反转录的模板合成第一链 cDNA。在合成 cDNA 时, 该方法充分利用反转录

酶的末端转移酶活性。在高浓度  $Mn^{2+}$  存在的条件下,反转录酶(SuperScript II)有极强的末端转移酶活性,它能在合成的第一链 cDNA 3' 端加上几个 dG,从而越过帽子结构,并使第一链 cDNA 延伸到人为

添加的寡聚核糖核酸位置,最终得到全长 cDNA。

有关用 Cap-jumping 法获得全长 cDNA 的原理最早见诸于 Efimov V A<sup>[15]</sup> 的报道中,而目前尚未见用这种方法构建全长 cDNA 文库的报道。

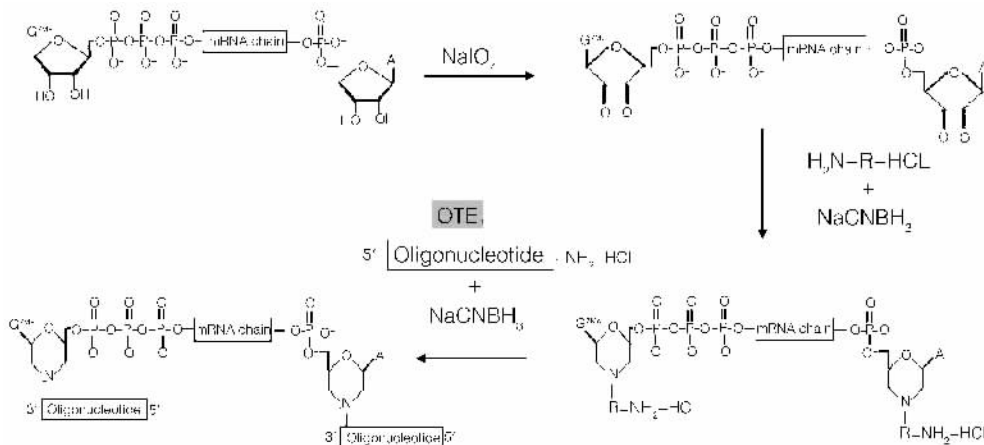


图 5 Cap-jumping 法原理(参考文献[16])

Fig.5 Schematic presentation of covalent binding of the OTE to the 5'-terminus of a capped mRNA(Reference [16])

### 1.6 mRNA 差减/均一化全长 cDNA 文库

全长 cDNA 文库在一定程度上解决了大规模获得全长 cDNA 的问题,但如果想通过大规模测序来发掘新基因,还有一个无法回避的现实问题——冗余序列。我们知道,生物体不同器官、不同组织,不同时期、不同生理条件下基因表达的模式不同,高拷贝的 mRNA 每种可达数千个,而与特定组织、器官的分化和发育及代谢相关的基因仅仅只有几个拷贝,因此通过测序的方法发现稀有表达基因,成功率非常低,而且不经济。为了提高发掘稀有表达新基因的效率,Carninci P 等<sup>[5]</sup> 在 Cap-trapper 法的基础上建立了基于 mRNA 差减/均一化全长 cDNA 文库构建方法(图 6),该方法不仅综合了构建全长和差减/均一化文库的技巧,而且又有许多独特的创新,克服了 DNA 杂交过程中的一些困难,文库质量得到较大提高。

根据差减/均一化所用的 driver 不同,可将其分为 DNA 差减/均一化法和 mRNA 差减/均一化法,前者用 DNA 作为差减/均一化的 driver,后者用 mRNA 作 driver。根据差减或均一化处理的时期不同又将其分为两类,一类是在文库构建的过程中进行差减或均一化处理,另一类是在文库构建完成以后进行上述操作。Carninci P 等<sup>[5]</sup> 提出的差减/均一化

全长 cDNA 文库的构建方法用 mRNA 作为差减/均一化的 Driver,在文库的构建过程中进行差减/均一化处理。该方法一定程度上解决了使用 DNA 进行差减处理遇到的难题——Driver 和 Tester 之间容易形成网状结构。影响网状结构形成的原因有两个方面:一类是由 Tester 和 Driver 自身的结构决定的;另一类可能是人为因素导致的。文库构建过程中,如果不控制 cDNA 5' 端 poly(G)和 3' 端 poly(A)的长度,在进行差减/均一化处理时,不同 cDNA 两端的 poly(A)和 poly(T),poly(G)和 poly(C)之间可能会结合,较链成网状结构,而使用 mRNA 作为 Tester,避免了人为因素的干扰。然而,由于 mRNA 比较容易降解,所以操作务必小心,谨防 RNase 污染。此外,由于使用 mRNA 作为 Driver,对 mRNA 的需求量非常大,所以在样品量比较少的情况下,这种方法不太适用。

mRNA 均一/差减法(mRNA normalization and subtraction of full-length cDNA library construction method)构建 cDNA 文库的过程可分为 3 个部分(图 6):全长 cDNA 的获得;差减/均一化处理;二链 cDNA 的合成及文库的获得。这 3 个部分中,除差减/均一化处理外,其他两部分的方法和步骤与 Cap-trapper 法都相同。在进行差减/均一化处理时,需

要将全长 ss-cDNA 与用生物素标记的过量 driver mRNA 杂交,用磁珠分离法将高表达的 cDNA 除去,

最终得到稀有表达的单链 cDNA(ss-cDNA)。

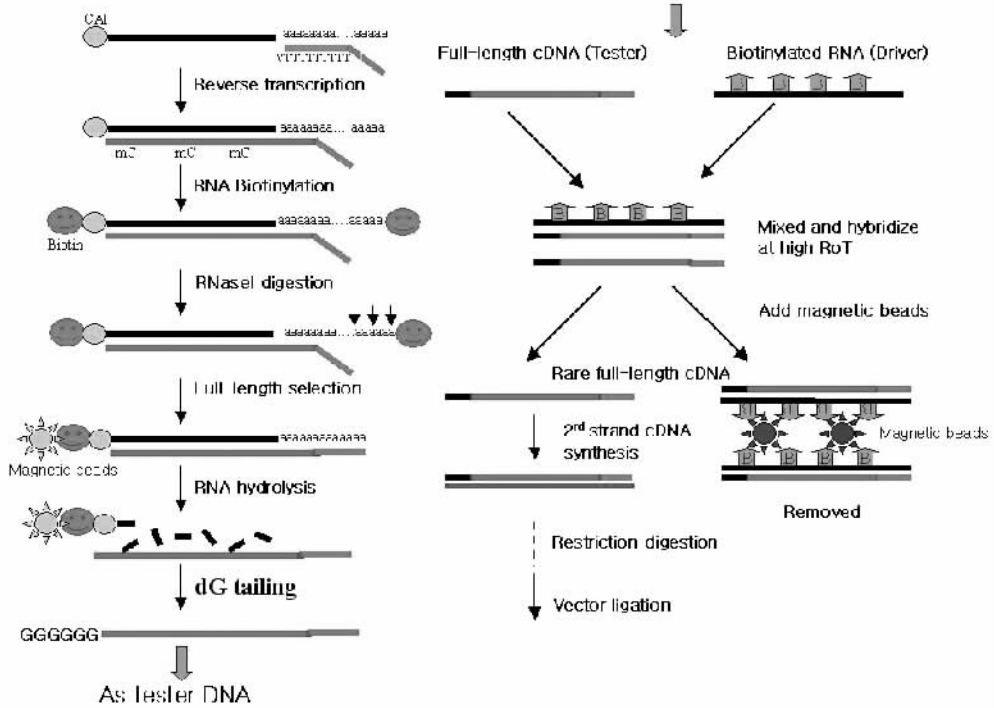


图 6 全长均一化/差减 cDNA 文库构建流程(参考文献[5])

Fig.6 Schematic diagram of the normalized-subtracted full-length cDNA preparation (Reference [5])

## 2 全长 cDNA 判定

全长 cDNA 的比例是影响全长 cDNA 文库质量的关键因素,如何判定 cDNA 序列的完整性是一个相当复杂的问题,它需要综合多方面因素进行全面评价。一般来说,来自 cDNA 文库的序列 3'端都带着 poly(A)尾巴,而 poly(A)尾巴就是 cDNA 3'端完整性的标志,所以 cDNA 的 3'端 UTR 大多都是完整的,而具体到 5'端,则需要采用不同的方法加以判断。根据判定方法的不同,可以将其分为两类:一类是实验验证法,另一类为生物信息学分析法。前者适于对少量数据的分析,后者适用于对大量数据的处理。本实验室开展了大规模小麦全长 cDNA 测序工作,得到 1 万多条小麦 cDNA 序列,在文库质量评价时我们选择后者用于序列完整性分析。

判定全长最直接、最迅速的方法是与 GenBank 数据库进行比较。如果数据库中存在与目的片段高度同源的全长基因,就可以确定该基因是否为全长,但这种方法也有例外,如果考察基因来自多拷贝基

因家族,尽管它们与 GenBank 中的同源序列在长度上差别较大,也不能草率断定目的基因是否为全长。

另一种方法是根据 mRNA 5'端的结构和碱基组成判定全长。mRNA 5'端 UTR 是影响 mRNA 转录的重要因素,起始密码子(ATG)上游-3位和下游+4位的碱基很大程度上决定了候选 ATG 位点转录起始的有效性<sup>[23~25]</sup>;起始密码子上游 UTR 的长度,G+C 含量<sup>[26]</sup>以及 ATG 的存在与否也会影响起始效率<sup>[27]</sup>;此外,研究发现高表达和稀有表达基因之间 5'UTR 的长度也存在较大差异<sup>[28]</sup>。Kochetov A V 等<sup>[29]</sup>对被子植物起始密码子两侧的序列作了全面的分析,并对单子叶和双子叶植物进行了统计和比较。他们发现,单子叶植物 Kozak 序列的一致序列为 gccRccATGGcg,5'UTR 的长度在 40~120 bp 之间,二碱基组合(R<sup>-3</sup>/G<sup>+4</sup>),在起始密码子两侧出现的频率很高。根据前人的研究结果,结合实际情况,选择合适的全长 cDNA 的判别方法,实现对基因全长的准确判定。

### 3 小 结

全长 cDNA 文库是高效、大规模获得基因全序列信息的基础,是开展功能基因组研究的一条重要途径。目前,在全长 cDNA 文库构建方面已经取得了很大的进展,总的来说,所有方法都着眼于真核生物 mRNA 5'端的帽子结构,但又各具独到之处。当前所有全长 cDNA 文库构建方法都存在一定的缺陷,在文库构建时应根据实验目的和实验室的工作条件选择适合自己的构建方法。

### 参 考 文 献 (References):

- [1] Haas B J, Volfvsky 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 Biol*, 2002, 3(6):RESEARCH0029.
- [2] Burge C B, Karlin S. Prediction of complete gene structure in human genomic DNA. *J Mol Bio*, 1997, 268:78~94.
- [3] Lukashin A V, Borodovsky M. GeneMark. hmm; new solutions for gene finding. *Nucleic Acids Res*, 1998, 26(4):1107~1115.
- [4] Salzberg S L, Pertea M, Delcher A L, Gardner M J, Tettelin H. Interpolated Markov models for eukaryotic gene finding. *Genomics*, 1999, 59(1):24~31.
- [5] 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 Res*, 2000, 10(10): 1617~1630.
- [6] Seki M, Carninci P, Nishiyama Y, Hayashizaki Y and Shinozaki K. High-efficiency cloning Arabidopsis full-length cDNA by biotinylated CAP trapper. *The plant Journal*, 1998, 15(5):707~720.
- [7] Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T, Ishii Y, Sasaki D, Bono H, Kondo S, Sugahara Y, Saito R, Osato N, Fukuda S, Sato K, Watahiki A, Hirozane-Kishikawa T, Nakamura M, Shibata Y, Yasunishi A, Kikuchi N, Yoshiki A, Kusakabe M, Gustincich S, Beisel K, Pavan W, Aidinis V, Nakagawara A, Held WA, Iwata H, Kono T, Nakauchi H, Lyons P, Wells C, Hume D A, Fagiolini M, Hensch TK, Brinkmeier M, Camper S, Hirota J, Mombaerts P, Muramatsu M, Okazaki Y, Kawai J, Hayashizaki Y. Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res*, 2003, 13(6B): 1273~1289.
- [8] Stapleton M, Liao G, Brokstein P, Hong L, Carninci P, Shiraki T, Hayashizaki Y, Champe M, Pacleb J, Wan K, Yu C, Carlson J, George R, Celniker S, Rubin G M. The Drosophila gene collection: Identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res*, 2002, 12(8): 1294~1300.
- [9] 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, Hashizume 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, Yasunishi A. Collection, mapping, and annotation of over 28 000 cDNA clones from japonica rice. *Science*, 2003, 301(5631):376~379.
- [10] Kato S, Sekine S, Oh S W, Kim N S, Umezawa Y, Abe N, Yokoyama-Kobayashi M, Aoki T. Construction of a human full-length cDNA bank. *Gene*, 1994, 150(2):243~250.
- [11] Ederly I, Chu L L, Sonenberg N, Pelletier J. An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). *Mol Cell Biol*, 1995, 15(6): 3363~3371.
- [12] Suzuki Y, Sugano S. Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol Biol*, 2003, 221(1): 73~91.
- [13] Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene*, 1997, 200(1-2):149~156.
- [14] Zhu Y Y, Machleder E M, Chenchik A, Li R, Siebert P D. Reverse transcriptase template switching; a SMART approach for full-length cDNA library construction. *Biotechniques*, 2001, 30(4):892~897.
- [15] Efimov V A, Chakhmakhcheva O G, Archdeacon J, Fernandez J, Fedorkin O N, Dorokhov Yu L, Atabekov J G. Detection of the 5'-cap structure of messenger RNAs with the use of the cap-jumping approach. *Nucl Acids Res*, 2001, 29(22): 4751~4759.
- [16] Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*, 1996, 37(3): 327~336.
- [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 captrapper. *DNA Res*, 1997, 4(1): 61~



- 66.
- [18] Clepet C, Le Clainche I, Caboche M. Improved full-length cDNA production based on RNA tagging by T4 DNA ligase. *Nucleic Acids Res*, 2004, 32(1):e6.
- [19] <http://dbtss.hgc.jp/docs/oligocap.html>.
- [20] <http://www.clontech.com>.
- [21] Wellenreuther R, Schupp I, Poustka A, Wiemann S. The German cDNA Consortium. SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics*, 2004, 5(1):36.
- [22] MAO Xin-Guo, KONG Xiu-Ying, ZHAO Gang-Yao, JIA Ji-Zeng. Construction of a Full-Length cDNA library of *Aegilops speltoides* tausch with optimized cap-trapper method. *Acta Genetica Sinica*, 2005, 32(8):811~817.  
毛新国, 孔秀英, 赵光耀, 贾继增. 利用改进的 Cap-trapper 法构建拟斯卑尔脱山羊草 (*Ae. speltoides*) 全长 cDNA 文库, 遗传学报, 2005, 32(8):811~817.
- [23] Pesole G, Gissi C, Grillo G, Licciulli F, Liuni S, Saccone C. Analysis of oligonucleotide AUG start codon context in eukaryotic mRNAs. *Gene*, 2000, 261(1):85~91.
- [24] Pesole G, Liuni S, Grillo G, Saccone C. Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene*, 1997, 205(1-2):95~102.
- [25] Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 2001, 276(1-2):73~81.
- [26] Mignone F, Gissi C, Liuni S and Pesole G. Untranslated region of mRNAs. *Genome Biol*, 2002, 3(3): reviews0004. 1~reviews0004. 10.
- [27] Rogozin I B, Kochetov A V, Kondrashov F A, Koonin E V, Milanese L. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics*, 2001, 17(10):890~900.
- [28] Kochetov A V, Ischenko I V, Vorobiev D G, Kel A E, Babenko V N, Kisselev L L, Kolchanov N A. Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features. *FEBS Lett*, 1998, 440(3):351~355.
- [29] Kochetov A V, Synchron O A, Rogozin I B. Context organization of mRNA 5'-untranslated regions of higher plants. *Mol Biol (Mosk)*, 2002, 36:649~656.

## 《分子生物学专业英语》一书的特色

哈尔滨工业大学出版社编辑 杜燕

《分子生物学专业英语》一书于 2006 年 1 月由哈尔滨工业大学出版社出版,主编程震龙是哈尔滨工业大学生命科学与工程系的中青年骨干教师,博士,他从事分子生物学教学和科研多年,尤其在双语教学领域具有丰富的实践经验,深受学生的喜爱与好评,曾获哈工大双语教学基金资助,并发表《分子生物学课程中的双语教学实践》等教学论文。

### 1. 内容全面新颖,系统性强

全书主要包括基础阅读与阅读材料两部分。基础阅读内容与国、内外通用的分子生物学教材的体系相一致,以权威教材为蓝本,精选主干内容,对其中的专业词汇予以注释,对文中的重点及难点内容予以解释说明。基础阅读部分共有 21 章内容,涉及生物小分子、DNA、RNA、蛋白质的结构与功能,基因、基因组、转录组、蛋白质组,复制、转录、翻译过程及其调控。

阅读材料部分是对基础阅读部分的扩展,供读者在完成基础阅读部分之后学习,或作为专业英语教学中的课外阅读材料使用。其中涉及分子进化、RNA 干涉与基因沉默、基因芯片、功能基因组、蛋白质组、生物信息学、艾滋病、基因治疗、疯牛病、转基因食品等,这些都是分子生物学发展的最新研究进展,与其他学科交叉的成果,具有学科前瞻性。

### 2. 材料重点突出,选择合理

该书在材料的选择上充分考虑到不同层次读者的需要。在基础阅读部分,主要选择国外通行的本科生及研究生教材中的相关内容,以生物大分子的结构与功能为重点,以中心法则为主线,将分子生物学的主要知识结构框架涵盖其中,重在讲解,使读者在掌握专业英语的同时掌握专业知识。

在阅读材料部分,本书的主要侧重点在于使读者了解学科的发展前沿与最新研究进展。这一部分在材料的选择上与基础阅读部分有所不同:前者重在使读者掌握分子生物学基础知识中的专业词汇及主要表达方式,内容结构更为严谨,材料多选自国外优秀教材;后者重在使读者了解更多、更新、更广的信息,因此选材更加灵活、新颖,内容也更具讨论性与趣味性。

### 3. 思想性、逻辑性强

该书思想正确,符合辩证唯物主义观点;层次分明,条理清楚,体系能够反映教材内容内在联系及本专业特有的思维方法。本书具体的内容安排是按着一生物大分子为重点,以中心法则为主线,从 DNA 到 RNA 到蛋白质,从 DNA 复制到转录、翻译,从基因组到转录组到蛋白质组的逻辑顺序编排,并扩展到相关重要的生命过程。体现出了生物大分子在生命活动过程中内在的相互联系、相互依存、相互制约的关系,同时又表现出了与环境条件的协调与统一。

本书对目前分子生物学研究领域中的不同观点都能够很好地包容,并给予客观地体现。对有争议的问题能够博采众家长,并体现辩证的观点,同时给予教师、学生与读者发挥自己思维的空间。

### 4. 语言生动流畅,注释准确

该书文字规范、简练、语言流畅、叙述生动。图文配合恰当,图表清晰、准确,符号、计量单位符号符合国家标准。对专业词汇及重点、难点段落的注释准确、明了。