

中华大蟾蜍精巢组织全长 cDNA 文库的构建及泛素延伸蛋白基因 *Bg-ubi52* 的获得

陈文 郑萍萍 聂刘旺*

(安徽师范大学生命科学学院 芜湖 241000)

摘要 运用 SMART 技术构建了中华大蟾蜍 (*Bufo bufo gargarizans*) 精巢全长 cDNA 文库。提取中华大蟾蜍精巢总 RNA, 用 Clontech 公司 SMARTTM cDNA 文库构建试剂盒反转录合成第一链 cDNA, LD-PCR 扩增获得全长 cDNA 双链, 经 *Sfi* I 酶切、层析柱分离后, 500 bp 以上的片段与 λ TriplEx2 载体连接并包装, 建成原始文库。经鉴定, 原始文库滴度为 2.21×10^6 pfu/ml, 重组率为 91%。文库扩增后的滴度为 2.94×10^9 pfu/ml, 重组率为 93.7%。插入片段大小分布于 0.4~2.0 kb 之间, 平均长度约为 1.0 kb, 说明已构建文库质量较高, 为进一步筛选、克隆精巢特异表达基因奠定了基础。从该文库中克隆到了泛素延伸蛋白基因, 全长 561 bp, 包含完整的 5' 和 3' 非编码区, 编码 128 个氨基酸, 即泛素的 76 个氨基酸后融合了 52 个氨基酸的核糖体 L40 蛋白。

关键词: 中华大蟾蜍 精巢 cDNA 文库 SMART 技术 泛素延伸蛋白基因

中图分类号: Q952 **文献标识码**: A **文章编号**: 0250-3263(2007)01-20-09

Construction of Testis cDNA Library and Sequence Analysis of Ubiquitin/L40e Extension Gene in *Bufo bufo gargarizans*

CHEN Wen ZHENG Ping-Ping NIE Liu-Wang*

(College of Life Sciences, Anhui Normal University, Wuhu Anhui 241000, China)

Abstract: A full-length cDNA library from the testis of Chinese Large Toad (*Bufo bufo gargarizans*) was constructed with the SMART (switching mechanism at 5' end of RNA transcript) technique. Total RNA was extracted from the testis and reversely transcribed into full-length cDNA using PowerscriptTM reverse transcriptase. First-strand cDNA was amplified by long-distance PCR. After *Sfi* I digestion and Chroma spin-400 fractionation, cDNAs (> 500 bp) were ligated to λ TriplEx2 vector and were packaged with Gigapack[®] III Gold Packaging Extract. The optimal primary library contained 2.21×10^6 pfu/ml clones and the amplified library had a titer of 2.94×10^9 pfu/ml, in which over 90% clones were recombinant and the average size of inserted cDNAs was about 1.0 kb. These results show that the testis cDNA library can be used for screening genes. A full-length gene with 5' and 3' untranslated regions was isolated from the cDNA library. Sequence analysis showed that this 561 bp cDNA encoded a 128-amino acid ubiquitin/L40e extension protein.

Key words: *Bufo bufo gargarizans*; Testis; cDNA library; SMART; Ubiquitin/L40e extension protein gene

尽管中华大蟾蜍 (*Bufo bufo gargarizans*) 精巢和卵巢的结构与功能、精子和卵子的结构与发生等在细胞水平已有许多研究。但是, 蟾蜍的性别决定、性别分化相关基因迄今未见报道。然而, 如同已在哺乳动物和其他动物中初步揭示的, 推测这些基因是在雄性性腺中特异表达, 与性腺发育密切相关, 如 DMRT1 和 DAX1 基因^[1-2]。随着分子生物学技术的发展和人类基因组计划的实施, 一系列寻找新基因的方法也应运而生, 而在众多方法中, 能得到完整的基因序列, 比较可行和可靠的主要还是 cDNA 文库筛选。cDNA 分子不含内含子序列和调控序列且与蛋白质的氨基酸序列具有对应性, 便于克隆和大量扩增。不像基因组 DNA 含有内含子很难表达, 可以从 cDNA 文库中筛选到所需的目的基因, 并直接用于该目的基因的表达。因此真核细胞的 cDNA 文库往往比基因文库更为有用, 构建 cDNA 文库是筛选新基因、分离已知基因、揭示某些因素对基因表达影响的基础^[3-5]。因此, 我们利用 SMART 技术, 构建了中华大蟾蜍的精巢全长 cDNA 文库, 为筛选性别决定、性别分化相关基因打下基础, 同时对于中华大蟾蜍这一物种遗传信息的保存及进化遗传学等方面的研究也有重要意义。对构建好的文库进行初步筛选, 分离得到一个具有 5' 和 3' 非编码区的泛素延伸蛋白基因。该基因在其他物种的成体精巢中也有大量表达。

1 材料与方 法

1.1 材料与试剂 中华大蟾蜍雄性个体采自安徽芜湖地区, 经活体解剖鉴定为性成熟个体。文库构建试剂盒 SMARTTM cDNA Library Construction Kit 购自 Clontech 公司, 噬菌体包装蛋白 Gigapack[®] III Gold Packaging Extract 购自 Stratagene 公司, RNA 提取试剂 Trizol[®] Reagent 购自 Invitrogen 公司, 其他常规分析纯试剂及耗材均为进口或国产。

1.2 总 RNA 提取 快速取新鲜的蟾蜍精巢组织入冰浴的匀浆管中, 按 Invitrogen 公司的 Trizol[®] Reagent 操作说明书分离总 RNA。1.1%

的琼脂糖凝胶电泳, 检测 RNA 完整性, 用紫外分光光度计测量 230 nm、260 nm 及 280 nm 处的 OD 值, 检测 RNA 纯度和得率。

1.3 cDNA 原始文库的构建 取 3 μ l (约 1 μ g) 蟾蜍精巢总 RNA, 反转录成 cDNA 第一链, 经 LD-PCR 合成双链 cDNA, 条件为 95 $^{\circ}$ C 20 s, (95 $^{\circ}$ C 5 s, 68 $^{\circ}$ C 6 min) 22 个循环。取 5 μ l PCR 产物和 2 kb DNA Marker, 用 1.1% 琼脂糖凝胶电泳鉴定双链 cDNA 的分布范围。双链 cDNA 经蛋白酶 K 消化、Sfi I 酶切、层析柱 (CHROMA SPIN-400) 纯化, 收集大于 500 bp 的 cDNA 片段。再与经 Sfi I 酶去磷酸化的载体 λ TriplEx2 在 16 $^{\circ}$ C 下, 按体积比 cDNA: 载体 = 1.5: 1 连接过夜。连接产物用 λ 噬菌体包装蛋白 Gigapack[®] III Gold Packaging Extract 进行体外包装, 22 $^{\circ}$ C 水浴 2 h, 然后加入 500 μ l 噬菌体稀释液和 20 μ l 氯仿, 混匀, 离心取上清, 即为蟾蜍精巢组织的原始文库。

1.4 原始文库的滴度及重组率测定 将原始文库分别按 1:5、1:10、1:20 3 个稀释度稀释, 各稀释物取 1 μ l 分别与 200 μ l 过夜培养的 XL1-Blue 菌液混合, 37 $^{\circ}$ C 温育 15 min 后加入 3 ml 融化的 LB/MgSO₄ 顶层培养基, 快速混匀铺在 90 mm LB/ MgSO₄ 平板上, 37 $^{\circ}$ C 培养 6 ~ 18 h, 定期检查确保噬菌斑生长。记数噬菌斑, 计算滴度 (pfu/ml) = 噬菌斑数 \times 稀释倍数 $\times 10^3$ / 噬菌体铺板体积 (μ l)。除在铺平板之前于顶层培养基中加入 75 μ l 的 IPTG (0.1 mol/L) 和 75 μ l 的 X-gal (0.1 mol/L) 外, 其余步骤同上述文库滴度测定方法, 统计蓝白斑数, 计算重组率 = 白斑数 / (白斑数 + 蓝斑数)。

1.5 扩增文库的建立及质量鉴定 以 XL1-Blue 为受体菌, 将全部原始文库按每个平板 6 $\times 10^4$ 个克隆数铺 150 mm 平板, 37 $^{\circ}$ C 培养 6 ~ 18 h, 待噬菌体生长融合后, 加入稀释缓冲液 4 $^{\circ}$ C 过夜。第二天摇床洗脱 1 h, 收集噬菌体洗脱液, 加入氯仿, 离心, 取上清即为扩增 cDNA 文库。扩增好的文库加 DMSO 至最终浓度为 7%, -70 $^{\circ}$ C 保存。取 1 μ l 上清按 1:100 000 稀释, 分别取 5 μ l、10 μ l、20 μ l 测定扩增文库的滴

度和重组率,方法同原始文库。随机挑取 40 个噬菌斑,连同琼脂分别放入 200 μ l SM 稀释缓冲液中,37 $^{\circ}$ C 温育 1 h,4 $^{\circ}$ C 过夜,然后进行 PCR 检测。取 2 μ l 噬菌体稀释液为模板,在 25 μ l 反应体系中加入 20 μ mol/L λ TriplEx2 插入子筛选引物 5'上游(5'-TCCGAGATCTGGACGAGC-3')和 3'下游(5'-TAATACGACTCACTATAGGG-3')各 0.5 μ l 和 1 U *Taq* DNA 聚合酶。PCR 反应循环参数为:95 $^{\circ}$ C 预变性 2 min,95 $^{\circ}$ C 30 s,58 $^{\circ}$ C 50 s,72 $^{\circ}$ C 2 min,35 个循环;72 $^{\circ}$ C 延伸 10 min。取 5 μ l 扩增产物,用 1.2% 琼脂糖凝胶电泳,鉴定插入 cDNA 片段的大小。

1.6 阳性克隆 cDNA 测序及其生物信息学分析 从文库平板上随机挑选噬菌斑,按试剂盒说明将阳性克隆噬菌体 λ TriplEx2 环化为 pTriplEx2,提取质粒 DNA。以 T7 和 SP6 为测序引物,委托上海英骏生物技术有限公司采用 ABI 3700 DNA Sequence 自动序列仪进行双向序列测定。将获得的基因序列输入 www.ncbi.nlm.nih.gov 网站,通过 BLAST 联机 GenBank 进行 DNA 及氨基酸序列相似性检索并命名。其他序列数据通过 BLAST 方法在线获取,全部源自 GenBank 数据库,具体见下文。通过 Clustal X1.81 软件对不同物种的氨基酸编码序列进行比对,所用矩阵为 Blosum,参数设置为 Gap Opening penalty :10 ;Gap Extension penalty :0.05 ;Delay Divergent Sequences :30%。系统进化树通过 MEGA V2.1 软件 NJ 法构建。用 Bootstrap 方法对构建的 NJ 树进行评估,参数设置为 random number generator seed :111 ;number of bootstrap trials :1 000。

2 结果与分析

2.1 中华大蟾蜍精巢组织总 RNA 的质和量 琼脂糖凝胶电泳结果显示,28 S、18 S 和 5 S rRNA 条带清晰,说明总 RNA 的完整性良好(图 1)。测得 $OD_{260}/OD_{280} = 2.12$, $OD_{260}/OD_{230} = 2.21$,RNA 总浓度达 0.94 μ g/ μ l,表明 RNA 的纯度很高。RNA 于 37 $^{\circ}$ C 下温浴 2 h(图 1:2),与对照(图 1:1)无明显的差异,表明 RNA 的稳定性良

好,完全符合建库要求。

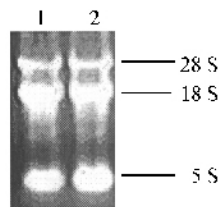


图 1 总 RNA 的 1.1% 琼脂糖胶电泳
Fig.1 Total RNA from testis of *Bufo bufo gargarizans* on 1.1% agarose gel

1. 总 RNA ; 2. 在 37 $^{\circ}$ C 下温浴 2 h 的总 RNA。
1. A sample of 5 μ l total RNA ; 2. A sample of 5 μ l total RNA incubated at 37 $^{\circ}$ C for 2 hours.

2.2 cDNA 的合成 图 2 中 22 个循环的 LD-PCR 产物电泳结果呈现从大到小均匀的 smear 涂布,大小在 0.4 ~ 2.0 kb 范围内,主要集中在 500 bp 以上。这表明中华大蟾蜍精巢的 RNA 成分具有多样性和复杂性。一般认为,非哺乳动物如植物、昆虫、酵母等的 PolyA⁺ RNA 分布在 0.15 ~ 3.0 kb 之间,由其合成的 cDNA 的分布范围也相应减小,本实验中用的是总 RNA,所以可能会有一些更小的片段。另外,不同组织 mRNA 的丰度变化也具有时空表达的特异性,这也会反应在 cDNA 分布范围上。因此在这个范围内的 cDNA 是比较完整的。

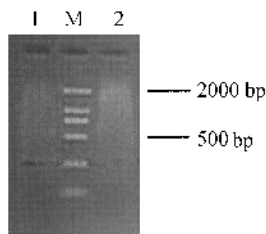


图 2 LD-PCR 产物电泳结果

Fig.2 The products of LD-PCR

M 2 kb 分子标记;1. 20 个循环的 LD-PCR 产物;
2. 22 个循环的 LD-PCR 产物。
M 2 kb marker ; 1. The product of LD-PCR with 20 cycles ; 2. The product of LD-PCR with 22 cycles.

2.3 所建文库的特性 将 cDNA 与 λ TriplEX2 载体按 1.5:1 比例连接后,经体外包装后所得原始文库的滴度为 2.21×10^6 pfu/ml,在涂有 X-

gal 和 IPTG 的平板上进行蓝白斑筛选,测定文库的重组率为 91%。中华大蟾蜍精巢 cDNA 文库经扩增后,测定文库的滴度为 2.94×10^9 pfu/ml 重组率为 93.7%。随机挑取 40 个噬菌斑进行 PCR,以鉴定 cDNA 插入片段大小,1.2% 琼

脂糖凝胶电泳结果显示,插入片段大小分布于 0.4~2.0 kb 之间,平均长度约为 1.0 kb,其中 14 个克隆的示意图见图 3,说明所构建的中华大蟾蜍精巢 cDNA 文库质量较高。

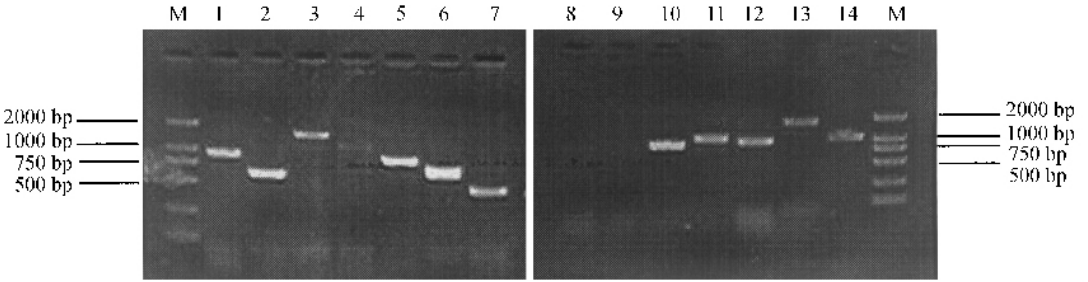


图 3 PCR 鉴定插入片段大小的电泳结果

Fig.3 Size identification of the inserted cDNA by PCR

M 2 kb 分子标记;1~14 随机挑取 14 个克隆的 PCR 产物。

M 2 kb marker; 1-14 PCR products of 14 random selected clones.

2.4 泛素延伸蛋白 ubi-52 的序列分析 随机挑选一个重组阳性克隆进行测序,经编辑后获得的 cDNA 片段全长 561 bp,其中 5'非翻译区有 93 个核苷酸,3'非翻译区有 94 个核苷酸,翻译区编码一个长为 128 个氨基酸残基的蛋白前体(图 4)。大多数真核生物基因起始密码子附近的序列为典型的 Kozak 结构 A/GNNATGG,其中 -3 位点的 A 和 +4 位点的 G 是高度保守的。该克隆的起始密码子附近序列为 AAGATGC,-3 位点为 A,但 +4 位点为 C。该基因的 5'非翻译区与编码核糖体蛋白质基因的 5'非翻译区有些共同点:GC 含量高,序列较短,在 5'端有一个富含嘧啶的序列^[6]。该基因富含嘧啶的序列为 -51 位到 -39 位的 TCCTTTCT-CCTT(图 4)。在 3'非编码区 polyA 的上游 20 bp 有一个 AATAAA 的多聚(A)加尾序列。通过与 GenBank 的核苷酸序列进行 Blast 比较,得到的基因为 ubiquitin/L40e extension protein 基因,又称为 ubiquitin-52aa extension protein 基因,根据中华大蟾蜍的拉丁学名 *Bufo bufo gargarizans*,命名为 *Bg-ubi52*。该基因序列已被 GenBank 数据库接受,登录号为 DQ855628。

进化地位物种的泛素延伸蛋白进行同源性比较,结果见图 5。原生动物的四膜虫、水稻及昆虫类的家蚕和蛾的泛素核糖体 L40 蛋白在泛素的 76 个氨基酸后融合了 53 个氨基酸,而从果蝇到人,包括酵母则在泛素之后融合了 52 个氨基酸。不同物种中前 76 个编码泛素的氨基酸是高度保守的,特别是参与泛素-蛋白复合体形成的 4 个功能位点 Lys-29、Lys-48、Lys-63 和 Gly-76 是绝对保守的,但有几个变化的位点值得注意。第 19 位,四膜虫、家蚕和蛾为 A,酵母、真菌、植物类的水稻和拟南芥为 S,其他为 P。第 20 位,除了鸡为 N,其他皆为 S。第 24 位,酵母、真菌、水稻和拟南芥为 D,其他为 E。第 28 位,只有酵母和真菌为 S,其他为 A。第 57 位,鸡、水稻和拟南芥为 A,而其他为 S。在泛素的三级结构中这几个变化的位点将聚在一起,形成一个和泛素的羧基化末端反方向的区域。据推测这个不保守的区域可能参与非蛋白水解,且在不同的物种间功能也不同^[7]。后 52/53 个编码核糖体 L40 的氨基酸包含一段富含碱性氨基酸的序列和一个 Zn 指结构,在不同进化地位的物种间也相当保守。C 末端 122~128 位这段富含碱

使用 ClustalX1.8 软件对真核生物中不同

```

                                -90
                                GGGCAGTAACGGGAGTGAAAACG
                                -30
                                -1
TCACTCCAGGTCGGGCCAGTCCTTTTCTCCTTAGTCGGGCCGCTGTGCGGGCAACAGAAAGCGCTCAAG

+1
ATG CAGATCTTTGTAAAAACCCTTACCGGGAAAACGATCACACTTGAGGTCGAGCCCAGTGACACTATTGA
M Q I F V K T L T G K T I T L E V E P S D T I E
1
90
AAATGTCAAGGCTAAGATTCAGGACAAGGAAGGTATCCACCTGACCAGCAGCGTCTGATCTTTGCTGGC
N V K A K I Q D K E G I P P D Q Q R L I F A G
30 10 40 20

150
AAGCAGCTTGAAGATGGACGTACCTTGTCTGACTACAACATCCAGAAAGAGTCCACTTTGCATCTGGTGTCT
K Q L E D G R T L S D Y N I Q K E S T L H L V L
50 60 70

230
CGCTTCTGTGGTGGTATCATTGAGCCTTCCCTCCGCCAGCTTGCCCAAGATACAATTGTGACAAGATGA
R L R G G I I E P S L R Q L A Q K Y N C D K M I
80 90

290
TTTGCCGCAAGTGCTATGCCCGTCTTACCCACGTGCTGTGAAGTCCGTAAGAAGAAGTGTGGCCACAC
C R K C Y A R L H P R A V N C R K K K C G H T
100 110

380
CAACAACCTGCGCCCAAGAAGAAGCTCAAATAACTCCATATGTCCTCTGCTATTGTITGAGTTTTGTTT
N N L R P K K K L K
120

430
AGAATAAAATAAAAACTAAAGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
460

```

图 4 中华大蟾蜍 *ubi52* cDNA 的核苷酸序列及推导的氨基酸序列

Fig.4 *Bg-ubi52* cDNA nucleotide and deduced amino acid sequence of *Bufo bufo gargarizans*

ATG 为起始密码子 ,TAA 为终止密码子 ,AATAAA 为 polyA 加尾信号。1~76 位氨基酸编码泛素蛋白 ,77~128 位(粗体字)氨基酸编码核糖体 L40 蛋白。

The positions of the nucleotides in the cDNA are given above the residues ; the positions of amino acids are designated below. Amino acids 1 through 76 are ubiquitin and 77 through 128 are ribosomal protein L40 ; the L40 sequence is in bold letters. ATG :initiation codon ;TAA : termination codon ; AATAAA indicates a polyadenylation signal sequence.

性氨基酸的序列 RPKKKL/VK 和一个核定位信号极为相似 ,这表明无论是整个泛素延伸蛋白还是 52/53 核糖体 L40 尾部蛋白 ,都可能会结合 DNA^[7]。Zn 指结构(96~115 位)更加强了这种可能性。Zn 指结构最早是在研究非洲爪蟾 5S rRNA 基因转录因子 TF III A 蛋白的氨基酸序列时发现的。典型的 TF III A Zn 指结构的序列为 Cys/His-X₂₋₄-Cys/His-X₂₋₁₅-Cys/His-X₂₋₄-Cys/His^[8]。相比较后 ,本文基因编码的蛋白质中的序列 Cys-X₂-Cys-X₁₀-Cys-X₄-Cys 符合这种结构 ,推测将会结合一个 Zn 原子 ,而 Zn 将使蛋白质分子形成指形构象的骨架与 DNA 结合^[9]。有研究推测 52/53 核糖体 L40 尾部蛋白作为整个延伸蛋白的一部分 ,在体内和特定的 DNA 相结

合 ,从而增加这些位点的泛素浓度 ,可能导致泛素化事件限定在染色体的某些特定部位^[7]。

以酿酒酵母的 *ubiquitin/S27A* 延伸蛋白基因的有 76 个氨基酸的核糖体 S27A 亚基为外群 ,基于从原生动物四膜虫、酵母到人的 *ubiquitin/L40* 延伸蛋白基因的 52 或 53 个氨基酸的核糖体 L40 亚基的氨基酸序列 ,用 MEGA V2.1 构建的 NJ 系统发生树如图 6 所示。动物界中最低等、最原始的原生动物位于树的基部。真菌类聚为一类 ,并且低等真菌酵母位于该类的下方 ,较早分化 ;高等真菌 ,如图 6 中的稻瘟真菌位于该类的上方 ,分化较晚。植物类形成一个独立的 ,比较古老的分支。无脊椎动物昆虫类聚在一起 ,和最进化的脊索动物从鱼到人

	20	40	60
Thale cress	MQIFVKTLTGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLADYD		
Rice	MQIFVKTLTGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLADYD		
Rat	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Toad	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Human	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Chicken	MQIFVKTLTGKTTITLEVEPNDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLADYD		
Fish	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Fruit fly	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Silkworm	MQIFVKTLTGKTTITLEVEASDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Miller	MQIFVKTLTGKTTITLEVEASDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Fungus	MQIFVKTLTGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Yeast	MQIFVKTLTGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
Tetrahym	MQIFVKTLTGKTTITLEVEASDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYD		
	*****:***:* . ***:***:*****:*****:*****:*****:***		
	80	100	120
Thale cress	IQKESTLHLVLRGGIIEPSLMLARKYNQDKMICRKYARLHPRAVNCRKKKCGHSNQ		
Rice	IQKESTLHLVLRGGIIEPSLQALARKYNQDKMICRKYARLHPRAVNCRKKKCGHSNQ		
Rat	IQKESTLHLVLRGGIIEPSLRQLAQKYNCDKMICRKYARLHPRAVNCRKKKCGHTNN		
Toad	IQKESTLHLVLRGGIIEPSLRQLAQKYNCDKMICRKYARLHPRAVNCRKKKCGHTNN		
Human	IQKESTLHLVLRGGIIEPSLRQLAQKYNCDKMICRKYARLHPRAVNCRKKKCGHTNN		
Chicken	IQKESTLHLVLRGGIIEPSLRQLAQKYNCDKMICRKYARLHPRAVNCRKKKCGHTNN		
Fish	IQKESTLHLVLRGGIIEPSLRQLAQKYNCDKMICRKYARLHPRAVNCRKKKCGHTNN		
Fruit fly	IQKESTLHLVLRGGIIEPSLRILAQKYNCDKMICRKYARLHPRATNCRKKKCGHTNN		
Silkworm	IQKESTLHLVLRGGIIEPSLRILAMKYNCEKMICRKYARLHPRATNCRKTKCGHTNN		
Miller	IQKESTLHLVLRGGIIEPSLRILAMKYNCDKMICRKYARLHPRATNCRKTKCGHTNN		
Fungus	IQKESTLHLVLRGGIIEPSLKALASKFNCDKQICRKYARLPPRATNCRKRKCGHTNQ		
Yeast	IQKESTLHLVLRGGIIEPSLKALASKYNCDKSVCRKYARLPPRATNCRKRKCGHTNQ		
Tetrahym	IQKESTLHLVLRGGMEPTIAALAKKYNCEKVCRCYARLPPKATNCRKRKCGHSNS		
	*****:***:*** ** *:* :** :**** *:* **** ***:**		
Thale cress	LRPKKKIK-		
Rice	LRPKKKIKN		
Rat	LRPKKKVK-		
Toad	LRPKKKVK-		
Human	LRPKKKVK-		
Chicken	LRPKKKVK-		
Fish	LRPKKKVK-		
Fruit fly	LRPKKKVK-		
Silkworm	LRPKKKIKD		
Miller	LRPKKKIKD		
Fungus	LRPKKKVK-		
Yeast	LRPKKKLR-		
Tetrahym	LRLKKPKKE		
	** *** :		

图 5 用 ClustalX1.8 软件比对不同来源的泛素延伸蛋白基因氨基酸序列

Fig.5 Multiple alignment of the aa sequences of ubiquitin/L40e extension proteins

参加比对的序列来自以下物种 拟南芥、水稻、大鼠、人、鸡、蟾蜍、文昌鱼、果蝇、蚕、甜菜夜蛾、稻瘟病菌、酵母菌、四膜虫；“*”和“:”分别表示绝对和高度保守的氨基酸残基。

The following sequences have been used for the alignment: *Arabidopsis thaliana* (Thale cress ; J05507) , *Oryza sativa* (Rice ; D12629) , *Rattus norvegicus* (Rat ; BC061544) , *Homo sapiens* (Human ; AF348700) , *Gallus gallus* (Chicken ; Z29958) , *Branchiostoma belcheri* (Fish ; AF395864) , *Drosophila melanogaster* (Fruit fly ; X53059) , *Bombyx mori* (Silkworm ; AY769312) , *Spodoptera exigua* (Miller ; AY149883) , *Magnaporthe grisea* (Fungus ; AAC13689) , *Saccharomyces cerevisiae* (Yeast ; Özkaynak et al. ,1987) , *Tetrahymena pyriformis* (Tetrahym ; CAA40021) .

相聚的这一支关系最近。值得注意的是,脊索动物这一支中鱼类又单独分离出来,位于该支

的基部。

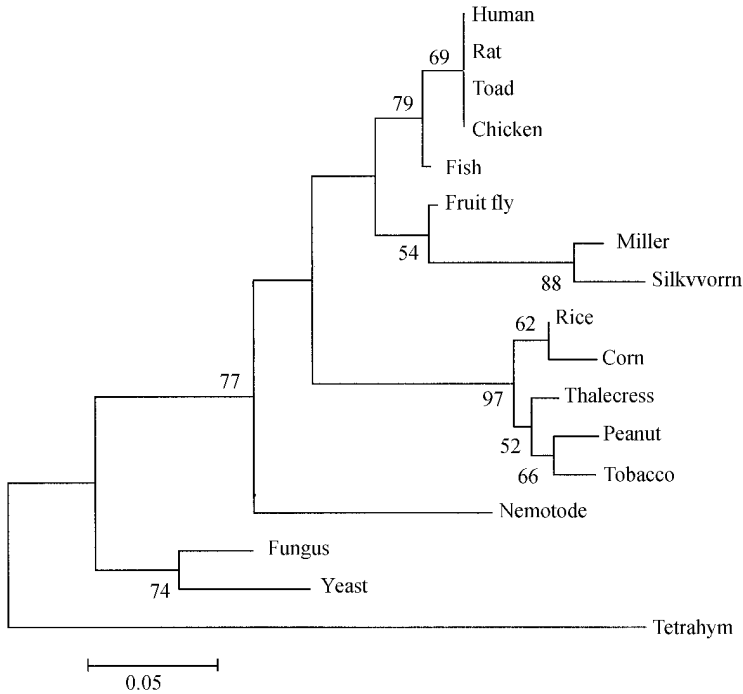


图 6 不同物种泛素延伸蛋白基因的核糖体 L40 亚基的系统进化树分析

Fig.6 Phylogenetic tree of ribosomal proteins from the small subunit(ubiquitin-ribosomal L40-protein)

以酿酒酵母 ubiquitin/S27A 的核糖体 S27A 亚基为外群,节点旁数据为 ML 分析 1 000 次 bootstrap 检验后的置信度(> 50% 的标出)。构建系统树的序列来自花生、玉米、烟草、线虫和图 5 中所用的物种。

Explanations of some sequences used for this tree are given in Figure 5. Other sequences have been used for this analysis :*Brassica rapa* (Peanut ; L21898),*Zea mays* (Corn ; BT018987),*Nicotiana sylvestris* (Tobacco ; M74100),*Caenorhabditis elegans* (Nematode ; P49632). Scale bar equals 0.05 (NJ) unit of Kimura 's two-parameter distance. The *Saccharomyces cerevisiae* ribosomal protein from the large subunit (ubiquitin-ribosomal S27A-protein) served as outgroup (AAB67466). Phylogenies were determined with MEGA V 2.1 by bootstrap analysis using neighbor-joining (1 000 replicates). Branches with less than 50% support have been collapsed .

3 讨论

评价一个文库是否有实用价值主要在于文库的含量和插入片段的大小两个方面。所构建的文库中必须有足够多的克隆数,这样才能确保基因组中的每一个序列至少有一个拷贝存在于重组文库中,文库的 cDNA 克隆数(独立克隆数)一般要求大于 $1 \times 10^6 \sim 2 \times 10^6$,本实验所构建的 cDNA 文库未扩增时滴度 2.21×10^6 pfu/ml,文库扩增后滴度为 2.94×10^9 pfu/ml,重组率达到 90% 以上,完全满足要求。同时为了鉴定

重组噬菌体中插入的 cDNA 片段的长度,本实验随机从文库中挑取 40 个噬菌斑进行 PCR 鉴定,电泳结果显示插入片段大小分布于 0.4 ~ 2.0 kb 之间,平均长度约为 1.0 kb。结果表明所构建的 cDNA 文库是高质量的。

本研究首次成功构建了中华大蟾蜍精巢全长 cDNA 文库,进一步筛选克隆获得有价值的 cDNA,将有助于了解精巢组织的基因表达谱,从分子水平探讨特异基因的表达。中华大蟾蜍精巢 cDNA 文库构建的技术特点 (1) 由于采用 Clontech 公司的 SMART 技术,该技术对接头引

物和反转录酶都进行了优化,可有效地去除不带多聚 A 尾的 RNA 和基因组 DNA,简化了 RNA 的纯化过程,而且能够合成完整的双链 cDNA,提高了文库中所含全长 cDNA 的比例(2)在合成第二链 cDNA 时,采用了 LD-PCR 技术,使得低丰度的基因表达序列在文库中得到富集。合成 cDNA 第二链时,PCR 扩增循环数的确定也很关键。循环数过多,容易产生非特异 PCR 产物,而循环数过少则产量不足。控制循环次数对基因在文库中拷贝数的分布也很重要,适当的循环次数可以防止低拷贝数的基因丢失,同时也可防止高拷贝基因过分放大^[10]。本研究在合成双链 cDNA 时,先采用较少的循环数,再逐渐增加,最终确定双链 cDNA 合成的最佳循环数为 22 个(图 2)(3)由于采用了具有稀有酶切位点的内切酶,建成的文库为定向文库,从而避免了钝端处理或甲基化保护内切酶位点不完全等建库方法的缺陷^[11,12](4)使用 λ TriplEx2 噬菌体作为载体,除了具有适用于低丰度或极低丰度的 mRNA 外,还有有序表达克隆插入序列及通过亚克隆可方便地将克隆从噬菌体转为质粒,给 cDNA 测序和其他研究带来了诸多方便^[13]。

在文库的初步筛选过程中分离到一个具有完整的 5'和 3'非编码区的泛素延伸蛋白基因。真核生物中泛素延伸蛋白基因分为两类:一类 C 末端与 52 或 53 个氨基酸的蛋白质融合,这类融合的蛋白与核糖体 L40 有很高的同源性;另一类与 76~80 个氨基酸的蛋白质融合,这类融合的蛋白质与核糖体 S27A 蛋白有很高的同源性^[14]。本文获得的基因是泛素基因后延伸一个 52 个氨基酸的蛋白质,属于第一类。泛素作为 76 个氨基酸的小多肽,在真核生物中广泛存在,是一种十分保守的蛋白,泛素基因后的核糖体 L40 延伸部分保守性也比较高,比如延伸部分 C 末端富含碱性氨基酸的序列 RPKKKL/VK 和 Zn 指结构,这暗示该基因在不同物种间可能行使相似甚至相同的功能。研究鸡的 *Ubi52* 基因发现,该基因在网织红细胞和肝脏中只有很微弱的表达,却偏爱在鸡的精子发生

过程中表达。在未成熟的精巢和成熟精巢中皆有大量表达,且在成熟精巢中的表达量显著高于未成熟精巢中的表达量。在精子生成不同时期的研究结果显示,*Ubi52* 基因在精子生成过程中表达量逐步提高,在精子生成的减数分裂后期和成熟的精子中皆有丰富的表达^[6,15]。本文获得的中华大蟾蜍 *Ubi52* 全长 cDNA 将为两栖动物中泛素基因家族在精子发生过程及早期胚胎中的研究提供分子资料。另外,考虑到泛素在真核生物中是高度保守的,并不适合用做系统发生方面的研究,本文选取泛素延伸蛋白基因的 52 或 53 个氨基酸的核糖体 L40 延伸部分来研究,因为延伸部分并不像泛素部分那么保守。在基于 Kimura 双参数模型的 NJ 系统发生分析中,较为正确地反映这些物种的系统发生关系,可见,泛素延伸基因 *Ubi52/53* 的延伸部分可作为生物物种进化的分子标志,值得进一步研究。这些均表明,本研究构建的文库,可以满足目的基因的克隆及其进一步比较分析的需要。

参 考 文 献

- [1] Shibata K, Takase M, Nakamura M. The DMRT1 expression in sex-reversed gonads of amphibians. *Gen Comp Endocrinol*, 2002, **127**(3): 232~241.
- [2] Sugita J, Takase M, Nakamura M. Expression of DAX1 during gonadal development of frog. *Gene*, 2001, **280**: 67~74.
- [3] 杨芳, 贺智敏, 詹显全等. 大鲵肝脏组织定向 cDNA 文库的构建及鉴定. *动物学报*, 2004, **50**(3): 475~478.
- [4] 李太武, 相建海, 刘瑞玉. 中国对虾 cDNA 文库的构建. *动物学报*, 1998, **44**(2): 237~238.
- [5] 林俊堂, 李玉昌, 张会勇等. 常氏肝癌细胞 cDNA 文库的构建及 ADAMs 相关基因的免疫筛选与序列分析. *遗传*, 2004, **26**(6): 793~796.
- [6] Mezquita J, Pau M, Mezquita C. Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids. *Gene*, 1997, **195**: 313~319.
- [7] Özakaynak E, Finley D, Solomon M J, et al. The yeast ubiquitin genes: a family of natural genes fusions. *EMBO J*, 1987, **6**(5): 1429~1439.
- [8] Berg J M. Potential metal-binding domains in nucleic acid binding proteins. *Science*, 1986, **232**(4749): 485~487.

- [9] 杨歧生编著. 分子生物学基础. 杭州 浙江大学出版社, 2001 92 ~ 93.
- [10] 左开井, 吴菲, 唐克轩. 海岛棉品种根部黄萎病菌诱导表达全长 cDNA 文库的构建. 棉花学报, 2002, **14**(5) : 291 ~ 294.
- [11] D 'Alessio J M, Gerard G F. Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H : the fate of information at the mRNA 5' terminus and the effect of *E. coli* DNA ligase. *Nucleic Acids Res*, 1988, **16**(5) : 1 999 ~ 2 014.
- [12] McClelland M, Nelson M, Raschke E. Effect of site - specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res*, 1994, **22**(17) : 3 640 ~ 3 659.
- [13] 王艺磊, 张子平. 日本对虾精巢和卵巢全长 cDNA 文库的构建. 动物学杂志, 2002, **38**(2) : 9 ~ 13.
- [14] Hector L C, Rosa B, Carmen A. Structure and expression of the *Drosophila* ubiquitin-52-amino-acid fusion-protein gene. *Biochemistry*, 1992, **286** : 281 ~ 288.
- [15] Rocamora N, Agell N. Methylation of chicken Ub I and Ub II polyubiquitin genes and their differential expression during spermatogenesis. *Biochemi J*, 1990, **267** : 821 ~ 829.