双阳梅花鹿三个 EST 序列的发现

关 敏　常雅萍　于永利

（白求恩医科大学免疫学教研室　长春　130021）

摘要：采用 HSV 1-2 TK 和人 IL-1β 特异性引物分别对鹿肝组织 cDNA 文库及鹿血 cDNA 进行 PCR 扩增，得到的特异性 cDNA 片段分别连入 TA 克隆载体测序，得到三个 EST，输入 GeneBank、Swissprot、dbEST 等数据库中进行同源比较分析。同源分析表明：获得双阳梅花鹿三个基因 EST。EST-1 为梅花鹿 α_2-抗纤溶酶同源基因 EST，EST-2 为梅花鹿水通道蛋白同源基因 EST，EST-3 为鹿血未知基因 EST，并含有一个 267 个碱基编码 88 个氨基酸的开放读码框架。

关键词：表达序列标签；双阳梅花鹿；cDNA；PCR

中图分类号：R392.2　**文献标识码**：A　**文章编号**：0250-3263(2001)05-28-05

Identification of Three ESTs of Shuangyang Sika Deer

GUAN Min　CHANG Ya-Ping　YU Yong-Li

(*Department of Immunology, Norman Bethune University of Medical Sciences　Changchun*　130021, *China*)

Abstract: We amplified the sika deer hepatic tissue cDNA library through PCR with HSV1-2 TK specific primer and obtained a cDNA fragment. Another two cDNA fragments of 500 bp and 1 000 bp were also isolated from sika deer blood cDNA through RT-PCR primed with a human IL-1β specific primer. These cDNA fragments were cloned into pGEM T Easy plasmid. Sequencing and BLAST search revealed that EST-1 is homologous to α_2-antiplasmin, EST-2 is homologous to aquaporin-4 (AQP-4), and EST-3 is a novel EST of sika deer containing a 267 bp ORF.

Key words: EST; Sika deer; cDNA; PCR

表达序列标签（expressed sequence tags, EST）是由大量随机取出的 cDNA 克隆一次测序得到的，是组织或细胞基因组 cDNA 的部分序列。EST 的概念是由 Adams 等人在 1991 年首次提出的[1]。1993 年美国 NCBI（National Center of Biotechnology Information）建立了一个专门的 EST 数据库（database of EST，dbEST）来保存和收集所有的 EST 数据[2]。EST 目前主要用于发现新基因和了解基因表达概况，是获取未知基因和发现新基因相对快速的方法。为了开发吉林双阳梅花鹿的基因资源，我们用两种引物从双阳型梅花鹿肝脏 cDNA 文库和鹿血中钓到三个 cDNA 克隆，经 DNA 测序后得到三个 EST。

1　材料与方法

1.1　实验材料

1.1.1　菌株和质粒　pGEM-T Easy Vector 购自美国 Promega 公司，菌种 JM109 由本实验室保存。

1.1.2　限制性内切酶及其它试剂　*Eco*R Ⅰ 购于大连宝泰克公司；逆转录酶、Rnase、*Taq* DNA 聚合酶购于上海 Promega 公司；Trizol 试剂购自 GIBCO /BRL 公司。逆转录试剂盒购自 GIBCO/

第一作者介绍　关敏，女，27 岁，硕士；研究方向：分子免疫学；

收稿日期：2000-10-26，修回日期：2001-03-05

BRL公司。Gene Clean 试剂盒购自白求恩医科大学分子生物学教研室。纯化质粒 DNA 试剂盒(Wizard)购自美国 BIO 101 公司。

1.1.3　实验标本　新鲜鹿血来自吉林省双阳鹿场。此梅花鹿属偶蹄目、鹿科、鹿属、东北亚种(*Cervus nippon hortulorum*)。

1.1.4　PCR 引物和 PCR 模板　人 IL-1β cDNA PCR 引物和 HSV 1-2 TK PCR 引物,由美国加洲大学合成。PCR 模板分别由双阳型梅花鹿肝组织 cDNA 文库和逆转录得到的 cDNA 为模板。人 IL-1β 引物:上游 AAACAGAAGTGCTCCTTC-AGG;下游 TGGAGAACACCACTTGTTGTT-GCTC-CA。HSV 1-2 TK 引物:上游 GCGAGAT ATCG-GCCGGGGAG;下游 TGCGGGCCCACAGCCTCCC。

1.2　方法

1.2.1　RT-PCR　从梅花鹿鹿角处采血约 0.5 ml,加入 1 ml Trizol 试剂裂解细胞。加 200 μl 氯仿,4℃ 12 000 r/min 离心 15 min,将上层水相中加 0.5 ml 的异丙醇,离心 10 min 使 RNA 沉淀,加 1 ml 75% 乙醇洗涤沉淀。溶总 RNA 沉淀于 9 μl DEPC 处理水中,加 1 μl Oligo-dT(0.05 μg/ml),65℃水浴 10 min。冰上 2 min,加 4.7 μl 5× 逆转录酶缓冲液,0.5 μl Rnasin(40 U/ml),8 μl dNTP(5 mmol/L),0.5 μl M-MLV 逆转录酶(200 U/μl)。混合后置 37℃水浴 60 min。

1.2.2　PCR 扩增　分别用鹿肝组织 PCR cDNA 文库为模板,HSV 1-2 TK 特异性引物和逆转录得到鹿血 cDNA 为模板,人 IL-1β 特异性引物进行 PCR 扩增:94℃ 60 s, 55℃ 60 s, 72℃ 90 s,共 35 个循环,72℃延伸 10 min。

1.2.3　cDNA 片段连入 TA 克隆载体　放大的特异性 cDNA 片段用 Gene Clean 试剂盒进行回收,取回收后的 cDNA 片段连入 pGEM T Easy 质粒,制备并转化新鲜的大肠杆菌 JM109 感受态细胞,扩大培养后以碱裂解法提取重组质粒。

1.2.4　插入 cDNA 片段测序　各取 2 μl 质粒进行 1% 琼脂糖凝胶电泳,观察是否有质粒滞后现象。挑选与从蓝色菌落中所提质粒相比滞后的质粒,用 *Eco*R Ⅰ 进行酶切鉴定。用 Wizard 试剂盒提取纯化重组 TA 质粒,经上海基康公司 DNA 序列分析仪进行自动测序。

2　结　果

2.1　EST-1 的获得和分析　鹿肝组织 cDNA 文库中 PCR 扩增得到一条特异性 cDNA 片段。连入 TA 载体后用 *Eco*R Ⅰ 酶切,释放出的 cDNA 片段大小为 400 bp,进行 DNA 测序,去除载体序列,得到 413 bp 的一段核苷酸序列,将其中含有开放读码框架(ORF)的序列,命名为 EST-1。在 GeneBank(http: www.ncbi.nlm.nih.gov/blast/)中进行相似性(Blastn)查找发现,与牛、小鼠、人 α_2-抗纤溶酶有 90% 以上的同源性。牛 α_2-抗纤溶酶全长 cDNA 为 2 216 个碱基,EST-1 与之 3′端 212 个碱基(2 004～2 216)有 90% 同源性(图 1)。序列输入 EST 数据库(dbEST),相似性查找发现 EST-1 3′端 66 个碱基与人胎肝 cDNA EST (gb/AA514190)有 98% 同源性。并在 GeneBank-dbEST 中登录,登录号为 AF274583 (AW990956)。

2.2　EST-2 的获得和分析　以 RT-PCR 方法,用人 IL-1β 特异性引物从鹿血 cDNA 中得到两条特异性 cDNA 片段,大小约为 500 bp 和 1 000 bp。测序后得到两个序列,去除载体序列,将其中含有开放读码框架(ORF)的序列,命名为 EST-2 和 EST-3。分析发现 EST-2 序列中含有 2 段 ORF,两个 ORF 产生的氨基酸序列是不同的。对应的氨基酸序列(134aa 和 138aa)经 Swissprot 中查找发现它们与牛肌钙网蛋白、大鼠钠通道蛋白、人钠通道蛋白等有较高同源性。提示 EST-2 可能编码参与膜转运的水通道蛋白或钠通道蛋白。GenBank 中查找发现(图 2),牛水通道蛋白-4-B(AQP-4-B)mRNA 全长为 5 641 bp,EST-2 有 201 个碱基(249,449)与 AQP-4-B mRNA 3′非编码区(4 358,4 557)有 90% 的同源性,与牛水通道蛋白-4-A(AQP-4-A)mRNA 3′非编码区(4 417,4 616)也有 90% 的同源性。在 GeneBank-dbEST 中登录,登录号为 AF274584 (AW990955)。

```
emb|X78436|BTA2A B.taurus mRNA for alpha-2-antiplasmin
 Length = 2216  Score =  244 bits (123)  Expect = 4e-62
Identities = 194/215 (90%) Gaps = 7/215 (3%)  Strand = Plus / Plus
Query: 89   cagcctccctcccactgtcacctcggtctccccgctggtccttggaggtcccaacactgc 148
            ||||||||| | |||||||||||| |||||||||||||||||||||||| ||||||||||
Sbjct: 2004 cagcctcccgctcactgtcacctcagtctccccgctggtccttggaggtaccaacactgc 2063
Query: 149  caagactcct-----tccttctcccctctgccccctctgcgcggcatctcagcagctgag 203
            ||||||||||     ||||||||||||||||||||||||||| |  ||||||  ||||||
Sbjct: 2064 caagactcctcttcttccttctcccctctgccccctctgcgccgtgtctcagtggctgag 2123
Query: 204  agggtggggggaagtggggggcattagctctcgaaggctcttttgtaaagtttttgtagtg 263
            ||||||||| ||| ||||||||||||||  | ||||||||||||||||||||||||||||
Sbjct: 2124 agggtgggg-aag-ggggggcattagctgccaaaggctcttttgtaaagtttttgtagtg 2181
Query: 264  atttttacaccacctgaataaagaatgaatgggcc 298
            |||||||| |||||||||||||||||||| |||||
Sbjct: 2182 atttttacgccacctgaataaagaatgaaggggcc 2216
```

图 1 EST-1 基因水平同源性比较

Fig.1 Blastn research in GenBank database of EST-1

```
BLASTN RESEARCH IN GENEBANK DATABASE:                          Score  E Value
dbj|AB028642.1|AB028642  Bos taurus mRNA for aquaporin-4-B, ...  240    4e-61
dbj|AB015947.2|AB015947  Bos taurus AQP-4 mRNA for aquaporin...  240    4e-61
dbj|AB028642.1|AB028642 Bos taurus mRNA for aquaporin-4-B, complete cds
 Length = 5641  Score =  240 bits (121), Expect = 4e-61
Identities = 182/201 (90%), Gaps = 1/201 (0%) Strand = Plus / Plus
Query: 249  gcttccctggtggctcagatggtaaagaatcctcctgcaatgtgggagacctgggtttga 308
            |||||||||||||||||||||||||||||||| |||||| || || || |||||||| ||
Sbjct: 4358 gcttccctggtggctcagatggtaaagaatccgcctgcattgcggaaggcctgggttcga 4417
Query: 309  tccctgggttgggaatatcccctgaaggaaggcttggcaacccactccagtattcttgct 368
            |||||||| |||||| |||||||| |||| ||| |||||| |||||||||||||||||||
Sbjct: 4418 tccctgggctgggaagatcccctggaggagggcatggcaatccactccagtattcttgct 4477
Query: 369  tggagaatccccatggacagaggaacctggtgggctacagtccatggggttgcaaagagc 428
            |||||||| ||||||||||||||| |||||||||||||||||| ||||||||||||||| 
Sbjct: 4478 tggagaat-cccatggacagaggagcctggtgggctacagtccctggggttgcaaagagt 4536
Query: 429  tggacatgactgagcgactaa 449
             ||||| |||||||| |||||
Sbjct: 4537 cggacacgactgagcaactaa 4557
dbj|AB015947.2|AB015947 Bos taurus AQP-4 mRNA for aquaporin-4-A, complete cds
 Length = 5700   Score =  240 bits (121), Expect = 4e-61
Identities = 182/201 (90%), Gaps = 1/201 (0%) Strand = Plus / Plus
Query: 249  gcttccctggtggctcagatggtaaagaatcctcctgcaatgtgggagacctgggtttga 308
            |||||||||||||||||||||||||||||||| |||||| || || || |||||||| ||
Sbjct: 4417 gcttccctggtggctcagatggtaaagaatccgcctgcattgcggaaggcctgggttcga 4476
Query: 309  tccctgggttgggaatatcccctgaaggaaggcttggcaacccactccagtattcttgct 368
            |||||||| |||||| |||||||| |||| ||| |||||| |||||||||||||||||||
Sbjct: 4477 tccctgggctgggaagatcccctggaggagggcatggcaatccactccagtattcttgct 4536
Query: 369  tggagaatccccatggacagaggaacctggtgggctacagtccatggggttgcaaagagc 428
            |||||||| ||||||||||||||| |||||||||||||||||| ||||||||||||||| 
Sbjct: 4537 tggagaat-cccatggacagaggagcctggtgggctacagtccctggggttgcaaagagt 4595
Query: 429  tggacatgactgagcgactaa 449
             ||||| |||||||| |||||
Sbjct: 4596 cggacacgactgagcaactaa 4616
```

图 2 EST-2 基因水平同源性比较

Fig.2 Blastn research in GenBank of EST-2

2.3　EST-3 序列分析　进入 GeneBank 和 dbEST,经相似性查找发现:EST-3 与任何已知基因或已知 EST 都不同源。将 EST-3 在 GeneBank 中登录,登录号为 AW990954。EST-3 基因序列见图 3。

分析发现 EST-3 序列中有 1 段 ORF,把这段 ORF 对应的氨基酸序列输入 Swissprot 查找相似性发现,其中有一段蛋白(47aa)与兔谷光甘肽 S 转移酶(GST)有 38%的同源性,如图 4。

```
TACTTAAGCT ATGCATCCAA CGCGTTGGGA GCTCTCCCAT ATGGTCGACC      50

TGCAGGCGGC CGCGAATTCA CTAGTGATTA AACAGATGAA GTGCTCCTTC     100
              EcoR I                    Il-1 primer 1
CAGGAGAAGG CTGAGAACAC AGCAGGGCCA GCTTTTGGCC TCAGACAGGA

TTATCCTTAT TACAGTAATA ATTGCCATTA TTAACAATTA TAGGAAATTA     200

TTCTTCACCT ATTTTCTAGC CATTAGCACT CAGCTGCCGT TGGGGAAGAG

ATTACAGTGT TGATTCTCAT TATCCTTCAA CACCTCCTTG CTGCCGGCCC     300

CAAAGTGCCT TGAGTGGGAG TTTATAACTC ATAGGATTCT TTCACGGGCA

GAAGGAAAAT TCTATCTTGG TTTCAAACGT TTCTCCAGTT TTAACATTAT     400

TTTTCCATCT ACACCGTGGA AGTAGTTCTT TCTTAAAATA CACCAGCCTC

CGGAACCTCT GAGAGGGTCT GTAGACTGAA AGACACACGA AGATGGAAAA     500

CAAAAGGTGA TATGCTTGGA GGTTCACAGG TGAGAGGTCA AGGGGGCAAG

GAACTATTGT CCAGTTGGTG ATTACTAACA CAGGATTGGA GCAACAAGTG     600
                                        IL-1 primer 2
GTGTTCTCCA AATCGAATTC CCGCGGCCGC CATGGCGGCC GGGAGCATGC
                EcoR I
```

图 3　序列 3 用 IL-1 引物钓取的 cDNA 片段 2 测序结果

Fig.3　The nucleotide sequence of the cDNA fragment 2 primed by hIL-1 β primers

GACAAC: EST-3(531 bp, 79～610); AACA: IL-1β primer; GAATTC: *Eco*R Ⅰ

```
sp|P46409|GTMU_RABIT GLUTATHIONE S-TRANSFERASE MU 1 (GST MU I)
Length = 218   Score = 28.2 bits (61), Expect = 4.3
Identities = 18/47 (38%), Positives = 24/47 (50%), Gaps = 1/47 (2%)
Query: 15   NWTIVPCPLDLSPVNLQAYHLLFSIFVCLSVYRPSQRFRRLVYFKKE 61
            N    VP  LD  P NL+ +H+ F      +S Y  S RF R+   F K+
Sbjct: 167  NRIFVPGCLDAFP-NLKDFHVRFEGLPKISAYMKSSRFIRVPVFLKK 212
```

图 4　EST-3 氨基酸水平同源性比较

Fig.4　Blastn research in swissprot database of EST-3

3　讨　论

目前,EST 主要用于寻找新基因和了解基因表达情况。通常先建立组织或细胞 cDNA 文库,从文库中随机取出足够量的 cDNA 克隆进行测序,测序结果进行生物信息学分析,即将所得的 EST 数据与 dbEST 等数据库的数据比较,确定哪些代表已知基因,哪些代表已知 EST,哪

些代表未知基因。然后进一步分析操作以得到未知基因或分析归类全部 EST 以获得组织或细胞的基因表达概况[3~5]。

本文采用人 IL-1β, HSV 1-2 TK 引物扩增鹿的 cDNA 文库,得到三个 EST,是在鹿的 cDNA 文库建立之后,利用手头现有的引物随意扩增得到的。得到的 EST 与引物来源基因无同源性,这与很多情况下的 PCR 结果相似,是一种非特异性扩增。

通过 PCR 非特异性扩增得到三个 cDNA 克隆,对之测序后去除载体序列得到三个 EST,均为两次测序结果。经 GeneBank、Swissprot 和 dbEST 数据库的同源分析表明,无已知基因或已知 EST 与得到的三个 EST 相同,说明它们均为未知基因 EST,我们已在 GenBank——dbEST 中进行登录。其中 EST-1 得自鹿肝组织 cDNA 文库,EST-2 和 EST-3 得自鹿血 cDNA。EST-1、EST-2 与已知基因高度同源,分别代表梅花鹿 α_2-抗纤溶酶(α_2-AP)同源 EST 和梅花鹿水通道蛋白-4(AQP-4)同源 EST。EST-1 钓自于鹿肝组织文库,与牛 α_2-抗纤溶酶高度同源,这与文献报道的 α_2-抗纤溶酶主要在肝脏中分泌表达相符合。来自鹿血的 EST-2 与水通道蛋白-4 同源也提示水通道蛋白基因在外周血中有一定表达。EST-3 与任何已知基因或已知 EST 均无同源性,推测它是鹿血中表达的未知基因,编码蛋白可能与鹿 GST 功能有关。

此外,在分析时发现,EST-1 3′端非编码区有标准的 polyA 加尾信号 AATAAA 和典型的多聚腺嘌呤核苷酸尾(polyA),可以推测其 3′端序列是完整的,而 5′端序列不完整。根据 cDNA 末端的快速扩增(RACE)方法[6],据此 3′端序列,我们设计了一 3′引物,用此引物作为下游引物,以 cDNA 文库 5′端通用引物为上游引物,进行 PCR 扩增,旨在钓取梅花鹿 α_2-抗纤溶酶(α_2-AP)全长 cDNA。目前,PCR 扩增工作正在进行中。

参 考 文 献

[1] Adams, M. D., J. M. Kelley, J. D. Gocayne *et al*. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 1991, **252**(5 301): 1 651 ~ 1 656.

[2] Boguski, M. S., M. J. Lowe, C. M. Tolstoshev. dbEST-database for "expressed sequence tags". *Nature Genet*, 1993, **4**(4): 332 ~ 333.

[3] 陆佳韵,王秀琴,吴旻. EST 及其应用. 生命科学, 1999, **11**(4): 186 ~ 188.

[4] 章卫平,曹雪涛,万涛等. 树突状细胞 cDNA 质粒文库的构建及其大规模随机体系的建立. 中国免疫学杂志, 1999, **6**(15): 245 ~ 249.

[5] Nelson, P. S., W. L. Ng, M. Schummer *et al*. An expressed-sequence-tag database of the human prostate: sequence analysis of 1168 cDNA clones. *Genomics*, 1998, **47**(1): 12 ~ 25.

[6] Frohman, M. A., M. K. Dush, G. R. Martin. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci*. USA, 1988, **85**(23): 8 998 ~ 9 002.