

猪 *TDRP1* 基因的电子克隆及生物信息学分析

张伟^{1,2}, 方梅霞³, 聂庆华^{1,2*}, 张细权^{1,2}

(1.华南农业大学 动物科学学院, 广东 广州 510642; 2.广东省农业动物基因组学与分子育种重点实验室, 广东 广州 510642; 3.暨南大学 医学院, 广东 广州 510632)

摘 要: 利用生物信息学和比较基因组学方法, 对猪 *TDRP1* 基因的结构、表达及功能进行分析, 结果表明: (1) 电子克隆获得猪 *TDRP1* 基因 cDNA 部分序列共 776 bp, 包括 119 bp 的 5'UTR、561 bp 的编码区和 96 bp 的 3'UTR, 共编码合成 187 个氨基酸多肽; (2) 猪 *TDRP1* 基因 CDS 序列与人、小鼠和大鼠的同源性分别为 85%、76.1%、77.1%。编码合成的猪 *TDRP1* 蛋白与人、小鼠、大鼠之间的同源性分别为 82.5%、71.1%和 70.1%; (3) 猪 *TDRP1* 蛋白的相对分子质量为 20 489.8, 不含信号肽, 为 1 个可溶性蛋白, 同时也是 1 个非分泌性蛋白, 存在 4 个二级结构区段; (4) 基于 cDNA 及 EST 数据库的检索显示, 猪 *TDRP1* 基因至少在卵巢和甲状腺等组织中表达。

关 键 词: 猪; *TDRP1* 基因; 电子克隆; 生物信息学

中图分类号: S828.8 文献标志码: A 文章编号: 1007-1032(2010)06-0672-06

Electronic cloning and bioinformatic analysis of the pig *TDRP1* gene

ZHANG Wei^{1,2}, FANG Mei-xia³, NIE Qing-hua^{1,2*}, ZHANG Xi-quan^{1,2}

(1.College of Animal Science, South China Agricultural University, Guangzhou 510642, China; 2. Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, Guangzhou 510642, China; 3. College of Medical, Jinan University, Guangzhou 510632, China)

Abstract: The structure, expression and function of pig *TDRP1* gene was predicted with bioinformatics and comparative genomics methods. Results showed that: (1) The electronically cloned pig *TDRP1* cDNA was 776 bp including a 119 bp 5'UTR, a 96 bp 3'UTR and an open reading frame (ORF) of 561 bp, which encoded a 187 amino acid peptide. (2) The CDS of pig *TDRP1* had 85%, 76.1% and 77.1% with that of human, mouse and rat, respectively. The translated pig *TDRP1* precursor had identities of 82.5%, 71.1% of identity homologous and 70.1% of identity homologous with its counterparts of human, mouse and rat, respectively. (3) The predicted molecular weight of pig *TDRP1* was 20 489.8 Daltons without signal peptide. The pig *TDRP1* might be a soluble and non-secretive protein, and probably forms four kinds of secondary structure. (4) Based on cDNA and EST library, the pig *TDRP1* gene expresses at least in ovary and thyroid gland tissues.

Key words: pig; *TDRP1* gene; electronic cloning; bioinformatics

培育出高繁殖力的家养动物, 一直都是养殖生产中待解决的问题。以往的研究重点趋向于雌性动物, 但只有品质优良的精子跟卵子结合, 加上后天的选育, 才有可能提升家养动物的繁殖力, 取得更加可观的经济效益。精子的发生是一个复杂的生理过程, 很大程度上决定于睾丸的发育状况。作为雄

性哺乳动物的生殖器官, 睾丸不仅是精子发生的场所, 而且是雄性激素产生的主要部位。有研究^[1-2]表明, 将雌性小鼠卵巢移植到雄性小鼠肾囊中, 对雄鼠精子生成及生殖机能无显著影响, 且移植后的卵巢功能可以保持长达 1 年^[3]的时间, 不过需要依靠雄性激素来维持功能的稳定性。可见, 睾丸对于动

收稿日期: 2010-06-24

基金项目: 农业部转基因生物新品种培育重大专项(2008ZX08006-005、2009ZX08009-145B); 粤港关键领域重点突破项目(2008A02)

作者简介: 张伟(1984—), 广东韶光人, 硕士研究生, 主要从事动物遗传育种研究, zwei618520@163.com; *通讯作者, nqinghua@scau.edu.cn

物繁殖起着不可替代的作用。Schultz等^[4]发现小鼠基因组中仅 4%左右参与了生殖细胞的特异性转录,而这些细胞中的 99%是减数分裂期及减数分裂后形成的。这预示着对精子发生的调控关键在于对减数分裂期的把握。近年来,研究人员发现了一个新的与精子发生相关的基因 *TDRP*(testis development-related protein) 或称 *C8orf42* (chromosome 8 open reading frame 42), 即睾丸发育相关蛋白。*TDRP* 基因在睾丸精原细胞中的表达很高,而精原细胞主要是由睾丸精子管上皮的原始生殖细胞经过多次有丝分裂形成的,与细胞减数分裂有关,说明 *TDRP* 基因跟细胞减数分裂有着一定的关联。*TDRP* 基因有 *TDRP1* 和 *TDRP2* 两种转录本,其中被认为起主要作用的是 *TDRP1* 转录本。当 *TDRP1* 基因缺失时,表现为精子生成功能停止^[5-7]。Wang等^[8]发现患有精子生成障碍症的男性 *TDRP1* 基因的表达量极低,接近于零水平。

迄今为止,关于 *TDRP1* 基因的研究主要集中在人和小鼠,针对其他动物的研究还鲜有报道。猪作为人类疾病动物模型中的重要成员和人类肉食

品的主要来源,对猪 *TDRP1* 基因的研究有助于深入了解诸如精子生成障碍等男性不育症的发病机制,同时也为提高猪繁殖性能奠定基础,因此,笔者采用生物信息学和比较基因组学方法,对猪 *TDRP1* 基因开展初步研究,以证实该基因的存在,分析该基因的结构、表达和生物学功能,为后续研究奠定基础。

1 材料与方法

1.1 蛋白质及 cDNA 序列获取

参考聂庆华等^[9]和方梅霞等^[10]的方法,从NCBI数据库^[11]中获取人等 11 个物种 *TDRP1* 的蛋白质序列及其相应 cDNA 序列(表 1)。将获得的蛋白质和 cDNA 用 FASTA 格式保存为文本文件。在所获得的高同源性序列中,只有人和小鼠的已明确命名为 *TDRP1* 基因,猕猴的推测为 *C8orf42* 基因(*TDRP1* 基因的另一命名),其余物种都只是预测蛋白质,还没有确切的命名。猪的 *TDRP1* 来源于电子克隆,即多条 EST 序列通过生物信息学软件拼接而来。

表 1 11 个物种 *TDRP1* 的蛋白质和 cDNA 序列
Table 1 Protein and cDNA sequences of *TDRP1* among 11 species

编号	中文名	英文名	拉丁文名	CDS 序列号	蛋白质序列号
1	狗	dog	<i>Canis familiaris</i>	XM_846720.1	XP_851813.1
2	牛	cattle	<i>Bos taurus</i>	XM_865548.3	XP_870641.3
3	猪	pig	<i>Sus scrofa</i>		
4	小鼠	house mouse	<i>Mus musculus</i>	BC145286.1	AAI45287.1
5	大鼠	Norway rat	<i>Rattus norvegicus</i>	NM_001100791	EDM08922.1
6	人	human	<i>Homo sapiens</i>	AY194292.2	AAO23974.2
7	猕猴	macaque	<i>Macaca mulatta</i>	XM_001118733.1	XP_001118733.1
8	狨猴	marmoset	<i>Callithrix jacchus</i>	XM_002756743.1	XP_002756789.1
9	斑马鱼	zebrafish	<i>Danio rerio</i>	XM_687868.4	XP_692960.1
10	非洲爪蛙	African clawed frog	<i>Xenopus laevis</i>	NM_001095454	NP_001088923.1
11	西部爪蛙	western clawed frog	<i>Xenopus (Silurana) tropicalis</i>	NM_001016270.2	NP_001016270.1

猪的 CDS 序列号由电子克隆;蛋白质序列号根据 CDS 翻译。

1.2 猪 *TDRP1* 基因的 cDNA 电子克隆

基于 NCBI 网站,利用人 *TDRP1* 基因 cDNA 编码区序列对猪 EST 文库进行 BLAST 分析,获取众多高度同源的猪 cDNA 克隆,主要有 368581 MARC(NCBI 序列号为 BI341215.1)、TOH 602768086.R1(BM658620.1)、OVR010062H03 (BW 965181.1)、scac0033.h.21 3prim(BX670425.2)、scac

0033.h.215prim(BX670426.2)、892264 MARC(CF 796048.1)、951463 MARC(CN162110.1)、952999 MARC (CN163491.1)、996750 MARC(CN165903.1)、UMC-pd10en3-016-a03(CO987410.1)、UMC-pd6end 2-010-a08 (CO993002.1)、1528306 MARC(DY 416075.1)、rpigcb_ 6170.y1(EW053808.2)、rfat0118_a6.y1(EW161089.2)、 ruio10_f4.y1(EW297528.2)、rnlu1806b_n23.y1

(EW455 561.2)、rsug30_n4.y1 (EW619691.2)、rthg24_f19.y1 (EW644128.2)、OVRT 10052E09 (FS690666.1)。利用 DNASTAR 软件的 SeqMan 程序进行序列拼接，分别向 5'和 3'端延伸，以尽可能地获得该基因的 cDNA 全长和猪 *TDRP1* 基因 cDNA 全长，对其开放阅读框 ORF 及翻译的蛋白质序列进行分析，并与其他 10 个物种进行同源性和系统聚类分析。

1.3 11 个物种 *TDRP1* 基因的同源性及系统进化分析

利用 DNASTAR 中的 MegAlign 程序对不同物种 *TDRP1* 的蛋白质及 CDS 进行 BLAST 分析，初步计算各物种序列间的同源性。再利用 MEGA4.1 程序^[12]计算物种间的遗传距离，并通过邻近归并法 (Neighbor-joining) 构建系统发生树。

1.4 猪 *TDRP1* 蛋白质结构及功能预测

进入 NCBI/BLAST 网站^[13]，将猪 *TDRP1* 蛋白质序列输入序列输入框内，选择“BLASTP”程序分析 *TDRP1* 蛋白质的功能属性。利用生物信息学网站^[14]分析猪 *TDRP1* 蛋白的物理性质，综合运用信号肽预

测网站^[15-16]和分析猪 *TDRP1* 可能的信号肽结构。再利用蛋白质二级结构预测网站^[17]分析 *TDRP1* 蛋白可能的二级结构。根据蛋白质结构域预测网站^[18]分析 *TDRP1* 蛋白可能存在的功能结构域。

1.5 猪 *TDRP1* 基因的组织表达

利用前面获得的猪 *TDRP1* 基因 EST 序列信息，根据这些 EST 序列的组织来源，综合分析猪 *TDRP1* 基因在不同组织中的表达情况。

2 结果与分析

2.1 猪 *TDRP1* 基因部分 cDNA 序列及开放阅读框 (ORF) 分析

基于猪 EST 文库的电子延伸，得到 2 条拼接序列，1 条长度约为 980 bp，另 1 条为 776 bp。通过与人的蛋白序列进行对比，确认猪 *TDRP1* 基因的 cDNA 为 776 bp 的这 1 条。得到猪 *TDRP1* 基因部分 cDNA 序列，其中包括 119 bp 的 5'UTR、96 bp 的 3'UTR 以及 561 bp 的编码区，共编码合成 187 个氨基酸多肽链(图 1)。

```
gtctagccgctccgtccttgcaggagccgcagctcgcagtcctcctcctcgggccagctgaggagggcgcgccacggagc
caggagcagcaggagcgcacggaccctgaagccgccccggagcATGTGGAAGCTGGGCAGGAGCCGGGTGCTTCTGGATGAGCC
M W K L G R S R V L L D E P
CCCAGGAGGAAGAGGGCCTGCCAGGAGGGCCGCGTCGGCTGCTGCCCCCGCTGCCGCCGCGCGCAGGTGCAGGGGGCGAGT
P E E E E G L P G G P P S A A A P A A A A A Q V Q G A S
TTCGGGGTGGAAAGAAGTGACTTCTCTGTTAAACAAGGATGATGAACAGCATCTGCTGAAAGATGCAAGTCCCCAAGTCC
F R G W K E V T S L F N K D D E Q H L L E R C K S P K S
AGAGGAACAACTTACGAGTGAAGAAGAGTTGAAGACAGAGAAGAAGTCTGGATTTGGGACAATTTGGTTTAAACAGAAC
R G T N L R V K E E L K T E K K S G F W D N L V L K Q N
GTGCGCTAAGAAACCAGATGAGATTGAAGTTGGGAGCCTCCAAAACCTTGTCTTGAAGATGTAGCCGCCGACTCAGGGGAC
V Q S K K P D E I E G W E P P K L V L E D V A A D S G D
CCCCGAGTACCGTCCATCCCGCCAGGCTGGGAGGAGGATGCCAAGGGCTCCACCAAGTACACCAGCCTGGCCGGCGCGGGG
P P S D R P S R P G W E E D A K G S T K Y T S L A G A G
AGCAGCTCCCGCTGGAGCCTCAAGTCGGCCGGAAGCTGGTCAGCATTCCCGGCAGAGCAAAGGCCACCTGACGGATAACTGG
S S S R W S L K S A G K L V S I R R Q S K G H L T D N W
GAAGAGTTGGAGTGAcgtgtccgaggagacacacagctcccaagtgtttccaaacaaaatttgcctaattttgcacctctgt
E E L E *
gtcatttcattcagattctgagtttgc
```

小写碱基为 5'UTR 和 3'UTR 区域序列，大写碱基为开放阅读框序列；阅读框下大写英文字母为编码氨基酸序列，*为翻译终止信号。

图 1 电子克隆获得的猪 *TDRP1* 基因 cDNA 序列及预测的氨基酸序列

Fig. 1 Pig *TDRP1* cDNA and its predicted amino acid sequences by ePCR

2.2 不同物种间 *TDRP1* 基因的同源性分析

根据人蛋白质序列 BLAST 分析，11 个物种间的 *TDRP1* 同源性差异较大，水生鱼类斑马鱼与其

余物种的同源性最低为 16.7% (牛)，最高为 29.7% (西
部爪蛙)。哺乳动物(人、猕猴、狨猴)间的同源性最
高，达到 94% 以上。啮齿类的大鼠和小鼠同源性为
90.5%。另外，两栖类的非洲爪蛙和西部爪蛙同源
性也超过 90% (表 2)。猪 *TDRP1* 与其余 10 个物种

的同源性除非洲爪蛙外，都高于 70%。比较 8 个哺乳动物(人、猕猴、狨猴、狗、猪、小鼠、大鼠、牛)的氨基酸后可以清楚地看到它们之间表现出的明

显差异，其中狗的蛋白质序列比其余物种要长(图 2)。这是不是由于为了适应环境、保持繁殖优势而自行进化造成的，还没有定论。

表 2 不同物种 *TDRP1* 基因 CDS 序列及氨基酸序列的同源性比较结果
Table 2 The homology of *TDRP1* CDS and its amino acid sequences among different species

Table with 12 columns (编号 1-11) and 12 rows (编号 1-11) showing homology percentages between species. Values range from 4.6% to 94.2%.

编号 1~11 分别代表非洲爪蛙、牛、狗、小鼠、人、猕猴、狨猴、大鼠、猪、西部爪蛙、斑马鱼等 11 个物种；对角线右上部分数值为氨基酸同源性，对角线左下部分数值为 CDS 同源性。

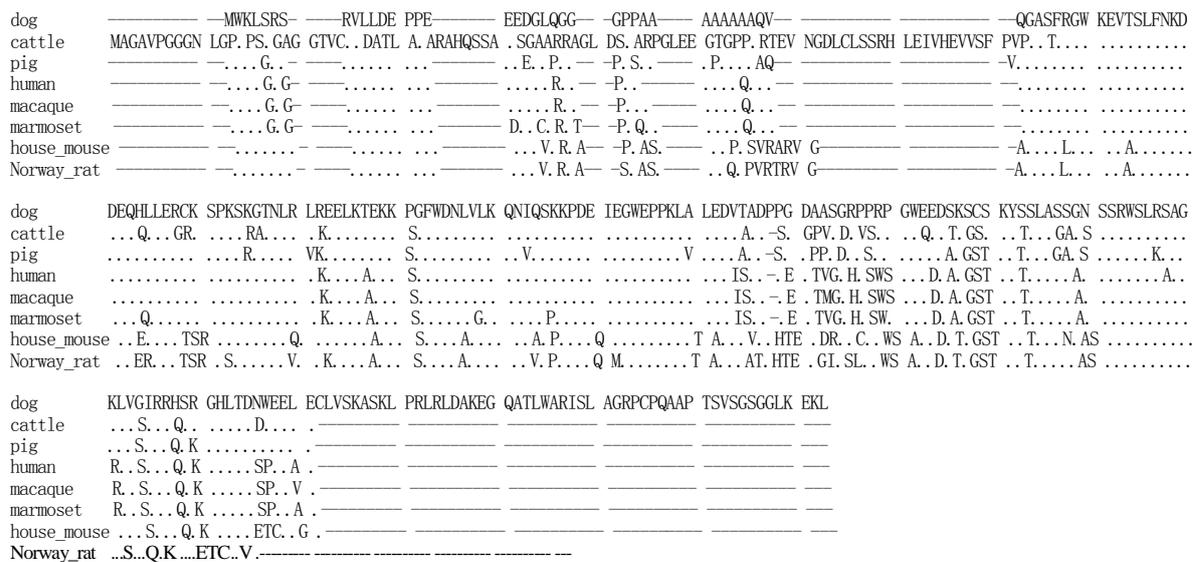


图 2 8 种哺乳动物 *TDRP1* 氨基酸序列 BLAST 分析结果

Fig.2 Homologous comparison of amino acid sequences of *TDRP1* among 8 species

TDRP1 基因的 CDS 序列 BLAST 后显示出类似的结果。斑马鱼的 CDS 序列与其他物种的同源性极低，哺乳动物间除牛之外，同源性几乎均高于 70%，尤其是人、猕猴与狨猴间，大鼠和小鼠间都大于 90%。此外，非洲爪蛙与西部爪蛙的同源性也超过 90%，达到 94.2%(表 2)。目的物种猪的 CDS 序列与人、猕猴、狨猴的 CDS 序列同源性都大于 80%，跟其余物种的同源性均比较低，为 40%~80%。

2.3 11 个物种 *TDRP1* 基因的系统进化分析

聚类分析结果显示，11 个物种被划分为界限清晰的 3 大类群，其中 8 个哺乳动物(人、猕猴、狨猴、狗、猪、小鼠、大鼠、牛)紧密聚合为一类，目的物种猪的进化程度与狗最接近；两栖类(非洲爪蛙、西部爪蛙)归为一类，斑马鱼作为水生鱼类单独为一类(图 3)。

法获得的序列确为猪 *TDRP1* 基因序列。此外,利用本研究结果,通过 RT-PCR 技术及分子克隆技术,本课题组对猪 *TDRP1* 基因进行了试验验证,并成功地在猪睾丸组织中获得了该基因的 CDS 全序列(GenBank 序列号 HM590584),与预测序列相似性达 98% 以上,进一步确认了 *TDRP1* 基因存在于猪体内。同时,本课题组对猪 *TDRP1* 基因进行了组织性表达规律、变异及标记性状关联分析,取得了一定的进展(成果论文待发表),为后续研究该基因的功能奠定了基础。

参考文献:

- [1] 彭南妮,薛立群,邓治邦,等. 卵巢移植对雄性受体鼠精子生成及生殖能力的影响[J]. 湖南农业大学学报:自然科学版, 2010, 36(1): 73-76.
- [2] 袁安文,许道军,王乃东,等. 移植时间对小鼠卵巢卵母细胞在雄性体内生长发育的影响[J]. 湖南农业大学学报:自然科学版, 2008, 34(1): 56-60.
- [3] Candy C J, Wood M J, Whittingham D G. Restoration of a normal reproductive lifespan after grafting of cryopreserved mouse ovaries[J]. Human Reproduction, 2000, 15(6): 1300-1304.
- [4] Schultz N, Hamra F K, Garbers D L. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets[J]. Proc Natl Acad Sci USA, 2003, 100(21): 12201-12206.
- [5] Brower J V, Rodic N, Seki T, et al. Evolutionarily conserved mammalian adenine nucleotide translocase 4 is essential for spermatogenesis[J]. J Biol Chem, 2007, 282(40): 29658-29666.
- [6] Dix D J, Allen J W, Collins B W, et al. Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility[J]. Proc Natl Acad Sci USA, 1996, 93(8): 3264-3268.
- [7] Dickins R A, Frew I J, House C M, et al. The ubiquitin ligase component Siah1a is required for completion of meiosis I in male mice[J]. Mol Cell Biol, 2002, 22(7): 2294-2303.
- [8] Wang X, Jiang H, Zhou W, et al. Molecular cloning of a novel nuclear factor, TDRP1, in spermatogenic cells of testis and its relationship with spermatogenesis[J]. Biochem Biophys Res Commun, 2010, 394(1): 29-35.
- [9] 聂庆华,方梅霞,李伟民,等. 鸡 *pATGL* 基因的电子克隆及比较基因组分析[J]. 华南农业大学学报, 2008, 29(2): 94-98.
- [10] 方梅霞,周晓宁,沈栩,等. 猪繁殖候选基因 *ERK2* 的电子克隆与生物信息学分析[J]. 广东农业科学, 2009(10): 11-15.
- [11] 李建华,陈锦华. 利用NCBI相关资源电子定位基因的方法和实践[J/OL]. 孝感学院学报, 2003, 23(6): 57-59[2003-11-06]. <http://www.ncbi.nlm.nih.gov>.
- [12] 于季芝. 基于距离的进化树构建算法研究[D/OL]. 哈尔滨: 哈尔滨工业大学, 2007[2007-07-25]. <http://www.megasoftware.net/mega41.html>.
- [13] 王哲,黄高升. NCBI的数据库资源及其应用[J/OL]. 生命科学, 2002, 14(1): 59-62[2002-02-15]. <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- [14] 冯雷. 人肿瘤坏死因子h-TNFF-a的生物信息学分析[D/OL]. 天津: 天津科技大学, 2009[2009-01-25]. <http://www.expasy.org/tools/protparam.html>.
- [15] 刘涛,申明霞. 五指山小型猪 *PMSA6* 基因cDNA克隆及生物信息学分析[J/OL]. 海南大学学报, 2009, 27(2): 134-143[2009-06-15]. <http://www.cbs.dtu.dk/services/SignalP>.
- [16] 孙平楠,周小玲. 信号肽生物信息学分析在 *Neurospora crassa phyA* 基因鉴定中的应用[J/OL]. 南方医科大学学报, 2009, 29(6): 1098-1101[2009-06-20]. http://bp.nuap.nagoya-u.ac.jp/sosui/sosuisignal/sosuisignal_submit.html.
- [17] 邹淑雪. 基于支持向量机的蛋白质结构域预测方法研究[D/OL]. 长春: 吉林大学, 2009[2009-06-02]. <http://bioinf4.cs.ucl.ac.uk:3000/psipred>.
- [18] 毕燕玲,宁秋. 癌症研究的生物信息学资源[J/OL]. 现代生物医学进展, 2010, 10(11): 2174-2186[2010-07-25]. <http://www.ebi.ac.uk/Tools/InterProScan>.
- [19] Ciccarelli F D, Doerks T, von Mering C, et al. Toward automatic reconstruction of a highly resolved tree of life[J]. Science, 2006, 311(5765): 1283-1287.
- [20] la Cour T, Kiemer L, Mølgaard A, et al. Analysis and prediction of leucine-rich nuclear export signals[J]. Protein Eng Des Sel, 2004, 17(6): 527-536.

责任编辑:王赛群
英文编辑:罗文翠