

青鳉DNA聚合酶 基因的电子克隆与分析

易乐飞, 周向红, 刘楚吾

(1. 淮海工学院海洋生物技术重点建设实验室, 江苏连云港222005; 2. 广东海洋大学水产学院, 广东湛江524025)

摘要 [目的] 为了进行青鳉DNA聚合酶 基因的电子克隆与分析的试验。[方法] 利用生物信息学手段克隆了青鳉DNA聚合酶 (OPd) 基因全长cDNA序列, 并对其启动子和编码蛋白结构进行分析。[结果] 该cDNA序列的全长1 716 bp, 编码含338个氨基酸残基的蛋白。OPd具有14个外显子和13个内含子。OPd启动子上发现了SP1等一些转录因子的潜在结合位点。[结论] OPd蛋白序列与其他脊椎动物的蛋白序列具有高度一致性。

关键词 青鳉; DNA聚合酶; 电子克隆; 生物信息学

中图分类号 Q785 文献标识码 A 文章编号 0517-6611(2007)34-11065-03

In Silico Cloning and Analysis of *Oryzias latipes* DNA Polymerase Gene

Yi Le-fei et al (Jiangsu Key Laboratory of Marine Biotechnology, Huaihai Institute of Technology, Lianyungang, Jiangsu 222005)

Abstract Using in silico cloning, we got the *Oryzias latipes* DNA polymerase (*OPd*) complete cDNA. The cDNA is 1 716 bp length and contains complete ORF frame. *OPd* contains 14 exons and 13 introns, encoding a putative protein of 338 amino acids. Some potential transcription factor binding sites are found in the promoter, such as SP1 binding site. *OPd* has high identity with that of other vertebrate. In conclusion, the *OPd* is highly consistent with the protein sequence of other vertebrate animals.

Key words *Oryzias latipes*; DNA polymerase beta; in silico cloning; Bioinformatics.

DNA聚合酶 (DNA polymerase beta, Pol) 属DNA聚合酶X家族成员, 是一种组成型表达的持家基因, 主要参与DNA碱基切除修复和DNA复制等过程^[1]。目前, 已在多种哺乳动物、两栖动物和鱼类中克隆了该基因^[2-5], 然而尚无青鳉 (*Oryzias latipes*) Pol 的相关报道。为此, 笔者利用生物信息学手段克隆了青鳉DNA聚合酶 (OPd) 基因全长cDNA序列, 并对其启动子和编码蛋白结构进行了分析。

1 材料与方法

1.1 材料 以青鳉为研究材料。

1.2 方法

1.2.1 电子克隆技术。以斑马鱼 Pol 的氨基酸序列对 GenBank 中青鳉EST数据库进行 Blastn 检索, 将检索到的高度匹配的青鳉EST序列用 CAP3^[6] 组装成 contig, 用 contig 检索青鳉DNA数据库得到了含 OPd 的 gDNA 序列。使用 Spidey^[7] 将青鳉gDNA与脊椎动物 Pol 编码区的一致序列进行比对, 结合内含子/外显子剪接位点信息以及 polyadq^[8] 的预测结果, 获得了完整的 OPd 的 cDNA 序列。

1.2.2 编码蛋白的相似性和一级结构分析。分析该 cDNA 序列行 ORF, 并以推导的氨基酸序列对 GenBank 中的非冗余的蛋白质数据库进行 Blastp 分析, 判断该蛋白是否为新序列, 并且推导蛋白功能。利用 ProtParam^[9] 进行编码蛋白的基本理化特性分析。使用 Clustal W^[10] 进行多序列比对。

1.2.3 启动子序列分析。选取起始密码子上游 1 kb 和下游 200 bp 的 gDNA 序列进行启动子分析。CpG 岛、转录起始位点和转录因子结合位点分析分别使用 CpGHot^[11]、NNPP^[12] 和 CONSITE^[13]。

2 结果与分析

2.1 青鳉DNA聚合酶 基因cDNA序列的电子克隆结果 用斑马鱼 Pol 的氨基酸序列对青鳉的 dbEST 进行 tBlastn 分析, 发现有 5 条青鳉EST 与之高度匹配 (E value < 10⁻⁵⁰)。将

它们进行组装, 得到了长 992 bp 的 contig。用该 contig 对青鳉 gDNA 数据库进行 Blastn 检索, 获得了含 OPd 的高度匹配 (E value = 6e⁻⁶⁶) 的 gDNA (BAAF03013489)。使用 Spidey 分析和 ORF 分析后, 发现该 contig 仅携带部分编码序列 (前 12 个外显子和外显子 13 的部分序列)。所以, 笔者使用 Spidey 将脊椎动物 Pol 编码区 3 端的一致序列与 BAAF03013489 比对, 再结合 GT-AG 规则以及 polyadq 预测的加尾信号, 得到了缺失的 3 端, 最后获得了长 1 716 bp 的 OPd 的 cDNA 完整序列 (图 1)。

>cDNA of *OPd* β

```
GCAGCTTGTAAATCGCTTTCGCACGCAAAATATTACTTGGGAATTTTTTTTAAAAACTACAATAGTGA
ATTACATCCACTATGAGCAAAAAGAAAAGCACCACAGGAATCTCCAAACGAAGGAATAACGGACTTTC
TCATCGAGTTGGCCAACTACGAGAAAAATGTCAACAGGGCAATACACAAGTACAATGCTTACAGGAA
AGCAGCATCTACCATTTCTAAATACCCAAGCAAGATTAGGAACGGAGAAGAAGCAAAAAAAGTGGAT
GGTGTGGTCCAAAATAGCAGAAAAGATTGATGAATTTCTACAACTGGCAACTAAGAAAAGTACAG
AAAAGATCCGAAATGATGATACCAGCTCTTCCATCAACTTCTCCTACTAGAGTTACTGGAATTTGGCCC
TGCTGCCCGCAGGAAGTTTTTGAAGAAGGGGTGAAGACATTAGATGATTTGAAAAAGGTTGAGCAC
AACTTAACCACCATCAAAAAGATTGGACTCAGGTACTTTGAAGAGTTTGAGAAAAGGATCCCAAGAG
CTGAAATGGAACAAATGGAGGCTCTTATCATTGCTGAGTTGGAGAAAATTTGATCCAGAATATATTGG
ACAATATGTGGAAGCTACAGAAGAGGTGCAGCGTGGAGTGGTATATTGATATTTTGGCTGACCCAC
CCAACTACACCTCGCAGACTGAGAAACAGCCCAAGCTCCTCCATGCTGTGGTTGAACACTTGGAGT
CCATTGGCTTTGTGACTGACACTCTGTCCAAAGGAGACACCAAGTTCATGGGTGCTGCCAACTGCA
GCAGAGCAATGATGATGAGGATGAGGAAGAATACCTTCACAGCGGTATTGACATTAGTTGATGCC
AAAGACCAGTACTACTGTGGAGTTTTGTATTTCACTGGAAGTGCATCTTTAACAAAAATATGAGAG
CTCATGCTTTGGAGAAGGGCTTCACTCTTAACGAATACACCATACGACCAGTGGTGTCACTGGTGT
GGCAGGAGAACCTCTGCCGGTGGCAGCGAGAGACATTTTTGAATACATCCAGTACAAGTTTAAA
GAGCCCAAGATCGCAGCCAGTAAAGGACAACAGTAAAGCCTCCAGCTAAAAGTATCTTTGAAATAG
AGCTGAACTGGAAAAAAAGCTGCGTTTTATGTTACTTTTTGAATAGTTTCGCTTGTGATACAACAA
AGGAACTACCGAGCGTTTTATTTAATGACTCAATAATGTTTTAATTGAGAAATGCACAGATTTGGA
GAAAAAATCTTTCTGGATGTGATCACTTTATTTGATGCTTATGAATCAATGTAAAAATTTGACCT
GTTTGGCTTTTCAATTTATCCCAAAGTGGTATTATCTTCTGAAACATGAGTTTGTTCAGAGA
AAAGTTTGTCTTATCATGATGACATGTTTTCTTTATTTGATTTACACTTTTTTTCTTGTGGGTT
ATTTTTTTTACATCCTTTATTTGGGGCAGACTGTATAGAAATGTTACAAACATACAATGTGATAGTT
TCAGTTTCCACTGTACTACTAAATTTCTAAAAAATTAAGTGAATCTTAAATATGTTTTTTTTTAA
TGTGCTTAAAAAAGTGAATCCGCCATGATTTTTTTGTTATGAATAAGCAATATGCAGTAGATCCCA
GTAATTTATCTTTGCTTTCAACATGATAAATATGTGTCA
```

图1 青鳉DNA聚合酶 基因的cDNA序列

与 kozak 规则相比, 该 cDNA 的第一个 ATG 密码子下游的 +4 位不是 G, 但是 -3 位是 A, 且上游 3 个读码框中都存在终止密码子, 再结合推导的氨基酸序列 N 端同源性比较结果, 可以认定第一个 ATG 是该基因的起始密码子。在 1 703 ~ 1 708 位存在典型的 Poly-A 加尾信号 (AATAAA), 而且其后还有由 2 个核苷酸组成的 CA 结构, 因此可以认为该试验电

作者简介 易乐飞 (1975-), 男, 湖北荆门人, 讲师, 从事分子遗传学方面的研究。

收稿日期 2007-06-28

子克隆的 QPd 的cDNA 序列完整,其5 UTR 长79 bp,3 UTR 长620 bp,CDS 长1 017 bp,编码蛋白质长338 个氨基酸残基。
2.2 青鳉DNA 聚合酶 的基因结构 将 QPd 的cDNA 序列与 BAAF03013489 进行对比,结果表明 QPd 全长约 4.3 kb,含有14 个外显子和13 个内含子,所有的外显子与内含子边界完全符合经典的GT- AG 规则,最长和最短外显子

分别是外显子14 和外显子6(图2)。外显子12 比人和斑马鱼分别多9 个和3 个核苷酸;由于5 UTR 和3 UTR 的相对保守性较低,外显子1 和14 的大小与人和斑马鱼的不符;此外,其他外显子的大小与人、斑马鱼完全一致。13 个内含子在不同物种中的保守性很差。该基因在不同物种中的大小差异主要由这些内含子引起。

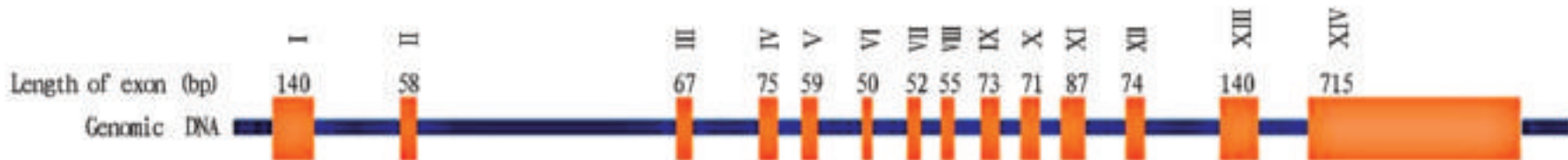


图2 青鳉DNA 聚合酶 的基因结构

2.3 启动子及转录因子结合位点 对选取的gDNA 序列进行C+ G 含量分析,结果表明整体C+ G 含量仅为36.8%,但在紧靠起始密码子上游104 ~176 bp 区间存在 CpG 岛,该区域C+ G 含量高达56.2%,说明紧靠起始密码子上游存在可能的启动子区域。NNPP 结果显示,在起始密码子上游- 117、- 126 和- 135 bp 等处可能存在多个转录起始位点。青鳉 QPd 的 EST 序列5 最末端的位置信息也反映出了类

似的结果,可能存在多个转录起始位点;其中,AM56108 和 AM50773 EST 序列的5 末端位置相同,所以可标记该转录起始位点为+1,当然并不能肯定该处是主要转录起始位点,但它至少比其他转录起始位点会更有优势。通过比较青鳉和斑马鱼的启动子序列,发现与其他持家基因一样在紧靠转录起始位点上游区域没有发现典型的TATA box 和CAAT box,但含有SP1 和Df3 等转录因子结合位点(表1)。

表1 青鳉DNA 聚合酶 基因启动子上的转录因子结合位点

转录因子	结合序列	起点	终点	得分	链	转录因子	结合序列	起点	终点	得分	链
FREAC 7	ATTTTATA	- 330	- 323	5.092	+	Ghf	CCCGCCCCCA	- 77	- 68	5.368	-
Df3	AAAATG	- 323	- 318	4.076	+	MZF_1-4	CCCCCA	- 73	- 68	6.469	-
FREAC 3	AAAATGTA	- 323	- 316	5.045	+	Df3	AACGTG	- 31	- 26	4.076	+
AP2alpha	GCCCCGCC	- 79	- 71	5.927	+	Df3	CACCTT	- 6	- 1	4.076	-
SP1	GCCCCGCC	- 79	- 70	8.598	-						

2.4 编码蛋白的相似性和一级结构 以推导的氨基酸序列对 GenBank 中的非冗余的蛋白质数据库进行 Blastp 分析,发现斑剑尾鱼、斑马鱼、非洲爪蛙、鼠、牛、狗和人等多种物种的 Pd 与青鳉具有极高度相似性(Identities > 77%,Positives > 90%),但在青鳉的蛋白质数据库没有找到与之高度相似的

序列,所以可以肯定该研究电子克隆的序列是一条新序列,且推导的蛋白就是青鳉的 DNA 聚合酶。该蛋白预测的分子量为38.7 kDa,其碱性氨基酸、酸性氨基酸、疏水性氨基酸和极性氨基酸的含量分别为52、55、102 和80 个,等电点为6.35。

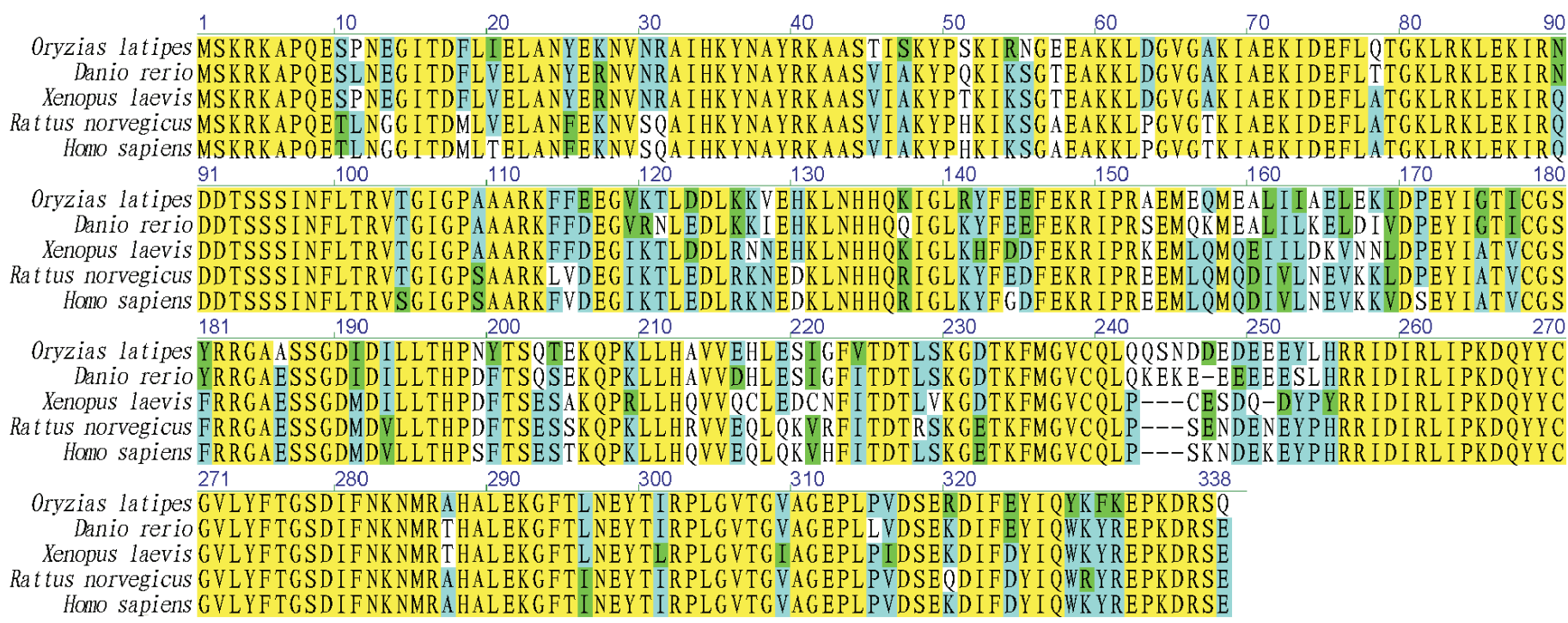


图3 青鳉、斑马鱼、非洲爪蛙、大鼠和人 Pd 的氨基酸序列比对结果

图3 表明,青鳉 QPd 比人、鼠和牛多3 个氨基酸残基,比非洲爪蛙和西部爪蛙多4 个氨基酸残基,比斑剑尾鱼和斑马鱼多1 个氨基酸残基。这些变化全部起因于外显子12 上的密码子数目的差异。青鳉 QPd 与斑马鱼、非洲爪蛙、大鼠和人的序列一致性(identity) 分别为87%、79%、77% 和

77%;若考虑功能相似的氨基酸之间的保守替换,那么青鳉 QPd 与它们的序列相似性(similarity) 分别为93%、89%、88% 和87%。

3 讨论

人、鼠和牛 Pol 的转录起始位点的研究表明,Pd 存在

多个转录起始位点,其中有1~2个为主要转录起始位点^[14-16]。这种多转录起始位点现象在其他持家基因中也较常见,例如人的 *H-ras*、*ADA* 和 *PGK*、小鼠的 *HPRT* 和 *DHFR* 等^[15]。该研究的分析结果也显示出了类似的结果,青鱼 β QPol 具有多个转录起始位点。

为了保证 DNA 的正常复制和修复,细胞内 Pd 的水平必须受到严格的调控。若 Pd 表达过低、突变或缺失,将造成 DNA 修复功能的缺陷;若 Pd 表达绝对或相对升高,将干扰或取代细胞内其他聚合酶的正常功能,导致遗传不稳定的发生^[17]。基因表达调控的最早和最中心的调控集中在转录起始水平,因此笔者分析了 QPol 的转录因子结合位点。为了避免假阳性位点,该研究采用进化足迹法分析。由于选择压力的作用,一段序列高度保守则说明该序列是有功能的,这样通过比较两种或以上的序列,就可以发现进化中的保守元件^[13]。该研究比较了青鱼和斑马鱼的启动子序列,结果显示 QPol 转录起始位点上游缺乏典型的 TATA box 和 CAAT box,但存在 SP1、Dof3 和 AP2 α 等转录调控因子的潜在结合位点。在许多缺乏 TATA box 的启动子中,SP1 对形成有效的转录起始复合物有重要的作用。缺乏 TATA box 的启动子通常含有多个 SH 元件。SP1 首先与启动子区富含 GC 的序列结合,接着 SP1 上富含谷氨酰胺的结构域与 TBP、hTAFII130 结合,C 端则与 hTAFII55 结合,通过蛋白与蛋白之间的多个相互作用,有效地将 THID 招募到核心启动子上,而且能稳定 THID 与核心启动子的结合,从而起始转录^[18]。青鱼也可能存在类似的机制,其他的转录因子可能参与诱导表达。

脊椎动物的 Pd 由2个结构域构成,其中一个分子量为 8 kDa 的 N 端结构域具有单链 DNA 结合和裂解酶活性,另一个分子量为 31 kDa 的 C 端结构域具有聚合酶活性。2 个结构域由 82~86 位点上的氨基酸残基相连^[11]。许多对酶活性非常重要的氨基酸残基在 QPol 中高度保守。例如,对裂解酶活性较重要的是保守的 H34、K35 和 Y39;一个潜在的结合 dNTP 的残基是保守的 K72,该残基可能通过对脱碱基位点的亲核攻击而参与裂解反应;能与聚合酶活性中心的 Mg^{2+} 结合并参与核苷酸转移反应的是保守的 D190、D259 和 D279;能稳定核苷酸结合口袋并促使引物、 Mg^{2+} 、dNTP 和其他活性氨基酸残基之间进行适当排列的是保守的 Y268;能与引入的 dNTP 作用并有助于错配核苷酸松动的是保守的 N282;与新生碱基对发生氢键和范德华力作用的是保守的 R286;与模板核苷酸发生氢键和范德华力作用的是保守的 K283^[15,19]。此外,青鱼与斑马鱼、非洲爪蛙、鼠和人等多种生物的氨基酸序列高度保守。所有这些一致性和相似性表明 QPol 结构上也可以划分为两个不同的功能区域,QPol 所执行的功能与哺乳动物一致。

不同生物 Pol 的氨基酸序列比对结果显示,在 242~251 区域出现了序列长度和组成的高度变异。这段高度变异区域位于名为 Loop II 的环中。Loop II 对维持聚合活性和聚合精确性具有重要作用。但只要 Loop II 含有 3 个或 3 个以上

的氨基酸残基就足以维持聚合活性,且维持聚合活性与 Loop II 的序列组成无关;Loop II 也可能折叠成一个控制 dNTP 进入活性中心的闸门,从而控制催化的精确性。这个闸门受到 Loop II 长度和可动性的影响^[20]。由此可见,青鱼 β DNA 聚合酶在 242~251 区域的变异并不会影响酶的聚合活性,但催化的精确性可能与其他物种存在差异。笔者利用电子克隆技术成功地克隆了青鱼 β DNA 聚合酶基因 cDNA 序列,该序列得到了青鱼 β EST 和 gDNA 数据的双重支持,而且推导出来的氨基酸序列也与其他脊椎动物的高度一致,表明得到的 cDNA 序列是真实存在的、有效的。与传统的基因克隆相比,电子克隆显现出了明显的优势。随着更多生物基因组测序的开展以及各类数据库的不断完善,电子克隆技术必将成为鱼类新基因克隆的重要手段。

参考文献

- [1] IDRIS H T, AL-ASSAR O, WILSON S H. DNA polymerase beta[J]. *Int J Biochem Cell Biol*, 2002, 34(4): 321-324.
- [2] SENGUPTA D N, ZMLDZKA B Z, KUMAR P, et al. Sequence of human DNA polymerase beta mRNA obtained through cDNA cloning[J]. *Biochem Biophys Res Commun*, 1986, 136(1): 341-347.
- [3] ZMLDZKA B Z, SENGUPTA D, MAISUKAGE A, et al. Structure of rat DNA polymerase beta revealed by partial amino acid sequencing and cDNA cloning[J]. *Proc Natl Acad Sci US A*, 1986, 83(14): 5106-5110.
- [4] REICHENBERGER S, PFEIFFER P. Cloning, purification and characterization of DNA polymerase beta from *Xenopus laevis*-studies on its potential role in DNA end joining[J]. *Eur J Biochem*, 1998, 251(1/2): 81-90.
- [5] OEHLERS L P, HEATER S J, RAINS J D, et al. Gene structure, purification and characterization of DNA polymerase beta from *Xiphophorus maculatus*[J]. *Comp Biochem Physiol C Toxicol Pharmacol*, 2004, 138(3): 311-324.
- [6] HUANG X, MADAN A. CAP3: A DNA sequence assembly program[J]. *Genome Res*, 1999, 9(9): 868-877.
- [7] WHEELANS J, CHURCH D M, OSTELL J M. Spidey: a tool for mRNA to genomic alignments[J]. *Genome Res*, 2001, 11(11): 1952-1957.
- [8] TABASKA J E, ZHANG M Q. Detection of polyadenylation signals in human DNA sequences[J]. *Gene*, 1999, 231(1/2): 77-86.
- [9] JOHN M W. The proteomics protocols[M]. US: Humana Press, 2005: 571-607.
- [10] FUKAMI KOBAYASHI K, SAITO N. How to make good use of CLUSTAL W[J]. *Tanpakushitsu Kakusan Koso*, 2002, 47(9): 1237-1239.
- [11] LARSEN F, GUNDERSEN G, LOPEZ R, et al. CpG islands as gene markers in the human genome[J]. *Genomics*, 1992, 13(4): 1095-1107.
- [12] OHLER U, LIAO G C, NIEMANN H, et al. Computational analysis of core promoters in the *Drosophila* genome[J]. *Genome Biol*, 2002, 3(12): 1-2.
- [13] SANDELLI N A, WASSERMAN W W, LENHARD B. ConSite: web-based prediction of regulatory elements using cross-species comparison[J]. *Nucleic Acids Res*, 2004, 32: 249-252.
- [14] WIDENS G, KEDAR P, WILSON S H. Human beta-polymerase gene. Structure of the 5' flanking region and active promoter[J]. *Journal of Biological Chemistry*, 1988, 263(32): 16992-16998.
- [15] YAMAGUCHI M, HIRSE F, HAYASHI Y, et al. Murine DNA polymerase beta gene: mapping of transcription initiation sites and the nucleotide sequence of the putative promoter region[J]. *Mol Cell Biol*, 1987, 7(5): 2012-2018.
- [16] CHEN K H, WOODT H E F, et al. The bovine DNA polymerase beta promoter: cloning, characterization and comparison with the human core promoter[J]. *Gene*, 1995, 164(2): 323-327.
- [17] 王谷亮, 余应年. DNA 聚合酶可能参与了更广泛的细胞反应并引起遗传不稳定[J]. *中国病理生理杂志*, 2002, 18(4): 427-432.
- [18] EMAMI K H, BURKE T W, SMALE S T. Sp1 activation of a TATA-less promoter requires a species-specific interaction involving transcription factor IID[J]. *Nucleic Acids Res*, 1998, 26(3): 839-846.
- [19] LIN P, PEDERSEN L C, BATRA V K, et al. Energy analysis of chemistry for correct insertion by DNA polymerase beta[J]. *PNAS*, 2006, 103(36): 13294-13299.
- [20] HING C, JAEGER J, SWEASY J B. Loop II of DNA polymerase beta is important for polymerization activity and fidelity[J]. *Nucleic Acids Research*, 2007, 35(9): 2924-2935.