

中华齿米虾胰蛋白酶基因的研究

张建业^{1,2}, 姜国良¹, 王 宁¹, 马欣荣¹

(1. 中国海洋大学 海洋生命学院, 山东 青岛 266003 2. 中国科学院 海洋研究所, 山东 青岛 266071)

摘要:以中华齿米虾(*Neocaridina denticulata sinensis*)基因组 DNA 为模板,根据胰蛋白酶基因保守序列设计简并引物,利用 PCR 技术获得一个基因。序列分析表明此基因与已报道的凡纳滨对虾(*Litopenaeus vannamei*)胰蛋白酶基因有较高的相似度,序列中包含一个内含子和两个不完整的外显子,编码 182 个氨基酸残基。该氨基酸序列与凡纳滨对虾胰蛋白酶氨基酸序列的相似度为 83.5%;含有胰蛋白酶所特有的 His、Asp 和 Ser 组成的活性三联体、决定胰蛋白酶底物专一性的 Asp、底物结合部位的 Gly 残基。综合分析认定该基因为胰蛋白酶基因片段。将其氨基酸序列与报道的其他动物胰蛋白酶氨基酸序列进行了系统进化分析。系统进化分析的结果与现有以表型特征为依据的虾分类结果是一致的。

关键词:中华齿米虾(*Neocaridina denticulata sinensis*);胰蛋白酶;基因;系统进化分析
中图分类号:Q78 **文献标识码:**A **文章编号:**1000-3096(2007)02-0040-04

近年来,国外学者针对甲壳动物基因,其中主要是经济类甲壳动物如虾、蟹及非经济类的卤虫(*Artemia*)基因开展了细致的研究工作,分离鉴定了约 30 多种功能基因。而国内在这方面的比较薄弱^[1]。胰蛋白酶是甲壳动物体内最丰富的一种蛋白水解酶,在甲壳动物的食物消化中起着重要的作用。目前凡纳滨对虾胰蛋白酶及其基因的研究已经较为详尽^[2]。但是其他甲壳动物胰蛋白酶基因研究的很少。

中华齿米虾(*Neocaridina denticulata sinensis*)属于一种经济虾,常在农村的集市上出售,也是肉食性鱼类的重要饵料。作者旨在克隆中华齿米虾胰蛋白酶基因,了解其基因结构,为将来表达胰蛋白酶基因并将之应用于生产提供资料。

1 材料与方法

1.1 材料

中华齿米虾购自青岛南山水产品市场,购回后立即清洗干净,取其尾节肌肉,称量后冻存于液氮中备用。DNA 快速纯化/回收试剂盒购自北京鼎国生物技术发展中心;DNA Marker DL2000、PMD18-T 载体和限制性内切酶购自日本 Takara 公司;Taq DNA 聚合酶、dNTP 等购自上海生工生物工程技术服务有限公司;其余药品及试剂均为进口和国产分析纯。PTC-100 型 PCR 仪为美国 Mjresearch, Inc 公司产品。

1.2 方法

1.2.1 中华齿米虾基因组 DNA 的提取

中华齿米虾基因组 DNA 的提取参照《分子克隆实验指南》^[3](略有改动)。

1.2.2 中华齿米虾胰蛋白酶基因保守片段的 PCR 扩增

根据 Genebank 中已报道的部分水生动物凡纳滨对虾(*Litopenaeus vannamei*),堪察加蟹(*Kamchatka crab*),红色帝王蟹(*Paralithodes camtschaticus*),螯虾(*Pacifastacus leniusculus*),海产七鳃鳗(*Petromyzon marinus*)和日本凤尾鱼(*Engraulis japonicus*)等胰蛋白酶氨基酸序列进行比对分析,找到两处保守域,设计出一对简并性引物:

上游引物:5'CTTCTGCGGMGCYTCCA TCTA-CA 3'

下游引物:5'RTARCCCCAGGASACRATGCC 3'

其中, M: C/A; Y: C/T; R: G/A; S: C/G。

PCR 反应体系:DNA 模板 1 μ L, dNTP (10 mmol/L) 1 μ L, 10 \times PCR buffer 5 μ L, MgCl₂ (25 mmol/L) 4 μ L, 上、下游引物(10 mmol/L) 各 1 μ L, Taq 酶(5 U/ μ L) 1 μ L, 补加 ddH₂O 到 50 μ L。石蜡油封口。PCR 反应程序:预变性 94 $^{\circ}$ C, 5 min, 变性 94 $^{\circ}$ C, 40 s; 退火 50 $^{\circ}$ C, 40 s; 延伸 72 $^{\circ}$ C, 60 s, 共循环 35 次, 充分延伸 72 $^{\circ}$ C, 10 min。

收稿日期:2005-05-22;修回日期:2005-11-10

作者简介:张建业(1971-),男,山西兴县人,博士后,讲师, E-mail: jyzhang2008@yahoo.com.cn;姜国良,通讯作者, E-mail: gjjiang@ouc.edu.cn

1.2.3 PCR 产物的回收及连接

PCR 产物的回收及连接按试剂说明书进行。

1.2.4 质粒转化、筛选及酶切检测

参照《分子克隆实验指南》^[21]。

1.2.5 克隆基因测序及序列分析

克隆基因由上海生工测序。利用 NCBI 的 Blast, Clustal x 1.8, DNA Tools 和 DNAStar 软件包进行结构、功能、同源性和系统进化分析。

2 结果与讨论

2.1 中华齿米虾胰蛋白酶基因的克隆

根据所设计的简并性 PCR 引物,直接扩增中华齿米虾基因组 DNA,所得到的 PCR 产物电泳结果如图 1 所示。扩增产物为一条清晰的亮带,分子量大小约为 700 bp。

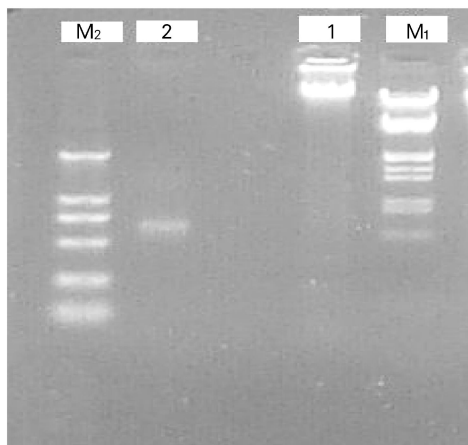


图 1 中华齿米虾胰蛋白酶基因片段 PCR 结果电泳检测

Fig. 1 The result of the trypsin gene fragments amplified from the genomic DNA of *Neocaridina denticulata sinensis*

M₁: Lambda DNA/ EcoR + Hind Marker; 1: 基因组 DNA; 2: 简并性 PCR 产物, 分子量约 700 bp; M₂: DL2000 Marker

M₁: Lambda DNA/ EcoR + Hind Marker; 1: genomic DNA; 2: degenerate PCR product; M₂: DL2000 Marker

2.2 酶切重组质粒电泳检测

对提取的重组质粒 DNA 进行 EcoR + Hind 双酶切、Hind 单酶切并与经 Hind 单酶切之 PUC-19 质粒 DNA 同时进行电泳,比较检测插入外源片段。结果显示,双酶切后的两个片段中小片段的大小约为 700 bp,恰为目的基因的长度;剩余大片段

19 相当,证明外援片段插入成功,见图 2。

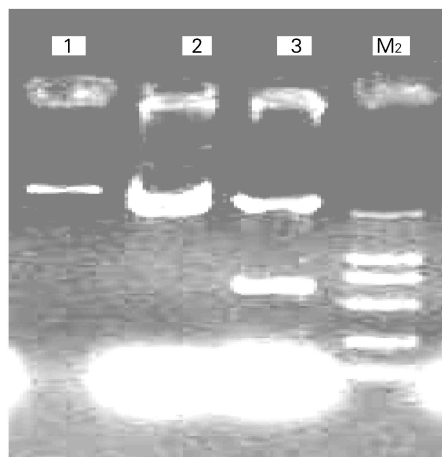


图 2 重组质粒 DNA 酶切检测外源片段结果

Fig. 2 Electrophoresis of recombinant plasmids digested with EcoR + Hind

M₂: DL2000 Marker; 1: PUC-19 质粒单酶切; 2: 重组质粒单酶切; 3: 重组质粒双酶切

M₂: DL2000 Marker; 1: PUC-19 plasmids/ Hind ; 2: recombinant plasmids/ Hind ; 3: recombinant plasmids/ EcoR + Hind

2.3 中华齿米虾胰蛋白酶基因序列分析

测序结果显示,片段大小 672 bp,去除 PCR 引物序列后长度为 629 bp。运用“GT-AG 法则”分析发现序列中含有一个内含子,位于 94 ~ 216 个核苷酸处,长度 122 bp;同时具有两个不完整的外显子区,分别为 1 ~ 94 和 217 ~ 672。利用 DNA Tools 软件进行分析得到一条长度为 182 个氨基酸残基的推断胰蛋白酶肽段,见图 3。

2.4 序列比对

分析发现克隆的基因与凡纳滨对虾胰蛋白酶核苷酸序列较高的一致度(61.2%)。推断出的氨基酸序列一致度为 83.5%,与三文鱼、鸡、牛以及人的胰蛋白酶蛋白质序列均有一定的相似度,分别为 41.8%、33.0%、35.7%和 37.9%。胰蛋白酶属于丝氨酸蛋白酶,推断的氨基酸序列具有丝氨酸蛋白酶的保守序列^[41]:(1)活性位点:含有在所有丝氨酸蛋白酶中都有的 His(17)、Asp(68)和 Ser(161)活性三联体。(2)底物结合位点:Asp(155),该残基位于底物结合沟的底部,并通过电荷作用而稳定底物裂解位点的 Lys 或 Arg 残基,因而也决定着胰蛋白酶的专一

```

1 TTCTGCGGGCGCCTCCATCTACAATGAGAAGTGGGCCATCTGTGCTGGACACTGTGTCCAG 60
1 F C G A S I Y N E N W A I C A G H C V Q 20
61 GGTGAAGACATGAATAACCCTGACTACCTCCGA GTAAGTTATCCAATTAAAGTATAATGG 120
21 G E D M N N P D Y L R 31
122 TGATATAATTGCACTTCAAATTATGTAAGACGCCTGTTTAGCTTTCTACTTTTCAGGATG 180
31 31
181 GGCCAGCTAGTAATGAAATAATATTTTTTATTCCAGGTTGTTGCTGGTGAGCACAACTTA 240
32 V V A G E H N L 39
241 GATGTTGATGAGGGTAATGAACAGGCCATTGTTCTCTCAAGGATCATTC AACACGAGAAT 300
40 D V D E G N E Q A I V L S R I I Q H E N 59
301 TACAATGGTTTCAGCATCAGCAACGACATCTCCCTTCTTCAGCTTTCTCAGCCATTGACT 360
60 Y N G F S I S N D I S L L Q L S Q P L T 79
361 TTCAACAACCTCGTTCAGCCGATTGCTCTTCTGCATCTGGTCACTTGCCACTGGTGAC 420
80 F N N F V Q P I A L P A S G H S A T G D 99
421 TGTGTTGTCTCCGGATGGGGTACAACCAGTGAAGGAGGCTCTACCCCATCTTCCCTCATG 480
100 C V V S G W G T T S E G G S T P S S L M 119
481 GCTGTGACAGTACCCGTCGTCAGTGATGATGAATGTCGCGCTGCCTATGGTCAGACTGAG 540
120 A V T V P V V S D D E C R A A Y G Q T E 139
541 GTTGAAGACTCCATGATCTGCGCTGGTCTTCTGAGGGAGGCAAGGACTCTTGCCAGGGA 600
140 V E D S M I C A G L P E G G K D S C Q G 159
601 GATTCTGGTGGCCCAATGGTCTGCTCTGATACTGGATCACCTTACCTGGCTGGCATCGTC 660
160 D S G G P M V C S D T G S P Y L A G I V 179
661 TCCTGGGGCTA 671
180 S W G *

```

图 3 中华齿米虾胰蛋白酶基因序列

Fig. 3 The sequence of trypsin of *Neocaridina denticulata sinensis*

序列大小为 672 bp,其中划线部分为设计的简并性 PCR 引物,去除引物部分后的序列全长为 629 bp。方框内序列为内含子区域,位于 94~216 个核苷酸处,长度 122 bp;外显子区域位于内含子两端,编码长度为 182 个氨基酸残基的胰蛋白酶肽段

The sequence is 672 bp nucleotide. Degenerate PCR primers are underlined. The sequence without these two primers is 629 bp. The sequence in frame is the site of intron(94~216) and its length is 122 bp; The two exons lie in both sides of the intron and they encode a peptide of Trypsin which includes 182 amino acids

性⁽⁵⁾。(3)催化位点:在催化位点附近的 Ala、Gly、Asp、Ser 等都是保守的,其中保守度最高的是 157 位~164 位胰蛋白酶催化位点所在的保守模体(motif),所有在本文中涉及到的物种胰蛋白酶在此部位均无变异。(4)含有 5 个在其他物种中都高度保守的 Cys,它们分别是 Cys₂、Cys₁₈、Cys₁₃₁、Cys₁₄₆和 Cys₁₅₇,可形成两个以上的二硫键。可进一步断定本实验所克隆得到的片段确为甲壳动物胰蛋白酶基因的一部分序列。同时,作者也发现其胰蛋白酶氨基酸序列与昆虫类的相似度则较低,以果蝇为例仅为 24.2%。推断其中原因,可能恰恰正是物种在进化过程中分化早晚的一种体现;也可能是由于它们生存环境尤其是食性差异所致,而其正确答案尚需进一步研究加以解释。

2.5 系统进化分析

利用 MegAlign 软件,将推断出的中华齿米虾胰蛋白酶氨基酸序列与 GeneBank 中已报到的 11 种代表性物种胰蛋白酶序列绘制系统进化树(图 4)。系统进化分析的结果能够反映各种甲壳动物的分类地位与进化上的关系:首先是甲壳动物与昆虫这两种节肢动物的最先分化,然后依次分化出了软骨鱼、硬骨鱼、两栖类和鸟类,最终分化出了哺乳动物。而在甲壳动物中则是挠足类动物最先分化出来,然后才是十足目与十足目的分化,最后是十足目内的分化。由此可以认为,利用胰蛋白酶做系统进化分析有可能作为经典分类方法的一项有益补充。

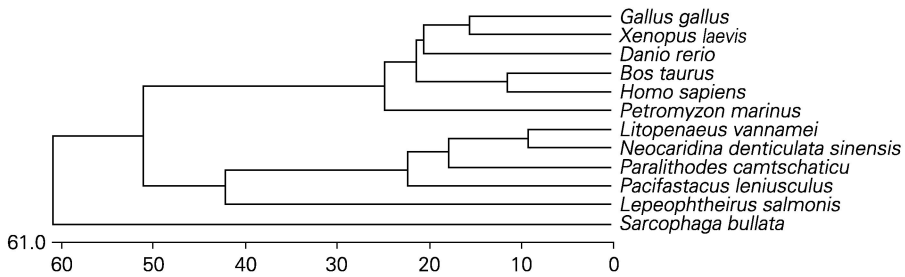


图 4 氨基酸序列系统进化树

Fig. 4 Phylogenetic analysis tree of amino acids sequence

参考文献:

- [1] 吴亚君,彭宣宪. 甲壳动物基因研究进展[J]. 海洋科学,2001,25(8):19-22.
- [2] Klein B, Sellos D, Van Wormhoudt A. Genomic organisation and polymorphism of a crustacean trypsin multi-gene family[J]. *Gene*,1998,216(1):123-129.
- [3] 萨姆布鲁克 J, 弗里奇 E F, 曼尼阿蒂斯 T. 分子克隆实验指南第二版[M]. 金冬雁,黎孟枫,译. 北京:科学出版社,2002.
- [4] Craik C S, Rocznik S, Largman C, et al. The catalytic role of the active site aspartic acid in serine proteases[J]. *Science*, 1987,237(4817):909-913.
- [5] Wang S, Magoulas C, Hickey D A. Isolation and characterization of a full-length trypsin-encoding cDNA clone from the Lepidopteran insect, *Choristoneura fumiferana* [J]. *Gene*, 1993,136(1-2):375-376.
- [6] Rudenskaya G N, Kislitsin Y A, Rebrikov D V. Col-lagenolytic serine protease PC and trypsin PC from king crab *Paralithodes camtschaticus*: cDNA cloning and primary structure of the enzymes[J]. *BMC Struct Biol*,2004,4(1):1-9.
- [7] Hernandez-Cortes P, Cerenius L, Garcia-Carreño F, et al. Trypsin from *Pacifastacus leniusculus* hepatopancreas: purification and cDNA cloning of the synthesized zymogen[J]. *Biol Chem*,1999,380(4):499-501.

Study on the trypsin gene from *Neocaridina denticulata sinensis*

ZHANG Jian-ye^{1,2}, JIANG Guo-liang¹, WANG Ning¹, MA Xin-rong¹

(1. College of Marine Life Science, Ocean University of China, Qingdao 266003, China; 2. Institute of Oceanology, the Chinese Academy of Sciences, Qingdao 266071, China)

Received: May, 22, 2005

Key words: *Neocaridina denticulata sinensis*; trypsin; gene; phylogeny

Abstract: According to the conserved domain of some trypsin genes reported in the Genbank, a pair of degenerate primers were designed and synthesized. Using genomic DNA of *Neocaridina denticulata sinensis* as a template polymerase chain reaction(PCR) was performed to clone a trypsin gene. Analysis of DNA sequence showed that the sequence was highly homologous with the trypsin gene of Pacific White Shrimp (*Litopenaeus vannamei*) and consisted of one intron and two incomplete exons. The deduced protein of the sequence has 182 amino acid residues and possesses all the key amino acids characteristics of the serine protease family. The catalytic active site is composed of the canonical triads of His, Asp and Ser and a specificity pocket occupied by Asp and the residues Gly, which define the trypsin specificity pocket. Phylogenetic analysis was processed with the amino acid sequence and protein sequences of the other 11 species' trypsin genes. The result is the same as the present taxonomy according to the characteristic surface configuration.

(本文编辑:张培新)