

# 马氏珠母贝(*Pinctada fucata*)组织蛋白酶 L 基因的克隆与表达分析\*

王忠良<sup>1,3</sup> 简纪常<sup>1,2</sup> 鲁义善<sup>1,2</sup> 丁 燊<sup>1,2</sup>  
王 蓓<sup>1,2</sup> 陈 刚<sup>1,3</sup> 吴灶和<sup>2,4</sup>

(1. 广东海洋大学水产学院 湛江 524088; 2. 广东省水产经济动物病原生物学及流行病学重点实验室 湛江 524088; 3. 广东海洋大学南海水产经济动物增殖重点实验室 湛江 524088; 4. 仲恺农业工程学院 广州 510225)

**提要** 根据已构建的溶藻弧菌(*Vibrio alginolyticus*)诱导的马氏珠母贝(*Pinctada fucata*)血淋巴 cDNA 差减文库得到的 ESTs 序列, 应用 cDNA 末端快速扩增(RACE)技术成功克隆了其组织蛋白酶 L 基因(*PFCatL*), 并对其进行了生物信息学分析; 应用实时荧光定量 PCR (Real-time PCR)技术, 研究了 *PFCatL* 基因在溶藻弧菌刺激前后马氏珠母贝足、外套膜、鳃、闭壳肌等 8 个组织中的表达变化。结果表明, *PFCatL* 基因 cDNA 全长 2004bp, 其中 5'非编码区(5'-UTR)50bp, 3'非编码区(3' -UTR)865bp, 开放阅读框(ORF)1089bp, 编码 362 个氨基酸, 其分子量计算值(MW)为 40.52kDa, 理论等电点(IP)为 5.20; 生物信息学分析表明, *PFCatL* 含有 16 个氨基酸残基组成的信号肽序列以及组织蛋白酶前体抑制功能域 I29; Clustalw2 多重比对发现 *PFCatL* 氨基酸序列在催化三联体 Cys-His-Asn、底物结合位点以及二硫键形成相关的半胱氨酸残基位点高度保守; Real-time PCR 研究发现, *PFCatL* 在马氏珠母贝各组织中均有表达, 但各组织间的表达量存在差异, 其中以肾和闭壳肌中的表达量最高; 溶藻弧菌感染 4h 后, 外套膜、鳃以及血淋巴中 *PFCatL* 基因的表达较感染前显著上调。

**关键词** 马氏珠母贝; 组织蛋白酶 L; cDNA 末端快速扩增; 实时荧光定量 PCR

**中图分类号** Q789

组织蛋白酶(Cathepsin)是一大类裂解肽键的水解酶。根据其作用方式不同可以分为外肽酶(Exopeptidase)和内肽酶(Endopeptidase), 内肽酶又根据其作用位点的催化基团不同分为: 丝氨酸蛋白酶(Serine proteinases)、半胱氨酸蛋白酶(Cysteine proteinases)、天冬氨酸蛋白酶(Aspartic proteinases)和金属蛋白酶(Metallo proteinases)。其中半胱氨酸类又包括组织蛋白酶 L、B、H、N、S、M、T 等不同的种类(Bond *et al*, 1987)。因其含有巯基又被称为巯基蛋白酶, 其活性中心含有活性必需的半胱氨酸残基, 与底物形成共价中间复合物, 具有普遍的蛋白水解活性。半胱氨酸蛋白酶主要存在于细胞质和溶酶体中

(杜欣军等, 2004)。

组织蛋白酶 L(Cathepsin L)为一种嗜酸性溶酶体蛋白水解酶, 是细胞溶酶体主要分泌蛋白, 它在溶酶体中的浓度高于 1mmol/L, 超过溶酶体总蛋白的 20% (Xing *et al*, 1998)。组织蛋白酶 L 兼有肽链内切酶和肽链外切酶的活性, 主要功能是降解蛋白, 水解某些前体蛋白酶原和激素原生成其活性形式, 或激活其它蛋白水解酶系统。在生理条件下, 通过裂解底物蛋白序列中远离氨基及羧基端的肽链降解从细胞外通过吞噬作用或从细胞内通过自噬作用进入细胞溶酶体系统的蛋白, 降解后释放的小分子肽及氨基酸弥散通过溶酶体膜并被细胞重新利用, 参与机体的多

\* 国家自然科学基金项目, 31202023 号; 国家科技支撑计划课题, 2012BAD17B00 号。王忠良, 博士, 讲师, E-mail: leong2006@126.com

通讯作者: 吴灶和, 教授, 博士生导师, E-mail: wuzaohe@126.com

收稿日期: 2012-05-31, 收修改稿日期: 2012-08-02

种生理活动, 如抗原的递呈、精子发生、排卵、角质细胞分化、胚胎形态发生、神经成熟和发育、骨基质的降解和中枢神经系统的成熟和发育等(曹新广, 2007)<sup>1)</sup>。病理状态下, 由于转录、翻译、修饰、定位、成熟、变化和与抑制剂间相互作用的异常而广泛参与多种疾病发生发展。如肿瘤的浸润与转移、关节炎、骨质疏松、阿尔茨海默病、多发性硬化症、寄生虫感染及其它慢性炎症性疾病有关(曹新广, 2007)<sup>1)</sup>。

目前, 组织蛋白酶 L 的研究主要集中在人和哺乳动物中, 并且已经对酶的结构和特性进行了全面细致的研究(McGrath *et al*, 1999; Turk *et al*, 2000)。近年来, 对来自无脊椎动物中的寄生种类组织蛋白酶 L 的研究也在增加, 这主要是由于其参与寄生虫在宿主体内引起的病理过程, 如日本血吸虫(*Schistosoma japonicum*)体内存在包括组织蛋白酶 L 在内的多种半胱氨酸蛋白酶, 这些半胱氨酸蛋白酶参与对人体血红蛋白的降解过程(雷智刚等, 2002)。本研究通过 cDNA 末端快速扩增技术(RACE)从马氏珠母贝(*Pinctada fucata*)血淋巴中成功克隆了组织蛋白酶 L 基因(*PFCatL*)的全长 cDNA 序列, 并利用实时荧光定量 PCR (Real-time PCR)技术探讨了其在溶藻弧菌(*Vibrio alginolyticus*)刺激前后马氏珠母贝不同组织中的表达变化规律, 以期了解组织蛋白酶 L 在马氏珠母贝体内的分布及其在病原感染下的表达模式, 为进一步研究软体动物组织蛋白酶 L 的功能奠定基础。

## 1 材料与方法

### 1.1 马氏珠母贝的暂养

马氏珠母贝(平均壳长 70mm)购自广东省湛江市徐闻迈陈珍珠贝养殖场, 暂养于 4 个室内玻璃钢水槽中(50cm × 60cm × 70cm), 每槽 20 只; 水温 25℃, 盐度 28, 连续充气; 暂养过程中以螺旋藻粉为饵料投喂, 每天换水(100%)一次; 室内暂养一周。

### 1.2 溶藻弧菌培养及感染

取-80℃保存的溶藻弧菌菌株, 平板划线后 28℃培养 18h, 挑取单菌落转移至 100mL TSB (2% NaCl)培养基中, 28℃振荡(120r/min)培养 18h, 5500g 离心 5min 收集菌体, 沉淀用 PBS 缓冲液悬浮, 紫外分光光度计测定 OD<sub>600</sub>, 确定感染浓度为 5.0 × 10<sup>7</sup> cfu/mL。用 1mL 无菌注射器, 在实验组马氏珠母贝闭壳肌注

射 0.1mL 上述浓度溶藻弧菌菌液, 同时对照组注射 0.1mL PBS 缓冲液。

### 1.3 马氏珠母贝血淋巴及各组织样品的采集

分别在溶藻弧菌感染 2、4、6、8、16、32h 后采集马氏珠母贝血淋巴。用 1mL 的无菌注射器在马氏珠母贝闭壳肌处采集血淋巴, 每只 0.5mL; 采集的血淋巴分装至 1.5mL 离心管中, 每管 1mL; 4℃ 800g 离心 10min 收集血细胞; 加入 1mL TRIzol 试剂, 移液器反复抽吸使血细胞充分裂解后与-80℃保存备用。另在溶藻弧菌感染 4h 后采集马氏珠母贝足、外套膜、鳃、闭壳肌、心脏、肾、胰腺腺以及血淋巴 8 个组织的样品。组织采集的量为 50—100mg, 立即置于盛液氮的研钵中充分研磨, 加入 1mL TRIzol 试剂, 转移至 1.5mL 离心管中, 移液器反复抽吸使组织细胞充分裂解后于-80℃保存备用。

### 1.4 总 RNA 提取

采用 TRIzol 试剂, 按照说明书方法进行总 RNA 提取。

### 1.5 *PFCatL* 基因 cDNA 末端快速扩增

使用 5'/3' RACE Kit, 2<sup>nd</sup> Generation 试剂盒进行 PFGal 基因 5'/3' cDNA 末端的克隆, 具体操作按试剂盒说明书方法进行(略作改动)。通过对已构建差减文库中获得的 EST 序列进行 BLASTx 同源性分析发现一条 249bp 的 EST 序列与其它多种无脊椎动物组织蛋白酶 L 具有较高的同源性(*e* 值从 3e-90 到 7e-104)。根据此 EST 序列, 应用 Primer Premier 5.0 软件设计如下 4 条基因特异性引物用于 5'/3' RACE-PCR 扩展(表 1)。

**1.5.1 3' cDNA 末端的 PCR 扩增** 根据已构建差减文库中获得的 EST 序列设计引物 cathnests2, 以及 RACE-PCR 扩增 3'PCR 锚定引物, 以 cDNA 一链为模板进行 PCR 扩增, 反应条件为 94℃预变性 2min; 94℃变性 30s, 58℃退火 45s, 72℃延伸 2min, 运行 30 个循

表 1 组织蛋白酶 L 基因(*PFCatL*)RACE-PCR 扩增所用引物

引物名称	引物序列
cathnests2	5'-GGACAGCACTTCAGACAGACAGG-3'
cathnesta2	5'-TCCTGTCTGTCTGAAGTGTCTGCC-3'
cathnesta3	5'-GCTGTCTGCTCACTCAGAGAAATCA-3'

1) 曹新广, 2007. Cathpsin L 和 Cathpsin B 的表达与大肠癌生物学行为的关系. 郑州: 郑州大学硕士学位论文, 24—25

环, 72°C 延伸 7min。将所得的目的片段回收纯化并连接到 pMD-18T 载体上, 转化到 *E. coli* DH5 $\alpha$  感受态细胞中, 筛选单克隆, 经菌落 PCR 检测为阳性后测序验证。

**1.5.2 5' cDNA 末端的 PCR 扩增** 根据已构建差减文库中获得的 EST 序列设计首轮引物 cathnesta3 及次轮引物 cathnesta2, 经两轮 PCR 反应扩增出 5' 端。首轮反应条件为 94°C 预变性 2min; 94°C 变性 15s, 55°C 退火 30s, 72°C 延伸 40s, 运行 10 个循环; 94°C 变性 15s, 55°C 退火 30s, 72°C 延伸 40s, 运行 25 个循环, 每一循环增加 20s; 72°C 延伸 7min。将首轮 PCR 产物稀释 10 倍作为第二轮 PCR 反应的模板, 反应条件为 94°C 预变性 2min, 94°C 30s, 58°C 45s, 72°C 1min, 运行 27 个循环, 72°C 7min。将所得的目的片段回收纯化并连接到 pMD-18T 载体上, 转化到 *E. coli* DH5 感受态细胞中, 挑选单克隆, 经菌落 PCR 检测为阳性后测序验证。

## 1.6 生物信息学分析

应用 BLAST 程序 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) 进行序列同源性比对和相似性分析; 以 DNAMAN 5.2.2 软件进行序列拼接; 开放阅读框 (ORF) 的确定以及氨基酸序列的翻译采用 DNASTar 和 DNATools 软件进行; 模体序列使用 InterPro 软件查找; 应用 Clustalw2 程序 (<http://www.ebi.ac.uk/Tools/clustalw2/>) 进行序列的多重比对; 使用 MEGA4 软件, 以邻位相连法 (neighbor-joining) 构建系统进化树; 应用 SignalP 3.0 程序 (<http://www.cbs.dtu.dk/services/SignalP/>) 进行信号肽序列查找; 应用 Prosite 程序分析氨基酸序列模体结构。

## 1.7 *PFCatL* 基因表达的实时荧光定量 PCR 分析

*PFCatL* 基因表达的实时荧光定量分析按 SYBR<sup>®</sup> Premix *Ex Taq*<sup>™</sup> (Perfect Real Time) 试剂盒进行。

**1.7.1 引物设计** 根据以上实验中获得的 *PFCatL* 基因全长 cDNA 序列, 设计 2 条基因特异性引物, 同

时以马氏珠母贝  *$\beta$ -actin* 和 *GAPDH* 基因作为双内参进行基因特异性异物设计, 见表 2。

**1.7.2 总 RNA 的 DNase 消化** 为减少总 RNA 中基因组 DNA 残留对实时荧光定量 PCR 结果的影响, 在合成 cDNA 一链前用 DNase I 对总 RNA 进行消化处理。

**1.7.3 扩增效率的确定** 取上述 DNase I 消化的总 RNA 合成的 cDNA 一链, 测定 OD 值, 计算浓度, 并按 10 倍系列稀释为模板标准品, 进行荧光定量 PCR 扩增。应用 7300 System SDS Software 软件 (Version 1.4, ABI) 自动进行数据分析, 并按实际扩增效率进行目标基因的荧光定量分析。

**1.7.4 *PFCatL* 在血淋巴中的时序表达及溶藻弧菌感染前后各组织中的表达差异** 将上述 DNase I 消化的总 RNA 合成的 cDNA 一链, 用无菌水稀释后作为荧光定量 PCR 的模板; 采用 SYBR Select Master Mix (Applied Biosystems), 根据操作说明在 ABI 7300 定量 PCR 仪上进行。反应总体积为 25 $\mu$ L, 内含 cDNA 模板 1 $\mu$ L, 引物各 0.5 $\mu$ L (引物浓度为 10 $\mu$ mol/L), 反应条件为: 95°C 10s, 95°C 5s, 60°C 31s, 共 40 个循环, 在最后一个循环结束后做溶解曲线, 确定 PCR 反应质量。

**1.7.5 数据统计与分析** 应用 7300 System SDS Software 软件, 采用 ddCt 法自动进行目标基因的荧光定量分析; 应用 SPSS 11.0 for Windows 软件对获得的实验数据进行 ANOVA 单因素方差分析和 Tukey HSD 多重比较, 检验不同处理间目标基因的表达差异, 以  $P < 0.05$  作为显著性差异。

## 2 结果与分析

### 2.1 各组织总 RNA 的提取

分别在溶藻弧菌感染 2、4、6、8、16 以及 32h 后采集马氏珠母贝血淋巴, TRIzol 法提取血淋巴总 RNA, 并进行 1.0% 琼脂糖凝胶电泳分析, 如图 1 所示,

表 2 实时荧光定量 PCR 分析中使用的引物  
Tab.2 Primers for the fluorescent quantitative real time-PCR

引物名称	扩增基因	引物序列	产物长度(bp)
DLS	<i><math>\beta</math>-actin</i>	5'-TGGTATGGGACAGAAGGAC-3'	94
DLA		5'-GACAATGCCGTGCTCAAT-3'	
PFGS	<i>GAPDH</i>	5'-TTTTGGCATTGAGGAAGGTTTG-3'	127
PFGA		5'-CAGTGGAGGATGGTATGATGTTAG-3'	
caths1	<i>PFCatL</i>	5'-GTATCGGTGGACTGGAGGGTGA-3'	54
catha1		5'-TTCCTTGCTTGGCTGTGTATGG-3'	

未见弥散条带, RNA 质量较好。

在溶藻弧菌感染 4h 后, 采集马氏珠母贝的足、外套膜、鳃、闭壳肌、心脏、肾脏、肝胰腺、血淋巴 8 个组织样品, 同时, 按同样方法采集对照组 8 个组织的样品。液氮中充分研磨后, 分别提取各组织的总 RNA。图 2a, b 分别为对照组和溶藻弧菌感染后各组织总 RNA 1.0% 琼脂糖凝胶电泳结果, 如图 2 所示, 条带清晰, 无弥散现象, 可用于后续实验。

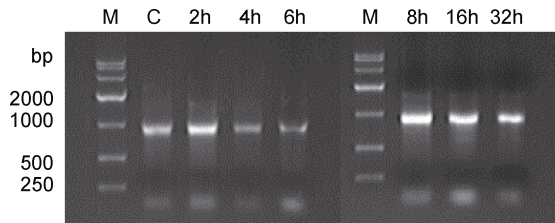


图 1 对照及溶藻弧菌感染后不同时间马氏珠母贝血淋巴总 RNA 电泳

Fig.1 Agarose electropherogram of total RNA purified from haemocytes of healthy and *V. alginolyticus* challenged *P. fucata* at different time points

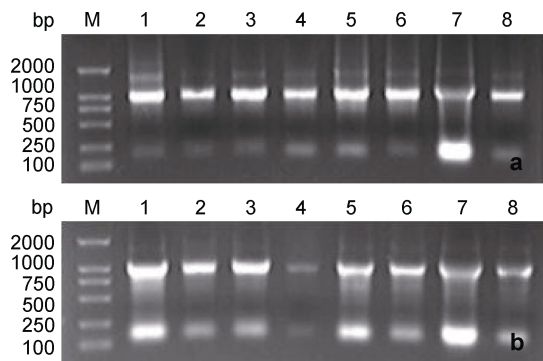


图 2 溶藻弧菌感染前(a)及感染 4h 后(b)马氏珠母贝各组织总 RNA 电泳

Fig.2 Agarose electropherogram of total RNA from tissues of healthy (a) and *V. alginolyticus* challenged (b) *P. fucata*  
M. Marker DL2000; 1. 足; 2. 外套膜; 3. 鳃; 4. 闭壳肌; 5. 心脏; 6. 肾; 7. 肝胰腺; 8. 血淋巴

## 2.2 *PFCatL* 基因全长 cDNA 序列的克隆

根据文库中获得的 EST 序列设计 4 条基因特异性引物分别进行 5'/3' RACE。通过 PCR 扩增分别获得 1423bp 的 *PFCatL* 基因 3'端 cDNA 序列和 605bp 的 *PFCatL* 基因 5'端 cDNA 序列(图 3)。应用 DNAMAN5.2.2 软件对 3'和 5'端 cDNA 序列进行拼接, 得到 2004bp 的 *PFCatL* 基因全长 cDNA 序列(GenBank 登录号: FJ501258)。

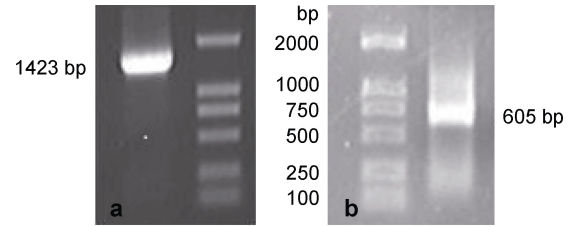


图 3 *PFCatL* 基因 3'RACE(a)及 5'RACE(b)产物琼脂糖凝胶电泳

Fig.3 Agarose electropherogram of 3' RACE (a) and 5' RACE (b) products of *PFCatL*

## 2.3 *PFCatL* 基因序列特征分析

*PFCatL* 基因 cDNA 序列全长 2004bp, 其中 5' UTR 50bp, 3' UTR 865bp, 含有一个加尾信号序列以及 poly(A)尾; *PFCatL* 基因 ORF 1089bp, 编码一个由 362 个氨基酸残基组成的蛋白质, 其分子量计算值为 40.52kDa, 理论等电点为 5.20; *PFCatL* 氨基酸序列 N 末端有一个由 16 个氨基酸残基组成的信号肽序列, 并含有一个组织蛋白酶前体抑制功能域 I29; 与其它组织蛋白酶 L 一样, *PFCatL* 氨基酸序列的催化三联体 Cys-His-Asn, 6 个形成二硫键的半胱氨酸残基以及 5 个底物结合相关的氨基酸残基高度保守, 但是在 *PFCatL* 成熟肽第二位为天冬氨酸残基(Asn), 而不是许多木瓜蛋白酶家族蛋白中的脯氨酸残基(Pro)(图 4)。

## 2.4 *PFCatL* 基因生物信息学分析

BLASTx 同源性比对分析发现, *PFCatL* 氨基酸序列与黄粉虫(*Tenebrio molitor*)等多种无脊椎动物组织蛋白酶 L 的同源性为 49%—58%, *e* 值为  $3e-90$ — $7e-104$ , 表明 *PFCatL* 为无脊椎动物组织蛋白酶 L 蛋白家族成员之一。

应用 Clustalw2 程序对 *PFCatL* 以及多种已知组织蛋白酶 L 的氨基酸序列进行多重比对(图 4)。结果显示, 所有组织蛋白酶 L 氨基酸序列间在催化三联体、底物结合位点以及二硫键形成相关的半胱氨酸残基位点高度保守, 总体上, *PFCatL* 的氨基酸序列成熟肽区序列保守性要高于前体肽的氨基酸序列。然而, 位于多种木瓜蛋白酶家族成员成熟蛋白酶第二位的脯氨酸残基(Pro)在 *PFCatL* 氨基酸序列中被天冬氨酸残基(Asp)取代。

根据多重比对的结果, 利用 MEGA4 软件以 NJ 法(neighbor-joining)构建多种脊椎动物和无脊椎动物组织蛋白酶 L 氨基酸序列的系统进化树。如图 5 所示, 系统进化树主要由两个主要的进化分支组成, 分别为脊椎动物和无脊椎动物组织蛋白酶 L; 在脊椎动物

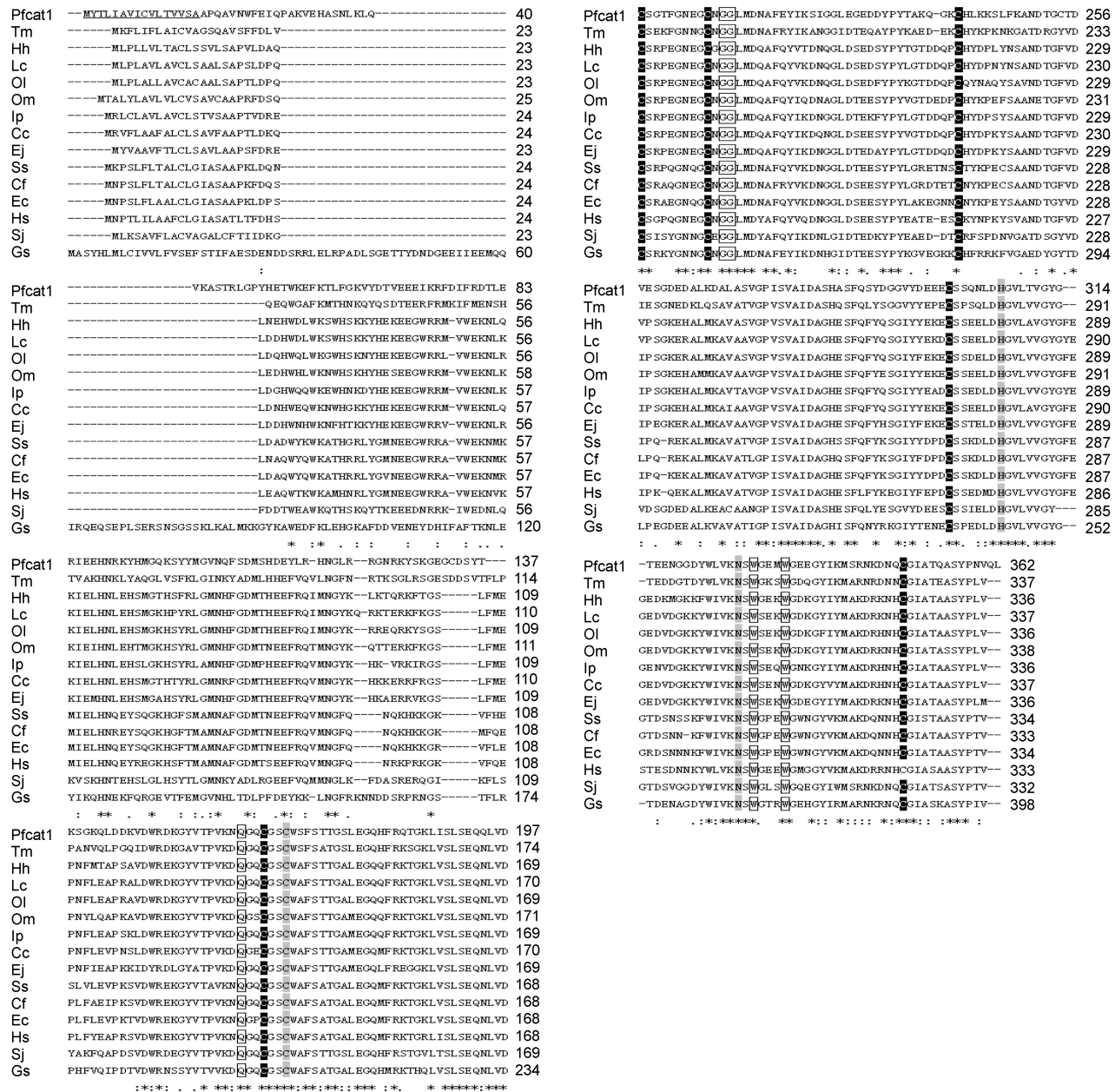


图4 PFCatL 基因氨基酸序列与其它已知组织蛋白酶 L 氨基酸序列多重比对结果

Fig.4 Multiple alignments of the deduced amino acid sequences of PFCatL with cathepsin L of other species

下划线表示信号肽序列; 灰色阴影表示催化三联体 Cys-His-Asn; 黑色阴影表示二硫键位点; 方框表示底物结合位点。多重比对的氨基酸序列来源及 GenBank 登录号分别为: Tm. 黄粉虫, AAP94046; Hh. 大西洋庸鲷, ABJ99858; Lc. 尖吻鲈, ABV59078; Ol. 青鳉, NP\_001098156; Om. 虹鳟, NP\_001117777; Ip. 斑点叉尾鲷, EU915299; Cc. 鲤, BAD08618; Ej. 鳊鱼, BAC16538; Ss. 野猪, NP\_999057; Cf. 家犬, NP\_001003115; Ec. 马, XP\_001494409; Hs. 人, AAH12612; Sj. 海参, ABW98676; Gs. 棘颚口线虫, ABY28387

分支中, 高等脊椎动物和较低等脊椎动物的组织蛋白酶 L 形成姐妹分支; 而在无脊椎动物分支中, PFCatL 与棘颚口线虫(*Gnathostoma spinigerum*)(Gs)聚在一起, 并与海参(*Stichopus japonicus*)(Sj)、黄粉虫(*Tenebrio molitor*)(Tm)形成姐妹分支。

2.5 PFCatL 基因的表达分析

应用 Real time-PCR 技术探讨了溶藻弧菌感染 2、

4、6、8、16、32h 后马氏珠母贝血淋巴中 PFCatL 基因的时序表达以及溶藻弧菌感染前后马氏珠母贝足、外套膜、鳃、闭壳肌、心脏、肾、肝胰腺和血淋巴 8 个组织中 PFCatL 基因的表达差异。

在马氏珠母贝血淋巴中, 溶藻弧菌感染 2h 后, PFCatL 基因的表达开始上调, 到 4h 时达到最大, 显著高于其它各时间点 PFCatL 基因的表达水平; 随着

时间的延长, 感染 6h 后 *PFCatL* 的表达量开始逐渐下降, 并在 16h 后 *PFCatL* 基因的表达恢复到对照组水平(图 6)。

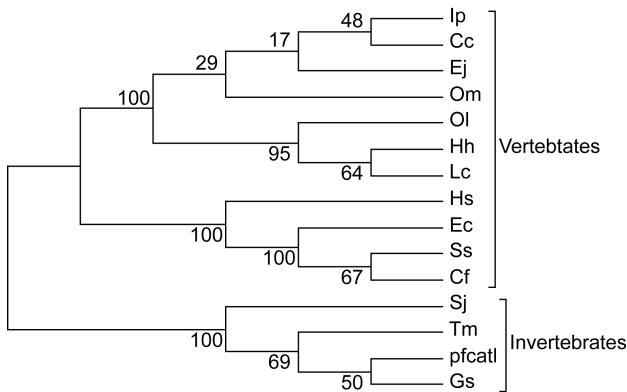


图 5 以 NJ 法构建的组织蛋白酶 L 家族成员氨基酸序列系统进化树

Fig.5 A phylogenetic tree of cathepsin L family members constructed with the neighbor-joining method 图中物种名称及 GenBank 序列登录号同图 4

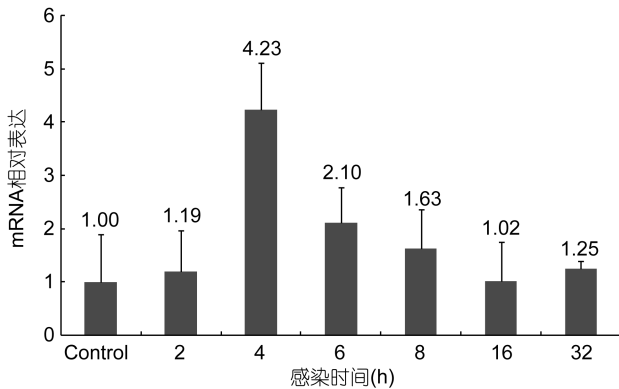


图 6 溶藻弧菌感染不同时间后马氏珠母贝血淋巴中 *PFCatL* 基因的表达变化

Fig.6 The *PFCatL* mRNA expression level in haemocytes of *P. fucata* at different time points after *V. alginolyticus* challenge

溶藻弧菌感染前, *PFCatL* 基因在马氏珠母贝的足、外套膜、鳃、闭壳肌、心脏、肾脏、肝胰腺、血淋巴 8 个组织中均有表达, 但各组织间的表达量存在差异, 其中以肾和闭壳肌中的表达量最高, 而其它组织中 *PFCatL* 基因的表达量相对较低(图 7); 溶藻弧菌感染 4h 后, 外套膜、鳃以及血淋巴中 *PFCatL* 基因的表达较感染前显著上调, 而闭壳肌、心脏以及肾中 *PFCatL* 基因的表达量较感染前稍有下降, 但二者间无显著差异(图 7)。

### 3 讨论

组织蛋白酶 L 基因已从多种脊椎动物和无脊椎

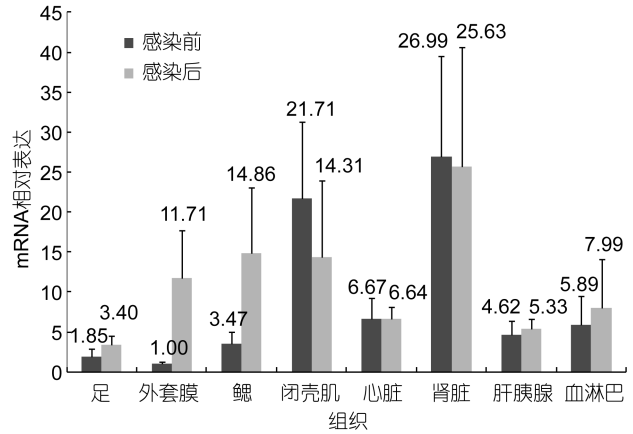


图 7 溶藻弧菌感染前后各组织中 *PFCatL* 基因的表达差异  
Fig.7 The *PFCatL* mRNA expression level in various tissues of healthy and *V. alginolyticus* challenged *P. fucata*

动物中克隆出来, 并已证实其在生理、病理过程中的多种功能(Boulay *et al*, 1996; Tsunemoto *et al*, 2004; Tingaud-Sequeira *et al*, 2007; Yeh *et al*, 2009)。本研究应用 EST 及 RACE 技术克隆了马氏珠母贝组织蛋白酶 L 基因的全长 cDNA 序列, 并分析了其序列结构特征及病原感染前后的组织表达差异。通过 BLAST 同源性比对、相似性分析及序列特征和系统进化分析的研究表明, *PFCatL* 基因编码的蛋白质为无脊椎动物组织蛋白酶 L 家族的成员。

与其它组织蛋白酶 L 相似, *PFCatL* 氨基酸序列由三部分结构域组成, N-末端由 16 个氨基酸残基组成的信号肽序列, 接着是 126 个氨基酸残基组成的前体肽, 最后是由 220 个氨基酸残基组成的成熟蛋白酶。值得注意的是, *PFCatL* 成熟蛋白酶的第二个氨基酸残基为天冬氨酸(Asp), 而不是其它木瓜蛋白酶家族成员中观察到的可防止 N-端蛋白不必要水解的脯氨酸残基(Pro)(Rawlings *et al*, 1994)。由于目前对于软体动物组织蛋白酶 L 的研究较少, 尚不清楚位于 *PFCatL* 成熟蛋白酶第二位的天冬氨酸残基是否具有防止 N-端蛋白不必要水解的功能还有待进一步论证, 以期能阐明软体动物组织蛋白酶 L 中特殊氨基酸残基的变异对其基因功能的影响。

在 *PFCatL* 氨基酸序列的氨基端具有组织蛋白酶前体抑制功能域 I29, 这类功能域作为蛋白酶前体发挥作用。I29 结构域也存在与其它木瓜蛋白酶前体中(Groves *et al*, 1996), 例如番木瓜(*Carica papaya*)的半胱氨酸蛋白酶 Caricain, 其前体形成 $\alpha$ -螺旋反向穿过底物结合位点, 与成熟酶的氨基端相连, 成熟酶因为无法与底物形成合适的构象而无法被激活, 通过水

解去除前体, 蛋白酶才能被激活。另一类抑制前体以多个拷贝存在, 具有更强的蛋白酶抑制活性(Olonen *et al.*, 2003)。

应用 Clustalw2 程序对 *PFCatL* 基因以及多种已知的脊椎动物和无脊椎动物组织蛋白酶 L 基因的氨基酸序列进行多重比对分析表明, *PFCatL* 的催化活性三联体 Cys-His-Asn、构建二硫键的半胱氨酸以及底物结合相关氨基酸残基与其它动物组织蛋白酶 L 高度保守。此外, 根据多重比对的结果, 应用 MEGA4 软件以 NJ 法构建了来自不同物种组织蛋白酶 L 的系统进化树, *PFCatL* 与无脊椎动物棘颚口线虫 (*Gnathostoma spinigerum*) 聚在一起, 并与海参 (*Stichopus japonicus*)、黄粉虫 (*Tenebrio molitor*) 形成姐妹分支, 进一步说明 *PFCatL* 属于无脊椎动物组织蛋白酶 L 家族成员。

组织蛋白酶在产生生物活性分子和清除生物活性分子方面起重要作用, 包括降解各种蛋白为机体提供营养, 清除细胞中残缺、不正常的多肽, 控制着细胞各种酶和活性肽的潜在浓度, 水解某些前体蛋白(酶原和激素原)生成其活性形式, 或激活其它蛋白水解酶系统而使其发挥作用等, 从而参与机体的多种生理和病理活动。Onishi 等(2004)研究发现, 当哺乳动物感染硕大利什曼原虫 (*Leishmania major*) 后, 组织蛋白酶 L 在 Th1 型免疫反应中起关键作用; 而抑制体内组织蛋白酶 L 活性或敲除组织蛋白酶 L 基因则会引起哺乳动物机能异常甚至引起死亡(Collette *et al.*, 2004); 在无脊椎动物中, 组织蛋白酶作为溶酶体蛋白水解系统的主要组分, 负责胞内蛋白特异性降解(Knop *et al.*, 1993)。本研究中, 溶藻弧菌感染后, 马氏珠母贝血淋巴中 *PFCatL* 基因的表达显著上调, 这与白斑综合症病毒(WSSV)感染后的凡纳滨对虾 (*Litopenaeus vannamei*) 肝胰腺中组织蛋白酶 L 基因上调的研究结果一致(Zhao *et al.*, 2007), 说明组织蛋白酶 L 参与了机体对外界病原入侵的免疫防御。

溶藻弧菌感染前, 马氏珠母贝 8 个组织中均检测到 *PFCatL* 基因的表达, 这与 Brix 等(2008)以及 Yeh 等(2009)对组织蛋白酶 L 基因在动物组织中特异表达的研究结果一致。溶藻弧菌感染 4h 后, 马氏珠母贝外套膜、鳃以及血淋巴中 *PFCatL* 基因的表达量显著上调, 而闭壳肌、心脏、肾、肝胰腺中 *PFCatL* 基因的表达较感染前变化不明显, 说明 *PFCatL* 可能通过外套膜、鳃以及血淋巴发挥对病原菌入侵的防御作用, 而且 *PFCatL* 基因在马氏珠母贝体内的表达可能存在

组成型和诱导型两种调控机制。

## 4 结语

本论文的研究结果为探讨马氏珠母贝免疫防御相关基因的作用机理和调控机制奠定了一定的基础; 丰富和发展了贝类分子免疫学的研究内容; 对于全面了解和认识贝类免疫防御的分子机制, 进而指导马氏珠母贝的遗传改良和抗病品系的培育都具有理论参考意义。

## 参 考 文 献

- 杜欣军, 邵红莲, 邵丁丁等, 2004. 棉铃虫组织蛋白酶 B 酶原在大肠杆菌中的表达及纯化. 农业生物技术学报, 12(2): 162—166
- 雷智刚, 孟锦绣, 何 蔼等, 2002. 日本血吸虫组织蛋白酶 L1 基因的编码区全序列分析及克隆. 中国寄生虫学与寄生虫病杂志, 20: 325—327
- Bond J S, Burler P E, 1987. Intracellular proteases. Annual Review of Biochemistry, 56: 331—361
- Boulay C L, Wormhoudt A V, Sellos D, 1996. Cloning and expression of cathepsin L-like proteinases in the hepatopancreas of the shrimp *Penaeus vannamei* during the intermolt cycle. Journal of Comparative Physiology B, 166: 310—318
- Brix K, Dunkhorst A, Mayer K *et al.*, 2008. Cysteine cathepsins: cellular roadmap to different functions. Biochimie, 90: 194—207
- Collette J, Bocock J P, Ahn K *et al.*, 2004. Biosynthesis and alternate targeting of the lysosomal cysteine protease cathepsin L. In: Jeon K W ed. International Review of Cytology, 251: 1—51
- Groves M R, Taylor M A J, Scott M *et al.*, 1996. The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. Structure, 4(10): 1193—1203
- Knop M, Schiffer H H, Rupp S *et al.*, 1993. Vacuolar/lysosomal proteolysis: proteases, substrates mechanisms. Current Opinion in Cell Biology, 5: 990—996
- McGrath M E, 1999. The lysosomal cysteine proteases. Annual Review of Biophysics and Biomolecular Structure, 28: 181—204
- Olonen A, Kalkkinen N, Paulin L, 2003. A new type of cysteine proteinase inhibitor—the salarin gene from Atlantic salmon (*Salmo salar* L) and Arctic charr (*Salvelinus alpinus*). Biochimie, 85(7): 677—681
- Onishi K, Li Y, Ishii K *et al.*, 2004. Cathepsin L is crucial for a Th1-type immune response during *Leishmania major* infection. Microbes and Infection, 6: 468—474
- Rawlings N D, Barrett A J, 1994. Families of cysteine peptidases. Methods in Enzymology, 24: 461—486
- Tingaud-Sequeira A, Cerdà J, 2007. Phylogenetic relationships

- and gene expression pattern of three different cathepsin L (Ctsl) isoforms in zebrafish: Ctsla is the putative yolk processing enzyme. *Gene*, 386: 98—106
- Tsunemoto K, Osatomi K, Nozaki Y *et al*, 2004. Molecular characterization of cathepsin L from hepatopancreas of the carp *Cyprinus carpio*. *Comparative Biochemistry and Physiology Part B*, 137: 107—114
- Turk B, Turk D, Turk V, 2000. Lysosomal cysteine proteases: more than scavengers. *Biochimica et Biophysica Acta*, 1477: 98—111
- Xing R, Addington A K, Mason R W, 1998. Quantification of cathepsins B and L in cells. *Biochemical Journal*, 322(Pt 2): 499—505
- Yeh H, Klesius P H, 2009. Channel catfish, *Ictalurus punctatus*, cysteine proteinases: Cloning, characterisation and expression of cathepsin H and L. *Fish and Shellfish Immunology*, 26: 332—338
- Zhao Z, Yin Z, Weng S P *et al*, 2007. Profiling of differentially expressed genes in hepatopancreas of white spot syndrome virus-resistant shrimp (*Litopenaeus vannamei*) by suppression subtractive hybridization. *Fish and Shellfish Immunology*, 52: 520—534

## CLONING AND EXPRESSION ANALYSIS OF THE CATHEPSIN L GENE FROM PEARL OYSTER *PINCTADA FUCATA*

WANG Zhong-Liang<sup>1,3</sup>, JIAN Ji-Chang<sup>1,2</sup>, LU Yi-Shan<sup>1,2</sup>, DING Yu<sup>1,2</sup>,  
WANG Bei<sup>1,2</sup>, CHEN Gang<sup>1,3</sup>, WU Zao-He<sup>2,4</sup>

(1. Fisheries College, Guangdong Ocean University, Zhanjiang, 524088; 2. Guangdong Provincial Key Laboratory of Pathogenic Biology and Epidemiology for Aquatic Economic Animals, Zhanjiang, 524088; 3. Key Laboratory of Aquaculture in South China Sea for Aquatic Economic Animal, Regular High Education Institute of Guangdong Province, Zhanjiang, 524088; 4. Zhongkai University of Agriculture and Engineering, Guangzhou, 510225)

**Abstract** To examine the expression patterns of the *PFCatL* gene in various tissues, including foot, mantle, gill, adductor, heart, kidney, hepatopancreas, and haemolymph of healthy and *Vibrio alginolyticus* challenged oyster *Pinctada fucata*, we cloned the cDNA of cathepsin L (*PFCatL*) in the oyster by rapid amplification cDNA ends (RACE) technique based on the ESTs that analyzed from SSH cDNA library, and employed fluorescent real-time quantitative PCR (RT-PCR). The results show that the full length of *PFCatL* cDNA was 2004bp, including a 5' 50bp UTR, a 3' 865bp UTR, and a 1089bp ORF encoding a polypeptide of 362 amino acids with an MW at 40.52 kDa, and a 5.20 IP. A signal peptide of 16 amino acids and a peptidase inhibitor domain I29 were detected in *PFCatL* by SignalP and Prosite analysis. The catalytic triad, substrate binding sites, and cysteine disulfide linkage sites were highly conserved in *PFCatL*, indicated by multiple alignment analysis. *PFCatL* transcripts were expressed ubiquitously in all tested tissues, while the expression levels were significantly different from each other; and the highest expression was detected in kidney and adductor. Four hours after injection with *V. alginolyticus*, the expression of *PFCatL* transcripts in mantle, gill, and haemolymph of *P. fucata* were significantly up-regulated relative to that of the control.

**Key words** *Pinctada fucata*; cathepsin L; rapid amplification of cDNA ends (RACE); fluorescent quantitative real-time PCR