

## 农业生物技术

# 家蚕组织蛋白酶基因家族的鉴定及表达特征分析

李懿, 周小英, 黎治浪, 李建伟, 陈世达, 郭超, 侯勇, 赵萍

家蚕基因组生物学国家重点实验室 西南大学, 重庆 400716

李懿, 周小英, 黎治浪, 等. 家蚕组织蛋白酶基因家族的鉴定及表达特征分析. 生物工程学报, 2015, 31(12): 1728–1740.  
Li Y, Zhou XY, Li ZL, et al. Identification and expression pattern of cathepsin family in silkworm (*Bombyx mori*). Chin J Biotech, 2015, 31(12): 1728–1740.

**摘要:** 家蚕是鳞翅目完全变态昆虫, 在其变态过程中伴随着巨大的形态变化, 包括旧组织的解离和新组织的形成, 在这过程中有多种组织蛋白酶参与。组织蛋白酶是一类细胞内蛋白酶, 广泛存在于各个物种中, 包括组织蛋白酶 B、H、L 等几个亚家族。对家蚕组织蛋白酶的研究将有利于阐明家蚕变态发育的详细过程。通过对家蚕基因组数据库进行筛选, 共在家蚕中鉴定到 13 种组织蛋白酶, 并对这 13 种组织蛋白酶的基本信息和表达模式进行了分析。另外, 利用家蚕基因芯片数据和荧光定量 PCR 分析, 鉴定编号为 BGIBMGA004622 的基因为卵巢特异表达的组织蛋白酶 L 亚家族基因。该基因全长 1 209 bp, 编码 402 个氨基酸。经过序列分析, 该酶与其他物种的组织蛋白酶 L 具有较高的同源性, 其活性位点高度保守, 且与鳞翅目的组织蛋白酶 L 在进化上聚为一支。同时, 对该基因进行克隆并原核表达, 结果显示重组蛋白以包涵体的形式表达。定量 PCR 结果显示, 该酶在蛹发育初期表达量逐渐升高, 至蛹 3 d 达到最高值, 推测其可能参与卵巢与卵母细胞的发育过程。

**关键词:** 家蚕, 组织蛋白酶, 变态发育, 表达特征分析, 卵巢

**Received:** February 2, 2015; **Accepted:** May 11, 2015

**Supported by:** National Basic Research Program of China (973 Program) (No. 2012CB114600), National High Technology Research and Development Program of China (863 Program) (No. 2011AA100306), National Natural Science Foundation (Nos. 31201853, 31172157), Basic and Advanced Research Project of Chongqing (No. cstc2015jcyA0469).

**Corresponding author:** Yong Hou. Tel: +86-23-6825-0099; E-mail: yhou@swu.edu.cn

国家重点基础研究发展计划 (973 计划) (No. 2012CB114600), 国家高技术研究发展计划 (863 计划) (No. 2011AA100306), 国家自然科学基金 (Nos. 31201853, 31172157), 重庆市基础与前沿研究计划 (No. cstc2015jcyA0469) 资助。

网络出版时间: 2015-06-01

网络出版地址: <http://www.cnki.net/kcms/detail/11.1998.Q.20150601.1604.001.html>

# Identification and expression pattern of cathepsin family in silkworm (*Bombyx mori*)

Yi Li, Xiaoying Zhou, Zhilang Li, Jianwei Li, Shida Chen, Chao Guo, Yong Hou, and Ping Zhao

Southwest University, State Key Laboratory of Silkworm Genome Biology, Chongqing 400716, China

**Abstract:** Silkworm is a holometabolous insect of *Lepidoptera*. During metamorphosis, significant morphological changes happen including the dissociation of old tissues and remodeling of new tissues. It has been reported that cathepsins are involved in these processes. Cathepsin is a kind of intracellular proteinase that exists in many species. It includes some subfamilies like cathepsin B, H and L. The studies on cathepsin are useful for clarifying the details of silkworm metamorphosis process. In total, 13 cathepsins were identified by screening the silkworm genome database. The basic information and the expression patterns about these genes were analyzed. Interestingly, an ovary-specific cathepsin L gene (Gene ID: BGIBMGA004622) was investigated by the data of silkworm microarray and real-time quantitative PCR (qPCR). The full-length cDNA is 1 209 bp, encoding a protein with 402 amino acids. Sequences alignment revealed that it has a high sequence similarity with cathepsin L of other species, and it is highly conserved in the active-site of the enzyme. The phylogenetic analysis showed that ovary-specific cathepsin L is clustered with other lepidopterous insects. Furthermore, this gene was cloned and prokaryotic expressed. Recombinant protein was present in inclusion body. Importantly, the qPCR result showed that the expression level of this gene is increasing during the early stage of pupal development and reaches the highest value at the 3<sup>rd</sup> day of pupal stage, which suggests that this gene may be involved in the process of development of the ovary and oocyte.

**Keywords:** silkworm, cathepsin, metamorphosis, expression pattern, ovary

生物体在其发育过程中,组织和器官会发生明显的阶段性变化。这一过程中,许多旧组织被程序性地水解掉,一些未分化的细胞发生分裂和生长,形成新的组织。这些变化伴随着丰富的组织蛋白酶的活动<sup>[1-2]</sup>。

组织蛋白酶是一类主要存在于溶酶体中的胞内蛋白酶,在弱酸性环境中易被活化<sup>[3]</sup>。自提出组织蛋白酶的概念以来,人们首先于20世纪40年代发现了组织蛋白酶C,在其后约40年的时间中只有组织蛋白酶B、H、L陆续被鉴定并测序。直至20世纪90年代组织蛋白酶B的晶体结构被成功解析后,有关组织蛋白酶的研究才有所进展,越来越多的组织蛋白酶被发现。

到目前为止从组织蛋白酶A到组织蛋白酶Z都已有报道<sup>[4-7]</sup>。

昆虫组织蛋白酶的研究始于20世纪80年代。1988年,Medina等在果蝇胚胎中发现了一种组织蛋白酶B,其与卵黄蛋白的降解和胚胎发生有着重要的关系。在成熟卵母细胞中,该酶以其非活性前体形式存在,并且会被另外一种同时存在于卵母细胞中的丝氨酸蛋白酶所活化。伴随着胚胎发育的过程,该酶的活性也在逐渐升高<sup>[8]</sup>。在棉铃虫中,徐夏莲、杨晓梅等对组织蛋白酶B的性质及纯化进行了报道,他们从棉铃虫卵中纯化出的这种组织蛋白酶B可能在胚胎发育中水解卵黄蛋白供给胚胎营养<sup>[9-10]</sup>。

2013年, Zhang 等对棉铃虫组织蛋白酶 L 的研究中发现, 该酶在棉铃虫化蛹时脂肪体解离中起到重要作用<sup>[11]</sup>。

家蚕作为鳞翅目昆虫的代表, 其组织蛋白酶参与变态发育过程的研究也一直广受关注。1990年, Kageyama 等从家蚕卵中纯化出一种组织蛋白酶 L, 并将其命名为家蚕酸性半胱氨酸蛋白酶 (BCP)。该酶在体外酸性环境下能够水解卵黄原蛋白 (Vg) 的重链、轻链, 以及卵特异蛋白 (ESP)<sup>[12]</sup>。1994年, 赵小凡等利用天然纯化的 BCP 制备抗血清, 对 BCP 的组织分布及合成部位进行了相关研究, 发现在雌蚕蛹期血淋巴和卵母细胞中存在大量的 BCP, 并且随着蛹的发育, 其成熟体逐渐增多, 活性也逐渐上升<sup>[13]</sup>。同年, Yamamoto 等通过免疫组化对 BCP 在家蚕卵中的分布进行了定位研究, 进一步证明了该酶对家蚕卵黄蛋白的降解有重要作用<sup>[14]</sup>。

我们前期对家蚕贮藏蛋白的晶体结构进行研究时, 发现贮藏蛋白具有木瓜蛋白酶的酶切位点<sup>[15]</sup>, 而木瓜蛋白酶与组织蛋白酶同属于半胱氨酸蛋白酶家族。为此, 我们调查了家蚕中组织蛋白酶家族的构成及基因的组织表达情况, 从中鉴定出了一个卵巢特异表达的组织蛋白酶 L 基因, 并对其表达特征进行了研究, 以期为后续研究奠定基础。

## 1 材料与方法

### 1.1 材料

供试家蚕品种由西南大学家蚕基因资源库提供。幼虫在 (25±1) °C、相对湿度 60%、自然光照条件下用桑叶进行饲养。待供试家蚕饲养至幼虫期 5 龄第 3 天时, 于冰冻条件下取头部、体壁、马氏管、中肠、脂肪体、丝腺、血淋巴、

卵巢及精巢, 并置于 -80 °C 冰箱保存。待幼虫化蛹后, 取 1 至 9 d 的雌蛹, 置于 -80 °C 冰箱保存。

### 1.2 试剂

总 RNA 提取试剂盒, 胶回收试剂盒及质粒提取试剂盒购自 Omega 公司; 高效反转录试剂盒购自 Promega 公司; 用于荧光定量 PCR 的试剂, DNA 聚合酶 *r-taq* 及原核表达所用限制性内切酶和连接酶购自宝生物工程 (大连) 有限公司; 克隆所用 DNA 聚合酶 HiFi-*taq* 与大肠杆菌 Trans1-T1, 及原核表达所用菌株 Transetta (DE3) 均购自北京全式金生物技术有限公司; 原核表达载体 pCold-SUMO 由本实验室保存; 用于琼脂糖电泳及 SDS-PAGE 相关试剂均采购自上海生工生物工程技术有限公司。

### 1.3 家蚕组织蛋白酶基因的鉴定及基本信息分析

将木瓜蛋白酶 (GenBank Accession No. AAB02650.1) 的氨基酸序列与 SilkDB 数据库<sup>[16]</sup> (<http://silkworm.genomics.org.cn/>) 进行比对, 获得候选家蚕组织蛋白酶基因。利用 Pfam (<http://pfam.xfam.org/search>) 分析候选基因编码的蛋白质结构域。同时, 对获得的氨基酸序列进行分子质量、等电点预测 ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)) 及信号肽分析 (<http://www.cbs.dtu.dk/services/SignalP/>)。

### 1.4 家蚕组织蛋白酶基因的组织表达模式分析

利用 SilkDB 数据库中家蚕 5 龄第 3 天各组织芯片数据<sup>[17]</sup>, 分析组织蛋白酶家族的基因表达模式。Cluster 3.0 软件用于基因表达水平归一化及聚类分析。

### 1.5 家蚕各组织及蛹、蛾期卵巢 RNA 提取, cDNA 的合成及荧光定量 PCR 检测

将5龄第3天家蚕各组织(包括头部、表皮、马氏管、中肠、脂肪体、丝腺、血细胞、卵巢及精巢)和蛹期及蛾期不同发育阶段的卵巢用液氮研磨,按照RNA提取试剂盒操作说明提取总RNA。使用高效反转录试剂盒合成cDNA并进行荧光定量PCR。根据SilkDB收录的基因编号为BGIBMGA004622的蛋白质编码区(CDS)序列,利用Primer Premier 5.0设计引物,引物序列见表1(①-④)。利用家蚕转录起始因子4A基因(BGIBMGA003186)的表达情况作为内参。引物交由上海生工生物工程技术有限公司合成。荧光定量PCR反应仪器为ABI 7500Fast,反应总体系为20  $\mu$ L(含SYBR预混液10  $\mu$ L,10  $\mu$ mol/L上下游引物各0.8  $\mu$ L,ROX染料0.4  $\mu$ L,cDNA模板2  $\mu$ L,灭菌水6  $\mu$ L),反应条件为:95  $^{\circ}$ C预变性30 s,之后进行40个循环,每个循环为95  $^{\circ}$ C 3 s,60  $^{\circ}$ C 30 s,其中每个模板进行3次重复实验,收集目的基因Ct值和内参基因Ct值进行数据分析。通过公式 $2^{-\Delta\Delta Ct}$ 计算目的基因的相对表达情况。由于使用家蚕转录起始因子4A基因为内参,因此得到的数据为目的基因相对于家蚕转录起始因子4A基因的相对表达量。

### 1.6 蛹期各组织 RT-PCR 检测

提取蛹期各组织(包括头部、表皮、血细胞、脂肪体、嗦囊、卵巢及精巢)提取RNA并反转录为cDNA,利用表1中①-④引物进行RT-PCR检测。反应总体系为25  $\mu$ L(含10 $\times$  r-taq buffer 2.5  $\mu$ L,dNTPs 1.5  $\mu$ L,10  $\mu$ mol/L上下游引物各0.5  $\mu$ L,cDNA模板1  $\mu$ L,r-taq DNA聚合酶0.5  $\mu$ L,灭菌水18.5  $\mu$ L),反应条件为:95  $^{\circ}$ C

预变性5 min,之后进行28个循环,每个循环为95  $^{\circ}$ C 40 s,60  $^{\circ}$ C 40 s,72  $^{\circ}$ C 30 s,最后72  $^{\circ}$ C延伸5 min。利用家蚕转录起始因子4A基因(BGIBMGA003186)的表达情况作为内参。PCR产物经1%琼脂糖电泳检测。

### 1.7 BGIBMGA004622 蛋白的多重序列比对和系统进化分析

帝王斑蝶、二化螟、黑翅果蝇等物种的组织蛋白酶L氨基酸序列下载于NCBI(<http://www.ncbi.nlm.nih.gov/>)。蛋白质多序列比对采用Clustal X软件,比对结果用GeneDoc软件分析。系统进化分析利用MEGA程序中的邻接法构建系统进化树(去除空位,采用1000次重复进行自举检验)。

### 1.8 BGIBMGA004622 基因的克隆及原核表达

根据NCBI收录的BGIBMGA004622基因的蛋白质编码区(CDS)序列,利用Primer Premier 5.0设计引物,引物序列见表1(⑤-⑥)。使用5龄第3天家蚕幼虫卵巢cDNA作为模板,进行PCR扩增,反应总体系为50  $\mu$ L(含10 $\times$  HiFi-taq 缓冲液5  $\mu$ L,dNTPs 3  $\mu$ L,10  $\mu$ mol/L上下游引物各1  $\mu$ L,cDNA模板1  $\mu$ L,HiFi-taq DNA聚合酶1  $\mu$ L,灭菌水38  $\mu$ L),反应条件为:95  $^{\circ}$ C预变性5 min,之后进行35个循环,每个循环为95  $^{\circ}$ C 40 s,60  $^{\circ}$ C 40 s,72  $^{\circ}$ C 90 s,最后72  $^{\circ}$ C延伸5 min。PCR产物经1%琼脂糖电泳后,目的片段使用胶回收试剂盒进行回收。将回收后DNA片段与pCold-SUMO载体分别用Xba I和Xho I内切酶酶切过夜,酶切产物再次进行琼脂糖电泳并切胶回收。回收后产物利用DNA连接酶进行连接并转化入Trans1-T1菌株

表 1 本研究所用引物

Table 1 Primers used in this study

Primer name	Primer sequence (5'-3')
① 4622-F	TCCTTGGATGGGTTTGTG
② 4622-R	TCTCGTGGTGCCTAGATG
③ 3186-F	TTCGTAAGGCTCTTCTCGT
④ 3186-R	CAAAGTTGATAGCAATTTCCCT
⑤ 4622- <i>Xho</i> I-F	CCGCTCGAGGTGAACATTACC GAAAATAC
⑥ 4622- <i>Xba</i> I-R	TGCTCTAGATTATTACAGTTCC ACATAGGTT

Underlined letters indicate restriction enzyme digestion sites.

中, 挑取单克隆并验证后送公司测序。将测序正确后的克隆扩大培养, 并提取质粒, 再将质粒转化入 *Transetta* (DE3) 菌株中, 挑取单克隆进行培养, 待  $OD_{600}$  达到 0.6 后, 加入 0.1 mmol/L IPTG 在 16 °C 继续培养 20 h。利用超声破碎仪对培养完成的大肠杆菌进行破碎, 并分别取离心后上清和沉淀进行 SDS-PAGE 检测。

## 2 结果与分析

### 2.1 家蚕组织蛋白酶家族成员基因的鉴定

通过对家蚕 SilkDB 数据库的筛选, 共鉴定到 15 个候选基因序列, 去除 2 个已经被报道为家蚕组织蛋白酶抑制剂的基因序列, 共得到 13 个组织蛋白酶基因 (表 2)。通过氨基酸序列同源比对发现, 13 个基因所编码的酶中有 5 种为 Cathepsin L 类型, 3 种为 Cathepsin O, 其余 4 种分属 Cathepsin B、K、F。对该家族基因编码的酶类的结构域进行分析, 除了编号为 BGIBMGA005131 具有多个重复的 Cystatin 结构域, 其他酶类均含有组织蛋白酶保守的结构域 Inhibitor I29 和/或 Peptidase C1。编码蛋白信号

肽预测显示, 该家族中只有 3 种酶具有信号肽结构。所编码的各蛋白相对分子量从 20 到 200 kDa 不等, 而等电点则从 4 到 10 不等。

### 2.2 家蚕组织蛋白酶基因组织表达模式

经过在 SilkDB 数据库中查询, 该家族中只有编号为 *BGIBMGA009146* 的基因没有 5 龄第 3 天芯片数据。其他芯片数据均下载于 SilkDB 的 Microarray 数据库中。通过 Cluster 软件对其进行归一化和聚类分析。发现该家族酶的组织表达模式可以分为 4 类, 分别命名为 I-IV (图 1)。其中, 第 I 类主要在头部、中肠、前/中部丝腺及后部丝腺有表达, *BGIBMGA009184* 还在表皮中有较高量表达。第 II 类主要在中肠和马氏管表达, *BGIBMGA009230* 在丝腺和血淋巴有表达, 而 *BGIBMGA009229* 则在卵巢中有所表达。第 III 类基因集中表达在精巢、卵巢、脂肪体和表皮中, 并且在血淋巴中 (除 *BGIBMGA007061*)、后部丝腺中 (除 *BGIBMGA011342*) 有一定量的表达; 而且 *BGIBMGA011342* 是该类基因中唯一一个在前/中部丝腺中表达的基因。第 IV 类基因中, *BGIBMGA004622* 和 *BGIBMGA004421* 表达模式比较接近, 均在生殖腺和马氏管中有较高量表达; *BGIBMGA007066* 除在卵巢中有所表达外, 还在脂肪体、表皮和前/中部丝腺中有表达。

### 2.3 *BGIBMGA004622* 基因在幼虫 5 龄第 3 天各组织的表达模式

荧光定量 PCR 结果显示, 在家蚕幼虫 5 龄第 3 天各组织中, *BGIBMGA004622* 在卵巢中的相对表达量为 0.0088, 而在其他组织中表达量极低, 可以认为此基因在卵巢中有较为特异地表达 (图 2)。该结果与 2.2 中芯片数据分析相一致。

表 2 组织蛋白酶基因家族基本信息分析

Table 2 Basic information of the cathepsin family

Gene ID	Gene location	Strand	Domain	Signal peptide	Relative molecular mass (kDa)	pI	Proteinase type
BGIBMGA005131	nscaf2823: 1 951 268–1 963 838	+	Cystatin Inhibitor_I29 Peptidase_C1	No	206.8	5.74	Cathepsin F
BGIBMGA011342	nscaf3026: 2 500 218–2 510 152	+	Inhibitor_I29 Peptidase_C1	Yes	38.1	5.90	Cathepsin L
BGIBMGA009139	nscaf2937: 327 740–343 618	–	Inhibitor_I29 Peptidase_C1	No	55.3	6.02	Cathepsin L
BGIBMGA009184	nscaf2938: 354 195–371 013	+	Peptidase_C1	No	77.6	6.37	Cathepsin L
BGIBMGA009230	nscaf2943: 1 775 657–1 778 622	–	Peptidase_C1	No	22.9	4.95	Cathepsin O
BGIBMGA004622	nscaf2801: 19 789–28 288	+	Inhibitor_I29 Peptidase_C1	No	27.8	9.66	Cathepsin L
BGIBMGA007061	nscaf2865: 826 993–828 006	+	Peptidase_C1 Peptidase_C1	Yes	37.6	6.00	Cathepsin B
BGIBMGA009146	nscaf2937: 476 541–478 797	+	Peptidase_C1	No	28.4	4.84	Cathepsin K
BGIBMGA004421	nscaf2795: 591 126–602 053	–	Peptidase_C1	No	27.2	5.53	Cathepsin B
BGIBMGA007066	nscaf2865: 1 161 265–1 162 530	+	Inhibitor_I29	No	25.9	7.67	Cathepsin O
BGIBMGA006893	nscaf2860: 3 057 768–3 061 906	–	Peptidase_C1	No	27.6	4.59	Cathepsin L
BGIBMGA009229	nscaf2943: 1 788 242–1 788 906	–	Inhibitor_I29	Yes	20.9	9.32	Cathepsin O
BGIBMGA009231	nscaf2943: 1 767 933–1 773 061	–	Inhibitor_I29 Peptidase_C1	No	23.9	9.12	Cathepsin O

## 2.4 *BGIBMGA004622* 基因在蛹期各组织及各时期表达模式

RT-PCR 结果显示, 在蛹期各组织中, *BGIBMGA004622* 也在卵巢有特异地表达 (图 3A), 这与其在幼虫 5 龄第 3 天各组织的表达模式一致。对其在蛹期及蛾期不同发育时期卵巢中的表达模式分析发现, 该基因的表达量在化蛹前期的卵巢中有一个明显的上升趋势, 且在

蛹 3 d 达到最高值, 随后逐渐降低。而到了蛹 5 d 后, 表达量又有所上升, 并在蛹 6 d 达到另外一个峰值, 随后表达量逐渐下降, 蛹 8 d 到蛾期基本不表达 (图 3B)。

## 2.5 *BGIBMGA004622* 与其他物种的多重序列比对与系统进化分析

通过将 *BGIBMGA004622* 基因编码的蛋白质氨基酸序列与其他物种的 Cathepsin L 氨基酸

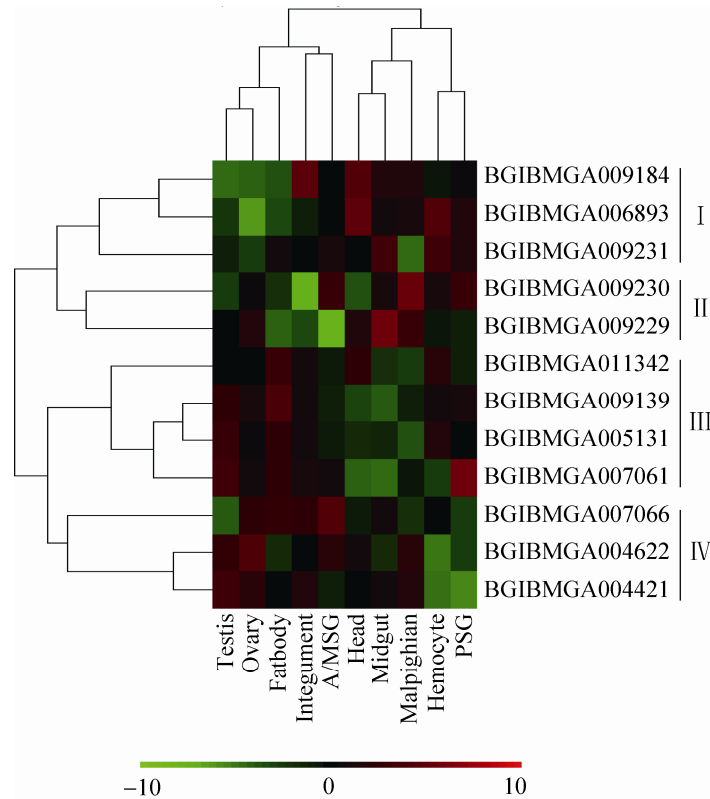


图 1 组织蛋白酶基因家族表达模式分析

Fig. 1 Expression patterns of the cathepsin family genes. A/MSG: anterior and middle part of silk gland; PSG: posterior part of silk gland.

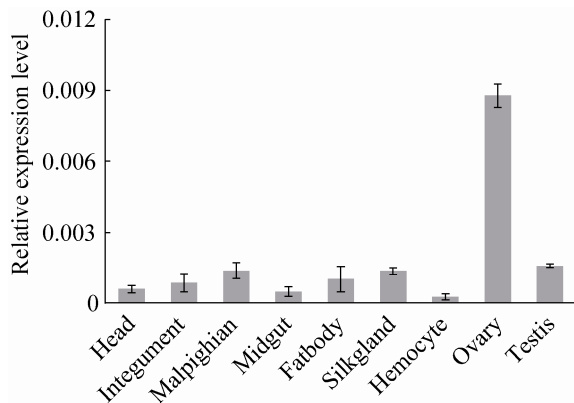


图 2 *BGIBMGA004622* 基因在幼虫期各组织的表达模式

Fig. 2 Expression pattern of *BGIBMGA004622* in larval tissues.

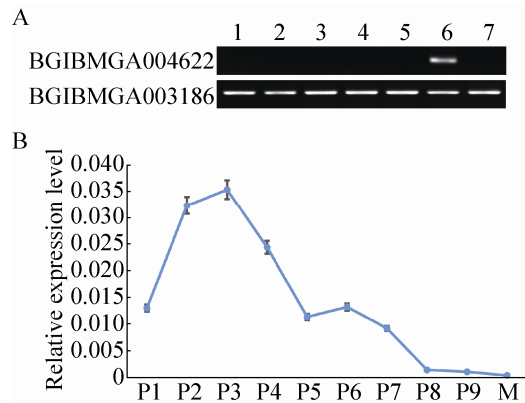


图 3 *BGIBMGA004622* 基因在蛹期各组织及各时期卵巢的表达模式

Fig. 3 Expression pattern of *BGIBMGA004622* in various pupal tissues and ovaries. (A) 1: head; 2: integument; 3: hemocyte; 4: fatbody; 5: crop; 6: ovary; 7: testis. (B) P1 to P9 represents the ovaries from the 1st day of pupae to the 9th day of pupae, respectively; M: ovaries from moths.

序列进行多重序列比对,发现这些酶 N 端结构差异较大,然而在 C 端却具有较高的同源性,并且在其两个结构域 Inhibitor I29 和 Peptidase C1 中,各个酶的氨基酸序列有着较高的保守性。而其催化活性位点(217 位的半胱氨酸,354 位的组氨酸,370 位的天冬酰胺),则表现出高度的保守性(图 4)。系统进化分析发现,家蚕与鳞翅目的帝王斑蝶亲缘关系很近,聚为一支,而与其他昆虫,黑翅果蝇和二化螟亲缘较近,与软体动物和哺乳动物的关系则较远(图 5)。

## 2.6 *BGIBMGA004622* 基因的克隆与原核表达

对 *BGIBMGA004622* 进行克隆,并将克隆片段与 pCold-SUMO 载体连接,构建重组质粒(图 5A)。重组质粒转化入 *Transetta* (DE3) 菌株中,进行诱导表达并检测。SDS-PAGE 结果显示,重组蛋白以包涵体的形式在沉淀中大量表达(图 6B, 3-4 泳道),而在上清中没有检测到重组蛋白的表达(图 6B, 1-2 泳道)。由于原核表达时所选用的 pCold-SUMO 载体上带有 SUMO 助溶标签,其分子量约为 17 kDa,再加上我们根据 NCBI 所报道的 *BGIBMGA004622* 基因的全长序列设计引物所获得的编码蛋白的理论分子量约为 44 kDa,因此融合蛋白理论分子量为 61 kDa,与体外表达的结果一致。

## 3 讨论

家蚕作为重要的鳞翅目经济昆虫,其相关研究一直广受关注。随着家蚕基因组框架图绘制完成<sup>[18-19]</sup>,越来越多的家族基因被鉴定并得到研究<sup>[20-21]</sup>。本研究通过对家蚕基因组数据库进行筛选,共得到 13 个编码组织蛋白酶的基因,涵盖组织蛋白酶 L、F、O、B、K 五个亚家族。

这些酶在其基因组成和蛋白性质上都有很大差别。通过对该家族在家蚕幼虫 5 龄第 3 天的组织表达模式进行分析,发现其表达模式也各有不同,进一步暗示了该家族伴随着家蚕的进化过程在功能上出现了巨大分化。通过聚类分析,所鉴定到的基因可以聚为 4 类。每一类基因都有其较为独特的表达模式,这些特殊的表达模式也为我们后期研究其功能提供了相应依据。其中编号 *BGIBMGA011342* 的基因最初被报道为家蚕酸性半胱氨酸蛋白酶 (BCP)<sup>[12]</sup>。因其在丝腺中表达量较高,且对丝素蛋白有水解作用,后被报道为丝素蛋白酶 (Fibroinase)<sup>[22-23]</sup>。后续研究表明该酶对丝胶蛋白也有水解作用<sup>[24]</sup>。除此之外,该家族中其他酶尚未见功能报道。

卵巢是雌性动物的生殖器官,伴随着动物的发育及卵母细胞的成熟过程,卵巢在其形态和功能上都会发生相应的变化。在许多动物中,卵巢组织蛋白酶也一直得到广泛地关注和研究。Oksjoki 等在对小鼠卵巢进行研究中发现,小鼠的卵巢中存在多种组织蛋白酶,且其表达模式也有着很大的差异。利用免疫定位杂交发现,不同的组织蛋白酶定位于小鼠卵巢的不同位置,表明其在功能上也有所不同。组织蛋白酶 L 主要存在于小鼠的黄体细胞中,推测其可能参与小鼠卵巢甾类物质的产生<sup>[25]</sup>。另有研究发现,组织蛋白酶对于卵母细胞中卵泡的发育及成熟有着重要作用<sup>[26-27]</sup>。在斑点叉尾鲷中的相关研究发现,卵黄生成及卵泡发育过程与组织蛋白酶的表达模式相一致<sup>[28]</sup>。

家蚕幼虫卵巢位于第 5 腹节背脉管的两侧,左右成对,略呈三角形。到幼虫 3 龄期,卵巢小室形成一个大型的端细胞。至 4 龄期,卵巢小室延长,弯曲成管状,盘旋在卵巢室内,称



<i>Bombyx mori</i>	MMLIVLLLQISLVNVCNITENTLKIKSLIMKYKNNRVHRSMPMTSLDGFVQDALMFFDRTERFKDVLNVY	: 69
<i>Tityus serrtlatus</i>	-----	: -
<i>Apostichopus japonicas</i>	-----	: -
<i>Crassostrea gigas</i>	-----	: -
<i>Hydra vulgaris</i>	-----	: -
<i>Longidorus elongates</i>	-----MIRIT	: 5
<i>Plautia stali</i>	-----	: -
<i>Danio rerio</i>	-----	: -
<i>Canis lupus familiaris</i>	-----MGAPAG	: 6
<i>Danaus plexippus</i>	-----MTSLP-----LFKKTARTRDYLSPF	: 20
<i>Chilo suppressalis</i>	-----RPPRS	: 5
<i>Drosaphila persimilis</i>	-----MVKILFLYG	: 9

<i>Bombyx mori</i>	NYSECVGDEGLMEKHVLLKGLLIEH-----LPRRHHEHYKAIHNKLY	: 111
<i>Tityus serrtlatus</i>	-MLREAVFALLVALAVAS-PFNKE-----LDRHHEIFQVHHEKQY	: 38
<i>Apostichopus japonicas</i>	-MLKSAVFLACVAGALCFTIIDKG-----FDDTWEAWKQTHSKQY	: 39
<i>Crassostrea gigas</i>	--MNVLWIVSVLAVARGATVQTNVQWFDLEAAQKHPEQLHLKAKAGINYPYEQAKKFKILHDKTY	: 67
<i>Hydra vulgaris</i>	--MKVFCALLLLGVTLAYTIER-----PVKDESIQWRMYHNKQY	: 38
<i>Longidorus elongates</i>	LLLHSIFLLGFVNSEQISQIQEHPRNLLIN-----HPYYPVWTFNKLKHAKSY	: 54
<i>Plautia stali</i>	---MKVFLVAAACLTLVYIAEAA-----ESTRFPSQYKQY	: 34
<i>Danio rerio</i>	-MMFGSLLFAVCCSAAALAHFNKN-----LDQHWELWKKHVKLY	: 38
<i>Canis lupus familiaris</i>	SFMKWLVLPLPCSYAVAQVHKDP-----TLDHHAHLWKKTYSKQY	: 47
<i>Danaus plexippus</i>	NESGVQRSREKCKYRKDYIVRVPEN-----LLDVHHEEYKTKYNKQY	: 62
<i>Chilo suppressalis</i>	--SGLYARVSWCRQLERSRVRI-----QTIPDKAY	: 33
<i>Drosaphila persimilis</i>	--FSTLSNINANYLENESLTDINN-----EHEQFHKINNKTY	: 44

Cathepsin propeptide inhibitor domain

<i>Bombyx mori</i>	SSTHHEMAALMKWRQNLRRVAREHNREYLAIQSYSLHINHFQDMHVTEYFGKV LKLIKAFPLFDPAED-	: 179
<i>Tityus serrtlatus</i>	QI-YBEGFRLITFEQNVIRINKHNLEYQLCRKSYRLGLNAYADLTPQEFVKQMGFIIRANRTSKA---	: 103
<i>Apostichopus japonicas</i>	TK-EEEDNRKRIWEDNLQKVKHNTHEHSLGLHSYTLGMNKYADLRGEEFVQMMNGLKFDASRE-RQ---	: 103
<i>Crassostrea gigas</i>	DALEEBRRRFEIARENVQKTEEHNKLYHLCKKSYLGVNQFSDLKHEEFVKYNG---LKKTSLKDG---	: 131
<i>Hydra vulgaris</i>	SHDGEETVRYTIWKDNEERRIREHN---LKGGEILKMNQFGDMTNSERKAFNG--YLSHKVNG---	: 97
<i>Longidorus elongates</i>	KTKDEELLRFQVFPASNHKVIEQHNIEYEACQHSFALSILNKEADMTNAEERQRMNGFKLPAKRLKASQP	: 123
<i>Plautia stali</i>	PSDSVBRYYRKKVYKQNEKQVREHNERYERCEVTYKMAINHLADMHPREEMATFLGFNRSRLRATNKVP--	: 101
<i>Danio rerio</i>	SCEDBEVGRRELWERNLELIAIHNLEASMGHSHYDLAINHMADMTTEETLQTLAVTRVPPGFKRP---	: 103
<i>Canis lupus familiaris</i>	KEENEVAVARRLIWEKNLKFMVLIHNLEHSMGMHSHYDLGMLNHLGDMTGEEVTSLMSGSLRVSPQWQ---	: 111
<i>Danaus plexippus</i>	QSRFHRSALSTWRSNMKRVAGHNOEYLAACKQAYTLHINHFQDWSIFSYIKQLLKLIRTLPLFDPAED-	: 130
<i>Chilo suppressalis</i>	DSHRDEMAALSQWRNRNKLIAAHNKRFLAGEISYTLHINQFADWHPEEYFKKILKLPDFTIPLLDPALD-	: 101
<i>Drosaphila persimilis</i>	SRNFDERTSLKAEVNYKIIKDKHNKYQDQOTTERLATNIMADMSTEGYLKNFLRLLKSQ--SNVADD-	: 110

Cla proteinase domain

<i>Bombyx mori</i>	--HHKTAYRHNRRCKVFKRTDWRDQCFKPRREEQWCCGAYAFVATHALCAQLYKRHG---EWNELSPQ	: 243
<i>Tityus serrtlatus</i>	---GRVPHRVQGVKLPDKVDWRDQGITVPTKDAQCGSOWAFSATGSLBQWAKAHGKSGLSLSEQ	: 168
<i>Apostichopus japonicas</i>	---GIKFLSYAKFQAPDSVDWRDQGYVTPVKDQGCCGSOWAFSTTGSLEBQOHRFTSG---VLTSLSEQ	: 163
<i>Crassostrea gigas</i>	----CSSYLAANLVEPDSVDWRKKGIVYTDVKNQGCCGSOWSFSTTGSLEBQOHRFKSG---KLVLSSES	: 195
<i>Hydra vulgaris</i>	----STFLTPNNFVAPDITVDWRNEGYVTPVKDQGCCGSOWAFSTTGSLEBQOHRFKTG---KLVLSSEQ	: 158
<i>Longidorus elongates</i>	LKEDGMIPEMPDNVTIPDSVDWRKQGYVTPVKDQGCCGSOWAFSATGSLBQOYKQOTG---KLVLSSEQ	: 189
<i>Plautia stali</i>	---EGIPFRHNKDAVIQKEVDWRQCGAISPVKDGHCSSOWAFSSGTALBAHTFLKKG---RRVLSSEQ	: 164
<i>Danio rerio</i>	----TAEYVSSSFVAVPDTLDWRDKQGYVTSVKNQGCCGSOWAFSSVGALEBQMLMKTTC---KLVDLSPQ	: 165
<i>Canis lupus familiaris</i>	---NVTYRSNSNQKLPDSVDWRKQCVTEVVKYQGCCGAWAFSAVGALEBAOLKLTG---KLVLSLQAQ	: 173
<i>Danaus plexippus</i>	---RRRTYRNTFDTRLPERVDWRKGERPRLEEQFHCCGAYAFATIHAVQAQVYKRRHG---DWRELSPO	: 194
<i>Chilo suppressalis</i>	--IHRSHYRHLANKAIPDRVDWASAKGEKPKLEEQWCCGAYAFVAHAHAQALYKHKHG---LWGELSPO	: 165
<i>Drosaphila persimilis</i>	--NIAEIVGSSQMTNIPESLDWRRKQFTTTPSQNQSCSGSYAFSIAESTEGQIFKRTG---KILSLSEQ	: 174



<i>Bombyx mori</i>	QIVDCS-IKDGNGCDGGSLRGAALRYAAREG-LVMEESHYPYVGKKEYCRVDSNLVRRRPRRATLPSGD	: 310
<i>Tityus serrtlatus</i>	NLVDCS-GAQCNLGCSSGLMDSAFDYTKANSGITDESQYPYALDGSCHFKKTCIGATLIGVVDIPLSGD	: 236
<i>Apostichopus japonicas</i>	NLVDCS-ISYGNNGCEGGLMDYAFQYIKDNLGIDTEDEKYPYEAEDDTCRSPDNVGAIDSGVVDVDSGD	: 233
<i>Crassostrea gigas</i>	QIVDCS-QSFGNCGCNGGLMDNAFKYIKSVGGLSEEDYYPKPKQCTCKEDDTKVAATDITGVVDVDSGS	: 261
<i>Hydra vulgaris</i>	NLVDCS-TAYCNGNCGDGLMDNAFTYIKENKGIIDSEASYPYTAEDCKCVFKKSSVAAEDTIGVVDIPEGN	: 226
<i>Longidorus elongates</i>	NLVDCD-VNGDDEGCNGGYMDGAFQYVETNKGIDTEASYPYKGRDCRORFKSEVDGATDITGVVDIPEGN	: 257
<i>Plautia stali</i>	NLIDCS-LNYGNNGCEGGLMEQAFQYVRDNDGIDTBEAYPYEGEDSECRKKNVNGATDAGEVTTIPLSGD	: 232
<i>Danio rerio</i>	NLVDCS-SKYGNLGCNGGYMSQAFQYVIDNNGIDSEESSYPYQGTQSCRVDPSQRAANCTSSKFFVSGGD	: 233
<i>Canis lupus familiaris</i>	NLVDCS-TEKYGNKGCNGGFMITAFQYIIDNNGIDSEASYPYKAMNCKORVDSKKRAATCSKYTELFPFSG	: 242
<i>Danaus plexippus</i>	QIVDCS-FKDGNEGCDGGSLQAALRYVARDG-LIRETYYPYIGQRCAHNSSESVSARVRRRSSLPPGD	: 261
<i>Chilo suppressalis</i>	QIVDCS-AADGNEGCDGGSLRGAALRYAARSQ-LVSEQYYPYTGKKGCHCKSSGLLARTKPKNAMLPPGD	: 232
<i>Drosaphila persimilis</i>	QIVDCS-VSHGNGGCTGGSLRNTLKYIQSTGIMRSDDYKYVSKKCKQIVRDLSSVNNITSSAILLVNND	: 242

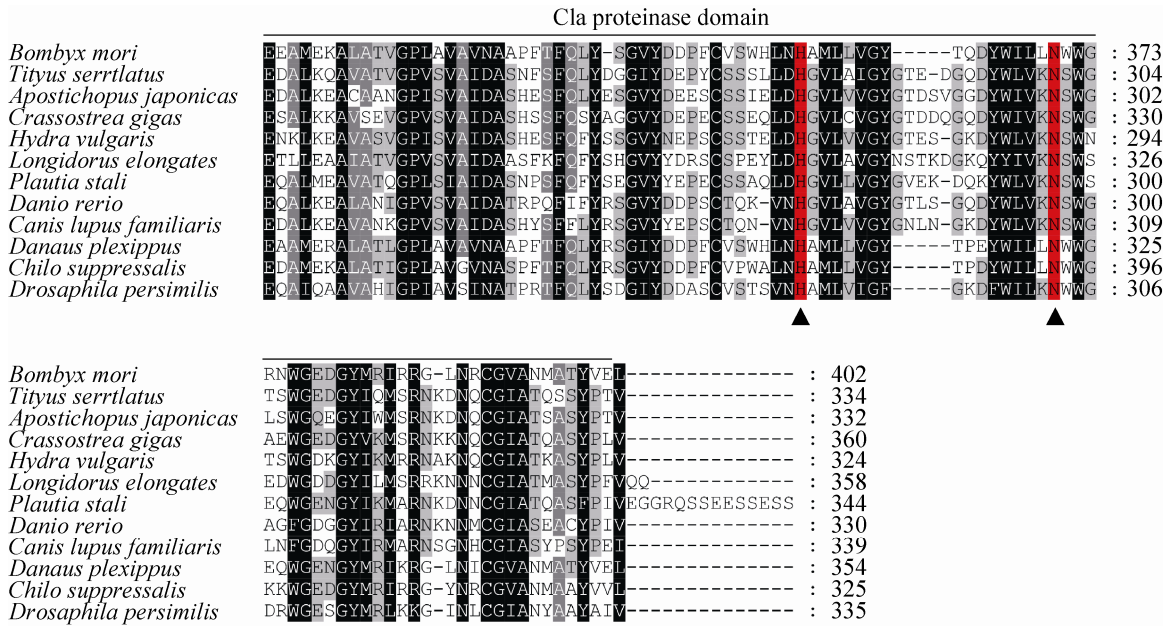


图 4 BGIBMGA004622 与其他物种组织蛋白酶 L 的多序列比对

Fig. 4 Sequence alignment of BGIBMGA004622 with other species' cathepsin L. *Bombyx mori* (NP\_001091806.1); *Danaus plexippus* (EHJ67628.1); *Chilo suppressalis* (AFQ01140.1); *Drosophila persimilis* (XP\_002014338.1); *Tityus serrtlatus* (CDJ26735.1); *Apostichopus japonicas* (ABW98676.1); *Crassostrea gigas* (EKC36430.1); *Hydra vulgaris* (AAO65603.1); *Longidorus elongatus* (AAY45871.1); *Plautia stali* (BAF94151.1); *Danio rerio* (NP\_001017661.1); *Canis lupus familiaris* (NP\_001002938.2); "▲" represents the active-site of cathepsin L.

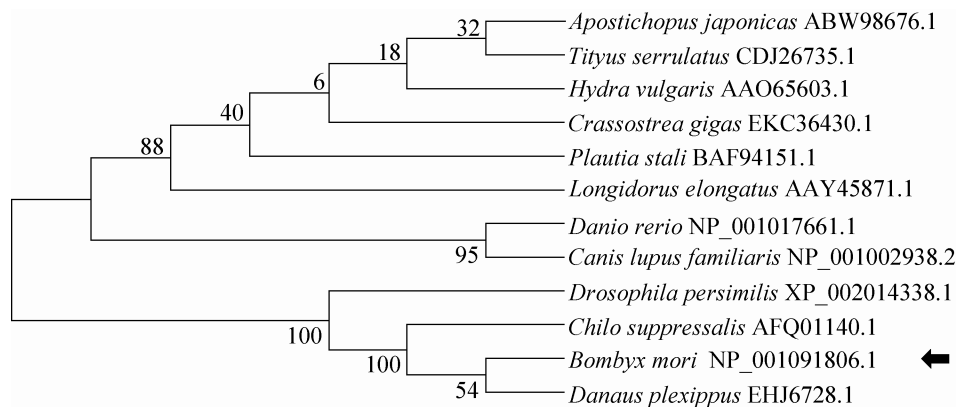


图 5 BGIBMGA004622 与其他物种组织蛋白酶 L 的系统进化树

Fig. 5 Cladogram analysis of BGIBMGA004622 with other species' cathepsin L. Black arrow indicates BGIBMGA004622.

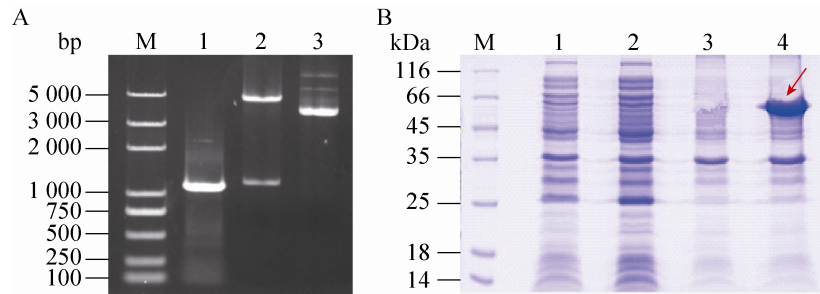


图 6 *BGIBMGA004622* 的克隆与原核表达

Fig. 6 Cloning and prokaryotic expression of *BGIBMGA004622*. (A) M: DNA marker; 1: *BGIBMGA004622* fragment; 2: double digestion (*Xba* I and *Xho* I) product of pCold-SUMO-4622; 3: pCold-SUMO vector. (B) M: protein marker; 1: supernatant (control); 2: supernatant (IPTG induced); 3: precipitation (control); 4: precipitation (IPTG induced). Arrow indicates the recombinant protein.

为卵巢管。化蛹后 2–3 d, 卵巢管显著伸长, 使卵巢膜破裂, 每个卵巢伸出 4 条卵巢管, 游离于体腔, 末端以端丝相连<sup>[29]</sup>。卵巢膜破裂的过程中伴随着组织的水解, 而我们的研究发现, *BGIBMGA004622* 基因在化蛹后表达量逐渐升高, 推测其可能参与了水解卵巢膜的过程, 从而利于卵巢管的伸出。家蚕卵室内的初级卵母细胞在幼虫期靠滋养细胞来提供其发育所需营养, 而在蛹期则靠卵泡细胞连续不断地供给营养物质, 从而不断生长, 逐渐由圆形变成椭圆形, 并不断下移。当初级卵母细胞生长完成时, 其周围的卵泡细胞就会分泌形成具有多层构造的卵黄膜。随后, 卵泡细胞又在卵黄膜的表面上, 逐步形成卵壳的结构。卵壳形成后, 卵上下两端的上皮细胞破裂, 上下相邻的卵排成列, 沿卵巢管下降到输卵管, 生长期完成, 时间大致在化蛹后 7–8 d。本研究同时发现 *BGIBMGA004622* 在蛹期第 3 天达到较高表达量, 并在蛹期 6 d 时达到另外一个较低峰值, 随后基本上不再表达, 暗示了其可能参与卵巢管伸长、卵泡细胞中营养物质的水解以

及初级卵母细胞的生长过程。

在对 *BGIBMGA004622* 基因进行原核表达时发现, 该酶在 pCold-SUMO 助溶标签的存在下, 并且在 16 °C 低温进行诱导, 重组蛋白仍以包涵体的形式表达。且后续实验中发现, 表达的重组蛋白在 8 mol/L 尿素或 6 mol/L 盐酸胍条件下也难以溶解, 这对该蛋白的纯化造成了很大困难。在接下来的研究中, 我们将考虑利用本实验室所构建好的家蚕杆状病毒表达系统<sup>[30]</sup>对该酶进行表达并进行相关研究。

本研究通过对家蚕基因组数据库进行筛选, 共鉴定到 13 种家蚕组织蛋白酶基因, 并对其基因和所编码蛋白的基本信息和表达模式进行了分析。同时鉴定到一种编码卵巢特异表达组织蛋白酶 L 的基因 *BGIBMGA004622*, 并分析了其在家蚕幼虫 5 龄第 3 天和蛹期及蛾期卵巢中的表达模式, 以及该酶与其他物种组织蛋白酶 L 的同源分析和进化分析。并且利用大肠杆菌对该酶进行了原核表达。研究内容将对后续研究提供基础信息。

## REFERENCES

- [1] Fonović M, Turk B. Cysteine cathepsins and extracellular matrix degradation. *Biochim Biophys Acta*, 2014, 1840(8): 2560–2570.
- [2] Liu J. Physiology function analysis of lysosomal cysteine proteases[D]. Hefei: USTC, 2006 (in Chinese).  
刘健. 溶酶体半胱氨酸蛋白酶生理功能的研究[D]. 合肥: 中国科技大学, 2006.
- [3] Yamahama Y, Uto N, Tamotsu S, et al. *In vivo* activation of pro-form *Bombyx* cysteine protease (BCP) in silkworm eggs: localization of yolk proteins and BCP, and acidification of yolk granules. *J Insect Physiol*, 2003, 49: 131–140.
- [4] Schreuder H, Liesum A, Kroll K, et al. Crystal structure of cathepsin A, a novel target for the treatment of cardiovascular diseases. *Biochem Biophys Res Commun*, 2014, 445(2): 451–456.
- [5] Turk D, Podobnik M, Kuhelj R, et al. Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett*, 1996, 384: 211–214.
- [6] Chen LZ, Zhou JL, Zhou YZ, et al. Molecular cloning of two *Rhipicephalus haemaphysaloides* cathepsin L-like cysteine proteinase gene. *Chin J Biotech*, 2004, 20(2): 203–208 (in Chinese).  
陈灵芝, 周金林, 周勇志, 等. 镰形扇头蜱两个组织蛋白酶 L-样半胱氨酸蛋白酶新基因的克隆与序列分析. *生物工程学报*, 2004, 20(3): 203–208.
- [7] Wang GF, Peng DL, Sun JH, et al. Cloning and sequence analysis of a new cathepsin L-like cysteine proteinase gene from *Ditylenchus destructor*. *Chin J Biotech*, 2011, 27(1): 60–68 (in Chinese).  
王高峰, 彭德良, 孙建华, 等. 马铃薯腐烂茎线虫 L 型半胱氨酸蛋白酶新基因 (*Dd-cpl-1*) 的克隆与序列分析. *生物工程学报*, 2011, 27(1): 60–68.
- [8] Medina M, Leon P, Vallejo CG. *Drosophila* cathepsin B-like proteinase: a suggested role in yolk degradation. *Arch Biochem Biophys*. 1988, 263(2): 355–363.
- [9] Xu XL, Zhao XF, Wang JX. Distribution and biosynthesis sites of cathepsin B in *Helicoverpa armigera*. *Zool Res*, 2001, 22(3): 242–245 (in Chinese).  
徐夏莲, 赵小凡, 王金星. 棉铃虫组织蛋白酶 B 组织分布与合成部位的研究. *动物学研究*, 2001, 22(3): 242–245.
- [10] Yang XM, Hou LJ, Dong DJ, et al. Expression and variations in activity of cathepsin B-like proteinase during the ontogenetic process of the cotton boll worm: *Helicoverpa armigera*. *J Shandong Univ: Sci Ed*, 2005, 40(5): 119–124 (in Chinese).  
杨晓梅, 侯立静, 董杜鹃, 等. 组织蛋白酶 B 在棉铃虫个体发育过程中的表达及活性研究. *山东大学学报: 理学版*, 2005, 40(5): 119–124.
- [11] Zhang Y, Lu YX, Liu J, et al. A regulatory pathway, ecdysone-transcription factor relish-cathepsin L, is involved in insect fat body dissociation. *PLoS Genet*, 2013, 9(2): e1003273.
- [12] Kageyama T, Takahashi S. Purification and characterization of a cysteine proteinase from silkworm eggs. *Eur J Biochem*. 1990, 193: 203–210.
- [13] Zhao XF, Gao QJ. Distribution and biosynthesis sites of acid cysteine proteinase in silkworm *Bombyx mori*. *Acta Zool Sin*, 1994, 40(1): 24–29 (in Chinese).  
赵小凡, 高桥进. 家蚕酸性半胱氨酸蛋白酶的分布及合成部位研究. *动物学报*, 1994, 40(1): 24–29.
- [14] Yamamoto Y, Zhao XF, Suzuki A, et al. Cysteine proteinase from the eggs of the silkworm, *Bombyx mori*: site of synthesis and a suggested role in yolk protein degradation. *J Insect Physiol*, 1994, 40(5): 447–454.
- [15] Hou Y, Li JW, Li Y, et al. Crystal structure of *Bombyx mori* arylphorins reveals a 3:3 heterohexamer with multiple papain cleavage sites. *Protein Sci*, 2014, 23(6): 735–746.
- [16] Duan J, Li RQ, Cheng DJ, et al. SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res*, 2010, 38(suppl 1):

- D453–D456.
- [17] Xia QY, Cheng DJ, Duan J, et al. Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. *Genome Biol*, 2007, 8(8): R162.
- [18] Xia QY, Zhou ZY, Lu C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 2004, 306(5703): 1937–1940.
- [19] Xia QY, Guo YR, Zhang Z, et al. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*, 2009, 326(5951): 433–436.
- [20] Zhao P, Dong Z, Duan J, et al. Genome-wide identification and immune response analysis of serine protease inhibitor genes in the silkworm, *Bombyx mori*. *PLoS ONE*, 2012, 7(2): e31168.
- [21] Zhao P, Wang GH, Dong ZM, et al. Genome-wide identification and expression analysis of serine proteases and homologs in the silkworm *Bombyx mori*. *BMC Genomics*, 2010, 11: 405.
- [22] Watanabe M, Kamei K, Sumida M. Fibroinase activity of silk gland in larval and early pupal development of the silkworm, *Bombyx mori* assayed with a fluorescent quenched peptide substrate. *J Insect Biotech Sericol*, 2006, 75(3): 115–126.
- [23] Watanabe M, Kotera T, Yura A, et al. Enzymatic properties of fibroinase of silk gland from day one pupa of the silkworm, *Bombyx mori*. *J Insect Biotech Sericol*, 2006, 75(1): 39–46.
- [24] Watanabe M, Kamei K, Sumida M. Sericin digestion by fibroinase, a cathepsin L-like cysteine proteinase, of *Bombyx mori* silk gland. *J Insect Biotech Sericol*, 2007, 76(1): 9–15.
- [25] Oksjoki S, Soderstrom M, Vuorio E, et al. Differential expression patterns of cathepsins B, H, K, L and S in the mouse ovary. *Mol Hum Reprod*, 2001, 7(1): 27–34.
- [26] Eykelbosh A, Kraak G. A role for the lysosomal protease cathepsin B in zebrafish follicular apoptosis. *Comp Biochem Physiol A Mol Integr Physiol*, 2010, 156: 218–223.
- [27] Carnevali O, Cionna C, Tosti L, et al. Role of cathepsins in ovarian follicle growth and maturation. *Gen Comp Endocrinol*, 2006, 146: 195–203.
- [28] Barrero M, Small B, D'Abramo L. Comparison of estradiol, testosterone, vitellogenin and cathepsin profiles among young adult channel catfish (*Ictalurus punctatus*) females from four selectively bred strains. *Aquaculture*, 2007, 264: 390–397.
- [29] Wu ZD. *Anatomy and Physiology of Silkworm*. Beijing: China Agriculture Press, 1989: 7 (in Chinese).  
吴载德. 蚕体解剖生理学. 北京: 中国农业出版社, 1989: 7.
- [30] Wang X, Li Y, Chen HF, et al. Expression of sacro/endoplasmic reticulum  $Ca^{2+}$ -ATPase from *Bombyx mori* by baculovirus expression system. *Chin J Biotech*, 2015, 31(3): 421–430 (in Chinese).  
王鑫, 李懿, 陈慧芳, 等. 利用杆状病毒表达家蚕肌质网型钙离子 ATP 酶蛋白. *生物工程学报*, 2015, 31(3): 421–430.

(本文责编 陈宏宇)