



刘莎, 梁文凯, 王玉琴, 李美娇, 陈芬莲, 罗云菲, 朱家颖. 蠊蛄组织蛋白酶基因鉴定及表达特征分析 [J]. 环境昆虫学报, 2024, 46 (5): 1161 - 1170. LIU Sha, LIANG Wen-Kai, WANG Yu-Qin, LI Mei-Jiao, CHEN Fen-Lian, LUO Yun-Fei, ZHU Jia-Ying. Identification and expression profiling of cathepsin genes in *Arma custos* [J]. *Journal of Environmental Entomology*, 2024, 46 (5): 1161 - 1170.

蠊蛄组织蛋白酶基因鉴定及表达特征分析

刘莎, 梁文凯, 王玉琴, 李美娇, 陈芬莲, 罗云菲, 朱家颖*

(西南林业大学生物多样性保护学院, 云南省森林灾害预警与重点实验室, 昆明 650224)

摘要: 组织蛋白酶在昆虫的消化、发育与变态中起重要作用, 是捕食性蝽唾液(毒液)中普遍存在的成分, 但其生理功能尚不清楚。本文利用同源比对的方法从蠊蛄 *Arma custos* 基因组中鉴定出组织蛋白酶基因, 使用生物信息学软件分析其序列特征和进化关系, 采用 RT-PCR 技术分析它们在成虫不同组织中的表达特征。结果表明, 蠊蛄基因组中有 37 个组织蛋白酶基因, 根据结构域分为组织蛋白酶 B (*AcCABI-4*)、组织蛋白酶 D (*AcCAD1-13*) 和组织蛋白酶 L (*AcCALI-20*)。多序列比对及结构域预测结果表明, 组织蛋白酶 B 和组织蛋白酶 L 均含有 Peptidase_C1 结构域、保守的酶催化位点(谷氨酰胺、半胱氨酸、组氨酸和天冬酰胺)以及 1 个由半胱氨酸残基(C)和组氨酸残基(H)构成的催化二联体, 而组织蛋白酶 D 中存在保守 Asp 结构域和 2 个保守的酶催化位点(天冬氨酸)。RT-PCR 结果显示, 除 *AcCAL7* 和 *AcCAB2* 外, 其余 33 个组织蛋白酶基因均在肠道中表达, 且绝大多数表现特异性或高表达。*AcCAL2*、*AcCAL5-7*、*AcCAL9* 等 20 个组织蛋白酶基因在唾液腺中表达, 且 *AcCAL2* 仅在主腺后叶和副腺表达, *AcCAL7* 仅在主腺前叶表达, *AcCAD13* 仅在主腺后叶表达, *AcCAB2* 仅在主腺前叶和主腺后叶表达。这些结果表明, 多数组织蛋白酶基因在蠊蛄口外消化和肠道消化中发挥消化功能。

关键词: 蠊蛄; 组织蛋白酶; 消化; 基因鉴定; 表达特征

中图分类号: Q968.1; S433

文献标识码: A

文章编号: 1674-0858 (2024) 05-1161-10

Identification and expression profiling of cathepsin genes in *Arma custos*

LIU Sha, LIANG Wen-Kai, WANG Yu-Qin, LI Mei-Jiao, CHEN Fen-Lian, LUO Yun-Fei, ZHU Jia-Ying*
(Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, College of Biodiversity Conservation, Southwest Forestry University, Kunming 650224, China)

Abstract: Cathepsins play crucial roles in the digestion, development, and metamorphosis of insects. They are commonly found components in the saliva (venom) of predatory bugs, while their physiological functions remain unclear. In this study, homology-based methods were employed to identify cathepsin genes from the genome of the stink bug *Arma custos*. Bioinformatics software was used to analyze their sequence characteristics and evolutionary relationships. RT-PCR technology was utilized to analyze their expression patterns in various adult tissues. The results revealed the presence of 37 protease genes in the *A. custos* genome, categorized into three groups based on structural domains: Cathepsin B (*AcCABI-4*), cathepsin D (*AcCAD1-13*), and cathepsin L (*AcCALI-20*). Multiple sequence alignment and domain prediction results indicated that cathepsin B and L both contain the Peptidase_C1 structural domain, conserved enzyme catalytic sites (glutamine, cysteine, histidine, and asparagine), and a catalytic dyad composed of cysteine (C) and histidine (H) residues. In contrast, cathepsin D possesses a conserved Asp domain and two conserved enzyme catalytic sites (aspartic acid). RT-PCR results demonstrated that, apart from *AcCAL7* and *AcCAB2*, the remaining 33 protease genes were expressed in the gut, with the majority

基金项目: 云南省农业基础研究联合专项重点项目(202101BD070001-024); 国家自然科学基金(32360686); 国家林草科技创新计划青年拔尖人才项目(2019132615); 云南省教育厅科技创新团队建设项目

作者简介: 刘莎, 女, 1999年生, 云南曲靖人, 硕士研究生, 研究方向为昆虫生理生化与分子生物学, E-mail: 3109328917@qq.com

*通讯作者 Author for correspondence: 朱家颖, 男, 博士, 教授, 研究方向为昆虫生理生化与分子生物学, E-mail: jy Zhu@swfu.edu.cn

收稿日期 Received: 2023-08-30; 接受日期 Accepted: 2024-02-02

showing tissue-specific or high expression. Twenty cathepsin genes, including *AcCAL2*, *AcCAL5-7*, and *AcCAL9*, were expressed in the salivary glands, with *AcCAL2* exclusively expressed in the main posterior gland and the accessory gland, *AcCAL7* exclusively expressed in the main anterior gland, *AcCAD13* exclusively expressed in the main gland posterior lobe, and *AcCAB2* exclusively expressed in the main anterior and posterior glands. These findings suggest that multiple cathepsin genes play roles in external and gut digestion in *A. custos*.

Key words: *Arma custos*; cathepsin; digestion; gene identification; expression profile

组织蛋白酶 (Cathepsin) 最早于 20 世纪 40 年代在哺乳动物的脾脏和肾中被发现 (Gutmann *et al.*, 1948), 由于组织蛋白酶最开始是在动物的溶酶体中作为能降解蛋白质的酶类被报道, 因此最初将所有的细胞内肽酶都称作组织蛋白酶 (Barrett, 1992)。目前, 已报道的组织蛋白酶数量繁多, 可根据活性中心的氨基酸类型将它们分为半胱氨酸蛋白酶 (组织蛋白酶 B、L、H、O、S、T、K、V、F)、天冬氨酸蛋白酶 (组织蛋白酶 D、E) 以及丝氨酸蛋白酶 (组织蛋白酶 A、G) (Conus *et al.*, 2010)。其中, 组织蛋白酶 B、D 和 L 的数量最多, 研究也最为深入 (Kim *et al.*, 2020; 商晓康等, 2023; Pasandideh *et al.*, 2023; Wang *et al.*, 2023; Yuan *et al.*, 2023)。其中, 组织蛋白酶 B 和组织蛋白酶 L 在还原性和弱酸性 (pH5~6) 条件下更具酶活性, 而组织蛋白酶 D 在 pH 介于 3.5~5 之间更具酶活性 (Conus *et al.*, 2010; Schmitz *et al.*, 2019)。此外, 组织蛋白酶 B 的氨基酸序列中具有由 12 个保守半胱氨酸形成的 6 个二硫键 (Martynov *et al.*, 2015), 以及由 20 个氨基酸形成的封闭环, 该环用于阻止肽酶抑制剂 (如胱氨酸蛋白酶) 进入酶活性位点 (Schmitz *et al.*, 2019)。组织蛋白酶 D 的典型特征是氨基酸序列内具有 2 个催化天冬氨酸残基以及 1 个脯氨酸环, 缺失该环的组织蛋白酶 D 的作用类似于脊椎动物的胃蛋白酶, 在消化过程中起作用 (Padilha *et al.*, 2009)。组织蛋白酶 L 在昆虫中以非活性形式 (酶原) 存在, 在信号肽后含有约 60 个氨基酸的前体肽 (通常含有抑制性 ERFNIN 基序), 通过分子内加工或其它酶切割后去除前体肽, 进而形成具有活性的成熟酶 (Santos *et al.*, 2023; Wiederanders, 2003)。

在哺乳动物中, 组织蛋白酶主要存在于溶酶体中, 在微酸性环境中具有最佳酶活性 (Yamahama *et al.*, 2003)。因组织蛋白酶在人类疾病中具有重要作用而被广泛研究 (Caculitan *et al.*,

2017; Fujiwara *et al.*, 2022; Ren *et al.*, 2023), 除参与降解蛋白质外, 还在组织分化和免疫反应中发挥重要功能 (Guan *et al.*, 2022)。在昆虫中, 组织蛋白酶最早在家蝇 *Musca domestica* 中作为“胃蛋白酶”被研究报道 (Greenberg & Paretsky, 1955)。组织蛋白酶除了作为昆虫肠道中的重要酶类, 在食物消化作用中发挥重要作用外 (Martynov *et al.*, 2015; Terra *et al.*, 2019; Dvoryakova *et al.*, 2022), 研究还发现组织蛋白酶能参与卵黄细胞降解、脂肪体解离、胚胎发育、翅发育等生物学过程, 在昆虫生长发育和变态中具有重要作用 (Medina *et al.*, 1988; Carnevali *et al.*, 2006; Eykelbosh *et al.*, 2010; Wang *et al.*, 2010; 李懿等, 2015; Sun *et al.*, 2018; Ferrara *et al.*, 2020)。

捕食性蝽作为重要的天敌昆虫类群, 在捕食的过程中通过口针将唾液腺合成的毒液注入到猎物体内, 起到杀死猎物和消化猎物组织的作用 (Cohen, 1998)。目前, 有少数关于捕食性蝽毒液成分的研究表明, 捕食性蝽毒液中含有丰富的酶类, 主要为丝氨酸蛋白酶和包括组织蛋白酶在内的其它水解酶, 但至今仍不清楚这些酶的具体生理功能 (Walker *et al.*, 2018; Qu *et al.*, 2023; Wu *et al.*, 2023)。本文以能捕食鳞翅目、鞘翅目、膜翅目等农林害虫, 且已经作为商品化的天敌昆虫应用于田间防治草地贪夜蛾 *Spodoptera frugiperda*、斜纹夜蛾 *Spodoptera litura*、烟青虫 *Heliothis assulta* 等重要害虫的蝽蝽 *Arma custos* (Hemiptera: Pentatomidae) 为研究对象 (邹德玉等, 2016; 马润国等, 2020; 唐艺婷等, 2020), 从该蝽基因组中鉴定出组织蛋白酶基因, 分析了不同类型组织蛋白酶基因的序列结构特征以及进化关系, 并通过 RT-PCR 分析了这些基因在成虫不同组织中的表达特征, 旨在鉴定出在唾液腺特异性或高表达的组织蛋白酶基因, 为今后研究其在蝽蝽捕食中的作用奠定基础。

1 材料与方 法

1.1 供试昆虫

供试蠋蝽为实验室饲养多代的种群, 使用黄粉虫高龄幼虫作为猎物, 饲养于 40 cm×30 cm×40 cm 的养虫笼中, 虫口密度保持在约 50 头/笼。饲养条件为温度 25℃±1℃, 光周期 16 L : 8 D, 相对湿度 65%±5%。

1.2 基因鉴定及序列分析

使用 Silva *et al.* (2022) 从多种昆虫中鉴定出的组织蛋白酶序列作为种子序列, 基于同源比对方法, 以 E-value<1×10⁻⁵ 作为阈值, 将引用的组织蛋白酶基因序列利用 TBtools 软件 (Chen *et al.*, 2020a) 中的 Blast 工具包搜索蠋蝽基因组数据。初步从蠋蝽基因组中鉴定出的潜在组织蛋白酶基因, 通过使用 NCBI 中的 BLASTp 搜索 Nr 库验证, 获得真正的组织蛋白酶基因。

鉴定出的蠋蝽组织蛋白酶基因编码氨基酸序列的等电点以及分子量使用 ExPasy (<https://web.expasy.org/protparam/>) 预测, 信号肽使用 SignalP-6.0 (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>) 进行预测, 结构域使用 SMART (<https://smart.embl.de/>) 预测。多序列比对采用 ClustalX 1.83 软件完成 (Chenna *et al.*, 2003), 比对结果使用 GeneDoc 软件进行着色。使用 TBtools 中的 IQtree2 插件, 最大似然法 (Maximum likelihood, ML) 构建系统发育树 (Chen *et al.*, 2020a)。系统发育树使用 FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software>

/Figtree/) 进行美化。

1.3 RT-PCR

取蠋蝽成虫, 于体视镜下在磷酸盐缓冲液中解剖获得成虫 (雌雄 1 : 1) 唾液腺不同组织 (主腺前叶、主腺后叶和副腺) 以及肠道、脂肪体、残体 (去除唾液腺、肠道和脂肪体后的虫体)。各组织样品收集到 1.5 mL 的离心管中, 液氮研磨后, 加入 1 mL Trizol 试剂 (Invitrogen), 保存在 -80℃ 冰箱备用。按照 Trizol 试剂 (Invitrogen) 说明书, 提取样品的总 RNA。使用 NanoDrop 2000 分光光度计 (Thermo Fisher Scientific), 于 260/280 nm 处检测 RNA 纯度, 并使用 1% 琼脂糖凝胶电泳检测其完整性。以提取的总 RNA 为模板, 参照 PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa) 试剂盒说明书合成 cDNA 模板。根据鉴定到的蠋蝽组织蛋白酶基因序列, 使用 Primer Premier 5.0 (Primer Biosoft International) 软件设计引物 (表 1)。基于茶翅蠋 *Halyomorpha halys* 的 40S 核糖体蛋白 S18 (40S ribosomal protein S18, 40S RP-S18) 基因, 通过同源比对从蠋蝽基因组中鉴定得到其 40S RP-S18 基因, 并作为内参用于 RT-PCR。RT-PCR 试验反应总体积为 25 μL, 其中 Dream Taq Green PCR Master Mix 12.5 μL, 上下游引物各 0.5 μL, cDNA 模板 1 μL, RNase-free H₂O 10.5 μL。PCR 反应条件为 95℃ 预变性 3 min, 95℃ 变性 30 s, 55℃ 退火 30 s, 72℃ 延伸 1 min, 35 个循环, 72℃ 延伸 10 min。PCR 产物经 1.2% 琼脂糖凝胶电泳检测并拍照分析。

表 1 本研究所用引物

Table 1 Primers used in this study

基因名称 Gene name	正向引物 (5'-3') Sense primer (5'-3')	反向引物 (5'-3') Anti-sense primer (5'-3')
<i>AcCAB1</i>	CACCAACTCCAACATCCCAG	CATTGCAGCCATCTCCACAG
<i>AcCAB2</i>	CATCCTACAATCTGAAATCCCG	CGCTAATAGCCAATAGGGAGTG
<i>AcCAB3</i>	ACTGACGGATGGCTCACTTC	CGCTAATAGCCAATAGGGAG
<i>AcCAB4</i>	GGCCGATACCAGATTCTTTC	CCTTCAACTGGTCCGTTATTC
<i>AcCAD1</i>	ACGAAGGGATCATCGGATTG	CGGCGAACCAAAGAAAGC
<i>AcCAD2</i>	GGAGGTTCTTGCTTCTCG	CAATATCCTCCGTCGAATGG
<i>AcCAD3</i>	TTGCTGAGGCGGTAAATG	TGATGGACCAGCAATGAGAC
<i>AcCAD4</i>	AGCCACCATCTCTTGACAGC	CCTAGGATCCATGGCTTGTC
<i>AcCAD5</i>	TAAGCCTGTCTTGTTCGGTG	AGACCAAACCTTCATTGAGGC

续表 1 Continued table 1

基因名称 Gene name	正向引物 (5'-3') Sense primer (5'-3')	反向引物 (5'-3') Anti-sense primer (5'-3')
<i>AcCAD6</i>	TCGCTCAAGCAGTCTTTTCC	GGTATTCTTTGCGGTTGATTG
<i>AcCAD7</i>	ATGGTCATCGTGGTGGTGC	GCGTTGATACCGTTCGTGTAG
<i>AcCAD8</i>	TCCGAGTCATTAGGCTGTGC	CGTTGGATTCTTCCTCTTCG
<i>AcCAD9</i>	TTACGCCATCGGCTACATG	CCGTCTTCGAGTTCTTCTGC
<i>AcCAD10</i>	TCACTGGACCAATCAACGAG	GGAACATCGTACCGAGAACC
<i>AcCAD11</i>	TGTGGAAGTGGAGGAAGTCA	TCAAGGAAGTGCCAGTATCG
<i>AcCAD12</i>	ACTTGCCACAAGGTACATCAG	CGTATCAACAAAACCAACTGC
<i>AcCAD13</i>	GTAACCCCCACAAACATTC	GACCCTGACATGAGTAGCACA
<i>AcCAL1</i>	TTCCTTCCTCCAGCAAACG	GGACGGTTCACAGGAATCAC
<i>AcCAL2</i>	GATGTTTACGTTGCACCAGC	ACACCATGATCGAGCTCATC
<i>AcCAL3</i>	AAACCTAGTGTCCCTCAGCG	AGCTCCGGAAGGTGTAGTTG
<i>AcCAL4</i>	GCTCCCAGAAGAAGTGACCC	TGCCACAGGCGTTGTTCTC
<i>AcCAL5</i>	ACGGTTGTGAAGGAGGTTTG	CACGAGTTCTTGACGAGCC
<i>AcCAL6</i>	TGTAGCCATTGACGCATCTC	AGCTTGGGTTGCTATTCCAC
<i>AcCAL7</i>	TCGATTGCTCATTGAACTACG	CCAGTATTTCTGGCCTTTCTC
<i>AcCAL8</i>	CAAGGAGTTGAAGTGGTGATG	TGAACAATCAACCAGGTTCTG
<i>AcCAL9</i>	GAAGGAACTGGCTGAAATGG	CAACCCGAACAGAGCAAATC
<i>AcCAL10</i>	GCATCGGCTGAGGTTTACG	TGGGATAGGCAATGGGTAGC
<i>AcCAL11</i>	AATACAGAAGTGTGCTGACG	GTTCCGTATCCAACAACCAG
<i>AcCAL12</i>	GAGCGAACAGAACTTGATGG	AGTTCATCAATAGCTGGGGC
<i>AcCAL13</i>	TGTTCTTGGAATACCGCTG	GGCGCATGATTACAGATTAG
<i>AcCAL14</i>	CCTGAAGAGGTCTCCATCCC	TTCTCCCCAGTTAAGTCCCC
<i>AcCAL15</i>	TCCAGGATCAAGGAGATTGC	ATCTGCATGGTCCATCCTTC
<i>AcCAL16</i>	TTGGTAGACTGTTCCGAGGC	TTCCTGTCCGTCCTCAACTC
<i>AcCAL17</i>	CTGTGGTGGTTTTCTGGTC	CCATCCTGTCCAAGGTATCC
<i>AcCAL18</i>	CTGTTGGGCTTTTAGTGCG	TTCTGCCCCAAAGTCTTCTCC
<i>AcCAL19</i>	TGATTCTGTTTGCCTGTTG	GCGTTGTTTCATATCCCTCC
<i>AcCAL20</i>	CTCCGGTCAAGGAGCAAAAG	GGGTTGTCCATTCTCCTTGC
<i>40S RP S18</i>	GGAAATTCCTGGCCTTACTGAC	TTCTTGCTCCTAGCCTTAGCT

2 结果与分析

2.1 蠋蝽组织蛋白酶基因鉴定及序列分析

通过使用其它昆虫的组织蛋白酶基因序列进行同源比对, 从蠋蝽基因组中鉴定出 37 个组织蛋

白酶基因, 包含 20 个组织蛋白酶 L (*AcCAL1-20*)、13 个组织蛋白酶 D (*AcCAD1-13*) 和 4 个组织蛋白酶 B (*AcCAB1-4*) 基因 (表 2)。这些组织蛋白酶基因编码蛋白的理论分子量的范围为 30.5~94.4 kDa, 等电点的范围为 5.03~9.1。信号肽预测结果表明, 有 23 个蠋蝽组织蛋白酶基因编码的

氨基酸序列中含有信号肽序列(表 2)。SMART 分析结果表明, 蠋蜻组织蛋白酶 B 和组织蛋白酶 L 同属半胱氨酸家族, 均拥有 1 个 Peptidase_C1 结构域, 区别在于组织蛋白酶 L 的保守结构域为

Peptidase_C1 或 Inhibitor_I29 和 Peptidase_C1, 而组织蛋白酶 B 的保守结构域为 Propeptide_C1 和 Peptidase_C1。组织蛋白酶 D 属于天冬氨酸蛋白酶, 其保守结构域为 Asp(表 2)。

表 2 蠋蜻基因组中鉴定出的组织蛋白酶基因及其序列结构特征

Table 2 Cathepsin genes identified in the genome of *Arma custos* and their sequence characteristics

蛋白酶类型 Proteinase type	基因名称 Gene name	结构域 Domain	信号肽 Signal peptide	分子量 (kDa) Molecular weight	等电点 pI
Cathepsin B	<i>AcCAB1</i>	Propeptide_C1、Peptidase_C1	是 Yes	37.0	5.78
	<i>AcCAB2</i>	Propeptide_C1、Peptidase_C1	是 Yes	37.9	4.76
	<i>AcCAB3</i>	Propeptide_C1、Peptidase_C1	是 Yes	37.6	4.99
	<i>AcCAB4</i>	Propeptide_C1、Peptidase_C1	是 Yes	37.0	7.53
Cathepsin D	<i>AcCAD1</i>	Asp	否 No	37.9	8.23
	<i>AcCAD2</i>	Asp	是 Yes	44.0	9.10
	<i>AcCAD3</i>	Asp	是 Yes	43.4	6.20
	<i>AcCAD4</i>	Asp	否 No	36.2	5.82
	<i>AcCAD5</i>	Asp	否 No	82.9	5.42
	<i>AcCAD6</i>	Asp	是 Yes	47.8	6.04
	<i>AcCAD7</i>	Asp	是 Yes	43.3	6.98
	<i>AcCAD8</i>	Asp	否 No	38.8	7.62
	<i>AcCAD9</i>	Asp	是 Yes	42.3	7.51
	<i>AcCAD10</i>	Asp	是 Yes	40.9	6.89
	<i>AcCAD11</i>	Asp	否 No	47.6	6.12
	<i>AcCAD12</i>	Asp	是 Yes	71.8	5.19
	<i>AcCAD13</i>	Asp	是 Yes	43.0	6.79
Cathepsin L	<i>AcCAL1</i>	Inhibitor_I29、Peptidase_C1	否 No	44.9	7.08
	<i>AcCAL2</i>	Inhibitor_I29、Peptidase_C1	是 Yes	36.9	5.13
	<i>AcCAL3</i>	Inhibitor_I29、Peptidase_C1	否 No	34.4	5.59
	<i>AcCAL4</i>	Inhibitor_I29、Peptidase_C1	是 Yes	36.1	5.49
	<i>AcCAL5</i>	Inhibitor_I29、Peptidase_C1	是 Yes	37.6	5.73
	<i>AcCAL6</i>	Peptidase_C1	是 Yes	36.1	5.51
	<i>AcCAL7</i>	Peptidase_C1	否 No	30.5	5.16
	<i>AcCAL8</i>	Inhibitor_I29、Peptidase_C1	否 No	56.3	5.14
	<i>AcCAL9</i>	Inhibitor_I29、Peptidase_C1	否 No	94.4	7.98
	<i>AcCAL10</i>	Inhibitor_I29、Peptidase_C1	是 Yes	74.9	5.41
	<i>AcCAL11</i>	Inhibitor_I29、Peptidase_C1	是 Yes	36.5	9.10
	<i>AcCAL12</i>	Inhibitor_I29、Peptidase_C1	是 Yes	67.3	5.97
	<i>AcCAL13</i>	Inhibitor_I29、Peptidase_C1	否 No	35.9	5.41
	<i>AcCAL14</i>	Inhibitor_I29、Peptidase_C1	是 Yes	37.1	6.26

续表 2 Continued table 2

蛋白酶类型 Proteinase type	基因名称 Gene name	结构域 Domain	信号肽 Signal peptide	分子量 (kDa) Molecular weight	等电点 pI
Cathepsin L	<i>AcCAL15</i>	Inhibitor_I29、Peptidase_C1	是 Yes	37.1	8.31
	<i>AcCAL16</i>	Inhibitor_I29、Peptidase_C1	是 Yes	36.7	5.03
	<i>AcCAL17</i>	Inhibitor_I29、Peptidase_C1	否 No	54.8	7.24
	<i>AcCAL18</i>	Inhibitor_I29、Peptidase_C1	否 No	39.9	6.03
	<i>AcCAL19</i>	Peptidase_C1	否 No	34.7	8.94
	<i>AcCAL20</i>	Inhibitor_I29、Peptidase_C1	是 Yes	38.1	5.14

将蠋蝽 B、D 和 L 与黑腹果蝇 *Drosophila melanogaster* 组织蛋白酶 B、D 和 L 的氨基酸序列分别进行多序列比对 (图 1~3), 结果显示组织蛋白酶 B 和 L 都有 4 个高度保守的酶催化位点 (谷氨酰胺、半胱氨酸、组氨酸和天冬酰胺), 组织蛋

白酶 D 中有 2 个高度保守的酶催化位点 (天冬氨酸)。在组织蛋白酶 B 和组织蛋白酶 L 中, 1 个半胱氨酸残基 (C) 和 1 个组氨酸残基 (H) 可构成催化二联体。

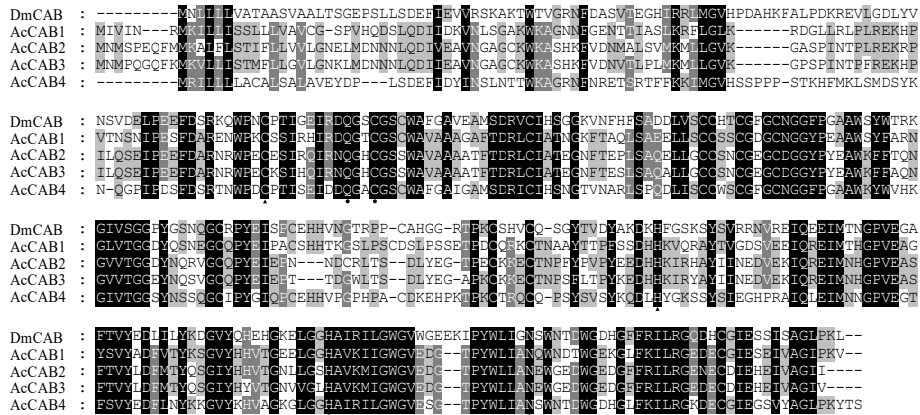


图 1 蠋蝽和黑腹果蝇组织蛋白酶 B 氨基酸多序列比对

Fig. 1 Multiple alignment of the amino acid of cathepsin B of *Arma custos* and *Drosophila melanogaster*

注: 深色阴影标记的是保守位点, 黑点表示高度保守的酶催化位点, 三角形表示催化二联体。Ac, 蠋蝽; Dm, 黑腹果蝇。图 2~3 同。Note: Dark shading indicated the conserved sites. Black dots represented the active-site of cathepsin. Tringles represented the catalytic diad. AC, *Arma custos*; Dm, *Drosophila melanogaster*. Same to Fig. 2~3.

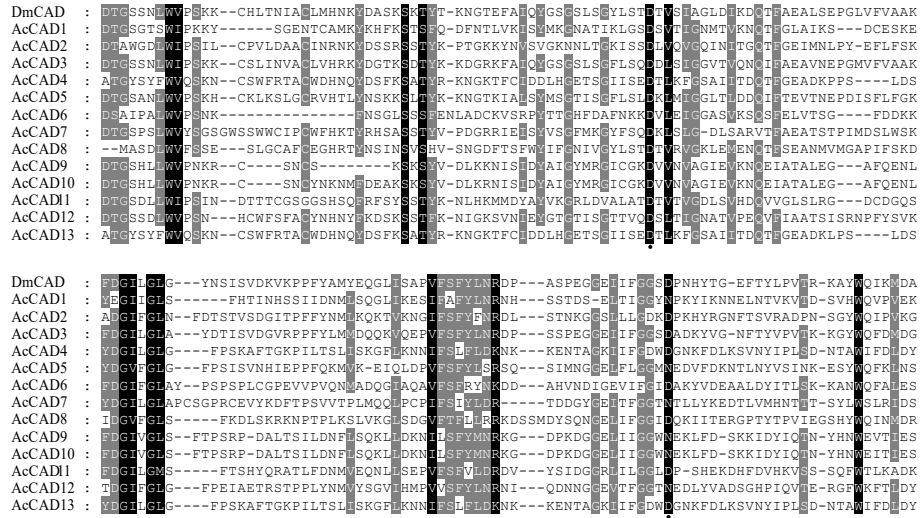


图 2 蠋蝽和黑腹果蝇组织蛋白酶 D 多序列比对

Fig. 2 Multiple alignment of the amino acid of cathepsin D of *Arma custos* and *Drosophila melanogaster*

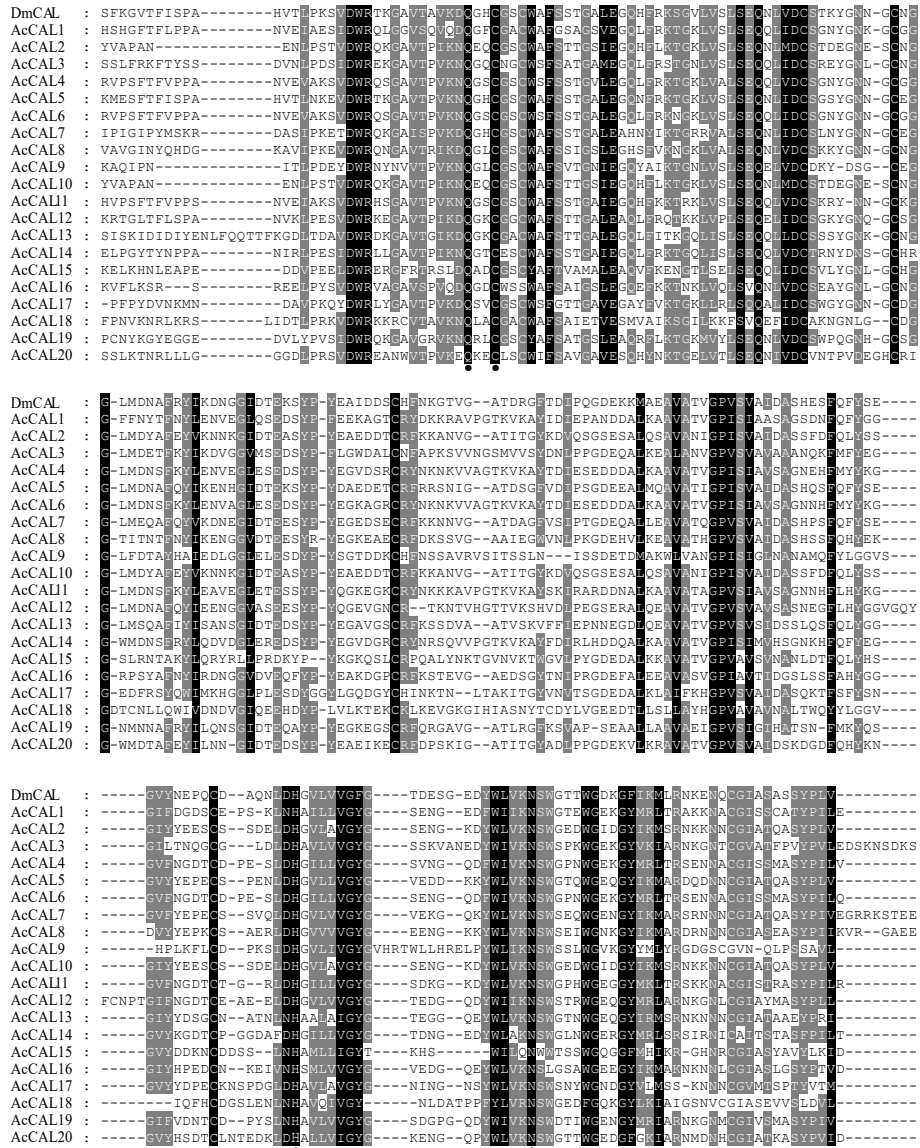


图 3 蠋蝽和黑腹果蝇组织蛋白酶 L 多序列比对

Fig. 3 Multiple alignment of the amino acid of cathepsin L of *Arma custos* and *Drosophila melanogaster*

系统发育树结果显示 (图 4), 来自蠋蝽和其它昆虫的组织蛋白酶 B、D 和 L 各自聚为独立的分支。就蠋蝽鉴定得到的 4 个组织蛋白酶 B 基因而言, 其中 AcCAB4 与家蚕 *Bombyx mori* 组织蛋白酶 B (BmCAB)、黑腹果蝇组织蛋白酶 (DmCAB) 及黄粉虫 *Tenebrio molitor* 组织蛋白酶 B (TmB17) 聚在一起, 为溶酶体组织蛋白酶。

2.2 蠋蝽组织蛋白酶基因表达特征

RT-PCR 结果显示 (图 5), 除 *AcCAL8* 和 *AcCAB3* 外, 蠋蝽其它 35 个组织蛋白酶基因在成虫组织中均检测到表达。在这些基因中, 仅 *AcCAL7* 和 *AcCAB2* 未检测到在肠道中表达外, 其余 33 个组织蛋白酶基因均在肠道中表达, 且绝大多数表现特异性或高表达。*AcCAL1~3*、*AcCAL5~7*、

AcCAL9~11、*AcCAL13~20* 等 29 个基因在脂肪体中有表达, 其中有 27 个基因在脂肪体和肠道中都有表达。其次, *AcCAL4*、*AcCAL12*、*AcCAD1* 和 *AcCAD4* 仅在肠道中检测到表达, *AcCAL1*、*AcCAL3*、*AcCAL4*、*AcCAL10*、*AcCAL14~16*、*AcCAL2*、*AcCAL5~7*、*AcCAL9*、*AcCAL11*、*AcCAL17~19* 等 20 个基因在唾液腺中表达。在这些基因中, *AcCAL2*、*AcCAL6*、*AcCAB1* 仅在主腺后叶和副腺表达, *AcCAL7* 仅在主腺前叶表达, *AcCAD13* 仅在主腺后叶表达, *AcCAB2* 仅在主腺前叶和主腺后叶中表达, 其余的 14 个组织蛋白酶基因在唾液腺的主腺前叶、主腺后叶和副腺中均有表达。

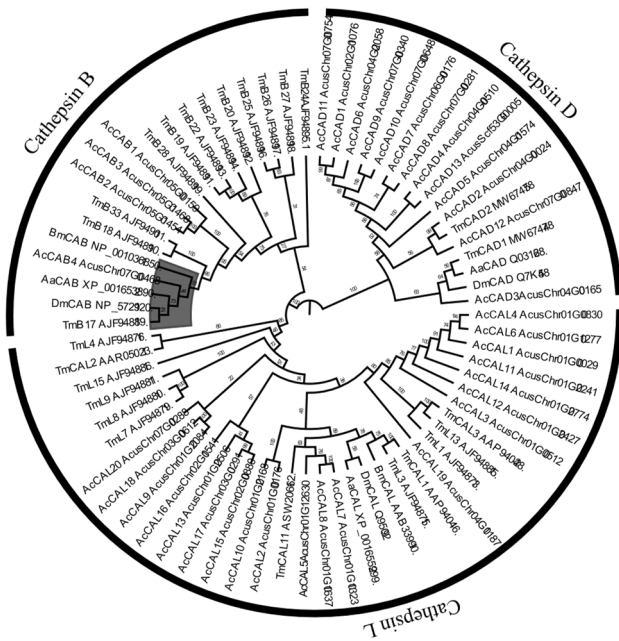


图 4 蠨螋与其它昆虫组织蛋白酶 B、D 和 L 系统发育分析
Fig. 4 Phylogenetic analysis of cathepsin B, D and L from *Arma custos* and other insects

注：Ac，蠨螋；Bm，家蚕；Tm，黄粉虫；Dm，黑腹果蝇。阴影覆盖区为溶酶体组织蛋白酶。Note: Ac, *Arma custos*; Bm, *Bombyx mori*; Tm, *Tenebrio molitor*; Dm, *Drosophila melanogaster*. Lysosomal cathepsin B were shaded.

3 结论与讨论

本研究通过同源比对的方法从蠨螋基因组中鉴定出 37 个组织蛋白酶基因，数量上明显多于现有报道的大多数鞘翅目基因组中的组织蛋白酶基因数量，如赤拟谷盗 *Tribolium castaneum* 中含有 24 个组织蛋白酶基因 (Silva *et al.*, 2022)。就各亚家族的数量而言，蠨螋基因组中的组织蛋白酶 L 的数量最多 (20 个)，组织蛋白酶 D 的数量其次 (13 个)，组织蛋白酶 B 的数量较少 (4 个)，而在一些鞘翅目昆虫 (如黄粉虫) 的组织蛋白酶 L 和组织蛋白酶 B 的数量均多于组织蛋白酶 D 的数量 (Silva *et al.*, 2022)。研究表明，组织蛋白酶的氨基酸序列中存在一个或多个高度保守的酶活性位点残基，如半胱氨酸、组氨酸、天冬酰胺、天冬氨酸，这些位点是确保它们具有酶活性的关键所在 (Rawlings *et al.*, 2008)。多序列比对结果显示，蠨螋组织蛋白酶 B 和 L 中，半胱氨酸残基 (C) 和组氨酸残基 (H) 可构成酶的催化二联体，与谷氨酰胺残基 (Q) 和天冬酰胺残基 (N) 一

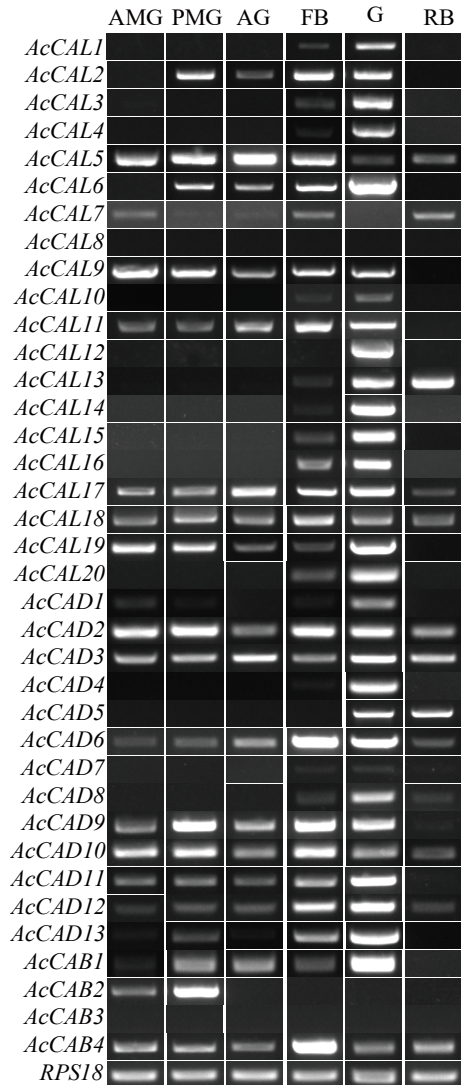


图 5 蠨螋组织蛋白酶基因在成虫不同组织中的表达特征
Fig. 5 Expression patterns of cathepsin genes in different tissues of the adult of *Arma custos*

注：AMG，主腺前叶；PMG，主腺后叶；AG，副腺；FB，脂肪体；GUT，肠道；RB，残体。Note: AMG, Anterior main gland; PMG, Posterior main gland; AG, Accessory gland; FB, Fat body; G, Gut; RB, Residual body.

起在酶的催化过程中起重要作用 (Lewis *et al.*, 1981)。组织蛋白酶 D 中含有 2 个天冬氨酸残基 (D)，是酶活性所必须的保守催化位点。系统发育树结果表明，蠨螋 AcCAB4 与家蚕、黑腹果蝇及黄粉虫的溶酶体蛋白酶聚在一起 (Cristofoletti *et al.*, 2005)，表明它应为溶酶体组织蛋白酶，具有消化功能。

蠨螋组织蛋白酶基因在成虫不同组织中的表达情况结果表明，它们在不同组织中呈现不同程度的表达，且多数基因主要在肠道和唾液腺中特

异性表达或高表达，这与李懿等（2015）对家蚕以及 Dvoryakova *et al.*（2022）对赤拟谷盗组织蛋白酶基因不同组织表达特征结果相似。绝大多数组织蛋白酶基因在蠋蝽肠道中特异性或高表达，表明它们具有消化功能。同时，*AcCAL4*、*AcCAL12*、*AcCAD1* 和 *AcCAD4* 仅在中肠中表达，因中肠是昆虫消化食物的重要场所，表明它们在对摄入的食物进行消化以及水解的过程中发挥重要作用（Dvoryakova *et al.*, 2022）。蠋蝽捕食时，通过将口针插入猎物体内，进而将唾液腺分泌的毒液注射到猎物中麻痹猎物并对猎物组织进行口外消化（Cohen, 1998）。组织蛋白酶是捕食性蝽毒液中的一种主要水解酶，可能起到分解猎物组织中复杂大分子的作用（Ghamari *et al.*, 2014）。蠋蝽中有 20 个组织蛋白酶基因在唾液腺中特异性或高表达，据此认为它们作为毒液成分在体外消化中发挥作用。而且，表达特征分析发现，*AcCAL2*、*AcCAL6*、*AcCAB1* 仅在主腺后叶和副腺表达，*AcCAL7* 仅在主腺前叶表达，*AcCAD13* 仅在主腺后叶表达，*AcCAB2* 仅在主腺前叶和主腺后叶表达。有研究表明，捕食性蝽主腺前叶、主腺后叶和副腺的毒液对猎物组织的裂解功能存在差异（Walker *et al.*, 2018; Fischer *et al.*, 2020; Wu *et al.*, 2023）。据此，在蠋蝽主腺前叶、主腺后叶和副腺中差异表达的组织蛋白酶基因，对猎物组织的裂解功能也应存在差异，各自具体生理功能有待进一步研究揭示。

参考文献 (References)

- Barret AJ. Cellular proteolysis an overview [J]. *Annals of the New York Academy of Sciences*, 1992, 674: 1-15.
- Bernard G, Paretsky D. Proteolytic enzymes in the House Fly, *Musca domestica* (L.) [J]. *Annals of the Entomological Society of America*, 1955, 48 (1-2): 46-50.
- Carnevali O, Cionna C, Tosti L, *et al.* Role of cathepsins in ovarian follicle growth and maturation [J]. *General and Comparative Endocrinology*, 2006, 146 (3): 195-203.
- Chen CJ, Chen H, Zhang Y, *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, 13 (8): 1194-1202.
- Chenna R, Sugawara H, Koike T, *et al.* Multiple sequence alignment with the Clustal series of programs [J]. *Nucleic Acids Research*, 2003, 31 (13): 3497-3500.
- Cohen AC. Solid-to-Liquid feeding: The inside (s)story of extra-oral digestion in predaceous Arthropoda [J]. *American Entomologist*, 1998, 44 (2): 103-117.
- Conus S, Simon HU. Cathepsins and their involvement in immune responses [J]. *Swiss Medical Weekly*, 2010, 140: w13042.
- Cristofolletti PT, Ribeiro AF, Terra WR. The cathepsin L-like proteinases from the midgut of *Tenebrio molitor* larvae: Sequence, properties, immunocytochemical localization and function [J]. *Insect Biochemistry and Molecular Biology*, 2005, 35 (8): 883-901.
- Dvoryakova EA, Vinokurov KS, Tereshchenkova VF, *et al.* Primary digestive cathepsins L of *Tribolium castaneum* larvae: Proteomic identification, properties, comparison with human lysosomal cathepsin L [J]. *Insect Biochemistry and Molecular Biology*, 2022, 140: 103679.
- Eykelbosh AJ, Van Der Kraak G. A role for the lysosomal protease cathepsin B in zebrafish follicular apoptosis [J]. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 2010, 156 (2): 218-223.
- Ferrara TFS, Schneider VK, Lima PS, *et al.* Gene expression studies and molecular characterization of a cathepsin L-like from the Asian citrus psyllid *Diaphorina citri*, vector of Huanglongbing [J]. *International Journal of Biological Macromolecules*, 2020, 158: 375-383.
- Fischer ML, Wielsch N, Heckel DG, *et al.* Context-dependent venom deployment and protein composition in two assassin bugs [J]. *Ecology and Evolution*, 2020, 17, 10 (18): 9932-9947.
- Fujiwara R, Ten H, Chen H, *et al.* Cathepsin D inhibits angiogenesis in pituitary neuroendocrine tumors [J]. *Acta Histochemica et Cytochemica*, 2022, 55 (6): 203-211.
- Guan Y, Yang X, Zhao R, *et al.* Characteristics of cathepsin members and expression responses to poly I: C challenge in Pacific cod (*Gadus macrocephalus*) [J]. *Fish & Shellfish Immunology*, 2022, 128: 484-493.
- Gutmann HR, Fruton JS. On the proteolytic enzymes of animal tissues; an intracellular enzyme related to chymotrypsin [J]. *Journal of Biological Chemistry*, 1948, 174 (3): 851-858.
- Kim JW, Cho JY, Kim J, *et al.* First report of cathepsin E in a teleost (Korean rose bitterling, *Rhodeus uyekii*): Molecular characterization and tissue distribution [J]. *Developmental & Comparative Immunology*, 2020, 106: 103607.
- Lewis SD, Johnson FA, Shafer JA. Effect of cysteine-25 on the ionization of histidine-159 in papain as determined by proton nuclear magnetic resonance spectroscopy. Evidence for a his-159-Cys-25 ion pair and its possible role in catalysis [J]. *Biochemistry*, 1981, 20 (1): 48-51.
- Li Y, Zhou XY, Li ZL, *et al.* Identification and expression pattern of cathepsin family in silkworm (*Bombyx mori*) [J]. *Chinese Journal of Biotechnology*, 2015, 31 (12): 1728-1740. [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 [J]. 生物工程学报, 2015, 31 (12): 1728-1740]
- Ma RG, Qin SJ, Zhang XP, *et al.* Functional response of *Arma chinensis* fallow predation to *Anomoneura mori* Schwarz [J]. *Science of Sericulture*, 2022, 48 (6): 559-564. [马润国, 秦世杰, 张新鹏等. 蠋蝽对桑木虱的捕食功能反应 [J]. 蚕业科学, 2022, 48 (6): 559-564]
- Martynov AG, Elpidina EN, Perkin L, *et al.* Functional analysis of C1 family cysteine peptidases in the larval gut of *Tenebrio molitor* and *Tribolium castaneum* [J]. *BMC Genomics*, 2015, 16 (1): 75.
- Medina M, Leon P, Vallejo CG. *Drosophila* cathepsin B-like proteinase: A suggested role in yolk degradation [J]. *Archives of Biochemistry and Biophysics*, 1988, 263 (2): 355-363.
- Padilha MH, Pimentel AC, Ribeiro AF, *et al.* Sequence and function of lysosomal and digestive cathepsin D-like proteinases of *Musca domestica* midgut [J]. *Insect Biochemistry and Molecular Biology*, 2009, 39 (11): 782-791.
- Pasandideh M, Harkinezhad T, Mohammadi L. A SNP in the ovine cathepsin K (*CTSK*) gene is associated with yearling growth performance in a crossbred sheep population [J]. *Animal Biotechnology*, 2023, 8: 1-5.
- Qu Y, Walker AA, Meng L, *et al.* The predatory stink bug *Arma custos* (Hemiptera: Pentatomidae) produces a complex proteinaceous venom to overcome caterpillar prey [J]. *Biology*, 2023, 12 (5): 691.
- Rawlings ND, Morton FR. The MEROPS batch BLAST: A tool to detect peptidases and their non-peptidase homologues in a genome [J]. *Biochimie*, 2008, 90 (2): 243-259.
- Ren CP, Zhou QW, Li SS, *et al.* Characterization, potential prognostic value, and immune heterogeneity of cathepsin C in diffuse glioma [J]. *Current Bioinformatics*, 2023, 18 (1): 76-91.

- Santos Correa KC, Moreira AC, Abd El-Raheem Ibrahim AG, *et al.* Identification and characterization of a recombinant cysteine peptidase (*AsCathL*) from leaf-cutting ant *Atta sexdens* Linnaeus, 1758 (Hymenoptera, Formicidae) [J]. *Protein Expression and Purification*, 2023, 201: 106174.
- Schmitz J, Gilberg E, Löser R, *et al.* Cathepsin B: Active site mapping with peptidic substrates and inhibitors [J]. *Bioorganic & Medicinal Chemistry*, 2019, 27 (1): 1-15.
- Shang XK, Zhang SM, Ni JJ. Research progress of cathepsin B in brain aging and Alzheimer's diseases [J]. *Hereditas*, 2023, 45 (3): 212-220. [商晓康, 张思萌, 倪建军. 组织蛋白酶 B 参与脑衰老及阿尔兹海默症发生发展研究进展 [J]. *遗传*, 2023, 45 (3): 212-220]
- Silva CP, Dias RO, Bernardes V, *et al.* Recruitment of lysosomal cathepsins B, L and D as digestive enzymes in Coleoptera [J]. *Insect Molecular Biology*, 2022, 31 (2): 225-240.
- Sun YX, Tang L, Wang P, *et al.* Cathepsin L-like protease can regulate the process of metamorphosis and fat body dissociation in *Antheraea pernyi* [J]. *Developmental & Comparative Immunology*, 2018, 78: 114-123.
- Tang, YT, Wang, MQ, Li, YY, *et al.* Predation of *Arma chinensis* on *Spodoptera litura* larvae [J]. *Chinese Tobacco Science*, 2020, 41 (1): 62-66. [唐艺婷, 王孟卿, 李玉艳等. 蠊蝽对斜纹夜蛾幼虫的捕食作用 [J]. *中国烟草科学*, 2020, 41 (1): 62-66]
- Terra WR, Dias RO, Ferreira C. Recruited lysosomal enzymes as major digestive enzymes in insects [J]. *Biochemical Society Transactions*, 2019, 47 (2): 615-623.
- Walker AA, Mayhew ML, Jin J, *et al.* The assassin bug *Pristhesancus plagipennis* produces two distinct venoms in separate gland lumens [J]. *Nature Communications*, 2018, 22, 9 (1): 755.
- Wang LF, Chai LQ, He HJ, *et al.* A cathepsin L-like proteinase is involved in moulting and metamorphosis in *Helicoverpa armigera* [J]. *Insect Molecular Biology*, 2010, 19 (1): 99-111.
- Wang YW, Xu H, Sun BG. Cathepsin H and cathepsin B of *Cynoglossus semilaevis* are involved in anti-bacterial immunity against *Edwardsiella tarda* [J]. *Fish & Shellfish Immunology*, 2023, 134: 108594.
- Wiederanders B, Kaulmann G, Schilling K. Functions of propeptide parts in cysteine proteases [J]. *Current Protein & Peptide Science*, 2003, 4 (5): 309-326.
- Wu CY, Li L, Wang YQ, *et al.* Morphological, functional, compositional and transcriptional constraints shape the distinct venom profiles of the assassin bug *Sycanus croceovittatus* [J]. *International Journal of Biological Macromolecules*, 2023, 250: 126162.
- Yamahama Y, Uto N, Tamotsu S, *et al.* In vivo activation of pro-form *Bombyx* cysteine protease (BCP) in silk moth eggs: Localization of yolk proteins and BCP, and acidification of yolk granules [J]. *Journal of Insect Physiology*, 2003, 49: 131-140.
- Yuan J, Gao YS, Liu DL, *et al.* PINK1-mediated mitophagy contributes to glucocorticoid-induced cathepsin K production in osteocytes [J]. *Journal of Orthopaedic Translation*, 2023, 38: 229-240.
- Zou DY, Xu WH, Liu BM, *et al.* Research progress and prospects of *Arma chinensis* Fallou (Hemiptera: Pentatomidae) [J]. *Journal of Environmental Entomology*, 2016, 38 (4): 857-865. [邹德玉, 徐维红, 刘佰明等. 天敌昆虫蠊蝽的研究进展与展望 [J]. *环境昆虫学报*, 2016, 38 (4): 857-865]