环境昆虫学报 2024, 46 (5) : 1161 - 1170



Journal of Environmental Entomology

刘莎,梁文凯,王玉琴,李美娇,陈芬莲,罗云菲,朱家颖.蠋蝽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.

莎,梁文凯,王玉琴,李美娇,陈芬莲,罗云菲,朱家颖* 刘 (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蠋蝽 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,蠋蝽 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(AcCAB1-4)、组织蛋白酶 D(AcCAD1-13)和 组织蛋白酶 L (AcCAL1-20)。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蠋蝽口外消化和肠道消化中发挥消化功能。 关键词: 蠋蝽; 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 (AcCAB1-4), cathepsin D (AcCAD1-13), and cathepsin L (AcCAL1-20). 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: jyzhu@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 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 等重要害虫的蠋蝽 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 后研究其在蠋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.ex pasy.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://sm art.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1mL 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 (Primier Biosoft International) 软件设 计引物(表1)。基于茶翅蝽 Halyomorpha halvs 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 试验反应总体系为 25 uL, 其中 Dream Tag 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
Tabla 1	Drimors used in this study

		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')
AcCAB1	CACCAACTCCAACATCCCAG	CATTGCAGCCATCTCCACAG
AcCAB2	CATCCTACAATCTGAAATCCCG	CGCTAATAGCCAATAGGGAGTG
AcCAB3	ACTGACGGATGGCTCACTTC	CGCTAATAGCCAATAGGGAG
AcCAB4	GGCCGATACCAGATTCTTTC	CCTTCAACTGGTCCGTTATTC
AcCAD1	ACGAAGGGATCATCGGATTG	CGGCGAACCAAAGAAAGC
AcCAD2	GGAGGTTCCTTGCTTCTCG	CAATATCCTCCGTCGAATGG
AcCAD3	TTGCTGAGGCGGTAAATG	TGATGGACCAGCAATGAGAC
AcCAD4	AGCCACCATCTCTTGACAGC	CCTAGGATCCATGGCTTGTC
AcCAD5	TAAGCCTGTCTTGTTCGGTG	AGACCAAACTTTCATTGAGGC

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')
AcCAD6	TCGCTCAAGCAGTCTTTTCC	GGTATTCTTTGCGGTTGATTG
AcCAD7	ATGGTCATCGTGGTGGTGC	GCGTTGATACCGTTCGTGTAG
AcCAD8	TCCGAGTCATTAGGCTGTGC	CGTTGGATTCTTCCTCTTCG
AcCAD9	TTACGCCATCGGCTACATG	CCGTCTTCGAGTTCTTCTGC
AcCAD10	TCACTGGACCAATCAACGAG	GGAACATCGTACCGAGAACC
AcCAD11	TGTGGAAGTGGAGGAAGTCA	TCAAGGAAGTGCCAGTATCG
AcCAD12	ACTTGCCACAAGGTACATCAG	CGTATCAACAAAACCAACTGC
AcCAD13	GTAACCCCCCACAAACATTC	GACCCTGACATGAGTAGCACA
AcCAL1	TTCCTTCCTCCAGCAAACG	GGACGGTTCACAGGAATCAC
AcCAL2	GATGTTTACGTTGCACCAGC	ACACCATGATCGAGCTCATC
AcCAL3	AAACCTAGTGTCCCTCAGCG	AGCTCCGGAAGGTGTAGTTG
AcCAL4	GCTCCCAGAAGAAGTGACCC	TGCCACAGGCGTTGTTCTC
AcCAL5	ACGGTTGTGAAGGAGGTTTG	CACGAGTTCTTGACGAGCC
AcCAL6	TGTAGCCATTGACGCATCTC	AGCTTGGGTTGCTATTCCAC
AcCAL7	TCGATTGCTCATTGAACTACG	CCAGTATTTCTGGCCTTTCTC
AcCAL8	CAAGGAGTTGAAGTGGTGATG	TGAACAATCAACCAGGTTCTG
AcCAL9	GAAGGAACTGGCTGAAATGG	CAACCCGAACAGAGCAAATC
AcCAL10	GCATCGGCTGAGGTTTACG	TGGGATAGGCAATGGGTAGC
AcCAL11	AATACAGAAGTGTCGCTGACG	GTTCCGTATCCAACAACCAG
AcCAL12	GAGCGAACAGAACTTGATGG	AGTTCATCAATAGCTGGGGC
AcCAL13	TGTTCTTGGGAATACCGCTG	GGCGGCATGATTCAGATTAG
AcCAL14	CCTGAAGAGGTCTCCATCCC	TTCTCCCCAGTTAAGTCCCC
AcCAL15	TCCAGGATCAAGGAGATTGC	ATCTGCATGGTCCATCCTTC
AcCAL16	TTGGTAGACTGTTCGGAGGC	TTCCTGTCCGTCCTCAACTC
AcCAL17	CTGTGGTGGTTTTCCTGGTC	CCATCCTGTCCAAGGTATCC
AcCAL18	CTGTTGGGCTTTTAGTGCG	TTCTGCCCAAAGTCTTCTCC
AcCAL19	TGATTCTGTTTGCCCTGTTG	GCGTTGTTCATATTCCCTCC
AcCAL20	CTCCGGTCAAGGAGCAAAAG	GGGTTGTCCATTCTCCTTGC
40S RP S18	GGAAATTCCTGGCCTTACTGAC	TTCTTGCCTCCTAGCCTTAGCT

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, 蠋蝽组织蛋白酶 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)。

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
	AcCAB1	Propeptide_C1、Peptidase_C1	是 Yes	37.0	5.78
Cathonsin D	AcCAB2	Propeptide_C1、Peptidase_C1	是 Yes	37.9	4.76
Cathepsin B	AcCAB3	Propeptide_C1、Peptidase_C1	是 Yes	37.6	4.99
	AcCAB4	Propeptide_C1、Peptidase_C1	是 Yes	37.0	7.53
	AcCAD1	Asp	否 No	37.9	8.23
	AcCAD2	Asp	是 Yes	44.0	9.10
	AcCAD3	Asp	是 Yes	43.4	6.20
	AcCAD4	Asp	否 No	36.2	5.82
	AcCAD5	Asp	否 No	82.9	5.42
	AcCAD6	Asp	是 Yes	47.8	6.04
Cathepsin D	AcCAD7	Asp	是 Yes	43.3	6.98
	AcCAD8	Asp	否 No	38.8	7.62
	AcCAD9	Asp	是 Yes	42.3	7.51
	AcCAD10	Asp	是 Yes	40.9	6.89
	AcCAD11	Asp	否 No	47.6	6.12
	AcCAD12	Asp	是 Yes	71.8	5.19
	AcCAD13	Asp	是 Yes	43.0	6.79
	AcCAL1	Inhibitor_I29、Peptidase_C1	否 No	44.9	7.08
	AcCAL2	Inhibitor_I29、Peptidase_C1	是 Yes	36.9	5.13
	AcCAL3	Inhibitor_I29、Peptidase_C1	否 No	34.4	5.59
	AcCAL4	Inhibitor_I29、Peptidase_C1	是 Yes	36.1	5.49
	AcCAL5	Inhibitor_I29、Peptidase_C1	是 Yes	37.6	5.73
	AcCAL6	Peptidase_C1	是 Yes	36.1	5.51
	AcCAL7	Peptidase_C1	否 No	30.5	5.16
Cathepsin L	AcCAL8	Inhibitor_I29、Peptidase_C1	否 No	56.3	5.14
	AcCAL9	Inhibitor_I29、Peptidase_C1	否 No	94.4	7.98
	AcCAL10	Inhibitor_I29、Peptidase_C1	是 Yes	74.9	5.41
	AcCAL11	Inhibitor_I29、Peptidase_C1	是 Yes	36.5	9.10
	AcCAL12	Inhibitor_I29、Peptidase_C1	是 Yes	67.3	5.97
	AcCAL13	Inhibitor_I29、Peptidase_C1	否 No	35.9	5.41
	AcCAL14	Inhibitor_I29、Peptidase_C1	是 Yes	37.1	6.26

表 2 蠋蝽基因组中鉴定出的组织蛋白酶基因及其序列结	构特征	E
----------------------------	-----	---

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
	AcCAL15	Inhibitor_I29、Peptidase_C1	是 Yes	37.1	8.31
	AcCAL16	Inhibitor_I29、Peptidase_C1	是 Yes	36.7	5.03
Cathonsin I	AcCAL17	Inhibitor_I29、Peptidase_C1	否 No	54.8	7.24
Cathepsin L	AcCAL18	Inhibitor_I29、Peptidase_C1	否 No	39.9	6.03
	AcCAL19	Peptidase_C1	否 No	34.7	8.94
	AcCAL20	Inhibitor_I29、Peptidase_C1	是 Yes	38.1	5.14

续表 2 Continued table 2

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.

DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5	: : : : : :	: DTGSSNLWVPSKKCHLTNIA_LMHNKTDASKSKTYT-KNGTEFALCIGS SLSSYLSTTTS AGLDIKD TAAEALSEPGLVFVAAK : DTGSGTS/TPKKYSGENTCAMKYKHFKSTS Q-DFNTLVKISYMKENATIKLGSGSVTICMMYKNET GLAIKSDCESKE DTAWGDWIPSILCPVLDAA INRNKTDSSRSSTVK-PTGKKYNISVGKNNLTCKISSGSVTICMMYKNET GLAIKSDCESKE DTGSSNLWIPSILCPVLDAA INRNKTDSSRSSTVK-PTGKYNISVGKNNLTCKISSGSVTICMMYKNET GLAIKSDCESKE DTGSSNLWIPSILCPVLDAA : AFGYSYFNVOCKNCSWFRTA WDHNQTDSFKSATVR-KNGKTFCIDDLESSISSFISFISTISFTSUKMIGLT GEATMPPSLDS : DTGSSNLWYPKKCKLKSLG RVHTIYNSKKSLTVK-KNGKTFCIDDLESSISSFISFISFTSUKMIGLTDDCITTEVTNEPDISFLFKK	
AcCAD0	:		
AcCAD8	÷	MASD WYFSSESLGCAF EGHRTWISINSVSHV-SNGDFTSFWIFCNIVGYLSTETWSVCKLEWENGTSSEANNVMGAPIFSKC)
AcCAD9	:	DTGSHLLWVPNKRCSNCSKSKSYV-DLKKNIS DYAISYMRCICGKSVVNVAGIEVKNCEIATALEGAFQENL	
AcCAD10	:	: DTGSHLLWVPNKRCSNCYNKNMEDEAK <mark>S</mark> KSYV-DLKRNISIDYAISYMRGICGK <mark>S</mark> VVNVAGIEVKNCEIATALEGAFQENL	
AcCAD11	:	: DTGSDLLWIPSIN-DTTTCGSGGSHSQERFSYSSTYK-NLHKMMDYAYVKCRLDYALATDTVTVCDLSVHDQVYGLSLRGDCDGQS	
AcCAD12	:	: DIGSSD_WVPSNHCWFSFA_YNHNYEKDSK <mark>S</mark> STEK-NIGKSVN ENGTCTISCTTVQDSDIGATSDAATSISRNPFYSVK	
AcCAD13	:	: AUGYSYF WY QSKNCSWFRTAGWDHNQYDSFKSAGYR-KNGKTFCDDLHGETSGIISECTEKFSSAIDTDTGEADKLPSLDS	
		-	
DmCAD	:	: FDGILGLGYNSISVDKVKPPFYAMYEQGL SAP <mark>VFSFYLNR</mark> DPASPEGCEIIFGGSBPNHYTG-EFTYLPVIR-KAY <mark>B</mark> QIKWDA	
DmCAD AcCAD1	:	: FDG1LGLGYNSISVDKVKPPFYAMYEQGLISAP <mark>V</mark> SSFYLNRDPASPEGGEIIFGSS <mark>P</mark> NHYTG-EFTYLPVR-KAYMQIKVDA : Y <mark>EGIIGL</mark> SFHTINHSSIIDNM <mark>-</mark> SQGLIKESIAFYLNRNHSSTDS-EITISCYMPKYIKNNELNTVKVD-SVHMQVPVEK	
DmCAD AcCAD1 AcCAD2	::	: FDGLGLGYNSISVDKVKPFFYAMYEQGLISAF <mark>V</mark> SFYLNRDPASPEGELIFGSDPNHYTG-EFTYLPV R-KAYMQIKVDA : MEGIGUSFHTINNSSIIDNM SQGLIKES (AFYLNRNHSSTDS-ELTIGGYMPKYIKNNELNTVKVMD-SVEMQVPVEK : A <mark>DGIFGLNFDTSTVSDGITPFFYNM</mark> KQKTVKNGISFYE <mark>NRDLSTNKGS</mark> ILL <mark>DKPKHYRGNFTSVRADEN-SGYMQIPVKG</mark>	
DmCAD AcCAD1 AcCAD2 AcCAD3	:::::::::::::::::::::::::::::::::::::::	: DGILGIGYNSISVDKVKPPFYAMYEQGLISAPVFSFYLNRDPASPEGGEIIFGSDPNHYTG-EFYLPVR-KAYMQIKVDA 1963FHTINNSSIIDNM SQGILKESIAFVINRNHSSTDS-EITIGNSKVIKNNELNTVKVD-SVHQOVPIK 2 AGIFGINFDTSVSDGITPFFYNM KQKTVKNGISSYERRENSISSIER BUD SVHQOVPIK 5 GGILGIAYDTISVDGVRPFFYLM DQQKVQEPVFSYLNRDPSSPEGGEIIFGSDADKYVG-NFTVPPVKK-KGVQEPVFS	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4	:::::::::::::::::::::::::::::::::::::::	: DGLGLGYNSISVDKVKPPFYAMYEQGL SAPVFSFYLNRDPASPEGEL IFGCSPNHYTG-EFYLPVnR-KAYMQIKVDA : YGCIIGLSFHTINRSSIDNM SQGLKESI AFYLNRNHSSTDG-ETIIGYNPKYKKNNELNTVKVD-SVHQVFVEK : AGGIGUNFDTSTVSDGITPFFYNM KQKTVKNGFSYPNRLSTNKGSSLLGKPKHYRONFTSVRADPN-SGYMQIPVG : GFGIGLAYDTISVDGVRPFYLM DQQKVGPVSTYNRDQKVGFVSGCELFFGSADKPKVG-NFTVVPKK-KGYQFDVDG : YDGLLGLGFPSKAFTGKPILTSLSKGFLKNNTSSLSLGKNKKENTACKIFGDMGNKFDLKSVNYIPLGD-NTAMIFDIDY	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5	: : : : : : : : : : : : : : : : : : : :	: PDG1LG1GYNSISVDKVKPPFYAMYEQGLISAPVOSFYLNRDPASPEGGE IFGSSPNHYTG-EFTYLPVR-KAYMQIKVDA : YEGIIGLSFHTINHSSIIDNM SQGLIKESIAFYINRNHSSTDS-EITISYNPKYIKNNELNTVKV D-SVHQVPVEK : ADG1GGMPDTSTVSDGITPFFYNM KQKTVKNGRSSYG KDLSTNKGSSLLGDKPKHYRGNFTSVRADPN-SGYMQIPVKK : FDG1LG4AYDTISVDGVRPFFYNM KQKVVQEVSSYLNRDPSSPEGGEIFFGSAAKYVG-NFTYVPV K-KGYMQFDNDG : YDG1G4GFPSISVNHIEPFFYNM KQKVVEFYSFYLNRDFSSPEGGEIFFGSAAKYVG-NFTYVPV K-KGYMQFDNDG : YDG1G4GFPSISVNHIEPFFQKM K-EIQLDPVSSYLSSYLSKAFTSKAFIFDDY : YDG4G4GFPSISVNHIEPFFQKM K-EIQLDPVSSYLSSYLS	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6	: : : : : : : : : : : : : : : : : : : :	TGILGIGYNSISVDKVKPPFYAMYEQGL SAPV-SFYIN RDPASPEGGE IF GSPNHYTG-EFYLPV R-KAYMQIKVDA "EGIIGUSFHTINNSSIIDNM SQGLFRES FAFVINRNSSTDS-ETI GYPKVIRNNELNTVKVD-SVHQOVPIR AFGIGUNPDISTVDGITPFFYNM KQKTVKNG SYFU GULSTNKGS LLGUAYDTISVDGVRPFYLM DQQKVQEPVSFYINRDPSSPEGE IF GSADKYVG-NFTYVPV K-KGYWQFDDG YGGIGGFPSKAFTGKFILTSLSKGFKNNSFBIJLSLKKGSKENTACKIF UGWGVGGEFPSKNFTGKFILTSLSKGFKNNSFBIJLSLKKGSSTNG-EFFLGMSGDVFDKNTNYVSIK-ESYGVFN STGIGGEFPSSSPLCGEVVPVQN ADQCIQAVSSFY SDGIFGLAYSSPSSPLCGEVVPVQN ADQCIQAVSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIQAVSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIQAVSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIQAVSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN ADQCIAQAYSSFY SDGIFGLAYSSPSSPLCGEVVVQN SDGIFGLAYSSPSFY SDGIFGLAYSSPSSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGLAYSSPSFY SDGIFGL	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD7 AcCAD8	: : : : : : : : : : : : : : : : : : : :	TOGILGIGYNSISVDKVKPPFYAMYEQGLISAPVISFYLNRDPASPEGGE IF GS PNHYTG-EFYLPV R-KAYMQIKVDA YEGIGGYNSISVDKVKPPFYAMYEQGLISAPVISFYLNRDPSSTDG-ETIGYPKVIKNNELNTVKVD-SVHQOVPVK A GIFGYNFDTSTVSDGITPFYNM KQKTVKNGIFSYL IGGIGGAYDTISVDGVRPFYLM DQQKVQEPVISFYLNRDPSSTDGE IF GS ADKYVG-NFTVYPV K-KGYMQFDVDG IGGIGGAPSSKAFGKFILTSL SKGFKNNFFBILTKNKKENTA KIF DWGNKFDLKSVNYIPID-DNTANIFDDY YDGVFGIGFPSISVNHIEPFFQKM K-EIQLDPVFSYLNRDPSIMNG BEFICDWGNKFDLKSVNYIPID-DNTANIFDDY IGGIGGAYSSTSUCHEFYQNMADGCAQAVSFRYKRDF-STMCGSVIFI IAKVVDEALDYITL K-KANGCAFES IGGIGGAYSSTGCFEVPVPVMADQCKVCFFYSVTFLMQUPPFFK	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD7 AcCAD8		<pre>E DG LG GYNSISVDKVKPPFYAMYEQGL SAPVSFYLN RDPASPEGE IFGS PNHYTG-EFYLPV R-KAYMQIKVDA H C IGLSFHTINHSSIDNM SQGL KESTA YLN RNHSSTDG-ETTIGYNPKYIKNNELNTVKV D-SVHQVPVEK A GG IGUNFDTSTVSDGTPFFYNM KQKTVKNG ISFYG RDLSTNKGS LL DKPKHYRGNFTSVRADPN-SGYNQTPVG E FG IGLAYDTISVDGVRPFYLM DQQKVQEPVSTYLN RDPSSEC EIFG SADKFVG-NFTYVPV K-KGYQCPDYDC 10G LGLGFPSKAFTGKPILTSL SKGFLKNN FSL LDKNKKENTA CK IF DWGGNKFDLKSVNYIPL D-NTANIFDIDY 10G LGLGFPSISVNHIEPFCKM K-BIQLDPVSYYL SRG0SINNG EFL CMREDVFDKNTLNYSINK-ESYNQFK INS 10G LGLACSGPRCEVYKDFTFSVVTLMQQL CAPSIS KI KDDAHVNDIGEVIF I DAKYDEAALDYITL K-KANNQFALSS ING 10G LGLACSGPRCEVYKDFTFSVVTLMQQL CPISEI YLN KTDDGYGE TFGGTTLLYKEDTLWHNTT-SYLASIR DS 11G LGLACSGPRCEVYKDFTFSVVTLMQQL CPISEI YLN KTDDGYGE TFGGTTLLYKEDTLWHNTT-SYLASIR DS 11G LGCAFFREDATSTDNTED COVID NUT SV KCA-CAPBUSYN E IFGGT CHOKITERGFTYTPVIESGHYQINNG TS CAPSING COVID NUT SVCA</pre>	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD7 AcCAD8 AcCAD9 AcCAD9	* * * * * * * * *	TGILGIGYNSISVDKVKPPFYAMYEQGL SAPV-SYYN RDPASPEGGE IF GS PNHYTG-EFYLPV R-KAYMQIK DA BGILGISFHTINNSSIIDNM SQGL KEST AFYL RNHSSTDS-E TI GY PKVIKNNELNTVKVD SVHQOVPIK AFGIGUFDISTVDGUTPFYLM DQQK QEPVSYIN RDPSSPEGE IF GS ADKYVG-NFTYVPV K-KGYWQFD DG TGGILGIAYDTISVDGVRPFYLM DQQK QEPVSYIN RDPSSPEGE IF GS ADKYVG-NFTYVPV K-KGYWQFD DG TGGILGIAYDTISVDGVRPFYLM DQQK QEPVSYIN RDPSSPEGE IF GS ADKYVG-NFTYVPV K-KGYWQFD DG TGGILGIAYDTISVDGVRPFYLM DQQK QEPVSYIN RDPSSPEGE IF GS ADKYVG-NFTYVPV K-KGYWQFD DG TGGILGIAFPSSNFTGKFILTSLSKCGFINN F-BILTKNKKENTACKIF DW GNKFDLKSVNYIPLD-NTANIFD DY TGGVGIGFFSISVNHIEPFFQKW K-BIQLDFVSYISSGSINGCEFFLGWSGDVFDKNTLVYSIK-ESYWGFK NS TGGIGIAYSSPSPLCGPEVVPVQN ADQGIAQAVSSRY NDDAHVNDIGEVIF GIAKYVDEAALDYITL K-KANNOFALES TGGIGISFFSISKFNFTSVVTPLMQQ PCHAEIT TEDDGVGE TFGID XKITERGFTTPVIGSKNY ON DR TGGVGISFFSRP-DALTSILDNF SQKLDKN LSYWRKGDFKDGYEFT TLVKKELFD-SKKIDVIG N-YHNEVTES FOGVGISFFSRPADITSILDNF SQKLDKN LSYWRKGDFKDGYEFT TI DYKKKIPD-SKKIDYC N-YHNEVTES	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD6 AcCAD7 AcCAD8 AcCAD9 AcCAD10 AcCAD10		TOGILGIGYNSISVDKVKPPFYAMYEQGLISAFVISFYLNRDPASPEGGE IF GS PNHYTG-EFYLPV R-KAYMQIK DA 'IBCIGYNSISVDKVKPPFYAMYEQGLISAFVISFYLNRDPSSTDG-E TIG GNPKVIKNNELNTVKVD-SVHQOVPUSK A GIFGINFDTSTVSDGITPFFYIM DQCKVGEPVISFIN RDPSSTDG-E IF GSADKYVG-NFTVVPV K-KGWQPVDS TOGIGIAYDTISVDGVRPFFYIM DQCKVGEPVISFIN RDPSSFEGGE IF GSADKYVG-NFTVVPV K-KGWQPVDS TOGIGIAFPSISVDHIEPPFYIM DQCKVGEPVISFIN RDPSSFEGGE IF GSADKYVG-NFTVVPV K-KGWQPVDS TOGIGIAFPSISVDHIEPPFYM K-EIOLDPVISFYL SSFEGGE IF GSADKYVG-NFTVVFW K-KGWQPVDS TOGIGIAFPSISVDHIEPPFYM K-EIOLDPVISFYL SSFEGGE IF GSADKYVG-NFTVPV K-KGWQPVDS TOGIGIAYPSSFICGEEVVFVQNADQG CAAVSBRY RDPAHVNDIGSVIS IF AKVVDBALDVITL K-KANGCALSS TOGIGIAYSSFSICGEEVVFVQNADQG CAAVSBRY RDPAHVNDIGSVIS IF TAKVVDBALDVITL K-KANGCALSS TOGIGIAYFSSFFICGEVVFVQNADQG CAAVSBRY RDPAHVNDIGSVIS IF TAKVVDBALDVITL K-KANGCALSS TOGIGIAYSSFSICGEVVFVQNADQG CAAVSBRY RDPAHVNDIGSVIS IF TAKVVDBALDVITL K-KANGCALSS TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKNIDVIG N-YHNETT IS S TOGIGSFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKNIDVIG N-YHNETT IS S TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKKIDVIG N-YHNETT IS S TOGIGISTFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKKIDVIG N-YHNETT IS S TOGIGISTFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKKIDVIG N-YHNETT IS S TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKKIDVIG N-YHNETT IS S TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGDFKDGGE II GWSKLFD-SKKIDVIG N-YHNETT IS S TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGSTOFKGE B TIG TAKVVKKKGSSKIFD-SKKIDVIG N-YHNETT IS S TOGIGISFFSSR-DALTSILDNF SQKLDKNN IS SYNTKKGSTOFKGGE TIG SYNTKKVSSKIFD-SKKIDVIG N-YHNETT IS S TOGIGISFFSSRP-DALTSILDNF SOKLDKN IS SYNTKGGDFKDGGE TIG SYNTKKYSSKIFD-SKKIDVIG N-YHNETT IS S TOGIGS SFFSSRP STORTISTING SYNTKKGGDFKDGGE TIG SYNTKKYSSKIFD-SKKIDVIG N-YHNETT IS S TOGIGS SFFSSRP STORTISTING SYNTKGGFYTFVIG N-YHNETT IS S TOGIGS SFFSSRP STO	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD6 AcCAD7 AcCAD8 AcCAD9 AcCAD10 AcCAD11		<pre>TOGLCICYNSISVDKVKPPFYAMYEQGLISAPVOSYINRDPASPECCE IF GS PNHYTG-EFTYLPV R-KAYMQIKVDA IS CIGCICFTUINRSIIDNM SQGLIKES FAYINRNHSSTDG-ETTIGYNEKVIKNRELNTVKVD-SVHQVPVEK ADGIFGINFDTSTVSDGITPFYIM DQQKVQEPVSYINRDPSSTDG-ETTIGYNEKVIKNRELNTVKVD-SVHQVPVEK IGGIGIAYDTISVDGVRPFYIM DQQKVQEPVSYINRDPSSTDGEIFGS ADKPKHYRGNFTSVRAPPN-SGYMQIPVKG IGGIGIGFPSKAFTCKFILTSISKGELKNNFSBILDKNKKENTAKIKIF DWONKFDLKSVNYIFLD-NTATIFDD VTGVFGIGFPSKAFTCKFILTSISKGELKNNFSBILDKNKKENTAKIKIF DWONKFDLKSVNYIFLD-NTATIFDD TOGLGIAFYSISVNHIEPFFQKMK-EIQLPVCSTYNRCBSHNGGEFFLOWONKFDLKSVNYIFLD-NTATIFDD TOGUGGISFPSKAFTCKFILTSISKGELKNNFSBILDKNKKENTAKIKIFFGCNKEDVFDKNTLNYVSINK-ESYMQFKINS IGGIGGLAYSPSELCGFEVVPVQNLADQCLQAPVSRYKKDDAHVNDIGEVIFFIFAXVDEAALDTITL K-KANGCFALSS IGGIGGISFFSKAFTCKFIFSVVTPLMQQLPCFCSIIILRTDDGYCE TFGCTNTLLYKEDTLVMHNT T-SYLSISLIDS IGGIGGISFFSKEPDALTSILDNFSQKLDKNFIFLRKKSSMDYSQNGEIFFGCTNLLYKEDTLVMNTTON-VNNFFTISS IGGIGGUGSFFFSRP-DALTSILDNFSQKLDKNFIFSVMRKGDFKDGCEIIGWNEKLFD-SKKIDYICNN-VNHNFTTISS IGGIGGIGSFFSRP-DALTSILDNFSQKLDKNFIFSVMRKGDFKDGCEIIGWNEKLFD-SKKIDYICNN-VNHNFTTISS IGGIGGUGSFFSRP-DALTSILDNFSQKLDKNFIFSVMRKGDFKDGCEIIGWNEKLFD-SKKIDYICNN-VNHNFFTTISS IGGIGGGFFFSRP-DALTSILDNFSQKLDKNFIFSVMRKGDFKDGCEIIGWNEKLFD-SKKIDYICNN-VNHNFFTTISS IGGIGGGFFFSRP-DALTSILDNFSQKLDKNFIFSVMRKG</pre>	
DmCAD AcCAD1 AcCAD2 AcCAD3 AcCAD4 AcCAD5 AcCAD6 AcCAD6 AcCAD7 AcCAD8 AcCAD9 AcCAD10 AcCAD11 AcCAD12		TIGLEGLGYNSISVDKVKPPFYAMYEQGL SAPVFSFYIN RDPASPEGGE IF GS PNHYTG-EFYLPV R-KAYMQIKUDA IGGIGFHTINNSSIIDNM SQGLFESFIAFYIN RDPSSTDS-ETIG GY PKVIKNNELNTVKVD SVHQOVFDK AFGIGUAYDTISVDGVRPFYLM DQQKVQEPVFSYIN RDPSSPEGE IF GS ADKYVG-NFTVPV K-KGYWQFDDG TYGIGGGFPSKAFTGKFILTSLSKGGLKNN SE BIT KKKKKNTACK IF DWGONFDLKSVNYIPID-D-NTAN IFD DY 100 FGGFPSSKNHEPFGKWK K-EIQLDPVFSYISSGNG EFGGL DWGONFDLKSVNYIVID D-NTAN IFD DY 100 FGGSFPSSKPTGKFILTSLSKGGLKNN SE BIT KKKKKNTACK IF DWGONFDLKSVNYIVID D-NTAN IFD DY 100 FGGSFPSSKPTGKFILTSLSKGGLKNN SE BIT KKKKKNTACK IF DWGONFDLKSVNYIVID D-NTAN IFD DY 100 FGGSFPSSKPTGKFILTSLSKGGLKNN SE BIT KKKKKNTACK IF DWGONFDLKSVNYIVIT K-SYMQKK NS 100 FGGSFPSSKP-DALTSILDNF SQKLDKN LSYMN KGCDFKGGE IF GT KLIFKGFTVID/GVIIT 100 FGS-FTFSRP-DALTSILDNF SQKLDKN LSYMN KGCDFKGGE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGSFFSSKP-DALTSILDNF SQKLDKN LSYMN KGCDFKGGE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGSFFSSKP-DALTSILDNF SQKLDKN LSYMN KGCDFKGGE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGGFFSKP-DALTSILDNF SQKLDKN LSYMN KGCDFKGGE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGGGFFSKP-DALTSILDNF SQKLDKN LSYMN KGCDFKGFE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGGGFFSKP-SLKSLFSKN SKN SKN SKGCDFKGGE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGGGFFSKFLSKLFSKN SKN SKGN SKDFGFE IFGT FFK FFK SKLFD-SKKIDVIC N-YHNGETIES 100 FGGGFFSKFLSKLFSKN SKN SKN SKGCDFKGFE II GWEKKFD-SKKIDVIC N-YHNGETIES 100 FGGGGFFSKFLSKLFSKN SKN SKN SKGN SKGC	

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

DmCAI

AcCAL1 AcCAL2 ACCAL3 AcCALS AcCALS AcCAL6 AcCAL7 ACCAL8 AcCAL9 AcCAL1 AcCAL11 AcCAL12 AcCAL12 AcCAL14 AcCAL15

AcCAL16

AcCAL17 AcCAL18

AcCAL19

AcCAL20

SEKGVTEISPA-

-HVTLPKSVDWRTKGAVTAVKDOGHCGSCWAFSSTGALDGOHFRKSGVLVSLSBONLVDCSTKYGNN-GCNG

ACCALI	. HSHGFIFLP	ANVE 17	100000000000000000000000000000000000000	SQVQDQGICGAC	WARGSARS WEG	OPERVICE RV OPE	E COLLECSONICKV-COCC
AcCAL2	YVAPAN	ENI	PSTV DWR QKGAV	TPIKNQEQCGSC	WAFSTTGSIEG	QHFLKTGKLVSL	EQNLMDCSTDEGNE-SCNG
AcCAL3	SSLERKETYS	SSDVNT	PDSTDWREKGAV	TPVKNOGOCNGC	WSESATEAMEG	OLFRSTGNLVSL	FOOLTDOSREYGNL-GONG
AcCAL4	DUDODEDDUDI						
ACCAL4	RVP5FIEVP1	ANVEV2	ANSVOWRQSGAV	IPVAN.G5CG5C	WSESSIGVING	ÖPL KVI GVP AP	I QQIIVDCSGNIGNN-GOGE
AcCAL5	: KMESFTFISH	2AHVTur	NKEV DWR TKGAV	TPVKNQGHCGSC	WARSSTCAURG	QNFRKTGKLVSL	PQNIIIDCSGSYGNN-GCEB
AcCAL6	RVPSFTFVP	PANVEV	AKSV DWR QSGAV	TPVKNQGSCGSC	WSFSSTGALEG	QLFRKNGKLVSL	EQQLIDCSGNYGNN-GCGG
AcCAL7	: IPIGIPYMSH	KRDAS	PKETDWROKGAI	SPVKDOGHOGSC	WARSSTGALPA	HNYIKTGRRWAL	DONI IDCSLNYGNN-GCEG
AcCAL8	VAVGINVOH	CKAVIII	PKEWDWPONCAW	TRIKE	MARSSTOSTRC	HSEVENCETWAL	TONI VDCSKKYCNN-CONS
ACALO	. VNOTDN	70 IUIT				OVATERCAL	
ACCAL9	KAQIPN===-		PDEIDWRNINVV	I PV ANGGLOGSC	WALSVIGNING	QIAIKIGNLVSL	INCLIVICORI-DSGOLG
AcCAL10	: YVAPAN	EN11	PSTVDWRQKIEAW	TPIKNQEQCGSC	WARSTTCSHDG	QHFLKTGKLVSL	EQNIMDCSTDEGNE-SONG
AcCAL11	: HVPSFTFVPI	2SNVEI	AKSV DWR HSGAV	TPVKNQGSCGSC	WAFSSTGAIDG	QHFKKTRKLVSL	EQQLVDCSKRY-NN-GCKG
AcCAL12	KRTGLTFLSE	PANVK	PESVDWRKEGAV	TPIKDOGKOGGO	WAFSTIGALPA	OLFROTKKLVPL	DELIDCSGKYGNO-GCSC
AcCAL13	SISKIDIDI	YENLEOOTTEKGD	TDAWDWRDKGAW	TGTKDOGKOGAC	WARSTTGALEG		FOOLLDCSSSYGNK-GONG
AcCAL14	ELDCYTYNDI	DANTDI	DESTOWETTONE	TRIVICTORS			TOOL WDG TRNYDNS - COUR
ACCAL14	. EDFGITINFI	ANIK	LOIDNRIDGAV		WARESIEMED	OTE KELIGOTI OT	EQUIVEERNIENS-GOHK
ACCALIS	KELKHNLEAR	PEDDVI	PEELDWRENGFR	IRSLOGADOGSO	MARIN VAMALIBA	QVERENGTISEL	BOORTDOSAFIGMT-COHE
AcCAL16	: KVFLKSR	-SREE	PYSVDWRVAGAV.	SPVQDQGDCWSS	WAFSAIGSLEG	QEFKKTNKLVQL	VQNLVDCSEAYGNL-GCNG
AcCAL17	- PFPYDVNKM	MNDAV	PKQY <mark>DWR</mark> LY <mark>G</mark> AV	TPVKDQSVCGSC	WSFGTTGAVEG	AYFVKTGKLLRL	QQALIDCSWGYGNN-GCDG
AcCAL18	FPNVKNRLKE	RSLIDT	PRKWOWRKKRCV	TAVKNOLACGAC	WARSAIETWES	MVAIKSGILKKF	VOEFIDCAKNGNLGCDC
AcCAL 19	DONVROVECO	CEDVI VI	DIVETOWDOVCAW				
A CAL 20					ATTONICO A		
ACCAL20	SSERINKEED		PRSVDWRLANWV	HEARTOVER TOP	MINERVERVED	QUINKIGELVIL.	NUMBER OF CHERT
DmCAL	G-LMDNAFR	INDNGGIDTEKSYI	P-YEAIDDSCHFI	NKGTVGATDR	GETDIPQGDEK	KMAE&V&TVGPV3	VAIDASHESFQFYSE
AcCAL1	G-FFNYTEN	LENVEGUQSEDSY	P-FEEKAGT <mark>C</mark> RY	DKKRAVPGTKVK	AYIDEPANDD	ALKAAVATVGPIS	IAASAGSDNFQFYGG
AcCAL2	-LMDYARE	WKNNKG DTPASY	-YEAEDDTCRF	KKANVGATIT	GWKDWOSGSES	ALOSAVANIGPIS	WAIDASSEDIOLMSS
ACCAL3	-LMDETRK	TKDVGGWMSEDSVI		APKSVVNGSMVV	SYDNI PPGDEO	ALKEALANVGPV	WAWABANOK MEVEG
A CALA							
ACCAL4	G-LMDNSEK	ENVEGNESEDST	P-TEGVDSRCRT	NKNKVVAGTKVK	AMTDIESEDDD	ALKAGVØTVGPIS	AVSAGNEHEMYNKG
AcCALS	G-LMDNAEQ	MENHGEDTEKSYI	P-YDAEDETCRF	RRSNIGATDS	GEVDIPSGDEE	ALMQOVATION	MAIDASHQSIQFWSE
AcCAL6	G-LMDNSFK	TLENVAGESEDSY:	P-YEGKAGR <mark>C</mark> RY	NKNKVVAGTKVK	AYTDIESEDDD	ALKAAVATVGPIS	SIAVSAGNNHEMYYKG
AcCAL7	G-LMEQARO	KDNEGIDTEESY	P-YEGEDSECRF	KKNNVGATDA	GEVSEPTGDEQ	ALLEAVATOGPUS	VAIDASHPSFOFYSE
ACCAL8	-TTTNTSN	TRENGONDTRESVI	-YEGKEAECEE	DKSSVGAATE	GOVNI PKGDEH	VIKEAVATHGEV	WATDASHSSTOHVEK
AcCALO				MCCAUDUCTTCC	INTOODER	DMARWEUANCET	
ACCAL9	- LF DI AHIA	ALL DIGGET LESDE		NSSAVRVSIISS	LNISSDEI	DWARWLVANGPI	GINNNAMQ ILGGVS
AcCAL10	G-LMDYAFE	WKNNKG DTDASYI	P-YEAEDDTCRF.	KKANVGATIT	GYKDVQSGSES	ALQS@V@NIGPIS	WAIDASSFDEQLMSS
AcCAL11	G-LMDNSFK	TEAVEGETESSY:	P – YQGKEGK <mark>C</mark> RYI	NKKKAVPGTKVK	AYSKIRARDDN	ALKAAVATAGPVS	IAVSAGNNHELHYKG
AcCAL12	G-LMDNARO	IEENGGWASEESY	-YQGEVGNCR-	-TKNTVHGTTVK	SHVDIPEGSER	ALQEAVATVGPV	VAVSASNEGFLHYGGVGQY
AcCAL13	- LMSOART	A SANSET DTEDSY	P-YEGAVGSORE	KSSDVAATVS	KVEF FERNEG	DLOEAVATVGPV	WSTDSSLOS OLVGG
AcCAL14	-WMDNCED			NDROUUDCTVUV	AMEDIBIUDDO	AT KANNA TVOLT	MUHSCNKH OFVEC
ACCAL14	- WHDN SER	ODV DE EREDSE	- LEGVDGRORI.	NKSQVVFGIKVK	ANTONKUNDOQ	ABRASVSIVOLIS	MUHSGNKH OFFEG
ACCALIS	-SLKNTAK	QRIRL PRDKIP	-MKGKQSLCRP	QALYNKTGVNVK	TWGVLPIGDED	ALKKOVØTVGPVA	A SANGUTDLEOTERHS
				KSTEVGAEDS	OTHER DRODER		
AcCAL16	G-RPSYAFN	IRDNGGVDVEQFY	- LARDGPORE		GTINIPRGDEF	ALLEENVISVGPIA	WTIDGSLSSTAHYGG
AcCAL16 AcCAL17	G-RPSYAFNY G-EDFRSYQU	YIRDNGGVDVEQFY WIMKHGGLPLESDY	GYLGQDGY C HI	NKTNLTAKIT	GYTNIPRGDEF	ALEEAVASVGP17 AlklaifkhGpv	AVTIDGSLSSFAHYGG VAIDASQKTFSFYSN
AcCAL16 AcCAL17 AcCAL18	G-RPSYAFN G-EDFRSYQ GDTCNLLQW	YIRDNGGVDVEQFYI NIMKHGGLPLESDYO IVDNDVGIQEBHDYI	-ILARDGFCRH GGYLGQDGYCHI -LVLKTEKCKLI	NKTNLTAKIT KEVGKGIHIASN	GYTNIPRGDEF GYVNVTSGDED YTCDYLVGEED	ALEEAVASVGPIA ALKLAIFKHGPV TLLSLLAYHGPVA	AVTIDGSLSSFAHYGG VAIDASQKTFSFYSN VAVNALTWOYYLGGV
AcCAL16 AcCAL17 AcCAL18 AcCAL19	G-RPSYAFN G-EDFRSYQ GDTCNLLQW	YIRDNGGVDVEQFY IMKHGGLPLESDY(IVDNDVGIQEEHDY ILONSGIDTEOAY	GYLGQDGYCHI -LVLKTEKCKL	NKTNLTAKIT KEVGKGIHIASN	GYTNIPRGDEF GYVNVTSGDED YTCDYLVGEED	ALEEAVASVGPI ALKLAIFKHGPV TLLSLLAYHGPV ALLAAVAFIGPV	AVTIDGSLSSFAHYGG VAIDASQKTFSFYSN VAVNALTWQYYLGGV
AcCAL16 AcCAL17 AcCAL18 AcCAL19	G-RPSYAFN G-EDFRSYQ GDTCNLLQW G-NMNNAFR	YIRDNGGVDVDQFY NIMKHGGLPLESDY IVDNDVGIQEBHDY YILQNSGIDTEQAYI	SGYLGQDGYCHI -LVLKTEKCKL -YEGKEGSCRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR	GYTNIPRGDEF GYVNVTSGDED YTCDYLVGEED GFKSVAP-SEA	ALEEAVASVGP1/ AlklaifkhGpv TllslayhGpv/ AllaavaeiGpv	AVTIDGSLSSFAHYGG VAIDASQKTESFYSN VAVNALTWQYYLGGV TGHHATSN-FMKYQS
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20	G-RPSYAFN G-EDFRSYQ GDTCNLLQW G-NMNNAFR G-WMDTAFE	YIRDNGGVDVBQFY MIMKHGGLPLBSDY IVDNDVGIQEDHDY MILQNSGIDTBQAY MILNN-GIDTBDSY	GGYLGQDGYCHI GGYLGQDGYCHI P-LVLKTEKCKL: P-YEGKEGSCRF P-YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT	GYTN PRGDEF GYVNVTSGDED YTCDYLVGEED GFKSVAP-SEA GYADLPPGDEK	AHEEAVASVGPI ALKLAIFKHGPV TLLSLLAYHGPV ALLAAVAEIGPV VLKRAVATVGPV	NTIDGSLSSBAHYGG VAIDASQKTSSFYSN NAVNALTWQYYLGGV IGIHATSN-BMKYQS VAIDSKDGDGQHYKN
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20	G-RPSYAFN G-EDFRSMQ GDTCNLLQWI G-NMNNAFR G-WMDTAFE	Y IRDNGCVDVBQFY /IMKHGCDPLBSD/ IVDNDVGIQEBHDY /ILQNSCIDTBQA /ILNN-CIDTBDS/	GYLGQDGYCHI GYLGQDGYCHI P-LVLKTEKCKL P-YEGKEGSCRF P-YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT	GYTN PRGDEF GYVNVTSGDED YTCDYLVGEED GFKSVAP-SEA GYADLPPGDEK	ALEENVASVGPIA ALKLAIFKHGPV TILSILAYHGPV ALLAAVAEIGPV VLKRAVATVGPV	NTEDGSLSSBAH GG WAIDESQKTESFUSE NAVELTWOYLGGV HGEH TSN-BMK 4QS WAIDSKDGDQHKEN
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20	G-RPSYAFN G-EDFRSYQ GDTCNLLQW G-NMNNAFR G-WMDTAFE	YIRDNGGVDVBQFY WIMKHGGBPLBSDY IVDNDVSIQEHDY YILQNSGIDTBQAY YILQNSGIDTBQAY	GYLGQDGYCHI -LVLKTEKCKL -YEGKEGSCRF -YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT	GYTN PRGDEF GYVNVTSGDED YTCDYLVGEED GEKSVAP-SEA GYADI PPGDEK	ALEEAVASVGPI ALKLAIFKHSPV TLLSILAYHSPV ALLAAVAEISPV VLKRAVATVSPV	NTH DGSLSSBAHNGG WAIDDSQKTSSFUSN NANN LHVQYLGGV IGHATSN-SMKNQS WAIDSKDGDQHNKN
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL	G-RPSYAFN G-EDFRSYQ GDTCNLLQWI G-NMNNAFR G-WMDTAFE	Y RDNGG VDVBQFY Y IMKHGG PLBSDY I VDNDVG QEBHDY Y ILQNSG IDTBQAY Y ILNNG IDTBQAY SPCCDAQNIDHG	GYLGQDGYCHI -IVLKTEKCKL -YEGKEGSCRF -YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT DESG-EDYWLVK	GYTN PRGDEF GYVN TSGDED YTCDYLVGEED GFKSVAP-SEA GYAD PPGDEK	ALEEAVASVGFI ALKLAIFKHGPV TLLSLLAYHGPV ALLAAVAEIGPV VLKRAVATVGPV	NTH DGSLSSBAH KGG VAID SQKTSSFISN NAVNALTWQYYLGGV IGHH TSN-EMK KSS VAIDSKDGDEQH KN VAIDSKDGDEQH KN
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL1	G-RPSYAN G-EDFRSYO GDTCNLLQW G-NMNNAR G-WMDTAE	Y HRUNGGVDVƏQFY MAKHGƏ PLƏSDY I VDNDVG IQE PHDY I LQNSG IDTƏQAY YILNN-GIDTƏQAY YILNN-GIDTƏDSY SPQQDAQNIDHG 3DSGE-PS-KINHA	JEARDGTO HI GYLGQDGYCHI -LVLKTEKCL YEGKEGSCRF -YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT DESG-EDYWLVK	GYTN PKGDEF GYVNWTSGDED YTCDYLVGEED GFKSWAP-SEA GYADI PPGDEK NSWGTTWGDKG	ALEEAVASVGFI ALKLAIFKHGPV TLLSILAYHGPV ALLAVAEIGPV VIKRAVATVGPV FIKMLRNKENOO	WT DESLSSAH CG VA DEGKTSFISSAS VA N. TWOYLGY IG H. TSN-BMK QS VA DSKDGD QH KN VA DSKDGD QH KN
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL1 AcCAL12	G-RPSYAN G-EDFRSYQ GDTCNLLQW G-NMNNAR G-WMDTAE	VERDAGEVOVOQFV AMKHGOPLOSDA VONDORQEBHDV VILQNSGIDTOQAY VILNN-GIDTOQAY SPQCDAQNIDHG SSSC-PS-KINHA	JEARDGTC HI GYLGQDGYCHI JEVLKTEKCKL YEGKEGSCRF YEAEIKECRF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT DESG-EDYWLVK ENGEDFWIIK	GYIN FRGDEF GYVN TSGDED YTCDYLVGEED GFKSWAP-SEA GYAD PPGDEK NSWGTTWGDKG NSWGTTWGDKG	ALEEAVASVGFI ALKLAIFKHGPV TLLSLAYHGPV ALLAVAEIGPV VLKRAVATVGPV FIKMLRNKENQC YMRLTRAKKNAC	NT DESLSS AH GG VA D SQKTSF SF SN VA ND TWOYYLGGV TG H TSN-SMK QS TA DSKDGDQH KN TA AS YD V S CA D LE
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL1 AcCAL1 AcCAL2	G-RPSYAN G-EDFRSYQ GDTCNLLQW G-NMNNAR G-WMDTAE G-WMDTAE	VIRDNGGVDVDQFY AMKHGGPLDSDY IDNDVGIQESHDY VILQNSGIDTDQAY VILQNSGIDTDQAY SEQCDAQNIDHG SDSGE-PS-KINHA 25SGS-SSELDHG	I EARDSFORMER GYLGQDGYCHI -LVLKTEKCKL YEGKEGSCRI YEAEIKECRF VLVVGFGT ILIVGYGSI VLVGYGSI	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATIT DESG-EDYWLVK ENGEDFWIIK ENGKDYWLVK	GYIN PRODE GYVN YTSGDED YTCDYLVGEED GEKSVAP-SEA GYAD PPGDEK NSWGTTWGDKG NSWGTEWGEKG NSWGEEWGEDG	ALEENVISVGPI ALKLOIFKHGPV THISILAVHGPV ALLAVHEIGPV VLKRAVATVGPV FIRMLRNKENQ YHRMRRNKENQ YHRMRRNKENQ YHRMRRNKENQ	VT DGSLSS AH GG VA DGSQKTSFISS VA NLTWOYLGGV IG HITSN-BMK QS VA DSKDGDGQHKN TASAS YD V TASAS YD V TASAS YD V
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL1 AcCAL1 AcCAL2 AcCAL3	G-RPSYAN G-EDFRSYQ GDTCNLLQW G-NMNNAR G-WMDTAF GIPD GIPD GILTM	IRDNG CVDV CFN IMKHGC PL SDY(ILWKHGC PL SDY(ILWN-CHDT CAY ILWN-CHDT CAY ILWN-CHDT CAY SPCCD-AQNI DHG SDSCE-PS-KNHA ISSS-SDEIDHG VGCCLDI DHA	CONTRACTOR	NRTNLTAKIT KEVGKGIHIASN QRGAVGATIR DPSKIGATIT DESG-EDYWLYK ENGEDFWIIK ENGKDYWLYK SKVANEDYWIYK	GYTN PRGDEP GYVN YTSGDED YTCDYLVGEED GFKS VAP-SEA GYADI PPGDEK NSWGTTWGDKG NSWGTTWGEKG NSWGEDWGIDG NSWSPKWGEKG	ALEENVISVEPI TUISILIYHEPV AULAIVAEIGPV VURRAVITVEPY FIRMLINKENQ VURLIAKKAA VIRMSINKKAN VIRMSINKKAN	NT DOSLSS AH GG VA DISOKTSFISF VA DISOKTSFISF VA ND TWOYYLGGV TG HITSN SMK QS TG SLSS QH KN TG SCA VIE TS CA VIE TA QA VIE VA QA VIE TS CA VIE TA QA VIE TA QA VIE
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL1 AcCAL2 AcCAL3 AcCAL4	G-RPSYAN G-EDFRSMO GDTCNLLQQ G-NMNNAR G-WMDTASE GIVN GIVN GIVN GIVT	$\label{eq:response} \begin{split} & \mathbb{R} \cdot \mathbb{D} N \in C N \cup \mathbb{C} \subset \mathbb{P} \\ & \mathbb{I} \setminus \mathbb{D} N \cup \mathbb{C} \subset \mathbb{D} \subset \mathbb{C} \\ & \mathbb{I} \setminus \mathbb{D} N \cup \mathbb{C} \subset \mathbb{D} \subset \mathbb{C} \\ & \mathbb{I} \setminus \mathbb{D} \setminus \mathbb{C} \subset \mathbb{D} \subset \mathbb{C} \\ & \mathbb{I} \setminus \mathbb{C} \subset \mathbb{C} \\ & \mathbb{P} \mathbb{C} \\$	SCHLEQDEYCHI -LVLKTEKCKLI -VEGKEGSCRF -VEAEIKECRF VLVVGFCS VLVVGYCS LLVVGYCS	NATNLTAKIT KEVGKGIHIASN QRGAVGATIR DPSKIGATIT DESG-EDYWLVK ENGEDFWLVK ENGEDFWLVK SKVANEDYWLVK VNGQDFWLVK	GYIN FRGDEP GYUN TSGDED YTCD'LVGEED GKSVAP-SEA GYAD PPGDEK NSWGTTWGDKG NSWGTTWGEKG NSWGFEWGEKG NSWGFPWGEKG	ALESIVISYOPI TULSILIYHOPY AULAIVIEIGPY VUKRIVIIGPY YUKRIVITY YUKLIAKKIA YIKITISKIA YIKITISKIA	NT DOSLSS AH GG VA DASQKT SSF JSN VA DISQKT SSF JSN IG HATSN-DMK QS VA DSKDGD QH KN IS CALY DIV
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL1 AcCAL2 AcCAL3 AcCAL4 AcCAL4	G-RPSYAN G-EDFRSMO GDTCNLQW G-NMNNAR G-WMDTAF G-WMDTAF G-GIPDO G-GIPDO G-GIPT G-GIPT G-GIPT	HRDNG CVDV CFL HMKHGC VPL SD (DDNDVC DC HD SD (LDNN-CD DT CAN SPQCDAQN DHG SDS D-PS-KINHA SCSC-SDE DHG QCCGLD DHA JDT O-PE-SD HG SPCS-SN DHG	SCHLEDDEYCHII CULKTEKCKII VLVUGFGT ULVGFGT ULVGYGS VLVGYGS ULVGYGS ULVGYGS ULVGYGS	NKTNLTAKIT KEVGKGIHIASN QRGAVGATIR DPSKIGATIT DESG-EDWDVK ENGEDEWIK ENGEDEWIK SKVANEDYWIV SKVANEDYWIV VNGQEWUV	GYIN, TSGDED YTCDYLVGED GFKSVAP-SEA GYADLPPGDEK NSWGTEWGEKG NSWGEDWGIDG NSWSFRWGEKG NSWGFVWGEKG	ALEANNEKSOP TILSILAYHGPY ALLANVAIGPY ALLANVAIGPY YEKRWATVGPY YEKRWATVGPY YEKRANA YIKKSNKKNN YVAITANKGNT YWALTSENNA YMALTSENNA	AT DOSLSSAN GG VA DOSCASSAN GG VA DOSCASSAN GG VA DOSCASSAN GG IG HATSN GMK QS IG HATSN GMK QS IG ASSOCASSAN GG
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL1 AcCAL2 AcCAL3 AcCAL4 AcCAL5 AcCAL6	G-RPSYAN G-EDFRSIQ GDTCNLLQW G-WMDTAFR G-WMDTAFR G-WMDTAFR G-WMDTA G-UVNN G-UVN	HRDNGCVDVCPL UMKHGCVDVCPLSDI HDNDVCDCEHDI HDNNCCDEHDI SDSCEPS-KINHA ESSC-SDEDBI SDSCE-PS-KINHA ESSC-SDEDBI SDCC-PS-SDBC SDCC-PS-SDBC	SCHARDSFEER GCHARDCYCHI LVLKTERCKL GCHARDCYCHI GCHARDCYCHI GCHARDCYCHI VEAEIKECRF VLVVGFC	NKTN-LTAKIT KEVGKGIHIASN QRGAVG-ATIT DESG-EDYWLVK ENG-EDFWLVK ENG-EDFWLVK SKVANEDYWLV VNG-QDFWLVK EDDKKYWLVK	G TIN PRODE G TVN TSODED YTCDYLVGEED G XSVAP-SEA G AD PPGDEK NS GTTWCDKC NS GTTWCDKC NS GEDWCEKC NS GEDWCEKC NS GTWCEKC NS GTWCEKC NS GTWCEKC	ALESIVISVER TULSILIYHEROV ALLAIVISIOPV VLKRVATVOOV YMLTIAKKA YMKSAKKA YKSAKKA YKIAKKA YKIAKKA YKIANKEN YMLTISENA	NT DOSLSS AH GG VA DOSCASS AH GG VA DOSCASS PARCON AN DITWOYLLGV IG HUTSN-WARGS IG HUTSN-WARGS IG HUTSN-WARGS IS CALVE SCALVE IS CALVE IS CALVE
AcCAL16 AcCAL17 AcCAL18 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL1 AcCAL2 AcCAL3 AcCAL4 AcCAL5 AcCAL6 AcCAL7	G-RPSYAN G-EDFRS/O GDTCNLQW G-NMNNAR G-WMDTAFE 	HRDNG CVD V CFU HMKHG CVD CE HDY I DDDV C CE HDY I LOND C CE HDY I LOND C DT CAY SEQUE - AQNI DHG SEC - SED HHG VGC G - LDI DHA SDT O - FE S DHG VGC G - SD DHG SDT O - FE S DHG DT O - FE S DHG DT O - FE S DHG	JARDJO CHI JULKTEKCKI VEGREGSCR VEALIKEC SUBJUST VGY	NKTN-LTAKIT KEVGKGIHIASN DESG-EDYWIK ENG-EDFWIK ENG-EDFWIK SKVANEDYWIY SKVANEDYWIY SKVANEDYWIY ENG-QDWIY ENG-QDWIY	G TIN PRODE GUYN TSODD YTCDYLVGED GISS AP-SEA GIAD PPGDEK NS GTTWGDKG NS GTDWGEKG NS GEDWGID NS SPRNGEKG NS GTWGEKG NS GTWGEKG NS GTWGEKG	ALEEL VISVEPI ALLA ITR SPV TILSIL VIEV ALLA VIEVEPV VIRIV TVSPV FINIT ARKIN VIRIV TR AKKA VISSINKEN VISSINKEN VIELSEN VIELSEN VIELSEN VIELSEN	AT D CSLSS AH CG VA D SQKTSF SF SN VA N. TWOYYLGGV TG H TSN-SMK QS TG A SS VD V SCA Y D V SCA Y D V SCA Y D V SCA Y D V SS NO V
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL2 AcCAL2 AcCAL3 AcCAL4 AcCAL4 AcCAL5 AcCAL6 AcCAL7	G-RPSYAN G-EDFRS G GTCCNLQW G-NMNNAR G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMTA G-G-VY G-G-G-VY G-G-V	HRDNGCVDVCPL UMKHGCVDVCPL IDNDVCPQE IDNOV <	I ANGEL C LEQDOY HI - LUKTERCIL - LEGKEGSOR - LEGKEGS	NRTNLTAKIT KEVGKGIHLASN QRGAVGATLR DESG-ED KUGED SKVANED VIV SKVANED VIV GDKK KU EDGKK KU CHGQD KIV EKGQK KU	GIN FRODE YTCDYLVGED G KSVAP-SEA G AD PPGDEK NSGEDWGEN NSGEDWGEN NSGEWGEN NSGEWGEN NSGEWGEN NSGEWGEN NSGEWGEN	ALESIVISVER TULSILIYHEROV ALEAIVAEIGPV ALEAIVAEIGPV YURROVIVOPV YURROVIVOPV YURROVIVOPV YURANKSNN YURANKSNN YURANSSNNA YURASSNA	T D DGLSS AH GG VA D SQKTSFSFSSG VA D SQKTSFSFSSG AN M LTWOYLGGV IG H_TSNMK QS IG H_TSNMK QS IS SCALYPIE SCALYPIE SCALYPIE A FPY TEDSKNSDKS SMAYPIE IS (ASYPIE) IS (ASYPIE) IS (ASYPIE) IS (ASYPIE) IS (ASYPIE) IS (ASYPIE)
AcCAL16 AcCAL17 AcCAL18 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL2 AcCAL2 AcCAL3 AcCAL4 AcCAL4 AcCAL6 AcCAL7 AcCAL8	G-RPSYAN G-EDFRSYO G-NMNNAR G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMDTA-E G-WMTA G-WMTA-E G-WMT	HRDNG CVD V CFU HMKHGC VE LSD (I DDDV C C C HDY I LDNS C DT CA SEQUE - AQNI DEG SDS C - PS KINHA SDS C - PS KINHA SDS C DD DHA SDT C - PE SI DHC SDT C - PE SI DHC	Image: Construction Image: Construction Image: Constretin Image: Constructin	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATLT DESG-ED WI SENGED WI SKVANED WI SKVANED WI SKVANED WI ENGQD WI ENG-QD WI ENGQC WI ENGKK WI	G TA PRODE G VN TSODD TTCDJUGED G KS AP-SA AD PPGDEK NS GTENGEKG NS GENWGEKG NS GENWGEKG NS GENWGEKG NS GENWGEKG NS GENWGEKG NS SEIWGENG	ALEE VISVEP ALLA IKK SPV TLEILAVESPV ALLAVESPV VLKRVATVSPV YKRVATVSPV YKRSVKKNN YVITARSNKKNN YVITANKKNN YVITANKNN YVITASNA YKALDODNN YMALSSNN YMA	NT D CSLSS AH CG NA D SQKTBSF SN NA N. TWOYYLGGV TG H FSN SMK QS TA DSKDGDQH KN SCA YD V SCA YD V SCA YD V SCA YD V A DSKDGDQH KN SCA YD V A QA YD V
AcCAL16 AcCAL17 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL1 AcCAL1 AcCAL2 AcCAL3 AcCAL4 AcCAL4 AcCAL5 AcCAL6 AcCAL7 AcCAL8 AcCAL8 AcCAL8	G-RPSYAN G-EDFRS O GTCCNLQW G-WMNNAR G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMTA	HRDNG CVD V CPU UMKHG CVD CQ HDI ULQNS CD CQ HDI ULQNS CD CQ HDI VLUNN - CD CQ SY SPQC D - AQNL DHG SDSC - PS KLWHG CQ CG - LDL DHA SDT C - PE SLOHG SDT C - PE SLOHG SPE S - SVQL DHG SPE S - SVQL DHG SPE S - AREL DHG SPE S - AREL DHG	I ANGEL C LEQDOY H I - ULKTERCIL - VEGKEGSOR - VEGKEG	NRTNLTAKIT KEVGKGIHLASN QRGAVGATLR DESG-ED WI ENGED WI SKVANED WI VNGQD WI VNGQD WI EDDKK WI EGDKK WI ENGQK WI LHRELF WI	GTAN PRODE YTCDYLVGEED G KSVAP-SEA GAD PPGDK NS GTDYGEK NS GEDWGEK NS GEDWGEK NS GENWGEK NS GENWGEK NS GENWGEK NS GENWGEK NS GENWGEK NS GENWGEK NS SEIWGNK SEIWGKK	ALESIVISYGPI TELSILIYHGPY ALLAITKSPY ALLAIVIEIGPY ALLAIVIEIGPY YLRIVISYGPY YLRIVISYGPY YLRIANGOT YLRIANGOT YLRIANGOT YLRIASENNA YLRIASENNA YLMASENNA YLMASENNA	TD DGSLSS AH GG VA DSQKTSFSFSFSG VA DSQKTSFSFSFSG IG HJTSN
AcCAL16 AcCAL17 AcCAL18 AcCAL18 AcCAL19 AcCAL20 DmCAL AcCAL2 AcCAL2 AcCAL3 AcCAL3 AcCAL4 AcCAL5 AcCAL6 AcCAL7 AcCAL9 AcCAL9 AcCAL9	G-RPSYAN G-EDFRS/G GDTCNLQW G-NMNNAR G-WMDTAFE G-WMDTAFE G-WMDTAFE G-WMTAFE G-WTAFEG G-WTA	HRDRGCVDVCPU JMKHGCVDVCPU JDNDVCDQE JDNDVCDQE HDNDVCDQE	ARAGE SLODORY HI LUNKTEKCI VECKEGSC BAEIKE VOF VOF VOF VOF VOF VOF VOF VOF VOF VOF VOF VOF VOF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR DPSKIGATLT DESG-ED WI ENGED WI SKVANED WIV SKVANED WIV SKVANED WIV EDDKK WI EDDKK WI ENGQD WIV ENGKK WI LLHRELP WI	GTAN FRGDED YTCDYLVGED G KSVAP-SEA NS GTTWGDKG NS GTTWGDKG NS GTTWGDKG NS GTWGEKG NS GTWGEKG NS GTWGEKG NS GTWGEKG NS SEIWGKG NS SEIWGKG NS SEIWGKG	ALEE VISVER ALLA IKR SPV TILSILAVISPV VERNVATVSPV YKRVATVSPV YKR VATVSPV YKR SKRKNN YKR SKRNN YKR SRNN YKR SRNN YKRA DODNN YKRA DODNN YKRA SRNN YKRA SRNN YKRA SRNN YKRA SRNN YKRA SRNN	NT DGSLSS AH GG NA DSQKTBSF SN NA N.T.WOYYLGGV TG TG H.T.SN-SMK QS TA DSKDGDQH KN TS CA DSKDGDQH KN TS CA Y NA PY Y LEDSKNSDKS SMA Y Y
AcCALI6 AcCALI8 AcCALI8 AcCALI9 AcCAL20 DmCAL AcCAL2 AcCAL2 AcCAL2 AcCAL3 AcCAL4 AcCAL5 AcCAL6 AcCAL7 AcCAL9 AcCAL9 AcCAL9 AcCAL9 AcCAL9	G - RPSYA N G - DFRS N G - DFRS N G - WMNNA R G - WMNNA R G - WMDTA B G - WMDTA B G - WMDTA B G - WMDTA B G - WMNTA B G - WMNT	HRDNG GVD V GPU IMAKG GVD G GE HDY IDNDVG GE HDY ILONSG DT GAS SPQG D - AQNL DHG SDS E PS-KINHA SES S - SDE DHG GG G - LD DHA SDT D - PE-SI DHG SPE S - SVU DHG	Image: Construction of the second s	NRTNLTAKIT KEVGKGIHLASN QRGAVGATLR DESG-ED WIL ENGED WIL SKVANED WI SKVANED WI SKVANED WI ENGQD WI ENG-QD WI ENG-QC WI LLHRELP WI LLHRELP WI LLHRELP WI	GTAN FRGDED YTCDYLVGEED G KSVAP-SEA G AD PPGDK NS GETENGEKG NS GETENGEKG NS GETENGEKG NS GETNGEGG NS GETNGEKG NS GETNGEKG NS SELWGKG NS SELWGKG NS SELWGKG	ALEE VISVEPI ALLA IKR SPV TILSIL VHEROV VIRAVISVES YNTANS YNTANSN YNTANSN YNTANSN YNTANSN YNTANSN YNTA	TD DGSLSS AH GG VA DSQKTS FSF SN VA NA TWOYYLGGV IG H.TSN SK SK IG H.TSN
AcCALI6 AcCALI8 AcCALI8 AcCALI8 AcCAL20 DmCAL AcCAL20 DmCAL AcCAL2 AcCAL3 AcCAL3 AcCAL3 AcCAL4 AcCAL4 AcCAL5 AcCAL6 AcCAL6 AcCAL6 AcCAL7 AcCAL8 AcCAL8 AcCAL8 AcCAL8 AcCAL9 AcCAL10 ACCAL10 ACC	G-RPSYAN G-DFRS/G GDTCNLQW G-WMDTAF G-WMDTAF G-WMDTAF G-WMDTAF G-WMDTAF G-WMDTAF G-WMDTAF G-WMTAF G-WMTAF G-	HRDNGCVDVCP HRENGCVDVCP JUNDVCDQE JUNDVCDQE JUNNCDQE	ARGGE LULKTEKCI LULKTEKCI LULKTEKCI LUCKTEKCI BAELIKEOR KUCF VCF VGCF	NKTNLTAKIT KEVGKGIHIASN QRGAVGATLR QPSKIGATLT DPSKIGATL ENGED WI SKVANED WI SKVANED WI SKVANED WI EDDKK WI EDDKK WI ENGQD WI ENGQD WI LLHRELP WI LLHRELP WI DKGKC WI DKGCD WI	GIAN FRODE YTCDYLVGED G KSVAP-SEA NS GTTWGERG NS GTTWGERG NS GTTWGERG NS GTPWGERG NS GTWGERG NS GTWGERG NS SEINGNK NS SEINGNK NS SEINGNK NS GTWGERG NS GEPWGERG	ALEEL VISVER ALLA IKR SPV TILSIL VK SPV VERSV TV SPV Y KR V TV SPV Y KR V TV SPV Y KR V TV SPV Y KR SV TV SPV Y KR SK A Y KR SR SR SR Y KR SR SR SR Y KR SR SR SR SR SR Y KR SR	NT DGSLSS AH CG
AccALl6 AccALl7 AccALl8 AccALl9 AccAL20 DmCAL AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL4 AccAL5 AccAL6 AccAL7 AccAL9 AccAL9 AccAL2 AccAL16 AccAL16 AccAL9 AccAL16 AccAL17 AccAL16 AccAL16 AccAL16 AccAL17 AccAL16 AccAL17 AccAL16 AccAL16 AccAL16 AccAL16 AccAL16 AccAL16 AccAL16 AccAL16 AccAL16 AccAL17 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL19 AccAL10 AccAL19 AccAL10 AccAL19 AccAL10	G-RPSYAN G-DFRSJO GTCCNLQW G-MMNNAR G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMDTA B G-WMTA B G-WMNTA B G	$\label{eq:restriction} \begin{array}{c} \text{IR} \text{RDRG SVD V} \in \text{FV} \\ \text{IMRG GV E L SDM (} \\ \text{ID} \text{RDNV C Q E HDY (} \\ \text{ID} \text{RDNV C Q E HDY (} \\ \text{ID} \text{RDNV C Q E HDY (} \\ \text{ID} \text{NSC D T 2 DSY (} \\ \text{ID} \text{NSC D SY (} \\ \text{ID} NSC D SY $	Image: Construction Image: Construction Image: Constretion Image: Construction	NRTNLTAKIT REVGRGIHLASN QRGAVGAID DSKIGAIT DSKIGAIT DSKIGED WI SKVANED WI SKVANED WI SKVANED WI ENGQD WI ENG-QD WI ENG-QC WI LLHRELP WI LLHRELP WI LLHRELP WI ENGKWI DGG-QD WI	GTAN FRGDED YTCDYLVGEED G KS MAPSEA G AD PPGDK NS GETENGEKG NS GETENGEKG NS GETWGEKG NS GETWGEKG NS GETWGEKG NS GETWGEKG NS SILWGKK NS SILWGKK S SILWGKK S SILWGKK S SILWGKK S GETWGECG NS GETWGECG	ALEEN VISVEPI ALLAITRESV TILSILAYHENO VIRAVIEIOPV ALLAIVEIOPV YIRAVITSPY YIRAVIEN YIRASNKKIN YIRASNKKIN YIRASNKKIN YIRASNKKIN YIRASNKKIN YIRASNKKIN YIRASNKKIN	T DGSLSS AH GG VA DGSQKTSFSFSSFSS VA VA DSKDGDQH KN TG H.TSN MC QS TG H.TSN MC QS
AccALl6 AccALl7 AccALl8 AccALl8 AccAL20 DmCAL AccAL20 DmCAL AccAL2 AccAL2 AccAL2 AccAL3 AccAL4 AccAL5 AccAL6 AccAL7 AccAL8 AccAL9 AccAL10 AccA	G-RPSYAN G-EDFRS 10 GDTCNLQW G-WMNNA R G-WMDTA B G-WMDTA B G-WMTA B G-WTA B G	$\begin{array}{c} \text{IR} \text{RDB}(\text{G}\text{VD}) \text{V} \text{OF} \\ \text{IM}(\text{RG}\text{O}\text{P} \mid \text{SD}) \text{V} \text{OF} \text{IS} \\ \text{IM}(\text{RG}\text{O}\text{P} \mid \text{SD}) \text{V} \text{OF} \text{OF} \\ \text{ID}(\text{D}\text{O}\text{V} \cap \text{OF} \mid \text{OF} \text{OF} \\ \text{ID}(\text{S}\text{O}\text{O}\text{P} \mid \text{OF} \text{OF} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{P} \mid \text{SD} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{P} \mid \text{SD} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{P} \mid \text{SD} \mid \text{SD} \text{OF} \\ \text{SPECS} \text{SD}(\text{O}\text{O}\text{S} \mid \text{SD} \text{OF} \\ \text{SPECS} \text{SD}(\text{O}\text{O}\text{S} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{P} \mid \text{SD} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \text{OF} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \text{O} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \text{SD} \ \text{O} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \ \text{SD} \ \text{SD} \ \text{SD} \ \text{SD} \ \text{O} \\ \text{SD}(\text{O}\text{O}\text{O}\text{O}\text{S} \mid \text{SD} \ S$	ANGS LUNKTEKCI LUNKTEKCI LUNKTEKCI LUNKTEKCI LEALIKEOR LUNGY	NKTNLTAKIT KEVGKGIHIASN QRGAVGATIR DPSKIG-ATIR DPSKIG-ATIR DPSKIG-KI KU ENGCD KI SKVANED WIV SKVANED WIV EDDKK KI EDDKK KI EKGQK WIV LHRELP WI LHRELP WI LHRELP WI DKGKU WIV DKGQE WI EGG-QE WI	GTAN FRGDEN YTCDYLVGED G KSYAP-SEA G AD PFGDEK NS GTEWGENG NS GTEWGENG NS GTEWGENG NS GTWGENG NS GTWGENG NS SEINGNNG NS SEINGNNG NS SEINGVNG NS GEPWGENG NS GTWGENG NS STRGENG NS STRGENG	ALEE VISVER ALLA ITRESPU TUESILAYEGPU VERSVITVEPU YIRSVITVEPU YIRSVITVEPU YIRSSNA YIRSSNA YIRSSNA YIRSSNA YIRASSNA YIMA DDIN YIMA DDIN YIMA DDIN YIMA DDIN YIMA DRINN YIMA SKIN	NT DGSLSS AH GG VA DGSLSS AH GG VA DSQKT SF SF SS VA DSKDGD QH KN IG HATSN
AccALl6 AccALl7 AccALl8 AccALl9 AccAL20 DmCAL AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL2 AccAL3 AccAL4 AccAL6 AccAL9 AccAL9 AccAL9 AccAL9 AccAL9 AccAL10 AccAL1	G - R P S YA N G - D F R S YA N G - D F R S N G - WM N N A R G - WM N N A R G - WM D T A R G - WM N C R G - W N R G	$\label{eq:response} \begin{split} & R = D C < D < Q < Q > Q > Q > Q > Q > Q > Q > Q > Q$	Image: Constraint of the second sec	NRTNLTAKIT REVGRGIHLASN QRGAVGATLR QRGAVGATLR DPSKIGATIT DESG-ED WI SKVANED WI VNGQD WI SKVANED WI UNG-QD WI EKGQC WI EKGQC WI LHRRLP WI LHRRLP WI LHRRLP WI LHRG-K WI DEGQC WI SGGCC WI NGGED WI SNGED WI	GTAN FRGDED YTCDYLVGED G KSVAP-SEA G AD PPGDK NS GTMCBKG NS GENGENGEN NS GENGENGEN NS GENGENGEN NS GENGENGEN NS SELWGKG NS SELWGKG NS SELWGKG NS GENGEC NS GENGEC NS GENGEC NS GENGEC S STRNEEC		T DGSLSS AH GG VA DGSQKTBSF SN VA DSQKTBSF SN TG H TSN TG SKDGDDQH KN

图 3 蠋蝽和黑腹果蝇组织蛋白酶 L 多序列比对

-VEDG--OF

-NLDAT

SDCPC

--NING

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。

HPED YDPE

-IQF

FVDN

YHSD

-KEI

NSPE

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, AcCAL20 和 AcCAD7 仅在脂肪体和肠道中表达。 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。

GSYP<mark>T</mark> PTY**VT**



图 4 蠋蝽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, 蠋蝽; 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蠋蝽基因组中 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, 蠋蝽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, 蠋蝽组织蛋白酶 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 蠋蝽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,蠋蝽 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。

蠋蝽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.

Cristofoletti 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 fallou 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, 倪军军. 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. 蠋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. 天敌昆虫蠋蝽的研究进展与展望 [J]. 环境昆虫学报, 2016, 38 (4): 857-865]