

10种昆虫过氧化氢酶基因的全基因组鉴定与特征预测^{*}

张莉娟, 尹华春, 李 勃, 陈 斌

(重庆师范大学 生命科学学院 昆虫与分子生物学研究所 媒介昆虫重庆市重点实验室, 重庆 401331)

摘要:【目的】在全基因组水平上鉴定10种具有代表性昆虫的过氧化氢酶(CAT)基因,并对鉴定所得CAT基因的基本特征和系统发育关系分别进行预测和分析。【方法】在NCBI数据库中下载CAT氨基酸序列作为询问序列,通过本地Blast方法在10种昆虫全基因组水平搜索和鉴定CAT基因并命名。通过生物信息学方法预测10种昆虫CAT家族基因的特征、氨基酸同源序列一致率、保守结构域等。通过MEGA-X软件用最大似然法推断10种昆虫的CAT基因的系统发育。【结果】共鉴定得到23个CAT基因,内华达古白蚁(*Zootermopsis nevadensis*)、东亚飞蝗(*Locusta migratoria*)、人虱(*Pediculus humanus*)、苜蓿蓟马(*Frankliniella occidentalis*)、茶翅蝽(*Halysmoda halys*)、赤拟谷盗(*Tribolium castaneum*)、黑腹果蝇(*Drosophila melanogaster*)、冈比亚按蚊(*Anopheles gambiae*)、家蚕(*Bombyx mori*)和意大利蜜蜂(*Apis mellifera*)基因组上分别有3,1,1,1,4,3,2,1,6和1个CAT基因,均具有全长的转录组序列,编码406~553个氨基酸,氨基酸分子量为46.0~63.7 kDa,等电点为5.21~9.19。鉴定得到的CAT基因共有122个外显子,99个内含子,且外显子与内含子分布模式在物种间差异较大而种内的CAT基因结构相似。有7个CAT基因编码有信号肽,信号肽包含17~22个氨基酸残基。所有CAT基因编码的氨基酸序列的N端具有1个由17个氨基酸组成的CAT近端活性区域,而C端由9个氨基酸组成的CAT血红素配体区域所组成;其中家蚕的BmCat2基因中单独含有1个细菌胞外溶质结合区域。10种昆虫的CAT1基因具有1:1的直系同源关系,家蚕出现特异性扩增独立形成支系,与CAT1支系互为姐妹群。【结论】研究结果提供了10个代表性昆虫CAT基因的基础信息,为进一步开展该基因的功能研究奠定了基础。

关键词:昆虫;过氧化氢酶;全基因组鉴定;特征;系统发育

中图分类号:Q966

文献标志码:A

文章编号:1672-6693(2019)03-0036-08

生物体在逆境胁迫条件下应激产生丰富的活性氧(Reactive oxygen species, ROS),如超氧阴离子自由基、单线态氧、过氧化氢、羟自由基等。ROS在昆虫的先天免疫系统中具有极为重要的作用^[1],它的动态平衡能够激活信号转导^[2]、调控细胞增长和凋亡^[3-4]等不同生理过程,但过量的ROS累积会破坏细胞结构和功能,严重损伤机体内的DNA、蛋白质等大分子物质。

过氧化氢酶(CAT)、超氧化物歧化酶(SOD)、过氧化物酶(POD)等3种抗氧化酶广泛分布于生物体内,由它们组成的保护酶系统可以清除生物体内过剩的ROS,保护生物体免受环境胁迫的危害^[5]。当ROS大量生产时,SOD也被大量诱导生成,催化ROS发生歧化反应生成水和过氧化氢,这是抗氧化作用的第一道防线^[6]。过氧化氢的不断累积会诱导CAT和POD基因表达上升,由此产生大量的CAT和POD以过氧化氢为底物并将它分解为对机体无害的水和氧气,最终消除超量ROS对生物体产生的氧化损伤,反应式为 $2\text{H}_2\text{O}_2 \rightarrow 2\text{H}_2\text{O} + \text{O}_2$ ^[7]。

Thenard于1881年发现CAT,之后Loew于1901年对该酶予以正式命名^[8]。迄今为止,有关酵母(*Aspergillus oryzae*)^[9]、拟南芥(*Arabidopsis thaliana*)^[10]、小鼠(*Mus musculus*)^[11]、智人(*Homo sapiens*)^[12]等生物的CAT基因的研究已有不少。在昆虫方面,Orr等人^[13]于1990年首先克隆了黑腹果蝇(*Drosophila melanogaster*)的CAT基因,为昆虫CAT基因研究的奠定了基础。此后对不同昆虫CAT基因的研究工作逐渐展开,例如在家蚕(*Bombyx mori*)^[14]、葱蝇(*Delia antiqua*)^[15]、意大利蜜蜂(*Apis mellifera*)^[16]、冈比亚按蚊(*Anopheles gambiae*)^[17]、异色瓢虫(*Harmonia axyridis*)^[18]等昆虫的CAT基因的功能和系统发育关系方面已有大量的研

^{*} 收稿日期:2018-11-12 修回日期:2019-03-18 网络出版时间:2019-05-09 19:30

资助项目:“两江学者”计划专项经费;国家自然科学基金(No. 31672363;No. 31372265);国际原子能机构DRP项目(No. 18268);国家科技基础性工作专项重点项目(No. 2015FY210300)

第一作者简介:张莉娟,女,研究方向为生物化学与分子生物学,E-mail:437619634@qq.com;通信作者:陈斌,男,教授,博士生导师,E-mail:bin.chen@cqnu.edu.cn

网络出版地址: <http://kns.cnki.net/kcms/detail/50.1165.N.20190509.1930.038.html>

究。鉴于 CAT 在昆虫的抗氧化作用中起到重要作用,是影响昆虫生长发育和生殖的一大主要因素,因此本研究在农林害虫、媒介医学昆虫以及经济昆虫中一共选出 10 种代表性昆虫作为研究对象,基于 NCBI 开放数据,利用生物信息学分析方法对它们的 CAT 基因家族在全基因组水平上进行了鉴定并进行命名。然后,对这些 CAT 基因编码的氨基酸序列的相似值和保守性、基因的结构、系统发育等方面进行了研究和分析,以便为进一步开拓昆虫中 CAT 功能的研究提供信息基础。

1 材料与方法

1.1 数据来源

10 种昆虫的基因组数据均源自 NCBI 的 GeneBank 数据库,它们的学名、隶属目名、种名缩写及组装 ID 号信息如表 1 所示。

表 1 10 种昆虫的基因组数据来源
Tab. 1 Genomic data sources for ten insect species

物种名	物种所在目名	种名缩写	组装 ID 号
内华达古白蚁(<i>Zootermopsis nevadensis</i>)	蜚蠊目(Blattaria)	Zn	GCF_000696155.1
东亚飞蝗(<i>Locusta migratoria</i>)	直翅目(Orthoptera)	Lm	GCA_000516895.1
人虱(<i>Pediculus humanus</i>)	虱目(Anoplura)	Ph	GCF_000006295.1
苜蓿蓟马(<i>Frankliniella occidentalis</i>)	缨翅目(Thysanoptera)	Fo	GCA_000697945.3
茶翅蝽(<i>Halymorpha halys</i>)	半翅目(Hemiptera)	Hh	GCF_000696795.2
赤拟谷盗(<i>Tribolium castaneum</i>)	鞘翅目(Coleoptera)	Tc	GCF_000002335.3
黑腹果蝇(<i>D. melanogaster</i>)	双翅目(Diptera)	Dm	GCF_000001215.4
冈比亚按蚊(<i>A. gambiae</i>)	双翅目(Diptera)	Ag	GCF_000005575.2
家蚕(<i>B. mori</i>)	鳞翅目(Lepidoptera)	Bm	GCF_000151625.1
意大利蜜蜂(<i>A. mellifera</i>)	膜翅目(Hymenoptera)	Am	GCF_003254395.2

1.2 基因鉴定与命名

为鉴定 10 种昆虫的 CAT 基因,从 NCBI 的 GeneBank 数据库中下载已经被鉴定出来 CAT 氨基酸序列作为本地 BLAST 的询问序列 Query。步骤如下:1) 通过 tBlastn 以已知序列作为 Query 搜索本地基因组数据库,阈值设为 1×10^{-5} ,获得可能的 CAT 基因序列。2) 将相似性大于 30% 的序列连同左右翼侧区(500 bp 或以上)输出,取出的基因组序列输入 Fgenesh+ (<http://www.softberry.com/berry.phtml>) 在线软件中进行基因预测。3) 为验证该序列是否为 CAT 的基因序列,将预测所得基因序列所翻译的氨基酸序列在 GenBank 中再次进行在线比对,将具有 CAT 结构域(Pfm 注释号为 PF00199)的基因作为 CAT 基因。

1.3 基因基本特征分析

用在线软件 Protparam(<http://web.expasy.org/protparam/>), TMHMM(<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) 和 SignalP(<http://www.cbs.dtu.dk/services/SignalP/>) 对 10 种昆虫 CAT 氨基酸序列的分子量、分子式、等电点、亚细胞定位、跨膜区和信号肽进行预测。使用 Bioedit 软件对这些 CAT 氨基酸序列进行多重比对和相似值(Identity)统计,将多重比对的结果通过在线软件 CSIC(<http://molevol.cmima.csic.es/castresana/index.html>) 进行不饱和预测,预测结果导入 GeneDoc 2.7.0 软件预测其保守区域。通过在线软件 GSDS(<http://gsds.cbi.pku.edu.cn/>) 预测 10 种昆虫的 CAT 基因结构,包括外显子和内含子的大小以及内含子的相位。通过 CorelDraw 8.0 绘制有关示意图。

1.4 基因系统发育关系

经过在线软件 Gblocks Server(<http://molevol.cmima.csic.es/castresana/index.html>) 计算 10 种昆虫 CAT 氨基酸序列的饱和度,结果导入 MEGA-X 软件计算这些 CAT 氨基酸序列的系统发育模型,基于赤池信息

量准则(Akaike information criterion, AIC)测算出 10 种昆虫的 CAT 氨基酸序列最佳进化模型为 LG+G。使用 MEGA-X^[19] 和 PHYML^[20] 在线网站(<http://www.atgc-montpellier.fr/>)基于这些昆虫的 CAT 氨基酸序列使用最大似然法(Maximum likelihood, ML)构建系统发育关系,进行 1 000 次重复计算,记录各分支 bootstrap 值,推断这些昆虫的系统发育关系。

2 结果

2.1 CAT 基因鉴定情况及所编码蛋白质的理化特性

基于 10 种代表性昆虫全基因组水平上的搜索与鉴定,共获得 23 个 CAT 基因。内华达古白蚁、东亚飞蝗、人虱、苜蓿蓟马、茶翅蝽、赤拟谷盗、黑腹果蝇、冈比亚按蚊、家蚕和意大利蜜蜂 10 个物种分别有 3,1,1,1,4,3,2,1,6 和 1 个 CAT 基因,其中隶属鳞翅目的家蚕 CAT 基因中出现了特异性扩增的现象。根据它们 Catalase 结构域信息及其在基因组上的位置,分别将上述物种的 CAT 基因命名,如表 2 所示。

表 2 10 种代表性昆虫 CAT 基因及所编码蛋白质特征
Tab. 2 Characteristics of CAT genes and their encoded proteins in 10 representative insects

基因名	基因 长度/bp	编码区 长度/bp	编码氨基酸 数目/个	Scaffold 位置分布	编码蛋白质相对 分子质量/kDa	等电点	信号肽在氨基酸 序列上的位置
<i>ZnCat1</i>	16 307	1 527	508	NW_019059701.1: 391 315~407 621	57.9	8.53	
<i>ZnCat2</i>	8 191	1 548	515	NW_019060838.1: 70 545~78 735	58.7	6.50	1~17
<i>ZnCat3</i>	6 073	1 578	525	NW_019060838.1: 62 488~68 560	59.8	5.78	1~20
<i>LmCat</i>	1 524	1 167	507	AVCP_010000197.1: 482~2 007	57.5	7.28	
<i>PhCat</i>	1 615	1 476	491	NW_002987345.1:88 685~90 299	55.9	6.84	
<i>FoCat</i>	11 234	1 563	506	NW_020292229.1: 666 678~ 677 911	57.0	8.41	
<i>HhCat1</i>	26 815	1 557	504	NW_020110941.1:2 978~29 792	56.0	5.21	
<i>HhCat2</i>	53 111	1 625	526	NW_020111069.1: 75 638~128 748	60.4	7.27	1~22
<i>HhCat3</i>	23 796	1 629	542	NW_020110941.1: 32 957~56 752	61.4	9.19	1~20
<i>HhCat4</i>	10 033	1 512	503	NW_020110221.1: 703 064~713 096	56.8	8.68	
<i>TcCat1</i>	9 608	1 506	501	NC_007418.3:11 506 553~11 516 160	56.2	8.21	
<i>TcCat2</i>	3 874	1 551	516	NW_015450233.1:549 758~553 631	58.7	6.23	
<i>TcCat3</i>	3 865	1 539	512	NW_015450233.1: 555 433~559 297	57.9	6.73	
<i>DmCat1</i>	5 113	1 563	506	NT_037436.4:18 822 786~18 827 898	57.1	8.39	
<i>DmCat2</i>	1 584	1 596	506	NT_033779.5:8 705 695~8 707 278	58.0	6.94	
<i>AgCat</i>	8 860	1 557	504	NT_078265.2:5 772 129~5 780 988	56.9	8.40	
<i>BmCat1</i>	2 320	1 566	507	NW_004582015.1:5 278 759~5 281 078	56.9	8.11	
<i>BmCat2</i>	1 949	1 708	553	NW_004582029.1:4 754 147~4 756 095	63.7	8.97	
<i>BmCat3</i>	1 787	1 592	515	NW_004581718.1:151 468~153 254	58.0	7.21	1~17
<i>BmCat4</i>	2 837	1 521	492	NW_004582179.1:10 240~13 076	55.5	8.19	
<i>BmCat5</i>	1 872	1 595	516	NW_004582179.1:23 047~24 918	57.1	8.48	1~17
<i>BmCat6</i>	3 726	1 255	406	NW_004582505.1:4 516~8 241	46.0	7.19	1~17
<i>AmCat</i>	10 555	1 542	513	NC_037643.1:8 659 926~8 670 480	58.0	8.39	

注:基因名中前两个字母代表的物种名见表 1,下同

鉴定所得 CAT 基因均具有蛋白质编码的全长序列,基因长度为 1 524~53 111 bp,编码区长度为 1 167~1 708 bp,编码氨基酸数为 406~553 个;编码蛋白质相对分子质量为 46.0~63.7 kDa,等电点为 5.21~9.19,亚

[illegible]

Fig. 1 The consistency of *CAT* gene with homologous sequence in ten insect species

封二彩图 2 显示, 23 个 CAT 基因共有 122 个外显子, 长度为 14~1 536 bp, 其中 *LmCat*, *PhCat*, *TcCat1*, *TcCat2*, *TcCat3*, *DmCat1*, *DmCat2*, *AgCat*, *BmCat1*, *BmCat2*, *BmCat3*, *BmCat4*, *BmCat5* 和 *BmCat16* 具有 2~3 个外显子, *ZnCat1*, *ZnCat2*, *ZnCat 3*, *FoCat*, *HhCat1*, *HhCat2*, *HhCat3*, *HhCat4* 和 *AmCat* 具有 8~11 个外显子。此外, 23 个 CAT 基因共有 99 个内含子, 长度为 52~21 018 bp。 *HhCat2* 基因中内含子总长度最大, 为 51 659 bp, 该基因序列长度也为最大。 10 种昆虫 CAT 的基因中 0 型、1 型和 2 型相位的内含子分别共有 65, 16 个和 18 个, 约占内含子总数的 66%, 16% 和 18%。

2.3 CAT 家族的保守结构域

2.4 CAT 基因的系统发育关系

利用 MEGA-X 软件以及在线软件 PHYML, 基于 10 种代表性昆虫 *CAT* 基因所编码的氨基酸序列分别构建了系统发育树。两种软件构建的系统发育关系具有相同的进化支系和自展值(bootstrap), 将系统发育支系分为 4 个群(图 4)。结果显示, 在群 I 和群 II 中 10 种昆虫的 *CAT1* 基因具有 1:1 的直系同源关系, bootstrap 值在 80% 以上。家蚕经过物种特异性扩增的 *CAT* 支系群 IV 与 10 种昆虫的 *CAT1* 支系群 II 和群 III 为姐妹群关系, bootstrap 自展值为 85%~100%。

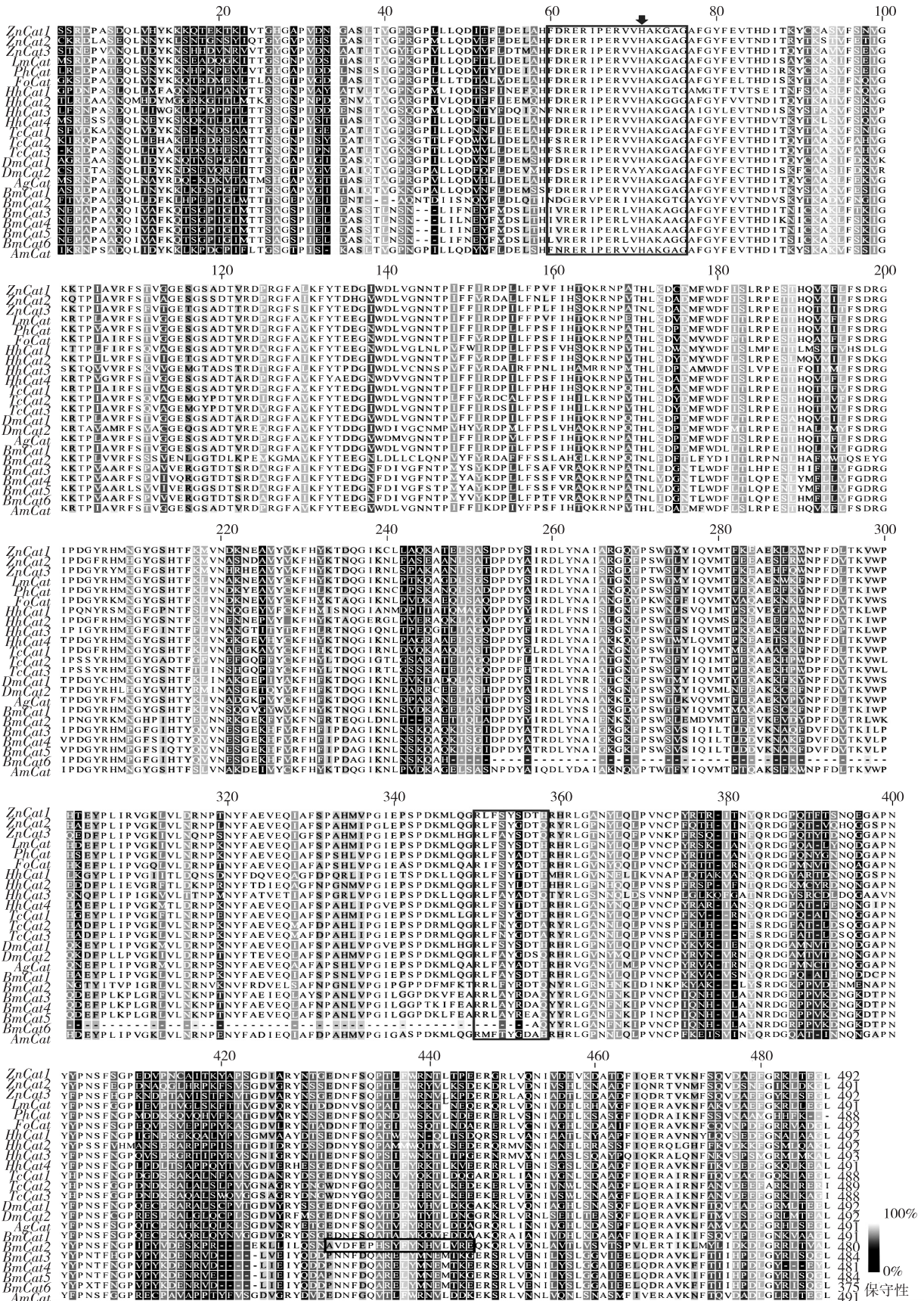


图 3 10 种代表性昆虫 CAT 氨基酸序列的多重比对及特征

Fig. 3 Multiple sequence alignment and characteristics of amino acid sequences of CAT in ten representative insects

在群 I 中, *HhCat4* 由 *HhCat3* 复制而来, *TcCat3* 由 *TcCat2* 复制而来。在群 II 中, *DmCat* 发生了 1 次基因复制事件。在群 III 中, *ZnCat* 基因中发生了 2 次基因复制事件: *ZnCat3* 由 *ZnCat2* 复制而来, *ZnCat2* 由 *ZnCat1* 复制而来。群 IV 是家蚕经过物种特异性扩增的 CAT 枝系, 推测其中包含的 5 个基因可能是由 *BmCat1* 基因复制而来。群 II 与群 III 互为姐妹群, 其中群 II 中的物种分类隶属于鞘翅目、鳞翅目和双翅目, 它们在分类地位上亲缘关系较近; 群 III 中的物种分类隶属于同翅目、膜翅目、直翅目、缨翅目、虱目和蜚蠊目, 在分化时间上出现较早, 推测它们应该面临比较相似的生活环境或生态位。

3 讨论

CAT 在昆虫的保护酶系统中扮演着重要角色, 能清除昆虫体内过剩的 ROS, 从而保护昆虫免受环境胁迫的危害。本研究在涉及农林学、媒介医学及经济类的 10 种代表性昆虫中共鉴定出 23 个 CAT 基因, 其中东亚飞蝗、人虱、苜蓿蓟马、冈比亚按蚊及意大利蜜蜂只含有 1 条 CAT 基因, 内华达古白蚁、茶翅蝽、赤拟谷盗和黑腹果蝇含有 2~4 个 CAT 基因, 而家蚕 CAT 基因存在特异性扩增现象, 因而数目达到 6 个。推测 CAT 基因的扩增可能与昆虫适应自身生活的环境有关^[17,21]。

10 种昆虫 CAT 基因编码的蛋白质相对分子质量介于 46.0~63.7 kDa 之间, 均含有 Catalase, Catalase-rel 结构域各 1 个, 且均不含跨膜域。 *ZnCat2*, *ZnCat3*, *HhCat2*, *HhCat3*, *BmCat3*, *BmCat5* 和 *BmCat6* 共 7 个基因具有信号肽, 信号肽包含 17~22 个氨基酸残基。CAT 是一种稳定性无信号肽的蛋白, 属于非分泌性蛋白^[22], 推测内华达古白蚁、茶翅蝽和家蚕中出现具有信号肽的 CAT 基因可能是它们在环境中受到的氧化应激更多, 为适应环境, 在基因复制过程中融合编码分泌蛋白的信号肽, 以保护体内各组织免受氧化应激反应^[23]。

本研究还分析了 10 个代表性昆虫物种 CAT 基因的系统发育关系, 可以看到 10 种昆虫的 CAT1 基因具有 1:1 的直系同源关系, 物种特异性扩增的家蚕 CAT 基因 (*BmCat2*, *BmCat3*, *BmCat4*, *BmCat5* 和 *BmCat6*) 组成的群 IV 与 10 种昆虫的 CAT1 基因支系群 II 和群 III 为姐妹群关系。其中, 具有特异性扩增的家蚕 CAT 基因中, 具有信号肽序列的 *BmCat3*, *BmCat5*, *BmCat6* 基因和具有细菌的分泌性 CAT 基因 *BmCat2* 没有聚集在一个进化枝上, 因此可以确定含有信号肽序列的家蚕 CAT 基因并不是在测序过程中混入的细菌基因, 推测在 *BmCat2* 中发生了基因的水平转移。此外, *BmCat2*, *BmCat3*, *BmCat4*, *BmCat5*, *BmCat6*, *TcCat2* 和 *TcCat3* 基因中缺失 CAT 血红素配体区域, 这可能是在 CAT 基因复制过程中, CAT 血红素配体区域随之缺失所致, 或许与昆虫特殊环境的适应相关。而意大利蜜蜂与家蚕 CAT 基因数目的差异可能反映了群居物种和独居物种在生活史和生态位上的差异^[16]。

先前的研究表明, CAT 基因在昆虫不同发育阶段和不同组织中都具有较大的差异化表达。如甜菜夜蛾 (*Spodoptera exigua*) CAT 基因的成虫期表达量最高达卵期表达量的 7 倍, 幼虫期表达量次之, 卵期表达量最低; 而在组织上, CAT 基因在 5 龄幼虫体壁、中肠、脂肪体和马氏管组织中都有表达, 其中脂肪体中表达量最高^[24]。白蜡窄吉丁 (*Agrilus planipennis*) CAT 基因在幼虫中肠有着较高表达量; 而在不同发育阶段上, 幼虫和预蛹期有着高表达量, 成虫时期表达量下降^[25]。冈比亚按蚊吸血后, 通过 RNAi 干扰技术沉默 CAT 基因, 成活率明显下降^[17]。上述研究揭示了 CAT 基因在昆虫的生长发育和繁殖等方面都发挥着重要作用。

综合上述, 本研究运用生物信息学方法鉴定了涉及农林学、媒介医学及经济类的 10 种代表性昆虫的 CAT 基因, 并进行了命名, 对基因序列的相似性和保守性、基因的结构、系统发育等方面进行了研究和分析。研究结

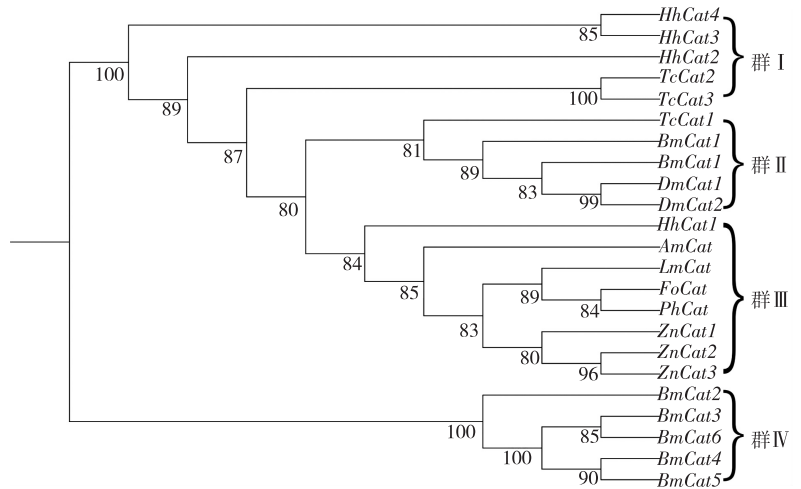


图 4 10 种代表性昆虫基于氨基酸序列的 CAT 基因的系统发育关系

Fig. 4 Phylogenetic relationships of the CAT gene in ten representative insects based on the amino acid sequences

果对进一步研究昆虫的 CAT 基因及相关免疫防御功能具奠定了基础。

参考文献:

- [1] BABIOR B M. Phagocytes and oxidative stress[J]. American Journal of Medicine, 2000, 109(1): 33-44.
- [2] RAY P D, HUANG B W, TSUJI Y. Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling [J]. Cellular Signalling, 2012, 24(5): 981-990.
- [3] STEFFENS B, STEFFEN-HEINS A, SAUTER M. Reactive oxygen species mediate growth and death in submerged plants [J]. Frontiers in Plant Science, 2013, 4(2): 179.
- [4] SIMON H U, HAJ-YEHIA A, LEVI-SCHAFFER F. Role of reactive oxygen species (ROS) in apoptosis induction [J]. Apoptosis, 2000, 5(5): 415-418.
- [5] SCHRADER M, FAHIMI H D. Peroxisomes and oxidative stress[J]. BBA-Molecular Cell Research, 2006, 1763(12): 1755-1766.
- [6] CORONA M, ROBINSON G E. Genes of the antioxidant system of the honey bee: annotation and phylogeny[J]. Insect Molecular Biology, 2010, 15(5): 687-701.
- [7] ZAMOCKY M, FURTMÜLLER P G, OBINGER C. Evolution of catalases from bacteria to humans[J]. Antioxidants & Redox Signaling, 2008, 10(9): 1527-1548.
- [8] 刘冰, 梁婵娟. 生物过氧化氢酶研究进展[J]. 中国农学通报, 2005, 21(5): 223-224.
LIU B, LIANG C J. Recent advances of catalase in organism[J]. Chinese Agricultural Science Bulletin, 2005, 21(5): 223-224.
- [9] HISADA H, HATA Y, KAWATO A, et al. Cloning and expression analysis of two catalase genes from *Aspergillus oryzae*[J]. Journal of Bioscience & Bioengineering, 2005, 99(6): 562-568.
- [10] DU Y Y, WANG P C, CHEN J, et al. Comprehensive functional analysis of the catalase gene family in *Arabidopsis thaliana* [J]. Journal of Integrative Plant Biology, 2008, 50(10): 1318-1326.
- [11] SCHISLER N J, SINGH S M. A quantitative genetic analysis of tissue-specific catalase activity in *Mus musculus* [J]. Biochemical Genetics, 1991, 29(1/2): 65.
- [12] QUAN F, KORNELUK R G, TROPAK M B, et al. Isolation and characterization of the human catalase gene[J]. Nucleic Acids Research, 1986, 14(13): 5321-5335.
- [13] ORR E C, BEWLEY G C, ORR W C. cDNA and deduced amino acid sequence of *Drosophila* catalase[J]. Nucleic Acids Research, 1990, 18(12): 3663.
- [14] YAMAMOTO K, BANNO Y, FUJII H, et al. Catalase from the silkworm, *Bombyx mori*: gene sequence, distribution, and overexpression[J]. Insect Biochemistry & Molecular Biology, 2005, 35(4): 277-283.
- [15] 胡甘雨, 司风玲, 车燕飞, 等. 葱蝇过氧化氢酶基因的克隆及生物信息学分析[J]. 西南大学学报(自然科学版), 2014, 36(1): 32-40.
HU G Y, SI F L, CHE Y F, et al. Cloning and bioinformatic analysis of catalase from onion maggot, *Delia antiqua* [J]. Journal of Southwest University (Natural Science Edition), 2014, 36(1): 32-40.
- [16] WEIRICH G F, COLLINS A M, WILLIAMS V P. Antioxidant enzymes in the honey bee, *Apis mellifera* [J]. Apidologie, 2002, 33(1): 3-14.
- [17] DEJONG R J, MILLER L M, MOLINACRUZ A. Reactive oxygen species detoxification by catalase is a major determinant of fecundity in the mosquito *Anopheles gambiae* [J]. Proceedings of the National Academy of Science of the United States of America, 2007, 104(7): 2122-2126.
- [18] 唐斌, 施佐堃, 郭红双, 等. 异色瓢虫过氧化氢酶(Catalase)的基因克隆及序列分析[J]. 应用昆虫学报, 2014, 51(1): 26-34.
TANG B, SHI Z K, GUO H S, et al. Cloning and sequence analysis of the *Harmonia axyridis* (Coleoptera: Coccinellinae) catalase gene[J]. Chinese Journal of Applied Entomology, 2014, 51(1): 26-34.
- [19] KUMAR S, STECHER G, LI M, et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms[J]. Molecular Biology & Evolution, 2018, 35(6): 1547-1549.
- [20] STÉPHANE G, FRANCK L, PATRICE D, et al. PHYML Online: a web server for fast maximum likelihood-based phylogenetic inference[J]. Nucleic Acids Research, 2005, 33: W557-W559.
- [21] WATERHOUSE R M, KRIVENTSEVA E V, MEISTER S, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes [J]. Science, 2007, 316 (5832): 1738-1743.
- [22] 董书伟, 高昭辉, 申小云, 等. 纳米铜对大鼠肝脏毒性相关蛋白过氧化氢酶的分离鉴定及生物信息学分析[J]. 中国农业科学, 2012, 45(14): 2981-2990.
DONG S W, GAO Z H, SHEN X Y, et al. Isolation, identification and bioinformatics analysis of CAT protein related with hepatotoxicity by copper nanoparticles in rats [J]. Scientia Agricultura Sinica, 2012, 45(14): 2981-2990.
- [23] 时桂芹. 家蚕抗氧化基因注释及功能研究[D]. 重庆: 重庆大学, 2013.
SHI G Q. Annotation and study on the function of antioxidant genes in silkworm, *Bombyx mori* [D]. Chongqing: Chongqing University, 2013.

- [24] 胡振,左洪亮,李亚楠,等. 甜菜夜蛾过氧化氢酶 cDNA 序列克隆、序列分析和表达特征[J]. 昆虫学报, 2011, 54(11): 1249-1257.
- HU Z, ZUO H L, LI Y N, et al. Cloning, sequence analysis and expression profiling of cDNA coding for catalase from the beet armyworm, *Spodoptera exigua* (Lepidoptera;

Noctuidae)[J]. Acta Entomologica Sinica, 2011, 54(11): 1249-1257.

- [25] RAJARAPU S P, MAMIDALA P, HERMS D A, et al. Antioxidant genes of the emerald ash borer (*Agrilus planipennis*): gene characterization and expression profiles [J]. Journal of Insect Physiology, 2011, 57(6): 819-824.

Animal Sciences

Genome-Wide Identification and Prediction of Characteristics on Catalase Genes in Ten Insect Species

ZHANG Lijuan, YIN Huachun, LI Bo, CHEN Bin

(Chongqing Key Laboratory of Vector Insects, Institute of Entomology and Molecular Biology, College of Life Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] The catalase genes of 10 representative insects were identified at the genome-wide level, and the basic characteristics and phylogenetic relationships of CAT genes identified were predicted and analyzed respectively. [Methods] The CAT amino acid sequences were downloaded from NCBI database as query sequences, and the CAT genes were searched and identified at the genome-wide level of 10 insects by local blast method. The bioinformatics methods were used to predict the characteristics, the consistency of amino acid homologous sequence and conserved domains of CAT family genes of 10 insect species. The phylogeny of CAT genes in 10 insect species was inferred by maximum likelihood method with MEGA-X software. [Findings] A total of 23 CAT genes were identified, the genomes of *Zootermopsis nevadensis*, *Locusta migratoria*, *Pediculus humanus*, *Frankliniella occidentalis*, *Halyomorpha halys*, *Tribolium castaneum*, *Drosophila melanogaster*, *Anopheles gambiae*, *Bombyx mori* and *Apis mellifera* have 3, 1, 1, 1, 4, 3, 2, 1, 6, and 1 CAT genes, respectively, all of which have full-length transcriptional sequence encoding 406~553 amino acids, with the molecular weight of 46.0~63.7 kDa and isoelectric point of 5.21~9.19. The identified CAT gene contain 122 exons and 99 introns, and the distribution patterns of exons and introns were significantly different among species, while the CAT gene structure within species was similar. Seven CAT genes have signal peptides containing 17~22 amino acid residues. The N-terminal of the amino acid sequence encoded by CAT genes has an CAT active domain composed of 17 amino acids, while the C-terminal is composed of a CAT heme-ligand domain composed of 9 amino acids and the *BmCat2* gene of *B. mori* contains a bacterial extracellular solute binding domain. *CAT1* gene of 10 insects have 1:1 orthologous relationship, and the *B. mori* shows specific expansion to form a branch independently, which is a sister group with the *CAT1* branch. [Conclusions] The results provide the basic information of CAT genes in 10 representative insect, and lay a foundation for further study on the function of this gene.

Keywords: insect; catalase; genome-wide identification; characteristics; phylogeny

(责任编辑 方 兴)

(接正文39页)

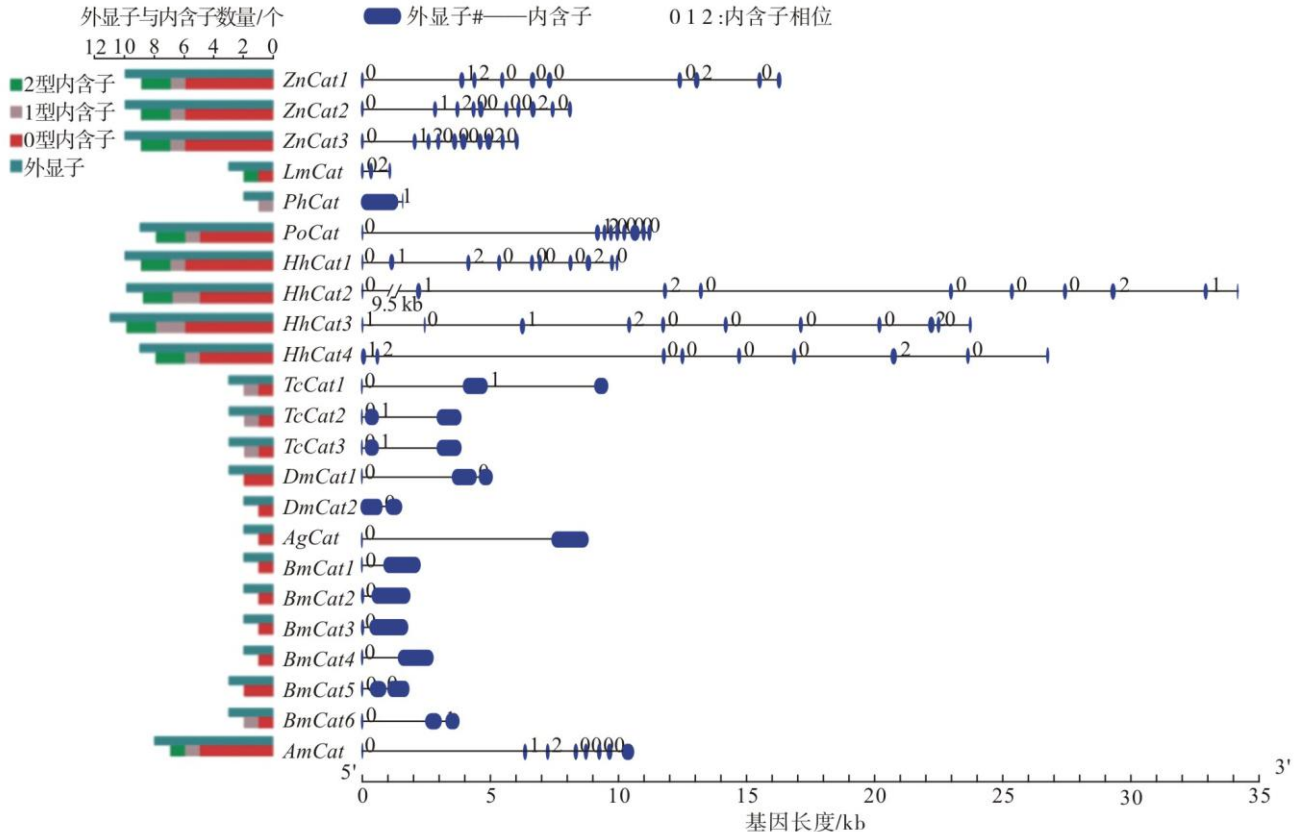
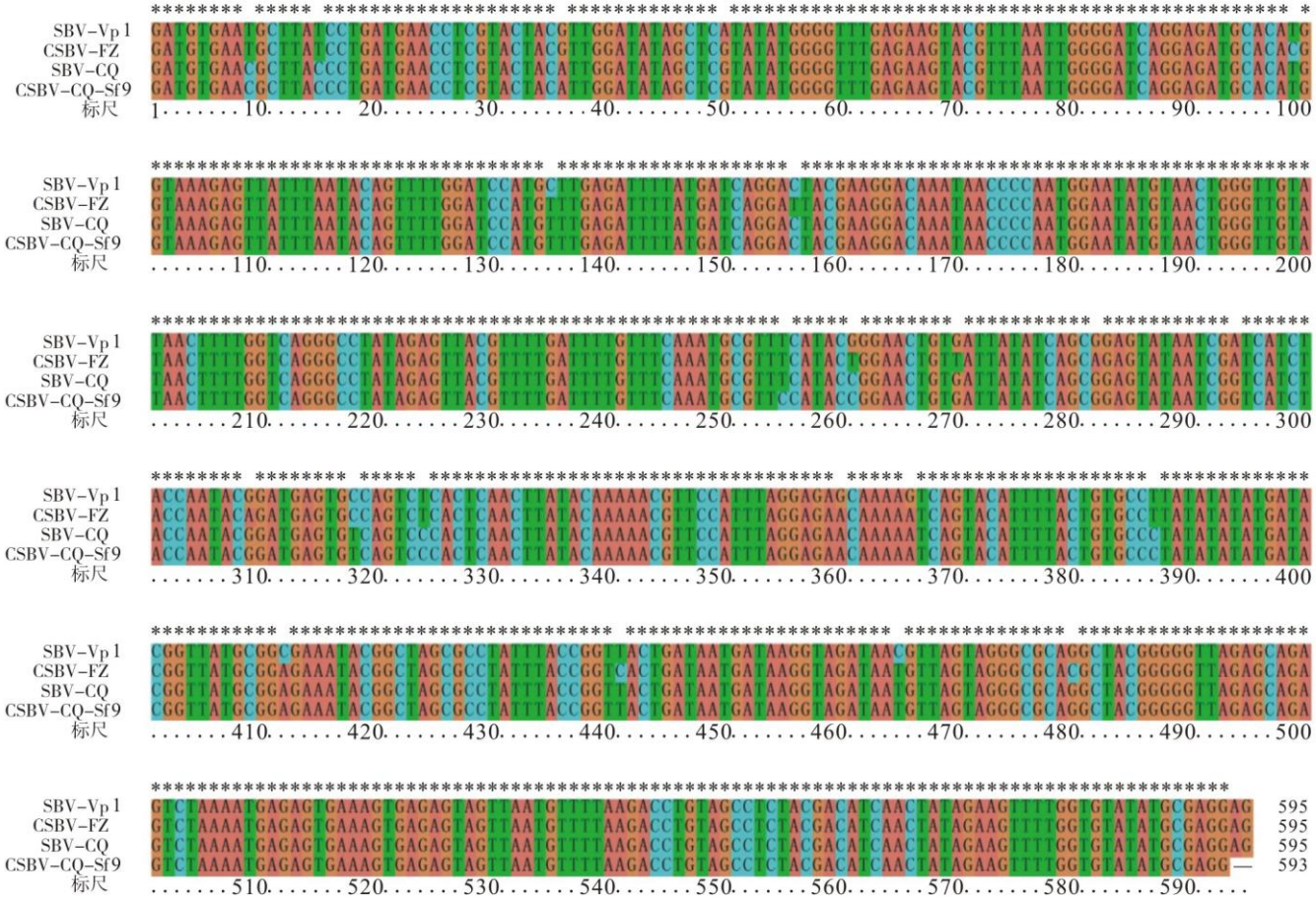


图2 10种昆虫*CAT*基因的结构及内含子相位
Fig. 2 Structure of *CAT* gene and intron phase in ten insect species

(接正文120页)



注: SBV-Vp1为分离自西方蜜蜂体内SBV的*vp1*基因, NCBI登录号为AF251124; CSBV-FZ为分离自中国福州中华蜜蜂体内SBV的*vp1*基因, NCBI登录号为KM495267; SBV-CQ为分离自中国重庆中华蜜蜂体内SBV的*vp1*基因, NCBI登录号为KJ716806; CSBV-CQ-Sf9为本研究感染Sf9细胞的CSBV重庆分离株的*vp1*基因序列; 碱基上方的星号代表相同的碱基

图5 CSBV的*Vp1*基因多重序列比对
Fig. 5 Sequence alignments of CSBV *Vp1* gene