

扶桑绵粉蚧钙调蛋白基因的克隆与生物信息学分析

罗梅 董章勇 宾淑英 林进添 吴仲真 李献锋

仲恺农业工程学院植物保护系/外来有害生物预警与控制研究所, 广州 510225

摘要 钙调蛋白(calmodulin, CaM)对生物体内多种 Ca^{2+} 依赖的细胞功能和酶体系都有重要的调节作用。为研究扶桑绵粉蚧的信号转导受体蛋白,首次克隆了扶桑绵粉蚧钙调蛋白基因 *PsCaM* 的 cDNA 全长序列,其开放阅读框(ORF)包含 447 bp 的片段,编码 148 个氨基酸。*PsCaM* 基因由 3 个内含子和 4 个外显子组成。3 个内含子的长度分别为 73、81、72 bp,分隔的 4 个外显子的长度分别为 33、133、183、98 bp。功能域分析结果显示:该蛋白具有 2 个 EF-hand 结构域,有 13 个 Ca^{2+} 结合位点;该蛋白的理论等电点是 6.21,属于稳定蛋白,且没有跨膜区域;通过同源建模获得了其蛋白的三维结构。多序列比较显示,*PsCaM* 基因相对较保守。

关键词 扶桑绵粉蚧; 钙调蛋白; 基因; 克隆; 生物信息学

中图分类号 Q 966; S 685.22 **文献标识码** A **文章编号** 1000-2421(2012)03-0320-05

钙调蛋白(calmodulin, CaM)是广泛存在于各种真核生物细胞中的多功能调节蛋白,为钙离子(Ca^{2+})的受体蛋白,对生物体内多种 Ca^{2+} 依赖的细胞功能和酶体系都有重要的调节作用,且对任何微量的钙都能敏感地捕获,参与的生化反应很多。CaM 是细胞内 Ca^{2+} 信号传导途径中的主要信号转导分子,介导调控由 Ca^{2+} 引起的一系列生理生化反应,参与并调节细胞的增生、分化、运动等基本代谢过程,通过激活下游靶蛋白调控基因转录、信号转导和肌肉收缩等生理过程^[1]。CaM 通常是以 Ca^{2+} -CaM 复合物形式参与调节细胞的各种活动^[2]。

扶桑绵粉蚧(*Phenacoccus solenopsis* Tinsley)隶属半翅目、粉蚧科、绵粉蚧亚科、绵粉蚧属(*Phenacoccus*)。2008 年 6 月在我国广州园林植物扶桑上发现该虫^[3],随后在海南三亚的扶桑上也有发现^[4]。2009 年 2 月,扶桑绵粉蚧正式被列入《中华人民共和国进境植物检疫性有害生物名录》。扶桑绵粉蚧繁殖能力强,能为害棉花、扶桑、向日葵、南瓜、番茄等多种重要经济作物。目前,国内外的研究主要集中在该虫的生物学特性和防治药剂的筛选,对其寄生机制和行为特性研究较少。扶桑绵粉蚧的行为受寄主植物和温度等条件的影响,笔者首次从扶桑绵粉蚧中克隆了钙调蛋白基因(*PsCaM*),并对

基因结构进行分析,同源建模获得了其蛋白的三维结构,旨在为采用分子生物学手段研究扶桑绵粉蚧的寄生机制和行为特性提供科学依据。

1 材料与方法

1.1 供试虫源和主要试剂

扶桑绵粉蚧为笔者所在实验室温室用扶桑植株饲养获得。取样后迅速放入液氮冷冻并保存于 $-70\text{ }^{\circ}\text{C}$ 冰箱备用。

RNA 提取试剂盒购于 Qiagen 公司,3'-RACE 和 5'-RACE 试剂盒购于 Invitrogen 公司, pMD18-T 克隆载体、反转录系统和 ExTaq DNA 聚合酶均购于 TaKaRa 公司,核酸琼脂糖凝胶回收试剂盒购于 Tiangen 公司。

1.2 总 RNA 的提取和反转录

按照 RNA 抽提试剂盒(Qiagen 公司)使用说明提取扶桑绵粉蚧混合种群的总 RNA。经琼脂糖凝胶电泳和紫外分光光度计检测后,按照反转录系统说明书分别以 3'-CDS Primer A 和 5'-CDS Primer A 为引物进行反转录获得第一链 cDNA,以此作为 PCR 模板进行 3'末端序列和 5'末端序列的扩增。

1.3 引物设计与合成

根据 GenBank 登录的昆虫钙调蛋白氨基酸保

收稿日期: 2011-11-12

基金项目: 农业部公益性行业科研专项(201103026)和仲恺农业工程学院科研基金项目(G3100004)

罗梅, 硕士, 讲师, 研究方向: 入侵生物. E-mail: 08luomei@163.com

通讯作者: 林进添, 博士, 教授, 研究方向: 昆虫行为调控. E-mail: linjtian@163.com

守序列,设计 2 对简并引物(CALF1 和 CALR1, CALF2 和 CALR2)进行 PCR 扩增,用于克隆扶桑绵粉蚧钙调蛋白基因的 cDNA 片段。根据 PCR 测序结果设计 5'-RACE(PsCR51、PsCR52 和 UPM、NUP)和 3'-RACE(PsCR3 和 UPM 及 NUP)反应引物(表 1),分别用于基因的 5'端和 3'端扩增。再根据序列拼接结果设计 1 对特异性引物(F1 和 R1)用于扩增扶桑绵粉蚧钙调蛋白基因全长阅读框,引物在广州英骏生物技术有限公司合成。

表 1 供试 PCR 和 RACE 反应的引物

引物 Primer	引物序列 Primer sequence
CALF1	5'-ACSTAWVGSTCARATHCG-3'
CALR1	5'-ATYDTCAWACDTAAYHAC-3'
CALF2	5'-AGTWSAAGAWATANCDCA-3'
CALR2	5'-TCCWCTTRRTTSCDCTC-3'
PsCR51	5'-CAGTTGAGTAACAAATGTGCAAGAT-3'
PsCR52	5'-CAACTTGACTATGAGTATGCATTACT-3'
PsCR3	5'-GTTGACGCATTTTAAAGCAAGT-3'
F1	5'-ATGGCTCATTATTTTCAGAG-3'
R2	5'-ATGGATGAGAAGAACGATA-3'

1.4 PCR 扩增与 RACE 反应

以合成的 cDNA 为模板,加入 10 × ExTaq DNA 聚合酶反应缓冲液 5 μL (含 Mg²⁺),正向和反向引物各 1 μL (10 mol/L),2.5 mmol/L dNTP 4 μL,ExTaq DNA 聚合酶 0.25 μL (5 U/μL),加水至 50 μL,混匀,离心,放入 PCR 仪扩增。PCR 反应程序:94 ℃变性 3 min;接着进行 35 个循环,循环条件为 94 ℃ 30 s,50 ℃ 45 s,72 ℃ 1 min。循环完毕后,72 ℃保温 10 min。RACE 反应参照试剂盒说明书进行。扩增产物用 1.0%琼脂糖凝胶电泳检测,回收目的片段。

1.5 PCR 和 RACE 产物克隆、鉴定及序列测定

RACE 产物经电泳回收纯化后,将回收产物连接至 pMD18-T 克隆载体后转化到感受态大肠杆菌 DH5α 中,经蓝白斑和氨苄筛选,重组质粒经扩大培养,提取质粒 DNA 用于检测。测序工作由广州英骏生物技术有限公司完成。

1.6 基因和蛋白的序列分析

利用 Blast 在 NCBI 网站进行序列相似性比对分析。通过比对,确认外显子、内含子、5'UTR 和 3'UTR 区域,进一步用 Sim4 在线工具进行验证。ORF 预测在 NCBI 上的 ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)上进行预测。蛋白翻译和基序分析及其理化性质预测采用蛋白组学 ExPASy([\[http://www.ch.embnet.org/software/TMPRED_form.html\]\(http://www.ch.embnet.org/software/TMPRED_form.html\)\)在线预测,跨膜预测采用 TMPRED \(\[http://www.ch.embnet.org/software/TMPRED_form.html\]\(http://www.ch.embnet.org/software/TMPRED_form.html\)\)及 TMHMM \(<http://www.cbs.dtu.dk/services/TMHMM-2.0>\)在线预测。利用 Clustal W 在线工具 \(<http://www.ebi.ac.uk/clustalw2>\)及 Clustal X 2.1 进行同源性分析和多序列比对。蛋白的功能域和二级结构分析采用在线工具 PredictProtein \(<http://www.predictprotein.org>\)上进行预测。蛋白质三维结构预测采用同源建模法,使用 SWISS-MODEL \(<http://swissmodel.expasy.org/>\)进行。](http://expasy.org/tools/prot-</p></div><div data-bbox=)

1.7 分子系统的进化分析

采用 Clustal X 2.1 进行多序列比对后,利用 MEGA 4.1 软件中的 Neighbor-Joining 方法,5 000 次重复构建分子系统进化树,并对系统进化关系进行分析。

2 结果与分析

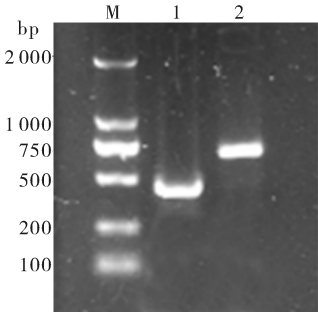
2.1 钙调蛋白基因全长 cDNA 克隆与序列分析

以扶桑绵粉蚧的 cDNA 为模板,以 2 对简并引物(CALF1 和 CALR1,CALF2 和 CALR2)进行巢式扩增后得到 200 bp 左右的特异条带,通过胶回收后连接到 pMD18-T 载体上并进行测序。将所得序列在 NCBI 核酸库中进行 Blast 分析,结果发现该序列与其他昆虫的钙调蛋白基因的相似性最高,推测所获得的片段是扶桑绵粉蚧钙调蛋白基因 cDNA 片段的一部分。根据所得的 cDNA 部分片段设计特异引物(PsCR51、PsCR52 和 PsCR3)分别进行基因 5'端和 3'端序列的扩增,测序拼接后得到扶桑绵粉蚧钙调蛋白基因的 cDNA 全长序列,命名为 *PsCaM*。为验证以上 PCR、3'-RACE 和 5'-RACE 片段是来源于同一个基因,本试验设计了 1 对特异性引物(F1 和 R1),PCR 扩增得到了 1 条约 500 bp 的特异性条带,测序结果与拼接序列一致,为 476 bp。该基因编码具有“AATAAA”的加尾信号。通过 NCBI 上的 ORF Finder 分析,该基因的开放阅读框包含 447 bp 的片段,编码 148 个氨基酸。用 Protparam 工具对该基因编码的蛋白进行预测,其分子质量为 17 392.9 ku;理论等电点为 6.21;分子式为 C₇₈₇H₁₂₂₃N₂₀₇O₂₂₆S₆,共由 2 449 个原子组成;稳定系数为 39.31,推测该蛋白是稳定存在的蛋白。利用 TMHMM-2.0 和 TMPRED 预测该蛋白均没有跨膜区域。

2.2 钙调蛋白基因 DNA 序列的获得与结构分析

用特异性引物 (F1 和 R1) 扩增扶桑绵粉蚧 DNA, 获得约 750 bp 的特异性条带 (图 1), 经克隆测序得到 702 bp 的序列。利用 GSDS (<http://gsds.cbi.pku.edu.cn/>) 构建基因结构图 (图 2), 该基因由 3 个内含子和 4 个外显子组成。内含子的长度分别为 73、81、72 bp; 分隔的 4 个外显子的长度分别为 33、133、183、98 bp。钙离子结合钙调蛋白时可以诱导 EF-hand 的构象发生变化, 导致靶蛋白激活或者失活, 而 EF-hands 通常是以成对出现或者是 2 个以上。

利用 InterPro (<http://www.ebi.ac.uk/interpro/>) 进行功能域分析, 结果显示该蛋白具有 2 个 EF-hand 结构域, 有 13 个 Ca^{2+} 结合位点 (图 2)。



M: DNA 分子质量标准 DNA molecular weight marker; 1: 引物 F1 和 R1 扩增 cDNA 产物 PCR products of *PsCaM*, cDNA as the templates; 2: 引物 F1 和 R1 扩增 DNA 产物 PCR products of *PsCaM*, DNA as the templates.

图 1 扶桑绵粉蚧钙调蛋白基因 cDNA 和 DNA 扩增

Fig. 1 Amplification of *PsCaM* from *Phenacoccus solenopsis*

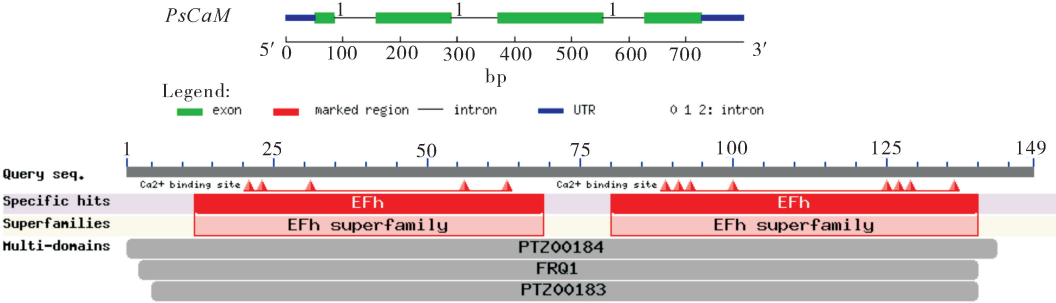


图 2 扶桑绵粉蚧钙调蛋白基因 *PsCaM* 结构和蛋白功能域

Fig. 2 Gene structure and protein functional domains analysis of *PsCaM* from *Phenacoccus solenopsis*

2.3 扶桑绵粉蚧钙调蛋白的二级结构预测

用 PredictProtein (<http://www.predictprotein.org/>) 对该蛋白的二级结构进行预测。149 个氨基酸组成的扶桑绵粉蚧钙网蛋白的二级结构螺旋由 α 螺旋 (alpha helix, Hh) 和 loop 结构 (L) 组成, 其中 Hh 占 59.7%, L 占 40.3%, 该蛋白为全 α 蛋白。该蛋白没有核定位信号 (nuclear localization signals, NLS) 的存在。1 个 N-糖基化位点 (N-glycosylation site), 第 125 位 NVSL; 3 个蛋白激酶 C 磷酸化位点 (protein kinase C phosphorylation site), 分别为第 90、第 97 和第 113 位的 SDK、SSK 和 SGK; 6 个酪蛋白激酶 II 磷酸化位点 (casein kinase II phosphorylation site), 分别为第 28、第 44、第 60、第 72、第 88 和第 113 位的 TLDE、TIFE、TFPE、SQVE、SDSD 和 SGKE; 3 个豆蔻酰位点 (N-myristoylation site), 分别为第 24、第 40 和第 94 位的

GQIRTL、GMSPTI 和 GLISSK。

2.4 钙调蛋白的多序列比对与三级结构预测

通过与其他 4 个昆虫的钙调蛋白序列比对可知, 扶桑绵粉蚧蛋白相对保守, 第 8~21、第 28~44、第 119~125 和第 135~148 个位点的氨基酸极保守 (图 3)。在 PDB 数据库现有的钙调蛋白数据库中, 扶桑绵粉蚧 *PsCaM* 与 PDB 号为 3o78A 的一致性最高, 为 40.816%, 以此为模板对 *PsCaM* 进行同源建模 (图 4)。

2.5 钙调蛋白基因的分子系统进化

分子系统进化分析的结果表明, 选取的 24 个与扶桑绵粉蚧钙调蛋白相似性较高的蛋白中, 可以大致聚为 4 类, 即植物、原生动物、脊椎动物和昆虫 (图 5)。*PsCaM* 与昆虫等节肢动物聚在一起, 与原生动物和植物相距较远, 其中与 *PsCaM* 关系最近的是豌豆蚜 (*Acyrtosiphon pisum*)。

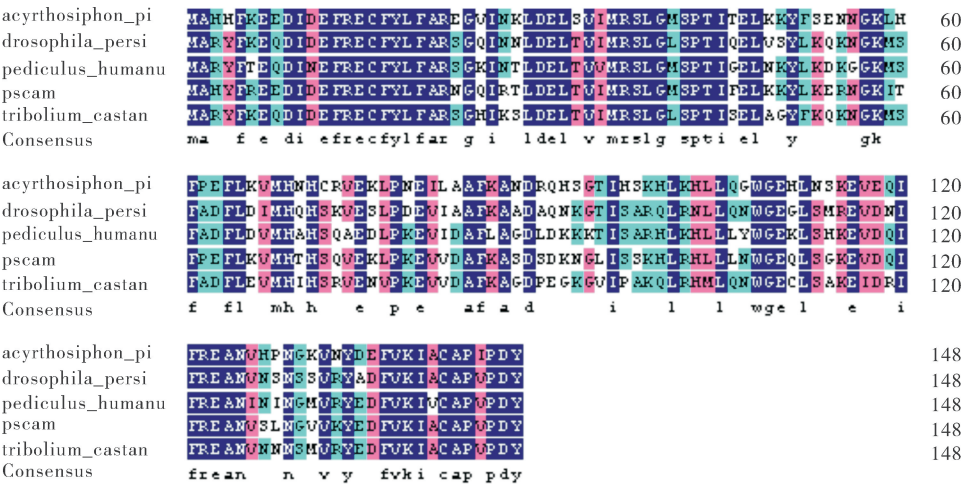


图 3 PsCaM 氨基酸序列同源性比对

Fig. 3 Amino acid sequence alignment of PsCaM with the known CaMs

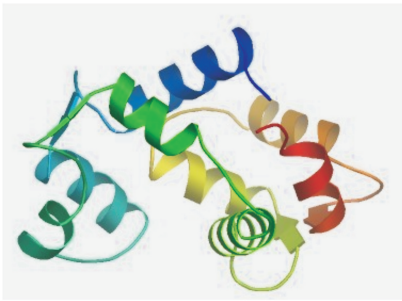


图 4 PsCaM 三级结构预测图

Fig. 4 Tertiary structure prediction map of PsCaM

3 讨 论

钙调蛋白基因序列非常保守,在脊椎动物中已知的 CaM 氨基酸水平完全一致^[1]。推测扶桑绵粉蚧钙调蛋白基因具有 2 个 EF 手结构。EF 手易于成对出现,形成一个分离的区域。配对的手之间可以相互沟通,EF 手之间显示出正向协同效应^[5]。钙调蛋白能随微量钙离子的变化而发生变化。扶桑绵粉蚧钙调蛋白基因具有 13 个 Ca²⁺ 结合位点,这为该基因提供了对钙离子反应敏感的条件。笔者通过

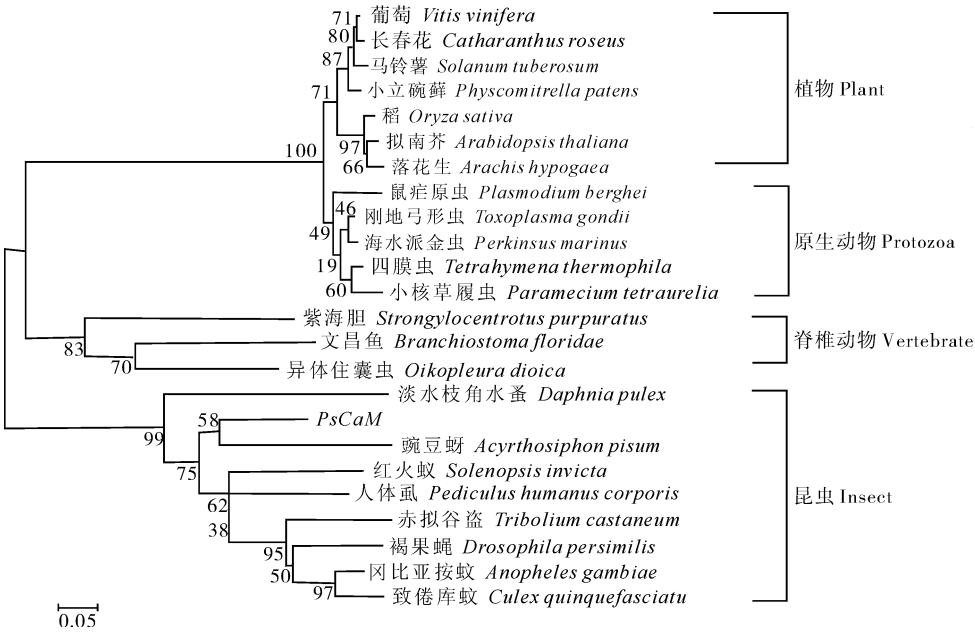


图 5 扶桑绵粉蚧钙调蛋白基因 PsCaM 系统进化树

Fig. 5 Phylogenetic tree of PsCaM from Phenacoccus solenopsis

同源建模获得了扶桑绵粉蚧钙调蛋白三维结构,可为进一步研究其蛋白结构和功能提供科学依据。

CaM 是一种能溶于水的酸性热稳定蛋白质,等电点(PI)约为 4.0。钙调蛋白为多基因家族,在扶桑绵粉蚧中可能存在钙调蛋白的不同亚型。本试验结果表明,扶桑绵粉蚧钙调蛋白是 PI 值约为 6.21 的稳定蛋白。CaM 普遍存在于真核细胞中,结构与功能在进化上高度保守。不同生物体来源的 CaM,其理化性质与生物学性质极为相似。选取与 *PsCaM* 基因同源的蛋白进行比对,结果表明其具有高度的保守性。Yagi 等^[6]认为钙调蛋白分子有 4 个钙离子结合域,分别位于第 8~40 氨基酸、第 44~76 氨基酸、第 81~113 氨基酸和第 117~148 氨基酸之间。这 4 个结构域中,Ⅰ/Ⅲ、Ⅱ/Ⅳ之间有很高的同源率。N 结构域(Ⅰ,Ⅱ)对钙离子浓度变化起反应,C 结构域(Ⅲ,Ⅳ)参与和靶蛋白的结合,只有完整的钙调蛋白分子,才能表达其生理功能。

马燕等^[7]的研究结果表明,DDT 处理使得家蝇(*Musca domestica vicina* Macquart)的环腺苷(cAMP)二次上升,推测其第二次上升可能与钙调蛋白和磷酸二酯酶有关。段江燕等^[8]的研究结果表明,番茄碱对棉铃虫钙调蛋白的影响很大,番茄碱的加入破坏了钙调蛋白-钙离子复合物的刚性,使蛋白

刚性下降。神经毒剂在昆虫的毒杀作用中,对神经系统蛋白质磷酸化的影响可能是一个重要机制,而蛋白磷酸化反应有的是依赖 Ca^{2+} 加钙调蛋白的蛋白激酶进行的。

参 考 文 献

- [1] YUASA H J, SUZUKI T, YAZAWA M. Structural organization of lower marine no nvertebrate calmodulin gene [J]. Gene, 2001, 279: 205.
- [2] SODERLING T R, CHANG B, BRICKEY D. Cellular signaling through multi-functional Ca^{2+} /calmodulin 2 dependent protein kinase II [J]. J Biol Chem, 2001, 276(5): 3719-3722.
- [3] 马骏, 胡学难, 刘海军, 等. 广州扶桑上发现扶桑绵粉蚧[J]. 植物检疫, 2009, 3(2): 35-36.
- [4] 徐卫, 付海滨, 龙琼华, 等. 海南省发现有害生物——扶桑绵粉蚧[J]. 植物检疫, 2009, 5(23): 33.
- [5] GIFFORD J L, WALSH M P, VOGEL H J. Structures and metalion-binding properties of the Ca^{2+} -binding helix-loop-helix EF-hand motifs [J]. Biochem, 2007, 405(2): 199.
- [6] YAGI K, YAZAWA M, KAKIUCHI S, et al. Identification of an activator protein formyosin light chain kinase as the Ca^{2+} -dependent modulator protein [J]. J Biol Chem, 1978, 253(5): 1338-1340.
- [7] 马燕, 张宗炳. 神经毒性杀虫药剂对家蝇头部环腺苷酸含量的影响[J]. 昆虫学报, 1989, 32(4): 393-398.
- [8] 段江燕, 杜黎明, 吴昊, 等. 番茄碱对棉铃虫的毒性作用机理初探[J]. 西北植物学报, 2006, 26(1): 117-120.

Molecular cloning and bioinformatics analysis of calmodulin genes in *Phenacoccus solenopsis* Tinsley

LUO Mei DONG Zhang-yong BIN Shu-ying LIN Jin-tian WU Zhong-zhen LI Xian-feng

Department of Plant Protection, Zhongkai University of Agriculture and Engineering/

Institute for Management of Invasive Alien Species, Guangzhou 510225, China

Abstract Calmodulin(CaM) plays an important regulatory role in a variety of *in vivo* Ca^{2+} -dependent cell functions and enzyme systems. It was the first time to clone the calmodulin gene from *Phenacoccus solenopsis* Tinsley. The full-length of open reading frame (ORF) is 447 bp, encoding 148 amino acid residues. *PsCaM* gene was constituted by three introns and four exons. The intron lengths were 73, 81, 72 bp; The lengths of the 4 separated exons were 33, 133, 183, 98 bp. Functional domain analysis showed that the protein got two EF-hand domains, with 13 Ca^{2+} binding sites. The theoretical protein isoelectric point is 6.21. It was stable protein, with no transmembrane region. The three-dimensional structure of its protein was obtained by homology modeling. Multiple sequence comparison revealed that the gene is relatively conservative. The research was the basis for further study of the functional mechanisms of calmodulin genes.

Key words *Phenacoccus solenopsis* Tinsley; calmodulin; gene; clone; bioinformatic

(责任编辑:陈红叶)