

doi:10.3969/j.issn.1672-5565.2016.04.03

家蝇 *Prohibitin* 基因的生物信息学分析

刘红美^{1*}, 张洁¹, 李玉权¹, 陶小买¹, 龚晓林²

(1. 贵州医科大学生物与工程学院, 贵阳 550004;

2. 贵州医科大学图书馆技术部, 贵阳 550004)

摘要: 为对家蝇 *Prohibitin* 蛋白序列进行生物信息学分析, 从而为该基因功能研究奠定基础。利用在线分析程序和相关工具软件分析 *Prohibitin* 蛋白的理化性质、结构域、并预测其空间结构和功能。结果表明家蝇 *Prohibitin* 蛋白由 277 个氨基酸组成, 分子量为 30.54 kDa, 理论等电点为 5.26, 为稳定蛋白, 有跨膜区, 但不含信号肽, 该蛋白属于 PHB 保守结构域家族, 亚细胞定位于细胞质, 二级结构以 α -螺旋为主。蛋白同源性比对结果显示, 昆虫中的 *Prohibitin* 蛋白具有较高的同源性。这些分析结果可为今后深入研究该蛋白的结构特征和功能提供参考。

关键词: 家蝇; *Prohibitin* 基因; 生物信息学; 进化树

中图分类号: Q343.1 文献标志码: A 文章编号: 1672-5565(2016)04-213-06

Bioinformatics analysis of *Prohibitin* gene of *Musca domestica* (Housefly)

LIU Hongmei^{1*}, ZHANG Jie¹, LI Yuquan¹, TAO Xiaomai¹, GONG Xiaolin²

(1. College of Biology and Engineering, Guizhou Medical University, Guiyang 550004, China;

2. Department of Technology of Library, Guizhou Medical University, Guiyang 550004, China)

Abstract: To perform the bioinformatics analysis of *Prohibitin* gene of *Musca domestica* (Housefly), and aims at providing the basis for its' function research. Internet online procedures and the related software are exploited to analyze the physical and chemical properties of *Prohibitin* protein and the domains of the protein, and predict the structure and functions of the protein. The *Prohibitin* protein sequence is composed of 277 amino acids with 30.54 kDa of the molecular weight. Theoretical isoelectric point is 5.26. The protein is stable and had structures of transmembrane and helical coils. The protein belongs to the family of PHB, and is located in cytoplasm and dominated by alpha helix in second structure. The phylogenetic tree reveals that the *Prohibitin* protein in insect has the highest homology. This work provides a systemic sequence analysis of *Prohibitin* protein of *Musca domestica* (Housefly). It will provide the useful reference for further investigation of its structure and function.

Keywords: *Musca domestica*; *Prohibitin* gene; Bioinformatics; Phylogenetic tree

Prohibitin 基因广泛分布于细菌、植物、酵母及哺乳动物等各种生物细胞中, 具有高度的同源性^[1]。哺乳动物中研究发现 *Prohibitin* (*phb*) 基因具有明显的抗细胞增殖^[2]、抗肿瘤作用^[3], 其表达产物 *Prohibitin* (PHB) 蛋白的氨基酸结构非常保守, 是一类新型的分子伴侣蛋白, 同时具有转录调控作用, 在生物体中发挥着非常重要的作用^[4-5]。但到目前为止, 昆虫中的 *Prohibitin* 基因的生物学功能

研究还未见深入报道。Atichat Kuadkitkan 等人发现 *Prohibitin* 蛋白是登革热 2 型病毒进入昆虫细胞的受体蛋白质^[6]。家蚕中的 *Prohibitin* 基因研究证明, *Prohibitin* 基因在家蚕的不同发育时期, 表达量存在明显差异, 暗示该基因在家蚕的生长发育中有重要的生理功能^[7]。课题组在对热诱导处理的家蝇全长 cDNA 文库中随机 EST 测序和序列分析发现, 其中一候选克隆可能为家蝇 *Prohibitin* 基因^[8]。根据

收稿日期: 2016-06-20; 修回日期: 2019-10-12.

基金项目: 贵州省高层次人才基金项目 (No.876)。

* 作者简介: 刘红美, 女, 博士, 副教授, 研究方向: 分子生物学与生物技术; E-mail: hmliu@gmc.edu.cn.

获得的 EST 序列,本文在已获得 *Prohibitin* 基因全长 cDNA 序列的基础上,应用生物信息学方法对该基因编码蛋白的基本理化性质、跨膜区、亲疏水性、二级及三级结构、功能域、同源性等进行了分析,拟为该基因的深入研究提供参考和理论依据。

1 材料方法

1.1 材料

课题组构建的热诱导家蝇 cDNA 文库和 EST 序列。

1.2 *Prohibitin* 基因全长 cDNA 的获得

挑取 *Prohibitin* 基因 EST 序列对应的阳性菌落进行培养并提取质粒,然后进行双向测序。

1.3 序列分析

采用 ORF 程序 (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) 查找序列的开放阅读框,用 Expasy 服务器上 ProtParam 程序分析蛋白的基本理化性质 (<http://au.expasy.org/tools/protparam.html>)。利用 ProtScale 在线程序^[9]对氨基酸序列进行亲疏水性分析 (<http://au.expasy.org/tools/protscale.html>),在线 TMHMM 2.0 服务器^[10]预测蛋白跨膜区 (<http://www.cbs.dtu.dk/services/TMHMM-2.0>),软件 SignalP4.0^[11]预测蛋白的信号肽 (<http://www.cbs.dtu.dk/services/signalp>),PSORT II 软件预测蛋白的亚细胞定位 (<http://psort.nibb.ac.jp/form2.html>),SOPMA 软件进行蛋白的二级结构预测 (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)。

结构域采用 NCBI 网站上的 conserved domains search service 在线分析工具预测 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)。蛋白质三级结构运用 Phyre2 程序进行 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)^[12],利用软件 MEGA7.0.14 进行氨基酸的多重序列比对和系统进化树的绘制^[13]。

2 结果分析

2.1 *Prohibitin* 全长 cDNA 的获得

挑取 *Prohibitin* 基因的候选克隆进行质粒提取和双向测序,测序结果经 Genbank 数据库进行 Blast 比对分析,确为 *Prohibitin* 基因的全长 cDNA 序列。

2.2 序列分析

应用 NCBI 的 ORF Finder 程序对全长 cDNA 序

列进行开放读码框的查找。结果显示,*Prohibitin* 基因开放读码框为 834 bp,共编码 277 个氨基酸,序列已提交到 GenBank(登录号为:ADT92002.1)。

具体氨基酸序列如下:

```
maaqqffnrigqglgvalvggvnsalynvdgghravifdrftgknevtgeg
thffipwvqrpriydirsqpmvvtgskdlqnvnitrlilyrpidqlpriyti
lgqdydervlpsiapevlkavvaqfdagelitqreivsqrvsdelterakqfghi
liddisithltfgreftqavemkvaqqaekarfvvekaeqqklaaiisaegd
aaaaellaksfaeagdgllvelrriaaediayqlsrsgvaylpgnqstllnlp
ntlaq
```

2.3 *Prohibitin* 蛋白基本理化性质

使用 ProtParam 程序预测该蛋白的分子式为 $C_{1366}H_{2182}N_{380}O_{409}S_2$,相对分子质量为 30.54 kDa,理论等电点为 5.26,不稳定参数为 38.39,根据不稳定参数的数值在 40 以下才是稳定蛋白的标准^[6],可推定 *Prohibitin* 基因编码的蛋白是稳定蛋白。氨基酸组成中 ALA 谷氨酸占最大比例(10.8%),带正电荷的残基(ARG+GLU)总数为 28 个,带负电荷的残基(ASP+GLU)总数 34 个。

2.4 *Prohibitin* 蛋白的结构域

根据蛋白亲疏水性分析结果(见图 1),该蛋白亲水性较强,为亲水蛋白。利用软件 TMHMM 软件预测得知此蛋白为跨膜蛋白,跨膜区域为 9-28(见图 2);图 3 表明该蛋白序列无信号肽,不是分泌性蛋白。使用 PSORTII 软件进行亚细胞定位分析表明,该蛋白定位于细胞质的可能性为 56.5%,定位于线粒体、细胞核的可能性分别为 13.0%。NCBI 网站的 conserved domains search service 在线分析工具预测结果显示,氨基酸在 26-187 区域含有一个 PHB 结构域。该蛋白具有保守的 PHB 结构域(见图 4)。

2.5 蛋白的结构预测

蛋白的 SOPMA 软件预测结果如图 5 所示,*Prohibitin* 蛋白二级结构元件主要以 α -螺旋(Hh)和延伸链结构(Ee)为主(分别占 45.49%,23.10%),其次是无规则卷曲(Cc)为 22.38%,最少的为 β -转角结构(Tt)为 9.03%。

2.6 家蝇 *Prohibitin* 蛋白的三级结构

利用在线分析工具 Phyre 2 对家蝇 *Prohibitin* 蛋白氨基酸序列进行三级结构预测。结果见图 6,结果显示,以 c3bk6C 结构为模板,*Prohibitin* 蛋白 277 个氨基酸序列中有 156 个氨基酸与模板达到 99.9% 的可信度。

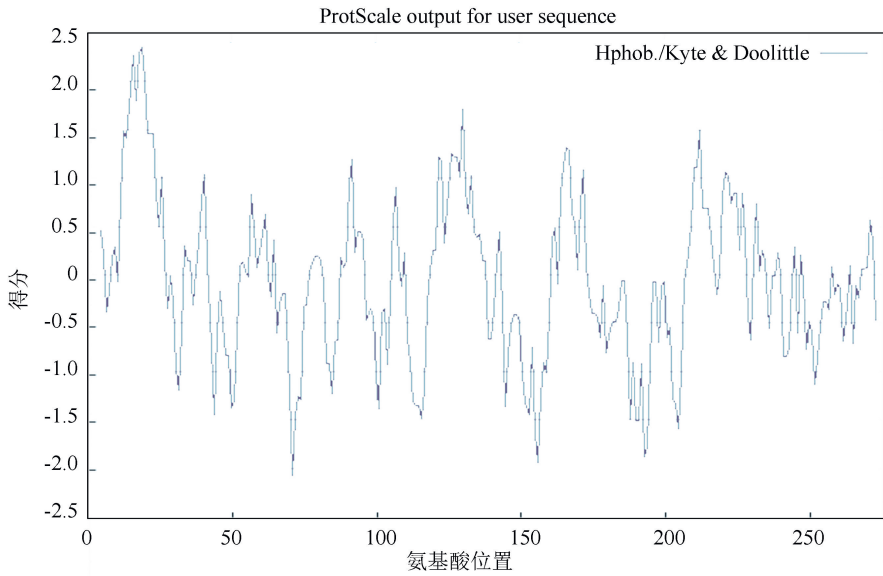


图 1 Prohibitin 蛋白亲疏水性分析

Fig.1 Hydrophobicity/hydrophilicity analysis of prohibitin protein

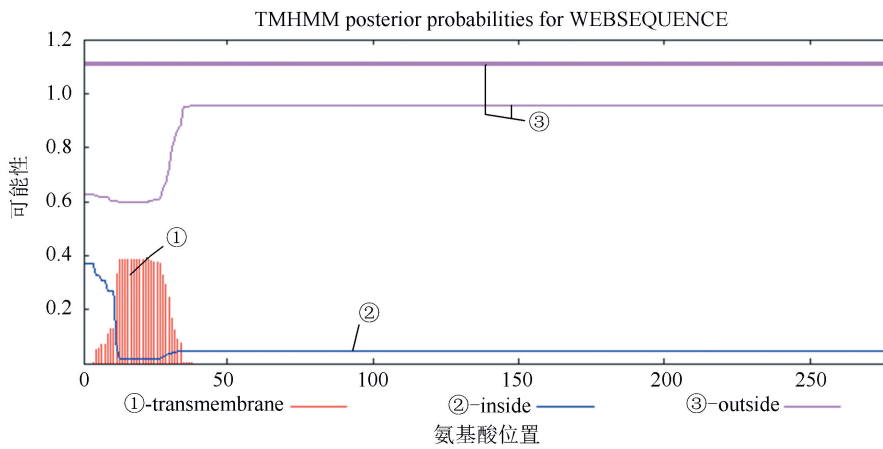


图 2 Prohibitin 蛋白跨膜区预测

Fig.2 Transmembrane region prediction of Prohibitin protein

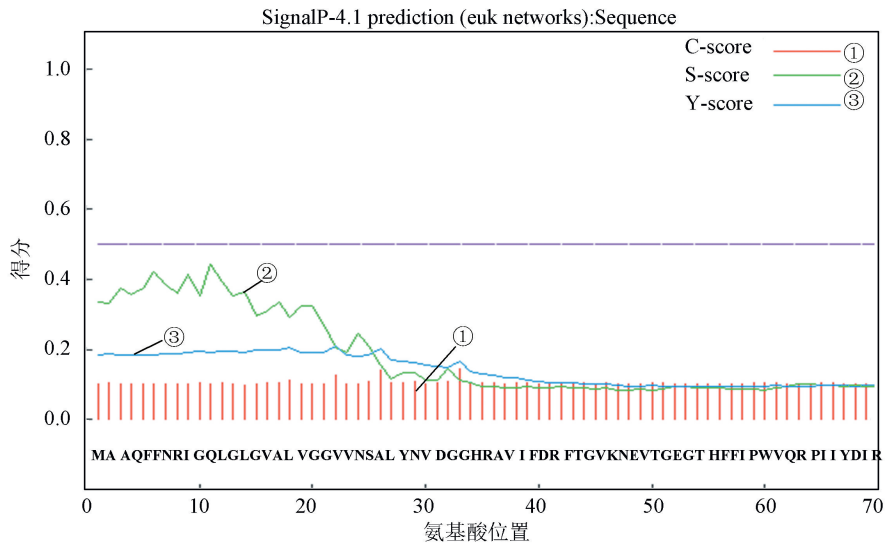


图 3 Prohibitin 蛋白信号肽预测

Fig.3 Signal peptide prediction of Prohibitin

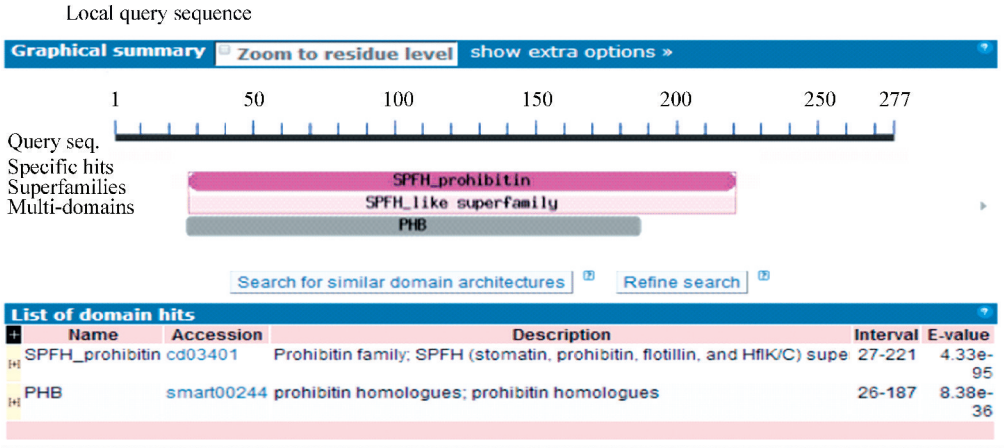


图 4 Prohibitin 蛋白功能域分析

Fig.4 Functional domain analysis of Prohibitin protein



图 5 Prohibitin 蛋白二级结构预测

Fig.5 Secondary structure prediction of Prohibitin protein

注:h;α-螺旋;e;延伸链;c;无规则卷曲;t;β转角

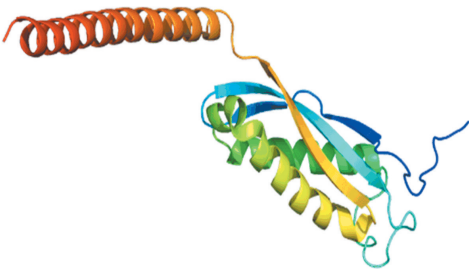


图 6 预测的 Prohibitin 蛋白三级结构

Fig.6 The tertiary structure prediction of Prohibitin

2.7 同源序列比对和系统发育

通过 Blastp 在线同源性比对, Prohibitin 蛋白与其他物种同源蛋白序列相似性都比较高,与厩螫蝇 (*Stomoxys calcitrans*) protein I(2)37Cc 蛋白序列相似

度为 97%,与美洲鸚鵡 (*Amazona aestiva*) 的 prohibitin isoform X1 蛋白序列相似度为 76%。利用 DNAMAN 本地版软件将家蝇 Prohibitin 蛋白序列与其他昆虫相似蛋白进行同源比对,结果显示(见图 7),昆虫中的 Prohibitin 蛋白具有较高相似性。

在 NCBI 数据库选择厩螫蝇、黑腹果蝇、埃及伊蚊、致倦库蚊、家蚕、丽金小蜂、大蜜蜂等 12 种昆虫 Prohibitin 同源蛋白序列,以人类 Prohibitin 蛋白为外群,构建系统进化树。系统发育分析发现,12 种昆虫中的 Prohibitin 蛋白可分为两大类,其中以家蝇、厩螫蝇、果蝇、埃及伊蚊、致倦库蚊等双翅目昆虫中的 Prohibitin 蛋白聚为一类,而另一大类则包含膜翅目昆虫亚类和鳞翅目昆虫亚类(见图 8)。

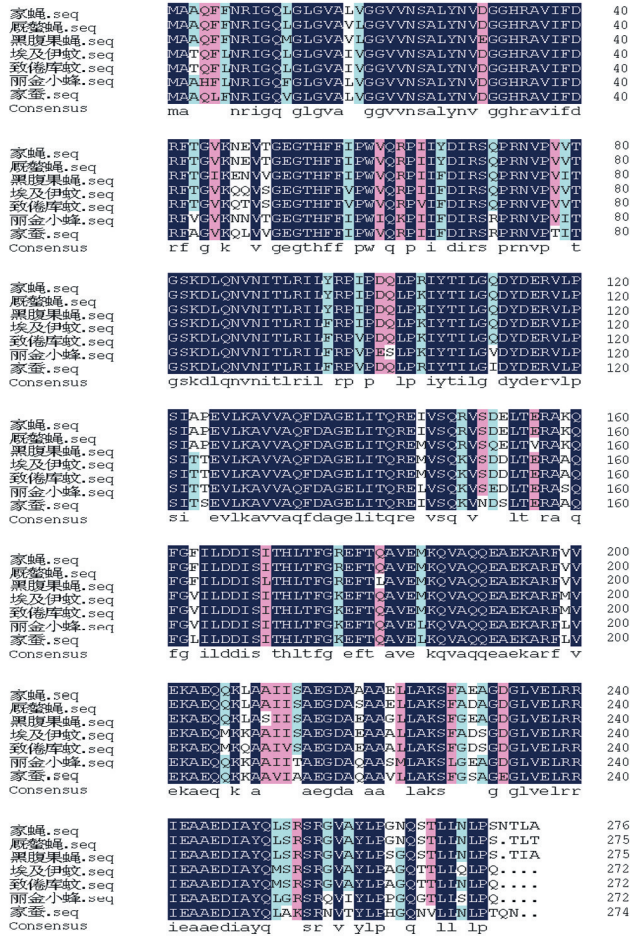


图 7 几种昆虫 Prohibitin 蛋白的多序列比对

Fig.7 Multiple sequence alignment of Prohibitin protein in different insects

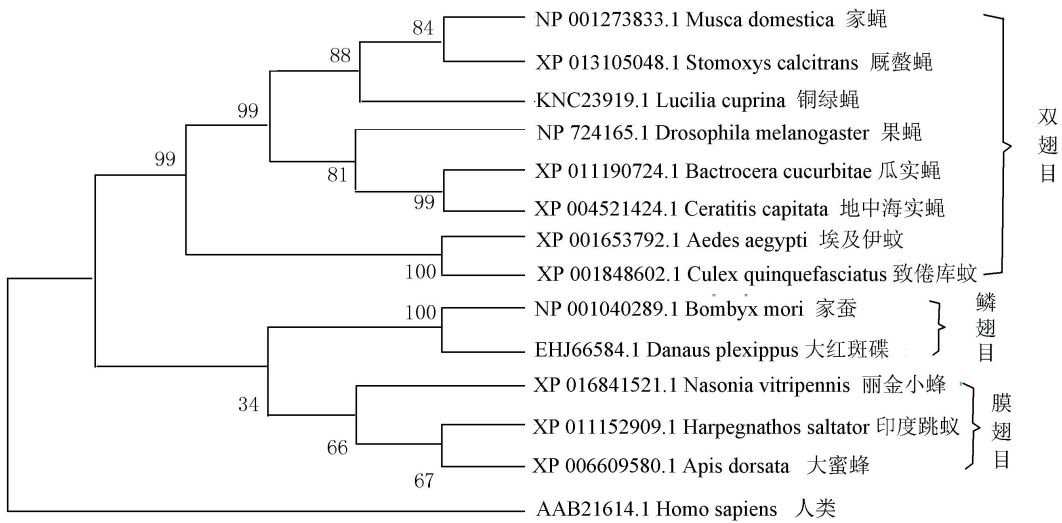


图 8 Prohibitin 蛋白的系统发育分析

Fig.8 Phylogenetic analysis of Prohibitin protein

3 讨论与结论

Prohibitin (PHB) 是一种进化保守分布广泛的

多功能蛋白。然而它具体的功能机制目前还不清楚。为更好的了解 PHB 蛋白的性质,我们从家蝇幼虫 cDNA 文库中克隆了 *Prohibitin* 基因的 cDNA 序列,该基因共编码 277 个氨基酸残基。其编码的蛋

白质分子量 30.54 kDa,理论等电点为 5.26,为稳定蛋白。通过预测,该蛋白具有 PHB 结构域,具有跨膜区,但不含信号肽,亚细胞定位分析显示该蛋白最大可能存在于细胞质中。该蛋白二级结构以 α -螺旋和延伸链为主。同源序列比对及系统发育分析显示该蛋白与其它昆虫的同源蛋白相似性较高。

本研究克隆得到 *Prohibitin* 基因的 cDNA 序列基础上,用生物信息学方法对蛋白序列进行了分析和预测,这些结果将为分子水平上进一步功能研究奠定了基础。

参考文献(References)

- [1] SATO T, SAKAMOTO T, TAKITA KI, et al. The human prohibitin (PHB) gene family and its somatic mutations in human tumors[J]. *Genomics*, 1993, 17(3): 762-764. DOI: 10.1006/geno.1993.1402.
- [2] CHOWDHURY I, THOMAS K, THOMPSON W E. Prohibitin(PHB) roles in granulosa cell physiology[J]. *Cell Tissue Research*, 2016, 363(1): 19-29. DOI: 10.1007/s00441-015-2302-9.
- [3] XU Y, YANG W, SHI J, et al. Prohibitin 1 regulates tumor cell apoptosis via the interaction with X-linked inhibitor of apoptosis protein [J]. *Journal of Molecular Cell Biology*, 2016, 8(3): 282-285. DOI: 10.1093/jmcb/mjw018.
- [4] ZHONG X, SONG X, WANG N, et al. Molecular identification and characterization of prohibitin from *Echinococcus granulosus*[J]. *Parasitol Research*, 2016, 115(2): 897-902. DOI: 10.1007/s00436-015-4846-8.
- [5] ANDE S R, NGUYEN K H, NYOMBA B L, et al. Prohibitin in adipose and immune functions. [J]. *Trends Endocrinol Metab*, 2016, 27(8): 531-541. DOI: 10.1016/j.tem.2016.05.003.
- [6] KUADKITKAN A, WIKAN N, FONGSARAN C, et al. Identification and characterization of prohibitin as a receptor protein mediating DENV-2 entry into insect cells[J]. *Virology*, 2010, 406(1): 149-161. DOI: 10.1016/j.virol.2010.07.015.
- [7] LV Z, ZHANG X, LIU L, et al. Characterization of a gene encoding prohibitin in silkworm, *Bombyx mori*[J]. *Gene*, 2012, 502(2): 118-24. DOI: 10.1016/j.gene.2012.03.035.
- [8] 刘红美, 张洁, 王贇, 等. 热胁迫后家蝇幼虫 cDNA 文库构建与随机 EST 测序分析[J]. *免疫学杂志*, 2010, 26(9): 772-775. DOI: 10.13431/j.cnki.immunol.j.20100183. LIU hongmei, ZHANG jie, WANG yun, et al. Construction of cDN A library from *Musca domestica* L. larvae following heat stress and the EST sequencing [J]. *Immunological Journal*, 2010, 26(9): 772-775. DOI: 10.13431/j.cnki.immunol.j.20100183.
- [9] GASTEIGER E, HOOGLAND C, GATTIKER A, et al. Protein identification and analysis tools on the ExPASy server; (In) John M. Walker (ed): *The proteomics protocols handbook* [M]. New York City: Humana Press, 2005, 571-607.
- [10] OLOF E, SREN B, GUNNAR H, et al. Locating proteins in the cell using TargetP, SignalP, and related tools[J]. *Nature Protocols*, 2007, 2(4): 953-971. DOI: 10.1038/nprot.2007.131.
- [11] THOMAS N P, SOREN B, GUNNAR H, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions[J]. *Nature Methods*, 2011, 8(1): 785-786. DOI: 10.1038/nmeth.1701.
- [12] KELLEY L A, MEZULIS S, YATES C M, et al. The Phyre2 web portal for protein modeling, prediction and analysis[J]. *Nature Protocols*, 2015, 10(6): 845-858. DOI: 10.1038/nprot.2015.053.
- [13] KUMAR S, STECHER G, TAMURA K. MEGA.7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Molecular Biology and Evolution*, 2016, 33(7): 1870-1874. DOI: 10.1093/molbev/msw054.