

# 白蜡虫雄虫真蛹转录组分析

于淑惠, 亓倩, 孙涛, 王雪庆, 杨璞\*, 冯颖

(中国林业科学研究院资源昆虫研究所, 国家林业局资源昆虫培育与利用重点实验室, 云南昆明 650224)

**摘要:** [目的] 蛹期对于白蜡虫雄虫发育和交配具有重要意义, 但目前缺乏对于白蜡虫真蛹基因转录情况的分析, 本研究旨在获得白蜡虫真蛹转录组数据, 了解其转录组特征。[方法] 采用 Illumina HiSeq 2000 进行高通量测序, 并进行生物信息学分析; 采用荧光定量 PCR (qRT-PCR) 检测 3 个热激蛋白基因 (*hsp*) 在不同虫态中的表达动态。[结果] 共获得 63 957 条 unigene, 63 272 个开放阅读框 (ORF), 9 561 个 SSR 分子标记, unigene 平均长度 674 bp, 共有 14 327 条 unigene 获得了注释, Gene Ontology (GO) 注释和 Kyoto Encyclopedia of Genes and Genomes (KEGG) 注释分析表明, 很多基因参与的功能与白蜡虫蛹期的生理特征吻合。[结论] 该转录组数据为白蜡虫基因功能鉴定和蛋白组学分析研究奠定了数据基础。

**关键词:** 白蜡虫; 蛹; 转录组; unigene; 热激蛋白

中图分类号: S899.1 文献标识码: A

## Transcriptome Analysis of Male White Wax Scale Pupae

YU Shu-hui, QI Qian, SUN Tao, WANG Xue-qing, YANG Pu, FENG Ying

(Research Institute of Resources Insects, Chinese Academy of Forestry, Key laboratory of Cultivating and Utilization of Resources Insects of State Forestry Administration, Kunming 650224, Yunnan, China)

**Abstract:** [Objective] To obtain the transcriptomic data and characterize the transcriptome of pupal stage of male white wax scale insect, *Ericerus pela*. [Method] The high-throughput sequencing was performed by Illumina HiSeq 2 000, and bioinformatics analysis were carried out subsequently. Three heat shock protein genes (*hsp*s) were selected for expression profile analysis cross different developmental stages using quantitative real-time PCR (qRT-PCR). [Result] The transcriptomic sequencing generated 63 957 unigenes, 63 272 open reading frame (ORF), and 9 561 simple sequence repeat (SSR) markers. The average length of the unigenes was 674 bp. A total of 14 327 unigenes were annotated in different databases. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that the putative function of many genes coincided the physiological characterization of pupal stage. [Conclusion] The data and analysis of the transcriptome could lay a foundation for further gene functional research and proteomic analysis of *E. pela*.

**Keywords:** *Ericerus pela*; pupae; transcriptome; unigene; heat shock protein

白蜡虫是我国特产资源昆虫, 二龄雄虫所分泌的白蜡是一种重要的工业原料。白蜡虫雌雄二型, 雌虫发育经历卵、幼虫、成虫 3 个阶段; 雄虫发育经

历卵、幼虫、前蛹、真蛹、成虫等阶段。雄虫和雌虫最主要的差别是雄虫有蛹期<sup>[1]</sup>。由于白蜡虫雄虫幼虫活动能力很差, 泌蜡特性又限制了雄虫的活

收稿日期: 2016-01-18

基金项目: 云南省应用基础研究重点项目 (2013FA052); 林业公益性行业科研专项 (201304808, 201504302); 国家自然科学基金 (31572337); 国家 863 计划 (2014AA021801)

作者简介: 于淑惠 (1984—), 女, 博士研究生, 主要从事分子生物学、森林保护研究。

\* 通讯作者: 杨璞, 副研究员。

动,蛹期的出现使雄虫可以羽化成为具有飞行能力的成虫,从而获得与雌虫交配的机会,那些具有较强飞行能力和更多活动空间的雄虫将繁殖出更多的后代。因此,蛹期对于白蜡虫的生存繁衍具有重要意义。

作者前期分析了白蜡虫二龄雄虫泌蜡高峰期转录组、雌雄虫发育过程差异表达基因、雄虫蛹期蛋白质组,但是对于雄虫真蛹转录组特征缺少深入的分析。研究表明,在昆虫的蛹期,发生了幼虫组织降解,成虫组织在此基础上发育重建等重要事件,涉及到细胞程序化死亡、细胞增殖、细胞分化等多种分子过程<sup>[2-7]</sup>。本研究对白蜡虫雄虫真蛹转录组进行高通量测序,对转录组进行注释分析,并采用 qRT-PCR 分析了 3 个 *hsp* 基因在不同虫态中的表达动态,为进一步深入了解白蜡虫蛹期形成提供参考。

## 1 材料与方法

### 1.1 供试昆虫

所用白蜡虫为昆明种群,采自中国林业科学研究院资源昆虫研究所(云南昆明)。

### 1.2 方法

1.2.1 RNA 提取及文库构建 从蜡花中挑选真蛹期雄虫 30 头,在解剖镜下检测,用灭菌双蒸水漂洗 3 次,Trizol (Invitrogen, U. S.) 匀浆。RNA 提取按照 Trizol 试剂说明书操作,采用 Agilent 2100 Bioanalyzer (Agilent Technologies) 检测提取 RNA 的质量。mRNA 分离、文库构建参考文献[1, 8-9]。

1.2.2 高通量测序、基因注释和分析 采用 Illumina HiSeq 2000 进行双末端(pair-ends)测序。对获得的原始 reads 进行过滤,去除接头和质量较差的 reads。采用 Trinity 软件进行序列组装,获得转录本序列(transcript),在聚类单元中选取最主要的转录本作为 unigene 序列。

获得的 unigene 使用 Getorf (<http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html>) 软件进行 ORF 预测。对每个基因的长度、深度、RPKM (Reads Per Kilobase per Million mapped reads)、Multi (多位置) 比对 reads 数目等进行统计分析。

将 unigene 序列用 BLAST<sup>[10]</sup> 软件与 Non-Redundant (NR)<sup>[11]</sup>、SwissProt<sup>[12]</sup>、GO<sup>[13]</sup>、cluster of orthologous groups (COG)<sup>[14]</sup>、KEGG<sup>[15]</sup> 数据库比对,获得注释信息 ( $E\text{-value} \leq e\text{-5}$ )。用 MISA<sup>[16]</sup> 对 1 kb 以上的 unigene 做 SSR 分析,参数设置为单碱基的重复数

$\geq 16$ ,双碱基的重复数 $\geq 6$ ,三至六碱基的重复数 $\geq 5$ 。

1.2.3 qRT-PCR 由于在白蜡虫雄虫蛹期转录组和蛋白组中均有大量热激蛋白(*hsp*)表达,故选取 3 个 *hsp* 基因 (*hsp40*、*hsp60*、*hsp90*),以  $\beta\text{-actin}$  基因为内参,采用相对定量的方法分析 3 个 *hsp* 基因在白蜡虫雌雄虫不同虫态中的表达变化。所用引物和具体方法参考文献[17], mRNA 相对含量的计算采用  $2^{-\Delta\Delta Ct}$  法<sup>[18]</sup>。数据分析采用 DPS 软件中的 LSD 法进行差异显著性分析 ( $P < 0.01$ )<sup>[19]</sup>。

## 2 结果与分析

### 2.1 高通量测序和组装

获得 29 014 801 条 reads,总碱基数为 5.86 G, GC(碱基)含量为 41.37%, Q30(碱基正确识别率 99.9%)为 84.91%。样品在当前的测序量条件下均已达到饱和。共组装获得 106 874 个转录本,进一步获得 63 957 条 unigene, unigene 平均长度 674 bp, N50 长度为 1 210 bp,长度在 1 kb 以上的 unigene 有 10 242 条,占 unigene 总数的 16.01% (图 1A)。共获得 63 272 个 unigene 的 ORF(开放阅读框)。

### 2.2 转录组注释

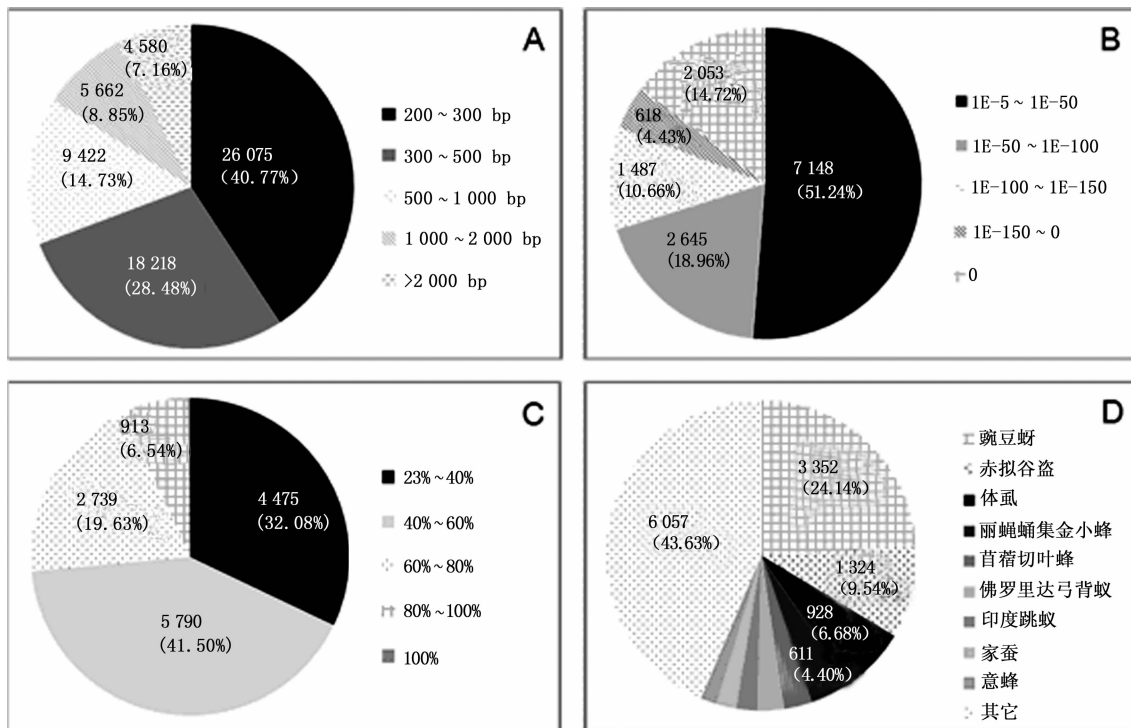
共有 14 327 个 unigene 获得了注释,在 Nr 数据库中注释的 unigene 数量最多,有 13 951 条(表 1),在 Nr 注释结果中,约一半的 unigene 的 E 值小于  $1E\text{-}50$  (图 1B),大多数 unigene 比对相似性在 50% 以上(图 1C),在获得注释的序列中,注释为豌豆蚜(*Acythosiphon pisum* Harris)基因的 unigene 数目高于其它物种(图 1D)。

表 1 unigene 注释统计

注释数据库	unigene 数目(比例/%)
COG	4 517(31.53)
GO	10 507(73.34)
KEGG	4 517(31.53)
Swissprot	11 449(79.91)
Nr	13 951(97.38)

### 2.3 GO、KEGG、COG 注释结果

GO 分析表明:属于结合(87.13%)、催化活性(69.98%)、细胞过程(90.25%)、代谢过程(83.37%)、生物过程(78.69%)等分类的 unigene 最多,还有很多 unigene 属于形态发生素活性(0.02%)、发育过程(72.18%)、细胞增殖(5.12%)、位置建立(50.52%)、色素沉积(1.38%)、表型转换(0.06%) (图 2)。



A: 长度分布; B: E 值分布; C: 相似性分布; D: 物种分布  
图 1 unigene 长度分析及在 Nr 数据库中的注释情况

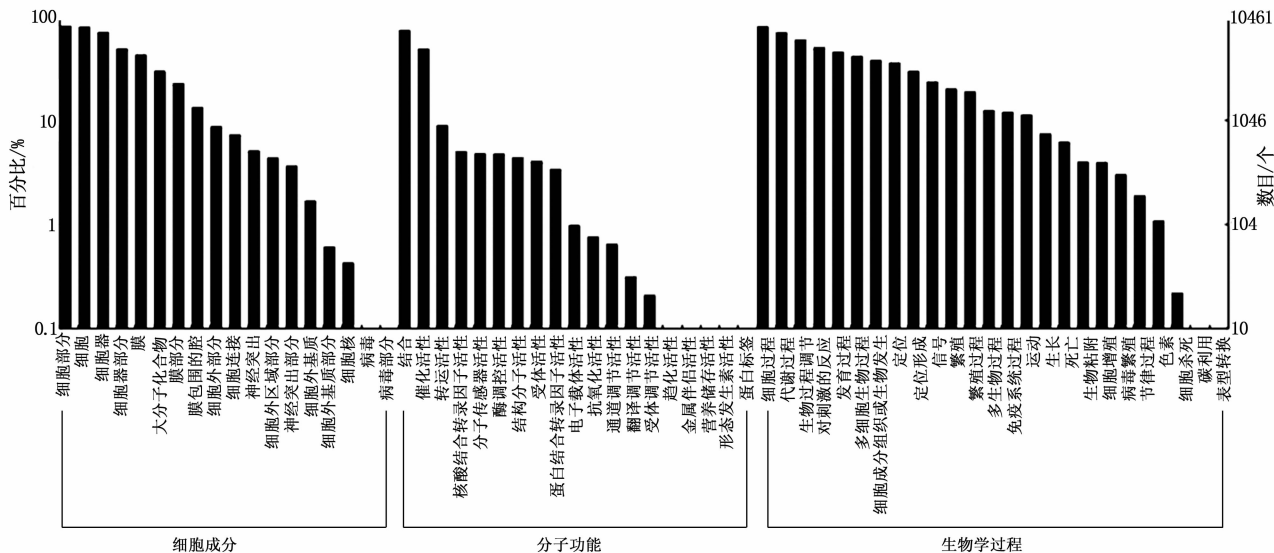


图 2 GO 注释分类图

KEGG 注释结果表明:参与 RNA 转运 (5.25%)、核糖体 (3.81%)、内质网上蛋白加工 (3.65%) 等通路的 unigene 最多,还有很多 unigene 参与内吞作用 (2.48%)、泛素介导的蛋白降解 (2.48%)、Wnt 信号通路 (2.32%) 等。

COG 注释结果表明:除 General function prediction only 之外,属于复制、重组和修复的 unigene

最多。

### 2.4 SSR 分析

共获得 SSR 分析标记 9 561 个。单碱基重复的最多,其次是三碱基重复 (表 2)。

### 2.5 基因表达分析

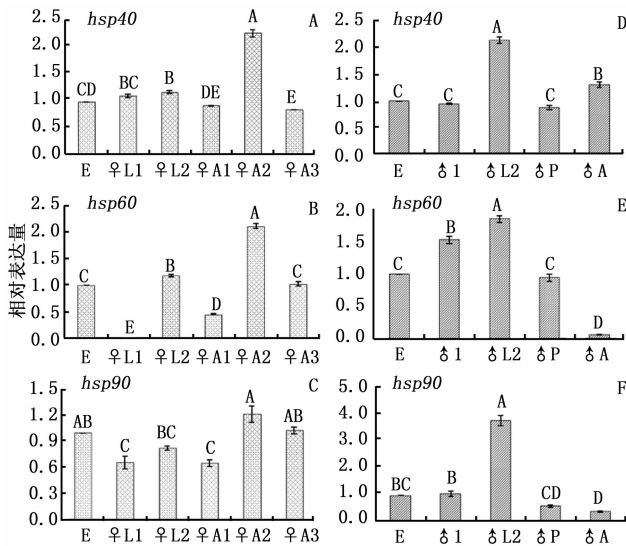
与卵期相比, *hsp40*、*hsp60* 在雌虫的发育中有不同程度地上调表达,在成虫期有明显上调表达;而

表2 SSR分析结果统计

搜索项	数量(比例/%)
检测序列的总数目	10 242
鉴定到的SSR总数目	9 561
含有SSR的序列数目	5 120
含单碱基重复的序列数	7 711(75.29)
含双碱基重复的序列数	538(5.25)
含三碱基重复的序列数	1 190(11.62)
含四碱基重复的序列数	115(1.12)
含五碱基重复的序列数	7(0.07)

*hsp90* 在雌虫的各个阶段均表达很低(图3 A~C)。

*hsp40*、*hsp60*、*hsp90* 在雄虫发育过程中均有不同程度地上调表达,它们的 mRNA 水平在二龄幼虫期达到最大,而在之后的发育阶段(蛹和成虫)中下降(图3 D~F)。



E:卵; ♀ L1:雌虫1龄幼虫; ♀ L2:雌虫2龄幼虫; ♀ A1:雌虫成虫初期; ♀ A2:雌虫成虫中期; ♀ A3:雌虫成虫后期; ♂ L1:雄虫1龄幼虫; ♂ L2:雄虫2龄幼虫; ♂ P:雄虫真蛹; ♂ A:雄虫成虫

图3 *hsp40*、*hsp60*、*hsp90* 在白蜡虫雌雄虫不同虫态中的表达变化

### 3 讨论

本研究共产生了 5.86 G 的数据量,共获得 63 957 条 unigene,样品在当前的测序量条件下均已达到饱和。获得的 unigene 数目多于白蜡虫 2 龄雄虫转录组 unigene,但与雌成虫以及所有虫态混合的转录结果相比,蛹期转录组获得的 unigene 基因数目略少,可能与白蜡虫在不同时期基因表达情况有关。蛹期转录组 unigene 在 1 kb 以上的有 10 242 条,说明组装效果很好<sup>[20]</sup>。注释结果表明,14 327 条 uni-

gene 在各个数据库中获得了注释,从另一方面间接说明了 unigene 的准确性。在 NR 数据库获得注释的 unigene 中,24.14% 的 unigene 注释为豌豆蚜相应的基因,在所注释物种中比例最高,可能是由于这两个物种在进化关系上比较接近的缘故。

GO 注释鉴定到的很多基因与蛹期生理活动相关,如细胞增殖、色素沉积、表型转换等,在 KEGG 注释中,发现很多基因参与蛋白降解、Wnt 信号通路等。这些基因对于成虫特征的形成具有重要意义。比如 Wnt 信号通路在马利筋长蜡(*Oncopeltus fasciatus* Dall) 身体分节中具有保守作用,但对于附肢发育并不起核心作用<sup>[19]</sup>。同时还鉴定到很多基因参与营养储存活性、繁殖,由于白蜡虫成虫期并不取食,交配后死亡,这些基因可能为雄虫交配繁殖过程提供营养<sup>[21]</sup>。

qRT-PCR 检测表明,*hsp40*、*hsp60*、*hsp90* 在雄虫发育过程中均有不同程度地上调表达,2 龄幼虫期达到最大,蛹期和成虫期又下降;而 *hsp90* 在雌虫的各个阶段均表达很低。已有研究表明,*hsp90* 基因参与昆虫的发育过程<sup>[22-26]</sup>,尤其是 *hsp90* 对于蜕皮激素受体 ultraspiracle 复合物的激活具有关键作用<sup>[24,26]</sup>。蛹期蛋白组中也鉴定大量 HSP 蛋白,这些 HSP 在白蜡虫的发育过程可能起到不同的作用,但他们的功能需要通过 RNA 干涉等实验进行验证。

### 4 结论

白蜡虫雄虫真蛹转录组分析获得 63 957 条 unigene,注释分析表明,一些 unigene 的生物学功能与白蜡虫雄虫蛹期的生理活动可能相关,该转录组数据为白蜡虫基因功能研究和蛋白组学分析研究奠定了数据基础。

### 参考文献:

- [1] Yang P, Chen X M, Liu W W, et al. Transcriptome analysis of sexually dimorphic Chinese white wax scale insects reveals key differences in developmental programs and transcription factor expression [J]. Sci Rep, 2015, 5:8141.
- [2] Truman J W, Riddiford L M. The origins of insect metamorphosis [J]. Nature, 1999, 401(6752):447-452.
- [3] White K P, Rifkin S A, Hurban P, et al. Microarray analysis of *Drosophila* development during metamorphosis [J]. Science, 1999, 286(5447):2179-2184.
- [4] Fu Q, Liu P C, Wang J X, et al. Proteomic identification of differentially expressed and phosphorylated proteins in epidermis involved in larval-pupal metamorphosis of *Helicoverpa armigera* [J]. BMC Ge-

- nomics, 2009, 10:600.
- [5] Man'e-Adr'os D, Cruz J, Vilaplana L, *et al.* The hormonal pathway controlling cell death during metamorphosis in a hemimetabolous insect[J]. *Dev Biol*, 2010, 346(1):150–160.
- [6] Gu J, Huang L X, Gong Y J, *et al.* De novo characterization of transcriptome and gene expression dynamics in epidermis during the larval-pupal metamorphosis of common cutworm[J]. *Insect Biochem Mol Biol*, 2013, 43(9):794–808.
- [7] Soares M P, Barchuk A R, Simoes A C, *et al.* Genes involved in thoracic exoskeleton formation during the pupal-to-adult molt in a social insect model, *Apis mellifera* [J]. *BMC Genomics*, 2013, 14:576.
- [8] Yang P, Zhu J Y, Gong Z J, *et al.* Transcriptome analysis of the Chinese white wax scale *Ericerus pela* with focus on genes involved in wax biosynthesis[J]. *PLoS One*, 2012, 7(4):e35719.
- [9] Yu S H, Yang P, Sun T, *et al.* Transcriptomic and proteomic analyses on the supercooling ability and mining of antifreeze proteins of the Chinese white wax scale insect[J]. *Insect Sci*, 2016, 1:1–8.
- [10] Altschul S F, Madden T L, Schäffer A A, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs[J]. *Nucleic Acids Res*, 1997, 25(17):3389–3402.
- [11] Deng Y Y, Li J Q, Wu S F, *et al.* Integrated nr database in protein annotation system and its localization[J]. *Computer Engineering*, 2006, 32(5):71–74.
- [12] Apweiler R, Bairoch A, Wu C H, *et al.* UniProt: the universal protein knowledgebase [J]. *Nucleic Acids Res*, 2004, 32:115–119.
- [13] Ashburner M, Ball C A, Blake J A, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium[J]. *Nat Genet*, 2000, 25(1):25–29.
- [14] Tatusov R L, Galperin M Y, Natale D A, *et al.* The COG database: a tool for genome-scale analysis of protein functions and evolution[J]. *Nucleic Acids Res*, 2000, 28(1):33–36.
- [15] Kanehisa M, Goto S, Kawashima S, *et al.* The KEGG resource for deciphering the genome[J]. *Nucleic Acids Res*, 2004, 32:277–280.
- [16] Arimoto M, Yamagishi K, Wang J, *et al.* Molecular breeding of lignin-degrading brown-rot fungus *Gloeophyllum trabeum* by homologous expression of laccase gene [J]. *AMB Express*, 2015, 5(1):81.
- [17] Liu W W, Yang P, Chen X M, *et al.* Cloning and expression analysis of four heat shock protein genes in *Ericerus pela* (Homoptera: Coccidae) [J]. *J Insect Sci*, 2014, 14(142):1–9.
- [18] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(–Delta Delta C (T)) Method[J]. *Methods*, 2001, 25(4):402–408.
- [19] Angelini D R, Kaufman T C. Functional analyses in the milkweed bug *Oncopeltus fasciatus* (Hemiptera) support a role for Wnt signaling in body segmentation but not appendage development[J]. *Dev Biol*, 2005, 283(2):409–423.
- [20] Zhu J Y, Yang P, Zhang Z, *et al.* Transcriptomic immune response of *Tenebrio molitor* pupae to parasitization by *Scleroderma guani* [J]. *PLoS One*, 2013, 8(1):e54411.
- [21] Yang P, Chen X M. Protein profiles of Chinese white wax scale, *Ericerus pela*, at the male pupal stage by high-throughput proteomics [J]. *Arch Insect Biochem Physiol*, 2014, 87(4):214–233.
- [22] Caplan A J, Jackson S, Smith D. Hsp90 reaches new heights. Conference on the Hsp90 chaperone machine [J]. *EMBO Rep*, 2003, 4(2):126–130.
- [23] Mahroof R, Zhu K Y, Neven L, *et al.* Expression patterns of three heat shock protein 70 genes among developmental stages of the red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae) [J]. *Mol Integr Physiol*, 2005, 141(2):247–256.
- [24] Tachibana S, Numata H, Goto S G. Gene expression of heat-shock proteins (Hsp23, Hsp70 and Hsp90) during and after larval diapause in the blow fly *Lucilia sericata* [J]. *J Insect Physiol*, 2005, 51(6):641–647.
- [25] de Andrade A, Siviero F, Rezende-Teixeira P, *et al.* Molecular characterization of a putative heat shock protein cognate gene in *Rhynchosciara Americana* [J]. *Chromosome Res*, 2009, 17(7):935–945.
- [26] Zhang Q R, Denlinger D L. Molecular characterization of heat shock protein 90, 70 and 70 cognate cDNAs and their expression patterns during thermal stress and pupal diapause in the corn earworm [J]. *J Insect Physiol*, 2010, 56(2):138–150.

(责任编辑:张玲)