Comparative Transcriptome Analysis of Three Bactrocera dorsalis (Diptera: Tephritidae) Organs to Identify Functional Genes in the Male Accessory Glands and Ejaculatory Duct

Authors: Chuan-Bei Tian, Dong Wei, Lin-Fan Xiao, Wei Dou, Huai Liu, et. al.

Source: Florida Entomologist, 100(1) : 42-51

Published By: Florida Entomological Society

URL: https://doi.org/10.1653/024.100.0108
Comparative transcriptome analysis of three Bactrocera dorsalis (Diptera: Tephritidae) organs to identify functional genes in the male accessory glands and ejaculatory duct

Chuan-Bei Tian, Dong Wei, Lin-Fan Xiao, Wei Dou, Huai Liu, and Jin-Jun Wang*

Abstract

The insect male accessory glands/ejaculatory duct (MAG/ED) are important tissues of the male reproductive system. The MAG/ED’s functions in reproduction have been well studied in Drosophila (Diptera: Drosophilidae) but remain largely unknown in the important agricultural pest Bactrocera dorsalis (Hendel) (Diptera: Tephritidae). In the present study, we re-assembled the transcriptome datasets of B. dorsalis’s fat body, testis, and MAG/ED and compared these tissue-specific transcriptomes. Clean reads from these transcriptome data sets were de novo re-assembled and clustered into 31,782 unigenes (average 922 bp). In total, 21,306 unigenes were functionally annotated by Blast against online databases. Comparative transcriptomic analysis identified numerous genes that were identified with the expressed tissue-bias patterns. Some MAG/ED-specific genes potentially involved in spermatozoa motility and capacitation (e.g., perlucina, glucose dehydrogenase, lipase), mating regulation (pheromone-binding protein–related protein), and immunity (lectin) were identified in B. dorsalis. The expressions of some of these genes were further validated by real-time quantitative polymerase chain reaction at transcriptional level. All of these identifications will help us to explore the physiological regulation of mating and reproduction in B. dorsalis in the future.

Key Words: oriental fruit fly; comparative transcriptome; accessory gland protein; seminal fluid protein; reproduction

Resumen

Las glándulas accesorias masculinas/conducto eyaculador (GAM/CE) de los insectos son tejidos importantes del sistema reproductivo masculino. Las funciones de GAM/CE en la reproducción han sido bien estudiadas en Drosophila (Diptera: Drosophilidae), pero siguen siendo poco conocidas en la importante plaga agrícola Bactrocera dorsalis (Hendel) (Diptera: Tephritidae). En el presente estudio, re-ensamblamos los datos de transcriptoma de la grasa del cuerpo, los testículos y los GAM/CE de B. dorsalis y comparamos estos transcriptomas específicos de tejidos. Las lecturas claras de estas conjuntos de datos de transcriptoma se volvieron a ensamblar de nuevo y se agruparon en 31,782 unigenes (promedio de 922 bp). En total, 21,306 unigenes fueron anotados funcionalmente por Blast contra bases de datos en línea. El análisis transcriptómico comparativo identificó numerosos genes que se identificaron con los patrones de sesgo de tejido expresado. En B. dorsalis se identificaron algunos genes específicos de GAM/CE potencialmente implicados en motilidad y capacitation de espermatozoides (por ejemplo, perlucina, deshidrogenasa de glucosa, lipasa), regulación del apareamiento (proteína relacionada con proteína de unión a feronomas) e inmunidad (lectina). Las expresiones de algunos de estos genes fueron aún más validadas por la reacción en cadena de la polimerasa cuantitativa en tiempo real a nivel transcripcional. Todas estas identificaciones nos ayudarán a explorar la regulación fisiológica del apareamiento y la reproducción en B. dorsalis en el futuro.

Palabras Clave: mosca de la fruta oriental; transcriptoma comparativo; proteína accesoria de las glándulas; proteína fluida seminal; reproducción

The male accessory glands (MAGs) are the source of a variety of secreted proteins and peptides (accessory gland proteins, Acps) that belong to a number of different functional categories (Davies & Chapman 2006; Dottorini et al. 2007). The importance of these proteins/peptides in modulating reproductive progress has become more evident with the increased understanding of protein identities and functions (Wolfner 1997, 2002; Gillott 2003). Acps are major components of the seminal fluid proteins that are transferred from male to female together with sperm during copulation. In addition to improving the male’s probability of paternity (Avila et al. 2011), they induce numerous post-mating physiological and behavioral changes in females (Gillott 2003), such as decreasing receptivity of remating (Miyatake et al. 1999; Radhakrishnan & Taylor 2008), affecting sperm storage parameters (Avila et al. 2010; King et al. 2011), increasing egg production and ovulation rate (Ram & Wolfner 2007; Xu & Wang 2011), modulating sperm competition by plug formation (Lung & Wolfner 2001), and changes in feeding behaviors (Lee & Klowden 1999) and longevity (Xu & Wang 2011). Most of these changes in females have been reported

Key Laboratory of Entomology and Pest Control Engineering, College of Plant Protection, Southwest University, Chongqing 400716, China; E-mail: tcxb8023@163.com (C.-B. T.), dongwei1988@yahoo.com (D. W.), a475391353@163.com (L.-F. X.), douwei80@swu.edu.cn (W. D.), liuhuai@swu.edu.cn (H. L.), jiwang7008@yahoo.com (J.-J. W.)

*Corresponding author; E-mail: jiwang7008@yahoo.com (J.-J. W.)

Supplementary material in Florida Entomologist 100(1) (Mar 2017) is online at http://purl.fcla.edu/fcla/entomologist/browse
to result from gene expression changes induced by Acps in females (Rogers et al. 2008; Thalayil et al. 2011). The ejaculatory duct (ED) also secretes proteins that are mixed with Acps, and transferred to the female during mating (Lung et al. 2001).

Recently, various molecular and genetic tools coupled with bioinformatics have been used widely to identify and analyze Acps and the encoding genes in insects (Braswell et al. 2006; Avila et al. 2011), particularly in Drosophila species (Diptera: Drosophilidae). Transcriptome data are being used widely to analyze the transcripts in male reproductive organs of many arthropod species such as Aedes aegypti (L.) (Diptera: Culicidae), Dermacentor variabilis Say (Ixodida: Ixodidae), Lutzomyia longipalpis (Lutz & Neiva) (Diptera: Psychodidae), and Te- leogryllus oceanicus (Le Guillou) (Orthoptera: Gryllidae) (Champagne & Brown 2007; Sonenshine et al. 2011; Azevedo et al. 2012; Bailey et al. 2013).

As one of most economically important insect pests, the oriental fruit fly, Bactrocera dorsalis (Hendel) (Diptera: Tephritidae), is distributed worldwide and has a powerful reproduction (Wan et al. 2012; Wei et al. 2015a). Many transcriptome sequencings were conducted in B. dorsalis to identify functional genes involved in development (Shen et al. 2011), reproduction (Zheng et al. 2012), and insecticide metabolism and resistance (Hsu et al. 2012). Tissue-specific transcriptomes of B. dorsalis midgut, fat body, and testis were characterized and studied in our previous studies (Shen et al. 2013; Yang et al. 2014; Wei et al. 2015b). It is difficult to identify individual Acps among the seminal fluid proteins due to their rapid evolution (Almeida & DeSalle 2008). Therefore, systematic gene expressions studies are limited, although several aspects of the reproductive biology in B. dorsalis have been studied (Ren et al. 2008; Wei et al. 2015a). Recently, the expression of some immune-related genes in MAGs was detected and studied (Lung et al. 2001; Belardinelli et al. 2005; Wong et al. 2008). In a previous transcriptome analysis of B. dorsalis MAGs, we identified many genes involved in immunity in the reproductive duct and several juvenile hormone-related genes (Wei et al. 2016). However, there is still an information gap in understanding the functional molecules and molecular genetics and evolution of Acps encoding genes in B. dorsalis.

In this study, we obtained the transcriptome data of B. dorsalis fat body, testis, and MAGs from the Sequence Read Archive in NCBI, and re-assembled the data. Systematic characterization and interpretation of the data allowed us to identify many MAG-specific genes potentially involved in sperm maturation, viability, mating regulation, and immunity that could advance our understanding of the reproductive biology of B. dorsalis.

Materials and Methods

SOURCES OF TISSUE-SPECIFIC TRANSCRIPTOME DATA

The raw data of B. dorsalis fat body, testis, and MAG/ED transcriptomes were sequenced and can be downloaded from the Sequence Read Archive of the National Center of Biotechnology Information (NCBI). The accession numbers of the datasets were SRR1026844 (Yang et al. 2014), SRR1032039 (Wei et al. 2015b), and SRR1168415 (Wei et al. 2016). All the samples were collected from the same laboratory strain under the same condition. The fat body sample was collected from the midgut, fat body, and testis transcriptomes. Overlapping sequences of specific lengths with no base calls (N) were combined in contigs to form longer fragments. De novo re-assembly of these transcriptomes was carried out with the short read assembling program Trinity based on clean reads (Grabherr et al. 2011). The Trinity program consists of 3 independent software modules: Inchworm, Chrysalis, and Butterfly, which were applied sequentially to process large volumes of RNA-seq reads. All of this process was conducted in the same way as in our previous transcriptome sequencing (Wei et al. 2015b).

The resulting sequences were distinct singletons, and then the sequence clustering software Tgicl was used to splice and remove the redundancy to obtain long, non-redundant unigenes (http://sourceforge.net/projects/tgicl/files/tgicl%20v2.1/). Then, all unigenes were divided into 2 classes, resulting in 1 cluster with the prefix CL followed by cluster ID with similarity more than 70%, and the 2nd cluster with the prefix Unigene followed by cluster ID. In the final step, Blastx alignment was performed between unigenes and proteins in databases such as NCBI nr, Swiss-Prot, and assigned to Kyoto encyclopedia of genes and genomes (KEGG) and clusters of orthologous groups (COG) with an E-value of 10⁻³, and the best alignment results were used to determine sequence direction of unigenes. The protein annotation was retrieved with the highest sequence similarity to unigenes. If results from different databases were conflicting, a priority order of NCBI nr, Swiss-Prot, KEGG, and COG was followed. If a unigene did not align to any of the above databases, the software of ESTScan was used to determine its sequence direction (Iseli et al. 1999). The direction from 5’ to 3’ ends was provided for unigenes with confirmed directions; for those without direction, sequences from the assembly software were provided.

FUNCTIONAL ANNOTATION OF UNIGENES

Gene annotation was conducted after re-assembly, which provided the necessary information of gene expression and function, including protein functional annotation, COG functional annotation, and gene ontology (GO) functional annotation. Unigene sequences were compared using Blastx to proteins in databases, namely NCBI nr, Swiss-Prot, KEGG, and COG with an E-value < 10⁻³. Using enzyme commission (EC) number terms, biochemical pathway information was collected by downloading relevant maps from the KEGG database, which contains a systematic analysis of inner cell metabolic pathways and functions of individual gene products (Kanehisa & Goto 2000). COG is a database in which orthologous gene products are classified. The functions of unigenes assembled from the 3 tissues were predicted and classified by aligning to COG database. Homology searches were then carried out by query of the NCBI nr database using the Blastx algorithm. After annotation, the Blast2Go program was used to obtain GO annotations (Conesa et al. 2005), and the software WEGO (Ye et al. 2006) was used to perform GO functional classification to understand the distribution of gene functions at the macro level.

DIFFERENTIAL GENE EXPRESSION ANALYSIS

To determine specific expression of genes, expression profiles were calculated using the reads per kilobase mapped (FPKM) method (Mortazavi et al. 2008), which does not include the length of different genes and sequencing level as parameters during calculation of gene expression. The false discovery rate (FDR) control statistical method was used to test multiple hypotheses to correct for P-values (P < 0.01) (Benjamini & Yekutieli 2001). The smaller the FDR, the larger the difference in the expression level between the 2 samples. To identify these differential genes expressed in each tissue in the current analysis, sequences were evaluated with FDR ≤ 0.001. Genes with FPKM < 0.1 were filtered as no expression in tissue. The absolute value of log2Ratio ≥ 1 was used to
judge the significance of differences in the initial gene expression analysis. After the analysis, we identified many genes that were only expressed or detectable in MAG/ED. Such genes were thereafter aligned to GO, COG, and KEGG assignments.

VALIDATION OF THE EXPRESSION OF MAG-SPECIFIC GENES

To validate the expression of MAG-specific genes, different tagmata and tissues including head (5 males), thorax (5 males), abdomen (5 males), midgut (20 males), fat body (50 males), Malpighian tubules (50 males), testis (50 males), and MAG/ED (100 males) were cut off or dissected from 3-d-old adult B. dorsalis males. Total RNA was isolated from each sample by using the TRIzol reagent (Invitrogen, Carlsbad, California). After DNA digestion with DNase (Promega, Madison, Wisconsin), total RNA (1 µg for each) was reverse transcribed into 1st-strand cDNA by using the PrimeScript RT Reagent Kit (TaKaRa, Dalian, China). The quantitative real-time polymerase chain reaction (qRT-PCR) included 5.0 µL GoTaq qPCR Master Mix (Promega, Madison, Wisconsin), 3.5 µL nuclease-free water, 0.5 µL template cDNA, and 0.5 µL of each primer (10 µM). The relative expression was determined by the StepOne Plus Real-Time PCR System (Life Technologies, Woodlands, Singapore). The thermal program was 1 cycle of 95 °C for 2 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. At the end of each run, a melting curve was generated to rule out the possibility of primer-dimer formation. The sequences of the specific primer sets were designed using online software Primer3 (version 4.0.0) (Table S1). A fragment of the ribosomal protein subunit 3 from B. dorsalis was also amplified as the inner reference (Wei et al. 2015b). Three biological replicates were performed for each sample. The relative expression levels were calculated using the 2−ΔΔCt method (Livak & Schmittgen 2001). Data were analyzed statistically by 1-way analysis of variance (ANOVA) for expression comparisons in the SPSS 16.0 software (SPSS Inc., Chicago, Illinois), and a value of P < 0.05 was considered to be statistically significant.

Results

ASSEMBLY AND ANALYSIS

Re-assembly of transcriptome data produced 31,782 unigenes with a mean length of 922 bp, and the total length of nucleotides was 29,300,417 bp (Table 1). The results indicated that the re-assembly sequences were longer than those in each transcriptome. A greater number of distinct clusters (7,426) suggested more isoforms were assembled in this study because more raw data were used in the assembly program. Of these unigenes, 3,655 (11.50%) unigenes were longer than 2,000 bp, whereas 1,257 unigenes in fat body, 2,438 unigenes in testis and 1,705 unigenes in MAG tissues were longer than 2,000 bp (Figure S1).

ANNUATION OF PREDICTED PROTEINS

After Blastx against the NCBI nr, nt, Swiss-Prot, KEGG, COG, and GO databases, 21,306 distinct sequences (67.04%) were matched to known genes encoding functional proteins (Table 1). Most of these sequences (20,180) were annotated in the NCBI nr database, whereas 12,456 were annotated in NCBI nt, 15,657 in Swiss-Prot, 13,782 in KEGG, 7,168 in COG, and 13,577 in GO. All the sequences annotated to NABI nr database were thereafter deposited at DDBJ/EMBL/GenBank with the Transcriptome Shotgun Assembly project accession number of GY50000000. Taking annotation in the nr database for the following analysis, there were almost 83.42% (16,835 out of 20,180) unigenes that showed strong similarity to 36 Drosophila species with 12.99% of the sequences matching homologous sequences from D. viridis Sturtevant, followed by D. willistoni Sturtevant (12.05%), D. melanogaster Meigen (11.73%), D. mojavensis Patterson (10.86%), D. ananassae Doleschall (8.19%), D. grimshawi Oldenberg (7.82%), and D. pseudoobscura pseudoobscura Frolova (7.22%) (Figure S2A). Only 306 sequences (1.52%) were annotated to the genus Bactrocera. The most important reason was the availability of the genome sequences of many Drosophila species, but the genome sequence of B. dorsalis had not been released when we performed the annotation. For the predicted proteins, over half of the hits had over 60% similarity with known proteins in the nr database (Figure S2B).

FUNCTIONAL ANNOTATION OF UNIGENES

Annotations of the unigenes provided information on their potential functions in each tissue of B. dorsalis. All of the re-assembled sequences were aligned to the COG database for functional prediction and classification. In total, 7,168 genes were assigned to 25 functional categories (Figure S3). Similar with each specific tissue transcriptome, the most common category was “general function prediction only,” which included 2,856 distinct sequences. The top 3 smallest groups were “RNA processing and modification” (1.20%), “extracellular structures” (1.00%), and “nuclear structure” (0.04%). For GO functional analysis, 13,577 sequences were categorized into 59 functional groups of 3 ontologies: biological process, cellular component, and molecular (Figure S4). There were 28 functional groups that contained more than 1,000 unigenes, including 17 biological processes, 8 cellular components, and 3 molecular functions. The most dominant groups in 3 ontologies were “cellular process” (9,880 unigenes) in biological process, “cell part” (7,839 unigenes) in cellular component, and “binding” (7,200 unigenes) in molecular function. KEGG annotation is usually used to categorize gene functions emphasizing biochemical pathways (Kanehisa & Goto 2000). In total, 13,782 sequences were remapped to 256 predicted pathways, of which 46 contained over 200 unigenes. Re-

Table 1. Assembly summary of transcriptomes of fat body, testis, and male accessory glands and ejaculatory duct (MAG/ED) from Bactrocera dorsalis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Fat body (Yang et al. 2014)</th>
<th>Testis (Wei et al. 2015b)</th>
<th>MAG/ED (Wei et al. 2016)</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (nt)</td>
<td>16,421,863</td>
<td>23,083,068</td>
<td>19,769,995</td>
<td>29,300,417</td>
</tr>
<tr>
<td>Number of unigenes</td>
<td>27,787</td>
<td>30,516</td>
<td>30,669</td>
<td>31,782</td>
</tr>
<tr>
<td>NS5</td>
<td>945</td>
<td>1,316</td>
<td>1,044</td>
<td>1,579</td>
</tr>
<tr>
<td>Mean length of unigene</td>
<td>591</td>
<td>756</td>
<td>645</td>
<td>922</td>
</tr>
<tr>
<td>Number of distinct clusters</td>
<td>3,166</td>
<td>5,487</td>
<td>4,586</td>
<td>7,426</td>
</tr>
<tr>
<td>Number of distinct singletons</td>
<td>24,621</td>
<td>25,029</td>
<td>26,083</td>
<td>24,356</td>
</tr>
<tr>
<td>Number of sequence with E-value &lt;10^-5</td>
<td>18,426</td>
<td>20,921</td>
<td>20,419</td>
<td>21,306</td>
</tr>
</tbody>
</table>
markably, the most abundant pathway was “metabolic pathways” including 1,919 sequences (Table S2).

**DIFFERENTIALLY EXPRESSED GENES (DEGS)**

Descriptive and quantitative transcriptome analysis is important to interpret the functional elements of genomes and reveal the molecular constituents of cells and tissues. To identify differentially expressed genes (DEGs) in MAG/ED, fat body, and testis tissues, we compared the transcriptomes from the 3 tissues and identified a number of genes highly presented in each transcriptome. Altogether, 25,054 unigenes were expressed in all the 3 tissues. In this study, transcriptome sequenceings were based on a mixture of total RNA from different age/development stages of male flies. Genes with tissue-specific expression may be critical to the function of each tissue. Among all the unigenes, there were 458 unigenes that were expressed specifically in MAG/ED, 385 in fat body, and 2,238 unigenes in testis tissues (Fig. 1). Another 207 unigenes were expressed lowly in the 3 tissues. In testis tissue, mitosis and meiosis take place in the process of spermatogenesis. It is possible that many specific unigenes are involved, explaining the particularly large number of unigenes expressed specifically in testis tissue.

In this study, there were 182 unigenes that were 100-fold higher expressed in MAG/ED than both fat body and testis tissues (Table S3). These unigenes were considered as functional genes highly expressed in MAG/ED. The MAG/ED-specific unigenes (641) were classified into 20 categories (Fig. 2) after COG annotation. Interestingly, the most prominent category was “cell wall/membrane/envelope biogenesis.” Categories of “cell cycle control, cell division, chromosome partitioning” and “transcription” also made up a large proportion. All these MAG/ED-specific highly expressed genes were also classified into 3 main categories and assigned to 45 subcategories by using the GO assignment (Fig. 3). Different from each tissue-specific transcriptome, the predominant groups in biological process were “cellular process,” “single-organism process,” and “multi-organism process.” All the MAG/ED-specific highly expressed genes were thereafter mapped to 6 groups of KEGG pathways including 105 pathways (Fig. 4). Remarkably, “metabolic pathways” was the most important pathway, which contained 32 unigenes. In total, there were 51 unigenes that were involved in immunity, e.g., “maturity onset diabetes of the young” involved in endocrine and metabolic diseases, “*Vibrio cholerae* infection” and “tuberculosis” involved in bacterial infection, “*Herpes simplex* infection” and “Epstein-Barr virus infection” involved in viral infection.

**MINING FOR FUNCTIONAL GENES HIGHLY EXPRESSED IN MAG/ED**

To identify MAG/ED-specific genes in this study, we compared the MAG/ED transcriptomes with those of fat body and testis, and identified a number of genes highly expressed in the MAG/ED. Because genes in MAG/ED have a rapid evolution frequency, most of the unigenes were annotated with non-specific functional descriptions; 294 out of 641 unigenes were distinct singleton sequences, and only 53 of them were predicted genes (Table S3). Notable unigenes specific to MAG/ED were circadian clock-controlled protein (*Unigene*4743), C-type lectin (*Unigene*12723), vitellogenin-1 (*Unigene*12065), venom serine protease (*Unigene*8523), perlin (*Unigene*5896), glucose dehydrogenase (*Unigene*9651), lipase 1 (*Unigene*5218), and pheromone-binding protein–related protein 5 (*Unigene*20808). These MAG/ED-specific genes may be involved directly in the male accessory proteins’ functions.

The expression profiling of several annotated unigenes (>500 bp) were also validated at transcriptional level by qRT-PCR. The results showed that all of these genes had high expressions in MAG/ED tissue (Fig. 5). Besides, 4 unknown distinct unigenes (>500 bp) with the largest FPKM values were determined in tissues of male adults, and all of the 4 unigenes were highly expressed in MAG/ED tissue (Fig. 6). Other MAG/ED-specific genes sequences of <500 bp or less then 100-fold expressed compared with other tissues were also identified, such as oocyte maturation arresting factor (*Unigene*21945), antigen 5 precursor (*Unigene*21744), N-acetylgalactosaminyltransferase (*Unigene*22691), reverse transcriptase polyprotein (*Unigene*21663), and potassium-dependent sodium–calcium exchanger (*Unigene*21750). These protein-coding genes also had high expression in MAG/ED tissue (Fig. S5). Regulatory mechanism determination of genes expressed differentially or specifically would provide new insights on the regulation of complex processes such as reproduction, development, and evolution.

**Discussion**

There is an increase in the studies on insect MAG and testis tissues to investigate the mating response mechanism and the function of specific seminal proteins through transcriptome (Scolari et al. 2012; Bailey et al. 2013) and proteome analysis (Simmons et al. 2013; Xu et al. 2013). Although the comparative analysis of reproductive tissues was studied in the cricket *T. oceanicus* before (Bailey et al. 2013), the current study is the first report to compare 3 tissue-specific transcriptomes by re-assembling the raw data in Tephritidae insects. Not much is known about the genes involved in the reproduction of this species, particularly the rapidly evolving Acps (Swanson et al. 2001; Braswell et al. 2006). This knowledge gap was previously suggested as the likely reason for the high proportion of unknown proteins in the testis and MAG transcriptomes of *T. oceanicus* males (Bailey et al. 2013). In each transcriptome, the number of unigenes may not necessarily reflect the real transcriptome complexity, as many of the assembled sequences may represent distinct non-overlapping regions of the same transcripts. Thus, we re-assembled sequences of *B. dorsalis* MAG/ED, fat body, and...
It is known that sperm function cannot be accomplished without Acps (La Vignera et al. 2011). This could explain why many functional genes involved in sperm motility, capacitation, and acrosome reaction were found highly expressed in MAG tissue. In insects, a remarkable array of proteins and peptides has been reported to occur in the male seminal fluids, several of which regulate critical female reproductive functions. Semen is particularly rich in fructose that provides nutrient energy for the spermatozoa and is secreted by the seminal vesicle. The positive correlation between the levels of fructose in the semen and the motility of sperms has previously been demonstrated (Patel et al. 1988). In the B. dorsalis MAG/ED transcriptome, aldose reductase unigenes, 6-phosphofructo kinase, fructose-1, 6-bisphosphatase, and fructose 1,6 bisphosphate-aldolase were identified (Wei et al. 2016). The transcripts of a glucose dehydrogenase (Unigene9651) and a lipase (Unigene5218) were highly detected in MAG/ED tissue, and because MAG/ED tissue actively secretes proteins and lipid, these organs may have a high demand for energy. In addition, sperms stay in the female genital tract for a long time, suggesting that some ATPases could be transported to females during mating. These ATPases, also identified as Acps (Sirot et al. 2011), are postulated to provide energy to sperms by hydrolyzing triglycerides (Walker et al. 2006). It has been reported that ATPase and UTPase amounts in MAGs are associated with the motility of sperms (Werner & Simmons 2008).

During the mating process, pathogens are introduced into the female reproductive tract (Lung et al. 2001). Therefore, to ensure successful fertilization, immunity-related proteins are likely to be present in the MAG/ED to protect seminal fluids from microbial infection. Lectins were mostly studied to investigate the immune response. In a previous study, 17 immuno- and C-type lectins transcripts were identified in the B. dorsalis fat body (Yang et al. 2014). Lectins were also identified in Tribolium castaneum (Herbst) (Coleoptera: Tenebrionidae), D. melanogaster, and Meligethes aeneus (F.) (Coleoptera: Nitidulidae) involved in immune response (Vogel et al. 2014). These types of genes can mediate the pathogen recognition in insects by phagocytosis (Pace & Baum 2002). Perlucin is a member of the C-type lectin family, which participates in the immune response to various stressors and defends against invading pathogens (Lin et al. 2013). Several immune functions have been proposed for C-type lectins, including cell adhesion, glycoprotein turnover, activation of prophenoloxidase cascade, hemocyte-mediated nodule formation, encapsulation, and opsonization (Zhang et al. 2011). Some of such genes were also identified in our previous study in MAGs of B. dorsalis, but only the genes involved in bacterial infection were identified (Wei et al. 2016). In this study, we identified a C-type lectin (Unigene12723) and a perlucin (Unigene5480 and Unigene5896) highly expressed in MAG/ED tissue. In addition, antigen 5 is a major allergen in the venom of vespids. This homologous gene or protein has been identified in many insect species (Hoffman 1993; Hawdon et al. 1996). This gene has also been in-
vestigated in mammal testis tissue and was found to be involved in spermatogenesis (Mizuki et al. 1992; King & Lu 1997), but the biological function of antigen 5 and its sequence-related proteins are still unknown. This study is the first to report the presence of antigen 5 in Bactrocera dorsalis.

![Gene ontology (GO) classification of unigenes expressed highly in male accessory glands and ejaculatory duct tissue of Bactrocera dorsalis.](https://bioone.org/journals/Florida-Entomologist)
Fig. 4. Kyoto encyclopedia of gene and genomes (KEGG) analysis of unigenes expressed highly in male accessory glands and ejaculatory duct tissue of *Bactrocera dorsalis*. Each category contains more than 1 unigene sequences.
antigen 5 in *B. dorsalis*. The gene's high-level expression in MAG/ED tissue suggests that it may exhibit an immune-related function in *B. dorsalis* reproductive tissues. Venom allergen and venom serine protease also were highly expressed in MAG/ED tissue of *B. dorsalis*.

Acps in insects evolve quickly and play an important role in ensuring the successful mating and fertilization. A comparative structural modeling approach indicates that major functional classes of mammalian and *Drosophila* seminal fluid proteins are conserved, suggesting a conserved function despite differences in reproductive strategies (Mueller et al. 2004). Evolutionary changes in these secretions may be driven by sexual selection and sexual conflict and have important implications in fertilization compatibility and evolution of reproductive isolation (Braswell et al. 2006). Indeed, in *D. melanogaster*, the evolutionary rate of many of these genes is so fast that they lack detectable orthologues even in other *Drosophila* species (Swanson et al. 2001; Mueller et al. 2005; Begun et al. 2006). Further investigation of the expression and regulation of these genes may provide new insights on the regulation of complex processes such as reproduction, development, and evolution.

In this comparative transcriptomic analysis, a number of functional genes possibly involved in immunity, sperm fertilization, and female mating regulation were identified in MAG/ED tissue. The identification of these unigenes could provide the foundation for further understanding the patterns and processes of mating regulation mechanisms and molecular
lar evolution among reproductive proteins in MAG/ED tissues of Tephritidae insects. This transcriptome comparison will advance our understanding of the function of the reproductive proteins in MAG/ED tissue that may play roles in the regulation of reproductive processes in *B. dorsalis*.

**Acknowledgments**

This work was supported in part by the National Natural Science Foundation of China (Grant No. 31601640), the National Student Innovation Training Program (201410635069), the earmarked fund for Modern Agro-industry (Citrus) Technology Research System (CARS-27), the Fundamental Research Funds for the Central Universities (XDJJK2016C081) of China, and the Project Funded by China Postdoctoral Science Foundation (2016T90828).

**References Cited**


King TP, Lu G. 1997. Hornet venom allergen antigen 5, Dol m 5: its T-cell epitopes in comparison with other Dol m 5 fragments. Allergy and Clinical Immunology 91: 71–78.


