A transcriptomic approach reveals the molecular basis of pre-pupal diapause of Red Banded Mango Caterpillar, *Deanolis sublimbalis*

Gandham Krishnarao[1], Avvaru Sujatha[1], Pola Sunitha[1], Meenal Vyas[2] and Pagadala Damodaram Kamala Jayanthi[2,*]

[1]College of Horticulture, Dr Y. S. R. Horticultural University, Venkataramannagudem 534 101, India
[2]Division of Crop Protection, ICAR-Indian Institute of Horticultural Research, Bengaluru 560 089, India

The Red Banded Mango Caterpillar (RBMC), *Deanolis sublimbalis* Snellen (Lepidoptera: Crambidae), a devastating monophagous pest of mango (*Mangifera indica* L.), enters a pre-pupal diapause in the absence of host fruits synchronizing its life cycle with seasonal fruiting across southeast Asia and Oceania. Considering its unique nature, a detailed *de novo* transcriptome analysis was carried out on different physiological stages of RBMC pupae to understand the mechanisms underlying diapause. A total of 102 differentially expressed unigenes were identified with altered expression patterns (55 upregulated and 47 downregulated) and consequently mapped to various pathways associated with diapause. Three major pathways, i.e. proteasome, Epstein–Barr virus infection and lipoic acid metabolism were significantly ($P < 0.01$) enriched during the diapause phase in *D. sublimbalis.* From the three pathways, 16 differentially expressed genes (15 up-regulated and 1 down-regulated) were identified to play a vital role in diapause management. To our knowledge, no earlier studies have identified diapause-related genes in *D. sublimbalis*. The information gained from the present study can be exploited to develop control strategies involving molecular tools.

Keywords: Developmental stages, diapause, metabolic pathways, Red Banded Mango Caterpillar, transcriptome analysis.

DIAPAUSE is a pre-programmed developmental (= metabolic) arrest in the life cycle of organisms that helps them tide over harsh conditions. This phenomenon can be observed in many insects, particularly to overcome environmental stress[1–3]. Many insect pests have evolved such ecophysiological modifications to survive unfavourable conditions. The process of diapause is dynamic with multiple, distinct phases. In the first or the induction phase, the insect receives an environmentally induced diapause stimulus and begins to prepare itself for the impending diapause. Physiological processes along with active development facilitate it for diapause (= developmental arrest) in the second phase. The third or the diapause initiation stage is associated with a cease in development and reduced metabolic rate. The fourth or the diapause phase constitutes the actual diapause and lasts until the environmental conditions turn favourable, leading to the final phase, i.e. diapause termination. The final phase initiates the resumption of development and active metabolism under favourable conditions[4]. Therefore, diapause initiation and termination help insects evade stressful environmental conditions and acclimatize to seasonal variations[1]. Despite its ecological significance, the crucial underlying molecular mechanisms responsible for diapause are widely unresolved.

The Red Banded Mango Caterpillar (hereafter RBMC), *Deanolis sublimbalis* Snellen (Lepidoptera: Crambidae), is a monophagous pest of mango fruits that synchronizes its development to the seasonal fruiting of its host using a strategic intermittent pre-pupal diapause. It is a destructive mango pest causing huge economic loss across India, Indonesia, Papua New Guinea, Myanmar, Thailand, China, Brunei, the Philippines and parts of Australia[5–14]. Earlier studies have reported that due to the restricted seasonal availability of mango fruits (usually fruiting occurs annually once), RBMC enters pre-pupal diapauses lasting around 8–9 months until the next season[10,15]. Therefore, the induction of a pre-pupal diapause aids in the survival of RBMC by allowing it to synchronize its life cycle with the seasonal fruiting of its host[15]. Subsequent studies also revealed that mature larvae undergo diapause in the bark of mango trees and physiological changes within the tree system initiate diapause termination[12,13]. Apart from biological processes, studies on molecular mechanisms of diapause in RBMC will facilitate an in-depth understanding of its physiological basis.

Similar studies carried out on *Aedes albopictus* (Skuse), *Culex pipiens* L., *Bombyx mori* L., *Drosophila montana* Meigen and *Bactrocera minax* (Enderlein) revealed multiple

*For correspondence. (e-mail: kamalajayanthi.pd@icar.gov.in)

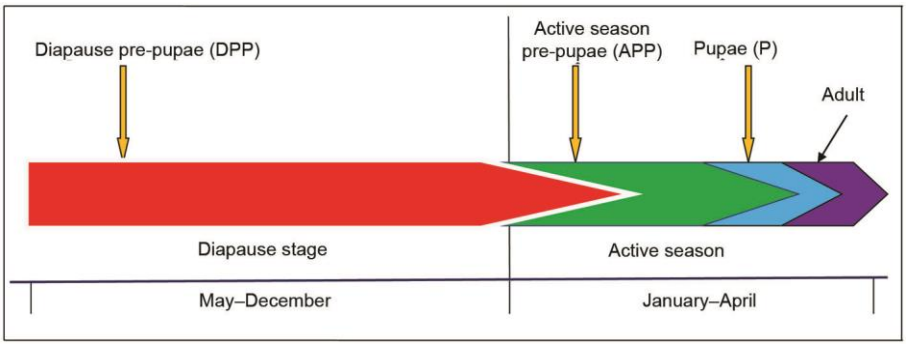

Figure 1. Schema of sampling of different stages of *Deanolis sublimbalis* for transcriptome sequencing.

diapause candidate genes that play crucial roles in the underlying physiological processes of insect diapause[16–24]. Although *D. sublimbalis* is an economically important pest, studies on the molecular aspects of its diapause are scarce. Elucidating the molecular mechanisms behind the process of diapause will help us better understand this notorious pest. In this context, the currently available high throughput sequencing technologies will help improve our understanding of the biological processes causing diapause in RBMC, and provide vast genomic information on its physiological systems.

In this study, a total of 179.7 million clean reads were obtained from the RBMC transcriptome by Illumina sequencing and a whole of 6218 unigenes with 367 KEGG pathways were identified. Further, differentially expressed genes associated with diapause and the transcripts exhibiting higher and lower expression levels in pairwise comparisons were examined. Our results emphasize the novel insights of molecular mechanisms underlying the pre-pupal diapause of this economically important pest.

Materials and methods

Insects

Three varied phenological stages of *D. sublimbalis*, i.e. diapause pre-pupae (DPP), active-season pre-pupae (APP) and pupae (P), were collected from Naguluru village, Krishna district, Andhra Pradesh, India (15°55′N; 81°10′E; 73 m amsl) (Figure 1). A set of five insects from each stage was used for experimental analysis. All samples were snap-chilled in liquid nitrogen and then stored at –80°C until RNA was extracted.

RNA isolation

RNA extraction was carried out using the Direct-zol™ RNA MiniPrep kit (Zymo Research, CA, USA) according to the manufacturer's instructions. RNA quantity was estimated using a Qubit RNA BR assay (Thermo Fisher Scientific, USA) and quality was confirmed using an Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA) and an RNA Nano kit (Agilent Technologies).

cDNA library construction and RNA sequencing

Sequencing libraries were prepared from 1 ug of RNA using NEB Next Ultra RNA Library Prep kit for Illumina (NEB #E7770, New England Biolabs, USA) sequencing after depleting the rRNA using NEB Next rRNA Depletion kit (Catlog #E7400S, New England Biolabs, USA). In brief, probes were hybridized to the RNA, followed by RNaseH digestion and DNase1 digestion. RNA purification was done after rRNA depletion using Agencourt RNA clean XP.

RNA fragmentation and cDNA synthesis were performed after the purification step. Subsequently, double-stranded cDNA was purified using 1.8× Agencourt AMPure XP beads. Adapter ligation was performed to cDNA library and the ligation reaction was purified using AMPure XP beads. This step was followed by PCR enrichment and purification of the PCR-enriched cDNA. Library quantity was checked with a Qubit ds DNA HS kit (Thermo Fisher Scientific) and quality was assessed on an Agilent Bioanalyzer 2100 and Agilent DNA 7500 kit (Agilent Technologies).

Illumina sequencing

The cDNA libraries were denatured after quality inspection and diluted to the Illumina recommended concentration of 1.8 pM using 0.2 nM NaOH and HT1 (hybridization buffer) reagent. The denatured and diluted libraries were loaded onto Nex Seq reagent cartridge.

Cluster generation and sequencing were performed on Illumina Next Seq 500 to generate 2×150 paired-end reads. Soon after the completion of run, demultiplexed sequence data were collected in FASTQ format and checked for data quality.

The sequence data quality was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC[25] for base quality distribution, per cent reads with average Q30 and Q20 and per cent GC and sequencing

adapter contamination. Raw sequence reads were processed to remove adapter sequences and low-quality bases using Trim-galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and Trimmomatic packages[26]. Whole transcriptome data of different developmental stages (DPP, APP and P) of *D. sublimbalis* are available under the SRA accession PRJNA691148.

De novo transcriptome assembly and functional annotation

De novo transcriptome assembly was generated using Trinity v.2.4.0 with default parameters[27]. QC (quality control)-passed sequence reads from all the samples (DPP, APP and P) were pooled and transcripts shorter than 300 bases were discarded.

The assembled transcripts were annotated employing UniProt/Swiss-Prot (UniProt), NCBI nr database and COG databases using BLASTx[28] and hits with *E* values <1e-05 and 1e-07 were considered as unigenes. Unannotated transcripts were further BLAST-searched against RNA central non-coding RNA sequences. The assembled sequences were analysed further and KEGG orthology (KO) annotations were assigned using KAAS (KEGG Automatic Annotation Server)[29]. The annotated transcripts were considered unigenes.

Differential expression analysis of DPP, APP and P

QC-passed reads were aligned onto the assembled transcriptome using Bowtie2 aligner with default parameters[30]. Duplicate reads were marked and removed using the Picard toolkit (http://broadinstitute.github.io/picard/). Transcript/unigene expression was estimated by extracting the read counts using SAM tools[31]. Differential gene/transcript expression was assessed using DESeq2 after normalizing the expression count data[32]. Comparisons were made between APP and P, DPP and P and DPP and APP.

Transcripts/unigenes that showed absolute differential expression in log two-fold change >2 and *P*-value <0.05 were considered statistically significant. Gene ontology (GO) enrichment analysis can determine the key biological functions and the Kyoto Encyclopedia of Genes and Genomes (KEGG) can determine the significant biochemical metabolic pathways and signal transduction pathways; therefore, the differentially expressed genes were subjected to these analyses. Differently expressed significant transcripts were annotated against GO. To check the biological process, cellular component and molecular functions of the transcripts, GO-term enrichment analysis was done using BINGO[33]. GO terms with hypergeometric *P*-value <0.05 were considered significant. The GO distribution graph was plotted using WEGO[29]. Similarly, significantly differentially expressed genes with KO annotation were considered for pathway enrichment analysis, which was carried out using a custom R script for the KEGG pathway.

Results

D. sublimbalis a serious mango pest, undergoes pre-pupal diapause during its off-season which helps the moth survive until the next mango fruiting season. Understanding and exploiting this weak link in the life cycle of RBMC might provide clues for its effective management.

Illumina sequencing and de novo assembly

Transcriptome analysis was performed on different developmental stages of *D. sublimbalis*, namely DPP, APP and P, each with two biological replicates. The corresponding six cDNA libraries yielded a total number of 193,553,980 raw reads, comprising 74,615,140; 57,502,348 and 61,436,368 reads for DPP, APP and P respectively (Supplementary Table 1). A total of 179,710,220 clean reads were obtained from all six libraries after adaptor trimming and quality filtering. In the DPP stage, 69,449,178 reads were obtained with a read length of 36–151 (bp) and Q20 (%), Q30 (%) and GC (%) values being 100, 99.34 and 53.00 respectively. In the APP and P stages, 53,193,758 and 57,067,284 reads were obtained respectively, with a read length of 36–151 (bp). The Q20 (%), Q30 (%) and GC (%) values were 100, 99.21, 53.00 and 100, 99.29, 52.00 respectively. The *de novo* assembly yielded 45,884 transcripts with an average length of 890 bp. The assembled transcripts had a minimum length of 301 bp and a maximum length of 28,177 bp. The GC content was 42.62%, with N50 transcripts being 1221 in number (Supplementary Table 2). However, the length of most transcripts (20,843) was around 200–500 bp, while a few transcripts (410) were >5000 bp length (Supplementary Figure 1).

Annotation of unigenes

A total of 24,148 from 45,884 transcripts (52.62% of the total transcripts) were annotated against the Swiss-Prot and NCBI nr databases using BLASTx. Among these transcripts, only those with an *E*-value of <1*e*-05 or 1*e*-07 were considered unigenes. The remaining unannotated transcripts were BLAST-searched against central non-coding RNA sequences, which yielded another 578 transcripts (Supplementary Table 2). The length of the assembled 7586 unigenes ranged from 200 to 500 bp. A total of 398 unigenes were found in the length range >5000 bp (Supplementary Figure 2). The *E*-value distribution of the peak hits with 48.94% homologous unigenes ranged between 1*e*-50 and 1*e*-05 and for the other unigenes at ranged between 1*e*-100 to 1*e*-50, 1*e*-150 to 1*e*-100, 0 to 1*e*-150 and 0 were 25.57%, 10.57%, 11.25%, 3.66% respectively (Supplementary Figure 3 *a*). As the genomic sequences of *D. sublimbalis* are not yet available, sequence alignments of the experimental unigenes were compared with the known genomes of other species. The unigenes showed maximum hits to

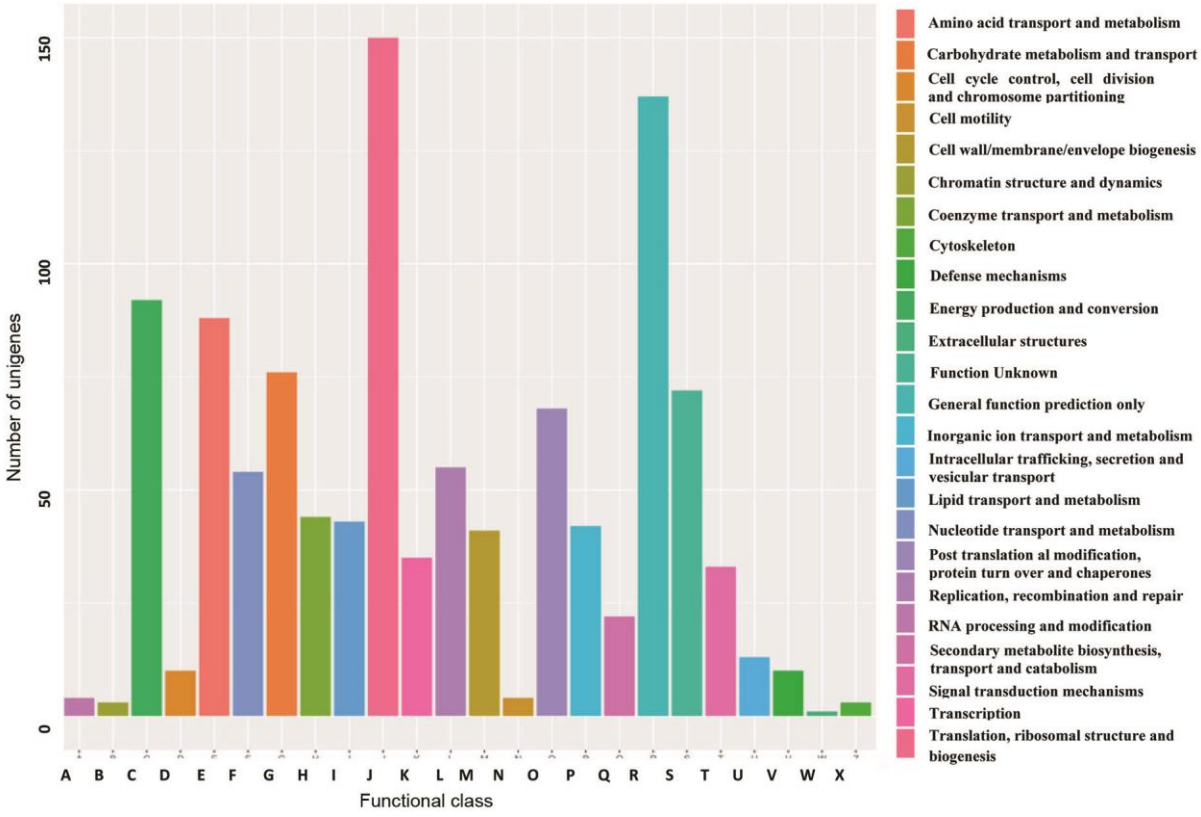

Figure 2. Cluster of orthologous groups (COG) functional categories in the *D. sublimbalis* transcriptome.

the orange worm, *Amyelois transitella* (Walker) (18.34%), followed by mulberry silk moth, *Bombyx mori* L. (7.65%), Asian yellow swallowtail, *Papilio xuthus* L. (6.82%), Old world swallowtail, *Papilio machaon* L. (5.74%), pink bollworm, *Pectinophora gossypiella* (Saunders) (5.67%), Monarch butterfly, *Danaus plexippus* L. (3.91%), common mormon, *Papilio polytes* L. (3.38%), diamondback moth, *Plutella xylostella* L. (3.06%), the winter moth, *Operophtera brumata* (L.) (2.97%), speckled wood butterfly, *Pararge aegeria* L. (1.59%) and others (40.87%) (Supplementary Figure 3 *b*).

Cluster of orthologous groups (COG) is one of the functional annotations usually performed to analyse the integrity of the libraries and the effectiveness of annotation by mapping the unigenes against the COG database using BLASTx. Approximately 25% of the unigenes were mapped onto the COG database, of which 17.7% could be assigned to 25 COG functional groups. The largest group in the cluster was translation, ribosomal structure and biogenesis (with 150 unigenes), followed by general function prediction (137), energy production and conversion (92), amino acid transport and metabolism (88), carbohydrate transport and metabolism (76), unknown functions (72), post-translational modification, protein turnover and chaperone (68), replication, recombination and repair (55), co-enzyme transport and metabolism (44), lipid transport and metabolism (43), inorganic transport and metabolism (42) and cell wall/membrane/envelope biogenesis (41), while the remaining groups contributed a total of 138 unigenes. These results demonstrate that the functional groups, translation, ribosomal structure and biogenesis were the most significant in *D. sublimbalis*, with a total of 150 mapped unigenes (Figure 2).

GO is an international standardization of the gene functional classification system, which defines gene functions and relationships under three major categories, namely molecular functions (molecular activities of gene products), cellular components (where gene products are active) and biological processes (pathways and larger processes associated with the activities of multiple gene products). These concepts play a key role in the bioinformatics annotation process. In the present study, cell and cell-part categories were the most abundant among the cellular components, whereas virion and virion parts were the least abundant. In the case of molecular functions, binding and catalytic activity accounted for the majority of unigenes, while electron carriers as well as protein tags had the least number of unigenes assigned. Within the category of biological processes, cellular and metabolic processes were the most abundant, whereas viral reproduction and locomotion were the least abundant groups (Figure 3).

Further, KEGG analysis was performed to examine the biological complexity of the genes. KEGG is a database for understanding high-level functions and utilities of biological

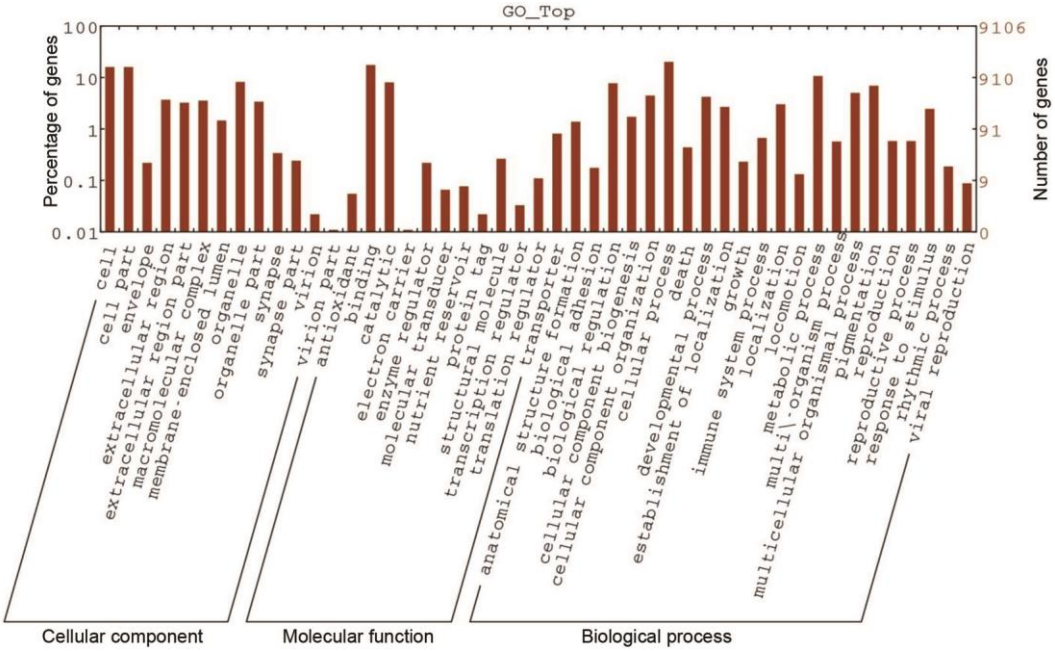

Figure 3. Gene ontology (GO) unigene categories for the *D. sublimbalis* transcriptome.

systems (such as cells, organisms, ecosystems) and molecular-level information, especially using high-throughput experimental technologies. Using KO, a total of 6218 unigenes could be mapped to 367 KEGG pathways and each assigned a KO number. Among these pathways, ribosome, endocytosis, spliceosome and RNA transport were the most represented with 120, 108, 102 and 101 unigenes respectively. Differentially expressed unigenes were classified into six categories through KEGG orthology, i.e. cellular process, environmental information processing, genetic information processing, human diseases, metabolism and organismal systems. The annotations signifying the areas specific to the present study were further identified and analysed (Figure 4).

Analysis of differentially expressed transcripts

To recognize the significant expression changes in transcripts, differential expression analysis was conducted through pairwise comparisons of the three dissimilar treatment conditions, viz. DPP, APP and P. Only the transcripts that showed absolute differential expression in log two-fold change >2 with *P*-value <0.05 were considered significant (Figure 5). A comparison of DPP with P showed 150 upregulated and 86 downregulated transcripts, whereas 64 upregulated and 99 downregulated transcripts were detected between DPP and APP stages. Further, 126 upregulated and 114 downregulated transcripts were uncovered between the APP and P stages (Figure 6). When DPP versus APP and DPP versus P comparisons were done, 15 up-regulated and 20 downregulated transcripts were identified. Furthermore, 40 upregulated and 27 downregulated transcripts were identified between DPP and P and APP and P.

To identify the differentially expressed genes (DEG) associated with diapause, transcripts showing higher and lower expression levels in pairwise comparisons were examined. A total of 55 upregulated and 47 downregulated transcripts in the diapause were identified based on the annotation results (Figure 6).

Functional enrichment analysis of differentially expressed transcripts

To comprehend the role of differentially expressed transcripts, GO terms allied with the three different stages were compared after mapping all the differentially expressed transcripts to the GO database. Genes that were upregulated and downregulated could be mapped to these GO terms. The DPP versus APP comparison revealed a total of 114 genes, including 78 biological processes (78 upregulated and none downregulated), eight cellular components (7 upregulated and 1 downregulated) and 28 molecular functions (5 upregulated and 23 downregulated). Whereas a comparison between DPP and P, revealed a total of 45 GO terms, including 26 biological processes (26 upregulated and none downregulated), 10 cellular components (10 upregulated and none downregulated) and 9 molecular functions (7 upregulated and 2 downregulated). Further, the APP versus P comparison revealed a total of 121 GO terms, including 58 biological processes (17 upregulated and 41 downregulated), one cellular component (one upregulated and none downregulated) and 62 molecular functions (57 upregulated and 5 downregulated). Finally, based on

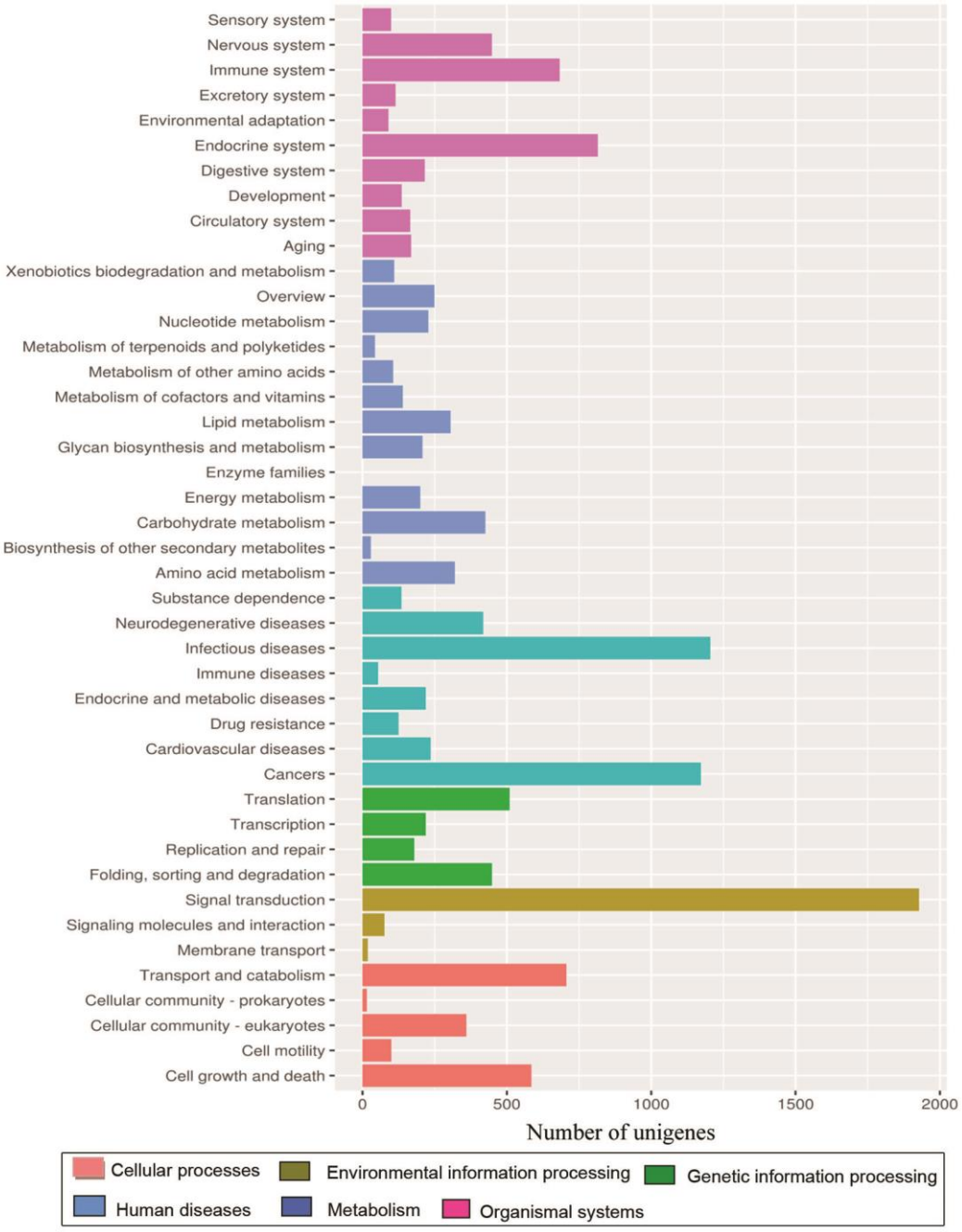

Figure 4. KEGG orthology classifications of differentially expressed genes in *D. sublimbalis* transcriptome.

the annotation results 55 upregulated and 47 downregulated transcripts selected in the diapause phase were identified (Figure 6). In addition, a hypergeometric test was conducted to analyse the significantly enriched pathways of the differentially expressed transcripts by comparing the whole genomic background during diapause. The results revealed 102 transcripts (55 upregulated and 47 downregulated) mapped with 56 pathways in the DPP versus P comparison and 55 pathways mapped with DPP versus APP. Among these, only three pathways, i.e. proteasome, Epstein–Barr virus infection and lipoic acid metabolism were significantly enriched ($P < 0.01$) (Supplementary Table 3). However, when the nominal P-value was set to <0.05, a total of 24 pathways were significantly enriched between DPP and APP. Further, 22 upregulated and 6 downregulated genes were identified among these two stages (Supplementary Table 4). Fifteen pathways were significantly ($P < 0.05$) synchronized between DPP versus P, and 36 upregulated and 9 downregulated genes were identified in these two stages (Supplementary Table 3). However, three pathways (proteasome, Epstein–Barr virus infection and lipoic acid metabolism) were more significantly ($P < 0.01$) enriched

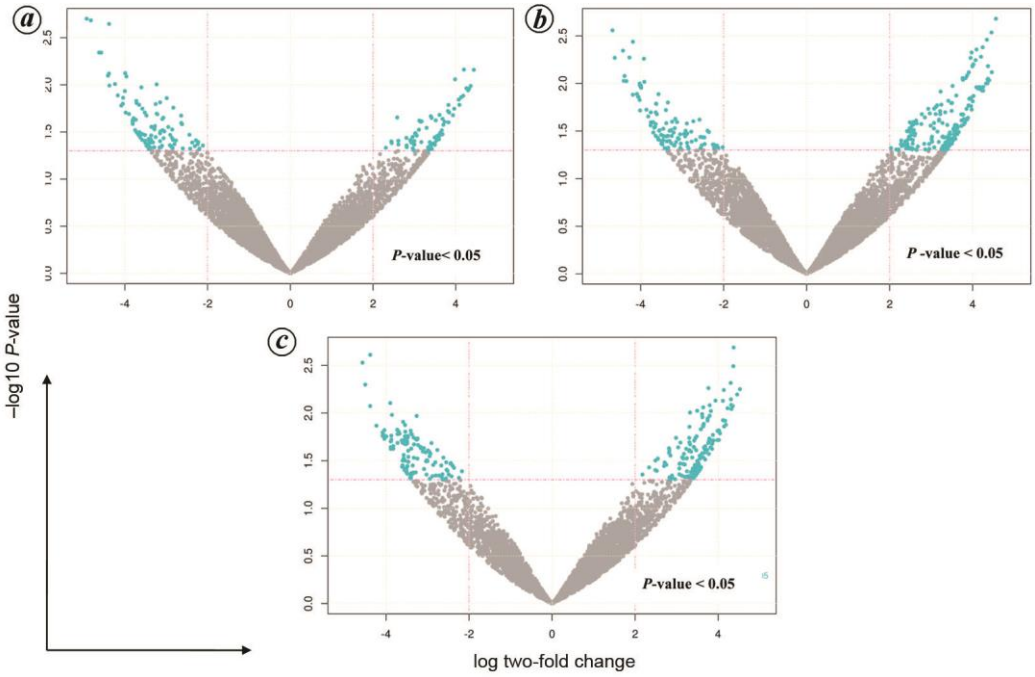

Figure 5. Volcano plots showing the *D. sublimbalis* transcripts in pairwise comparison. ***a***, Diapause pre-pupae (DPP) versus active season pre-pupae (APP). ***b***, DPP versus pupae (P). ***c***, APP versus P.

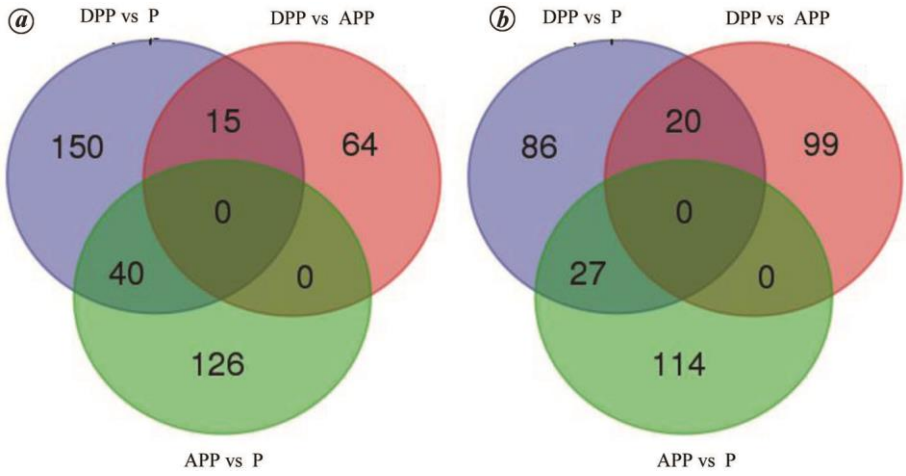

Figure 6. Summary of differentially expressed genes in three libraries (DPP, APP, AP) through pairwise comparisons in the *D. sublimbalis* transcriptome. ***a***, Overexpressed genes. ***b***, Underexpressed genes.

among these stages, with 25 upregulated genes and 1 downregulated gene identified during diapause (Table 1).

Identification of diapause-associated genes and metabolic pathways in D. sublimbalis

A total of 102 differentially expressed transcripts during diapause were discerned from the expression data of the three stages (DPP, APP and P). Among these transcripts, a total of 55 upregulated and 47 downregulated unigenes were identified in the diapause (Supplementary Tables 3 and 4). These unigenes were distributed over 113 pathways. However, three pathways (proteasome, Epstein–Barr virus infection and lipoic acid metabolism) were more significantly enriched (*P*-value < 0.01) than the others in the diapause phase (Supplementary Figure 4 and Table 3). However, the proteasome and Epstein–Barr virus infection pathways were interlinked. Both pathways involve similar proteins and one common associated pathway, i.e. ubiquitin-mediated proteolysis (Table 1 and Supplementary Figure 4 *a* and *b*).

From these three pathways, 25 upregulated genes and 1 downregulated gene were identified in the diapause (Table 1).

Table 1. Genes and proteins involved in the regulation of diapause in *Deanolis sublimbalis*

Pathway	Protein	Genes
Proteasome	26S proteasome regulatory subunit N5	*Psmd12/Rpn5***
	26S proteasome regulatory subunit N6	*Psmd11/Rpn6***
	26S proteasome regulatory subunit N8	*Psmd7/Rpn8***
	26S proteasome regulatory subunit N9	*Psmd13/Rpn9***
	26S proteasome regulatory subunit N1	*Psmd2/Rpn1***
	26S proteasome regulatory subunit N13	*Rpn13***
	26S proteasome regulatory subunit T1	*Psmc2/Rpt1***
	26S proteasome regulatory subunit T4	*Psmc6/Rpt4***
	26S proteasome regulatory subunit T5	*Psmc3/Rpt5***
	26S proteasome regulatory subunit T3	*Psmc4/Rpt3***
	20S proteasome subunit alpha 2	*Psma2***
	20S proteasome subunit alpha 3	*Psma4***
	20S proteasome subunit alpha 4	*Psma7***
	20S proteasome subunit alpha 5	*Psma5***
	20S proteasome subunit beta 6	*Psmb1***
Epstein–Barr virus infection	26S proteasome regulatory subunit T1	*Psmc2/Rpt1***
	26S proteasome regulatory subunit T3	*Psmc4/Rpt3***
	26S proteasome regulatory subunit T4	*Psmc6/Rpt4***
	26S proteasome regulatory subunit T5	*Psmc3/Rpt5***
	26S proteasome regulatory subunit N5	*Psmd12/Rpn5***
	26S proteasome regulatory subunit N6	*Psmd11/Rpn6***
	26S proteasome regulatory subunit N8	*Psmd7/Rpn8***
	26S proteasome regulatory subunit N9	*Psmd13/Rpn9***
`	26S proteasome regulatory subunit N13	*Rpn13***
	26S proteasome regulatory subunit N1	*Psmd2/Rpn1***
Lipoic acid metabolism	Lipoyl transferase 1	*Lipt1**

**Upregulated; *Downregulated.

Though the same genes were involved in the pathways of proteasome and Epstein–Barr virus infection, a total of 15 upregulated genes and one downregulated gene, i.e. *Psmd2/Rpn1*, *Psmd7/Rpn8*, *Psmd11/Rpn6*, *Psmd12/Rpn5*, *Psmd13/Rpn9*, *Psmc2/Rpt1*, *Psmc3/Rpt5*, *Psmc4/Rpt3*, *Psmc6/Rpt4*, *Psma2*, *Psma4*, *Psma5*, *Psma7*, *Psmb1*, *Rpn13* and *Lipt1* were significantly enriched during diapause of *D. sublimbalis.*

Discussion

Diapause is a complex, eco-physiological phenomenon with multiple phenological phases that helps insects deal with the harsh environmental conditions[1,34–36]. The process could be triggered by several factors such as abrupt changes in temperature, variations in photoperiod, scarcity of food supply, etc. Once initialized, an insect undergoing diapause experiences complete developmental arrest and maintains minimal metabolism. This behaviour enables it to acclimatize to the local environment by overcoming prolonged periods of unfavourable conditions. Different species of insects exhibit diapause to varying extents and at distinct developmental stages of their lives.

D. sublimbalis, undergoes diapause at an early pre-pupal stage in the absence of host fruits (during off-season), enabling the synchronization of its life stages to seasonal fruiting[1,15,37]. Understanding the molecular basis of diapause regulation and its associated genes in *D. sublimbalis* will provide invaluable information that could potentially be used to design management strategies and help explore key molecular traits of this perplexing process.

In the present study, illumina sequencing of various developmental stages of *D. sublimbalis*, namely DPP, APP and P resulted in a total of 193,553,856 reads that were subsequently assembled into 45,884 transcripts. Several unigenes from these transcripts were later identified as homologous to multiple functional protein-encoding genes in the UniProt, Swiss-Prot and NCBI nr databases. As a member of the order Lepidoptera, *D. sublimbalis* displays homology to other close lepidopterans, including 18.34% for *A. transitella* (Pyralidae), followed by 7.65% towards the silk moth, *B. mori* (Bombycidae) (Supplementary Figure 3 *b*). However, 47% of transcripts were unannotated, with none of the unigenes being annotated to *D. sublimbalis*, due to a lack of molecular data on this insect, highlighting the significance of the present study.

Differential expression profiling and functional enrichment of the transcripts resulted in a total of 1 downregulated and 15 upregulated genes. All except the *Lipt1* gene were found to be upregulated and none of them has been reported to play a role in insect diapause to date. All upregulated genes encode subunits of proteasome 26S in *D. sublimbalis* (Table 1). Fujiwara *et al.*[38] have shown that the *Lipt1* gene influences lipid metabolism and is known to catalyse the transfer of the lipoyl group from lipoyl-AMP (adenosine monophosphate) to the specific lysine residue of

lipoyl domains of lipoate-dependent enzymes. However, there has been no report suggesting a possible function of *Lipt1* in insect diapauses. Although most of the available knowledge on insect lipid metabolism revolves around *Drosophila*, insects, particularly those with obligate or facultative diapause, have a great potential to elucidate lipid metabolism. In this context, the present study on *D. sublimbalis* might steer us in the right direction towards a better understanding of the functional basis of insect diapause.

Lipids are major energy reservoirs in insects[39] and considering the importance of lipid metabolism during reproduction, flight, starvation and diapause, the downregulation of *Lipt1* is quite intriguing[40]. In general, the deficiency of *Lipt1* was found to impede lipogenesis (a *de novo* process of synthesizing fatty acids as a primary energy storage form, mostly from carbohydrates), but increase fatty acid oxidation (FAO; a major source of ATP production). Lipogenesis and FAO are mutually exclusive processes, the latter is generally high during fasting and low in nourished animals. Oxidation of fatty acids yields twice the amount of energy (9 kcal/g) compared to carbohydrates/proteins (4 kcal/g), which is particularly helpful during diapause when carbohydrate supply is limited.

The differential expressed genes mentioned above were mapped to three major pathways (proteasome, Epstein–Barr virus infection and lipoic acid metabolism; Supplementary Figure 4) and may have a possible role in insect diapause. Of these, the proteasome and Epstein–Barr virus infection pathways were found to involve similar proteins and one common associated pathway (ubiquitin-mediated proteolysis). Proteasomes are the general protein complexes present in all eukaryotes and a few bacteria, while ubiquitin is a small protein that is tagged to other proteins for degradation. Thus, the proteasomal degradation pathway and adjustment of proteins by ubiquitination is an indispensable and fundamental mechanism for various key cellular processes (regulation of gene expression, response to oxidative stress, cell-cycle regulation, DNA repair, antigen presentation, cell-to-cell communication and cell differentiation)[41,42].

In the present study, the identified genes, namely *Rpn5*, *Rpn6* and *Rpn9* were reported as the structurally inter-related subunits among themselves and to the *COP9* (constitutive photomorphogenesis 9) complex and *eIF3* (eukaryotic initiation factor 3). The subunits accumulate to form a horseshoe-like configuration enfolding the *Rpn8/Rpn11* heterodimer (known to be a deubiquitinating enzyme)[43]. Furthermore, ubiquitin plays a major role in the degradation of proteasomes[42,44]. It has been established recently that components of the 26S proteasome (via ubiquitin domains) are involved in proteasome activity[43] and degradation of phosphorylated proteins through the SCF complex (Cu11/Skp/F-box protein), which leads to the regulation of the cell cycle, apoptosis and circadian rhythms[45,46]. Moreover, phosphorylation of proteins is involved in regulating insect diapause[47–51], especially in *B. mori*[49] and *B. minax*[46]. Denlinger[1] reported the degradation of proteins in non-diapause eggs through ubiquitin domains. In the onion fly, *Delia antique* (Meigen), a total of 45 genes were recognized in ubiquitination and their potential role in insect diapause was also highlighted[52].

We identified MAPK (mitogen-activated protein kinase), a signalling mechanism, as one of the integral parts of the Epstein–Barr virus infection pathway, which is one of the differential expression enriched pathways in the transcriptome of *D. sublimbalis* (Supplementary Figure 4 *b*). Earlier experiments demonstrated the role of the MAPK family in insect diapause and cold acclimatization[50–57]. Fujiwara and Denlinger[58] explored a possible role of MAPKs in low-temperature signalling that elicits rapid cold hardening. Thus, proteasome and Epstein–Barr virus infection pathways via ubiquitin-mediated proteolysis might be important in regulating diapause in *D. sublimbalis*. From these two pathways, a total of 15 genes were identified to play a crucial role in the diapause of *D. sublimbalis*. With a possible role in the degradation of enzymes and ubiquitination, these genes could mediate cell-to-cell communication involved in the regulation of diapause.

In the present study, we also identified the gene *Lipt1* (involved in lipoic acid metabolism)[59] as an important candidate gene in the diapause of *D. sublimbalis*. As lipids and fats are the most common forms of energy storage during diapause, we identified lipoic acid metabolism as another important differential expression enriched pathway in the transcriptome analysis of *D. sublimbalis*. The present study reveals that this pathway has a potential role in *D. sublimbalis* diapause as the lipoic acid metabolism starts from fatty acid biosynthesis (Supplementary Figure 4 *c*). Most diapausing insects receive little to no food, thereby hiding and accumulating adequate reserves of energy in the pre-diapause period to meet their metabolic requirements to complete development and resume activity at the termination of diapause. Clark and Chadbourne[60] reported that diapause-associated larvae accumulate greater lipid reserves than non-diapause larvae, since fats serve as pre-dominant energy reserves during the diapause period[61]. Further, it has been demonstrated that fatty acid synthase plays a central role in lipid accumulation in both vertebrates and invertebrates[62–64], and has also been found to influence diapause[65]. Qi *et al.*[66] reported that fatty acid biosynthesis is a vital pathway for diapause of *Coccinella septumpunctata* (Linnaeus) through *de novo* transcriptome studies. All these studies have established that lipoic acid metabolism is one of the prime mechanisms in regulating diapause, which has been provided.

The transcriptomic analyses of RBMC in the present study have provided substantial molecular information and significantly improved our understanding of diapause-associated genes in this economically important frugivorous pest. We have identified the candidate genes and pathways responsible for the regulation of diapause and diverse physiological activities in *D. sublimbalis* by comparing the

transcriptomes of different developmental stages. The genes and pathways identified in the present study provide a platform for further research on diapause-related molecular mechanisms in *D. sublimbalis* and other insects.

Declaration of competing interest: The authors declare that they have no conflict of interest.

1. Denlinger, D. L., Regulation of diapause. *Annu. Rev. Entomol*., 2002, **47**, 93–122.
2. MacRae, T. H., Gene expression, metabolic regulation and stress tolerance during diapause. *Cell. Mol. Life Sci*., 2010, **67**, 2405–2424.
3. Liu, J. Y. and Lin, J. R., Diapause induction and termination of *Bombyx mori*. *Guangdong Canye*, 2011, **45**, 35–38.
4. Kostal, V., Eco-physiological phases of insect diapause. *J. Insect Physiol.*, 2006, **52**, 113–127.
5. Senguptha, G. C. and Behura, B. K., Some new records of crop pests from India. *Indian J. Entomol*., 1955, **17**, 283–285.
6. Tipon, H. T., Seed borer in mango. In Paper present at the Second National Fruit Crop Symposium, Cebu City, Philippines, 12–14 December 1979.
7. Kalshoven, L. G. E., *The Pests of Crops in Indonesia*, PT Ichtiar Baru – Van Hoeve Jakarta, Indowara, 1981, p. 701.
8. Golez, H. G., Bionomics and control of mango seed borer *Noorda albizonalis* Hampson (Pyralidae: Lepidoptera). *Acta Hortic*., 1991, **291**, 418–424.
9. Jha, S. and Sarkar, A., *Mango in Malda*, Bidhan Chandra Krishi Viswavidyalaya, Noida, 1991, p. 13.
10. Zaheruddeen, S. M. and Sujatha, A., Record of *Deanolis albizonalis* (Hampson.) (Pyralidae: Odontinae) as mango fruit borer in Andhra Pradesh. *J. Bombay Natl. Hist. Soc*., 1993, **90**, 528.
11. Waterhouse, D. F., *Biological Control of Insect Pest: Southeast Asian Prospects*, ACIAR Monograph, 1998, vol. 5, p. 548.
12. Krull, S. M. E., Studies on the mango-ecosystem in Papua New Guinea with special reference to the ecology of *Deanolis sublimbalis* Snellen (Lepidoptera, Pyralidae) and to the biological control of *Ceroplastes rubens* (Homoptera, Coccidae). Ph D thesis, Justus-Liebig-Universitat Gieben, Germany, 2004, p. 190.
13. Pinese, B., Biology, damage levels and control of red-banded mango caterpillar in Papua New Guinea and Australia – project update. Australian Centre for International Agricultural Research (on-line), 2005; http://www.aciar.gov.au/web.nsf/doc/ACIA-68K3JQ (accessed on 11 January 2006).
14. Tenakanai, D., Dori, F. and Kurika, K., Red-banded mango caterpillar, *Deanolis sublimbalis* Snellen. (Lepidoptera: Pyralidae: Odontinae), in Papua New Guinea. In *Pest and Disease Incursions: Risks, Threats and Management in Papua New Guinea*, ACIAR Technical Reports, 2006, no. 62, pp. 161–165.
15. Fenner, T., Red-banded mango caterpillar: biology and control prospects. Northern Territory Department of Primary Industry and Fisheries, Queensland, 1997, p. 2.
16. Poelchau, M. F., Reynolds, J. A., Denlinger, D. L., Elsik, C. G., Armbruster, P. A. and Denovo, A., Transcriptome of the Asian tiger mosquito, *Aedes albopictus* to identify candidate transcripts for diapause preparation. *BMC Genomics*, 2011, **12**, 619.
17. Poelchau, M. F., Reynolds, J. A., Denlinger, D. L., Elsik, C. G. and Armbruster, P. A., Transcriptome sequencing as a platform to elucidate molecular components of the diapause response in the Asian tiger mosquito *Aedes albopictus*. *Physiol. Entomol*., 2013, **38**, 173–181.
18. Poelchau, M. F., Reynolds, J. A., Elsik, C. G., Denlinger, D. L. and Armbruster, P. A., Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, *Aedes albopictus*. *Proc. R. Soc. London, Ser. B*, 2013, **280**, 20130–20143.
19. Dong, Y., Desneux, N., Lei, C. and Niu, C., Transcriptome characterization analysis of *Bactrocera minax* and new insights into its pupal diapause development with gene expression analysis. *Int. J. Biol. Sci*., 2014, **10**, 1051–1063.
20. Cheolho, S. and Denlinger, D. L., Insulin signaling and FOXO regulate the overwintering diapause of the mosquito *Culex pipiens*. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 6777–6781.
21. Fukuda, S. and Takeuchi, S., Diapause factor-producing cells in the suboesophageal ganglion of the silkworm, *Bombyx mori* L. *Proc. Jpn. Acad*., 1967, **43**, 51–56.
22. Fukuda, S. and Takeuchi, S., Studies on the diapause factors producing cells in the suboesophageal ganglion of the silkworm, *Bombyx mori* L. *Embryologia*, 1967, **4**, 333–353.
23. Imai, K., Konno, T., Nakazawa, Y., Komiya, T., Isobe, M. and Koga, K., Isolation and structure of diapause hormone of the silkworm, *Bombyx mori*. *Proc. Jpn. Acad. Ser. B*, 1991, **67**, 98–101.
24. Kankare, M., Parker, D., Merisalo, M., Salminen, T. S. and Hoikkala, A., Transcriptional differences between diapausing and non-diapausing *D. montana* females reared under the same photoperiod and temperature. *PLoS ONE*, 2016, **11**, 0161852.
25. Ewels, P., Magnusson, M., Lundin, S. and Kaller, M., MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 2016, **32**, 3047–3048.
26. Bolger, A. M., Lohse, M. and Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014, **30**, 2114–2120.
27. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A. and Amit, I., Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol*., 2011, **29**, 644–652.
28. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J. and Bealer, K., BLAST+: architecture and applications. *BMC Bioinform.*, 2009, **10**, 421.
29. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J. and Zhang, Z., WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.*, 2006, **34**, 293–297.
30. Langmead, B., Aligning Short Sequencing Reads with Bowtie, *Curr. Protoc. Bioinform.*, 2010, **11**, 7; https://doi.org/10.1002/0471250953.bi1107s32.
31. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. and Homer, N., The Sequence Alignment/Map format and SAM tools. *Bioinformatics*, 2009, **25**, 2078–2079.
32. Love, M. I., Huber, W. and Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*., 2014, **15**, 550.
33. Maere, S., Heymans, K. and Kuiper, M., BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 2005, **21**, 3448–3449.
34. Hodek, I., Diapause development, diapause termination and the end of diapause. *Eur. J. Entomol.*, 1996, **93**, 475–487.
35. Tauber, M. J. and Tauber, C. A., Insect seasonality: diapause maintenance, termination, and post diapause development. *Annu. Rev. Entomol.*, 1976, **21**, 81–107.
36. Tauber, M. J. Tauber, C. A. and Masaki, S., *Seasonal Adaptations of Insects*, Oxford University Press, NY, USA, 1986.
37. Sujatha, A. and Zaheruddeen, S. M., Biology of Pyralid fruit borer *Deanolis albizonalis* (Hampson): a new pest of mango. *J. Appl. Zool. Res*., 2002, **13**, 1–5.
38. Fujiwara, K., Suzuki, M., Okumachi, Y., Okamura Ikeda, K., Fujiwara, T., Takahashi, E. I. and Motokawa, Y., Molecular cloning, structural characterization and chromosomal localization of human lipoyl transferase gene. *Eur. J. Biochem*., 1999, **260**, 761–767.
39. Arrese, E. L. and Soulages, J. L., Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol*., 2010 **55**, 207–225; doi:10.1146/annurev-ento-112408-085356.
40. Toprak, U., Hegedus, D., Dogan, C. and Guney, G., A journey into the world of insect lipid metabolism. *Arch. Insect Biochem. Physiol*., 2020, **104**(2), e21682.

41. Bheda, A., Shackelford, J. and Pagano, J. S., Expression and functional studies of ubiquitin C-terminal hydrolase L1 regulated genes. *PLoS ONE*, 2009, **4**, 6764.
42. Li, Y., Zhou, Z., Shen, M., Ge, L. and Liu, F., Ubiquitin-conjugating enzyme E2 E inhibits the accumulation of Rice Stripe virus in *Laodelphax striatellus* (Fallen). *Viruses*, 2020, **12**, 908.
43. Walters, K. J., Goh, A. M., Wang, Q., Wagner, G. and Howley, P. M., Ubiquitin family proteins and their relationship to the proteasome: a structural perspective. *Biochim. Biophys. Acta*, 2004, **1695**, 73–87.
44. Ikeda, F., Ubiquitin conjugating enzymes in the regulation of the autophagy-dependent degradation pathway. *Matrix Biol*., 2021, **100–101**, 23–29; doi:10.1016/j.matbio.2020.11.004.
45. Donaldson, T. D., Noureddine, M. A., Reynolds, P. J., Bradford, W. and Duronio, R. J., Targeted disruption of *Drosophila* Roc1b reveals functional differences in the Roc subunit of cullin-dependent E3 ubiquitin ligases. *Mol. Biol. Cell*, 2004, **15**, 4892–4903.
46. Wang, J., Ran, L. L., Li, Y. and Liu, Y. H., Comparative proteomics provides insights into diapause program of *Bactrocera minax* (Diptera: Tephritidae). *PLoS ONE*, 2020, **15**(12), e0244493.
47. Williams, K. D., Busto, M., Suster, M. L., So, A. K. C., Ben-Shahar, Y. and Leevers, S. J., Natural variation in *Drosophila melanogaster* diapause due to the insulin-regulated PI3-kinase. *Proc. Nat. Acad. Sci. USA*, 2006, **103**, 15911–15915.
48. Hao, K., Ullah, H., Jarwar, A. R., Nong, X. Q., Tu, X. B. and Zhang, Z. H., Molecular identification and diapause-related functional characterization of a novel dual-specificity kinase gene, MPKL, in *Locusta migratoria*. *FEBS Lett.*, 2019, **593**, 3064–3074.
49. Lin, J. L., Lin, P. L. and Gu, S. H., Phosphorylation of glycogen synthase kinase-3β in relation to diapause processing in the silkworm, *Bombyx mori*. *J. Insect Physiol.*, 2009, **55**, 593–598.
50. Kidokoro, K., Iwata, K., Takeda, M. and Fujiwara, Y., Involvement of ERK/MAPK in regulation of diapause intensity in the false melon beetle, *Atrachya menetriesi*. *J. Insect Physiol*., 2006, **52**, 1189–1193.
51. Fujiwara, Y. and Shiomi, K., Distinct effects of different temperatures on diapause termination, yolk morphology and MAPK phosphorylation in the silkworm, *Bombyx mori*. *J. Insect Physiol*., 2006, **52**, 1194–1201.
52. Hao, Y. J., Zhang, Y. J., Si, F. L., Fu, D. Y., He, Z. B. and Chen, B., Insight into the possible mechanism of the summer diapause of *Delia antiqua* (Diptera: Anthomyiidae) through digital gene expression analysis. *Insect Sci*., 2016, **23**, 438–451.
53. Duan, T. F., Li, L., Tan, Y., Li, Y. Y. and Pang, B. P., Identification and functional analysis of microRNAs in the regulation of summer diapause in *Galeruca daurica*. *Comp. Biochem. Physiol. D*, 2020, 100786.
54. Iwata, K. I., Fujiwara, Y. and Takeda, M., Effects of temperature, sorbital, alanine and diapause hormone on embryonic development in *Bombyx mori*: *in vitro* tests of old hypothesis. *Physiol. Entomol*., 2005, **30**, 317–323.
55. Fujiwara, Y., Shindome, C., Takeda, M. and Shiomi, K., The roles of ERK and P38 MAPK signaling cascades on embryonic diapause initiation and termination of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.*, 2006, **36**, 47–53.
56. Kidokoro, K., Iwata, K., Fujiwara, Y. and Takeda, M., Effects of juvenile hormone analogs and 20-hydroxyecdysone on diapause termination in eggs of *Locusta migratoria* and *Oxya yezoensis*. *J. Insect Physiol*., 2006, **52**, 473–479.
57. Fujiwara, Y., Tanaka, Y., Iwata, K., Rubio, R. O., Yaginuma, T., Yamashita, O. and Shiomi, K., ERK/MAPK regulates ecdysteroid and sorbitol metabolism for embryonic diapause termination in the silkworm, *Bombyx mori*. *J. Insect Physiol*., 2006, **52**, 569–575.
58. Fujiwara, Y. and Denlinger, D. L., MAPK is a likely component of the signal transduction pathway triggering rapid cold hardening in the flesh fly *Sarcophaga crassipalpis*. *J. Exp. Biol*., 2007, **210**, 3295–3300.
59. Cronan, J. E., Progress in the enzymology of the mitochondrial diseases of lipoic acid requiring enzymes. *Front. Genet*., 2020, **11**, 510.
60. Clark, E. W. and Chadbourne, D. S., A comparative study of the weight, lipid, and water content of the pink bollworm. *Ann. Entomol. Soc. Am*., 1962, **55**, 225–228.
61. Hahn, D. A. and Denlinger, D. L., Meeting the energetic demands of insect diapause: nutrient storage and utilization. *J. Insect Physiol*., 2007, **53**, 760–773.
62. Griffin, M. and Sul, H. S., Insulin regulation of fatty acid synthase gene transcription: roles of USF and SREBP-1c. *IUBMB Life*, 2004, **56**, 595–600.
63. Alabaster, A., Isoe, J., Zhou, G., Lee, A., Murphy, A., Day, W. A. and Miesfeld, R. A., Deficiencies in acetyl-CoA carboxylase and fatty acid synthase 1 differentially affect eggshell formation and blood meal digestion in *Aedes aegypti*. *Insect Biochem. Mol. Biol*., 2011, **41**, 946–955.
64. Majerowicz, D. and Gondim, K. C., *Insect Lipid Metabolism: Insights into Gene Expression Regulation. Recent Trends in Gene Expression*, Nova Science Publishers, 2013, pp. 147–190.
65. Xue, F., Spieth, H. R., Li, A. and Ai, H., The role of photoperiod and temperature in determination of summer and winter diapause in the cabbage beetle, *Colaphellus bowringi* (Coleoptera: Chrysomelidae). *J. Insect Physiol*., 2002, **48**, 279–286.
66. Qi, X., Zhang, L., Han, Y., Ren, X., Huang, J. and Chen, H., *De novo* transcriptome sequencing and analysis of *Coccinella septempunctata* L. in non-diapause, diapause and diapauses terminated states to identify diapauses associated genes. *BMC Genomics*, 2015, **16**, 1086.

ACKNOWLEDGEMENTS. We thank the Director, Indian Institute of Horticultural Research, Bengaluru for providing the necessary research facilities. The present work was financially supported by Dr Y.S.R. Horticultural University, Venkataramannagudem and an ICAR-funded National Fellow Project. We also thank Dr Rajshekhar Reddy, Chief Scientific Officer, Clevergene, Benglauru for inputs on bioinformatic analysis of the data, and H. K. Maurya, Narayana Olympaid, Bengaluru for final editing of the draft.

Received 9 March 2022; revised accepted 18 May 2022

doi: 10.18520/cs/v123/i3/471-481