

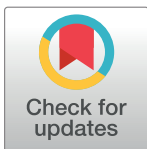
RESEARCH ARTICLE

Differential transcriptome analysis reveals genes related to cold tolerance in seabuckthorn carpenter moth, *Eogystia hippophaecolus*

Mingming Cui¹, Ping Hu¹, Tao Wang², Jing Tao^{1*}, Shixiang Zong^{1*}

1 Beijing Key Laboratory for Forest Pest Control, Beijing Forestry University, Beijing, P.R. China, **2** Mentougou Forestry Station, Beijing, China

* taojing1029@hotmail.com (JT); zongsx@126.com (SXZ)



OPEN ACCESS

Citation: Cui M, Hu P, Wang T, Tao J, Zong S (2017) Differential transcriptome analysis reveals genes related to cold tolerance in seabuckthorn carpenter moth, *Eogystia hippophaecolus*. PLoS ONE 12(11): e0187105. <https://doi.org/10.1371/journal.pone.0187105>

Editor: J Joe Hull, USDA Agricultural Research Service, UNITED STATES

Received: June 22, 2016

Accepted: October 15, 2017

Published: November 13, 2017

Copyright: © 2017 Cui et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data are available from the SRA database (accession number SRR4409152) and the GEO database (accession number GSE92686).

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 31470651).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Seabuckthorn carpenter moth, *Eogystia hippophaecolus* (Lepidoptera: Cossidae), is an important pest of sea buckthorn (*Hippophae rhamnoides*), which is a shrub that has significant ecological and economic value in China. *E. hippophaecolus* is highly cold tolerant, but limited studies have been conducted to elucidate the molecular mechanisms underlying its cold resistance. Here we sequenced the *E. hippophaecolus* transcriptome using RNA-Seq technology and performed *de novo* assembly from the short paired-end reads. We investigated the larval response to cold stress by comparing gene expression profiles between treatments. We obtained 118,034 unigenes, of which 22,161 were annotated with gene descriptions, conserved domains, gene ontology terms, and metabolic pathways. These resulted in 57 GO terms and 193 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. By comparing transcriptome profiles for differential gene expression, we identified many differentially expressed proteins and genes, including heat shock proteins and cuticular proteins which have previously been reported to be involved in cold resistance of insects. This study provides a global transcriptome analysis and an assessment of differential gene expression in *E. hippophaecolus* under cold stress. We found seven differential expressed genes in common between developmental stages, which were verified with qPCR. Our findings facilitate future genomic studies aimed at improving our understanding of the molecular mechanisms underlying the response of insects to low temperatures.

Introduction

Insects are poikilothermic animals and thus temperature is one of the most important abiotic factors influencing their distribution and behavior [1,2]. Nearly every aspect of an insect's life is affected by temperature, including survival, development, reproduction, and longevity [3–6]. Nevertheless, insects have evolved to flourish in a wide variety of thermal environments [3]. They have achieved this through the natural selection of a variety of strategies to cope with

changes in ambient temperatures. Some insects overwinter by migrating to warmer climes or by moving to protected habitats [7–9]. Most insects, however, survive through the winter as a result of a variety of physiological and biochemical adaptations that can be grouped into two general strategies: freeze intolerance or freeze tolerance [10–12]. Freeze-intolerant insects are very sensitive to ice formation. They avoid freezing by supercooling to survive at low temperatures [1]. They achieve this by reducing intra-cellular water content, synthesizing cryoprotectants, and actively removing ice-nucleating agents (INAs) [7,12,13]. Compared with freeze intolerance, the freeze-tolerant strategy is more advanced because it permits extracellular ice formation within tissues. Ice formation is triggered by INAs such as insect ice nucleating proteins (INPs), which are often lipoproteins [14, 15].

To date, considerable research has been conducted on the biochemical mechanisms of insect cold resistance. Increasingly, genomic and proteomic screening methods are being applied to discover the genes and proteins that enable winter survival, which is greatly broadening our understanding of important molecular processes. Two key proteins have so far been recognized to directly contribute to insect cold hardiness: anti-freeze proteins (AFPs) and INPs [16]. Other proteins contribute to cold hardiness indirectly, including heat-shock proteins (HSPs), antioxidant enzymes, and enzymes involved in cryoprotectant synthesis and degradation [17]. In the transcriptomic studies of Dunning et al., (2013) and Dennis et al., (2015), a set of structural cuticular genes were identified to be associated with stick insect adaptation to the alpine environment. [18, 19]. They also identified three candidate genes that responded to low temperature: prolyl 4-hydroxylase subunit alpha-1, staphylococcal nuclease domain-containing protein 1 and a cuticular protein gene [20]. Durant et al's (2015) research on transcription profiling of the brain in *Ceratina calcarata* showed that changes in gene expression associated with overwintering, included an underrepresentation of genes related to muscle fibers and an overrepresentation of genes related to lipid metabolic processes [21].

Because RNA transcript levels generally reflect physiological demand, changes in transcript abundance should identify genes involved in an organism's response to extrinsic stimuli [22]. Transcriptome analysis using Next Generation technologies has been proven to be a powerful method for exploring the underlying basis of the cold response in a variety of insect lineages. Examples include the study of transcriptional profiling of overwintering gene expression in a carpenter bee [21], the identification of cold-responsive genes in an alpine stick insect [20], the analysis of divergent transcriptional responses in alpine and lowland stick insects [18], and the study of cold adaptation in eight genera stick insects [19]. Compared with traditional differential genes expression methods such as expressed sequence tags (EST), microarrays, serial analysis of gene expression (SAGE), or massively parallel signature sequencing (MPSS), a complete transcriptome sequence and analysis can be completed for any species with high efficiency without a reference genome. Moreover, transcripts normally expressed at very low levels can be detected and quantified [23–26]. While transcriptome analysis have been used in studies of the cold response of *Drosophila melanogaster* and *D. subobscura* [27–29], its application to study cold hardiness in other insects has been quite limited.

The seabuckthorn carpenter moth *Eogystia hippophaecolus* (Hua, Chou, Fang & Chen, 1990) is the dominant boring insect pest of sea buckthorn (*Hippophae rhamnoides* L.). Sea buckthorn is a shrub that is distributed across northern and western China. The plant grows well in sandy soils in cold climates. It has been extensively planted to prevent soil erosion and desertification, and also to serve as a source for food and medicine products [30]. *E. hippophaecolus* completes a generation every three to four years and most of its life history is spent in 16 larval stages lasting about 31 months [31]. Eggs are mainly deposited in wounds or in cracks of the bark. The egg stage lasts 7–30 days. The early larvae feed on trunk phloem at first. Small larvae often congregate in large numbers and later migrate to the base of the trunk or to the

roots before winter. Larvae are very cold tolerant and can survive in temperatures lower than -30°C . The larvae obstruct water transport in sea buckthorn plants by burying into the trunk and roots. The shape of the galleries is irregular and most are not connected to each other. Each gallery is filled with sawdust and fecula. The mature larvae will climb out of the worm-hole, and then burrow 15 cm deep into the soil in June. The pupal stage lasts about 30 days and is followed by eclosion, mating, egg-laying, and hatching in July, followed by overwintering in October [32]. Its voraciousness, high reproduction rate, and inconspicuous habit make the pest difficult to control. In China, seabuckthorn carpenter moth is considered a major threat to sea buckthorns, and it can also feed on other plants, such as elms [33,34]. In preliminary studies measuring physiological indices (e.g., SCP, freezing point, etc.) of cold resistance in *E. hippophaecolus*, we classified the seabuckthorn carpenter moth as a highly freeze-tolerant organism [35].

In this study, we used RNA-Seq and *de novo* transcriptome assembly to generate transcriptomes and examine the changes in the regulation of transcription associated with cold treatment. Detailed differential expression analysis revealed a number of candidate genes that are potentially related to the cold tolerance of *E. hippophaecolus*. Our results provide new genetic information about this insect pest and provide a rich resource for the discovery and identification of novel genes involved in the cold stress response.

Materials and methods

Insect material and RNA extraction

E. hippophaecolus larvae were collected from Jianping County, Chaoyang City, Liaoning Province, China, in late December 2014. The latitude and longitude of the collection site are 119.637°E and 41.3954°N . Given the long life history and many larval stages of this species, we chose three developmental stages (seventh instar, tenth instar, and thirteenth instar) [31] and two treatment temperatures to characterize variation in gene expression. Healthy larvae in different developmental stages were treated as presented in Table 1. Prior to treatment, we cleaned the moths with sterile water and starved them for three days. All of the instruments used to handle the moths were sterilized before use. After housing for 48 h at room temperature (25°C), three samples of each development stage were treated differently: one stayed at room temperature as a control and the other two were kept in 5°C and -5°C , respectively, for 10 h, and then snap-frozen in liquid nitrogen and stored at -80°C until the RNA extraction step. Total RNA was extracted using TRIzol reagent (Ambion) and the RNeasy Plus Mini Kit (No. 74134; Qiagen, Hilden, Germany) following the manufacturer's instructions. A NanoDrop 8000 (Thermo, Waltham, MA, USA) was used to quantify the RNA. The mRNA samples were purified and fragmented using the TruSeq RNA Sample Preparation Kit v2-Set A (No. RS-122-2001; Illumina, San Diego, CA, USA) to remove rRNA. After additional quality

Table 1. Schematic table of the experimental design used to investigate effects of cold stress.

Developmental stage	control group		Treatment group	
	25°C		5°C	5°C
7th instar	T1		T4	
10th instar	T2		T5	
13th instar	T3		T6	

Note: There was one larva in each different treatment. After treatment, the total RNA of larvae under 5°C and 5°C were mixed for sequencing library construction. Samples were named T1–T6.

<https://doi.org/10.1371/journal.pone.0187105.t001>

control, cDNA library construction, Illumina sequencing, and *de novo* transcriptome assembly were performed by the Beijing Biomarker Biotechnology Co. (Beijing, China). In view of the costs, 5°C and -5°C samples were pooled by instar to construct the cDNA library. This method has been used in previous transcriptome studies [36, 37]. No specific permits were required for the described field studies. The location is not privately owned or protected, and the field studies did not involve endangered or protected species.

Illumina sequencing and *de novo* transcript assembly

We sequenced the *E. hippophaecolus* transcriptome on an Illumina HiSeq 2500 platform and performed *de novo* assembly from the short paired-end reads (200bp). Before transcript assembly, all raw reads were processed to remove low-quality and adaptor sequences by Trimmomatic 0.32 (<http://www.usadellab.org/cms/index.php?page=trimmomatic>) using the following parameters (Phred64 LEADING:3 TRAILING:3 SLIDING WINDOW:4:15 MINLEN:36), while the unpaired reads were discarded [38]. The clean reads were then assembled into contigs and the contigs were subsequently combined into transcripts based on paired-end information using the Trinity platform (Version: r2013.11.10) [39]. CD-HIT was used to exclude redundant sequences and the clustering was implemented using TGICL (2.1) [40] to combine sequences with more than 90% similarity and an overlapping length longer than 35bp. The completeness assessment was performed by BUSCO (version 2), using the insecta lineage [41]. The transcripts were blasted to various RNA databases including Silva, GtRNadb by BLAST software (version 2.2.26) for filtering ($e\text{-value} \leq 10^{-5}$). After filtering, ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), microRNA (miRNA), small nucleolar RNA (snoRNA), repeat sequences, and other non-coding RNA (ncRNA) were removed. The list of the sequences identified as ncRNA has been shown in S1 Table.

The raw data has been submitted to the NCBI Sequence Read Archive (SRA) with accession number SRR4409152. The mapped read counts have been submitted to the Gene Expression Omnibus (GEO) with accession number GSE92686.

Gene functional annotation

We utilized BLAST software (version 2.2.26) [42] to obtain homologous sequences ($e\text{-value} \leq 10^{-5}$) by searching several public protein databases, including the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database (downloaded on 02/2015) [43], the Swiss-Prot protein database 51.0 [44], the Gene Ontology (GO) database (downloaded on 12/2014) [45], the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database 73.1 [46], and the Cluster of Orthologous Groups (COG) database (downloaded on 2013) [47]. We used the single best blast hit only to identify the contig and to extract GO information, while non-arthropod contigs were filtered out.

After predicting the amino-acid sequences of unique genes by using Transdecoder (downloaded on 2015), we obtained functional information about the unigenes by comparing these to the Pfam database using HMMER software ($e\text{-value} \leq 10^{-10}$) [48].

Differential gene expression analysis

Bowtie (4.4.7) [49] was used to align reads to reference sequences. Gene expression levels were measured in the RNA-Seq analysis as reads per kilobase of exon model per million mapped reads (RPKM) [50]. The EBSeq software was used to identify differentially expressed genes in pair-wise comparisons and separate models were fit for each development stage [51]. The results of all statistical tests were corrected for multiple testing with the Benjamini–Hochberg false discovery rate with the following parameter values: FDR < 0.01. Sequences were deemed

to be significantly differentially expressed if the adjusted P value obtained by this method was <0.001 and there was at least a twofold change (>1 or <-1 in \log_2 ratio value which were calculated by the average RPKM value of treatment libraries divided by the average RPKM value of control libraries). K-Means clustering of gene expression was performed using the R Package software v.3.1.3 to examine the relationships between samples. GO annotation and KEGG pathway analysis were also performed to assess the differential expression results. To annotate the gene with gene ontology (GO) terms, the nr BLAST results were imported into the Blast2GO program (Conesa, Gotz et al. 2005). GO annotations for the genes were obtained by Blast2GO. This analysis mapped all of the annotated genes to GO terms in the database and counted the number of genes associated with each term. A Perl script was then used to plot GO functional classification for the unigenes with a GO term hit to view the distribution of gene functions. The gene sequences were also aligned to the Clusters of Orthologous Group (COG) database to predict and classify functions (Tatusov, Galperin et al. 2000). We used KOBAS (<http://kobas.cbi.pku.edu.cn/home.do>) to annotate a set of sequences with KO terms and identify both the most frequent and the statistically significantly enriched pathways. KEGG pathways were tested for enrichment using the right-sided Fisher's exact test. We removed the T3 transcript and its corresponding control, T6, from the DEG analysis because of large, unexplained variation or low correlation between the T3 and other samples.

Gene expression validation by real-time quantitative PCR. The seven genes DE in response to cold treatment in both developmental stages were chosen for validation using qRT-PCR. Total RNA was reverse transcribed using the PrimeScript RT Reagent Kit with gDNA Eraser (No. RR047A; TaKaRa, Shiga, Japan). Gene-specific primers were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>). qRT-PCR was performed on a Bio-Rad CFX96 real-time PCR System (Hercules, CA, USA) using SYBR Premix Ex-Taq II (No. RR820A; TaKaRa), according to the manufacturer's instructions.

The template cDNAs were diluted 20-fold with nuclease-free deionized water. Of this, 2 μl was added to the reaction mixture (25 μl), which included 12.5 μl of SYBR Premix Ex Taq II, 1 μl of each primer (10 μM), and 8.5 μl of ddH₂O. The PCR was performed with the following thermal cycle: 95°C for 30 s, 40 cycles at 95°C for 5 s, 60°C for 30 s and 65°C to 95°C in increments of 0.5°C for 5 s to generate melting curves. Each reaction was performed on three biological and four technical replicates. The average threshold cycle was calculated to estimate the relative gene expression levels using the $2^{-\Delta\Delta\text{Ct}}$ method [52]. The data were expressed as the mean \pm SE, and all primer information is listed in S2 Table. A t-test was performed for the qPCR data. β -Actin was used as the reference gene to normalize target gene expression [53].

Results

Sequencing and assembly

A total of 39.00 Gb of clean data passed the Illumina quality filter after transcriptome sequencing of six cDNA samples. The number of the paired-end reads and the length of these reads are provided in S3 Table. The Q30 and GC contents were 88.30% and 44.93%, respectively. The quality control result of six samples was seen in S3 Table. The *de novo* assembly using Trinity software and standard parameter values yielded 191,200 transcripts (none shorter than 200 bp) with an N50 of 1,687 bp. The unigene dataset included 118,034 sequences with an N50 of 904bp. The length distribution of unigenes closely followed the length distribution of transcripts (S1 Fig). Coverage of clean reads mapped to the assembly was provided in S3 Table. The completeness assessment result showed that 90.0% of BUSCO genes were "complete", of which 84.3% were single-copy BUSCOs and 5.7% were duplicated BUSCOs. 6.2% were "fragmented", while the remaining 3.8% were "missing".

Table 2. Statistics of annotation analysis of unigenes.

Databases	Unigene	300<length<1000	length≥1000
COG_Annotation	4,452	692	3,272
GO_Annotation	10,525	2,613	5,578
KEGG_Annotation	7,611	1,708	4,257
KOG_Annotation	10,784	2,266	6,763
Pfam_Annotation	12,413	2,893	7,857
Swissprot_Annotation	21,452	6,629	9,654
nr_Annotation	21,232	6,540	9,621
All	22,161	6,967	9,723

<https://doi.org/10.1371/journal.pone.0187105.t002>

Annotation of predicted proteins

After annotation, genes with a significant blast hit to arthropods were identified. In total, 22,161 unigenes were found in at least one public database, such as the nr database, the SwissProt protein database, the GO database, the KEGG database, and the COG database. The SwissProt database (21,452 annotated unigenes, 21.47%) had the most matches, followed by the nr database (21,232, 21.25%). A total of 77,735 (77.82%) unigenes were not found in the public databases (Table 2). Overall, most of the unigenes either could not be annotated or their descriptions were uninformative (e.g., putative, unknown, hypothetical, or unnamed protein). Among the annotated unigenes, more than 70% of them matched with other lepidopteran species. The best homologous species match in the nr database for each sequence is shown in S2 Fig. The top matched species were the silkworm (*Bombyx mori*), with 8627 unigenes (40.73%), followed by the monarch butterfly (*Danaus plexippus*). The blast information for each contig was provided in S4 Table.

Global functional classification of unigenes by GO and KEGG

One unigene can have different functional annotations; therefore, 38,435 unigenes in total with a match in the GO database were classified by their GO terms. The unigenes could be categorized into 57 functional sub-groups of the three main GO groups: 15,798 unigenes fell in the biological process group, 6,721 in the cellular component group, and 15,916 in the molecular function group (Fig 1). The most frequent GO terms were: metabolic process (6,071 unigenes), binding (5,492 unigenes), cellular process (5,258 unigenes), catalytic activity (5,067 unigenes), single-organism process (4,354 unigenes), cell part (3,159 unigenes), cell (3,156 unigenes).

Pathway enrichment analysis based on the KEGG database was used to identify enriched metabolic pathways or signal transduction pathways. In total, 5,990 unigenes were assigned to 193 pathways (Fig 2). The pathway categories with the largest number of unigenes were purine metabolism and ribosome (both 208 unigenes), followed by oxidative phosphorylation (174), RNA transport (170), Protein processing in endoplasmic reticulum (158), spliceosome (138), carbon metabolism (137), peroxisome (133), and lysosome (115).

DEGs in *E. hippophaecolus* larvae under cold stress

The correlations between samples were shown in S3 Fig. Based on the correlations, the clustering of samples was presented in S4 Fig. It can be seen sample T3 was not correlated well with other samples (r^2 all less than 0.6). Use of such samples is likely to result in the spurious detection of DEG, and so sample T3 and its corresponding control, T6, were excluded from the

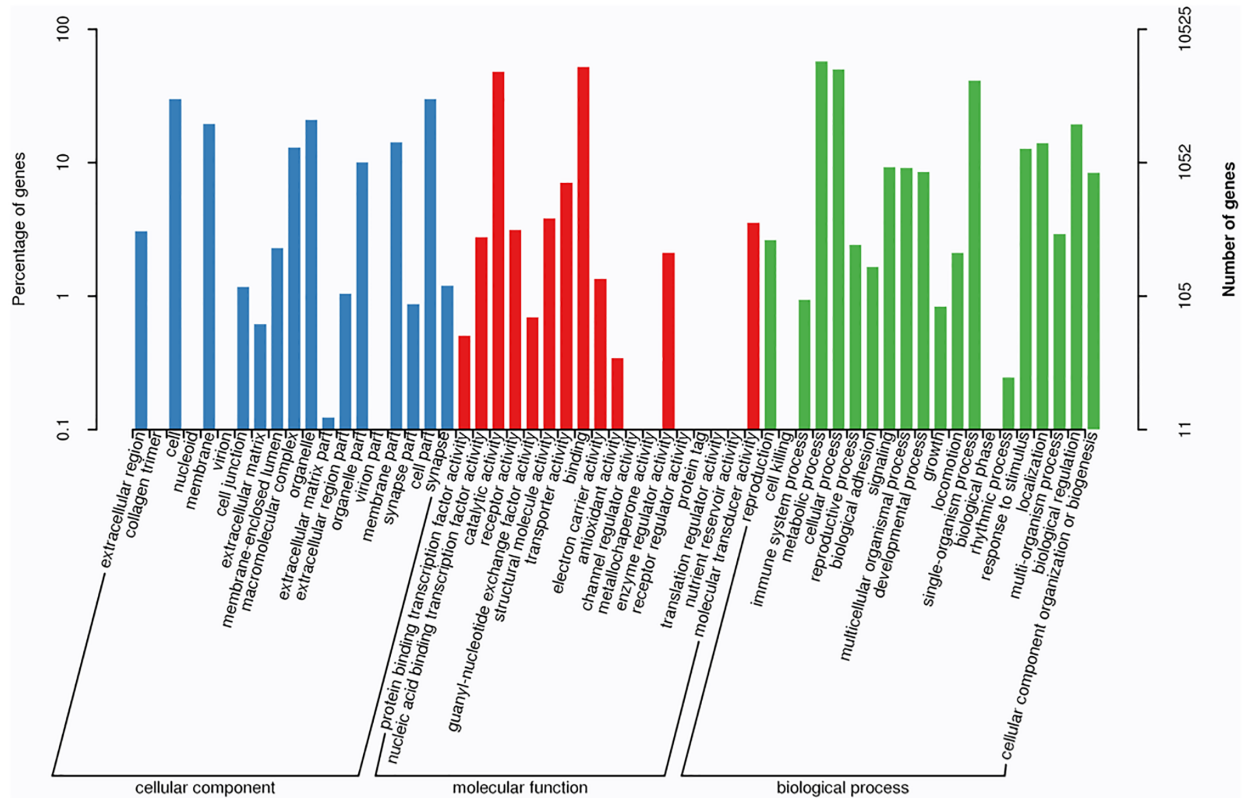


Fig 1. Gene ontology (GO) annotations of the annotated unigenes. 38,435 unigenes were assigned to three GO categories containing 57 functional subcategories. The y-axis on the right shows the number of genes in category, and that on the left shows the percentage of a specific category of genes in that main category.

<https://doi.org/10.1371/journal.pone.0187105.g001>

DEG analysis. The remaining two groups (T1-T4, T2-T5) were then used to explore the genes involved in cold resistance. The DE information for each contig was provided in [S5 Table](#).

We applied a rigorous procedure to identify DEGs (see [Methods](#)) and identified 427 unigenes differentially expressed between group T1 and T4, and 71 between group T2 and T5

- Purine metabolism(3.47%)
- Ribosome(3.47%)
- Oxidative phosphorylation(2.90%)
- RNA transport(2.83%)
- Protein processing in endoplasmic reticulum(2.63%)
- Spliceosome(2.30%)
- Carbon metabolism(2.28%)
- Peroxisome(2.22%)
- Other(75.94%)

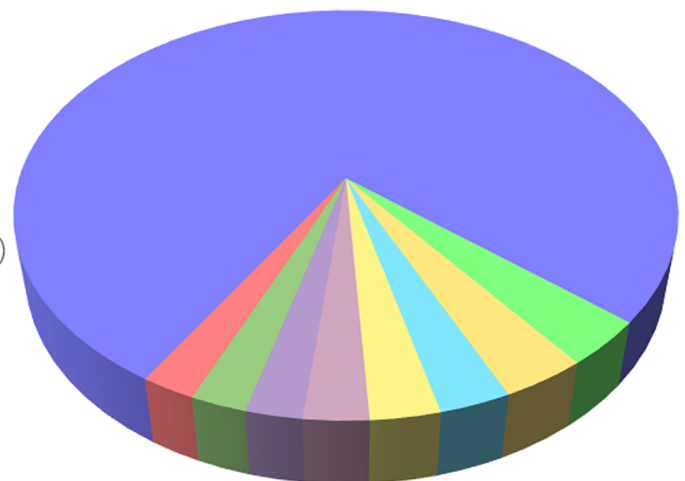


Fig 2. Pathway assignment based on KEGG mapping.

<https://doi.org/10.1371/journal.pone.0187105.g002>

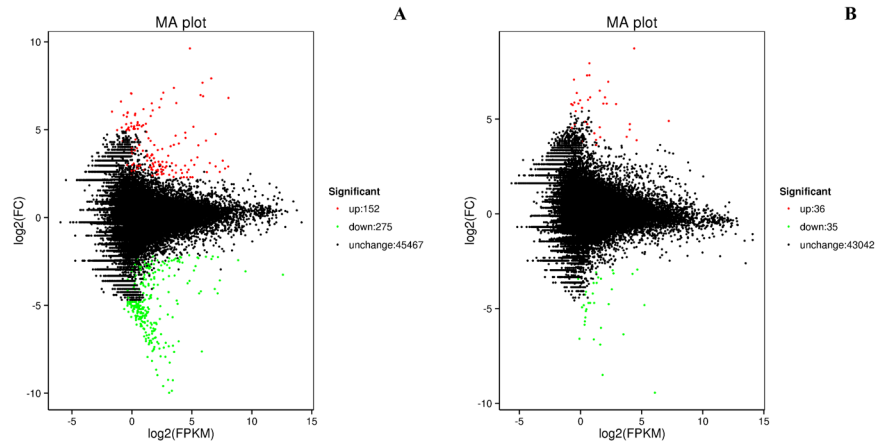


Fig 3. Distribution of DEGs between samples. (A) Distribution of DEGs between T1 and T4. (B) Distribution of DEGs between T2 and T5.

<https://doi.org/10.1371/journal.pone.0187105.g003>

(Fig 3). About half of the DEGs had annotations. The annotations of the DEGs were shown in S6 Table. Only seven DEGs were shared by the two groups, three of which were annotated. The blast results of these seven genes were shown in Table 3. GO annotation was used to evaluate the function of DEGs that were significantly different between treatments (Fig 4). Between T1 and T4, the most significant items were “Biological Process: metabolic process”, “Molecular Function: catalytic activity”, “Biological Process: cellular process”, “Molecular Function: binding”, and “Cellular Component: cell part” (Fig 4a). When comparing T2 and T5, unigenes were assigned to “Biological Process: metabolic process”, “Molecular Function: binding”, “Molecular Function: catalytic activity”, “Biological Process: cellular process” and “Molecular Function: structural molecule activity” (Fig 4b).

Table 3. Common DEGs differentially expressed genes in two groups (T1-T4 and T2-T5) from RNAseq.

Unigene ID	Blast Blastx Match				Log ₂ FC ^a	Regulated ^a	Log ₂ FC ^b	Regulated ^b
	annotation	Acc. Number	E-value	Identity				
c79869.graph_c1	endonuclease-reverse transcriptase [<i>Bombyx mori</i>]	ADI61830.1	0	64%	3.567980225	up	3.829149437	up
c76833.graph_c0	PREDICTED: uncharacterized protein LOC105570877 [<i>Vollenhovia emeryi</i>]	XP_011883719.1	1.00E-152	47%	3.587111514	down	4.534550285	up
c73192.graph_c0	PREDICTED: uncharacterized protein LOC106708429 isoform X1 [<i>Papilio machaon</i>]	XP_014355420.1	0	64%	6.026349523	up	7.944625711	up
c59923.graph_c0	putative axonemal leucine-rich repeat protein [<i>Danaus plexippus</i>]	EHJ70571.1	5.00E-128	67%	2.600939044	normal	2.969087192	normal
c33229.graph_c0	--	--	--	--	9.625223032	up	8.727798365	up
c36418.graph_c0	--	--	--	--	2.841664353	up	5.818570542	up
c68914.graph_c0	--	--	--	--	2.459464841	up	6.968321714	up

Note: FC^a indicates the ratio of RPKM of T4 sample divided by the RPKM of T1 sample; FC^b indicates the ratio of RPKM of T5 sample divided by the RPKM of T2 sample; Regulated^a indicates the expression level of DEGs in T4 compared to T1, including three situations, up-regulated, down-regulated or kept unchanged; other columns are functional annotations of unigenes; Regulated^b indicates the expression level of DEGs in T5 compared to T2; “--” means no information.

<https://doi.org/10.1371/journal.pone.0187105.t003>

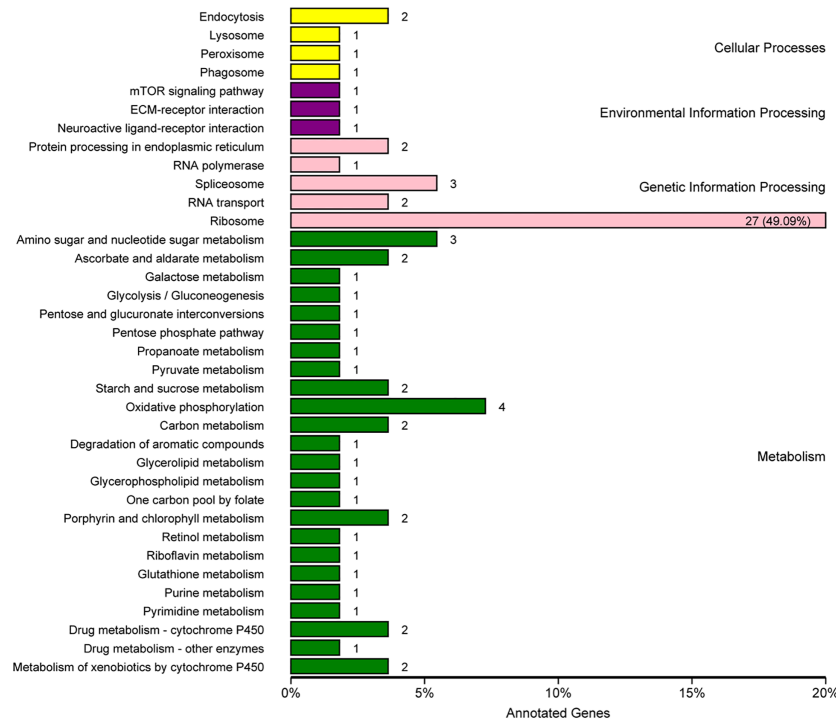


Fig 5. Pathways enriched for DEGs in T1-T4. The y-axis indicates the pathways and the x-axis indicates the percentage of unigenes in each pathway.

<https://doi.org/10.1371/journal.pone.0187105.g005>

QRT-PCR validation

This experiment involved RT-PCR (reverse transcription PCR), gel electrophoresis, Sanger sequencing, and qPCR of seven common DEGs identified in the RNA sequence data. The non-redundant transcripts from the transcriptome data and the RT-PCR fragments showed high sequence identities, validating the accuracy of the Illumina sequencing and *de novo* assembly. The qPCR results are presented in Fig 6. It can be seen that expression changes were in the same direction for the qPCR, but the magnitude differed. The FKPM of these genes were shown in Table 3. In general, the results were consistent with the Illumina sequencing data, corroborating the reliability and accuracy of the transcriptome analysis. This ensures the feasibility and sustainability of our further research on these or other DEGs from the transcriptome data.

Discussion

In this study, we conducted a comprehensive transcriptome analysis and characterized the gene expression profiles of *E. hippophaecolus* under cold stress. Through the analysis of DEGs sets, we identified transcriptome changes in *E. hippophaecolus* larvae.

From the assembly result, it can be seen that the length of 57.79% of the transcripts and 73.76% of the unigenes are less than 500 bp. N50 length of the transcripts is 1,687 bp, while the length of the unigenes is 904 bp. The completeness assessment by BUSCO indicates the comparatively high quality of the assembly according to the study of Simão et al. (2015) [41]. Our results are in line with other insect transcriptome projects using Illumina technology [36, 54].

Only 22.39% of the unigenes produced a significant blast hit to an arthropod sequence in the nr database. This can be partly attributed to the limited information available on Lepidoptera

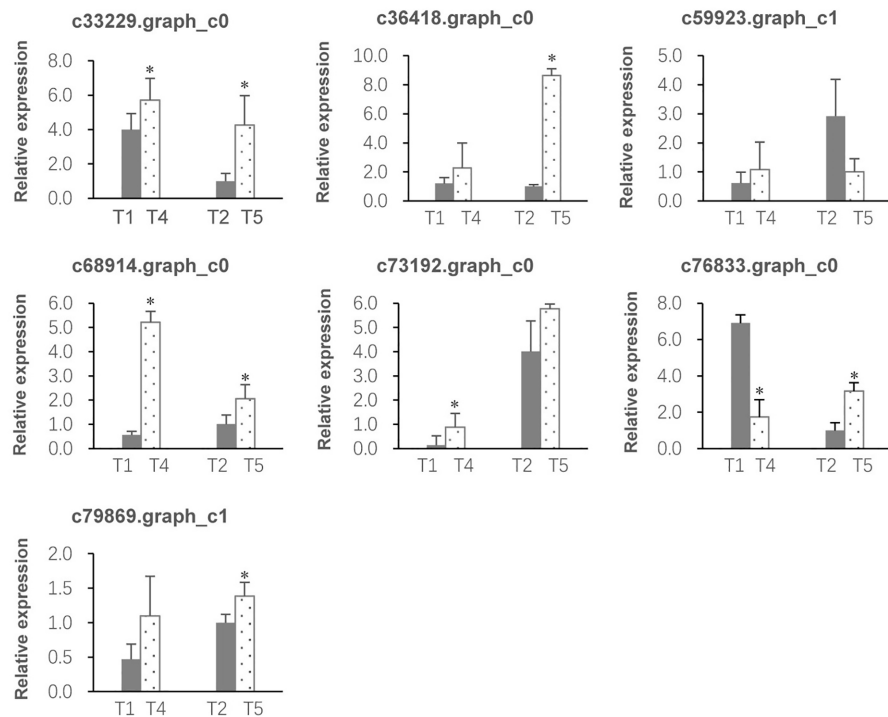


Fig 6. Validation of the DEGs by qRT-PCR. c33229.graph_c0~ c79869.graph_c1 are unigene IDs. T1&T2 represent the control groups; T4&T5 represent the low temperature groups. Relative expression values are presented as the mean ± SE of three biological replicates. Differences statistically significant are indicated by asterisk ($P < 0.05$).

<https://doi.org/10.1371/journal.pone.0187105.g006>

transcriptomes and genomes. Moreover, some genes in this moth could be highly diverged, thereby exceeding the blast threshold. Unsurprisingly, most of the *E. hippophaecolus* unigenes matched sequences from other lepidopteran species.

In this study, DEG libraries from cold treated and untreated *E. hippophaecolus* larvae were compared to identify functional genes at the whole-organism level. There were small-scale changes in gene expression. The number of DEGs from the T1-T4 comparison was larger than in the T2-T5 comparison. This difference may due to the different stages of larvae. Young larvae of *E. hippophaecolus* may not be as innately cold tolerant as older larvae to resist cold, so they may have to regulate related genes intensively to compensate. It has been previously noted in *Drosophila montana* that genes implicated in increasing cold tolerance are different between diapausing and non-diapausing flies, and that this could be due to different energy budgets at different life stages [55]. There were seven DE genes in common between developmental stages. Little previous work has been conducted on these genes, but our work suggests they are important for cold tolerance. Although there was little overlap in the genes between stages, there were some overlaps in the processes involved. Processes such as “binding”, “RNA binding”, “structural constituent of cuticle”, “N-acetylmuramoyl-L-alanine amidase activity” and “peptidoglycan catabolic process”, in which genes involved were all up-regulated; “nucleic acid binding” and “nucleic acid binding transferase activity” in which genes involved were all down-regulated were found in common through GO analysis.

From the pathway analysis, we found some *hsp*s were differentially expressed. HSPs are well-known chaperone proteins that play important roles in cold hardiness and organism responses to other stressors (e.g., heat, low oxygen levels, UV radiation, bacterial and viral

infection, heavy metals) that can affect the folding and functional conformation of proteins [56, 57]. That not all HSP genes were upregulated is consistent with previous studies [58, 59] and highlights that the mechanisms regulating these genes need further research. The “oxidative phosphorylation” pathway, which involves several enzymes, including cytochrome oxidase subunit, NADH dehydrogenase subunit, and ATP synthase subunit, was significantly downregulated under cold treatment. Antioxidant defense is a cell preservation strategy in response to environmental challenges [60] (e.g., heat, cold, UV radiation, osmotic challenge, toxins). Genes associated with the “metabolism of xenobiotics by cytochrome P450” pathways were modestly downregulated. It has been previously reported that cytochrome P-450 is related to temperature regulation [61].

In addition, genes for a number of cuticular proteins were also upregulated. These proteins are essential during insect cuticle formation and in the development of insect body structures and organs. Previous studies have shown that some insects upregulate the expression of cuticular protein genes to resist cold or other environmental stressors [19, 62–64]. To investigate the changes of structure or amount of the cuticle in response to cold could be an avenue for future work.

In this research, we also discovered many DEGs and the roles they play in cold tolerance in insects are unknown. Though it is hard to elucidate the molecular mechanisms underlying a complex process from the action of a few genes, the functions of these genes are worth being verified in future studies.

Curiously, the proteins most associated with insect cold hardiness, AFPs, were not found in the transcriptome profile. There may be several reasons for this. Firstly, although AFPs have been well studied, only about 200 insect AFP genes have been registered in GenBank. Secondly, the relationship between mRNA transcript levels and protein levels is complex. Integrated proteome/transcriptome studies have shown that mRNA levels are poorly associated with protein levels for all but the most abundant proteins [65]. Thirdly, it may be that more genes/proteins would be stimulated or expressed in the chill recovery phase [56]. Fourthly, genes involved in increasing cold tolerance have been shown to differ even between closely related species [66]. Finally, it is generally presumed that AFPs play a greater role in the cold resistance of freeze intolerant insects; for freeze tolerant species, on the other hand, AFPs may not be important [12].

During the transcriptome construction, we took into account the development stage of the moth, but neglected the replicates, thus might cause problems such as false positives when conducted DE analysis. Mixing samples in one RNA-seq sample may cause inconsistent of gene expression. The parallel experiment within different development stages could provide a reference when we choose interesting genes, thus we focus on the common DEGs in the parallel tests. Moreover, further validation results of the DEGs using qRT-PCR showed the consistency of the transcriptome data. This study provides a profile of the candidate genes that are suspected to be involved with insect cold hardiness. Many further studies should be carried out to test and verify the function the candidate genes.

In conclusion, this study is the first to characterize the transcriptome of *E. hippophaecolus* larvae without a reference genome. Our data will facilitate future genomic research and in-depth annotation of insect genomes. Moreover, the preliminary identification of genes and pathways exhibiting differential expression under cold stress may further our general understanding of cold resistance in insects.

Supporting information

S1 Fig. Statistics of *de novo* assembly of the transcriptome. (a) Distribution of all unigenes length. (b) Distribution of all transcripts lengths. The x-axis indicates the length and y-axis

indicates the number of unigenes or transcripts.
(TIF)

S2 Fig. Characteristics of the homology search of Illumina sequences in the nr database.
(TIF)

S3 Fig. Correlation between samples.
(TIF)

S4 Fig. Correlations matrix heat map.
(TIF)

S1 Table. The list of the sequences identified as ncRNA.
(XLSX)

S2 Table. Primers for real-time PCR.
(PDF)

S3 Table. Part of the sequencing and assembly results.
(DOCX)

S4 Table. The blast information for each contig.
(XLSX)

S5 Table. DE information for each contig. FDR: false discovery rate; Log₂FC: log₂ fold change of DE genes; regulated: the expression level of DEGs in T04/T05 compared to T01/T02, including three situations, up-regulated, down-regulated or normal; T01~T05 are FPKM values of unigenes in each sample.
(XLSX)

S6 Table. Annotations of the DEGs. _FPKM: FPKM values of unigenes in each sample; FDR: false discovery rate; Log₂FC: log₂ fold change of DE genes; regulated: the expression level of DEGs in T04/T05 compared to T01/T02, including three situations, up-regulated, down-regulated or normal; other columns are functional annotations of unigenes.
(XLSX)

S1 File. KEGG pathway maps of DEGs in T1-T4.
(ZIP)

Acknowledgments

We would like to thank Bin Tian, Chenglong Gao and Lixiang Wang for their contributions in insect sampling.

Author Contributions

Investigation: Mingming Cui, Shixiang Zong.

Methodology: Mingming Cui, Ping Hu, Jing Tao.

Writing – original draft: Mingming Cui.

Writing – review & editing: Ping Hu, Tao Wang, Jing Tao, Shixiang Zong.

References

1. Bale JS. Cold hardiness and overwintering of insects. In: Gordon ER, editor. Biology and population dynamics of invertebrate crop pests. Andover; 1989. p. 305–340.

2. Bale JS, Masters GJ, Hodkinson ID, Awmack C, Bezemer TM, Brown VK, et al. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biol.* 2002; 8(1):1–16.
3. Richard E, Lee J. Principles of insect low temperature tolerance. In: Richard E. et al., editors. *Insects at low temperature*. Springer US; 1991. p. 17–46.
4. Turnock W, Lamb R, Bodnaryk R. Effects of cold stress during pupal diapause on the survival and development of *Mamestra configurata* (Lepidoptera: Noctuidae). *Oecologia.* 1983; 56(2–3):185–192. <https://doi.org/10.1007/BF00379690> PMID: 28310194
5. Bale JS. Insect cold hardiness: freezing and supercooling—an ecophysiological perspective. *J Insect Physiol.* 1987; 33(12):899–908.
6. Hutchinson L, Bale J. Effects of sublethal cold stress on the aphid *Rhopalosiphum padi*. *J Appl Ecol.* 1994:102–108.
7. Lee RE. Insect cold-hardiness: To freeze or not to freeze. *Biosci.* 1989; 39(5):308–313.
8. Oswood MW, Miller LK, Irons JG. Overwintering of freshwater benthic macroinvertebrates. In: Richard E. et al., editors. *Insects at low temperature*. Springer US; 1991. p. 360–375.
9. Irons JG, Miller LK, Oswood MW. Ecological adaptations of aquatic macroinvertebrates to overwintering in interior Alaska (USA) subarctic streams. *Can J Zool.* 1993; 71(1):98–108.
10. Salt R. Principles of insect cold-hardiness. *Annu Rev Entomol.* 1961; 6:55–74.
11. Baust J, Rojas R. Review—insect cold hardiness: facts and fancy. *J Insect Physiol.* 1985; 31(10):755–9.
12. Zachariassen KE. Physiology of cold tolerance in insects. *Physiol Rev.* 1985; 65(4):799–832. PMID: 3903795
13. Salt RW. Role of glycerol in producing abnormally low supercooling and freezing points in an insect, *Bracon cephi* (Gahan). *Nature.* 1958; 181(4618):1281–1281
14. Sinclair BJ, Addo-Bediako A, Chown SL. Climatic variability and the evolution of insect freeze tolerance. *Biol Rev.* 2003; 78(2):181–195. <https://doi.org/10.1017/s1464793102006024> PMID: 12803420
15. Trautsch J, Rosseland BO, Pedersen SA, Kristiansen E, Zachariassen KE. Do ice nucleating lipoproteins protect frozen insects against toxic chemical agents? *J Insect Physiol.* 2011; 57(8):1123–1126.
16. Duman JG. Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annu Rev Physiol.* 2001; 63(1):327–357.
17. Storey KB, Storey JM. Insect cold hardiness: metabolic, gene, and protein adaptation. *Can J Zool.* 2012; 90(4):456–475.
18. Luke TD, Alice BD, Sinclair BJ, Newcomb RD, Buckley TR. Divergent transcriptional responses to low temperature among populations of alpine and lowland species of New Zealand stick insects (*Micrarchus*). *Mole. Eco.* 2014; 23(11):2712–26.
19. Dennis AB, Dunning LT, Sinclair BJ, Buckley TR. Parallel molecular routes to cold adaptation in eight genera of new zealand stick insects. *Sci. Rep.* 2015; 5.
20. Dunning LT, Dennis AB, Park D, Sinclair BJ, Newcomb RD, Buckley TR. Identification of cold-responsive genes in a new zealand alpine stick insect using rna-seq. *Com. Biochem. Phys. D.* 2013; 8(1), 24–31.
21. Durant DR, Berens AJ, Toth AL, Rehan SM. Transcriptional profiling of overwintering gene expression in the small carpenter bee, *Ceratina calcarata*. *Apidologie.* 2015; 47(4), 572–582.
22. Tittiger C. Functional genomics and insect chemical ecology. *J Chem Eco.* 2004; 30(12):2335–2358.
23. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research.* 2008; 18(9):1509–1517. <https://doi.org/10.1101/gr.079558.108> PMID: 18550803
24. Ramani AK, Nelson AC, Kapranov P, Bell I, Gingeras TR, Fraser AG. High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective *Caenorhabditis elegans*. *Genome Biol.* 2009; 10(9):R101. <https://doi.org/10.1186/gb-2009-10-9-r101> PMID: 19778439
25. Yassour M, Kaplan T, Fraser HB, Levin JZ, Pfiffner J, Adiconis X, et al. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *P Natl Acad Sci USA.* 2009; 106(9):3264–3269.
26. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 2008; 320:1344–1349. <https://doi.org/10.1126/science.1158441> PMID: 18451266
27. Qin W, Neal S, Robertson R, Westwood J, Walker V. Cold hardening and transcriptional change in *Drosophila melanogaster*. *Insect Mol Biol.* 2005; 14(6):607–613. <https://doi.org/10.1111/j.1365-2583.2005.00589.x> PMID: 16313561

28. Sørensen JG, Nielsen M, Loeschcke V. Gene expression profile analysis of *Drosophila melanogaster* selected for resistance to environmental stressors. *J Evolution Biol.* 2007; 20(4):1624–1636.
29. Laayouni H, García-Franco F, Chávez-Sandoval BE, Trotta V, Beltran S, Corominas M, et al. Thermal evolution of gene expression profiles in *Drosophila subobscura*. *BMC evolutionary biology.* 2007; 7(1):42.
30. Li TS, Schroeder W. Sea buckthorn (*Hippophae rhamnoides* L.): a multipurpose plant. *HortTechnology.* 1996; 6(4):370–380.
31. Zong S, Luo Y, Xu Z, Wang T. Preliminary study on different larval instars of seabuckthorn carpenter moth, *Holcocerus hippophaecolus*. *Chin Bull Entomol.* 2006; 43(5):626–627.
32. Zong S. Studies on the Bio-ecological characteristics of seabuckthorn carpenter moth: *Holcocerus hippophaecolus* (Lepidoptera: Cossidae). Diss. Beijing forestry university. 2006.
33. Luo Y, Lu C, Xu Z. Control strategies on a new serious forest pest insect seabuckthorn carpenterworm, *Holcocerus hippophaecolus*. *Forest Pest Dis.* 2003; 22(5):25–28.
34. Fang YL, Sun JH, Zhao CH, Zhang ZN. Sex pheromone components of the sandthorn carpenterworm, *Holcocerus hippophaecolus*. *J Chem Ecol.* 2005; 31(1):39–48. PMID: [15839478](https://pubmed.ncbi.nlm.nih.gov/15839478/)
35. Tian B, Xu L, Zhang M, Feng YQ, Zong SX. Supercooling capacity and cryoprotectants of overwintering larvae from different populations of *Holcocerus hippophaecolus*. *Cryo Letters.* 2016; 37(3):206. PMID: [27393957](https://pubmed.ncbi.nlm.nih.gov/27393957/)
36. Wu W, Sun H, Guo J, Jiang F, Liu X, Zhang G. *De novo* transcriptome characterization of the ghost moth, *Thitarodes pui*, and elevation-based differences in the gene expression of its larvae. *Gene.* 2015; 574(1):95–105. <https://doi.org/10.1016/j.gene.2015.07.084> PMID: [26235680](https://pubmed.ncbi.nlm.nih.gov/26235680/)
37. Han XJ, Wang YD, Chen YC, Lin LY, Wu QK. Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. *PLoS One.* 2013; 8(10):e76890. <https://doi.org/10.1371/journal.pone.0076890> PMID: [24130803](https://pubmed.ncbi.nlm.nih.gov/24130803/)
38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: [24695404](https://pubmed.ncbi.nlm.nih.gov/24695404/)
39. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full length transcriptome assembly from RNA Seq data without a reference genome. *Nat Biotechnol.* 2011; 29: 644–652. <https://doi.org/10.1038/nbt.1883> PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
40. Perteza G, Huang X, Liang F, Perteza G, Huang X, Liang F, et al. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics.* 2003; 19(5):651–652. PMID: [12651724](https://pubmed.ncbi.nlm.nih.gov/12651724/)
41. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015; 31(19):3210–212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: [26059717](https://pubmed.ncbi.nlm.nih.gov/26059717/)
42. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* 1997; 25(17): 3389–3402. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
43. Deng YY, Li JQ, Wu SF, Zhu Y, Chen Y, Fuchu HE, et al. Integrated nr Database in Protein Annotation System and Its Localization. *Comput Eng.* 2006; 32(5):71–74.
44. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004; 32:D115–119. <https://doi.org/10.1093/nar/gkh131> PMID: [14681372](https://pubmed.ncbi.nlm.nih.gov/14681372/)
45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat genet.* 2000; 25(1): 25–29. <https://doi.org/10.1038/75556> PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
46. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Research.* 2004; 32(suppl_1):D277–D280.
47. Tatusov RL, Galperin MY, Natale DA. The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000; 28(1):33–36. PMID: [10592175](https://pubmed.ncbi.nlm.nih.gov/10592175/)
48. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011; 39(suppl_2):29–37.
49. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:1–10.
50. Trapnell C; Williams BA; Perteza G; Mortazavi A; Kwan G; Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2015; 28(28):511–515.
51. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics.* 2013; 29:1035–1043. <https://doi.org/10.1093/bioinformatics/btt087> PMID: [23428641](https://pubmed.ncbi.nlm.nih.gov/23428641/)

52. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *Methods*. 2001; 25(4):402–408. <https://doi.org/10.1006/meth.2001.1262> PMID: 11846609
53. Hu P, Tao J, Cui MM, Gao CL, Lu PF, Luo YQ. Antennal transcriptome analysis and expression profiles of odorant binding proteins in *Eogystia hippophaecolus* (Lepidoptera: Cossidae). *BMC Genomics*. 2016; 17:651. <https://doi.org/10.1186/s12864-016-3008-4> PMID: 27538507
54. Zhu JY, Zhao N, Yang B. Global transcriptome profiling of the pine shoot beetle, *Tomicus yunnanensis* (Coleoptera: Scolytinae). *PloS one*. 2012; 7(2):e32291, <https://doi.org/10.1371/journal.pone.0032291> PMID: 22384206
55. Parker DJ, Ritchie MG, Kankare M. Preparing for winter: the transcriptomic response associated with different day lengths in *Drosophila montana*. *G3-Genes Genom Genet*. 2016; 6(5):1373–1381.
56. Clark MS, Worland MR. How insects survive the cold: molecular mechanisms—a review. *J Comp Physiol B*. 2008; 178(8):917–933. <https://doi.org/10.1007/s00360-008-0286-4> PMID: 18584182
57. Storey KB, Storey J. Heat shock proteins and hypometabolism: adaptive strategy for proteome preservation. *Res Rep Biol*. 2011; 2:57–68.
58. Zhang G, Storey JM, Storey KB. Chaperone proteins and winter survival by a freeze tolerant insect. *J Insect Physiol*. 2011; 57(8):1115–1122. <https://doi.org/10.1016/j.jinsphys.2011.02.016> PMID: 21382374
59. Wang H, Li K, Zhu JY, Fang Q, Ye GY. Cloning and expression pattern of heat shock protein genes from the endoparasitoid wasp, *Pteromalus puparum* in response to environmental stresses. *Arch Insect Biochem*. 2012; 79:247–263.
60. Zhang J, Goyer C, Pelletier Y. Environmental stresses induce the expression of putative glycine-rich insect cuticular protein genes in adult *Leptinotarsa decemlineata* (Say). *Insect Mol Biol*. 2008; 17(3):209–216. <https://doi.org/10.1111/j.1365-2583.2008.00796.x> PMID: 18477239
61. Nakashima T, Harada Y, Miyata S, Kiyohara T. Inhibitors of cytochrome P-450 augment fever induced by interleukin-1 beta. *American Journal of Physiology*. 1996; 271(5): R1274–1277.
62. Purać J, Burns G, Thorne MA, Grubor-Lajšić G, Worland MR, Clark MS. Cold hardening processes in the Antarctic springtail, *Cryptopygus antarcticus*: clues from a microarray. *J Insect Physiol*. 2008; 54(9):1356–1362. <https://doi.org/10.1016/j.jinsphys.2008.07.012> PMID: 18703067
63. Teets NM, Denlinger DL. Surviving in a frozen desert: environmental stress physiology of terrestrial Antarctic arthropods. *Journal of Experimental Biology*. 2014; 217(1):84–93.
64. Storey KB, Storey JM. Tribute to PL Lutz: putting life on 'pause'—molecular regulation of hypometabolism. *J Exp Biol*. 2007; 210(10):1700–1714.
65. Sharabiani MT, Siermala M, Lehtinen TO, Vihinen M. Dynamic covariation between gene expression and proteome characteristics. *BMC Bioinformatics*. 2005; 6(1):1–18.
66. Parker DJ, Vesala L, Ritchie MG, Laiho A, Hoikkala A, Kankare M. How consistent are the transcriptome changes associated with cold acclimation in two species of the *Drosophila virilis* group? *Heredity*. 2015; 115:13–21. <https://doi.org/10.1038/hdy.2015.6> PMID: 25669607