Gene 668 (2018) 229–236

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Transcriptome sequencing for identification of diapause-associated genes in fall webworm, *Hyphantria cunea* Drury

Yu Deng[a,1], Fei Li[a,1], Lynne K. Rieske[b], Li-Li Sun[a], Shou-Hui Sun[a,b,*]

[a] College of Forestry, Shenyang Agricultural University, Shenyang 110866, China
[b] Department of Entomology, University of Kentucky, Lexington 40546, USA

ARTICLE INFO

Keywords:
Hyphantria cunea
Transcriptome
Diapause
Functional enrichment analysis
Differentially expressed genes

ABSTRACT

Fall webworm, *Hyphantria cunea* Drury (Lepidoptera: Arctiidae) is extremely adaptable and highly invasive in China as a defoliator of ornamental and forest trees. Both voltinism and diapause strategies of fall webworm in China are variable, and this variability contributes to it invasiveness. Little is known about molecular regulation of diapause in fall webworm. To gain insight into possible mechanisms of diapause induction, high-throughput RNA-seq data were generated from non-diapause pupae (NDP) and diapause pupae (DP). A total of 58,151 unigenes were assembled and researched against nine public databases. In total, 29,013 up-regulated and 3451 down-regulated unigenes were differentially expressed by DP when compared with those of NDP. Genes encoding proteins such as UDP-glycosyl transferase (UGT), cytochrome P450 and Hsp70 were predicted to be involved in diapause. Moreover, GO function and KEGG pathway enrichments were performed on all differentially expressed genes (DEGs) and showed that cell cycle and insulin signaling pathways may be related to the diapause of the fall webworm. This study provides valuable information about the fall webworm transcriptome for future gene function research, especially as it relates to diapause.

1. Introduction

Fall webworm, *Hyphantria cunea* Drury (Lepidoptera: Arctiidae), a polyphagus pest that feeds on about 160 species of broad leaf trees including mulberry, oak, hickory, pecan, walnut, elm, alder, willow, sweetgum, and poplar (Liao et al., 2010) is an important quarantine pest that is highly adaptable and invasive in China. Fall webworm originates in North America, and spread into central Europe and eastern Asia in the 1940s (Warren and Tadic, 1970; Masaki et al., 1968). In China it was first found in Dandong (124°N/40°E) of Liaoning Province in 1979, and now has spread southwards to Wuhu (118°N/31°E) of Anhui Province, westwards to Xianyang (108°N/34°E) of Shanxi Province and northwards to Siping (123°N/42°E) of Jilin Province (Ji et al., 2003; Zhang et al., 2016). There are commonly two or three generations per year in China, and fall webworm can pupate in both the summer and winter. In Shenyang, there are just two generations per year.

The annual arrival of winter in temperate regions represents a fundamental challenge to the survival and reproduction of many insects (Poelchau et al., 2011), who typically respond to the harsh conditions of winter by undergoing diapause, a process in which the insects grow slowly or not at all, showing low metabolism or developmental inactivity for a long period in the absence of a specific stimulus to break diapause (Qi et al., 2015). Generally, in the insect lifecycle, diapause is accomplished by the dynamic change of developmental, behavioral and physiological events and influenced by environmental factors such as humidity, nutrition, temperature and light (Saravanakumar et al., 2008). Furthermore, the processes related to regulation of development, metabolic depression, stress tolerance and nutrient storage appear to be important physiological components of the diapause response. While many physiological and ecological aspects of insect diapause are known, such as water loss (Benoit and Denlinger, 2007), nutrient storage, and utilization (Hahn and Denlinger, 2007; Hahn and Denlinger, 2011), knowledge of insect diapause at the molecular level remains largely unknown, including for the fall webworm.

Recently, next-generation sequencing technology has greatly facilitated genetic and genomic analyses, especially for insects without reference genome sequences (Ragland et al., 2010; Ekblom and Galindo, 2011; Liu et al., 2014). Next generation sequencing has already led to exciting progress on the transcriptome in several insect species, such as *Bombyx mori* (Xia et al., 2004), *Danaus plexippus* (Zhan et al., 2011), *Heliconius melpomene* (THG Consortium, 2012) and *Plutella*

Abbreviations: NDP, non-diapause pupae; DP, diapause pupae

* Corresponding author at: College of Forestry, Shenyang Agricultural University, Shenyang 110866, China.
E-mail address: sysshh@163.com (S.-H. Sun).

[1] Yu Deng and Fei Li contributed equally to this work.

https://doi.org/10.1016/j.gene.2018.05.023
Received 26 October 2017; Received in revised form 3 May 2018; Accepted 7 May 2018

xylostella (You et al., 2013). In this study, we used transcriptome sequencing to compare the gene expression profiles of the fall webworm at diapausing and non-diapausing stages, and identified differentially and specifically expressed unigenes following diapause using Illumina sequencing technology. This partially characterized fall webworm transcriptome provides an important molecular resource for further genomic studies of diapause induction.

2. Materials and method

2.1. H. cunea pupae

All samples were collected from the campus of Shenyang Agricultural University (SY strain, 117°38′E, 41°11′N) on June (2013) and October (2013) to February (2014). Whole pupae at diapause or non-diapause stages, including early-, middle-, and late-stage were freshly frozen and stored at −80 °C until RNA extraction. For RNA-seq library preparation, all stages (early-, middle-, and late-stage) were combined equally as a single mixed sample and sequenced to increase the coverage of the transcriptome. For analysis of differentially expressed genes (DEGs), diapause pupae (DP) and non-diapause pupae (NDP) libraries were prepared to compare the gene expression profiles. The RNA from 6 individuals (3 male and 3 female) was pooled as a single sample.

2.2. RNA-seq library preparation

Total RNA was extracted from DP and NDP samples using total RNA extractor (Sangon Biotech, Shanghai, China) according to the manufacturer's instructions. The quantity, purity and integrity of RNA were determined by spectrophotometry and gel electrophoresis. Samples were pooled due to limited resources. Equal amounts of high-quality RNA of different stages from the fall webworm were pooled for cDNA synthesis. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Ipswich, MA USA) following the manufacturer's instructions, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. The double-stranded cDNA was used for end reparation, poly (A) addition, and ligation with sequencing adaptors. Following purification and size selection, the suitable cDNA fragments were enriched with PCR to generate the final cDNA library. Sequencing was performed on Illumina HiSeq 2500 platform with the paired-end read module (Sangon Biotech Co., Ltd., Shanghai, China).

2.3. Sequence analysis, de novo assembly and annotation

For transcriptome assembly, the sequenced raw reads were filtered by removing adaptor sequences of 3′ end, low-quality sequences (Q ≤ 20 bases), and sequences with ambiguous bases (N). The remaining high-quality reads were *de novo* assembled into transcripts using Trinity (2013-02-25) by paired-end joining method. Transcripts with a minimum length of 200 bp were clustered. For each cluster (representing transcriptional complexity for the same gene), the longest sequence was preserved and designated as unigene.

The resulting unigenes were blasted against NCBI Nr (NCBI non-redundant protein database), Swiss-Prot (a manually annotated and reviewed protein sequence database), TrEMBL (Translated EMBL Nucleotide Sequence Data Library), CDD (Conserved Domain Database), Pfam (Protein family) and KOG (EuKaryotic Orthologous Groups) databases (E-value ≤ 1e−5). Functional annotation by GO (Gene Ontology database) was analyzed using the Blast2GO program (Conesa et al., 2005; Götz et al., 2008). The KEGG (Kyoto Encyclopedia of Genes and Genomes) was also used to predict and classify possible functions.

2.4. Analysis and annotation of DEGs

We aligned the high-quality cleaned RNA-seq reads to the assembled fall webworm transcripts with the Bowtie program, allowing one mismatch. Following alignments, raw counts for each sample were derived and normalized to reads per kilobase of exon model per million mapped reads (RPKM). For differential expression analysis, only samples with RNA-seq reads > 200 bp in length were used. DEGs (fold changes ≥ 2 and adjusted false discovery rate (FDR) ≤ 0.001) between DP and NDP were identified with the DESeq package (Audic and Claverie, 1997). For the enrichment analysis, the hypergeometric distribution algorithm (Phyper) is used to calculate the P value and corrected by FDR. All of the DEGs were mapped to the GO and KEGG pathway terms and significantly enriched terms were filtered by P ≤ 0.05.

2.5. Experimental validation

To evaluate the quality of the sequence assembly and expression profile, seven up-regulated genes and seven down-regulated genes that were expressed during diapause were validated *via* real-time quantitative PCR (qPCR). 14 pairs of primers (Additional file 1) were designed using Primer Premier 5.0. 1 μg of RNA extracted from the same samples at the same time was employed for first-strand cDNA synthesis using the Prime ScriptII 1st Strand cDNA Synthesis Kit (Takara, Dalian, China) according to the manufacturer's protocol. qPCR was conducted using SYBR FAST qPCR Kit Master Mix(2×) Universal (KAPA Biosystems, Woburn, MA, USA) under the following conditions: 95 °C for 3 min, followed by 45 cycles of 95 °C for 20 s and 60 °C for 1 min. The melting curve was analyzed from 60 °C to 95 °C to detect non-specific product amplification. The *β-Actin* and *RPL13* were used as the internal genes. All of the data obtained through qPCR were analyzed *via* the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

3. Results

3.1. Tag identification and quantification

Using data obtained from sequencing through the Illumina Hiseq2500 PE125 platform, we performed routine transcriptome analysis on pupae of *H. cunea* in non-diapause and diapause stages, encompassing a total of 119,859,478 raw reads, comprising 78,908,094 and 40,951,384 reads for non-diapause and diapause groups, respectively. After adaptor trimming and quality filtering, a total of 78,349,608 and 40,830,528 clean reads of each group were filtered, and assembled into 89,213 transcripts with a mean length of 1284.32 bp and N50 length of 2912 bp using Trinity software through pair-end reads. The transcripts were further assembled into 58,151 unigenes with a mean length of 787.1 bp (Table 1). Of these unigenes, 46,992 (80.81%) were 200–1000 bp in length and 5500 (9.46%) were > 2000 bp, with most unigenes falling between 200 bp and 500 bp (56.76%). The unigene length distribution closely followed the transcript length distribution.

3.2. Sequence annotation

All unigenes were annotated (Table 2). The NR database had the largest match (20,872, 35.89%), followed by the TrEMBL (20,771, 35.72%) and GO (16,734, 28.78%) database. The E-value distribution of the top hits in the NR database showed that 56.93% of the mapped sequences had strong homology (E-value < 1.0e−50). Because the