Significant transcriptional changes in mature daughter *Varroa destructor* mites during infestation of different developmental stages of honeybees

Jiangli Wu[a], Hany M. Elsheikha[b], Yangyang Tu[a], Awraris Getachew[a], Huaiyu Zhou[c], Chunxue Zhou[c*], Shufa Xu[a*]

a *Key Laboratory of Pollinating Insect Biology, Ministry of Agriculture and Rural Affairs, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, PR China*

b *Faculty of Medicine and Health Sciences, School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK*

c *Department of Parasitology, Shandong University School of Basic Medicine, Jinan 250100, PR China*

∗ *Correspondence to: CX Zhou, Department of Parasitology, Shandong University School of Basic Medicine, Jinan 250100 China. E-mail: zhouchunxue23@163.com; SF Xu, Key Laboratory of Pollinating Insect Biology, Ministry of Agriculture, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093 China, E-mail: xushufa@caas.cn*

ABSTRACT

BACKGROUND: *Varroa destructor* is considered a major cause of honeybee (*Apis mellifera*) colony losses worldwide. Although *V. destructor* mites exhibit preference behavior for certain honeybee lifecycle stages, the mechanism underlying host finding and preference remains largely unknown.

RESULTS: By using a *de novo* transcriptome assembly strategy, we sequenced the mature daughter *V. destructor* mite transcriptome during infestation of different stages of honeybees (brood cells, newly emerged bees and adult bees). A total of 132,779 unigenes were obtained with an average length of 2,745 bp and N50 of 5,706 bp. About 63.1% of the transcriptome could be annotated based on sequence homology to the predatory mite *Metaseiulus occidentalis* proteins. Expression analysis revealed that mature daughter mites had distinct transcriptome profiles after infestation of different honeybee stages, and that the majority of the differentially expressed genes (DEGs) of mite infesting adult honeybees were down-regulated compared to that infesting the sealed brood cells. Gene Ontology and KEGG pathway enrichment analyses showed that a large number of DEGs were involved in cellular process and metabolic process, suggesting that *Varroa* mites undergo metabolic adjustment to accommodate the cellular, molecular and/or immune response of the honeybees. Interestingly, in adult honeybees, some mite DEGs involved in neurotransmitter biosynthesis and transport were identified and their levels of expression were validated by qPCR.

CONCLUSION: These results provide evidence for transcriptional reprogramming in mature daughter *Varroa* mites during infestation of honeybees, which may be relevant to understanding the mechanism underpinning adaptation and preference behavior of these mites for honeybees.

Keywords: *Varroa destructor*; *Apis mellifera*; transcriptomic analysis; infestation; neurotransmitter

1 INTRODUCTION

Honeybees are important economic insects because they are the main crop pollinator and honey producer. However, these activities are seriously compromised by the spread of *Varroa destructor*, an ectoparasite of honeybee that is responsible for significant losses in honeybee populations.[1] The *Varroa* mites damage the honeybee colony by feeding on the fat body and hemolymph of honeybee brood.[2] Additionally, these mites can act as a vector for multiple viruses, leading to more damage to the bee colony.[3] The damage caused by mites accelerates maturation of the infested worker honeybees and reduces their lifespan. A previous study showed that honeybee colonies heavily infested with the mites die within 1-2 years.[4] Heavy *Varroa* mite infestation results in an unbalanced demographic structure and even collapse of the colony.[5] Due to the acaricide resistance that is rapidly evolving, mite infestations have become more difficult to control. Additionally, synthetic miticides used to treat honeybees against *Varroa* mite infestation can contaminate honey and other hive products.[6, 7]

During the *Varroa* life cycle, the mite switches between adult and brood stage of the honeybee host. As a parasite without a free living phase, the *Varroa* mites prefer living in the dark nest of honeybees, especially in the sealed brood cells.[8] For their reproductive success, after leaving the brood cell on young honeybees, the mites have to infest suitable adult bees in order to spread to new brood cells. The development of genomic tools has considerably facilitated the elucidation of the molecular mechanisms underlying the honeybee-*Varroa* mite interactions and provided important means to diagnose and manage bee diseases.[9, 10] A previous study showed that mite infestation perturbed the gene expression patterns and enhanced the immune response of honeybees.[11] *Varroa* mite infestation has been also shown to change the expression levels of some antibacterial peptides, such as abacein and defensin, in the bee host.[12]

Chemical orientation is essential for the reproductive success of *Varroa* mite. Freshly emerged infested bees are less attractive to *Varroa* mite and the nurse bees of middle age are the most infested in breeding colonies.[13] A previous study showed that extensive remodeling of *Varroa* transcriptome occurs during adult mite life cycle, where gene

expression profiles of different adult life stages of *Varroa* suggested mite adaptation to its host.[14] However, there is a lack of data on the global dynamic transcriptome of mature daughter *Varroa* mite during its transition from young bees to newly emerged bees, to adult bees.

In this study, we performed a *de novo* transcriptome assembly and annotation of mature daughter *V. destructor* mites infesting three different developmental stages of bees (brood cells, newly emerged bees and adult bees). Our data show the complexity of transcriptional changes that occur in adult *V. destructor* mites as they infest and interact with different lifecycle stages of honeybees.

2 MATERIALS AND METHOD

2.1 Honeybee colonies and *V. destructor* mites

Artificially mated *A. mellifera* queens were purchased from Chengde honeybee breeding station, Hebei Province, China, and kept for colony build-up at the apiary of the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences in Beijing. A total of 6 honeybee colonies with mated queens of the same age, similar colony strength and high level of parasitism were used in this study. Meanwhile, another 9 non-infested honeybee colonies with mated queens of the same age, similar colony strength were also bred at the apiary of the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences in Beijing. No acaricide treatment was applied in this study in order to avoid pesticide bias.

2.2 Sample collection

Adult female *V. destructor* mites with the same age were equally obtained from 6 honeybee colonies with a high level of parasitism. These mites were delivered into brood cells in non-infested colonies shortly before capping. Mature daughter mites were collected from sealed brood cells in the comb (S1), newly emerged bees from brood cells within one day (S2) and adult bees (emerged from the cell after 7 days, S3). For S1 group, mature daughter *Varroa* mites were harvested from soon-to-emerge bees. For

S2 and S3, mother mites were removed from brood cells at pupal stage and the cell sealed with melted beeswax. Each group contained 3 replicates in 3 non-infested colonies. Each replicate consists of 100 adult mites. All the mite samples were frozen in liquid nitrogen and stored at -80℃ until RNA extraction.

2.3 RNA extraction and quantification

Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. The RNA concentration was measured using Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA), and RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

2.4 Transcriptome library preparation and sequencing

About 1.5 μg RNA/sample was used for library construction. Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and fragmented using divalent cations under elevated temperature in NEBNext First-Strand Synthesis Reaction Buffer (5X). First and second-strand cDNA synthesis were synthesized. After adenylation of 3' ends of DNA fragments, NEBNext Adaptors with hairpin loop structure were ligated to prepare for hybridization, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then, PCR was performed and purified, and the quality of the library was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform and paired-end reads were generated.

2.5 Sequencing data processing, *de novo* assembly and annotation

Using in-house Perl scripts, raw reads were filtered by removing the low-quality reads

and reads that contains adapter or poly-N. Also, Q30, GC-content and sequence duplication level were calculated based on the clean reads in order to evaluate the sequencing quality. Sequence contaminants from other sources, such as bacteria and fungi, present in the samples were removed prior to functional annotation analysis. The pair-end short reads were assembled into contigs via Trinity software.[15] The Benchmarking Universal Single Copy Orthologs (BUSCO) v2 were used to evaluate the quality and completeness of transcriptome assembly obtained in this study.[16] The *de novo* transcriptome served as the reference. Then, high-quality clean reads were mapped back to the assembled transcriptome sequences using Bowtie 2 software.[17] Unigenes were annotated against the non-redundant protein sequences database (Nr), nucleotide sequences database (Nt), Pfam, clusters of orthologous groups, eukaryotic ortholog groups and Swissprot databases.

2.6 Identification of SSRs

V. destructor assembled transcriptome was scanned for the identification of simple sequence repeats (SSRs) using MISA (http://pgrc.ipk-gatersleben.de/misa) using the default parameters.[18] The minimum number of repeat units for mono-nucleotide was 10 and for di-nucleotide was 6, whereas for tri-, tetra-, penta- and hexa-nucleotide, the minimum number of repeat units was >5 in the MISA search criteria.

2.7 Gene expression analysis

To calculate the transcript expression, we used the reads per kilobase of the exon model per million mapped reads (RPKM) method. The differentially expressed genes (DEGs) were identified using the DESeq R package based on the negative binomial distribution.[19] The *P* values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). An adjusted *P*-value <0.05 along with at least two-fold change was used to identify significantly differential expression of the transcripts. The heatmap showing the differential unigenes was generated via TIGR MultiExperiment Viewer (MeV, v4.8).[20]

2.8 Gene ontology (GO) and KEGG pathway analysis

Gene Ontology (GO) enrichment analyses of the DEGs were performed using web-based GO software (http://www.geneontology.org) for gene ontology (GO) annotation and enrichment analysis.[21] The GO project includes three main categories: biological process, cellular component and molecular function. Also, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted using a web-based database (http://www.genome.p/kegg).[22] Protein-protein interaction (PPI) networks were built based on the publicly available program, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database.[23] The PPI networks of these DEGs were visualized using a Cytoscape software.[24]

2.9 Quantitative real-time RT-PCR (qRT-PCR) verification

To validate the result of RNA-Seq, total RNA was extracted using TRIzol method (Invitrogen) and was reverse-transcripted to single strand cDNA using GoScriptTM Reverse Transcription System (Promega, MI, USA) according to the manufacturer's instructions. The gene-specific primer pairs were designed using Primer Premier 5.0 software (**Table S1**), and the qPCR was performed on LightCycler 480 (Roche Diagnostics, Tokyo, Japan). The amplification reactions were performed with the following conditions: 2 min at 95°C, 40 cycles of 95 °C for 5s, 60°C for 30s. The experiment was repeated three times using three independently isolated RNA samples. The glyceraldehyde-3-phosphatedehydrogenase (GAPDH) gene was used as a reference, and relative expression levels were calculated using the $2^{-\Delta\Delta CT}$ method.

3 RESULTS

3.1 Transcriptome sequencing and *de novo V. destructor* transcriptome assembly

Nine cDNA libraries were generated with mRNA from three groups: S1 (mature daughter mites collected from the sealed brood cells), S2 (mature daughter mites obtained from newly emerged bees) and S3 (mites on the adult bees). Each group included 3 biological replicates. These cDNA libraries were subjected to high

throughput sequencing. As shown in **Table 1**, we acquired > 46 million 150bp paired-end-seq raw reads from each cDNA library. After eliminating adapters, ambiguous nucleotides and low-quality sequences, > 44 million clean reads (Q20>95%) were retained, which accumulated to > 6.75 Giga bases (Gb) read length with a GC percentage > 41%. The raw paired-end sequence dataset has been deposited at the National Center for Biotechnology Information (NCBI) Short Read Archive under BioProject ID: PRJNA486893.

The clean reads from the above nine cDNA libraries were assembled by Trinity program and 249,505 transcripts were generated with an average length of 1,603 bp and an N50 length of 5,069 bp. Among the transcripts, 130,086 (52.14%) are < 500 bp long, and 50,509 (20.24%) are > 2,000 bp (**Fig. 1**). These transcripts were further subjected to cluster and assembly analyses. A total of 132,779 unigenes were obtained with a mean length of 2,745 bp and an N50 value of 5,706 bp (**Table 2**). Among the assembled unigenes, 17,996 (13.55%) are between 200 bp and 500 bp long, and 50,507(38.04%) are > 2,000 bp (**Fig. 1**).

3.2 Sequence annotations

In order to identify the putative functions of the unigenes, BLAST programs (e value< 1.0E-5) was employed to search against public databases (Nr, Nt, KO, Swiss-Prot, PFAM, GO and KOG), which were used for gene annotations. As shown in **Table 3**, the results showed that 82,068 (61.8% of 132,779) unigenes were matched to one or more databases. A total of 68,940 unigenes were found to have homologs in the NR database with an e-value < the cutoff (e-value=1E-5). The e-value distribution analysis of the hit unigenes showed that 52.7% of *V. destructor* unigenes have the highest homology with an e-value cut-off < 1E-100 (**Fig. 2A**). Likewise, the similarity distribution showed that 56.9% of all the unigenes had a similarity > 80%, whereas 42.9% of unigenes had similarity that ranged from 40% to 80% and only 0.1% had similarity below 40% (**Fig. 2B**). As anticipated, the top unigene hit was found in the arthropod genomes. *Metaseiulus occidentalis* (63.1%), *Apis mellifera* (15.7%), *Apis dorsata* (4.7%), *Apis florea* (3.6%), and *Ixodes scapularis* (1.9%) had the top five

counts of unigenes with NR annotation (**Fig. 2C**). Evaluation of the quality of assembly and completeness of annotations of the transcriptome was performed using BUSCO software. For comparative purposes, we have included a recently published version of *V. destructor* transcriptome assembly performed by Mondet et al..[16] Our assembly is ~ 98.8% complete (309 complete single-copy and 744 complete duplicated BUSCO), while only 0.8% of contigs were fragmented (9 BUSCOs) and 0.4% were missing (4 BUSCOs) (**Fig. 3**).

All unigenes were aligned to the Cluster of Orthologus Groups (COG) database for functional prediction and classification. A total of 44,672 unigenes were assigned to appropriate COG clusters, which could be classified into 25 functional categories. As shown in **Fig. 4**, 'General function prediction only' was the largest category (7,698 unigenes); followed by 'signal transduction mechanisms' (7,604 unigenes), and 'posttranslational modification, protein turnover, chaperones' (4,545 unigenes).

3.3 SNP detection and characterization of simple sequence repeats (SSRs) markers.

A total of 23,030 high-quality SNPs were identified among all unigenes (**Table S2**). The predicted SNPs included 19,284 transitions (9,509 C/T and 9,775 A/G transitions) and 3,746 transversions (925 A/T, 904 A/C, 887 T/G and 1030 C/G transversions). To investigate new molecular markers, all the unigenes found in this study were used to identify SSRs in the *V. destructor* transcriptome. A total of 95,470 SSRs were discovered in 52,332 unigenes (39.4%). As shown in **Fig. S1**, the most abundant motifs detected were mononucleotide (47% of the total SSRs), followed by dinucleotide motifs (29.7%). Among the identified SSRs, A (T) (45.2% of the total SSRs) accounted for 96.2% of the mononucleotide repeats, whereas AT (AT), AC (GT) and AG (CT) together accounted for 29.5% of the total SSRs (**Fig. 5**).

3.4 Expression analysis and identification of differentially expressed genes (DEGs)

RSEM (RNA-Seq by Expectation-Maximization) was used to identify DEGs in *V. destructor* after infestation of different bee developmental stages. Nine digital gene

expression libraries were constructed after Illumina deep sequencing. Through the alignment to the assembled *V. destructor* transcriptome, > 87% clean reads were mapped uniquely (**Table S3**). The rest of the clean reads that were not mapped to the reference transcriptome assembly were filtered as multiple aligned and not included in subsequent analyses. The FPKM (Fragments per Kilobase of transcript per million mapped reads) values were used to profile the expression level of each transcript. As shown in **Fig. 6A**, the distribution profile for all transcripts showed that the value of FPKM of transcripts in S1 group was higher than the other two groups. In the FPKM density distributions, the maximum density of $\log_{10}$ (FPKM+1) was approximately -0.5, and the FPKM increased as the density gradually decreased (**Fig. 6B**). FPKM interval and total genes in each library are shown in **Fig. 6C**. Results showed that most FPKM of reads were between 0.3 and 3.57 in S1, and 0-0.1 in S2 and S3.

DEGs were determined by applying the screening thresholds of $\log_2$ fold change and *p*adj＜0.05. Based on the method, the number of up-regulated DEGs were 2,016, 824 and 34, and the down-regulated DEGs were 11,138, 7,475 and 121 in 'S2 *vs.* S1', 'S3 *vs.* S1' and 'S3 *vs.* S2', respectively (**Fig. 7A**, **Fig. S2A** and **S2B**). The three comparison groups shared a total of 10 DEGs, and 5,615 (S2 *vs.* S1), 742 (S3 *vs.* S1) and 49 (S3 *vs.* S2) DEGs were uniquely differentially expressed (**Fig. 7B**). To define the gene expression profiles of DEGs in different mite transcriptomes, a hierarchical clustering of DEGs was constructed based on FPKM of RNA-seq data. As shown in **Fig. 7C**, the unsupervised hierarchical clustering showed that the three groups displayed distinct gene expression patterns and gene expression profile in S2 showed higher identity to S3.

3.5 GO and KEGG pathway enrichment analyses of DEGs

To better understand the biological regulatory mechanisms underlying mite infestation, we performed GO annotation analysis of the DEGs identified in this study. Significantly enriched GO terms were identified using an adjusted *P*-value based on hypergeometric distribution. There were 60 significantly enriched GO terms between S2 and S1, and 49 between S3 and S1. However, there were no significantly enriched GO terms

between S3 and S2. In terms of S2 group *vs.* S1 group, the GO terms with the maximum number of DEGs in biological process, cellular component and molecular function were 'cellular process', 'cell', and 'binding', respectively (**Fig. 8A**). When comparing S3 group and S1 group, the maximum number categories of DEGs enrichment in the GO three categories also were 'cellular process', 'cell', and 'binding', respectively (**Fig. S3**).

Next, we performed the KEGG pathway enrichment analysis based on these DEGs. As shown in **Table S4**, there were 297 enriched KEGG pathways in S2 *vs.* S1. The top three overrepresented pathways were: 'Proteasome', 'Oxidative phosphorylation' and 'Protein processing' in the endoplasmic reticulum (**Fig. 8B**). As shown in **Table S5**, there were 296 enriched KEGG pathways between S3 and S1. 'Proteasome', 'Oxidative phosphorylation' and 'Parkinson's disease' were the top three most significantly enriched KEGG pathways (**Fig. S4)**.

The role of these DEGs in related metabolic pathways was also studied through the KEGG pathway enrichment analysis. As shown in **Fig. 9A**, 'Arginine and proline metabolism', 'Cysteine and methionine metabolism' and 'Valine, leucine and isoleucine degradation' were the top 3 most enriched amino acid metabolic pathways between S2 and S1. Also, comparison between S2 and S1 showed 'Glycerophospholipid metabolism', 'Glycerolipid metabolism' and 'Fatty acid degradation' as the top 3 most affected lipid metabolism pathways (**Fig. 9B**).

3.6 Candidate genes with putative functions in neurotransmitters regulation

In both invertebrates and vertebrates, the neurotransmitters act as biological mediators of intracellular communication by the activation of certain receptors and other second messengers in neurons. The control of physiology and behavior is achieved through the involvement of neurotransmitter signaling. As shown in **Fig. 10**, four candidate genes encoding enzymes in *V. destructor* were down-regulated between S2 and S1, including glutamine synthetase, glutamate dehydrogenase, glutamate synthase, and phenylalanine hydroxylase. Also, two genes encoding glutamate decarboxylase and

acetycholinesterase were up-regulated. Five genes encoding neural transporters involved in neurotransmitters biosynthesis and transport were up-regulated, which included choline transporter, mitochondrial aspartate/glutamate transporter, glial high affinity glutamate transporter, excitatory amino acid transporter and vesicular amine transporter. Genes encoding GABA transporter and vesicular glutamate transporter were up-regulated. The relative transcript abundances of these different enzyme encoding genes and neural transporter genes in the central nervous system were validated and characterized by qPCR, and the results showed that the direction of the expression was consistent between RNA seq and qPCR. However, fold change was more significant in RNA seq than qPCR for 3 genes: GS, GDH and PAH.

4 DISCUSSION

The mite *Varroa destructor* can cause severe mortality in honeybee populations worldwide. Despite the significant economic impact, our understanding about the genetic basis underlying the adaptation of *V. destructor* to honeybees is limited. To address this knowledge gap, the transcriptional profiles of mature daughter *V. destructor* mites infesting three different developmental stages of honeybees (brood cells, newly emerged bees and adult bees) were compared.

Our sequencing analysis revealed ~ 51 million raw reads in each sequenced mite sample, and 97.38% of these were clean reads (~ 7.49 Gb reads). Compared with the genome of *V. destructor* (368 Mb), the transcriptome sequence data represented ~12.5-fold coverage of the genome of *V. destructor*. This large dataset of transcripts provides new opportunities for further gene identification and development of molecular markers. These reads were assembled into 132,779 unigenes with an average length of 1,603 bp, a maximum length of 41,024 bp, and an N50 of 5,069 bp. The results of BUSCO analysis of *V. destructor* transcriptome obtained in the present study compared to a previously published transcriptome showed that our transcriptome assembly has a better quality and is more complete. Also, 23,030 high-quality SNPs and 95,470 SSRs were obtained, which can be used for the construction of high-quality genetic map. The

annotated genes were highly matched with those of the predatory mite *Metaseiulus occidentalis*, which was anticipated because both of *M. occidentalis* and *V. destructor* belong to order Parasitiformes.

In our study, 155 DEGs were found between S2 and S3. Also,1,492 DEGs were previously detected in mites collected from capped brood cells containing developing bees ready to emerge compared to mites infesting adult bees.[16] The majority of the DEGs in *V. destructor* mites infesting newly emerged bees and adult bees were down-regulated compared to mites infesting the sealed brood cells, suggesting that infestation of the more mature honeybee stages seems to be associated with suppression rather than activation of *V. destructor* genes. These results agree with a previous study, which showed that mites have a preference to adult honeybees, but seems inconsistent with the fact that more mites are found in the sealed brood cells than on adult bees.[25]

There were a small number of genes whose expression was discordant between mature daughter mites infesting the newly emerged bees and the adult bees. These DEGs were involved in cellular and metabolic processes, and were largely up-regulated between S2 *vs.* S3, suggesting increased dysregulation of biological functions and processes that were already dysregulated in mites infesting newly emerged bees. Preference for specific honeybee development stage can influence the mite's food intake and metabolism. A previous study showed that *V. destructor* mites consume fat body tissue rather than honeybee's hemolymph.[2] In our study, active fatty acid metabolism was detected, and numerous lipid metabolic pathways were affected during *V. destructor* infestation of different stages of honeybees.

The GO enrichment analysis showed that DEGs identified in the comparison between S2 *vs.* S1 and S3 *vs.* S1 were strongly associated with metabolic process and binding activities, suggesting that transcriptional changes observed in mites infesting different development stages of honeybees are related to metabolic processes. Oxidative phosphorylation is another pathway that was significantly affected, indicating that energy metabolism is involved in mite-honeybee interaction. Compared to S1, DEGs in both S2 and S3 were significantly involved in the proteasome pathway. Proteasome, found in the nucleus and cytoplasm of eukaryotic cells, provides the main

pathway for degradation of intracellular proteins.[26] Perturbed proteasome activities are associated with altered protein metabolism, which was also detected in the present study. This perturbation might facilitate adaption of mites to new hosts and as they shift between different stages of honeybees.

Varroa mites have expanded its host range by shifting from the eastern honeybee *Apis cerana* to the western honeybee *Apis mellifera.* It is sensible to anticipate the relationship between *Varroa* mites and their new bee *Apis mellifera* host to be less balanced and more harmful compared to their relationship with the original bee *Apis cerana* that has been established over a long period of co-evolution. Thus, a balanced *A. mellifera*-*V. destructor* relationship demands adaption strategies from both *V. destructor* and the honeybee host. Mite infestation significantly altered honeybee genes involved in embryonic development, cell metabolism and immunity.[27] *Varroa* sensitive hygiene (VSH) behavior is a crucial strategy employed by honeybees to detect and remove the brood infested by *Varroa* mites. Comparisons between the antennal transcriptome of VSH and non-VSH honeybees showed that the majority DEGs in VSH bees were up-regulated, and 30% of these DEGs were related to metabolism.[28] This suggests that differentially expressed metabolism-related genes of honeybees may contribute to honeybee's defense against mite infestation. On the other hand, transcriptome analysis of mature daughter mites infesting different developmental stages of honeybees showed that a large number of genes involved in energy and metabolic processes were affected, suggesting that adaptation of *Varroa* to honeybees is mediated by differential regulation of bioenergetic-related genes. Taken together, these data suggest that metabolic modulation is a key adaptation mechanism that shape *A. mellifera*-*V. destructor* relationship.

Another intriguing observation of our study was the alterations in the neurotransmitter regulation of mites. The γ-aminobutyric acid (GABA) is the major inhibitory neurotransmitter in insects, and GABA receptors are a common target of different insecticides.[29] Acetylcholine (ACh) is a major excitatory neurotransmitters in the insect central nervous system and glutamate is present at the neuromuscular junction.[30] Glutamine synthetase is mainly located in the cytosol of astrocytes and can

catalyze the ATP-dependent synthesis of glutamine.[31] Glutamate synthase is an essential enzyme involved in glutamine metabolism and can convert α-ketoglutarate and ammonia to glutamate.[32] Both glutamine synthetase and glutamate synthase were found down-regulated in this study. Glutamate decarboxylase, which catalyzes the production of GABA, was found up-regulated.[33] Genes encoding GABA transporter and vesicular glutamate transporter were up-regulated. These alterations in the neurotransmitter regulation might be involved in the mite's preference behavior for certain honeybee stages. Further assessment of the potential role of neurotransmitter related enzymes and transporters as pharmacological targets for future insect control strategies is warranted.

In conclusion, using state-of-the-art sequencing approach, we elucidated the transcriptome of mature daughter *V. destructor* mite during infestation of three developmental stages of honeybees. Information related to the expressed genes and their biological functions significantly expand the currently known gene repertoire of *V. destructor* and can guide further genetic studies on *V. destructor*. Our data may inform future studies on the molecular mechanisms underlying the relationship between transcriptional changes and adaptation of *V. destructor* mites to novel honeybee hosts.

ACKNOWLEDGEMENTS

This research was supported by the Modern Agro-industry Technology Research System (CARS-44-KXJ-6), the Agricultural Science and Technology Innovation Program (CAASASTIP-2020-IAR), the National Natural Science Foundation of China (Grant No. 31902221), the Central Public interest Scientific Institution Basal Research Fund (No.Y2019PT17-03) and the Open Foundation of Key Laboratory of Pollination Insect Biology, Ministry of Agriculture and Rural Affairs (No. 2018MFNZS04).

REFERENCES

1 Sammataro D, Gerson U and Needham G, Parasitic mites of honey bees: life history, implications, and impact. *Annu Rev Entomol* **45**:519-548 (2000).

2 Ramsey SD, Ochoa R, Bauchan G,Gulbronson C, Mowery JD, Cohen A,Lim D, Joklik J, Cicero JM, Ellis JD, Hawthorne D and vanEngelsdorp D. *Varroa* destructor feeds primarily on honey bee fat body tissue and not hemolymph. *Proc Natl Acad Sci U S A*, **116** (5): 1792-1801 (2019).

3 Nazzi F and Le Conte Y, Ecology of *Varroa destructor*, the major ectoparasite of the western honey bee, *Apis mellifera*. *Annu Rev Entomol* **61**:417-432 (2016).

4 Locke B, Natural *Varroa* mite-surviving *Apis mellifera* honeybee populations. *Apidologie* **47**: 467-482 (2016).

5 Barron AB, Death of the bee hive: understanding the failure of an insect society. *Curr Opin Insect Sci* **10**:45-50 (2015).

6 Wallner K, Varroacides and their residues in bee products. *Apidologie* **30**:235-248 (1999).

7 Maggi M, Damiani N, Ruffinengo S, De Jong D, Principal J and Eguaras M, Brood cell size of *Apis mellifera* modifies the reproductive behavior of *Varroa destructor*. *Exp Appl Acarol* **50**(3): 269-279 (2010).

8 Bogdanov S, Contaminants of bee products. *Apidologie* **37**:1–18 (2006).

9 Grozinger CM and Robinson GE, The power and promise of applying genomics to honey bee health. *Curr Opin Insect Sci* **10**:124-132 (2015).

10 Zanni V, Galbraith DA, Annoscia D, Grozinger CM and Nazzi F, Transcriptional signatures of parasitization and markers of colony decline in *Varroa*-infested honey bees (Apis mellifera). *Insect Biochem Mol Biol* **87**:1-13 (2017).

11 Zhang Y, Liu X, Zhang W and Han R, Differential gene expression of the honey bees *Apis mellifera* and *A. cerana* induced by *Varroa destructor* infection. *J Insect Physiol* **56**: 1207-1218 (2010).

12 Gregory PG, Evans JD, Rinderer T and de Guzman L, Conditional immune-gene suppression of honeybees parasitized by *Varroa* mites. *J Insect Sci* **5**:7 (2005).

13 Rosenkranz P, Aumeier P and Ziegelmann B, Biology and control of *Varroa destructor*. *J Invertebr Pathol* **103**:S96-119 (2010).

14 Mondet F, Rau A, Klopp C, Rohmer M, Severac D, Le Conte Y and Alaux C, Transcriptome profiling of the honeybee parasite *Varroa destructor* provides new biological insights into the mite adult life cycle. *BMC Genomics* **19**:328 (2018).

15 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**:644-652 (2011).

16 Simão F A, Waterhouse R M, Ioannidis P, Kriventseva EV and Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**: 3210-3212 (2015).

17 Langmead B and Salzberg SL, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359 (2012).

18 Thiel T, Michalek W, Varshney RK and Graner A, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* **106**:411-422 (2003).

19 Anders S and Huber W, Differential expression analysis for sequence count data. *Genome*

Biol **11**:R106 (2010).

20 Howe E. HK, Nair S., Schlauch D., Sinha R., Quackenbush J. *MeV: MultiExperiment Viewer*, In: Ochs M., Casagrande J., Davuluri R. (eds) Biomedical Informatics for Cancer Research. Springer, Boston, MA, pp. 267–277 (2010).

21 Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C and Apweiler R, The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* **37**: D396-403 (2009).

22 Kanehisa M, Goto S, Sato Y, Furumichi M and Tanabe M, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**:D109-114 (2012).

23 Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen LJ, STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**:D808-815 (2013).

24 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**:2498-2504 (2003).

25 LeDoux MN, Pernal SF, Higo HA and Winston ML. Development of a bioassay to test the orientation behaviour of the honey bee ectoparasite, *Varroa jacobsoni*. *J Apic Res* **39**: 47-54 (2000),.

26 Adams J. The proteasome: a suitable antineoplastic target. *Nature Rev Cancer* **4**: 349 (2004).

27 Rosenkranz P, Aumeier P, Ziegelmann B. Biology and control of Varroa destructor. *J Invertebr Pathol* **103**: S96-S119 (2010),.

28 Mondet F, Alaux C, Severac D, Rohmer M, Mercer AR and Le Conte Y, Antennae hold a key to *Varroa*-sensitive hygiene behaviour in honey bees. *Sci Rep* **5**:10454 (2015).

29 Ménard C, Folacci M, Brunello L,Charreton M, Collet C, Mary R,Rousset M, Thibaud JB, Vignes M, Charnet P and Cens Tl. Multiple combinations of RDL subunits diversify the repertoire of GABA receptors in the honey bee parasite *Varroa destructor*. *J Bio Chem* **293**: 19012-19024(2018).

30 Gauthier M, Grünewald B, Neurotransmitter Systems in the Honey Bee Brain: Functions in Learning and Memory. In: Galizia CG, Eisenhardt D, Giurfa M (eds) Honeybee neurobiology and behavior. Springer, Dordrecht, pp. 155–169 (2011).

31 Silvina B, Nicolás C, Mauricio M, Federico I, Ari Z, Christiane O, Tilman G, Darío E and Rafael R, Insights into the mechanisms of peroxynitrite-mediated inactivation of human glutamine synthetase. *Free Radical Biology and Medicine* **120**:S41 (2018).

32 Temple SJ, Vance, CP., Gantt JS, Glutamate synthase and nitrogen assimilation. *Trends Plant Sci* **3**:51-56 (1998).

33 Kaufman D, Houser C and Tobin A, Two forms of the gamma-aminobutyric acid synthetic enzyme glutamate decarboxylase have distinct intraneuronal distributions and cofactor interactions. *J Neurochem* **56**: 720-723 (1991).

Legends to figures:

Figure 1. Length distribution of the assembled sequences. The length of unigenes and transcripts are shown on the x-axis, and the number is indicated on the y-axis.

Figure 2. Characteristics of homology search of unigenes against the NR nucleotide database. (A) E-value distribution of BLAST hits for each unigene with a cut off E-value of 1.0E-5. (B) Similarity distribution of the top BLAST hits for each unigene. (C) Species distribution is shown as percentage of the total homologous gene hits.

Figure 3. A comparison of the completeness of the newly sequenced *Varroa* transcriptome to a previously published *Varroa* mite transcriptome. The results of the BUSCO analysis showed that the assembled transcriptome in the present study is more complete than the previously published transcriptome.

Figure 4. Cluster of Orthologous Groups (COG) analysis of the identified genes in *V. destructor*. The x-axis label represents different COG classes. The y-axis labels represents the number of unigenes.

Figure 5. Frequency of the most abundant simple sequence repeats (SSRs) motifs identified in *V. destructor* transcriptome.

Figure 6. Comparisons of gene expression levels among the different test groups. (A) Box plot showing the distribution of expression values of all the transcripts. x-axis label indicates different groups; y-axis label indicates $\log_{10}$ (FPKM+1) values in RNA-Seq data. (B) FPKM density distribution for all transcripts. (C) Number of genes at different expression levels.

Figure 7. Differently expressed Genes (DEGs) in differently compared libraries.

(A) Volcano plot of the DEGs between S1 and S2. (B) Venn diagram of the DEGs among S1, S2 and S3 groups. (C) Heatmap of DEGs among all transcripts based on FPKM units. Blue color indicates low expression and red color denostes high expression.

Figure 8. Functional analysis of the DEGs. (**A**) Gene ontology analysis of DEGs. Unigenes were summarized into three main categories, namely biological processes [BP], cellular components [CC], and molecular function [MF]. The X-axis label denotes the number of unigenes, whereas the Y-axis label represents the unigenes' respective GO terms. (**B**) The top 20 significantly enriched KEGG pathways of the DEGs. The x-axis label shows the rich factor. Rich factor represents the number of DEGs/total number of genes in the KEGG pathway. The larger the value, the greater the enrichment. The y-axis label shows the KEGG pathways. The color of the dots represents *q* value and the size of the dot represents the number of DEGs enriched in the pathway.

Figure 9. Statistics of the number of DEGs involved in (A) amino acid metabolism and (B) lipid metabolism.

Figure 10. Differentially expressed genes (DEGs) encoding neurotransmitter enzymes were verified by qPCR. The expression trends of three DEGs, including glutamate synthase (GLT), acetycholinesterase (ACHE) and glutamate decarboxylase (GAD), were found to be similar to those obtained by RNA-seq, suggesting that the RNA-seq data reliably reflected the gene expression trends. However, fold change was more significant in RNA seq than qPCR for glutamine synthetase (GS), glutamate dehydrogenase (GDH), phenylalanine hydroxylase (PAH).

SUPPORTING INFORMATION

Figure S1. Distribution of simple sequence repeats (SSRs) in the different classes.

Figure S2. Volcano plots of relative gene expression. (A) Volcano plot of the DEGs between S3 and S1. (B) Volcano plot of the DEGs between S3 and S2.

Figure S3. Gene ontology analysis of DEGs between S3 and S1. Unigenes were categorized according to the three categories biological processes (BP), cellular components (CC) and molecular function (MF). The x-axis label denotes the number of unigenes and the y-axis label represents the unigenes' respective GO terms.

Figure S4. KEGG enrichment analysis of the DEGs between S3 and S1.

Figure S5. Differentially expressed genes encoding neurotransmitter transporters verified by qPCR.

Table S1. List of primers used in real-time quantitative PCR analysis.

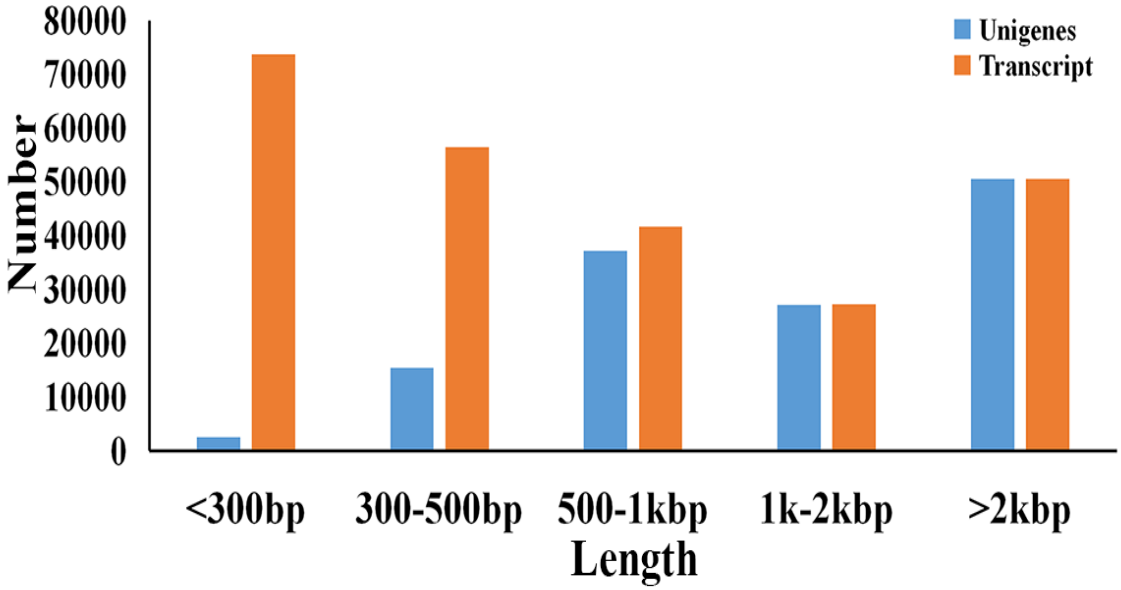
Table S2. Summary of single nucleotide polymorphisms (SNPs) identified from the RNA-Seq data.

Table S3. Mapping rate.

Table S4. KEGG pathway analysis of the DEGs identified between S2 and S1.

Table S5. KEGG pathway analysis of the DEGs identified between S3 and S1.

Figure 1



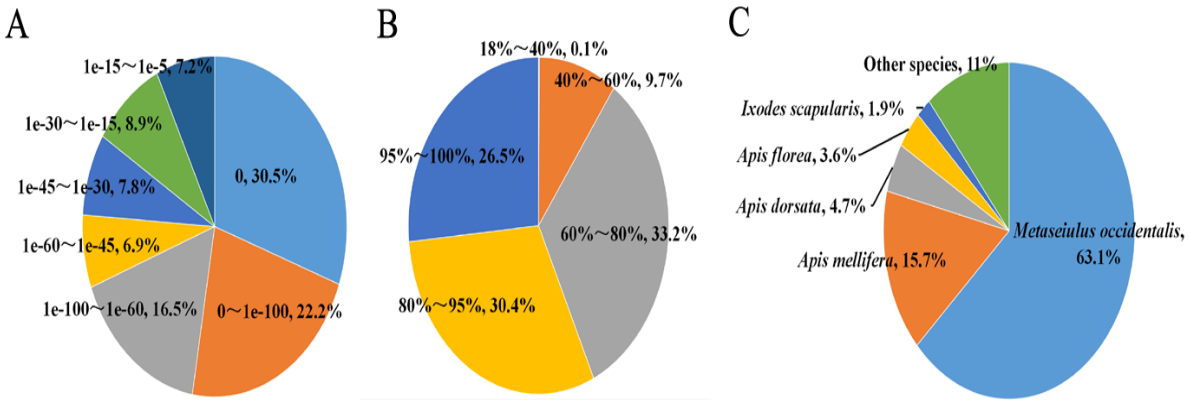
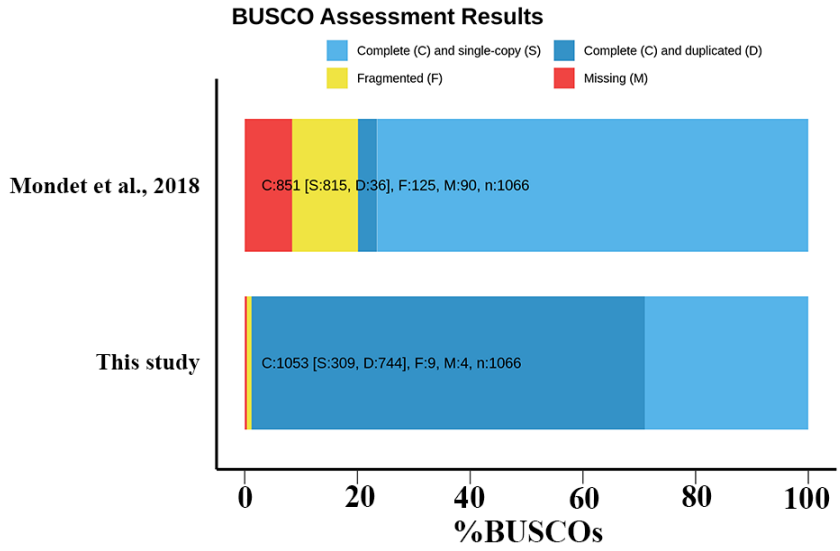


Figure 2

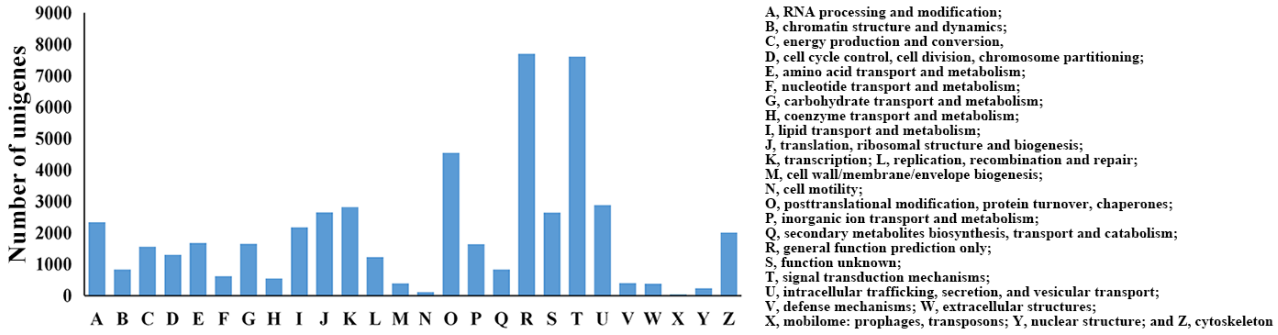


613

614 **Figure 3**

615

616



617

618

619 **Figure 4**

620

621

622

623

624

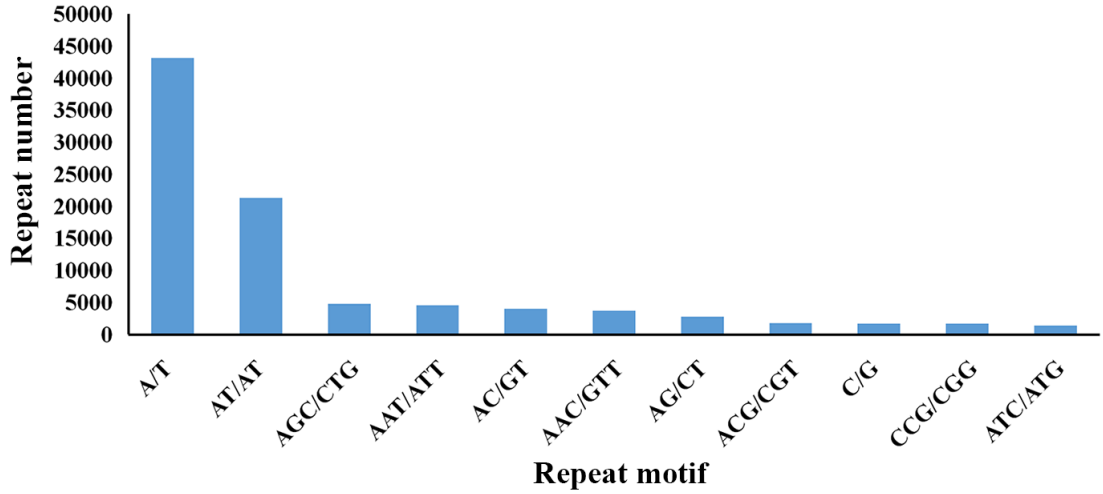
625

626

627

628

629



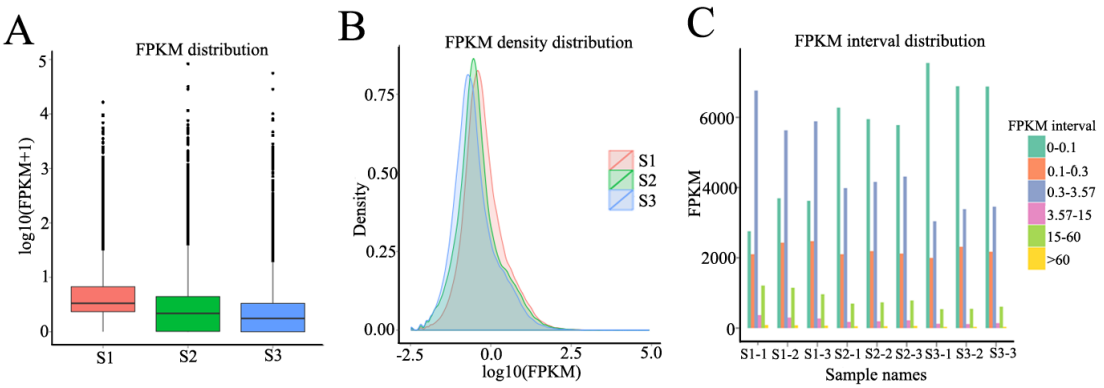
630

631 **Figure 5**

632

633

634



635

636

637 **Figure 6**

638

639

640

641

642

643

644

645

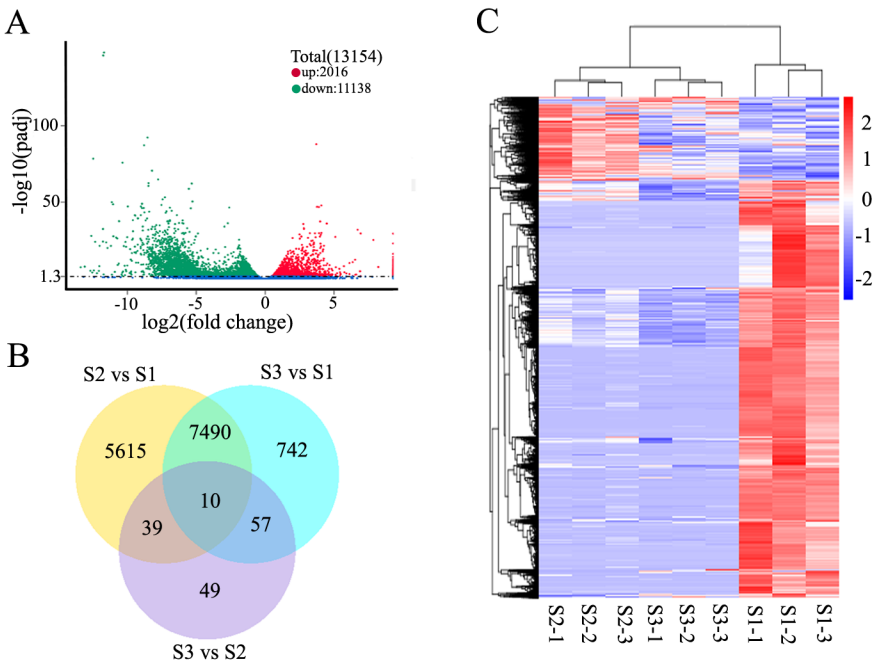


Figure 7

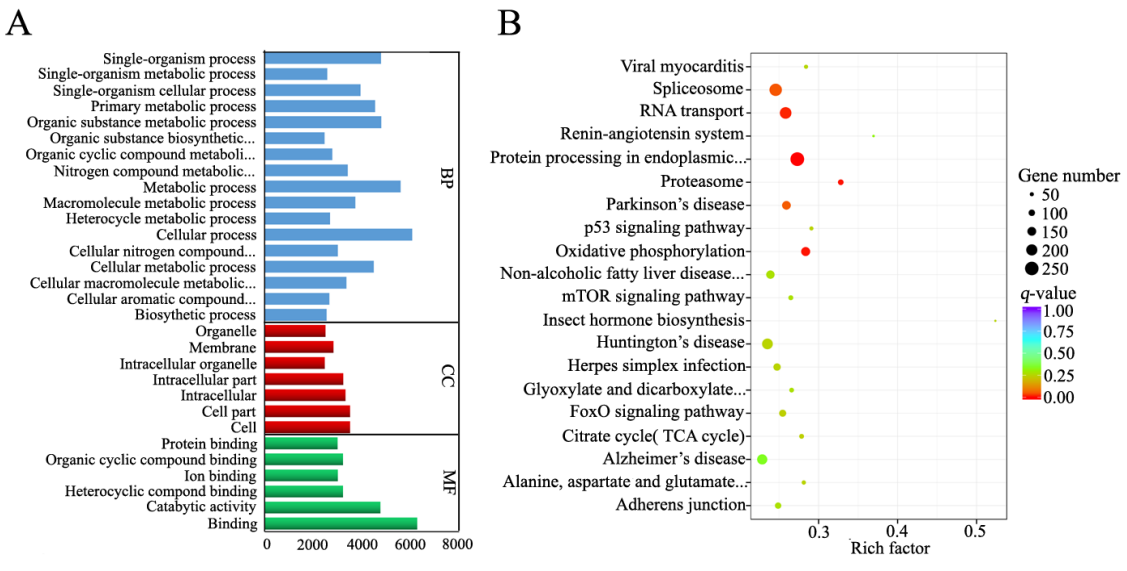


Figure 8

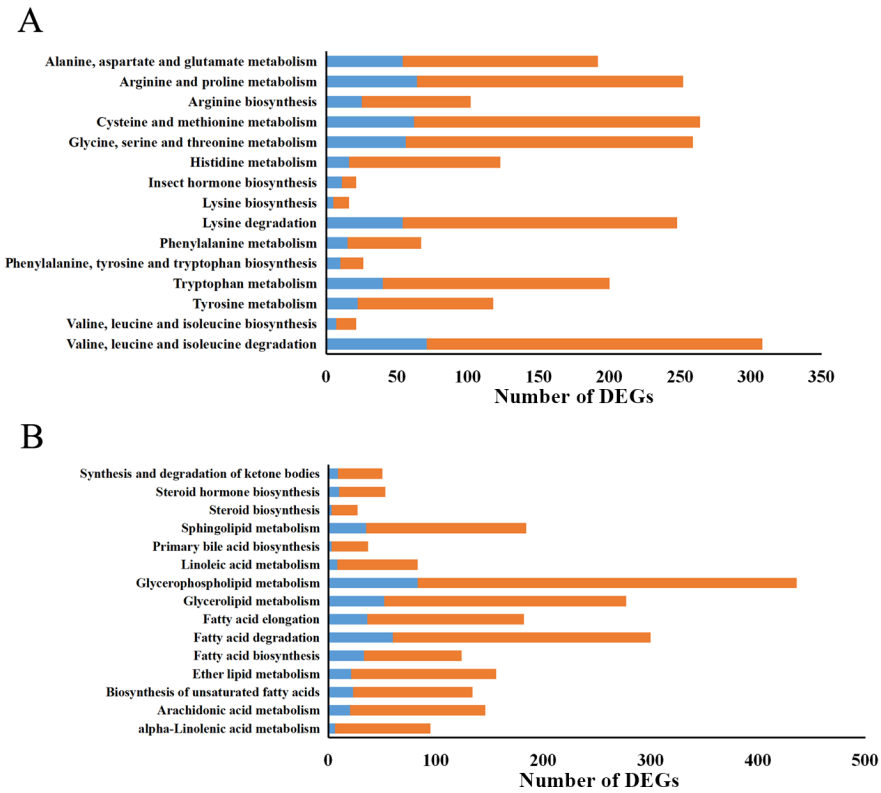


Figure 9

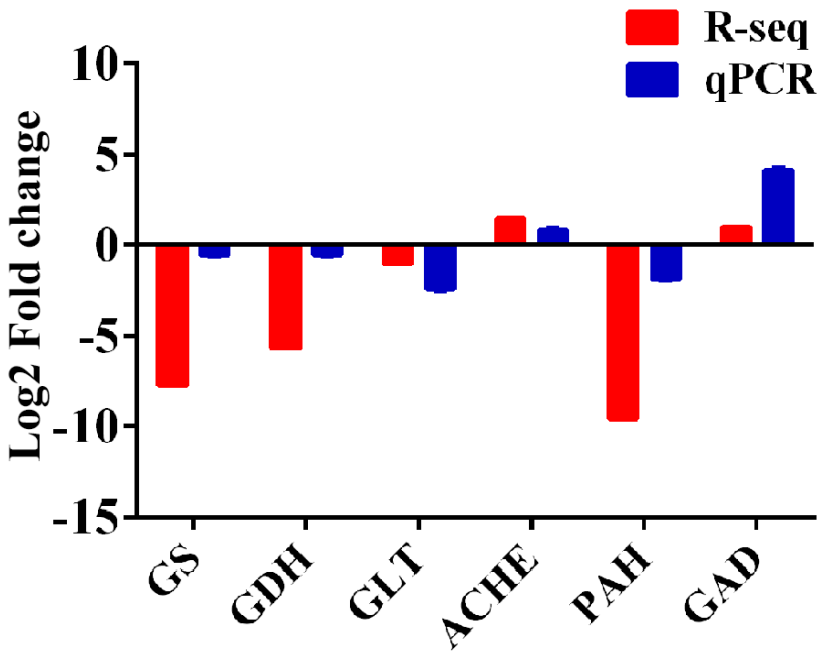


Figure 10

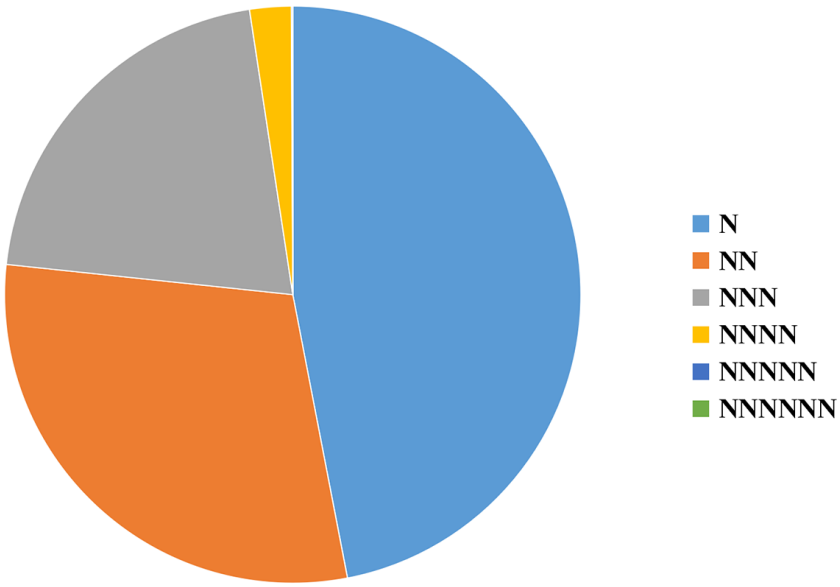


Figure S1

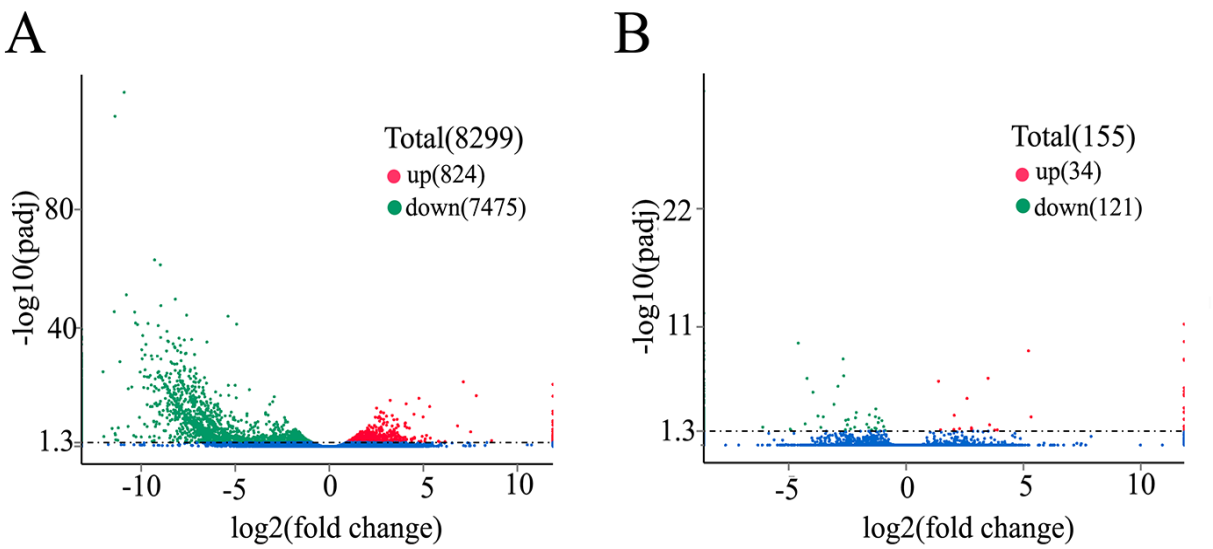
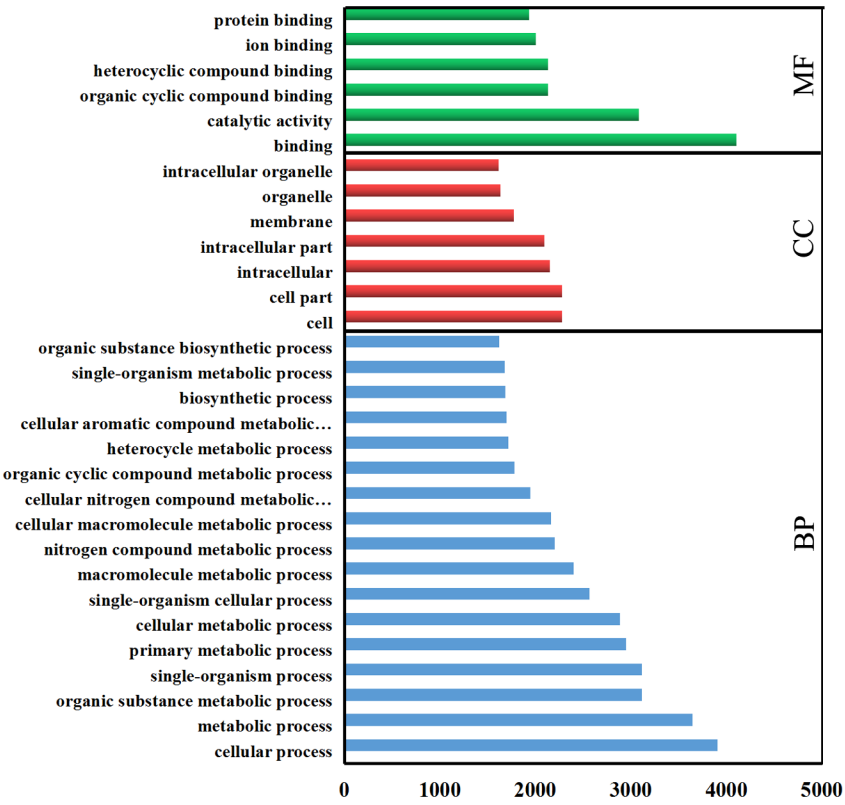
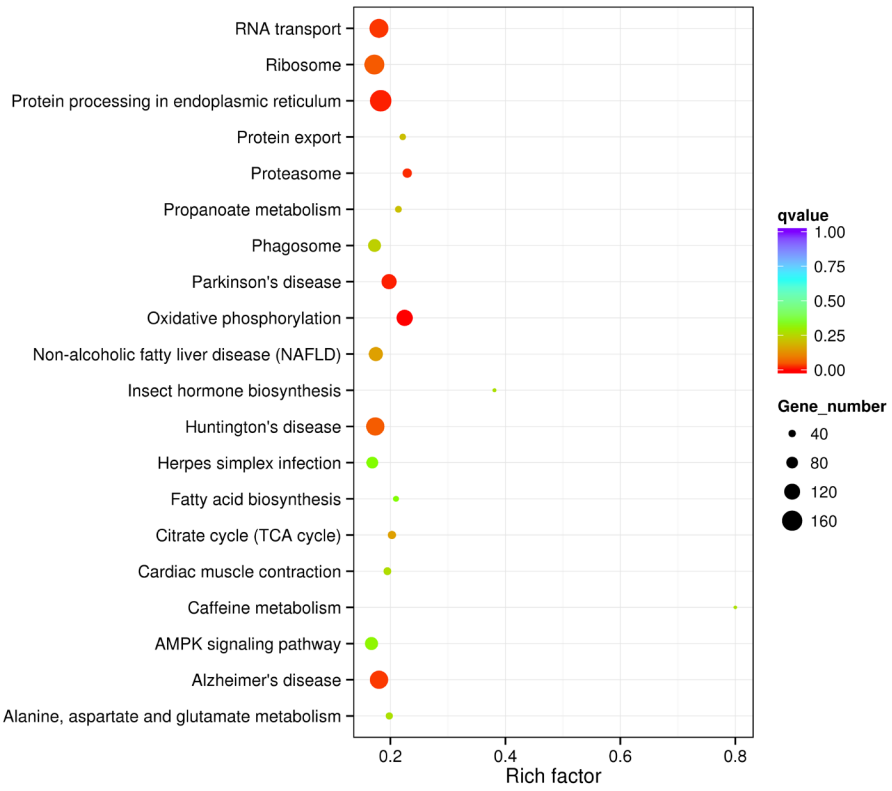


Figure S2



678

679 **Figure S3**



680

681 **Figure S4**

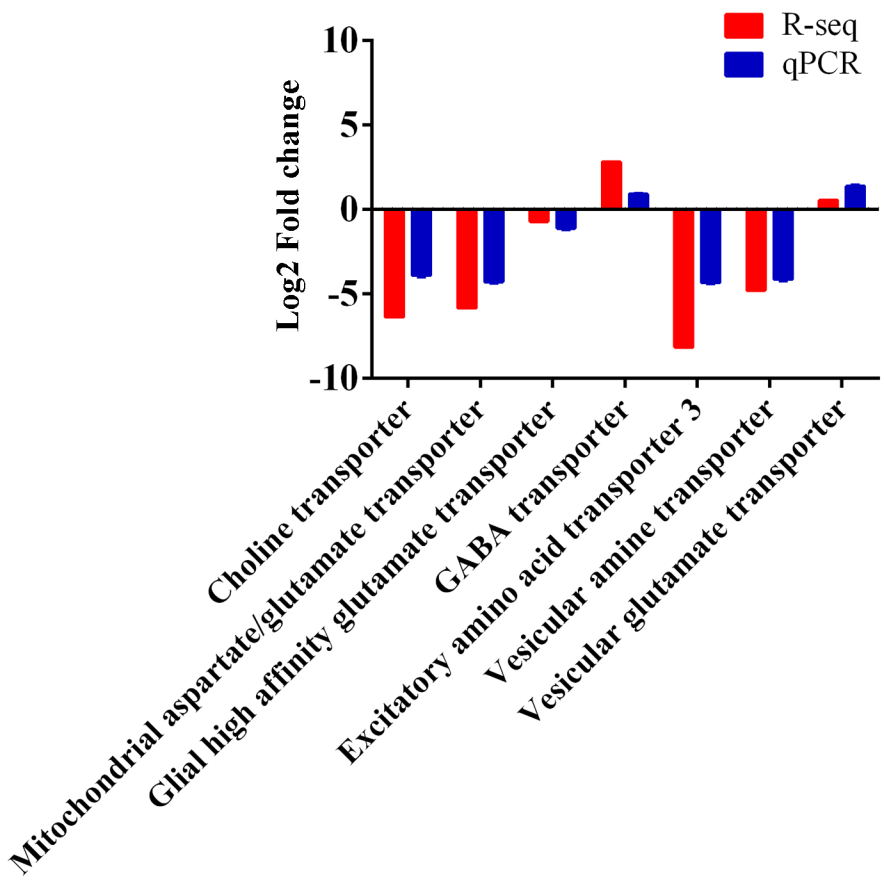

R-seq
qPCR
Log2 Fold change
10
5
0
-5
-10
Choline transporter
Mitochondrial aspartate/glutamate transporter
Glial high affinity glutamate transporter
GABA transporter
Excitatory amino acid transporter 3
Vesicular amine transporter
Vesicular glutamate transporter

Figure S5

705

706 **Table 1. Overall quality chrematistics of the obtained sequences.**

Sample	Raw reads	Clean reads	Clean bases	Error (%)	Q20 (%)	Q30 (%)	GC (%)
S1-1	54,378,534	52,106,480	7.82G	0.02	95.74	89.84	43.60
S1-2	51,026,944	49,799,612	7.47G	0.02	96.71	91.71	41.93
S1-3	51,252,838	49,975,004	7.5G	0.02	96.68	91.60	42.19
S2-1	47,660,264	46,543,174	6.98G	0.02	96.86	91.97	41.83
S2-2	46,079,074	44,978,124	6.75G	0.02	96.91	92.06	41.91
S2-3	47,945,260	46,826,616	7.02G	0.02	96.84	91.94	42.77
S3-1	54,792,998	53,466,454	8.02G	0.02	97.02	92.30	40.62
S3-2	56,699,156	55,317,900	8.3G	0.02	97.16	92.59	40.50
S3-3	51,391,404	50,129,476	7.52G	0.02	97.02	92.33	40.75

707

Table 2. Summary statistics of RNA sequencing data obtained from *V. destructor.*

Transcript	Number of transcripts	249,505
	Mean length of transcripts	1,603
	N50 of transcripts	5,069
	Minimum length	201
	Mean length	1,603
	Median length	470
	Maximum length	41,024
Unigenes	Number of unigenes	132,779
	Mean length of unigenes	2,745
	N50 of unigenes	5,706
	Minimum length	201
	Mean length	2,745
	Median length	1,304
	Maximum length	41,025

730 **Table 3 The number of unigenes annotated in public database.**

Annotated in NR	68,940
Annotated in NT	28,984
Annotated in KO	37,658
Annotated in Swiss-Prot	55,713
Annotated in Pfam	64,334
Annotated in GO	64,509
Annotated in KOG	44,672
Annotated in all databases	12,832
Annotated in at least one database	82,068

731

732

733

734 **Table S1 List of primers used in real-time quantitative PCR analysis.**

Gene name	GeneBank accession	Primers	Sequences
Glutamine synthetase	XM_001123051.4	F	5'-AATAAATACCTCGACCTGCCTCA -3'
		R	5'-TGTAAATTGCCACTGGACAAAG-3'
Mitochondrial aspartate/glutamate transporter	XM_006564542.2	F	5'-AATCACGAAACGCTTAGCAGA-3'
		R	5'-CCACCTATGGACCCTAGAACG-3'
Acetylcholinesterase	XM_022801084.1	F	5'-GAAGCCGTCACAATGCGTTTG-3'
		R	5'-GATTTGCGAAGTGCTGGTGGG-3'
GABA transporter	XM_022794855.1	F	5'-GCGTACACTCTTTACTACTTTGTGGC-3'
		R	5'-CGATGTTCCGTCGGTTCTATT-3'
Phenylalanine hydroxylase	XM_623297.5	F	5'-ATGGGCAACCAATACCAAGAGT-3'
		R	5'-GCGAAGAAAGGAGACCAGCAAC-3'
Glutamate dehydrogenase	XM_392776.5	F	5'-GAAGATAAATTGGTCGAAGATATTGGC-3'
		R	5'-TCGCGTCTCAGAGGAAAGGAAGT-3'
Vesicular amine transporter	XM_016916695.1	F	5'-CGTAGCACCTTGTTCCCATCT -3'
		R	5'-CTTCAACCGTCCGTCGTGT-3'
Excitatory amino acid transporter	XM_395840.6	F	5'-TCAAAGATGAACCGTGGACAAA-3'
		R	5'-GCTAGCGACGATTAAAGGCAAT-3'
Glial high affinity glutamate transporter	XM_022805322.1	F	5'-CTCATGGACCTCATCAGAAATG-3'
		R	5'-CAGGGCGATACAGAACACG-3'
Choline transporter	XM_017058422.1	F	5'-ATGTGGACTCAACTTGGATAAA-3'
		R	5'-CACAAGCTACAGAACCTAAGTGAT-3'
Glutamate synthase	XM_022799273.1	F	5'-TCCACGCTCGCTTCTCAAC-3'
		R	5'-GGCTCTACGACGGGATACAA-3'
Glutamate decarboxylase	XM_022844535.1	F	5'-GTCGTTTGATACGGGTGATA-3'
		R	5'-TTGGTCCCTTAACCTTTGG-3'
Vesicular glutamate transporter	XM_022812530.1	F	5'-AATCAAACATTATGACCCGAACC -3'
		R	5'-TGCCAAAGCCCATCACAGA-3'
GAPDH	XM_022799398.1	F	5'-ATGGGTGTCAACCACGAGAAGT -3'
		R	5'-TGTTGCATGTACCGTTGTCATTAG-3'

735
736
737
738
739
740
741
742
743

744

745 **Table S2 Summary of single nucleotide polymorphisms (SNPs) identified from the**

746 **RNA-Seq data.**

Type		Number of counts
Transition	C/T	9,509
	A/G	9,775
Transversion	A/T	925
	A/C	904
	T/G	887
	C/G	1,030

747

748

749

750

751

752

753 **Table S3 Mapping rate**

Sample name	Total reads	Total mapped
S1-1	45,508,724	39,674,964 (87.18%)
S1-2	38,097,826	33,963,030 (89.15%)
S1-3	39,057,800	33,751,264 (86.41%)
S2-1	31,554,312	28,541,324 (90.45%)
S2-2	32,017,116	28,905,060 (90.28%)
S2-3	37,136,436	33,247,750 (89.53%)
S3-1	30,272,770	27,888,034 (92.12%)
S3-2	33,716,384	31,150,580 (92.39%)
S3-3	29,729,392	27,250,768 (91.66%)

754