Infection, Genetics and Evolution 75 (2019) 104008

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

De novo transcriptome sequencing and comparative analysis of *Haemaphysalis flava Neumann*, 1897 at larvae and nymph stages

Jiaying Guo[a,c,1], Yali Sun[a,c,1], Xiaoying Luo[a,c], Muxiao Li[a,c], Pei He[a,c], Lan He[a,b,c,*], Junlong Zhao[a,b,c,*]

[a] State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei 430070, China
[b] Key Laboratory of Animal Epidemical Disease and Infectious Zoonoses, Ministry of Agriculture, Huazhong Agricultural University, Wuhan, Hubei 430070, China
[c] Key Laboratory of Preventive Veterinary Medicine in Hubei Province, Wuhan, Hubei 430070, China

ARTICLE INFO

Keywords:
Haemaphysalis flava Neumann, 1897
Transcriptome
Larvae
Nymph

ABSTRACT

Haemaphysalis flava Neumann, 1897 (*H. flava*) is of public health significance due to its capability of transmitting several pathogens such as *Rickettsia*, *Ehrlichia*, *Bartonella* and *Francisella tularensis*. However, lack of complete genome, transcriptome and proteome information limits the understanding of the biology of *H. flava*. Here, the total RNA of *H. flava* was collected separately at larvae and nymph stages and analyzed with high-throughput RNA sequencing technology. The obtained data were assembled and annotated based on the near origin species in the Nr database. The functions of the unigenes were annotated and classified by seven databases, including Nr, Nt, Pfam, KOG, Swiss-Prot, GO and KEGG. A total of 61,850,967 and 79,579,368 clean reads were obtained with a data bulk of 9.28 G and 11.94 G in larvae and nymph stages, respectively. The number of unigenes was 440,896, with 48.6% of them being matched to the Nr database and 51.4% remaining unknown. Additionally, 1,776,404 SNPs were identified in the unigenes. Differential analysis revealed 80 differentially expressed genes (DEGs), including 56 up-regulated genes and 24 down-regulated genes in the nymph versus larvae. qPCR confirmed 4 of the 56 up-regulated genes and 4 of the down-regulated genes. KEGG analysis of the DEGs showed that aldehyde dehydrogenase and sorbitol dehydrogenase, two up-regulated unigenes in nymph versus larvae, were both matched to the top three enriched pathways: "chloroalkane and chloroalkene degradation", "fatty acid degradation" and "glycolysis and gluconeogenesis". This is the first report on the whole transcriptome of *H. flava* at larvae and nymph stages. This study contributes to the understanding of *H. flava* at the gene expression level in different developmental stages and provides a theoretical basis for the development of vaccines against *H. flava*.

1. Introduction

Ticks are blood-feeding ectoparasites with medical and veterinary importance and survive in the four growth forms of eggs, larvae, nymphs and adults (Tissot Dupont and Raoult, 1993; Kettle, 1995). There are many different species in ticks, and to date, approximately 900 species have been described, which can be divided into three families: the Argasidae (191 species), the Ixodidae (701 species) and the Nuttalliellidae (only one species, *Nuttalliella namaqua*) (Horak et al., 2002). Among them, *Haemaphysalis flava Neumann*, 1897 (*H. flava*) belongs to the order Acarina of the class Arachnida and the three-host tick (Ozawa et al., 1982; Fujimoto, 1987; Kakuda et al., 1990). It exhibits the three-host life cycle with (larvae, nymph and adult) each feeding once on their respective host during the life cycle, followed by moulting of the detached larvae and nymph to the nymph and adult, respectively, and finally, the oviposition of the engorged female, which mated on the host, on the ground after detachment (Qiu et al., 2014; Xu et al., 2016; Liu et al., 2017). Previous studies have shown that it can transmit different pathogen species on the Asiatic continent (Japan, Tailand, Tibet …) (Furuno et al., 2017; Jung et al., 2019). Furthermore, its parasitic hosts are widely distributed, but the main host is dog (Xu et al., 2015; Duan and Cheng, 2017). In addition, it is also found on

Abbreviations: H. flava, *Haemaphysalis flava Neumann*, 1897; DEGs, differentially expressed genes; cDNA, complementary DNA; CDS, coding sequences; SSR, simple sequence repeats; SNP, single nucleotide polymorphism; qPCR, quantitative real-time PCR; *B. microplus*, *Boophilus microplus*

* Corresponding authors at: State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei 430070, China.
E-mail addresses: helan@mail.hzau.edu.cn (L. He), zhaojunlong@mail.hzau.edu.cn (J. Zhao).
[1] The authors did the equal work.

https://doi.org/10.1016/j.meegid.2019.104008
Received 26 February 2019; Received in revised form 11 August 2019; Accepted 16 August 2019
Available online 19 August 2019

wild animals such as giant panda in the Qinling Mountain area of China (Cheng et al., 2013).

The ticks can carry a variety of pathogenic microorganisms such as *Rickettsia*, *Ehrlichia*, *Bartonella* and *Francisella tularensis*, resulting in the transmission of diseases in humans, domestic and wild animals (Pfaffle et al., 2013; Ejiri et al., 2018). Moreover, during the feeding process, they can induce inadvertent damage to the host's blood vessels and tissues at the bite site. They can also directly cause severe toxic conditions leading to irritation, allergy and paralysis of the host (Ishigaki et al., 2012). Therefore, it is crucial and necessary to inhibit the growth and spread of the ticks. Currently, polyester, organophosphorus, and amidine are widely used as broad-spectrum insecticides for controlling the tick populations in animal husbandry (Pfaffle et al., 2013). However, these drugs may cause the pollution of environment and animal by-products, create high antibiotic-resistance in ticks and increase their prevention and control cost. This suggests that the development of new strategies against the ticks has become an extremely urgent task at present.

With the rapid development of sequencing, genome and transcriptome combined with bioinformatics analysis have grown enormously and made great contributions to the identification and annotation of new genes (Verjovski-Almeida et al., 2003; Egekwu et al., 2014; Yin et al., 2016; Charrier et al., 2018). For example, the genome of *Ixodes scapularis* was analyzed as a member of the subphylum Chelicerata (Miller et al., 2018). In 2006, the genome project was finished for *Boophilus microplus* (*B. microplus*), the southern cattle tick as the most economically important tick all over the world (Guerrero et al., 2006). However, there is no information available about the genome of *H. flava,* and lack of transcriptome and proteome information limits the understanding of *H. flava* in many aspects, such as molecular biology, biochemistry and biological evolution.

In this study, the transcriptome of *H. flava* at larvae and nymph stages was sequenced and compared, and the differentially expressed genes (DEGs) in different stages were obtained, which may contribute to developing new targets for controlling *H. flava* and facilitate the understanding of the pathogenesis of *H. flava* at different stages.

2. Materials and methods

2.1. Tick rearing and sample collection

Adult ticks (23 fully engorged females, 99 males) were collected from a wild hedgehog in Hubei province of China. The engorged ticks were reared in the dark at a temperature of 28 ± 1 °C in an 88% relative humidity biochemical desiccator for culture of next generation until the oviposition of the female ticks. The larvae and nymph samples were collected separately when eggs molted to larvae and engorged larvae molted to the nymph. Meanwhile, *H. flava* larvae were fed on white New Zealand rabbit.

2.2. RNA extraction, cDNA library construction and sequencing

The total RNA was extracted from a pool of 30 *H. flava* larvae and 30 nymphs using the TRIzol Reagent (Invitrogen, CA, USA). The total RNA was quantified by agarose gel electrophoresis to test RNA degradation and contamination and the RNA integrity and quantity was further tested using agilent 2100 bioanalyzer (Agilent technologies, CA, USA). Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, MA, USA) following manufacturer's recommendations 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 (Epicenter, Madison, WI, USA) and the rRNA was removed by using NEBNext rRNA depletion kit (NEB, MA, USA). Then 2 μg mRNA was fragmented randomly in fragmentation buffer, followed by complementary DNA (cDNA) synthesis and the addition of a custom second-strand synthesis buffer (Illumina, CA, USA) to generate the second strand by nick-translation with the reaction system (50 μl) of 1 μl dNTPs Mixture (10 mM), 4 μl RNase H (5 U/μl), 5 μl second strand cDNA synthesis buffer (10×), 4 μl *Escherichia coli* ligase (10 U/μl), 4 μl *E. coli* polymerase I (5 U/μl) and DEPC water. The cDNA was purified using AMPure XP beads (Beckman Coulter), and the cDNA library was constructed after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment. The DNA concentration was determined by a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and diluted to 1 ng/μl before insert size analysis by agilent 2100 Bioanalyzer (Agilent technologies, USA), followed by quantitative PCR (library activity > 2 nM). The library was sequenced on an Illumina Hiseq X platform with 150-base paired-end reads (Novogene Bioinformatics Technology Company, Beijing, China). The sequencing data obtained in this study were deposited in the National Center for Biotechnology Information (NCBI, accession number: SRR8426798).

2.3. Data filtering, de novo assembly and functional annotation

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. The used software was ng_qc designed by Novogene Bioinformatics Technology Company. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N (the proportion > 0.1%) and low quality reads that the number of bases with a mass value of Qphred ≤ 20 accounts for > 50% of the total reads. Because of the absence of *H. flava* genome as a reference, clean reads are de novo assembled with the reported genome sequence of *Ixodes scapularis* by Trinity software (r20140413p1) for transcriptome reconstruction (Grabherr et al., 2011). A comprehensive gene functional annotation was performed using seven databases including NCBI non-redundant protein sequences (Nr) (NCBI blast 2.2.28+, e-value = 1e-5), NCBI nucleotide sequences (Nt) (NCBI blast 2.2.28+, e-value = 1e-5), Protein family (Pfam) (HMMER 3.0 package, hmmscan, e-value = 0.01), Cluster of Orthologous Groups of proteins (COG) and euKaryotic Orthologous Groups (KOG) (NCBI blast 2.2.28+, e-value = 1e-3), Swiss-Prot (NCBI blast 2.2.28+, e-value = 1e-5), Kyoto Encyclopedia of Genes and Genome (KEGG) (KAAS, KEGG Automatic Annotation Server, e-value = 1e-10), and Gene Ontology (GO) (Blast2GO v2.5, e-value = 1e-6). The genes were annotated with each database and the predicted protein coding sequences (CDS) of the unigenes were blasted based on the priority of NR and Swiss-prot databases. If the information matched, CDS is extracted from unigene sequences and translated into peptide sequences based on the standard codon table (from 5′ to 3′). If not, unigenes with no hits in BLAST are analyzed with ESTScan (3.0.3) to predict their coding regions and determine their sequence direction. MISA (v1.0) was used for the simple sequence repeats (SSR) detection of the unigenes. GATK3 was used to carry out single nucleotide polymorphism (SNP). Clean reads were mapped back to the transcriptome by Bowtie (v2.0) for quantification of the expression level by RSEM (Li and Dewey, 2011). Finally, the read count was obtained for each gene of each sample and normalized to the FPKM value.

2.4. Gene expression difference analysis

Several methods for analyzing the differential expression genes (DEGs) were compared in terms of distribution, housekeeping, clustering, intra-variance and false-positive rate. DESeq and TMM were found to be more standardized than FPKM, although the latter was regarded as the standardized method for analysis of differential genes. In this study, DEGs were screened in the readcount data by DESeq, with a screening threshold (padj) of < 0.05. Meanwhile, the DEGs were ranked by $\log_2$ fold-change values and the functions of the DEGs were analyzed using the GO (Blast2GO v2.5, e-value = 1e-6) and KEGG (KAAS, KEGG Automatic Annotation Server, e-value = 1e-10) databases.

Table 1
Primers for validation of transcriptome sequencing by qPCR.

Gene name	Forward primer (5′-3′)	Reverse primer (3′-5′)
Actin	CAAGGAGAAGCTTTGCTACGTC	CCGATGGTGATGACCTGAC
C326998_g1	CACAAGATGTGGCTTGTTCA	CACACTTGGTACGATGGCTT
C313786_g1	TGTGTGTTCAATCCTGATGC	CAGGTCAGTAVGGTTGTTGC
C320142_g1	TCAGCGAGGTTGAATACGAT	GTCAGCCTTCGGAAGTTCTC
C290670_g1	TGTTCTACAGGAACGGCATC	CTTCTCCGTATCCGACCAAT
C215729_g2	ACTCTTCTGGGACCCTTCCT	GGACGTACAGGGTGTCCTCT
C311129_g3	ATGGTGCAGGAATCGTTTG	ACATCCGAGGGTGTAGCAG
C301668_g1	GTTTCATGTCACCAATCCCA	TCGAGCATCTTCACTACGGT
C94915_g1	TGAAGCCACAACTTCCTTTG	AGCCGAAGAAGGCAAGAATA

2.5. cDNA synthesis and quantitative real-time PCR

cDNA was synthesized with the RNA of *H. flava* at larvae and nymph stages using PrimeScript™RT reagent Kit (Takara Biotechnology, Beijing, China). The cDNA was stored at −20 °C for further use. Quantitative real-time PCR (qPCR) was performed to validate the transcriptome sequencing results. The primers for qPCR were designed based on eight DEGs, including four up-regulated and four down-regulated genes in the nymph versus larvae. The four up-regulated genes included endochitinase, cysteine proteinase, cytochrome and chitin, and four down-regulated genes included lipase, cystatin and other two hypothetical protein genes. The actin gene was used as a reference gene to standardize and optimize the expression level. The primers were designed using GenScript Real-time PCR (TaqMan) PrimerDesign (http://www.genscript.com/SSI-bin/app/primer). All primers and targeted genes are shown in Table 1. The cDNA was 5-fold serially diluted (2 μl, 0.4 μl and 0.08 μl) using DEPC water to test the sensitivity of the primers. For the specificity of the primers, the cDNA from *Rhipicephalus haemaphysaloides* was regarded as a negative control and the H_2O was regarded as a blank control. qPCR was performed with SYBR Premix Ex Taq™II kit (TaKaRa) according to the manufacturer's protocol under the conditions of 95 °C for 30 s and 40 cycles of 95 °C for 5 s and 60 °C for 30 s. The qPCR reaction system (25 μl) consisted of 12.5 μl SYBR Premix Ex Taq™II (2×), 0.5 μl ROX referenceII (50×), 1 μl forward primer (10 μmol), 1 μl reverse primer (10 μmol), 2 μl cDNA, and 8 μl DEPC water. The qPCR was performed on 96 wells plate and each reaction was conducted in triplicate. The relative expression level of different genes was measured by $2^{-\Delta\Delta Ct}$.

3. Results

3.1. RNA sequencing and de novo assembly

In this study, the transcriptome of *H. flava* at larvae and nymph stages was sequenced, assembled and annotated. A total of 64,474,326 and 81,612,022 raw reads were produced in larvae and nymph stages, respectively. After removing the reads containing adapter, poly-N and with low quality, 61,850,967 and 79,579,368 clean reads were obtained, with a data volume of 9.28 G and 11.94 G for *H. flava* in larvae and nymph stages, respectively. Q20, Q30 and GC-content were also calculated based on the clean data. In the larvae stage, the GC content was 49.45%, and the Q20 and Q30 were 96.58 and 92.22%, respectively. In the nymph stage, the percentage of Q20, Q30 and GC was 97.26, 93.3 and 52.69%, respectively (Table 2).

After sequencing, the transcriptome was assembled with Trinity due to its advantage of reconstructing the transcriptome without a reference genome. After removing the transcripts with a length of < 200 bp, we obtained a total of 536,336 and 440,896 transcripts and unigenes with an average length of 509 and 447 bp and an N50 of 641 bp and 482 bp, respectively. The length distributions of the transcripts and unigenes are shown in Fig. 1, and most of them were < 301 bp, accounting for 76.4% and 81.23%, respectively. The maximum length of the unigenes was 26,722 bp and the minimum length was 201 bp.

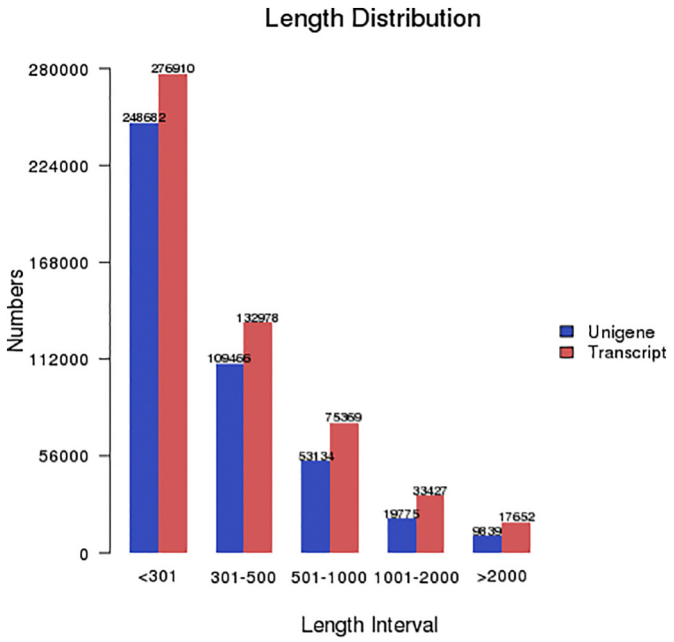

Fig. 1. The length distribution of the unigenes and transcripts. The X axis indicates the length interval. The Y axis indicates the number of unigenes and transcripts in blue and red, respectively. The unigenes are shown in blue and transcripts in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Gene functional annotation and classification

All the transcripts were annotated by seven databases (Fig. 2A). As the genome of *H. flava* had not been sequenced, the transcripts obtained in this study were annotated with other near origin species in the Nr database. The annotation results showed that 35.7% of the transcripts were matched to *Ixodes scapularis*, 5.0% to *Stegodyphus mimosarum*, 3.0% to *Nuttalliella namaqua*, 2.5% to *Metaseiulus occidentalis*, 2.3% to *Strongylocentrotus purpurtus* and 51.4% were unknown genes that had not been annotated yet in any other species (Fig. 2B). Apparently, unknown genes account for more than half of the total genes.

The functions of genes were annotated and classified in terms of biological process, cellular component and molecular function based on the GO database (Fig. 3). In the GO term of biological process, most annotated genes were assigned to cellular process (GO:0009987), metabolic process (GO:0008152) and single-organism process (GO:0044699), which were obviously far more than other processes. For organisms, most of the biological processes are similar, although different organisms may possess their own special biological activity process. For *H. flava*, some processes are more significant than other

Table 2
Summary of the transcriptome of *Haemaphysalis flava Neumann* at larvae and nymph stages for assembling.

Sample	Raw reads	Clean reads	Clean bases	Error (%)	Q20 (%)	Q30 (%)	GC (%)
Larvae	64,474,326	61,850,967	9.28 G	0.01	96.58	92.22	49.45
Nymph	81,612,022	79,579,368	11.94 G	0.01	97.26	93.3	52.69

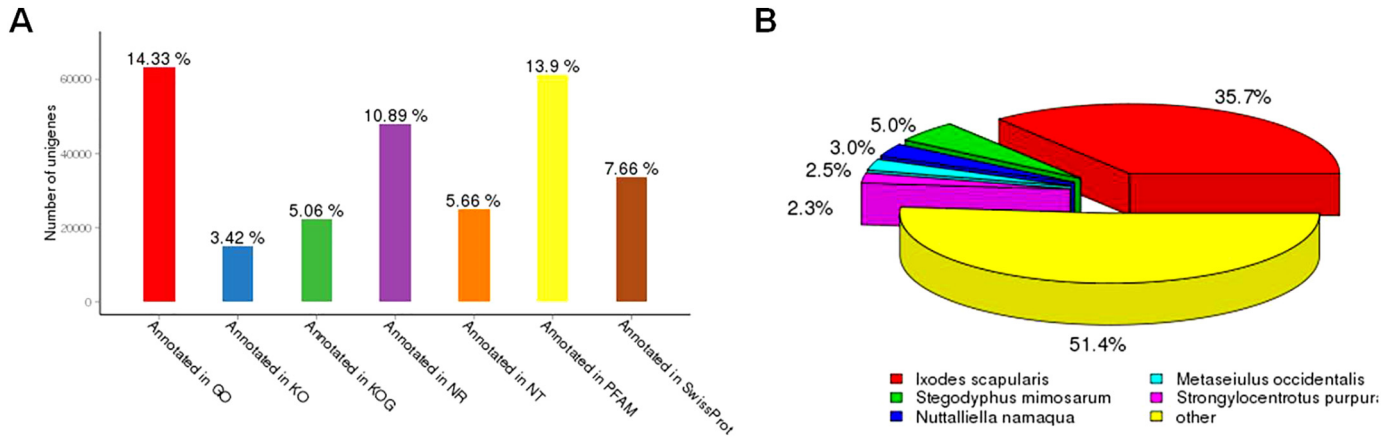

Fig. 2. (A) The transcripts of *Haemaphysalis flava Neumann* were annotated by seven databases. (B) The transcripts obtained in this study were annotated with the near origin species in the Nr database. Among them, the transcripts annotated to *Ixodes scapularis* were the most, accounting for 35.7% of the total.

processes. For example, in cellular process, cellular metabolic process may play an important role by focusing on macromolecular, nitrogen compound and aromatic compound metabolic processes, indicating the involvement of these compounds in some metabolic processes significant and unique to *H. flava*. In single-organism process, cell communication and organic acid metabolic processes are dominant relative to other processes, while in metabolic process, organic substance metabolic process is more important than other processes.

In molecular function, binding and catalytic activity were apparently more than other processes in level 2. In binding process, more genes were assigned to ion binding (anion and cation binding) and heterocylic compound binding, which mainly included the nucleoside phosphate binding, nucleoside binding and nucleic acid binding processes. However, in cellular component, genes that were assigned to the cell and cell junction were slightly more than those of other processes. For cellular component, the annotated genes were mainly involved in cell (GO: 0005623) and cell junction (GO:0030054). In the GO term of molecular function, the genes were mainly annotated to binding (GO: 0005488) and catalytic activity (GO:0003824).

Furthermore, a total of 22,347 unigenes were annotated and classified by KOG database, which can be divided into 25 categories (Fig. 4), with the top three KOG categories including “General function prediction only”, “posttranslational modification, protein turnover, chaperones” and “signal transduction mechanisms”, accounting for 22%, 9.9% and 9.4%, respectively. Moreover, a total of 15,112 unigenes were annotated by the KEGG database and assigned to the five categories of cellular processes, genetic information processing, metabolism, environmental information processing, and organismal systems (Fig. 5). Among them, 1804 unigenes (12%) participate in signal transduction, 1450 genes (9.5%) play a role in translation, and 1212 genes (8%) take part in transport and catabolism, which are the top three pathways the unigenes were involved in. Specifically, the top three enriched pathways were “chloroalkane and chloroalkene degradation”, “fatty acid degradation” and “glycolysis and gluconeogenesis” and all of them were related with the two up-regulated unigenes in the nymph versus larvae: aldehyde dehydrogenase and sorbitol dehydrogenase. Additionally, one hypothetical protein, a down-regulated unigene, was matched to “ubiquitin-mediated proteolysis” pathway.

Single Nucleotide Polymorphisms were screened and a total of 1,776,404 SNPs were identified in the unigenes, with non-coding SNP and coding SNP accounting for 76.40% and 23.60%, respectively. The numbers of transitions and transversions were 1,124,651 and 651,752,

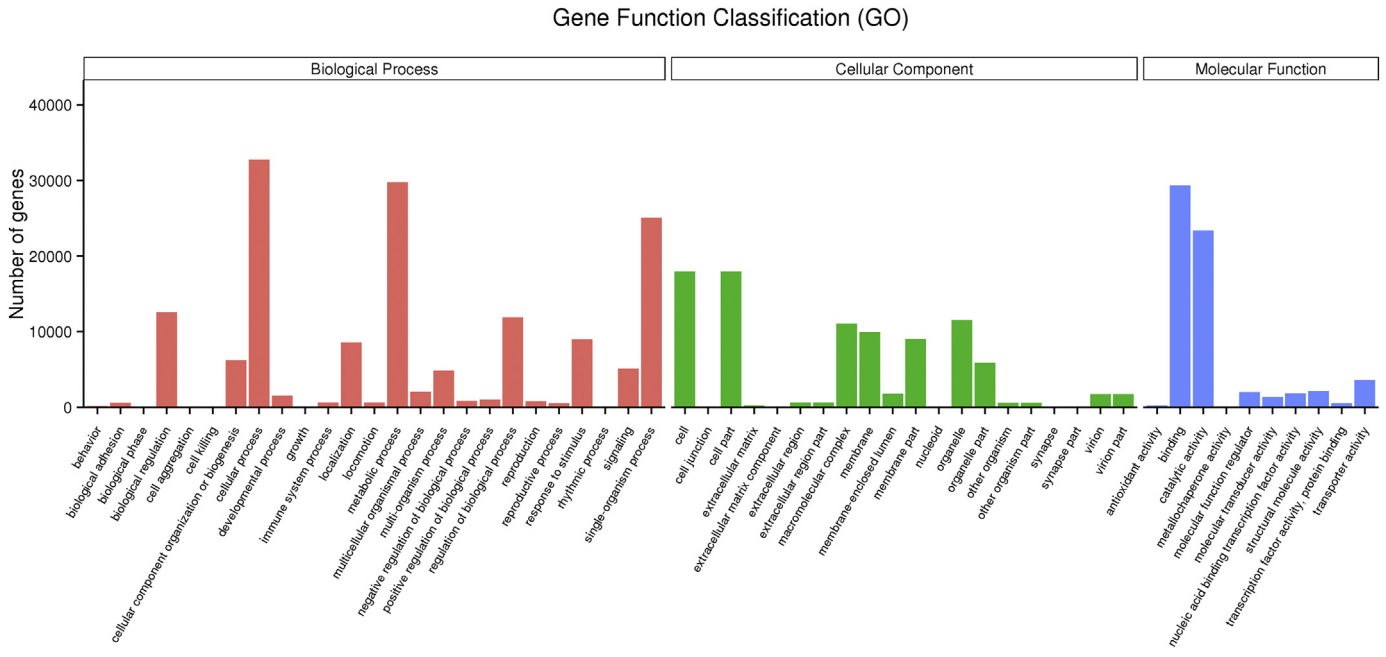

Fig. 3. Annotation and classification of unigenes by GO database in terms of biological process, cellular component and molecular function.

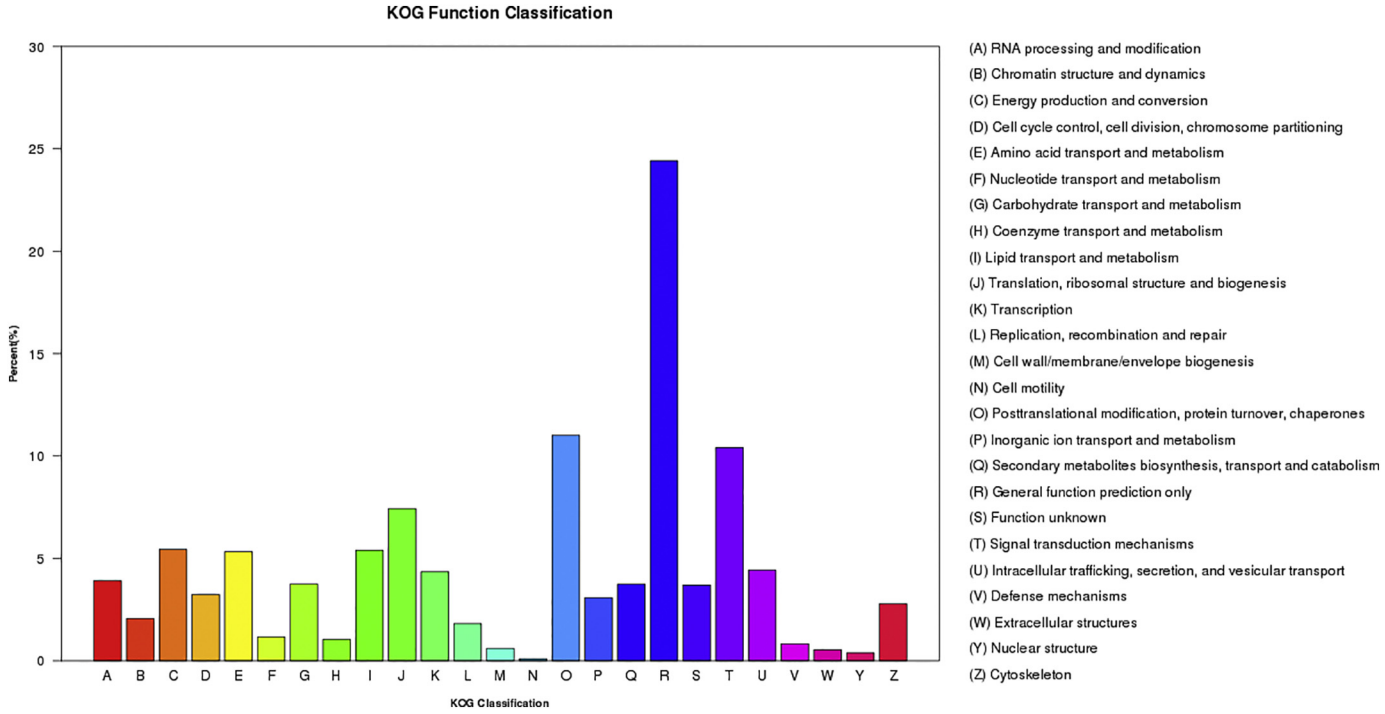

Fig. 4. Annotation and classification of unigenes by KOG database. A total of 22,347 unigenes were divided into 25 categories, with "general function prediction only" as the largest category, accounting for 22% of the total.

respectively, resulting in 278,646 (15.69%) and 140,509 (7.91%) of synonymous and nonsynonymous nucleotide substitutions, respectively.

3.3. Differential expression analysis of H. flava in the larvae and nymph stages and experimental validation by qPCR

Differential expression analysis revealed 80 DEGs of *H. flava* in the larvae and nymph stages, with 56 of them up-regulated and 24 down-regulated (Fig. 6). In the 56 up-regulated genes (nymph versus larvae), 30 (53.5%) were annotated to hypothetical proteins of unknown function and 26 (46.5%) were annotated to cysteine, cysteine-rich secreted protein, cysteine proteinase and secreted proteins in the Nr database. In the 24 down-regulated genes, 17 (70.8%) were annotated to hypothetical proteins of unknown function and only 7 (29.2%) were annotated to lipase, methyltransferase, two types of cystatins and secreted proteins. Among them, hypothetical proteins were in the majority, probably due to the fact that the genome of most of the tick species has not been sequenced, spliced or annotated.

The results of RNA sequencing were confirmed by qPCR with a total of eight DEGs. The relative expression level of the four up-regulated genes was about 3.5-fold higher in the nymph stage than in the larvae stage. The relative expression level of the four down-regulated genes was 5-fold lower in the nymph stage than in the larvae stage. The qPCR result was consistent with that of RNA sequencing.

4. Discussion

In this study, the transcriptome of *H. flava* at larvae and nymph stages was obtained by using the high-throughput RNA sequencing method. A total of 536,336 and 440,896 transcripts and unigenes were obtained, respectively. Due to lack of genome information of *H. flava*, the transcripts had to be annotated with other near origin species. Annotation results showed that only 48.6% of the transcripts match with the reported near species, and 51.4% are unknown genes, suggesting the necessity for further research of ticks. Moreover, more than half of the genes were annotated to hypothetic proteins of unknown function. Ticks were of veterinary significance, which can transmit several pathogens and lead to enormous economic loss in livestock industry. In recent years, increasing attention has been paid to ticks such as *Ixodes persulcatus*, *Rhipicephalus sanguineus*, *Rhipicephalus microplus* and so on (Anatriello et al., 2010; Ali et al., 2014). To date, only the genome, transcriptome and proteome of *Ixodes scapularis* and *Boophilus microplus* have been sequenced and annotated completely (Guerrero et al., 2006; Miller et al., 2018). This phenomenon can be attributed to the following reasons: (1) RNA collection was difficult due to the small size of ticks; (2) blood-feeding may lead to host pollution in sequencing; (3) the existence of so many repeated sequences in the genome increases the difficulty of splicing and annotation (Gibson et al., 2013; Vayssier-Taussat et al., 2013).

In this article, we obtained for the first time the transcriptome of *H. flava* at larvae and nymph stages, and 80 DEGs were identified, including 56 up-regulated genes and 24 down-regulated genes in the nymph versus larvae. The functions of these DEGs were analyzed by different databases and the majority (47 DEGs) were annotated to hypothetical proteins of unknown function. Among the 24 down-regulated genes with up-regulated expression in the larvae stage, lipase was the highest expressed based on the $\log_2$ fold change, revealing that lipase plays an important role in the larvae stage. According to the GO annotation, the genes that encode the lipase match to several processes, including lipid catabolic process, cholesterol, secondary alcohol, sterol, alcohol, steroid metabolic process and lipid catalytic activity. While all the GO annotations were matched to lipid-related activity, previous studies have also shown that lipase had the digestion activity in the larvae stage of the American lobster *Homarus americanus* Milne Edwards (Brahimi-Horn et al., 1989; Biesiot and Capuzzo, 1990). To date, no studies have focused on the digestion system at the larvae stage, so whether lipase on the larvae of *H. flava* has the digestion activity needs to be further studied. Ticks are exclusively (obligate) hematophagous ectoparasites (Willadsen and Riding, 1980). Therefore, blood from the host was the most essential source for the survival, growth and reproduction of ticks. The blood digestion process is not only vital but also very complicated. The system consists of a series of proteases, including serine, cysteine, and aspartic proteases, which

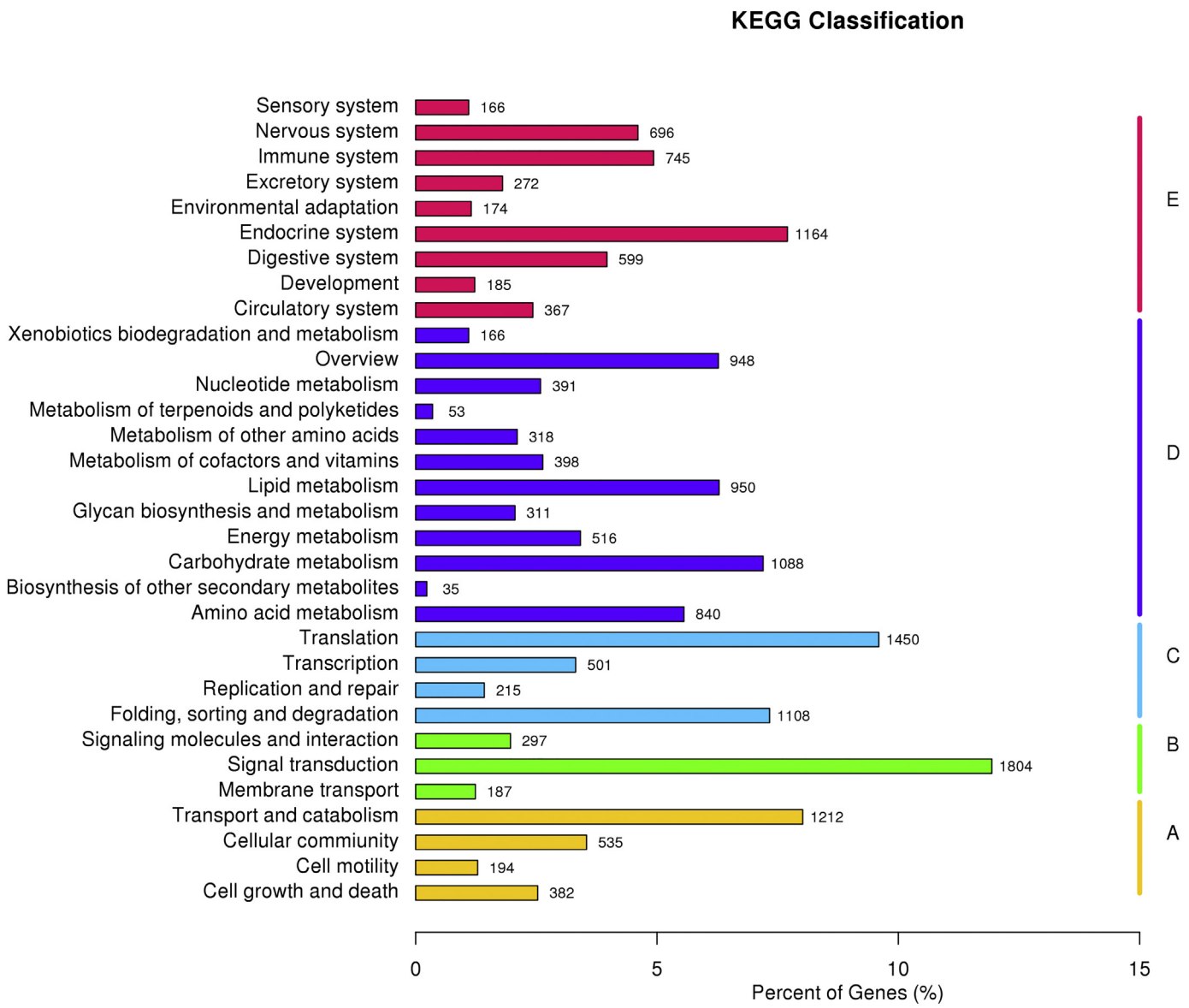

Fig. 5. Annotation and classification of unigenes by KEGG database. The 15,112 unigenes were annotated and assigned into five categories, including cellular process, genetic information processing, metabolism, environmental information processing, and organismal systems.

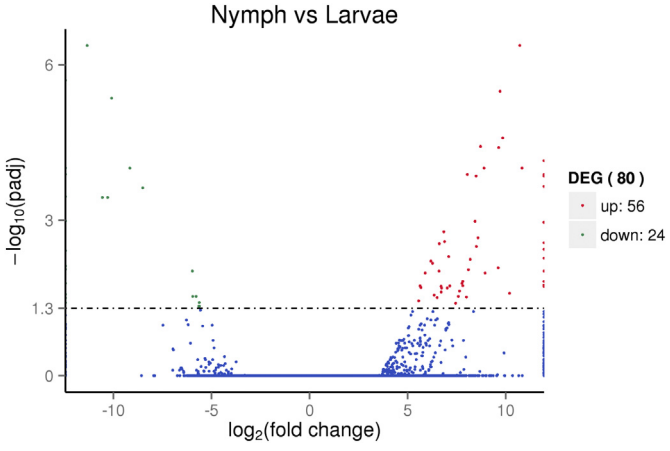

Fig. 6. 80 differentially expressed genes (DEGs) identified between nymph and larvae stages, with 56 up-regulated and 24 down-regulated in nymph versus larvae.

have been reported in previous studies (Sojka et al., 2008; Horn et al., 2009; Shaw et al., 2012; Zhang et al., 2015). For example, the network of cysteine and aspartic peptidases play a concerted role in the blood digestion in the gut of *Ixodes ricinus* (Sojka et al., 2008). In this study, one gene that was annotated to midgut cysteine proteinase 4 (cathepsin L), a member of peptidase C1-like family, was up-regulated in nymph versus larvae. This was confirmed in our previous study by Sun et al., who demonstrated that the transcriptional level of cathepsin L was significantly higher in nymph and adult than in egg and larvae, even though it could be transcribed in all the development stages, including egg, larvae, nymph, and adult (Sun et al., 2019). Cathepsin L was also identified to be localized in the gut through the anti-cathepsin L serum, which also revealed that cathepsin L plays a key role in blood digestion of ticks in the gut. Several studies have also been performed on cathepsin L-like (Clara et al., 2011) in *Haemaphysalis longicornis* (HlCPL-A), *B. microplus* (BmCL1), *Ixodes ricinus* (IrCL1) and *Rhipicephalus (Boophilus) annulatus* (RaCL1) (Lima et al., 2006; Sojka et al., 2008; Zhou et al., 2009; Franta et al., 2011; Zhang et al., 2015; Saidi et al., 2016). The mechanism for blood digestion of ticks is complex. In the present study, the high expression of cysteine proteinase 4 in the

nymph, but low expression in the larvae stage indicated that the ticks may utilize different types of proteinase for proteolysis at different stages, due to the limitation of growth and regulation. This suggests that midgut cysteine proteinase 4 may play a key role in the process of blood digestion in nymph and some substitutions may exist in larvae, even though the process of proteolysis in the midgut is a network and needs the incorporation of a battery of proteases.

Proteinase inhibitors are also present and have a series of physiological functions, especially in the control of proteolytic enzymes and the interaction between host and parasite. However, some proteinase inhibitors only produce effects at a specific stage rather than throughout the whole life cycle. For example, the concentration of inhibitor of trypsin and chymotrypsin of *B. microplus* varied in the different stages of the life cycle, which was very high in egg and unfed stages and decreased rapidly after contact with the host (Willadsen and McKenna, 1983; Lima et al., 2006). Similarly, in the present study, two different cystatins were up-regulated in the larvae versus nymph, which has been rarely reported in tick-related studies. Cystatin superfamily, the strong inhibitor of cysteine proteinase, consists of stefins (intracellular), cystatins and the kininogens (extracellular), which have been widely described in many organisms, including animals, plants, insects and protozoa (Willadsen and Riding, 1980; Zhou et al., 2006; Zhou et al., 2009; Salat et al., 2010; Kolb et al., 2015; Liang et al., 2015; Valdes-Rodriguez et al., 2015; Wang et al., 2015; Christova et al., 2018; Melo et al., 2018). They can form tight and reversible contact with the cysteine proteinase, but their functions remained to be further elucidated. In *B. microplus*, cystatin (type 1) was characterized in fat body (Lima et al., 2006; Parizi et al., 2013). In ovary, cystatin was also detected and speculated to have roles in modulating endogenous proteolysis. In salivary gland of *B. microplus* and *Amblyomma americanum*, cystatin-like protein was detected and was supposed to regulate the interactions between host and parasite (Lima et al., 2006; Schwarz et al., 2012; Parizi et al., 2013). Overall, the physiological functions of cystatin are complicated and diverse in different developmental stages of ticks (Schwarz et al., 2012), and the two different up-regulated cystatins of *H. flava* may have specific functions in the larvae stage. Specifically, the high expression of cysteine proteinase 4 in nymph, and the high expression of cystatins (inhibitors of cysteine proteinase) in larvae both confirmed that cysteine proteinase plays a key role in nymph but not in larvae. However, how the cysteine proteinase works in the nymph and what proteinase plays a key role in larvae remain to be further studied.

To date, the knowledge for the *H. flava* remains limited and rather fragmented. Most of the studies were focused on the adult stage, including the female and male, fully engorged and partially engorged. The transcriptome studies of midgut and salivary gland of *H. flava* identified the genes that were involved in or associated with blood digestion, feeding and defending from pathogens. The proteomics research on the midgut detected some significant proteins implicated in preventing blood meal coagulation, facilitating intracellular digestion (Xu et al., 2015, 2016; Liu et al., 2018a,b). However, the knowledge is still limited about the whole transcriptome of *H. flava* at the larvae and nymph stages. This study performed a comparative analysis of the transcriptome at the two different stages and obtained 80 DEGs between the larvae and nymph stages. It may provide new insight into the biology of *H. flava* and contribute to the development of effective therapies against *H. flava* or the diseases transmitted by *H. flava*.

5. Conclusion

This is the first report about the transcriptome sequences of *H. flava* at larvae and nymph stages. The transcripts obtained were annotated and classified based on seven databases, with more attention being given to GO, KOG and KEGG analyses. A total of 80 DEGs were identified between larvae and nymph stages, including 56 up-regulated genes and 24 down-regulated genes. This study provides the first insights into *H. flava* at the gene expression level in different developmental stages and has provided a theoretical basis for the development of new strategies for the control of *H. flava*.

Ethics approval and consent to participate

The experimental animals were housed and treated in accordance with the stipulated rules for the regulation of the administration of affair concerning experimental animals of the P. R. China. All experiments were performed under the approval of Laboratory Animals Research Centre of Hubei Province and the Ethics Committee of Huazhong Agricultural University (Permit number: HZAUCA-2016-007).

Consent for publication

Not applicable.

Availability of data and materials

All data obtained in this study had been deposited to the National Center for Biotechnology Information (NCBI) with the accession number SRR8426798.

Declaration of Competing Interests

The authors declare that they have no competing interests.

Funding

This study was supported by the National Key Basic Research Program (973 Program) of China (Grant No. 2015CB150302), the National Key Research and Development Program of China (Grant No. 2017YFD0500402 and 2017YFD0501201), the National Natural Science Foundation of China (Grant No. 31772729) and the Natural Science Foundation of Hubei Province (Grant No. 2017CFA020).

Authors' contributions

Performed the experiments: JG, YS and PH. Participated in the data analysis: ML, LH, JG, YS and PH. Helped with the diagnostic assays: PH and XL. Edited the manuscript: JG, YS, LH and JZ. All authors have read and approved the final manuscript.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.104008.

References

Ali, A., Tirloni, L., Isezaki, M., Seixas, A., Konnai, S., Ohashi, K., et al., 2014. Reprolysin metalloproteases from Ixodes persulcatus, *Rhipicephalus sanguineus* and *Rhipicephalus microplus* ticks. Exp. Appl. Acarol. 63 (4), 559–578. https://doi.org/10.1007/s10493-014-9796-9.

Anatriello, E., Ribeiro, J.M., de Miranda-Santos, I.K., Brandao, L.G., Anderson, J.M., Valenzuela, J.G., et al., 2010. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. BMC Genomics 11, 450. https://doi.org/10.1186/1471-2164-11-450.

Biesiot, P.M., Capuzzo, J.M., 1990. Changes in digestive enzyme-activities during early development of the American lobster Homarus-Americanus Milne-Edwards. J. Exp. Mar. Biol. Ecol. 136 (2), 107–122. https://doi.org/10.1016/0022-0981(90)90190-N.

Brahimi-Horn, M.C., Guglielmino, M.L., Sparrow, L.G., Logan, R.I., Moran, P.J., 1989. Lipolytic enzymes of the digestive organs of the crown-of-thorns starfish (*Acanthaster planci*): comparison of the stomach and pyloric caeca. Comp. Biochem. Physiol. B 92

(4), 637–643.
Charrier, N.P., Couton, M., Voordouw, M.J., Rais, O., Durand-Hermouet, A., Hervet, C., et al., 2018. Whole body transcriptomes and new insights into the biology of the tick Ixodes ricinus. Parasit. Vectors 11 (1), 364. https://doi.org/10.1186/s13071-018-2932-3.
Cheng, W.Y., Zhao, G.H., Jia, Y.Q., Bian, Q.Q., Du, S.Z., Fang, Y.Q., et al., 2013. Characterization of Haemaphysalis flava (Acari: Ixodidae) from Qingling subspecies of giant panda (*Ailuropoda melanoleuca qinlingensis*) in Qinling Mountains (Central China) by morphology and molecular markers. PLoS One 8 (7), e69793. https://doi.org/10.1371/journal.pone.0069793.
Christova, P.K., Christov, N.K., Mladenov, P.V., Imai, R., 2018. The wheat multidomain cystatin TaMDC1 displays antifungal, antibacterial, and insecticidal activities in planta. Plant Cell Rep. 37 (6), 923–932. https://doi.org/10.1007/s00299-018-2279-4.
Clara, R.O., Soares, T.S., Torquato, R.J., Lima, C.A., Watanabe, R.O., Barros, N.M., et al., 2011. Boophilus microplus cathepsin L-like (BmCL1) cysteine protease: specificity study using a peptide phage display library. Vet. Parasitol. 181 (2–4), 291–300. https://doi.org/10.1016/j.vetpar.2011.04.003.
Duan, D., Cheng, T., 2017. Determination of the microbial community features of Haemaphysalis flava in different developmental stages by high-throughput sequencing. J. Basic Microbiol. 57 (4), 302–308. https://doi.org/10.1002/jobm.201600557.
Egekwu, N., Sonenshine, D.E., Bissinger, B.W., Roe, R.M., 2014. Transcriptome of the female synganglion of the black-legged tick Ixodes scapularis (Acari: Ixodidae) with comparison between Illumina and 454 systems. PLoS One 9 (7), e102667. https://doi.org/10.1371/journal.pone.0102667.
Ejiri, H., Lim, C.K., Isawa, H., Yamaguchi, Y., Fujita, R., Takayama-Ito, M., et al., 2018. Isolation and characterization of Kabuto Mountain virus, a new tick-borne phlebovirus from Haemaphysalis flava ticks in Japan. Virus Res. 244, 252–261. https://doi.org/10.1016/j.virusres.2017.11.030.
Franta, Z., Sojka, D., Frantova, H., Dvorak, J., Horn, M., Srba, J., et al., 2011. IrCL1 - the haemoglobinolytic cathepsin L of the hard tick, Ixodes ricinus. Int. J. Parasitol. 41 (12), 1253–1262. https://doi.org/10.1016/j.ijpara.2011.06.006.
Fujimoto, K., 1987. Ecological studies on ixodid ticks. 2. Comparison of the seasonal occurrence of three ixodid ticks, Haemaphysalis flava, Ixodes ovatus and I. nipponensis in the south-western part of Saitama prefecture. Jpn. J. Sanit. Zool. 38, 7–12.
Furuno, K., Lee, K., Itoh, Y., Suzuki, K., Yonemitsu, K., Kuwata, R., et al., 2017. Epidemiological study of relapsing fever borreliae detected in Haemaphysalis ticks and wild animals in the western part of Japan. PLoS One 12 (3), e0174727. https://doi.org/10.1371/journal.pone.0174727.
Gibson, A.K., Smith, Z., Fuqua, C., Clay, K., Colbourne, J.K., 2013. Why so many unknown genes? Partitioning orphans from a representative transcriptome of the lone star tick Amblyomma americanum. BMC Genomics 14, 135. https://doi.org/10.1186/1471-2164-14-135.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652. https://doi.org/10.1038/nbt.1883.
Guerrero, F.D., Nene, V.M., George, J.E., Barker, S.C., Willadsen, P., 2006. Sequencing a new target genome: the Boophilus microplus (Acari: Ixodidae) genome project. J. Med. Entomol. 43 (1), 9–16.
Horak, I.G., Camicas, J.L., Keirans, J.E., 2002. The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida): a world list of valid tick names. Exp. Appl. Acarol. 28 (1–4), 27–54.
Horn, M., Nussbaumerova, M., Sanda, M., Kovarova, Z., Srba, J., Franta, Z., et al., 2009. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem. Biol. 16 (10), 1053–1063. https://doi.org/10.1016/j.chembiol.2009.09.009.
Ishigaki, Y., Nakamura, Y., Oikawa, Y., Yano, Y., Kuwabata, S., Nakagawa, H., et al., 2012. Observation of live ticks (Haemaphysalis flava) by scanning electron microscopy under high vacuum pressure. PLoS One 7 (3), e32676. https://doi.org/10.1371/journal.pone.0032676.
Jung, M., Kho, J.W., Lee, W.G., Roh, J.Y., Lee, D.H., 2019. Seasonal occurrence of Haemaphysalis longicornis (Acari: Ixodidae) and Haemaphysalis flava, vectors of severe fever with thrombocytopenia syndrome (SFTS) in South Korea. J. Med. Entomol. https://doi.org/10.1093/jme/tjz033.
Kakuda, H., Shiraishi, S., Uchida, T.A., 1990. Seasonal fluctuations of populations and effects of temperatures on development and growth in the tick, Haemaphysalis flava. J. Fac. Agric. Kyushu Univ. 35 (1/2), 17–26.
Kettle, D.S., 1995. Medical and Veterinary Entomology. CAB International, Wallingford, Oxon, UK.
Kolb, P., Vorreiter, J., Habicht, J., Bentrop, D., Wallich, R., Nassal, M., 2015. Soluble cysteine-rich tick saliva proteins Salp15 and Iric-1 from E. coli. FEBS Open Bio. 5, 42–55. https://doi.org/10.1016/j.fob.2014.12.002.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 12, 323. https://doi.org/10.1186/1471-2105-12-323.
Liang, J., Wang, Y., Ding, G., Li, W., Yang, G., He, N., 2015. Biotic stress-induced expression of mulberry cystatins and identification of cystatin exhibiting stability to silkworm gut proteinases. Planta 242 (5), 1139–1151. https://doi.org/10.1007/s00425-015-2345-x.
Lima, C.A., Sasaki, S.D., Tanaka, A.S., 2006. Bmcystatin, a cysteine proteinase inhibitor characterized from the tick Boophilus microplus. Biochem. Biophys. Res. Commun. 347 (1), 44–50. https://doi.org/10.1016/j.bbrc.2006.06.018.
Liu, L., Cheng, T.Y., Yang, Y., 2017. Cloning and expression pattern of a heat shock cognate protein 70 gene in ticks (Haemaphysalis flava). Parasitol. Res. 116 (6), 1695–1703. https://doi.org/10.1007/s00436-017-5444-8.
Liu, L., Cheng, T.Y., He, X.M., 2018a. Proteomic profiling of the midgut contents of Haemaphysalis flava. Ticks Tick Borne Dis. 9 (3), 490–495. https://doi.org/10.1016/j.ttbdis.2018.01.008.
Liu, L., Liu, Y.S., Liu, G.H., Cheng, T.Y., 2018b. Proteomics analysis of faecal proteins in the tick Haemaphysalis flava. Parasit. Vectors 11 (1), 89. https://doi.org/10.1186/s13071-018-2673-3.
Melo, P.M.S., El Chamy Maluf, S., Azevedo, M.F., Paschoalin, T., Budu, A., Bagnaresi, P., et al., 2018. Inhibition of plasmodium falciparum cysteine proteases by the sugarcane cystatin CaneCPI-4. Parasitol. Int. 67 (2), 233–236. https://doi.org/10.1016/j.parint.2017.12.005.
Miller, J.R., Koren, S., Dilley, K.A., Harkins, D.M., Stockwell, T.B., Shabman, R.S., et al., 2018. A draft genome sequence for the Ixodes scapularis cell line, ISE6. F1000Res 7, 297. https://doi.org/10.12688/f1000research.13635.1.
Ozawa, A., Yamaguchi, N., Hayakawa, K., Matsuo, I., Niizuma, K., Ohkido, M., 1982. A case of tick bite (Haemaphysalis flava)–consideration of tularemia infection through tick bite. Nihon Hifuka Gakkai zasshi. Japan. J. Dermatol. 92 (13), 1415.
Parizi, L.F., Githaka, N.W., Acevedo, C., Benavides, U., Seixas, A., Logullo, C., et al., 2013. Sequence characterization and immunogenicity of cystatins from the cattle tick Rhipicephalus (Boophilus) microplus. Ticks Tick Borne Dis. 4 (6), 492–499. https://doi.org/10.1016/j.ttbdis.2013.06.005.
Pfaffle, M., Littwin, N., Muders, S.V., Petney, T.N., 2013. The ecology of tick-borne diseases. Int. J. Parasitol. 43 (12 − 13), 1059–1077. https://doi.org/10.1016/j.ijpara.2013.06.009.
Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F., Sugimoto, C., 2014. Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. PLoS One 9 (8), e103961. https://doi.org/10.1371/journal.pone.0103961.
Saidi, S., Nabian, S., Ebrahimzade, E., Najafi, A., Moghaddam, M.M., Sazmand, A., et al., 2016. Identification and characterization of a cathepsin L-like cysteine protease from Rhipicephalus (Boophilus) annulatus. Exp. Appl. Acarol. 68 (2), 251–265. https://doi.org/10.1007/s10493-015-9993-1.
Salat, J., Paesen, G.C., Rezacova, P., Kotsyfakis, M., Kovarova, Z., Sanda, M., et al., 2010. Crystal structure and functional characterization of an immunomodulatory salivary cystatin from the soft tick Ornithodoros moubata. Biochem. J. 429 (1), 103–112. https://doi.org/10.1042/BJ20100280.
Schwarz, A., Valdes, J.J., Kotsyfakis, M., 2012. The role of cystatins in tick physiology and blood feeding. Ticks Tick Borne Dis. 3 (3), 117–127. https://doi.org/10.1016/j.ttbdis.2012.03.004.
Shaw, D.K., Hyde, J.A., Skare, J.T., 2012. The BB0646 protein demonstrates lipase and haemolytic activity associated with Borrelia burgdorferi, the aetiological agent of Lyme disease. Mol. Microbiol. 83 (2), 319–334. https://doi.org/10.1111/j.1365-2958.2011.07932.x.
Sojka, D., Franta, Z., Horn, M., Hajdusek, O., Caffrey, C.R., Mares, M., et al., 2008. Profiling of proteolytic enzymes in the gut of the tick Ixodes ricinus reveals an evolutionarily conserved network of aspartic and cysteine peptidases. Parasit. Vectors 1 (1), 7. https://doi.org/10.1186/1756-3305-1-7.
Sun, Y., He, L., Yu, L., Guo, J., Nie, Z., Liu, Q., et al., 2019. Cathepsin L-a novel cysteine protease from Haemaphysalis flava Neumann, 1897. Parasitol. Res. 118 (5), 1581–1592. https://doi.org/10.1007/s00436-019-06271-4.
Tissot Dupont, H., Raoult, D., 1993. Tick-borne diseases. Rev. Med. Interne 14 (5), 300–306.
Valdes-Rodriguez, S., Galvan-Ramirez, J.P., Guerrero-Rangel, A., Cedro-Tanda, A., 2015. Multifunctional amaranth cystatin inhibits endogenous and digestive insect cysteine endopeptidases: a potential tool to prevent proteolysis and for the control of insect pests. Biotechnol. Appl. Biochem. 62 (5), 634–641. https://doi.org/10.1002/bab.1313.
Vayssier-Taussat, M., Moutailler, S., Michelet, L., Devillers, E., Bonnet, S., Cheval, J., et al., 2013. Next generation sequencing uncovers unexpected bacterial pathogens in ticks in western Europe. PLoS One 8 (11), e81439. https://doi.org/10.1371/journal.pone.0081439.
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimarães, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., et al., 2003. Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni. Nat. Genet. 35 (2), 148–157. https://doi.org/10.1038/ng1237.
Wang, Y., Yu, X., Cao, J., Zhou, Y., Gong, H., Zhang, H., et al., 2015. Characterization of a secreted cystatin from the tick Rhipicephalus haemaphysaloides. Exp. Appl. Acarol. 67 (2), 289–298. https://doi.org/10.1007/s10493-015-9946-8.
Willadsen, P., McKenna, R.V., 1983. Trypsin-chymotrypsin inhibitors from the tick, Boophilus microplus. Aust. J. Exp. Biol. Med. Sci. 61, 231–238 Pt 2.
Willadsen, P., Riding, G.A., 1980. On the biological role of a proteolytic-enzyme inhibitor from the ectoparasitic tick Boophilus microplus. Biochem. J. 189 (2), 295–303.
Xu, X.L., Cheng, T.Y., Yang, H., Yan, F., Yang, Y., 2015. De novo sequencing, assembly and analysis of salivary gland transcriptome of Haemaphysalis flava and identification of sialoprotein genes. Infect. Genet. Evol. 32, 135–142. https://doi.org/10.1016/j.meegid.2015.03.010.
Xu, X.L., Cheng, T.Y., Yang, H., Liao, Z.H., 2016. De novo assembly and analysis of midgut transcriptome of Haemaphysalis flava and identification of genes involved in blood digestion, feeding and defending from pathogens. Infect. Genet. Evol. 38, 62–72. https://doi.org/10.1016/j.meegid.2015.12.005.
Yin, F., Sun, P., Wang, J., Gao, Q., 2016. Transcriptome analysis of dormant tomonts of the marine fish ectoparasitic ciliate cryptocaryon irritans under low temperature. Parasit. Vectors 9 (1), 280. https://doi.org/10.1186/s13071-016-1550-1.
Zhang, H., Yang, S., Gong, H., Cao, J., Zhou, Y., Zhou, J., 2015. Functional analysis of a novel cysteine-rich antimicrobial peptide from the salivary glands of the tick Rhipicephalus haemaphysaloides. Parasitol. Res. 114 (10), 3855–3863. https://doi.org/10.1007/s00436-015-4615-8.
Zhou, J., Ueda, M., Umemiya, R., Battsetseg, B., Boldbaatar, D., Xuan, X., et al., 2006. A secreted cystatin from the tick Haemaphysalis longicornis and its distinct expression patterns in relation to innate immunity. Insect Biochem. Mol. Biol. 36 (7), 527–535. https://doi.org/10.1016/j.ibmb.2006.03.003.
Zhou, J., Liao, M., Ueda, M., Gong, H., Xuan, X., Fujisaki, K., 2009. Characterization of an intracellular cystatin homolog from the tick Haemaphysalis longicornis. Vet. Parasitol. 160 (1–2), 180–183. https://doi.org/10.1016/j.vetpar.2008.10.086.