Transcriptome Analysis and Molecular Characterization of Olfactory Binding Protein genes in Parasitoid wasp Anagrus nilaparvatae(Hymenoptera: Mymaridae)

Yin Ma[1], Tingfa Huang[2], Binjie Tang[2], Bingyang Wang[2], Jianbai Liu[3], and Qiang Zhou[1]

[1]Sun Yat-sen University State Key Laboratory of Biocontrol
[2]School of Life Sciences, Sun Yat-sen University
[3]Guangdong Academy of Sciences

September 24, 2021

Abstract

Anagrus nilaparvatae is an important egg parasitoid wasps of rice pests rice planthopper. Based on the powerful olfactory system of sensing chemical information in nature, A. nilaparvatae shows complicated life activities and behaviors, such as feeding, mating and hosting. In this study, we constructed a full-length transcriptome library and further to identify the characteristics of olfactory binding proteins, the first participant in the olfactory system. Through full-length transcriptome sequencing, splicing, assembly, and data correction by Illumina, we obtained 163.59Mb of transcriptome data and 501,179 items of annotation information, and performed GO functional classification of unigenes of the transcriptome. We analyzed the sequence characteristics of olfactory binding protein genes, and 8 genes (AnilOBP2, AnilOBP9 AnilOBP23, AnilOBP56, AnilOBP83, AnilCSP5, AnilCSP6 and AnilNPC2) were identified. After sequence alignment and conserved domain prediction, the 8 proteins were consistent with the typical characteristics of OBPs, CSPs and NPC2s in insects. The phylogenetic tree analysis showed that the 8 genes share low homology relationship with other species in Hymenopteran. Finally, RT-qPCR was used to analyze the expression responses of the 8 genes in different genders and stimulated by volatiles. The relative expression levels of AnilOBP9, AnilOBP26, AnilOBP83, AnilCSP5 and AnilNPC2 in males were significantly higher than those in female, while the relative expression levels of AnilCSP6 were opposite. The expression levels of AnilOBP9 and AnilCSP6 were significantly altered by the stimulation of β-caryophylene, suggesting the two genes may be related to host searching. In this study, the transcriptome data of parasitoid wasps A. nilaparvatae could provide a reference for the molecular biology research of the parasitoids, and the identification and analysis of olfactory binding proteins not only help us further clarify the physiological characteristics and parasitic mechanism of the parasitoids, but also promote the utilization of natural enemy resources.

Transcriptome Analysis and Molecular Characterization of Olfactory Binding Protein Genes inParasitoid wasp *Anagrus nilaparvatae*(Hymenoptera: Mymaridae)

Ying Ma[1, 2], Tingfa Huang [3], Bingjie Tang[3], Bingyang Wang[3], Jianbai Liu[4*] and Qiang Zhou [1, 3*]

1. State Key Laboratory for Biocontrol, Sun Yat-Sen University, Guangzhou 510275, China
2. School of Agriculture, Sun Yat-sen University, Guangzhou 510275, China.
3. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China.
4. Institute of Nanfan Seed Industry, Guangdong Academy of Sciences, Guangzhou 510316, China.

Ying Ma and Tingfa Huang have contributed equally to this work and as co-first authors

*Correspondence:*lsszhou@mail.sysu.edu.cn* ; liujb5@mail2.sysu.edu.cn

Abstract

Anagrus nilaparvatae is an important egg parasitoid wasp of rice pest rice planthopper. Based on the powerful olfactory system of sensing chemical information in nature,*A. nilaparvatae* shows complicated life activities and behaviors, such as feeding, mating and hosting. In this study, we constructed a full-length transcriptome library and further to identify the characteristics of olfactory binding proteins, the first participant in the olfactory system. Through full-length transcriptome sequencing, splicing, assembly, and data correction by Illumina, we obtained 163.59Mb of transcriptome data and 501,179 items of annotation information, and performed GO functional classification of unigenes of the transcriptome. We analyzed the sequence characteristics of olfactory binding protein genes, and 8 genes (*AnilOBP2* , *AnilOBP9AnilOBP23* , *AnilOBP56* , *AnilOBP83* , *AnilCSP5* ,*AnilCSP6* and *AnilNPC2*) were identified. After sequence alignment and conserved domain prediction, the 8 proteins were consistent with the typical characteristics of OBPs, CSPs and NPC2s in insects. The phylogenetic tree analysis showed that the 8 genes share low homology relationship with other species in Hymenopteran. Finally, RT-qPCR was used to analyze the expression responses of the 8 genes in different genders and stimulated by volatiles. The relative expression levels of *AnilOBP9* , *AnilOBP26* , *AnilOBP83* ,*AnilCSP5* and *AnilNPC2* in males were significantly higher than those in female, while the relative expression levels of*AnilCSP6* were opposite. The expression levels of *AnilOBP9*and *AnilCSP6* were significantly altered by the stimulation of β-caryophylene, suggesting the two genes may be related to host searching. In this study, the transcriptome data of parasitoid wasps*A. nilaparvatae* could provide a reference for the molecular biology research of the parasitoids, and the identification and analysis of olfactory binding proteins not only help us further clarify the physiological characteristics and parasitic mechanism of the parasitoids, but also promote the utilization of natural enemy resources.

Keywords *Anagrus nilaparvatae* , Olfactory binding protein, Transcriptome, Molecular Characterization

Introduction

The external environment of insects is complex and changeable, and the volatiles in the environment can transmit important information related to their survival and reproduction, such as food, location of the opposite sex, alarm, spawning location and habitat environment, etc. (Stocker, 1994; Grosse-Wilde et al., 2011). Insects rely on a powerful sensory system to help them use chemical information in these environments quickly and efficiently (Leal, 2013; Turlings Erb, 2018).

The olfactory system of insects includes two parts: the central olfactory system and the peripheral olfactory system (Ong Stopfer, 2012). When odor or pheromone molecules diffuse into specialized receptors on the insect's body surface, olfactory binding proteins transport these molecules to the olfactory receptors on the membrane of the peripheral nerve dendrites (Benton et al., 2007); After the interaction of the signal molecules with the olfactory receptors, the chemical signals will be converted into electrical signals that stimulate the dendritic nerves, which will finally be transmitted to the central nerve to control the insect's behavior and physiological response (Zhou, 2010; Leal, 2013). Olfactory binding proteins, the first participant in the olfactory system, are main expression in the sense of smell peripheral lymph system, responsible for the identification and transmission of smell and pheromone molecules reach to olfactory receptors.

In recent years, with the application of a variety of new biological technologies, a large number of insect olfactory binding proteins have been identified and their various functions have been revealed. Olfactory binding proteins mainly include three major families, odorant-binding proteins (OBPs), chemosensory proteins (CSPs), and Niemann-pick type C2 proteins (NPC2s). The OBPs family is the first olfaction binding protein to be identified in insects (Vogt Riddiford, 1981). Its functions include odor recognition, assistance in transporting odor molecules, degradation and removal of odor molecules, etc. (Kaissling, 1986; Pelosi et al., 1995; Krieger et al., 1996; Vogt et al., 1999). OBPs are a kind of small molecules and water-soluble spherical proteins. Almost all insect OBPs have highly conserved cysteines, which connected disulfide bonds are main factors affecting and maintaining its protein structure. According to the number of conserved cysteines, OBPs were classified into five classes: "Classical OBPs"、"Dimer OBPs"、"Minus- C OBPs"、"Plus-C OBPs" and "Atypical OBPs"(Cui et al., 2017; Qu et al., 2021; Fan et al., 2011). CSPs are thought to be the

carrier of odor molecules and chemicals, and can bind to chemical messages (Peng et al., 2017). CSP family have been suggested to be the ancestor of the OBP genes (Vieira Rozas, 2011) although these two families do not show significant sequence homology within Entognatha. CSPs and OBPs are similar in many ways, they both are expressed in large quantities in antennae and participate in the process of olfactory recognition, and both are small compact polypeptides, mainly made of α-helical domains which define a hydrophobic binding cavity (Sandler et al., 2000; Campanacci et al., 2003; Tegoni et al., 2004). CSPs are smaller than OBPs (approximately 12 kDa), contain four conserved cysteines forming two uncrossed disulfide Bridges (Cys1-Cys2, Cys3-Cys4), and contain 5-6 α helices (Sanchez-Gracia et al., 2009). CSPs are evolutionally conserved compared to OBPs, and the high conservatism may be one of the reasons why there are fewer CSPs than OBPs (Pelosi et al., 2018; Wanner et al., 2004). CSPs in distantly evolved insects also tend to have 40-50% similar amino acid residues, compared with 10-15% for OBPs (Pelosi et al., 2006). Therefore, it is assumed that CSPs are less specific for selective binding of compounds and have a wider binding range and more flexible binding ability (Wang et al., 2019). NPC2s in insects found in recent years are similar to OBPs in functionality (Pelosi et al., 2014; Ishida et al., 2014; Zheng et al., 2018). Evolutional analysis showed that NPC2s were soluble protein molecules that appeared earlier than OBPs and CSPs in arthropods (Vizueta et al., 2020). Different from OBPs and CSPs, the secondary structure of NPC2 in insects was mainly β-folded based which form a larger internal binding cavity (Ishida et al., 2014). There are also conserved cysteines in NPC2 sequence, which connected 2 to 3 disulfide bonds to maintain its stable three-dimensional structure (Zhu et al., 2018).

Anagrus nilaparvatae Pang et Wang (Hymenoptera: Mymaridae) is the main egg parasitic wasp of rice pest rice planthopper, which is widely used as a biological control resource in production practice (Zheng et al., 2017). In order to find hosts and supplement food, these parasitoids need to receive and process information of plant volatiles in the complex rice ecosystem, and regulate its behavior to adapt to the environment. Many studies have shown that *A. nilaparvatae* can not only distinguish the volatiles released by rice, but also use the information compounds produced by rice fed by brown planthopper *Nilaparvata lugens*(Stål) to locate the eggs (Lou et al., 2005; Lou et al., 2006).*A. nilaparvatae* can also locate the eggs of the brown planthopper using rice volatiles such as (E)-2 hexenal, methyl salicylate, caryophylene and linalool (Xiao et al., 2012), and some plant essential oils could be attractants of*A. nilaparvatae* for natural enemies to control pests (Mao et al., 2018). In addition, it could identify the necessary wintertime habitat and food source from the vegetation volatiles on the field, for example, they could accurately locate the *Impatiens balsamena* ,*Emilia sonchifolia* and *Sesamum Indicum* , access to essential food supplements to increase longevity and parasitic efficiency (Zhu et al., 2013).

Although there has been previous work on the effects of plant volatiles on the behavior of *A. nilaparvatae* , related to molecular biology research is only on mitochondrial COI and 28s, 5.8s ribosomal genes (Triapitsyn et al., 2018), and molecular feature analysis has not been mentioned. The purpose of this study was to study the characteristics of olfactory binding protein genes in*A. nilaparvatae* , so as to provide theoretical basis for revealing the mechanism of searching hosts and even parasitism. We constructed a full-length transcriptome library, and obtained the olfactory binding protein genes by keyword search, and analyzed the sequence characteristics of these genes. Finally, the expression patterns of these genes in different genders and stimulated by volatiles were quantitatively analyzed. This study provides a basis for the study of the molecular characteristics of the parasitoid, and provides a reference for further revealing the molecular mechanism behind its behavior.

Materials and Methods

2.1. Insects

The hosts *N. lugens* were collected from paddy fields at the farm of South China Agricultural University in Guangdong Province (N 23°9'3", E 113deg20'2") in 2016 and reared with rice hydroponic seedlings.*A. nilaparvatae* were collected in the paddy field of the farm of South China Agricultural University in 2018, and were stably cultured for 60 generations by the rice seedlings with the eggs of *N. lugens* in insect cage (120 mesh gauze). The insect cage was placed in an artificial climate chamber (GXZ-380D, Zhejiang Ningbo

Jiangnan Instrument Factory), and rearing conditions were as follows: photoperiod 14 L:10 D, temperature 25, and Humidity 80%.

2.2. RNA Extraction

150 adult wasps from male and female of *A. nilaparvatae* were used for RNA extraction. For each sample, total RNA was extracted using Eastep$^{(r)}$Super reagents (Promega, Shanghai, China) according to the manufacturer's instructions. RNA concentration and purity were detected by Nanodrop 2000c (Thermofisher, USA), and sample integrity was detected by Agilent 2100 (Rin value d¿7, RNA¿2μg).

2.3. Transcriptome library construction, sequencing and functional annotation

Transcriptome sequencing and library construction were done by Tiangen Biochemical Technology Company (Beijing, China).

Second-generation sequencing: mRNA was enriched by magnetic beads with Oligo(dT) and broken into short fragments by Fragmentation buffer under high temperature; Using mRNA as template, the first cDNA strand was synthesized by adding six base random primers, and then synthesized second cDNA strand. The ends of the double-stranded cDNA were repaired and polyA was added to the 3 ends; The cDNA fragments with connectors were enriched by PCR amplification and sequenced on Illumina HiSeq^{TM}4000 platform. Third-generation sequencing, the experimental process is carried out according to the standard protocal provided by Oxford Nanopore Technologies (ONT) (Jain et al., 2016): Primer annealing, reverse transcription into cDNA, and add switch oligo; RNA digestion, second strand synthesis; DNA damage repair and terminal repair, magnetic bead purification. Add the sequencing connector and sequence it on the machine.

Full-length reads were obtained after used the software Pychopper to filter short fragments and low-quality reads and removed joints of raw fastq data from Nanopore sequencing. Then, the software ONclust2 was utilized to cluster and correct the consensus sequences obtained from the reads. Finally, CD-HIT was used to cluster the full-length transcripts and remove the redundant sequences with more than 90% similarity. Raw image Data files obtained by Illumina sequencing were transformed into original Sequenced Reads/Raw Data by Base Calling analysis. The software TrimMomati was used to remove the joint sequence of reads. After filtering the second-generation short sequence data, compared them to the obtained full-length transcript sequence by the software BWA, and then sort the comparison results. Finally, the full-length transcript was corrected by Pilon according to the comparison results of the second-generation data.

The software Transdecode was used to predict potential coding sequence (CDS) in transcriptional. In order to obtain comprehensive gene function information, six major databases were annotated, including Pfan, Uni-prot, NR, NT, GO, KEGG and TF.

2.4. Screening and structural analysis of olfactory binding protein genes

After annotating the amino acid sequence of unigenes, the olfactory binding protein genes were obtained from the annotation of NR, KEGG and Uni-prot database. The obtained sequences were compared in NCBI BLAST (*https://blast.ncbi.nlm.nih.gov/Blast.cgi*). The open reading frame (ORF) and the amino acid sequence of the proteins expressed by these genes were predicted by NCBI ORF Finder (*https://www.ncbi.nlm.nih.gov/orffinder/*). The molecular size and isoelectric point of protein were predicted by ProtParam tool (*https://web.expasy.org/protparam/*). SignalP (*http://www.cbs.dtu.dk/services/SignalP/*) was used to predict the signal peptide of those proteins. SWISS-MODEL (*https://swissmodel.expasy.org/interactive*) was used to predict the three-dimensional structure of the proteins, and Pfam (*http://pfam.xfam.org/*) was used to search for the conserved domain of protein sequences.

2.5. Phylogenetic analysis

Homologous sequences of eight olfactory binding genes were retrieved using online tool BLAST, and a developmental evolutionary tree was constructed using MEGA 7.0 in combination with olfactory binding

proteins from the published database of Hymenoptera species. Node support was assessed using a bootstrap procedure with 1000 replicates (Tamura et al., 2013).

2.6. RT-qPCR

Three groups of samples were taken: 1) untreated female wasps; 2) female wasps stimulated by 0.01 g/L β-caryophyllene for 1h; 3) the untreated male wasps. The cDNA was obtained in Light Cycler 480 (Roche, USA) according to the reverse transcription kit Gotaq®qPCR Master Mix A6002 and GoScript™ Reverse Transcription System (Promega, USA) operating instructions, each set consisting of 100 wasps, with three replicates per set. The reaction procedure was as follows: 25 , 5min; 42 , 75 min; 70 , 15 min; 4, [?]. The procedure for quantitative PCR reaction was as follows: pre-deformation at 95 for 30 s; Denaturation at 95 , 5 s, annealing at 53 , 15 s; 72 extension, 20 s, 40 cycles; Dissolution curve at 65 , 15s. After the reaction, Light CyCler480 software was used to analyze the Real Time PCR amplification curve and dissolution curve, and the relative expression of the target gene was analyzed according to 2^{-MCt}method. The primer design using Primer3(v.0.4.0) (*https://bioinfo.ut.ee/primer3-0.4.0/*). The primer sequences are showed in Supplementary Table 1.

2.7. Statistical methods

The differences of expression levels of 8 olfactory binding protein genes in response to β- caryophyllene stimulation and between male and female wasps were determined by t -tests (SPSS ver.20.0).

Results

3.1. Transcriptome analysis

163.59Mb Data were obtained after clustering and correcting Raw Data obtained from Nanopore transcriptome sequencing. A total of 224,251 unigenes were obtained, the longest sequence was 11,203 bp, the average length was 729.27 bp, and the N50 was 998 bp. After Illumina correction, the longest sequence was 11,225 bp, with an average length of 729.49 bp and N50 was 998 bp. Statistical results of Nanopore sequencing data and data corrected by Illumina are showed in Supplementary Table 2. Among these unigenes, 151,454 (67.79%) are between 200 and 700bp in length, and 46,207 (20.68%) are more than 1,000 bp in length. Due to full-length transcriptome sequencing, the sequencing was complete. After multiple corrections, the quality of the data group was higher, and the original transcriptome data results and length distribution statistics were shown in Table 1.

3.2. Functional annotation of unigenes

All the sequences of unigenes were compared in NT, NR, Uni-prot, GO, KEGG, PFAM and TF databases, and the results showed database annotation information in the NCBI official nucleic acid database, protein database, studied protein database, GO functional classification, KEGG metabolic pathway, protein family database, NCBI protein database and the transcription factor database. A total of 501,179 items of annotation information was obtained from the transcriptome of the adult wasps, and the annotation information matched 76,326 sequences in the database (Table 2).

Among the above database comparison results, NR database has the most identical sequences, so NR database could be better to cover the output results and fully analyze the sequence homology. According to the distribution statistics of the data with E-value less than 1.0E-5, 17.5% (0-1E-50) alignment sequences showed strong homology, 39.16% (1E-50-1E-20) alignment sequences showed strong homology, and the rest 43.32%(1E-20˜1E-5) showed moderate homology (Figure 1A). Among the sequences alignment, the similarity between 7.48% of the sequences and NR database is higher than 80%, 25.97% of the sequences are between 60% and 80%, and 54.59% of the sequences are less than 60% (Figure 1B). In the comparison results based on NR data,*Ceratosolen solmsi* Marchali (60.53%) had the largest number of matches, followed by *Nasonia vitripennis* (7.08%),*Trichomalopsis sarcophagae* (5.52%) and *Trichogramma pretiosum* (3.74%) (Figure 1C).

A total of 89008 (39.69%) unigenes from the transcriptome were annotated into the GO database. Among all the unigenes, 303,636 correspond to Biological processes, 166,144 correspond to Cellular components,

and 65,857 correspond to Molecular functions. These three categories are further divided into 45 secondary functional annotations, among which, Biological processes are divided into 22 secondary functional items, Cellular components into 11 secondary functional items, and Molecular functions into 12 secondary functional items. In Biological processes, the most unigenes are commented to the Cellular process (52544), followed by Metabolic processes (45,901) and Cellular component organizations or biogenesis (34,715); Among the Cellular components, the most unigenes were annotated to Cell (53422) and Organelle (50859); Among Molecular functions, Binding (24,952) and Structural molecule activity (20,805) have the most notes (Figure 2).

3.3. Identification and bioinformatics analysis of olfactory binding protein genes

By searching for the key words of olfactory binding protein genes, 21 *OBP* s, 5 *CSP* s and 2 *NPC* s were found in PFAM, 10 *OBP* s and 1 *CSP* found in NR, 18 *OBP* s and 1 *NPC* 2 found in UNI-PROT, 15 *OBP* s,1 *CSP* and 1 *NPC2* found in KEGG.

For the retrieved sequences, 34 *OBP* s, 5 *CSP* s and 2 *NPC* s were left after the unigenes with the same sequence number were removed. Then software Mega 7.0 was used to compare nucleic acid similarity and eliminate the repeated unigenes, leaving 5 *OBP* s, 2 *CSP* s and 1 *NPC2* . Finally, ORF Finder was used to obtain the ORF of the sequence and the protein sequence, Pfam was used to retrieve the protein sequence domain, and BLAST was used to retrieve the homologous proteins to name the gene (Table 3). The transcriptome sequencing received 4 *OBP* s, 2 *CSP* s and 1 *NPC2* , namely *AnilOBP2* , *AnilOBP9* , *AnilOBP26* ,*AnilOBP56* , *AnilOBP83a* , *AnilCSP5* , *AnilCSP6* and *AnilNPC2* , respectively corresponding to predict protein for AnilOBP2, AnilOBP9 AnilOBP26, AnilOBP56, AnilOBP83a, AnilCSP5, AnilCSP6 and AnilNPC2.

3.4. Sequence characterization of olfactory binding protein genes

After the ORF prediction of the nucleic acid sequence, the molecular weight, isoelectric point, hydrophilicity and signal peptide of the proteins were predicted, and the results were shown in Table 4.

The ORF of the five *AnilOBP* s has 312-441 bases and encode 103-146 amino acids. The molecular weight of the four AnilOBPs is 11.56-16.54 kD. AnilOBP26 and AnilOBP56 are hydrophobic proteins with acidic isoelectric points, while AnilOBP2, AnilOBP9 and AnilOBP83 are hydrophilic proteins with alkaline isoelectric point. Except for AnilOBP83, other 4 OBPs have signal peptide sequences at the N terminal. The conserved domain of OBPs, namely GOBP family, has 6 conserved cysteines, and these 6 cysteines form three disulfide bonds, and the bond formation rules are Cys1-Cys4, Cys2-Cys5 and Cys3-Cys6. Among the 6 predicted amino acid sequences, except AnilOBP83, the remaining AnilOBPs all have intact conserved domains. Specifically, the cysteine spacing pattern between amino acid 24 and 138 of AnilOBP2 is C1-X26-C2-X3-C3-X39-C4-X12-C5-X8-C6 (X means any amino acid); The cysteine spacing pattern of AnilOBP9 was C1-X29-C2-X3-C3-X18-C4-X22-C5-X11-C6 between amino acids at amino acid positions 22 to 130. The cysteine spacing pattern of AnilOBP26 was C1-X27-C2-X3-C3-X39-C4-X8-C5-X8-C6 between amino acid sites 19 and 131. The cysteine spacing pattern of AnilOBP56 was C1-X27-C2-X3-C3-X38-C4-X8-C5-X8-C6 between amino acid positions 15 and 125. The cysteine spacing pattern of AnilOBP83 was C1-X1-C2-X1-C3-X41-C4-X9-C5-X8-C6 between amino acid positions 1 and 95. The predicted three-dimensional structures of the five AnilOBPs proteins were all OBP proteins, which were mainly composed of α helix structure (Figure 3A-E).

The ORF of *AnilCSP5* and *AnilCSP6* have 234 and 483 bases, encoding 77 and 160 amino acids, with the molecular weights of 8.76 kD and 18.89 kD, respectively. The isoelectric points of AnilCSP5 and AnilCSP6 are alkalic and hydrophilic proteins. There are 29 amino acid residues at the N terminal of AnilCSP5, which are signal peptide sequences, while AnilCSP6 is not annotated to signal peptide. The conserved domain of CSPs, insect pheromone-binding family A10/OS-D, consists of four conserved cysteines sites (Wanner et al., 2004), forming two disulfide bonds. The conserved domains of two AnilCSPs were predicted by Pfam. The conserved domains of AnilCSP5 were between amino acids 4 and 63, but not complete. The conserved domain of AnilCSP6 was located between amino acids 26 and 115, and the cysteine spacing pattern was

C1-X6-C2-X18-C3-X2-C4. The predicted three-dimensional structure of the two AnilCSPs were consistent with the CSP proteins, which were mainly composed of α helix structure (Figure 3F-G).

The ORF of *AnilNPC2* has 465 bases and encodes 154 amino acids, and the molecular weight of AnilNPC2 is 17.07 kD. The isoelectric point of AnilNPC2 is 8.51, and is a hydrophobic protein. The conserved domain prediction results showed that it belonged to the lipid binding protein family and possessed the ML domain (MD-2-related lipid recognition domain), which was located between amino acids 19 and 153, and had 6 conserved cysteines. The cysteine spacing pattern was C1-X15-C2-X4-C3-X46-C4-X12-C5-X39-C6. The three-dimensional structure of AnilNPC2 protein is mainly composed of β folds (Figure 3H).

3.5. Phylogenetic analysis of olfactory binding proteins

In order to understand the differences and similarities between the amino acid sequences of the olfactory binding proteins of *A. nilaparvatae* and its related species, three evolutionary trees, OBPs of 9 species (Figure 4), CSPs of 9 species (Figure 5) and NPC2s of 11 species (Figure 6) were constructed according to the Neighbor-joining (NJ) method. The results show that AnilOBP2 in *A. nilaparvatae* is a clade with MpulOBP7 and MuplOBP12 (*Meteorus pulchricornis*), and then a clade with AnilOBP9; AnilOBP83 with NvitOBP83 (*Nasonia vitripennis*), AnilOBP26 with NvitOBP26, and AnilOBP56 with AbamOBP48 (*Aenasius Bambawalei*) are merged into one branch. The homology between AnilOBP2 and AnilOBP9 is closer than that of other AnilOBPs. AnilCSP5 in *A. nilaparvatae* is far from other CSPs; AnilCSP6 is a clade with AmelCSP5 (*Apis mellifera*) and AcerCSP2 (*Apis ceran*), and then a clade with MpulCSP6 (*Meteorus pulchricornis*). AnilNPC2 in *A. nilaparvatae* is a clade with CsolNPC2b (*Ceratosolen solmsi*), TpreNPC2(*Trichogramma pretiosum)* and NvitNPC2a (*Nasonia vitripennis*) belonged to the same clade, but AnilNPC2 was far from the others.

3.6. Expression of olfactory binding protein genes response to β-caryophyllene

In order to understand the expression response of olfactory binding proteins to plant volatilization stimulation, we used β-caryophyllene, a volatile substance that can attract *A. nilaparvatae* as the stimulus, to compare the changes of the expression levels of eight olfactory binding proteins. The results showed that the relative expression levels of*AnilOBP2* , *AnilOBP26* , *AnilOBP56* , *AnilOBP83* ,*AnilCSP5* and *AnilNPC2* were not significantly different, while *AnilOBP9* was significantly decreased and *AnilCSP6*was significantly increased in response to β-caryophylene stimulation (Figure 7).

3.7. Expression of olfactory binding protein genes in different Genders

In order to understand the expression pattern of olfactory binding proteins in *A. nilaparvatae* between male and female, we quantified the differences in the expression levels of 8 genes between male and female. The results showed that there was no significant difference in the relative expression levels of *AnilOBP2* and*AnilOBP56* between male and female wasps. The relative expression levels of *AnilOBP9* , *AnilOBP26* , *AnilOBP83* ,*AnilCSP5* and *AnilNPC2* in male were significantly higher than that in female. By contrast, the relative expression level of*AnilCSP6* in female was significantly higher than that in male (Figure 8).

Discussion

In this study, a full-length transcriptome database of *A. nilaparvatae* , an important natural enemy of rice planthopper, was constructed. A total of 10,405,444 reads with an average length of 695.59 bp were obtained by Nanopore sequencing. Compared to 4,3657748 reads with an average quality of 35.88 bp were obtained by Illumina sequencing, Nanopore transcriptome sequencing technology can obtain more data, and the read length is longer, indicating the unique advantages of Nanopore in identifying gene sequences. However, the third-generation sequencing platform has the disadvantage of high single base error rate. In order to improve the base accuracy, the sequencing data of Illumina was combined with the correction of Nanopore sequencing data, and 224,251 reads with an average length of 729.49 bp were finally obtained. This is the first time that third-generation transcriptome sequencing has been used in parasitic wasps.

By comparing the obtained unigenes with NT, NR, Uni-Prot and other public databases, 501179 items of

annotation information were obtained successfully. In the NR database, the sequence of *A. nilaparvatae*is high similar with *C. solmsi* , but the transcription sequence annotation amount was low and the similarity degree of most sequences was lower than 60%, which provided reference for the subsequent data mining of *A. nilaparvatae* . A large number of unigenes remain to be annotated in the full-length transcripts. The main reason is that the annotation information of transcriptome is derived from known insect genes, while the genomes of parasitoid wasps are less studied (Branstetter et al., 2018).

Based on the annotated full-length transcriptome data, eight olfactory binding protein genes were obtained by keyword Screening and sequence alignment, including five *OBP* s,*AnilOBP2* , *AnilOBP9* , *AnilOBP26* , *AnilOBP56* , and *AnilOBP83* , two *CSP* s, *AnilCSP5* and*AnilCSP6* and one *NPC2* , *AnilNPC2* . The transcripts of this study annotated a small number of olfactory binding protein genes, because, as mentioned above, there has not been a report on the genomic information of *A. nilaparvatae* , so there is too little annotated information available. On the other hand, olfactory binding protein share low conservatism, for example, OBPs only share on average 10–15% of their residues between species, and CSPs often share 40–50% identical residues between orthologues from phylogenetically distant species (Pelosi et al., 2006; Wang et al., 2019). Finally, the number of genes encoding in insects for olfactory binding protein is highly variable among species, ranging from 12 to about 100 for OBPs and from 4 to 70 for CSPs (Vieira et al., 2012; Pelosi et al., 2014; Qi et al., 2015; Zhou et al., 2015).

Structure and function of proteins correspond, so analyzing the structure of a protein can help predict its function. To data, the structures of more than 20 OBPs have been solved by X-ray crystallography and/or nuclear magnetic resonance (NMR) spectroscopy, some also complexed with ligands (Brito et al., 2016), and the structures of three CSPs are available (Lartigue et al., 2002; Tomaselli et al., 2006; Jansen et al., 2007; Pelosi et al., 2018), which are all spherical structures based on α-helices. By contrast, only one structure of NPC2 was analyzed in*Camponotus japonicus* among insects (Ishida et al., 2014), which is spherical structures based on β-folded. Now, the three-dimensional structure of proteins could be predicted by software based on its amino acid sequence, so as to analyze the spatial conception of the protein and predict its possible function. In our study, after sequence alignment and conserved domain prediction, the eight proteins were identified as the typical characteristics of OBPs, CSPs and NPC2s in insects. The predicted of AnilOBPs and AnilCSPs are spherical structures formed by α-helices, and AnilNPC2 are spherical structures formed by β-folded which is similar to that of CjapNPC2 in C. japonicus (Ishida et al., 2014). Except for AnilOBP83 and AnilCSP6, the other six proteins all contained N-terminal signal peptides, which are predicted to have the function of information binding and transport.

Phylogenetic analysis is helpful to discover the evolutionary relationships of proteins and analyze the homology of species. By evolutionary tree analysis, we found the evolutionary distance of olfactory binding protein genes is far to other insects in Hymenoptera, and they are relatively independent even within the same species. The possible reason is that both OBPs and CSPs seem to have undergone much duplication and differentiation in Hexapoda during evolution (Pelosi et al., 2014), and NPC2s seems earlier than OBPs and CSPs in arthropods (Vizueta et al., 2020). There is low conservatism of olfactory binding protein genes (Pelosi et al., 2006; Wang et al., 2019).

The expression profiles of olfactory binding proteins in parasitic wasps correspond to their biological functions. Therefore, the study of olfactory binding protein expression profiles at different levels is helpful to promote the understanding of olfactory system in parasitic wasps at the molecular level. β-caryophyllene is a volatile released by rice to attract *A. nilaparvatae* seaching eggs of rice planthopper when rice is infested by rice planthoppers (Lou et al., 2005) and behavioral experiments have also shown that β-caryophyllene is attractive to rice planthopper wasps (Xiao et al., 2012). The expression levels of *AnilOBP9* and*AnilCSP6* in female responded significantly to the stimulation of β-caryophylene, in which the expression of AnilOBP9 decreased significantly, and the expression of*AnilCSP6* increased significantly, suggesting that *AnilOBP9*and *AnilCSP6* may have the function of sensing the stimulation of β-caryophylene.

At the same time, the expression of olfactory binding protein genes also showed some gender differences. Except for *AnilOBP2* and*AnilOBP56* , the expression of other genes was different between male and female.

The expression level of *AnilCSP6* in female was significantly higher than that in male, with the high expression stimulated by β-caryophylene, speculating that *AnilCSP6* may play an important role in the host searching in *A. nilaparvatae*females. The differences were common in other parasitoid wasps. For example, the expression levels of two genes of *MmedNPC2* in male were both higher than that in female (Zheng et al., 2018), suggesting that *MmedNPC2* in *M.mediator* may play a role in the perception of plant volatiles. *AnilOBP9* , *AnilOBP26* ,*AnilOBP83* , *AnilCSP5* and *AnilNPC2* were expressed higher in males than in females, we speculated that these genes may encode proteins involved in sexspecific behaviors, including selectively sensing and transporting sex pheromones released by females in the process of molecular recognition and searching for suitable mates. A pattern found in other insects, such as *Locusta migratoria* , where males have a lot of CSPs in their genitals, the high abundance of LmigCSP91 could well bind ∝-naphthalonitrile and β-naphthalonitrile (Ban et al., 2013; Zhou et al., 2013), and in *Adelphocoris suturalis*Jakovlev, *AsutCSP1* was expressed higher in the male antennae than in the female antennae (Cui et al., 2017).

Conclusion

In this study, the first high-quality full-length transcriptome database of *A. nilaparvatae* was constructed. This study provides key data for understanding the complexity of the transcriptome of *A. nilaparvatae* and the sequence and functional annotation information of the subsequent complete reference genome. The molecular characteristics of olfactory binding proteins in *A. nilaparvatae* were further revealed. Eight olfactory binding proteins were screened and identified, and their structures and phylogenetics analysis were conducted. It was speculated that *AnilCSP6* might be raleted to the host reaching in female by RT-qPCR, while the specific functions still need to be further studied.

Acknowledgements

We thank Tiangen Biochemical Technology Company (Beijing, China) help us complete transcriptome sequencing and library construction. This work was supported by the Special Foundation for Key Research Program of Guangzhou (Grant No. 201804020062), the Agricultural Science and Technology Innovative and Promotion Program of Guangdong Province (2020KJ113).

Author contributions

Tingfa Huang and Qiang Zhou designed the study. Ying Ma and Tingfa Huang performed all experiments and analyses, and wrote the first draft of the manuscript. Bingyang Wang and Bingjie Tang performed the experiment and helped to collect data. Jianbai Lui assisted to amend the manuscript. All authors read, commented on and approved the manuscript.

Funding

This work was supported by the Special Foundation for Key Research Program of Guangzhou (Grant No. 201804020062), the Agricultural Science and Technology Innovative and Promotion Program of Guangdong Province (2021KJ113).

Data availability Statement

Transcriptome sequences are available in NCBI accessions SRX11519542 and SRX11519541 (*BioProject ID: PRJNA748197*).

Competing Interests Statement

The authors declare that they have no competing interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Benton R, Vannice KS, Vosshall LB (2007). An essential role for a CD36-related receptor in pheromone detection in *Drosophila* . Nature 450(7167): 289–293.*https://doi.org/10.1038/nature06328*

Branstetter MG, Childers AK, Cox-Foster D, Hopper KR, Kapheim KM, Toth AL, Worley KC (2018). Genomes of the Hymenoptera. Curr Opin Insect Sci 25: 65–75.*https://doi.org/10.1016/j.cois.2017.11.008*

Brito NF, Moreira MF, Melo ACA (2016). A look inside odorant-binding proteins in insect chemoreception. Journal of Insect Physiology 95: 51–65.*https://doi.org/10.1016/j.jinsphys.2016.09.008*

Campanacci V, Lartigue A, Hallberg BM, Jones TA, Giudici-Orticoni MT, Tegoni M, Cambillau C (2003). Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. Proc Natl Acad Sci U S A 100(9): 5069–5074.*https://doi.org/10.1073/pnas.0836654100*

Cui HH, Gu SH, Zhu XQ, Wei Y, Liu HW, Khalid HD, Guo YY, Zhang YJ (2017). Odorant-binding and chemosensory proteins identified in the antennal transcriptome of *Adelphocoris suturalis* jakovlev. Comp Biochem Physiol Part D Genomics Proteomics 24: 139–145.*https://doi.org/10.1016/j.cbd.2016.03.001*

Fan J, Francis F, Liu Y, Chen JL, Cheng DF (2011). An overview of odorant-binding protein functions in insect peripheral olfactory reception. Genet Mol Res 10(4): 3056–3069.*https://doi.org/10.4238/2011.December.8.2*

Grosse-Wilde E, Kuebler LS, Bucks S, Vogel H, Wicher D, Hansson BS (2011). Antennal transcriptome of *Manduca sexta* . Proceedings of the National Academy of Sciences of the United States of America 108(18): 7449–7454.*https://doi.org/10.1073/pnas.1017963108*

Ishida Y, Tsuchiya W, Fujii T, Fujimoto Z, Miyazawa M, Ishibashi J, Matsuyama S, Ishikawa Y, Yamazaki T (2014). Niemann-pick type C2 protein mediating chemical communication in the worker ant. Proc Natl Acad Sci U S A 111(10): 3847–3852.*https://doi.org/10.1073/pnas.1323928111*

Jansen S, Chmelik J, Zidek L, Padrta P, Novak P, Zdrahal Z, Picimbon JF, Lofstedt C, Sklenar V (2007). Structure of *Bombyx mori*chemosensory protein 1 in solution. Arch Insect Biochem Physiol 66(3): 135–145.*https://doi.org/10.1002/arch.20205*

Kaissling KE (1986). Chemo-electrical transduction in insect olfactory receptors. Annu Rev Neurosci 9: 121–145.*https://doi.org/10.1146/annurev.ne.09.030186.001005*

Krieger J, von Nickisch-Rosenegk E, Mameli M, Pelosi P, Breer H (1996). Binding proteins from the antennae of *Bombyx mori* . Insect Biochem Mol Biol 26(3): 297–307.*https://doi.org/10.1016/0965-1748(95)00096-8*

Lartigue A, Campanacci V, Roussel A, Larsson AM, Jones TA, Tegoni M, Cambillau C (2002). X-ray structure and ligand binding study of a moth chemosensory protein. J Biol Chem 277(35): 32094–32098.*https://doi.org/10.1074/jbc.M204371200*

Leal WS (2013). Odorant reception in insects: Roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol 58: 373–391.*https://doi.org/10.1146/annurev-ento-120811-153635*

Lou YG, Du MH, Turlings TC, Cheng JA, Shan WF (2005). Exogenous application of jasmonic acid induces volatile emissions in rice and enhances parasitism of *Nilaparvata lugens* eggs by the parasitoid*Anagrus nilaparvatae* . J Chem Ecol 31(9): 1985–2002.*https://doi.org/10.1007/s10886-005-6072-9*

Lou YG, Hua XY, Turlings TCJ, Cheng TA Chen XX, Ye GY (2006). Differences in induced volatile emissions among rice varieties result in differential attraction and parasitism of *Nilaparvata lugens* eggs by the parasitoid *Anagrus nilaparvatae* in the field. J Chem Ecol, 32(11):2375–2387.*https://doi.org/10.1007/s10886-006-9151-7*

Mao GF, Mo XC, Fouad H, Abbas G, Mo JC (2018). Attraction behaviour of*Anagrus nilaparvatae* to remote lemongrass (cymbopogon distans) oil and its volatile compounds. Nat Prod Res 32(5): 514–520.*https://doi.org/10.1080/14786419.2017.1326486*

Ong RC, Stopfer M (2012). Peripheral and central olfactory tuning in a moth. Chem Senses 37(5): 455–461.*https://doi.org/10.1093/chemse/bjr127*

Pelosi P, Iovinella I, Felicioli A, Dani FR (2014). Soluble proteins of chemical communication: An overview across arthropods. Front Physiol 5: 320.*https://doi.org/10.3389/fphys.2014.00320*

Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR (2018). Beyond chemoreception: Diverse tasks of soluble olfactory proteins in insects. Biol Rev Camb Philos Soc 93(1): 184–200.*https://doi.org/10.1111/brv.12339*

Pelosi P, Maida R (1995). Odorant-binding proteins in insects. Comp Biochem Physiol B Biochem Mol Biol 111(3): 503–514.*https://doi.org/10.1016/0305-0491(95)00019-5*

Pelosi P, Zhou JJ, Ban LP, Calvello M (2006). Soluble proteins in insect chemical communication. Cell Mol Life Sci 63(14): 1658-1676.*https://doi.org/10.1007/s00018-005-5607-0*

Jain M, Olsen HE, Paten B and Akeson M (2016). The Oxford Nanopore Minion: Delivery of nanopore sequencing to the genomics community. Genome Biol 17(1): 239.*https://doi.org/10.1186/s13059-016-1103-0*

Peng Y, Wang SN, Li KM, Liu JT, Zheng Y, Shan S, Yang YQ, Li RJ, Zhang YJ, Guo YY (2017). Identification of odorant binding proteins and chemosensory proteins in *Micropli-tis mediator* as well as functional characterization of chemosensory protein 3. PLoS One 12(7): e0180775.*https://doi.org/10.1371/journal.pone.0180775*

Qi Y, Teng Z, Gao L, Wu S, Huang J, Ye G, Fang Q (2015). Transcriptome analysis of an endoparasitoid wasp *Cotesia chilonis* (Hymenoptera: Braconidae) reveals genes involved in successful parasitism. Arch Insect Biochem Physiol 88(4): 203–221.*https://doi.org/10.1002/arch.21214*

Qu C, Wang Ran, Che WN, Li FQ, Zhao HP, Wei YY, Luo C, Xue M (2021). Identification and tissue distribution of odorant binding protein genes in *Harmonia axyridis* (Coleoptera: Coccinellidae). Journal of Integrative Agriculture 20(8): 2204–2213.*https://doi.org/10.1016/S2095-3119(20)63297-X*

Sanchez-Gracia A, Vieira FG, Rozas J (2009). Molecular evolution of the major chemosensory gene families in insects. Heredity 103: 208–216.*https://doi.org/10.1038/hdy.2009.55*

Sandler BH, Nikonova L, Leal WS, Clardy J (2000). Sexual attraction in the silkworm moth: Structure of the pheromone-binding-protein-bombykol complex. Chem Biol 7(2): 143–151.*https://doi.org/10.1016/s1074-5521(00)00078-8*

Stocker RF (1994). The organization of the chemosensory system in*Drosophila melanogaster* : A review. Cell Tissue Res 275(1): 3–26.*https://doi.org/10.1007/BF00305372*

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). Mega6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12): 2725–2729.*https://doi.org/10.1093/molbev/mst197*

Tegoni M, Campanacci V, Cambillau C (2004). Structural aspects of sexual attraction and chemical communication in insects. Trends Biochem Sci 29(5): 257–264.*https://doi.org/10.1016/j.tibs.2004.03.003*

Tomaselli S, Crescenzi O, Sanfelice D, Ab E, Wechselberger R, Angeli S, Scaloni A, Boelens R, Tancredi T, Pelosi P, Picone D (2006). Solution structure of a chemosensory protein from the desert locust*Schistocerca gregaria* . Biochemistry 45(35): 10606–10613.*https://doi.org/10.1021/bi060998w*

Triapitsyn SV, Rugman-Jones PF, Tretiakov PS, Shih HT, Huang SH (2018). New synonymies in the *Anagrus incarnatus* Haliday 'species complex' (Hymenoptera: Mymaridae) including a common parasitoid of economically important planthopper (Hemiptera: Delphacidae) pests of rice in Asia. Journal of Natural History 52: 2795–2822.*https://doi.org/10.1080/00222933.2018.1552333*

Turlings TCJ, Erb M (2018). Tritrophic interactions mediated by herbivore-induced plant volatiles: Mechanisms, ecological relevance, and application potential. Annu Rev Entomol 63: 433–452.*https://doi.org/10.1146/annurev-ento-020117-043507*

Vieira FG, Foret S, He X, Rozas J, Field LM, Zhou JJ (2012). Unique features of odorant-binding proteins of the parasitoid wasp*Nasonia vitripennis* revealed by genome annotation and comparative analyses. PLoS

One 7(8): e43034.*https://doi.org/10.1371/journal.pone.0043034*

Vieira FG, Rozas J (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: Origin and evolutionary history of the chemosensory system. Genome Biol Evol 3: 476–490.*https://doi.org/10.1093/gbe/evr033*

Vizueta J, Escuer P, Frias-Lopez C, Guirao-Rico S, Hering L, Mayer G, Rozas J, Sanchez-Gracia A (2020). Evolutionary history of major chemosensory gene families across PanArthropoda. Mol Biol Evol 37(12): 3601-3615.*https://doi.org/10.1093/molbev/msaa197*

Vogt RG, Riddiford LM (1981). Pheromone binding and inactivation by moth antennae. Nature 293(5828): 161–163.*https://doi.org/10.1038/293161a0*

Vogt RG, Callahan FE, Rogers ME, Dickens JC (1999). Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris*(Hemiptera, Heteroptera). Chem Senses 24 (5): 481–495.*https://doi.org/10.1093/chemse/24.5.481*

Wang Q, Zhou JJ, Liu JT, Huang GZ, Xu WY, Zhang Q, Chen JL, Zhang YJ, Li XC, Gu SH (2019). Integrative transcriptomic and genomic analysis of odorant binding proteins and chemosensory proteins in aphids. Insect Mol Biol 28(1): 1–22.*https://doi.org/10.1111/imb.12513*

Wanner KW, Willis LG, Theilmann DA, Isman MB, Feng Q, Plettner E (2004). Analysis of the insect *OS -D* -like gene family. J Chem Ecol 30(5): 889–911.*https://doi.org/10.1023/b:joec.0000028457.51147.d4*

Xiao Y, Wang Q, Erb M, Turlings TC, Ge L, Hu L, Li J, Han X, Zhang T, Lu J, Zhang G, Lou Y (2012). Specific herbivore-induced volatiles defend plants and determine insect community composition in the field. Ecol Lett 15(10): 1130–1139.*https://doi.org/10.1111/j.1461-0248.2012.01835.x*

Zheng Y, Wang SN, Peng Y, Lu ZY, Shan S, Yang YQ, Li RJ, Zhang YJ, Guo YY (2018). Functional characterization of a Niemann-pick type C2 protein in the parasitoid wasp *Microplitis mediator* . Insect Sci 25(5): 765–777.*https://doi.org/10.1111/1744-7917.12473*

Zhou CX, Min SF, Yan-Long T, Wang MQ (2015). Analysis of antennal transcriptome and odorant binding protein expression profiles of the recently identified parasitoid wasp, *Sclerodermus* sp.. Comp Biochem Physiol Part D Genomics Proteomics 16: 10–19.*https://doi.org/10.1016/j.cbd.2015.06.003*

Zheng XS, Lu YH, Zhu PY, Zhang FC, Tian JC, Xu HX; Chen GH, Nansen C, Lu ZX (2017). Use of banker plant system for sustainable management of the most important insect pest in rice fields in China. Scientific reports 7: 45581.*https://doi.org/10.1016/S2095-3119(20)63297-X*

Zhou JJ (2010). Odorant-binding proteins in insects. Vitam Horm 83: 241–272.*https://doi.org/10.1016/S0083-6729(10)83010-9*

Zhou XH, Ban L P, Iovinella I, Zhao LJ, Gao Q, Felicioli A, Sagona S, Pieraccini G, Pelosi P, Zhang L, Dani FR (2013). Diversity, abundance, and sex-specific expression of chemosensory proteins in the reproductive organs of the locust Locusta *Migratoria manilensis* . Biol Chem 394(1): 43–54.*https://doi.org/10.1515/hsz-2012-0114*

Zhu J, Guo M, Ban L, Song LM, Liu Y, Pelosi P, Wang G (2018). Niemann-pick C2 proteins: A new function for an old family. Front Physiol 9: 52.*https://doi.org/10.3389/fphys.2018.00052*

Zhu P, Gurr GM, Lu ZX., Chen GH, Zheng XS, Xu HX, Yang YJ (2013). Laboratory screening supports the selection of sesame (*Sesamum indicum*) to enhance *Anagrus* spp. parasitoids (Hymenoptera: Mymaridae) of rice planthoppers. Biological Control 64(1): 83–89.*https://doi.org/10.1016/j.biocontrol.2012.09.014*

Table 1 Statistical results of transcriptome sequencing of *Anagrus nilaparvatae*

Length range	Raw data	Polished
200-700	7506513(72.24 %)	151454(67.79 %)

Length range	Raw data	Polished
700-1200	1438724(13.66 %)	37465(16.52 %)
1200-1700	705180(6.7 %)	17363(7.66 %)
1700-2200	429750(4.08 %)	9262(4.08 %)
2200-2700	169499(1.61 %)	4491(1.98 %)
2700+	141574(1.36 %)	3369(1.51 %)
Number of contig	10405444	224251
Large contig(>=1000bp)	1882079	46207
N50 length(bp)	880	998
Mean contig length(bp)	695.59	729.49

Note: Length range: The Length range of the transcript; Raw data: Nanopore transcriptome sequencing data; Polished: Unigenes obtained from the transcriptome sequencing data after being corrected by Illumina; Number of contig: Number of effective reads; Large contig(>=1000 bp) : Number of reads longer than 1000 bp; N50 Length (bp): Sent the obtained unigene in order of length from large to small, and successively add up the length of unigenes until the length is no less than 50% of the total length; Mean Contig Length (bp) : Mean length of unigenes.

Table 2 Unigenes annotated in different databases

Database	Query number	Percentage	Target number
KEGG	73529	32.79 %	10177
NR	82111	36.62 %	20983
NT	142990	63.76 %	12157
Pfam	22482	10.03 %	3040
TF	6324	2.82 %	192
Uni-prot	84735	37.79 %	17760
GO	89008	39.69 %	12017

Note: Database: Name of Database; Query number: Unigenes number; Percentage: Percentage of unigenes compared to the database; Target number: Number of sequences match in the database.

Table 3 List of olfactory binding protein genes alignment of

Anagrus nilaparvatae

Gene name	Homology search with known protein	Homology search with known protein	Homology search with known
	Scientific name	E-value	Identity
AnilOBP2	*Trichogramma dendrolimi*	6E-73	78.63 %
AnilOBP9	*Apis mellifera*	0.001	26.61 %
AnilOBP26	*Nasonia vitripennis*	2.20E-20	43.91 %
AnilOBP56	*Nylanderia fulva*	2E-28	42.86 %
AnilOBP83	*Nasonia vitripennis*	1E-32	45.63 %
AnilCSP5	*Trichogramma dendrolimi*	2E-20	69.49 %
AnilCSP6	*Apis mellifera*	4.10E-09	30.68 %
AnilNPC2	*Trichogramma pretiosum*	9E-45	52.32 %

Table 4 Characteristic of eight olfactory binding proteins of

Anagrus nilaparvatae

Gene name	ORF length (nt)	Protein length (aa)	Molecular weigth (Da)	pI	Grand average of hydropathicity (GRAVY)	Signal peptide
AnilOBP2	441	146	16542.92	4.74	-0.049	N'(19aa)
AnilOBP9	431	132	15272.52	5.64	-0.573	N'(21aa)
AnilOBP26	420	139	15859.42	8.96	0.631	N'(16aa)
AnilOBP56	396	131	13948.57	8.17	0.037	N'(16aa)
AnilOBP83	312	103	11556.32	4.78	-0.206	-
AnilCSP5	234	77	8760.29	9.05	-0.375	N'(29aa)
AnilCSP6	483	160	18896.61	9.61	-0.838	-
AnilNPC2	465	154	17070.87	8.51	0.185	N'(19aa)