

Identification of the Differences in Olfactory System Between Male and Female Oriental Tobacco Budworm *Helicoverpa assulta*

Haichao Li

Henan Agricultural University

Weizheng Li

Henan Agricultural University

Changjian Miao

Henan Agricultural University

Gaoping Wang

Henan Agricultural University

Man Zhao

Henan Agricultural University

Guohui Yuan

Henan Agricultural University

Xianru Guo (✉ guoxianru@126.com)

Henan Agricultural University

Research article

Keywords: olfactory system, *Helicoverpa assulta*, odorant binding proteins (OBPs), chemosensory proteins (CSPs), pheromone-binding proteins (PBP), odorant receptors (OR), sensory neuron membrane proteins (SNMPs)

Posted Date: January 21st, 2020

DOI: <https://doi.org/10.21203/rs.2.21173/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The olfactory system of insects facilitates their search for host and mates, hence it plays an essential role for insect survival and reproduction. Insects recognize odor substances through olfactory neurons and olfactory genes. Previous studies showed that there are significant sex-specific differences in how insects identify odorant substances, especially sex pheromones. However, whether the sex-specific recognition of odorant substances is caused by differences in the expression of olfaction-related genes between males and females remains unclear.

Results: In order to clarify this problem, the whole transcriptome sequence of the adult *Helicoverpa assulta*, an important agricultural pest of tobacco and other Solanaceae plants, was obtained using PacBio sequencing. RNA-seq analysis showed that there were 27 odorant binding proteins (OBPs), 24 chemosensory proteins (CSPs), 4 pheromone-binding proteins (PBPs), 68 odorant receptors (OR) and 2 sensory neuron membrane proteins (SNMPs) genes that were expressed in the antennae of male and female *H. assulta*. Females had significantly higher expression of *GOBP1-like*, *OBP*, *OBP3*, *PBP3* and *SNMP1* than males, while males had significantly higher expression of *GOBP1*, *OBP7*, *OBP13*, *PBP2* and *SNMP2*.

Conclusions: Our findings improve our understanding of olfactory genes in *H. assulta*, and can be used to further study pheromone identification, mate search, and sex differences in an insect's ability to search for hosts. These results improve our understanding of mate search and host differentiation in *H. assulta*.

Background

A sensitive olfactory system is necessary for the survival of insects and the olfactory system regulates daily behavioral decisions, such as host seeking, mating, ovipositing and avoiding toxic substances (Li et al., 2017; Yi et al., 2018; Zhang et al., 2019). In order to recognize and discriminate against a variety of chemicals in the environment, insects have evolved a family of odorant-gated ion channels comprised of a highly conserved co-receptor (Orco) and a divergent odorant receptor (OR) that confers chemical specificity (Butterwick et al., 2018). Response to an odorant is mediated by the activation of a series of chemosensory receptor proteins, including odorant binding proteins (OBPs), chemosensory proteins (CSPs), pheromone-binding proteins (PBPs), odorant receptors (ORs), odorant-degrading enzymes (ODEs), gustatory receptors (GRs), and ionotropic receptors (IRs).

Binding of odorants to olfactory receptors (ORs) elicits downstream signals, which are further processed to perceive odor in the brain (Bak et al., 2018). ORs are the core odorant detectors for insects to identify their host and mates. Each insect species has a unique repertoire of ORs that have functional properties that match the species' ecological needs, though little is known about the molecular basis of olfaction outside Diptera (de Fouchier et al., 2017).

High-throughput sequencing technology allows for the analysis of gene expression profiles across different tissues or developmental time points. Many studies have analyzed the expression of olfactory

related genes in insects using transcriptomes. For example, the transcriptome of the German cockroach *Blattella germanica*'s antennae contained 73 putative chemosensory genes, with 62 unique genes in this species. These chemosensory genes included 48 OBPs, 9 CSPs, 6 SNMPs, 5 ORs and 5 Irs. Notably, Plus-C OBPs accounted for an exceptionally high proportion (39.58%) of the total 48 OBPs in *Blattella germanica* (Niu et al., 2016). The mealworm *Tenebrio molitor* (Coleoptera: Tenebrionidae) possess 19 OBPs, 12 CSPs, 20 ORs, 6 IRs and 2 SNMPs (Liu et al., 2015). In *Lissorhoptrus oryzophilus* Kuschel (Coleoptera: Curculionidae), 60 candidate chemosensory receptor genes were identified, including 41 ORs, 10 IRs and 9 GRs (Zhang et al., 2019). *Galleria mellonella* Linnaeus (Lepidoptera: Galleriinae) has 46 ORs, 25 IRs, 2 SNMPs, 22 OBPs and 20 CSPs (Zhao et al., 2019). *Bactrocera minax*, a dipteran insect, has one CSPs, 21 OBPs, 53 ORs, 29 IRs and 4 SNMPs (Xu et al., 2019). Identifying evolutionarily shared and novel olfactory genes across different species provides insight into how different species detect odor substances, and thereby alter their behavior.

Sex-specific differences in the olfactory system affect how insects seek hosts and mates. In *Lobesia botrana*, transcriptome analysis of males and females identified 35 OBPs and 18 CSPs, and of these, 20 OBPs were significantly expressed in female antennae, while 5 were up-regulated in males (Rojas et al., 2018). In Asian long-horned beetle *Anoplophora glabripennis* Motschulsky, 12 *AglaOBPs* were expressed specifically in antennae, and seven *AglaOBPs* were highly expressed in the male antennae. These genes may function in the detection of female sex pheromones (Wang et al., 2019).

Helicoverpa assulta is an oligophagous insect that feeds on Solanaceae plants, such as tobacco and hot pepper. In order to understand the olfactory system of *H. assulta*, we generated the first whole transcriptome in *H. assulta* male and female adults using PacBio technology, and compared expression differences in genes involved in olfaction between males and females.

Results

Acquisition of PacBio Data

We constructed two PacBio ISO-Seq libraries, which contained 358,000 and 219,000 Polymerase Reads. The maximum lengths of reads were 119.98kb and 122.36kb, and N50 lengths were 53.98kb and 60.22kb (Supplemental Table 1). Subreads were obtained after removing low-quality sequences (Supplemental Table 2). The maximum read was 99.84 kb and 87.35kb, and the N50 length was 1.68 kb and 1.74 kb, respectively. The length of the reads was primarily distributed between 200 and 2500 bp, which indicated that the data quality was relatively high. Finally, we obtained 1,024,079,888 bp of reads of insert data from all subreads (557,226 reads) (Supplemental Table 3). Mean reads length of the insert was 1.8 kb. By classifying all of the reads that were obtained, we found that the sequence after reads of insert (ROI) could be divided into four categories: full-length non-chimeric sequence, chimeric sequence, non-full-length sequence and short reads. The full-length non-chimeric and non-full-length sequences can be seen in Supplemental Figure 1. The consistent sequence output by Quiver was divided into low quality sequence (LQ) and high quality sequence (HQ) (Figure 1). The information of reads clustering was

summarized in Supplemental Table 4. For subsequent analysis, we only used sequences with consistent Quality values (QV, a per base accuracy estimate). The high quality sequences were merged, and 60,116 transcripts were obtained, with total length of 11,605,404 bp, N50 of 2,104 bp, N90 of 1,178 bp. The longest sequence was 8,052 bp and the shortest sequence was 293 bp. The average GC content of the sequence was 39.49%. The length distributions of the de-redundant transcripts are listed in Supplemental Figure 2A. BUSCO analysis revealed that the transcriptome assembly was relatively complete (Supplemental Figure 2B).

Gene Ontology

A total of 60,116 isoforms were annotated, and 55,860 valid annotations were obtained with the annotation rate of 92.92% (Supplemental Table 5). In the NT database, the annotation rate was 89.79%. Comparing the annotation of *H. assulta* to other insect species (Figure 2A), we found that the sequence consistency of *H. assulta* and *H. armigera* was the highest (74.92%), followed by *Bombyx mori* (3.52%), *Amyelols translata* (2.90%) and *Papilio machaon* (2.42%). Similarly, we annotated the isoforms in the KOG database and compared the isoforms from 25 functional KOG groups (Figure 2B). The results showed that the top five categories of gene data obtained were as follows: General function prediction only (8,738), Signal transduction mechanisms (5,043), Function unknown (3,161), and Posttranslational modification (3,161), and Protein turnover, chaperones (2,986). The Blast2GO software was used to annotate all isoform results from NR to the GO database, and annotations were made in Biological process (8,759), Cellular component (14,394), and Molecular function (8,010). The results showed that Cell (2,703), Cell part (2,663); and Membrane (2,448) were the top three categories with the most genes enriched in cellular component. Cellular process (2,640), Metabolic process (2,207) and Biological regulation (807) were the top three categories with the most genes enriched in Biological process, while Binding (3,548), Catalytic activity (3,111) and Transporter activity (551) are the top three categories with the most genes enriched in molecular function (Figure 2C). Genes expressed in the antennae were mainly enriched in the following KEGG categories: Signal transduction (5,738), Global and overview maps (5600), Cancers: overview (4,012), Endocrine system (3,977), and Transport and catabolism (3,412) (Figure 2D). Taken together, annotations of the transcriptome suggest that genes expressed in the antennae of *H. assulta* are likely involved in signal transduction, recognition, and transport. A summary Venn diagram of the annotated genes is shown in Figure 2E.

Acquisition of RNA-Seq data from male and female antennae of *H. assulta*

We next conducted RNA-seq of the antennae of male and female *H. assulta*, and the resulting data from each sample exceeded 1GB. The Q20 was above 97%, Q30 was above 90%, and 95% of the reads were clean reads (Supplement Table 6). After removing redundant sequences, 60,116 valid unigenes were obtained, and more than 87% of the unigenes were mapped to the PacBio data. Only 10% of the data did not map to the PacBio data (Supplement Table 7).

Sex-specific expression of genes in the antennae of *H. assulta*

We found that 51,558 unigenes were expressed in both males and females. There were 9,979 differential genes, including 3,924 up-regulated unigenes and 6,055 down-regulated unigenes in males and females (Figure 3A). Of the unigenes, 1,734 were specifically expressed in males, and 2,350 were specifically expressed in females (Figure 3B). The differentially expressed genes were clustered and classified into GO functional categories (Figure 3C-D, Supplemental Figure 3), and the most enriched functional classifications were odorant binding, olfactory receptor activity, transmembrane signaling receptor activity, signaling receptor activity and receptor activity. KEGG analysis showed that differentially expressed genes were mainly distributed in signal transduction, global and overview maps, endocrine system and immune system (Supplemental Figure 3). These results suggest that genes expressed in a sex-specific manner in the antennae are involved in the olfactory system, signal recognition, immunity, transportation and hormone regulation.

There were 9799 differentially expressed unigenes in the antennae of males and females (Figure 4A). Among them, there were 1,768 unigenes related to olfactory system genes, including 801 ORs, 673 OBPs, 184 CSPs, 586 PBPs, 375 SNMPs (Supplement table 8). In addition, we identified multiple differentially expressed odor-degrading enzymes, including 7,385 cytochrome oxidase P450 family transcripts, 529 glutathione S-transferases (GSTs), 105 epoxide hydrolase-related lipases, and 20 aldehyde oxidase (AOX). However, we did not obtain any sensilla esterase (SE) related genes (Supplemental Table 9).

The differentially expressed genes were enriched in the GO category of biological functions, such as binding (587), catalytic activity (506), membrane (408), membrane part (384), cell (364) and cell process (352) (Figure 4B). KEGG analysis indicated that the differentially expressed genes were mainly concentrated in global and overview maps pathways (1,018) of the metabolism process, and signal transduction pathway (1,034), endocrine system (748) and transport and catabolism (633) from the environmental information process (Figure 4C). Genes involved in olfactory recognition, transport and metabolism accounted 34.40% of the differentially expressed genes. The differentially expressed genes contained 39 ORs, 71 OBPs, one CSP, 33 PBP, 26 SNMPs, 573 cytochrome P450, 86 GSTs, 28 epoxide hydrolase-related lipases (Epoxide lipase) and five AOXs.

The odorant receptors *Or83b*, *OR2*, *OR47*, and *OR4-like* were differentially expressed in males and females. *Or83b* expression was significantly higher in females, while *OR4-like* was enriched in males. Of the OBPs that were identified, *GOBP 83a-like*, *GOBP1* and *OBP3* were enriched in females and *OBP7*, *GOBP1-like*, *OBP34*, *OBP7*, *GOBP56a-like*, and *OBP13* were enriched in males. Interestingly, PBPs and SNMPs were not differentially expressed, and only one CSP was differentially expressed.

Discussion

Olfaction is essential for insects to identify host substrates, mates and oviposition sites (Qi et al., 2018). OBPs in the insect antennae transport pheromones and other semiochemicals to activate olfactory signals across the sensillum lymph to the ORs (He et al., 2010). To better understand the olfactory system of *H. assulta*, we generated an antennal PacBio transcriptome. *H. assulta* has a very narrow host range

compared to other species in Noctuidae. We identified 60,116 unigenes in the antennae of *H. assulta*, and of these, 55,860 unigenes were annotated using seven functional databases. We also compared gene expression patterns in male and female antennal tissues RNA-Seq. We obtained 9979 differentially expressed unigenes in the male and female antennae, which accounted for 17% of all unigenes. These results suggest that there is high level of sex-specific gene expression in *H. assulta* antennae. The differentially expressed genes were involved in environmental signal recognition, signal transduction, signal transit, signaling molecules and interaction, cell growth and death, translation, the immune system, amino acid metabolism, and developmental pathways. These results are consistent with the known function of the antenna for signal recognition in insects. We also found that genes involved in olfaction were significantly differentially expressed in male and female antennae. For example, there were 170 OBP and OR genes that were differentially expressed in males and females, accounting for 1.70% of the differentially expressed genes. In addition, 9.62% of the differentially expressed genes were olfactory related unigenes and 6.93% of the differentially expressed unigenes were involved in smell degradation. These findings indicate that there are significant differences in the expression of olfactory genes between males and females in the antennae.

Odorant-Binding Proteins (OBPs) and Chemosensory Proteins (CSPs)

OBPs and CSPs play essential roles in insect chemosensory recognition (Wang et al., 2019; Zhang et al., 2018). In *Carposina sasakii* Matsumura, GOBPs and PBPs have stronger binding affinities to some sex pheromones, and PBP1 has high affinity to sex pheromones and some plant volatiles, OBP1-2 has high binding affinity to general plant volatiles (Tian et al., 2019). In the silkworm *Bombyx mori*, OBPs are subdivided into three main subfamilies, PBPs, GOBPs and antennal-binding proteins (ABPs). PBPs and GOBPs are thought to be involved in both sensing and transporting the sex pheromone bombykol and host volatiles, respectively (Zhou et al., 2009). *BmorPBP1* and *BmorGOBP2* are expressed specifically and at high levels in the antennae, consistent with their involvement in olfaction (He et al., 2010). *BmorPBP1*, *BmorPBP2*, *BmorGOBP2* and *BmorABPx* all bind to bombykol (Zhou et al., 2009). OBPs in *Holotrichia oblita* could mediate its interaction with organic fertilizer-derived attractants, such as skatole, p-cresol, and 4-allylanisole (Wang et al., 2019).

In *Myzus persicae*, three OBP genes (OBP3/7/8), as well as *CSP1/4/6*, *CSP2/9* and *CSP5/8* are located in tandem in the genome (Wang et al., 2019). In honeybees, CSP4 interacts with volatile pheromones that are the primary signals larvae send to the nurses to receive care (Wu et al., 2019). In *Cnaphalocrocis medinalis*, CmedCSP33 can bind to seven out of 32 rice volatiles. Fluorescence quenching experiments revealed that CmedCSP33 forms a stable complex with nerolidol and β -ionone, and these two compounds cause conformational changes in CmedCSP33 that can be detected by circular dichroism (CD) spectra (Duan et al., 2019).

We obtained 27 OBP genes, 24 CSP genes and 4 PBP genes in male and female *H. assulta* adult antennae (Supplemental Table 8). PBPs recognize sex pheromones, and we found that *PBP2* was expressed strongly in males, while *PBP3* was female-specific. CSP was expressed in both males and

females. Similarly, *GOBP1*, *OBP7*, and *OBP13* were highly expressed in males, while *GOBP1-like*, *OBP*, and *OBP3* were highly expressed in females. These differentially expressed genes may be involved in mate recognition.

Olfactory Receptor (OR) Genes

Currently, a large amount of Lepidoptera OR repertoires have been identified from transcriptome and genome sequencing, and these studies shown that broadly defined environmental adaptations (e.g., terrestrial vs. aquatic) are correlated with the number of functional and non-functional OR genes that are retained by a species (Hughes et al., 2018). Molecular analyses of OR gene families have demonstrated that duplications in the OR family 5/8/9 were closely related to the dietary niche of herbivores (Hughes et al., 2018). In our study, we found 68 OR genes, and only eight of these genes showed sex-specific expression patterns. Therefore, whether these genes are required for recognizing sex-specific pheromones requires further study.

SNMPS

In insects, olfactory sensory neurons (OSNs) detect male and female pheromone signals by expressing SNMP1. SNMP1 is thought to be the common receptor involved in transmitting pheromones to neighboring pheromone receptors (Blankenburg et al., 2019). In the present study, we found that SNMP1 was mainly expressed in females and SNMP2 was mainly expressed in males of *H. assulta*. In *Drosophila*, absence of SNMP1 can significantly reduce the response time of the female to male sex pheromones (Li et al., 2014), indicating that SNMP1 is involved in sex pheromone recognition.

Conclusion

We generated full-length transcriptomes from antennae of *H. assulta* using PacBio sequencing and RNA-Seq, and obtained a relatively complete pattern of olfactory gene expression. In total, 27 OBPs genes, 24 CSPs genes, four PBPs genes, 68 ORs genes and two SNMPS genes were identified, and genes expressed in a sex-specific manner were further analyzed. We found that *GOBP1-like*, *OBP*, *OBP3*, *PBP3*, *SNMP1* were enriched in females and *GOBP1*, *OBP7*, *OBP13*, *PBP2* and *SNMP2* were enriched in males. We also found that although there were many OR genes that were differentially expressed in males and females, most of these patterns were not significantly. Our findings improve our understanding of olfactory genes in *H. assulta*, and can be used to further study pheromone identification, mate search, and sex differences in an insect's ability to search for hosts.

Methods

Insects and Samples

Larvae of *H. assulta* were obtained from the tobacco field in Xuchang campus of Henan Agricultural University, Henan province, China. In the laboratory, larvae were reared on an artificial diet at 26 ± 1 °C

with a photoperiod of 14:10 (L: D). Female and male pupae were kept separately. After emergence, moths were provided with a 10 % (w/w) honey solution for feeding. Ten moths of each sex were collected every day from 1d to 5d after emergence, and the antennae of the moths were dissected and immediately frozen in liquid nitrogen for 10 minutes and then stored at - 80 °C until [RNA extraction](#). Each 10 pairs of antennae from males or females were mixed into a pool respectively as a repeat. Samples were collected for each sex three times.

RNA Extraction, PCR, PacBio and RNA sequencing

Total RNA was extracted from antennae samples using TRIzol Reagent following the manufacturer's protocol. The purity and concentration of the total RNA was measured by a NanoPhotometer and Qubit 3.0 Fluorometer (Implen, USA), respectively. Integrity of the RNA was evaluated using an Agilent Bioanalyzer 2100 system (Agilent, USA) according to the manufacturer's protocol.

PacBio Sequencing carried out at the BGI using High-seq2500 (BGI-Shenzhen, China). To obtain complete information of all transcripts, full length transcriptome sequencing was conducted. The best RNA sample out of the three replicates was selected for male and female adult antennal samples of *H. assulta*, then mixed together as one sample at a 1:1 ratio for PacBio SMRT-seq. The full-length cDNAs were synthesized using the SMRTer PCR cDNA Synthesis Kit (BGI-Shenzhen, China). Two SMAT cells were run on the PacBio RSII platform. The resulting library was sequenced using the Iso-Seq function of the PacBio RSII system and RNA-Seq was conducted using BGISEQ500 (BGI-Shenzhen, China). Intact total RNA samples were used for constructing cDNA libraries using the NEBNext Ultra™ RNA Library Prep Kit from Illumina following the manufacturer's protocol. The cDNA libraries were sequenced on a BGISEQ500 platform using the BGISEQ500 Reagent Kit according to the manufacturer's protocol, and 150 bp paired-end raw reads were generated. Reads were only defined as full-length if the 5' and 3' terminal joints and polyA of the original sequence were detected. After the ROI treatment, only full-length non-chimeric and non-full-length sequences were used for further analysis.

Bioinformatics and differential expression analysis

For SMRT-seq, reads of insert (ROI) sequences were extracted. Consistent sequences were polished using the Quiver algorithm to cluster non-full-length sequences, and low-quality sequences were corrected using RNA-seq data.

Clean reads were obtained by removing reads containing adapters and low-quality reads with more than 50% N bases using the QoRTs software. After removing redundant sequences, we annotated unigene isoforms using seven functional databases (NR, NT, GO, KOG, Pfam, KEGG and SwansPprot). Gene expression levels were estimated using the RSEM software (version: 1.2.25), and FPKM (fragments per kilobase of transcript per million mapped reads) values were used to describe the expression levels of genes.

Differential expression analysis

Differential expression analysis of male and female antennae was performed using DESeq. DESeq provides statistical tests for determining differential expression of digital gene expression data using a model based on a negative binomial distribution (Hughes et al., 2018). The P values were adjusted using the Benjamini and Hochberg's approach to account for the false discovery rate (FDR). Genes that had a two fold difference (the absolute value of log₂ ratio ≥ 1) between two stages and a FDR less than 0.01 were categorized as differentially expressed genes (DEGs).

Supplementary Information

- **Additional files1: Figure S1. Reads classification.** The sequences after insert recognition (Reads of insert, ROI) processing can be divided into four categories: full-length non-chimeric, chimeric, non-full-length and short reads. A full-length read is defined to meet three indexes (5' and 3' ends of the original sequence and a poly A signal)., must be able to detect all three to define a full-length read. After ROI treatment, only full-length non-chimeric sequences and non-full-length sequences were used for the next analysis. A. Hassan A01 library; B. Hassan D01 library.
- **Additional files2: Figure S2. Merge libraries and remove redundancy.** The high-quality sequences obtained after clustering and error correction for each library are finally merged together to remove redundancy, and the final transcript sequence obtained after deduplication. A. Length distribution of de-redundant transcripts. B. Correspondence before and after Isoform clustering. The quality of assembled transcripts was assessed using a single copy ortholog database BUSCO Gene comparison, to some extent, illustrates the integrity of transcriptome assembly.
- **Additional files3: Table S1. Summary of Polymerase Reads.**
- **Additional files4: Table S2. Summary of Subreads.**
- **Additional files5: Table S3. Summary of PacBio sequence.**
- **Additional files6: Table S4. Information summary table of the Read cluster of each library.**
- **Additional files7: Table S5. Information summary table of the Read cluster of each library.**
- **Additional files8: Table S6. Summary of RNA-seq Reads.**
- **Additional files9: Table S7. Summary of gene mapping.**
- **Additional files10: Table S8. Olfactory Related Transcript Information**
- **Additional files11: Table S9. Possible odor-degrading enzyme transcript information**

Abbreviations

Orco: odorant-gated ion channels comprised of a highly conserved co-receptor,

OBPs: odorant binding proteins,

CSPs: chemosensory proteins,

PBPs: pheromone-binding proteins,

ORs: odorant receptors,

ODEs: odorant-degrading enzymes,

GRs: gustatory receptors,

IRs: ionotropic receptors,

BUSCO: Benchmarking Universal Single-Copy Orthologs,

SRA: Sequence Read Archive,

KEGG: Kyoto Encyclopedia of Genes and Genomes,

GO: Gene Ontology,

NR: Non-Redundant Protein Sequence Database,

NT: Nucleotide Sequence Database,

SWISS-PROT: SwissProt Database,

Pfam: Protein families,

KOG: EuKaryotic Orthologous Groups,

CDS: the candidate coding regions,

PacBio: PacBio Sequel Systems are powered by Single Molecule.

Declarations

Acknowledgements

We thank MogoEdit, for editing the English text of a draft of this manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 31572331 and 31772520), Provincial Natural Science Foundation of Henan Province (grant no. 142300410019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available in the additional files or from the corresponding author on reasonable request.

Author information

College of Plant Protection, Henan Agricultural University, Zhengzhou 450002, PR China.

Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200032, PR China.

Author Contributions

HC and WZ performed the experiments. HC and XR and GH designed and wrote the manuscript. HC and CJ analyzed the data. XR revised the manuscript. All authors read and approved the manuscripts.

Ethics approval and consent to participate

The insect material used in our laboratory breeding.

Consent for publication

Not applicable.

Conflict Of Interest

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

Competing interests

No Competing interests.

References

- Bak, J. H., Jang, S., and Hyeon, C. (2018). Implications for human odor sensing revealed from the statistics of odorant-receptor interactions. *PLoS Comput Biol.* 14:e1006175. doi:10.1371/journal.pcbi.1006175
- Blankenburg, S., Cassau, S., and Krieger, J. (2019). The expression patterns of SNMP1 and SNMP2 underline distinct functions of two CD36-related proteins in the olfactory system of the tobacco budworm *Heliothis virescens*. *Cell Tissue Res.* 378, 485-497. doi:10.1007/s00441-019-03066-y
- Butterwick, J. A., del Marmol, J., Kim, K. H., Kahlson, M. A., Rogow, J. A., Walz, T., et al. (2018). Cryo-EM structure of the insect olfactory receptor Orco. *Nature.* 560, 447-452. doi:10.1038/s41586-018-0420-8
- de Fouchier, A., Walker, W. B., Montagne, N., Steiner, C., Binyameen, M., Schlyter, F., et al. (2017). Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat.*

Commun. 8, 15709. doi:10.1038/ncomms15709

Duan, S. G., Li, D. Z., Wang, M. Q. (2019). Chemosensory proteins used as target for screening behaviourally active compounds in the rice pest *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae). *Insect Mol. Biol.* 28, 123-135. doi: 10.1111/imb.12532

He, X., Tzotzos, G., Woodcock, C., Pickett, J.A., Hooper, T., Field, L. M., et al. (2010). Binding of the general odorant binding protein of *Bombyx mori* BmorGOBP2 to the moth sex pheromone components. *J. Chem. Ecol.* 36, 1293-1305. doi:10.1007/s10886-010-9870-7

Hughes, G. M., Boston, E. S. M., Finarelli, J. A., Murphy, W. J., Higgins, D. G., Teeling, E. C.(2018). The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Mol. Biol. Evol.* 35, 1390-1406. doi: 10.1093/molbev/msy028

Li, K.B, Wei, H. S, Shu, C. L, Zhang, S., Cao, Y. Z, Luo, C., et al. (2017). Identification and comparison of candidate odorant receptor genes in the olfactory and non-olfactory organs of *Holotrichia oblita* Faldermann by transcriptome analysis. *Comp Biochem Physiol Part D Genomics Proteomics.* 24, 1-11. doi: 10.1016/j.cbd.2017.07.001

Li, Z. Z, Ni, J. D., Huang, J., and Montell, C. (2014). Requirement for *Drosophila* SNMP1 for rapid activation and termination of pheromone-induced activity. *PLoS Genet.* 10:e1004600. Epub 2014/09/26. doi: 10.1371/journal.pgen.1004600

Liu, S., Rao, X. J., Li, M. Y., Feng, M. F., He, M. Z., and Li, S. G. (2015). Identification of candidate chemosensory genes in the antennal transcriptome of *Tenebrio molitor* (Coleoptera: Tenebrionidae). *Comp Biochem Physiol Part D Genomics Proteomics.* 13, 44-51. doi: 10.1016/j.cbd.2015.01.004.

Niu, D. J., Liu, Y., Dong, X. T., and Dong, S. L. (2016). Transcriptome based identification and tissue expression profiles of chemosensory genes in *Blattella germanica* (Blattaria: Blattidae). *Comp Biochem Physiol Part D Genomics Proteomics.* 18, 30-43. doi: 10.1016/j.cbd.2016.03.002

Qi, M. M, Wei, S. S, and Wei, C. (2018). Identification of candidate olfactory genes in cicada *Subpsaltria yangi* by antennal transcriptome analysis. *Comp Biochem Physiol Part D Genomics Proteomics.* 28, 122-133. doi: 10.1016/j.cbd.2018.08.001

Rojas, V., Jimenez, H., Palma-Millanao, R., Gonzalez-Gonzalez, A., Machuca, J., Godoy, R., et al. (2018). Analysis of the grapevine moth *Lobesia botrana* antennal transcriptome and expression of odorant-binding and chemosensory proteins. *Comp Biochem Physiol Part D Genomics Proteomics.* 27, 1-12. doi: 10.1016/j.cbd.2018.04.003

Tian, Z. Q., Qiu, G. S., Li, Y. Y., Zhang, H. J., Yan, W. T., Yue, Q., et al. (2019). Molecular characterization and functional analysis of pheromone binding proteins and general odorant binding proteins from *Carposina sasakii* Matsumura (Lepidoptera: Carposinidae). *Pest Manag. Sci.* 75, 234-245. doi: 10.1002/ps.5107

- Wang, C. Q., Li, J. Q., Li, E. T., Nyamwasa, I., Li, K. B., Zhang, S., et al. (2019). Molecular and functional characterization of odorant-binding protein genes in *Holotrichia oblita* Faldermann. *Int J. Biol. Macromol.* 136, 359-367. doi: 10.1016/j.ijbiomac.2019.06.013
- Wang, J. Z., Gao, P., Luo, Y. Q., and Tao, J. (2019). Characterization and expression profiling of odorant-binding proteins in *Anoplophora glabripennis* Motsch. *Gene.* 693, 25-36. doi: 10.1016/j.gene.2018.12.075
- Wang, Q., Zhou, J. J., Liu, J. T., Huang, G. Z., Xu, W. Y., Zhang, Q., et al. (2019). Integrative transcriptomic and genomic analysis of odorant binding proteins and chemosensory proteins in aphids. *Insect Mol. Biol.* 28, 1-22. doi: 10.1111/imb.12513
- Wu, F., Feng, Y. L., Han, B., Hu, H., Feng, M., Meng, L. f., et al. (2019). Mechanistic insight into binding interaction between chemosensory protein 4 and volatile larval pheromones in honeybees (*Apis mellifera*). *Int. J. Biol. Macromol.* 141,553-563. doi: 10.1016/j.ijbiomac.2019.09.041
- Xu, P. H., Wang, Y. H., Akami, M., and Niu, C. Y. (2019). Identification of olfactory genes and functional analysis of BminCSP and BminOBP21 in *Bactrocera minax*. *PLoS One.*14:e0222193. doi: 10.1371/journal.pone.0222193
- Yi, J. K., Yang, S., Wang, S., Wang, J., Zhang, X. X., Liu, Y., et al. (2018). Identification of candidate chemosensory receptors in the antennal transcriptome of the large black chafer *Holotrichia parallela* Motschulsky (Coleoptera: Scarabaeidae). *Comp Biochem Physiol Part D Genomics Proteomics.* 28, 63-71. doi: 10.1016/j.cbd.2018.06.005
- Zhang, K. P., Feng, Y. L., Du, L. X., Gao, S. S., Yan, H., Kun, L., et al. (2019). Functional analysis of MsepOR13 in the oriental armyworm *Mythimna separata* (Walker). *Front. Physiol.* 10:367. doi: 10.3389/fphys.2019.00367
- Zhang, X. X., Yang, S., Zhang, J. H, Wang, X., Wang, S., Liu, M. N, et al. (2019). Identification and expression analysis of candidate chemosensory receptors based on the antennal transcriptome of *Lissorhoptus oryzophilus*. *Comp Biochem Physiol Part D Genomics Proteomics.* 30, 133-142. doi: 10.1016/j.cbd.2019.02.007
- Zhang, Y., Tan, Y., Zhou, X. R., and Pang, B. P. (2018). A whole-body transcriptome analysis and expression profiling of odorant binding protein genes in *Oedaleus infernalis*. *Comp Biochem Physiol Part D Genomics Proteomics.* 28, 134-141. doi: 10.1016/j.cbd.2018.08.003
- Zhao, H. X., Xiao, W. Y., Ji, C. H., Ren, Q., Xia, X. S., Zhang, X. F., et al. (2019). Candidate chemosensory genes identified from the greater wax moth, *Galleria mellonella*, through a transcriptomic analysis. *Sci. Rep.* 9, 10032. doi: 10.1038/s41598-019-46532-x
- Zhou, J. J., Robertson, G., He, X. L., Dufour, S., Hooper, A. M., Pickett, J. A., et al. (2009). Characterisation of *Bombyx mori* odorant-binding proteins reveals that a general odorant-binding protein discriminates

Figures

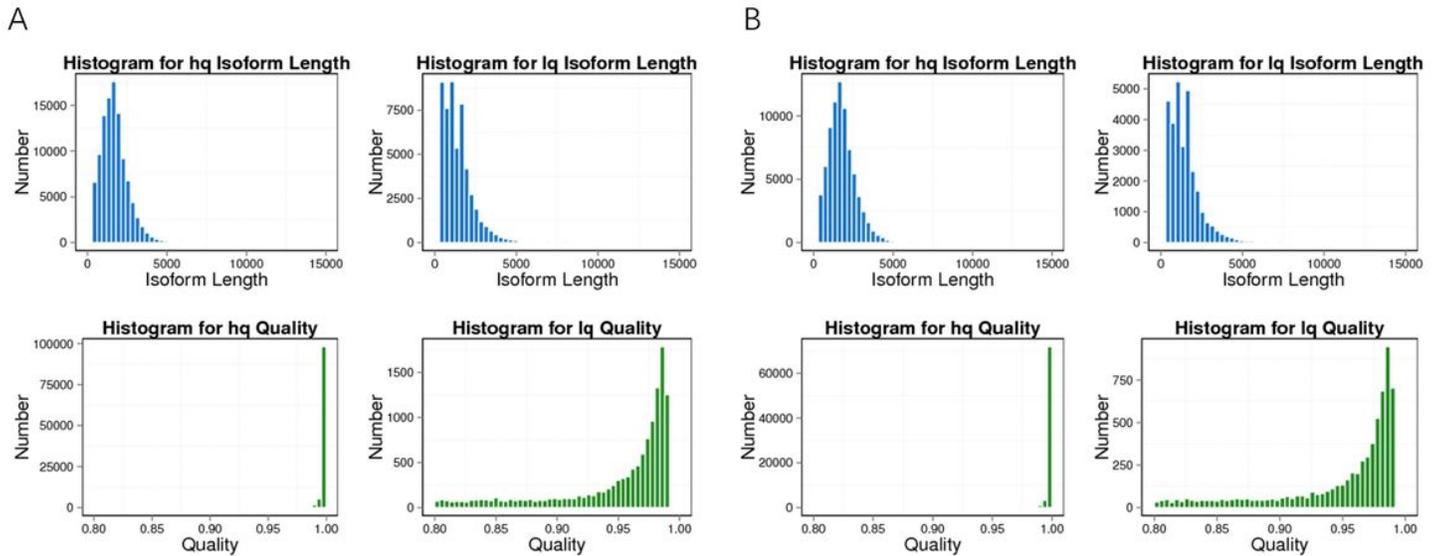


Figure 1

Comparison diagram of QV sequences on sequence length and sequence mass value. Data processing uses the SMRT analysis suite to identify incorporation fragments, Reads Classify, Reads Cluster and Reads Quvier, and finally obtain high-quality full-length consensus sequences. Further, each insert is identified based on the data Subreads data, and then integrated to generate inserts Consensus form of the sequence, the sequence length and sequence quality distribution of high-quality (hq) and low-quality (lq) isoforms in each library are illustrated. A. Low QV, B. High QV. In the Histogram for Isoform Length graph, the x-axis represents the read length, and the left y-axis represents the number of reads; In the Histogram for Quality graph, the x-axis represents the read quality value, and the left y-axis represents the number of reads.

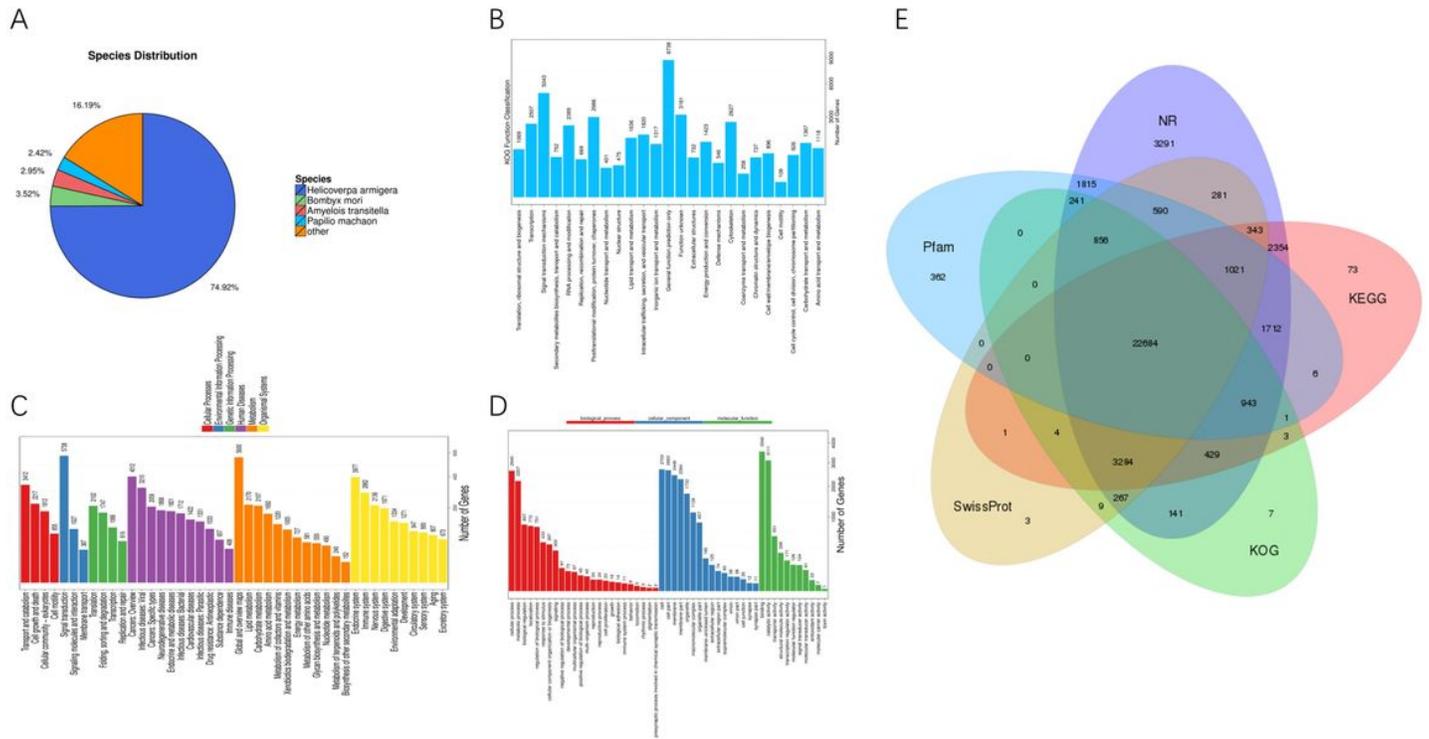


Figure 2

Overall analysis of PacBio data of adult antennae of *H. assulta*. After the de-redundancy is completed, we performed an annotation on the seven major functions of the Uniq Isoform obtained by the de-redundancy. A. According to the annotation results of the NR database, the proportion of different species on the isoform annotation is counted, and the species distribution map is drawn; B. Annotate Isoform to the KOG database, and compare the statistics of Isoform in the 25 functional groups of the KOG database; C. Compare all Isoforms to the KEGG database, and count the Isoforms on the level1 and level2 levels of the KEGG database, and draw a statistical diagram of the KEGG function distribution; D. Classification diagram. Blast2GO software was used to annotate the Isoform results of all comparisons to the NR database to the GO database, and statistically annotate to the three aspects of GO: Biological Process, Cellular Component, and Molecular Function; E. Use Venn diagrams to show annotated results for NR, KOG, KEGG Pfam, and SwissProt.

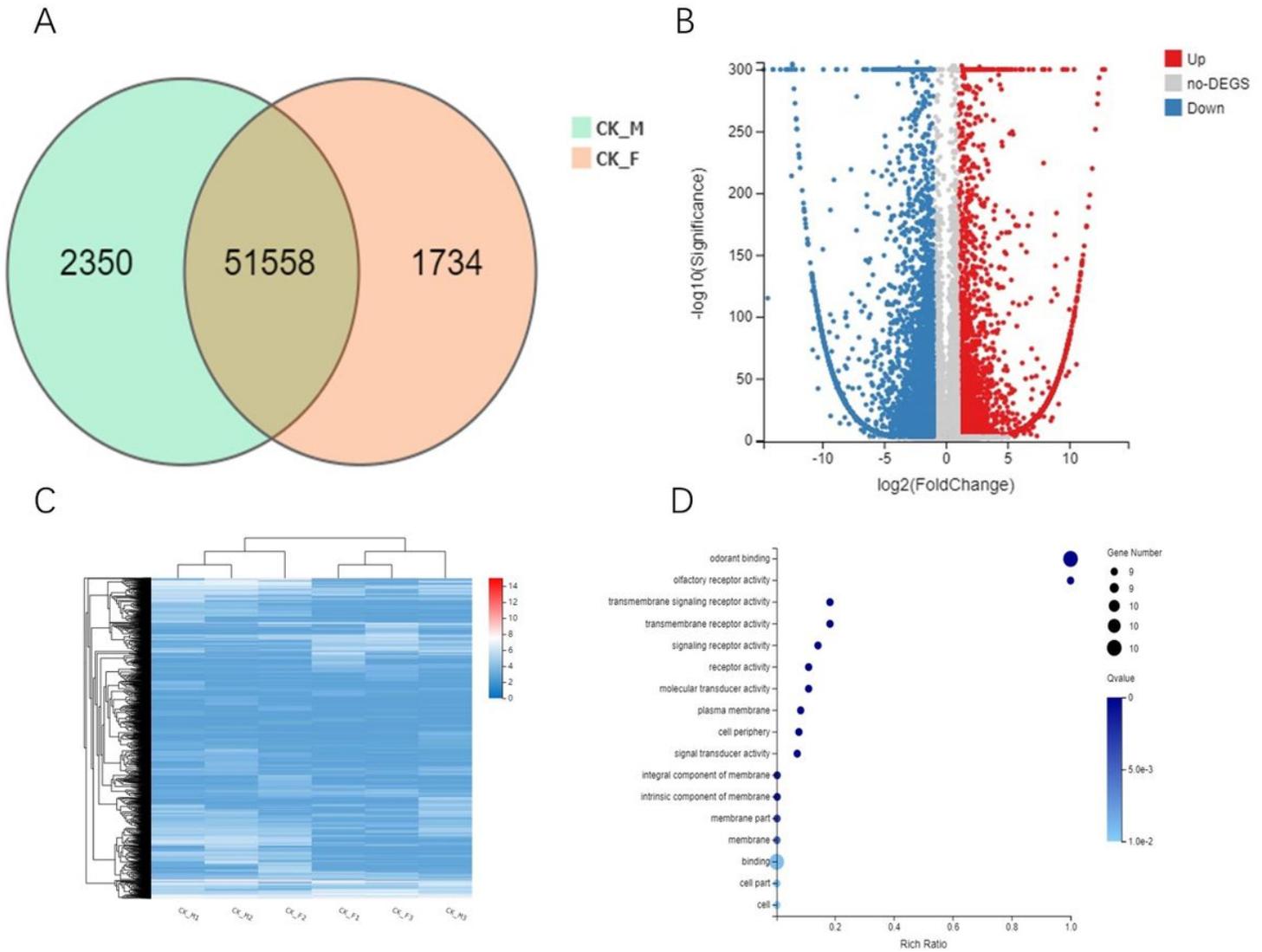


Figure 3

Differential expression analysis of olfactory genes between male and female of *H. assulta*. Bowtie2 was used to align clean reads to the PacBio sequence of *H. assulta*, and then RSEM was used to calculate the gene expression level of each sample. A smoothed estimate of the intensity-dependent noise level is done by loess regression, and converted to local standard deviations (SDs) of M conditioned on A, under the assumption of normal distribution. The local SDs are then used to identify the difference of the gene expression between the two samples (Fold Change ≥ 2 and Adjusted P value ≤ 0.001). A. Venn diagram of differentially expressed genes in females and males of *H. assulta*; B. Volcano map of differentially expressed genes between female and male antennae of *H. assulta*; C. Cluster analysis of differentially expressed genes in female and male antennae of *H. assulta*; D. Enrichment analysis of KEGG pathway differentially expressed genes in female and male antennae of *H. assulta*.

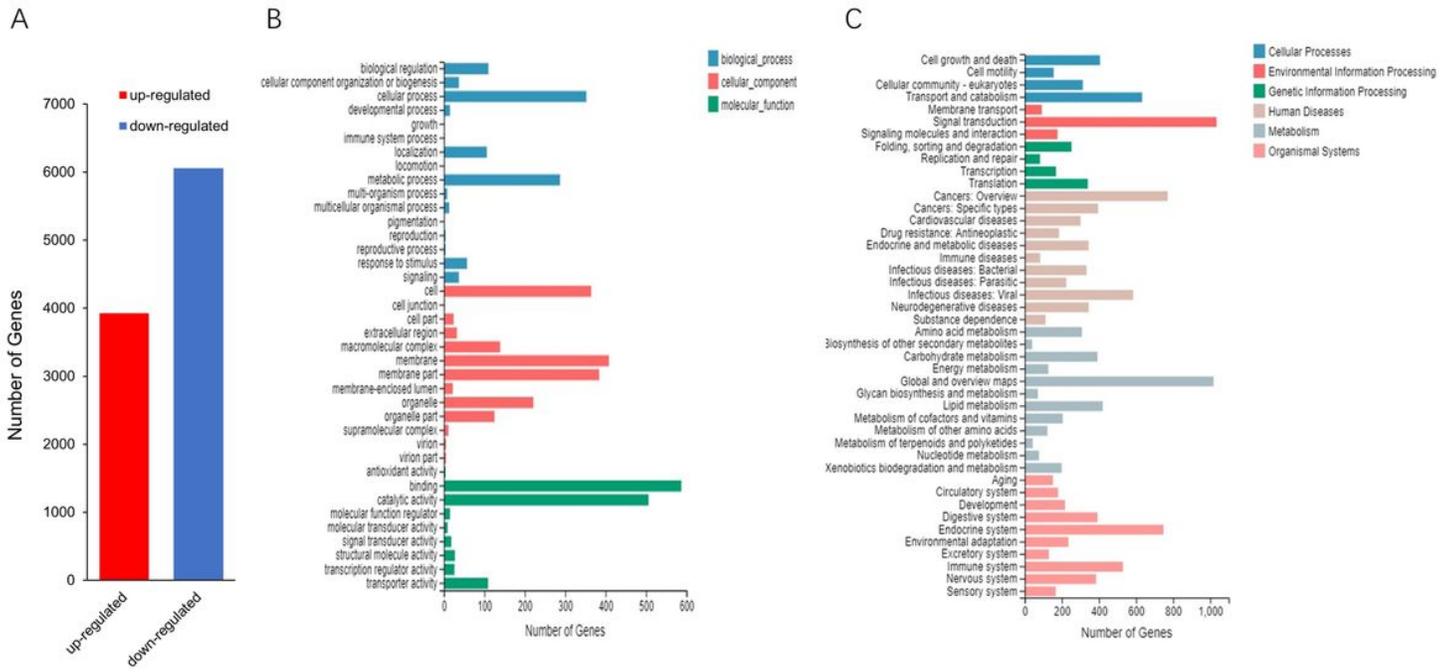


Figure 4

Functional enrichment analysis of antennae differentially expressed genes in female and male of *H. assulta*. A. Up-regulated and Down-regulated genes in differentially expressed antennae of female and male of *H. assulta* (P value ≤ 0.001); B. The GO classification of differentially expressed genes in female and male antennae of *H. assulta* (P value ≤ 0.001 ; FDR ≤ 0.01), Gene Ontology is divided into three functional categories: molecular function, cellular component and biological process. Functional classification based on differential gene test results. There are sub-categories for each level under each major category; C. The KEGG in the differentially expressed genes of female and male antennae of *H. assulta* (P value ≤ 0.001 ; FDR ≤ 0.01). Classification based on the results of KEGG notes and official classifications. We classify the differential genes into biological pathways and divide the genes involved in the KEGG metabolic pathway into 6 branches: Cellular Processes, Environmental Information Processing, Genetic Information Processing, Human Disease (animal only), Metabolism, Organic Systems. Statistics are further classified under each branch.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfiles39.docx](#)
- [Additionalfiles1.JPG](#)
- [Additionalfiles2.JPG](#)
- [Additionalfiles11.xlsx](#)
- [Additionalfiles10.xlsx](#)