

SCIENTIFIC REPORTS

OPEN

Characterization of the Adult Head Transcriptome and Identification of Migration and Olfaction Genes in the Oriental Armyworm *Mythimna separate*

Hai-Xu Bian¹, Hong-Fang Ma¹, Xi-Xi Zheng¹, Ming-Hui Peng¹, Yu-Ping Li¹, Jun-Fang Su², Huan Wang¹, Qun Li¹, Run-Xi Xia¹, Yan-Qun Liu¹ & Xing-Fu Jiang³

The oriental armyworm *Mythimna separate* is an economically important insect with a wide distribution and strong migratory activity. However, knowledge about the molecular mechanisms regulating the physiological and behavioural responses of the oriental armyworm is scarce. In the present study, we took a transcriptomic approach to characterize the gene network in the adult head of *M. separate*. The sequencing and *de novo* assembly yielded 63,499 transcripts, which were further assembled into 46,459 unigenes with an N50 of 1,153 bp. In the head transcriptome data, unigenes involved in the 'signal transduction mechanism' are the most abundant. In total, 937 signal transduction unigenes were assigned to 22 signalling pathways. The circadian clock, melanin synthesis, and non-receptor protein of olfactory gene families were then identified, and phylogenetic analyses were performed with these *M. separate* genes, the model insect *Bombyx mori* and other insects. Furthermore, 1,372 simple sequence repeats of 2–6 bp in unit length were identified. The transcriptome data represent a comprehensive molecular resource for the adult head of *M. separate*, and these identified genes can be valid targets for further gene function research to address the molecular mechanisms regulating the migratory and olfaction genes of the oriental armyworm.

The oriental armyworm, *Mythimna separate* Walker (Lepidoptera: Noctuidae), is a polyphagous, migratory pest of high economic importance and can attack >300 plant species in nearly 100 families of crops, such as corn (*Zea mays*), sorghum (*Sorghum bicolor*), and rice (*Oryza sativa*). This pest is widely distributed between latitudes 45°N and 45°S and between longitudes 60°E and 170°W and has been documented in China, Russia, Japan, India, Eastern Australia, New Zealand, and some parts of the Pacific Islands. China, India, and Australia have experienced periodic outbreaks of oriental armyworm infestation¹. Larvae of *M. separate* feed mainly on leaves, and the older larvae of 4–6 instars cause the main harm, typically leaving only the midrib uneaten. This results in heavy losses. *M. separate* overwinters during the pupa stage in the soil but also sometimes during the larval or adult stage. The moth of *M. separate* can migrate up to 1500 km. The moth undertakes a seasonal, long-distance, multi-generation roundtrip migration between southern and northern China each year². To date, chemical control is still the major strategy to protect crops from damage by the oriental armyworm, leading to the development of resistance to many chemical insecticides. This species has been known to be a model migratory pest moth and is used as a test insect to explore new agricultural pesticides³.

¹Insect Resource Center for Engineering and Technology of Liaoning Province, College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, 110866, China. ²School of Basic Medicine, Guangzhou University of Chinese Medicine, Guangzhou, 510006, China. ³State Key laboratory for Biology of Plant Diseases and Insect Pest, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, China. Hai-Xu Bian, Hong-Fang Ma, Xi-Xi Zheng, Ming-Hui Peng and Yu-Ping Li contributed equally to this work. Correspondence and requests for materials should be addressed to Y.-Q.L. (email: liuyanqun@syau.edu.cn) or X.-F.J. (email: xfjiang@ippcaas.cn)

At the start of this work, the genomic data available in public databases for *M. separate* are particularly scarce, and only 190 sequences, including 65 mRNA sequences, were accessible in the NCBI database for this economical pest. Even for the *Mythimna* genus that contains more than 288 species worldwide, only 102 mRNA sequences out of 1240 nucleotide sequences were available in NCBI. To address this issue and obtain a comprehensive understanding of *M. separate* molecular biology, more recent work based on *de novo* transcriptomic sequencing has led to the identification of insecticide resistance-related genes⁴. However, little is known regarding the underlying molecular mechanisms regulating the oriental armyworm physiological and behavioural responses.

Here, we provide a *de novo* transcriptome analysis of adult heads from *M. separate* using the Illumina HiSeq platform. An objective of this study is to characterize the transcriptome in the *M. separate* adult head. We chose the head because it is the sensory and feeding centre and supports the antennae. The brain inside of the head is the most important component of the central nervous system, playing a vital role in insect behaviour. We specifically focused on genes involved in migration, including the circadian clock and melanin synthesis pathway genes, as well as non-receptor genes related to olfaction. We provide a comprehensive list of genes related to these key processes in *M. separate*. In the monarch butterfly, *Danaus plexippus*, the circadian clock in the brain plays an important role in migration by providing the timing component of the time-compensated sun compass orientation^{5,6}. Therefore, it is possible that the circadian clock in the brain is involved in the induction of the oriental armyworm migration, as in *D. plexippus*⁷. As an important neurotransmitter, dopamine modulates various aspects of insect behaviour, such as locomotor activity, decision making, phase change, learning and memory^{8–10}, and it is also a precursor of melanin, which is a predominant insect pigment. Olfaction plays a key role in various insect behaviours, such as those related to locating suitable hosts, avoiding predators, identifying oviposition sites, and finding sexual partners¹¹. In the present study, a total of 63,499 transcripts were generated by the adult head transcriptome data. Among these transcripts, we identified 14 homologs of the genes involved in the *Drosophila* clock, 20 melanin synthesis genes, and 53 olfaction-related genes by comprehensive phylogenetic analysis. Furthermore, the adult heads from multiple individuals also allowed us to examine the simple sequence repeat (SSR) feature in *M. separate*. These genes identified here can be valid targets for further gene function research to address the molecular mechanisms regulating oriental armyworm migration and olfaction.

Results

Sequencing and *de novo* assembly. To date, there remains no published genome for *M. separate*. Thus, for this study, we *de novo*-assembled and blast-annotated the head transcriptome for this species. Twenty-eight adult heads, excluding the eyes, were collected from a pool of animals in the morning, afternoon and evening to create the cDNA library. Transcriptomic sequence data were generated using the head cDNA library and Illumina HiSeqTM2500/MiSeq technology. We acquired 27,208,038 raw reads from the head transcriptome. After removing adapters, ambiguous nucleotides and low quality sequences, 3.27 Gbp of clean sequence data in 26,128,167 clean reads remained, with a Q20 value of 95.94% and a GC content of 46.02%. These clean reads were then randomly clipped into 25-mers for sequence assembly using Trinity software. The assembly resulted in 63,499 transcripts longer than 200 bp, which were further assembled into 46,459 unigenes, with an N50 of 1,153 and mean length of 690 bp, including 8,739 unigenes larger than 1,000 bp (Table S1). Paired-end reads that do not contain ambiguous bases were deposited in the NCBI Sequence Read Archive under the accession no. SRX2430648, and the assembled sequences were deposited in the NCBI Transcriptome Shotgun Assembly (TSA) under accession no. GFC00000000, associated with Bioproject PRJNA357654.

To validate the reliability of the transcriptome sequence assembly, the nucleotide sequences of the 20 annotated unigenes from Illumina sequencing were selected to compare with those of the cDNAs obtained from Sanger sequencing (Table S2). These 20 cDNAs that were previously identified in *M. separate* by RT-PCR cloning were obtained from the NCBI nucleotide database. As we expected, the pairwise comparison exhibited a comparable alignment with 95–100% nucleotide identities. The results not only testified to the reliability of the assembly of the transcriptome and transcript annotation but also indicated that it could be useful for further research.

Functional annotation of unigenes. For functional annotation of the *M. separate* head transcriptome, we searched all 46,459 unigene sequences against the non-redundant NCBI protein database using the Blastx tool with a cut-off E-value of 10^{-5} . Using this approach, 19,973 unigenes (42.99% of all distinct sequences) returned a Blast hit in the Nr database, 12,703 unigenes (27.34%) had specific matches in the Swiss-Prot database, and 13,461 unigenes had matches in the Pfam database. In total, 22,334 (48.07%) unigenes were annotated in at least one database. Matches to the Nr database indicated that the silkworm *Bombyx mori* had the highest share of matches with 46.3%, followed by the butterfly *D. plexippus* with 30.0% (Figure S1).

Gene Ontology (GO) assignments were further employed to classify the functions of the *M. separate* unigenes. A total of 14,850 unigenes could be categorized into 59 functional groups (Fig. 1a). Biological processes, molecular functions, and cellular components were associated with 37,161 unigenes, 18,583 unigenes, and 26,024 unigenes, respectively. Within the biological process category, 'cellular process' (8,259 unigenes; 22.22%), 'metabolic process' (7,506 unigenes; 20.20%), and 'single-organism process' (6,434 unigenes; 17.31%) were the most abundant groups, whereas 'biological phase' (12 unigenes), 'rhythmic process' (8 unigenes), and 'hormone secretion' (5 unigenes) were the least abundant groups. In the cellular component category, the 'cell' (5181 unigenes; 19.91%) and 'cell part' (5179 unigenes; 19.90%) categories were highly represented, followed by the 'organelle' (3424 unigenes; 13.16%), 'macromolecular complex' (3030 unigenes; 11.64%), and 'membrane' (2687 unigenes; 10.32%) categories, whereas 'nucleoid' (2 unigenes), 'collagen trimer' (1 unigene), and 'symplast' (1 unigene) were the least abundant categories. In terms of molecular functions, 'binding' (8,637 unigenes; 46.48%) and 'catalytic activity' (6,119 unigenes; 32.93%) were enriched, whereas 'receptor regulator activity' (5 unigenes), 'metallochaperone activity' (4 unigenes), and 'translation regulator activity' (2 unigenes) were the least abundant categories.

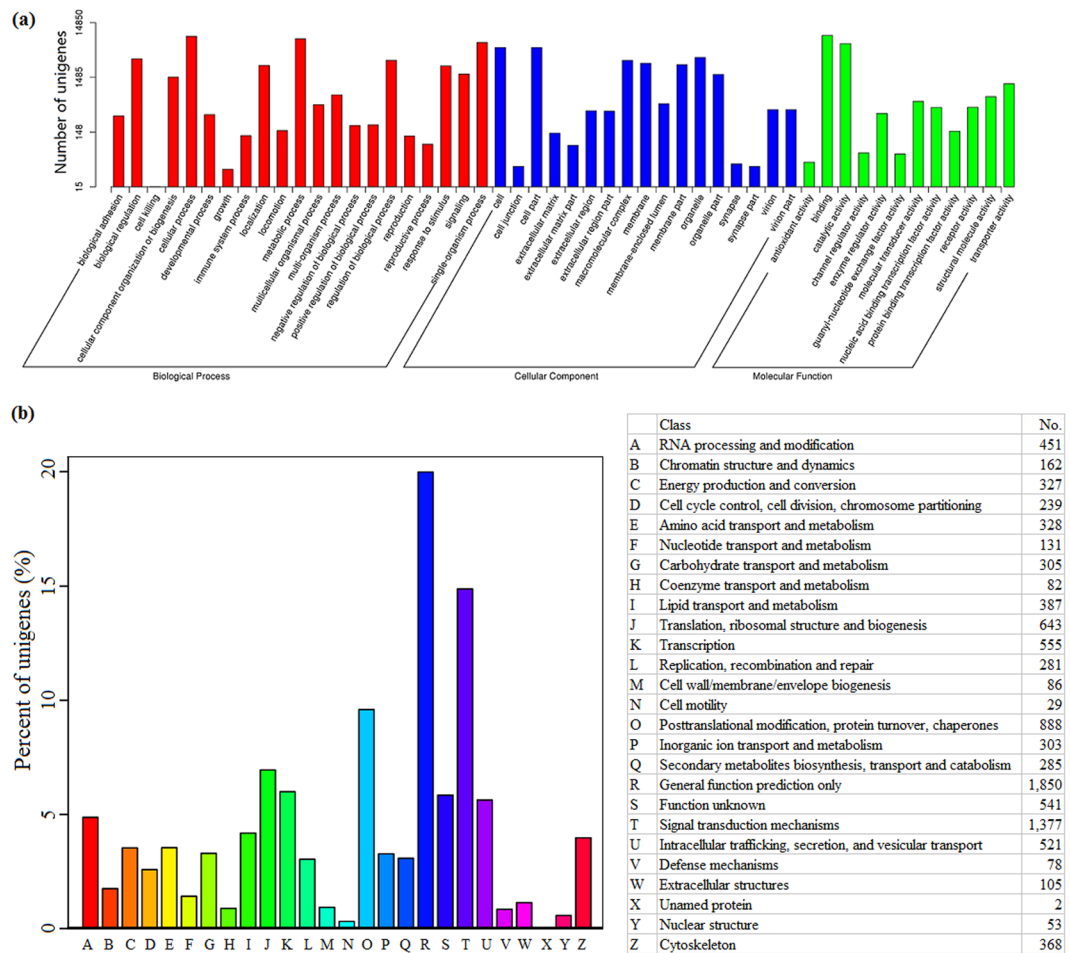


Figure 1. GO classification (a) and KOG categories (b) for the *M. separate* head transcriptome. The number of unigenes is shown on the right column for KOG categories.

The unigenes of *M. separate* were also characterized by KOG (euKaryotic Ortholog Groups) to enable conceptualization of its transcripts into potential functional groups. In total, 9,256 unigenes were annotated to 26 KOG categories (Fig. 1b). The KOG classification indicated that except 'general function prediction', genes involved in 'signal transduction mechanisms' (14.88%), 'post translational modification, protein turnover, chaperones' (9.59%), and 'translation, ribosome structure and biogenesis' (6.95%) were the most abundant.

Next, the unigenes were mapped to reference canonical pathways in the KEGG database, and 7,552 unigenes were assigned to KEGG Orthology (KO) terms and grouped into 264 pathways. These annotated pathways were clustered into five major categories (Fig. 2a). Among these pathways, 'signal transduction pathways' constituted the largest category, which contained 937 unigenes (12.41%), followed by 'translation' (743, 9.84%) and the 'endocrine system' (505, 6.69%) category. The 937 unigenes in signal transduction were assigned to 22 signalling pathways (Fig. 2b). Of these, the P13K-Akt, cAMP, and MAPK signalling pathways were ranked first to third according to the number of KEGG assignments. In addition, the nervous system (269), environmental adaptation (134), and sensory system (101) categories were well represented.

Circadian clock genes. In the transcriptome data, fourteen homologs of the genes involved in the *Drosophila* clock were well represented, including *cryptochrome 1*, *cryptochrome 2*, *cycle*, *clock*, *vriille*, *timeless*, *slimb*, *period*, *double time*, *shaggy*, *PAR-domain protein 1*, *casein kinase 2 alpha*, *casein kinase 2 beta*, and *methoprene-tolerant* (Table S3). Ten of the 14 clock genes were full-length. The Blastx results indicated that the proteins encoded by these genes shared relatively high amino acid identities (65–98%) with lepidopteran species. We further confirmed these genes by comparing them with other known insect clock genes using phylogenetic analysis (Fig. 3). The RPKM values of these genes were also evaluated and, generally, *shaggy* was the most abundant.

Melanin synthesis genes. In this study, several transcript-encoding enzymes involved in the melanin synthesis pathway were identified in the *M. separate* transcriptome, including tyrosine hydroxylase, DOPA decarboxylase, arylalkylamine N-acetyl transferase, NBAD hydrolase, NBAD synthase, GTP cyclohydro-lase, prophenoloxidase and laccase (Table S3). Two isoforms are found for genes encoding DOPA decarboxylase, arylalkylamine N-acetyl transferase, and prophenoloxidase. All the genes, except *prophenoloxidase 1*, are full-length. We also obtained all the yellow family genes in the *M. separate* transcriptome, excluding *yellow-e*,

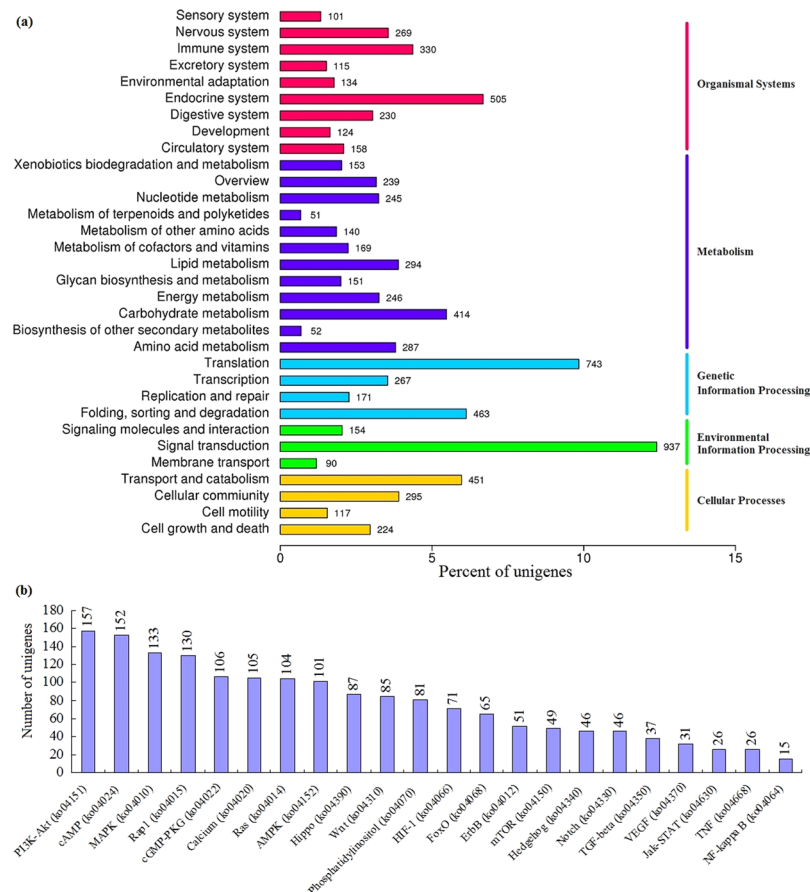


Figure 2. KEGG Orthology (KO) classification (a) and third-tier KEGG pathways in the signal transduction category (b) for *M. separate* head transcriptome. The number of unigenes is shown in the column. P13K-Akt exhibits the highest number of assigned unigenes in the signal transduction category.

and the protein of each identified gene contained a conserved major royal jelly protein domain (Table S3). Six of the 8 yellow genes are full-length. The *yellow-x* gene showed the highest expression level in the yellow family genes. All these melanin synthesis genes were confirmed by phylogenetic analysis (Fig. 4a and b). The Blastx results indicated that the proteins encoded by these genes shared relatively high amino acid identities (51–98%) with lepidopteran species. These genes identified in *M. separate* represented almost all known enzymes involved in melanin synthesis in *Drosophila*. A brief schematic based on the melanin synthesis pathway in *Drosophila*¹² is shown in Fig. 4c.

Olfactory genes. Due to the high sequence diversity of olfactory genes, their identification has largely only been possible with insects for which genomic data are available. It is very difficult to identify these genes by traditional homology-based methods. Recently, advances in RNA-Seq technology have opened up the possibility for such identifications in non-model organisms^{13–15}. With respect to *M. separate*, only one pheromone-binding protein and three olfactory receptor genes have been identified. In this study, three non-receptor protein families involved in the detection of volatile substances were identified, including odorant binding proteins (OBPs), chemosensory proteins (CSPs), and sensory neuron membrane proteins (SNMPs) in the *M. separate* transcriptome (Table S3). We confirmed these olfactory genes by phylogenetic analyses based on the alignment of protein sequences from four lepidopteran species, including *M. separate*, *Spodoptera exigua*, *Dendrolimus kikuchii*, and *B. mori*.

We identified two SNMPs (SNMP1 and SNMP2) in *M. separate*. Both were full-length genes. The SNMPs of *M. separate* were grouped with orthologues from other insect species (Fig. 5a) and shared more than 88% sequence identity with lepidopteran species.

A second group of non-receptor proteins is CSPs. We identified 18 transcripts encoding candidate CSPs in the *M. separate* transcriptome, a number comparable to 16 CSPs in *B. mori*¹⁶, 17 in *D. kikuchii*¹⁴ and *S. exigua*. Fifteen of 18 CSPs likely represent full-length proteins. The Blastx results indicated the 17 MsCSPs shared relatively high amino acid identities (59–99%) with lepidopteran CSPs. All the full-length amino acid sequences possessed a signal peptide and highly conserved four-cysteine profile (Figure S2). Although no obvious branches could be defined in the phylogenetic tree, five orthologues (1:1:1:1) could be found among four species (Fig. 5b).

A third class of non-receptor proteins identified in *M. separate* is the OBPs. A total of 38 candidate OBP genes were identified through Blastx analyses. A conserved domain search (CD Search) was further performed to

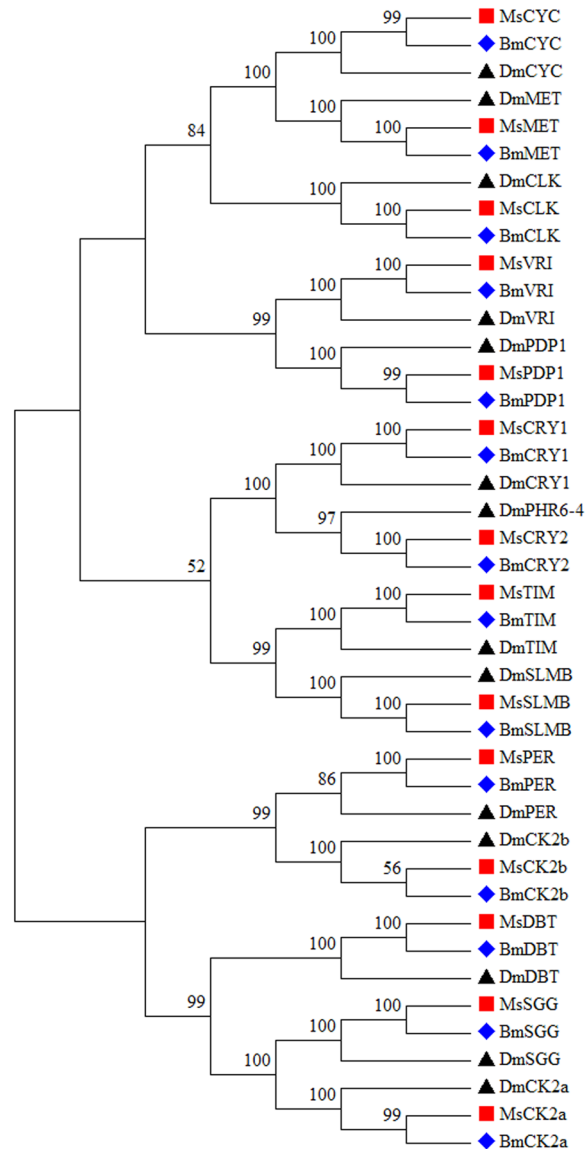


Figure 3. Neighbour-joining tree based on the amino acid sequences of the clock genes. CYC, cycle; MET, methoprene-tolerant; CLK, clock; VRI, vrille; PDP1, PAR-domain protein 1; CRY, cryptochrome; TIM, timeless; SLMB, slimb; PER, period; CK2b, casein kinase 2 beta; DBT, double time; SGG, shaggy; CK2a, casein kinase 2 alpha.

confirm the superfamily of the candidate OBPs. Through a CD search, 5 candidates that contained the conserved JHBP domain (cl12117) were excluded, although each of them had a high sequence similarity at the protein level with the odorant binding proteins. Therefore, we identified 33 OBP genes in the *M. separate* transcriptome data, a number comparable to the 34 OBPs in *S. exigua*. Thirty of the 33 OBPs genes have intact ORFs, and the remaining three (*MsPBP2*, *MsPBP3*, *MsOBP9*) lack the 3' end. The signal sequences are present at the hydrophobic N-terminus for all 33 OBPs. The Blastx results indicated that the 33 MsOBPs shared relatively high amino acid identities (33–98%) with lepidopteran OBPs. Both MsOBP25 and MsOBP28 have longer protein sequences than others, but each of them exhibits 79% and 69% identities with the OBP9 of *Spodoptera litura* (ALD65883) and OBP10 of *Ostrinia furnacalis* (BAV56797) and are associated with a query coverage of 100% and 91%, respectively. In the present study, six subfamilies of OBPs were defined based on their sequence comparison (Figure S2) and phylogenetic analysis (Fig. 6a). Following the description in *B. mori*¹⁶ and *D. melanogaster*¹⁷, we also named six subfamilies PBP-GOBP, CRLBP, ABP-I, ABP-II, Plus-C and Minus-C. The spacing pattern of the conserved six-cysteines in the *M. separate* OBP family is similar to those in *B. mori* and *D. melanogaster*. MsOBP18, MsOBP19 and MsOBP20 belong to the Minus-C subfamily, which is missing the conserved cysteines C2 and C5. Seven MsOBPs are defined as the Plus-C subfamily, but four of these do not have the PBP-GOBP motif (cl11600) when blasted in the CD Search (MsOBP22, MsOBP24, MsOBP26, MsOBP27). The conserved C2 and C3 of the four Plus-C OBPs (MsOBP22, MsOBP23, MsOBP24, MsOBP27) are separated by 4 amino acid residues rather than the usual 3 of the other OBPs, as observed in *S. litura* OBPs¹⁸. In the phylogenetic tree, many

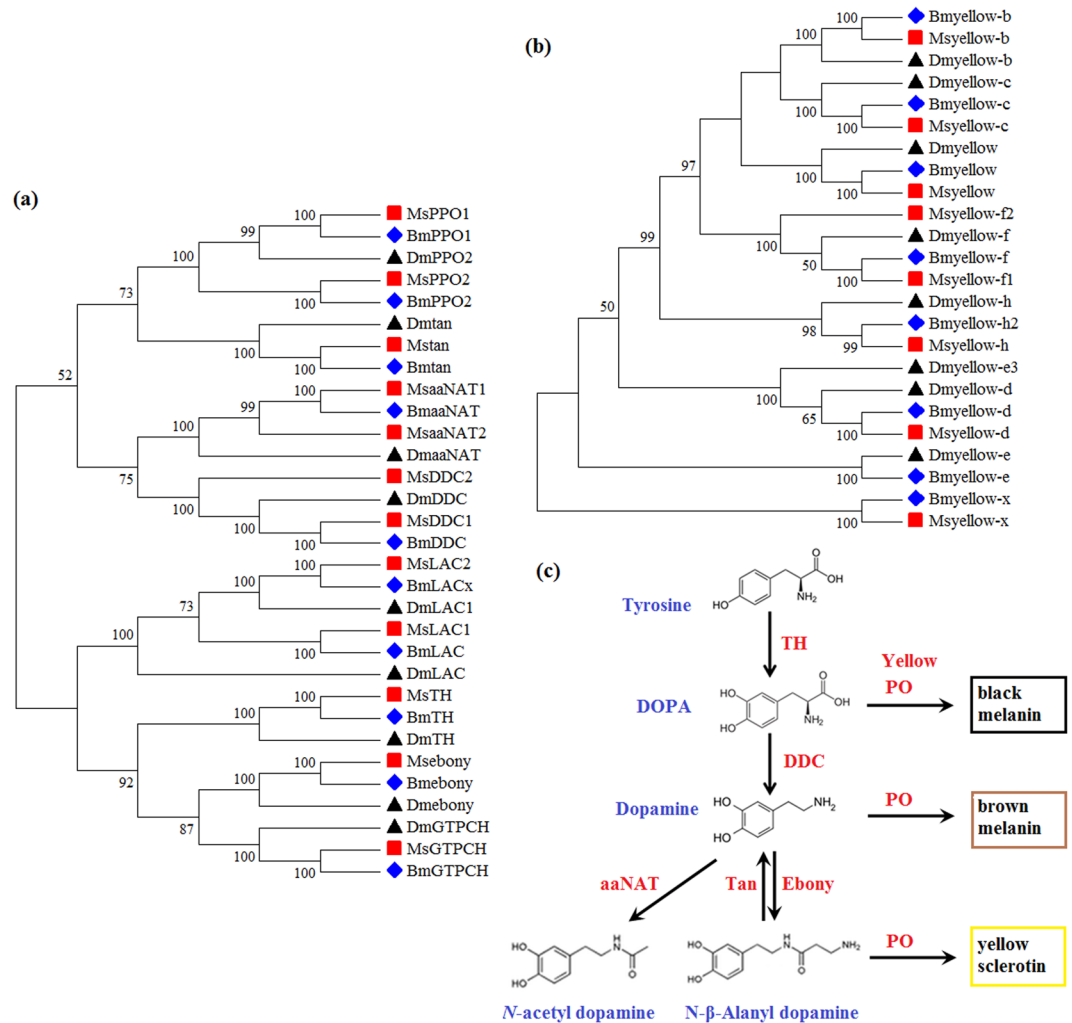


Figure 4. Neighbour-joining tree based on the amino acid sequences of melanin synthesis/related genes (a,b) and a schematic representation of the melanin synthesis pathway (c) found in *M. separate*. PPO, prophenoloxidase; tan, NBAD hydrolase; aaNAT, arylalkylamine N-acetyl transferase; DDC, DOPA decarboxylase; LAC, laccase; TH, tyrosine hydroxylase; ebony, NBAD synthase; GTPCH, GTP cyclohydrolase.

terminal relationships and four subfamilies (PBP-GOBP, Minus-C, ABP-I and ABP-II) were supported by high bootstrap values. Although the four lepidopteran species covered in this study are closely related, only twelve orthologues (1:1:1:1) could be found among the four species.

SSR discovery. The identification of microsatellite polymorphisms is helpful for population genetics of this pest. Presently, no SSR markers are available for this species, although a preliminary information analysis of the SSR loci in *M. separate* has been reported¹⁹. In that study, only 400 potential SSR loci were identified in 372 unigenes. As we used 28 individual moths to construct the cDNA library, certain polymorphism levels are expected to be present in the transcriptome data. Here, the 46,459 unigenes assembled in this study were analysed to mine potential SSRs using the MISA package. A total of 5,267 potential SSRs were identified in 4,499 unigene sequences, of which 604 sequences contained more than 1 SSR. Among these SSRs, 228 were represented in compound formation, and 3895 were mononucleotide SSRs. As it is difficult to distinguish true mononucleotide repeats from polyadenylation sites, we did not include mononucleotide repeats in the following analysis. Therefore, 1,372 SSRs of 2–6 bp in unit length were identified (Table 1), which suggested a frequency of about one SSR per 23.35 kb of expressed sequences. Among all SSRs identified, trinucleotide repeats (63.78%) represented the most abundant microsatellite repeat units, followed by dinucleotide repeats (32.87%). Ten types of trinucleotide repeats were observed. Of these, CCG/CGG was the most abundant, followed by ATC/ATG and AAT/TTA. The CG/GC sequence was the most common among the dinucleotide repeat motifs, followed by AC/GT, AT/TA, and AG/CT. The AAAT/TTTA sequence was the most abundant in the tetranucleotide repeats. Most SSRs (48.91%) exhibited a repeat number of 5, and 37.24% of the SSRs had a repeat number of 6. The SSRs predicted in this study could lay a platform for better understanding the molecular ecology of this species. The primer information for the potential SSRs should be sent as a query.

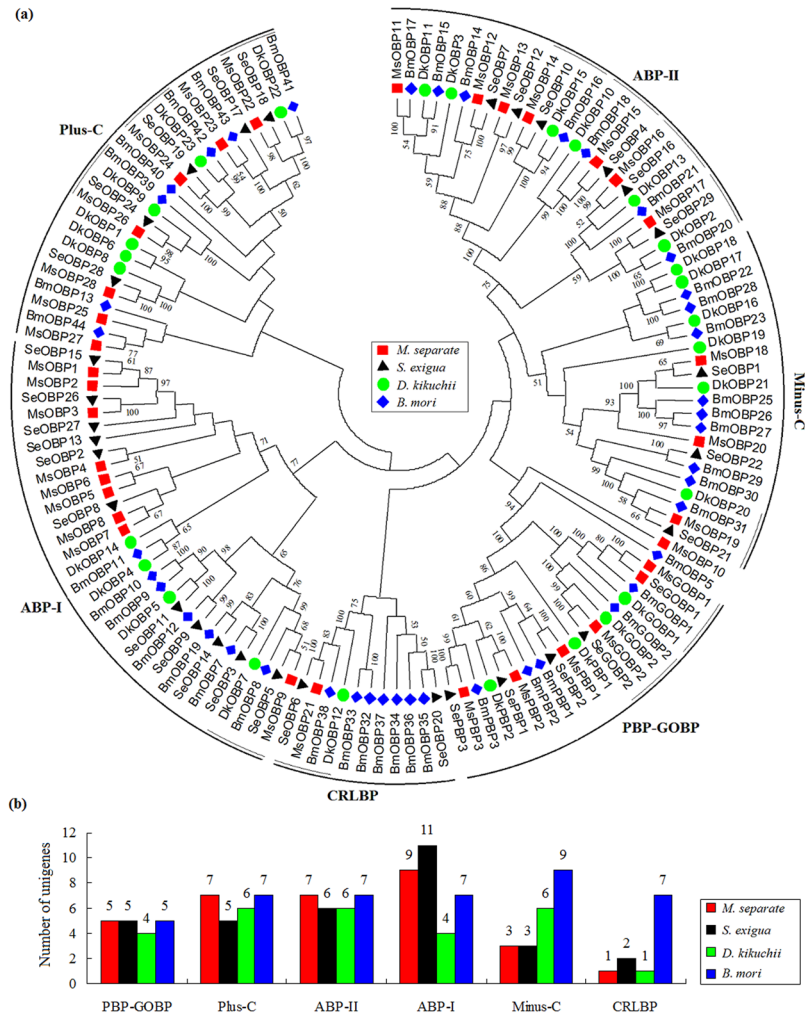


Figure 6. Neighbour-joining tree (a) and gene number comparison (b) of candidate OBP proteins among four lepidopteran species. Six groups, including PBP-GOBP, CRLBP, ABP-I, ABP-II, Plus-C and Minus-C, are marked on the dendrogram. Twelve orthologues (1:1:1:1) among four lepidopteran species are also marked.

To date, no studies have focused on the circadian clock of *M. separate*, although two clock genes, *cryptochrome 1* and *2*, have previously been isolated. We hypothesized that the circadian clock might be involved in the induction of oriental armyworm migration, as was found in *D. pleippus*⁷. A nearly complete set of clock genes identified in this study indicated the presence of the clock in the head of *M. separate*. The identified genes provide a template for exploring how the circadian clock affects the future migration of the oriental armyworm.

Insect melanism is one of the common polymorphisms in nature. This was also reported in *M. separate*²⁸. Adults of the melanic form are almost completely darkened in comparison with those of the typical form, and the inheritance of melanic characteristics follows a Mendelian law. Although the melanic population possesses many developmental and reproductive advantages, it exhibits a reduced migration activity compared with the typical population²⁹. To date, knowledge on the molecular mechanism controlling melanism in the oriental armyworm is very limited. Previous studies have indicated that insect melanism is tightly associated with the melanin synthesis pathway^{12, 30}. The melanin synthesis genes identified here will contribute to further functional research on the molecular mechanism controlling melanism and influencing migration in the oriental armyworm.

Olfaction is critical for insects because of its use in most key insect behaviours. SNMPs, CSPs and OBPs are three members of the non-receptor olfactory gene families. Among the three olfactory gene families, SNMPs and CSPs are more conserved than OBPs^{14, 31}, which was also confirmed by our results. We identified comparable numbers of CSPs with those found in *B. mori*, *D. kikuchii* and *S. exigua*. Generally, SNMPs have two orthologues in insects³², and we also found both in *M. separate*. The exact functions of OBPs are still unknown, although their most important function is suggested to be involved in the capture and delivery of outside odorants to the odorant receptors. In this study, we identified 33 OBP genes from the *M. separate* head transcriptome, of which 32 *MsOBPs*, except *MsPBP1*, are reported here for the first time. The number of *M. separate* OBPs identified in this study was comparable with the numbers from the antennal transcriptomes of *Agrotis ipsilon* (33)³³, *Spodoptera littoralis* (36)³⁴, and *S. litura* (38)¹⁸. The topology of OBPs in our phylogenetic tree is largely consistent with the phylogenetic relationship established from six lepidopteran species, including *B. mori*, *A. ipsilon*, *Helicoverpa*

SSR type/Repeats	Repeat numbers							Total	Percent %
	5	6	7	8	9	10	11~24		
Dinucleotide								451	32.9
AC/GT		88	25	17	9	4	5	148	
AG/CT		32	10	5	4	1		52	
AT/TA		60	14	7	10	2		93	
CG/CG		143	15					158	
Trinucleotide								875	63.8
AAC/GTT	21	5	1					27	
AAG/CTT	80	23		2				105	
AAT/TTA	136	17	7	2				162	
ACC/GGT	18	5	2					25	
ACG/CGT	25	9	2	1				37	
ACT/AGT	6	2						8	
AGC/CTG	32	8	3	1			1	45	
AGG/CCT	7	1	1					9	
ATC/ATG	116	43	14	2	1			176	
CCG/CGG	192	69	17	3				281	
Tetranucleotide								43	3.13
AAAC/GTTT	2							2	
AAAG/CTTT	7	1						8	
AAAT/TTTA	17	3						20	
AACC/GGTT	1							1	
AATC/ATTG	3							3	
ACAG/CTGT	2							2	
ACAT/ATGT	1				1			2	
ACCT/AGGT	1			1				2	
ACTC/AGTG	1							1	
ACTG/AGTC		1						1	
AGCG/CGCT	1							1	
Pentanucleotide								2	0.15
AAATC/ATTG	1							1	
AATCG/ATTCG	1							1	
Hexanucleotide								1	0.07
AGCCTG/AGGCTC		1						1	
Total	671	511	111	41	25	7	6	1372	
Percent %	48.91	37.24	8.09	2.99	1.82	0.51	0.44		100

Table 1. Repeat numbers and unit length distribution of SSRs in the transcriptome of *M. separata* unigenes.

armigera and three *Spodoptera* species¹⁸. For the OBP numbers, subfamilies of PBP-GOBP, Plus-C and ABP-II are comparable among the four species, whereas each of the Minus-C and CRLBP subfamilies show lineage-specific expansion and diversification in *B. mori* (Fig. 7b). The increase of OBP number in the two subfamilies in *B. mori* may be related to its long-term artificial selection or different adaptation responses. These genes could be valid targets for further gene function research. Further research on the molecular mechanism of olfaction in *M. separata* based on the data in this study will be helpful for the more efficient control of this pest.

Materials and Methods

Insects. A *M. separata* colony has been reared continuously in the laboratory of the pest group at the Institute of Plant Protection, Chinese Academy of Agricultural Sciences (Beijing, China). The larvae were reared on fresh maize leaves at $25 \pm 2^\circ\text{C}$ and $70 \pm 5\%$ relative humidity under a 12:12 light:dark photoperiod. A total of 28 adult heads, excluding eyes, were collected for construction of the cDNA library. Heads were collected in the morning (10), afternoon (9) and evening (9). Then, head samples were immediately frozen in liquid nitrogen and subsequently stored at -80°C until use.

RNA isolation and sequencing library preparation. Frozen head samples were shipped to Novogene (Beijing, China) for RNA isolation and library construction. Total RNA isolation was extracted from the heads of *M. separata* using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). RNA degradation and contamination was monitored on 1% agarose gels. RNA purity was checked with a NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using a Qubit RNA Assay Kit with a Qubit 2.0 Fluorometer (Life Technologies, CA, USA). The RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used to assess the RNA integrity. A total amount of $3\mu\text{g}$ of RNA was used as input

material for the RNA sample preparations. Sequencing libraries were generated with NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB, USA). To select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). The size-selected, adaptor-ligated cDNA fragments were enriched by PCR with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. Lastly, PCR products were purified (AMPure XP system), and the quality of the library was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, CA, USA).

de novo assembly and functional annotation. The cDNA library was sequenced on an Illumina HiSeq 2500 platform, and paired-end reads of 2×125 bp in size were generated. The *de novo* assembly pipeline for the *M. separate* head transcriptome was outlined in Figure S3. Raw fastq data (raw reads) were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing the adaptor, ploy-N and low quality reads from the raw data. At the same time, Q20, Q30, GC-content and the sequence duplication level of the clean data were calculated. The left files (read1 files) from all the libraries/samples were pooled into one big left.fq file, and the right files (read2 files) into one big right.fq file. Transcriptome assembly was accomplished based on the left.fq and right.fq files using Trinity³⁵ with a Kmer_length of 25 and min_kmer_cov set to 2, and all other parameters set to default. Function annotation was performed based on the following databases: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins), Swiss-Prot (A manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), and GO (Gene Ontology).

Sequence analysis. The similarity searches were performed with the NCBI-Blast network server (<http://blast.ncbi.nlm.nih.gov/>). Putative N-terminal signal peptides of proteins were predicted by the Signal IP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>)³⁶. The detection of the conserved protein domains was carried out with a batch CD-search tool in NCBI³⁷. The expression abundance of the unigene was calculated based on the reads per kilobase per million mapped reads (RPKM) method³⁸.

Phylogenetic analysis. The accession numbers of sequences used for phylogenetic analysis are listed in Table S4. For phylogenetic analysis of migration-related genes, we included *M. separate* (Ms), and two model insects, *B. mori* (Bm) and *Drosophila melanogaster* (Dm). For olfactory genes, we included four lepidopteran species, *M. separate*, *B. mori*, *D. kikuchii* (Dk) and *S. exigua* (Se). Amino acid sequences were aligned with ClustalX 1.83³⁹, and unrooted trees were constructed with MEGA6.0⁴⁰ using the neighbour-joining method, with Poisson correction of distances and bootstrap replications set at 1000.

SSR detection. Picard - tools v1.41 and samtools v0.1.18 were used to sort, remove duplicated reads and merge the bam alignment results. SSRs were identified using MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>), and primers for each SSR were designed using Primer3 (<http://primer3.sourceforge.net/releases.php>).

References

- Wang, G. P., Zhang, Q. W. & Ye, Z. H. The role of nectar plants in severe outbreaks of armyworm *Mythimna separata* (Lepidoptera: Noctuidae) in China. *Bull. Entomol. Res.* **96**, 445–455 (2006).
- Jiang, X. F., Luo, L. Z., Zhang, L., Sappington, T. W. & Hu, Y. Regulation of migration in *Mythimna separata* (Walker) in China: A review integrating environmental, physiological, hormonal, genetic, and molecular factors. *Environ. Entomol.* **40**, 516–533, doi:10.1603/EN10199 (2011).
- Fan, L., Guo, Y., Zhi, X., Yu, X. & Xu, H. Stereoselective synthesis of 2 α -chloro-picropodophyllotoxins and insecticidal activity of their esters against oriental armyworm, *Mythimna separata* Walker. *J. Agric. Food Chem.* **62**, 3726–3733, doi:10.1021/jf405316w (2014).
- Liu, Y., Qi, M., Chi, Y. & Wuriyanghan, H. De novo assembly of the transcriptome for oriental armyworm *Mythimna separata* (Lepidoptera: Noctuidae) and analysis on insecticide resistance-related genes. *J. Insect Sci.* **16**, 92, doi:10.1093/jisesa/iew079 (2016).
- Perez, S. M., Taylor, O. R. & Jander, R. A sun compass in monarch butterflies. *Nature* **387**, 29–29, doi:10.1038/387029a0 (1997).
- Sauman, I. *et al.* Connecting the navigational clock to sun compass input in monarch butterfly brain. *Neuron* **46**, 457–467, doi:10.1016/j.neuron.2005.03.014 (2005).
- Zhu, H., Casselman, A. & Reppert, S. M. Chasing migration genes: A brain expressed sequence tag resource for summer and migratory monarch butterflies (*Danaus plexippus*). *Plos One* **3**, e1345, doi:10.1371/journal.pone.0001345 (2008).
- Draper, I., Kurshan, P. T., McBride, E., Jackson, F. R. & Kopin, A. S. Locomotor activity is regulated by D2-like receptors in *Drosophila*: an anatomic and functional analysis. *Dev. Neurobiol.* **67**, 378–393, doi:10.1002/dneu.v67:3 (2007).
- Ma, Z. Y., Guo, W., Guo, X. J., Wang, X. H. & Kang, L. Modulation of behavioral phase changes of the migratory locust by the catecholamine metabolic pathway. *Proc. Natl. Acad. Sci. USA* **108**, 3882–3887, doi:10.1073/pnas.1015098108 (2011).
- Xu, G. *et al.* De novo assembly and characterization of central nervous system transcriptome reveals neurotransmitter signaling systems in the rice striped stem borer, *Chilo suppressalis*. *BMC Genomics* **16**, 525, doi:10.1186/s12864-015-1742-7 (2015).
- Leal, W. S. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu. Rev. Entomol.* **58**, 373–391, doi:10.1146/annurev-ento-120811-153635 (2013).
- Wittkopp, P. J. & Beldade, P. Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Semin. Cell Dev. Biol.* **20**, 65–71, doi:10.1016/j.semcdb.2008.10.002 (2009).
- Grosse-Wilde, E. *et al.* Antennal transcriptome of *Manduca sexta*. *Proc. Natl. Acad. Sci. USA* **108**, 7449–7454, doi:10.1073/pnas.1017963108 (2011).
- Zhang, S., Zhang, Z., Wang, H. & Kong, X. Antennal transcriptome analysis and comparison of olfactory genes in two sympatric defoliators, *Dendrolimus houi* and *Dendrolimus kikuchii* (Lepidoptera: Lasiocampidae). *Insect Biochem. Mol. Biol.* **52**, 69–81, doi:10.1016/j.ibmb.2014.06.006 (2014).
- Leitch, O., Papanicolaou, A., Lennard, C., Kirkbride, K. P. & Anderson, A. Chemosensory genes identified in the antennal transcriptome of the blowfly *Calliphora stygia*. *BMC Genomics* **16**, 255, doi:10.1186/s12864-015-1466-8 (2015).
- Gong, D. P., Zhang, H. J., Zhao, P., Xia, Q. Y. & Xiang, Z. H. The odorant binding protein gene family from the genome of silkworm *Bombyx mori*. *BMC Genomics* **10**, 332, doi:10.1186/1471-2164-10-332 (2009).
- Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J. & Tanouye, M. A. Genomewide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* **12**, 1357–1369, doi:10.1101/gr.239402 (2009).

18. Gu, S. H. *et al.* Identification and comparative expression analysis of odorant binding protein genes in the tobacco cutworm *Spodoptera litura*. *Sci. Rep* **5**, 13800, doi:10.1038/srep13800 (2015).
19. Hu, Y. H. *et al.* The information analysis of SSR loci in the *Mythimna separate* (Walker) transcriptome. *J. Shanxi Agric. Univ. (Natural Science Edition)* **35**, 484–489 (2015).
20. Zhang, J. *et al.* Antennal transcriptome analysis and comparison of chemosensory gene families in two closely related noctuid moths, *Helicoverpa armigera* and *H. assulta*. *PLoS One* **10**, A840 (2015).
21. Nascimento, A. R. B. D. *et al.* Comparative transcriptome analysis of lufenuron-resistant and susceptible strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *BMC Genomics* **16**, 1–12 (2015).
22. Li, L. *et al.* An analysis of the *Athetis lepigone* transcriptome from four developmental stages. *PLoS One* **8**, e73911, doi:10.1371/journal.pone.0073911 (2013).
23. Das, S. *et al.* Transcriptome analysis of the molting gland (Y-organ) from the blackback land crab, *Cecarcinus lateralis*. *Comp. Biochem. Physiol. Part D* **17**, 26–40, doi:10.1016/j.cbd.2015.11.003 (2016).
24. Tian, C. *et al.* Characterization of *Adelphocoris suturalis* (Hemiptera: Miridae) transcriptome from different developmental stages. *Sci. Rep* **5**, 11042, doi:10.1038/srep11042 (2015).
25. Liu, Q. N. *et al.* Characterization of the complete mitochondrial genome of the oriental armyworm, *Mythimna separata* (Lepidoptera: Noctuidae). *Eur. J. Entomol.* **112**, 399–408, doi:10.14411/eje.2015.055 (2015).
26. Rezende, V. B. *et al.* Head transcriptomes of two closely related species of fruit flies of the *Anastrepha fraterculus* group reveals divergent genes in species with extensive gene flow. *G3* **6**, 3283–3295, doi:10.1534/g3.116.030486 (2016).
27. Huang, Q. Y. *et al.* Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus* (Shiraki). *PLoS One* **7**(11), e150383, doi:10.1371/journal.pone.0050383 (2012).
28. Liu, H. B. & Luo, L. Z. Morphological characteristics and inheritance of the melanic form of the oriental armyworm, *Mythimna separata* (Walker) (Lepidoptera:Noctuidae). *Acta Entomol. Sin* **47**, 287–292 (2004).
29. Jiang, X. F., Luo, L. Z. & Zhang, L. Relative fitness of near isogenic lines for melanic and typical forms of the oriental armyworm, *Mythimna separata* (Walker). *Environ. Entomol.* **36**, 1296–1301, doi:10.1093/ee/36.5.1296 (2007).
30. Koch, P. B., Behnecke, B. & Ffrench-Constan, R. H. The molecular basis of melanism and mimicry in a swallowtail butterflies. *Curr. Biol* **10**, 591–594, doi:10.1016/S0960-9822(00)00494-2 (2000).
31. Vieira, F. G. & Rozas, J. 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, doi:10.1093/gbe/evr033 (2011).
32. Nichols, Z. & Vogt, R. G. The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* **38**, 398–415, doi:10.1016/j.ibmb.2007.11.003 (2008).
33. Gu, S. H. *et al.* Molecular characterization and differential expression of olfactory genes in the antennae of the black cutworm moth *Agrotis ipsilon*. *PLoS One* **9**(8), e103420, doi:10.1371/journal.pone.0103420 (2014).
34. Poivet, E. *et al.* A comparison of the olfactory gene repertoires of adults and larvae in the noctuid moth *Spodoptera littoralis*. *PLoS One* **8**(4), e60263, doi:10.1371/journal.pone.0060263 (2013).
35. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652, doi:10.1038/nbt.1883 (2011).
36. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786, doi:10.1038/nmeth.1701 (2011).
37. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–226, doi:10.1093/nar/gku1221 (2015). (Database issue).
38. Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628, doi:10.1038/nmeth.1226 (2008).
39. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882, doi:10.1093/nar/25.24.4876 (1997).
40. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol* **30**, 2725–2729, doi:10.1093/molbev/mst197 (2013).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (No. 31372372), the scientific research program for agricultural commonweal from Ministry of Agriculture (No. 201403031) and the China Agriculture Research System (CARS-22).

Author Contributions

Y.Q.L. and X.F.J. conceived and designed the experiments. H.X.B., H.F.M., X.X.Z. and M.H.P. performed the experiments and analysed the data. M.H.P. performed the sample collection. J.E.S., H.W., Q.L., R.X.X. contributed to the experiments, data analyses and data interpretation. Y.P.L. revised the manuscript. Y.Q.L. wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02513-6

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017