# Genome assembly and methylome analysis of the white wax scale insect provides insight into sexual differentiation of metamorphosis in hexapod

Hang Chen[1], Qin Lu[1], Xiaoming Chen[1], Xiaofei Ling[1], Pengfei Liu[1], Ni Liu[1], Weiwei Wang[1], Jinwen Zhang[1], Qian Qi[1], Weifeng Ding[1], Xin Zhang[1], Ying Feng[1], Yurong Zhang[2], Ming-Shun Chen[3], and Kirst King-jones[4]

[1]Research Institute of Resource Insects, Chinese Academy of Forestry
[2]Hunan Academy of Forestry
[3]Kansas State University
[4]University of Alberta

November 13, 2020

## Abstract

Scale insects are hemimetabolous, showing "incomplete" metamorphosis and no true pupal stage. Ericerus pela, commonly known as the white wax scale insect (hereafter, WWS), is a wax-producing insect found in Asia and Europe. WWS displays dramatic sexual dimorphism, with notably different metamorphic fates in males and females. Males develop into winged adults, while females are neotenic and maintain a nymph-like appearance, which are flightless and remain stationary. Here we report the de novo assembly and analysis of the WWS genome. From these data, we constructed a robust phylogenetic analysis of 24,923 gene families from 16 representative insect genomes, which indicates that holometabola evolved from hemimetabolous insects in the Late Carboniferous, about 50 million years earlier than previously thought. To study the distinct development of males and females, we analyzed the methylome landscape in either sex. Surprisingly, WWS displayed high levels of methylation (4.42%) when compared to other insects. We observed differential methylation patterns for genes involved in steroid and sesquiterpenoids production as well as related fatty acid metabolism pathways. We show here that both males and females produce distinct profiles of ecdysone (the principal insect steroid hormone) and juvenile hormone (a sesquiterpenoid), consistent with their different development fates. Our results provide a comprehensive genomic and epigenomic resource of scale insects that provide new insights into the evolution of metamorphosis and sexual dimorphism in insects.

### Genome assembly and methylome analysis of the white wax scale insect provides insight into sexual differentiation of metamorphosis in hexapod

Hang Chen[1]*, Qin Lu[1], Xiaoming Chen[1,2]*, Xiaofei Ling[1], Pengfei Liu[1], Ni Liu[1], Weiwei Wang[1], Jinwen Zhang[1], Qian Qi [1], Weifeng Ding[1], Xin Zhang[1], Ying Feng[1], Yurong Zhang[3], Ming-Shun Chen[4], Kirst King-Jones[5]*

[1] Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, 650224, China.

[2] The Key Laboratory of Cultivating and Utilization of Resources Insects, State Forestry Administration, Kunming, 650224, China.

[3] Hunan Academy of Forestry, Changsha, 410001, China

[4] Department of Entomology, Kansas State University, Manhattan, KS 66506, USA

1

[5] Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

* Correspondence: stuchen6481@gmail.com, Cafcxm@139.com or *kingjone@ualberta.ca*

**running title:** Genome and methylome of the white wax insects

## Abstract

Scale insects are hemimetabolous, showing "incomplete" metamorphosis and no true pupal stage.*Ericerus pela* , commonly known as the white wax scale insect (hereafter, WWS), is a wax-producing insect found in Asia and Europe. WWS displays dramatic sexual dimorphism, with notably different metamorphic fates in males and females. Males develop into winged adults, while females are neotenic and maintain a nymph-like appearance, which are flightless and remain stationary. Here we report the *de novo* assembly and analysis of the WWS genome. From these data, we constructed a robust phylogenetic analysis of 24,923 gene families from 16 representative insect genomes, which indicates that holometabola evolved from hemimetabolous insects in the Late Carboniferous, about 50 million years earlier than previously thought. To study the distinct development of males and females, we analyzed the methylome landscape in either sex. Surprisingly, WWS displayed high levels of methylation (4.42%) when compared to other insects. We observed differential methylation patterns for genes involved in steroid and sesquiterpenoids production as well as related fatty acid metabolism pathways. We show here that both males and females produce distinct profiles of ecdysone (the principal insect steroid hormone) and juvenile hormone (a sesquiterpenoid), consistent with their different development fates. Our results provide a comprehensive genomic and epigenomic resource of scale insects that provide new insights into the evolution of metamorphosis and sexual dimorphism in insects.

**Key words:** Methylome, Hi-C assembly, Hormone metabolism, RNAi, Hemimetabolous, Holometabola, Sexual differentiation, Developmental transcriptome

## 1 — INTRODUCTION

Since the launch of the 5,000 Insect Genome Project (i5k) in 2011, at least 497 insect genomes have been sequenced and annotated according to the NCBI database (Robinson et al., 2011). Most of the sequenced insect genomes are either important agricultural pests, vectors for pathogens, or species that serve as model organisms (Xianhui et al., 2014). Silkworm (*Bombyx mori* ), honeybee (*Apis cerana* ) and WWS (*Ericerus pela* ) are the top three resource insects that have provided significant economic or medical benefits to humans for millennia. The genomes of both honeybee and silkworm have been sequenced and extensively annotated, whereas high-quality genomic data for WWS is still absent so far (Consortium, 2006; Xia et al., 2004). Found in the subtropical and temperate regions of China, Japan and Russia (Supplementary Fig. 1), WWS-derived white wax has been extensively used in printing, pharmaceutical, food and cosmetic industries because of its special chemical properties (Ma et al., 2018; Wang et al., 2017).

The evolutionary origin of the insect metamorphosis is a much-debated mystery in biology. Wing development in insects occurs during metamorphosis, however, the evolutionary drivers led up to the invention of metamorphosis are still poorly understood (Misof et al., 2014). Juvenile hormone mediates metamorphosis in a range of insect species (Azinna et al., 2018; Stillwell et al., 2010). Research on firebugs has shown how juvenile hormone counteracts ecdysone during the nymph instar in hemimetabolous insects (Barbora et al., 2011; Slama & Williams, 1965). During the evolution of metamorphosis, a heterochronic shift in the time of the appearance of juvenile hormone during embryogenesis appears to have interrupted the ancestral growth trajectory in the developing insect (Gilbert, 1994). Previous studies noted that juvenile hormone plays a crucial role in mediating neoteny of different insect species. Scale insects such as WWS convergently acquired a unique mode of metamorphosis referred to as neometaboly, which is reminiscent of holometaboly (Vea et al., 2016). WWS males produce wax during the nymph stages and form winged adults, resembling a holometabolous fate (Supplementary Fig. 2). Females, however, remain stationary and appear to follow a hemimetabolous route (Qi et al., 2019). Accordingly, WWS represents an attractive research model not only to investigate the adaptive mechanisms by which a herbivorous parasite overcomes plant defense systems, but it also serves as an excellent model for the evolutionary relationships between hemimetabolous

2

and holometabolous insects. In this study, we comprehensively analyzed the developmental differences between males and females using high-quality whole genome data of WWS, transcriptome profiling and DNA methylation. Our study paves the way to understand the developmental processes during metamorphosis and the hormone antagonism effect the sexual differentiation of metamorphosis in insects.

## 2 — MATERIALS AND METHODS

### 2.1 — Samples

Samples of WWS were maintained at the Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, China (N25@3'47", E102@45'18", H1,950m). For *de novo* sequencing and analysis, we collected ~1,500 individuals of female larvae in first instar from one colony into three micro-tubes and sink in liquid nitrogen for DNA isolation and whole genome sequencing. For DNA methylation and analysis, we collected 500 female and male individuals, respectively, to make sure that the samples are sufficient for total DNA isolation and the following genetic analyses. The samples for detecting hormone and RNA extraction were collected every 3 days from first instar larvae to aging adults and stored in liquid nitrogen until used.

### 2.2 — Genomic sequencing and assembly

High quality DNA was extracted from the whole first instar female larvae by the QIAamp DNA Blood Mini kit (Qiagen, Germantown, USA). Illumina and third generation sequencing (PacBio) were used respectively. For Illumina sequencing, a library with average insert size ~350 bp was constructed and sequenced on an Illumina HiSeq 2500 platform at Novogene (Novogene, Beijing, China). For third generation sequencing, a library with average insert size ~20 Kb was constructed. The libraries were sequenced on PacBio Sequel platform with Sequel SMRT cells 1M v2 at Novogene.

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. Clean data (reads) were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality. The genome was assembled by a short-read assembly method, SOAPdenovo2 package. A de Bruijn graph was built by splitting the reads into K-mers, from the short insert size libraries (<1kb), without making use of pairing information. After a series of graph simplifications, the reads were assembled into contigs. All available paired-end reads were realigned onto the contig sequences to construct the linkage between contigs. The linkage was removed if it was supported by an unreliable weight of paired-end relationships. We used the strategy of subgraph linearization to simplify the contig linkage graph, by means of extracting unambiguously linear paths. The scaffolding process was iterated in the order of estimated insert size by our in-house scaffolder GOBOND. Finally, for filling the intra-scaffold gaps, a local assembly was performed to locate the reads in the gap region by GapCloser 1.12, with the other end uniquely mapped to the contig.

### 2.3 — Gene annotation

We used a homology-based method to predict the protein-coding gene structure. Firstly, we built a non-redundant protein database of WWS and other sepcies. Then protein sequences were aligned to the genome by using TBlastN with an E-value cutoff by 1E-5. For each blast hit, genewise was used to predict the exact gene structure in the corresponding genomic regions. Finally, RNA-seq data were mapped to genome using Tophat (version 2.0.8). Then cufflinks (version 2.1.1) (http://cufflinks. cbcb.umd.edu/) was used to assemble transcripts to gene models. We used the transcript information to revise the gene set.

Functional annotation of protein-coding genes was evaluated by BLASTP (E-value:1E-05). Protein domains were annotated by searching InterPro (V32.0) and Pfam (V27.0) databases, using InterProScan (V4.8) and Hmmer (V3.1) respectively. Gene Ontology (GO) terms for each gene were obtained from the corresponding InterPro or Pfam entry. The pathways were assigned by blast against the KEGG database, with an E-value cutoff of 1E-05. The tRNA genes were identified by tRNAscan-SE software. The rRNA fragments were predicted by aligning to the rRNA sequences database using BlastN at E-value of 1E-10. The miRNA and snRNA genes were predicted by INFERNAL software against the Rfam database.

## 2.4 — Chromosome conformation capture sequencing (Hi-C)

The Hi-C libraries were sequenced on the Illumina HiSeq X Ten platform (San Diego, CA, United States) with 150PEmode. Subsequently, the Hi-C sequencing data were aligned to the assembled scaffolds by BWA-mem50 (bwa-0.7.17) and the scaffolds were clustered onto chromosomes with LACHESIS (*http://shendurelab.github.io/LACHESIS/*), the final genome was 638.30Mb and the contig and scaffold N50 were 631.79Mb and 69.68Mb, respectively (Tables 1 and Supplementary Tables 9).

## 2.5 — Identification of contracted and expanded gene families

Gene families were constructed through a hierarchical clustering algorithm and 'all against all' BLASTP. The alignments with high-scoring segment pairs (HSPs) were conjoined for each gene pair by solar. To identify homologous gene-pairs, more than 30% coverage of the aligned regions in both homologous genes was required. We determined the expansion and contraction of orthologous gene families by comparing the cluster size differences between the ancestor and other species using the CAFÉ (Version 1.6) program. A random death model was used to study changes of gene families along each lineage of phylogenetic tree. A probabilistic graphical model (PGM) was introduced to the probability of transitions in gene family size from parent to child nod phylogeny.

## 2.6 — RNA interference

DNA fragments of candidate genes were amplified with Taq PCR Master Mix (2X, with Blue Dye) (Sangon Biotech, Shanghai, China) and inserted into the RNAi vector L4440 plamsid (Addgene, USA) following the Gateway procedure. The resulting constructs were verified by complete sequencing, then recombinant plasmids were transfected into DH5α (TianGen Biotech Co. LTD, Beijing, China). DH5α cells were cultured in SOC solid medium containing 2mg/L Amp+ (pH=7.0) for 12 hours. The cells were then transferred to a MS liquid medium for another 12 hours. Plasmid DNA for dsRNA production was isolated using E.Z.N.A.TM Plasmid Mini Kit I (Sangon Biotech, Shanghai, China), and was then transfected into the HT115 cells through Thermal Activation. HT115 cells with a dsRNA production vector were cultured in SOC solid medium containing 2mg/L Amp+ and 10ml/L Tet (PH=7.0) for 12 hours, and then were transferred into 2YT liquid medium containing 1.2 ml/L Amp and 2 ml/L Tet for obtaining large amounts of dsRNAs. After purification, dsRNA was sprayed onto early second male nymphs. Silencing of dsRNA was examined using qRT-PCR at 12, 24, 48, and 72h after RNAi treatments.

## 2.7 — Phylogenetic reconstruction

The phylogenetic tree was reconstructed by using shared single-copy genes. Protein sequences for these single-copy genes were aligned using MUSCLE, then protein sequence alignments were transformed back to CDS alignments. CDS alignments of single-copy genes were then concatenated into a super matrix. Based on this super matrix, A phylogenetic tree was constructed using the ML (maximum likelihood) algorithm as implemented in RAxML software (version 7.2.3) and in MEGA Version 5 software (Alexandros & Stamatakis, 2006; Tamura et al., 2011).

## 2.8 — RNA sequencing and transcriptome assembly

Total RNA was extracted, separately, from female nymphs of first instar (FF), female nymphs of second instar (SF), , male first instar larvae (FM), and male second instar larvae (SM) using the TRIzol kit (Life Technologies, Camarillo, USA) following the procedure provided. A total amount of 3 µg RNA per sample was used as input material for the RNA sample preparations. Libraries were constructed and sequenced according to the Illumina protocol. Sequencing libraries were generated using NEBNext® Ultra RNA Library Prep Kit for Illumina(r) (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNaseH). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into

4

blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structures were ligated to prepare for hybridization. To select cDNA fragments of preferentially, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µlUSER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality 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 (Illumia) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq 2500 platform and paired-end reads were generated.

## 2.9 — DNA Methylation

Total DNA was extracted by the methods mentioned previously. Cytosine-methylated barcodes were ligated to sonicated DNA as per manufacturer's instructions. Then DNA fragments were treated twice with bisulfite using EZ DNA Methylation-GoldTM Kit (Zymo Research). Library concentration was quantified using a Qubit(r) 2.0 Flurometer (Life Technologies, CA, USA) and quantitative PCR. Insert size was analyzed on an Agilent Bioanalyzer 2100 system. Transcript libraries were sequenced on an Illumina Hiseq 2500/4000 platform. Image analysis and base calling were performed with Illumina CASAVA pipeline. Bismark software (version 0.16.3) was used to perform alignments of bisulfite-treated reads to a reference genome (Krueger & Andrews, 2011). The reference genome was firstly transformed into bisulfite-converted version and then indexed using bowtie2 (Langmead & Salzberg, 2012). The results of methylation extractor were transformed into bigWig format for visualization using IGV browser. The sodium bisulfite non-coversion rate was calculated as the percentage of cytosine sequenced at cytosine reference positions in the lambda genome. Differentially methylated regions (DMRs) were identified using the DSS software (Hao et al., 2014; Park & Wu, 2016; Wu et al., 2015). Gene Ontology (GO) enrichment analysis of genes related to DMRs was implemented using the GOseq R package (Young et al., 2010), in which gene length bias was corrected. GO terms with corrected P-value less than 0.05 were considered significantly enriched by DMR-related genes. KOBAS software (Mao et al., 2005) was used to test the statistical enrichment of DMR related genes in KEGG pathways.

## 2.10 — Determination of hormone titers

Juvenile hormone abundance in insects was determined using a high performance liquid chromatography (HPLC) (Feyereisen & Tobe, 1981). Authentic JH-III (Sigma; 25, 50, 100, 200 and 400ng*µl$^{-1}$) was used for a standard curve (Supplementary Figure 20). Approximately 5mg females or males were homogenized in 2ml methanol/diethyl ether (1:1) for 3-5min on ice, and centrifuged for 10min at 4000rpm/min, and supernatants were collected. Supernatants were dried under high purity N2 and there then dissolved in methanol/water (75:25). The operating conditions were a $C^{18}$ column (ID 45µm, 4.6 mm x 150mm in length); a LC200 HPLC pump with an ML-425 micro injection system (with a 10 µL sampling loop); and a UV-monitor with peak monitoring at 218nm. Recording was made at a 1 mV full scale. Methanol /water mixture (75: 25) was used as the developing solvent at a 0.6ml·min$^{-1}$ flow rate and 6-7 kg.cm$^{-2}$ pressure.

Ecdysones in males and females were quantified using an enzyme-linked immunosorbent assay (ELISA) as previously described (Kelly et al., 1992). Authentic 20E (Sigma; 0.625, 1.25, 2.5, 5 and10 ng·ml$^{-1}$) was used to generate a standard curve (Supplementary Fig. 20). Approximately 5mg insects in each assay was homogenized thoroughly using a motorized blue pestle (10sec) in 200µl of methanol and centrifuged for 5 min (13.3K rpm) at room temperature. The extracts were pooled and dried under SpeedVac centrifuging conditions and were then dissolved into 50µl EIA buffer for least 2 hours or at 4 overnight. The samples were then analyzed via ELISA.

## 3 — RESULTS

### 3.1 — Genomic sequences

5

We sequenced the genome of a female WWS using a whole-genome shotgun strategy. Cleaned-up reads provided 107-fold average coverage across 88.23 Gb of assembled sequence (Supplementary Fig. 3 and Supplementary Tables 1 and 2), with N50 values of 411.09 kb and 1.25 Mb respectively, for contigs and scaffolds in the final assembly (generated using SOAPdenovo4; Supplementary Figs. 4-5 and Supplementary Tables 3-5). We identified a total of 14,020 protein-coding genes (82.7% of which were functionally classified; Figure 1a, Supplementary Figs. 6-7 and Supplementary Tables 6-8). The WWS assembly was further refined by using high throughput chromosome conformation capture (Hi-C) data, comprising 9 scaffolds with scaffold N50 of 69.68Mb (Table 1). As a result, 98.16% contigs assembly of 638.30 Mb were distributed across 9 chromosome-level pseudomolecules (Fig. 2a, b; Table 1 and Supplementary Table 9).

When we compared WWS with other insect species, we found 4615 gene families in common and 497 gene families unique to WWS (Figure 1a). Using these protein-coding genes, we generated a timescale for insect evolution (Figure 1b). Comparative analyses were carried out on gene clusters among WWS and other representative insect species along with the *Daphnia pulex* (water flea, a crustacean) as a non-insect outgroup (Figure 1b, Supplementary Table 10). A total of 24,923 gene families were identified among these species with 553 clusters as common singlets (Supplementary Fig. 7). Mobile elements comprised about 37.3% repeat content of the WWS genome and the percent of long terminal repeats (LTR) was 28.5%, much higher than other transposons (DNA, LINE and SINE) (Supplementary Fig. 8 and Supplementary Tables 11-12). In addition, we identified noncoding RNA (ncRNA) genes, including 357 miRNA, 188 tRNA, 42 rRNA, and 46 snRNA genes (Supplementary Table 13).

### 3.2 — Gene families

We identified 17 gene families that were expanded, and three families were contracted in the WWS genome when compared to the other 15 arthropod species we used in this study (Figure 2c). We analyzed the functional properties of the gene family expanded in the WWS genome and identified 497 unique gene families that might be related to adaptations consistent with plant sap consumption (Figure 1a). InterPro classification of the genes from 17 significantly expanded gene families (276 genes) showed significant enrichment of these genes in categories mainly involved in xenobiotic detoxification, including metabolism of xenobiotics by cytochrome P450 enzymes (P<0.01) and other drug-metabolizing enzymes (P<0.01) (all Fisher's exact test) (Supplementary Table 14). This finding is consistent with the idea that a parasitic lifestyle, and the concomitant exposure to toxic compounds in the plant sap, would drive the expansion of genes encoding enzymes involved in xenobiotic metabolism.

Four groups of 24 genes involved in fatty acid synthesis and transport also underwent expansion. We examined all genes that may play major roles in fatty acid and lipid metabolic pathways and compared them with those from other insect species. A total of 704 WWS genes are predicted to act in fatty acid and lipid metabolic pathways, compared to only 239 related genes in the Hessian fly genome (Supplementary Figs. 9 and Supplementary Table 14). This result suggests that the expansion of fatty acid and lipid metabolism genes is linked to the production of wax in WWS (Kunst & Samuels, 2003; Teerawanichpan et al., 2010).

Two major groups of genes were found to have been fewer members compared to other insect genomes, namely odorant receptor genes and genes encoding dynein proteins. Specifically, there are seven dynein-encoding genes, 31 genes encoding odorant receptors, and 12 genes encoding odorant binding proteins in the WWS genome while the Hessian fly genome harbors 18 dynein genes, 122 genes encoding odorant receptors, and 32 genes encoding odorant binding proteins (Supplementary Figs. 9). Dynein, a motor protein, converts the chemical energy contained in ATP into the mechanical energy of movement (Supplementary Table 15). The reduction in dynein and the olfactory system for WWS might have coincided with the recession of foraging in this species as it feeds on its host branch without movement from the second instar nymph to adult. The reduction of odorant receptor genes might be explained by its parasitic nature and sessile lifestyle of females.

The expansion of xenobiotic detoxification genes is consistent with the idea that this could lead to the ability of WWS to effectively neutralize secondary compounds present in plant phloem. Similarly, the expansion of

6

genes related to fatty acid and lipid synthesis is likely explained by the high production levels of white wax in WWS, an effective adaption for WWS to evade predators and cope with harsh environmental conditions.

### 3.3 — Methylome

Recent studies have highlighted the importance of DNA methylation for understanding insect phenotypic plasticity and biological complexity (Lyko et al., 2011; Wang et al., 2014). To investigate the potential involvement of epigenetic regulation in metamorphosis of WWS, we first performed a comparative methylome analysis for male and female first instar nymphs by whole genome bisulfite sequencing (WGBS).

We assessed the DNA methylation patterns of the WWS genome and assessed that 4.42% of genomic cytosines are methylcytosines in males and 3.90% in females (Figure 3a). There are three types of methylation sites in the WWS genome (CG, CHH, CHG), and nearly all methylcytosines were found in CG dinucleotides, which were substantially enriched in gene bodies (Supplementary Figs. 12-13, Supplementary Table 16). CG and non-CG (CHH, CHG) methylation analysis indicated a mosaic pattern that fluctuated drastically across the linkage groups (Figure 3b, Supplementary Figs. 14-18, Supplementary Table 17), with high methylated domains interspersed with regions of low methylation. Interestingly, we found that repetitive elements were highly methylated and exons had higher methylation levels than introns (Figure 3a), which is different from what has been observed from other insect species in which introns have higher levels of methylation except for in *Locusta migratoria* (Wang et al., 2014). Although WWS appears to harbor only one copy of the methyltransferase gene (*Dnmt1* ), we observed much higher methylation levels (~4.2%) in comparison with what (0.1-1.6%) have been reported from other insect species such as silkworm and honeybees (Phalke et al., 2009; Regev et al.,1998; Xiang et al., 2010) (Figure 3b, Supplementary Table 18).

Through WGBS, a total of 385,025,000 CpG sites were identified, covering 14,020 genes in the WWS genome (Supplementary Table 17). Among the methylated genes, 1,699 genes showed significant methylation differences between males and females. The differentially methylated genes were distributed throughout the genome, and were mainly classified into 14 categories, including metabolic process, developmental process and response to stimulus (Figure 3c, Supplementary Figs. 19, Supplementary Table 19). Many of these categories have been documented in various hexapoda species to have crucial roles in lipid and protein metabolism, tissue morphogenesis and organ development, which are all tied to metamorphosis in insects and other animals (Williams & Carrol, 2009). Term enrichment analysis using KEGG pathways yielded seven terms that were significantly different, including i) the renin–angiotensin system (RAS), ii) renin secretion, iii) glutathione metabolism, iv) fatty acid biosynthesis, v) starch and sucrose metabolism, vi) neuroactive ligand-receptor and vii) insulin secretion (Figure 3d). Intriguingly, the genes associated with the top seven KEGG terms all displayed higher methylation levels in males compared to females (Figure 3d, Supplementary Table 20). Our results suggest that males undergo more physiological changes in the first instar nymph, with concomitant high DNA methylation levels, while females appear to continue having the same methylation levels in different developmental stages. In addition, our data indicated that differential expression of genes with roles in hormone and energy metabolism were the basis for sexually dimorphic development. Methylation differences in muscle development genes were enriched in the first instar nymph, which may be linked with dimorphic development, particularly in wing-formation in males (which requires wing muscles) and in leg degeneration in females (which equates the loss of muscle mass) (Supplementary Figs. 2, Supplementary Table 21-22).

### 3.4 — Hormone analysis

Hormone titers of WWS males and females have been determined from nymphs to adults (Figure 4 a, c). Unlike previous results reported from other insect species, our study suggests that ecdysone (20E) antagonizes juvenile hormone (JH) during WWS metamorphosis. Throughout the life cycle, JH titers in WWS females and males appeared to display a downward trend, with minimal JH levels detectable at the adult stage (Figure 4 a, c). Ecdysone was much higher than JH in males (holometabola). Even in females (hemimetabola), ecdysone was much higher than JH except in early first larval stage (Figure 4 a, c; Supplementary Figure 10). During the second larval and adult stages, neoteny females also showed

much higher ecdysone. Our results is different from previous studies that showed JH mediates neoteny in insects (AZinna et al., 2018; Stillwell et al., 2010), indicating that there are different mechanisms in insect metamorphosis regulation.

In addition to the comparative genomic analyses, we performed a transcriptomic analysis of WWS-based RNA-seq data from nymphs of both sexes (Figure 4b). Differentially expressed genes were more pronounced in males at different stages than females when the second instar nymphs were compared with larvae (Figure 4b). This may reflect the differences in developmental regulation between males and females. A hierarchical clustering analysis of the differentially expressed genes related to hormone metabolism also showed significant differences between males and females, particularly at the second instars. In general, genes involved in ecdysone and JH synthesis were expressed at higher levels in second instar females, whereas genes involved in ecdysone and JH degradation were expressed at lower levels. For example, *CYP314A1 (shade)* and *CYP307A1 (spook)* , which encode cytochrome P450 enzymes involved in ecdysone synthesis; and *CYP15A1,* which encodes an enzyme involved in synthesizing JHIII from methyl farnesoate (Helvig, Koener, Unnithan, & Feyereisen, 2004); were expressed at higher levels in second instar females. On the other hand, a gene encoding a JH esterase (JHE), which plays a role in breaking down JH; and *CYP18A1* , which encodes an enzyme to break down ecdysone; were expressed at lower levels in the second instar females. The expression patterns of genes related to hormone metabolism were consistent with a significant drop in JH in second instar males (Figure 4b, d) (Helvig et al., 2004) (Figure 4b, d). The overall higher expression levels of hormone synthesis genes in the second-instar than in first-instar nymphs suggest that both JH and 20E production are ramped up in the second instars.

Based on their expression patterns and putative roles in hormone metabolism, the five putative genes *CYP314A1* , *CYP307A1* , *CYP15A1* , *CYP18A1* , and the JHE-encoding gene were selected for follow-up RNAi studies to provide direct evidence for their roles in hormone metabolism. Interference effect of RNAi on gene expression was observed between 24 to 72 hours (72h) (Figure 4e, f). Silencing *CYP18A1* slowed down growth of females (Figure 4g, h), suggesting that *CYP18A1* depletion interfered with growth of WWS larvae.

### 3.5 — Evolutionary metamorphosis

According to gene clustering analyses, 24,923 gene families could be used for evolutionary comparison between WWS and other 15 representative insects (Supplementary Figure 7, Supplementary Table 10). A phylogenetic tree was generated with robust and consistent results on key taxa from extant insect orders representing hemimetabolic, holometabolic and ametabolic species, and estimated divergence dates based on validated fossils (Figure 2c). Regardless of the reference genomes and tree-building methods used, the genome phylogeny strongly supported the interspecific relationships shown in the tree based on the analyses of reference genomes (Figure 2c). Estimation of divergence times suggests that the aphids *Acyrthosiphon pisum* and *Diuraphis noxia* form a sister group and diverged from the WWS in the Middle Triassic ( 218.0 $\pm$ 60.7 Mya). Further, our analysis indicated that the last common ancestor of Holometabola and Hemimetabola separated in the Late Carboniferous ( 301.6 +- 9.7 Mya). This differs from a previous report, which found that fossils of holometabolic insects were first seen at the end of Permian ( 252 Mya) and constitute a monophyletic group (Kristensen, 1975; Nel et al., 2013). Our research indicates that complete metamorphosis emerged in the Late Carboniferous ( 301.6 +- 9.7 Mya, Figure 2c), at least 50 million years earlier than the previous study, and followed by a period were terrestrial plants diversified dramatically (Misof et al., 2014). Based on the phylogenetic analysis, ametabolous represent a primitive group located in the basal branch of the tree, whereas hemiptera and holometabola reside in the lower and upper branches separately. WWS represents an important transitional model between hemimetaboly and holometaboly. Our study provides evidence that holometamorphosis derived from hemimetamorphosis in the Late Carboniferous. Clearly, there is an evolutionary trend from ametabolous to hemimetabolous insect development, followed by the emergence of holometabola (Figure 1b, 2c).

### 4 — DISCUSSION

Here we report a high-quality genome sequence assembled from an important scale insect species. Among the 14,020 protein-encoding genes predicted from this genome sequence, 6,189 genes were unique and 5,599 of them were supported by transcripts, indicating that these genes encode real products (Supplementary Figure 11). Previous molecular phylogenetic studies only used a number of nuclear and mitochondrial genes to construct the evolutionary tree, which might have strong bias in data sampling and could not well reflect the evolutionary relationship between different species in insects (Misof et al., 2014). The molecular evolution tree based on whole genome information can avoid some of the problems (Figure 2c). The large number of unique genes discovered from this study provides a useful resource for future WWS studies and should also be useful for future comparative studies with other organisms.

Expansion and reduction of genes are usually associated with adaptation of species to specific ecological requirements in their respective habitats (Harris & Hofmann, 2004; Lespinet et al., 2002). Parasitic insect species typically display expanded gene families that allow for feeding of the hosts (Oliva et al., 2015; Rays et al., 2013; Zhao et al., 2015). Plant pathogens have expanded gene families that function in cell wall degradation to facilitate invasion to their host plants (Morales-Cruz et al., 2015). Compared to other insect genomes, there are several categories of genes that underwent significant expansion in the WWS genome. One of these categories corresponds to fatty acid metabolism, including fatty acid synthases and low-density lipoprotein. UDP-glucuronosyltransferases, involved in removing xenobiotics, are also expanded significantly, which may have allowed the insect to eliminate endogenous and exotic toxic chemicals.

One of the striking features of the WWS genome is its unique DNA methylation pattern with the highest methylation level (Supplementary Table 18) in the class of hexapoda that has been reported so far (Phalke et al., 2009; Regev et al., 1998; Xiang et al., 2010). The high methylation of the WWS genome may have allowed the insect for high phenotypic plasticity to cope with environmental challenges since this females insect species is immobile at the adult stage, and as such, hard to evade harsh environmental conditions.

Two families of hormones, the ecdysteroids and the juvenile hormones, control molting and metamorphosis during postembryonic life (Gilbert, 1994). Ecdysone induces the production of a new cuticle, whereas JH regulates the character of molting (Truman & Riddiford, 1999). During molts, JH levels are high, but fall dramatically to practically undetectable in the pre-adult stage (Belles, 2011). JH represses metamorphosis, and its downregulation is required for metamorphosis to occur (Riddiford, 2008). In our study, 20E titers in male development shows five distinct peaks, while females only display two, consistent with the idea the male development is more complex. Remarkably, females exhibit a pronounced peak of JH in the 1st stage nymph, with a corresponding drop in ecdysone. As such, the determinant of female development appears to occur early in development, since the JH titers are dramatically higher in FF compared to FM samples and the relationship between JH and ecdysone is reversed in males (Figure 4a-c). In fact, the female 1st stage nymph is the only time in either of the sexes where JH eclipses absolute ecdysone concentrations. Therefore, we propose that the key developmental events that lead to female development are established already in 1st stage nymphs. Later in development, the additional ecdysone peaks observed in males are likely associated with the development of morphological features specific to males.

Insect metamorphosis is a fascinating adaptation that allows the transformation of nymphs into reproductive adults. Ancestral insect species did not undergo metamorphosis and there are still some extant species that lack metamorphosis or only undergo partial metamorphosis (Truman & Riddiford, 1999). Both hemimetamorphosis and complete metamorphosis exist in the male and female of white wax insect, respectively (Supplementary Figs. 2). This suggests that WWS represents a transitional position between hemimetamorphosis and holometabolic insects. The main benefits of sexually dimorphic metamorphosis in WWS are likely the ability of larvae and adults to utilize different habitats and food resources, and this may have serves as a force to accelerate the evolution of rapid life cycles and enhance their chance for survival effectively.

From this respect, WWS represents a fascinating model not only to study the adaptive mechanisms of plant-sap feeding parasites but also to examine the evolutionary relationships between hemimetaboly and holometaboly in different insect orders. The phylogenetic status, evolution of WWS gene families presented here provide a key case and framework for furtherly understanding the evolution of insect metamorphosis

9

and functional herbivory adaption.

Our results provided comprehensive genome and epigenome resource of scale insects and shed new insights into the different evolution paths of metamorphosis between males and females within the same species. It represents a conceptual framework for future studies that examine the evolution of hexapod metamorphosis.

## ACKNOWLEDGMENTS

## AUTHOR CONTRIBUTIONS

X.C., K.K. and H.C. designed and led the project. N.L. and P.L. collected samples and determined hormone titers. H.C., X.L. and Q.L. accomplished the whole-genome sequencing and Hi-C assembly. H.C., Q.L., Y.F., K.K. and M.C. performed whole-genome assembly quality control and methylome analyses. Q.L., W.W. and P.L. afforded photos of fresh collection and designed the figures; X.L., N.L., X.Z., Q.L. and P.L. performed the functional experiments of RNA interference. H.C., M.C. and X.C. wrote the manuscript.; M.C. and K.K. improved the manuscript. All authors read and approved the final manuscript.

## ORCID

Hang Chen https://orcid.org/0000-0002-8690-4099

Qin Lu https://orcid.org/0000-0001-5902-5390

Xiaoming Chen https://orcid.org/0000-0003-3989-9925

Jinwen Zhang https://orcid.org/0000-0001-8554-9593

Qian Qi https://orcid.org/0000-0002-7863-958X

Ming-Shun Chen https://orcid.org/0000-0002-0636-808x

Kirst King-Jones https://orcid.org/0000-0002-9089-8015

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

The datasets generated and analysed during the current study are available in the Sequence Read Archive repository (SRA;*https://www.ncbi.nlm.nih.gov/sra/*). The RNA-seq data generated in this study have been submitted to the NCBI BioProject database (*https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA551363*) under accession number PRJNA551363, Hi-C data have been submitted to the Sequence Read Archive (SRA;*https://www.ncbi.nlm.nih.gov/sra/?term=SRP263073*) under accession number SRP263073.

The Whole Genome Shotgun project for WWS generated in this study have been submitted to the Nucleotide database of NCBI (*https://www.ncbi.nlm.nih.gov/nuccore/WOFM00000000*) under accession number WOFM00000000. The assembled genome and the merged methylome data were given on Dryad (*https://datadryad.org/stash/share/dIas3tyCEDAD1jJnCpZ6S7ZsqI9WFMLN7QiasQBOlR4*).

## REFERENCES

Alexandros, & Stamatakis. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22* , 2688-2690.

Azinna, R., HirokiGotoh, TakaakiKojima, & TeruyukiNiimi. (2018). Recent advances in understanding the mechanisms of sexually dimorphic plasticity: insights from beetle weapons and future directions. *Current Opinion in Insect Science, 25* , 35-41.

Barbora, K., Vlastimil, S., Marek, J., & Amit, S. (2011). Common and Distinct Roles of Juvenile Hormone Signaling Genes in Metamorphosis of Holometabolous and Hemimetabolous Insects. *PLoS One, 6* (12), e28728-.

Belles, X. (2011). *Origin and Evolution of Insect Metamorphosis* : John Wiley & Sons, Ltd.

Consortium, H. G. S. (2006). Honey Bee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee Apis mellifera. *Nature, 444* (7118), 512-512.

Feyereisen, R., & Tobe, S. S. (1981). A rapid partition assay for routine analysis of juvenile hormone release by insect corpora allata. *Analytical Biochemistry, 111* (2), 372-375.

Gilbert, L. I. (1994). *Insect Hormones* (Vol. 39): Princeton Univ. Press.

Hao, F., Conneely, K. N., & Hao, W. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data.*Nucleic Acids Research, 8* (8), 8.

Harris, R. M., & Hofmann, H. A. (2004). Seeing is believing: Dynamic evolution of gene families: Fig. 1.*Proceedings of the National Academy of Sciences of the United States of America, 112* (5), 1252-1253.

Helvig, C., Koener, J. F., Unnithan, G. C., & Feyereisen, R. (2004). CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. *Proc Natl Acad Sci U S A, 101* (12), 4024-4029.

Kelly, T. J., Masler, E. P., Thyagaraja, B. S., Bell, R. A., & Imberski, R. B. (1992). Development of an in vitro assay for prothoracicotropic hormone of the gypsy moth,Lymantria dispar(L.) following studies on identification, titers and synthesis of ecdysteroids in last-instar females. *Journal of Comparative Physiology Biochemical Systemic and Environmental Physiology, 162* (7), 581-587.

Kristensen, N. P. (1975). The phylogeny of hexapod "orders". A critical review of recent accounts.*Journal of Zoological Systematics & Evolutionary Research, 13* (1), 1-44.

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics, 27* (11), 1571-1572.

Kunst, L., & Samuels, A. L. (2003). Biosynthesis and secretion of plant cuticular wax. *Progress in Lipid Research, 42* (1), 51-80.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9* (4), 357-359.

Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family exapnsion in the evolution of eukryotes. *Genome Research, 12* , 1048-1059.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., & Maleszka, R. (2011). The honey bee epigenomes: differential methylation of brain DNA in queens and workers.*PLoS Biol, 9* (1), e1000506.

Ma, J., Liyi, M., Hong, Z., Zhongquan, Z., Youqiong, W., Kai, L., . . . A., L. D. (2018). Policosanol fabrication from insect wax and optimization by response surface methodology. *PLoS One, 13* (5), e0197343-.

Ma, J., Ma, L., Zhang, Z., Li, K., & Zhang, H. (2018). In vivo evaluation of insect wax for hair growth potential. *PLoS One, 13* (2), e0192612.

Mao, X., Tao, C., Olyarchuk, J. G., & Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary.*Bioinformatics, 21* (19), 3787-3793.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., & Xin, Z. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science, 346* (6210), 763-767.

Morales-Cruz, A., Amrine, K. C. H., Blanco-Ulate, B., Lawrence, D. P., Travadon, R., Rolshausen, P. E., . . . Cantu, D. (2015). Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics, 16* , 469.

Nel, A., Roques, P., Nel, P., Prokin, A. A., Bourgoin, T., Prokop, J., . . . Wappler, T. (2013). The earliest-known holometabolous insects. *Nature, 503* (7475).

Oliva, R. F., Cano, L. M., Raffaele, S., Win, J., Bozkurt, T. O., Belhaj, K., . . . Kamoun, S. (2015). A Recent Expansion of the RXLR Effector Gene Avrblb2 Is Maintained in Global Populations of Phytophthora infestans Indicating Different Contributions to Virulence. *Mol. Plant-Microbe Interact, 28* (8), 901.

Park, Y., & Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics, 32* (10), 1-10.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., & Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics, 41* (6), 696-702.

Qi Q, Lv P, Chen X-M, Chen H, Chen M-S, Yang P. 2019. Sexual Dimorphism in Wax Secretion Offers Ecological Adaptability During Ericerus pela (Hemiptera: Coccidae) Evolution. Environmental Entomology 48: 410-418.

Rays H. Y. Jiang , Irene de Bruijn , Brian J. Haas , Rodrigo Belmonte, Lars Lobach, James Christie, . . . West, P. v. (2013). Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica. *Plos Genetics, 9* (6), e1003272.

Regev, A., Lamb, M. J., & Jablonka, E. (1998). The role of DNA methylation in invertebrates: Developmental regulation or genome defense? *Mol. Biol. Evol., 15* (7), 880–891.

Riddiford, L. M. (2008). Juvenile hormone action: A 2007 perspective. *Journal of Insect Physiology, 54* (6), 895-901.

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., . . . Schneider, D. J. (2011). Creating a Buzz About Insect Genomes.*Science, 331* (6023), 1386-1386.

Slama, K., & Williams, C. M. (1965). Juvenile hormone activity for the bug Pyrrhocoris apterus.*Proceedings of the National Academy of Sciences, 54* , 411-414.

Stillwell, R. C., Blanckenhorn, W. U., Teder, T., Davidowitz, G., & Fox, C. W. (2010). Sex Differences in Phenotypic Plasticity Affect Variation in Sexual Size Dimorphism in Insects: From Physiology to Evolution. *Annual Review of Entomology, 55* , 227-245.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol., 28* (10), 2731-2739.

Teerawanichpan, P., Robertson, A. J., & Qiu, X. (2010). A fatty acyl-CoA reductase highly expressed in the head of honey bee (Apis mellifera) involves biosynthesis of a wide range of aliphatic fatty alcohols. *Insect Biochemistry Molecular Biology, 40* (9), 0-649.

Truman, J. W., & Riddiford, L. M. (1999). The origins of insect metamorphosis. *Nature, 401* (6752), 447-452.

Vea, I. M., Tanaka, S., Shiotsuki, T., Jouraku, A., Tanaka, T., & Minakuchi, C. (2016). Differential Juvenile Hormone Variations in Scale Insect Extreme Sexual Dimorphism.*PLoS One, 11* (2), e0149459.

Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C et al. 2014. The locust genome provides insight into swarm formation and long-distance flight Nature Communications 5: 1-9.

Wang, Z. D., Feng, Y., Ma, L.-y., Li, X., Ding, W.-f., & Chen, X.-m. (2017). Hair growth promoting effect of white wax and policosanol from white wax on the mouse model of testosterone-induced hair loss.*Biomedicine & Pharmacotherapy, 89* , 438-446.

Williams, T. M., & Carrol, S. B. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics, 10* , 797-804.

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., . . . Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Researc, 43* (21), 141.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., . . . Ma, C. (2004). A Draft Sequence for the Genome of the Domesticated Silkworm (Bombyx mori). *Science, 306* (5703), 1937-1940.

Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., . . . Dong, Y. (2010). Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology, 28* (5), 516–520.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias.*Genome Biol., 11* (2), 1-12.

Zhao, C., NavarroEscalante, L., HangChen, R.Benatti, T., Qu, J., Chellapilla, S., . . . MarcusCoyle. (2015). A Massive Expansion of Effector Genes Underlies Gall-Formation in the Wheat Pest Mayetiola destructor. *Current Biology, 25* (5), 613-620.
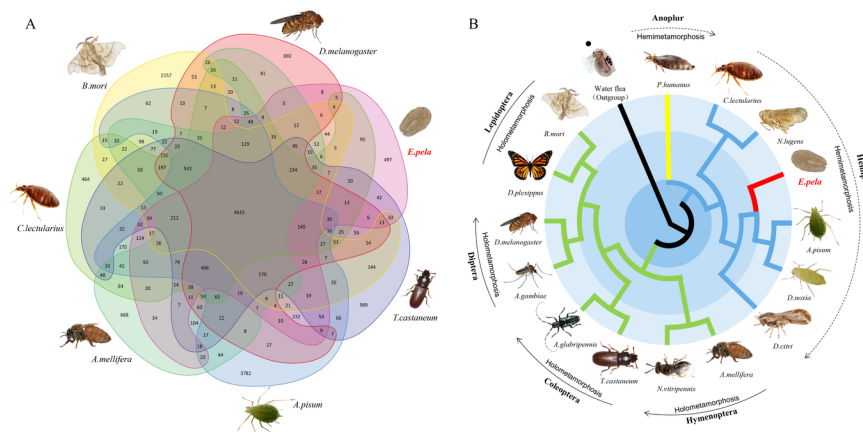
## FIGURE LEGENDS

**FIGURE 1** Orthologous genes and phylogenetic tree of insects. (a) Venn diagram of shared orthologous genes in WWS and six other insect species. All the numbers refer to the numbers of genes. Numbers in overlapped circles are the numbers of common genes between or among different species. (b) Phylogenetic tree and estimated evolutionary path for WWS and other invertebrates. The species with dotted lines represent hemimetamorphosis located in the lower branches while species with solid lines represent holometabola located in the upper branches. The circles shaded blue, light blue and gray areas represent the boundaries for the Ordovicain-Permian, Permian-Triassic and Jurassic to the present, respectively. The arrows on different order represent the direction of insect evolution.
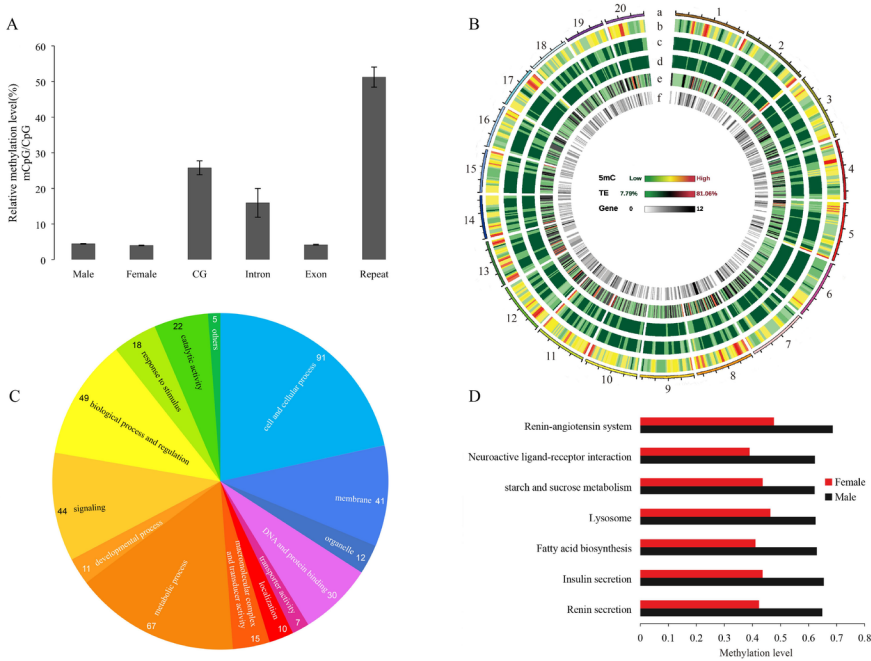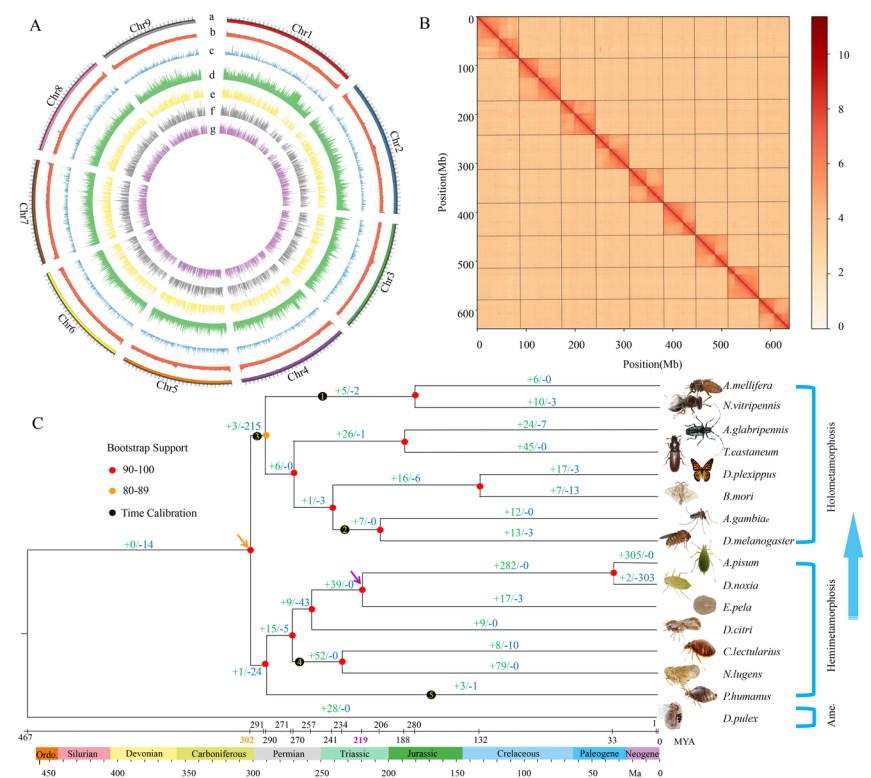
**FIGURE 2** Genome evolution of *E. pela* . (a) Distribution of WWS genomic features. Track a: Circular representation of the linkage chromosomes of the WWS genome. Track b–d: GC content (100kb window), gene density (100kb window) and percentage of repeats (100kb window). Track e-g: Mean expression of annotated genes calculated as log (FPKM + 1) of the mean expression value in three repeat female samples in the first larvae stage. (b) Genome wide Hi-C interaction map with 500k resolution. The color from light to deep indicates the increase of interaction intensity, and the deeper the color represent the stronger the interaction. The coordinates represent its position on the genome. The nine squares are the nine chromosomes of the WWS. (c) Insect metamorphosis and its evolutionary path inferred based on comparison of 24,923 gene families from WWS and 15 other representative invertebrate genomes. The numbers on each branch correspond to the numbers of gene families that have expanded (green) and contracted (blue) in the WWS genome. All nodes in red received support levels of 90-100% and 80-89% in orange with five black dots representing time calibration by fossils and references. The estimated divergence times are included under the nodes. MYA, million years ago. The arrow and the number in orange color indicate the divergence time of the hemimetabolous separated from the holometabolan and those in purple color show the time of WWS diverge from aphids. The arrow in light blue on different order indicate the direction of insect evolution is ametabolous to hemimetabolous, then to Holometabola. Ame. is the abbreviation of ametabolous.
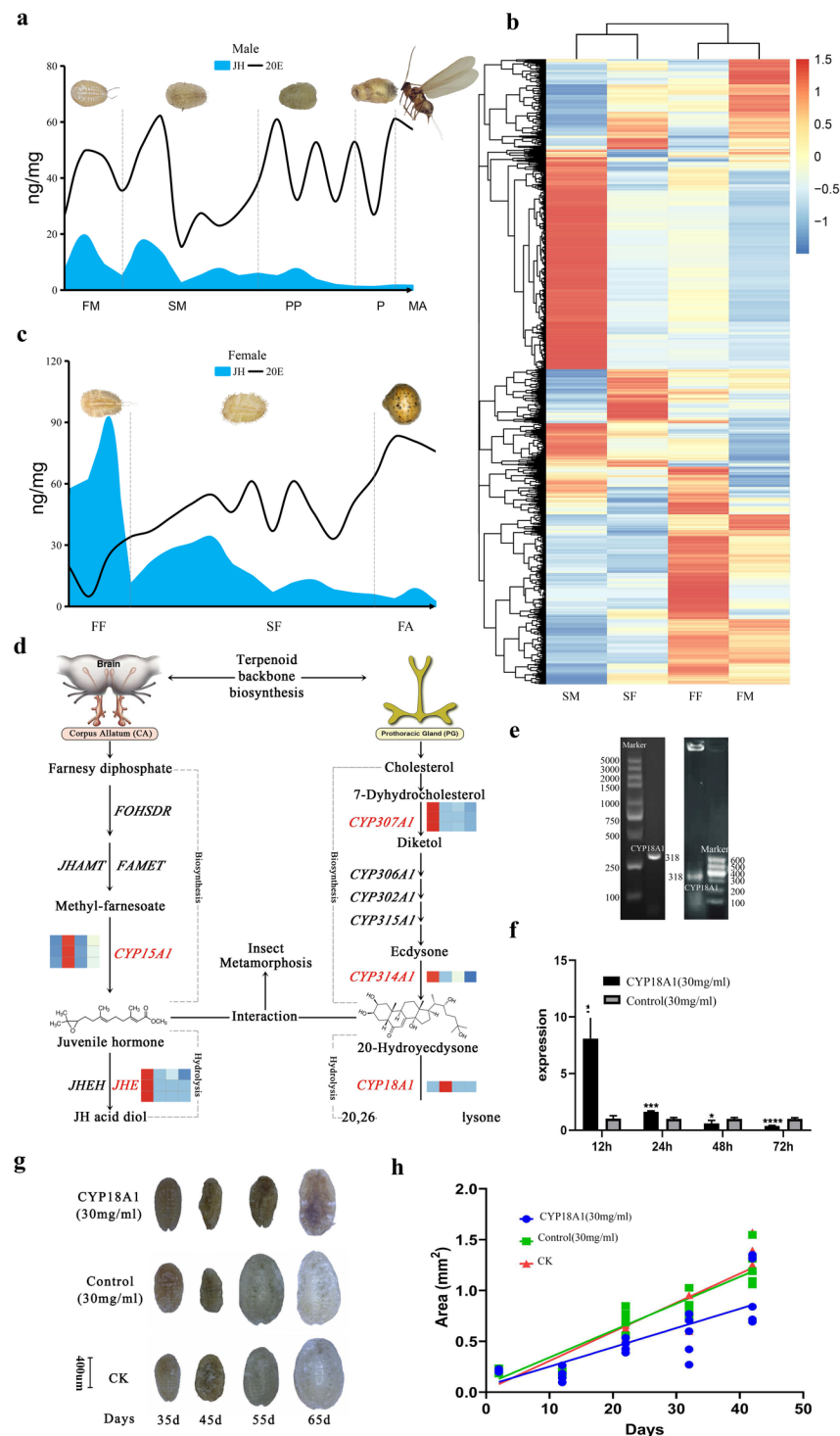
**FIGURE 3** Methylomic analyses between females and males of WWS at the first instar nymph stage. (a) Relative CG methylation levels of different genomic regions through a WGBS analysis. The ratio was

defined as the number of methylated CG to the number of total CG and was calculated using the CG on the reads that map to the defined regions. The results were derived from the data of first-instar nymphs. (b) Distribution of methylation density on the WWS whole genome. Track a: 20 linkage groups of the WWS genome. Track b: CG sequence methylation density. Track c: CHG sequence methylation density. Track d: CHH sequence methylation density. Track e: TE represents the proportion heat map of repeated sequences. Track f: Gene number density. (c) Functional classification of differentially methylated genes (DMGs) between male and female in first-instar nymphs. (d) Enrichment analysis of KEGG of DMR related genes. The bar represents the proportion of gene methylation of each pathway in the male (black) and female (red).

**FIGURE 4** Juvenile hormone and ecdysone titers with gene regulation comparison between females and males. (a), (c) The change in JH and 20E titers during different development stages of male (a) and female (c) development. Males undergo through first and second instar nymphs (FM, SM), pseudo-prepupae (PP), pseudo-pupae (P), and adults (MA). Females undergo through first to second instar nymph (FF, SF) and adults (FA). (b) A heat map of genes with elevated expression in nymphs. The heat map was generated using hierarchical clustering and complete linkage of the top 500 most highly expressed orthologous genes from males and females. Different colors represent different levels of gene expression in FF, FM, SF, SM, respectively. (d) Model of molecular regulation involved in hormone metabolism pathways. Five genes are shown in red involving JH and ecdysone synthesis and degradation. The expression patterns of the genes *CYP18A1* , *CYP314A1* and *CYP307A1* , having roles in the metabolism of ecdysone whereas the patterns of the genes *JHE* and *CYP15A1* , having roles in the metabolism of juvenile hormone. (e) Detection of *CYP18A1* target gene by PCR. Left lane is PCR amplification results of WWS cDNA whereas right lane is PCR amplification results of HT115 (*CYP18A1* -L4440) bacterial solution. (f) Expression levels of *CYP18A1* at 12, 24, 48 and 72 hours after RNAi treatment in the early second instar of nymphs. Control is WWS female under HT1115 treatment without target gene *CYP18A1* (L4440/HT115). (g) Morphological changes of WWS female at 35, 45, 55 and 65 days were monitored after RNAi treatment. Control is WWS female under HT1115 treatment without target gene *CYP18A1* (L4440/HT115). CK is WWS female under natural status without any treatment. (h) Variation trend of WWS female body area at 2, 12, 22 and 42 days after RNAi treatment. Control is WWS female under HT1115 treatment without target gene *CYP18A1* (L4440/HT115) and CK is WWS female under natural status without any treatment.

**Hosted file**

`Table 1.pdf` available at https://authorea.com/users/375617/articles/492784-genome-assembly-and-methylome-analysis-of-the-white-wax-scale-insect-provides-insight-into-sexual-

differentiation-of-metamorphosis-in-hexapod