

SCIENTIFIC REPORTS



OPEN

De novo characterization of *Phenacoccus solenopsis* transcriptome and analysis of gene expression profiling during development and hormone biosynthesis

Surjeet Kumar Arya^{1,2}, Yogeshwar Vikram Dhar^{1,2}, Santosh Kumar Upadhyay³,
Mehar Hasan Asif^{1,2} & Praveen Chandra Verma^{1,2}

The cotton mealybug *Phenacoccus solenopsis* is a devastating pest of cotton causing tremendous loss in the yield of crops each year. Widespread physiological and biological studies on *P. solenopsis* have been carried out, but the lack of genetic information has constrained our understanding of the molecular mechanisms behind its growth and development. To understand and characterize the different developmental stages, RNA-Seq platform was used to execute *de-novo* transcriptome assembly and differential gene expression profiling for the eggs, first, second, third instar and adult female stages. About 182.67 million reads were assembled into 93,781 unigenes with an average length of 871.4 bp and an N50 length of 1899 bp. These unigenes sequences were annotated and classified by performing NCBI non-redundant (Nr) database, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Groups (COG), Gene ontology (GO), the Swiss-Prot protein database (Swiss-Prot), and nearest related organism *Acyrtosiphon pisum* (pea aphid) database. To get more information regarding the process of metamorphosis, we performed a pairwise comparison of four developmental stages and obtained 29,415 differentially expressed genes. Some of the differentially expressed genes were associated with functional protein synthesis, anti-microbial protection, development and hormone biosynthesis. Functional pathway enrichment analysis of differentially expressed genes showed the positive correlation with specific physiological activities of each stage, and these results were confirmed by qRT-PCR experiments. This study gives a valuable genomics resource of *P. solenopsis* covering all its developmental stages and will promote future studies on biological processes at the molecular level.

The cotton mealybug, *Phenacoccus solenopsis* (Pseudococcidae) is a genetically diverse and highly destructive invasive insect pest of many horticultural and agricultural crops, worldwide¹. It infests more than 200 different plant species and causes significant yield loss^{2–6}. It affects the plants by feeding and vectoring debilitating plant viruses, as well⁷. It feeds phloem-sap and promotes the growth of damaging fungi on honeydew excretions deposited on plant parts⁵. Annual worldwide crop loss is estimated around hundreds of millions of dollars⁵. Other mealybug species infesting cotton in India include *Maconelliococcus hirsutus* (pink hibiscus mealybug), *Paracoccus marginatus* Williams and Ganara de willink (Papaya mealybug) and *Nipaecoccus viridis* (spherical mealybug)⁵. However, the cotton mealybug *P. solenopsis* is the most destructive species, which affects 65% of cotton producing area in India^{8–10}. Due to the lack of efficient mealybug resistance germplasm, the complexity of plant-mealybug

¹CSIR-National Botanical Research Institute, (Council of Scientific and Industrial Research) Rana Pratap Marg, Lucknow, UP-226001, India. ²Academy of Scientific and Innovative Research (AcSIR), Anusandhan Bhawan, Room No: 310, 2-Rafi Marg, New Delhi, India. ³Department of Botany, Panjab University, Chandigarh, 160014, India. Correspondence and requests for materials should be addressed to P.C.V. (email: praveenverma@nbri.res.in)

interactions got increased along with the development of resistant pest biotypes and outbreaks of mealybug causing substantial crop losses are being regularly reported^{11–13}. Cotton breeders and seed plant growers are in continuous efforts to find an efficient genetic strategy for their control in cotton plants.

Hemi-metabolous insects, such as aphids, whiteflies, leafhoppers and true bugs undergoes morphogenesis as similar to that seen in the larva-to-pupa and pupa-to-adult transitions of holo-metabolous insects, to produce mature external wings and genitalia. However, the change of one form to another is not prominent because crawler or nymphs look similar in appearance to their adult form. In spite of all these differences in metamorphosis, both holo-metabolous and hemi-metabolous process are regulated by general mechanisms involving molting steroid 20-hydroxyecdysone (20E) and the sesquiterpenoid, juvenile hormone (JH)^{14–16}. JH regulates many processes of insect physiology, such as reproduction, development, diapauses, and metamorphosis^{16–18}. Subsequently, regulation of JH levels is crucial throughout an insect's life.

The transcriptome sequencing data can be used to infer the information related to the development and other metabolic processes in an organism. In insect, growth, molting and other regulatory processes are mostly regulated by the hormonal balance between acyclic sesquiterpenoid produced at the corpus allatum (CA), and ecdysteroid synthesized at the prothoracic glands¹⁹. The family of sesquiterpenoid primarily limited to insects²⁰. Eight different types of JHs have been identified in insects²¹. They are synthesized through the mevalonate pathway (MVAP) and involve thirteen discrete enzymatic reactions. The synthesis of JH is conventionally divided into two steps, early (MVAP) and late (JH branch) steps. In early steps, MVAP has to form FPP through a series of sequential steps that involves the critical roles of major and minor enzymatic reactions. Mevalonate first transformed into isopentenyl diphosphate (IPP) in the presence of three enzymatic steps. FPP synthase (fpps), a short-chain prenyltransferase, generates a FPP by completing two sequential couplings reactions. In the first reaction, IPP and dimethylallyl pyrophosphate (DMAPP) condense in a head to tail manner to produce geranyl diphosphate (GPP). Then, condensation reaction occurs repeatedly where GPP reacts with IPP and yields FPP. Many FPP synthases have been identified in several insects and typically active as homodimers^{22,23}. The late steps of JHs biosynthesis involve the conversion of FPP to farnesol (FOL) that is catalyzed by an FPP phosphatase (fppase or fppp). Afterward, farnesol undergoes two sequential oxidation reactions that generate farnesal and farnesoic acid (FA). Further, biosynthetic step involves methyl esterification and epoxidation which are catalyzed by an acid methyltransferase (Jhamt) and an epoxidase (epox). The last step of the biosynthetic reaction is usually considered to be JH-specific^{21,24}.

The steroid hormone ecdysone has roles in insect molting and metamorphosis through its timely release into the circulating hemolymph from the prothoracic gland^{19,24}. In case of ecdysteroid, phylum arthropoda lacks the squalene synthase enzyme, and its synthesis relied on the continuous supply of external sterol source derived from the diet. It usually requires dealkylation and a series of hydroxylation steps¹⁹. Previous studies reported that the most active form of ecdysone is 20-hydroxyecdysone (20E) which binds effectively to the nuclear receptor (EcR) of the target tissue to elicit specific changes in the transcription of genes. The complete biosynthesis of 20E from cholesterol is catalyzed by cytochrome P450 enzymes encoded by the halloween family of genes. These families of genes include: phantom/Cyp306A1 (phm), spook/Cyp307A1 (spo), spookier/Cyp307A2 (spok), shadow/Cyp315A1 (sad)^{24–27}.

In this study, we analyzed the high throughput RNA sequencing data (RNA-Seq) to establish transcriptome profile of four developmental stages of *P. solenopsis*, enabling the identification of potential target genes based on differential expression. We generated over 100 million bases of high-quality RNA sequence using Illumina technology, which was assembled into 93,781 distinct sequences. RNA-Seq data from various developmental stages enabled identification of vital differentially expressed genes (DEGs) involved in the development and various biosynthetic pathways. Those essential genes may also be a potential target for the control of mealybug in plants through RNA interference (RNAi).

Materials and Methods

Insect rearing. *P. solenopsis* was reared on cotton plants at 25–28 °C, 70–80% relative humidity and photoperiods of 16:8 (L: D) h for biological studies in the laboratory at CSIR-National Botanical Research Institute, Lucknow, India, as described earlier²⁸. Egg, first instar, second instar and third instar nymphs, and female adults) of *P. solenopsis* were collected separately from the reared population. Due to the limited number of biological samples, egg and first instar nymphs were pooled and considered as a single sample as EggI, while second instar, third instar and female adults were taken, separately.

Construction of cDNA library, Illumina sequencing and assembly. Total RNA from above samples were extracted separately using TRIzol reagent (Sigma Aldrich, USA) following the manufacturer's instructions. DNA contaminants were removed using DNase enzyme (Invitrogen) digestion, followed by rRNA removal using Ribo-Minus (Invitrogen). The RNA samples were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, USA) and qualified by Agilent Bioanalyzer (Agilent Technologies, USA). The processed RNAs with RNA integrity number (RIN) values (≥ 7.00) were sent for sequencing using Illumina HiSeq 2500 platform (2 × 100 bp strategy). The cDNA libraries were constructed according to Illumina TruSeq™ Stranded mRNA sample preparation guide for Illumina Paired-End sequencing service provided by Scigenome, India. Raw reads generated through HiSeq 2500 were filtered to exclude low quality reads. The transcriptome data of *P. solenopsis* was processed through a filtering process while considering the Q-value of 30 using the NGS QC Toolkit²⁹. The clean reads from four cDNA libraries were pooled and subjected to *de-novo* assembly using Trinity software³⁰ with default parameters (Supplementary Fig. S1). High-quality reads had a Phred score over 30 across more than 70% of the bases.

Sample name	Raw reads	Clean reads	% of HQ reads	Q > 30 (%)	GC content (%)
A-R1	28055790	25311685	98.6	90.23	39.72
A-R2	28055790	25311685	96.9	84.35	39.48
B-R1	22904667	20021109	98.42	88.88	40.12
B-R2	22904667	20021109	96.25	80.13	39.7
C-R1	24200312	20147343	98.12	88.12	39.23
C-R2	24200312	20147343	95.31	74.78	38.77
D-R1	16183365	14853656	91.78	71.23	40.12
D-R2	16183365	14853656	91.78	65.34	39.95

Table 1. Summary of statistical data for the transcriptomes of *P. solenopsis* developmental stages.

Functional annotation and differential gene expression analysis. Assembled unigenes were annotated by a BLASTx search (E-value < 10^{−5}) against the NCBI non-redundant (Nr) database (<ftp://ftp.ncbi.nih.gov/blast/db/>), Clusters of Orthologous Groups (COG) (<http://www.ncbi.nlm.nih.gov/COG/>) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>), Gene ontology (GO) (<http://www.geneontology.org/>), the Swiss-Prot protein database (Swiss-Prot) (<http://www.uniprot.org/>), and nearest related organism *Acyrtosiphon pisum* (pea-aphid)³¹. AgriGO (<http://bioinfo.cau.edu.cn/agriGO/>) online tool was used for GO mapping of the annotated contigs. The gene expression abundance was calculated using RSEM³² and compared using edgeR software³³. The differentially expressed genes were distinguished on the basis of stage-specific genes with their log₂ Fold change value in the respective developmental stages (if P < 0.05, |log₂| ≥ 2 and FDR < 0.05 to be considered statistically significant).

Pathway enrichment analysis of the differentially expressed genes (DEGs). To obtain a comprehensive perspective on the molecular basis of development, we focused on differential genes that have a role in hormone biosynthesis pathway and development. The enrichment analysis of GO and KEGG pathways was performed with DEGs by using KOBAS 3.0 software [<http://kobas.cbi.pku.edu.cn/home.do>]^{34,35}. Pathway Enrichment analysis finds the most significantly enriched pathways by applying a hypergeometric test to map all DEGs to give terms in the GO database by comparing it with the whole genome background. The mapped enriched pathways include metabolic or signal transduction. The formula used for the calculation was:

$$P = 1 - \sum_{i=0}^{M-1} \frac{\binom{m}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

here, N represents the number of all genes that with KEGG annotation, n is the number of DEGs in N, M is the number of all genes annotated to specific pathways, and m is the number of DEGs in M. The pathway with Q value 0.05 is defined as the enriched one. For the Detailed method one can refer to the earlier reports^{36,37}. Visualization of data was accomplished using the JMP 13.0 software³⁸.

qRT-PCR analyses. The expression of 10 selected genes during developmental stages in *P. solenopsis* was validated using qRT-PCR. The gene-specific primers were designed using Primer 3.0 program (Table S1). The qRT-PCR reactions were performed on 7500 Real-Time PCR Systems (Applied Biosystems, USA) using the following program cycles with step involves: an initial denaturation step of 20 s at 50 °C, 95 °C for 10 min., and 40 cycles of 15 s at 95 °C, and 1 min at 60 °C. Reaction was followed by melt curve analysis at default parameters to check the PCR specificity by steady increase in temperature from 60 °C to 90 °C. The expression level was estimated using the 2^{−ΔΔCT} method and normalized with three selected reference genes (*β-tubulin*, *α-tubulin*, and *rpl32*) based on expression studies in different development stages of *P. solenopsis*²⁸. Values for each developmental stage were taken as an average of three technical replicates of each biological replicate. Statistical analysis was done with using SPSS 16.0.

Data Accessibility. The transcriptomic dataset for different developmental stages of *Phenacoccus solenopsis* has been in deposited in the NCBI SRA database, under the SRA accession number of SRP133470.

Results

Whole-transcriptome sequencing and data assembly. A total of 56.11, 45.80, 48.40 and 32.36 million raw-reads were generated from EggI, second instar, third instar and adult females respectively. After removing low quality and adaptor reads, 50.62, 40.04, 40.29 and 29.70 million clean-reads were obtained for each. The base sequence of the total data was approximately 7.27, 11.6, 10.98 and 13.74 gigabases, respectively. The entire data were deposited in the NCBI SRA database, under the SRA accession number of SRP133470. These reads were assembled into 93,781 distinct contigs with an average length of 871 bp and N50 length of 1899 bp (Tables 1 and 2).

Gene Annotation. A total of 38,725 contigs were mapped to NR, 31965 to Swiss-prot, 16,095 to COG, 7,839 to KEGG, and 31904 to *A. pisum* databases (Fig. 1). Overall, a total of 38,725 unique contigs were annotated (Table 3). A total of 11,896 contigs were annotated by a BLAST hit with 18 insect species (Fig. 2). As expected, the top hits were found in the insect genomes, especially hemipteran and lepidopteran. *A. pisum* and *Tribolium*

Number of genes	GC percentage	N50	Average contig lengths	Total transcripts assembled
93,781	35.69	1899	871.4	1,11,871

Table 2. Summary of assembly statistics for the different developmental stages of *P. solenopsis*. N50 is the 50% length of all unigenes.

Sequence File	NR	Swiss-prot	COG	KEGG	<i>A. pisum</i>	Total
Unigenes	38,725	31,965	16,095	7,839	31,904	29,415

Table 3. Functional annotations of total unigenes in *P. solenopsis*. Nr: non-redundant database; Swiss-Prot: Swiss Protein database; KOG: Clusters of Orthologous Groups of proteins; KEGG: Kyoto Encyclopedia of Genes and Genomes; *A. pisum*.

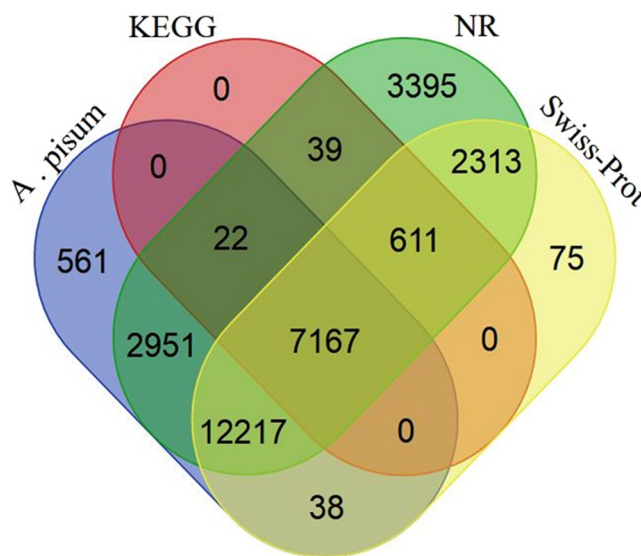


Figure 1. A Venn diagram illustrating shared and unique DEGs annotated in nr, Swissprot, COG, GO and KEGG public databases. Among 33,831 DGEs, 29,415 annotated in at least one of the public databases, including in nr, Swiss-prot, COG KEGG and GO databases, respectively.

castaneum were the first and second top-hit species with the annotations of 4,182 and 1,933 contigs followed by *Pediculus humanus corporis*, *Nasonia vitripennis*, and *Aspergillus oryzae* RIB40 (Table S2).

GO mapping and pathway analysis. The GO mapping was performed using *A. pisum* data available at AgriGO database. A total of 102 biological processes, 68 molecular functions, and 27 cellular components were mapped. Cellular process (2167 genes), metabolic process (2348 genes), primary metabolic process (1977 genes), cellular metabolic process (1756 genes), macromolecule metabolic process (1685 genes), nitrogen compound metabolic process (1087 genes), biosynthetic process (999 genes), nucleobase, nucleoside, nucleotide and nucleic acid metabolism (994 genes) were the major biological processes (Fig. 3, Tables S3–S5). In the KEGG assignment, 7839 unigenes were mapped to 259 different pathways (Supplementary excel file 1). Out of them, 20.16% pathways were classified into the pathway related to metabolism, with most of them involved in oxidative phosphorylation (27 unigenes), purine and pyrimidine metabolism (104 unigenes), carbon metabolism (154 unigenes), biosynthesis of amino acids (37 unigenes), Wnt signaling pathway (78 unigenes), and FoxO signaling pathway (69 unigenes) (Supplementary excel file 2).

Differential gene expression analyses among *P. solenopsis* developmental stages. To analyze the differential expression levels of genes between the major developmental stages, up and down-regulated genes were estimated between each pair of *P. solenopsis* such as EggI vs. second instar, second vs. third instar, third instar vs. adult female, adult female vs. EggI. Genes with two-fold variation in expression with statistical significance between two or more developmental stages were considered as differentially expressed. The expressions profile of the differential gene was changed as the insect move to the different instar of the stages. The top 10 genes for each comparison are listed in Table S6. We then performed four comparisons viz. EggI vs. second, second vs. third and third vs. adult female, and adult female vs EggI and identified the genes showing significant variation in expression.

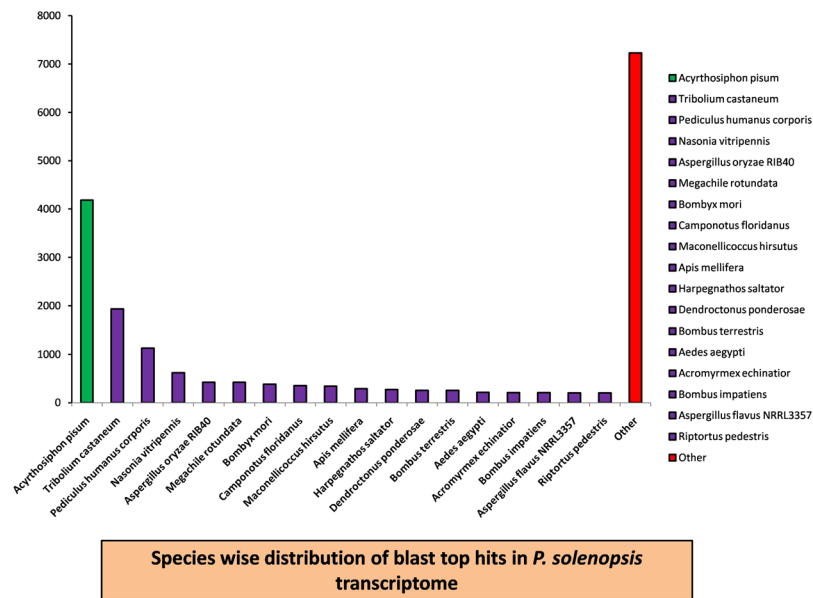


Figure 2. The top-eighteen species distribution of the BLASTX results. Unigenes were aligned with the NCBI-Nr protein database with a cutoff E value 10^{-5}. Different colors represent different species.

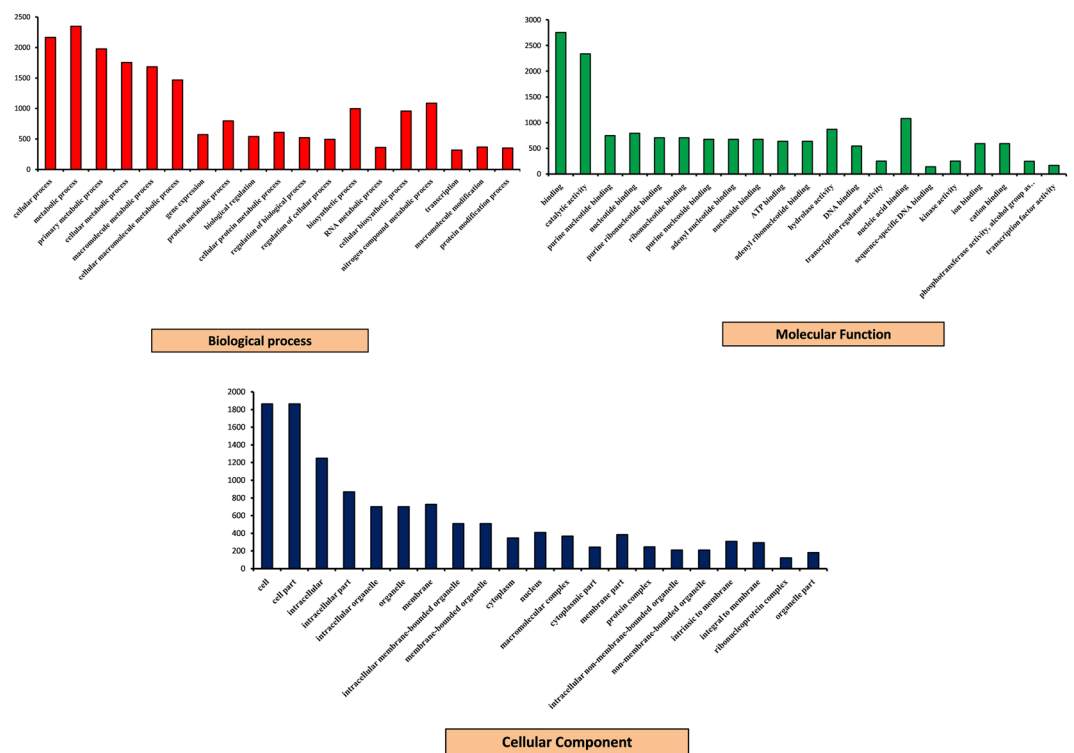


Figure 3. GO unigene categories. The unigenes were annotated in three main categories: biological process (blue), cellular component (red) and molecular function (green). The left side of the y-axis represents the percentage of a specific category of genes for a main category. The right side of the y-axis represents the number of genes in a category.

In total, 31,640 different genes showed significant expression change in hatching-to-second instar process, including 10,059 up-regulated and 21,581 down-regulated genes (Fig. 4). Most of the up-regulated genes in the second instar categorized into the cellular process, metabolic process, organic substance metabolic process, and biosynthetic process (Supplementary excel file 3). Other categories, such as organic cyclic compound biosynthetic process (biological process categories), membrane-bound organelle (cellular component categories), hydrolase activity, oxido-reductase activity, etc. were also enriched. On the other hand, the high expressing genes

Differential Gene Expression Statistics

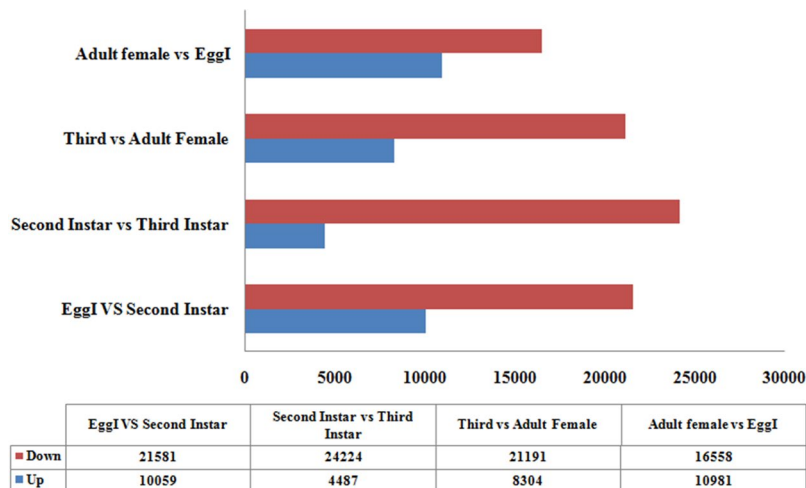


Figure 4. Comparison of the numbers of unigenes in different developmental stages in *P. solenopsis*. Up-regulated unigenes are marked in red, and down-regulated unigenes are marked in green.

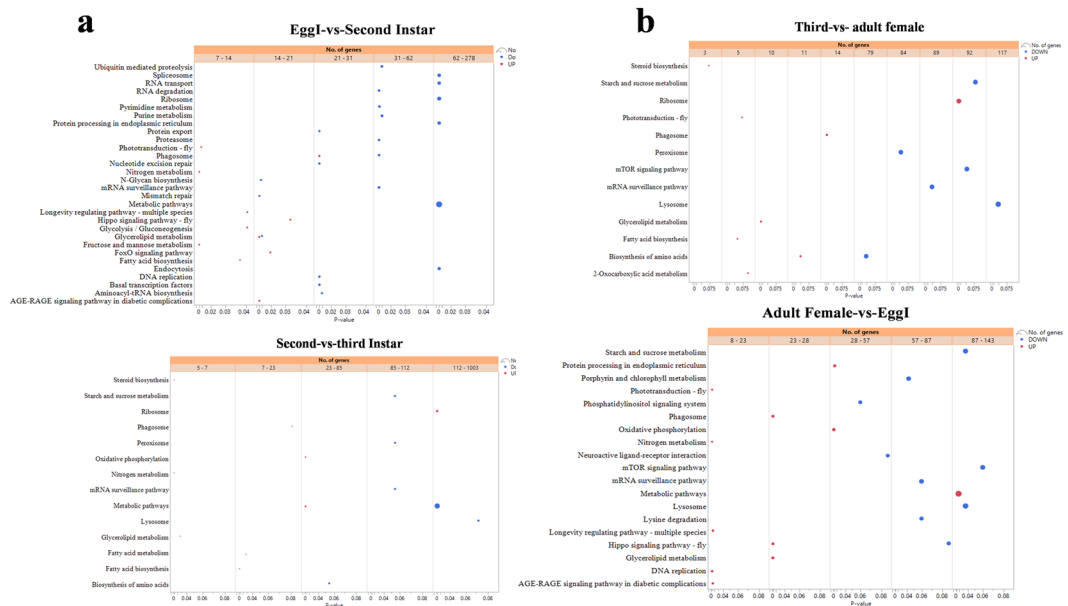


Figure 5. Enrichment analysis of KEGG pathways in comparisons of different stages. The x-axis indicates the p value calculated in enrichment test. Numbers of genes that were up-regulated (red) or down-regulated (blue) in comparisons of (a) EggI vs. Second instar and Second vs. third instar, (b) Third vs. adult and Adult female vs. EggI are shown. The size of circles indicates the number of genes in that pathway. Red circles represented up-regulated genes, while blue circles represented down-regulated genes.

in the egg stage were mainly enriched for DNA integration, protein folding, and protein synthesis processes such as ATP binding and ribosome. In down-regulated genes, the representative abundant enrichment GO terms included developmental process, biosynthetic process, cell differentiation, and biological regulation, intracellular membrane-bound organelle, (cellular component ontology), nucleotide acid binding, DNA binding, transcription factor activity, transcriptional regulation activity (molecular function ontology). Furthermore, we also performed KEGG enrichment analysis (Fig. 5a, Supplementary excel 4 and Table S7). FoxO signaling pathway, Hippo signaling pathway, fructose and mannose metabolism, glycerolipid metabolism, nitrogen metabolism, and glycolysis/gluconeogenesis were up-regulated, and glycerolipid metabolism, N-Glycan biosynthesis, and protein export were down-regulated.

A total of 28,711 significantly expressed genes (including 4,487 up-regulated genes and 24,224 down-regulated genes) were obtained in the third instar vs. second instar library (Fig. 4). According to GO enrichment analysis (Supplementary excel file 3), when second instar mealybug molt to the third instar many metabolic process genes

were found to be up-regulated and categorized into RNA catabolic process, ribosome biogenesis, cellular nitrogen compound catabolic process, organic cyclic compound catalytic process, etc. For the down-regulated genes, the representative abundant GO terms included developmental process, biosynthetic process, cell differentiation, and biological regulation, intracellular membrane bound organelle, (cellular component ontology), nucleotide acid binding, DNA binding, transcription factor activity, transcriptional regulation activity (molecular function ontology). In the KEGG enrichment analysis (Fig. 5a, Supplementary excel file 4 and Table S7), the three up-regulated pathways include oxidative phosphorylation, ribosome, and fatty acid biosynthesis. The down-regulated genes were primarily involved in biosynthesis of amino acids and starch and sucrose metabolism.

A comparison between third instar and adult female revealed 29,495 significantly differentially expressed genes, including 8,304 up-regulated genes and 21,191 down-regulated genes (Fig. 4). According to GO enrichment analysis (Supplementary excel file 3), when second instar molt to the third instar many metabolic process genes were found to be up-regulated and categorized into cellular process, biosynthetic process, single-organism process, macromolecular process, ribonucleoprotein complex (cellular component categories), structural molecule activity (molecular function categories). For the down-regulated genes, the abundant enrichment GO terms included biological regulation, primary metabolic process (biological process ontology), membrane-bounded vesicle, (cellular component ontology), transport, protein binding (molecular function ontology). In the KEGG analysis (Fig. 5b, Supplementary excel file 4 and Table S7), the three up-regulated pathways include phagosome, biosynthesis of amino acids, glycerolipid metabolism, ribosome, photo-transduction, and 2-oxocarboxylic acid metabolism. The down-regulated genes were primarily involved in mTOR signaling pathway and starch and sucrose metabolism.

When differential gene expression was analyzed between the adult female and EggI transcriptome, 27,539 genes exhibited significantly differential expression. Of these genes, 10,981 were up-regulated, and 16,558 were down-regulated in the adult transcriptome (Fig. 4). Based on GO enrichment analysis (Supplementary excel file 3), most of the up-regulated genes in the adult female categorized into the metabolic process, protein metabolic process and cellular biosynthetic process. Other categories, such as gene expression, cellular process (biological process ontology), ribosomal subunit (cellular component ontology), and structural molecule activity, nucleic acid binding and organic cyclic compound binding (molecular function ontology) were also enriched. On the other part, the high expressing genes in the egg stage were mainly enriched for DNA integration, protein folding, and protein synthesis process such as ribonucleoprotein complex, ATP binding, ribosome. For the down-regulated genes, the representative abundant enrichment GO terms included cell cycle process, cell differentiation (cellular component strategies), nucleoplasm part, non-membrane-bounded organelle, and chromosome organization (cellular component categories), and nucleoside phosphate binding (molecular function categories). Furthermore, we also performed KEGG analysis (Fig. 5b, Supplementary excel file 4 and Table S7), Hippo signaling pathway, oxidative phosphorylation, metabolic pathways, and glycerolipid metabolism were up-regulated, and lysine degradation, sucrose, and starch metabolism and neuroactive ligand-receptor interaction were down-regulated.

Venn diagrams of the DEG's among developmental stages allowed the identification of transcripts exclusive or common to the various stages of *P. solenopsis*. The first Venn diagram comprising DEG's between EggI and all analyzed instar identified 10,059, 8,439 and 10,981 DEG's between EggI and the second, third and adult stages, respectively. By comparing, the EggI with second instar, third instar, and adult 3,500 DEG's were identified that shared between all EggI, second, third and adult female stages (Fig. 6a). Of the 10,059 DEG'S between EggI and second instar, 3803 were exclusive whereas 1172 DEG's were shared between second and third instar. The remaining was shared among the second, third and adult female (Fig. 6a). Of the 8439 DEG's between EggI and 3rd instar, 2341 were exclusive, 1426 were shared between the third and adult stages, and 1584 were shared between the second and adult female stages (Fig. 6a).

A Venn diagram of DEG'S between second, third instar stages showed that there were 4,487, 7694 and 8304 DEG's between second and third, adult female stages, respectively (Fig. 6b). In this case, 30 DEG'S were commonly found between second, third and adult stages (Fig. 6b). Of the 4487 DEG'S between second and the third instar stage, 2733 were exclusive (Fig. 6b). Of the 7694 DEG'S between adults and second stage, 2097 were exclusive, and 3843 were also differentially expressed between adults and third instar. Of the 4487 DEG'S between second and third instar stage, 4431 were exclusive (Fig. 6b). Among the 587 DEG'S between adults and EggI stage, 117 were exclusive to this stage (Fig. 6b).

The euclidian distance matrix was estimated using normalized gene expression values of all the libraries based on transcripts profiles to create a dendrogram and a heatmap describing the similarities among the developmental stages of *P. solenopsis* (Supplementary Fig. S2). The heatmap and the clustered groups clearly indicated a similarity gradient between the developmental stages, with the transcript profile from second instar more similar to that of the third instar, and from the third instar more similar to the adult females. The gene expression profile from EggI was the most distinct from all the other stages, particularly from adults. There was less similarity among EggI and second instar than for third instar and an adult female. Among the different instars, the closer the stages, the higher the similarity of transcripts profiles.

Validation of differently expressed transcripts (DEG'S) during development using qRT-PCR.

qRT-PCR analyses were performed on ten randomly selected genes: TRINITY_DN6817_c0_g2, TRINITY_DN5360_c0_g1, TRINITY_DN25340_c3_g1, TRINITY_DN22738_c0_g1, TRINITY_DN11167_c0_g1, TRINITY_DN62673_c0_g1, TRINITY_DN35383_c0_g1, TRINITY_DN2037_c0_g2, TRINITY_DN65539_c0_g1 and TRINITY_DN73421_c0_g1. (Fig. 7 and Table S1). TRINITY_DN6817_c0_g2, TRINITY_DN5360_c0_g1, TRINITY_DN25340_c3_g1, TRINITY_DN22738_c0_g1, TRINITY_DN11167_c0_g1, TRINITY_DN62673_c0_g1, TRINITY_DN35383_c0_g1, and TRINITY_DN65539_c0_g1 were primarily expressed in all stages, whereas TRINITY_DN2037_c0_g2 & TRINITY_DN73421_c0_g1 were not showing any expression in adult stage. The high confirmation rate of the unigenes indicated the reliability of the transcriptome data.

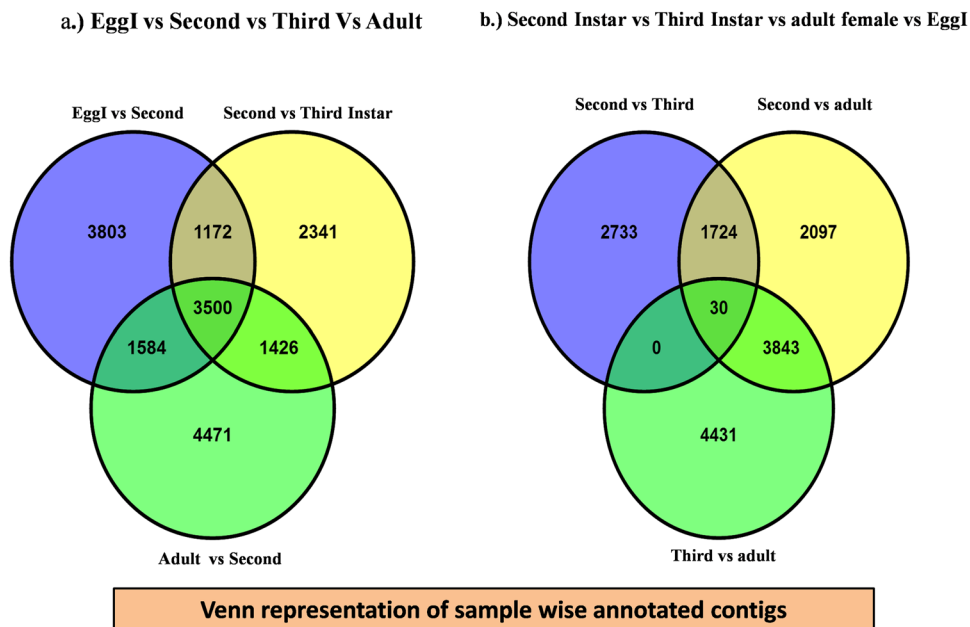


Figure 6. Venn diagrams representing number of differentially expressed genes among developmental stages. (a) Number of genes differentially expressed between the EggI vs. Second Instar vs. Third Instar vs. Adult female, commonly shared or not among stages. (b) Number of genes expressed between Second vs. Third vs. adult female, commonly shared or not among stages.

Identification of differentially expressed genes involved in hormone biosynthesis. The contigs related to the genes involved in the biosynthesis of sesquiterpenoid juvenile hormone (JH) and ecdysteroid pathways were identified by blast search against the reported genes from *A. pisum*. For JH II biosynthesis, six contigs were mapped to farnesylpyrophosphate synthase genes (Fps, Fpps2), four to juvenile hormone acid methyltransferase (Jhamt), and twelve to cytochrome P450 family 15, subfamily A, polypeptide 1 (CYP15A1) of *A. pisum* with significant similarity (identity > 60% and $E < e^{-10}$). For JH III degradation, thirteen contigs showed similarity to juvenile hormone esterase (Jhe), thirteen to juvenile hormone epoxide hydrolase (Jheh), and twenty-one to juvenile hormones esterase-like protein (Jdhk) (Fig. 8a). List of identified contigs and their expression values is shown in Table S8.

In case of ecdysteroid biosynthesis, two contigs were identified for neverland gene (cholesterol dehydrogenase). A subsequent search was performed to identify the contigs for cytochrome P450 enzymes encoded by halloween genes. A total of thirteen contigs showed similarity to spook (CYP307A), eleven to phantom (CYP306A1), nine to disembodied (CYP302A1), nine to shadow (CYP315A1), and eleven to shade (CYP314A1). For ecdysteroid inactivation, fourteen contigs showed similarity to 26-dehydroxylase (CYP18A1), which is putatively responsible for 20, 26-dihydroxyecdysone production. Twenty contigs showed similarity with 3-dehydroecdysone-3- α -reductase and fifteen with ecdysone oxidase, which is putatively responsible for 3-epiecdysone production, were identified (Fig. 8b). Complete information about the gene listed for ecdysteroid biosynthesis with mapped contigs and their expression values are given in Table S8.

Discussion

The mealybug, *P. solenopsis* is a major invasive pest of cotton destroying a vast range of agriculturally important cash crop. However, their detail genomics study has not been done so far. In the absence of complete genome sequences, de novo-transcriptome analysis can help in estimating future gene expression and functional analysis on *P. solenopsis*. That can also enhance our understanding of its biological processes and molecular mechanisms. In the current study, RNA-Seq technology was applied to reveal the comparative transcriptome profiling of four developmental stages (EggI, second, third and adult female) of *P. solenopsis*. The de-novo assembly of short read sequence without having reference genome poses a challenge with existing bioinformatics tools³⁹. Here, we used trinity software which is a well-accepted methodology for its precision and quality of the assembled sequences²⁹. Before doing assembly, the four datasets (EggI, second, third instar and adult female) were merged, and read abundance estimation was normalized to $50 \times$ coverage using the *in-silico* normalization tool of Trinity to improve assembly duration and minimize the memory need. Filtering and normalization reduced the dataset to 43.59 GB. About 182.67 million normalized read pairs were assembled using Trinity at default parameters. The 38,725 contigs annotated using NR database showed the highest similarity of *P. solenopsis* with the *A. pisum* rather than *B. tabaci*. This disparity was observed due to the close evolutionary relationship between *A. pisum* and *P. solenopsis*, which referred a common selection of genes between closely related insects. This pattern is different from the transcriptome studies of hemipterans, *B. tabaci*^{40,41}, *A. mellifera*⁴², *Myzus persicae*⁴³ and *A. gossypii*⁴⁴. These observations are likely due to non-availability of sufficient transcriptome data for the hemipteran species published during the study period.

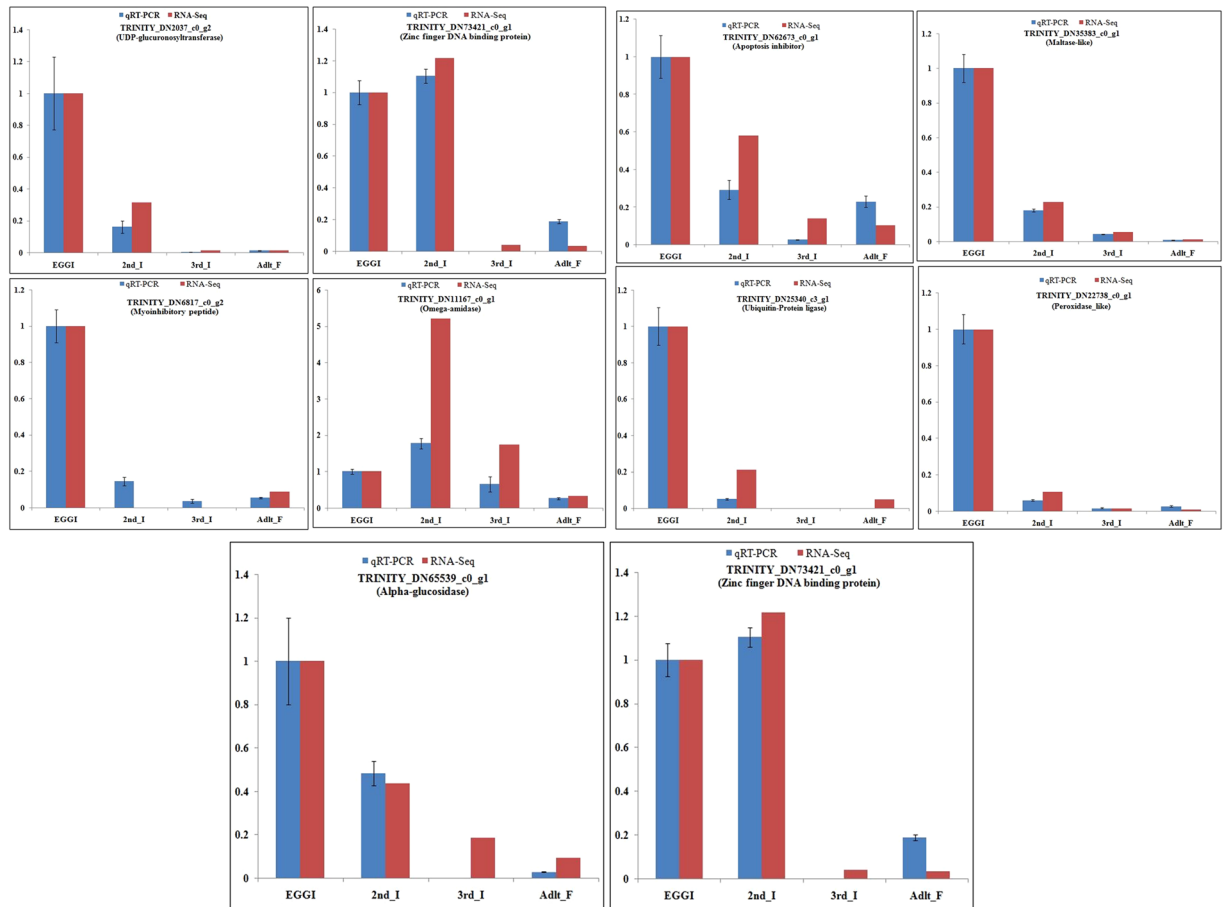


Figure 7. Comparison of qRT-PCR (red bar) and transcriptome (blue bar) expression data for ten randomly selected genes to confirm expression patterns indicated by transcriptome sequencing. Three technical replicates were performed for each of the three biological replicates. The height of each bar chart represents the mean average of sample-specific $2^{-\Delta\Delta C_t}$ values. “1” represents the egg stage, “2” represents the 2nd Instar stage, “3” represents the 3rd instar stage, and “4” represents the adult stages.

Further, we compared the gene expression profiles between four developmental stages by constructing DEGs libraries, which revealed 29,415 differentially expressed genes. The total number of unigenes of the *de-novo* assembled assembly was higher than expected total number of genes found in other hemiptera and lepidoptera species^{41,43,45}. Recently, an article was published that showed transcriptome data analysis of *P. solenopsis* to identify SSR markers and genes involved in sex pheromone metabolism^{46,47}. The reported study was basically done on female adult mealybugs, but no comparative differential analysis was performed on the insect developmental stages and hormone biosynthesis. The present study is the first report on *P. solenopsis* that explained the differential expression of the genes involved in insect development and hormone biosynthesis pathway. The information generated through the analysis provides molecular resources for the study of other hemipteran and related species that establish a framework to interpret the changes in gene expression during insect development.

An interesting finding from this study is the identification of many differentially expressed genes across the developmental stages that enriched current knowledge of *P. solenopsis* gene expression profiles. It will be useful for future research related to various developmental, physiological and metabolic pathways in this as well as in other related insects. It may also help to identify vital RNAi target for the control of *P. solenopsis*, as reported in the case of other hemipteran insects^{48–51}. During comparison between different developmental libraries, a large number of genes showed specific life stage-related expression profiles (e.g., ecdysteroid and juvenile hormone related) that were likely involved in developmental differentiations. The developmental characteristics of *P. solenopsis* are very typical and showed hemi-metabolous metamorphosis; which means insects develop from eggs to first instar nymphs, and then to second-third instar, eventually emerge as either adult male or female^{4,5,52}. The Adult female is considered as the most devastating stage for infesting host plants that lay eggs and start a new life cycle^{4,53}. During the developmental growth from eggs to first–second instars, 10,059 up-regulated genes and 21,581 down-regulated genes, third to adult stage showed 4,487 up-regulated genes, and 24,224 down-regulated genes were found. From the available differential expressed genes (DEGs), the search was performed to identify the ten-most up-regulated and down-regulated genes in different developmental comparison. The top up-regulated genes in EggI to adult female stage was found to be putative antimicrobial knottin protein of *B. tabaci*⁵⁴, eukaryotic

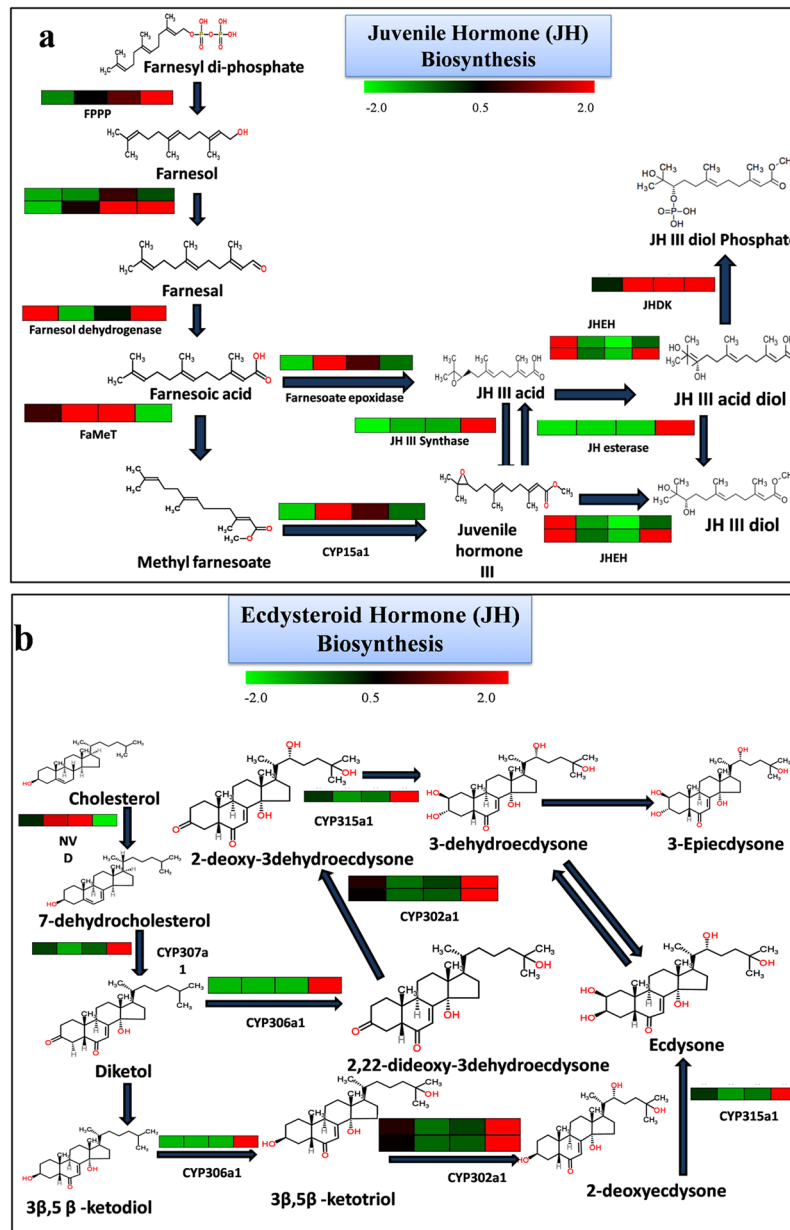


Figure 8. Juvenile hormone and ecdysteroid hormone biosynthesis with number of hits from the *P. solenopsis* transcriptome. (a) The analysis of the transcriptome identified homologues genes coding for enzymes of the Juvenile Hormone biosynthetic pathway, including Farnesylpyrophosphate synthase (Fpps); Juvenile Hormone Acid Methyltransferase (Jhamt); and Cytochrome P450 (CYP15A1); and for degradation of the Juvenile Hormone, with Juvenile Hormone Epoxide Hydrolase (Jheh); and Juvenile Hormone Esterase 1 (jhe1), with respective number of identified hits at $e < e-30$ by Blastx. Chemical structures obtained from KEGG. (b) The analysis of the transcriptome identified homologues genes coding for enzymes of the Ecdysteroid hormone biosynthetic pathway, including Cholesterol 7-dehydrogenase (Neverland); and a series of cytochrome P450 enzymes encoded by the Halloween genes, such as spook (CYP307A); phantom (CYP306A1); disembodied (CYP302A1); shadow (CYP315A1); and shade (CYP314A1). For ecdysteroid inactivation, a 26-dehydroxylase (CYP18A1), with respective number of identified hits at $e < e-30$ by Blastx. Chemical structures obtained from KEGG.

translation initiation factor 3, serine proteinase of *P. humanus corporis*⁵⁵, DnaJ domain-containing protein of *B. mori*⁵⁶⁻⁵⁸, 60 S ribosomal protein L18 of *A. pisum*, tubulin beta-3 chain-like of *A. pisum*⁵⁹, Rad21 CG17436-PA of *T. castaneum*^{60,61}, sodium-independent sulfate anion transporter-like isoform X1 of *Apis dorsata*⁶² and putative antimicrobial knottin protein Btk-4 of *B. tabaci*. The down-regulated genes were tumor protein p63 isoform alpha 2-like genes of *A. pisum*, similar to Secretory Phospholipase A2^{63,64} of *T. castaneum*, zinc finger protein SLUG⁶⁵ of *P. humanus corporis*, regulator of G-protein signaling 9-like isoform 1^{66,67} of *P. humanus corporis*, and chromobox protein homolog 5-like isoform 1⁶⁸ of *Nasonia vitripennis*. Furthermore, to identify significant GO categories and KEGG pathways in different developmental stages of *P. solenopsis* GO, and KEGG enrichment analyses

were performed. GO enrichment analysis show that up regulated developmental specific genes belong to ‘protein metabolic process’, ‘biosynthetic process’, ‘organic substance metabolic process’, ‘ribosome biogenesis’, ‘protein binding’, ‘structural molecule activity’ and ‘protein targeting’. According to the KEGG pathway analysis, fifteen, six, and eleven pathways were observed as significantly enriched for DEGs of EggI vs. second instar, second vs. third instar, and third instar vs. adult females, respectively. The pathway enriched for different developmental stages were glycerolipid metabolism⁶⁹, FoxO-signaling pathway^{20,70–73}, Hippo signaling pathway^{74–76}, fatty acid biosynthesis^{77–79}, biosynthesis of amino acids^{41,80–85} and steroid biosynthesis^{86–88} suggesting that genes in these pathways seem to be involved in early developmental process. Likewise, we have identified all the genes coding for enzymes involved in juvenile and ecdysteroid hormone biosynthetic pathways and found many of them as potential RNAi targets for silencing (Fig. 8). In JH III biosynthesis pathway, there were twelve to twenty contigs homologous to each gene from *A. pisum* and *M. hirsutus*, while 10–13 contigs were recognized for the degrading enzymes for JHE and JHEH. In case of ecdysteroid biosynthesis which completely depends on sterols obtained from the diet, involves consecutive action of dehydrogenase (neverland), followed by a series of cytochrome P450 enzymes encoded by the halloween genes. In the insect hormone biosynthesis pathway, all the genes that encode enzymes of the ecdysteroid and sesquiterpenoid juvenile hormone biosynthetic pathways can be a vital target for RNAi and other silencing mechanism. By regulating the developmental pathway gene expression would be a challenging and attractive RNAi method to control the invasiveness of sap-sucking pest, *P. solenopsis*.

This study generated four transcriptome libraries and gave information regarding the genomic resources for further evaluation of *P. solenopsis*. The information generated through analyses will help in future to identify the potential targets in relation to developmental pathways. Moreover, many unigenes and expression profiles of high-quality transcriptome datasets of *P. solenopsis* serve as a step towards understanding the role of different stages in the spread of mealybug and their infestation on agriculturally important cash crops for their control.

References

- Hodgson, C. J., Abbas, G., Arif, M. J., Saeed, S. & Karar, H. *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Coccoidea: Pseudococcidae), an invasive mealybug damaging cotton in Pakistan and India, with a discussion on seasonal morphological variation. *Zootaxa* **1913**, 1e35 (2008).
- Abbas, G., Arif, M. J. & Saeed, S. Systematic status of a new species of the genus *Phenacoccus* Cockerell (Pseudococcidae), a serious pest of cotton, *Gossypium hirsutum* L., in Pakistan. *Pak. Entomol.* **27**, 83e84 (2005).
- Abbas, G., Arif, M. J., Ashfaq, M., Aslam, M. & Saeed, S. Host plants, distribution and overwintering of cotton mealybug (*Phenacoccus solenopsis*) Homoptera: Pseudococcidae. *Int. J. Agric. Biol.* **12**, 421e425 (2010).
- Vennila, S. *et al.* Biology of the mealybug, *Phenacoccus solenopsis* on cotton in the laboratory. *Journal of Insect Science* **10**(1), 115 (2010).
- Nagrare, V. S. *et al.* Compendium of Cotton Mealybugs. *Central Institute for Cotton Research* Nagpur, India, 42 pp (2011).
- Arif, M. I., Rafiq, M., Wazir, S., Mehmood, N. & Ghaffar, A. Studies on cotton mealybug, *Phenacoccus solenopsis* (Pseudococcidae: Homoptera), and its natural enemies in Punjab, Pakistan. *International Journal of Agriculture & Biology* **14**(4) (2012).
- Tang, Y., He, M. & Lu, Y. Acquisition, retention and transmission risk of Cotton leaf curl Multan virus by cotton mealybug *Phenacoccus solenopsis* Tinsley. *Journal of Environmental Entomology* **38**(4), 736–741 (2016).
- Dhawan, A. K. *et al.* Incidence and damage potential of mealy bug, *Phenacoccus solenopsis* Tinsley on cotton in Punjab. *Indian Journal of Ecology* **34**, 166–72 (2007).
- Jhala, R. C., Bharpoda, T. M. & Patel, M. G. *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae), the mealy bug species recorded first time on cotton and its alternate host plants in Gujarat, India. *Uttar Pradesh Journal of Zoology* **28**(3), 403–406 (2008).
- Nagrare, V. S. *et al.* Widespread infestation of the exotic mealybug species, *Phenacoccus solenopsis* (Tinsley) (Hemiptera: Pseudococcidae), on cotton in India. *Bulletin of entomological research* **99**(5), 537–541 (2009).
- Afzal, M. B. & Shad, S. A. Resistance inheritance and mechanism to emamectin benzoate in *Phenacoccus solenopsis* (Homoptera: Pseudococcidae). *Crop Protection* **71**, 60–65 (2015).
- Saddiq, B., Abbas, N., Shad, S. A., Aslam, M. & Afzal, M. B. Deltamethrin resistance in the cotton mealybug, *Phenacoccus solenopsis* Tinsley: Cross-resistance to other insecticides, fitness cost analysis and realized heritability. *Phytoparasitica* **44**(1), 83–90 (2016).
- Ejaz, M. & Ali, S. S. Spirotetramat Resistance Selected in the *Phenacoccus solenopsis* (Homoptera: Pseudococcidae): Cross-Resistance Patterns, Stability, and Fitness Costs Analysis. *Journal of Economic Entomology* **110**(3), 1226–1234 (2017).
- Truman, J. W. & Riddiford, L. M. Endocrine insights into the evolution of metamorphosis in insects. *Annual review of entomology* **47**(1), 467–500 (2002).
- Riddiford, L. M. How does juvenile hormone control insect metamorphosis and reproduction? *General and comparative endocrinology* **179**(3), 477–484 (2012).
- Ishimaru, Y. *et al.* TGF- β signaling in insects regulates metamorphosis via juvenile hormone biosynthesis. *Proceedings of the National Academy of Sciences* **113**(20), 5634–9 (2016).
- Nijhout, H. F. *Insect Hormones*. Princeton Univ. Press, Princeton, NJ (1994).
- Riddiford, L. M. Cellular and molecular actions of juvenile hormone I. General considerations and premetamorphic actions. *Advances in insect physiology* **24**, 213–274 (1994).
- Klowden, M. J. *Physiological systems in insects*. Academic Press (2013).
- Noriega, F. G. Juvenile hormone biosynthesis in insects: what is new, what do we know, and what questions remain? *International scholarly research notices* (2014).
- Goodman, W. G., Granger, N. A., Gilbert, L. I., Iatrou, K. & Gill, S. S. The juvenile hormones. *Insect Dev.: Morphogenes. Molting Metamorph* **8**, 305 (2009).
- Sen, S. E., Ewing, G. J. & Thursten, N. Characterization of lepidopteran prenyltransferase in *Manduca sexta* corpora allata. *Archives of insect biochemistry and physiology* **32**(3/4), 315–332 (1996).
- Cusson, M. *et al.* Characterization and tissue-specific expression of two lepidopteran farnesyl diphosphate synthase homologs: Implications for the biosynthesis of ethyl-substituted juvenile hormones. *Proteins: Structure, function, and bioinformatics* **65**(3), 742–758 (2006).
- Pankotai, T. *et al.* Genes of the ecdysone biosynthesis pathway are regulated by the dATAC histone acetyltransferase complex in *Drosophila*. *Molecular and cellular biology* **30**(17), 4254–66 (2010).
- Warren, J. T. *et al.* Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* **99**(17), 11043–8 (2002).
- Gilbert, L. I. Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*. *Mol. Cell Endocrinol* **215**, 1–10 (2004).

27. De Loof, A. Ecdysteroids, juvenile hormone and insect neuropeptides: recent successes and remaining major challenges. *General and comparative endocrinology* **155**(1), 3–13 (2008).
28. Arya, S. K. *et al.* Reference genes validation in *Phenacoccus solenopsis* under various biotic and abiotic stress conditions. *Scientific Reports* **7**(1), 13520 (2017).
29. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS one* **7**(2), e30619 (2012).
30. Grabherr, M. G. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**(7), 644–652 (2011).
31. Kanehisa, M. KEGG for linking genomes to life and the environment. *Nucleic acids research* **36**(suppl_1), D480–D484 (2007).
32. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**(1), 323 (2011).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1), 139–140 (2010).
34. Wu, J., Mao, X., Cai, T., Luo, J. & Wei, L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic acids research* **34**(suppl_2), W720–W724 (2006).
35. Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research* **39**(suppl_2), W316–W322 (2011).
36. Hao, Q. N. *et al.* Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. *BMC genomics* **12**(1), 525 (2011).
37. Shen, Y. *et al.* Genome expression profile analysis of the immature maize embryo during dedifferentiation. *PLoS one* **7**(3), e32237 (2012).
38. JMP[®], Version 13.0. SAS Institute Inc., Cary, NC (1989–2007).
39. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nature Reviews Genetics* **12**(10), 671–682 (2011).
40. Wang, X. W. *et al.* De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC genomics* **11**(1), 400 (2010).
41. Upadhyay, S. K. *et al.* Whitefly genome expression reveals host-symbiont interaction in amino acid biosynthesis. *PLoS one* **10**(5), e0126751 (2015).
42. Wu, Y. *et al.* Comparative transcriptome analysis on the synthesis pathway of honey bee (*Apis mellifera*) mandibular gland secretions. *Scientific Reports* **7**(1), 4530 (2017).
43. Ji, R. *et al.* Transcriptome analysis of green peach aphid (*Myzus persicae*): Insight into developmental regulation and inter-species divergence. *Frontiers in plant science* **7** (2016).
44. Li, Z. Q. *et al.* Ecological adaption analysis of the cotton aphid (*Aphis gossypii*) in different phenotypes by transcriptome comparison. *PLoS one* **8**(12), e83180 (2013).
45. Park, Y. *et al.* Analysis of transcriptome data in the red flour beetle. *Tribolium castaneum*. *Insect biochemistry and molecular biology* **38**(4), 380–6 (2008).
46. Luo, M. *et al.* High-throughput discovery of SSR genetic markers in the mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae), from its transcriptome database. *Acta Entomologica Sinica* **57**(4), 395–400 (2014).
47. Hu, J. *et al.* Transcriptome analysis of the cotton mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae). *Acta Entomologica Sinica* **60**(1), 9–17 (2017).
48. Upadhyay, S. K. *et al.* RNA interference for the control of whiteflies (*Bemisia tabaci*) by oral route. *Journal of biosciences* **36**(1), 153–61 (2011).
49. Upadhyay, S. K. *et al.* siRNA machinery in whitefly (*Bemisia tabaci*). *PLoS One* **8**(12), e83692 (2013).
50. Shukla, A. K. *et al.* Expression of an insecticidal fern protein in cotton protects against whitefly. *Nature biotechnology* **34**(10), 1046–51 (2016).
51. Thakur, N. *et al.* Enhanced whitefly resistance in transgenic tobacco plants expressing double stranded RNA of v-ATPase A gene. *PLoS one* **9**(3), e87235 (2014).
52. Nikam, N. D., Patel, B. H. & Korat, D. M. Biology of invasive mealy bug, *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae) on cotton. *Karnataka Journal of Agricultural Sciences* **23**(4), 649–651 (2010).
53. Khan, A. M., Ashfaq, M., Khan, A. A., Naseem, M. T. & Mansoor, S. Evaluation of potential RNA-interference-target genes to control cotton mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae). *Insect science* (2017).
54. Shatters, R. G. Jr. *et al.* A knottin-like putative antimicrobial gene family in the whitefly *Bemisia tabaci* biotype B: Cloning and transcript regulation. *Journal of Insect Science* **8** (2008).
55. Bao, Y. Y. *et al.* Comparative analysis of *Bombyx mori* nucleopolyhedrovirus responsive genes in fat body and haemocyte of B. mori resistant and susceptible strains. *Insect molecular biology* **19**(3), 347–358 (2010).
56. Li, Y. *et al.* Cloning and analysis of DnaJ family members in the silkworm. *Bombyx mori*. *Gene* **576**(1), 88–98 (2016).
57. Sirigineedi, S., Vijayagowri, E., Murthy, G. N., Rao, G. & Ponnuvel, K. M. Molecular characterization of DnaJ 5 homologs in silkworm *Bombyx mori* and its expression during egg diapause. *Insect science* **21**(6), 677–686 (2014).
58. Tang, X. *et al.* Identification of sumoylated proteins in the silkworm *Bombyx mori*. *International journal of molecular sciences* **15**(12), 22011–27 (2014).
59. Dettman, R. W., Turner, F. R., Hoyle, H. D. & Raff, E. C. Embryonic expression of the divergent *Drosophila* β 3-tubulin isoform is required for larval behavior. *Genetics* **158**(1), 253–263 (2011).
60. Zhang, N. *et al.* Characterization of the interaction between the cohesin subunits Rad21 and SA1/2. *PLoS One* **8**(7), e69458 (2013).
61. Hallson, G. *et al.* The *Drosophila* cohesin subunit Rad21 is a trithorax group (trxG) protein. *Proceedings of the National Academy of Sciences* **105**(34), 12405–10 (2008).
62. Wu, M. C., Chang, Y. W., Lu, K. H. & Yang, E. C. Gene expression changes in honey bees induced by sublethal imidacloprid exposure during the larval stage. *Insect Biochemistry and Molecular Biology* **88**, 12–20 (2017).
63. Stanley, D. The non-venom insect phospholipases A 2. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1761**(11), 1383–1390 (2006).
64. Shrestha, S. & Kim, Y. Biochemical characteristics of immune-associated phospholipase A 2 and its inhibition by an entomopathogenic bacterium. *Xenorhabdus nematophila*. *The Journal of Microbiology* **47**(6), 774–782 (2009).
65. Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nature reviews. Molecular cell biology* **3**(3), 155 (2002).
66. Cai, M. J. *et al.* G-protein-coupled receptor participates in 20-hydroxyecdysone signaling on the plasma membrane. *Cell Communication and Signaling* **12**(1), 9 (2014).
67. Bai, H. & Palli, S. R. Identification of G protein-coupled receptors required for vitellogenin uptake into the oocytes of the red flour beetle, *Tribolium castaneum*. *Scientific reports* **6**, 27648 (2016).
68. Caragata, E. P. *et al.* The transcriptome of the mosquito *Aedes fluviatilis* (Diptera: Culicidae), and transcriptional changes associated with its native *Wolbachia* infection. *BMC genomics* **18**(1), 6 (2017).
69. XueKe, G. *et al.* Lipidomics and RNA-Seq Study of Lipid Regulation in *Aphis gossypii* parasitized by *Lysiphlebia japonica*. *Scientific Reports* **7**(1), 1364 (2017).
70. Sim, C. & Denlinger, D. L. Insulin signaling and the regulation of insect diapause. *Frontiers in physiology* **4** (2013).
71. Vafopoulou, X. The coming of age of insulin-signaling in insects. *Frontiers in physiology* **5** (2014).

72. Zhang, X. S., Wang, T., Lin, X. W., Denlinger, D. L. & Xu, W. H. Reactive oxygen species extend insect life span using components of the insulin-signaling pathway. *Proceedings of the National Academy of Sciences* **114**(37), E7832–E7840 (2017).
73. Zeng, B. *et al.* The FOXO transcription factor controls insect growth and development by regulating juvenile hormone degradation in the silkworm, *Bombyx mori*. *Journal of Biological Chemistry* jbc-M117 (2017).
74. Gotoh, H. *et al.* The Fat/Hippo signaling pathway links within-disc morphogen patterning to whole-animal signals during phenotypically plastic growth in insects. *Developmental Dynamics* **244**(9), 1039–45 (2015).
75. Hayashi, S., Tamura, K. & Yokoyama, H. Yap1 Transcription regulator in the Hippo signaling pathway, is required for *Xenopus* limb bud regeneration. *Developmental biology* **388**(1), 57–67 (2014).
76. Dong, D. J., Jing, Y. P., Liu, W., Wang, J. X. & Zhao, X. F. The Steroid Hormone 20-Hydroxycydysone Up-regulates Ste-20 Family Serine/Threonine Kinase Hippo to Induce Programmed Cell Death. *Journal of Biological Chemistry* **290**(41), 24738–24746 (2015).
77. Tan, Q. Q., Liu, W., Zhu, F., Lei, C. L. & Wang, X. P. Fatty acid synthase 2 contributes to diapause preparation in a beetle by regulating lipid accumulation and stress tolerance genes expression. *Scientific reports* **7**, 40509 (2017).
78. Haritos, V. S. *et al.* The convergent evolution of defensive polyacetylenic fatty acid biosynthesis genes in soldier beetles. *Nature communications* **3**, ncomms2147 (2012).
79. Li, X. *et al.* Molecular characterization of two fatty acyl-CoA reductase genes from *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae). *Journal of Insect Science* **16**(1) (2016).
80. Dhadialla, T. S. & Raikhel, A. S. Biosynthesis of mosquito vitellogenin. *Journal of Biological Chemistry* **265**(17), 9924–9933 (1990).
81. Douglas, A. E. How multi-partner endosymbioses function. *Nature Reviews Microbiology* **14**(12), 731–743 (2016).
82. Russell, C. W., Bouvaine, S., Newell, P. D. & Douglas, A. E. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Applied and environmental microbiology* **79**(19), 6117–6123 (2013).
83. Chen, P. S. Amino acid and protein metabolism in insect development. *Advances in insect physiology* **3**, 53–132 (1966).
84. Wilson, A. C. & Duncan, R. P. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proceedings of the National Academy of Sciences* **112**(33), 10255–10261 (2015).
85. Hansen, A. K. & Moran, N. A. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences* **108**(7), 2849–2854 (2011).
86. Niwa, Y. S. & Niwa, R. Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. *Development, growth & differentiation* **58**(1), 94–105 (2016).
87. Yamanaka, N. *et al.* Regulation of insect steroid hormone biosynthesis by innervating peptidergic neurons. *Proceedings of the National Academy of Sciences* **103**(23), 8622–7 (2006).
88. Ou, Q. *et al.* The insect prothoracic gland as a model for steroid hormone biosynthesis and regulation. *Cell reports* **16**(1), 247–62 (2016).

Acknowledgements

P.C.V. is grateful to the SERB-Department of Science & Technology (DST), Government of India for the financial support. S.K.A. is grateful to UGC for Senior Research Fellowship. Y.V.D. is thankful to CSIR-Empower project for financial support. S.K.U. is thankful to DST for DST-INSPIRE Faculty Fellowship (IFA12-LSPA-09).

Author Contributions

S.K.A. performed sample collection, RT-qPCR, data analysis, lab experiments and drafted the manuscript. Y.V.D. performed data analysis. S.K.U. and M.H.A. helped in preparation of manuscripts and lab experiments. P.C.V. conceived the study, participated in its design and coordination of work, data analysis and interpretation and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-25845-3>.

Competing Interests: The authors declare 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) 2018