Data in Brief 59 (2025) 111366

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Walnut husk transcriptome dataset of codling moth (*Cydia pomonella*) infestation at different times

Xiaoyan Cao[a], Xiaoqin Ye[b], Adil Sattar[b,*]

[a] College of Horticulture, Xinjiang Agricultural University, China
[b] College of Forestry and Landscape Architecture, Xinjiang Agricultural University, China

ARTICLE INFO

Article history:
Received 20 October 2024
Revised 13 January 2025
Accepted 30 January 2025
Available online 6 February 2025

Dataset link: RNA sequencing of walnut husk infected by codling moth at different times (Original data)

Keywords:
Walnut husk
Illumina NovaSeq 6000
RNA-seq
Transcriptome
Differential gene expression

ABSTRACT

Walnuts, along with almonds, cashews, and hazelnuts, are renowned as the world's "four famous nuts," with walnuts being the foremost among them. Walnut fruit is rich in nutrients, including proteins, fats, polyphenols, sugars, phospholipids, melatonin, sterols, flavonoids, iron, zinc, manganese, and other trace elements, as well as dietary fiber. However, the codling moth poses a significant threat to walnut fruits as a major pest. Despite its importance, the transcriptomic changes in walnut husk at different times of codling moth infestation have not been fully explored. In this study, we employed the Illumina NovaSeq 6000 platform to sequence the transcriptome of walnut husk at various time points (0, 12, 24, 36, 48, and 72 hours) after codling moth infestation. The RNA-seq libraries yielded between 41,402,492 and 48,358,932 clean reads, resulting in a total of 120.34 Gb of clean data after filtering out low-quality reads. In total, 936 million reads were generated, with approximately 90% aligning uniquely to the reference genome. Differential expression analysis revealed the number of differentially expressed genes (DEGs) at each time point, including 21 genes associated with plant hormone synthesis. The results of this study provide new insights into the transcriptional changes in walnut husk induced by codling moth infestation and lay a foundation for future research on walnut husk defense mechanisms. The

* Corresponding author at: College of Forestry and Landscape Architecture, Xinjiang Agricultural University, Urumqi 830052, China.
E-mail address: adl1968@126.com (A. Sattar).

https://doi.org/10.1016/j.dib.2025.111366

raw FASTQ files from this transcriptome experiment are publicly available in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1140835.

Specifications Table

Subject	Horticulture
Specific subject area	Transcriptome analysis of codling moth damage on walnut husk at different times
Type of data	table, figure
Data collection	RNA sequencing, Illumina NovaSeq 6000
Data source location	Institution: Xinjiang Agricultural Universit, Urumqi, China
Data accessibility	Repository name: NCBI SRA Data identification number: PRJNA1140835 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1140835/
Related research article	Cao X, Ye X, Sattar A. 2024. Transcriptomic and coexpression network analyses revealed the regulatory mechanism of *Cydia pomonella* infestation on the synthesis of phytohormones in walnut husks. PeerJ 12: e18130 https://doi.org/10.7717/peerj.18130.

1. Value of the Data

- Transcriptome data sets provide valuable insights into the transcriptional regulation of key genes of walnut husk following damage by codling moth at various time points. These genes are also involved in defense-related metabolic pathways, such as biosynthesis of salicylic acid, abscisic acid, and jasmonic acid.
- Transcriptome data sets can help study the defense mechanisms of walnut husk produced at different times by codling moth infestation. It can serve as reference data for researchers studying walnut peel defenses.
- These data can be utilized in subsequent comparative genomics studies aimed at understand the defense response of walnut husk to codling moth infestation.

2. Background

Walnuts (*Juglans regia* L.) are a strategically important species and a vital nutritional resource for humans, owing to the high protein and oil content in their kernels. These trees are widely distributed throughout China and are commercially cultivated in various regions, including Yunnan, Shaanxi, Beijing, Xinjiang [1]. The codling moth (*Cydia pomonella*) is a worldwide quarantine pest. In 2018, we first discovered a codling moth infestation in walnuts in Hotan, Xinjiang. However, there is very limited information on the defence response of walnut husk produced at different times by codling moth infestation. Currently, no studies have reported on the gene expression profiles of walnut husk by codling moth infestation at different times. The gene expression changes in walnut husk caused by codling moth infestation at different time points were investigated using RNA-seq analysis, and the expression of genes related to metabolite synthesis pathways was compared across different infestation times.

3. Data Description

In this study, walnut husks infested by codling moth were sampled for sequencing at 0, 12, 24, 36, 48, and 72 hours. Each time point was biologically replicated three times, result-

Table 1
RNA-seq samples and read metrics.

Sample name	Raw Reads	Clean Reads	Reads mapped	Unique mapped	Q20(%)
0h-1	45,746,320	44,433,842	42,597,184	40,954,562	97.38
0h-2	45,431,288	44,154,760	42,266,546	40,816,598	97.33
0h-3	47,224,592	45,950,178	44,151,284	42,415,399	97.58
12h-1	45,626,090	44,191,214	42,297,920	40,664,156	97.23
12h-2	43,515,072	42,135,160	40,245,980	38,616,503	97.31
12h-3	45,636,166	44,043,688	41,886,696	40,266,051	96.86
24h-1	45,503,982	42,871,842	41,027,411	39,411,438	97.68
24h-2	47,447,984	44,412,704	42,000,597	40,466,907	96.64
24h-3	46,727,046	43,779,874	41,778,376	39,886,722	97.11
36h-1	44,010,636	41,402,492	39,491,654	37,800,246	97.37
36h-2	48,329,132	45,531,266	43,546,565	41,993,724	97.2
36h-3	49,044,344	47,480,188	45,374,439	43,707,649	97.27
48h-1	48,298,040	46,945,492	44,847,718	42,546,922	97.32
48h-2	49,500,666	48,358,932	46,404,137	43,605,727	97.55
48h-3	45,353,270	43,417,344	41,114,352	39,120,456	96.88
72h-1	47,800,876	46,046,684	43,833,187	41,878,278	97.04
72h-2	46,374,098	43,799,000	41,463,698	39,362,466	97.15
72h-3	45,567,394	43,363,084	41,398,571	39,559,972	97.29

Table 2
The identified phytohormone-related genes, along with their corresponding enzymes and the number of associated transcripts involved in the defense response of walnut husk.

Phytohormone	Gene	Enzyme	No of Transcripts
ABA	ZEP	zeaxanthin epoxidase	2
	NCED	9-cis-epoxy carotenoid dioxygenase	3
	ABA2	xanthoxin dehydrogenase	3
JA	LOX	lipoxygenase	1
	AOS	allene oxide synthase	2
	AOC	allene oxide cyclase	1
	OPR	12-oxophytodienoate reductase	4
SA	PAL	phenylalanine ammonia lyase	5

ing in a total of 18 sequencing libraries. Sequencing was performed using the Illumina NovaSeq 6000 platform. The raw sequence read data were deposited at Sequence Read Archive (SRA) database of NCBI BioProject under the accession of PRJNA1140835. The raw and filtered read data were summarized in Table 1. The number of differentially expressed genes (DEGs) was obtained according to the mapping of the filtered reads on walnut reference genome and are represented in the Fig. 1. The top 20 KEGG pathways for each time point are shown in Fig. 2. The phytohormone-related genes and the number of transcripts associated with them are listed in Table 2.

4. Experimental Design, Materials and Methods

4.1. Sample collection

The walnut fruit that had been infested at different times by codling moth was sampled in March 2022 at Hotan County. The damaged walnut husk was removed from the 1 cm^2 area around the borehole with a scalpel, immediately frozen in liquid nitrogen, and stored at -80°C until used for RNA extraction.

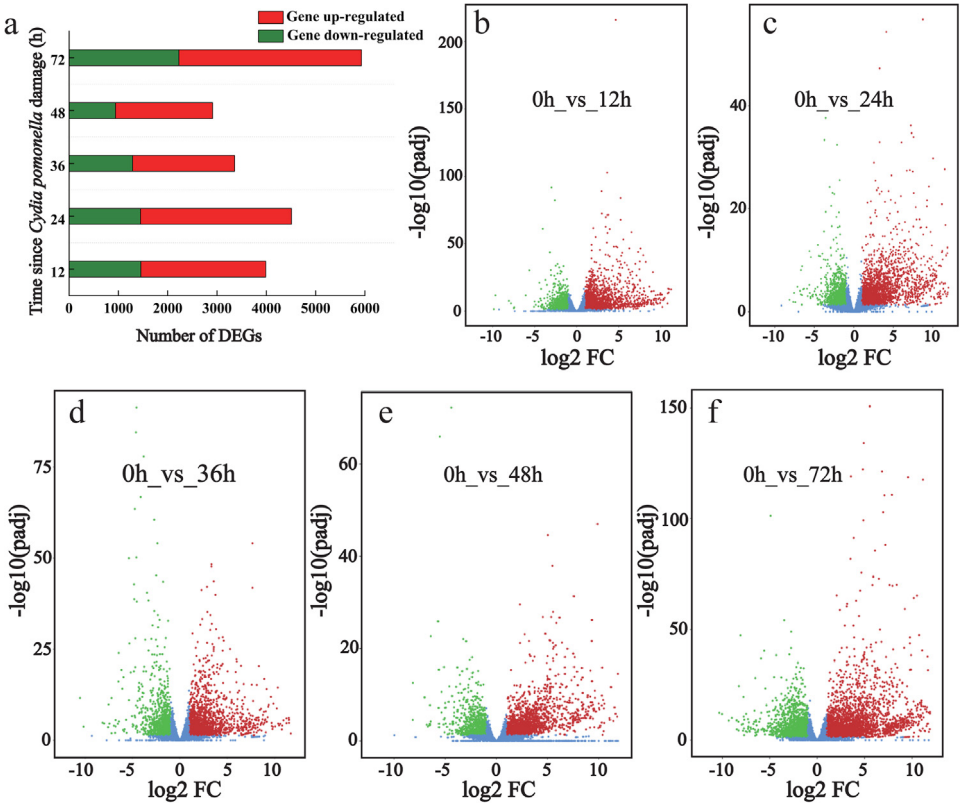

Fig. 1. Overview of differentially expressed genes (DEGs). (a) The number of differentially expressed genes (DEGs). (b-f) Volcano plot comparing transcripts expression of walnut husk at different times of damage by codling moth. The red and green dots represent significantly differentially expressed transcripts; whereas the blue dots were not significantly differentially expressed.

4.2. Transcriptome sequencing

Total RNA of the walnut husk was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA). The RNA quality was detected by a NanoPhotometer spectrophotometer (IMPLEN, CA, USA), Qubit 2.0 Fluorometer (Life Technologies, CA, USA), and Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The poly(A) mRNA was enriched by magnetic beads with oligo (dT). The mRNA was randomly fragmented. First-strand cDNA was synthesized using the M-MuLV reverse transcriptase system. The RNA strand was then degraded by RNase H, and second-strand cDNA was synthesized using DNA polymerase. The double-stranded cDNAs were ligated to sequencing adapters. The cDNAs (~200 bp) were screened using AMPure XP beads. After amplification and purification, cDNA libraries were obtained and sequenced using the Illumina Novaseq6000 system [2,3].

4.3. Data analysis

4.3.1. Data quality control

Use fastp v 0.19.3 to filter the original data, mainly to remove reads with adapters; when the N content in any sequencing reads exceeds 10% of the base number of the reads, remove

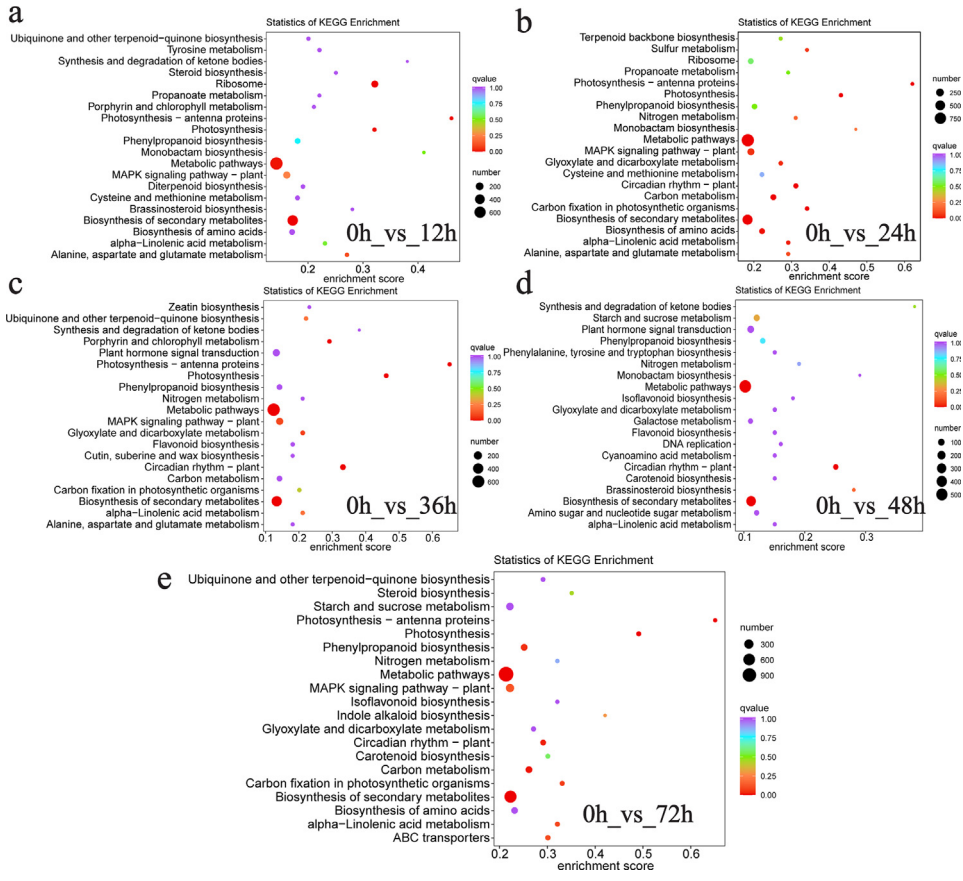

Fig. 2. Top 20 KEGG enrichment pathways at different time points.

the paired reads; when any sequencing reads when the number of low-quality (Q≤20) bases contained in reads exceeds 50% of the bases of the reads, this paired reads will be removed and obtained clean reads.

4.3.2. Differentially expressed genes

HISAT v2.1.0 was used to construct the index and compare the clean reads to the reference genome [4]. DESeq2 v1.22.1 was used to analyze the differential expression between the two groups, and the P value was corrected using the Benjamini & Hochberg method. The DEGs were identified by an absolute value of log2(fold change) ≥1 and a false discovery rate (FDR) <0.05, and further utilised to form a volcano plot to visualize and highlight the degree of which genes are significantly differentially expressed.

4.3.3. Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis

The KEGG Mapper –Search and Color Pathway software was employed for pathway analysis. KOBA software was used for Pathway annotation, and the Pathway in KEGG database was taken as the unit, hypergeometric test was used to find out the Pathway significantly enriched in differentially expressed genes compared with the whole genome background, and then Pathway significant enrichment analysis was conducted [5].

Limitations

None.

Ethics Statement

This study does not involve human subjects, animal experiments, and any data collected from social media platforms.

Data Availability

RNA sequencing of walnut husk infected by codling moth at different times (Original data) (NCBI).

CRediT Author Statement

Xiaoyan Cao: Conceptualization, Investigation, Formal analysis, Writing – original draft; **Xiaoqin Ye:** Formal analysis; **Adil Sattar:** Writing – review & editing.

Acknowledgements

This study was funded by the Key Research and Development Projects in the Xinjiang Uygur Autonomous Region (2021B02004).

Declaration of Competing Interest

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

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2025.111366.

References

[1] G. Rao, K. Sui, J. Zhang, Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts (*Juglans regia* L.), Biol. Open 6 (2016) 829–836, doi:10.1242/bio.017863.
[2] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (2016) 1–19, doi:10.1186/s13059-016-0881-8.
[3] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (1) (2009) 57–63, doi:10.1038/nrg2484.
[4] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360, doi:10.1038/nmeth.3317.
[5] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, KEGG for linking genomes to life and the environment, Nucl. Acids Res. 36 (suppl_1) (2007) D480–D484 no., doi:10.1093/nar/gkm882.