



OPEN

DATA DESCRIPTOR

Gene expression atlas of the Colorado potato beetle (*Leptinotarsa decemlineata*)

Léonore Wilhelm¹, Yangzi Wang^{1,2} & Shuqing Xu¹✉

The Colorado potato beetle (CPB) is a major pest of potato crops, known for its remarkable ability to develop resistance to more than 50 pesticides. For decades, CPB has served as a model species for studying insecticide resistance, insect physiology, diapause, reproduction, and evolution. However, research progress on CPB has been hindered by the lack of comprehensive genomic and transcriptomic resources. Here, leveraging a recently established chromosome-level genome assembly, we constructed a gene expression atlas of CPB using transcriptomic data from 61 samples representing major organs and developmental stages. By integrating short- and long-read sequencing technologies, we enhanced the genome annotation and identified 6,623 additional genes that were previously undetected. Furthermore, we developed a web portal to facilitate the search and visualization of the gene expression atlas, providing an accessible resource for the research community. The CPB gene expression atlas offers valuable tools and comprehensive data that will accelerate future research in pest control and insect biology.

Background & Summary

The Colorado potato beetle (CPB) is a devastating pest originating from North America that has spread in Europe and Asia in the 20th century. It causes substantial damage to solanaceous plants, in particular potatoes, as one CPB can consume around 40 cm² of leaf tissue during the larval stages and 100 cm² leaf area every 10 days during its adult life¹. The CPB exhibits remarkable resistance to pesticides. More than 50 active compounds proved to be ineffective against various tested populations (Mota-Sanchez and Wise 2017 Arthropod Pesticide Resistance Database). Studies have shown that resistance to many insecticides evolved rapidly in response to the applications of insecticides and showed geographical variation^{2–6}. This is largely due to high genetic variation in the population⁷, likely as the result of CPB's high fecundity. A mated female CPB can produce 25 eggs per day on average⁸ and up to 724 eggs in 30 days⁹, generating a large pool of individuals upon which selection can act. Adaptation to pesticides appears to be polygenic, involving genes related to detoxification, cuticle composition, and neuronal receptors^{7,10}.

The CPB has been the object of extended studies, most of which are related to understanding the mechanisms of pesticide resistance. Agronomic studies on the CPB include bioassays with potential pesticides^{11,12}, use of biocontrol agents¹³, and RNA interference^{14–17}. Recent studies have focused on finding alleles linked with resistance^{18,19} and gene expression linked with resistance²⁰. The CPB has also been used as a model to understand insect development and physiology^{21–23}, insect diapause^{24–27}, and more recently insect lipid metabolism and calcium signalling^{28,29} as well as behaviour³⁰.

Despite the abundance of research topics addressed using the CPB, the research in CPB is currently constrained by the lack of accessible genomic and transcriptomic data. The chromosome-level reference genome has been released recently and several studies on population genomics of CPB have been carried out^{31,32}. Yet, we currently have no publicly available gene expression data among tissues and developmental stages, which is vital for understanding the genetic mechanisms of most phenotypic traits in CPB. Here, we sequenced the transcriptomes of 12 tissues in adults (five from female, seven from male), five tissues in the last instar larvae, as well as the eggs and entire body of each larval stage. Using these data, we further improved the genome annotation and established a gene expression atlas of CPB (Table 1). The updated genome annotation contains 34,350 genes with

¹Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University, 55128, Mainz, Germany. ²Institute for Evolution and Biodiversity, University of Münster, 48161, Münster, Germany. ✉e-mail: shuqing.xu@uni-mainz.de

Annotation version	Data used	Gene number	Busco score
The i5k Initiative	<ul style="list-style-type: none"> • 13 RNA-seq samples of whole-body of adults, larvae and tissues. • Protein homologs from the Insecta GenBank database and 7 RefSeq proteomes 	16533	C:93.9%[S:92.2%,D:1.7%],F:3.3%,M:2.7%,n:2124
Yan <i>et al.</i> ³¹	<ul style="list-style-type: none"> • 13 RNA-seq samples of whole-body of adults and 1st instar larvae • Protein homologs of all insects downloaded from OrthoDB 	27692	C:93.5%[S:86.0%,D:7.5%],F:1.0%,M:5.5%,n:1367
This paper	<ul style="list-style-type: none"> • 17 RNA-seq samples from different tissues and developmental stages (Supplementary Table 2) • Protein homologs from <i>D. melanogaster</i> and <i>T. castaneum</i> downloaded from Uniprot • Long reads from 21 sample types • Genes annotated in Yan <i>et al.</i> but absent in our current annotation were added to the annotation. 	34315	C:96.2%[S:88.1%,D:8.1%],F:0.9%,M:2.9%,n:1367

Table 1. Comparison of the functional annotation from Yan *et al.* with the annotation from this paper.

a BUSCO score of 96.2%, which is 3% higher than the previous annotation. To provide easy access to the expression data, we established a web-based portal allowing users to search, download and visualize the expression of genes among organs or developmental stages. The CPB atlas can be accessed via <https://cpb-atlas.uni-mainz.de/>.

Methods

The insects were maintained in a climate chamber with 24 °C, 70% humidity and a photoperiod of 16 h light and 8 h dark per day. Larvae, as well as adults, were fed with leaves from organically grown potato plants (Annabelle variety, purchased from Ellenberg's Kartoffelvielfalt GmbH & Co. KG, Barum, Germany), similar to our previous work³⁰. The E06 strain, which was originally collected from Spain in 2012, was used for all samples³³. The experimental design comprises the following steps (Fig. 1): dissection of the specimens, RNA extraction, RNA sequencing, data processing, and generation of the functional annotations.

Dissection and RNA extraction. The CPB atlas includes transcriptomes of tissues from fourth instar larvae (L4) and adults, as well as the whole body from different stages, from egg to third instar larva (Fig. 2). The tissues were dissected in PBS under a stereoscopic microscope. Dissection time was limited to 10 min to avoid RNA degradation, and the tissues were subsequently snap-frozen in liquid nitrogen and stored in a -80 °C freezer. The whole-body samples were similarly snap-frozen before being stored to ensure identical processing of the entire dataset. The eggs were pooled into groups of five. Three L1, L2, and L3 larvae were pooled to create each sample, respectively. We pooled individuals because the extracted RNA from each individual was not sufficient for sequencing. The L4 larvae and adult tissues are from single individuals. The L4 larvae were dissected on the 19th day of development from egg laying. Adults were kept in individual petri dishes from emergence and dissected seven days later. RNA was extracted using the RNeasy Mini kit from Qiagen (Venlo, The Netherlands), following the manufacturer's protocol. Lysis was performed with micro pestles. To obtain the full-length transcript of all the genes using an iso-seq approach, we pooled 200 ng RNA from one replicate of each sample type (except two due to their low RNA abundance; Supplementary table 1). The 61 RNA-seq samples and one iso-seq sample were sequenced using Illumina NovaSeq and PacBio Sequel sequencers, respectively, at Novogene (Cambridge, UK).

For RNA-seq library preparation, the Novogene NGS RNA Library Prep Set (PT042) was used. The magnetic beads of oligos d(T)25 were used to enrich the mRNA. Subsequently, mRNA was randomly fragmented, and cDNA synthesis was carried out using random hexamers and the reverse transcriptase enzyme. The second chain was synthesized with the addition of an Illumina buffer (non-directional library preparation). The resulting products were purified, end-repaired, and ligated with adapters. Fragments of the appropriate size are enriched by PCR, where indexed P5 and P7 primers are introduced, and final products are purified. The libraries were checked with Qubit 2.0 and real-time PCR for quantification and bioanalyzer Agilent 2100 for size distribution detection. Quantified libraries will be pooled and sequenced on the Illumina Novaseq. 6000 platform, according to effective library concentration and data amount, using the paired-end 150 strategy (PE150).

For Iso-seq, the SMRTbell library was prepared. The first strand cDNA was synthesized from the polyA tail by NEBNext's reverse transcriptase. Then the enzyme's terminal transferase activity added a few extra nucleotides to the 3' end of the cDNA. Subsequently, template switching, extension and amplification were performed to generate double-stranded cDNA. The cDNA was ligated with universal hairpin adapters. After purification with magnetic beads, the sequencing primer was annealed to the SMRTbell templates, followed by binding of the sequencing polymerase to the annealed templates. The SMRTbell library was sequenced in the PacBio Sequel IIe platform, in CCS mode. We used three replicates for each sample except the white fat body and the yellow fat body of the L4 larva, for which two samples were used.

Iso-seq data processing. Iso-seq data were processed using the command-line tools from PacBio's SMRT Link software (available at <https://www.pacb.com/>). First, consensus reads were generated with the function *css*. We demultiplexed reads with *lima* and then removed polyA tails with *isoseq3 refine* using default parameters. The reads were clustered with *isoseq3 cluster*. The Full-Length Non-Concatemer (FLNC) reads obtained were mapped to the genome³¹ using *pbbmm2*, the *pacbio* wrapper for *minimap2*³⁴. Finally, they were collapsed using *Cupcake* (v0.1.4, https://github.com/Magdoll/cDNA_Cupcake) to produce non-redundant full-length transcripts.

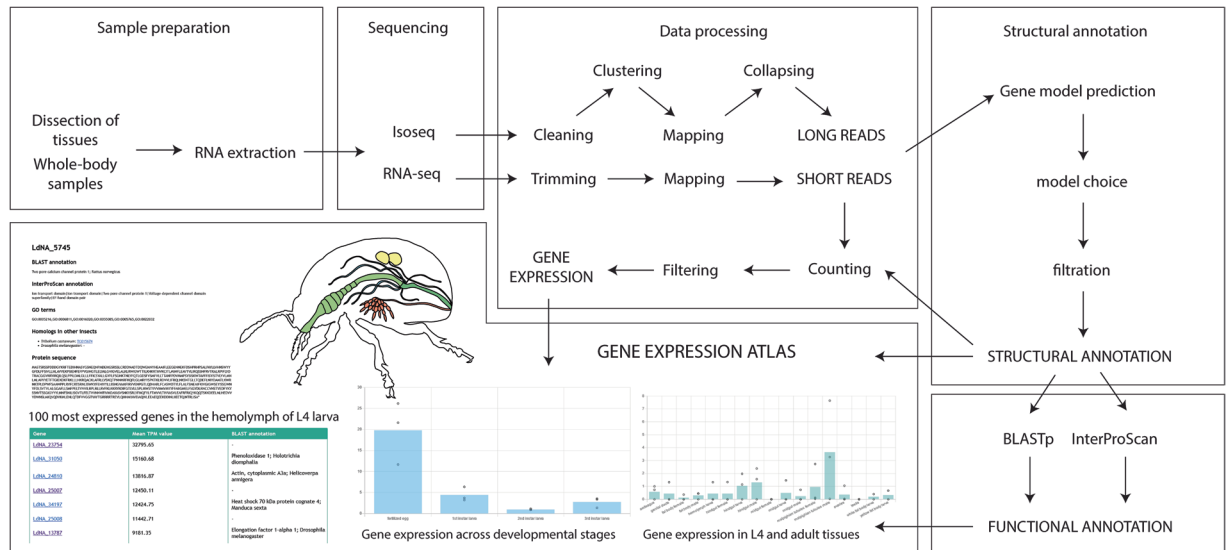


Fig. 1 Graphical representation of the CPB atlas workflow. The samples were first prepared (either dissected or directly frozen), RNA was extracted, then sequenced. The data from RNA-seq and Isoseq were processed separately. Both were used to construct the structural annotation i.e. position of the genes and their exons and CDS on the genome assembly. The structural annotation was employed later to build the functional annotation i.e. function of each gene, by a search with BLASTp and InterProScan separately. The short reads were mapped to the structural annotation to obtain an expression count for the genes. The resulting gene expression were displayed on a web portal in the form of bar plots.

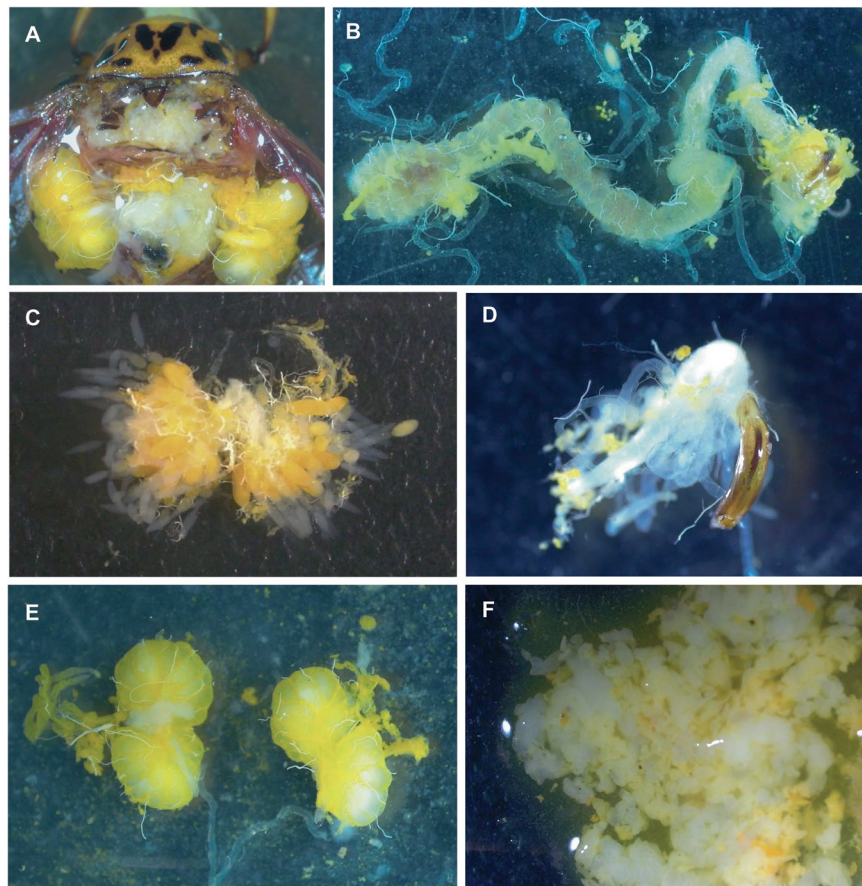


Fig. 2 Dissection of tissues: (A) Male incised dorsally. (B) Midgut, hindgut and malpighian tubules; (C) Ovaries; (D) Aedeagus and genital ducts; (E) Testes; (F) Fat body (white and yellow) of a fourth instar.

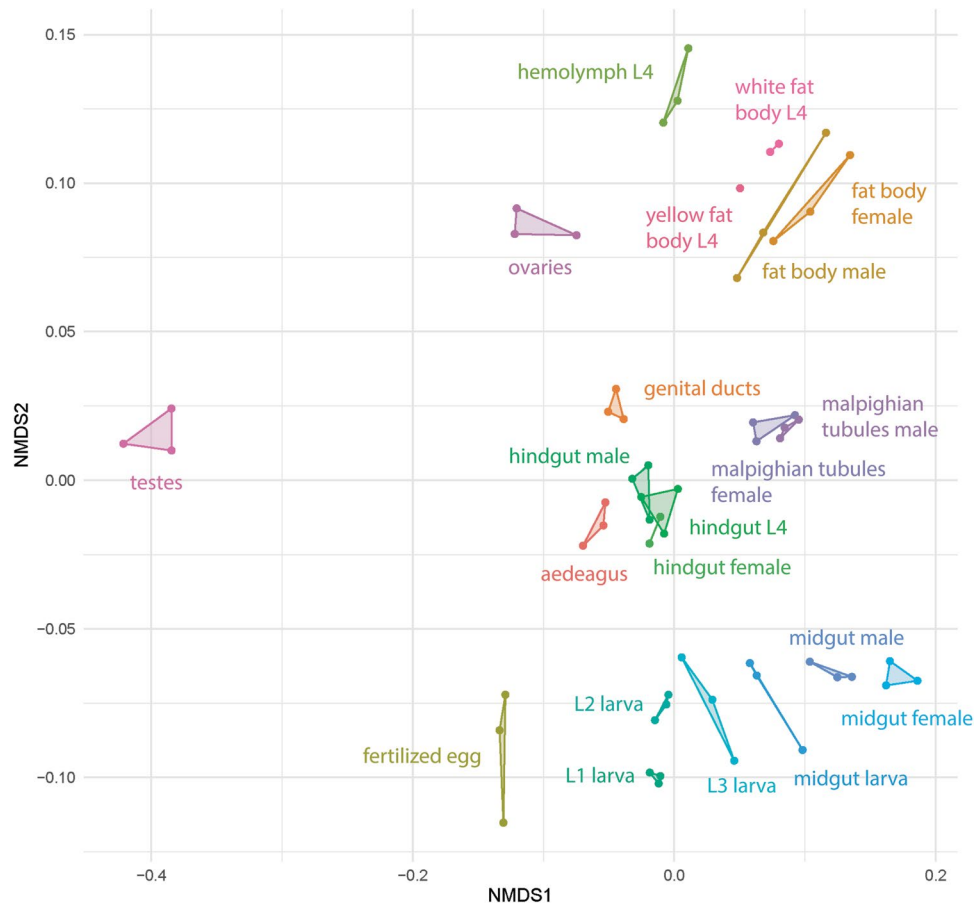


Fig. 3 Non-metric multidimensional scaling (NMDS) of the samples based on $\log_2(\text{TPM} + 1)$ values. Biological replicates are connected.

RNA-seq data processing. Trimmomatic was used to remove the Illumina adapters, drop the reads with a low quality or a short length (SLIDINGWINDOW:4:15 MINLEN:36) and remove the first ten bases of the reads (HEADCROP:10). To assess the quality both before and after this filtering process, we utilized FastQC (v0.11.9, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The reads were mapped to a chromosome-level genome assembly³¹ using STAR³⁵ (version 2.7.8a). FeatureCounts (Subread 2.0.5) was used to count the reads mapped to the annotated genes, with the parameters “-p -countReadPairs” which indicates that the reads are paired. The genes that were not expressed were filtered out based on their transcript per million (TPM) values. We kept only the transcript having a TPM value above one in at least two samples (https://github.com/Xu-lab-Evolution/CPB_gene_expression_atlas/blob/main/TPMfiltration.R). After filtration, 15,578 genes out of 34,350 were kept in the gene expression atlas.

Protein-coding gene annotation. Prior to genome annotation, we removed the contaminants using FCS GX³⁶ (v 0.5.0, <https://github.com/ncbi/fcs-gx>). For the prediction of protein-coding genes, we employed a modified BRAKER^{37–41} pipeline (<https://github.com/Gaius-Augustus/BRAKER/tree/master>). In brief, this approach integrates gene models predicted based on both short-read RNA-seq transcriptome (BRAKER1 method^{42–47}) and protein homologs from *Drosophila melanogaster* and *Tribolium castaneum* (BRAKER2 method^{40,43,44,48–52}). Subsequently, TSEBRA⁴¹ was used to compare these predictions against full-length transcripts obtained from Iso-seq data, thereby determining the most accurate gene models.

When running BRAKER1, we combined paired-end RNA-seq reads from 17 different tissue types (Supplementary table 2) and used the repeat soft-masked CPB reference genome³¹. The short reads facilitated the automatic training of GeneMark-ET by using the spliced alignment data to assist gene model predictions by AUGUSTUS⁵³. For BRAKER2, a similar process was conducted for the automatic training of GeneMark-EP+⁵². However, in this step, BRAKER2 employed information on protein-coding exon boundaries derived from the alignment of homologous protein sequences from *Drosophila melanogaster* and *Tribolium castaneum* (downloaded from UniProt⁵⁴) to assist AUGUSTUS in gene prediction. We used GeneMarkS-T⁵⁵ to identify the protein-coding regions within each full-length transcript. We then subsequently merged gene models from BRAKER1 and BRAKER2 and compared them against the Iso-seq full-length transcripts using TSEBRA. We retained only the longest isoform for each gene model. Using AGAT (v0.9.0, <https://github.com/NBISweden/AGAT>) we removed genes that are shorter than 100 bp and single-exon genes that lacked complete start or stop codons. Unique gene models that were present in the previous annotation version³¹ and had no overlap with any

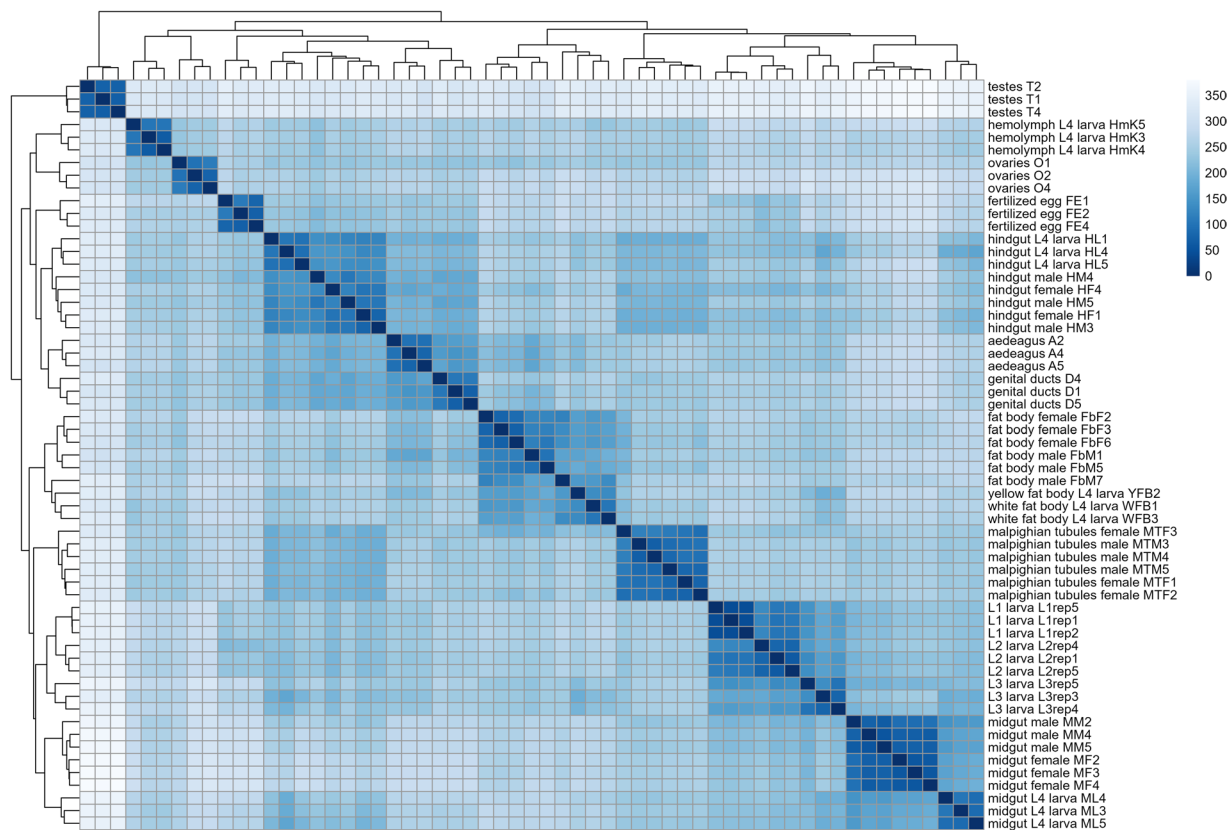


Fig. 4 Heatmap showing the pairwise comparisons of $\log_2(\text{TPM})$ values across the 59 samples of the gene expression atlas. All the tissues are correctly clustering together showing an absence of contaminations from other tissues. Some tissues also segregate by developmental stage (e.g. midgut of L4 larvae vs adults) showing a modification of the transcriptome through age. On the contrary there are no clear separation of the tissues between the sexes in adult, showing the absence of sex-specific transcriptome for tissues common to both sexes.

models predicted in the current version were retained in the new annotation. For the functional annotations, protein sequences were aligned to the UniProtKB²⁰ database using *blastp* (BLAST + v2.12.0)⁴⁴ with “-evalue 1e-6 -max_hsps 1 -max_target_seqs 1 -outfmt 6”. The results were further processed using InterProScan (5.71–102.0)⁵⁶ with the options “-goterms -iprlookup”. The KEGG pathways were added using BlastKOALA⁵⁷.

Weighted gene co-expression network analysis. We performed a weighted gene co-expression network analysis using the WGCNA package in R⁵⁸. The data were normalized using counts adjusted by TMM factors (CTF), we then applied a hyperbolic arcsine (asinh) transformation following the procedure described by Johnson and Krishnan⁵⁹. We used tissue samples from adults and L4 larvae and kept genes with TPM values greater than one in at least one sample. The network was constructed with the blockwiseModules function and the parameters TOMType = “signed” and power = 9. The genes were divided into 52 modules; the smaller one (Salmon4) contained 20 genes and the largest (Turquoise) contained 3,002 genes. We performed correlation analysis to link modules eigengenes to tissues using the Pearson coefficient (Supplementary figure 3).

Tissue-specificity of the genes. Tissue-specificity of the genes has been tested with the extended tau score^{60,61}. This score attributes a value from 0 (ubiquitous) to 1 (specific) to each gene of the dataset. We first calculated the median TPM value for each gene in every tissue. Those values were \log_2 transformed, and the tau score was calculated using the formula described in Yanai *et al.*⁶⁰. The genes with a tau score greater than 0.85 and with a median TPM value above 10 in all tissues were considered specific. These genes were then assigned tissues using the extended tau score method, which determines a threshold based on the maximum expression value, the standard deviation and a Z-value derived from a Fuzzy c-means clustering of lower and upper bounds expression.

A total of 2487 genes (16% of the expressed genes) were tissue-specific, amongst which 2,246 were specific to one tissue, 227 to two tissues, and 14 to three tissues. The testes were by far the tissue presenting the more tissue-specific genes, with a number of 1,102 genes. Oppositely, the fat body displayed only 61 specific genes.

Gene Ontology enrichment analysis. To check whether the RNA-seq data accurately reflects the transcriptome of each tissue, a gene ontology enrichment analysis was performed on the tissue-specific genes identified previously. The Enricher function of the ClusterProfiler package on R was used separately on the GO terms belonging to the “molecular function” and the “biological process” ontologies⁶². The results suggest that the GO terms enriched among tissue-specific genes are related to the function of each tissue (Supplementary table 3).

For example, testes-specific genes were, amongst others, enriched in the biological processes of cilium assembly (GO:0060271) and cilium movement (GO:0003341). The fat body specific genes were enriched in fatty acid biosynthetic process (GO:0006633) and long-chain fatty-acyl-CoA metabolic process (GO:0035336).

Data Records

The raw data and the analysis results have been deposited in Figshare⁶³. The raw sequencing data are also available in the NCBI Sequence Read Archive (SRA) under accession numbers SAMN39526582 to SAMN39526643, associated with BioProject PRJNA1067435. The expression data can be accessed under GEO accession number GSE285883⁶⁴. The genome annotation, the transcriptome and the various analysis results are available on the web application: <https://cpb-atlas.uni-mainz.de/resources>.

Technical Validation

The sample integrity, purity, and quantification were verified using an Agilent 5400 device. Samples were visualized on an NMDS and a pairwise heatmap to detect potential outliers using the $\log_2(\text{TPM} + 1)$ values (Supplementary figures 1, 2). Two outliers were detected: a yellow fat body and a female hindgut. As this is likely due to contamination during dissection and sample collection steps, we removed these two samples from the CPB atlas. After the removal of outliers, all the visualization methods were applied again and showed that replicates cluster closely to each other (Figs. 3, 4).

Code availability

The scripts used to process short reads are available at the address: https://github.com/Xu-lab-Evolution/CPB_gene_expression_atlas/releases/tag/release_v1.0. The genome annotation has been generated using the same method described at https://github.com/Xu-lab-Evolution/Waterlily_aphid_genome_project⁶⁵.

Received: 22 April 2024; Accepted: 11 February 2025;

Published online: 19 February 2025

References

1. Ferro, D. N., Logan, J. A., Voss, R. H. & Elkinton, J. S. Colorado potato beetle (Coleoptera: Chrysomelidae) temperature-dependent growth and feeding rates. *Environ Entomol* **14**, 343–348 (1985).
2. Huseth, A. S. *et al.* Spatial and temporal potato intensification drives insecticide resistance in the specialist herbivore, *Leptinotarsa decemlineata*. *PLoS One* **10**, e0127576 (2015).
3. Chen, Y. H., Cohen, Z. P., Bueno, E. M., Christensen, B. M. & Schoville, S. D. Rapid evolution of insecticide resistance in the Colorado potato beetle, *Leptinotarsa decemlineata*. *Curr Opin Insect Sci* **55**, 101000 (2023).
4. Tisler, A. M. & Zehnder, G. W. Insecticide resistance in the Colorado potato beetle (Coleoptera: Chrysomelidae) on the Eastern Shore of Virginia. *J Econ Entomol* **83** (1990).
5. Szendrei, Z., Grafius, E., Byrne, A. & Ziegler, A. Resistance to neonicotinoid insecticides in field populations of the Colorado potato beetle (Coleoptera: Chrysomelidae). *Pest Manag Sci* **68** (2012).
6. Scott, I. M., Tolman, J. H. & Macarthur, D. C. Insecticide resistance and cross-resistance development in Colorado potato beetle *Leptinotarsa decemlineata* Say (Coleoptera: Chrysomelidae) populations in Canada 2008–2011. *Pest Manag Sci* **71** (2015).
7. Péliissié, B. *et al.* Genome resequencing reveals rapid, repeated evolution in the Colorado potato beetle. *Mol Biol Evol* **39**, msac016 (2022).
8. Fargues, J., Delmas, J.-C., Augé, J. & Lebrun, R. A. Fecundity and egg fertility in the adult Colorado beetle (*Leptinotarsa decemlineata*) surviving larval infection by the fungus *Beauveria bassiana*. *Entomol Exp Appl* **61**, 45–51 (1991).
9. Weber, D. C. & Ferro, D. N. Flight and fecundity of Colorado potato beetles (Coleoptera: Chrysomelidae) fed on different diets. *Ann Entomol Soc Am* **89**, 297–306 (1996).
10. Crossley, M. S., Chen, Y. H., Groves, R. L. & Schoville, S. D. Landscape genomics of Colorado potato beetle provides evidence of polygenic adaptation to insecticides. *Mol Ecol* **26**, 6284–6300 (2017).
11. Kordali, S., Kesdek, M. & Cakir, A. Toxicity of monoterpenes against larvae and adults of Colorado potato beetle, *Leptinotarsa decemlineata* Say (Coleoptera: Chrysomelidae). *Ind Crops Prod* **26**, 278–297 (2007).
12. Clements, J. *et al.* Conjugated linoleic acid as a novel insecticide targeting the agricultural pest *Leptinotarsa decemlineata*. *PLoS One* **14**, e0220830 (2019).
13. Trdan, S., Vidrih, M., Andjus, L. & Laznik, Ž. Activity of four entomopathogenic nematode species against different developmental stages of Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera, Chrysomelidae). *Helminthologia* **46**, 14–20 (2009).
14. Zhu, F., Xu, J., Palli, R., Ferguson, J. & Palli, S. R. Ingested RNA interference for managing the populations of the Colorado potato beetle, *Leptinotarsa decemlineata*. *Pest Manag Sci* **67**, 175–182 (2011).
15. Clements, J. *et al.* RNA interference of three up-regulated transcripts associated with insecticide resistance in an imidacloprid resistant population of *Leptinotarsa decemlineata*. *Pestic Biochem Physiol* **135**, 35–40 (2017).
16. Kalsi, M. & Palli, S. R. Transcription factor cap n collar C regulates multiple cytochrome P450 genes conferring adaptation to potato plant allelochemicals and resistance to imidacloprid in *Leptinotarsa decemlineata* (Say). *Insect Biochem Mol Biol* **83**, 1–12 (2017).
17. Mishra, S. *et al.* Selection for high levels of resistance to double-stranded RNA (dsRNA) in Colorado potato beetle (*Leptinotarsa decemlineata* Say) using non-transgenic foliar delivery. *Sci Rep* **11**, 6523 (2021).
18. Zichová, T., Kocourek, F., Salava, J., Nadřová, K. & Stará, J. Detection of organophosphate and pyrethroid resistance alleles in Czech *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae) populations by molecular methods. *Pest Manag Sci* **66**, 853–860 (2010).
19. Rinkevich, F. D. *et al.* Multiple evolutionary origins of knockdown resistance (kdr) in pyrethroid-resistant Colorado potato beetle, *Leptinotarsa decemlineata*. *Pestic Biochem Physiol* **104**, 192–200 (2012).
20. Clements, J., Schoville, S., Peterson, N., Lan, Q. & Groves, R. L. Characterizing molecular mechanisms of imidacloprid resistance in select populations of *Leptinotarsa decemlineata* in the Central Sands region of Wisconsin. *PLoS One* **11**, e0147844 (2016).
21. Koopmanschap, A. B., Oouchi, H. & de Kort, C. A. D. Effects of a juvenile hormone analogue on the eggs, post-embryonic development, metamorphosis and diapause induction of the Colorado potato beetle, *Leptinotarsa decemlineata*. *Entomol Exp Appl* **50**, 255–263 (1989).
22. Shen, C.-H. *et al.* Ecdysis triggering hormone is essential for larva–pupa–adult transformation in *Leptinotarsa decemlineata*. *Insect Mol Biol* **30**, 241–252 (2021).
23. de Loof, A. & de Wilde, J. Hormonal control of synthesis of vitellogenic female protein in the Colorado beetle, *Leptinotarsa decemlineata*. *J Insect Physiol* **16**, 1455–1466 (1970).

24. Lefevre, K. S., Koopmanschap, A. B. & De Kort, C. A. D. Changes in the concentrations of metabolites in haemolymph during and after diapause in female Colorado potato beetle, *Leptinotarsa decemlineata*. *J Insect Physiol* **35**, 121–128 (1989).
25. Yocum, G. D., Rinehart, J. P., Chirumamilla-Chapara, A. & Larson, M. L. Characterization of gene expression patterns during the initiation and maintenance phases of diapause in the Colorado potato beetle, *Leptinotarsa decemlineata*. *J Insect Physiol* **55**, 32–39 (2009).
26. Lehmann, P., Margus, A. & Lindström, L. Inheritance patterns of photoperiodic diapause induction in *Leptinotarsa decemlineata*. *Physiol Entomol* **41**, 218–223 (2016).
27. Doğan, C. *et al.* Two calcium-binding chaperones from the fat body of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae) involved in diapause. *Arch Insect Biochem Physiol* **106**, e21755 (2021).
28. Güneş, G. *et al.* A look into Colorado potato beetle lipid metabolism through the lens of lipid storage droplet proteins. *Insect Biochem Mol Biol* **133**, 103473 (2021).
29. Doğan, C. *et al.* Characterization of calcium signaling proteins from the fat body of the Colorado Potato Beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae): Implications for diapause and lipid metabolism. *Insect Biochem Mol Biol* **133**, 103549 (2021).
30. Edison, A. *et al.* Evidence of active oviposition avoidance to systemically applied imidacloprid in the Colorado potato beetle. *Insect Sci* (2024).
31. Yan, J. *et al.* Chromosome-level genome assembly of the Colorado potato beetle, *Leptinotarsa decemlineata*. *Sci Data* **10** (2023).
32. Schoville, S. D., Cohen, Z. P. & Crossley, M. S. Population Genomic Insights into Insecticide Resistance in the Colorado potato beetle. in *Population Genomics* (2021).
33. Mehlhorn, S. G., Geibel, S., Bucher, G. & Nauen, R. Profiling of RNAi sensitivity after foliar dsRNA exposure in different European populations of Colorado potato beetle reveals a robust response with minor variability. *Pestic Biochem Physiol* **166** (2020).
34. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34** (2018).
35. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29** (2013).
36. Astashyn, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol* **25** (2024).
37. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7** (2006).
38. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24** (2008).
39. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods in Molecular Biology* **1962** (2019).
40. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3** (2021).
41. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22** (2021).
42. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32** (2016).
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215** (1990).
44. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10** (2009).
45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25** (2009).
46. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stümborg, M. P. & Marth, G. T. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27** (2011).
47. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* **42** (2014).
48. Lomsadze, A., Ter-Hovhannisyán, V., Chernoff, Y. O. & Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* **33** (2005).
49. Iwata, H. & Gotoh, O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res* **40** (2012).
50. Gotoh, O., Morita, M. & Nelson, D. R. Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment. *BMC Bioinformatics* **15** (2014).
51. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12** (2014).
52. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* **2** (2020).
53. Stanke, M. *et al.* AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34** (2006).
54. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* **47** (2019).
55. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* **43** (2015).
56. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30** (2014).
57. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428** (2016).
58. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9** (2008).
59. Johnson, K. A. & Krishnan, A. Robust normalization and transformation techniques for constructing geneco-expression networks from RNA-seq data. *Genome Biol* **23** (2022).
60. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21** (2005).
61. Lüleci, H. B. & Yılmaz, A. Robust and rigorous identification of tissue-specific genes by statistically extending tau score. *BioData Min* **15** (2022).
62. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16** (2012).
63. Wilhelm, L., Wang, Y. & Xu, S. Gene expression atlas of the Colorado potato beetle (*Leptinotarsa decemlineata*). *figshare* <https://doi.org/10.6084/m9.figshare.28255721> (2025).
64. NCBI. *GEO* <https://identifiers.org/geo:GSE285883> (2025).
65. Wang, Y. & Xu, S. A high-quality genome assembly of the waterlily aphid *Rhopalosiphum nymphalae*. *Sci Data* **11** (2024).

Acknowledgements

We acknowledge Cansu Doğan for her advice on identifying and dissecting the CBP tissues, Ursula Martiné for her considerable help in the RNA extraction, Christina Sternara Tsiara and Pablo Duchén for creating the heatmaps displayed on the web application, Alex Garel for deploying the web application. This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2526 (Project number 407023052).

Author contributions

Conceived and designed the experiments: S.X., L.W. Performed the experiments: L.W. Analyzed the data: L.W., Y.W. Built the web application: L.W., Wrote the paper: L.W., Y.W. Substantially revised the paper: S.X. Supervision and grant acquisition: S.X.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04607-7>.

Correspondence and requests for materials should be addressed to S.X.

Reprints and permissions information is available at www.nature.com/reprints.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025