



Data Article

Transcriptomic dataset of the development and maturation of the *Rhipicephalus microplus* ovary

Raquel Cossío-Bayúgar^{a,*}, Estefan Miranda-Miranda^a,
Hugo Aguilar-Díaz^a, Verónica Narváez-Padilla^b, Enrique Reynaud^{c,*}

^a Centro Nacional de Investigaciones Disciplinarias en Salud Animal e Inocuidad, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP). Boulevard Cuauhnahuac No. 8534, Jiutepec, Morelos 62574, Mexico

^b Centro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Chamilpa, Cuernavaca, 62209, Mexico

^c Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología Universidad Nacional Autónoma de México, Cuernavaca, Morelos. Av. Universidad 2001, Chamilpa, Cuernavaca, 62209, Mexico

ARTICLE INFO

Article history:

Received 20 March 2024

Revised 11 June 2024

Accepted 17 June 2024

Available online 24 June 2024

Dataset link: [Transcriptomic dataset of the development and maturation of the *Rhipicephalus microplus* ovary \(Original data\)](#)

Dataset link: [Rhipicephalus microplus raw ovary gene expression sequence reads \(Original data\)](#)

Keywords:

Cattle tick

Ovogenesis

Gametogenesis

Ovary development

Differential gene expression

Tick reproduction

ABSTRACT

To conduct differential gene expression analysis, ovaries from the cattle tick *Rhipicephalus microplus* were dissected at three distinct developmental stages (preingurgitated, immature ingurgitated, and mature ingurgitated). Additionally, undissected intact mature males and complete ingurgitated female ticks without ovaries (carcasses) were also collected to serve as reference samples for analysis. To perform total RNA purification, tissue from ten individuals representing each of the five previously described conditions was pooled. mRNA was isolated from the purified total RNA using the oligo (dT) method. Following fragmentation, double stranded cDNA was synthesized and ligated to sequencing adapters. Suitably sized fragments were subsequently used for PCR amplification. Libraries were analyzed and quantified using an Agilent 2100 Bioanalyzer and an ABI StepOnePlus Real-Time PCR System. A total of 45.64 Gb bases were sequenced using the Illumina HiSeq sequencing platform. After assembling the samples and correcting for abundance, we obtained 82,877 unigenes. The total length, average length, N50, and GC content

* Corresponding authors.

E-mail addresses: cossio.raquel@inifap.gob.mx (R. Cossío-Bayúgar), enrique.reynaud@ibt.unam.mx (E. Reynaud).

of the unigenes were 89,754,828 bp, 1,082 bp, 2,068 bp and 49.04 % respectively. For functional annotation, the unigenes were aligned with 7 functional databases. The number of unigenes identified in the functional databases were as follows: 32,518 (NR:39.24 %), 10,259 (NT:12.38 %), 23,624 (Swissprot:28.50 %), 22,203 (KOG:26.79 %), 25,072 (KEGG:30.25 %), 17,435 (GO:21.04 %), and 23,220 (InterPro:28.02 %). Unigene candidate coding regions (CDS) among the unigenes were predicted using TransDecoder software and 42,143 CDS were detected. We also detected 10,522 simple sequence repeats (SSRs) distributed on 8,126 unigenes, and predicted 4,672 transcription factors (TF) coding unigenes. Our data can be used to identify genes that are important for male and female tick and arachnid reproduction and tick general physiology.

© 2024 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license

(<http://creativecommons.org/licenses/by-nc/4.0/>)

Specifications Table

Subject	Biological sciences, Omics: Transcriptomics
Specific subject area	Differential gene expression between maturing ovaries, males and female carcasses of the cattle tick <i>Rhipicephalus microplus</i> using tissue specific transcriptomic data
Type of data	Table Image Chart Graph Figure
Data collection	RNA was purified with Trizol, analyzed with the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Kit; measured parameters: RIN value, 28S/18S ratio, and fragment length distribution. Libraries were sequenced in the Illumina HiSeq platform. The bioinformatics workflow included filtration of low-quality reads, assembly of clean reads into unigenes, functional annotation, SSR detection, calculation of unigene expression levels, SNP detection, identification of differentially expressed genes between samples (DEGs), clustering analysis, and functional annotations.
Data source location	<ul style="list-style-type: none"> · Institution: "Centro Nacional de Investigación en Salud Animal e Inocuidad" (CENID-SAI, National Center for Research in Animal Health and Safety) · City/Region: Jiutepec, Morelos · Country: México
Data accessibility	Repository name: NCBI BioProject Data identification number: PRJNA884635 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA884635 Reynaud, Enrique (2024), "Transcriptomic dataset of the development and maturation of the <i>Rhipicephalus microplus</i> ovary.", Mendeley Data, V1, doi: 10.17632/t9h88383xd.1 https://data.mendeley.com/datasets/t9h88383xd/1
Related research article	

1. Value of the Data

- *Rhipicephalus microplus* is a blood sucking ectoparasite that costs billions to the cattle industry.
- *R. microplus* causes stress in cattle, damages their skin, reduces milk and meat production and transmits bovine diseases such as anaplasmosis and babesiosis.

- These data provide a comprehensive transcriptome analysis of the developmental stages of *Rhipicephalus microplus* ovaries.
- Transcriptomic data from the reproductive tissue of *R. microplus* are highly valuable because they allow the identification of genes involved in tick reproduction, which can become potential targets for controlling ticks.
- These data will benefit the scientists who are interested in arthropod reproduction, tick biology and pest control.
- These data are versatile and can be utilized to explore various aspects of tick and acarid biology, including reproduction, development and evolution. Additionally, these finding could lead to valuable insights into potential gene targets for the development of new pesticides. Moreover, these data can be used and reused for continuous research in these areas.

2. Background

The objective of this study was to identify differentially expressed genes during the development of the germ line in the cattle tick *R. microplus*. This ectoparasite inflicts significant economic losses on the cattle industry. To achieve this goal, we isolated ovaries from *R. microplus* at three distinct developmental stages and extracted RNA for transcriptome sequencing. Our aim was to identify crucial genes involved in ovarian and oocyte development. To distinguish reproductive genes from housekeeping and somatic tissue genes, we sequenced the transcriptome of female ticks without ovaries (carcasses). Additionally, we sequenced the transcriptome of undissected intact males, taking advantage that their testicles account for a considerable portion of their body mass. Genes related to spermatogenesis and male reproduction can be identified by the subtraction of the genes expressed in female gonads and carcasses from the genes that are expressed in males.

3. Data Description

Table 1 shows the quality metrics of the raw reads and the generated assembly and annotation of the information.

Table 2 shows unigene Functional Annotation found in 7 functional databases (NR, NT, GO, KOG, KEGG, SwissProt and InterPro) and the intersection between all of them.

Table 1
Clean reads quality metrics.

Sample	Total raw reads(M)	Total clean reads (M)	Total clean bases (Gb)	Clean reads Q20 (%)	Clean reads Q30 (%)	Clean reads ratio (%)
SRR21725969 Female carcass	91.84	90.25	9.03	98.52	95.67	98.27
SRR21725968 Preingurgitated ovary	94.03	91.69	9.17	98.6	95.98	97.52
SRR21725966 Immature ingurgitated ovary	94.03	92.29	9.23	98.59	95.92	98.15
SRR21725967 SRR21725967 Mature ingurgitated ovary	94.03	91.56	9.16	98.61	95.99	97.37
SRR21725965 male	91.84	90.51	9.05	98.55	95.77	98.55

Sample: Sample name

Total Raw Reads (Mb): The number of reads before filtering.

Total Clean Reads (Mb): The number of reads after filtering.

Total Clean Bases (Gb): The total base amount after filtering.

Clean Reads Q20 (%): The percentage of bases whose quality was greater than 20 in the clean reads.

Clean Reads Q30 (%): The percentage of bases whose quality was greater than 30 in the clean reads.

Clean Reads Ratio (%): The percentage of the number of clean reads.

Table 2

Unigene functional annotation were found in 7 functional databases (NR, NT, GO, KOG, KEGG, SwissProt and InterPro).

Values	Total	Nr	Nt	SwissProt	KEGG	KOG	InterPro	GO	Intersection	Overall
Number	82,877	32,518	10,259	23,624	25,072	22,203	23,220	17,435	3906	35,070
Percentage	100 %	39.24 %	12.38 %	28.50 %	30.25 %	26.79 %	28.02 %	21.04 %	4.71 %	42.32 %

Intersection: The number of unigenes annotated in all 7 functional databases.

Overall: The number of unigenes annotated in any of the 7 functional databases.

Table 3

Species distribution of the genes found in the nonredundant nucleotide database.

Species	Phylum or Subphylum	Class	Subclass	Order	%
<i>Ixodes scapularis</i>	Chelicerata	Arachnida		Ixodida	47.87 %
<i>Limulus polyphemus</i>	Chelicerata	Merostomata		Xiphosurida	9.17 %
<i>Parasteatoda tepidariorum</i>	Chelicerata	Arachnida		Araneae	3.64 %
<i>Stegodyphus mimosarum</i>	Chelicerata	Arachnida		Araneae	3.01 %
<i>Galendromus occidentalis</i>	Arthropoda	Arachnida	Acari	Mesostigmata	2.13 %
<i>Nuttalliella namaqua</i>	Chelicerata	Arachnida	Acari	Ixodida	1.61 %
<i>Amblyomma variegatum</i>	Chelicerata	Arachnida		Ixodida	1.57 %
<i>Rhipicephalus microplus</i>	Arthropoda	Arachnida		Ixodida	1.53 %
<i>Tropilaelaps mercedesae</i>	Arthropoda	Arachnida	Acari	Ixodida	1.21 %

Distribution of homologous genes of other annotated species found in the nonredundant nucleotide database. All identifiable organisms whose represented genes had a ratio $\geq 1\%$ in our database are Chelicerata or *Arachnida*.

Table 4

Differentially expressed genes (DEGs) using female carcass as reference.

2-fold change in expression (log2 fold change of ± 1)		
Tissue	Up regulated	Down regulated
PIO	23,757	4403
IO	22,885	4294
FMIO	23,161	4204
Male	18,769	4902
Benjamini-Hochberg procedure adjusted pseudo P-values		
PIO	16,746	2292
IO	15,147	2076
FMIO	15,729	2187
Male	10,756	2087

Up and down regulated genes using the 2-fold change in expression (log2 fold change of ± 1) criterion and the stricter Benjamini-Hochberg procedure adjusted pseudo P-values criterion.

Table 3 shows the species distribution of the genes found in the nonredundant nucleotide database for which $\geq 1\%$ of the genes were represented ratio in our database.

Table 4 shows up regulated and down regulated genes using the 2-fold change in expression (log2 fold change of ± 1) and the corrected up and down regulation using the Benjamini-Hochberg procedure.

Fig. 1 shows the functional annotation of homologous genes of other annotated species.

Fig. 2 shows the representation of homologous genes in different databases, CDS size distribution and transcription factor family classification.

Fig. 3 shows gene expression and principal component analysis.

Supplementary files that contain all the raw data related to all the tables, graphs, images and charts are provided at: Reynaud, Enrique (2024), "Transcriptomic dataset of the development and maturation of the Rhipicephalus microplus ovary.", Mendeley Data, V1, doi:10.17632/t9h88383xd.1

<https://data.mendeley.com/datasets/t9h88383xd/1>

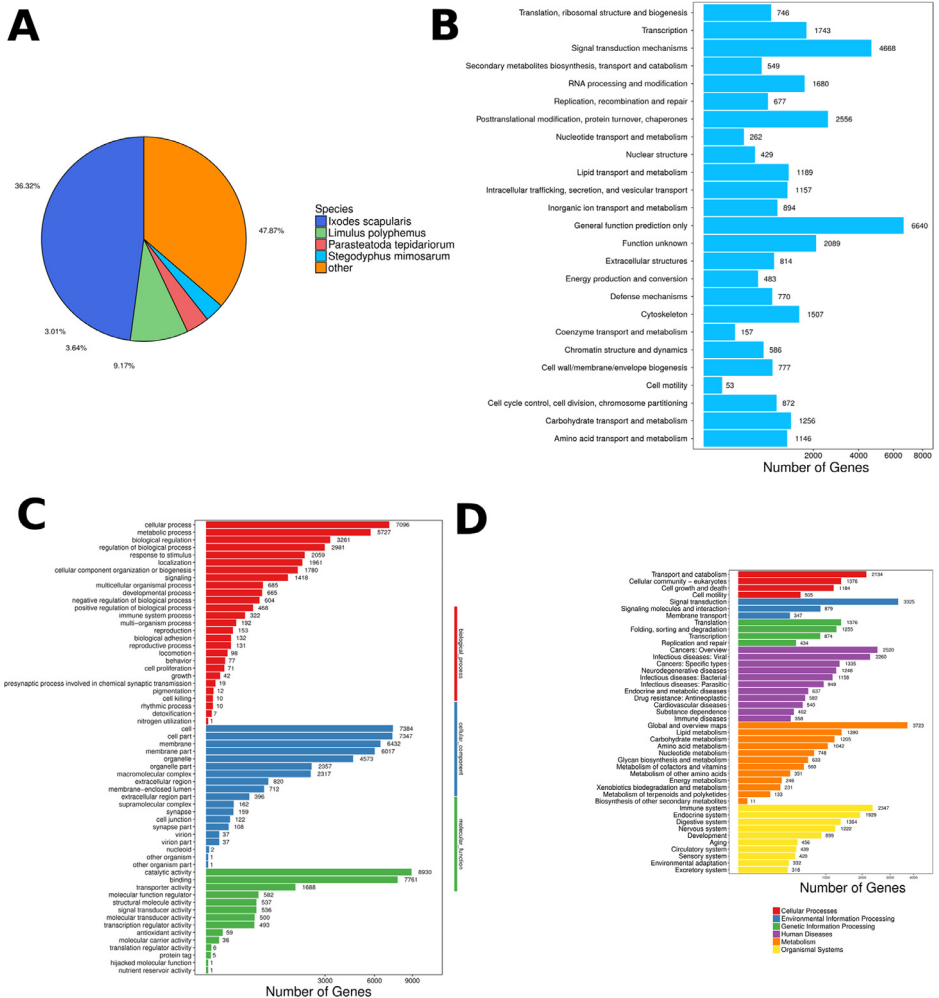


Fig. 1. Functional annotation of homologous genes of other annotated species. A) Distribution of homologous genes from other annotated species identified in the nonredundant nucleotide database revealed that all organisms with genes represented at a ratio of $\geq 1\%$ in our database belonged to the *Chelicerata* or *Arachnida* taxa B) Eukaryotic Orthologous Groups of proteins (KOG) functional distribution. C) Gene Ontology (GO) functional distribution. D) Kyoto Encyclopedia of Genes and Genomes (KEGG) functional distribution.

Bioproject and BioSample data were uploaded to NCBI according to instructions [1]. The accession numbers were assigned as follows: PRJNA884635, SAMN31034487, SAMN31034488, SAMN31034489, SAMN310344890, and SAMN31034491.

4. Experimental Design, Materials and Methods

RNA extraction. Ticks at appropriate developmental stages and sex were selected, and ovaries were extracted from females. Ticks were washed with distilled water to remove any extraneous debris. A transversal cut was performed between the first and second leg pairs to remove the entire anterior area. The internal organs were extruded in Jan & Jan30 solution (NaCl 128 mM,

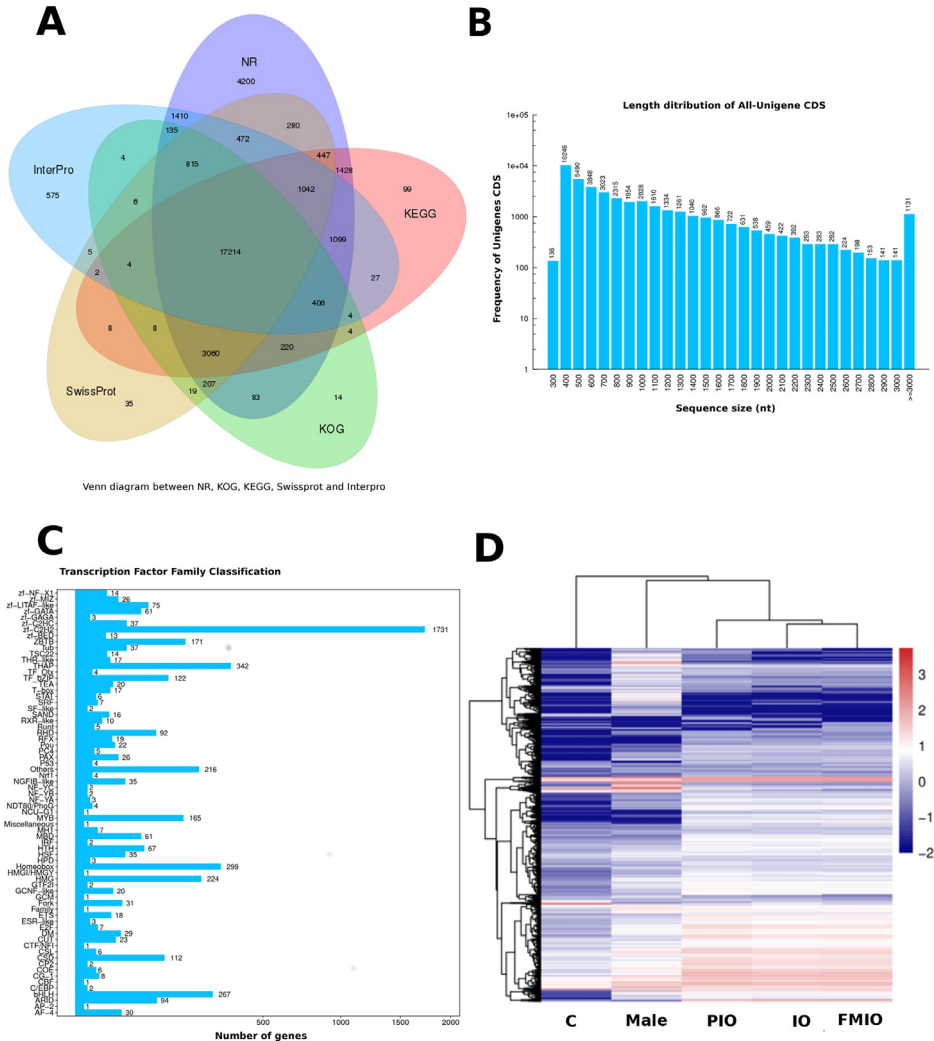


Fig. 2. Representation of homologue genes in different databases, CDS size distribution and transcription factor family classification. A) Venn diagram of the NR, KOG, KEGG, SwissProt, and Interpro databases. Extensive functional overlap can be found among the different databases; however, some genes are only represented in unique databases. B) Length distribution of all unigene CDSs. C) Transcription Factor family classification of unigenes. D) Heatmap showing the distribution of transcription factor expression levels according to tissue. Transcription factor expression clearly clusters ovarian tissue regardless of its maturity (C = Carcass; Male = Undissected intact males; PIO = Preingurgitated Ovary; IO = Iningurgitated ovary; FMIO = Fully Matured Iningurgitated Ovary) separately from males, which are also clearly differentiated from gonadless somatic tissue (carcasses).

KCl 2 mM, MgCl₂ 4 mM, Sucrose 36 mM, HEPES 5 mM, pH 7.3) [2]. Complete ingurgitated female ticks without ovaries (carcasses), complete males, and extracted ovaries were rapidly frozen in liquid nitrogen and then finely

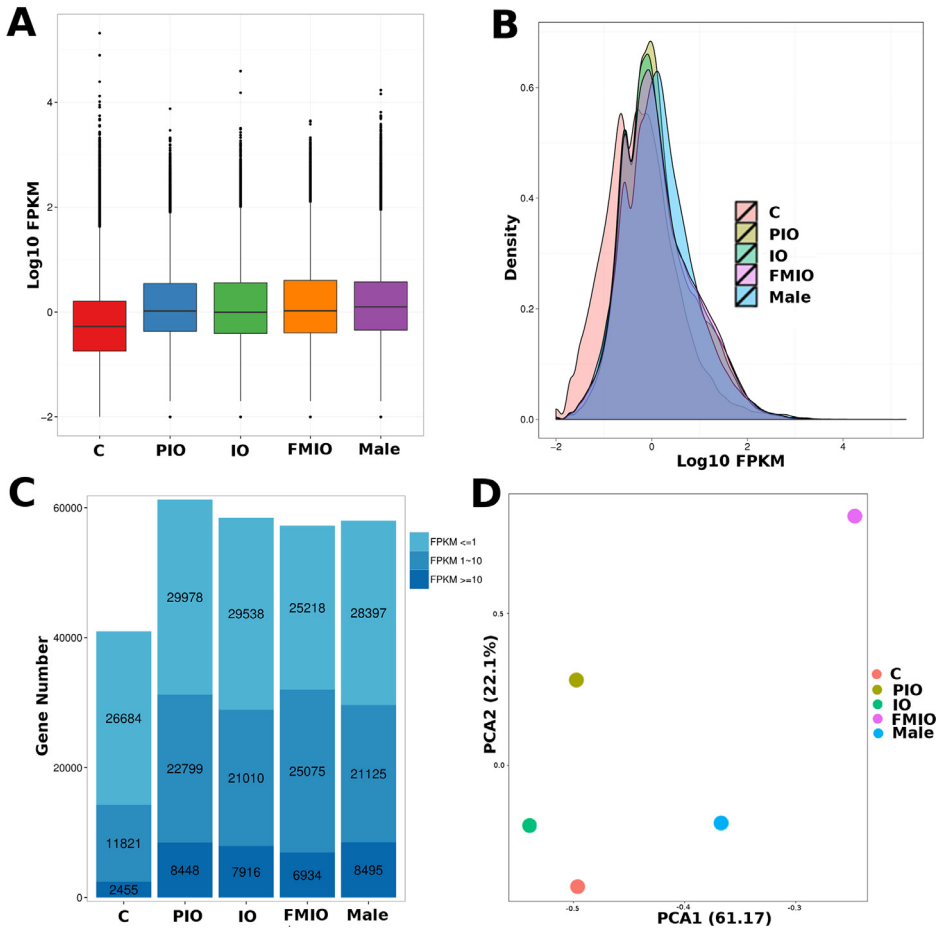


Fig. 3. Gene expression and principal component analysis. A) Box plot of gene expression showing the distribution and dispersion of gene expression levels in the different tissues analyzed. B) Gene expression distribution plot illustrating gene expression across the analyzed tissues. The distribution of the genes expressed in each tissue exhibited distinct peaks, indicating tissue-specific gene expression patterns. The shape of the distributions of the different tissues are similar, suggesting homogenous overall gene expression levels across the analyzed tissues. C) Gene expression distribution for each of the tissues analyzed. D) PCA scatter plot displaying the projection of gene expression data onto two principal components (PCs). The x-axis represents PCA1, which explains 61.17 % of the total variance in the dataset, while the y-axis represents PCA2, explains 22.1 % of the total variance. Each point on the scatter plot corresponds to a tissue sample, and its position reflects the tissue's gene expression profile in the reduced-dimensional space defined by the two selected PCs. C = Carcass, Male = Undissected intact males, PIO = Preingurgitated ovary; IO = Ingurgitated ovary; FMIO = Fully Matured Ingurgitated Ovary.

incubated for 2–3 min. The sample was centrifuged for 15 min at $12,000 \times g$ at 4°C , resulting in separation into a lower phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. The aqueous phase containing the RNA was transferred to a new tube, and 0.5 mL of isopropanol was added per 1 mL of TRIzol reagent used for lysis. After incubating for 10 min at 4°C , the sample was centrifuged for 10 min at $12,000 \times g$ at 4°C . The RNA precipitate formed a white gel-like pellet, which was resuspended in 1 mL of 75 % ethanol per 1 mL of TRIzol reagent used for lysis. The sample was briefly vortexed and centrifuged for 5 min at $7500 \times g$ at 4°C . The supernatant was discarded, and the RNA pellet was air-dried for 5–10 min. Finally, the pellet was resuspended in 20–50 μL of RNase-free water, 0.1 mM EDTA, or 0.5 % SDS solution by

pipetting up and down. The RNA could be stored in 75 % ethanol for at least 1 year at -20°C or at least 1 week at 4°C .

Library construction. Total RNA samples were subjected to quality control (QC) using an Agilent 2100 Bioanalyzer with an Agilent RNA 6000 Nano Kit. The QC assessment involved measuring RNA concentration and RIN value, evaluating the 28S/18S ratio, and analyzing the fragment length distribution. mRNA isolation from total RNA was performed using the oligo(dT) method. Subsequently, the mRNA was fragmented and fragments of 200 to 300 bases were purified. First strand cDNA and second strand cDNA were then synthesized. The resulting cDNA fragments were purified and treated with EB buffer for end repair, followed by the addition of a single nucleotide (adenine). Adapters were ligated to the cDNA fragments, and PCR amplification was conducted to obtain cDNA fragments of the appropriate size. The Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System were utilized to quantify and qualify the resulting libraries.

Library sequencing: Libraries were sequenced using the Illumina HiSeq sequencing platform. Base quality values were assessed using the Illumina GA Pipeline v1.5. The Illumina HiSeq2000/2500 system's quality value system was Phred+64. The sequencing data were stored in the Sanger FASTQ file format [3], which includes quality scores for each sequence.

Sequence reads filtering. The sequence reads were filtered to remove low-quality reads, reads containing adaptors, and reads with a high content of unknown bases (N). The filtering steps were as follows: Reads containing adaptors were removed and reads with more than 5 % of unknown bases (N) were discarded. Low-quality reads were filtered out using the following criterion: if the percentage of bases with a quality score less than 10 exceeded 20 % within a read, it was considered a low-quality read and discarded. The filtered reads, referred to as "Clean Reads," were retained for downstream analyses. The clean reads were stored in the FASTQ format for further analysis.

Tables, graphs, images and charts plotting. Gene expression data was obtained and tabulated in the form of Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values. To ensure accurate analysis and visualization, zero FPKM values were replaced with NaNs to avoid issues with logarithmic transformations.

For the identification of differentially expressed genes, we calculated \log_2 fold changes in gene expression for each genotype relative to the control. Genes with at least a 2-fold change in expression (\log_2 fold change of ± 1) were identified. Pseudo p-values were calculated based on these fold changes, and the Benjamini-Hochberg procedure [4] was applied to control the false discovery rate (FDR) across multiple comparisons.

All visualizations and statistical analyses were conducted using Python with the pandas, numpy, seaborn, matplotlib, and statsmodels libraries. Up regulated and down regulated genes are reported in Table 4.

Bioinformatics workflow. Low-quality reads were filtered using the previously described criterion. Next, the clean reads were assembled into unigenes, followed by functional annotation. The workflow also included the detection of simple sequence repeats (SSRs) and calculating unigene expression levels. Differentially expressed genes (DEGs) between samples were identified in the final stages. Clustering analysis was performed, and functional annotations of the DEGs were carried out to gain further insights into their biological relevance.

De novo assembly. *De novo* assembly was performed using Trinity on the clean reads, where PCR duplicates were removed to enhance efficiency [5]. The three Trinity independent software modules (Inchworm, Chrysalis, and Butterfly) were used.

Inchworm assembled reads into unique transcript sequences, including full-length transcripts for the most frequent isoforms and unique portions of alternatively spliced transcripts. Chrysalis clustered Inchworm contigs. Butterfly was used for reporting full-length transcripts of alternatively spliced isoforms and distinguishing transcripts corresponding to paralogous genes.

The resulting assembly from the Trinity of a list of transcripts was identified. TGICL [6] was used for gene family clustering, yielding the final unigenes. As multiple samples were analyzed, TGICL was executed separately for each unigene to obtain the final set for downstream analysis. The unigenes were categorized into two types: clusters, identified with the prefix "CL," followed

by a cluster ID (representing several unigenes with >70 % similarity), and singletons, denoted as unigenes.

software: <https://github.com/trinityrnaseq/trinityrnaseq/wiki/Output-of-Trinity-Assembly>

Unigene functional annotation. The following functional annotation methods were used: NT, NR, GO, KOG, KEGG, SwissProt, and InterPro. Unigenes were annotated to the NT, NR, KOG, KEGG, and SwissProt databases using Blastn, Blastx [7], or Diamond [8]. GO annotation was performed using Blast2GO [9] with NR annotation, and InterPro annotation was conducted using InterProScan5 [10]

The software versions and parameters used were as follows:

Blast: Version v2.2.23, default parameters. Website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Diamond: Version v0.8.31, default parameters. Website: <https://github.com/bbuchfink/diamond>

Blast2GO: Version v2.5.0, default parameters. Website: <https://www.blast2go.com>

InterProScan5: Version v5.11–51.0, default parameters. Website: <https://code.google.com/p/interproscan/wiki/Introduction>

The databases used for annotation were:

NT: Nucleotide database comprising sequences from various sources. Website: <ftp://ftp.ncbi.nlm.nih.gov/blast/db>

NR: Protein database including sequences from multiple sources. Website: <ftp://ftp.ncbi.nlm.nih.gov/blast/db>

GO: Gene Ontology database representing knowledge of gene functions. Website: <http://geneontology.org>

KOG: EuKaryotic Orthologous Groups database for identifying ortholog and paralog proteins. Website: <ftp://ftp.ncbi.nlm.nih.gov/pub/COG/KOG>

KEGG: Kyoto Encyclopedia of Genes and Genomes database for genomic, pathway, disease, and drug information. Website: <http://www.genome.jp/kegg>

SwissProt: Manually annotated protein sequence database providing high-quality information. Website: <http://ftp.ebi.ac.uk/pub/databases/swissprot>

InterPro: Resource for functional analysis of protein sequences, classifying them into families and predicting important sites. Website: <http://www.ebi.ac.uk/interpro>

Unigene CDs prediction. Transdecoder was used to identify candidate coding regions. The longest open reading frame (ORF) was extracted, followed by a search for Pfam protein homologous sequences using Blast to SwissProt and Hmmscan to predict the coding region.

The software versions and parameters used were as follows:

TransDecoder:

Version: v3.0.1

Parameters: default

Website: <https://transdecoder.github.io>

Unigene TF prediction. The unigenes were mapped to the AnimalTFDB2.0 database to identify corresponding TF families, Ensembl gene IDs, and database linkages. Through these linkages, we obtained access to the TF families' to obtain genetic information, functions, and binding sites. The sequence of each unigene was obtained using getorf, after which the ORFs were aligned to TF domains from the AnimalTFDB2.0 database using hmmsearch. TFs were identified based on the descriptions annotated in the AnimalTFDB2.0 database [11,12].

The software versions, parameters, and database used were as follows:

getorf:

Version: EMBOSS:6.5.7.0

Parameters: -minsize 150

Website: <http://genome.csdb.cn/cgi-bin/emboss/help/getorf>

hmmsearch:

Version: v3.0

Parameters: default

Website: <http://hmmer.org>

Website: <http://www.bioguo.org/AnimalTFDB/>

Unigene expression analysis. The clean reads were mapped to the unigenes using Bowtie2 [13], and calculating the gene expression levels were calculated using RSEM [14]. Hierarchical clustering analysis was performed using the hclust function, and PCA analysis was conducted using the princomp function in R software.

The software versions and parameters used were as follows:

Bowtie2 Version: v2.2.5

Parameters: -q -phred64 -sensitive -dpad 0 -gbar 99999999 -mp 1,1 -np 1 -score-min L,0,-0.1 -I 1 -

X 1000 -no-mixed -no-discordant -p 1 -k 200

Website: <https://bowtie-bio.sourceforge.net/index.shtml>

RSEM:

Version: v1.2.12

Parameters: default

Website: <https://deweylab.github.io/RSEM/>

Limitations

As each of the samples of RNA sequenced came from a pool of individual samples the relative expression levels represent the average expression levels in the population.

Ethics Statement

This work did not involve human subjects or data collected from social media platforms. Animal management was performed accordingly to the ethical guidelines of our institutions. Animal care and use was according to the Mexican norm NOM-062-ZOO-1999 and its technical specifications for producing, caring, and use laboratory animals.

CRediT Author Statement

Raquel Cossío-Bayúgar: Conceptualization, writing-Original draft preparation, sample collection and purification. **Hugo Aguilar-Díaz:** Gene expression analysis, data curation. Estefan Miranda-Miranda: Data curation, Investigation, writing-Reviewing and editing. **Verónica Narváez-Padilla:** Data curation, investigation, writing-Reviewing and Editing. **Enrique Reynaud:** Conceptualization, writing-Original draft preparation, sample collection, and purification.

Data Availability

Transcriptomic dataset of the development and maturation of the *Rhipicephalus microplus* ovary (Original data) (Mendeley Data).

Rhipicephalus microplus raw ovary gene expression sequence reads (Original data) (NCBI Bio-Project)

Acknowledgments

We thank M.C. Rene Hernandez Vargas and Dr. Iván Sanchez Díaz for their technical support.

Funding

This research received funding DGAPA/UNAM, PAPIIT-IN210124.

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.

References

- [1] K. Clark, K. Pruitt, T. Tatusova, I. Mizrahi, BioProject, The NCBI Handbook, 2nd ed., National Center for Biotechnology Information (US), 2013 <https://www.ncbi.nlm.nih.gov/sites/books/NBK169438/> (Accessed 15 February 2024).
- [2] R. Cossío-Bayúgar, E. Miranda-Miranda, V.N. Padilla, F. Olvera-Valencia, E. Reynaud, Perturbation of tyraminer-gic/octopaminergic function inhibits oviposition in the cattle tick *Rhipicephalus (Boophilus) microplus*, *J. Insect Physiol.* 58 (2012) 628–633.
- [3] P.J.A. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, *Nucleic Acids Res.* 38 (2010) 1767, doi:10.1093/nar/gkp1137.
- [4] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Series B* 57 (1995) 289–300, doi:10.1111/j.2517-6161.1995.tb02031.x.
- [5] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644–652, doi:10.1038/nbt.1883.
- [6] G. Pertea, X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai, J. Quackenbush, TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets, *Bioinformatics* 19 (2003) 651–652, doi:10.1093/bioinformatics/btg034.
- [7] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410, doi:10.1016/S0022-2836(05)80360-2.
- [8] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods* 12 (2015) 59–60, doi:10.1038/nmeth.3176.
- [9] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics* 21 (2005) 3674–3676, doi:10.1093/bioinformatics/bti610.
- [10] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier, *Nucleic Acids Res.* 33 (2005) W116–W120, doi:10.1093/nar/gki442.
- [11] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, *Trends Genet.* 16 (2000) 276–277, doi:10.1016/s0168-9525(00)02024-2.
- [12] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions, *Nucleic Acids Res.* 41 (2013) e121, doi:10.1093/nar/gkt263.
- [13] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359, doi:10.1038/nmeth.1923.
- [14] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics* 12 (2011) 323, doi:10.1186/1471-2105-12-323.