Published on Jan 02, 2025 by Arcadia Science **DOI**: 10.57844/arcadia-6833-4117

An interactive visualization tool for *Amblyomma americanum* differential expression data

We analyzed RNA-seq data from *Amblyomma americanum* to explore gene expression linked to skin manipulation during tick feeding. We built an interactive app to explore the differential expression results and find patterns related to tick sex, tissue, and time in blood meal.

Contributors (A-Z)

Audrey Bell, Adair L. Borges, Feridun Mert Celebi, Elizabeth A. McDaniel, Taylor Reiter, Emily C.P. Weiss

Version 1 · Mar 31, 2025

Purpose

Ticks have evolved to feed on host blood undetected. Female ticks take long “blood meals” that can last over a week. These ticks use molecules in their saliva to manipulate host pathways and evade the immune system. Some of these molecules may have therapeutic benefits for humans, particularly in managing itch and inflammation. These molecules are likely produced in the female tick salivary glands,

potentially at higher levels than in males or other tissues. Investigating differential gene expression could help identify anti-itch or anti-inflammatory molecules.

We re-analyzed public RNA-seq data from *A. americanum* ticks, focusing on variables such as sex, tissue type, and feeding time. Though batch effects and a lack of replicates limited the number of samples we could analyze, we were able to compare 20 of 56 RNA-seq samples using two differential expression models. The first model compared different tissues within and between sexes, while the second also included time since the start of a blood meal. We developed an interactive application to explore the results, aiming to identify tick molecules that manipulate skin pathways.

Our primary audience is researchers interested in identifying new therapeutic proteins or molecules in female tick salivary glands. We envision these researchers using this tool as a complement to other genetic or molecular discovery approaches. For example, a researcher who's identified protease inhibitor genes in the *A. americanum* genome could narrow this list down to those most likely to interact with the host by using the app to identify which protease inhibitors are expressed in the salivary gland.

- This pub is part of the **project**, "Ticks as treasure troves: Molecular discovery in new organisms." Visit the project narrative for more background and context.
- You can find **code** for the creation of the differential expression models and for the Shiny app, along with usage instructions, in this GitHub repo.

The context

Ticks (order *Ixodida*) are parasitic insects that feed on the blood of animals. There are over 800 recognized species of ticks on the planet **[1]**. Ticks have evolved many ways to evade host detection during a blood meal **[2]**. The blood meal can last up to a week in hard-bodied, slow-feeding ticks. Only females participate in these long blood meals — males may feed intermittently from a host but primarily seek out hosts to mate with females **[3]**.

Prolonged female feeding requires many adaptations for the tick to remain attached to the host undetected and to maintain blood flow throughout the meal. These include

strategies to maintain blood supply by overcoming platelet aggregation and blood coagulation and to hide from the host by blocking itch, pain, and some immune system activities **[2]**. Saliva delivers molecules and proteins that achieve these actions.

While ticks use these adaptations to get host blood, we expect that many of the molecules they use to manipulate host pathways would have therapeutic benefits if we could co-opt them for use in humans. At Arcadia, we're particularly interested in molecules or proteins that manipulate skin pathways involved in itch and inflammation. Examples of tick saliva molecules that manipulate skin pathways include votucalis, which sequesters histamine to attenuate itch **[4]**, evasin P991_AMBCA, which binds to inflammation-causing chemokines **[5]**, and a carboxypeptidase that cleaves the vasodilator and pain-inducing peptide bradykinin **[6]**.

We expect the salivary glands of female ticks to express the most proteins with potential therapeutic activities. These organs produce saliva, which is secreted at the feeding site and manipulates the biology of the host. Given that the biology of tick feeding varies during feeding **[3][7]** — manipulating distinct biological pathways at different points — the temporal expression of a gene may offer insights into its functional role. Similarly, for candidate proteins of therapeutic interest, understanding a gene's expression pattern can guide us in determining the optimal time to harvest tick tissues if this is relevant to the experiment.

Taken together, gene expression in female tick salivary glands may offer clues for uncovering the molecules that manipulate human skin pathways.

What expression data do we have to work with

The relative affordability of next-generation sequencing combined with publishing mandates for open data have produced an abundance of public sequencing data. RNA sequencing in particular is a popular modality in part because it does not require a reference genome to gain insight into an organism's gene expression. While RNA-seq data is often generated to answer specific research questions, the comprehensive nature of the data means it can be reanalyzed or repurposed to investigate other biological questions beyond the scope of the original research. This makes RNA-seq data valuable and reusable for different studies. At the same time, RNA-seq data often have strong batch effects (non-biological dataset-to-dataset variation) from things like sample handling, RNA extraction protocol or kit, sequencer, and genomic

heterozygosity of a species **[8]**. No matter their source, major batch effects prevent comparison between samples. Biological replicates (minimum two) and balanced experimental designs are also necessary to compare many samples with differential expression. For example, we have to discard some samples if a condition doesn't have a replicate or we don't have a mirrored condition sample to compare it to (e.g., male vs. female).

RNA-seq experiments have been popular in *Amblyomma americanum*, the lone star tick **[9][10][11][12]**. *A. americanum* has a range that covers most of the Eastern United States and is in part responsible for the doubling of tick-borne diseases from 2004–2016 **[13][14][15]**. Given the potential importance of tick saliva in the transmission of tick-borne pathogens and its role in the development of alpha-gal syndrome, many of these RNA-seq studies include samples from female salivary glands as well as other tissues, such as the mid-gut and samples from male ticks. Contrasting these samples may highlight gene expression profiles specific to female salivary glands and uncover key mechanisms of host manipulation. However, batch effects may prevent unified analysis because these samples originated from different studies. This is particularly true for *A. americanum*, which has high genetic diversity that clusters by population **[16]**.

We wanted to assess whether we could use public data to investigate the genes *A. americanum* expresses when interacting with a host. Focusing on variables such as sex, tissue type, and time during the blood meal, our goal was to develop differential expression models with DESeq2. Differential expression models make statistical comparisons between normalized gene counts to determine which genes are induced or repressed in different conditions. These models may help us identify specific tick molecules expressed in the salivary glands of females, which likely manipulate host skin pathways at different feeding stages **[3][7]**, providing insights into tick biology and host manipulation.

Our approach to visualizing tick differential expression

We re-analyzed public RNA-seq data from *A. americanum* for differential expression analysis. We identified samples that clustered according to biological variables rather than the originating study. These samples allowed us to perform differential expression

modeling based on variables like sex, tissue, and time in blood meal (the number of hours the sample was taken after feeding began). We then developed an interactive R Shiny app to make it easy to explore these results. The application includes key RNA-seq analyses and visualizations like principal component analysis plots, volcano plots, MA plots (log ratio mean average plots), and gene plots, as well as tools to summarize gene expression by condition. Users can control differential expression results by metrics like $\log_2$ fold change, p-value, or average gene count per gene.

The resource

> You can find **code** for the creation of the differential expression models and for the Shiny app, along with usage instructions, in this GitHub repo (DOI: 10.5281/zenodo.14548891).

Building differential expression models

Finding samples

We started our analysis by identifying publicly available Illumina RNA-seq samples. Using the NCBI Taxonomy page, we searched for “Amblyomma americanum” in August 2023 and followed the Entrez records link to SRA Experiments. We then used the SRA filtering tools to limit the results to RNA-seq sequencing data from Illumina sequencing chemistries: `txid6943[Organism:noexp] AND "biomol rna"[Properties] AND "platform illumina"[Properties] AND "strategy RNA-Seq"[All Fields]`. A summary of the samples is included in Table 1.

SRA study accession	Number of samples	Number of samples included in differential expression analysis	Reference
SRP051699	4	4	**[7]**
SRP091404	6	2 (replicates)	**[12]**
SRP052078; SRP052091; SRP052108; SRP052106; SRP052114; SRP052123; SRP052145; SRP052154	8	0 (batch effects)	Arthropod cell line
SRP032795	14	14	**[10]**
SRP446981	24	0 (batch effects)	**[9]**

Table 1

Summary of publicly available RNA-seq data analyzed in this project.

In cases where we had to omit samples from our analysis, we've noted the reason for omission in parentheses in the third column.

View the complete set of samples and metadata analyzed in this project.

Creating gene counts

We next processed these RNA-seq samples into gene counts.

We first **downloaded reads** with SRA Tools (version 3.0.6) `fasterq-dump` and quality- and adapter-trimmed reads with fastp (version 0.23.4) **[17]**.

We **quantified transcripts** using Salmon (version 1.10.2) **[18]** against an *A. americanum* transcriptome assembly ("Amblyomma_americanum_transcriptome_assembly_data.tar.gz") **[19]** to quantify read counts.

While Salmon produces transcript (isoform) counts, differential expression results are more accurate when comparing gene counts **[20]**. The most common way to assign transcript isoform-level counts to their parent genes is to use a transcript-to-gene

mapping file. The R package tximport uses a `tx2gene` file to sum the counts for all transcripts that encode the same gene and to report the gene-level counts **[20]**.

To generate a transcript-to-gene map (`tx2gene` file) for gene-level quantification, we first **mapped the reference transcripts back to the genome** using uLTRA (version 0.1) **[21]**.

Next, we **assigned a gene name to a transcript when it overlapped with part of the gene's interval** as annotated in the *A. americanum* genome annotation GFF3 ("Amblyomma_americanum_annotation_data.tar.gz") **[19]**. After making these files, we imported transcript counts and summarized them to gene counts using the tximport package (version 1.28.0) function `tximport()` with the parameter `type = salmon` **[20]**.

> View the **workflow code** for creating gene counts from RNA-seq accessions on GitHub.

Picking samples to include in the differential expression analysis

We next assessed which samples we could compare via differential expression analysis. Using the gene counts generated above, we assessed similarities between samples as well as conditions captured in the metadata. Our exclusion criteria included samples without replicates (minimum of n = 2 per condition) and batch effects that led to samples clustering more strongly by study than by condition (as determined by eye).

We first eliminated four samples from study SRP091404 **[12]**. This study investigated changes to the transcriptome of *A. americanum* during infection with *Ehrlichia chaffeensis* **[12]**, a tick-borne pathogen primarily transmitted by *A. americanum* **[22]**. Two of the six samples captured whole, uninfected ticks while the other four captured infected ticks. Since no RNA-seq samples in other studies were exposed to this pathogen, we could not account for infection with *E. chaffeensis* as a variable in a differential expression model, so we eliminated these four samples. Further, these samples didn't have replicates within the study, which also prevented analysis by differential expression.

We next used principal component analysis to determine if batch effects led to samples clustering more by study than by biological condition (Figure 1). This analysis excluded 32 samples from two studies (SRP446981 and “Arthropod cell line”).

We eliminated all 24 samples from SRP446981 **[9]**. This study analyzed the transcriptome response of *A. americanum* to *Escherichia coli* challenge. Initially used for differential expression analysis, it featured unfed female ticks injected with either phosphate-buffered saline or *E. coli* and analyzed whole (i.e., sampling all tissues). All samples from the study clustered tightly together and away from whole-tick samples from other studies, indicating the batch effects were too strong to make cross-study comparisons. It’s possible that the injections caused a biological impact that led to these batch effects, but we can’t evaluate this with the available data.

Last, we eliminated eight samples from the “Arthropod cell line” sequencing effort. These eight samples originate from two *A. americanum* cell lines. The samples all cluster tightly together and away from other samples. Since we have no other cell line data from other studies, there’s no way to evaluate whether these samples cluster alone because they have different expression or strong batch effects.

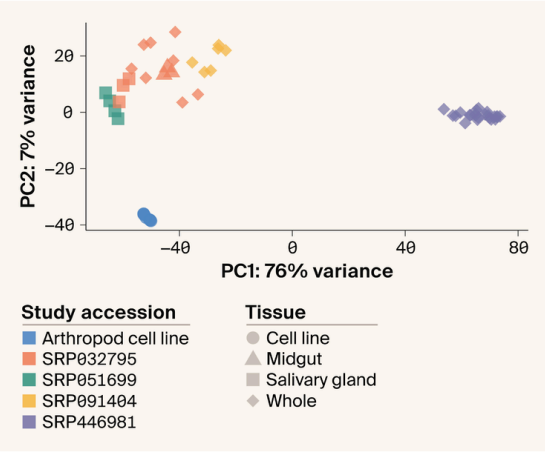

Figure 1

Principal component analysis of normalized gene counts for the top 500 genes by variance across RNA-seq samples.

Sample color corresponds to the original study that published the RNA-seq data (see Table 1) while sample shape corresponds to the type of tick tissue from which the sample originated. “Arthropod cell line” and SRP446981 data group by study accession instead of by tissue type, while other samples group by tissue type. PC: Principal component. The percentage of the variance explained by PC1 and PC2 is reported in each axis label.

Building the differential expression models

After filtering the data, samples from different tissues, sexes, and taken at different times during a blood meal remained. We next performed a differential expression analysis so we could compare these samples. We built differential expression models using the DESeq2 package (version 1.40.2) using commands `DESeqDataSetFromTximport()` and `DESeq()` **[23]**. We experimented with different model structures using the `design` parameter to maximize the number of samples and conditions that we could compare. In the end, we were able to optimize these two factors with two models. In both cases, we combined the variables we included in the model. In DESeq2 analysis, the model matrix must be "full rank" to prevent variables from being redundant, which can skew the results. We simplified the model by combining variables, ensuring each variable is distinct. This lets DESeq2 accurately calculate and attribute effects to each variable independently **[23]**.

We were able to include all samples when we combined the variables “sex” and “tissue” (Table 2).

Model “sex_tissue”	Number of samples
female_x_midgut	3
female_x_salivary_gland	7
female_x_whole	5
male_x_whole	5

Table 2

The variables and number of samples included in the “sex_tissue” model.

The combined variables are separated by an “x.”

While this model included all samples that we could compare, it gives no insight into how gene expression varies based on time in the blood meal. Given this, we built a second model that included time in the blood meal (by hour), only including samples with replicates for different times in the blood meal (Table 3).

Model “sex_tissue_blood_meal_hour”	Number of samples
female_x_midgut_x_72_144	2
female_x_salivary_gland_x_12_48	2
female_x_salivary_gland_x_72_144	3
female_x_whole_x_72_144	2
male_x_whole_x_72_144	3

Table 3

The variables and number of samples included in the “sex_tissue_blood_meal_hour” model.

The combined variables are separated by an “x.” The numbers in the condition names indicate the range of hours in the blood meal from which samples were taken.

In general, it’s better to analyze all samples in a single model, but given the limitations of working with this data, we worked within the bounds of what was statistically possible in order to achieve the most biological insight.

We don’t present detailed results here as they change based on filtering and the specific conditions compared. However, we include a summary of the number of differentially expressed genes using default filters ($\log_2$ fold change = 2, false discovery rate = 0.05, base mean count = 10), focused on differential expression in female salivary glands (Table 4). We observed expected ranges of expression and genes with larger expression differences across more distinct conditions, giving us confidence that our models are useful.

Condition 1	Condition 2	Induced	Repressed
Female salivary gland	Female midgut	183	358
Female salivary gland	Female whole tick	240	631
Female salivary gland	Male whole tick	256	946
Female salivary gland, 72–144 h	Female salivary gland, 12–48 h	5	1
Female salivary gland, 72–144 h	Female whole tick, 72–144 h	12	48
Female salivary gland, 72–144 h	Male whole tick, 72–144 h	31	190

Table 4

The number of differentially expressed genes in female salivary glands when compared to different conditions.

The first three contrasts are from the “sex_tissue” model while the last three are from the “sex_tissue_blood_meal_hour” model. Induced

and repressed genes have positive and negative $\log_2$ fold changes, respectively. Only genes with a base mean count greater than 10 and a false discovery rate less than 0.05 are included.

Interactive application for exploring gene expression in *A. americanum*

We built the above differential expression models to facilitate insights into *A. americanum* gene expression. We included the maximum number of variables and conditions possible given what is available in public data. However, we wanted to let others explore these results with different biological questions in mind. For example, researchers could use our data to identify the conditions under which their gene of interest is most highly expressed. Likewise, it could be used to evaluate whether a gene of interest exhibits stronger sex-associated or time-associated effects. These individual biological applications would be difficult to anticipate and share in written pub format. To give researchers flexible access to this data, we wrote an interactive

Shiny app to explore the gene count and differential expression results. Instructions for using the app are available on the GitHub repository along with a mapping table (shiny/mapped_gene_names_GCA_030143305.2.csv) that enables conversion of GenBank protein IDs from the *A. americanum* assembly to the gene names used in the app.

The application features several analytical tools separated into different tabs. Users first select which model they want to explore (Figure 2). They can then visualize the samples in the model in a metadata table and a principal component analysis plot colored by variables from the model (Figure 2). The “Differential Expression (DE) Analysis” tab offers filtering capabilities based on $\log_2$ fold change, p-value, and mean count of the gene, supporting detailed comparisons across different conditions (Figure 2). Users can also upload their data in this tab to highlight whether specific genes of interest are differentially expressed.

Figure 2

Preview of the "Differential Expression Explorer" for *Amblyomma americanum*.

The Shiny app allows users to explore two differential expression models: one built on the variables sex and tissue and one built on the additional variable time since blood meal started. Users can toggle between the two models using a drop-down menu. Four tabs present different information from these models. The first, "PCA Plot," gives an overview of the samples included in each model using a principal component analysis and a metadata table with information about each sample. The second tab, "DE Analysis," includes the differential expression analysis capabilities and visualizations. Users can select which conditions to contrast and thresholds for filtering results. The differentially expressed genes are then plotted in interactive plots. The third tab, "Gene," allows users to visualize normalized counts per gene. These plots are helpful to determine if a few outlier samples drive differential expression of a gene of interest. The last tab, "Expression by Condition," allows users to see which genes are expressed in each condition. The toggle on this tab allows users to highlight genes that have higher relative expression.

The application includes functionality for gene-specific inquiries where users can input a gene name to generate a boxplot displaying expression across different conditions, offering a granular view of gene activity (Figure 2). Furthermore, the "Expression by

Condition" tab provides a table that reports gene expression thresholds and percentiles, allowing users to filter and download gene expression data (Figure 2). This tab is particularly useful for identifying genes that are consistently or exclusively expressed in specific tissues like salivary glands.

Currently, the transcript and gene names used in our pipeline are the bespoke annotations assigned by intermediate tools in different pipelines. However, GenBank recently accepted our genome gene-boundary annotations **[19]**. Given this, we have also provided a mapping table that can be used to map NCBI protein identifiers to the names used in our app. The mapping table is available in the GitHub repository.

Additional methods

We used ChatGPT as a starting point to put our code into a Shiny app and adjusted ChatGPT's outputs. It also suggested wording ideas and edits, and we picked and chose which bits to use.

We also provided Notion AI with starting text and had it rearrange that text to fit the structure of one of our pub templates, and then edited that output.

Key takeaways

1. Re-analyzing public RNA-seq data from the tick species *Amblyomma americanum* let us construct two distinct models for assessing differential gene expression, though major batch effects and lack of sufficient replicates limited the number of samples we could include in our analysis.
2. Our differential expression models let users compare variables like sex (male, female), tissue type (whole tick, midgut, and salivary gland), and timing in the blood meal. Users can identify gene expression patterns potentially linked to skin pathway manipulation by comparing these variables.
3. Our interactive Shiny app provides a user-friendly platform for exploring differential expression experiments. This application features tools for visualization and analysis, including principal component analysis, volcano plots, and gene count plots, and allows for results-filtering based on significance or expression levels.

Next steps

We'd like to improve this resource in two specific ways:

1. **Updating the analysis and Shiny app to include *A. americanum* gene GenBank identifiers**: Currently, the transcript and gene names used in our pipeline are the bespoke annotations assigned by intermediate tools in different pipelines. However, GenBank recently accepted our genome-gene boundary annotations **[19]**. Given this, we may update the Shiny app to include these identifiers to make the analysis experience more consistent across public resources. As a quick fix, we provide a mapping file that allows a user to correspond NCBI gene identifiers to our internal gene names.
2. **Adding new RNA-seq samples to the models**: In March 2024, the National Institute of Allergy and Infectious Disease released 21 new RNA-seq samples from the mid-gut of *A. americanum*. As more RNA-seq samples are released, or if we sequence samples ourselves, we could add these new samples to this analysis. We would have to check new samples for batch effects and assess whether the current model matrix could include more samples.

If you use our Shiny app, analysis, or any part of our code, we'd love to hear how it works for you. Any feedback on issues or potential useful features to add is welcome.

References

1 BARKER SC, MURRELL A. (2004). Systematics and evolution of ticks with a list of valid genus and species names. https://doi.org/10.1017/s0031182004005207

2 Mans BJ, Neitz AWH. (2004). Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. https://doi.org/10.1016/j.ibmb.2003.09.002

3 Aljamali MN, Ramakrishnan VG, Weng H, Tucker JS, Sauer JR, Essenberg RC. (2009). Microarray analysis of gene expression changes in feeding female and

male lone star ticks, *Amblyomma americanum* (L). https://doi.org/10.1002/arch.20318

4 Alrashdi I, Alsubaiyel A, Chan M, Battell EE, Ennaceur A, Nunn MA, Weston-Davies W, Chazot PL, Obara I. (2022). Votucalis, a Novel Centrally Sparing Histamine-Binding Protein, Attenuates Histaminergic Itch and Neuropathic Pain in Mice. https://doi.org/10.3389/fphar.2022.846683

5 Singh K, Davies G, Alenazi Y, Eaton JRO, Kawamura A, Bhattacharya S. (2017). Yeast surface display identifies a family of evasins from ticks with novel polyvalent CC chemokine-binding activities. https://doi.org/10.1038/s41598-017-04378-1

6 Ribeiro JMC, Mather TN. (1998). Ixodes scapularis:Salivary Kininase Activity Is a Metallo Dipeptidyl Carboxypeptidase. https://doi.org/10.1006/expr.1998.4296

7 Kim TK, Tirloni L, Pinto AFM, Diedrich JK, Moresco JJ, Yates JR, da Silva Vaz I, Mulenga A. (2020). Time-resolved proteomic profile of Amblyomma americanum tick saliva during feeding. https://doi.org/10.1371/journal.pntd.0007758

8 Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. https://doi.org/10.1038/nrg2825

9 Lyu B, Li J, Niemeyer B, Anderson DM, Beerntsen B, Song Q. (2023). Integrative analysis highlights molecular and immune responses of tick Amblyomma americanum to Escherichia coli challenge. https://doi.org/10.3389/fcimb.2023.1236785

10 Radulović ŽM, Kim TK, Porter LM, Sze S-H, Lewis L, Mulenga A. (2014). A 24-48 h fed Amblyomma americanum tick saliva immuno-proteome. https://doi.org/10.1186/1471-2164-15-518

11 Karim S, Ribeiro JMC. (2015). An Insight into the Sialome of the Lone Star Tick, Amblyomma americanum, with a Glimpse on Its Time Dependent Gene Expression. https://doi.org/10.1371/journal.pone.0131292

12 Kim D, Jaworski DC, Cheng C, Nair ADS, Ganta RR, Herndon N, Brown S, Park Y. (2018). The transcriptome of the lone star tick, Amblyomma americanum, reveals molecular changes in response to infection with the pathogen, Ehrlichia chaffeensis. https://doi.org/10.1016/j.aspen.2018.05.009

13 Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR. (2018). *Vital Signs*: Trends in Reported Vectorborne

Disease Cases — United States and Territories, 2004–2016. https://doi.org/10.15585/mmwr.mm6717e1

14 Paules CI, Marston HD, Bloom ME, Fauci AS. (2018). Tickborne Diseases — Confronting a Growing Threat. https://doi.org/10.1056/nejmp1807870

15 The expanding spectrum of disease caused by the Lone Star Tick, Amblyomma americanum. (2021). https://doi.org/10.53854/liim-2903-8

16 Monzón JD, Atkinson EG, Henn BM, Benach JL. (2016). Population and Evolutionary Genomics of*Amblyomma americanum*, an Expanding Arthropod Disease Vector. https://doi.org/10.1093/gbe/evw080

17 Chen S, Zhou Y, Chen Y, Gu J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. https://doi.org/10.1093/bioinformatics/bty560

18 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. https://doi.org/10.1038/nmeth.4197

19 Celebi FM, Chou S, McDaniel EA, Reiter T, Weiss ECP. (2024). Predicted genes from the Amblyomma americanum draft genome assembly. https://doi.org/10.57844/ARCADIA-9602-3351

20 Soneson C, Love MI, Robinson MD. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. https://doi.org/10.12688/f1000research.7563.2

21 Sahlin K, Mäkinen V. (2021). Accurate spliced alignment of long RNA sequencing reads. https://doi.org/10.1093/bioinformatics/btab540

22 Childs JE, Paddock CD. (2003). The ascendancy of Amblyomma americanum as a vector of pathogens affecting humans in the United States. https://doi.org/10.1146/annurev.ento.48.091801.112728

23 Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. https://doi.org/10.1186/s13059-014-0550-8