5-2015

Differential gene expression between two deveopmental stages of the Gené's organ in the cattle fever tick, Rhipicephalus (Boophilus) microplus

Jason P. Tidwell
University of Texas-Pan American

DIFFERENTIAL GENE EXPRESSION BETWEEN TWO DEVEOPMENTAL STAGES OF THE GENÉ'S ORGAN IN THE CATTLE FEVER TICK, *Rhipicephalus (Boophilus) microplus*

A Thesis

By

JASON P. TIDWELL

Submitted to the Graduate School of
The University of Texas-Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2015

Major Subject: Biology

DIFFERENTIAL GENE EXPRESSION BETWEEN TWO DEVEOPMENTAL STAGES OF THE GENÉ'S ORGAN IN THE CATTLE FEVER TICK, *Rhipicephalus microplus*

A Thesis
by
JASON TIDWELL

COMMITTEE MEMBERS

Dr. Matthew Terry
Chair of Committee

Dr. Christopher Vitek
Committee Member

Dr. Robert Gilkerson
Committee Member

Dr. Donald Thomas
Committee Member

May 2015

ABSTRACT

Tidwell, Jason, Differential Gene Expression Between Two Developmental Stages Of The Gené's Organ In The Cattle Fever Tick, *Rhipicephalus microplus*. Master of Science (MS), May 2015, 49 pp., 6 tables, 28 figures, references, 84 titles.

The southern cattle fever tick (SCFT), *Rhipicephalus microplus*, is considered the most economically important ectoparasite of livestock worldwide and ranks sixth among the most pesticide resistant arthropods globally. This single-host hard tick is a vector of the infectious agents causing bovine babesiosis and anaplasmosis. The estimated impact of these SCFT-borne diseases on the livestock industry in the world is $7 billion annually. This project focuses on the Gené's organ and addresses the need to identify novel targets that can serve as the basis for sustainable SCFT control strategies. Wax secreted by the Gené's organ protects tick eggs from desiccation, bacterial and fungal infections, inorganic chemicals, and pesticides. We identified 1,513 new transcripts to add to the *R. microplus* genome database. A comparison of the two developmental stages of the Gene's organ through quantitative RNA-seq analysis revealed 5,454 differentially expressed transcripts. The RNA-seq data was confirmed through RT-PCR.

DEDICATION

The road to completion of a Master's degree in Biology has had some detours. It is thanks to the constant support from my family that I was able to accomplish this goal in my life. I would like to thank my parents, Steve and Sarah Tidwell. The example of my brothers and sisters and their encouragement were greatly appreciated. Without all of you, this would not have been possible. Thank you for putting up with me.

ACKNOWLEDGEMENTS

I want to thank Dr. Matthew Terry for his help and support during my time under his tutelage. Since both of us were relatively new to the realm of next generation sequencing and bioinformatic analysis of sequence data, we were both learning together how to overcome the obstacles that surfaced along the way. He is a great mentor and feel honored for the time that I had studying under him.

I am grateful to the United States Department of Agriculture for their support in my studies at the University of Texas Pan-American and their financial support in my research. Dr. Adalberto A Pérez de León was very supportive in my educational pursuits. He is the laboratory director for the Knipling-Bushland U.S. Livestock Insects Research Lab which has a sister lab outside of Edinburg at Moore Air Base, the Cattle Fever Tick Research Laboratory (CFTRL). The Cattle Fever Tick Research Laboratory was where the *Rhipicephalus microplus* colony is kept that was used in my research. Dr. Pérez de León was instrumental in introducing me to the iPlant Collaborative, which helped me immensely in cutting the time down in my bioinformatic analysis of the Gené's organ transcriptome data.

I want to thank Dr. Dannett Bishop, Kevin Temeyer, Felix Guerrero, and Evan Braswell with letting me bounce ideas off of them and giving me their expert advice that helped in the process of developing the experimental design for this research project.

The beginning of the bioinformatic analysis would not have happened if not for Graham Toal. He helped set up the high-performance computing cluster at the UTPA for bioinformatic research including trouble shooting the software when it wasn't working. I would also like to thank Drs. Miller, Thomas, and Vitek in their support and help along the way. Dr. Megan Keniry was extremely helpful in teaching me the particulars of real-time PCR. I would also like to thank my graduate committee for their support throughout my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Babesiosis, Cattle Fever

Babesiosis is a disease that is caused by an intraerythrocytic protozoan of the genus *Babesia* that infect a wide range of animals and occasionally man. It is a tick-borne disease that has a worldwide distribution. The cattle industry is mostly affected by babesiosis with *Babesia bovis* and *B. bigemina* being the two most important species[1].

Theobold Smith and Fred Kilborne were the first to identify the organism that caused 'Texas Fever' in 1893 which they named *Pyrosoma bigeminum* later becoming *Babesia bigemina.* They demonstrated that transmission of the disease organism could be done from an arthropod to a mammalian host when they determined that *B. bigemina* was transmitted by *Boophilus annulatus* to cattle[2].

Babesia bovis and *B. bigemina* have a range between 40˚N and 32˚S which coincides with its host natural range, the *Boophilus* ticks. In 2003 revision of the *Boophilus* tick species incorporated them into the *Rhipicephalus* genus[3]. Because of the amount of literature available and the common use of the name in the livestock community, *Boophilus* and *Rhipicephalus* both continue to be used when discussing cattle fever ticks. In this paper, *Rhipicephalus microplus* and *R. annulatus* will be used.

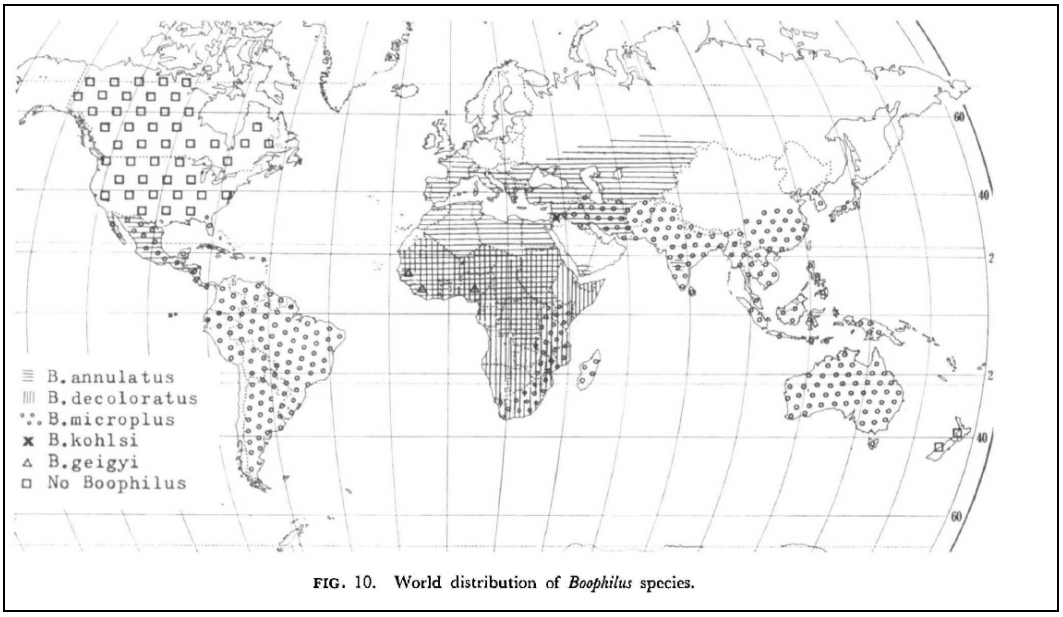

FIG. 10. World distribution of *Boophilus* species.

Figure 1: World distribution of *Rhipicephalus (Boophilus)* species[4].

Most of the cattle of the world lie within the range of *R. microplus* and *R. annulatus*, exposing them to babesiosis. When ranchers bring non-native bovine species into their herds that have not built up resistance to *Babesia*, costs due to babesiosis exposure increase. Mortality rates increase, abortions, loss of milk and meat production, and control measures are all costs due to babesiosis[1]. It is estimated to cost the agriculture industry worldwide $7 billion dollars per year[5].

The United States implemented the Cattle Fever Tick Eradication Program (CFTEP) in 1906; directed by the Animal and Plant Health Inspection Service[6]. The CFTEP required all cattle to be dipped in an organophosphate acaricide and it successfully pushed the range of the ticks out of the continental United States by 1943 and established a quarantine zone along the southern border with Mexico.

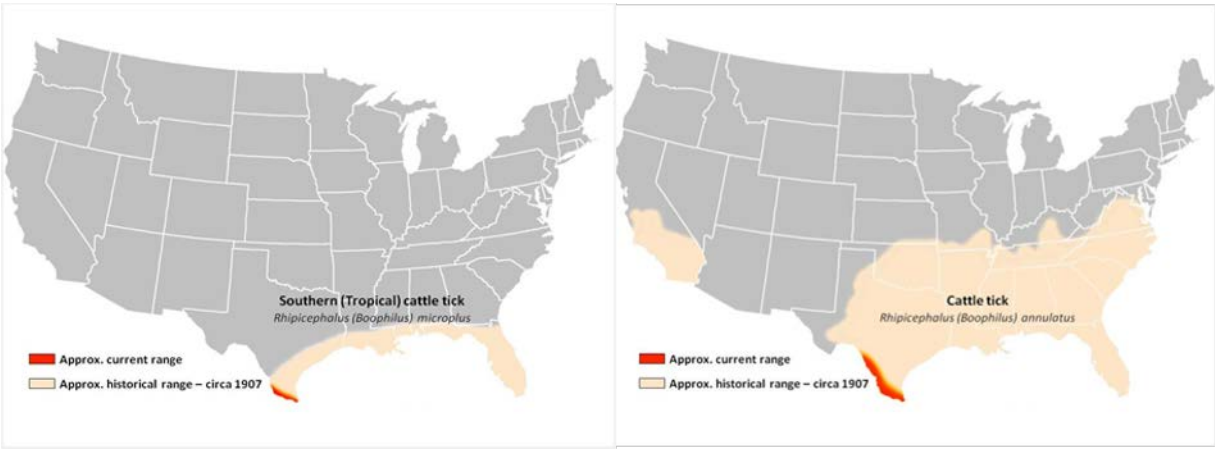

Figure 2: Historical and current distribution of *R. microplus* and *R. annulatus* in the United States[7].

There are four classes of acaricides used to treat cattle: organophosphates, pyrethroids, formamidines, and macrocyclic lactones. Every acaricide is a neurological agent that affects the nerve impulses of ticks and mites. Organophosphates affect acetylcholinesterase, which breaks down the neurotransmitter acetylcholine. Pyrethroids target the voltage-gated sodium channel. Formamidines bind to octopamine receptors. The majority of macrocyclic lactones target glutamate-gated chloride channels[8]. Each one is effective against susceptible strains, but a single point mutation in the sodium channel can make a cattle fever tick resistant to pyrethroids[9].

Mexico implemented an eradication program from 1974 to 1984 where organophosphates were the main method of control [10]. Resistance to organophosphates led to the use of pyrethroids and amitraz to combat the resistance to organophosphates. Due to the cost of amitraz, pyrethroids were more heavily used at first. When resistance to pyrethroids became known in 1993, amitraz was used to combat the resistant strains [11]. In 2001, Soberanes reported the first case of Amitraz resistance in *R. microplus* from Mexico [12]. Fipronil resistance was found in *R. microplus* from the Northern States of Mexico in 2013 [13]. Resistant strains of *R. microplus* to multiple acaricides were found in 15 states of Mexico [14]. In the United States, only one acaricide, the organophosphate coumaphos, is approved to be used by the Food and Drug

Administration. Already there have been strains tested that have shown up to a 53.2% survival rate after the first dipping treatment. After the second dipping, the control rate increased to 92.9% [15,16]. The control rate established by the Cattle Fever Tick Eradication Program is 95%.

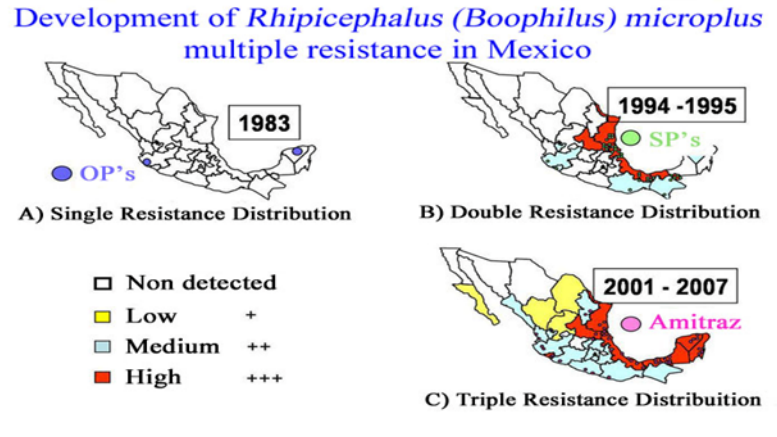

Figure 3: Map of distribution of acaricide resistance in Mexico from 1983 to 2007 with A) showing the first cases of resistance to organophosphates followed by dual resistance to synthetic pyrethroids in 1994-1995 (B). Map C shows the locations of *R. microplus* stains that acquired amitraz resistance along with OP and SP resistance [84].

Creating new derivations of neurological targeting acaricides may be one avenue to combat the growing resistance in *R. microplus.* The need for an alternative method to control *R. microplus* is imperative as the organism continues to evolve under the pressure of the control methods currently being used. The objective of this project is to better understand the biological pathways used in the Gené's organ and identify possible targets that could be used as a novel method of control of the southern cattle tick.

CHAPTER II

RHIPICEPHALUS MICROPLUS AND THE ORGAN OF GENÉ

Rhipicephalus microplus

In the suborder of Ixodida there are two families of ticks, "hard"ticks (Ixodidae) and "soft" ticks (Argasidae) which refers to the flexibility of their exoskeletons. The current hypothesis is that ticks evolved from being scavengers to obligate hematophagous organisms around 370 million years ago[17]. "Hard" tick life stages include egg, larva, nymph, and adult. There are also two different types of feeding characteristics among "hard" ticks in that some are multi-host (2- or 3-host) ticks and some are one-host ticks[18]. Ticks have four developmental stages in their life cycle: egg, larva, nymph, and adult[18]. Two- and three-host ticks complete each developmental life stage on a different host usually feeding on their host for one week to reach engorgement. The 3-host tick would then drop off their host and undergo ecdysis on the ground. They would then search for another host to attach and complete the next developmental stage. Once the adult finds a host, the female will not feed to repletion until it has been mated. Once mated, the female will feed until engorged and fall off the host to lay their egg mass that can contain upwards of 2000 eggs depending on the species of tick. *Rhipicephalus microplus* is a one-host tick and completes a majority of its development on a single host.

Tick anatomy is simple upon first inspection. The major male organs include the salivary glands, the gut, the testes, the rectal sac, and synganglion. The major female organs are the salivary glands, the gut, the ovaries, the rectal sac, synganglion, and the organ of Gené [19]. This study will focus on Gené's organ. In 1848, one of the earliest observations of ticks ovipositing described the effect of disturbing the organ, which caused the eggs to shrivel [20]. This organ is unique in that it everts itself outside the body of the tick to cover the eggs during oviposition. The organ of Gené is an evolutionary novelty unique to hard and soft ticks. Mites, the other suborder in Acari, do not have an organ of Gené. It doesn't become fully developed until the female is fully engorged. It takes the adult female seven days to reach full engorgement and drop off its host. The Gené's organ isn't fully developed until day 7 of adult feeding and after dropping from the host (personal observations). It is surmised that the trigger for the female tick to enter its rapid feeding cycle for full engorgement could be the same trigger for the tubular glands of Gené's organ to develop[21,22].

The different parts of the Gené's organ are the horn, the corpus, the tubular glands, the epithelial sac, and the lumen (Figure 5). The horn consists of inner cuticle that is compressed in layers so that it can evert from the camerostomal aperture during ovipositioning (Figure 7).

The wax composition produced by the Gené's organ in *Ixodes ricinus*[23], *Hyalomma savignyi*[20], and *Ornithordorus moubata*[24] have different compositions that correspond to the environments in which the species thrive. For example, *Ixodes ricinus* inhabit humid zones with a relative humidity of 80% and their egg wax composition is adapted to this type of environment; if these eggs are left in a dry environment, they lose 27% of their mass [20].

When30% of the weight of an egg is lost the mortality rate is approximately 50% and increases to 100% once the egg mass has lost 40% of its weight [25].

The protective wax that the Gené's organ produces is necessary for the survival of the egg even when the conditions are optimal [26]. The wax that is produced has several characteristics with its primary function as a water-proofing agent. It also has antibacterial and antifungal properties even such that it prevents biofilms of *Pseudomonas aeruginosa* from forming [23,27–30]. The wax is composed of branched and non-branched chain alkanes, steroids, fatty acids and alcohol [26,31,32]. Our project will use transcriptomics to identify the genes and biological pathways responsible for the generation of the waxes of the Gené's organ.

CHAPTER III

GENÉ'S ORGAN TRANSCRIPTOME ASSEMBLY

Introduction

Gené's organ is found in the craniodorsal region of the podosoma directly beneath the anterior part of the scutum. The morphology of Gené's organ is similar across species of ticks in that the structure consists of a double-sac structure [26,33,22,34]. The outer epithelial sacs (ES) and the inner cuticular sacs (CS) are separated forming a luminal area between them. The camerostomal aperture is where the Gené's organ everts to coat the eggs with its protective wax. The camerostomal aperture can be found between the posterior edge of the basis capituli and to the anterior edge of the scutum (Figure 4). The Gené's organ in unfed and ovipositing

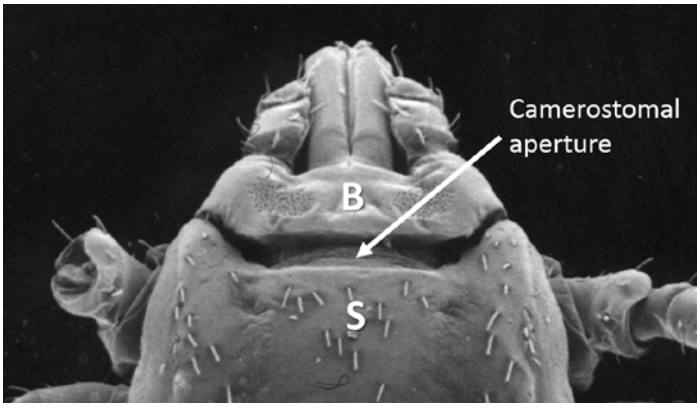

Figure 4: Electron micrograph of a female *R. microplus* displaying the location of the camerostomal aperture from which the Gené's organ emerges to coat the eggs during oviposition. B, basis capitulum; S, scutum (photo credit: @National Tick Collection 2004/www.discoverlife.org).

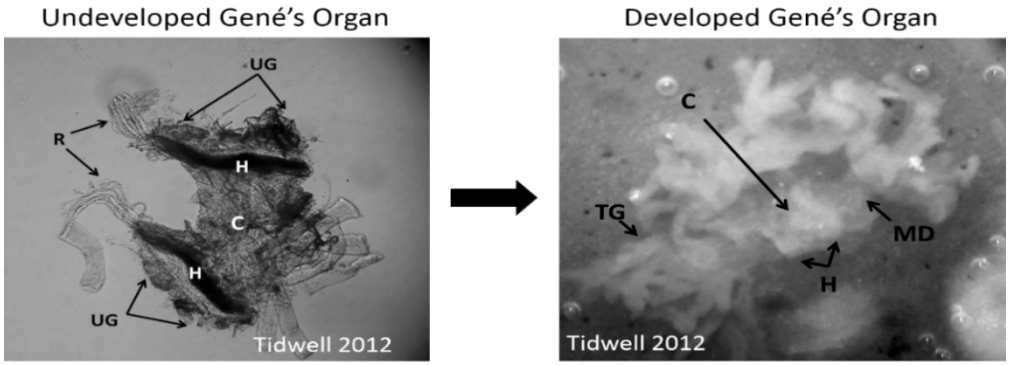

Figure 5: A comparison of the morphologically similarities and differences of the Gené's organ in unfed and ovipositing female *R. microplus*. R, retractor muscles; UG, undeveloped glands; H, horns; C, corpus; TG, tubular glands; and MD, main efferent duct of tubular glands.

female ticks consists of a corpus and 2 posterior horns. In unfed ticks, the Gené's organ has 4 small buds of undeveloped glands on either side (Figure 5). In ovipositing ticks, the undeveloped tubular glands have matured to form complex tubular glands with a duct into the lumen between the CS and ES (Figures 5,8-9)[35]. The developed tubular glands are similar to class 3 glands in arthropods [22].

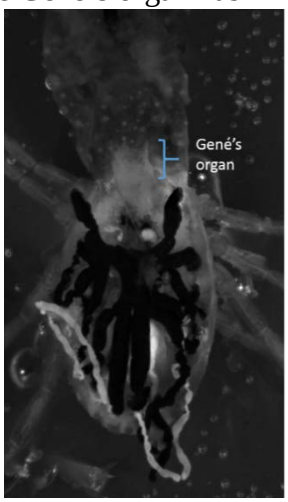

Figure 6: Photograph of unfed female adult *R. microplus* highlighting the location and size of the undeveloped Gené's organ.

In order to better understand the biological pathways and genes that are used in the Gené's organ, the transcriptome of two developmental stages of the organ were sequenced using next-generation sequencing technology. *Rhipicephalus microplus* is a non-model organism; a draft genome is currently available, however it only has a 0.7x coverage [36–38]. Because of the low coverage of the genome assembly, we did a *de novo* assembly of the transcriptome. Current efforts are focusing on sequencing transcriptomes of the organism in order to increase the coverage of the genome. Our Gené's organ transcriptome is the first transcriptome of Gené's organ in any hard or soft tick.

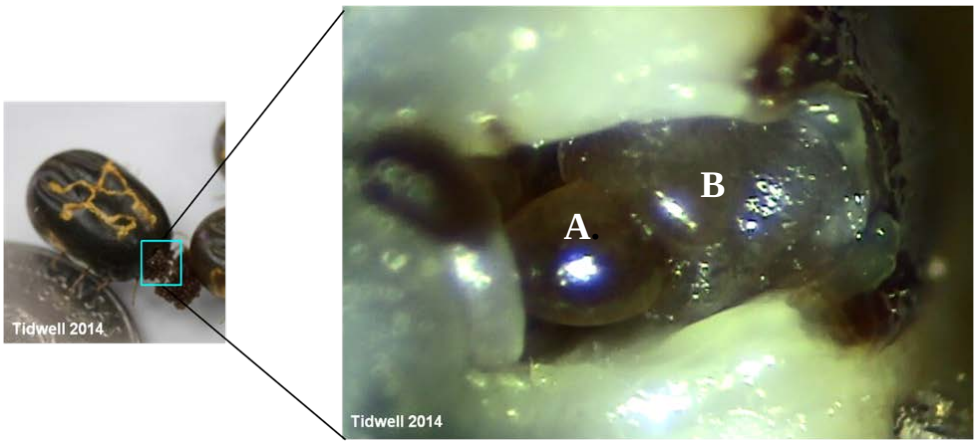

Figure 7: Ovipositing *R. microplus* female with the Gené's organ everted coating an egg. A) Egg leaving the genital pore. B) Everted Gené's organ.

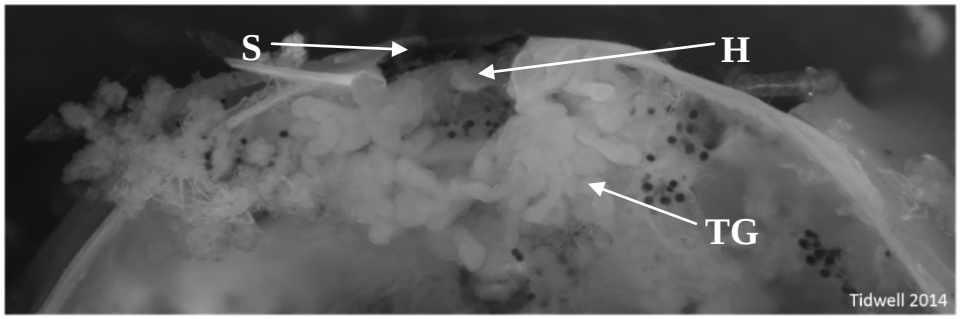

Figure 8: Picture of a developed Gené's organ in an engorged *R. microplus* female. S) Scutum. H) horns of Gené's organ. TG) tubular glands.

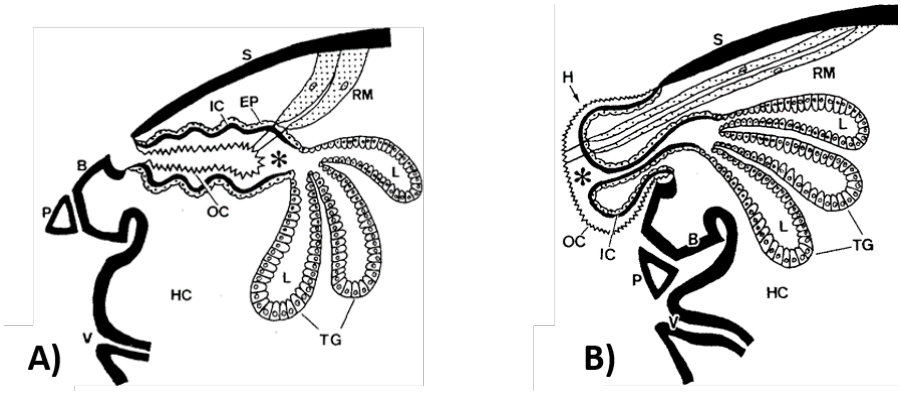

Figure 9: Diagram of the function and anatomy of Gené's organ. A) Relaxed form Gené's organ. B) Everted Gené's organ. The surface of the horn (H) covered with the everted outer cuticle (OC), the lumen (*) between the outer and inner (IC) cuticles filled with secretion of the tubular glands. B, basis capituli; EP, epithelial layer; HC, haemocoel; L, lumen of the tubular gland; P, palp; RM, retractor muscle of Gené's organ; S, scutum; TG, tubular gland; V, vulva [26].

Methods

The Deutsch strain of *Rhipicephalus microplus* was established from ticks collected in Webb County, TX in 2001. The tick colony has been maintained at the USDA-ARS Cattle Fever Tick Research Laboratory at 27˚C, 95% RH, and at a photoperiod of 12:12 (L:D) h [39] for fifty generations. The ticks and calves were determined to be free of *Babesia bovis* and *B. bigemina* [40]. The F_{47} generation was used in this experiment.

Patches of muslin cloth were glued to the shaved sides of a stanchioned Hereford calf (Figure 10). The first patch was infested with 250 mg of *R. microplus* Deutch F_{47} generation larvae with the same amount of larvae infested in the second patch two days later. Each life stage of the tick takes about 7 days to complete. Engorged nymphs were pulled from the calf 14 days after infestation and incubated at 26.7˚C and 95% RH. They were incubated until ecdysis had occurred.

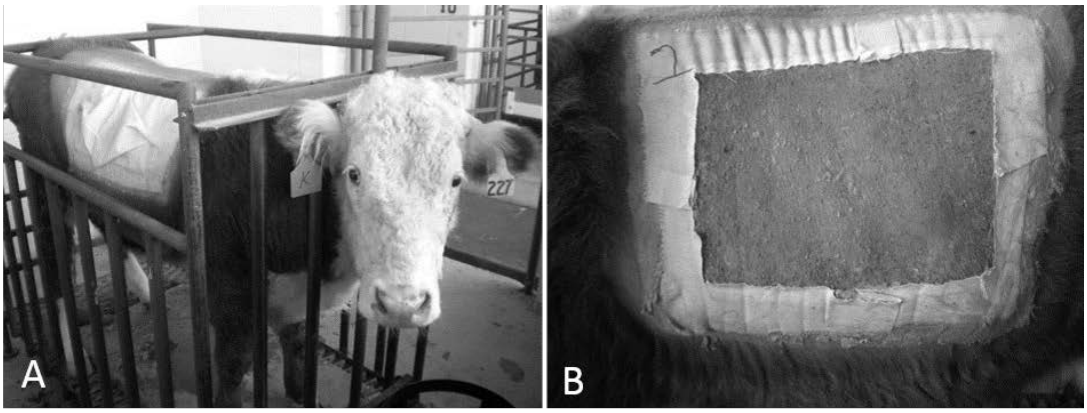

Figure 10: Pictures of example of the method for rearing of *R. microplus* by infesting a stanchioned calf. A) Stanchioned calf with patches used for *R. microplus* tick colony maintenance. B) Cut away of patch for access to the ticks at different developmental stages.

Freshly molted adults are termed unfed ticks as they have yet to have a blood meal in the adult life stage. Dissection of unfed female adults was done within 24 hours of ecdysis under cold Dulbecco's Phosphate Buffered Saline (PBS), pH 7.47. The undeveloped Gené's organs were

stored in RNAlater and kept on ice until stored at 4˚C overnight and then moved to -20˚C. Fully engorged female ticks were collected and allowed to oviposit. Upon ovipositing, Gené's organs were dissected under the same conditions as the unfed female adult ticks.

RNAlater was removed from the samples and Trizol was added. The samples were then homogenized by grinding with a pestle and incubated at room temperature for 5 minutes. Chloroform was added to cause phase separation of the homogenate and vortexed for 15 seconds. The samples were then incubated at room temperature for 3 minutes and then centrifuged at 12,000 x g for 15 minutes at 4˚C. The supernatant was transferred to a new 1.5 ml microcentrifuge tube where 100% isopropanol was added to precipitate the DNA. The samples were mixed well and incubated at room temperature for ten minutes. The DNA was then pelleted by centrifuging the samples at 12,000 x g for ten minutes at 4˚C. The supernatant was carefully decanted and 70% ethanol was added to wash the DNA. The samples were centrifuged again at 7,500 x g for 5 minutes at 4˚C. They were decanted again and allowed to air dry in order to remove any residual ethanol. The DNA pellets were resuspended in TE buffer. The concentration of the RNA in the samples was determined by NanoDrop 1000 (Thermo Scientific) and the purity was determined by a fragment analyzer.

Libraries were prepared using the TruSeq Stranded mRNA LT kit (Illumina) and barcode primers were attached to the inserts with a median insert size of 200 bp. The libraries of undeveloped and developed Gené's organ were pooled to be sequenced on an Illumina MiSeq. The sequencing was run to 600 cycles with the MiSeq Reagent kit v3 (Illumina) which gave an estimated 15x coverage.

The raw sequence reads were generated in a fastq format. The raw reads were uploaded to the Discovery Environment (DE) of the iPlant Collaborative[41]. The iPlant Collaborative is a cloud

based computing center for biologists not necessarily trained in bioinformatics. They have available in the Discovery Environment hundreds of applications that have been used in the literature for their appropriate function. The quality of the raw reads were assessed through the HTProcess_prepare_directories_and_fastqc application and individual files were made into a library that can be used for further analysis beginning with the fastqc program. FastQC (ver 0.10.1) was used to run a statistical assessment of the quality of the reads before and after processing[42].

With the assessment of the raw reads by FastQC, the raw reads were filtered and trimmed by using Trimmomatic[43]. FastQC determined that some of the reads still retained adapter sequences that were ligated to the reads during the library preparation. A text file was created that included all of the relevant adapter sequences. The program referenced our adapter file to trim them from the reads along with the first 9 basepairs from the leading end of the reads and 20 base pairs from the trailing end. These numbers were used from the FastQC results showing the quality at the beginning and end of the reads. A sliding window of 5 base pairs where the average phred score needed to be 20 or higher was used to better the per base phred score. After the processing of the reads, they were put through FastQC again. The quality of the reads was based on the Per base sequence quality, Per sequence quality score, Per base sequence content, Per base GC content, Per sequence GC content, Per base N content, Sequence length distribution, Overrepresented sequences and Kmer content.

We compared three *de novo* transcriptome assemblers: Trinity[44](r2013-08-14), Soapdenovo-Trans[45](v1.0), and Velvet-Oasis[46](v1.2.075). Trinity uses an automatic k-mer value of 25, so the same k-mer value was used when running the other programs to assemble the

Gené's organ transcriptome. The contig N50, number of contigs, number of assembled bases, average contig length, and largest contig were the metrics used to assess the different assemblies.

The Trinity assembly was then run through the Trinity utility TransDecoder[47]. It predicts open reading frames from the transcripts and possible protein-coding regions. TransDecoder also uses CD-HIT[48] to remove redundant transcripts by consolidating and clustering the transcripts.

The predicted amino acid sequences were used to search protein databases for homologous sequences. The National Center for Biotechnology Information (NCBI) BLASTp program was used with an E-value of $1e^{-20}$ with the maximum target sequence set at 1. The translated Gené's organ transcriptome was compared to the UniProt annotated protein database. The BLASTp results were then uploaded into Blast2GO[49] (3.0) for further annotation. The Blast results were transferred to the Blast2GO (3.0) program and mapped to Gene Ontology terms. Blast2GO (3.0) uses the program InterProScan[50] which searches for similarities of the amino acid sequences to many different databases. The different search programs used within InterProScan are BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, PartterScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther, Gene3D, Phobius, and Coils[50]. BlastProDom scans the families in the ProDom database. ProDom is a comprehensive set of protein domain families automatically generated from the UniProtKb/Swiss-Prot and UniProtKB/TrEMBL sequence databases using psi-blast. FPrintScan scans groups of motifs that are more potent together in their biological context in the PRINTS database. HMMPIR scans the Hidden Markov Models (HMMs) that are present in the PIR Protein Sequence Database (PSD) of functionally annotated protein sequences, PIR-PSD. HMMPfam scans the HMMs that are present in the PFAM Protein families database.

HMMSmart scans the HMMS that are present in the SMART domain families database. HMMTigr scans the HMMs that are present in the TIGRFAMs protein families database. Profiles based on weight matrices are more sensitive for the detection of divergent protein families which are PROSITE profiles. ProfileScan and PartterScan scan against PROSITE profiles. PartterScan is a newer version of the Profile Scan search software. HAMAP profiles are similar to PROSITE profiles but they are specific for bacterial archaeal and plastid-encoded protein families. The HAMAP profile database was scanned with HAMAP. Superfamily scans a library of profile HMMs that represent all proteins of known structure. SignalPHMM predicts the presence and location of signal peptide cleavage sites in amino acid sequences. TMHMM predicts transmembrane helices in proteins. HMMPanther defines protein families and subfamilies by their divergence of function. Gene3D scans a collection of CATH protein domain assignments for ENSEMBL genomes and Uniprot sequences. Phobius is another program for predicting transmembrane and signaling characteristics of the protein from the amino acid sequence. Coils predicts the coiled regions in proteins. The annotations derived from InterProScan are then mapped to Gene Ontology (GO) terms. The GO-Slim program was run to reduce the amount of GO terms per transcript which makes for a simpler and more manageable functional dataset[51].

The GO terms were mapped to enzyme codes that are used in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[52]. The enzyme codes were used to identify biological pathways.

The Gené's organ transcriptome was also compared to the *R. microplus* and *Ixodes scapularis* genome assemblies using the same parameters for search as above. The *R. microplus* genome assembly was used as a measure to help determine the quality of the transcriptome and

to identify any transcripts that were novel. We used the genome assembly CCG_RMi_1.0, which was from the Deutsch strain and published on NCBI in 2012 that was submitted by the Center for Comparative Genomics at Murdoch University. It contains 175,208 contigs with a contig N50 of 827 and a genome coverage of 3.41x. For the same reason we compared the transcriptome assembly to the *I. scapularis* genome as it is the only other tick genome assembly publicly available. The *I. scapularis* genome assembly JCVI_ISG_i3_1.0 was published in 2008 and has 369,492 scaffolds and 570,637 contigs with a contig N50 of 2,942 and a genome coverage of 6x. In addition to comparing the two genome assemblies, we added the *R. microplus* gut and larval transcriptomes that are available at the Cattle Tick Base – *Rhipicephalus microplus* website which is run by the Centre for Comparative Genomics at Murdoch University. The *R. microplus* gut and larval transcriptomes contain 11,333 and 6,082 contigs respectively. The comparison of the Gené's organ transcriptome to these will allow for better identification of novel transcripts and addcoverage to the *R. microplus* genome.

Results

High quality de novo assembly of *R. microplus* Gené's organ transcriptome. A total of 18,407,658 Illumina MiSeq reads from undeveloped and developed Gené's organ were generated. After trimming the adapters and running the reads through a quality filter, 17,057,546 (Table 1) reads were assembled with Trinity [44,47,53], Velvet-Oasis [54–56], and Soapdenovo-Trans [47,57–59]. The Trinity assembly resulted in 107,520 contigs with an N50 of 1,241 bases. The

Table 1: RNA-Seq reads before and after quality filtering and trimming of reads.

	Undeveloped Gené's Organ	Developed Gené's Organ	Total
Raw Reads	8,514,978	9,892,680	18,407,658
Reads After Quality Filtering and Trimming	7,837,138	9,220,408	17,057,546

largest contig length was 24,577 bases with an average of 707 bases. Velvet-Oasis assembled 63,224 contigs with an N50 of 402 bases. Soapdenovo-Trans assembled 209,575 contigs with an N50 of 415. The longest contig length in the Velvet-Oasis and Soapdenovo-Trans assemblies was 4,338 and 16,113 bases respectively (Table 2). The Trinity assembly was chosen as the better assembly because it had the higher contig N50, average contig length, max contig length, total assembled bases of the *de novo* assemblers. Trinity generated 68,398 putative unigenes (Table 3). A single gene can produce several different isoforms and the total Trinity transcripts include possible isoforms of several genes. Trinity condenses the isoforms into putative genes, or 'unigenes'. Transcripts that did not have any comparable isoform are called singletons. The Trinity assembly generated 31,717 singletons with a total assembly of 76,039,780 bases (Figure 3).

Table 2: Summary statistics of three different *de novo* transcriptome assemblies.

	Trinity	Velvet-Oases	Soapdenovo-Trans
contigs	107,520	63,224	209,575
Total assembled bases	76,039,780	24,409,089	59,018,925
contig N50	1,241	402	415
minimum contig length	201	200	100
max contig length	24,577	4,338	16,113
average contig length	707	386	282

Table 3: Statistics of the *R. microplus* Gené's organ transcriptome Trinity assembly.

Statistics of *Rhipicephalus microplus* Gené's organ transcriptome assembly	
Total trinity 'genes'	68,398
Singletons	31,717
Total trinity transcripts	107,520
Percent GC	47
Longest contig (b)	24,577
Shortest contig (b)	201
Transcript contig N10	4,561
Transcript contig N20	3,163
Transcript contig N30	2,323
Transcript contig N40	1,714
Transcript contig N50	1,241
Median contig length	353
Average contig	707
Total assembled bases	76,039,780

Table 4: Statistics of the completeness of two Trinity de novo transcriptome assemblies of the Gené's organ based on 248 CEGs.

		# Proteins	%Completeness	#Total	Average	%Orthologs
Trinity assembly of raw reads	Complete	76	**30.65**	157	2.07	61.84
	Group 1	15	22.73	38	2.53	80.00
	Group 2	7	12.50	12	1.71	57.14
	Group 3	25	40.98	57	2.28	64.00
	Group 4	29	44.62	50	1.72	51.72
	Partial	111	**44.76**	228	2.05	58.56
	Group 1	19	28.79	52	2.74	84.21
	Group 2	13	23.21	30	2.31	69.23
	Group 3	31	50.82	67	2.16	61.29
	Group 4	48	73.85	79	1.65	43.75
Trinity assembly of filtered and trimmed reads	Complete	199	**80.24**	493	2.48	72.36
	Group 1	42	63.64	117	2.79	78.57
	Group 2	46	82.14	105	2.28	67.39
	Group 3	50	81.97	124	2.48	74.00
	Group 4	61	93.85	147	2.41	70.49
	Partial	226	**91.13**	598	2.65	75.66
	Group 1	51	77.27	140	2.75	72.55
	Group 2	54	96.43	138	2.56	74.07
	Group 3	58	95.08	152	2.62	81.03

	Group 4	63	96.92	168	2.67	74.60

Another quality assessment of the Gené's organ transcriptome assembly was done through a computational method of using a list of core eukaryotic genes (CEGMA) that occur in a wide variety of eukaryotes. By comparing and identifying protein sequences to the CEGMA core genes, the assembly can be assessed as to its completeness [60]. CEGMA was run on a Trinity assembly of the raw reads and an assembly after the quality filtering and trimming. The Trinity assembly of the raw reads had 76 complete sequences of the 248 ultraconserved core eukaryotic genes and 111 partial sequences with a transcriptome completeness of coverage at 44.76%. The Trinity assembly of the quality filtered and trimmed reads showed a 91.13% transcriptome completeness having 226 of the 248 ultraconserved core eukaryotic genes. The percentage of orthologs for the identified CEGs increased with the Trinity assembly of the quality filtered and trimmed reads from 58.56% to 75.66% (Table 4).

Functional Annotation of Gené's organ transcriptome. The Gené's organ transcriptome assembly was then clustered, condensed, and translated by using the software TransDecoder which resulted in 27,388 amino acid sequences. The amino acid sequences were compared to the Uniprot database which yielded 18,408 hits with an e-value < 1E-20 (Figure 18). The average amino acid sequence length was 315 (Figure 17). The range of

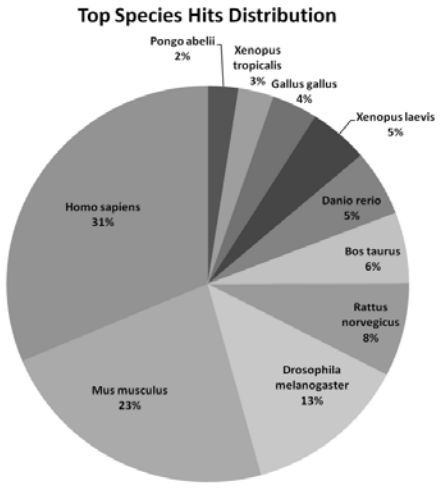

Figure 11: Protein BLAST top hit species for the Gene's organ transcriptome. The transcriptome Gene's organ proteins were compared to the Uniprot database in search for homologous proteins.

alignment similarity was from 27% to 100% (Figure 16). The species with the most hits were *Homo sapiens* (4970), *Mus musculus* (3630), *Drosophila melanogaster* (2082), and *Rattus norvegicus* (1209) (Figure 11).

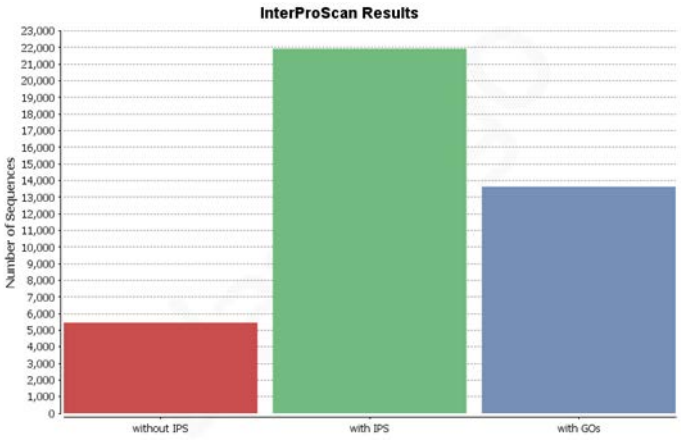

Figure 12: InterProScan results showing the increase in the number of sequences with InterProScan annotations as compared to GO ontology terms and sequences without any annotations.

The Blast results were transferred to the Blast2GO program. Blast2GO contains the program InterProScan which searches for similarities of the amino acid sequences to many different databases. The InterProScan increased the number of descriptions to the amino acid sequences from 13,632 to 21,933 a 60.89% increase (Figure 12).

KEGG and Gene ontology pathway analysis. The Gené's organ transcriptome protein sequences were run through Blast2GO software where the results of the InterProScan and the BLASTp search were used to assign GO annotation. Gene ontology uses the annotations of the homologous sequences assigned to the transcriptome sequences by the BLAST search and the InterProScan results to assign GO terms. The first level of GO annotation is divided into three groups:

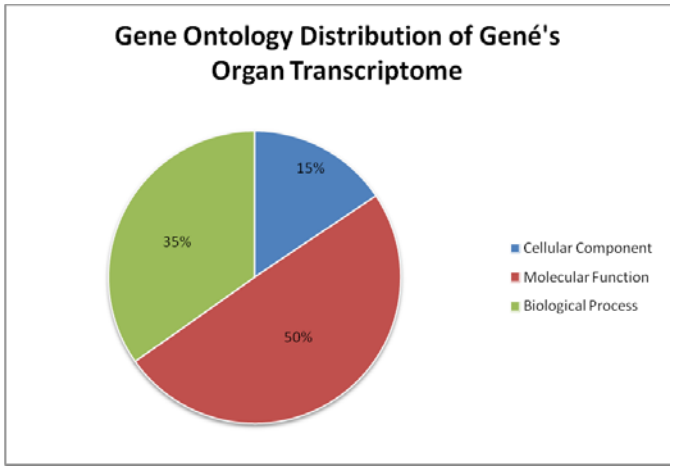

Figure 13: A graphical representation of the distribution of level one GO annotations of the Gene's organ transcriptome.

biological process, cellular component, and molecular function. Since a protein could have multiple roles, it can also have more than one GO annotation. The Gené's organ transcriptome had 61% of the protein sequences with a molecular function, 24% with a biological process, and 15% had a cellular component (Figure 13).

The translated transcriptome is divided into subcategories defining the putative role of the protein. The GO terms mapped to the transcript can be several layers. In the Gené's organ transcriptome, the most GO terms for a single transcript was 11 (Figure 14).

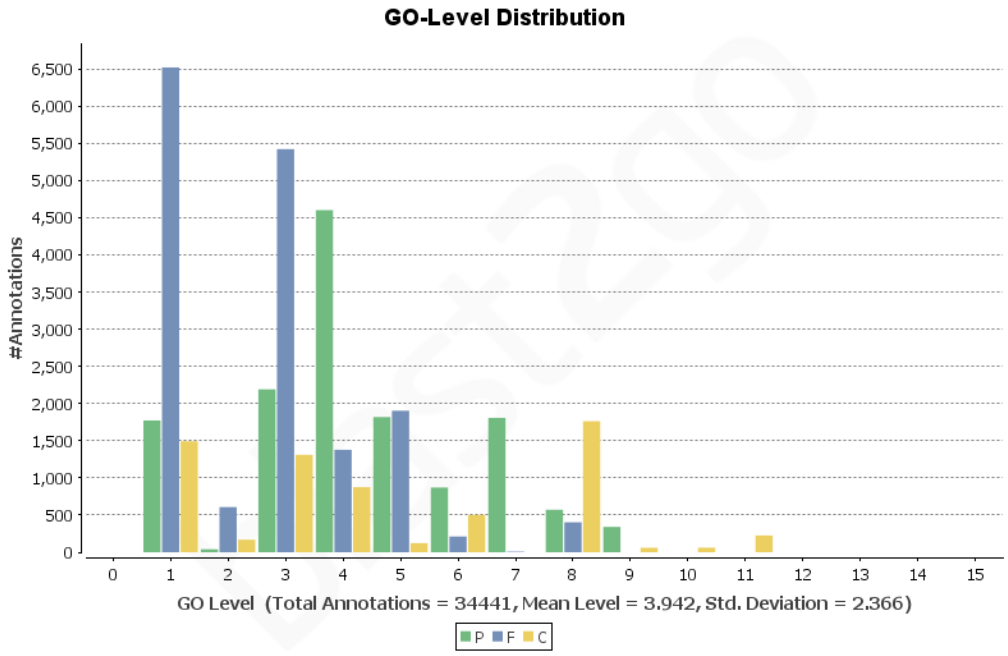

Figure 14: The GO terms mapped to the Gené's organ transcriptome with some transcripts receiving multiple GO terms because of their homology to well-studied proteins.

Within the category of molecular function, 4,510 transcripts coded for a binding protein followed by having a catalytic activity function with 3,630 transcripts (Figure 19). The binding category can be divided into multiple subcategories: protein binding, DNA binding, metal ion binding, nucleotide binding and RNA binding. The most prevalent molecular function in the Gené's organ transcriptome was protein binding (27.75%) (Figure 15). Cellular component had

seven main categories: cell, organelle, macromolecular complex, membrane-closed lumen, extracellular region, membrane, and extracellular matrix (Figure 20). The transcripts were categorized into subcategories with an integral component of membrane (9.46%) as the most prominent followed by a nuclear (7.75%) and membrane (6.04%) function. The GO terms related to biological processes correlated to metabolic processes, cellular process, and single-organism process as the top three with 5,583, 5,436, and 4,110 annotations respectively (Figure 21). The subcategories related to biological process were cell redox homeostasis (0.37%), intracellular signal transduction (0.33%), and protein folding (0.29%) (Figure 15).

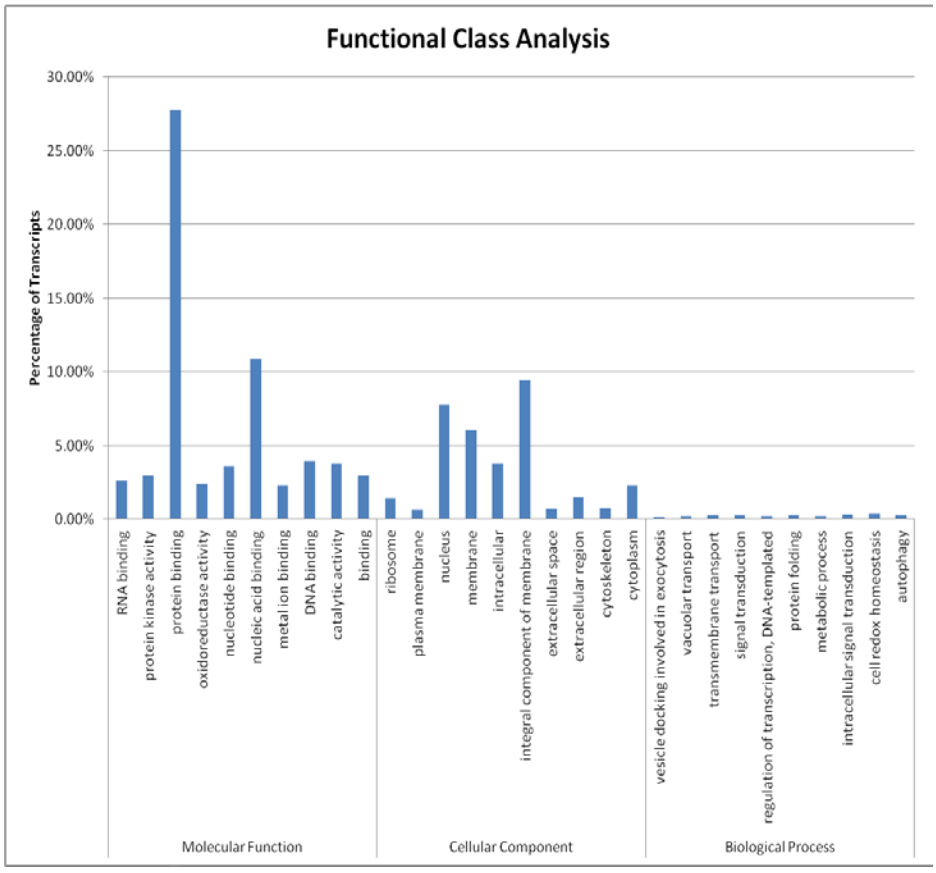

Figure 15: A distribution of the annotations for the Gené's organ breaking the level 2 GO terms into their subcategories. The majority of the GO terms were related to Molecular Function.

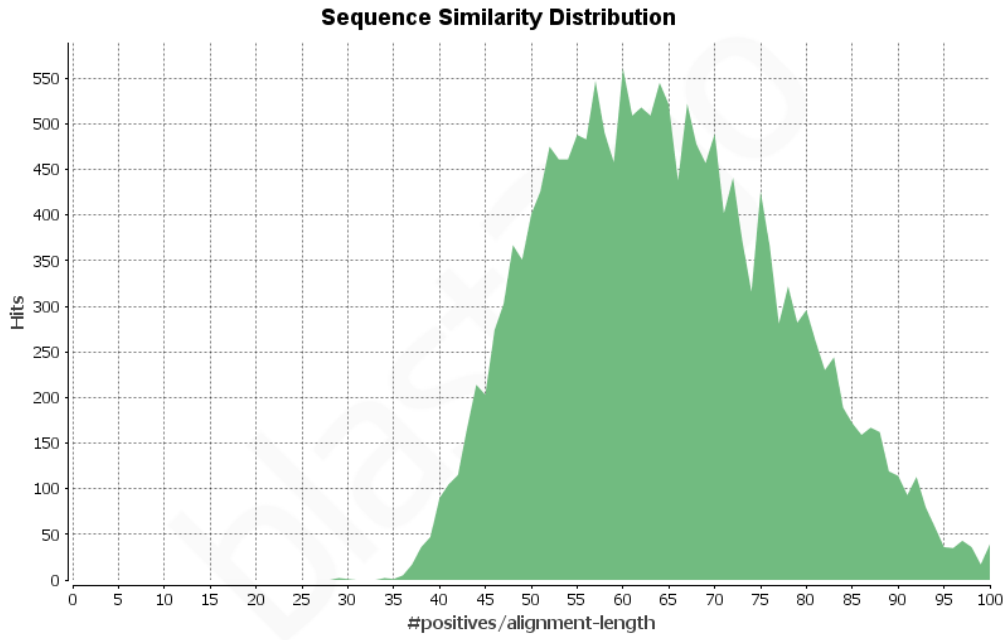

Figure 16: The statistical representation of the number of transcripts that aligned with a homologous protein sequence in the Uniprot database and the percentage of homology.

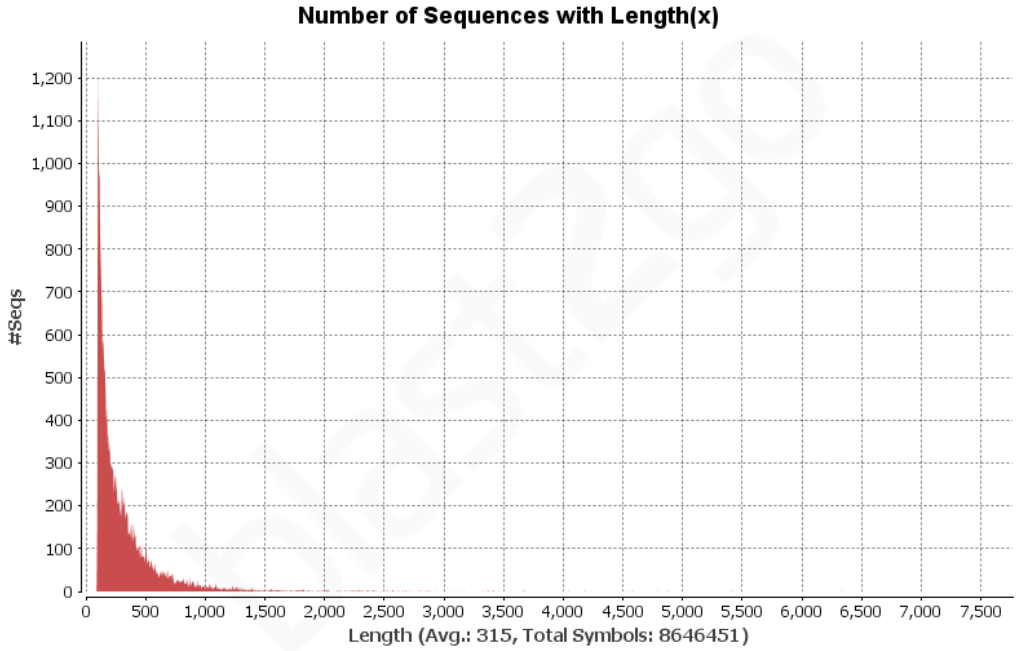

Figure 17: The number of sequences in the Gené's organ transcriptome and their length.

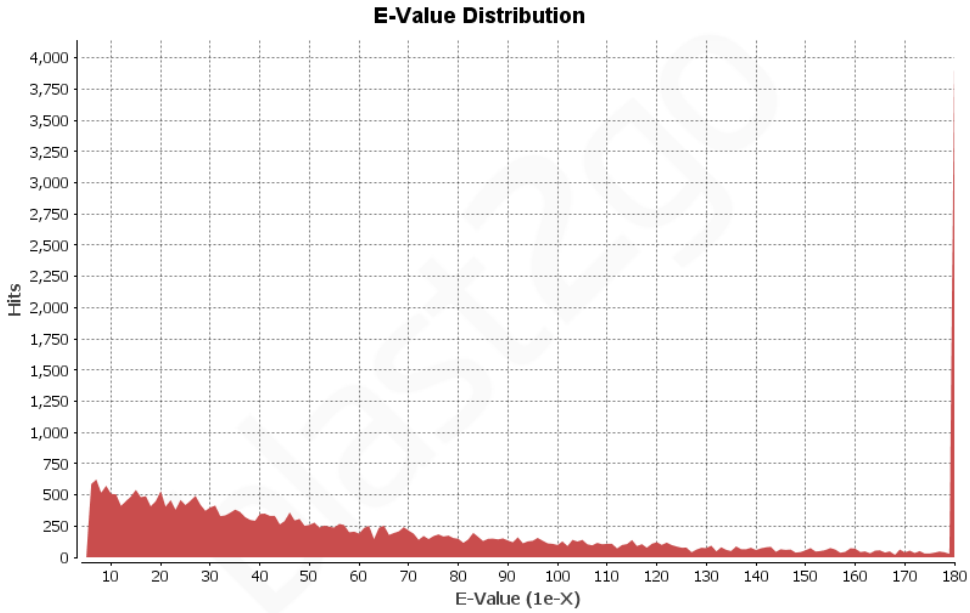

Figure 18: The range of e-values in the Blastp search of the Uniprot database to the Gené's organ transcriptome.

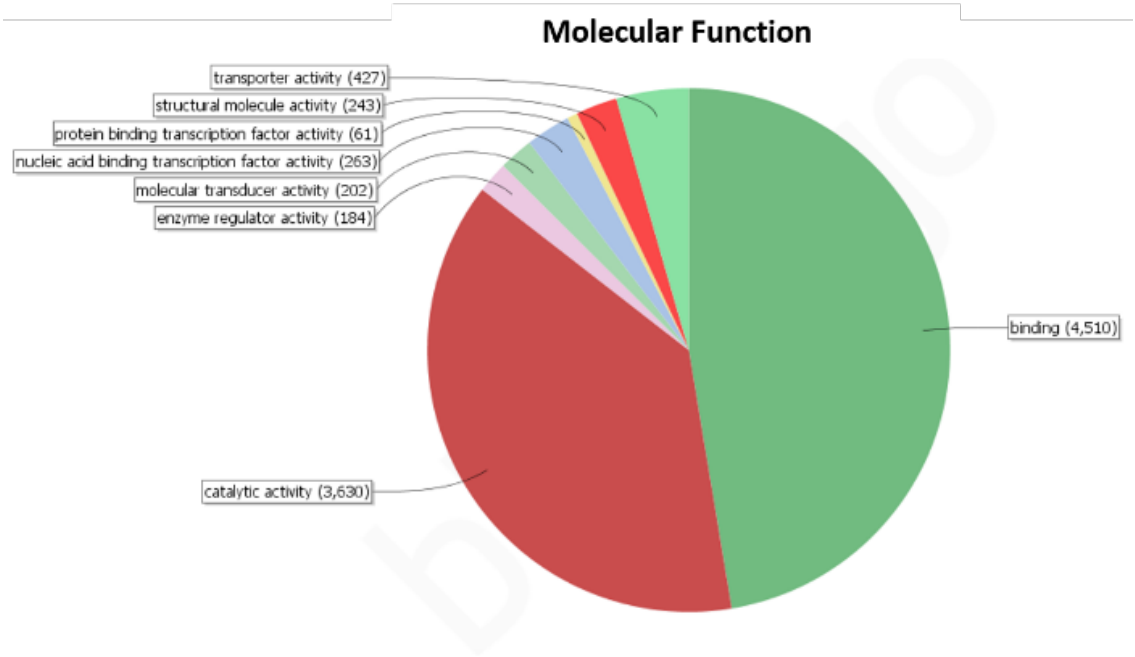

Figure 19: A distribution of the level 2 GO terms related to the Molecular Function of the Gené's organ transcriptome.

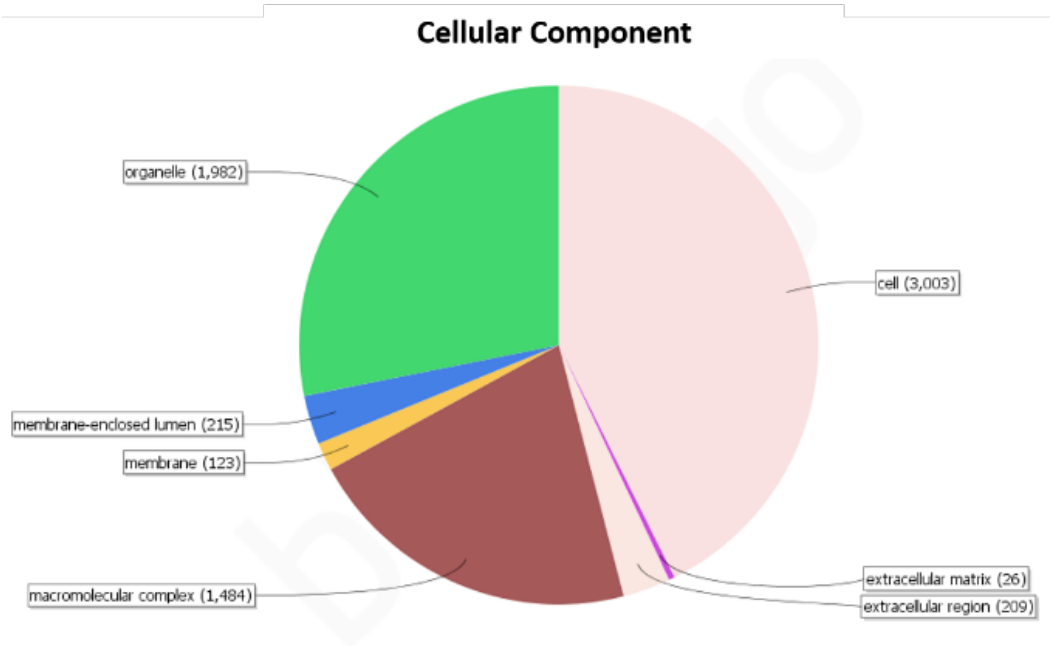

Figure 20: A distribution of the level 2 GO terms related to the Cellular Component of the Gené's organ transcriptome.

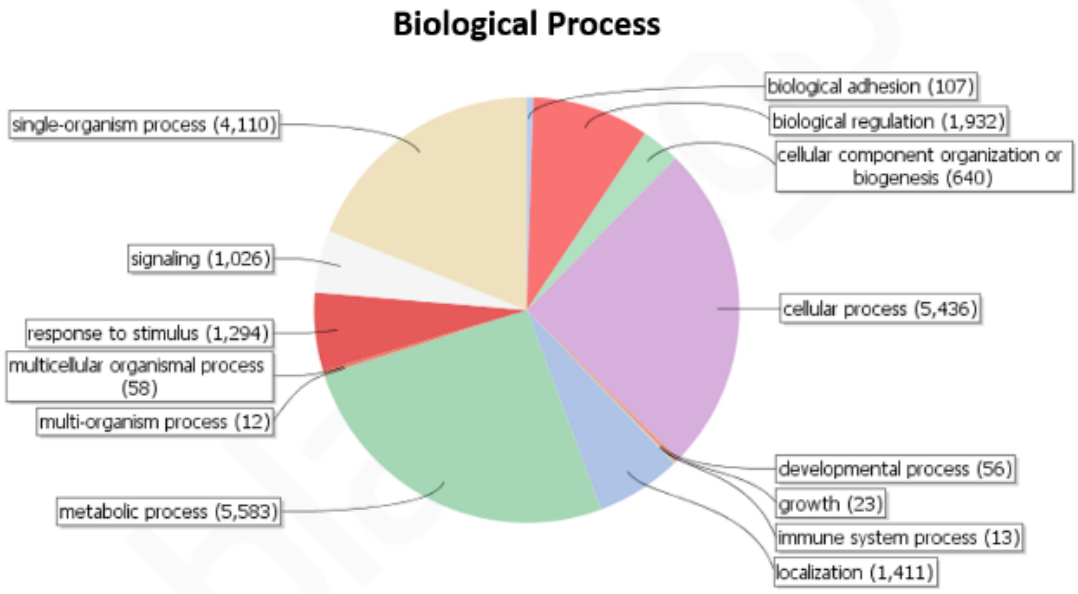

Figure 21: A distribution of the level 2 GO terms related to Biological Processes for the Gené's organ transcriptome.

The mapping of the translated transcriptome sequences to Kyoto Encyclopedia of Genes and Genomes (KEGG) enzyme codes allowed for the identification of orthologs in *R. microplus* in the KEGG pathway database. The biological pathway for purine metabolism had 594

sequences and 33 different enzymes attributed to the pathway (Table 5). There were a total of 113 different KEGG pathways that had at least one enzyme and sequence (Supplemental Data). The KEGG pathway for the biosynthesis of antibiotics had the most diverse amount of enzymes with 70 (Table 5). Figure 22 shows an example of a KEGG biological pathway map with codes for the enzymes in the pathway. In the oxidative phosphorylation pathway 6 of the 11 enzymes were identified in the Gené's organ transcriptome.

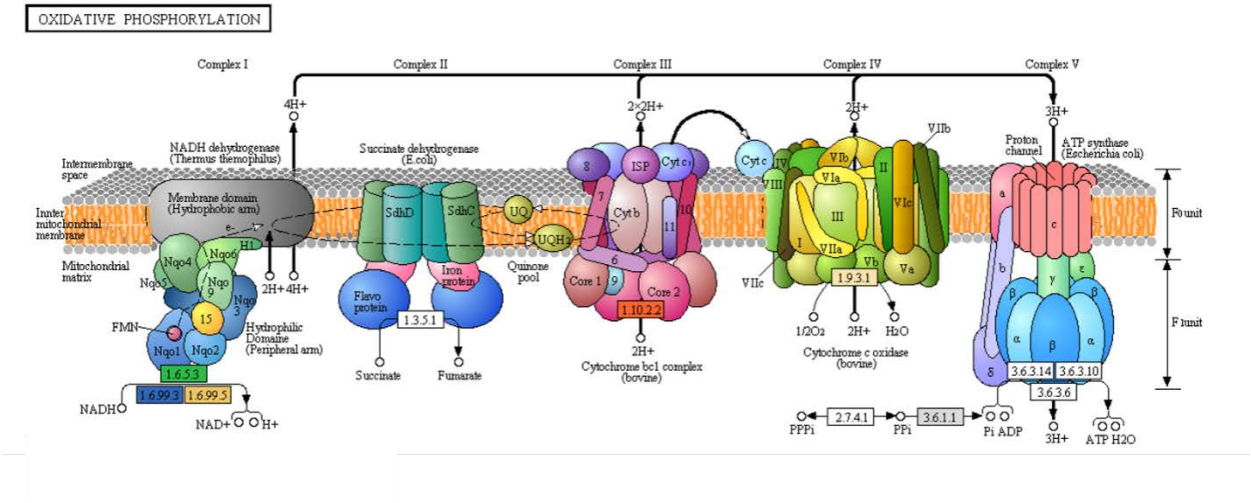

Figure 22: An example of a KEGG biological pathway with the enzyme codes in boxes. The colored boxes are the enzymes identified in the Gené's organ transcriptome.

The Gené's organ transcriptome was compared to two tick genome assemblies and two transcriptomes with BLASTn[61]. Of the 175,208 contigs that are contained in the *R. microplus* genome assembly, 47,054 aligned with the *R. microplus* Gené's organ transcriptome assembly. 67% of those contigs only aligned with the *R. microplus* genome and the Gené's organ transcriptome. There were 963 transcripts that were homologous to all four of the datasets. The *R. microplus* Gené's organ transcriptome was compared to the 570,637 contigs contained in the *I. scapularis* genome assembly and identified 1,513 contigs that were not in the other datasets (Figure 22). Out of those 1,513 contigs unique to the *I. scapularis* genome assembly and the *R.*

microplus transcriptome assembly, 215 were annotated with descriptions from the Uniprot database mapping the transcripts to GO terms. An additional 449 transcripts received GO terms through InterProScan results and subsequent mapping.

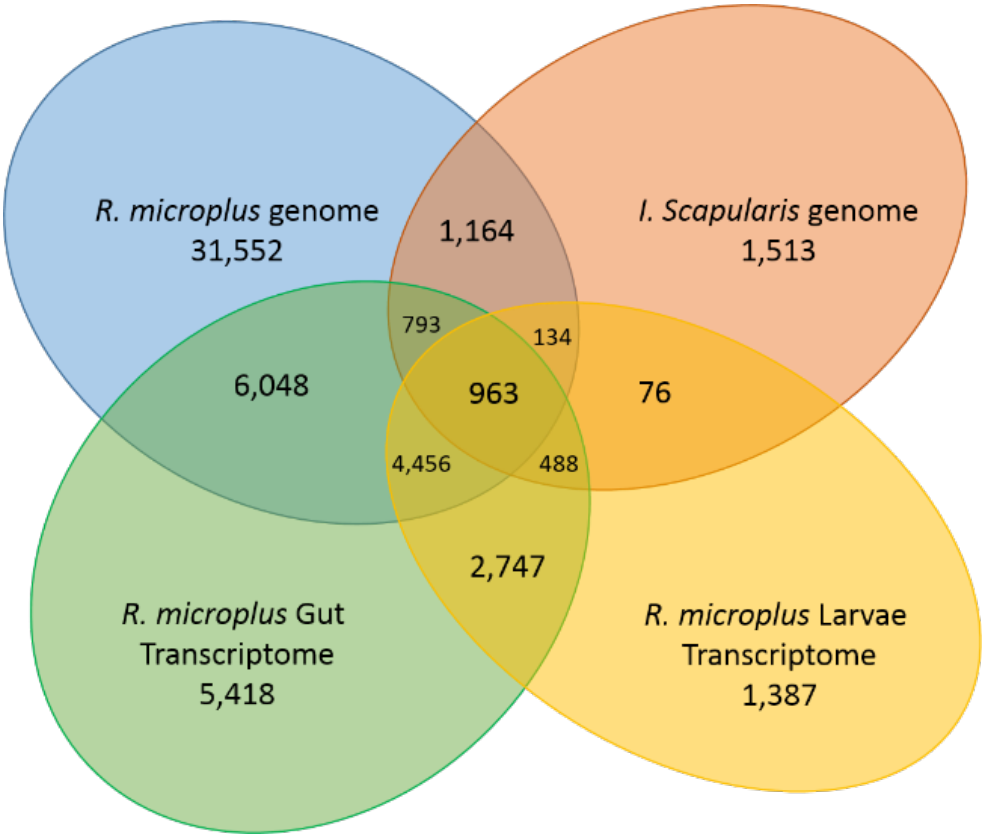

Figure 23: Sequence comparisons with Blastn between contigs from *R. microplus* genome (175,208 total contigs), *I. scapularis* (570,637), *R. microplus* gut transcriptome (11,333), and *R. microplus* larval transcriptome (6,082).

Table 5: Top KEGG Pathways with the number of protein sequences identified in the Gené's organ transcriptome and the different number of enzymes in each pathway.

Num	KEGG Pathways	#Seqs in Pathway	#Enzymes
1	Purine metabolism	594	33
2	Thiamine metabolism	454	1
3	Biosynthesis of antibiotics	176	70
4	Aminobenzoate degradation	151	4
5	Pyrimidine metabolism	104	16
6	T cell receptor signaling pathway	82	2
7	Phosphatidylinositol signaling system	74	8
8	Aminoacyl-tRNA biosynthesis	69	18
9	Drug metabolism - other enzymes	58	7
10	Phenylalanine metabolism	54	5
11	Other glycan degradation	53	5
12	Inositol phosphate metabolism	53	9
13	Citrate cycle (TCA cycle)	52	15
14	Phenylpropanoid biosynthesis	50	1
15	Carbon fixation pathways in prokaryotes	48	14
16	Amino sugar and nucleotide sugar metabolism	48	16
17	Glutathione metabolism	43	11
18	Riboflavin metabolism	38	2
19	Methane metabolism	33	13
20	Pyruvate metabolism	32	15
21	Tryptophan metabolism	32	8
22	Pentose phosphate pathway	32	12
23	Lysine degradation	31	5
24	Drug metabolism - cytochrome P450	31	4
25	Glycolysis / Gluconeogenesis	29	16
26	Terpenoid backbone biosynthesis	27	8
27	Glycerophospholipid metabolism	27	10
28	Propanoate metabolism	26	8
29	Various types of N-glycan biosynthesis	26	5
30	Valine, leucine and isoleucine degradation	25	9

Discussion

Rhipicephalus (Boophilus) microplus is one of the most important vectors for disease in the agriculture industry[62]. The genome of the tick has been problematic in sequencing because of its size and repetitiveness[63]. Like transcriptome assemblers, genome assemblers use the same principle in fragmenting large segments of cDNA libraries in order to sequence them. They are then parsed back together through de Brujin graphing[64]. A genome with multiple repeats will make this technique difficult to do with next-generation sequencing technologies. A recent advancement in sequencing technology utilizes longer reads and forgoes fragmenting the genomic or transcriptomic libraries[64–66]. The long read sequencing will be able to overcome the repeats in the sequence and give a more accurate genome assembly. The technique is still costly because of the novelty of the technology. Another method to overcome the problems associated with sequencing the *R. (B.) microplus* genome is through sequencing the transcriptomes of the different organs and life stages of the tick[67–70].

The Gené's organ has a major role in the preservation of tick eggs and has been mentioned in the literature as a possible target for tick control. Our transcriptome assembly is the first for a Gené's organ in any tick species. The purine metabolic pathway was the main biological pathway identified as active in the Gené's organ transcriptome. This pathway is responsible for generating purine and pyrimadine nucleotides. It is also responsible for synthesizing ATP, GTP, CTP, and UTP[71]. Adenosine triphosphate is the main energy currency in the cell. Guanadine triphosphate is used in protein synthesis and CTP is used in lipid metabolism[72]. With the Gené's organ actively producing a protective wax in its developed state, the prevalence of the purine metabolism pathway would be expected. It's interesting that the next most prevalent pathway is the thiamine metabolism pathway with a high number of

sequences devoted to it and yet only one enzyme is identified. This suggests that although the enzyme is connected to the thiamine metabolism pathway, its role could be for another biological process.

Through this research, we were able to add 664 contigs with annotations to the *R. microplus* genome. There are still 849 transcripts that were unique to only the Gené's organ transcriptome and *I. scapularis* genome. Comparing these transcripts to other databases such as the NCBI's non-redundant protein database and Uniprot-TREMBL could produce insights into their role and function.

CHAPTER IV

DIFFERENTIAL GENE EXPRESSION ANALYSIS OF THE GENÉ'S ORGAN TRANSCRIPTOME

Methods

Analysis of the gene expression from the Gené's organ transcriptome was done by mapping the raw reads back onto the transcriptome assembly. The aligned reads were then counted through the program RSEM [73]. RNA-Seq by expectation maximization (RSEM) is unique to other differential expression software packages in that it does not require a reference genome [73]. It utilizes the assembled transcripts from a *de novo* assembly in order to quantify gene expression. The results of RSEM are analyzed by the software R package edgeR [74]. The edgeR package was used for its ability to account for biological variability when there are only one or two replicate samples. In our case, we only had one sample for each developmental stage of the Gené's organ. The edgeR package analyzed the RSEM data and expressed the differentially expressed genes by a heat map and a volcano map. The transcripts were grouped together through hierarchical clustering allowing for identification of possible transcripts that are biologically linked.

The parameters for edgeR were adjusted to a P-value of 1e-6 and a fold change of 2^{15}. Transcripts were selected from each cluster with seven having a high expression level in thedeveloped Gené's organ and four with repressed expression levels (Table 5). The transcripts with the higher expression levels were chosen because of their roles in lipid metabolism and

transport pathways which may be viable targets for disrupting the function of the Gené's organ. We used actin as our reference gene. Primers were designed for the selected transcripts by Primer3Plus [75].

Table 6: Primers used in the relative quantification of *R. microplus* Gené's organ mRNAs by real-time PCR.

Name	DNA primer/probe sequence (5' -> 3')	Putative Function of Transcript	Expression in Developed Gené's Organ
comp23240_c0_seq14_F	GATGTCGCTCTTCGTGTTCA	protein binding	low
comp23240_c0_seq14_R	TGCCGAGATTGTGTCAGAAG		
comp23811_c0_seq4_F	TGTTTGCATTGACCAACGAT	unknown	low
comp23811_c0_seq4_R	TTTGCTTGTCAGCATTCCAG		
comp22624_c0_seq1_F	GGTGGTAACGGCAAGAAGAA	unknown	low
comp22624_c0_seq1_R	TGAACAGAAACGTCGACTGG		
comp23315_c0_seq1_3_F	GGCTTTGGAGGTGGTACAGA	unknown	low
comp23315_c0_seq1_3_R	AGCTCGTAGCTCCAGCGTAG		
comp20572_c0_seq1_F	GGTGCTGGTGGTGTAGAGGT	lipid transport	high
comp20572_c0_seq1_R	AGCTTCGACAAGACCCTCAA		
comp23006_c0_seq1_F	GCTGAGGTTACTGCCAGGAG	transmembrane transport	high
comp23006_c0_seq1_R	AATTCCAGTTCAGGCAATCG		
comp21272_c0_seq3_F	CTCGATGACGAACTCTGTCG	elongation of very long fatty acid chains	high
comp21272_c0_seq3_R	CGCAAACAAAGGAGTCCCTA		
comp15748_c0_seq1_F	ATATTCCAACGCGTTTCTGG	cytochrome P450	high
comp15748_c0_seq1_R	TCACATCACGTTCCACGTTT		
comp20991_c0_seq4_F	CATGGTGTCGATGAAGAACG	lipid metabolic process	high
comp20991_c0_seq4_R	CCACCTGGCTGGTAAACAGT		
comp19087_c0_seq1_F	ATCGCGAGCCTTTCAAGTTA	fatty acid elongation	high
comp19087_c0_seq1_R	CCACGAGTGCAAGTTTACGA		
comp23430_c0_seq1_F	GAAATGTGAAGCACGAAGCA	unknown	high
comp23430_c0_seq1_R	CAATGGTGGCAATGTCTACG		
actin_R [76]	TGATCTGCGTCATCTTCTCG	Reference gene	
actin_F [76]	CACGGTATCGTCACCAACTG		

Real-time PCR was used to validate the differential expression levels of the RNA-seq analysis of the *R. microplus* Gené's organ transcriptome. All real-time PCR reactions were done in triplicate. Standard curves were determined for each primer set with a DNA quantity rangingfrom 1.56 x 10^{-2} ng to 1 ng. The equation of the standard curve for each primer set was

used to determine the efficiency of the PCR reaction with those primers. High resolution melt curve analysis was conducted on each primer set to determine the specificity of the primers. Relative quantification was determined through Pfaffl's method[77,78] taking into account differences in PCR efficiency.

Results

The RSEM software mapped the reads to the transcriptome assembly in order to identify the expression levels of the transcripts in the samples. The results of RSEM are then taken by the R package, edgeR, where the counts of the reads are statistically analyzed into the categories logFC, logCPM, p-value, FDR, and -1*log10(FDR). The category logFC is the logarithmic value of the fold-change of the expression between the two developmental stages of the Gené's organ. The logCPM is the logarithmic value of counts-per-million of the read. The p-value is a statistical measure of how likely a result occurred simply by chance. Because the sampled data is tested multiple times, a false discovery rate (FDR) is used to adjust the p-value in order to reduce the number of false positives. There were 2,529 transcripts in the undeveloped Gené's organ sample that had at least a logFC of 2 or greater. The developed Gené's organ sample had 4,282 transcripts with at least a logFC of 2 or greater (Figure 25).

Two main clusters of transcripts can be seen in a heatmap generated by edgeR (Figure 26) which shows the differential expression levels of transcripts in the Gené's organ at the two developmental stages sampled. One group has a high expression level in one developmental stage and the expression is repressed in the other stage and vice versa for the other group. The transcripts were grouped into 10 clusters by similar expression levels under the assumption that genes under the same regulation would have similar expression levels.

All primer sets had clean melt curves where there was a single peak except for transcripts 21272 and 23430 (Figure 27). The PCR efficiency for these two primer sets was also lower at 55.73% and 36.32% respectively. The amplification efficiencies of the primer sets ranged from 36.32% to 16,966% (Figure 24). Using the ΔΔCt method to find a qualitative expression level ratio between the two developmental stages found that the primer set for transcript 23811 was the only one not to be statistically significant (Figure 28). Pfaffl's method was used to verify the results of the ΔΔCt method.

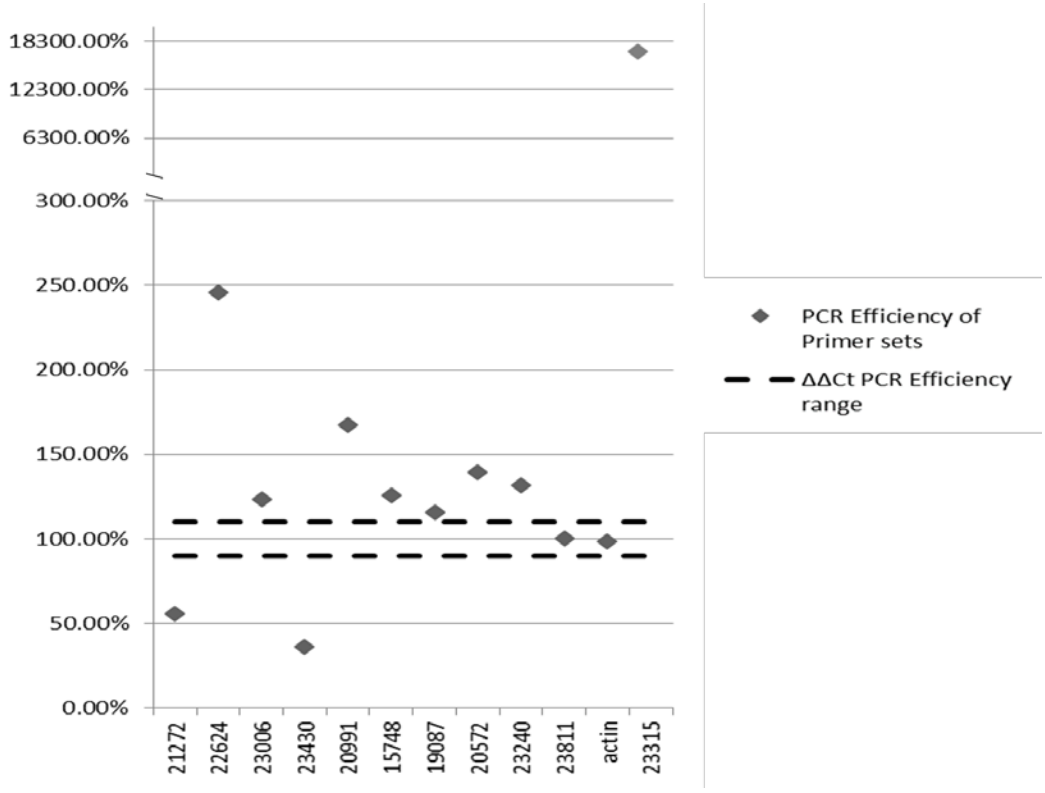

Figure 24: PCR efficiencies of the primers sets are represented with the range needed to use the ΔΔCt method for qualitative gene expression measurements. The ΔΔCt method makes the assumption that the PCR efficiencies of the primer sets are between 90-110%.

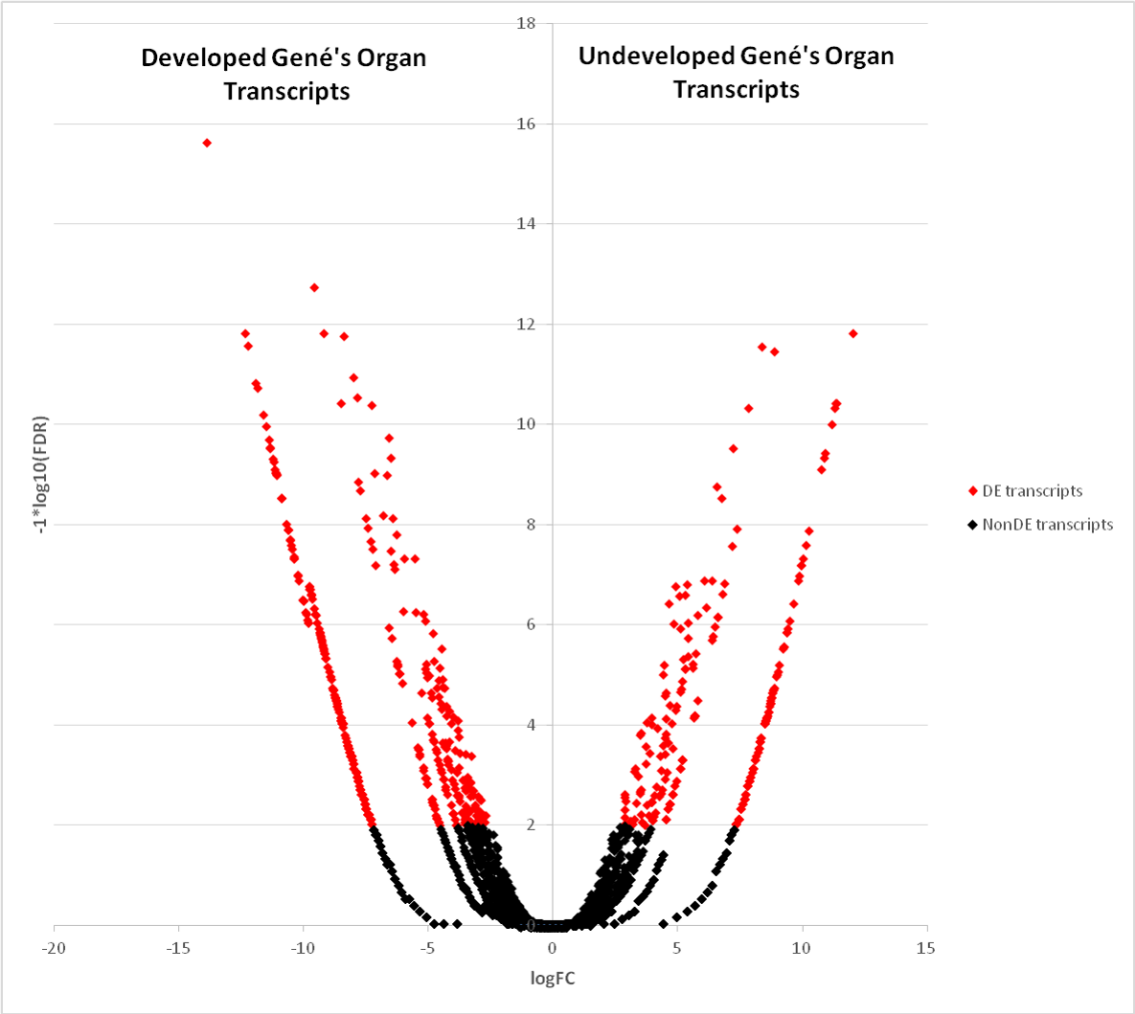

Figure 25: Volcano plot of the Gené's organ transcriptome assembly. The black transcripts have less than two fold change.

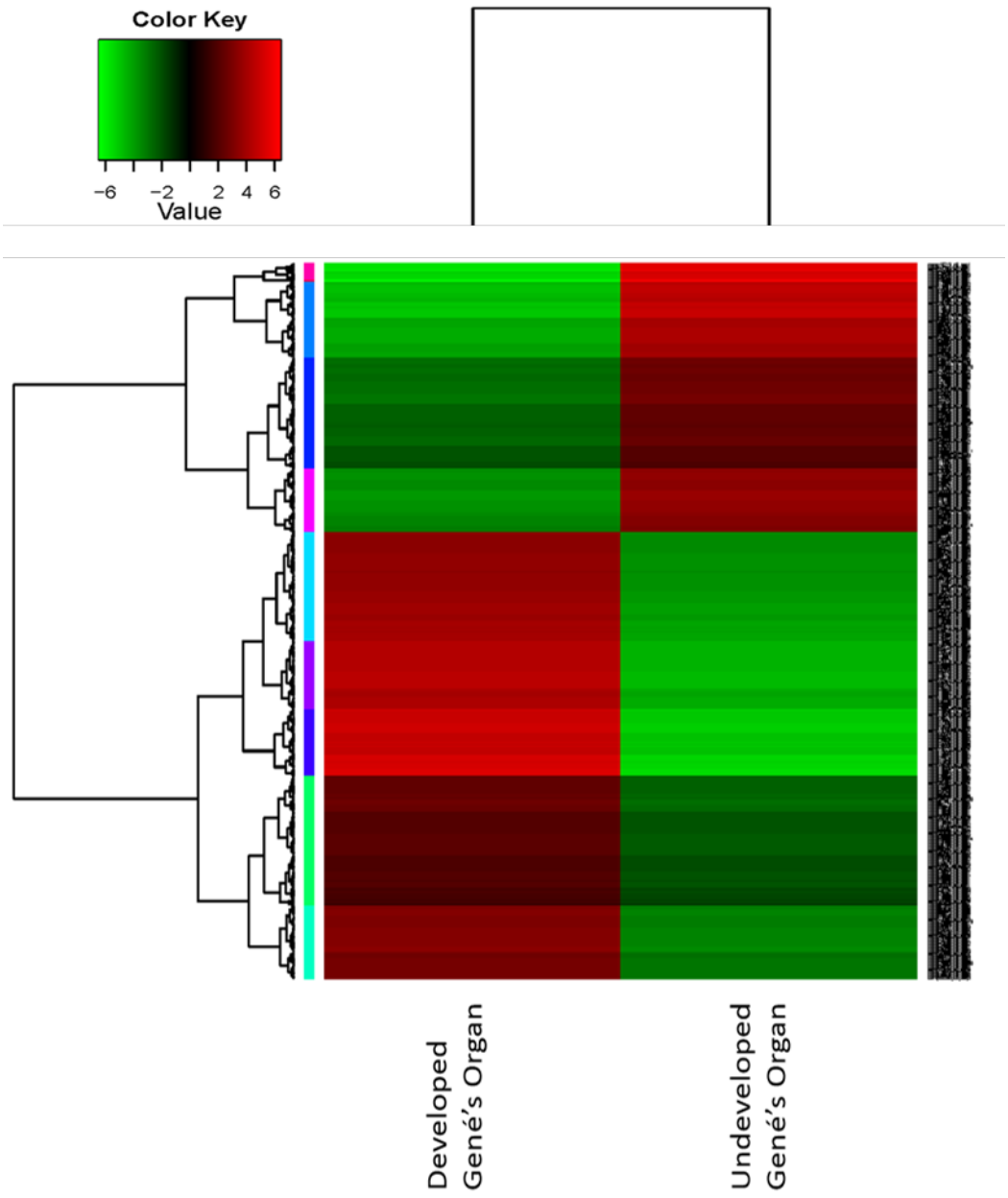

Figure 26: Heatmap of 10,000 differentially expressed transcripts of the *R. microplus* Gené's organ comparing two developmental stages of the organ.

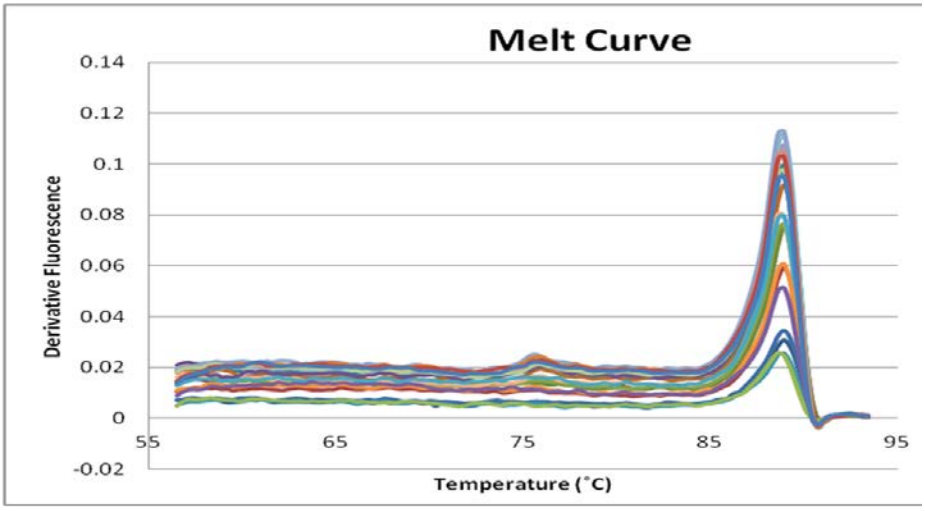

Figure 27: Example of a melt curve graph of the actin gene. The single peak denotes the specificity of the primers amplifying a single product.

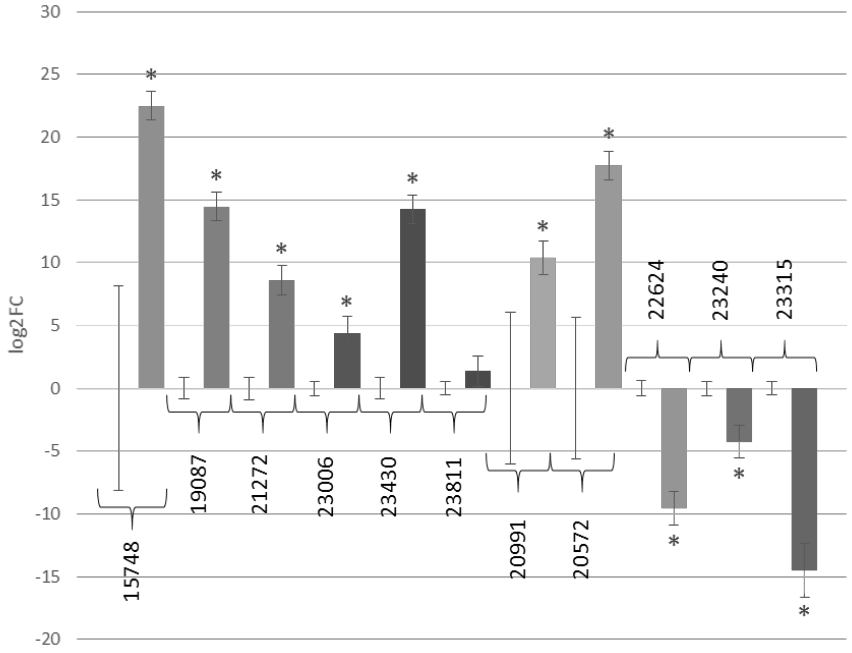

Figure 28: Absolute quantification of the transcripts chosen for verification of the RNA-seq analysis through real-time PCR using the ΔΔCt method. All primer sets were normalized to the expression levels of the undeveloped Gené's organ. The first bar is not present because of the normalization of the expression levels but the variance of the replicates is shown in the error bars showing the standard deviations of the samples. The * denotes if the difference in expression levels of the transcript between the two developmental stages of the Gené's organ was statistically significant (P < 0.05).

CHAPTER V

CONCLUSION

We assembled the transcript sequences for two developmental stages of the Gené's organ and identified new transcripts to add to the *Rhipicephalus microplus* genome project. These transcripts will help increase the coverage of the *R. microplus* genome assembly. The majority of these transcripts had a molecular function for protein binding. The two main biological process GO term annotations were "regulation of transcription, DNA-templated" and "transmembrane transport." The two main GO term annotations for the cellular component category were "integral component of membrane" and "chromatin." There are 849 transcripts that did not receive any annotation but still aligned to the *Ixodes scapularis* genome assembly. Annotating these transcripts by using multiple protein databases will further increase the quality of the genome assembly of *R. microplus* and *I. scapularis.*

In this study, we did a *de novo* transcriptome assembly for the Gené's organ. We used this method because of the low coverage of the *R. microplus* genome assembly at 3.41x. There are multiple alternative methods that could have been used for our pipeline. One would be to use Soapdenovo-trans or Velvet-Oasis as *de novo* assemblers and use multiple k-mers to build a meta-transcriptome. A Trinity assembly would be incorporated into the meta-transcriptome so that different isomers can be included. The program Cap3 could then be used to remove the redundant transcripts allowing for the teasing out of long and short length transcripts[79]. Another

method would be to use a reference genome for the alignment of the reads. Comparing our *de novo* Gene's organ transcriptome assembly to one based on the *R. microplus* genome as a reference is something that could be done. The reads that didn't align to the genome could then be assembled *de novo*[53]. The Gene's organ transcriptome assembly that we present here had 91% coverage. By using a reference genome for alignment, it may help to increase the coverage by 9%.

The organ of Gené is unique to hard ticks[28,35]. This organ has not been studied as in depth as the other organs of hard or soft ticks. It is only recently that the wax produced by the Gené's organ is beginning to receive some attention identifying its antimicrobial properties[23,27,29,30,80]. The waxes produced by different species of ticks correspond to the environment that they inhabit and are inherently different[29,30]. Our transcriptome assembly will help enhance our understanding of these differences with further study.

Further study can be done with our research as to identify the biological pathways that were used in the different developmental stages of the Gené's organ. Comparison analysis of these pathways to wax producing glands in other insects can be used to better understand the biology of the Gené's organ and possibly elucidate methods that could be used to disrupt its function as a novel control method against ticks.

The results of our research may have many downstream applications that open some exciting avenues of study. Because of the severe infestation problems in Brazil, plant extracts are tested for possible acaricidal properties. Ribeiro found that some of these extracts have a detrimental effect on the Gené's organ[81,82]. Booth also tested the plant based chemical, precocene, and showed that it damaged the functionality of the Gené's organ in engorged female

R. microplus[83]. With our transcriptome assembly of the Gené's organ, we can further investigate how precocene and other plant extracts affect the Gené's organ at the gene expression level. This can lead to further research as the results of these studies would lead to possible candidates for tick control targets. Another RNA-seq analysis of the Gené's organ with differing treatment groups following the experimental pipeline in this study would be a method to identify candidate genes to target. Identifying new targets for tick control will allow for ranchers, farmers, and animal health agencies the ability to combat multiple acaricide resistant tick populations which could increase the milk production and lower mortality rates in herds infested with *R. microplus* decreasing the economic impact that this species has around the world.

REFERENCES

1. Bock, R., Jackson, L., De Vos, a. & Jorgensen, W. Babesiosis of cattle. *Parasitology* 129, S247–S269 (2004).

2. Smith, T. & Kilborne, F. L. *Investigations into the nature, causation and prevention of Southern cattle fever. Ninth Annu. Rep. Bur. Anim. Ind. Year 1892* (1893).

3. Murrell, A. & Barker, S. C. Synonymy of *Boophilus* Curtice, 1891 with *Rhipicephalus* Koch, 1844 (Acari: Ixodidae). *Syst. Parasitol*. 56, 169–72 (2003).

4. Feldman-Muhsam, B. & Shechter, R. Some notes on the genus *Boophilus* (Ixodidae), with special reference to species found in Israel. *J. Med. Entomol*. 7, 677–686 (1970).

5. Guerrero, F. D., Lovis, L. & Martins, J. R. Acaricide resistance mechanisms in *Rhipicephalus (Boophilus) microplus*. *Bras. J. Vet. Parasitol*. 21, 1–6 (2012).

6. Service, U. and P. H. I. Tick Disease Information. (2015). at <http://www.aphis.usda.gov/wps/portal/?1dmy&urile=wcm%3Apath%3A/APHIS_Content_Library/SA_Our_Focus/SA_Animal_Health/SA_Animal_Disease_Information/SA_Cattle_Health/SA_Ticks>

7. Pérez de León, A. a et al. Integrated Strategy for Sustainable Cattle Fever Tick Eradication in USA is Required to Mitigate the Impact of Global Change. *Front. Physiol.* 3, 195 (2012).

8. Rodríguez-vivas, R. I. et al. *Rhipicephalus (Boophilus) microplus* resistant to acaricides and ivermectin in cattle farms of Mexico. Bras. *J. Vet. Parasitol*. 23, 113–122 (2014).

9. Guerrero, F. D., Li, A. Y. & Hernandez, R. Molecular Diagnosis of Pyrethroid Resistance in Mexican Strains of *Boophilus microplus* (Acari : Ixodidae). *J. Med. Entomol*. 39, 770–776 (2002).

10. Trapapa, B. J. The campaign against *Boophilus microplus* in Mexico, benefit, problems and prospects. Z. Garcia-Vesquez (ed.), Anim. Prod. Heal. (1989).

11. Fragoso-Sanchez, H. & Garcia-Vazquez, Z. Response of Mexican *Rhipicephalus (Boophilus) microplus* ticks to selection by amitraz and genetic analysis of attained resistance. *J. Entomol*. 8, 218–228 (2011).

12. Soberanes, N., Santamaría, M. & Fragoso, H. First case reported of Amitraz resistance in the cattle tick *Boophilus microplus* in Mexico. *Técnica Pecu. en México* 40, 81–92 (2002).

13. Miller, R. J. et al. First report of fipronil resistance in *Rhipicephalus (Boophilus) microplus* of Mexico. *Vet. Parasitol*. 191, 97–101 (2013).

14. Santamaria, V. M. et al. Analisis de la situacion actual mediante el montoreo de susceptibilidad a ixodicidas en *Boophilus microplus* de 1993 a 1999 y medidas preventivas para retardar la resistencia al amitraz en Mexico. in *IV Semin. Int. Parasitol. Anim. Control la Resist. en Garrapatas y Moscas Importancia Vet. y Enfermedades que Transm*. (1999).

15. Li, A. Y. Status of resistance to acaricides in Mexican strains of the southern cattle tick *Boophilus microplus* (Acari: Ixodidae). *Resist. Pest Manag. Newsl*. 13, 7–12 (2004).

16. Davey, R. B., George, J. E. & Miller, R. J. Control of an organophosphate-resistant strain of *Boophilus microplus* (Acari: Ixodidae) infested on cattle after a series of dips in coumaphos applied at different treatment intervals. *J. Med. Entomol*. 41, 524–528 (2004).

17. Jeyaprakash, A. & Hoy, M. a. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp. Appl. Acarol*. 47, 1–18 (2009).

18. *Biology of Ticks*. (Oxford University Press, 2014).

19. Sonenshine, D. E. & Roe, R. M. in *Biol. Ticks* (Sonenshine, D. E. & Roe, R. M.) 74–98 (Oxford University Press, 2014).

20. Lees, A. D. & Beament, J. W. L. An egg-waxing organ in ticks. Q. *J. Microsc. Sci*. 89, 291–322 (1948).

21. Oliver, J. H. & Dotson, E. M. Hormonal Control of Molting and Reproduction in Ticks. *Am. Zool*. 33, 384–396 (1993).

22. Booth, T. F. Wax lipid secretion and ultrastructural development in the egg-waxing (Gené's) organ in Ixodid ticks. *Tissue Cell* 21, 113–122 (1989).

23. Arrieta, M. C., Leskiw, B. K. & Kaufman, W. R. Antimicrobial activity in the egg wax of the African cattle tick *Amblyomma hebraeum* (Acari: Ixodidae). *Exp. Appl. Acarol.* 39, 297–313 (2006).

24. Edelmann, B., Gobel, E., Schol, H. & Gothe, R. Morphology and structural organization of Gene's organ in *Argas walkerae*. *Med. Vet. Entomol.* 15, 422–432 (2001).

25. Sutherst, R. W. & Bourne, a S. The effect of desiccation and low temperature on the viability of eggs and emerging larvae of the tick, *Rhipicephalus (Boophilus) microplus* (Canestrini) (Ixodidae). *Int. J. Parasitol.* 36, 193–200 (2006).

26. Kakuda, H., Mori, T. & Shiraishi, S. Functional morphology of Gene's organ in *Haemaphysalis longicornis* (Acari:Ixodidae). *Exp. Appl. Acarol.* 16, 263–275 (1992).

27. Zimmer, K. R., Macedo, A. J., Nicastro, G. G., Baldini, R. L. & Termignoni, C. Egg wax from the cattle tick *Rhipicephalus (Boophilus) microplus* inhibits *Pseudomonas aeruginosa* biofilm. *Ticks Tick. Borne. Dis.* 4, 366–76 (2013).

28. Booth, T. F. Observation on the composition and biosynthesis of egg wax lipids in the cattle tick, *Boophilus microplus*. *Exp. Appl. Acarol.* 14, 137–149 (1992).

29. Esteves, E. et al. Antimicrobial activity in the tick *Rhipicephalus (Boophilus) microplus* eggs: Cellular localization and temporal expression of microplusin during oogenesis and embryogenesis. *Dev. Comp. Immunol.* 33, 913–919 (2009).

30. Yu, Z. et al. Antimicrobial activity in the egg wax of the tick *Amblyomma hebraeum* (Acari: Ixodidae) is associated with free fatty acids C16:1 and C18:2. *Exp. Appl. Acarol.* 58, 453–470 (2012).

31. Gilby, a R. Studies of cuticular lipides of arthropods. III. The chemical composition of the wax from *Boophilus microplus*. *Arch. Biochem. Biophys.* 67, 320–324 (1957).

32. McCamish, M., Cannell, G. & Cherry, L. The nonpolar egg wax lipids of the cattle tick, *Boophilus microplus* (Canestrini). *Lipids* 12, 182–187 (1977).

33. Booth, T. F., Beadle, D. J. & Hart, R. J. Ultrastructure of the accessory glands of Gene's organ in the cattle tick, *Boophilus*. *Tissue Cell* 16, 589–599 (1984).

34. Schöl, H., Sieberz, J., Göbel, E. & Gothe, R. Morphology and structural organization of gené's organ in *Dermacentor reticulatus* (Acari: Ixodidae). *Exp. Appl. Acarol.* 25, 327–52 (2001).

35. Ogihara, M. H. & Taylor, D. in *Biol. Ticks* (Sonenshine, D. E. & Roe, R. M.) 456–457 (Oxford University Press, 2014).

36. Guerrero, F. D. et al. BmiGI: a database of cDNAs expressed in *Boophilus microplus*, the tropical/southern cattle tick. *Insect Biochem. Mol. Biol*. 35, 585–95 (2005).

37. Moolhuijzen, P. M. et al. The complexity of *Rhipicephalus (Boophilus) microplus* genome characterised through detailed analysis of two BAC clones. *BMC Res. Notes* 4, 254 (2011).

38. Ullmann, A. J., Lima, C. M. R., Guerrero, F. D., Piesman, J. & Black, W. C. Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. *Insect Mol. Biol*. 14, 217–22 (2005).

39. Davey, R. B., Thompson, G. D. & Drummond, R. O. Ovipostional Biology of the Cattle Tick, *Boophilus annulatus* (Acari: Ixodidae), in the Laboratory. *J. Med. Entomol*. 17, 287–289 (1980).

40. Guerrero, F. D., Bendele, K. G., Davey, R. B. & George, J. E. Detection of *Babesia bigemina* infection in strains of *Rhipicephalus (Boophilus) microplus* collected from outbreaks in south Texas. *Vet. Parasitol*. 145, 156–63 (2007).

41. Oliver, S. L., Lenards, A. J., Barthelson, R. a., Merchant, N. & McKay, S. J. Using the iPlant collaborative discovery environment. *Curr. Protoc. Bioinforma*. 1–26 (2013). doi:10.1002/0471250953.bi0122s42

42. Babraham Bioinformatics: FastQC A Quality Control tool for High Throughput Sequence Data. at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

43. Bolger, a. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible read trimming tool for Illumina NGS data. *Bioinformatics* 1–7 (2014).

44. Grabherr, M., Haas, B. & Yassour, M. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat.* … 29, 644–652 (2013).

45. Xie, Y. et al. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **Bioinforma,** 1–8 (2014).

46. Zerbino, D. Using the Velvet de novo Assembler for Short-Read Sequencing Technologies. *Curr. Protoc. Bioinforma*. 1–13 (2010). doi:10.1002/0471250953.bi1105s31.Using

47. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc*. 8, 1–43 (2013).

48. Zhao, Q.-Y. et al. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* 12 Suppl 1, S2 (2011).

49. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–6 (2005).

50. Quevillon, E. et al. InterProScan: Protein domains identifier. *Nucleic Acids Res*. 33, 116–120 (2005).

51. Hirschman, J. E. et al. Genome Snapshot: a new resource at the *Saccharomyces* Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res*. 34, D442–D445 (2006).

52. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 32, D277–D280 (2004).

53. Ghaffari, N. et al. Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci. Rep*. 4, 7081 (2014).

54. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–92 (2012).

55. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–77 (2011).

56. Strickler, S. R., Bombarely, A. & Mueller, L. A. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *Am. J. Bot*. 99, 257–66 (2012).

57. Neil, S. T. O. & Emrich, S. J. Assessing De-Novo Transcriptome Assemblies. (2013).

58. Yang, Y. & Smith, S. A. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* 14, 328 (2013).

59. Vijay, N., Poelstra, J. W., Künstner, A. & Wolf, J. B. W. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol. Ecol*. 22, 620–34 (2013).

60. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067 (2007).

61. Ye, J., McGinnis, S. & Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res*. 34, W6–W9 (2006).

62. Rodríguez-Vivas, R. I. et al. Red deer (*Cervus elaphus*) as a host for the cattle tick *Rhipicephalus microplus* (Acari: Ixodidae) in Yucatan, Mexico. *Exp. Appl. Acarol*. 60, 543–52 (2013).

63. Guerrero, F. D. et al. Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, *Rhipicephalus (Boophilus) microplus*. *BMC Genomics* 11, 374 (2010).

64. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet*. 12, 671–82 (2011).

65. Garg, R. & Jain, M. in Legum. *Genomics Methods Protoc*. 1069, 43–58 (Humana Press, 2013).

66. Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet*. 15, 121–32 (2014).

67. Heekin, A. M. et al. The ovarian transcriptome of the cattle tick, *Rhipicephalus (Boophilus) microplus*, feeding upon a bovine host infected with Babesia bovis. *Parasit. Vectors* 6, 276 (2013).

68. Heekin, A. M. et al. Gut transcriptome of replete adult female cattle ticks, *Rhipicephalus (Boophilus) microplus*, feeding upon a Babesia bovis-infected bovine host. *Parasitol. Res*. 112, 3075–90 (2013).

69. Egekwu, N., Sonenshine, D. E., Bissinger, B. W. & Michael Roe, R. Transcriptome of the female synganglion of the black-legged tick *Ixodes scapularis* (Acari: Ixodidae) with comparison between Illumina and 454 systems. *PLoS One* 9, (2014).

70. Sonenshine, D. E. et al. First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). *PLoS One* 6, e24711 (2011).

71. Curto, R., O. Voit, E., Sorribas, A. & Cascante, M. Mathematical models of purine metabolism in man. *Math. Biosci*. 151, 1–49 (1998).

72. Berg, J., Tymoczko, J. & Stryer, L. in *Biochemistry* (2002). at <http://www.ncbi.nlm.nih.gov/books/NBK22408/>

73. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323 (2011).

74. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2009).

75. Untergasser, A. et al. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res*. 35, W71–W74 (2007).

76. Pohl, P. C. et al. ABC transporter efflux pumps: a defense mechanism against ivermectin in *Rhipicephalus (Boophilus) microplus. Int. J. Parasitol*. 41, 1323–33 (2011).

77. Pfaffl, M. Quantification strategies in real-time PCR Michael W . Pfaffl. *A-Z Quant. PCR* 87–112 (2004). doi:http://dx.doi.org/10.1007/s10551-011-0963-1

78. Pfaffl, M. W. Relative quantification. *Real-time PCR* 63–82 (2004). doi:10.1186/1756-6614-3-5

79. Melicher, D., Torson, A. S., Dworkin, I. & Bowsher, J. H. A pipeline for the de novo assembly of the *Themira biloba* (Sepsidae: Diptera) transcriptome using a multiple k-mer length approach. *BMC Genomics* 15, 188 (2014).

80. Alduini, N., Silva, M., Franzolin, M., Mendonça, R. & Lima-netto, S. Antimicrobial activity from ticks eggs waxes. *BMC Proc*. 8, P156 (2014).

81. Ribeiro, V. L. S. et al. Acaricidal properties of the essential oil from *Hesperozygis ringens* (Lamiaceae) on the cattle tick *Riphicephalus (Boophilus) microplus*. *Bioresour. Technol*. 101, 2506–9 (2010).

82. Ribeiro, V. L. S. et al. Chemical composition and larvicidal properties of the essential oils from *Drimys brasiliensis Miers* (Winteraceae) on the cattle tick *Rhipicephalus (Boophilus) microplus* and the brown dog tick *Rhipicephalus sanguineus*. *Parasitol. Res*. 102, 531–5 (2008).

83. Booth, T. F., Beadle, D. J. & Hart, R. J. The Effects of Precocene Treatment on Egg Wax Production in Gene's Organ and Egg Viability in the Cattle Tick *Boophilus microplus* (Acarina Ixodidae): An Ultrastructural Study. *Exp. Appl. Acarol*. 2, 187–198 (1986).

84. Rosario-Cruz, R. et al. Genetic basis and impact of tick acaricide resistance. *Front. Biosci*. 14, 2657-2665 (2009).

BIOGRAPHICAL SKETCH

Jason P. Tidwell graduated from the University of Texas – Pan American in May 2015 with a Master of Science in Biology. He received an Associate in Arts degree from Iowa Central Community College in 2004. The following semester he entered Northern Arizona University where he received his Bachelors of Science in Microbiology in 2006. After completing his bachelor's degree, he entered into the master's program at Northern Arizona University working on cancer genetics research. After two years of his work at NAU, he accepted a position at the USDA-ARS Cattle Fever Tick Research Laboratory as a microbiologist since 2009. While he has been in the Rio Grande Valley, he has been active in supporting the Boy Scouts of America as an Assistant Scoutmaster for two years and presently as Scoutmaster for troop 106. His mailing address is 214 ½ S. 12th Ave, Apartment 2, Edinburg, TX 78539.