RAFAEL MAZIOLI BARCELOS

IS *Rickettsia amblyommii* ABLE TO REGULATE LONG NON-CODING RNA EXPRESSION IN *Amblyomma sculptum* TICK? AN *in silico* APPROACH

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2017

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T

B242i
2017

Barcelos, Rafael Mazioli, 1985-

Is *Rickettsia amblyommii* able to regulate long non-coding RNA expression in *Amblyomma sculptum* tick? : An *in silico* approach / Rafael Mazioli Barcelos. – Viçosa, MG, 2017.

x, 41f. : il. (algumas color.) ; 29 cm.

Orientador: Cláudio Lisias Mafra de Siqueira.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Carrapato. 2. *Amblyomma sculptum*. I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-graduação em Bioquímica Aplicada. II. Título.

CDD 22 ed. 595.429

RAFAEL MAZIOLI BARCELOS

IS *Rickettsia amblyommii* ABLE TO REGULATE LONG NON-CODING RNA EXPRESSION IN *Amblyomma sculptum* TICK? AN *in silico* APPROACH

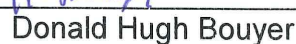
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Doctor Scientiae*.

APROVADA: 28 de junho de 2017.

Jorge Abdala Dergam dos Santos

Maximiller Dal-Bianco Lamas Costa

Raphael Contelli Klein



Donald Hugh Bouyer

Cláudio Lísias Mafra de Siqueira
(Orientador)

“Ninguém é responsável pelo meu fracasso. Ninguém é responsável pela minha felicidade”.

(Leandro Karnal)

AGRADECIMENTOS

Pela realização deste trabalho, gostaria de agradecer:

- À Deus, por ter guiado e abençoado todo o meu caminho até aqui;
- À minha mãe, Eliana, e ao meu pai, Geronimo, eternos amigos e companheiros responsáveis por toda a minha determinação, caráter e inspiração. Pelo apoio incondicional nos momentos difíceis durante a realização deste trabalho;
- Aos meus irmãos, Guilherme e Nathália, pelo apoio, torcida e amizade que levarei durante toda a minha vida. À minha linda afilhada Lara, um presente de Deus em minha vida;
- Ao Prof. Dr. Cláudio Mafra por ter proporcionado toda a infra-estrutura e orientação para a realização e conclusão deste trabalho;
- Aos amigo(a)s irmã(o)s Mari, Grazi, Cynhia, Natasha, Filippe, Michele por todo apoio, conselhos e ombro durante todo o processo desta fase.
- À família Barcelos e Mazioli, pelo apoio e torcida sempre;
- À Melina Pena, por toda ajuda profissional, conselhos e me fazer enxergar e enfrentar as dificuldades da vida de modo otimista e racional;
- Thank you Galveston and UTMB friends: Nicole, Chris, Guang, Paula, Inaia, Lorenzo, Tom, Patricia and Pep for all friendship and good times in Galveston, UTMB and USA. “Screw you, we’re from TEXAS!”;
- Aos professores Dr. Jorge Dergam (DBA/UFV), Dr. Raphael Contelli Klein (UFOB), Maximiller Dal-Bianco Lamas Costa (DBB/UFV) que, solicitamente, aceitaram fazer parte da banca examinadora desta tese de doutorado;
- Thank you Professor Dr. Donald H. Bouyer and Miguel Grimaldo, for all advices and friendship when I was in Galveston;
- Aos companheiros do LAPEM, em especial Carlos, Gabriel, Paulão, Fernanda e Adriano;
- Aos colegas do NUBIOMOL, em especial Pedro, pela imensa ajuda neste trabalho e Edvaldo. Obrigado pelos conselhos e amizade de ambos;
- Aos colegas do DBB;
- Ao secretário da Pós-Graduação do DBB, Eduardo Monteiro, por ser sempre solícito, paciente e eficiente;
- À CAPES e CNPq, pela bolsa de incentivo para realização deste trabalho;

SUMÁRIO

LISTA DE FIGURAS v
LISTA DE TABELAS vi
LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS vii
ABSTRACT ix
RESUMO x
1. GENERAL INTRODUCTION 1
CHAPTER 1: BACKGROUND 3
1. Amblyomma sculptum tick 3
2. Tissues RNA-seq 6
2.1. Midgut 6
2.2. Ovary 7
3. RNA-seq of A. sculptum (Amblyomma cajennense complex) 8
4. Long non-coding RNA (lncRNA) 9
5. Rickettsia amblyommii 12
6. REFERENCES 15
CHAPTER 2: Is Rickettsia amblyommii able to regulate long non-coding RNA expression in Amblyomma sculptum tick? An in silico approach 21
1. INTRODUCTION 21
2. METHODS 23
3. RESULTS 26
4. DISCUSSION 32
4.1. Midgut upregulation 32
4.2. Midgut downregulation 34
4.3. Ovary differential expression 35
5. CONCLUSIONS 36
6. PERSPECTIVES 37
7. ACKNOWLEDGMENTS 38
8. REFERENCES 39

LISTA DE FIGURAS

Figure 1. Three-host ixodid ticks have a life cycle that spends three years, generally, although some species can complete the cycle in only two years. Adult females drop off the third host to lay eggs after feeding (1), usually in the fall. Eggs hatch into six-legged larvae (2) and overwinter in the larval stage. In the spring, the larvae seeks the first host to attach it, usually a small mammal (3). Later in the summer, engorged larvae leave the first host (4) and molt into nymphs (5), usually in the fall. The ticks overwinter in this stage. During the following spring, the nymphs seeks the second host (6), usually another small mammal. The nymphs feed on the second host and drop off later in the summer (7). Nymphs molt into adults (7a) – (7b) off the host in the late summer or fall, and overwinter in this stage. The next spring, adults seeks the third host, which is usually a big mammals or human (8). The adults feed and mate on the third host during the summer. Females drop off the host in the fall to continue the cycle. Females may reattach and feed multiple times. The three hosts do not necessarily have to be different species, or even different individuals. Also, humans may serve as first, second or third hosts. Source: Centers for Diseases Control and Prevention (CDC) at https://www.cdc.gov/dpdx/ticks/index.html. 5

Figure 2. Some lncRNA genes contain embedded miRNA sequences (red boxes), located in two different regions: an exon (blue box) or an intron (line) of the gene. The three sources result in very different types of primary transcript but the pathways converge at the level of pre-miRNA structure. Legend: miRNA, microRNA; pri-miRNA, primary miRNA; pre-miRNA, precursor miRNA. Source: (DYKES et al., 2017). 11

Figure 3. Method workflow showing how to reach the putatives lncRNAs. The asterisk shows the filter steps to eliminate any coding sequence....................25

Figure 4. Filter workflow to select lncRNAs. Two assembly software were tested: Trinity and CLC Genomics. The numbers inside the boxes represent how many transcripts left after each filter while the numbers next to the arrows show how many transcripts increased or decreased. 26

Figure 5. Heatmap of some putatives lncRNAs up and downregulated from Trinity assembly. The numbers on the right side are the sequences numbers in the Table 3. All putative lncRNA are located in supplemental tables. 29

Figure 6. Heatmap of some lncRNAs up and downregulated from CLC Genomics assembly. The numbers on the right side are the sequences numbers in the Table 3. All putative lncRNA are located in supplemental tables. 30

LISTA DE TABELAS

Table 1. Number of sequences upregulated and downregulated infected and non-infected tissues**..27**

Table 2. The top species hits from Trinity and CLC annotations. Numbers indicate the quantity of hits found. **...27**

Table 3. Annotation of putative lncRNAs upregulated and downregulated in midgut and an unique annotated downregulated ovary infected sequence**...........................31**

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

LAPEM	Laboratório de Parasitologia e Epidemiologia Molecular
UFV	Universidade Federal de Viçosa
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CAPES	Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior
DBB/UFV	Departamento de Bioquímica e Biologia Molecular – UFV
RNA-seq	Sequência de ácido ribonucléico
RT-qPCR	Reverse Transcriptase Quantitative Polimerase Chain Reaction
cDNA	Ácido desoxirribonucleico complementar
qPCR	Real time Polimerase chain reaction quantitative
SeqID	idential residues alignment sequence
EMBL-EBI	European Bioinformatcs Institute
TrEMBL	Translate Sequence for European Molecular Data Library
UniProt	Resource Protein Universal
NIH	National Institute Health
EMBL	European Molecular Biology Laboratory
UniProtKB	UniProt Knowledge Base
UniRef	UniProt Refernce Clusters
UniParc	UniProt Archive
NuBiomol	Nucleo de biomoléculas da Universidade Federal de Viçosa
PFAM	The Protein Family Database
CDD	Conserved Domains Database
ncRNA	Non-coding RNA
lncRNA	Long non-coding RNA
EST	Expressed sequence tag
RNA-seq	Sequenciamento de RNA
CDS	Coding sequence
rRNA	RNA ribossomal
tRNA	RNA transportador
RNA pol II	RNA polimerase II
lincRNA	Long intergenic non-coding RNA

SFG	Spotted Fever Group
IgG	Imunoglobulina G
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTx	Basic Local Alignment Search Tool for translated nucleotide query
RPSBLAST	Reverse Position-Specific BLAST
CPC	Coding Potential Calculator
GtRNAdb	Genomic tRNA Database
RefSeq RNA	Reference Sequence RNA database
NONCODE	Non-coding Sequence database
mRNA	RNA mensageiro
PIN	Partially intronic non-coding
TIN	Totally intronic non-coding
PTC	Premature translation-termination codon
NMD	Nonsense-mediated mRNA decay
UTR	Untranslated region
ORF	Open Reading frame
UTMB	University of Texas Medical Branch
TACC	Texas Advanced Computing Center
sORF	Small ORF
ROS	Reactive Oxygen Species

ABSTRACT

BARCELOS, Rafael Mazioli, D.Sc., Universidade Federal de Viçosa, June, 2017. ***Is Rickettsia amblyommii* able to regulate long non-coding RNA expression in *Amblyomma sculptum* tick? An *in silico* approach.** Adviser: Cláudio Lísias Mafra de Siqueira.

The *Amblyomma sculptum* tick belongs to *Amblyomma cajennnense* complex, the most important taxon in transmission and host for *Rickettsia rickettsii* bacteria, the main etiologic agent of Brazilian Spotted Fever in Brazil and Rocky Mountain Spotted Fever in USA. As a member of this species complex, *A. sculptum* tick is actually very important under human and veterinary subject acting as vector for this zoonotic agent in Brazil. Its role in biological cycle of rickettsia strains, as well *Rickettsia*-tick interactions, still need to be better understood since recent studies found these strains infecting *A. sculptum* ticks. As putative modulators of these pathogen-host interaction, the roles of long non-coding RNAs (lncRNA) of ticks are unknown and need to be discovered and characterized. The lncRNAs are involving in modulation of a plethora of cell activities as transcription, gene silencing and chromosome opening, for example. Thus, herein we present an *in silico* approach for analyze the modulation of lncRNAs by *R. amblyommii* in *A. sculptum* tick. Using previously published data sets of *A. sculptum* transcriptomes, we identified putative lncRNAs and evaluated for differential expression in midgut and ovaries in this tick specie. Two assemblers were tested, Trinity and CLC Genomics, to construct the contigs and a pipeline to select the sequences with lncRNA characteristics. We obtained 31,888 and 23,733 contigs of putative lncRNAs using Trinity and CLC Genomics, respectively. We further identified more than 500 sequences of putative lncRNA that significantly aligned to sequences of RefSeq RNA, RNA Central and NONCODE databases. The transcriptome analysis further suggests that *R. amblyommii* is inducing differential expression of putative lncRNAs in midgut and ovary tissues. The work herein contributes for tick lncRNA database increasing and the initials insights of which lncRNA sequences are involving in *Rickettsia*-tick relationship.

RESUMO

BARCELOS, Rafael Mazioli, D.Sc., Universidade Federal de Viçosa, junho de 2017. **A *Rickettsia amblyommii* é capaz de regular a expressão de RNA não codificante de cadeia longa no carrapato *Amblyomma sculptum*? Uma abordagem *in silico***. Orientador: Cláudio Lísias Mafra de Siqueira.

O carrapato *Amblyomma sculptum* pertence ao complexo *Amblyomma cajennnense*, taxon mais importante na transmissão e como hospedeiro da bactéria *Rickettsia rickettsii*, principal agente etiológico da Febre Maculosa Brasileira e da Febre Maculosa das Montanhas Rochosas nos EUA. Como membro deste complexo de espécies, o carrapato *A. sculptum* é atualmente muito importante sob o aspecto médico veterinário atuando como vetor deste agente zoonótico no Brasil. Seu papel no ciclo biológico da riquétsia, bem como a relação *Rickettsia*-carrapato, precisam ser bem compreendidos pois estudos recentes encontraram esta espécie de bactéria no carrapato *A. sculptum*. Como parte desta relação, as funções de RNAs não codificantes de cadeias longas (lncRNA) são desconhecidas e precisam ser descobertas. Os lncRNAs estão envolvidos em uma infinidade de atividades celulares tais como transcrição, silenciamento gênico e abertura cromossômica, por exemplo. Dessa forma, apresentamos aqui uma abordagem *in silico* da modulação de lncRNAs pela *Rickettsia amblyommii* em carrapatos da espécie *A. sculptum*. Tomando como dados trabalhos publicados anteriormente nós identificamos e avaliamos possíveis lncRNAs para expressão diferencial em intestinos e ovários desta espécie de carrapato. Dois montadores de sequencias foram testados, Trinity e CLC Genomics, para construção de *contigs* e um filtro a partir de um *pipeline* caseiro para selecionar apenas sequências com características de lncRNA. Um total de 31.888 e 23.733 *contigs* foram montados pelo Trinity e CLC Genomics, respectivamente. Nós anotamos mais de 500 sequências possíveis de lncRNA contra os bancos de dados RefSeq RNA, RNA Central e NONCODE. Nossos resultados sugerem que a *R. amblyommii* está induzindo expressão diferencial de lncRNAs nos tecidos dos intestinos e ovários. Nosso trabalho contribui para o aumento do banco de dados de lncRNA de carrapatos bem como fornecer idéias iniciais de quais sequências de lncRNA estão envolvidas na relação *Rickettsia*-carrapato.

1. GENERAL INTRODUCTION

Ticks hold a position of high importance among the species considered as sources for biomolecule discovery, since they produce a complex repertoire of peptides with a wide range of biological activities (EGEKWU et al., 2014; GARCIA et al., 2014). In the Ixodidae family, a new group classified as *Amblyomma sculptum* (formerly denominated as the *Amblyomma cajennense* specie) stands out as a very aggressive tick with a trioxen life cycle that is a parasite for a wide range of vertebrate hosts, from wild and domestic species to human beings. *Amblyomma sculptum* has a large geographic distribution in Latin America, especially in peri-urban areas, being characterized as an important zoonotic and epidemiological agent (BARROS-BATTESTI et al., 2006; NAVA et al., 2014).

Nowadays, transcriptomic sequencing has emerged as an approach for the high-throughput detection of mRNA molecules, providing useful information about the gene content of species (EGEKWU et al., 2014; MARDIS, 2008). Furthermore, non-coding RNAs (ncRNA) are in spotlight since their multiple functions and localizations inside cells (KAPRANOV et al., 2007). The long non-coding RNA (lncRNA), is a ncRNA type and, as rule of thumb, they have more than 200bp length, produced by RNA pol II, have RNA processing events as metazoan cells and ORFs shorter than 100 amino acids but don’t be able to encode protein (KAPRANOV et al., 2007; WANG et al., 2011). A few classifications of lncRNA have been made by empirical attributes used to detect them: (1) based on transcript length; (2) based on association with annotated protein-coding genes; (3) based on association with other DNA elements of known function; (4) based on protein-coding RNA resemblance; (5) based on association with repeats; (6) based on association with a biochemical pathway or stability; (7) based on sequence and structure conservation; (8) based on expression in different biological states; (9) based on association with subcellular structures; (10) based on function (ST.LAURENT et al., 2015). Despite all these lncRNA classes, a lot of functions, how they regulate and are regulated in different conditions and cell compartments remains scant.

In addition, *A. sculptum* organs, such as the midgut and ovaries, still require further study in ncRNA field. The largest interface between ticks and pathogens occurs in the midgut portion, and the ovaries are the place where rickettsiae amplification events and its transovarian transmission takes place, a key process inspotted fevers epidemiology (LABRUNA et al., 2008, 2011). Since different tissues have distinct gene regulation, RNA-Seq analysis of different tissues can expand the *A. sculptum* lncRNA database and shed light on its biological processes.

Thus, the work herein aims the different expression of lncRNAs through RNA-seq sequencing of infected and non-infected *A. sculptum* by *Rickettsia amblyommii* in different tissues.

CHAPTER 1: BACKGROUND

1. *Amblyomma sculptum* tick

At the beginning of the second half of the last century, several morphologically similar individuals were grouped into the *A. cajennense* taxon as synonyms (FONSECA et al., 1953). However, ecological differences were not evaluated at that time indicating that this tick species could include genetically diverse populations (subspecies) (GUGLIELMONE A, MANGOLD A, 1992), which means different synonymies existing within these species, characterizing a species complex. Thus, distinct species and consequently genetic differences would cover this complex currently known as *A. cajennense*.

In a breeding experiment in Argentina, which consisted of crossing two geographically distinct *A. cajennense* populations (El Rey National Park in the Province of Salta and Copo National Park in the Province of Santiago del Estero) was analyzed the reproductive compatibility between them (MASTROPAOLO et al., 2011). It was observed that the larvae born was extremely low when males and females from different populations were crossed. In a similar work, tests with *A. cajennense* populations from Colombia (Villeta, Cundinamarca), Argentina (El Rey National Park, province of Salta) and Brazil (states of São Paulo and Rondônia) achieving similar results, with populations belonging to Colombia and Brazil presenting low fertility when crossed (LABRUNA et al., 2011).

By phylogenetic and phylogeographic analyzes, was observed a genetic structuring of six geographically distinct clades of *A. cajennense* in samples from South American countries and some from North and Central America under different habitats and vegetation characteristics of the place (BEATI et al., 2013). Similar results were obtained through the High Resolution Melting (HRM) and phylogeographic analysis of *A. cajennense* ticks from Minas Gerais, Paraná, Mato Grosso and Goiás states in Brazil, corroborating these previous results (BARCELOS,

2013). A meticulous morphological analysis associated with molecular findings confirmed the existence of six species in *A. cajennense* taxon (NAVA et al., 2014).

Thus, the six species that make up the *A. cajennense* complex are: *A. cajennense* sensu stricto, located in the Amazon region of the South American countries Venezuela, the Guianas and in the Brazilian states Pará, Roraima and Tocantins; *A. interandinum*, reported in the Andean region of Peru; *A. mixtum* from Texas, USA, to the west of Ecuador; *A. patinoi* present so far only in the eastern Cordillera of Colombia; *A. tonelliae*, distributed in the dry areas of the central-north Chaco region of Argentina, Bolivia and Paraguay; and *A. sculptum* spreading from the wetlands of northern Argentina, Bolivia and Paraguay, to the Amazonian areas of Brazil, located in the states of Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo, southern Paraná, Pernambuco, Mato Grosso, Mato Grosso do Sul and the center of Goiás (NAVA et al., 2014).

The ticks of this complex are obligate hematophages, belonging to the class Arachnida, subclass Acarina and order Ixodida (NAVA et al., 2007). They are classified as trioxene ticks, requiring three hosts to complete their life cycles. This characteristic, combined with low parasitic specificity, contributes to the fact that these ticks play a central role in the transmission of potential pathogens from wild reservoirs to humans, which confers to this arthropod a great zoonotic importance to public health (BARROS-BATTESTI et al., 2006). From an epidemiological point of view, they play a major role as spotted fever transmitters and other zoonosis of medical and veterinary importance. It has been shown to be one of the most common parasite species of vertebrates in the world. From the point of reproduction view, a single engorged female can lay between 5,000 and 10,000 eggs, which after 30 days will hatch, generating between 5,000 and 10,000 larvae (BARROS-BATTESTI et al., 2006). Some days after the larvae hatch in the soil, as soon as the exoskeleton matures, they rise in the leaves of the grass, forming numerous agglomerates of larvae, waiting for a vertebrate host. In the immature stages, they do not have a marked parasitical specificity attaching to various vertebrate species, among them humans. Once on the host, the larvae feed on a period of 3 to 7 days. After this period, they fall out on the ground, where they perform the ecdise to the nymph stage. The pre-molting period can take from 4 to 8 weeks. After 15 days of molting period, they will start seeking a new host to feed (BARROS-BATTESTI et al., 2006).

At the time they infest a new host, the nymphs feed again for a period of 3 to 7 days, falling to the ground where they become adults after a period of 4 to 8 weeks. The Figure 1 shows a typic three hosts tick cycle like *A sculptum*. As soon as exoskeleton hardening occurs (approximately 2 weeks after molting), adults are able to infest other, often large, vertebrate hosts, being prepared to remain for several months without feeding. After fertilized and engorged, the female will fall out on the soil to perform the oviposition, which, after a period of approximately 30 days, will hatch the eggs and a new generation of the larvae close the cycle. Males generally continue on the vertebrate host for 3 to 4 weeks feeding and reproducing with other females (BARROS-BATTESTI et al., 2006).

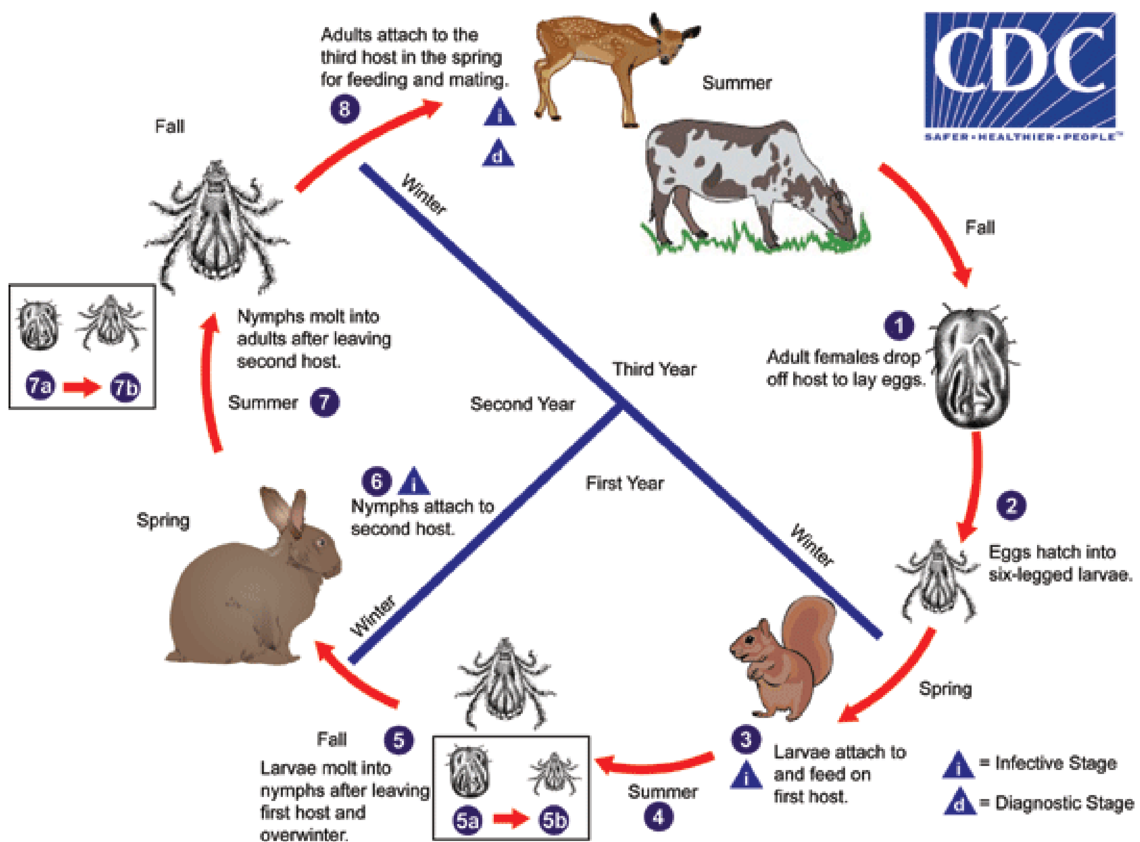


Figure 1. Three-host ixodid ticks have a life cycle that spends three years, generally, although some species can complete the cycle in only two years. Adult females drop off the third host to lay eggs after feeding (1), usually in the fall. Eggs hatch into six-legged larvae (2) and overwinter in the larval stage. In the spring, the larvae seeks the first host to attach it, usually a small mammal (3). Later in the summer, engorged larvae leave the first host (4) and molt into nymphs (5), usually in the fall. The ticks overwinter in this stage. During the following spring, the nymphs seeks the second host (6), usually another small mammal. The nymphs feed on the second host and drop off later in the summer (7). Nymphs molt into adults (7a) – (7b) off the host in the late summer or fall, and overwinter in this stage. The next spring, adults seeks the third host, which is usually a big mammals or human (8). The adults feed and mate on the third host during the summer. Females drop off the host in the fall to continue

the cycle. Females may reattach and feed multiple times. The three hosts do not necessarily have to be different species, or even different individuals. Also, humans may serve as first, second or third hosts. Source: Centers for Diseases Control and Prevention (CDC) at https://www.cdc.gov/dpdx/ticks/index.html.

2. Tissues RNA-seq

2.1. Midgut

At the present time more than 30 scientific articles have been published with the subject RNA-seq of ticks. Of these, the organ preferentialy studied was the salivary gland of the tick, followed by intestine and ovary. A few species of ticks have already been contemplated with transcriptional analyzes, including: *Rhipicephalus (Boophilus) microplus* (HEEKIN et al., 2013), *R. appendiculatus* (DE CASTRO et al., 2016), *R. pulchellus* (TAN et al., 2015), *R. sanguineus* (ANATRIELLO et al., 2010), *Ixodes ricinus* (SCHWARZ et al., 2014), *I. pacificus* (FRANCISCHETTI et al., 2005), *I. scapularis* (LEWIS et al., 2015), *Ornithodoros parkeri* (FRANCISCHETTI et al., 2008a), *O. moubata* (OLEAGA et al., 2017), *O. coriaceus* (FRANCISCHETTI et al., 2008b), *Dermacentor variabilis* (ANDERSON et al., 2008), *D. andersoni* (ALARCON-CHAIDEZ et al., 2007), *Antricola delacruzi* (RIBEIRO et al., 2012); *Argas monolakensis* (MANS et al., 2008), *Haemaphysalis longicornis* (NAKAJIMA et al., 2005), *H. flava* (XU et al., 2016) and *Hyalomma marginatum rufipes* (FRANCISCHETTI et al., 2011).

The midgut, largest organ of the tick, is where the blood of the host undergoes digestion. Exceptionally for erythrocytes, which are digested in the lumen, all blood is digested intra-cellularly, unlike other hematophagous arthropods (SONENSHINE et al., 2013). Due it occurs within cells, the digestion is slowly performed with storage of nutrients during the process. In this way, the large volume of blood ingested is stored in the intestinal lumen. This leads to prolonged life of these arachnids, which can survive for several months after blood feeding (SONENSHINE et al., 2013). The mialome (intestine transcriptome) is exploited because it contains innumerable molecules related to digestion (e.g. proteases and lipases), as well as defense

molecules against infecting micro-organisms. As a complex mixture, blood requires several molecules for effective digestion as the blood composition containing plasma proteins, minerals, vitamins, hormones and other biomolecules (BALASHOV, 1972). With the presence of blood proteins, proteases are commonly found in mialomes, for example, serine proteinases, cysteines proteinases, aspartyl proteinases and carboxipeptidases found in transcriptomes of *R. (B.) microplus* and *I. ricinus* (HEEKIN et al., 2013; KOTSYFAKIS et al., 2015). Another consequence of the absence of digestive enzymes in the intestinal lumen is the creation of an environment conducive to the development of infectious agents from hosts. The protozoan of the *Babesia* genus, also transmitted by the tick species *R. (B.) microplus*, entails expression of stress-related genes such as peroxynectin, cytochrome P450, Toll receptors, glutathione S-transferase and cytochrome c oxidase (HEEKIN et al., 2013). Enzymes that metabolize lipids and carbohydrates also are found in the midgut environment. For the metabolism of carbohydrates, enzymes such as glycosyl hydrolases, galactosidases. A-glycosidases, α-amylases and chitinases have already been observed. For the lipid metabolism, the following enzymes have already been found: pancreatic lipases, phospholipases, sphingomyelinases, triacylglycerol lipase and others (ANDERSON et al., 2008; HEEKIN et al., 2013; KOTSYFAKIS et al., 2015).

2.2. Ovary

The ovary, the female reproductive organ, is responsible for the production of oocytes by oogenesis. Unlike insects, the ovary of ticks is not segmented in different zones in which they have differentiated rates of reproductive development (SONENSHINE et al., 2013). Vitellin or calf (VT) is formed in post-fertilization oocytes which has vitellogenin as a precursor. VT is the main component of the egg yolk sac. After the process of fertilization and feeding of the female, the embryogenesis begins. At this stage, in addition to the production of VT, numerous molecules must be gathered so that there is no lack of nutrients for the development of the embryo. Among the molecules we have proteins, lipids, sugars among others. These components are stored in special organelles, the yolk granules (ATELLA et al., 2005). In addition to

the granules, there must be accumulation of vitello, also called vitellogenesis, which will accumulate inside the oocytes. After the process of vitellogenesis the embryo will have all the nutrients necessary for the embryogenesis to occur when these eggs grow in the environment (LOGULLO et al., 2002; SEIXAS et al., 2010). Since VT and vitellogenin are one of the main components for the development of the embryo, transcripts related to these proteins, is more frequent in transcriptomes involving ovaries (HEEKIN et al., 2013; KOTSYFAKIS et al., 2015; STUTZER et al., 2013). In oogenesis and embryogenesis several mitotic and meiotic processes occur. Thus, transcripts related to DNA replication such as DNA polymerases, RNA polymerases, histones, chromosome organization related proteins, cyclin-dependent kinases, cyclins among other molecules are observed in ovarian transcriptomes (STUTZER et al., 2013). The ovaries are one of the major target organs of pathogens such as bacteria and protozoa of the genus *Rickettsia* and *Babesia*, respectively. This preference is due to transovarian transmission of these pathogens to the offspring. Thus, as in mialome studies, several are the transcripts related to tick immunity found in ovary transcriptomes (HEEKIN et al., 2013; KOTSYFAKIS et al., 2015; STUTZER et al., 2013).

3. RNA-seq of *A. sculptum* (*Amblyomma cajennense* complex)

The *Amblyomma* genus has been extensively explored by transcriptome analyzes but only protein coding sequences were further discussed. In other hand, the non-coding RNA sequences were only counted and not explored with more details about annotations. Within *Amblyomma* genus, the following species had transcriptome studied: *A. americanum, A. variegatum, A. cajennense, A. maculatum, A. aureolatum, A. triste, A. sculptum* and *A. parvum* (GARCIA et al., 2014; KARIM et al., 2015, 2011; MARTINS et al., 2017a; MOREIRA et al., 2016; RIBEIRO et al., 2011). All studies involving this genus were mostly in salivary glands followed by midgut. Despite the aim of all these studies were protein coding genes, the unknown sequences found by them might have non-coding RNAs (ncRNA). For example, around 1,806 unknown CDS in salivary glands, midgut and ovaries tissues in *A. sculptum* transcriptome (MOREIRA et al., 2016); and around 842 ESTs in *A.*

cajennense with unknown functions (BATISTA et al., 2008). From an RNA-seq experiment involving three *Amblyomma* species, *A. cajennense* (nowadays *A. sculptum*), *A. parvum* and *A. triste*, a *A. sculptum* showed 762 unknown CDS, while *A. parvum* 631 and *A. triste* 2,416 CDS a total of 3,809 sequences (GARCIA et al., 2014). A different research with *A. sculptum* and *A. aureolatum* infected by *R. rickettsii* showed a high unknown CDS number too, around 1,968 sequences for *A. sculptum* and 2,475 for *A. aureolatum*, both important *Rickettsia*-transmitter species in Brazil (MARTINS et al., 2017). The latter work was in midgut which highlight the importance of putative non-coding RNAs in all tissues nor only in salivary gland. It is worth of note in *I. scapularis* genomic approach around 4,439 ncRNA genes were found (GULIA-NUSS et al., 2016). We know human beings have approximately 50-70% non-coding regions in DNA which suggests that ticks might have a high non-coding regions coverage too, but different when compared to other organisms (DJEBALI et al., 2012; IYER et al., 2015). This information show us a great importance of non-coding RNA and also a much more ncRNA sequences need to be discovered by future RNA-seq works.

4. Long non-coding RNA (lncRNA)

The long non-coding RNAs (lncRNA) belong to non-coding RNA class such as tRNA and rRNA. As rule of thumb, they have more than 200bp length produced by RNA pol II, with RNA processing events like metazoan cells and ORFs shorter than 100 amino acids but incapable to encode protein (KAPRANOV et al., 2007; WANG et al., 2011). However, this classification is limited and classes have been risen with distinct functions. A few classifications of lncRNA have been made by empirical attributes used to detect them: (1) based on transcript length; (2) based on association with annotated protein-coding genes; (3) based on association with other DNA elements of known function; (4) based on protein-coding RNA resemblance; (5) based on association with repeats; (6) based on association with a biochemical pathway or stability; (7) based on sequence and structure conservation; (8) based on expression in different biological states; (9) based on association with subcellular structures; (10) based on function (ST.LAURENT et al., 2015). Despite all these

lncRNA classes, a lot of functions, how they regulate and are regulated in different conditions and cell compartments remain scant.

The lncRNA can be found regulating transcription (DYKES et al., 2017), chromatin opening (BÖHMDORFERR et al., 2015), gene silencing (ZHAO et al., 2010), chromosome inactivation (LEE, 2011) and translation regulations (VAN HEESCH et al., 2014). The best lncRNA study case is the X chromosome inactivation (XCI) as a dosage compensation example. One of the two X chromosomes in females is heterochromatinized resulting in silencing (Xi or inactive X, or Xi) and the other X remains active and is expressed in each female cell (Xa or active X) (LEE, 2011; LYON, 1961). XCI in placental mammals is largely controlled by a cluster of lncRNA loci known as the X-inactivation center (*Xic*) (BROWN et al., 1991). A lncRNA called *Xist* (17-kb X (inactive)-specific transcript) is highly expressed from Xi during the onset of XCI, but not from Xa. The silencing form is composed by *Xist* RNA grouped with X chromosome and forms an "*Xist* cloud" (BROWN et al., 1992; CLEMSON et al., 1996), acting as scaffold recruitment of silencing factors. To activate it, *Xist* itself is regulated by other lncRNAs. *Tsix*, an antisense transcript in opposite orientation from a promoter downstream of *Xist*, is highly expressed before XCI initiates but is active in Xa showing reverse pattern as *Xist* expression (LEE et al., 1999). *Tsix* has been demonstrated to coordinate X chromosome pairing to generate epigenetic asymmetry within the *Xist* locus (BACHER et al., 2006; XU, 2006) and to downregulate *Xist* by a number of potential mechanisms (OGAWA et al., 2008; SADO et al., 2005; SUN et al., 2006; ZHAO et al., 2008).

Some lncRNA can act as micro RNA (miRNA) precursor, also called host gene, such as the tumour suppressor miR-31 that is downregulated in breast cancer. *MIR-31* gene is located at the first intron of the lncRNA LOC554202 and the methylation of the host gene promoter controls the transcription of this gene (AUGOFF et al., 2012). The Figure 2 shows some examples of how lncRNA results in miRNA (DYKES et al., 2017).

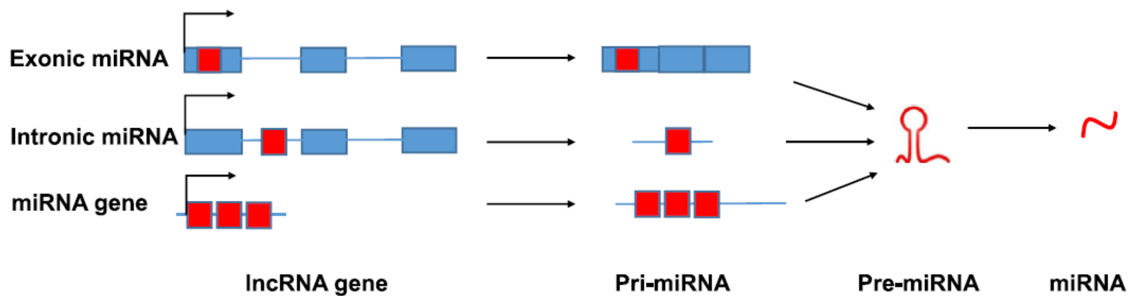


Figure 2. Some lncRNA genes contain embedded miRNA sequences (red boxes), located in two different regions: an exon (blue box) or an intron (line) of the gene. The three sources result in very different types of primary transcript but the pathways converge at the level of pre-miRNA structure. Legend: miRNA, microRNA; pri-miRNA, primary miRNA; pre-miRNA, precursor miRNA. Source: (DYKES et al., 2017).

In arthropods, we have few cases of empirical studies, such as *Drosophila melanogaster* (CHEN et al., 2016), *Aedes aegypti* (ETEBARI et al., 2016), *Bombyx mori* (WU et al., 2016), *Apis cerana* and *A. mellifera* (JAYAKODI et al., 2015). One class of lncRNA, a lincRNA (long intergenic non-coding RNA) was studied in some of these arthropods. In *A. aegypti* study, infection by dengue virus serotype 2 (DENV-2) and *Wolbachia* was evaluated which generated 3,482 putative lincRNAs, while in fifth instar silkworm larvae *Bombyx mori*, the researchers found around 11,810 lncRNAs sequences; for *Apis* species infected by virus causative of SBV disease, a 2,470 lincRNAs were found from *A. cerana* and 1,514 lincRNAs in *A. mellifera.* While in *Drosophila melanogaster* study, 1,077 lncRNAs were found in 27 developmental processes. In each specie, treatment and instar level study were generated unique and new lncRNAs when compared with previous works, when available such as for *Drosophila melanogaster* (LI et al., 2012) and *Anopheles gambiae* (JENKINS et al., 2015), another mosquito specie, which highlight the diverse lncRNA universe and its specificity.

Some of these studies had proved experimentally the existence of lncRNA. For example, in *Drosophila melanogaster* study, was verified that *Yellow-achaete intergenic RNA (yar)* lncRNA is a regulator of yellow and achaete gene transcription with sleep-regulation-link in a normal circadian rhythm (SOSHNEV et al., 2011). In this study, the lncRNA gene was measured by qPCR in *Drosophila yar* mutants. The flies lacking the gene showed shortened sleep bouts within a normal circadian sleep–wake cycle. A genome-wide characterization utilizing RNA-seq experiment data of *Apis* genus infected by SBV virus disease, six lincRNA genes were tested, *Nb-1, Ks-*

1, AncR-1, and *kakusei* which are involved in behavior and brain related (JAYAKODI et al., 2015). The putative lincRNA expressions were validated by RT-qPCR of sequences generated by RNA-seq. In an *A. aegypit* study, potential lincRNAs were selected by RNA-seq experiment, validated by RT-qPCR and functional tested with RNAi knockdown the selected targets (ETEBARI et al., 2016). Twelve lincRNAs were tested which 6 for dengue virus serotype 2 and 6 for *Wolbachia*. A lincRNA named lincRNA_1317 may be involved in DENV-2 replication.

5. *Rickettsia amblyommii*

The *R. amblyommii* is a Gram-negative bacteria intracellular obligate, α-Proteobacteria as every *Rickettsia* of spotted fever group (SFG). The SFG is maintained in environment by ticks, causing severe diseases in humans. In this group, several outer membranes proteins are known to help the bacteria invasion in host cells, e.g. rompA and rompB (ELLISON et al., 2008; FELSHEIM et al., 2009). Its pathogenicity is undetermined despite the presence of invasion outer membrane proteins and evidences in tick *Rickettsia*-infected areas in Arkansas and Virginia states, USA, in a 12 members of a military unit (DASCH et al., 1993). They developed mild illnesses and had antibodies to SFG Rickettsiae when their sera were tested in western blots (DASCH et al., 1993). Interestingly, five of the 12 cases showed specific reactivity to major surface-protein antigens of strain 85-1034, a strain of *R. amblyommi* in USA, indicating their infection with the rickettsia. Another study showed patients diagnosed with *R. rickettsii* but was found high titer of IgG antibody against *R. amblyommii* antigens and nor *R. rickettsia*. This study was in a high abundance tick *A. americanum*-infected area by *R. amblyommii* (APPERSON et al., 2008). But at the present no *R. amblyommii* has been isolated from humans. Phylogenetic studies show that *R. amblyommiii* is closely related to the others highly pathogenic rickettsias which highlight its importance (MERHEJ et al., 2014). *Rickettsia amblyommii* is found in *Amblyomma* genus ticks, specifically in *A. coelebs* (LABRUNA et al., 2004; PAROLA et al., 2007)*, A. sculptum* (former *A. cajannenese*) (LABRUNA et al., 2004), *A. neumanni* (LABRUNA et al., 2007), *A. longirostre*

(MCINTOSH et al., 2015), *A. geayi* (OGRZEWALSKA et al., 2010), *A. auricularium* (SARAIVA et al., 2013) in South America.

The *R. amblyommi* was first reported in Brazil from the western Amazon forest, in Rondônia state (LABRUNA et al., 2004). Within several *Amblyomma* ticks infected, the *A. cajennense* and *A. coelebs* had minimum infection of 26.8% (11/41) and 14.3% (1/7), respectively, with *R. amblyommii*, all of them showing 100% similarity. Other studies showed different geographic locations in Brazil with ticks infected by this rickettsia organism, for example, the states Pará (OGRZEWALSKA et al., 2010), Bahia (OGRZEWALSKA et al., 2011), Pernambuco (SARAIVA et al., 2013), São Paulo (OGRZEWALSKA et al., 2008), and Minas Gerais (NUNES et al., 2015). The latter was the first report of *A. sculptum* infected by *R. amblyommii* after the *A. cajennense* complex reassessment. In another study in French Guyana, 15.4% (2/13) of *A. coelebs* had 98% of similarity with *R. amblyommii*, suggesting the bacteria presence in this country and also proved by close phylogenetic relationship (PAROLA et al., 2007). In Argentina, *A. neumanni* was investigated for *Rickettsia* and found infected (23.6%, 13/55) by *Candidatus Rickettsia amblyommiii* and phylogenetically close related to others SFG rickettsia (LABRUNA et al., 2007). *Amblyomma neumanni* can be found parasitizing humans as well as domestic animals in Argentina which increase the importance of *R. amblyommii* pathogenicity investigation.

In Central America, Costa Rica, 10 of 15 *A. cajennense* ticks were infected by *R. amblyommii* (HUN et al., 2011). Also in domestic animals, in Panama, *R. amblyommii* DNA was found in *A. cajennense, D. nitens*, and *R. sanguineus*, correlated to seroreactive-positive horses and dogs (BERMÚDEZ et al., 2011).

In North America *R. amblyommii* is resident of many populations of Lone Star ticks, *A. americanum*. Eighty percent (80%) of *A. americanum* surveyed throughout the southern and eastern of USA had shown *R. amblyommii* (MIXSON et al., 2006). Other study showed that *A. maculatum* collected in Arkansas for rickettsiae demonstrated 26% of *R. amblyommii* infection (TROUT et al., 2010). Due it, was proposed that the mostly of seroreactivity to SFG antigens in the USA is caused by moderate infections with *R. amblyommii* (APPERSON et al., 2008). However, only *Amblyomma*-associated SFG *Rickettsia* confirmed is *R. parkeri* infecting humans.

As we can see, tick vector species and host distribution through America continent is vast. The advent of Next Generation Sequencing (NGS) and other technologies show a huge advent in *Rickettsia* life cycle knowledge as well as a complex network between the vectors and hosts. Some works involving vertebrates hosts infested with infected ticks, e.g., the capybara *Hydrochoerus hydrochaeris* parasitized with *A. sculptum* infected by *R. rickettsia* (former *A. cajennense*) showed the bacteria cycle in this hosts and ticks' susceptibility, all monitored by serum samples and PCRs (SOUZA et al., 2009). A RNA-seq experiment with ticks infected with *R. rickettsia* (MARTINS et al., 2017) and *R. amblyommii* (MOREIRA et al., 2016) could identify what sequences are differential expressed and their putative function within this relationship.

6. REFERENCES

ALARCON-CHAIDEZ, F. J.; SUN, J.; WIKEL, S. K. Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). **Insect Biochemistry and Molecular Biology**, v. 37, n. 1, p. 48–71, 2007.

ANATRIELLO, E. et al. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. **BMC Genomics**, v. 11, p. 450, 2010.

ANDERSON, J. M.; SONENSHINE, D. E.; VALENZUELA, J. G. Exploring the mialome of ticks: An annotated catalogue of midgut transcripts from the hard tick, *Dermacentor variabilis* (Acari: Ixodidae). **BMC Genomics**, v. 9, n. 1, p. 552, 2008.

APPERSON, C. S. et al. Tick-Borne Diseases in North Carolina: Is *"Rickettsia amblyommii"* a Possible Cause of Rickettsiosis Reported as Rocky Mountain Spotted Fever? **Vector-Borne and Zoonotic Diseases**, v. 8, n. 5, p. 597–606, out. 2008.

ATELLA, G. C. et al. Oogenesis and egg development in triatomines: A biochemical approach. **Anais da Academia Brasileira de Ciencias**, v. 77, n. 3, p. 405–430, set. 2005.

AUGOFF, K. et al. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. **Molecular Cancer**, v. 11, n. 1, p. 5, 2012.

BACHER, C. P. et al. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. **Nature Cell Biology**, v. 8, n. 3, p. 293–299, 24 mar. 2006.

BALASHOV, Y. S. Bloodsucking ticks (Ixodoidea) — vectors of diseases of man and animals. **Miscellaneous publications of the entomological society of America**, v. 8, p. 161–376., 1972.

BARCELOS, R. M. **Análise filogeográfica e filogenética do carrapato *Amblyomma cajennense* (Fabricius, 1787)**. [s.l: s.n.].

BARROS-BATTESTI, D.; ARZUA, M.; BECHARA, H. **Carrapatos de Importância Medico-Veterinaria da Região Neotropical: Um guia ilustrado para identificação de especiés**. [s.l.] Vox/ICTTD-3/Butantan, 2006.

BATISTA, I. F. C. et al. Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari: Ixodidae). **Toxicon**, v. 51, n. 5, p. 823–834, 2008.

BEATI, L. et al. *Amblyomma cajennense* (Fabricius, 1787) (Acari: Ixodidae), the Cayenne tick: phylogeography and evidence for allopatric speciation. **BMC Evolutionary Biology**, v. 13, n. 1, p. 267, 2013.

BERMÚDEZ, C. S. E. et al. Rickettsial infection in domestic mammals and their ectoparasites in El Valle de Antón, Coclé, Panamá. **Veterinary Parasitology**, v. 177, n. 1–2, p. 134–138, abr. 2011.

BÖHMDORFERR, G.; WIERZBICKI, A. T. **Control of Chromatin Structure by Long Noncoding RNATrends in Cell Biology**, out. 2015. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0962892415001269>. Acesso em: 5 jul. 2017

BROWN, C. J. et al. Localization of the X inactivation centre on the human X chromosome in Xq13. **Nature**, v. 349, n. 6304, p. 82–84, 3 jan. 1991.

BROWN, C. J. et al. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. **Cell**, v. 71, n. 3, p. 527–542, 30 out. 1992.

CHEN, B. et al. Genome-wide identification and developmental expression profiling

of long noncoding RNAs during *Drosophila metamorphosis*. **Scientific Reports**, v. 6, n. 1, p. 23330, 2016.
CLEMSON, C. M. et al. XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure. **Journal of Cell Biology**, v. 132, n. 3, p. 259–275, fev. 1996.
DASCH, G. A. et al. No Titl. **American Journal of Tropical Medicine and Hygiene**, v. 49, n. 3_Suppl, p. 3–262, 1993.
DE CASTRO, M. H. et al. De novo assembly and annotation of the salivary gland transcriptome of *Rhipicephalus appendiculatus* male and female ticks during blood feeding. **Ticks and Tick-borne Diseases**, v. 7, n. 4, p. 536–548, 2016.
DJEBALI, S. et al. Landscape of transcription in human cells. **Nature**, v. 489, n. 7414, p. 101–108, 2012.
DYKES, I.; EMANUELI, C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. **Genomics, Proteomics & Bioinformatics**, v. 15, n. 3, p. 177–186, 2017.
EGEKWU, N. et al. Transcriptome of the Female Synganglion of the Black-Legged Tick *Ixodes scapularis* (Acari: Ixodidae) with Comparison between Illumina and 454 Systems. **PLoS ONE**, v. 9, n. 7, p. e102667, 30 jul. 2014.
ELLISON, D. W. et al. Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. **Infection and Immunity**, v. 76, n. 2, p. 542–550, 2008.
ETEBARI, K. et al. Identification of *Aedes aegypti* Long Intergenic Non-coding RNAs and Their Association with *Wolbachia* and Dengue Virus Infection. **PLoS Neglected Tropical Diseases**, v. 10, n. 10, p. 1–18, 2016.
FELSHEIM, R. F.; KURTTI, T. J.; MUNDERLOH, U. G. Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: Identification of virulence factors. **PLoS ONE**, v. 4, n. 12, 2009.
FONSECA, F. DA; ARAGÃO, H. B. **Notas de ixodologia: VI - Descrição da fêmea de *Amblyomma multipunctum*, Neumann 1899 e redescrição do macho (Acari: Ixodídae) Memórias do Instituto Oswaldo Cruz**, 1953.
FRANCISCHETTI, I. M. B. et al. The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). **Insect Biochemistry and Molecular Biology**, v. 35, n. 10, p. 1142–1161, 2005.
FRANCISCHETTI, I. M. B. et al. An insight into the sialome of the soft tick, *Ornithodorus parkeri*. **Insect Biochemistry and Molecular Biology**, v. 38, n. 1, p. 1–21, 2008a.
FRANCISCHETTI, I. M. B. et al. An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. **Journal of Proteomics**, v. 71, n. 5, p. 493–512, 2008b.
FRANCISCHETTI, I. M. B. et al. An insight into the sialotranscriptome and proteome of the coarse bontlegged tick, *Hyalomma marginatum rufipes*. **Journal of Proteomics**, v. 74, n. 12, p. 2892–2908, 2011.
GARCIA, G. et al. The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. **Parasites & Vectors**, v. 7, n. 1, p. 430, 2014.
GUGLIELMONE A, MANGOLD A, O. B. **Ciclo de vida del *Amblyomma cajennense* (Fabricius, 1787) (Acari: Ixodidae) en condiciones de laboratorioRevista de Medicina Veterinaria**, 1992.
GULIA-NUSS, M. et al. Genomic insights into the Ixodes scapularis tick vector of

Lyme disease. **Nature Communications**, v. 7, n. May 2015, p. 10507, 2016.
HEEKIN, A. M. et al. Gut transcriptome of replete adult female cattle ticks, *Rhipicephalus (Boophilus) microplus*, feeding upon a *Babesia bovis*-infected bovine host. **Parasitology Research**, v. 112, n. 9, p. 3075–3090, 2013.
HUN, L. et al. First report of the isolation and molecular characterization of *Rickettsia amblyommii* and *Rickettsia felis* in Central America. **Vector Borne Zoonotic Dis**, v. 11, n. 10, p. 1395–1397, 2011.
IYER, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. **Nature Genetics**, v. 47, n. 3, p. 199–208, 2015.
JAYAKODI, M. et al. Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. **BMC Genomics**, v. 16, n. 1, p. 680, 2015.
JENKINS, A. M.; WATERHOUSE, R. M.; MUSKAVITCH, M. A. Long non-coding RNA discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. **BMC Genomics**, v. 16, n. 1, p. 337, 23 dez. 2015.
KAPRANOV, P. et al. RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. **Science**, v. 316, n. 5830, p. 1484–1488, 8 jun. 2007.
KARIM, S.; RIBEIRO, J. M. C. An insight into the sialome of the lone star tick, *Amblyomma americanum*, with a glimpse on its time dependent gene expression. **PLoS ONE**, v. 10, n. 7, 2015.
KARIM, S.; SINGH, P.; RIBEIRO, J. M. C. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. **PLoS ONE**, v. 6, n. 12, 2011.
KOTSYFAKIS, M. et al. Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. **Scientific Reports**, v. 5, n. 1, p. 9103, 2015.
LABRUNA, M. B. et al. *Rickettsia bellii* and *Rickettsia amblyommii* in *Amblyomma* ticks from the State of Rondônia, Western Amazon, Brazil. **Journal of medical entomology**, v. 41, n. 6, p. 1073–1081, 2004.
LABRUNA, M. B. et al. Infection by *Rickettsia bellii* and *Candidatus "Rickettsia amblyommii"* in *Amblyomma neumanni* ticks from Argentina. **Microbial Ecology**, v. 54, n. 1, p. 126–133, 2007.
LABRUNA, M. B. et al. Comparative susceptibility of larval stages of *Amblyomma aureolatum*, *Amblyomma cajennense*, and *Rhipicephalus sanguineus* to infection by *Rickettsia rickettsii*. **Journal of medical entomology**, v. 45, n. 6, p. 1156–1159, 2008.
LABRUNA, M. B. et al. Cross-mating experiments with geographically different populations of *Amblyomma cajennense* (Acari: Ixodidae). **Experimental and Applied Acarology**, v. 54, n. 1, p. 41–49, 1 maio 2011.
LEE, J.; DAVIDOW, L. S.; WARSHAWSKY, D. Tsix, a gene antisense to Xist at the X-inactivation centre. **Nature Genetics**, v. 21, n. 4, p. 400–404, 1 abr. 1999.
LEE, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. **Nature Reviews Molecular Cell Biology**, v. 12, n. 12, p. 815–826, 23 nov. 2011.
LEWIS, L. A. et al. Identification of 24h Ixodes scapularis immunogenic tick saliva proteins. **Ticks and Tick-borne Diseases**, v. 6, n. 3, p. 424–434, 2015.
LI, M. et al. The novel long non-coding RNA CRG regulates *Drosophila* locomotor behavior. **Nucleic Acids Research**, v. 40, n. 22, p. 11714–11727, dez. 2012.
LOGULLO, C. et al. Binding and storage of heme by vitellin from the cattle tick,

Boophilus microplus. **Insect Biochemistry and Molecular Biology**, v. 32, n. 12, p. 1805–1811, 2002.
MANS, B. J. et al. Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behavior. **Insect Biochemistry and Molecular Biology**, v. 38, n. 1, p. 42–58, 2008.
MARDIS, E. R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, v. 9, n. 1, p. 387–402, set. 2008.
MARTINS, L. A. et al. The Distinct Transcriptional Response of the Midgut of *Amblyomma sculptum* and *Amblyomma aureolatum* Ticks to *Rickettsia rickettsii* Correlates to Their Differences in Susceptibility to Infection. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. April, p. 129, 28 abr. 2017.
MASTROPAOLO, M. et al. Biological differences between two allopatric populations of *Amblyomma cajennense* (Acari: Ixodidae) in Argentina. **Experimental and Applied Acarology**, v. 53, n. 4, p. 371–375, 2011.
MCINTOSH, D. et al. Detection of *Rickettsia bellii* and *Rickettsia amblyommii* in *Amblyomma longirostre* (Acari: Ixodidae) from Bahia state, Northeast Brazil. **Brazilian Journal of Microbiology**, v. 46, n. 3, p. 879–883, 2015.
MERHEJ, V. et al. Genotyping, evolution and epidemiological findings of Rickettsia species. **Infection, Genetics and Evolution**, v. 25, p. 122–137, 2014.
MIXSON, T. R. et al. Prevalence of *Ehrlichia*, *Borrelia*, and *Rickettsial agents* in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states. **Journal of medical entomology**, v. 43, n. 6, p. 1261–1268, 1 nov. 2006.
MOREIRA, H. N. S. et al. A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of the Amblyomma sculptum tick. **Parasitology International**, v. 66, n. 2, p. 64–73, 2016.
NAKAJIMA, C. et al. Random Sequencing of cDNA Library Derived from Partially-Fed Adult Female *Haemaphysalis longicornis* Salivary Gland. **Journal of Veterinary Medical Science**, v. 67, n. 11, p. 1127–1131, 2005.
NAVA, S. et al. The ticks (Acari: Ixodida: Argasidae, Ixodidae) of Paraguay. **Annals of Tropical Medicine & Parasitology**, v. 101, n. 3, p. 255–270, 2007.
NAVA, S. et al. Reassessment of the taxonomic status of *Amblyomma cajennense* (Fabricius, 1787) with the description of three new species, *Amblyomma tonelliae* n. sp., *Amblyomma interandinum* n. sp. and *Amblyomma patinoi* n. sp., and reinstatement. **Ticks and Tick-borne Diseases**, v. 5, n. 3, p. 252–276, 2014.
NUNES, E. D. C. et al. *Rickettsia amblyommii* infecting *Amblyomma sculptum* in endemic spotted fever area from southeastern Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 8, p. 1058–1061, 2015.
OGAWA, Y.; SUN, B. K.; LEE, J. T. Intersection of the RNA Interference and X-Inactivation Pathways. **Science**, v. 320, n. 5881, p. 1336–1341, 6 jun. 2008.
OGRZEWALSKA, M. et al. Ticks (Acari: Ixodidae) infesting wild birds in an Atlantic forest area in the state of São Paulo, Brazil, with isolation of rickettsia from the tick Amblyomma longirostre. **Journal of medical entomology**, v. 45, n. 4, p. 770–774, 2008.
OGRZEWALSKA, M.; UEZU, A.; LABRUNA, M. B. Ticks (Acari: Ixodidae) infesting wild birds in the Atlantic Forest in northeastern Brazil, with notes on rickettsial infection in ticks. **Parasitology Research**, v. 108, n. 3, p. 665–670, 2010.
OGRZEWALSKA, M.; UEZU, A.; LABRUNA, M. B. Ticks (Acari: Ixodidae) infesting wild birds in the Atlantic Forest in northeastern Brazil, with notes on rickettsial infection in ticks. **Parasitology Research**, v. 108, n. 3, p. 665–670, 2011.

OLEAGA, A. et al. Ticks and Tick-borne Diseases Functional annotation and analysis of the Ornithodoros moubata midgut genes di ff erentially expressed after blood feeding. **Ticks and Tick-borne Diseases**, n. November 2016, p. 0–1, 2017.
PAROLA, P. et al. A tick-borne rickettsia of the spotted-fever group, similar to *Rickettsia amblyommii*, in French Guyana. **Annals of Tropical Medicine and Parasitology**, v. 101, n. 2, p. 185–188, 2007.
RIBEIRO, J. M. et al. A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. **BMC Genomics**, v. 12, n. 1, p. 136, 2011.
RIBEIRO, J. M. C. et al. The sialotranscriptome of *Antricola delacruzi* female ticks is compatible with non-hematophagous behavior and an alternative source of food. **Insect Biochemistry and Molecular Biology**, v. 42, n. 5, p. 332–342, 2012.
SADO, T.; HOKI, Y.; SASAKI, H. Tsix silences Xist through modification of chromatin structure. **Developmental Cell**, v. 9, n. 1, p. 159–165, jul. 2005.
SARAIVA, D. G. et al. *Rickettsia amblyommii* Infecting *Amblyomma auricularium* Ticks in Pernambuco, Northeastern Brazil: Isolation, Transovarial Transmission, and Transstadial Perpetuation. **Vector-Borne and Zoonotic Diseases**, v. 13, n. 9, p. 615–618, 2013.
SCHWARZ, A. et al. A Systems Level Analysis Reveals Transcriptomic and Proteomic Complexity in *Ixodes Ricinus* Midgut and Salivary Glands During Early Attachment and Feeding. **Molecular & Cellular Proteomics**, v. 13, n. 10, p. 2725–2735, 2014.
SEIXAS, A. et al. Endocrinology and control of tick vitellogenesis. **Acta Scientiae Veterinariae**, v. 38, n. 2, p. 95–111, 2010.
SONENSHINE, D. E.; ROE, R. M. **Biology of Ticks: Volume 1**. 2nd. ed. [s.l.] Oxford University Press, 2013.
SOSHNEV, A. A. et al. A conserved long noncoding RNA affects sleep behavior in *Drosophila*. **Genetics**, v. 189, n. 2, p. 455–68, out. 2011.
SOUZA, C. E. et al. Experimental infection of capybaras *Hydrochoerus hydrochaeris* by *Rickettsia rickettsii* and evaluation of the transmission of the infection to ticks *Amblyomma cajennense*. **Veterinary Parasitology**, v. 161, n. 1–2, p. 116–121, abr. 2009.
ST.LAURENT, G.; WAHLESTEDT, C.; KAPRANOV, P. The Landscape of long noncoding RNA classification. **Trends in Genetics**, v. 31, n. 5, p. 249–251, 2015.
STUTZER, C. et al. Gene expression profiling of adult female tissues in feeding *Rhipicephalus microplus* cattle ticks. **International Journal for Parasitology**, v. 43, n. 7, p. 541–554, 2013.
SUN, B. K.; DEATON, A. M.; LEE, J. T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. **Molecular Cell**, v. 21, n. 5, p. 617–628, 3 mar. 2006.
TAN, A. W. L. et al. Sexual differences in the sialomes of the zebra tick, *Rhipicephalus pulchellus*. **Journal of Proteomics**, v. 117, p. 120–144, 2015.
TROUT, R. et al. Rickettsiae in gulf coast ticks, Arkansas, USA. **Emerging Infectious Diseases**, v. 16, n. 5, p. 830–832, maio 2010.
VAN HEESCH, S. et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. **Genome Biology**, v. 15, n. 1, p. R6, 7 jan. 2014.
WANG, K. C.; CHANG, H. Y. Molecular Mechanisms of Long Noncoding RNAs. **Molecular Cell**, v. 43, n. 6, p. 904–914, 16 set. 2011.
WU, Y. et al. Systematic identification and characterization of long non-coding RNAs

in the silkworm, *Bombyx mori*. **PLoS ONE**, v. 11, n. 1, p. 1–25, 2016.
XU, N. Transient Homologous Chromosome Pairing Marks the Onset of X Inactivation. **Science**, v. 311, n. 5764, p. 1149–1152, 24 fev. 2006.
XU, X. L. et al. De novo assembly and analysis of midgut transcriptome of *Haemaphysalis flava* and identification of genes involved in blood digestion, feeding and defending from pathogens. **Infection, Genetics and Evolution**, v. 38, p. 62–72, 2016.
ZHAO, J. et al. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. **Science**, v. 322, n. 5902, p. 750–756, 31 out. 2008.
ZHAO, J. et al. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. **Molecular Cell**, v. 40, n. 6, p. 939–953, 22 dez. 2010.

CHAPTER 2: Is *Rickettsia amblyommii* able to regulate long non-coding RNA expression in *Amblyomma sculptum* tick? An *in silico* approach

1. INTRODUCTION

As time goes by, new functions of long non-coding RNAs (lncRNA) are discovered. They belong to non-coding RNA (ncRNA) group and their roles mostly are unknown, despite some characteristics are available such as promoter conservation, splice junctions, exons, predicted structures, genome, association with chromatin specific structures which indicate active transcription; molecular keys signal and transcription factors; dynamic expression and alternative splicing during differentiation; tissue- and cell-specific expressions patterns and cellular localization; and altered expression or splicing pattern in cancer and other diseases (MATTICK, 2009).

In the Arthropod phylum, a few previous studies have already shown the existence and function of long non-coding RNAs in some insects like *Drosophila*, *Apis mellifera*, *Bombyx mori* and *Anopheles gambiae* (JAYAKODI et al., 2015; JENKINS et al., 2015). The functions are diversified, with functions such as heat shock response, sleep regulation with circadian cycle relationship and the locomotion of *Drosophila* flies. (DENG et al., 2006; SOSHNEV et al., 2011).

Non-coding RNAs have been identified in tick transcriptomes (GULIA-NUSS et al., 2016; MOREIRA et al., 2016) but not characterized. The *Amblyomma sculptum* tick species belongs to *Amblyomma cajennense* complex (NAVA et al., 2014), and is the main vector taxon of transmission for Brazilian Spotted Fever whose etiological agent is *Rickettsia rickettsii.* In seeking better understanding of this new complex, and the roles of lncRNAs modulation of *Rickettsia*-tick interaction, more studies are needed on the influence level of *Rickettsia* species with non-coding RNAs in the tick's biological cycle. This knowledge could guide the development of new strategies to eliminate these ectoparasites as well as the discovery of new molecules acting in *A. sculptum-Rickettsia* interaction. Thus, the main goal of the present work is identify

putative lncRNAs in the transcriptomes of *A. sculptum* and analyze their differential expression in different tissues infected by *Rickettsia amblyommii*.

2. METHODS

To identify the lncRNAs in *A. sculptum*, we downloaded the sequence data of nine transcriptomes available in Short Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). These transcriptomes were sequenced by Garcia et al. (2015) from salivary glands (SRA accession numbers SRR1201285, SRR1201288, SRR1201289 and SRR1201290, Roche 454 data) and by Moreira et al. (2017) from midgut (SRR3176713, Ion Torrent PGM; and SRR3120106, Illumina HiSeq) and ovaries (SRR3126275, Illumina HiSeq) of *A. sculptum*. In addition, we also analyzed the transcriptomes of midgut (SRR3198618) and ovaries (SRR3198621) infected with *Rickettsia amblyommii* (Illumina HiSeq data) (Moreira et. al, 2017). The libraries are from adult and females of *A. sculptum* fed in rabbits.

The reads were analyzed for quality in FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) (quality score threshold: Q20) and adapter removal using Trimmomatic version 0.36 (Bolger et al. 2014) for Illumina HiSeq 2000 data, Trim Galore! version 0.4.4 (https://github.com/FelixKrueger/TrimGalore) for Ion Torrent PGM data, and PRINSEQ version 0.20.4 (Schmieder and Edwards 2011) for Roche 454 data.

After trimming, the reads were de novo assembled using CLC Genomics Workbench version 8.5.3 (Qiagen) (parameters: Kmer size=23; bubble size=5,000) and Trinity version 2.4.0 (Grabherr et al. 2013) (default parameters). Then, the following pipeline was applied to analyze the transcript contigs and find putative lncRNAs:

1. *Size selection.* lncRNAs have more than 200nt in sequence length and contain ORFs with maximum size of 300nt (100aa). This filtering step were performed using Galaxy Tool (Blankenberg et al. 2007) and all the six frames were used to search the ORFs;
2. *Removal of partial or potential coding sequences*. To eliminate potential coding sequences, the Coding Potential Calculator (CPC) (Kong et al. 2007) was used as primary filter. Then, similarities searches against proteins of Acari species (taxonomy identifier: 6933) available at UniProt Knowledgebase (UniProtKB) using BLASTx tool of BLAST software (e-value threshold=10^{-5})

(Camacho et al. 2009) was performed. An additional in-house script was used to filter only unmatched sequences in these analyzes;

3. *Removal of mitochondrial, rRNA and tRNA sequences.* The filtered sequences were subjected to similarity searches against the mitochondrial genome of *Amblyomma sculptum* (Lima et al., 2017, submitted) and the databases of rRNA (SILVA database, https://www.arb-silva.de/) and tRNA (GtRNAdb, http://gtrnadb.ucsc.edu/) sequences by BLASTn tool to eliminate any protein sequence as well as ribosomal sequences (e-value threshold=10^{-5});
4. *Removal of sequences with signatures of coding sequences.* For the last steps and to eliminate any leftover of protein sequences, the filtered sequences were subjected to profile searches using against the Protein Families database (Pfam) and Conserved Domains Database (CDD) (Finn et al. 2016) using RPSBLAST tool (e-value threshold=10^{-5});
5. *Non-coding annotation.* After these filtering steps, the putative lncRNAs sequences were annotated by similarities searches against NONCODE (http://www.noncode.org/), RNA Central (http://rnacentral.org/) and RefSeq-RNA databases from NCBI using BLASTn tool (e-value threshold=10^{-5}). The selected databases were due downloadable content as we have a bunch of sequences. The NONCODE and RNA Central are only for non-coding RNA while RefSeq RNA only curated RNA sequences and bigger database for annotation;
6. *Differential expression analysis of the identified putative lncRNAs.* The differential expression analysis was performed using EdgeR (Robinson et al. 2009) (a part of Bioconductor R package) and using Trinity scripts (https://github.com/trinityrnaseq/trinityrnaseq/wiki/Trinity-Differential-Expression) and using and from The Biostar Handbook (Albert 2017) we generated heatmap and differential expression to visualize most expressed transcripts from both assemblies. We choose the p-value cut off of 0.05;

The Figure 3 shows the method workflow to reach the putative lncRNAs sequences.

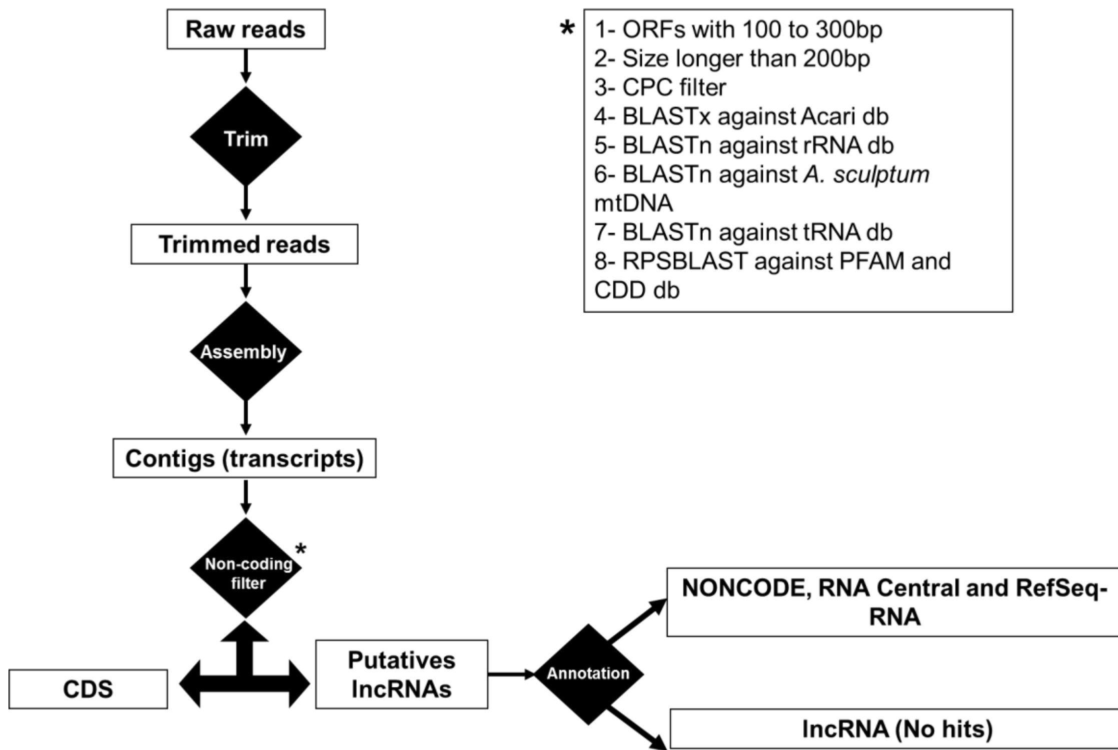


Figure 3. Method workflow showing how to reach the putatives lncRNAs. The asterisk shows the filter steps to eliminate any coding sequence.

3. RESULTS

We assembled a total of 128,410 and 77,998 contigs for Trinity and CLC Genomics, respectively. After all filtering steps, we obtained 31,188 and 23,733 putative lncRNAs from Trinity and CLC Genomics analysis, respectively. For Trinity assembly, the annotation against NONCODE database identified 118 sequence matches while from RefSeq RNA and RNA Central found 560 and 133 matches, respectively. For CLC, we found 57, 533 and 161 matches for NONCODE, RefSeq RNA and RNA Central databases, respectively. Figure 4 shows the filter workflow with the numbers of contigs filtered for each step. Supplemental Table 1 shows all NONCODE transcripts annotation while Supplemental Table 2 and Supplemental Table 3 are for RefSeq RNA and RNA Central annotations, respectively.

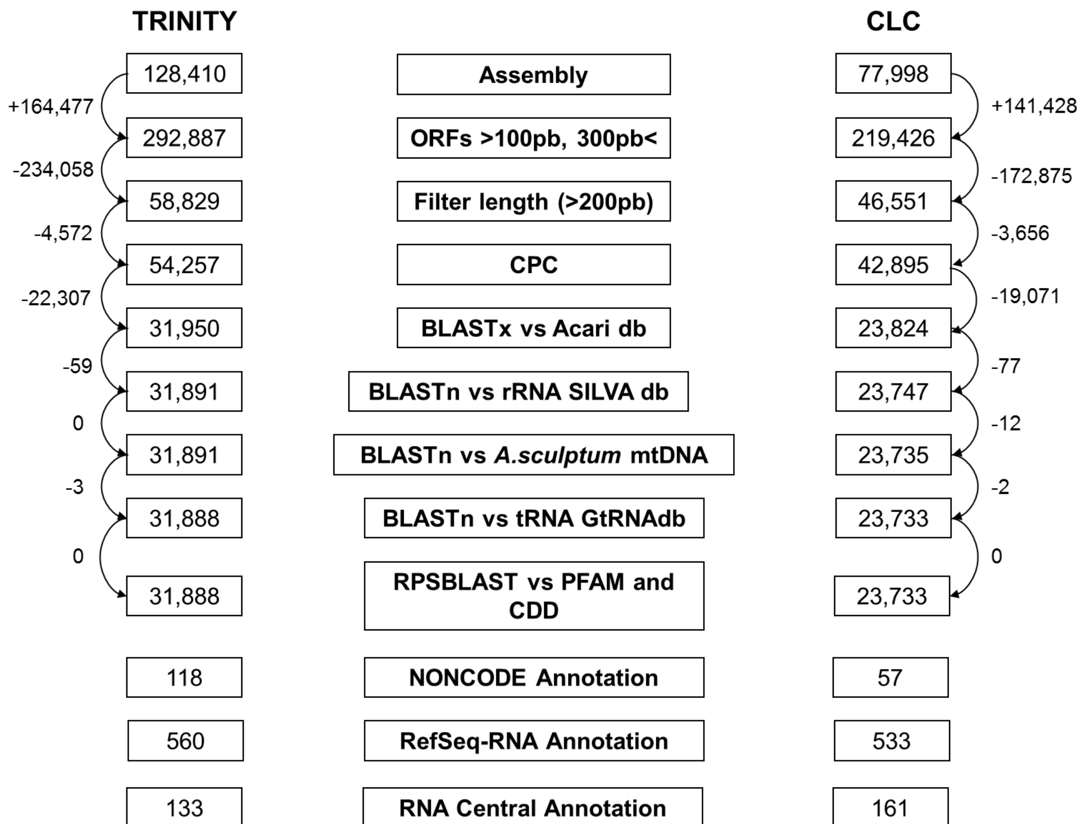


Figure 4. Filter workflow to select lncRNAs. Two assembly software were tested: Trinity and CLC Genomics. The numbers inside the boxes represent how many transcripts left after each filter while the numbers next to the arrows show how many transcripts increased or decreased.

In both tissues the differential expression results showed higher upregulated sequences than downregulated (Table 1).

Table 1. Number of sequences upregulated and downregulated infected and non-infected tissues

Tissue comparison	Number of sequences		TOTAL
Upregulated	**Trinity**	**CLC**	-
Infected midgut	938	581	1,519
Infected ovary	841	481	1,322
TOTAL	1,779	1,062	2,841
Downregulated			
Infected midgut	764	447	1,211
Infected ovary	597	236	833
TOTAL	**1,361**	**683**	**2,044**

Top species hits from Trinity, CLC and RNA Central annotations are showed in the Table 2

Table 2. The top species hits from Trinity and CLC annotations. Numbers indicate the quantity of hits found.

Species	Organism	Trinity	CLC	Total
Brassica rapa	Plant	134	115	249
Oryctolagus cuniculus	Rabbit	139	339	359
Brassica napus	Plant	80	61	141
Homo sapiens	Human	137	68	96
Ixodes scapularis	Tick	53	25	68
Gorilla gorilla gorilla	Gorilla	20	7	20
Pan troglodytes	Chimpanzee	10	4	14

The resulting heatmaps show both the up and downregulated contigs from the midgut and ovaries from the Trinity (Figure 5) and CLC Genomics (Figure 6) assemblers. We observed that both organs had similar differential expression patterns. The up and downregulated sequences from midgut and ovary are indicated in Table 3. We identified multiple putative lncRNAs, but only a few of the sequences have been annotated due to a lack of data in the tick non-coding databases. Therefore, our discussion will focus in only the annotated sequences because the functions have not been assigned and we don’t know the functions of the annotated

contigs. Also, as we desire to know whether *R. amblyomii* is capable of inducing lncRNA differential expression, only infected tissues over non-infected will be discussed, not the vice-versa.

In general, both assembly strategies showed similar annotations with some particular putative lncRNAs from each one. From our comparative analysis of upregulated infected midgut sequences over midgut non-infected we can highlight the cleavage and polyadenylation specificity factor (trinity_6256), cuticle protein: putative (trinity_52197), Misexpression suppressor of KSR: putative (trinity_24167), multifunctional chaperone: putative (trinity_32938), RNA-binding protein (trinity_49811) and shock protein 90 alpha family class B member 1 (hsp90ab1) (trinity_28498), peroxidase 42 (clc_24629) and secreted salivary gland peptide: putative (clc_26358) as being prominent sequences in the upregulated organs..

The downregulated contigs we could highlight were protein lateral stalk subunit P2 (RPLP2) (trinity_27204), 3'(2'), 5-bisphosphate nucleotidase 1 (BPNT1): transcript variant X3 (trinity_24155), peroxisomal biogenesis factor 5-like (PEX5L) (trinity_33412) and cytochrome P450: family 8: subfamily B: polypeptide 1 (CYP8B1) (clc_44720).

In ovaries, no annotated upregulated sequences were found despite upregulation of 100 contigs. Only one annotated sequence was found to be downregulated from Trinity assemblage, an uncharacterized mRNA (trinity_34676). All sequences can be found in Supplemental Tables 4, 5, 6 and 7.

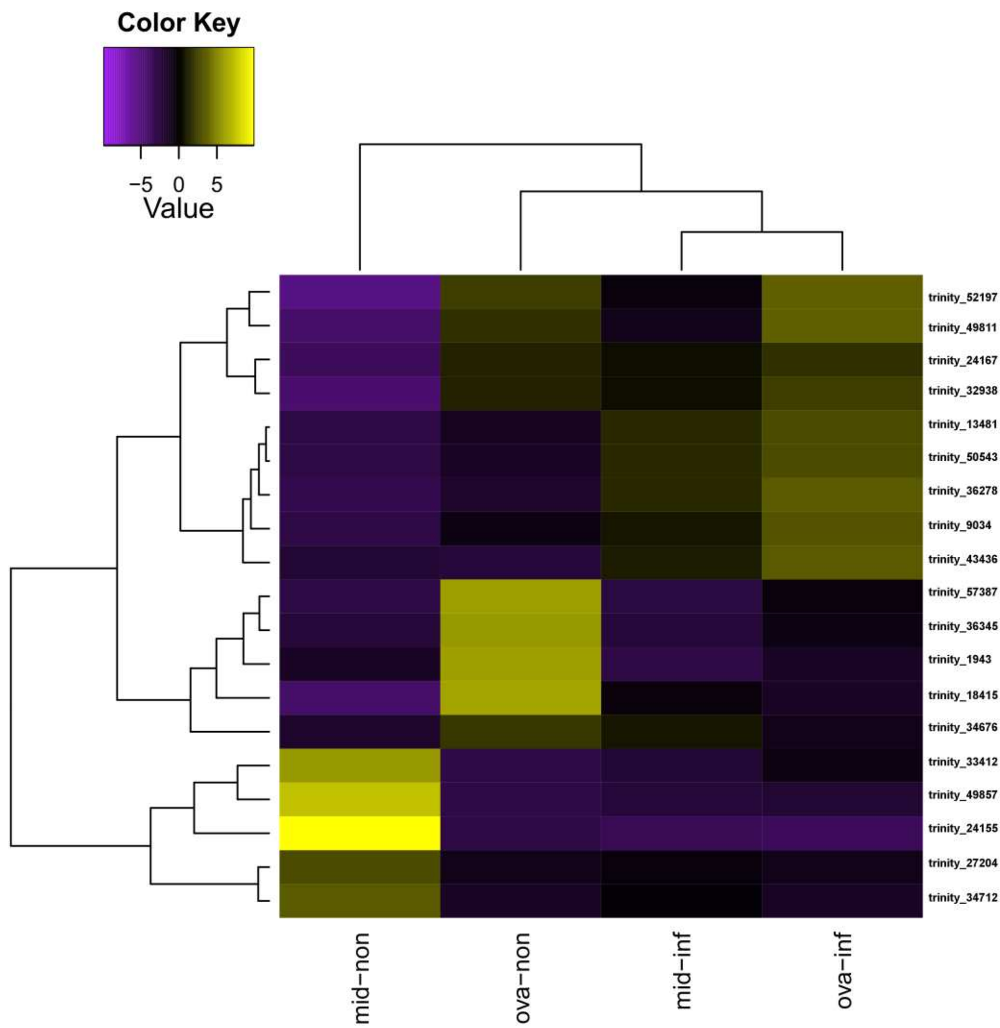


Figure 5. Heatmap of some putatives lncRNAs up and downregulated from Trinity assembly. The numbers on the right side are the sequences numbers in the Table 3. All putative lncRNA are located in supplemental tables.

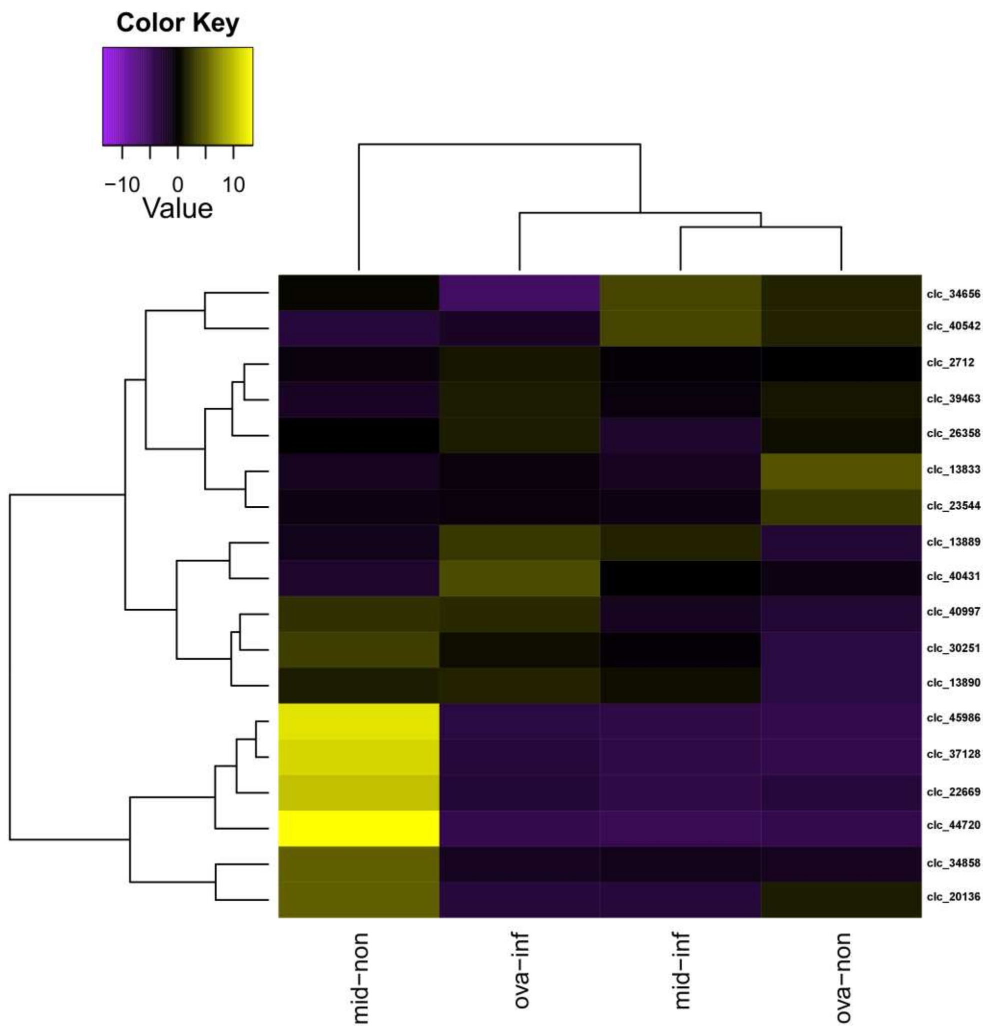


Figure 6. Heatmap of some lncRNAs up and downregulated from CLC Genomics assembly. The numbers on the right side are the sequences numbers in the Table 3. All putative lncRNA are located in supplemental tables.

Despite hundreds of sequences annotated, we still had few compared database's size of other arthropods, which indicate the poor tick`s database. It is worth of note that all these sequences are putative and the names might change in future since the contigs herein have lncRNA characteristics. For example, the contig named trinity_27204 was annotated as Ribosomal protein lateral stalk subunit P2 (RPLP2): mRNA, while the same contig from RNA Central database showed a long non-coding RNA.

Table 3. Annotation of putative lncRNAs upregulated and downregulated in midgut and an unique annotated downregulated ovary infected sequence.

Query **Upregulated**	logFC	RefSeq RNA	RNA Central	NONCODE
trinity_6256	3,18	*Ixodes scapularis* cleavage and polyadenylation specificity factor: putative: mRNA	-	-
trinity_52197	5,36	*Ixodes scapularis* cuticle protein: putative: mRNA	-	-
trinity_24167	5,35	*Ixodes scapularis* Misexpression suppressor of KSR: putative: mRNA	-	-
trinity_32938	5,97	*Ixodes scapularis* multifunctional chaperone: putative: mRNA	-	-
trinity_49811	4,21	*Ixodes scapularis* RNA-binding protein: putative: mRNA	-	-
trinity_28498	4,42	*Larimichthys crocea* heat shock protein 90 alpha family class B member 1 (hsp90ab1): mRNA	-	-
clc_24629	3,98	PREDICTED: *Brassica napus* peroxidase 42 (LOC106435511): mRNA	-	-
clc_26358	2,93	*Ixodes scapularis* secreted salivary gland peptide: putative: mRNA	-	-
Downregulated				
trinity_27204	-4,30	*Homo sapiens* ribosomal protein lateral stalk subunit P2 (RPLP2): mRNA	*Homo sapiens* long non-coding RNA NONHSAT017332.2	NONHSAT017332.2
trinity_24155	-14,01	PREDICTED: *Oryctolagus cuniculus* 3'(2'), 5-bisphosphate nucleotidase 1 (BPNT1): transcript variant X3: mRNA	-	-
trinity_33412	-7,18	PREDICTED: *Oryctolagus cuniculus* peroxisomal biogenesis factor 5-like (PEX5L): mRNA	-	-
trinity_49857	-10,43	PREDICTED: *Oryctolagus cuniculus* uncharacterized LOC108178806 (LOC108178806): ncRNA	-	-
trinity_34712	-4,07	PREDICTED: Pan troglodytes uncharacterized LOC107972887 (LOC107972887): ncRNA	*Homo sapiens* long non-coding RNA NONHSAT000090.2	NONHSAT222393.1
clc_44720	-14,97	*Oryctolagus cuniculus* cytochrome P450: family 8: subfamily B: polypeptide 1 (CYP8B1): mRNA	*Oryctolagus cuniculus* U2 spliceosomal RNA	-
Ovary downregulated trinity_34676	-4,38	PREDICTED: *Brassica oleracea* var. *oleracea* uncharacterized LOC106298029 (LOC106298029): mRNA	-	-

Legend 1. logFC: log2 of fold change; NONCODE db prefix: NONHSAT: Non-coding from Homo sapiens.

4. DISCUSSION

It is currently well known that long non-coding RNAs have important roles inside the cells and that “junk DNA” is becoming forgotten (NIAZI et al., 2012; PALAZZO et al., 2015). The research in this field is increasingly growing and new lncRNA functions have been discovered as well as a subsequent increase in database size. Despite of the lack of a specific name or function for multiple sequences annotations, we could observe that *R. amblyommii* was able to change the gene expressions inside the tick *A. sculptum.* Table 1 shows that infected tissues have an increased number of upregulated sequences when compared to sequences from non-infected tissues.

4.1. Midgut upregulation

Infected midguts had more upregulated sequences than downregulated ones. From our analysis using both assembly strategies for midgut upregulated contigs, we observed the following lncRNAs which are involved in transcription and alternative splicing regulations: misexpression suppressor of KSR (MESK gene) (trinity_24167), a known gene that participate in RAS1/MAPK pathway signaling in *Drosophila* and one of the most important cell signaling pathways in metazoa (HUANG et al., 2000); a lncRNA transcript variant X2; cleavage and polyadenylation specificity factor: putative (trinity_6256); and RNA-binding protein: putative (trinity_49811). We believe for sequences from *Homo sapiens* (humans) annotations, like the ribosomal protein lateral stalk subunit P2 (RPLP2) (trinity_27204); or that from other organisms like *Brassica oleracea* (cabbage) and *Oryctolagus cuniculus* (rabbit), two hypothesis may explain the observed results: (1) once the ticks fed on rabbits, some rabbit’s lncRNA can be extracted and isolated by our in-house script. We observed a high number of putative lncRNA from *Oryctolagus cuniculus* in Table 2; (2) the poor tick annotated non-coding RNA database which can be explained by low annotated sequences number from ticks species (Table 2).

Our analysis has also identified additional contigs that are important during rickettsial infection. For example, heat-shock proteins (heat shock protein 90 alpha family class B member 1 (hsp90ab1), trinity_28498 sequence) usually are linked to cell stress (SCHLESINGER, 1990). When ticks are suffering rickettsial infection, their cells respond to the pathogens with heat-shock proteins i

ncreasing as observed with *Anaplasma phagocytophilum* infection in *I. scapularis* (BUSBY et al., 2012). The upregulation of peroxidase (clc_24629) suggests oxidative stress inside the tick's cells. It has been shown that reactive oxygen species (ROS) in insect cells are an immune response against pathogens like *Wolbachia* (ZUG et al., 2015). As infection by *R. amblyommii* upregulates several putative lncRNAs with relationships with peroxidase, it suggests the bacteria somehow is trying to evade the tick immune system through neutralization of the ROS generated by the host. Similar results were seen during studies with *A. sculptum* infected by *R. rickettsia* with the upregulation of enzymes involved with ROS production in midgut during infection (MARTINS et al., 2017).

Chaperones are known to correctly fold new translated proteins (BRUCE ALBERTS et al., 2016). We suspect that *R. amblyomii* induces protein folding in *A. sculptum* by stimulating upregulation of a putative multifunctional chaperone (trinity_32938). In a temperature shift experiment with *R. rickettsia* infecting *A. aureolatum* males, the co-chaperone HscB was induced by increase of temperature which shows somehow *Rickettsia* bacteria is able to modulate chaperones (GALLETTI et al., 2016). Additionally, an in depth research is needed to identify what proteins are suffering modulating and delineate the mechanism and identify the lncRNAs that are involved in this process. In humans, lncRNAs are involved in targeting HJURP/CENP-A chaperone complex to the centromeric area in chromosomes (QUÉNET et al., 2014). The cuticle protein (trinity_52197) is suspected is part of exoskeleton synthesis and its extension during blood feeding since the female has approximately 100-fold increase in size (SONENSHINE et al., 2013). Its upregulation need to be studied once *A. sculptum* and *A. aureolatum* showed cuticle protein downregulation by *R. rickettsia* infection (MARTINS et al., 2017). Also, it seems that lncRNA can be co-expressed with protein coding genes involved in stressful conditions like insecticide resistance in moth *Plutella xylostella* (L.) (ETEBARI et al., 2015). As with every pathogen transmitted by ticks, they have to

surpass the salivary gland barrier to be able to be transmitted to vertebrate hosts. The upregulation of secreted salivary gland peptide (putative clc_26358) suggests upregulation of lncRNAs involved with the immune system response of the tick against the *Rickettsia* or potentially the ability of salivary glands proteins to facilitate pathogen invasions as observed with *Borrelia burgdorferi* (HAJDUŠEK et al., 2013) and *A. phagocytophilum* (SULTANA et al., 2010). For example, these bacteria interact with salivary gland proteins Salp15 and SALP16, respectively.

4.2. Midgut downregulation

Despite a higher upregulated contigs than downregulated, we still had a plethora of sequences belonging to midgut downregulated sequences. Within them we can find ribosomal proteins like ribosomal protein lateral stalk subunit P2 (RPLP2) (trinity_27204) and 3'(2'), 5-bisphosphate nucleotidase 1 (BPNT1): transcript variant X3 (trinity_24155). Both sequences are present in the nucleotide machinery as the former in transcription and the latter in nucleotide metabolism. In other words, the *R. amblyommii* is downregulating transcription machinery and nucleotide metabolism inside the tick. The peroxisomal biogenesis factor 5-like (PEX5L) (trinity_33412) is responsible for importation of molecules to the peroxisome organelle to degrade ROS (KUNZE et al., 2015). As mentioned above, the upregulation of peroxidase degrades hydrogen peroxide and assists *R. amblyommii* in its attempts to evade the tick's immune system. However, in conjunction with the downregulation of PEX5L, *A. sculptum* seems to combat the bacteria by reducing the degradation of ROS, which is important to pathogen death.

Finally, the cytochrome P450 contig is part of a multifunctional protein family in arthropods (GULIA-NUSS et al., 2016; MARTINS et al., 2017). It is integral in the binding of juvenile hormone in insects and also in the detoxification of multiple xenobiotic factors. Also can be found related to structure and function of *Drosophila* sensory organs (WILLINGHAM et al., 2004), and pesticide metabolism as well as hormonal/chemosensory in *Metaseiulus occidentalis* (WU et al., 2016). We found cytochrome P450 family 8: subfamily B: polypeptide 1 (CYP8B1) (clc_44720)

downregulated in *A. sculptum*. Interestingly, Martins et al. (2017) found the same gene upregulated in *A. sculptum* and *A. aureolatum* infected by *R. rickettsia*.

4.3. Ovary differential expression

Only one sequence could be annotated for infected ovaries, an uncharacterized mRNA from *Brassica oleracea* var. *oleracea* species.

Although our study was able to bring some clarity to the process, we only focused our study on the annotated lncRNA sequences in the databases. As more contigs are annotated and their actual functions discovered, an in depth analysis can be conducted. In spite of the fact that several research groups have focused on coding proteins sequences, there has been little attention given to non-coding RNAs (GARCIA et al., 2014; KOTSYFAKIS et al., 2015; MARTINS et al., 2017; MOREIRA et al., 2016).

Despite different annotated sequences numbers, we found no more than 560 putative lncRNA in our work and more than 20,000 unknown, putative long non-coding RNAs without annotations (Supplemental tables). This suggests that the database of lncRNA for Arthropods phylum is underdeveloped as our analysis had more matches from mammals than ticks or other invertebrates. In addition, “non-traditional” model organisms like ticks have poor ncRNAs databases. The low number of annotated sequences and some matches between annotations and up and downregulated transcripts emphasizes the need of additional research of long non-coding RNAs in tick biology. The high percentage of non-coding space in human genome (around 50-70%) (DJEBALI et al., 2012; IYER et al., 2015) and the plethora of lncRNAs’ functions and variations (ST.LAURENT et al., 2015), show us an unknown world of non-coding RNAs yet to be discovered. It is very important from now on make detailed analysis such as qPCR of the up-regulated transcripts and analysis of heterologous expression to determine the likelihood of the assemblies and annotations. Also, the rise of bifunctionality of lncRNAs will uncover a new field of study since not only mRNA can be translated to proteins but also ncRNAs.

5. CONCLUSIONS

The non-coding RNA world represents a promising field for future study because of the diversified role of these molecules inside the cells. The interface between ticks and *Rickettsia* species is always important to understand the pathways, molecules and their functions within this interaction. Our analysis identified multiple upregulated putative lncRNAs in infected midguts and ovaries. This suggests that *R. amblyommii* is able to regulate the lncRNAs inside the tick *A. sculptum* by modulation of a diversified lncRNA with distinct functions like transcription controlling, interacting with ribosomes and *trans*-acting factors, ROS metabolism, all of these are important for bacterial cell invasion.

In this project, we identified more than 20,000 sequences with long non-coding RNA characteristics which add to our current knowledge of lncRNA. Unfortunately, fewer than 600 of the sequences are annotated which highlights the importance of studies. Further experiments need to be conducted to delineate the functions of these lncRNAs and their relationships with the molecules that they encode. There is also a need to increase the annotated ncRNA database for ixodid ticks.

6. PERSPECTIVES

The advent of new technologies, as NGS, resulted new biological findings each research published. New discussions are risen and the molecular biology dogmas have been broken and being in the science spotlight. One of them, that transcription of mRNA will produce a protein has been on firing line. The new ncRNA functions has been risen years ago and they are showing that cell biology is much more complicated and fascinating than imagined before. Maybe a change in gene concepts might change? The transcripts nor genes are operational units from genome? The progression in evolution is becoming more and more complex and molecular studies are even more important than before.

The long non-coding RNAs discussed herein are on the spotlight nowadays due their diverse functions and specificity for each organism type. As we could observe, the annotation is the beginning and the steps of how they work *in vivo* is current happening in new researches. It has been proven that the lncRNAs are involved in many cell pathways and they can control them or be controlled. The protein coding genes characteristics has been put aside and *cis*-elements importance are now increased. The "junk DNA" was left behind and now its importance is equal to proteins. These vast majority of genomic space still is obscure and unexplored. The "omics" studies are helping to discover this unknown world and specially sequences that do not have conserved parts when compared to protein coding genes.

When applied to pathogen field, the lncRNAs increase their importance since they can be modulated by these microorganisms. Bacteria with undetermined pathogenicity, like *R. amblyommii*, are shown to be able to induce differential expression of ncRNAs inside their hosts and show us a very sophisticated functionality inherited by the ancestors. Once again, the evolution study is showing to the most pessimists researchers that it is important to understand the ancestry and how they evolved through the history. The high diversity of hosts, such as dogs, horses, birds and small mammlas, for example, as well as multiple tick species as pointed here, indicate the capacity of *R. amblyommii* adapt to each one. Of course,

we can extrapolate to other *Rickettsia* species, specially the most virulent ones. Furthermore, only one mammal was tested for *R. amblyommii* virulence, the guinea pig, which show the need of study with others animal models and add the role of lncRNA found herein and in others works. The new century is about to begin and we, scientists, are showing that a lot of information and knowledge can be generated and discovered on Earth.

7. ACKNOWLEDGMENTS

The authors are grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Foundation, Ministry of Education of Brazil to funding the doctorate sandwich scholarship in University of Texas Medical Branch (UTMB), Galveston, USA. Fundação de Amparo à Pesquisa de Minas Gerais (Fapemig) for funding the work. The Núcleo de Análise de Biomoléculas (Nubiomol) of Universidade Federal de Viçosa, Brazil, UTMB, Texas Advanced Computing Center (TACC) in USA and the UTMB researchers Saravanan Thangamani and Naomi Forrester for providing the facilities for the conduction of the experiments. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

8. REFERENCES

ALBERT, I. **The Biostar Handbook: Bioinformatics data analysis guide**.
BLANKENBERG, D. et al. A framework for collaborative analysis of ENCODE data: Making large-scale analyses biologist-friendly. **Genome Research**, v. 17, n. 6, p. 960–964, jun. 2007.
BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.
BRUCE ALBERTS et al. **Molecular Biology of the Cell**. 6. ed. [s.l.] Garland Publishing, 2016.
BUSBY, A. T. et al. Expression of heat shock proteins and subolesin affects stress responses, *Anaplasma phagocytophilum* infection and questing behaviour in the tick, *Ixodes scapularis*. **Medical and Veterinary Entomology**, v. 26, n. 1, p. 92–102, mar. 2012.
CAMACHO, C. et al. BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, p. 421, 2009.
DENG, X.; MELLER, V. H. **Non-coding RNA in fly dosage compensationTrends in Biochemical Sciences**Elsevier, , set. 2006. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16890440>. Acesso em: 22 fev. 2017
DJEBALI, S. et al. Landscape of transcription in human cells. **Nature**, v. 489, n. 7414, p. 101–108, 2012.
ETEBARI, K.; FURLONG, M. J.; ASGARI, S. Genome wide discovery of long intergenic non-coding RNAs in Diamondback moth (*Plutella xylostella*) and their expression in insecticide resistant strains. **Scientific reports**, v. 5, n. November, p. 14642, 28 set. 2015.
FINN, R. D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research**, v. 44, n. D1, p. D279–D285, 4 jan. 2016.
GALLETTI, M. F. B. M. et al. Virulence genes of *Rickettsia rickettsii* are differentially modulated by either temperature upshift or blood-feeding in tick midgut and salivary glands. **Parasites & vectors**, v. 9, n. 1, p. 331, 2016.
GARCIA, G. R. et al. The sialotranscriptome of Amblyomma triste, Amblyomma parvum and Amblyomma cajennense ticks, uncovered by 454-based RNA-seq. **Parasites & vectors**, v. 7, p. 430, 2014.
GRABHERR, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology**, v. 29, n. 7, p. 644–652, 2013.
GULIA-NUSS, M. et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. **Nature Communications**, v. 7, n. May 2015, p. 10507, 2016.
HAJDUŠEK, O. et al. Interaction of the tick immune system with transmitted pathogens. **Frontiers in Cellular and Infection Microbiology**, v. 3, p. 26, 2013.
HUANG, A. M.; RUBIN, G. M. A misexpression screen identifies genes that can modulate RAS1 pathway signaling in *Drosophila melanogaster*. **Genetics**, v. 156, n. 3, p. 1219–1230, nov. 2000.
IYER, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. **Nature genetics**, v. 47, n. 3, p. 199–208, 2015.
JAYAKODI, M. et al. Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. **BMC Genomics**, v. 16, n. 1, p. 680, 2015.
JENKINS, A. M.; WATERHOUSE, R. M.; MUSKAVITCH, M. A. Long non-coding RNA

discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. **BMC genomics**, v. 16, n. 1, p. 337, 23 dez. 2015.
KONG, L. et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. **Nucleic Acids Research**, v. 35, n. SUPPL.2, p. 345–349, 2007.
KOTSYFAKIS, M. et al. Tissue- and time-dependent transcription in Ixodes ricinus salivary glands and midguts when blood feeding on the vertebrate host. **Nature**, v. 5, n. 9103, p. 1–10, 2015.
KUNZE, M. et al. Mechanistic insights into PTS2-mediated peroxisomal protein import: The co-receptor PEX5L drastically increases the interaction strength between the cargo protein and the receptor PEX7. **Journal of Biological Chemistry**, v. 290, n. 8, p. 4928–4940, 20 fev. 2015.
MARTINS, L. A. et al. The Distinct Transcriptional Response of the Midgut of Amblyomma sculptum and Amblyomma aureolatum Ticks to Rickettsia rickettsii Correlates to Their Differences in Susceptibility to Infection. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. April, p. 129, 28 abr. 2017.
MATTICK, J. S. The genetic signatures of noncoding RNAs. **PLoS Genetics**, v. 5, n. 4, 2009.
MOREIRA, H. N. S. et al. A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of the *Amblyomma sculptum tick*. **Parasitology International**, v. 66, n. 2, p. 64–73, 2017.
NAVA, S. et al. Reassessment of the taxonomic status of *Amblyomma cajennense* (Fabricius, 1787) with the description of three new species, *Amblyomma tonelliae* n. sp., *Amblyomma interandinum* n. sp. and *Amblyomma patinoi* n. sp., and reinstatement. **Ticks and Tick-borne Diseases**, v. 5, n. 3, p. 252–276, 2014.
NIAZI, F.; VALADKHAN, S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. **Rna**, v. 18, n. 4, p. 825–843, 2012.
PALAZZO, A. F.; LEE, E. S. Non-coding RNA: What is functional and what is junk? **Frontiers in Genetics**, v. 5, n. JAN, p. 2, 26 jan. 2015.
QUÉNET, D.; DALAL, Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. **eLife**, v. 3, p. e03254, 12 ago. 2014.
ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 2009.
SCHLESINGER, M. J. **Heat shock proteinsJournal of Biological Chemistry**, 25 jul. 1990. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/2197269>. Acesso em: 26 maio. 2017
SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, v. 27, n. 6, p. 863–864, 15 mar. 2011.
SONENSHINE, D. E.; ROE, R. M. **Biology of Ticks: Volume 1**. 2nd. ed. [s.l.] Oxford University Press, 2013.
SOSHNEV, A. A. et al. A conserved long noncoding RNA affects sleep behavior in *Drosophila*. **Genetics**, v. 189, n. 2, p. 455–468, 2011.
ST.LAURENT, G.; WAHLESTEDT, C.; KAPRANOV, P. The Landscape of long noncoding RNA classification. **Trends in Genetics**, v. 31, n. 5, p. 249–251, 2015.
SULTANA, H. et al. *Anaplasma phagocytophilum* induces actin phosphorylation to selectively regulate gene transcription in *Ixodes scapularis* ticks. **The Journal of**

Experimental Medicine, v. 207, n. 8, p. 1727–1743, 2 ago. 2010.
WILLINGHAM, A. T.; KEIL, T. A tissue specific cytochrome P450 required for the structure and function of *Drosophila* sensory organs. **Mechanisms of Development**, v. 121, n. 10, p. 1289–1297, 2004.
WU, K. et al. The Glutathione-S-Transferase, Cytochrome P450 and Carboxyl/Cholinesterase Gene Superfamilies in Predatory Mite *Metaseiulus occidentalis*. **PLOS ONE**, v. 11, n. 7, p. e0160009, 28 jul. 2016.
ZUG, R.; HAMMERSTEIN, P. *Wolbachia* and the insect immune system: What reactive oxygen species can tell us about the mechanisms of *Wolbachia*-host interactions. **Frontiers in Microbiology**, v. 6, n. OCT, p. 1201, 27 out. 2015.