Sialotranscriptomics of the brown ear ticks, *Rhipicephalus appendiculatus* Neumann, 1901 and *R. zambeziensis* Walker, Norval and Corwin, 1981, vectors of Corridor disease

by

MINIQUE HILDA DE CASTRO

submitted in accordance with the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in the subject

LIFE SCIENCES

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF. B.J. MANS

CO-SUPERVISOR: PROF. D.J.G. REES

NOVEMBER 2017

“It always seems impossible until it’s done”

— Nelson Mandela —

DEDICATION

I would like to dedicate this thesis to my late ouma (grandmother), whom on her sick bed encouraged me to do this degree. I miss you, ouma.

And I thank you.

DECLARATION

Name: Minique Hilda de Castro
Student number: 53321529
Degree: Doctor of Philosophy in Life Sciences

Thesis title:
Sialotranscriptomics of the brown ear ticks, *Rhipicephalus appendiculatus* Neumann, 1901 and *R. zambeziensis* Walker, Norval and Corwin, 1981, vectors of Corridor disease

I, Minique Hilda de Castro, declare that the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I have not previously submitted this work, or part of it, for examination at UNISA for another qualification or at any other higher education institution.

_____ _____

SIGNATURE DATE

ACKNOWLEDGEMENTS

The path to complete this PhD was long and sometimes very rocky and it would not have been possible without the support of a number of people that I would like to express my sincerest gratitude to:

My supervisor, Prof. Ben Mans, for valuable insights into the project, financial support and the longest meetings ever (!), where we were able to discuss research questions and problems as peers. I really appreciated that. It was a privilege to see how your brain works and I hope to still learn a great deal from you in future. My co-supervisor, Prof. Jasper Rees, for introducing me to the worlds of Macintosh and Next generation sequencing. Thank you for giving me the opportunity to do this degree, for financial support and of course, for 'lab space'.

My colleagues and co-authors from the Epidemiology, Parasites and Vectors group, Daniel De Klerk, Siyamcela Genu, Philasande Gaven and Dr. Abdalla Latif. And especially to Ronel Pienaar, for your valuable advice and assistance during the progression of the degree. My colleagues and friends at the Biotechnology Platform, especially the students from the 'First Wave', for support, smiles and payday 'braais' (barbeques). Past and present staff at the Biotechnology Platform Sequencing Facility, Marija Kvas, Jonathan Featherston, Kerry-Anne Pillay and Jurgens de Bruin, for preforming the sequencing for the project. Charles Hefer for bioinformatic support and guidance. And to Lerato Mpye, whom we all know runs the Platform, for administrative assistance and advice during the project.

My family without whom all of this would not have been possible. My mother (Corine de Castro) and father (Alex), for your unconditional love, patience, understanding and support, both emotional and financial, throughout the course of the degree. My little sister (Therese), for our long Skype conversations, even though you did most of the talking (☺), they kept me sane and focused on what is important. My sister (Alexa) and other sister (Lizelle van den Berg), for your support and allowing me to have a home away from home (that also has a beach!). My second family (Karien and Johan Vermeulen), for unconditional love and support and constant packed dinners.

My friends both near and far. I have had the privilege to walk this road with a number of amazing friends that I would like to thank. Vinet Coetzee, for tireless support and guidance, editorial advice, giving me a roof over my head when I didn't have one anymore and gluten-fee pizza's on nights when I was near my limits. And of course little Gracy Coetzee, whom with the little patter of her feet kept mine firmly on the ground and my head in the game. I will always remember that, in the end, it was you that made me complete this degree. Ritesh Mewalal, for the break from writing (and boredom) our Skype conversations promised. I love you and miss you my friend. The 'Golden Girls', Annie Gous, Stephanie Cornelissen and Erika Viljoen, for endless motivational (and some not so) conversations on our Whatsapp group and at 'Jocks'. You guys have a special place in my heart and I would not have been here without you. Simon Lashmar, whom I only know for a short while but already love dearly. Tracey-Leigh Prigge (and the rest of the Prigge-clan), Nicky Creux, Timothy Bray, Alex Jansen van Rensburg, Isa-Rita Russo, Shani Bekker, Sunel and Rane Oelofse, Louise Swart and the rest of the Dining Club, Magretha Wang, Bronwyn Gover, Adrian Bantich, Marja and Karel Mostert and Marius Riekert. Thank you for being my friend even after I disappeared into hermit-mode, for listening to my rantings over a 'wyntjie' (a glass of wine) or two and for supporting me through this process.

THESIS SUMMARY

Sialotranscriptomics of the brown ear ticks, *Rhipicephalus appendiculatus* Neumann, 1901 and *R. zambeziensis* Walker, Norval and Corwin, 1981, vectors of Corridor disease

Minique Hilda de Castro

Degree: Doctor of Philosophy

Subject: Life Sciences

Supervisors: Prof. B.J. Mans and Prof. D.J.G. Rees

Corridor disease is an economically important tick-borne disease of cattle in southern Africa. The disease is caused by *Theileria parva* and transmitted by the vectors, *Rhipicephalus appendiculatus* and *R. zambeziensis*. There is currently no vaccine to protect cattle against *T. parva* that is permitted in South Africa. To develop recombinant anti-tick vaccines against Corridor disease, comprehensive databases of genes expressed in the tick's salivary glands are required. Therefore, in Chapters 2 and 3, mRNA from the salivary glands of *R. appendiculatus* and *R. zambeziensis* was sequenced and assembled using next generation sequencing technologies. Respectively, 12 761 and 13 584 non-redundant protein sequences were predicted from the sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* and uploaded to public sequence domains. This greatly expanded the number of sequences available for the two vectors, which will be invaluable resources for the selection of vaccine candidates in future. Further, in Chapter 3, differential gene expression analysis in *R. zambeziensis* revealed dynamic expression of secretory protein transcripts during feeding, suggestive of stringent transcriptional regulation of these proteins. Knowledge of these intricate expression profiles will further assist vaccine development in future. In Chapter 4, comparative sialotranscriptomic analyses were performed between *R. appendiculatus* and *R. zambeziensis*. The ticks have previously shown varying vector competence for *T. parva* and this chapter presents the search for correlates of this variance. Phylogenetic analyses were performed using these and other publically available tick transcriptomes, which indicated that *R. appendiculatus* and *R. zambeziensis* are closely related but distinct species. However, significant expression differences were observed between the two ticks, specifically of genes involved in tick immunity or pathogen transmission,

signifying potential bioinformatic signatures of vector competence. Furthermore, nearly four thousand putative long non-coding RNAs (lncRNAs) were predicted in each of the two ticks. A large number of these showed differential expression and suggested a potential transcriptional regulatory function of lncRNA in tick blood feeding. LncRNAs are completely unexplored in ticks. Finally, in Chapter 5, concluding remarks are given on the potential impact the *R. appendiculatus* and *R. zambeziensis* sialotranscriptomes may have on future vaccine developments and some future research endeavours are discussed.

Key terms:
Rhipicephalus appendiculatus; *Rhipicephalus zambeziensis*; Corridor disease; Tick salivary glands; *de novo* transcriptome assembly; Next generation sequencing; Sialotranscriptomics; Secretory proteins; Differential gene expression; Comparative transcriptomics; Species phylogeny; Vector competence; Long non-coding RNA.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1

Literature Review: Corridor disease and its control in South Africa

1.1 General introduction

Ticks are blood feeding ectoparasites that serve as vectors for a variety of human and veterinary diseases worldwide (Jongejan and Uilenberg, 2004; Dennis and Piesman, 2005). Of the nearly nine hundred tick species, approximately 10% are known disease vectors affecting livestock, humans and domestic animals (Jongejan and Uilenberg, 2004). Ticks transmit more bacterial, viral and protozoan pathogens than any other arthropod group. In humans, ticks are the second most important disease vector, only surpassed by mosquitos (Sonenshine, 1991). In livestock, ticks transmit a number of economically important tick-borne diseases (babesiosis, anaplasmosis and theileriosis) and severe tick infestations can reduce animal weight, lower milk production and reduce hide qualities (De Castro, 1997). Annual global losses due to ticks and tick-borne diseases in cattle was estimated at US$13.9 - 18.7 billion in 1996 (De Castro, 1997), but considering inflation rates, these losses could be closer to US$22 - 30 billion per annum (Lew-Tabor and Rodriguez-Valle, 2016). In South Africa alone, the losses due to ticks and tick-borne diseases have been estimated at roughly US$31.6 million per year (Minjauw and McLeod, 2003). Chemical acaricides have always been the most effective mechanism of tick control (Willadsen, 2006), but the emergence of acaricide-resistance (Abbas *et al.*, 2014) and health concerns due to chemical residues in meat, milk and the environment (Graf *et al.*, 2004) have recently shifted the focus of tick biologists to the development of recombinant anti-tick vaccines (Guerrero *et al.*, 2012; Marcelino *et al.*, 2012; de la Fuente *et al.*, 2016a; Lew-Tabor and Rodriguez-Valle, 2016).

1.2 Corridor disease

Corridor disease (CD), also known as buffalo disease, is an economically important cattle disease in southern Africa, which obtained its name after being (re)discovered in the ‘corridor’ between the then Hluhluwe and Umfolozi game reserves (now incorporated together in the Hluhluwe-iMfolozi Park) in KwaZulu-Natal (Neitz *et al.*, 1955). It was shown to be the same disease that was previously reported in Zimbabwe twenty years before (Lawrence, 1979). Corridor disease is caused by the protozoan parasite *Theileria parva*, which is transmitted from the natural reservoir host, African buffalo (*Syncerus caffer*), to cattle (Uilenberg, 1999). It is a tick-borne disease spread by the brown ear ticks, *Rhipicephalus appendiculatus*, *R. zambeziensis* and *R. duttoni*

(Norval *et al.*, 1992). Corridor disease is acute, often fatal, and cattle can die within four days from the onset of symptoms. The disease is characterised by fever, enlarged lymph nodes, pulmonary oedema and laboured breathing and has been shown to result in mortality rates of more than 90% (Neitz *et al.*, 1955; Potgieter *et al.*, 1988).

In South Africa, CD is a controlled disease (Animal Disease Act 1984, Act No. 35) and is endemic to the Kruger National Park, Hluhluwe-iMfolozi, Eastern Shores, Ndumu and Tembe game reserves and adjacent buffalo farms (Potgieter *et al.*, 1988). However, the main CD vector, *R. appendiculatus*, has a much wider distribution range in South Africa (Estrada-Pena, 2003), highlighting the risk that CD could rapidly spread through South Africa if not strictly controlled. The game industry is expanding in South Africa (Sibeko *et al.*, 2008) and with it the demand for wildlife animals such as buffalo, increasing the risk of CD outbreaks. Corridor disease outbreaks occur on cattle farms adjacent to game farms where cattle come into close contact with *T. parva*-carrying buffalo (Sibeko *et al.*, 2008; Mbizeni *et al.*, 2013). Recently, surveillance of 15 farms (communal and commercial) in KwaZulu-Natal, reported 31 CD outbreaks over a six-year period (Mbizeni *et al.*, 2013).

1.2.1 Cattle-associated theileriosis

Two additional disease syndromes are also caused by *T. parva*, East Coast fever (ECF) and January disease (Zimbabwean theileriosis). The diseases differ from CD by their etiologies: ECF and January disease are transmitted between cattle, whereas CD is transmitted from buffalo to cattle. East Coast fever is an economically important, fatal disease spread throughout central and eastern Africa, killing more than a million animals per year and amounting in approximately US$168 million worth of damages in 1989 alone (Mukhebi *et al.*, 1992). The disease symptoms are similar to CD, although disease progression is slower, about two weeks (ranging from one to three weeks) and the cattle usually die from severe pulmonary oedema (Lawrence, 1979). East Coast fever was introduced into South Africa in 1902 by infected cattle imported from Tanzania and Kenya to restock depleted cattle numbers after the rinderpest epidemic of 1896 (Lawrence, 1979; Stoltsz, 1989). The ensuing ECF epidemic resulted in an estimated 5.5 million cattle deaths (Potgieter *et al.*, 1988). East Coast fever was finally eradicated from South Africa in the 1950's through rigorous tick control, quarantine and

slaughtering measures (Neitz, 1957; Norval *et al.*, 1992). Shortly after the eradication of ECF, CD was identified (Neitz *et al.*, 1955) and has persisted in South Africa ever since. January disease is a less severe, seasonal disease occurring in Zimbabwe during the rainy months of January to April, coinciding with adult *R. appendiculatus* activity (Lawrence, 1979). The disease was first recognised in 1936 (Lawrence, 1979) and annually causes a significant number of cattle deaths. Stringent control measures employed by Zimbabwe, such as dipping and alternate grazing pastures during dry months have alleviated the disease threat (Latif *et al.*, 2001).

1.2.2 *Theileria parva*

The causative agent of CD, ECF and January disease is the apicomplexan protozoan parasite, *Theileria parva*, Theiler, 1904 (Uilenberg, 1999). *Theileria parva* is the most economically important tick-borne pathogen of cattle in Africa (Minjauw and McLeod, 2003) and is distributed through eastern, central and southern Africa. *Theileria parva* also asymptomatically infects the African buffalo and is believed to have co-evolved in buffalo long before the introduction of cattle (Uilenberg, 1981; Young, 1981).

The life cycle of *T. parva* is complex and involves the completion of various stages in both the tick and vertebrate hosts (Figure 1.1) and has been reviewed on numerous occasions (e.g. Norval *et al.*, 1992; Bishop *et al.*, 2004; Mans *et al.*, 2015; Tretina *et al.*, 2015; Nene *et al.*, 2016). Herein follows a shortened version of only the main stages. *Theileria parva* is a transstadially-transmitted parasite, whereby the tick vector obtains the parasite by a blood meal in one life stage and transmits it to the vertebrate host of its ensuing life stage after moulting. The parasite remains in the salivary glands of the tick during moulting and sporogony (multiplication) is only initiated when the tick attaches to the next vertebrate host. Infective sporozoites are released into the host via the tick saliva during feeding. The sporozoites invade the lymphocytes of the vertebrate host and form schizonts. The infection transforms the lymphocytes into immortalised lymphoblasts that undergo clonal expansion and is the major cause of pathology and disease in the vertebrate host (reviewed in Dobbelaere and Heussler, 1999). Following, merogony occurs in the infected lymphocytes, which results in cell rupture and release of merozoites. The merozoites infect erythrocytes and are transformed into piroplasms, the infective stage to tick vectors. In the gut of the tick,

gametogenesis occurs that forms zygotes, the only diploid stage (Gauer *et al.*, 1995). The zygotes invade the gut epithelial cells and are released into the haemolymph in the form of motile kinetes. The kinetes move to and invade the salivary glands, specifically the *e* cells of the type III acini (Fawcett *et al.*, 1982), where they remain until the following tick instar.

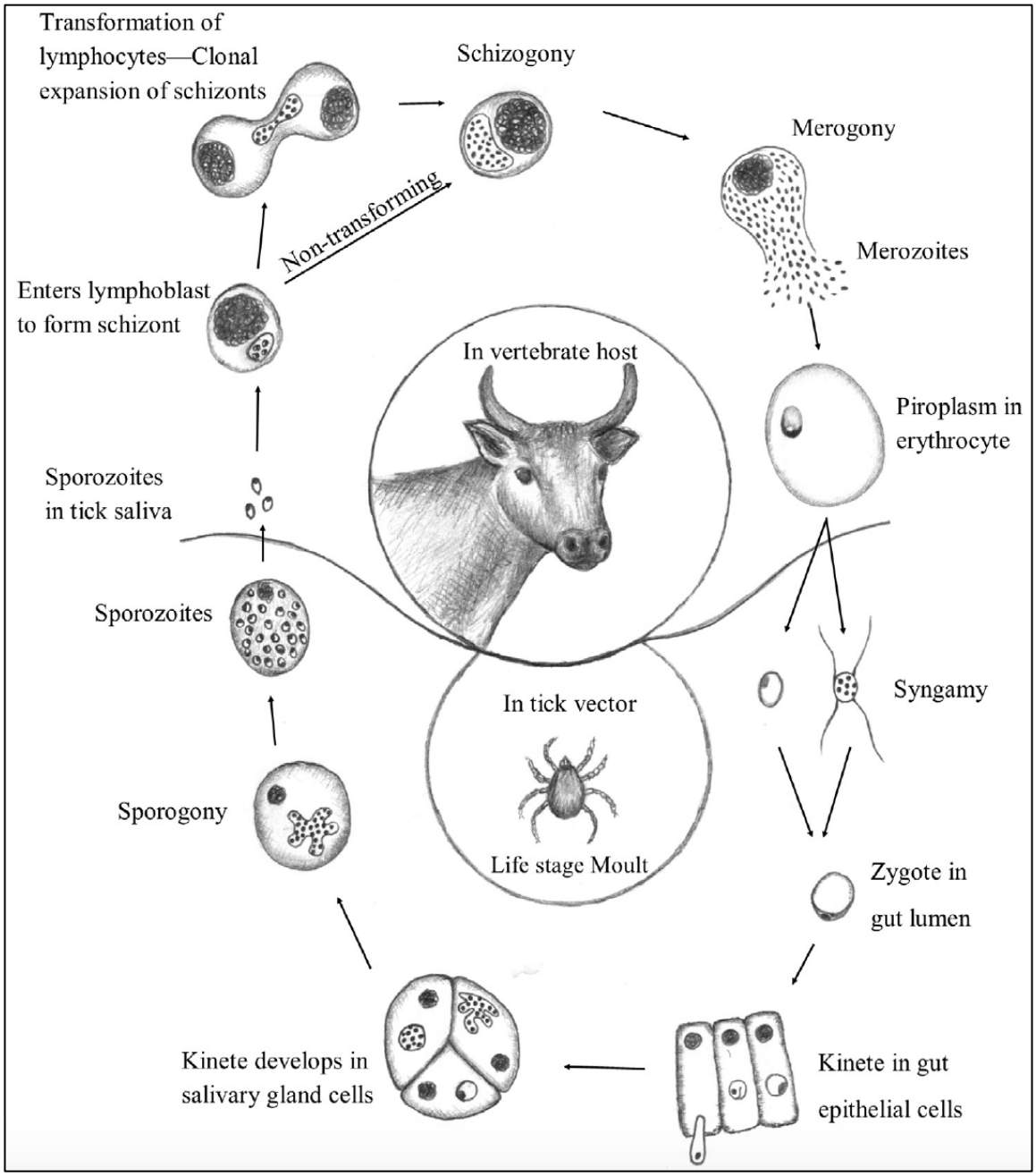


Figure 1.1 Life cycle of *T. parva* in the vertebrate and invertebrate hosts. Source of diagram: Mans *et al.* (2015).

The genomes of a number of *T. parva* strains have been sequenced to date, showing high variability between the strains (Gardner *et al.*, 2005; Henson *et al.*, 2012; Hayashida *et al.*, 2013; Norling *et al.*, 2015). Previously, *T. parva* causing CD, ECF and January disease were classified as subspecies: *T. parva parva*, causing ECF, transmitted between cattle and characterised by high schizont and piroplasm loads in the infected cattle; *T. parva lawrencei*, causing CD, transmitted from the African buffalo to cattle and characterised by few schizonts and very few to no piroplasms in the infected cattle; and *T. parva bovis*, an intermediate parasite causing the intermediate epidemiology of the less pathogenic January disease, also transmitted between cattle (Neitz *et al.*, 1955; Lawrence, 1979; Potgieter *et al.*, 1988; Uilenberg, 1999). This nomenclature was abandoned (Perry and Young, 1993) after it was determined that the genetic variability of *T. parva* in nature was more complex and rather represented a range of diversity in the species than three distinct subspecies. Today, they are merely referred to as buffalo- or cattle-derived *T. parva*, with buffalo-derived *T. parva* strains showing higher levels of diversity than cattle-derived *T. parva* (Conrad *et al.*, 1987; Pelle *et al.*, 2011).

1.2.3 Transformation of buffalo-derived *T. parva* to cattle-derived *T. parva*

Due to the rapid disease progression of CD, caused by buffalo-derived *T. parva*, it is assumed that CD is predominantly self-limiting in cattle, since cattle die before the *T. parva* parasite reaches the tick infective piroplasm stage (Neitz *et al.*, 1955; Neitz, 1957; Norval *et al.*, 1992). Under experimental conditions, it was shown that repeated passage of buffalo-derived *T. parva* through cattle, transformed the parasite to be behaviourally similar to that of cattle-derived *T. parva* (Barnett and Brocklesby, 1966; Maritim *et al.*, 1992). Even so, the topic of ‘transformation’ is highly controversial as many attempts to transform buffalo-derived *T. parva* in South Africa have been unsuccessful (Neitz *et al.*, 1955; Neitz, 1957; Potgieter *et al.*, 1988). In cattle populations in South Africa, it has been shown that a small percentage of cattle can survive infection (Thompson *et al.*, 2008; Yusufmia *et al.*, 2010; Mbizeni *et al.*, 2013), with the potential to become carriers of *T. parva* and infective to other cattle. However, under experimental conditions, ticks did not become infected when fed on these recovered cattle, and a carrier-state in the cattle could not be proven (Thompson *et al.*, 2008; Mbizeni *et al.*, 2013). The transformation of buffalo-derived *T. parva* to cattle-derived *T. parva* will have serious

implications for the control of CD in South Africa and might lead to a situation similar to the original ECF epidemic (Yusufmia *et al.*, 2010).

1.2.4 Vectors of *T. parva*

Theileria parva is transmitted by the brown ticks, *R. appendiculatus, R. zambeziensis* and *R. duttoni* (Lawrence *et al.*, 1983; Stoltsz, 1989; Norval *et al.*, 1992). The ticks are three-host ticks that obtain obligate blood meals from different hosts during each life stage (larvae, nymphs and adults). *Rhipicephalus appendiculatus* and *R. zambeziensis* are highly similar, even though the two species can still be successfully differentiated using phylogenetic (Wouters *et al.*, 1987; Mtambo *et al.*, 2007a) and morphological (Walker *et al.*, 2005) analyses. A notable difference between the two species is, however, their ability to transmit *T. parva* (Potgieter *et al.*, 1988; Stoltsz, 1989; Ochanda *et al.*, 1998). *Rhipicephalus duttoni*'s natural distribution is restricted to south-western Angola (Gomes and Wouters, 1991) and due to its absence from South Africa it is out of scope of the current review.

1.2.4.1 *Rhipicephalus appendiculatus*

The main vector of *T. parva* is *Rhipicephalus appendiculatus* Neumann, 1901 (Norval *et al.*, 1992). The tick also transmits *T. taurotragi* causing benign bovine theileriosis, *Anaplasma marginale* causing bovine anaplasmosis, Thogoto virus causing Nairobi sheep disease and *Rickettsia conorii* causing tick typhus in humans (Walker *et al.*, 2005). The adults mainly infest domestic animals, such as cattle, goats and sheep, and wild animals, such as buffalo, waterbuck and eland (Walker *et al.*, 2005). The immature life stages are found on smaller mammals, such as impala and hares. The adult ticks feed on and in the ears of their hosts and in cattle, high levels of infestation can cause severe damage to the earlobes and collapse of the immune system resulting in susceptibility to other diseases (Jongejan and Uilenberg, 2004).

Rhipicephalus appendiculatus naturally occurs throughout central, eastern and southern Africa (Figure 1.2) and this wide distribution makes it a very economically important vector on the African continent (Norval *et al.*, 1992; Walker *et al.*, 2003). In South Africa, *R. appendiculatus* is present in Limpopo, Northwest, Gauteng, Mpumalanga, KwaZulu-Natal and the coastal regions of the Eastern Cape (Estrada-

Pena, 2003). Based on projections of future climate change, a distribution shift of *R. appendiculatus* from west to east has been predicted, based on predicted temperature increases in the already hot and dry western areas of Africa (Olwoch *et al.*, 2008) This could increase tick population numbers in Botswana, Malawi, South Africa and Zimbabwe (Olwoch *et al.*, 2008), impacting the control measures of CD in southern Africa.

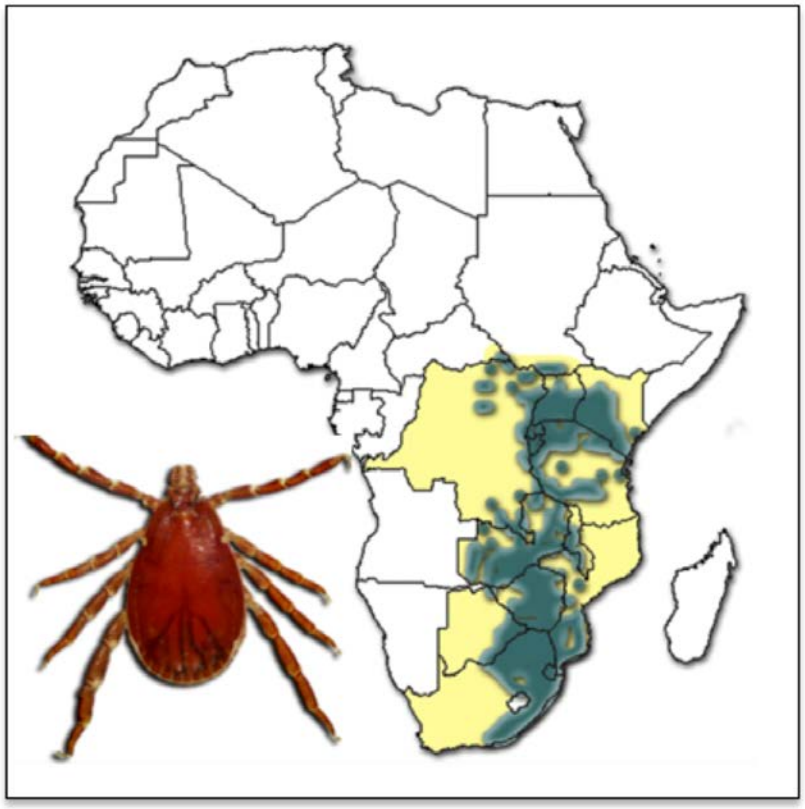


Figure 1.2 Natural distribution of *R. appendiculatus* in Africa. The colours denote; green for the localities and yellow for the countries where the tick has been identified. A dorsal view of a male *R. appendiculatus* is shown to the left of the diagram. Source: www.afrivip.org.

Rhipicephalus appendiculatus species have been clustered into three groups based on morphological, behavioural and genetic differences: the eastern African group (Kenya, Rwanda, Burundi, Tanzania and Uganda); the southern African group (southern province of Zambia, Zimbabwe and South Africa); and an intermediate 'transition' group (eastern province of Zambia), sharing characteristics with both the eastern and

southern groups (Madder *et al.*, 1999). The tick body of *R. appendiculatus* changes in size according to latitude, where ticks from the southern group have greater body size than those from the eastern group (Speybroeck *et al.*, 2004). Lower survival rates under unfavourable conditions have been reported for *R. appendiculatus* with small body sizes (Chiera *et al.*, 1985). Southern Africa has marked wet and dry climatic seasons and Speybroeck *et al.* (2004) ascribed the changes in body size to larger ticks requiring survival advantages in these distinct seasonal conditions. The distinct seasons are also believed to be the reason that southern *R. appendiculatus* ticks enter diapause, whereas eastern *R. appendiculatus* do not (Madder *et al.*, 2002). In eastern Africa, *R. appendiculatus* can be found on animals throughout the year, completing two or more life cycles annually (Kaiser *et al.*, 1982). Whereas in southern Africa, a single annual life cycle of *R. appendiculatus* is observed (Short and Norval, 1981; Rechav, 1982), with complete absence of ticks in the dry season when the ticks are in diapause (Speybroeck *et al.*, 2002). Based on molecular data, the mitochondrial *cytochrome c oxidase subunit I* (*cox1*) and *12S ribosomal RNA* (*rRNA*) genes divided the eastern and southern *R. appendiculatus* groups into two genetically differentiated groups (Mtambo *et al.*, 2007b). Further, the *cox1* gene separated the southern African group from the 'transition' group (Mtambo *et al.*, 2007c), indicating that the three classified *R. appendiculatus* groups are genetically distinct. Only low levels of genetic variation was observed within the eastern African group (Kenya), where the *cox1* and *12S rRNA* genes showed no phylogeographic structure (Kanduma *et al.*, 2016a) and microsatellite markers showed little genetic differentiation (Kanduma *et al.*, 2016b).

1.2.4.2 *Rhipicephalus zambeziensis*

Rhipicephalus zambeziensis Walker, Norval and Corwin, 1981 is distributed through eastern and southern Africa (Figure 1.3) (Walker *et al.*, 1981). In South Africa it can be found in the provinces of Limpopo, Northwest and Mpumalanga (Walker *et al.*, 2005). The tick prefers hot and dry river valley systems, in environments which are not as arid as semi-desert or desert areas (Walker *et al.*, 1981). In geographic regions that overlap with the distribution of *R. appendiculatus*, *R. zambeziensis* occurs at low altitude in regions not suitable for *R. appendiculatus* survival. *Rhipicephalus zambeziensis* is a vector for *T. parva*, *T. taurotragi* and *A. bovis* (Walker *et al.*, 2005). The adult and immature stages of *R. zambeziensis* can be found on similar hosts as *R. appendiculatus*,

but in domesticated animals they preferentially feed on cattle (Norval *et al.*, 1982). The tick has not been shown to undergo behavioural diapause (Berkvens *et al.*, 1995), potentially allowing a second tick generation in favourable climatic conditions. Berkvens *et al.* (1995) further noted that in the absence of diapause, the manner by which *R. zambeziensis* survives the harsh conditions (high ambient temperatures and low humidity) of the southern African climatic cycles is unknown.

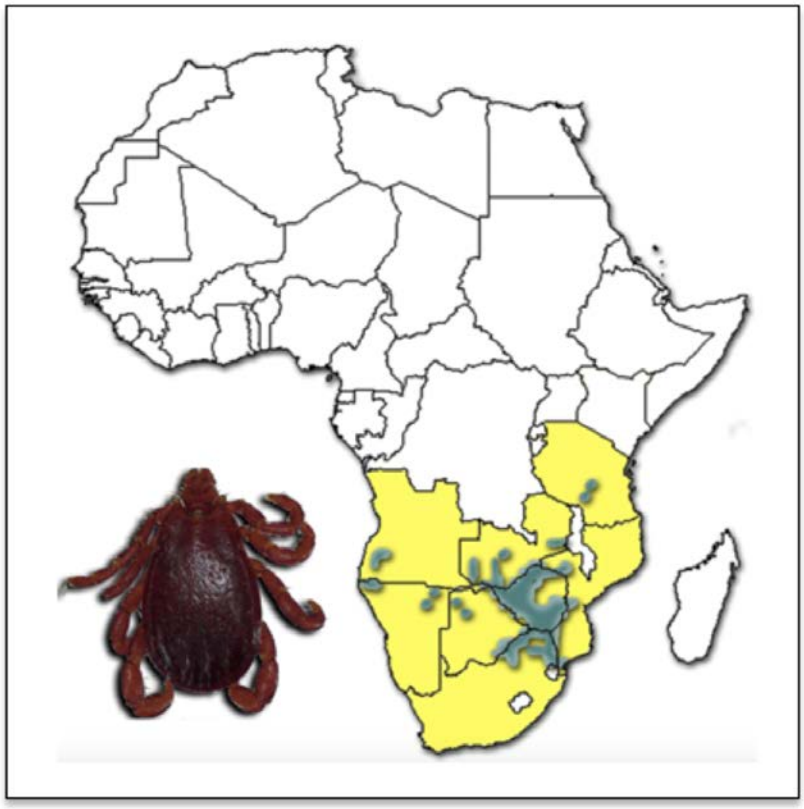


Figure 1.3 Geographic distribution of *R. zambeziensis* in Africa. Green indicates localities and yellow, countries in which the tick has been identified. A male specimen is also shown. From www.afrivip.org.

1.2.4.3 Differences between *R. appendiculatus* and *R. zambeziensis*

Rhipicephalus appendiculatus and *R. zambeziensis* ticks are morphologically highly similar (Walker *et al.*, 1981), but they have been proven to be bonafide species using phylogenetic analyses (Mtambo *et al.*, 2007a). The ticks are difficult to distinguish in the field, but *R. zambeziensis* has more prominent punctuations on its scutum. Experimental interbreeding produced fertile offspring for crosses between *R.*

zambeziensis females and *R. appendiculatus* males, but not reciprocally (Zivkovic *et al.*, 1986). The developmental stages of *R. zambeziensis* take longer (Walker *et al.*, 1981), the females lay more eggs (Zivkovic *et al.*, 1986) and the species is more adapted to extreme experimental environmental conditions (Madder *et al.*, 2005) as compared to *R. appendiculatus*. Indeed, *R. zambeziensis* is known to naturally occur in hotter, drier regions than *R. appendiculatus* (Walker *et al.*, 1981). Competitive displacement between *R. appendiculatus* and *R. zambeziensis* was shown to occur in regions where the species distribution overlap (Mooring and Mazhowu, 1995) and in years of extreme temperature rises, *R. zambeziensis* replaced *R. appendiculatus* in drier areas (Norval *et al.*, 1982; Madder *et al.*, 2005). In subsequent years as the temperature stabilised, *R. appendiculatus* returned as the predominant vector.

Maybe the most pertinent difference between the tick species is the variability in vector competence of *T. parva*, where more *R. zambeziensis* ticks had infected salivary glands and of a higher infection load (more infected acini) than *R. appendiculatus* after being fed on *T. parva*-infected cattle (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989). Moreover, when comparing vector competence of two strains of *T. parva*, Muguga (sampled from Kenya) and Boleni (sampled from Zimbabwe), it was observed that *R. zambeziensis* salivary glands showed higher infection rates than five *R. appendiculatus* strains when the Boleni (southern African) strain was used as infection agent (Ochanda *et al.*, 1998). These studies suggested that *R. zambeziensis* has better vector competence for *T. parva* than *R. appendiculatus*, especially when *T. parva* strains naturally occurring in the southern parts of Africa and sharing geographical distribution with *R. zambeziensis* are concerned (Walker *et al.*, 1981).

Climatic changes, such as fluctuations in temperature and rainfall, could change the natural distribution of a tick species as suitable environments are established in other regions (Dantas-Torres, 2015). With the projected increase in temperature and decrease in rainfall in sub-Saharan Africa, the distribution of *R. appendiculatus* has been predicted to change (Olwoch *et al.*, 2008), and similarly *R. zambeziensis* can be expected to change distribution as it prefers hot and dry environments (Norval *et al.*, 1982). The risk of *R. zambeziensis*, a highly competent vector of *T. parva*, that has high egg production, that has a propensity for hot and dry conditions and shown to move into

areas when *R. appendiculatus* departs, spreading to a larger distribution due to predicted climatic changes, will have serious implications for the control of CD in future.

1.3 Vector competence

Ticks require large blood meals that they ingest over days to weeks to complete their life cycle (Sonenshine, 1991; Jongejan and Uilenberg, 2004). Extended feeding times result in suitable environments for microorganism growth and as a result, ticks inadvertently carry infections of a number of different pathogens, many of which are harmful to their vertebrate hosts and are responsible for some of the most economically important vector-borne diseases (Jongejan and Uilenberg, 2004). The capability of a tick to transmit a pathogen relies on a number of factors, including the pathogen's ability to infect the tick, pathogen survival and multiplication in tick tissues and successful transmission of the pathogen to the host; this is referred to as vector competence (Lane, 1994; Beerntsen *et al.*, 2000; de la Fuente *et al.*, 2017). Vector competence is the genetic component of the broader terminology of vector capacity that refers to vector-pathogen-host association, and includes environmental and behavioural factors, such as tick densities, feeding preferences, longevity, and host and pathogen availability (Beerntsen *et al.*, 2000).

1.3.1 Evolution of vector-pathogen interactions

Defence against pathogen invasion has driven the evolution of a basic innate immune system in ticks. This immune system has been reviewed extensively before (e.g. Taylor, 2006; Kopáček *et al.*, 2010; Hajdušek *et al.*, 2013; Baxter *et al.*, 2017; de la Fuente *et al.*, 2017). Broadly, the immune response consists of the recognition of pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs), the activation of signalling cascades, antimicrobial peptide (AMP) secretions, pathogen opsonisation and phagocytosis and apoptosis to remove infected cells. Genes involved in tick innate immunity, being the sole means of protection against infecting pathogens, have shown high allelic variation and accelerated evolution (Kotsyfakis *et al.*, 2015b; Baxter *et al.*, 2017).

Pathogen infections affect different tick species in varying ways as was seen in a meta-analysis of ticks infected by *Anaplasma* species, where ticks that were natural vectors of a specific pathogen elicited less severe stress responses to infection than ticks that were non-natural vectors, probably due to co-evolution (Villar *et al.*, 2010). Pathogens and ticks have co-evolved together to mutually benefit both organisms (de la Fuente *et al.*, 2016b), e.g.: ticks can benefit from pathogen infection by obtaining enhanced survival and in turn the pathogen receives the benefit of being transmitted to the vertebrate host. This has been shown in ticks by the overexpression, in response to infection, of an antifreeze glycoprotein giving cold tolerance (Neelakanta *et al.*, 2010) and heat shock proteins assisting with stress generated from heat shock and questing behaviour (Busby *et al.*, 2012; Villar *et al.*, 2015). Pathogens can successfully be transmitted and in some cases assisted during transmission (Ramamoorthi *et al.*, 2005; Nuttall and Labuda, 2008; Dai *et al.*, 2010), to complete their next life cycle.

Pathogen transmission to the vertebrate host is achieved by exploitation of the immune modulatory properties of the vector (Nuttall *et al.*, 2000). The enhancement of parasite transmission in the presence of a salivary gland extract (SGE) has been found for a number of different tick and pathogen interactions (reviewed in Nuttall and Labuda, 2004) and was first observed through the enhancement of Thogoto virus transmission by *R. appendiculatus* SGE (Nuttall and Jones, 1991). The authors referred to this enhancement of transmission as saliva-assisted transmission (SAT). It has been shown that SAT is highly species-specific for the combination of vector and pathogen, potentially indicating the involvement of a small set of highly specialised proteins (Jones *et al.*, 1992).

1.3.2 Interactions between *T. parva* and *R. appendiculatus*

As mentioned earlier in this review, experimental evidence has shown that *R. zambeziensis* exhibits better vector competence of *T. parva* than *R. appendiculatus* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998). However, the molecular mechanism of this difference in vector competence is completely unknown. The potential effect on the control of CD if *R. zambeziensis* spreads to a wider distribution, warrants a better understanding of *T. parva*-vector interactions and the differences between the responses of *T. parva* vectors to infection. Previously, studies

have focused on *T. parva-R. appendiculatus* interactions, of which some will briefly be discussed in this section. To date, no studies have been performed on *T. parva-R. zambeziensis* interactions.

A functional role for haemagglutinin/ lectin, which functions as a pattern-recognition receptor (PRR), was shown in *T. parva-R. appendiculatus* interactions (Kamwendo *et al.*, 1995; Kibuka-Sebitosi, 2006). Haemagglutinin activity increased during the infection of the ticks (Kibuka-Sebitosi, 2006) and the inhibition of haemagglutinin resulted in an increase in infection of the salivary glands (Kamwendo *et al.*, 1995). Additionally, Kibuka-Sebitosi (2006) showed that *R. pulchellus*, a species refractory to *T. parva*, exhibited higher levels of haemagglutinin activity than *R. appendiculatus* that vectors *T. parva*. This highlights a potential role of haemagglutinin activity in determining vector competence of *T. parva*.

Further, as mentioned before, *R. appendiculatus* SGE demonstrated evidence of SAT through the enhancement of Thogoto virus transmission (Nuttall and Jones, 1991). Additional evidence of the presence of SAT properties in the salivary glands of *R. appendiculatus* was obtained by the four-fold enhancement of tick-borne encephalitis virus transmission as compared to transmission in the absence of SGE (Labuda *et al.*, 1993). Furthermore, *R. appendiculatus* SGE were shown, in combination with interleukin-2, to enhanced the susceptibility of lymphocytes to *T. parva* infection (Shaw *et al.*, 1993). It would be invaluable to understand the mechanisms and proteins involved in SAT in *R. appendiculatus,* to determine the manner by which they assist *T. parva* transmission and whether they might influence the competency of the vectors.

Lastly, in a study analysing gene expression differences between *T. parva*-infected and uninfected *R. appendiculatus* ticks, no significant differences were found in the salivary glands based on conventional cDNA sequencing technologies (Nene *et al.*, 2004). It is known that *T. parva* infects only the *e* cells of type III salivary gland acini (reviewed in Bishop *et al.*, 2004) and at an overall low infectivity rate [only between 3 - 24 infected acini in *R. appendiculatus* and 19 - 43 in *R. zambeziensis* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998)]. Low levels of parasitemia, combined with the sequence depth limitations of conventional sequencing technologies, most probably confounded the results obtained by Nene *et al.* (2004) at that time, more

than ten years ago. Improvements in sequencing technologies over the last decade have increased sequence depth considerably (Wang *et al.*, 2009), which would better illuminate *T. parva-R. appendiculatus* infection studies in future.

1.4 Recombinant anti-tick vaccines

In the late 1930s, William Trager conducted a range of ground-breaking experiments that led to some of the biggest breakthroughs in the field of tick biology (Trager, 1939a, 1939b). He showed that guinea pigs acquired immune resistance against *Dermacentor variabilis* ticks, even after a single exposure. Moreover, he found that this acquired resistance could be transferred from a resistant guinea pig to a naive one by passive transfer. His experiments also showed that the acquired resistance cross-protected guinea pigs against a closely related tick species, *D. andersoni*. Trager was also the first to artificially immunise naive guinea pigs by injecting them with *D. variabilis* larvae extract. He also hypothesised that the salivary glands are the most relevant organs during feeding, because they are the only organs in contact with the host. To test this hypothesis, he inoculated naive guinea pigs with salivary gland and gut extracts of partially fed female ticks. He found that the guinea pigs inoculated with the salivary gland extracts showed a greater degree of immune resistance compared to the guinea pigs inoculated with guts. His findings have been replicated multiple times since then and established the concept of recombinant anti-tick vaccines to protect vertebrate hosts against tick infestations.

Forty years later, Allen and Humphreys (1979) showed that immunity could be achieved in the host after exposure to only the antigens of the tick. The authors injected cattle and guinea pigs with extracted antigens from tick guts, which protected the animals from tick challenges. Since then, a number of studies have investigated the protection efficiencies of different recombinant tick proteins, some of which showed very promising efficacy values in hosts (reviewed in Merino *et al.*, 2013a; Lew-Tabor and Rodriguez-Valle, 2016). Recombinant vaccines are attractive vaccines as they are cost-effective, environmentally friendly and result in rapid immune responses (Nascimento and Leite, 2012). Even though recombinant anti-tick vaccines are appealing for the control of ticks, the practical development of the vaccines has proven

to be challenging as only two vaccines have been commercialised to date, both targeting the same Bm86 gut protein of *R. microplus* (Willadsen *et al.*, 1995; Canales *et al.*, 1997). One of these vaccines has already been discontinued due to impracticalities and non-adoption by the Australian beef industry (Lew-Tabor and Rodriguez-Valle, 2016).

1.4.1 'exposed' vs. 'concealed' antigens

It was initially considered that 'concealed' antigens, such as gut proteins that are not presented to the immune system of the host, would perform better as vaccine targets than 'exposed' antigens from the salivary glands (Willadsen *et al.*, 1989; Riding *et al.*, 1994). This was based on the high levels of conservation in 'concealed' antigens that were not under strong selective pressure to evolve in order to evade host immune mechanisms, which can be expected to result in vaccines with broader species effectiveness. An inherent problem with this rationale was that 'concealed' antigens, which are not exposed to the host's immune system, do not induce long-lived immune memory and the vaccines required a number of boosts to achieve optimal protection (García-García *et al.*, 1999; Lew-Tabor and Rodriguez-Valle, 2016). Additionally, 'concealed' antigens influence tick population sizes by affecting tick physiology post meal acquisition, e.g. blood digestion, fecundity and egg survival (Nuttall *et al.*, 2006). This means that ticks still complete feeding, inadvertently transmitting any pathogens they might be carrying. Salivary gland 'exposed' antigens, on the other hand, result in natural immune boosts every time ticks feed on the immunised host and may affect essential tick feeding mechanisms, whereby rejection of ticks can occur during feeding and prior to pathogen transmission (Nuttall *et al.*, 2006). For these reasons, re-interest in 'exposed' antigens from tick salivary glands has been generated in recent years.

1.4.2 Secretory proteins as attractive vaccine targets

Ticks are highly adapted to complete their blood meals (Binnington, 1978) without being noticed by their hosts. The salivary glands, central to this adaptation and the organs in closest proximity to the host during feeding, produce a complicated cocktail of proteins and molecules, referred to as the sialome. These proteins are secreted into the host to create a stable feeding environment, enabling unimpeded blood ingestion and evasion of host immune surveillance. Secretory proteins modulate haemostatic (e.g. vasoconstriction, platelet aggregation and blood coagulation) and immune (e.g.

inflammation, the complement system, T- and B-cell activation) responses of the host (reviewed in Mans and Neitz, 2004; Francischetti *et al.*, 2009; Mans, 2011; Oliveira *et al.*, 2011; Kotál *et al.*, 2015; Mans, 2016; Mans *et al.*, 2016). Continuously exposure to the host's immune system and involvement in crucial functions during feeding make secretory proteins attractive targets for vaccine development. Additionally, secretory proteins facilitate pathogen transmission, lending further support of their attractiveness as potential vaccine candidates (Merino *et al.*, 2013a). The availability of sequences of the proteins expressed in tick salivary glands will support the selection of secretory proteins ('exposed' antigens) as potential anti-tick vaccine candidates in future.

1.4.3 Reverse vaccinology

Traditionally, recombinant vaccines have been developed by empirically screening individual pathogen proteins for eliciting immune responses in the host. In the post-genomic era and with the availability of numerous whole genomes of pathogens, reverse vaccinology has emerged as an effective strategy for high-throughput vaccine development (Rappuoli, 2000; He *et al.*, 2010; Seib *et al.*, 2012). The reverse vaccinology approach relies on bioinformatic tools to search the entire genomic sequence of the pathogen for the prediction of a whole genome repertoire of antigens. Promising antigens are selected and a small subset screened for immunogenic properties in the host (He *et al.*, 2010; Ribas-Aparicio *et al.*, 2017). Some of the advantages of this approach are the reduction in time and cost of developing vaccines and the identification of antigens that might be present in small amounts or only expressed at certain stages during development, which would have hindered their purification based on traditional methods (Ribas-Aparicio *et al.*, 2017). Reverse vaccinology has successfully been applied on a number of different pathogens (reviewed in Seib *et al.*, 2012), and recently the availability of tick genomes has enabled tick biologists to apply this approach in the prediction of tick antigenic candidates (Guerrero *et al.*, 2012; Marcelino *et al.*, 2012; de la Fuente *et al.*, 2016a; Kuleš *et al.*, 2016; Lew-Tabor and Rodriguez-Valle, 2016).

Through the collaborative effort of a global tick consortium, the genome sequence of the first tick species, *Ixodes scapularis*, has recently been completed (de la Fuente *et al.*, 2016c; Gulia-Nuss *et al.*, 2016). Draft genomes have also been released

for *R. microplus* (Barrero *et al.*, 2017) and *I. ricinus* (Cramaro *et al.*, 2017). The large estimated genome sizes of hard tick species, which are in the range 2.0 - 7.1 Gbp (Ullmann *et al.*, 2005; Geraci *et al.*, 2007) and the large proportion of repetitive DNA in their genomes, approximately 70% (Barrero *et al.*, 2017), restrict the sequence and assembly process of tick genomes. This could be seen when the Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Simão *et al.*, 2015) was used to analyse the assembled genomes. BUSCO searches for the presence and completeness of a set of 1066 orthologous protein sequences that are found in all arthropods. Based on the assessment the *I. scapularis* genome is 85% complete, whereas the *R. microplus* and *I. ricinus* genomes are only, respectively 40 and 56% complete. These technical implications, together with the financial requirements involved in sequencing an entire genome, make the sequencing of sialotranscriptomes a more feasible strategy to obtain a gene catalogue in ticks, given the current technologies. Tick sialotranscriptomes can be used as reference sequences in reverse vaccinology antigenic searches. Maruyama *et al.* (2017) described the successful use of sialotranscriptomics and bioinformatics to identify efficacious vaccine candidates in *R. microplus*. An advantage of using sialotranscriptomes to genomes is the added availability of gene expression data generated by transcriptomics, e.g. spatiotemporal gene expression and expression variation due to tick-pathogen or tick-host interactions.

1.4.4 Characterising tick sialotranscriptomes

Conventional cDNA cloning and sequencing to produce expressed sequence tags (ESTs) was an early method of choice for the large scale characterisation of expressed genes in tick salivary glands. However, Mans (2011) observed poor correlation between gene expression data (based on ESTs) and proteomic data and questioned the accuracy of these sequence databases. It was later revealed that the sequence depth achieved by EST sequencing had been insufficient to cover the full complexity of expression in tick salivary glands, resulting in sequence biases towards abundantly expressed genes. One such generated EST sequence database that suffered from these same shortcomings was that of *R. appendiculatus* (Nene *et al.*, 2004). Nevertheless, EST sequencing provided a deep insight into salivary gland complexity of a variety of tick species (Nene *et al.*, 2002; Valenzuela *et al.*, 2002; Nene *et al.*, 2004; Santos *et al.*, 2004; Francischetti *et al.*, 2005b; Ribeiro *et al.*, 2006; Alarcon-Chaidez *et al.*, 2007; Chmelař *et al.*, 2008;

Francischetti *et al.*, 2008; Mans *et al.*, 2008a; Anatriello *et al.*, 2010; Zivkovic *et al.*, 2010a; Ribeiro *et al.*, 2011; Gibson *et al.*, 2013).

With the advent of next generation sequencing (NGS), these shortcomings have been overcome. Next generation sequencing, which produces millions of sequencing reads, achieves sequence depths required to elucidate even lowly expressed genes of complex protein families (Wang *et al.*, 2009). To exploit the advances of NGS technologies, a number of *de novo* tick sialotranscriptomes have recently been generated (Karim *et al.*, 2011; Schwarz *et al.*, 2013; Garcia *et al.*, 2014; Mudenda *et al.*, 2014; Karim and Ribeiro, 2015; Tan *et al.*, 2015a; Xu *et al.*, 2015; Yu *et al.*, 2015; Ong *et al.*, 2016; Martins *et al.*, 2017; Maruyama *et al.*, 2017; Moreira *et al.*, 2017; Ribeiro *et al.*, 2017; Rodriguez-Valle *et al.*, 2017), including the two generated for this PhD study (de Castro *et al.*, 2016; de Castro *et al.*, 2017). The availability of these transcriptomes affords unprecedented insight into tick salivary gland biology and blood feeding evolution. With the transcriptomes sequenced of only a handful of the nearly nine hundred known tick species (Jongejan and Uilenberg, 2004), it can be expected that this knowledge will expand even further in the coming years.

1.4.4.1 Secretory protein families are multi-genic

In recent years with the advancement of NGS, it has been realised that tick salivary glands are much more diverse and contain much larger secretory protein families than initially anticipated (Table 1.1) (reviewed in Mans, 2016; Mans *et al.*, 2016). It has been shown that secretory proteins are under positive selection (Dai *et al.*, 2012; Kotsyfakis *et al.*, 2015b) and have been subjected to gene duplications, resulting in lineage-specific expansions (Mans and Neitz, 2004; Mans *et al.*, 2008a; Francischetti *et al.*, 2009; Mans, 2011; Mans *et al.*, 2017) and large multi-genic functionally redundant protein families (Chmelař *et al.*, 2016; Mans, 2016). The release of the first NGS tick sialotranscriptome in 2011 nearly doubled the total number of tick protein sequences available for ticks at that time (Karim *et al.*, 2011; Mans, 2016). Since then it has been shown that tick salivary glands express thousands of proteins, of which around one to three thousand can be classified as secretory proteins, that contain multi-genic families that may have as many as 500 members (Table 1.1). The multi-genic secretory protein families have shown to contain some of the most abundantly expressed genes in the salivary glands,

indicating their functional importance during feeding and host immune evasion (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Ribeiro *et al.*, 2017). Whether these large protein numbers are a true reflection of the number of genes in the genomes, the number of transcripts being expressed for each of the genes, or an artefact of short read NGS technologies (Martin and Wang, 2011), need to be addressed in future when sequencing of tick genomes become a more practical option.

1.4.4.2 Dynamic expression and antigenic variation of secretory proteins

Members of multi-genic secretory protein families have been shown to alternate expression during different feeding phases (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Bullard *et al.*, 2016; Kim *et al.*, 2016). These dynamic expression patterns have been proposed as a manner by which the tick changes its antigenicity presented to the host's immune system, while maintaining the function required during feeding as members of multi-genic families have shown to be functionally redundant (Chmelař *et al.*, 2016; Mans, 2016). The alternating expression of secretory protein family members has been termed 'sialome switching', a form of antigenic variation in ticks (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Bullard *et al.*, 2016; Kim *et al.*, 2016). Antigenic variation has been well described before in microorganisms that systematically change their surface proteins to remain undetected by the host immune system (reviewed in Deitsch *et al.*, 2009). In addition to alternating expression of secretory protein family members, ticks can also weakly express a number of family members concurrently. This could result in immune evasion due to low immunogenicity caused by each family member, while the additive effect of all the expressed members results in the function being retained (Chmelař *et al.*, 2016). Functionally redundant multi-genic secretory protein families will constrain vaccine development, where immunisation against a protein in a multi-genic family may be bypassed by the expression of other functionally similar members from the same family (Guerrero *et al.*, 2012). This indicates that clever design strategies will be required to develop effective vaccines against secretory protein families and therefore better insight into the complexities of tick salivary glands are required.

Table 1.1 Summary of tick transcriptomes generated by next generation sequencing technologies indicating the number of protein sequences in the dataset, the number of secretory and housekeeping proteins and the number of proteins in four of the largest expanded multi-genic tick secretory protein families. The table has been amended from our previous work (Mans *et al.*, 2016) and updated with additional tick transcriptomes released since then.

Species	Bioproject	Tissue source[a]	Protein Nr[b]	SP[c]	HKP[c]	KI[c]	BT[c]	LC[c]	MP[c]	Reference[d]
Ixodidae: Metastriata										
Amblyomma americanum	PRJNA218793	SG	3139	849	1814	110	34	213	71	(Karim and Ribeiro, 2015)
A. maculatum	PRJNA72241	SG	4849	886	3325	88	24	304	147	(Karim *et al.*, 2011)
A. cajennense	PRJNA241272	SG	5770	1362	3479	187	125	275	109	(Garcia *et al.*, 2014)
A. parvum	PRJNA241271	SG	2838	476	1977	61	19	109	38	(Garcia *et al.*, 2014)
A. triste	PRJNA241269	SG	8098	1761	4991	135	87	589	115	(Garcia *et al.*, 2014)
A. aureolatum	PRJNA344771	SG,G	7999	643	6413	50	23	101	61	(Martins *et al.*, 2017)
A. sculptum	PRJNA343654	SG,G	4246	240	3516	20	7	30	28	(Martins *et al.*, 2017)
A. sculptum	PRJNA309641	SG,G,O	16248*	543	5543	22	5	63	37	(Moreira *et al.*, 2017)
Rhipicephalus microplus	LYUQ00000000	Genome	24758	1018	17053	62	34	86	143	(Barrero *et al.*, 2017)
R. microplus	PRJNA288687	GO	34028*	422	7427	11	7	25	28	(Tidwell *et al.*, unpublished)
R. pulchellus	PRJNA170743	SG	11227	1414	8160	196	39	331	107	(Tan *et al.*, 2015a)
R. annulatus	PRJNA255770	SG	17536*	437	4111	25	6	53	13	(de la Fuente *et al.*, unpublished)
R. sanguineus	PRJEB8914	L	34944*	888	12949	44	27	65	91	(De Marco *et al.*, 2017)
Hyalomma excavatum	PRJNA311286	SG	5337	415	4404	28	14	72	45	(Ribeiro *et al.*, 2017)

Ixodidae: Prostriata										
Ixodes scapularis	ABJB000000000	Genome	20486	1368	13042	99	49	65	342	(Gulia-Nuss *et al.*, 2016)
I. ricinus	PRJNA177622	SG	8685	3882	3537	512	310	564	564	(Schwarz *et al.*, 2013)
I. ricinus	PRJNA217984	SG,G	16002	3670	9891	478	406	568	427	(Schwarz *et al.*, 2014b; Kotsyfakis *et al.*, 2015b)
I. ricinus	PRJNA183509	HL	2854	625	1712	47	66	126	48	(Kotsyfakis *et al.*, 2015a)
I. persulcatus	PRJNA263101	WB	35300*	866	7920	75	33	63	82	(Zhang *et al.*, unpublished)
Argasidae										
Ornithodoros rostratus	PRJNA270484	SG,G	16299*	396	5600	21	22	34	60	(Araujo *et al.*, unpublished)
O. turicata	PRJNA281459	S	52817*	853	13328	41	26	41	188	(Egekwu *et al.*, 2016)
O. moubata	PRJNA377416	G	8493	265	7189	10	4	1	50	(Oleaga *et al.*, 2017)
O. mimon	PRJNA310605	G	1389	70	986	3	0	0	3	(Landulfo *et al.*, 2017)

[a] Tissue source of transcriptomes: SG: salivary glands; G: gut; HL: haemolymph; L: larvae; GO: gene's organ; S: synganglia; WB: whole body; O: ovaries.

[b] Protein Nr refers to the number of proteins downloaded from the Bioproject accession. In cases where proteins were not available, transcript sequences were downloaded, open reading frames larger than 240 nucleotides predicted, translated into proteins and redundancy removed. These transcriptomes are indicated by *.

[c] Protein classification was based on BLASTp analysis against an in-house Acari database containing all available mite and tick sequences (as described in de Castro *et al.*, 2016). SP: secretory protein; HKP: Housekeeping protein; KI: Kunitz/BPTI inhibitor; BT: Basic tail secretory protein; LC: Lipocalin; MP: Metalloprotease.

[d] The references that are indicated as 'unpublished' are of tick transcriptomic datasets that have been released without associated publications.

1.4.4.3 Predominant secretory protein families in tick salivary glands

Next generation sequencing has provided the ability to obtain members of multi-genic families with even dynamic expression ranges, which is resulting in much better characterisation of these large protein families. From tick sialotranscriptomes that have been sequenced before (see Table 1.1 and references therein), it has been observed that the largest multi-genic secretory protein families are the Kunitz domain/Bovine pancreatic trypsin inhibitor (BPTI), Basic tail secreted protein, Lipocalin and Metalloprotease families. In the following section these important protein families will be discussed briefly.

The initial host immune defences, including cellular immunity, inflammation, blood clotting, platelet aggregation and complement activation are mediated by proteases. Ticks modulate these processes by secreting protease inhibitors into the feeding site. A number of protease inhibitors have been found in tick salivary glands and of these, Basic tail secreted proteins and Kunitz domain/BPTIs have been shown to be the largest families. Basic tail secreted proteins contain repeats of basic amino acids on their carboxyl-termini that could potentially associate with negatively charged phospholipids (Stevenson and Poller, 1982; Andersen *et al.*, 2004) serving as scaffolds of proteinase complexes (Francischetti, 2010). The family was initially identified in abundance in *I. scapularis* and shown to inhibit the Xa clotting factor (Narasimhan *et al.*, 2002). The family is also expanded in other Prostriates, but smaller family sizes are found in Metastriates (Table 1.1).

Bovine pancreatic trypsin inhibitor was the first Kunitz domain-containing protein to be characterised functionally (Kunitz and Northrop, 1936). The Kunitz domain consists of about 50 - 60 amino acids, typically containing six cysteine residues that form three disulfide bonds, although some variations have been observed (Paesen *et al.*, 2007; Paesen *et al.*, 2009; Valdés *et al.*, 2013). Kunitz domain-containing proteins have been found in a number of tick species and about 15 have been functionally characterised in ticks (described by Schwarz *et al.*, 2014a). Tandem repeats of the domain occur in ticks and based on the number of domains, the proteins are classified as monolaris, bilaris, trilaris and so forth. A monolaris protein from *R. appendiculatus* was found to function as a potassium channel blocker (Paesen *et al.*, 2009) and bi- and pentalaris proteins from *I. scapularis* function as clotting inhibitors of the extrinsic

pathway (Francischetti *et al.*, 2002; Francischetti *et al.*, 2004; Monteiro *et al.*, 2005; Nazareth *et al.*, 2006; Monteiro *et al.*, 2008).

Lipocalins are small proteins widely found in nature that share limited sequence homology, although conserved regions are found that differentiate family members (Flower *et al.*, 1993). The structure of Lipocalins, consisting of an eight-stranded antiparallel beta-sheet, is however highly conserved (Flower, 1996). The structure typically results in a lipophilic barrel-like structure that binds lipids (Paesen *et al.*, 1999; Flower *et al.*, 2000; Mans *et al.*, 2008b), hence the name Lipocalin literally meaning 'cup of lipid'. Apart from binding lipid compounds, such as leukotrienes and thromboxane A2 (Beaufays *et al.*, 2008; Mans and Ribeiro, 2008a, 2008b), in ticks Lipocalins have also shown to bind biogenic amines, such as serotonin and histamine (Paesen *et al.*, 1999; Paesen *et al.*, 2000; Sangamnatdej *et al.*, 2002; Mans, 2005; Mans and Ribeiro, 2008a; Mans *et al.*, 2008b). Lipocalins are therefore involved in evasion of host immune surveillance by scavenging the agonists of inflammation. Additionally, Lipocalins have also shown functionalities in the inhibition of the clotting (Ribeiro *et al.*, 1995) and complement (Nunn *et al.*, 2005; Mans and Ribeiro, 2008a) systems.

Metalloproteases can be classified into two main families: Neprilysin-like or membrane metallo-endopeptidase proteases and Reprolysin-like or ADAM (short for A Disintegrin And Metalloproteinase) proteases. Neprilysin-like metalloproteases are typically extracellular membrane-bound proteins that function in the inactivation of hormone peptides in mammalian systems (Turner *et al.*, 2000). In ticks, a protein in this family has been shown to degrade bradykinin, a mediator of inflammation (Ribeiro and Mather, 1998). These proteases have also been associated with *Borrelia* transmission and dissemination in the vertebrate host (Gebbia *et al.*, 2001). The predominant metalloprotease family in tick salivary glands are Reprolysins. Reprolysin-like metalloproteases are ubiquitously found in animals and are involved in a number of functions, including connective tissue remodelling, embryonic development, inflammation and angiogenesis (Wolfsberg *et al.*, 1995). Reprolysins in ticks have been implicated in disrupting the haemostatic system of the host by exhibiting fibrin(ogen)olytic and gelatinase activities, inhibition of angiogenesis and disaggregation of platelets (Francischetti *et al.*, 2003, 2005a; Harnnoi *et al.*, 2007; Decrem *et al.*, 2008a). These functions are associated with feeding site maintenance and

might suggest that tick Reprolysins act directly on the extracellular matrix of the host (Alarcon-Chaidez *et al.*, 2007; Harnnoi *et al.*, 2007). The interference of Reprolysins with blood clotting has also been observed in snake and spider venoms, where members of the family are involved in activities that disrupt homeostasis and cause haemorrhaging in the prey (Hati *et al.*, 1999; Ramos *et al.*, 2003).

1.4.5 Vaccine developments in tick secretory proteins

Secretory proteins expressed in the salivary glands during feeding are known as 'exposed' antigens due to their presentation to the host immune system. Several studies using 'exposed' antigens isolated from salivary glands or saliva have shown great promise towards the protection of vertebrate hosts against tick infestations from, for example: *Haemaphysalis longicornis* (Mulenga *et al.*, 1999), *Ornithodoros moubata* (García-Varas *et al.*, 2010), *R. microplus* (Merino *et al.*, 2013b; Ali *et al.*, 2015), and *I. ricinus* (Prevot *et al.*, 2007; Decrem *et al.*, 2008b). Several studies that reported protection in vertebrates from *I. scapularis* also reported reduced levels of pathogen transmission, when the tick histamine release factor (tHRF) (Dai *et al.*, 2010), salivary protein of 15kDa (Salp15) (Dai *et al.*, 2009) and tick salivary lectin pathway inhibitor (TSLPI) (Schuijt *et al.*, 2011) proteins were targeted. Studies involving *R. appendiculatus* have also been performed. Protection was shown using a putative *R. appendiculatus* cement protein, 64P (Trimnell *et al.*, 2002), which also resulted in cross-protection against *R. sanguineus* and *I. ricinus* infestations (Trimnell *et al.*, 2005), indicating its potential as a broad-spectrum anti-tick vaccine. Further, during the protection from *I. ricinus* ticks, the vaccine also restricted pathogen transmission (Labuda *et al.*, 2006). Another two *R. appendiculatus* proteins, an Immunoglobulin-binding protein (IGBP) (Wang *et al.*, 1998) and the *Rhipicephalus* Immuno-dominant Molecule 36 (RIM36) protein (Bishop *et al.*, 2002) resulted in strong antibody responses in the host, signifying their potential as vaccine candidates.

Even though a number of vaccine candidates have been identified in recent years, none of them have entered the commercial market (Guerrero *et al.*, 2012). It could be speculated that this might be due to partial protection or limited cross-species protection achieved by these vaccine candidates. Multi-antigen vaccines that combine a number of partially protective antigens have been suggested as a better alternative for

the production of effective anti-tick vaccines (Parizi *et al.*, 2012). Furthermore, as mentioned in the previous section, a number of secretory protein families, from which 'exposed' antigen vaccines are developed, are known to be large, functionally redundant, multi-genic families (Chmelař *et al.*, 2016; Mans, 2016) that alter antigenicity while maintaing protein function (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Bullard *et al.*, 2016; Kim *et al.*, 2016). It could therefore be assumed that the tick can alternate its protein expression from one targeted by a vaccine-immunised host to a functionally redundant member not recognised by the host, thereby escaping the immune response elucidated by the vaccine and resulting in partial or no protection (Guerrero *et al.*, 2012). The availability of comprehensive tick sialotranscriptomes, containing sequences of entire multi-genic protein families, will assist in this regard by revealing common protein regions in the families. This could redirect anti-tick vaccine development towards targeting common antigens of protein families rather than antigens of single proteins alone.

1.5 Control of Corridor disease in South Africa

The Department of Agriculture, Forestry and Fisheries (DAFF) have put strict regulations in place to prevent the spread of CD in South Africa (Animal Disease Act 1984, Act No. 35). These regulations control the cattle-wildlife interface and restrict the movement of *T. parva* infected buffalo to non-endemic regions. It requires veterinary control of buffalo in endemic regions, securely fenced and registered farms where buffalo are kept, frequent dipping of cattle and strict animal monitoring to identify early signs of an outbreak. Despite these regulations, outbreaks of CD still occur in South Africa (Thompson *et al.*, 2008; Mbizeni *et al.*, 2013).

Chemical acaricides are still the most widely used control measure against ticks (Willadsen, 2006). Regular tick dipping programmes are strictly enforced in South Africa, but even after more than a century of dipping, ticks have not been eradicated from the country or even diminished in population numbers. This can mainly be ascribed to the high level of acaricide-resistance observed in tick populations continuously exposed to chemical acaricides (Mekonnen *et al.*, 2002; Abbas *et al.*, 2014). Due to increased acaricide-resistance and concerns of chemical residues lingering

in the environment and potentially in the milk and meat products of the cattle (Graf *et al.*, 2004), alternative measures of tick control are being explored.

A live attenuated vaccine, the Muguga cocktail vaccine, has been developed to protect against ECF (Radley *et al.*, 1975). The vaccine, developed from three *T. parva* strains, Muguga, Kiambu-5 and Serengeti-transformed, provides broad-spectrum immunity against ECF and is used effectively in a number of eastern African countries (Uilenberg *et al.*, 1977; Morzaria *et al.*, 2000). The vaccine uses the 'infection-and-treatment' method (ITM) in which the cattle are injected with live *T. parva* sporozoites and simultaneously treated with long-acting oxytetracycline. The animals show mild disease symptoms, then develop immunity during recovery and become carriers for the specific strains of *T. parva* (Radley *et al.*, 1975). A major concern of the vaccine, impeding its wide-scale implementation, is the introduction of *T. parva* strains into regions where they were not previously found (McKeever, 2007). Additionally, the vaccine is difficult and expensive to produce and requires a liquid nitrogen cold chain (Nene *et al.*, 2016). Furthermore, when investigating the efficacy of the vaccine against buffalo-derived *T. parva*, only limited protection (Radley *et al.*, 1979) or no protection (Sitt *et al.*, 2015) was observed and the vaccinated cattle died of CD (Sitt *et al.*, 2015). Due to the potential risk that buffalo-derived *T. parva* might transform to cattle-derived *T. parva*, it is not permitted by law to use any treatments or vaccines that will result in a *T. parva* carrier-state in animals in South Africa (Animal Disease Act 1984, Act No. 35). Similarly, great success has been obtained by the treatment of ECF using anti-theileria drugs, such as Buparvaquone (Butalex, Coopers Animal Health), but the treatment also results in carrier-state in recovered cattle (Dolan, 1986; Potgieter *et al.*, 1988), and accordingly cannot be implemented in South Africa. By law, cattle that recover from CD infection have to be slaughtered in South Africa.

A potential alternative control strategy for managing CD in South Africa is the development of recombinant vaccines, which would reduce the risk of generating a carrier-state in the protected cattle and has a smaller likelihood of resulting in tick resistance. Research towards the development of recombinant vaccines against *T. parva* has shown that the surface protein, p67, showed a great deal of promise as a vaccine candidate due to its low diversity in a set of cattle-derived *T. parva* strains (Nene *et al.*, 1996; Nene *et al.*, 2016). Due to the high recombination rate and sequence divergence

observed among *T. parva* strains (Gardner *et al.*, 2005; Katzer *et al.*, 2011; Henson *et al.*, 2012; Hayashida *et al.*, 2013; Norling *et al.*, 2015), an efficient vaccine candidate will need to protect against a number of these strains and therefore proteins showing lower variability is advantageous. However, when buffalo-derived *T. parva* strains were investigated, p67 showed a high degree of sequence diversity (Sibeko *et al.*, 2010), raising doubts about its potential efficacy as a vaccine candidate against buffalo-derived *T. parva* and CD.

An alternative strategy to vaccine design against *T. parva* is to develop a vaccine that will protect against its vectors. An advantage of such an approach is that the immunised cattle would be protected against all tick-borne diseases carried by the vector, even the yet undiscovered ones (Willadsen, 2004; Nuttall *et al.*, 2006). Accordingly, recombinant vaccine development against tick vectors has become a very attractive strategy and some candidates have emerged that show great promise (reviewed in Merino *et al.*, 2013a; Lew-Tabor and Rodriguez-Valle, 2016), including a number of *R. appendiculatus* proteins (Wang *et al.*, 1998; Bishop *et al.*, 2002; Trimnell *et al.*, 2002). None of these vaccines have been commercialised to date and partial protection due to alternating expression of functionally redundant multi-genic family members might be a reason for this. To develop more efficacious anti-tick vaccines, a comprehensive understanding of tick blood feeding, evasion of host immune defences and transmission of pathogens are required. This level of understanding will require discovery of all of the genes involved in these processes together with their expression dynamics and NGS technologies is a promising approach in this regard.

1.6 Justification and Research Aims of the study

Corridor disease is an economically important cattle disease in South Africa, which is controlled by the Department of Agriculture, Forestry and Fisheries. Yet, sporadic outbreaks occur where cattle and buffalo interact and in recent years the expansion of the South African game industry has put cattle at a re-emerging risk of contracting CD. *Theileria parva* causes CD and is spread by *R. appendiculatus* and *R. zambeziensis* from buffalo to cattle. There is variation in the *T. parva* transmission rates of the vectors and climate change might increase the risk of the tick with higher vector competence, *R.*

zambeziensis, spreading to a wider distribution. Chemical acaricides have been predominantly used for vector control, but tick-resistance is becoming a serious problem, causing the investigation of alternative control measures. The potential risk of CD changing aetiology from buffalo-cattle transmission to cattle-cattle transmission is of great concern in South Africa. Therefore, vaccines or treatment that causes a *T. parva* carrier-state in cattle that might become infective to other cattle, is by law not permitted to be used in South Africa. This makes the development of recombinant anti-tick vaccines that will protect the cattle without resulting in a carrier-state, an important research endeavour. The development of recombinant vaccines requires in depth understanding of tick feeding biology and proteins involved in evading the immune responses of the host. The first step towards this goal is to develop complete sequence datasets of transcripts from salivary glands, which would allow the identification of potential vaccine candidates. Furthermore, *R. appendiculatus* and *R. zambeziensis* differ in their disease transmission and it would be of interest to investigate whether the reason for this could be explained by the compositions of their salivary glands, as much of *T. parva*'s vector-based life cycle occurs in these organs. These findings will assist in the control of CD in South Africa.

Research Aim 1: To assemble and annotate the *R. appendiculatus* sialotranscriptome during feeding.

Research Aim 2: To assemble and annotate the *R. zambeziensis* sialotranscriptome during feeding.

Research Aim 3: To compare the sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* to determine whether morphological and behavioural differences between the species can be explained on a molecular level.

CHAPTER 2

De novo assembly and annotation of the salivary gland transcriptome of *Rhipicephalus appendiculatus* male and female ticks during blood feeding

Minique H. de Castro [1,2,3], Daniel de Klerk [3], Ronel Pienaar [3,4], Abdalla A. Latif [3,5], D. Jasper G. Rees [1,2] and Ben J. Mans [1,3,4]

[1] *College of Agriculture and Environmental Sciences, University of South Africa, Johannesburg, South Africa*

[2] *Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa*

[3] *Epidemiology, Parasites and Vectors, Onderstepoort Veterinary Research, Agricultural Research Council, Onderstepoort, South Africa*

[4] *Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa*

[5] *School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa*

This chapter is largely based on work that was published in a peer-reviewed journal:

de Castro MH, De Klerk D, Pienaar R, Latif AA, Rees DJG & Mans BJ (2016) Ticks and Tick-Borne Diseases. 7, 536-548. doi: 10.1016/j.ttbdis.2016.01.014.

The version presented here is not the final print version.

2.1 Abstract

Tick secretory proteins modulate inflammation, haemostasis and immune responses of vertebrate hosts and are attractive candidates for recombinant anti-tick vaccines. Yet, many have not been characterised due to the limited number of sequences available for homology-based annotation of arthropods, such as ticks. To address this limitation, the salivary glands of the economically important *Rhipicephalus appendiculatus* tick was sequenced during feeding. The Illumina RNA sequencing reads were quality filtered and *de novo* assembled to construct a *R. appendiculatus* sialotranscriptome of 21 410 transcripts. A non-redundant set of 12 761 *R. appendiculatus* proteins was predicted from the transcripts, including 2134 putative secretory and 8237 putative housekeeping proteins. Secretory proteins accounted for most of the expression in the salivary gland transcriptome (63%). Of the secretory protein class, the Glycine rich superfamily contributed 66% and the Lipocalin family 12% of the transcriptome expression. Differential expression analysis identified 1758 female and 2346 male up-regulated transcripts, suggesting varying blood feeding mechanisms employed between female and male ticks. The sialotranscriptome assembled in this work, greatly improves on the sequence information available for *R. appendiculatus* and is a valuable resource for potential future vaccine candidate selection.

Keywords: *Rhipicephalus appendiculatus*; Next generation sequencing; *de novo* transcriptome assembly; Sialotranscriptomics; Salivary glands; Secretory proteins.

2.2 Introduction

Rhipicephalus appendiculatus, one of the most economically important tick species in Africa, transmits the protozoan parasite *Theileria parva*, that causes the related cattle diseases: Corridor disease (CD), East Coast fever (ECF) and January disease (Stoltsz, 1989; Lawrence *et al.*, 1994; Uilenberg, 1999). The *T. parva* parasite that causes CD is transmitted from African buffalo (*Syncerus caffer*) to cattle and is appropriately referred to as buffalo-derived *T. parva* (Lawrence *et al.*, 1994; Uilenberg, 1999). Corridor disease is believed to be self-limiting in cattle due to the parasite's inability to be transmitted from cattle to cattle (Norval *et al.*, 1992). The potential transformation of buffalo-derived *T. parva* to cattle-derived *T. parva* (cattle transmission enabled) will have serious implications for the control of CD in South Africa (Yusufmia *et al.*, 2010).

Due to the potential risk of transformation, treatment or vaccination resulting in a *T. parva* carrier-state in cattle, that might become infective to other cattle, is not permitted to be used in South Africa. The Muguga cocktail vaccine, used in other Africa countries to protect cattle against *T. parva*, is based on the 'infection-and-treatment' method and results in a carrier-state in the recovered animals (Radley *et al*., 1975; Boulter and Hall, 1999). Similarly, the anti-theileria drug, Buparvaquone also results in a carrier-state in recovered cattle (Dolan, 1986; Potgieter *et al.*, 1988). Accordingly, neither of these can be implemented in South Africa. Therefore, the only manner by which cattle can be protected from CD in South Africa is by means of recombinant anti-tick vaccines, as these vaccines cannot result in *T. parva* carrier-state in cattle.

The development of a recombinant anti-tick vaccine requires a comprehensive understanding of tick blood feeding behaviour, host immune evasion and modulation, and the genes involved in these processes. Additionally, vaccine development requires comprehensive genome or transcriptome sequence databases for use as references in reverse vaccinology approaches (Seib *et al.*, 2012). Previously, an expressed sequence tag (EST) library was generated from *R. appendiculatus* salivary glands, the *R. appendiculatus* gene index (RaGI) (Nene *et al.*, 2004). Due to technical limitations of the technology at the time, transcripts below 1000 nucleotides were excluded from the dataset resulting in underrepresentation of smaller genes and an incomplete transcriptome (Nene *et al.*, 2004). Additionally, with the availability of recent tick sialotranscriptomic studies (Karim *et al.*, 2011; Schwarz *et al.*, 2013; Garcia *et al.*, 2014; Mudenda *et al.*, 2014; Tan *et al.*, 2015a; Xu *et al.*, 2015), it has been realised that conventional EST sequencing could not achieve the sequence depths required to ascertain all the complexities found in tick salivary gland transcriptomes. The advancement of next generation sequencing (NGS) provided the possibility to generate millions of sequence reads cost effectively (Collins *et al.*, 2008; Ekblom and Galindo, 2011) and offered enough sequence depth to assemble complete tick salivary gland transcriptomes that represent all the expressed genes.

The aim of the current study was then to sequence and *de novo* assemble a representative, comprehensive gene catalogue of *R. appendiculatus* using NGS technologies that improves the publically available sequence information of this species. Further, genes putatively involved in blood feeding were identified by investigating the

expression abundance and differential expression of secretory protein families between female and male ticks. This is the first report of a *de novo* assembled and annotated sialotranscriptome of *R. appendiculatus* using NGS. The sialotranscriptome is an invaluable resource to the tick community and will facilitate future comparative studies with other tick transcriptomes to elucidate the biology of tick feeding and aid in resolving complex tick protein families. Moreover, this transcriptome will be a valuable source for future vaccine candidate selection to protect South African cattle from CD.

After the completion and publication of the current study, an *R. appendiculatus* salivary gland transcriptome was released on NCBI (Bioproject PRJNA309182, Jore *et al*., unpublished). To date the transcriptome has not been annotated or published and is of limited value in its current state.

2.3 Methods

2.3.1 Ethics statement

All animals used in this study were housed at the Agricultural Research Council - Onderstepoort Veterinary Research (ARC-OVR) Institute. Ethics approval was obtained from the University of South Africa, College of Agriculture and Environmental Sciences - Animal Ethics Review Committee (approval number: 2014/CAES/098) and from the Onderstepoort Veterinary Research Institute - Animal Ethics Committee (approval numbers: AEC01.15 and AEC12.11) Approval was further obtained from the Department of Agriculture, Forestry & Fisheries to do research in terms of Section 20 of the Animal diseases act, 1984 (Act no. 35 of 1984, approval number: 12/11/1/1). Approval letters can be found in Appendix A.

2.3.2 Ticks

Ticks were obtained from a parasite-free colony initiated from adult *R. appendiculatus* ticks collected from the Marakele National Park, South Africa. The colony was maintained under standard tick-rearing procedures at ARC-OVR: temperature of 26 °C (± 1 °C), relative humidity of 75% (± 5%) and photoperiod of 12-hour light/ 12-hour dark (according to Heyne *et al.*, 1987). Adult ticks were fed in customised feeding bags on the backs of disease-free Hereford cattle (*Bos taurus*).

2.3.3 Salivary gland dissection and RNA extraction

About 20 male and 20 female ticks were carefully removed from the bovine at different times during feeding (two and five days post attachment), without disrupting their mouthparts. Unfed male and female ticks were also obtained from the laboratory colony. Ticks were dissected under a stereomicroscope using sterile conditions and salivary glands removed and stabilised in RNAlater (Qiagen, Valencia, CA) according to the manufacturer's specifications. Salivary glands were pooled by sex, resulting in one sample for female and one sample for male ticks. Total RNA was extracted from each pooled sample using the RNeasy Protect Mini Kit (Qiagen) followed by residual genomic DNA removal with *DNase I* (Qiagen) digestion. RNA quantity was estimated using the Qubit fluorometer 2.0 (Life Technologies, Carlsbad, CA) and RNA integrity using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

2.3.4 Library preparation and RNA sequencing

Two slightly different library preparation procedures were followed due to differences in the read lengths generated by the HiSeq 2000 and MiSeq Illumina instruments. For HiSeq 2000 sequencing libraries, total RNA was used in the TruSeq RNA Sample Preparation kit (Illumina, San Diego, CA) according to the manufacturer's specifications. Briefly, poly-A mRNA was isolated, fragmented (for 8 minutes), converted to double stranded cDNA, followed by adaptor ligation and amplification. The final libraries were size selected by agarose gel electrophoreses, to excise the ± 300 bp fragment fractions. For MiSeq library preparation, the RNA samples were pooled and fragmented for a shorter time (3 minutes, to facilitate the generation of longer fragments) followed by excision of a high molecular weight fraction (± 600 - 1200 bp). Sequencing was performed at the next generation sequencing facility of the ARC - Biotechnology Platform (South Africa).

2.3.5 Read quality filtering and *de novo* transcriptome assembly

Illumina adaptor sequences and low quality bases were removed from the sequence reads using cutadapt v1.0, parameters: -e 0.02 -O 5 -m 20 -q 20 (Martin, 2011) and the FASTQ Quality Filter package of the FASTX-Toolkit v0.0.13, parameters: -q 20 -p 95 (hannonlab.cshl.edu/fastx_toolkit/). The quality-filtered sequence reads of each sample were pooled to generate a single transcriptome assembly of the *R. appendiculatus*

salivary glands representing both sexes. The Trinity software package, release 2014-07-17, was used to *de novo* assemble transcripts longer than 300 bp, at k-mer size of 25 (Grabherr *et al.*, 2011; Haas *et al.*, 2013). A minimum k-mer coverage of two was used (to reduce erroneous k-mers being built into the de Bruijn graphs) and transcripts with low expression levels (that likely represent artefacts) were removed by filtering with a Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) value of one (Mortazavi *et al.*, 2008). By using an expression level threshold (FPKM value ≥ 1) as a proxy for functionally active transcripts, transcripts with higher confidence were selected above background expression or incorrectly assembled transcripts (Gan *et al.*, 2010; Hebenstreit *et al.*, 2011). No assumptions are made that true transcription does not occur below this threshold, just that at such low levels it is not easily distinguishable from background noise.

2.3.6 Transcriptome assembly quality assessment

Internal validation was performed by mapping the paired end sequence reads back to the transcriptome [reads mapped back to transcript (RMBT)] using Bowtie2 v2.2.3 (Langmead and Salzberg, 2012) to estimate whether the transcriptome represented the reads. For external validation, an EST dataset of 7970 *R. appendiculatus* gene index (RaGI) sequences (Nene *et al.*, 2004) was BLASTn (Basic Local Alignment Search Tool) aligned (E-value $\leq$ E-20) to a local sequence database of the assembled transcripts. Four reference-based metrics were generated as proposed by Martin and Wang (2011): accuracy (percentage of identical bases between the transcripts and reference alignment), completeness (percentage of reference sequences that have more than 80% of their lengths covered by the transcriptome), contiguity (percentage of reference sequences that are represented by a single longest transcript covering more than 80% of the length of reference) and chimerism (percentage of transcripts that aligned to more than one reference sequence over more than 80% of the reference length). The transcriptome completeness was also measured by the Core Eukaryotic Genes Mapping Approach (CEGMA v2.5), which uses hidden Markov models (HMMs) to search for the presence of 248 ultra-conserved core Eukaryotic genes (CEGs) in the transcriptome (Parra *et al.*, 2007).

2.3.7 Transcriptome annotation

Transcriptome annotation was performed using BLASTx similarity searches (E-value < E-05) against a number of protein sequence databases: NCBI non-redundant (NR) database (retrieved 23/12/2014), UniProt Knowledgebase translated EMBL-Bank (UniProtKB/TrEMBL, retrieved 23/12/2014), predicted peptides from the *Ixodes scapularis* genome (IscaW1.4, retrieved 20/05/2015, www.vectorbase.org), all *Rhipicephalus* protein sequences from NCBI (retrieved 10/06/2015), an in-house curated database of available Acari (mites and ticks) protein sequences from NCBI and the EuKaryotic Orthologous Groups (KOG) dataset (Tatusov *et al.*, 2003), retrieved 04/06/2015 (ftp.ncbi.nih.gov/pub/mmdb/cdd/little_endian). The search results obtained from the BLASTx alignment to the NR database was submitted to the BLAST2GO software package (Conesa *et al.*, 2005) to retrieve Gene Ontology (GO) terms and Enzyme Commission (EC) numbers. Web Gene Ontology Annotation Plot (WEGO) was used to visualize the GO terms (level 2) present in more than 100 transcripts (Ye *et al.*, 2006). Kyoto Encyclopedia of Genes and Genomes (KEGG) mapping was used for pathway analysis and KEGG Automatic Annotation Server (KAAS) to assign *I. scapularis* KEGG orthology (KO) identifiers to the transcripts (Moriya *et al.*, 2007). The protein-coding potential of the transcripts were determined by the Coding Potential Calculator (CPC) package (Kong *et al.*, 2007), Coding-Potential Assessment Tool (CPAT v1.2.2) package (Wang *et al.*, 2013) and Predictor of lncRNAs and mRNAs based on k-mer scheme (PLEK v1.2) package (Li *et al.*, 2014).

2.3.8 Open reading frame prediction, annotation and evaluation

Open reading frames (ORFs) of the transcripts were predicted using the orffinder.pl script (github.com/vikas0633/perl). Conserved domains were identified by similarity searches against the Pfam database (Finn *et al.*, 2016) and NCBI's Conserved Domain Database (CDD) (Marchler-Bauer *et al.*, 2015). Putative signal peptide and transmembrane topology were predicted using the SignalP 4.0 (Petersen *et al.*, 2011), Phobius (Käll *et al.*, 2007) and TMHMM 2.0 (Krogh *et al.*, 2001) servers. The predicted amino acid sequences were BLASTp aligned against the same protein search databases stated above. A priority order of Acari database, NR, UniProtKB/TrEMBL, *I. scapularis* proteins, *Rhipicephalus* proteins, Pfam database and CDD database was used for annotation. Predicted proteins were kept in the dataset if a significant BLASTp or

domain-based match was obtained. Lastly, CD-HIT v4.5.4 (Li and Godzik, 2006) was used to remove the shortest of two or more amino acid sequences at 100% similarity to reduce redundancy in the final set of predicted proteins. To evaluate similarity to other tick proteins, the predicted *R. appendiculatus* proteins were compared against two known tick datasets: predicted peptides from the *I. scapularis* genome (IscaW1.4), representative of a 'near complete' tick genome (Pagel Van Zee *et al.*, 2007) and predicted protein sequences from the *R. pulchellus* transcriptome (NCBI Bioproject PRJNA170743) (Tan *et al.*, 2015a), representative of proteins expressed in tick salivary glands during feeding.

2.3.9 Tick protein family characterisation

An in-house curated database of Acari (tick and mite) protein sequences (AcariDB) was used for tick protein family characterisation. The Acari database consisted of 166 901 protein sequences downloaded from NCBI (retrieved 01/08/2014 and updated as new sequences were released). In cases of sequences only available in EST datasets, putative ORFs were predicted using orffinder.pl as described. For Acari database annotation, protein sequences were submitted to the KEGG database (Moriya *et al.*, 2007), while secretory families were manually annotated using Position-Specific Iterative (PSI)-BLAST analysis (Altschul *et al.*, 1997). The annotated Acari sequences were transformed into a local BLAST database and non-annotated Acari sequences were aligned to these by BLASTp (Altschul *et al.*, 1990), to assign annotations. Searching against the final curated Acari database, classified the *R. appendiculatus* proteins into four main classes; putative secretory proteins (with indications of cell secretion), putative housekeeping proteins (important in basic cell functional processes), unknown function proteins (function unknown), and no hit proteins (proteins that obtained no significant match in the database).

2.3.10 Expression analysis of transcripts

Transcripts per million (TPM) values were determined for the transcripts using the Bowtie 2 v2.2.3 (Langmead and Salzberg, 2012) and RNA-Seq by Expectation-Maximization (RSEM) v1.2.15 software packages (Li and Dewey, 2011). The TPM values were calculated for the entire transcriptome as well as for female and male ticks separately. Differentially expressed genes were determined by the Bioconductor/

Empirical analysis of digital gene expression data in R (edgeR) software package (Robinson *et al.*, 2010), at a fold change of > 2 and false discovery rate (FDR) p-value of < 0.001. Chi-square test with Bonferroni correction was used for significance testing.

2.3.11 Comparison to publically available sequences of *R. appendiculatus*

The transcriptome was compared to the RaGI gene index of 7970 *R. appendiculatus* ESTs (Nene *et al.*, 2004) to evaluate whether the assembled transcriptome improved on the publically available *R. appendiculatus* sequence dataset. Completeness of the RaGI dataset was evaluated by CEGMA analysis and compared to the *R. appendiculatus* transcriptome. Mutual coverage of the transcriptomes was estimated by BLASTn alignment (E-value ≤ E-05) against one another. Additionally, the protein classes in each transcriptome were compared. A further comparison was performed of the proteins assembled in this study against previously identified *R. appendiculatus* quantitative reverse transcriptase (RT) PCR reference genes (Nijhof *et al.*, 2009) and functionally characterised or validated *R. appendiculatus* proteins (Wang and Nuttall, 1995; Paesen *et al.*, 1999; Bishop *et al.*, 2002; Trimnell *et al.*, 2002; Mulenga *et al.*, 2003c; Mulenga *et al.*, 2003d; Paesen *et al.*, 2007; Paesen *et al.*, 2009; Imamura *et al.*, 2013; Preston *et al.*, 2013). The sequences of the proteins were obtained from NCBI and used in a BLASTp database to retrieve homologous sequences in the assembled *R. appendiculatus* protein set. For further characterisation of the serine proteinase inhibitors (serpins), a previously generated phylogeny was reproduced (Tirloni *et al*., 2014b). MUSCLE v3.8.31 was used for multiple sequence alignment (Edgar, 2004) and IQ-TREE for maximum likelihood phylogeny inference with 1000 bootstrap replicates (Minh *et al.*, 2013; Chernomor *et al.*, 2016). The phylogenetic tree was visualised in FigTree v1.4.3 (tree.bio.ed.ac.uk/software/figtree/).

2.3.12 Availability of supporting data

Raw sequence reads were deposited in the NCBI Short Read Archive (SRA, SRR2568016 - 9) under Bioproject accession number PRJNA297811. The transcripts have been deposited in the Transcriptome Shotgun Assembly project at DDBJ/EMBL/GenBank under accession GEDV00000000. The version described in this chapter is the first version, GEDV01000000. The predicted protein sequences were deposited in NCBI under the accession numbers JAP75796.1 - JAP88556.1.

2.4 Results

2.4.1 *Rhipicephalus appendiculatus de novo* transcriptome assembly, validation and annotation

In total, approximately 430 million paired end reads, ranging in size from 100 - 250 bp, were generated for the *R. appendiculatus* salivary glands (Appendix B: Table S1). Rigorous adapter trimming and quality filtering discarded between 12 - 19% of the reads, resulting in about 380 million read 1 and 340 million read 2 sequences that were used for transcriptome assembly. In total, 87 688 transcripts were assembled using the Trinity software package, which were reduced to 21 410 high confidence transcripts (Table 2.1) after filtering based on transcript abundance (FPKM value $\geq$ 1). The reference-based metrics indicated that the transcript sequences were highly accurate (99%), most transcripts were near full-length (83% completeness), many transcripts were intact (58% contiguity) and few showed evidence of chimerism (6%). The CEGMA analysis (Parra *et al.*, 2007) showed that 242 (98%) of the core Eukaryotic genes (CEGs) were present in the transcriptome and 236 (95%) of the CEGs were complete. The read mapping-based assessment denoted that 82% of the reads mapped back to the transcripts, indicating that most of the sequence reads were used in the assembly process. Overall, the evaluation metrics indicated that a highly representative, high confidence transcriptome of *R. appendiculatus* was assembled.

Table 2.1 Summary of *R. appendiculatus* transcriptome assembly statistics.

Assembly statistics	***R. appendiculatus* transcriptome**
Number of transcripts	21 410
Transcripts larger than 500 bp	18 892
Transcripts larger than 1 Kb	13 702
Transcripts larger than 10 Kb	115
Shortest transcript (bp)	301
Longest transcript (bp)	16 259
Mean transcript length (bp)	2060.3
Median transcript length (bp)	1443
Transcript N50 (bp)	3134
Total number of bases in assembly (Mb)	44.1
GC content (%)	49.0

By BLASTx searching the 21 410 transcripts against six protein databases, 15 645 transcripts (73%) were functionally annotated based on sequence similarity to proteins in at least one of the search databases (Table 2.2; annotations for all transcripts can be found in Appendix B: Table S2). KOG analysis assigned categories to 8282 transcripts, of which the 'General function prediction only' category was the largest, followed by 'Signal transduction mechanisms' and 'Posttranslational modification, protein turnover, chaperones' (Appendix B: Figure S1). Additionally, 63 757 GO terms were assigned to 10 111 transcripts and these were classified into 31 276 biological processes, 13 731 cellular components and 18 750 molecular functions (Appendix B: Figure S2). On the second level GO classification, the transcripts assigned to biological processes were predominantly characterised as 'Cellular process', 'Metabolic process' and 'Biological regulation'. The 'Cell' and 'Cell part' sub-classes were highly represented in the cellular component and 'Binding' and 'Catalytic activity', in the molecular function categories. KEGG pathways analysis assigned *I. scapularis* KO identifiers to 4647 transcripts (Appendix B: Figure S3). The most represented pathways were 'Ribosome', 'RNA transport', 'Protein processing in endoplasmic reticulum' and 'Spliceosome'.

Table 2.2 Summary of the functional annotation of the transcriptome of *R. appendiculatus.* Transcripts were BLASTx searched against locally configured databases with a cut-off E-value < E-05. Details of the datasets can be obtained in the Methods section.

Annotation databases	Number of transcripts	Percentage of transcripts
Transcriptome	21 410	100
BLASTx against NR	11 812	55.2
BLASTx against UniProtKB/TrEMBL	13 659	63.8
BLASTx against *I. scapularis* predicted peptides	11 123	52.0
BLASTx against *Rhipicephalus* protein sequences	12 485	58.3
BLASTx against EuKaryotic Orthologous Groups	8282	38.7
BLASTx against Acari in-house curated protein database	15 548	72.6
Functionally annotated in at least one database	15 645	73.1
Functionally annotated in all databases	7568	35.3
Assigned with Gene Ontology (GO) terms[a]	10 111	47.2
Assigned with Enzyme Commission (EC) numbers[a]	2882	13.5
Assigned with KEGG orthology (KO) identifiers[b]	4647	21.7

[a] GO terms and EC numbers were assigned by the BLAST2GO software package.

[b] Assigned from the *I. scapularis* genome using the KEGG (Kyoto Encyclopedia of Genes and Genomes) Automatic Annotation Server (KAAS).

2.4.2 Open reading frame prediction, annotation and comparison with *I. scapularis* and *R. pulchellus*

A total of 14 433 ORFs were predicted, which together represented 13 996 (65%) of the *R. appendiculatus* transcripts (Appendix B: Table S3 shows annotations of all predicted *R. appendiculatus* proteins). No ORFs were predicted for the remaining 7414 transcripts. The transcripts with no predicted ORFs were, on average, smaller (size range of 301 - 9726 bp and average size of 1166 bp) than the transcripts for which ORFs were predicted (302 - 16 259 bp, average 2534 bp). Of the transcripts with no ORFs, 97% were predicted to be putative non-protein-coding transcripts by at least two of the coding potential prediction software packages. In 2% of cases, more than one ORF was predicted per transcript, representing either miss-assembly or polycistronic transcripts. A final set of 12 761 non-redundant *R. appendiculatus* proteins was translated from the

ORFs. Eighty seven percent (11 034) of the non-redundant *R. appendiculatus* proteins were likely full-length (e.g. contained predicted start and stop codons) and most had significant BLASTp matches to protein search databases (79% and 89% against NR and UniProtKB/TrEMBL, respectively). Signal peptides were predicted for 3548 of the proteins and a total of 2593 proteins contained a transmembrane helix. The predicted proteins were searched against the Pfam database and 13 246 (3546 unique) Pfam domains were identified, categorising 7630 proteins. The most frequently observed domains were the 'Kunitz/bovine pancreatic trypsin inhibitor domain', 'Immunoglobulin I-set domain', and 'RNA recognition motif ' (Appendix B: Figure S4). The in-house curated Acari BLAST database classified the *R. appendiculatus* proteins into 2134 secretory, 8237 housekeeping, 1697 unknown function and 693 no hit proteins. Most of the putative secretory proteins (71%) had a signal peptide signature and of these, 97% started with a Methionine codon. Of the secretory proteins for which no signal peptides were predicted, only 66% initiated with Methionine codons, indicating potential truncations in some of these proteins.

The *R. appendiculatus* predicted proteins were compared to two publically available tick datasets. The *R. appendiculatus* (12 761) protein set had a similar number of proteins to *R. pulchellus* (11 227) that also represented salivary gland derived transcripts, and fewer proteins than *I. scapularis* (20 486) that represented predicted proteins from a whole tick genome. The *R. appendiculatus* protein set (70 - 4966 predicted amino acid [aa] residues, average size of 400 aa) was similar in length to the proteins of the other species (*R. pulchellus*: 66 - 6645 aa, average 472 aa; *I. scapularis:* 32 - 4588 aa, average 224 aa). To compare the composition of the three protein datasets, Pfam domains were predicted for each dataset and the 500 most frequently occurring domains, in each dataset, were compared to each other (Figure 2.1). Most of the domains (279) were shared between all three species and *I. scapularis* contained slightly more unique tissue- or species-specific Pfam domains (115 compared to 94 and 108, for *R. appendiculatus* and *R. pulchellus* respectively). These comparisons indicated that the predicted *R. appendiculatus* proteins were similar to proteins from other tick species, and even more so to proteins expressed in tick salivary glands.

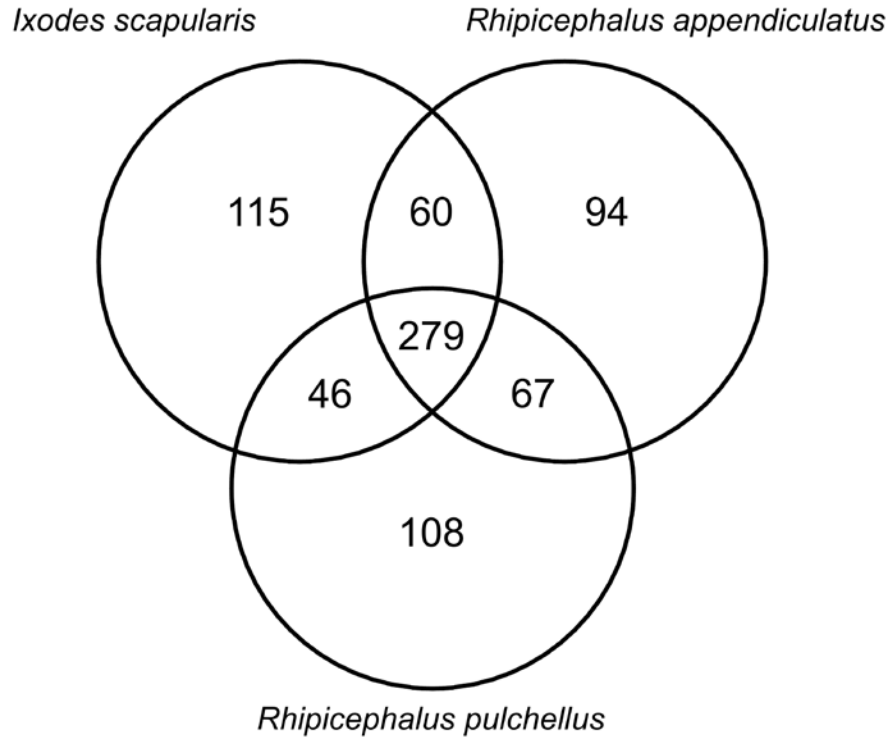


Figure 2.1 Pfam domain comparison of *R. appendiculatus, R. pulchellus and I. scapularis*. The 500 most represented Pfam domains in each species were used for the comparison analysis. Pfam searches were performed against the Pfam database and the Venn diagram drawn with Venny 2.0 (bioinfogp.cnb.csic.es/tools/venny/). Datasets used: 12 761 predicted non-redundant *R. appendiculatus* proteins (assembled in this study), 20 486 *I. scapularis* predicted peptides (IscaW1.4) and 11 227 *R. pulchellus* predicted proteins (NCBI Bioproject PRJNA170743).

2.4.3 Expression profiling in the *R. appendiculatus* transcriptome

The *R. appendiculatus* transcriptome had a large dynamic range of expression, from the lowest expressed transcript (0.7 transcripts per million [TPM]) to the most abundant (43 988 TPM) (Appendix B: Table S2). Less than 4% (775) of the transcripts accounted for 85% of the total mapped reads and the twenty most abundant transcripts accounted for 37% of the expression (Appendix B: Table S4). Nine of these twenty transcripts were annotated as belonging to the Glycine rich superfamily and represented 19% of the total expression in the transcriptome. Six additional transcripts were annotated as unknown function or retrieved no significant BLAST result and accounted for 9% of the total expression. The second most abundantly expressed transcript in the *R. appendiculatus*

transcriptome (c53945_g1_i1), representing 4% of the total transcriptome expression, was annotated as 16S ribosomal RNA. All transcripts without predicted ORFs, including rRNA molecules and putative non-coding RNAs, accounted for 12% of the expression in the transcriptome. The secretory protein class represented a disproportionally large part of the total transcriptome expression (63%), given its relatively small number of proteins (2134, 17%; Figure 2.2a). Conversely, the housekeeping class of 8237 proteins represented the majority of proteins (65%) but only 23% of the total transcriptome expression. The unknown function and no hit protein classes, accounting for the smallest fraction of the total transcriptome expression, had higher average transcript expression levels (average TPM values of 53 - 54) as compared to the housekeeping class (TPM value of 24). This indicated that the unknown function and no hit protein classes contained uncharacterised proteins that were expressed at potentially biologically meaningful levels in the transcriptome.

Families within the secretory protein class were expressed at varying average TPM levels in the transcriptome (ranging from 1 to 3072; Table 2.3). No correlation was observed between the number of proteins in a family and the percentage the family contributed to the total transcript expression in the secretory protein class (Figure 2.2b; Table 2.3). The Glycine rich superfamily, representing only 6% (119 proteins) of the total number of secretory proteins, contributed 66% of the total transcript expression in the secretory protein class. The second largest transcript expression contributor to the secretory protein class, at 12%, was that of the largest secretory family, Lipocalin (containing 24% of the secretory proteins). Moreover, members of these two families were the most abundantly expressed transcripts in the transcriptome (Figure 2.2c). All other secretory protein families (1499 proteins) resulted in the remaining 22% of the expression in the secretory protein class.

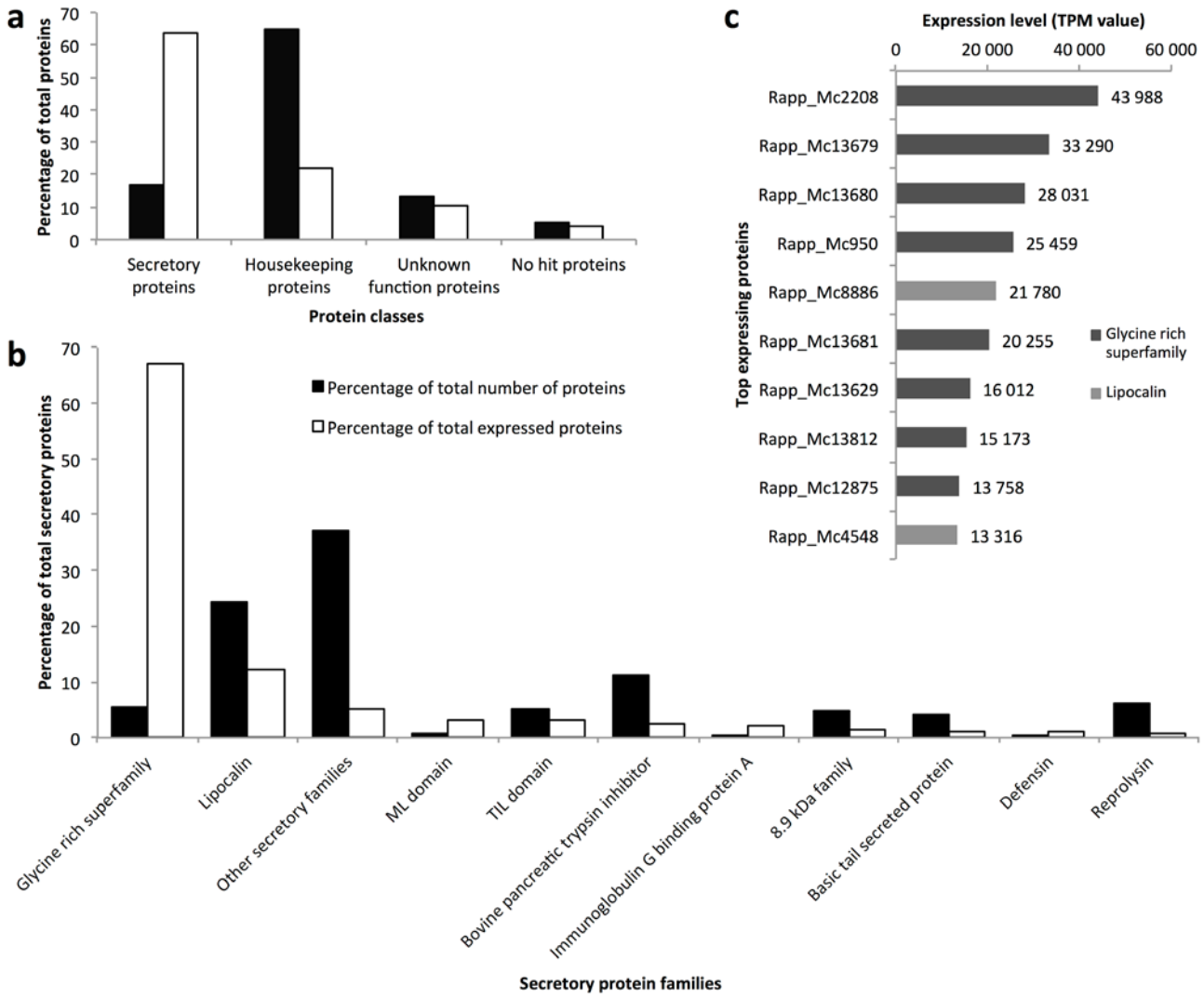


Figure 2.2 Expression analysis in the transcriptome of *R. appendiculatus.* **a** The percentage of transcripts in each protein class and the expression contribution of those classes to the total expression in the *R. appendiculatus* transcriptome. **b** The percentage of transcripts within each protein family of the secretory protein class and the contribution of those families to the total expression in the secretory protein class. Black indicates protein numbers and white, expression contribution. Expression was measured by TPM (transcripts per million). **c** TPM values of the top ten expressing proteins in the transcriptome of *R. appendiculatus.* Dark grey represents members of the Glycine rich protein family and light grey, members of the Lipocalin family.

Table 2.3 Characterisation of the tick secretory protein family expression in the *R. appendiculatus* transcriptome.

Secretory protein family	Number of family members	Proportion of the total number of secretory proteins represented by this family (%)	Protein family average TPM value	Proportion of the secretory protein class expression represented by this family (%)	ORF ID of the top expressing member in the family	TPM value of the top expressing member in the family	Proportion of the protein family represented by the top expressing member (%)
Lipocalin	516	24.18	133.1	12.47	Rapp_Mc8886	21 779.9	31.7
Bovine pancreatic trypsin inhibitor	236	11.06	58.4	2.50	Rapp_Mc8896	2447.7	17.8
Reprolysin	133	6.23	34.9	0.84	Rapp_Mc5881	628.4	13.5
Glycine rich superfamily	119	5.58	3072.0	66.34	Rapp_Mc2208	43 988.3	11.6
TIL domain	108	5.06	164.0	3.21	Rapp_Mc1646	2899.0	16.4
8.9 kDa family	102	4.78	84.0	1.55	Rapp_Mc13118	1185.6	13.8
Basic tail secreted protein	90	4.22	75.8	1.24	Rapp_Mc4488	879.8	12.9
Evasin	68	3.19	55.5	0.69	Rapp_Mc9039	619.8	16.4
Ixodegrin B	57	2.67	30.1	0.31	Rapp_Mc823	450.2	26.2
Gluzincin	52	2.44	7.6	0.07	Rapp_Mc4972	115.1	29.0
Mucin	52	2.44	44.9	0.42	Rapp_Mc417	876.5	37.6
Digestive system (including Serine proteases)	50	2.34	25.0	0.23	Rapp_Mc1191	137.7	11.0
Cystatin	47	2.20	54.4	0.46	Rapp_Mc13730	776.5	30.4
Folding, sorting and degradation (including Cathepsins)	40	1.87	96.8	0.70	Rapp_Mc945	1498.5	38.7
28 kDa Metastriate family	31	1.45	50.7	0.29	Rapp_Mc2646	557.1	35.5
Chitin-binding proteins	30	1.41	23.0	0.13	Rapp_Mc9698	223.6	32.4
Serpin	27	1.27	9.0	0.04	Rapp_Mc5185	74.7	30.8
DA-P36 family	25	1.17	23.9	0.11	Rapp_Mc8808	340.8	57.1

Transport and catabolism	25	1.17	36.1	0.16	Rapp_Mc2177	535.9	59.3
One of each family	23	1.08	9.4	0.04	Rapp_Mc3057	42.5	19.6
Lipid metabolism	22	1.03	4.3	0.02	Rapp_Mc1456	12.1	12.9
Carboxypeptidase inhibitor	22	1.03	42.9	0.17	Rapp_Mc10222	388.5	41.2
5’-Nucleotidase	16	0.75	14.3	0.04	Rapp_Mc6697	47.6	20.8
Microplusin	16	0.75	65.4	0.19	Rapp_Mc1964	434.7	41.6
ML domain	16	0.75	1111.8	3.23	Rapp_Mc774	11 353.3	63.8
Antigen 5 family	13	0.61	188.9	0.45	Rapp_Mc1903	1015.0	41.3
Signaling molecules and interaction	13	0.61	1.6	0.00	Rapp_Mc8916	2.8	13.6
Translation	13	0.61	11.0	0.03	Rapp_Mc13622	43.1	30.2
24 kDa family	12	0.56	24.2	0.05	Rapp_Mc9762	91.6	31.5
Defensin	12	0.56	464.4	1.01	Rapp_Mc8698	1900.0	34.1
8 kDa Amblyomma family	11	0.52	28.8	0.06	Rapp_Mc13004	154.3	48.7
Glycan biosynthesis and metabolism	11	0.52	14.0	0.03	Rapp_Mc6227	80.5	52.3
Sphingomyelinase	9	0.42	10.7	0.02	Rapp_Mc837	22.9	23.8
Signal transduction	8	0.37	7.8	0.01	Rapp_Mc1131	15.0	24.0
Transcription	8	0.37	4.4	0.01	Rapp_Mc1617	9.5	26.8
Carbohydrate metabolism	7	0.33	4.3	0.01	Rapp_Mc5896	7.6	25.2
Fibrinogen-related domain	7	0.33	52.6	0.07	Rapp_Mc9028	248.6	67.6
Secretory - unknown function	7	0.33	12.8	0.01	Rapp_Mc9124	51.4	65.0
Immunoglobulin G binding protein A	6	0.28	2051.8	2.23	Rapp_Mc1190	4721.8	38.4
Metabolism of other amino acids	6	0.28	29.9	0.03	Rapp_Mc5888	67.4	37.6
Phospholipase A2	6	0.28	22.5	0.02	Rapp_Mc8892	44.3	32.8
Replication and repair	6	0.28	5.3	0.01	Rapp_Mc2861	9.2	29.2
7DB family	5	0.23	26.5	0.02	Rapp_Mc5571	66.1	49.8
Metalloprotease	5	0.23	12.1	0.01	Rapp_Mc12946	38.9	64.4
SALP15	4	0.19	10.3	0.01	Rapp_Mc1541	28.7	70.0
Astacin	3	0.14	1.8	0.00	Rapp_Mc7012	3.4	63.8

Cell growth and death	3	0.14	6.8	0.00	Rapp_Mc3897	17.4	84.9
Dermacentor 9 kDa expansion	3	0.14	14.7	0.01	Rapp_Mc1065	26.0	59.0
Histidine rich	3	0.14	2.9	0.00	Rapp_Mc450	4.0	45.8
14 kDa family	3	0.14	25.1	0.01	Rapp_Mc8740	54.4	72.3
Kazal domain	3	0.14	21.1	0.01	Rapp_Mc421	59.7	94.3
Kazal/ vWf domain	3	0.14	18.6	0.01	Rapp_Mc2515	35.5	63.5
TELEM	3	0.14	2.0	0.00	Rapp_Mc5946	2.3	38.9
Thyropin	3	0.14	67.6	0.04	Rapp_Mc1844	99.5	49.1
Cysteine rich	2	0.09	1.0	0.00	Rapp_Mc1691	1.1	53.7
Energy metabolism	2	0.09	15.8	0.01	Rapp_Mc6151	25.1	79.3
Hirudin	2	0.09	310.2	0.11	Rapp_Mc11642	418.3	67.4
Bovine pancreatic trypsin inhibitor - Lipocalin	1	0.05	3.6	0.00	Rapp_Mc3211	3.6	100.0
Chitin deacetylase activity	1	0.05	1.2	0.00	Rapp_Mc2536	1.2	100.0
Cell motility	1	0.05	3.4	0.00	Rapp_Mc4124	3.4	100.0
Cysteine rich hydrophobic domain 2	1	0.05	13.4	0.00	Rapp_Mc7198	13.4	100.0
Fatty acid-binding protein	1	0.05	40.4	0.01	Rapp_Mc2582	40.4	100.0
Histamine release factor	1	0.05	1211.6	0.22	Rapp_Mc12631	1211.6	100.0
Immune system	1	0.05	24.2	0.00	Rapp_Mc8912	24.2	100.0
26 kDa family	1	0.05	7.2	0.00	Rapp_Mc5668	7.2	100.0
Kazal/ SPARC domain	1	0.05	65.0	0.01	Rapp_Mc1895	65.0	100.0

2.4.4 Expression differences between female and male *R. appendiculatus* ticks

Mapping of 159 069 158 paired end sequence reads for female and 164 148 238 for male salivary glands enabled the calculation of separate transcript expression values within the different sexes. In the female and male transcriptomes, the 100 most abundant transcripts accounted for the majority of the expression in each transcriptome (55% of the total expression observed in the female transcriptome and 75% in the male transcriptome; Appendix B: Table S2). Many of the families in the secretory protein class had significantly skewed expression between males and female transcriptomes after a Bonferroni corrected p-value < 0.0008 (Figure 2.3). The Glycine rich superfamily, the most abundant family, was expressed twice as much in the male transcriptome (476 036 TPM) than in the female transcriptome (228 994 TPM). These high expression values of the Glycine rich superfamily accounted for the majority of the secretory protein class expression in the transcriptome of each gender (72% and 55% for the male and females, respectively). Some families were almost exclusively expressed in one of the sexes. For example, the 28 kDa Metastriate, Hirudin, DA-P36 and One of each families showed female-predominant transcriptome expression, while the ML domain, Immunoglobulin G binding protein A and Cystatin families showed male-predominant transcriptome expression (Figure 2.3).

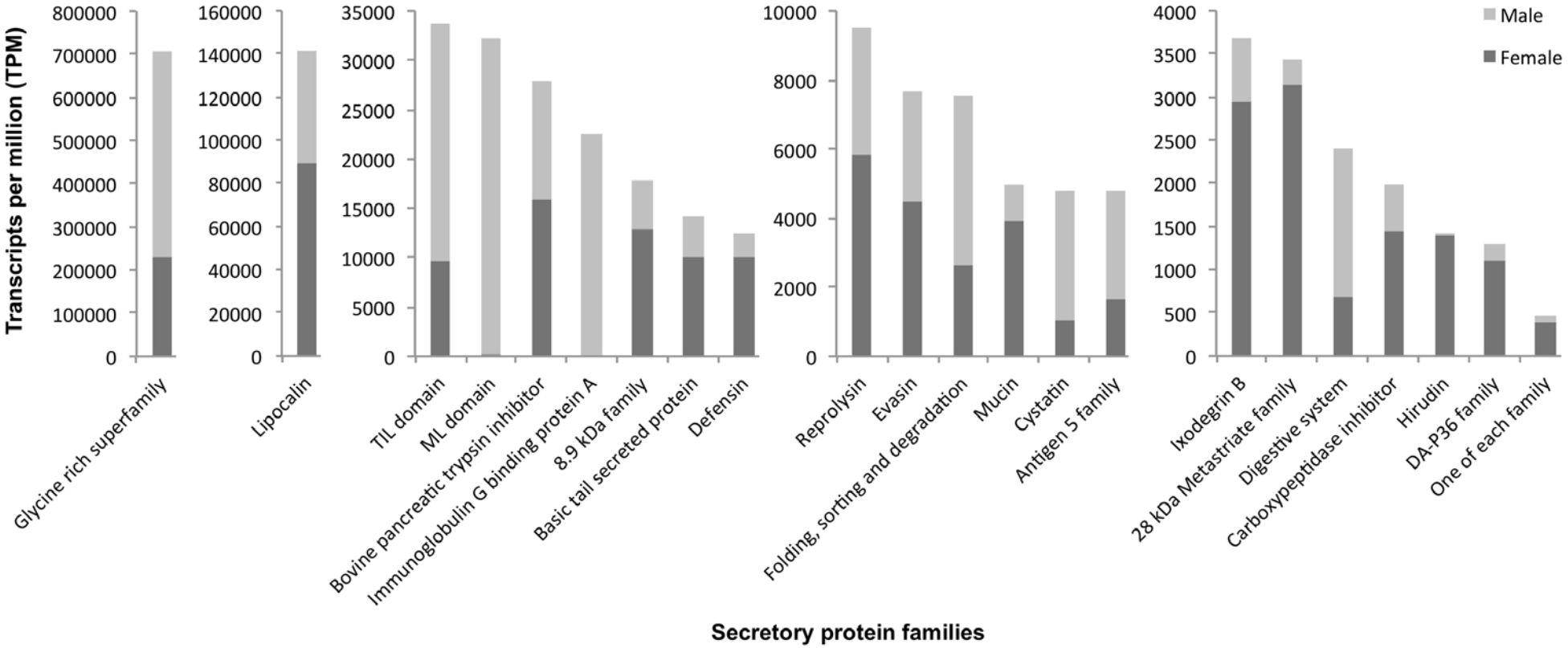


Figure 2.3 Gender-skewed expression of secretory protein families in the *R. appendiculatus* transcriptome. Secretory protein families with significant (Chi-square test with Bonferroni corrected p-value < 0.0008) gender-biased transcriptome expression are indicated. Expression was measured as transcripts per million (TPM). Female expression is indicated by dark grey and male expression by light grey.

A total of 1758 and 2346 transcripts were differentially up-regulated (at least a two-fold increase) in the female and male transcriptomes, respectively (Appendix B: Table S5). Of these, 570 (32%) and 553 (24%) were annotated as putative secretory proteins in each of the female and male transcriptomes. Significantly more transcripts of the Ixodegrin B (36 female vs. 7 male transcripts) and One of each (17 vs. 1) families were up-regulated in the female transcriptome while more of the Digestive system (including Serine proteases; 3 vs. 25) and Gluzincin (2 vs. 33) families were up-regulated in the male transcriptome after a Bonferroni corrected p-value < 0.0010. Additional large gender differences, albeit not significant, were observed in the DA-P36 (15 vs. 2) and 28 kDa Metastriate (18 vs. 4) families. A large number of the differentially expressed transcripts (727 and 920 for females and males, respectively) were transcripts without predicted ORFs. Ninety one percent of these were predicted to be putative non-coding RNA molecules and likely involved in tick feeding regulatory functions between female and males. Many of the differentially expressed transcripts encoded for proteins of which the function has not yet been elucidated (10%) or transcripts with no significant BLAST matches to protein databases (7%), indicating the large number of proteins involved in tick feeding that are still uncharacterised.

2.4.5 Comparison of the assembled transcriptome to publically available sequences of *R. appendiculatus*

The *R. appendiculatus* transcriptome assembled in this study contained more transcripts (21 410 vs. 7970) of longer length (average length of 2060 bp vs. 853 bp) than the *R. appendiculatus* EST gene index (RaGI) dataset (Nene *et al.*, 2004). Moreover, the level of completeness of the *R. appendiculatus* transcriptome (98% of the core Eukaryotic genes were present and 95% were complete) exceeded that of the RaGI gene set (only 34% were present and 22% complete). The BLASTn alignment showed that 81% of the transcripts in the RaGI set (6437) were represented by the *R. appendiculatus* transcriptome, whereas only 31% of the transcripts in the *R. appendiculatus* transcriptome (6693) were represented by the RaGI gene set. This indicated that many new transcripts were present in the *R. appendiculatus* transcriptome. In addition, 4.5 times more secretory proteins (2134 *R. appendiculatus* proteins vs. 459 RaGI sequences) and two times more housekeeping proteins (8237 vs. 3796) were assembled in the current transcriptome (Figure 2.4a). Comparison of the protein families within the

secretory class revealed that for most of the families more proteins were found in the *R. appendiculatus* protein set compared to the RaGI set (Figure 2.4b). Particularly large differences in the number of proteins were observed for many of the families: e.g. the Lipocalin (516 compared to 24 for *R. appendiculatus* and RaGI, respectively), Bovine pancreatic trypsin inhibitor (236 vs. 26) and Evasin (68 vs. 4) protein families. These comparisons revealed that the transcriptome assembled in this study is a substantial improvement on the publically available sequences of *R. appendiculatus.*

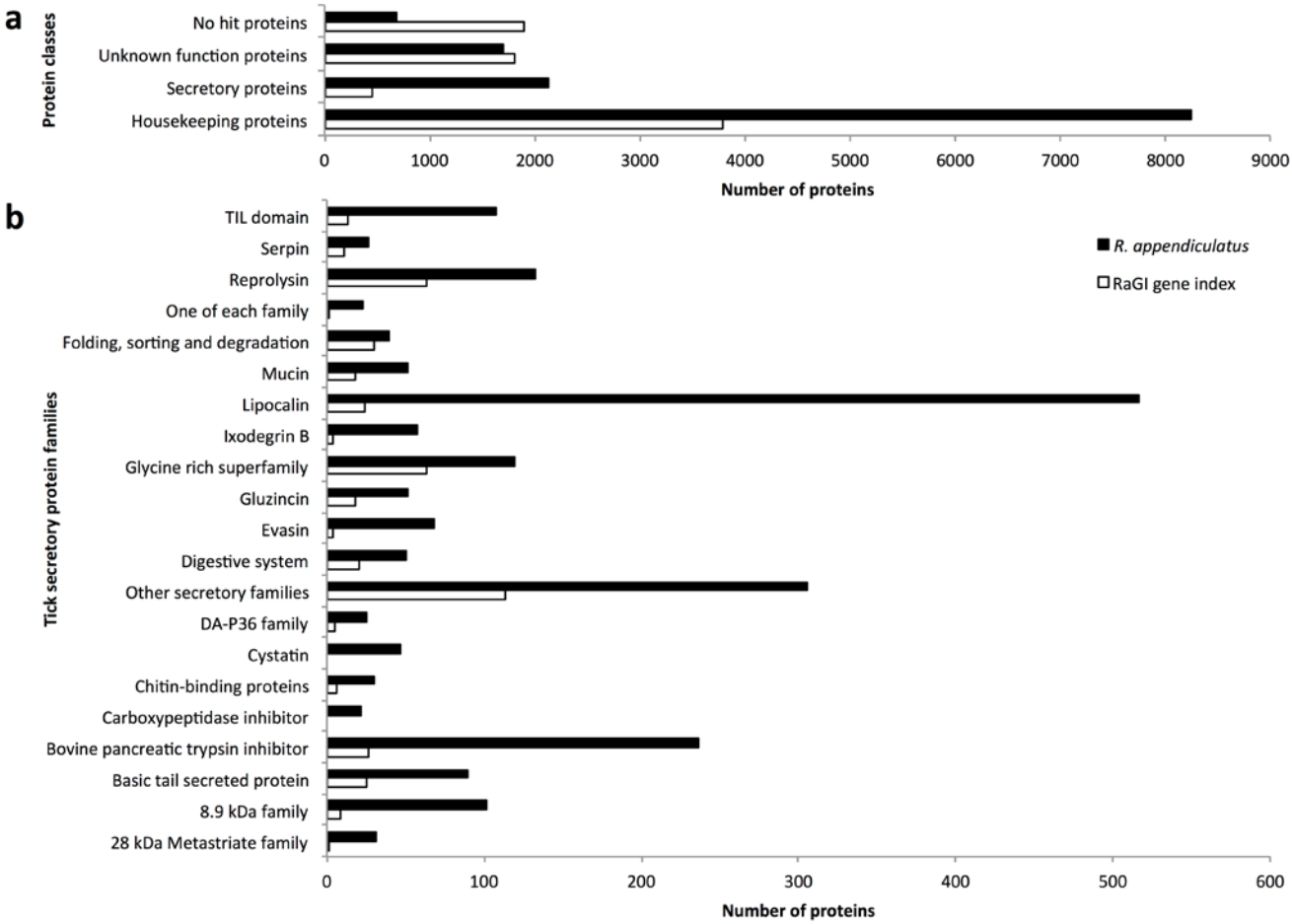


Figure 2.4 Comparison of tick protein family sizes in the *R. appendiculatus* and RaGI gene sets. a Number of members in each tick protein class of the *R. appendiculatus* transcriptome in comparison to the previously generated *R. appendiculatus* gene index (RaGI) (Nene *et al.*, 2004). **b** The number of members in the largest tick secretory protein families of the *R. appendiculatus* transcriptome and RaGI. The *R. appendiculatus* transcriptome is indicated in black and the RaGI, in white.

Predominantly full-length versions of previously characterised *R. appendiculatus* proteins were assembled in the *R. appendiculatus* transcriptome. All nine quantitative RT-PCR reference genes (Nijhof *et al.*, 2009) were identified in the assembled transcriptome (Table 2.4). Many of these genes were previously only available as EST fragments, but here full-length versions were assembled for all, except Beta actin (85% complete). The genes showed varying degrees of gender-skewed expression, which would affect their suitability as expression reference genes in certain experimental designs. As expected, midgut specific proteins: gut cystatin, Ra-cyst-1 (Imamura *et al.*, 2013); midgut serine proteinases, RAMSPs (Mulenga *et al.*, 2003c); and Bm86-like protein, Ra86-1 (Nijhof *et al.*, 2009); were either not observed in the *R. appendiculatus* salivary gland transcriptome, or present at a very low expression levels. The putative cement protein, *Rhipicephalus* immuno-dominant molecule 36 (RIM36), was the only previously characterised protein not assembled in a single transcript. Two non-overlapping RIM36 transcripts were assembled, c15622_g1_i1 and c33374_g1_i1, that each coded for a truncated peptide with 100% protein identity to the RIM36 protein (AAK98794.1) sequenced by Bishop *et al.* (2002). The sequence reads did not support connection of the transcripts and they were kept as RIM36 fragments in the transcriptome. Protein identities between 48% and 88% were observed for four proteins; Serine proteinase inhibitors RAS-1 and RAS-4 (Mulenga *et al.*, 2003d), Male-specific histamine-binding protein, HBPM (Paesen *et al.*, 1999) and Japanin (Preston *et al.*, 2013). These four proteins represented 16% of the previously characterised proteins, which was in range with the percent identity observed between the *R. appendiculatus* transcriptome and the RaGI gene set (15% of best BLAST hits had identities $\leq$ 90%). No proteins with significant homology were identified for the Japanin-like-RA1, JL-RA1 (Preston *et al.*, 2013) or the Tryptase inhibitor, TdP1 (Paesen *et al.*, 2007) proteins. In order to validate the low identity observed or absence from the dataset of some of the proteins, a subset of 22.4 million reads (about 5% of the sequence reads) was mapped to the nucleotide sequences of the previously characterised *R. appendiculatus* proteins and the sequences assembled in this transcriptome. Not a single sequence read mapped to the HBPM, JL-RA1 and TdP1 sequences downloaded from NCBI. In contrast, 288 649 sequence reads mapped to the transcript of Rapp_Mc8886, the homolog of HBPM assembled in the *R. appendiculatus* transcriptome. Similarly, significantly lower mapping was observed for the published RAS-1, RAS-4 and Japanin sequences compared to homologous sequences assembled in the *R. appendiculatus* transcriptome.

Table 2.4 Previously characterised *R. appendiculatus* proteins and their annotation in the assembled *R. appendiculatus* transcriptome.

Protein name	Protein description	Accession number	Reference	Protein ID	Identity (%)	Full-length	Combined TPM	Female TPM	Male TPM
PPIA	Cyclophilin	CD793819	(Nijhof *et al.*, 2009)	Rapp_Mc8751	100	Complete	278.8	444.3	144.8
ELF1A	Elongation factor 1-alpha - fragment	CD797149	(Nijhof *et al.*, 2009)	Rapp_Mc1620	99	Complete	2382.7	3145.6	1766.0
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	CD791831	(Nijhof *et al.*, 2009)	Rapp_Mc3915	98	Complete	401.0	675.9	178.2
GST	Glutathione S-transferase	CD789942	(Nijhof *et al.*, 2009)	Rapp_Mc2803	98	Complete	741.4	1626.6	27.9
H3F3A	H3 Histone	CD795637	(Nijhof *et al.*, 2009)	Rapp_Mc10419	100	Complete	49.2	68.1	33.9
RPL4	Ribosomal protein L4 - fragment	CD794864	(Nijhof *et al.*, 2009)	Rapp_Mc2580	100	Complete	614.7	756.7	503.5
TBP	TATA box binding protein - fragment	CD780134	(Nijhof *et al.*, 2009)	Rapp_Mc2537	96	Complete	1.9	2.6	1.4
BTUB	Beta tubulin - fragment	CD781348	(Nijhof *et al.*, 2009)	Rapp_Mc425	100	Complete	142.8	267.8	39.6
ACTB	Beta actin	AAP81256.1	(Nijhof *et al.*, 2009)	Rapp_Mc4908	100	Fragment[a]	733.7	1172.0	381.6
Ra-cyst-1	Gut cystatin	AGB35873.1	(Imamura *et al.*, 2013)	Rapp_Mc817	100	Complete	4.5	2.5	6.2
64P	Salivary gland-associated protein 64P	AAM09648.1	(Trimnell *et al.*, 2002)	Rapp_Mc13701	98	Complete	424.3	207.0	583.2
RIM36	Putative cement protein RIM36	AAK98794.1	(Bishop *et al.*, 2002)	Rapp_Mc2208	100	Fragment[b]	43 988	18 954	63 919
				Rapp_Mc13680	100	Fragment[b]	28 031	12 227	40 535
RAS-1	Serine proteinase inhibitor serpin-1	AAK61375.1	(Mulenga *et al.*, 2003d)	Rapp_Mc7014	88	Complete	6.5	6.8	6.2
RAS-2	Serine proteinase inhibitor serpin-2	AAK61376.1	(Mulenga *et al.*, 2003d)	Rapp_Mc6400	92	Complete	11.3	14.6	8.5

RAS-3	Serine proteinase inhibitor serpin-3	AAK61377.1	(Mulenga *et al.*, 2003d)	Rapp_Mc5185	95	Complete	74.7	136.7	24.2
RAS-4	Serine proteinase inhibitor serpin-4	AAK61378.1	(Mulenga *et al.*, 2003d)	Rapp_Mc4940	72	Complete	18.5	39.1	1.6
HBP1	Female-specific histamine-binding protein 1	O77420	(Paesen *et al.*, 1999)	Rapp_Mc4548	99	Complete	13 316	29 685	0.5
HBP2	Female-specific histamine-binding protein 2	O77421	(Paesen *et al.*, 1999)	Rapp_Mc3118	98	Complete	3527.8	7847.9	0.2
HBPM	Male-specific histamine-binding protein	O77422	(Paesen *et al.*, 1999)	Rapp_Mc8886	48	Complete	21 780	1.8	39 107
IGBP-MA	Immunoglobulin G binding protein A	AAB68801.1	(Wang and Nuttall, 1995)	Rapp_Mc1190	100	Complete	4721.8	0.6	8630.8
IGBP-MB	Immunoglobulin G binding protein B	AAB68802.1	(Wang and Nuttall, 1995)	Rapp_Mc2702	99	Complete	6350.9	0.4	11 556
IGBP-MC	Immunoglobulin G binding protein C	AAB68803.1	(Wang and Nuttall, 1995)	Rapp_Mc774	99	Complete	11 353	0.7	20 524
Japanin	Japanin precursor	AGF70149.1	(Preston *et al.*, 2013)	Rapp_Mc9023	86	Complete	177.5	395.5	0.0
JL-RA1[c]	Japanin-like-RA1 precursor	AGF70151.1	(Preston *et al.*, 2013)	np	np	np	np	np	np
JL-RA2	Japanin-like-RA2 precursor	AGF70152.1	(Preston *et al.*, 2013)	Rapp_Mc378	99	Complete	1186.9	2641.9	0.1
Ra-KLP	Kunitz/BPTI-like protein precursor	ACM86785.1	(Paesen *et al.*, 2009)	Rapp_Mc8716	100	Complete	980.3	2278.1	0.1
TdP1[c]	Tryptase inhibitor precursor	AAW32666.1	(Paesen *et al.*, 2007)	np	np	np	np	np	np

[a] Deduced protein sequence of Beta actin had an N-terminal truncation of 55 amino acids.

[b] RIM36 was assembled into two non-overlapping transcripts, c15622_g1_i1 and c33374_g1_i1, encoding a 137 aa peptide (Rapp_Mc13680) and a 140 aa peptide (Rapp_Mc2208), respectively.

[c] The protein homologs of JL-RA1 and TdP1 were not present (np) in the *R. appendiculatus* transcriptome.

The HBPM homolog, at 21 780 TPM, is the 6th highest expressed transcript in the *R. appendiculatus* transcriptome (Appendix B: Table S4) and almost exclusively expressed in the male salivary gland transcriptome (TPM of 39 107 vs. 1.8 for males and females, respectively). A signal peptide signature between position 15 and 16, no transmembrane helices, and the complete tick histamine binding protein domain, pfam02098 (E-value = E-33), were observed for the assembled protein (Appendix B: Table S3). Searching against the NCBI protein database (PDB) retrieved the Chain A, Histamine Binding Protein From Female Brown Ear *R. appendiculatus* (pdb|1QFT|A) sequence with significant similarity (E-value = E-35) and identity (36%). These are all typical characteristics of histamine binding proteins as characterised by Paesen *et al.* (1999). The Japanin homolog assembled in the *R. appendiculatus* transcriptome, Rapp_Mc9023, contained a signal peptide signature, lacked a transmembrane domain and had no protein domain similarities to known protein databases (Appendix B: Table S3). This was similar to what was shown for Japanin (Preston *et al.*, 2013). The conserved cysteine residues and the tick Lipocalin motive, previously characterised in Japanin, were also observed in Rapp_Mc9023. The four assembled serpin proteins were observed at varying levels of protein identities to the previously characterised proteins (72 - 95%, Table 2.4). The complete SERPIN cd00172 domain and the conserved reactive centre loop (RCL), typical to serpins, were identified in all four sequences (Appendix B: Table S3). When the serpin phylogeny from Tirloni *et al.* (2014b) was reproduced, Rapp_Mc7014, Rapp_Mc6400 and Rapp_Mc5185 grouped closest to their respective homologs (RAS-1 to RAS-3, Appendix B: Figure S5). Further, Rapp_Mc4940 grouped in the same cluster as its homolog, RAS-4 (72% protein identity), but more closely to *R. pulchellus* serpin RpS-2 (JAA54310.1, 95% identity) and *R. microplus* RmS-17 (AHC98668.1, 92% identity).

2.5 Discussion

The aim of the study was to *de novo* assemble a salivary gland transcriptome of *R. appendiculatus* that represented the feeding stages of both female and male ticks. During *de novo* assembly, an expression catalogue is constructed from short sequence reads without any prior reference, making it essential to assess whether the assembled transcriptome reflects the actual transcripts in the biological sample. In this study, two

transcriptome quality assessments, read mapping-based and reference sequence-based assessments (Martin and Wang, 2011; O'Neil and Emrich, 2013), indicated that the assembled transcriptome was representative of the sequence reads and showed a high level of accuracy, completeness and contiguity. The estimated chimerism statistic was somewhat suboptimal and could be attributed to using an EST reference set for the transcriptome evaluation. In EST datasets, sequences are usually short, incomplete and not a true representation of the expressed transcriptome. When examining the number of transcripts for which multiple ORFs were predicted, the level of chimerism decreases to a much more appropriate level of only two percent. Moreover, the assembled transcriptome was a vast improvement on the available RaGI gene index of *R. appendiculatus* (Nene *et al.*, 2004) due to the improvement in transcript length, completeness and representation of previously unidentified *R. appendiculatus* transcripts. The similarity of the *R. appendiculatus* proteins to two known tick protein sequence datasets: *I. scapularis* (Pagel Van Zee *et al.*, 2007), representing a complete set of expected proteins in ticks; and *R. pulchellus* (Tan *et al.*, 2015a), representing proteins expected to be expressed in the salivary glands of feeding ticks; indicated that the protein prediction was comprehensive and resulted in a fully representative set of tick salivary gland proteins. Annotation of non-model organisms and especially arthropods is challenging due to the small number of completed genomes and limited publically available protein sequences for homology searches. Previous tick sialome studies that compared transcripts to a single search database returned few functional annotations, 29% in *Amblyomma americanum* (Gibson *et al.*, 2013) and 37% in *Haemaphysalis flava* (Xu *et al.*, 2015). In contrast, a higher overall annotation rate of 73% was observed in the transcriptome of *R. appendiculatus* assembled here. The higher rate of annotation was attributed to searching against more than one protein database and the removal of lowly expressed (FPKM value < 1) transcripts from the transcriptome (Mortazavi *et al.*, 2008). No ORFs were predicted for 35% of the transcripts. A small percentage of these transcripts had BLASTx annotations and biologically relevant expression levels, indicating that some putative proteins might not have been be predicted in these transcripts, albeit at a low percentage. These transcripts were mainly predicted to have low protein-coding potential and might be representative of long non-coding RNA (lncRNA). Long ncRNAs are RNA molecules longer than 200 bp that contain no ORFs for translation and have shown functions in transcriptional regulation, RNA processing and protein scaffolding (reviewed in Wilusz *et al.*, 2009). In

Drosophila, over a thousand lncRNA molecules have been identified and more lncRNAs were sex-specifically expressed compared to protein-coding genes in adult flies (Young *et al.*, 2012). Similarly, it was observed that 40% of the differentially expressed transcripts between male and female *R. appendiculatus* ticks had no predicted open reading frames. Their classification as true lncRNAs remains to be experimentally determined.

Seventeen percent of the predicted *R. appendiculatus* proteins were characterised as putative secretory proteins, which was in the same range, 13 - 37%, as in other tick sialotranscriptomes (Karim *et al.*, 2011; Garcia *et al.*, 2014; Karim and Ribeiro, 2015; Tan *et al.*, 2015a). The same authors also reported a wide range of expression percentages of secretory proteins in the sialotranscriptomes (between 17 - 49%), slightly less than what was observed for *R. appendiculatus* (63%). Additionally, the *R. appendiculatus* transcriptome had a large dynamic expression range, with few transcripts accounting for most of the expression in the salivary glands. Most of these highly expressed transcripts contained ORFs that coded for secretory proteins (i.e. 52% of the 50 highest expressing transcripts were classified as secretory protein families). Similarly, previous studies reported that secretory proteins were the most highly expressed transcripts in the *I. ricinus* salivary glands when compared to midgut tissues (Schwarz *et al.*, 2014b; Kotsyfakis *et al.*, 2015b). High expression levels of secretory protein transcripts were not surprising as salivary glands are actively producing and secreting proteins into the host that facilitate tick feeding by altering the host's haemostasis, inflammation and immune response.

Transcripts of the Glycine rich superfamily were expressed at particularly high levels and contributed 66% of the secretory protein class expression in *R. appendiculatus*. Similarly, 48% of the secretory class expression in *R. pulchellus* was of Glycine rich transcripts (Tan *et al.*, 2015a). This was opposed to low levels, of between 3 - 28%, observed in the secretory class of the sialotranscriptomes of *Amblyomma* ticks (Garcia *et al.*, 2014; Karim and Ribeiro, 2015). The mouthparts of the ticks might offer one plausible explanation for the high levels of Glycine rich transcripts in the *R. appendiculatus* and *R. pulchellus* sialotranscriptomes, compared to the levels observed in *Amblyomma* ticks. Glycine rich proteins with adhesive and tensile characteristics form part of the cement-cone or 'glue' that adheres ixodid ticks to their hosts to assist

uninterrupted feeding (Binnington and Kemp, 1980; Sonenshine, 1991). *Rhipicephalus appendiculatus* and *R. pulchellus* ticks are classified as Brevirostrata ticks, which have short mouthparts that barely penetrate the host's epidermis and therefore require wide and deep cement-cones to facilitate adhesion. *Amblyomma* (Longirostrata ticks), on the other hand, have longer mouthparts that penetrate the skin more deeply, requiring a smaller cement-cone to facilitate adhesion. Indeed, Maruyama *et al.* (2010) found that ticks with short mouthparts expressed elevated levels of Glycine rich transcripts when compared to ticks with long mouthparts. Interestingly, a larger abundance of Glycine rich transcripts was observed in the male compared to female *R. appendiculatus* transcriptomes, a finding also observed in *R. pulchellus* (Tan *et al.*, 2015a). One would expect, that female ticks require a larger cement-cone and consequently more Glycine rich proteins than male ticks, given their prolonged feeding time and substantial increase in body size (Sonenshine, 1991). Yet, the male *R. appendiculatus* and *R. pulchellus* (Tan *et al.*, 2015a) ticks expressed larger quantities of Glycine rich transcripts in their salivary glands than females, suggesting an additional function of Glycine rich proteins in male salivary glands. The mating and feeding behaviour of male ticks that attach, detach and re-attach to where females are feeding (Sonenshine, 1991) might require a constant supply of secretory proteins, such as Glycine rich proteins, in the salivary glands of male ticks. Furthermore, tick cement proteins have been identified as potential vaccine candidates due to the strong immune response they cause in their hosts (Bishop *et al.*, 2002; Trimnell *et al.*, 2005). It is therefore also possible that some male Glycine rich proteins may facilitate immune evasion by acting as decoy antigens, thereby enhancing female feeding (Wang *et al.*, 1998).

Similar to previous studies (Aljamali *et al.*, 2009; Xiang *et al.*, 2012; Tan *et al.*, 2015a), in this study various differences were observed in gene expression between the male and female salivary transcriptomes, suggestive of different feeding or host immune evading mechanisms employed by the different sexes. Three such differentially expressed genes, the Immunoglobulin G binding protein - Male A - C (IGBP-MA - C), were exclusively expressed in the male salivary transcriptome and have been shown to enable male *R. appendiculatus* ticks to assist co-feeding females by altering the feeding site (Wang *et al.*, 1998). Notably, transcripts of protease and protease inhibitors such as Serine proteases, Peptidases (Gluzincin), and Cystatins, were up-regulated in the *R. appendiculatus* male transcriptome, similar to observations in the *R. pulchellus* males

(Tan *et al.*, 2015a). These proteins might play a role in reproduction, since they are abundant in seminal fluid (Findlay *et al.*, 2008; Sonenshine *et al.*, 2011) and Tan *et al.* (2015a) proposed that seminal fluid-like proteins present in male saliva could assist in copulation. Interestingly, many of the differentially expressed transcripts were annotated as putative proteins of which the functions have yet to be elucidated, indicating the large numbers of proteins important in tick feeding that are still uncharacterised.

The assembled *R. appendiculatus* transcriptome was surveyed for the presence and similarity of proteins previously characterised in *R. appendiculatus*. Sex-skewed expression was observed in the nine reference genes used for quantitative RT-PCR analysis (Nijhof *et al.*, 2009), indicating a further level of consideration when selecting reference genes for expression analysis. Next generation sequencing has the advantage of globally investigating the expression profile of many potentially stable genes over a variety of conditions and has previously been used to select reference genes for RT-PCR analysis (Brooks *et al.*, 2011; Tan *et al.*, 2015b). Little to no expression in the salivary tissues was reported for the midgut proteins RAMSPs (Mulenga *et al.*, 2003c) and Ra86-1 (Nijhof *et al.*, 2009), in accordance to the inability to assemble transcripts of these genes in the *R. appendiculatus* salivary transcriptome. In contrast, the Gut cystatin, Ra-cyst-1, was expressed at very low levels in the *R. appendiculatus* transcriptome, even though Imamura *et al.* (2013) showed a lack of expression in female salivary glands. This highlights the dynamic range and sensitivity of NGS and RNA sequencing technologies when compared to conventional sequencing technologies (Wang *et al.*, 2009). However, one of the technical limitations of NGS is the difficulty the software algorithms face when assembling repeat regions (Wang *et al.*, 2009), such as the low complexity repeat regions found in Glycine rich proteins. For this reason the RIM36 gene was assembled in two fragmented transcripts. Without sufficient read support, it was not possible to join the transcripts and they remained as fragmented versions of RIM36 in the final transcriptome.

Many of the *R. appendiculatus* genes that were previously functionally characterised, showed highly gender-specific expression profiles, indicating unique functions required by male and female ticks during feeding. Similar to previous studies - which found male-specific expression for HBPM (Paesen *et al.*, 1999) and IGBP-M (Wang and Nuttall, 1995) genes and female-specific expression for HBP1, HBP2

(Paesen *et al.*, 1999), Kunitz/BPTI-like protein, Ra-KLP (Paesen *et al.*, 2009) and Japanin (Preston *et al.*, 2013) genes - male-specific expression for HBPM and IGBP-Ms and female-specific expression for HBP1, HBP2, Ra-KLP and Japanin were found in the assembled transcriptome. Therefore, expression profiles generated in this study corroborated the publically available knowledge of *R. appendiculatus* genes. However, lower than expected protein identity percentages (although still in range with the comparison between the assembled transcriptome and RaGI) were observed for the assembled RAS-4 (72%) and HBPM (48%) proteins compared to previous work (Paesen *et al.*, 1999; Mulenga *et al.*, 2003d). Also, no homologous proteins were assembled for JL-RA1 (Preston *et al.*, 2013) or TdP1 (Paesen *et al.*, 2007) in this transcriptome. The successful mapping of the sequence reads to the assembled genes but not to the sequences downloaded from NCBI, clearly indicated that the assembly of the low identity copies or the omission of two genes from the transcriptome were not technical assembly errors, but a true reflection of the reads, and by proxy the genes, in the transcriptome. Even though RAS-4 and HBPM have low levels of identity compared to the original protein sequences, they are full-length and contain the expected complete functional domains, motifs and signal signatures. The HBPM homolog assembled here has the same abundant male-specific expression profile as the previously published HBPM (Paesen *et al.*, 1999) and the assembled RAS-4 homolog clustered into the expected phylogenetic clade of serpins (Appendix B: Figure S5). The assembled proteins therefore seem to be functional, although their functions remain to be empirically determined. The absence of JL-RA1 and TdP1 from the *R. appendiculatus* transcriptome might indicate that other proteins have acquired their functions in the salivary glands and these proteins still remain to be identified. Alternatively, Preston *et al.* (2013) and Paesen *et al.* (2007) have shown that the TdP1 gene is not constitutively expressed and the JL-RA1 protein not constitutively active during feeding. If the genes display a very restricted expression pattern at a very specific time during feeding, it might be possible that the absence of the genes from the *R. appendiculatus* transcriptome could be due to the sampling protocol used in the study missing the 'snapshot' of expression of the genes. The sampling design, that included three separate time points to represent unfed, prefed and partially fed ticks, should compensate for this variability in expression, but more exhaustive sampling of additional time points might yet uncover the presence of these genes in the transcriptome.

The variation in some of the assembled salivary proteins compared to the available *R. appendiculatus* protein sequences, especially some very promising proteins for tick control, was unexpected. Out of 17 previously functionally characterised proteins, six were either present at low protein identities (ranging from 48 - 88%) or absent from the salivary gland transcriptome altogether. One possible explanation for the divergence observed in the salivary proteins of the *R. appendiculatus* transcriptome is the presence of positive selection. Positive selection in salivary proteins involved in arthropod blood feeding has previously been reported for mosquitos (Chagas *et al.*, 2013; Arcà *et al.*, 2014) and ticks (Dai *et al.*, 2012; Kotsyfakis *et al.*, 2015b). Co-evolution with the host and constant adaptation to the host's immune system would drive rapid expansion and divergence in the tick's salivary protein families, which could explain the variation observed in these proteins. A second explanation for the variation observed in some of the salivary protein sequences in the *R. appendiculatus* transcriptome is the natural variability observed in *R. appendiculatus* populations due to different geographical distribution and climate variation. Differences in diapause behaviour (Madder *et al.*, 2002), size of the tick body (Speybroeck *et al.*, 2004) and vector competence (Ochanda *et al.*, 1998) has caused naturally occurring *R. appendiculatus* to be clustered into three groups: eastern African, southern African and an intermediate 'transition' group (Madder *et al.*, 1999). On a molecular level, the *cytochrome c oxidase subunit I* (*cox1*) gene separated the southern and 'transition' groups into two genetically differentiated clades (Mtambo *et al.*, 2007c) and the *cox1* and *12S ribosomal RNA* (*rRNA*) genes separated the southern and eastern African groups (Mtambo *et al*., 2007b). These results indicated that the three *R. appendiculatus* groups classified by Madder *et al.* (1999) are genetically distinct from each other, albeit without enough support to be classified as subspecies. Ticks sequenced in this study were from a laboratory maintained stock at ARC-OVR, South Africa (southern African *R. appendiculatus* group), while most of the previously characterised *R. appendiculatus* proteins originated either from tick stocks maintained at the International Livestock Research Institute (ILRI, Kenya, eastern African *R. appendiculatus* group) or were undisclosed. Knowing that *R. appendiculatus* from different geographic locations are genetically distinct may explain the variability observed in the salivary proteins sequenced in this study compared to those sequenced in previous studies. Furthermore, similar to the ticks sequenced in the current study, most of the previous studies have been preformed on laboratory tick colonies (or as mentioned, were undisclosed).

Previously, a high degree of genetic divergence has been observed between different laboratory-bred *R. appendiculatus* tick colonies (Kanduma *et al*., 2016b), which might be a third possible explanation for the differences observed between the sequences reported here and publically available sequences that originated from different tick colonies. Kanduma *et al*. (2016b) showed that the laboratory-bred tick colonies exhibited high levels of inbreeding and most were significantly divergent from each other and the wild populations they were initially sampled from. The finding that a number of very important salivary proteins differed in sequence between the current assembled *R. appendiculatus* transcriptome and previously characterised proteins of *R. appendiculatus*, could have serious implications for the control of the tick by means recombinant vaccines. It emphasises the importance of knowing which salivary protein sequences are present in the tick populations from which protection is desired as anti-tick vaccines designed against protein sequences not present in the tick population, or against proteins that have diverged significantly, might render the resultant vaccines ineffective.

2.6 Conclusions

This is the first study that assembled a *de novo* sialotranscriptome of *R. appendiculatus* female and male ticks using NGS sequencing technologies. The transcriptome is of high quality and improves on the previously generated sequence dataset of *R. appendiculatus*. Differences in the abundance of certain secretory families resulted in unique salivary protein compositions for female and male ticks. Some of the most abundantly expressed transcripts were proteins of unknown function, highlighting the current shortfalls in the understanding of tick feeding. Sequence differences were observed in some of the previously functionally characterised proteins assembled in the *R. appendiculatus* transcriptome, potentially resulting from positive selection, natural *R. appendiculatus* population diversification or genetic isolation due to inbreeding in tick colonies. These differences will have serious implications for the control strategies of *R. appendiculatus* using recombinant vaccines. The transcriptome of *R. appendiculatus* is one of only a small number of tick sialotranscriptomes available to date that together will assist in the characterisation of tick proteins and protein families and will therefore improve the understanding of tick feeding, host-interaction and tick biology as a whole.

2.7 Authors' contributions

Conception and design of the study: MHD and BJM. Supply of funding: BJM. Tick colony maintenance, tick feeding and dissections: DD, RP and BJM. Execution of experimental work, data acquisition, bioinformatic analyses, data interpretation and manuscript drafting: MHD. Revision of the manuscript: MHD DD RP AAL DJGR BJM. All authors read and approved the final version of the manuscript.

2.8 Acknowledgments

We kindly thank Dr. Vinet Coetzee of the Department of Genetics, University of Pretoria, South Africa for suggestions and critical review of the manuscript. MHD was supported by a National Research Foundation/ Department of Science and Technology - Professional Development Program (NRF/DST-PDP) scholarship. BJM was funded by the NRF Incentive Funding for Rated Researchers (NRF-Mans) and the Economic Competitive Support Programme (30/01/V010).

CHAPTER 3

Salivary gland transcriptomics of *Rhipicephalus zambeziensis* ticks during feeding reveal dynamic transcription of secretory proteins, suggestive of stringent temporal regulation

Minique H. de Castro [1,2,3], Daniel de Klerk [3], Ronel Pienaar [3,4], D. Jasper G. Rees [1,2] and Ben J. Mans [1,3,4]

[1] *College of Agriculture and Environmental Sciences, University of South Africa, Johannesburg, South Africa*

[2] *Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa*

[3] *Epidemiology, Parasites and Vectors, Onderstepoort Veterinary Research, Agricultural Research Council, Onderstepoort, South Africa*

[4] *Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa*

This chapter is largely based on work that was published in a peer-reviewed journal:

de Castro MH, De Klerk D, Pienaar R, Rees DJG & Mans BJ (2017) Parasites & Vectors. 10, 384. doi: 10.1186/s13071-017-2312-4.

The version presented here is not the final print version.

3.1 Abstract

Ticks secrete a diverse mixture of secretory proteins into the host to evade its immune response and facilitate blood feeding, making secretory proteins attractive targets for the production of recombinant anti-tick vaccines. The largely neglected tick species, *Rhipicephalus zambeziensis*, is an efficient vector of *Theileria parva* in southern Africa but its available sequence information is limited. Next generation sequencing has advanced sequence availability for ticks in recent years and has assisted the characterisation of secretory proteins. This study focused on the *de novo* assembly and annotation of the salivary gland transcriptome of *R. zambeziensis* and the temporal expression of secretory protein transcripts in female and male ticks, before the onset of feeding and during early and late feeding. The salivary gland transcriptome of *R. zambeziensis* yielded 23 631 transcripts from which 13 584 non-redundant proteins were predicted. Eighty six percent of these contained a predicted start and stop codon and were estimated to be putatively full-length proteins. A fifth (2569) of the predicted proteins were annotated as putative secretory proteins and explained 52% of the expression in the transcriptome. Expression analyses revealed that 2832 transcripts were differentially expressed among feeding time points and 1209 between the tick sexes. The expression analyses further indicated that 57% of the annotated secretory protein transcripts were differentially expressed. Dynamic expression profiles of secretory protein transcripts were observed during feeding of female ticks. Whereby a number of transcripts were up-regulated during early feeding, presumably for feeding site establishment and then during late feeding, 52% of these were down-regulated, indicating that transcripts were required at specific feeding stages. This suggested that secretory proteins are under stringent transcriptional regulation that fine-tunes their expression in salivary glands during feeding. No open reading frames were predicted for 7947 transcripts. This class represented 17% of the differentially expressed transcripts, suggesting a potential transcriptional regulatory function of long non-coding RNA in tick blood feeding. The assembled sialotranscriptome greatly expands the sequence availability of *R. zambeziensis*, assists in the understanding of the transcription of secretory proteins during blood feeding and will be a valuable resource for future vaccine candidate selection.

Keywords: *Rhipicephalus zambeziensis*; Next generation sequencing; Tick salivary glands; *de novo* transcriptome assembly; Sialotranscriptomics; Differential expression; Secretory proteins.

3.2 Introduction

Rhipicephalus zambeziensis is naturally distributed through eastern and southern Africa and is a vector of *Theileria parva*, the causative agent of Corridor disease (CD), East Coast fever (ECF) and January disease (Walker *et al.*, 1981; Lawrence *et al.*, 1983; Norval *et al.*, 1992). *Rhipicephalus appendiculatus*, the predominant vector of *T. parva*, has a much broader distribution through central, eastern and southern Africa (Norval *et al.*, 1992; Walker *et al.*, 2003). *Rhipicephalus zambeziensis* is better adapted to extreme environmental conditions (Madder *et al.*, 2005) and naturally occurs in hotter and drier regions than *R. appendiculatus* (Walker *et al.*, 1981). One of the most pertinent differences between *R. zambeziensis* and *R. appendiculatus* is the variability in vector competence of *T. parva*. During infection experiments, more *R. zambeziensis* ticks were infected and at higher infection loads than *R. appendiculatus* ticks (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998), indicating, at least experimentally, that *R. zambeziensis* is a better vector of *T. parva*. Furthermore, climate projections of increased temperature and decreased rainfall in sub-Saharan Africa (Olwoch *et al.*, 2008), suggest that environmental conditions may become better suited for *R. zambeziensis* and that the tick might spread to a broader distribution. In a similar situation, the competent vector and invasive species *R. microplus* displaces *R. decoloratus* on a wide scale following climatic gradients, resulting in serious tick-borne disease control issues (Lynen *et al.*, 2008). The risk associated with the expansion of *R. zambeziensis*, a highly competent vector of *T. parva*, to a wider distribution due to projected climatic changes, may therefore have serious implications for the control of ECF, CD and January disease in southern Africa.

Corridor disease is an economically important cattle disease in southern Africa, and is caused by buffalo-derived *T. parva* transmitted from the African buffalo (*Syncerus caffer*) to cattle (Uilenberg, 1999). In South Africa, CD is a controlled disease through regular cattle dipping regimes and restricted movement of buffalo. Yet, sporadic outbreaks occur on cattle farms in close proximity to game farms (Mbizeni *et al.*, 2013). Corridor disease is self-limiting in cattle and cattle-to-cattle transmissions have not been observed in South Africa (Neitz *et al.*, 1955; Neitz, 1957; Potgieter *et al.*, 1988; Thompson *et al.*, 2008; Mbizeni *et al.*, 2013). Still, the risk of changing to a cattle-to-cattle transmission disease could have serious implications (Yusufmia *et al.*, 2010) and within the South African boarders no treatments or vaccines resulting in a carrier-state

in cattle may be used. The only alternative to protect cattle from CD in South Africa is to design anti-tick vaccines that would protect cattle against *R. zambeziensis* infestations and the first step towards this goal is the compilation of a comprehensive sequence dataset of *R. zambeziensis*.

Prior to the start of this study, only 33 *R. zambeziensis* sequences (nucleotide and protein) were publically available. To alleviate the sequence shortcomings and generate a comprehensive sequence dataset of *R. zambeziensis*, the aim of the current study was to assemble and annotate a high quality sialotranscriptome of female and male ticks during different feeding stages. The assembled transcriptome was used as reference to determine abundances and differential expression of transcripts and protein families during feeding. This is the first report of a *de novo* transcriptome of *R. zambeziensis* using next generation sequencing (NGS). This transcriptome will likely prove invaluable for future comparative studies analysing multi-genic secretory protein phylogenies and to broaden the understanding of tick biology and the proteins involved in blood feeding and host immune modulation. Additionally, the comprehensive transcriptome will assist in the selection of candidates for future vaccine developmental studies.

3.3 Methods

3.3.1 Ethics statement

Ethical approval was obtained from the College of Agriculture and Environmental Sciences - Animal Ethics Review Committee of the University of South Africa (approval number: 2014/CAES/098), the Agricultural Research Council - Onderstepoort Veterinary Research (ARC-OVR) Animal Ethics Committee (approval numbers: AEC01.15 and AEC12.11), and the Department of Agriculture, Forestry & Fisheries in terms of doing research under Section 20 of the Animal diseases act, 1984 (Act no. 35 of 1984, approval number: 12/11/1/1). Approval letters can be found in Appendix A.

3.3.2 Tick feeding and salivary gland dissection

A parasite-free *R. zambeziensis* tick colony is maintained under standard laboratory and tick-rearing protocols at ARC-OVR (Heyne *et al.*, 1987). The colony was initiated from ticks collected from vegetation in the Marakele National Park, South Africa. Adult ticks were fed in feeding bags on disease-free Hereford (*Bos taurus*) cattle and carefully removed at three and five days after attachment. Unfed ticks, from the same batch of engorged nymphs as the fed ticks, were also processed. Twenty male and twenty female ticks were removed at each time point. Ticks were dissected directly after removal from the bovine and the salivary glands extracted and stabilised in RNAlater (Qiagen, Valencia, CA) according to the manufacturer's specifications until RNA extraction.

3.3.3 RNA extraction, library preparation and sequencing

Total RNA was extracted using the RNeasy Protect Mini Kit (Qiagen) and residual genomic DNA removed by *DNase I* (Qiagen) digestion. Approximately 2 µg of total RNA was used for library generation in the TruSeq RNA Sample Preparation kit (Illumina, San Diego, CA). Fragment fractions of ± 300 bp were excised after agarose gel electrophoreses to prepare the libraries for HiScanSQ 100 bp paired end sequencing. Additionally, a pooled library from an equimolar mixture of all six samples was generated for MiSeq sequencing. This library was prepared from longer RNA fragments by a shortened RNA fragmentation step (changed from 8 to 3 minutes) in the TruSeq Sample Preparation kit, according to the manufacturer's specifications. A high molecular weight library fraction (± 600 - 1000 bp) was excised for subsequent MiSeq sequencing, generating 300 bp paired end reads. Sequencing was performed on the HiScanSQ and MiSeq Illumina instruments at the Biotechnology Platform Sequencing Facility (Agricultural Research Council, South Africa).

3.3.4 Read quality filtering, *de novo* transcriptome assembly and evaluation

Different software packages than the ones used during *R. appendiculatus* assembly (Chapter 2) were used in the assembly process of *R. zambeziensis*, due to improvements in algorithms of read quality filtering and transcriptome evaluation software. Trimmomatic version 0.32 (Bolger *et al.*, 2014) was used to remove Illumina adaptor sequences and low quality bases from the sequence reads using the parameters, ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:20 TRAILING:20

SLIDINGWINDOW:4:20 AVGQUAL:15 MINLEN:50. The SLIDINGWINDOW parameter was adjusted to 7:15 for MiSeq generated sequences. Following quality filtering, paired end MiSeq sequences that overlapped by at least 20 bases, were merged into a single sequence using the CLC Genomics Workbench version 7.5.1 (Qiagen). All quality-filtered and merged reads were pooled together to assemble a single *de novo* transcriptome of *R. zambeziensis* using the Trinity software package version 2.1.1 (Grabherr *et al.*, 2011; Haas *et al.*, 2013). Standard k-mer size of 25 was used and a minimum k-mer coverage of two was selected to reduce the incorporation of potential sequencing errors. An expression level threshold of Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) value of one was used to select against lowly expressed transcripts, that possibly represent assembly artefacts or background expression (Mortazavi *et al.*, 2008; Gan *et al.*, 2010; Hebenstreit *et al.*, 2011). The quality of the assembled transcriptome was evaluated using the Transrate version 1.0.0 software package (Smith-Unna *et al.*, 2016) that estimates the quality of each transcript based on the mapping coverage of sequence reads and the alignment to a closely related reference sequence. The set of well-curated predicted proteins from the sequenced tick genome, *I. scapularis* (IscaW1.4) (Gulia-Nuss *et al.*, 2016), was used as reference. The Transrate transcript evaluation was used as an additional transcript selection process to remove low quality transcripts and improve the confidence in the final transcriptome.

3.3.5 Transcriptome annotation

The transcriptome was BLASTx aligned (E-value < E-05) against the following protein databases: the National Center for Biotechnology Information (NCBI) non-redundant (NR) and Transcriptome Shotgun Assembly (TSA)-NR databases (retrieved April 2016), the UniProt Knowledgebases (UniProtKB/TrEMBL and UniProtKB/Swiss-Prot, retrieved July 2016), the *I. scapularis* predicted peptides (IscaW1.4, retrieved July 2016) (Gulia-Nuss *et al.*, 2016), an Acari database (AcariDB, containing all available mite and tick sequences, as described in Chapter 2) retrieved July 2016, and the EuKaryotic Orthologous Groups (KOG) dataset, retrieved July 2016 (Tatusov *et al.*, 2003). Additionally, the transcriptome was BLASTn searched against the NCBI non-redundant nucleotide (Nt) database (retrieved April 2016). The BLAST2GO software package (Conesa *et al.*, 2005) was used to search the NR BLASTx results for Gene Ontology (GO) terms and visualised by the Web Gene Ontology Annotation Plot (WEGO) (Ye *et*

al., 2006). Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) identifiers from the *I. scapularis* genome were assigned to the transcripts using the KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007). The protein-coding potential of the transcripts were determined by three software packages: Predictor of lncRNAs and mRNAs based on k-mer scheme v1.2 (PLEK) (Li *et al.*, 2014); Coding Potential Calculator (CPC) (Kong *et al.*, 2007); and Coding-Potential Assessment Tool (CPAT) (Wang *et al.*, 2013).

3.3.6 Open reading frame prediction and annotation

Translation frames obtained from transcripts with significant BLASTx searches against AcariDB were used to predict putative open reading frames (ORFs) of 240 bp or longer using the OrfPredictor Server (Min *et al.*, 2005). A second round of prediction for transcripts where ORFs were not predicted was performed using the orffinder.pl script (github.com/vikas0633/perl). The predicted ORFs were translated into amino acid sequences, BLASTp searched against the AcariDB database and compared to the BLASTx results of the transcripts. In cases where the BLAST results were not similar the ORFs were manually inspected and, where possible, ORFs were predicted using the Expert Protein Analysis System (ExPASy) Translate tool (Gasteiger *et al.*, 2003). Predicted proteins were BLASTp searched against the search databases mentioned above as well as against NCBI's Conserved Domain Database (CDD) (Marchler-Bauer *et al.*, 2015) and the Pfam database (Finn *et al.*, 2016). SignalP 4.1 (Petersen *et al.*, 2011) and Phobius (Käll *et al.*, 2007) were used to identify putative signal peptide signatures and TMHMM 2.0 (Krogh *et al.*, 2001), to predict the transmembrane topology of the amino acid sequences. ORFs with no BLASTp or domain-based matches were discarded, after which CD-HIT v4.5.4 (Li and Godzik, 2006) was used to obtain a final non-redundant set of predicted proteins. These proteins were annotated in a priority order of: AcariDB, NR, UniProtKB/TrEMBL, *I. scapularis* proteins, Pfam database and CDD database. The AcariDB was used to classify the proteins into four main classes: putative secretory, housekeeping, unknown function and no hit (proteins for which no significant matches were found in the database) proteins. The final set of predicted proteins was evaluated for completeness by searching for the presence of a conserved set of 1066 arthropod single-copy orthologous proteins using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v2 software (Simão *et al.*,

2015). To evaluate whether the methodology pipeline followed in this study produced a representative set of *Rhipicephalus* salivary gland proteins, the predicted *R. zambeziensis* proteins were compared to the predicted proteins of two closely related tick sialotranscriptomes: *R. pulchellus* (NCBI Bioproject PRJNA170743) (Tan *et al.*, 2015a) and the *R. appendiculatus* transcriptome assembled in Chapter 2 (NCBI Bioproject PRJNA297811) (de Castro *et al.*, 2016). In Chapter 2, the *R. appendiculatus* transcriptome was evaluated by an assortment of quality assessment criteria, resulting in a high quality transcriptome. Therefore, by using the *R. appendiculatus* transcriptome as reference, similarity to it would indicate that the assembled *R. zambeziensis* transcriptome is also of high quality. As a further validation of the assembled transcriptome and predicted proteins, 17 previously characterised *R. appendiculatus* protein sequences (Wang and Nuttall, 1995; Paesen *et al.*, 1999; Bishop *et al.*, 2002; Trimnell *et al.*, 2002; Mulenga *et al.*, 2003d; Paesen *et al.*, 2007; Paesen *et al.*, 2009; Preston *et al.*, 2013) were used in a BLASTp database to identify putative representatives in the *R. zambeziensis* transcriptome.

3.3.7 Differential expression analysis

The *de novo* assembled transcriptome of *R. zambeziensis* was used as reference for differential expression analysis, whereby the quality-filtered sequence reads of each time point was analysed against it separately. Additionally, the filtered reads within each sex were combined to perform differential expression analysis between male and female ticks. The Trinity software package contains custom scripts to facilitate differential expression analysis (Grabherr *et al.*, 2011; Haas *et al.*, 2013). Initially, the reads were mapped to the transcriptome using Bowtie 2 v2.2.3 (Langmead and Salzberg, 2012), followed by abundance estimation with the RNA-Seq by Expectation-Maximization (RSEM) v1.2.31 software package (Li and Dewey, 2011) using the trimmed mean of M values (TMM) normalisation (Robinson and Oshlack, 2010). The transcripts per million (TPM) values were determined and used to estimate the relative abundance of a transcript, also referred to as its expression level (Wagner *et al.*, 2012). Differential expression analysis was performed using the Bioconductor/ Empirical analysis of digital gene expression data in R (edgeR) v3.14.0 software package (Robinson *et al.*, 2010), using the fixed dispersion value of 0.4 (as recommended for biologically non-replicated samples), fold change of > 4 and false discovery rate (FDR) p-value of < 0.01. Chi-

square tests with Bonferroni corrections were used for significance testing in comparisons of protein family expression.

3.3.8 Availability of supporting data

Sequence data supporting the findings of this study have been deposited in public sequence databases. Raw sequence reads have been deposited in the NCBI Short Read Archive (SRA, SRR5438376 - 82) under Bioproject accession number PRJNA381085. The transcript sequences have been deposited in the NCBI Transcriptome Shotgun Assembly (TSA) project under accession number GFPF00000000. The transcriptome version described in this chapter is the first version, GFPF01000000. The predicted protein sequences were deposited in NCBI under the accession numbers MAA11163.1 - MAA24730.1.

3.4 Results

3.4.1 *De novo* assembly and validation of the *R. zambeziensis* transcriptome

To assemble the salivary gland transcriptome of *R. zambeziensis*, between 22 and 37 million paired HiScanSQ sequencing reads were generated for each time point, together with about 22 million paired MiSeq reads from a pool of all time points (Appendix C: Table S1). After adapter trimming and quality filtering, 81% of the HiScanSQ reads were retained in a paired end format and 7% as single ends. Of the MiSeq reads, 31% were retained as paired and 35% as single ends reads. All quality-filtered reads were combined to create a dataset containing approximately 192 million read 1 and 146 million read 2 sequencing reads, which was used for *de novo* assembly of the *R. zambeziensis* transcriptome. In total, 140 703 transcripts were assembled that were filtered based on expression level (FPKM values ≥ 1) and read mapping confidences [as estimated by the transcriptome evaluation software, Transrate (Smith-Unna *et al.*, 2016)]. This resulted in a final transcriptome of 23 631 high confidence transcripts (Table 3.1). Of the sequenced reads, 94% mapped to the initial and 91% to the final transcriptome. This indicated that the nearly one hundred and twenty thousand transcripts excluded from the final assembly represented less than 3% of the total reads, likely representing mostly low quality transcripts. Furthermore, 79% of the transcripts in

the final assembly were near full-length, showing a larger than 75% alignment coverage to their best BLAST matches in the *I. scapularis* protein set. The proteins predicted from the final transcriptome were analysed for assembly and annotation completeness by the BUSCO software package that searches for the presence and completeness of a conserved set of arthropod proteins (Simão *et al.*, 2015). Eighty six percent of the conserved proteins were both present and full-length and only 1.7% of the proteins were fragmented and not assembled in full-length copies. These metrics indicated that a high quality *R. zambeziensis* transcriptome was assembled, which was close to completion, mostly full-length and representative of the sequence reads from which it was build.

3.4.2 Annotation and characterisation of the *R. zambeziensis* transcriptome and predicted proteins

The assembled *R. zambeziensis* transcriptome was annotated by sequence similarity searches against a number of databases, which resulted in 18 311 transcripts (78% of the transcriptome) obtaining a significant match (E-value < E-05) to at least one of the databases (Table 3.1; complete transcript annotation can be found in Appendix C: Table S2). Nearly half of the transcripts were assigned Gene Ontology terms, which included 12 805 cellular components, 20 458 biological processes and 14 603 molecular functions (Appendix C: Figure S1). Based on KOG functional categories, most *R. zambeziensis* transcripts were classified as 'Post translational modification, protein turnover, chaperones', 'General function prediction only' or 'Signal transduction mechanisms' (Appendix C: Figure S2). The KEGG pathway analysis revealed that most transcripts belonged to the 'Ribosome', 'RNA transport' and 'Spliceosome' pathways (Appendix C: Figure S3). Additionally, alignment against the NCBI Nt database retrieved full-length copies of all four ribosomal RNA molecules in the *R. zambeziensis* transcriptome (Appendix C: Table S2).

Table 3.1 Summary of the *R. zambeziensis* transcriptome assembly and annotation statistics.

Transcriptome statistics	**Value[a]**
Transcriptome assembly statistics	
Total number of transcripts	23 631
Number of transcripts > 500 bp	19 903
Number of transcripts > 1 Kb	13 330
Number of transcripts > 10 Kb	80
Shortest transcript length (bp)	201
Longest transcript length (bp)	17 108
Mean length of transcripts (bp)	1793.5
Median length of transcripts (bp)	1193
Transcript N50 (bp)	2807
Total bases in assembly (Mb)	42.4
Ambiguous base calls (Ns)	0
GC content (%)	49
Number of non-redundant predicted proteins	13 584
Transcriptome annotation statistics[b]	
BLASTx against NR	12 756 (54.0%)
BLASTx against UniProtKB/TrEMBL	16 451 (69.6%)
BLASTx against UniProtKB/Swiss-Prot	10 572 (44.7%)
BLASTx against TSA-NR	16 711 (70.7%)
BLASTx against *Ixodes scapularis* predicted peptides	11 804 (50.0%)
BLASTx against EuKaryotic Orthologous Groups (KOG)	9620 (40.7%)
BLASTx against AcariDB	18 245 (77.2%)
Assigned with Gene Ontology (GO) terms[c]	11 360 (48.1%)
Assigned with Enzyme Commission (EC) numbers[c]	3493 (14.8%)
Assigned with KEGG orthology (KO) identifiers[d]	4869 (20.6%)
Annotated in at least one database	18 311 (77.5%)

[a] Value indicating either the number of transcripts, proteins or bases, the transcript length or percentage, as indicated in the table.

[b] Number (and %) of transcripts annotated based on significant matches (E-value < E-05) against databases as detailed in the Methods section.

[c] GO terms and EC numbers assigned with BLAST2GO.

[d] From the KEGG (Kyoto Encyclopedia of Genes and Genomes) Automatic Annotation Server (KAAS) using the *I. scapularis* genome.

In total, 15 737 open reading frames (ORFs) were predicted from 66% (15 684) of the *R. zambeziensis* transcripts. A small fraction of the transcripts (0.2%) encoded for more than one ORF. The predicted ORFs were translated into amino acid (aa) sequences and reduced, based on 100% aa sequence similarity, to a non-redundant set of 13 584 predicted proteins (predicted protein annotation can be found in Appendix C: Table S3). Eighty six percent of these were estimated to be putatively full-length proteins as they contained both a predicted start and stop codon. Of the transcripts for which no ORFs could be predicted (7947 transcripts), 98% were assigned as putative non-coding sequences by at least two coding potential databases (Appendix C: Table S2). Signal peptide signatures were observed in 3488 predicted proteins, 2169 proteins contained transmembrane domains and 8061 proteins showed similarity to at least a single Pfam domain (the thirty most abundant R. *zambeziensis* Pfam domains can be found in Appendix C: Figure S4). AcariDB was used to classify the predicted proteins into: 8139 housekeeping, 2569 secretory, 1706 unknown function and 1170 no hit proteins. Protein annotation was performed based on a priority database order (see Methods section). Finally, full-length gene copies of all thirteen *R. zambeziensis* mitochondrial proteins were predicted and annotated (Appendix C: Table S3).

3.4.3 Comparison of the *R. zambeziensis* transcriptome to other *Rhipicephalus* sialotranscriptomes

The predicted proteins from the assembled *R. zambeziensis* transcriptome were compared to publically available sialotranscriptomes of two closely related *Rhipicephalus* species, of which one was the high quality *R. appendiculatus* transcriptome assembled in Chapter 2. The number of predicted *R. zambeziensis* proteins (13 584) were similar to the number of predicted proteins in the *R. appendiculatus* (12 761) and *R. pulchellus* (11 227) transcriptomes. The lengths of the *R. zambeziensis* proteins (57 - 4928 aa, average 341 aa) were slightly shorter than those predicted for *R. appendiculatus* (70 - 4966 aa, average 400 aa) and *R. pulchellus* (66 - 6645 aa, average 471 aa). However, from the length distribution of the predicted proteins it was evident that the *R. zambeziensis* transcriptome was enriched for shorter proteins without compromising the distribution of longer proteins (Figure 3.1a). Pfam domain prediction and comparison among the species showed that most of the domains (64%) were shared among all the transcriptomes (Figure 3.1b). More domains were

shared between the phylogenetically closer species, *R. zambeziensis* and *R. appendiculatus* (15%), than *R. pulchellus* and any of the two species (about 5% shared with each species). These results showed that the assembled *R. zambeziensis* transcriptome and resulting predicted proteins were comparable to published tick sialotranscriptomes and therefore representative of proteins expected in the salivary glands of feeding *Rhipicephalus* ticks.

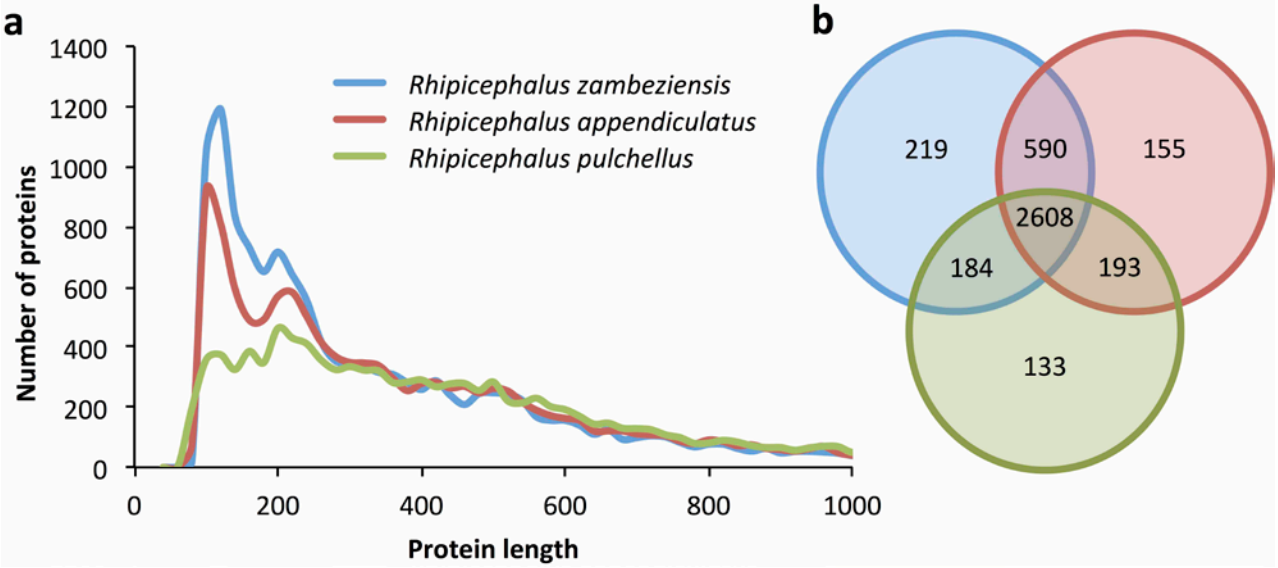


Figure 3.1 Comparisons of the predicted *R. zambeziensis* proteins to proteins of two closely related *Rhipicephalus* species. a Length distribution of the predicted proteins of *R. zambeziensis, R. appendiculatus* and *R. pulchellus*. The number of proteins are indicated based on a protein length sliding window of 20 amino acids (aa), showing a maximum length of 1000 aa. **b** Pfam domain comparison of the three *Rhipicephalus* species. Datasets used: 13 584 predicted proteins from the assembled *R. zambeziensis* transcriptome, 12 761 *R. appendiculatus* proteins (de Castro *et al.*, 2016) and 11 227 *R. pulchellus* proteins (Tan *et al.*, 2015a). Blue represents *R. zambeziensis*; red, *R. appendiculatus;* and green, *R. pulchellus*.

3.4.4 Putative orthologues of functionally characterised proteins in the *R. zambeziensis* transcriptome

To verify that the assembled *R. zambeziensis* transcriptome could be used as a resource in future, representatives of 17 previously characterised *R. appendiculatus* proteins were searched for in the transcriptome. To date no functional studies have been performed in *R. zambeziensis*, although a number of functional studies have been performed in *R. appendiculatus* and based on its close relation to *R. zambeziensis*, it was chosen as target for the homology searches. Representative *R. zambeziensis* sequences were identified for thirteen *R. appendiculatus* proteins based on more than 70% protein identity to the target sequences (Appendix C: Table S4). Nine putative orthologues were obtained when the minimum protein identity was increased to 90%, representing the identification of a putative orthologous sequence for more than half (53%) of the targeted sequences. These included the: Immunoglobulin G binding protein - Male A - C (IGBP-MA - C) (Wang and Nuttall, 1995); Female-specific histamine-binding protein 1 (HBP1) (Paesen *et al.,* 1999); *R. appendiculatus* serine proteinase inhibitor serpins 1 - 3 (RAS-1 to RAS-3) (Mulenga *et al.,* 2003d); *Rhipicephalus* immuno-dominant molecule 36 (RIM36) (Bishop *et al.,* 2002); and Japanin-like-RA1 precursor (JL-RA1) (Preston *et al.,* 2013). Four *R. zambeziensis* proteins shared less than 70% protein identity to their best match, indicating either that these proteins were not assembled in the *R. zambeziensis* transcriptome, that *R. zambeziensis* does not contain orthologues of these proteins or that protein divergence in these families changed the sequences considerably. It is of importance to mention that actual protein function can only be determined by functional protein characterisation, which remains to be performed for the *R. zambeziensis* proteins. Then, except for RIM36 that was assembled in two separate non-overlapping transcripts and RAS-2 that contained no predicted stop codon, all the other *R. zambeziensis* proteins contained both a predicted start and stop codon and were predicted to be full-length. This indicated that the *R. zambeziensis* transcriptome is a valuable resource from which full-length sequences of potential vaccine candidates can be selected in future.

3.4.5 Expression composition in the feeding phases and sexes of the *R. zambeziensis* transcriptome

Between 19 and 30 million clean, paired sequence reads were obtained for each time point after quality filtering (Appendix C: Table S1) and mapped to the assembled transcriptome of *R. zambeziensis* to estimate transcript abundance. Suitable mapping rates of around 90% were achieved for each time point. Overall, a wide expression range of 0.4 - 46 919 TPM was observed in the *R. zambeziensis* transcriptome and only a few transcripts (560 transcripts, 2% of the transcriptome) accounted for 80% of the total expression (Appendix C: Table S2). Two thirds of the assembled transcripts were predicted to be protein-coding and represented 84% of the expression in the transcriptome (Figure 3.2a). Only a small fraction of the predicted proteins were classified as secretory proteins (19%; 2569 proteins), but this protein class represented more than half of the transcription in the coding fraction of the salivary glands (52%, Figure 3.2b). Conversely, the largest protein class at 60%, the housekeeping proteins, represented only 36% of the transcript expression. Within the secretory protein class, 71% of the expression was as a result of a single protein family, the Glycine rich superfamily (Table 3.2). Some of the other large contributors to the secretory protein class were the Lipocalin (6%), Secretory - unknown function (5%) and Bovine pancreatic trypsin inhibitor (3%) families. Further examination of the secretory protein families showed that their expression composition changed greatly over time, resulting in unique secretory protein compositions for each time point in female and male ticks (Figure 3.3a - b; Appendix C: Table S5). The Glycine rich superfamily was the most abundant family in four of the six time points, where it ranged from 67 - 84% of the expression, reaching a maximum contribution to expression at the third day of feeding in each sex. In unfed females the most predominant family was the Histamine release factor (HRF) family (38%) and in day 5 fed females, Lipocalin was the most abundant (26%, Figure 3.3a). Notably, the HRF family that contributed largely to the expression in the salivary glands of unfed female and male ticks consisted of only a single family member (Rzam_Mc198). At a TPM value of 6057, this transcript was the 20th highest expressed transcript in the *R. zambeziensis* transcriptome.

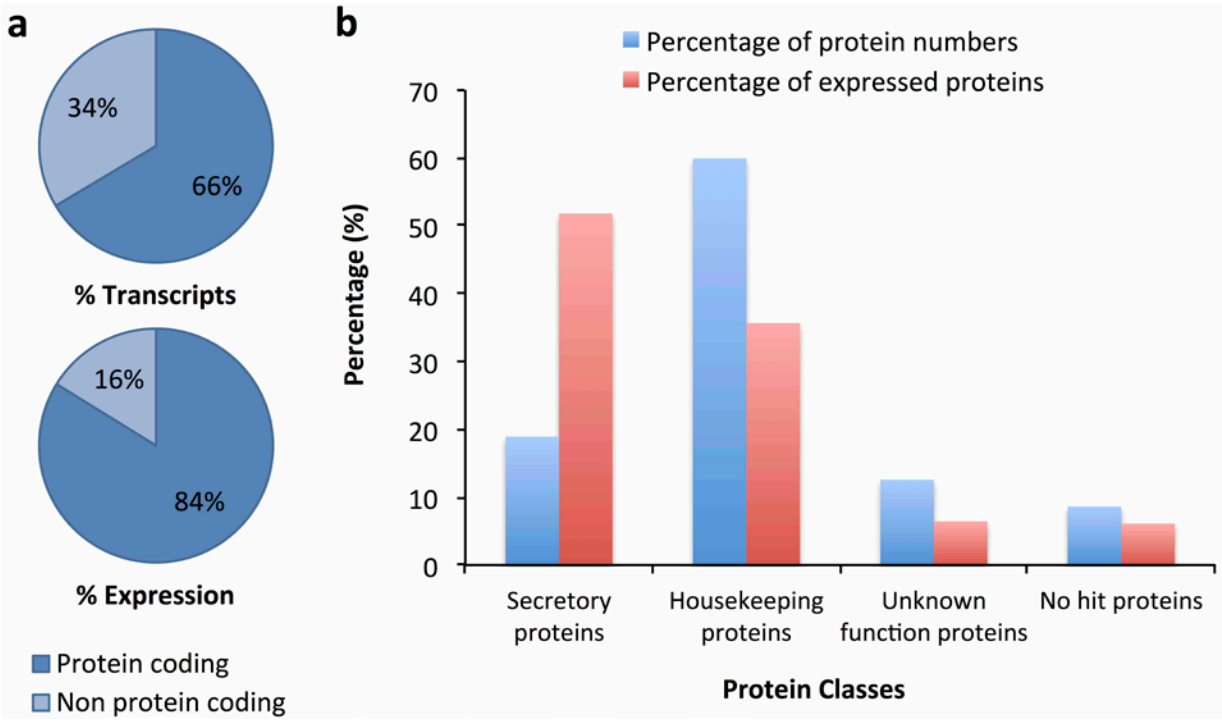


Figure 3.2 Classification and expression analyses in the *R. zambeziensis* transcriptome. a Proportions of predicted protein-coding (indicated by dark blue colouring) and predicted non protein-coding (light blue) transcripts and their contribution to total expression. **b** Proportions of protein numbers (blue) and expression contribution (red) of the four predicted protein classes to the total protein-coding fraction of the transcriptome. Expression was estimated by transcripts per million.

Table 3.2 Characterisation of the secretory protein families in the *R. zambeziensis* transcriptome.

Secretory protein family	Number of family members	Proportion of family members (%)	Average TPM[a] value	Sum of the family expression (TPM[a])	Proportion of family expression (%)
Lipocalin	588	22.90	39.9	23 452.4	5.62
Bovine pancreatic trypsin inhibitor	307	11.95	34.1	10 458.5	2.51
Reprolysin	213	8.29	19.4	4123.6	0.99
TIL domain	177	6.89	38.1	6735.1	1.61
Glycine rich superfamily	161	6.27	1842.4	29 6618.8	71.07
Basic tail secreted protein	111	4.32	43.0	4777.5	1.14
8.9 kDa family	91	3.54	74.3	6758.6	1.62
Digestive system (including Serine proteases)	87	3.39	9.6	833.0	0.20
Mucin	62	2.41	47.1	2922.9	0.70
28 kDa Metastriate family	56	2.18	31.7	1773.5	0.42
Evasin	55	2.14	33.3	1832.9	0.44
Secretory - unknown function	46	1.79	455.6	20 958.6	5.02
Folding, sorting and degradation (including Cathepsins)	46	1.79	88.2	4058.2	0.97
Cystatin	45	1.75	17.3	777.1	0.19
Gluzincin	43	1.67	7.3	315.5	0.08
Serpin	33	1.29	6.3	208.0	0.05
Ixodegrin B	32	1.25	55.3	1770.0	0.42
One of each family	28	1.09	3.7	103.7	0.02
Carboxypeptidase inhibitor	27	1.05	16.0	432.4	0.10
Chitin-binding proteins	26	1.01	20.3	527.9	0.13
5’-Nucleotidase	25	0.97	6.6	163.9	0.04
Transport and catabolism	24	0.93	35.2	845.5	0.20
24 kDa family	21	0.82	25.0	524.9	0.13
7DB family	19	0.74	11.3	215.5	0.05
DA-P36 family	19	0.74	10.2	192.9	0.05
Defensin	19	0.74	211.7	4022.2	0.96
ML domain	17	0.66	417.4	7095.6	1.70
Antigen 5 family	14	0.55	57.6	805.6	0.19
Microplusin	14	0.55	47.6	666.6	0.16
8 kDa Amblyomma family	13	0.51	24.5	318.8	0.08

Sphingomyelinase	11	0.43	2.9	32.2	0.01
Glycan biosynthesis and metabolism	10	0.39	9.4	93.6	0.02
Lipid metabolism	10	0.39	5.9	58.8	0.01
Transcription	10	0.39	3.8	37.5	0.01
Translation	8	0.31	10.6	84.9	0.02
Serine/ threonine protein kinase	8	0.31	6.6	52.7	0.01
Carbohydrate metabolism	8	0.31	3.5	28.2	0.01
Thyropin	7	0.27	22.7	158.7	0.04
Fibrinogen-related domain	7	0.27	20.8	145.9	0.03
Glutathione metabolism	7	0.27	16.6	116.4	0.03
Metalloprotease	7	0.27	8.3	57.9	0.01
Dermacentor 9 kDa expansion	6	0.23	11.8	70.9	0.02
Replication and repair	6	0.23	5.6	33.4	0.01
Immunoglobulin G binding protein A	5	0.19	989.9	4949.4	1.19
Phospholipase A2	5	0.19	9.5	47.7	0.01
Kazal/ vWf domain	4	0.16	4.8	19.1	0.01
Hirudin	3	0.12	108.7	326.2	0.08
SALP15/ Ixostatin	3	0.12	35.5	106.4	0.03
14 kDa family	3	0.12	9.8	29.3	0.01
Kazal domain	3	0.12	9.3	27.9	0.01
Signal transduction	3	0.12	2.2	6.7	0.00
Madanin	2	0.08	153.5	307.1	0.07
Energy metabolism	2	0.08	13.1	26.3	0.01
CDIV	2	0.08	9.4	18.9	0.01
EF hand domain	2	0.08	5.3	10.7	0.00
Histamine release factor	1	0.04	6057.3	6057.3	1.45
Fatty acid-binding protein	1	0.04	70.2	70.2	0.02
Kazal/ SPARC domain	1	0.04	70.1	70.1	0.02
Immune system	1	0.04	21.9	21.9	0.01
Cysteine rich hydrophobic domain 2	1	0.04	10.8	10.8	0.00
26 kDa family	1	0.04	5.5	5.5	0.00
Proline rich	1	0.04	5.1	5.1	0.00
Cysteine rich	1	0.04	2.2	2.2	0.00

[a] TPM (transcripts per million) values were used to estimate expression.

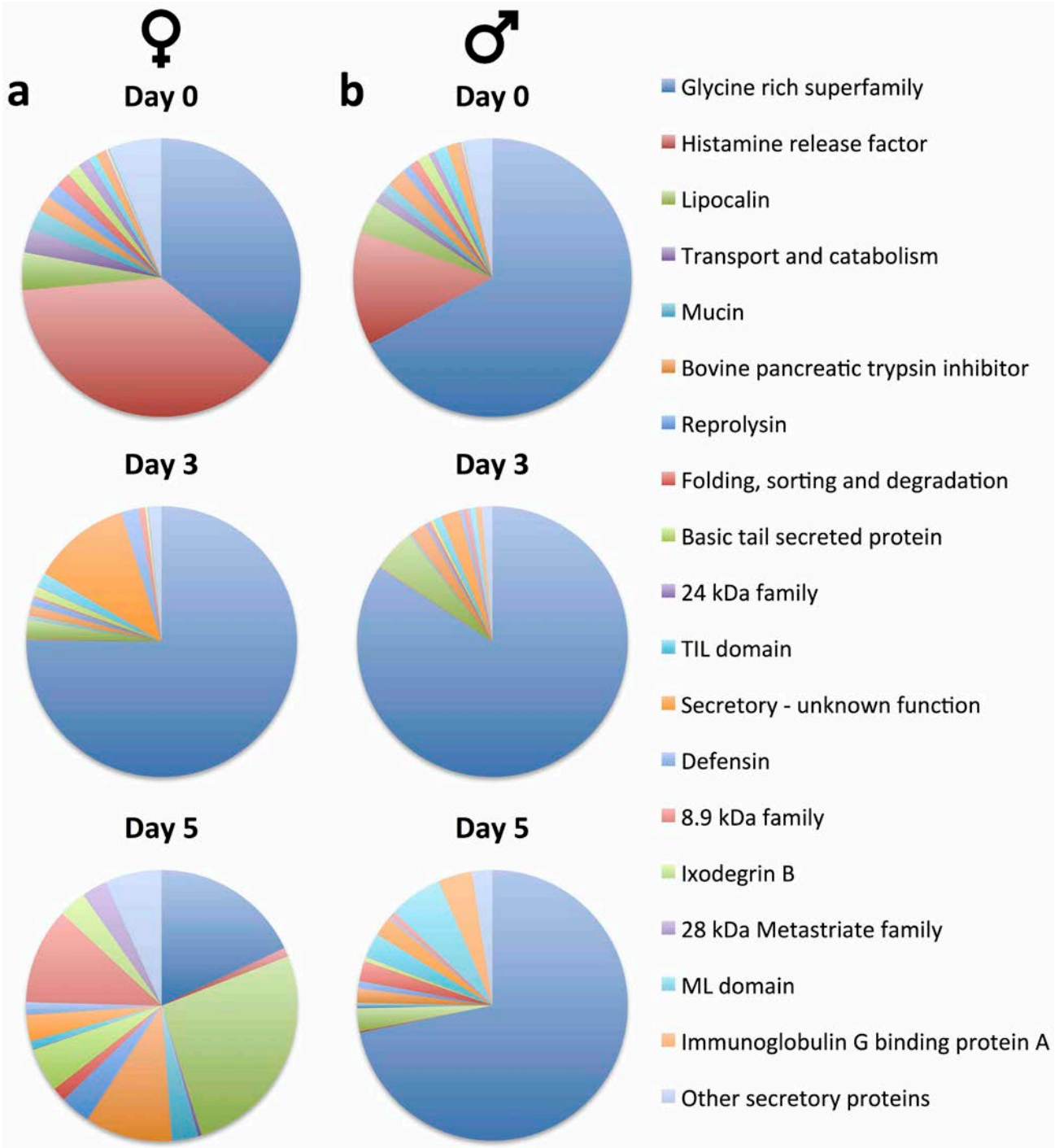


Figure 3.3 Expression proportions of the *R. zambeziensis* secretory protein families during feeding. The proportions of the highest contributing secretory protein families of **a** female and **b** male ticks at different feeding time points are indicated. Expression levels were measured by transcripts per million (TPM). Colour key representing the protein families is indicated. Expression values can be found in Appendix C: Table S5.

3.4.6 Differential expression in the *R. zambeziensis* salivary glands

Pairwise comparisons between time points within each sex identified 2832 significantly differentially expressed transcripts, of which 1927 were specific to the feeding phases of female ticks, 663 were male feeding-specific and 242 were shared between the feeding phases of the sexes. A between-gender comparison was performed by combining the reads of all the time points in each sex prior to mapping and expression analysis. This resulted in 1209 differentially expressed transcripts, of which 641 were up-regulated in females and 568, in males. Half of the differentially expressed transcripts (1470, 49%) were classified as belonging to the secretory protein class (Figure 3.4a). The remaining 572 (19%), 350 (11%), 113 (4%) and 510 (17%) transcripts belonged to the housekeeping, unknown function, no hit or no ORF classes, respectively. The secretory protein families with the largest number of differentially expressed transcripts were the Lipocalin (342 transcripts; 23% of the differentially expressed transcripts in the secretory protein class), Bovine pancreatic trypsin inhibitor (184; 13%), Reprolysin (134; 9%), and Glycine rich (93; 7%) families. Remarkably, 57% of the transcripts classified as secretory proteins were differentially expressed in the salivary glands of *R. zambeziensis* (Figure 3.4b), indicating that ticks use a large repertoire of secretory proteins during feeding. Much smaller proportions, 6 - 21%, of the other classes showed variation in expression and in total 13% of the assembled transcripts were differentially expressed. Furthermore, differentially expressed transcripts accounted for a substantial contribution (45%) of the total expression observed in the salivary glands. This was more pronounced in the secretory protein class, where 72% of the expression was as a result of differentially expressed transcripts. This suggested that the variation observed in the expression composition of the secretory protein families during feeding (Figure 3.3) is under strict transcriptional regulation that continually fine-tunes the expression in the salivary glands.

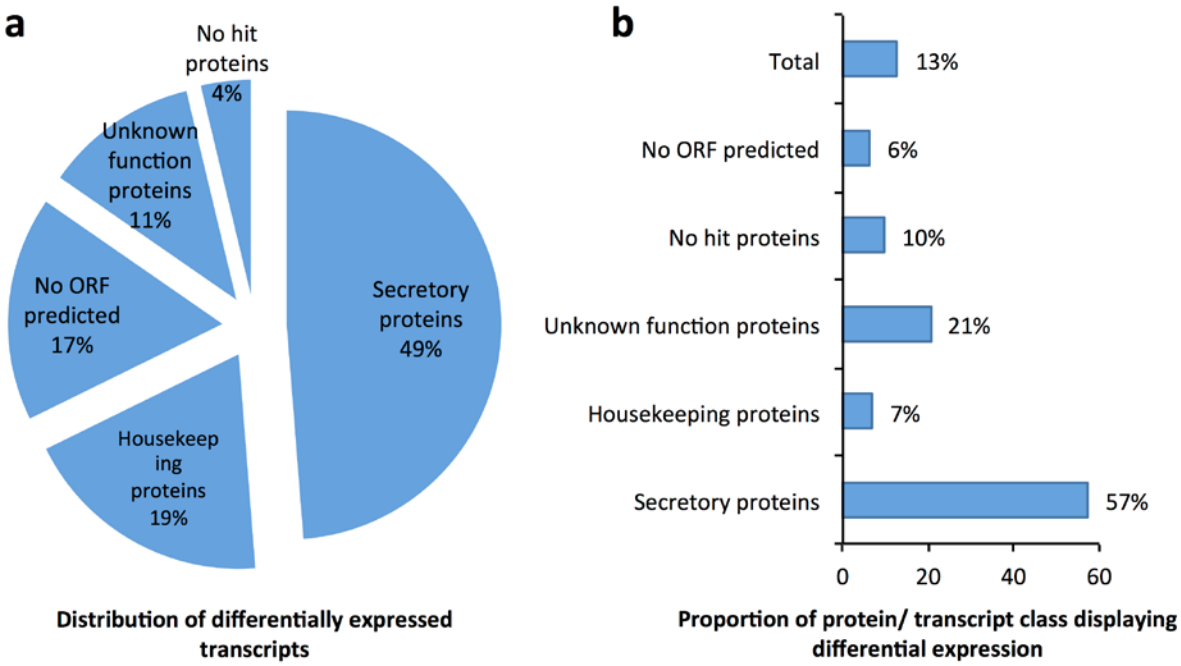


Figure 3.4 Overview of differential expression in the *R. zambeziensis* sialotranscriptome. a Classification of differentially expressed transcripts into different protein or transcript classes. **b** Proportion of differential expression observed within each protein or transcript class. Differential expression analyses were performed using the edgeR software package (with the parameters: fixed dispersion = 0.4, fold change > 4 and FDR p-value < 0.01).

Of the differentially expressed transcripts identified in the between-gender comparison, 376 female and 259 male transcripts were classified as belonging to the secretory protein class (Appendix C: Table S6). Significantly more transcripts of the Lipocalin (140 female vs. 42 male transcripts), 8.9 kDa (36 vs. 12), Reprolysin (27 vs. 3) and 28 kDa Metastriate (23 vs. 2) families were up-regulated in females after implementing a Bonferroni corrected p-value < 0.0013. In males, the TIL (Trypsin Inhibitor-like) domain (4 vs. 29), Folding, sorting and degradation (including Cathepsins; 1 vs. 17), Digestive system (including Serine proteases; 0 vs. 41), Cystatin (0 vs. 26) and 7DB (0 vs. 13) families showed significantly more up-regulated transcripts when compared to females. Of these, Digestive system, Cystatin and 7DB families exhibited male-specific up-regulation, as no family members were up-regulated in females. Other secretory families showed large differences between the sexes, albeit not significant after Bonferroni correction (Appendix C: Table S6).

3.4.7 Dynamic expression patterns of the *R. zambeziensis* secretory protein families

Many transcripts belonging to secretory protein families were significantly up-regulated during early feeding phases (from unfed to day 3 fed ticks) of both female (541 transcripts) and male (335) ticks (Figure 3.5a - b; Appendix C: Table S7). When considering late feeding (feeding progression from day 3 to 5), in specifically females, most of the differentially expressed secretory transcripts were down-regulated (Figure 3.5c). Surprisingly, about 52% of these down-regulated transcripts were the same transcripts that were up-regulated during early feeding, implying that certain secretory transcripts were only required during early feeding stages in females. Hardly any significant differential expression was observed in the male late feeding phase (Figure 3.5d), demonstrating that the male day 3 and 5 time points were highly similar in their secretory family expression. Comparisons between unfed and day 5 fed female and male ticks revealed a large number of up-regulated secretory transcripts in both sexes, together with some down-regulated transcripts in females (Figure 3.5e - f). Only about 42% of these up-regulated transcripts in female ticks were shared with the transcripts up-regulated during early feeding (day 0 to 3), likely due to more than half of the transcripts up-regulated in early feeding undergoing down-regulation in female late feeding. Conversely, in males, due to the few significant expression differences observed between day 3 and 5, a larger overlap of 74% was seen between the up-regulated transcripts in the day 0 to 3 and day 0 to 5 comparisons. Non-significant expression differences between male day 3 and day 5 time points, which would not have been picked up by the current analyses, might explain the differences observed in the up-regulation profiles of the day 0 to 3 and day 0 to 5 comparisons in males (Figure 3.5b and f). Furthermore, comparisons between the number of up-regulated transcripts in each secretory family between day 0 to 3 and day 0 to 5 revealed that significantly more transcripts of the Glycine rich (61 vs. 17 transcripts, p-value < 0.0001) and TIL domain (38 vs. 11, p-value $= 0.00011$) families were up-regulated in earlier feeding points in females (Figure 3.5a and e; Appendix C: Table S7). Also, significantly more of the Digestive system (including Serine proteases, 12 vs. 41, p-value < 0.0001) transcripts were up-regulated in later feeding points in males (Figure 3.5b and f). These results suggested that in tick salivary glands, genes of the secretory protein class are temporally regulated to alter the protein cocktail that is secreted into the host as feeding progresses.

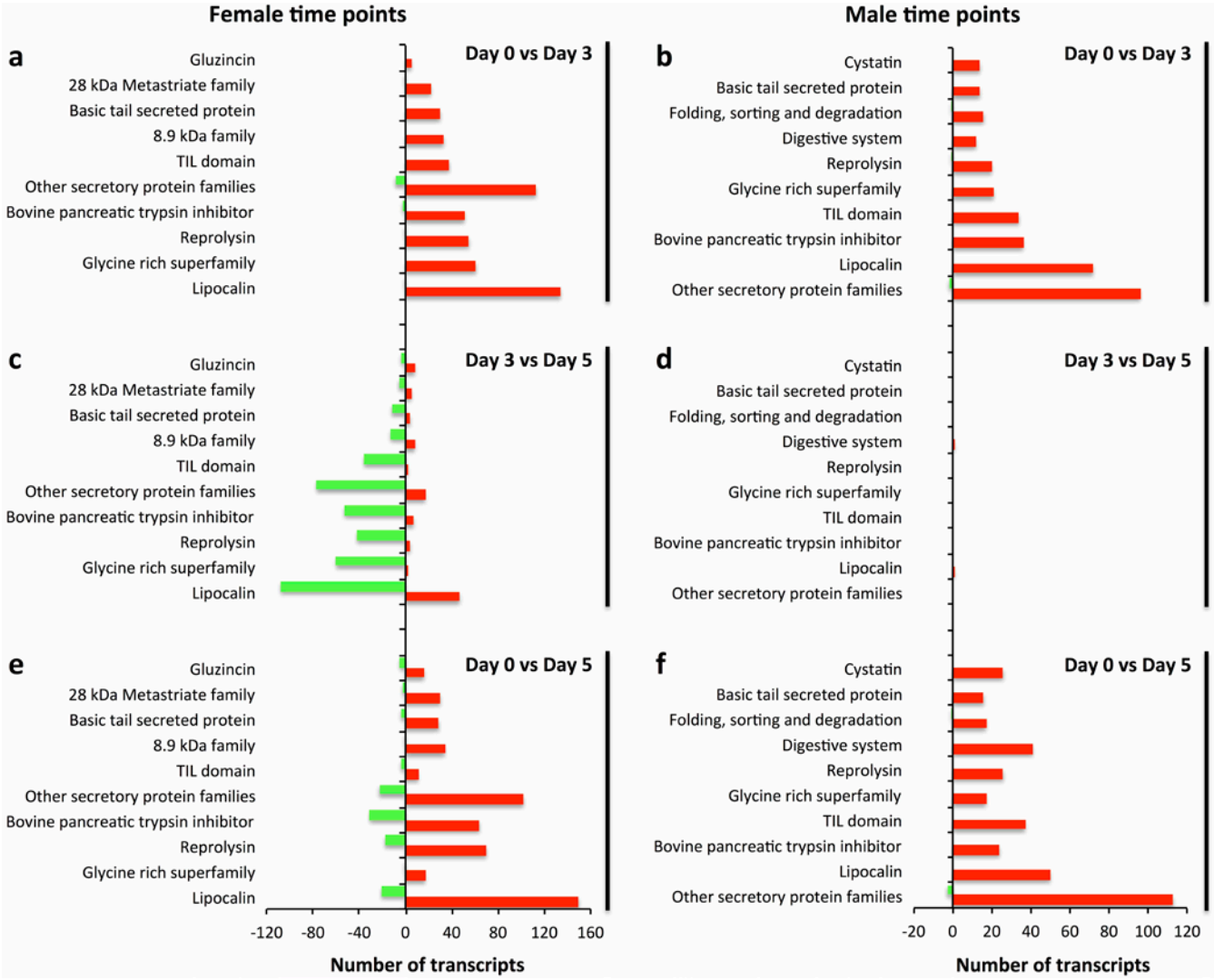


Figure 3.5 Differentially expressed transcripts of secretory protein families in *R. zambeziensis*. The numbers of up- (red colour) and down-regulated (green) secretory protein transcripts, after pairwise comparisons between different feeding time points, are represented. Pairwise comparisons are shown for female: **a** day 0 vs. day 3, **c** day 3 vs. day 5 and **e** day 0 vs. day 5; and male ticks: **b** day 0 vs. day 3, **d** day 3 vs. day 5 and **f** day 0 vs. day 5. The pairwise comparisons are represented as a progression of feeding and show how transcript expression changed from the earlier to the later time point. The edgeR software package (fixed dispersion = 0.4, fold change > 4 and FDR p-value < 0.01) was used for differential expression.

3.5 Discussion

The main aim of this study was the *de novo* assembly, annotation and characterisation of the sialotranscriptome of *R. zambeziensis*, a vector of *T. parva*. Thus far, *R. zambeziensis* has been a largely neglected tick species due to its limited distribution through eastern and southern Africa (Walker *et al.*, 1981), which was reflected by the small number of publically available sequences of *R. zambeziensis* prior to the start of this study; only 31 nucleotide and two protein sequences were available in GenBank. In the present study this shortfall was alleviated by the transcriptome assembly and deposition of 13 584 annotated predicted proteins of *R. zambeziensis* into the public domain. This set of predicted proteins is a valuable resource for future vaccine candidate selection in *R. zambeziensis*, as was shown by the availability of mostly full-length versions of previously characterised proteins. The *R. zambeziensis* transcriptome was constructed from sequence reads from male and female ticks, unfed and representative phases of early and late feeding ticks, to represent a large proportion of genes involved in adult tick feeding. Great care was taken to assemble a high quality *R. zambeziensis* transcriptome that was representative of the reads from which it was assembled, near complete, containing a large proportion of full-length protein sequences and similar to the closely related tick sialotranscriptomes of *R. pulchellus* (Tan *et al.*, 2015a) and *R. appendiculatus*, assembled in Chapter 2 (de Castro *et al.*, 2016). Similarity to the high quality transcriptome of *R. appendiculatus* (assembled and evaluated based on a number of quality assessment criteria in Chapter 2), further indicated that the assembled *R. zambeziensis* transcriptome was of high quality. Furthermore, the number of predicted protein sequences in the *R. zambeziensis* transcriptome was similar to the 11 105 genes predicted to be expressed in the salivary glands of *I. scapularis*, the only tick with a completely sequenced genome (de la Fuente *et al.*, 2016c; Gulia-Nuss *et al.*, 2016). Provisional functions were assigned to the assembled *R. zambeziensis* transcriptome such as histamine (Paesen *et al.*, 1999) and immunoglobulin (Wang and Nuttall, 1995) binding, serine proteinase inhibition (Mulenga *et al.*, 2003d) and immunomodulation (Preston *et al.*, 2013). These functions have been predicted based on the presence of proteins in the *R. zambeziensis* transcriptome showing high protein identity (≥ 90%) to previously characterised *R. appendiculatus* proteins, but the functions of these proteins will require empirical characterisation to be confirmed in *R. zambeziensis*. Furthermore, the predicted *R. zambeziensis* proteins were annotated and 19% were classified as putative secretory proteins, in accordance with 13 - 37% annotated in previously

assembled tick sialotranscriptomes (Karim *et al.*, 2011; Schwarz *et al.*, 2013; Garcia *et al.*, 2014; Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Tan *et al.*, 2015a; de Castro *et al.*, 2016; Ong *et al.*, 2016; Ribeiro *et al.*, 2017). In total, 8139 of the predicted proteins were classified as putative housekeeping proteins, similar to the number of proteins identified in other metastriate transcriptomes (Karim *et al.*, 2011; Garcia *et al.*, 2014; de Castro *et al.*, 2016) and the predicted number of core housekeeping genes in Chelicerata species, approximately 7000 (Mans *et al.*, 2016). Full-length sequences of all 13 mitochondrial genes and all four rRNA molecules were also assembled and can be used in future phylogenetic analyses.

The sequence reads of each time point were independently mapped to the assembled transcriptome to estimate transcript abundance during feeding. During RNA extraction, all ticks in a time point were pooled to obtain enough tissue for extraction and sequencing, similar to methodology followed in other tick transcriptomic studies (Schwarz *et al.*, 2013; Mudenda *et al.*, 2014; Schwarz *et al.*, 2014b; Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Xu *et al.*, 2015; Yu *et al.*, 2015). The edgeR software package can accommodate non-replicated samples using strict parameters (Robinson *et al.*, 2010), resulting in the identification of fewer false positives but also fewer differentially expressed transcripts. Consequently, only large expression differences were considered significant within this work, resulting in high confidence in the assigned differentially expressed transcripts but also the possibility that important feeding genes might have been missed if they had smaller variable expression levels. Nevertheless, 13% of the assembled transcripts showed differential expression and these accounted for nearly half (45%) of the expression in the salivary glands.

The work in this chapter further focussed on the transcriptional response of secretory proteins to tick feeding and these proteins were therefore investigated in greater detail. While the secretory protein class represented only 19% of the predicted proteins in the *R. zambeziensis* salivary glands, this class accounted for the majority of transcript expression (52%). Similar transcript abundance of a small proportion of secretory proteins was seen in other tick sialotranscriptomes: 17% and 63% of the proportions of respectively, predicted proteins and expression were classified as secretory proteins in *R. appendiculatus* (as estimated in Chapter 2); 23% and 62% in *Hyalomma excavatum* (Ribeiro *et al.*, 2017); and 37% and 49% in *Amblyomma*

americanum (Karim and Ribeiro, 2015). Nearly half (49%) of the differentially expressed transcripts in *R. zambeziensis* were classified as secretory proteins. Previous studies have also shown that the majority of differentially expressed transcripts in tick salivary glands are classified as secretory proteins (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b). The abundance of transcripts coding for secretory proteins in tick salivary glands was not unexpected as the salivary glands are in close contact with the host into which secretory proteins are being secreted to escape host immune defences. Remarkably, most (57%) of the predicted *R. zambeziensis* secretory proteins showed significant differential expression, even using the stringent parameters used in this study, and accounted for 72% of the expression observed in the secretory proteins class. These large variations in the *R. zambeziensis* secretory protein expression during feeding resulted in each time point resembling a unique transcriptome with different expression proportions of secretory protein families. These differences were more pronounced in female ticks that require more on-host time to feed to repletion. Similarly, changes in secretory protein family compositions during feeding were seen in the saliva of *R. microplus* (Tirloni *et al.*, 2014a) and *I. scapularis* (Kim *et al.*, 2016).

Differential expression analysis of the *R. zambeziensis* secretory protein families revealed dynamic profiles of transcript expression during feeding of female ticks. Most of the differentially expressed secretory transcripts were up-regulated during early feeding whereas during late feeding most were down-regulated. A similar expression pattern was also observed in *A. americanum* salivary glands (Aljamali *et al.*, 2009). However, in *R. zambeziensis* 52% of the up-regulated transcripts in early feeding were shared with the down-regulated transcripts in late feeding, indicating that these transcripts were specifically required only during early feeding and the establishment of the feeding site in female ticks. Further assessment of the expression regulation in the most differentially expressed multi-genic secretory protein families (Lipocalin, Bovine pancreatic trypsin inhibitor, Reprolysin and Glycine rich), revealed similar patterns of intricate transcript up- and down-regulation and a trend towards preferential expression of transcripts in a single time point in female salivary glands. Correspondingly, members of multi-gene secretory protein families showed substantial differential expression in the salivary glands of *A. americanum* (Karim and Ribeiro, 2015; Bullard *et al.*, 2016) and *I. ricinus* (Kotsyfakis *et al.*, 2015b), where certain transcripts were uniquely expressed during different developmental stages or feeding time points.

Similar profiles were described in *I. scapularis* ticks based on proteomic analyses of saliva collected in 24-hour intervals during feeding (Kim *et al.*, 2016), indicating that transcriptomic profiles of salivary glands are transferable to the proteome being secreted into the host. These results suggest that members of tick secretory protein families are under tight transcriptional regulation that result in complex compositions of secretory proteins in different feeding phases, equipping ticks with the diversity in secretory proteins to evade host immune defences. Indeed, the authors of the above-mentioned studies refer to the expression dynamics observed in tick salivary glands as 'sialome switching' or a form of antigenic variation of secretory proteins to evade host immune recognition while still maintaining host immune modulation to feed successfully. Antigenic variation is a mechanism by which infecting microorganisms systematically alter or 'switch' their surface proteins to remain unnoticed by the host immune defences (reviewed in Deitsch *et al.*, 2009). In ticks, antigenic variation will rely on functional redundancy of the multi-genic secretory protein families, where alternate members could be sequentially expressed to evade immune recognition but still retain the same function or a number of members could be expressed at low levels concurrently to result in an additive functional effect while maintaining low immunogenicity to each protein (Chmelař *et al.*, 2016). Functional redundancy remains to be empirically proven for most of the proteins of multi-genic secretory protein families. The ability of ticks to preserve crucial blood feeding functionalities while simultaneously evading the host immune response by altering exposed antigens, make ticks remarkably well adapted to feed on their hosts for extended periods of time.

The above-mentioned studies were based on results generated from female ticks (Karim and Ribeiro, 2015; Kotsyfakis *et al.*, 2015b; Bullard *et al.*, 2016; Kim *et al.*, 2016). Both female and male ticks were analysed in the *R. zambeziensis* sialotranscriptome, showing that the expression profiles of secretory proteins were much less complex in males compared to females during feeding. Similar to the low expression variation in male *A. americanum* ticks during feeding (Aljamali *et al.*, 2009), only five differentially expressed transcripts were observed between the third and fifth day of feeding of male *R. zambeziensis* ticks, indicating that the male sialotranscriptome necessary for successful feeding had mostly already been established by the third day of feeding. Furthermore, a number of transcripts were differentially expressed between female and male ticks, indicating that some secretory protein families were differentially

regulated between the sexes and that female and male ticks feed and evade host immune defences using different mechanisms. This was similar to what was observed for *R. appendiculatus* in Chapter 2 (de Castro *et al*., 2016) as well as in other studies (Aljamali *et al.*, 2009; Xiang *et al.*, 2012; Tan *et al.*, 2015a). Male ticks feed by continuously attaching and de-attaching from the host to be in close proximity to female ticks (Sonenshine, 1991) in order to mate and assist with establishing a stable feeding cavity from where females can feed to repletion to complete their life cycles. This might mean that male ticks need a constant supply of secretory proteins to be ready to feed at any time or to secrete decoy antigens into the host to assist females during feeding, which has been shown for male-specific Immunoglobulin-binding proteins (Wang *et al.*, 1998). Furthermore, the Serine protease, Cathepsin and Cystatin families, were found to be nearly exclusively up-regulated in male *R. zambeziensis* salivary glands when compared to females, similar to families found up-regulated in males in *R. appendiculatus* (Chapter 2, de Castro *et al*., 2016) and *R. pulchellus* (Tan *et al*., 2015a). During mating, male ticks salivate on their spermatophores before inserting them into the female genital pore using their mouthparts (Feldman-Muhsam *et al.*, 1970). Proteases and protease inhibitors have abundantly been found in insect seminal fluid (Findlay *et al.*, 2008; Sonenshine *et al.*, 2011) and the up-regulation of their transcripts in male salivary glands might suggest an additional function for these secretory proteins in tick reproduction (as proposed by Tan *et al.*, 2015a).

In order to alter the exposed antigens presented to their hosts and to feed unnoticed until repletion, ticks must have evolved stringent temporal regulatory processes to sequentially express secretory proteins, of which the mechanism is still largely unknown. Adamson *et al*. (2013) showed that *SDS3*, a component of the Sin3 histone deacetylase corepressor complex involved in histone modification and repression of transcription, was particularly down-regulated concurrently with a number of expression differences in secretory protein transcripts in *A. maculatum* ticks. The authors proposed that tick secretory proteins might, to some degree, be under epigenetic regulation by histone modification and chromatin remodeling. Based on these results, 34 genes associated with histone modification were identified in the transcriptome of *I. ricinus* (Kotsyfakis *et al.*, 2015b). Additionally, recent surveillance of five histone and 34 histone modifying enzyme transcripts in *I. scapularis* indicated that *Anaplasma phagocytophilum* altered tick epigenetics to assist pathogen infection and multiplication

during infection (Cabezas-Cruz *et al.*, 2016a), which changed the expression of the tick's salivary transcripts (Ayllón *et al.*, 2015). The relationship between epigenetic gene regulation and long non-coding RNA (lncRNA) has been well-established in recent years (as reviewed in Kung *et al.*, 2013; Mercer and Mattick, 2013; Engreitz *et al.*, 2016), by which lncRNAs bind to and act as scaffolds between specific sequences in the genome to be transcriptionally regulated and chromatin-modifying enzymes. Long ncRNAs are RNA molecules larger than 200 bp that contain no open reading frames (> 300 bp) for protein translation (Dinger *et al.*, 2008). These lncRNAs have been largely unexplored in ticks, although a set of 4439 predicted non-coding RNA genes were annotated in the recently completed *I. scapularis* genome (Gulia-Nuss *et al.*, 2016). Also, in Chapter 2, 7414 of the assembled *R. appendiculatus* transcripts contained no predicted ORFs and this set of transcripts represented a striking 40% of the differentially expressed transcripts in the salivary glands (de Castro *et al.*, 2016). Similarly, in the current transcriptome assembly of *R. zambeziensis*, no ORFs could be predicted for a third of the assembled transcripts (7947 transcripts). These transcripts represented 16% of the total expression in the salivary glands and most (98%) showed low protein-coding potential (Kong *et al.*, 2007; Wang *et al.*, 2013; Li *et al.*, 2014). These no ORF-containing transcripts also represented 17% (510 transcripts) of the differentially expressed transcripts in *R. zambeziensis* during feeding. This set of transcripts with no predicted ORFs likely contains putative lncRNAs, but also probably protein-coding transcripts for which the predicted ORFs were too small to be retained or potentially misassembled sequences of no biological significance. These protein-coding transcripts are unfortunately not easily distinguishable from lncRNAs and warrants further examination. The large number of differentially expressed transcripts in feeding *R. zambeziensis* ticks, for which no ORFs could be predicted, suggests that lncRNA molecules might be involved in tick blood feeding, potentially through the transcriptional regulation of tick secretory proteins. This could either be as a complementary mechanism to the proposed epigenetic regulation by Adamson *et al*. (2013) or by other more direct regulatory functions of lncRNAs, such as: acting as transcriptional co-regulators or co-repressors, binding to transcription factors to act as decoys, controlling alternative splice variants, stabilising mRNA by sequestering micro RNAs (miRNAs) away from targets, to name but a few (reviewed in Kung *et al.*, 2013; Engreitz *et al.*, 2016).

The role of lncRNAs in the regulation of the vertebrate immune system has been well established in the last years (reviewed in Aune and Spurlock, 2016; Zhang and Cao, 2016), although the function in arthropod immunity remains to be determined. Recently, a potential role for lncRNAs in vector immunity has been identified by the increase of a number of *Aedes aegypti* lncRNAs in response to virus and endosymbiont infection (Etebari *et al.*, 2016). It has also been shown that the vertebrate immune system can be modulated by viral lncRNAs to enhance virus survival during infection of the host (Rossetto and Pari, 2011; Scaria and Pasha, 2013). Apart from normal secretory processes, tick salivary glands also excrete molecules into the saliva by a mechanism known as apocrine secretion, where whole pieces of the cells are shed and cytoplasmic contents excreted into the lumen of the acinus (Shaw and Young, 1995; Farkaš, 2015; Mans *et al.*, 2016). It is conceivable that should it be proven that ticks express lncRNAs that target host defences, this excretion mechanism might be the entry point of the lncRNAs into tick saliva. The functions of lncRNAs in arthropods are still elusive, but the significant number of differentially expressed no ORF-containing transcripts observed in *R. zambeziensis*, the extensive transcriptional regulatory functions described for lncRNAs and the potential host immunomodulation by parasite-derived lncRNAs, warrant further investigations of these important RNA molecules in ticks.

Similar to observations made in Chapter 2 about *R. appendiculatus*, it was found that the Glycine rich superfamily was an exceptionally large contributor (71%) to the total expression of the secretory protein class in *R. zambeziensis*. Maruyama *et al*. (2010) previously showed that tick species with short mouthparts, for e.g. *Rhipicephalus*, require large quantities of Glycine rich proteins, presumably to form the cement-cone for adhesion to the host's skin (Binnington and Kemp, 1980; Sonenshine, 1991). The findings of Maruyama *et al*. (2010) were based on only a few species with limited expressed sequence tag (EST) data, but the availability of NGS generated tick sialotranscriptomes have extended the hypothesis to more species with better confidence. Accordingly, the expression contribution of the Glycine rich superfamily to the secretory protein class in ticks with short mouthparts was between 48 - 71% based on three *Rhipicephalus* species (Tan *et al*., 2015a; Chapter 2; current study) and between 3 - 28% for ticks with long mouthparts based on four *Amblyomma* (Garcia *et al.*, 2014; Karim and Ribeiro, 2015) and a single *Hyalomma* (Ribeiro *et al.*, 2017) species. *Rhipicephalus* (0.34 mm) has much shorter mouthparts than *Hyalomma* (0.62 mm) and

Amblyomma (1.27 mm) ticks, with no overlap between the three groups (Slovák *et al.*, 2014). Another observation in the *R. zambeziensis* transcriptome was that the Glycine rich superfamily was expressed nearly twice as much in male than female ticks (382 258 TPM in males and 202 292 TPM in females), an observation shared with *R. appendiculatus* (Chapter 2, de Castro *et al*., 2016) and *R. pulchellus* (Tan *et al*., 2015a). A large contribution of Glycine rich proteins was observed in early feeding females, followed by down-regulation in late feeding, as would be expected for the establishment of the cement-cone, which is generally created within the first three to four days after attachment (Moorhouse and Tatchell, 1966). Contrary, in male salivary glands, Glycine rich proteins remained disproportionally overrepresented in all time points. In light of the cement-cone, the abundance of the Glycine rich superfamily in males is counterintuitive as one would expect female ticks, considering their extended feeding times and exceptional expansions in body size (Sonenshine, 1991), to require a larger cement-cone, and therefore more Glycine rich protein expression. However, the Glycine rich superfamily is a large family to which a number of alternative functions have been ascribed that might be involved during feeding and employed by male ticks to maintain the feeding site. For instance, some Glycine rich proteins have shown to function as antimicrobial peptides in insects (reviewed in Yi *et al.*, 2014) and might assist prolonged tick feeding by keeping the feeding cavity free from microbial infections. Additionally, some Glycine rich proteins contain RNA-recognition motifs to bind RNA during transcriptional regulation and splicing (reviewed in Rogelj *et al.*, 2012) or to potentially bind nucleic extrusions from neutrophil extracellular traps of the host defence response [as proposed by Maruyama *et al*. (2010)]. Furthermore, in ticks, some Glycine rich proteins have shown to be immuno-dominant (Bishop *et al.*, 2002; Trimnell *et al.*, 2005) and might be used by the males as decoy antigens to preoccupy the host defences, thereby assist female feeding (Wang *et al.*, 1998).

In addition, the *R. zambeziensis* data revealed that even though the expression level of the Glycine rich superfamily was double in males than females, the number of differentially expressed transcripts were considerably fewer: six male and 18 female differentially expressed transcripts for the gender-specific comparison; and 22 male and 89 female differentially expressed transcripts for the feeding-specific comparisons. Similarly, in *R. pulchellus*, two male and 19 female-specific up-regulated transcripts were found (Tan *et al.*, 2015a). These data suggest that female ticks require a more

diverse set of Glycine rich proteins in their salivary glands, potentially indicative of the generation of a more complex, longer lasting cement-cone that might facilitate the prolonged feeding times and extreme body size expansions seen in female ticks (Sonenshine, 1991). Alternatively, the Glycine rich protein diversity observed in female ticks might be a mechanism of antigenic variation. Antigenic variation has been proposed as the mechanism by which *R. microplus*, a one-host tick that remains on the same host for all of its life stages, evades host immune defences by displaying a larger variety of Glycine rich proteins when compared to two- or three-host ticks (Maruyama *et al.*, 2010). As mentioned before in Chapter 2, the nucleotide sequences of the Glycine rich proteins are low in complexity and contain multiple repeats. A technical limitation of assembly algorithms when using short read NGS sequencing data, such as Illumina sequences, is the difficulty in assembling low complexity and repeat sequences (Wang *et al.*, 2009). This becomes apparent when considering that from all the predicted *R. zambeziensis* proteins only 5% had no predicted methionine start codons, but when assessing only the predicted Glycine rich proteins, 35% had no predicted start codons, indicating a number of fragmented or incomplete sequences in this family. An example of this was seen by the inability to assemble the RIM36 (Bishop *et al.*, 2002) Glycine rich orthologue of *R. zambeziensis* into a single transcript. Interestingly, in the *R. appendiculatus* transcriptome (Chapter 2), the RIM36 protein was also assembled into two separate non-overlapping transcripts. The inability to assemble this protein into a single transcript by two separate transcriptome sequencing attempts highlights the difficulty of assembling repeat regions using short sequence reads. Alternatively, it might also be an indication that the initial RIM36 reported by Bishop *et al*. (2002) might have been erroneous. The true integrity of RIM36 remains to be determined in future. Third generation sequencing technologies, Pacific BioSciences Iso-Seq and Oxford Nanopore MinION, generate sequence reads with much longer lengths (> 10 Kb) than Illumina and are used to improve general transcriptome assembly complications, e.g. repeat regions and alternative splice variants (Bolisetty *et al.*, 2015; Rhoads and Au, 2015). In future, the use of these technologies will assist with the assembly of full-length transcripts for members of the Glycine rich superfamily, e.g. RIM36, and tick multi-gene families in general.

Another major contributing family to the transcriptome expression of specifically unfed ticks was the Histamine release factor (HRF) family, representing 38% of female and 13% of male transcript expression in unfed ticks. This family consisted of a single protein, Rzam_Mc198, which showed 73% protein identity to the *I. scapularis* tick histamine release factor (tHRF, AAY66972.1) (Dai *et al.*, 2010). Tick HRF binds to basophil cells and induces the release of histamine (Mulenga *et al.*, 2003a; Dai *et al.*, 2010), a molecule integral to the vertebrate inflammatory response. Ticks are highly sensitive to histamine during the early stages of feeding, but the sensitivity diminishes towards the end of feeding (Kemp and Bourne, 1980) and histamine becomes required for rapid engorgement presumably by facilitating vascular permeability and blood flow (Dai *et al.*, 2010). Contrary to the increasing expression of *tHRF* in *I. scapularis* as feeding progressed (Dai *et al.*, 2010), a considerable overexpression of *tHRF* was observed in unfed *R. zambeziensis* ticks. Knowing the detrimental effect of histamine on feeding site establishment (Kemp and Bourne, 1980), it is likely that tHRF has a different function prior to feeding, in the free-living stages of *R. zambeziensis*. Histamine release factor proteins are highly conserved and found in a number of eukaryotes and, apart from involvement in inflammation by histamine release, have functions in early developmental processes and cell survival by anti-apoptotic activity in response to stressors such as heat shock and oxidative stress (reviewed in Bommer, 2012). Desiccation, due to low environmental humidity and high temperatures, is one of the main constraints of free-living ticks surviving the many months between feeding stages, resulting in ticks requiring adaptations to survive in certain climatic regions (Randolph, 2004). *Rhipicephalus zambeziensis* naturally occurs in hot, dry regions with low humidity in southern Africa (Walker *et al.*, 1981; Norval *et al.*, 1982) and has shown high tolerance for extreme experimental abiotic conditions, including low humidity and high temperature (Madder *et al.*, 2005). Additionally, the marked wet and dry climatic seasons in southern Africa results in an extended period of inactivity of *R. zambeziensis* where only few of the life stages are found on hosts in the long, dry season (Norval *et al.*, 1982). Yet, the tick does not undergo behavioural diapause (Berkvens *et al.*, 1995) and it is unclear how it survives these long periods. It is conceivable that the overexpression of *tHRF* in unfed *R. zambeziensis* might contribute to the survival of the tick during free-living and questing periods, potentially by means of increased tolerance to stressors such as water or moisture deficiencies, an association that remains to be experimentally verified.

Another function of tHRF was previously observed during pathogen transmission, where the *I. scapularis tHRF* gene was up-regulated during *Borrelia burgdorferi*-infection. Silencing of the gene by RNA interference or immunisation with recombinant tHRF resulted in reduced transmission rates to the vertebrate host (Dai *et al.*, 2010), indicating that tHRF functions as a saliva-assisted transmission (SAT) molecule (Nuttall and Labuda, 2008). *Rhipicephalus zambeziensis* and *R. appendiculatus* are vectors of the protozoan parasite, *T. parva* (Lawrence *et al.*, 1983; Norval *et al.*, 1992), and it has experimentally been shown that *R. zambeziensis* has better vector competence for *T. parva* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998). Integral to the life cycle of *T. parva* is the tick's salivary gland environment where the pathogen multiplies and awaits transmission to the vertebrate host (Bishop *et al.*, 2004). Important vector-pathogen interactions may be expected to occur in these organs. Indeed, the overall level of *tHRF* expression in the salivary glands of *R. zambeziensis* (TPM value of 6057) was about five-times more than the expression seen for the *R. appendiculatus tHRF* gene (TPM value of 1212, Chapter 2). As tHRF has been shown to function as a SAT molecule enhancing pathogen transmission, the variation in expression between the *tHRFs* of *R. zambeziensis* and *R. appendiculatus* might, at least to some degree, explain the variation observed in the vector competence of *T. parva* of the two tick species.

3.6 Conclusions

In this chapter the first *de novo* transcriptome assembly of *R. zambeziensis* has been reported. The deposition of 13 584 annotated predicted proteins into GenBank vastly improves the sequence availability of the tick and will assist future studies in this largely neglected tick species. Similar to what was found in Chapter 2, an abundance of secretory protein transcription was seen in the salivary glands, the organs in closest proximity to the host that are actively expressing proteins to orchestrate host immune modulation. A large number of these secretory transcripts showed significant differential expression resulting in intricate expression profiles that changed considerably over feeding stages. Dynamic secretory protein transcription is possibly a result of antigenic variation, a mechanism previously proposed by which exposed tick antigens are 'switched' while evading host immune detection. Interchangeable expression of multi-

genic secretory proteins will have serious implications for recombinant vaccine development. Tick sialotranscriptomes, such as *R. zambeziensis*, will assist recombinant vaccine development by having the sequences of all members of a multi-gene family available to design vaccines against potentially shared antigens. Furthermore, a large number of transcripts for which no ORFs could be predicted were differentially expressed in the *R. zambeziensis* transcriptome, suggesting a role for lncRNAs in tick blood feeding and specifically, transcriptional regulation of secretory proteins, and motivates further investigation of this potentially important RNA molecule in ticks. Additionally, the transcriptome will enhance the understanding of the biology of *R. zambeziensis*, which is of major importance for the control of CD in southern Africa, as the tick is an efficient vector of *T. parva*, well adapted to extreme environments and, with the intensification of climate change, predicted to expand its natural distribution in future.

3.7 Authors' contributions

Conception and design of the study: MHD and BJM. Supply of funding: BJM. Tick colony maintenance, tick feeding and dissections: DD, RP and BJM. Execution of experimental work, data acquisition, bioinformatic analyses, data interpretation and manuscript drafting: MHD. Revision of the manuscript: MHD DD RP DJGR BJM. All authors read and approved the final version of the manuscript.

3.8 Acknowledgements

Dr. Vinet Coetzee of the Department of Genetics, University of Pretoria, South Africa is thanked for suggestions and critical review of the manuscript. MHD was supported by a National Research Foundation/ Department of Science and Technology - Professional Development Program (NRF/DST-PDP) scholarship. BJM was funded by the NRF Incentive Funding for Rated Researchers (NRF-Mans) and the Economic Competitive Support Programme (30/01/V010).

CHAPTER 4

Comparison between the expressed gene catalogues of *Rhipicephalus appendiculatus* and *R. zambeziensis* salivary glands reveals putative transcriptomic signatures of the variance in vector competence of the tick species

Minique H. de Castro [1,2,3], D. Jasper G. Rees [1,2] and Ben J. Mans [1,3,4]

[1] *College of Agriculture and Environmental Sciences, University of South Africa, Johannesburg, South Africa*

[2] *Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa*

[3] *Epidemiology, Parasites and Vectors, Onderstepoort Veterinary Research, Agricultural Research Council, Onderstepoort, South Africa*

[4] *Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa*

4.1 Abstract

Ticks require large blood meals, which means they feed for extended periods of time, resulting in ticks being exposed to a number of microorganisms. Like other arthropods, ticks rely on a basic innate immune system to combat these microorganisms. *Rhipicephalus appendiculatus* and *R. zambeziensis* are known vectors of *Theileria parva* and have shown varying vector competence for the parasite. To search for transcriptomic correlates of the variance in vector competence, the assembled sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* (from Chapters 2 and 3) were compared. Phylogenetic inference using 76 single copy orthologues indicated that the two ticks are very closely related but distinct species. Even though the species were shown to be genetically highly similar, significant differences were observed in their expression levels, specifically of genes involved in tick immunity or pathogen transmission. Moreover, bioinformatic signatures were observed that suggested that *R. appendiculatus* has a stronger immune system than *R. zambeziensis*, that genes used by pathogens for transmission and infection are highly expressed in *R. zambeziensis* and that the salivary glands of *R. zambeziensis* might be more structurally pliable for expansion during *T. parva* multiplication. Predicted long non-coding RNAs (lncRNAs) were identified in both species and their expression levels compared. *Rhipicephalus appendiculatus* showed elevated expression of putative lncRNA orthologues between the species, suggesting that some of the immune responses in *R. appendiculatus* might be governed by lncRNAs. Additionally, a reference datasets of 138 SCOs were compiled that can be used by tick biologists in future phylogenetic reconstructions.

Keywords: *Rhipicephalus appendiculatus*; *Rhipicephalus zambeziensis*; Tick salivary glands; Comparative sialotranscriptomics; Tick species phylogeny; Vector competence; Long non-coding RNA.

4.2 Introduction

In the previous chapters, the salivary gland transcriptomes of *Rhipicephalus appendiculatus* and *R. zambeziensis*, two vectors of Corridor disease, were assembled and annotated. Previous molecular phylogenetic analyses classified *R. appendiculatus* and *R. zambeziensis* as two separate species (Mtambo *et al.*, 2007a). Even so, the species have shown to be highly similar morphologically (Walker *et al.*, 1981), albeit

with some differences observed in developmental rates (Walker *et al.*, 1981), egg laying capacity (Zivkovic *et al.*, 1986) and adaptation to extreme environmental conditions (Madder *et al.*, 2005). Furthermore, differences in vector competence of *Theileria parva*, the causative agent of Corridor disease, have also been observed (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998). With the prediction of changes in climatic conditions and suitable geographic regions for these ticks (Olwoch *et al.*, 2008) and to improve future control measures of Corridor disease in South Africa, it is of importance to have a clear understanding of the differences between the species. The aim of this chapter was then to compare the sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* to find potential molecular explanations for the morphological and behavioural differences between the species. During the progression of the study, much of the focus was directed towards the differences in vector competence for *T. parva* of the ticks.

Vector competence is measured as the capability of a tick to successfully transmit a pathogen. It relies on the ability of the pathogen to infect the tick, survive and multiply in its tissues and get transmitted to the host (Lane, 1994; Beerntsen *et al.*, 2000; de la Fuente *et al.*, 2017). Ticks, like other arthropods, rely on a basic innate immune system to combat pathogen invasion. The immune response broadly consists of the self/ non-self recognition of infecting microorganisms or pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs), triggering signalling cascades that leads to pathogen opsonisation and elimination by phagocytosis, secretion of a number of antimicrobial peptides (AMPs), activation of apoptosis to remove infected cells and initiation of effector responses and proteins by signals such as reactive oxygen species (reviewed in Taylor, 2006; Kopáček *et al.*, 2010; Hajdušek *et al.*, 2013; Baxter *et al.*, 2017; de la Fuente *et al.*, 2017). These innate immune responses are most commonly observed in the tick midgut and haemolymph, but expression or activity of PRRs and AMPs have also been shown in salivary glands (e.g. Kamwendo *et al.*, 1993; Kocan *et al.*, 2008).

Salivary glands are integral to the life cycle of pathogens as many pathogens develop in the salivary glands before being transferred to the host through tick salivation (reviewed in Kazimírová and Štibrániová, 2013; Liu and Bonnet, 2014). During *T. parva* infection of the tick vector, the parasite infects the gut epithelial cells and then

migrates through the haemolymph to the salivary glands. The parasite awaits the blood meal of the ensuing tick instar for multiplication in the salivary glands followed by transmission to the vertebrate host via tick saliva (reviewed in Bishop *et al.*, 2004). Furthermore, salivary glands express proteins that are secreted into the vertebrate host for host immunomodulation and evasion (Francischetti *et al.*, 2009). These proteins can be hijacked by pathogens to promote their transmission and survival in the host, known as saliva-assisted transmission (SAT) (Nuttall and Labuda, 2008). This has been demonstrated before where crude extracts of *R. appendiculatus* salivary glands enhanced the transmission of *T. parva* (Shaw *et al.*, 1993) and Thogoto virus (Nuttall and Jones, 1991) to the vertebrate host. A few of the secretory proteins involved in SAT have been successfully characterised in ticks (Ramamoorthi *et al.*, 2005; Dai *et al.*, 2010), but many still remain unknown.

Considering the importance of tick salivary glands in the development and multiplication of *T. parva* and the key role tick secretory proteins play in the transmission of pathogens, it was hypothesised that a comparison between the salivary gland transcriptomes of *R. appendiculatus* (de Castro *et al.*, 2016, Chapter 2) and *R. zambeziensis* (de Castro *et al.*, 2017, Chapter 3) might yield correlates of their variance in vector competence of *T. parva*. Knowing the complexities involved in *de novo* transcriptome assembly (Wang *et al.*, 2009), the first thing that had to be determined was whether a comparative analysis would be suitable between the assembled transcriptomes. Two control experiments were performed in this regard. The first was the reassessment of the genetic relationship between *R. appendiculatus* and *R. zambeziensis* using the assembled transcriptomes. This enabled a multi-gene phylogenetic inference of the species compared to only two regions, internal transcribed spacer 2 (ITS2) and mitochondrial 12S ribosomal RNA (rRNA), used in the previously phylogenetic analysis (Mtambo *et al.*, 2007a). The second control experiment was the assessment of the quality of the assembled transcriptomes. Even though a number of transcriptome quality assessments have been performed in the previous chapters, it was of importance to evaluate whether the transcriptomes had been sampled at similar depths to support comparative observations. In the absence of similar sampling depths, it would not have been possible to directly compare the transcriptomes.

4.3 Methods

4.3.1 Tick, mite and horseshoe crab protein datasets

Seventeen transcriptome and genome datasets of hard (Ixodidae) and soft (Argasidae) tick species were downloaded from NCBI (Table 4.1) and used in the molecular phylogeny. The genomes of *Metaseiulus occidentalis* (Parasitiformes mite), *Tetranychus urticae* (Acariformes mite) and the Atlantic horseshoe crab, *Limulus polyphemus* from the order Xiphosurida, were also downloaded. For datasets where only transcript nucleotide sequences were available (i.e. the predicted proteins were not publically available), the proteins were predicted following an in-house protocol. This entailed: downloading the transcripts from the Transcriptome Shotgun Assembly (TSA) database; predicting open reading frames (ORFs) larger than 240 nucleotides containing a start and stop codon using the orffinder.pl script (github.com/vikas0633/perl); translating ORF sequences to amino acids; and removing redundant proteins based on 100% identity using CD-HIT v4.5.4 (Li and Godzik, 2006). The quality of each dataset was evaluated for presence and completeness of a core set of proteins expected to be found in arthropods using the Benchmarking Universal Single-Copy Orthologs v3 (BUSCO) software (Simão *et al.*, 2015).

4.3.2 Determining tick single copy orthologues

Tick orthologous protein clusters were determined from a core set of nine tick species representing both hard and soft ticks using the OrthoMCL v2.0.9 algorithm (Li *et al.*, 2003). The core set contained proteins of eight hard tick species, including three *Rhipicephalus*, three *Amblyomma* and two *Ixodes* species, and the predicted proteins (using our in-house prediction protocol) of the *Ornithodoros rostratus* transcriptome (Table 4.1). All the sequences were combined in a single dataset for all-versus-all BLASTp analysis, filtered (based on E-value < E-05 and percent match length > 50%) and orthologous groups determined by the Markov Cluster (MCL) algorithm (Van Dongen, 2000), with an inflation value of 1.5. Sequences of the single copy orthologous (SCO) groups, containing a single protein from each species, were extracted using the ExtractSCOs.sh and ExtractSeq.sh scripts (github.com/halexand/Ehux_HD/blob/master/orthoMCL_output/). The SCO protein representatives of *I. scapularis* were extracted and used as a reference SCO set (Appendix D: Table S1) for downstream analyses.

Table 4.1 Publically available tick, mite and horseshoe crab transcriptomic and genomic datasets used for selecting single copy orthologues and species phylogeny.

Species	Sampling tissue	Number of proteins[a]	TSA accession[a]	NCBI Bioproject/ WGS accession	Reference[b]
Ixodidae: Metastriata					
Rhipicephalus appendiculatus[c]	Salivary glands	12 761		PRJNA297811	(de Castro *et al.*, 2016, Chapter 2)
R. zambeziensis[c]	Salivary glands	13 583		PRJNA381085	(de Castro *et al.*, 2017, Chapter 3)
R. pulchellus[c]	Salivary glands	11 227		PRJNA170743	(Tan *et al.*, 2015a)
R. sanguineus	Larvae	34 944*	HACP01, HACW01	PRJEB8914	(De Marco *et al.*, 2017)
R. microplus	Gene's organ	34 028*	GELJ01	PRJNA288687	(Tidwell *et al.*, unpublished)
R. annulatus	Salivary glands	17 536*	GBJT01	PRJNA255770	(de la Fuente *et al.*, unpublished)
Hyalomma excavatum	Salivary glands	5337		PRJNA311286	(Ribeiro *et al.*, 2017)
Amblyomma cajennense[c]	Salivary glands	5770		PRJNA241272	(Garcia *et al.*, 2014)
A. maculatum[c]	Salivary glands	4849		PRJNA72241	(Karim *et al.*, 2011)
A. triste[c]	Salivary glands	8098		PRJNA241269	(Garcia *et al.*, 2014)
A. aureolatum	Salivary glands & Gut	7999		PRJNA344771	(Martins *et al.*, 2017)
Ixodidae: Prostriata					
Ixodes scapularis[c]	Genome	20 486		ABJB000000000	(Gulia-Nuss *et al.*, 2016)
I. ricinus[c]	Salivary glands & Gut	16 002		PRJNA217984	(Schwarz *et al.*, 2014b; Kotsyfakis *et al.*, 2015b)
I. persulcatus	Whole body	35 300*	GBXQ01	PRJNA263101	(Zhang *et al.*, unpublished)
Argasidae					
Ornithodoros rostratus[c]	Salivary glands & Gut	16 299*	GCJJ01	PRJNA270484	(Araujo *et al.*, unpublished)
O. moubata	Gut	8493		PRJNA377416	(Oleaga *et al.*, 2017)
O. turicata	Synganglia	52 817*	GDIC01, GDIE01	PRJNA281459	(Egekwu *et al.*, 2016)
Mite: Parasitiformes					
Metaseiulus occidentalis	Genome	11 982		AFFJ00000000	(Hoy *et al.*, 2016)
Mite: Acariformes					
Tetranychus urticae	Genome	15 054		CAEY00000000	(Grbić *et al.*, 2011)
Atlantic horseshoe crab: Xiphosurida					
Limulus polyphemus	Genome	23 287		AZTN00000000	(Simpson *et al.*, 2017)

[a] Number of proteins downloaded from the Bioproject accession or predicted from the transcripts downloaded from the Transcriptome Shotgun Assembly (TSA) project. Protein prediction strategy can be found in the Methods section. Species for which proteins were predicted are indicated by *.

[b] References indicated as 'unpublished' refer to datasets that have been released without associated publications.

[c] The nine datasets used to determine single copy orthologous (SCO) tick groups using OrthoMCL (Li *et al.*, 2003).

4.3.3 Molecular phylogenetic analysis

To expand the SCO dataset from the core set of nine tick species (used in 4.3.2) and include sequences of additional species in the phylogenetic analysis, the set of reference *I. scapularis* SCO proteins was used to generate a BLASTp database. Representative SCO protein sequences of the additional species were obtained by reciprocally searching the proteins of each species against the reference SCO BLASTp database (E-value < E-05) using the Conditional Reciprocal Best BLAST (CRB-BLAST) v0.6.9 program (Aubry *et al.*, 2014). The results were manually filtered to retain SCOs where representatives of at most two species were missing and where, at the same time, more than 80% of the SCOs were present in each of the species (that passed the filtering process). Based on these criteria, from an initial set of 32 tick datasets downloaded from NCBI, only 17 passed filter and could be used in the species phylogeny (Table 4.1). Multiple sequence alignment was performed per SCO group using MUSCLE v3.8.31 (Edgar, 2004) and the alignments trimmed with trimAl v1.2 (Capella-Gutiérrez *et al.*, 2009) using the -automated1 parameter (optimised for maximum likelihood phylogenies). ModelFinder (Kalyaanamoorthy *et al.*, 2017), implemented in IQ-TREE v1.5.4 (Nguyen *et al.*, 2014), was used to estimate the optimal evolutionary protein model for each SCO group alignment, based on the Bayesian information criterion (BIC). Following, alignments were concatenated into a single super alignment using FASconCAT v1.0 (Kück and Meusemann, 2010), parameters -i -p -p -s. Then, IQ-TREE was used to infer the maximum likelihood phylogeny from the concatenated alignment using the edge-linked proportional partition (Chernomor *et al.*, 2016) and the optimal protein models predicted for each SCO group. Branch support was obtained by the ultrafast bootstrap (UFBoot) method (Minh *et al.*, 2013) with 1000 replicates. The

phylogenetic tree was visualised in FigTree v1.4.3 (tree.bio.ed.ac.uk/software/figtree/) and the outgroup set as *L. polyphemus*.

4.3.4 Comparative analyses between *R. appendiculatus* and *R. zambeziensis*

The previously assembled salivary gland transcriptomes of female and male *R. appendiculatus* (de Castro *et al.*, 2016, Chapter 2) and *R. zambeziensis* (de Castro *et al.*, 2017, Chapter 3) ticks were used for the comparative analyses. As reported in Chapters 2 and 3, the transcriptomes were assembled from dissected salivary gland samples of unfed, early feeding and late feeding ticks and represented annotated secretory protein families and estimations of expression levels [estimated through the metric transcripts per million (TPM) (Wagner *et al.*, 2012)]. These datasets will be further analysed in this chapter and will henceforth be referred to simply as the *R. appendiculatus* and *R. zambeziensis* transcriptomes. To detect differences between the secretory protein family structures (number of family members and average expression levels) of the two tick species, Chi-square comparisons, with Bonferroni corrections, were performed. Additionally, Pfam database (Finn *et al.*, 2016) searches were performed on each set of proteins to identify significant differences between the species. Furthermore, protein sequences of tick immunity-related genes and tick genes altered by pathogen transmission were mined from three review articles (Kopáček *et al.*, 2010; Hajdušek *et al.*, 2013; Liu and Bonnet, 2014). Protein sequences of *R. appendiculatus* and *R. zambeziensis* were BLASTp (E-value < E-20) searched against these target proteins to identify putative orthologous tick immunity or pathogen transmission proteins in each tick species. The expression levels of the orthologues were subsequently compared to identify significant differences between the species using the Chi-square test with Bonferroni correction.

4.3.5 Long non-coding RNA annotation and analysis

During the assembly process of *R. appendiculatus* and *R. zambeziensis*, a large number of transcripts were annotated as No ORF transcripts (transcripts for which ORFs could not be confidently predicted, refer to Chapters 2 and 3). To differentiate putative lncRNA transcripts from coding transcripts missed during the initial ORF prediction, the transcriptomes of *R. appendiculatus* and *R. zambeziensis* were filtered through a lncRNA prediction pipeline (based on Hezroni *et al.*, 2015; Ulitsky, 2016). Based on the

pipeline, transcripts were discarded on the following criteria: 1) length shorter than 200 bp; 2) containing an ORF longer than 100 amino acids as predicted in all frames from any start codon using orffinder.pl; 3) retrieving a significant BLASTx result (E-value < E-05) from any of the following databases: NCBI non-redundant (Nr) database, UniProt Knowledgebase (UniProtKB/TrEMBL), predicted peptides from the *I. scapularis* genome (IscaW1.4) (Gulia-Nuss *et al.*, 2016), AcariDB database (containing mite and tick sequences, as described in Chapter 2) or the EuKaryotic Orthologous Groups (KOG) dataset (Tatusov *et al.*, 2003); or 4) predicted as protein-coding by two or more of the following software: Predictor of lncRNAs and mRNAs based on k-mer scheme v1.2 (PLEK) (Li *et al.*, 2014), Coding Potential Calculator (CPC) (Kong *et al.*, 2007) and Coding-Potential Assessment Tool (CPAT) (Wang *et al.*, 2013). The transcripts that were retained after the filtering process were annotated as putative lncRNAs. The putative lncRNAs of *R. appendiculatus* and *R. zambeziensis* were reciprocally BLASTn searched against each other to obtain reciprocal best BLAST hits (RBBH, E-value < E-20, alignment coverage > 50%). The expression levels of the RBBH lncRNAs were compared between the species by Chi-square test and Bonferroni correction.

4.4 Results

4.4.1 Generating a tick single copy orthologue reference dataset

Single copy orthologous sequences, present in a diverse range of ticks, are required to infer molecular species phylogenies on a number of evolutionally divergent tick species. Towards this goal, OrthoMCL (Li *et al.*, 2003) was used to predict orthologues from a core set of nine tick species, including species from the major hard tick clades, *Rhipicephalus*, *Amblyomma* and *Ixodes*, and a representative of soft ticks, *O. rostratus* (Table 4.1). The datasets had to be publically available prior to the start of the analysis and the predicted protein sequences had to be available for download, to be included in the core prediction set. *Ornithodoros rostratus* was the only Argasidae species with any next generation transcriptomic sequence data available at that time, even though protein sequences were not available. As it was necessary to include a soft tick species in the core prediction set, the proteins were predicted in-house, resulting in 16 299 putative protein sequences for *O. rostratus*. The core prediction dataset, consisting of a total of 109 075 protein sequences, were clustered by OrthoMCL and 138 groups were found

that contained a single sequence from each of the nine species, the SCO groups. The 138 tick SCOs mainly consisted of housekeeping and uncharacterised proteins and the *I. scapularis* representative sequence of each SCO was extracted and used in a reference database for downstream analyses (Appendix D: Table S1).

4.4.2 Selecting tick databases to be included in the species phylogeny

By means of reciprocal best BLAST, the SCO reference sequence database was used as a toolkit to include additional tick species into the molecular phylogeny dataset. Additional to the initial nine species, 23 tick transcriptome or genome datasets were downloaded from NCBI. A trade-off was desired between the maximum number of SCOs to base the phylogeny on, the maximum number of tick species to include in the phylogeny and the minimum number of missing data points. Datasets that were either too small, contained too few SCO sequences or were duplicates of species already included in the dataset, were discarded. In the end, this analysis resulted in the selection of a final set of 76 SCOs based on allowing missing data for at most two species per SCO group (Appendix D: Table S2) and 17 tick and three outgroup species, which each contained > 80% of the selected SCOs (Table 4.1). Of the selected tick species, protein sequences were not publically available for six of the species and these were predicted in-house (labelled by * in Table 4.1). The quality of the transcriptome or genome of each of the 20 species to be included in the phylogeny was determined by BUSCO assessment of core proteins expected to be present in all arthropod species (Figure 4.1). As would be expected, the four genomes showed the highest number of complete and the least number of missing core proteins. The quality of the transcriptomes varied considerably, from 40 - 90% for the level of completeness and 53 - 8% for the percentage of missing proteins for respectively, *R. annulatus* and *R. appendiculatus*.

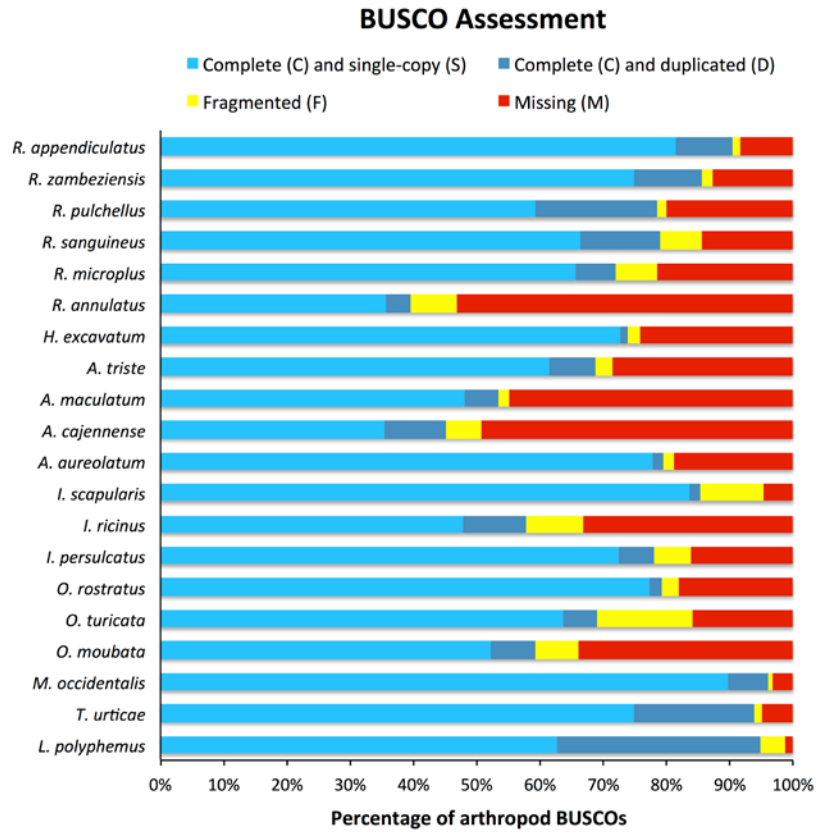


Figure 4.1 Quality assessment of the transcriptomic and genomic datasets used in the molecular phylogeny. The presence and completeness of a core set of 1066 arthropod proteins were evaluated in the datasets using the Benchmarking Universal Single-Copy Orthologs (BUSCO) software. Colour key is indicated at the top.

4.4.3 Species phylogeny

Multiple sequence alignments and evolutionary model predictions were performed within each of the 76 selected SCOs. The concatenated, partitioned alignment, totalling 23 659 amino acids in length with 9194 informative sites (39%), was used for maximum likelihood phylogenetic inference by IQ-TREE (details of the alignment can be found in Appendix D: Table S2). Per species gaps ranged from 0.1 - 16% (average 7%) and per species missing data, from 0 - 20% (average 6%). The inferred phylogeny clustered the Ixodidae and Argasidae tick families into distinct monophyletic clades with 100% bootstrap support (Figure 4.2). In Ixodidae, the Metastriata and Prostriata ticks grouped separately and all species of each genus grouped together in unique monophyletic clades with 100% bootstrap support. The Atlantic horseshoe crab, *L. polyphemus*, was set as outgroup and the Parasitiformes mite, *M. occidentalis*, clustered closer to ticks than *T. urticae* (Acariformes).

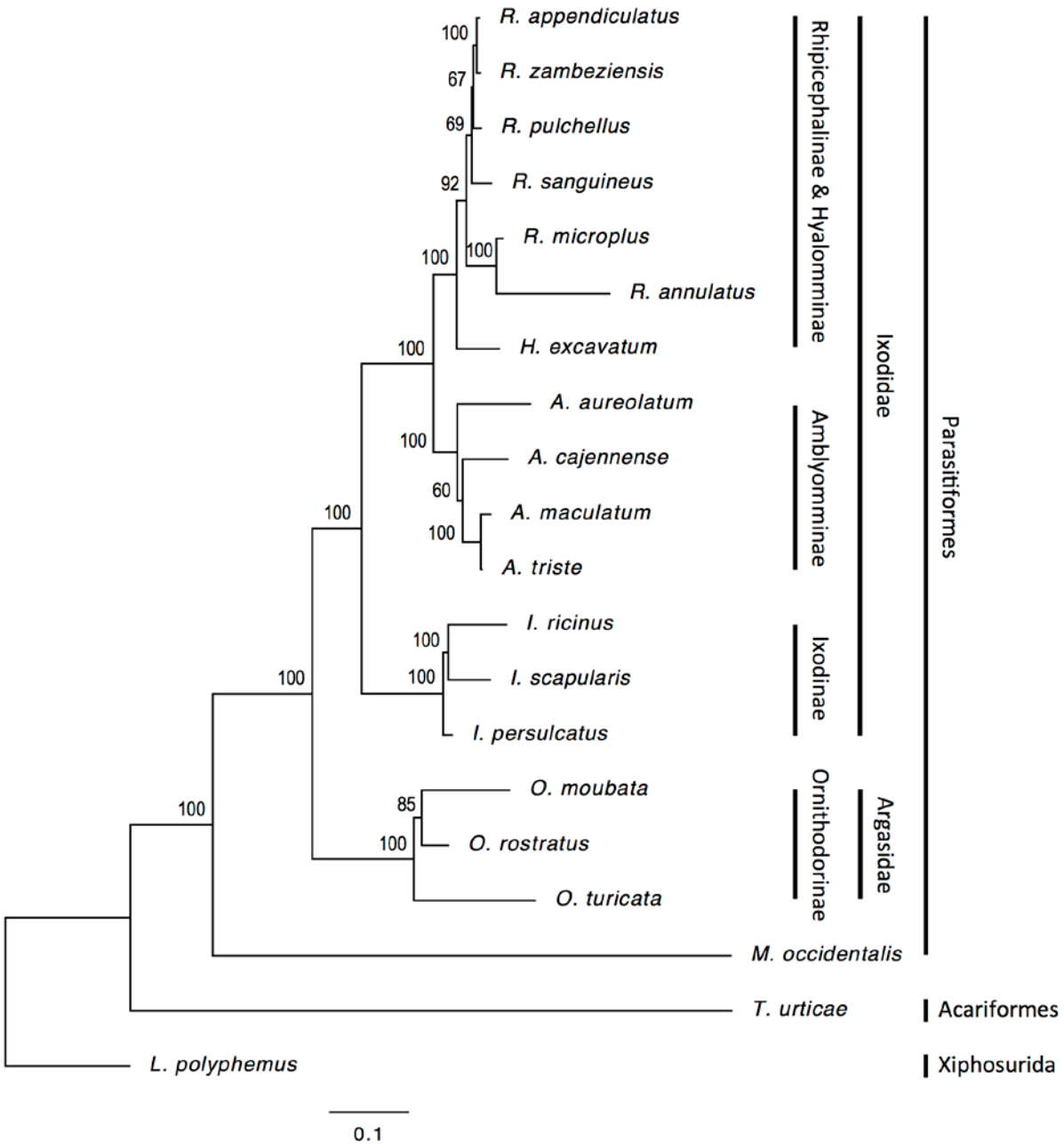


Figure 4.2 Maximum likelihood species phylogeny inferred from 76 single copy orthologues (SCOs). The SCOs were aligned separately using optimised evolutionary models and IQ-TREE was used to infer the phylogeny from the concatenated alignment. Between 18 and 20 species were aligned per SCO group. The ultrafast bootstrap method with 1000 replicates was used for branch support and values ≥ 60% are indicated. *Limulus polyphemus* was set as outgroup. Orders, families and genera are indicated.

4.4.4 Assessment of the quality of the *R. appendiculatus* and *R. zambeziensis* transcriptomes

To evaluate whether comparative analyses between the transcriptomes of *R. appendiculatus* and *R. zambeziensis* were feasible, quality assessment of the transcriptomes were performed. *Rhipicephalus appendiculatus* and *R. zambeziensis* were the two best performing transcriptomes based on BUSCO metrics (Figure 4.1). Of the 1066 core genes, 964 were both present and complete in *R. appendiculatus* (completeness level of 90%). In *R. zambeziensis*, 912 of the core genes were present and complete (86% complete). Only 8% and 13% of the genes were missing from respectively, *R. appendiculatus* and *R. zambeziensis*. The two datasets are transcriptomic and it can be expected that some of the core genes might not be expressed in tick salivary gland tissues. This was evident from the high levels of completeness observed in genomic datasets that theoretically will contain all the core genes: *M. occidentalis* (96%), *L. polyphemus* (95%), *T. urticae* (94%), and *I. scapularis* (85%). The high level of completeness and the similarity in the number of core genes obtained between *R. appendiculatus* and *R. zambeziensis*, suggested that no genes were missed due to insufficient sequence coverage or assembly procedures in either species and that the core genes that were reported missing could have merely been as a result of not being expressed in the salivary glands. These results indicated transcriptomes with similar quality measures and adequate sampling depth to allow downstream comparative analyses.

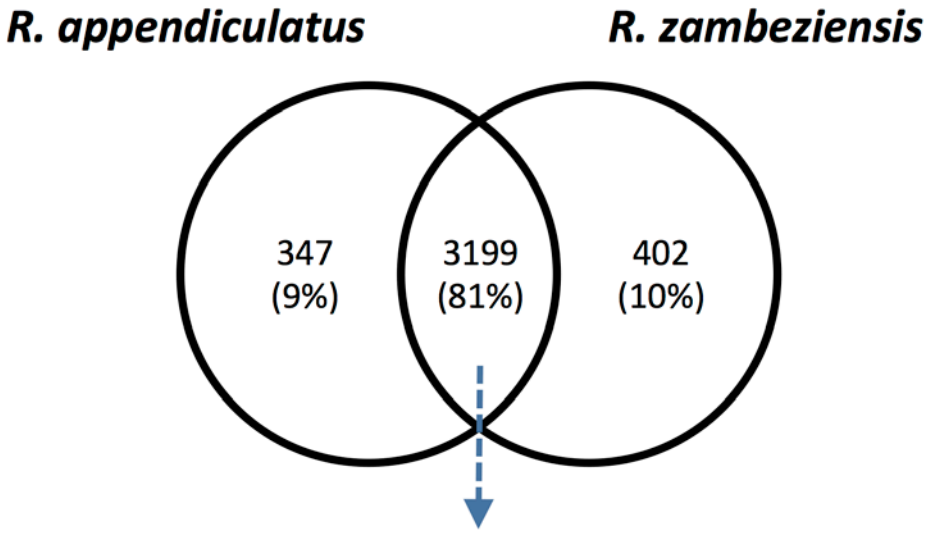
4.4.5 Assessment of the genetic relationship between *R. appendiculatus* and *R. zambeziensis*

From the inferred molecular phylogeny it was possible to see that *R. appendiculatus* and *R. zambeziensis* grouped as separate species with 100% bootstrap, but that they were genetically highly similar due to the short branch lengths of the tree. The pair-wise distance estimated between them was 0.0078, which was the smallest of any of the pair-wise species comparisons. Indeed, the percentage protein identity between *R. appendiculatus* and *R. zambeziensis* of the 76 SCO proteins from which the phylogeny was inferred was on average 99%. This comparison was extended to include all reciprocal best BLAST protein hits between the species. Based on stringent best BLAST hit thresholds of E-value < E-20 and alignment coverage > 75%, 5901 proteins were

shared in a 1:1 manner between the two species. These represented 46% and 43% of respectively, the *R. appendiculatus* and *R. zambeziensis* proteins, and showed on average 96% protein identity between the species. To confirm that the species are unique, the *cytochrome c oxidase subunit I (cox1)* gene of each species was BLASTn aligned against the NCBI Nt database (blast.ncbi.nlm.nih.gov/Blast.cgi). The most significant match in each case was the *R. appendiculatus* mitochondrial genome (KC503257.1) at sequence identities of 99.9% and 91% for respectively, *R. appendiculatus* and *R. zambeziensis*. No mitochondrial genome (or *cox1* gene) of *R. zambeziensis* was available on NCBI at the time of the analysis. The data indicated that *R. appendiculatus* and *R. zambeziensis* are closely related but distinct species.

4.4.6 Pfam domain analysis in *R. appendiculatus* and *R. zambeziensis*

The Pfam domain analysis revealed that 3948 domains were present at least once in *R. appendiculatus* or *R. zambeziensis* (the 100 most variable domains are tabulated in Appendix D: Table S3). Eighty-one percent (3199) of the Pfam domains were shared between the species and around 9 - 10% were unique in either species (Figure 4.3). Comparison of the Pfam domain occurrences between *R. appendiculatus* and *R. zambeziensis* showed three domains, the Spectrin repeat (PF00435), Fibronectin type III (PF00041) and Laminin EGF (PF00053) domains, with significant enrichment in either one of the two species (after a Bonferroni corrected p-value < 0.00003). The Fibronectin type III and Laminin EGF domains were more prevalent in *R. appendiculatus* and the Spectrin repeat domain, in *R. zambeziensis*. Another six significantly different Pfam domains were identified, but they formed part of larger domain families and when all family members were considered, did not remain significant.



Different domain occurrences	*Ra*	*Rz*	*χ2*	p-value
Spectrin repeat (PF00435)	100	195	30.59	< 0.0001
Fibronectin type III (PF00041)	141	76	19.47	< 0.0001
Laminin EGF (PF00053)	84	38	17.34	< 0.0001

Figure 4.3 Pfam domain comparison between *R. appendiculatus* and *R. zambeziensis*. The proportions of shared and unique domains are indicated together with the domains that showed significantly different occurrences between the species. Significance was estimated by Chi-square test with Bonferroni correction, p-value < 0.00003, *df* = 1. *Ra, R. appendiculatus; Rz, R. zambeziensis*.

4.4.7 Comparisons of the secretory protein family sizes and average expression levels between *R. appendiculatus* and *R. zambeziensis*

Comparing the number of members in each secretory protein family between the two tick species showed that the Ixodegrin B family (57 vs. 32 members, in respectively, *R. appendiculatus* and *R. zambeziensis*, p-value = 0.0003) and the Secretory - unknown function family (7 vs. 46, p-value < 0.0001) were significantly expanded in *R. appendiculatus* and *R. zambeziensis*, respectively. Additionally, large differences, albeit not significant after Bonferroni correction (p-value < 0.0007), were seen in the Reprolysin (133 vs. 213, p-value = 0.0061) and TIL (Trypsin Inhibitor-like) domain (108 vs. 177, p-value = 0.0072) families. Comparisons between the average transcripts per million (TPM) values of secretory protein families of *R. appendiculatus* and *R.*

zambeziensis revealed that 16 families had significantly different expression levels after a Bonferroni corrected p-value < 0.0007 (Table 4.2). Four families, the Histamine release factor (HRF), Secretory - unknown function, Madanin and Ixostatin families, had higher average expression levels in *R. zambeziensis* compared to *R. appendiculatus*. The remaining families with significantly different expression levels had higher average expression values in *R. appendiculatus.*

Table 4.2 Secretory protein families with significantly different average expression levels between *R. appendiculatus* and *R. zambeziensis* after Bonferroni corrected p-value < 0.0007 (*df* = 1). Average expression levels were estimated by transcripts per million (TPM).

Secretory protein family	***R. appendiculatus***	***R. zambeziensis***	***χ2***	**p-value**
Histamine release factor	1211.6	6057.3	3230.41	0
Secretory - unknown function	12.8	455.6	418.59	< 0.0001
Immunoglobulin G binding protein A	2051.8	989.9	370.75	< 0.0001
ML domain	1111.8	417.4	315.35	< 0.0001
Glycine rich superfamily	3072.0	1842.4	307.70	< 0.0001
Hirudin	310.2	108.7	96.87	< 0.0001
Madanin	23.8	153.5	94.89	< 0.0001
Defensin	464.4	223.1	84.70	< 0.0001
TIL domain	164.0	38.1	78.51	< 0.0001
Antigen 5	188.9	57.6	69.98	< 0.0001
Lipocalin	133.1	39.9	50.25	< 0.0001
Thyropin	67.6	22.7	22.35	< 0.0001
Cystatin	54.4	17.3	19.20	< 0.0001
Ixostatin superfamily	10.3	35.5	13.90	0.0002
Fibrinogen-related domain	52.6	20.8	13.70	0.0002
Carboxypeptidase inhibitor	42.9	16.0	12.27	0.0005

4.4.8 Putative tick immunity and pathogen transmission orthologues of *R. appendiculatus* and *R. zambeziensis* and their expression analysis

Proteins previously shown to have involvement in tick immunity or pathogen transmission were selected from literature (Kopáček *et al.*, 2010; Hajdušek *et al.*, 2013; Liu and Bonnet, 2014) and used as target sequences in a BLASTp database. Using these targets, nineteen putative orthologues were identified in each of *R. appendiculatus* and *R. zambeziensis* (Appendix D: Table S4). As the assignment of putative orthology was based solely on sequence similarity, a high protein identity limit of 70% was used to identify the proteins. All, except two *R. appendiculatus* proteins, were predicted to be full-length. The 19 proteins identified in each species were reciprocal best hits of each other when *R. appendiculatus* and *R. zambeziensis* were BLAST compared to each other, varying in percent identity from 97 - 100%. Then, by comparing the TPM values of the potential orthologues between *R. appendiculatus* and *R. zambeziensis*, seven showed significantly different expression levels after implementing a Bonferroni corrected p-value < 0.0026 (Table 4.3; Appendix D: Table S4). The spectrin α chain (fodrin), varisin, mitochondrial porin (T2) and factor D-like orthologues, had higher expression levels in *R. appendiculatus*, whereas the tHRF, glutathione S-transferase (GST, AAL99403.1) and subolesin orthologues showed elevated expression in *R. zambeziensis*.

4.4.9 Annotation and expression analysis of putative long non-coding RNAs in *R. appendiculatus* and *R. zambeziensis*

Putative lncRNA molecules were predicted from the transcripts for which no ORFs could be obtained during the transcriptome assemblies of *R. appendiculatus* and *R. zambeziensis* (refer to Chapters 2 and 3). Using the prediction pipeline (see Methods) 7414 *R. appendiculatus* No ORF transcripts were filtered to 3830 putative lncRNAs and 7947 *R. zambeziensis* No ORF transcripts, to 3980 putative lncRNAs (Figure 4.4). During the initial assembly processes (Chapters 2 and 3), ORFs were predicted and retained in the datasets only after implementing very stringent criteria, in order to have higher confidence that the predicted ORFs potentially represented proteins. Here, ORFs were predicted using less stringent criteria to eliminate any potentially coding transcripts from the lncRNA datasets, in order to have higher confidence that the remaining transcripts are of non-coding nature. To compare the lncRNAs of *R. appendiculatus* and

R. zambeziensis, reciprocal best BLAST analysis was performed that identified 439 putative lncRNAs that were shared between the species. Comparisons of the TPM values showed that 23 putatively shared lncRNAs had significantly different expression levels between the species after a Bonferroni corrected p-value < 0.0001 (Figure 4.4; Appendix D: Table S5). More lncRNA transcripts were expressed at higher levels in *R. appendiculatus*, where 18 putative lncRNAs showed higher expression, than in *R. zambeziensis*, where only five lncRNAs showed higher expression. Furthermore, *R. appendiculatus* showed significantly higher overall expression in the putatively shared lncRNAs, with a total TPM value of 9460 and average TPM value of 22. Equivalent TPM values for *R. zambeziensis* were 2895 and 7 (Figure 4.4).

Table 4.3 Significantly different expression levels between *R. appendiculatus* and *R. zambeziensis* of putative tick immunity and pathogen transmission orthologues. Expression level was measured by transcripts per million (TPM) and orthologues with significantly different expression levels (p-values < 0.0026) are indicated. Details of all orthologues and their expression comparisons can be found in Appendix D: Table S4.

Protein name (Accession nr)	***R. appendiculatus***	***R. zambeziensis***	**p-value**	**Reference[a]**
Tick histamine release factor (tHRF, AAY66972.1)	1211.6	6057.3	0	(Dai *et al.*, 2010)
Glutathione S-transferase (GST, AAL99403.1)	741.4	1124.0	< 0.0001	(Rosa de Lima *et al.*, 2002)
Subolesin (AAV67034.2)	16.5	67.7	< 0.0001	(de la Fuente *et al.*, 2006)
Spectrin α chain (fodrin, XP_002433506.1)	87.6	33.4	< 0.0001	(Ayllón *et al.*, 2013)
Varisin (AAO24323.1)	29.3	6.0	< 0.0001	(Kocan *et al.*, 2008)
Mitochondrial porin (T2, XP_002408065.1)	130.8	80.9	0.0006	(Ayllón *et al.*, 2013)
Factor D-like (AAN78224.1)	26.6	7.2	0.0008	(Simser *et al.*, 2004b)

[a] Reference implicating the protein in tick immunity or pathogen transmission.

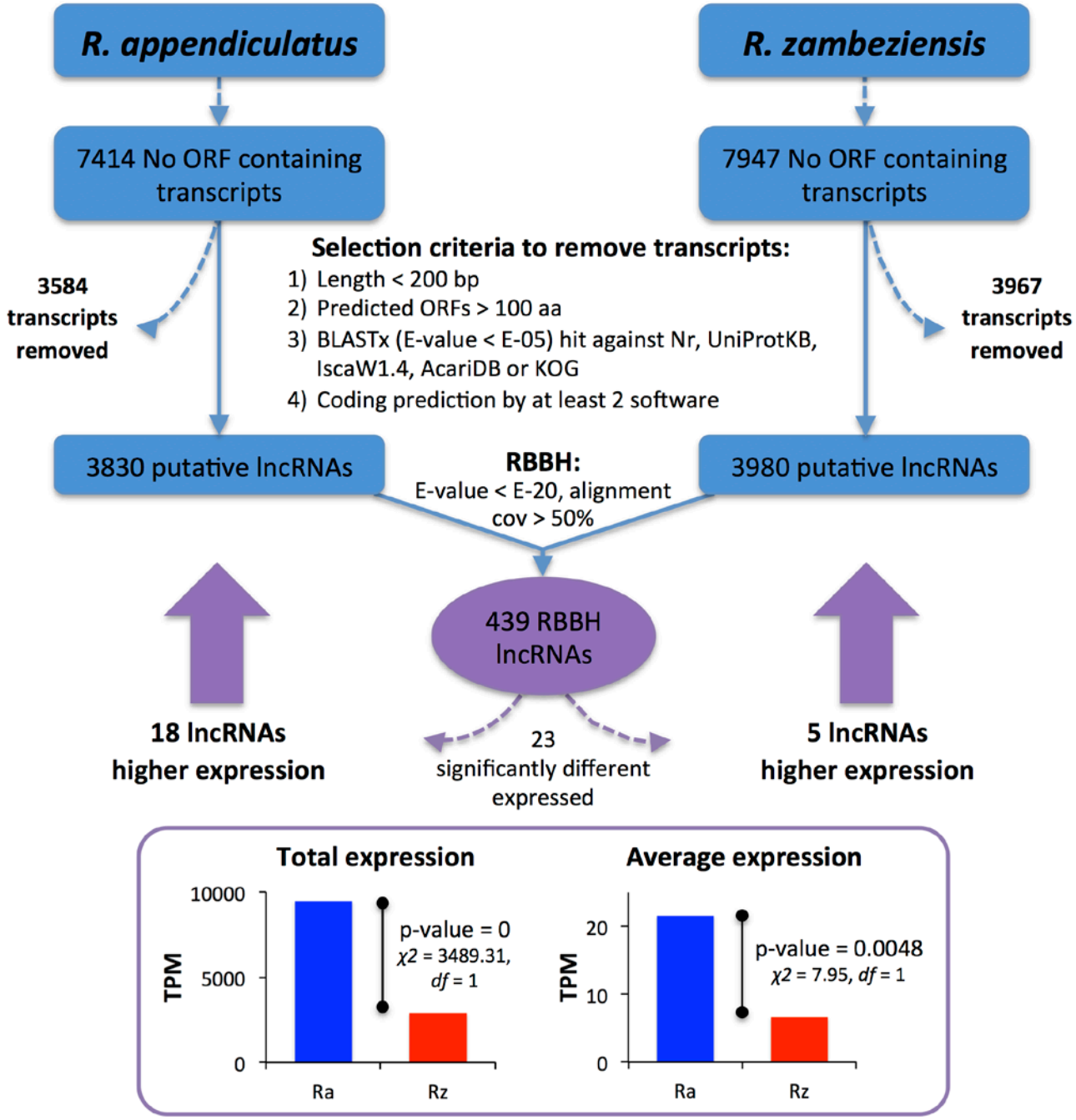


Figure 4.4 Identification and comparison of putative long non-coding RNAs in *R. appendiculatus* and *R. zambeziensis*. The pipeline used to identify putative lncRNAs is based on a process of elimination of potential coding transcripts, details of which are indicated in the blue section. Reciprocal best BLAST lncRNA matches were identified between *R. appendiculatus* and *R. zambeziensis* and used for expression analyses, as highlighted in purple (see Appendix D: Table S5 for details on significantly different lncRNAs). Significant differences between total and average expression of *Ra*, *R. appendiculatus* (blue) and *Rz*, *R. zambeziensis* (red) are also indicated. RBBH, reciprocal best BLAST hits; ORF, open reading frame; TPM, transcript per million.

4.5 Discussion

In this chapter, the recently assembled transcriptomes of *R. appendiculatus* (de Castro *et al.*, 2016, Chapter 2) and *R. zambeziensis* (de Castro *et al.*, 2017, Chapter 3) were compared in an attempt to find genetic differences that might explain the behavioural and morphological differences previously observed between the two species (Walker *et al.*, 1981; Zivkovic *et al.*, 1986; Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998; Madder *et al.*, 2005). Based on quality assessment, we were confident that the transcriptomes were sampled at the same depth to assure that genes were not excluded from either dataset due to sequence coverage or assembly technicalities, making direct comparisons between the species possible. Previous phylogenetic reconstruction based on ITS2 and 12S rRNA classified *R. appendiculatus* and *R. zambeziensis* as two separate species (Mtambo *et al.*, 2007a). To investigate whether this relationship is maintained with the inclusion of additional genes, a species phylogeny was reconstructed from genes gathered from available tick genomes and transcriptomes. The aim was to include as many available sialotranscriptomes as possible and for species without available sialotranscriptomes, transcriptomes of other tissues were included if available. Unfortunately, for some transcriptome publications, the predicted protein sequences, assembled transcripts or even raw sequence data were not made publically available and could therefore not be included in the analysis. The issue of unavailability of data and subsequent exclusion from protein databases and comparative analyses has been addressed before (Mans *et al.*, 2016). A total of 32 tick transcriptomic and genomic datasets were downloaded, which were reduced to 17 based on the size of the dataset, the number of SCOs present, and the prior availability of the species in the set (resulting in no species duplications in the final dataset). As it is known that paralogous sequences hinder phylogenetic reconstruction (Koonin, 2005) and that tick secretory protein families have experienced extensive gene duplications and lineage-specific expansions (Mans and Neitz, 2004; Mans *et al.*, 2008a; Francischetti *et al.*, 2009; Mans, 2011; Mans, 2016; Mans *et al.*, 2017), it was decided to infer the species phylogeny on sequences occurring only once in each species, the set of so called single copy orthologues (SCOs). By using reciprocal best BLAST analysis, this set of proteins was used to extend the number of species included in the phylogeny and propose its use in future as a toolbox for tick phylogenetic reconstruction.

The inferred tick species phylogeny was similar to previously observed evolutionary relationships among tick species based on rRNA or mitochondrial gene sequences (e.g. Klompen and Oliver, 1993; Murrell *et al.*, 2001; Estrada-Peña *et al.*, 2005; Barker and Murrell, 2008; Mans *et al.*, 2012; Burger *et al.*, 2014a; Burger *et al.*, 2014b; Abdullah *et al.*, 2016; Landulfo *et al.*, 2017). Ixodidae and Argasidae families clustered into distinct monophyletic clades. Within Argasidae, only the Ornithodorinae subfamily was represented. *Ornithodoros rostratus* and *O. moubata* have been classified as being part of *Ornithodoros* sensu stricto (s.s.) and clustered together (Klompen and Oliver, 1993; Burger *et al.*, 2014b). Within Ixodidae, the Metastriata and Prostriata formed distinct monophyletic clades and *Ixodes* (subfamily Ixodinae) is the only genus in the Prostriata group. Based on 28S rRNA sequences, Kovalev and Mukhacheva (2012) found that *I. scapularis* grouped basal to *I. ricinus* and *I. persulcatus*. In the current phylogeny based on 76 SCOs, *I. persulcatus* was shown to group basal to *I. scapularis* and *I. ricinus* with 100% bootstrap support, indicating that inferences based on a large number of genes may improve inferred relationships between tick species. Within the Metastriata group the three genera, *Rhipicephalus, Hyalomma* and *Amblyomma*, grouped into three distinct monophyletic clades with 100% bootstrap values. The structure of the *Amblyomma* species relationships, (*A. aureolatum*, (*A. cajennense*, (*A. maculatum*, *A. triste*))), was similarly previously observed for 16S rRNA sequences (Estrada-Peña *et al.*, 2005). The *Hyalomma* species, *H. excavatum*, formed its own monophyletic clade, sister to the *Rhipicephalus* monophyletic clade. This close relationship with the Rhipicephalinae has been observed before (Abdullah *et al.*, 2016) and it has even been suggested that Hyalomminae should be included in the Rhipicephalinae (Murrell *et al.*, 2001; Barker and Murrell, 2008). The relationships of the *Rhipicephalus* species used in the current phylogeny were similar to what has been shown before: that *R. appendiculatus, R. zambeziensis* and *R. pulchellus* formed a single clade with *R. appendiculatus* and *R. zambeziensis* being the closest, and that *R. microplus* and *R. annulatus* grouped together in a distinct clade with a 100% bootstrap value (Murrell *et al.*, 2001; Barker and Murrell, 2008). *Rhipicephalus microplus* and *R. annulatus,* species of the subgenus *Boophilus* (Murrell *et al.*, 2003), form part of the *R. microplus* species complex (Burger *et al.*, 2014a). In the current phylogeny, the branch length of *R. annulatus* was noticeably long, differing from the previous report by Burger *et al.* (2014a), and potentially could be ascribed to the quality of the *R. annulatus* dataset. Based on the BUSCO core gene analysis, only 40% of the genes were complete

and 53% of the genes were missing, resulting in 16% gaps in the *R. annulatus* sequences in the super alignment. A recent tick species phylogeny was inferred from 425 homologous gene families in 20 transcriptomic datasets (Landulfo *et al.*, 2017). The authors, however, did not implement a single copy gene strategy as was followed during the phylogenetic reconstruction in the crrent study. The strategy implemented in this study was to limit the inclusion of paralogous sequences with undefined evolutionary histories in the phylogeny as these could hinder phylogenetic inferences (Koonin, 2005). Nevertheless, the inferred species phylogeny in the current study was similar in structure to the phylogeny reconstructed by Landulfo *et al.* (2017) and signifies the future of tick taxonomy based on larger sets of genes.

Previous phylogenetic reconstructions classified *R. appendiculatus* and *R. zambeziensis* as two separate species (Mtambo *et al.*, 2007a). Based on the 100% bootstrap value separating the two species in the phylogeny constructed in the current study and the large difference in *cox1* gene sequences between the species, the results obtained in this study also motivate maintaining the separate species status of *R. appendiculatus* and *R. zambeziensis*. Even so, from the species phylogeny and similar to previous reports (Murrell *et al.*, 2001; Barker and Murrell, 2008), it was evident that *R. appendiculatus* and *R. zambeziensis* are genetically very similar due to their close grouping, very short branches and low pair-wise distance values. Almost half of each transcriptome could potentially be considered as orthologous to the other species, with very high average protein identities (> 95%) observed within these orthologues. Additionally, 81% of the Pfam domains are shared between the species and only two of the secretory protein families significantly differed in size between the two species. From these results it became apparent that the sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* were genetically highly similar, leaving the question unanswered about what is causing the differences observed between the two species. The answer to this question might be found in the difference in expression of certain genes in the salivary glands of the two species. Comparing the average expression levels of the secretory protein families, 16 were found to be significantly different between *R. appendiculatus* and *R. zambeziensis.* Interestingly, for a number of these an association with tick immunity or pathogen transmission was found. Therefore, it was decided to focus the search towards finding more of these kinds of genes and employed a BLAST database of protein sequences extracted from review articles in the field (Kopáček *et al.*,

2010; Hajdušek *et al.*, 2013; Liu and Bonnet, 2014). Nineteen putative tick immunity or pathogen transmission orthologues were observed in each species and seven of these showed significant expression variation between the species. It was therefore concluded that the largest differences between the sialotranscriptomes of *R. appendiculatus* and *R. zambeziensis* seemed to be between expression levels of putative tick immunity or pathogen transmission orthologues, which potentially might explain the differences in the observed vector competence of the ticks. Herewith follows a discussion of the genes that showed varying expression levels between the species and their previously shown involvement in tick immunity or pathogen transmission based on literature.

The tick immune system roughly consists of the identification of non-self molecules by PRRs, secretion of AMPs to combat infection and the activation of apoptosis to remove the pathogens and/ or infected cells (reviewed in Taylor, 2006; Kopáček *et al.*, 2010; Hajdušek *et al.*, 2013; Baxter *et al.*, 2017; de la Fuente *et al.*, 2017). Expression level comparisons of genes or protein families within this group revealed that *R. appendiculatus* had elevated expression levels for tick immunity-related genes when compared to *R. zambeziensis*. Firstly, these included potential PRRs, of which Fibrinogen-related domain and Myeloid Differentiation-2-related lipid-recognition (ML) domain families showed varying expression. Fibrinogen-related domain proteins are classified as lectins/ haemagglutinins and are thought to play key roles in self/ non-self recognition (Kovář *et al.*, 2000; Rego *et al.*, 2006). The inhibition of haemagglutinin resulted in increased *T. parva* infections in the salivary glands of *R. appendiculatus* (Kamwendo *et al.*, 1995) and it was observed that haemagglutinin activity increased in *T. parva* infected *R. appendiculatus* ticks, indicating a functional role for haemagglutinins in *R. appendiculatus* - *T. parva* interactions (Kibuka-Sebitosi, 2006). ML domain proteins were shown to interact with bacterial lipopolysaccharides through a pathogen recognising Toll-like receptor (Hyakushima *et al.*, 2004), suggesting a role in pathogen recognition. The infection of *I. ricinus* with *Borrelia burgdorferi* resulted in the up-regulation of an ML domain gene (Rudenko *et al.*, 2005) and a number of ML domain genes were overexpressed in the *I. ricinus* haemocyte transcriptome (Kotsyfakis *et al.*, 2015a). Secondly, significant expression differences were observed for a number of genes or protein families classified as potential AMPs. These included the varisin and factor D-like genes and the Defensin, Cystatins and TIL domain families. The Defensin family is a well-known antimicrobial peptide family

found in many vertebrates and invertebrates, including a number of tick species (reviewed in Taylor, 2006). Defensins show antimicrobial activity against Gram-positive bacteria (Yi *et al.*, 2014), Gram-negative bacteria and fungi (Tonk *et al.*, 2014; Tonk *et al.*, 2015) and protozoan parasites (Tsuji *et al.*, 2007; Cabezas-Cruz *et al.*, 2016b). A Defensin isolated from *Dermacentor variabilis*, varisin, showed antimicrobial activity against *B. burgdorferi* (Johns *et al.*, 2001b) and down-regulation by RNA interference (RNAi) caused a reduction in the antimicrobial activity (Hynes *et al.*, 2008). Factor D-like is a CLIP-domain containing serine protease that exhibits antimicrobial activity in arthropods (Kawabata *et al.*, 1996; Kanost and Jiang, 2015). Previously, a factor D-like gene was significantly up-regulated when *D. variabilis* was challenged by either bacterial or fungal pathogens (Simser *et al.*, 2004b; Jaworski *et al.*, 2010). Cystatins (cysteine protease inhibitors) showed antimicrobial activity in *Tachypleus tridentatus*, the horseshoe crab (Agarwala *et al.*, 1996) and *Bombyx mori*, silk moth (Yamamoto *et al.*, 1999), indicating a potential role in arthropod innate immunity. A role in tick immunity has also been shown where the *Haemaphysalis longicornis* cystatin, *Hlcyst-2*, showed increased expression in *Babesia gibsoni*-infected larval ticks and the recombinant protein showed anti-*Babesial* activity (Zhou *et al.*, 2006). The *R. microplus* TIL domain containing protein, BmSI-7, showed strong antimicrobial activity, suggesting that TIL domain containing proteins might have a role to play in tick innate immunity (Sasaki *et al.*, 2008). Furthermore, four putative TIL domain containing proteins were highly abundant in the *I. ricinus* haemocyte transcriptome (Kotsyfakis *et al.*, 2015a). Thirdly, significant differences in expression levels were observed between *R. appendiculatus* and *R. zambeziensis* of potential apoptotic orthologues, mitochondrial porin (T2) and spectrin α chain (fodrin). Previously, orthologues of the *I. scapularis* spectrin α chain gene (XP_002433506.1) could not be found in other tick species (Ayllón *et al.*, 2013), but that was prior to the availability of the *R. appendiculatus* (de Castro *et al.*, 2016) and *R. zambeziensis* (de Castro *et al.*, 2017) transcriptomes. Invading pathogens target apoptosis signalling pathways of the host to prevent cell death (Ayllón *et al.*, 2013; Ayllón *et al.*, 2015; Ribet and Cossart, 2015) and protozoan parasites, including *Theileria*, have shown to inhibit apoptosis and promote cell survival in vertebrate hosts (Heussler *et al.*, 2006; Bruchhaus *et al.*, 2007). During *Anaplasma phagocytophilum* infection of *I. scapularis*, the pathogen down-regulated the expression of the mitochondrial porin and spectrin α chain genes to reduce apoptosis, thereby enhancing infection levels (Ayllón *et al.*, 2013). The

mitochondrial porin, a voltage-dependent anion channel, is responsible for movement of the metabolic flux across the outer membrane of the mitochondria (Colombini *et al.*, 1996) and early on during apoptosis the channel is closed leading to mitochondrial shutdown (Lemasters and Holmuhamedov, 2006). Down-regulation of the gene by RNAi has resulted in increased infection in tick cell lines and salivary glands (Ayllón *et al.*, 2013; Ayllón *et al.*, 2015). Spectrin α chain is a membrane-associated cytoskeletal protein important in maintaining normal membrane structure (Bennett and Gilligan, 1993). The protein is cleaved by caspases during apoptosis, presumably resulting in the defined apoptotic feature of membrane blebbing (Martin *et al.*, 1995). A dual functionality of the spectrin α chain gene during pathogen infection was demonstrated by Ayllón *et al.* (2013), where *A. phagocytophilum* reduced the gene's expression in the *I. scapularis* salivary glands, inhibiting apoptosis, and increased the expression in the midguts causing structural reorganisation, which may assist pathogens during invasion (Shimada *et al.*, 1999; Ribet and Cossart, 2015; Cardoso *et al.*, 2016). The protein families or gene orthologues discussed above are all involved in some aspect of tick innate immunity, be it pathogen recognition, antimicrobial (including antiprotozoal) activity or apoptosis activation. Considering that *R. appendiculatus*, the tick with lower vector competence for *T. parva* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998) showed higher expression levels for all of these genes, it might suggest that *R. appendiculatus* has a stronger innate immune system compared to *R. zambeziensis* and consequently might be less susceptible to infection by *T. parva*.

Four genes or protein families involved in different aspects of pathogen transmission or infectivity rates showed more abundant expression values in *R. zambeziensis* when compared to *R. appendiculatus*. These included the putative SAT molecules, Ixostatin and HRF families (including the *tHRF* gene); the transcription factor, subolesin; and the antioxidant gene, GST. The first differently expressed family between *R. appendiculatus* and *R. zambeziensis* was the Ixostatin superfamily. This family is predominantly expanded in *Ixodes* lineages and contains two SAT proteins, salivary protein of 15kDa (SALP15) and salivary protein of 16kDa (SALP16). During infection of *I. scapularis*, the outer surface protein C (OspC) of *B. burgdorferi* was shown to be bound to SALP15, which protected the parasite from antibody-mediated elimination by the vertebrate host, aiding parasite transmission (Ramamoorthi *et al.*, 2005). The *B. burgdorferi* infection caused an increase in the transcription of SALP15 in

the salivary glands and RNAi mediated down-regulation resulted in reduced spirochete load in the vertebrate host (Ramamoorthi *et al.*, 2005). Similarly, *A. phagocytophilum* infection caused SALP16 transcription increases in the *I. scapularis* salivary glands and silencing by RNAi resulted in reduced salivary gland infection rates (Sukumaran *et al.*, 2006). It was further shown that the up-regulation of SALP16 was not as a result of a tick defence mechanism but mediated by the *A. phagocytophilum* pathogen to aid its transmission (Sultana *et al.*, 2010). The second secretory protein family that showed expression variation between the two species was the HRF family. In both ticks, the HRF family consisted of a single member, the orthologue of the *tHRF* gene. The *tHRF* gene of *I. scapularis* was shown to be up-regulated during infection by *B. burgdorferi* and immunisation with recombinant tHRF or RNAi silencing of *tHRF* resulted in reduced parasite transmission rates to the host, which indicated that the protein has SAT functions (Dai *et al.*, 2010). The expression of *tHRF* increased as feeding progressed, potentially facilitating blood flow by vascular permeability and *B. burgdorferi* dissemination through the vertebrate host. The *tHRF* of *D. variabilis* also showed up-regulation in response to *Rickettsia montanensis* infection (Mulenga *et al.*, 2003b). The third gene that showed elevated expression in *R. zambeziensis* compared to *R. appendiculatus* was subolesin. The subolesin transcription factor has been shown to function in the immune response during infection of a number of tick species (Zivkovic *et al.*, 2010b; Merino *et al.*, 2011a; Merino *et al.*, 2011b). *Anaplasma marginale* infection resulted in the up-regulation of subolesin, which increased infection of salivary glands but not midguts (Zivkovic *et al.*, 2010b). Gene knockdown or immunisation of cattle with a recombinant protein decreased tick infection levels (de la Fuente *et al.*, 2006; Merino *et al.*, 2011a; Merino *et al.*, 2011b). The fourth significantly different expressed gene was GST. Defence mechanisms targeted towards pathogen invasion results in imbalanced homeostasis or oxidative stress from reactive oxygen species in the host cells (Nappi and Vass, 1998; Pereira *et al.*, 2001). Glutathione S-transferases, together with other antioxidant enzymes, detoxify xenobiotic compounds to protect cells against oxidative and chemical stress (Hayes *et al.*, 2005; Yan *et al.*, 2013). It has been shown that GST expression levels increase during pathogen invasion (Rudenko *et al.*, 2005; de la Fuente *et al.*, 2007; Cotté *et al.*, 2014) and that knockdown studies resulted in reduced infection rates (de la Fuente *et al.*, 2007; Kocan *et al.*, 2009). These results suggested that pathogens might require GSTs to reduce infection-induced oxidative stress, making a better-suited environment for pathogen infection (Kocan *et al.*, 2009).

The abundance in *R. zambeziensis* of SAT genes (Ixostatin and *tHRF*) that assist pathogen transmission and genes that alter infection rates in ticks, such as subolesin and GST, might explain the higher infectivity rates and vector competence seen for this species when compared to *R. appendiculatus* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998).

On a different note, it is known that *R. zambeziensis* has extreme environmental adaptability, as *R. zambeziensis* occurs in hotter and drier regions (Walker *et al.*, 2005) and showed better tolerance for different abiotic experimental conditions, including low humidity and high temperature (Madder *et al.*, 2005), when compared to *R. appendiculatus*. The higher expression of GST and *tHRF* in *R. zambeziensis* compared to *R. appendiculatus* might assist in *R. zambeziensis*' ability to survive in arid environments. Firstly, oxidative stress can be increased by environmental factors including temperature (Lushchak, 2011) and the expression of GST genes has been induced during experimental abiotic stressors such as heat (Yan *et al.*, 2013). Furthermore, the transgenic over-expression of a GST in *Caenorhabditis elegans* led to resistance against heat shock and an overall increase in the lifespan of the worms (Ayyadevara *et al.*, 2005). Secondly, HRF orthologues have been shown to function in early developmental processes and as anti-apoptotic agents in response to heat shock and oxidative stress in eukaryotes (reviewed in Bommer, 2012). Indeed in *Arabidopsis*, the overexpression of the HRF orthologue, *AtTCTP* (*A. thaliana* translationally controlled tumor protein), enhanced drought tolerance of the transgenic plants (Kim *et al.*, 2012). Furthermore, *Arabidopsis* and *Drosophila* TCTP orthologues reciprocally rescued the knockdown mutants of the other species, indicating that the functions of the TCTP proteins in insects and plants are highly conserved (Brioudes *et al.*, 2010) and that the insect orthologue might be involved in drought tolerance as well. These data suggest that the higher abundance of *tHRF* and GST in *R. zambeziensis* might assist the tick in its adaptability to extreme environments.

Even though the Pfam domain comparison showed that most (81%) of the domains were shared between *R. appendiculatus* and *R. zambeziensis*, the occurrence of three domains significantly differed between the species. These were: the Spectrin repeat domain that forms part of the spectrin α chain gene, a previously discussed gene involved in the cytoskeletal structure and membrane blebbing during apoptosis (Bennett

and Gilligan, 1993; Martin *et al.*, 1995); and two domains found in proteins involved in the extracellular matrix, the Fibronectin type III (Narasimhan *et al.*, 2014) and Laminin EGF (Vlachou *et al.*, 2001) domains. Extracellular matrix proteins have shown to be targeted by adhesive surface proteins of pathogens, adhesins, to invade host cells (Henderson *et al.*, 2011; Ribet and Cossart, 2015). Fibronectin (containing multiple Fibronectin type III domains) is an extracellular matrix glycoprotein used for cell structure, adhesion and signalling (Xu and Mosher, 2011), and is targeted by a large number of bacterial fibronectin-binding proteins to facilitate infection (Henderson *et al.*, 2011). Interestingly, even though fibronectin orthologues were absent from the *Drosophila* genome, a large number of Fibronectin type III domains were found, indicating that this domain could still be extensively used by invertebrates for cell adhesive and signalling receptors (Hynes and Zhao, 2000). Concordantly, an *I. scapularis* gut protein containing a number of Fibronectin type III domains has been shown to interact with *B. burgdorferi* during infection and down-regulation of the gene resulted in reduced pathogen load in the tick and reduced transmission to the vertebrate host (Narasimhan *et al.*, 2014). Interactions have also previously been shown between fibronectin of the vertebrate host and the *B. burgdorferi* surface proteins, BBK32 (Fikrig *et al.*, 2000) and RevA (Brissette *et al.*, 2009a). Another extracellular matrix protein, Laminin, containing a number of Laminin EGF domains, has been shown to be a major component of basal lamina of *D. melanogaster* (Hynes and Zhao, 2000) and functions in adhesion and cell structuring and shape (Xu and Mosher, 2011). In the *Anopheles gambiae* midgut, a number of *Plasmodium* surface proteins have shown to be bound to laminin during multiplication (Vlachou *et al.*, 2001; Warburg *et al.*, 2007) and the down-regulation of a laminin gene resulted in significant reduction in the development of *Plasmodium* oocysts (Arrighi *et al.*, 2005). Laminin is also the target of *B. burgdorferi* during murine infection by means of interaction with the pathogen outer-surface proteins, RevA (Brissette *et al.*, 2009a), BmpA (Verma *et al.*, 2009) and ErpX (Brissette *et al.*, 2009b). Analyses of the *R. appendiculatus* and *R. zambeziensis* transcriptomes, revealed an overrepresentation of the Fibronectin type III and Laminin EGF domains in *R. appendiculatus*, contradictory to what would be expected because of its lower vector competence for *T. parva*. This indicates the current limited understanding of pathogen transmission in ticks, but might also suggest that a greater diversity of extracellular matrix proteins exists in *R. appendiculatus*, potentially supporting the strong defence responses of this tick species against pathogens. Similarly,

polymorphisms in human Toll-like receptors (TLRs) resulted in variations in the immune responses launched against *B. burgdorferi* (Rahman *et al.*, 2016). An alternative possibility for the differences observed in the domains of extracellular matrix proteins, taken together with the aforementioned cytoskeletal Spectrin repeat domain, is that they might infer structural differences in the salivary glands of *R. appendiculatus* and *R. zambeziensis*. It has been shown that the development of *T. parva* in tick salivary glands resulted in the infected acini greatly increasing in size to harbour all the sporozoites (Purnell and Joyner, 1968). Such an increase in cell size would require membrane rearrangement and restructuring by cytoskeletal and/ or extracellular matrix proteins (Xu and Mosher, 2011), and may have implications for the competency of the ticks to vector *T. parva*. Structural variation in the salivary glands of *R. appendiculatus* and *R. zambeziensis* and the potential involvement with *T. parva* transmission remain to be experimentally confirmed.

Long non-coding RNA molecules, as the name suggests, are transcripts longer than 200 nucleotides and non-coding, i.e. no ORF longer than 100 amino acids can be predicted from the sequence (Dinger *et al.*, 2008). Long ncRNAs have shown to act as scaffolds for protein complexes, control different aspects of transcription (reviewed in Kung *et al.*, 2013; Engreitz *et al.*, 2016) and regulate the vertebrate immune system (reviewed in Aune and Spurlock, 2016; Zhang and Cao, 2016). Recently, the up-regulation of a number of lncRNAs during virus and endosymbiont infection of *Aedes aegypti*, has suggested a possible role in vector immunity for lncRNAs (Etebari *et al.*, 2016). Furthermore, a number of non-coding RNAs have been identified in the *I. scapularis* genome (Gulia-Nuss *et al.*, 2016). During the transcriptome assemblies of *R. appendiculatus* and *R. zambeziensis*, respectively 7414 and 7947 transcripts were obtained for which no ORFs could be predicted (refer to Chapters 2 and 3). A lncRNA filtering pipeline was employed to have higher confidence that the remaining transcripts after the filtering process might indeed be non-coding and potentially lncRNAs. Close to four thousand putative lncRNAs were predicted in each species and of these 439 were identified as potential orthologues shared between *R. appendiculatus* and *R. zambeziensis*. Interestingly, *R. appendiculatus* showed higher overall expression for the set of putative lncRNA orthologues, based on total and average expression and the number of transcripts with higher expression levels. Considering the functionality of lncRNA in immunity (Aune and Spurlock, 2016; Etebari *et al.*, 2016; Zhang and Cao,

2016), it is conceivable that *R. appendiculatus*, with its lower infectivity rates of *T. parva* and seemingly superior immune system compared to *R. zambeziensis*, might employ lncRNAs as part of its immune system. As lncRNAs have been implicated in diverse functions, the functions of the putative lncRNAs identified in the current study remain unknown and will require future experimental evaluations.

Apart from involvement in tick immune responses or pathogen transmission and infection rates, a number of the genes/ families discussed in the aforementioned sections have also been implicated in tick blood feeding and host immune evasion, e.g. cystatins (Schwarz *et al.*, 2012), subolesin (de la Fuente *et al.*, 2006), SALP15 (Garg *et al.*, 2006), SALP16 (Hidano *et al.*, 2014) and tHRF (Dai *et al.*, 2010). Therefore, validation will be required to determine whether the differences observed between *R. appendiculatus* and *R. zambeziensis* are as a result of vector competence or species differences in blood feeding. Furthermore, the largest limitation of the current study is that assumptions are made based on data from uninfected ticks. Vectors and pathogens have co-evolved for millions of years (de la Fuente *et al.*, 2016b) and ticks have shown to respond differently to pathogen infection. For example, in *D. variabilis*, that is refractory to *B. burgdorferi*, the parasite is rapidly cleared from the tick compared to the establishment of infection in *I. scapularis*, a natural vector of *B. burgdorferi* (Johns *et al.*, 2001a). It is therefore of importance to perform infection studies of specific pathogens with their specific vectors to obtain a clear picture of the vector-pathogen interaction. However, infection studies of *T. parva* in *R. appendiculatus* and *R. zambeziensis* are confined by the low infectivity rates in tick salivary glands [ranges of 3 - 24 infected acini in *R. appendiculatus* and 19 - 43 in *R. zambeziensis* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998)], and the fact that the parasite infects only a single cell type of a single acini type (the e cells of acini III, reviewed in Bishop *et al.*, 2004). This will result in transcriptomic studies with very low parasite to tick transcript ratios, which could explain the results obtained in a previous study that found no significant differences between *T. parva*-infected and non-infected *R. appendiculatus* ticks using conventional technologies (Nene *et al.*, 2004). To better understand the responses of *R. appendiculatus* and *R. zambeziensis* during *T. parva* infection, we propose that rather than conventional sampling, infected cells should be separated from the background uninfected salivary gland tissues by either flow cytometry or laser capture microdissection. This will result in better parasite to tick

transcript ratios, which would assist analyses. Another limitation of the current study is that only salivary gland gene catalogues were compared. Even though it is known that *T. parva* spends much of its life cycle in the tick salivary glands (reviewed in Bishop *et al.*, 2004), sampling of other tissues would broaden the understanding of vector-*T. parva* interactions. For instance, it is not clear whether the difference observed in vector competence between *R. appendiculatus* and *R. zambeziensis* is due to *T. parva* not being able to infect the midguts of *R. appendiculatus* or if the selection occurs at the salivary gland interface. Further studies are therefore needed to address the effect that other organs might have on *T. parva* infection.

It is of note to point out that the fact that *R. appendiculatus* has shown lower vector competence of *T. parva* than *R. zambeziensis*, does not preclude it from being a extremely successful vector of the parasite. *Rhipicephalus appendiculatus* is widely distributed throughout central, eastern and southern Africa, making it one of the most economically important vectors in Africa (Norval *et al.*, 1992; Walker *et al.*, 2003). It is also the main vector of *T. parva* (Norval *et al.*, 1992). Vector capacity is an umbrella term when describing a vector's ability to transmit a parasite and includes: genetic components, such as vector competence; and environmental and behavioural components, such as tick densities and distribution (Beerntsen *et al.*, 2000). In this sense *R. appendiculatus* has a very high vector capacity for *T. parva*.

4.6 Conclusions

The evaluation of the breadth of the assembled *R. appendiculatus* and *R. zambeziensis* transcriptomes with regard to containing core orthologous genes was used as a validation that the transcriptomes were similarly sampled and opened the possibility to compare the transcriptomes with each other. In the comparison between the salivary gland transcriptomes of the species, the organs in which a large proportion of the life cycle of *T. parva* occurs, bioinformatic signatures were observed that contributed to the explanation of the variance in vector competence of the species. The transcriptomes of *R. appendiculatus* and *R. zambeziensis* were genetically very close but a number of differences were observed in the expression of immunity- and pathogen transmission-related secretory protein families and genes. The salivary glands of *R. appendiculatus*

showed greater expression of general tick immunity genes or families, such as antimicrobial/ antiprotozoal, pathogen recognising and apoptotic genes, suggesting that *R. appendiculatus* has a stronger innate immune system compared to *R. zambeziensis* and resultantly might be less susceptible to parasite infection, such as *T. parva.* In *R. zambeziensis*, higher expression was observed in genes that have previously been manipulated by parasites to enhance their life cycle in the tick vector and transmission to the animal host, which might also be advantageous to *T. parva*. Additionally, potential evidence was obtained that *R. zambeziensis* salivary glands might be structurally more adaptable to size increases, as would be expected for successful multiplication of *T. parva* sporozoites and the repertoire of *R. appendiculatus* extracellular matrix proteins may have higher diversity assisting as a defence barrier during pathogen infection, such as *T. parva*. Further, the comparison between the two tick species revealed putative evidence to support the differences observed in environmental adaptation between *R. appendiculatus* and *R. zambeziensis*, where *R. zambeziensis* is adapted to survival in hotter and drier conditions. Overall higher expression for a set of putative lncRNA orthologues was observed in *R. appendiculatus*, which might indicate that the tick employs lncRNAs as part of its immune defence mechanisms against pathogen infections, such as *T. parva*. However, the true functions of the lncRNAs remain to be experimentally proven. Further, a species phylogeny was constructed from the transcriptomic and genomic data of 17 tick species using a SCO reference dataset, that is offered as a toolbox for future phylogenetic inferences in ticks.

4.7 Authors' contributions

BJM supplied the funding. MHD and BJM designed and conceived the study. MHD performed the bioinformatic analyses and drafted and wrote the manuscript. MHD, DJGR and BJM revised the manuscript. The final version of the manuscript was read and approved by all the authors.

4.8 Acknowledgements

The authors thank Mr. Jonathan Featherston of the Biotechnology Platform, Agricultural Research Council, South Africa for performing the OrthoMCL analysis.

CHAPTER 5

Conclusions

Corridor disease is a devastating disease of southern Africa in which cattle can die within four days from showing clinical symptoms. The disease is transmitted by the *Theileria parva* parasite from buffalo to cattle in regions where the animals are in close contact (Uilenberg, 1999). Corridor disease is a controlled disease in South Africa based on stringent dipping regimes and highly regulated movement control of buffalo (Animal Disease Act 1984, Act No. 35). Yet, sporadic outbreaks occur (Sibeko *et al.*, 2008; Mbizeni *et al.*, 2013) and this can be expected to increase in coming years as the game industry, and as a result the demand for buffalo, is increasing. There are currently no treatments for the disease that are permitted in South Africa as these result in a carrier-state in the recovered animals (Animal Disease Act 1984, Act No. 35). Additionally, cattle that do survive the disease must, by law, be slaughtered due to the risk of the disease changing aetiology from buffalo-cattle transmission to cattle-cattle transmission. It has been speculated that this change in aetiology could result in an epidemic of similar proportions to what was seen with the introduction of East Coast fever in South Africa more than a century ago (Yusufmia *et al.*, 2010). Further, the currently available vaccine that protects cattle against *T. parva*, the 'infection-and-treatment' method, may not be used in South Africa, as this also results in a carrier-state in cattle (Radley *et al.*, 1975). Since the current methods of Corridor disease control in South Africa, acaricides and buffalo movement restrictions, are become less successful, the only manner by which cattle can be protected from Corridor disease in South Africa is by the development of recombinant vaccines that will prevent ticks from feeding and transmitting pathogens to the cattle.

With the advent of the genomics revolution and the recent availability of completed and draft tick genomes, new-generation vaccines and reverse vaccinology has obtained much attention in the tick vaccine development field (Guerrero *et al.*, 2012; Marcelino *et al.*, 2012; de la Fuente *et al.*, 2016a; Lew-Tabor and Rodriguez-Valle, 2016). Reverse vaccinology is a bioinformatic approach that searches genomic sequence for the presence of antigenic regions that could potentially be attractive vaccine candidates (Seib *et al.*, 2012). The selection of appropriate vaccine candidates, before the onset of any experimental analyses, is perhaps the most important step in increasing the success rate of anti-tick vaccine developments. Using current technologies, genome sequencing of a great number of tick species is impracticable due to the large predicted sizes of tick genomes (Ullmann *et al.*, 2005; Geraci *et al.*, 2007). Therefore, sequencing

of tick transcriptomes to generate all the gene sequences expressed in a relevant tick organ, have recently become a more widely adopted strategy by tick biologists (reviewed in Mans *et al.*, 2016). Recently, this strategy of generating and using an expressed gene catalogue to search for vaccine candidates has proven successful by identified efficacious vaccine candidates in ticks (Maruyama *et al.*, 2017). Directed towards this, and as a first step towards the long-term aim of developing recombinant anti-tick vaccines that will protect cattle from Corridor disease, this PhD aimed to sequence, assemble and annotate the sialotranscriptomes of *Rhipicephalus appendiculatus* and *R. zambeziensis,* two vectors of Corridor disease. A further aim was to compare the transcriptomes to find molecular correlates of the differences between the species, with special focus on variability in vector competence. A clear understanding of the variability between the vectors of Corridor disease would assist the selection of vaccine candidates that potentially could result in cross-species protection.

The assembled transcriptomes of *R. appendiculatus* and *R. zambeziensis* are valuable resources for the selection of future vaccine candidates in the quest to find vaccines that will protect cattle against Corridor disease and will be permitted to be used in South Africa. The basis for this assumption is on a number of different reasons that will be explained henceforth. First, the transcriptomes contain most of the important genes expressed in the salivary glands during feeding and can function as comprehensive search databases for reverse vaccinology approaches. In Chapters 2 and 3, the transcriptome of each species was assembled from a diverse sample set, representing female and male ticks and the main feeding stages: unfed, early feeding and late feeding. Additionally, based on a number of criteria the transcriptomes were shown to be of high quality, nearly complete and representative of other tick sialotranscriptomes. Furthermore, previous studies have largely focussed on assembling the gene repertoires of feeding females and only a few have investigated males (Aljamali *et al.*, 2009; Xiang *et al.*, 2012; Tan *et al.*, 2015a; de Castro *et al.*, 2016). Males assist females during feeding by maintaining the feeding site and it has been shown that females feed faster in the presence of male ticks (Wang *et al.*, 1998). With the great degree of secretory protein transcript expression observed in both *R. appendiculatus* and *R. zambeziensis* male salivary glands, it can be expected that the sialomes of male ticks have large effects on the immune response at feeding sites. The

inclusion of the male transcriptomes in the databases for future vaccine development will result in additional avenues for vaccine candidate selection.

Second, vaccine candidates selected from these transcriptomes would potentially disrupt important tick feeding processes and result in long-lasting immune responses in the host. Chapters 2 and 3 focussed on sequencing the salivary glands of the tick species, since proteins secreted into the host are essential for maintaining the feeding site, by promoting blood feeding and evading host immune responses (Francischetti *et al.*, 2009). It is important that vaccines target proteins with critical functions in the tick, whereby the disruption of function will result in termination of feeding and tick rejection. The additional advantage of focussing on the salivary glands is that vaccines targeting proteins from these organs, known as 'exposed' antigens, would be naturally presented to the host's immune system during feeding resulting in continued priming of the immune system (Nuttall *et al.*, 2006). These vaccines would potentially not require immune booster injections to achieve adequate protection, which would assist implementation by cattle farmers.

Third, in combination, the *R. appendiculatus* and *R. zambeziensis* transcriptomes can be used to select candidates for potential cross-species protection. *Rhipicephalus appendiculatus* and *R. zambeziensis* are morphologically very similar (Walker *et al.*, 1981), but have been classified as two separate species based on phylogenetic reconstructions (Mtambo *et al.*, 2007a). In Chapter 4 it was shown that the two species are genetically highly similar, with half of each transcriptome showing orthology between the two species and many of the orthologous proteins showing larger than 95% similarity. This level of similarity opens the possibility that the selection of certain vaccine candidates could result in simultaneous protection from both Corridor disease vectors.

Fourth, the assembled transcriptomes can be used to identify vaccine candidates that potentially interfere with pathogen multiplication and transmission. Another strategy followed by researchers in the past has been to select vaccine candidates that, additional to disrupting crucial feeding functions, also impede pathogen transmission (Merino *et al.*, 2013a). In this way both the vector and the pathogen are targeted at the same time. In Chapter 4, the comparative analysis between the transcriptomes of *R.*

appendiculatus and *R. zambeziensis* indicated that even though the species are genetically highly similar, their gene expression profiles differed substantially. *Rhipicephalus zambeziensis* has been shown to have better vector competence for *T. parva* than *R. appendiculatus* (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998) and correspondingly, a number of the transcriptional differences observed between the species were of secretory proteins and protein families previously associated with tick immunity or pathogen transmission. This revealed bioinformatic signatures of the difference in vector competence of the tick species. In the discussion section of Chapter 4, the limitation of this comparison without the inclusion of data from *T. parva*-infected ticks was acknowledged, since it is known that ticks have evolved different mechanisms to combat pathogen infections (Johns *et al.*, 2001a; Villar *et al.*, 2010). Low salivary gland infectivity rates (Potgieter *et al.*, 1988; Blouin and Stoltsz, 1989; Ochanda *et al.*, 1998), together with the infection of only a single cell in a single type of acini (reviewed in Bishop *et al.*, 2004), result in very low ratios of *T. parva*-infected-to-non-infected salivary gland tissues. This currently limits the analysis of *T. parva*-infected ticks using conventional methods of sequencing the entire salivary gland. The use of specialised technologies, such as flow cytometry or laser capture microdissection, that can separate infected cells from the background uninfected salivary gland tissue will be required to analyse *T. parva*-infected ticks. This important future research endeavour will provide biological validation of the results obtained from the bioinformatic comparison performed in Chapter 4.

Fifth, the comprehensive transcriptomes containing all or most of the members of large multi-genic secretory protein families, will assist in the selection of 'exposed' antigens as vaccine candidates by searching for family representative antigens that could potentially protect the animal against an entire secretory protein family and not just a single member. As mentioned above, naturally eliciting an immune response during tick feeding is a strong attraction for the use of 'exposed' antigens, or secretory proteins, as vaccine candidates (Nuttall *et al.*, 2006). At the same time, this constant exposure to the host's immune system has resulted in adaptations in these proteins to evade the immune responses, and in most cases this involved expansions of the families (Mans and Neitz, 2004; Francischetti *et al.*, 2009). In Chapter 3, the comparison between the different feeding phases of *R. zambeziensis* revealed dynamic expression of the transcripts of the secretory protein families. This expression pattern has previously been termed 'sialome

switching', which is an antigenic variation mechanism employed by ticks to avoid immune responses of the host while still preserving important functions during feeding. This fascinating adaptation of ticks allows them to feed undetected on their hosts for extended time periods. Both, the large number of proteins and the dynamic expression in the secretory protein families are substantial confining factors when using secretory proteins as targets in vaccine developments, since functionally redundant family members could be expressed to maintain the crucial function while removing the antigenicity of a vaccine targeted protein, resulting in vaccine escape (Guerrero *et al.*, 2012). In this regard, the transcriptomes of *R. appendiculatus* and *R. zambeziensis* be used in a strategy for selecting vaccine candidates based on protection from an entire, or close to entire, secretory protein family opposed to a single protein. Multiple-alignments of the secretory protein family members will reveal conserved sequence regions among the family members. If these coincide with regions showing antigenicity predictions, vaccines could hypothetically be developed to protect against a secretory protein family (family representative antigens) and not just a single protein.

Apart from the assembly of valuable resources for future vaccine candidate selection, some additional observations were made during the PhD that merit further investigation in future. The first being the regulatory processes involved in antigenic variation. It is anticipated that a finely tuned process, such as antigenic variation, will be under very stringent transcriptional regulation. It has been suggested that epigenetics is associated with the regulation of secretory protein expression (Adamson *et al.*, 2013), but the mechanism of this is still completely unknown. During the assembly of the transcriptomes of *R. appendiculatus* and *R. zambeziensis*, numerous transcripts were assembled for which open reading frames could not be predicted. In Chapter 4, these transcripts were examined in greater detail and approximately 4000 putative long non-coding RNA (lncRNA) molecules were predicted in each species, with about 11% of these shared between the species. The involvement of lncRNAs in a number of transcriptional regulatory processes has been well established (reviewed in Kung *et al.*, 2013; Engreitz *et al.*, 2016), but the functions of lncRNAs in ticks are completely unexplored and warrant future characterisation. Knowledge of the regulation of secretory protein expression is fundamental to the understanding of tick feeding and host immune evasion, and would result in better-informed vaccine candidate selections.

Another observation was the large divergences found in certain secretory protein sequences between different geographic strains of *R. appendiculatus* (Chapter 2): the southern African strain assembled during this PhD vs. publically available *R. appendiculatus* sequences (most probably from eastern African strains). The southern and eastern African *R. appendiculatus* strains have been shown to differ behaviourally (Madder *et al.*, 2002), morphologically (Speybroeck *et al.*, 2004) and genetically (Mtambo *et al.*, 2007b). However, it was unexpected to observe differences in very important secretory proteins that have previously been shown to have crucial functions in blood acquisition and induced strong antibody responses in hosts. These differences highlighted the importance of having a clear understanding of the genetic background of the natural tick population from which vaccine protection is required. Furthermore, it has previously been shown that laboratory-bred tick stocks have diverged significantly from the populations they were initially sampled from (Kanduma *et al.*, 2016b). In context of vaccine development, it is integral that the laboratory tick stock from which experimental evidence will be obtained needs to represent the genes and proteins circulating in the natural tick population in order to achieve successful protection. In this respect, a future research endeavour would be to characterise the diversity in the South African *R. appendiculatus* tick populations and to validate that the laboratory tick stock housed at the Onderstepoort Veterinary Research Institute represents the natural *R. appendiculatus* populations of South Africa. It would also be of value to perform this analysis in *R. zambeziensis* ticks.

It was further observed, in Chapters 2 and 3, that the Glycine rich superfamily contained a number of truncated proteins. This was ascribed to technical limitations of assembly algorithms when short read sequences are used to assemble low complexity and repeat regions (Wang *et al.*, 2009), such as members of the Glycine rich superfamily. This family represented the largest proportion of secretory protein expression in the transcriptomes of both *R. appendiculatus* and *R. zambeziensis*, making it essential to have the family well characterised. Oxford Nanopore and Pacific BioSciences, the so called 3rd generation sequencing technologies, produce long sequence reads that negate the assembly process, and associated caveats, entirely (Bolisetty *et al.*, 2015; Rhoads and Au, 2015). These technologies would improve the characterisation of the Glycine rich superfamily as well as other truncated or

misassembled sequences in the transcriptomes and would be a meaningful implementation in future transcriptome improvement efforts.

In summary, in this PhD the sialotranscriptomes of two vectors of Corridor disease, *R. appendiculatus* and *R. zambeziensis*, were assembled and annotated. The transcriptomes are of high quality, near complete and valuable resources for the future selection of vaccine candidates. Dynamic expression of secretory protein families - representing antigenic variation - was observed during feeding, putative lncRNAs were predicted and proposed to be involved in transcriptional regulation of secretory proteins and bioinformatic signatures were observed that potentially explain the variability in vector competence observed between *R. appendiculatus* and *R. zambeziensis*. Certain results obtained and conclusions made during the PhD were based on bioinformatic algorithms and require future empirical functional validation. Regardless, the assembled transcriptomes will advance the understanding of the evolution of tick blood feeding and the mechanisms involved in the regulation thereof and will be valuable resources for future vaccine candidate selection.

APPENDICES

APPENDIX A

Ethics approval letters

AEC 12.11

ARC • LNR

Onderstepoort Veterinary Institute

Animal Ethics

Decision of the Animal Ethics Committee for the use of living vertebrates for research, diagnostic procedures and product development

APPROVAL PERIOD: 2011/ 2012

PROJECT NUMBER:	Routine tick colony maintenance					
PROJECT TITLE:	Routine tick colony maintenance					
PROJECT LEADER:	Arthur Spickett					
DIVISION:	PVVD					
CATEGORY:	B					
SPECIES OF ANIMAL:	Bovines	Ovines	Chickens	Rodents	Rabbits	Reptiles
NUMBER OF ANIMALS:	10 Annually	10 Annually	10 Annually	20 Annually	20 Annually	10 Annually
NOT APPROVED:						
APPROVED:	APPROVED					

PLEASE NOTE: Should the number or species of animal(s) required, or the experimental procedure(s) change, please submit a revised animal ethics clearance form to the animal ethics committee for approval before commencing with the experiment

SIGNATURE: Dr L. Lopez
CHAIRPERSON ANIMAL ETHICS COMMITTEE

DATE: 23/05/11

Onderstepoort Veterinary Institute

AEC 01.15

Animal Ethics

Decision of the Animal Ethics Committee for the use of living vertebrates for research, diagnostic procedures and product development

APPROVAL PERIOD: 2015/2016

PROJECT NUMBER:	Routine tick colony maintenance					
PROJECT TITLE:	Routine tick colony maintenance					
PROJECT LEADER:	Ben Mans					
DIVISION:	PVVD					
CATEGORY:	B					
SPECIES OF ANIMAL:	Bovines	Ovines	Chickens	Rodents	Rabbits	Reptiles
NUMBER OF ANIMALS:	10 Annually	10 Annually	10 Annually	20 Annually	20 Annually	10 Annually
NOT APPROVED:						
APPROVED:	APPROVED					

PLEASE NOTE: Should the number or species of animal(s) required, or the experimental procedure(s) change, please submit a revised animal ethics clearance form to the animal ethics committee for approval before commencing with the experiment

SIGNATURE: Dr P. Mutowembwa
CHAIRPERSON ANIMAL ETHICS COMMITTEE

DATE: 15/04/2015

UNISA university of south africa

2014-05-15

Ref. Nr.: 2014/CAES/098

To:
Student: M De Castro **Student nr: 53321529**
Supervisor: Prof BJ Mans
Department of Environmental Science
College of Agriculture and Environmental Sciences

Dear Prof Mans and Ms De Castro

Request for Ethical approval for the following research project:

Sialotranscriptomics of Rhipicephalus appendiculatus and R. zambeziensis infected with Theileria parva

The application for ethical clearance in respect of the above mentioned research has been reviewed by the Animal Research Ethics Review Committee of the College of Agriculture and Environmental Sciences, Unisa. Ethics clearance for the above mentioned project (Ref. Nr.: 2014/CAES/098) **is given** for the duration of the research project.

However, the applicant is reminded that a Section 20 permit is required for the infectious disease phase of the study. This permit must be submitted to the Committee as soon as it becomes available.

Please be advised that should any part of the research methodology change in any way as outlined in the Ethics application (Ref. Nr.: 2014/CAES/098), it is the responsibility of the applicant to inform the CAES Animal Ethics Committee. In this instance a memo should be submitted to the Animal Ethics Committee in which the changes are identified and fully explained.

The Animal Ethics Committee wishes you all the best with this research undertaking.

Kind regards,

Prof E Kempen,
CAES Animal Ethics Review Committee Chair

PLEASE NOTE THE PERMIT MUST BE SUBMITTED TO THE ETHICS COMMITTEE

Prof MJ Linington
Executive Dean: College of Agriculture and Environmental Sciences

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsmile: +27 429 12 429 4150
www.unisa.ac.za

agriculture,
forestry & fisheries

Department:
Agriculture, Forestry and Fisheries
REPUBLIC OF SOUTH AFRICA

Directorate Animal Health, Department of Agriculture, Forestry and Fisheries
Private Bag X138, Pretoria 0001

Enquiries: Mr Herry Gololo • Tel: +27 12 319 7532 • Fax: +27 12 319 7470 • E-mail: HerryG@daff.gov.za
Reference: 12/11/1/1

Prof B J Mans
ARC-OVI
Onderstepoort

RE: Permission to do research in terms of Section 20 of the ANIMAL DISEASES ACT, 1984 (ACT NO. 35 of 1984)

Dear Prof Mans

Your fax / memo / letter/ Email dated 25 May 2015, requesting permission under Section 20 of the Animal Disease Act, 1984 (Act No. 35 of 1984) to perform a research project or study, refers.

I am pleased to inform you that permission is hereby granted to perform the following research/study, with the following conditions :

Conditions:

1. This permission does not relieve the researcher of any responsibility which may be placed on him by any other act of the Republic of South Africa;
2. No field ticks may be collected from outside of South Africa;
3. Ticks may only be imported after having obtained prior written amendment to this Section 20 approval;
4. Animals used for the feeding of immature field collected ticks for identification purposes, must be euthanized and the carcasses incinerated;
5. If animals used for maintaining clean tick colonies are kept for re-feeding, these animals must be treated with a registered acaracide after each feeding to ensure no risk of environmental contamination with ticks if the animals are moved to open pens inbetween feeding;
6. Theileria positive cattle may not be used for the maintenance of tick colonies;
7. Masking tape must be incorporated at the sides of the white steel containers used for the handling of ticiks, to prevent possible escape of ticks;
8. Vervet monkey control measures to be implemented to prevent access to the stable complex;
9. Ticks may not be outsourced withour prior written approval from DAFF.

Title of research/study: Maintenance of tick colonies
Researcher (s): Prof Mans
Institution: ARC-OVI
Your Ref./ Project Number: AEC 01.15
Our ref Number: 12/11/1/1

Kind regards,

DR. MPHO MAJA
DIRECTOR OF ANIMAL HEALTH

Date: 2015 -10- 19

APPENDIX B

Supplementary files of Chapter 2

Appendix B: Table S1 Summary of the library preparation, sequencing and quality filtering of the sequence data of *R. appendiculatus.*

Dataset	Library preparation (concentration of starting total RNA)	Library preparation (RNA fragmentation time)	Library preparation (number of amplification cycles)	Library preparation (size selection by excision from agarose gel)	Illumina instrument used for sequencing	Sequence read length	Number of raw sequence reads (read 1/ read 2)	Size of raw sequence reads (bp)	Number of quality filtered sequence reads (read 1/ read 2)	Average size of quality filtered sequence reads (bp)	Percentage of reads discarded (read 1/ read 2)
HiSeq 2000 generated sequence reads											
HiSeq	4 ug	8 min	15	± 300 bp	HiSeq 2000	100 x 100	413 323 262/ 413 323 262	100	366 810 605/ 338 340 792	20 - 100	11.3/ 18.1
MiSeq generated sequence reads											
MiSeq SE*	4 ug	8 min	15	± 300 bp	MiSeq	240 (SE)	3 855 867	240	2 961 283	20 - 240	23.2
MiSeq PE*	3.1 ug	3 min	12	± 600 - 1200 bp	MiSeq	250 x 250	13 216 382/ 13 216 382	250	8 781 175/ 6 297 010	20 - 250	33.6/ 52.4
Total MiSeq data							**17 072 249/ 13 216 382**	**150 - 250**	**12 565 276/ 5 474 192**	**20 - 250**	**26.4/58.6**
Total generated sequence reads											
Total sequence data (HiSeq and MiSeq)							430 395 511/ 426 539 644	100 - 250	379 375 881/ 343 814 984	20 - 250	11.9/19.4

* SE = single end sequencing; PE = paired end sequencing.

Appendix B: Table S2 Annotation of *R. appendiculatus* transcripts.

The file can be found using the following link:

https://ars.els-cdn.com/content/image/1-s2.0-S1877959X16300139-mmc3.xlsx

Appendix B: Table S3 Annotation of *R. appendiculatus* predicted proteins.

The file can be found using the following link:

https://ars.els-cdn.com/content/image/1-s2.0-S1877959X16300139-mmc5.xlsx

Appendix B: Table S4 Top expressing transcripts in the *R. appendiculatus* transcriptome.

Expression rank *	Transcript ID	ORF ID	Annotation	TPM value	Percentage of transcriptome
1	c33374_g1_i1	Rapp_Mc2208	Glycine rich superfamily: RIM36	43 988	4.40
2	c53945_g1_i1	No ORF predicted	16S ribosomal RNA	40 496	4.05
3	c43993_g1_i2	Rapp_Mc13679	Glycine rich superfamily	33 290	3.33
4	c15622_g1_i1	Rapp_Mc13680	Unknown function	28 031	2.81
5	c22478_g1_i1	Rapp_Mc950	Glycine rich superfamily	25 459	2.55
6	c53938_g1_i1	Rapp_Mc8886	Lipocalin family: Male-specific histamine-binding salivary protein	21 780	2.18
7	c46457_g2_i1	Rapp_Mc13681	Glycine rich superfamily	20 255	2.03
8	c37026_g1_i1	Rapp_Mc13629	Glycine rich superfamily	16 012	1.60
9	c43993_g1_i1	Rapp_Mc13812	Glycine rich superfamily	15 173	1.52
10	c41649_g1_i1	Rapp_Mc12875	Unknown function	13 758	1.38
11	c41162_g1_i1	Rapp_Mc4548	Lipocalin family: Female-specific histamine-binding protein 1	13 316	1.33
12	c41649_g1_i2	Rapp_Mc12173	Unknown function	13 177	1.32
13	c43993_g1_i3	Rapp_Mc10553	Glycine rich superfamily	13 151	1.32
14	c48158_g1_i1	Rapp_Mc8700	No hit	12 495	1.25
15	c36384_g1_i1	Rapp_Mc13682	Glycine rich superfamily	12 489	1.25
16	c17798_g1_i1	Rapp_Mc774	ML domain: Immunoglobulin G binding protein C	11 353	1.14
17	c39014_g2_i1	Rapp_Mc9768	Glycine rich superfamily	10 864	1.09
18	c36396_g1_i1	Rapp_Mc9443	No hit	9103	0.91
19	c50957_g1_i1	Rapp_Mc13626	Unknown function	8589	0.86
20	c1612_g1_i1	Rapp_Mc13700	Energy metabolism: Cytochrome c oxidase subunit 1	8340	0.83

* Transcripts ranked based on TPM (transcripts per million) value.

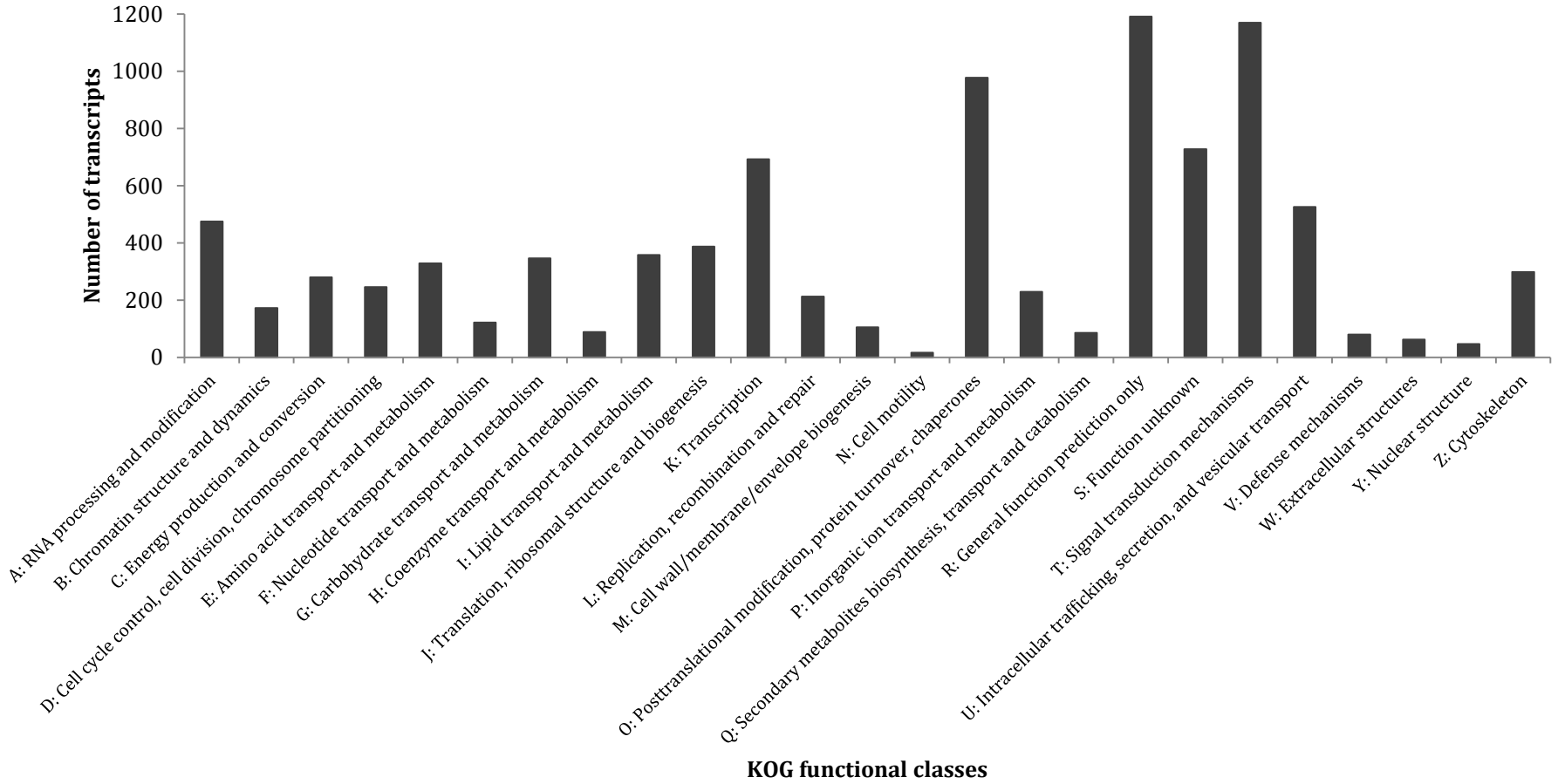
Appendix B: Table S5 Differential expression between female and male ticks in the salivary transcriptome of *R. appendiculatus.*

Protein families	**Female up-regulated**	**Male up-regulated**	$\chi 2$	**p-value**
Secretory protein families	*570*	*553*	*0.26*	*0.6119*
24 kDa family	3	4	0.14	0.7055
28 kDa Metastriate family	18	4	8.91	0.0028
5'-Nucleotidase	3	2	0.20	0.6547
7DB family	0	1	1.00	0.3173
8 kDa Amblyomma family	2	6	2.00	0.1573
8.9 kDa family	45	28	3.96	0.0466
Antigen 5 family	1	2	0.33	0.5637
Astacin	0	2	2.00	0.1573
Basic tail secreted protein	36	21	3.95	0.0469
Bovine pancreatic trypsin inhibitor	61	71	0.76	0.3841
Carbohydrate metabolism	2	0	2.00	0.1573
Carboxypeptidase inhibitor	6	5	0.09	0.7630
Cell motility	0	1	1.00	0.3173
Chitin-binding proteins	0	5	5.00	0.0253
Cystatin	19	15	0.47	0.4927
Cysteine rich	1	1	0.00	1.0000
DA-P36 family	15	2	9.94	0.0016
Defensin	2	5	1.29	0.2568
Dermacentor 9 kDa expansion	0	3	3.00	0.0833
Digestive system (including Serine proteases)	3	25	17.29	< 0.0001*
Evasin	36	16	7.69	0.0055
Fibrinogen-related domain	2	0	2.00	0.1573
Folding, sorting and degradation (including Cathepsins)	1	8	5.44	0.0196
Gluzincin	2	33	27.46	< 0.0001*
Glycan biosynthesis and metabolism	2	0	2.00	0.1573
Glycine rich superfamily	14	31	6.42	0.0113

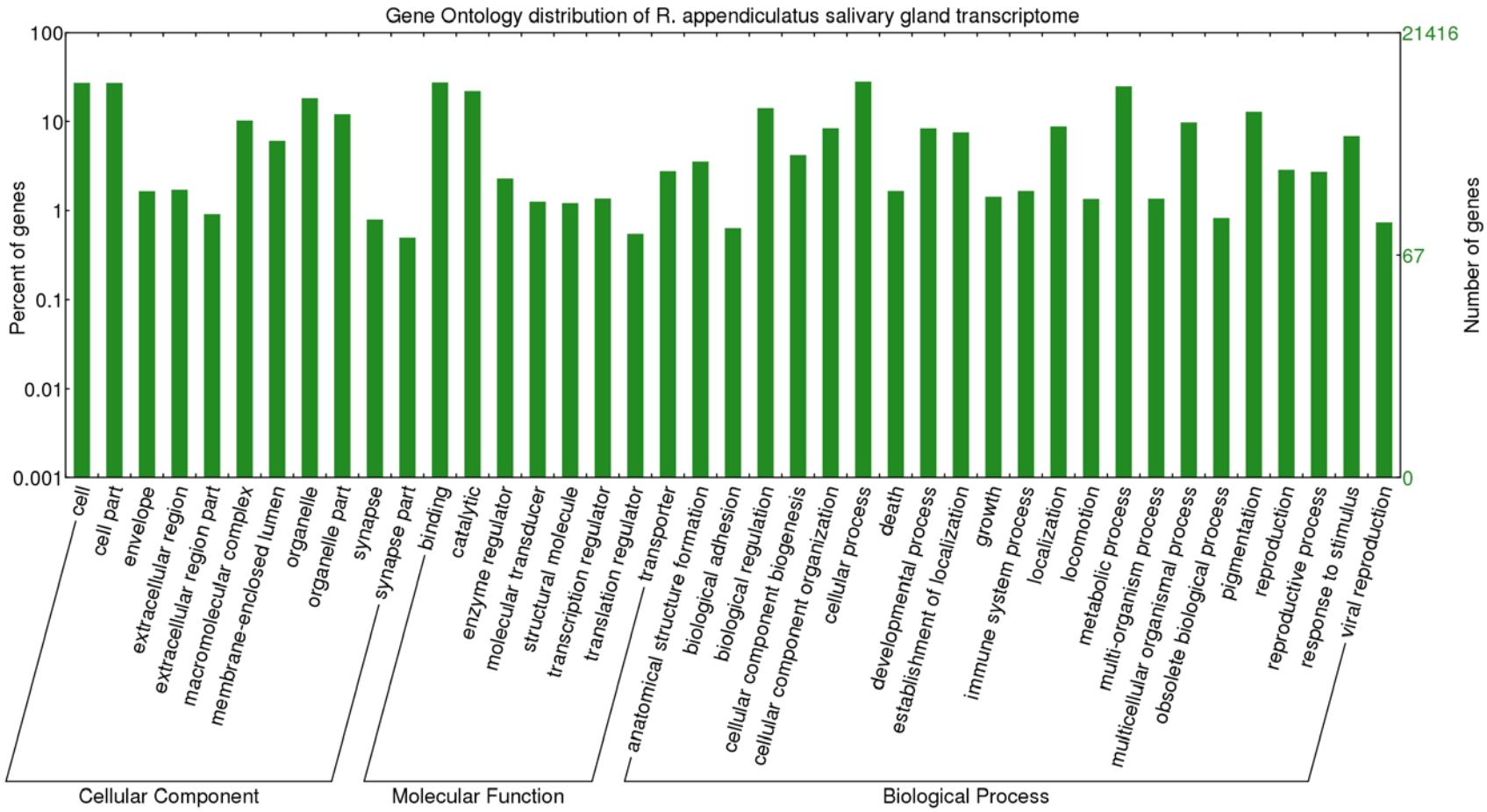
Hirudin	2	0	2.00	0.1573
Histidine rich	3	0	3.00	0.0833
Immunoglobulin G binding protein A	0	6	6.00	0.0143
Ixodegrin B	36	7	19.56	< 0.0001*
Kazal domain	2	1	0.33	0.5637
Kazal/ vWf domain	1	1	0.00	1.0000
Lipid metabolism	7	5	0.33	0.5637
Lipocalin	157	154	0.03	0.8649
Microplusin	3	5	0.50	0.4795
ML domain	1	8	5.44	0.0196
Mucin	11	3	4.57	0.0325
No hit	0	1	1.00	0.3173
One of each family	17	1	14.22	0.0002*
Phospholipase A2	0	2	2.00	0.1573
Reprolysin	27	15	3.43	0.0641
SALP15	0	1	1.00	0.3173
Secretory - unknown function	0	2	2.00	0.1573
Serpin	3	5	0.50	0.4795
Signal transduction	1	0	1.00	0.3173
Signaling molecules and interaction	3	0	3.00	0.0833
Sphingomyelinase	0	5	5.00	0.0253
TIL domain	20	34	3.63	0.0568
Transport and catabolism	2	6	2.00	0.1573
Housekeeping protein class	*220*	*413*	*58.85*	*< 0.0001**
Unknown function protein class	*130*	*288*	*59.72*	*< 0.0001**
No hit protein class	*111*	*172*	*13.15*	*0.0003**
Transcripts without predicted ORFs	*727*	*920*	*22.62*	*< 0.0001**
Total	**1758**	**2346**	**84.25**	**< 0.0001***

Transcripts were estimated as up-regulated [fold change > 2, false discovery rate (FDR) p-value of < 0.001] by edgeR (Empirical analysis of digital gene expression data in R). Up-regulated transcripts were compared between female and male ticks using Chi-square. $\chi 2$ - and p-values are indicated (df = 1).

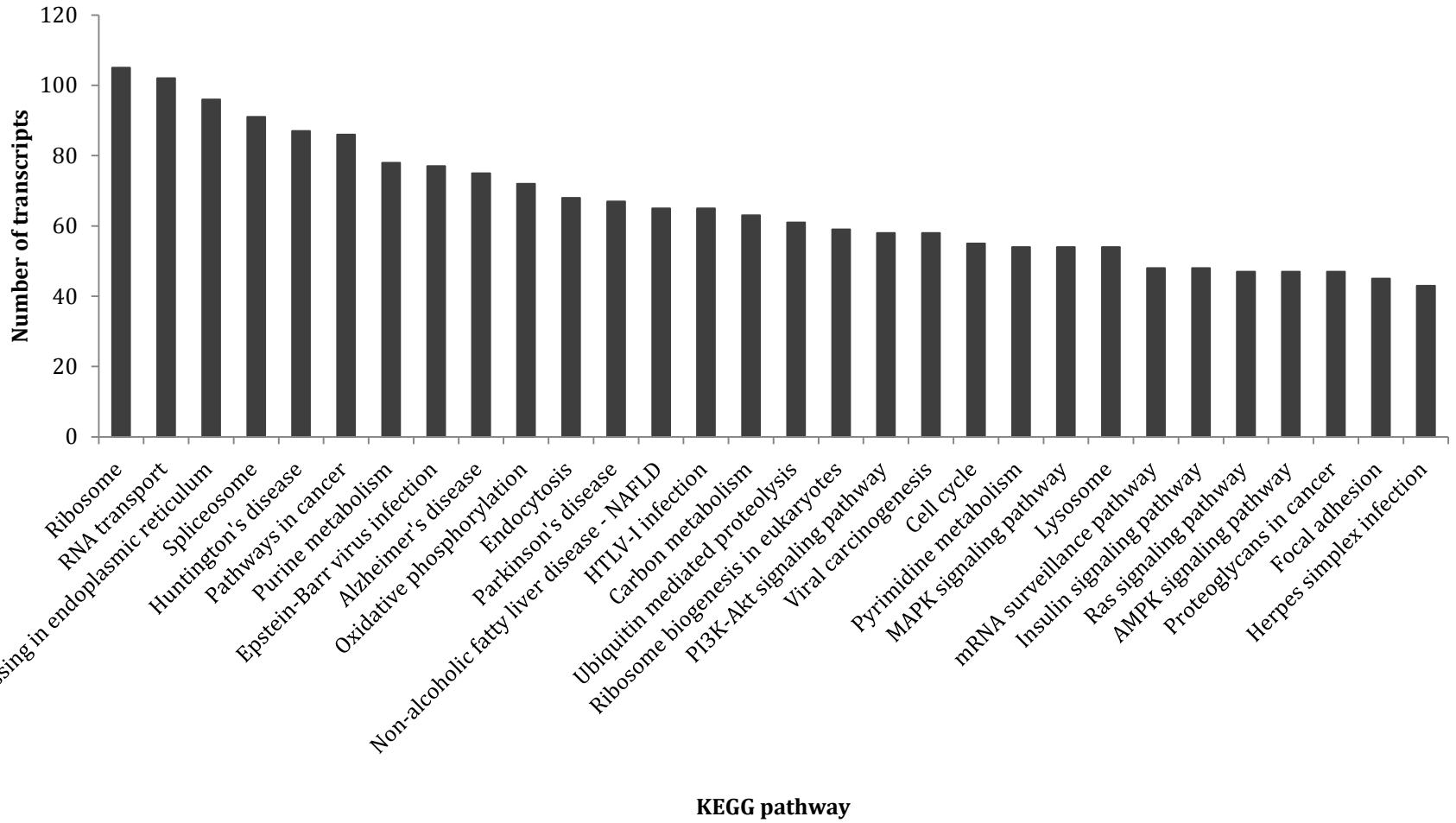
* Significant Chi-square test (Bonferroni corrected p-value < 0.0010).



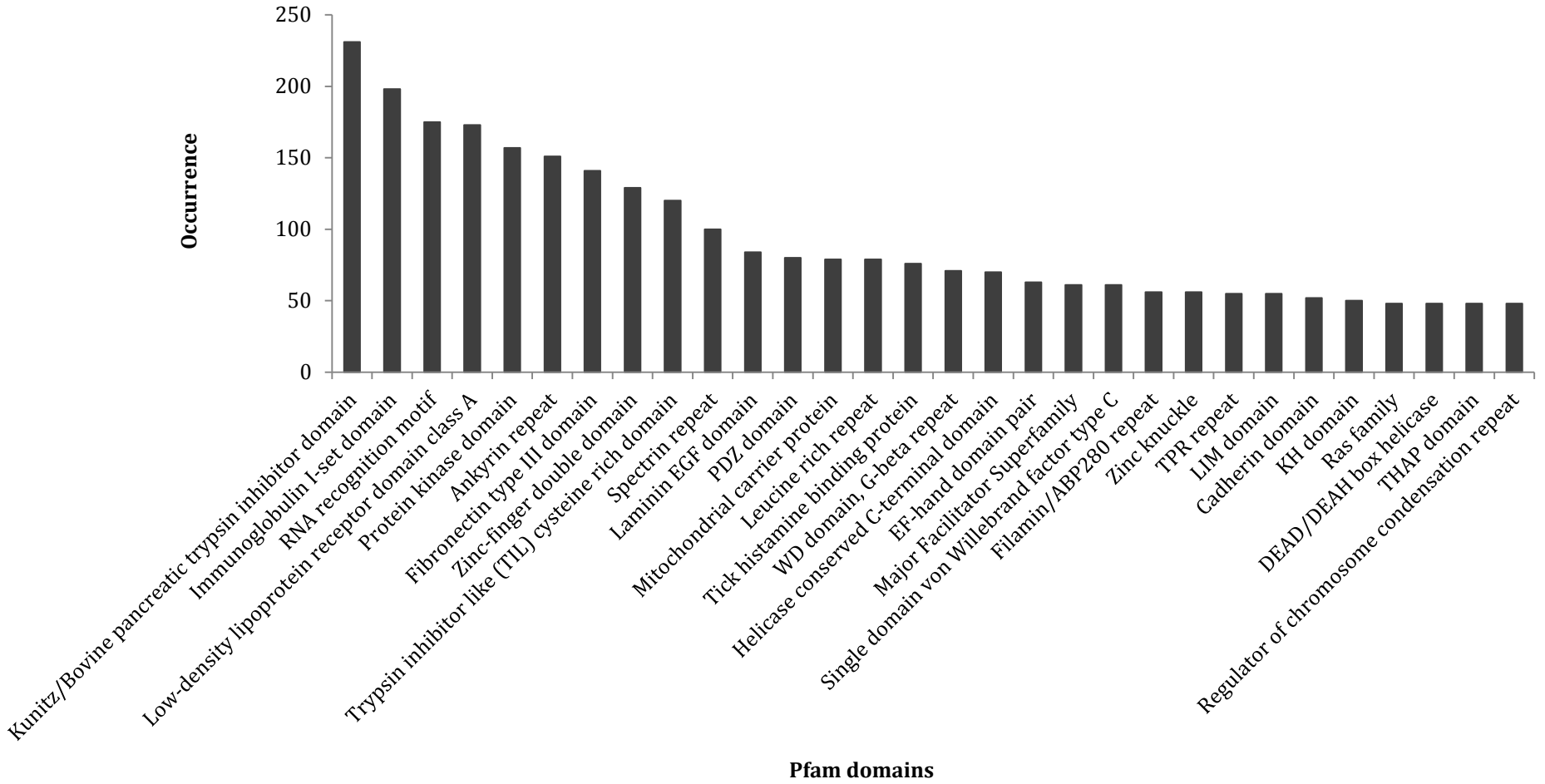
Appendix B: Figure S1 KOG functional classification of *R. appendiculatus* transcripts. Number of transcripts with a significant BLASTx result (E-value < E-05) in each functional Eukaryotic Clusters of Orthologs (KOG) class. In total, 8282 transcripts obtained a significant BLASTx match and could be assigned to 25 KOG functional categories, with 3791 unique KOG terms.



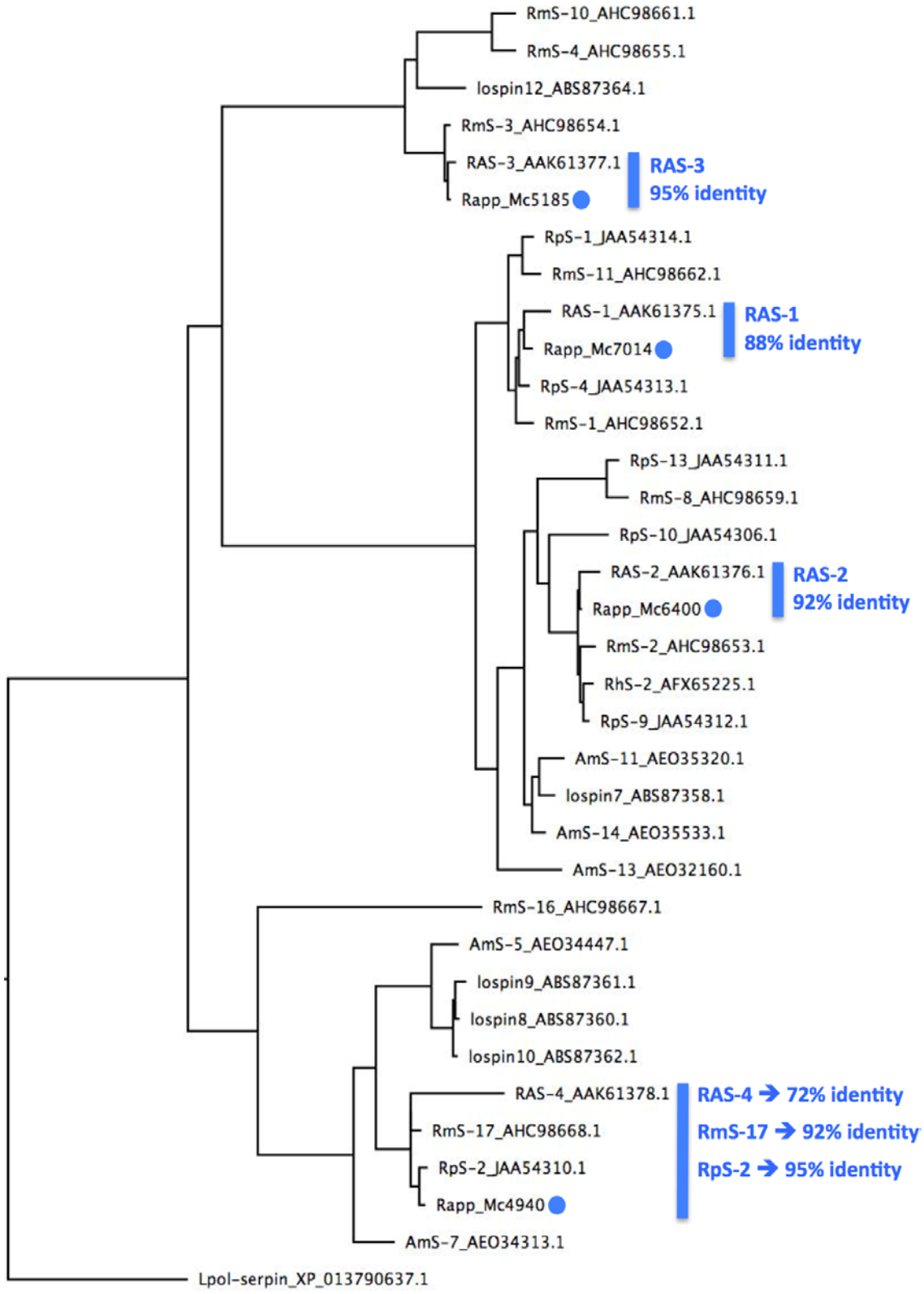
Appendix B: Figure S2 Gene Ontology distribution of the *R. appendiculatus* transcriptome. Proportion of transcripts classified into categories of biological processes, molecular functions and cellular components are indicated. Web Gene Ontology Annotation Plot (WEGO) was used to plot level two GO terms present in at least 100 transcripts.



Appendix B: Figure S3 Identified KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways in the transcriptome of *R. appendiculatus*. In total, 4647 transcripts were allocated to 321 *I. scapularis* KEGG pathways. The top 30 most abundant pathways are shown.



Appendix B: Figure S4 Pfam domains identified in the predicted proteins of the *R. appendiculatus* transcriptome. In total, 13 246 Pfam domains were identified, categorising 7630 of the *R. appendiculatus* proteins. The 30 most occurring Pfam domains are indicated.



Appendix B: Figure S5 Serine proteinase inhibitor (serpin) maximum likelihood phylogeny indicating the clustering of the RAS-1 to RAS-4 orthologues. The phylogeny was inferred from a subset of sequences from a previous serpin phylogeny (Tirloni *et al.*, 2014b). Additionally, serpin B6-like from *Limulus polyphemus*

(XP_013790637.1) was set as outgroup. Ultrafast bootstrap with 1000 replicates was performed and values above 60% indicated. Blue circles denote the orthologues assembled in the *R. appendiculatus* transcriptome of the previously characterised *R. appendiculatus* serpins (RAS-1 to RAS-4) (Mulenga *et al.*, 2003d). Also indicated in blue are the orthologous clusters and the protein identities to the respective proteins. Rapp_Mc7014, Rapp_Mc6400 and Rapp_Mc5185 clustered with to their respective homologs (RAS-1 to RAS-3) and Rapp_Mc4940 clustered together with RAS-4, RpS-2 and RmS-17. RmS, *R. microplus* serpins (Tirloni *et al.*, 2014b); RpS, *R. pulchellus* serpins (Tan *et al.*, 2015a); RhS, *R. haemaphysaloides* serpins (Yu *et al.*, 2013); AmS, *A. maculatum* serpins (Karim *et al.*, 2011); lospin, Lone Star tick serpins of *A. americanum* (Mulenga *et al.*, 2007).

APPENDIX C

Supplementary files of Chapter 3

Appendix C: Table S1 *Rhipicephalus zambeziensis* library preparation specifications and the sequence reads pre- and post-quality filtering.

Dataset	Library preparation (concentration of starting total RNA)	Library preparation (RNA fragmentation time)	Library preparation (number of amplification cycles)	Library preparation (size selection by excision from agarose gel)	Illumina instrument used for sequencing	Number of raw sequence reads (read 1/ read 2)	Size of raw sequence reads (bp)	Number of quality filtered paired end sequence reads (read 1/ read 2)	Number of quality filtered single end sequence reads	Size range of quality filtered sequence reads (bp)	Percentage of reads retained (paired end formation/ single end formation)
HiScanSQ generated sequence reads											
F0: Female - unfed	2 ug	8 min	12	± 300 bp	HiScanSQ	32 027 583/ 32 027 583	100	25 391 810/ 25 391 810	5 007 301	50 - 100	79.3/ 7.8
F3: Female - 3 days feeding	2 ug	8 min	12	± 300 bp	HiScanSQ	32 142 442/ 32 142 442	100	25 321 431/ 25 321 431	5 123 605	50 - 100	78.8/ 8.0
F5: Female - 5 days feeding	2 ug	8 min	12	± 300 bp	HiScanSQ	37 528 275/ 37 528 275	100	29 918 808/ 29 918 808	5 786 823	50 - 100	79.7/ 7.7
M0: Male - unfed	2 ug	8 min	12	± 300 bp	HiScanSQ	27 591 382/ 27 591 382	100	23 335 530/ 23 335 530	3 236 827	50 - 100	84.6/ 5.9
M3: Male - 3 days feeding	2 ug	8 min	12	± 300 bp	HiScanSQ	26 047 465/ 26 047 465	100	21 694 635/ 21 694 635	3 232 318	50 - 100	83.3/ 6.2
M5: Male - 5 days feeding	2 ug	8 min	12	± 300 bp	HiScanSQ	22 145 769/ 22 145 769	100	18 518 042/ 18 518 042	2 702 203	50 - 100	83.6/ 6.1
Total HiScanSQ sequence data						**177 482 916/ 177 482 916**	**100**	**144 180 256/ 144 180 256**	**25 089 077**	**50 - 100**	**81.2/ 7.1**

MiSeq generated sequence reads											
Equimolar mix of all six samples	2 ug	3 min[a]	12	± 600 - 1000 bp	MiSeq	22 653 340/ 22 653 340	300	6 901 796/ 6 901 796	15 631 696	50 - 300	30.5/ 34.5
Merging of paired end MiSeq reads[b]						6 901 796/ 6 901 796	50 - 300	2 254 236/ 2 254 236	4 647 560	50 - 580	32.7/ 67.3
Total MiSeq sequence data						**22 653 340/ 22 653 340**	**300**	**2 254 236/ 2 254 236**	**20 279 256**	**50 - 580**	**10.0/ 44.8**
Total generated sequence reads											
Total sequence data						**200 136 256/ 200 136 256**	**100 - 300**	**146 434 492/ 146 434 492**	**45 368 333**	**50 - 580**	**73.2/ 11.3**

[a] Varying RNA fragmentation time was used during the preparation of the MiSeq sequencing library to facilitate the generation of larger fragments for sequencing on the longer read Miseq instrument.

[b] Due to the long Miseq sequencing reads, many of the paired end sequences overlapped and were merged into a single read when an overlap of 20 bp was observed.

Appendix C: Table S2 Annotation of the *R. zambeziensis* transcriptome.

The file can be found using the following link:

https://static-content.springer.com/esm/art%3A10.1186%2Fs13071-017-2312-4/MediaObjects/13071_2017_2312_MOESM2_ESM.xlsx

Appendix C: Table S3 Annotation of the *R. zambeziensis* predicted proteins.

The file can be found using the following link:

https://static-content.springer.com/esm/art%3A10.1186%2Fs13071-017-2312-4/MediaObjects/13071_2017_2312_MOESM4_ESM.xlsx

Appendix C: Table S4 Putative orthologues in the *R. zambeziensis* transcriptome of previously characterised proteins.

Protein name	Protein description	Accession number	*R. zambeziensis* Protein ID	Identity (%)	Full-length[b]	Combined TPM[d]	Female TPM[d]	Male TPM[d]
IGBP-MA	Immunoglobulin G binding protein - Male A	AAB68801.1	Rzam_Mc5608	97	Complete	2113.1	1.1	4076.8
IGBP-MB	Immunoglobulin G binding protein - Male B	AAB68802.1	Rzam_Mc3335	92	Complete	2588.6	1.1	5065.3
IGBP-MC	Immunoglobulin G binding protein - Male C	AAB68803.1	Rzam_Mc3213	98	Complete	4434.2	1.5	8720.0
HBP1	Female-specific histamine-binding protein 1	O77420	Rzam_Mc9434	91	Complete	250.3	512.5	0
HBP2	Female-specific histamine-binding protein 2	O77421	Rzam_Mc5492	75	Complete	11.6	24.0	0
HBPM	Male-specific histamine-binding salivary protein	O77422	Rzam_Mc12946	42	Complete	24.4	0	47.4
RIM36	*Rhipicephalus* immuno-dominant molecule 36	AAK98794.1	Rzam_Mc6473	98	Fragment[c]	10 590.3	5205.0	15 850.2
			Rzam_Mc186	96	Fragment[c]	6645.5	2908.9	10 290.5
64P	Salivary gland-associated protein 64P	AAM09648.1	Rzam_Mc354	85	Complete	497.5	95.3	884.3
RAS-1	*R. appendiculatus* serine proteinase inhibitor serpin-1	AAK61375.1	Rzam_Mc7211	90	Complete	15.6	7.0	23.7
RAS-2	*R. appendiculatus* serine proteinase inhibitor serpin-2	AAK61376.1	Rzam_Mc9158	90	Fragment	7.8	9.7	6.1
RAS-3	*R. appendiculatus* serine proteinase inhibitor serpin-3	AAK61377.1	Rzam_Mc8493	94	Complete	39.7	69.9	10.3
RAS-4	*R. appendiculatus* serine proteinase inhibitor serpin-4	AAK61378.1	Rzam_Mc5334	71	Complete	14.8	29.4	0.6
TdP1	Tryptase inhibitor precursor	AAW32666.1	Rzam_Mc366	34	Complete	49.5	92.1	9.4
Ra-KLP	*R. appendiculatus* Kunitz/BPTI-like protein	ACM86785.1	Rzam_Mc262	86	Complete	4.4	9.1	0

JL-RA1	Japanin-like-RA1 precursor	AGF70151.1	Rzam_Mc2302	90	Complete	1.8	3.6	0.1
JL-RA2	Japanin-like-RA2 precursor	AGF70152.1	Rzam_Mc3898	58	Complete	1.7	2.1	1.4
Japanin	Japanin precursor	AGF70149.1	Rzam_Mc597	32	Complete	3.2	6.6	0

[a] Referencing based on numbering in manuscript.

[b] Full-length as classified by containing a predicted start and stop codon in the deduced amino acid sequence.

[c] RIM36 was assembled into two separate transcripts, each coding for its own fragmented open reading frame.

[d] Expression was estimated as TPM (transcripts per million).

Protein references: IGBP-MA (AAB68801.1), IGBP-MB (AAB68802.1) and IGBP-MC (AAB68803.1) (Wang and Nuttall, 1995); HBP1 (O77420), HBP2 (O77421) and HBPM (O77422) (Paesen *et al.*, 1999); RIM36 (AAK98794.1) (Bishop *et al.*, 2002); 64P (AAM09648.1) (Trimnell *et al.*, 2002); RAS-1 (AAK61375.1), RAS-2 (AAK61376.1), RAS-3 (AAK61377.1) and RAS-4 (AAK61378.1) (Mulenga *et al.*, 2003d); TdP1 (AAW32666.1) (Paesen *et al.*, 2007); Ra-KLP (ACM86785.1) (Paesen *et al.*, 2009); JL-RA1 (AGF70151.1), JL-RA2 (AGF70152.1) and Japanin (AGF70149.1) (Preston *et al.*, 2013).

Appendix C: Table S5 Proportions of the most abundant secretory protein families in the *R. zambeziensis* transcriptome during feeding.

Secretory protein families	Female day 0 (TPM)	Female day 0 (%)	Female day 3 (TPM)	Female day 3 (%)	Female day 5 (TPM)	Female day 5 (%)	Male day 0 (TPM)	Male day 0 (%)	Male day 3 (TPM)	Male day 3 (%)	Male day 5 (TPM)	Male day 5 (%)
Glycine rich superfamily	22 529.8	35.66	425 477.1	75.20	47 415.7	18.00	86 953.4	67.12	526 364.9	84.26	411 893.1	71.77
Histamine release factor	23 926.9	37.88	705.4	0.12	2830.4	1.07	17 073.3	13.18	538.3	0.09	792.2	0.14
Lipocalin	2758.9	4.37	13 092.3	2.31	68 973.3	26.18	5059.6	3.91	31 440.1	5.03	15 824.5	2.76
Transport and catabolism	1768.9	2.80	327.6	0.06	1052.2	0.40	1684.1	1.30	318.1	0.05	654.1	0.11
Mucin	1499.0	2.37	2301.2	0.41	8328.5	3.16	1600.8	1.24	1421.9	0.23	3041.1	0.53
Bovine pancreatic trypsin inhibitor	1155.3	1.83	7072.8	1.25	27 400.5	10.40	2808.4	2.17	12 669.7	2.03	10 024.5	1.75
Reprolysin	1093.6	1.73	5051.5	0.89	9538.2	3.62	1256.0	0.97	2436.6	0.39	4853.2	0.85
Folding. sorting and degradation (including Cathepsins)	1076.0	1.70	1367.4	0.24	4347.2	1.65	1438.8	1.11	2151.0	0.34	13 834.8	2.41
Basic tail secreted protein	984.9	1.56	6464.1	1.14	13 846.5	5.26	1811.5	1.40	2425.1	0.39	2953.6	0.51
24 kDa family	848.4	1.34	645.1	0.11	200.7	0.08	934.6	0.72	321.1	0.05	373.1	0.07
TIL domain	597.1	0.95	9405.6	1.66	2612.1	0.99	1982.0	1.53	5756.7	0.92	16 662.8	2.90
Secretory - unknown function	524.1	0.83	66 351.4	11.73	8185.8	3.11	1683.1	1.30	13 897.2	2.22	13 318.4	2.32

Defensin	7.3	0.01	11 841.0	2.09	3973.6	1.51	83.8	0.06	3099.7	0.50	1694.7	0.30
8.9 kDa family	274.6	0.43	4540.0	0.80	30 232.9	11.47	470.5	0.36	3131.9	0.50	3513.3	0.61
Ixodegrin B	148.1	0.23	1629.0	0.29	8695.3	3.30	209.2	0.16	253.7	0.04	282.7	0.05
28 kDa Metastriate family	137.9	0.22	662.9	0.12	7921.2	3.01	141.9	0.11	1547.3	0.25	559.2	0.10
ML domain	126.2	0.20	38.7	0.01	143.4	0.05	118.8	0.09	4300.9	0.69	36 254.4	6.32
Immunoglobulin G binding protein A	0.0	0.00	5.2	0.00	1.2	0.00	2.5	0.00	4487.0	0.72	23 233.5	4.05
Other secretory proteins	3716.1	5.88	8822.4	1.56	17 780.8	6.75	4236.9	3.27	8107.6	1.30	14 176.4	2.47
Total secretory protein portion	63 173.1	100	565 800.4	100	263 479.3	100	129 549.1	100	624 668.7	100	573 939.3	100

The expression level (measured in transcripts per million, TPM) and proportion (%) of each secretory protein family was estimated per time point. Proportions were visual represented in Figure 3.3 of Chapter 3.

Appendix C: Table S6 Comparison between the differential expression of *R. zambeziensis* female and male ticks.

Protein classes/ families	**Female up-regulated**	**Male up-regulated**	***χ2***	**p-value**
Secretory protein class	*376*	*259*	*21.89*	< 0.0001*
Lipocalin	140	42	52.77	< 0.0001*
Digestive system (including Serine proteases)	0	41	41.00	< 0.0001*
Cystatin	0	26	26.00	< 0.0001*
Reprolysin	27	3	19.20	< 0.0001*
TIL domain	4	29	18.94	< 0.0001*
28 kDa Metastriate family	23	2	17.64	< 0.0001*
Folding, sorting and degradation (including Cathepsins)	1	17	14.22	0.0002*
7DB family	0	13	13.00	0.0003*
8.9 kDa family	36	12	12.00	0.0005*
DA-P36 family	12	1	9.31	0.0023
Bovine pancreatic trypsin inhibitor	37	17	7.41	0.0065
Ixodegrin B	10	1	7.36	0.0067
One of each family	7	0	7.00	0.0082
Glycine rich superfamily	18	6	6.00	0.0143

Gluzincin	8	1	5.44	0.0196
Immunoglobulin G binding protein A	0	5	5.00	0.0253
ML domain	0	4	4.00	0.0455
Mucin	0	4	4.00	0.0455
Hirudin	3	0	4.00	0.0455
Chitin-binding proteins	0	3	3.00	0.0833
Carboxypeptidase inhibitor	3	0	3.00	0.0833
Evasin	12	5	2.88	0.0896
Antigen 5 family	0	2	2.00	0.1573
Basic tail secreted protein	13	8	1.19	0.2752
5’-Nucleotidase	1	3	1.00	0.3173
8 kDa Amblyomma family	0	1	1.00	0.3173
Sphingomyelinase	0	1	1.00	0.3173
24 kDa family	1	0	1.00	0.3173
Dermacentor 9 kDa expansion	1	0	1.00	0.3173
Glycan biosynthesis and metabolism	1	0	1.00	0.3173
Kazal domain	1	0	1.00	0.3173

Lipid metabolism	1	0	1.00	0.3173
Serine/ threonine protein kinase	1	0	1.00	0.3173
Secretory - unknown function	9	6	0.60	0.4386
Transport and catabolism	1	2	0.33	0.5637
Defensin	2	1	0.33	0.5637
Microplusin	2	2	0.00	1.0
Serpin	1	1	0.00	1.0
Housekeeping protein class	*87*	*69*	*2.08*	*0.1495*
Unknown function protein class	*56*	*91*	*8.33*	*0.0039*
No hit protein class	*21*	*26*	*0.53*	*0.4658*
Transcripts without predicted ORFs	*101*	*123*	*2.16*	*0.1416*
Total	**642**	**568**	**4.53**	**0.0334**

Differential expression analysis using edgeR (Empirical analysis of digital gene expression data in R) software package (parameters: fixed dispersion of 0.4, FDR p-value of < 0.01 and fold change of > 4) was performed.

Chi-square test was performed to compare differences between female and male ticks. $\chi 2$ - and p-values are indicated. $df = 1$.

* Significant Chi-square test (Bonferroni corrected p-value < 0.0013).

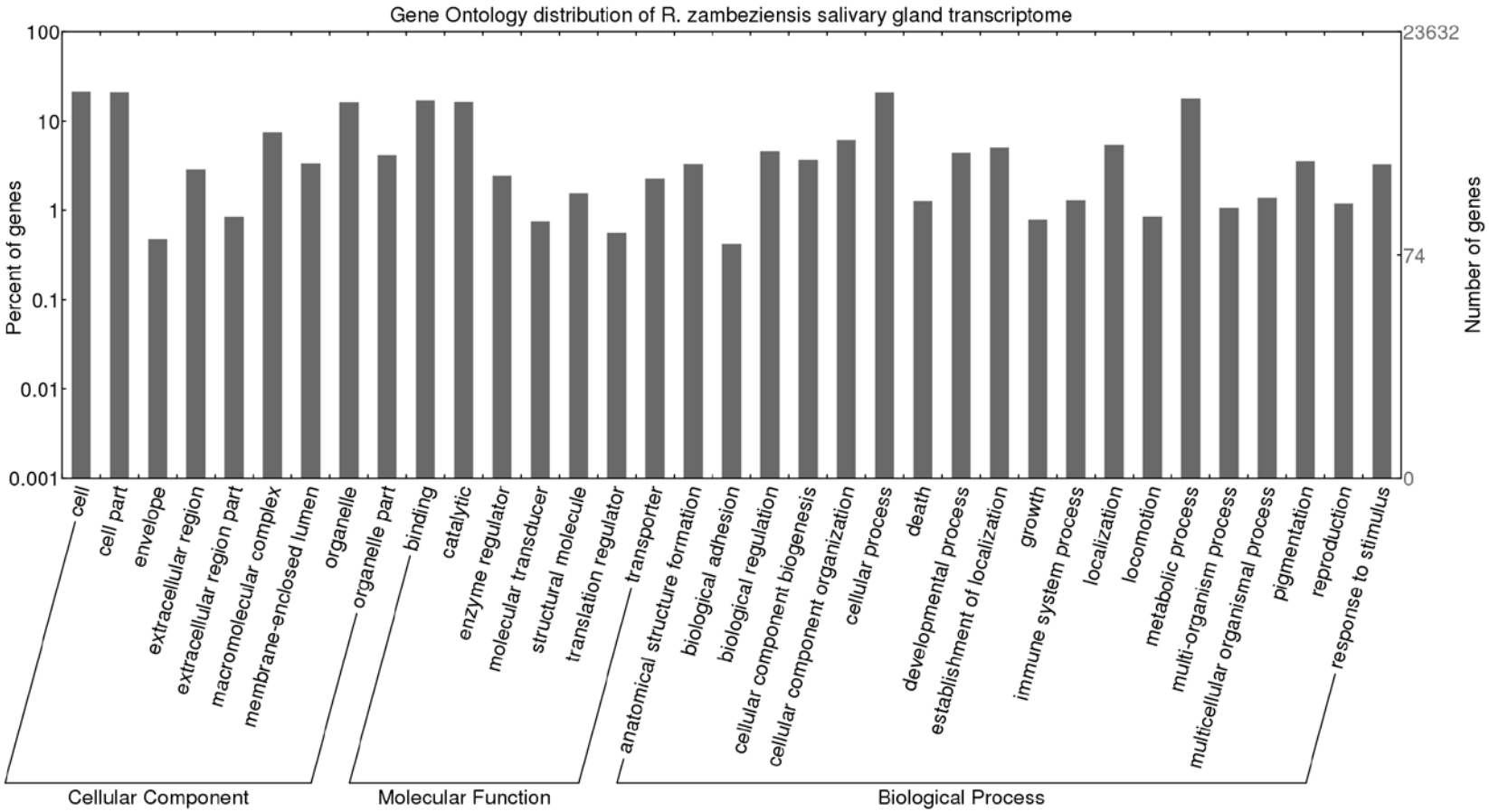
Appendix C: Table S7 Differentially expressed transcripts in *R. zambeziensis* of the different protein classes and families during feeding.

Protein classes/ families	**F0 vs. F3 (F0 Up)**	**F0 vs. F3 (F3 Up)**	**F0 vs. F5 (F0 Up)**	**F0 vs. F5 (F5 Up)**	**F3 vs. F5 (F3 Up)**	**F3 vs. F5 (F5 Up)**	**M0 vs. M3 (M0 Up)**	**M0 vs. M3 (M3 Up)**	**M0 vs. M5 (M0 Up)**	**M0 vs. M5 (M5 Up)**	**M3 vs. M5 (M3 Up)**	**M3 vs. M5 (M5 Up)**
Secretory protein class	*15*	*541*	*111*	*522*	*411*	*103*	*4*	*335*	*4*	*367*	*0*	*2*
Lipocalin	1	134	21	150	108	47	0	72	0	50	0	1
Glycine rich superfamily	0	61	1	17	60	1	0	21	0	17	0	0
Reprolysin	1	55	18	70	42	4	1	20	0	26	0	0
Bovine pancreatic trypsin inhibitor	3	51	31	64	53	7	0	36	0	24	0	0
TIL domain	0	38	4	11	36	2	0	34	0	37	0	0
8.9 kDa family	0	33	1	34	13	8	0	15	0	16	0	0
Basic tail secreted protein	0	30	4	29	12	4	0	14	0	16	0	0
28 kDa Metastriate family	1	22	3	30	6	5	0	3	0	5	0	0
Gluzincin	0	5	5	16	4	8	0	1	0	2	0	0
Evasin	0	17	0	17	5	2	0	6	0	7	0	0
Secretory - unknown function	1	12	3	8	10	1	0	8	0	10	0	0
Ixodegrin B	0	10	2	2	8	0	0	2	0	2	0	0
Carboxypeptidase inhibitor	0	10	0	2	12	0	0	4	0	3	0	0
Defensin	0	9	0	7	3	0	0	9	0	6	0	0
Mucin	0	4	1	1	11	0	1	4	0	4	0	0
DA-P36 family	1	3	2	12	1	7	0	1	0	1	0	0
One of each family	0	1	0	9	0	3	0	0	0	0	0	0
Serpin	0	1	0	7	0	0	0	1	0	5	0	0
Hirudin	0	3	0	3	0	0	0	0	0	0	0	0
Thyropin	0	3	0	4	0	0	0	0	0	0	0	0

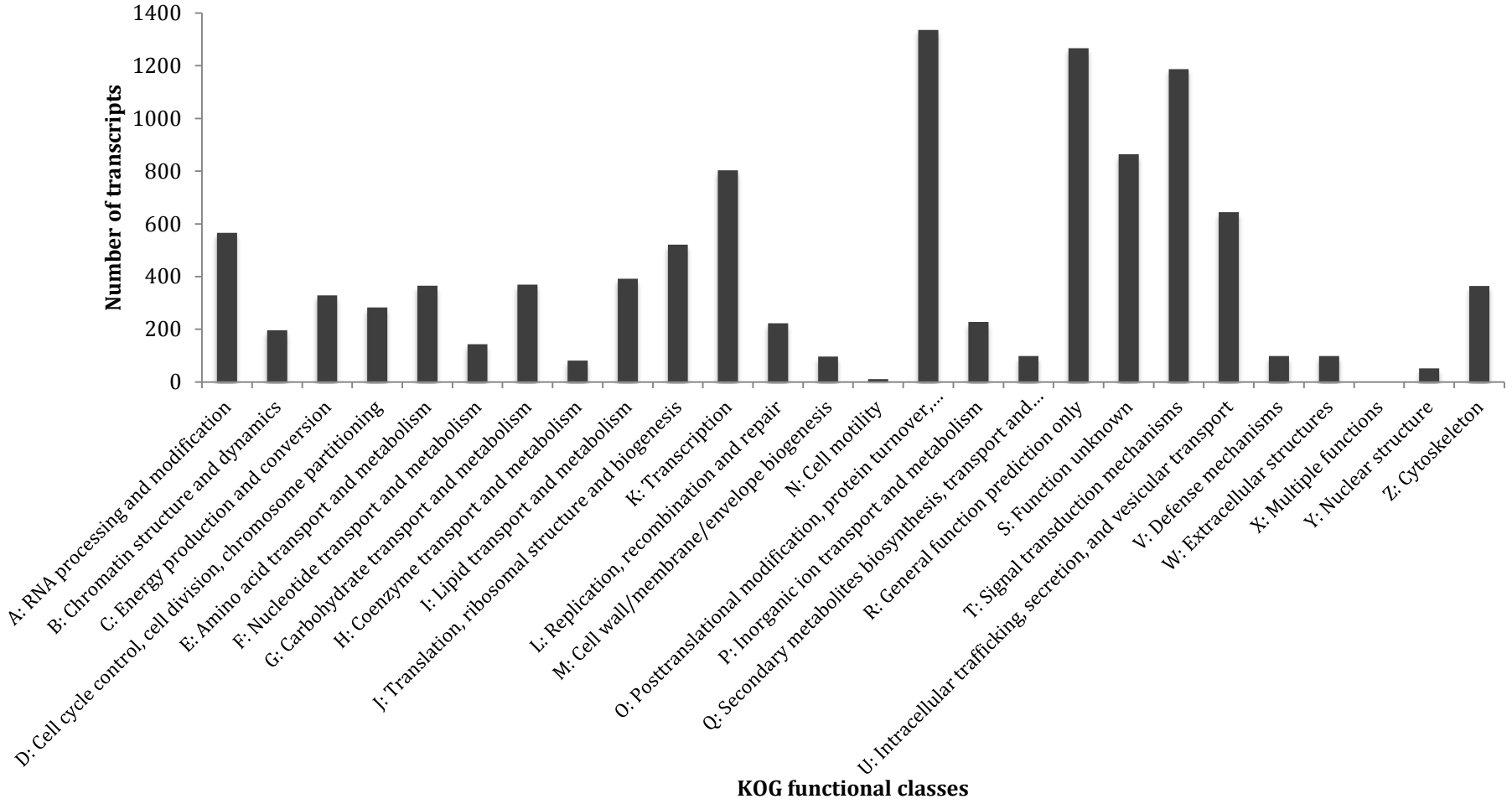
Glycan biosynthesis and metabolism	1	3	1	3	0	0	0	0	1	0	0	0
Dermacentor 9 kDa expansion	0	1	0	2	2	1	0	2	0	2	0	0
Fibrinogen-related domain	0	6	0	2	1	0	0	4	0	4	0	0
Folding, sorting and degradation (including Cathepsins)	1	0	0	2	3	0	1	16	1	17	0	0
Kazal domain	0	0	0	2	0	1	0	0	0	0	0	0
Lipid metabolism	0	1	0	2	1	0	0	0	0	0	0	0
Microplusin	0	3	1	2	0	0	0	2	0	1	0	0
SALP15/ Ixostatin	0	0	0	2	0	0	0	1	0	1	0	0
24 kDa family	0	3	5	1	1	0	0	0	0	2	0	0
5’-Nucleotidase	0	0	1	1	3	0	0	0	0	0	0	0
Antigen 5 family	0	2	0	1	1	0	0	4	0	4	0	0
Chitin-binding proteins	0	0	0	1	0	0	0	6	0	6	0	0
Cystatin	0	4	0	1	4	0	0	14	0	26	0	0
Kazal/ SPARC domain	0	1	0	1	0	0	0	0	0	0	0	0
ML domain	0	0	0	1	0	0	0	3	0	4	0	0
Phospholipase A2	0	3	0	1	2	0	0	1	0	1	0	0
Signal transduction	0	0	0	1	0	0	0	0	0	0	0	0
Serine/ threonine protein kinase	0	0	0	1	0	1	0	0	0	0	0	0
Transport and catabolism	0	1	0	1	0	0	0	2	0	2	0	0
Translation	1	1	1	1	0	0	0	0	0	0	0	0
8 kDa Amblyomma family	0	3	0	0	6	0	0	3	0	3	0	0
Digestive system (including Serine proteases)	0	4	1	0	3	0	0	12	0	41	0	1

Immunoglobulin G binding protein A	0	1	0	0	0	0	0	5	0	5	0	0
14 kDa family	0	1	0	0	0	0	0	0	0	0	0	0
Madanin	0	1	0	0	0	0	0	1	0	1	0	0
Metalloprotease	1	0	0	0	0	0	0	0	1	0	0	0
Sphingomyelinase	2	0	5	0	0	0	1	1	0	1	0	0
Transcription	1	0	0	0	0	1	0	0	1	0	0	0
7DB family	0	0	0	0	0	0	0	7	0	15	0	0
Housekeeping protein class	*53*	*126*	*79*	*219*	*66*	*36*	*47*	*65*	*23*	*72*	*0*	*1*
Unknown function protein class	*8*	*95*	*38*	*64*	*126*	*11*	*2*	*95*	*3*	*98*	*0*	*2*
No hit protein class	*4*	*29*	*13*	*26*	*26*	*5*	*3*	*32*	*1*	*28*	*0*	*0*
Transcripts without predicted ORFs	*23*	*136*	*55*	*149*	*110*	*29*	*7*	*96*	*5*	*116*	*0*	*0*
Total	**103**	**927**	**296**	**980**	**739**	**184**	**63**	**623**	**36**	**681**	**0**	**5**

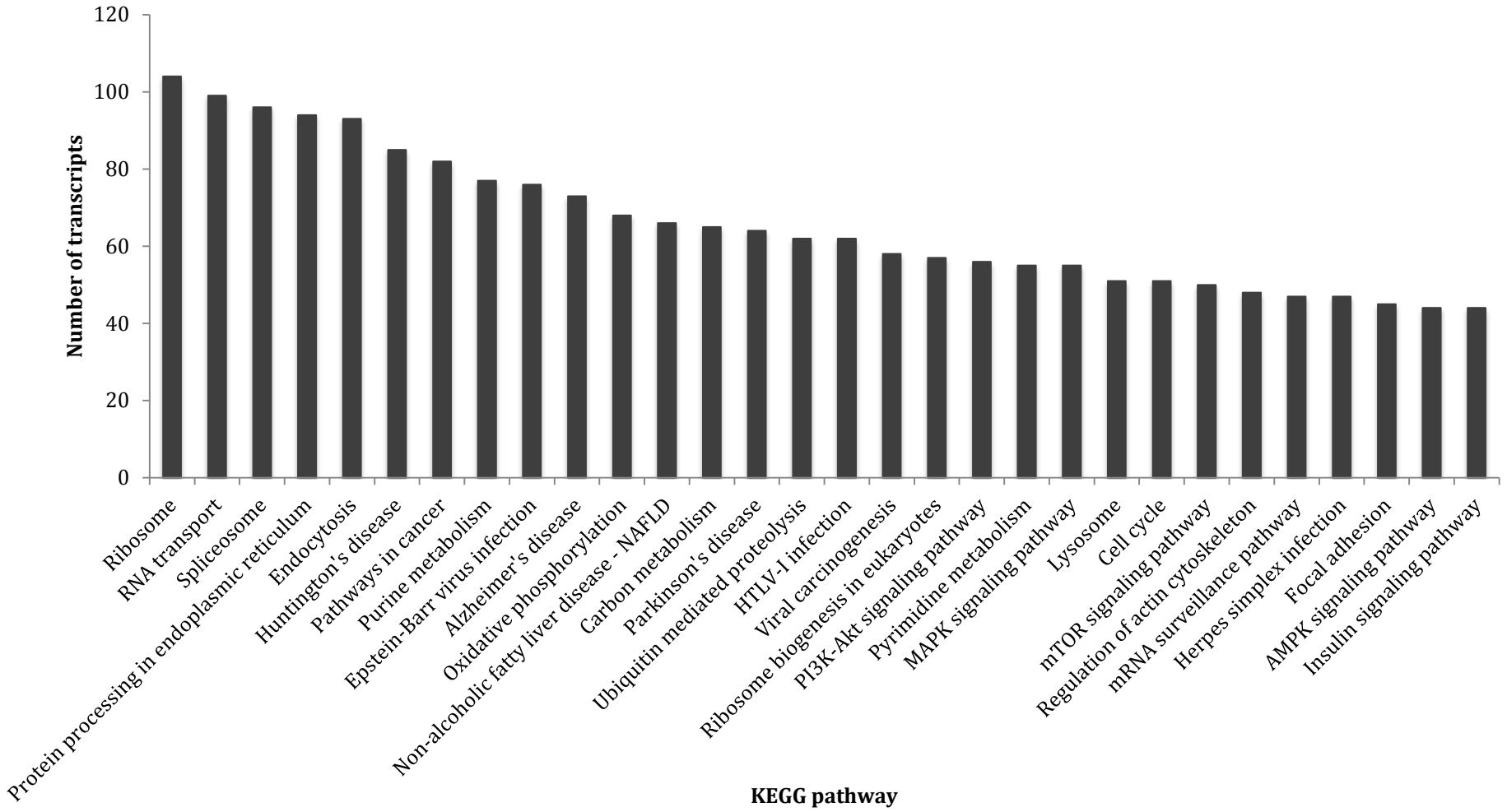
The edgeR (Empirical analysis of digital gene expression data in R) software package was used for differential expression analysis.



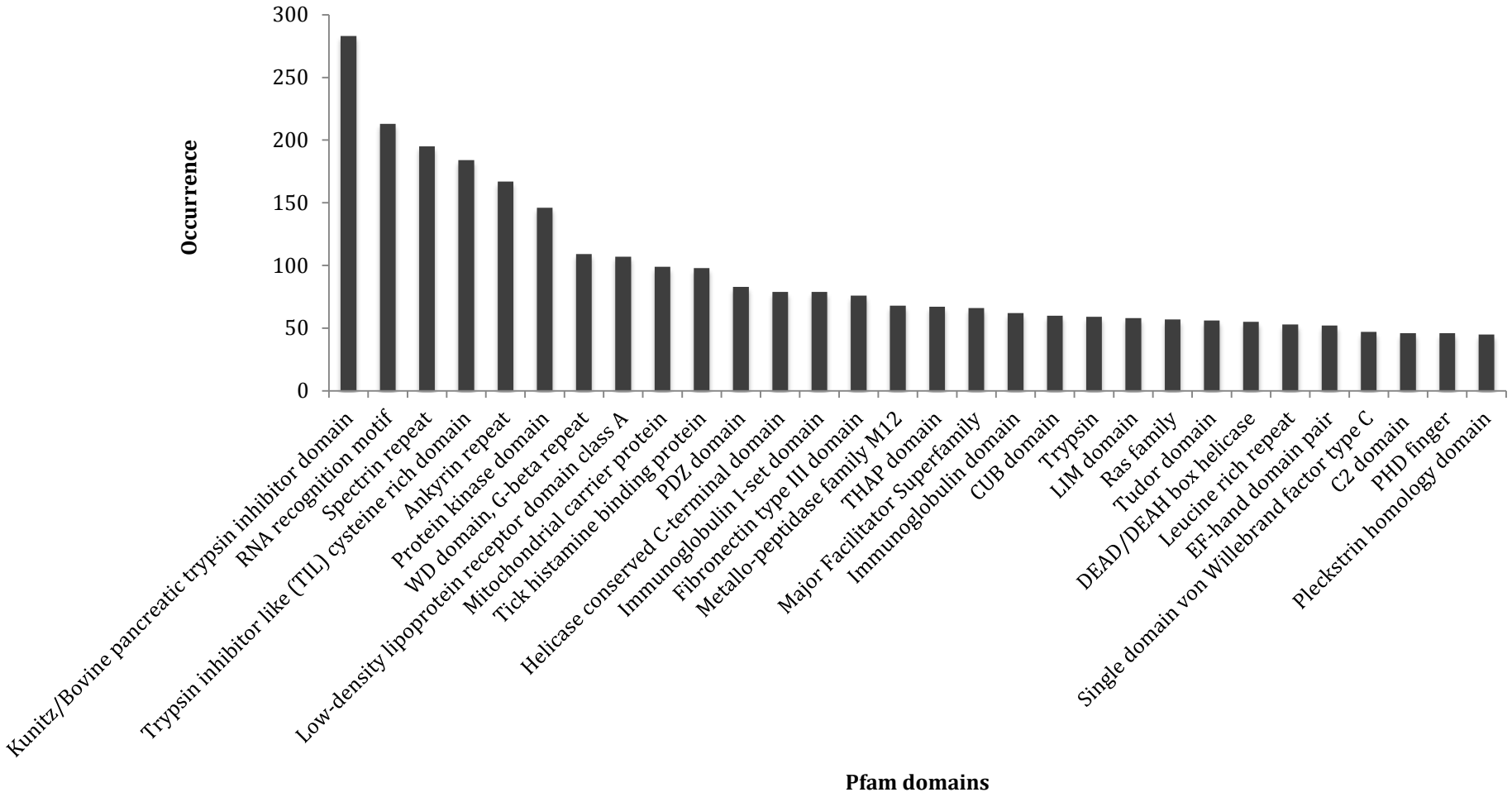
Appendix C: Figure S1 Gene Ontology (GO) characterisation of *R. zambeziensis*. Second level GO terms of biological processes, molecular functions and cellular components were analysed by the online software, WEGO. In total, 20 487 biological processes, 9659 molecular functions and 18 436 cellular components were obtained.



Appendix C: Figure S2 Eukaryotic Clusters of Orthologs (KOG) clustering of *R. zambeziensis* transcripts. A total of 9620 transcripts were assigned to 25 KOG categories**.**



Appendix C: Figure S3 The most abundant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways identified in *R. zambeziensis*.
In total, 4869 transcripts were assigned to 338 pathways.



Appendix C: Figure S4 Thirty most abundant Pfam domains in *R. zambeziensis*. A total of 8061 of the *R. zambeziensis* proteins contained at least a single Pfam domain.

APPENDIX D

Supplementary files of Chapter 4

Appendix D: Table S1 Descriptions of the 138 single copy orthologues present in all nine tick species. The *Ixodes scapularis* representative of each orthologous group is shown with annotations from VectorBase (www.vectorbase.org/organisms/ixodes-scapularis).

SCO	*I. scapularis* accession	Protein annotation
OG3547	ISCW021748-PA	Iron-containing alcohol dehydrogenase, putative (Fragment)
OG3550	ISCW020558-PA	NudC domain-containing protein, putative
OG3552	ISCW017009-PA	Membrane protein, putative
OG3554	ISCW013992-PA	Putative uncharacterized protein
OG3555	ISCW020688-PA	Dactylin, putative (Fragment)
OG3558	ISCW004107-PA	Membrane protein involved in ER to golgi transport, putative
OG3559	ISCW008843-PA	Steroid reductase, putative
OG3560	ISCW005013-PA	Parkinson disease 7 domain-containing protein, putative
OG3561	ISCW005878-PA	Ribosomal protein L2, putative
OG3568	ISCW006034-PA	Secreted salivary gland peptide, putative
OG3569	ISCW017709-PA	Heat shock protein 20.6, putative
OG3570	ISCW002622-PA	RNA polymerase II transcription elongation factor, putative
OG3572	ISCW022108-PA	Cytochrome B5, putative
OG3573	ISCW018532-PA	Protein phosphatase 1, regulatory (Inhibitor) subunit PPP1R2, putative
OG3576	ISCW002471-PA	Putative uncharacterized protein
OG3578	ISCW013037-PA	Ribosomal protein S26, putative
OG3579	ISCW016884-PA	Ribosomal protein S16, putative
OG3580	ISCW004473-PA	Histone 2A, putative
OG3581	ISCW015484-PA	Ribosomal protein S15Aa, putative
OG3582	ISCW018635-PA	cGMP-phosphodiesterase, putative (Fragment)
OG3583	ISCW020076-PA	Hepatocellular carcinoma-associated antigen, putative
OG3585	ISCW002944-PA	Chloride channel protein 1,2, putative
OG3586	ISCW019323-PA	Ubiquitin conjugating enzyme, putative
OG3589	ISCW006588-PA	Suppressor of actin, putative
OG3592	ISCW004820-PA	DNA-directed RNA polymerase subunit E', putative (Fragment)
OG3594	ISCW002129-PA	Molecular chaperone, putative
OG3595	ISCW020733-PA	Putative uncharacterized protein (Fragment)
OG3597	ISCW007981-PA	Ubiquitin protein ligase, putative
OG3598	ISCW023283-PA	Protective antigen 4D8
OG3600	ISCW002560-PA	N-acetyltransferase, putative
OG3601	ISCW006586-PA	Membrane protein, putative
OG3602	ISCW019130-PA	Ribosomal protein L13A, putative
OG3604	ISCW018020-PA	Rheb: GTP-binding protein Rheb

OG3606	ISCW010282-PA	RAS protein, putative (Fragment)
OG3607	ISCW015263-PA	E3 ubiquitin ligase, putative
OG3609	ISCW003337-PA	NIPSNAP1 protein, putative
OG3610	ISCW016281-PA	Prefoldin, putative
OG3611	ISCW006948-PA	20S proteasome, regulatory subunit beta, putative
OG3613	ISCW019380-PA	Ca2+ sensor, putative
OG3614	ISCW018508-PA	Ubiquitin protein ligase, putative
OG3617	ISCW021969-PA	RAB-9 and, putative
OG3618	ISCW010148-PA	Zinc finger protein, putative
OG3619	ISCW019597-PA	GDP-mannose pyrophosphorylase/mannose-1-phosphate guanylyltransferase
OG3620	ISCW011991-PA	Coiled-coil domain-containing protein, putative
OG3622	ISCW006856-PA	Membrane protein, putative
OG3624	ISCW017002-PA	RAS-related protein, putative
OG3625	ISCW021213-PA	Translation initiation factor 4F, cap-binding subunit, putative
OG3626	ISCW013805-PA	RAS-related protein, putative (Fragment)
OG3628	ISCW007783-PA	Aminoimidazole-4-carboxamide ribonucleotidetransformylase/IMP
OG3633	ISCW005463-PA	Prosalpha7: 26S proteasome alpha 7 subunit
OG3634	ISCW014398-PA	Alternative splicing factor ASF/ SF2, putative
OG3635	ISCW018497-PA	TPR domain-containing protein, putative
OG3636	ISCW008685-PA	Glucose de-repression and pre-vacuolar sorting protein, putative
OG3637	ISCW021807-PA	Transport and golgi organization protein, putative
OG3639	ISCW004969-PA	Survival motor neuron protein, putative (Fragment)
OG3641	ISCW018678-PA	Rwd domain-containing protein, putative
OG3645	ISCW012814-PA	Proteasome subunit alpha type, putative
OG3646	ISCW020246-PA	Reductase, putative
OG3648	ISCW009069-PA	Translation initiation factor, putative
OG3649	ISCW000036-PA	Syntaxin, putative
OG3650	ISCW014574-PA	Phosphomannomutase, putative
OG3651	ISCW019075-PA	Vacuolar H+-ATPase V1 sector, subunit D, putative
OG3652	ISCW024619-PA	Putative uncharacterized protein (Fragment)
OG3653	ISCW001107-PA	Ubiquitin protein ligase, putative
OG3654	ISCW021330-PA	Zinc finger protein, putative (Fragment)
OG3655	ISCW011816-PA	Putative uncharacterized protein
OG3660	ISCW014567-PA	Golgi protein, putative (Fragment)
OG3663	ISCW012982-PA	Putative uncharacterized protein
OG3664	ISCW024444-PA	Mammary gland family protein
OG3665	ISCW022135-PA	Protein phosphatase 1 binding protein, putative
OG3666	ISCW007146-PA	RNA-binding protein musashi, putative
OG3667	ISCW005210-PA	Dehydrogenase kinase, putative
OG3669	ISCW023170-PA	arp2/ 3, putative

OG3670	ISCW016533-PA	DEAD box protein abstrakt, putative
OG3671	ISCW012298-PA	Putative uncharacterized protein
OG3672	ISCW001649-PA	Putative uncharacterized protein (Fragment)
OG3673	ISCW020032-PA	Structure-specific recognition protein, putative
OG3674	ISCW017582-PA	Putative uncharacterized protein (Fragment)
OG3676	ISCW017493-PA	Putative uncharacterized protein
OG3677	ISCW001590-PA	U5 snRNP-associated RNA splicing factor, putative
OG3678	ISCW007193-PA	Phytanoyl-CoA alpha-hydroxylase, putative
OG3679	ISCW015357-PA	GDP-mannose pyrophosphorylase, putative
OG3680	ISCW010259-PA	Rat sphingolipid delta 4 desaturase, putative
OG3681	ISCW014414-PA	Putative uncharacterized protein (Fragment)
OG3682	ISCW007970-PA	Putative uncharacterized protein
OG3683	ISCW008840-PA	WD-repeat protein, putative
OG3684	ISCW017897-PA	Putative uncharacterized protein
OG3685	ISCW019115-PA	Branched chain alpha-keto acid dehydrogenase, putative (Fragment)
OG3686	ISCW022123-PA	Actin, putative
OG3687	ISCW010016-PA	Enoyl-CoA hydratase, putative
OG3688	ISCW004267-PA	WD-repeat protein, putative
OG3689	ISCW020742-PA	mRpL37: 39S mitochondrial ribosomal protein L37
OG3690	ISCW002422-PA	UBX domain-containing protein, putative
OG3691	ISCW020516-PA	GOT2 aspartate aminotransferase, putative
OG3692	ISCW018975-PA	COP9 signalosome, subunit CSN3, putative
OG3694	ISCW002616-PA	Microfibril-associated protein, putative
OG3695	ISCW008801-PA	Putative uncharacterized protein
OG3696	ISCW020455-PA	RNA-binding protein musashi, putative (Fragment)
OG3697	ISCW000261-PA	G patch domain and KOW motifs-containing protein, putative
OG3698	ISCW021592-PA	Sorting nexin, putative (Fragment)
OG3700	ISCW016836-PA	Nuclear matrix protein, putative (Fragment)
OG3701	ISCW016001-PA	Hepatocellular carcinoma-associated antigen, putative
OG3703	ISCW001753-PA	Serine/ threonine protein kinase rio2, putative (Fragment)
OG3706	ISCW009216-PA	Exosome component, putative
OG3708	ISCW020925-PA	CDGSH iron sulfur domain-containing protein, putative
OG3709	ISCW018296-PA	Transmembrane protein, putative
OG3711	ISCW014200-PA	Beclin, putative
OG3712	ISCW007279-PA	Boca, putative
OG3713	ISCW001550-PA	Transmembrane protein Tmp21, putative
OG3714	ISCW020156-PA	Transmembrane protein 41A, putative
OG3715	ISCW006264-PA	Putative uncharacterized protein
OG3716	ISCW013450-PA	Carbonic anhydrase, putative (Fragment)
OG3719	ISCW006181-PA	Valacyclovir hydrolase, putative

OG3721	ISCW000585-PA	Fam43, putative
OG3722	ISCW017282-PA	Oxodicarboxylate carrier protein, putative
OG3723	ISCW018585-PA	Isocitrate dehydrogenase, putative
OG3725	ISCW012625-PA	Putative uncharacterized protein
OG3726	ISCW009187-PA	CD151 antigen, putative
OG3727	ISCW004488-PA	Tryptophan-rich protein, putative
OG3728	ISCW023370-PA	Thiamin pyrophosphokinase, putative (Fragment)
OG3732	ISCW003698-PA	Putative uncharacterized protein
OG3734	ISCW006277-PA	RNA polymerase II proteinral transcription factor BTF3, putative
OG3735	ISCW019638-PA	NADH-ubiquinoe oxidoreductase subunit, putative
OG3736	ISCW018346-PA	Vacuolar H+ ATPase
OG3737	ISCW011809-PA	Putative uncharacterized protein
OG3739	ISCW016991-PA	MPN domain-containing protein, putative (Fragment)
OG3740	ISCW001758-PA	Putative uncharacterized protein
OG3742	ISCW016111-PA	Mediator of RNA polymerase II transcription subunit, putative
OG3744	ISCW011885-PA	Putative uncharacterized protein
OG3745	ISCW006520-PA	Putative uncharacterized protein
OG3746	ISCW013792-PA	Putative uncharacterized protein
OG3747	ISCW009370-PA	Putative uncharacterized protein
OG3748	ISCW019117-PA	Pyridoxine kinase, putative (Fragment)
OG3749	ISCW007588-PA	Carrier protein, putative (Fragment)
OG3751	ISCW006672-PA	Pseudouridylate synthase, putative
OG3752	ISCW018771-PA	Glutamine synthetase, putative
OG3753	ISCW005792-PA	Putative uncharacterized protein
OG3757	ISCW013828-PA	Secreted protein, putative

Appendix D: Table S2 Summary of the partitioned alignment of the species phylogeny indicating sequence range, number of informative sites and alignment model. Seventy-six of the single copy orthologues were selected for the molecular phylogeny based on presence in at least 18 of the species.

SCO	Number of species[a]	Alignment length (aa)	Sequence range	Predicted alignment model[b]	Number of informative sites	Informative sites/ Alignment length (%)
OG3547	18	456	1 - 456	LG+G4	114	25.0
OG3554	20	533	457 - 989	LG+G4	313	58.7
OG3558	19	161	990 - 1150	LG+R2	50	31.1
OG3568	18	106	1151 - 1256	LG+G4	52	49.1
OG3569	20	169	1257 - 1425	LG+G4	35	20.7
OG3570	20	111	1426 - 1536	LG+I	28	25.2
OG3573	19	136	1537 - 1672	WAG+G4	90	66.2
OG3576	18	124	1673 - 1796	LG+G4	42	33.9
OG3579	18	118	1797 - 1914	LG+G4	56	47.5
OG3580	20	128	1915 - 2042	PMB	3	2.3
OG3585	18	895	2043 - 2937	LG+R3	549	61.3
OG3594	20	148	2938 - 3085	JTT+G4	95	64.2
OG3598	20	159	3086 - 3244	JTT+G4	41	25.8
OG3600	18	175	3245 - 3419	LG+I	10	5.7
OG3602	19	171	3420 - 3590	WAG+G4	51	29.8
OG3606	19	184	3591 - 3774	VT+G4	12	6.5
OG3607	20	177	3775 - 3951	JTTDCMut+G4	46	26.0
OG3610	19	189	3952 - 4140	JTTDCMut+G4	74	39.2
OG3611	18	205	4141 - 4345	LG+R2	70	34.1
OG3613	19	191	4346 - 4536	LG+G4	64	33.5
OG3614	18	196	4537 - 4732	VT+G4	88	44.9
OG3617	18	204	4733 - 4936	LG+G4	11	5.4
OG3618	19	191	4937 - 5127	LG+G4	44	23.0
OG3619	19	358	5128 - 5485	LG+R2	66	18.4
OG3624	19	214	5486 - 5699	LG+G4	129	60.3
OG3628	18	591	5700 - 6290	LG+G4	98	16.6
OG3633	20	252	6291 - 6542	LG+G4	72	28.6
OG3635	20	324	6543 - 6866	LG+G4	163	50.3
OG3636	18	215	6867 - 7081	LG+R2	48	22.3
OG3637	18	224	7082 - 7305	LG+G4	113	50.4
OG3639	20	211	7306 - 7516	JTT+R3	143	67.8

OG3641	18	230	7517 - 7746	LG+G4	104	45.2
OG3646	19	242	7747 - 7988	LG+G4	150	62.0
OG3648	18	163	7989 - 8151	LG+G4	24	14.7
OG3650	20	248	8152 - 8399	LG+G4	115	46.4
OG3651	19	247	8400 - 8646	LG+G4	57	23.1
OG3654	20	269	8647 - 8915	JTTDCMut+G4	152	56.5
OG3660	18	281	8916 - 9196	LG+G4	46	16.4
OG3665	20	280	9197 - 9476	LG+G4	132	47.1
OG3667	18	403	9477 - 9879	LG+R3	204	50.6
OG3670	18	622	9880 - 10 501	LG+R3	130	20.9
OG3672	19	314	10 502 - 10 815	JTT+R3	104	33.1
OG3673	19	686	10 816 - 11 501	LG+F+R3	188	27.4
OG3674	19	454	11 502 - 11 955	JTT+G4	221	48.7
OG3679	18	415	11 956 - 12 370	LG+G4	156	37.6
OG3681	18	347	12 371 - 12 717	LG+R2	180	51.9
OG3682	20	354	12 718 - 13 071	LG+R3	127	35.9
OG3683	18	492	13 072 - 13 563	JTT+G4	27	5.5
OG3687	20	390	13 564 - 13 953	LG+G4	176	45.1
OG3688	19	401	13 954 - 14 354	JTT+R2	216	53.9
OG3689	19	403	14 355 - 14 757	LG+G4	227	56.3
OG3690	18	422	14 758 - 15 179	LG+G4	229	54.3
OG3691	20	425	15 180 - 15 604	LG+G4	134	31.5
OG3692	19	426	15 605 - 16 030	JTT+G4	107	25.1
OG3694	19	438	16 031 - 16 468	LG+F+R3	124	28.3
OG3696	19	390	16 469 - 16 858	WAG+G4	320	82.1
OG3697	19	475	16 859 - 17 333	JTT+G4	233	49.1
OG3698	19	496	17 334 - 17 829	JTT+R2	205	41.3
OG3700	19	669	17 830 - 18 498	JTT+R3	317	47.4
OG3701	20	506	18 499 - 19 004	LG+G4	290	57.3
OG3706	19	587	19 005 - 19 591	LG+G4	277	47.2
OG3708	19	98	19 592 - 19 689	LG+G4	47	48.0
OG3713	19	205	19 690 - 19 894	JTT+G4	67	32.7
OG3714	18	241	19 895 - 20 135	LG+G4	125	51.9
OG3716	19	308	20 136 - 20 443	VT+I+G4	244	79.2
OG3721	18	300	20 444 - 20 743	JTT+G4	73	24.3
OG3722	19	302	20 744 - 21 045	JTTDCMut+G4	64	21.2
OG3723	18	379	21 046 - 21 424	LG+R3	63	16.6
OG3726	20	214	21 425 - 21 638	LG+G4	173	80.8
OG3736	19	154	21 639 - 21 792	cpREV+R2	18	11.7
OG3739	19	448	21 793 - 22 240	LG+G4	138	30.8

OG3740	19	234	22 241 - 22 474	LG+R2	33	14.1
OG3742	18	236	22 475 - 22 710	JTTDCMut+G4	30	12.7
OG3745	19	235	22 711 - 22 945	JTTDCMut+G4	91	38.7
OG3749	19	311	22 946 - 23 256	LG+G4	148	47.6
OG3752	19	403	23 257 - 23 659	WAG+G4	138	34.2
Concatenated alignment	20	23 659	1 - 23 659	Partitioned models	9194	38.9

[a] Number of species that have sequence data for the single copy orthologous (SCO) group and are included in the particular alignment. Groups missing sequence data of more than two species were not included in the final concatenated alignment.

[b] The optimal evolutionary alignment model for each SCO was predicted using ModelFinder of the IQ-TREE v1.5.4 package (Nguyen *et al.*, 2014; Kalyaanamoorthy *et al.*, 2017).

Appendix D: Table S3 The 100 most variable Pfam domains between *R. appendiculatus* and *R. zambeziensis.*

Pfam domain name	***R. appendiculatus***	***R. zambeziensis***	***χ2***	**p-value**
zf-H2C2_2[a]	129	0	129.00	< 0.0001*
I-set[a]	198	79	51.12	< 0.0001*
RRM_6[a]	42	0	42.00	< 0.0001*
Ig_3[a]	10	62	37.56	< 0.0001*
zf-C2H2[a]	4	44	33.33	< 0.0001*
TPR_11[a]	55	10	31.15	< 0.0001*
Spectrin	100	195	30.59	< 0.0001*
fn3 (Fibronectin type III)	141	76	19.47	< 0.0001*
Laminin_EGF	84	38	17.34	< 0.0001*
Reprolysin_5	28	68	16.67	< 0.0001
adh_short_C2	0	16	16.00	< 0.0001
Ldl_recept_a	173	107	15.56	< 0.0001
Cys_rich_FGFR (fibroblast growth factor receptor)	0	15	15.00	0.0001
TIL	120	184	13.47	0.0002
Ank_5	31	9	12.10	0.0005
HAD	12	0	12.00	0.0005
CUB	28	60	11.64	0.0006
Reprolysin_2	38	14	11.08	0.0009
Abhydrolase_5	11	0	11.00	0.0009
GCC2_GCC3	11	0	11.00	0.0009
COMM_domain	0	10	10.00	0.0016
Ig_2	21	5	9.85	0.0017
zf-H2C2_5	2	14	9.00	0.0027
HCaRG	9	0	9.00	0.0027
zf-C3HC4_2	1	11	8.33	0.0039
WD40	71	109	8.02	0.0046
Granulin	9	25	7.53	0.0061
Tnp_P_element	1	10	7.36	0.0067
Mucin2_WxxW	12	2	7.14	0.0075
Methyltransf_25	0	7	7.00	0.0082
rve	17	36	6.81	0.0091
Biotin_carb_N	0	6	6.00	0.0143
MCM_OB	0	6	6.00	0.0143
Methyltr_RsmB-F	0	6	6.00	0.0143
DUF1136	6	0	6.00	0.0143

CaMKII_AD	16	5	5.76	0.0164
G_glu_transpept	4	14	5.56	0.0184
HEAT_2	14	4	5.56	0.0184
TSP_1	1	8	5.44	0.0196
C8	8	1	5.44	0.0196
Clathrin	3	12	5.40	0.0201
Filamin	56	34	5.38	0.0204
Kunitz_BPTI	231	283	5.26	0.0218
LRR_8	79	53	5.12	0.0236
Ephrin_rec_like	0	5	5.00	0.0253
Glyco_hydro_30C	0	5	5.00	0.0253
GST_C_6	0	5	5.00	0.0253
HS1_rep	0	5	5.00	0.0253
SAPS	0	5	5.00	0.0253
WSD	0	5	5.00	0.0253
AA_permease_C	5	0	5.00	0.0253
CPSase_L_chain	5	0	5.00	0.0253
PH_13	5	0	5.00	0.0253
Thymosin	5	0	5.00	0.0253
Sushi	34	18	4.92	0.0265
Glyco_transf_7N	3	11	4.57	0.0325
TPR_1	6	16	4.55	0.0330
Hydrolase_4	1	7	4.50	0.0339
OATP	1	7	4.50	0.0339
WHEP-TRS	1	7	4.50	0.0339
Hydrolase	2	9	4.45	0.0348
TUDOR	36	56	4.35	0.0371
Prenyltrans	14	5	4.26	0.0389
Asp_protease	0	4	4.00	0.0455
BEN	0	4	4.00	0.0455
Guanylate_cyc	0	4	4.00	0.0455
Sel1	0	4	4.00	0.0455
Acetyltransf_7	4	0	4.00	0.0455
Nol1_Nop2_Fmu	4	0	4.00	0.0455
Kelch_1	41	25	3.88	0.0489
Thyroglobulin_1	3	10	3.77	0.0522
Methyltransf_11	10	3	3.77	0.0522
MBOAT	16	29	3.76	0.0526
RRM_1	175	213	3.72	0.0537
MIT_C	2	8	3.60	0.0578

DUF4749	8	2	3.60	0.0578
Neuralized	8	2	3.60	0.0578
Peptidase_S10	17	30	3.60	0.0579
GCIP	1	6	3.57	0.0588
Med7	1	6	3.57	0.0588
Str_synth	1	6	3.57	0.0588
GIT_SHD	6	1	3.57	0.0588
PA	6	1	3.57	0.0588
Ribonuclease_3	6	1	3.57	0.0588
Lipase	13	5	3.56	0.0593
Laminin_G_2	40	25	3.46	0.0628
Peptidase_C1	12	23	3.46	0.0630
zf-CCHC	56	38	3.45	0.0634
THAP	48	67	3.14	0.0764
Glyco_transf_7C	3	9	3.00	0.0833
Methyltransf_31	9	3	3.00	0.0833
ABC_trans_CmpB	0	3	3.00	0.0833
CCDC84	0	3	3.00	0.0833
CDC45	0	3	3.00	0.0833
Clathrin_propel	0	3	3.00	0.0833
CLCA	0	3	3.00	0.0833
Exostosin	0	3	3.00	0.0833
Glyco_transf_92	0	3	3.00	0.0833
Lipase_GDSL_2	0	3	3.00	0.0833
NUDE_C	0	3	3.00	0.0833

Chi-square test was performed and $\chi 2$ - and p-values are indicated. $df = 1$.

* Significant Chi-square test (Bonferroni corrected p-value < 0.00003).

[a] Pfam domains with significant p-values forming part of larger Pfam domain families. When analysed individually, p-values were significant, but when analysed as Pfam domain families, significance was not maintained.

Appendix D: Table S4 Expression level comparison between putative tick immunity or pathogen transmission orthologues of *R. appendiculatus* and *R. zambeziensis*. Putative orthologues were determined by BLASTp analyses against previously characterised tick immunity proteins and Chi-square tests were performed between the expression levels (transcripts per million; TPM) of each potential orthologue in each species.

Accession number	**Protein name**	**Size (aa)**	***R. appendiculatus* orthologue**	**Size (aa)**	**Identity to target (%)**	***R. zambeziensis* orthologue**	**Size (aa)**	**Identity to target (%)**	**Identity between *R. appendiculatus* and *R. zambeziensis* (%)**	**TPM: *R. appendiculatus***	**TPM: *R. zambeziensis***	***χ2***	**p-value**
AAY66972.1	Tick histamine release factor (tHRF)	173	Rapp_Mc12631	173	74.0	Rzam_Mc198	173	73.4	99.4	1211.6	6057.3	3230.41	0*
AAL99403.1	Glutathione S-transferase	220	Rapp_Mc2803	220	94.6	Rzam_Mc5179	220	94.1	98.2	741.4	1124.0	78.48	< 0.0001*
AAV67034.2	Subolesin	165	Rapp_Mc1474	161	93.8	Rzam_Mc550	161	93.8	98.8	16.5	67.7	31.06	< 0.0001*
XP_002433506.1	Spectrin α chain (fodrin)	2368	Rapp_Mc158	2417	95.1	Rzam_Mc1908	2417	95.1	100	87.6	33.4	24.25	< 0.0001*
AAO24323.1	Varisin	74	Rapp_Mc11752	117	81.1	Rzam_Mc13584	74	79.7	97.3	29.3	6.0	15.37	< 0.0001*
XP_002408065.1	Mitochondrial porin (T2)	282	Rapp_Mc4024	282	86.2	Rzam_Mc13585	282	86.5	99.7	130.8	80.9	11.76	0.0006*
AAN78224.1	Factor D-like	374	Rapp_Mc13562	444	96.5	Rzam_Mc674	373	96.5	100	26.6	7.2	11.18	0.0008*
AAD15991.1	Glutathione S-transferase	223	Rapp_Mc12689	223	98.7	Rzam_Mc2231	223	99.1	99.6	56.5	30.3	7.89	0.0050
AAR29939.1	Calreticulin	411	Rapp_Mc2933	412	98.5	Rzam_Mc8227	412	98.8	99.8	642.5	547.2	7.65	0.0057
AAK97814.1	Salivary protein 25D (Salp25D)	221	Rapp_Mc62	221	90.5	Rzam_Mc3872	221	90.5	99.1	45.8	29.3	3.61	0.0573
BAF43801.1	Longipain	341	Rapp_Mc14147	334[a]	76.9	Rzam_Mc5941	347	78.0	98.8	26.4	20.1	0.83	0.3616
AAO92279.1	Glutathione S-transferase	215	Rapp_Mc2764	226	70.2	Rzam_Mc12649	226	71.6	97.4	3.4	1.7	0.58	0.4449
AAO23571.1	C-type lysozyme	139	Rapp_Mc3600	140	88.5	Rzam_Mc4607	140	87.8	97.1	2.8	1.4	0.45	0.5039
XP_002406129.1	α 1, 3-fucosyltransferase[c]	324	Rapp_Mc5325	399	72.3	Rzam_Mc8155	404	72.0	98.3	2.7	1.4	0.37	0.5440
XP_002403528.1	Vacuolar H+ ATPase (vATPase)[d]	519	Rapp_Mc663	348	90.8	Rzam_Mc12576	348	90.8	100	29.7	25.6	0.30	0.5830
ADU86241.1	Signal transducer and activator of transcription (STAT)[c]	752	Rapp_Mc7017	797	86.1	Rzam_Mc10339	807	86.1	100	3.0	3.9	0.10	0.7546
ADK23790.1	Phosphoinositide 3-kinase (pak1)[c]	290	Rapp_Mc7734	456	87.9	Rzam_Mc9852	456	87.6	99.6	3.4	2.8	0.05	0.8226
B7Q1P7	p21-activated kinase (pi3k)	864	Rapp_Mc4176	697[b]	84.0	Rzam_Mc8060	916	84.3	99.9	14.2	13.4	0.02	0.8820
ACJ26770.1	α2-macroglobulin (IrAM)	1486	Rapp_Mc7464	1484	81.4	Rzam_Mc13586	1497	80.9	98.5	34.1	41.3	0.69	0.4068

[a] C-terminal truncation in *R. appendiculatus* orthologue (21 aa).

[b] N-terminal truncation in *R. appendiculatus* orthologue (189 aa).

[c] The *R. appendiculatus* and *R. zambeziensis* orthologues of α 1, 3-fucosyltransferase, STAT and pak1 had either extended C- or N-terminal sequences that were identical in both species and the orthologues were considered full-length.

[d] *R. appendiculatus* and *R. zambeziensis* orthologue sequences of vATPase started at a later region than the target sequence (from *I. scapularis*). Investigation on UniProtKB (www.uniprot.org) revealed that the N-terminal extension in the *I. scapularis* protein might be a species-specific modification. Based on this, the *R. appendiculatus* and *R. zambeziensis* orthologues of vATPase were considered full-length.

* Significant p-values < 0.0026 after Bonferroni correction (*df* = 1).

Gene references: AAY66972.1, *I. scapularis* tick histamine release factor (tHRF) (Dai *et al.*, 2010); AAL99403.1, *R. microplus* glutathione S-transferase (Rosa de Lima *et al.*, 2002); AAV67034.2, *D. variabilis* subolesin (de la Fuente *et al.*, 2006); XP_002433506.1, *I. scapularis* spectrin α chain (fodrin) (Ayllón *et al.*, 2013); AAO24323.1, *D. variabilis* varisin (Kocan *et al.*, 2008); XP_002408065.1, *I. scapularis* mitochondrial porin (T2) (Ayllón *et al.*, 2013); AAN78224.1, *D. variabilis* factor D-like (Simser *et al.*, 2004b); AAD15991.1, *R. microplus* glutathione S-transferase (He *et al.*, 1999); AAR29939.1, *R. annulatus* calreticulin (Antunes *et al.*, 2012); AAK97814.1, *I. scapularis* salivary protein 25D (Salp25D) (Narasimhan *et al.*, 2007); BAF43801.1, *H. longicornis* longipain (Tsuji *et al.*, 2008); AAO92279.1, *D. variabilis* glutathione S-transferase (Dreher-Lesnick *et al.*, 2006); AAO23571.1, *D. variabilis* c-type lysozyme (Simser *et al.*, 2004a); XP_002406129.1, *I. scapularis* alpha 1, 3-fucosyltransferase (Pedra *et al.*, 2010); XP_002403528.1, *I. scapularis* vacuolar H+ ATPase (vATPase) (de la Fuente *et al.*, 2007); ADU86241.1, *I. scapularis* signal transducer and activator of transcription (STAT) (Liu *et al.*, 2012); ADK23790.1, *I. scapularis* phosphoinositide 3-kinase (pak1) (Sultana *et al.*, 2010); B7Q1P7, *I. scapularis* p21-activated kinase (pi3k) (Sultana *et al.*, 2010); and ACJ26770.1, *I. ricinus* alpha2-macroglobulin (IrAM) (Buresova *et al.*, 2009).

Appendix D: Table S5 The 23 reciprocal best BLAST transcripts, annotated as putative long non-coding RNAs, of *R. appendiculatus* and *R. zambeziensis* that showed significantly different expression levels (Bonferroni corrected p-value < 0.0001; *df* = 1). Expression levels indicated as TPM, transcript per million.

R. appendiculatus **transcript**	**Length (bp)**	**TPM**	***R. zambeziensis*** **transcript**	**Length (bp)**	**TPM**	χ^2	**p-value**
c15911_g1_i1	341	5063.5	TRINITY_DN23428_c0_g1_i1	354	126.9	4695.32	0.0
c60750_g1_i1	355	516.0	TRINITY_DN47004_c0_g1_i1	319	87.4	304.36	< 0.0001
c59789_g1_i1	449	214.2	TRINITY_DN54189_c0_g1_i1	578	16.8	168.78	< 0.0001
c35159_g1_i1	415	209.9	TRINITY_DN28874_c0_g3_i1	332	18.0	161.64	< 0.0001
c32358_g1_i1	618	153.7	TRINITY_DN24067_c0_g1_i1	371	2.2	147.21	< 0.0001
c36128_g1_i1	410	9.6	TRINITY_DN29797_c0_g1_i1	610	129.6	103.47	< 0.0001
c444_g1_i1	382	99.2	TRINITY_DN20140_c0_g2_i1	382	6.2	82.06	< 0.0001
c43081_g1_i1	415	78.6	TRINITY_DN34107_c6_g21_i1	394	2.0	72.69	< 0.0001
c29220_g1_i1	413	51.3	TRINITY_DN34572_c10_g1_i1	431	179.5	71.27	< 0.0001
c32160_g1_i1	408	57.9	TRINITY_DN67250_c0_g1_i1	325	1.8	52.67	< 0.0001
c38959_g1_i1	1136	2.1	TRINITY_DN31069_c0_g2_i1	1105	50.3	44.29	< 0.0001
c43149_g4_i1	648	55.3	TRINITY_DN34679_c0_g4_i1	646	10.2	31.03	< 0.0001
c38676_g1_i1	527	4.5	TRINITY_DN24964_c1_g2_i1	550	42.1	30.30	< 0.0001
c38941_g3_i1	1312	44.3	TRINITY_DN36562_c7_g3_i1	1413	5.9	29.33	< 0.0001
c46136_g1_i1	926	32.4	TRINITY_DN26193_c1_g1_i1	468	1.4	28.58	< 0.0001
c42302_g4_i1	726	33.8	TRINITY_DN36048_c2_g3_i1	1238	2.1	27.89	< 0.0001
c43256_g2_i1	720	68.6	TRINITY_DN35854_c2_g2_i1	743	23.7	21.91	< 0.0001
c44750_g4_i1	921	31.8	TRINITY_DN36562_c4_g1_i2	918	4.8	19.80	< 0.0001
c36127_g1_i2	311	29.1	TRINITY_DN48262_c0_g1_i1	304	4.6	17.80	< 0.0001
c39687_g1_i1	890	26.1	TRINITY_DN34606_c3_g1_i1	999	3.3	17.63	< 0.0001
c30489_g1_i3	787	26.1	TRINITY_DN2030_c0_g1_i1	640	3.5	17.19	< 0.0001
c41523_g1_i2	649	2.0	TRINITY_DN37637_c2_g3_i2	496	22.1	16.63	< 0.0001
c32675_g1_i1	637	24.0	TRINITY_DN15956_c0_g1_i1	511	3.1	16.24	< 0.0001

APPENDIX E

Curriculum Vitae

Minique de Castro

Agricultural Research Council	Email: decastrom@arc.agric.za
100 Old Soutpan Road	Mobile: +27721464954
Onderstepoort, Pretoria, 0110	Date of birth: 28 September 1980
South Africa	Gender: Female

Qualifications:

PhD Life Sciences, University of South Africa, South Africa, 2017
MSc Genetics, University of Pretoria, South Africa, 2006 (distinction)
BSc (Hons) Genetics, University of Pretoria, South Africa, 2002
BSc Human genetics, University of Pretoria, South Africa, 2001

Dissertations:

PhD: Sialotranscriptomics of the brown ear ticks, *Rhipicephalus appendiculatus* Neumann, 1901 and *R. zambeziensis* Walker, Norval and Corwin, 1981, vectors of Corridor disease. Supervisors: Prof. B.J. Mans & Prof. D.J.G. Rees (Agricultural Research Council and University of South Africa)

MSc: Allelic diversity in the *CAD2* and *LIM1* lignin biosynthetic genes of *Eucalyptus grandis* Hill ex Maiden and *E. smithii* R. T. Baker. Supervisors: Prof. A.A. Myburg & Prof. P. Bloomer

BSc (Hons): 1) The disproportionate contribution of pollen parents in inter-specific polymix crosses of *Eucalyptus*. Supervisor: Prof AA Myburg. 2) The effect on cytotoxicity of a transmembrane domain 2 mutant of African horsesickness virus non-structural NS3 protein. Supervisors: Dr .W.C. Fick & Prof. H. Huismans

Employment history:

Jan 2011 - Dec 2012: Agricultural Research Council - Next Generation Sequencing Facility.
May 2007 - Nov 2010: University of Pretoria - Research assistant.

Awards and travel grants:

Whitehead Scientific Travel Award, to attend the 9th Tick and Tick-borne Pathogen & 1st Asia Pacific Rickettsia Joint Conference (2017).

Scherago International Student Travel Grants, to attend the Plant and Animal Genome Conference (2014 & 2016).

Best Poster Award, 8th Tick and Tick-borne Pathogen & 12th Biennial Society for Tropical Veterinary Medicine Joint Conference (2014).

Best Animal Science Oral (PhD), ARC Professional Development Program Conference (2014).

Publications:

de Castro MH, de Klerk D, Pienaar R, Rees DJG and Mans BJ (2017). Sialotranscriptomics of *Rhipicephalus zambeziensis* reveal intricate expression profiles of secretory proteins and suggest tight temporal transcriptional regulation during blood feeding. Parasites & Vectors. 10: 384.

Mans BJ, Featherston J, de Castro MH and Pienaar R (2017). Gene duplication and protein evolution in tick-host interactions. Frontiers in Cellular and Infection Microbiology. 7: 413.

de Castro MH, de Klerk D, Pienaar R, Latif AA, Rees DJG and Mans BJ (2016). *De novo* assembly and annotation of the salivary gland transcriptome of *Rhipicephalus appendiculatus* male and female ticks during blood feeding. Ticks Tick-Borne Diseases. 7(4): 536-548.

Mans BJ, de Castro MH, Pienaar R, de Klerk D, Gaven P, Genu S and Latif AA (2016). Ancestral reconstruction of tick lineages. Ticks Tick-Borne Diseases. 7(4): 509-535.

Mans BJ, de Klerk D, Pienaar R, de Castro MH and Latif AA. (2015). Next-generation sequencing as means to retrieve tick systematic markers, with the focus on *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae). Ticks Tick-Borne Diseases. 6(4): 450-462.

Creux* NM, de Castro* MH, Ranik M, Maleka MF and Myburg AA. (2013). Diversity and cis-element architecture of the promoter regions of cellulose synthase genes in Eucalyptus. Tree Genetics & Genomes 9(4): 989-1004. * *Joint first authorship.*

Mans BJ, de Klerk D, Pienaar R, de Castro MH and Latif AA. (2012). The mitochondrial genomes of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae) and *Argas africolumbae* (Ixodoidae: Argasidae): estimation of divergence dates for the major tick lineages and reconstruction of ancestral blood-feeding characters. PLoS One 7(11): e49461.

Abolnik C, de Castro M and Rees J. (2012). Full genomic sequence of an African Avian Paramyxovirus Type 4 strain isolated from a wild duck. Virus Genes 45(3): 537-541.

Creux NM, Ranik M, De Castro MH, Spokevicius A, Bossinger G, Maritz-Olivier C and Myburg AA. (2012). Understanding the transcriptional regulation of the *Eucalyptus cellulose synthase1* gene. South African Journal of Botany. 79: 182.

Severn-Ellis A, Schoeman MH, Willemse S, Sippel A, Rees J and de Castro M. (2012). Towards guava wilt disease resistance in South Africa. Acta Hortic. 959: 73-79.

Creux NM, De Castro MH, Ranik M, Spokevicius AV, Bossinger G, Maritz-Olivier C and Myburg AA. (2011). *In silico* and functional characterization of the promoter of a *Eucalyptus* secondary cell wall associated cellulose synthase gene (*EgCesA1*). BMC Proceedings 2011, 5(Suppl 7): P107.

Myburg A, Bradfield J, Cowley E, Creux N, de Castro M, Hatherell T-L, Mphahlele M, O'Neill M, Ranik M, Solomon L, Victor M, Zhou H, Galloway G, Horsley T, Jones N, Stanger T, Bayley A, Edwards N and Janse B. (2008). Forest and fibre genomics: Biotechnology tools for applied tree improvement. Southern Forests: Journal of Forest Science 70(2): 59-68.

Presentations:

de Castro MH, de Klerk D, Pienaar R, Rees DJG and Mans BJ. (2017). Sequencing the salivary gland transcriptome of *Rhipicephalus zambeziensis* during blood feeding. 9th Tick and Tick-borne Pathogen & 1st Asia Pacific Rickettsia Joint Conference, August 27 - September 1, Cairns, Australia.

de Castro MH, de Klerk D, Pienaar R, Latif A, Rees DJG and Mans BJ. (2016). Sialotranscriptomics of *Rhipicephalus appendiculatus* male and female ticks. Plant and Animal Genome Conference XXIV, January 9 - 13, San Diego, California.

de Castro MH, de Klerk D, Pienaar R, Latif AA, Rees DJG and Mans BJ. (2016). Sialotranscriptomics of *Rhipicephalus appendiculatus* and *R. zambeziensis*, vectors of Corridor disease. Agricultural Research Council - Professional Development Program Conference, October 24 - 25, Pretoria, South Africa.

de Castro MH, de Klerk D, Pienaar R, Latif A, Rees DJG and Mans BJ. (2014). *de novo* transcriptome assembly of *Rhipicephalus appendiculatus* salivary glands. 8th Ticks and Tick-borne Pathogen & Biennial Society for Tropical Veterinary Medicine Joint Conference, August 24 - 29, Cape Town, South Africa. (Poster).

de Castro MH, de Klerk D, Pienaar R, Latif A, Rees DJG and Mans BJ. (2014). Transcriptome assembly and annotation of the brown ear tick, *Rhipicephalus appendiculatus*. South African Genetics Society & South African Society for Bioinformatics Joint Congress, September 23 - 26, Kwalata Game Ranch, South Africa.

de Castro MH, de Klerk D, Pienaar R, Latif A, Rees DJG and Mans BJ. (2014). *de novo* transcriptome assembly of *Rhipicephalus appendiculatus*. Agricultural Research Council - Professional Development Program Conference, June 27, Pretoria, South Africa.

de Castro MH, de Klerk D, Pienaar R, Latif A, Rees DJG and Mans BJ. (2014). *de novo* assembly of an expressed gene catalog from the tick, *Rhipicephalus appendiculatus*. Plant and Animal Genome Conference XXII, January 11 - 15, San Diego, California. (Poster).

REFERENCE LIST

Abbas RZ, Zaman MA, Colwell DD, Gilleard J & Iqbal Z 2014: Acaricide resistance in cattle ticks and approaches to its management: the state of play. *Veterinary Parasitology.* 203, 6-20.

Abdullah HHAM, El-Molla A, Salib FA, Allam NAT, Ghazy AA & Abdel-Shafy S 2016: Morphological and molecular identification of the brown dog tick *Rhipicephalus sanguineus* and the camel tick *Hyalomma dromedarii* (Acari: Ixodidae) vectors of Rickettsioses in Egypt. *Veterinary World.* 9, 1087-1101.

Adamson SW, Browning RE, Budachetri K, Ribeiro JMC & Karim S 2013: Knockdown of selenocysteine-specific elongation factor in *Amblyomma maculatum* alters the pathogen burden of *Rickettsia parkeri* with epigenetic control by the Sin3 histone deacetylase corepressor complex. *PLoS One.* 8, e82012.

Agarwala KL, Kawabata S, Hirata M, Miyagi M, Tsunasawa S & Iwanaga S 1996: A cysteine protease inhibitor stored in the large granules of horseshoe crab hemocytes: purification, characterization, cDNA cloning and tissue localization. *Journal of Biochemistry.* 119, 85-94.

Alarcon-Chaidez FJ, Sun J & Wikel SK 2007: Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). *Insect Biochemistry and Molecular Biology.* 37, 48-71.

Ali A, Parizi LF, Guizzo MG, Tirloni L, Seixas A, Vaz Ida S & Termignoni C 2015: Immunoprotective potential of a *Rhipicephalus* (*Boophilus*) *microplus* metalloprotease. *Veterinary Parasitology.* 207, 107-114.

Aljamali MN, Ramakrishnan VG, Weng H, Tucker JS, Sauer JR & Essenberg RC 2009: Microarray analysis of gene expression changes in feeding female and male lone star ticks, *Amblyomma americanum* (L). *Archives of Insect Biochemistry and Physiology.* 71, 236-253.

Allen JR & Humphreys SJ 1979: Immunisation of guinea pigs and cattle against ticks. *Nature.* 280, 491-493.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ 1990: Basic local alignment search tool. *Journal of Molecular Biology.* 215, 403-410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ 1997: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 25, 3389-3402.

Anatriello E, Ribeiro JMC, de Miranda-Santos IKF, Brandão LG, Anderson JM, Valenzuela JG, Maruyama SR, Silva JS & Ferreira BR 2010: An insight into the

sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. *BMC Genomics.* 11, 450.

Andersen JF, Gudderra NP, Francischetti IMB, Valenzuela JG & Ribeiro JMC 2004: Recognition of anionic phospholipid membranes by an antihemostatic protein from a blood-feeding insect. *Biochemistry.* 43, 6987-6994.

Antunes S, Galindo RC, Almazán C, Rudenko N, Golovchenko M, Grubhoffer L, Shkap V, do Rosário V, de la Fuente J & Domingos A 2012: Functional genomics studies of *Rhipicephalus* (*Boophilus*) *annulatus* ticks in response to infection with the cattle protozoan parasite, *Babesia bigemina*. *International Journal for Parasitology.* 42, 187-195.

Arcà B, Struchiner CJ, Pham VM, Sferra G, Lombardo F, Pombi M & Ribeiro JMC 2014: Positive selection drives accelerated evolution of mosquito salivary genes associated with blood-feeding. *Insect Molecular Biology.* 23, 122-131.

Arrighi RBG, Lycett G, Mahairaki V, Siden-Kiamos I & Louis C 2005: Laminin and the malaria parasite's journey through the mosquito midgut. *Journal of Experimental Biology.* 208, 2497-2502.

Aubry S, Kelly S, Kümpers BMC, Smith-Unna RD & Hibberd JM 2014: Deep evolutionary comparison of gene expression identifies parallel recruitment of *trans*-factors in two independent origins of C_4 photosynthesis. *PLoS Genetics.* 10, e1004365.

Aune TM & Spurlock CF 2016: Long non-coding RNAs in innate and adaptive immunity. *Virus Research.* 212, 146-160.

Ayllón N, Villar M, Busby AT, Kocan KM, Blouin EF, Bonzón-Kulichenko E, Galindo RC, Mangold AJ, Alberdi P, Pérez de la Lastra JM, Vázquez J & de la Fuente J 2013: *Anaplasma phagocytophilum* inhibits apoptosis and promotes cytoskeleton rearrangement for infection of tick cells. *Infection and Immunity.* 81, 2415-2425.

Ayllón N, Villar M, Galindo RC, Kocan KM, Šíma R, López JA, Vázquez J, Alberdi P, Cabezas-Cruz A, Kopáček P & de la Fuente J 2015: Systems biology of tissue-specific response to *Anaplasma phagocytophilum* reveals differentiated apoptosis in the tick vector *Ixodes scapularis*. *PLoS Genetics.* 11, e1005120.

Ayyadevara S, Engle MR, Singh SP, Dandapat A, Lichti CF, Beneš H, Shmookler Reis RJ, Liebau E & Zimniak P 2005: Lifespan and stress resistance of *Caenorhabditis elegans* are increased by expression of glutathione transferases capable of

metabolizing the lipid peroxidation product 4-hydroxynonenal. *Aging Cell.* 4, 257-271.

Barker SC & Murrell A 2008: Systematics and evolution of ticks with a list of valid genus and species names. In Bowman A S & Nuttall P A (eds.): *Ticks: Biology, Disease and Control.* Cambridge: Cambridge University Press, 1-39.

Barnett SF & Brocklesby DW 1966: The passage of "*Theileria lawrencei* (Kenya)" through cattle. *British Veterinary Journal.* 122, 396-409.

Barrero RA, Guerrero FD, Black M, McCooke J, Chapman B, Schilkey F, Pérez de León AA, Miller RJ, Bruns S, Dobry J, Mikhaylenko G, Stormo K, Bell C, Tao Q, Bogden R, Moolhuijzen PM, Hunter A & Bellgard MI 2017: Gene-enriched draft genome of the cattle tick *Rhipicephalus microplus*: assembly by the hybrid Pacific Biosciences/Illumina approach enabled analysis of the highly repetitive genome. *International Journal for Parasitology.* 47, 569-583.

Baxter RHG, Contet A & Krueger K 2017: Arthropod innate immune systems and vector-borne diseases. *Biochemistry.* 56, 907-918.

Beaufays J, Adam B, Menten-Dedoyart C, Fievez L, Grosjean A, Decrem Y, Prévôt P-P, Santini S, Brasseur R, Brossard M, Vanhaeverbeek M, Bureau F, Heinen E, Lins L, Vanhamme L & Godfroid E 2008: Ir-LBP, an *Ixodes ricinus* tick salivary LTB4-binding lipocalin, interferes with host neutrophil function. *PLoS One.* 3, e3987.

Beerntsen BT, James AA & Christensen BM 2000: Genetics of mosquito vector competence. *Microbiology and Molecular Biology Reviews.* 64, 115-137.

Bennett V & Gilligan DM 1993: The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane. *Annual Review of Cell Biology.* 9, 27-66.

Berkvens DL, Pegram RG & Brandt JRA 1995: A study of the diapausing behaviour of *Rhipicephalus appendiculatus* and *R. zambeziensis* under quasi-natural conditions in Zambia. *Medical and Veterinary Entomology.* 9, 307-315.

Binnington KC 1978: Sequential changes in salivary gland structure during attachment and feeding of the cattle tick, *Boophilus microplus*. *International Journal for Parasitology.* 8, 97-115.

Binnington KC & Kemp DH 1980: Role of tick salivary glands in feeding and disease transmission. *Advances in Parasitology.* 18, 315-339.

Bishop R, Lambson B, Wells C, Pandit P, Osaso J, Nkonge C, Morzaria S, Musoke A & Nene V 2002: A cement protein of the tick *Rhipicephalus appendiculatus*, located in

the secretory *e* cell granules of the type III salivary gland acini, induces strong antibody responses in cattle. *International Journal for Parasitology.* 32, 833-842.

Bishop R, Musoke A, Morzaria S, Gardner M & Nene V 2004: *Theileria*: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology.* 129, S271-S283.

Blouin EF & Stoltsz WH 1989: Comparative infection rates of *Theileria parva lawrencei* in salivary glands of *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis*. *Onderstepoort Journal of Veterinary Research.* 54, 211-213.

Bolger AM, Lohse M & Usadel B 2014: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30, 2114-2120.

Bolisetty MT, Rajadinakaran G & Graveley BR 2015: Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biology.* 16, 204.

Bommer UA 2012: Cellular function and regulation of the translationally controlled tumour protein TCTP. *Open Allergy Journal.* 5, 19-32.

Boulter N & Hall R 1999: Immunity and vaccine development in the bovine theilerioses. *Advances in Parasitology.* 44, 41-97.

Brioudes F, Thierry AM, Chambrier P, Mollereau B & Bendahmane M 2010: Translationally controlled tumor protein is a conserved mitotic growth integrator in animals and plants. *Proceedings of the National Academy of Sciences.* 107, 16384-16389.

Brissette CA, Bykowski T, Cooley AE, Bowman A & Stevenson B 2009a: *Borrelia burgdorferi* RevA antigen binds host fibronectin. *Infection and Immunity.* 77, 2802-2812.

Brissette CA, Verma A, Bowman A, Cooley AE & Stevenson B 2009b: The *Borrelia burgdorferi* outer-surface protein ErpX binds mammalian laminin. *Microbiology.* 155, 863-872.

Brooks MJ, Rajasimha HK, Roger JE & Swaroop A 2011: Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes. *Molecular Vision.* 17, 3034.

Bruchhaus I, Roeder T, Rennenberg A & Heussler VT 2007: Protozoan parasites: programmed cell death as a mechanism of parasitism. *Trends in Parasitology.* 23, 376-383.

Bullard RL, Williams J & Karim S 2016: Temporal gene expression analysis and RNA silencing of single and multiple members of gene family in the lone star tick *Amblyomma americanum*. *PLoS One.* 11, e0147966.

Buresova V, Hajdusek O, Franta Z, Sojka D & Kopacek P 2009: IrAM-an α_2-macroglobulin from the hard tick *Ixodes ricinus*: characterization and function in phagocytosis of a potential pathogen *Chryseobacterium indologenes*. *Developmental & Comparative Immunology.* 33, 489-498.

Burger TD, Shao R & Barker SC 2014a: Phylogenetic analysis of mitochondrial genome sequences indicates that the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*, contains a cryptic species. *Molecular Phylogenetics and Evolution.* 76, 241-253.

Burger TD, Shao R, Labruna MB & Barker SC 2014b: Molecular phylogeny of soft ticks (Ixodida: Argasidae) inferred from mitochondrial genome and nuclear rRNA sequences. *Ticks and Tick-Borne Diseases.* 5, 195-207.

Busby AT, Ayllón N, Kocan KM, Blouin EF, de la Fuente G, Galindo RC, Villar M & de la Fuente J 2012: 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.* 26, 92-102.

Cabezas-Cruz A, Alberdi P, Ayllón N, Valdés JJ, Pierce R, Villar M & de la Fuente J 2016a: *Anaplasma phagocytophilum* increases the levels of histone modifying enzymes to inhibit cell apoptosis and facilitate pathogen infection in the tick vector *Ixodes scapularis*. *Epigenetics.* 11, 303-319.

Cabezas-Cruz A, Tonk M, Bouchut A, Pierrot C, Pierce RJ, Kotsyfakis M, Rahnamaeian M, Vilcinskas A, Khalife J & Valdés JJ 2016b: Antiplasmodial activity is an ancient and conserved feature of tick defensins. *Frontiers in Microbiology.* 7, 1682.

Canales M, Enríquez A, Ramos E, Cabrera D, Dandie H, Soto A, Falcón V, Rodríguez M & de la Fuente J 1997: Large-scale production in *Pichia pastoris* of the recombinant vaccine Gavac™ against cattle tick. *Vaccine.* 15, 414-422.

Capella-Gutiérrez S, Silla-Martínez JM & Gabaldón T 2009: TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25, 1972-1973.

Cardoso R, Soares H, Hemphill A & Leitão A 2016: Apicomplexans pulling the strings: manipulation of the host cell cytoskeleton dynamics. *Parasitology.* 143, 957-970.

Chagas AC, Calvo E, Rios-Velásquez CM, Pessoa FAC, Medeiros JF & Ribeiro JMC 2013: A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. *BMC Genomics.* 14, 875.

Chernomor O, von Haeseler A & Minh BQ 2016: Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology.* 65, 997-1008.

Chiera JW, Newson RM & Cunningham MP 1985: The effect of size on feeding and breeding performance of *Rhipicephalus appendiculatus* Neumann. *International Journal of Tropical Insect Science.* 6, 555-560.

Chmelař J, Anderson JM, Mu J, Jochim RC, Valenzuela JG & Kopecký J 2008: Insight into the sialome of the castor bean tick, *Ixodes ricinus*. *BMC Genomics.* 9, 233.

Chmelař J, Kotál J, Kopecký J, Pedra JHF & Kotsyfakis M 2016: All for one and one for all on the tick-host battlefield. *Trends in Parasitology.* 32, 368-377.

Collins LJ, Biggs PJ, Voelckel C & Joly S 2008: An approach to transcriptome analysis of non-model organisms using short-read sequences. *Genome Informatics.* 21, 3-14.

Colombini M, Blachly-Dyson E & Forte M 1996: VDAC, a channel in the outer mitochondrial membrane. *Ion Channels.* 4, 169-202.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M & Robles M 2005: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 21, 3674-3676.

Conrad PA, Stagg DA, Grootenhuis JG, Irvin AD, Newson J, Njamunggeh REG, Rossiter PB & Young AS 1987: Isolation of *Theileria* parasites from African buffalo (*Syncerus caffer*) and characterization with anti-schizont monoclonal antibodies. *Parasitology.* 94, 413-423.

Cotté V, Sabatier L, Schnell G, Carmi-Leroy A, Rousselle JC, Arsène-Ploetze F, Malandrin L, Sertour N, Namane A, Ferquel E & Choumet V 2014: Differential expression of *Ixodes ricinus* salivary gland proteins in the presence of the *Borrelia burgdorferi* sensu lato complex. *Journal of Proteomics.* 96, 29-43.

Cramaro WJ, Hunewald OE, Bell-Sakyi L & Muller CP 2017: Genome scaffolding and annotation for the pathogen vector *Ixodes ricinus* by ultra-long single molecule sequencing. *Parasites & Vectors.* 10, 71.

Dai J, Narasimhan S, Zhang L, Liu L, Wang P & Fikrig E 2010: Tick histamine release factor is critical for *Ixodes scapularis* engorgement and transmission of the Lyme disease agent. *PLoS Pathogens.* 6, e1001205.

Dai J, Wang P, Adusumilli S, Booth CJ, Narasimhan S, Anguita J & Fikrig E 2009: Antibodies against a tick protein, Salp15, protect mice from the Lyme disease agent. *Cell Host & Microbe.* 6, 482-492.

Dai S-X, Zhang A-D & Huang J-F 2012: Evolution, expansion and expression of the Kunitz/BPTI gene family associated with long-term blood feeding in *Ixodes scapularis*. *BMC Evolutionary Biology.* 12, 4.

Dantas-Torres F 2015: Climate change, biodiversity, ticks and tick-borne diseases: the butterfly effect. *International Journal for Parasitology: Parasites and Wildlife.* 4, 452-461.

De Castro JJ 1997: Sustainable tick and tickborne disease control in livestock improvement in developing countries. *Veterinary Parasitology.* 71, 77-97.

de Castro MH, De Klerk D, Pienaar R, Latif AA, Rees DJG & Mans BJ 2016: *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.* 7, 536-548.

de Castro MH, De Klerk D, Pienaar R, Rees DJG & Mans BJ 2017: Sialotranscriptomics of *Rhipicephalus zambeziensis* reveals intricate expression profiles of secretory proteins and suggests tight temporal transcriptional regulation during blood-feeding. *Parasites & Vectors.* 10, 384.

de la Fuente J, Almazán C, Blouin EF, Naranjo V & Kocan KM 2006: Reduction of tick infections with *Anaplasma marginale* and *A. phagocytophilum* by targeting the tick protective antigen subolesin. *Parasitology Research.* 100, 85-91.

de la Fuente J, Antunes S, Bonnet S, Cabezas-Cruz A, Domingos AG, Estrada-Peña A, Johnson N, Kocan KM, Mansfield KL, Nijhof AM, Papa A, Rudenko N, Villar M, Alberdi P, Torina A, Ayllón N, Vancova M, Golovchenko M, Grubhoffer L, Caracappa S, Fooks AR, Gortazar C & Rego ROM 2017: Tick-pathogen interactions and vector competence: identification of molecular drivers for tick-borne diseases. *Frontiers in Cellular and Infection Microbiology.* 7, 114.

de la Fuente J, Blouin EF, Manzano-Roman R, Naranjo V, Almazán C, Pérez de la Lastra JM, Zivkovic Z, Jongejan F & Kocan KM 2007: Functional genomic studies of tick cells in response to infection with the cattle pathogen, *Anaplasma marginale*. *Genomics.* 90, 712-722.

de la Fuente J, Kopáček P, Lew-Tabor A & Maritz-Olivier C 2016a: Strategies for new and improved vaccines against ticks and tick-borne diseases. *Parasite Immunology.* 38, 754-769.

de la Fuente J, Villar M, Cabezas-Cruz A, Estrada-Peña A, Ayllón N & Alberdi P 2016b: Tick-host-pathogen interactions: conflict and cooperation. *PLoS Pathogens.* 12, e1005488.

de la Fuente J, Waterhouse RM, Sonenshine DE, Roe RM, Ribeiro JMC, Sattelle DB & Hill CA 2016c: Tick genome assembled: new opportunities for research on tick-host-pathogen interactions. *Frontiers in Cellular and Infection Microbiology.* 6, 103.

De Marco L, Epis S, Comandatore F, Porretta D, Cafarchia C, Mastrantonio V, Dantas-Torres F, Otranto D, Urbanelli S, Bandi C & Sassera D 2017: Transcriptome of larvae representing the *Rhipicephalus sanguineus* complex. *Molecular and Cellular Probes.* 31, 85-90.

Decrem Y, Beaufays J, Blasioli V, Lahaye K, Brossard M, Vanhamme L & Godfroid E 2008a: A family of putative metalloproteases in the salivary glands of the tick *Ixodes ricinus*. *FEBS Journal.* 275, 1485-1499.

Decrem Y, Mariller M, Lahaye K, Blasioli V, Beaufays J, Zouaoui Boudjeltia K, Vanhaeverbeek M, Cérutti M, Brossard M, Vanhamme L & Godfroid E 2008b: The impact of gene knock-down and vaccination against salivary metalloproteases on blood feeding and egg laying by *Ixodes ricinus*. *International Journal for Parasitology.* 38, 549-560.

Deitsch KW, Lukehart SA & Stringer JR 2009: Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Reviews Microbiology.* 7, 493-503.

Dennis DT & Piesman JF 2005: Overview of tick-borne infections of humans. In Goodman J L, Dennis D T & Sonenshine D E (eds.): *Tick-borne diseases of humans.* Washington: American Society of Microbiology Press, 3-11.

Dinger ME, Pang KC, Mercer TR & Mattick JS 2008: Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Computational Biology.* 4, e1000176.

Dobbelaere D & Heussler V 1999: Transformation of leukocytes by *Theileria parva* and *T. annulata*. *Annual Reviews in Microbiology.* 53, 1-42.

Dolan TT 1986: Chemotherapy of East Coast fever: the long term weight changes, carrier state and disease manifestations of parvaquone treated cattle. *Journal of Comparative Pathology.* 96, 137-146.

Dreher-Lesnick SM, Mulenga A, Simser JA & Azad AF 2006: Differential expression of two glutathione S-transferases identified from the American dog tick, *Dermacentor variabilis*. *Insect Molecular Biology.* 15, 445-453.

Edgar RC 2004: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* 32, 1792-1797.

Egekwu N, Sonenshine DE, Garman H, Barshis DJ, Cox N, Bissinger BW, Zhu J & Roe RM 2016: Comparison of synganglion neuropeptides, neuropeptide receptors and neurotransmitter receptors and their gene expression in response to feeding in *Ixodes scapularis* (Ixodidae) vs. *Ornithodoros turicata* (Argasidae). *Insect Molecular Biology.* 25, 72-92.

Ekblom R & Galindo J 2011: Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity.* 107, 1-15.

Engreitz JM, Ollikainen N & Guttman M 2016: Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology.* 17, 756-770.

Estrada-Pena A 2003: Climate change decreases habitat suitability for some tick species (Acari: Ixodidae) in South Africa. *Onderstepoort Journal of Veterinary Research.* 70, 79-93.

Estrada-Peña A, Venzal JM, Mangold AJ, Cafrune MM & Guglielmone AA 2005: The *Amblyomma maculatum* Koch, 1844 (Acari: Ixodidae: Amblyomminae) tick group: diagnostic characters, description of the larva of *A. parvitarsum* Neumann, 1901, 16S rDNA sequences, distribution and hosts. *Systematic Parasitology.* 60, 99-112.

Etebari K, Asad S, Zhang G & Asgari S 2016: Identification of *Aedes aegypti* long intergenic non-coding RNAs and their association with *Wolbachia* and Dengue virus infection. *PLoS Neglected Tropical Diseases.* 10, e0005069.

Farkaš R 2015: Apocrine secretion: new insights into an old phenomenon. *Biochimica et Biophysica Acta.* 1850, 1740-1750.

Fawcett DW, Büscher G & Doxsey S 1982: Salivary gland of the tick vector of East Coast fever. IV. Cell type selectivity and host cell responses to *Theileria parva*. *Tissue and Cell.* 14, 397-414.

Feldman-Muhsam B, Borut S & Saliternik-Givant S 1970: Salivary secretion of the male tick during copulation. *Journal of Insect Physiology.* 16, 1945-1949.

Fikrig E, Feng W, Barthold SW, Telford SR & Flavell RA 2000: Arthropod- and host-specific *Borrelia burgdorferi bbk32* expression and the inhibition of spirochete transmission. *Journal of Immunology.* 164, 5344-5351.

Findlay GD, Yi X, MacCoss MJ & Swanson WJ 2008: Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biology.* 6, e178.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J & Bateman A 2016: The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research.* 44, D279-D285.

Flower DR 1996: The lipocalin protein family: structure and function. *Biochemical Journal.* 318, 1-14.

Flower DR, North ACT & Attwood TK 1993: Structure and sequence relationships in the lipocalins and related proteins. *Protein Science.* 2, 753-761.

Flower DR, North ACT & Sansom CE 2000: The lipocalin protein family: structural and sequence overview. *Biochimica et Biophysica Acta.* 1482, 9-24.

Francischetti IMB 2010: Platelet aggregation inhibitors from hematophagous animals. *Toxicon.* 56, 1130-1144.

Francischetti IMB, Mather TN & Ribeiro JMC 2003: Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. *Biochemical and Biophysical Research Communications.* 305, 869-875.

Francischetti IMB, Mather TN & Ribeiro JMC 2004: Penthalaris, a novel recombinant five-Kunitz tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick vector of Lyme disease, *Ixodes scapularis*. *Thrombosis and Haemostasis.* 91, 886-898.

Francischetti IMB, Mather TN & Ribeiro JMC 2005a: Tick saliva is a potent inhibitor of endothelial cell proliferation and angiogenesis. *Thrombosis and Haemostasis.* 94, 167-174.

Francischetti IMB, Meng Z, Mans BJ, Gudderra N, Hall M, Veenstra TD, Pham VM, Kotsyfakis M & Ribeiro JMC 2008: An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. *Journal of Proteomics.* 71, 493-512.

Francischetti IMB, Pham VM, Mans BJ, Andersen JF, Mather TN, Lane RS & Ribeiro JMC 2005b: The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). *Insect Biochemistry and Molecular Biology.* 35, 1142-1161.

Francischetti IMB, Sá-Nunes A, Mans BJ, Santos IM & Ribeiro JMC 2009: The role of saliva in tick feeding. *Frontiers in Bioscience.* 14, 2051-2088.

Francischetti IMB, Valenzuela JG, Andersen JF, Mather TN & Ribeiro JMC 2002: Ixolaris, a novel recombinant tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick, *Ixodes scapularis*: identification of factor X and factor Xa as scaffolds for the inhibition of factor VIIa/tissue factor complex. *Blood.* 99, 3602-3612.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K & Chen X 2010: Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Research.* 20, 763-783.

Garcia GR, Gardinassi LG, Ribeiro JMC, Anatriello E, Ferreira BR, Moreira HNS, Mafra C, Martins MM, Szabó MPJ, de Miranda-Santos IKF & Maruyama SR 2014: The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. *Parasites & Vectors.* 7, 430.

García-García JC, Gonzalez IL, González DM, Valdés M, Méndez L, Lamberti J, D'agostino B, Citroni D, Fragoso H, Ortiz M, Rodríguez M & de la Fuente J 1999: Sequence variations in the *Boophilus microplus* Bm86 locus and implications for immunoprotection in cattle vaccinated with this antigen. *Experimental and Applied Acarology.* 23, 883-895.

García-Varas S, Manzano-Román R, Fernández-Soto P, Encinas-Grandes A, Oleaga A & Pérez-Sánchez R 2010: Purification and characterisation of a P-selectin-binding molecule from the salivary glands of *Ornithodoros moubata* that induces protective anti-tick immune responses in pigs. *International Journal for Parasitology.* 40, 313-326.

Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, Wilson RJ, Sato S, Ralph SA, Mann DJ, Xiong Z, Shallom SJ, Weidman J, Jiang L, Lynn J, Weaver B, Shoaibi A, Domingo AR, Wasawo D, Crabtree J, Wortman JR, Haas B, Angiuoli SV, Creasy TH, Lu C, Suh B, Silva JC, Utterback TR, Feldblyum TV, Pertea M, Allen J, Nierman WC, Taracha EL, Salzberg SL, White OR, Fitzhugh HA, Morzaria S, Venter JC, Fraser CM & Nene V

2005: Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science.* 309, 134-137.

Garg R, Juncadella IJ, Ramamoorthi N, Ashish, Ananthanarayanan SK, Thomas V, Rincón M, Krueger JK, Fikrig E, Yengo CM & Anguita J 2006: Cutting edge: CD4 is the receptor for the tick saliva immunosuppressor, Salp15. *Journal of Immunology.* 177, 6579-6583.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD & Bairoch A 2003: ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research.* 31, 3784-3788.

Gauer M, Mackenstedt U, Mehlhorn H, Schein E, Zapf F, Njenga E, Young A & Morzaria S 1995: DNA measurements and ploidy determination of developmental stages in the life cycles of *Theileria annulata* and *T. parva*. *Parasitology Research.* 81, 565-574.

Gebbia JA, Coleman JL & Benach JL 2001: *Borrelia* spirochetes upregulate release and activation of matrix metalloproteinase gelatinase B (MMP-9) and collagenase 1 (MMP-1) in human cells. *Infection and Immunity.* 69, 456-462.

Geraci NS, Johnston JS, Robinson JP, Wikel SK & Hill CA 2007: Variation in genome size of argasid and ixodid ticks. *Insect Biochemistry and Molecular Biology.* 37, 399-408.

Gibson AK, Smith Z, Fuqua C, Clay K & Colbourne JK 2013: Why so many unknown genes? Partitioning orphans from a representative transcriptome of the lone star tick *Amblyomma americanum*. *BMC Genomics.* 14, 135.

Gomes AF & Wouters G 1991: Species identification of *Rhipicephalus duttoni* in relation to the *Rhipicephalus appendiculatus* group (Acarina: Ixodidae). *Journal of Medical Entomology.* 28, 16-18.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A 2011: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology.* 29, 644-652.

Graf JF, Gogolewski R, Leach-Bing N, Sabatini GA, Molento MB, Bordin EL & Arantes GJ 2004: Tick control: an industry point of view. *Parasitology.* 129, S427-S442.

Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Ngoc PCT, Ortego F, Hernández-Crespo P, Diaz I, Martinez M, Navajas M, Sucena É, Magalhães S, Nagy L, Pace RM, Djuranović S, Smagghe G, Iga M, Christiaens O, Veenstra JA, Ewer J, Villalobos RM, Hutter JL, Hudson SD, Velez M, Yi SV, Zeng J, Pires-daSilva A, Roch F, Cazaux M, Navarro M, Zhurov V, Acevedo G, Bjelica A, Fawcett JA, Bonnet E, Martens C, Baele G, Wissler L, Sanchez-Rodriguez A, Tirry L, Blais C, Demeestere K, Henz SR, Gregory TR, Mathieu J, Verdon L, Farinelli L, Schmutz J, Lindquist E, Feyereisen R & Van de Peer Y 2011: The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature.* 479, 487-492.

Guerrero FD, Miller RJ & Pérez de León AA 2012: Cattle tick vaccines: many candidate antigens, but will a commercially viable product emerge? *International Journal for Parasitology.* 42, 421-427.

Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, de la Fuente J, Ribeiro JM, Megy K, Thimmapuram J, Miller JR, Walenz BP, Koren S, Hostetler JB, Thiagarajan M, Joardar VS, Hannick LI, Bidwell S, Hammond MP, Young S, Zeng Q, Abrudan JL, Almeida FC, Ayllón N, Bhide K, Bissinger BW, Bonzon-Kulichenko E, Buckingham SD, Caffrey DR, Caimano MJ, Croset V, Driscoll T, Gilbert D, Gillespie JJ, Giraldo-Calderón GI, Grabowski JM, Jiang D, Khalil SM, Kim D, Kocan KM, Koči J, Kuhn RJ, Kurtti TJ, Lees K, Lang EG, Kennedy RC, Kwon H, Perera R, Qi Y, Radolf JD, Sakamoto JM, Sánchez-Gracia A, Severo MS, Silverman N, Šimo L, Tojo M, Tornador C, Van Zee JP, Vázquez J, Vieira FG, Villar M, Wespiser AR, Yang Y, Zhu J, Arensburger P, Pietrantonio PV, Barker SC, Shao R, Zdobnov EM, Hauser F, Grimmelikhuijzen CJ, Park Y, Rozas J, Benton R, Pedra JH, Nelson DR, Unger MF, Tubio JM, Tu Z, Robertson HM, Shumway M, Sutton G, Wortman JR, Lawson D, Wikel SK, Nene VM, Fraser CM, Collins FH, Birren B, Nelson KE, Caler E & Hill CA 2016: Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nature Communications.* 7, 10507.

Haas BJ, Papanicolaou A, Yassour M, Grabherr MG, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N & Regev A 2013: *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols.* 8, 1494-1512.

Hajdušek O, Šíma R, Ayllón N, Jalovecká M, Perner J, de la Fuente J & Kopáček P 2013: Interaction of the tick immune system with transmitted pathogens. *Frontiers in Cellular and Infection Microbiology.* 3, 26.

Harnnoi T, Sakaguchi T, Nishikawa Y, Xuan X & Fujisaki K 2007: Molecular characterization and comparative study of 6 salivary gland metalloproteases from the hard tick, *Haemaphysalis longicornis*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology.* 147, 93-101.

Hati R, Mitra P, Sarker S & Bhattacharyya KK 1999: Snake venom hemorrhagins. *Critical Reviews in Toxicology.* 29, 1-19.

Hayashida K, Abe T, Weir W, Nakao R, Ito K, Kajino K, Suzuki Y, Jongejan F, Geysen D & Sugimoto C 2013: Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. *DNA Research.* 20, 209-220.

Hayes JD, Flanagan JU & Jowsey IR 2005: Glutathione transferases. *Annual Review of Pharmacology and Toxicology.* 45, 51-88.

He H, Chen AC, Davey RB, Ivie GW & George JE 1999: Characterization and molecular cloning of a glutathione S-transferase gene from the tick, *Boophilus microplus* (Acari: Ixodidae). *Insect Biochemistry and Molecular Biology.* 29, 737-743.

He Y, Rappuoli R, De Groot AS & Chen RT 2010: Emerging vaccine informatics. *Journal of Biomedicine and Biotechnology.* 2010, 218590.

Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A & Teichmann SA 2011: RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular Systems Biology.* 7, 497.

Henderson B, Nair S, Pallas J & Williams MA 2011: Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiology Reviews.* 35, 147-200.

Henson S, Bishop RP, Morzaria S, Spooner PR, Pelle R, Poveda L, Ebeling M, Küng E, Certa U, Daubenberger CA & Qi W 2012: High-resolution genotyping and mapping of recombination and gene conversion in the protozoan *Theileria parva* using whole genome sequencing. *BMC Genomics.* 13, 503.

Heussler V, Sturm A & Langsley G 2006: Regulation of host cell survival by intracellular *Plasmodium* and *Theileria* parasites. *Parasitology.* 132, S49-S60.

Heyne H, Elliott EGR & Bezuidenhout JD 1987: Rearing and infection techniques for *Amblyomma* species to be used in heartwater transmission experiments. *Onderstepoort Journal of Veterinary Research.* 54, 461-471.

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP & Ulitsky I 2015: Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports.* 11, 1110-1122.

Hidano A, Konnai S, Yamada S, Githaka N, Isezaki M, Higuchi H, Nagahata H, Ito T, Takano A, Ando S, Kawabata H, Murata S & Ohahsi K 2014: Suppressive effects of neutrophil by Salp16-like salivary gland proteins from *Ixodes persulcatus* Schulze tick. *Insect Molecular Biology.* 23, 466-474.

Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, Pomerantz AF, Simão FA, Thomas J, Jiggins FM, Murphy TD, Pritham EJ, Robertson HM, Zdobnov EM, Gibbs RA & Richards S 2016: Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized Hox genes and superdynamic intron evolution. *Genome Biology and Evolution.* 8, 1762-1775.

Hyakushima N, Mitsuzawa H, Nishitani C, Sano H, Kuronuma K, Konishi M, Himi T, Miyake K & Kuroki Y 2004: Interaction of soluble form of recombinant extracellular TLR4 domain with MD-2 enables lipopolysaccharide binding and attenuates TLR4-mediated signaling. *Journal of Immunology.* 173, 6949-6954.

Hynes RO & Zhao Q 2000: The evolution of cell adhesion. *Journal of Cell Biology.* 150, F89-F96.

Hynes WL, Stokes MM, Hensley SM, Todd SM & Sonenshine DE 2008: Using RNA interference to determine the role of varisin in the innate immune system of the hard tick *Dermacentor variabilis* (Acari: Ixodidae). *Experimental and Applied Acarology.* 46, 7-15.

Imamura S, Konnai S, Yamada S, Parizi LF, Githaka N, Vaz Ida S, Murata S & Ohashi K 2013: Identification and partial characterization of a gut *Rhipicephalus appendiculatus* cystatin. *Ticks and Tick-Borne Diseases.* 4, 138-144.

Jaworski DC, Zou Z, Bowen CJ, Wasala NB, Madden R, Wang Y, Kocan KM, Jiang H & Dillwith JW 2010: Pyrosequencing and characterization of immune response genes from the American dog tick, *Dermacentor variabilis* (L.). *Insect Molecular Biology.* 19, 617-630.

Johns R, Ohnishi J, Broadwater A, Sonenshine DE, De Silva AM & Hynes WL 2001a: Contrasts in tick innate immune responses to *Borrelia burgdorferi* challenge:

immunotolerance in *Ixodes scapularis* versus immunocompetence in *Dermacentor variabilis* (Acari: Ixodidae). *Journal of Medical Entomology.* 38, 99-107.

Johns R, Sonenshine DE & Hynes WL 2001b: Identification of a defensin from the hemolymph of the American dog tick, *Dermacentor variabilis*. *Insect Biochemistry and Molecular Biology.* 31, 857-865.

Jones LD, Hodgson E, Williams T, Higgs S & Nuttall PA 1992: Saliva activated transmission (SAT) of Thogoto virus: relationship with vector potential of different haematophagous arthropods. *Medical and Veterinary Entomology.* 6, 261-265.

Jongejan F & Uilenberg G 2004: The global importance of ticks. *Parasitology.* 129, S3-S14.

Kaiser MN, Sutherst RW & Bourne AS 1982: Relationship between ticks and Zebu cattle in southern Uganda. *Tropical Animal Health and Production.* 14, 63-74.

Käll L, Krogh A & Sonnhammer EL 2007: Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Research.* 35, W429-W432.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A & Jermiin LS 2017: ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods.* 14, 587-589.

Kamwendo SP, Ingram GA, Musisi FL & Molyneux DH 1993: Haemagglutinin activity in tick (*Rhipicephalus appendiculatus*) haemolymph and extracts of gut and salivary gland. *Annals of Tropical Medicine & Parasitology.* 87, 303-305.

Kamwendo SP, Musisi FL, Trees AJ & Molyneux DH 1995: Effects of haemagglutination (lectin) inhibitory sugars on *Theileria parva* infection in *Rhipicephalus appendiculatus*. *International Journal for Parasitology.* 25, 29-35.

Kanduma EG, Mwacharo JM, Githaka NW, Kinyanjui PW, Njuguna JN, Kamau LM, Kariuki E, Mwaura S, Skilton RA & Bishop RP 2016a: Analyses of mitochondrial genes reveal two sympatric but genetically divergent lineages of *Rhipicephalus appendiculatus* in Kenya. *Parasites & Vectors.* 9, 353.

Kanduma EG, Mwacharo JM, Mwaura S, Njuguna JN, Nzuki I, Kinyanjui PW, Githaka N, Heyne H, Hanotte O, Skilton RA & Bishop RP 2016b: Multi-locus genotyping reveals absence of genetic structure in field populations of the brown ear tick (*Rhipicephalus appendiculatus*) in Kenya. *Ticks and Tick-Borne Diseases.* 7, 26-35.

Kanost MR & Jiang H 2015: Clip-domain serine proteases as immune factors in insect hemolymph. *Current Opinion in Insect Science.* 11, 47-55.

Karim S & Ribeiro JMC 2015: An insight into the sialome of the lone star tick, *Amblyomma americanum*, with a glimpse on its time dependent gene expression. *PLoS One.* 10, e0131292.

Karim S, Singh P & Ribeiro JMC 2011: A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS One.* 6, e28525.

Katzer F, Lizundia R, Ngugi D, Blake D & McKeever D 2011: Construction of a genetic map for *Theileria parva*: identification of hotspots of recombination. *International Journal for Parasitology.* 41, 669-675.

Kawabata S, Tokunaga F, Kugi Y, Motoyama S, Miura Y, Hirata M & Iwanaga S 1996: Limulus factor D, a 43-kDa protein isolated from horseshoe crab hemocytes, is a serine protease homologue with antimicrobial activity. *FEBS Letters.* 398, 146-150.

Kazimírová M & Štibrániová I 2013: Tick salivary compounds: their role in modulation of host defences and pathogen transmission. *Frontiers in Cellular and Infection Microbiology.* 3, 43.

Kemp DH & Bourne A 1980: *Boophilus microplus*: the effect of histamine on the attachment of cattle-tick larvae - studies *in vivo* and *in vitro*. *Parasitology.* 80, 487-496.

Kibuka-Sebitosi E 2006: Potential role of lectins in ticks: *Rhipicephalus appendiculatus* and *Rhipicephalus pulchellus* (Acari: Ixodidae). *Systematic and Applied Acarology Special Publications.* 21, 1-15.

Kim TK, Tirloni L, Pinto AFM, Moresco J, Yates III JR, da Silva Vaz Jr I & Mulenga A 2016: *Ixodes scapularis* tick saliva proteins sequentially secreted every 24 h during blood feeding. *PLoS Neglected Tropical Diseases.* 10, e0004323.

Kim YM, Han YJ, Hwang OJ, Lee SS, Shin AY, Kim SY & Kim JI 2012: Overexpression of *Arabidopsis* translationally controlled tumor protein gene *AtTCTP* enhances drought tolerance with rapid ABA-induced stomatal closure. *Molecules and Cells.* 33, 617-626.

Klompen JSH & Oliver JH 1993: Systematic relationships in the soft ticks (Acari: Ixodida: Argasidae). *Systematic Entomology.* 18, 313-331.

Kocan KM, de la Fuente J, Manzano-Roman R, Naranjo V, Hynes WL & Sonenshine DE 2008: Silencing expression of the defensin, varisin, in male *Dermacentor variabilis* by RNA interference results in reduced *Anaplasma marginale* infections. *Experimental and Applied Acarology.* 46, 17-28.

Kocan KM, Zivkovic Z, Blouin EF, Naranjo V, Almazán C, Mitra R & de la Fuente J 2009: Silencing of genes involved in *Anaplasma marginale*-tick interactions affects the pathogen developmental cycle in *Dermacentor variabilis*. *BMC Developmental Biology*. 9, 42.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L & Gao G 2007: CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research.* 35, W345-W349.

Koonin EV 2005: Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*. 39, 309-338.

Kopáček P, Hajdušek O, Burešová V & Daffre S 2010: Tick innate immunity. In Söderhäll K (ed.) *Invertebrate immunity*. New York: Springer, 137-162.

Kotál J, Langhansová H, Lieskovská J, Andersen JF, Francischetti IMB, Chavakis T, Kopecký J, Pedra JHF, Kotsyfakis M & Chmelař J 2015: Modulation of host immunity by tick saliva. *Journal of Proteomics.* 128, 58-68.

Kotsyfakis M, Kopáček P, Franta Z, Pedra JHF & Ribeiro JMC 2015a: Deep sequencing analysis of the *Ixodes ricinus* haemocytome. *PLoS Neglected Tropical Diseases.* 9, e0003754.

Kotsyfakis M, Schwarz A, Erhart J & Ribeiro JMC 2015b: Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. *Scientific Reports.* 5, 9103.

Kovalev SY & Mukhacheva TA 2012: Phylogeographical structure of the tick *Ixodes persulcatus*: a novel view. *Ticks and Tick-Borne Diseases.* 3, 212-218.

Kovář V, Kopáček P & Grubhoffer L 2000: Isolation and characterization of Dorin M, a lectin from plasma of the soft tick *Ornithodoros moubata*. *Insect Biochemistry and Molecular Biology.* 30, 195-205.

Krogh A, Larsson B, Von Heijne G & Sonnhammer ELL 2001: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology.* 305, 567-580.

Kück P & Meusemann K 2010: FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution.* 56, 1115-1118.

Kuleš J, Horvatić A, Guillemin N, Galan A, Mrljak V & Bhide M 2016: New approaches and omics tools for mining of vaccine candidates against vector-borne diseases. *Molecular BioSystems.* 12, 2680-2694.

Kung JTY, Colognori D & Lee JT 2013: Long noncoding RNAs: past, present, and future. *Genetics.* 193, 651-669.

Kunitz M & Northrop JH 1936: Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *Journal of General Physiology.* 19, 991-1007.

Labuda M, Jones LD, Williams T & Nuttall PA 1993: Enhancement of tick-borne encephalitis virus transmission by tick salivary gland extracts. *Medical and Veterinary Entomology.* 7, 193-196.

Labuda M, Trimnell AR, Ličková M, Kazimírová M, Davies GM, Lissina O, Hails RS & Nuttall PA 2006: An antivector vaccine protects against a lethal vector-borne pathogen. *PLoS Pathogens.* 2, e27.

Landulfo GA, Patané JSL, Silva DGN, Junqueira-de-Azevedo ILM, Mendonca RZ, Simons SM, Carvalho E & Barros-Battesti DM 2017: Gut transcriptome analysis on females of *Ornithodoros mimon* (Acari: Argasidae) and phylogenetic inference of ticks. *Revista Brasileira de Parasitologia Veterinária.* 26, 185-204.

Lane RS 1994: Competence of ticks as vectors of microbial agents with an emphasis on *Borrelia burgdorferi*. In Sonenshine D E & Mather T N (eds.): *Ecological dynamics of tick-borne zoonoses.* New York: Oxford University Press, 45-67.

Langmead B & Salzberg SL 2012: Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 9, 357-359.

Latif AA, Hove T, Kanhai GK, Masaka S & Pegram RG 2001: Epidemiological observations of Zimbabwean theileriosis: disease incidence and pathogenicity in susceptible cattle during *Rhipicephalus appendiculatus* nymphal and adult seasonal activity. *Onderstepoort Journal of Veterinary Research.* 68, 187-195.

Lawrence JA 1979: The differential diagnosis of the bovine theilerias of southern Africa. *Journal of the South African Veterinary Association.* 50, 311-313.

Lawrence JA, de Vos AJ & Irvin AD 1994: Corridor disease. In Coetzer J A W, Thomson G R & Tustin R C (eds.): *Infectious diseases of livestock: with special reference to southern Africa.* Cape Town: Oxford University Press, 326-328.

Lawrence JA, Norval RAI & Uilenberg G 1983: *Rhipicephalus zambeziensis* as a vector of bovine theileriae. *Tropical Animal Health and Production.* 15, 39-42.

Lemasters JJ & Holmuhamedov E 2006: Voltage-dependent anion channel (VDAC) as mitochondrial governator - thinking outside the box. *Biochimica et Biophysica Acta.* 1762, 181-190.

Lew-Tabor AE & Rodriguez-Valle M 2016: A review of reverse vaccinology approaches for the development of vaccines against ticks and tick borne diseases. *Ticks and Tick-Borne Diseases.* 7, 573-585.

Li A, Zhang J & Zhou Z 2014: PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme. *BMC Bioinformatics.* 15, 311.

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

Li L, Stoeckert CJ & Roos DS 2003: OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research.* 13, 2178-2189.

Li W & Godzik A 2006: Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 22, 1658-1659.

Liu L, Dai J, Zhao YO, Narasimhan S, Yang Y, Zhang L & Fikrig E 2012: *Ixodes scapularis* JAK-STAT pathway regulates tick antimicrobial peptides, thereby controlling the agent of human granulocytic anaplasmosis. *Journal of Infectious Diseases.* 206, 1233-1241.

Liu XY & Bonnet SI 2014: Hard tick factors implicated in pathogen transmission. *PLoS Neglected Tropical Diseases.* 8, e2566.

Lushchak VI 2011: Environmentally induced oxidative stress in aquatic animals. *Aquatic Toxicology.* 101, 13-30.

Lynen G, Zeman P, Bakuname C, Di Giulio G, Mtui P, Sanka P & Jongejan F 2008: Shifts in the distributional ranges of *Boophilus* ticks in Tanzania: evidence that a parapatric boundary between *Boophilus microplus* and *B. decoloratus* follows climate gradients. *Experimental and Applied Acarology.* 44, 147-164.

Madder M, Speybroeck N, Bilounga A, Helleputte D & Berkvens D 2005: Survival of unfed *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis* adults. *Medical and Veterinary Entomology.* 19, 245-250.

Madder M, Speybroeck N, Brandt J & Berkvens D 1999: Diapause induction in adults of three *Rhipicephalus appendiculatus* stocks. *Experimental and Applied Acarology.* 23, 961-968.

Madder M, Speybroeck N, Brandt J, Tirry L, Hodek I & Berkvens D 2002: Geographic variation in diapause response of adult *Rhipicephalus appendiculatus* ticks. *Experimental and Applied Acarology.* 27, 209-221.

Mans BJ 2005: Tick histamine-binding proteins and related lipocalins: potential as therapeutic agents. *Current Opinion in Investigational Drugs.* 6, 1131-1135.

Mans BJ 2011: Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *Journal of Innate Immunity.* 3, 41-51.

Mans BJ 2016: Glandular matrices and secretions: blood-feeding arthropods. In Cohen E & Moussian B (eds.): *Extracellular composite matrices in arthropods.* Switzerland: Springer, 625-688.

Mans BJ, Andersen JF, Francischetti IMB, Valenzuela JG, Schwan TG, Pham VM, Garfield MK, Hammer CH & Ribeiro JMC 2008a: Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. *Insect Biochemistry and Molecular Biology.* 38, 42-58.

Mans BJ, de Castro MH, Pienaar R, de Klerk D, Gaven P, Genu S & Latif AA 2016: Ancestral reconstruction of tick lineages. *Ticks and Tick-Borne Diseases.* 7, 509-535.

Mans BJ, de Klerk D, Pienaar R, de Castro MH & Latif AA 2012: The mitochondrial genomes of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae) and *Argas africolumbae* (Ixodoidae: Argasidae): estimation of divergence dates for the major tick lineages and reconstruction of ancestral blood-feeding characters. *PLoS One.* 7, e49461.

Mans BJ, Featherston J, de Castro MH & Pienaar R 2017: Gene duplication and protein evolution in tick-host interactions. *Frontiers in Cellular and Infection Microbiology.* 7, 413.

Mans BJ & Neitz AWH 2004: Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochemistry and Molecular Biology.* 34, 1-17.

Mans BJ, Pienaar R & Latif AA 2015: A review of *Theileria* diagnostics and epidemiology. *International Journal for Parasitology: Parasites and Wildlife.* 4, 104-118.

Mans BJ & Ribeiro JMC 2008a: Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. *Insect Biochemistry and Molecular Biology.* 38, 841-852.

Mans BJ & Ribeiro JMC 2008b: A novel clade of cysteinyl leukotriene scavengers in soft ticks. *Insect Biochemistry and Molecular Biology.* 38, 862-870.

Mans BJ, Ribeiro JMC & Andersen JF 2008b: Structure, function, and evolution of biogenic amine-binding proteins in soft ticks. *Journal of Biological Chemistry.* 283, 18721-18733.

Marcelino I, de Almeida AM, Ventosa M, Pruneau L, Meyer DF, Martinez D, Lefrançois T, Vachiéry N & Coelho AV 2012: Tick-borne diseases in cattle: applications of proteomics to develop new generation vaccines. *Journal of Proteomics.* 75, 4232-4250.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C & Bryant SH 2015: CDD: NCBI's conserved domain database. *Nucleic Acids Research.* 43, D222-D226.

Maritim AC, Young AS, Lesan AC, Ndungu SG, Stagg DA & Ngumi PN 1992: Transformation of *Theileria parva* derived from African buffalo (*Syncerus caffer*) by tick passage in cattle and its use in infection and treatment immunization. *Veterinary Parasitology.* 43, 1-14.

Martin JA & Wang Z 2011: Next-generation transcriptome assembly. *Nature Reviews Genetics.* 12, 671-682.

Martin M 2011: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 17, 10-12.

Martin SJ, O'Brien GA, Nishioka WK, McGahon AJ, Mahboubi A, Saido TC & Green DR 1995: Proteolysis of fodrin (non-erythroid spectrin) during apoptosis. *Journal of Biological Chemistry.* 270, 6425-6428.

Martins LA, Galletti MFBM, Ribeiro JMC, Fujita A, Costa FB, Labruna MB, Daffre S & Fogaça AC 2017: 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.* 7, 129.

Maruyama SR, Anatriello E, Anderson JM, Ribeiro JMC, Brandão LG, Valenzuela JG, Ferreira BR, Garcia GR, Szabó MPJ, Patel S, Bishop R & de Miranda-Santos IKF 2010: The expression of genes coding for distinct types of glycine-rich proteins varies according to the biology of three metastriate ticks, *Rhipicephalus* (*Boophilus*) *microplus*, *Rhipicephalus sanguineus* and *Amblyomma cajennense*. *BMC Genomics.* 11, 363.

Maruyama SR, Garcia GR, Teixeira FR, Brandão LG, Anderson JM, Ribeiro JMC, Valenzuela JG, Horackova J, Veríssimo CJ, Katiki LM, Banin TM, Zangirolamo AF, Gardinassi LG, Ferreira BR & de Miranda-Santos IKF 2017: Mining a differential sialotranscriptome of *Rhipicephalus microplus* guides antigen discovery to formulate a vaccine that reduces tick infestations. *Parasites & Vectors.* 10, 206.

Mbizeni S, Potgieter FT, Troskie C, Mans BJ, Penzhorn BL & Latif AA 2013: Field and laboratory studies on Corridor disease (*Theileria parva* infection) in cattle population at the livestock/game interface of uPhongolo-Mkuze area, South Africa. *Ticks and Tick-Borne Diseases.* 4, 227-234.

McKeever DJ 2007: Live immunisation against *Theileria parva*: containing or spreading the disease? *Trends in Parasitology.* 23, 565-568.

Mekonnen S, Bryson NR, Fourie LJ, Peter RJ, Spickett AM, Taylor RJ, Strydom T & Horak IG 2002: Acaricide resistance profiles of single- and multi-host ticks from communal and commercial farming areas in the Eastern Cape and North-West provinces of South Africa. *Onderstepoort Journal of Veterinary Research.* 69, 99-105.

Mercer TR & Mattick JS 2013: Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*. 20, 300-307.

Merino O, Alberdi P, Pérez de la Lastra JM & de la Fuente J 2013a: Tick vaccines and the control of tick-borne pathogens. *Frontiers in Cellular and Infection Microbiology.* 3, 30.

Merino O, Almazán C, Canales M, Villar M, Moreno-Cid JA, Estrada-Peña A, Kocan KM & de la Fuente J 2011a: Control of *Rhipicephalus* (*Boophilus*) *microplus* infestations by the combination of subolesin vaccination and tick autocidal control after subolesin gene knockdown in ticks fed on cattle. *Vaccine.* 29, 2248-2254.

Merino O, Almazán C, Canales M, Villar M, Moreno-Cid JA, Galindo RC & de la Fuente J 2011b: Targeting the tick protective antigen subolesin reduces vector infestations and pathogen infection by *Anaplasma marginale* and *Babesia bigemina*. *Vaccine.* 29, 8575-8579.

Merino O, Antunes S, Mosqueda J, Moreno-Cid JA, Pérez de la Lastra JM, Rosario-Cruz R, Rodríguez S, Domingos A & de la Fuente J 2013b: Vaccination with proteins involved in tick-pathogen interactions reduces vector infestations and pathogen infection. *Vaccine.* 31, 5889-5896.

Min XJ, Butler G, Storms R & Tsang A 2005: OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research.* 33, W677-W680.

Minh BQ, Nguyen MAT & von Haeseler A 2013: Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution.* 30, 1188-1195.

Minjauw B & McLeod A 2003. Tick-borne diseases and poverty: the impact of ticks and tick-borne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa. Research report DFID Animal Health Programme: Centre for Tropical Veterinary Medicine, University of Edinburgh.

Monteiro RQ, Rezaie AR, Bae JS, Calvo E, Andersen JF & Francischetti IMB 2008: Ixolaris binding to factor X reveals a precursor state of factor Xa heparin-binding exosite. *Protein Science.* 17, 146-153.

Monteiro RQ, Rezaie AR, Ribeiro JMC & Francischetti IMB 2005: Ixolaris: a factor Xa heparin-binding exosite inhibitor. *Biochemical Journal.* 387, 871-877.

Moorhouse DE & Tatchell RJ 1966: The feeding processes of the cattle-tick *Boophilus microplus* (Canestrini): a study in host-parasite relations. *Parasitology.* 56, 623-632.

Mooring MS & Mazhowu W 1995: Geographic distribution of ticks and comparison of tick load on male and female impala from three sites in Zimbabwe. *Zimbabwe Veterinary Journal.* 26, 76-85.

Moreira HNS, Barcelos RM, Vidigal PMP, Klein RC, Montandon CE, Maciel TEF, Carrizo JFA, de Lima PHC, Soares AC, Martins MM & Mafra C 2017: A deep insight into the whole transcriptome of midguts, ovaries and salivary glands of the *Amblyomma sculptum* tick. *Parasitology International.* 66, 64-73.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC & Kanehisa M 2007: KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research.* 35, W182-W185.

Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B 2008: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods.* 5, 621-628.

Morzaria S, Nene V, Bishop R & Musoke A 2000: Vaccines against *Theileria parva*. *Annals of the New York Academy of Sciences.* 916, 464-473.

Mtambo J, Madder M, Van Bortel W, Berkvens D & Backeljau T 2007a: *Rhipicephalus appendiculatus* and *R. zambeziensis* (Acari: Ixodidae) from Zambia: a molecular reassessment of their species status and identification. *Experimental and Applied Acarology.* 41, 115-128.

Mtambo J, Madder M, Van Bortel W, Chaka G, Berkvens D & Backeljau T 2007b: Further evidence for geographic differentiation in *R. appendiculatus* (Acari: Ixodidae) from eastern and southern provinces of Zambia. *Experimental and Applied Acarology.* 41, 129-138.

Mtambo J, Madder M, Van Bortel W, Geysen D, Berkvens D & Backeljau T 2007c: Genetic variation in *Rhipicephalus appendiculatus* (Acari: Ixodidae) from Zambia: correlating genetic and ecological variation with *Rhipicephalus appendiculatus* from eastern and southern Africa. *Journal of Vector Ecology.* 32, 168-175.

Mudenda L, Pierlé SA, Turse JE, Scoles GA, Purvine SO, Nicora CD, Clauss TRW, Ueti MW, Brown WC & Brayton KA 2014: Proteomics informed by transcriptomics identifies novel secreted proteins in *Dermacentor andersoni* saliva. *International Journal for Parasitology.* 44, 1029-1037.

Mukhebi AW, Perry BD & Kruska R 1992: Estimated economics of theileriosis control in Africa. *Preventive Veterinary Medicine.* 12, 73-85.

Mulenga A, Khumthong R & Blandon MA 2007: Molecular and expression analysis of a family of the *Amblyomma americanum* tick Lospins. *Journal of Experimental Biology.* 210, 3188-3198.

Mulenga A, Macaluso KR, Simser JA & Azad AF 2003a: The American dog tick, *Dermacentor variabilis*, encodes a functional histamine release factor homolog. *Insect Biochemistry and Molecular Biology.* 33, 911-919.

Mulenga A, Macaluso KR, Simser JA & Azad AF 2003b: Dynamics of *Rickettsia*-tick interactions: identification and characterization of differentially expressed mRNAs in uninfected and infected *Dermacentor variabilis*. *Insect Molecular Biology.* 12, 185-193.

Mulenga A, Misao O & Sugimoto C 2003c: Three serine proteinases from midguts of the hard tick *Rhipicephalus appendiculatus*; cDNA cloning and preliminary characterization. *Experimental and Applied Acarology.* 29, 151-164.

Mulenga A, Sugimoto C, Sako Y, Ohashi K, Musoke A, Shubash M & Onuma M 1999: Molecular characterization of a *Haemaphysalis longicornis* tick salivary gland-associated 29-kilodalton protein and its effect as a vaccine against tick infestation in rabbits. *Infection and Immunity.* 67, 1652-1658.

Mulenga A, Tsuda A, Onuma M & Sugimoto C 2003d: Four serine proteinase inhibitors (serpin) from the brown ear tick, *Rhiphicephalus appendiculatus*; cDNA cloning and

preliminary characterization. *Insect Biochemistry and Molecular Biology.* 33, 267-276.

Murrell A, Campbell NJH & Barker SC 2001: A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. *Molecular Phylogenetics and Evolution.* 21, 244-258.

Murrell A, Campbell NJH & Barker SC 2003: The value of idiosyncratic markers and changes to conserved tRNA sequences from the mitochondrial genome of hard ticks (Acari: Ixodida: Ixodidae) for phylogenetic inference. *Systematic Biology.* 52, 296-310.

Nappi AJ & Vass E 1998: Hydrogen peroxide production in immune-reactive *Drosophila melanogaster*. *Journal of Parasitology.* 84, 1150-1157.

Narasimhan S, Coumou J, Schuijt TJ, Boder E, Hovius JW & Fikrig E 2014: A tick gut protein with fibronectin III domains aids *Borrelia burgdorferi* congregation to the gut during transmission. *PLoS Pathogens.* 10, e1004278.

Narasimhan S, Koski RA, Beaulieu B, Anderson JF, Ramamoorthi N, Kantor F, Cappello M & Fikrig E 2002: A novel family of anticoagulants from the saliva of *Ixodes scapularis*. *Insect Molecular Biology.* 11, 641-650.

Narasimhan S, Sukumaran B, Bozdogan U, Thomas V, Liang X, DePonte K, Marcantonio N, Koski RA, Anderson JF, Kantor F & Fikrig E 2007: A tick antioxidant facilitates the Lyme disease agent's successful migration from the mammalian host to the arthropod vector. *Cell Host & Microbe.* 2, 7-18.

Nascimento IP & Leite LCC 2012: Recombinant vaccines and the development of new vaccine strategies. *Brazilian Journal of Medical and Biological Research.* 45, 1102-1111.

Nazareth RA, Tomaz LS, Ortiz-Costa S, Atella GC, Ribeiro JMC, Francischetti IMB & Monteiro RQ 2006: Antithrombotic properties of Ixolaris, a potent inhibitor of the extrinsic pathway of the coagulation cascade. *Thrombosis and Haemostasis.* 96, 7-13.

Neelakanta G, Sultana H, Fish D, Anderson JF & Fikrig E 2010: *Anaplasma phagocytophilum* induces *Ixodes scapularis* ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. *Journal of Clinical Investigation.* 120, 3179-3190.

Neitz WO 1957: Theileriosis, gonderioses and cytauxzoonoses: a review. *Onderstpoort Journal of Veterinary Research.* 27, 275-430.

Neitz WO, Canham AS & Kluge EB 1955: Corridor disease: a fatal form of bovine theileriosis encountered in Zululand. *Journal of the South African Veterinary Association.* 26, 79-87.

Nene V, Kiara H, Lacasta A, Pelle R, Svitek N & Steinaa L 2016: The biology of *Theileria parva* and control of East Coast fever - current status and future trends. *Ticks and Tick-Borne Diseases.* 7, 549-564.

Nene V, Lee D, Kang'a S, Skilton R, Shah T, de Villiers E, Mwaura S, Taylor D, Quackenbush J & Bishop R 2004: Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. *Insect Biochemistry and Molecular Biology.* 34, 1117-1128.

Nene V, Lee D, Quackenbush J, Skilton R, Mwaura S, Gardner MJ & Bishop R 2002: AvGI, an index of genes transcribed in the salivary glands of the ixodid tick *Amblyomma variegatum*. *International Journal for Parasitology.* 32, 1447-1456.

Nene V, Musoke A, Gobright E & Morzaria S 1996: Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. *Infection and Immunity.* 64, 2056-2061.

Nguyen L-T, Schmidt HA, von Haeseler A & Minh BQ 2014: IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution.* 32, 268-274.

Nijhof AM, Balk JA, Postigo M & Jongejan F 2009: Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus* (*Boophilus*) *microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. *BMC Molecular Biology.* 10, 112.

Norling M, Bishop RP, Pelle R, Qi W, Henson S, Drábek EF, Tretina K, Odongo D, Mwaura S, Njoroge T, Bongcam-Rudloff E, Daubenberger CA & Silva JC 2015: The genomes of three stocks comprising the most widely utilized live sporozoite *Theileria parva* vaccine exhibit very different degrees and patterns of sequence divergence. *BMC Genomics.* 16, 729.

Norval RAI, Perry BD & Young AS 1992: *The epidemiology of theileriosis in Africa.* London: Academic Press.

Norval RAI, Walker JB & Colborne J 1982: The ecology of *Rhipicephalus zambeziensis* and *Rhipicephalus appendiculatus* (Acarina, Ixodidae) with particular reference to Zimbabwe. *Onderstepoort Journal of Veterinary Research.* 49, 181-190.

Nunn MA, Sharma A, Paesen GC, Adamson S, Lissina O, Willis AC & Nuttall PA 2005: Complement inhibitor of C5 activation from the soft tick *Ornithodoros moubata*. *Journal of Immunology.* 174, 2084-2091.

Nuttall PA & Jones LD 1991: Non-viraemic tick-borne virus transmission: mechanisms and significance. In Dusbábek F & Bukva V (eds.): *Modern acarology Vol. 2.* The Hague: SPB Academic Publishing, 3-6.

Nuttall PA & Labuda M 2004: Tick-host interactions: saliva-activated transmission. *Parasitology.* 129, S177-S189.

Nuttall PA & Labuda M 2008: Saliva-assisted transmission of tick-borne pathogens. In Bowman A S & Nuttall P A (eds.): *Ticks: Biology, Disease and Control.* Cambridge: Cambridge University Press, 205-219.

Nuttall PA, Paesen GC, Lawrie CH & Wang H 2000: Vector-host interactions in disease transmission. *Journal of Molecular Microbiology and Biotechnology.* 2, 381-386.

Nuttall PA, Trimnell AR, Kazimirova M & Labuda M 2006: Exposed and concealed antigens as vaccine targets for controlling ticks and tick-borne diseases. *Parasite Immunology.* 28, 155-163.

O'Neil ST & Emrich SJ 2013: Assessing *de novo* transcriptome assembly metrics for consistency and utility. *BMC Genomics.* 14, 465.

Ochanda H, Young AS, Medley GF & Perry BD 1998: Vector competence of 7 rhipicephalid tick stocks in transmitting 2 *Theileria parva* parasite stocks from Kenya and Zimbabwe. *Parasitology.* 116, 539-545.

Oleaga A, Obolo-Mvoulouga P, Manzano-Román R & Pérez-Sánchez R 2017: Functional annotation and analysis of the *Ornithodoros moubata* midgut genes differentially expressed after blood feeding. *Ticks and Tick-Borne Diseases.* 8, 693-708.

Oliveira CJF, Sá-Nunes A, Francischetti IMB, Carregaro V, Anatriello E, Silva JS, de Miranda Santos IKF, Ribeiro JMC & Ferreira BR 2011: Deconstructing tick saliva: non-protein molecules with potent immunomodulatory properties. *Journal of Biological Chemistry.* 286, 10960-10969.

Olwoch JM, Reyers B, Engelbrecht FA & Erasmus BFN 2008: Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. *Journal of Arid Environments.* 72, 108-120.

Ong CT, Rodriguez-Valle M, Moolhuijzen PM, Barrero RA, Hunter A, Szabo T, Bellgard MI & Lew-Tabor AE 2016: Exploring the transcriptomic data of the

Australian paralysis tick, *Ixodes holocyclus*. *GSTF Journal of Veterinary Science.* 3, 1.

Paesen GC, Adams PL, Harlos K, Nuttall PA & Stuart DI 1999: Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Molecular Cell.* 3, 661-671.

Paesen GC, Adams PL, Nuttall PA & Stuart DL 2000: Tick histamine-binding proteins: lipocalins with a second binding cavity. *Biochimica et Biophysica Acta.* 1482, 92-101.

Paesen GC, Siebold C, Dallas ML, Peers C, Harlos K, Nuttall PA, Nunn MA, Stuart DI & Esnouf RM 2009: An ion-channel modulator from the saliva of the brown ear tick has a highly modified Kunitz/BPTI structure. *Journal of Molecular Biology.* 389, 734-747.

Paesen GC, Siebold C, Harlos K, Peacey MF, Nuttall PA & Stuart DI 2007: A tick protein with a modified Kunitz fold inhibits human tryptase. *Journal of Molecular Biology.* 368, 1172-1186.

Pagel Van Zee J, Geraci NS, Guerrero FD, Wikel SK, Stuart JJ, Nene VM & Hill CA 2007: Tick genomics: the *Ixodes* genome project and beyond. *International Journal for Parasitology.* 37, 1297-1305.

Parizi LF, Reck J, Oldiges DP, Guizzo MG, Seixas A, Logullo C, de Oliveira PL, Termignoni C, Martins JR & Vaz Ida S 2012: Multi-antigenic vaccine against the cattle tick *Rhipicephalus* (*Boophilus*) *microplus*: a field evaluation. *Vaccine.* 30, 6912-6917.

Parra G, Bradnam K & Korf I 2007: CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics.* 23, 1061-1067.

Pedra JHF, Narasimhan S, Rendić D, DePonte K, Bell-Sakyi L, Wilson IBH & Fikrig E 2010: Fucosylation enhances colonization of ticks by *Anaplasma phagocytophilum*. *Cellular Microbiology.* 12, 1222-1234.

Pelle R, Graham SP, Njahira MN, Osaso J, Saya RM, Odongo DO, Toye PG, Spooner PR, Musoke AJ, Mwangi DM, Taracha ELN, Morrison WI, Weir W, Silva JC & Bishop RP 2011: Two *Theileria parva* CD8 T cell antigen genes are more variable in buffalo than cattle parasites, but differ in pattern of sequence diversity. *PLoS One.* 6, e19015.

Pereira LS, Oliveira PL, Barja-Fidalgo C & Daffre S 2001: Production of reactive oxygen species by hemocytes from the cattle tick *Boophilus microplus*. *Experimental Parasitology.* 99, 66-72.

Perry BD & Young AS 1993: The naming game: the changing fortunes of East Coast fever and *Theileria parva*. *Veterinary Record.* 133, 613-616.

Petersen TN, Brunak S, von Heijne G & Nielsen H 2011: SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods.* 8, 785-786.

Potgieter FT, Stoltsz WH, Blouin EF & Roos JA 1988: Corridor disease in South Africa: a review of the current status. *Journal of the South African Veterinary Association.* 59, 155-160.

Preston SG, Majtán J, Kouremenou C, Rysnik O, Burger LF, Cabezas-Cruz A, Guzman MC, Nunn MA, Paesen GC, Nuttall PA & Austyn JM 2013: Novel immunomodulators from hard ticks selectively reprogramme human dendritic cell responses. *PLoS Pathogens.* 9, e1003450.

Prevot PP, Couvreur B, Denis V, Brossard M, Vanhamme L & Godfroid E 2007: Protective immunity against *Ixodes ricinus* induced by a salivary serpin. *Vaccine.* 25, 3284-3292.

Purnell RE & Joyner LP 1968: The development of *Theileria parva* in the salivary glands of the tick, *Rhipicephalus appendiculatus*. *Parasitology.* 58, 725-732.

Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Kirimi IM, Purnell RE & Young AS 1975: East Coast fever: 1. Chemoprophylactic immunization of cattle against *Theileria parva* (Muguga) and five theilerial strains. *Veterinary Parasitology.* 1, 35-41.

Radley DE, Young AS, Grootenhuis JG, Cunningham MP, Dolan TT & Morzaria SP 1979: Further studies on the immunization of cattle against *Theileria lawrencei* by infection and chemoprophylaxis. *Veterinary Parasitology.* 5, 117-128.

Rahman S, Shering M, Ogden NH, Lindsay R & Badawi A 2016: Toll-like receptor cascade and gene polymorphism in host-pathogen interaction in Lyme disease. *Journal of Inflammation Research.* 9, 91-102.

Ramamoorthi N, Narasimhan S, Pal U, Bao F, Yang XF, Fish D, Anguita J, Norgard MV, Kantor FS, Anderson JF, Koski RA & Fikrig E 2005: The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature.* 436, 573-577.

Ramos OHP, Carmona AK & Selistre-de-Araujo HS 2003: Expression, refolding, and in vitro activation of a recombinant snake venom pro-metalloprotease. *Protein Expression and Purification.* 28, 34-41.

Randolph SE 2004: Tick ecology: processes and patterns behind the epidemiological risk posed by ixodid ticks as vectors. *Parasitology.* 129, S37-S65.

Rappuoli R 2000: Reverse vaccinology. *Current Opinion in Microbiology.* 3, 445-450.

Rechav Y 1982: Dynamics of tick populations (Acari: Ixodidae) in the Eastern Cape province of South Africa. *Journal of Medical Entomology.* 19, 679-700.

Rego ROM, Kovar V, Kopáček P, Weise C, Man P, Šauman I & Grubhoffer L 2006: The tick plasma lectin, Dorin M, is a fibrinogen-related molecule. *Insect Biochemistry and Molecular Biology.* 36, 291-299.

Rhoads A & Au KF 2015: PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics.* 13, 278-289.

Ribas-Aparicio RM, Castelán-Vega JA, Jiménez-Alberto A, Monterrubio-López GP & Aparicio-Ozores G 2017: The impact of bioinformatics on vaccine design and development. In Afrin F, Hemeg H & Ozbak H (eds.): *Vaccines.* Rijeka: InTech, 7.

Ribeiro JMC, Alarcon-Chaidez F, Francischetti IMB, Mans BJ, Mather TN, Valenzuela JG & Wikel SK 2006: An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochemistry and Molecular Biology.* 36, 111-129.

Ribeiro JMC, Anderson JM, Manoukis NC, Meng Z & Francischetti IMB 2011: A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. *BMC Genomics.* 12, 136.

Ribeiro JMC & Mather TN 1998: *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. *Experimental Parasitology.* 89, 213-221.

Ribeiro JMC, Schneider M & Guimaraes JA 1995: Purification and characterization of prolixin S (nitrophorin 2), the salivary anticoagulant of the blood-sucking bug *Rhodnius prolixus*. *Biochemical Journal.* 308, 243-249.

Ribeiro JMC, Slovák M & Francischetti IMB 2017: An insight into the sialome of *Hyalomma excavatum*. *Ticks and Tick-Borne Diseases.* 8, 201-207.

Ribet D & Cossart P 2015: How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes and Infection.* 17, 173-183.

Riding GA, Jarmey J, McKenna RV, Pearson R, Cobon GS & Willadsen P 1994: A protective "concealed" antigen from *Boophilus microplus*. Purification, localization, and possible function. *Journal of Immunology.* 153, 5158-5166.

Robinson MD, McCarthy DJ & Smyth GK 2010: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26, 139-140.

Robinson MD & Oshlack A 2010: A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology.* 11, R25.

Rodriguez-Valle M, Moolhuijzen P, Barrero RA, Ong CT, Busch G, Karbanowicz T, Booth M, Clark R, Koehbach J, Ijaz H, Broady K, Agnew K, Knowles AG, Bellgard MI & Tabor AE 2017: Transcriptome and toxin family analysis of the paralysis tick, *Ixodes holocyclus*. *International Journal for Parasitology*. In press, doi: 10.1016/j.ijpara.2017.1007.1007.

Rogelj B, Godin KS, Shaw CE & Ule J 2012: The functions of glycine-rich regions in TDP-43, FUS and related RNA-binding proteins. In Lorkovic Z J (ed.) *RNA Binding Proteins.* Austin: Landes Bioscience, 1-17.

Rosa de Lima MF, Sanchez Ferreira CA, Joaquim de Freitas DR, Valenzuela JG & Masuda A 2002: Cloning and partial characterization of a *Boophilus microplus* (Acari: Ixodidae) glutathione S-transferase. *Insect Biochemistry and Molecular Biology.* 32, 747-754.

Rossetto CC & Pari GS 2011: Kaposi's sarcoma-associated herpesvirus noncoding polyadenylated nuclear RNA interacts with virus- and host cell-encoded proteins and suppresses expression of genes involved in immune modulation. *Journal of Virology.* 85, 13290-13297.

Rudenko N, Golovchenko M, Edwards MJ & Grubhoffer L 2005: Differential expression of *Ixodes ricinus* tick genes induced by blood feeding or *Borrelia burgdorferi* infection. *Journal of Medical Entomology.* 42, 36-41.

Sangamnatdej S, Paesen GC, Slovak M & Nuttall PA 2002: A high affinity serotonin- and histamine-binding lipocalin from tick saliva. *Insect Molecular Biology.* 11, 79-86.

Santos IK, Valenzuela JG, Ribeiro JMC, De Castro M, Costa JN, Costa AM, da Silva ER, Neto OB, Rocha C, Daffre S, Ferreira BR, da Silva JS, Szabó MP & Bechara GH 2004: Gene discovery in *Boophilus microplus*, the cattle tick: the transcriptomes of ovaries, salivary glands, and hemocytes. *Annals of the New York Academy of Sciences.* 1026, 242-246.

Sasaki SD, de Lima CA, Lovato DV, Juliano MA, Torquato RJS & Tanaka AS 2008: BmSI-7, a novel subtilisin inhibitor from *Boophilus microplus*, with activity toward

Pr1 proteases from the fungus *Metarhizium anisopliae*. *Experimental Parasitology.* 118, 214-220.

Scaria V & Pasha A 2013: Long non-coding RNAs in infection biology. *Frontiers in Genetics.* 3, 308.

Schuijt TJ, Coumou J, Narasimhan S, Dai J, DePonte K, Wouters D, Brouwer M, Oei A, Roelofs JJTH, van Dam AP, van der Poll T, Van't Veer C, Hovius JW & Fikrig E 2011: A tick mannose-binding lectin inhibits the vertebrate complement cascade to enhance transmission of the Lyme disease agent. *Cell Host & Microbe.* 10, 136-146.

Schwarz A, Cabezas-Cruz A, Kopecký J & Valdés JJ 2014a: Understanding the evolutionary structural variability and target specificity of tick salivary Kunitz peptides using next generation transcriptome data. *BMC Evolutionary Biology.* 14, 4.

Schwarz A, Tenzer S, Hackenberg M, Erhart J, Gerhold-Ay A, Mazur J, Kuharev J, Ribeiro JMC & Kotsyfakis M 2014b: A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. *Molecular & Cellular Proteomics.* 13, 2725-2735.

Schwarz A, Valdés JJ & Kotsyfakis M 2012: The role of cystatins in tick physiology and blood feeding. *Ticks and Tick-Borne Diseases.* 3, 117-127.

Schwarz A, von Reumont BM, Erhart J, Chagas AC, Ribeiro JMC & Kotsyfakis M 2013: *De novo Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *FASEB Journal.* 27, 4745-4756.

Seib KL, Zhao X & Rappuoli R 2012: Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clinical Microbiology and Infection.* 18, 109-116.

Shaw MK, Tilney LG & McKeever DJ 1993: Tick salivary gland extract and interleukin-2 stimulation enhance susceptibility of lymphocytes to infection by *Theileria parva* sporozoites. *Infection and Immunity.* 61, 1486-1495.

Shaw MK & Young AS 1995: Differential development and emission of *Theileria parva* sporozoites from the salivary gland of *Rhipicephalus appendiculatus*. *Parasitology.* 111, 153-160.

Shimada O, Ishikawa H, Tosaka-Shimada H & Atsumi S 1999: Rearrangements of actin cytoskeleton during infection with *Escherichia coli* O157 in macrophages. *Cell Structure and Function.* 24, 237-246.

Short NJ & Norval RAI 1981: The seasonal activity of *Rhipicephalus appendiculatus* Neumann 1901 (Acarina: Ixodidae) in the highveld of Zimbabwe Rhodesia. *Journal of Parasitology.* 67, 77-84.

Sibeko KP, Geysen D, Oosthuizen MC, Matthee CA, Troskie M, Potgieter FT, Coetzer JAW & Collins NE 2010: Four p67 alleles identified in South African *Theileria parva* field samples. *Veterinary Parasitology.* 167, 244-254.

Sibeko KP, Oosthuizen MC, Collins NE, Geysen D, Rambritch NE, Latif AA, Groeneveld HT, Potgieter FT & Coetzer JAW 2008: Development and evaluation of a real-time polymerase chain reaction test for the detection of *Theileria parva* infections in Cape buffalo (*Syncerus caffer*) and cattle. *Veterinary Parasitology.* 155, 37-48.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV & Zdobnov EM 2015: BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31, 3210-3212.

Simpson SD, Ramsdell JS, Watson WH & Chabot CC 2017: The draft genome and transcriptome of the Atlantic horseshoe crab, *Limulus polyphemus*. *International Journal of Genomics.* 2017, 7636513.

Simser JA, Macaluso KR, Mulenga A & Azad AF 2004a: Immune-responsive lysozymes from hemocytes of the American dog tick, *Dermacentor variabilis* and an embryonic cell line of the Rocky Mountain wood tick, *D. andersoni*. *Insect Biochemistry and Molecular Biology.* 34, 1235-1246.

Simser JA, Mulenga A, Macaluso KR & Azad AF 2004b: An immune responsive factor D-like serine proteinase homologue identified from the American dog tick, *Dermacentor variabilis*. *Insect Molecular Biology.* 13, 25-35.

Sitt T, Poole EJ, Ndambuki G, Mwaura S, Njoroge T, Omondi GP, Mutinda M, Mathenge J, Prettejohn G, Morrison WI & Toye P 2015: Exposure of vaccinated and naive cattle to natural challenge from buffalo-derived *Theileria parva*. *International Journal for Parasitology: Parasites and Wildlife.* 4, 244-251.

Slovák M, Štibrániová I, Hajnická V & Nuttall PA 2014: Antiplatelet-derived growth factor (PDGF) activity in the saliva of ixodid ticks is linked with their long mouthparts. *Parasite Immunology.* 36, 32-42.

Smith-Unna R, Boursnell C, Patro R, Hibberd JM & Kelly S 2016: TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Research.* 26, 1134-1144.

Sonenshine DE 1991: *Biology of ticks.* Oxford: Oxford University Press.

Sonenshine DE, Bissinger BW, Egekwu N, Donohue KV, Khalil SM & Roe RM 2011: First transcriptome of the testis-vas deferens-male accessory gland and proteome of

the spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). *PLoS One.* 6, e24711.

Speybroeck N, Madder M, Thulke HH, Mtambo J, Tirry L, Chaka G, Marcotty T & Berkvens D 2004: Variation in body size in the tick complex *Rhipicephalus appendiculatus*/*Rhipicephalus zambeziensis*. *Journal of Vector Ecology.* 29, 347-354.

Speybroeck N, Madder M, Van Den Bossche P, Mtambo J, Berkvens N, Chaka G, Mulumba M, Brandt J, Tirry L & Berkvens D 2002: Distribution and phenology of ixodid ticks in southern Zambia. *Medical and Veterinary Entomology.* 16, 430-441.

Stevenson KJ & Poller L 1982: The procoagulant activity of partial thromboplastin extracts: the role of phosphatidyl serine. *Thrombosis Research.* 26, 341-350.

Stoltsz WH 1989: Theileriosis in South Africa: a brief review. *Revue Scientifique et Technique (International Office of Epizootics).* 8, 93-102.

Sukumaran B, Narasimhan S, Anderson JF, DePonte K, Marcantonio N, Krishnan MN, Fish D, Telford SR, Kantor FS & Fikrig E 2006: An *Ixodes scapularis* protein required for survival of *Anaplasma phagocytophilum* in tick salivary glands. *Journal of Experimental Medicine.* 203, 1507-1517.

Sultana H, Neelakanta G, Kantor FS, Malawista SE, Fish D, Montgomery RR & Fikrig E 2010: *Anaplasma phagocytophilum* induces actin phosphorylation to selectively regulate gene transcription in *Ixodes scapularis* ticks. *Journal of Experimental Medicine.* 207, 1727-1743.

Tan AWL, Francischetti IMB, Slovak M, Kini RM & Ribeiro JMC 2015a: Sexual differences in the sialomes of the zebra tick, *Rhipicephalus pulchellus*. *Journal of Proteomics.* 117, 120-144.

Tan Q-Q, Zhu L, Li Y, Liu W, Ma W-H, Lei C-L & Wang X-P 2015b: A *de novo* transcriptome and valid reference genes for quantitative real-time PCR in *Colaphellus bowringi*. *PLoS One.* 10, e0118693.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL & Nikolskaya AN 2003: The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 4, 41.

Taylor D 2006: Innate immunity in ticks: a review. *Journal of the Acarological Society of Japan.* 15, 109-127.

Thompson BE, Latif AA, Oosthuizen MC, Troskie M & Penzhorn BL 2008: Occurrence of *Theileria parva* infection in cattle on a farm in the Ladysmith district,

KwaZulu-Natal, South Africa. *Journal of the South African Veterinary Association.* 79, 31-35.

Tirloni L, Reck J, Terra RMS, Martins JR, Mulenga A, Sherman NE, Fox JW, Yates III JR, Termignoni C, Pinto AFM & Vaz Ida S 2014a: Proteomic analysis of cattle tick *Rhipicephalus (Boophilus) microplus* saliva: a comparison between partially and fully engorged females. *PLoS One.* 9, e94831.

Tirloni L, Seixas A, Mulenga A, Vaz Ida S & Termignoni C 2014b: A family of serine protease inhibitors (serpins) in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus*. *Experimental Parasitology.* 137, 25-34.

Tonk M, Cabezas-Cruz A, Valdés JJ, Rego ROM, Chrudimská T, Strnad M, Šíma R, Bell-Sakyi L, Franta Z, Vilcinskas A, Grubhoffer L & Rahnamaeian M 2014: Defensins from the tick *Ixodes scapularis* are effective against phytopathogenic fungi and the human bacterial pathogen *Listeria grayi*. *Parasites & Vectors.* 7, 554.

Tonk M, Cabezas-Cruz A, Valdés JJ, Rego ROM, Grubhoffer L, Estrada-Pena A, Vilcinskas A, Kotsyfakis M & Rahnamaeian M 2015: *Ixodes ricinus* defensins attack distantly-related pathogens. *Developmental & Comparative Immunology*. 53, 358-365.

Trager W 1939a: Acquired immunity to ticks. *Journal of Parasitology*. 25, 57-81.

Trager W 1939b: Further observations on acquired immunity to the tick *Dermacentor variabilis* Say. *Journal of Parasitology.* 25, 137-139.

Tretina K, Gotia HT, Mann DJ & Silva JC 2015: *Theileria*-transformed bovine leukocytes have cancer hallmarks. *Trends in Parasitology.* 31, 306-314.

Trimnell AR, Davies GM, Lissina O, Hails RS & Nuttall PA 2005: A cross-reactive tick cement antigen is a candidate broad-spectrum tick vaccine. *Vaccine.* 23, 4329-4341.

Trimnell AR, Hails RS & Nuttall PA 2002: Dual action ectoparasite vaccine targeting ‘exposed’ and ‘concealed’ antigens. *Vaccine.* 20, 3560-3568.

Tsuji N, Battsetseg B, Boldbaatar D, Miyoshi T, Xuan X, Oliver JH & Fujisaki K 2007: Babesial vector tick defensin against *Babesia* sp. parasites. *Infection and Immunity.* 75, 3633-3640.

Tsuji N, Miyoshi T, Battsetseg B, Matsuo T, Xuan X & Fujisaki K 2008: A cysteine protease is critical for *Babesia* spp. transmission in *Haemaphysalis* ticks. *PLoS Pathogens.* 4, e1000062.

Turner AJ, Brown CD, Carson JA & Barnes K 2000: The neprilysin family in health and disease. *Advances in Experimental Medicine and Biology.* 447, 229-240.

Uilenberg G 1981: Theilerial species of domestic livestock. In Irvin A D, Cunningham M P & Young A S (eds.): *Advances in the control of Theileriosis.* Dordrecht: Springer, 4-37.

Uilenberg G 1999: Immunization against diseases caused by *Theileria parva*: a review. *Tropical Medicine & International Health.* 4, A12-A20.

Uilenberg G, Silayo RS, Mpangala C, Tondeur W, Tatchell RJ & Sanga HJ 1977: Studies on Theileriidae (Sporozoa) in Tanzania. X. A large-scale field trial on immunization against cattle theileriosis. *Tropenmedizin und Parasitologie.* 28, 499-506.

Ulitsky I 2016: Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics.* 17, 601-614.

Ullmann AJ, Lima CMR, Guerrero FD, Piesman J & Black WCT 2005: Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the southern cattle tick, *Boophilus microplus*. *Insect Molecular Biology.* 14, 217-222.

Valdés JJ, Schwarz A, Cabeza de Vaca I, Calvo E, Pedra JHF, Guallar V & Kotsyfakis M 2013: Tryptogalinin is a tick Kunitz serine protease inhibitor with a unique intrinsic disorder. *PLoS One.* 8, e62562.

Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Mather TN & Ribeiro JMC 2002: Exploring the sialome of the tick *Ixodes scapularis*. *Journal of Experimental Biology*. 205, 2843-2864.

Van Dongen SM 2000: Graph clustering by flow simulation. PhD thesis. University of Utrecht.

Verma A, Brissette CA, Bowman A & Stevenson B 2009: *Borrelia burgdorferi* BmpA is a laminin-binding protein. *Infection and Immunity.* 77, 4940-4946.

Villar M, Ayllón N, Alberdi P, Moreno A, Moreno M, Tobes R, Mateos-Hernández L, Weisheit S, Bell-Sakyi L & de la Fuente J 2015: Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. *Molecular & Cellular Proteomics.* 14, 3154-3172.

Villar M, Ayllón N, Busby AT, Galindo RC, Blouin EF, Kocan KM, Bonzón-Kulichenko E, Zivkovic Z, Almazán C, Torina A, Vazquez J & de la Fuente J 2010: Expression of heat shock and other stress response proteins in ticks and cultured tick cells in response to *Anaplasma* spp. infection and heat shock. *International Journal of Proteomics.* 2010, 657261.

Vlachou D, Lycett G, Sidén-Kiamos I, Blass C, Sinden RE & Louis C 2001: *Anopheles gambiae* laminin interacts with the P25 surface protein of *Plasmodium berghei* ookinetes. *Molecular and Biochemical Parasitology.* 112, 229-237.

Wagner GP, Kin K & Lynch VJ 2012: Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences.* 131, 281-285.

Walker AR, Bouattour A, Camicas JL, Estrada-Peña A, Horak IG, Latif AA, Pegram RG & Preston PM 2003: *Ticks of domestic animals in Africa: a guide to identification of species.* Edinburgh: Bioscience Reports.

Walker JB, Keirans JE & Horak IG 2005: *The genus Rhipicephalus (Acari, Ixodidae): a guide to the brown ticks of the world.* Cambridge: Cambridge University Press.

Walker JB, Norval RAI & Corwin MD 1981: *Rhipicephalus zambeziensis* sp. nov., a new tick from eastern and southern Africa, together with a redescription of *Rhipicephalus appendiculatus* Neumann, 1901 (Acarina, Ixodidae). *Onderstepoort Journal of Veterinary Research.* 48, 87-104.

Wang H & Nuttall PA 1995: Immunoglobulin-G binding proteins in the ixodid ticks, *Rhipicephalus appendiculatus*, *Amblyomma variegatum* and *Ixodes hexagonus*. *Parasitology.* 111, 161-165.

Wang H, Paesen GC, Nuttall PA & Barbour AG 1998: Male ticks help their mates to feed. *Nature.* 391, 753-754.

Wang L, Park HJ, Dasari S, Wang S, Kocher J-P & Li W 2013: CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research.* 41, e74.

Wang Z, Gerstein M & Snyder M 2009: RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics.* 10, 57-63.

Warburg A, Shtern A, Cohen N & Dahan N 2007: Laminin and a *Plasmodium* ookinete surface protein inhibit melanotic encapsulation of Sephadex beads in the hemocoel of mosquitoes. *Microbes and Infection.* 9, 192-199.

Willadsen P 2004: Anti-tick vaccines. *Parasitology.* 129, S367-S387.

Willadsen P 2006: Tick control: thoughts on a research agenda. *Veterinary Parasitology.* 138, 161-168.

Willadsen P, Bird P, Cobon GS & Hungerford J 1995: Commercialisation of a recombinant vaccine against *Boophilus microplus*. *Parasitology.* 110, S43-S50.

Willadsen P, Riding GA, McKenna RV, Kemp DH, Tellam RL, Nielsen JN, Lahnstein J, Cobon GS & Gough JM 1989: Immunologic control of a parasitic arthropod. Identification of a protective antigen from *Boophilus microplus*. *Journal of Immunology.* 143, 1346-1351.

Wilusz JE, Sunwoo H & Spector DL 2009: Long noncoding RNAs: functional surprises from the RNA world. *Genes & Development.* 23, 1494-1504.

Wolfsberg TG, Primakoff P, Myles DG & White JM 1995: ADAM, a novel family of membrane proteins containing a disintegrin and metalloprotease domain: multipotential functions in cell-cell and cell-matrix interactions. *Journal of Cell Biology.* 131, 275-278.

Wouters G, Brandt J & Berkvens D 1987: Species characterisation of *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis* (Ixodidae: Acarina) by enzyme-electrophoresis. *Annales de la Societe Belge de Medecine Tropicale.* 67, 267-270.

Xiang FY, Zhou YZ & Zhou JL 2012: Identification of differentially expressed genes in the salivary gand of *Rhipicephalus haemaphysaloides* by the suppression subtractive hybridization approach. *Journal of Integrative Agriculture.* 11, 1528-1536.

Xu J & Mosher D 2011: Fibronectin and other adhesive glycoproteins. In Mecham R P (ed.) *The extracellular matrix: an overview.* Berlin: Springer, 41-75.

Xu X-L, Cheng T-Y, Yang H, Yan F & Yang Y 2015: *De novo* sequencing, assembly and analysis of salivary gland transcriptome of *Haemaphysalis flava* and identification of sialoprotein genes. *Infection, Genetics and Evolution.* 32, 135-142.

Yamamoto Y, Watabe S, Kageyama T & Takahashi SY 1999: Purification and characterization of *Bombyx* cysteine proteinase specific inhibitors from the hemolymph of *Bombyx mori*. *Archives of Insect Biochemistry and Physiology.* 42, 119-129.

Yan H, Jia H, Gao H, Guo X & Xu B 2013: Identification, genomic organization, and oxidative stress response of a sigma class glutathione S-transferase gene (AccGSTS1) in the honey bee, *Apis cerana cerana*. *Cell Stress and Chaperones.* 18, 415-426.

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L & Wang J 2006: WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research.* 34, W293-W297.

Yi HY, Chowdhury M, Huang YD & Yu XQ 2014: Insect antimicrobial peptides and their applications. *Applied Microbiology and Biotechnology*. 98, 5807-5822.

Young A 1981: The epidemiology of theileriosis in East Africa. In Irvin A D, Cunningham M P & Young A S (eds.): *Advances in the control of Theileriosis.* Dordrecht: Springer, 38-55.

Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu J-L & Ponting CP 2012: Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biology and Evolution.* 4, 427-442.

Yu X, Gong H, Zhou Y, Zhang H, Cao J & Zhou J 2015: Differential sialotranscriptomes of unfed and fed *Rhipicephalus haemaphysaloides*, with particular regard to differentially expressed genes of cysteine proteases. *Parasites & Vectors.* 8, 597.

Yu Y, Cao J, Zhou Y, Zhang H & Zhou J 2013: Isolation and characterization of two novel serpins from the tick *Rhipicephalus haemaphysaloides*. *Ticks and Tick-Borne Diseases.* 4, 297-303.

Yusufmia SB, Collins NE, Nkuna R, Troskie M, Van den Bossche P & Penzhorn BL 2010: Occurrence of *Theileria parva* and other haemoprotozoa in cattle at the edge of Hluhluwe-iMfolozi Park, KwaZulu-Natal, South Africa. *Journal of the South African Veterinary Association.* 81, 45-49.

Zhang Y & Cao X 2016: Long noncoding RNAs in innate immunity. *Cellular & Molecular Immunology.* 13, 138-147.

Zhou J, Ueda M, Umemiya R, Battsetseg B, Boldbaatar D, Xuan X & Fujisaki K 2006: A secreted cystatin from the tick *Haemaphysalis longicornis* and its distinct expression patterns in relation to innate immunity. *Insect Biochemistry and Molecular Biology.* 36, 527-535.

Zivkovic D, Pegram RG, Jongejan F & Mwase ET 1986: Biology of *Rhipicephalus appendiculatus* and *R. zambeziensis* and production of a fertile hybrid under laboratory conditions. *Experimental and Applied Acarology.* 2, 285-298.

Zivkovic Z, Esteves E, Almazán C, Daffre S, Nijhof AM, Kocan KM, Jongejan F & de la Fuente J 2010a: Differential expression of genes in salivary glands of male *Rhipicephalus* (*Boophilus*) *microplus* in response to infection with *Anaplasma marginale*. *BMC Genomics.* 11, 186.

Zivkovic Z, Torina A, Mitra R, Alongi A, Scimeca S, Kocan KM, Galindo RC, Almazán C, Blouin EF, Villar M, Nijhof AM, Mani R, La Barbera G, Caracappa S, Jongejan F & de la Fuente J 2010b: Subolesin expression in response to pathogen infection in ticks. *BMC Immunology.* 11, 7.

– The end –