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

Minique H. de Castro [a,b,c], Daniel de Klerk [a], Ronel Pienaar [a], Abdalla A. Latif [a,d], D. Jasper G. Rees [b,c] and Ben J. Mans [a,c,e*]

[a] Parasites, Vectors and Vector-borne Diseases, Onderstepoort Veterinary Institute, Agricultural Research Council, Onderstepoort, South Africa

[b] Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa

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

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

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

* Corresponding author

Tel.: +27 12 529 9200

Fax: +27 12 529 9434

E-mail address: mansb@arc.agric.za (B.J. Mans)

Abstract

Tick secretory proteins modulate haemostasis, inflammation and immune responses of the host and are attractive recombinant anti-tick vaccine candidates. Yet, many of the proteins have not been characterised due to the limited sequence availability for ticks and other arthropods for homology-based annotation. To address this limitation, we sequenced the salivary glands of the economically important adult male and female *Rhipicephalus appendiculatus* ticks during feeding. The quality filtered Illumina sequencing reads were *de novo* assembled to generate 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; *De novo* transcriptome assembly; Salivary glands; Sialotranscriptomics; Secretory proteins; Next generation sequencing.

Introduction

Ticks are blood-feeding ectoparasites that serve as vectors for a variety of human and veterinary diseases worldwide (Dennis and Piesman, 2005; Jongejan and Uilenberg, 2004). 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 a number of economically important tick-borne diseases in livestock and severe tick infestations can reduce animal weight, lower milk production and reduce hide qualities (De Castro, 1997). Ticks evolved highly adaptive mechanisms to feed unnoticed on their host for extended periods of time (Binnington, 1978; Binnington and Stone, 1981). The salivary glands produce numerous proteins to create a stable feeding environment by modulating the host's haemostasis, inflammation and immune response (reviewed in Fontaine et al., 2011; Francischetti et al., 2009; Mans, 2011). Currently, the most effective control agents for tick infestations are chemical acaricides (reviewed in Ghosh et al., 2007; Willadsen, 2006), but tick resistance to acaricides is becoming a global problem (reviewed in Abbas et al., 2014). Tick vaccines as an alternative control method is becoming more attractive and both exposed (secreted proteins) and hidden antigens may be targets (Nuttall et al., 2006; Willadsen, 2004).

Rhipicephalus appendiculatus, one of the most economically important tick species in Africa, transmits the protozoan parasite *Theileria parva,* causing the related cattle diseases; East Coast fever (ECF), Corridor disease (CD) and January disease (Lawrence et al., 1994; Stoltsz, 1989; Uilenberg, 1999). East Coast fever is a cattle to cattle transmitted economically important and devastating disease throughout central and eastern Africa, killing more than a million animals per annum and amounting in 168 million US dollars worth of damages in 1989 (Mukhebi et al., 1992). Corridor disease is a clinically similar disease that has shown

mortality rates of more than 90% (Neitz, 1955; Potgieter et al., 1988), whereas January disease is a less severe, seasonal disease occurring in Zimbabwe (Lawrence et al., 2004). East Coast fever was eradicated from South Africa in the 1950's through rigorous quarantine and slaughtering control measures (Neitz, 1957; Norval et al., 1992), but CD still persists as a controlled disease of cattle that is regulated by the Department of Agriculture, Forestry and Fisheries (Animal Disease Act 1984, Act No. 35). The *T. parva* parasite causing CD is transmitted from African buffalo (*Syncerus caffer*), the natural reservoir host, to cattle (Lawrence et al., 1994; Uilenberg, 1999). Conversion of CD from buffalo-cattle to cattle-cattle transmission will have serious implications for the control of this disease in South Africa. In addition, immunisation of cattle against ECF by the infection and treatment method, in which the live parasite is injected into the cattle followed by treatment with oxytetracycline, results in a *T. parva* carrier-state in cattle (Boulter and Hall, 1999; Radley et al., 1975). In South Africa, treatment or immunisation resulting in a *T. parva* carrier-state in cattle is not permitted due to the risk of buffalo-derived *T. parva* adapting to cattle-cattle transmission. An alternative for the management of CD risk in South Africa is the production of recombinant vaccines against the tick vector. The development of a recombinant vaccine requires a comprehensive understanding of tick feeding, host immune evasion and the genes involved in these processes.

Expressed sequence tag (EST) sequencing of cDNA libraries have previously been used to generate a gene index for *R. appendiculatus* (Nene et al., 2004). Due to technical limitations of the technology at the time, transcripts below 1000 bp were excluded from the dataset resulting in underrepresentation of smaller genes. Nevertheless, EST sequencing has provided a deeper insight into salivary gland complexity of a variety of tick species, including *R. sanguineus* (Anatriello et al., 2010), *Argas monolakensis* (Mans et al., 2008),

Amblyomma variegatum (Ribeiro et al., 2011), and *R. (Boophilus) microplus* (Zivkovic et al., 2010), to mention only a few. Regardless of the insights gained, the depth of sequencing achieved by these technologies is insufficient to cover the full complexity of sialomes as high levels of divergence in secretory proteins and extensive gene duplications in highly abundant protein families have been observed in ticks (Mans, 2011). This diversity is likely to modulate the host immune response by having multiple proteins to fulfil the same function, each slightly different to escape detection by the host's immune response (Francischetti et al., 2009; Mans, 2011). Recent advances in next generation sequencing (NGS) and RNA sequencing (RNAseq) technologies have paved the way to study many non-model organisms cost-effectively (Collins et al., 2008; Ekblom and Galindo, 2011). Next generation sequencing, which produces millions of sequencing reads, can achieve the sequence depth required to elucidate even lowly expressed genes of complex protein families. To exploit these advances, a number of *de novo* sialotranscriptomes (sets of RNA molecules in tick salivary glands) have recently been generated using NGS technologies for a number of tick species; *A. maculatum* (Karim et al., 2011), *Ixodes ricinus* (Schwarz et al., 2013), *Dermacentor andersoni* (Mudenda et al., 2014), *A. triste, A. parvum* and *A. cajennense* (Garcia et al., 2014), *Haemaphysalis flava* (Xu et al., 2015) and *R. pulchellus* (Tan et al., 2015a). These transcriptomes highlighted the true expansion within salivary protein families in the different tick species and identified new candidate genes involved in feeding. Yet, no such in depth transcriptome has yet been generated for *R. appendiculatus*, despite it being one of the most economically important ticks in southern and eastern Africa.

The aims of this study were therefore to (i) *de novo* assemble a representative, comprehensive gene catalogue of *R. appendiculatus* that improves the publically available sequence information of this species, (ii) provide a high quality annotation and characterisation of tick

secretory proteins, (iii) identify genes putatively involved in blood feeding by investigating the expression abundance of protein families and the differential expression between female and male ticks, and (iv) compare the assembled genes to sequences of known functional *R. appendiculatus* genes. To our knowledge, this is the first report of a *de novo* assembled sialotranscriptome of *R. appendiculatus* using next generation sequencing. We believe this transcriptome is an invaluable resource to the tick community and will facilitate future comparative studies with other tick salivary transcriptomes to elucidate the biology of tick feeding and aid in resolving complex tick protein families. Moreover, this transcriptome can also be used for future proteomic studies in *R. appendiculatus* and will be a valuable source for potential vaccine candidate selection.

Materials and Methods

Ethics statement

All animals used in this study were housed at the Onderstepoort Veterinary Institute (OVI). Ethical approval was obtained from the Onderstepoort Veterinary Institute Animal Ethics Committee for the feeding of various life stages of ticks (approval number: AEC12.11, extended to AEC01.15; Tick Feeding and Colony Maintenance Project) and from the University of South Africa, College of Agriculture and Environmental Sciences Animal Ethics Review Committee (approval number: 2014/CAES/098).

Ticks

Ticks (Pongola strain; South Africa) were obtained from a parasite-free colony maintained under standard laboratory and tick-rearing protocols at OVI (according to Heyne et al., 1987).

The ticks were maintained at 26 °C (± 1 °C), relative humidity of 75% (± 5%) and a 12-hour light/ 12-hour dark photoperiod. Adult ticks were fed in feeding bags on the backs of Hereford (*Bos taurus*) cattle originating from the disease-free cattle stock maintained at OVI.

Salivary gland dissection and RNA extraction

About 20 male and 20 female ticks were carefully removed from the bovine at different times during feeding (2 and 5 days post attachment), without disrupting their mouthparts. Unfed male and female ticks were obtained from the laboratory colony. Ticks were dissected under a stereomicroscope using sterile conditions and the salivary glands 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* digestion (Qiagen). 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).

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, before excising 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 production of longer reads) followed by excision of a high molecular weight fraction (±600 bp - 1200 bp). Sequencing was performed at the Biotechnology Platform Sequencing Facility (Agricultural Research Council, South Africa).

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, with commands: -e 0.02 -O 5 -m 20 -q 20 (Martin, 2011) and the FASTQ Quality Filter package of the FASTX-Toolkit v0.0.13, commands: -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 bps, 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 lowly expressed transcripts (that likely represent artefacts) were removed by filtering with a Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) value of 1 (Mortazavi et al., 2008). By using an expression level threshold (FPKM > 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.

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 ≤ 1e-20) to a local sequence database of the transcripts. Four reference-based metrics were generated; 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) as proposed by Martin and Wang (2011). The transcriptome completeness was also measured with 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).

Transcriptome annotation

The transcriptome was annotated using BLASTx similarity searches (E-value < e-5) 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).

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 (http://pfam.xfam.org/) database (Finn et al., 2014) and NCBI's Conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/cdd) (Marchler-Bauer et al., 2015). Putative signal peptide and transmembrane topology were predicted using the SignalP 4.0 (Petersen et al., 2011), Phobius (Kall 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. The predicted *R. appendiculatus* proteins were compared against two known tick protein datasets: predicted peptides from the *I. scapularis* genome (IscaW1.4, www.vectorbase.org), 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), representative of proteins expressed in salivary glands during tick feeding (Tan et al., 2015a), to evaluate similarity to other tick proteins.

Tick protein family characterisation

An in-house curated database of Acari (tick and mite) protein sequences 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).

Expression analysis of transcripts

Transcripts per million (TPM) values were determined for the transcripts using the Bowtie2 v2.2.3 (Langmead and Salzberg, 2012) and RNA-Seq by Expectation-Maximization (RSEM v1.2.15) software packages (Li and Dewey, 2011). 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). Chi-square test with Bonferroni correction was used for significance testing.

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 ≤ 1e-5) against one another. Additionally, the protein classes in each transcriptome were compared. A further comparison of the proteins assembled in this study against proteins previously identified as quantitative reverse transcriptase (RT) PCR reference genes (Nijhof et al., 2009) and functionally characterised or validated *R. appendiculatus* proteins (Bishop et al., 2002; Imamura et al., 2013; Mulenga et al., 2003a, 2003b; Paesen et al., 1999, 2007, 2009; Preston et al., 2013; Trimnell et al., 2002; Wang and Nuttall, 1995) was performed. The sequences of the proteins were obtained from NCBI and used in a BLASTp database to retrieve homologous sequences in the *R. appendiculatus* protein set.

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 paper is the first version, GEDV01000000.

Results

R. 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 (Supplemental Table A1). 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 1) after filtering based on transcript abundance (FPKM value ≥ 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 1: Summary of *R. appendiculatus* transcriptome assembly statistics.

	R. appendiculatus transcriptome
Number of transcripts	21 410
Number of transcripts > 500 bp	18 892
Number of transcripts > 1 Kb	13 702
Number of transcripts > 10 Kb	115
Shortest transcript length (bp)	301
Longest transcript length (bp)	16 259
Mean length of transcripts (bp)	2060.3
Median length of transcripts (bp)	1443
Transcript N50 (bp)	3134
Total bases in assembly (Mb)	44.1
Ambiguous base calls (Ns)	0
GC content (%)	49.0

BLASTx alignment of the 21 410 transcripts against six protein databases, functionally annotated 15 645 transcripts (73%) based on sequence similarity to proteins in at least one of the search databases (Table 2; annotations for all transcripts can be found in Supplemental Table B). 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" (Supplemental Fig. C1). 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 (Supplemental Fig. C2). 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

Table 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-5. Details of the datasets can be obtained in the Materials and Methods section.

	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 *Ixodes scapularis* predicted peptides	11 123	52.0
BLASTx against *Rhipicephalus* protein sequences	12 485	58.3
BLASTx against EuKaryotic Orthologous Groups (KOG) dataset	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).

activity", in the molecular function categories. KEGG pathways analysis assigned *I. scapularis* KO identifiers to 4647 transcripts (Supplemental Fig. C3). The most represented pathways were "Ribosome", "RNA transport", "Protein processing in endoplasmic reticulum" and "Spliceosome".

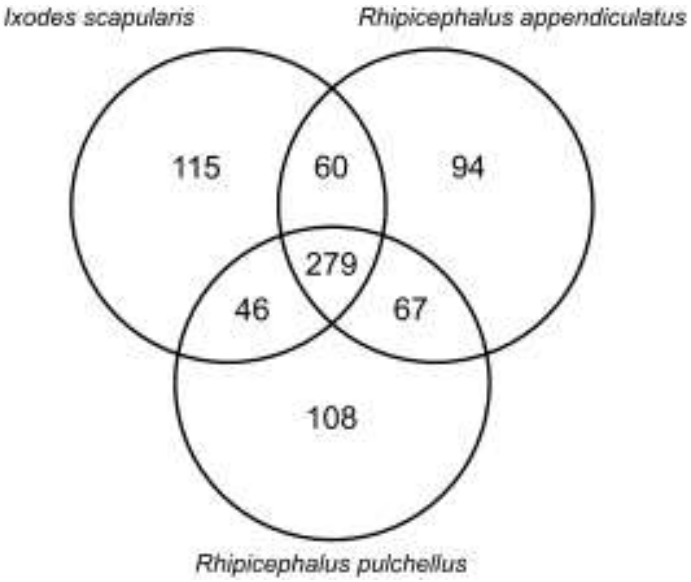
Open reading frame prediction, annotation and comparison with *I. scapularis* and *R. pulchellus*

A total of 14 433 ORFs were predicted, which together represented 13 996 (65%) of the *R. appendiculatus* transcripts (Supplemental Table D 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 - 16259 bp, average 2534 bp). Of the transcripts with no ORFs, 97% were predicted to be putative non-protein coding transcripts by at least two 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 a predicted start and stop codon) and most had significant BLASTp matches to protein search databases (79% and 89% for 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 " (Supplemental Fig. C4). 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 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 (Fig. 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, even more so to proteins expressed in tick salivary glands.

Fig. 1. Pfam domain comparison between *Ixodes scapularis, Rhipicephalus appendiculatus* and *R. pulchellus*. The 500 most represented Pfam domains in each species were used for the comparison analysis. Pfam searches were performed against the Pfam database (http://pfam.xfam.org/) and the Venn diagram drawn with Venny 2.0 (http://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, www.vectorbase.org) and 11 227 *R. pulchellus* predicted proteins (NCBI Bioproject PRJNA170743).

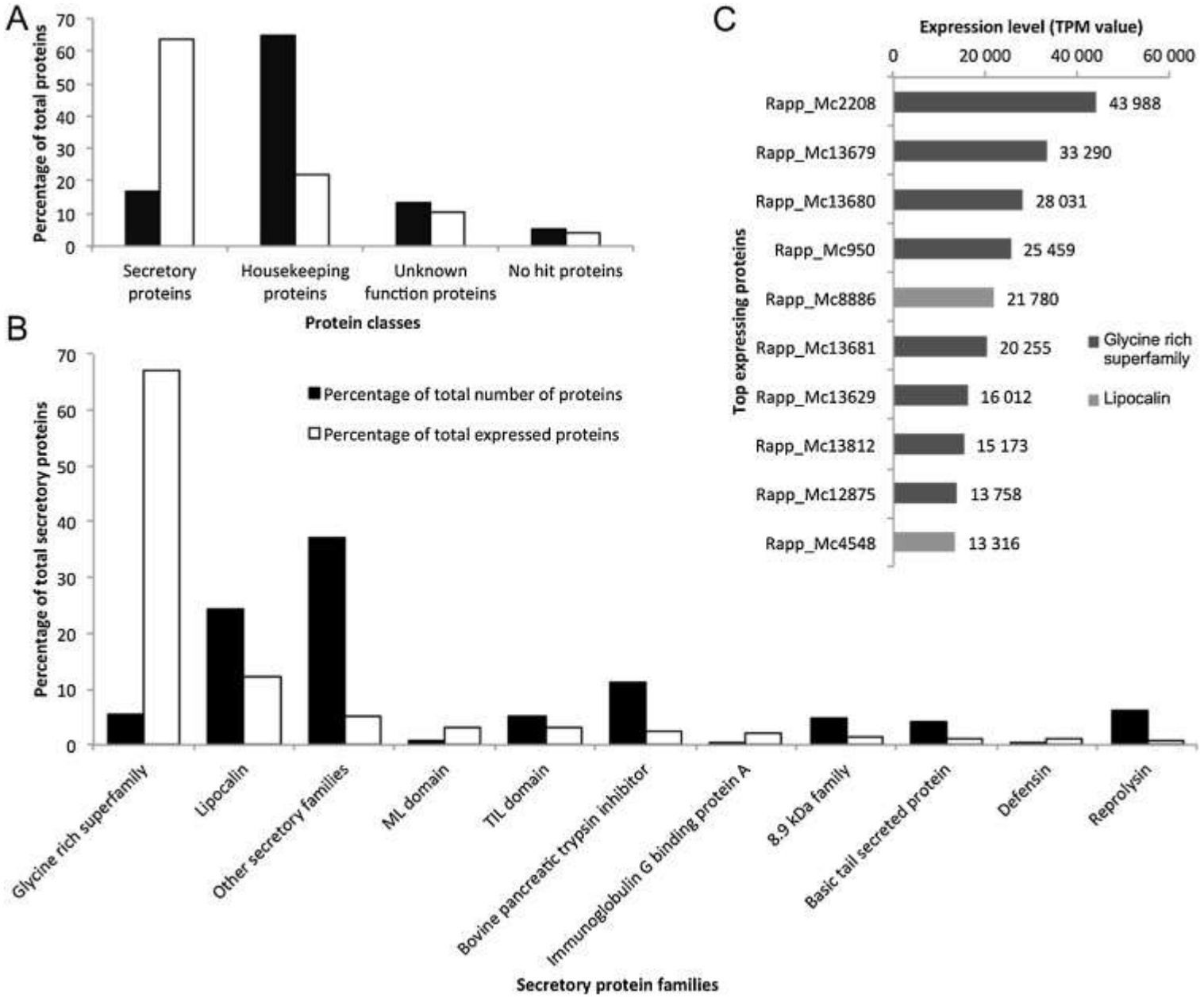


Expression profiling in the *R. appendiculatus* transcriptome

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) (Supplemental Table B). 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 (Supplemental Table A2). 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%; Fig. 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; Supplemental Table A3). 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 (Fig. 2B). 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 (Fig. 2C). All other secretory protein families (1499 proteins) resulted in the remaining 22% of the transcript expression in the secretory protein class.

Fig. 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.

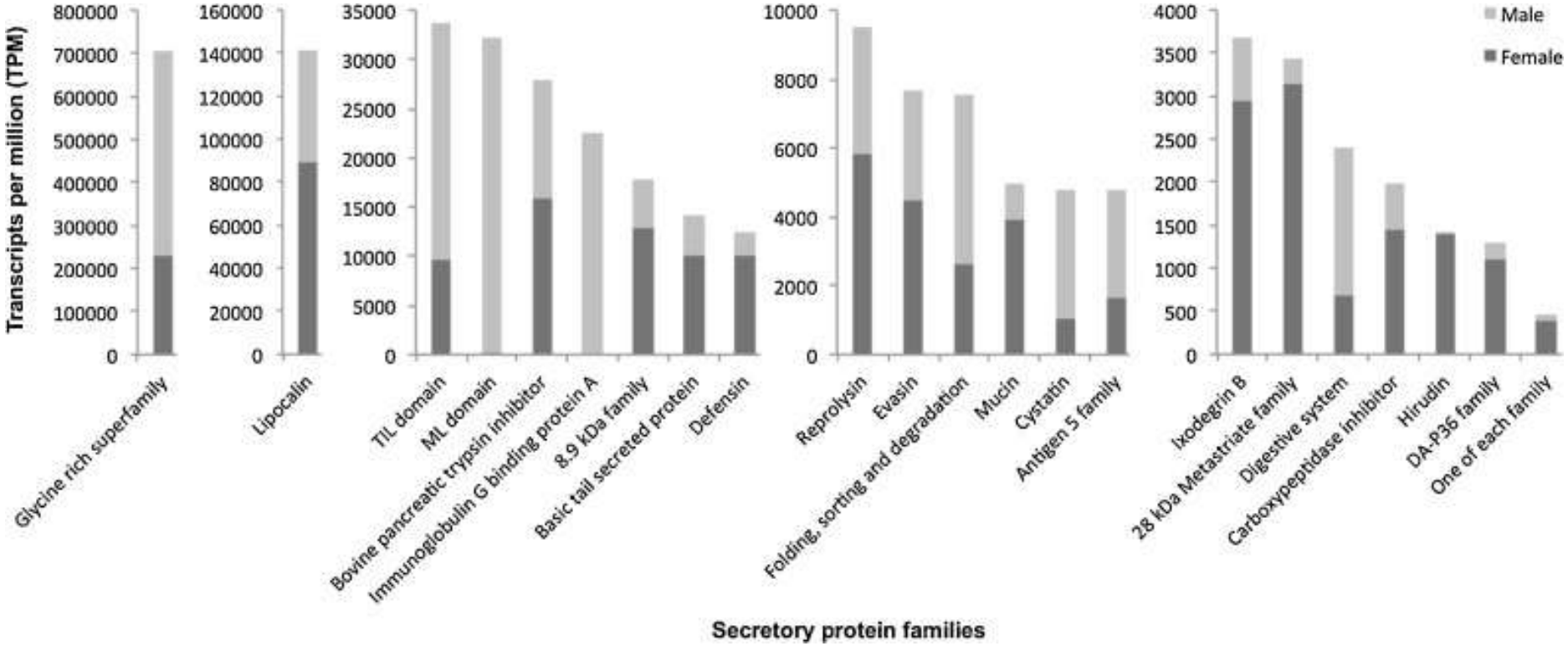


Expression differences between female and male sialotranscriptomes of *R. appendiculatus*

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; Supplemental Table B). Many of the families in the secretory protein class had significantly skewed expression between males and female transcriptomes after Bonferroni correction (Fig. 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 family 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 (Fig. 3).

A total of 1758 and 2346 transcripts were differentially up regulated (at least a 2-fold increase) in the female and male transcriptomes, respectively (Supplemental Table A4). 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

Fig. 3. Gender-skewed expression of secretory protein families in the *R. appendiculatus* transcriptome. Secretory protein families with significant ($p < 0.05$, Chi-square test with Bonferroni correction) 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.

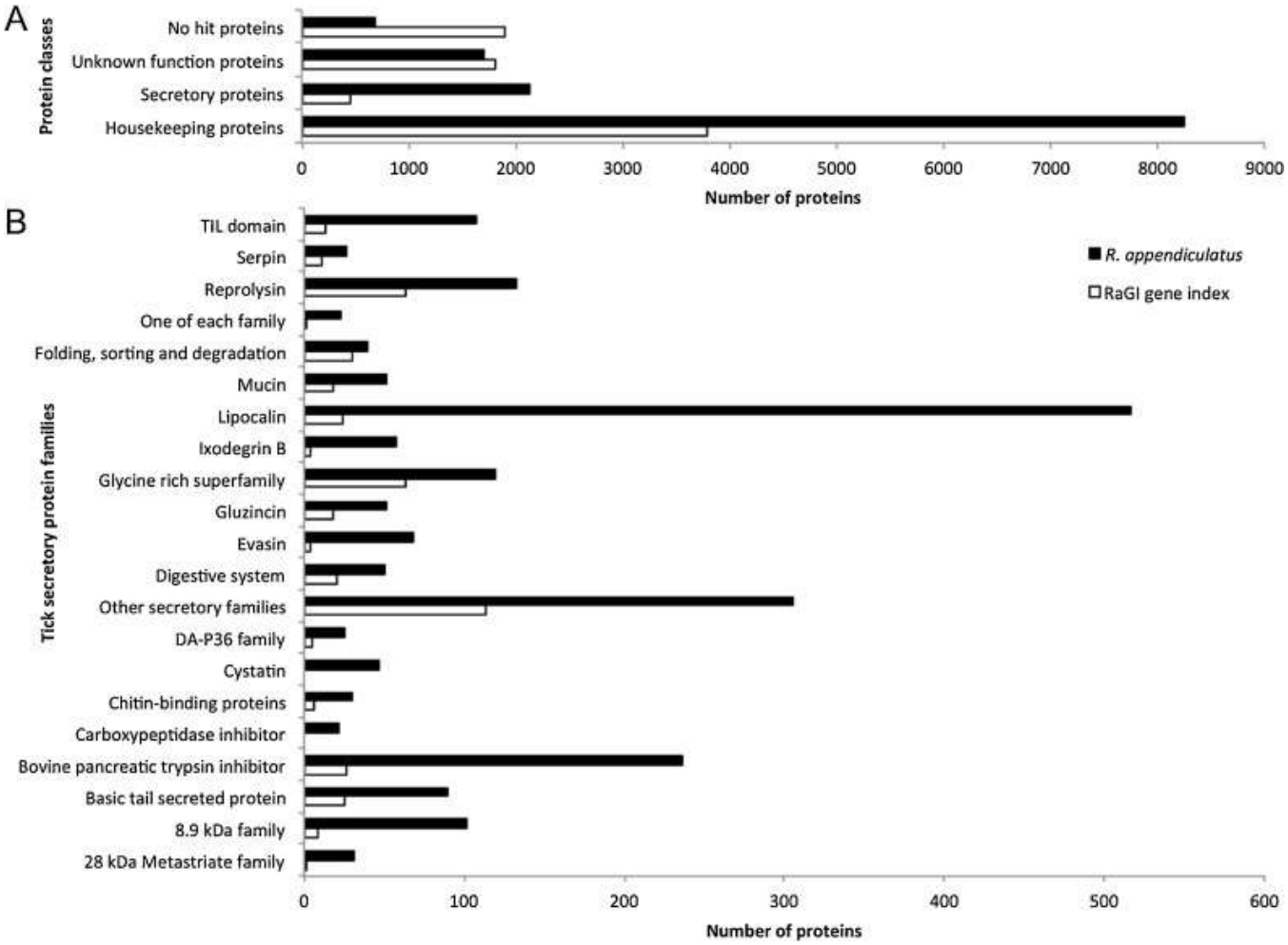


transcriptome while more of the Digestive system (including Serine proteases; 3 vs. 25) and Gluzincin (2 vs. 33) families were up regulated in the male transcriptome after Bonferroni correction. Additional large gender differences, albeit not significant, were observed in the DA-P36 (15 vs. 2) and 28 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.

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)

Fig. 4. Distribution of tick protein families 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 gene index. The *R. appendiculatus* transcriptome is indicated in black and the RaGI gene index in white.



were assembled in the current transcriptome (Fig. 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 (Fig. 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.*

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 transcriptome (Table 3). Many of these genes were previously only available as EST fragments, but 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., 2003a) 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 level. 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.

Appendix A. Supplemental Tables:

Sequence reads and expression.

Supplemental Table A1: 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

Supplemental Table A2: 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	9 103	0.91
19	c50957_g1_i1	Rapp_Mc13626	Unknown function	8 589	0.86
20	c1612_g1_i1	Rapp_Mc13700	Energy metabolism: Cytochrome c oxidase subunit 1	8 340	0.83

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

Supplemental Table A3: 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	133.12	12.47	Rapp_Mc8886	21 779.87	31.71
Bovine pancreatic trypsin inhibitor	236	11.06	58.43	2.50	Rapp_Mc8896	2447.72	17.75
Reprolysin	133	6.23	34.93	0.84	Rapp_Mc5881	628.36	13.52
Glycine rich superfamily	119	5.58	3072.04	66.34	Rapp_Mc2208	43 988.23	11.55
TIL domain	108	5.06	164.00	3.21	Rapp_Mc1646	2898.99	16.37
8.9 kDa family	102	4.78	84.00	1.55	Rapp_Mc13118	1185.58	13.84
Basic tail secreted protein	90	4.22	75.77	1.24	Rapp_Mc4488	879.84	12.90
Evasin	68	3.19	55.54	0.69	Rapp_Mc9039	619.77	16.41
Ixodegrin B	57	2.67	30.13	0.31	Rapp_Mc823	450.18	26.21
Gluzincin	52	2.44	7.63	0.07	Rapp_Mc4972	115.13	29.02
Mucin	52	2.44	44.87	0.42	Rapp_Mc417	876.51	37.57
Digestive system (including Serine proteases)	50	2.34	25.01	0.23	Rapp_Mc1191	137.69	11.01
Cystatin	47	2.20	54.35	0.46	Rapp_Mc13730	776.50	30.40
Folding, sorting and degradation (including Cathepsins)	40	1.87	96.78	0.70	Rapp_Mc945	1498.52	38.71
28 kDa Metastriate family	31	1.45	50.68	0.29	Rapp_Mc2646	557.06	35.46
Chitin-binding proteins	30	1.41	23.00	0.13	Rapp_Mc9698	223.55	32.39
Serpin	27	1.27	8.98	0.04	Rapp_Mc5185	74.71	30.81

DA-P36 family	25	1.17	23.86	0.11	Rapp_Mc8808	340.81	57.13
Transport and catabolism	25	1.17	36.14	0.16	Rapp_Mc2177	535.87	59.31
One of each family	23	1.08	9.44	0.04	Rapp_Mc3057	42.46	19.57
Lipid metabolism	22	1.03	4.28	0.02	Rapp_Mc1456	12.14	12.89
Carboxypeptidase inhibitor	22	1.03	42.89	0.17	Rapp_Mc10222	388.49	41.17
5'-Nucleotidase	16	0.75	14.33	0.04	Rapp_Mc6697	47.63	20.77
Microplusin	16	0.75	65.37	0.19	Rapp_Mc1964	434.68	41.56
ML domain	16	0.75	1111.82	3.23	Rapp_Mc774	11 353.27	63.82
Antigen 5 family	13	0.61	188.87	0.45	Rapp_Mc1903	1014.98	41.34
Signaling molecules and interaction	13	0.61	1.56	0.00	Rapp_Mc8916	2.75	13.60
Translation	13	0.61	10.99	0.03	Rapp_Mc13622	43.10	30.17
24 kDa family	12	0.56	24.22	0.05	Rapp_Mc9762	91.56	31.50
Defensin	12	0.56	464.44	1.01	Rapp_Mc8698	1899.98	34.09
8 kDa Amblyomma family	11	0.52	28.78	0.06	Rapp_Mc13004	154.27	48.73
Glycan biosynthesis and metabolism	11	0.52	14.00	0.03	Rapp_Mc6227	80.50	52.27
Sphingomyelinase	9	0.42	10.67	0.02	Rapp_Mc837	22.89	23.84
Signal transduction	8	0.37	7.80	0.01	Rapp_Mc1131	14.98	24.01
Transcription	8	0.37	4.42	0.01	Rapp_Mc1617	9.48	26.79
Carbohydrate metabolism	7	0.33	4.31	0.01	Rapp_Mc5896	7.62	25.24
Fibrinogen-related domain	7	0.33	52.55	0.07	Rapp_Mc9028	248.58	67.58
Secretory - unknown function	7	0.33	12.81	0.01	Rapp_Mc9124	51.43	65.04
Immunoglobulin G binding protein A	6	0.28	2051.82	2.23	Rapp_Mc1190	4721.78	38.35
Metabolism of other amino acids	6	0.28	29.87	0.03	Rapp_Mc5888	67.37	37.59
Phospholipase A2	6	0.28	22.53	0.02	Rapp_Mc8892	44.33	32.79
Replication and repair	6	0.28	5.26	0.01	Rapp_Mc2861	9.23	29.23
7DB family	5	0.23	26.52	0.02	Rapp_Mc5571	66.08	49.84
Metalloprotease	5	0.23	12.08	0.01	Rapp_Mc12946	38.89	64.40

SALP15	4	0.19	10.25	0.01	Rapp_Mc1541	28.69	69.99
Astacin	3	0.14	1.77	0.00	Rapp_Mc7012	3.39	63.84
Cell growth and death	3	0.14	6.83	0.00	Rapp_Mc3897	17.39	84.91
Dermacentor 9 kDa expansion	3	0.14	14.72	0.01	Rapp_Mc1065	26.04	58.95
Histidine rich	3	0.14	2.87	0.00	Rapp_Mc450	3.95	45.82
14 kDa family	3	0.14	25.10	0.01	Rapp_Mc8740	54.41	72.25
Kazal domain	3	0.14	21.10	0.01	Rapp_Mc421	59.70	94.33
Kazal/vWf domain	3	0.14	18.64	0.01	Rapp_Mc2515	35.48	63.46
TELEM	3	0.14	1.99	0.00	Rapp_Mc5946	2.32	38.93
Thyropin	3	0.14	67.58	0.04	Rapp_Mc1844	99.53	49.09
Cysteine rich	2	0.09	1.03	0.00	Rapp_Mc1691	1.10	53.66
Energy metabolism	2	0.09	15.83	0.01	Rapp_Mc6151	25.10	79.28
Hirudin	2	0.09	310.16	0.11	Rapp_Mc11642	418.25	67.42
Bovine pancreatic trypsin inhibitor - Lipocalin	1	0.05	3.56	0.00	Rapp_Mc3211	3.56	100.00
Chitin deacetylase activity	1	0.05	1.17	0.00	Rapp_Mc2536	1.17	100.00
Cell motility	1	0.05	3.36	0.00	Rapp_Mc4124	3.36	100.00
Cysteine rich hydrophobic domain 2	1	0.05	13.39	0.00	Rapp_Mc7198	13.39	100.00
Fatty acid-binding protein	1	0.05	40.42	0.01	Rapp_Mc2582	40.42	100.00
Histamine release factor	1	0.05	1211.56	0.22	Rapp_Mc12631	1211.56	100.00
Immune system	1	0.05	24.15	0.00	Rapp_Mc8912	24.15	100.00
26 kDa family	1	0.05	7.18	0.00	Rapp_Mc5668	7.18	100.00
Kazal/SPARC domain	1	0.05	64.96	0.01	Rapp_Mc1895	64.96	100.00

Supplemental Table A4: Differential expression between female and male ticks in the salivary transcriptome of *R. appendiculatus.*

Protein families	Female up regulated *	Male up regulated *
Secretory protein families	570	553
24 kDa family	3	4
28 kDa Metastriate family	18	4
5'-Nucleotidase	3	2
7DB family	0	1
8 kDa Amblyomma family	2	6
8.9 kDa family	45	28
Antigen 5 family	1	2
Astacin	0	2
Basic tail secreted protein	36	21
Bovine pancreatic trypsin inhibitor	61	71
Carbohydrate metabolism	2	0
Carboxypeptidase inhibitor	6	5
Cell motility	0	1
Chitin-binding proteins	0	5
Cystatin	19	15
Cysteine rich	1	1
DA-P36 family	15	2
Defensin	2	5
Dermacentor 9 kDa expansion	0	3
Digestive system (including Serine proteases)	3	25
Evasin	36	16
Fibrinogen-related domain	2	0
Folding, sorting and degradation (including Cathepsins)	1	8
Gluzincin	2	33
Glycan biosynthesis and metabolism	2	0
Glycine rich superfamily	14	31
Hirudin	2	0
Histidine rich	3	0
Immunoglobulin G binding protein A	0	6
Ixodegrin B	36	7
Kazal domain	2	1
Kazal/vWf domain	1	1
Lipid metabolism	7	5
Lipocalin	157	154
Microplusin	3	5
ML domain	1	8
Mucin	11	3
No hit	0	1
One of each family	17	1
Phospholipase A2	0	2
Reprolysin	27	15

SALP15	0	1
Secretory - unknown function	0	2
Serpin	3	5
Signal transduction	1	0
Signaling molecules and interaction	3	0
Sphingomyelinase	0	5
TIL domain	20	34
Transport and catabolism	2	6
Housekeeping protein class	220	413
Unknown function protein class	130	288
No hit protein class	111	172
Transcripts without predicted ORFs	727	920
Total	**1758**	**2346**

* Transcripts estimated as up regulated (fold change > 2) by the edgeR (Empirical analysis of digital gene expression data in R) software package.

Table 3: 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., 2003b	Rapp_Mc7014	88	Complete	6.5	6.8	6.2
RAS-2	Serine proteinase inhibitor serpin-2	AAK61376.1	Mulenga et al., 2003b	Rapp_Mc6400	92	Complete	11.3	14.6	8.5
RAS-3	Serine proteinase inhibitor serpin-3	AAK61377.1	Mulenga et al., 2003b	Rapp_Mc5185	95	Complete	74.7	136.7	24.2

RAS-4	Serine proteinase inhibitor serpin-4	AAK61378.1	Mulenga et al., 2003b	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.

Protein identities between 48% and 88% were observed for four proteins; Serine proteinase inhibitors RAS-1 and -4 (Mulenga et al., 2003b), 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 ≤ 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 the homologous sequences assembled in the *R. appendiculatus* transcriptome.

The HBPM homolog, at 21 780 TPM, is the 6th highest expressed transcript in the *R. appendiculatus* transcriptome (Supplemental Table A2) 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 of 2e-33), were observed for the assembled protein (Supplemental Table D). Searching against the NCBI PDB protein database retrieved the Chain A, Histamine Binding Protein From Female Brown Ear *R. appendiculatus* (pdb|1QFT|A) sequence with significant similarity (E-value 2e-

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 (Supplemental Table D). 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 3). The complete SERPIN cd00172 domain and the conserved reactive centre loop (RCL), typical to serpins, were identified in all four sequences (Supplemental Table D). When the serpin phylogeny from Tirloni et al. (2014) was reproduced, Rapp_Mc7014, Rapp_Mc6400 and Rapp_Mc5185 grouped closest to their respective homologs (RAS1-3, data not shown). On the other hand, Rapp_Mc4940 grouped in the same cluster as its homolog, RAS-4 (72% protein identity), but more closely to *R. pulchellus* serpin (JAA54310.1, 95% identity) and *R. (B.) microplus* RmS17 (AHC98668.1, 92% identity).

Discussion

The main aim of this study was the *de novo* assembly of a salivary transcriptome of *R. appendiculatus* that represented the feeding stages of both female and male ticks. During *de novo* assembly, an expression catalogue was constructed from short sequence reads without any prior reference, making it essential to assess whether it reflected the actual transcripts in the biological sample. Two transcriptome assessments, read mapping-based and reference-based (Martin and Wang, 2011; O'Neil and Emrich, 2013), indicated that the assembled

transcriptome was representative of the sequence reads and showed high 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 length and representation of previously unidentified *R. appendiculatus* transcripts. The similarity of the *R. appendiculatus* proteins to two known tick 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 tick-representing set of proteins. Annotation of non-model organisms and especially arthropods is challenging due to few 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 *A. americanum* (Gibson et al., 2013) and 37% in *H. flava* (Xu et al., 2015). In contrast, a higher overall annotation of 73% was observed in the transcriptome of *R. appendiculatus* assembled here. The higher annotation was attributed to searching against more than one protein database and the removal of lowly expressed (FPKM < 1) transcripts from the transcriptome (Mortazavi et al., 2008). No open reading frames 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 be unpredicted in these transcripts, albeit at low percentage. These transcripts were mainly predicted to have low

protein-coding potential and might be representative of long intergenic noncoding RNA (lincRNA); RNA molecules longer than 200 bp that contain no open reading frames for translation. LincRNAs have shown functions in transcriptional regulation, RNA processing and protein scaffolding (reviewed in Wilusz et al., 2009). In *Drosophila*, over a thousand lincRNA molecules have been identified and more lincRNAs were sex-specifically expressed compared to protein coding genes in adult flies (Young et al., 2012). Similarly, we observed that 40% of the differentially expressed transcripts in the *R. appendiculatus* transcriptome had no predicted open reading frames. Their classification as lincRNAs 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 as other tick sialotranscriptomes of about 13 - 37% (Garcia et al., 2014; Karim and Ribeiro, 2015; Karim et al., 2011; Tan et al., 2015a). The same authors also reported a wide range of expression proportions for secretory protein transcripts 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 high expressing transcripts contained ORFs that coded for secretory proteins (i.e. 52% of the 50 highest expressing transcripts were classified as belonging to secretory protein families). Similarly, previous studies in *I. ricinus* reported that secretory proteins were the highest expressed transcripts in the salivary glands when compared to midgut tissues (Kotsyfakis et al., 2015; Schwarz et al., 2014). 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 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). *R. 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, female ticks to 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. 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; Tan et al., 2015a; Xiang et al., 2012), we observed various differences 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 proteins (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-RCR 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., 2003a) and Ra86-1 (Nijhof et al., 2009), in accordance to our inability to assemble the 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 RNAseq 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, we were unable 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 - we found male-specific expression for HBPM and IGBP-Ms and female-specific expression for HBP1, HBP2, Ra-KLP and Japanin in our transcriptome. Expression profiling corroborated the publically available knowledge of *R. appendiculatus* genes. However, lower than expected protein identity percentages (though still in range with the transcriptome comparison with RaGI) were observed for the assembled RAS-4 (72%) and HBPM (48%) proteins compared to previous work (Mulenga et al., 2003b; Paesen et al., 1999). 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 identities 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 (Tirloni et al., 2014). The assembled proteins therefore seem to be functional, though their functions remain to be determined. The exclusion 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 also 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 tight 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 our sampling missing the 'snapshot' of expression of the genes. Our sampling design, that included three separate time points (0, 2 and 5 days feeding), 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 (Arcà et al., 2014; Chagas et al., 2013) and ticks (Dai et al., 2012; Kotsyfakis et al., 2015). Constant adaptation to host immune evasion and co-evolution with hosts would drive rapid expansion and divergence in salivary protein families. Another alternative explanation is the variability observed in naturally occurring *R. appendiculatus* populations due to different geographical distribution and climatic changes. 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 the molecular level, the *cytochrome oxidase subunit I* gene separated the southern and 'transition' groups into two genetically differentiated clades, albeit without enough support to be classified a subspecies (Mtambo et al., 2007). The east African group was not included in the study, which potentially would

have resulted in more pronounced separation. The third plausible explanation for the differences observed in the sequences of the salivary proteins reported here as compared to publically available sequences is the high divergence observed between laboratory-bred *R. appendiculatus* tick colonies. Based on microsatellite marker analysis, Kanduma et al. (2015) showed strong evidence that different *R. appendiculatus* laboratory breeding populations were genetically distinct. The laboratory-bred stocks exhibited high levels of inbreeding and most were significantly divergent from each other and the wild populations they were initially sampled from. Ticks sequenced in this work were from a laboratory maintained stock at OVI, South Africa (southern 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 group) or were undisclosed. These findings could have serious implications for the control strategies of *R. appendiculatus* by anti-tick recombinant vaccines. Currently, immunisation against the *T. parva* parasite is by means of the common infection and treatment method, which results in a *T. parva* carrier-state in cattle (Boulter and Hall, 1999; Radley et al., 1975). Similarly, treatment using parvaquone or buparvaquone also results in a carrier-state in cattle (Dolan, 1986). The current form of control in South Africa is a quarantine and slaughter policy to prevent the possible establishment of a carrier-state in cattle. Therefore, in South Africa, the only viable possibility for protection against Corridor disease is the development of a vaccine against the tick vector and knowledge of the salivary proteins present in locally occurring *R. appendiculatus* populations are invaluable towards this endeavour.

In conclusion, this is the first study to assemble a *de novo* sialotranscriptome of *R. appendiculatus* female and male ticks using next generation sequencing technologies. The transcriptome is of high quality and improves on the previously generated sequence dataset

for *R. appendiculatus*. Transcriptome expression profiles are complex and 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. Differences in some of the previously functionally characterised proteins were observed, potentially resulting from positive selection, natural *R. appendiculatu*s 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 protein vaccines. The transcriptome of *R. appendiculatus* is one of only a small number of tick sialotranscriptomes available to date that together will assist in characterising tick proteins and protein families and improve our understanding of tick feeding, host-interaction and tick biology as a whole.

Acknowledgments

We kindly thank Dr. Vinet Coetzee for suggestions and critical review of the manuscript and Dr. Eshchar Mizrachi for inputs during initial experimental design. This work was supported by the Economic Competitive Support Programme (30/01/V010) and Incentive Funding for Rated Researchers (NRF-Mans).

Author Contributions

Conception and design of the work: MD BM. Acquisition of data: MD DK RP AL JR BM. Analysis and interpretation of data: MD BM. Drafting the article: MD BM. Revising the article and final approval: MD DK RP AL JR BM.

References

Abbas, R.Z., Zaman, M.A., Colwell, D.D., Gilleard, J., Iqbal, Z., 2014. Acaricide resistance in cattle ticks and approaches to its management: The state of play. Vet. Parasitol. 203, 6-20.

Aljamali, M.N., Ramakrishnan, V.G., Weng, H., Tucker, J.S., Sauer, J.R., Essenberg, R.C., 2009. Microarray analysis of gene expression changes in feeding female and male lone star ticks, *Amblyomma americanum* (L). Arch. Insect Biochem. Physiol. 71, 236-253.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Anatriello, E., Ribeiro, J.M.C., de Miranda-Santos, I.K.F., Brandão, L.G., Anderson, J.M., Valenzuela, J.G., Maruyama, S.R., Silva, J.S., Ferreira, B.R., 2010. An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. BMC Genomics 11, 450.

Arcà, B., Struchiner, C., Pham, V., Sferra, G., Lombardo, F., Pombi, M., Ribeiro, J., 2014. Positive selection drives accelerated evolution of mosquito salivary genes associated with blood- feeding. Insect Mol. Biol. 23, 122-131.

Binnington, K.C., 1978. Sequenctial changes in salivary gland structure during attachment and feeding of the cattle tick, *Boophilus microplus*. Int. J. Parasitol. 8, 97-115.

Binnington, K.C., Kemp, D.H., 1980. Role of tick salivary glands in feeding and disease transmission. Adv. Parasitol. 18, 315-339.

Binnington, K.C., Stone, B.F., 1981. Developmental changes in morphology and toxin content of the salivary gland of the Australian paralysis tick *Ixodes holocyclus*. Int. J. Parasitol. 11, 343-351.

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. Int. J. Parasitol. 32, 833-842.

Boulter, N., Hall, R., 1999. Immunity and vaccine development in the bovine theilerioses. Adv. Parasitol. 44, 41-97.

Brooks, M.J., Rajasimha, H.K., Roger, J.E., Swaroop, A., 2011. Next-generation sequencing facilitates quantitative analysis of wild-type and *Nrl*$^{-/-}$ retinal transcriptomes. Mol. Vision 17, 3034-3054.

Chagas, A.C., Calvo, E., Rios-Velásquez, C.M., Pessoa, F.A., Medeiros, J.F., Ribeiro, J.M., 2013. A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. BMC Genomics 14, 875.

Collins, L.J., Biggs, P.J., Voelckel, C., Joly, S., 2008. An approach to transcriptome analysis of non-model organisms using short-read sequences. Genome Inform. 21, 3-14.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

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 Evol. Biol. 12, 4.

De Castro, J.J., 1997. Sustainable tick and tick-borne diseases control in livestock improvement in developing countries. Vet. Parasitol. 71, 77-97.

Dennis, D.T., Piesman, J.F., 2005. Overview of tick-borne infections of humans, in: Goodman, J.L., Dennis, D.T., Sonenshine, D.E. (Eds.), Tick-borne diseases of humans. American Society for Microbiology Press, Washington, DC, pp. 3-11.

Dolan, T.T., 1986. Chemotherapy of East Coast fever: the long term weight changes, carrier state and disease manifestations of parvaquone treated cattle. J. Comp. Pathol. 96, 137-146.

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

Findlay, G.D., Yi, X., MacCoss, M.J., Swanson, W.J., 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. PLoS Biol. 6, e178.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. The Pfam protein families database. Nucleic Acids Res. (Database Issue) 42, D222-D230.

Fontaine, A., Diouf, I., Bakkali, N., Missé, D., Pagès, F., Fusai, T., Rogier, C., Almeras, L., 2011. Implication of haematophagous arthropod salivary proteins in host-vector interactions. Parasit. Vectors 4, 187.

Francischetti, I.M., Sa-Nunes, A., Mans, B.J., Santos, I.M., Ribeiro, J.M., 2009. The role of saliva in tick feeding. Front. Biosci. 14, 2051-2088.

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 Res. 20, 763-783.

Garcia, G.R., Gardinassi, L.G., Ribeiro, J.M., Anatriello, E., Ferreira, B.R., Moreira, H.N., Mafra, C., Martins, M.M., Szabó, M.P., de Miranda-Santos, I.K.F., Maruyama, S.R., 2014.

The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. Parasit. Vectors 7, 430.

Ghosh, S., Azhahianambi, P., Yadav, M.P., 2007. Upcoming and future strategies of tick control: a review. J. Vector Borne Dis. 44, 79-89.

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

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494-1512.

Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., van Oudenaarden, A., Teichmann, S.A., 2011. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. 7, 497.

Heyne, H., Elliott, E.G., Bezuidenhout, J.D., 1987. Rearing and infection techniques for *Amblyomma* species to be used in heartwater transmission experiments. Onderstepoort J. Vet. Res. 54, 461-471.

Imamura, S., Konnai, S., Yamada, S., Parizi, L.F., Githaka, N., Vaz, I.D.S., Murata, S., Ohashi, K., 2013. Identification and partial characterization of a gut *Rhipicephalus appendiculatus* cystatin. Ticks Tick Borne Dis. 4, 138-144.

Jongejan, F., Uilenberg, G., 2004. The global importance of ticks. Parasitol. 129, S3-S14.

Kall, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. Nucleic Acids Res. (Database Issue) 35, (Web Server Issue), W429-W432.

Kanduma, E.G., Mwacharo, J.M., Mwaura, S., Njuguna, J.N., Nzuki, I., Kinyanjui, P.W., Githaka, N., Heyne, H., Hanotte, O., Skilton, R.A., 2015. Multi-locus genotyping reveals absence of genetic structure in field populations of the brown ear tick (*Rhipicephalus appendiculatus*) in Kenya. Ticks Tick Borne Dis. DOI:10.1016/j.ttbdis.2015.08.001.

Karim, S., Ribeiro, J.M.C., 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, J.M.C., 2011. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS One 6, e28525.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., Gao, G., 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345-W349.

Kotsyfakis, M., Schwarz, A., Erhart, J., Ribeiro, J.M., 2015. Tissue-and time-dependent transcription in Ixodes ricinus salivary glands and midguts when blood feeding on the vertebrate host. Sci. Rep. 5, 9103.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567-580.

Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Met. 9, 357-359.

Lawrence, J.A., de Vos, A.J., Irvin, A.D., 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. Oxford University Press, Oxford, UK, pp. 326-328.

Lawrence, J.A., Perry, B.D., Williamson, S.M., 2004. Zimbabwe theileriosis, in: Coetzer, J.A.W., Tustin, R.C. (Eds.), Infectious Diseases of Livestock, 2nd edition. Oxford University Press, Cape Town, SA, pp. p472-474.

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, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

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.

Madder, M., Speybroeck, N., Brandt, J., Berkvens, D., 1999. Diapause induction in adults of three *Rhipicephalus appendiculatus* stocks. Exp. Appl. Acarol. 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. Exp. Appl. Acarol. 27, 209-221.

Mans, B.J., 2011. Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. J. Innate Immun. 3, 41-51.

Mans, B.J., Andersen, J.F., Francischetti, I.M., Valenzuela, J.G., Schwan, T.G., Pham, V.M., Garfield, M.K., Hammer, C.H., Ribeiro, J.M., 2008. Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. Insect Biochem. Mol. Biol. 38, 42-58.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, D222-D226.

Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. Nat. Rev. Genet. 12, 671-682.

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

Maruyama, S.R., Anatriello, E., Anderson, J.M., Ribeiro, J.M., Brandão, L.G., Valenzuela, J.G., Ferreira, B.R., Garcia, G.R., Szabó, M.P., Patel, S., Bishop, R., de Miranda-Santos, I.K., 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.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182-W185.

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

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

Mudenda, L., Pierlé, S.A., Turse, J.E., Scoles, G.A., Purvine, S.O., Nicora, C.D., Clauss, T.R., Ueti, M.W., Brown, W.C., Brayton, K.A., 2014. Proteomics informed by transcriptomics identifies novel secreted proteins in *Dermacentor andersoni* saliva. Int. J. Parasitol. 44, 1029-1037.

Mukhebi, A.W., Perry, B.D., Kiusk, R., 1992. Estimated economics of theileriosis control in Africa. Prev. Vet. Med. 12, 73-85.

Mulenga, A., Misao, O., Sugimoto, C., 2003a. Three serine proteinases from midguts of the hard tick *Rhipicephalus appendiculatus*; cDNA cloning and preliminary characterization. Exp. Appl. Acarol. 29, 151-164.

Mulenga, A., Tsuda, A., Onuma, M., Sugimoto, C., 2003b. Four serine proteinase inhibitors (serpin) from the brown ear tick, *Rhiphicephalus appendiculatus*; cDNA cloning and preliminary characterization. Insect Biochem. Mol. Biol. 33, 267-276.

Neitz, W.O., 1955. Corridor disease: A fatal form of bovine theileriosis encountered in Zululand. Bull. Epizoot. Dis. Afr. 3, 121-123.

Neitz, W.O., 1957. Theileriosis, gonderiosis and cytauxzoonosis. A review. Onderstepoort J. Vet. Res. 27, 275-430.

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 Biochem. Mol. Biol. 34, 1117-1128.

Nijhof, A.M., Balk, J.A., 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 Mol. Biol. 10, 112.

Norval, R.A.I., Perry, B.D., Young, A.S., 1992. The epidemiology of theileriosis in Africa. Academic Press, London, UK.

Nuttall, P., Trimnell, A., Kazimirova, M., Labuda, M., 2006. Exposed and concealed antigens as vaccine targets for controlling ticks and tick- borne diseases. Parasite Immunol. 28, 155-163.

O'Neil, S.T., Emrich, S.J., 2013. Assessing *De Novo* transcriptome assembly metrics for consistency and utility. BMC Genomics 14, 465.

Ochanda, H., Young, A., Medley, G., Perry, B., 1998. Vector competence of 7 rhipicephalid tick stocks in transmitting 2 *Theileria parva* parasite stocks from Kenya and Zimbabwe. Parasitology 116, 539-545.

Paesen, G., Adams, P., Harlos, K., Nuttall, P., Stuart, D., 1999. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. Mol. Cell 3, 661-671.

Paesen, G.C., Siebold, C., Dallas, M.L., Peers, C., Harlos, K., Nuttall, P.A., Nunn, M.A., Stuart, D.I., Esnouf, R.M., 2009. An ion-channel modulator from the saliva of the brown ear tick has a highly modified Kunitz/BPTI structure. J. Mol. Biol. 389, 734-747.

Paesen, G.C., Siebold, C., Harlos, K., Peacey, M.F., Nuttall, P.A., Stuart, D.I., 2007. A tick protein with a modified Kunitz fold inhibits human tryptase. J. Mol. Biol. 368, 1172-1186.

Pagel Van Zee, J., Geraci, N.S., Guerrero, F.D., Wikel, S.K., Stuart, J.J., Nene, V.M., Hill, C.A., 2007. Tick genomics: the *Ixodes* genome project and beyond. Int. J. Parasitol. 37, 1297-1305.

Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061-1067.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785-786.

Potgieter, F.T., Stoltsz, W.H., Blouin, E.F., Roos, J.A., 1988. Corridor disease in South Africa: a review of the current status. J. S. Afr. Vet. Assoc. 59, 155-160.

Preston, S.G., Majtán, J., Kouremenou, C., Rysnik, O., Burger, L.F., Cabezas Cruz, A., Chiong Guzman, M., Nunn, M.A., Paesen, G.C., Nuttall, P.A., 2013. Novel immunomodulators from hard ticks selectively reprogramme human dendritic cell responses. PLoS Pathog. 9, e1003450.

Radley, D.E., Brown, C.G.D., Cunningham, M.P., Kimber, C.D., Musisi, F.L., Payne, R.C., Purnell, R.E., Stagg, S.M., Young, A.S., 1975. East Coast fever Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. Vet. Parasitol. 1, 51-60.

Ribeiro, J.M., Anderson, J.M., Manoukis, N.C., Meng, Z., Francischetti, I.M., 2011. A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. BMC Genomics 12, 136.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

Schwarz, A., Tenzer, S., Hackenberg, M., Erhart, J., Gerhold-Ay, A., Mazur, J., Kuharev, J., Ribeiro, J.M., Kotsyfakis, M., 2014. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. Mol. Cell. Proteomics 13, 2725-2735.

Schwarz, A., von Reumont, B.M., Erhart, J., Chagas, A.C., Ribeiro, J.M.C., Kotsyfakis, M., 2013. *De novo Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. FASEB J. 27, 4745-4756.

Sonenshine, D.E., 1991. Biology of ticks. Oxford University Press, New York, Oxford.

Sonenshine, D.E., Bissinger, B.W., Egekwu, N., Donohue, K.V., Khalil, S.M., Roe, R.M., 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, H., Mtambo, J., Tirry, L., Chaka, G., Marcotty, T., Berkvens, D., 2004. Variation in body size in the tick complex *Rhipicephalus appendiculatus*/*Rhipicephalus zambeziensis*. J. Vector Ecol. 29, 347-354.

Stoltsz, W.H., 1989. Theileriosis in South Africa: a brief review. Revue Scientifique et Technique, Office International des Épizooties 8, 93-102.

Tan, A.W., Francischetti, I.M., Slovak, M., Kini, R.M., Ribeiro, J.M., 2015a. Sexual differences in the sialomes of the zebra tick, *Rhipicephalus pulchellus*. J. 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, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.

Tirloni, L., Seixas, A., Mulenga, A., da Silva Vaz, I., Termignoni, C., 2014. A family of serine protease inhibitors (serpins) in the cattle tick *Rhipicephalus (Boophilus) microplus*. Exp. Parasitol. 137, 25-34.

Trimnell, A.R., Davies, G.M., Lissina, O., Hails, R.S., Nuttall, P.A., 2005. A cross-reactive tick cement antigen is a candidate broad-spectrum tick vaccine. Vaccine 23, 4329-4341.

Trimnell, A.R., Hails, R.S., Nuttall, P.A., 2002. Dual action ectoparasite vaccine targeting ‘exposed’ and ‘concealed’ antigens. Vaccine 20, 3560-3568.

Uilenberg, G., 1999. Immunization against diseases caused by *Theileria parva*: a review. Trop. Med. Int. Health 4, A12-20.

Wang, H., Nuttall, P., 1995. Immunoglobulin-G binding proteins in the ixodid ticks, *Rhipicephalus appendiculatus*, *Amblyomma variegatum* and *Ixodes hexagonus*. Parasitology 111, 161-165.

Wang, H., Paesen, G.C., Nuttall, P.A., Barbour, A.G., 1998. Male ticks help their mates to feed. Nature 391, 753-754.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., Li, W., 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 41, e74.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57-63.

Willadsen, P., 2004. Anti-tick vaccines. Parasitology 129, S367-S387.

Willadsen, P., 2006. Tick control: Thoughts on a research agenda. Vet. Parasitol. 138, 161-168.

Wilusz, J.E., Sunwoo, H., Spector, D.L., 2009. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23, 1494-1504.

Xiang, F.-Y., Zhou, Y.-Z., Zhou, J.-l., 2012. Identification of differentially expressed genes in the salivary gand of *Rhipicephalus haemaphysaloides* by the suppression subtractive hybridization approach. J. Integr. Agr. 11, 1528-1536.

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. Infect. Genet. Evol. 32, 135-142.

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 Res. (Database Issue) 34, (Web Server Issue), W293-W297.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.-L., Ponting, C.P., 2012. Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. Genome Biol. Evol. 4, 427-442.

Zivkovic, Z., Esteves, E., Almazán, C., Daffre, S., Nijhof, A.M., Kocan, K.M., Jongejan, F., de la Fuente, J., 2010. Differential expression of genes in salivary glands of male *Rhipicephalus (Boophilus) microplus* in response to infection with *Anaplasma marginale*. BMC Genomics 11, 186.