HHS Public Access

Author manuscript

Ticks Tick Borne Dis. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:

Ticks Tick Borne Dis. 2017 February ; 8(2): 201–207. doi:10.1016/j.ttbdis.2016.08.011.

An insight into the sialome of *Hyalomma excavatum*

José M.C. Ribeiro[a], **Mirko Slovák**[b], and **Ivo M.B. Francischetti**[a]

[a]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Bethesda MD 20852, USA

[b]Institute of Zoology, Slovak Academy of Sciences, Dúbravská cesta 9, 84506, Bratislava, Slovakia

Abstract

Tick saliva contains hundreds or thousands of proteins that help blood feeding by impairing their hosts' hemostasis, inflammation and immunity. Salivary gland transcriptomes allow the disclosure of this pharmacologically active potion that consists of several multi-gene families, many of which are tick-specific. We here report the "de novo" assembly of ~ 138 million reads deriving from a cDNA library from salivary glands of adult male and female *Hyalomma excavatum* leading to the public deposition of 5,337 coding sequences to GenBank. Among the deducted putative secreted proteins, metalloproteases, glycine rich proteins, mucins, anticoagulants of the madanin family and lipocalins were the most expressed. Novel protein families were identified. These sequences will permit proteomic studies aiming at identification of target antigens, epidemiological markers or salivary pharmaceuticals of interest, and contribute to our understanding of the fast evolution of the tick sialome.

Keywords

tick; salivary gland; saliva; transcriptome; sialome

Introduction

Tick saliva is represented by a complex assemblage of peptidic and non-peptidic compounds that assist blood feeding by disarming their host hemostasis and immunity (Chmelar et al., 2012; Francischetti et al., 2009; Kotal et al., 2015). Transcriptome studies indicate that ticks express hundreds or thousands of different polypeptides in their saliva that can be grouped into several multi-gene families, such as lipocalins, serpins, Kunitz-domain containing peptides and many other tick-specific families of unknown function (Chmelar et al., 2016; Francischetti et al., 2009). The diversity of these proteins is large, and probably derives from positive selection as their gene products benefit from mutations that evade, even temporarily, their hosts' immune system.

Correspondence to: José M.C. Ribeiro.

There are currently 64,150 protein sequences deposited on GenBank (as of Feb 11, 2016) annotated as deriving from tick salivary glands. While 26,370 are from the genus *Ixodes*, most from *I. ricinus*, the metastriate group counts 25,582 sequences for members of the *Amblyomma* genus and 11,513 for *Rhipicephalus*, most being for *R. pulchellus*. The genus *Hyalomma* is represented by only 100 sequences, 96 of which are from *H. marginatus rufipes.* In the present work we report the public deposition of 5,337 sequences derived from the sialotranscriptome of adult male and female *H. excavatum* ticks that fed for different amounts of time on white rabbits. *H. excavatum* (Apanaskevich and Horak, 2005; Hoogstraal and Kaiser, 1959) is of veterinary importance as a vector of *Theileria* to cattle and sheep in Africa (Friedhoff, 1997), and is a suspected vector of Crimean-Congo hemorrhagic fever (Khan et al., 1997).

Material and Methods

Ticks—Ticks were reared at the Institute of Zoology, Slovak Academy of Sciences. The original stock was the kind gift from Dr. Michael Samish, Kimron Veterinary Institute, Bait Dagan, Israel. White rabbits were used to rear all stages of this tick in the laboratory, as previously described (Slovak et al., 2002). The ticks were maintained in desiccators filled with saturated KCl solution to provide a RH of 85–90 %, at a photoperiod of 16:8 (L:D) and temperature 24 + 2 °C. The salivary glands were dissected from adult ticks originated from F3 laboratory generation that were unfed or fed at different times as indicated in Table 1. Glands were stored in RNAlater (Qiagen, Valencia CA) until used for mRNA extraction. The glands were pooled before RNA extraction. The usage of animals in these experiments was approved by the State Veterinary and Food Administration of the Slovak Republic (permit numbers 928/10-221 and 1335/12-221).

RNA extraction, library preparation and sequencing

RNA preparation, library construction and sequencing were performed essentially as described previously (Ribeiro et al., 2014). mRNA library construction and sequencing were done by the NIH Intramural Sequencing Center. The salivary gland (SG) library was constructed using the TruSeq RNA sample prep kit, v2 (Illumina Inc., San Diego, CA). The resulting cDNA was fragmented using a Covaris E210 (Covaris, Woburn, MA). Library amplification was performed using eight cycles to minimize the risk of over-amplification. Sequencing was performed on a HiSeq 2000 (Illumina) with v. 3 flow cells and sequencing reagents. One lane of the HiSeq machine was used for this and four other libraries, distinguished by bar coding. Libraries from the triatomine bug *Panstrongylus megistus*, the horse fly *Tabanus bromius* and the bat *Diphila ecaudata* were co-sequenced with *Hyalomma*, and we found that some cross-contamination of sequences occurred between the libraries (see below). Researchers reanalyzing the raw data should take this possibility of contamination into consideration. A total of 138,144,530 sequences of 101 nucleotides in length were obtained for the *Hyalomma* library. A paired-end protocol was used.

Bioinformatic tools used

The pipeline used has been described before (Ribeiro et al., 2015). Briefly, raw data were processed using RTA 1.12.4.2 and CASAVA 1.8.2. Reads were trimmed of low quality

regions and were assembled with the ABySS software (Genome Sciences Centre, Vancouver, BC, Canada) (Birol et al., 2009; Simpson et al., 2009) using various kmer (k) values (every tenth from 21 to 91) and SOAPdenovo-Trans assembler (Luo et al., 2012). The resulting assemblies were joined by an iterative BLAST and cap3 assembler (Karim et al., 2011). Sequence contamination between bar-coded libraries were identified and removed when their sequence identities were over 98% but their abundance of reads were > 10 fold between libraries. Coding sequences (CDS) were extracted using an automated pipeline based on similarities to known proteins or by obtaining CDS containing a signal peptide (Nielsen et al., 1999). CDS and their protein sequences were mapped into a hyperlinked Excel spreadsheet (presented as Supplemental File 1). Signal peptide, transmembrane domains, furin cleavage sites, and mucin-type glycosylation were determined with software from the Center for Biological Sequence Analysis (Technical University of Denmark, Lyngby, Denmark) (Duckert et al., 2004; Julenius et al., 2005; Nielsen et al., 1999; Sonnhammer et al., 1998). Reads were mapped into the contigs using blastn (Altschul et al., 1997) with a word size of 25, masking homonucleotide decamers and allowing mapping to up to three different CDS if the BLAST results had the same score values. Mapping of the reads was also included in the Excel spreadsheet. Values of the reads per kilobase of transcript per million mapped reads (RPKM) (Trapnell et al., 2012) for each coding sequence were also mapped to the spreadsheet. To compare relative expression of transcripts, we use the “expression index” defined as the number of reads mapped to a particular CDS divided by the largest found number of reads mapped to a single CDS, which in the case of this transcriptome was a value of 1,354,561 mapped to a single madanin coding sequence. Automated annotation of proteins was based on a vocabulary of nearly 350 words found in matches to various databases, including Swissprot, Gene Ontology, KOG, Pfam, and SMART, Refseq-invertebrates and the acari subset of the GenBank sequences obtained by querying acari [organism] and retrieving all protein sequences. Detailed bioinformatics analysis of our pipeline can be found in our previous publication (Karim et al., 2011). For determination of synonymous and non-synonymous sites within coding sequences, the tool BWA aln (Li and Durbin, 2010) was used to map the reads to the CDS, producing SAI files that were joined by BWA sampe module, converted to BAM format, and sorted. The sequence alignment/map tools (samtools) package (Li et al., 2009) was used to do the mpileup of the reads (samtools mpileup), and the binary call format tools (bcftools) program from the same package was used to make the final vcf file containing the single-nucleotide polymorphic (SNP) sites, which were only taken if the site coverage was at least 100 (–D100), the quality was 20 or better and the SNP frequency was 5 or higher (default). Determination of whether the SNPs lead to a synonymous or non-synonymous codon change was achieved by a program written in Visual Basic by JMCR, the results of which are mapped into the Excel spreadsheet and color visualized in hyperlinked rtf files within Additional File 1. Sequence alignments were done with the ClustalX software package (Thompson et al., 1997). Phylogenetic analysis and statistical neighbor-joining bootstrap tests (1,000 iterations) of the phylogenies were done with the Mega package (Kumar et al., 2004).

Data access

The raw reads were deposited on the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under bioproject ID PRJNA311286. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GEFH00000000. The version described in this paper is the first version, GEFH01000000. Hyperlinked excel spreadsheets containing the coding sequences and their annotation are available at http://exon.niaid.nih.gov/transcriptome/Hexcav/Hyaexcav-web.xlsx (hyperlinked excel spreadsheet, 21 MB).

Results and Discussion

Overview of the sialotranscriptome of *Hyalomma excavatum*

Following assembly of 138,144,530 reads, a total of 53,228 contigs were obtained (Supplemental file S1), from which we extracted 7,875 coding sequences. These coding sequences mapped 57,440,028 reads, or 42% of the total reads. Their average length was 1,125 nucleotides (nt) with 3,273 CDS being equal or larger than 1,000 nt. These CDS were classified into four classes: “secreted” (S), “housekeeping” (H), “unknown” (U) and “transposable elements” (TE) (Table 2). The S class had 1,796 assigned CDS, and mapped 61% of all reads in accordance with the secretory nature of the organ. The H class produced 5,511 CDS, mapping 36% of the reads. TE’s accounted for 2.24% of the CDS and 0.94% of the reads, a typical finding when comparing to other sialotranscriptomes. Finally, 390 CDS were not able to be classified, representing 1.12% of the reads.

The housekeeping CDS were further classified by their function (Table 3), not surprisingly showing the category “protein synthesis” to be the most expressed and accruing 19% of the reads of the H class, followed by the protein modification category, with 11.7% of the reads of the H class.

The classification of the putative secreted proteins is shown in table 4. Enzymes accounted for 7.8% of the reads, metalloproteases of the reprolysin family (Francischetti et al., 2003, 2005) accruing the majority of these reads. Glycine rich proteins, associated with glue proteins (Bishop et al., 2002; Maruyama et al., 2010) acquired 16% of the S class reads. Contigs coding for protease inhibitor domains collected 21 % of the reads, 4.6 % from single domain Kunitz proteins (Francischetti et al., 2009) and 13.6% from members of the Madanin anti-thrombin peptide family (Nakajima et al., 2006). Lipocalins accrued 12.8% of the S class reads. Immunosuppressor proteins of the DAP-36 (Bergman et al., 2000) and evasin (Deruaz et al., 2008; Frauenschuh et al., 2007) families accounted for 5% of the reads. Sixteen percent of the reads mapped to tick-specific families of unknown function, including eight new multigenic families that are disclosed in Table 4, one of which (5.6 kDa family) (Pichu et al., 2009) accumulated 4.6 % of the S class reads.

An insight into a selected sample of expanded families in the Hyalomma excavatum sialome

Reprolysins—Phylogenetic analysis of 45 sequences from *H. excavatum* encoding metalloproteases of the reprolysin family (Supplemental figure S1) aligned to their best

matches by blastp against the acari database from NCBI revealed 14 strong clades, several of which have strong sub clades to provide a total of 45 branches. The 45 *H. excavatum* proteins populate 23 of these branches, indicating this to be the minimum number of reprolysin coding genes for this tick, and emphasizing the expansion of this protein family in ticks which were associated with fibrinolytic properties and angiogenesis inhibition (Francischetti et al., 2003, 2005; Harnnoi et al., 2007). Notice that clade XIV, with three sub clades is the only one to contain sequences from the predatory mite *Metaseiulus occidentalis*, and may represent the ancestral gene that in ticks generated the other clades by gene or genome duplications. Proteins of this family have been targeted as anti-tick vaccines providing partial protection (Ali et al., 2015).

Lipocalins—Tick lipocalins have been shown to function as histamine, serotonin, thromboxane and cysteinyl-leukotriene kratagonists (Mans and Ribeiro, 2008; Paesen et al., 2000; Sangamnatdej et al., 2002), as well as anti-complement (Mans and Ribeiro, 2008; Nunn et al., 2005). Supplemental file S1 displays 201 lipocalin sequences having over 150 amino acids in length. Attempts to produce protein alignment-base phylogenies of these sequences were unsuccessful, as they produce too many nodes with poor bootstrap support. These proteins, however, can be grouped in 16 clusters of similarity as indicated in columns EJ-FM of supplemental file S1, and one group of 26 singletons. The larger cluster groups 73 sequences that share at least 40% similarity at a stretch of at least one half of the length of the larger sequence of the compared pair. Phylogeny analysis of these sequences indicate 3 super clades with strong bootstrap support and a minimum of 19 branches with sequences having more than 25% amino acid divergence (Supplemental figure S2). Group IX has the best match to the CDD histamine binding motif, having 4 sequences with only 30% identity and 78 % similarity. The role of these broad families of lipocalins in *Hyalomma* remains to be determined.

One-of-each family—The assembled sialome of *H. excavatum* reveals a cluster of 14 related sequences that has matches to uniquely tick proteins first described as “one of each” family because at the time only one protein member was found in each of the tick species analyzed (Francischetti et al., 2009). However, expanded transcriptomes revealed the multi-gene character of this family within single tick species (Karim et al., 2011). The *H. excavatum* sequences have relatively low degree of expression, the most expressed having an EI value of 0.15 and RPKM of 211. Alignment of these and related acari sequences (Fig S3A) reveals a single pair of conserved cysteines plus a relatively conserved series of mostly hydrophobic amino acids producing the block C-x(13,16)-[LFIV]-x(2)-[LIFVM]-x(16,18)-[ILMFV]-x(9)-[ILMV]-x(2)-[LFMIV]-x(8)-[FYH]-C-x(44,46)-[LF]. The deducted phylogenetic tree (Supplemental Fig S3) indicates nine robust clades, several of which further contain robust sub clades. Clade V contains seven *Hyalomma* sequences in three distinct subclades indicating possible genes that arose by tandem duplications. Overall, the 14 *H. excavatum* sequences populate nine individual clades, all except clade Va having at least one additional species, indicating the ancient duplication of this gene family. All *Ixodes*-genus derived sequences, as expected, are on a single distinct clade (VIII).

Novel 22 kDa family—Six deduced protein sequences from the assembled sialome of *H. excavatum* cluster at 35% similarity, having overall identities of only 6% and 16% similarity. They show a block of eight conserved cysteines plus a few other amino acids to produce the pattern C-x(7)-C-Y-x(35,39)-C-x(6,7)-C-x(9,11)-E-x-G-Y-x(21)-C-x(8)- C-x(25,31)-C-x(2)-E-x(4,6)-C-x(16,17)-A. They do not produce any significant matches to the non-redundant protein database from NCBI, but produces matches to tick proteins deposited in the TSA database. Alignment of these tick sequences (Fig S4A) produces a similar pattern block C-x(7)-C-Y-x(53,64)-C-x(6,7)-C-x(9,11)-E-x-G-Y-x(42,43)-C-x(8)-C-x(24,31)-C with one less Cys, despite all proteins having eight such amino acids. Phylogenetic analysis indicates seven robust clades under three super clades (Fig S4B). The *H. excavatum* proteins are relatively poorly expressed, with EI values equal or below 0.012.

Sialome polymorphism analysis

The deep coverage of high quality reads mapped to the deducted coding sequences allows for identification of single nucleotide polymorphisms and determination of synonymous (S) and non-synonymous (NS) mutations within the diverse functional classes of the *H. excavatum* sialome. Toward this end we calculated the average number of S and NS mutations per 100 codons in transcripts accumulating a minimum average of 100 fold nucleotide base coverage. Results (Table 5) indicate that the Unknown, secreted and transposable element classes have the highest ratio of NS to S mutations, supporting the fast rate of evolution of tick sialomes.

We have also compared the degree of similarity of the deducted CDS from *H. excavatum* to other proteins from Acari (excluding H. excavatum) downloaded from GenBank. In a set of 5,898 CDS that produces at least 75% coverage of the best match by blastp, again the Unknown, Secreted and Transposable Element classes appear as those having the least similarities to their counterparts (Table 6), as expected from the larger NS/S rate of these gene classes.

Conclusions

The assembly of the *H. excavatum* sialome following an Illumina protocol contributed to the public disclosure of 5,337 CDS and proteins to the TSA archive of GenBank, thus helping to enlarge the databank of acari sequences in general and those of salivary secreted proteins in particular. Because tick salivary secreted proteins are at a fast pace of evolutionary change, each tick species needs its own sialome solved if it is to be used as a source of vaccine targets or epidemiological markers of exposure. Complete or near complete sialomes make possible proteomics experiments to attempt molecular identification of antigens or other products of interest, as well as producing new targets of interest in unrelated evolutionary studies (Kasuya et al., 2016).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

JMCR and IMBF were supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases. MS was partially supported by the Slovak Research and Development Agency (APVV-0737-12) and by Slovak VEGA grant 2/0089/13. Because JMCR and IMBF are government employees and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. You can establish rights outside of the U.S. subject to a government use license.

References cited

Ali A, Parizi LF, Guizzo MG, Tirloni L, Seixas A, Vaz da IS Jr, Termignoni C. Immunoprotective potential of a *Rhipicephalus (Boophilus) microplus* metalloprotease. Veterinary parasitology. 2015; 207:107–114. [PubMed: 25480468]

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]

Apanaskevich DA, Horak IG. The genus Hyalomma Koch, 1844. II Taxonomic status of *H (Euhyalomma) anatolicum* Koch, 1844 and *H (E.) excavatum* Koch, 1844 (Acari, Ixodidae) with redescriptions of all stages. Acarina. 2005; 13:181–197.

Bergman DK, Palmer MJ, Caimano MJ, Radolf JD, Wikel SK. Isolation and molecular cloning of a secreted immunosuppressant protein from *Dermacentor andersoni* salivary gland. J Parasitol. 2000; 86:516–525. [PubMed: 10864249]

Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, Horsman DE, Connors JM, Gascoyne RD, Marra MA, Jones SJ. De novo transcriptome assembly with ABySS. Bioinformatics (Oxford, England). 2009; 25:2872–2877.

Bishop R, Lambson B, Wells C, Pandit P, Osaso J, Nkonge C, Morzaria S, Musoke A, Nene V. A cement protein of the tick *Rhipicephalus appendiculatus*, located in the secretory e cell granules of the type III salivary gland acini, induces strong antibody responses in cattle. International journal for parasitology. 2002; 32:833–842. [PubMed: 12062554]

Chmelar J, Calvo E, Pedra JH, Francischetti IM, Kotsyfakis M. Tick salivary secretion as a source of antihemostatics. Journal of proteomics. 2012; 75:3842–3854. [PubMed: 22564820]

Chmelar J, Kotal J, Kopecky J, Pedra JH, Kotsyfakis M. All For One and One For All on the Tick-Host Battlefield. Trends in parasitology. 2016

Deruaz M, Frauenschuh A, Alessandri AL, Dias JM, Coelho FM, Russo RC, Ferreira BR, Graham GJ, Shaw JP, Wells TN, Teixeira MM, Power CA, Proudfoot AE. Ticks produce highly selective chemokine binding proteins with antiinflammatory activity. The Journal of experimental medicine. 2008; 205:2019–2031. [PubMed: 18678732]

Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel. 2004; 17:107–112. [PubMed: 14985543]

Francischetti IM, Mather TN, Ribeiro JM. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. Biochemical and biophysical research communications. 2003; 305:869–875. [PubMed: 12767911]

Francischetti IM, Mather TN, Ribeiro JM. Tick saliva is a potent inhibitor of endothelial cell proliferation and angiogenesis. Thrombosis and haemostasis. 2005; 94:167–174. [PubMed: 16113800]

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

Frauenschuh A, Power CA, Deruaz M, Ferreira BR, Silva JS, Teixeira MM, Dias JM, Martin T, Wells TN, Proudfoot AE. Molecular cloning and characterization of a highly selective chemokine-binding protein from the tick *Rhipicephalus sanguineus*. The Journal of biological chemistry. 2007; 282:27250–27258. [PubMed: 17640866]

Friedhoff KT. Tick-borne diseases of sheep and goats caused by Babesia, Theileria or Anaplasma spp. Parassitologia. 1997; 39:99–109. [PubMed: 9530692]

Harnnoi T, Sakaguchi T, Nishikawa Y, Xuan X, Fujisaki K. Molecular characterization and comparative study of 6 salivary gland metalloproteases from the hard tick, *Haemaphysalis longicornis*. Comparative biochemistry and physiology. 2007; 147:93–101. [PubMed: 17292650]

Hoogstraal H, Kaiser MN. Observations on Egyptian Hyalomma ticks (Ixodoidea, Ixodidae). 5 Biological notes and differences in identity of *H anatolicum* and its subspecies *anatolicum* Koch and *excavatum* Koch among Russian and other workers Identity of *H lusitanicum*. Koch Annals of the Entomological Society of America. 1959; 52:243–261.

Julenius K, Molgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology. 2005; 15:153–164. [PubMed: 15385431]

Karim S, Singh P, Ribeiro JM. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS ONE. 2011; 6:e28525. [PubMed: 22216098]

Kasuya G, Fujiwara Y, Takemoto M, Dohmae N, Nakada-Nakura Y, Ishitani R, Hattori M, Nureki O. Structural Insights into Divalent Cation Modulations of ATP-Gated P2X Receptor Channels. Cell reports. 2016; 14:932–944. [PubMed: 26804916]

Khan AS, Maupin GO, Rollin PE, Noor AM, Shurie HH, Shalabi AG, Wasef S, Haddad YM, Sadek R, Ijaz K, Peters CJ, Ksiazek TG. An outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates, 1994–1995. The American journal of tropical medicine and hygiene. 1997; 57:519–525. [PubMed: 9392589]

Kotal J, Langhansova H, Lieskovska J, Andersen JF, Francischetti IM, Chavakis T, Kopecky J, Pedra JH, Kotsyfakis M, Chmelar J. Modulation of host immunity by tick saliva. Journal of proteomics. 2015; 128:58–68. [PubMed: 26189360]

Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in bioinformatics. 2004; 5:150–163. [PubMed: 15260895]

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2010; 26:589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25:2078–2079.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012; 1:18. [PubMed: 23587118]

Mans BJ, Ribeiro JM. Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. Insect biochemistry and molecular biology. 2008; 38:841–852. [PubMed: 18694828]

Maruyama SR, Anatriello E, Anderson JM, Ribeiro JM, Brandao LG, Valenzuela JG, Ferreira BR, Garcia GR, Szabo MP, Patel S, Bishop R, de Miranda-Santos IK. 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. 2010; 11:363. [PubMed: 20529354]

Nakajima C, Imamura S, Konnai S, Yamada S, Nishikado H, Ohashi K, Onuma M. A novel gene encoding a thrombin inhibitory protein in a cDNA library from *Haemaphysalis longicornis* salivary gland. The Journal of veterinary medical science /the Japanese Society of Veterinary Science. 2006; 68:447–452.

Nielsen H, Brunak S, von Heijne G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein engineering. 1999; 12:3–9. [PubMed: 10065704]

Nunn MA, Sharma A, Paesen GC, Adamson S, Lissina O, Willis AC, Nuttall PA. Complement inhibitor of C5 activation from the soft tick *Ornithodoros moubata*. J Immunol. 2005; 174:2084–2091. [PubMed: 15699138]

Paesen GC, Adams PL, Nuttall PA, Stuart DL. Tick histamine-binding proteins: lipocalins with a second binding cavity. Biochim Biophys Acta. 2000; 1482:92–101. [PubMed: 11058751]

Pichu S, Ribeiro JM, Mather TN. Purification and characterization of a novel salivary antimicrobial peptide from the tick, *Ixodes scapularis*. Biochemical and biophysical research communications. 2009; 390:511–515. [PubMed: 19852941]

Ribeiro JM, Chagas AC, Pham VM, Lounibos LP, Calvo E. An insight into the sialome of the frog biting fly, *Corethrella appendiculata*. Insect biochemistry and molecular biology. 2014; 44:23–32.

Ribeiro JMC, Schwarz A, Francischetti IMB. A deep insight into the sialotranscriptome of the Chagas disease vector, *Panstrongylus megistus* (Hemiptera: Heteroptera). J Med Entomol. 2015; 52:351–358. [PubMed: 26334808]

Sangamnatdej S, Paesen GC, Slovak M, Nuttall PA. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. Insect Mol Biol. 2002; 11:79–86. [PubMed: 11841505]

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome research. 2009; 19:1117–1123. [PubMed: 19251739]

Slovak M, Labuda M, Marley SE. Mass laboratory rearing of *Dermacentor reticulatus* ticks (Acarina, Ixodidae). Biologia (Bratisl). 2002; 57:261–266.

Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998; 6:175–182. [PubMed: 9783223]

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25:4876–4882. [PubMed: 9396791]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012; 7:562–578. [PubMed: 22383036]

Table 1

Time of feeding, number of ticks and tick gender from which salivary glands were removed into RNAlater for library construction.

Fed	Females	Males
0	4	4
60–90 min	4	4
230–260 min	4	4
6–6.5 hr	4	4
15 hr	4	4
1 day	4	4
2 days	3	3
3 days	3	3
4 days	3	3
5 days	3	3
6 days	3	3
7 days	3	3
8 days	3	3
9 days - females dropped off	3	3
Total	48	48

Table 2

Functional classification of the coding sequences (CDS) found in the sialotranscriptome of *Hyalomma excavatum*

Class	Number of CDS	Number of reads	% of CDS	% of reads	Reads /CDS
Secreted	1,796	35,383,379	22.81	61.60	19,701
Housekeeping	5,511	20,872,968	70.00	36.34	3,788
Transposable elements	176	539,446	2.24	0.94	3,065
Unknown	390	644,161	4.95	1.12	1,652
Total	7,873	57,439,954	100	100	

Table 3

Functional classification of the putative housekeeping sialotranscriptome of *Hyalomma excavatum.*

Classification	Number of CDS	Number of reads	% of Reads
Protein synthesis machinery	347	4,007,931	19.20
Protein modification	266	2,451,117	11.74
Signal transduction	780	2,396,282	11.48
Unknown conserved	1039	2,227,867	10.67
Energy metabolism	170	1,258,093	6.03
Transcription machinery	543	1,211,667	5.80
Protein export	344	1,057,686	5.07
Cytoskeletal proteins	216	856,980	4.11
Transporters and channels	286	845,006	4.05
Extracellular matrix	129	761,140	3.65
Carbohydrate metabolism	159	631,972	3.03
Proteasome machinery	227	613,300	2.94
Lipid metabolism	230	426,228	2.04
Nuclear regulation	223	404,755	1.94
Detoxification	62	294,937	1.41
Nuclear export	34	256,277	1.23
Amino acid metabolism	82	243,965	1.17
Transcription factor	130	216,481	1.04
Storage	10	201,302	0.96
Nucleotide metabolism	97	166,480	0.80
Immunity	50	146,911	0.70
Oxidant metabolism/Detoxification	36	119,669	0.57
Intermediary metabolism	51	76,922	0.37
Total	5,511	20,872,968	100.00

Table 4

Functional classification of the putative secreted sialotranscriptome of *Hyalomma excavatum.*

Classification	Number of CDS	Number of reads	% of Reads
Enzymes			**7.80**
Apyrase	10	83,312	0.24
Endonuclease	8	27,588	0.08
Lipases	13	16,645	0.05
Metalloprotease			
ADAMS	1	646	0.00
M13 metalloprotease	13	79,942	0.23
Salivary reprolysin	46	2,217,522	**6.27**
Serine carboxypeptidase	5	196,280	0.55
Other peptidases	27	139,444	0.39
Glycine rich proteins			**16.14**
Glycine rich 35-2-62 family	39	5,661,276	16.00
GRP-3	4	47,638	0.13
Other Gly rich proteins	1	259	0.00
Mucins	26	143,948	0.41
Protease inhibitor domains			**20.90**
Kunitz domains			
Monolaris	40	1,632,033	4.61
Monolaris - non canonical	4	2,410	0.01
Bilaris	17	152,257	0.43
Pentalaris	2	22,988	0.06
Kunitz-like	9	343,155	0.97
Madanin	9	4,824,396	13.63
TIL domain			
MonoTil	21	171,899	0.49
BiTil	12	99,335	0.28
Tritil	2	101,388	0.29
Other inhibitor domains			
Kazal	2	4,398	0.01
Serpin	10	10,749	0.03
Cystatin	5	20,875	0.06
Thyropin	1	6,544	0.02
Tick carboxypeptidase inhibitor	2	3,282	0.01
Lipocalins			**12.70**
Group I	73	1,309,162	3.70
Group II	26	104,781	0.30
Group III	9	13,735	0.04
Group IV	7	232,112	0.66
Group V	5	359,476	1.02

Classification	Number of CDS	Number of reads	% of Reads
Group VI	5	227,600	0.64
Group VII	4	224,764	0.64
Group VIII	4	1,829	0.01
Group IX	4	66,727	0.19
Group X	4	3,029	0.01
Group XI	9	30,581	0.09
Group XII	12	432,721	1.22
Group XIII	12	659,589	1.86
Other lipocalins	31	827,246	2.34
Antigen-5	7	30,097	**0.09**
Immunity related			**4.97**
Evasin	34	478,632	1.35
Evasin 35–29	12	239,664	0.68
DAP-36	21	1,039,088	2.94
Antimicrobial			**0.48**
Defensin, truncated	1	9,218	0.03
5.3 kDa/defensin family	7	20,414	0.06
Microplusin family	8	65,093	0.18
Lysozyme	3	43,730	0.12
Pathogen recognition motifs			
Peptidoglycan binding protein	4	26,770	0.08
Ixoderin	1	627	0.00
ML domain containing protein	6	4,011	0.01
Unknown function, tick specific			**16.32**
Basic Tail	9	999,985	2.83
23 kDa family	3	2,191	0.01
8.9 kDa family	57	1,443,831	**4.08**
18.3 kDa family	8	160,949	0.45
28 kDa metastriate family	8	204,111	0.58
Amblyomma 40–33 family member	10	15,220	0.04
One of each	14	34,322	0.10
New families			
New 8 kDa family	12	1,032,674	2.92
Divergent 8 kDa family	9	15,553	0.04
New 5.6 kDa protein family	29	1,628,455	**4.60**
New 11 kDa family	10	23,987	0.07
New 15 kDa family	6	118,230	0.33
New 22.5 kDa family	11	91,768	0.26
New 22 kDa family	7	2,693	0.01
Conserved secreted proteins of unknown function	80	1,172,112	**3.31**
Other secreted proteins	885	5,976,393	**16.89**
Total	1,796	35,383,379	100.00

Table 5

Synonymous and Non-synonymous single nucleotide polymorphisms in the *Hyalomma excavatum* sialome.

Class	Average read coverage[1]	SE	Average Syn/100 codons	SE	Average NS/100 codons	SE	NS/Syn	N
Unknown	2,574.9	864.6	0.4437	0.1056	1.4468	0.2482	3.2610	84
Secreted	5,231.1	404.5	0.5542	0.0407	0.6129	0.0438	1.1060	842
Transposable element	935.7	560.9	0.7037	0.2300	0.5817	0.1707	0.8267	26
Extracellular matrix/cell adhesion	868.8	419.0	0.7685	0.1522	0.3004	0.0904	0.3909	57
Oxidant metabolism/detoxification	1,244.2	703.3	0.9031	0.2521	0.3521	0.1272	0.3899	40
Protein modification machinery	1,292.7	439.1	0.6474	0.0824	0.2178	0.0512	0.3364	147
Unknown, conserved	533.5	104.0	0.8842	0.0759	0.2228	0.0270	0.2520	378
Metabolism	805.5	69.1	1.0034	0.0746	0.2124	0.0255	0.2116	368
Protein synthesis machinery	2,680.7	243.4	1.0669	0.0971	0.2229	0.0270	0.2089	250
Signal transduction	741.8	255.6	0.7747	0.0785	0.1497	0.0308	0.1933	250
Nuclear regulation	546.7	121.8	0.6635	0.2055	0.1205	0.0450	0.1816	58
Transcription machinery	377.8	57.6	0.7316	0.0765	0.1306	0.0300	0.1786	205
Transporters/storage	531.0	103.5	0.9838	0.1308	0.1584	0.0338	0.1610	107
Cytoskeletal	779.4	226.5	1.2110	0.2085	0.1570	0.0370	0.1297	83
Transcription factor	363.6	83.8	1.1184	0.2585	0.1231	0.0419	0.1101	41
Proteasome machinery	408.1	69.1	0.8273	0.1363	0.0878	0.0249	0.1061	116
Immunity	560.2	229.3	0.6837	0.2757	0.0502	0.0356	0.0734	14
Protein export machinery	511.9	53.3	0.8751	0.0923	0.0546	0.0152	0.0624	193

[1]Only contigs with average read coverage depth of 100 or higher were used in this analysis.

Table 6

Identities between *Hyalomma excavatum* proteins and their best matching Acari proteins.

Class	Average % identity [1]	Standard Error	Number of individuals
Unknown	35.9574	1.8024	47
Secreted	58.8467	0.7033	972
Transposable element	63.7174	2.0130	92
Extracellular matrix/cell adhesion	80.2430	1.8518	107
Oxidant metabolism/detoxification	83.2500	1.4899	84
Immunity	83.9211	2.8959	38
Transcription factor	84.3738	1.5261	107
Unknown, conserved	85.4036	0.4901	887
Protein modification machinery	86.8193	0.8867	249
Nuclear export	88.2500	2.4064	24
Metabolism	88.6545	0.3950	715
Nuclear regulation	88.9133	0.8533	196
Transporters/storage	89.1783	0.7875	230
Cytoskeletal	89.8652	0.9794	178
Transcription machinery	90.0167	0.5001	479
Signal transduction	90.1700	0.4314	641
Protein synthesis machinery	90.3363	0.4692	333
Proteasome machinery	91.2150	0.7205	200
Protein export machinery	93.8401	0.4158	319

[1] Comparisons only made if *H. excavatum* query covered at least 75% of the best match.