



TickSialoFam (TSFam): A Database That Helps to Classify Tick Salivary Proteins, a Review on Tick Salivary Protein Function and Evolution, With Considerations on the Tick Sialome Switching Phenomenon

OPEN ACCESS

José M. C. Ribeiro ^{1*} and Ben J. Mans ^{2,3,4}

Edited by:

Brice Rotureau,
Institut Pasteur, France

Reviewed by:

Jindrich Chmelar,
University of South Bohemia in Ceské
Budějovice, Czechia
Daniel Sojka,
Institute of Parasitology
(ASCR), Czechia
James Jason Valdes,
Institute of Parasitology
(ASCR), Czechia

*Correspondence:

José M. C. Ribeiro
jrbeiro@niaid.nih.gov

Specialty section:

This article was submitted to
Parasite and Host,
a section of the journal
Frontiers in Cellular and Infection
Microbiology

Received: 28 March 2020

Accepted: 17 June 2020

Published: 24 July 2020

Citation:

Ribeiro JMC and Mans BJ (2020)
TickSialoFam (TSFam): A Database
That Helps to Classify Tick Salivary
Proteins, a Review on Tick Salivary
Protein Function and Evolution, With
Considerations on the Tick Sialome
Switching Phenomenon.
Front. Cell. Infect. Microbiol. 10:374.
doi: 10.3389/fcimb.2020.00374

¹ Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Rockville, MD, United States, ² Epidemiology, Parasites and Vectors, Agricultural Research Council - Onderstepoort Veterinary Research, Pretoria, South Africa, ³ The Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa, ⁴ Department of Life and Consumer Sciences, University of South Africa, Pretoria, South Africa

Tick saliva contains a complex mixture of peptides and non-peptides that counteract their hosts' hemostasis, immunity, and tissue-repair reactions. Recent transcriptomic studies have revealed over one thousand different transcripts coding for secreted polypeptides in a single tick species. Not only do these gene products belong to many expanded families, such as the lipocalins, metalloproteases, Antigen-5, cystatins, and apyrases, but also families that are found exclusively in ticks, such as the evasins, Isac, DAP36, and many others. Phylogenetic analysis of the deduced protein sequences indicate that the salivary genes exhibit an increased rate of evolution due to a lower evolutionary constraint and/or positive selection, allowing for a large diversity of tick salivary proteins. Thus, for each new tick species that has its salivary transcriptome sequenced and assembled, a formidable task of annotation of these transcripts awaits. Currently, as of November 2019, there are over 287 thousand coding sequences deposited at the National Center for Biotechnology Information (NCBI) that are derived from tick salivary gland mRNA. Here, from these 287 thousand sequences we identified 45,264 potential secretory proteins which possess a signal peptide and no transmembrane domains on the mature peptide. By using the psiblast tools, position-specific matrices were constructed and assembled into the TickSialoFam (TSF) database. The TSF is a rpsblastable database that can help with the annotation of tick sialotranscriptomes. The TSA database identified 136 tick salivary secreted protein families, as well as 80 families of endosomal-related products, mostly having a protein modification function. As the number of sequences increases, and new annotation details become available, new releases of the TSF database may become available.

Keywords: tick, saliva, transcriptome, annotation, sialome switching, feeding, salivary glands

INTRODUCTION

Tick salivary glands serve multiple physiological roles: While off their hosts, special acini (type I) produce a hygroscopic saliva that spreads over the ticks' palps, absorbs atmospheric water vapor, and when re-ingested helps to keep the tick hydrated (Bowman and Sauer, 2004). The salivary glands are also a major excretion/water balance organ in ixodid ticks, as the majority of the water ingested during a blood meal is pumped back into their hosts as saliva (Binnington and Kemp, 1980). In ixodid males, due to their peculiar and successful oral sex reproduction mode, special male acini (type IV) contribute to salivary products that are uniquely male and are molecular homologs of insect seminal gland or vertebrate prostate gland proteins (Tan et al., 2015). Finally, and the focus of this article, tick salivary glands help blood feeding by secreting products that help tick attachment to the host's skin and that inhibit host hemostasis (a physiological response that encompasses platelet aggregation, vasoconstriction, and blood clotting), inhibit the vertebrate tissue repair response, and modulate host immunity (Francischetti et al., 2009).

There are 899 known tick species divided into two major families, the soft tick (*Argasidae*) and the hard tick (*Ixodidae*) families (Guglielmone et al., 2010). A third family, *Nuttalliellidae*, exists, but with a single species. Soft ticks feed relatively rapid (typically <1 h), while hard ticks feed for several days or weeks. Adult soft ticks can feed many times, while adult hard ticks feed only once. Hard ticks are further divided into Prostriate and Metastriate. Metastriate ticks have relatively short mouthparts, producing copious amounts of cement, while Prostriate ticks have longer mouthparts, producing less abundant cement. Adult and nymphal soft ticks do not produce cement. Blood is the only nutritious food for all ticks. It is estimated that there are over 15,000 species of arthropods that feed on blood, and this mode of feeding evolved independently at least 20 times (Mans, 2011)—thus creating a scenario of convergent evolution. Ticks, mosquitoes, sand flies, kissing bugs, and fleas evolved blood feeding independently, but they all share the presence of a salivary apyrase activity, an enzyme (ATP-diphosphohydrolase) that breaks down ATP and ADP released by damaged cells and trigger platelet and neutrophil activation. However, three different gene families were recruited for this task: (1) the 5'-nucleotidase in ticks (Stutzer et al., 2009), some genera of triatomine bugs, and mosquitoes; (2) the Cimex-type apyrase (CD73) in sand flies and the kissing bug genus *Rhodnius*; (3) the CD39 apyrase in fleas (Ribeiro and Arca, 2009). These types of enzymes are ubiquitous in eukaryotes, normally extracellular and bound to the membrane or intracellular and associated to endosomes; their transcripts always display a signal peptide indicative of secretion. Evolution of salivary apyrase, as an adaptation to blood feeding, thus happened by processes of gene duplication of a salivary gland expressed gene with loss of its membrane anchor (Champagne et al., 1995). Another common evolutionary process of adaptation to blood feeding involved further gene duplication events of already established salivary genes, which initially increased the mRNA dosage, but allowed for further functional or antigenic diversification. For example, sand flies and mosquitoes have several related salivary expressed

genes of the D7 family, which belong to the odorant binding protein group, and are associated with binding of agonists of hemostasis and inflammation (Valenzuela et al., 2002a; Mans et al., 2007; Calvo et al., 2009; Alvarenga et al., 2010). Ticks and blood sucking Hemiptera had a large expansion of the lipocalin family, which are also associated with binding of agonists of hemostasis and inflammation (Andersen and Ribeiro, 2017). Fleas expanded members of the acid phosphatase family (function unknown) (Andersen et al., 2007; Ribeiro et al., 2012).

While Adult sand fly saliva has <50 polypeptides and mosquito saliva has near 100, hard tick saliva has several hundred, or thousands of polypeptides (Ribeiro and Arca, 2009; Ribeiro et al., 2010). The increased number of components in hard tick saliva was initially thought to derive from the prolonged period of tick feeding, where it would face not only the hemostatic host response, but also its tissue repair and immune responses. However, deep sequencing of salivary transcriptomes at different times of feeding uncovered that ticks change their salivary repertoire frequently, within hours; a process named "sialome switching" (Valenzuela et al., 2002b; Karim and Ribeiro, 2015; Perner et al., 2018). Thus, while in mosquitoes and other non-tick hematophagous organisms, all members of a salivary protein family are simultaneously expressed; in hard ticks, individual genes of the same family, or paralogs, are expressed at different times during the feeding process. Accordingly, when the host mounts an antibody response to a particular antigen, a process that takes a few days, chances are that the antigenic molecule has been substituted by a not antigenically recognizable paralogs. Sialome switching is thus thought to be a mode of immune evasion. It may also serve ticks when they feed on different host species, as they may "find" the best sialome when feeding on a lizard, which may be different then when feeding on a mouse. Sialome switching may thus be a mechanism adapting ticks to feed on different hosts.

The study of the function of saliva in blood feeding by arthropods in the last 30 years was transformed by the revolution of cheap DNA sequencing. In the beginning it was the "grind and find" way: salivary homogenates were used to discover activities determined by a bioassay; followed by chromatographic methods of purification of the biological activity; in the case it was a polypeptide, obtaining Edman sequences allowing the construction of degenerate DNA probes that were used to amplify fragments of the salivary cDNA; that would be used to find, by hybridization, and with luck, a full length clone from which the primary sequence of the studied protein could be determined. The research flow was then from the bioassay to the protein, and from the protein to the DNA or mRNA.

Currently, for <300 \$, we can obtain 20 million sequences of 150 nucleotides (nt) in length that can be assembled "de novo" (in the absence of a genomic sequence to serve as an assembly scaffold), producing high quality transcript sequences that can be converted to their coding protein sequences. Accordingly, the "sialome" (from the Greek σίελος = saliva) can be obtained inexpensively and in a relatively short amount of time (Ribeiro and Francischetti, 2003). These sialomes have revealed a surprising number of novel protein sequences; they frequently have no similar matches when comparing their

primary sequence to a bank of known proteins. Less frequently, they produce matches to ubiquitous protein families, usually enzymes, but also members of the antigen-5 family or families of protease inhibitors, such as those containing cystatin, serpin, Kazal, or Kunitz domains. These deduced protein sequences can be further used in two ways: First, in the absence of the organism's genomic sequences, they will serve as a data bank for proteomic studies, a technology that requires a reliable "a priori" defined set of sequences; the tandem mass spectrometry (MS/MS) protocol maps the partial sequences.

Theoretically, if the genome sequence of the organism under study is known, then its deduced protein sequences could serve this purpose. However, most blood sucking insects and ticks still do not have a reliable genome sequence. The genome of the tick *Ixodes scapularis* is at best 50% complete (Gulia-Nuss et al., 2016). Even with better genomic data information, such as the anopheline mosquitoes (Holt et al., 2002; Neafsey et al., 2015), a detailed comparison of the protein sequences known to be salivary expressed with the deduced genomic protein sequences indicated that there was an error on ~50% of the predicted genomic sequences; either because they were missed, or because the intron-exon boundaries were incorrect (Arca et al., 2017). Proteomic studies of the salivary glands of blood sucking arthropods thus rely on carefully curated protein sequences deduced from the salivary transcriptome assembly. Second, the deduced protein sequences can guide production of recombinant proteins to be tested in bioassays to determine their functions, or as immunological markers of vector exposure. Thus, the research flows today from the mRNA to the protein, and from the protein to the bioassay.

Another approach exists, namely the immuno-proteome method, which is based on high throughput cloning and sequencing methods aimed at anti-tick vaccine discovery (Das et al., 2001; de la Fuente et al., 2006; Narasimhan et al., 2007; Radulovic et al., 2014; Becker et al., 2015; Lewis et al., 2015; Garcia et al., 2017). First, an animal is made hyperimmune to tick salivary proteins by previous tick exposure, or injection of salivary homogenates. The hyperimmune serum, and its control, are then used to scan expression cDNA libraries made from tick salivary glands. The positive clones are then sequenced, recombinantly expressed, and tested for their biological activity and as a vaccine to disrupt tick feeding. It has succeeded in discovering and characterizing many tick salivary proteins (Narasimhan et al., 2007; Dai et al., 2009; Schuijt et al., 2010). Somewhat surprisingly, many intracellular tick proteins, not having a signal peptide indicative of secretion, were discovered to be good antigen candidates; as was the case of the conserved tick protein, subolesin (de la Fuente et al., 2006), a member of the Akirin transcription regulator (de la Fuente et al., 2011). Similarly, a tick protein similar to vertebrate histamine releasing factor was found (Mulenga et al., 2003), and named histamine release factor (HRF). This is also a conserved member of the cytoskeletal family named "Translationally controlled tumor protein" (TCTP) and was responsible for the activity of human tissue homogenates to trigger histamine release by mast cells (Xiao et al., 2017). Human TCTP thus has a dual function: as a cytoskeletal protein: and as a cytokine-like protein when cells lyse

and their contents are released to the extracellular milieu. ADP and ATP similarly are primarily intracellular molecules; however when released to the extracellular compartment after cellular disruption or secretion, they similarly trigger pro inflammatory reactions. It thus appears that tick saliva contains proteins that are secreted through the classic merocrine pathway, as well as non-conventional secretion pathways, including apocrine or holocrine secretions (Farkas, 2015). Indeed, merocrine and apocrine secretion has been characterized by ultrastructural studies in soft tick salivary glands (Coons and Roshdy, 1981). Hard tick salivary glands suffer a degeneration process after a blood meal (Harris and Kaufman, 1981; Friesen and Kaufman, 2009), and it is possible that this process starts while the tick is still feeding and thus generating a holocrine salivary secretion. Additionally, there is recent evidence that exosomes may be secreted in tick saliva (Hackenberg and Kotsyfakis, 2018; Zhou et al., 2018; Chavez et al., 2019). Exosomes can fuse with host cells and deliver their products intracellularly, including microRNAs that may inhibit translation of some protein types (Keller et al., 2006). The study of salivary exosomal secretion in tick feeding is still in its infancy, and the next decade should uncover its more precise role in feeding by ticks.

Ten years ago, a review on tick sialomes based on 3,500 tick salivary proteins identified more than 30 protein families, most having unknown function (Francischetti et al., 2009). That review was done when the first sialomes were uncovered with the now obsolete Sanger DNA sequencing method. At that point in time, the now defunct 454 and surviving Illumina protocols appeared, and the number of salivary-derived tick protein-coding sequences deposited to the National Center for Biotechnology Information (NCBI) soared, reaching over 287,000 in November/2019 (**Figure 1**). These sequences derive from 44 species within 10 genera; only 19 species from 7 genera are represented with more than one thousand sequences (**Table 1**). Considering that there are 899 tick species from 20 different genera (Guglielmone et al., 2010), we have achieved 1,000 sequences for only 2.7% of the known tick species, which covers 35% of the known genera. Clearly, we are still far from knowing the complete sialome repertoire of ticks (the sialoverse), a task that should take one or more decades for its fulfillment.

To facilitate the task of annotating newly assembled transcriptomes, we present a rpsblastable database of tick salivary families, named TickSialoFam or TSFam. Together with an update on the classification of tick salivary families as proposed in 2009 (Francischetti et al., 2009), we provide further insights into the evolution of the genes coding for these proteins.

METHODS

Sequence Retrieval and Organization

The nucleotide database from NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide/>) was queried with the expression "ixodida[organism] AND salivary," and the resulting records were downloaded in the gb (GenBank) format. Sequences with "mitochondr" in their description were excluded, as well as those results related to unassembled EST's, and those from the REFSEQ database. This resulted in a set of 287,343 sequences

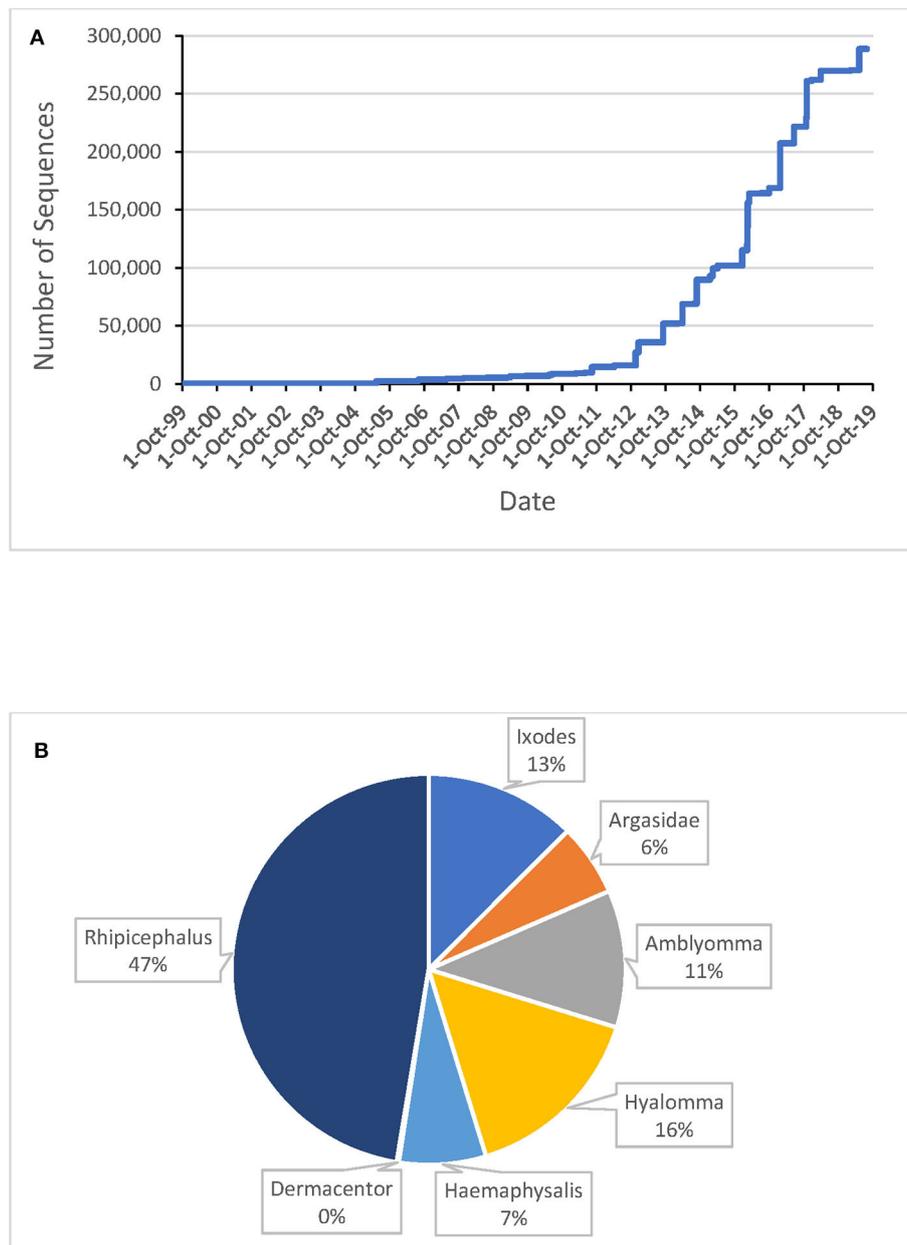


FIGURE 1 | Tick salivary protein coding sequences (excluding EST's) deposited at the National Center for Biotechnology Information (NCBI). **(A)** Number of sequences deposited, 1993–2019. **(B)** Breakdown of sequences available in Nov/2019 by genus or family.

(Table 1). It should be noted that up to mid-2018, all protein sequences submitted to the NCBI Transcriptome Shotgun Annotation (TSA) database were deposited in the NCBI protein database, as were the nucleotide sequences deposited to the NCBI nucleotide database. After this time, all TSA submissions appear only in the nucleotide database, solely as a link leading to downloading the nucleotide and protein sequences. These TSA databases were then downloaded and added to those of the nucleotide database. A script written in visual basic version 6 (VB6) extracted the individual protein sequence in fasta format; this

script also built a table that included the sequence accession code, its description, author's, date of publication of the sequence, and bibliographical information (including link to PubMed, when available). When the protein sequence was unavailable, the larger open reading frame was translated. These data were imported into a hyperlinked Excel spreadsheet available as **Supplemental Spreadsheet 1**.

Since this sequence set contained both those related to housekeeping, as well as a salivary secreted function (from now on just called as “housekeeping” or “secreted,”) we ran these

TABLE 1 | Tick salivary coding sequences retrieved from NCBI on November 2019.

A		B	
Species	Number of Sequences	Species	Number of Sequences
<i>Amblyomma americanum</i>	3,967	<i>Rhipicephalus appendiculatus</i>	49,146
<i>Amblyomma aureolatum</i>	93	<i>Rhipicephalus bursa</i>	39,435
<i>Amblyomma cajennense</i>	7,098	<i>Hyalomma dromedarii</i>	38,518
<i>Amblyomma maculatum</i>	4,855	<i>Ixodes ricinus</i>	33,250
<i>Amblyomma parvum</i>	2,838	<i>Rhipicephalus annulatus</i>	20,753
<i>Amblyomma rotundatum</i>	355	<i>Haemaphysalis longicornis</i>	20,592
<i>Amblyomma sculptum</i>	4,249	<i>Rhipicephalus zambeziensis</i>	14,393
<i>Amblyomma triste</i>	8,098	<i>Rhipicephalus pulchellus</i>	11,227
<i>Amblyomma tuberculatum</i>	67	<i>Amblyomma triste</i>	8,098
<i>Amblyomma variegatum</i>	743	<i>Ornithodoros turicata</i>	7,560
<i>Antricola delacruzi</i>	317	<i>Amblyomma cajennense</i>	7,098
<i>Argas monolakensis</i>	1,059	<i>Ornithodoros rostratus</i>	6,587
<i>Carios mimon</i>	4	<i>Hyalomma excavatum</i>	5,338
<i>Dermacentor andersoni</i>	505	<i>Amblyomma maculatum</i>	4,855
<i>Haemaphysalis longicornis</i>	20,592	<i>Amblyomma sculptum</i>	4,249
<i>Haemaphysalis qinghaiensis</i>	2	<i>Amblyomma americanum</i>	3,967
<i>Hyalomma anatolicum anatolicum</i>	88	<i>Amblyomma parvum</i>	2,838
<i>Hyalomma asiaticum asiaticum</i>	1	<i>Ixodes scapularis</i>	2,451
<i>Hyalomma dromedarii</i>	38,518	<i>Argas monolakensis</i>	1,059
<i>Hyalomma excavatum</i>	5,338	<i>Rhipicephalus microplus</i>	969
<i>Hyalomma rufipes</i>	558	<i>Rhipicephalus sanguineus</i>	761
<i>Ixodes affinis</i>	1	<i>Amblyomma variegatum</i>	743
<i>Ixodes holocyclus</i>	7	<i>Hyalomma rufipes</i>	558
<i>Ixodes pacificus</i>	118	<i>Ornithodoros parkeri</i>	544
<i>Ixodes persulcatus</i>	17	<i>Dermacentor andersoni</i>	505
<i>Ixodes ricinus</i>	33,250	<i>Ornithodoros brasiliensis</i>	440
<i>Ixodes scapularis</i>	2,451	<i>Amblyomma rotundatum</i>	355
<i>Ixodes sinensis</i>	5	<i>Antricola delacruzi</i>	317
<i>Ornithodoros brasiliensis</i>	440	<i>Ornithodoros coriaceus</i>	272
<i>Ornithodoros coriaceus</i>	272	<i>Ixodes pacificus</i>	118
<i>Ornithodoros kalahariensis</i>	2	<i>Amblyomma aureolatum</i>	93
<i>Ornithodoros moubata</i>	59	<i>Hyalomma anatolicum anatolicum</i>	88
<i>Ornithodoros parkeri</i>	544	<i>Amblyomma tuberculatum</i>	67
<i>Ornithodoros rostratus</i>	6,587	<i>Ornithodoros moubata</i>	59
<i>Ornithodoros turicata</i>	7,560	<i>Ixodes persulcatus</i>	17
<i>Rhipicephalus annulatus</i>	20,753	<i>Ixodes holocyclus</i>	7
<i>Rhipicephalus appendiculatus</i>	49,146	<i>Ixodes sinensis</i>	5
<i>Rhipicephalus bursa</i>	39,435	<i>Carios mimon</i>	4
<i>Rhipicephalus haemaphysaloides</i>	1	<i>Haemaphysalis qinghaiensis</i>	2
<i>Rhipicephalus microplus</i>	969	<i>Ornithodoros kalahariensis</i>	2
<i>Rhipicephalus pulchellus</i>	11,227	<i>Hyalomma asiaticum asiaticum</i>	1
<i>Rhipicephalus sanguineus</i>	761	<i>Ixodes affinis</i>	1
<i>Rhipicephalus zambeziensis</i>	14,393	<i>Rhipicephalus haemaphysaloides</i>	1
Total	287,343	Total	287,343

(A) Data sorted by species and (B) data sorted by number of sequences.

sequences through the SignalP program (version 3.0) (Nielsen et al., 1999) to identify whether they had a predicted signal sequence indicative of secretion. We also ran them as well as through the TMHMM program (Sonnhammer et al., 1998) to determine their predicted transmembrane domains. Of the 287,343 sequences, 45,264 had a signal sequence, excluding those having one or more transmembrane domain outside the signal sequence; thus obtaining a set coding for putative classical secreted proteins (Table 2). To identify protein families, we clusterized these sequences by blasting each one against the remaining set and joining those that reached variable degrees of identity (varying from 25 to 95% in 5% increments), over at least 75% of the length of the larger sequence. Notice that if a polypeptide A satisfies the rule with polypeptide B, and B with polypeptide C, the cluster ABC would be formed, even though polypeptide A does not satisfy the rule with polypeptide C. For each cluster, we obtained a fasta file with its members. In the cases where the cluster had 5 or more sequences, a clustal (Larkin et al., 2007) alignment was obtained. The resulting fasta files and clustal alignments were hyperlinked to the spreadsheet with the indication of the cluster number (ordered from the most to the least abundant in sequences) and its number of sequences (Supplemental Spreadsheet 2).

From the original set of 287,343 sequences, we extracted the 45,264 sequences classified as secreted. The remaining sequences were submitted to the secretomeP program to identify those sequences that could qualify as secreted through non-classical pathways, as indicated by a secretomeP score larger than 0.6 (Bendtsen et al., 2004). The resulting 125,197 sequences are displayed on Supplemental Spreadsheet 3.

TickSialoFam Model Construction

A program was written in VB6 that inspected each alignment file from the various clusters of the “Secreted” data set (Supplemental Spreadsheet 2). In the cases where it had five or more sequences, and if the alignment indicated at least five sites of identity or conservation, then the fasta file was used to construct a PSIBlast-based model (Altschul et al., 1997; Schaffer et al., 2001). First, all sequences of the cluster were blasted against themselves, and the one that accrued the largest sum of scores (excluding self-blast) was elected as the cluster centroid. The program blastpgp, from the blast package (Psi-blast), was then run using the centroid sequence as input; the formatted cluster fasta sequences (using the blast suit program formatdb) was used as a target databank, with the parameters -W 2 (blastp word size), and -j 5 (maximum number of iterations). The -h switch (inclusion e value after the first blast step) was variable, according to the size of the centroid: the chosen values were 1e-15 for length (l) larger than 200; 1e-8 for 100 < l < 200; 1e-6 for 60 < l < 100; and 1e-4 for l < 60. These values were arbitrarily adjusted according to size of the sequence: because a smaller sequence accrues a smaller score value than a large sequence and thus has a higher e value. The resulting matrix file was saved using the -C switch. These models were named according to the cluster parameters; for e.g., a model named 35-123 derived from the sequences found in cluster number 123 from the clusterization at 35% identity.

TABLE 2 | Putative secreted class of tick salivary sequences retrieved from NCBI on November 2019.

A		B	
Species	Number of Sequences	Species	Number of Sequences
<i>Amblyomma americanum</i>	1,369	<i>Ixodes ricinus</i>	9,776
<i>Amblyomma aureolatum</i>	27	<i>Rhipicephalus appendiculatus</i>	6,690
<i>Amblyomma cajennense</i>	2,072	<i>Hyalomma dromedarii</i>	3,833
<i>Amblyomma maculatum</i>	917	<i>Rhipicephalus bursa</i>	3,190
<i>Amblyomma parvum</i>	721	<i>Rhipicephalus zambeziensis</i>	3,017
<i>Amblyomma rotundatum</i>	164	<i>Amblyomma triste</i>	2,542
<i>Amblyomma sculptum</i>	307	<i>Haemaphysalis longicornis</i>	2,493
<i>Amblyomma triste</i>	2,542	<i>Amblyomma cajennense</i>	2,072
<i>Amblyomma tuberculatum</i>	26	<i>Rhipicephalus pulchellus</i>	1,969
<i>Amblyomma variegatum</i>	267	<i>Rhipicephalus annulatus</i>	1,467
<i>Antricola delacruzi</i>	83	<i>Amblyomma americanum</i>	1,369
<i>Argas monolakensis</i>	301	<i>Amblyomma maculatum</i>	917
<i>Dermacentor andersoni</i>	154	<i>Ixodes scapularis</i>	890
<i>Haemaphysalis longicornis</i>	2,493	<i>Ornithodoros turicata</i>	818
<i>Hyalomma anaticum</i>	27	<i>Amblyomma parvum</i>	721
<i>Hyalomma asiaticum asiaticum</i>	1	<i>Ornithodoros rostratus</i>	657
<i>Hyalomma dromedarii</i>	3,833	<i>Hyalomma excavatum</i>	397
<i>Hyalomma excavatum</i>	397	<i>Amblyomma sculptum</i>	307
<i>Hyalomma rufipes</i>	133	<i>Argas monolakensis</i>	301
<i>Ixodes affinis</i>	1	<i>Amblyomma variegatum</i>	267
<i>Ixodes holocyclus</i>	7	<i>Rhipicephalus microplus</i>	257
<i>Ixodes pacificus</i>	74	<i>Rhipicephalus sanguineus</i>	224
<i>Ixodes persulcatus</i>	14	<i>Amblyomma rotundatum</i>	164
<i>Ixodes ricinus</i>	9,776	<i>Ornithodoros parkeri</i>	162
<i>Ixodes scapularis</i>	890	<i>Dermacentor andersoni</i>	154
<i>Ixodes sinensis</i>	5	<i>Hyalomma rufipes</i>	133
<i>Ornithodoros brasiliensis</i>	61	<i>Ornithodoros coriaceus</i>	111
<i>Ornithodoros coriaceus</i>	111	<i>Antricola delacruzi</i>	83
<i>Ornithodoros kalahariensis</i>	2	<i>Ixodes pacificus</i>	74
<i>Ornithodoros moubata</i>	37	<i>Ornithodoros brasiliensis</i>	61
<i>Ornithodoros parkeri</i>	162	<i>Ornithodoros moubata</i>	37
<i>Ornithodoros rostratus</i>	657	<i>Amblyomma aureolatum</i>	27
<i>Ornithodoros turicata</i>	818	<i>Hyalomma anaticum</i>	27
<i>Rhipicephalus annulatus</i>	1,467	<i>Amblyomma tuberculatum</i>	26
<i>Rhipicephalus appendiculatus</i>	6,690	<i>Ixodes persulcatus</i>	14
<i>Rhipicephalus bursa</i>	3,190	<i>Ixodes holocyclus</i>	7
<i>Rhipicephalus haemaphysaloides</i>	1	<i>Ixodes sinensis</i>	5
<i>Rhipicephalus microplus</i>	257	<i>Ornithodoros kalahariensis</i>	2
<i>Rhipicephalus pulchellus</i>	1,969	<i>Hyalomma asiaticum asiaticum</i>	1
<i>Rhipicephalus sanguineus</i>	224	<i>Ixodes affinis</i>	1
<i>Rhipicephalus zambeziensis</i>	3,017	<i>Rhipicephalus haemaphysaloides</i>	1
Total	45,264	Total	45,264

(A) Data sorted by species and (B) data sorted by number of sequences.

These procedures are similar to those used for construction of the COG and KOG databases (Tatusov et al., 2003).

Another method to create matrices was done by psiblasting each sequence against the whole Ixodida database, using the same parameters indicated above. These models are named by the NCBI accession number of the parent protein sequence. These matrix files were then combined with the program makeprofiledb with the parameters -threshold 9.82 -scale 100.0 -dbtype rps -index true, forming the initial TSFam database. Finally, the number of models used in the final database was reduced by selecting those that produced a match with a maximum *e*-value of 1e-20.

Sequence Functional Classification

To further help classification of the protein sequences, possible proline-hydroxylation (Rhoads and Udenfriend, 1969; Bohmer, 1971; Kivirikko et al., 1972) and tyrosine sulfation (Nicholas et al., 1999) sites were determined, including the percentages of glycines, prolines, and tyrosines of each protein sequence, and their number of cysteine residues. These features helped to classify the families within the glycine-rich proteins (GRP) group. Determination of the presence of glycosyl-phosphate-inositol (GPI) anchors were determined with the program DGPI (Pierleoni et al., 2008). This helped to identify secreted proteins that were bound to membranes. The presence of furin cleavage sites were also detected (Duckert et al., 2004), identifying polyproteins with various cleavage sites. The percentages of serines and threonines were determined. The number of possible mucin-type galactosylation sites being measured with the program NetOGlyc (Hansen et al., 1998; Julenius et al., 2005) helps to further identify and classify the mucins. The set of secreted sequences were compared by blastp; to a subset of proteins from the NCBI NR database; to the Swissprot database; to the Enzyme Commission set from the KEGG database; to the salivary Ixodida sequences retrieved as indicated above; to the top 100 proteins associated with exosomes from exocarta (<http://exocarta.org/download>); and to rpsblast comparisons to the Conserved Domains database (CDD), Kog, Pfam, Smart, and TSF databases. All these comparisons were hyperlinked to the corresponding spreadsheets and allowed manual annotation of the sequences.

The annotations were transferred to the TSF models, which included in their description the following fields of information: (1) model name: as indicated in the previous section (e.g., 35-123 or JAP81818.1). (2) Group: This classification may include non-phylogenetically related sequences, such as the glycine-rich proteins (GRP) and mucins but can include groups such as the Lipocalins or Serpins. For this reason, this field is named group and not family. (3) Family: A subdivision of the group, containing phylogenetically related sequences. This field is empty in many models. (4) H/S/I: Indicates whether the sequence is classified as Secreted, Housekeeping, or in a few cases, Indeterminate. (5) Other: Includes additional information codified as E, enzyme; AM, antimicrobial; and PI, protease inhibitor. These fields are separated by “|” characters, which may be changed to tabs to produce the display shown in the hyperlinked spreadsheets.

The publicly available formatted TSFam database (**Supplemental File 1**) can be used in a rpsblast search using as an input a fasta file of tick salivary proteins, using a command line such as: “[x]blastall -p rpsblast -d [y]TSF -i [z]in.fasta -o [z]tsf.blc -IT -JT -v20 -b10 -e1e-4 -FF.” Where [x], [y], and [z] are the paths to the blast program, the TSFam database, and the input fasta and the output files, respectively. We recommend accepting the classifications using the TSFam database when the coverage of the model by the queried protein is at least 2/3 (>66.6%) and having a maximum *e* value of 1e-4. To access the blast programs, use version 2.2.26 available at <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/legacy.NOTSUPPORTED/2.2.26/>.

Tools That Help With Mining the Database

Three tools, written in visual basic and compiled to run under the Windows environment, are available upon email request to jribeiro@niaid.nih.gov. The tool, *fasta2tbl*, reads a fasta file and creates a hyperlinked table that can be imported into a spreadsheet. The tool, *rpsblast2tbl*, reads a rpsblast result file and maps the results and annotation into a hyperlinked table that can be imported into the fasta table done with *fasta2tbl*. The tool, *GetFSA*, takes a list of sequence names found in the fasta table or spreadsheet and creates a fasta file that can be used for phylogenetic or other studies.

Phylogenetic Analysis

Phylogenetic analysis were done with Mega7 (Kumar et al., 2016). Alignments were done with Clustal (Larkin et al., 2007) and refined with Muscle (Edgar, 2004) using the MEGA7 program (Kumar et al., 2016). The program FUBAR (Murrell et al., 2013) of the HYPHY (Pond et al., 2005) package was used to find positively selected sites within the in-frame nucleotide sequences. Positively selected codons were identified by their assigned probabilities of the parameter α being smaller than beta. The program RDP4 was used to determine the number of recombination breakpoints within the nucleotide alignments (Martin and Rybicki, 2000).

RESULTS

Table 1 lists the number of retrieved sequences by species, sorted by species name (A), and number of sequences (B); **Table 2** lists the same information for the subset of sequences classified as “secreted.”

Because the sequences under this study derive mainly from transcriptome assembly studies, many of the sequences are truncated either at their 5′ or 3′ ends. Only 57% of the 45,264 protein sequences start with a methionine (and even if it starts with one, there is no guarantee that it is the starting methionine); stop codons are found for 57% of the sequences. Both a starting methionine and stop codon occur in 49% of the sequences. Accordingly, identification of signal peptide indicative of secretion can only be found with confidence in a maximum of 57% of the set. Despite these flaws, truncated protein sequences are still important, as they may later serve as sequence templates for MS/MS studies that may include extension of selected coding sequences.

Putative Secreted Proteins

The TSFam database of motifs allowed classification of 18,963 sequences from the 45,264 salivary-derived Ixodida sequences from NCBI, classified as putative secreted proteins (**Supplemental Spreadsheet 2**). These were organized into 136 groups, as indicated in the **Supplemental Table 1**. The criteria for classification of these sequences required an *e*-value <1e-4 and a model coverage of at least 67% (2/3 of the model length). The classifications can be further grouped according to their similarities, or lack of similarities, to other known proteins or protein domains; they can also be grouped according to their known function, leading to the categories “Enzymes,” “Protease inhibitors,” “Ubiquitous families,” “Tick-specific anti-clotting peptides,” “Immune-related,” “Antimicrobials,” and “Tick-specific families” (**Figure 2**). Notice that the category “Tick-specific families” is the larger one, however, most of its members have no known function. Many of the groups are complex in nature and may contain phylogenetically unrelated families; for instance, the “Glycine-rich family,” is further subdivided into 19 families in **Supplemental Spreadsheet 2**. Six of the groups contains the majority of sequences (**Figure 3**).

Considering all 136 groups of **Supplemental Table 1**, only 37 have been characterized with at least one member having been recombinantly expressed and its function determined by bioassay, enzymatic assay, or proteomic detection from tick saliva. These studies are referenced in **Supplemental Table 1**. From these 37 characterized families, four have been only enzymatically characterized. Accordingly, the responsible tick protein is only presumed; nine have been found by saliva proteomic studies, but their function has not been characterized. It is also possible that many of the deduced CDS derive from pseudogenes (Kalyana-Sundaram et al., 2012). Thus, we have relative confidence that we know the molecular nature and function of at least one member from 24 of the 136 groups. Confidence should be defined in this case as: (1) Ascertaining that the protein is expressed in the saliva; (2) determining the biological activity of the recombinant protein. Very few protein families have at least one member satisfying these two requirements. Accordingly, this database consists mostly of a hypothetical set to be further characterized by proteomic and functional studies and should be updated as new studies are published.

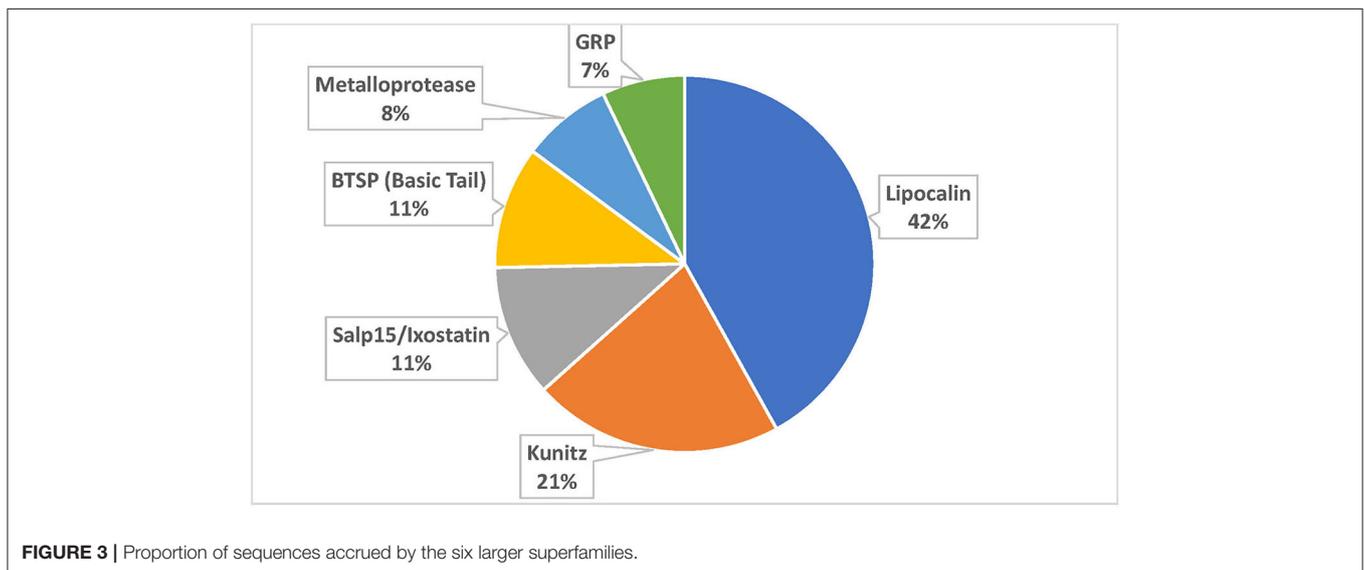
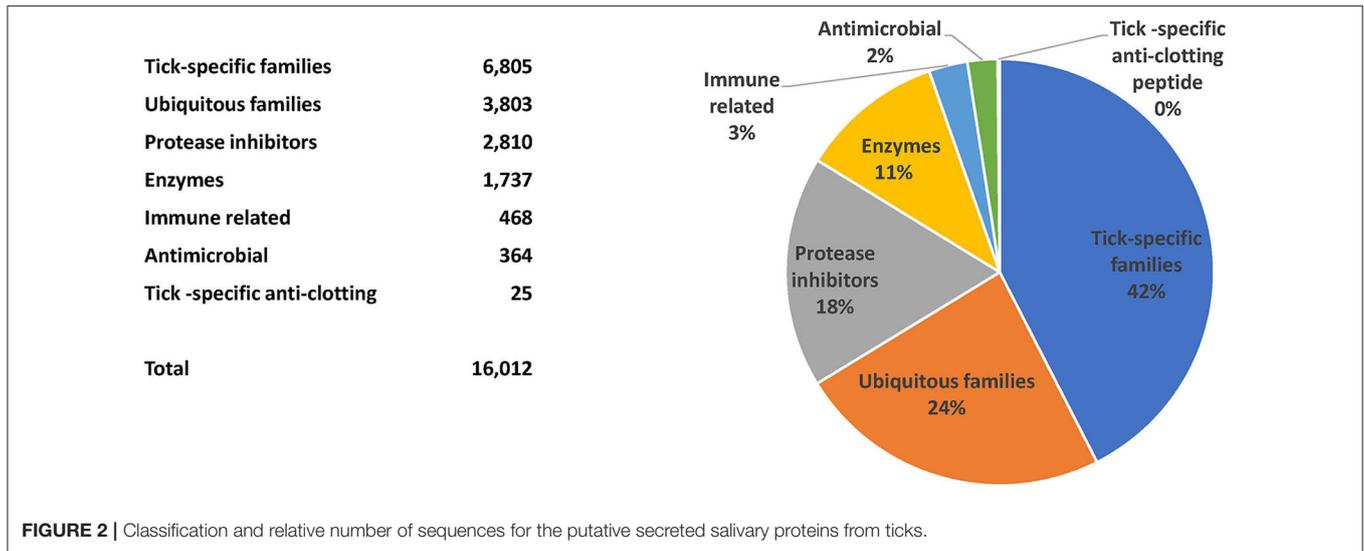
This new data set provides the mining rounds for further analysis into the evolution of previously known families, including characterization of novel protein families by recombinant expression and functional assay of the predicted proteins.

Evolutionary Insights on Tick Salivary Proteins

What follows is the phylogenetical analysis of selected protein families. Some large families have complex composition (such as, Kunitz, basic tail, Salp15) and will not be analyzed.

Apyrase/5′-Nucleotidase Family

As stated in the introduction, the saliva of unrelated blood feeding arthropods has the ability to hydrolyze ADP and



ATP, thus inhibiting important pathways of platelet and neutrophil aggregation and other pro-inflammatory reactions (Francischetti et al., 2009). The **Supplemental Spreadsheet 2** and **Table 1** provide for 50 full length sequences belonging to the 5'-nucleotidase family; four of which have a GPI anchor, indicating they are not secreted and could represent the homolog ancestral genes that gave origin to gene versions coding for secreted products. **Supplemental Spreadsheet 1** indicate the presence of 287 annotations for 5'-nucleotidase; fifty of which are full-length. Phylogenetic analysis of these 50 sequences (**Supplemental Figure 1**) confirms this scenario: The GPI-containing sequences group together in a single clade, while the remaining sequences group into three major clades, Clades I and II, with Metastriate and Prostriate sequences, and an Argasidae clade. Clade I has two main branches, one containing Prostriate sequences and another containing only Metastriate sequences. The Prostriate sub clade further divides into two branches, indicative of an ancient gene duplication event. Clade

II is similarly organized, with two main branches, and containing a single Prostriate sequence. It is also relevant that the depth of amino acid divergence on the clades I and II are at least 2-fold of the depth of the classic GPI clade, which contains one sequence from *Ixodes*, two from *Rhipicephalus*, and one from *Amblyomma*. Notice that the analysis has many more sequences derived from *I. ricinus* than from other species, reflecting the larger read coverage for the *Ixodes* transcriptomes. As coverage is increased for other species, novel genes coding for products of the 5'-nucleotidase/Apyrase should appear. The phylogram indicates the existence of at least seven genes coding for secreted apyrases from *I. ricinus*, based on the branches having 100% bootstrap support; the amino acid divergence of the secreted proteins is larger than that of the housekeeping, GPI anchored sequences, indicating faster evolution of the genes coding for secreted salivary genes; this was either caused by neutral evolution of these genes when compared to a purifying evolution scenario of the housekeeping genes, or due to positive

selection pressure on the genes coding for secreted enzymes, driven by the immune pressure of their hosts (Mans et al., 2017). Submission of the apyrase coding enzymes to the FUBAR analysis (Murrell et al., 2013) indicated no sites under positive selection. RDP analysis (Martin and Rybicki, 2000) using the same sequence alignment indicated three recombination breakpoints (**Supplemental Spreadsheet 4**).

Lipocalins

There are 3,689 sequences classified as lipocalins in **Supplemental Spreadsheet 2** (considering >66% coverage with a TSFam lipocalin match). A subset of these sequences that had the PFAM His bind domain (with at least 80% coverage) amounted to 1,050 sequences. Attempts to construct a phylogeny (based on the faster Neighbor-Joining algorithm) with this subset were not informative, with most clades having poor bootstrap support. The phylogeny reconstruction of a smaller subset of 32 *I. ricinus* sequences, that were clustered at 40% sequence identity (**Supplemental Figure 2**), shows two main clades that further subdivide in eight smaller clades containing 2 or more sequences. FUBAR analysis indicated 16 codons under positive selection. RDP analysis indicated 11 recombination breakpoints (**Supplemental Spreadsheet 4**).

Metalloproteases

The TSF algorithm indicates 677 sequences as being metalloproteases in **Supplemental Spreadsheet 2**. Eighty-seven of these sequences, from *I. ricinus*, *A. americanum*, and *R. appendiculatus* were chosen for alignment and phylogenetic analysis based on their completeness and match to the CDD domain cd04272 named ZnMc_salivary_gland_MPs. The cladogram displays five Metastriate-specific clades with strong bootstrap support, plus five Prostriate ones; each clade further subdividing in two or more subclades (**Supplemental Figure 3**). We can conclude that in *I. ricinus*, and possibly in other ticks, there are at least 10 genes coding for this subdivision of the salivary metalloproteases. FUBAR analysis found no sites under positive selection. RDP indicated 3 recombination breakpoints (**Supplemental Spreadsheet 4**).

DAP36 Family

The founding member of this family was identified as an immunosuppressant salivary protein from *Dermacentor andersoni* (Bergman et al., 1995, 1998). In 2009 this family was characterized as a diverse Metastriate-specific family. However, here we find it is also composed of *Ixodes ricinus* members, as depicted in the phylogram of **Supplemental Figure 4**, built with a total of 121 sequences showing eight clades with strong bootstrap support, each containing subclades. These clades are composed of sequences originating from the same genus. FUBAR indicated strong positive selection for seven sites RDP analysis indicated six recombination breakpoints (**Supplemental Spreadsheet 4**).

Cystatins

Cystatins are peptide inhibitors of cysteine proteases that are widespread in vertebrates (Abrahamson et al., 2003). They were first found in tick saliva by screening a salivary transcriptome

of *I. scapularis* (Ribeiro et al., 2006). Different from lipocalins, metalloproteases, and Kunitz products, salivary cystatins are poorly transcribed. A recombinant peptide indicated its specificity for cathepsin L, and the protein was named sialostatin L. The expected activity was found in the tick saliva and expression of the protein was confirmed by reactive antibodies raised against the recombinant protein (Kotsyfakis et al., 2006). Naturally infested guinea pigs do not develop antibodies against sialostatin L, but vaccination of guinea pigs with sialostatin L reduced the feeding of ticks, and a boost reaction was observed (Kotsyfakis et al., 2008). A second *I. scapularis* salivary cystatin, named sialostatin L2, has been characterized (Chen et al., 2014; Lieskovska et al., 2015a). The effects of sialostatin L and L2 in mammalian immunity has been well-studied (Sa-Nunes et al., 2009; Kotsyfakis et al., 2010; Horka et al., 2012; Schwarz et al., 2012; Bruhl et al., 2014; Chen et al., 2014; Klein et al., 2015; Lieskovska et al., 2015a,b; Wang et al., 2016; Kotal et al., 2019). **Supplemental Figure 5** depicts the phylogram of 82 tick cystatins, which indicates at least 17 clear clades each being mostly genus specific. FUBAR analysis does not recognize any codon site under positive selection, but RDP indicates the existence of one recombination breakpoint (**Supplemental Spreadsheet 4**).

Proteins Possibly Secreted by Alternative Pathways

Supplemental Spreadsheet 3 contains proteins without a classical secretion signal but containing a secretomeP score indicative of secretion via non-classical pathways. Among these there are the subolesins, described in the introduction. Phylogenetic analysis, based on the alignments of 16 coding sequences, shows three clades of FUBAR indicates no sites under positive selection. RDP analysis indicates one recombination breakpoint.

Housekeeping Proteins

Annotation of the SignalP-containing sequences in **Supplemental Spreadsheet 2** by the TSA algorithm identifies 1,931 coding sequences, organized into 80 groups (**Supplemental Table 2** and **Supplemental Spreadsheet 2**) that probably have a housekeeping function associated with the ER or Golgi compartments. Notably, 1,409 sequences representing 72% of the total group of housekeeping sequences are enzymes.

TABLE 3 | Recombination breakpoints and positive selected sites predicted by RDP4 (Martin et al., 2015) and FUBAR (Murrell et al., 2013).

Family	Number of recombination breakpoints	Number of positive selected sites
Lipocalin	11	16
DAP36	6	7
Apyrase	3	1
Metalloprotease	3	0
Cystatin	1	0
Subolesin	1	0

Among these enzymes, 187 code for glycosyltransferases, most probably associated with glycosylating the salivary secretome. The glycosylation of tick salivary proteins has been under scrutiny recently, due to the epidemics of α -Gal allergy triggered by consumption of beef proteins (Commins and Platts-Mills, 2013). Human cases are associated with previous exposure to tick bites, and spatially correlated with tick abundance (Commins et al., 2011; Mateo-Borrega et al., 2019). A recent search for enzymes coding for α -Gal transferases (GALT) in ticks (Cabezas-Cruz et al., 2018) showed that the classical α 1-3 GALTs that produce the Gal α 1-3Gal β 1-(3)4GlcNAc-R (α -Gal) are absent in ticks; their function is taken by other transferases of the α 1-4 and β 1-4 GALT families. These enzymes are recognizable in **Supplemental Spreadsheet 2** by the TSFam group of Glycosyltransferases and subfamily Lactosylceramide 4-alpha-galactosyltransferase. Other previously identified enzymes that are associated with protein modification (Ribeiro et al., 2006; Francischetti et al., 2009) include those promoting proline hydroxylation and tyrosine sulfation. While tyrosine sulfation has been determined in tick salivary anticoagulant peptides (Thompson et al., 2017; Watson et al., 2019), proline hydroxylation, a common modification in collagen (Bohmer, 1971), has not been confirmed in ticks despite the high abundance of salivary collagen-like cement proteins that display the motifs triggering proline hydroxylation (Rhoads and Udenfriend, 1969; de Jong et al., 1991; Shimizu et al., 2005). Proteomic studies should include these possible modifications in the databases used for identification of peptide fragments obtained by MS/MS.

DISCUSSION

The TSFam database, and its associated hyperlinked spreadsheets of tick salivary proteins and associated nucleotide sequences, helps the annotation of tick salivary secreted proteins, as well as the retrieval of these sequences for phylogenetic studies. It has been reported that the genes coding for salivary proteins of blood sucking arthropods show high evolutionary rates; either due to the relaxed constraint of the genes allowing for less negative selection pressure, or for the more rare occurrence of positive selection (Daix et al., 2007; Schroeder et al., 2007; Decrem et al., 2008; Dai et al., 2012; Arca et al., 2014; Mans et al., 2017). The phylogenetic analysis, derived from the alignment of tick salivary coding sequences from the lipocalin, metalloprotease, cystatin, apyrase, DAP36, and subolesin families (**Supplemental Figures 1–6** and **Table 3**), indicate a high number of recombination breakpoints on the genes coding for lipocalins (11 sites) and DAP36 (6 sites); three recombination breakpoint sites for the apyrase and metalloproteases; and one recombination breakpoint on the genes coding for the cystatins and subolesin. No positive selection sites were identified for the metalloprotease, cystatin, and subolesin genes. It is to be noted that the occurrence of recombination breakpoints could lead to false predictions of positive selected sites (Posada and Crandall, 2002; Bay and Bielawski, 2011). While positive selection is associated with fast

gene evolution and divergence, intragenic recombination can strongly contribute to genome diversity among individuals, a situation that might be occurring within the lipocalin and DAP36 coding genes (Lauer et al., 2018; Olabode et al., 2019; Salim et al., 2019). As a note of caution, it should be considered that the diversity of transcripts from these multi-gene families could be artifactual due to the “*de novo*” assembly from small reads which could create false recombinant assemblies (Salmon et al., 2010). A possible way to verify the frequency of these artifacts would be to compare the resulting lipocalin transcripts derived from single tick transcriptomes to those derived from the combined libraries. On the other hand, if recombination exists and leads to new protein variants, is it a classical meiotic recombination derived from a high number of intra-genic breakpoints, or does it occur in a somatic environment, such as occurs with vertebrate immunoglobulins. This question could be answered by attempting the genome mapping of each transcript and identifying whether the transcript derives from a mosaic of exons or a “well-behaved” sequence of exons.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JR conceived the work and did the bioinformatic analysis. BM contributed in the data analysis and interpretation. All authors contributed to writing the manuscript.

FUNDING

JR was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (Vector-Borne Diseases: Biology of Vector Host Relationship, Z01 AI000810-19). BM was supported by the National Research Foundation of South Africa (Grant No. 118540).

ACKNOWLEDGMENTS

This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). We are grateful to helpful comments on the manuscript by Drs. Lucas Tirloni and John Andersen. We would like to thank Brian Brown, NIH Library Editing Service, for reviewing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2020.00374/full#supplementary-material>

REFERENCES

- Abrahamson, M., Alvarez-Fernandez, M., and Nathanson, C. M. (2003). Cystatins. *Biochem. Soc. Symp.* 70, 179–199. doi: 10.1042/bss0700179
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Alvarenga, P. H., Francischetti, I. M., Calvo, E., Sa-Nunes, A., Ribeiro, J. M., and Andersen, J. F. (2010). The function and three-dimensional structure of a thromboxane A₂/cysteinyl leukotriene-binding protein from the saliva of a mosquito vector of the malaria parasite. *PLoS Biol.* 8:e1000547. doi: 10.1371/journal.pbio.1000547
- Andersen, J. F., Hinnebusch, B. J., Lucas, D. A., Conrads, T. P., Veenstra, T. D., Pham, V. M., et al. (2007). An insight into the sialome of the oriental rat flea, *Xenopsylla cheopis* (rosts). *BMC Genomics* 8:102. doi: 10.1186/1471-2164-8-102
- Andersen, J. F., and Ribeiro, J. M. (2017). “Chapter 4 - Salivary kratagonists: scavengers of host physiological effectors during blood feeding,” in *Arthropod Vector: Controller of Disease Transmission* (London: Elsevier), 51–63. doi: 10.1016/B978-0-12-805360-7.00004-6
- Arca, B., Lombardo, F., Struchiner, C. J., and Ribeiro, J. M. (2017). Anopheline salivary protein genes and gene families: an evolutionary overview after the whole genome sequence of sixteen *Anopheles* species. *BMC Genomics* 18:153. doi: 10.1186/s12864-017-3579-8
- Arca, B., Struchiner, C. J., Pham, V. M., Sferra, G., Lombardo, F., Pombi, M., et al. (2014). Positive selection drives accelerated evolution of mosquito salivary genes associated with blood-feeding. *Insect Mol. Biol.* 23, 122–131. doi: 10.1111/imb.12068
- Bay, R. A., and Bielawski, J. P. (2011). Recombination detection under evolutionary scenarios relevant to functional divergence. *J. Mol. Evol.* 73, 273–286. doi: 10.1007/s00239-011-9473-0
- Becker, M., Felsberger, A., Frenzel, A., Shattuck, W. M. C., Dyer, M., Kugler, J., et al. (2015). Application of M13 phage display for identifying immunogenic proteins from tick (*Ixodes scapularis*) saliva. *BMC Biotechnol.* 15:43. doi: 10.1186/s12896-015-0167-3
- Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* 17, 349–356. doi: 10.1093/protein/gzh037
- Bergman, D. K., Ramachandra, R. N., and Wikel, S. K. (1995). *Dermacentor andersoni*: salivary gland proteins suppressing T-lymphocyte responses to concanavalin A *in vitro*. *Exp. Parasitol.* 81, 262–271. doi: 10.1006/expr.1995.1117
- Bergman, D. K., Ramachandra, R. N., and Wikel, S. K. (1998). Characterization of an immunosuppressant protein from *Dermacentor andersoni* (Acari: Ixodidae) salivary glands. *J. Med. Entomol.* 35, 505–509. doi: 10.1093/jmedent/35.4.505
- Binnington, K. C., and Kemp, D. H. (1980). Role of tick salivary glands in feeding and disease transmission. *Adv. Parasitol.* 18, 316–340. doi: 10.1016/S0065-308X(08)60403-0
- Bohmer, T. (1971). The biosynthesis of collagen. *Acta Rheumatol. Scand.* 17, 209–222. doi: 10.3109/rhe1.1971.17.issue-1-4.28
- Bowman, A. S., and Sauer, J. R. (2004). Tick salivary glands: function, physiology and future. *Parasitology* 129 (Suppl), S67–81. doi: 10.1017/S0031182004006468
- Bruhl, T. J., Klein, M., Staudt, V., Reuter, S., Grebe, N., Gerlitzki, B., et al. (2014). A new target for the treatment of mast cell dependent asthma? *Respirology* 19:14. doi: 10.1111/resp.12257
- Cabezas-Cruz, A., Espinosa, P. J., Alberdi, P., Simo, L., Valdes, J. J., Mateos-Hernandez, L., et al. (2018). Tick galactosyltransferases are involved in alpha-Gal synthesis and play a role during *Anaplasma phagocytophilum* infection and *Ixodes scapularis* tick vector development. *Sci. Rep.* 8:14224. doi: 10.1038/s41598-018-32664-z
- Calvo, E., Mans, B. J., Ribeiro, J. M., and Andersen, J. F. (2009). Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3728–3733. doi: 10.1073/pnas.0813190106
- Champagne, D. E., Smartt, C. T., Ribeiro, J. M., and James, A. A. (1995). The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5'-nucleotidase family. *Proc. Natl. Acad. Sci. U.S.A.* 92, 694–698. doi: 10.1073/pnas.92.3.694
- Chavez, A. S. O., O'Neal, A. J., Santambrogio, L., Kotsyfakis, M., and Pedra, J. H. F. (2019). Message in a vesicle - trans-kingdom intercommunication at the vector-host interface. *J. Cell Sci.* 132:jcs224212. doi: 10.1242/jcs.224212
- Chen, G., Wang, X., Severo, M. S., Sakhon, O. S., Sohail, M., Brown, L. J., et al. (2014). The tick salivary protein sialostatin L2 inhibits caspase-1-mediated inflammation during *Anaplasma phagocytophilum* infection. *Infect. Immun.* 82, 2553–2564. doi: 10.1128/IAI.01679-14
- Commins, S. P., James, H. R., Kelly, L. A., Pochan, S. L., Workman, L. J., Perzanowski, M. S., et al. (2011). The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose-alpha-1,3-galactose. *J. Allerg. Clin. Immunol.* 127, 1286–1293 E1286. doi: 10.1016/j.jaci.2011.02.019
- Commins, S. P., and Platts-Mills, T. A. E. (2013). Tick bites and red meat allergy. *Curr. Opin. Allergy Clin. Immunol.* 13, 354–359. doi: 10.1097/ACI.0b013e3283624560
- Coons, L. B., and Roshdy, M. A. (1981). Ultrastructure of granule secretion in salivary glands of *Argas (Persicargas) arboreus* during feeding. *Parasitol. Res.* 65, 225–234. doi: 10.1007/BF00929188
- Dai, J., Wang, P., Adusumilli, S., Booth, C. J., Narasimhan, S., Anguita, J., et al. (2009). Antibodies against a tick protein, Salp15, protect mice from the Lyme disease agent. *Cell Host Microbe* 6, 482–492. doi: 10.1016/j.chom.2009.10.006
- Dai, S. X., Zhang, A. D., and 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. doi: 10.1186/1471-2148-12-4
- Daix, V., Schroeder, H., Praet, N., Georin, J. P., Chiappino, I., Gillet, L., et al. (2007). *Ixodes* ticks belonging to the *Ixodes ricinus* complex encode a family of anticomplement proteins. *Insect Mol. Biol.* 16, 155–166. doi: 10.1111/j.1365-2583.2006.00710.x
- Das, S., Banerjee, G., DePonte, K., Marcantonio, N., Kantor, F. S., and Fikrig, E. (2001). Salp25D, an *Ixodes scapularis* antioxidant, is 1 of 14 immunodominant antigens in engorged tick salivary glands. *J. Infect. Dis.* 184, 1056–1064. doi: 10.1086/323351
- de Jong, L., van der Kraan, I., and de Waal, A. (1991). The kinetics of the hydroxylation of procollagen by prolyl 4-hydroxylase. Proposal for a processive mechanism of binding of the dimeric hydroxylating enzyme in relation to the high kcat/Km ratio and a conformational requirement for hydroxylation of -X-Pro-Gly-sequences. *Biochim. Biophys. Acta* 1079, 103–111. doi: 10.1016/0167-4838(91)90030-4
- de la Fuente, J., Almazan, C., Blas-Machado, U., Naranjo, V., Mangold, A. J., Blouin, E. F., et al. (2006). The tick protective antigen, 4D8, is a conserved protein involved in modulation of tick blood ingestion and reproduction. *Vaccine* 24, 4082–4095. doi: 10.1016/j.vaccine.2006.02.046
- de la Fuente, J., Moreno-Cid, J. A., Canales, M., Villar, M., de la Lastra, J. M. P., Kocan, K. M., et al. (2011). Targeting arthropod subolesin/akirin for the development of a universal vaccine for control of vector infestations and pathogen transmission. *Vet. Parasitol.* 181, 17–22. doi: 10.1016/j.vetpar.2011.04.018
- Decrem, Y., Beaufays, J., Blasioli, V., Lahaye, K., Brossard, M., Vanhamme, L., et al. (2008). A family of putative metalloproteases in the salivary glands of the tick *Ixodes ricinus*. *FEBS J.* 275, 1485–1499. doi: 10.1111/j.1742-4658.2008.06308.x
- Duckert, P., Brunak, S., and Blom, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.* 17, 107–112. doi: 10.1093/protein/gzh013
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Farkas, R. (2015). Apocrine secretion: new insights into an old phenomenon. *BBA Gen. Subj.* 1850, 1740–1750. doi: 10.1016/j.bbagen.2015.05.003
- Francischetti, I. M. B., Sa-Nunes, A., Mans, B. J., Santos, I. M., and Ribeiro, J. M. C. (2009). The role of saliva in tick feeding. *Front. Biosci.* 14, 2051–2088. doi: 10.2741/3363
- Friesen, K. J., and Kaufman, W. R. (2009). Salivary gland degeneration and vitellogenesis in the ixodid tick *Amblyomma hebraeum*: surpassing a critical weight is the prerequisite and detachment from the host is the trigger. *J. Insect Physiol.* 55, 936–942. doi: 10.1016/j.jinsphys.2009.06.007
- Garcia, G. R., Maruyama, S. R., Nelson, K. T., Ribeiro, J. M. C., Gardinassi, L. G., Maia, A. A. M., et al. (2017). Immune recognition of salivary proteins from the cattle tick *Rhipicephalus microplus* differs according to the genotype of the bovine host. *Parasit. Vector* 10:144. doi: 10.1186/s13071-017-2077-9

- Guglielmo, A. A., Robbins, R. G., Apanaskevich, D. A., Petney, T. N., Estrada-Pena, A., Horak, I. G., et al. (2010). The *Argasidae*, *Ixodidae* and *Nuttalliellidae* (*Acari: Ixodida*) of the world: a list of valid species names. *Zootaxa* 10, 95–136. doi: 10.11646/zootaxa.2528.1.1
- Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., et al. (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* 7:10507. doi: 10.1038/ncomms10507
- Hackenberg, M., and Kotsyfakis, M. (2018). Exosome-mediated pathogen transmission by arthropod vectors. *Trends Parasitol.* 34, 549–552. doi: 10.1016/j.pt.2018.04.001
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* 15, 115–130. doi: 10.1023/A:1006960004440
- Harris, R. A., and Kaufman, W. R. (1981) Hormonal control of salivary gland degeneration in the ixodid tick *Amblyomma hebraeum*. *J. Insect Physiol.* 27, 241–248. doi: 10.1016/0022-1910(81)90057-3
- Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129–149. doi: 10.1126/science.1077061
- Horka, H., Staudt, V., Klein, M., Taube, C., Reuter, S., Dehzad, N., et al. (2012). Inhibits the Th9-derived production of the asthma-promoting cytokine IL-9 and is effective in the prevention of experimental asthma. *J. Immunol.* 188, 2669–2676. doi: 10.4049/jimmunol.1100529
- Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15, 153–164. doi: 10.1093/glycob/cwh151
- Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D. R., Wu, Y.-M., Cao, X., et al. (2012). Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* 149, 1622–1634. doi: 10.1016/j.cell.2012.04.041
- Karim, S., and Ribeiro, J. M. (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. doi: 10.1371/journal.pone.0131292
- Keller, S., Sanderson, M. P., Stoeck, A., and Altevogt, P. (2006). Exosomes: from biogenesis and secretion to biological function. *Immunol. Lett.* 107, 102–108. doi: 10.1016/j.imlet.2006.09.005
- Kivirikko, K. I., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1972). Hydroxylation of (X-Pro-Gly)_n by procollagen proline hydroxylase. Effect of chain length, helical conformation and amino acid sequence in the substrate. *Biochim. Biophys. Acta* 271, 347–356. doi: 10.1016/0005-2795(72)90209-7
- Klein, M., Bruhl, T. J., Staudt, V., Reuter, S., Grebe, N., Gerlitzki, B., et al. (2015). Tick salivary sialostatin L represses the initiation of immune responses by targeting IRF4-dependent transcription in murine mast cells. *J. Immunol.* 195, 621–631. doi: 10.4049/jimmunol.1401823
- Kotal, J., Stergiou, N., Busa, M., Chlastakova, A., Berankova, Z., Rezacova, P., et al. (2019). The structure and function of Iristatin, a novel immunosuppressive tick salivary cystatin. *Cell. Mol. Life Sci.* 76, 2003–2013. doi: 10.1007/s00018-019-03034-3
- Kotsyfakis, M., Anderson, J. M., Andersen, J. F., Calvo, E., Francischetti, I. M., Mather, T. N., et al. (2008). Cutting edge: Immunity against a “silent” salivary antigen of the Lyme vector *Ixodes scapularis* impairs its ability to feed. *J. Immunol.* 181, 5209–5212. doi: 10.4049/jimmunol.181.8.5209
- Kotsyfakis, M., Horka, H., Salat, J., and Andersen, J. F. (2010). The crystal structures of two salivary cystatins from the tick *Ixodes scapularis* and the effect of these inhibitors on the establishment of *Borrelia burgdorferi* infection in a murine model. *Mol. Microbiol.* 77, 456–470. doi: 10.1111/j.1365-2958.2010.07220.x
- Kotsyfakis, M., Sa-Nunes, A., Francischetti, I. M., Mather, T. N., Andersen, J. F., and Ribeiro, J. M. (2006). Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick *Ixodes scapularis*. *J. Biol. Chem.* 281, 26298–26307. doi: 10.1074/jbc.M513010200
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lauer, S., Avcilla, G., Spealman, P., Sethia, G., Brandt, N., Levy, S. F., et al. (2018). Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS Biol.* 16:e3000069. doi: 10.1371/journal.pbio.3000069
- Lewis, L. A., Radulovic, Z. M., Kim, T. K., Porter, L. M., and Mulenga, A. (2015). Identification of 24 h *Ixodes scapularis* immunogenic tick saliva proteins. *Ticks Tick Borne Dis.* 6, 424–434. doi: 10.1016/j.ttbdis.2015.03.012
- Lieskovska, J., Palenikova, J., Langhansova, H., Campos Chagas, A., Calvo, E., Kotsyfakis, M., et al. (2015a). Tick sialostatin L and L2 differentially influence dendritic cell responses to *Borrelia* spirochetes. *Parasit. Vectors* 8:275. doi: 10.1186/s13071-015-0887-1
- Lieskovska, J., Palenikova, J., Sirmarova, J., Elsterova, J., Kotsyfakis, M., Campos Chagas, A., et al. (2015b). Tick salivary cystatin sialostatin L2 suppresses IFN responses in mouse dendritic cells. *Parasit. Immunol.* 37, 70–78. doi: 10.1111/pim.12162
- Mans, B. J. (2011). Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *J. Innate Immun.* 3, 41–51. doi: 10.1159/000321599
- Mans, B. J., Calvo, E., Ribeiro, J. M., and Andersen, J. F. (2007). The crystal structure of D7r4, a salivary biogenic amine-binding protein from the malaria mosquito *Anopheles gambiae*. *J. Biol. Chem.* 282, 36626–36633. doi: 10.1074/jbc.M706410200
- Mans, B. J., Featherston, J., de Castro, M. H., and Pienaar, R. (2017). Gene duplication and protein evolution in tick-host interactions. *Front. Cell. Infect. Microbiol.* 7:413. doi: 10.3389/fcimb.2017.00413
- Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563. doi: 10.1093/bioinformatics/16.6.562
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003
- Mateo-Borrega, M. B., Garcia, B., Larramendi, C. H., Azofra, J., Gonzalez-Mancebo, E., Alvarado, M. I., et al. (2019). IgE-mediated sensitization to galactose- α -1,3-Galactose (α -Gal) in urticaria and anaphylaxis in Spain: geographical variations and risk factors. *J. Invest. Allerg. Clin.* 29, 436–443. doi: 10.18176/jiaci.0373
- Mulenga, A., Macaluso, K. R., Simser, J. A., and Azad, A. F. (2003). The American dog tick, *Dermacentor variabilis*, encodes a functional histamine release factor homolog. *Insect Biochem. Mol. Biol.* 33, 911–919. doi: 10.1016/S0965-1748(03)00097-3
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., et al. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30, 1196–1205. doi: 10.1093/molbev/mst030
- Narasimhan, S., Deponte, K., Marcantonio, N., Liang, X., Royce, T. E., Nelson, K. F., et al. (2007). Immunity against *Ixodes scapularis* salivary proteins expressed within 24 hours of attachment thwarts tick feeding and impairs *Borrelia* transmission. *PLoS ONE* 2:E451. doi: 10.1371/journal.pone.0000451
- Neafsey, D. E., Waterhouse, R. M., Abai, M. R., Aganezov, S. S., Alekseyev, M. A., Allen, J. E., et al. (2015). Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* 347:1258522. doi: 10.1126/science.1258522
- Nicholas, H. B., Jr., Chan, S. S., and Rosenquist, G. L. (1999). Reevaluation of the determinants of tyrosine sulfation. *Endocrine* 11, 285–292. doi: 10.1385/ENDO.11.3.285
- Nielsen, H., Brunak, S., and von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 12, 3–9. doi: 10.1093/protein/12.1.3
- Olabode, A. S., Avino, M., Ng, G. T., Abu-Sardana, F., Dick, D. W., and Poon, A. F. Y. (2019). Evidence for a recombinant origin of HIV-1 Group M from genomic variation. *Virus Evol.* 5:vey039. doi: 10.1093/ve/vey039
- Perner, J., Kropackova, S., Kopacek, P., and Ribeiro, J. M. C. (2018). Sialome diversity of ticks revealed by RNAseq of single tick salivary glands. *PLoS Negl. Trop. Dis.* 12:e0006410. doi: 10.1371/journal.pntd.0006410
- Pierleoni, A., Martelli, P. L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* 9:392. doi: 10.1186/1471-2105-9-392

- Pond, S. L., Frost, S. D., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- Posada, D., and Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J. Mol. Evol.* 54, 396–402. doi: 10.1007/s00239-001-0034-9
- Radulovic, Z. M., Kim, T. K., Porter, L. M., Sze, S. H., Lewis, L., and Mulenga, A. (2014). A 24–48 h fed *Amblyomma americanum* tick saliva immuno-proteome. *BMC Genomics* 15:518. doi: 10.1186/1471-2164-15-518
- Rhoads, R. E., and Udenfriend, S. (1969). Substrate specificity of collagen proline hydroxylase: hydroxylation of a specific proline residue in bradykinin. *Arch. Biochem. Biophys.* 133, 108–111. doi: 10.1016/0003-9861(69)90493-7
- Ribeiro, J. M., Alarcon-Chaidez, F., Francischetti, I. M., Mans, B. J., Mather, T. N., Valenzuela, J. G., et al. (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mol. Biol.* 36, 111–129. doi: 10.1016/j.ibmb.2005.11.005
- Ribeiro, J. M., Assumpcao, T. C., Ma, D., Alvarenga, P. H., Pham, V. M., Andersen, J. F., et al. (2012). An insight into the sialotranscriptome of the cat flea, *Ctenocephalides felis*. *PLoS ONE* 7:e44612. doi: 10.1371/journal.pone.0044612
- Ribeiro, J. M., and Francischetti, I. M. (2003). Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol.* 48, 73–88. doi: 10.1146/annurev.ento.48.060402.102812
- Ribeiro, J. M., Mans, B. J., and Arca, B. (2010). An insight into the sialome of blood-feeding *Nematocera*. *Insect Biochem. Mol. Biol.* 40, 767–784. doi: 10.1016/j.ibmb.2010.08.002
- Ribeiro, J. M. C., and Arca, B. (2009). From sialomes to the sialoverse: An insight into the salivary potion of blood feeding insects. *Adv Insect Physiol* 37, 59–118. doi: 10.1016/S0065-2806(09)37002-2
- Salim, B., Amin, M., Igarashi, M., Ito, K., Jongejan, F., Katakura, K., et al. (2019). Recombination and purifying and balancing selection determine the evolution of major antigenic protein 1 (map 1) family genes in *Ehrlichia ruminantium*. *Gene* 683, 216–224. doi: 10.1016/j.gene.2018.10.028
- Salmon, A., Flagel, L., Ying, B., Udall, J. A., and Wendel, J. F. (2010). Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol.* 186, 123–134. doi: 10.1111/j.1469-8137.2009.03093.x
- Sa-Nunes, A., Bafica, A., Antonelli, L. R., Choi, E. Y., Francischetti, I. M., Andersen, J. F., et al. (2009). The immunomodulatory action of sialostatin I on dendritic cells reveals its potential to interfere with autoimmunity. *J. Immunol.* 182, 7422–7429. doi: 10.4049/jimmunol.0900075
- Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., et al. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29, 2994–3005. doi: 10.1093/nar/29.14.2994
- Schroeder, H., Daix, V., Gillet, L., Renaud, J. C., and Vanderplassen, A. (2007). The paralogous salivary anti-complement proteins IRAC I and IRAC II encoded by *Ixodes ricinus* ticks have broad and complementary inhibitory activities against the complement of different host species. *Microbes Infect* 9, 247–250. doi: 10.1016/j.micinf.2006.10.020
- Schuijt, T., Dai, J. F., van Dam, A., van de Poll, T., and Fikrig, E. (2010). The tick salivary protein Salp9-like inhibits complement-mediated killing of *Borrelia burgdorferi* sensu lato isolates. *Inflamm. Res.* 59, S23–S23. doi: 10.1128/IAI.00232-08
- Schwarz, A., Valdes, J. J., and Kotsyfakis, M. (2012). The role of cystatins in tick physiology and blood feeding. *Ticks Tick Borne Dis.* 3, 117–127. doi: 10.1016/j.ttbdis.2012.03.004
- Shimizu, M., Igasaki, T., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., et al. (2005). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *Plant J.* 42, 877–889. doi: 10.1111/j.1365-313X.2005.02419.x
- Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.
- Stutzer, C., Mans, B. J., Gaspar, A. R., Neitz, A. W., and Maritz-Olivier, C. (2009). *Ornithodoros savignyi*: soft tick apyrase belongs to the 5'-nucleotidase family. *Exp. Parasitol.* 122, 318–327. doi: 10.1016/j.exppara.2009.04.007
- Tan, A. W., Francischetti, I. M., Slovak, M., Kini, R. M., and Ribeiro, J. M. (2015). Sexual differences in the sialomes of the zebra tick, *Rhipicephalus pulchellus*. *J. Proteomics* 117, 120–144. doi: 10.1016/j.jprot.2014.12.014
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Thompson, R. E., Liu, X. Y., Ripoll-Rozada, J., Alonso-Garcia, N., Parker, B. L., Pereira, P. J. B., et al. (2017). Tyrosine sulfation modulates activity of tick-derived thrombin inhibitors. *Nat. Chem.* 9, 909–917. doi: 10.1038/nchem.2744
- Valenzuela, J. G., Charlab, R., Gonzalez, E. C., de Miranda-Santos, I. K., Marinotti, O., Francischetti, I. M., et al. (2002a). The D7 family of salivary proteins in blood sucking diptera. *Insect Mol. Biol.* 11, 149–155. doi: 10.1046/j.1365-2583.2002.00319.x
- Valenzuela, J. G., Francischetti, I. M. B., Pham, V. M., Garfield, M. K., Mather, T. N., and Ribeiro, J. M. C. (2002b). Exploring the sialome of the tick, *Ixodes scapularis*. *J. Exp. Biol.* 205, 2843–2864.
- Wang, X., Shaw, D. K., Sakhon, O. S., Snyder, G. A., Sundberg, E. J., Santambrogio, L., et al. (2016). The tick protein sialostatin L2 binds to annexin A2 and inhibits NLR4-mediated inflammasome activation. *Infect. Immun.* 84, 1796–1805. doi: 10.1128/IAI.01526-15
- Watson, E. E., Ripoll-Rozada, J., Lee, A. C., Wu, M. C. L., Franck, C., Pasch, T., et al. (2019). Rapid assembly and profiling of an anticoagulant sulfoprotein library. *Procee. Natl. Acad. Sci. U.S.A.* 116, 13873–13878. doi: 10.1073/pnas.1905177116
- Xiao, B., Yang, Z. M., Xiao, M., Chen, H., and Ma, Q. (2017). The structure and biological functions of translationally controlled tumor protein and its role in tumor progression. *Prog Biochem Biophys.* 44, 148–154.
- Zhou, W. S., Woodson, M., Neupane, B., Bai, F. W., Sherman, M. B., Choi, K. H., et al. (2018). Exosomes serve as novel modes of tick-borne flavivirus transmission from arthropod to human cells and facilitates dissemination of viral RNA and proteins to the vertebrate neuronal cells. *PLoS Pathog* 14:e1006764. doi: 10.1371/journal.ppat.1006764

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ribeiro and Mans. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.