Gene Duplication and Protein Evolution in Tick-Host Interactions

Ben J. Mans [1,2,3*], Jonathan Featherston [4], Minique H. de Castro [1,3,4] and Ronel Pienaar [1]

[1] Epidemiology, Parasites and Vectors, Agricultural Research Council-Onderstepoort Veterinary Research, Onderstepoort, South Africa, [2] Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa, [3] Department of Life and Consumer Sciences, University of South Africa, Pretoria, South Africa, [4] Agricultural Research Council-The Biotechnology Platform, Onderstepoort, South Africa

Ticks modulate their hosts' defense responses by secreting a biopharmacopiea of hundreds to thousands of proteins and bioactive chemicals into the feeding site (tick-host interface). These molecules and their functions evolved over millions of years as ticks adapted to blood-feeding, tick lineages diverged, and host-shifts occurred. The evolution of new proteins with new functions is mainly dependent on gene duplication events. Central questions around this are the rates of gene duplication, when they occurred and how new functions evolve after gene duplication. The current review investigates these questions in the light of tick biology and considers the possibilities of ancient genome duplication, lineage specific expansion events, and the role that positive selection played in the evolution of tick protein function. It contrasts current views in tick biology regarding adaptive evolution with the more general view that neutral evolution may account for the majority of biological innovations observed in ticks.

Keywords: tick evolution, gene duplication, protein family evolution, salivary gland, blood-feeding evolution, hematophagy

EVOLUTION OF HEMATOPHAGY IN TICKS

Ticks (Ixodida) are obligate blood-feeding ecto-parasites that evolved a blood-feeding lifestyle >250 million years ago (MYA) (Mans et al., 2011, 2012, 2016). The ancestral hematophagous lineage diverged into extant families: Argasidae (soft ticks ∼200 species), Ixodidae (hard ticks ∼700 species), and the monotypic Nuttalliellidae (*Nuttalliella namaqua*) (Guglielmone et al., 2010). A parasitic blood-feeding lifestyle entails interaction with the vertebrate host, necessitating the evolution of mechanisms to ensure successful acquisition of a blood meal, described as the four stages of blood-feeding evolution (**Figure 1**), namely: host-detection, host-attachment, host-interaction, and blood meal processing (Mans, 2014). For the current study, tick-host interactions and lineage specific innovation that occurred after divergence of the main tick families are of interest. This is most apparent in the comparison of the salivary gland repertoires secreted into the host during feeding. The evolution of salivary gland protein families and protein function is mediated by gene duplication and a major issue is whether tick-host evolution is adaptive or neutral. The current review will examine these issues and aim to illuminate the general theories of gene duplication and functional evolution with tick specific examples.

FIGURE 1 | Gene duplication in the evolution of blood-feeding in ticks. The four main stages in the evolution of blood-feeding includes host detection, host attachment, tick-host interaction, and blood meal processing (Mans, 2014). Tick salivary glands play a central role in tick-host interactions by secreting the bioactive molecules that modulates host defense mechanisms. Some of these secreted proteins belong to large protein families expanded by gene duplication that can count from tens to hundreds of different genes per tick species. BPTI (Basic pancreatic trypsin inhibitor), BTSP (Basic tail secretory proteins), LTB4 (leukotriene B4), LTC4 (leukotriene C4), TXA2 (thromboxane A2). Information on protein families summarized from Mans et al. (2016).

COMPARATIVE TRANSCRIPTOMICS BETWEEN HEMATOPHAGOUS ARTHROPODS

Independent evolution of hematophagy is evident given the paraphyletic nature of blood-feeding arthropod lineages that originated at least 20 times (Ribeiro et al., 2010; Mans, 2011; Mans et al., 2016). Convergent evolution is apparent from the various solutions that blood-feeders found for their common problem, i.e., how to control the vertebrate host's defense mechanisms. A number of generalizations regarding salivary gland transcriptomes suggest that evolutionary mechanisms may be universal for all blood-feeders (Mans, 2011). As such, all blood-feeders secrete bioactive components from their salivary glands into the feeding site that target similar key host defense mechanisms. This underscores the universality of vertebrate host inflammatory and hemostatic mechanisms, since these evolved (>350 MYA), before most arthropod lineages adapted to blood feeding (Delvaeye and Conway, 2009; Conway, 2015). Unique salivary gland protein families in different arthropod lineages, suggests that salivary gland repertoires differed before adaptation to blood feeding depending on the specific lifestyle (scavenger, parasites, plant feeder, or predator) of the non-hematophagous ancestral lineage. Closely related hematophagous lineages have similar salivary gland protein families, even if they do not have similar functions and these same families are prone to gene duplication (Mans, 2011). These generalizations have predictive value for the study of uncharacterized blood-feeding lineages, where we expect targeting of key host processes, unique salivary gland repertoires and limited numbers of protein families that have been expanded by gene duplication.

TICK SALIVARY GLAND TRANSCRIPTOMES

Hundreds to thousands of proteins synthesized in the salivary glands of ticks target key host defense mechanisms involved in inflammation and hemostasis (Francischetti et al., 2009). Extensive tables summarizing hemostatic and immune inhibitors, host targets, protein family affiliations, and distribution of orthologs in genera and lineages were recently updated (Mans, 2016; Mans et al., 2016). Summaries of sequenced transcriptomes, secretory, and house-keeping proteins, protein family members, and largest expanded families were also tabulated (Mans, 2016; Mans et al., 2016). All tick families possess the same major salivary gland secretory protein families (**Figure 1**). These include lipocalins, kunitz-basic pancreatic trypsin inhibitors (BPTI), metalloprotease, serpin, trypsin inhibitor-like (TIL), cystatin, and basic tail secretory (BTSP) families (Mans et al., 2008a, 2016). However, few orthologs are shared between the different tick families and lineage specific expansions of these gene families are common (Mans et al., 2008a, 2016; Dai et al., 2012; Schwarz et al., 2014a). Of interest, is the sudden rise in described secretory family members after introduction of next-generation sequencing technologies (NGS), with an ∼10-fold increase for secretory protein family members from conventional cDNA libraries (10–100s family members) to NGS (100–1,000s family members)

sequenced transcriptomes. The numbers of secretory family members found in the *Ixodes scapularis* genome is closer to that observed for cDNA libraries. Whether the high levels of lineage specific expansion observed in NGS transcriptomes are real due to alternative splicing or gene duplication, or artefactual due to transcriptome assembly, taxonomic sampling, or sequence depth bias cannot be determined yet (Mans, 2016; Mans et al., 2016). However, functional analysis of the major host interacting proteins confirm that they belong to different protein families in the different tick families and therefore evolved these functions independently (Mans et al., 2002, 2016; Mans and Neitz, 2004a; Mans, 2011). The current study will discuss scenarios and evolutionary mechanisms by which these novel and different functions evolved.

SPECIATION, GENE DUPLICATION, AND THE FATES OF GENES

Members of protein families are homologous, i.e., share a common ancestor (Koonin et al., 2002). Homology may be due to vertical descent (speciation) and such genes are orthologs or orthologous. Homology may also be due to gene duplication resulting in two or more members of the same gene within the genome and such genes are paralogs or paralogous (Fitch, 1970, 2000; Koonin, 2005; Gabaldón and Koonin, 2013). During subsequent speciation, these paralogs will become orthologs in the descendant lineages. Gene loss may occur in orthologs and paralogs, or gene duplication may occur in selective descendant lineages and can lead to phylogenetic relationships that are complex and difficult to resolve, given our dependence on information from extant lineages. Attempts to classify these relationships led to creation of designations such as co-orthologs (a gene from one species is collectively orthologous to duplicated genes in other species), inparalogs (lineage-specific gene duplications occurring after speciation), and outparalogs (lineage-specific gene duplications occurring before speciation) (Sonnhammer and Koonin, 2002; Koonin, 2005). It becomes very difficult to delineate these without a broad taxonomic sampling of closely related and divergent lineages and for most discussions, the general concepts of orthologous and paralogous genes are sufficient (Jensen, 2001). Confounding factors in ortholog identification includes domain-shuffling, acquisition/loss of new domains and alternative splicing (Gabaldón and Koonin, 2013). Most small tick proteins involved at the tick-host interface belong to single domain families (BPTI, cystatin, lipocalin, serpin, TIL), or when multi-domains are present they are generally oligomers of the same domain (e.g., BPTI, BTSP) (Francischetti et al., 2009). Even so, given the fact that extensive lineage-specific expansions occur in ticks (existence of co-orthologs and inparalogs), the identification of orthologs remains problematic.
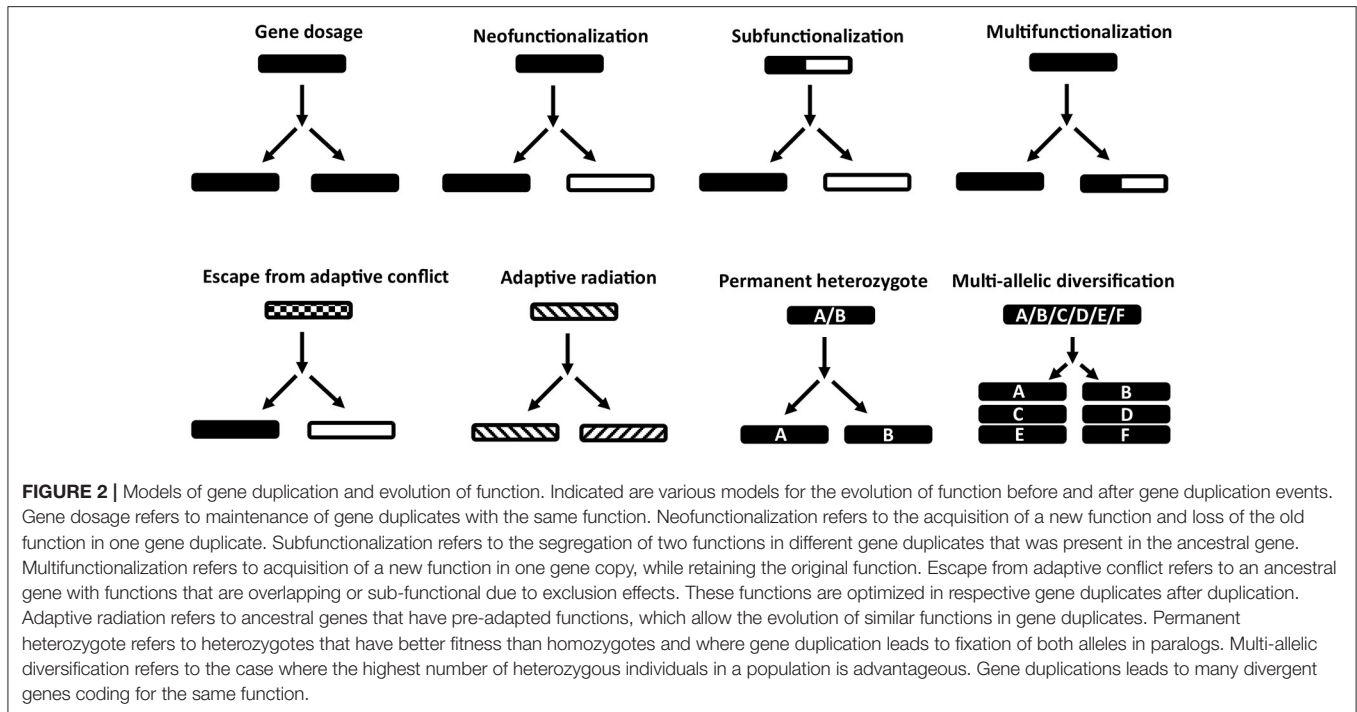
ORTHOLOGOUS GENES IN TICKS

Orthologs generally possess similar molecular structure, function, mechanism of action, conserved residues involved in molecular interaction and domain architecture across species or lineages and can be traced to the last common ancestral lineage where this function originated (Gabaldón and Koonin, 2013). This is the basis for the universality of general cell biological processes such as transcription, translation, cellular localization, secretion, transport, metabolism, and our ability to annotate genes by homology as encapsulated in the ortholog conjecture (Gabaldón and Koonin, 2013; Rogozin et al., 2014). Since most orthologs perform vital functions in general cell biology or development, they are fixed in populations or species by negative selection. Even so, gene losses occur, orthologs acquire new or additional functions, domains are exapted for a new function (Gabaldón and Koonin, 2013). Exaptation of house-keeping functions at the tick-host interface have occurred as seen for glycolytic enolase, that also function as plasminogen activator (Díaz-Martín et al., 2013). Numerous orthologs specifically involved in tick-host interaction exist. Apyrases, biogenic amine binding proteins (BABP), enolase, metalloproteases, and defensins evolved in the last common ancestor to all ticks (Mans et al., 2016). BPTI-thrombin inhibitors, BPTI-fibrinogen receptor antagonists, and cysteinyl leukotriene scavengers evolved in the last common ancestor to soft ticks (Mans and Ribeiro, 2008a; Mans et al., 2008b). Ixodegrins, serpin-thrombin inhibitors, BPTI-thrombin inhibitors, and immunoglobulin-binding proteins evolved in the last common ancestor to hard ticks (Mans et al., 2016). Similarly, orthologs restricted to different genera exist, for example, the moubatin-clade thus far restricted to the genus *Ornithodoros* that are paralogous to the biogenic amine binding clade (Mans and Ribeiro, 2008b).

PARALOGOUS GENES IN TICKS

Once a gene duplicates several possibilities exist for its fate (**Figure 2**). One copy may rapidly acquire deleterious mutations and become pseudogenized (non-functionalization) (Innan and Kondrashov, 2010). Paralogs may retain the same function and if maintained by negative selection, are considered to be due to requirements of a dosage effect in house-keeping proteins, i.e., high concentrations are necessary to fulfill the requirements of the cell (Innan and Kondrashov, 2010; Rogozin et al., 2014). Genes in this category include histone and 18S, 5.8S, and 28S ribosomal RNA cassettes that occur in hundreds of copies in metazoans, including ticks (Mans et al., 2015). All copies are highly similar and maintained via concerted evolution, gene conversion or strong purifying selection (Nei and Rooney, 2005; Innan and Kondrashov, 2010). Examples of paralogs maintained for dosage effects at the tick-host interface are the savignygrins and monogrins secreted at high concentrations to target the abundant fibrinogen receptor on platelets (Mans et al., 2003a, 2008b).

Mutations may accumulate in one copy resulting in loss of the old function and evolution of a new function (neo-functionalization) and leads to the generalization that paralogs have different functions (Gabaldón and Koonin, 2013). Functions may be similar in a general sense, i.e., members of the Ran-GTP or karyopherin families all function in cellular transport and trafficking (Mans et al., 2004); tick lipocalins function

FIGURE 2 | Models of gene duplication and evolution of function. Indicated are various models for the evolution of function before and after gene duplication events. Gene dosage refers to maintenance of gene duplicates with the same function. Neofunctionalization refers to the acquisition of a new function and loss of the old function in one gene duplicate. Subfunctionalization refers to the segregation of two functions in different gene duplicates that was present in the ancestral gene. Multifunctionalization refers to acquisition of a new function in one gene copy, while retaining the original function. Escape from adaptive conflict refers to an ancestral gene with functions that are overlapping or sub-functional due to exclusion effects. These functions are optimized in respective gene duplicates after duplication. Adaptive radiation refers to ancestral genes that have pre-adapted functions, which allow the evolution of similar functions in gene duplicates. Permanent heterozygote refers to heterozygotes that have better fitness than homozygotes and where gene duplication leads to fixation of both alleles in paralogs. Multi-allelic diversification refers to the case where the highest number of heterozygous individuals in a population is advantageous. Gene duplications leads to many divergent genes coding for the same function.

as scavengers of diverse bioactive molecules such as biogenic amines (Mans et al., 2008c), cysteinyl leukotrienes (Mans and Ribeiro, 2008a), thromboxane A2 (TXA2), and leukotriene B4 (LTB4) (Mans and Ribeiro, 2008b); or BPTI inhibitors that target different serine proteases such as fXa and thrombin (Mans et al., 2002), but differ in target specificity. Functions may also differ completely, such as the platelet aggregation inhibitors that belong to the BPTI protein family, but target integrin receptors (Mans et al., 2008b).

Some genes may have multiple functions and after duplication, each gene loses one function but retains the other (sub-functionalization) (Force et al., 1999; Lynch and Force, 2000). An example in ticks may be the histamine and serotonin binding proteins, monotonin and monomine in *Argas* ticks that probably derived from a lipocalin capable of binding both histamine and serotonin in the same site (Mans et al., 2008c).

Some paralogs may also retain the original function and evolve new functions, thereby becoming multifunctional (multi-functionalization). An example in ticks would be the complement C5 inhibitors from *Ornithodoros* ticks that retained the ancestral function of scavenging LTB4, but evolved C5 complement binding activity on their βH-α2 loop (Mans and Ribeiro, 2008b).

If multiple functions are overlapping in mechanism, hindering functional optimization, gene duplication, and subsequent divergence can optimize functions; the escape from adaptive conflict model (Hughes, 1994; Des Marais and Rausher, 2008). In ticks, an example of this specific model has not yet been found, although candidates exist that would certainly benefit from this, e.g., moubatin that binds LTB4 (inhibition of neutrophil migration) and TXA2 (inhibition of platelet aggregation) in the

same binding pocket at similar affinities (Mans and Ribeiro, 2008b). This increases the risk of competitive exclusion and neutralization of both functions if both antagonists were to be present in the feeding site at the same time.

Proteins with multiple functions or broad specificities may already be predisposed to evolve a new function once duplicated, the adaptive radiation model (Francino, 2005). Tick inhibitors with similar general functions may fall in this class. Positive selection has originally been promoted as the driving force for gene-level adaptive radiations, but it should be noted that neutral or even mildly deleterious mutations can also accumulate in genes and become fixed within a population through processes such as drift. It may be argued that neutral evolutionarily processes play a particularly important role for generating gene diversity and novel functions in genes with broad specificities because mutations in genes with promiscuous activity are more likely to be neutral than they are to have either a positive or negative effect on fitness.

Where heterozygous alleles have better fitness than homozygotes, gene duplication can create a permanent heterozygote (Proulx and Phillips, 2006). An extension of this model, the multi-allelic diversifying selection model, occurs when heterozygosity is advantageous at the population level (Innan and Kondrashov, 2010). In such instances, gene duplication leads to fixation of different heterozygotes under positive selection, accumulation of new alleles and gene expansion. For both these models, hyper-variable genes in ticks that possess the same function would fit in these models as proposed for genes evading the host immune system (Chmelař et al., 2016). However, even where related extant taxa have been well-sampled, in most instances it is challenging to determine

which model best accounts for specific instances of family expansions because while the underlying causes for family diversification may differ the outcomes may appear similar. For example, when considering the multi-allelic diversifying model, the gene dosage model may instead be considered for gene duplicates where the active or binding site is internally located or the functional residues are restricted to a small area of the protein structure and where purifying selection only acts on these regions to conserve functionality and protein fold structure. Equally, neutral evolution and accumulation of mutations also lead to extremely diverse families that have the same or similar functions that would resemble the multi-allelic diversification model, but without the need for positive selection.

RATES OF GENE DUPLICATION

Gene duplication gives rise to paralogs and is considered the major mechanism to generate genetic diversity and new functions (Lynch, 2002). The general rate of gene duplication is ~1% per gene per million years, while 50% of duplicated genes are lost every 4 million years (Lynch, 2002). In a species with ~20,000 genes (such as ticks), ~1,000 genes will become fixed over 10 million years. With regard to ticks, *I. scapularis* is the only genome available (Gulia-Nuss et al., 2016), with 2–3% predicted genomic and 7–22% transcriptomic paralogs (Van Zee et al., 2016). Paralog pairs (two duplicated genes) make up ~80% of paralogs found in the genome, indicating a single duplication event for these genes. The majority of these duplications occurred <6 MYA (Van Zee et al., 2016), which fit the expected gene duplication rate. The transcriptome data suggest a higher rate of gene duplication than expected. Genome and transcriptome differences are partly due to a higher number of genes in salivary gland transcriptomes (Valenzuela et al., 2002; Ribeiro et al., 2006), not present in the genome assembly, which only extracted gene annotations from ~57% of the genome (Gulia-Nuss et al., 2016). Conversely, other measures of completeness such as the core eukaryotic genes mapping approach (CEGMA, 248 conserved genes) and benchmarking universal single-copy orthologs (BUSCO, 2,675 conserved genes), indicate ~80 and ~69% genes represented in the *I. scapularis* genome, respectively (Hoy et al., 2016). A lack of gene coverage alone cannot account for these differences.

The only related tick species with extensive genetic data is *I. ricinus* for which salivary, midgut, and hemocyte transcriptomes were sequenced and a draft genome assembled (Chmelař et al., 2008; Schwarz et al., 2013, 2014b; Cramaro et al., 2015, 2017; Kotsyfakis et al., 2015a,b; Perner et al., 2016). Molecular clock analysis based on the mitochondrial proteins, suggests that *I. scapularis* and *I. ricinus* diverged ~9 MYA, suggesting the majority of the gene duplications detected by Van Zee et al. (2016) occurred after their divergence. It is therefore of interest that ticks exhibit extensive duplicated gene families in their salivary gland transcriptomes which seem to be lineage or species specific expansions, i.e., phylogenetic analysis indicate protein family clades that effectively consist of genes from only one species (Mans et al., 2008a, 2016; Schwarz et al., 2013). Reciprocal best hit analysis of non-redundant datasets (clustered

at 95% protein identity to remove possible alleles, resulting in 20,869 *I. scapularis* and 30,641 *I. ricinus* proteins), indicates that 10,105 best hits (orthologs) occur between *I. scapularis* and *I. ricinus* ($<E^{-10}$) (Mans et al., 2016). This corresponds with ~10,000 orthologs found between *I. scapularis* and other parasitiform mites (Dong et al., 2017). Unique proteins from each dataset with significant hits ($<E^{-10}$) in their respective orthologous sets, comprised 23% (4,738 proteins) and 53% (16,000 proteins), respectively for *I. scapularis* and *I. ricinus*, indicating that these paralogs duplicated after speciation, or were lost in different species. In the case of *I. ricinus*, it indicates a larger gene duplication rate than *I. scapularis*. This could be due to genes missing from the *I. scapularis* assembly, true duplications in *I. ricinus* or artifacts from *de novo* assembled transcriptomes derived from NGS (Mans et al., 2016). The genome size of *I. ricinus* (~2.71 Gbp) is larger than *I. scapularis* (~2.2 Gbp) (Geraci et al., 2007; Cramaro et al., 2017), and may explain some of the paralog differences. However, for both species the number and rate of gene duplications are higher than expected, differing by ~5–16-fold from the expected number of gene duplications. With regard to major secretory protein families, the Kunitz-BPTI, BTSP, and lipocalin protein family members of *I. scapularis* possess 69 (39%, 176 total), 49 (40%, 123 total), and 47 (39%, 119 total) unique genes (duplicated per species), respectively. For *I. ricinus* there are 846 (80%, 1,058 total), 682 (82%, 834 total), and 1,011 (79%, 1,283 total) unique genes for the Kunitz-BPTI, BTSP, and lipocalin family members, respectively (Mans et al., 2016). This ranges from 40 to 80% of the total members for each family and is higher than the overall percentage of duplicated genes (7–22%) observed in the genome of *I. scapularis* (Van Zee et al., 2016). This differs by orders of magnitude from the average rate of gene duplication observed in other arthropods (Gulia-Nuss et al., 2016; Hoy et al., 2016; Dong et al., 2017), and suggests elevated gene duplication rates in ticks and most especially for secretory proteins. The question raised, is whether this number of gene duplications could have occurred since speciation, with the implication that these duplicates are an evolutionary response of the tick to a blood-feeding environment.

MOLECULAR ARMS RACES

The diversity in gene family members may be an adaptive response to escape the host immune system described as an escalating host-parasite arms race that leads to innovation from either side to counter-act new functions (Mans, 2011; Chmelař et al., 2016). This may take the form of new functions evolving, making defense systems on both sides more redundant (i.e., the Red Queen hypothesis). It could take the form of expansion of the same protein family to yield antigenic variants that retain the same function. These variants, secreted concurrently at low levels during feeding, evade the immune system, while still achieving a concentration necessary for immune system inhibition: the varying epitope hypothesis (Couvreur et al., 2008; Chmelař et al., 2016). The varying epitope hypothesis is a variant of the multi-allelic diversifying selection model (Innan and Kondrashov, 2010). It may also take the form of an expanded family that escapes the host's immune system by differential

expression during feeding: the antigenic shift during feeding hypothesis (Chmelař et al., 2016). Supporting evidence for this may be heightened non-synonymous substitution rates (Kn/Ks > 1, positive selection), indicative of adaptive selection (Kotsyfakis et al., 2015b; Ribeiro et al., 2017).

Paralogous proteins from the same clade with similar functions such as BABPs, LTB4 scavengers or GPIIbIIIa inhibitors have appreciable sequence similarity, excluding antigenic variation to escape the immune system of the host (Mans et al., 2003a, 2008b,c; Mans and Ribeiro, 2008b). These paralogs probably fulfill a dosage requirement, since functions like scavenging of bioactive molecules or targeting abundant platelet receptors, require high concentrations of inhibitors. The only study that characterized an extensive multigene family that shows the same function, is the Salp20 group of complement inhibitors that disrupts the active C3 convertase (C3bBbP) complex by binding properdin (Valenzuela et al., 2000; Daix et al., 2007; Tyson et al., 2007, 2008; Couvreur et al., 2008). Different Salp20 genes are expressed in different individuals across the time course of feeding (Couvreur et al., 2008). This family would represent a mixed model of antigenic shift during feeding and varying epitopes hypotheses. This family has a number of gene members in both *I. scapularis* and *I. ricinus* and shows overall sequence identity of 35–75%. The members of this family are all highly N- and O-linked glycosylated, while thrombospondin repeats of properdin specifically bind sulfated glycoconjugates and glycosaminoglycans (Couvreur et al., 2008; Tyson et al., 2008). The mechanism of action among the various gene duplicates may in certain instances be sequence independent (i.e., glycosylation dependent), relaxing purifying selective constraints on the protein sequences, while retaining their function. This family would then be a very specific case of functional conservation due to post-translational modification, and not necessarily an example of antigenic shift or varying epitopes and may not be good support in favor of these hypotheses. Similarly, scavengers of bioactive molecules such as the lipocalins, may show extensive sequence variation, since their conserved binding sites are internally located and antibody responses would not necessarily block activity. Their sequence diversity may therefore not be due to antigenic variation, but may be due to genetic drift and relaxation of purifying selection.

Co-evolution of tick and host in an escalating arms race was proposed to explain the diversity of tick salivary gland proteins (Mans, 2011; Chmelař et al., 2016). In this, positive selection of secretory proteins plays an important role as a means to evade the host's immune system (Kotsyfakis et al., 2015b; Ribeiro et al., 2017). The Red Queen hypothesis was also invoked to explain gene duplication and protein evolution in salivary glands as part of an escalating arms race (Schwarz et al., 2014a). The Red Queen hypothesis is an extension of co-evolution that specifically indicates that as ticks evolve numerous functions, the host should respond in kind. Whereas, extensive gene duplication and genetic diversity is observed for ticks, the same cannot be said for the host hemostatic or even immune systems. The basic hemostatic and immune

functions of the host were evolved before adaptation of ticks to a blood-feeding lifestyle and are generally conserved in mammals (Delvaeye and Conway, 2009; Conway, 2015). The Red Queen hypothesis cannot therefore be strictly applied to tick-host interactions.

POSITIVE SELECTION IN THE FACE OF NEUTRAL EVOLUTION AND GENETIC DRIFT

The current dominant view in evolutionary biology is that neutral evolution and genetic drift may have been responsible for many evolved functions and that neutral evolution should be considered as the null hypothesis that must first be rejected before adaptive explanations are considered (Lynch, 2007; Nielsen, 2009; Koonin, 2016). Positive selection, where the rate of non-synonymous substitutions in orthologs (i.e., different species) is greater than synonymous substitutions (Kn/Ks > 1), is the gold standard measure of adaptive signal (Yang and Bielawski, 2000; Nielsen, 2005; Kondrashov, 2012). When comparing closely related taxa, the majority of genes found in eukaryotic genomes have a Kn/Ks ratio ≤1, which is indicative of neutral evolution or negative/purifying selection. Overall, few genes have been identified that exhibit signatures of positive selection (Yang and Bielawski, 2000; Kondrashov, 2012). In ticks, positive selection has been detected in many secretory proteins supporting the hypothesis that ticks have adapted to a blood-feeding environment (Daix et al., 2007; Couvreur et al., 2008; Dai et al., 2012; Kotsyfakis et al., 2015b; Van Zee et al., 2016; Ribeiro et al., 2017). All of these studies calculated Kn/Ks ratios between paralogous proteins within single species or used mapping and variant detection of NGS data, where higher Kn/Ks ratios may reflect non-specific mapping of reads to paralogs. These approaches are similar to measures of volatility and liable to suffer from the same drawbacks from which no conclusion can be made regarding positive selection (Nielsen and Hubisz, 2005). In contrast, if orthologs and paralogs between *I. ricinus* and *I. scapularis* from the Salp20 families are compared, ortholog Kn/Ks ratios are below 1 (Daix et al., 2007). Detection of positive selection is complicated by the presence of paralogs in a dataset (Brieuc and Naish, 2011), and increased Kn/Ks ratios may be expected between paralogs with different functions, since their divergence entails non-synonymous substitutions. Relaxation of purifying selection in recently duplicated genes (i.e., increased rates of mutation) is also observed, but is not necessarily due to positive selection (Kondrashov et al., 2002). In ticks, the average Kn/Ks obtained for secreted proteins ranges from Kn/Ks of 1.1–1.3 (Kotsyfakis et al., 2015b; Ribeiro et al., 2017), which may be interpreted as slightly above neutral signal. However, this may be due to relaxation of purifying selection rather than positive selection, with the caveat that differences between paralogs may still be more extensive than between orthologs. An alternative definition of neutral evolution, where mutations that do not alter gene function appreciably are considered neutral, Kn/Ks-values slightly above or below 1 are still considered neutral

(Nei, 2005). As yet, we cannot conclude that positive selection is evident in secretory proteins at genome or transcriptome scale for ticks until the Kn/Ks ratios for orthologs between closely related species such as *I. scapularis* and *I. ricinus* have been determined. Therefore, the null hypotheses of neutral evolution cannot be rejected in favor of an adaptive interpretation for tick-host interactions.

NEUTRAL EVOLUTION AND GENETIC DRIFT AS DRIVERS OF BLOOD-FEEDING EVOLUTION

If neutral evolution and genetic drift are major causes of diversity in ticks and adaptive selection is limited, the questions remain how ticks evolved the impressive array of functions exhibited at the feeding site and why elevated levels of gene duplication are observed in salivary gland families (Mans et al., 2016). One possibility lies in the opportunity that salivary gland expression offers for the evolution of new functions (Mans, 2011, 2016). Proteins destined for secretion need to maintain signals for secretion and sorting, i.e., signal peptides and sorting signals for secretory granules (Nielsen et al., 1997; Gomez-Navarro and Miller, 2016). Once a protein enters the secretory pathway, deleterious effects on cellular activities or its own function will be negated by containment within cellular and extracellular compartments, such as the secretory granule where it is packaged with other secretory proteins before being secreted into the external environment of the feeding site. Since the energetic cost of maintaining gene duplicates is relatively inexpensive (Lynch and Marinov, 2015), and deleterious mutations will not affect other cellular processes, ticks can maintain such sub-functional or even non-functional genes until they are fixed by neutral evolution and genetic drift. Neutral mutations may accumulate in these secreted sub-functional (or even non-functional genes), and be fixed as long as the secretory signals or structural determinants are maintained (Lynch et al., 2016). Once a mutation arises that is beneficial, followed by subsequent mutations that enhances this new functionality, this function may be maintained by purifying selection. This is the general case of neofunctionalization, where new functions evolve due to random neutral mutations (Innan and Kondrashov, 2010). Evolution of new function may even be more subtle, when redundant pathways or processes are targeted, starting with maintenance of the original function in the new gene duplicate until a mutation arises that changes its functional specificity slightly and allows binding to related targets with related functions. Subsequent optimization of function will lead to loss of the original function and gain of a new function. This will be a combination of the multi-functionalization, neofunctionalization, subfunctionalization, escape from adaptive conflict, and adaptive radiation models. Another possibility is for genes with broad target specificities that duplicate and subfunctionalize due to loss-of-function mutations, which result in maintenance of gene duplicates with different functions (Lynch and Force, 2000). These scenarios may explain why so many members of the same protein family

target similar members in a host family (BPTI➔serine proteases, serpins➔serine proteases, cystatins➔cysteine proteinases, evasins➔chemokines). It also implies that many proteins in salivary transcriptomes may have sub-optimal or no functions, but are not pseudogenes in the classical sense (Mudge and Harrow, 2016), since they are still transcribed, expressed and secreted. In this regard, pseudogenes may compose a large part of multigene families in genomes ranging from 25 to 60% (Nei and Rooney, 2005). It may also explain in part the lineage specific expansions observed in salivary gland protein families, since duplicated genes may rarely be lost as observed for house-keeping duplicates. The birth and growth of multigene families approximate power law behavior (Koonin et al., 2002; Koonin, 2011), with those families that start to expand due to gene duplication, accumulating more gene duplications over time. In this scenario, multigene families can rapidly expand without recourse to adaptive selection. Instead of an adaptive arms race with the host, on the battlefield of the tick-feeding interface (Chmelař et al., 2016), the salivary glands and their secretory proteins become a vast experimental playground, where the building blocks can be tested and changed by neutral evolution without being discarded. The latter scenario posits the null hypothesis of neutral evolution in terms of tick-host interactions and the general argument may hold for blood-feeding arthropods in general, since all show expansion of lineage specific protein families (Mans, 2011).

DATING OF MULTIGENE FAMILIES

Gene duplication linked with molecular sequences and phylogenetic analysis offers the promise of dating gene duplication events (Kumar, 2005). This could allow determination of gene duplication rates and estimation of the emergence time of new functions. However, without accurate constraints on nodes, the high level of divergence in salivary gland paralogs could lead to over-estimation of divergence times using molecular clock models. While this is an acknowledged problem in molecular clock dating techniques (Kumar, 2005), scenarios arise where divergence estimations for *I. ricinus* and *I. scapularis* based on analysis of the BPTI family indicated multiple divergence dates for these species in different clades of the same protein family (Schwarz et al., 2014a). As such, dates ranged from 10 to 66 MYA for divergence of *I. ricinus* and *I. scapularis*, and from 8 to 60 MYA for clades that were clearly lineage specific expansions in *I. ricinus* (Schwarz et al., 2014a). This implies non-clock-like behavior in different clades (Mans et al., 2016). Since *I. scapularis* and *I. ricinus* are genetically closely related and estimated to have diverged ∼9 MYA, only one divergence date is possible for these species, while lineage specific expansion likely occurred <6 MYA in both species (Van Zee et al., 2016). This negates most of the evolutionary conclusions made in the Schwarz et al. (2014a) study. It was suggested that clade G6 is fundamental in understanding evolution of the BPTI inhibitors in *I. ricinus* (Schwarz et al., 2014a). This clade is the largest monophyletic clade (lineage specific expansion) and the fastest evolving clade (diverging around 70 MYA as estimated with the molecular clock). However, as indicated, the divergence observed

could only have occurred <9 MYA and therefore represents an interesting view on the gene duplication rate in this family. It was suggested that accelerated evolution is evident from this analysis based on the number of members (Schwarz et al., 2014a). However, counts of the numbers of orthologs (15), paralogs (150) and gene duplications per orthologous clade (Average = 4.9, Range = 1–22), indicate that any given orthologous clade did not expand excessively (≤25 duplications/paralogous clade in 5 million years). Given power law behavior this number of gene duplications may be well within the expected norm for the major secretory protein families. If it is taken into account that some genes may be pseudogenes (Nei and Rooney, 2005), that some orthologous groups may be underrepresented in the *I. scapularis* genome, or that the *I. ricinus* transcriptome may be artefactually inflated, then the proposed accelerated evolution is not obvious. What this section highlights is the need for robust and careful analysis when multigene families are analyzed, specifically with regard to divergence time estimates. Efforts to establish which genes are truly expressed as proteins and functional at the feeding site would also assist in making phylogenetic analysis more robust. Taxonomic sampling of closely related lineages to assist in the identification of orthologous clades to prevent artificial interpretations of species specific expansions is necessary and important to correlate divergence dates for clades expected to have diverged at the time of speciation. Conversely, inclusion of highly divergent sequences from distantly related species does not increase nodal confidence, since small families such as the BPTI fold are likely to have accumulated saturated substitutions. In these cases phylogenetic trees may be obtained that given the low nodal support are basically non-informative (Dai et al., 2012). Given the expansion of tick sequences, future analyses of multigene families may need to focus exclusively on specific lineages or genera, with the emphasis that adequate taxon sampling in these lineages or genera is necessary to delineate orthologous relationships and lineage specific expansions. Inclusion of soft and hard tick sequences in the same analysis leads to the interesting conclusion that argasid and prostriate ticks diverged 155 or 100 MYA, while metastriate ticks diverged 193 MYA (Schwarz et al., 2014a), or that hard and soft tick sequences are interspersed throughout the tree with low nodal support and no evidence for orthology (Dai et al., 2012). This may be expected when divergent paralogous genes are analyzed, which beyond sharing a common ancestor in the ancestral tick lineage, may be separated by multiple gene duplication events. Analysis of multigene families should therefore be performed within a biological context, taking into account recognized phylogenetic relationships (Null hypothesis) and where phylogenetic data deviate from this, data that is not robust should be discarded. Conversely, multigene families and paralogous clades offer the opportunity to reduce uncertainty in molecular clock estimates by using the principle that divergence of paralogous clades from different species by definition occur at the same time. Different nodes can then be constrained by cross-calibration or cross-bracing to ensure that they will have similar ages, thereby improving time estimates across the whole tree (Shih and Matzke, 2013; Zhaxybayeva et al., 2013).

PHYLOGENETIC ANALYSIS OF FUNCTIONAL EVOLUTION

While adaptive selection may not be a major phenomenon in the evolution of blood-feeding behavior, ancestral reconstruction of functional evolution allows for a historical description of "adaptation to a blood-feeding environment." This approach aims to answer the how, when and where functions evolved in different tick lineages (Mans et al., 2016). Phylogenetic analysis linked with functional annotation allow for the reconstruction of the evolution of functions in paralogous genes or testing of orthologous relationships to define the origins of functions in ancestral lineages. However, most phylogenetic analyses of salivary gland protein families are used to summarize or visualize transcriptome data and not to construct explicit functional evolution scenarios. Little information regarding evolution of function is extracted from these analyses beyond confirmation of lineage specific expansion and assignment of proteins to clades (Valenzuela et al., 2002; Francischetti et al., 2005, 2008a,b, 2011; Couvreur et al., 2008; Mans et al., 2008a; Anatriello et al., 2010; Karim et al., 2011; Ribeiro et al., 2011, 2012; Garcia et al., 2014; Tan et al., 2015). This is partly due to the paucity in available functional information and, given the increase in transcriptome studies and the lag in functional studies are likely to continue into the future. However, it is also due to the extensive lineage specific expansions observed in protein families, making identification of orthologs in different genera difficult, since the families are now becoming so large that accurate multiple alignments and well-supported phylogenetic trees are not feasible goals anymore (Mans et al., 2016). On the other hand, phylogenetic analyses have also been performed to analyze protein families during their functional characterization. This was done in the BPTI family for hemostatic (Mans et al., 2002, 2003a, 2008b; Francischetti et al., 2004), ion channel (Paesen et al., 2009), and tryptase inhibitors (Valdés et al., 2013). Analysis of the BPTI family in general was performed for *I. ricinus* (Schwarz et al., 2014b; Valdés and Moal, 2014), *I. scapularis* (Dai et al., 2012), and *R. microplus* (Louw et al., 2013). For the lipocalin family it has been performed for the biogenic amine binding proteins (Mans et al., 2003b, 2008c; Mans and Neitz, 2004b; Mans and Ribeiro, 2008a,b; Díaz-Martín et al., 2011; Valdés et al., 2016), LTB4 and TXA2 binding proteins (Mans and Ribeiro, 2008b), LTC4 binding proteins (Mans and Ribeiro, 2008a; Manzano-Román et al., 2016), japanin (Preston et al., 2013), and savicalin (Cheng et al., 2010). Lipocalins have also been analyzed in general for *I. ricinus* (Beaufays et al., 2008; Konnai et al., 2011), *Ornithodoros savignyi* (Mans et al., 2003b; Mans and Neitz, 2004b), and *R. microplus* (Rodriguez-Valle et al., 2013). Evolution of the serpin family has been studied in *A. americanum* (Mulenga et al., 2007; Porter et al., 2015), *I. scapularis* (Mulenga et al., 2009), and *R. microplus* (Tirloni et al., 2014). Cystatins were analyzed in *I. scapularis* (Kotsyfakis et al., 2006; Ibelli et al., 2013) and novel families described using phylogenetic analysis (Mulenga et al., 2013). It is clear that phylogenetic analysis has become a central part of salivary gland protein family characterization. However, the majority of these studies still use phylogenetic analysis to place their proteins of

interest within a context of other proteins with known function, predict function by homology, or attempt to describe domain evolution in general. Reconstruction of functional evolution of paralogous clades is still limited and mainly due to the lack of knowledge regarding the functions and functional mechanisms of the majority of secretory proteins.

Studies that addressed scenarios for the evolution of function between paralogs, include the clotting and platelet aggregation inhibitors from soft ticks (Mans et al., 2002, 2008b), evolution of biogenic amine binding (Mans et al., 2008c) and evolution of LTB4, TXA2, and complement C5 inhibitors in the moubatin clade of soft tick inhibitors (Mans and Ribeiro, 2008a). For BPTI proteins it has been proposed that soft tick thrombin, fXa and fibrinogen receptor inhibitors share a common evolutionary pathway (Mans et al., 2002, 2008b). Both thrombin and platelet aggregation inhibitors have orthologs in the *Argas* and *Ornithodoros* genera, implying that blood-clotting and platelet aggregation inhibitors evolved in the last common ancestor to soft ticks, which was dated at ~234 ± 25 MYA using mitochondrial protein analysis (Mans et al., 2008b, 2012).
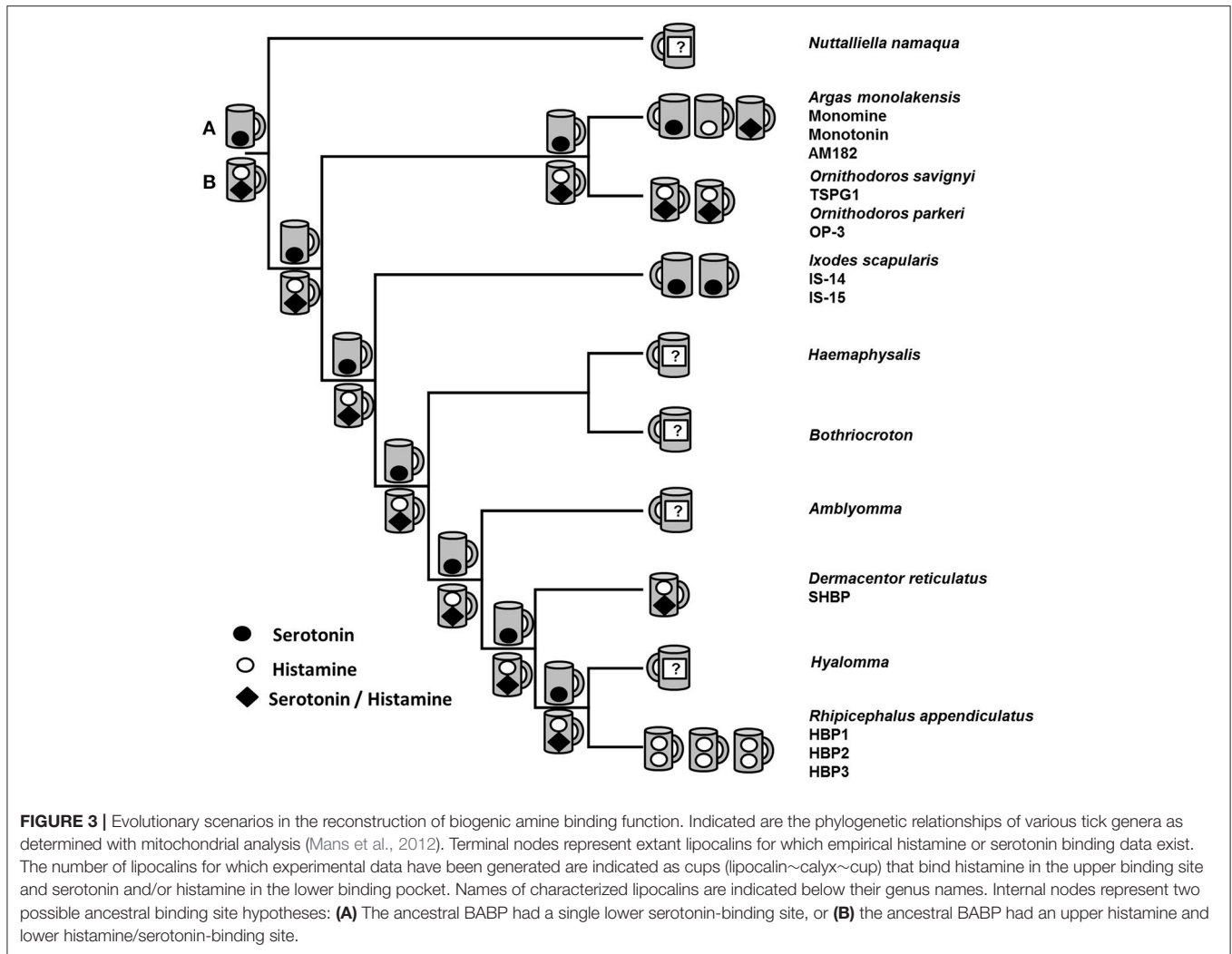
The biogenic amine binding clade represents an interesting model for functional evolution since members exist that can bind histamine in an upper binding pocket (or not) and histamine and/or serotonin in a lower binding pocket (Paesen et al., 1999, 2000; Sangamnatdej et al., 2002; Mans et al., 2008c). Several models for evolution of biogenic amine binding therefore exist that cannot yet be adequately resolved (**Figure 3**). Parsimony analysis of the BABPs suggested that the ancestral biogenic amine bound in the lower binding pocket of the lipocalin barrel was serotonin and that histamine binding originated several times independently, both in the lower and upper binding sites (Mans et al., 2008c). This would represent the neofunctionalization model of biogenic amine binding. Alternatively, the ancestral BABP possessed both an upper histamine-binding and a lower histamine/serotonin-binding site as retained in lipocalins from *Ornithodoros* and *Dermacentor* (Sangamnatdej et al., 2002; Mans et al., 2008c). Subsequent gene duplication and subfunctionalization (or escape from adaptive conflict, since histamine and serotonin shares the same lower binding pocket) led to loss of the upper binding site and lipocalins that either binds histamine or serotonin in the lower binding site as observed for monotonin and monomine from *Argas*, or the *Ixodes* serotonin binding lipocalins (Mans et al., 2008c). In the case of *I. scapularis*, the histamine-binding gene may have been lost after subfunctionalization, although a much larger screening of genes is necessary to confirm this. A similar scenario exists for the histamine-binding lipocalins from *Rhipicephalus appendiculatus* that seem to bind only histamine (Paesen et al., 1999). These alternative scenarios may be resolved once more members are empirically verified from different genera, emphasizing the need for a broad taxon sampling in evolutionary studies. The evolutionary models presented here aim to distill our current knowledge on BABP evolution from experimentally verified data. It should be noted that the evolutionary pathways may be more complex, since an expansion of proteins with the biogenic amine binding motif is observed in all tick transcriptomes, leading to potentially tens to hundreds of biogenic amine binders in each species. It is likely that all different models of functional evolution

after gene duplication are present in this family, making it a particularly interesting study model.

Another clade with detailed functional analysis that allows for dissection of evolutionary scenarios are the moubatin-clade of LTB4, TXA2 and complement C5 inhibitors (Mans and Ribeiro, 2008a). This clade shares a sister relationship with the BABP clade and is found only in the Ornithodorinae. It therefore represents paralogous gene duplication that occurred in the last common ancestor of the Ornithodorinae. The residues conserved for TXA2 and LTB4 binding are conserved across the family, suggesting that this is the ancestral function of this clade. Interestingly enough, this is a potential case for escape from adaptive conflict. In this regard, glycine 85 located inside the lipocalin beta-barrel is important for TXA2 binding, since other more bulky residues such as arginine interferes sterically with binding. TSGP2 that possesses an arginine at this position is unable to bind TXA2, but can be rescued by a R85G mutation. The same R85G mutation occurs in *Ornithodoros moubata* complement inhibitor (OMCI) and would suggest that LTB4 binding has been optimized in these two members of the clade. Both of these lipocalins (and TSGP3) target complement C5 using a very specific βH-α2 loop (Mans et al., 2008a; Jore et al., 2016). This would suggest multi-functionalization and escape from adaptive conflict by specialization of LTB4 and TXA2 binding. Recent research suggests that targeting of the complement cascade, neutrophils and platelet aggregation are closely interlinked (Deppermann and Kubes, 2016). Multi-functional proteins that target related mechanisms may evolve, not because of adaptive selection, but rather because of neutral evolution resulting in redundancy ultimately leading to beneficial mutations.

WHOLE GENOME DUPLICATION IN TICKS

Gene duplication may occur for individual genes in a stochastic manner across the genome. On the other hand, large-scale gene duplication events that include chromosome or whole genome duplications give rise to many gene duplicates at once and are an important source for the generation of gene copies and as a catalyst for adaptation to new environments and speciation (Nei, 2005; Van de Peer et al., 2009). In ticks, the extensive duplication events observed in transcriptomes were suggested to be due to possible genome duplications in hard ticks (Ribeiro et al., 2006; Dai et al., 2012). This may be partly supported by the increase in chromosome numbers observed in ticks when compared to other parasitiform mites, with mesostigmatid mites having 2- to 3-fold fewer chromosomes ($2n = 8 \pm 0.3$), compared to soft ($2n = 22 \pm 4$) and hard ($2n = 22 \pm 2$) ticks (Oliver, 1977; Mans and Neitz, 2004a). Analysis of the *I. scapularis* genome found no evidence for a genome duplication event (Van Zee et al., 2016). However, the analysis focused on recent gene duplication events dated <6 MYA. Given the similarities in chromosome numbers of hard and soft ticks, genome duplication in ticks may have occurred in the ancestral tick lineage (>290 MYA) and given that the majority of duplicated genes go extinct, may not be easily detectable since detection of whole genome duplication becomes increasingly difficult with age (Vanneste et al., 2013). The major contributor to genome size in ticks are repetitive elements with

FIGURE 3 | Evolutionary scenarios in the reconstruction of biogenic amine binding function. Indicated are the phylogenetic relationships of various tick genera as determined with mitochondrial analysis (Mans et al., 2012). Terminal nodes represent extant lipocalins for which empirical histamine or serotonin binding data exist. The number of lipocalins for which experimental data have been generated are indicated as cups (lipocalin~calyx~cup) that bind histamine in the upper binding site and serotonin and/or histamine in the lower binding pocket. Names of characterized lipocalins are indicated below their genus names. Internal nodes represent two possible ancestral binding site hypotheses: **(A)** The ancestral BABP had a single lower serotonin-binding site, or **(B)** the ancestral BABP had an upper histamine and lower histamine/serotonin-binding site.

66% contributing to the genome size of *I. scapularis* and 69% for *R. microplus* (Ullmann et al., 2005). It is not yet clear what proportion of the expanded genome size of *I. ricinus* may be attributed to repetitive elements (Cramaro et al., 2017), but it is likely that most of the genome size differences observed between tick species are not due to gene duplications, but rather the universal abundance of repetitive elements and non-coding DNA (Mans et al., 2016). Larger genome sizes of ticks compared to other mites is not evidence of genome duplication *per se* (Geraci et al., 2007). We may tentatively propose that chromosome numbers suggest a whole genome duplication in the ancestral lineage of ticks and this may have been associated with adaptation to blood-feeding. Subsequent divergence and stochastic gene duplication may explain more recent differences observed in tick transcriptomes.

CONCLUSIONS

Gene duplication remains a major factor in the evolution of blood-feeding behavior of ticks, even if most new functions and the evolution of secretory families in ticks may undergo neutral evolution. Hypotheses on adaptive evolution should first reject the null hypothesis that evolution did not occur via neutral evolution. More rigorous analysis is also necessary to characterize paralogous gene families, especially where molecular dating is concerned. Analysis of the literature indicates that our understanding of functional evolution is still limited, since few studies aim to reconstruct evolutionary pathways. Even so, the increase in taxon sampling of salivary gland transcriptomes holds the promise that we will be able to get an accurate estimate of the number of orthologous clades distributed across various tick lineages and the number of functions encoded in the tick sialoverse.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Anatriello, E., Ribeiro, J. M., de Miranda-Santos, I. K., Brandão, L. G., Anderson, J. M., Valenzuela, J. G., et al. (2010). An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. *BMC Genomics* 11:450. doi: 10.1186/1471-2164-11-450

Beaufays, J., Adam, B., Decrem, Y., Prévôt, P. P., Santini, S., Brasseur, R., et al. (2008). *Ixodes ricinus* tick lipocalins: identification, cloning, phylogenetic analysis and biochemical characterization. *PLoS ONE* 3:e3941. doi: 10.1371/journal.pone.0003941

Brieuc, M. S., and Naish, K. A. (2011). Detecting signatures of positive selection in partial sequences generated on a large scale: pitfalls, procedures and resources. *Mol. Ecol. Res.* 11(Suppl. 1), 172–183. doi: 10.1111/j.1755-0998.2010.02948.x

Cheng, P. H., Mans, B. J., Neitz, A. W., and Gaspar, A. R. (2010). Savicalin, a lipocalin from hemocytes of the soft tick, *Ornithodoros savignyi*. *Exp. Appl. Acarol.* 52, 313–326. doi: 10.1007/s10493-010-9368-6

Chmelař, J., Anderson, J. M., Mu, J., Jochim, R. C., Valenzuela, J. G., and Kopecký, J. (2008). Insight into the sialome of the castor bean tick, *Ixodes ricinus*. *BMC Genomics* 9:233. doi: 10.1186/1471-2164-9-233

Chmelař, J., Kotál, J., Kopecký, J., Pedra, J. H., and Kotsyfakis, M. (2016). All for one and one for all on the tick-host battlefield. *Trends Parasitol.* 32, 368–377. doi: 10.1016/j.pt.2016.01.004

Conway, E. M. (2015). Reincarnation of ancient links between coagulation and complement. *J. Thromb. Haemost.* 13, S121–S132. doi: 10.1111/jth.12950

Couvreur, B., Beaufays, J., Charon, C., Lahaye, K., Gensale, F., Denis, V., et al. (2008). Variability and action mechanism of a family of anticomplement proteins in *Ixodes ricinus*. *PLoS ONE* 3:e1400. doi: 10.1371/journal.pone.0001400

Cramaro, W. J., Hunewald, O. E., Bell-Sakyi, L., and Muller, C. P. (2017). Genome scaffolding and annotation for the pathogen vector *Ixodes ricinus* by ultra-long single molecule sequencing. *Parasit. Vectors* 10:71. doi: 10.1186/s13071-017-2008-9

Cramaro, W. J., Revets, D., Hunewald, O. E., Sinner, R., Reye, A. L., and Muller, C. P. (2015). Integration of *Ixodes ricinus* genome sequencing with transcriptome and proteome annotation of the naïve midgut. *BMC Genomics* 16:817. doi: 10.1186/s12864-015-1981-7

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., Georgin, 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

Delvaeye, M., and Conway, E. M. (2009). Coagulation and innate immune responses: can we view them separately? *Blood* 114, 2367–2374. doi: 10.1182/blood-2009-05-199208

Deppermann, C., and Kubes, P. (2016). Platelets and infection. *Semin. Immunol.* 28, 536–545. doi: 10.1016/j.smim.2016.10.005

Des Marais, D. L., and Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454, 762–765. doi: 10.1038/nature07092

Díaz-Martín, V., Manzano-Román, R., Oleaga, A., Encinas-Grandes, A., and Pérez-Sánchez, R. (2013). Cloning and characterization of aplasminogen-binding enolase from the saliva of the argasid tick *Ornithodoros moubata*. *Vet. Parasitol.* 191, 214–301. doi: 10.1016/j.vetpar.2012.09.019

Díaz-Martín, V., Manzano-Román, R., Siles-Lucas, M., Oleaga, A., and Pérez-Sánchez, R. (2011). Cloning, characterization and diagnostic performance of the salivary lipocalin protein TSGP1 from *Ornithodoros moubata*. *Vet. Parasitol.* 178, 163–172. doi: 10.1016/j.vetpar.2010.12.014

Dong, X., Armstrong, S. D., Xia, D., Makepeace, B. L., Darby, A. C., and Kadowaki, T. (2017). Draft genome of the honey bee ectoparasitic mite, *Tropilaelaps mercedesae*, is shaped by the parasitic life history. *Gigascience* 6, 1–17. doi: 10.1093/gigascience/gix008

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systemat. Zool.* 19, 99–106. doi: 10.2307/2412448

Fitch, W. M. (2000). Homology a personal view on some of the problems. *Trends Genet.* 16, 227–231. doi: 10.1016/S0168-9525(00)02005-9

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

Francino, M. P. (2005). An adaptive radiation model for the origin of new gene functions. *Nat. Genet.* 37, 573–577. doi: 10.1038/ng1579

Francischetti, I. M., Anderson, J. M., Manoukis, N., Pham, V. M., and Ribeiro, J. M. (2011). An insight into the sialotranscriptome and proteome of the coarse bontlegged tick, *Hyalomma marginatum* rufipes. *J. Proteomics* 74, 2892–2908. doi: 10.1016/j.jprot.2011.07.015

Francischetti, I. M., Mans, B. J., Meng, Z., Gudderra, N., Veenstra, T. D., Pham, V. M., et al. (2008a). An insight into the sialome of the soft tick, *Ornithodorus parkeri*. *Insect Biochem. Mol. Biol.* 38, 1–21. doi: 10.1016/j.ibmb.2007.09.009

Francischetti, I. M., Mather, T. N., and Ribeiro, J. M. (2004). Penthalaris, a novel recombinant five-Kunitz tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick vector of Lyme disease, *Ixodes scapularis*. *Thromb. Haemost.* 91, 886–898. doi: 10.1160/TH03-11-0715

Francischetti, I. M., Meng, Z., Mans, B. J., Gudderra, N., Hall, M., Veenstra, T. D., et al. (2008b). An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. *J. Proteomics* 71, 493–512. doi: 10.1016/j.jprot.2008.07.006

Francischetti, I. M., My Pham, V., Mans, B. J., Andersen, J. F., Mather, T. N., Lane, R. S., et al. (2005). The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). *Insect Biochem. Mol. Biol.* 35, 1142–1161. doi: 10.1016/j.ibmb.2005.05.007

Francischetti, I. M., Sa-Nunes, A., Mans, B. J., Santos, I. M., and Ribeiro, J. M. (2009). The role of saliva in tick feeding. *Front. Biosci.* 14:3363. doi: 10.2741/3363

Gabaldón, T., and Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14, 360–366. doi: 10.1038/nrg3456

Garcia, G. R., Gardinassi, L. G., Ribeiro, J. M., Anatriello, E., Ferreira, B. R., Moreira, H. N., et al. (2014). The sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma cajennense* ticks, uncovered by 454-based RNA-seq. *Parasit. Vectors* 7:430. doi: 10.1186/1756-3305-7-430

Geraci, N. S., Johnston, S. J., Robinson, P. J., Wikel, S. K., and Hill, C. A. (2007). Variation in genome size of argasid and ixodid ticks. *Insect Biochem. Mol. Biol.* 37, 399–408. doi: 10.1016/j.ibmb.2006.12.007

Gomez-Navarro, N., and Miller, E. (2016). Protein sorting at the ER-Golgi interface. *J. Cell Biol.* 215, 769–778. doi: 10.1083/jcb.201610031

Guglielmone, A. A., Robbins, R. G., Apanaskevich, D. A., Petney, T. N., Estrada-Pe-a, A., Horak, I. G., et al. (2010). The argasidae, ixodidae and nuttalliellidae (Acari: Ixodida) of the world: a list of valid species names. *Zootaxa* 2528, 1–28. doi: 10.5281/zenodo.196488

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

Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., et al. (2016). Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized hox genes and superdynamic intron evolution. *Genome Biol. Evol.* 8, 1762–1775. doi: 10.1093/gbe/evw048

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci. Lond. B* 256, 119–124. doi: 10.1098/rspb.1994.0058

Ibelli, A. M., Hermance, M. M., Kim, T. K., Gonzalez, C. L., and Mulenga, A. (2013). Bioinformatics and expression analyses of the *Ixodes scapularis* tick cystatin family. *Exp. Appl. Acarol.* 60, 41–53. doi: 10.1007/s10493-012-9613-2

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689

Jensen, R. A. (2001). Orthologs and paralogs - we need to get it right. *Genome Biol.* 2, 1002.1–1002.3. doi: 10.1186/gb-2001-2-8-interactions1002

Jore, M. M., Johnson, S., Sheppard, D., Barber, N. M., Li, Y. I., Nunn, M. A., et al. (2016). Structural basis for therapeutic inhibition of complement C5. *Nat. Struct. Mol. Biol.* 23, 378–386. doi: 10.1038/nsmb.3196

Karim, S., Singh, P., and Ribeiro, J. M. (2011). A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS ONE* 6:e28525. doi: 10.1371/journal.pone.0028525

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* 279, 5048–5057. doi: 10.1098/rspb.2012.1108

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biol.* 3, research0008.1–0008.9. doi: 10.1186/gb-2002-3-2-research0008

Konnai, S., Nishikado, H., Yamada, S., Imamura, S., Ito, T., Onuma, M., et al. (2011). Molecular identification and expression analysis of lipocalins from blood feeding taiga tick, *Ixodes persulcatus* Schulze. *Exp. Parasitol.* 127, 467–474. doi: 10.1016/j.exppara.2010.10.002

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–338. doi: 10.1146/annurev.genet.39.073003.114725

Koonin, E. V. (2011). Are there laws of genome evolution? *PLoS Comput. Biol.* 7:e1002173. doi: 10.1371/journal.pcbi.1002173

Koonin, E. V. (2016). Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol.* 14:3382. doi: 10.1186/s12915-016-0338-2

Koonin, E. V., Wolf, Y. I., and Karev, G. P. (2002). The structure of the protein universe and genome evolution. *Nature* 420, 218–223. doi: 10.1038/nature01256

Kotsyfakis, M., Kopáček, P., Franta, Z., Pedra, J. H., and Ribeiro, J. M. (2015a). Deep sequencing analysis of the *Ixodes ricinus* haemocytome. *PLoS Negl. Trop. Dis.* 9:e0003754. doi: 10.1371/journal.pntd.0003754

Kotsyfakis, M. S., Nunes, A., Francischetti, I. M., Mather, T. N., Andersen, J. F., Ribeiro, J. M. et al. (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

Kotsyfakis, M., Schwarz, A., Erhart, J., and Ribeiro, J. M. (2015b). Tissue- and time-dependent transcription in Ixodes ricinus salivary glands and midguts when blood feeding on the vertebrate host. *Sci. Rep.* 5:9103. doi: 10.1038/srep09103

Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* 6, 654–662. doi: 10.1038/nrg1659

Louw, E., van der Merwe, N. A., Neitz, A. W., and Maritz-Olivier, C. (2013). Evolution of the tissue factor pathway inhibitor-like Kunitz domain-containing protein family in *Rhipicephalus microplus*. *Int. J. Parasitol.* 43, 81–94. doi: 10.1016/j.ijpara.2012.11.006

Lynch, M. (2002). Genomics. Gene duplication and evolution. *Science* 297, 945–947. doi: 10.1126/science.1075472

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8597–8604. doi: 10.1073/pnas.0702207104

Lynch, M., Ackerman, M. S., Gout, J. F., Long, H., Sung, W., Thomas, W. K., et al. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* 17, 704–714. doi: 10.1038/nrg.2016.104

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473. Available online at: http://www.genetics.org/content/154/1/459

Lynch, M., and Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15690–15695. doi: 10.1073/pnas.1514974112

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. (2014). "Heme processing and the evolution of hematophagy," in *Biology of Ticks, Vol. 1, 2nd Edn.*, eds D. E. Sonenshine and R. M. Roe (New York, NY: Oxford University Press), 220–239.

Mans, B. J. (2016). "Glandular matrices and secretions: blood- feeding arthropods," in *Extracellular Composite Matrices in Arthropods*, eds E. Cohen and B. Moussian (Switzerland: S*pringer*), 625–688.

Mans, B. J., Anantharaman, V., Aravind, L., and Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3, 1612–1637. doi: 10.4161/cc.3.12.1316

Mans, B. J., and Neitz, A. W. (2004a). Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem. Mol. Biol.* 34, 1–17. doi: 10.1016/j.ibmb.2003.09.002

Mans, B. J., and Neitz, A. W. (2004b). Exon-intron structure of outlier tick lipocalins indicate a monophyletic origin within the larger lipocalin family. *Insect Biochem. Mol. Biol.* 34, 585–594. doi: 10.1016/j.ibmb.2004.03.006

Mans, B. J., and Ribeiro, J. M. (2008a). A novel clade of cysteinyl leukotriene scavengers in soft ticks. *Insect Biochem. Mol. Biol.* 38, 862–870. doi: 10.1016/j.ibmb.2008.06.002

Mans, B. J., and Ribeiro, J. M. (2008b). Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. *Insect Biochem. Mol. Biol.* 38, 841–852. doi: 10.1016/j.ibmb.2008.06.007

Mans, B. J., Andersen, J. F., Francischetti, I. M., Valenzuela, J. G., Schwan, T. G., Pham,V. M., et al. (2008a). Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. *Insect Biochem. Mol. Biol.* 38, 42–58. doi: 10.1016/j.ibmb.2007.09.003

Mans, B. J., Andersen, J. F., Schwan, T. G., and Ribeiro, J. M. (2008b). Characterization of anti-hemostatic factors in the argasid, *Argas monolakensis*: implications for the evolution of blood-feeding in the soft tick family. *Insect Biochem. Mol. Biol.* 38, 22–41. doi: 10.1016/j.ibmb.2007.09.002

Mans, B. J., de Castro, M. H., Pienaar, R., de Klerk, D., Gaven, P., et al. (2016). Ancestral reconstruction of tick lineages. *Ticks Tick Borne Dis.* 7, 509–535. doi: 10.1016/j.ttbdis.2016.02.002

Mans, B. J., de Klerk, D., Pienaar, R., and Latif, A. A. (2011). *Nuttalliella namaqua*: a living fossil and closest relative to the ancestral tick lineage: implications for the evolution of blood-feeding in ticks. *PLoS ONE* 6:e23675. doi: 10.1371/journal.pone.0023675

Mans, B. J., de Klerk, D., Pienaar, R., de Castro, M. H., and Latif, A. A. (2012). The mitochondrial genomes of *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae) and *Argas africolumbae* (Ixodoidea: Argasidae): estimation of divergence dates for the major tick lineages and reconstruction of ancestral blood-feeding characters. *PLoS ONE* 7:e49461. doi: 10.1371/journal.pone.0049461

Mans, B. J., de Klerk, D., Pienaar, R., de Castro, M. H., and Latif, A. A. (2015). Next-generation sequencing as means to retrieve tick systematic markers, with the focus on *Nuttalliella namaqua* (Ixodoidea: Nuttalliellidae). *Ticks Tick Borne Dis.* 6, 450–462. doi: 10.1016/j.ttbdis.2015.03.013

Mans, B. J., Louw, A. I., and Neitz, A. W. (2002). Evolution of hematophagy in ticks: common origins for blood coagulation and platelet aggregation inhibitors from soft ticks of the genus *Ornithodoros*. *Mol. Biol. Evol.* 19, 1695–1705. doi: 10.1093/oxfordjournals.molbev.a003992

Mans, B. J., Louw, A. I., and Neitz, A. W. (2003a). The influence of tick behavior, biotope and host specificity on concerted evolution of the platelet aggregation inhibitor savignygrin, from the soft tick *Ornithodoros savignyi*. *Insect Biochem. Mol. Biol.* 33, 623–629. doi: 10.1016/S0965-1748(03)00047-X

Mans, B. J., Louw, A. I., and Neitz, A. W. (2003b). The major tick salivary gland proteins and toxins from the soft tick, *Ornithodoros savignyi*, are part of the tick Lipocalin family: implications for the origins of tick toxicoses. *Mol. Biol. Evol.* 20, 1158–11567. doi: 10.1093/molbev/msg126

Mans, B. J., Ribeiro, J. M., and Andersen, J. F. (2008c). Structure, function, and evolution of biogenic amine-binding proteins in soft ticks. *J. Biol. Chem.* 283, 18721–18733. doi: 10.1074/jbc.M800188200

Manzano-Román, R., Díaz-Martín, V., Oleaga, A., Obolo-Mvoulouga, P., and Pérez-Sánchez, R. (2016). TSGP4 from *Ornithodoros moubata*: molecular cloning, phylogenetic analysis and vaccine efficacy of a new member of the lipocalin clade of cysteinyl leukotriene scavengers. *Vet. Parasitol.* 227, 130–137. doi: 10.1016/j.vetpar.2016.08.005

Mudge, J. M., and Harrow, J. (2016). The state of play in higher eukaryote gene annotation. *Nat. Rev. Genet.* 17, 758–772. doi: 10.1038/nrg.2016.119

Mulenga, A., Khumthong, R., and Blandon, M. A. (2007). Molecular and expression analysis of a family of the *Amblyomma americanum* tick Lospins. *J. Exp. Biol.* 210, 3188–3198. doi: 10.1242/jeb.006494

Mulenga, A., Khumthong, R., and Chalaire, K. C. (2009). *Ixodes scapularis* tick serine proteinase inhibitor (serpin) gene family: annotation and transcriptional analysis. *BMC Genomics* 10:217. doi: 10.1186/1471-2164-10-217

Mulenga,. A., Kim,. T. K., and Ibelli, A. M. (2013). Deorphanization and target validation of cross-tick species conserved novel *Amblyomma americanum* tick saliva protein. *Int. J. Parasitol.* 43, 439–451. doi: 10.1016/j.ijpara.2012.12.012

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* 22, 2318–23142. doi: 10.1093/molbev/msi242

Nei, M., and Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. doi: 10.1146/annurev.genet.39.073003.112240

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot. Sci.* 10, 1–6. doi: 10.1093/protein/10.1.1

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420

Nielsen, R. (2009). Adaptionism-30 years after Gould and Lewontin. *Evolution* 63, 2487–2490. doi: 10.1111/j.1558-5646.2009.00799.x

Nielsen, R., and Hubisz, M. J. (2005). Evolutionary genomics: detecting selection needs comparative data. *Nature* 433:E6. doi: 10.1038/nature03222

Oliver, J. H. Jr. (1977). Cytogenetics of mites and ticks. *Annu. Rev. Entomol.* 22, 407–429. doi: 10.1146/annurev.en.22.010177.002203

Paesen, G. C., Adams, P. L., Harlos, K., Nuttall, P. A., and Stuart, D. I. (1999). Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Mol. Cell* 3, 661–671. doi: 10.1016/S1097-2765(00)80359-7

Paesen, G. C., Adams, P. L., Nuttall, P. A., and Stuart, D. L. (2000). Tick histamine-binding proteins: lipocalins with a second binding cavity. *Biochim. Biophys. Acta* 1482, 92–101. doi: 10.1016/S0167-4838(00)00168-0

Paesen, G. C., Siebold, C., Dallas, M. L., Peers, C., Harlos, K., Nuttall, P. A., et al. (2009). An ion-channel modulator from the saliva of the brown ear tick has a highly modified Kunitz/BPTI structure. *J. Mol. Biol.* 389, 734–747. doi: 10.1016/j.jmb.2009.04.045

Perner, J., Provazník, J., Schrenková, J., Urbanová, V., Ribeiro, J. M., and Kopáček, P. (2016). RNA-seq analyses of the midgut from blood- and serum-fed Ixodes ricinus ticks. *Sci. Rep.* 6:36695. doi: 10.1038/srep36695

Porter, L., Radulović, Ž., Kim, T., Braz, G. R., Da Silva Vaz, I. Jr. and Mulenga, A. (2015). Bioinformatic analyses of male and female *Amblyomma americanum* tick expressed serine protease inhibitors (serpins). *Ticks Tick Borne Dis.* 6, 16–30. doi: 10.1016/j.ttbdis.2014.08.002

Preston, S. G., Majtán, J., Kouremenou, C., Rysnik, O., Burger, L. F., Cabezas Cruz, A., et al. (2013). Novel immunomodulators from hard ticks selectively reprogramme human dendritic cell responses. *PLoS Pathog.* 9:e1003450. doi: 10.1371/journal.ppat.1003450

Proulx, S. R., and Phillips, P. C. (2006). Allelic divergence precedes and promotes gene duplication. *Evolution* 60, 881–892. doi: 10.1111/j.0014-3820.2006.tb01168.x

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., Anderson, J. M., Manoukis, N. C., Meng, Z., and Francischetti, I. M. (2011). A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. *BMC Genomics* 12:136. doi: 10.1186/1471-2164-12-136

Ribeiro, J. M., Labruna, M. B., Mans, B. J., Maruyama, S. R., Francischetti, I. M., Barizon, G. C., et al. (2012). The sialotranscriptome of *Antricola delacruzi* female ticks is compatible with non-hematophagous behavior and an alternative source of food. *Insect Biochem. Mol. Biol.* 42, 332–342. doi: 10.1016/j.ibmb.2012.01.003

Ribeiro, J. M., Mans, B. J., and Arcà, 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., Slovák, M., and Francischetti, I. M. (2017). An insight into the sialome of *Hyalomma excavatum*. *Ticks Tick Borne Dis.* 8, 201–207. doi: 10.1016/j.ttbdis.2016.08.011

Rodriguez-Valle, M., Moolhuijzen, P., Piper, E. K., Weiss, O., Vance, M., Bellgard, M., et al. (2013). *Rhipicephalus microplus* lipocalins (LRMs): genomic identification and analysis of the bovine immune response using *in silico* predicted B and T cell epitopes. *Int. J. Parasitol.* 43, 739–752. doi: 10.1016/j.ijpara.2013.04.005

Rogozin, I. B., Managadze, D., Shabalina, S. A., and Koonin, E. V. (2014). Gene family level comparative analysis of gene expression in mammals validates the ortholog conjecture. *Genome Biol. Evol.* 6, 754–762. doi: 10.1093/gbe/evu051

Sangamnatdej, S., Paesen, G. C., Slovak, M., and Nuttall, P. A. (2002). A high affinity serotonin- and histamine-binding lipocalin from tick saliva. *Insect Mol. Biol.* 11, 79–86. doi: 10.1046/j.0962-1075.2001.00311.x

Schwarz, A., Cabezas-Cruz, A., Kopecký, J., and Valdés, J. J. (2014a). Understanding the evolutionary structural variability and target specificity of tick salivary Kunitz peptides using next generation transcriptome data. *BMC Evol. Biol.* 14:4. doi: 10.1186/1471-2148-14-4

Schwarz, A., Tenzer, S., Hackenberg, M., Erhart, J., Gerhold-Ay, A., Mazur, J., et al. (2014b). A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. *Mol. Cell Proteomic* 13, 2725–2735. doi: 10.1074/mcp.M114.039289

Schwarz, A., von Reumont, B. M., Erhart, J., Chagas, A. C., Ribeiro, J. M., and Kotsyfakis, M. (2013). *De novo Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *FASEB J.* 27, 4656–4745. doi: 10.1096/fj.13-232140

Shih, P. M., and Matzke, N. J. (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci U.S.A.* 110, 12355–12360. doi: 10.1073/pnas.1305813110

Sonnhammer, E. L., and Koonin, E. V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* 18, 619–620. doi: 10.1016/S0168-9525(02)02793-2

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

Tirloni, L., Seixas, A., Mulenga, A., Vaz Ida, S. Jr., and Termignoni, C. (2014). A family of serine protease inhibitors (serpins) in the cattle tick *Rhipicephalus (Boophilus) microplus*. *Exp. Parasitol.* 137, 25–34. doi: 10.1016/j.exppara.2013.12.001

Tyson, K. R., Elkins, C., and de Silva, A. M. (2008). A novel mechanism of complement inhibition unmasked by a tick salivary protein that binds to properdin. *J. Immunol.* 180, 3964–3968. doi: 10.4049/jimmunol.180.6.3964

Tyson, K., Elkins, C., Patterson, H., Fikrig, E., and de Silva, A. (2007). Biochemical and functional characterization of Salp20, an *Ixodes scapularis* tick salivary protein that inhibits the complement pathway. *Insect Mol. Biol.* 16, 469–479. doi: 10.1111/j.1365-2583.2007.00742.x

Ullmann, A. J., Lima, C. M., Guerrero, F. D., Piesman, J., and Black, V. I., W.C. (2005). Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. *Insect Mol. Biol.* 14, 217–222. doi: 10.1111/j.1365-2583.2005.00551.x

Valdés, J. J., and Moal, I. H. (2014). Prediction of Kunitz ion channel effectors and protease inhibitors from the *Ixodes ricinus* sialome. *Ticks Tick Borne Dis.* 5, 947–950. doi: 10.1016/j.ttbdis.2014.07.016

Valdés, J. J., Cabezas-Cruz, A., Sima, R., Butterill, P. T., Ružek, D. and Nuttall, P. A. (2016). Substrate prediction of Ixodes ricinus salivary lipocalins differentially expressed during Borrelia afzelii infection. *Sci. Rep.* 6:32372. doi: 10.1038/srep32372

Valdés, J. J., Schwarz, A., Cabeza de Vaca, I., Calvo, E., Pedra, J. H., Guallar, V., et al. (2013). Tryptogalinin is a tick Kunitz serine protease inhibitor with a unique intrinsic disorder. *PLoS ONE* 8:e62562. doi: 10.1371/journal.pone.0062562

Valenzuela, J. G., Charlab, R., Mather, T. N., and Ribeiro, J. M. (2000). Purification, cloning, and expression of a novel salivary anticomplement protein from the tick, *Ixodes scapularis*. *J. Biol. Chem.* 275, 18717–18723. doi: 10.1074/jbc.M001486200

Valenzuela, J. G., Francischetti, I. M., Pham, V. M., Garfield, M. K., Mather, T. N., and Ribeiro, J. M. (2002). Exploring thesialome of the tick *Ixodes*

scapularis. *J. Exp. Biol.* 205, 2843–2864. Available online at: http://jeb.biologists. org/content/205/18/2843.short

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600

Van Zee, J. P., Schlueter, J. A., Schlueter, S., Dixon, P., Sierra, C. A., and Hill, C. A. (2016). Paralog analyses reveal gene duplication events and genes under positive selection in *Ixodes scapularis* and other ixodid ticks. *BMC Genomics* 17:2. doi: 10.1186/s12864-015-2350-2

Vanneste, K., Van de Peer, Y., and Maere, S. (2013). Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* 30, 177–190. doi: 10.1093/molbev/mss214

Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15, 496–503. doi: 10.1016/S0169-5347(00)01994-7

Zhaxybayeva, O. (2013). Anciently duplicated genes reduce uncertainty in molecular clock estimates. Proc. Natl. Acad. Sci. U.S.A. 110, 12168-12169. doi: 10.1073/pnas.1310930110