Sialomes and Mialomes: A Systems-Biology View of Tick Tissues and Tick-Host Interactions

Jindrich Chmelar  
*University of South Bohemia*

Jan Kotal  
*University of South Bohemia*

Shahid Karim  
*University of Southern Mississippi*

Petr Kopacek  
*Czech Academy of Science*

Ivo M.B. Francischetti  
*NIAID*

*See next page for additional authors*

Follow this and additional works at: [https://aquila.usm.edu/fac_pubs](https://aquila.usm.edu/fac_pubs)

**Recommended Citation**
Available at: [https://aquila.usm.edu/fac_pubs/17537](https://aquila.usm.edu/fac_pubs/17537)
Authors
Jindrich Chmelar, Jan Kotal, Shahid Karim, Petr Kopacek, Ivo M.B. Francischetti, Joao H.F. Pedra, and Michail Kotsyfakis

This article is available at The Aquila Digital Community: https://aquila.usm.edu/fac_pubs/17537
Sialomes and mialomes: a systems biology view of tick tissues and tick-host interactions

Jindřich Chmelař1,2, Jan Kotál1,3, Shahid Karim4, Petr Kopacek3, Ivo M.B. Francischetti5, Joao H.F. Pedra6, and Michail Kotsyfakis3

1Faculty of Science, University of South Bohemia in České Budějovice, Budweis, Czech Republic
2Department of Clinical Pathobiochemistry, Technische Universität Dresden, Dresden, Germany
3Institute of Parasitology, Biology Center, Czech Academy of Sciences, Budweis, Czech Republic
4Department of Biological Sciences, The University of Southern Mississippi, Hattiesburg, Mississippi, United States of America
5Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, USA
6Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

Abstract

Tick saliva facilitates tick feeding and infection of the host. Gene expression analysis of tick salivary glands and other tissues involved in host-pathogen interactions has revealed a wide range of bioactive tick proteins. Transcriptomic analysis has been a milestone in the field and has recently been enhanced by next generation sequencing. Furthermore, the application of quantitative proteomics to ticks with unknown genomes has provided deeper insights into the molecular mechanisms underlying tick hematophagy, pathogen transmission, and tick-host-pathogen interactions. Here we review current knowledge on the transcriptomics and proteomics of tick tissues from a systems biology perspective and discuss future challenges in the field.

Keywords

Sialomes; systems biology; next generation sequencing; tick-borne pathogens

Correspondence to: mich_kotsyfakis@yahoo.com.

Illumina: https://www.illumina.com/technology/next-generation-sequencing.html
PacBio: http://www.pacificbiosciences.com/
Nanopore sequencing: https://nanoporetech.com/applications/dna-nanopore-sequencing

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Ticks, hosts, and pathogens

Ticks are obligatory ectoparasitic blood feeders that parasitize reptiles, birds, and mammals. Ticks are medically important since they transmit a plethora of pathogenic agents that cause human diseases including anaplasmosis, ehrlichiosis, babesiosis, rickettsiosis, and others (http://www.cdc.gov/ticks/diseases/). Lyme borreliosis is a common tick-borne disease worldwide, while tick-borne encephalitis is a public health concern in Europe and Asia (http://ecdc.europa.eu/en/healthtopics/emerging_and_vector-borne_diseases/tick_borne_diseases/tick_borne_encephalitis/pages/index.aspx). Ticks are divided into two major groups: soft ticks (family Argasidae) and hard ticks (family Ixodidae), which differ in their life cycles and blood-feeding strategies [1,2] and, as a consequence, are exposed to different host homeostatic responses. Hemostasis and acute inflammation are common responses to both groups of ticks and form the basis of the host anti-tick response. Hard ticks, however, must also counteract chronic inflammatory responses and specific humoral and cellular immunity [3].

Bellum omnium contra omnes

Dynamic, multi-directional interactions occur between ticks, hosts, and transmitted pathogens in both the tick and host environments, affecting all three members (Figure 1). These can be regarded as a continuous bellum omnium contra omnes, or war of all against all. When a tick ingests host blood, hemoglobin is digested and detoxified in the tick gut [4-6] and proteases of host or pathogenic origin are neutralized [7]. Tick midgut proteins and cells interact with ingested tick-borne pathogens, which migrate via the midgut and haemocoel [8-10] to invade the salivary glands, proliferate, and acquire salivary proteins on their surface. For example, the midgut proteins TROSPA and Ixofin3D bind to Borrelia spirochetes and facilitate midgut colonization and subsequent pathogen transmission to the host [11,12]. Proteins of the Salp15-like multigenic family are produced in the tick salivary glands and bind to Borrelia spirochetes to modulate host immunity, thus facilitating infection of the host [13,14]. Tick saliva secreted into the host suppresses local host immune responses, primarily to enable blood acquisition; however, the resulting host immunosuppression facilitates host infection [15-18]. Since tick salivarysecretions are the main mediators of host immunosuppression or immunomodulation, salivary composition plays a critical role in tick-borne pathogen transmission and represents a major topic of interest to researchers in the field [3].

High-throughput approaches such as transcriptomics and proteomics have facilitated the systematic characterization of salivary composition and gene expression dynamics throughout tick feeding. Moreover, high-throughput technologies are useful for investigating the effects of other biological factors such as the sex or developmental stage of ticks or the presence/absence of pathogens in their tissues. Here we review and discuss the new high-throughput techniques used to study tick-host-pathogen interactions.

On the path to sialome analysis

Tick saliva research has steadily progressed over the last three decades (Figure 2). The known immunomodulatory properties of tick saliva or salivary gland extracts (SGE)
(recently reviewed by Kotál and colleagues [19]) has enabled the adoption of a “function to protein” approach (Figure 2A), in which crude tick saliva fractions or SGEs that retain the biological activity of the starting material have been purified and isolated [20,21]. However, in the best-case scenario, this approach requires large amounts of starting material and only leads to the identification of individual salivary proteins [22,23].

Early reverse genetics approaches (Figure 2B) enabled the search for specific genes by nucleic acid hybridization-based screening of cDNA libraries produced from tick salivary glands [24-26]. Protein-coding cDNAs of interest were cloned, overexpressed using various systems, and their function characterized in bioassays [27,28].

These two low-throughput approaches were subsequently supplemented and supplanted by the rapid development of high-throughput approaches. These have led to the discovery of a hugely diverse set of salivary and midgut proteins acting at the interface of pathogen transmission in both the vector and host (Figure 2C). The terms ‘sialome’ and ‘mialome’ (see Glossary) were introduced to describe projects that identified hundreds of transcribed genes in tick salivary glands and the midgut, respectively, which were then extensively annotated and catalogued [29-36]. As the tick research community started to embrace transcriptomics many sialomes were published and hundreds of sequences were disclosed in GenBank, which was a real breakthrough: for the first time, the complexity of gene regulation in tick salivary glands had been itemized. Similar to earlier reverse genetics strategies, genes of interest were then expressed as recombinant proteins and extensively characterized at the biochemical and biological levels [37-42].

The next important development in studying tick-host-pathogen interactions was the introduction of next generation sequencing (NGS; Figure 2D and Glossary). Compared to classical Sanger sequencing, NGS platforms such as Illumina or 454 provided unprecedented transcriptome coverage, making them pioneering tools for quantitative analysis of gene expression dynamics in different tick tissues (see below). Some transcriptomics projects were complemented with proteomics (Figures 2C and D). Early proteomic analysis of tick saliva employed Edman degradation protein sequencing; in most cases, individual SDS-PAGE gel bands were used for subsequent protein sequencing [29,43]. As transcriptomic coverage improved using NGS, more sensitive mass spectrometry (MS) methods such as liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) or reverse phase liquid chromatography followed by tandem mass spectrometry (RP-LC-MS/MS) (see Glossary) have simultaneously allowed more thorough and comprehensive analysis of protein expression dynamics [44-46]. This combination of transcriptomics and proteomics is referred to as a “systems biology approach” or systems-level analysis.

In the following sections we summarize the knowledge gained over 13 years of tick transcriptomics and proteomics research and discuss the questions that can now be answered using new high-throughput analyses. Each advance in available technology over time was significant in its own right. As the theme of this review attests, the newest tools are always heralded as the greatest advances. Due to space constraints, we focus on tick salivary glands, the midgut, and hemocytes.

*Trends Parasitol.* Author manuscript; available in PMC 2017 March 01.
Finding missing pieces of the puzzle

Early transcriptomics projects on *Ixodes* spp. [29-31,33,35] provided insights into the qualitative aspects of tick salivary gene expression. Hundreds of transcripts were identified and the mechanisms and trends in salivary protein evolution were described. This provided a foundation for building a comprehensive overview of the molecular interface between ticks, hosts, and transmitted pathogens, which was further facilitated by the functional characterization of the discovered transcripts and proteins. The first tick sialome work by Valenzuela and colleagues revealed various mechanisms of host immune system evasion mediated by the salivary secretions of *Ixodes scapularis* and the existence of multigenic protein families in secreted saliva [29]. These protein families were subsequently confirmed in other tick species [30,32,47,34-36]. The major groups of secreted proteins common in most tick species are described in Figure 3. Despite efforts to identify as many transcripts as possible using transcriptomics, new family members were still being identified using more specific approaches such as cDNA library screening with gene-specific probes or RT-PCR with degenerate primers [48,38]. It was clear, therefore, that coverage of genetic diversity using expensive Sanger sequencing alone was insufficient. Nevertheless, the annotated sequencing data supported the adoption of “proteomics informed by transcriptomics (PIT)”, i.e., the identification of salivary proteins (mainly by Edman degradation) based on the discovered transcript sequences. The main qualitative improvement brought by NGS was the ability to detect large numbers of novel transcripts due to its capacity to detect even weakly expressed genes by extensive transcriptomic coverage. NGS has thus provided a more complete picture of tick gene expression and its regulation, and in doing so confirmed the presence of major protein families across tick species, the most represented (in terms of diversity and transcription rate) being Kunitz-domain proteins, lipocalins, metalloproteases, and basic tail proteins [49-51]. These families contain tens to hundreds of members of varying sequence similarity. The initial discovery of multigenic families in early transcriptomic studies provided the basis for efforts to decipher the function of each individual family member at the protein level [52,17,53]. The higher transcriptomic coverage and greater insights into the expression dynamics of individual transcripts provided by NGS significantly improved our ability to predict the function and importance of individual family members in tick survival and pathogen transmission.

Evolutionary insights

High sequence similarity between members of multigenic salivary protein families suggests that they originate from common ancestors that underwent multiple gene duplication events and subsequent mutations with divergent evolution [2,54,55]. Recent NGS-based *Ixodes ricinus* transcriptome analysis provided convincing experimental support for this hypothesis [56]. The authors analyzed synonymous (Syn) and non-synonymous (NS) mutation rates for several multigenic families. The highest NS mutation rate was detected in genes coding for unknown, secreted, and immunity-related proteins, suggesting an accelerated mutation rate in these gene groups. Phylogenetic analysis across tick species showed that certain branches (clades) of the phylogenetic tree were preferentially occupied by family members originating from a single tick species [49,50], i.e., most gene duplications (and thus most proteins) in multigenic families are tick lineage specific, in accordance with previously
published studies [55]. Thanks to the much higher transcriptomic coverage by NGS and new computational and bioinformatics approaches, phylogenetic analyses provide more information about multigenic protein family evolution [57]. Furthermore, two NGS projects provided solid evidence that accelerated tick evolution is related to hematophagy. In the first, Ribeiro and colleagues showed that non-hematophagous adult *Antricola delacruzi* females displayed the highest divergence from other hematophagous species with respect to secreted salivary proteins [58]. In the second, a comparison of male and female *Rhipicephalus pulchellus* revealed a list of hematophagy-related genes, since males do not feed on blood [46]. These examples show that large transcriptomic datasets can provide the foundation for studying the evolution of hematophagy in arthropods.

**When more is better**

The main advantage of NGS-based transcriptomics, however, lies in its ability to quantitatively describe transcriptome dynamics. Gene discovery from classical Sanger sequencing and NGS projects are compared in Table 1. The most extensive Sanger sequencing analysis revealed 13,643 and 12,319 unique transcripts in mixed libraries from *Rhipicephalus microplus* [47] and *Amblyomma americanum* [59], respectively; the usual number of unique sequences in similar projects is around one thousand. NGS-based transcriptome projects, on the other hand, can produce hundreds of thousands of assembled contigs [50] and over 50,000 unique transcripts [60,46]. On average, around 16,000 unique transcripts have been discovered in NGS projects on ticks, around ten-fold higher than Sanger sequencing in terms of novel transcript identification and around 100-200 fold higher in terms of total contigs. The latter number is particularly important for the quantitative analysis of tick physiology: the unprecedented transcriptome coverage by NGS enables statistically reliable analysis of gene expression dynamics of secreted salivary proteins throughout the course of tick feeding, the comparison of tissue- and developmental stage-specific transcript accumulation, and metabolic pathway analysis [55, 56]. Furthermore, different physiological conditions can be compared, for example changes in transcriptional regulation in the presence or absence of tick-borne pathogens and the influence of different host species on gene regulation in different tick tissues (see below and Figure 4).

**Gene expression dynamics during tick feeding**

One of the most important questions in tick-host-pathogen interactions is how tick salivary gland gene expression contributes to host homeostasis, pathogen transmission, and disease. In 2006, Ribeiro and colleagues observed that there were 20 genes at least two-times over-represented than expected in the cDNA library originating from the salivary glands of adult *I. scapularis* females 18-24 hours after attachment [31]. These transcripts were collagen-like proteins, Kunitz-domain containing proteins, basic tail proteins, and several proteins of unknown function. Similarly, seven genes significantly differed from random in ticks 3-4 days after host feeding, but each transcript belonged to a different family. Interestingly, different members of the same protein family were expressed at different time points of tick feeding [31]. Similarly, individual collagen-like protein family members were preferentially transcribed at specific tick feeding phases in the first *I. ricinus* sialome project [35]. This time-dependent preferential gene expression was recently confirmed in an NGS
transcriptional analysis of metalloproteases, Kunitz-domain containing proteins, and lipocalins [56]. In the same study, an over ten-fold tick feeding time-dependent difference was observed for 1447 genes expressed in the salivary glands, of which 1135 encoded secreted proteins. The most represented protein families were Kunitz domain-containing proteins followed by Salp15/ixostatin family members, lipocalins, metalloproteases, and several novel protein family members of unknown function [56]. A very recent NGS study analyzed Amblyomma americanum salivary glands at four feeding time points and confirmed time-dependent preferential gene expression in another medically important tick [61]. The genetic mechanisms underpinning sequential expression of individual members of a multigene family are unknown. However, transposable elements and genes of viral origin are consistently being detected in sialomes, suggesting that there are active changes in the tick genomic structure [49,56]. Together with the observed influence of histone modifications on gene expression in Amblyomma maculatum [62], we can hypothesize that sequential transcriptional regulation is epigenetically regulated [56]. The benefit of shifts in sequential gene expression between different members of the same protein family is still only speculative but may reflect the need to evade immune recognition by the host.

**Tissue- and life stage-specific transcriptional regulation**

Quantitative transcriptomic analysis has also been performed in I. ricinus according to tissue (salivary glands vs. midgut) and developmental stage (nymphs vs. adults). Multiple pairwise comparisons between nymph and adult female midguts and salivary glands at different tick feeding time points revealed over 8300 genes with at least ten-fold differences in gene expression in salivary glands and midguts [56]. A systems biology analysis of I. ricinus salivary glands and midguts revealed some discrepancies between the transcriptomic and proteomic analyses [44]: of a total of 1510 genes expressed at both transcriptomic and proteomic levels in the specific tick tissues, 373 proteins were more abundant in the salivary glands than in the midgut, but only 110 of these displayed corresponding transcript accumulation in the same tissue. Conversely, 217 proteins were significantly upregulated in the tick midgut vs. salivary glands, but only 93 had a similar transcriptional pattern.

The authors explained this discrepancy by the delay between activation of the transcriptional and translational machinery or due to tissue-specific pre-synthesis (or secretion) of some tick proteins. More detailed analysis revealed that the majority of over-represented salivary gland proteins were secreted proteins or connected to the protein modification machinery, while over-represented midgut proteins were mainly metabolic [44]. While transcriptomics provides only putative (although often precise) information on salivary composition, proteomics reveals the actual composition, making these two approaches complementary.

The experimental design of transcriptomic projects largely depends on the question being asked. Therefore, we can expect future NGS transcriptomic or systems level analyses to focus on specific questions concerning individual genes, protein families, or physiological phenomena related to the tick lifecycle. For instance, Lewis and colleagues constructed and sequenced a cDNA library of tick immunogenes by phage display library screening of I. scapularis females fed for 24 hours with serum collected from rabbits sensitized by repeat exposure to ticks. They found 182 contigs that were considered immunogenic and usable for
potential vaccine development [63]. A similar approach was used to identify 895 potential immunogens in the salivary glands of *Amblyomma americanum* [64]. The influence of stress on unfed *Dermacentor reticulatus* gene expression was studied using a systems biology approach and revealed hundreds of stress-activated genes that could be potentially targeted to develop novel tick control methods [45].

In addition to the midgut and salivary glands, hemocytes also appear to be an important interface between the tick and pathogens. Characterization of the *I. ricinus* hemocyte transcriptome (hemocytome) revealed the existence of important immune-related proteins in these ancient phagocytic ancestors of mammalian leukocytes [65]. The identified transcripts encoded proteins similar to defensins, pattern recognition receptors, proteases, protease inhibitors, and others. However, in contrast to 8300 genes with more than 10-fold difference in expression in salivary glands or the midgut, only 327 were expressed at least five-times more in hemocytes compared to other tissues [65].

**Ticks versus pathogens**

The question of how pathogens affect tick gene expression has been addressed in several studies. In 2004, Nene and colleagues compared the sialomes of *Rhipicephalus appendiculatus* ticks with or without *Theileria parva* infection but did not find any significant differences in gene expression [66]. Ribeiro and colleagues identified ten differentially expressed contigs in *I. scapularis* nymph salivary glands with or without *Borrelia burgdorferi* infection [31] belonging to the 5.3kDa family, the basic tail protein family, and histamine-binding proteins in the lipocalin superfamily. Notably, some lipocalin genes were overexpressed in infected ticks and others in uninfected ticks. Jaworski and colleagues used 454 pyrosequencing to characterize the immune response of *Dermacentor variabilis* after injection with different bacterial species [67]. In whole body sample analyses, the authors identified over 30 immune responsive genes including genes encoding serpins, calreticulin, superoxide dismutase, galectin, and defensins. Interestingly, transcriptional upregulation in response to bacterial infection was only confirmed in seven genes by RT-PCR [67]. Similarly, 26 differentially expressed genes were identified by cDNA library subtraction and classical Sanger sequencing after infection of *Rhipicephalus microplus* with *Babesia bovis* [68]. The upregulated genes were mostly related to metabolism and tick immunity, indicating that *Babesia bovis* is a physiological burden in infected ticks. In contrast to the low number of genes identified in these studies, infection of *R. microplus* with *Anaplasma marginale* affected the expression of 888 midgut genes and 146 salivary gland genes, mostly of unknown function, using microarray analysis [69]. In another study, infection of *I. ricinus* with *Bartonella henselae* resulted in transcriptional upregulation of 829 genes and downregulation of 517 genes in salivary glands [70]. Similar to Ribeiro’s observation [31], these genes belonged to the same multigenic families, with certain members upregulated and others downregulated on infection. In the same study, Liu and colleagues revealed that IrSPI protein, a Kunitz protease inhibitor, was most upregulated after infection with *B. henselae*. IrSPI was shown to facilitate tick feeding and the proliferation of the pathogen in the tick salivary glands [70]. A very thorough NGS-based analysis focused on apoptotic pathway changes in *I. scapularis* infected with *Anaplasma phagocytophilum* [71]. The authors revealed pathogen-driven inhibition of apoptosis that

*Trends Parasitol*. Author manuscript; available in PMC 2017 March 01.
facilitated establishment of the pathogen via upregulation of Janus kinase/signal transducer and activator of transcription (JAK/STAT) signaling in the tick. This is an excellent example of how huge NGS datasets can be used to answer specific questions. Such analyses of specific metabolic and signaling pathways can be useful not only for disclosing differences in gene expression between two physiological states (e.g., non-infected vs. infected ticks), but also for detailed characterization of the molecular pathways that mediate tick physiology.

An NGS approach was also successfully used to analyze the spectrum of pathogens present in I. ricinus [72]. In total, 12 bacterial pathogens were identified including known pathogens like Borrelia afzelii, B. garini, B. burgdorferi s.s., Anaplasma phagocytophilum, but also new and unexpected species like Borrelia miyamotoi or Bartonella henselae, Bartonella graham, and Rickettsia felis. The same group used the same dataset to identify protozoan parasites [73]. Similarly, both known and unknown species infecting I. ricinus were revealed by data analysis. Carpi and colleagues combined I. ricinus 454 pyrosequencing data and cDNA libraries to identify over 100 bacterial genera, both pathogenic and symbiotic [74]. These projects highlight the potential of NGS for tick-borne pathogen identification. NGS has also been used several times in tick research to analyze symbiotic bacteria and the midgut microbiome, as recently reviewed by Narasimhan and Fikrig [75].

Concluding Remarks

Tick transcriptomics has undergone rapid and impressive progress due to technical developments in NGS and proteomics; this, in turn, has allowed not only the identification of tick transcripts but also the detailed analysis of transcriptional and translational dynamics. It is now feasible to describe the detailed proteome dynamics of tick salivary glands and midguts for over 1500 proteins in ticks with unknown genomes [44]. Therefore, the latest technical developments have resulted in at least an order of magnitude increase in the number of identified transcripts and proteins compared to early sialome and mialome projects. There is no doubt that there is room for further research in this field, regardless of whether the focus is on the evolution of the tripartite interaction between ticks, hosts, and tick-borne pathogens, the gene expression dynamics of hematophagy-related genes, or even the physiological or ecological aspects of tick biology. Furthermore, there remains an open question about the real-time changes occurring in tick tissues in response to gene and protein expression changes in the host skin and bite site draining lymph nodes. More importantly, there are still many topics in disease vector genomics that remain barely investigated, such as the role of epigenetics and non-coding genomic regions and non-coding RNAs in the tripartite interaction. Finally, we anticipate the development of research projects that employ single-cell/single molecule sequencing methodologies and do not require nucleic acid amplification. Furthermore, other ‘-omics’ approaches such as metabolomics may shed additional light on the physiology of this tripartite interaction. The increasing impact of vector-borne diseases in human and veterinary public health mandates the use of cutting-edge technologies to rapidly develop novel control methods and tools.
Acknowledgments

We thank Nextgenediting (www.nextgenediting.com) for providing editorial assistance. This work was supported by the Grant Agency of the Czech Republic (GACR grant P502/12/2409 to MK and grant 13-11043S to PK), the Czech Academy of Sciences (grant no. RV060077344 to the Biology Center—Institute of Parasitology), and the 7th Framework Program of the European Union (EU FP7; Marie Curie Reintegration grant PIRG07GA-2010-268177 to MK). This publication reflects only the authors’ views and the European Union is not liable for any use that may be made of the information contained herein. MK and JHFP were supported by the National Institutes of Health (RO1 AI093653) and JHFP was supported by start-up funds provided by the University of Maryland-Baltimore, School of Medicine. IMBF was supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH; USA). MK received support from the Czech Academy of Sciences (Jan Evangelista Purkyne Fellowship). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References


**Glossary**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig</td>
<td>Consensual DNA sequence assembled from several ESTs during bioinformatics analysis of transcriptomes. Contigs that encode proteins of the same family form clusters. The number of contigs per cluster is used in quantitative analyses, such as for describing gene expression dynamics or comparisons of gene expression at different tick developmental stages.</td>
</tr>
<tr>
<td>Edman degradation</td>
<td>Edman Sequencing is still the most robust and fastest approach to sequencing the N-terminus of peptides or proteins. Developed by Pehr Edman, in this method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues.</td>
</tr>
<tr>
<td><strong>Expressed sequence tag (EST)</strong></td>
<td>A single sequence read, usually used in the context of classical Sanger sequencing.</td>
</tr>
<tr>
<td>Hemocytome</td>
<td>A hemocyte transcriptome.</td>
</tr>
<tr>
<td>Mialome</td>
<td>A midgut transcriptome sometimes coupled with its proteomic analysis. Derived from sialome and midgut.</td>
</tr>
<tr>
<td>Next generation sequencing (NGS)</td>
<td>General term for a group of sequencing platforms that enable massive sequencing of RNA or DNA. Platforms commonly used in this field are 454 pyrosequencing (Roche), Illumina dye sequencing (Illumina, Solexa), and SOLiD sequencing (Life Technologies). Their strength lies in the production of huge numbers of short sequence reads (70-150 bp for Illumina, 200-500bp for 454) and powerful assembly software that allows the assembly of individual reads into contigs.</td>
</tr>
<tr>
<td>Proteomics informed by transcriptomics (PIT)</td>
<td>Protein identification based on a transcript sequence database.</td>
</tr>
<tr>
<td><strong>(Reverse phase) liquid chromatography followed by tandem mass spectrometry (RP)-LC-MS/MS</strong></td>
<td>A high-throughput proteomic analysis technique in which proteins are separated by liquid chromatography, followed by two rounds of mass spectrometry for detection.</td>
</tr>
<tr>
<td>Sanger sequencing</td>
<td>The classical sequencing method used in most biological applications prior to NGS development. Sanger sequencing is based on the incorporation of fluorescently labeled dideoxynucleotides that stop polymerase activity and label DNA fragments, which are subsequently subjected to capillary electrophoresis and laser detection of fluorescence. The advantage of this method is the resulting long reads; one read can cover up to 1000 base pairs.</td>
</tr>
<tr>
<td>Sialome</td>
<td>A salivary gland transcriptome, sometimes coupled with its proteome analysis, usually used in the context of hematophagous arthropods. Sialome is a composite of the Greek word for saliva (sialos) and ‘transcriptome’.</td>
</tr>
<tr>
<td>Unique transcript</td>
<td>A cDNA sequence originating from the reverse transcription of mRNA and coding for a single protein.</td>
</tr>
</tbody>
</table>
Outstanding Questions Box

- Will the new, emerging nucleotide sequence analysis methods improve assembly of transcriptome reads into contigs for organisms with an unknown genome by providing longer sequencing reads?

- What changes in knowledge can we expect from transcriptomic and proteomic analyses at the individual tick level, which are now becoming feasible due to the decreased amount of starting material needed in such analyses?

- Decreasing costs will boost the adoption of systems biology approaches and genome sequencing projects in this field. How much of an impact will this have on our understanding of the tick lifecycle and hematophagy in molecular and evolutionary terms?

- What real-time changes occur in tick tissues in response to gene and protein expression changes in the host skin and bite site draining lymph nodes?

- How soon will we be able to employ other ‘-omics’ approaches, such as metabolomics, to this field?
Trends Box

- The development of high-throughput next generation sequencing methodologies revolutionized research at the vector-pathogen-host interface.

- The decrease in sequencing costs and in the quantity of required starting material, as well as higher transcriptome coverage improved drastically our understanding of the gene expression regulation in tick tissues involved (or not) in pathogen transmission.

- High throughput quantitative proteomics are now feasible even for disease vectors with an unknown genome and have provided deeper insights into the molecular mechanisms underlying hematophagy, pathogen transmission, and tick-pathogen-host interactions.

- Emerging high-throughput gene sequencing methodologies and other ‘-omics’ methodologies, such as metabolomics, may soon be applied to this field.
Figure 1. The complex interactions between ticks, hosts, and tick-borne pathogens

In ticks, the toxic properties of hemoglobin and the deleterious activities of leukocyte-derived proteins are attenuated by protease inhibition, enzymatic digestion of hemoglobin, and toxic iron scavenging [7,4,76]. Ingested pathogens interact with midgut proteins [11,12] and host blood-derived factors before migrating through the tick midgut and hemolymph to interact with the tick's innate immune system [9,8]. Next, the pathogens migrate to the tick salivary glands, where they proliferate and acquire salivary proteins on their surface [13,14]. Pathogens are then injected into the host along with tick saliva, and tick salivary components begin to suppress the local host homeostatic response that is immediately raised against the tick bite-induced injury, the ‘foreign’ tick salivary antigens, and the tick-borne pathogens [77,16,19,3]. The overall outcome is facilitation of tick feeding and pathogen colonization of the tick, the host, or both.
Figure 2. A schematic of the four methodological pipelines for tick salivary protein identification
(A) Activity to protein identification approach. (B) cDNA library screening for specific genes of interest. (C) Random cDNA library sequencing combined with proteomics. (D) Direct RNA sequencing with next generation sequencing combined with advanced reverse phase liquid chromatography-tandem mass spectrometry ((RP)-LC-MS/MS) proteomics. See main text for details.
Figure 3. Major multigenic families identified by high-throughput transcriptomic analysis of tick salivary glands and their effect on host homeostasis
These families include serine protease inhibitors (serpins, Kunitz, trypsin inhibitor-like (TIL)-domain), cystatins, lipocalins (histamine-binding proteins), disintegrins, metalloproteases, and several novel protein families with little or no similarity to other proteins: basic tail, Salp15, and evasins. Many other multigenic families, although less abundant, are found in sialomes. Yellow-red rectangles represent the tick salivary contents.
Figure 4. Schematic of some comparisons that can be performed by next generation sequencing

(A) Comparisons were made between various tick developmental stages [51,56] and as a function of tick blood feeding progress [50,64,56,60,61]. Differences in gene expression regulation were analyzed in the presence (infected ticks shown with green color in the figure) or absence of pathogens in ticks [78,70,71]. Transcriptional regulation was also compared between male and female adult ticks [63] and between adult female ticks upon primary (feeding on naive host) or secondary infestation of the same host (the host was exposed to ticks for the second time) [46]. Comparisons of transcription regulation in tick salivary glands (SG), midgut (MG), and hemocytes (HC) (see inset) were performed at different tick developmental stages infected or non-infected with various tick-borne pathogens [44,65,56]. Other tick tissues analyzed by transcriptomics include synganglia and male and female reproductive organs; due to space restraints we do not discuss these projects further but details can be found in [79-84].

(B) The analysis of various tick-dwelling microorganisms (tick microbiome analysis and exploring the host species-dependent pathogen diversity in ticks) revealed a high number of bacteria and protozoa in ticks [72,73,75].
<table>
<thead>
<tr>
<th>Year</th>
<th>Sequencing Method</th>
<th>Tissue</th>
<th>Spices</th>
<th>Total Reads</th>
<th>Average Read Length</th>
<th>Good Quality EST</th>
<th>Unique Sequences</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Sanger</td>
<td>SG</td>
<td>Isodes scapularis</td>
<td>735</td>
<td>-</td>
<td>735</td>
<td>410</td>
<td>[29]</td>
</tr>
<tr>
<td>2004</td>
<td>Sanger</td>
<td>SG</td>
<td>Rhipicephalus appendiculatus</td>
<td>28 416</td>
<td>670/780</td>
<td>19 046</td>
<td>7 359</td>
<td>[66]</td>
</tr>
<tr>
<td>2004</td>
<td>Sanger</td>
<td>SG</td>
<td>Rhipicephalus micropus</td>
<td>-</td>
<td>-</td>
<td>324</td>
<td>188</td>
<td>[83]</td>
</tr>
<tr>
<td>2004</td>
<td>Sanger</td>
<td>HC</td>
<td>Rhipicephalus micropus</td>
<td>-</td>
<td>-</td>
<td>196</td>
<td>157</td>
<td>[83]</td>
</tr>
<tr>
<td>2005</td>
<td>Sanger</td>
<td>SG</td>
<td>Isodes pacificus</td>
<td>-</td>
<td>487</td>
<td>1 068</td>
<td>557</td>
<td>[30]</td>
</tr>
<tr>
<td>2006</td>
<td>Sanger</td>
<td>SG</td>
<td>Isodes scapularis</td>
<td>8 150</td>
<td>-</td>
<td>7 476</td>
<td>3 020</td>
<td>[31]</td>
</tr>
<tr>
<td>2007</td>
<td>Sanger</td>
<td>SG</td>
<td>Dermacentor andersoni</td>
<td>1 440</td>
<td>600</td>
<td>1 299</td>
<td>762</td>
<td>[32]</td>
</tr>
<tr>
<td>2007</td>
<td>Sanger</td>
<td>ML</td>
<td>Rhipicephalus micropus</td>
<td>42 512</td>
<td>-</td>
<td>-</td>
<td>13 643</td>
<td>[47]</td>
</tr>
<tr>
<td>2008</td>
<td>Sanger</td>
<td>SG</td>
<td>Amblyomma cajennense</td>
<td>1 920</td>
<td>472</td>
<td>1 754</td>
<td>1 234</td>
<td>[85]</td>
</tr>
<tr>
<td>2008</td>
<td>Sanger</td>
<td>MG</td>
<td>Dermacentor variabilis</td>
<td>2 304</td>
<td>-</td>
<td>1 679</td>
<td>835</td>
<td>[33]</td>
</tr>
<tr>
<td>2008</td>
<td>Sanger</td>
<td>SG</td>
<td>Isodes icinus</td>
<td>2 304</td>
<td>503</td>
<td>1 881</td>
<td>1 274</td>
<td>[35]</td>
</tr>
<tr>
<td>2008</td>
<td>Sanger</td>
<td>SG</td>
<td>Ornithodoros coriaeus</td>
<td>-</td>
<td>-</td>
<td>1 089</td>
<td>726</td>
<td>[43]</td>
</tr>
<tr>
<td>2008</td>
<td>Sanger</td>
<td>SG</td>
<td>Ornithodoros parkeri</td>
<td>-</td>
<td>-</td>
<td>1 529</td>
<td>649</td>
<td>[34]</td>
</tr>
<tr>
<td>2009</td>
<td>Sanger</td>
<td>SG</td>
<td>Amblyomma americanum</td>
<td>-</td>
<td>-</td>
<td>3 868</td>
<td>2 002</td>
<td>[36]</td>
</tr>
<tr>
<td>2010</td>
<td>Sanger</td>
<td>SG</td>
<td>Rhipicephalus sanguineus</td>
<td>-</td>
<td>-</td>
<td>2 034</td>
<td>1 024</td>
<td>[86]</td>
</tr>
<tr>
<td>2010</td>
<td>454</td>
<td>AF</td>
<td>Dermacentor variabilis</td>
<td>233 335</td>
<td>203</td>
<td>-</td>
<td>38 683</td>
<td>[67]</td>
</tr>
<tr>
<td>2011</td>
<td>Sanger</td>
<td>SG</td>
<td>Hyalomma marginatum niptes</td>
<td>-</td>
<td>-</td>
<td>2 084</td>
<td>1 167</td>
<td>[87]</td>
</tr>
<tr>
<td>2011</td>
<td>454</td>
<td>SG</td>
<td>Amblyomma maculatum</td>
<td>1 626 969</td>
<td>-</td>
<td>190 646</td>
<td>15 814</td>
<td>[49]</td>
</tr>
<tr>
<td>2011</td>
<td>454</td>
<td>AF</td>
<td>Isodes icinus</td>
<td>60 186</td>
<td>227</td>
<td>-</td>
<td>-</td>
<td>[74]</td>
</tr>
<tr>
<td>2012</td>
<td>Sanger</td>
<td>Larvae</td>
<td>Rhipicephalus micropus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>775</td>
<td>[68]</td>
</tr>
<tr>
<td>2012</td>
<td>Sanger</td>
<td>SG</td>
<td>Antricola delaeucruci</td>
<td>-</td>
<td>-</td>
<td>1 147</td>
<td>923</td>
<td>[58]</td>
</tr>
<tr>
<td>2013</td>
<td>Sanger</td>
<td>ML</td>
<td>Amblyomma americanum</td>
<td>20 256</td>
<td>-</td>
<td>15 390</td>
<td>12 319</td>
<td>[59]</td>
</tr>
<tr>
<td>2013</td>
<td>Sanger</td>
<td>MG</td>
<td>Rhipicephalus micropus</td>
<td>5 000</td>
<td>-</td>
<td>4 054</td>
<td>1 628</td>
<td>[78]</td>
</tr>
<tr>
<td>2013</td>
<td>Illumina</td>
<td>Nymphs</td>
<td>Isodes icinus</td>
<td>162 000 000</td>
<td>101</td>
<td>-</td>
<td>-</td>
<td>[72]</td>
</tr>
<tr>
<td>2013</td>
<td>454</td>
<td>SG</td>
<td>Isodes icinus</td>
<td>441 381</td>
<td>518</td>
<td>93 331</td>
<td>34 560</td>
<td>[50]</td>
</tr>
<tr>
<td>2013</td>
<td>Illumina</td>
<td>SG</td>
<td>Isodes icinus</td>
<td>67 703 183</td>
<td>90</td>
<td>269 600</td>
<td>34 560</td>
<td>[50]</td>
</tr>
<tr>
<td>Year</td>
<td>Sequencing Method</td>
<td>Tissue</td>
<td>Spices</td>
<td>Total Reads</td>
<td>Average Read Length</td>
<td>Good Quality EST</td>
<td>Unique Sequences</td>
<td>Refs</td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>--------</td>
<td>--------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>2014</td>
<td>Illumina</td>
<td>SG</td>
<td><em>Amblyomma americanum</em></td>
<td>18 800 000</td>
<td>-</td>
<td>-</td>
<td>17 593</td>
<td>[64]</td>
</tr>
<tr>
<td>2014</td>
<td>Illumina</td>
<td>Larvae+MG</td>
<td><em>Dermacentor reticulatus</em></td>
<td>21 677 414</td>
<td>201</td>
<td>18 946</td>
<td>3 808</td>
<td>[45]</td>
</tr>
<tr>
<td>2014</td>
<td>Illumina</td>
<td>SG+MG</td>
<td><em>Ixodes ricinus</em></td>
<td>585 000 000</td>
<td>-</td>
<td>-</td>
<td>198 504</td>
<td>[44]</td>
</tr>
<tr>
<td>2014</td>
<td>454</td>
<td>SG</td>
<td><em>Amblyomma cajennense</em></td>
<td>67 677</td>
<td>-</td>
<td>-</td>
<td>4 604</td>
<td>[51]</td>
</tr>
<tr>
<td>2014</td>
<td>454</td>
<td>SG</td>
<td><em>Amblyomma parvum</em></td>
<td>104 817</td>
<td>-</td>
<td>-</td>
<td>3 796</td>
<td>[51]</td>
</tr>
<tr>
<td>2014</td>
<td>454</td>
<td>SG</td>
<td><em>Amblyomma triste</em></td>
<td>442 756</td>
<td>-</td>
<td>-</td>
<td>1 124</td>
<td>[51]</td>
</tr>
<tr>
<td>2014</td>
<td>454</td>
<td>SG</td>
<td><em>Ixodes ricinus</em></td>
<td>778 598</td>
<td>379</td>
<td>-</td>
<td>24 539</td>
<td>[70]</td>
</tr>
<tr>
<td>2015</td>
<td>454</td>
<td>HC</td>
<td><em>Ixodes ricinus</em></td>
<td>926 596</td>
<td>498</td>
<td>-</td>
<td>15 716</td>
<td>[65]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>HC</td>
<td><em>Ixodes ricinus</em></td>
<td>49 328 982</td>
<td>148</td>
<td>-</td>
<td>15 716</td>
<td>[65]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>SG</td>
<td><em>Amblyomma americanum</em></td>
<td>344 909 378</td>
<td>101</td>
<td>-</td>
<td>5 792</td>
<td>[61]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>SG+MG</td>
<td><em>Ixodes ricinus</em></td>
<td>268 914 130</td>
<td>-</td>
<td>-</td>
<td>25 808</td>
<td>[56]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>SG</td>
<td><em>Ixodes scapularis</em></td>
<td>28 000 000</td>
<td>101</td>
<td>-</td>
<td>11 105</td>
<td>[71]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>MG</td>
<td><em>Ixodes scapularis</em></td>
<td>26 000 000</td>
<td>101</td>
<td>-</td>
<td>12 651</td>
<td>[71]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>Nymphs</td>
<td><em>Ixodes scapularis</em></td>
<td>31 000 000</td>
<td>101</td>
<td>-</td>
<td>16 083</td>
<td>[71]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>AF</td>
<td><em>Ixodes scapularis</em></td>
<td>3 700 000</td>
<td>-</td>
<td>-</td>
<td>9 134</td>
<td>[63]</td>
</tr>
<tr>
<td>2015</td>
<td>Illumina</td>
<td>SG</td>
<td><em>Rhipicephalus pulchellus</em></td>
<td>241 229 128</td>
<td>-</td>
<td>-</td>
<td>50 460</td>
<td>[46]</td>
</tr>
</tbody>
</table>

*a* AF – Adult female, whole body; SG - salivary glands; MG – midgut; ML - mixed library from several tissues; HC – hemocytes; EST – expressed sequence tag; “-” - not indicated in the study. The background color indicates the sequencing method: white – Sanger sequencing, light grey – 454 pyrosequencing, dark grey – Illumina sequencing.

*b* Sum of transcripts identified by both 454 and Illumina sequencing.