School of doctoral studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Comparative transcriptomics of *Ixodes ricinus* tick life stages

Ph.D. Thesis

RNDr. Pavlína Věchtová

Supervisor: Prof. RNDr. Libor Grubhoffer, CSc., Dr.h.c. mult.

University of South Bohemia, Faculty of Science
&
Biology Centre of the Academy of Science of the Czech Republic,
Institute of Parasitology

České Budějovice 2020

This thesis should be cited as:

Věchtová, P., 2020: Comparative transcriptomics of *Ixodes ricinus* tick life-stages. Ph. D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 165 pp.

Annotation

The proposed thesis describes stage-specific transcription of *Ixodes ricinus* tick. Based on the transcriptome assemblies of *I. ricinus* tick life stages this work describes processes typical for each *I. ricinus* life stage and infers the significance of methylation and glycosylation for ticks.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 22. 4. 2020

.....

RNDr. Pavlína Věchtová

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre CAS, supporting doctoral studies in the Molecular and Cell Biology and Genetics study programme.

Přírodovědecká fakulta
Faculty of Science

Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Parazitologický ústav, Biologické centrum AV ČR

Institute of Parasitology Biology Centre, AS CR

Financial support

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic project C4SYS (LM2015055), Czech-Bioimaging (LM2015062), CEITEC 2020 project (LQ1601), INTER-ACTION projects (LTARF18021, LTAUSA18040), ERD Funds (project CePaViP; CZ.02.1.01/0.0/0.0/16_019/0000759, Postdok_BIOGLOBE (CZ.1.07/2.3.00/30.0032), by the Czech Science Foundation (15-03044S, 18-27204S), Grant Agency of University of South Bohemia in České Budějovice, Czech Republic – GA JU (GAJU 04-155/2013/P) and by the European Union FP7 project ANTIDotE (602272-2) and ANTOGONE (278976).

Acknowledgements

This thesis is not a piece of work of just one person. It is in fact a product of collective efforts of all the people who were supporting me throughout my life or influenced me to some extent so that I ended up here, writing these last rows of on of the greatest achievements of my life.

First, I want to express my deepest gratitude to my supervisor Libor who accepted me to his big family in his laboratory and allowed me to work on this project.

Second, I want to thank Janko, my mentor, who were guiding me through the entire time of my PhD studies and who supported me with all the advisory that a PhD student could have possibly wished for. Additionally, I also want to thank Janko for providing me with the possibility to work on so many projects that have been underway in our lab and which, besides all, helped me to learn so many new and exciting things and facilitated solving so many problems otherwise impossible to puzzle out.

I also want to express my gratitude to Vláďa Beneš, the head of the GeneCore facility, for having me in his lab for my PhD internship and also for the sequencing data that form a basis of this thesis.
I want to also thank to my entire family, and in particular to my parents, for supporting me through my life and during my studies and also to my brother and grandparents, most of whom are no longer with us. I want to thank a lot to all my lab colleagues who I could always asked for help and advisory and who were also a great mental support to me, namely to Martin, velká Hanka, Káťa, Dima, malá Hanka, malá Pája, Jarka and all others who have been around.
I also thank to all my friends just for being friends, because friendship is a priceless valuable and a real friend is a person who you can always rely on and who supports you unconditionally, which I appreciate the most, and which I had never taken for granted. Namely I want to thank to Zolík, my best and most loyal friend, and also to Péťulka, Káťa, Váša, Mary, Danuš, Peťánek, velkej trpaslík, my friends from our volleyball group and all others, it is such a shame that I cannot name all of them!
Finally, I want to thank to my boyfriend Péťa, he is the most patient and tolerant person in the world and I cannot express more how great support he has been to me, especially during the last few months when I was finishing my thesis. It has been such a stressful period of my life and I cannot even imagine how I would survive that without his patience and support.
Thanks to All of You!

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. Vechtova, P., Sterbova, J., Sterba, J., Vancova, M., Rego, R.O.M., Selinger, M., Strnad, M., Golovchenko, M., Rudenko, N., Grubhoffer, L. 2018. A bite so sweet: the glycobiology interface of tick-host-pathogen interactions. *Parasites and Vectors* 11(1):594. doi: 10.1186/s13071-018-3062-7.

Pavlína Věchtová was in charge of writing chapters "Glycosylation in the Borrelia infection cycle", "A short overview of host ECM proteins", "Vector-host glycosylated interactions", "Tick glycans" and in manuscript preparation, organisation, and revisions.

II. Kotsarenko, K., Vechtova, P., Hammerova, Z., Langova, N., Malinovska, L., Wimmerova, M., Sterba, J., Grubhoffer, L., 2019. Newly identified DNA methyltransferases of Ixodes ricinus ticks. *Ticks and Tick Borne Diseases*. 11(2):101348. doi: 10.1016/j.ttbdis.2019.101348.

Pavlína Věchtová was in charge of bioinformatic and experimental procedures and participated in writing and revision of the manuscript.

III. Vechtova, P., Fussy, Z., Cegan, R., Sterba, J., Erhart, J., Benes, V., Grubhoffer, L. Catalogue of stage-specific transcripts in *Ixodes ricinus* and their potential functions during tick life-cycle. *Parasites and Vectors.* Accepted.

Pavlína Věchtová was in charge of experimental procedures and bioinformatic analyses, writing the manuscript and organizing and curation of the article.

Table of Contents

1 Introduction

1.1 Methylation and glycosylation in *Ixodes ricinus* transcriptome

1.1.1 Perspectives for sequencing and analysis of transcriptome

Gene expression and synthesis of the respective proteins is regulated on multiple levels, starting with gene transcription where multiple regulation mechanisms control not only the quantity but also the quality of mRNA being expressed in cellular nucleus at a given time. This is happening in all phases of the transcription process mediated by a range of regulation factors and epigenetic modifications which influence the chromatin state of a given gene, initiation, progress, and attenuation of its transcription. The quality of mRNA is post-transcriptionally modulated during transcriptional splicing, which in Eukaryotes, extends a coding capacity of the genome and provides a broader variety of transcripts and consequently a larger diversity of proteins being synthesized within a single cell from a single genomic locus.

Consequent regulation of protein synthesis happens analogically in all phases of mRNA translation to protein co- or post-translationally. The control over protein synthesis in each step is driven by a network of different factors whose activity finely tunes the quantity and quality of proteins being synthesized at given time and under ever-changing environmental conditions, which enables the cell to respond very dynamically and promptly to various intra or extracellular signals.

The existence of such a complex gene expression regulation is reflected in the need to promptly react to the changes in environment and renders better adaptability to the host organism. All of these processes, although currently well understood and described in fine details, are still happening in the black box hidden behind a cellular membrane and their investigation is largely done via indirect observations such as a characterization of the population of transcripts or proteins that are present within the cell.

The knowledge of gene transcription in an organism can thus contribute to a better understanding of processes that are underway at that moment in the organism and

may also provide an insight into the impulses that triggered the regulation of transcription under current physiological conditions.

Molecular techniques used in the field of cellular biology have advanced over the past few decades and the scientific community currently exploits the possibilities of batch techniques over laborious and time-consuming series of individual experiments used at the emergence and development of molecular biology.

The existing state-of-the-art technologies, like next-generation sequencing producing large amounts of data, give more thorough and comprehensive answers to questions formulated within each research project.

1.1.2 Comparative transcriptomics for the identification of stage-specific transcripts in the tick *Ixodes ricinus*

Ixodes ticks are ectoparasitic blood-feeding arthropods whose life cycle is well adapted to the climatic conditions of central Europe. In the last two decades, their habitat has been expanding due to an acceleration of climate changes and so does the prevalence of the pathogens that ticks transmit **[1]**.

Tick life cycle is tightly bound to the spreading of several clinically relevant pathogens amongst which the most featured are the tick-borne encephalitis (TBE) causing tick-borne encephalitis virus and the Lyme disease causative agent *Borrelia burgdorferi sensu lato*. Growing number of TBE and Lyme disease cases motivates tick parasitologists to encompass a range of methodological approaches which support our current understanding of tick physiology and mechanisms underlying the spread and propagation of tick-borne agents in the context of tick life-cycle [2,3].

Comparably to other scientific fields, also tick-directed research has been experiencing a boom of publications dealing with “omics” data. Public databases are bursting with collections of transcriptomic data generated for projects which conceptually aim at characterisation and/or quantification of transcripts that underlie the tick capacity to transmit tick-borne pathogens in organs important in tick feeding and immunity. The list of *I. ricinus* Transcriptome Shotgun Assembly (TSA) projects currently available in public databases is given in Table 1.

Table 1. List of published *I. ricinus* transcriptome sequencing projects ordered chronologically.

TSA accession	BioProject Accession	source (stage; tissue)	feeding status	# contigs	reference
GADI01	PRJNA177622	adult female and nymphs; salivary gland	fed	8,685	[4]
GANP01	PRJNA217984	adult and nymph; salivary gland and midgut	partially fed	16,002	[5]
GBIH01	PRJNA183509	adult female; salivary glands, midgut, haemocytes	semi-engorged female	2,854	[6]
GCJO01	PRJNA270980	adult; midgut	unfed	59,924	[7]
GEFM01	PRJNA311553	adult; gut	fed	7,215	[8]
GEGO01	PRJNA312361	adult; salivary gland	fed	7,692	[9]
GFVZ01	PRJNA395009	nymphs, adult males and females; whole body	unfed, partially fed and fully fed	179,316	[10]

1.1.3 Potential role of methylation and glycosylation in response to pathogens

Methylation and glycosylation are involved in the regulation of gene expression. Control over gene expression is done using a diversity of cellular regulatory mechanisms. Glycosylation and methylation essentially represent a form of gene expression regulation, each of which stands on the opposite end of gene expression pathway. While DNA methylation regulates expression of gene on the level of genetic information itself, protein glycosylation, among others, is employed in the regulation of protein translation which further influences protein folding, stability, and transport to its destination (Fig 1).

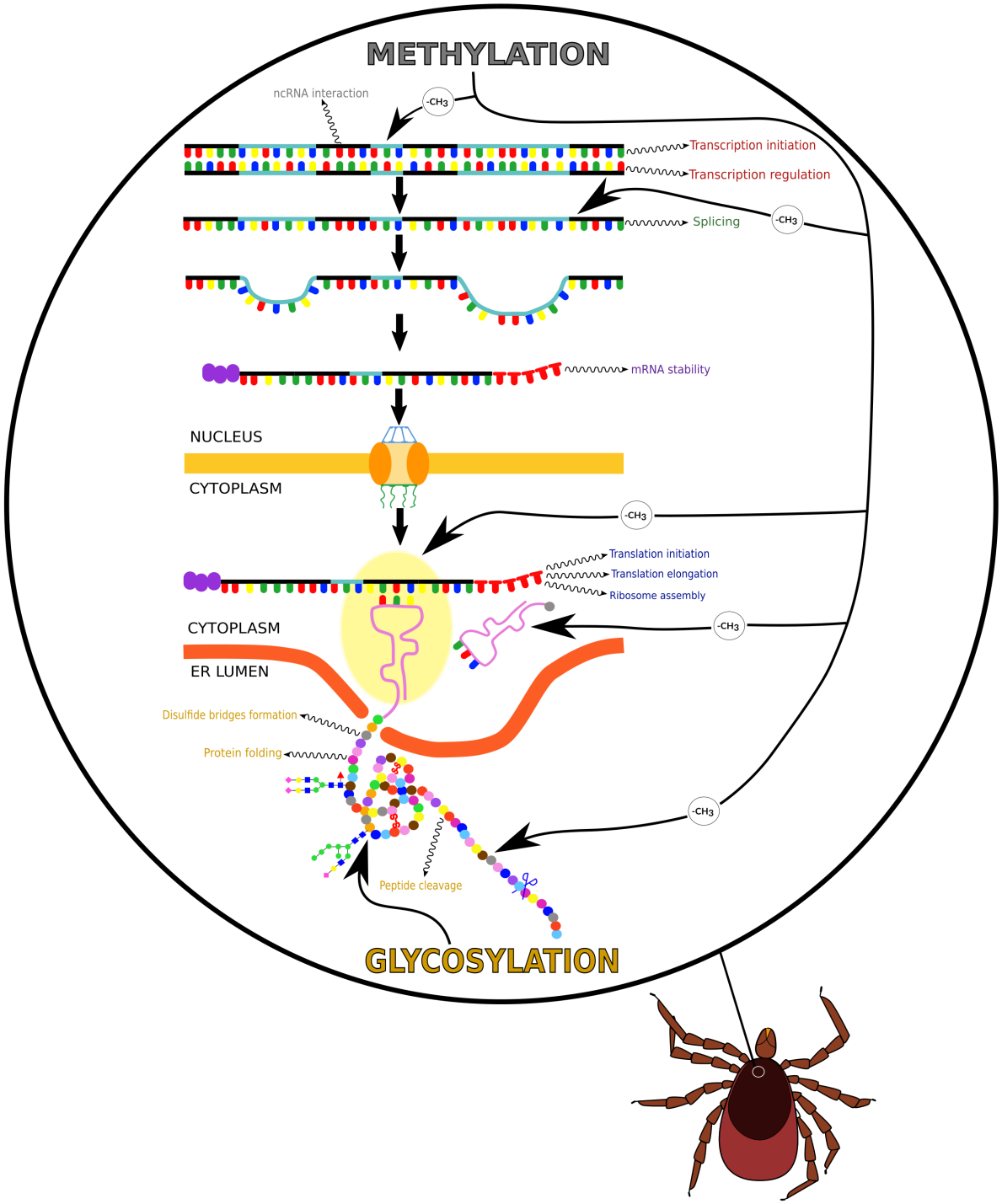


Figure 1. Schematic representation of the eukaryotic gene expression regulation. Molecules and processes that are subjected to regulation by methylation and glycosylation are indicated by arrows

Studies exploring a linkage of reciprocal methylation changes in the host-parasite relationship in very diverse organisms suggest a potential of finding
a similar linkage in other examples of parasitic organisms, including ticks. The existence of applied research directed against methylation in clinically important pathogens shows a great potential for extending this research to other parasitic organisms (Tab. 1).

Very different mechanisms and motivations in the exploitation of glycosylation apparatus are seen in host-pathogen interaction research. Glycans are usually the first molecules that the host immune system encounters when detecting the invading pathogens [31]. In many instances, they mediate an interaction of invader with host cells and represent a molecular anchor for successful pathogen infection [32]. Glycans are inherently immunogenic due to their exposure on the external face of the cellular membrane. Many research groups have already appreciated these glycan properties and described a phenomenon of molecular mimicry when the invading pathogen hides in a coat of host-like glycan structures and conceals its presence from the host immune system. Other studies revealed an ability of the invading pathogen to modulate a host immune response through the alteration of glycan coating on the parasite surface using its own or host-synthesized monosaccharides. Tab. 2 presents a selection of research or review publications dealing with specific examples of statements presented above.

These represent just a fragment of evidence which also initiated this study and encouraged an investigation of DNA methylation and glycosylation in the context of *I. ricinus* biology. Those two interesting but also very important regulatory mechanisms certainly play an essential role in diverse cellular processes and mediate an implementation of changes triggered by a complex interplay of different cellular signalling cascades. Lots of these changes manifest as stimuli for gene expression regulation which ultimately lead to the activation of adaptive mechanisms responding to new environment. These also include adaptations to presence of pathogens and are implicated in very complex processes such as host-parasite co-evolution, heritable changes in virulence factors or host resistance [12,33–35]. Identifying their significance for ticks and also their potential role in the tick-host-pathogen relationship is thus of great interest.

Table 2. Examples of basic and applied research investigating or reviewing the role of methylation and glycosylation in a host/vector-parasite relationship.

Parasite species (order)	host/*vector	Title	Reference
role of methylation in the host-parasite relationship			
Leishmania donovani (Euglenozoa)	*Homo sapiens* (Mammalia)	*Leishmania donovani* infection causes distinct epigenetic DNA methylation changes In host macrophages.	[11]
non-specific Fungi and Bacteria	*Galleria mellonella*, *Tribolium castaneum* (Arthropoda)	The role of epigenetics in host–parasite coevolution: lessons from the model host insects *Galleria mellonella* and *Tribolium castaneum*	[12]
Schistosoma mansoni (Platyhelminthes)	*Biomphalaria glabrata* (Mollusca)	The *Biomphalaria glabrata* DNA methylation machinery displays spatial tissue expression, is differentially active in distinct snail populations and is modulated by interactions with *Schistosoma mansoni*	[13]
Cotesia plutellae (Arthropoda)	*Plutella xylostella* (Arthropoda)	An endoparasitoid wasp influences host DNA methylation	[14]
Schistosoma haematobium (Platyhelminthes)	*Homo sapiens* (Mammalia)	Schistosomiasis induces persistent DNA methylation and tuberculosis-specific immune changes	[15]
Fasciola gigantica (Platyhelminthes)	*Bos taurus* (Mammalia)	DNA methylation and hydroxymethylation profiles reveal possible role of highly methylated TLR signaling on *Fasciola gigantica* excretory/secretory products (FgESPs) modulation of buffalo dendritic cells.	[16]
Metarhizium anisopliae (Ascomycota)	*Rhipicephalus microplus* (Arthropoda)	Genome-wide DNA methylation analysis of *Metarhizium anisopliae* during tick mimicked infection condition	[17]
Plasmodium falciparum (Protozoa)	*Homo sapiens* (Mammalia)	DNA methylation bisubstrate inhibitors are fast-acting drugs active against Artemisinin-resistant *Plasmodium falciparum* parasites	[18]
	Review	Epigenetic effects of infection on the phenotype of host offspring: parasites reaching across host generations	[19]
	Review	Epigenetics of host–pathogen interactions: The road ahead and the road behind	[20]
role of glycosylation in a host/*vector-parasite relationship			
Trypanosoma brucei (Euglenozoa)	**Glossina* spp. (Arthropoda)	Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors.	[21]
Trypanosoma cruzii (Euglenozoa)	*Homo sapiens* (Mammalia)	Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole.	[22]
Schistosoma	*Biomphalaria glabrata*	Structural characterization of N-glycans	[23]

mansoni (Platyhelminthes)	(Mollusca)	from the freshwater snail *Biomphalaria glabrata* cross-reacting with *Schistosoma mansoni* glycoconjugates.	
Toxoplasma gondii (Apicomplexa)	*Homo sapiens* (Mammalia)	N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii.*	[24]
Chlamydia trachomatis (Chlamydiae)	*Homo sapiens* (Mammalia)	Glycosylation-dependent galectin–receptor interactions promote *Chlamydia trachomatis* infection	[25]
	Review	Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response?	[26]
	Review	Glycoconjugates in host-helminth interactions.	[27]
	Review	Glycan-protein interactions in viral pathogenesis	[28]
	Review	Glycans in the roles of parasitological diagnosis and host-parasite interplay.	[29]
	Review	Exploitation of glycosylation in enveloped virus pathobiology.	[30]

1.1.4 Methylation and glycosylation significance for tick cell-cycle

The tick-transmitted pathogens' successful spread and survival is tightly correlated with tick life cycle. It is mainly the complexity of the tick life cycle, with an integration of mechanisms facilitating pathogen transmission to final host, that draws a lot of attention. Therefore, a research of functions whose significance is life cycle-dependent and developmentally specific should also be more appreciated.

Search for the transcripts of genes involved in the investigated cellular processes in *I. ricinus* is certainly a point to start with. Similarity search based on the information from related taxa provides a first probe into the extent and diversity of molecular factors underlying the process in question.

Interrogation of tick stage-specific transcriptomes for enzymes and other proteins involved in methylation and glycosylation provides an additional insight into the modality of these processes from the developmental perspective.

Identification of repertoire of enzymes and other interacting proteins involved in methylation and glycosylation may elucidate their extent and character. Interrogation of analogical research in related organisms may also help to infer their significance for the tick.

1.2 Eukaryotic protein glycosylation

Sugars or carbohydrates are omnipotent molecules serving in various ways and for different purposes. Chemically they are defined as organic polymers constituted of monosaccharide units represented by different species of hydroxy aldehydes and hydroxy ketones. Carbohydrates in the context of living organisms are employed in lots of processes. The most obvious is their integral role in energetic metabolism where they work as mediators of storage and transfer of energy like starch or glucose, respectively. They are indispensable in cell anabolism where they are used as building stones in the formation of body skeletons as for instance cellulose or chitin. However, sugars have also crucial yet still underestimated role as protein-binding molecules called glycans.

1.2.1 Glycans, an indispensable part of the currently existing cellular life

Glycans are the main object of interest in formerly rather underestimated glycomics, a scientific field that describes an array of glycoconjugates, primarily protein-bound sugars, within living organisms.

They can be homo- or heteropolymers of monosaccharide residues and they are attached to proteins either through the nitrogen in an amino group of asparagine side chain yielding the *N*-glycans or through oxygen of a serine/threonine hydroxyl group in the case of *O*-glycans. In *N*-glycans, the asparagine is always a part of a consensus sequence Asn-Xxx-Ser/Thr, where Xxx is any amino acid except proline. *O*-glycans do not share any consensus sequence; however, a vicinity of proline in -1 or +3 position with respect to serine or threonine in protein sequence is often required.

Glycans, in terms of eukaryotic cell, are formed on a variety of proteins co- or post-translationally in endoplasmic reticulum (ER) and further modified in Golgi apparatus (GA) during a process called glycosylation. The structure of glycans attached to proteins or lipids is often complex due to a combination of different monosaccharide units, a system resembling the DNA code. Basically, there is a limited pool of donor molecules commonly used to create *N*-glycan structures within eukaryotes. The limited structural diversity could be caused by the essential character of *N*-

glycosylation. *N*-glycans mediate proper folding and stability of essential proteins and affect directly their proper function. Thus, only a subtle change could alter a 3D structure of the proteins involved in crucial cell processes [36]. On the other hand, relatively limited subset of *N*-glycan monosaccharides and the conserved character of *N*-glycosylation mechanisms facilitate the characterization of *N*-glycosylation and understanding of the underlying processes.

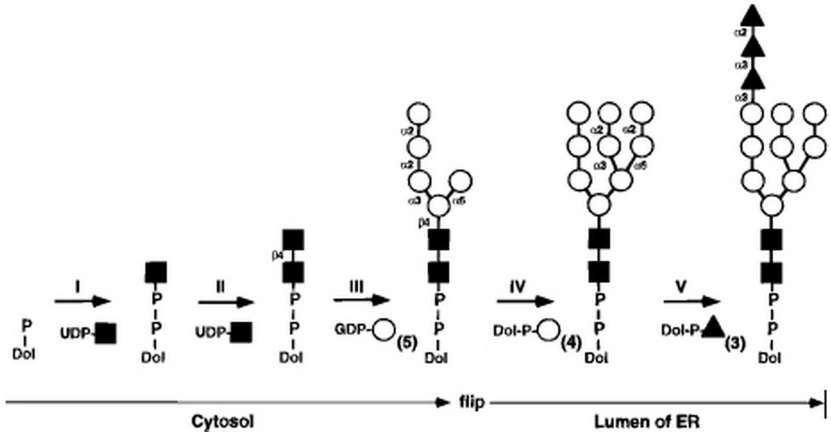


Figure 2. Biosynthesis of the dolichol oligosaccharide precursor [37].

All *N*-linked glycans are synthesized from a single core structure which is further trimmed and modified by *N*-glycosylation machinery. Formation of the core structure proceeds on a lipid molecule called dolichol pyrophosphate (Dol-PP) which is located in the ER membrane. The formation of dolichol-linked *N*-glycan precursor is initiated on the outer face of ER membrane. When the seventh monosaccharide residue is attached, the precursor molecule is flipped to the luminal face of the reticular membrane where its synthesis is finished. The complete precursor $Glc_3Man_9GlcNAc_2$-PP-Dol counts fourteen monosaccharide residues, which applies for all higher Eukaryotes, and is ready for *en-bloc* transfer to the nascent protein. All of the following modifications of *N*-glycans are performed on the protein by means of orderly (de-)glycosylation steps within ER and GA yielding in mature glycoprotein. The details of dolichol-linked precursor synthesis and *N*-glycan maturation are given in Fig. 2 and Fig. 3, respectively.

Within Eukaryotes, three structural classes of *N*-glycans have been established: high-mannose, hybrid, and complex glycans. Their classification depends on a degree of

processing of *N*-glycan precursor in GA and selection of activated sugar donors that are used in glycan structure (see Fig. 3).

When it comes to *O*-glycosylation, the level of structural diversity reaches different dimensions. The repertoire of structures formed via *O*-glycosylation is much wider in both Eukaryota and Prokaryota. The employment of a wider repertoire of sugar donors provides more sugar combinations and consequently broader range of *O*-glycan structures. Moreover, *O*-glycans do not share one common core structure. Instead, there are four most common core structures (Fig. 4) and lots of rare ones were discovered as well [38].

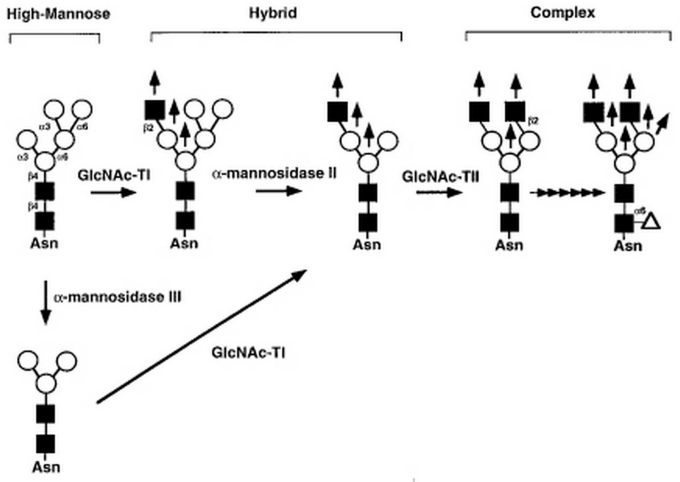


Figure 3. Vertebrate *N*-glycan diversification in GA [37].

Examination of bacterial glycosylation apparatus revealed a much wider overall variety of glycan structures in glycans compared to Eukaryotes. Prokaryotic glycosylation was formed over considerably longer evolutionary time that Prokaryotes had compared to Eukaryotes, hence more elaborate glycosylation mechanisms had a chance to evolve. Moreover, Prokaryotes are more versatile organisms in relation to ecological niches that they inhabit along with their shorter generation time and ability to exchange their DNA material by means of lateral transfer even among phylogenetically distant species [39].

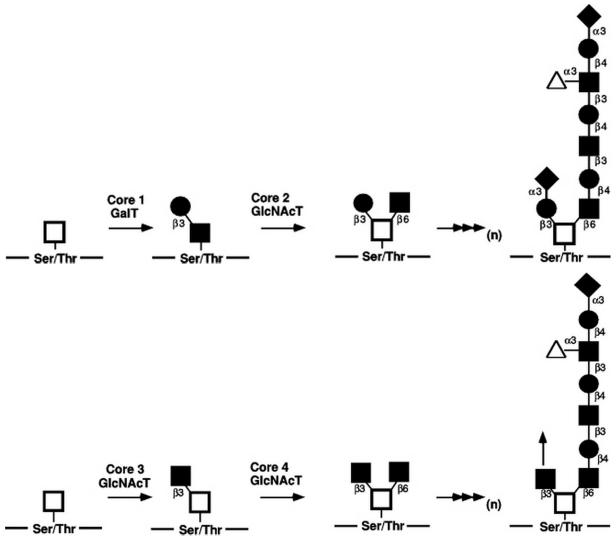


Figure 4. Core 1, 2, 3, and 4 O-glycan subtype formation [38].

1.2.2 Glycan functions

The glycan function is implemented even before the synthesis of the respective glycosylated protein is finished. Glycans help with glycoprotein folding and refolding of misfolded proteins, thus maintaining their functional integrity. Proper glycan synthesis and protein folding are checked in every step of the glycoprotein synthetic pathway throughout ER. For this purpose, glycans interact with ER-residing carbohydrate-binding proteins or lectins that reveal possible miss-folded state of the glycoproteins and arrange their refolding or sentence them for degradation in cytosolic protease complexes.

Some glycoproteins are not transported to GA but their signal sequence guides them for example to lysosome as is the case of lysosomal enzymes. However, most glycoproteins are subjected to further processing in GA where their glycans are modified considerably. These glycoproteins are predominantly exported outside the cell or into an outer membrane where they function as signalling molecules [40]. Here they are engaged in cell signalling pathways leading in regulation of various cellular processes that are responsive to external stimuli or in self-molecules recognition [41].

Importance of glycosylation is even more evident when studying altered phenotypes of organisms with glycosylation dysfunctions [42]. Reversely, altered glycosylation pattern can be a symptom of a disease. Alteration in glycosylation apparatus can

manifest in a reduction or absence of glycosylation in glycoproteins leading to a synthesis of dysfunctional proteins and subvital or lethal phenotypes [42].

More profound knowledge background of glycosylation mechanisms will help us in understanding of human congenital disorders of glycosylation and possibly in development of an appropriate treatment. It may also contribute to better comprehension of mechanisms of pathogen infection and propagation which can contribute to a development of molecules of biotechnological interest, various biotherapeutic agents like vaccines or human synthetic substitutes [41].

1.2.3 Glycosyltransferases, a toolkit for glycan assembly

The population of glycan structures in the cell is not primarily given by the availability of donor molecules or by the protein to be glycosylated but most importantly by the presence and number of genes underlying particular glycosyltransferases (GT) and their expression.

GTs represent a large and heterogeneous group of enzymes processing variable donor and substrate sugar molecules. However, their catalytic mechanism shares similar features. Nature of their employment in glycan synthesis lies in a catalytic transfer of a glycosyl moiety of an activated donor molecule to an acceptor substrate molecule with either retention or inversion of the anomer.

The process of *N*-glycosylation is separated into several phases and different sets of GTs and glycosidases are employed in each phase. First phase presents a set of GTs and glycosidases working on an *N*-glycan core structure formation. Details of the process are depicted in Fig. 2. Second group represents GTs located within GA which participate on a branching and diversification of high-mannose, hybrid, and complex glycans. The scheme is provided in Fig. 3. The third set of GTs is in fact a quite diverse group of enzymes classified according to donor molecule specificity. In terms of eukaryotic cell, the most common donor monosaccharides are added by means of mannosyltransferases, fucosyltranferases, sialyltransferases, galactosyltranferases, glucuronyltransferases. An integral role in synthesis of *N*-glycans has the oligosaccharyltransferase mediating *en bloc* transfer of Dol-PP linked *N*-glycan precursor to the nascent protein in ER and its characterization will be discussed in a separate chapter.

While sequence similarity classifies GTs into more than 90 families according to CAZy (Carbohydrate-Active enZYmes) database [43], the 3-D structural simulations along with information from crystal structure analyses assign them only 3 fold types. Despite a limited number of structural folds, the list of donor and acceptor specificities is quite large yielding in an enormous number of products. Moreover, combination of donor and acceptor specificities even extends the variety of products synthesized by GTs. Some of them specialize in different donors and share the same specificity for one acceptor molecule. Inversely, there are GTs specializing in one donor but recognizing more than one acceptor molecule. The promiscuity in donor and acceptor binding induce rather stochastic process of competition of several different GTs for one substrate molecule or attachment of one donor to a range of acceptor molecules yielding in several glycoforms of an identical protein. The details of the enzyme competition are depicted in Fig. 5.

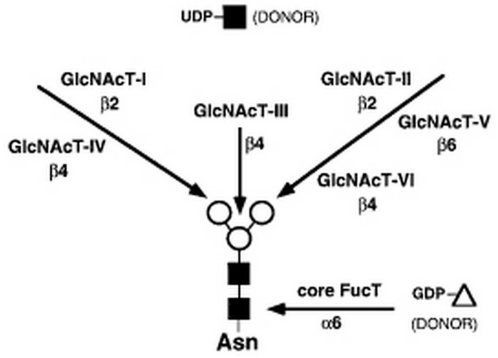


Figure 5. UDP-GlcNAc transferases in N-glycan branching and the GDP-α1–6 fucosyltransferase involved in core fucosylation in animal cells [37].

Information about glycan profiles in many organisms is still insufficient and there is not a general rule that could help with their prediction [40]. However, some findings indicate that comparison of the actual glycan structures combined with the presence and sequence similarity of genes for GTs could be a good lead to reveal evolution of glycosylation mechanism within life.

An inference of the glycosylation profile based on a search of genes or transcripts for glycosyltransferases is not sufficient as not all genes for GTs are expressed, i.e. the gene for the GT could be present but its expression could be restricted to a specific cell type, developmental period or physiological state and their transcription can thus escape its detection. The genome inquiry can yield ambiguous results, such as for

example a discovery of non-functional newly formed pseudogene that can still keep a high degree of sequence homology. Moreover, even a small change in amino acid sequence can result in an alteration of the donor or acceptor specificity, giving rise to a homologue with different function. Underlying genetic information can be also changed in response to differential temporal and spatial gene expression as well as in response to epigenetic factors or to the change of environmental conditions that are all to be regulated by particular cell and correspond to current needs of the cell. All above mentioned factors change GTs expression pattern and can even induce the expression of different GTs seemingly latent in a genome [44].

1.2.4 Oligosaccharyltransferase

One of the crucial, evolutionarily conserved steps in *N*-glycosylation is represented by a transfer of *N*-glycan precursor from Dol-PP to a nascent protein. The transfer is mediated by dolichyl-diphospholigosaccharide:protein GTs, commonly denoted oligosaccharyltranferases (OST). OSTs mediate a very specific and highly conserved chemical reaction of transfer of tetradecasaccharide ($GlcNAc_2Man_9Glc_3$) from a Dol-PP to an acceptor asparagine in a consensus sequence of nascent protein. The structure of the transferred precursor is conserved among higher Eukaryotes in terms of both composition and the number of the sugar building blocks that OST transfers to a protein. Different composition of the precursor has been observed in lower Eukaryotes; for example, in *Trypanosoma* and *Leishmania* the number of monosaccharides is reduced to eleven and eight, respectively [45–47]. The enzyme itself comprises a single catalytic subunit in Bacteria, Archea, and Lower Eukaryotes. Higher Eukaryotes require at least eight subunits to assemble the fully functional enzyme. However, the substitution of multimeric enzyme in yeast with single subunit enzyme of *Leishmania* was sufficient to complement its full function [48,49].

In *Saccharomyces cerevisiae*, the enzyme was studied the most comprehensively. The yeast OST is a multimeric protein complex composed of nine subunits (Ost1p, Ost2p, Ost3p or Ost6p, Ost4p, Ost5p, Wbp1p, Swp1p, and Stt3p) which assemble into three subcomplexes [50,51]. The number and similarity of particular protein subunits can slightly differ among Eukaryotes. Of the nine subunits, only five of them were proven

to be essential for the cell, based on experiments with deletions of genes in *S. cerevisiae* [52–56].

In order to fully characterize the OST, a catalytic subunit had to be identified. A search for catalytic subunit was based on several approaches. One approach involved DNA sequence comparison of subunits from various Eukaryotes ranging from lower Eukaryotes to vertebrates. The comparison revealed that the STT subunit was the most conserved of all polypeptides of the OST complex. This finding was later supported by the discovery of STT3-like gene being the only homologue of the OST subunits in archaebacterial genome. Further analysis of eukaryotic and archaeal STT3 identified invariant sequence (WWDXG motif) that was found in all STT3 homologues analysed so far and which represents the catalytic site of the enzyme [51]. In addition, at least one STT3 homologue is present in every eukaryotic genome analysed so far. The highest variation in number of STT3 genes has been observed in lower Eukaryotes; particularly 4 and 3 genes are present in genomes of *Leishmania* and *Trypanosoma*, respectively, [57] while yeast possesses one STT3 gene only. All higher Eukaryotes have two isoforms of the STT3 subunit, STT3A and STT3B. Phylogenetic analysis showed that STT3B isoform of higher Eukaryotes bears similarity to the STT3 subunit of fungi and nematodes while the STT3A isoform seems to be related to the one found in *D. discoideum* [58]. A reason why an organism would keep both isoforms in its genome is not entirely clear; however, an analysis of their expression revealed tissue-specific expression patterns for both of them. Moreover, both isoforms exhibit differential activities. This implies that their differential expression is a form of regulation of glycosylation [58].

Further attempts to characterise both STT3 isoforms managed to distinguish them functionally. While STT3A manifests its activity co-translationally, STT3B works on the protein post-translationally, although the unfolded state of the protein is required for proper activity of both isoforms. Moreover, their activity is rather “site-selective” as the STT3B preferentially glycosylates *N*-terminally located glycosylation sites [59].

1.2.5 Diversity of *N*-glycosylation mechanism

N-glycosylation occurs in a wide spectrum of organisms of all kingdoms, which gives evidence of its essential character for organism as well as of an age and complexity of

the processes participating in the formation of complex sugar residues. Thanks to differences in existing glycosylation mechanisms, it is not surprising that variability of glycan structures reached different degree of development for a given cell architecture.

It is evident that the largest body of information is available for *N*-linked glycans thanks to a high conservancy of asparagine-linked structures present in Eukaryota. Such a high conservation is underlaid by the essential role of *N*-linked glycans in protein folding and other functions that eukaryotic cell developed through its evolution.

It was not until recently when glycosylation has been assigned solely to eukaryotic kingdom and any ongoing research focused on a characterisation of eukaryotic glycosylation at the time. The very first reference on the existence of a glycoprotein outside Eukaryotes comes from survey of S-layer in archaebacterial gram-negative *Halobacterium salinarium* where the glycosylated bacterial proteins were first identified [60]. It took more than twenty years before the first reports on prokaryotic glycosylation emerged [61]. The prokaryotic *N*-glycosylation system was first characterized in gram-negative *Campylobacter jejuni* where the first genes involved in *N*-linked glycoprotein synthesis were identified [62]. Research of prokaryotic *N*-glycosylation advanced substantially as of its discovery.

N-glycosylation reached the highest degree of its diversity and specialization in Eukaryotes. Their *N*-glycans are formed in ER and their actual diversity is shaped in GA, both uniquely eukaryotic structures. Glycans entering GA are subjected to a series of trimming reactions and further addition of a variety of monosaccharide donors and branching reactions all mediated by GA glycosidases and GTs. Thus, the existence and further development of eukaryotic organelles probably converged with a development of *N*-glycosylation to reach high degree of specialization and diversity of *N*-linked glycans that we observe in Eukaryotes today.

Prokaryotic cells lack such organelles and their protein processing is restricted to periplasm and cell surface. Moreover, it was discovered that the *en bloc* transfer of *N*-glycan precursor in Prokaryotes does not occur until the target protein has already fully folded, which implies that *N*-glycosylation in Prokaryotes mediates other tasks but protein folding as it is in Eukaryotes. Most bacterial glycans rather assume different roles. The best-studied glycoproteins are found in S-layers, pilins, flagella,

and are surface and secreted proteins which are known to be involved in adhesion and/or biofilm formation. In the case of parasitic bacteria, some glycoproteins facilitate the processes of invasion and infection of the host organism [63].

The modifications and further processing of *N*-glycans that commonly occur in Eukaryotes were not observed in prokaryotic cells leaving all the *N*-glycan diversity to be shaped before the *en bloc* transfer of the glycan to the bacterial protein [63]. Moreover, the glycan assembly *per se* occurs on an undecaprenol, a prokaryotic counterpart of the eukaryotic Dol-PP. In the best-described model, the final core structure counts seven sugar residues and uses some prokaryote-specific units, the amino- and deoxy-sugars like bacillosamine and its derivatives [64].

Despite these great differences in glycosylation mechanisms present among Eukaryota and Prokaryota, some strikingly similar structures were discovered that some Prokaryotes share with Eukaryotes. Most of these Prokaryotes were identified as parasitic species, thus these similarities were assigned to the phenomenon of "molecular mimicry" that the pathogenic organisms developed to diminish a risk of their detection by the host immune system. This ability of pathogenic bacteria to develop enzymatic pathways for the production of eukaryotic-like glycans suggests a participation of molecular mimicry in accelerating the process of adaptation of pathogenic bacteria to their hosts and consequently, it contributed to the evolution of Prokaryotes [40].

1.2.6 Glycosylation in Eukaryotes

Even though the *N*-glycosylation pathway is shared by all members of the domain Eukaryota, only vertebrates are fully able to exploit a structural potential of *N*-glycosylation machinery. In fact, higher Eukaryotes are equipped with enzymatic machinery that allows them to process the *N*-glycan precursor completely by mannose (Man) residue trimming, giving rise to the *N*-glycan Man3GlcNAc2 core structure, which provides a substrate for glycosyltransferases generating hybrid and complex glycans. In lower Eukaryotes an absence of these enzymes led to a development of their glycans in different direction. For instance, *Saccharomyces* processes its glycans to further develop its Man branches in order to shape large high-mannose structures while slime mould trims its Mans [65].

Development of hybrid and complex glycans also involves inclusion of more types of sugar residues and thus higher Eukaryotes display structurally more diverse *N*-glycans often structurally characteristic for a particular taxonomic group. It is known that insects typically include a1,3-linked fucose (Fuc) to their proximal core GlcNAc, while plants and also some invertebrates often incorporate a bisecting xylose. Vertebrates, on the other hand, are typical of their a1,6-linked core Fuc as well as of the presence of terminal sialic acid in their glycans (Fig. 6). It is not surprising that both formerly mentioned a1,3-core Fuc and bisecting xylose were proven to be immunogenic when exposed to the vertebrate immune system. Such findings stress even more an integral role of glycans in a vertebrate immune system [40].

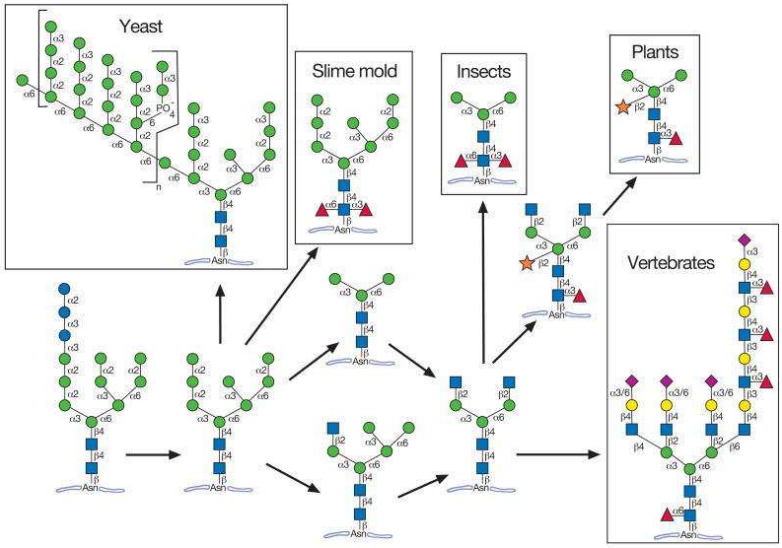


Figure 6. Dominant pathways of N-glycan processing among different taxa [65].

1.2.7 Glycosylation in invertebrates

Despite a diversity of invertebrates covering ~97 % of all eukaryotic species on Earth, the majority of research focused on glycomics and glycoproteomics has been done in mammals. In invertebrates, most research groups aim at characterisation of glycosylation in the model species *D. melanogaster* and *Ca. elegans*. Thus, the conclusions made about the significance of this posttranslational modification in invertebrates are by large build upon findings made in these two models.

Gůycosylation is a very heterogenosu modification in terms of both enzymatic apparatus and a glycome that is produced. The comparison of the two most widespread, the *N*- and *O*-protein glycosylation pathways shows more common

features among Eukaryotes in *N*-glycosylation which was also shown in some invertebrate species with what is known from vertebrates.

Although arthropods possess the genes for GTs participating in hybrid and complex *N*-linked glycan synthesis, these glycans represent only about 10 % of all arthropod *N*-glycome [66–68]. Such discrepancy was assigned to an activity of hexosaminidase which removes *N*-acetylglucosamine (GlcNAc) added by *N*-acetylglucosaminyltransferase I from a common substrate for hybrid and complex glycan synthesis pathway (Fig. 2). The continuous removal of GlcNAc significantly lowers an overall amount of the substrate for enzymes acting in hybrid and complex glycan synthesis contributing substantially to an observed ratio in favour of high-mannose glycans.

Consistently, specific glycan structures were described to be specific for invertebrate lineage. For example, arthropod glycosylation is represented by an addition of Fuc to the proximal GlcNAc in a1,3-linkage which elicits the above mentioned immune response in vertebrates.

Caenorhabditis elegans also produces all common *N*-glycan types [69]. However, some atypical glycan forms such as fuco-paucimanosidic or bisecting fucose-galactose branched glycans were identified in the glycomes of this nematode [69–71].

Conversely, detection of sialic acid, a typical vertebrate structure, was reported in invertebrate cells. Although many studies dealt with detection of this structure in invertebrates, only a few pieces of evidence of its presence have been brought forward. The *Drosophila* genome sequence project revealed a presence of sialyltransferase-like gene in a fruit fly genome [72] and experiments with transfection of insect cell lines with mammalian sialyltransferase and galactosyltransferase yielded sialylated glycoproteins [73]. An interrogation of mosquito glycome revealed a presence of sulphate and glucuronic acid residues in the glycan branches [74].

The discovery of complex glycans in insects also brought about the detection of previously unforeseen diversity of different monosaccharide residues originally attributed only to higher Eukaryotes. For example, a re-evaluation of inclusion of charged glycan moieties, a feature typical of complex vertebrate glycans, is now in place in invertebrate glycoproteomes [75].

N-glycans modified by zwitterions phosphorylcholine and phosphoethanolamine were detected in some lepidopteran species [76] and some phosphorylcholine complemented with Fuc and GlcA were also found in a mollusc glycome [77,78].
Schistosoma is apparently able to attach xylose to its *N*-glycan core and a honey bee synthesizes glucuronylated glycans for a royal jelly production [79,80].
N-glycosylation is functionally implicated mainly in the processes that require their cell membrane association like cell-cell adhesion and signal transduction that are ultimately involved in body growth, embryogenesis, and organ development [81]. Functional studies shed more light on details of *N*-glycan involvement in specific cellular functions of invertebrates. For example, an alg5 GT, attaching terminal glucose to an *N*-glycan precursor in the early steps of *N*-glycan formation was found crucial for epidermal differentiation during *Drosophila* late embryonic development and a mutant GT causes disruption in the formation of lateral epidermis and mesoderm during *Drosophila* gastrulation [82].
Screening of *O*-glycosylation is much more challenging due to an absence of a consensus glycosylated amino acid sequence and a high variability of core structures that are synthesized directly on a target protein in a sequential manner compared to an *en bloc* transfer of a prebuild common core of *N*-glycosylation pathway. Lack of common structural features in *O*-glycans hinders the development of universal screening techniques exploiting common deglycosylation enzymes or enrichment methods prior MS detection. Moreover, the stability of different *O*-glycan bonds differs and the *O*-glycan presence and further detection in biological samples can be eliminated during routine sample preparation pipelines.
Therefore, the majority of findings, collected about *O*-glycosylation is based on the most widely distributed O-GalNAcylation, which was first described in mucin and hence the products of this modification are better known as mucin-type glycans. Mucins are one of the most heavily *O*-glycosylated proteins and are also the most widespread in Prokaryotes. Investigation of *Drosophila O*-glycans showed ~90 % of entire *O*-glycoproteome repertoire being of mucin-type. Similarly to *N*-glycans, also O-linked mucin-type glycans are primarily targeted to extracellular side of the cell membrane or are secreted outside the cell in invertebrates [83].
Functional manifestation of glycans in development and their roles associated with multicellular life were demonstrated in mutant models of underlying enzymes.

For example, a mutation of gene for a PGANT family member *O*-GalNAcylating enzyme caused an impairment of cell adhesion and Golgi associated protein trafficking [84].

Significance of *O*-fucosylation and *O*-glucosylation in epidermal growth factor domains and thrombospodin type I domains was established upon observation of phenotypes of mutants in the underlying enzymes *O*-fucosyltransferase and *N*-acetylglucosaminyltransferase (O-GlcNAc-transferase) that caused multiple organ defects followed by death in the former and defects in eye development, adult eclosion, and shortened survival in the latter case [85–88]. Concerted action of xylose and glucose attaching enzymes during glycosylation of Notch in *Drosophila* are essential for its proper transport and Notch-mediated signal transduction [89,90].

O-mannosyltransferases are a widespread group of enzymes, whose pathway and activity is conserved in Eukaryotes. Their importance for human physiology was demonstrated by description of severe phenotypes, such as impairment in brain, eye, and muscle, caused by mutations in the underlying enzymes [91]. Drosophila flies bearing mutations in two mannosyltransferases, POMT1 and POMT2, exhibited defects in synaptic transmission, abdomen development, and muscle dystrophy ascribing these enzymes a role in a normal muscle development [92,93].

O-GlcNAc modification assumes a specific position among other glycosylation reactions. *O*-GlcNAc-transferases are cytosolic proteins and their action is decoupled from traditional ER- or GA-associated glycosylation pathways. GlcNAc attachment hence occurs rather in cytosol or in nucleus and the reaction is thus more dynamic and prone to complementary modification by other cytosolic posttranslational modifications. Its involvement in cell signalling, stress response, and energy metabolism was suggested [94]. Mutations of GlcNAc-transferase (GlcNAc) or *O*-GlcNAcylase (OGA) in *Ca. elegans* led to a dysregulation of hundreds of transcripts, which was linked to an alteration of RNA polymerase II (RNAPII) activity and also affected basal transcription complex, which are both targeted by this modification [95]. Detailed description of altered phenotypes of GlcNAc mutant revealed decrease in life span and impaired regulation of metabolism [96], which was also to a certain extent described in *Drosophila* [97,98].

1.2.8 Conclusions

Glycosylation represents a modification that widens a repertoire of protein forms whose structure and properties are encoded in the sequence of DNA. It is able to finely tune the protein structure and contribute to the modification of protein properties which predefines it for specific cellular functions. Glycosylation is highly variable modification which can yield a repertoire of different glycoforms that can be employed in a fine-tuning of different cellular functions and contribute to an enrichment of cellular regulome. The variable nature of this modification was also reflected in the diversity of different glycan structures that evolved within Eukaryotes in a lineage-specific manner. Although the mechanism of protein glycosylation is more or less conserved and well described, the diversity of resulting glycoproteomes is as large as is the number of species on Earth. Therefore, it is difficult to make general conclusions for particular phylogenetic group based on the observations in a few selected species. Still, common features can be ascribed to invertebrate glycosylation, many of which are applicable Eukaryote-wide. It is widely accepted that majority of glycoproteins are cell membrane-associated and their participation in intercellular adhesion or integration in cell signal transduction were demonstrated in numerous functional studies in model invertebrates. The essential role of glycosylation for invertebrates was proven by reduced vitality and lethality of phenotypes resulting from mutation of underlying glycosylating enzymes. Some of these observations found their counterparts in phylogenetically distant species and was instrumental in an elucidation of nature of human diseases. Importantly, both the resulting glycoproteome and its role for its carrier protein are highly contextual and more data needs to be collected in order to fully comprehend this complex feature of cellular life due to the inherent heterogeneity of this post-translational modification.

1.3 Eukaryotic methylation

DNA methylation refers to a covalent and reversible modification of genomic DNA without alteration of its sequence. Along with other epigenetic modifications, it represents a form of heritable genome regulation that is maintained across cellular division and is able to dynamically reflect changes of intracellular or environmental conditions into alteration of phenotypes. The three main mechanisms of epigenetic modification currently recognized act on chromatin via methylation of cytosine or adenine, histones via a range of post-translational modifications, and non-coding RNAs expression. Of the three mechanisms, methylation is probably the most stable and faithful epigenetic modification and is also the most well documented.

Existence of DNA methylation in Eukaryotic genomes is known as long as is the structure of DNA itself. From an evolutionary perspective, methylation in eukaryotic cell was adopted from prokaryotic system, where it is used for selective degradation of exogenous infectious DNA (bacteriophages), and has evolved to a widespread feature of eukaryotic genomes ever since [99].

It was postulated that Eukaryotes adapted methylation in response to genome duplication that happened at the nascence of eukaryotic lineage. The very primary function of methylation in Eukaryotes was also an inactivation of parasitic sequences, although of intragenomic origin. Repression of repetitive DNA became indispensable for genome integrity maintenance potentially compromised by overly active mobile elements and for the regulation of growing repetitive DNA stretches which often constitute a majority of genomic content in a eukaryotic cell. Such behaviour, originally targeting repetitive portion of the genome was later adopted for reduction of an effective size of large eukaryotic genomes and was also extended to genes. Methylation targeting transcriptional units was harnessed for aberrant transcription regulation and also became a control mechanism of contextually dependent on/off switch for spatiotemporal transcription regulation and took control over gene expression in Eukaryotes.

Methylation was widely adapted by eukaryotic organisms owing very large genomes such as those in vertebrate lineage. The importance of parsimonious genome management is apparent as seen on a positive correlation of genome size and a degree of methylation observed in several studied eukaryotes (Tab. 2) [100]

Table 2. Positive correlation of genome size and the degree of methylation in the species with known methylation pattern. The table is ordered according to genome size in descending order. The colour of DNA methylation type column correlates with the degree of genome methylation in the corresponding species. The colour key is ordered from red, corresponding to most methylated to green which corresponds to the least methylated. "Global" and "Methylated" methylation types both indicate a high degree of methylation. "Global" methylation type moreover indicates that a methylation pattern was evident from the source publication [100].

species	eukaryotic group	genome size (Mb)	# coding genes	DNA methylation type	reference
Homo sapiens	Chordata	3,098.83	20,805	Global	[101]
Macaca mulatta	Chordata	2,971.31	21,748	Methylated	[102]
Mus musculus	Chordata	2,730.87	22,515	Methylated	[103]
Equus caballus	Chordata	2,506.97	20,955	Global	[104]
Blattella germanica	Arthropoda	1,960.00	29,216	Methylated	[105]
Danio rerio	Chordata	1,373.47	25,592	Global	[106]
Gallus gallus	Chordata	1,065.37	16,878	Methylated	[107]
Glycine max	Embryophyta	978.49	55,897	Methylated	[108]
Solanum lycopersicum	Embryophyta	827.75	34,658	Methylated	[109]
Crassostrea gigas	Mollusca	557.72	26,101	Mosaic	[110]
Vitis vinifera	Embryophyta	486.27	29,971	Methylated	[111]
Bombyx mori	Arthropoda	481.80	14,623	Mosaic	[112]
Physcomitrella patens	Bryophyta	471.85	32,234	Methylated	[113]
Solenopsis invicta	Arthropoda	396.02	14,466	Methylated	[114]
Oryza sativa	Embryophyta	375.05	37,960	Mosaic	[115]
Nematostella vectensis	Cnidaria	356.61	24,773	Mosaic	[116]
Apis mellifera	Arthropoda	250.27	15,314	Mosaic	[117]
Selaginella moellendorffii	Embryophyta	212.65	34,799	Low methylated	[118]
Daphnia pulex	Arthropoda	197.21	30,590	Mosaic	[119]
Ciona intestinalis	Tunicata	177.00	11,616	Mosaic	[120]
Tribolium castaneum	Arthropoda	165.93	16,590	Not methylated	[121]
Drosophila melanogaster	Arthropoda	143.73	13,947	Not methylated	[122]
Arabidopsis thaliana	Embryophyta	119.67	27,655	Global	[123]
Trichinella spiralis	Nematoda	63.53	16,380	Methylated	[124]
Aspergillus flavus	Fungi	36.89	13,485	Not methylated	[125]
Cordyceps militaris	Fungi	32.27	9,651	Low methylated	[126]

The selective elimination of large portions of genome results in a lower amount of DNA available for scanning and binding by a limited population of cellular transcription complexes and regulatory factors. This facilitated the development of a more effective genome management and was also implicated in the execution of spatiotemporal gene expression and cellular and tissue differentiation for efficient implementation of complex vertebrate body plans.

Different situation, however, developed in invertebrate lineage. Unlike vertebrates with dense methylation pattern covering on average 80 % of their genomes, most invertebrates show thinly methylated pattern and the methylation abundance varies considerably among different lineages. Thus, theories elucidating the role of methylation in invertebrates are as numerous as is the diversity of invertebrate lineages.

1.3.1 Extent of methylation in Eukaryotic genomes

With a growing number of characterized methylomes of Eukaryotes, it is becoming increasingly evident that the pattern and extent of methylation exhibit lineage-specific features. It is generally accepted that vertebrates methylate heavily and quite densely, their genomes only avoiding rather long stretches of unmethylated CpG islands, which are by rule colocalized with promoter regions of transcription units [123].

An exploitation of bisulphite sequencing projects revealed a gradient of average methylation abundances in genomes of invertebrate protostomes ranging from zero to approximately 0.15 and 0.25-0.3 in invertebrate deuterostomes. With increasing genome size and complexity, a trend in the expansion of methylation continued and hence is the abundance of methylation of 0.42 at its lowest and extending 0.75 at its highest observed in vertebrates [124].

Arthropods methylate their genomes considerably less extensively than vertebrates. Still, the portion of methylated genomes varies substantially across Arthropod taxa. The most conspicuous difference in the degree of methylation is observed in vertebrates in comparison to thinly methylated genomes of holometabolous insects [125]. Moreover, records of losses of methylation and its underlying enzymatic apparatus in several arthropods imply that this modification may serve a different purpose than in vertebrate genomes.

Consistently with methylation density, also a pattern of methylation is surprisingly uniform within vertebrate species. Methylation in vertebrate genome forms a continuous cover while the distribution of methylation in arthropod genomes resemble rather a mosaic pawing. Such pattern is reflected in the observation that arthropod methylation is targeted to active genes forming the mosaic genomic landscape of methylated regions of active genes interspersed with large intergenic stretches of non-methylated DNA [123,126–128].

1.3.2 Role of methylation in Eukaryotes

1.3.2.1 Regulation of gene expression

Role of gene body methylation

Gene body methylation is a feature typical of all eukaryotic species studied so far and is positively correlated in constitutive gene expression and transcriptional noise reduction. However, vertebrates methylate all gene bodies irrespective of their transcriptional activity [129]. The transcriptional activity in vertebrates is reflected in the methylation state of promoter regions where the presence of methylation is negatively correlated with the transcriptional activity of the respective gene. Thereby it seems that the major role of gene body methylation in vertebrates is in the regulation of illegitimate transcription.

Inversely, only some gene bodies are methylated in invertebrate genomes [130,131] exhibiting bimodal methylation pattern. This refers to a situation when gene bodies targeted by methylation show very dense methylation pattern while the other gene bodies are not methylated at all. Invertebrates, in general, show a positive correlation of gene body methylation and gene expression. On the other hand, no linkage of gene promoter methylation with gene expression was found in invertebrates [132]. This demonstrates that methylation of gene bodies is associated with transcriptional regulation in invertebrates while methylation of gene promoters controls gene expression in vertebrates [133–136].

Recently, it has been postulated that promoter hypomethylation is not a trigger of transcriptional activity but rather serves as a primer for transcriptional activation that is factually conferred by methylation of enhancer or regulatory sequences controlling

expression of the respective genes [137]. Combined functional annotations and comprehensive phylogenetic analysis revealed that orthologues of methylated gene bodies in invertebrates are unmethylated in their promoter regions in vertebrates and vice versa [138].

However, there seems to be a gradient in this contextually dependent linkage of methylation to transcriptional activation of genes in Eukaryotes. Evidence comes from an analysis of *Ciona intestinalis* genome methylation, a representative of the closest invertebrate outgroup of the vertebrate lineage. Although not genome-wide and contextually dependent on the methylation state of the nearby gene bodies, *Ci. intestinalis* exhibits a linkage of a gene expression to gene promoter methylation typical for vertebrates [138]. The only other example of CpG promoter methylation linkage to gene expression regulation in invertebrates was observed in a *homeobox* gene promoter of mollusc *Crassostrea gigas* [139–141]. This raises a question about the function of gene promoter methylation in invertebrates.

Ci. intestinalis here could provide evidence about the transition in the role of methylation between vertebrates and invertebrates. From an evolutionary perspective, it really seems that the vertebrate lineage developed a promoter methylation to regulate gene expression developmentally and tissue specifically, while gene body methylation dominates as apparently ancestral mechanism of control over gene regulation in all Eukaryotes.

A gradient of methylation involvement in genome regulation from invertebrates to vertebrates is also evident in the degree of parental methylome reprogramming during embryogenesis. It is widely accepted that methylation plays a key role in embryogenesis, cell differentiation and parental imprinting in the vertebrates and a reset of genome methylation is crucial for proper implementation of parental methylome in an offspring. Clear evidence of methylation reprogramming during vertebrate embryogenesis comes from well-characterized genomes of zebrafish and mammals that exhibit a sharp difference among methylation levels in gametes and somatic cells in comparison to embryonic tissue. Inversely, there was no evidence for methylation fluctuation in the representatives of pre-bilaterian or protostome lineages of invertebrates as shown in sea anemone or honeybee. An absence of linkage between methylation and embryonic development was demonstrated in these invertebrates. In contrast, methylation levels dynamics during development of

gametes to an embryo was described in the species of invertebrate deuterostomes, sea urchin and sea squirt [124]. Methylation pattern of HOX genes that drive a segmentation and body plan implementation during development of metazoans is consistent with these findings. Methylation dynamics of HOX genes is absent in invertebrates, while dramatic methylation dynamics of these gene clusters is seen in different vertebrate organs during embryogenesis showing a clear association of HOX gene methylation and developmental regulation in the vertebrate lineage. Similarly, key developmental genes, like Notch, Hedgehog, or TGF-b are permanently unmethylated in their promoter regions in the vertebrates and at the same time exhibit a significant methylation dynamics of their gene bodies throughout embryogenesis and development. However, a permanent unmethylated state of these genes throughout the entire life cycle, as well as an identification of a single hox gene cluster in invertebrates, support previous findings that show no connection between methylation and developmental regulation in this group of organisms [124].

The above findings indicate that vertebrates use methylation as a major regulatory mechanism that helps in the regulation of their genomes whose large size also partially expanded due to a genome duplication at the evolutionary emergence of the vertebrate lineage. This is also associated with the regulation of complex developmental processes that evolved with the evolution of vertebrates and that are largely controlled by DNA methylation. Conversely, different regulatory processes prevailed in invertebrates and the control of embryonic regulation has most probably been taken over by histone modification, non-coding RNA or other epigenetic mechanisms.

Remarkable conclusions can be inferred from a comparison of gene methylation in species from distantly related lineages of invertebrates. The genome-wide analysis of methylomes of several invertebrate species revealed gradient in the share of the methylated fraction of genomes from very low in insect species to moderately high in sea anemone and sea squirt. At the gene level resolution, it seems that the studied insect species methylate markedly smaller share of their genes compared to other invertebrate lineages. From an evolutionary perspective, it appears that highly methylated gene body is an ancestral state in invertebrates and gradually changed to low methylation in insect lineage [131]. Closer investigation of methylation in insects reveals that gene methylation status gained lineage-specific tendencies. Many genes

enriched in specialized cell signalling pathways assume an inverse methylation state in different insect taxa. Most of these genes underlie developmental, immune, and defence mechanisms or cellular respiration and phosphorylation. The above stated might provide certain indications that insects' response to altered environmental cues as well as lineage-specific adaptations may have been accomplished via differential methylation of given functional group of genes.

Role of CpG doublets

Methylation of cytosine in CpG dinucleotide is a predominant type of methylation in Eukaryotes and is found in a symmetrical constellation on both DNA strands [100].

It is widely accepted that CpG methylation colocalizes with transcriptionally active genes in both vertebrates and invertebrates and is clearly associated with sequences of mobile and repetitive DNA in the vertebrates and plants genomes causing repression of their activity [103].

As mentioned previously, methylation was recruited from prokaryotic defensive restriction-modification system for control over transcriptional regulation of eukaryotic genes as well as for transcriptional silencing of mobile elements and genomic repeats. Both functions of DNA methylation are widespread across eukaryotes and thus are considered evolutionarily ancient; however, their extent, implementation, and purpose took different directions in different eukaryotic lineages. Universally, irrespective of taxonomic affiliation, the nature of transcriptional regulation depends on the position of CpG methylation within the transcriptional unit. Methylation of the gene body is usually associated with transcriptional activation [142]. Methylation of promoter or enhancer region, on the other hand, is a typical marker of transcriptional repression [143–145]. Consistently, promoters of the most active genes in vertebrates contain stretches of CpG (CpG islands) that are permanently demethylated, presumably with the exception of a short developmental period of time in the germline. It is not surprising then, that promoters of active genes are the only CpG rich portions of the genome while the other genomic regions are poor or depleted of CpGs.

Currently, widely accepted hypothesis talks about CpG depletion as a consequence of frequent mutations of 5 methylcytosine (5mC) in CpGs to thymine. Considering relatively high mutation rate of 5mCs in CpG doublets by deamination to TpGs, it is

only logical that the only CpGs left in a genome are those in the gene promoters that are only transiently methylated and thus least frequently threatened by methyl-CpG to TpG substitution [146–148].

Considerably less information is available for promoter methylation in invertebrate species. Thus, a correlation in the pattern of promoter and gene body methylation in vertebrates and invertebrates was investigated. Incidence of CpG rich tracts correlates with an unmethylated fraction of gene-rich genomic areas in studied vertebrate and invertebrate species. Unmethylated CpG rich tracts colocalize with promoters of genes coding for proteins that perform core cellular functions which are evolutionarily more conserved than temporally and spatially expressed genes showing variable methylation in their promoter regions [130,146,149,150].

The discrepancy in CpG content in promoters and gene bodies is indeed consistent with CpG depletion in methylated regions caused by a high rate of methylcytosine deamination events [148]. Conversely, gene bodies of genes with core functions are enriched in methylated DNA patches, a feature shared even with plant genomes [131,151]. This is in contradiction with a high mutation rate of methylated CpGs and one would expect little conservation of methylation rich sequences. The mechanism preserving a high degree of homology in these conserved orthologous genes among distant invertebrate lineages remains poorly understood at present [130,151,152]. Similar methylation pattern within transcriptional units ascribes analogous function to methylation in both vertebrates and invertebrates [130].

Mechanism of epigenetic transcription regulation

As discussed earlier, the transcriptional units of active genes are characterized by promoter regions containing unmethylated CpG islands and methylated gene bodies. Conversely, the methylation of CpGs at gene promoter causes transcriptional silencing. Interestingly, CpG methylation in both contexts are associated with H3K9me3, a marker of transcriptional repression. Thus, the methylation of gene promoters causes transcriptional repression which is essentially implemented via a concerted action of different transcription regulatory factors leading to histone modifications causing the formation of nucleosomes tightly packaging DNA molecule that cannot serve as a substrate for transcription initiation complex assembly.

The chromatin remodelling here is a result of the cooperation of DNA and histone-modifying enzymes that either add or remove respective functional groups like methyl, acetyl, ubiquitin or phosphate ultimately leading to chromatin remodelling, which contributes to chromatin condensation and transcription repression [153–156]. Inversely, histone modification can affect the methylation of the interacting DNA. Histone modifications can support or impair binding and activity of DNA methyltransferases (DNMTs) [157]. For example, H3K4me3, a marker of high transcriptional activity occurs in high abundances in CpG islands in promoter regions of actively transcribed genes. The presence of this histone modification inhibits binding of DNA methyltransferases to the respective genomic regions and contributes for example to the maintenance of an unmethylated state of the CpG islands [158–160].

1.3.2.2 Regulation of alternative splicing

Recent bisulphite sequencing studies releasing high-resolution data on genome methylation distribution have revealed a previously unrecognized pattern of methylation within gene bodies and preferential methylation of exons with a gradient of methylation at exon-intron boundaries. This observation, in the context of nucleosome positioning data or RNAPII kinetics regulation [144,145,161,162], is implicated in the regulation of gene expression on the level of transcriptional elongation and may also be involved in the regulation of alternative splicing [163] proposed for both vertebrates [161], which was also supported with functional study [164] and invertebrates [152,165–167]. Interestingly, while studied vertebrate species show enrichment of methylation in exons [163], different preferences for both introns and exons methylation was observed in studied invertebrates. In oyster, honey bee, a parasitoid wasp, or a burying beetle, an enrichment was observed in exons [167–170]. Comparison of exon methylation with spliced transcripts in oyster revealed that highly methylated exons are preferentially included in the final transcripts while significantly hypomethylated exons are skipped by the splicing apparatus [167]. Contrarily, preferential methylation of cytosines in introns was described in *Tribolium* and *Drosophila* [166,171].

1.3.2.3 Repression of mobile DNA activity

Direct association of methylation and mobile DNA repression has been shown for vertebrates and plants. Hypermethylation of intergenic regions colocalized with mobile elements has been reported and associated with the silencing of their activity. Similarly, methylation of gene bodies is correlated with both regulation of transcription initiation from alternative internal promoters [172,173] as well as with silencing of intragenic repetitive elements often colocalized with introns [103,113]. Conversely, methylation in most invertebrates is not preferentially targeted to mobile or repetitive DNA.

Although not highly abundant, still cytosine methylation of mobile DNA differs among invertebrate lineages. For example, a moderate number of 5mCs was colocalized with repetitive elements in invertebrate deuterostome *Ci. intestinalis* [138] or in protostome lineage of mollusc *Cr. gigas* [174]. However, no correlation with repetitive DNA was shown in another protostome lineage of insects such as silkworm, honey bee or parasitoid wasp [132,168,169].

In T. castaneum and *D. melanogaster*, a preferential methylation of intergenic regions in comparison to gene bodies was described. Intergenic regions accommodate many functional elements, like transcriptional regulatory sites or intergenic mobile elements and repeats. Moreover, only 10 and 50 % of cytosines are methylated in these sites in *T. castaneum and D. melanogaster*, respectively, possibly being temporally or spatially restricted. This methylation dynamics is suggestive of possible involvement in transposable elements silencing or regulation of gene expression via methylation of respective transcriptional regulatory sites [171].

Detailed investigation of methylation in the fruit fly shows different 5mC enrichment in various types of genomic repeats; there was no evidence of methylation in transposable elements while an increased abundance of 5mC was observed in simple sequence repeats in *Drosophila* [171]. In red flour beetle, a significant correlation of methylation with repetitive elements was described. In particular, selective methylation of different classes of repetitive elements was observed and simple repeats were amongst the most highly methylated in agreement with a survey made in fruit fly [166].

1.3.2.4 Enigmatic role of Non-CpG methylation

Examination of DNA methylation has been focusing on the characterization of CpG methylation as a major substrate for methyltransferase activity in Eukaryotes and there are many credible indications elucidating its function in the vertebrates or in plants. Cytosines methylated outside the CpG context were long considered rather obscure due to their rare occurrence and an absence of theories proposing the mechanism and role of this modification. First observation of this type of methylation was reported in plants and was long considered a plant-specific phenomenon. The mechanism of non-CpG methylation is largely unknown; however, its dynamics and management were at least partially elucidated by an identification of the respective demethylation enzyme in plant genome [175,176]. Further investigation described this modification in mammals, although confined to embryonic cells, stem cells, or gametes [148,177]. The estimates of non-CpG methylation were ranging between 15-52 % in embryonic cells and rather negligible levels were described in somatic tissues.

Recent advances in the evaluation of genome methylation sequencing projects and development of unbiased epigenetic approaches, however, uncovered a much wider extent of this modification in different organisms and tissues. For example, updated estimates of non-CpG methylation raised to 25-35 % in various mammalian cell types and share more than half of the detected methylations of DNA in neural tissue [143]. Although the function of this type of modification is still an object of investigation [178–183] its existence is only now starting to be recognized and appreciated.

Reports of cytosine methylation in non-CpG context are considerably less frequent in invertebrates and are mainly restricted to model species. Non-CpG 5mCs distribution was colocalized with that of methyl-CpG and is also restricted to gene bodies of invertebrate chordate model *Ci. intestinalis* avoiding the regions of CpG rich stretches, resembling those of vertebrate CpG islands in promoter regions [118]. Investigation of DNA methylation in the insect model described a presence of 5mC in non-CpG doublets also in *D. melanogaster* [184,185].

Ultra-low methylation levels observed in many insect species were usually under the detection limit of formerly used methods for detection of methylation. This also disabled more detailed and unbiased characterization of methylation in these species, if detected. Recent development of in-depth bisulphite sequencing method in

combination with unbiased approaches for 5mC detection enabled re-evaluation of previously published data on methylation in insect [166].

An updated detailed insight into the non-CpG methylation in insects was reported in *D. melanogaster*. Not only was cytosine methylation found predominantly in non-CpG regions, but it was also associated with repetitive DNA and other intergenic regions while completely absent from gene bodies. An observation of negatively correlated gene body methylation and transcriptional activity is also in sharp contrast with the methylation pattern reported in other invertebrates [171].

Similarly, re-evaluation of methylation in *Tribolium castaneum* revealed a presence of 5mC in its genome, which is in contradiction with former studies reporting the absence of methylation in *Tribolium* genome [113,170]. New data confirmed ultra-low methylation predicted for *T. castaneum* ranging from 0.08-0.26 % of all genomic cytosines. However, consistently with *D. melanogaster* [171] and a red paper wasp *Polistes canadensis* [186], the predominance of non-CpG methylation was observed in the *Tribolium* genome comprising nearly 75 % of all methylated cytosines. Moreover, a methylation distribution was more frequent in intergenic regions and introns which is in contradiction to methylation of the coding portion of genome typical for the majority of eukaryotic species studied so far [113]. Additionally, a gradient of methylation of both CpGs and non-CpGs was observed at the flanking regions and in introns of the gene body which is again in agreement with observations in *Drosophila* methylation pattern but is in contradiction with most other Eukaryotes.

Different distribution of non-CpG and symmetrical CpG methylation was also reported in the genome of *T. castaneum*. This indicates that methylation of cytosines in these two instances is differentially regulated and is most probably catalysed by different enzymes [166].

1.3.3 Writers and Readers of Eukaryotic methylation

1.3.3.1 5mC “writer” proteins (DNA methyltransferases)

Methyltransferases (METTs) in Eukaryotes are classified into a family of proteins characterized by their common catalytic domain. The catalytic activity of DNMTs originates in the prokaryotic enzymes involved in the defensive restriction-

modification system [99]. These type II restriction METTs were adopted by eukaryotic cell and upon fusion with a eukaryotic gene with a chromatin-binding function, they evolved into the diversity of sequence families present in Eukaryotes today. Currently, there are 6 METTs sequence subfamilies that diverged from its common prokaryotic ancestor in Eukaryotes [187] of which DNMT1 and DNMT3 METTs are canonical enzymes catalysing the most extensive maintenance and de novo methylation of genomic DNA in most animals. DNMT2 does not exhibit DNA methylation activity; however, it shares a highly conserved C-terminal catalytic domain with remaining DNMTs and is present across Eukaryotes. All METTs family members share N-terminal regulatory and C-terminal catalytic domains both of which differ in their subdomain architecture in each family member. Figure 7 shows a domain architecture of DNA METTs shared among members of Metazoa.

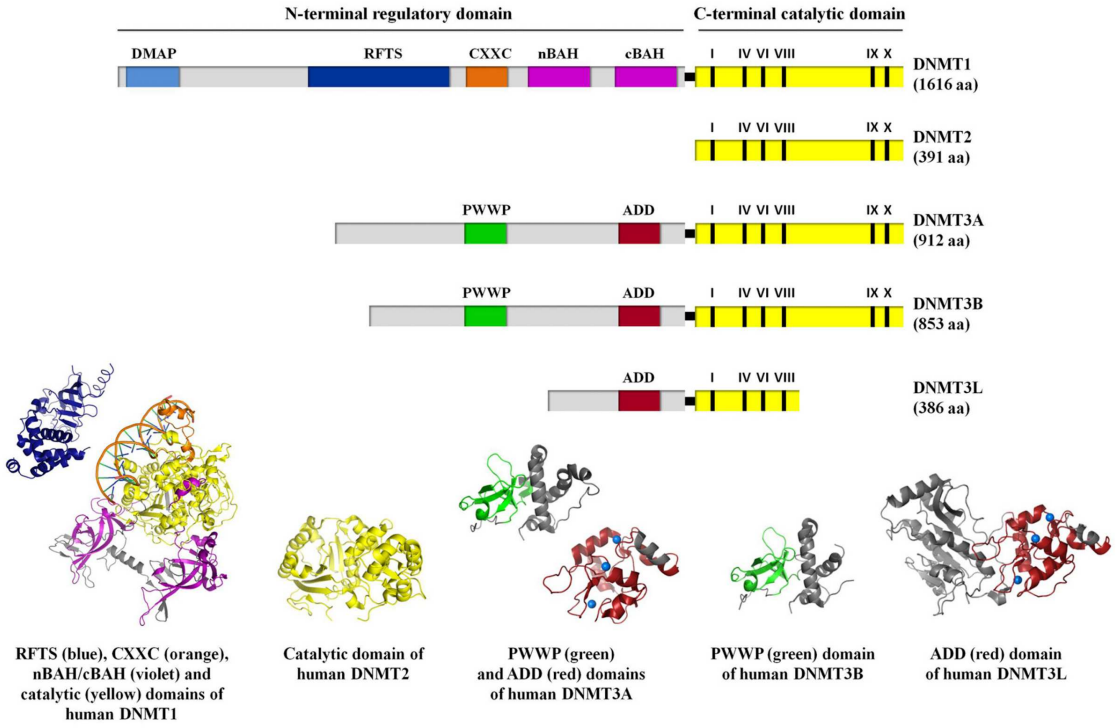
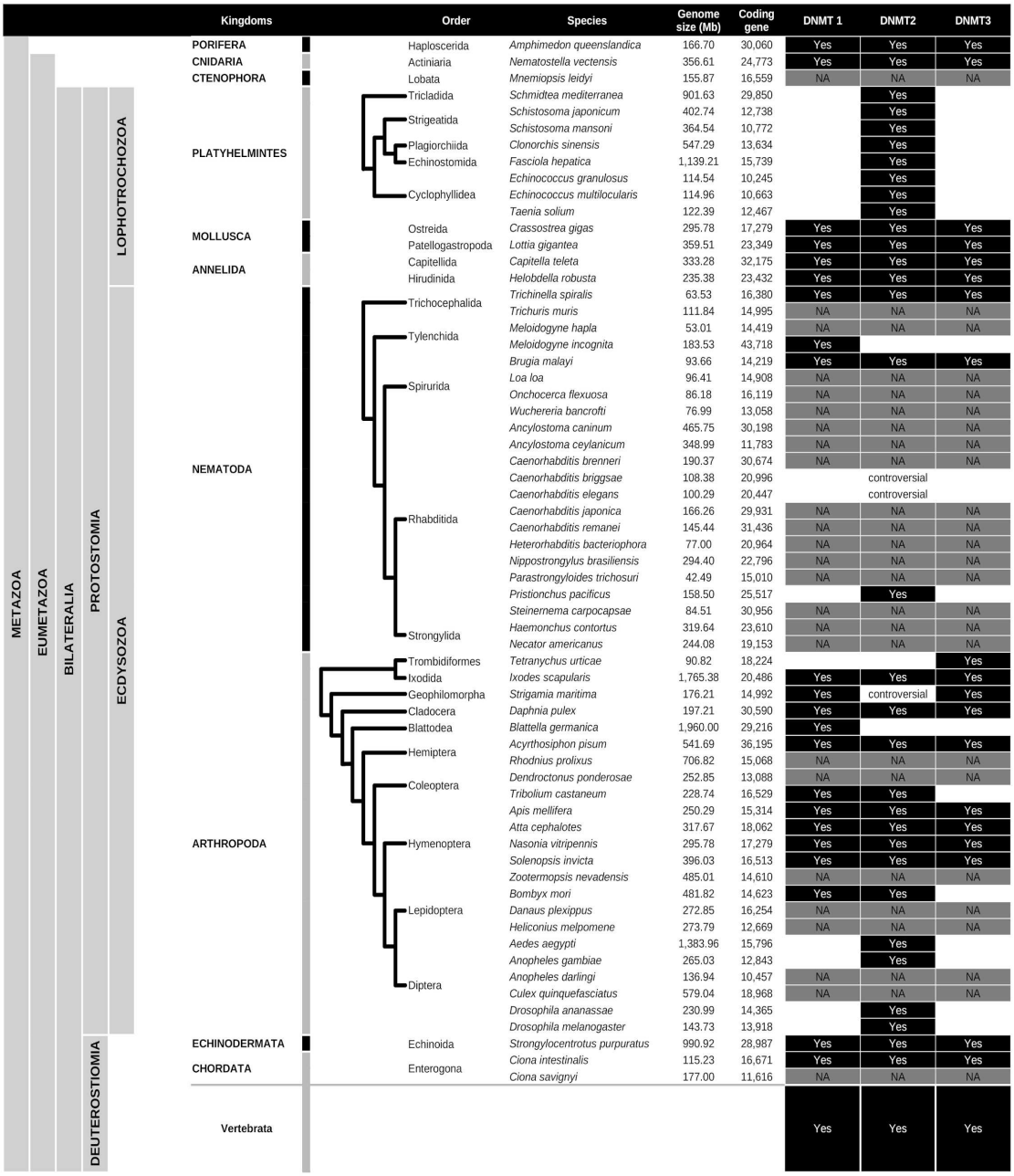


Figure 7. Domain architecture of human DNA methyltransferases. DMAP, DNA methyltransferase associated protein; RFTS, replication foci targeting sequence; CXXC, bromo-adjacent homology (nBAH/cBAH); PWWP, Pro-Trp-Trp-Pro domain; ADD, DNMT3L histone H3 binding domain [188].

All three enzyme classes show a different degree of conservancy as by their evolution within Eukaryotes. The indispensability of DNMT1 and DNMT3 turned out to be

exclusive to vertebrate lineage as many completed invertebrate genome sequencing projects revealed several losses of one or both METTs only retaining DNMT2 (Fig 8). Interestingly, both DNMT1 and DNMT3 diverged into several paralogous sequences within Metazoa while DNMT2 was retained as a single-copy gene across all taxa, where present (Fig. 9).



Kingdoms	Order	Species	Genome size (Mb)	Coding gene	DNMT 1	DNMT2	DNMT3
PORIFERA	Haploscerida	*Amphimedon queenslandica*	166.70	30,060	Yes	Yes	Yes
CNIDARIA	Actiniaria	*Nematostella vectensis*	356.61	24,773	Yes	Yes	Yes
CTENOPHORA	Lobata	*Mnemiopsis leidyi*	155.87	16,559	NA	NA	NA
PLATYHELMINTES	Tricladida	*Schmidtea mediterranea*	901.63	29,850		Yes	
	Strigeatida	*Schistosoma japonicum*	402.74	12,738		Yes	
		Schistosoma mansoni	364.54	10,772		Yes	
	Plagiorchiida	*Clonorchis sinensis*	547.29	13,634		Yes	
	Echinostomida	*Fasciola hepatica*	1,139.21	15,739		Yes	
		Echinococcus granulosus	114.54	10,245		Yes	
	Cyclophyllidea	*Echinococcus multilocularis*	114.96	10,663		Yes	
		Taenia solium	122.39	12,467		Yes	
MOLLUSCA	Ostreida	*Crassostrea gigas*	295.78	17,279	Yes	Yes	Yes
	Patellogastropoda	*Lottia gigantea*	359.51	23,349	Yes	Yes	Yes
ANNELIDA	Capitellida	*Capitella teleta*	333.28	32,175	Yes	Yes	Yes
	Hirudinida	*Helobdella robusta*	235.38	23,432	Yes	Yes	Yes
NEMATODA	Trichocephalida	*Trichinella spiralis*	63.53	16,380	Yes	Yes	Yes
		Trichuris muris	111.84	14,995	NA	NA	NA
	Tylenchida	*Meloidogyne hapla*	53.01	14,419	NA	NA	NA
		Meloidogyne incognita	183.53	43,718	Yes		
	Spirurida	*Brugia malayi*	93.66	14,219	Yes	Yes	Yes
		Loa loa	96.41	14,908	NA	NA	NA
		Onchocerca flexuosa	86.18	16,119	NA	NA	NA
		Wuchereria bancrofti	76.99	13,058	NA	NA	NA
		Ancylostoma caninum	465.75	30,198	NA	NA	NA
		Ancylostoma ceylanicum	348.99	11,783	NA	NA	NA
		Caenorhabditis brenneri	190.37	30,674	NA	NA	NA
		Caenorhabditis briggsae	108.38	20,996		controversial	
		Caenorhabditis elegans	100.29	20,447		controversial	
	Rhabditida	*Caenorhabditis japonica*	166.26	29,931	NA	NA	NA
		Caenorhabditis remanei	145.44	31,436	NA	NA	NA
		Heterorhabditis bacteriophora	77.00	20,964	NA	NA	NA
		Nippostrongylus brasiliensis	294.40	22,796	NA	NA	NA
		Parastrongyloides trichosuri	42.49	15,010	NA	NA	NA
		Pristionchus pacificus	158.50	25,517		Yes	
		Steinernema carpocapsae	84.51	30,956	NA	NA	NA
	Strongylida	*Haemonchus contortus*	319.64	23,610	NA	NA	NA
		Necator americanus	244.08	19,153	NA	NA	NA
ARTHROPODA	Trombidiformes	*Tetranychus urticae*	90.82	18,224			Yes
	Ixodida	*Ixodes scapularis*	1,765.38	20,486	Yes	Yes	Yes
	Geophilomorpha	*Strigamia maritima*	176.21	14,992	Yes	controversial	Yes
	Cladocera	*Daphnia pulex*	197.21	30,590	Yes	Yes	Yes
	Blattodea	*Blattella germanica*	1,960.00	29,216	Yes		
	Hemiptera	*Acyrthosiphon pisum*	541.69	36,195	Yes	Yes	Yes
		Rhodnius prolixus	706.82	15,068	NA	NA	NA
	Coleoptera	*Dendroctonus ponderosae*	252.85	13,088	NA	NA	NA
		Tribolium castaneum	228.74	16,529	Yes	Yes	
		Apis mellifera	250.29	15,314	Yes	Yes	Yes
	Hymenoptera	*Atta cephalotes*	317.67	18,062	Yes	Yes	Yes
		Nasonia vitripennis	295.78	17,279	Yes	Yes	Yes
		Solenopsis invicta	396.03	16,513	Yes	Yes	Yes
		Zootermopsis nevadensis	485.01	14,610	NA	NA	NA
	Lepidoptera	*Bombyx mori*	481.82	14,623	Yes	Yes	
		Danaus plexippus	272.85	16,254	NA	NA	NA
		Heliconius melpomene	273.79	12,669	NA	NA	NA
	Diptera	*Aedes aegypti*	1,383.96	15,796		Yes	
		Anopheles gambiae	265.03	12,843		Yes	
		Anopheles darlingi	136.94	10,457	NA	NA	NA
		Culex quinquefasciatus	579.04	18,968	NA	NA	NA
		Drosophila ananassae	230.99	14,365		Yes	
		Drosophila melanogaster	143.73	13,918		Yes	
ECHINODERMATA	Echinoida	*Strongylocentrotus purpuratus*	990.92	28,987	Yes	Yes	Yes
CHORDATA	Enterogona	*Ciona intestinalis*	115.23	16,671	Yes	Yes	Yes
		Ciona savignyi	177.00	11,616	NA	NA	NA
Vertebrata					Yes	Yes	Yes

Figure 8. Shows the distribution of all DNMTs across Metazoan invertebrates. The data presented in the figure are sourced from Aliaga, 2019 [100]. The missing data were supplemented from Ensembl genome database (http://ensemblgenomes.org/) and WormBase ParaSite (https://parasite.wormbase.org/).

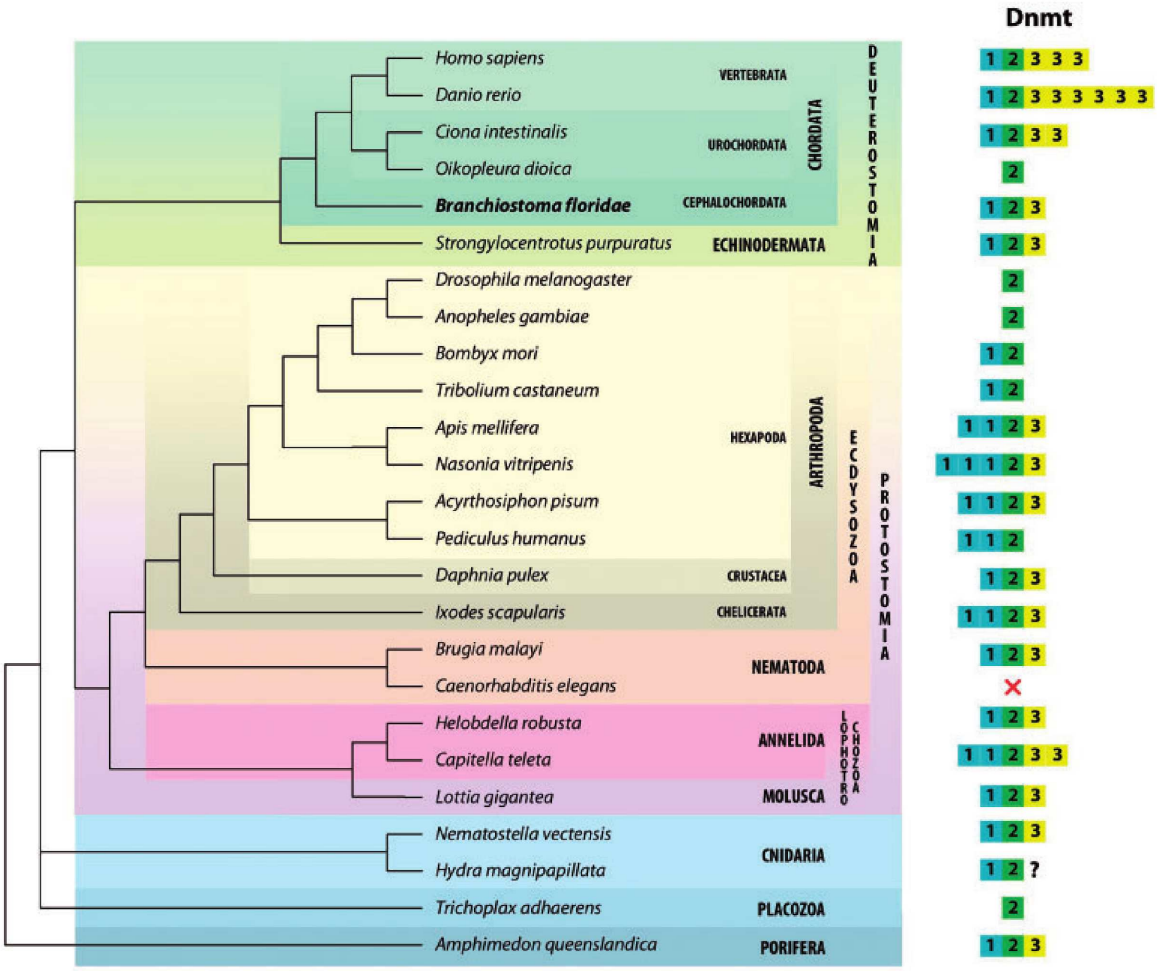


Figure 9. The distribution of DNMT paralogues in different Eukaryotic lineages. The DNMT groups are shown for only some selected animal species. Numbered boxes represent DNMT1, DNMT2, or DNTM3 orthologue. The number of boxes indicates the number of each orthologue in the respective species [189].

DNMT1

DNMT1 is a Eukaryotic enzyme whose prominent role was shown in a maintenance of DNA methylated state established by *de novo* methylating DNMT3 during early embryonic development in most vertebrates [190]. In concert with UHRF1 protein, DNMT1 targets the hemimethylated DNA strand following the replication fork [191]. Two paralogous genes DNMT1a and DNMT1b diversified in some fish species [192]. Independently, the diversification of DNMT1 into two paralogous genes (identically denoted DNMT1a and DNMT1b) also occurred in some insect orders (Fig. 10) [105]. The sequence and domain organization of DNMT1 is distinct from remaining transferases and the presence and function of particular domains reflect its

methylation maintaining function. Replication focus targeting sequence (RTFS) contains a UHRF1 binding site which targets the catalytic domain of DNMT1 to the hemimethylated loci produced by the replication fork. CXXC domain interacts with unmethylated CpG substrates holding it in an appropriate position to the active site of the catalytic domain. The mechanism of BAH1 and BAH2 domains participation in DNMT1 activity has not been described to date, although the deletion of both BAH domains disabled targeting of DNMT1 to the replication foci even in the presence of RTFS domain. Moreover, mutants of lysine-glycine rich stretch (GK repeats) in C terminus of BAH2 showed *de novo* deposition of methylation by DNMT1, which implies the role of BAH domain in the maintenance function of DNMT1 and is implicated in the regulation of imprinting in mammals [193]. This also suggests potential promiscuity of DNMT1 function and by extent may provide an alternative mechanism for *de novo* methylation in invertebrate species missing DNMT3 gene in their genome.

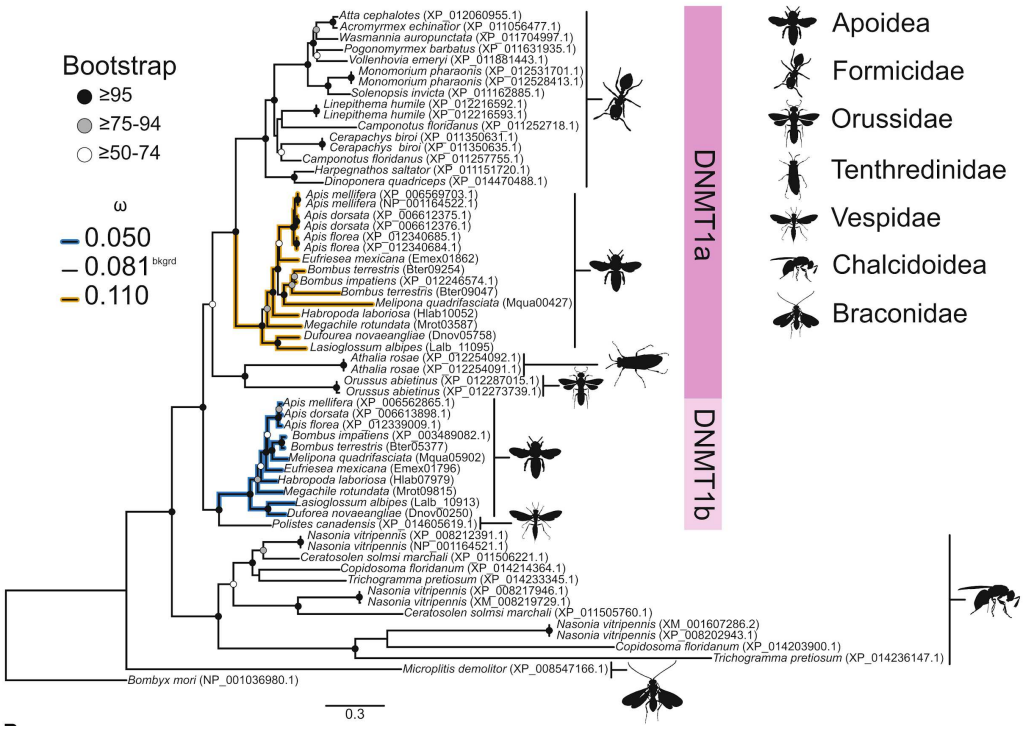


Figure 10. Diversification of DNMT1a and DNMT1b orthologues within several insect orders as by phylogenetic reconstruction. Hypothetical secondary loss of DNMT1b orthologue occurred in *Formicidae* (data not shown). *Bombyx mori* DNMT1 was used to root the tree [105].

DNMT2 (TrDMT1)

DNMT2 is the most conserved and widespread enzyme with DNA METTs catalytic domain in Eukaryotes [194]. Although originally considered a DNA methylating enzyme, DNMT2 lacks a DNA methylating activity, which led to the identification of C38 position of the tRNA Asp as its main substrate in *A. thaliana*, *M. musculus*, and *D. melanogaster* [195]. Later investigation revealed tRNA Gly, tRNA Val, and under specific conditions also tRNA Glu as additional substrates in these organisms [196–198]. Reports of DNMT2-dependent methylation of 5mC DNA are still controversial [171,199–201]. Similarly unclear is also the importance of DNMT2 activity for the cell and the organism. Besides the suggested role in development in vertebrates, based on knockdown of DNMT2 of zebrafish [202], most DNMT2 knockouts did not display any discernible alterations from physiological phenotype [195,203]. Due to a participation of tRNA in protein translation, the role of DNMT2 methylation was suggested in the regulation of protein synthesis, which was further functionally verified in mouse [204,205]. Several proposed theories arise from an investigation of *D. melanogaster* DNMT2 knockout mutants. Without any visible phenotypes, a detailed investigation revealed a significant overexpression of retrotransposons in *Drosophila* embryos [206]. Consistently, a suppression of tandem repeat silencing was also dysregulated in DNMT2 mutant flies [207]. Collectively, this suggests that DNMT2 activity plays a role in the regulation of repetitive sequences activity and propagation in *D. melanogaster* [208].

Despite its substrate specificity, DNMT2 catalytic domain is homologous to that of other DNA methylating enzymes [209]. In line, multiple sequence alignment of DNA METT genes, DNMT2, and other RNA-cytosine C5 METT genes showed little homology of DNMT2 with RNA METTs. This suggests a common ancestry of DNMT2 with DNA methylating enzymes [194]. High sequence homology shared with DNMT1 subfamily proposed that DNMT2 is an ancestral gene of DNMT1 and DNMT2 in Eukaryotes as RNA methylation activity was considered more ancient. However, a phylogenetic analysis of Eukaryotic DNMT sequences do not support such conclusions. Instead, it seems that all 3 gene family members evolved from distinct Prokaryotic ancestors (Fig 11A), which is also in line with a phylogenetic reconstruction of separate sequence clusters of all 3 Eukaryotic DNMT members and corresponding DNMTs in Prokaryotic genome (Fig 11B.)

It is peculiar that DNMT2 is the only single-copy gene of the DNMT family of enzymes within Eukaryotes, which implies a high degree of conservation in Eukaryotic genomes. Attempts to explain the function of this enzyme brought inconclusive results [208] and thus the role of this enzyme for a genome epigenetic regulation still awaits elucidation.

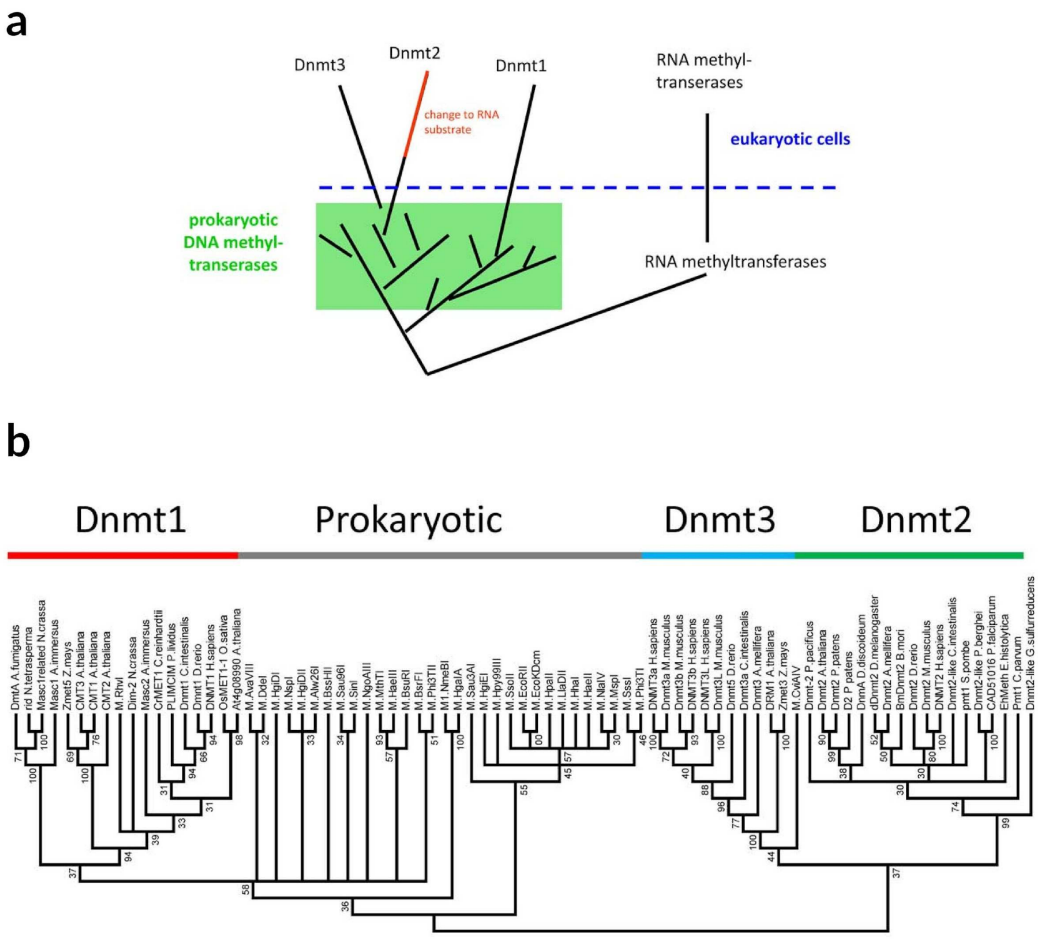


Figure 11. Phylogenetic reconstruction of DNA methyltransferase catalytic domain protein family. (a) Consensus phylogenetic model indicating an independent origin of DNMT1, 2, and 3 in the prokaryotic DNA methyltransferases ancestors. The phylogenetic model also depicts an independent phylogeny and origin of DNMT2 and other RNA methyltransferases. (b) Phylogenetic tree showing relationship of prokaryotic and eukaryotic DNA methyltransferases and DNMT2 [194].

DNMT3

Along with DNMT1, also DNMT3 is considered a canonical DNA methyltransferase that is responsible for *de novo* methylation upon epigenetic reprogramming that occurs early in the embryonic development. Of the 3 family members, DNMT3 seems to be the most evolutionarily diverse and most dispensable gene as evident in the diversification of DNMT3 orthologues in many vertebrates and independent losses in several lineages of invertebrate animals (Fig. 9). The diversification of DNMT3 in vertebrates was suggested to be a successive event after whole-genome duplication that occurred at the nascence of vertebrate lineage. Function of the DNMT3 paralogues largely overlaps; however, a preferential methylation of single-copy genes or repetitive sequences by different DNMT3 paralogues was described to some extent in mouse [210]. Such diversification in DNMT3 repertoire enabled fine-tuning of genome methylation pattern underlying embryonic developmental regulation in the respective organisms.

The DNMT3 family members share 3 conserved domains. Besides the C-terminal conserved catalytic domain, shared by all DNMT enzymes, DNMT3 contains N-terminal regulatory PWWP chromatin binding domain (binding H3K36me3 histone tail) characterized by a conserved proline-tryptophan motif, and ADD domain interacting with histone H3 tails unmethylated at lysine K4 and also contributes to a chromatin binding and comprises 3 subdomains: cysteine-rich GATA-like zinc finger, PHD finger zinc finger domain, and a C-terminal a-helix. Recently, an interaction of ADD with other binding partners involved in the epigenetic regulation was also described, which could affect DNMT3 activity [211].

1.3.3.2 5mC “reader” proteins

The importance of DNA methylation can be fully appreciated only in the context of network of proteins that are able to read this epigenetic marks and mediate an appropriate cellular response. The most prominent functions in the 5mC recognition are conferred by two protein families containing conserved 5mC recognition domains; a family of methyl-CpG binding domain (MBD) proteins and SRA domain family of proteins, which is by large represented by UHRF1 protein.

UHRF

Maintenance of an established methylation pattern and its heritability is underlaid by a cooperation of many important interplayers of DNA methylation apparatus. The essential step in methylation maintenance includes a recognition of hemimethylated state which is ultimately secured by E3 ubiquitin-protein ligase (UHRF1) protein besides DNMT1. Additionally, UHRF1 exhibits a functional promiscuity in epigenetic regulation as it also binds to H3K9me2/3, a marker of newly assembled nucleosomes, which also associates UHRF1 binding with pericentric heterochromatin. Both binding activities were found important for the reliable maintenance of DNA methylation pattern. Link between the two binding activities of UHRF1 has not been fully explained; however, mutations in Tudor and SRA domains (Fig. 12) responsible for binding of H3K9me2/3 and hemi-mCpG substrates respectively, showed that both domains are important for targeting of DNMT1 to the replication foci [191].

UHRF is a fairly conserved protein found across Eukaryotes. As its activity promotes DNMT1 mediated maintenance of DNA methylation, it is not surprising then that its phylogenetic distribution in Eukaryotes is correlated with distribution of DNMT1 [212,213]. The only known paralogous gene, UHRF2 was identified in vertebrate genome and its role was implicated in the regulation of active demethylation [214].

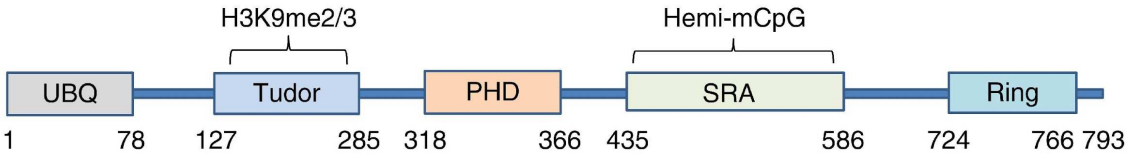


Figure 12. UHRF proteins domain architecture. UBQ, Ubiquitin domain; PHD, Plant Homeodomain; SRA, SET and RINGER finger associated domain; Ring, Really Interesting New Gene [191].

MBD proteins

Unlike SRA protein family, the members of MBD family represent a large and diverse group of proteins whose divergence follows different Eukaryotic lineages, mainly in animals.

Based on the identification of specific modules in the identified MBD proteins, differential activities as well as their classification were proposed. Currently, three MBD classes are recognized (Fig. 13) [215]. Most MBD proteins were found in

multiple subunit protein complexes where they secure a recognition of a target genomic 5mC loci by the respective protein complex of the epigenetic apparatus. The activities of MBD proteins were implicated in diverse epigenetic functions where they perform a readout of DNA methylation in specific genomic contexts and recruit various chromatin remodellers and histone deacetylases and methylases that can be associated for instance with gene repression [216]. Several case examples describe the linkage of MBD proteins with specific functions in epigenetic regulation. Group I MBD proteins, containing PHD (Plant Homeodomain), were assigned a role in proper function of histone acetyltransferase forming a linkage between DNA methylation and histone acetylation [217]. Members of Group II MBD proteins were identified in nucleolar remodelling complex, functioning in a repression of RNA polymerase I transcription [218].

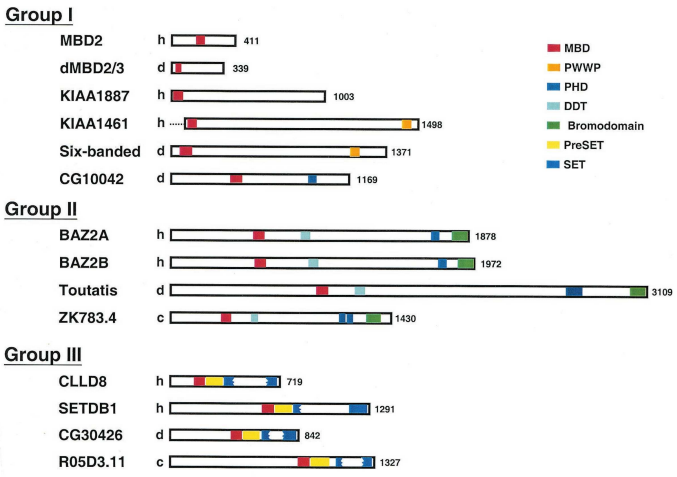


Figure 13. Classification and domain architecture of MBD proteins. The classification is based on the arrangement of domains and specific sequence motifs indicated by coloured boxes in each protein sequence illustration. h, human; d, Drosophila; c, C. elegans; MBD; Methyl-CpG-binding domain; PWWP, Pro-Trp-Trp-Pro motif domain; PHD, Plant Homeodomain; DDT, DNA binding homeobox and Different Transcription factor domain; PreSET, N-terminal to some SET domain; SET, Su(var)3-9, Enhancer-of-zeste and Trithorax domain. [215].

TET proteins

Key players in DNA methylation dynamics and management are undoubtedly members of TET (Ten-Eleven Translocation) family of dioxygenases that are in charge of conversion of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) hence representing a major active form of demethylation in eukaryotic cells. The process initiated by TET demethylase starts by a TET mediated oxidation of 5mC forming

5hmC which is subsequently metabolized to 5-formylcytosine and 5-carboxylcytosine. Unmodified cytosine base is eventually introduced during replication coupled with base excision repair. Although several other mechanisms underlie the process of active demethylation, the TET mediated demethylation is the most prominent one.

The family of TET proteins exhibit distinct sequence features (Fig. 14). The C-terminal catalytic domain comprises of DSBH core which executes the 5mC oxidation and a Cys-rich insert which wraps around the catalytic core and stabilizes the TET-DNA interaction. Interestingly, the catalytic domain alone is able to localize to nucleus and perform 5mC oxidation itself [219]. The CXXC domain at N-terminus reportedly exhibits a chromatin-binding activity, although this statement is controversial due to contradictory conclusions of different studies [220,221]

Phylogenetic investigation classified TET proteins into the family shared with JBP (J-binding protein) group of proteins identified in Kinetoplastids. Large-scale phylogenetic screening further revealed JBP homologues in diverse Eukaryotes and bacteriophages.

Three paralogues of TET protein were identified in mammalian genomes (Fig. 14) and their activity was proven crucial for proper embryonic reprogramming of developing embryo. Its increased activity was also observed in neural tissue of adult individuals [222]. Recent studies described a TET1-mediated demethylation involvement in a proper function of DNA repair mechanism, which extends to the global effect of genomic stability.

Additionally, the activity of TET1 was traced to the promoters of active genes where they were implicated in the maintenance of unmethylated status of CpG islands in the active genes. This predetermines the DNA methylation as a dynamic tool for control over gene expression [223].

Most invertebrates, on the other hand, share a single TET homologue and its specificity seems to be more relaxed with respect to the selection of methylated substrate [125]. This seems to be mainly the case of insects with reduced or absent DNA methylation levels where TET homologues took over the role of 6-methyladenine (6mA) DNA and 5mC mRNA demethylases as seen in *D. melanogaster* [224,225]. In line with vertebrates, the activity of TET protein was also observed in neural tissue in *Drosophila* and its role was specifically implicated in an axonal development in neural tissue in developing larva [226].

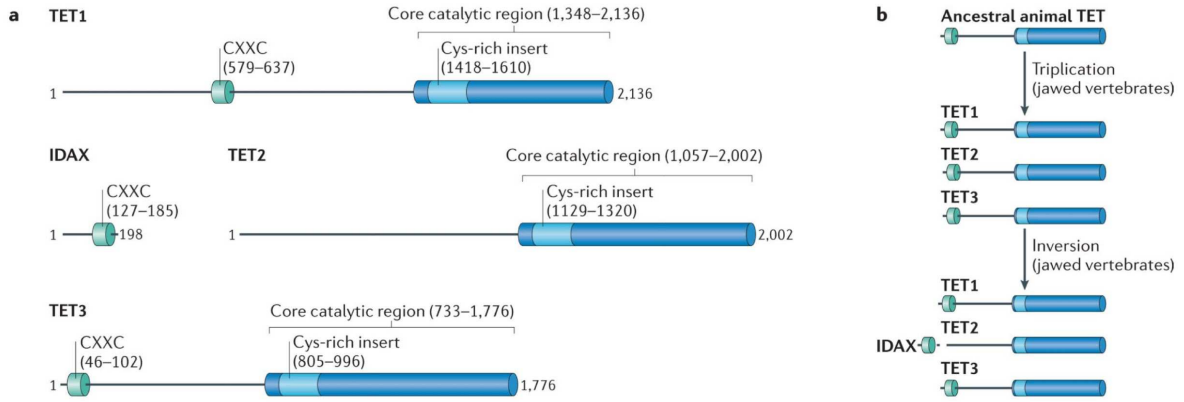


Figure 14. Characterisation of TET protein family domains. (a) Domain structure of conserved mammalian orthologues of TET protein family. (b) Evolution of domain structure in TET proteins of Metazoa lineage [227].

1.3.4 N6-adenine DNA methylation

Methylation of adenine in DNA is a modification widely distributed in prokaryotic genomes as a part of their defensive restriction-modification system. Its presence and function in Eukaryotes is still largely undescribed and rather enigmatic. In fact, the methylation of adenine was originally described in different RNA classes in eukaryotic cell. Its function was ascribed to the regulation of gene expression on the level of both transcription and translation. Defect in methylation of adenine in RNA produced more or less severe phenotypes in brewer‘s yeast, fruit fly, arabidopsis, or human cells [228–231].

Discovery of 6mA in investigated eukaryotic genomes was hindered by an availability of only low-sensitivity techniques for detection of DNA methylation. Recent development of more sensitive techniques for 6mA detection encouraged a research in this area. However, reports on this type of epigenetic modification are still sparse and restricted to only few model species. Fig. 15 shows levels of 6mA and 5mC across tree of life [232]. Underrepresentation of eukaryotic taxa in the tree indicates that adenine methylation is by large a privilege of prokaryotic and archaeal domains. It also

supports the observation about low levels of 6mA in Eukaryotes and demonstrates the difficulty of 6mA detection by conventional methods.

Thorough investigation of methylation in *Ca. elegans* using more sensitive techniques revealed ultra-low levels of 5mC [233] that were under the detection limits of conventional methods and thus neglected in previous studies (e.g., [234]). 6mA was also detected in *Ca. elegans* genome, and even in higher amounts compared to 5mC [235,236]. Consistently, the presence of N6 adenine methyl-transferase (DAMT-1) was confirmed in the *Ca. elegans* genome and DAMT-1 knockout individuals exhibited increased 6mA levels and sterility [236].

Similarly, high-sensitivity detection techniques were also able to detect low amounts of both adenine and cytosine methylation in the genome of *Drosophila* with 6mA levels being lower by orders of magnitude to those of 5mC [224,237]. The methylation of adenine was found to be important in the regulation of transposition as seen in an enrichment of 6mA in transposon gene bodies and an increased transposon transcription in DMAD mutant flies. Additionally, this also suggest an involvement of adenine methylation in gene expression regulation. Oscillation of 6mA levels were observed throughout embryonic development in *D. melanogaster* being 70 x higher in early development, which also corresponded to an activity of adenine DNA demethylase (DMAD), a TET protein homologue which in *Drosophila* took over the role of methyl-adenine demethylase. DMAD activity was also proven essential for embryogenesis as DMAD mutants produced phenotypes lethal in pupa stage.

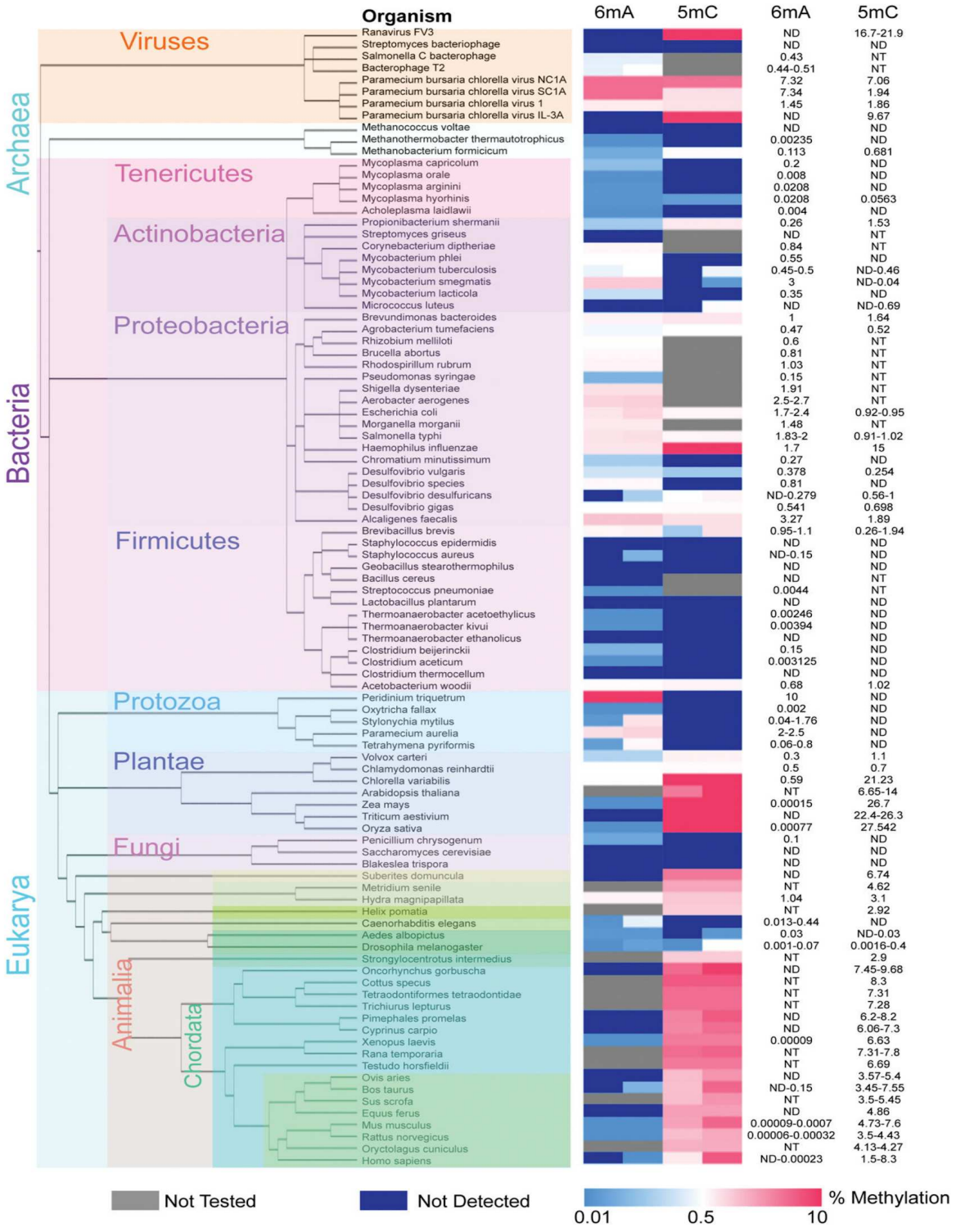


Figure 15. Comparison of RNA and DNA methylation levels in all life domains. [232].

An investigation of enzymatic machinery underlying DNA adenine methylation is widely biased towards the characterization of the adenine demethylating enzymes with adenine methyltransferase being largely overlooked. DNA adenine demethylating enzyme was successfully identified in *Ca. elegans* [236]. Further characterisation of this protein classified it to the protein family of RNA 6mA methyltransferases [238]. This family is characterized by the presence of a conserved MT-A70 domain (Fig. 16) which is present in a wide range of proteins found across Eukaryotes [239]. Besides a variety of paralogous proteins identified in the vertebrates, two orthologues denoted DAMT-1 and METTL4 were also found in insects. METTL4 orthologue, phylogenetically more conserved in Eukaryotes [240] was also functionally characterized and its DNA adenine methylation activity was confirmed in *Bombyx mori* METTL4 knockout cells.

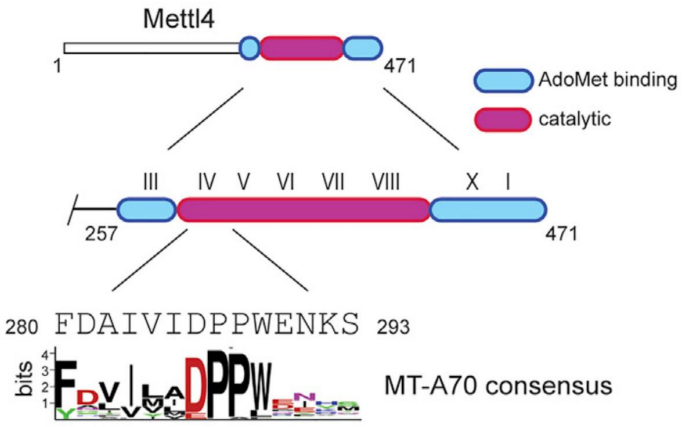


Figure 16. Structure of MT-A70 domain in METTL4 orthologue of adenine DNA methylating protein family [241].

Distribution of 6mA within vertebrate genomes shows high variation and differs in each newly investigated vertebrate species. Screening of several representative vertebrates reported a preferential occupancy of 6mA in introns, which was also confirmed tissue-wide [242]. Conversely, an enrichment of 6mA in zebrafish was observed in exons and repetitive elements while avoiding intronic, intergenic, and promoter regions [243].

A range of different functions was also attributed to the presence of genomic 6mA and to the activity of the corresponding methylation enzyme in the vertebrates. The effects of adenine methylation range from chromosome-specific gene silencing in

human embryonic stem cells [244] to the positive correlation of 6mA levels with repetitive elements expression decreasing with the progress of development in zebrafish [243]. These initial case reports ascribe 6mA rather incoherent and species-specific functions without any hints of more systematic functional classification. It is the future investigation of 6mA in other species that will shed more light on the significance of this epigenetic modification for Eukaryotic cell.

1.3.5 N6-adenine RNA methylation

Unlike in DNA, adenine in RNA is commonly methylated in Eukaryotes, with mRNA accounting for the majority of the total RNA population in the cell targeted by methylation. In fact, methylation of adenine in RNA is one of the most common RNA epigenetic modifications widespread in Eukaryotes.

The reports of RNA adenine methylation had long been very sparse and fragmented until the recent development of epitranscriptomic techniques which enabled large-scale detection and characterization of this epigenetic modification in numerous species. Thus, the majority of information that was collected about 6mA is a matter of past decade when a 6mA mapping method was published [245,246].

The distribution of 6mA is most prominent in the 3‘ UTR and near stop codons of mRNA molecules. Targeting of mRNA is also very selective with only specific subpopulations of transcripts being methylated. Most transcripts receive a single methylation modification of adenine and the targeted adenines appear to be constitutively methylated in different tissues or cell lines. Functional annotation of methylated transcripts ascribed this epigenetic modification a role in developmental regulation and cell fate specification [245–247]. Consistently, highly orthologous or housekeeping gene transcripts were depleted of 6mA [248].

First reports ascribed 6mA a role in mRNA stability using radioisotopic labelling and monitoring a half-life of 6mA containing transcripts in comparison to non-methylated molecules [249]. This was complemented by later observation that depletion of Df2 cytosolic 6mA binding protein involved in destabilization of methylated mRNA prolongs a half-life of the target mRNA. Contrastingly, silencing of RNA methyladenine transferase is associated with increased mRNA stability, which implies more complex mechanisms underlying this aspect of 6mA function.

More conclusive findings were made with linking 6mA to the translation upregulation. The effect of 6mA is always coupled with the interaction of cytosolic 6mA reader proteins which directly or indirectly recruit elongation factors or other translation associated molecules and initiate an assembly of ribosomal complexes and translation activation [250].

In *D. melanogaster*, an intronic 6mA was implicated in the regulation of splicing in gene for sex determination [251–253]. However, attempts to correlate 6mA distribution with exon-intron structure failed to bring congruent results [254–256].

The enzyme that catalyses adenine methylation in RNA substrates is a multi-subunit protein complex. In Eukaryotes, a range of different adenine methyltransferase protein complexes developed that recognize different RNA substrates. For example, 28S rRNA is methylated by ZCCHC4 [257,258] 18S rRNA by the METTL5-TRMT112 protein complex, and U6 snRNA methylation is secured by METTL16 [259]. The methylation of adenine in mRNA is a product of METTL3-METTL14 heterodimer activity. This protein complex is formed by multiple subdomains of which the most important is METTL3 partner representing a catalytic subunit and METTL14, an allosteric adaptor protein which contains an RNA-binding site and functions as an allosteric activator of METTL3 catalytic activity. Both orthologues share common MT-A70 (MTase) domain which harbours catalytic site of the proteins (Fig. 17). The resulting protein domains, one of which is a SAM-binding domain, carry a disrupted SAM binding motif in METTL14 partner and thus makes it catalytically inactive.

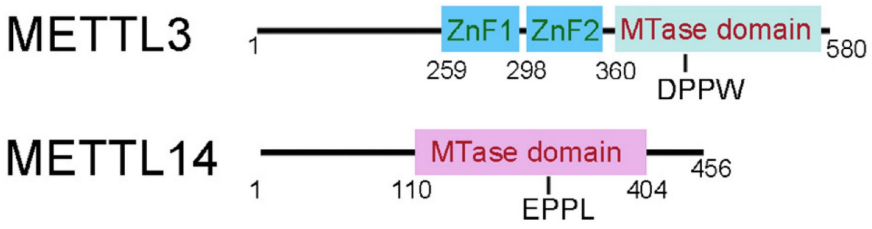


Figure 17. Domain structure of the two most important subunits of METTL3-METTL14 adenine RNA methylation complex [260].

1.3.6 Methylation in chelicerates

It is evident that cytosine DNA methylation in invertebrates is not an essential modification and that processes regulated by DNA methylation crucial for a survival in the vertebrates play a subsidiary role or are regulated by different mechanisms in invertebrates. This is evident upon inspection of 5mC levels in DNA of many invertebrate organisms (Fig. 18) or when searching for orthologues of DNA methylating enzymes in their genome (Fig. 8).

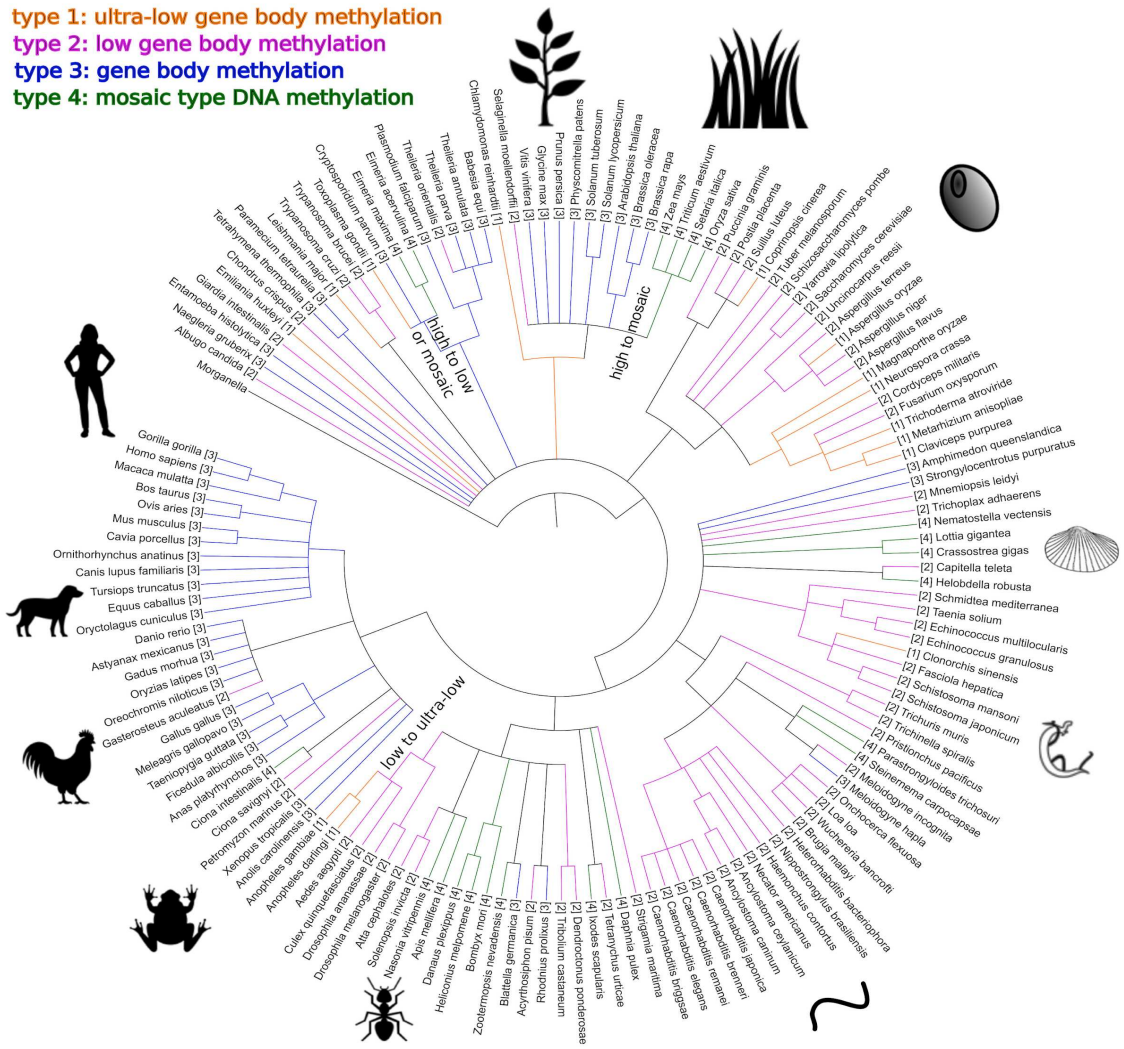


Figure 18. The tree of life constructed from 147 species. The different colour of branches and numbers in square brackets correspond to four types of methylation as identified by Aliaga, 2019 [100].

Several publications have emerged in the last couple of years in order to elucidate the role of this modification for Chelicerates. Majority of studies are focused on chelicerate parasite models like *Tetranychus urticae* or *Varroa destructor* to name just a few (Tab. 3).

Table 3. List of publications dealing with DNA methylation in Chelicerates.

species	title	publication
Tetranychus urticae	Expression level and immunolocalization of *de novo* methyltransferase 3 protein (TuDNMT3) in adult females and males of the two-spotted spider mite, *Tetranychus urticae*.	[261]
Varroa destructor	A new detection method for a newly revealed mechanism of pyrethroid resistance development in *Varroa destructor*.	[262]
Metaseiulus occidentalis	Genome Sequencing of the Phytoseiid Predatory Mite *Metaseiulus occidentalis* Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution	[263]
Tetranychus urticae	Divergent methylation pattern in adult stage between two forms of *Tetranychus urticae* (Acari: Tetranychidae).	[264]
Dermatophagoides farinae	Rewired RNAi-mediated genome surveillance in house dust mites.	[265]
Stegodyphus dumicola	DNA Methylation Patterns in the Social Spider, *Stegodyphus dumicola*.	[266]

Up to now, the genome of a single species was analysed for the methylation pattern using the direct method of bisulphite sequencing. The inference of genome methylation in the remaining species was based on the CpG O/E method applied to genome assemblies [266].

Comparably to other invertebrate lineages, also chelicerate genomes contain variable DNA methylation profiles and its significance for epigenetic regulation oscillates in different species (Fig. 19). In line with other arthropods, contradictory findings were also reported in a distribution of methylation in different genomic elements of Chelicerates.

This could be ascribed partially to previous postulations that reduction of DNA methylation in invertebrates and loss of methylation in some insects and nematodes occurred as the benefits of its adaptive properties were outweighed by its mutational leverage to the genome [113].

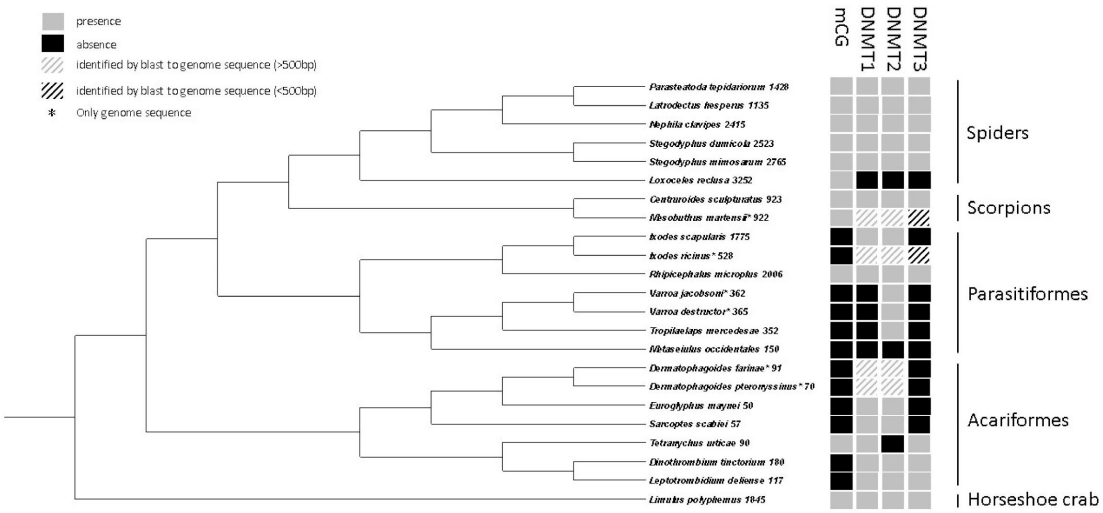


Figure 19. Presence and distribution of DNMT1, DNMT2, and DNMT3 genes in Chelicerate species with known genomes. The presence and level of genomic DNA methylation in the corresponding species were inferred by $CpG_{O/E}$ method [266].

Specific evidence of methylation redundancy in chelicerate genomes was found in a study where 5mC enrichment of transposable elements (TE) was observed in social spider *Stegodyphus dumicola*, which is in contrast to general hypomethylation of TEs observed in a number of other invertebrates [103,113,130].

Another chelicerate species, dust mite *Dermatophagoides farina*, also does not methylate TEs in its genome. Functional characterization of dust mite small RNAs revealed an enrichment of their reads that are targeted to TEs genomic locations [265]. Moreover, dust mite lacks piRNA pathway commonly used for TEs silencing [267]. This indicates that mites may use siRNA-like pathway for regulation of TEs activity rather than methylation. This conclusion is also consistent with evidence of frequent genome methylation in species lacking piRNA pathway [268,269].

2 Aims and objectives

This thesis aims at the identification of stage-specific transcription of *Ixodes ricinus*. It emphasizes the significance and role of methylation and glycosylation for tick development and regulation of gene expression.

The individual aims were:

1. To produce stage-specific transcriptome assemblies of *I. ricinus* and submit the sequencing data to standard bioinformatic pipelines, including annotation and GO enrichment analyses.
2. To identify the transcripts that are specifically expressed in different *I. ricinus* life-stages and evaluate their significance for tick development.
3. To identify transcripts that are involved in methylation and glycosylation pathways in *I. ricinus* based on data from public databases from related species.
4. To identify and characterise the enzymes involved in methylation pathways in the *I. ricinus* transcriptome and infer the significance and extent of methylation for the *I. ricinus* tick.

3 Research papers

Vechtova, P., Sterbova, J., Sterba, J., Vancova, M., Rego, R.O.M., Selinger, M., Strnad, M., Golovchenko, M., Rudenko, N., Grubhoffer, L. 2018. A bite so sweet: the glycobiology interface of tick-host-pathogen interactions. *Parasites and Vectors* 11(1):594. doi: 10.1186/s13071-018-3062-7.

Vechtova et al. Parasites & Vectors (2018) 11:594
https://doi.org/10.1186/s13071-018-3062-7

Parasites & Vectors

REVIEW Open Access

A bite so sweet: the glycobiology interface of tick-host-pathogen interactions

Pavlina Vechtova[1,2*†], Jarmila Sterbova[1,2†], Jan Sterba[1,2], Marie Vancova[1,2], Ryan O. M. Rego[1,2], Martin Selinger[1,2], Martin Strnad[1,2], Maryna Golovchenko[1], Nataliia Rudenko[1] and Libor Grubhoffer[1,2]

Abstract

Vector-borne diseases constitute 17% of all infectious diseases in the world; among the blood-feeding arthropods, ticks transmit the highest number of pathogens. Understanding the interactions between the tick vector, the mammalian host and the pathogens circulating between them is the basis for the successful development of vaccines against ticks or the tick-transmitted pathogens as well as for the development of specific treatments against tick-borne infections. A lot of effort has been put into transcriptomic and proteomic analyses; however, the protein-carbohydrate interactions and the overall glycobiology of ticks and tick-borne pathogens has not been given the importance or priority deserved. Novel (bio)analytical techniques and their availability have immensely increased the possibilities in glycobiology research and thus novel information in the glycobiology of ticks and tick-borne pathogens is being generated at a faster pace each year. This review brings a comprehensive summary of the knowledge on both the glycosylated proteins and the glycan-binding proteins of the ticks as well as the tick-transmitted pathogens, with emphasis on the interactions allowing the infection of both the ticks and the hosts by various bacteria and tick-borne encephalitis virus.

Keywords: Tick, Pathogen, Host, Glycan, Lectin, Glycobiology, *Borrelia*, *Anaplasma*, TBEV, Carbohydrate-binding

Background

Vector-borne diseases constitute 17% of all infectious diseases in the world [1]. Pathogenic viruses, bacteria, and protozoa are carried by blood-feeding arthropods on just about all the continents and both livestock and people tend to be affected by these. This becomes a large economic burden on the animal health sector and on the public health system of various countries. Ticks are the first among blood-feeding vectors in terms of the number of pathogens that they can transmit. Unfortunately, there are next to no vaccines against the tick-transmitted bacterial and protozoan diseases and very few against tick-borne viruses [2]. The only successful anti-tick vaccine, based on the glycoprotein Bm86 from the cattle tick *Rhiphicephalus microplus*, has been shown to be efficient against ticks of the genus *Rhiphicephalus* and Bm86 homologue vaccines have had some efficiency against at least two species of the genus *Hyalomma*, but this is not the case for other ticks and the pathogens they transmit [3]. The European Centre for Disease Prevention and Control suggests that there will be a rise in tick-borne diseases based on changes in various factors including the environment and socio-economics [4]. Research efforts to combat tick-borne diseases have usually centred, as with most other infectious diseases, on determining the Achilles' heel of the pathogen. Most endeavours have focussed on understanding host-pathogen interactions, primarily at the vertebrate level. Protein-carbohydrate interactions between the pathogen and the host cell are of primary importance, in terms of attachment and/or invasion of the cell, whether in an invertebrate or vertebrate host. The observation that there is conservation in the protein-carbohydrate recognition strategies can be used as part of novel approaches for intervention [5]. It has been shown that many regulatory mechanisms are mediated by post-translational modifications (PTM). One example of a PTM that regulates protein degradation and signaling in eukaryotes is ubiquitination. Pathogens are known to exploit ubiquitination to infect mammalian cells and it

* Correspondence: vechtp00@jcu.cz
†Pavlina Vechtova and Jarmila Sterbova contributed equally to this work.
[1]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[2]Faculty of Science, University of South Bohemia, Branišovská 1760, CZ-37005 České Budějovice, Czech Republic

has been shown that the ubiquitination machinery is present in the tick *Ixodes scapularis*. It was identified that the E3 ubiquitin ligase XIAP restricted bacterial colonization of the vector and *xiap* silencing significantly increased tick colonization by the bacterium *Anaplasma phagocytophilum*, the causative agent of human granulocytic anaplasmosis [6].

Over the last decade, there has been a slow increase in the knowledge of vector-host-pathogen interactions which start from the time a pathogen invades the vector within a blood meal and attaches to the tick midgut lumen. Later it traverses to the tick salivary glands and completes its life-cycle by transmission to a new mammalian host during the subsequent tick feeding [2].

Four possible routes that may facilitate pathogen survival and transmission by most arthropod vectors including ticks have been pointed out. These include: (i) pathogen carbohydrate-binding adhesins that attach to receptors in the tick midgut; (ii) the attachment of carbohydrate-binding proteins of the arthropod to the pathogen as part of its innate immunity; (iii) carbohydrate-binding proteins that are soluble and form a link between the pathogen and midgut surfaces; and (iv) the use of co-receptors to enhance the interactions within the vector [5].

In this review, we would like to highlight the glycobiological aspects of all four of these specific mechanisms that come into play when looking at the vector-pathogen interactions as well as glycobiology-associated processes between the mammalian host and the pathogen (see Fig. 1). Glycobiology of ticks and tick-borne pathogens is developing together with the increased availability and sensitivity of analytical methods; a short overview is listed below, together with some relevant references for readers seeking deeper knowledge.

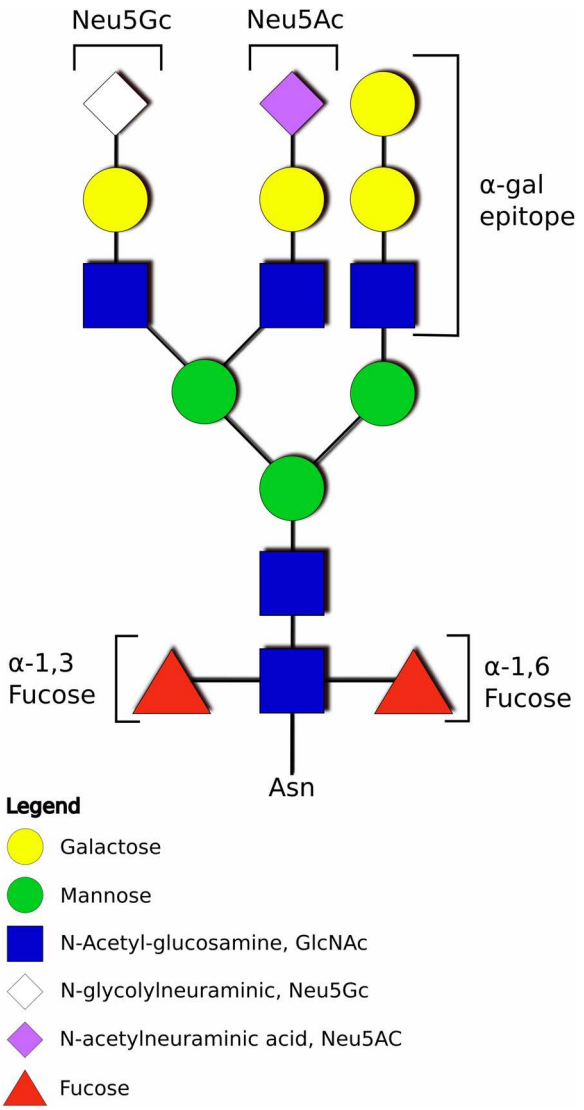


Fig. 1 Pathogen-tick and pathogen-host interactions. The scheme represents carbohydrate mediated interactions depicting in particular the most well-known interacting partners of *Borrelia*, tick vector and host (**a**) and *Anaplasma*, tick vector and host (**b**). Both *Borrelia* and *Anaplasma* produce adhesion molecules recognizing either a specific glycoprotein (such as TROSPA, decorin) or a specific glycan (core α1,3-fucose, sialylated glycans) in the tick and the host. Furthermore, *Borrelia* produce proteins interacting with host glycoproteins regulating its immune system. Two examples of the recognized glycans are shown in (**b**): an *O*-glycan bearing both α1,3-bound fucose and also a sialic acid, and an *N*-glycan with its core modified by the α1,3-bound fucose. The used symbol nomenclature is based on the Consortium for Functional Glycomics (http://www.functionalglycomics.org/)

Importance of glycosylation for protein functions

Post-translational modifications can be found in both prokaryotic and eukaryotic organisms; among them, glycosylation is one of the most abundant and most important. Protein glycosylation affects all the functions of proteins - their structure, activity, interactions with other molecules, half-life in the cell or organism; immune recognition is also dependent on the interaction of immune cells and receptors with glycosylated molecules. A wide variety of possible glycan structures and linkages increases the functionality of proteins [7].

The importance of carbohydrates for the function of proteins can be simply shown on complement proteins. The complement system comprises of more than 30 plasma- or membrane-bound components. Most of them are glycosylated to the various extent and the type of glycosylation generates tissue- or cell-specific population of glycoforms of each complement molecule. The specific glycoform population then secures functions that are required for a particular cell type or tissue. The repertoire of functions secured by glycans ranges from the control of protein folding, proper assembly within the endoplasmic reticulum, mechanistic shielding of the protein backbone

against protease degradation, preventing inappropriate protein-protein interactions or formation of proper spatial protein conformation or participation in recognition epitope formation [8]. Specific examples show interesting ways in which glycans influence or modulate the complement cascade.

C1q is a recognition molecule of classical complement pathway and mediates initiation of the pathway by binding to the antibody-antigen complexes. The proper function of the C1q is conditioned by the appropriate triple helix formation within C1q monomer and the formation of C1q hexamer whose spatial conformation may be secured by the presence of N-linked glycan of each monomer [9].

Glycosylation was also proven important in complement regulatory factors where factor H (fH) glycosylation mediates its proper folding within the endoplasmic reticulum (ER). The absence of glycosylation or its malfunctioning leads to in fH misfolding and retention in ER causing clinical symptoms in children in form of hypocomplementemic renal disease [10].

C1-inhibitor is a plasma glycoprotein and, along with other members of the serpin proteases, its inhibitory activities are enhanced by binding of negatively charged polysaccharides. Most of the polysaccharides binding to the C1-inhibitor induce allosteric changes of the inhibitor molecule causing potentiation of the attachment to the C1 proteases or, as in case of dermatan sulphate, the potentiation is caused by the formation of a negatively charged polysaccharide-mediated linkage between positively charged portions of the C1-inhibitor and C1 protease molecules [11]. Structural characterization of the C1-inhibitor reveals extensive O-glycosylation with a high number of sialylated glycans. The trials for functional characterization of C1-inhibitor glycosylation showed an increased resistance of highly O-glycosylated region against proteolytic degradation [12] and highlighted the importance of sialylation for prolonged serum half-life [13].

Advances in bioanalytical methods for glycobiology

The most frequently used methods in glycobiology are mass spectrometry in combination with chromatography or capillary electrophoresis, glycan/lectin microarrays, or lectin staining. All of these developed greatly recently; for example, the increasing number of mass spectrometers available throughout the world and the development of more sensitive instruments and specifically the introduction of the Orbitrap mass spectrometers, greatly advanced the possibilities for glycan and glycoprotein analysis [14, 15]. The number of commercially available microarrays is also increasing and nowadays allows more or less specific detection of almost any kind of glycan. The availability of lectins together with the possibility to synthesize specific glycan molecules allows also the preparation of in-house glycan- or lectin-arrays; another possibility is the service provided by the Consortium for Functional Glycomics (http://www.functionalglycomics.org/).

Here, we review the current knowledge on how pathogens have evolved "sweet" strategies to overcome the immune responses within the vector and the mammalian host and the use of carbohydrate-binding properties to perpetuate their transmission and dissemination into a vertebrate host (Tables 1 and 2). We also provide a near comprehensive catalogue of all carbohydrate molecules that play a part in the disease cycle that have been characterized to date, be it within the tick or the mammalian host. We would like this to be the start of a renewed interest in the glycobiology of ticks and tick-borne diseases.

Glycosylation in the *Borrelia* infection cycle

Compared to eukaryotes, glycosylation in bacteria produces a much more diverse repertoire of glycoconjugate structures which are often species- or strain-specific. Most of the bacterial glycoconjugates are an integral part of the bacterial cell wall and provide the bacterial cell structural integrity. Additionally, the bacterial glycosylated cell surface structures mediate adhesion and interaction with its environment or host. Although the structural features of bacterial surface glycans have been well described, the function of many of them, including those in pathogenic bacteria, remain unexplored. In principle, pathogenic bacteria use glycosylation for two reasons; they synthesize host-like glycan structures to hide from the host immune system and, conversely, they produce glycosylated proteins that are able to bind more effectively the host immune molecules and thus influence their activity [16].

Since all *Borrelia* species are host-propagated bacteria that shuttle between a vertebrate host and an arthropod vector, these spirochetes have developed strategies to adjust to these diverse environments [17]. This is achieved by regulating the level of gene expression in response to changes in temperature, pH, salts, nutrient content, and other host- and vector-dependent factors. A significant number of *Borrelia* proteins mediate the interactions with host/vector molecules and thus enable *Borrelia* to complete its infectious cycle. Recent findings highlight the importance of carbohydrate moieties in these interactions and in the overall pathogenesis of this infectious spirochete.

Borrelia/tick glycosylated interactions

When entering the vertebrate host during tick-feeding, *Borrelia* must overcome several barriers to successfully invade and disseminate in the host body. The invasion of the host is difficult as it requires the interaction of the existing *Borrelia* surface structures with host tissues without being noticed by the host immune system. *Borrelia* have developed many elaborate strategies to recognize diverse host molecules and cell types to promote

Table 1 Summary of carbohydrate-binding proteins of *Borrelia* spp. recognizing tick or host receptors. The carbohydrate-binding proteins from *Borrelia* spp. are listed together with the recognized molecule from the vector or the host. Glycoproteins or glycans are listed as the recognized molecules depending on the available information. Majority of proteins from Lyme borreliosis spirochetes are listed; in the case of relapsing fever *Borrelia*, the bacterial species is defined

Borrelia spp. protein	Tick binding partner	Reference
Borrelia vs tick		
OspA	TROSPA	[236]
OspC	SALP15	[27]
TSLPI/P8	Mannose binding lectin (MBL)	[42]
Vsp33 (*B. hermsii*)	Unknown receptor in tick SG	[62]
Borrelia vs host		
Bgp (p26)	GAG	[294]
DbpA (p20)	Decorin/dermatansulfate	[48, 295]
DbpB (p19)	Decorin/dermatansulfate/chondroitinsulfate	[48, 295]
Bbk32	Fibronectin /heparansulfate/dermatansulfate	[63]
P66	Integrins	[87]
OspA	Plasminogen	[296]
OspC	Plasminogen	[297, 298]
Enolase	Plasminogen	[299]
Erps (OspE/F related proteins)	Factor H or FHL protein	[105]
CRASPs	Factor H	[105]
PAMPs	Mannose receptor on dendritic cells	[300]
Unknown	Neolacto-(Gal4GlcNAc3Gal4Glc1)-carrying glycoconjugates in human erythrocytes	[301]
VspB (*B.turicatae*)	GAG	[61]

Table 2 Summary of carbohydrate-binding proteins of *Anaplasma* recognizing tick or host receptors. The carbohydrate-binding proteins from *Anaplasma* are listed together with the recognized molecule from the vector or the host. Glycoproteins or glycans are listed as the recognized molecules depending on the available information

Anaplasma protein	Binding partner	Reference
MSP1a (MSP1 complex)	Vector binding partner: Unknown receptor in IDE8 tick cells	[189, 190]
Unknown molecule	Vector binding partner: Core α 1,3-fucose glycoprotein	[203]
Unknown adhesin-like molecule	Host binding partner: 1,3-Fuc and Sia in sialyl Lewis X, PSGL-1 in human neutrophils	[199, 200]
Unknown adhesin-like molecule	Host binding partner: 1,3-Fuc and Sia in sialyl Lewis X, PSGL-1 in murine neutrophils	[199]
Unknown adhesin-like molecule	Host binding partner: 1,3-Fuc and Sia in sialyl Lewis X, PSGL-1 in human myeloid HL-60 cells	[197, 201]
Unknown molecule of *A. phagocytophilum* NCH-1 strain	Host binding partner: 1,3-fucose in murine bone marrow-derived mast cells (BMMCs), murine peritoneal mast cells	[198]
Unknown molecule of *A. phagocytophilum* NCH-1 strain	Host binding partner: α 1,3-Fuc in human skin-derived mast cells	[198]
AmOmpA	Host binding partner: α2,3-sialylated and α1,3-fucosylated glycan of the sialyl Lewis x in myeloid cells	[182]
AmOmpA	Host binding partner: α2,3-sialylated and α1,3-fucosylated glycan of the 6- sulfo-sialyl Lewis x in endothelial cells	[180, 182]
AmOmpA	Host binding partner: α2,3-sialylated, α2,6-sialylated, α1,3-fucosylated glycan receptors in human and murine myeloid HL-60 cells, 6- sulpho-sialyl Lewis x in endothelial cells	[180–182]
Unknown	Host binding partner: α1,3-fucose	[203]

dissemination and chronic infection [18], and to overcome host immune system surveillance [19]. The concerted action of these structurally and functionally diverse *Borrelia* surface molecules helps the spirochete to successfully adapt and multiply in the host body.

The interacting molecules of *Borrelia* and the tick are often modified by glycosylation producing a diverse pool of structures. Moreover, glycosylation is a dynamic modification and can be readily altered upon environmental cues [20].

The presence of glycoconjugates on the surface of cultured *B. burgdorferi* has been demonstrated by the ability of *Borrelia* to bind a number of lectins [21]. In search of *B. burgdorferi* glycosylation patterns, increased attention has been paid to outer surface proteins that are produced at different stages of the *Borrelia* transmission cycle and represent points of interaction between the spirochetes and their hosts/vectors.

Borrelia outer surface proteins

Borrelia outer surface proteins A and B (*osp*A and *osp*B) are encoded on a bicistronic operon and extensively expressed on the surface of spirochetes in unfed ticks. OspA is one of the major and most comprehensively studied *Borrelia* proteins. While OspA mediates *Borrelia* attachment to the tick midgut when spirochetes are acquired by ticks during blood-feeding, OspB plays a key role in successful colonization of the tick midgut. OspA downregulation is important for *Borrelia* detachment, multiplication, and migration from the tick midgut to salivary glands [22–25]. When ticks are fed to engorgement, *Borrelia* clears OspA and OspB from the surface expressing instead another outer surface protein C (OspC) [22, 26].

OspC, encoded by *bbb19* mapped to the cp26 plasmid, is one of the most divergent genes in *Borrelia* genome, and is crucial for the early stages of mammalian host infection by the spirochete, but not required for acquisition of spirochetes by tick, tick colonization or migration from salivary glands to the gut [27–32].

Erps (OspE/F-related proteins) are a family of surface integrins with high affinity to factor H and encoded by *erp*-loci localised on each of the cp32 plasmids. Lyme disease spirochetes control Erp synthesis throughout the bacterial infectious tabacycle, producing the proteins during the infection of the host but downregulating their synthesis during tick infection stage. The best-characterized members are OspE and OspF proteins [33], their paralogues OspE/F-related proteins [34] and a group of OspE/F-like leader peptides (Elps) [35].

OspA, OspB and TROSPA

Earlier work had indicated that OspA and OspB are the major *Borrelia* glycosylated proteins [36], yet a later study showed that the suggested *N*-linked glycosylation does not occur [37]. Colonization of ticks by spirochetes requires the involvement of tick receptor(s). Although a tick receptor for OspB has not yet been identified, the tick receptor for OspA (TROSPA) is located in the tick gut and is heavily glycosylated. The blockade of TROSPA by TROSPA antisera or by downregulation of TROSPA *via* RNAi reduced *B. burgdorferi* adherence to the tick gut, hampering the spirochete transmission to the mammalian host. The number of potential posttranslational modification sites in TROSPA is unusually high (> 30), with a predominance of *O*-glycosylation sites [25].

OspC and Salp15

When transmitted from the tick vector to the host, *Borrelia* are delivered within the tick saliva. Tick saliva contains a plethora of bioactive molecules, which have been shown to be important for immunosuppression of the host responses [28]. One of the secreted salivary proteins is *Salp15* [29]. This protein specifically interacts with *B. burgdorferi* OspC which results in the protection of *Borrelia* from antibody-mediated killing and plays a critical role in establishing *B. burgdorferi* infection [27]. Whereas no data exists about the potential glycosylation of OspC, the glycosylation of Salp15 was demonstrated experimentally [30]. Salp15 from *I. ricinus* did not deliver the same protection to *B. garinii* and *B. afzelii* against antibody-mediated killing [31], presumably suggesting that the Salp15 binding for some species is an advantage for surviving in nature [31]. An explanation may lie in a different structural or spatial organization of the OspC or Salp15 molecule causing better access to the binding sites of each of the molecules in *B. burgdorferi*. Another hypothesis claims that *B. burgdorferi* OspC holds differently charged areas which interact in a way that favour formation of OspC multimers or even a lattice [38]. This structure, along with bound Salp15, might form a protective coating on the bacteria preventing an access of anti-OspC antibodies or *B. burgdorferi* antiserum [27]. In addition to the direct interplay between Salp15 and *B. burgdorferi*, Salp15 indirectly facilitates the host invasion by inhibiting dendritic cell activation by binding to the receptor/lectin DC-SIGN, localized on the surface of macrophages and dendritic cells [32].

Ixofin3D and Ixodes scapularis dystroglycan-like protein

Ixofin3D and *I. scapularis* dystroglycan-like protein (ISDLP) are glycoproteins expressed on the surface of midgut cells which were identified as candidate tick midgut binding partners of *B. burgdorferi* using a yeast surface display assay [39]. The expression of both Ixofin3D and ISDLP was elevated in *Borrelia*-infected tick midgut during feeding. Ixofin3D and ISDLP interact with spirochete cells as was confirmed *in vitro* by immunofluorescence assay and RNA interference. The RNAi-mediated reduction in expression of Ixofin3D and ISDLP resulted in decreased

spirochete burdens in the tick salivary glands and in the murine host as well [39, 40]. The full-length Ixofin3D contains four putative fibronectin type III domains. Ixofin3D is glycosylated as shown experimentally by periodic-acid Schiff's staining of a recombinant protein produced in *Drosophila* cells. Even though the importance of Ixofin3D for *Borrelia* infection was shown, the borrelial binding partner for Ixofin3D has not yet been identified [39].

The binding partner for ISDLP is also yet to be discovered. Like Ixofin3D, ISDLP silencing did not reduce the spirochete numbers in the gut but the transmigration process from gut to salivary glands was impaired. The mechanism remains unknown although the collected evidence implies that ISDLP may facilitate gut tissue remodelling or reduced barrier for spirochete transfer to salivary glands [40].

TSLPI

Tick salivary lectin pathway inhibitor (TSLPI) is a secreted salivary protein that protects *Borrelia* from complement-mediated killing. TSLPI facilitates spirochete transmission and acquisition through interference with the host mannose-binding lectin (MBL) and inhibition of the host lectin complement pathway [41]. *N*-linked glycosylation of recombinant *Drosophila*-expressed TSLPI appeared to be vital for its function as a lectin pathway inhibitor, suggesting that TSLPI *N*-glycans are involved in its binding to MBL carbohydrate recognition domains [42, 43].

Borrelia adhesins and extracellular matrix

Adhesion is the first and basic event in establishing an infection. The *Borrelia* cell surface is, at the time of host invasion, covered by adhesion proteins that can recognize and bind to various host cell types and/or extracellular matrix (ECM) components and thus promote *Borrelia* dissemination and settlement in various corners of the host body. Although *Borrelia* adhesins are not glycosylated, the presence of glycosylation has been confirmed in their tick receptors, suggesting a significant role of glycosylation in adhesin-receptor interaction.

Borrelia burgdorferi encodes a variety of adhesins and their characterization and role in the *Borrelia* infection cycle using different approaches was thoroughly described in a review by Coburn et al. [44]. With regard to their overlapping roles in *Borrelia* adhesion to the host tissue, it is important to note that only the concerted action of various adhesins guarantees an effective adhesion and transmigration of spirochetes to different hosts and their tissues [44].

A short overview of host ECM proteins

Glycosaminoglycans

Several ECM-associated molecules are specifically targeted by *Borrelia* adhesins. Exceptionally important seem to be the glycosaminoglycans (GAGs), large linear polysaccharides constructed of repeating disaccharide units (e.g. hyaluronan, chondroitin, dermatan, heparan, keratan) that decorate the ECM proteins. GAG chains are abundantly modified by sulphurylation, which imparts them a strong negative charge [45]. Numerous studies have shown that binding of GAGs by *B. burgdorferi* enables colonization of the host [46].

Fibronectin

A relevant ECM-associated molecule for *B. burgdorferi* attachment is fibronectin (Fn), a high-molecular weight, dimeric glycoprotein found in body fluids and in the ECM. *Borrelia burgdorferi* fails to bind to the ECM *in vitro* upon exposure to an anti-Fn antibody, implicating the Fn involvement in *Borrelia* attachment [47].

Decorins

Decorins are ubiquitous ECM proteoglycans, which are associated with collagen fibrils in the mammalian connective tissues [48]. Decorins are complex glycoproteins; apart from serine linked GAG chain, they are also modified by up to 3 N-glycans [49, 50]. Numerous studies associate decorins with *Borrelia* adhesin attachment and interestingly, the binding is promoted by intact decorin proteoglycan molecule rather than its protein core or GAG chain itself [48].

Laminin

Laminin is a large extracellular matrix multidomain glycoprotein. It is a critical molecule in the basement membrane assembly and, by extension, in tissue formation in the developing organism [51]. Besides its role in basement membrane architecture, it also mediates cellular interactions and provides a dense network for various cellular signalling and attachment events. The existence of different laminin isoforms gives space to developmental regulations mediated by differential responses to cells and newly forming tissues. Laminin, as well as other ECM forming molecules, possesses numerous glycosylation sites and its molecule is modified up to 32 % by N-linked glycans [52].

The carbohydrate portion of Laminin was proven to be a mediator of attachment in several bacterial species [53]. Laminin is also a potent target of several borrelial adhesins [54–56] although the direct involvement of the laminin carbohydrate moiety has not been reported yet.

Integrins

Integrins are glycosylated cell surface receptors mediating cell adhesions to the extracellular matrix and some important cell-cell interactions [57]. The presence of *N*-linked glycans in integrin molecule proves integral for the stability of the domain conformation and consequently affects integrin adhesive properties [58]. Integrins possess

a typical heterodimeric structure combining different α and β polypeptide chains and their combination determines the specificity of integrins [59]. Integrins are expressed on all mammalian cells except erythrocytes. The expression of different integrin subtypes produces unique cell surface signature of each cell type [60].

Borrelia adhesins

Vsps

Relapsing fever *Borrelia*, unlike Lyme disease-causing *Borrelia*, are vectored by soft ticks of the genus *Ornithodoros*. They are present in the blood of the mammalian host in high numbers which lead to high fevers followed by bouts of relapses. They recognize glycosaminoglycans (GAG), which mediates the attachment of *Borrelia* to mammalian cells. GAG recognition is partly dependent on the presence of some of the variable small proteins (Vsps). *Borrelia turicatae*, a relapsing fever borrelia that is vectored by *O. turicata*, recognizes GAGs *via* VspB which allows binding of *B. turicatae* to cultured mammalian cells as well as increased spread and replication in the mammalian host. *Borrelia hermsii* also attaches to cultured mammalian cells *via* GAGs; however, Vsps are not essential for this binding [61].

After the feeding of *O. hermsi* with *B. hermsii*-infected blood, the bacteria switched from expression of many bloodstream outer surface variable major proteins (Vmps) to a unique protein, variable tick protein (Vtp, Vsp33) [62].

BBK32, RevA and C1-inhibitor

Borrelia burgdorferi expresses at least two fibrinogen-binding proteins, BBK32 [63] and RevA [64]. BBK32 is a protein, whose attenuation does not block the spirochete transmission from the tick to the host [65], but lowers the bacterial loads in different tissues at different time points of infection [66]. *Borrelia burgdorferi* also attaches to endothelium in the vascular system through fibrinogen (Fn) and this interaction becomes stronger with increasing blood flow, allowing the spirochete to overcome fluid shear stress [67]. These stabilizing interactions are sustained by catch bond properties of BBK32 [68]. Following the binding to Fn, BBK32 binds to various kinds of GAGs, including heparin sulphates and dermatan sulphates of the host ECM [69–71]. It also seems to be involved in the modulation of the innate immunity. In particular, BBK32 binds the C1 complex of the classical innate immunity pathway, preventing its activation and thus obstructing classical pathway-mediated *Borrelia* lysis [72].

As *B. burgdorferi* BBK32 mutants are still able to bind fibronectin, an additional Fn-binding protein, RevA, was identified [58]. RevA expression on the *Borrelia* cell surface was upregulated in mammalian host compared to the tick vector. Furthermore, *Borrelia*-infected patients produced anti-RevA antibodies throughout various stages of Lyme disease suggesting its involvement in Lyme disease establishment and persistence in the host. RevA appears to have multiple binding sites which *Borrelia* uses to bind host cells *via* Fn [46, 73].

DbpA/DbpB

Decorin-binding proteins A and B (DbpA and DbpB) are adhesins found on the surface of *B. burgdorferi* [20, 45]. These proteins are critical for the virulence of *B. burgdorferi* [74, 75]. New data suggests that the decorin-binding proteins actually do not bind directly to the decorin protein core but interact with decorin *via* GAGs that are attached to the protein [76–78]. The binding studies of DbpA and DbpB from different *Borrelia* genospecies showed that there are clear differences in the decorin binding activity and that these differences may ultimately lead to the differences in tissue tropism and clinical manifestations associated with particular *Borrelia* genospecies [76, 79]. *In vivo* functional studies demonstrated the importance of DbpA/B adhesins for *Borrelia* invasion of the mammalian host especially in the early stages of infection [80].

Bgp

Borrelia burgdorferi glycosaminoglycan binding protein (Bgp) is a surface-exposed protein on intact spirochetes [70]. Recombinant Bgp bound the same GAG as the whole spirochete, agglutinated erythrocytes and inhibited binding of *B. burgdorferi* to the mammalian cells. A transposon mutant of the *Bgp* gene had less ability to adhere to host endothelial and epithelial cells *in vitro* and to colonize host target tissues leading to the reduced inflammatory manifestation of Lyme disease in the mouse model. The adherence was not fully disrupted due to the existence of other GAG-binding adhesins which facilitate host colonization and also highlights the importance of *Borrelia* GAG-binding ability for the completion of the infection cycle [81]. Although the Bgp attachment to GAG is not essential for disease establishment, the protein appears to be involved in the formation of an initial infectious niche in the host. Different spirochetes strains possess different GAG-binding preferences and their binding ability to multiple cells depends on the GAGs that they express [76].

BmpA

BmpA (*Borrelia* membrane protein A) and its three paralogues B, C, and D are all laminin-binding borrelial outer surface proteins [82]. Like other *Borrelia* surface proteins, BmpA is also antigenic. All bmp genes are located on the *Borrelia* chromosome, arranged in clusters that are differentially regulated [83]. The involvement in the development of arthritis in the mouse model was described for two Bmp proteins, BmpA and BmpB [84].

Borrelia adhesins and integrin-mediated interactions

Borrelia binds to host endothelial cells *via* the interaction of integrins $\alpha_{IIb}\beta_3$, $\alpha_V\beta_3$, and $\alpha_V\beta_1$ with *Borrelia* surface adhesins [59, 85]. It was also described that the causative agent of relapsing fever, *B. hermsii*, binds to human platelets promoted by the platelet glycoprotein integrin $\alpha_{IIb}\beta_3$ and is diminished by $\alpha_{IIb}\beta_3$ antagonists or by a genetic defect in this integrin [86].

P66

P66 is one of the candidate ligands for β_3-chain integrins (e.g. $\alpha_{IIb}\beta_3$, $\alpha_V\beta_3$) [87]. P66 also functions as a porin [88, 89], and structural predictions, as well as some experimental data, present the molecule as porin assuming the structure of β-barrel [90].

P66 mutants showed a dramatically reduced ability to attach to integrin $\alpha_V\beta_3$ [91]. Endothelial cells responded to wild-type *Borrelia* infection by upregulation of endothelial growth factor compared to a control infection with a P66 deletion mutant. The ability of P66 mutants to transmigrate through the cell monolayer was impaired, which suggests the role of P66 in *Borrelia* transendothelial migration, although its porin function does not play a role in the migration process [92].

Mammalian integrins typically contain an RGD (Arg--Gly-Asp tripeptide) consensus sequence in their binding domain, where aspartic acid is a key binding amino acid. P66 lacks this sequence; however, residues 205 and 207 of its 203–209 binding region are both aspartic acid [93]. P66 deletion mutants applied subcutaneously are readily cleared out of the site of infection, which refers to the possible involvement of the innate immune system and confirms the importance of this protein for host colonization together with other studies [94]. However, tick colonization is shown to be P66-independent [94].

BB0172

BB0172 is an outer membrane protein containing von Willebrand factor A domain which mediates intercellular and protein-protein interactions in ECM. It is, for example, involved in the attachment of platelets to the ECM in the site of damaged endothelial epithelium *via* platelet surface glycoprotein [95]. BB0172 showed a weak interaction with ECM-associated fibronectin. Importantly, a strong affinity was observed in the attachment of BB0172 to $\alpha_{III}\beta_1$ integrin. Moreover, the affinity was much stronger than the one observed in the interaction of borrelial P66 adhesins with β_3 chain integrins [95].

Borrelia adhesins interacting with mammalian complement

Mammalian innate immunity is alerted by a variety of surface-exposed molecules of invading pathogens. The first encounter of host antibodies with potentially harmful intruder activates the complement system which assists in tagging of the pathogen for destruction and also acts on pathogen clearance itself by the formation of membrane attack complex. Different pathways of the complement system progress in a cascade-like manner and its brisk response to pathogen invasion must be under the control of regulating mechanisms preventing complement from attacking host cells.

Invading a host organism, the pathogens have evolved different strategies to circumvent the immune response. Many of these strategies are in fact directed against components of the complement system. The most widespread strategy employs molecules recruiting or mimicking the complement regulators, including the direct interaction of pathogens with complement proteins leading to the modulation or inhibition of their function or indirectly to the activation of complement proteins enzymatic degradation [96].

The complement regulators are represented by several serum proteins that are able to dampen the activity of complement and prevent host self-destruction. Two of them, complement factor H (FH) and its splice homologue Factor H-like (FHL) inhibit the alternative complement pathway response using host-|specific surface patterns like sialic acid or GAGs and thus promoting self-recognition processes [97, 98].

FH is a plasma glycoprotein containing 9 glycosylation sites [99] bearing complex, predominantly diantennary disialylated, fucosylated, and nonfucosylated glycans at eight of the nine glycosylation sites [100]. Similarly, FHL is also a plasma glycoprotein [101]. Both proteins possess a RGD motif which is assigned cell adhesive properties and thus can modulate cell adhesion. Additionally, FHL promotes anchorage-dependent cell attachment and spreading [102].

CRASPs and ERPs

The two complement regulators, FH and FHL are bound by *Borrelia* surface proteins hence preventing the activation at the central step of the complement cascade. Serum resistant *Borrelia* express adhesins on their surface, which are capable of interfering with different components of the host complement system leading to the modulation of host immune response and hampering the complement-mediated spirochete lysis [103, 104].

The two well-characterized types of complement interfering adhesins, complement regulator-acquiring surface proteins (CRASP) 1 and 2 [105], control the complement activity by binding complement regulating molecules such as FH and FHL-1 [104, 106]. Up to now, five different CRASPs (CRASP-1 to CRASP-5) have been described and each of them presents a different binding ability to FH, FHL-1, or plasminogen [98, 104, 106].

CRASP-1 (CspA, BBA68) has been studied the most extensively. It shows a strong affinity to the complement regulators which inactivate the complement response very efficiently [106, 107].

The expression of CRASP-1 is repressed in the tick vector and increases in the mammalian host, which suggests its role in spirochete transmission and evasion of the host immune response [108, 109]. CRASP-1 also confers serum resistance to *B. burgdorferi*. The role of CRASP-1 in complement inactivation is evident in the CRASP-1 knockout-mutants which inefficiently bound human FHL and attracted complement constituents more readily [110, 111].

Apart from *B. bavariensis*, all studied *Borrelia* species possess CRASP-1 orthologues conferring complement inactivation [112, 113]. The orthologues belong to the same protein family although the encoding genes do not share the same locus with the *B. burgdorferi* CspA [98].

CRASP-2 (CspZ) is another *Borrelia* adhesin binding both FH and FHL-1 independently of CRASP-1 and reinforcing *Borrelia* complement resistance [114, 115]. The CRASP-2 expression fluctuates in a somewhat similar manner to CRASP-1 during the *Borrelia* infectious cycle. Like CRASP-1, CRASP-2 is also upregulated during an established mammalian infection and is able to activate antibody-mediated immune response [116], which makes this adhesin important for the diagnosis of Lyme disease infection. The triggered immune response does not, however, provide the host with protective immunity and has no effect on spirochete dissemination [117].

Three members of the polymorphic Erp (OspE/F-related protein) protein family, ErpA (BBP38, CRASP-5), ErpC (CRASP-4) and ErpP (BBN38, CRASP-3), are plasminogen binding proteins that can simultaneously bind to FH and FH-related proteins [103, 107, 118–122].

The Erp proteins are most probably involved in different reservoir hosts infection due to differential binding abilities of particular Erp paralogues [123, 124]. Despite their complement regulator binding properties, none of the Erp proteins are necessary for the protection of *Borrelia* spirochetes against complement-mediated killing; CRASP-1/CRASP-2 deletion mutants expressing all Erp proteins were susceptible to serum mediated lysis [119, 120, 125].

Erps are not upregulated during *Borrelia* transmission but their expression gradually increases during Lyme disease progression, suggesting their role during mammalian infection [125]. Interestingly, *Borrelia* can regulate the expression of both Erps and CRASPs very dynamically as different isolates of *B. burgdorferi* (*s.l.*) reacted differently to complement-mediated killing [56, 126]. Moreover, some of the Erp members present multiple functions during *Borrelia* infection. For example, ErpX ability to bind complement regulators is complemented by its laminin binding properties [56]. The overlapping activities of *Borrelia* surface molecules enhance the overall infectious potential of the spirochete.

Borrelia-specific host pattern-recognition receptors and lectins

Toll-like receptors

Recognition of pathogens is mediated by a set of pattern-recognition receptors (PRRs). The group of glycosylated proteins that comprise the Toll or Toll-like receptors family (TLRs) are transmembrane receptors that function as PRRs in mammals [127]. So far, eleven members that potentially participate in the recognition of invading pathogens have been identified in mammalian genomes [128] and glycosylation was shown to have a critical role in TLR presentation on the cell surface [129, 130].

There are several TLR members, whose role in spirochete recognition has been identified. The well-characterized TLR2 is presented on antigen-presenting cells, epithelial and endothelial cells [131]. It was able to recognize a variety of ligands and was important for macrophage activation and further triggering of the immune response in *Borrelia*-infected mammalian hosts when stimulated by OspA [132]. The signal transduction through TLR1/2 in response to *B. burgdorferi* invasion can elicit opposite immunoregulatory effects in the blood and CNS immune cells, affecting the different susceptibility of these compartments to infection [127].

TLR4 is expressed in macrophages and dendritic cells [130] and is upregulated upon *Borrelia* infection or stimulation by OspC [133, 134] and its main ligands are lipopolysaccharides (LPS) from gram-negative bacteria [135]. The role of TLR4 in *Borrelia* recognition remains unclear as *B. burgdorferi* does not express LPS on its surface.

TLR9 is responsible for recognition and further endosomal/lysosomal internalization of CpG motifs in bacterial DNA [136]. This process has been observed in sonicated *Borrelia*, which promoted the activation of murine cells *via* TLR9 [137].

Nucleotide-oligomerization domain-like receptors

Nucleotide-oligomerization domain-like receptors (NOD-like receptors or NLR) are a group of intracellular PRRs, capable of binding bacterial muropeptides, the molecules derived from bacterial peptidoglycans [138]. Together with TLRs, NOD-like receptors are crucial for recognition of *Borrelia* species. Contrary to other PRR families, NLRs bind bacterial ligands intracellularly, i.e. they are able to recognize the pathogen-associated molecular patterns (PAMPs) that enter the cell *via* phagocytosis or through the membrane pores induced during cellular stress [139].

There are several NLR protein members that can bind carbohydrate-associated PAMPs, although only a few of them were directly observed to be involved in Lyme

disease. NOD1 and NOD2 receptors are the most extensively investigated major PRRs [138, 140].

Borrelia-infected primary murine astrocytes upregulated NOD-proteins upon exposure to some TLR-ligands [138], while murine primary microglia infected by *Borrelia* only upregulated NOD2 and not the NOD1 [141]. NOD2 activation by *Borrelia* stimulated inflammatory cytokines release. Their activities are assigned to a host proinflammatory response, although their particular role in Lyme disease establishment remains unknown [142]. NOD2 stimulation by *Borrelia* induces inflammation during the early stages of Lyme disease but induces tolerance and suppresses *B. burgdorferi*-mediated Lyme arthritis and carditis in mice during later phases of infection [143]. *Borrelia* recognition in the host is conferred by the combined action of TLR and NOD2. The activation of both receptors at a time by *Borrelia* species is essential for an effective cytokine release. It has been concluded that TLR2 and NOD2 co-recognition of *Borrelia* surface receptors leads to both induction of a proper immune response and to inflammatory-induced pathology [144].

C-type lectin receptors

A family of calcium-dependent receptors that bind carbohydrate ligands include both soluble and cell-associated (transmembrane) lectins in vertebrates. C-type lectin receptors (CLRs) expressed by dendritic cells are crucial for tailoring immune response to pathogens. The transmembrane type is predominantly expressed by antigen-presenting cells functioning as PRRs recognizing PAMPs in bacteria [128]. Currently, 17 CLR subfamilies are described in vertebrates.

Mannose receptor represents a subgroup of CLRs binding mannose-containing bacterial transmembrane PAMPs. CLRs are involved in the recognition and phagocytosis of several microorganisms including *B. burgdorferi*. In particular, CLRs were upregulated in dendritic cells after *B. burgdorferi* activation and facilitated phagocytosis of *B. burgdorferi* by monocytes and macrophages [128]. However, the recognized borrelial protein is yet to be identified.

Surface glycolipids of *Borrelia burgdorferi*

Borrelia have an unusual composition of glycolipids in their outer membrane; they synthesize mono-α-galactosyl-diacylglycerol (MGalD) and cholesterol derived glycolipids cholesteryl-β-D-galacto-pyranoside, cholesteryl 6-O-acyl-β-D-galactopyranoside (ACG), or cholesteryl 6-O-palmitoyl-β-D-galactopyranoside (ACGal/BbGL-1) [145–147].

The *Borrelia* glycolipids induce inflammatory reactions; in particular, two glycolipids ACGal/BbGL-I and MGalD/BbGL-II, are probably immunogenic [145, 148]. The immunogenic epitope is recognized in the lipid part of the glycolipids [149]. An important constituent of the immunogenic epitope is the α-linked terminally bound galactose which is recognized by the T-cell receptor of invariant natural killer T cells (NKT) [150]. This then promotes their activation as well as the proliferation of Lyme disease-directed antibodies [151–153] which recognize glycolipids in the cell membrane of *Borrelia* but also *Ehrlichia* [154]. Importantly, the induced antibodies against the glycolipid fraction cross-react with gangliosides, which explains the phenomenon of neuroborreliosis [155].

The glycolipid recognition by invariant NKT cells seems to be an alternative system for innate immune system activation by bacteria lacking LPS, an otherwise typical antigenic determinant of most gram-negative bacteria [156].

Borrelia bind to GalCer (galactosylceramide) on Schwan cells [157], LacCer (lactosylceramide), ceramide trihexoside and gangliosides GD1a and GT1b. Moreover, *Borrelia* displays a specific affinity to disialoganglioside GD1a and trisialoganglioside GT1b carrying sialic acid. The ability to bind such a wide range of glycosphingolipids might provide an explanation for its ability to adhere to a wide spectrum of different cell types [158]. *Borrelia* did not bind gangliosides GM1, GD1b, GM2, GM3 and asialo-GM1 implying the requirement for terminally bound sialic acid in ganglioside recognized epitope and demonstrates the specific character of *Borrelia* and acidic gangliosides interaction. Interestingly, adhesion to GD1a and GT1b, as well as GalCer or LacCer was not compromised by free sialic acid, galactose or lactose, respectively [158, 159]. Conversely, GalCer-binding sites were saturable using free GalCer in CHO-K1 cells preventing spirochetes from attachment [148].

Vector-host glycosylated interactions

Similarly to *Borrelia*, the tick's successful evasion of the host response depends on its ability to conceal its activities from the host immune system. The pursuit of successful feeding drove ticks to equip their saliva with multiple pharmacologically active molecules which feature immunomodulatory activities. The myriad of diverse functions include cytolysis, vasodilatation, anticoagulation, anti-inflammation and immunosuppression. The comprehensive list of tick pharmacologically active salivary gland molecules is presented in a recent review [28].

P672 and CCL8

P672 is a chemokine binding protein (evasin). Evasins bind to multiple chemokines of different origin and their effects are thus pleiotropic. To date, several evasins originating in tick saliva have been identified [160, 161] and they inhibit responses of many chemokine sensitive molecules including neutrophiles or macrophages, which have been demonstrated in several tick species [28]. P672 was originally identified in *Rhipicephalus pulchellus* and its promiscuous binding abilities assign it 13 different chemokine partners showing different dissociation constants.

Mass spectrometric characterisation revealed the presence of several *N*-linked glycans and their deprival negatively influences the affinity of P672 to CCL8, although the underlying mechanism of this observation is yet to be uncovered [162].

Protease inhibitors

Many of the tick salivary proteins are glycosylated [163]. While the exact structure of the glycans attached to these proteins has not been studied, research has concentrated on the role of glycosylation with regard to the recognition of glycans by host immune systems. The importance of the glycan part for antibody recognition was shown for several proteins, such as AamS6 serpin [164], *R. microplus* serpins [165] or evasins 1 and 3 [166] confirming the need to use of glycosylated recombinant proteins in anti-tick vaccine preparations.

For proteins, where the role of glycosylation for the protein function was not confirmed, masking of the tick proteins antigenic epitopes and thus minimization of the immune response was speculated as the role of glycosylation [166].

Serpin 19

Serpin 19 is a serine protease inhibitor identified in the saliva of *Amblyomma americanum*. Serpin 19 displays a broad range of inhibitory activities: it interferes with the host homeostasis, coagulation and the development of inflammatory response. Importantly, the activity of many serine proteases is both positively and negatively regulated when bound to GAGs [167–169] and serpin 19 also contains several predicted GAG binding motives [170]. The functional validation further confirmed its GAG-binding properties and also extended the list of binding partners with heparin sulphate and heparin [170].

Variegin

The inhibition of blood coagulation cascade represents an important property of tick saliva that facilitates successful engorgement on the host. Variegin is a small thrombin-binding oligopeptide isolated from *A. variegatum* salivary glands. During tick feeding, variegin binds thrombin and disables its fibrinolytic activity and thus blocks the blood coagulation cascade. Despite its small size, variegin possesses a single *O*-linked glycan [171]. The synthetic *O*-glycosylated variegin analogues show significantly higher affinity to thrombin and consequently lower reaction kinetics of thrombin-mediated fibrinogenolysis compared to the non-glycosylated form, confirming the importance of its glycosylation. The functional analysis of the inhibition mechanism using macromolecular docking revealed the formation of some favourable hydrogen bonds between hydroxyl groups of the glycan and the allosterically important sites of thrombin [172].

Glycosylation in the *Anaplasma* infection cycle

Anaplasma is a genus of gram-negative rickettsial bacteria. They are obligate intracellular parasites infecting mammals including many domestic animals. The infection causes a reduction of the animal's body weight, abortions, reduces milk production and frequently leads to death [173–175]. In humans, *A. phagocytophilum* is the only confirmed pathogenic species causing human granulocytic anaplasmosis. Patients suffer from fever, headache, myalgias, chills, leukopenia, thrombocytopenia and liver damage manifested by elevated liver enzymes in serum [176]. The symptoms are usually mild but for some individuals, e.g. patients with a weakened immune system, it can be fatal. The infected vertebrate host serves as a reservoir where the bacterium can proliferate for many years and infect naïve ticks [177].

The main vectors of the genus *Anaplasma* are ticks, especially species of the genera *Ixodes*, *Dermacentor*, *Rhipicephalus* and *Amblyomma* [178, 179]. The initial phase of the infection during colonization of the host is the recognition of a suitable cell, attachment onto this cell, and entry into it. This process is facilitated by several specialized bacterial proteins (adhesins/invasins) that can recognize host surface molecules including glycans and glycoproteins and initiate signalling cascades to promote pathogen internalization. *Anaplasma* spp. express several surface proteins which are involved in binding to glycosylated host cells receptors and thus in the infection of the host and tick cells. These differ in glycan specificity and importance for the infection of various hosts and host cell types.

Anaplasma glycoprotein-binding surface proteins

As an intracellular pathogen, *Anaplasma* depends on a host cell to survive. *Anaplasma* infects two different types (groups) of organisms: the tick vector and the mammalian hosts, with various cell types being infected by the pathogen. Recognition of the cell type and of the infected organism is provided through binding of surface glycan epitopes or even several epitopes on a glycoprotein molecules.

Two groups of *Anaplasma* surface proteins were shown to recognize tick or host glycoproteins: outer membrane proteins (Omps) and major surface proteins (MSPs).

OmpA belongs to highly conserved genes among *A. phagocytophilum* isolates and is transcriptionally induced during feeding of *A. phagocytophilum*-infected ticks on mice and also during the invasion of mammalian but not tick cells [180, 181]. Pre-treatment of *A. phagocytophilum* or *A. marginale* bacteria with the respective OmpA antiserum reduces their ability to infect mammalian cells [181, 182]. Also, preincubation of mammalian cells with a recombinant ApOmpA effectively inhibits *A. phagocytophilum* infection of host cells.

Glycoproteins containing α1,3-fucose and either sLex or 6-sulfo sLex on host cells are recognized by the outer

membrane protein A (ApOmpA) of *A. phagocytophilum* [180, 181]. On the other hand, OmpA of *A. marginale* (AmOmpA), a species non-pathogenic for humans, binds only α1,3-fucose and sLex but not 6-sulfo-sLex glycans. *Anaplasma marginale* also produces AmOmpA in both the infected mammalian and tick cells. Pre-treatment of host cells with sialidase or trypsin reduces or nearly eliminates OmpA adhesion. Therefore, AmOmpA interacts with sialylated glycoproteins *via* an adhesin-receptor pair. Thus, both AmOmpA and ApOmpA recognize different receptor molecules even though these receptors share some structural similarity and thus provide a similar function to these two bacterial species [182].

Structures of *A. marginale* and *A. phagocytophilum* OmpA proteins are very similar and their binding domains are structurally conserved. The OmpA binding domain was identified within amino acids 59 to 74 and it is responsible for the recognition of α2,3-bound sialic acid and α1,3-fucose [180]. A recent study by Hebert et al. [182] describes the OmpA receptor-binding domain between the amino acids 19 to 74.

Another group of surface proteins interacting with host (glycosylated) molecules are the major surface proteins (MSPs) that are involved in the adhesion of host cells and the immunological reaction of the host [183–187]. MSP1 protein with its variants α, β1, and β2 and the MSP3 protein are present in *A. marginale*, while MSP2 and MSP4 in both *A. marginale* and *A. phagocytophilum* [188]. The MSP1 complex consists of two polypeptides MSP1a and MSP1b and both polypeptides participate in adhesion processes to both tick cells and bovine erythrocytes [183–186].

Similarly to OmpA, these proteins show glycan-binding activity. *Anaplasma marginale* MSP1 and MSP2 can hemagglutinate bovine erythrocytes [184] suggesting recognition of some erythrocyte surface saccharide molecules. Recombinant forms of the MSP1 isoforms are glycosylated; MSP1a recombinant glycoprotein contains glucose, galactose, mannose and xylose, while MSP1b contains glucose, galactose and mannose. The functional domain of MSP1a contains tandemly repeated peptides that are important for adhesion to tick cells and bovine erythrocytes. The MSP1a polypeptide backbone alone shows binding to tick cell extract proteins and the glycan in its N-terminus enhances this binding [189, 190]. The MSP2 protein binds to the mammalian PSGL-1 [191] and thus can be responsible for the above mentioned *Anaplasma* recognition of sialic acid on this protein. A hypervariable region is present in the middle of the MSP2 gene which allows the bacterium to express various paralogs of the protein on its surface, possibly enhancing immune system evasion [192, 193]. However, the glycan binding abilities of the various MSP2 paralogs were not studied.

In addition to the above-described receptor molecules, two other proteins, Asp14 and AipA, were found to be acting together with OmpA during the infection of host cells. However, neither of these two proteins were shown to bind glycans nor to be glycosylated [180]. Finally, during the past ten years, other novel *A. phagocytophilum* surface proteins Asp55, Asp62 and APH_1235, with possible function as adhesins and invasins have been identified [194–196]; however, their receptor molecules remain unknown.

Anaplasma-host interactions

A confirmation of *Anaplasma* recognition of host-surface glycans came by Goodman et al. [197] showing binding of *A. phagocytophilum* to the cell surface of the promyelocytic leukaemia cell line HL-60. Bacterial binding to the cell surface correlates with the expression of the sialyl Lewis x (sLex) or a closely related 6-sulpho sLex glycan-containing molecules glycan and α1,3-fucosylated molecules. On the other hand, α1,3-fucosylated glycans but not sialylated glycans, are essential for the infection of murine and human mast cells by *A. phagocytophilum* [180, 198]. These glycan epitopes are important for *Anaplasma in vivo* as has been shown by Carlyon et al. [199].

The protein part bearing the recognized glycans can be also important; thus, not any glycan molecule is recognized, only the one found on a specific protein. In humans and animal hosts, *A. phagocytophilum* exhibits, amongst others, a tropism for myeloid cells. As an adhesion molecule involved in the binding to the surface of human neutrophils, the P-selectin glycoprotein ligand-1, PSGL-1, has been identified [197, 199–202]. In the case of human PSGL-1, *A. phagocytophilum* cooperatively binds to a short amino acid sequence in its *N*-terminal region and an *O*-glycan containing a sialyl Lewis x (sLex) on PSGL-1 (NeuAcα2,3Galβ1,4[Fucα1,3]GlcNac) [202] or on another molecule. On the other hand, PSGL-1 is not the major ligand in mice [199, 200]. Thus, the terminal or core α1,3-fucosylated glycans seem to be a generally recognized receptor, while sialylated glycans and PSGL-1 enhance the infection of diverse types of mammalian host cells.

Anaplasma-vector interactions

In the pathogen-tick relationship, several tick glycosylated molecules can be induced in the presence of a pathogen in the tick tissues and help the pathogen to colonize the tick or enhance its infection. For example, α1,3-core-fucosylated glycans are required for tick colonization by *A. phagocytophilum* and silencing of the responsible fucosyltransferases results in the absence of *Anaplasma* in the infected ticks. To increase the number of its receptors in the tick, *A. phagocytophilum* induces the expression of α1,3-fucosyltransferases to enhance the colonization of *I. scapularis* ticks. Therefore, α 1,3-fucose is a unifying

determinant that *A. phagocytophilum* targets to infect its natural murine and arthropod reservoirs and accidental human hosts as well. In addition, the presence or absence of these glycans does not affect the transmission of the pathogen from the tick vector to the vertebrate host. While the infection of the tick by *Anaplasma* depends on the presence of α1,3-core-fucosylated glycans, these epitopes do not seem to be important for the infection by another tick-borne pathogen, *B. burgdorferi* [203].

Furthermore, tandem repeat peptides of the MSP1a functional domain are important for the adhesion of bacteria to tick cells and the glycosylation of MSP1a probably plays a role during the adhesion of *A. marginale* to tick cells [189, 190].

Colonization of the tick by pathogens depends on the tick life-cycle; one of the crucial steps is the colonization of the midgut or survival in the midgut in the process of the blood meal digestion. For successful colonization, the tick midgut peritrophic matrix (PM) and bacterial biofilms formed in the midgut are critical. The PM forms a barrier between the midgut lumen and the epithelial cells lining the luminal side of the midgut and is formed by a thick matrix of mostly chitin with various proteins, such as chitin deacetylase, and glycoproteins [204]. One of the bacteria depending on the biofilm formation in the *I. scapularis* tick midgut is *A. phagocytophilum*. The presence of this bacterium affects the midgut microbial community and biofilm composition and it also decreases the expression of several genes for the glycoprotein peritrophin, one of the major PM components. This results also in decreased PM thickness. Furthermore, RNAi silencing of these genes significantly enhanced *Anaplasma* colonization of the tick [205]. *Anaplasma* further enhances its chances for a successful colonization of *I. scapularis* ticks by induction of an antifreeze protein (IAFGP) during the infection of ticks [205]. This secreted antifreeze glycoprotein inhibits bacterial biofilm formation through binding to the D-alanine residue of some bacteria peptidoglycan and was induced in response to *Anaplasma* infection [206, 207]. IAFGP expression resulted in thinning of the tick midgut PM and RNAi silencing of *iafgp* gene resulted in the absence of *Anaplasma* in the tick midgut [205].

Tick lectins

Ticks, like other arthropods, lack specific adaptive immunity. To defend themselves against invading microorganisms, ticks use the evolutionarily older nonspecific innate immune system, including both cellular and humoral immune responses. Cellular immune reactions involve haemocytes capable of phagocytosis, encapsulation or nodulation of foreign microorganisms and particles. The humoral immune response involves a range of non-specific pathogen-recognizing defence systems: PRRs, lectins, complement-like molecules, pro-phenoloxidase activation, haemolymph coagulation factors, antimicrobial peptides, reactive oxygen species, etc. Some of these molecules which function as mediators in the innate immune response are glycosylated and/or may recognize glycan-containing epitopes, e.g. recognition receptors for pathogens, complement-related molecules, or lectins (Table 3). In mammals, lectins play an important role in the recognition of specific glycosylated surface molecules of a variety of pathogens (PAMPs) and subsequent activation of the lectin pathway [208, 209]. MBL or ficolins known to recognize *N*-acetyl groups [210] serve as the recognition molecules, which are further integrated with the MBL-associated serine proteases to trigger the complement activation.

Fibrinogen-related proteins

Invertebrates contain a variety of fibrinogen-related proteins (FReps), all of them sharing structural similarity with fibrinogen. A common feature of FReps is their glycan-binding activity as they recognize the invading pathogen through its specific glycan epitopes. Their expression increases upon infection of the invertebrate by parasites or by pathogens [211, 212] with possibly a specific role in complement activation [213]. However, some of the tick FReps family proteins (such as ixoderins described below) may have various other functions (Table 3).

Dorin M from the soft tick *Ornithodoros moubata*, the first lectin purified and characterized from any tick species, shows a strong similarity to ficolins but lacks the N-terminal collagen domain [214, 215]. Dorin M and its closest homologue OMFREP, also from *O. moubata*, share sequence similarity with the innate immune FReps Tachylectin 5A and B from the horseshoe crab, *Tachypleus tridentatus* [215, 216]. It has a binding activity for sialic acid [214], its conjugates and *N*-acetyl-hexosamines. The protein has three *N*-glycosylation sites modified by high-mannose type glycans and core-fucosylated paucimannose glycans [217]. Other FReps were later identified in the haemolymph of *D. marginatus*, *R. appendiculatus*, *R. pulchellus* and *R. sanguineus* based on the cross-reactivity with sera directed against Dorin M [218].

The hard tick *I. ricinus* contains several FReP encoding sequences in its genome (ixoderins A, B and C) and their analogues are present in *I. scapularis* as well. While proteins similar to ixoderins A and C are present also in other tick species, ixoderin B-like proteins are found only in the genus *Ixodes*. All these proteins contain predicted glycosylation sites and they contain the fibrinogen-like domain with carbohydrate-binding properties [213, 215]. In *I. ricinus*, the expression of ixoderin A is restricted to haemocytes, salivary glands, and midgut while ixoderin B is only expressed in salivary glands [215]. As expected based on published information on other invertebrate

Table 3 Overview of identified tick lectins. Lectins identified in different tick species are listed including the tissue where the lectin was identified. Lectin binding specificity, its function and molecular weight are also listed if known

Lectin	Species	Tick tissue	Specificity	MW (kDa)	Function	Reference
Galectins (OmGalec)	*O. moubata*	Haemocytes, midgut, SG, ovaries	Lactosamine-like disaccharides	37.4	Putative functions in tick development, immunity, and vector-pathogen interaction	[221]
Dorin M	*O. moubata*	Haemocytes	*N*-acetyl-D-hexosamines and Sialic acid specific	na	Pattern recognition molecules	[214]
OMFREP	*O. moubata*	Hemolymph, salivary glands	Probably similar to Dorin M	na	Probably similar to Dorin M	[215]
Ixoderin A	*I. ricinus*	Hemolymph, salivary glands, midgut	Peptidoglycan recognition protein?	na	Putative defence protein, identification of self-/non-self tissues	[215, 219]
Ixoderin B	*I. ricinus*	Salivary glands	Unknown	na	Unknown putative immunomodulatory function	[215, 219]
Hemelipoglycoprotein	*D. marginatus*	Haemocytes, salivary glands, gut	Galactose- and mannose-binding specificity	290, 2 subunits	Putative innate immunity	[220]
Unknown lectin	*I. ricinus*	Gut, hemolymph	Sialic acid, *N*-acetyl-glucosamine	85	Putative recognition molecule	[233]
Unknown lectin	*I. ricinus*	SGs	Sialic acid	70	Unknown	[233]
TSLPI	*I. scapularis*	Unknown	Mannan	na	Unknown	[42]
HlCLec	*H. longicornis*	Midgut, ovary	Unknown	60.2	Unknown	[223]
Serpin 19	*A. americanum*	Saliva	GAGs	43.0	Serine protease inhibitor	[170]

Abbreviations: *MW* molecular weight, *na* not available

FReps, ixoderins are also involved in defence against pathogens. Namely, ixoderins A and B are involved in phagocytosis of some pathogens as shown for *Candida albicans* [219]. On the other hand, knockdown of these two ixoderins did not affect the phagocytosis of the tick-transmitted *B. afzelii* and knockdown of all three ixoderins does not affect its transmission [219]. The reason can be the missing protein glycosylation and thus the binding site for these lectins on the *Borrelia* surface [37]. Ixoderins and FReps can be involved in other processes as well; ixoderin B may be involved in the matrix attachment processes and angiogenesis inhibition. Alternatively, it may antagonize the effect of host ficolin [215].

Finally, one of the tick storage proteins, hemelipoglycoprotein, from several hard tick species seems to share a structural similarity to FReps with its primary sequence showing a high similarity to the fibrinogen domain [218, 220].

Other tick lectins

OmGalec from the soft tick *O. moubata* is the first member of galectin family identified in ticks with the specificity towards β1-3 and β1-4 bound galactose to GlcNAc, and Glc and α1-3 bound galactose to GalNAc [221]. Similar proteins are also present in *R. appendiculatus* and *I. scapularis* [188]. OmGalec contains two carbohydrate-binding domains which share low sequence similarity and thus possibly possesses a different saccharide-specificity. The protein is expressed in various life-stages and tissues, with the highest expression in haemocytes, midguts and ovaries [221]. It has been shown that galectins play a vital role in immune homeostasis by being pathogen recognition receptors [222].

C-type lectins are also present in the available tick genomes and transcriptomes [223, 224]. The only characterized C-type lectin from *Haemaphysalis longicornis* (HlCLec) contains three various carbohydrate-binding domains. Each of them has been shown to recognize the bacteria *E. coli* and *S. aureus* and participate in the tick defence against gram-negative bacteria, but they do not have a direct effect on bacterial growth. HlCLec also affects the blood-feeding process and affects larvae hatching and mortality. Expression of this lectin is increased during blood-feeding and is the highest in the midgut and ovary [223]. In mosquitoes, C-type lectins influence the midgut colonization by bacteria midgut microbiome [225] and facilitate infection with West Nile and dengue viruses [226, 227].

Calreticulin (CRT), a lectin chaperone responsible mainly for the control and proper folding of glycoproteins, is conserved in all tick species and is even used as the biomarker for human tick bites in *I. scapularis* [228]. In blood-feeding parasites, CRTs participate in evasion of the host defence mechanisms, namely the complement by binding the initiator of this pathway, the C1q protein, or factor Xa participating in the blood coagulation [229]. In

mammals, CRT on the surface of neutrophils also binds C1q as well as other immune-related lectins [230]. Similarly, the salivary secreted CRT from *A. americanum* binds host C1q. On the other hand, it does not bind the factor Xa and does not inhibit the activation of the classical complement cascade and host haemostasis. The *A. americanum* CRT shares a very high sequence similarity with other tick CRTs and thus similar functions of tick CRT can be expected [231].

Several other lectins are characterized in *I. ricinus*, but have not been identified to date: the 37, 60, 65, and 73 kDa lectins from midgut showing haemagglutination activity [163, 232]. The 37 kDa lectin has a binding specificity towards β1-3 glucan, while the 65 kDa protein binds bovine submaxillary mucin, containing a complicated mixture of various glycan structures and more specifically binds free sialic acid. Another lectin is present in haemolymph/haemocytes with a molecular weight of 85 kDa. It is a C-type lectin with specificity towards sialic acid and GlcNAc [233]. Several other lectins with haemagglutination activity have also been described in other ticks including *R. appendiculatus* [234, 235], *O. tartakovskyi*, *O. tholozani* and *A. polonicus* [233].

Tick glycans

Regarding the glycans and glycoproteins of blood-feeding arthropods, several studies describe these molecules using lectin staining and other indirect methods. Lectin studies show the presence of both *N*- and *O*-glycosylated proteins in tick tissues and some glycoproteins have been shown to be antigenic determinants for the immune response of the host [236–239]. In recent years, the direct determination of glycan structures and composition, mostly using mass spectrometry, has also been published, either from tick tissues and cells [203, 240] or purified proteins [220]. The three most interesting glycan structures related to host-parasite interaction and host immune system reaction are described below; representation of these structures in a glycan molecule is shown in Fig. 2. An overview of tick glycans with known structures is listed in Table 4.

Alpha-galactose epitope

Alpha-galactose epitopes (Galα1-3Gal; αGal) are abundant on glycolipids and glycoproteins of plants, arthropods and non-primate mammals [241]. αGal is a novel allergen identified first during clinical trials in 2004 in patients treated with cetuximab, a medical preparation for metastatic colorectal cancer treatment. Several cases of hypersensitivity reaction were registered soon after cetuximab administration into the blood due to the presence of αGal in its structure. The majority of sensitive individuals come from a population in south-eastern USA [242]. Furthermore, the geographical distribution of cases with cetuximab hypersensitivity corresponded to the distribution of red meat allergy cases and tick prevalence. Additionally, patients with red meat allergy experienced a tick bite in the months preceding the allergy symptoms. The causative agent of the αGal sensitization in the south-eastern region of the USA is the lone star tick *A. americanum* [243]. Red meat allergy is also linked with *I. holocyclus* tick bite in the Australian population [244] . Conversely, a bite by the *I. scapularis* tick from the same genus in the USA does not seem to result in red meat allergy [243]. Lastly, Chinuki et al. [245] described the allergy development upon *H. longicornis* bite in Japan. Direct evidence on αGal epitopes presence in *I. ricinus* is provided by Hamsten et al. [246], specifically in the tick midgut. However, the presence of αGal just in the tick saliva is what is important for patient sensitization. In this regard, the presence of undigested complete host proteins and glycoproteins was described in the tick body and, importantly, in the tick saliva [240, 246, 247] and thus the presence of αGal originating in the blood of non-mammalian hosts from the previous blood-feeding can be expected in the saliva. The αGal epitope is only known to be present in the saliva of *A. sculptum*, a tick that until now has not been connected with red meat allergy cases [248].

Core α1,3-fucosylation

The allergenic core α1,3-fucose (α1,3-Fuc) attached on the proximal GlcNAc residue is widely present in plants and arthropods and is one of the well-known possible human allergens as it is usually absent in mammals. It can induce production of specific IgE antibodies associated with IgE-mediated allergic immune responses, which is mostly described for schistosomes or venoms of some species of the order Hymenoptera. However, such a response is not described after a tick bite [249–251]. It is rather surprising, as α1,3-fucosylated structures are present in the tick salivary glands as well as in saliva of both *I. ricinus* and *I. scapularis* [203, 252]. This can be explained by the structural features of the allergenic epitopes; for example, in the case of core α1,3-Fuc, terminal GlcNAc weakens the immune response [253]. Additionally, more than one epitope has to be present to trigger the allergic reaction and the presence of blocking IgG4 antibodies against this epitope can lower the immune reaction [249].

The α1,3-Fuc modification of the *N*-linked glycan core mediates an entrance of one of the tick-transmitted pathogens, *A. phagocytophilum*, into *I. scapularis* midgut cells, but it is not required for the transmission of the pathogen to a vertebrate host. Furthermore, *Anaplasma* increases the expression of α1,3-fucosyltransferases in the tick, further increasing its ability to infect the tick. On the other hand, the infection of the tick by *B. burgdorferi* was not affected by the presence or absence of core α1,3-Fuc [203].

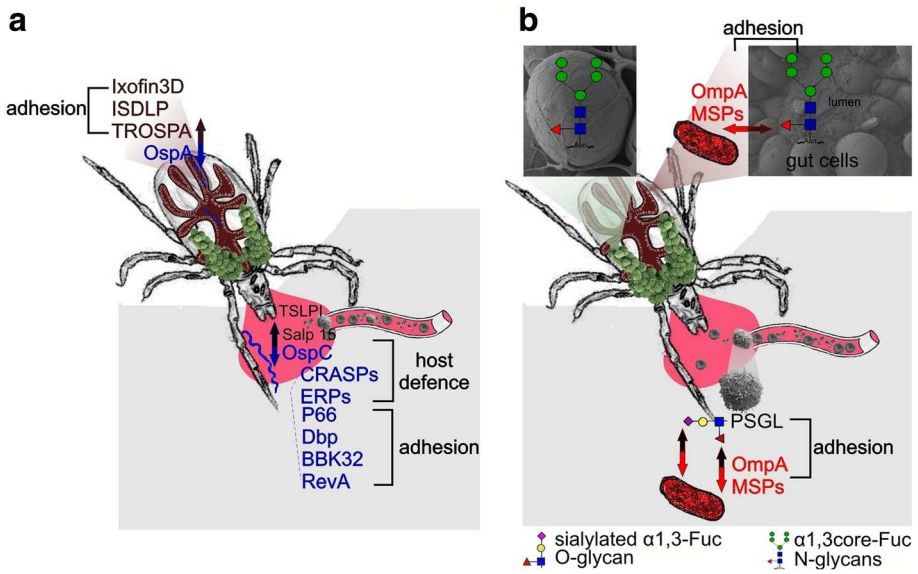


Fig. 2 Scheme of a model complex-type glycan showing presented oligosaccharide structures. An example of an *N*-glycan bearing the three glycoepitopes immunogenic in the mammalian hosts are shown. The αGal epitope is formed by a terminal galactose bound to another galactose *via* an α1-3 bond. In the case of the core α1,3-fucose, both the specific α1,3 bond and the core (not terminal localization of fucose are important for the immunogenicity in mammals. Two types of sialic acid are present in Eukaryotes: the *N*-acetylneuraminic acid (Neu5Ac) and the *N*-glycolylneuraminic acid (Neu5Gc). As humans do not possess the enzymatic apparatus for the production of Neu5Gc into glycans, glycans terminated with this type of sialic acid are immunogenic in humans. The used symbol nomenclature is based on the Consortium for Functional Glycomics (http://www.functionalglycomics.org/)

Sialic acids

Sialic acids (Sia) are found typically in the terminal position of vertebrate complex *N*- or *O*-linked glycans. In insects, some studies have shown the ability of sialylation [254, 255] and the importance of sialylation for insect-development, even though the abundance of sialylated glycans is very low [256].

N-glycans terminated with Sia are present also in the organs of the tick *I. ricinus*, namely in the gut, salivary glands, ovary and Malpighian tubules [240, 257]. However, the sialylated proteins in the adult ticks originate most probably from the host blood [258]. Hypothetically, sialoglycans present in the tick organs and in the secreted tick saliva can be engaged in molecular mimicry. We suppose that sialic acid is produced also by the tick itself in the ovary and eggs and later in larvae; the exact role of the tick sialylated proteins for the physiology and development of ticks is not yet known (unpublished results). Both eukaryotic types of sialic acids, *N*-acetyl-neuraminic acid and *N*-glycolyl-neuraminic acid (Fig. 2), were detected in the ticks [240].

N-linked glycans of flaviviruses

Tick-borne encephalitis virus (TBEV), a member of the genus *Flavivirus*, can cause serious infections in humans, which may result in encephalitis/meningoencephalitis. The viral single-stranded genomic RNA of positive polarity contains one open reading frame, which encodes a single polyprotein that is co-translationally and post-translationally cleaved by viral and cellular proteases into three structural and seven non-structural proteins (Table 4) [259].

Flaviviral non-structural proteins

Non-structural proteins of the flavivirus family (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) do not have their precise role elucidated, but they are generally considered as the effectors of viral replication, which occurs in close association with cellular membranes. Dramatic changes in the intracellular membrane structures including convoluted membranes, vesicle packets or paracrystalline arrays were observed as a result of dengue virus (DENV), TBEV or West Nile virus (WNV) replication [260–262]. Recently, NS1, NS4A and possibly also NS2A, were described to be involved in the formation of vesicle packets [263, 264]. Moreover, novel functions in terms of virus-host interactions were recently described for particular NS proteins; for example, TBEV NS5 protein acts as an inhibitor of interferon-activated Jak-STAT signalling [265].

Flaviviral structural proteins

Apart from seven non-structural proteins, the flaviviral genome encodes three structural proteins (C, M, E). The flaviviral nucleocapsid is composed of (+) ssRNA genome and the capsid protein (C), whereas the host-derived

Table 4 Overview of identified tick glycan structures composition. Monosaccharide compositions of the identified N-glycans are shown. Note that in some cases, the same composition can define various structures. For each glycan, the protein or the tick samples is listed, in which it was identified by mass spectrometry

Glycan composition				Protein/sample	Reference
Paucimannose glycans					
HexNAc	Hex	dHex	Sia		
2	3	0	0	Dorin M (position ^{41}NHS, ^{171}NGS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	4	0	0	Dorin M (position ^{41}NHS, ^{171}NGS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
High-mannose glycans					
HexNAc	Hex	dHex	Sia		
2	5	0	0	Dorin M (position ^{41}NHS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	6	0	0	Dorin M (position ^{41}NHS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	7	0	0	Dorin M (position ^{41}NHS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	8	0	0	Dorin M (position ^{41}NHS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	9	0	0	Dorin M (position ^{41}NHS, ^{129}NHS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				D. marginatus Hemelipoglycoprotein	[220]
				I. scapularis salivary gland	[203]
2	10	0	0	*D. marginatus* Hemelipoglycoprotein	[220]
				I. scapularis salivary gland	[203]
Core-fucosylated glycans					
HexNAc	Hex	dHex	Sia		
2	3	1	0	Dorin M (position ^{171}NGS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
2	4	1	0	Dorin M (position ^{171}NGS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]

Table 4 Overview of identified tick glycan structures composition. Monosaccharide compositions of the identified N-glycans are shown. Note that in some cases, the same composition can define various structures. For each glycan, the protein or the tick samples is listed, in which it was identified by mass spectrometry *(Continued)*

Glycan composition				Protein/sample	Reference
				I. scapularis salivary gland	[203]
2	5	1	0	Dorin M (position ^{171}NGS) from *O. moubata*	[217]
				I. ricinus fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
3	2	1	0	*I. ricinus* fed female salivary glands	[240]
3	3	1	0	*I. ricinus* fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
				I. scapularis salivary gland	[203]
4	3	1	0	*I. ricinus* fed female midgut	[240]
				I. scapularis salivary gland	[203]
4	4	1	0	*I. scapularis* salivary gland	[203]
4	5	1	0	*I. ricinus* fed female midgut	[240]
				I. scapularis salivary gland	[203]
4	6	1	0	*I. ricinus* fed female midgut	[240]
5	3	1	0	*I. ricinus* fed female salivary glands	[240]
				I. scapularis salivary gland	[203]
5	5	1	0	*I. ricinus* fed female midgut	[240]
6	6	1	0	*I. ricinus* fed female midgut	[240]
Complex glycans					
HexNAc	Hex	dHex	Sia		
3	4	0	0	*I. ricinus* fed female salivary glands	[240]
4	5	0	0	*I. ricinus* fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
4	6	0	0	*I. ricinus* fed female midgut	[240]
4	7	0	0	*D. marginatus* Hemelipoglycoprotein	[220]
4	8	0	0	*D. marginatus* Hemelipoglycoprotein	[220]
5	6	0	0	*I. ricinus* fed female midgut	[240]
6	2	0	0	*I. ricinus* fed female midgut	[240]
Sialylated glycans (containing either Neu5Ac or Neu5Gc)					
HexNAc	Hex	dHex	Sia		
4	5	0	1	*I. ricinus* fed female salivary glands	[240]
				I. ricinus fed female midgut	[240]
4	5	0	2	*I. ricinus* fed female midgut	[240]
5	6	0	1	*I. ricinus* fed female midgut	[240]
5	6	0	2	*I. ricinus* fed female midgut	[240]

Abbreviations: *HexNAc N*-acetyl-hexosamine (*N*-acetyl-glucosamine or *N*-acetyl-galactosamine), *Hex* hexose (mannose, glucose, galactose), *dHex* deoxyhexose (fucose), *Sia* sialic acid (*N*-acetyl-neuraminic acid, *N*-glycolyl-neuraminic acid)

envelope contains two glycoproteins, the membrane (prM/M) protein and the envelope (E) protein [266].

The E glycoprotein is localized in the viral envelope and is the main antigenic determinant of TBEV inducing a humoral immune response. It mediates fusion of TBEV with host cell membrane and thus facilitates the virus entry to the host cell. It forms heterodimers with the prM protein; the prM-E protein interaction is essential for proper folding of E protein [267]. The heterodimers then migrate to the ER membrane and eventually

bud off as nucleocapsid-containing immature virions [260, 268].

The M glycoprotein is an integral part of the viral envelope together with E protein. It forms heterodimers with the E protein and functions as a chaperon ensuring proper folding of E protein [267]. Non-infectious immature virions containing prM/E homodimers undergo maturation process in late trans-Golgi network by cleavage of pr part from the prM protein by host protease furin. The cleavage produces protein M and triggers re-organization of protein E to form homodimers [268, 269].

TBEV proteins glycosylation

The glycosylation of viral proteins increases their folding efficiency and promotes their intracellular transport by the interaction with host lectins [270]. In several viruses that cross the endoplasmic reticulum during their life-cycle, the protein glycosylation was proven important for virus growth, budding, secretion, and pathogenicity (reviewed in [271–273]). So far, only membrane glycoproteins prM/M and E are known to be *N*-glycosylated in TBEV (Table 5) [274–276]. α1,3-core fucosylated, high-mannose and hybrid N-glycans were shown to be present in the E protein of TBEV produced in chicken embryos by affinoblots [277]. E protein glycosylation was also confirmed recently also using cryo-electron microscopy, even though the exact glycan structure was not defined [278]. Another two sites (N130 and N207) in the NS1 protein are *N*-glycosylated in the case of dengue virus [274, 279]. One of these two *N*-glycosylation sites (N207) is present also in the TBEV NS1 protein; however, its glycosylation has not yet been shown (Table 5).

The E glycoprotein is a viral surface protein and thus contains major antigenic epitopes responsible for triggering the host immune system [280]. The *N*-glycosylation site at position N154 is present in the majority of TBEV strains and other flaviviruses. Moreover, a potential N361 glycosylation site is present in TBEV E protein as well. Depending on the strain, zero to two glycans are attached [280, 281]. The investigation of the presence and position of E protein glycans showed that an increased number of E protein glycans elevate its expression. Conversely, the E protein glycosylation deletion mutants showed reduced E protein production, suggesting the importance of glycosylation for the viral life-cycle [276].

Importantly, the absence of E-protein glycosylation affects the E protein conformation and further the TBEV infectivity only in the mammalian host, but not in the tick vector. Different temperatures in the host (37 °C) and the vector (23 °C) do not affect the stability of the deglycosylated E protein [282]. In light of this evidence, the E protein glycosylation seems to represent one of the factors conferring different vector and host competence. Interestingly, the investigation of mosquito-borne flaviviruses, DENV and WNV, shows that glycans modifying E protein are important for virus propagation in both vector and host cells. However, the role of the particular *N*-linked glycosylation site varies depending on the invertebrate/vertebrate host [283–285]. For example, the importance of *N*-linked glycan at 154 aa of WNV E protein was proved in case of the vector (*Culex pipiens* and *Cx. tarsalis* mosquitoes) as well as the bird host (*Gallus gallus*) [286, 287].

The prM protein encodes for the precursor of membrane protein M and also contains one *N*-linked glycosylation consensus sequence in N32 position (Table 5) [267, 276]. During TBEV maturation, the structural proteins prM and E form heterodimer, where prM has a chaperone-like role in the folding and maturation of E protein [267], although the biological role of TBEV prM glycosylation, has not yet been elucidated. However, Goto et al. [279] suggest the participation of carbohydrate-mediated interaction for prM-E heterodimer formation; glycosylation-deficient mutant of prM reduces the secretion of E protein to 60% in comparison to the wild-type prM. Further evidence for the crucial role of prM glycan was described in the case of WNV.

Table 5 TBEV protein glycosylation overview. List of TBEV proteins and their functions. Identified or predicted N-linked glycosylation sites are listed as well. NetNGlyc 1.0 Server was used for N-linked glycosylation site prediction

	Protein	Function	*N*-linked glycosylation	Reference
Structural	C	Capsid protein; forming of nucleocapsid	None	
	prM/M	Envelope protein; E protein chaperone	N32	[265, 302]
	E	Envelope protein; binding and fusion	N154, N361	[263–265]
Non-structural	NS1	Replication	Predicted: N85, N207	
	NS2A	Assembly, replication	None	
	NS2B	NS3 serine-protease cofactor	None	
	NS3	Serine-protease, helicase, replication RNA triphosphatase	Predicted: N160, N499, N555	
	NS4A	Assembly, replication	None	
	NS4B	Assembly, induction of membrane rearrangements	Predicted: N188	
	NS5	Methyltransferase, RNA-dependent RNA polymerase	Predicted: N18, N175, N215	

The prM glycosylation-deficient mutants decreased the formation and release of virus-like particles as well as genome copies. However, the infectivity of prM glycosylation-deficient mutants was not affected in mosquito, avian or mammalian cell lines [284].

In summary, *N*-linked glycosylation of TBEV prM and E proteins represents a multifaceted factor which is involved in many steps of the viral life-cycle, especially in virion assembly/secretion, and host/vector competence. Despite various studies, there are many aspects which need to be elucidated, especially the role of viral protein *N*-linked glycans within tick vectors. Moreover, the presence of other *N*-linked glycans in predicted sites of NS1, NS3, NS4B, and NS5 remains to be determined as well as their potential function.

Conclusions

The recent decades have provided an outstanding amount of new data about glycoconjugates and a growing line of evidence highlights the importance of carbohydrate-based interactions in the complex pathogen-host environment [288]. Glycoconjugates have an enormous structural diversity in the glycan moieties and therefore fulfil a variety of biological roles [289]. Given the fact that glycoconjugates are the major components of the outer surface of animal cells [290], it is likely that all interactions of microbial pathogens with their hosts/vectors are affected to a certain degree by the pattern of glycans and glycan-binding molecules that each produces. Despite the fact that protein glycosylation in the field of tick-borne pathogens has become a subject of increased attention in the last decade [37, 218, 291], there is still a deep knowledge gap regarding the nature and the specific roles of the glycoconjugates in the infectious cycle of these pathogens. All the findings mentioned in this review have tackled the important, yet still inadequately explored, the field of the carbohydrate-based interactions at the pathogen-tick-host interface. The basis of these interactions needs to be further addressed to gain clearer insights into the intricate strategies that the parasites employ to successfully finish their life-cycles. Ultimately, the common goal of scientists working in any field dealing with infectious diseases is to find an effective countermeasure against the particular threat. Ticks transmit a great variety of bacterial, viral and protozoan pathogens and therefore the search for a potent vaccine against each of these parasites costs an enormous amount of effort and money. One of the most promising strategies to cope with all pathogens transmitted by ticks is the development of a general anti-tick vaccine [292]. The potentially important role of sugar moieties in such a tick vaccine has already been suggested, showing, for instance, the tick midgut protein Bm86 to be more immunogenic in glycosylated form than non-glycosylated [291]. However, the progress in this matter is still insufficient and intense analysis of glycosylation needs to be addressed in future studies in order to be applied to the development of new therapeutics. Modern glycan sequencing technologies and strategies that allow site-specific mass-spectrometric identification of proteins with glycan modifications in a complex biological sample have shown that glycosylation could be much more extensive than previously thought [293].

Abbreviations

ACG: Cholesteryl 6-O-acyl-β-D-galactopyranoside; ACGal/Bb-GL-1: Cholesteryl 6-O-palmitoyl-β-D-galactopyranoside; AmOmp: *Anaplasma marginale* outer membrane protein; AP: Alternative complement pathway; ApOmp: *Anaplasma phagocytophilum* outer membrane protein; Bgp: *Borrelia* GAG-binding protein; CLR: C-type lectin receptor; CP: Classical complement pathway; CRASP: Complement regulator-acquiring surface protein; CRT: Calreticulin; DAF: Decay accelerating factor; DbpA: Decorin-binding protein; DENV: Dengue virus; ECM: Extracellular matrix; Erp: OspE/F-related proteins; FH: Complement factor H; FHL: Factor H-like; Fn: Fibrinogen; FREP: Fibrinogen-related protein; Fuc: Fucose; GAG: Glycosaminoglycan; GalCer: Galactosylceramide; IAFGP: *Ixodes scapularis* antifreeze glycoprotein; ISDLP: *Ixodes scapularis* dystroglycan-like protein; LacCer: Lactosylceramide; LP: Lectin complement pathway; LPS: Lipopolisaccharides; MAC: Membrane attack complex; MBL: Mannose-binding lectin; MGalD: Mono-α-galactosyl-diacylglycerol; MSP: Major surface proteins; MW: Molecular weight; na: Not available; NKT: Natural killer T cells; NLR/NOD-like receptors: Nucleotide-oligomerization domain-like receptors; Omp: Outer membrane protein; Osp: Outer surface protein; PAMP: Pathogen-associated molecular pattern; PM: Peritrophic matrix; PRR: Pattern recognition receptor; PSGL–1: P-selectin glycoprotein ligand-1; RGD: Arg-Gly-Asp tripeptide, a binding motif of fibronectin; Salp: Salivary protein; SG: Salivary glands; Sia: Sialic acid; TBEV: Tick-borne encephalitis virus; TLR: Toll-like receptors; TROSPA: Tick receptor for OspA; TSLPI: Tick salivary lectin pathway inhibitors; Vmp: Variable major protein; Vsp: Variable small protein; Vtp: Variable tick protein; WNV: West Nile virus

Funding

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic projects Postdok_BIOGLOBE (CZ.1.07/2.3.00/30.0032), C4SYS (LM2015055), and Czech-BioImaging (LM2015062), by the Czech Science Foundation (15-03044S), INTER-ACTION project No. LTARF18021, and by the European Union FP7 projects ANTIDotE (602272-2) and ANTIGONE (278976).

Authors' contributions

All authors contributed to the manuscript preparation. PV and JS contributed to this work equally. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 5 December 2017 Accepted: 14 August 2018
Published online: 14 November 2018

References

1. WHO. http://www.who.int/en/. 1948. Accessed 28 Mar 2018.
2. de la Fuente J, Antunes S, Bonnet S, Cabezas-Cruz A, Domingos AG, Estrada-Pena A, et al. Tick-pathogen interactions and vector competence:

identification of molecular drivers for tick-borne diseases. Front Cell Infect Microbiol. 2017;7:114.
3. Coumou J, Wagemakers A, Trentelman JJ, Nijhof AM, Hovius JW. Vaccination against Bm86 homologues in rabbits does not impair *Ixodes ricinus* feeding or oviposition. PLoS One. 2014;10:e0123495.
4. Semenza JC, Suk JE. Vector-borne diseases and climate change: a European perspective. FEMS Microbiol Lett. 2018;365:2.
5. Dinglasan RR, Jacobs-Lorena M. Insight into a conserved lifestyle: protein-carbohydrate adhesion strategies of vector-borne pathogens. Infect Immun. 2005;73:7797–807.
6. Severo MS, Choy A, Stephens KD, Sakhon OS, Chen G, Chung DW, et al. The E3 ubiquitin ligase XIAP restricts *Anaplasma phagocytophilum* colonization of *Ixodes scapularis* ticks. J Infect Dis. 2013;208:1830–40.
7. Varki ACR, Esko JD, et al. Essentials of Glycobiology. 3rd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2017.
8. Ritchie GE, Moffatt BE, Sim RB, Morgan BP, Dwek RA, Rudd PM. Glycosylation and the complement system. Chem Rev. 2002;102:305-20-19.
9. Bann JG, Bachinger HP. Glycosylation/hydroxylation-induced stabilization of the collagen triple helix. 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. J Biol Chem. 2000;275:24466–9.
10. Ault BH, Schmidt BZ, Fowler NL, Kashtan CE, Ahmed AE, Vogt BA, et al. Human factor H deficiency. Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. J Biol Chem. 1997;272:25168–75.
11. Dijk M, Holkers J, Voskamp P, Giannetti BM, Waterreus WJ, van Veen HA, et al. How dextran sulfate affects C1-inhibitor activity: a model for polysaccharide potentiation. Structure. 2016;24:2182–9.
12. Stavenhagen K, Kayili HM, Holst S, Koeleman C, Engel R, Wouters D, et al. *N*- and *O*-glycosylation analysis of human C1-inhibitor reveals extensive mucin-type *O*-glycosylation. Mol Cell Proteomics. 2017;17:1225–38.
13. Minta JO. The role of sialic acid in the functional activity and the hepatic clearance of C1-INH. J Immunol. 1981;126:245–9.
14. Novotny MV, Alley WR. Recent trends in analytical and structural glycobiology. Curr Opin Chem Biol. 2013;17:832–40.
15. Oswald DM, Cobb BA. Emerging glycobiology tools: a Renaissance in accessibility. Cell Immunol. 2018; https://doi.org/10.1016/j.cellimm.2018.04.010.
16. Tytgat HL, van Teijlingen NH, Sullan RM, Douillard FP, Rasinkangas P, Messing M, et al. Probiotic gut microbiota isolate interacts with dendritic cells *via* glycosylated heterotrimeric pili. PLoS One. 2016;11:e0151824.
17. Barbour AG, Hayes SF. Biology of *Borrelia* species. Microbiol Rev. 1986; 50:381–400.
18. Coburn J, Fischer JR, Leong JM. Solving a sticky problem: new genetic approaches to host cell adhesion by the Lyme disease spirochete. Mol Microbiol. 2005;57:1182–95.
19. Berndtson K. Review of evidence for immune evasion and persistent infection in Lyme disease. Int J Gen Med. 2013;6:291–306.
20. Cohen M. Notable aspects of glycan-protein interactions. Biomolecules. 2015;5:2056–72.
21. Vancova M, Nebesarova J, Grubhoffer L. Lectin-binding characteristics of a Lyme borreliosis spirochete *Borrelia burgdorferi sensu stricto*. Folia Microbiol. 2005;50:229–38.
22. Stevenson B, Schwan TG, Rosa PA. Temperature-related differential expression of antigens in the Lyme disease spirochete, *Borrelia burgdorferi*. Infect Immun. 1995;63:4535–9.
23. Pal U, de Silva AM, Montgomery RR, Fish D, Anguita J, Anderson JF, et al. Attachment of *Borrelia burgdorferi* within *Ixodes scapularis* mediated by outer surface protein A. J Clin Invest. 2000;106:561–9.
24. Schwan TG, Piesman J. Vector interactions and molecular adaptations of Lyme disease and relapsing fever spirochetes associated with transmission by ticks. Emerg Infect Dis. 2002;8:115–21.
25. Fikrig E, Pal U, Chen M, Anderson JF, Flavell RA. OspB antibody prevents *Borrelia burgdorferi* colonization of *Ixodes scapularis*. Infect Immun. 2004;72:1755–9.
26. Schwan TG, Piesman J, Golde WT, Dolan MC, Rosa PA. Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. Proc Natl Acad Sci USA. 1995;92:2909–13.
27. Ramamoorthi N, Narasimhan S, Pal U, Bao F, Yang XF, Fish D, et al. The Lyme disease agent exploits a tick protein to infect the mammalian host. Nature. 2005;436:573–7.
28. Simo L, Kazimirova M, Richardson J, Bonnet SI. The essential role of tick salivary glands and saliva in tick feeding and pathogen transmission. Front Cell Infect Microbiol. 2017;7:281.
29. Anguita J, Ramamoorthi N, Hovius JWR, Das S, Thomas V, Persinski R, et al. Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4(+) T cell activation. Immunity. 2002;16:849–59.
30. Kolb P, Vorreiter J, Habicht J, Bentrop D, Wallich R, Nassal M. Soluble cysteine-rich tick saliva proteins Salp15 and Iric-1 from *E. coli*. FEBS Open Bio. 2015;5:42–55.
31. Hovius JW, Schuijt TJ, de Groot KA, Roelofs JJ, Oei GA, Marquart JA, et al. Preferential protection of *Borrelia burgdorferi sensu stricto* by a Salp15 homologue in *Ixodes ricinus* saliva. J Infect Dis. 2008;198:1189–97.
32. Hovius JW, de Jong MA, den Dunnen J, Litjens M, Fikrig E, van der Poll T, et al. Salp15 binding to DC-SIGN inhibits cytokine expression by impairing both nucleosome remodeling and mRNA stabilization. PLoS Pathog. 2008;4:e31.
33. Lam TT, Nguyen TP, Montgomery RR, Kantor FS, Fikrig E, Flavell RA. Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. Infect Immun. 1994;62:290–8.
34. Casjens S, van Vugt R, Tilly K, Rosa PA, Stevenson B. Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. J Bacteriol. 1997;179:217–27.
35. Akins DR, Caimano MJ, Yang X, Cerna F, Norgard MV, Radolf JD. Molecular and evolutionary analysis of *Borrelia burgdorferi* 297 circular plasmid-encoded lipoproteins with OspE- and OspF-like leader peptides. Infect Immun. 1999;67:1526–32.
36. Sambri V, Stefanelli C, Cevenini R. Detection of glycoproteins in *Borrelia burgdorferi*. Arch Microbiol. 1992;157:205–8.
37. Sterba J, Vancova M, Rudenko N, Golovchenko M, Tremblay TL, Kelly JF, et al. Flagellin and outer surface proteins from *Borrelia burgdorferi* are not glycosylated. J Bacteriol. 2008;190:2619–23.
38. Eicken C, Sharma V, Klabunde T, Lawrenz MB, Hardham JM, Norris SJ, et al. Crystal structure of Lyme disease variable surface antigen VlsE of *Borrelia burgdorferi*. J Biol Chem. 2002;277:21691–6.
39. Narasimhan S, Coumou J, Schuijt TJ, Boder E, Hovius JW, Fikrig E. A tick gut protein with fibronectin III domains aids *Borrelia burgdorferi* congregation to the gut during transmission. PLoS Pathog. 2014;10:e1004278.
40. Coumou J, Narasimhan S, Trentelman JJ, Wagemakers A, Koetsveld J, Ersoz JI, et al. *Ixodes scapularis* dystroglycan-like protein promotes *Borrelia burgdorferi* migration from the gut. J Mol Med (Berl). 2016;94:361–70.
41. Narasimhan S, Santiago F, Koski RA, Brei B, Anderson JF, Fish D, et al. Examination of the *Borrelia burgdorferi* transcriptome in *Ixodes scapularis* during feeding. J Bacteriol. 2002;184:3122–5.
42. Schuijt TJ, Coumou J, Narasimhan S, Dai J, Deponte K, Wouters D, et al. A tick mannose-binding lectin inhibitor interferes with the vertebrate complement cascade to enhance transmission of the Lyme disease agent. Cell Host Microbe. 2011;10:136–46.
43. Schuijt TJ, Narasimhan S, Daffre S, DePonte K, Hovius JW, Van't Veer C, et al. Identification and characterization of *Ixodes scapularis* antigens that elicit tick immunity using yeast surface display. PLoS One. 2011;6:e15926.
44. Coburn J, Leong J, Chaconas G. Illuminating the roles of the *Borrelia burgdorferi* adhesins. Trends Microbiol. 2013;21:372–9.
45. Leong JM, Robbins D, Rosenfeld L, Lahiri B, Parveen N. Structural requirements for glycosaminoglycan recognition by the Lyme disease spirochete, *Borrelia burgdorferi*. Infect Immun. 1998;66:6045–8.
46. Brissette CA, Gaultney RA. That's my story, and I'm sticking to it - an update on *B. burgdorferi* adhesins. Front Cell Infect Microbiol. 2014;4:41.
47. Szczepanski A, Furie MB, Benach JL, Lane BP, Fleit HB. Interaction between *Borrelia burgdorferi* and endothelium *in vitro*. J Clin Invest. 1990;85:1637–47.
48. Guo BP, Norris SJ, Rosenberg LC, Hook M. Adherence of *Borrelia burgdorferi* to the proteoglycan decorin. Infect Immun. 1995;63:3467–72.
49. Choi HU, Johnson TL, Pal S, Tang LH, Rosenberg L, Neame PJ. Characterization of the dermatan sulfate proteoglycans, DS-PGI and DS-PGII, from bovine articular cartilage and skin isolated by octyl-sepharose chromatography. J Biol Chem. 1989;264:2876–84.
50. Rosenberg LC, Choi HU, Tang LH, Johnson TL, Pal S, Webber C, et al. Isolation of dermatan sulfate proteoglycans from mature bovine articular cartilages. J Biol Chem. 1985;260:6304–13.
51. Sasaki T, Fassler R, Hohenester E. Laminin: the crux of basement membrane assembly. J Cell Biol. 2004;164:959–63.
52. Kumar AP, Nandini CD, Salimath PV. Structural characterization of N-linked oligosaccharides of laminin from rat kidney: changes during diabetes and modulation by dietary fiber and butyric acid. FEBS J. 2011;278:143–55.
53. Singh B, Fleury C, Jalalvand F, Riesbeck K. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol Rev. 2012;36:1122–80.

54. Cabello FC, Hulinska D, Godfrey HP. Molecular Biology of Spirochetes. Amsterdam: IOS Press; 2006.
55. Brissette CA, Verma A, Bowman A, Cooley AE, Stevenson B. The *Borrelia burgdorferi* outer-surface protein ErpX binds mammalian laminin. Microbiology. 2009;155:863–72.
56. Brissette CA, Cooley AE, Burns LH, Riley SP, Verma A, Woodman ME, et al. Lyme borreliosis spirochete Erp proteins, their known host ligands, and potential roles in mammalian infection. Int J Med Microbiol. 2008;298(Suppl. 1):257–67.
57. Hynes RO. Integrins: a family of cell surface receptors. Cell. 1987;48:549–54.
58. Cai X, Thinn AMM, Wang Z, Shan H, Zhu J. The importance of *N*-glycosylation on β3 integrin ligand binding and conformational regulation. Sci Rep. 2017;7:4656.
59. Coburn J, Magoun L, Bodary SC, Leong JM. Integrins αvβ3 and α5β1 mediate attachment of Lyme disease spirochetes to human cells. Infect Immun. 1998;66:1946–52.
60. Hynes RO. Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992;69:11–25.
61. Magoun L, Zuckert WR, Robbins D, Parveen N, Alugupalli KR, Schwan TG, et al. Variable small protein (Vsp)-dependent and Vsp-independent pathways for glycosaminoglycan recognition by relapsing fever spirochaetes. Mol Microbiol. 2000;36:886–97.
62. Porcella SF, Raffel SJ, Anderson DE Jr, Gilk SD, Bono JL, Schrumpf ME, et al. Variable tick protein in two genomic groups of the relapsing fever spirochete *Borrelia hermsii* in western North America. Infect Immun. 2005;73:6647–58.
63. Probert WS, Johnson BJ. Identification of a 47 kDa fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. Mol Microbiol. 1998;30:1003–15.
64. Brissette CA, Bykowski T, Cooley AE, Bowman A, Stevenson B. *Borrelia burgdorferi* RevA antigen binds host fibronectin. Infect Immun. 2009; 77:2802–12.
65. Li X, Liu X, Beck DS, Kantor FS, Fikrig E. *Borrelia burgdorferi* lacking BBK32, a fibronectin-binding protein, retains full pathogenicity. Infect Immun. 2006;74:3305–13.
66. Hyde JA, Weening EH, Chang M, Trzeciakowski JP, Hook M, Cirillo JD, et al. Bioluminescent imaging of *Borrelia burgdorferi in vivo* demonstrates that the fibronectin-binding protein BBK32 is required for optimal infectivity. Mol Microbiol. 2011;82:99–113.
67. Niddam AF, Ebady R, Bansal A, Koehler A, Hinz B, Moriarty TJ. Plasma fibronectin stabilizes *Borrelia burgdorferi*-endothelial interactions under vascular shear stress by a catch-bond mechanism. Proc Natl Acad Sci USA. 2017;114:E3490–E8.
68. Ebady R, Niddam AF, Boczula AE, Kim YR, Gupta N, Tang TT, et al. Biomechanics of *Borrelia burgdorferi* vascular interactions. Cell Rep. 2016;16:2593–604.
69. Lin YP, Chen Q, Ritchie JA, Dufour NP, Fischer JR, Coburn J, et al. Glycosaminoglycan binding by *Borrelia burgdorferi* adhesin BBK32 specifically and uniquely promotes joint colonization. Cell Microbiol. 2015;17:860–75.
70. Parveen N, Robbins D, Leong JM. Strain variation in glycosaminoglycan recognition influences cell-type-specific binding by Lyme disease spirochetes. Infect Immun. 1999;67:1743–9.
71. Moriarty TJ, Shi M, Lin YP, Ebady R, Zhou H, Odisho T, et al. Vascular binding of a pathogen under shear force through mechanistically distinct sequential interactions with host macromolecules. Mol Microbiol. 2012;86:1116–31.
72. Garcia BL, Zhi H, Wager B, Hook M, Skare JT. *Borrelia burgdorferi* BBK32 inhibits the classical pathway by blocking activation of the C1 complement complex. PLoS Pathog. 2016;12:e1005404.
73. Brissette CA, Rossmann E, Bowman A, Cooley AE, Riley SP, Hunfeld KP, et al. The borrelial fibronectin-binding protein RevA is an early antigen of human Lyme disease. Clin Vaccine Immunol. 2010;17:274–80.
74. Shi Y, Xu Q, McShan K, Liang FT. Both decorin-binding proteins A and B are critical for the overall virulence of *Borrelia burgdorferi*. Infect Immun. 2008;76:1239–46.
75. Salo J, Jaatinen A, Soderstrom M, Viljanen MK, Hytonen J. Decorin binding proteins of *Borrelia burgdorferi* promote arthritis development and joint specific post-treatment DNA persistence in mice. PLoS One. 2015;10:e0121512.
76. Benoit VM, Fischer JR, Lin YP, Parveen N, Leong JM. Allelic variation of the Lyme disease spirochete adhesin DbpA influences spirochetal binding to decorin, dermatan sulfate, and mammalian cells. Infect Immun. 2011;79:3501–9.
77. Morgan A, Sepuru KM, Feng W, Rajarathnam K, Wang X. Flexible linker modulates glycosaminoglycan affinity of decorin binding protein A. Biochemistry. 2015;54:5113–9.
78. Wang X. Solution structure of decorin-binding protein A from *Borrelia burgdorferi*. Biochemistry. 2012;51:8353–62.
79. Salo J, Loimaranta V, Lahdenne P, Viljanen MK, Hytonen J. Decorin binding by DbpA and B of *Borrelia garinii*, *Borrelia afzelii*, and *Borrelia burgdorferi sensu stricto*. J Infect Dis. 2011;204:65–73.
80. Imai DM, Samuels DS, Feng S, Hodzic E, Olsen K, Barthold SW. The early dissemination defect attributed to disruption of decorin-binding proteins is abolished in chronic murine Lyme borreliosis. Infect Immun. 2013;81:1663–73.
81. Schlachter S, Seshu J, Lin T, Norris S, Parveen N. The *Borrelia burgdorferi* glycosaminoglycan binding protein Bgp in the B31 strain is not essential for infectivity despite facilitating adherence and tissue colonization. Infect Immun. 2018;86:e00667–17.
82. Verma A, Brissette CA, Bowman A, Stevenson B. *Borrelia burgdorferi* BmpA is a laminin-binding protein. Infect Immun. 2009;77:4940–6.
83. Antonara S, Ristow L, Coburn J. Adhesion mechanisms of *Borrelia burgdorferi*. Adv Exp Med Biol. 2011;715:35–49.
84. Pal U, Wang P, Bao F, Yang X, Samanta S, Schoen R, et al. *Borrelia burgdorferi* basic membrane proteins A and B participate in the genesis of Lyme arthritis. J Exp Med. 2008;205:133–41.
85. Coburn J, Barthold SW, Leong JM. Diverse Lyme disease spirochetes bind integrin $\alpha_{IIb}\beta_3$ on human platelets. Infect Immun. 1994;62:5559–67.
86. Alugupalli KR, Michelson AD, Barnard MR, Robbins D, Coburn J, Baker EK, et al. Platelet activation by a relapsing fever spirochaete results in enhanced bacterium-platelet interaction *via* integrin $\alpha_{IIb}\beta_3$ activation. Mol Microbiol. 2001;39:330–40.
87. Coburn J, Chege W, Magoun L, Bodary SC, Leong JM. Characterization of a candidate *Borrelia burgdorferi* β3-chain integrin ligand identified using a phage display library. Mol Microbiol. 1999;34:926–40.
88. Skare JT, Mirzabekov TA, Shang ES, Blanco DR, Erdjument-Bromage H, Bunikis J, et al. The Oms66 (p66) protein is a *Borrelia burgdorferi* porin. Infect Immun. 1997;65:3654–61.
89. Barcena-Uribarri I, Thein M, Sacher A, Bunikis I, Bonde M, Bergstrom S, et al. P66 porins are present in both Lyme disease and relapsing fever spirochetes: a comparison of the biophysical properties of P66 porins from six *Borrelia* species. Biochim Biophys Acta. 1798;2010:1197–203.
90. Kenedy MR, Luthra A, Anand A, Dunn JP, Radolf JD, Akins DR. Structural modeling and physicochemical characterization provide evidence that P66 forms a β-barrel in the *Borrelia burgdorferi* outer membrane. J Bacteriol. 2014;196:859–72.
91. Coburn J, Cugini C. Targeted mutation of the outer membrane protein P66 disrupts attachment of the Lyme disease agent, *Borrelia burgdorferi*, to integrin αVβ3. Proc Natl Acad Sci USA. 2003;100:7301–6.
92. Ristow LC, Bonde M, Lin YP, Sato H, Curtis M, Wesley E, et al. Integrin binding by *Borrelia burgdorferi* P66 facilitates dissemination but is not required for infectivity. Cell Microbiol. 2015;17:1021–36.
93. Defoe G, Coburn J. Delineation of *Borrelia burgdorferi* p66 sequences required for integrin αIIbβ3 recognition. Infect Immun. 2001;69:3455–9.
94. Ristow LC, Miller HE, Padmore LJ, Chettri R, Salzman N, Caimano MJ, et al. The β3-integrin ligand of *Borrelia burgdorferi* is critical for infection of mice but not ticks. Mol Microbiol. 2012;85:1105–18.
95. Wood E, Tamborero S, Mingarro I, Esteve-Gassent MD. BB0172, a *Borrelia burgdorferi* outer membrane protein that binds integrin α3β1. J Bacteriol. 2013;195:3320–30.
96. Lambris JD, Ricklin D, Geisbrecht BV. Complement evasion by human pathogens. Nat Rev Microbiol. 2008;6:132–42.
97. Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. Nat Immunol. 2010;11:785–97.
98. Kraiczy P. Hide and seek: how Lyme disease spirochetes overcome complement attack. Front Immunol. 2016;7:385.
99. Rodriguez de Cordoba S, Esparza-Gordillo J, Goicoechea de Jorge E, Lopez-Trascasa M, Sanchez-Corral P. The human complement factor H: functional roles, genetic variations and disease associations. Mol Immunol. 2004;41:355–67.
100. Fenaille F, Le Mignon M, Groseil C, Ramon C, Riande S, Siret L, et al. Site-specific *N*-glycan characterization of human complement factor H. Glycobiology. 2007;17:932–44.
101. Kraiczy P, Wallich R. Borrelial complement-binding proteins. In: Embers ME, editor. The Pathogenic Spirochetes: Strategies for Evasion of Host Immunity and Persistence. Boston: Springer US; 2012. p. 63–88.

102. Hellwage J, Kuhn S, Zipfel PF. The human complement regulatory factor-H-like protein 1, which represents a truncated form of factor H, displays cell-attachment activity. Biochem J. 1997;326:321–7.
103. Metts MS, McDowell JV, Theisen M, Hansen PR, Marconi RT. Analysis of the OspE determinants involved in binding of factor H and OspE-targeting antibodies elicited during *Borrelia burgdorferi* infection in mice. Infect Immun. 2003;71:3587–96.
104. Wallich R, Pattathu J, Kitiratschky V, Brenner C, Zipfel PF, Brade V, et al. Identification and functional characterization of complement regulator-acquiring surface protein 1 of the Lyme disease spirochetes *Borrelia afzelii* and *Borrelia garinii*. Infect Immun. 2005;73:2351–9.
105. Kraiczy P, Skerka C, Kirschfink M, Brade V, Zipfel PF. Immune evasion of *Borrelia burgdorferi* by acquisition of human complement regulators FHL-1/reconectin and Factor H. Eur J Immunol. 2001;31:1674–84.
106. Kraiczy P, Hellwage J, Skerka C, Becker H, Kirschfink M, Simon MM, et al. Complement resistance of *Borrelia burgdorferi* correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins. J Biol Chem. 2004;279:2421–9.
107. Haupt K, Kraiczy P, Wallich R, Brade V, Skerka C, Zipfel PF. Binding of human factor H-related protein 1 to serum-resistant *Borrelia burgdorferi* is mediated by borrelial complement regulator-acquiring surface proteins. J Infect Dis. 2007;196:124–33.
108. von Lackum K, Miller JC, Bykowski T, Riley SP, Woodman ME, Brade V, et al. *Borrelia burgdorferi* regulates expression of complement regulator-acquiring surface protein 1 during the mammal-tick infection cycle. Infect Immun. 2005;73:7398–405.
109. Bykowski T, Woodman ME, Cooley AE, Brissette CA, Brade V, Wallich R, et al. Coordinated expression of *Borrelia burgdorferi* complement regulator-acquiring surface proteins during the Lyme disease spirochete's mammal-tick infection cycle. Infect Immun. 2007;75:4227–36.
110. Kenedy MR, Vuppala SR, Siegel C, Kraiczy P, Akins DR. CspA-mediated binding of human factor H inhibits complement deposition and confers serum resistance in *Borrelia burgdorferi*. Infect Immun. 2009;77:2773–82.
111. Hallstrom T, Siegel C, Morgelin M, Kraiczy P, Skerka C, Zipfel PF. CspA from *Borrelia burgdorferi* inhibits the terminal complement pathway. MBio. 2013;4:e00481-13.
112. Bhide MR, Travnicek M, Levkutova M, Curlik J, Revajova V, Levkut M. Sensitivity of *Borrelia* genospecies to serum complement from different animals and human: a host-pathogen relationship. FEMS Immunol Med Microbiol. 2005;43:165–72.
113. Herzberger P, Siegel C, Skerka C, Fingerle V, Schulte-Spechtel U, van Dam A, et al. Human pathogenic *Borrelia spielmanii* sp. nov. resists complement-mediated killing by direct binding of immune regulators factor H and factor H-like protein 1. Infect Immun. 2007;75:4817–25.
114. Hartmann K, Corvey C, Skerka C, Kirschfink M, Karas M, Brade V, et al. Functional characterization of BbCRASP-2, a distinct outer membrane protein of *Borrelia burgdorferi* that binds host complement regulators factor H and FHL-1. Mol Microbiol. 2006;61:1220–36.
115. Siegel C, Schreiber J, Haupt K, Skerka C, Brade V, Simon MM, et al. Deciphering the ligand-binding sites in the *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 required for interactions with the human immune regulators factor H and factor H-like protein 1. J Biol Chem. 2008;283:34855–63.
116. Kraiczy P, Seling A, Brissette CA, Rossmann E, Hunfeld KP, Bykowski T, et al. *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 (CspZ) as a serological marker of human Lyme disease. Clin Vaccine Immunol. 2008;15:484–91.
117. Coleman AS, Yang X, Kumar M, Zhang X, Promnares K, Shroder D, et al. *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 does not contribute to complement resistance or host infectivity. PLoS One. 2008;3:3010e.
118. Alitalo A, Meri T, Lankinen H, Seppala L, Lahdenne P, Hefty PS, et al. Complement inhibitor factor H binding to Lyme disease spirochetes is mediated by inducible expression of multiple plasmid-encoded outer surface protein E paralogs. J Immunol. 2002;169:3847–53.
119. Hammerschmidt C, Hallstrom T, Skerka C, Wallich R, Stevenson B, Zipfel PF, et al. Contribution of the infection-associated complement regulator-acquiring surface protein 4 (ErpC) to complement resistance of *Borrelia burgdorferi*. Clin Dev Immunol. 2012;2012:349657.
120. Siegel C, Hallstrom T, Skerka C, Eberhardt H, Uzonyi B, Beckhaus T, et al. Complement factor H-related proteins CFHR2 and CFHR5 represent novel ligands for the infection-associated CRASP proteins of *Borrelia burgdorferi*. PLoS One. 2010;5:e13519.
121. Hovis KM, Tran E, Sundy CM, Buckles E, McDowell JV, Marconi RT. Selective binding of *Borrelia burgdorferi* OspE paralogs to factor H and serum proteins from diverse animals: possible expansion of the role of OspE in Lyme disease pathogenesis. Infect Immun. 2006;74:1967–72.
122. McDowell JV, Wolfgang J, Tran E, Metts MS, Hamilton D, Marconi RT. Comprehensive analysis of the factor H binding capabilities of *Borrelia* species associated with Lyme disease: delineation of two distinct classes of factor H binding proteins. Infect Immun. 2003;71:3597–602.
123. Stevenson B, Bono JL, Schwan TG, Rosa P. *Borrelia burgdorferi* erp proteins are immunogenic in mammals infected by tick bite, and their synthesis is inducible in cultured bacteria. Infect Immun. 1998;66:2648–54.
124. Fikrig E, Narasimhan S. *Borrelia burgdorferi*-traveling incognito? Microbes Infect. 2006;8:1390–9.
125. Brooks CS, Vuppala SR, Jett AM, Alitalo A, Meri S, Akins DR. Complement regulator-acquiring surface protein 1 imparts resistance to human serum in *Borrelia burgdorferi*. J Immunol. 2005;175:3299–308.
126. van Dam AP, Oei A, Jaspars R, Fijen C, Wilske B, Spanjaard L, et al. Complement-mediated serum sensitivity among spirochetes that cause Lyme disease. Infect Immun. 1997;65:1228–36.
127. Sun J, Duffy KE, Ranjith-Kumar CT, Xiong J, Lamb RJ, Santos J, et al. Structural and functional analyses of the human Toll-like receptor 3. Role of glycosylation. J Biol Chem. 2006;281:11144–51.
128. Berende A, Oosting M, Kullberg BJ, Netea MG, Joosten LA. Activation of innate host defense mechanisms by *Borrelia*. Eur Cytokine Netw. 2010;21:7–18.
129. Leifer CA, Medvedev AE. Molecular mechanisms of regulation of Toll-like receptor signaling. J Leukocyte Biol. 2016;100:927–41.
130. Medzhitov R, Preston-Hurlburt P, Janeway CA Jr. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature. 1997;388:394–7.
131. Muzio M, Bosisio D, Polentarutti N, D'Amico G, Stoppacciaro A, Mancinelli R, et al. Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: selective expression of TLR3 in dendritic cells. J Immunol. 2000;164:5998–6004.
132. Wooten RM, Ma Y, Yoder RA, Brown JP, Weis JH, Zachary JF, et al. Toll-like receptor 2 is required for innate, but not acquired, host defense to *Borrelia burgdorferi*. J Immunol. 2002;168:348–55.
133. Bernardino AL, Myers TA, Alvarez X, Hasegawa A, Philipp MT. Toll-like receptors: insights into their possible role in the pathogenesis of Lyme neuroborreliosis. Infect Immun. 2008;76:4385–95.
134. Salazar JC, Pope CD, Moore MW, Pope J, Kiely TG, Radolf JD. Lipoprotein-dependent and -independent immune responses to spirochetal infection. Clin Diagn Lab Immunol. 2005;12:949–58.
135. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y, et al. Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. J Immunol. 1999;162:3749–52.
136. Hacker H, Mischak H, Miethke T, Liptay S, Schmid R, Sparwasser T, et al. CpG-DNA-specific activation of antigen-presenting cells requires stress kinase activity and is preceded by non-specific endocytosis and endosomal maturation. EMBO J. 1998;17:6230–40.
137. Hemmi H, Takeuchi O, Kawai T, Kaisho T, Sato S, Sanjo H, et al. A Toll-like receptor recognizes bacterial DNA. Nature. 2000;408:740–5.
138. Inohara N, Nunez G. NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol. 2003;3:371–82.
139. Franchi L, Warner N, Viani K, Nunez G. Function of Nod-like receptors in microbial recognition and host defense. Immunol Rev. 2009;227:106–28.
140. Kanneganti TD, Lamkanfi M, Nunez G. Intracellular NOD-like receptors in host defense and disease. Immunity. 2007;27:549–59.
141. Sterka D Jr, Marriott I. Characterization of nucleotide-binding oligomerization domain (NOD) protein expression in primary murine microglia. J Neuroimmunol. 2006;179:65–75.
142. Sterka D Jr, Marriott I. Functional expression of NOD2, a novel pattern recognition receptor for bacterial motifs, in primary murine astrocytes. Glia. 2006;53:322–30.
143. Petnicki-Ocwieja T, DeFrancesco AS, Chung E, Darcy CT, Bronson RT, Kobayashi KS, et al. Nod2 suppresses *Borrelia burgdorferi* mediated murine Lyme arthritis and carditis through the induction of tolerance. PLoS One. 2011;6:e17414.
144. Oosting M, Berende A, Sturm P, Ter Hofstede HJ, de Jong DJ, Kanneganti TD, et al. Recognition of *Borrelia burgdorferi* by NOD2 is central for the induction of an inflammatory reaction. J Infect Dis. 2010;201:1849–58.

145. Stubs G, Fingerle V, Wilske B, Gobel UB, Zahringer U, Schumann RR, et al. Acylated cholesteryl galactosides are specific antigens of *Borrelia* causing Lyme disease and frequently induce antibodies in late stages of disease. J Biol Chem. 2009;284:13326–34.
146. Ben-Menachem G, Kubler-Kielb J, Coxon B, Yergey A, Schneerson R. A newly discovered cholesteryl galactoside from *Borrelia burgdorferi*. Proc Natl Acad Sci USA. 2003;100:7913–8.
147. Radolf JD, Caimano MJ, Stevenson B, Hu LT. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. Nat Rev Microbiol. 2012;10:87–99.
148. Jones KL, Seward RJ, Ben-Menachem G, Glickstein LJ, Costello CE, Steere AC. Strong IgG antibody responses to *Borrelia burgdorferi* glycolipids in patients with Lyme arthritis, a late manifestation of the infection. Clin Immunol. 2009;132:93–102.
149. Comstock LE, Fikrig E, Shoberg RJ, Flavell RA, Thomas DD. A monoclonal antibody to OspA inhibits association of *Borrelia burgdorferi* with human endothelial cells. Infect Immun. 1993;61:423–31.
150. Wang J, Li Y, Kinjo Y, Mac TT, Gibson D, Painter GF, et al. Lipid binding orientation within CD1d affects recognition of *Borrelia burgorferi* antigens by NKT cells. Proc Natl Acad Sci USA. 2010;107:1535–40.
151. Garcia Monco JC, Wheeler CM, Benach JL, Furie RA, Lukehart SA, Stanek G, et al. Reactivity of neuroborreliosis patients (Lyme disease) to cardiolipin and gangliosides. J Neurol Sci. 1993;117:206–14.
152. Garcia-Monco JC, Seidman RJ, Benach JL. Experimental immunization with *Borrelia burgdorferi* induces development of antibodies to gangliosides. Infect Immun. 1995;63:4130–7.
153. Weller M, Stevens A, Sommer N, Dichgans J, Kappler B, Wietholter H. Ganglioside antibodies: a lack of diagnostic specificity and clinical utility? J Neurol. 1992;239:455–9.
154. Venkataswamy MM, Porcelli SA. Lipid and glycolipid antigens of CD1d-restricted natural killer T cells. Semin Immunol. 2010;22:68–78.
155. Hossain H, Wellensiek HJ, Geyer R, Lochnit G. Structural analysis of glycolipids from *Borrelia burgdorferi*. Biochimie. 2001;83:683–92.
156. Smith DG, Williams SJ. Immune sensing of microbial glycolipids and related conjugates by T cells and the pattern recognition receptors MCL and Mincle. Carbohydr Res. 2016;420:32–45.
157. Garcia Monco JC, Fernandez Villar B, Rogers RC, Szczepanski A, Wheeler CM, Benach JL. *Borrelia burgdorferi* and other related spirochetes bind to galactocerebroside. Neurology. 1992;42:1341–8.
158. Backenson PB, Coleman JL, Benach JL. *Borrelia burgdorferi* shows specificity of binding to glycosphingolipids. Infect Immun. 1995;63:2811–7.
159. Kaneda K, Masuzawa T, Yasugami K, Suzuki T, Suzuki Y, Yanagihara Y. Glycosphingolipid-binding protein of *Borrelia burgdorferi sensu lato*. Infect Immun. 1997;65:3180–5.
160. Hajnicka V, Kocakova P, Slavikova M, Slovak M, Gasperik J, Fuchsberger N, et al. Anti-interleukin-8 activity of tick salivary gland extracts. Parasite Immunol. 2001;23:483–9.
161. Frauenschuh A, Power CA, Deruaz M, Ferreira BR, Silva JS, Teixeira MM, et al. Molecular cloning and characterization of a highly selective chemokine-binding protein from the tick *Rhipicephalus sanguineus*. J Biol Chem. 2007;282:27250–8.
162. Eaton JRO, Alenazi Y, Singh K, Davies G, Geis-Asteggiante L, Kessler B, et al. The N-terminal domain of a tick evasin is critical for chemokine binding and neutralization and confers specific binding activity to other evasins. J Biol Chem. 2018; https://doi.org/10.1074/jbc.RA117.000487.
163. Uhlir J, Grubhoffer L, Borsky I, Dusbabek F. Antigens and glycoproteins of larvae, nymphs and adults of the tick *Ixodes ricinus*. Med Vet Entomol. 1994;8:141–50.
164. Mulenga A, Kim T, Ibelli AM. *Amblyomma americanum* tick saliva serine protease inhibitor 6 is a cross-class inhibitor of serine proteases and papain-like cysteine proteases that delays plasma clotting and inhibits platelet aggregation. Insect Mol Biol. 2013;22:306–19.
165. Tirloni L, Kim TK, Coutinho ML, Ali A, Seixas A, Termignoni C, et al. The putative role of *Rhipicephalus microplus* salivary serpins in the tick-host relationship. Insect Biochem Mol Biol. 2016;71:12–28.
166. Deruaz M, Frauenschuh A, Alessandri AL, Dias JM, Coelho FM, Russo RC, et al. Ticks produce highly selective chemokine binding proteins with antiinflammatory activity. J Exp Med. 2008;205:2019–31.
167. Pratt CW, Church FC. Heparin binding to protein C inhibitor. J Biol Chem. 1992;267:8789–94.
168. Rein CM, Desai UR, Church FC. Serpin-glycosaminoglycan interactions. Methods Enzymol. 2011;501:105–37.
169. Tollefsen DM. Heparin Cofactor II. In: Church FC, Cunningham DD, Ginsburg D, Hoffman M, Stone SR, Tollefsen DM, editors. Chemistry and Biology of Serpins (Advances in Experimental Medicine and Biology, Vol. 425). New York: Springer Science+Business Media; 1997.
170. Radulovic ZM, Mulenga A. Heparan sulfate/heparin glycosaminoglycan binding alters inhibitory profile and enhances anticoagulant function of conserved *Amblyomma americanum* tick saliva serpin 19. Insect Biochem Mol Biol. 2017;80:1–10.
171. Koh CY, Kazimirova M, Trimnell A, Takac P, Labuda M, Nuttall PA, et al. Variegin, a novel fast and tight binding thrombin inhibitor from the tropical bont tick. J Biol Chem. 2007;282:29101–13.
172. Shabareesh PRV, Kumar A, Salunke DM, Kaur KJ. Structural and functional studies of differentially *O*-glycosylated analogs of a thrombin inhibitory peptide - variegin. J Pept Sci. 2017;23:880–8.
173. Stuen S, Bergstrom K, Palmer E. Reduced weight gain due to subclinical *Anaplasma phagocytophilum* (formerly *Ehrlichia phagocytophila*) infection. Exp Appl Acarol. 2002;28:209–15.
174. Splitter EJ, Twiehaus MJ, Castro ER. Anaplasmosis in sheep in the United States. J Am Vet Med Assoc. 1955;127:244–5.
175. Melendez RD. Future perspectives on veterinary hemoparasite research in the tropics at the start of this century. Ann N Y Acad Sci. 2000;916:253–8.
176. Dumler JS, Bakken JS. Human ehrlichioses: newly recognized infections transmitted by ticks. Annu Rev Med. 1998;49:201–13.
177. Keesing F, Hersh MH, Tibbetts M, McHenry DJ, Duerr S, Brunner J, et al. Reservoir competence of vertebrate hosts for *Anaplasma phagocytophilum*. Emerg Infect Dis. 2012;18:2013–6.
178. Parola P, Raoult D. Ticks and tickborne bacterial diseases in humans: an emerging infectious threat. Clin Infect Dis. 2001;32:897–928.
179. Stafford KC, Station CAE. Tick management handbook: an integrated guide for homeowners, pest control operators, and public health officials for the prevention of tick-associated disease. New Haven: Connecticut Agricultural Experiment Station; 2007.
180. Seidman D, Hebert KS, Truchan HK, Miller DP, Tegels BK, Marconi RT, et al. Essential domains of *Anaplasma phagocytophilum* invasins utilized to infect mammalian host cells. PLoS Pathog. 2015;11:e1004669.
181. Ojogun N, Kahlon A, Ragland SA, Troese MJ, Mastronunzio JE, Walker NJ, et al. *Anaplasma phagocytophilum* outer membrane protein A interacts with sialylated glycoproteins to promote infection of mammalian host cells. Infect Immun. 2012;80:3748–60.
182. Hebert KS, Seidman D, Oki AT, Izac J, Emani S, Oliver LD Jr, et al. *Anaplasma marginale* outer membrane protein A is an adhesin that recognizes sialylated and fucosylated glycans and functionally depends on an essential binding domain. Infect Immun. 2017;85:e00968-16.
183. Barbet AF, Allred DR. The msp1β multigene family of *Anaplasma marginale*: nucleotide sequence analysis of an expressed copy. Infect Immun. 1991;59:971–6.
184. McGarey DJ, Allred DR. Characterization of hemagglutinating components on the *Anaplasma marginale* initial body surface and identification of possible adhesins. Infect Immun. 1994;62:4587–93.
185. de la Fuente J, Garcia-Garcia JC, Blouin EF, Kocan KM. Differential adhesion of major surface proteins 1a and 1b of the ehrlichial cattle pathogen *Anaplasma marginale* to bovine erythrocytes and tick cells. Int J Parasitol. 2001;31:145–53.
186. McGarey DJ, Barbet AF, Palmer GH, McGuire TC, Allred DR. Putative adhesins of *Anaplasma marginale*: major surface polypeptides 1a and 1b. Infect Immun. 1994;62:4594–601.
187. Contreras M, Alberdi P, Mateos-Hernandez L, Fernandez de Mera IG, Garcia-Perez AL, Vancova M, et al. *Anaplasma phagocytophilum* MSP4 and HSP70 proteins are involved in interactions with host cells during pathogen infection. Front Cell Infect Microbiol. 2017;7:307.
188. NCBI. 1988. https://www.ncbi.nlm.nih.gov/. Accessed 30 Nov 2017.
189. de la Fuente J, Garcia-Garcia JC, Blouin EF, Kocan KM. Characterization of the functional domain of major surface protein 1a involved in adhesion of the rickettsia *Anaplasma marginale* to host cells. Vet Microbiol. 2003;91:265–83.
190. Garcia-Garcia JC, de la Fuente J, Bell-Eunice G, Blouin EF, Kocan KM. Glycosylation of *Anaplasma marginale* major surface protein 1a and its putative role in adhesion to tick cells. Infect Immun. 2004;72:3022–30.
191. Park J, Choi KS, Dumler JS. Major surface protein 2 of *Anaplasma phagocytophilum* facilitates adherence to granulocytes. Infect Immun. 2003;71:4018–25.
192. Rejmanek D, Foley P, Barbet A, Foley J. Antigen variability in *Anaplasma phagocytophilum* during chronic infection of a reservoir host. Microbiology. 2012;158:2632–41.

193. Rejmanek D, Foley P, Barbet A, Foley J. Evolution of antigen variation in the tick-borne pathogen *Anaplasma phagocytophilum*. Mol Biol Evol. 2012;29:391–400.
194. Ge Y, Rikihisa Y. Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. J Bacteriol. 2007;189:7819–28.
195. Mastronunzio JE, Kurscheid S, Fikrig E. Postgenomic analyses reveal development of infectious *Anaplasma phagocytophilum* during transmission from ticks to mice. J Bacteriol. 2012;194:2238–47.
196. Troese MJ, Kahlon A, Ragland SA, Ottens AK, Ojogun N, Nelson KT, et al. Proteomic analysis of *Anaplasma phagocytophilum* during infection of human myeloid cells identifies a protein that is pronouncedly upregulated on the infectious dense-cored cell. Infect Immun. 2011;79:4696–707.
197. Goodman JL, Nelson CM, Klein MB, Hayes SF, Weston BW. Leukocyte infection by the granulocytic ehrlichiosis agent is linked to expression of a selectin ligand. J Clin Invest. 1999;103:407–12.
198. Ojogun N, Barnstein B, Huang B, Oskeritzian CA, Homeister JW, Miller D, et al. *Anaplasma phagocytophilum* infects mast cells *via* α1,3-fucosylated but not sialylated glycans and inhibits IgE-mediated cytokine production and histamine release. Infect Immun. 2011;79:2717–26.
199. Carlyon JA, Akkoyunlu M, Xia L, Yago T, Wang T, Cummings RD, et al. Murine neutrophils require α1,3-fucosylation but not PSGL-1 for productive infection with *Anaplasma phagocytophilum*. Blood. 2003;102:3387–95.
200. McEver RP, Cummings RD. Role of PSGL-1 binding to selectins in leukocyte recruitment. J Clin Invest. 1997;100(Suppl. 11):S97–103.
201. Yago T, Leppanen A, Carlyon JA, Akkoyunlu M, Karmakar S, Fikrig E, et al. Structurally distinct requirements for binding of P-selectin glycoprotein ligand-1 and sialyl Lewis x to *Anaplasma phagocytophilum* and P-selectin. J Biol Chem. 2003;278:37987–97.
202. Herron MJ, Nelson CM, Larson J, Snapp KR, Kansas GS, Goodman JL. Intracellular parasitism by the human granulocytic ehrlichiosis bacterium through the P-selectin ligand, PSGL-1. Science. 2000;288:1653–6.
203. Pedra JH, Narasimhan S, Rendic D, DePonte K, Bell-Sakyi L, Wilson IB, et al. Fucosylation enhances colonization of ticks by *Anaplasma phagocytophilum*. Cell Microbiol. 2010;12:1222–34.
204. Kariu T, Smith A, Yang X, Pal U. A chitin deacetylase-like protein is a predominant constituent of tick peritrophic membrane that influences the persistence of Lyme disease pathogens within the vector. PLoS One. 2013;8:e78376.
205. Abraham NM, Liu L, Jutras BL, Yadav AK, Narasimhan S, Gopalakrishnan V, et al. Pathogen-mediated manipulation of arthropod microbiota to promote infection. Proc Natl Acad Sci USA. 2017;114:E781–E90.
206. Heisig M, Abraham NM, Liu L, Neelakanta G, Mattessich S, Sultana H, et al. Antivirulence properties of an antifreeze protein. Cell Rep. 2014;9:417–24.
207. Neelakanta G, Sultana H, Fish D, Anderson JF, Fikrig E. *Anaplasma phagocytophilum* induces *Ixodes scapularis* ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. J Clin Invest. 2010;120:3179–90.
208. Ip WK, Takahashi K, Ezekowitz RA, Stuart LM. Mannose-binding lectin and innate immunity. Immunol Rev. 2009;230:9–21.
209. Matsushita M. Ficolins: complement-activating lectins involved in innate immunity. J Innate Immun. 2010;2:24–32.
210. Krarup A, Thiel S, Hansen A, Fujita T, Jensenius JC. L-ficolin is a pattern recognition molecule specific for acetyl groups. J Biol Chem. 2004;279:47513–9.
211. Kairies N, Beisel HG, Fuentes-Prior P, Tsuda R, Muta T, Iwanaga S, et al. The 2. 0-A crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems. Proc Natl Acad Sci USA. 2001;98:13519–24.
212. Hanington PC, Zhang SM. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J Innate Immun. 2011;3:17–27.
213. Kopacek P, Hajdusek O, Buresova V. Tick as a model for the study of a primitive complement system. Adv Exp Med Biol. 2012;710:83–93.
214. Kovar V, Kopacek P, Grubhoffer L. Isolation and characterization of Dorin M, a lectin from plasma of the soft tick *Ornithodoros moubata*. Insect Biochem Mol Biol. 2000;30:195–205.
215. Rego ROM, Hajdusek O, Kovar V, Kopacek P, Grubhoffer L, Hypsa V. Molecular cloning and comparative analysis of fibrinogen-related proteins from the soft tick *Ornithodoros moubata* and the hard tick *Ixodes ricinus*. Insect Biochem Mol Biol. 2005;35:991–1004.
216. Rego ROM, Kovar V, Kopacek P, Weise C, Man P, Sauman I, et al. The tick plasma lectin, Dorin M, is a fibrinogen-related molecule. Insect Biochem Mol Biol. 2006;36:291–9.
217. Man P, Kovar V, Sterba J, Strohalm M, Kavan D, Kopacek P, et al. Deciphering Dorin M glycosylation by mass spectrometry. Eur J Mass Spectrom. 2008;14:345–54.
218. Sterba J, Dupejova J, Fiser M, Vancova M, Grubhoffer L. Fibrinogen-related proteins in ixodid ticks. Parasit Vectors. 2011;4:127.
219. Honig Mondekova H, Sima R, Urbanova V, Kovar V, Rego ROM, Grubhoffer L, et al. Characterization of *Ixodes ricinus* fibrinogen-related proteins (Ixoderins) discloses their function in the tick innate immunity. Front Cell Infect Microbiol. 2017;7:509.
220. Dupejova J, Sterba J, Vancova M, Grubhoffer L. Hemelipoglycoprotein from the ornate sheep tick, *Dermacentor marginatus*: structural and functional characterization. Parasit Vectors. 2011;4:4.
221. Huang X, Tsuji N, Miyoshi T, Nakamura-Tsuruta S, Hirabayashi J, Fujisaki K. Molecular characterization and oligosaccharide-binding properties of a galectin from the argasid tick *Ornithodoros moubata*. Glycobiology. 2007;17:313–23.
222. Vasta GR, Ahmed H, Nita-Lazar M, Banerjee A, Pasek M, Shridhar S, et al. Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. Front Immunol. 2012;3:199.
223. Maeda H, Miyata T, Kusakisako K, Galay RL, Talactac MR, Umemiya-Shirafuji R, et al. A novel C-type lectin with triple carbohydrate recognition domains has critical roles for the hard tick *Haemaphysalis longicornis* against Gram-negative bacteria. Dev Comp Immunol. 2016;57:38–47.
224. Smith AA, Pal U. Immunity-related genes in *Ixodes scapularis* - perspectives from genome information. Front Cell Infect Microbiol. 2014;4:116.
225. Pang X, Xiao X, Liu Y, Zhang R, Liu J, Liu Q, et al. Mosquito C-type lectins maintain gut microbiome homeostasis. Nat Microbiol. 2016;1:16023.
226. Cheng G, Cox J, Wang P, Krishnan MN, Dai J, Qian F, et al. A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes. Cell. 2010;142:714–25.
227. Neelakanta G, Sultana H. Viral receptors of the gut: vector-borne viruses of medical importance. Curr Opin Insect Sci. 2016;16:44–50.
228. Alarcon-Chaidez F, Ryan R, Wikel S, Dardick K, Lawler C, Foppa IM, et al. Confirmation of tick bite by detection of antibody to *Ixodes* calreticulin salivary protein. Clin Vaccine Immunol. 2006;13:1217–22.
229. Schroeder H, Skelly PJ, Zipfel PF, Losson B, Vanderplasschen A. Subversion of complement by hematophagous parasites. Dev Comp Immunol. 2009;33:5–13.
230. Eggleton P, Lieu TS, Zappi EG, Sastry K, Coburn J, Zaner KS, et al. Calreticulin is released from activated neutrophils and binds to C1q and mannan-binding protein. Clin Immunol Immunopathol. 1994;72:405–9.
231. Kim TK, Ibelli AM, Mulenga A. *Amblyomma americanum* tick calreticulin binds C1q but does not inhibit activation of the classical complement cascade. Ticks Tick Borne Dis. 2015;6:91–101.
232. Uhlir J, Grubhoffer L, Volf P. Novel agglutinin in the midgut of the tick *Ixodes ricinus*. Folia Parasitol (Praha). 1996;43:233–9.
233. Grubhoffer L, Kovar V, Rudenko N. Tick lectins: structural and functional properties. Parasitology. 2004;129:S113–S25.
234. Kamwendo SP, Ingram GA, Musisi FL, Molyneux DH. Haemagglutinin activity in tick (*Rhipicephalus appendiculatus*) haemolymph and extracts of gut and salivary gland. Ann Trop Med Parasitol. 1993;87:303–5.
235. Kamwendo SP, Ingram GA, Musisi FL, Trees AJ, Molyneux DH. Characteristics of tick, *Rhipicephalus appendiculatus*, glands distinguished by surface lectin binding. Ann Trop Med Parasitol. 1993;87:525–35.
236. Pal U, Li X, Wang T, Montgomery RR, Ramamoorthi N, Desilva AM, et al. TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. Cell. 2004;119:457–68.
237. Grubhoffer L, Dusbabek F. Lectin binding analysis of *Argas polonicus* tissue glycoproteins. Vet Parasitol. 1991;38:235–47.
238. Vancova M, Zacharovova K, Grubhoffer L, Nebesarova J. Ultrastructure and lectin characterization of granular salivary cells from *Ixodes ricinus* females. J Parasitol. 2006;92:431–40.
239. Grubhoffer L, Hajdusek O, Vancova M, Sterba J, Rudenko N. Glycobiochemistry of ticks, vectors of infectious diseases: carbohydrate-binding proteins and glycans. FEBS J. 2009;276:141.
240. Vancova M, Sterba J, Dupejova J, Simonova Z, Nebesarova J, Novotny MV, et al. Uptake and incorporation of sialic acid by the tick *Ixodes ricinus*. J Insect Physiol. 2012;58:1277–87.
241. Thall A, Galili U. Distribution of Galα1-3Galβ1-4GlcNAc residues on secreted mammalian glycoproteins (thyroglobulin, fibrinogen, and immunoglobulin G) as measured by a sensitive solid-phase radioimmunoassay. Biochemistry. 1990;29:3959–65.
242. Chung CH, Mirakhur B, Chan E, Le QT, Berlin J, Morse M, et al. Cetuximab-induced anaphylaxis and IgE specific for galactose-α-1,3-galactose. N Engl J Med. 2008;358:1109–17.

243. Commins SP, James HR, Kelly LA, Pochan SL, Workman LJ, Perzanowski MS, et al. The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose-α-1,3-galactose. J Allergy Clin Immunol. 2011;127:1286–93.e6.
244. Van Nunen SA, O'Connor KS, Clarke LR, Boyle RX, Fernando SL. An association between tick bite reactions and red meat allergy in humans. Med J Aust. 2009;190:510–1.
245. Chinuki Y, Ishiwata K, Yamaji K, Takahashi H, Morita E. *Haemaphysalis longicornis* tick bites are a possible cause of red meat allergy in Japan. Allergy. 2016;71:421–5.
246. Hamsten C, Starkhammar M, Tran TA, Johansson M, Bengtsson U, Ahlen G, et al. Identification of galactose-α-1,3-galactose in the gastrointestinal tract of the tick *Ixodes ricinus*; possible relationship with red meat allergy. Allergy. 2013;68:549–52.
247. Wang H, Nuttall PA. Excretion of host immunoglobulin in tick saliva and detection of IgG-binding proteins in tick haemolymph and salivary glands. Parasitology. 1994;109:525–30.
248. Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Mather TN, Ribeiro JMC. Exploring the sialome of the tick *Ixodes scapularis*. J Exp Biol. 2002;205:2843–64.
249. Araujo RN, Franco PF, Rodrigues H, Santos LCB, McKay CS, Sanhueza CA, et al. *Amblyomma sculptum* tick saliva: α-Gal identification, antibody response and possible association with red meat allergy in Brazil. Int J Parasitol. 2016;46:213–20.
250. Altmann F. The role of protein glycosylation in allergy. Int Arch Allergy Immunol. 2007;142:99–115.
251. van Die I, Gomord V, Kooyman FN, van den Berg TK, Cummings RD, Vervelde L. Core α1->3-fucose is a common modification of *N*-glycans in parasitic helminths and constitutes an important epitope for IgE from *Haemonchus contortus* infected sheep. FEBS Lett. 1999;463:189–93.
252. Altmann F. Coping with cross-reactive carbohydrate determinants in allergy diagnosis. Allergo J Int. 2016;25:98–105.
253. Vancova M, Nebesarova J. Correlative fluorescence and scanning electron microscopy of labelled core fucosylated glycans using cryosections mounted on carbon-patterned glass slides. PLoS One. 2015;10:e0145034.
254. Bencurova M, Hemmer W, Focke-Tejkl M, Wilson IB, Altmann F. Specificity of IgG and IgE antibodies against plant and insect glycoprotein glycans determined with artificial glycoforms of human transferrin. Glycobiology. 2004;14:457–66.
255. North SJ, Koles K, Hembd C, Morris HR, Dell A, Panin VM, et al. Glycomic studies of *Drosophila melanogaster* embryos. Glycoconjugate J. 2006;23:345–54.
256. Koles K, Irvine KD, Panin VM. Functional characterization of *Drosophila* sialyltransferase. J Biol Chem. 2004;279:4346–57.
257. Repnikova E, Koles K, Nakamura M, Pitts J, Li H, Ambavane A, et al. Sialyltransferase regulates nervous system function in *Drosophila*. J Neurosci. 2010;30:6466–76.
258. Sterba J, Vancova M, Sterbova J, Bell-Sakyi L, Grubhoffer L. The majority of sialylated glycoproteins in adult *Ixodes ricinus* ticks originate in the host, not the tick. Carbohydr Res. 2014;389:93–9.
259. Gritsun TS, Lashkevich VA, Gould EA. Tick-borne encephalitis. Antiviral Res. 2003;57:129–46.
260. Welsch S, Miller S, Romero-Brey I, Merz A, Bleck CK, Walther P, et al. Composition and three-dimensional architecture of the dengue virus replication and assembly sites. Cell Host Microbe. 2009;5:365–75.
261. Gillespie LK, Hoenen A, Morgan G, Mackenzie JM. The endoplasmic reticulum provides the membrane platform for biogenesis of the flavivirus replication complex. J Virol. 2010;84:10438–47.
262. Offerdahl DK, Dorward DW, Hansen BT, Bloom ME. A three-dimensional comparison of tick-borne flavivirus infection in mammalian and tick cell lines. PLoS One. 2012;7:e47912.
263. Miorin L, Romero-Brey I, Maiuri P, Hoppe S, Krijnse-Locker J, Bartenschlager R, et al. Three-dimensional architecture of tick-borne encephalitis virus replication sites and trafficking of the replicated RNA. J Virol. 2013;87:6469–81.
264. Yu L, Takeda K, Gao Y. Characterization of virus-specific vesicles assembled by West Nile virus non-structural proteins. Virology. 2017;506:130–40.
265. Best SM, Morris KL, Shannon JG, Robertson SJ, Mitzel DN, Park GS, et al. Inhibition of interferon-stimulated JAK-STAT signaling by a tick-borne flavivirus and identification of NS5 as an interferon antagonist. J Virol. 2005;79:12828–39.
266. Lindenbach BD, Rice CM. Molecular biology of flaviviruses. Adv Virus Res. 2003;59:23–61.
267. Lorenz IC, Allison SL, Heinz FX, Helenius A. Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. J Virol. 2002;76:5480–91.
268. Mackenzie JM, Westaway EG. Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. J Virol. 2001;75:10787–99.
269. Stadler K, Allison SL, Schalich J, Heinz FX. Proteolytic activation of tick-borne encephalitis virus by furin. J Virol. 1997;71:8475–81.
270. Mandl CW. Steps of the tick-borne encephalitis virus replication cycle that affect neuropathogenesis. Virus Res. 2005;111:161–74.
271. Heinz FX, Allison SL. Flavivirus structure and membrane fusion. Adv Virus Res. 2003;59:63–97.
272. Hammond C, Braakman I, Helenius A. Role of *N*-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. Proc Natl Acad Sci USA. 1994;91:913–7.
273. Yoshii K, Yanagihara N, Ishizuka M, Sakai M, Kariwa H. *N*-linked glycan in tick-borne encephalitis virus envelope protein affects viral secretion in mammalian cells, but not in tick cells. J Gen Virol. 2013;94:2249–58.
274. Winkler G, Heinz FX, Kunz C. Studies on the glycosylation of flavivirus E proteins and the role of carbohydrate in antigenic structure. Virology. 1987;159:237–43.
275. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature. 1995;375:291–8.
276. Lorenz IC, Kartenbeck J, Mezzacasa A, Allison SL, Heinz FX, Helenius A. Intracellular assembly and secretion of recombinant subviral particles from tick-borne encephalitis virus. J Virol. 2003;77:4370–82.
277. Grubhoffer L, Guirakhoo F, Heinz F, Kunz C. Interaction of tick-borne encephalitis virus protein E with labelled lectins. In: Lectins: Biology, Biochemistry and Clinical Biochemistry, vol. 7. St Louis: Sigma Chemical Company; 1990. p. 313–9.
278. Fuzik T, Formanova P, Ruzek D, Yoshii K, Niedrig M, Plevka P. Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody. Nat Commun. 2018;9:436.
279. Goto A, Yoshii K, Obara M, Ueki T, Mizutani T, Kariwa H, et al. Role of the *N*-linked glycans of the prM and E envelope proteins in tick-borne encephalitis virus particle secretion. Vaccine. 2005;23:3043–52.
280. Putnak JR, Charles PC, Padmanabhan R, Irie K, Hoke CH, Burke DS. Functional and antigenic domains of the dengue-2 virus nonstructural glycoprotein NS-1. Virology. 1988;163:93–103.
281. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus genome organization, expression, and replication. Annu Rev Microbiol. 1990;44:649–88.
282. Heinz FX, Allison SL. Structures and mechanisms in flavivirus fusion. Adv Virus Res. 2000;55:231–69.
283. Hanna SL, Pierson TC, Sanchez MD, Ahmed AA, Murtadha MM, Doms RW. *N*-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. J Virol. 2005;79:13262–74.
284. Bryant JE, Calvert AE, Mesesan K, Crabtree MB, Volpe KE, Silengo S, et al. Glycosylation of the dengue 2 virus E protein at N67 is critical for virus growth *in vitro* but not for growth in intrathoracically inoculated *Aedes aegypti* mosquitoes. Virology. 2007;366:415–23.
285. Mondotte JA, Lozach PY, Amara A, Gamarnik AV. Essential role of dengue virus envelope protein *N*-glycosylation at asparagine-67 during viral propagation. J Virol. 2007;81:7136–48.
286. Moudy RM, Zhang B, Shi PY, Kramer LD. West Nile virus envelope protein glycosylation is required for efficient viral transmission by *Culex* vectors. Virology. 2009;387:222–8.
287. Murata R, Eshita Y, Maeda A, Maeda J, Akita S, Tanaka T, et al. Glycosylation of the West Nile virus envelope protein increases *in vivo* and *in vitro* viral multiplication in birds. Am J Trop Med Hyg. 2010;82:696–704.
288. Hyde JA. *Borrelia burgdorferi* keeps moving and carries on: a review of borrelial dissemination and invasion. Front Immunol. 2017;8:114.
289. Hooper LV, Gordon JI. Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology. 2001;11:1R–10R.
290. Nizet V, Esko JD. Bacterial and viral infections. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor: Cold Spring Harbour Laboratory Press; 2009.
291. de la Fuente J, Canales M, Kocan KM. The importance of protein glycosylation in development of novel tick vaccine strategies. Parasite Immunol. 2006;28:687–8.

292. Sprong H, Trentelman J, Seemann I, Grubhoffer L, Rego RO, Hajdusek O, et al. ANTIDotE: anti-tick vaccines to prevent tick-borne diseases in Europe. Parasit Vectors. 2014;7:77.
293. Boysen A, Palmisano G, Krogh TJ, Duggin IG, Larsen MR, Moller-Jensen J. A novel mass spectrometric strategy "BEMAP" reveals extensive *O*-linked protein glycosylation in enterotoxigenic *Escherichia coli*. Sci Rep. 2016;6:32016.
294. Parveen N, Leong JM. Identification of a candidate glycosaminoglycan-binding adhesin of the Lyme disease spirochete *Borrelia burgdorferi*. Mol Microbiol. 2000;35:1220–34.
295. Guo BP, Brown EL, Dorward DW, Rosenberg LC, Hook M. Decorin-binding adhesins from *Borrelia burgdorferi*. Mol Microbiol. 1998;30:711–23.
296. Fuchs H, Wallich R, Simon MM, Kramer MD. The outer surface protein A of the spirochete *Borrelia burgdorferi* is a plasmin(ogen) receptor. Proc Natl Acad Sci USA. 1994;91:12594–8.
297. Lagal V, Portnoi D, Faure G, Postic D, Baranton G. *Borrelia burgdorferi sensu stricto* invasiveness is correlated with OspC-plasminogen affinity. Microbes Infect. 2006;8:645–52.
298. Onder O, Humphrey PT, McOmber B, Korobova F, Francella N, Greenbaum DC, et al. OspC is potent plasminogen receptor on surface of *Borrelia burgdorferi*. J Biol Chem. 2012;2870:16860–8.
299. Floden AM, Watt JA, Brissette CA. *Borrelia burgdorferi* enolase is a surface-exposed plasminogen binding protein. PLoS One. 2011;6:e27502.
300. Cinco M, Cini B, Murgia R, Presani G, Prodan M, Perticarari S. Evidence of involvement of the mannose receptor in adhesion of *Borrelia burgdorferi* to monocyte/macrophages. Infect Immun. 2001;69:2743–7.
301. Guo BP, Teneberg S, Munch R, Terunuma D, Hatano K, Matsuoka K, et al. Relapsing fever *Borrelia* binds to neolacto glycans and mediates rosetting of human erythrocytes. Proc Natl Acad Sci USA. 2009;106:19280–5.
302. Kuismanen E, Hedman K, Saraste J, Pettersson RF. Uukuniemi virus maturation: accumulation of virus particles and viral antigens in the Golgi complex. Mol Cell Biol. 1982;2:1444–58.

Kotsarenko, K., Vechtova, P., Hammerova, Z., Langova, N., Malinovska, L., Wimmerova, M., Sterba, J., Grubhoffer, L., 2019. Newly identified DNA methyltransferases of Ixodes ricinus ticks. *Ticks and Tick Borne Diseases*. 11(2):101348. doi: 10.1016/j.ttbdis.2019.101348.

Ticks and Tick-borne Diseases xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis

Original article

Newly identified DNA methyltransferases of *Ixodes ricinus* ticks

Kateryna Kotsarenko[a,b,*], Pavlina Vechtova[a,b], Zuzana Hammerova[c], Natalia Langova[c], Lenka Malinovska[c,d], Michaela Wimmerova[c,d,e], Jan Sterba[a,b], Libor Grubhoffer[a,b]

[a] Faculty of Science, University of South Bohemia, Branisovska 1760, 37005, Ceske Budejovice, Czech Republic
[b] Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Branisovska 31, 37005, Ceske Budejovice, Czech Republic
[c] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlarska 2, Brno, 61137, Czech Republic
[d] Central European Institute of Technology, Masaryk University, Kamenice 5, Brno, 62500, Czech Republic
[e] Department of Biochemistry, Faculty of Science, Masaryk University, Kotlarska 2, Brno, 61137, Czech Republic

ARTICLE INFO

Keywords:
DNA methylation
Methyltransferase
5-Methylcytosine
N6-Methyladenine
Epigenetics
Ixodes tick

ABSTRACT

DNA methylation at the fifth position of cytosine (5mC) and at the sixth position of adenine (6 mA) plays an important role in the regulation of the gene expression and, in eukaryotes, is essential for normal development. For *Ixodes ricinus*, the most common European arthropod vector of human and animal pathogens, the DNA methylation profile and the role of DNA methylation in tick development are still under discussion. Our goal was to analyze the status of *I. ricinus* DNA methylation at different life stages and identify enzymes that produce this type of DNA modification. We found that 5mC and 6mA are present in *I. ricinus* genomic DNA at all life stages. In the transcriptome of *I. ricinus*, we identified the sequences of the putative IrDNMT1, IrDNMT3, and IrDAMT enzymes, and bioinformatic analysis and three-dimensional modeling predicted their DNA methylation activity. This confirms that *I. ricinus* possesses a complete DNA methylation toolkit. Our results suggest that DNA methylation is important for the physiology and transstadial development of ticks.

1. Introduction

DNA methylation represents a fundamental epigenetic modification that regulates chromatin architecture and gene transcription in eukaryotes (Callebaut et al., 1999; Messerschmidt et al., 2014; Yarychkivska et al., 2018). In last decade, the arthropod epigenetics has become a fast-developing field of science and many recent studies have focused on the localization of DNA methylation in arthropod genomes and understanding of its functional significance (Glastad et al., 2014; Bewick et al., 2017; Hunt et al., 2013).

In contrast to the heavily methylated mammalian genomes, invertebrate genomes are only sparsely methylated showing rather 'mosaic' distribution of active transcription-related methylation across the coding sequences (Wedd and Maleszka, 2016). The role of epigenetic modifications in embryogenesis and behavior of insects has already been presented in several publications (Li-Byarlay, 2016; Holman et al., 2016; Wang et al., 2013; Zhang et al., 2015b; Yan et al., 2015). DNA methylation has been described also in some crustaceans and arachnids, suggesting the importance of epigenetic mechanisms for regulation of gene expression in arthropods in general (Hunt et al., 2013; Lyko and Maleszka, 2011; Wedd and Maleszka, 2016).

The description of arthropod epigenome, however, has not covered an important group of the blood-feeding ectoparasitic group of arthropods, the ticks. In spite of the global medical and veterinary importance of *Ixodes* tick species, relatively little is known about their genome organization and epigenetic modification. For example, Cabezas-Cruz et al. (2016) identified and characterized 5 histones and 34 histone-modifying enzymes in the tick vector *I. scapularis* (Cabezas-Cruz et al., 2016).

Glastad et al. (2013) showed that correlations between CpG observed/expected values in termite libraries and the gene set IscaW1.2 (http://www.vectorbase.org) of the blacklegged tick *I. scapularis* were

Abbreviations: 5mC, 5-Methylcytosine; 6mA, N6-Methyladenine; BAH domain, bromo adjacent homology domain; CXXC zinc finger domain, Cytosine – X - X-cytosine zinc finger domain; DAMT, DNA adenine methyltransferase; DCM domain, DNA cytosine methylase domain; DNMT, DNA (cytosine-5) methyltransferase; DPPF motif, Asp-Pro-Pro-Phe motif; SAM, S-Adenosyl methionine; SAH, S-Adenosylhomocysteine; METTL protein, methyltransferase-like protein; MT-A70 domain, methyltransferase-A70 domain; PWWP motif, Pro-Trp-Trp-Pro motif

* Corresponding author at: Faculty of Science, University of South Bohemia, Branisovska 1760, 37005, Ceske Budejovice, Czech Republic.
E-mail addresses: kkotsarenko@prf.jcu.cz (K. Kotsarenko), vechtova@prf.jcu.cz (P. Vechtova), 393906@mail.muni.cz (Z. Hammerova), 451140@mail.muni.cz (N. Langova), malinovska@mail.muni.cz (L. Malinovska), michaw@chemi.muni.cz (M. Wimmerova), sterbaj@prf.jcu.cz (J. Sterba), liborex@paru.cas.cz (L. Grubhoffer).

https://doi.org/10.1016/j.ttbdis.2019.101348
Received 9 May 2019; Received in revised form 5 November 2019; Accepted 28 November 2019

nearly as high as those in other invertebrates with DNA methylation (*Acyrthosiphon pisum, Apis mellifera*, and *Ciona intestinalis*) (Glastad et al., 2013). Thus, it was suggested that DNA methylation is also present in *I. scapularis*. De Meeûs et al. (2004) specifically methylated cytosine in CpG island using the M.SssI methyltransferase. Methylation was confirmed with antibodies against 5mC using Dot blot assay in both *I. ricinus* males and females. These results indicated the presence of CpG islands which could be methylated in the DNA of *Ixodes* ticks (de Meeûs et al., 2004). Meyer et al. (2010) showed the presence of digestion resistant DNA in the *I. scapularis* genome using the methylation-sensitive restriction enzyme HpaII. Long stretches (> 20 kb) of gDNA containing tandem repeats were resistant to the HpaII treatment (Meyer et al., 2010). Furthermore, Zhang et al. (2015b) predicted the presence of the DNMT1 and DNMT3 methylation enzymes in *I. scapularis* ticks using a bioinformatic approach (Zhang et al., 2015b).

Based on these data we hypothesize that *Ixodes* ticks similarly to other eukaryotes use epigenetic mechanisms for regulation of gene expression. To confirm this hypothesis, we searched for the transcripts of DNA methyltransferases in the *I. ricinus* transcriptome, estimated their relative expression in different life stages, and verified the presence of methylated 5mC and 6 mA in the tick genomic DNA.

2. Materials and methods

2.1. I. ricinus transcriptome assembly

Total RNA was isolated from the egg, larvae, partially fed nymph, adult female, and partially fed female of *I. ricinus* using NucleoSpin RNA II (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. The concentration of RNA was measured using NanoPhotometer Pearl (Implen, Munchen, Germany) and the quality of RNA was determined using 2100 Bioanalyzer with RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA, USA). cDNA and sequencing library construction and sequencing of RNA samples were done by the GeneCore facility of EMBL in Heidelberg. cDNA synthesis and sequencing library construction were carried out using TruSeq DNA Sample Prep kit v2 (PE50 reads) (Illumina, San Diego, CA, USA). Sequencing was performed in HiSeq 2000 sequencing platform. All raw reads were trimmed off sequencing adapters and short reads using Trimmomatic (Bolger et al., 2014). Trimmed reads were assembled using Trinity assembler v2.0.6 (Haas et al., 2013).

GenBank database was used to search the homologs of DNMT1, DNMT3, and DAMT in *Ixodes* ticks. Output sequences of the blast search were used to identify homologous sequences in the *I. ricinus* transcriptome in each life stage using tblastn program v2.2.31+ within blast + command-line application with default parameters and tabular output format (-outfmt 7 switch).

2.2. Bioinformatic analyses

The conserved domains Dcm, PWWP and MT-A70 of *I. ricinus* putative DNA methyltransferases were downloaded from the CDD database (Marchler-Bauer et al., 2011) and used to search for methyltransferase sequences of arthropod species available in the GenBank database (downloaded 19 October 2018). The phylogenetic relationship of *I. ricinus* methyltransferase and other known arthropod DNMT1, DNMT3, and DAMT enzymes were constructed. The alignment of the protein sequences was performed using Muscle (Chojnacki et al., 2017) with default parameters. The phylogenetic reconstruction was performed using maximum likelihood estimation in IQTREE, v. 1.6.6 using MFT parameter of -m option which tests the submitted data for the best fit model (Kalyaanamoorthy et al., 2017) and UFBoot parameter of –bb option which performs ultrafast bootstrap approximation of the computed branches (Hoang et al., 2018). The final tree was visualized and adjusted in Interactive Tree of Life, v 4.2.3 (Letunic and Bork, 2016).

The ProtParam program (Gasteiger et al., 2005) was used to describe the physicochemical properties of the corresponding proteins and program WoLF PSORT (Horton et al., 2007) was used to identify their intracellular localization. DiANNA (Ferrè and Clote, 2005) was used to predict the presence of disulfide bridges. I-TASSER (Yang and Zhang, 2015; Roy et al., 2010; Zhang et al., 2017) and COACH (Yang et al., 2012; Yang et al., 2013) programs were used to predict enzyme activity and the binding sites. I-TASSER was also used to predict the secondary and tertiary structures. We modeled tertiary structures of the putative IrDNMT1, DNMT3, and DAMT (whole sequences) and also parts of sequences corresponding to conserved domains DNA cytosine methylase (Dcm), Pro-Trp-Trp-Pro (PWWP) and Methyltransferase-A70 (MT-A70) for the putative IrDNMT1, DNMT3, and DAMT, respectively. The resulting structures were displayed using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

2.3. qRT-PCR

RNA samples from different life stages of *I. ricinus* (freshly laid egg, larvae, partially fed nymph, adult female, and partially fed female) were isolated using the NucleoSpin® RNA kit (Macherey-Nagel, Düren, Germany) and treated with dsDNase (ThermoFisher Scientific, Waltham, MA, USA). The synthesis of cDNA was done using ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs, Ipswich, MA, USA). qRT-PCR reaction was performed using the qPCR 2x SYBR Master Mix (Top-Bio, Vestec, Czech Republic) in LightCycler ®480 Real-Time PCR System (Roche, Mannheim, Germany). The primers used in the experiments are listed in Table 1.

The resulting Ct values were recorded and processed with LightCycler 480 Software release 1.5.0 SP4 (Roche, Mannheim, Germany). The Ct values of IrDNMT1, IrDNMT3 and IrDAMT transcripts were normalized by the normalization factor calculated using 6 tick housekeeping genes, *EF1α*, *ferritin*, *β-tubulin*, *rp49*, *rps4*, and *ppia* as suggested previously (Nijhof et al., 2009; Koci et al., 2013) using the formula, presented by (Schmittgen and Livak, 2008).

2– (Ct gene of interest – Ct internal control),

Where Ct is the threshold cycle and internal control means the housekeeping genes.

2.4. Preparation of DNA samples

Genomic DNA samples were isolated from different life stages of *I. ricinus* (freshly laid eggs, larvae, partially fed nymphs, adult non-fed females, and partially fed females) using NucleoSpin® Tissue kit (Macherey-Nagel, Düren, Germany) and treated with RNase A according to the NucleoSpin® Tissue protocol. The guts were removed from *I. ricinus* nymphs, adult females, and partially fed females prior to isolation of DNA to avoid the contamination with host DNA (and thus false-positive results in the detection of DNA methylation).

Genomic DNA isolated from onion (*Allium cepa*) was used as a positive control (highly methylated CpG islands were previously described for the onion genome (Suzuki et al., 2010) and synthetic amplicons – as the negative control. Isolation of genomic DNA from onion was performed according to slightly modified protocol of (Dellaporta et al., 1983). In brief, 20 *g* of homogenized onion tissue was mixed with 10 ml

Table 1
Oligonucleotide primer sequences used in the study.

Primers	Sequence 5'→3'	Product length
DNMT1_F8	CGCAGCACAAGGTGACGC	140 bp
DNMT1_R9	GGCAGACGGGTCCTCC	
DNMT3_F2	CACAAGCACCACAAGCACAAGG	222 bp
DNMT3_R7	GCTCGCCAGCAGGTCG	
DAMT_F1	GGGTGGTTTGTATCTCATGAT	138 bp
DAMT_R1	CGCAGAATCCATTTGAAAAGG	

Table 2
List of hits from blast search of *Ixodes* nr database.

Ixodes ticks protein accession number	*I.ricinus* transcript accession number	Translated protein	Identity	Query cover	E-value
JAB78989.1	MH795939	Putative *I. ricinus* adenine-methyltransferase 1	30 %	10 %	4.1
JAB78989.1	MH795940	Putative *I. ricinus* adenine-methyltransferase 3	26 %	96 %	1e-11
JAB78989.1	MH926033	Putative *I. ricinus* adenine-methyltransferase 4	99 %	100 %	8e-137
EEC00968.1	MH795941	Putative *I. ricinus* cytosine-methyltransferase 1	37 %	46 %	5e-59
EEC00969.1	MH795942	Putative *I. ricinus* cytosine-methyltransferase 2 isoform 1	45 %	90 %	9e-49
EEC00969.1	MH795943	Putative *I. ricinus* cytosine-methyltransferase 2 isoform 2	45 %	90 %	8e-49
EEC00969.1	MH795944	Putative *I. ricinus* cytosine-methyltransferase 3	54 %	81 %	8e-72
EEC00969.1	MH795945	Putative *I. ricinus* cytosine-methyltransferase 5 isoform 1	99 %	98 %	6e-168
EEC00969.1	MH795946	Putative *I. ricinus* cytosine-methyltransferase 5 isoform 2	99 %	66 %	9e-70
EEC00968.1	MH795948	Putative *I. ricinus* cytosine-methyltransferase 8	94 %	55 %	0.0
JAB81312.1	MH926034	Putative *I. ricinus* cytosine-methyltransferase 10	100 %	100 %	0.0
EEC19658.1	MH795950	Putative *I. ricinus* cytosine-methyltransferase 11 isoform 1	97 %	96 %	0.0
EEC19658.1	MH795951	Putative *I. ricinus* cytosine-methyltransferase 11 isoform 2	97 %	96 %	0.0

extraction buffer with the following incubation at 65 °C for 45 min. 5 M potassium acetate (3 ml) was added to the mixture and incubated for 20 min on ice and centrifuged for 15 min at 6000 × g. The supernatant was transferred into a fresh tube and mixed with 1 vol of 100 % isopropanol. The mixture was incubated for 30 min at −20 °C. Precipitated DNA was transferred into a fresh tube with 1 ml of 70 % ethanol, gently mixed, and centrifuged at 13 000 x g for 5 min. The pellet was dried and dissolved in PCR-grade water. gDNA was incubated with RnaseA (Carl Roth, Karlsruhe, Germany) at 37 °C for 1 h to digest RNA and avoid positive signals caused by methylation of RNA. After the treatment, the gDNA sample was purified with ethanol precipitation and dissolved in NucleoSpin® Tissue elution buffer (Macherey-Nagel, Düren, Germany).

Mixture of amplicons of firefly luciferase (1653 bp) and internal transcribed spacer (1200 bp) from *Ixodes* ticks were used as a negative control of DNA methylation. Purification of PCR products was done using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany).

All DNA samples were diluted in the elution buffer to a concentration of 100 ng/µl. Concentration was measured using NanoPhotometer Pearl (Implen, Munchen, Germany).

2.5. Dot blot assay

Genomic DNA samples (5 ul of DNA) from ticks and onion (positive control), and synthetic amplicons (negative control) were spotted on a pre-wet nitrocellulose membrane (10600003, Amersham, Little Chalfont, Great Britain), allowed to air-dry, crosslinked in a UV crosslinker. The membranes were blocked in a blocking buffer (5 % skimmed milk in PBS) for 2 h at RT. To detect adenine methylation, the membrane was incubated overnight at 4 °C with polyclonal Anti-N6-methyladenosine antibody (1:5000) (202 003, Synaptic Systems, Goettingen, Germany) followed by an incubation with Alkaline Phosphatase-conjugated Anti-Rabbit IgG secondary Antibody (1:1000) (AP-1000, Vector Laboratories, CA, USA). To detect cytosine methylation, the membrane was incubated overnight at 4 °C with Anti-5-Methylcytosine antibody, clone 33D3 (1:1000) (MABE146, Millipore, CA, USA), followed by the incubation with Alkaline Phosphatase-conjugated Anti-Mouse IgG secondary Antibody (1:1000) (AP-2000, Vector Laboratories, CA, USA). Signals were developed with BCIP/NBT development solution. BCIP and NBT reagents were purchased from AppliChem, BioChemica (Darmstadt, Germany). The input DNA samples were directly stained with GelRed Nucleic Acid Stain (Biotrend, Köln, Germany), and observed under G: BOX Chemi XX6 gel documentation system (Syngene, Cambridge, UK).

3. Results and discussion

3.1. Identification of I. ricinus methyltransferases

The GenBank database was used to search for the known sequences of methyltransferases within *I. scapularis* genome (accession ID: ABJB010000000) (Gulia-Nuss et al., 2016) and partial transcriptome (accession ID: GBBN01000001.1) (Egekwu et al., 2014), and *I. ricinus* transcriptomes (accession ID: GBIH00000000.1 (Kotsyfakis et al., 2015); accession ID: GCJO00000000.1 (Cramaro et al., 2015); accession ID: GANP00000000.1, (Schwarz et al., 2014). We found 23 hypothetical proteins encoding putative DNA methyltransferases in *I. ricinus* transcriptomes and *I. scapularis* genome (Table S1). However, BLAST alignment revealed that only 9 of them had similarities to cytosine and adenine DNA methyltransferases from other eukaryote species: accession ID JAP72831.1, JAB78989.1, JAB81312.1, EEC00472.1, EEC00969.1, XP_002433408.1, XP_002411630.1, EEC13877.1, and EEC00968.1. The other 13 proteins had similarity to RNA or protein methylating enzymes.

Potential *Ixodes* DNA methyltransferases were used to search similar transcripts in the *I. ricinus* transcriptome which was produced in our laboratory. In total, we identified 13 transcripts sharing more than 26 % identity with the *Ixodes* putative DNA methyltransferases. We submitted the 13 identified sequences of *I. ricinus* DNA methyltransferases to the GenBank database and the accession numbers are listed in Table 2.

The highest identity and query cover (98–100 %) with the putative *Ixodes* DNA methyltransferases were shown for three *I. ricinus* transcripts: MH926033, MH795945, and MH926034 (Table 2) which share similarity with Methyltransferase-like protein 4, DNA (cytosine-5)-methyltransferase 1, and PWWP domain-containing protein, respectively. Therefore, we named the protein putative *I. ricinus* adenine-methyltransferase 4, translated from transcript MH926033, as putative IrDAMT. Protein putative *I. ricinus* cytosine-methyltransferase 5 isoform 1, translated from transcript MH795945, was named by us as putative IrDNMT1 and protein putative *I. ricinus* cytosine-methyltransferase 10, translated from transcript MH926034, was named as putative IrDNMT3.

3.2. Domain organization of the I. ricinus DNA methyltransferases and their evolutionary relationships with other arthropod species

We identified ORF of the chosen transcripts and performed a conserved domain search using the CDD database v.3.16 of NCBI (Marchler-Bauer et al., 2011) with default parameters and a standard display result mode. We found that putative IrDNMT1 contains Bromo Adjacent Homology (BAH) domains (intervals 102–229 bp and 333–450 bp) and DNA cytosine methylase (Dcm) domain (interval 490–947 bp) (Fig. 1).

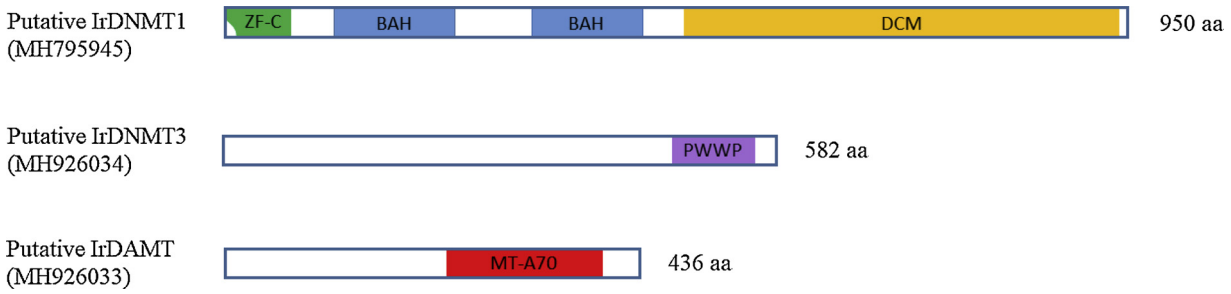


Fig. 1. Schematic presentation of conserved domains identified for the putative DNA methyltransferases, obtained from *I. ricinus* transcriptome. Putative IrDNMT1: green - CXXC zinc finger domain, blue – Bromo adjacent homology (BAH) domain, orange – DCM domain; putative IrDNMT3: purple - PWWP domain; putative IrDAMT: red – MT-A70 domain.

BAH domain plays an important role in linking DNA methylation, replication, and transcriptional regulation (Yarychkivska et al., 2018; Callebaut et al., 1999) and Dcm domain targets CCAGG and CCTGG sites to methylate cytosine at the C5 position (Marinus and Løbner-Olesen, 2014). In addition, the N-terminal part of the sequence starts with a part of the Zinc finger domain (Zf-CXXC), which indicates that the sequence of transcript MH795945 is probably incomplete. According to Pfam database, which contains 147 sequences of DNA (cytosine-5)-methyltransferases, Zf-CXXC, 2x BAH and Dcm domain organization is typical only for cytosine-specific DNA methyltransferases (El-Gebali et al., 2018). This supports our assumption that the putative IrDNMT1 possesses functions of a DNA-methylating enzyme.

Another protein - putative IrDNMT3, contains the PWWP domain (interval 482–568 bp). The conserved Pro-Trp-Trp-Pro motif within the PWWP domain underlies the establishment of DNA methylation patterns during embryogenesis and gametogenesis (Qiu et al., 2002).

Putative IrDAMT contains the MT-A70 domain with a highly conserved DPPF motif (interval 194–357 bp). We performed BLAST analysis to specify the origin of our putative IrDAMT and found homology of this protein to the enzyme Methyltransferase-like protein 4 (METTL4). METTL4 enzyme, as well as METTL3, contains conserved DPPW motif, however, instead of RNA methylation (Wang et al., 2016), it methylates adenine in DNA. The role of METTL4 in DNA adenine methylation and development control was shown previously in lepidopteran *Bombyx mori* (Wang et al., 2018).

We aimed to show the evolutionary relationships between DNA methyltransferases of the *I. ricinus* and other arthropod species. Therefore, the conserved domains of the putative IrDNMT1, IrDNMT3, and IrDAMT were used to search for their homologs within arthropods which share at least 30 % of sequence identity and can be considered as homologous sequences (Pearson, 2013). Therefore, we chose the homologs of DNMT1 sharing more than 68 % identity with the conserved sequence of the Dcm domain, DNMT3 homologs sharing more than 61 % identity with the PWWP domain and DAMT homologs sharing more than 39 % identity with the MT-A70 domain. The phylogenetic reconstruction of the complete sequences of putative IrDNMT1, IrDNMT3, and IrDAMT, found in *I. ricinus* transcriptome and homologous sequences identified in other arthropods was based on Muscle alignment and computed using maximum-likelihood estimation (Fig. 2).

Species from Bivalvia class were used as outgroups. The phylogenetic reconstruction of both putative IrDNMT1 (Fig. 2A) and putative IrDNMT3 (Fig. 2B) sequences clustered the arthropod species into their corresponding taxonomic classes. The putative IrDNMT1 (transcript MH795945) and IrDNMT3 (transcript MH926034) share the same node with the sequences of the closely related species within Arachnids. A similar evolutionary relationship was calculated for the DAMT arthropod sequences; however, the position of the putative IrDAMT (transcript MH926033) does not assign it the common ancestor with other arachnid species. Conversely, the putative IrDAMT is placed in a sister branch of the cluster of Maxillopoda, Branchiopoda, and Insecta (Fig. 2C).

Interestingly, two phylogenetic trees (Fig. 2A and C) did not show relatives in *I. scapularis* for IrDNMT1 and for IrDAMT. In the case of IrDNMT1, it was caused by the exclusion of *I. scapularis* DNMT1 sequence (EEC00969.1, genome accession ID: ABJB010000000 (Gulia-Nuss et al., 2016) that was found to be shorter than *I. ricinus* protein - 230 amino acids compared to 950 amino acids, respectively. This indicates that the sequence of *I. scapularis* DNMT1 might be incomplete. However, the length of IrDNMT1 correlates with that of DNMT1 proteins from other arthropod species. In the case of IrDAMT, we did not find any *I. scapularis* orthologues in publicly available databases.

3.3. Physical and chemical properties and 3D structure of the putative *I. ricinus* DNA methyltransferases

The relation of the newly identified *I. ricinus* putative proteins to the DNA methyltransferases was also confirmed by the prediction of their physical and chemical properties, and 3D modeling. All these putative proteins were predicted to have a nuclear localization (Table 3), which supports their hypothetical assignment to the methyltransferase group of enzymes.

The instability index, which provides an estimate of the protein stability in a test tube (Gasteiger et al., 2005), was higher than 40 for all three studied proteins. Therefore, all of them were predicted to be unstable. Disulfide bridges predicted in putative IrDNMT1, DNMT3, and DAMT probably help to maintain the shape and stability of the molecule and their substrate specificity. COACH program predicted several binding sites for DNA in all three proteins. Furthermore, SAM-binding sites for the donor of methyl groups (S-adenosyl-L-methionine molecule) were found in the putative IrDNMT1 and DAMT. The lack of SAM-binding site in the putative IrDNMT3 can be caused by the lack of structural information about DNMT3 enzymes or imply that the putative IrDNMT3 may be a partial sequence, although this hypothesis requires further verification.

The I-TASSER program was used to predict the tertiary structure of the putative *I. ricinus* DNA methyltransferases. For each target, I-TASSER simulations generated a large ensemble of structural conformations, called decoys. All decoys were clustered based on the pairwise structure similarity. The confidence of each resulting model was quantitatively measured by the C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [−5, 2], where a C-score of a higher value signifies a model with higher confidence and vice-versa (Wu et al., 2007). The C-scores for the predicted 3D models of putative IrDNMT1, IrDNMT3, and IrDAMT proteins were in this range and reached 1.20, −1.48, and −0.92, respectively. When large multidomain proteins are modeled, some parts of the protein can be modeled at the expense of another part, therefore, we modeled also the 3D structure of conserved domains Dcm, PWWP, and MT-A70 with C-score of 1.64, 0.64, and −0.76, respectively. Predicted model of the putative IrDNMT1 protein had higher

Fig. 2. Phylogenetic reconstruction of the arthropod DNA methyltransferases based on muscle alignment and maximum-likelihood estimation. The listed species are colored according to their classification to particular arthropod taxonomic classes. **A**. Illustration of the phylogenetic relationship of arthropod DNMT1 sequences, the tree was rooted using *Crassostrea virginica* (Mollusca); **B**. Illustration of phylogenetic relationship of arthropod DNMT3 sequences, the tree was rooted using *Mizuhopecten yessoensis* (Mollusca); **C**. Illustration of phylogenetic relationship of arthropod DAMT sequences, the tree was rooted using *Mizuhopecten yessoensis* (Mollusca).

Table 3
Predicted properties of the putative *I. ricinus* DNA methyltransferases.

Predicted Properties	Putative IrDNMT1 (MH795945)	Putative IrDNMT3 (MH926034)	Putative IrDAMT (MH926033)
The Number of Amino Acids	950	581	372
Molecular Weight	106.9 kDa	64.0 kDa	40.9 kDa
Theoretical Isoelectric Point (pI)	8.23	9.53	9.14
Hydrophobic Index	−0.611	−0.625	−0.062
Instability Index	45.91 (unstable)	54.47 (unstable)	55.44 (unstable)
Aliphatic Index	65.66	71.72	89.46
Location in Cell	Nucleus	Nucleus	Nucleus
Disulfide Bridges	5-692, 8–579, 11–32, 27–275, 168–904, 180–946, 246–413, 249–647, 291–501, 461–840, 517–574, 522–830, 544–828, 608–765, 639–836, 707-850	135–571, 138–441, 142–171, 165–474, 324–424	7–286, 26–106, 145–339, 155–233, 237–243
SAM Binding Site	498, 499, 500, 502, 503, 504, 521, 522, 523, 542, 543, 544, 576, 578, 600, 929, 930, 931	–	181, 182, 183, 200, 202, 214, 215, 217, 320, 321, 322, 343, 345, 346, 356, 357
DNA Binding Site	16, 24, 25, 26, 30, 585, 586, 589, 591, 628	478, 483, 484, 485, 501, 529, 531	319, 321, 322, 323, 341, 345

confidence than models of putative IrDNMT3 and DAMT and the models of conserved domains were more reliable than the models of whole sequences.

The I-TASSER program suggested several structural templates for the 3D model of the putative IrDNMT1 (Fig. 3A). The best template was the structure of the DNMT1 methyltransferase from mouse (PDB ID: 3PT6), and the 3D models of IrDNMT1 protein and Dcm domain were built using this structure separately. Also, the following binding sites were predicted for IrDNMT1: DNA, Zn^{2+}, and S-adenosylhomocysteine (SAH), where SAH is the binding site for the donor molecule (SAM) (Zhang and Zheng, 2016).

I-TASSER identified also several structural templates for the 3D model of the putative IrDNMT3 (Fig. 4), one of which was a human PWWP Domain-Containing Protein 2B (PDB ID: 4LD6). The model of the PWWP domain alone was found to be more reliable than IrDNMT3 (the C-score of the domain model was 0.64, compared to -1.48 of the whole sequence model).

Several structural templates were used for IrDAMT and MT-A70 domain modeling (Fig. 5), one of which was the human METTL3-METTL14 complex for N6-adenosine methylation (PDB ID: 5IL0). I-TASSER predicted several different functions for IrDAMT, mainly acyltransferases activity or as a proteins utilizing Zn^{2+} as a cofactor. However, for the model of the MT-A70 domain, the predicted function was the adenine-specific DNA-methyltransferase.

Considering the fact that the sequence of putative IrDNMT1 is likely incomplete, we aimed to predict the structure of the complete sequence of *I. ricinus* DNMT1. We performed the search of *I. ricinus* transcriptome data from NCBI database and found the putative *I. ricinus* BAH domain protein (GenBank: JAC94558.1; mRNA sequence - GenBank: GBIH01000152.1) which shares a 99 % identity and 72 % query coverage in amino acid sequence with putative IrDNMT1 according to the BLAST alignment. Thus, the putative *I. ricinus* BAH domain protein represents also DNA cytosine methyltransferase and is a homolog of the putative IrDNMT1. This sequence is 420 aa longer than the putative IrDNMT1 and is complete at the N-terminus. The putative *I. ricinus* BAH domain protein contains the region that corresponds to the conserved domain DNMT1-RFD (Fig. 6 A), and also three other domains which were identified in the putative IrDNMT1 (Fig. 1): Zf-CXXC, BAH, and Dcm. This domain topology (DNMT1-RFD, Zf-CXXC, 2x BAH and Dcm) is typical for DNMT1 enzymes, according to the Pfam database (El-Gebali et al., 2018). Moreover, the identified *I. ricinus* transcript MH795948 also shared homology with putative *I. ricinus* BAH domain protein and contained the DNMT1-RFD conserved domain. Amino acid alignment revealed that putative *I. ricinus* BAH domain protein represents the complete sequence of *I. ricinus* enzyme DNMT1, and transcripts MH795948 and MH795945 belong to this one gene (Fig. 6A).

We predicted the physicochemical characteristics for the putative *I. ricinus* BAH domain protein (Table S2), among them nuclear localization, several disulfide bridges, and SAM, DNA, and Zn^{2+} binding sites. We also designed the 3D model for the putative *I. ricinus* BAH domain protein to compare it with the structure of the putative IrDNMT1 (Fig. 6A–C). The I-TASSER program predicted several structural templates for the 3D model of putative *I. ricinus* BAH domain protein with C-score = -1.65. The best template was the structure of DNMT1 from mouse (PDB ID: 3PT6) and from human (PDB ID: 4WXX). Furthermore, I-TASSER predicted the DNA methylation function for this protein, too.

3.4. Expression of DNA methyltransferases and DNA methylation profile at different life stages of I. ricinus

The bioinformatic analysis confirmed our assumption about the presence of DNA methylating machinery in *I. ricinus*. Therefore, we aimed to evaluate the expression of newly identified putative *I. ricinus* DNA methyltransferases genes at different tick life stages and compare them to the DNA methylation profile (Fig. 7).

Our results revealed a relatively weak expression of the *IrDNMT3* gene (coding for a *de novo* methylation enzyme) during different *I. ricinus* life stages (Fig. 7A). Expression of the *IrDNMT1* gene (coding methylation maintaining enzyme) was similarly weak in larvae, fed and unfed females. The increased expression, however, was observed in eggs and fed nymphs. These data correspond with the results of Dot blot analysis (Fig. 7B), which confirmed the presence of cytosine methylation in *I. ricinus* ticks, but the level of methylation was significantly lower than that of hypermethylated onion gDNA (Fig. 7B).

The highest level of cytosine methylation was detected in partially fed nymph followed by lower signals of immunodetection in eggs and unfed females. Larvae and partially fed females showed only slightly visible signals. Our presented results of cytosine methylation are in agreement with the data from other arthropod species. For instance, cytosine methylation levels in *Tribolium castaneum* and bee *Osmia rufa* were increased during the transition from the new-born to mature stages (Song et al., 2017; Strachecka et al., 2017). However, the results of *IrDNMT1* and *IrDNMT3* expression differ from those for other arthropod species: DNMT1 and DNMT3 enzymes of *Nilaparvata lugens* were highly expressed in mated and gravid females but weakly in larvae, male adults, and virgin female adults, suggesting that mating induces their expressions (Zhang et al., 2015b); age-related increase in *Dnmt3* expression was also shown for *B. terrestris* workers, suggesting a novel association between aging and methylation (Lockett et al., 2016).

High expression levels of *IrDAMT* gene in eggs (Fig. 7A) are in agreement with the data from other arthropod species. For instance, the

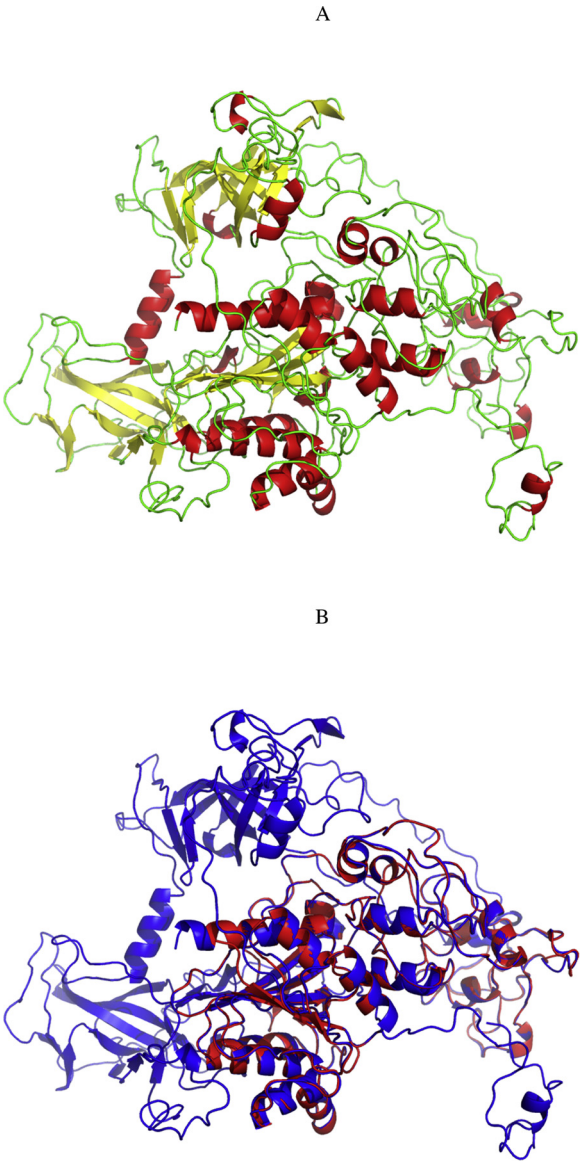


Fig. 3. 3D structures of putative IrDNMT1 and Dcm domain predicted by I-TASSER: **A.** Putative IrDNMT1, C-score = 1.20; α-helixes are colored in red, β-sheets in yellow and random coils in green. **B.** Comparison of predicted 3D model of putative IrDNMT1 with the model of Dcm domain, C-score = 1.64; the whole sequence model is colored in blue, the domain model in red.

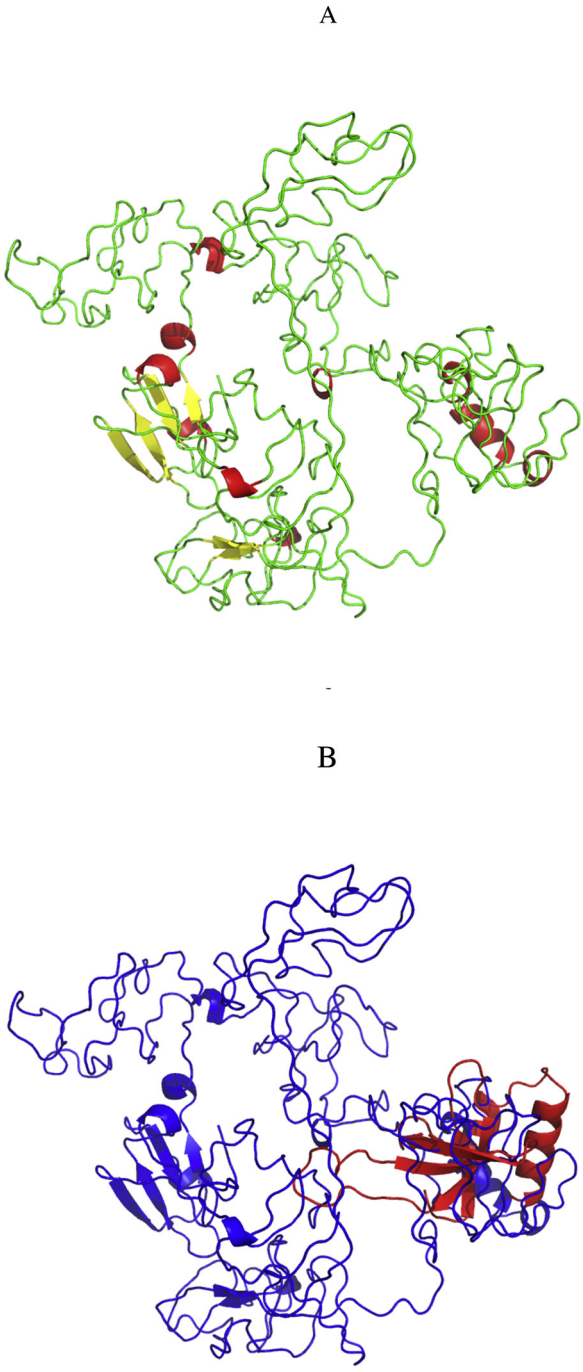


Fig. 4. 3D structures of putative IrDNMT3 and PWWP domain predicted by I-TASSER: **A.** Putative IrDNMT3, C-score = -1.48; α-helixes are colored in red, β-sheets in yellow and random coils in green. **B.** Comparison of predicted 3D model of putative IrDNMT3 with the model of PWWP domain, C-score = 0.64; the whole sequence model is colored in blue, the domain model in red.

highest expression levels of the adenine methyltransferase gene *Mettl4* in another blood-feeding parasite - *Aedes aegypti* (mosquito), were detected in the embryo and adult male stages (Falckenhayn et al., 2016). At the same time, DNA adenine methylation generated the strongest signal in *I. ricinus* females, and it was significantly higher when compared to other life stages of the tick (Fig. 7B). However, the level of adenine methylation was relatively high also in eggs and larvae that is in agreement with the data observed in insects. For instance, a strong 6 mA signal was detected in the DNA of embryos of *Drosophila melanogaster* and *Bombyx mori* at the very early stages (Zhang et al., 2015a; Wang et al., 2018), however, the signals were relatively weak in adult tissues and in late-stage embryos (Zhang et al., 2015a).

Thus, our results suggest that differences in the expression of DNA methyltransferases are associated with the complex life cycle of *I. ricinus* tick. Moreover, it was previously shown that the presence of heme leads to changes in the expression of several tick genes (Perner et al., 2018; Rodriguez-Valle et al., 2010). In our experiments, we observed extremely low levels of expression of the studied genes in fed females compared to their expression in unfed females. Additionally, the antibody staining showed only low DNA methylation in fed females. Therefore, we hypothesized that methyltransferase genes may be those whose expression is affected by blood meal and, more specifically, by ingestion of heme. However, this assumption requires further study, as well as the role of DNA demethylation enzymes in maintaining the DNA methylation profile in *I. ricinus*.

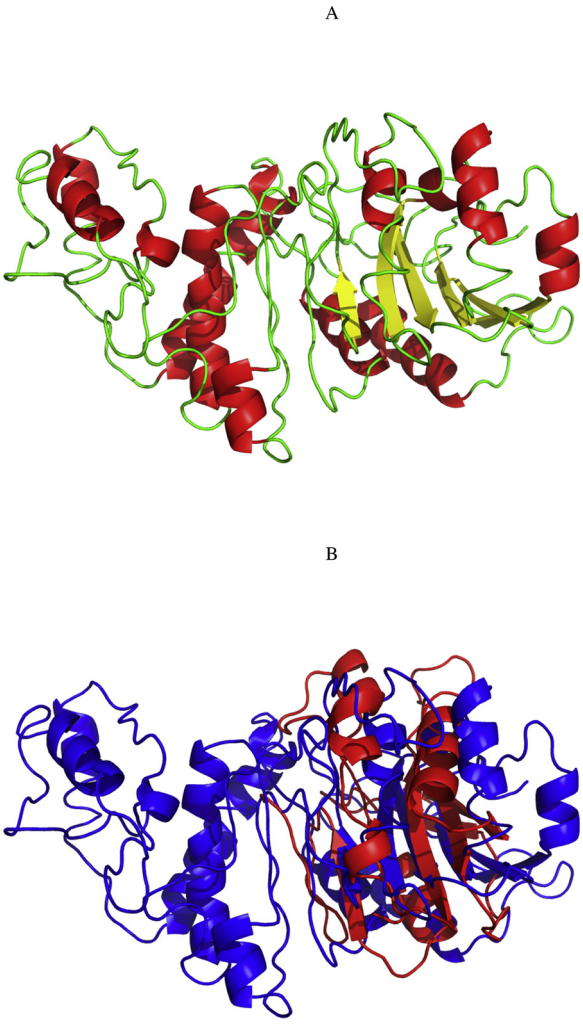


Fig. 5. 3D structure of putative IrDAMT and MT-A70 domain predicted by I-TASSER: **A.** Putative IrDAMT, C-score = -0.92; α-helixes are colored in red, β-sheets in yellow and random coils in green. **B.** Putative IrDAMT with the model of the MT-A70 domain, C-score = -0.76. The whole sequence model is colored in blue, the domain model in red.

4. Conclusions

In the presented study, we searched for the homologs of DNA methyltransferases in the *I. ricinus* transcriptome at different life stages – egg, larvae, fed nymph, adult female and fed female. We identified 13 transcripts, of which 10 transcripts were homologous to cytosine-methylating enzymes – DNMT1 and DNMT3, and 3 transcripts shared homology with adenine-methylating enzymes – DAMT. Thus, our findings indicate that *I. ricinus* expresses all of the enzymes producing typical eukaryotic gDNA methylation. The expression of the DNA methyltransferase differs among *I. ricinus* life stages. This observation can be caused by the complex life cycle of ticks. As we have demonstrated, *I. ricinus* possess the system of epigenetic regulation of gene expression and we aim to investigate the role of DNA methylation in the *I. ricinus* development and describe more in depth the character and degree of methylation in *I. ricinus* in our further work.

Funding

This study was supported by the Ministry of Education, Youth and Sports of the Czech Republic CEITEC 2020 project (LQ1601), INTERACTION project (LTARF 18021), and the Grant Agency of the Czech Republic (15-03044S, 18-27204S). Access to instruments and other

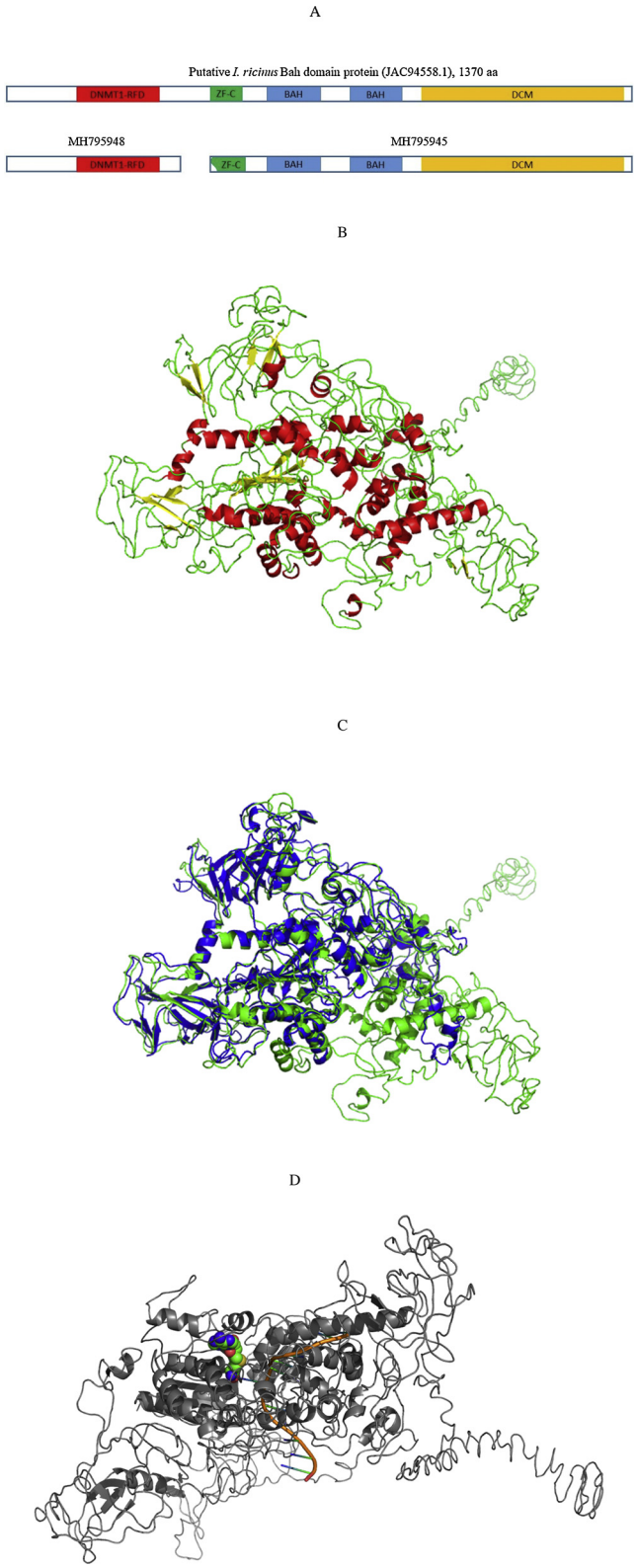


Fig. 6. Conserved domains and 3D structure of the putative *I. ricinus* BAH domain protein (GenBank: JAC94558.1). **A.** Schematic presentation of DNMT1-RFD, Zf-C, BAH and Dcm domains. **B.** 3D structure of the putative *I. ricinus* BAH domain protein predicted by I-TASSER, C-score = -1.65. **C.** Comparison of the models of the putative IrDNMT1 (blue color) and the putative *I. ricinus* BAH domain protein (green color). **D.** The model of the putative *I. ricinus* BAH domain protein with predicted binding sites for SAH and DNA.

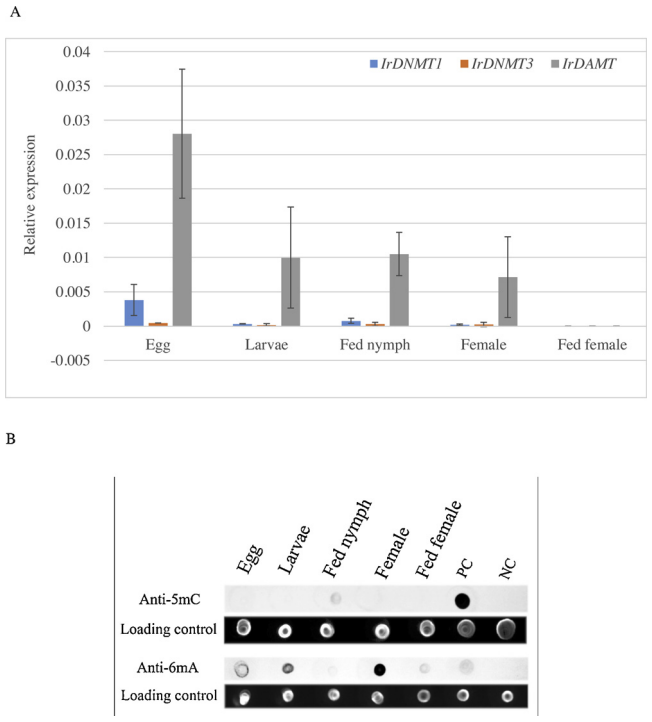


Fig. 7. Expression of DNA methyltransferases during different *I. ricinus* life stages (egg, larvae, partially fed nymph, female and partially fed female). **A.** Relative expression of *IrDNMT1*, *IrDNMT3* and *IrDAMT* measured using qRT-PCR. The Ct values of all three genes were scaled by normalization factor calculated from Ct values of 6 tick housekeeping genes, *EF1α*, *ferritin*, *β-tubulin*, *rp49*, *rps4*, and *ppia*. **B.** Identification of DNA methylation in *I. ricinus* using Dot blot assay with specific anti-5mC and anti-6 mA antibodies. Figures are representative of three independent biological replicates with three technical replicates per sample. Input DNA samples were stained with GelRed Nucleic Acid Stain for the loading control. PC – positive control, *A. cepa* gDNA; NC – negative control, PCR synthetized amplicons.

facilities was supported by the Czech research infrastructure for systems biology C4SYS project (LM2015055).

Acknowledgments

We thank Jan Erhart (Institute of Parasitology, Biology Centre of ASCR) for maintaining the tick colonies. Computational resources were supplied by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects CESNET (Project No. LM2015042) and CERIT-Scientific Cloud (Project No. LM2015085) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ttbdis.2019.101348.

References

Bewick, A.J., Vogel, K.J., Moore, A.J., Schmitz, R.J., 2017. Evolution of DNA methylation across insects. Mol. Biol. Evol. 34. https://doi.org/10.1093/molbev/msw264.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Cabezas-Cruz, A., Alberdi, P., Ayllón, N., Valdés, J.J., Pierce, R., Villar, M., de la Fuente, J., 2016. *Anaplasma phagocytophilum* increases the levels of histone modifying enzymes to inhibit cell apoptosis and facilitate pathogen infection in the tick vector *Ixodes scapularis*. Epigenetics 11, 303–319. https://doi.org/10.1080/15592294.2016.1163460.

Callebaut, I., Courvalin, J.-C., Mornon, J.-P., 1999. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. FEBS Lett. 446, 189–193. https://doi.org/10.1016/S0014-5793(99)00132-5.

Chojnacki, S., Cowley, A., Lee, J., Foix, A., Lopez, R., 2017. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic Acids Res. 45, W550–W553. https://doi.org/10.1093/nar/gkx273.

Cramaro, W.J., Revets, D., Hunewald, O.E., Sinner, R., Reye, A.L., Muller, C.P., 2015. Integration of *Ixodes ricinus* genome sequencing with transcriptome and proteome annotation of the naïve midgut. BMC Genomics 16, 871. https://doi.org/10.1186/s12864-015-1981-7.

de Meeûs, T., Humair, P.-F., Grunau, C., Delaye, C., Renaud, F., 2004. Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. Int. J. Parasitol. 34, 943–950. https://doi.org/10.1016/j.ijpara.2004.04.006.

Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. 1, 19–21. https://doi.org/10.1007/BF02712670.

Egekwu, N., Sonenshine, D.E., Bissinger, B.W., Roe, R.M., 2014. Transcriptome of the female synganglion of the black-legged tick *Ixodes scapularis* (Acari: ixodidae) with comparison between Illumina and 454 systems. PLoS One 9, e102667. https://doi.org/10.1371/journal.pone.0102667.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2018. The Pfam protein families database in 2019. Nucleic Acids Res. https://doi.org/10.1093/nar/gky995.

Falckenhayn, C., Carneiro, V.C., de Mendonça Amarante, A., Schmid, K., Hanna, K., Kang, S., Helm, M., Dimopoulos, G., Fantappié, M.R., Lyko, F., 2016. Comprehensive DNA methylation analysis of the *Aedes aegypti* genome. Sci. Rep. 6, 36444. https://doi.org/10.1038/srep36444.

Ferrè, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res. 33, W230–2. https://doi.org/10.1093/nar/gki412.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook. Humana Press, Totowa, NJ, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571.

Glastad, K.M., Hunt, B.G., Goodisman, M.A., 2014. Evolutionary insights into DNA methylation in insects. Curr. Opin. Insect Sci. 1, 25–30. https://doi.org/10.1016/J.COIS.2014.04.001.

Glastad, K.M., Hunt, B.G., Goodisman, M.A.D., 2013. Evidence of a conserved functional role for DNA methylation in termites. Insect Mol. Biol. 22, 143–154. https://doi.org/10.1111/imb.12010.

Gulia-Nuss, M., Nuss, A.B., Meyer, J.M., Sonenshine, D.E., Roe, R.M., Waterhouse, R.M., Sattelle, D.B., de la Fuente, J., Ribeiro, J.M., Megy, K., Thimmapuram, J., Miller, J.R., Walenz, B.P., Koren, S., Hostetler, J.B., Thiagarajan, M., Joardar, V.S., Hannick, L.I., Bidwell, S., Hammond, M.P., Young, S., Zeng, Q., Abrudan, J.L., Almeida, F.C., Ayllón, N., Bhide, K., Bissinger, B.W., Bonzon-Kulichenko, E., Buckingham, S.D., Caffrey, D.R., Caimano, M.J., Croset, V., Driscoll, T., Gilbert, D., Gillespie, J.J., Giraldo-Calderón, G.I., Grabowski, J.M., Jiang, D., Khalil, S.M.S., Kim, D., Kocan, K.M., Koči, J., Kuhn, R.J., Kurtti, T.J., Lees, K., Lang, E.G., Kennedy, R.C., Kwon, H., Perera, R., Qi, Y., Radolf, J.D., Sakamoto, J.M., Sánchez-Gracia, A., Severo, M.S., Silverman, N., Šimo, L., Tojo, M., Tornador, C., Van Zee, J.P., Vázquez, J., Vieira, F.G., Villar, M., Wespiser, A.R., Yang, Y., Zhu, J., Arensburger, P., Pietrantonio, P.V., Barker, S.C., Shao, R., Zdobnov, E.M., Hauser, F., Grimmelikhuijzen, C.J.P., Park, Y., Rozas, J., Benton, R., Pedra, J.H.F., Nelson, D.R., Unger, M.F., Tubio, J.M.C., Tu, Z., Robertson, H.M., Shumway, M., Sutton, G., Wortman, J.R., Lawson, D., Wikel, S.K., Nene, V.M., Fraser, C.M., Collins, F.H., Birren, B., Nelson, K.E., Caler, E., Hill, C.A., 2016. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. Nat. Commun. 7, 10507. https://doi.org/10.1038/ncomms10507.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hoang, D.T., Vinh, L.S., Flouri, T., Stamatakis, A., von Haeseler, A., Minh, B.Q., 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol. Biol. 18, 11. https://doi.org/10.1186/s12862-018-1131-3.

Holman, L., Trontti, K., Helanterä, H., 2016. Queen pheromones modulate DNA methyltransferase activity in bee and ant workers. Biol. Lett. 12, 20151038. https://doi.org/10.1098/rsbl.2015.1038.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35, W585–7. https://doi.org/10.1093/nar/gkm259.

Hunt, B.G., Glastad, K.M., Yi, S.V., Goodisman, M.A.D., 2013. The function of intragenic DNA methylation: insights from insect epigenomes. Integr. Comp. Biol. 53, 319–328. https://doi.org/10.1093/icb/ict003.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Koci, J., Simo, L., Park, Y., 2013. Validation of internal reference genes for real-time quantitative polymerase chain reaction studies in the tick, *Ixodes scapularis* (Acari: ixodidae). J. Med. Entomol. 50, 79–84. https://doi.org/10.1603/me12034.

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

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245. https://doi.org/10.1093/nar/gkw290.

Li-Byarlay, H., 2016. The function of DNA methylation marks in social insects. Front. Ecol. Evol. 4, 57. https://doi.org/10.3389/fevo.2016.00057.

Lockett, G.A., Almond, E.J., Huggins, T.J., Parker, J.D., Bourke, A.F.G., 2016. Gene expression differences in relation to age and social environment in queen and worker bumble bees. Exp. Gerontol. 77, 52–61. https://doi.org/10.1016/J.EXGER.2016.02.007.

Lyko, F., Maleszka, R., 2011. Insects as innovative models for functional studies of DNA methylation. Trends Genet. 27, 127–131. https://doi.org/10.1016/J.TIG.2011.01.003.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., Bryant, S.H., 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, D225–D229. https://doi.org/10.1093/nar/gkq1189.

Marinus, M.G., Løbner-Olesen, A., 2014. DNA methylation. EcoSal Plus 6. https://doi.org/10.1128/ecosalplus.ESP-0003-2013.

Messerschmidt, D.M., Knowles, B.B., Solter, D., 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. 28, 812–828. https://doi.org/10.1101/gad.234294.113.

Meyer, J.M., Kurtti, T.J., Van Zee, J.P., Hill, C.A., 2010. Genome organization of major tandem repeats in the hard tick, *Ixodes scapularis*. Chromosome Res. 18, 357–370. https://doi.org/10.1007/s10577-010-9120-4.

Nijhof, A.M., Balk, J.A., Postigo, M., Jongejan, F., 2009. Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus (Boophilus)* microplus and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. BMC Mol. Biol. 10, 112. https://doi.org/10.1186/1471-2199-10-112.

Pearson, W.R., 2013. An introduction to sequence similarity ("homology") searching. Curr. Protoc. Bioinf. https://doi.org/10.1002/0471250953.bi0301s42. Chapter 3, Unit3.1.

Perner, J., Kotál, J., Hatalová, T., Urbanová, V., Bartošová-Sojková, P., Brophy, P.M., Kopáček, P., 2018. Inducible glutathione S-transferase (IrGST1) from the tick *Ixodes ricinus* is a haem-binding protein. Insect Biochem. Mol. Biol. 95, 44–54. https://doi.org/10.1016/J.IBMB.2018.02.002.

Qiu, C., Sawada, K., Zhang, X., Cheng, X., 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat. Struct. Biol. 9, 217–224. https://doi.org/10.1038/nsb759.

Rodriguez-Valle, M., Lew-Tabor, A., Gondro, C., Moolhuijzen, P., Vance, M., Guerrero, F.D., Bellgard, M., Jorgensen, W., 2010. Comparative microarray analysis of *Rhipicephalus (Boophilus) microplus* expression profiles of larvae pre-attachment and feeding adult female stages on Bos indicus and Bos taurus cattle. BMC Genomics 11, 437. https://doi.org/10.1186/1471-2164-11-437.

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. https://doi.org/10.1038/nprot.2010.5.

Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108. https://doi.org/10.1038/nprot.2008.73.

Schwarz, A., Tenzer, S., Hackenberg, M., Erhart, J., Gerhold-Ay, A., Mazur, J., Kuharev, J., Ribeiro, J.M.C., Kotsyfakis, M., 2014. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. Mol. Cell Proteomics 13, 2725–2735. https://doi.org/10.1074/mcp.M114.039289.

Song, X., Huang, F., Liu, J., Li, C., Gao, S., Wu, W., Zhai, M., Yu, X., Xiong, W., Xie, J., Li, B., 2017. Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. DNA Res. 24, 445–457. https://doi.org/10.1093/dnares/dsx016.

Strachecka, A., Chobotow, J., Paleolog, J., Łoś, A., Schulz, M., Teper, D., Kucharczyk, H., Grzybek, M., 2017. Insights into the biochemical defence and methylation of the solitary bee *Osmia rufa* L: a foundation for examining eusociality development. PLoS One 12, e0176539. https://doi.org/10.1371/journal.pone.0176539.

Suzuki, G., Shiomi, M., Morihana, S., Yamamoto, M., Mukai, Y., 2010. DNA methylation and histone modification in onion chromosomes. Genes Genet. Syst. 85, 377–382. https://doi.org/10.1266/ggs.85.377.

Wang, P., Doxtader, K.A., Nam, Y., 2016. Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. Mol. Cell 63, 306–317. https://doi.org/10.1016/j.molcel.2016.05.041.

Wang, X., Li, Z., Zhang, Q., Li, B., Lu, C., Li, W., Cheng, T., Xia, Q., Zhao, P., 2018. DNA methylation on N6-adenine in lepidopteran *Bombyx mori*. Biochim. Biophys. Acta - Gene Regul. Mech. 1861, 815–825. https://doi.org/10.1016/j.bbagrm.2018.07.013.

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.-H., Colbourne, J.K., Clark, A.G., Werren, J.H., 2013. Function and evolution of DNA methylation in *Nasonia vitripennis*. PLoS Genet. 9, e1003872. https://doi.org/10.1371/journal.pgen.1003872.

Wedd, L., Maleszka, R., 2016. DNA Methylation and Gene Regulation in Honeybees: From Genome-Wide Analyses to Obligatory Epialleles. Springer, Cham, pp. 193–211. https://doi.org/10.1007/978-3-319-43624-1_9.

Wu, S., Skolnick, J., Zhang, Y., 2007. Ab Initio Modeling of Small Proteins by Iterative TASSER Simulations. https://doi.org/10.1186/1741-7007-5-17.

Yan, H., Bonasio, R., Simola, D.F., Liebig, J., Berger, S.L., Reinberg, D., 2015. DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu. Rev. Entomol. 60, 435–452. https://doi.org/10.1146/annurev-ento-010814-020803.

Yang, J., Roy, A., Zhang, Y., 2013. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29, 2588–2595. https://doi.org/10.1093/bioinformatics/btt447.

Yang, J., Roy, A., Zhang, Y., 2012. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. Nucleic Acids Res. 41, D1096–D1103. https://doi.org/10.1093/nar/gks966.

Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43, W174–81. https://doi.org/10.1093/nar/gkv342.

Yarychkivska, O., Shahabuddin, Z., Comfort, N., Boulard, M., Bestor, T.H., 2018. BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate *de novo* and maintenance methylation *in vivo*. J. Biol. Chem. https://doi.org/10.1074/jbc.RA118.004612. jbc.RA118.004612.

Zhang, C., Freddolino, P.L., Zhang, Y., 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. Nucleic Acids Res. 45, W291–W299. https://doi.org/10.1093/nar/gkx366.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., Li, C., Liu, B., Luo, Y., Zhu, Y., Zhang, N., He, S., He, C., Wang, H., Chen, D., 2015a. N6-methyladenine DNA modification in Drosophila. Cell 161, 893–906. https://doi.org/10.1016/j.cell.2015.04.018.

Zhang, J., Xing, Y., Li, Y., Yin, C., Ge, C., Li, F., 2015b. DNA methyltransferases have an essential role in female fecundity in brown planthopper, *Nilaparvata lugens*. Biochem. Biophys. Res. Commun. 464, 83–88. https://doi.org/10.1016/j.bbrc.2015.05.114.

Zhang, J., Zheng, Y.G., 2016. SAM/SAH analogs as versatile tools for SAM-Dependent methyltransferases. ACS Chem. Biol. 11, 583–597. https://doi.org/10.1021/acschembio.5b00812.

Vechtova, P., Fussy, Z., Cegan, R., Sterba, J., Erhart, J., Benes, V., Grubhoffer, L. Catalogue of stage-specific transcripts in *Ixodes ricinus* and their potential functions during tick life-cycle. *Parasites and Vectors*. Accepted.

Catalogue of stage-specific transcripts in *Ixodes ricinus* and their potential functions during tick life-cycle

Pavlina Vechtova[1,2,+], Zoltan Fussy[1,3], Radim Cegan[4], Jan Sterba[1,2], Jan Erhart[2], Vladimir Benes[5], Libor Grubhoffer[1,2]

[1]University of South Bohemia, Faculty of Science, Ceske Budejovice, Czech Republic
[2]Biology Centre CAS, Institute of Parasitology, Ceske Budejovice, Czech Republic
[3]Charles University, Faculty of Science, BIOCEV, Prague, Czech Republic
[4]Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic
[5]GeneCore, EMBL, Heidelberg, Germany

*Correspondence: vechtova@prf.jcu.cz
E-mails:
PV: vechtova@prf.jcu.cz
ZF: zoltan.fussy@gmail.com
RC: cegan@ibp.cz
JS: sterbaj@prf.jcu.cz
JE: erhart@paru.cas.cz
VB: benes@embl.de
LG: liborex@paru.cas.cz

Abstract

Background: Castor bean tick *Ixodes ricinus* is an important vector of several clinically important diseases whose prevalence increases with accelerating global climate changes. Characterization of a tick life cycle is thus of great importance. However, researchers mainly focus on specific organs of fed life stages while early development of this tick species is largely neglected.

Methods: In an attempt to better understand the life cycle of this wide-spread arthropod parasite, we sequenced the transcriptomes of four life stages (egg, larva, nymph, and adult female),

including unfed and partially blood-fed individuals. To enable a more reliable identification of transcripts and their comparison in all five transcriptome libraries, we validated an improved-fit set of five *I. ricinus*-specific reference genes for internal standard normalization of our transcriptomes. Then, we mapped biological functions to transcripts identified in different life stages (clusters) to elucidate life stage-specific processes. Finally, we drew conclusions from the functional enrichment of these clusters specifically assigned to each transcriptome, also in the context of recently published transcriptomic studies in ticks.

Results: We found that reproduction-related transcripts are present in both fed nymph and fed female, underlining the poorly documented importance of ovaries as moulting regulators in ticks. Additionally, we identified transposase transcripts in tick eggs suggesting elevated transposition during embryogenesis, co-activated with factors driving developmental regulation of gene expression. Our findings also highlight the importance of the regulation of energetic metabolism in tick eggs during embryonic development and glutamate metabolism in nymphs.

Conclusions: Our study presents novel insights into stage-specific transcriptomes of *I. ricinus* and extends the current knowledge of this medically important pathogen, especially in the early phases of its development.

Keywords

Ixodes ricinus; Tick development; Transcriptome assembly; Reference gene validation; Life stage

Background

Predominantly European tick *Ixodes ricinus* is a common blood-feeding arthropod transmitting several widespread human pathogens, including the spirochaete *Borrelia burgdorferi* causing Lyme disease, tick-borne encephalitis virus (TBEV), the causative agent of human encephalitis and meningitis, *Anaplasma phagocytophilum*, an intracellular alpha-proteobacterium causing granulocytic anaplasmosis, and *Rickettsia* spp. causing spotted fever syndrome [1–4]. As the natural distribution and activity of *I. ricinus* have been augmenting over past decades, so have the emergence and manifestation of tick-borne diseases. Worldwide, there are about 10,000 cases of tick-borne encephalitis [5] and 85,000 cases of Lyme disease [6], reported annually and these epidemiological numbers have been raising attention with respect to public health, economy or

tourism [7–11]. The life cycle of arthropod-borne pathogens is tightly bound to the life cycle of their vectors/hosts and thus understanding the life cycle of a vector organism often reveals important aspects of vector-pathogen dynamics including the factors influencing disease transmission to final hosts.

Unlike other haematophagous arthropods, like mosquitoes or flies, ticks exhibit a complex and rather long life cycle and they usually require feeding on several host organisms for its completion. *Ixodes* ticks are able to finish their life cycle within 3 to 6 years in wildlife depending on environmental conditions [12]. During its development, the tick hatches from an egg and undergoes metamorphosis and moults to the next active life stage: larva, nymph, or adult [13,14]. Each moulting is preceded by blood-feeding on the respective host; the selection of host species is perhaps the broadest of all ticks ranging from small mammals, birds, and reptiles in immature stages to large mammals in adult ticks [15]. The time intervals for the completion of each life stage vary and are greatly influenced by many factors such as season, host abundance, selection of host species, or climatic conditions [16]. The feeding itself lasts 3-5 days in a larval stage, 4-7 days in nymphs, and 7-11 days in adult females [17]. Adult females mate with adult males during feeding on the host to accomplish reproduction. Laid egg batch contains, on average, 2,000-2,500 eggs [18], but the number of eggs in one batch can reach up to 4,000 [19]. The reason for an extraordinarily long life cycle of *I. ricinus* is arguably its three-host life cycle, in particular when the tick drops off the host after each blood meal and undergoes metamorphosis and moulting off the host [14]. Furthermore, the absence of a host or suboptimal microclimatic conditions (e.g., low temperature) drive the tick to enter a diapause, which can be induced in any life stage and contributes to the extension of its life cycle [20].

Longevity along with the blood-feeding ectoparasitic life strategy of ticks must have been preceded by many adaptations, differing from those of blood-feeding insects and including features in the regulation of development and metamorphosis that are yet a matter of debate [13]. The efforts to describe factors controlling the development of other arthropod vectors, such as vectors of malaria (*Anopheles gambiae*) and yellow fever (*Aedes aegypti*), are mainly focused on the blood-feeding and reveal an upregulation of genes associated with blood meal processing, peritrophic matrix formation, egg development, and immunity on the organismal level [21,22] or in salivary glands [23–26], the latter being regarded as a crucial mediator of pathogen transmission to the mammalian host. Several transcriptomic studies were focusing on molecules

that might directly influence feeding, such as haem utilization in ticks [27] or arthropod proteases, both being essential factors enabling a haematophagous life strategy [28]. Transcriptional regulation of the entire life cycle, controlling tick ontogenesis and development has not been fully covered to date and existing research has only focused on a specific organ [29], or a life stage [30].

Due to an increase in its epidemiological importance, *I. ricinus* tick has become a species featured in many recent transcriptomic studies. Majority of them focused on the transcription in salivary glands and/or midgut, which are the key organs in the tick-borne pathogens' life cycle. Studying these organs in response to blood feeding can be instrumental for the identification of factors that underlie survival and dissemination of pathogens within their vector and their transmission into the final host [24,31–37]. More specifically, expression analyses using tick haemocytes, the main actors in tick immunity, can unveil the character of the immunity barrier for tick-borne pathogens [29]. Organism-level transcriptomes of feeding stages, on the other hand, can provide a picture of global changes induced by blood meal [30], hence a more thorough description of factors driving tick developmental processes throughout its life cycle will be instrumental in understanding the process of host-seeking and blood-feeding as an integral event in tick development.

In this study, we focused on transcripts associated with development, aiming at the presentation of significant new data of the main processes linked to specific life stages of *I. ricinus* and functions that are stably expressed. We present a catalogue of transcripts in transcriptome assemblies of several life stages of *I. ricinus* to provide an outline of transcription important for specific time points of tick development and functions in particular life stages. Our data identified transcripts involved in tick embryonic development thereby providing a source of information for research in tick cell lines, a tick model for in vitro research derived from tick embryonic cells [38]. This represents a significant contribution, which facilitates an initiation and development of methods largely applicable by means of *in vitro* models such as double-stranded RNA post-transcriptional gene silencing (RNAi) or targeted genome editing using CRISPR/Cas9.

Methods

Sample preparation and next-generation sequencing

Both partially fed and unfed life stages of *I. ricinus* were collected in the tick rearing facility of Institute of Parasitology, Biology Centre CAS; partially fed life stages were fed on laboratory guinea pigs obtained at the animal rearing facility therein. The partially fed nymphs were feeding 3 to 4 days and partially fed females 5 to 6 days. Under rearing facility conditions, fed females start laying eggs 4 weeks after feeding and the process takes approximately 2 weeks. The eggs were collected immediately after laying and thus represent an early stage of embryogenesis. Larvae were collected after complete hatching of an egg clutch, which usually takes 2 to 4 weeks. Unfed females were hatching 7 to 9 weeks after the full engorgement of feeding nymphs. Females were collected after all females had moulted from a batch of nymphs that were feeding simultaneously on laboratory animals. Fed stages were not dissected to remove host blood as haemolymph containing haemocytes and possibly other cells would be lost in the process.

Total RNA was isolated from 3 halves of egg clutches (3 × ~600 eggs), 3 batches of larvae hatched from 3 halves of egg clutches (3 × ~600 individuals), partially fed nymphs (3 × 10 individuals), adult (unfed) females (3 × 7 individuals) and partially fed females (3 × 3 individuals) of *I. ricinus* tick using NucleoSpin RNA II (Macherey-Nagel, Duren, Germany) according to manufacturer's instructions. The concentration of RNA was measured using Implen NanoPhotometer (Implen, Munchen, Germany) and the quality of RNA was determined using 2100 Bioanalyzer with RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA, USA). The 3 RNA samples of each life stage were pooled to obtain a single RNA sample per life stage. Each sample is a mixture of three different cohorts of ticks collected in a tick rearing facility in order to cover the genetic variability existing between individuals and to reconstruct the maximum number of transcripts that can be possibly expressed in each life stage. cDNA synthesis and library preparation was carried out using TruSeq DNA Sample Prep kit v2 (PE50 reads) (Illumina, San Diego, CA, USA), followed by sequencing on the HiSeq 2000 platform; both the library preparation and sequencing were performed by the GeneCore facility of the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

Transcriptome assembly and annotation

The PE50 raw reads of all five stages were trimmed off sequencing adapters, short and low quality reads using Trimmomatic (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 trimmomatic parameters) [40]. Trimmed reads of all five libraries were used to build a single assembly using Trinity assembler v2.1.1 [39] in order to obtain a highly comprehensive catalogue of complete and high-quality transcripts. Assembly completeness was assessed by mapping the raw reads to the assembly using Bowtie2 v2.3.0 [40] and Samtools flagstat v1.6 [41], followed by BUSCO v3.0.2 search [42] using the catalogue of conserved Arthropoda orthologues in protein mode (database last accessed 7 Jul 2017). Contigs showing high similarity to taxonomic groups other than Panarthropoda according to the BlobTools v1.0 pipeline [43] were omitted as possible contaminants.

The contigs of transcriptome assembly were conceptually translated into protein sequences in all six frames. To retrieve annotations, contigs were queried in protein space against TrEMBL, Swiss-Prot, and non-redundant protein database (nr) (downloaded 30 Jul 2017) using blastp of BLAST v2.6.0+ with default parameters except for the e-value set to 1×10^{-4}, returning a maximum of 10 best hits (-max_target_seqs 10). Using a custom Python3 script, the data were integrated with the annotations retrieved by InterProScan v5.36-75.0 [44], which also provided the gene ontology (GO) terms for contigs. Some annotations and GO terms were transferred from the *I. ricinus* proteome at UniProt, with the criterion for an assignment being >90% sequence identity on the protein level.

Validation of assembly completeness

Assembly completeness and comprehensiveness was assessed by the search of transcripts showing tissue-specific expression. The query protein sequences were downloaded from GeneBank (database last accessed 11 Feb 2020) (see Table 1 for accession numbers). Blastp of BLAST v2.6.0+ with default parameters was used to search a transcriptome assembly translated into protein sequences in all six frames. The best blast hits were collected and their corresponding nucleotide sequences were retrieved from *I. ricinus* transcriptome assembly. Alignments of query sequences and their best blast hits were constructed using MAFFT 1.4.0 [45] integrated within Geneious Prime 2020.0.5 (https://www.geneious.com).

Table 1 Genes used for the validation of assembly completeness. The tissue specificity and supporting references of the listed genes in *I. ricinus* are also provided.

gene	accession number of nucleotide/ protein query sequence	tissue	reference
IrCD 1	EF428204.1/ABO26561.1	midgut	[46]
IrCD 2	HQ615697.1/ADU03674.1	midgut	[46]
IrCD 3	HQ615698.1/ADU03675.1	midgut	[46]
Iris	AJ269658.2/CAB55818.2	salivary glands	[47]
Ixoderin A	AY341424.1/AAQ93650.1	haemocytes, salivary glands, midgut	[48]
Ixoderin B	AY643518.3/AAV41827.2	salivary glands	[48]

Reference genes validation

Three individual batches (biological replicates) of eggs (3 × ~600 individuals), larvae (3 × ~600 individuals), nymphs (3 × 15 individuals), partially fed nymphs (3 × 10 individuals), females (3 × 7 individuals), and partially fed females (3 × 3 individuals) of *I. ricinus* were collected in the tick rearing facility as above. Partially fed stages were dissected and washed off host blood in Ringer physiological solution according to Glaser (1917) [49]. Removal of host blood was important as it can inhibit the PCR reaction; since the expression of reference genes is expected to be at similar levels in all tissues, the absence of haemolymph, in this case, did not affect the results. Ticks were homogenized in Mixer mill MM 400 (Retsch, Haan, Germany) with steel beads in LBS buffer and RNA was isolated using NucleoSpin RNA Plus (Macherey-Nagel, Duren, Germany) according to manufacturer's instructions. The concentration of RNA was measured using Implen NanoPhotometer (Implen, Munich, Germany). cDNA was synthesized using ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs, Ipswich, MA, USA). Primers for reference genes were designed according to *I. ricinus* transcripts showing the highest sequence identity with publicly available sequences of *I. scapularis* (see Additional file 1: Table S1 for primer sequences) as determined by blastn of BLAST v2.6.0+ with default parameters. qPCR reactions for reference genes expression assay in each life stage were prepared using the qPCR 2× SYBR Master Mix (Top-Bio, Vestec, Czech Republic) with 20 ng of total RNA transcribed to cDNA as an input in each reaction. The qRT-PCR reaction and fluorescence acquisition were done in a LightCycler 480 Real-Time PCR System cycler (Roche, Basel, Switzerland) and the resulting Ct values were recorded and retrieved in LightCycler 480 Software release 1.5.0 SP4 (Roche, Mannheim, Germany). The validation of reference genes was done using BestKeeper v.1 Microsoft Excel-based tool [50]. The gene stability ranking is based

on the calculation of pairwise variation of candidate reference genes among samples of different life stages and their biological replications as the standard deviation of $\log_2$-transformed expression ratios [51].

The comparison of transcriptomes of *Ixodes ricinus* life stages

The reads of the common assembly were redistributed among the tick life stages and their read counts were assigned to the transcripts of the individual life stages in order to facilitate identification and annotation of the transcripts and to enable enrichment analysis and visualization. Transcript quantification was not applied as biological triplicates are necessary for statistical support. Instead, read count data were used for the annotation of the tick transcriptome assemblies and GO enrichment analysis through identification of correct transcripts and their comparison among individual assemblies. The read count estimation was performed using RSEM (RNA-Seq by Expectation-Maximization) [52] implemented in the Trinity Transcript Quantification pipeline (accessed 15 Oct 2019, [39]). The contigs with low mapping counts were removed in order to avoid the presence of chimeric, fragmented, or biologically insignificant transcripts (cpm=2 cut-off). BUSCO v3.0.2 search [42] was performed with the cpm=2 filtered dataset and compared with BUSCO search performed with unfiltered transcriptome described in the “Transcriptome assembly and annotation” section. The mapping counts of each transcriptome were cross-sample normalized to the reference gene counts in order to identify transcripts that are present and specific for each transcriptome. The reference gene scaling factor was calculated using 5 normalization methods: TMM (Trimmed Mean of M-values), TMMwsp (Trimmed Mean of M-values with singleton pairing), UQ (Upper Quartile), and RLE (Relative Log Expression) implemented in edgeR v3.12.0 as suggested by [53] and Median normalization (MED) implemented in edgeR as described in [54]. The reference genes’ read mapping counts were retrieved from the scaled matrices and their geometric means and geometric standard deviations were calculated in order to select the most efficient normalization method.

Contig clustering and enrichment analysis

The counts of mapped reads in each transcriptome were transformed into a $\log_2$-scaled matrix in order to perform an enrichment analysis. Library-specific transcripts were inferred by hierarchical clustering of the Morpheus matrix analysis software

(https://software.broadinstitute.org/morpheus/, last accessed 30 Nov 2019) using complete linkage and 1-Pearson correlation metrics. Based on the resulting dendrogram, transcripts were assigned to clusters that are represented in one or more transcriptome assemblies. 2-50 clusters were built and a custom Python3 script was used to calculate within-cluster read count variability as a sum of squares function of Euclidean distances from respective cluster centroids. The frozen code is available on GitHub (Fussy, Z. (2019), GitHub repository, https://github.com/morpholino/PYTHON/blob/master/clustering_metrics_kliste.py). We used the elbow method to find the inflexion point where a minimal number of clusters explains ~90% of the total variability of the complete read count matrix. The transcripts assigned to each library were then subjected to GO enrichment analysis by goa-tools v0.8.12 [55] with Benjamini/Hochberg false discovery rate correction; the GO terms relationship file (go-basic.obo) from geneontology.org was last accessed 30 Nov 2019. A custom Python3 script was used to visualize the enrichment inferred for each transcriptome assembly.

Results

Transcriptome assembly and annotation

A total of 430,604,850 paired-end 50-bp reads were obtained by sequencing of all five *I. ricinus* life stages. Trimming and quality filtering yielded 428,514,142 reads (see Additional file 1: Table S1 for details). The transcriptome assembly with Trinity assembler produced 117,583 “Trinity isoforms” and 83,534 “Trinity genes”. According to samtools flagstat, the Bowtie2 mapping rate of all libraries is around 80 %, which corresponds to good quality assemblies with little information being lost in unmapped reads (see quality and mapping statistics available in Additional file 2: Table S2).

The RSEM package was used to distribute mapped reads among the five transcriptome assemblies of *I. ricinus.* Reads with low mapping counts (cpm=2 threshold) were removed from the dataset of mapped reads as potentially misassembled or chimaeric. The number of transcripts after this low count filtering dropped from 83,534 to 25,872, which roughly corresponds to the genome assembly report of the related tick species, *I. scapularis* (23,340 transcripts) [56].

Read mapping counts were calculated in order to identify transcripts that are library-specific, demonstrating the presence or absence of individual transcripts across life stages. This does not

necessitate biological replicates compared to statistical analyses required by quantitative RNAseq pipelines.

To enforce a more robust comparison, we scaled the five life-stage libraries to reference transcript counts selected through quantitative reference gene validation (see "Reference gene validation" section below).

Of the five different normalization techniques employed, the RLE method showed the lowest dispersion of reference gene variability among the five assemblies (see Additional file 2: Table S2.). The RLE was used to calculate reference gene normalization factor using mapping counts of five most stable reference genes of intermediate read mapping counts (*rps4, RpL32, rpl4, ferritin, RpL13A*), to which the matrix of read mapping counts of the five life stages was scaled.

Of 25,872 contigs of the pooled transcriptome assembly, whose quality and biological relevance was supported by a high number of mapped reads (i.e. contigs passing the cpm=2 threshold), we could find annotation for 13,626 using InterProScan and 9,510 of these were assigned GO terms. For some of these, and additional 653 contigs, we could find annotation using BLAST against the *I. ricinus* proteome deposited at UniProt, totalling 14,279 and 11,282 contigs with IPS and/or GO terms, respectively.

BUSCO search of Arthropoda conserved orthologues within the unfiltered 83,534-contig assembly as input reported 95.7% completeness. The assembly of contigs passing the cpm=2 threshold exhibited 95.6 % completeness, which shows that cpm=2 filtering effectively removed low quality, fragmented, or chimeric contigs introducing false information or carrying insignificant biological role for the *I. ricinus* assembly. A detailed BUSCO report is given in Additional file 3: Table S3.

Whole-body transcriptomes may lack mRNAs of lowly expressed genes. This concern motivated us to perform additional verification of a comprehensiveness of our assembly. Our test was based on an assumption that genes that exhibit tissue-specific expression can be expressed in minute abundances compared to the genes expressed constitutively and organism-wide. Thus, their expression can be undetected if sequencing depth is insufficient or the assembly is of poor quality. We searched publications dealing with *I. ricinus* tissue-specific expression and selected randomly 3 tissue-specific genes and their paralogues. The list of sequences, their accession numbers, and corresponding publications showing their tissue-specific expression profile in *I.*

ricinus are provided in Table 1. In particular, we chose 3 paralogues of Cathepsin D (IrCDs) with expression restricted to tick midgut [46], a single sequence of the family of tick Serine protease inhibitors (Serpins) identified in *I. ricinus* salivary glands (Iris) [47], and two paralogues of fibrinogen-related protein, Ixoderin A and Ixoderin B showing tissue-specific expression profiles. Expression of Ixoderin A was detected in haemocytes, salivary glands and midgut and transcripts of Ixoderin B were found in salivary gland tissue only [57]. Thus, a higher abundance and consequently a higher chance of full recovery of Ixoderin A transcript was expected in comparison to Ixoderin B. Importantly, our assembly contains sequences corresponding to all six queries (see Additional files 5 – 10 for alignments). We found four isoforms of Iris; all four sequences showed similarly high sequence identity. Of note, for the salivary gland-specific paralogue Ixoderin B we also found a transcript, albeit 5' truncated. Intriguingly, the Ixoderin B query and hit showed only 70% identity on the amino acid level. However, best hits within *I. ricinus*-specific nr and tsa_nr databases (last accessed 11 Jun 2020) also returned a best hits having 75% (ABO09954.1) and 89% (JAB75084.1) identity, respectively, both annotated as Ixoderin B5. Nucleotide sequences of these two hits were included in the Ixoderin B alignment to support the identity of the putative Ixoderin B5 sequence from our assembly.

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GIDG00000000. The version described in this article is the first version, GIDG01000000.

Reference genes validation

The selection of candidate reference genes for validation should be done carefully, ideally based on existing data. Within Acari, only 4 species were analysed for the reference gene selection and their validation: *Tetranychus urticae* [59], *Ripicephalus microplus, R. appendiculatus* [58], and *I. scapularis* [61]. Our study focused on the identification of transcripts in tick life stages with recommendations from previous publications [58,59,61] (Table 2). In total, 13 reference genes were selected for testing; *EF1a, ferritin, GAPDH, H3F3A, ppiA, RpL32, RpL13a, rpl4, rps4, sdha, TBP, TUBB, v-ATPase*. Of these, *sdha* (succinate dehydrogenase complex flavoprotein subunit A) and *TBP* (TATA-Box Binding Protein) were excluded from further analysis due to nonspecific amplification during qRT-PCR assay optimization (data not shown). The best BLAST hits in our *I. ricinus* transcriptome assembly, using published *I. scapularis* references as

queries, were used for primer design. These sequences were manually annotated and uploaded to GenBank. The list of *I. ricinus* reference genes used in reference gene validation assay, their accession numbers, and their respective *I. scapularis* queries with references are given in Table 2. The quantification of *I. ricinus* candidate reference genes expression was performed using qRT-PCR assay. The expression levels of reference genes and their variation in each life stage are given in Fig. 1A and B, respectively.

Table 2 Genes used in reference gene validation assay. The genes were selected based on validation in the referenced works.

gene	**annotation**	**reference**	**query sequence**	***I. ricinus* homologue sequence**
EF1α	elongation factor 1α	[58] [59]	GU074814.1	MN728895
ferritin	ferritin (somatic)	[60]	AY277906.1	MN728896
GAPDH	glyceraldehyde 3-phosphate dehydrogenase	[61] [58] [59]	XM_002434302.1	MN728904
H3F3A	H3 histone family 3A	[58]	XM_002399526.1	MN728905
RpL13A	ribosomal protein L13A	[61]	XM_002436237.1	MN728901
ppiA	cyclophilin-type peptidylprolyl cis-trans isomerase A	[58]	XM_002407873.1	MN728903
RpL32	ribosomal protein L32	[59]	XM_002399465.1	MN728898
rpl4	ribosomal protein L4	[58]	XM_002402278.1	MN728900
rps4	ribosomal protein S4	[61]	DQ066214.1	MN728897
sdha	succinate dehydrogenase complex subunit A	[59]	XM_002408156.1	MN879329
TBP	TATA-Box binding protein	[61] [58]	XM_002402081.1	MN879330
TUBB	β-tubulin	[58] [59]	GQ411364	MN728899
v-ATPase	vacuolar-type H[+]-ATPase	[59]	XM_002413524.1	MN728902

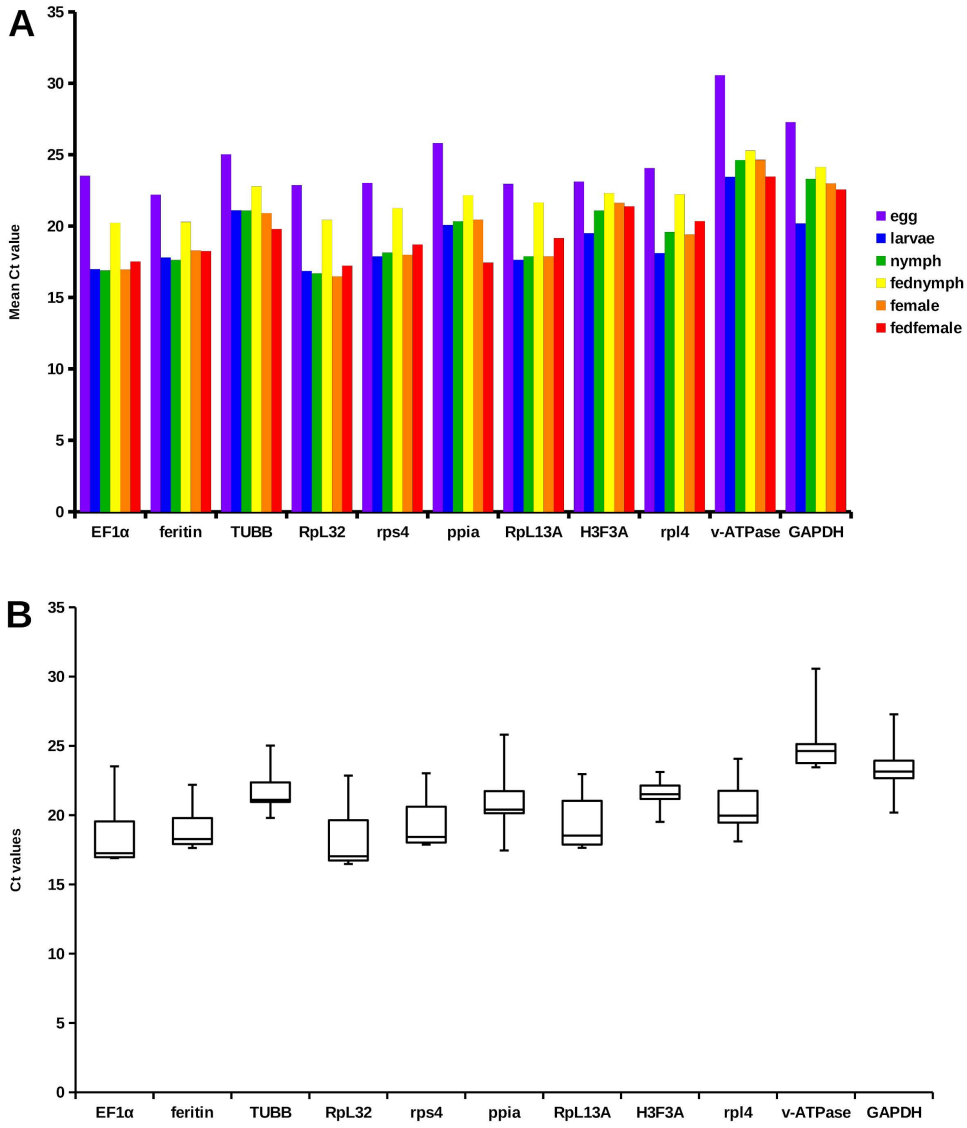


Fig. 1 Reference gene expression levels and stability in different life stages. (A) Comparison of abundances of reference gene transcripts in six *I. ricinus* life stages. Each bar represents the geometric mean of Ct values calculated from biological triplicate. (B) Transcript abundance variation of each reference gene calculated for six *I. ricinus* life stages. Individual box plot elements represent the following variables: line = median, boxes = upper and lower quartile, whiskers = range of the dataset.

The Ct values of each reference gene for each life stage and their biological replicates were integrated in BestKeeper v.1 [50]. The list of reference genes and their pairwise correlation coefficients (r) are listed in Table 3. All of the genes show a strong positive correlation (r) ranging between 0.85-0.99. Moreover, the first eight genes in Table 3 show a very strong positive correlation with r value above 0.9.

Table 3 Genes used in reference gene validation assay. Genes are ordered according to their pairwise correlation coefficient (r) in descending order.

rank	gene	Pairwise correlation coefficient (r)	p-value
1	*rps4*	0.98500	0.00100
2	*RpL32*	0.98333	0.00100
3	*EF1a*	0.98333	0.00105
4	*ferritin*	0.96467	0.00233
5	*rpl4*	0.95067	0.00447
6	*RpL13A*	0.94867	0.00523
7	*v-ATPase*	0.94133	0.00667
8	*TUBB*	0.93400	0.00939
9	*ppiA*	0.89333	0.01930
10	*GAPDH*	0.86467	0.03381
11	*H3F3A*	0.85867	0.04036

The pairwise correlation coefficient of all eleven tested reference genes indicates a high positive correlation. For normalization, we selected only the most stable reference genes as recommended previously [51,62–64].

Furthermore, we excluded *EF1a* gene from the calculation of HK factor for being a highly abundant transcript (8-12K cpm) (data not shown). Medium to highly abundant transcripts are more suitable reference genes due to their clear and reliable detection in every sample [65]; whereas highly abundant transcripts tend to express less stably and thus their selection as reference genes should be made with precaution [66].

Identification of transcripts specific for *Ixodes ricinus* life stages and Gene Ontology enrichment analysis

A Pearson correlation matrix assigning the transcripts to one or more of the transcriptome datasets was constructed. Of 25,872 transcripts, 10,266 were identified as life stage-specific. The remaining transcripts, identified in all assemblies, were classified as “housekeeping genes”, important throughout tick life stages. These transcripts were removed from further analysis as developmentally nonspecific (see “HK_transcripts” column in Additional file 4: Table_S4).

Additionally, transcripts supported with very low read mapping counts (counts per million (cpm) < 2) as by Pearson correlation analysis, were removed from individual libraries prior to the identification of stage-specific transcripts. Numbers of library-specific, as well as housekeeping transcripts upon low read mapping count filtering, are presented in Fig. 2.

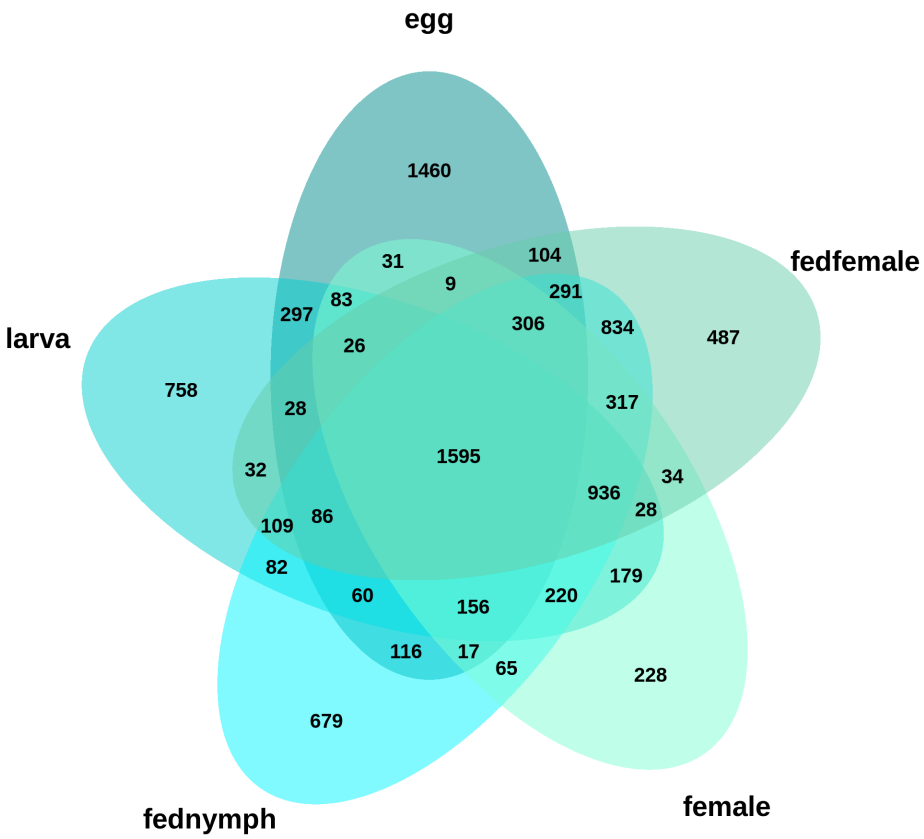


Fig. 2 Graphical illustration of transcripts that are present in different transcriptome libraries. The diagram also shows the number of "housekeeping transcripts" present in all transcriptomes in the middle intersections. The numbers in the outer intersections show transcripts that are common for more but not all libraries.

Note the high number of transcripts present in transcriptomes of both feeding stages and of the egg transcriptome assembly. Transcripts in each stage-specific library were organized according to the number of mapped reads in descending order (Additional file 4: Table_S4) and five of the transcripts that were assigned the most reads in each library (Top 5 transcripts) were selected for a detailed annotation and characterisation (Tab. 3). The presence of these transcripts in each life stage is illustrated in Fig. 3. It is evident that transcripts present in fed nymphs are specific for feeding stages and thus can be also found in fed females. Similarly, transcripts identified in

unfed females are also detectable in the transcriptome of unfed larval stages. The Top 5 transcripts in *I. ricinus* eggs, on the other hand, seem to be only present in the egg transcriptome as evident from the heatmap colour intensity in the remaining four libraries.

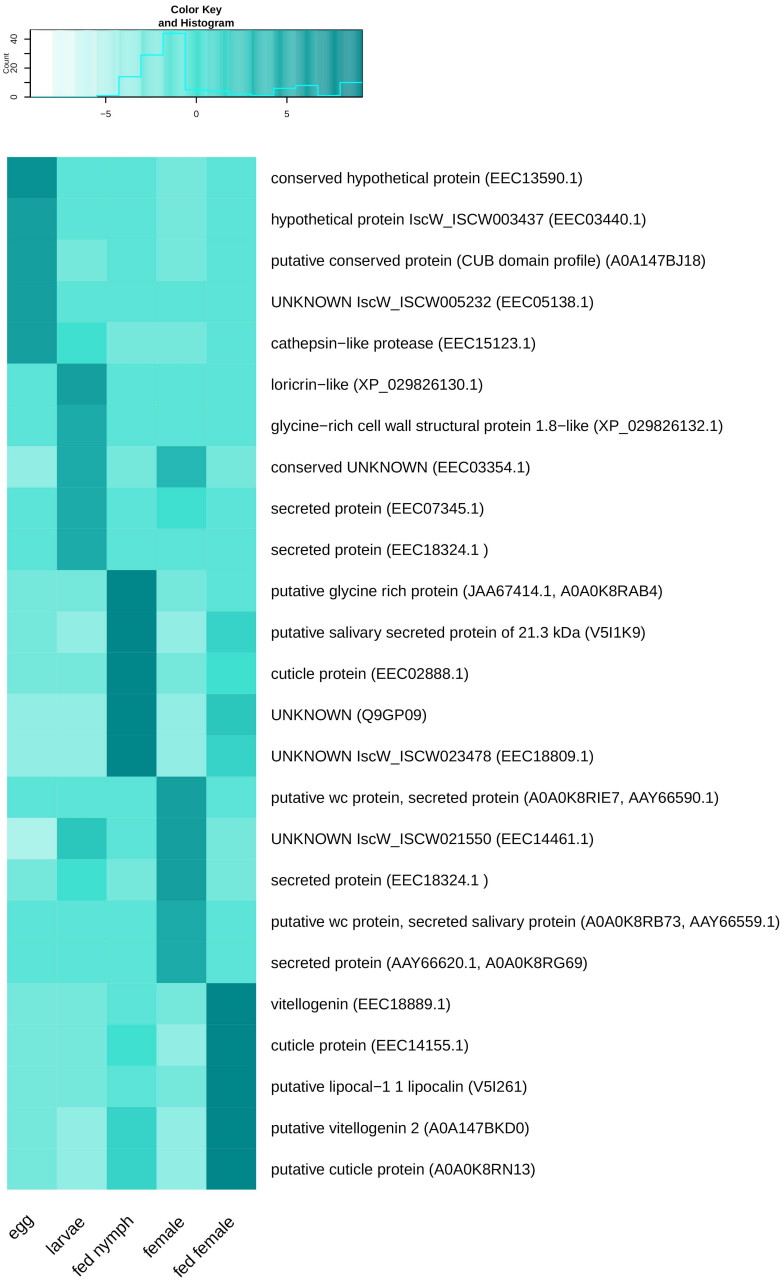


Fig. 3 The presence of Top 5 transcripts in the *I. ricinus* common assembly used for GO enrichment and annotation. The heatmap is clustered according to transcripts read mapping counts in cpm in each life stage in descending order. Transcript descriptions are supplied with an accession number of the best blast hit in brackets.

For the purposes of GO enrichment analysis and annotation, the matrix of read mapping counts of only the stage-specific transcripts was constructed and subjected to hierarchical clustering using complete linkage and 1-Pearson correlation metrics. The resulting dendrogram was cut to 2-50 clusters and their internal variability was calculated (see Methods). As per the elbow method, 10 clusters were the optimal number that accounted most efficiently for the total variability within the matrix (90.33 % explained). Fig. 4A presents the graphical illustration of the selected clusters of transcripts and shows their affiliation to particular stage-specific transcriptomes. Gene Ontology enrichment was determined for all transcript clusters, including an additional cluster of transcripts identified as “housekeeping genes”, which were removed prior clustering, and visualized as a categorized bubble plot (Fig. 4B). Of 10,266 stage-specific transcripts, 3,839 were assigned a GO term; altogether, there were 964 unique GO terms, of which 112 are categorized to cellular components, 391 to biological processes, and 451 to molecular function. Transcripts present in all transcriptome assemblies with assignable GO terms (n=7,443) were added for comparison (see cluster 11 of Fig. 4B). Figure 4B presents a detailed visualization of the enrichment of each GO term per stage and cluster.

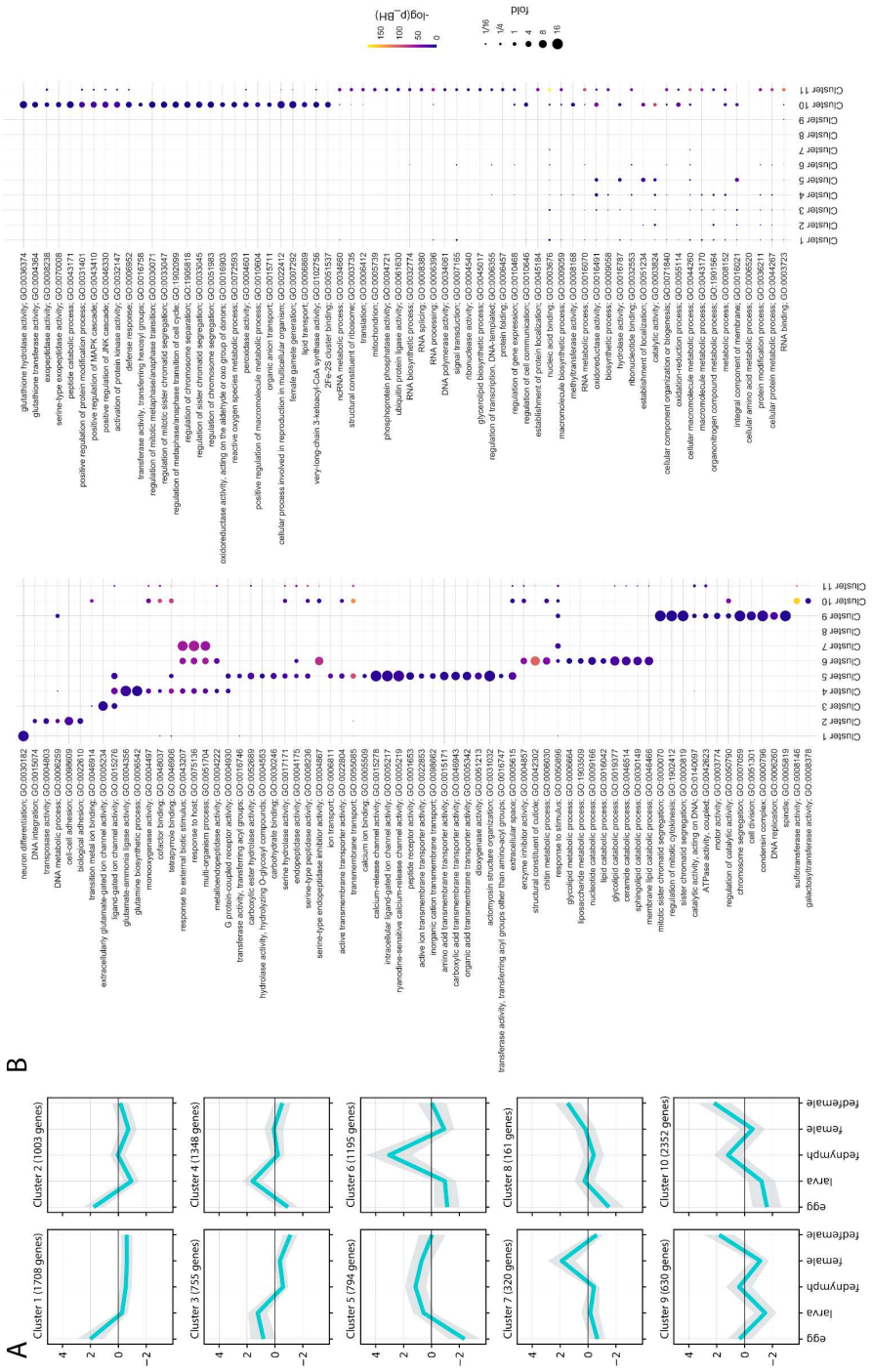


Fig. 4 Illustration of stage-specific transcripts present in ten clusters identified with hierarchical clustering and intra-cluster variability evaluation. (A) The distribution of stage-specific transcripts within 10 transcript clusters. Mean log2-scaled read mapping counts are plotted per cluster and life stage, shades represent standard deviation. (B) Gene ontology terms enrichment for ten stage-specific clusters, including the cluster of "housekeeping gene" transcripts in the common transcriptome assembly of the five tick stage-specific libraries (Cluster 11). The colour and size of data points represent -log10 value of Benjamini/Hochberg FDR-corrected p-value and fold enrichment compared to the whole set of transcripts, respectively (see graphical legends).

GO enrichment of transcript clusters in egg and larva life stages

Clusters 1 - 4 contain transcripts that were enriched only in egg and larva libraries. In particular, clusters 1 and 2 involve transcripts specific for egg development. Cluster 1 contains transcripts that are almost exclusively enriched in term GO:0030182 which entails the biological process "neuronal differentiation". Cluster 2 is the most enriched with terms describing processes associated with DNA regulation and processing, possibly involving transposon activity, and with cellular adhesion. Cluster 3 presents transcripts identified in both egg and larva assemblies, enriched with GO terms describing transmembrane signal transduction via ligand-gated ion channels (GO:0005234 and GO:0015276). Transcripts in cluster 4 were identified in larval stage only and the most enriched GO terms in this cluster describe cellular energetic metabolism involving glutamine as a main source of energy (GO:0004356 and GO:0006542) or perhaps glutamatergic neuronal activity, in line with glutamate ion channel enrichment in cluster 3.

The annotation of Top 5 transcripts in egg identified the transcript c76180_g1 which is highly homologous to cathepsin-like protease (EEC15123.1) that plays a major role in egg yolk degradation and is thus implicated in egg development [67]. Another transcript found in egg library (c80189_g5) is homologous to a conserved protein (A0A147BJ18) containing an extracellular CUB domain (abbreviation derived from the complement C1r/C1s, Uegf, Bmp1), which is common for extracellular or cell membrane-associated proteins that are involved in developmentally regulated processes [68].

The Top 5 transcripts identified in larval transcriptome are related to salivary gland cement proteins as demonstrated by transcript c73350_g2 highly similar to loricrin-like protein XP_029826130.1 identified in *I. scapularis* [69,70] and transcript sharing homology with cell wall components (c72525_g1 identical to XP_029826132.1). The production of salivary gland related proteins in unfed stages implies an increased production of saliva, which is a typical physiological feature of questing unfed stages preparing for an upcoming blood meal especially after a prolonged period of starvation [71,72].

Fed nymph library-specific enrichment

Transcripts in cluster 5 and 6 were identified in fed nymph transcriptome; Cluster 5 aggregates transcripts that are present in fed nymphs, larvae and females. The processes identified for transcripts in this cluster extend a wide range of functions, playing a role in very diverse cellular

functions and may underlie processes that have later developmental onset but are neglected in a feeding female where most effort and resources are put into blood meal processing and primarily into reproduction. The most dominant GO terms in this cluster describe cellular signalization via ligand-dependent calcium channels, transmembrane transporter activities (e.g. GO:0015278, GO:005217, GO:005219) or metabolic processes (e.g. GO:0030246, GO:0004175, GO:0017171) involving lipidic molecules and by extent cell membrane structure and integrity (e.g. GO:0052689, GO:0016746), consistent with the observation in larvae (see below). Cluster 6 presents transcripts enriched in fed nymph only. The most conspicuous activities associated with this cluster are involved in cuticle formation, catabolism of lipids, and enzymatic inhibition (e.g. GO:0042302, GO:0019377, GO:0004867). These activities are directly linked to tick feeding resulting in the fast growth of a tick cuticle-covered body which secures an efficient blood meal accommodation and also to blood digestion and its regulation respectively. Additionally, the signalling pathways responding to a foreign organism are also activated in feeding nymphs (e.g. GO:0075136, GO:0043207).

GO enrichment of unfed female

Cluster 7 comprises transcripts enriched in unfed females. The GO terms typical for this cluster describe cofactor/tetrapyrrole binding activity (GO:0048037, GO:0046906) and signalling pathways triggered by external biotic stimuli or host (GO:0043207, GO: 0075136, GO:0051704). The enrichment of these terms also occurred in larvae and fed nymphs, however, to a lower extent than in unfed females. Two of the Top 5 transcripts identified in the female library (c69267_g3 and c69267_g2) are homologous to salivary gland wc proteins identified previously in both *I. ricinus* and *I. scapularis* (accessions A0A0K8RIE7/AAZ66590.1 and A0A0K8RB73/AAY66559.1, respectively).

Fed female library-specific enrichment

Transcript clusters 8 and 9 are specific for the fed female life stage. Interestingly, no enrichment occurred in cluster 8, possibly as it is the smallest transcript cluster (161 contigs, i.e. 1.57% of transcripts assigned to different stage-specific libraries) and none of the class of transcripts was large enough to enable enrichment. Cluster 9 accommodates transcripts whose functions, as by enrichment, are nearly exclusively related to cell division processes including transcripts driving

mitosis, DNA replication, chromatid segregation, etc. (e.g. GO:0007059, GO:0006260, GO:0005819).

Feeding responsive transcript clusters

Cluster 10 includes transcripts that are typical for both, fed nymph and fed female, and extends the widest range of GO terms of all ten transcript clusters. The most enriched terms describe processes which feeding stages utilize for blood meal processing and regulation, including protease activity, lipid processing and transport, also comprising lipid metabolism-related to cell wall integrity and composition (e.g. GO:0070008, GO:0036374, GO:0102756, GO:0006869). The blood-feeding is also associated with activities that are required for processing of toxic products of blood meal degradation (GO:0004601, GO:0072593) and for the elimination of foreign molecules and potentially pathogens incoming with the blood meal, thus inducing immune defence pathways (GO:0006952). Similarly to cluster 9, the transcripts associated with mitosis, cell division, and related processes are enriched (e.g. GO:0030071, GO:0033047, GO:0033045). Also related are terms enriched for transcripts underlying factors regulating cell cycle (GO:0043410, GO:0046330). Another group of related GO terms describe transcripts influencing ovarian development, gametogenesis and reproduction (GO:0022412, GO:0007292). The Top 5 transcripts identified in both libraries of fed stages are mainly related to salivary gland-related functions triggered during feeding and cuticle protein synthesis to support blood meal accommodation within the tick. For example, transcript (c81210_g1) shows high similarity to the sequence (Q9GP09) described in a study of salivary gland factors and their expression induced by blood meal in *I. ricinus* [73]. Further characterisation of this transcript showed its high sequence identity (over 80 %) with a glycine-rich protein identified in *I. scapularis*, related to fibroin heavy chain protein (XP_029833024) involved in a build-up of cement which the tick uses for attachment of its hypostome to the host during feeding [70]. Similarly, the remaining annotated transcripts present in both libraries of feeding stages are of salivary gland origin or cuticle-related. Additionally, fed females of haematophagous parasites typically highly expressed transcripts related to reproduction. This corresponds to the presence of transcripts homologous to vitellogenins (c79317_g2 identical to EEC18889.1 and c80433_g1 identical to A0A147BKD0), expressed in egg yolk, that were found among the Top 5 transcripts in the fed female library.

Also present in fed female transcriptome is a transcript homologous to lipocalin 1, which is a protein typically produced in tick salivary glands and is a tick-specific evolutionary adaptation to blood-feeding [74].

The transcripts typical for both unfed larvae and females are mainly related to the processes involved in cell growth and cell cycle control or metabolism. An example is transcript c72929_g1 highly similar to the *I. scapularis* secreted protein (EEC07345) of unknown identity whose homologous transcript XP_029841102.1 underlies the production of a conserved cwc2 group pre-mRNA-splicing factor. Other two transcripts (c82554_g1 and c79150_g3, homologous to *I. scapularis* EEC03354 and EEC14461, respectively) of unknown function each contain an Acyltransferase 3 family domain and an NRF superfamily domain (related GO term GO:0016747). The activity of proteins containing both these domains is usually associated with lipid metabolism, processing, or transport and could be, by extent, also related to cell membrane structure and integrity and cell membrane functional components, which was also observed in the model organism *C. elegans* [75,76].

Discussion

Feeding-responsive transcripts

Four transcript clusters (6, 8, 9, 10) recovered from our analysis are specific for fed stages. Functional classification of most typical transcription was linked to cuticle formation and chitin metabolism as well as to the activation of blood meal processing enzymes, all in response to a blood meal. These findings were anticipated and are in line with previous observations [30,31,33,77].

Factors involved in mitosis and cell division are also commonly seen in feeding stages. Specific expression of these factors reflects the precocious cellular growth occurring shortly after a blood meal. Consistently, two GO terms (GO:0022412 and GO:0007292) involved in reproduction are enriched in both feeding stages. Their association with feeding females is fairly straightforward, underlying factors driving ovarian development and gamete generation. Enrichment of these transcripts in feeding nymphs is not self-evident; however, a lead to its functional elucidation in this non-adult stage can be found in thinly documented ecdysteroid hormonal regulation in ticks. The regulation of ecdysis is essential for the proper development of all arthropods, underlined by the existence of an organ dedicated to ecdysteroids production. In

chelicerates, ecdysteroidogenesis and its regulation is poorly understood but possibly restricted to ovarian tissue in both its mature and immature life stages as in soft tick *Ornithodoros moubata* [78]. This study suggests that factors driving reproduction, ovaries development, and ecdysis might be tightly regulated. The involvement of tick nymphal ovaria in processes decoupled from reproduction is demonstrated by their insensitivity to vitellogenin which plays a principal role in egg yolk formation and thus in gametogenesis and reproduction in arthropods [79,80]. Collectively, these findings along with an enrichment of reproductive transcripts in our fed nymph specific library strongly support the importance of ovaria for ecdysteroidogenesis in ticks, thus corroborating the enrichment of reproduction-related transcripts in feeding nymph. We suggest that transcripts functionally linked to reproduction in the library of fed nymph are in fact involved in the development of ovaria which clearly play a role in the regulation of moulting in other ticks and our findings are the first to describe this connection in *I. ricinus* tick.

Transcripts identified in unfed stages

The transcripts enriched in basic metabolic processes are present in larvae (glutamine-based metabolism; GO:0005234, GO:0004356, GO:0006542). Most of the related processes are also active in other life stages but their transcription is presumably overshadowed by processes more relevant for these stages. However, we suggest a linkage of some of these seemingly basic processes to specific functions in unfed tick stages. Perhaps, glutamate-ammonia ligase activity may point out GABA-ergic neuronal activity promoting larval motility [81]. Response to host or to external biotic stimulus is present in both females and larvae, though host-induced pathways were also observed for fed nymphs and thus might not be exclusive to unfed stages [82–85]. The presence of host response transcripts in larvae seems to be consistent with transcripts underlying glutamine metabolism of GABA-ergic neurons and host-seeking and was also observed in an unfed larva transcriptome library in other ticks [86].

Transcripts highly similar to *I. scapularis* secreted wc proteins were found prominent in unfed females (Table 4). This group of peptides, characterised by the presence of Trp-Cys dipeptide motif (thus “WC” proteins) at their C termini, are apparently specific to ticks and their function is yet to be elucidated [77,87,88].

Table 4 Highest-ranking transcripts of the read mapping counts matrix in each of the five stage-specific assemblies. Accession numbers of the best blast hits and their description are supplied. GO term and/or IPS annotations including their respective descriptions are also listed when available.

Contig ID	Best blast hit	Description
egg		
c77007_g1	**EEC13590.1**	**conserved hypothetical protein**
c80053_g1	**EEC03440.1**	**hypothetical protein IscW_ISCW003437**
c80189_g5	**A0A147BJ18** (IPR000859)	**putative conserved protein** (CUB domain profile)
c77756_g1	**EEC05138.1**	**UNKNOWN IscW_ISCW005232**
c76180_g1	**EEC15123.1**(GO:0006508; GO:0008234; IPR000169)	**cathepsin-like protease** (proteolysis; cysteine-type peptidase activity; Eukaryotic thiol (cysteine) proteases asparagine active site)
larva		
c73350_g2	**XP_029826130.1** (PR01217)	**loricrin-like (LOC115311581)** (Proline rich extensin signature)
c72525_g1	**XP_029826132.1** (PR01228)	**glycine-rich cell wall structural protein 1.8-like** (Eggshell protein signature)
c82554_g1	**EEC03354.1** (GO:0016747; IPR002656)	**conserved UNKNOWN** (transferase activity, transferring acyl groups other than amino-acyl groups; Acyltransferase family)
c72929_g1	**EEC07345.1**	**secreted protein**
c69133_g1	**EEC18324.1**	**secreted protein**
fed nymph		
c23277_g1	**JAA67414.1, A0A0K8RAB4** (PR01217)	**putative glycine rich protein** (Proline rich extensin signature)
c80070_g2	**V5I1K9**	**putative salivary secreted protein of 21.3 kDa (Fragment)**
c78019_g4	**EEC02888.1** (GO:0042302; IPR000618)	**cuticle protein** (structural constituent of cuticle; Chitin-binding type R&R domain profile. Insect cuticle protein)
c81210_g1	**Q9GP09**	**UNKNOWN**

c75846_g6	**EEC18809.1** (PR01217)	**UNKNOWN IscW_ISCW023478** (Proline rich extensin signature)
female		
c69267_g3	**A0A0K8RIE7 (AAY66590.1)**	**putative wc protein, secreted protein**
c79150_g3	**EEC14461.1**	**UNKNOWN IscW_ISCW021550**
c75091_g1	**EEC18324.1**	**secreted protein**
c69267_g2	**A0A0K8RB73 (AAY66559.1)**	**putative wc protein, secreted salivary protein**
c1565_g1	**AAY66620.1, A0A0K8RG69**	**secreted protein**
fed female		
c79317_g2	**EEC18889.1** (GO:0005319; GO:0006869; IPR001747; IPR001846)	**vitellogenin** (lipid transporter activity; lipid transport; VWFD domain profile.von Willebrand factor type D domain Vitellogenin domain profile)
c79647_g1	**EEC14155.1** (GO:0042302; IPR000618)	**cuticle protein** (structural constituent of cuticle; Chitin-binding type R&R domain profile. Insect cuticle protein)
c73798_g1	**V5I261** (GO:0030682; GO:0043176; IPR002970)	**putative lipocal-1 1 lipocalin (Fragment)** (evasion or tolerance of host defenses; amine binding; Tick histamine binding protein)
c80433_g1	**A0A147BKD0** (GO:0005319; GO:0006869; IPR001747; IPR001846)	**putative vitellogenin 2 (Fragment)** (lipid transporter activity; lipid transport; VWFD domain profile. von Willebrand factor type D domain Vitellogenin domain profile)
c74644_g1	**A0A0K8RN13** (GO:0042302; IPR000618)	**putative cuticle protein (Fragment)** (structural constituent of cuticle; Chitin-binding type R&R domain profile. Insect cuticle protein)

Transcription specific for egg development and tick embryogenesis

Despite numerous efforts in basic and applied tick research [89–93], few studies have concentrated on early stages of tick development [94]. A paucity of early-stage information

hinders the implementation of targeted approaches, such as RNA interference or characterisation of vaccine candidates [95,96]. Our study is among the first to provide a comprehensive and biologically relevant catalogue of transcripts for future research aiming at controlling the population of ticks in the early stages of their development.

Apart from an anticipated functional enrichment in neuronal development (GO:0030182), observed in the embryonic development of other arthropods as well [97], cell adhesion-related transcripts (GO:0098609), clearly important for embryonic morphogenesis, were also highly enriched [98].

Of interest was an enrichment of factors unleashing mobile DNA transposition and consequent DNA integration which are typically reactivated during embryonic development. An elevated expression of transposase associated with transposon activity is usually less regulated in order to enable an implementation of developmental regulation of host DNA during embryogenesis [99]. Transposition-related activities can thus be considered markers of tick early embryonic development.

An increased expression of cathepsin protease observed in the egg library was also anticipated as cathepsin proteases take part in egg yolk proteolysis and thus are crucial for the energetic metabolism during embryogenesis [67,96,100,101]. Consistently, cathepsins have been in the scope as targets for tick control [102,103]. CUB domain proteins were also found among egg specific transcripts and their expression in egg is most typically manifested in embryonic developmental factors [68]. CUB domains are found conserved in Metazoa and most often occur in cell surface proteins that mediate interactions of embryonal morphogenetic proteins and metalloproteases [104–106].

Reference gene validation assay

Reference gene validation using qRT-PCR is an essential step for proper quantitative data normalization. Currently, this is the most recognized methodology to provide reliable gene expression comparisons [107]. It is also a method of choice in cross-sample normalization of read count data or for the evaluation of normalization techniques in studies dealing with transcriptomic NGS data [53,108,109].

All genes tested in our study show a high degree of stability among tick life stages as by their correlation coefficient (r). Still, it is crucial to carefully select candidate reference genes

with respect to specific physiological circumstances given by the experimental design. Related to the tick development, specific examples can be found in haematophagous arthropods experiencing one or more blood meals in their life cycle. This represents a radical alteration of physiological conditions and profoundly affects the expression of many genes, including those that are considered fairly stable in a majority of other organisms as is the case of β-actin in the kissing bug *Rhodnius prolix* [110] or ribosomal components of *A. aegypti* during vitellogenesis triggered by blood meal [111]. In our study, several ribosomal genes (*RpL13A, RpL32, rpl4, rps4*) were tested and all of them proved highly stable. We ascribe such discordance to fundamental differences in life cycle among ticks and mosquitoes. In particular, the duration of blood-feeding is substantially different, taking approximately 2 minutes in the mosquito in comparison to several days in the tick [79,112], which certainly has implications to the initiation and progress of vitellogenesis [22,78,113] and hence the expression stability of ribosomal genes. Accordingly, inspection of partially fed stages may have also contributed to ribosomal transcript stability in this work. Thus, the set of reference genes, rather than just one of them, can be applied in future studies comparing gene expression between tick life stages. However, the robustness of our reference gene set can lower under different study designs, for instance monitoring gene expression under specific experimental conditions, tissue-specific expression, or time-lapse expression in time points of tick development different than here, such as fully fed stages. We, therefore, highlight the importance of *de novo* validation of selected reference genes for specific experimental designs.

Cataloguing a stage-specific transcription of *Ixodes ricinus*

The main mission of our project was to produce a comprehensive catalogue of transcripts of the tick *I. ricinus* and identify transcripts specific for developmental stages of interest by their cross-stage comparison. Using entire bodies for sequencing was thus of importance. This approach allowed us to collect all transcripts that are typical of each life stage. The main limitation here is the risk of losing information about very lowly abundant transcripts whose expression is either tissue-specific, or is restricted to a short period during development, or is induced by a specific physiological state. Sequencing of transcriptomes derived from individual tissues or time points, on the other hand, suffers from compositional bias which is manifested in the overrepresentation of functionally specific populations of transcripts, disregarding expression in other tissues or

developmental periods. This is evident from the identification of transcripts exclusive for RNA libraries originating in different life stages of *I. ricinus* salivary glands [31].

In order to detect lowly abundant transcripts, we performed sequencing of each library to a very high read depth using the HiSeq2000 Illumina platform. For each of our libraries, we received 72-120 M of reads (PE50). Additionally, we prepared our transcriptome assembly from reads combined from all five stage-specific libraries. This approach should secure an even more comprehensive catalogue of transcripts representing the overall coding potential of the *I. ricinus* genome. Comparison of transcript number in our assembly (25,872) with the coding capacity of *I. scapularis* (23,340) [56], the closest relative species with publicly available genome data, also provided an indication about the high completeness of our assembly.

We additionally confirmed a good recovery of lowly abundant transcripts in our assembly by searching for transcripts of genes whose expression was found to be tissue-specific and thus presumably very low in the population of all transcripts of the whole-body transcriptome. Based on previously published studies, we selected five genes whose expression had been found restricted to a specific tissue in the tick. Our blast search yielded 100 % of query transcripts, which again validates the high completeness of our *I. ricinus* assembly. Evaluation of another tissue-specific marker, however, revealed that the sequence of the putative Ixoderin B is 5' partial. Ixoderin B is a salivary gland-specific transcript in contrast to its A isoform which exhibits organism-wide transcription and whose corresponding transcript was fully assembled in our transcriptome. We admit limitations to our dataset in terms of both transcript presence and CDS completeness and encourage to inquire public databases to recover complete sequences of lowly expressed genes in the future. It is however very likely that the majority of lowly abundant transcripts we removed from the raw assembly of 83,534 sequences are truncated sequences which do not represent biologically relevant transcription of the interrogated life stage. Our dataset thus constitutes a representative transcriptome of *I. ricinus* across life stages and the removal of truncated low-quality transcripts does not negatively influence an estimation of stage-specific transcription.

Our study was designed so to cover the interpopulation expression variability of *I. ricinus* without losing the chance of recovery of lowly expressed transcripts or rare transcript isoforms underlying an interpopulation expression variability of this species. We thus pooled triplicated samples collected from ticks originating in different tick populations. This allowed us to produce

a high per sample read coverage without compromising the full assembly of rare transcripts. This was done on the expense of reducing the reliability of transcriptome quality evaluation and general significance of conclusions that we derive from our sequencing project. However, with a series of quality filtering and evaluation tests as for example the BUSCO assessment, we were able to present a list of indicators supporting high efficiency of our bioinformatic approach and good quality of our output data. The same experimental design was also presented in previous publications where a single library per sample was sequenced on the expense of high sequence coverage [31,33]. Using well-established bioinformatic pipelines along with a number of quality evaluation tests, the authors were able to construct reliable time- and tissue-dependent tick transcriptome assemblies supported by a single sequencing library per sample. This facilitated a postulation of firm conclusions based on well-organized analyses focused on the proper characterization of particular sample-specific assemblies, their reciprocal comparison, and an inference and functional annotation of sample-specific transcription.

With precaution, we performed *de novo* assembly instead of reference driven transcriptome reconstruction. Despite the existence of annotated reference genome of *I. scapularis*, a genome reference of related species may have obstructed assembly of *I. ricinus* sequences transcribed from highly interspecifically variable loci.

Our study was also organized to produce a representative catalogue of stage-specific transcripts that do not provide strong quantitative information about gene expression. Our transcriptome assembly was however submitted to a series of quality testing, normalization, and functional annotations. The bioinformatic pipeline that was eventually applied to our sequencing data was used upon thorough consultation of existing studies dealing with stage-specific transcription and with general bioinformatic guidelines [51,53,86,109]. Due to a unique design of our study, we organized our work by a combination of different approaches based on assumptions derived from the type and nature of our data and, at the same time, from biological questions postulated in our project. The resulting outputs provided valuable information about processes and functions typical for each life stage. Moreover, our transcriptome database represents a priceless pool of information for initiation of future research projects that can be built upon the mere knowledge about transcription in *I. ricinus* followed by the associated validation of transcripts sequences and their cross-referencing with public databases. Moreover,

the collection of transcripts specifically expressed in eggs can be exploited in the development of *in vitro* methods in tick-derived cell cultures originating in tick embryonic cells [38].

The descriptive nature of our study poses certain limits for an interpretation of our data in comparison with other quantitatively designed studies. On the other hand, putting our data into proper context, even within quantitative studies, can still present valid points for discussion with perspective for further confirmation by experimental approaches where required.

Conclusions

Our work presents and discusses new findings about developmentally specific transcripts identified in *I. ricinus* eggs, larvae, partially fed nymphs, females, and partially fed females. Proper normalization of our transcriptome assemblies was performed using a set of reference genes whose stability was verified by a quantitative gene validation essay of eleven candidate reference genes also presented in this study. Our data confirm the identity of transcripts involved in a tick feeding presented in previous studies and support the role of ovaries in ecdysteroidogenesis also in *I. ricinus* as was previously suggested in related tick species. Additionally, we describe processes and specific transcripts apparently important for embryogenesis, whereby most conspicuously energetic metabolism, developmental mobile DNA reactivation, and subsequent initiation of morphogenetic processes are crucial. Our study presents new insights into early and late tick development, consistently with previously published research and draws our findings into new biological contexts. As a whole, our work extends the collection of important information for further investigation in both basic and applied tick research, including the development of tick-targeted population control programs.

Additional files

Supplementary data submitted as separate files:

Additional file 1: Table_S1. Description of primers used in housekeeping gene validation assay.

Additional file 2: Table_S2. Summary of library mapping statistics and cross-sample normalization.

Additional file 3: Table_S3. Summary of *de novo* assembly quality statistics.

Additional file 4: Table_S4. List of stage-specific and housekeeping transcripts.

Additional file 5: Alignment of Cathepsin D1 (EF428204.1) query sequence and a corresponding transcript recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c79321_g3_i1).

Additional file 6: Alignment of Cathepsin D2 (HQ615697.1) query sequence and a corresponding transcript recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c81800_g1_i2).

Additional file 7: Alignment of Cathepsin D3 (HQ615698.1) query sequence and a corresponding transcript recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c81927_g1_i1).

Additional file 8: Alignment of Iris (AJ269658.2) query sequence and four corresponding transcripts representing Trinity assembler isoforms recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c83951_g1_i2, c83951_g1_i5, c83951_g1_i3, c83951_g1_i1).

Additional file 9: Alignment of Ixoderin A (AY341424.1) query sequence and a corresponding transcript recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c80994_g1_i1).

Additional file 10: Alignment of Ixoderin B (AY341424.1) query sequence and a truncated corresponding transcript recovered from *Ixodes ricinus* stage-specific transcriptome assembly (c82323_g13_i1).

Abbreviations

cpm: counts per million

geoSTDEV: geometric standard deviation

GO: Gene Ontology

MED: Median normalization

RLE: Relative Log Expression

RSEM: RNA-Seq by Expectation-Maximization

TBEV: Tick-borne encephalitis virus

TMM: Trimmed Mean of M-values

TMMwsp: Trimmed Mean of M-values with singleton pairing

UQ: Upper Quartile

Acknowledgements

We acknowledge computation resources provided by CERIT-SC and MetaCentrum, Brno, Czech Republic.

Declarations

Ethics approval and consent to participate

All animal experiments presented in this study were in accordance with the Animal Protection Law of the Czech Republic (§17, Act No. 246/1992 Sb) and with the approval of the Czech Academy of Sciences (approval no. 161/2010).

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. The Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession **GIDG00000000**. The version described in this paper is the first version, **GIDG01000000**.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Czech Science Foundation grants 18-27204S; by the Ministry of Education, Youth and Sports of the Czech Republic (projects LM2015055, LTARF 18021, and LTAUSA18040); and ERD Funds (the project CePaViP; CZ.02.1.01/0.0/0.0/16_019/0000759, Postdok_BIOGLOBE CZ.1.07/2.3.00/30.0032).

Authors' contributions

PV, JS and LG conceived the study. PV wrote the manuscript, organized the article, managed and produced tick samples used in this study, performed reference gene validation assay and

made a bioinformatic processing and evaluation of the tick transcriptomes. ZF supervised and corrected manuscript writing and performed annotation and GO enrichment analysis of the transcriptomic data. JE is in charge of a tick rearing facility and provided all the tick individuals used in this study. VB was in charge of library preparation and production of sequencing data. PV and RC were processing the sequencing data including quality evaluation, assembly and mapping. All authors read and approved the final manuscript.

References

1. Heyman P, Cochez C, Hofhuis A, Van Der Giessen J, Sprong H, Porter SR, et al. A clear and present danger: Tick-borne diseases in Europe. Expert Rev Anti Infect Ther. 2010;8:33–50.

2. Michelet L, Delannoy S, Devillers E, Umhang G, Aspan A, Juremalm M, et al. High-throughput screening of tick-borne pathogens in Europe. Front Cell Infect Microbiol. Frontiers Research Foundation; 2014;4.

3. Vayssier-Taussat M, Kazimirova M, Hubalek Z, Hornok S, Farkas R, Cosson JF, et al. Emerging horizons for tick-borne pathogens: From the “one pathogen-one disease” vision to the pathobiome paradigm. Future Microbiol. 2015. p. 2033–43.

4. Honig V, Svec P, Halas P, Vavruskova Z, Tykalova H, Kilian P, et al. Ticks and tick-borne pathogens in South Bohemia (Czech Republic) - Spatial variability in *Ixodes ricinus* abundance, Borrelia burgdorferi and tick-borne encephalitis virus prevalence. Ticks Tick Borne Dis. 2015;6:559–67.

5. Amicizia D, Domnich A, Panatto D, Lai PL, Cristina ML, Avio U, et al. Epidemiology of tick-borne encephalitis (TBE) in Europe and its prevention by available vaccines. Hum Vaccines Immunother. 2013;9:1163–71.

6. Lindgren E, Jaenson, Thomas, G. T. Lime borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures. Copenhagen: WHO Regional Office for Europe Copenhagen; 2006.

7. Heinz FX, Stiasny K, Holzmann H, Grgic-Vitek M, Kriz B, Essl A, et al. Vaccination and tick-borne encephalitis, central Europe. Emerg Infect Dis. 2013;19:69–76.

8. Bogovic P. Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management. World J Clin Cases; 2015;3:430.

9. Steffen R. Epidemiology of tick-borne encephalitis (TBE) in international travellers to Western/Central Europe and conclusions on vaccination recommendations. J. Travel Med. 2016;23(4):taw018.

10. vvan den Wijngaard CC, Hofhuis A, Simões M, Rood E, van Pelt W, Zeller H, et al. Surveillance perspective on Lyme borreliosis across the European Union and European Economic Area. Euro Surveill. 2017;22.

11. Sykes RA, Makiello P. An estimate of Lyme borreliosis incidence in Western Europe. J Public Heal. 2017;39:74–81.

12. Cerny V. Sezónní dynamika klíštěte Ixodes risinus v divokém místě zaklíštění. Československá Parazitol. 1957;4:57–8.

13. Roe RM, Donohue K V., Khalil SMS, Sonenshine DE. Hormonal regulation of metamorphosis and reproduction in ticks. Front. Biosci. 2008;13(18):7250–68.

14. Sonenshine DE, Roe RM, editors. Biology of Ticks Volume 1. second. New York: Oxford University Press (OUP); 2014.

15. Walker AR. Review of “Ticks: Biology, Disease and Control” by Alan Bowman & Patricia Nuttall (eds.). Parasit Vectors. Springer Nature; 2009;2:1.

16. Foldvari G. Life cycle and ecology of *Ixodes ricinus*: the roots of public health importance. In: Braks MAH, Van Wieren SE, Takken W and Sprong H (eds.) Ecology and prevention of Lyme borreliosis. Ecology and Control of Vector-borne diseases, Volume 4. Wageningen Academic Publishers, Wageningen, the Netherlands, pp. 31-40.

17. Balashov YS. Ixodid Ticks: Parasites and Vectors of Infections. St. Petersburg: Nauka; 1998.

18. Balashov YS. Bloodsucking ticks (Ixodoidea)-vectors of disease in man and animals. Misc. Publ. Entomol. Soc. Am. 1972.

19. Honzakova E, Olejnicek J, Cerny V, Daniel M, Dusbabek F. Relationship between number of eggs deposited and body weight of engorged *Ixodes ricinus* female. Folia Parasitol. 1975;22:37–42.

20. Gray JS, Kahl O, Lane RS, Levin ML, Tsao JI. Diapause in ticks of the medically important *Ixodes ricinus* species complex. Ticks Tick Borne Dis; 2016. p. 992–1003.

21. Dana AN, Hong YS, Kern MK, Hillenmeyer ME, Harker BW, Lobo NF, et al. Gene expression patterns associated with blood-feeding in the malaria mosquito *Anopheles gambiae*. BMC Genomics. 2005;6:5.

22. Reid WR, Zhang L, Liu N. Temporal gene expression profiles of pre blood-fed adult females immediately following eclosion in the southern house mosquito *Culex quinquefasciatus*. Int J Biol Sci. 2015;11:1306–13.

23. Thangamani S, Wikel SK. Differential expression of *Aedes aegypti* salivary transcriptome upon blood feeding. Parasites and Vectors. 2009;2:34.

24. Karim S, Singh P, Ribeiro JMC. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. PLoS One. 2011;6:e28525.

25. Schwarz A, Medrano-Mercado N, Schaub GA, Struchiner CJ, Bargues MD, Levy MZ, et al. An Updated Insight into the Sialotranscriptome of *Triatoma infestans*: Developmental Stage and Geographic Variations. PLoS Negl Trop Dis. 2014;8(12):e3372.

26. Bensaoud C, Nishiyama MY, Ben Hamda C, Lichtenstein F, Castro De Oliveira U, Faria F, et al. De novo assembly and annotation of *Hyalomma dromedarii* tick (Acari: Ixodidae) sialotranscriptome with regard to gender differences in gene expression. Parasites and Vectors. 2018;11:314.

27. Perner J, Provaznik J, Schrenkova J, Urbanova V, Ribeiro JMC, Kopacek P. RNA-seq analyses of the midgut from blood- and serum-fed *Ixodes ricinus* ticks. Sci Rep. 2016;6:36695.

28. Santiago PB, De Araújo CN, Motta FN, Praça YR, Charneau S, Bastos IMD, et al. Proteases of haematophagous arthropod vectors are involved in blood-feeding, yolk formation and immunity - a review. Parasites and Vectors. 2017;10:79.

29. Kotsyfakis M, Kopáček P, Franta Z, Pedra JHF, Ribeiro JMC. Deep Sequencing Analysis of the *Ixodes ricinus* Haemocytome. PLoS Negl Trop Dis. Public Library of Science; 2015;9(5):e0003754.

30. Charrier NP, Couton M, Voordouw MJ, Rais O, Durand-Hermouet A, Hervet C, et al. Whole body transcriptomes and new insights into the biology of the tick *Ixodes ricinus*. Parasit Vectors;11:364.

31. Schwarz A, Von Reumont BM, Erhart J, Chagas AC, Ribeiro JMC, Kotsyfakis M. De novo *Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. FASEB J. 2013;27:4745–56.

32. Schwarz A, Tenzer S, Hackenberg M, Erhart J, Gerhold-Ay A, Mazur J, et al. A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. Mol Cell Proteomics; 2014;13:2725–35.

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

34. Mans BJ, Andersen JF, Francischetti IMB, Valenzuela JG, Schwan TG, Pham VM, et al. Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behavior. Insect Biochem Mol Biol. 2008;38:42–58.

35. Esteves E, Maruyama SR, Kawahara R, Fujita A, Martins LA, Righi AA, et al. Analysis of the salivary gland transcriptome of unfed and partially fed *Amblyomma sculptum* ticks and descriptive proteome of the saliva. Front Cell Infect Microbiol. 2017;7:476.

36. Landulfo GA, Patané JSL, Silva DGN da, Junqueira-de-Azevedo ILM, Mendonca RZ, Simons SM, et al. Gut transcriptome analysis on females of Ornithodoros mimon (Acari: Argasidae) and phylogenetic inference of ticks. Rev Bras Parasitol Vet. 2017;26:185–204.

37. Araujo RN, Silva NCS, Mendes-Sousa A, Paim R, Costa GCA, Dias LR, et al. RNA-seq analysis of the salivary glands and midgut of the Argasid tick *Ornithodoros rostratus*. 2019;9:6764.

38. Bell-Sakyi L, Zweygarth E, Blouin EF, Gould EA, Jongejan F. Tick cell lines: tools for tick and tick-borne disease research. Trends Parasitol. 2007;23(9):450–7.

39. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.

40. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

42. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.

43. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research. 2017;6.

44. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

45. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47:W5–10.

46. Sojka D, Franta Z, Frantová H, Bartošová P, Horn M, Váchová J, et al. Characterization of gut-associated cathepsin D hemoglobinase from tick *Ixodes ricinus* (IrCD1). J Biol Chem. 2012;287:21152–63.

47. Leboulle G, Crippa M, Decrem Y, Mejri N, Brossard M, Bollen A, et al. Characterization of a novel salivary immunosuppressive protein from *Ixodes ricinus* ticks. J Biol Chem. 2002;277:10083–9.

48. Rego ROM, Hajdusek O, Kovar V, Kopacek P, Grubhoffer L, Hypsa V. Molecular cloning and comparative analysis of fibrinogen-related proteins from the soft tick *Ornithodoros moubata* and the hard tick *Ixodes ricinus*. Insect Biochem Mol Biol. 2005;35:991–1004.

49. Glaser, R. W. “Ringer" solutions and some notes on the physiological basis of their ionic composition. Comp Biochem Physiol. 1917;2:241–89.

50. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15.

51. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:RESEARCH0034.

52. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

53. Evans C, Hardin J, Stoebel DM. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform. 2018;19:776–92.

54. Maza E, Frasse P, Senin P, Bouzayen M, Zouine M. Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments: A matter of relative size of studied transcriptomes. Commun Integr Biol. 2013;6:e25849.

55. Klopfenstein D V., Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018;8:10872.

56. Pagel Van Zee J, Geraci NS, Guerrero FD, Wikel SK, Stuart JJ, Nene VM, et al. Tick genomics: The *Ixodes* genome project and beyond. Int J Parasitol. 2007;37:1297–305.

57. Rego ROM, Hajdusek O, Kovar V, Kopacek P, Grubhoffer L, Hypsa V. Molecular cloning and comparative analysis of fibrinogen-related proteins from the soft tick *Ornithodoros moubata* and the hard tick *Ixodes ricinus*. Insect Biochem Mol Biol. 2005;35:991–1004.

58. Nijhof AM, Balk JA, Postigo M, Jongejan F. Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus (Boophilus) microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. BMC Mol Biol. 2009;10:112.

59. Yang C, Pan H, Liu Y, Zhou X. Stably expressed housekeeping genes across developmental stages in the two-spotted spider mite, *Tetranychus urticae*. PLoS One. 2015;10:e0120833.

60. Hajdusek O, Almazán C, Loosova G, Villar M, Canales M, Grubhoffer L, et al. Characterization of *ferritin 2* for the control of tick infestations. Vaccine. 2010;28:2993–8.

61. Koci J, Simo L, Park Y. Validation of Internal Reference Genes for Real-Time Quantitative Polymerase Chain Reaction Studies in the Tick, *Ixodes scapularis* (Acari: Ixodidae). J Med Entomol. 2013;50:79–84.

62. Bionaz M, Loor JJ. Identification of reference genes for quantitative real-time PCR in the bovine mammary gland during the lactation cycle. Physiol Genomics. 2007;29:312–9.

63. Kaur R, Sodhi M, Sharma A, Sharma VL, Verma P, Swami SK, et al. Selection of suitable reference genes for normalization of quantitative RT-PCR (RT-qPCR) expression data across twelve tissues of riverine buffaloes (*Bubalus bubalis*). PLoS One. 2018;13:e0191558.

64. Razavi SA, Afsharpad M, Modarressi MH, Zarkesh M, Yaghmaei P, Nasiri S, et al. Validation of Reference Genes for Normalization of Relative qRT-PCR Studies in Papillary Thyroid Carcinoma. Sci Rep. 2019;9:15241.

65. Caracausi M, Piovesan A, Antonaros F, Strippoli P, Vitale L, Pelleri MC. Systematic identification of human housekeeping genes possibly useful as references in gene expression studies. Mol Med Rep. 2017;16:2397–410.

66. Chim SSC, Wong KKW, Chung CYL, Lam SKW, Kwok JSL, Lai CY, et al. Systematic selection of reference genes for the normalization of circulating RNA transcripts in pregnant women based on RNA-Seq data. Int J Mol Sci. 2017;18:E1709.

67. Sojka D, Francischetti IMB, Calvo E, Kotsyfakis M. Cysteine proteases from bloodfeeding arthropod ectoparasites. Adv Exp Med Biol. 2011;712:177–91.

68. Bork P, Beckmann G. The CUB domain: A widespread module in developmentally regulated proteins. J Mol Biol. 1993;231:539–45.

69. Francischetti IMB, Sa-Nunes A, Mans BJ, Santos IM, Ribeiro JMC. The role of saliva in tick feeding. Front Biosci. 2009;14:2051–88.

70. Suppan J, Engel B, Marchetti-Deschmann M, Nürnberger S. Tick attachment cement – reviewing the mysteries of a biological skin plug system. Biol Rev. 2018;93:1056–76.

71. Tirloni L, Kim TK, Berger M, Termignoni C, da Silva Vaz I, Mulenga A. *Amblyomma americanum* serpin 27 (AAS27) is a tick salivary anti-inflammatory protein secreted into the host during feeding. PLoS Negl Trop Dis. 2019;13:e0007660.

72. Rosendale AJ, Dunlevy ME, McCue MD, Benoit JB. Progressive behavioural, physiological and transcriptomic shifts over the course of prolonged starvation in ticks. 2019;28:49–65.

73. Leboulle G, Rochez C, Louahed J, Rutti B, Brossard M, Bollen A, et al. Isolation of *Ixodes ricinus* salivary gland mRNA encoding factors induced during blood feeding. Am J Trop Med Hyg. 2002;66:225–33.

74. Ganfornina MD, Kayser H, Sanchez D. Lipocalins in Arthropoda: Diversification and Functional Explorations. In: Akerstrom B, Borregaard N, Flower DR, Salier JP, editors. Lipocalins. Landes Bioscience; 2006. p. 49–74.

75. Choy R, Thomas JH. Fluoxetine-resistant mutants in *C. elegans* define a novel family of transmembrane proteins. Mol Cell. 1999;4:143–52.

76. Choy RKM, Kemner JM, Thomas JH. Fluoxetine-resistance genes in *Caenorhabditis elegans* function in the intestine and may act in drug transport. Genetics. 2006;172:885–92.

77. Chmelar J, Anderson JM, Mu J, Jochim RC, Valenzuela JG, Kopecky J. Insight into the sialome of the castor bean tick, *Ixodes ricinus*. BMC Genomics. 2008;9:233.

78. Ogihara MH, Hikiba J, Suzuki Y, Taylor DM, Kataoka H. Ovarian ecdysteroidogenesis in both immature and mature stages of an acari, *Ornithodoros moubata*. PLoS One. 2015;10:e0124953.

79. Richards MH. Vitellogenin and vitellogenin-like genes: not just for egg production. Insects Soc. 2019;66:505–6.

80. Taylor D, Chinzei Y, Ito K, Higuchi N, Ando K. Stimulation of Vitellogenesis by Pyrethroids in Mated and Virgin Female Adults, Male Adults, and Fourth Instar Females of *Ornithodoros moubata* (Acari: Argasidae). J Med Entomol. 1991;28:322–9.

81. Lumaret J-P, Errouissi F, Floate K, Rombke J, Wardhaugh K. A Review on the Toxicity and Non-Target Effects of Macrocyclic Lactones in Terrestrial and Aquatic Environments. Curr Pharm Biotechnol. 2012;13:1004–60.

82. Hodzic E, Fish D, Maretzki CM, De Silva AM, Feng S, Barthold SW. Acquisition and transmission of the agent of human granulocytic ehrlichiosis by *Ixodes scapularis* ticks. J Clin Microbiol. 1998;36:3574–8.

83. Kocan KM, de la Fuente J, Coburn LA. Insights into the development of *Ixodes scapularis*: a resource for research on a medically important tick species. Parasit Vectors. 2015;8:592.

84. Scott JD, Clark KL, Coble NM, Ballantyne TR. Detection and Transstadial Passage of Babesia Species and *Borrelia burgdorferi* Sensu Lato in Ticks Collected from Avian and Mammalian Hosts in Canada. Healthcare. 2019;7:155.

85. Liu XY, Bonnet SI. Hard tick factors implicated in pathogen transmission. PLoS Negl Trop Dis. Public Library of Science; 2014;8:e2566.

86. Villar M, Popara M, Ayllón N, De Fernández Mera IG, Mateos-Hernández L, Galindo RC, et al. A systems biology approach to the characterization of stress response in *Dermacentor reticulatus* tick unfed larvae. PLoS One. 2014;9(2):e89564.

87. Ribeiro JMC, Alarcon-Chaidez F, Ivo IM, Mans BJ, Mather TN, Valenzuela JG, et al. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. Insect Biochem Mol Biol. 2006;36:111–29.

88. Kim YH, Islam MS, You MJ. Proteomic screening of antigenic proteins from the hard tick, *Haemaphysalis longicornis* (Acari: Ixodidae). Korean J Parasitol. 2015;53:85–93.

89. Nuttall PA, Trimnell AR, Kazimirova M, Labuda M. Exposed and concealed antigens as vaccine targets for controlling ticks and tick-borne diseases. Parasite Immunol. 2006;28(4):155–63.

90. Labuda M, Trimnell AR, Ličková M, Kazimírová M, Davies GM, Lissina O, et al. An Antivector Vaccine Protects against a Lethal Vector-Borne Pathogen. PLoS Pathog. 2006;2(4):e27.

91. Merino O, Alberdi P, Pérez De La Lastra JM, de la Fuente J. Tick vaccines and the control of tick-borne pathogens. Front. Cell. Infect. Microbiol. 2013;3:30.

92. Sprong H, Trentelman J, Seemann I, Grubhoffer L, Rego ROM, Hajdušek O, et al. ANTIDotE: Anti-tick vaccines to prevent tick-borne diseases in Europe. Parasites and Vectors. 2014;7:77.

93. Rego ROM, Trentelman JJA, Anguita J, Nijhof AM, Sprong H, Klempa B, et al. Counterattacking the tick bite: Towards a rational design of anti-tick vaccines targeting pathogen transmission. Parasites and Vectors. 2019;12:229.

94. Santos VT, Ribeiro L, Fraga A, de Barros CM, Campos E, Moraes J, et al. The embryogenesis of the Tick *Rhipicephalus (Boophilus) microplus*: The establishment of a new chelicerate model system. Genesis. 2013;51:803–18.

95. Seixas A, Oliveira P, Termignoni C, Logullo C, Masuda A, da Silva Vaz I. *Rhipicephalus (Boophilus) microplus* embryo proteins as target for tick vaccine. Vet Immunol Immunopathol. 2012;148:149–56.

96. Hussein HE, Johnson WC, Taus NS, Suarez CE, Scoles GA, Ueti MW. Silencing expression of the *Rhipicephalus microplus* vitellogenin receptor gene blocks *Babesia bovis* transmission and interferes with oocyte maturation. Parasites and Vectors. 2019;12:7.

97. Artieri CG, Fraser HB. Transcript length mediates developmental timing of gene expression across drosophila. Mol Biol Evol. 2014;31:2879–89.

98. Shawky JH, Davidson LA. Tissue mechanics and adhesion during embryo development. Dev Biol. 2015;401(1):152–64.

99. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. 2018;19:199.

100. Estrela A, Seixas A, Termignoni C. A cysteine endopeptidase from tick (*Rhipicephalus (Boophilus) microplus*) larvae with vitellin digestion activity. Comp Biochem Physiol - B Biochem Mol Biol. 2007;148:410–6.

101. Zhang TT, Qiu ZX, Li Y, Wang WY, Li MM, Guo P, et al. The mRNA expression and enzymatic activity of three enzymes during embryonic development of the hard tick Haemaphysalis longicornis. Parasites and Vectors. 2019;12:96.

102. Seixas A, Leal AT, Nascimento-Silva MCL, Masuda A, Termignoni C, da Silva Vaz I. Vaccine potential of a tick vitellin-degrading enzyme (VTDCE). Vet Immunol Immunopathol. 2008;124:332–40.

103. Martins R, Ruiz N, Fonseca RN da, Vaz Junior I da S, Logullo C. The dynamics of energy metabolism in the tick embryo. Rev Bras Parasitol Vet. 2018;27:259–66.

104. Song JL, Wong JL, Wessel GM. Oogenesis: Single cell development and differentiation. Dev Biol. 2006;300:385–405.

105. Lee HX, Mendes FA, Plouhinec JL, De Robertis EM. Enzymatic regulation of pattern: BMP4 binds CUB domains of Tolloids and inhibits proteinase activity. Genes Dev. 2009;23:2551–62.

106. Nunes da Fonseca R, van der Zee M, Roth S. Evolution of extracellular Dpp modulators in insects: The roles of tolloid and twisted-gastrulation in dorsoventral patterning of the *Tribolium* embryo. Dev Biol. 2010;345:80–93.

107. Radonić A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun. 2004;313:856–62.

108. Zyprych-Walczak J, Szabelska A, Handschuh L, Górczak K, Klamecka K, Figlerowicz M, et al. The Impact of Normalization Methods on RNA-Seq Data Analysis. Biomed Res Int. Hindawi Limited; 2015;2015:621690.

109. Wang Z, Lyu Z, Pan L, Zeng G, Randhawa P. Defining housekeeping genes suitable for RNA-seq analysis of the human allograft kidney biopsy tissue. BMC Med Genomics. 2019;12:86.

110. Paim RM, Pereira MH, Di Ponzio R, Rodrigues JO, Guarneri AA, Gontijo NF, et al. Validation of reference genes for expression analysis in the salivary gland and the intestine of *Rhodnius prolixus* (Hemiptera, Reduviidae) under different experimental conditions by quantitative real-time PCR. BMC Res Notes. 2012;5:128.

111. Niu LL, Fallon AM. Differential regulation of ribosomal protein gene expression in *Aedes aegypti* mosquitoes before and after the blood meal. Insect Mol Biol. 2000;9:613–23.

112. Chadee DD, Beier JC, Mohammed RT. Fast and slow blood-feeding durations of *Aedes aegypti* mosquitoes in Trinidad. J Vector Ecol. 2002;27:172–7.

113. Valzania L, Mattee MT, Strand MR, Brown MR. Blood feeding activates the vitellogenic stage of oogenesis in the mosquito *Aedes aegypti* through inhibition of glycogen synthase kinase 3 by the insulin and TOR pathways. Dev Biol. Elsevier Inc.; 2019;454:85–95.

4 Conclusions and future perspectives

Whole-body transcriptome sequencing of *Ixodes ricinus* tick life stages was initiated with a perspective to support several lines of tick-directed research conducted in our laboratory. A reconstruction of an unbiased and comprehensive catalogue of *I. ricinus* transcriptome required a more thorough consideration. Most importantly, the whole bodies of all stages were used. This approach allowed us to obtain transcripts from all tick tissues in the used developmental stages, including those that are active only during tick feeding. Moreover, blood removal from feeding stages was not performed, as is done in most similar analyses in ticks, as the majority of haemocytes would be lost during tick dissection.

Leaving all blood in the sample brings about a few issues. Not only are blood-derived samples potent inhibitors of most in vitro enzymatic reactions, but they also represent a source of host mRNA that may consume a substantial portion of sequencing reads otherwise dedicated to tick transcripts. This may also result in reduced sequencing sensitivity as some very lowly abundant transcripts may hence escape detection. These difficulties are usually overcome by application of some well-established RNA isolation techniques which should efficiently prevent co-isolation of haem, the most prolific enzymatic inhibitor associated with blood-derived samples [270].

In order to detect the lowly abundant transcripts, a deep sequencing of each sample was also employed. Each sample was dedicated at least 70M 50 bp pair-end reads. All five datasets of stage-specific sequenced reads were further combined in order to reconstruct a common transcriptome assembly. This approach is usually a method of choice if an assembly of lowly abundant transcripts or isoforms is in question. Redistribution of transcripts to stage-specific libraries was secured by an established bioinformatic pipeline. The dataset was submitted to multiple quality check-ups and filtering to remove the host-derived transcripts as well as truncated and chimeric sequences carrying no biological significance. The resulting stage-specific transcriptomes were confronted with existing literature and their presence supported the relevance of our approach and outcoming results.

Our comprehensively validated transcriptome assembly was further used for a homology search of transcripts involved in methylation and glycosylation pathways in arthropods.

Based on the literature review of methylation in Eukaryotes presented in the Introduction chapter, it is evident that inference of extent, role, and significance of 5mC methylation for particular invertebrate lineages cannot be specified just based on the observations from the related taxa. Unlike in the vertebrates, cytosine DNA methylation in invertebrates apparently evolved into a mechanism implementing non-essential adaptive changes in response to changing environment by fine regulation of genome architecture, gene expression, and associated processes. DNA methylation in inquired invertebrates plays a subsidiary role and ultra-low levels or absence of 5mC along with loss of canonical METT genes from genomes in several invertebrate lineages support this hypothesis [100]. Conversely, methylation of adenine in RNA seems to be a widely used modification in all Eukaryotes and several lines of evidence ascribe this epigenetic modification a role in all phases of gene expression pathway [250]. We, therefore, investigated both 5mC and 6mA methylation in *I. ricinus* genome. Screening an extent of methylation in studied Chelicerates brought similarly inconsistent results as for remaining invertebrates [266]. Thus. the first probe into methylation pathway in *I. ricinus* was done by the HMM search of our *I. ricinus* transcriptome assembly based on HMM profile constructed from consensus alignments of conserved domains of DNMT1, DNMT3, and DAMT genes of a publicly available database of Arthropoda sequences. We recovered all three homologues of these enzymes in *I. ricinus* transcriptome and extended the list of Chelicerate species possessing the full complement of canonical adenine and cytosine DNA methylating enzymes. We additionally verified the presence of their products in tick using 5mC and 6mA antibodies and indirectly confirmed their activity in tick. These data, however, represent a first insight into the methylation pathway in *I. ricinus* tick and further research is required to provide a more direct evidence, as for example, measurement of METTs activity with specific substrates and a characterization of 6mA methylation also for mRNA substrates.

Additionally, a search for other important members of methylation pathway, such as DNMT2, UHRF1, TET, and MBD proteins has already been done and also these data are a part of the project extending our research focused on methylation in the *I. ricinus* tick.

Substantially more data is available regarding glycosylation in ticks. This important post-translational modification was proven important in many aspects of tick life-

cycle and the significance of carbohydrate-based interactions has already been implicated in a tick-host-pathogen interface.

The presence of many glycans, as well as lectins in ticks, has been proven experimentally. However, the origin of some glycan structures is, at least in the case of blood-feeding arthropods, more than speculative. Thus, the existence of the respective synthetic apparatus, transcriptionally active in tick, could elucidate the importance of these structures for the tick physiology and interaction with the host organism.

Lectins represent a large group of molecules with very diverse functions in all Eukaryotes. In invertebrates, and in particular in the parasitic ones, the major attention is drawn to a group of lectins involved in a humoral innate immunity, the only defence system playing an important role in recognition of invading pathogens.

Identification of these molecules in *I. ricinus* transcriptome represent a promising perspective that will advance our understanding of the character and extent of tick innate immunity. It will provide more details about the way the tick deals with loads of host-derived molecules acquired during blood meal as well as with pathogens that target the tick either as a final host or as a vector organism.

Investigation of glycosylation and methylation in the context of tick life-cycle and development represents one of the research directions which complement the common efforts of parasitologist community in tackling the threads to a public health that tick and tick-borne pathogens pose. The mechanism of methylation in tick has been so far largely unexplored and this study presents one of the first insights into the methylation pathway in *I. ricinus*. Glycosylation in tick has been investigated in larger details in tick-host-pathogen interface. However; the large complexity of this mechanism requires much more efforts to acquire at least a rough concept of a role that glycosylation plays in the tick relationship to a host and pathogen both being an integral part of its life-cycle. This study provides an overview of the carbohydrate-based interactions in tick-host-pathogen interface and also present new perspectives for studying a mechanism of carbohydrate-based interactions on the level of expression of underlying glycosylation machinery and carbohydrates recognizing proteins.

5 References

1. Alkishe AA, Peterson AT, Samy AM. Climate change influences on the potential geographic distribution of the disease vector tick *Ixodes ricinus*. Wooten RM, editor. PLoS One. 2017;12:e0189092.

2. Amicizia D, Domnich A, Panatto D, Lai PL, Cristina ML, Avio U, et al. Epidemiology of tick-borne encephalitis (TBE) in Europe and its prevention by available vaccines. Hum Vaccines Immunother. 2013;9:1163–71.

3. Lindgren E, Jaenson TGT. Lyme borreliosis in Europe. Influences of climate and climate change, epidemiology, ecology and adaptation measures. Copenhagen: WHO Regional Office for Europe; 2006.

4. Schwarz A, Von Reumont BM, Erhart J, Chagas AC, Ribeiro JMC, Kotsyfakis M. De novo *Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. FASEB J. 2013;27:4745–56.

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

6. Kotsyfakis M, Kopáček P, Franta Z, Pedra JHF, Ribeiro JMC. Deep Sequencing Analysis of the *Ixodes ricinus* Haemocytome. PLoS Negl Trop Dis. 2015;9:e0003754.

7. Cramaro WJ, Revets D, Hunewald OE, Sinner R, Reye AL, Muller CP. Integration of *Ixodes ricinus* genome sequencing with transcriptome and proteome annotation of the naïve midgut. BMC Genomics. 2015;16.

8. Perner J, Provazník J, Schrenková J, Urbanová V, Ribeiro JMC, Kopáček P. RNA-seq analyses of the midgut from blood- and serum-fed *Ixodes ricinus* ticks. Sci Rep. 2016;6:36695.

9. Perner J, Kropáčková S, Kopáček P, Ribeiro JMC. Sialome diversity of ticks revealed by RNAseq of single tick salivary glands. PLoS Negl Trop Dis. 2018;12:e0006410.

10. Charrier NP, Couton M, Voordouw MJ, Rais O, Durand-Hermouet A, Hervet C, et al. Whole body transcriptomes and new insights into the biology of the tick *Ixodes ricinus*. Parasit Vectors. 2018;11:364.

11. Marr AK, MacIsaac JL, Jiang R, Airo AM, Kobor MS, McMaster WR. Leishmania donovani Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages. Horn D, editor. PLoS Pathog. 2014;10:e1004419.

12. Vilcinskas A. The role of epigenetics in host–parasite coevolution: lessons from the model host insects *Galleria mellonella* and *Tribolium castaneum*. Zoology. 2016. p. 273–80.

13. Geyer KK, Niazi UH, Duval D, Cosseau C, Tomlinson C, Chalmers IW, et al. The *Biomphalaria glabrata* DNA methylation machinery displays spatial tissue expression, is differentially active in distinct snail populations and is modulated by interactions with *Schistosoma mansoni.* Pimenta PF, editor. PLoS Negl Trop Dis. 2017;11:e0005246.

14. Kumar S, Kim Y. An endoparasitoid wasp influences host DNA methylation. Sci Rep. Nature Publishing Group; 2017;7:1–16.

15. DiNardo AR, Nishiguchi T, Mace EM, Rajapakshe K, Mtetwa G, Kay A, et al. Schistosomiasis Induces Persistent DNA Methylation and Tuberculosis-Specific Immune Changes. J Immunol. 2018;201:124–33.

16. Mei XF, Shi W, Zhang YY, Zhu B, Wang YR, Hou LJ, et al. DNA methylation and hydroxymethylation profiles reveal possible role of highly methylated TLR signaling on *Fasciola gigantica* excretory/secretory products (FgESPs) modulation of buffalo dendritic cells. Parasites and Vectors. 2019;12:358.

17. Sbaraini N, Bellini R, Penteriche AB, Guedes RLM, Garcia AWA, Gerber AL, et al. Genome-wide DNA methylation analysis of *Metarhizium anisopliae* during tick mimicked infection condition. BMC Genomics. 2019;20:836.

18. Nardella F, Halby L, Hammam E, Erdmann D, Cadet-Daniel V, Peronet R, et al. DNA Methylation Bisubstrate Inhibitors Are Fast-Acting Drugs Active against Artemisinin-Resistant *Plasmodium falciparum* Parasites. ACS Cent Sci. 2020;6:16–21.

19. Poulin R, Thomas F. Epigenetic effects of infection on the phenotype of host offspring: Parasites reaching across host generations. Oikos. 2008. p. 331–5.

20. Gómez-Díaz E, Jordà M, Peinado MA, Rivero A. Epigenetics of Host–Pathogen Interactions: The Road Ahead and the Road Behind. Chitnis CE, editor. PLoS Pathog. 2012;8:e1003007.

21. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz-Renggli C, Maeda Y, et al. Surface sialic acids taken from the host allow trypanosome survival in Tsetse fly vectors. J Exp Med. 2004;199:1445–50.

22. Rubin-de-Celis SSC, Uemura H, Yoshida N, Schenkman S. Expression of trypomastigote trans-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole. Cell Microbiol. 2006;8:1888–98.

23. Lehr T, Geyer H, Maass K, Doenhoff MJ, Geyer R. Structural characterization of *N*-glycans from the freshwater snail *Biomphalaria glabrata* cross-reacting with *Schistosoma mansoni* glycoconjugates. Glycobiology. 2007;17:82–103.

24. Luk FCY, Johnson TM, Beckers CJ. *N*-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. Mol Biochem Parasitol. 2008;157:169–78.

25. Lujan AL, Croci DO, Gambarte Tudela JA, Losinno AD, Cagnoni AJ, Mariño K V., et al. Glycosylation-dependent galectin–receptor interactions promote *Chlamydia trachomatis* infection. Proc Natl Acad Sci U S A. 2018;115:E6000–9.

26. van Die I, Cummings RD. Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response? Glycobiology. 2010;20:2–12.

27. Prasanphanich NS, Mickum ML, Heimburg-Molinaro J, Cummings RD. Glycoconjugates in host-helminth interactions. Front Immunol. 2013;4:240.

28. Raman R, Tharakaraman K, Sasisekharan V, Sasisekharan R. Glycan–protein interactions in viral pathogenesis. Curr. Opin. Struct. 2016. p. 153–62.

29. Veríssimo CDM, Graeff-Teixeira C, Jones MK, Morassutti AL. Glycans in the roles of parasitological diagnosis and host-parasite interplay. Parasitology. 2019. p. 1217-32.

30. Watanabe Y, Bowden TA, Wilson IA, Crispin M. Exploitation of glycosylation in enveloped virus pathobiology. Biochim. Biophys. Acta - Gen. Subj. 2019. p. 1480–97.

31. Poole J, Day CJ, Von Itzstein M, Paton JC, Jennings MP. Glycointeractions in bacterial pathogenesis. Nat. Rev. Microbiol. 2018. p. 440–52.

32. Cohen M. Notable aspects of glycan-protein interactions. Biomolecules. 2015. p. 2056–72.

33. Kasuga T, Gijzen M. Epigenetics and the evolution of virulence. Trends Microbiol. 2013. p. 575–82.

34. Al Akeel R. Role of epigenetic reprogramming of host genes in bacterial pathogenesis. Saudi J. Biol. Sci. 2013. p. 305–9.

35. Mukherjee K, Dubovskiy I, Grizanova E, Lehmann R, Vilcinskas A. Epigenetic mechanisms mediate the experimental evolution of resistance against parasitic fungi in the greater wax moth *Galleria mellonella*. Sci Rep. 2019;9:1–11.

36. Varki A, Kornfeld S. Historical Background and Overview. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 3rd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009. p. 1–21.

37. Stanley P, Schachter H, Taniguchi N. N-glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials Glycobiol. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009. p. 1–14.

38. Marth JD. *O*-Glycans. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1999.

39. Nevo E. Evolution in action across life at “evolution canyons”, Israel. Trends Evol Biol. 2009;1:12–34.

40. Varki A, Lowe JB. Biological Roles of Glycans. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.

41. Roth Z, Yehezkel G, Khalaila I. Identification and Quantification of Protein Glycosylation. Stoytcheva M, editor. Int J Carbohydr Chem. 2012;2012:640923.

42. Lowe JB. Naturally Occurring Genetic Disorders of Glycosylation. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1999. p. 479–98.

43. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 2009;37.

44. Zuber C, Roth J. N-glycosylation. In: Gabius HJ, editor. sugar code Fundam Glycosci. Weinheim: Weinheim; 2009. p. 87–110.

45. Parodi AJ. *N*-glycosylation in trypanosomatid protozoa. Glycobiology. 1993;3:193–9.

46. Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, et al. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets glycosyltranferases. Proc Natl Acad Sci U S A. 2005;102:1548–53.

47. Wilson IBH, Paschinger K, Rendic D. Glycosylation of model and “model” organisms. In: Gabius HJ, editor. sugar code Fundam Glycosci. Weinheim: Willey – VCh; 2009. p. 139–54.

48. Nasab FP, Schulz BL, Gamarro F, Parodi AJ, Aebi M. All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. Mol Biol Cell. 2008;19:3758–68.

49. Hese K, Otto C, Routier FH, Lehle L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*. Glycobiology. 2009;19:160–71.

50. Karaoglu D, Kelleher DJ, Gilmore R. The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. J Biol Chem. 1997;272:32513–20.

51. Spirig U, Glavas M, Bodmer D, Reiss G, Burda P, Lippuner V, et al. The STT3 protein is a component of the yeast oligosaccharyltransferase complex. Mol Gen Genet. 1997;256:628–37.

52. Pathak R, Parker CS, Imperiali B. The essential yeast NLT1 gene encodes the 64 kDa glycoprotein subunit of the oligosaccharyl transferase. FEBS Lett. 1995;362:229–34.

53. Knauer R, Lehle L. The *N*-oligosaccharyltransferase complex from yeast. FEBS Lett. 1994;344:83–6.

54. Pathak R, Imperiali B. A dual affinity tag on the 64-kDa Nlt1p subunit allows the rapid characterization of mutant yeast oligosaccharyl transferase complexes. Arch Biochem Biophys. 1997;338:1–6.

55. Kelleher DJ, Gilmore R. The *Saccharomyces cerevisiae* oligosaccharyltransferase is a protein complex composed of Wbp1p, Swp1p, and four additional polypeptides. J Biol Chem. 1994;269:12908–17.

56. Knauer R, Lehle L. The oligosaccharyltransferase complex from Saccharomyces cerevisiae. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. J Biol Chem. 1999;274:17249–56.

57. Izquierdo L, Schulz BL, Rodrigues JA, Güther MLS, Procter JB, Barton GJ, et al. Distinct donor and acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. EMBO J. 2009;28:2650–61.

58. Kelleher DJ, Karaoglu D, Mandon EC, Gilmore R. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. Mol Cell. 2003;12:101–11.

59. Ruiz-Canada C, Kelleher DJ, Gilmore R. Cotranslational and Posttranslational *N*-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. Cell. 2009;136:272–83.

60. Mescher MF, Strominger JL. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*. J Biol Chem. 1976;251:2005–14.

61. Messner P. Bacterial glycoproteins. Glycoconj J. 1997;14:3–11.

62. Szymanski CM, Ruijin Y, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol. 1999;32:1022–30.

63. Dell A, Galadri A, Sastre F, Hitchen P. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. Int J Microbiol. 2010;2010:148178.

64. Mäki M, Renkonen R. Biosynthesis of 6-deoxyhexose glycans in bacteria. Glycobiology. 2004;14:1R-15R.

65. Varki A, Freeze HH, Gagneux P. Evolution of Glycan Diversity. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.

66. ten Hagen KG, Zhang L, Tian E, Zhang Y. Glycobiology on the fly: developmental and mechanistic insights from Drosophila. Glycobiology. 2009;19:102–11.

67. Tiemeyer M, Nakato H, Esko JD. Arthropoda. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 3rd ed. Cold Spring Harbor: 2018. p. 1–18.

68. Geisler C, Jarvis DL. Identification of genes encoding *N*-glycan processing β-*N*-acetylglucosaminidases in *Trichoplusia ni* and *Bombyx mori*: Implications for glycoengineering of baculovirus expression systems. Biotechnol Prog. 2010;26:34–44.

69. Cipollo JF, Costello CE, Hirschberg CB. The fine structure of *Caenorhabditis elegans N*-glycans. J Biol Chem. 2002;277:49143–57.

70. Yan S, Brecker L, Jin C, Titz A, Dragosits M, Karlsson NG, et al. Bisecting galactose as a feature of *N*-glycans of wild-type and mutant *Caenorhabditis elegans*. Mol Cell Proteomics. 2015;14:2111–25.

71. Yan S, Wilson IBH, Paschinger K. Comparison of RP-HPLC modes to analyse the *N*-glycome of the free-living nematode *Pristionchus pacificus.* Electrophoresis. 2015;36:1314–29.

72. Kim K, Lawrence SM, Park J, Pitts L, Vann WF, Betenbaugh MJ, et al. Expression of a functional *Drosophila melanogaster N*-acetylneuraminic acid (Neu5Ac) phosphate synthase gene: evidence for endogenous sialic acid biosynthetic ability in insects. Glycobiology. 2002;12:73–83.

73. Hollister JR, Jarvis DL. Engineering lepidopteran insect cells for sialoglycoprotein production by genetic transformation with mammalian beta 1,4-galactosyltransferase and alpha 2,6-sialyltransferase genes. Glycobiology. 2001;11:1–9.

74. Kurz S, Aoki K, Jin C, Karlsson NG, Tiemeyer M, Wilson IBH, et al. Targeted release and fractionation reveal glucuronylated and sulphated *N*- and *O*-glycans in larvae of dipteran insects. J Proteomics. 2015;126:172–88.

75. Paschinger K, Wilson IBH. Anionic and zwitterionic moieties as widespread glycan modifications in non-vertebrates. Glycoconj. J. 2020. p. 27–40.

76. Stanton R, Hykollari A, Eckmair B, Malzl D, Dragosits M, Palmberger D, et al. The underestimated *N*-glycomes of lepidopteran species. Biochim Biophys Acta - Gen Subj. 2017;1861:699–714.

77. Paschinger K, Wilson IBH. Analysis of zwitterionic and anionic N-linked glycans from invertebrates and protists by mass spectrometry. Glycoconj J. 2016;33:273–83.

78. Eckmair B, Jin C, Abed-Navandi D, Paschinger K. Multistep fractionation and mass spectrometry reveal zwitterionic and anionic modifications of the *N*- and *O*-glycans of a marine snail. Mol Cell Proteomics. 2016;15:573–97.

79. Hykollari A, Malzl D, Eckmair B, Vanbeselaere J, Scheidl P, Jin C, et al. Isomeric separation and recognition of anionic and zwitterionic *N*-glycans from royal jelly glycoproteins. Mol Cell Proteomics. 2018;17:2177–96.

80. Khoo KH, Huang HH, Lee KM. Characteristic structural features of schistosome cercarial *N*-glycans: expression of Lewis X and core xylosylation. Glycobiology. 2001;11:149–63.

81. Zielinska DF, Gnad F, Schropp K, Wiśniewski JR, Mann M. Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. Mol Cell. 2012;46:542–8.

82. Zhang Y, Kong D, Reichl L, Vogt N, Wolf F, Großhans J. The glucosyltransferase Xiantuan of the endoplasmic reticulum specifically affects E-Cadherin expression and is required for gastrulation movements in Drosophila. Dev Biol. 2014;390:208–20.

83. Zhu F, Li D, Chen K. Structures and functions of invertebrate glycosylation. Open Biol. 2019. p. 180232.

84. Walski T, De Schutter K, Van Damme EJM, Smagghe G. Diversity and functions of protein glycosylation in insects. Insect Biochem. Mol. Biol. 2017. p. 21–34.

85. Okajima T, Irvine KD. Regulation of Notch signaling by *O*-linked fucose. Cell. 2002. p. 893–904.

86. Sasamura T, Sasaki N, Miyashita F, Nakao S, Ishikawa HO, Ito M, et al. Neurotic, a novel maternal neurogenic gene, encodes an *O*-fucosyltransferase that is essential for Notch-Delta interactions. Development. 2003;130:4785–95.

87. Dominguez M, Ferres-Marco D, Gutierrez-Aviño FJ, Speicher SA, Beneyto M. Growth and specification of the eye are controlled independently by Eyegone and Eyeless in *Drosophila melanogaster*. Nat Genet. 2004;36:31–9.

88. Dönitz J, Schmitt-Engel C, Grossmann D, Gerischer L, Tech M, Schoppmeier M, et al. iBeetle-Base: a database for RNAi phenotypes in the red flour beetle *Tribolium castaneum*. Nucleic Acids Res. 2015;43:D720-5.

89. Harvey BM, Rana NA, Moss H, Leonardi J, Jafar-Nejad H, Haltiwanger RS. Mapping sites of *O*-glycosylation and fringe elongation on Drosophila Notch. J Biol Chem. 2016;291:16348–60.

90. Ishio A, Sasamura T, Ayukawa T, Kuroda J, Ishikawa HO, Aoyama N, et al. *O*-fucose monosaccharide of Drosophila Notch has a temperature-sensitive function and cooperates with *O*-glucose glycan in Notch transport and Notch signaling activation. J Biol Chem. 2015;290:505–19.

91. Nakamura N, Lyalin D, Panin VM. Protein *O*-mannosylation in animal development and physiology: From human disorders to Drosophila phenotypes. Semin. Cell Dev. 2010. p. 622–30.

92. Ichimiya T, Manya H, Ohmae Y, Yoshida H, Takahashi K, Ueda R, et al. The twisted abdomen phenotype of Drosophila POMT1 and POMT2 mutants coincides with their

heterophilic protein *O*-mannosyltransferase activity. J Biol Chem. 2004;279:42638-47.

93. Wairkar YP, Fradkin LG, Noordermeer JN, DiAntonio A. Synaptic defects in a Drosophila model of congenital muscular dystrophy. J Neurosci. 2008;28:3781–9.

94. Hart GW, Akimoto Y. The O-GlcNAc Modification. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials Glycobiol. 3rd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009. p. 2015–2017.

95. Hardivillé S, Hart GW. Nutrient regulation of signaling, transcription, and cell physiology by *O*-GlcNAcylation. Cell Metab. 2014. p. 208–13.

96. Rahman MM, Stuchlick O, El-Karim EG, Stuart R, Kipreos ET, Wells L. Intracellular protein glycosylation modulates insulin mediated lifespan in *C. elegans*. Aging (Albany NY). 2010;2:678–90.

97. Mariappa D, Selvan N, Borodkin VS, Alonso J, Ferenbach AT, Shepherd C, et al. A mutant *O*-GlcNAcase as a probe to reveal global dynamics of protein *O*-GlcNAcylation during Drosophila embryonic development. Biochem J. 2015;470:255–62.

98. Ingham PW. A gene that regulates the bithorax complex differentially in larval and adult cells of Drosophila. Cell. 1984;37:815–23.

99. Bestor TH. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 1990. p. 179–87.

100. Aliaga B, Bulla I, Mouahid G, Duval D, Grunau C. Universality of the DNA methylation codes in Eucaryotes. Sci Rep. 2019;9:1–11.

101. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38:1378–85.

102. Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, et al. Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas. PLoS Genet. 2015;11.

103. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010;107:8689–94.

104. Gim JA, Hong CP, Kim DS, Moon JW, Choi Y, Eo J, et al. Genome-wide analysis of DNA methylation before- and after exercise in the thoroughbred horse with MeDIP-seq. Mol Cells. 2015;38:210–20.

105. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA methylation across insects. Mol Biol Evol. 2017;34:654–65.

106. MacKay AB, Mhanni AA, McGowan RA, Krone PH. Immunological detection of changes in genomic DNA methylation during early zebrafish development. Genome. 2007;50:778–85.

107. Li Q, Li N, Hu X, Li J, Du Z, Chen L, et al. Genome-wide mapping of DNA methylation in chicken. PLoS One. 2011;6:e19428.

108. Schmitz RJ, He Y, Valdés-López O, Khan SM, Joshi T, Urich MA, et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res. 2013;23:1663–74.

109. Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol. 2013;31:154–9.

110. Gavery MR, Roberts SB. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). BMC Genomics. 2010;11:483.

111. Schellenbaum P, Mohler V, Wenzel G, Walter B. Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera L.*). BMC Plant Biol. 2008;8:78.

112. Patel C V., Gopinathan KP. Determination of trace amounts of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. Anal Biochem. 1987;164:164–9.

113. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010;328:916–9.

114. Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, et al. The genome of the fire ant *Solenopsis invicta.* Proc Natl Acad Sci U S A. 2011;108:5679-84.

115. Takata M, Kiyohara A, Takasu A, Kishima Y, Ohtsubo H, Sano Y. Rice transposable elements are characterized by various methylation environments in the genome. BMC Genomics. 2007;8:469.

116. Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, et al. Functional CpG methylation system in a social insect. Science. 2006;314:645–7.

117. Asselman J, De Coninck DIM, Pfrender ME, De Schamphelaere KAC. Gene Body Methylation Patterns in Daphnia Are Associated with Gene Family Size. Genome Biol Evol. 2016;8:1185–96.

118. Simmen MW, Leitgeb S, Charlton J, Jones SJM, Harris BR, Clark VH, et al. Nonmethylated transposable elements and methylated genes in a chordate genome. Science. 1999;283:1164–7.

119. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell. 2006;126:1189–201.

120. Gao F, Liu X, Wu XP, Wang XL, Gong D, Lu H, et al. Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. Genome Biol. 2012;13:R100.

121. Gowher H, Ehrlich KC, Jeltsch A. DNA from *Aspergillus flavus* contains 5-methylcytosine. FEMS Microbiol Lett. 2001;205:151–5.

122. Wang YL, Wang ZX, Liu C, Wang SB, Huang B. Genome-wide analysis of DNA methylation in the sexual stage of the insect pathogenic fungus *Cordyceps militaris*. Fungal Biol. 2015;119:1246–54.

123. Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol. 1997;17:1469–75.

124. Xu X, Li G, Li C, Zhang J, Wang Q, Simmons DK, et al. Evolutionary transition between invertebrates and vertebrates via methylation reprogramming in embryogenesis. Natl Sci Rev. 2019;6:993–1003.

125. Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola. Genome Biol Evol. 2018;10:1185–97.

126. Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R. Epigenetic regulation of the honey bee transcriptome: Unravelling the nature of methylated genes. BMC Genomics. 2009;10:472.

127. Glastad KM, Arsenault S V, Vertacnik KL, Geib SM, Kay S, Danforth BN, et al. Variation in DNA Methylation Is Not Consistently Reflected by Sociality in Hymenoptera. Genome Biol Evol. 2017;9:1687–98.

128. Glastad KM, Gokhale K, Liebig J, Goodisman MAD. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. Sci Rep. 2016;6:37110.

129. Jjingo D, Conley AB, Yi S V., Lunyak V V., King Jordan I. On the presence and role of human gene-body DNA methylation. Oncotarget. 2012;3:462–74.

130. Suzuki MM, Kerr ARW, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 2007;17:625–31.

131. Sarda S, Zeng J, Hunt BG, Yi S V. The evolution of invertebrate gene body methylation. Mol Biol Evol. 2012;29:1907–16.

132. Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol. 2010;28:516–20.

133. Antequera F. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 2003. p. 1647–58.

134. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.

135. Lou S, Lee HM, Qin H, Li JW, Gao Z, Liu X, et al. Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome Biol. 2014;15.

136. Mendizabal I, Yi S V. Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. Hum Mol Genet. 2016;25:69–82.

137. de Mendoza A, Lister R, Bogdanovic O. Evolution of DNA Methylome Diversity in Eukaryotes. J Mol Biol. 2019;S0022-2836(19)30659-X.

138. Keller TE, Han P, Yi S V. Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary. Mol Biol Evol. 2016;33:1019–28.

139. Olson CE, Roberts SB. Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes. Front Physiol. 2014;5 JUN:224.

140. Riviere G, Wu GC, Fellous A, Goux D, Sourdaine P, Favrel P. DNA Methylation Is Crucial for the Early Development in the Oyster *C. gigas*. Mar Biotechnol. 2013;15:739–53.

141. Saint-Carlier E, Riviere G. Regulation of Hox orthologues in the oyster *Crassostrea gigas* evidences a functional role for promoter DNA methylation in an invertebrate. FEBS Lett. 2015;589:1459–66.

142. Yang X, Han H, DeCarvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell. 2014;26:577–90.

143. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341:1237905.

144. Chodavarapu RK, Feng S, Bernatavichute Y V., Chen PY, Stroud H, Yu Y, et al. Relationship between nucleosome positioning and DNA methylation. Nature. 2010;466:388–92.

145. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479:74–9.

146. Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 1980;8:1499–504.

147. Simmen MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. Genomics. Academic Press; 2008;92:33–40.

148. Smallwood SA, Tomizawa SI, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43:811–4.

149. Gundersen G, Kolsto AB, Larsen F, Prydz H. Tissue-specific methylation of a CpG island in transgenic mice. Gene. 1992;113:207–14.

150. Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, Edinger M, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res. 2009;19:1165–74.

151. Takuno S, Gaut BS. Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. Mol Biol Evol. 2012;29:219–27.

152. Park J, Peng Z, Zeng J, Elango N, Park T, Wheeler D, et al. Comparative analyses of DNA methylation and sequence evolution using Nasonia genomes. Mol Biol Evol. 2011;28:3345–54.

153. Bird AP, Wolffe AP. Methylation-induced repression-belts, braces, and chromatin. Cell. 1999. p. 451–4.

154. Jones PL, Veenstra GJC, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet. 1998;19:187–91.

155. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 1998;393:386–9.

156. Phillips JE, Corces VG. CTCF: Master Weaver of the Genome. Cell. 2009. p. 1194–211.

157. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013. p. 23–38.

158. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448:714–7.

159. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010;38:4246–53.

160. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.

161. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Res. 2010;20:320–31.

162. Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nat Struct Mol Biol. 2009;16:990–5.

163. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 2012. p. 484–92.

164. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. RNA. 2018;24:1351–62.

165. Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ, Amdam G V. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics. 2012;13:480.

166. Song X, Huang F, Liu J, Li C, Gao S, Wu W, et al. Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. DNA Res. 2017;24:445–57.

167. Song K, Li L, Zhang G. The association between DNA methylation and exon expression in the Pacific oyster *Crassostrea gigas*. PLoS One. 2017;12.

168. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. PLoS Biol. 2010;8.

169. Wang X, Wheeler D, Avery A, Rago A, Choi JH, Colbourne JK, et al. Function and Evolution of DNA Methylation in *Nasonia vitripennis*. PLoS Genet. 2013;9:e1003872.

170. Cunningham CB, Ji L, Wiberg RAW, Shelton J, McKinney EC, Parker DJ, et al. The Genome and Methylome of a Beetle with Complex Social Behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). Genome Biol Evol. 2015;7:3383–96.

171. Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L, et al. Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. Genome Res. 2014;24:821–30.

172. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, Dsouza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010;466:253–7.

173. Illingworth RS, Bird AP. CpG islands - “A rough guide.” FEBS Lett. 2009. p. 1713–20.

174. Wang X, Li Q, Lian J, Li L, Jin L, Cai H, et al. Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. BMC Genomics. 2014;15:1119.

175. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001;292:2077–80.

176. Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T. An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. EMBO J. 2009;28:1078–86.

177. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 2000;97:5237–42.

178. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. Nat. Rev. Genet. 2013. p. 204–20.

179. Patil V, Ward RL, Hesson LB. The evidence for functional non-CpG methylation in mammalian cells. Epigenetics. 2014. p. 823–8.

180. Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, et al. Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. PLoS Genet. 2011;7:e1002389.

181. Barrès R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, et al. Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density. Cell Metab. 2009;10:189–98.

182. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;471:68–73.

183. Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature. 2014;511:177-83.

184. Field LM, Lyko F, Mandrioli M, Prantera G. DNA methylation in insects. Insect Mol. Biol. 2004. p. 109–15.

185. Mandrioli M. Epigenetic tinkering and evolution: Is there any continuity in the role of cytosine methylation from invertebrates to vertebrates? Cell. Mol. Life Sci. 2004. p. 2425–7.

186. Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, et al. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc Natl Acad Sci U S A. 2015;112:13970–5.

187. Ponger L, Li W-H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. Mol Biol Evol. 2005;22:1119–28.

188. Maresca A, Zaffagnini M, Caporali L, Carelli V, Zanna C. DNA methyltransferase 1 mutations and mitochondrial pathology: Is mtDNA methylated? Front Genet. 2015;5:90.

189. Albalat R, Martí-Solans J, Cañestro C. DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. Brief Funct Genomics. 2012;11:142-55.

190. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001. p. 1089–93.

191. Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun. 2013;4.

192. Liu J, Hu H, Panserat S, Marandel L. Evolutionary history of DNA methylation related genes in chordates: new insights from multiple whole genome duplications. Sci Rep. 2020;10:970.

193. Yarychkivska O, Shahabuddin Z, Comfort N, Boulard M, Bestor TH. BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation *in vivo*. J Biol Chem. 2018;293:19466–75.

194. Jurkowski TP, Jeltsch A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One. 2011;6:e28104.

195. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science. 2006;311:395–8.

196. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev. 2010;24:1590–5.

197. Becker M, Müller S, Nellen W, Jurkowski TP, Jeltsch A, Ehrenhofer-Murray AE. Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. Nucleic Acids Res. 2012;40:11648–58.

198. Müller S, Windhof IM, Maximov V, Jurkowski T, Jeltsch A, Förstner KU, et al. Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). Nucleic Acids Res. 2013;41:8615–27.

199. Li S, Du J, Yang H, Yin J, Ding J, Zhong J. Functional and structural characterization of DNMT2 from *Spodoptera frugiperda*. J Mol Cell Biol. 2013;5:64–6.

200. Raddatz G, Guzzardo PM, Olova N, Fantappié MR, Rampp M, Schaefer M, et al. Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc Natl Acad Sci U S A. 2013;110:8627–31.

201. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. Anal Chem. 2014;86:3697–702.

202. Rai K, Chidester S, Zavala C V., Manos EJ, James SR, Karpf AR, et al. Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. Genes Dev. 2007;21:261–6.

203. Jeltsch A. Molecular enzymology of mammalian DNA methyltransferases. Curr Top Microbiol Immunol. 2006. p. 203–25.

204. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol. 2012;19:900–5.

205. Shanmugam R, Fierer J, Kaiser S, Helm M, Jurkowski TP, Jeltsch A. Cytosine methylation of tRNA-Asp by DNMT2 has a role in translation of proteins containing poly-Asp sequences. Cell Discov. 2015;1:15010.

206. Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G. Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat Genet. 2009;41:696–702.

207. Dorer DR, Henikoff S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell. 1994;77:993–1002.

208. Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, et al. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. RNA Biol. 2017. p. 1108–23.

209. Jurkowski TP, Meusburger M, Phalke S, Helm M, Nellen W, Reuter G, et al. Human DNMT2 methylates tRNAAsp molecules using a DNA methyltransferase-like catalytic mechanism. RNA. 2008;14:1663–70.

210. Shen L, Gao G, Zhang Y, Zhang H, Ye Z, Huang S, et al. A single amino acid substitution confers enhanced methylation activity of mammalian Dnmt3b on chromatin DNA. Nucleic Acids Res. 2010;38:6054–64.

211. Jeltsch A, Jurkowska RZ. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic Acids Res. 2016;44:8556-75.

212. Huff JT, Zilberman D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell. 2014;156:1286–97.

213. Rošić S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T, et al. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat Genet. 2018;50:452-9.

214. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013;152:1146–59.

215. Hung MS, Shen CKJ. Eukaryotic methyl-CpG-binding domain proteins and chromatin modification. Eukaryot. Cell. 2003. p. 841–6.

216. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: Readers of the epigenome. Epigenomics. 2015. p. 1051–73.

217. Kalkhoven E, Teunissen H, Houweling A, Verrijzer CP, Zantema A. The PHD Type Zinc Finger Is an Integral Part of the CBP Acetyltransferase Domain. Mol Cell Biol. 2002;22:1961–70.

218. Zhou Y, Santoro R, Grummt I. The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. EMBO J. 2002;21:4632–40.

219. Wu X, Zhang Y. TET-mediated active DNA demethylation: Mechanism, function and beyond. Nat. Rev. Genet. 2017. p. 517–34.

220. Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenöder S, et al. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS One. 2011;6:e16627.

221. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Mol Cell. 2011;42:451–64.

222. Zhao H, Chen T. Tet family of 5-methylcytosine dioxygenases in mammalian development. J. Hum. Genet. 2013. p. 421–7.

223. Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep. 2012. p. 28–35.

224. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015;161:893–906.

225. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016;351:282–5.

226. Ismail JN, Badini S, Frey F, Abou-Kheir W, Shirinian M. Drosophila Tet Is Expressed in Midline Glia and Is Required for Proper Axonal Development. Front Cell Neurosci. 2019;13:252.

227. Pastor WA, Aravind L, Rao A. TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. Nat. Rev. Mol. Cell Biol. 2013. p. 341–56.

228. Clancy MJ. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res. 2002;30:4509–18.

229. Hongay CF, Orr-Weaver TL. Drosophila inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Natl Acad Sci U S A. 2011;108:14855–60.

230. Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, et al. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell. 2008;20:1278–88.

231. Bokar JA. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In: Grosjean H, editor. Fine-tuning RNA Funct by Modif Ed. Berlin: Springer; 2005. p. 141–7.

232. O'Brown ZK, Greer EL. N6-methyladenine: A conserved and dynamic DNA mark. Adv Exp Med Biol. 2016;945:213–46.

233. Hu CW, Chen JL, Hsu YW, Yen CC, Chao MR. Trace analysis of methylated and hydroxymethylated cytosines in DNA by isotope-dilution LC-MS/MS: First evidence of DNA methylation in *Caenorhabditis elegans*. Biochem J. 2015;465:39–47.

234. Simpson VJ, Johnson TE, Hammen RF. *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res. 1986;14:6711–9.

235. Wenzel D, Palladino F, Jedrusik-Bode M. Epigenetics in *C. elegans:* Facts and challenges. Genesis. 2011. p. 647–61.

236. Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizábal-Corrales D, et al. DNA methylation on N6-adenine in *C. elegans.* Cell. 2015;161:868–78.

237. Lyko F, Maleszka R. Insects as innovative models for functional studies of DNA methylation. Trends Genet. 2011. p. 127–31.

238. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol. 2014;10:93–5.

239. Li Z, Zhao P, Xia Q. Epigeneticmethylations on n6-adenine and n6-adenosine with the same input but different output. Int. J. Mol. Sci. 2019;20(12):E2931.

240. Wang X, Li Z, Zhang Q, Li B, Lu C, Li W, et al. DNA methylation on N6-adenine in lepidopteran *Bombyx mori.* Biochim Biophys Acta - Gene Regul Mech. 2018;1861:815–25.

241. Kweon SM, Chen Y, Moon E, Kvederaviciutė K, Klimasauskas S, Feldman DE. An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing. Mol Cell. 2019;74:1138-1147.e6.

242. Koziol MJ, Bradshaw CR, Allen GE, Costa ASH, Frezza C, Gurdon JB. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat Struct Mol Biol. 2016;23:24–30.

243. Liu J, Zhu Y, Luo GZ, Wang X, Yue Y, Wang X, et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat Commun. 2016;7:13052.

244. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature. 2016;532:329–33.

245. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell. 2012;149:1635–46.

246. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012;485:201–6.

247. Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AAF, Kol N, Salmon-Divon M, et al. Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science. 2015;347:1002–6.

248. Schwartz S, Mumbach MR, Jovanovic M, Wang T, Maciag K, Bushkin GG, et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep. 2014;8:284–96.

249. Sommer S, Lavi U, Darnell JE. The absolute frequency of labeled *N*-6-methyladenosine in HeLa cell messenger RNA decreases with label time. J Mol Biol. 1978;124:487–99.

250. Zaccara S, Ries RJ, Jaffrey SR. Reading, writing and erasing mRNA methylation. Nat. Rev. Mol. Cell Biol. 2019. p. 608–24.

251. Kan L, Grozhik A V., Vedanayagam J, Patil DP, Pang N, Lim KS, et al. The m 6 A pathway facilitates sex determination in Drosophila. Nat Commun. 2017;8:15737.

252. Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, et al. M6 A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. Nature. 2016;540:301–4.

253. Lence T, Akhtar J, Bayer M, Schmid K, Spindler L, Ho CH, et al. M6A modulates neuronal functions and sex determination in Drosophila. Nature. 2016;540:242–7.

254. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature. 2015;518:560–4.

255. Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vågbø CB, Geula S, et al. m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. 2017;31:990–1006.

256. Louloupi A, Ntini E, Conrad T, Ørom UAV. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. Cell Rep. 2018;23:3429–37.

257. Ma H, Wang X, Cai J, Dai Q, Natchiar SK, Lv R, et al. N 6-Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. Nat Chem Biol. 2019;15:88–94.

258. van Tran N, Ernst FGM, Hawley BR, Zorbas C, Ulryck N, Hackert P, et al. The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. Nucleic Acids Res. 2019;47:7719–33.

259. Pendleton KE, Chen B, Liu K, Hunter O V., Xie Y, Tu BP, et al. The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell. 2017;169:824-835.e14.

260. Huang J, Yin P. Structural Insights into N6-methyladenosine (m6A) Modification in the Transcriptome. Genomics, Proteomics Bioinforma. 2018. p. 85–98.

261. Yang SX, Guo C, Zhang YK, Sun JT, Hong XY. Expression level and immunolocalization of de novo methyltransferase 3 protein (TuDNMT3) in adult females and males of the two-spotted spider mite, *Tetranychus urticae.* Exp Appl Acarol. 2015;67:381–92.

262. Strachecka A, Borsuk G, Olszewski K, Paleolog J. A new detection method for a newly revealed mechanism of pyrethroid resistance development in *Varroa destructor*. Parasitol Res 2015;114(11):3999–4004.

263. Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite *Metaseiulus occidentalis* Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biol Evol. 2016;8:1762–75.

264. Yang SX, Guo C, Zhao XT, Sun JT, Hong XY. Divergent methylation pattern in adult stage between two forms of *Tetranychus urticae* (Acari: Tetranychidae). Insect Sci. 2018;25:667–78.

265. Mondal M, Klimov P, Flynt AS. Rewired RNAi-mediated genome surveillance in house dust mites. PLoS Genet. 2018;14.

266. Liu S, Aagaard A, Bechsgaard J, Bilde T. DNA methylation patterns in the social spider, *Stegodyphus dumicola*. Genes (Basel). 2019;10.

267. Vagin V V., Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. Science. 2006;313:320–4.

268. Sarkies P, Selkirk ME, Jones JT, Blok V, Boothby T, Goldstein B, et al. Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. Hurst LD, editor. PLOS Biol. 2015;13:e1002061.

269. Geyer KK, Chalmers IW, MacKintosh N, Hirst JE, Geoghegan R, Badets M, et al. Cytosine methylation is a conserved epigenetic feature found throughout the phylum Platyhelminthes. BMC Genomics. 2013;14:462.

270. Schwochow D, Serieys LEK, Wayne RK, Thalmann O. Efficient recovery of whole blood RNA - a comparison of commercial RNA extraction protocols for high-throughput applications in wildlife species. BMC Biotechnol. 2012;12:33.

6 List of abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5 Methylcytosine
6mA	6-methyladenine
ADD	DNMT3L histone H3 binding domain
CAZy	Carbohydrate-Active enZYmes
CXXC	Bromo-adjacent homology
DAMT-1	N6-adenine methyl-transferase
DMAD	Adenine DNA demethylase
DMAP	DNA methyltransferase associated protein
DNMT	DNA methyltransferase
DDT	DNA binding homeobox and Different Transcription factor domain
Dol-PP	Dolichol pyrophosphate
ER	Endoplasmic Reticulum
Fuc	Fucose
GA	Golgi apparatus
GK repeats	Lysine-glycine rich stretch
GlcNAc	N-acetylglucosamine
GT	Glycosyltransferase
JBP	J-binding protein
Man	Mannose
MBD	Methyl-CpG binding domain
METT	Methyltransferase
OGA	O-GlcNAcylase
GlcNAc	GlcNAc-transferase
OST	Oligosaccharyltranferase
PHD	Plant homeodomain
PreSET	N-terminal to some SET domain
PWWP	Pro-Trp-Trp-Pro domain
RFTS	Replication foci targeting sequence
RNAPII	RNA polymerase II
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax domain

TBE	Tick-borne encephalitis
TE	Transposable element
TET	Ten-eleven translocation
TSA	Transcriptome Shotgun Assembly
UHRF1	E3 ubiquitin-protein ligase

7 Curriculum vitae

RNDr. Pavlína Věchtová

e-mail: vechtpova@prf.jcu.cz, vechtova@paru.cas.cz

Date and place of birth: 17. 12. 1984, Pardubice, Czech Republic

Education:

2008: Bsc. in Biologie, Faculty of Science, University of South Bohemia in České Budějovice v Českých Budějovicích, Czech Republic

2011: Msc. in Experimental Biology, Faculty of Science, University of South Bohemia in České Budějovice v Českých Budějovicích, Czech Republic

2011: RNDr. in Experimental Biology, Faculty of Science, University of South Bohemia in České Budějovice v Českých Budějovicích, Czech Republic

2011-: PhD studies in Molecular and cell biology and genetics Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic

Work experience:

2012-: R&D scientist, Institute of Parasitology, Biology Centre CAS

2016-: Research assistant, Institute of Chemistry, Faculty of Science, University of South Bohemia in České Budějovice v Českých Budějovicích, Czech Republic

Research internship:

2015: (1 month): GeneCore facility, EMBL, Heidelberg, Germany

Teaching activities:

Faculty of Science JU:

Biochemistry practicals (laboratory courses; in Czech)

Advanced Biochemistry practicals (laboratory courses; in Czech)

Biochemistry laboratory I (laboratory courses; in English)

Biochemistry laboratory II (laboratory courses; in English)

Pharmacology and Toxicology (laboratory courses; in Czech)

Introduction to Toxicology (laboratory courses; in Czech)

Xenobiochemistry and Toxicology (laboratory courses; in English)

Forensic genetics (laboratory courses; in Czech)
Experimental methods (laboratory courses; in Czech)

Attended courses:

2012 Akreditovaný kvalifikační kurz Odborné zdravotnické laboratorní metody; Osvědčení o získání odborné způsobilosti k výkonu zdravotnického povolání pod odborným dohledem (Praha, Czech Republic; 31. 1. - 15. 5. 2012)

2013 Bioinformatické vyhodnocování pyrosekvenačních dat (České Budějovice, Czech Republic; 04. - 07. 6. 2013)

2014 Workshop on Genomics (Český Krumlov, Czech Republic; 12. - 24. 1. 2014)

2015 EMBL Advanced Course: Whole Transcriptome Data Analysis (Heidelberg, Germany; 29. 9. - 01. - 10. 2015)

2016 Osvědčení o odborné způsobilosti k navrhování pokusů a projektů pokusů podle §15d odst. 3 zákona č. 246/1992 Sb., na ochranu zvířat proti týrání, ve znění pozdějších předpisů (Brno, Czech Republic; 15. 11. 2016)

Grant applications:

2009-2011: Identification of genes for STT3 subunit of oligosaccharyltransferase in tick species and analysis of their expression – Grant Agency of the University of South Bohemia in České Budějovice, Czech Republic – GA JU – Principal Investigator

Publications with impact factor:

Zurovcova M., Havelka J., Stary P., Vechtova P., Chundelova D., Jarosova A., Kucerova L., 2010. "DNA barcoding" is of limited value for identifying adelgids (Hemiptera: *Adelgidae*) but supports traditional morphological taxonomy. *European Journal of Entomology* 107(2): 147-156. doi: 10.14411/eje.2010.020

Vechtova P., Dalíkova M., Sykorová M., Zurovcova M., Fussy Z., Zrzava M. 2016. CpSAT-1, a transcribed satellite sequence from the codling moth, *Cydia pomonella*. *Genetica* 144(4): 385-95. doi: 10.1007/s10709-016-9907-0

Vechtova, P., Sterbova, J., Sterba, J., Vancova, M., Rego, R.O.M., Selinger, M., Strnad, M., Golovchenko, M., Rudenko, N., Grubhoffer, L. 2018. A bite so sweet: the glycobiology interface of tick-host-pathogen interactions. *Parasites & Vectors* 11(1):594. doi: 10.1186/s13071-018-3062-7

Loginov, D., Loginova Y., Dycka, F., Böttinger, K., Vechtova, P., Sterba, J. 2019. Tissue-specific signatures in tick cell line MS profiles. *Parasites & Vectors* 12:212. doi: 10.1186/s13071-019-3460-5

Selinger, M., Tykalova, H., Sterba, J., Vechtova, P., Vavruskova, Z, Lieskovska, J., Kohl, A., Schnettler, E., Grubhoffer, L., 2019. Tick-borne encephalitis virus downregulates rRNA synthesis and host protein production in human cells of neuronal origin. *PloS Neglected Tropical Diseases* 13(9):e0007745. doi: 10.1371/journal.pntd.0007745

Xin, M., Vechtova, P., Shaliutina-Kolesova, A., Fussy, Z., Loginov, D., Dzyuba, B., Linhart, O., Boryshpolets, S., Rodina, M., Li, P., Loginova, Y., Sterba, J. 2019. Transferrin Identification in Sterlet (*Acipenser ruthenus*) Reproductive System. *Animals (Basel)* 9(10):753. doi: 10.3390/ani9100753.

Kotsarenko, K., Vechtova, P., Hammerova, Z., Langova, N., Malinovska, L., Wimmerova, M., Sterba, J., Grubhoffer, L., 2019. Newly identified DNA methyltransferases of *Ixodes ricinus* ticks. *Ticks and Tick Borne Diseases* 11(2):101348. doi: 10.1016/j.ttbdis.2019.101348

D. Loginov, K. Böttinger, Y. Loginova, F. Dycka, P. Vechtova, J. Sterba. 2020. Biotyping of TBEV-infected IRE/CTVM19 tick cell line. *Ticks and Tick-borne Diseases.* in press. 10.1016/j.ttbdis.2020.101420

Number of publications in WoS: 9
Citations without self-citations (6. 3. 2020): 25
H-index: 2