

# Positive Selection in Tick Saliva Proteins of the Salp15 Family

Petra Catalina Schwalie · Jörg Schultz

Received: 15 September 2008 / Accepted: 9 December 2008 / Published online: 22 January 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** When taking their blood meal on the mammalian host, ticks transfer a multitude of different proteins from their saliva into the host. Some of these proteins are hijacked by pathogens for their own purposes. *Borrelia burgdorferi*, the Lyme disease agent, is critically dependent on the presence of the tick protein Salp15 when infecting the host. Similarly, *Anaplasma phagocytophilum*, which causes anaplasmosis, needs Salp16, a homologue of Salp15, to get transferred from the host into the tick. Here we analyzed whether adaptive evolution has shaped the Salp15 protein family. Using site-specific estimates of  $K_A/K_S$  ratios, we identified different positions within the Salp15 protein family which have undergone a phase of positive selection. Additionally, we analyzed the *B. burgdorferi* protein interacting with Salp15, OspC. Again, sites showing signs of positive selection were identified, although they are more likely a result of the antigenic features of OspC than of the influence of Salp15. The identification of probably functionally relevant sites in the Salp15 family might direct the detailed experimental analysis of their interaction with human and bacterial proteins.

**Keywords** Outer surface protein C · *Borrelia burgdorferi* · *Ixodes scapularis* · Salp16

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-008-9194-1) contains supplementary material, which is available to authorized users.

P. C. Schwalie · J. Schultz (✉)  
Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany  
e-mail: joerg.schultz@biozentrum.uni-wuerzburg.de

## Introduction

Ticks are ectoparasites feeding for several days on the blood of their host while engorging from about 3 mm up to, in extreme cases, 4 cm. While taking their blood meal, the arthropods introduce saliva facilitating their persistence on the host. Among other functions, the saliva contains immunosuppressive, anticomplement, and antihemostatic activity (Gillespie et al. 2000; Valenzuela et al. 2000). Proteins within the saliva are of interest to research not only for a deeper understanding of the tick infection, but also as a basis for the development of vaccines. Indeed a tick immunity was described in 1939 (Trager 1939).

In addition to feeding on its blood, ticks also transfer different pathogens into the host. This includes *Borrelia burgdorferi*, the Lyme disease agent, *Anaplasma phagocytophilum*, which causes anaplasmosis, and pathogens responsible for tick-borne encephalitis. Some of these pathogens have hijacked tick saliva proteins for their own purposes. In the case of *B. burgdorferi*, the tick protein Salp15 is crucial for infection. This protein binds both the outer surface protein C (OspC) of *B. burgdorferi* and the CD4 receptor of a mammalian host. The latter interaction inhibits CD4+ cell activation by repressing calcium fluxes triggered by T cell receptor ligation (Anguita et al. 2002; Motameni et al. 2004). Furthermore, Salp15 binds to the C-type lectin DC-SIGN in dendritic cells and inhibits the pro-inflammatory cytokine production suppressing the T-cell-stimulatory role of dendritic cells (Hovius et al. 2008). Together, these effects seem to shield *B. burgdorferi* from the host's immune system (Schuijt et al. 2008). Similarly, *A. phagocytophilum* utilizes the tick protein Salp16 for transmission from the mammalian host into the tick (Foley and Nieto 2007). Both Salp15 and Salp16 belong to the same protein family, although they exhibit different functions: the

first one enables the transfer of a pathogen from the tick into the host, while the latter facilitates transmission from the host into the tick. Here we set out to analyze whether this family has undergone a phase of adaptive evolution and to identify sites of importance for the molecular function of these proteins.

## Materials and Methods

### Sequence Data Collection

The amino acid sequence of Salp15 from *Ixodes scapularis* (gi:15428294) (Das et al. 2001) was used to search for related *Ixodes scapularis* salivary gland proteins (PSI-BLAST at NCBI against nr until convergence; e-value cutoff, 0.005; <http://www.ncbi.nih.gov/BLAST>). OspC protein sequences from *Borrelia burgdorferi* were available in the NCBI Protein Database. Amino acid sequences were aligned using MUSCLE (Edgar 2004) and the resulting alignment was manually optimized. The corresponding DNA sequences were collected by accession number at NCBI (<http://www.ncbi.nih.gov/>) and the protein alignment was translated in cDNA-alignment using Pal2Nal (Suyama et al. 2006). To improve the highly divergent *Ixodes scapularis* multiple alignment, we submitted the initial sequences to Gblocks (Castresana 2000), and building on the conserved blocks obtained in the output, we created the new alignment Salp15\_cut (Initial alignment positions 22–96, 123–148, and 204–226 were removed), as well as Salp15\_gBlocks, where only the conserved blocks were maintained (100% in Gblocks output).

### Site-Specific $K_A/K_S$ Analysis

Phylogenetic trees were built using a maximum likelihood method (proml) of the PHYLIP package 3.66 (Felsenstein 2005) using the Jones-Taylor-Thornton model of amino acid changes. Site-specific  $K_A/K_S$  analysis was carried out using codeml from the PAML package (v. 3.15) (Yang 1997). Codeml uses maximum likelihood to predict sites in a group of cDNA sequences that have been subject for positive selection. Simple models with sites of a  $K_A/K_S$  or  $\omega$  ratio between 0 and 1 were compared with more complex models that allowed for ratios  $>1$ . Log likelihood values (l) were calculated for each model by maximum likelihood. This enabled a likelihood ratio test (LRT) to provide a statistically significant comparison between a simple model and a more complex one. If the complex model indicated an estimated  $\omega$  ratio  $>1$ , and the test statistic ( $2\Delta l$ ) was greater than the critical values of the chi-square distribution with the appropriate degree of freedom (Yang 1998), positive selection was inferred. An empirical

Bayes (EB) approach was used to calculate the posterior probability that each site was from a particular site class, and sites with high posterior probabilities coming from the class with  $\omega > 1$  (with  $p > 0.90$ ) were inferred to be under positive selection. This made it possible to detect positive selection and to identify sites under positive selection even if the average  $\omega$  ratio over all sites was  $\ll 1$ . Three LRTs were used to test for positive selection. The first test compared M1a and M2a; the second, M7 and M8; and the third, the null hypothesis M8a ( $\beta\omega_s = 1$ ) and M8 (Swanson et al. 2003; Wong et al. 2004). First, we tested the LRT for the null model M1a (NearlyNeutral), which assumes two site classes in proportions  $p_0$  and  $p_1 = 1 - p_0$  with  $0 < \omega_0 < 1$  and  $\omega_1 = 1$ , and the alternative model M2a (PositiveSelection), which adds a proportion  $p_2$  of sites with  $\omega_2 > 1$  estimated from the data. Those are slight modifications of models M1 (neutral) and M2 (selection), which had  $\omega_0 = 0$  fixed (Nielsen and Yang 1998). The second LRT compares the null model M7 ( $\beta$ ), which assumes a beta distribution for  $\omega$  (in the interval  $0 < \omega < 1$ ), and the alternative model M8 ( $\beta\omega$ ), which adds an extra class of sites with positive selection ( $\omega_s > 1$ ). The degrees of freedom used for the chi-square in these first two tests was  $df = 2$ . Then M8a was specified using NSsites = 8, fix\_omega = 1, omega = 1. The critical values for the M8a-M8 comparison that we used were 2.71 at 5% and 5.41 at 1%.

### Protein Structure Analysis

The amino acid sequence of the crystallized OspC from *Borrelia burgdorferi* (PDB access code 1GGQ [Eicken et al. 2001]) was aligned with the analyzed OspC sequences using MUSCLE. PyMOL (<http://pymol.sourceforge.net/>) was used for structural manipulations and to generate images. Alignment positions under positive selection with  $p > 0.90$  were mapped onto the protein structure.

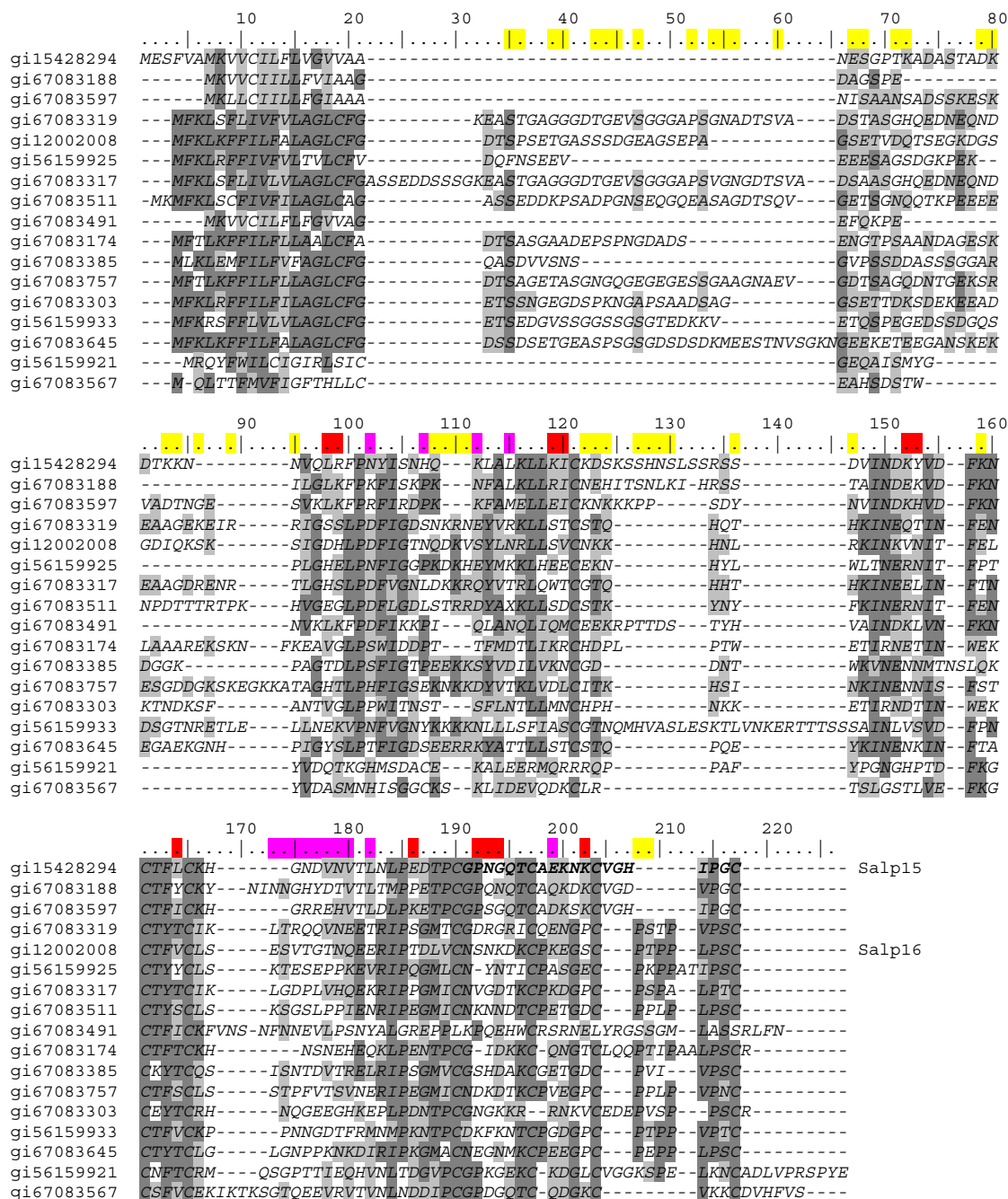
## Results

### Positive Selection in *Ixodes* Saliva Proteins

As Salp15 (gi: 15428294) is a key protein within the transmission of *Borrelia burgdorferi* from the tick into its new host, we used PSI-Blast (Altschul et al. 1997) with default parameters to identify further members of the family. Indeed, 17 paralogues were found in *Ixodes scapularis*, including Salp16 (gi: 12002008), all belonging to the previously classified groups of Cys-rich ixostatins and the LPTS coxy family (Ribeiro et al. 2007). Their alignment revealed a conserved signal sequence at the N-terminus, followed by conserved blocks and gap regions

(Fig. 1). To exclude positions which cannot be aligned with certainty, highly divergent sections of the alignment were cut out as described under “Materials and Methods”, generating the alignments Salp15\_cut (103 amino acids; Supplementary Fig. S2) and Salp15\_gBlocks (65 amino acids; Supplementary Fig. S3). The highly conserved signal sequence remained in all alignments, as the used site

specific models are able to cope with different selective pressure for each site. For each of these three alignments, a phylogenetic tree was calculated by proml. As the most reliable tree was the one calculated on the complete sequences (Supplementary Fig. S1), this was further used for analyzing the complete alignment and Salp15\_cut. For the Salp15\_gBlocks analysis the tree based on the Gblocks



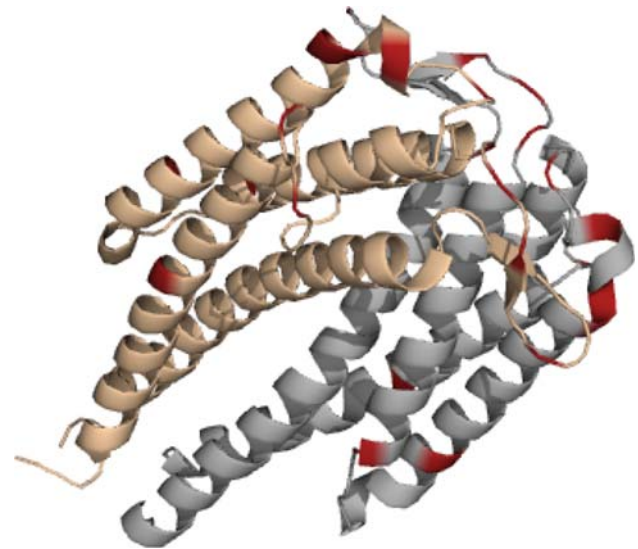
**Fig. 1** Sequence alignment of the 17 *Ixodes scapularis* Salp15-related proteins, named with the GenBank Identifier. Positive selection sites detected by codeml  $>0.90$  are highlighted in yellow if detected by a single analysis, in pink if predicted by two assays, and

in red if predicted by all three methods. Fifty-four sites were detected after a complete sequence analysis, 31 after submitting the Salp15\_cut.aln and 17 when considering only the Gblocks conserved fragments. The region of Salp15 interacting with CD4 is in boldface

alignment was used (Supplementary Fig. S4). To identify regions which have undergone positive selection, maximum likelihood estimates of  $K_A/K_S$  ratios were performed using three different approaches: (1) analysis of complete sequence alignment and tree based on complete sequence alignment; (2) analysis of Salp15\_cut and tree based on complete sequence alignment; and (3) analysis of Salp15\_gBlocks and tree based on the Gblocks alignment. As the average  $K_S$  in all three versions of the alignment was  $<0.42$  (complete alignment,  $0.4 \pm 0.27$ ; Salp15\_cut,  $0.42 \pm 0.41$ ; Salp15\_gblocks,  $0.32 \pm 0.27$ ), saturation of synonymous sites was not observed. Models taking into account sites under positive selection fitted the data significantly better than others (with  $p < 0.001$  for M1a-M2a, M7-M8, and M8a-M8 in all three models except for  $p = 0.05$  in analysis 1, M8a-M8) (Table 1). Moreover, the positive selection sites detected in the three approaches ( $p > 0.90$ ) by Bayes Empirical Bayes (BEB) using M8 were redundant, having the 17 Salp15\_gBlocks sites included in the 31 Salp15\_cut sites and both included in the 54 complete alignment analysis (Fig. 1).

Do the identified sites anyhow relate to the molecular function of the family? To date, mainly Salp15 has been studied in detail. Its C-terminal 20 amino acids are responsible for the interaction with CD4 (Garg et al. 2006) (highlighted in Fig. 1). Indeed, four of the sites identified in all of our analyses fell into this region. As these positions have been under adaptive selection in the whole family, this region might be involved in protein interactions also in other members. If this is the case, the identified positions might be responsible for the specificity of the interaction. Another protein of the human immune system interacting with Salp15 is DC-SIGN (Hovius et al. 2008). This interaction relies on a carbohydrate structure on Salp15. Therefore, one would not expect to see any signs of positive selection in the amino acid sequence of the protein.

Then what is the function of the further sites under adaptive selection? One possibility is that they are involved in the interaction with the bacterial OspC. If this were the case, interacting sites in OspC might also show signs of adaptive evolution.



**Fig. 2** The OspC dimer structure (1GGQ.pdb) with positive selection sites highlighted in red in cartoon view. Thirteen  $\omega +$  sites have been detected, mostly in the surface-exposed loop regions and the shorter  $\alpha$ -helices (positions 51, 57, 59, 93, 114, 125, 140, 141, 145, 148, 164, 167, and 182). The two long  $\alpha$ -helices from each subunit form the core of the structure and extend the entire length of the molecule, having only two positive selection spots

#### Positive Selection in *Borrelia* OspC

We therefore tested several evolutionary models for the borrelian OspC protein, which not only binds to Salp15 (Anguita et al. 2002) but also is up-regulated during tick engorgement and involved in colonizing the mammalian host. It has been noted previously that regions of OspC have undergone phases of positive selection (Theisen et al. 1995). Using site-specific models we were able to search for positions which were the target sites of this selection. We aligned 17 OspC sequences from *Borrelia burgdorferi* with MUSCLE (Supplementary Fig. S5), calculated a phylogenetic tree (Supplementary Fig. S6), and performed a codeml analysis. Indeed, models considering sites under positive selection fitted the data significantly better than models that did not ( $p < 0.001$  for M1a-M2a, M7-M8, and M8a-M8) (Table 1). In total, 13 sites under positive

**Table 1** Codeml results for *Ixodes scapularis* Salp15-like sequences and for *Borrelia burgdorferi* OspC:  $K_A/K_S$  ratio average across all sites under PAML model M1a (neutral);  $p_1$  proportion of sites in  $\omega > 1$ ;  $\omega_1$  estimate of  $\omega$  for this class

	$K_A/K_S$	Average $K_S$	$p$ -value			Selection parameters, M8
			M1a vs. M2a	M7 vs. M8	M8a vs. M8	
Salp15						
Complete alignment	0.6203	0.398	$<0.001$	$<0.001$	$<0.05$	$p_1 = 0.436, \omega_1 = 1.476$
Cut	0.7375	0.418	$<0.01$	$<0.001$	n.s.	$p_1 = 0.537, \omega_1 = 1.24$
gBlocks	0.4918	0.319	$<0.001$	$<0.001$	$<0.001$	$p_1 = 0.292, \omega_1 = 1.886$
OspC	0.3531	0.11	$<0.001$	$<0.001$	$<0.001$	$p_1 = 0.089, \omega_1 = 2.49$

selection were identified ( $p > 0.90$ ). The three-dimensional structure of OspC allowed us to check whether these sites are spatial related (Eicken et al. 2001). Indeed,  $\omega+$  sites lay mostly in the surface-exposed region of the protein (Fig. 2). Eight fell into the epitope-containing region from helix 3 to helix 5 (Earnhart and Marconi 2007). The remaining five sites under positive selection outside of this region might be candidates for an interaction with a ligand like Salp15.

## Discussion

Analyzing the rate of synonymous versus nonsynonymous substitutions in the Salp15 family of tick saliva proteins revealed that this protein family as well as one interaction partner in bacteria, OspC, has undergone a period of adaptive evolution. This finding might be of relevance for two aspects of research on tick mediated *Borrelia* infection. First, it sheds some light on the evolutionary constraints shaping such a delicate system involving three species, the arthropod tick, the bacterial pathogen, and the mammalian host. How are the bacteria that utilize these proteins affected by the rapid evolution of the Salp15 family proteins? To date, the only bacterial protein known to interact with a tick saliva protein of the analyzed group is OspC, which binds directly to Salp15. Indeed, we identified sites having undergone adaptive evolution in OspC. In a possible model, the host's immune system could drive the adaptive evolution of Salp15. As *B. burgdorferi* is dependent on the interaction of OspC with Salp15 for infection of the host, this, again, might drive the evolution of OspC. Obviously, this model is highly speculative and it will be challenging, if possible at all, to dissect the evolutionary forces on OspC, as this protein is itself a major target of the human immune system. A first step in this direction would be the identification of sites involved in the interaction of OspC and Salp15. If this region on OspC is not involved in interactions with the host's immune system but does harbor positions identified here, this could indicate that the host's immune system does not drive the adaptive evolution of this part of OspC but, rather, Salp15 itself does. Although some of the identified sites in Salp15 fall into the region interacting with CD4, reasons for the adaptive evolution of Salp15 other than evasion of the host's immune system are conceivable. For example, members of this protein family might have been involved in the adaptation to new hosts or there could be pressure to evolve different functional features.

Second, our findings might direct further experimental research on proteins of the Salp15 family and their interaction with bacterial as well as mammalian proteins. In the case of Salp15, identified sites in the region interacting

with CD4 might be good candidates for site-directed mutagenesis to further understand the details of this interaction. Similarly, it would be interesting to test whether OspC binds to the other identified sites or whether there are further proteins responsible for their adaptive evolution. Finally, the presence of positively selected sites throughout the whole family might indicate the importance of other members as exemplified by Salp16. Indeed all of the members are expressed between 6 and 72 h post host attachment (Ribeiro et al. 2007). As our analyses relied solely on sequence data, this study illustrates how we can benefit from the soon-to-come genome sequence of the tick (Hill and Wikel 2005).

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anguita J, Ramamoorthi N, Hovius JW, Das S, Thomas V, Persinski R, Conze D, Askenase PW, Rincon M, Kantor FS, Fikrig E (2002) Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4(+) T cell activation. *Immunity* 16:849–859
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
- Das S, Banerjee G, DePonte K, Marcantonio N, Kantor FS, Fikrig E (2001) Salp25D, an *Ixodes scapularis* antioxidant, is 1 of 14 immunodominant antigens in engorged tick salivary glands. *J Infect Dis* 184:1056–1064
- Earnhart CG, Marconi RT (2007) OspC phylogenetic analyses support the feasibility of a broadly protective polyvalent chimeric Lyme disease vaccine. *Clin Vaccine Immunol* 14: 628–634
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
- Eicken C, Sharma V, Klabunde T, Owens RT, Pikas DS, Hook M, Sacchetti JC (2001) Crystal structure of Lyme disease antigen outer surface protein C from *Borrelia burgdorferi*. *J Biol Chem* 276:10010–10015
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle
- Foley J, Nieto N (2007) Anaplasma phagocytophilum subverts tick salivary gland proteins. *Trends Parasitol* 23:3–5
- Garg R, Juncadella IJ, Ramamoorthi N, Ashish, Ananthanarayanan SK, Thomas V, Rincon M, Krueger JK, Fikrig E, Yengo CM, Anguita J (2006) Cutting edge: CD4 is the receptor for the tick saliva immunosuppressor, salp15. *J Immunol* 177:6579–6583
- Gillespie RD, Mbow ML, Titus RG (2000) The immunomodulatory factors of bloodfeeding arthropod saliva. *Parasite Immunol* 22:319–331
- Hill CA, Wikel SK (2005) The *Ixodes scapularis* Genome Project: an opportunity for advancing tick research. *Trends Parasitol* 21:151–153
- Hovius JW, de Jong MA, den Dunnen J, Litjens M, Fikrig E, van der Poll T, Gringhuis SI, Geijtenbeek TB (2008) Salp15 binding to DC-SIGN inhibits cytokine expression by impairing both

- nucleosome remodeling and mRNA stabilization. *PLoS Pathog* 4:e31
- Motameni AR, Juncadella IJ, Ananthanarayanan SK, Hedrick MN, Huet-Hudson Y, Anguita J (2004) Delivery of the immunosuppressive antigen Salp15 to antigen-presenting cells by *Salmonella enterica* serovar Typhimurium aroA mutants. *Infect Immun* 72:3638–3642
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936
- Ribeiro JM, Arca B, Lombardo F, Calvo E, Phan VM, Chandra PK, Wikel SK (2007) An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*. *BMC Genomics* 8:6
- Schuijt TJ, Hovius JW, van Burgel ND, Ramamoorthi N, Fikrig E, van Dam AP (2008) The tick salivary protein Salp15 inhibits the killing of serum-sensitive *Borrelia burgdorferi* sensu lato isolates. *Infect Immun* 76:2888–2894
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612
- Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20:18–20
- Theisen M, Borre M, Mathiesen MJ, Mikkelsen B, Lebech AM, Hansen K (1995) Evolution of the *Borrelia burgdorferi* outer surface protein OspC. *J Bacteriol* 177:3036–3044
- Trager W (1939) Acquired immunity to ticks. *J Parasitol* 25:57–81
- Valenzuela JG, Charlab R, Mather TN, Ribeiro JM (2000) Purification, cloning, and expression of a novel salivary anticomplement protein from the tick, *Ixodes scapularis*. *J Biol Chem* 275:18717–18723
- Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573