Infection, Genetics and Evolution 68 (2019) 153–160

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Evolution and diversity of the EMA families of the divergent equid parasites, *Theileria equi* and *T. haneyi*

L.N. Wise[b,c,*,1], L.S. Kappmeyer[a,1], D.P. Knowles[b], S.N. White[a,b,d,**]

[a] Animal Disease Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Pullman, WA, USA
[b] Department of Veterinary Microbiology and Pathology, College of Veterinary Medicine, Washington State University, Pullman, WA, USA
[c] School of Veterinary Medicine, St. George's University, True Blue, Grenada
[d] Center for Reproductive Biology, Washington State University, Pullman, WA, USA

ARTICLE INFO

Keywords:
Theileria equi
Theileria haneyi
Merozoite antigen

ABSTRACT

The equine parasite *Theilera equi* continues to curtail global equine commerce due primarily to its ability to persist indefinitely in the immunocompetent horse. Details regarding the parasite life cycle, pathogenesis and mechanism of persistence remain unclear. The recently discovered *T. haneyi* is also capable of persistence in the horse, creating a potential reservoir for additional infections. These two divergent parasites share a unique gene family that expresses surface merozoite antigens, or equi merozoite antigens (EMAs). The EMA family was maintained in number and size in both parasites despite a species divergence of over 30 million years ago. This family is unique amongst *Theilerias* in number, structure and biochemical properties. In silico analysis revealed no evidence of selection for diversity within this family, indicating a role in host adaptation and persistence rather than antigenic variation and immune escape. Biochemical analysis revealed the presence of a conserved domain, homologous to the hemolysin toxin found in cobra venom. This finding combined with data from protein interaction prediction models may indicate interaction with the structural components of the host erythrocyte and a role in merozoite entry or escape. Additional predicted protein interactions focus on disruption of the enzymatic functions of the host cell, potentially resulting in enhanced parasite survival.

1. Introduction

Theileria equi is a tick-borne, apicomplexan parasite of equids and one of the causative agents of Equine Piroplasmosis. Naïve animals experience clinical symptoms of hemolytic anemia and associated systemic sequelae. Resolution of initial infection invariably results in lifelong persistence and transmission risk. Global equine commerce is curtailed due to restricted movement of infected horses (Wise et al., 2013). The mechanisms by which *T. equi* persists remain unknown due to voids in knowledge concerning parasite's life cycle, ability to undergo antigenic variation and host immune responses. Data support the hypothesis that the erythrocytic merozoite stage of *T. equi* causes erythrocyte rupture and is the persistent form of the parasite (Melhorn and Schein, 1998).

Infected horses produce high levels of antibodies to merozoite surface antigens, termed equi merozoite antigens or EMAs (Knowles et al., 1992; Melhorn and Schein, 1998). Antibody responses to EMA proteins are the basis for several diagnostic and regulatory tests, and EMA1-3 are known to interact with the internal cytoskeleton of the erythrocyte during infection (Kumar et al., 2004; Silva et al., 2013; Wise et al., 2017). Completion of the genome sequence of *T. equi* led to the discovery that the EMA proteins were not encoded by single copy genes as is the case of the orthologs found in *T. parva* (mMPSA), *T. annulata* (TAMS1) and *T. orientalis* (MSPS), but instead members of a larger, nine member multigene family (*ema*1–9) (Kappmeyer et al., 2012). These single copy orthologs of the *ema* genes found in other *Theileria* species are immunodominant and functional roles in antigenic variation and erythrocyte interaction have been hypothesized and reported (Gubbels et al., 2000; Jenkins and Bogema, 2016; Hayashida et al., 2013, Skilton et al., 2000). In *T. orientalis*, MPSP was demonstrated to be involved in merozoite entry into the erythrocyte and TAMS1 in *T. annulata* has been thoroughly investigated for its role in antigenic variation and immune escape (Katzer et al., 2002; Takemae et al., 2014).Additional information about the uniquely structured *ema* family was obtained with the

* Correspondence to: L. N. Wise, Post Office Box 7, True Blue, St Georges, Grenada.
** Correspondence to: S. N. White, PO Box 646630, Pullman, WA 99164, USA.
E-mail addresses: lwise1@sgu.edu (L.N. Wise), Stephen.White@ars.usda.gov (S.N. White).
[1] Wise and Kappmeyer are co-first authors.

https://doi.org/10.1016/j.meegid.2018.12.020
Received 12 October 2018; Received in revised form 4 December 2018; Accepted 17 December 2018
Available online 18 December 2018

recent discovery of the closely related, novel, equid parasite *T. haneyi* (Knowles et al., 2018). As with *T. equi*, *T. haneyi* also has 9 *ema* family genes, rather than the single copy genes found in all other *Theileria* species. Phylogenetic and syntenic analysis indicates that the ema family underwent expansion prior to the speciation event that led to separation of *T. equi* and *T. haneyi* approximately 33 million years ago. In the *T. haneyi* genome, six of the nine *ema* homologs identified are syntenic orthologs to *T. equi*. The *T. equi* ema genes, *ema*1, *ema*3, and *ema*4 have no detectable orthologs in *T. haneyi* but *T. haneyi* contains three distinct *ema*s, named *ema*11–13, bringing the total number to nine as in *T. equi*. The infection dynamics of *T. haneyi* in the horse remain unclear but the erythrocytic merozoite stage predominates in cases of inapparent, persistent infection in the horse (Knowles et al., 2018).

The purpose of this study was to investigate the expanded *ema* families of both *T. equi* and *T. haneyi* in regard to genetic diversity, structure and potential function. Since roles in antigenic variation leave signatures of historical diversifying selection to achieve immune escape, we will leverage the 30 million years of divergent evolution between *T. equi* and *T. haneyi* to test for signatures of positive selection. We will also investigate signatures of negative selection (conservation) in ema genes at unprecedented resolution by applying codon site-wise analysis, and utilize protein interaction partner predictions to further refine understanding of ema function.

2. Materials & methods

2.1. Genomic analysis

The published genomes of *T. equi* and *T. haneyi* underpin the advanced analyses in this study (Kappmeyer et al., 2012, Knowles et al., 2018). Alignment of the *Theileria sp* EMA sequences was done using CLC Genomics Workbench version 10.1.1 (Qiagen Bioinformatics, Redwood City, CA, USA) Create Alignment function. Once an alignment of the full length *T. equi* and *T. haneyi* family members was complete, segments of the aligned proteins overlapping the EMA1 and EMA2 hemolysin domain were displayed as a subset alignment (Knowles et al., 1997). Total protein length, isoelectric points, and molecular mass, and domains were determined through the ExPASy Bioinformatics Resource Portal (https://web.expasy.org/compute_pi/). To assess the predicted protein structure for each EMA in *T. equi* and *T. haneyi*, sequences were analyzed using InterPro (https://www.ebi.ac.uk/interpro), ExPASy Bioinformatics Resource Portal (https://web.expasy.org/compute_pi/) and KohGPI (http://gpi.unibe.ch) (Fankhauser and Mäser, 2005; Finn et al., 2017).

2.2. Codon site-wise analysis for historical positive and negative selection

Evidence for signatures of historical selection (positive and negative) acting on the *ema* gene family expansion in *T. equi* was tested using gene sequences from the *T. equi* EMA1-EMA9 and *T. haneyi* EMA2, EMA5-EMA9, and EMA11-EMA13 (Accession numbers in Table S1) (Kappmeyer et al., 2012; Knowles et al., 2018). Additional analyses of EMA5 were performed using sequences from *T. equi* and *T. haneyi*, as well as *T. annulata*, *T. parva*, and *T. orientalis* (Hayashida et al., 2013; Bogema et al., 2018; Pain et al., 2005). Phylogeny-aware codon-based multiple alignment was performed using PRANK software specifically to minimize concerns about possible historical insertion-deletion events on positive selection analysis (Jordan and Goldman, 2012; Löytynoja and Goldman, 2008; Löytynoja and Goldman, 2010). Codon site-wise analysis for positive selection was performed using analytic measures from two separate software packages. Specifically, the likelihood ratio tests for site models (M1a versus M2a and M7 versus M8 models) implemented in PAML 4 software (Yang, 2007) and the site-wise likelihood ratio (SLR) test (Massingham and Goldman, 2005; Sánchez et al., 2011) were both used for assessing positive selection. Codon-based alignment by MUSCLE (Edgar, 2004) in MEGA (Kumar et al., 2016) supported codon site-wise negative (conservative) selection analysis by SLR. For this analysis, Hochberg step-up adjusted *P*-values were used to determine support for conservative selection (Massingham and Goldman, 2005).

2.3. Host protein interaction partner prediction for EMA proteins

The Red Blood Cell Collection (RBCC) database provided a list of 1286 proteins expressed in red blood cells with high confidence (Hegedűs et al., 2015). These peptide sequences were queried for potential interaction partners for EMA proteins using the machine learning-based system HOPITOR (Host Pathogen Interaction predictor) (Basit et al., 2018). Hopitor leverages machine-learning from known host-pathogen protein interactions, and those known interactions are biased in favor of human host proteins. Furthermore, human gene exon definition and protein sequence annotation are much more complete than for equids, so human red blood cell protein sequences as well as horse protein sequences were utilized to maximize discovery of all possible protein interaction partners for the EMA proteins (Jónsson et al., 2014; Wade et al., 2009; Zerbino et al., 2018). Gene Ontology term overrepresentation analysis was performed using PANTHER with the GO database release date of September 6, 2018 (Mi et al., 2016). All host genes predicted by HOPITOR to interact with 15 or more EMA family proteins were included in analyses compared to the input list utilizing a binomial test with Bonferroni correction for multiple testing. Annotation datasets included GO molecular function, GO biological process, and GO cellular component.

3. Results

3.1. Genomic analysis

Although the alignment of the 18 EMA polypeptides (Fig. 1a) demonstrates the amino acid variability that extends throughout the length of the comparison, the overall primary structure of each predicted protein is conserved. A comparison of the predicted peptide length of each of the 6 full length shared genes as well as the 3 genes unique to each organism is quite similar, ranging from 267 to 289 amino acids. As previously reported, the predicted amino acid identity amongst all EMA family members of *T. equi* and *T. haneyi* ranges from 34% to 81% with *ema*11–13 being the most similar (Knowles et al., 2018). *Ema*9 is truncated in *T. equi* yet is full length in *T. haneyi*. In contrast, EMA 7 is truncated in *T. haneyi* but full length in *T. equi*.

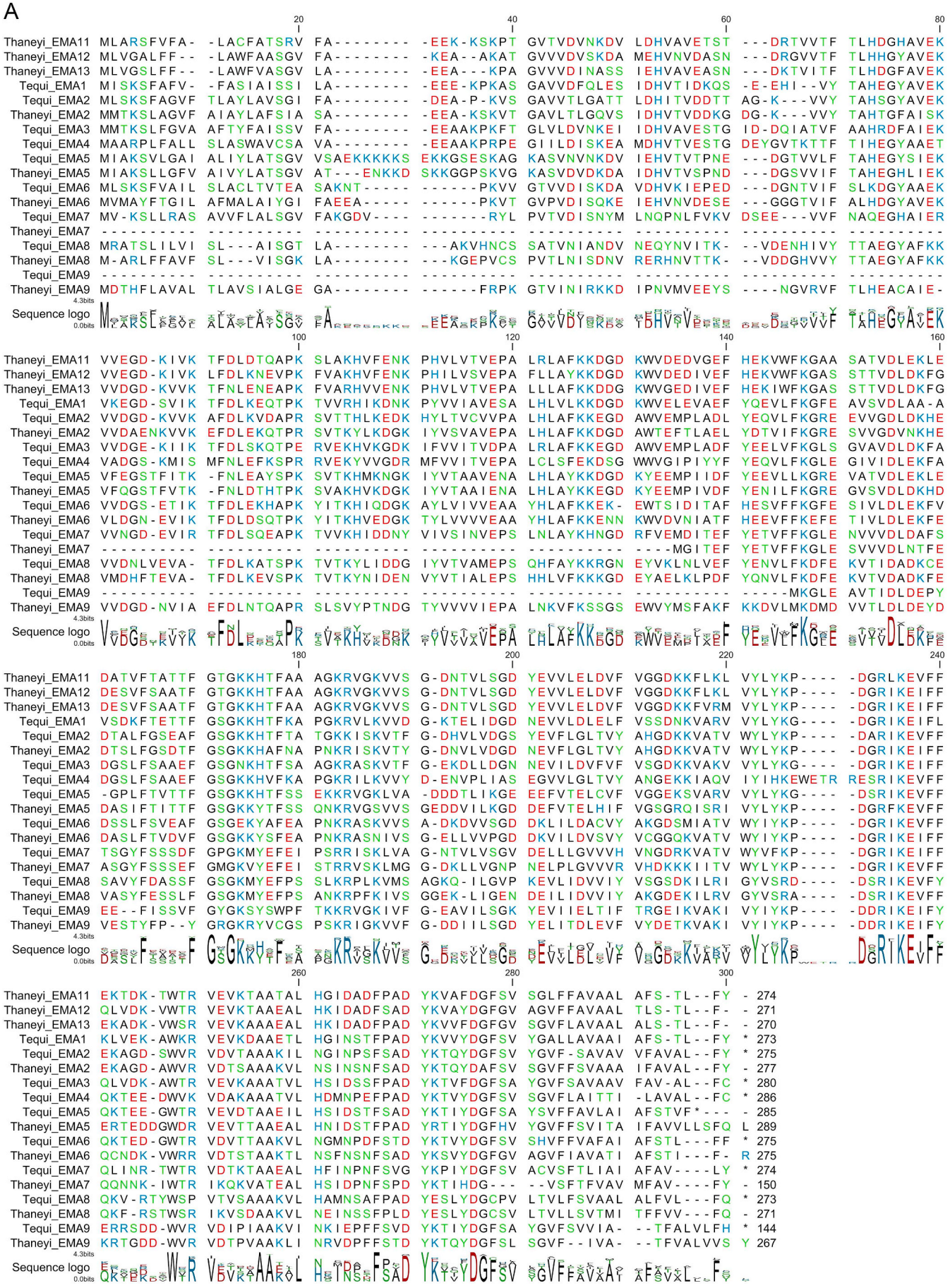
Analysis of the predicted structure of the EMA proteins in both species revealed that each of the full length proteins contain a signal peptide, a C terminal transmembrane domain and a GPI anchor. This is consistent with the previous findings for *T. equi* EMA1 and EMA2 (Kappmeyer et al., 1993; Knowles et al., 1997).

All 18 *ema* family members have an overall acidic isoelectric point with the exception of the *ema*7 in *T. haneyi*, the truncated EMA member of that species. Within these acidic predicted proteins, a basic domain at approximately the 140-170aa location (Table 1) was identified in all 18 family members. This domain was previously identified in *T. equi ema*1, ema2, *T. parva* mMPSA and *T. annulata* TAMS1 as a homolog to an elapid venom hemolysin toxin (Fig. 1b) (Katzer et al., 2002; Knowles et al., 1997).

3.2. Codon site-wise analysis for historical positive and negative selection

Repeated tests failed to detect signatures of historical positive selection at any amino acid site in the EMA gene family. Neither likelihood ratio test from the PAML package (M1a versus M2a, M7 versus M8) detected evidence for positive selection (all $P > 0.05$) for EMA within *T. equi*, within *T. haneyi*, or in the omnibus analysis including EMA from both *T. equi* and *T. haneyi*. Further, the SLR test did not detect evidence for any site with historical positive selection in any of

A



(caption on next page)

Fig. 1. a/b: Alignment of EMA family polypeptides predicted from genome coding sequences for *Theileria equi* (Tequi) and *Theileria haneyi* (Thaneyi) with EMA common name designations and comparative lengths. Fig. 1a shows the entire alignment. Fig. 1b shows only the alignment of the hemolysin domains. Gaps are indicated with (−). A sequence logo indicating predominant residues at each position is in the style of Schneider and Stephens, and with a polarity color scheme where nonpolar (A, C, F, I, L,M, V, W) are black and uncharged (N, Q, S, T, Y) residues are green, while charge-contributing residues emphasized with acidic residues in red, and basic residues in blue. The values 4.2 and 0 bits indicates the sequence information content per position based on the comparable height of the residue designations from calculation in CLC Genomics Workbench alignment tool (Schneider and Stephens, 1990) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

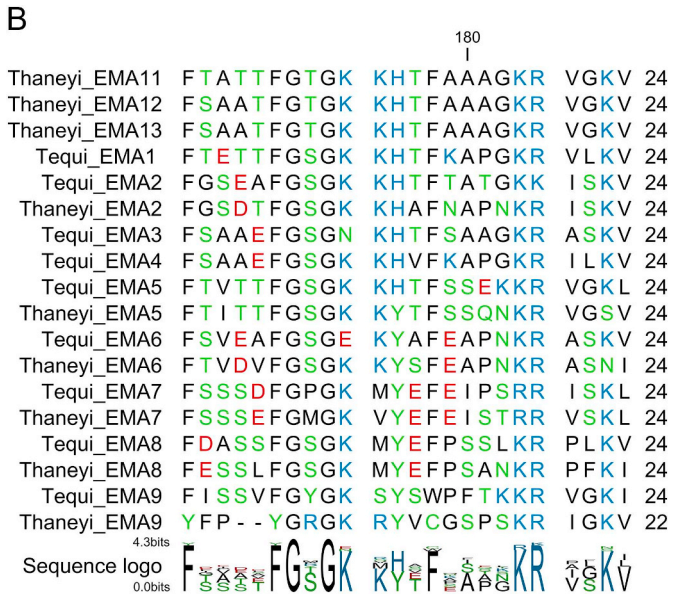


Fig. 1. (*continued*)

these analyses (all nominal $P > 0.05$ and Hochberg P > 0.05).

However, negative selection analysis did identify numerous codon sites with evidence of historical negative selection (Supplementary Figs. S1 & S2). Supplementary Fig. S1 shows an overall alignment of all EMA family members from both *T. equi* and *T. haneyi.* This analysis highlights evidence for conservative selection at 134 sites with at the 95% confidence and 25 sites with 99% confidence. Supplementary Fig. S2 shows an alignment of EMA5 orthologs across 5 *Theileria* species, and highlights evidence for conservative selection at 13 sites at 95% confidence and 18 at 99% confidence.

Table 1
EMA family member comparison for *T. equi* and *T. haneyi*. **T. equi ema9* and *T. haneyi ema7* are truncated in that species but full length in the other species.

Common Name	Predicted peptide length	Molecular mass (kda)	Overall PI	PI – hemolysin domain only
ema1	272 aa	28.1	5.38	10.58
ema2 equi	274 aa	27.4	5.71	10.18
ema2 haneyi	277aa	30.1	6.09	10.46
ema3	280 aa	28.1	5.24	10.29
ema4	275 aa	28.3	5.13	10.58
ema5 equi	286 aa	29.6	5.81	10.58
ema5 haneyi	289aa	32	5.48	10.46
ema6 equi	285 aa	28.1	5.12	8.43
ema6 haneyi	275aa	30.6	4.92	9.53
ema7 equi	274 aa	29.1	4.98	8.52
*ema7** *haneyi*	150aa	17.2	7.89	10.00
ema8 equi	273aa	28.2	6.14	9.53
ema8 haneyi	271aa	30.7	5.82	9.53
*ema9** *equi*	144 aa	16.4	5.89	10.30
ema9 haneyi	267aa	30	4.87	10.30
ema11	274aa	30.1	5.93	11.33
ema12	271aa	29.8	5.78	11.33
ema13	270aa	29.4	5.33	11.33

3.3. Host protein interaction partner prediction for EMA proteins

HOPITOR analysis results, using 1286 known red blood cell proteins and each of the 18 EMAs, identified 905 human proteins and 523 horse proteins that had a predicted interaction with at least one member of the EMA family (Tables S2 and S3). In the human analysis, 69 proteins interacted with at least 15 of the EMAs and in the horse, 25 fit this criterion. Eighteen of these proteins in both the human and horse genomes were predicted to interact with at least 15 of the 18 members of the EMA family, the most notable of which are described in Table 2.

Gene Ontology analysis for GO cellular component terms with human proteins from the HOPITOR interaction prediction set highlighted three terms with significant enrichment, cullin-RING ubiquitin ligase complex ($P = 0.015$, 10.1-fold enrichment), nuclear chromosome part ($P = 0.049$, 5.5-fold enrichment), and Unclassified ($P < 0.001$, 1.85-fold enrichment). Genes highlighted for the term cullin-RING ubiquitin ligase complex included CAND1, CUL1, CUL2, DDB1, FBXL20, and SKP1. Genes highlighted for the term nuclear chromosome part included only histone genes (Table S2). Similar analysis from the horse protein interaction prediction set showed two significantly enriched terms, cullin-RING ubiquitin ligase complex ($P = 0.031$, 15.5-fold enrichment) and Unclassified ($P < 0.001$, 1.85-fold enrichment). Genes highlighted for the term cullin-RING ubiquitin ligase complex included CUL1, CUL2, DDB1, and SKP1. No terms showed significant enrichment for GO molecular function or GO biological process in either predicted interaction set (all $P > 0.05$).

4. Discussion

Infections with *T. equi* continue to have a significant, negative impact on global equine commerce and trade. The mechanism by which this parasite persists within the immunocompetent horse remains unclear. The life stage that invades the horse via an infected tick is the sporozoite. Like the closely related cattle parasites *Theileria parva* and *T. annulata*, it first invades peripheral blood mononuclear cells yet the role of PBMC infection in the pathogenesis of *T. equi* is unknown (Ramsay et al., 2013). Data supports the hypothesis that the persistent stage of *T. equi* is the merozoite (Melhorn and Schein, 1998). Parasite encoded antigens have yet to be identified on the surface of infected erythrocytes and the lysis of infected erythrocytes is not through adaptive immune recognition of erythrocyte surface embedded parasite antigens (Knowles et al., 1994). Thus persistent infection is maintained through asexual reproduction within the erythrocyte prior to rupture of the cell, release of daughter merozoites, which then enter additional erythrocytes. If the acute infection is not fatal, horses enter a carrier phase of infection where they are generally asymptomatic yet serve as a reservoir for transmission.

Identification of the mechanisms by which the parasite persists is of paramount importance in the pursuit of prevention and treatment of this insidious disease. The role that the recently discovered *T. haneyi* plays in the equine piroplasmosis disease complex has yet to be fully elucidated but given its similarities to *T. equi*, is likely to be an integral aspect. Genomic analysis of *T. equi* and *T. haneyi* did not indicate that the *ema* family possesses characteristics consistent with a typical apicomplexan multi-gene family capable of antigenic variation, yet interspecies conservation of this unique gene family lends itself to functional importance for the parasite. This body of research compared the ema

Table 2
HOPITOR Analysis Results (See Tables S2 and S3 for complete analysis results). Proteins from both the human and horse database strongly predicted to interact with the EMA proteins.

Gene (Protein) name	Number of EMA proteins predicted to interact	Average probability score*	Function of the protein within the erythrocyte
CALR (Calreticulin)	17	0.767	Ubiquitin binding ligase
DDB1 (Damage specific DNA binding)	18	0.858	Cullin RING ubiquitin ligase complex components[a]
SKP1 (S phase kinase)	18	0.908	
CAL1 (Cullin 1)	18	0.959	
CAL2 (Cullin 2)	17	0.847	
FBXL20 (F box/Leucine rich repeat)	18	0.853	
GAPDH (glyceraldehyde dehydrogenase)	17	0.888	Glycolytic enzyme; vesicular transport
TUBA1A (Alpha tubulin)	18	0.935	Microtubule proteins
TUBB (Tubulin beta)	18	0.904	
TUBA1C (Alpha tubulin)	18	0.953	

* Predicted probability of interaction is considered significant at > 0.5.
[a] Additional ubiquitin associated proteins (CAND1 – human; YOD1 – horse) were also identified but not in both species.

families of *T. equi* and *T. haneyi*, including putative function, by exploring the potential effects of natural selection.

The nine-member *ema* family of *T. equi* was originally described as a 10-member family (*ema*1–10). The ema10 gene model (BEWA_047350) was clustered with the first 9 by TribeMCL based only on conserved N-terminal residues typical of a hydrophilic signal peptide, and hydrophobic C-terminal residues typical of a transmembrane domain (Kappmeyer et al., 2012). Intervening sequence bears no resemblance in sequence length and conserved residues seen in all other EMAs, and lacks the hemolysin domain. This further analysis leads the authors to exclude it as an EMA family member, and likely has no ancestor genetic relationship to the other EMAs. Further, with discovery of the EMA gene family in *T. haneyi*, no ortholog to BEWA_047350 was found to exist, and those authors determined to name unique family members EMA 11–13 to avoid confusion with the previous reference to EMA 10 (Knowles et al., 2018). A detailed cross-species, ema gene-based phylogenetic analysis was presented in their work, especially in Fig. 3 (Knowles et al., 2018). Numerous *T. equi* isolates are present across the world and analyses of geographic diversity have been presented previously (Cunha et al., 2002; Knowles et al., 1997), but the ema family remains highly conserved regardless of location (Hall et al., 2013; Knowles et al., 1991a). Originally, *T. haneyi* was thought to be a new isolate of *T. equi* but phylogenetic evidence supports that *T. haneyi* diverged from *T. equi* approximately 33 million years ago. Data indicates that the ema family first expanded within *T. equi* prior to the speciation event that occurred resulting in *T. haneyi* (Knowles et al., 2018). Despite this divergence the ema family maintained the same number, biochemical structure and presumably function in both species.

Sequence analysis demonstrates that *T. haneyi* ema11–13 duplicated from the existing ema genes. Predicted amino acid identity amongst all EMA family members of *T. equi* and *T. haneyi* ranges from 34% to 81% with *ema*11–13 being the most similar at 74–81%. Based on phylogeny, ema1 gave rise to all ema other *ema* genes and *ema*11–13 is most closely related to *ema*5 (Knowles et al., 2018). EMA1 and EMA2 have been previously described in the literature as immunodominant, surface expressed proteins. The highly conserved EMA1 is the basis for the nPCR test and for the regulatory cELISA diagnostic test due to the fact that infected horses globally have high titer responses to EMA1 in both the acute and chronic phases of disease (Knowles et al., 1991b). Antibodies against EMA1 and 2 are correlated with protection, yet they are unable to eliminate the parasite completely (Cunha et al., 2006). Little information is currently available regarding the EMAs of *T. haneyi* during infection in equids but its lack of EMA1 at least partially explains why infected horses test negative on the cELISA and the nPCR for *T. equi*. EMAs in both species have a unique yet conserved amino acid structure. Analysis for predicted protein structure supports the previous reports for EMA1 and 2; that the EMA proteins are surface expressed and membrane bound via a GPI anchor (Kappmeyer et al., 1993; Knowles et al., 1997).

PAML and SLR are complementary programs that use coding DNA to detect patterns at each codon site within a given alignment. The $\omega = dN/dS$ ratios for each site are calculated by comparing the rates of nonsynonymous (codon altering) substitutions to synonymous (silent) substitutions. Diversifying selection leads to many more nonsynonymous substitutions than silent changes, so the dN/dS ratio is > 1. Conversely, purifying selection leads to more silent substitutions than nonsynonymous changes, so the ratio is < 1. PAML uses a maximum likelihood approach and a likelihood ratio test to determine significance for positive (diversifying) or negative (purifying) selection. SLR uses a Bayesian approach, but both provide codon site-wise insight into types of historical selection.

None of the tests for genetic diversity found evidence for positive selection during the expansion of the EMA family of *T. equi* or *T. haneyi* at any codon site. This finding is not consistent with the hypothesis that the ema family's primary role in *T. equi* and *T. haneyi* persistence is antigenic variation. Additionally, there were many sites identified with significant evidence of negative (purifying) selection (Supplementary Fig. S1). This is consistent with either general conservation of roles for the EMA proteins as the family expanded, or diversification of roles for the EMA1–9 proteins without selection pressure. Signatures such as these indicate a role in host adaptation and persistence rather than antigenic variation and immune escape, supporting the hypotheses addressed herein.

Of the 18 EMAs, EMA5 is the closest syntenic ortholog to the merozoite proteins identified in *T. annulata, T. parva* and *T. orientalis*. A separate analysis and comparison of EMA5 to its orthologs in these species revealed no evidence of positive diversifying selection at any position but several sites with significant evidence for conservation. This is largely in agreement with conservation identified in the EMA family members within *T. equi* and *T. haneyi* (Supplementary Fig. S2). This finding strengthens the assumption that these proteins in the examined *Theileria* species consistently lack evidence of diversifying selection. Previous publications regarding *T. parva* and *T. anuulata* provided evidence of negative and positive selection but the analyses were focused on the gene level and did not include the same codon sitewise analysis included in this manuscript (Hayashida et al., 2013; Pain et al., 2005; Weir et al., 2010). Overall, the finding of historical negative, purifying selection focuses inquiry on more evolutionarily stable roles for EMAs.

To further investigate the potential interaction of the EMA proteins with the erythrocyte, human and equine proteins were analyzed alongside the EMA proteins for predicted interactions and the results support that the EMA family members are most likely to interact with structural and enzymatic proteins within the erythrocyte. The HOPITOR analysis is a computational approach to determine protein-to-protein interactions. This technique uses an improved machine-learning technique with a database of known protein interactions to offer a probability of interaction between host and pathogen proteins

from outside the training set. The positive controls are created using proteins that have been shown to interact using more traditional, biochemical methods. The predicted probability of an interaction is considered significant at a value of > 0.5 (Basit et al., 2018). This analysis identified at least three classes of predicted host-pathogen protein interactions with important functional properties: cullin-RING ubiquitin ligase complexes, GAPDH, and microtubule proteins (Table 2).

The first group of predicted protein interactions include components and regulators of the cullin-RING ubiquitin ligase complexes. In general, these complexes add ubiquitin groups that target proteins for a variety of functions including degradation within the host cell (Bosu and Kipreos, 2008). In *T. annulata*, this protein group exhibits an apparent role in host cell proliferation and transformation and in *Plasmodium falciparum*, the ubiquitin system is important to parasite persistence and proliferation (Hamilton et al., 2014; Marsolier et al., 2015; Zhao et al., 2017). These proteins were significantly enriched in predicted interactions from both human ($P = 0.0073$) and horse ($P = 0.031$) erythrocyte proteins, and the large term enrichment (10.9-fold in human, 15.5-fold in horse) further supports the importance of these complexes. *T. equi* and *T. haneyi* might employ EMA family proteins in part to combat degradation of parasite proteins and/or host proteins important for parasite survival and reproduction.

The second is GAPDH, a glycolytic enzyme in the erythrocyte membrane, which was demonstrated to mediate the connection between merozoite ligands of *P. falciparum* and the host erythrocyte (Pal-Bhowmick et al., 2012). This protein is also known to have antimicrobial properties for organisms as diverse as Gram-negative bacteria and fungi (Wagener et al., 2013). It is reasonable to consider the possibility that GAPDH might also inhibit growth of *Theileria*, and having a way to circumvent this limitation might provide a strong advantage for *T. equi* and *T. haneyi*.

The third group includes both alpha and beta tubulin, components of microtubules expressed in the erythrocyte. Microtubule binding has been implicated in the process that occurs in *Theileria* species during host cell entry and specifically, a *Theileria annulata* protein (TaSE) binds to alpha tubulin with a proposed function related to parasite localization within the cell (Jalovecka et al., 2018; Schneider et al., 2007). For *T. equi* and *T. haneyi*, the details regarding merozoite entry, localization within and escape from the equid erythrocyte are unclear. Merozoites from other closely related *Theileria* parasites have been shown to encounter their host cell randomly and enter through a rapid “zippering” process that allows the parasite to be free within the host cell cytoplasm while being associated with the microtubule system (Jalovecka et al., 2018). The EMA ortholog in *T. oritentalis*, MSPS, binds to the red blood cell surface through heparin binding components (Takamae et al., 2014). Using IFAT, both EMA1 and 2 were identified on the surface of the horse erythrocyte prior to *T. equi* invasion (Kumar et al., 2004). The overall negative charge on the surface of the erythrocyte would, in theory, make it difficult for a protein with a low isoelectric point, like an EMA (average PI of 5) to interact with the cell membrane. If the proteins are involved in erythrocyte binding like *T. oritentalis* MSPS, the EMAs ability to interact may be explained by the presence of the conserved basic domain in each EMA. This basic domain, with a PI range of 8.4–10.6, was originally reported for EMA 1 and 2 and this domain has identity to elapid cytotoxin proteins found in cobra venom (Knowles et al., 1997). The ringhals cobra species *Hemachatus haemachatus* produces a variety of toxins within its venom including cytotoxins 1 and 2 (also reported as hemolytic toxin 12B and A respectively) (Joubert, 1977). These toxins have numerous biological functions, one of which is to cause hemolysis by acting as a phospholipase (Condrea, 1979). The hemolytic toxins are non-enzymatic and are classified as 3 finger toxins because of the unique secondary structure that resembles 3 “fingers”. The fingers are formed by the disulfide bonds that arise between the frequently occurring cysteine residues (Utkin, 2013). The amino acid structure of this domain in the EMA proteins lacks any cysteine residues, indicating an alternate structure, perhaps one influenced by the conserved arginine and lysine residues, is present.

It is also possible that rather than or in addition to being involved in entry into the erythrocyte, the hemolysin domain and EMA proteins function in parasite escape from the erythrocyte and hemolysis after maturation. In other hemolysis inducing *Theileria* organisms like *T. annulata* and *T. orientalis*, the details of pathogenesis also remain unclear. During infection with either parasite, normal protective bovine erythrocyte antioxidant compounds are decreased which, in conjunction with an increase in lipid peroxidase activity, results in membrane asymmetry and is hypothesized to lead to hemolysis (Razavi et al., 2011; Shiono et al., 2003). Similar findings of biochemical alterations to the erythrocyte surface were reported in donkeys infected with *T. equi* (Ambawat et al., 1999).

The localization of the EMA proteins during the intra-erythrocytic phase of infection is important in determining the potential role in merozoite escape. EMA 1 and 2 are expressed simultaneously on the surface of the intra-erythrocytic merozoite after invasion. During the later maturation phases within the erythrocyte, EMA2 was observed to be shed into the cytoplasm and to be adhered to the inner membrane of the host cell. This was not observed for EMA1 (Kumar et al., 2004). Similarly, EMA3 was identified on the surface of the merozoite and to be adhered to the inner surface of the infected erythrocyte after invasion (Ikadai et al., 2006). Additionally, EMA1 and 2 were shown to interact with the erythrocyte structural membrane proteins spectrin and actin (Kumar et al., 2012). Interestingly in our protein interaction analysis, several isoforms of actin and spectrin were tested and while a number of interactions were detected, none of these proteins were predicted to bind with > 50% of the EMAs and none of the isoforms were predicted to interact with EMA1 or 2 (Tables S2 and S3). This topic warrants additional investigation.

Interestingly, both *T. equi* and *T. haneyi* infect equid red blood cells, which are prone to form rouleaux through attractive forces (Lording, 2008; Skalak et al., 1981). It is possible that these forces complicate entry and/or escape from the red blood cells by occluding surface area and/or reinforcing membranes to increase the total force needed for penetration (Griffith and Eng, 1921). This could explain why only *Theilerias* of equids possess 9 EMA family genes, but more work will be required to address this possibility.

5. Conclusions

The results of this study indicate that the EMA family of *T. equi* and *T. haneyi* demonstrates no evidence of historical diversifying selection. This family maintained the same number and biochemistry despite the speciation event that occurred at least 33 million years ago. Structurally their mixture of amino acid conservation and variability through their length is unique but not related to antigenic variation. When this analysis was extended to compare the orthologs in other closely related *Theilerias* at the codon level, the findings were consistent and demonstrated only evidence of purifying selection at numerous sites. The identification of the hemolysin domain and predicted interaction with structural components of the erythrocyte may indicate a role in merozoite invasion, localization or escape. Additionally, this predicted interaction with important enzymatic mechanisms may indicate function related to parasite survival and propagation. This information warrants additional investigation into the specific function of this important family in both parasites' life cycle and persistence in the host.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.12.020.

Acknowledgements

The authors thank Codie Durfee and Caylee Birge for technical assistance.

Funding

This work was supported by the Agricultural Research Service (ARS), USA, CRIS#2090-320000-03900D.

References

Ambawat, H.K., Malhotra, D.V., Kumar, S., Dhar, S., 1999. Erythrocyte associated haemato-biochemical changes in *Babesia equi* infection experimentally produced in donkeys. Vet. Parasitol. 85, 319–324.

Basit, A.H., Abbasi, W.A., Asif, A., Gull, S., Minhas, F.U.A.A., 2018. Training host-pathogen protein-protein interaction predictors. J. Bioinforma. Comput. Biol. https://doi.org/10.1142/S0219720018500142.

Bogema, D.R., Micallef, M.L., Liu, M., Padula, M.P., Djordjevic, S.P., Darling, A.E., Jenkins, C., 2018. Analysis of *Theileria orientalis* draft genome sequences reveals potential species-level divergence of the Ikeda, Chitose and Buffeli genotypes. BMC Genomics 19, 298.

Bosu, D.R., Kipreos, E.T., 2008. Cullin-RING ubiquitin ligases: global regulation and activation cycles. Cell Div 3, 7.

Condrea, E., 1979. Hemolytic effects of snake venoms. In: Lee, C.Y. (Ed.), Snake Venoms. Springer Publishing, New York, pp. 450–459.

Cunha, C.W., Kappmeyer, L.S., McGuire, T.C., Dellagostin, O.A., Knowles, D.P., 2002. Conformational dependence and conservation of an immunodominant epitope within the Babesia equi erythrocyte-stage surface protein equi merozoite antigen 1. Clin. Diagn. Lab. Immunol. 9, 1301–1306.

Cunha, C., McGuire, T., Kappmeyer, L., Hines, S., Lopez, A., Dellagostin, O., Knowles, D., 2006. Development of specific immunoglobulin Ga (IgGa) and IgGb antibodies correlates with control of parasitemia in *Babesia equi* infection. Clin. Vaccine Immunol. 513, 297–300.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Fankhauser, N., Mäser, P., 2005. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics 21, 1846–1852.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuca, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C., Wu, C.H., Xenarios, I., Yeh, L.S., Young, S.Y., Mitchell, A.L., 2017. InterPro in 2017 – beyond protein family and domain annotations. Nucleic Acids Res. 45, D190–D199.

Griffith, A.A., Eng, M., 1921. VI. The phenomena of rupture and flow in solids. Philos. Trans. Roy. Soc. Lond. A 221, 163–198.

Gubbels, M., Katzer, F., Hide, G., Jongejan, F., Shiels, B.R., 2000. Generation of a mosaic pattern of diversity in the major merozoite-piroplasm surface antigen of *Theileria annulata*. Mol. Biochem. Parasitol. 110, 23–32.

Hall, C., Busch, J., Scoles, G., Palma-Cagle, K., Ueti, M., Kappmeyer, L., Wagner, D., 2013. Genetic characterization of *Theileria equi* infecting horses in North America: evidence for a limited source for U.S. introductions. Parasit. Vectors 6, 35.

Hamilton, M.J., Lee, M., Le Roch, K.G., 2014. The ubiquitin system: an essential component to unlocking the secrets of malaria parasite biology. Mol. BioSyst. 10, 715–723.

Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D., Sugimoto, C., 2013. Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. DNA Res. 20, 209–220.

Hegedűs, T., Chaubey, P.M., Várady, G., Szabó, E., Sarankó, H., Hofstetter, L., Sarkadi, B., 2015. Inconsistencies in the red blood cell membrane proteome analysis: generation of a database for research and diagnostic applications. Database. https://doi.org/10.1093/database/bav056.

Ikadai, H., Ishida, H., Sasaki, M., Taniguchi, K., Miyata, N., Koda, M., Igarashi, I., Oyamadaa, T., 2006. Molecular cloning and partial characterization of Babesia equi EMA-3. Mol. Biochem. Parasitol. 150, 371–373.

Jalovecka, M., Hajdusek, O., Sojka, D., Kopacek, P., Malandrin, L., 2018. The complexity of piroplasms life cycles. Cell Infect. Microbiol. 8, 248.

Jenkins, C., Bogema, D.R., 2016. Factors associated with seroconversion to the major piroplasm surface protein of the bovine haemoparasite *Theileria orientalis*. Parasit Vectors 9, 106.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Lear, T., 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. PNAS 111, 18655–18660.

Jordan, G., Goldman, N., 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol. Biol. Evol. 29, 1125–1139.

Joubert, F., 1977. Snake venom toxins. The aminoacid sequences of three toxins (9B, 11 and 12A) from Hemachatus haemachatus (Ringhals) Venom. Euro. J. Biochem. 74, 387–396.

Kappmeyer, L.S., Perryman, L.E., Knowles, D.P., 1993. A *Babesia equi* gene encodes a surface protein with homology to Theileria species. Mol. Biochem. Parasitol. 62, 121–124.

Kappmeyer, L.S., Thiagarajan, M., Herndon, D.R., Ramsay, J.D., Caler, E., Djikeng, A., Gillespie, J.J., Lau, A.O.T., Roalson, E.H., Silva, J., Suarez, C.E., Ueti, M.W., Nene, V.M., Mealey, R.H., Knowles, D.P., Brayton, K.A., 2012. Comparative genomic analysis and phylogenetic position of *Theileria equi*. BMC Genomics 13, 603.

Katzer, F., McKellar, S., Ferguson, M., d'Oliveira, C., Shiels, B.R., 2002. A role for tertiary structure in the generation of antigenic diversity and molecular association of the Tams1 polypeptide in *Theileria annulata*. Mol. Biochem. Parasitol. 122, 55–67.

Knowles Jr., D.P., Perryman, L.E., Goff, W.L., Miller, C.D., Harrington, R.D., Gorham, J.R., 1991a. A monoclonal antibody defines a geographically conserved surface protein epitope of *Babesia equi* merozoites. Infect. Immun. 59, 2412–2417.

Knowles Jr., D.P., Perryman, L.E., Kappmeyer, L.S., Hennager, S.G., 1991b. Detection of equine antibody to *Babesia equi* merozoite proteins by a monoclonal antibody-based competitive inhibition enzyme-linked immunosorbent assay. J. Clin. Microbiol. 29, 2056–2058.

Knowles Jr., D.P., Kappmeyer, L.S., Stiller, D., Hennager, S.G., Perryman, L.E., 1992. Antibody to a recombinant merozoite protein epitope identifies horses infected with *Babesia equi*. J. Clin. Microbiol. 30, 3122–3126.

Knowles Jr., D.P., Kappmeyer, L.S., Perryman, L.E., 1994. Specific immune responses are required to control parasitemia in *Babesia equi* infection. Infect. Immun. 62, 1909–1913.

Knowles, D.P., Kappmeyer, L.S., Perryman, L.E., 1997. Genetic and biochemical analysis of eryth-rocyte-stage surface antigens belonging to a family of highly conserved proteins of *Babesia equi* and Theileria species. Mol. Biochem. Parasitol. 90, 69–79.

Knowles, D.P., Kappmeyer, L.S., Haney, D., Herndon, D.R., Fry, L.M., Munro, J.B., Sears, K., Ueti, M.W., Wise, L.N., Silva, M., Schneider, D.A., Grause, J., White, S.N., Tretina, K., Bishop, R.P., Odongo, D.O., Pelzel-McCluskey, A.M., Scoles, G.A., Mealey, R.H., Silva, J.C., 2018. Discovery of a novel species, *Theileria haneyi* n. sp., infective to equids, highlights exceptional genomic diversity within the genus Theileria: implications for apicomplexan parasite surveillance. Int. J. Parasitol. 48 (9–10), 679–690.

Kumar, S., Yokoyama, N., Kim, J., Bork-Mimm, S., Inoue, N., Xuan, X., Igarashi, I., Sugimoto, C., 2012. Theileria equi merozoite antigen-2 interacts with actin molecule of equine erythrocyte during their asexual development. Exp. Parasit. 4, 508–512.

Kumar, S., Yokoyama, N., Kim, J.-Y., Huang, X., Inoue, N., Xuan, X., Igarashi, I., Sugimoto, C., 2004. Expression of *Babesia equi* EMA-1 and EMA-2 during merozoite developmental stages in erythrocyte and their interaction with erythrocytic membrane skeleton. Mol. Biochem. Parasitol. 233, 221–227.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lording, P.M., 2008. Erythrocytes. VCNA Eq. 24, 225–237.

Löytynoja, A., Goldman, N., 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science 320, 1632–1635.

Löytynoja, A., Goldman, N., 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinform. 11, 579.

Marsolier, J., Perichon, M., DeBarry, J., Villoutreix, B., Chluba, J., Lopez, T., Garrido, C., Zhou, X., Lu, K., Fritsch, L., Ait-Si-Ali, A., Mhadhbi, M., Medjkane, S., Weitzman, J., 2015. Theileria parasites secrete a prolyl isomerase to maintain host leukocyte transformation. Nature 520, 378–382.

Massingham, T., Goldman, N., 2005. Detecting amino acid sites under positive selection and purifying selection. Genetics 169, 1753–1762.

Melhorn, H., Schein, E., 1998. Redescription of *Babesia equi* Laveran, 1901 as *Theileria equi*. Parasitol. Res. 84, 467–475.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2016. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C.A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R.M., Cronin, A., de Villiers, E.P., Fraser, A., Fosker, N., Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D.E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neil, S., Price, C., Quail, M.A., Rabbinowitsch, E., Rawlings, N.D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A.R., Woodward, J., Dobbelaere, D.A., Langsley, G., Rajandream, M.A., McKeever, D., Shiels, B., Tait, A., Barrell, B., Hall, N., 2005. Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. Science 309, 131–133.

Pal-Bhowmick, I., Andersen, J., Srinivasan, P., Narum, D.L., Bosch, J., Miller, L.H., 2012. Binding of aldolase and glyceraldehyde-3-phosphate dehydrogenase to the cytoplasmic tails of *Plasmodium falciparum* merozoite duffy binding-like and reticulocyte homology ligands. MBio 3 e00292–12.

Ramsay, J.D., Ueti, M.W., Johnson, W.C., Scoles, G.A., Knowles, D.P., Mealey, R.H., 2013. Lymphocytes and macrophages are infected by *Theileria equi*, but T cells and B cells are not required to establish infection in vivo. PLoS One 8, e76996.

Razavi, S.M., Nazifi, S., Bateni, M., Rakhshandehroo, E., 2011. Alterations of erythrocyte antioxidant mechanisms: antioxidant enzymes, lipid peroxidation and serum trace elements associated with anemia in bovine tropical theileriosis. Vet. Parasitol. 180, 209–214.

Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutíerrez, S., Huerta-Cepas, J., Gabaldón, T., Dopazo, J., Dopazo, H., 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res. 39, W470–W474.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.

Schneider, I., Haller, D., Kullmann, B., Beyer, D., Ahmed, J.S., Seitzer, U., 2007. Identification, molecular characterization and subcellular localization of a *Theileria annulata* parasite protein secreted into the host cell cytoplasm. Parasitol. Res. 101, 1471–1482.

Shiono, H., Yagi, Y., Chikayama, Y., Miyazaki, S., Nakamura, I., 2003. Oxidative damage

and phosphatidylserine expression of red blood cells in cattle experimentally infected with *Theileria sergenti*. Parasitol. Res. 89, 228.
Silva, M., Graca, T., Suarez, C., Knowles, D., 2013. Repertoire of *Theileria equi* immunodominant antigens bound by equine antibody. Mol. Biochem. Parasitol. 188, 109–115.
Skalak, R., Zarda, P.R., Jan, K.M., Chien, S., 1981. Mechanics of rouleau formation. Biophys. J. 35, 771–781.
Skilton, R., Musoke, A., Wells, C., Yagi, Y., Nene, V., Spooner, P., Gachanja, J., Osaso, J., Bishop, R.P., Morzaria, S., 2000. A 32 kDa surface antigen of Theileria parva: characterization and immunization studies. Parasitology 120, 553–564.
Takemae, H., Sugi, T., Kobayashi, K., Murakoshi, F., Recuenco, F.C., Ishiwa, A., Inomata, A., Horimoto, T., Yokoyama, N., Kato, K., 2014. Analyses of the binding between *Theileria orientalis* major piroplasm surface proteins and bovine red blood cells. Vet. Rec. 175, 149.
Utkin, Y.N., 2013. Three-finger toxins, a deadly weapon of elapid venom–milestones of discovery. Toxicon 62, 50–55.
Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Blöcker, H., 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326, 865–867.
Wagener, J., Schneider, J.J., Baxmann, S., Kalbacher, H., Borelli, C., Nuding, S., Braunsdorf, C., 2013. A peptide derived from the highly conserved protein GAPDH is involved in tissue protection by different antifungal strategies and epithelial immunomodulation. J. Invest. Dermatol. 133, 144–153.
Weir, W., Karagenç, T., Baird, M., Tait, A., Shiels, B.R., 2010. Evolution and diversity of secretome genes in the apicomplexan parasite *Theileria annulata*. BMC Genomics 11, 42.
Wise, L.N., Kappmeyer, L.S., Mealey, R.H., Knowles, D.P., 2013. Review of equine piroplasmosis. J. Vet. Intern. Med. 27, 1334–1346.
Wise, L.N., Kappmeyer, L.S., Silva, M.G., White, S.N., Grause, J.F., Knowles, D.P., 2017. Verification of post-chemotherapeutic clearance of Theileria equi through concordance of nested PCR and immunoblot. Ticks Tick Borne Dis. 9, 135–140.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Gil, L., 2018. Ensembl 2018. Nucleic Acids Res. 46, D754–D761.
Zhao, S., Guan, G., Liu, J., Liu, A., Li, Y., Yin, H., Luo, J., 2017. Screening and identification of host proteins interacting with *Theileria annulata* cysteine proteinase (TaCP) by yeast-two-hybrid system. Parasit. Vectors 10, 536.