Ticks and Tick-borne Diseases 17 (2026) 102656

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

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

Original article

Theileria equi sensu lato genetic diversity and phylogeny: new insights from mitochondrion, apicoplast and nucleus multilocus gene analysis

Mickaël Mège, Florian Berquier, Claire Bonsergent, Laurence Malandrin *

Oniris, INRAE, UMR 1300 BIOEPAR, 44300 Nantes, France

ARTICLE INFO

Keywords:
Equine piroplasmosis
Cryptic species
Theileria haneyi
Lineage
ema genes

ABSTRACT

Equine piroplasmosis is a significant tick-borne disease with a global distribution. Among the two causative agents, *Theileria equi* has undergone numerous taxonomic reclassifications over time, reflecting advances in our understanding of its biology, molecular characteristics, and phylogenetic relationships. Formerly named *Babesia equi*, now classified in the Equus group apart from *Babesia* and *Theileria* species, its genus name will again certainly change. The genetic diversity also within *T. equi* lead to the renaming of one lineage as *Theileria haneyi*. *Theileria equi* should better now be considered as *T. equi* sensu lato (including *T. haneyi*) with a set of different cryptic species awaiting for formal taxonomical changes to be taken. A better knowledge within this group is therefore essential as this diversity has an impact on the diagnostic of this disease which is submitted to strict regulations governing the international movement of horses. In the present study, we designed primers to study the same five genes from the nucleus (*hsp70, clamp* and *ama-1*), the mitochondria (*cox1*) and the apicoplast (*tufA*), on a set of 36 mono-infected samples of the lineages A, C (*T. haneyi*) and E of *T. equi* sensu lato. The phylogenetic separation in lineages A, C and E based on 18S rRNA sequencing was confirmed by this multi-locus analysis. Genetic diversity and phylogenetic separation within lineage A into two sub-lineages A1 and A2 were demonstrated on the basis of all five genes, with an additional A3 sub-lineage on the basis of nuclear (*hsp70*) and mitochondrial (*cox1*) markers. All these sub-lineages would deserve the rank of species based on *cox1* divergence confirmed by other markers. Phylogenetic and haplotype networks analyses were also conducted on genes used in molecular or serological diagnostic, *ema-1* and *ema-11*, to evaluate the impact of this diversity on currently used tests. We confirmed the absence of molecular detection of the lineage E DNA with any of these two gene targets. We linked the sub-lineages A1 and A3 to the *ema-1* clades A and B, with a worldwide distribution. We confirmed the link of the rare A2 sub-lineage from France to a new *ema-1* clade D, with additional Eurasian sequences from Croatia, Mongolia and China. A fifth *ema-1* clade E was also confirmed. Separation of lineage C (*T. haneyi*) into at least two sub-lineages was also confirmed with the few *ema-11* sequences available, based on their genetic diversity, and their phylogenetic and haplotype network analyses. From this study, it becomes clear that the division of *T. equi* sensu lato into species corresponding either to the lineage (5 species) or to the sub-lineage (at least 10) should await a more comprehensive knowledge of its genetic diversity worldwide, as it may impact the diagnostic tools to be developed.

1. Introduction

Equine piroplasmosis is significant tick-borne disease affecting equids (horses, donkeys, mules, and zebras). It is caused by intra-erythrocytic apicomplexan parasites, primarily *Theileria equi* and *Babesia caballi* that can be transmitted by various tick genera (De Waal, 1992; Scoles and Ueti, 2015). From a health perspective, the multiplication of these parasites can lead to a wide spectrum of clinical outcomes in equids, ranging from mild symptoms like lethargy, anorexia, and dyspnea to severe and potentially fatal acute manifestations such as fever, colic, anemia and jaundice (Onyiche et al., 2019; Tirosh-Levy et al., 2020; Wise et al., 2013). None of these symptoms are pathognomonic for *B. caballi* or *T. equi* infection, which complicates differential diagnosis. Following infection, horses can become asymptomatic carriers. While *B. caballi* parasites may be eliminated after several years (approximately 4 years) or following treatment, *T. equi* establishes a persistent, lifelong infection, even after therapeutic intervention (Onyiche et al., 2019; Tirosh-Levy et al., 2020). Economically, equine

* Corresponding author at: UMR BIOEPAR, Oniris, site de la Chantrerie 44300 Nantes, France.
E-mail address: laurence.malandrin@inrae.fr (L. Malandrin).

https://doi.org/10.1016/j.ttbdis.2026.102656
Received 30 September 2025; Received in revised form 30 April 2026; Accepted 1 May 2026
Available online 22 May 2026

piroplasmosis incurs losses due to treatment costs, animal fatalities, reduced performance, and significant negative impacts on international trade (Mendoza et al., 2024; Rothschild, 2013; Tamzali, 2013).

Equine piroplasmosis is now considered endemic in most tropical, subtropical, and temperate regions worldwide (Onyiche et al., 2019; Tirosh-Levy et al., 2020). Conversely, countries such as Australia, Canada, Japan, and New Zealand report being free of the disease, while the United States maintains a non-endemic status characterized by limited outbreaks following the importation of infected horses (Tirosh-Levy et al., 2020). Consequently, the World Organisation for Animal Health (WOAH) has classified equine piroplasmosis as a notifiable disease, mandating screening for horses participating in international events or being exported to disease-free countries (OIE, 2024). For reliable diagnosis, a combined approach is recommended, integrating serological techniques such as Indirect Fluorescent Antibody Test (IFAT) and Competitive Enzyme-Linked Immunosorbent Assay (cELISA), with molecular biology methods like nested Polymerase Chain Reaction (nested PCR) and quantitative Polymerase Chain Reaction (qPCR) (Mendoza et al., 2024; OIE, 2024; Scoles and Ueti, 2015).

The advancement of molecular tools has revolutionized our understanding of haemoprotozoan parasites, shifting the focus from simple pathogen detection to high-resolution molecular epidemiology and characterization. Molecular identification has historically relied on the 18S rRNA gene, and this approach has unveiled substantial intra-specific diversity. Five lineages (genotypes) of *T. equi* (A to E) and three of *B. caballi* (A, B1, and B2) are currently recognized, following the nomenclature proposed by Tirosh-Levy et al. (2020). The term lineage will be preferred in the whole manuscript to replace the term genotype used in the literature in agreement with and as recommended by Schnittger et al. (2022). Analysis of numerous sequences demonstrates a correlation between certain *T. equi* 18S rRNA lineages and their geographical origin. For instance, lineage B has been detected in Africa and the Mediterranean region; lineage D in Africa, the Mediterranean, and the Middle East; and lineage E primarily in Asia and Europe. Conversely, lineages A and C exhibit a global distribution (Tirosh-Levy et al., 2020).

The recent naming of *Theileria haneyi,* corresponding to the 18S rRNA lineage C, suggests the potential presence of multiple cryptic species or subspecies within the broader *T. equi* complex (Bhoora et al., 2010; Dahmana et al., 2019; Knowles et al., 2018; Mège et al., 2025), a complex that should be viewed as *Theileria equi* sensu lato until appropriate data is produced and taxonomy changes are validated. To overcome the limited resolution of 18S rRNA sequencing, other genetic markers, such as *ema-1* gene (*T. equi*) or *rap-1* (*B. caballi*) are often employed. These genes also serve as targets for serological and molecular diagnostic techniques. However, their occurrence appears to be restricted to specific lineages in both *T. equi* and *B. caballi*, potentially leading to inaccurate diagnoses (Knowles et al., 2018; Bhoora et al., 2010; Mège et al., 2025). Furthermore, in certain endemic areas, the frequent occurrence of co-infections with different lineages complicates the clear attribution of sequences to an 18S rRNA lineage (Bhoora et al., 2020; Mège et al., 2025).

Most Apicomplexan parasites possess three distinct genomes: nuclear, mitochondrial and apicoplastic. The two latter ones, remnants of secondary endosymbiosis, exhibit different inheritance strategies and evolutionary histories (Arisue and Hashimoto, 2015; Berná et al., 2021; Lamb et al., 2023). In the present study, we selected genetic markers based on the four key criteria: representing the three distinct genomes, being present in only one copy per cell, being involved in essential cellular functions or erythrocyte invasion to ensure their presence across all lineages, and exhibiting different evolutionary rates, as inferred from studies on other organisms. We applied these markers to a selection of *T. equi* sensu lato samples from asymptomatic horses that were previously confirmed as being mono-infected, thereby avoiding confusion in sequence attribution (Mège et al., 2025).

Specifically, the *tufA* gene (elongation factor Tu) and *cox-1* gene (cytochrome C oxidase subunit 1) were selected to represent the apicoplast and the mitochondrion genomes, respectively (Saxena et al., 2012; Schreeg et al., 2016; Sivakumar et al., 2022). Both genes code for cellular essential function (translation and electron transport chain respectively), with *cox-1* also serving as a barcoding reference gene for species delineation (Antil et al., 2023; Herbert et al., 2003). Their nuclear counterpart selected was *hsp70* (heat shock protein 70), a gene target frequently used for phylogenetic reconstruction (Paulino et al., 2024; Schnittger et al., 2012). The two other selected nuclear genes, *ama-1* (apical membrane antigen 1) and *clamp* (claudin-like apicomplexan microneme protein), were chosen due to their critical roles in parasite entry into host cells, their significance as targets for vaccine development and their use in phylogenetic studies (Bonsergent et al., 2021; Remarque et al., 2008; Triglia et al., 2000; Tyler et al., 2011). AMA-1 is a transmembrane protein central to moving junction formation, while CLAMP facilitates parasite progression through the host cell membrane, a key step for establishing infection (Onzere et al., 2021). CLAMP has also been described as essential for the invasion of mosquito vector salivary glands by *Plasmodium* (Loubens et al., 2023).

On these well-characterized samples, we chose more specific gene targets (*ema-1* and *ema-11*) to analyze sublineage genetic diversity. The *ema-1* and *ema-11* genes are members of the merozoite surface antigens (EMAs) family, which may have a role in erythrocyte invasion, parasite survival and propagation (Wise et al., 2019). These antigens are immunologically significant and commonly used in serological diagnostics due to their strong immunogenicity, although they have the known inconvenience of exhibiting lineage-selective occurrence (Bhoora et al., 2010; Knowles et al., 2018).

In summary, the integrated analysis of these seven genes on 18S rRNA-typed samples should provide an essential foundation for improving diagnostic methods, refining taxonomy, and guiding prevention and control efforts against equine piroplasmosis globally.

2. Materials and methods

2.1. Sampling

For this study, we selected 36 *T. equi* infected blood samples collected from 2020 to 2024, from asymptomatic horses, mainly from metropolitan France but also from Guadeloupe in the Caribbeans, to create a sample set representative of the genetic diversity of *T. equi*. Most of these samples (21/36) were genotyped with 18S rRNA sequencing from three previous studies (Hermans et al., 2025; Jouglin et al., 2025; Mège et al., 2025). In metropolitan France, 15 additional samples were selected either to be representative of the geographical repartition (lineage E), or to add underrepresented lineages (A and C) to the study. The final dataset includes eight blood samples infected with *T. equi* lineage A (six from metropolitan France and two from Guadeloupe), seven with *T. equi* lineage C (six from Guadeloupe and one from metropolitan France), and 21 with *T. equi* lineage E from metropolitan France. Three supplementary blood samples from Guadeloupe identified as infected with the lineage A were included in the *ema-1* analysis (Mège et al., 2025). Detailed description of the sample origins is given in Table 1 and in Fig. 1.

2.2. DNA extraction

The blood samples used in this study were collected from asymptomatic horses and prepared as previously described (Hermans et al., 2025; Jouglin et al., 2025; Mège et al., 2025).

For each sample, genomic DNA was extracted from 100 µL of packed erythrocytes and buffy coat diluted 1:1 in 1X PBS, using the NucleoSpin® Blood kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. The extracted DNA was then stored at −20 °C until further use. DNA quantification and purity assessment were not performed, as the total genomic DNA concentration in these samples is

Table 1
Horse blood sample references, origins and lineages as determined by the 18S rRNA sequencing and by the other analyzed genes for the sub-lineages A1 and A2.

Sample Reference	County	Date	Lineage or sub-lineage
CVA20–26	Yonne	2020	A1
CVG24–24	Guadeloupe	2024	
CVG24–38	Guadeloupe	2024	
CVG24–71[a]	Guadeloupe	2024	
CVG24–106[a]	Guadeloupe	2024	
CVG24–130[a]	Guadeloupe	2024	
PS163–1	Dordogne	2022	
PS269–1	Loir-et-Cher	2022	
J23–20	Eure	2023	
CVT21–14	Gers	2021	A2
CVL21–18	Rhône	2021	
PS103–1	Essonne	2021	C
CVG24–30	Guadeloupe	2024	
CVG24–34	Guadeloupe	2024	
CVG24–74	Guadeloupe	2024	
CVG24–86	Guadeloupe	2024	
CVG24–108	Guadeloupe	2024	
CVG24–136	Guadeloupe	2024	
CV19–49	Indre-et-Loire	2019	E
CV20–61	Deux-Sèvres	2020	
CVT20–12	Haute-Garonne	2020	
CV20–62	Maine-et-Loire	2020	
PS131–1	Ardèche	2021	
PS90–1	Doubs	2021	
PS237–1	Drôme	2021	
CVL21–16	Loire	2021	
CVT21–62	Lot	2021	
CVA21–02	Paris	2021	
CVT21–77	Pyrénées-Orientales	2021	
PS184–1	Yonne	2021	
CVA21–08	Yvelines	2021	
PS399–11	Aisne	2022	
PS207–4[b]	Bas-Rhin	2022	
PS1–1	Gironde	2022	
PS71–2	Ille-et-Vilaine	2022	
PS185–1[b]	Meurthe-et-Moselle	2022	
PS23–2	Var	2022	
PS276–2	Cher	2023	
PS372–3	Creuse	2023	

[a] *ema-1* and 18S rRNA amplification only.
[b] no amplification of *cox 1*.

predominantly composed of host (equine) DNA. Given the carrier status and expected low parasitemia of the sampled horses, total DNA measurements do not accurately reflect the quantity or quality of the target parasite DNA available for downstream applications.

2.3. 18S rRNA genotyping

For this study, the 18S rRNA gene of all 36 samples was amplified. Briefly, a first PCR was used to amplify piroplasms 18S rRNA partial gene using CRYPTOF and CRYPTOR primers (Malandrin et al., 2010). A nested PCR was then used to separately amplify *B. caballi* 18S rRNA with the specific primers BabF4 and BabR4 (856 bp) or *T. equi* 18S rRNA with the primers TeqF3 and TeqR5 (1470 bp) (Jouglin et al., 2025).

To minimize the risk of including co-infected samples, which could complicate sequence attribution, 18S rRNA chromatograms were carefully analysed and all specimens were pre-screened using lineage-specific primers for A and C, as described by Mège et al. (2025).

2.4. Primer design

Due to the low concentration of target DNA in all samples, a nested PCR approach was developed to amplify each gene. During primer design, the main challenge was the lack of available data regarding the target genes across the different lineages included in our study, particularly for lineage E. The only sources providing comprehensive information for all genes of interest were the three genomes available in public databases: WA and NVSL354 for lineage A, and Eagle Pass for lineage C/*T. haneyi* (Kappmeyer et al., 2012; Grimsley et al., 2023; Knowles et al., 2018).

Our approach therefore consisted of identifying, for each target gene, conserved regions between *T. equi* lineages A and C based on the available genomes, in order to maximize the chances of successfully amplifying these genes in lineage E as well.

The first step was to search for the probable sequences of the *tufA* (elongation factor Tu), *cox1* (cytochrome c oxidase subunit 1), *hsp70* (heat shock protein 70), *ama-1* (apical membrane antigen 1), and *clamp* (claudin-like microneme protein in Apicomplexa) genes in the data from Kappmeyer et al. (2012) available through the public PiroplasmaDB database (VEuPathDB project).

In the second step, to ensure that the retrieved sequences corresponded to the target genes, they were aligned with the genomes of *T. equi* WA and NVSL354 (GCA_000342415 and GCA_024610995), as well as with that of *T. equi/T. haneyi* (GCA_002189625), using the Bowtie2 mapping tool (v2.4.5) integrated in Geneious Prime® software v2025.2.2 (Dotmatics, USA). Open reading frames (ORFs) corresponding to the target genes were predicted. The resulting complete sequences were then verified through a BLASTX alignment against the NCBI database to confirm that the predicted ORFs matched the expected gene products.

For the third step, gene sequences from the three genomes were aligned using Clustal Omega (v1.2.2) within Geneious. Based on these alignments, sets of gene-specific primers were manually designed to amplify at least 400 bp of each gene after the nested PCR. Several couples of primers were designed and tested to select the most appropriate sets allowing amplification of each gene for the three lineages A, C and E. PCR positive amplification controls consisted in lineage A and lineage C DNA samples.

Intronic regions in the *clamp* gene were also identified at this stage, as the reference sequence from PiroplasmaDB originated from transcriptomic data.

For the *ema-1* gene, the previously nPCR protocol (Baptista et al., 2013) was not selected first because it raises a 229 bp fragment, that we considered as being too short; and second because we were previously not able to amplify *ema-1* from *T. equi* lineage E with this primer set (Mège et al., 2025). We therefore tested several other published or home-made primers at different locations on the gene that will be detailed further in Section 3.3.3., to get a longer fragment for the lineage A infected samples and to further highlight the absence of *ema-1* amplification in the lineage E infected samples.

For the *ema-11* gene, a different strategy was adopted. External primers previously described by Bhoora et al. (2025) were used, and internal primers were designed to enable nested PCR (nPCR).

All primers were evaluated for potential self-dimers and hairpin structures using the OligoCalc web tool (Kibbe, 2007). Melting temperatures (Tm) were estimated using Promega's Tm Calculator, with the magnesium ion concentration set to 2 mM.

The description of selected primers is given in Table 2.

2.5. Gene amplification and sequencing

All nested PCR (nPCR) reactions were performed using a SimpliAmp™ thermal cycler (ThermoFisher Scientific, USA) and the GoTaq® G2 Flexi DNA Polymerase kit (Promega, USA).

The first-round PCR was carried out in a 30 µL reaction volume containing: 1X GoTaq® Flexi colorless reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTP mix (Promega, USA), 0.5 µM of each of forward and reverse primers, 1 unit of GoTaq® Flexi DNA polymerase, 7 µL of genomic DNA template, and nuclease-free water (Promega, USA) to adjust the final volume.

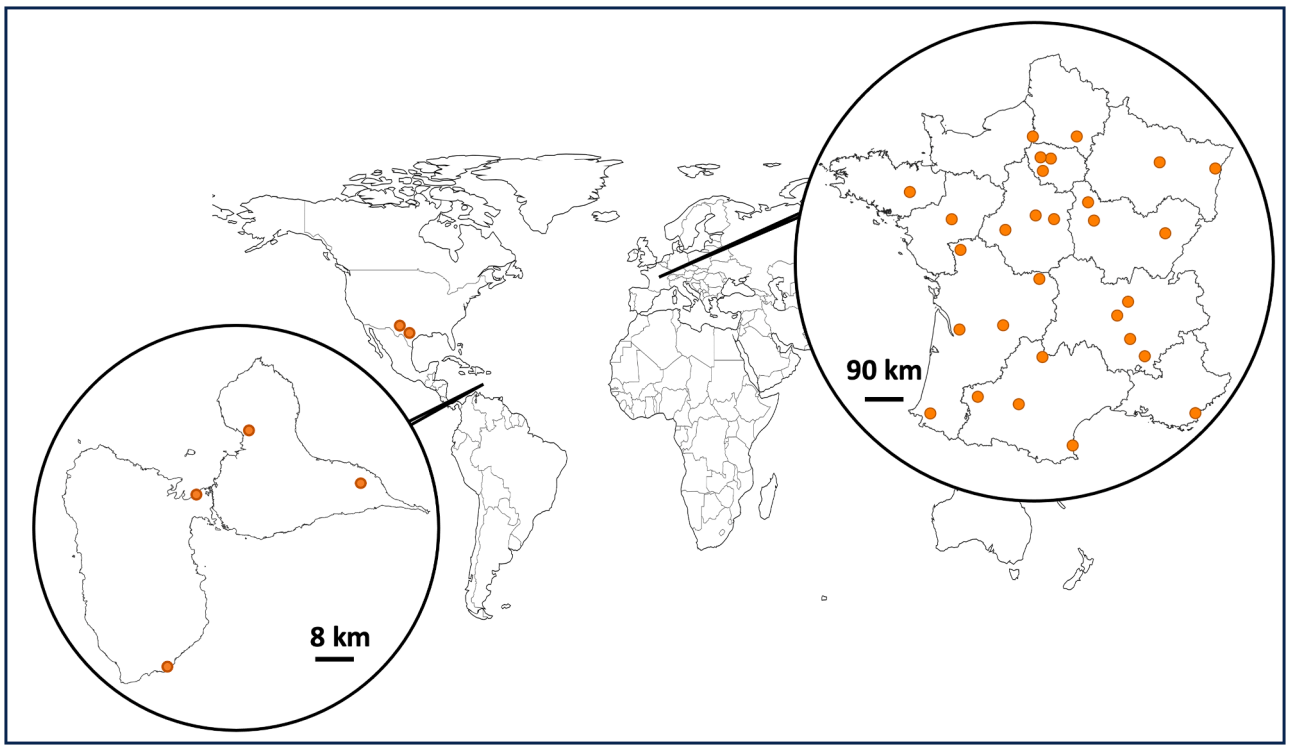

Fig. 1. Maps showing the localization of the samples included in the multilocus sequence analysis, from metropolitan France, Guadeloupe, and USA (reference genome sequences).

Table 2
Nested PCR primers, amplification conditions and amplicon sizes of *Theileria equi* sensu lato genes analyzed in this study. All primers were designed in this study, except ThEMA11* (Bhoora et al., 2025).

Genome	Gene	Primer name	Primer sequence (5′ - 3′)	Primer length (bp)	Tm	Elongation time	Amplicon size (bp)
Apicoplast	*tufA*	tufA_Out_F1	GGAACCATAGGACACATAGAC	21	57 °C	1 min 15 s	949
		tufA_Out_R2	GTTTATGTCTTCCTCCTTCTTCT	23			
		tufA_Inn_F3	AAGACACTACGCACATATAGACT	23		45 s	477
		tufA_Inn_R3	CCTCTACCTGTTATGGAGAAACT	23			
Mitochondria	*cox-1*	cox-1_Out_F4	GTAAYCACAAGATAATMGGATTAGG	25	57 °C	1 min	881
		cox-1_Out_R3	ACTTCTRCTATCAATTTCCATACC	24			
		cox-1_Inn_F5	AGRACWGAATTAAGTATGAGTGG	23		45 s	619
		cox-1_Inn_R6	CTTCAGGATGMCCAAAYAAC	20			
Nucleus	*hsp70_chr1*	hsp70_chr1_Out_F2	TCGATGCYAAGAGGTTGATTG	21	57 °C	2 min	1710
		hsp70_chr1_Out_R2	CAGTAGGACCACCACTTGG	19			
		hsp70_chr1_Inn_F6	TTCCAGGGTCARAAGAAGACA	21		1 min 30 s	1334
		hsp70_chr1_Inn_R1	CTCATGCTGTAGCAGTAATTTTC	23			
	ama-1	ama-1_Out_F1	GACGACTACAACAAGACCTG	20	57 °C	1 min 15 s	1034
		ama-1_Out_R5	TTTGAGATTCTCACGGGTGTG	21			
		ama-1_Inn_F6	AAGTACTGTGAYGCTGATGGT	21		1 min	958
		ama-1_Inn_R6	GTGTGACTTCAGAGGCTC	18			
	clamp	clamp_Out_F1	TGCCCACATTGAAGGTGTATT	21	57 °C	1 min 15 s	1026
		clamp_Out_R2	ACAACTATATGCACAAGGGAAG	22			
		clamp_Inn_F5	ATTTGCATCATRGGAGCYAAGTT	23		1 min	874
		clamp_Inn_R1	CATCGTACAAGAAACCAGATATG	23			
	ema-1	ema-1_Out_F1	GCATCCTCGCCGAGGAG	17	58 °C	1 min	724
		ema-1_Out_R1	GCGCCATAGACGGAGAAG	18			
		ema-1_Inn_F2	TCGTCGACTTCCAGTTGGA	19		45 s	552
		ema-1_Inn_R2	AGATTTCCTTGATTCTGCCATC	22			
	ema-11	ThEMA11_for*	GGAGACAGAGAAGCCATCAAA	21	58 °C	1 min	739
		ThEMA11_rev*	CTTTGCCACCTCCAGAGTATT	21			
		ThEMA11_Inn_F1	CACTCTGGTCCATGTCTTG	19		45 s	579
		ThEMA11_Inn_R1	AAGGATGTATTGGATCATGTG	21			

Thermal cycling conditions included an initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 30 s, annealing at either 57 °C or 58 °C for 30 s, and extension at 72 °C for a gene-specific duration (Table 2). A final extension step was performed at 72 °C for 5 min.

Ten microliters of the 1:10 diluted first-round PCR product was then used as the template for the second-round (nested) PCR, which was performed in a 30 µL reaction volume, replacing the colourless buffer with the green reaction buffer provided in the same kit. Reaction components and thermal cycling conditions were identical to those used in the first round, with gene-specific adjustments to the extension time (Table 2). Second-round PCR products were visualized by electrophoresis on 1.5% agarose gels containing SYBR® Safe DNA Gel Stain (Invitrogen, USA). Gel images were acquired using a Gel Doc XR+ imaging system (Bio-Rad, USA).

PCR products were subsequently purified using the ExoSAP-IT™ PCR Product Cleanup kit (Applied Biosystems, USA) and submitted for Sanger sequencing (Eurofins Genomics, Germany). Amplicons were sequenced in both directions.

Consensus sequences for each target gene were generated using Geneious.

2.6. Sequence and phylogenetic analyses

Nucleotidic consensus sequences generated for each target gene were first aligned using Clustal Omega v1.2.2, integrated into Geneious, to allow for manual correction prior to downstream analysis. These alignments enabled the determination of nucleotide sequence homology percentages.

These nucleotide sequences were translated with Geneious using the correct open reading frame, and the resulting amino acid sequences were aligned by Clustal Omega v1.2.2. This step was used to calculate the number of identical sites and the pairwise percentage identity.

In the case of *clamp*, the introns' position of our amplified sequences was determined according to the available intron-less sequence of the lineage A from the WA strain (XM_004830748). In case of apparent variation on their length and position, the intron positions were determined based on amino acid alignments with the lineage A sequence, and the presence of the dinucleotide GU at its 5′ end and an AG at its 3′ end.

Phylogenetic analyses were conducted using the FastTree/OneClick workflow available on the NGPhylogeny.fr platform (Lemoine et al., 2018). Briefly, this pipeline involves the following steps: initial sequence alignment with MAFFT (Multiple Alignment using Fast Fourier Transform), alignment trimming using BMGE (Block Mapping and Gathering with Entropy), and tree inference via maximum likelihood using FastTree.

While default parameters were used for the initial processing steps, tree inference was specifically conducted using the GTR+G substitution model for nucleotide sequences and the LG+G substitution model for protein sequences. The reliability of the branches was assessed with 1000 bootstrap replicates, and final trees were exported in Newick format.

For the *tufA, cox-1, hsp70, ama-1*, and *clamp* genes, sequences from *B. caballi* genome were used as outgroups to root the phylogenetic trees. In the case of the *T. equi ema-1* gene, the tree was rooted using the *T. equi ema-11* gene sequence, and vice versa. This approach, utilizing closely related paralogs from the *ema* multigene family, was chosen to ensure a stable root and high sequence homology for alignment, while clearly differentiating the two distinct gene clusters.

The resulting trees were then formatted and visualized using TreeViewer v2.2.0 (Bianchini and Sánchez-Baracaldo, 2024).

2.7. Haplotype network construction

The network was generated and visualized using Hapsolutely v.0.2.3 (Vences et al., 2024), employing the TCS algorithm (Templeton, Crandall, and Sing network). For the *ema-1* gene, the network was built using 232 sequences (221 from public databases and 11 from our study) covering a 438 bp region of the gene. For the *ema-11* gene, 22 sequences were used (15 from public databases and 7 from our study) covering a 515 bp region of the gene. Details on the origin of all used sequences are provided in the supplementary Table sheet 1 (*ema-1* haplotype network).

3. Results

3.1. Amplification and features of the 5 genes present in all lineages

In this study, the set of ten primers successfully amplified partial sequences of the *tufA, cox1, hsp70, ama-1*, and *clamp* genes by nested PCR for all 36 selected samples from lineages A, C, and E. Only two *cox1* amplicons from lineage E infected samples could not be obtained. The accession numbers are detailed in the supplementary Table sheet 2.

Detailed amplification conditions, as well as the sizes of the amplicons that were trimmed for further analysis, are provided in Tables 2 and 3. The sequences included in the analysis ranged from 400 to 1135 base pairs, covering 32% to 59% of the full length of the different genes. We generated concatenated sequences of approximately 3.5 kb for all three lineages (Table 3).

Variable sequence lengths were observed for *tufA, ama-1*, and *clamp*. For *tufA*, this variation is due to a 3-base pair (bp) deletion in lineage A, resulting in sequences of 400 bp, while lineages C and E have sequences of 403 bp (supplementary Fig. 1A). For *ama-1*, sequence lengths varied between 827 bp and 836 bp. Two deletions of 3 bp and 6 bp occurred in lineage A sequences, resulting in the shortest sequence (827 bp). The 3 bp deletion was also present in lineage E, yielding an 833 bp sequence, whereas no deletions were found in the 836 bp lineage C sequence. The positions of these mutations are shown in the (supplementary Fig. 1B).

The amplified partial *clamp* sequence successfully covered the five introns previously described in lineage A (XM_004830748). We found that the second intron (41 bp) was absent exclusively in lineage E (Fig. 2). Furthermore, the sizes of the introns varied between lineages, and even within lineage A, as seen with the fourth intron, which was either 31 or 32 bp long.

Table 3
Amplicon coverage rate relative to the full length of the genes studied in *Theileria equi* sensu lato. The reference gene sizes for each gene correspond to those in the WA genome of lineage A or the *Theileria haneyi* genome (*ema-11*).

Gene	Gene ORF size (bp)	Amplicon size (bp)	Trimmed sequence size (bp)	Coverage (%)
tufA	1230	477	400–403	32.5–32.8
cox-1	1605	619	553	34.5
hsp70	1926	1334	1135	58.9
ama-1	1717	958	827–836	48.2–48.7
clamp	1443	874	766–787 (with introns)	53.1–54.5
	1226		562 (without introns)	45.8
ema-1	819	552	485	59.2
ema-11	885	579	515	58.2

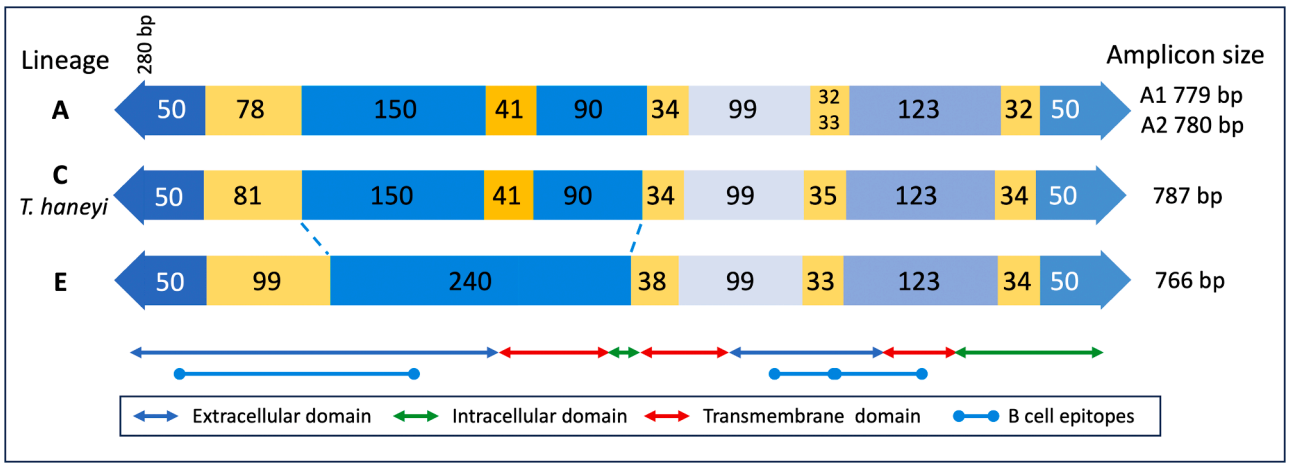

Fig. 2. Schematic representation of the partial amplified *clamp* gene for the three *Theileria equi* sensu lato lineages A, C and E. The position of the sequence compared to the whole gene of the genome reference isolate WA is indicated (280 bp). The lengths and positions of introns (yellow-orange) and exons (blue) are indicated. The rough positions of the protein domains as well as B epitopes are also shown (according to Onzere et al., 2021).

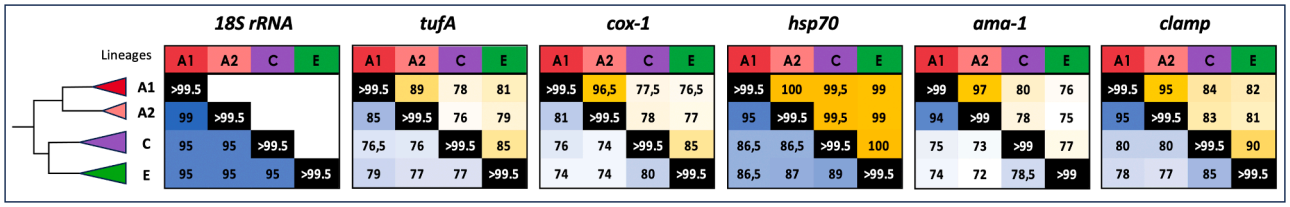

Fig. 3. Simplified phylogenetic representation of the three studied lineages A (A1 and A2), C (*Theileria haneyi*) and E (left) and similarity matrices of nucleotide (blue background) and protein (orange background) homology percentages for each studied lineage, categorized by analyzed genes. The color intensity varies according to the similarity percentage. The data includes our dataset and the available sequences from the three sequenced genomes (WA, NVSL354 and Eagle Pass). The similarity values are averaged as multiple and variable sequences are represented in each lineage and sub-lineage.

3.2. Molecular characterization and phylogenetic differentiation of lineages A, C, and E

The sequences from each gene were first compared to evaluate the intra-lineage genetic diversity. For lineages C and E, sequences from all genes—apicoplast (*tufA*), mitochondrial (*cox1*), and nuclear (*hsp70, ama-1*, or *clamp*)—were highly similar, with a similarity rate often exceeding 99.5% except *ama-1* which is detailed below (Fig. 3). In fact, most of these sequences were identical, and the observed variability was primarily due to unique substitutions in single sequences.

Within lineage E, *ama-1* was the most variable gene. Eight positions within the 21 *ama-1* sequences showed polymorphisms (supplementary Fig. 1B), separating the samples into different subgroups without any clear structure (phylogenetic tree presented in supplementary Fig. 2). The sequence similarities for *ama-1* therefore ranged from 99.04% to 100% within the lineage E. This translates to a single amino acid change in one 277-amino-acid sequence. For the other genes, only a few polymorphic positions were found: two in *clamp* and one in *tufA, cox1*, and *hsp70*. Notably, all of these changes were silent mutations, meaning they did not result in a change to the amino acid sequence.

In this study, we found that the sequences of all five genes from the seven blood samples classified as being infected with *T. haneyi* (lineage C) were highly similar, and often identical, to the *T. haneyi* genome reference sequence (Eagle Pass GCA_002189625). Within lineage C, two polymorphic positions were found in the eight *ama-1* sequences (including the Eagle Pass sequence), with similarities ranging therefore from 99.76% to 100%, and no resulting amino acid differences. We also identified silent nucleotide substitutions in *clamp* (2) and *tufA* (1), but none in *hsp70* or *cox1*.

Within lineage A, and for all five studied genes, the sequences separated into two distinct groups, which we named A1 and A2 (Fig. 3) and that will be further detailed in the following Section 3.3.

Phylogenetic analysis, performed on the concatenated sequences of all genes (*tufA, cox1, hsp70, ama-1*, and *clamp*) (Fig. 4) as well as on each separate gene clearly showed that lineages A, C, and E are well supported (Figs. 5 and 6 and supplementary Figs. 2, 3 and 4). The concatenated tree separates into two major branches: one containing only lineage A, and a sister branch containing lineages C and E, which then further segregate.

This topology was consistent across all analyses, with the exception of the tree generated from *clamp* sequences (see supplementary Figs. 2, 3 and 4). In that tree, the branching separates lineages A and C from lineage E, which is a different arrangement. This is notable despite the fact that lineages C and E have a higher sequence similarity (85%) when compared to the similarities between A and C (80%) and A and E (77%).

3.3. Lineage A is a polyphyletic group

3.3.1. Lineage A separates into two sub-lineages A1 and A2 according to multilocus sequence analysis

Based on the multi-locus sequence analysis, we previously noticed that the lineage A sequences from our study separated into two distinct sub-lineages, A1 and A2. The A1 group consisted of six infected samples (four from mainland France and two from Guadeloupe), along with the two genome reference sequences (NVSL354 and WA). The A2 group was detected in two samples from mainland France (CVT21–14 and CVL21–18). Within each of these groups, sequences were highly similar, and most were identical. The sequence differences between these two sub-lineages varied depending on the gene analyzed. The similarities were around 95% for the nuclear genes (*hsp70, clamp*, and *ama-1*), but dropped to 85% for the apicoplast marker (*tufA*) and 81% for the mitochondrial marker (*cox1*; Fig. 3). In comparison, similarities between

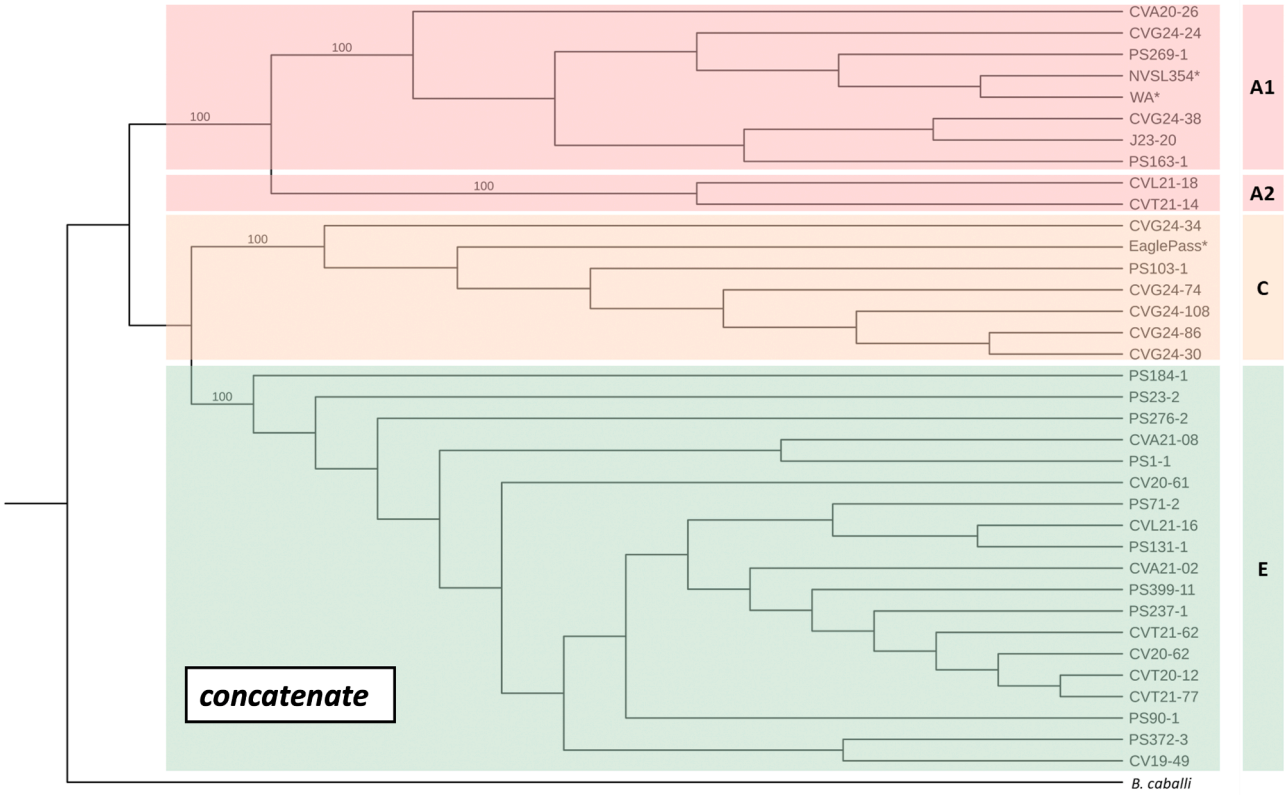

Fig. 4. A maximum likelihood phylogenetic tree of *Theileria equi* based on five partial gene sequences (*cox-1, tufA, hsp70, ama-1, clamp*). The tree includes 37 sequences with 3489 analyzed positions. As there are no GenBank accession numbers for the concatenated sequences, the blood samples' names were used (see Table 1). The GTR+G model has been used with 1000 repetitions. Significant bootstrap values (>70) are indicated on the branches. Sequences from reference genomes for the lineages A and C are indicated with *. Previously described lineages (A-E) as described in Tirosh-Levy et al. (2020) are indicated with different colored backgrounds. The subdivision of the lineage A in sub-lineages A1 and A2 is indicated with strong support values.

lineages A (A1/A2), C, and E were much lower—below or equal to 80% for all genes except *hsp70* (86.5–89%) and for *clamp* between the lineages C and E (85%).

The separation of these two sub-lineages was strongly supported by all individual phylogenetic trees and by the tree constructed from the concatenated sequences (Figs. 4–6 and supplementary Figs. 2, 3 and 4).

3.3.2. *Lineage A separates into three sub-lineages according to hsp70 and cox1 analysis*

A recent study from Brazil provided a set of *T. equi* sensu lato *hsp70* sequences that we were able to add to our analysis due to sequence overlap (Paulino et al., 2024). Most of these sequences (36/40) were identical to our A1 sub-lineage sequences. A second set of sequences (4/40) formed a third group, which we named A3, that also includes a sequence from Mongolia (BequUSDA AB248743; Yamasaki et al., 2007) and is distinct from A2. This new A3 group is closer to A1 (97% sequence similarity) than it is to A2 (95%). The three sub-lineages, A1, A2, and A3, are all well supported in the *hsp70* phylogenetic tree, with high bootstrap values of 95% and 100% (Fig. 5).

We retrieved and analyzed complementary *cox1* sequences from GenBank. Two sequences from France, named *Theileria* sp. "Europa" (MF510494–95), belonged to lineage E, showing a high similarity of 98–99% to our samples. In the same study, two other sequences from France/Corsica and Senegal, named *Theileria* sp. "Africa" (MF510492 and MF510493), belonged to the A1 sub-lineage, with a similarity exceeding 99.5%. Another set of sequences, including the BequUSDA *cox1* sequence (AB499091) (Xuan et al., 2001), a sequence from France/Corsica (MF510491) (Dahmana et al., 2019), and the sequence from the initial *B. equi* infected sample used to study the development of the parasite in ticks (OL672235) (Schnittger et al., 2022; Zapf and Schein, 1994), was also analyzed. These three sequences were identical to each other but distinct from both the A1 and A2 sub-lineages. They were closer to A1 (83.4% similarity) than to A2 (80.5%). The inclusion of the BequUSDA sequence (AB499091) in the *cox1* phylogenetic analysis separates the sub-lineage A3 from A1 and A2 with strong bootstrap values (Fig. 6). The other two sequences were not included in the phylogenetic analysis because of their lengths and/or partial overlapping with our sequences, but we can confidently group these three sequences within the A3 sub-lineage as they are identical over the comparable length.

3.3.3. *Congruent ema-1 and other genes phylogenetic analysis: lineage A separates into at least four sub-lineages*

We aimed to get a better and deeper picture of lineage A by characterizing *ema-1* sequences and classifying them into the clades A, B and C previously described for this gene (reviewed in Tirosh-Levy et al., 2020). Despite using several sets of primers (supplementary Fig. 5), including commonly used ones (Baptista et al., 2013; Coultous et al., 2022), we were never able to amplify *ema-1* from any of the 7 lineage C or 21 lineage E infected samples. Using the newly designed nPCR, we successfully amplified a 485 bp *ema-1* fragment from all 11 samples of lineage A (A1 or A2).

The *ema-1* sequences from lineage A1 (9 sequences) were either identical or nearly identical (99.8%) (Fig. 7A). They were also identical to many sequences available in GenBank from a wide range of countries and clustered within a commonly named clade A (*ema-1* clade A; Tirosh-Levy et al., 2020). We named this clade A/A1 to link the *ema-1* A clade and the A1 sub-lineage from our multi-locus analysis. In contrast, the two identical *ema-1* sequences from the A2 sub-lineage differed significantly from A/A1 (88% nucleotide similarity) and formed a new

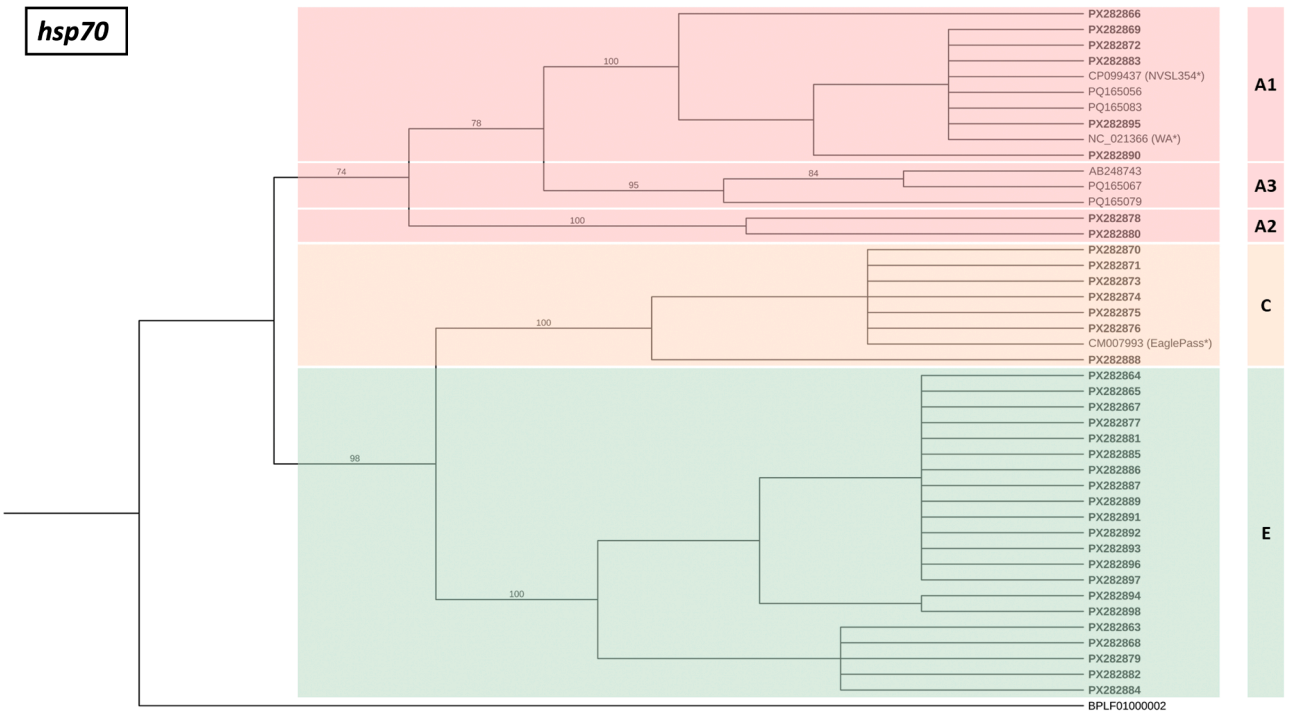

Fig. 5. A maximum likelihood phylogenetic tree of *Theileria equi* based on partial *hsp70* gene sequences. The tree includes 45 sequences with 1135 analyzed positions. The GTR+G model has been used with 1000 repetitions. Significant bootstrap values (> 70) are indicated on the branches. Representative sequences obtained in the present study are indicated in bold. Sequences from reference genomes for the lineages A and C are indicated with *. Previously described lineages (A-E) as described in Tirosh-Levy et al. (2020) are indicated with different colored backgrounds. The subdivision of the lineage A in sub-lineages A1, A2, and A3 is indicated with strong support values.

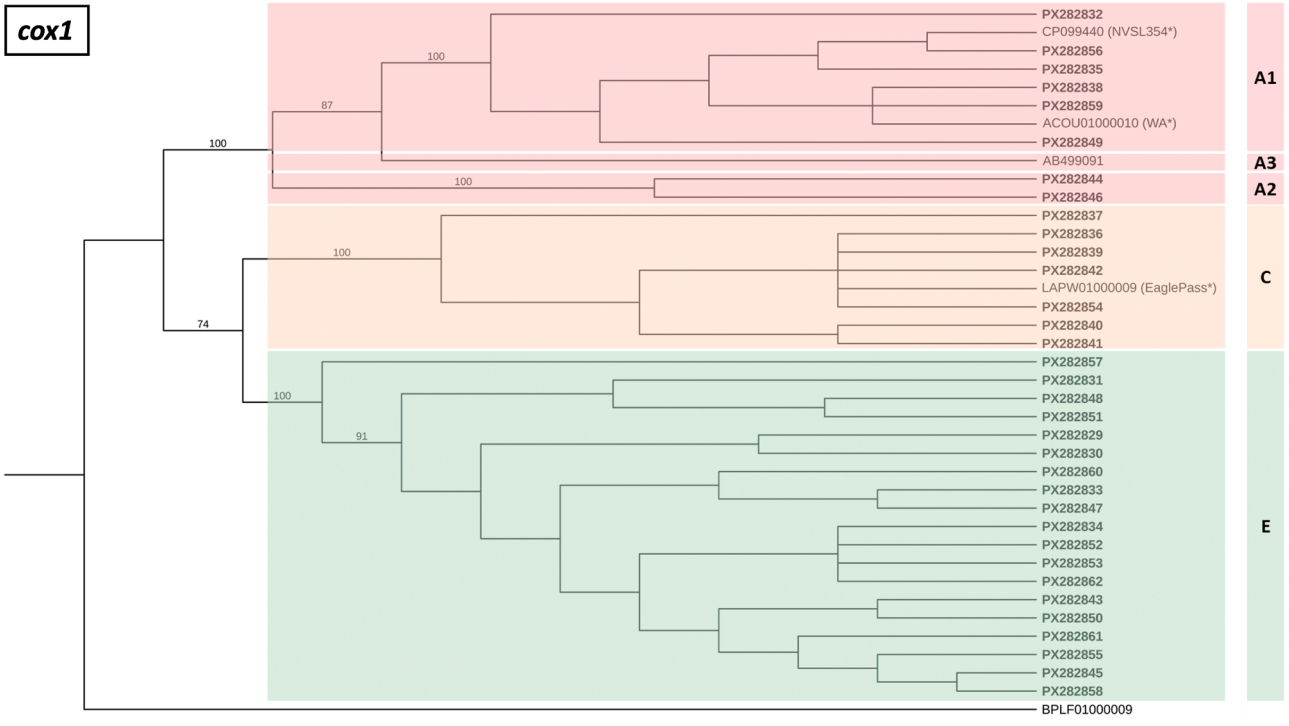

Fig. 6. A maximum likelihood phylogenetic tree of *Theileria equi* based on partial *cox1* gene sequences. The tree includes 39 sequences with 553 analyzed positions. The GTR+G model has been used with 1000 repetitions. Significant bootstrap values (> 70) are indicated on the branches. Sequences from reference genomes for the lineages A and C are indicated with *. Previously described lineages (A-E) as described in Tirosh-Levy et al. (2020) are indicated with different colored backgrounds. The subdivision of the lineage A in sub-lineages A1, A2, and A3 is indicated.

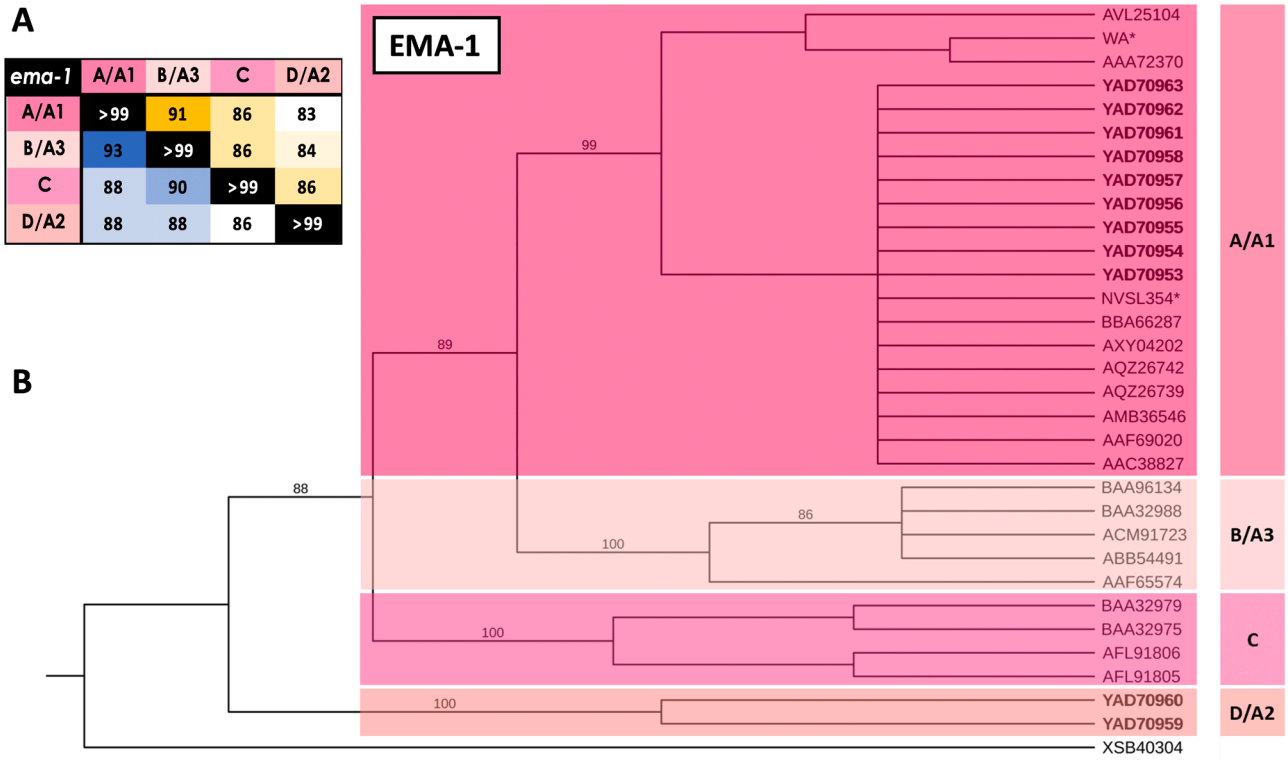

Fig. 7. A. Matrix of nucleotide (blue background) and protein (orange background) similarity percentages for partial *ema-1* sequences (485 bp). The color intensity varies according to the similarity percentage. The data includes our dataset and the available sequences from public database. The similarity values are averaged as multiple and variable sequences are represented in each sub-lineage. **B.** A maximum likelihood phylogenetic tree of *Theileria equi* based on partial EMA-1 protein sequences. The tree includes 32 sequences with 161 analyzed amino acids. The four EMA-1 clades (A to D) are indicated as well as their corresponding sub-lineages (A1 to A3). The LG+G model has been used with 1000 repetitions. Significant bootstrap values (>70) are indicated on the branches. Sequences from reference genomes for the lineage A are indicated with *.

clade, which we named D/A2. This new clade is distinct from the three previously described clades (A, B, and C; Tirosh-Levy et al., 2020). A similarity table is available (Fig. 7A). The available *ema-1* sequence of the BequUSDA strain (AB043618), which we classified as A3 based on *hsp70* and *cox1* similarities, clustered within *ema-1* clade B. This clade is more closely related to A/A1 (93% nucleotide similarity) than to the other two clades (90% and 88% with D/A2 and clade C, respectively), a proximity that was already noticed for *cox-1* and *hsp70*. We named this clade B/A3 to link the *ema-1* B clade and the A3 sub-lineage. A phylogenetic tree based on the EMA-1 amino acid sequences confirmed the presence of these four sub-lineages with strong bootstrap support (Fig. 7B and supplementary Table sheet 3).

To include more *ema-1* sequences from diverse geographical and study origins and to get a better overview of the relationships between the sequences of the four sub-lineages, we constructed a haplotype network including available sequences from GenBank. Sequences were selected, balancing length, coverage, and representativity, which resulted in a final dataset of 232 sequences, 438 bp long, from 17 different countries. A detailed description of this sequence set is available in the supplementary Table sheet 1. This analysis clearly separated the four sub-lineages, previously defined by phylogenetic analysis, into four distinct and distantly linked clusters (Fig. 8). The A/A1 group gathered the majority of sequences (166 sequences), featuring an apparent ancestral core of 118 identical sequences and a high number of satellite haplotypes in a star-like shape. These sequences originated from 16 different countries, reflecting a broad geographical distribution. A similar star-shaped pattern was found for the second most abundant sub-lineage, B/A3 (42 sequences), with a central core of 24 identical sequences. In contrast, there was no clear structure within the *ema-1* clade C cluster (13 sequences), which primarily consists of sequences from South Africa (10/13), but also includes sequences from Japan and Brazil. The D/A2 sub-lineage, with only two sequences from our study, appears to have stemmed from the major A/A1 sub-lineage, from which it is quite distant (51 mutations). The closer proximity of B/A3 with A/A1 was also highlighted in the network (22 mutational events compared to 48 and 51 with C/A4 and D/A2 respectively).

3.3.4. *Are there more potential sub-lineages within the lineage A?*

We compared a supplementary dataset of 108 shorter *ema-1* sequences (167–430 bp) not included in the previous haplotype network to our four delineated sub-lineage sequences to look for other clades and new members of the A2 sub-lineage (Suppl. Table sheet 4). Within a sub-lineage, similarity was over 99%, but between sub-lineages, it was below 94%. As was already shown in the haplotype network, most of these sequences belonged to the A/A1 sub-lineage (65/108). We also found sequences belonging to the B/A3 clade (7/108) and one for clade C.

We did find 15 sequences corresponding to the D/A2 sub-lineage, primarily in Mongolia (11 sequences), in China (one sequence representing 20 infected samples) but also in Croatia (3 sequences), where they were previously designated as clade D (Manna et al., 2018), group 4 (Asia) and group E (Croatia), respectively (Coultous et al., 2022; Munkhjargal et al., 2013; Wu et al., 2023). Additionally, at least one supplementary clade, which we named E, was found in Croatia, with 20 identical sequences that Coultous et al. (2022) had named D. The SNPs that differentiate these five clades are shown on the alignment in Fig. 9.

3.4. *Is ema-11 gene a target to evaluate genetic diversity within the lineage C/T. haneyi?*

We designed new internal primers to develop an *ema-11* nPCR based

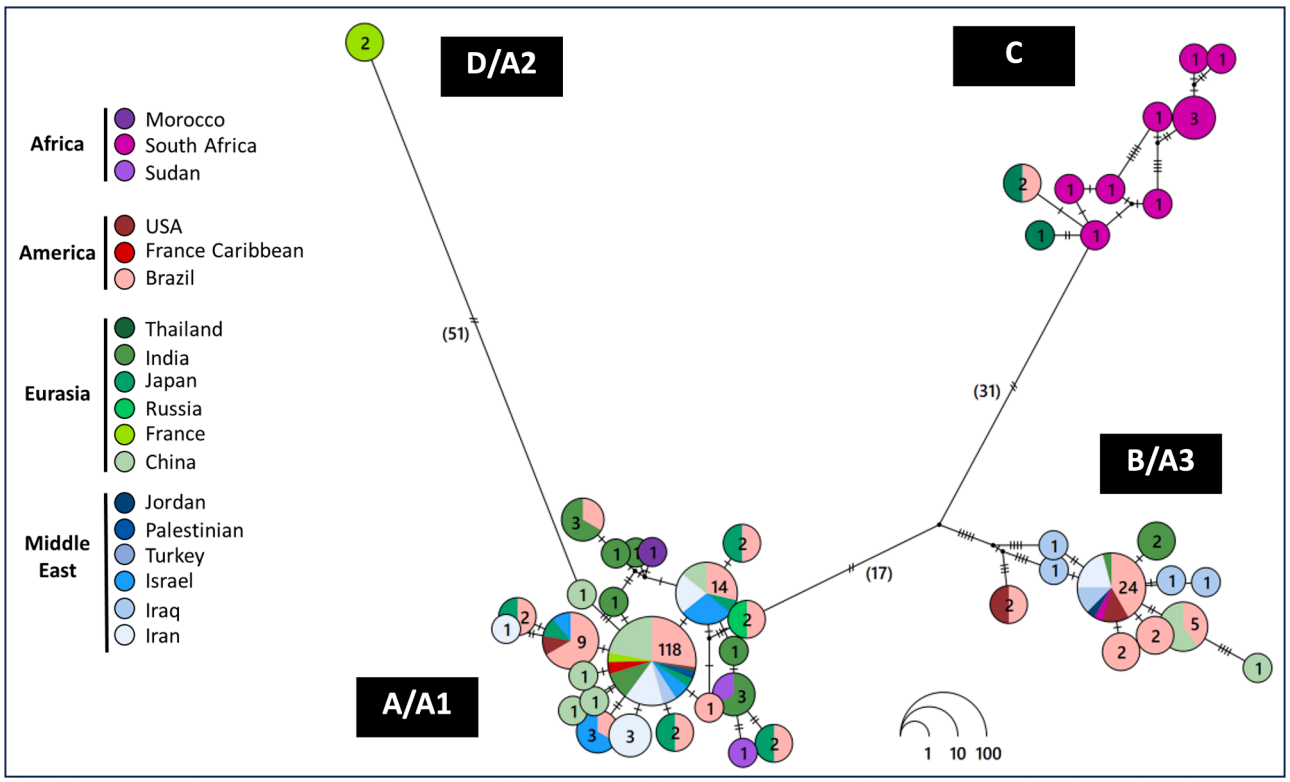

Fig. 8. Haplotype TCS network of 232 *Theileria equi ema-1* sequences (438 bp) characterized in equids (Genbank and this study). The size of each circle is proportional to the number of individual samples for each haplotype. A log scale is used to allow representation of high number of sequences. Nucleotide variations are denoted by the hatch marks across the lines connecting the haplotypes with each bar representing a single nucleotide variation. When this number exceeds 5, the number of SNPs is indicated close to the branch for clarity. The lengths of the branches aren't strictly proportional to the number of mutational events, and were adjusted manually to improve figure readability. Colors correspond to the geographic distribution of the different haplotypes and clades/sub-lineages' names are displayed in bold within black squares.

Clade/	10	60	217//	25	85	167
Sub-lineage	51	102	171	231	291	353
A/A1	GCGGATTTGCATAGTCACACCTACAAAGTCCCACCCCGCTTAGGTTGCCGTCGCGAGTGGGTCTCAGCACATCAC					
B/A3	ACGGGTCCGCTCCGCCACACCTACTAAGCCCCACCCCGTCCCTGCTGCCCTCCCGAGCGTACCCAAACACATCAT					
C[a]	ACAAACCGAATCCACCTACACCACAACGCTCCATCCCGTCCAGGCCGTAGCCGGTCCCCTACTAAAGCGCGACAC					
C	ACAAACAGAATCCACCTACACCACAACACCCCACCCCGTCCAGGCCGTGGCCGGACCTGTACTCAAGCGCGACAC					
D/A2[b]	ATGGATCGGCTTCACTCCAATTTTCGCGTTTTGCAATACTTAGGTTCCCGTTGCGACTGTACTCAGACGCATAAC					
Group 4	GCGGATCGGCTTCACTCCAATTTTCGCGTTTTGCAATACTTA----------					
Group E	----------GTTCCCGTTGCGACTGTACTCAGACGCATAAC					
E/Group D[c]	----------ACCGCCGTTGCGATTGTACTTAGATGTATCCC					

Fig. 9. Alignment of single nucleotide polymorphisms (75) detected over the comparable *ema-1* sequence lengths (353 bp) that discriminate between the 5 *ema-1* clades A to E of *T. equi* lineage A, noted in red bold characters.

Variable positions only were kept from the alignments of the consensus sequences obtained for each lineage or sub-lineage. The nomenclature used is: A-C as referred to in Tirosh-Levy et al., 2020, D as referred to in the present study and in Manna et al., 2018, and E as referred to in the present study. Short sequences from Coultous et al., 2022 (167 bp) and from other studies of interest on parasites infecting Mongolian and Chinese equids (Munkhjargal et al., 2013; Wu et al., 2023) overlapp only on 4 conserved positions. SNPs highlighted in yellow are specific to one clade.

[a] Clade C sequences further separate at least into two groups with differences highlighted in blue: from South Africa with reference sequences JQ782605.1, JQ782606.1 and from Japan with reference sequences AB015208.1, AB015212.1.

[b] Alignment with the longer sequence obtained in this study allowed assignation of group 4 according to Munkhjargal et al., 2013 and Wu et al., 2023 (reference sequences AB731192.3, MK568953.1) and group E according to Coultous et al., 2022 (reference sequences ON502854.1, ON502856.1) to the same sub-lineage A2b indicated in green bold (or *ema-1* clade D).

[c] clade E corresponds to group D according to Coultous et al., 2022 (reference sequences ON502855.1, ON502857.1.)

The positions indicated just above the alignment correspond to the sequences of the present study and above to the positions in the short sequences (219 bp[b] and 167 bp[b,c]).

on previously described *ema-11* sequence from *T. haneyi* genome and primers (Bhoora et al., 2025; Knowles et al., 2018). Amplification was successful for all lineage C infected samples (7/7) with the expected 579 bp partial fragments, but not from any of the eight lineage A or 21 lineage E infected samples.

The seven *ema-11* sequences showed high similarity (99.3% to

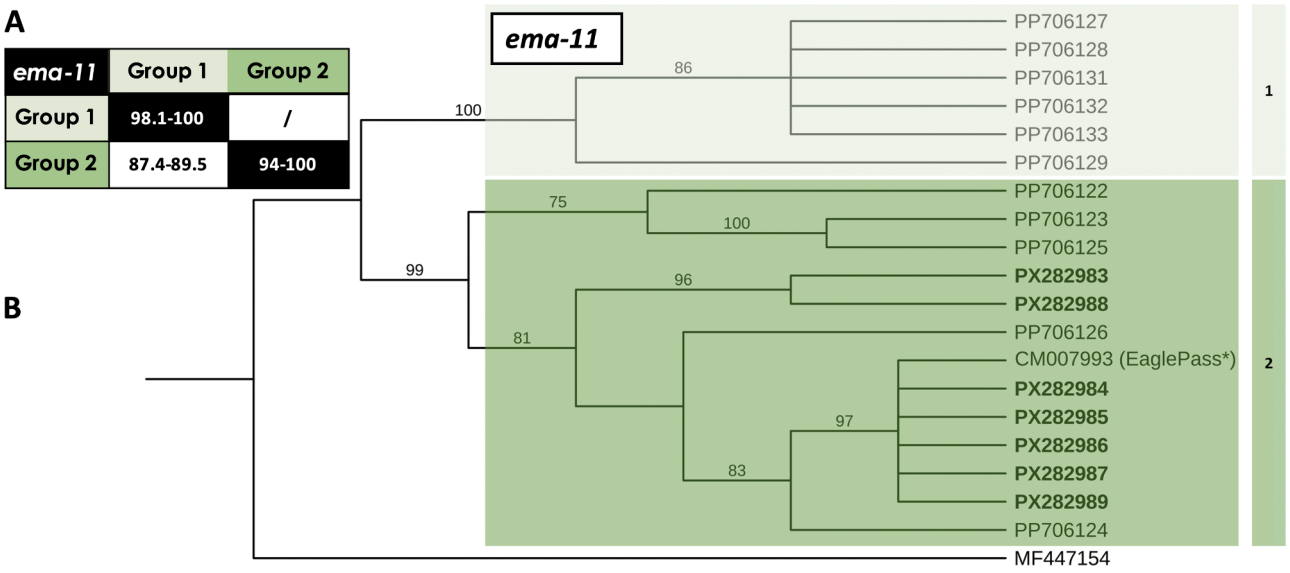

Fig. 10. A. Similarity matrices of nucleotide homology percentages for partial *ema-11* gene (515 bp). The data includes our dataset and the available sequences from public database deposed by Bhoora et al., 2025. A range of similarity values are indicated as multiple and variable sequences are represented in each group. **B.** A maximum likelihood tree of *Theileria equi* based on *ema-11* gene partial sequences. The tree includes 20 sequences with 515 analyzed positions. The GTR+G model has been used with 1000 repetitions. Significant bootstrap values (>70) are indicated on the branches. Sequences obtained in the present study are indicated in bold. Sequence from reference genome for the lineage C are indicated with *. Previously described lineages (Group 1 and 2) as described in Bhoora et al. in 2025 are indicated with different colored backgrounds.

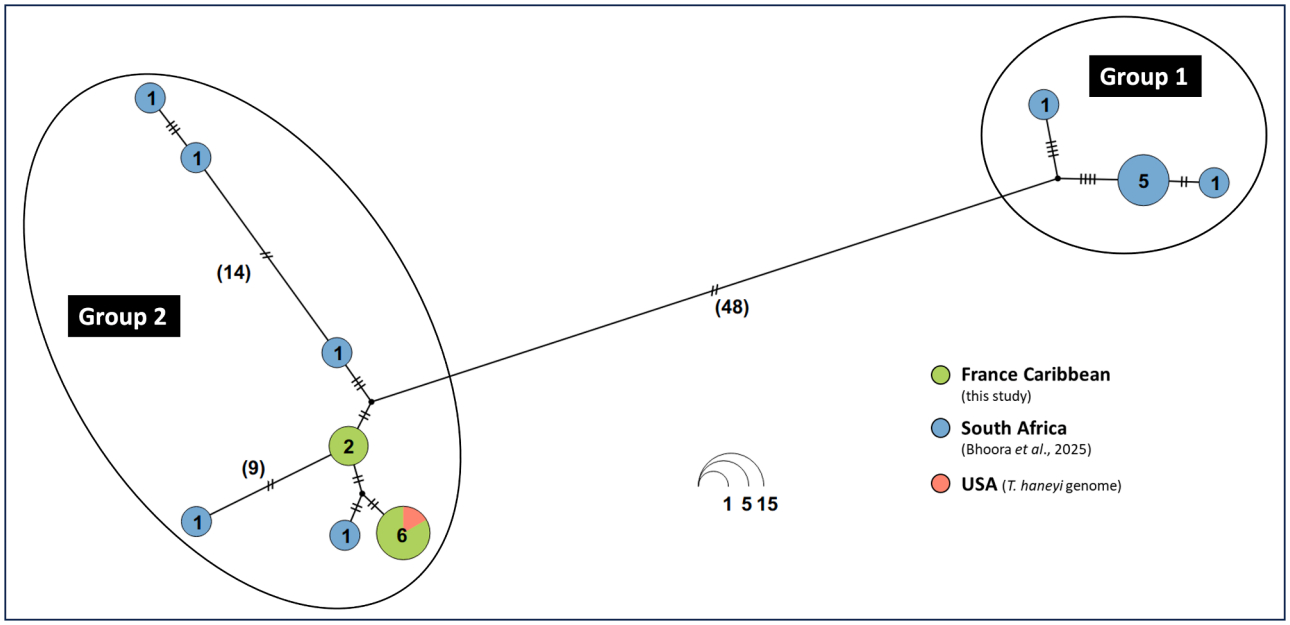

Fig. 11. Haplotype TCS network of 20 *Theileria equi ema-11* sequences (515 bp) characterized in equids infected by genotype C/*T. haneyi* (Genbank and this study). The size of each circle is proportional to the number of individual samples for each haplotype. A log scale is used to allow representation of high number of sequences. Nucleotide variations are denoted by the hatch marks across the lines connecting the haplotypes with each bar representing a single nucleotide variation. The lengths of the branches are proportional to the number of mutational events. When this number exceeds 5, the number of SNPs is indicated close to the branch for clarity. Colors correspond to the geographic distribution of the different haplotypes. Clades' names are displayed in bold within black squares.

100%), with four single nucleotide polymorphisms (SNPs) differentiating two haplotypes. The seven Caribbean sequences had strong homologies (94–100%) with South African *ema-11* sequences described as group 2 by Bhoora et al., 2025 (Fig. 10A). Similarities with the group 1 ranged from 87.4% to 89.5%. In both the phylogenetic analysis and the haplotype network, the Caribbean *ema-11* sequences clustered within the group 1, alongside five South African *ema-11* sequences (Figs. 10B and 11; Bhoora et al., 2025). The haplotype network, supported by high bootstrap values in the phylogenetic analysis, more clearly evidenced that the previously defined group 2 appears to split into two separate clades (Figs. 10B and 11).

4. Discussion

This study aimed to investigate the genetic diversity within *Theileria equi* sensu lato by targeting genes from the parasite's three distinct

genomes. We designed primer sets for the partial amplification of five genes with varying evolutionary rates: *cox1* (mitochondrial), *tufA* (apicoplast), and *hsp70, ama-1*, and *clamp* (nuclear). The amplification was successfully performed on a panel of 36 infected blood samples, representing three of the five described *T. equi* lineages (A, C/*T. haneyi*, and E). The resulting sequences were concatenated to form a robust dataset of approximately 3.5 kb for subsequent phylogenetic analyses. Prior to this, parasite from each infected blood sample had been assigned to its specific lineage based on partial 18S rRNA sequences, which were meticulously checked to confirm the absence of co-infections (Mège et al., 2025).

A horse infected with the lineage C (*T. haneyi*) was discovered for the first time in France. However, this horse lived in Guadeloupe and was brough back to metropolitan France 12 years before the sampling performed for this study. This horse was most probably infected there and remained carrier of this lineage from the Caribbeans where it is abundant (Mège et al., 2025).

This study reports the first partial sequences of five novel genes derived from an extensive panel of lineage E infected blood samples, enabling a direct and comprehensive comparison with lineage A and *T. haneyi* (lineage C). Our findings, based on these five genetic markers, confirm that the *T. equi* parasites from mainland France and the Caribbean, previously assigned to lineages A and C based on 18S rRNA sequences, clearly group with their respective lineage genome references. The 21 lineage E infected samples showed high genetic similarity and also formed a well-defined lineage. We have now provided five additional reference genes for lineage E. These markers have demonstrated significant intra-lineage variability within lineage A, proving more suitable for future genetic diversity studies than the highly conserved 18S rRNA gene.

A key advantage of our primer design is its ability to amplify the five genes across genetically distant lineages (A, C, and E), suggesting their potential utility for analyzing the remaining lineages, B and D. A potential limitation, however, is that these primers may co-amplify all variants in mixed infections, leading to unusable sequences. Nevertheless, the sequence dataset we have generated can serve as a valuable resource to design lineage-specific primers.

In this study, our phylogenetic analyses were congruent across all genes—regardless of their origin (mitochondrial, apicoplast, or nuclear)—confirming the separation of *T. equi* into three distinct lineages with high bootstrap support. However, the tree topologies varied slightly depending on the individual gene analyzed. While two major lineages consistently diverged, their precise relationships differed from those described by Schnittger et al. (2022) based on the 18S rRNA gene. Our analysis showed that lineage A consistently diverged independently from lineages C and E for all genes except *clamp*, though with low bootstrap support for *tufA* and *ama-1*. In the concatenated sequence analysis, lineage A was more distantly related to lineages C/E, with high bootstrap values, supporting its distinct position. The low bootstrap support previously observed in 18S rRNA analyses (Schnittger et al., 2022) and the different topologies reported in other studies (Chatanga et al., 2025; Mège et al., 2025; Otgonsuren et al., 2024; Tirosh-Levy et al., 2020) highlight the ongoing uncertainty in *T. equi* sensu lato phylogeny. To definitively improve our understanding, complementary sequence data for these five genes from lineages B and D would be crucial.

Once we had confirmed that our 36 *T. equi* infected blood samples clustered into three distinct lineages with no evidence of co-infection, we proceeded to link the presence and sequences of *ema-1* and *ema-11* genes to their respective lineages, A and C. We designed new nested PCR primer sets and successfully amplified *ema-1* exclusively in the 11 lineage A infected samples and *ema-11* in the seven lineage C infected samples (*T. haneyi*). Neither of these two genes was amplified from the 21 lineage E infected samples, which corroborates our previous findings (Mège et al., 2025). This suggests that lineage E likely possesses its own set of *ema* genes, as already described for the lineages A and C through the genome sequencing (Grimsley et al., 2023; Kappmeyer et al., 2012; Knowles et al., 2018).

In other studies, *ema-1* sequences have been attributed to 18S rRNA lineage E infected samples (e.g., Coultous et al., 2022; Munkhjargal et al., 2013). However, these *ema-1* sequences, despite their short length, were highly similar to lineage A's *ema-1* (either A1 or A2) and contained SNPs characteristic of the lineage A as shown in our alignments. In such cases, *T. equi* from both lineages A and E co-occurred in the same geographical area, increasing the possibility of co-infection. It's plausible that the 18S rRNA PCR preferentially amplified the most abundant lineage (e.g., lineage E), while the *ema-1* nPCR detected a less abundant, co-infecting lineage A parasite which moreover, is the only one to possess this *ema-1* gene, as we demonstrated previously. This could lead to the misattribution of an *ema-1* sequence to a lineage E 18S rRNA sequence.

Similarly, 18S rRNA sequences identified as *T. haneyi* (lineage C) were positioned within lineage E (Yang et al., 2025). The authors' strategy assumed that an *ema-1*-negative/*chr1sco*-positive (the latter being specific of *T. haneyi*) sample was exclusively infected with *T. haneyi*. However, based on our current and previous results (Mège et al., 2025), a co-infected sample with *T. haneyi* and lineage E would yield the same result. The absence of true *T. haneyi* 18S rRNA sequences in the Yang and colleagues' study is notable. Co-infections with different *T. equi* isolates are extremely frequent and have been quantified using lineage-specific detection PCRs (Ahedor et al., 2023; Coultous et al., 2020; Mège et al., 2025). These co-infections contribute to blurring our understanding of *T. equi* sensu lato phylogeny.

In this study, all five genetic markers clearly highlighted the presence of two genetically distinct clades, named A1 and A2, within lineage A. This distinction had previously been suspected based on 18S rRNA sequences (Jouglin et al., 2025). The analysis of *hsp70* sequences from a Brazilian study (Paulino et al., 2024), along with available *cox-1* sequences (Schnittger et al., 2022; Xuan et al., 2001; Yamasaki et al., 2007), and our current dataset confirmed the presence of a third, phylogenetically well-supported sub-lineage, that we named A3.

For the first time, our analysis of *ema-1* sequences directly linked the previously described *ema-1* clades A, D, and B (Manna et al., 2018; Tirosh-Levy et al., 2020) to the multi-locus sub-lineages A1, A2, and A3, respectively. The entire dataset supports the conclusion that each of these sub-lineages within lineage A should be considered a distinct species. This conclusion is based on four key lines of evidence: [1] high bootstrap values in the phylogenetic analysis; [2] the significant genetic distance between the sub-lineages, with only 81% similarity for the mitochondrial *cox1* gene, which is a common species-barcoding marker (Antil et al., 2023; Herbert et al., 2003; Schreeg et al., 2016); [3] the consistent results from five other genes (*hsp70, ama-1, clamp, ema-1*, and *tufA*) with different evolutionary traits, confirming the same phylogenetic relationships; [4] the structure of the haplotype network, which shows well-separated, star-shaped *ema-1* clades.

Despite this strong genetic evidence, there is no clear geographical structuring among these candidate species, and they all infect the same hosts (equids). This suggests that these candidate species may have originated in different parts of the world and spread globally through the trading of domestic horses. Moreover, there might be more of these candidate species to be confirmed (corresponding to *ema-1* clades C and E) or discovered.

The fundamental question is at what level to delineate species within *T. equi* sensu lato. Should it be at the lineage level, as anticipated for *T. haneyi* (Knowles et al., 2018) and proposed by Schnittger et al. (2022) for other lineages, or at the sub-lineage level? The available, albeit limited, *ema-11* data (Bhoora et al., 2025; Knowles et al., 2018) already suggest that *T. haneyi* itself may comprise -multiple species. This is a notable divergence from the historical species delineation in *Babesia* and *Theileria* sensu stricto, which was often based on host specificity before molecular data drew their clustering (Schnittger et al., 2022). For *T. equi* sensu lato, the process is inverted: with no apparent host or geographical

structuring, species delineation is slowly emerging through the accumulation of molecular data. While 18S rRNA is a reliable phylogenetic marker, its conservation and limited resolution make it problematic for distinguishing sub-lineages, particularly in the presence of frequent co-infections. Therefore, we advise against rushing to assign formal species names, as such changes are adopted slowly and can cause confusion for horse owners and diagnostic companies, especially when related to export limitations and control measures.

The significant genetic diversity and lack of geographic structuring between and within *T. equi* lineages indicate that a single vaccine would likely fail to provide broad protection. Therefore, successful vaccine design must consider a multi-antigenic approach to account for the distinct variants of surface proteins found in each sub-lineage.

5. Conclusion

Through the generation of a comprehensive multi-gene dataset for three *T. equi* sensu lato lineages (A, C, and E), this work significantly refines the species' phylogeny and establishes a foundational framework for the development of high-resolution diagnostic tools. We not only confirmed the 18S rRNA lineage structure but also, for the first time, linked specific *ema-1* clades to the sub-lineages within lineage A, which we also distinguished through our multi-locus gene analysis. While these findings represent a significant step forward, additional data are crucial to gain a more complete picture. We need more sequences for lineages B and D, as well as more *ema-11* sequences within lineage C. Expanding the study of lineage E to include more geographically diverse samples beyond France would also be vital to enhance our understanding.

Funding

The study was funded by IFCE (Institut Français du Cheval et de l'Équitation [French Institute for Horses and Riding]), Fonds Eperon and Institut Carnot France Futur Elevage.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, LM used Gemini in order to improve the English language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethical statement

All procedures were approved by the Ethics Committee on Animal Experimentation from Antilles and Guyane (CEMEAAG), process number HC-69–2024–4, or by the Oniris Ethics Committee for Veterinary Clinical and Epidemiological Research for the metropolitan France sampling (CERVO-2019–1-V) or by the VetAgro Sup Research Ethics Committee (Approval number: 2211), depending of the sample origin.

Data availability

The datasets used and/or analyzed during the current study are available from L.M. as well as sample DNA reference, on reasonable request.

CRediT authorship contribution statement

Mickaël Mège: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Florian Berquier:** Writing – review & editing, Investigation, Formal analysis. **Claire Bonsergent:** Writing – review & editing, Validation, Resources. **Laurence Malandrin:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

We extend our sincere gratitude to all the participants in this study. This research would not have been possible without the invaluable collaboration of the numerous citizens, veterinarians, technicians and associations (RESPE: Réseau d'Epidémio-Surveillance en Pathologie Equine; Association pour le développement des Sciences Equines) who spent time and effort to this project. We particularly thank them for their commitment and enthusiasm, which were essential to its success.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ttbdis.2026.102656.

References

Ahedor, B., Otgonsuren, D., Zhyldyz, A., Guswanto, A., Ngigi, N.M.M., Valinotti, M.F.R., Kothalawala, H., Kalaichelvan, N., Silva, S.S.P., Kothalawala, H., Acosta, T.J., Sivakumar, T., Yokoyama, N., 2023. Development and evaluation of specific polymerase chain reaction assays for detecting *Theileria equi* genotypes. Parasit. Vect. 16, 435. https://doi.org/10.1186/s13071-023-06045-z.

Antil, S., Abraham, J.S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., Bhagat, P., Gupta, R., Sood, U., Lal, R., Toteja, R., 2023. DNA barcoding, an effective tool for species identification: a review. Mol. Biol. Rep. 50, 761–775. https://doi.org/10.1007/s11033-022-08015-7.

Arisue, N., Hashimoto, T., 2015. Phylogeny and evolution of apicoplasts and apicomplexan parasites. Parasitol. Int. 64, 254–259. https://doi.org/10.1016/j.parint.2014.10.005.

Baptista, C., Lopes, M.S., Tavares, A.C., Rojer, H., Kappmeyer, L., Mendonça, D., da Câmara Machado, A., 2013. Diagnosis of *Theileria equi* infections in horses in the Azores using cELISA and nested PCR. Ticks Tick. Borne Dis. 4, 242–245. https://doi.org/10.1016/j.ttbdis.2012.11.008.

Berná, L., Rego, N., Francia, M.E., 2021. The elusive mitochondrial genomes of Apicomplexa: where are we now? Front. Microbiol. 12, 751775. https://doi.org/10.3389/fmicb.2021.751775.

Bhoora, R., Quan, M., Matjila, P.T., Zweygarth, E., Guthrie, A.J., Collins, N.E., 2010. Sequence heterogeneity in the equi merozoite antigen gene (*ema-1*) of *Theileria equi* and development of an *ema-1*-specific TaqMan MGB assay for the detection of *T. equi*. Vet. Parasitol. 172, 33–45. https://doi.org/10.1016/j.vetpar.2010.04.025.

Bhoora, R.V., Collins, N.E., Schnittger, L., Troskie, C., Marumo, R., Labuschagne, K., Smith, R.M., Dalton, D.L., Mbizeni, S., 2020. Molecular genotyping and epidemiology of equine piroplasmids in South Africa. Ticks Tick. Borne Dis. 11, 101358. https://doi.org/10.1016/j.ttbdis.2019.101358.

Bhoora, R.V., Mbaba, T.V., Troskie, M., Ackermann, R.E., Collins, N.E., 2025. Quantitative detection of *Theileria haneyi* in South African horses. Ticks Tick. Borne Dis. 16, 102487. https://doi.org/10.1016/j.ttbdis.2025.102487.

Bianchini, G., Sánchez-Baracaldo, P., 2024. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. Ecol. Evol. 14, e10873. https://doi.org/10.1002/ece3.10873.

Bonsergent, C., de Carné, M.C., de la Cotte, N., Moussel, F., Perronne, V., Malandrin, L., 2021. The new human *Babesia* sp. FR1 is a european member of the *Babesia* sp. MO1 clade. Pathogens 10, 1433. https://doi.org/10.3390/pathogens10111433.

Chatanga, E., Ahedor, B., Atabek, B., Kainga, H., Kapalamula, T., Razemba, T., Nakao, R., Nonaka, N., Sivakumar, T., Yokoyama, N., 2025. An epidemiological survey of equine piroplasmosis in donkeys and horses in Malawi. Vet. Parasitol. Reg. Stud. Rep. 63, 101315. https://doi.org/10.1016/j.vprsr.2025.101315.

Coultous, R.M., McDonald, M., Raftery, A.G., Shiels, B.R., Sutton, D.G.M., Weir, W., 2020. Analysis of *Theileria equi* diversity in The Gambia using a novel genotyping method. Transbound. Emerg. Dis. 67, 1213–1221. https://doi.org/10.1111/tbed.13454.

Coultous, R., Gotić, J., McCann, M., Sutton, D., Beck, R., Shiels, B., 2022. Novel equi merozoite antigen (*ema-1*) gene heterogeneity in a geographically isolated *Theileria equi* population in Croatia. Parasit. Vect. 15, 401. https://doi.org/10.1186/s13071-022-05484-4.

Dahmana, H., Amanzougaghene, N., Davoust, B., Normand, T., Carette, O., Demoncheaux, J.P., Mulot, B., Fabrizy, B., Scandola, P., Chik, M., Fenollar, F., Mediannikov, O., 2019. Great diversity of piroplasmida in Equidae in Africa and Europe, including potential new species. Vet. Parasitol. Reg. Stud. Rep. 18, 100332. https://doi.org/10.1016/j.vprsr.2019.100332.

De Waal, D.T., 1992. Equine piroplasmosis: a review. Br. Vet. J. 148, 6–14. https://doi.org/10.1016/0007-1935(92)90061-5.

Grimsley, M., Hicks, J., Zeineldin, M., Murphy, G., Sigafoose, T., 2023. Complete genome sequence of *Theileria equi* NVSL354. Microbiol. Resour. Announc. 12, e00809–22. https://doi.org/10.1128/mra.00809-22.

Hebert, P.D., Cywinska, A., Ball, S., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270, 313–321. https://doi.org/10.1098/rspb.2002.2218.

Hermans, L.M., Bonsergent, C., Josson, A., Rocafort-Ferrer, G., Le Guyader, M., Pessey, S. A., Leblond, A., Malandrin, L., 2025. Evaluation of *Theileria equi* vertical transmission rate and routes in a cohort of asymptomatic mares and their foals. Ticks Tick. Borne Dis. 16, 102432. https://doi.org/10.1016/j.ttbdis.2024.102432.

Jouglin, M., Bonsergent, C., de la Cotte, N., Mège, M., Bizon, C., Bohec, A., Couroucé, A., Guibert, O., Lallemand, E., Lemonnier, L., Leblond, A., Leroux, A., Marano, I., Muzard, A., Prévôt, A., Quéré, E., Thirouin, M., Toussaint, M., Agoulon, A., Malandrin, L., 2025. Equine piroplasmosis in different geographical areas in France: prevalence heterogeneity of asymptomatic carriers and low genetic diversity of *Theileria equi* and *Babesia caballi*. Ticks Tick. Borne Dis. 16, 102434. https://doi.org/10.1016/j.ttbdis.2024.102434.

Kappmeyer, L.S., Thiagarajan, M., Herndon, D.R., Ramsay, J.D., Caler, E., Djikeng, A., Gillespie, J.J., Lau, A.O., Roalson, E.H., Silva, J.C., Silva, M.G., 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 Genom. 13, 603. https://doi.org/10.1186/1471-2164-13-603.

Kibbe, W.A., 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic. Acids. Res. 35, W43–W46. https://doi.org/10.1093/nar/gkm234.

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, 679–690. https://doi.org/10.1016/j.ijpara.2018.03.010.

Lamb, I.M., Okoye, I.C., Mather, M.W., Vaidya, A.B., 2023. Unique properties of apicomplexan mitochondria. Annu. Rev. Microbiol. 77, 541–560. https://doi.org/10.1146/annurev-micro-032421-120540.

Lemoine, F., Domelevo Entfellner, J.B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., Gascuel, O., 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556, 452–456. https://doi.org/10.1038/s41586-018-0043-0.

Loubens, M., Marinach, C., Paquereau, C.E., Hamada, S., Hoareau-Coudert, B., Akbar, D., Franetich, J.F., Silvie, O., 2023. The claudin-like apicomplexan microneme protein is required for gliding motility and infectivity of *Plasmodium* sporozoites. PLoS Pathog. 19, e1011261. https://doi.org/10.1371/journal.ppat.1011261.

Malandrin, L., Jouglin, M., Sun, Y., Brisseau, N., Chauvin, A., 2010. Redescription of *Babesia capreoli* (Enigk and Friedhoff, 1962) from roe deer (*Capreolus capreolus*): isolation, cultivation, host specificity, molecular characterisation and differentiation from *Babesia divergens*. Int. J. Parasitol. 40, 277–284. https://doi.org/10.1016/j.ijpara.2009.08.008.

Manna, G., Cersini, A., Nardini, R., Bartolomé Del Pino, L.E., Antognetti, V., Zini, M., Conti, R., Lorenzetti, R., Veneziano, V., Autorino, G.L., Scicluna, M.T., 2018. Genetic diversity of *Theileria equi* and *Babesia caballi* infecting horses of Central-Southern Italy and preliminary results of its correlation with clinical and serological status. Ticks Tick. Borne Dis. 9, 1212–1220. https://doi.org/10.1016/j.ttbdis.2018.05.005.

Mège, M., Bonsergent, C., Viry, L., Dhune, M., Lecollinet, S., Malandrin, L., 2025. Genetic diversity of equine piroplasmosis agents in Guadeloupe (Caribbeans): first report of *Theileria haneyi*, evaluation of diagnostic tools and impact of horse movement. Ticks Tick. Borne Dis. 16, 102547. https://doi.org/10.1016/j.ttbdis.2025.102547.

Mendoza, F.J., Pérez-Écija, A., Kappmeyer, L.S., Suarez, C.E., Bastos, R.G., 2024. New insights in the diagnosis and treatment of equine piroplasmosis: pitfalls, idiosyncrasies, and myths. Front. Vet. Sci. 11, 1459989. https://doi.org/10.3389/fvets.2024.1459989.

Munkhjargal, T., Sivakumar, T., Battsetseg, B., Nyamjargal, T., Aboulaila, M., Purevtseren, B., Bayarsaikhan, D., Byambaa, B., Terkawi, M.A., Yokoyama, N., Igarashi, I., 2013. Prevalence and genetic diversity of equine piroplasms in Tov province, Mongolia. Infect. Genet. Evol. 16, 178–185. https://doi.org/10.1016/j.meegid.2013.02.005.

OIE, 2024. Infection with *Theileria equi* and *Babesia caballi* (equine piroplasmosis). Terrestrial Animal Health Code. https://www.woah.org/en/what-we-do/standards/codes-and-manuals/terrestrial-code-online-access/?id=169&L=0&htmfile=chapitre_equine_piroplasmosis.htm.

Onyiche, T.E., Suganuma, K., Igarashi, I., Yokoyama, N., Xuan, X., Thekisoe, O., 2019. A review on equine piroplasmosis: epidemiology, vector ecology, risk factors, host immunity, diagnosis and control. Int. J. Env. Res. Publ. Health 16, 1736. https://doi.org/10.3390/ijerph16101736.

Onzere, C.K., Fry, L.M., Bishop, R.P., Silva, M.G., Bastos, R.G., Knowles, D.P., Suarez, C. E., 2021. *Theileria equi* claudin like apicomplexan microneme protein contains neutralization-sensitive epitopes and interacts with components of the equine erythrocyte membrane skeleton. Sci. Rep. 11, 9301. https://doi.org/10.1038/s41598-021-88902-4.

Otgonsuren, D., Amgalanbaatar, T., Narantsatsral, S., Enkhtaivan, B., Munkhgerel, D., Zoljargal, M., Davkharbayar, B., Myagmarsuren, P., Battur, B., Battsetseg, B., Sivakumar, T., Yokoyama, N., 2024. Epidemiology and genetic diversity of *Theileria equi* and *Babesia caballi* in Mongolian horses. Infect. Genet. Evol. 119, 105571. https://doi.org/10.1016/j.meegid.2024.105571.

Paulino, P.G., Amaral, F.B., de Oliveira, R.T., de Andrade, S.G., Rabello, C.A., Meirelles, N., de Souza Santana, M., Galdino, K.C.P., Jacob, J.C.F., Peckle, M., Massard, C.L., Santos, H.A., 2024. Genetic diversity and phylogenetic relationships of *Theileria equi* inferred from heat shock protein 70 kDa gene sequences in Brazilian horses. Vet. Res. Commun. 49, 16. https://doi.org/10.1007/s11259-024-10599-3.

Remarque, E.J., Faber, B.W., Kocken, C.H., Thomas, A.W, 2008. Apical membrane antigen 1: a malaria vaccine candidate in review. Trends. Parasitol. 24, 74–84. https://doi.org/10.1016/j.pt.2007.12.002.

Rothschild, C.M., 2013. Equine piroplasmosis. J. Equine Vet. Sc. 33, 497–508. https://doi.org/10.1016/j.jevs.2013.03.189.

Saxena, V., Garg, S., Tripathi, J., Sharma, S., Pakalapati, D., Subudhi, A.K., Boopathi, P. A., Saggu, G.S., Kochar, S.K., Kochar, D.K., Das, A., 2012. *Plasmodium vivax* apicoplast genome: a comparative analysis of major genes from Indian field isolates. Acta Trop. 122, 138–149. https://doi.org/10.1016/j.actatropica.2012.01.007.

Schnittger, L., Rodriguez, A.E., Florin-Christensen, M., Morrison, D.A., 2012. *Babesia*: a world emerging. Infect. Genet. Evol. 12, 1788–1809. https://doi.org/10.1016/j.meegid.2012.07.004.

Schnittger, L., Ganzinelli, S., Bhoora, R., Omondi, D., Nijhof, A.M., Florin-Christensen, M., 2022. The Piroplasmida *Babesia, Cytauxzoon*, and *Theileria* in farm and companion animals: species compilation, molecular phylogeny, and evolutionary insights. Parasitol. Res. 121, 1207–1245. https://doi.org/10.1007/s00436-022-07424-8.

Schreeg, M.E., Marr, H.S., Tarigo, J.L., Cohn, L.A., Bird, D.M., Scholl, E.H., Levy, M.G., Wiegmann, B.M., Birkenheuer, A.J., 2016. Mitochondrial genome sequences and structures aid in the resolution of Piroplasmida phylogeny. PLoS. One 11, e0165702. https://doi.org/10.1371/journal.pone.0165702.

Sivakumar, T., Tuvshintulga, B., Otgonsuren, D., Batmagnai, E., Ahedor, B., Kothalawala, H., Vimalakumar, S.C., Silva, S.S.P., Yamagishi, J., Yokoyama, N., 2022. Phylogenetic analyses of the mitochondrial, plastid, and nuclear genes of *Babesia* sp. Mymensingh and its naming as *Babesia naoakii* n. sp. Parasit. Vect. 15, 299. https://doi.org/10.1186/s13071-022-05374-9.

Scoles, G.A., Ueti, M.W., 2015. Vector ecology of equine piroplasmosis. Ann. Rev. Entomol. 60, 561–580. https://doi.org/10.1146/annurev-ento-010814-021110.

Tamzali, Y., 2013. Equine piroplasmosis: an updated review. Eq. Vet. Edu. 25, 590–598. https://doi.org/10.1111/eve.12070.

Tirosh-Levy, S., Gottlieb, Y., Fry, L.M., Knowles, D.P., Steinman, A., 2020. Twenty years of equine piroplasmosis research: global distribution, molecular diagnosis, and phylogeny. Pathogens 9, 926. https://doi.org/10.3390/pathogens9110926.

Triglia, T., Healer, J., Caruana, S.R., Hodder, A.N., Anders, R.F., Crabb, B.S., Cowman, A. F., 2000. Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. Mol. Microbiol. 38, 706–718. https://doi.org/10.1046/j.1365-2958.2000.02175.x.

Tyler, J.S., Treeck, M., Boothroyd, J.C., 2011. Focus on the ringleader: the role of AMA1 in apicomplexan invasion and replication. Trends. Parasitol. 27, 410–420. https://doi.org/10.1016/j.pt.2011.04.002.

Vences, M., Patmanidis, S., Schmidt, J.C., Matschiner, M., Miralles, A., Renner, S.S., 2024. Hapsolutely: a user-friendly tool integrating haplotype phasing, network construction, and haploweb calculation. Bioinform. Adv. 4, vbae083. https://doi.org/10.1093/bioadv/vbae083.

Wise, L.N., Kappmeyer, L.S., Mealey, R.H., Knowles, D.P., 2013. Review of equine piroplasmosis. J. Vet. Intern. Med. 27, 1334–1346. https://doi.org/10.1111/jvim.1216.

Wise, L.N., Kappmeyer, L.S., Knowles, D.P., White, S.N., 2019. Evolution and diversity of the EMA families of the divergent equid parasites, *Theileria equi* and *T. haneyi*. Infect. Genet. Evol. 68, 153–160. https://doi.org/10.1016/j.meegid.2018.12.020.

Wu, J., Cui, Y., Yu, F., Muhatai, G., Tao, D., Zhao, A., Ning, C., Qi, M., 2023. Prevalence and genetic characterization of *Theileria equi* and *Babesia caballi* in grazing horses in Xinjiang, northwestern China. Parasitol. Res. 122, 387–394. https://doi.org/10.1007/s00436-022-07749-4.

Xuan, X., Larsen, A., Ikadai, H., Tanaka, T., Igarashi, I., Nagasawa, H., Fujisaki, K., Toyoda, Y., Suzuki, N., Mikami, T.J., 2001. Expression of *Babesia equi* merozoite antigen 1 in insect cells by recombinant baculovirus and evaluation of its diagnostic potential in an enzyme-linked immunosorbent assay. Clin. Microbiol. 39, 705–709. https://doi.org/10.1128/JCM.39.2.705-709.2001.

Yamasaki, M., Inokuma, H., Sugimoto, C., Shaw, S.E., Aktas, M., Yabsley, M.J., Yamato, O., Maede, Y., 2007. Comparison and phylogenetic analysis of the heat shock protein 70 gene of *Babesia* parasites from dogs. Vet. Parasitol. 145, 217–227. https://doi.org/10.1016/j.vetpar.2007.01.003.

Yang, G., Chen, Y., Chen, K., Hu, Z., Li, J., Wang, J., Guo, W., Wang, X., Du, C., 2026. Prevalence and molecular epidemiology of the novel equine parasite *Theileria haneyi* in China. Equine Vet. J. 58, 638–649. https://doi.org/10.1111/evj.14505.

Zapf, F., Schein, E., 1994. The development of *Babesia* (*Theileria*) *equi* (Laveran, 1901) in the gut and the haemolymph of the vector ticks, *Hyalomma* species. Parasitol. Res. 80, 297–302. https://doi.org/10.1007/BF02351869.