



Contents lists available at SciVerse ScienceDirect

## International Journal for Parasitology

journal homepage: [www.elsevier.com/locate/ijpara](http://www.elsevier.com/locate/ijpara)

# Evolution of the tissue factor pathway inhibitor-like Kunitz domain-containing protein family in *Rhipicephalus microplus*

Elizabeth Louw<sup>a</sup>, Nicolaas A. van der Merwe<sup>b</sup>, Albert W.H. Neitz<sup>a</sup>, Christine Maritz-Olivier<sup>b,\*</sup>

<sup>a</sup> Department of Biochemistry, Faculty of Natural and Agricultural Sciences, University of Pretoria, Lynnwood Road, Pretoria 0002, South Africa

<sup>b</sup> Department of Genetics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0002, South Africa

## ARTICLE INFO

## Article history:

Received 6 September 2012

Received in revised form 14 November 2012

Accepted 15 November 2012

Available online 7 December 2012

## Keywords:

*Rhipicephalus microplus*

Kunitz domain

Tissue factor pathway inhibitor (TFPI)

Phylogenetic

Bayesian Markov Chain Monte Carlo

Duplication

## ABSTRACT

One of the principle mechanisms utilised by ticks to obtain a blood meal is the subversion of the host's haemostatic response. This is achieved through the secretion of saliva containing anti-haemostatic proteins into the feeding lesion. Lineage-specific expansion of predicted secretory protein families have been observed in all previously studied ticks and occurred in response to adaptation to a blood-feeding environment. Of these, the predominant families are common between both hard and soft ticks. One of these families, namely the Kunitz domain-containing protein family, includes proven tissue factor pathway inhibitor-like (TFPI-like) anti-haemostatics such as ixolaris and penthalaris that play a crucial role during tick feeding. Although Kunitz-type proteins have been found in *Rhipicephalus microplus*, the TFPI-like Kunitz protein family has not yet been studied. We report a comprehensive search for TFPI-like Kunitz domain-containing proteins in *R. microplus* expressed sequence tag libraries, resulting in the identification of 42 homologues. The homologues were bioinformatically and phylogenetically studied, including the application of an intensive Bayesian Markov Chain Monte Carlo (MCMC) analysis of the individual Kunitz domain nucleotide sequences. We show that the *R. microplus* TFPI-like Kunitz protein family groups into two main clades that presumably underwent ancient duplication, which indicates that a whole genome duplication event occurred at least 150 million years ago. Evidence for recent and ancient gene and domain duplication events was also found. Furthermore, the divergence times of the various tick lineages estimated in this paper correspond with those presented in previous studies. The elucidation of this large protein family's evolution within *R. microplus* adds to current knowledge of this economically important tick.

© 2012 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

*Rhipicephalus microplus* is considered to be the most important tick parasite of cattle worldwide, resulting in profit losses and financial expenditures due to both direct and indirect damage to the animals. Infestation by these ticks causes reductions in weight gain and milk production. This tick species is also responsible for indirect damage to cattle due to its vectoring capability for anaplasmosis and babesiosis pathogens. Losses due to *R. microplus* infestations have been estimated at US \$7 billion annually (Peconick et al., 2008) and these losses are expected to rise due to expansion of the tick's habitable range into non-endemic areas (White et al., 2003; Estrada-Peña et al., 2005; Lynen et al., 2008). Currently, only one tick vaccine has been developed and is marketed as Gavac in Central and South America, and as TickGARD

in Australia (de la Fuente et al., 2007a). It is a recombinant vaccine based on a *R. microplus* gut antigen (Bm86) and while it has been successfully used in conjunction with other acaricides it exhibits limited and varying efficacy against different tick species and strains (reviewed by Willadsen, 2004). Therefore, the identification of novel tick vaccine candidates is of great importance to the global cattle industry.

Recently, tick research has benefitted greatly from genetic technologies such as expressed sequence tag (EST) studies and whole-genome sequencing. Before the advent of such technologies, gene discovery had been achieved through time-consuming and laborious experiments, often resulting in the identification of only a single gene. The recent release of first-draft genome sequencing data for *R. microplus* (Bellgard et al., 2012), together with genome data already available for *Ixodes scapularis* (Hill and Wikel, 2005; van Zee et al., 2007), has stimulated gene discovery in these and other ticks. However, these sequencing data sets need to undergo additional analyses before useful information may be extracted. Additionally, bioinformatic analyses and results should be confirmed by biological evidence such as protein characterisation and functional studies.

\* Corresponding author. Address: Room 7-30, Department of Genetics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Lynnwood Road, Pretoria 0002, South Africa. Tel.: +27 12 420 3945/3258; fax: +27 12 362 5327/5302.

E-mail address: [christine.maritz@up.ac.za](mailto:christine.maritz@up.ac.za) (C. Maritz-Olivier).

Previous EST studies showed lineage-specific expansion of several predicted secretory protein families for both hard and soft ticks (Mans et al., 2002b, 2008a; Mans and Neitz, 2004). These studies pointed to a series of gene duplication events after the divergence of the Ixodidae and Argasidae families. Such duplications provide the genetic variance that enables phenotypic hyper-variability and may result in many paralogues with slightly altered specificities that can target a range of host and native proteins (Ohta, 1994, 2000; Christeller, 2005). The predominant predicted secretory protein families, i.e. lipocalins, basic tail secretory protein, Kunitz domain-containing peptides and metalloproteases are common between hard and soft ticks. Of these, the Kunitz domain-containing protein family alone accounts for between 5% and 20% of all predicted secretory proteins (Ribeiro et al., 2006; Chmelař et al., 2008; Francischetti et al., 2008a,b; Mans et al., 2008a). In *I. scapularis*, arguably the most intensively studied tick in North America due to its vector capacity for Lyme disease, the Kunitz domain-containing protein family comprises 10% of all predicted secretory proteins (Ribeiro et al., 2006). Even though several Kunitz proteins have been identified and/or characterised in *R. microplus* (Tanaka et al., 1999; de Miranda Santos et al., 2004; Sasaki et al., 2004; Rachinsky et al., 2007; Macedo-Ribeiro et al., 2008), the larger Kunitz domain-containing protein family has not been specifically studied within this tick species.

Kunitz-type proteins also have significant roles within other parasites. In the black fly, *Simulium vittatum*, they have been shown to inhibit coagulation and have a possible role in modulating inflammatory responses (Tsujimoto et al., 2012). In the flatworm, *Echinococcus granulosus*, a family of Kunitz domain-containing proteins have been implicated in the initial stages of infection (González et al., 2009), while many other parasites have recruited the Kunitz-scaffold into their venom proteins (reviewed by Fry et al. (2009)). In hookworms such as *Ancylostoma ceylanicum* it has been shown that Kunitz-type inhibitors protect the parasite against digestive serine proteases present in the host's small intestine (Costa et al., 2009; Tsujimoto et al., 2012). Immunisation against the *A. ceylanicum* Kunitz-type inhibitors confers partial protection against hookworm-associated growth delay, lending support for their possible use as parasite vaccine candidates (Chu et al., 2004).

The Kunitz protein family members, of which bovine pancreatic trypsin inhibitor (BPTI) is the type member (Burck et al., 1967; Cerwinsky et al., 1967), are typically serine protease inhibitors. One of their crucial functions in vertebrates is the control of haemostasis. Human tissue factor pathway inhibitor (TFPI) is a potent anticoagulant that achieves inhibition of essential serine proteases of the blood coagulation cascade, such as factor Xa (fXa) and factor VIIa (fVIIa), through its three Kunitz domains (Lwaleed and Bass, 2006). As blood coagulation is one of the primary and most robust defences of vertebrates against blood-feeding parasites, it is unsurprising that all ticks secrete an assortment of anti-haemostatic proteins (reviewed by Maritz-Olivier et al. (2007) and Corral-Rodríguez et al. (2009)). These anticoagulants serve to keep blood in its fluid form in both the feeding lesion and the tick's midgut, enabling the parasite to both obtain and digest its blood meal.

Various tick-derived Kunitz proteins that inhibit the serine proteases of the blood coagulation cascade, as well as soft tick-derived platelet aggregation inhibitors, have been identified. Ornithodorin and savignin are both bi-Kunitz thrombin inhibitors isolated from the soft ticks, *Ornithodoros moubata* (van de Locht et al., 1996) and *Ornithodoros savignyi* (Mans et al., 2002a). Despite having similarity to BPTI, they do not bind to thrombin via the Kunitz substrate-binding loop, but instead insert their N- and C-terminal residues into thrombin's active site cleft and basic fibrinogen recognition exosite, respectively. Similar binding modes have been inferred for the hard tick-derived thrombin inhibitors, boophilin from *R. microplus* (Macedo-Ribeiro et al., 2008) and hemalin from

*Haemaphysalis longicornis* (Liao et al., 2009), as well as for the soft tick-derived fX(a) inhibitors, tick anticoagulant protein from *O. moubata* (Wei et al., 1998) and fXa inhibitor (fXaI) from *O. savignyi* (Joubert et al., 1998). Multi-Kunitz inhibitors of fX(a) have also been isolated from the hard tick *I. scapularis* (Francischetti et al., 2002, 2004). Platelet aggregation inhibitors from soft ticks that use the Kunitz domain substrate-binding loop to present an RGD integrin recognition motif have also been characterised. Savignygrin isolated from *O. savignyi* is the most well-known of these (Mans et al., 2002c), but several savignygrin homologues have been identified in other soft ticks, including monogrin in *Argas monolakensis* (Mans et al., 2008b), disagregin in *O. moubata* (even though it presents RED instead of RGD (Karczewski et al., 1994; Karczewski and Connolly, 1997)) and RGD-presenting Kunitz proteins have also been found in *Ornithodoros parkeri* (Francischetti et al., 2008a).

Two TFPI homologues, ixolaris and penthalaris, have been identified and characterised in *I. scapularis* and have been shown to be potent anticoagulants (Francischetti et al., 2002, 2004; Monteiro et al., 2005, 2008; Nazareth et al., 2006). They share sequence homology with TFPI and have anticoagulant capabilities within the nanomolar range. In the case of ixolaris, this is achieved through inhibition of the tissue factor/fVIIa complex by using fX(a) as a scaffold (Francischetti et al., 2002). Penthalaris and ixolaris also form part of the larger Kunitz domain-containing protein family in *I. scapularis* (Ribeiro et al., 2006) which has developed through lineage-specific expansion.

Lineage-specific expansion of the Kunitz protein family has been previously observed in all studied ticks (Valenzuela et al., 2002; Francischetti et al., 2005, 2008a,b; Ribeiro et al., 2006; Alarcon-Chaidez et al., 2007; Batista et al., 2008; Chmelař et al., 2008; Mans et al., 2008a; Dai et al., 2012). Previous studies have shown that this protein family has evolved through gene duplication within the main tick families but after the divergence of hard and soft ticks (Mans et al., 2002b, 2008a,b; Mans and Neitz, 2004). The Kunitz protein family has therefore expanded in response to the blood-feeding environment that the Ixodidae and Argasidae encountered individually. Recently, Dai et al. (2012) showed that a greater expansion of the Kunitz-type proteins within hard ticks, specifically *I. scapularis*, likely facilitates the long-term blood-feeding behaviour of these ticks. It is likely that a similar, expanded Kunitz domain-containing protein family is present in *R. microplus*, with the members showing homology to the TFPI-like proteins, ixolaris and penthalaris.

Due to a lack of knowledge regarding a TFPI-like Kunitz protein family in the economically important *R. microplus*, we undertook to identify TFPI-like Kunitz domain-containing proteins in the cattle tick. Our preliminary identification of this protein family, using available *R. microplus* sequence databases, revealed a family of 42 homologues. Phylogenetic analysis uncovered both ancient and recent gene- and domain duplications, supporting previous observations that lineage-specific expansions of the Kunitz protein family occurred in all studied ticks. More exhaustive studies on this family of proteins will be of great significance once the complete genome of the cattle tick becomes available. This will allow intensive cross-genera analyses, facilitating a deeper insight into the evolution of ticks and their independent adaptation to a blood-feeding lifestyle.

## 2. Materials and methods

### 2.1. Identification of TFPI-like Kunitz domain-containing proteins in *R. microplus*

In order to assemble a data set of TFPI-like Kunitz domain-containing homologues, the Basic Local Alignment Search Tool

(BLAST) (Altschul et al., 1997) was used to identify sequences with-in available *R. microplus* sequence databases that had similarity with the amino acid sequences of BPTI, TFPI, ixolaris and penthalaris (Table 1). BLAST searches were performed against a contig database (BmiContig compiled by C. Maritz-Olivier) assembled from the *R. microplus* gene index (BmiGi) (available at the Dana-Farber Cancer Gene Index Project, DFCI, at <http://compbio.dfci.harvard.edu/tgi/tgipage.html>), as well as against *R. microplus* sequences available from the National Centre for Biotechnology Information (NCBI, available at <http://www.ncbi.nlm.nih.gov>). Sequences that showed significant similarities ( $E\text{-value} \leq 5 \times 10^{-2}$ ) to at least two of the four query sequences were included in further analyses. The correct reading frame for nucleotide sequences retrieved during BLAST searches was determined by aligning the various translated open reading frames with the amino acid sequence of the TFPI protein.

## 2.2. Bioinformatic analysis of homologues

The assembled sequences were submitted for domain-, signal peptide- and transmembrane region-prediction to online bioinformatic tools available through the ExPASy molecular biology server (available online at <http://www.expasy.ch>). Domain prediction was performed using both SMART version 6 (Simple Modular Architecture Research Tool, (Schultz et al., 1998; Letunic et al., 2008)) and Pfam version 24.0 (Finn et al., 2010) software. Signal peptide and transmembrane regions were predicted by SignalP version 3.0 (Nielsen et al., 1997; Bendtsen et al., 2004) and TMHMM version 2.0 (Kahsay et al., 2005) software, respectively. Sequences containing premature stop codons were discarded, except in the case of one (TC1750pt) where it was apparent that such a premature stop codon was the result of a nonsense mutation within the Kunitz domain. Remaining sequences were curated by truncation to the starting methionine, if one was apparent, and by discarding parts of the sequence that were separated from predicted Kunitz domains with stop codons. Forty-two TFPI-like Kunitz domain-containing homologues were identified in this manner.

## 2.3. Phylogenetic analysis of complete homologue sequences

The full amino acid sequences of all the homologues were aligned to the four initial query sequences using the ClustalX program, version 2.0.7 (Larkin et al., 2007) and phylogenetic trees were inferred from the alignment with the use of the MEGA4 program, build# 4028 (Tamura et al., 2007). Visualisation of the underlying relationships between the various sequences was achieved by employing the maximum parsimony (MP), minimum evolution (ME), neighbour-joining (NJ) and unweighted pair group mean average (UPGMA) methods. The Poisson correction model for amino acid substitution was employed and, due to varying sequence

lengths of homologues, all positions that contained missing data or alignment gaps were eliminated only in pairwise sequence comparisons. The reliability of the generated phylogenetic trees was evaluated by means of 10,000 bootstrap replicates. For all other analysis options the program default settings were used. The correct topology and analysis method was accepted as the method that yielded a tree that had the greatest number of nodes with fair support (present in at least 50% of bootstrap replicates).

## 2.4. Phylogenetic analysis of individual Kunitz domains from homologues

The 42 homologues were divided into their individual core Kunitz domains by truncating the domains at positions 4 and 56 (BPTI numbering according to Schechter and Berger (1967)). Partial Kunitz domains were not included. To be able to infer divergence times of the *R. microplus* Kunitz protein family and to show its lineage-specific expansion, an outgroup of *Argasidae* single Kunitz domain-containing sequences was selected and processed in the same manner and included in the analysis. These were savignygrin from *O. savignyi*, monogrin 1A and 1B from *A. monolakensis* and two savignygrin-like transcripts from *O. parkeri* herein referred to as ORNIPARK9 and ORNIPARK10 (Table 1). The final data set contained 98 sequences of which 93 were from homologues retrieved from online *R. microplus* databases.

Alignment of codons within the various Kunitz domains was manually refined by using the amino acid sequence alignments as guides. Phylogenetic analyses were performed using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) version 1.5.3 (Drummond and Rambaut, 2007). Three taxon sets with normal prior distributions were calibrated with divergence times from Jeyaprakash and Hoy (2009). These were Ixodida, which included all sequences ( $300 \pm 27$  million years ago (MYA)), Argasidae containing the five soft tick sequences ( $214 \pm 28$  MYA) and 93 *R. microplus* sequences representing the Metastriate taxon set ( $134 \pm 22$  MYA). The General Time Reversible (GTR) substitution model was employed, base frequencies were estimated and the site heterogeneity model "Gamma + Invariant Sites" with four gamma categories was selected. Two other models were also evaluated, one using the SRD06 model and another partitioning the codons separately, but the best fitting model was selected as one that correctly resolved the outgroup. A relaxed uncorrelated lognormal clock was used with the Yule tree prior for speciation and a starting tree was generated with the UPGMA method. Ten simulation runs, each with a Markov Chain Monte Carlo (MCMC) chain length of 10,000,000, were independently performed with parameter logging at every 1,000th generation. Result files from independent analyses were analysed using Tracer (version 1.5.0) and software tools included in the BEAST package. The resultant tree was visualised with FigTree, version 1.3.1.

In order to obtain bootstrap support for the Bayesian tree, phylogenies were also generated using MEGA4. All methods (MP, ME, NJ and UPGMA) were employed with complete deletion of missing data or alignment gaps and inclusion of all codon positions (1st + 2nd + 3rd + non-coding). The nucleotide substitution model was the Maximum Composite Likelihood (MCL) and 1,000 bootstrap replicates were performed to infer confidence. The validity of the trees were evaluated against the Bayesian tree, and the tree with the highest number of nodes supported by either the posterior probability from BEAST and/or the bootstrap support from MEGA4 was selected as the most appropriate. This was the tree recovered by the ME method.

## 2.5. Analysis of family expansion as a result of domain duplication

To ascertain whether duplication events occurred that allowed expansion of the TFPI-like Kunitz domain-containing protein

**Table 1**

Accession numbers of transcripts used in this study. The GenBank accession numbers of the amino acid sequences of transcripts used as queries (BPTI, TFPI, ixolaris and penthalaris) or as outgroups for phylogenetic studies (savignygrin, monogrin 1A and 1B, ORNIPARK9 and ORNIPARK 10) are shown.

| Transcript name                            | GenBank accession number |
|--|--------------------------|
| Bovine pancreatic trypsin inhibitor (BPTI) | 115114                   |
| Ixolaris                                   | 15077002                 |
| Monogrin 1A                                | 114152966                |
| Monogrin 1B                                | 114152968                |
| ORNIPARK9                                  | 149287172                |
| ORNIPARK10                                 | 149286898                |
| Penthalaris                                | 22164266                 |
| Savignygrin                                | 21435981                 |
| Tissue factor pathway inhibitor (TFPI)     | 125932                   |



family in *R. microplus*, we analysed the level of similarity and identity within several clades identified during phylogenetic analysis of the individual Kunitz domains. We defined ancient duplications as sequences that have an amino acid identity of 40% or higher and a nucleotide identity of at least 50%. Recent duplications were defined as those with an amino acid identity of no less than 85% and a minimum nucleotide identity of 90%. The data set comprised of sub-clades containing at least three sequences, with a maximum divergence time of 55 MYA between the sequences.

### 3. Results

#### 3.1. Identification of TFPI-like Kunitz proteins in *R. microplus* and bioinformatic analysis

Forty-two sequences with significant similarity ( $E\text{-value} \leq 5 \times 10^{-2}$ ) to at least two query sequences (BPTI, TFPI, ixolaris and penthalaris) were identified during BLAST searches of *R. microplus* sequences in the BmiContig and NCBI databases. Analyses of these sequences revealed the presence of at least one Kunitz domain per sequence, up to a maximum of nine Kunitz domains (as for TC4684), with an average of 2.3 Kunitz domains per sequence (Fig. 1). Only three of the retrieved homologues were annotated in NCBI, while remaining sequences were of unknown affinity. EU636772 was described as a putative chymotrypsin inhibitor precursor (GenBank accession number EU636772, (Lima et al., 2010)), TC4205 corresponded to the two boophilin isoforms (G2 and H2, GenBank accession numbers 17529564 and 17529566, respectively) and TC6750 was described as a putative secreted protein known as penthaplus (GenBank accession number ABA55034.1).

The Bilaris family was the predominant group, constituting almost half of all Kunitz domain-containing proteins (17 sequences, 44% of the total), followed by the Monolaris and Trilaris families consisting of 14 sequences (33%) and five sequences (14%), respectively (Figs. 2–4). However, five of the 42 sequences contained partial Kunitz domains predicted at the N-terminal (CK192299, CV441191, TC4684, TC6491, TC13756) and others did not present with an initial methionine.

Approximately 40% of the full Kunitz domains presented a lysine or arginine in the P1 position, similar to BPTI and TFPI. The alignment of the individual Kunitz domain amino acid sequences

is shown in Supplementary Fig. S1. Signal peptides were predicted for 17 sequences (41%) with 24 remaining non-secretory (60%). Only a single sequence had a predicted signal anchor (TC334). Twelve of the 17 secretory sequences (71%) were contained within the Bilaris group. Furthermore, transmembrane regions were predicted for four of the 42 sequences (TC334, TC2049, TC6502 and TC8213), amounting to 10% of the total. The signal peptide predicted for TC8213 spanned most of the transmembrane region predicted for this transcript. In the case of TC334 the predicted signal anchor also spanned part of the transmembrane region.

Upon considering the alignment of the TFPI-like Kunitz proteins (Figs. 2–4), the disulphide-bonded cysteine residues of the Kunitz domain remained the most conserved residues, with partial conservation only for the second and fourth cysteines (Figs. 3 and 4). This loss of conservation of C<sub>2</sub> and C<sub>4</sub> was also present in two of the initial query sequences, ixolaris and penthalaris, which both originate from the tick, *I. scapularis*.

#### 3.2. Phylogenetic analyses of complete amino acid sequences

Phylogenetic analyses of full-length amino acid sequences using the MEGA4 program revealed that the NJ method yielded a consensus tree (Fig. 5) with the greatest degree of support. Of the 43 nodes, 28 displayed fair support (present in at least 50% of bootstrap replicates) and 16 of these had good support (present in at least 75% of bootstrap replicates).

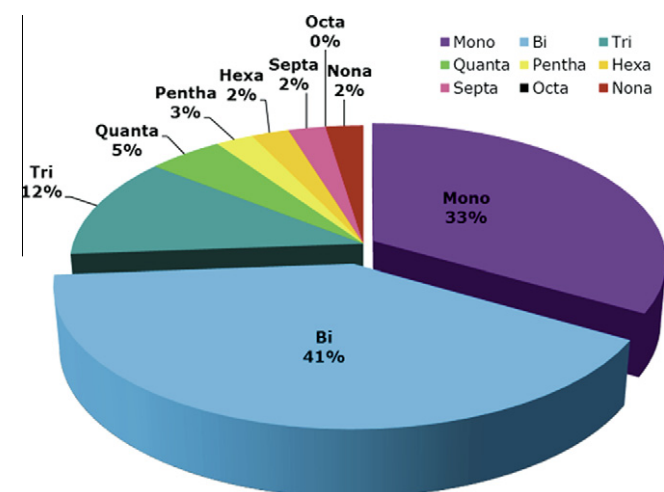
The homologues were grouped into four major clades (Clades I–IV) with 16, nine, four and 13 homologue sequences, respectively (Fig. 5). Clade I was the largest and most diverse clade, containing sequences from every multi-Kunitz family as well as two of the initial query sequences, namely TFPI and penthalaris. The sub-clade, Clade Ia, had good support for all nodes. Both ixolaris and BPTI grouped into Clade II, which was composed almost exclusively of Monolaris sequences, with the exception of three Bilaris sequences (ixolaris, TC3418, TC5548). The smallest of the major clades was Clade III, which contained three Bilaris and one Monolaris sequence, and which was basal to both Clades I and II. Clade IV consisted mainly of Bilaris sequences but also contained a Monolaris sub-assembly (Clade IVa), as well as three Trilaris and one penthalaris sequence scattered within Clade IVc. The Monolaris sub-clade (Clade IVa) was located basal to the rest of the sub-clades in Clade IV.

#### 3.3. Phylogenetic analysis of individual Kunitz domains

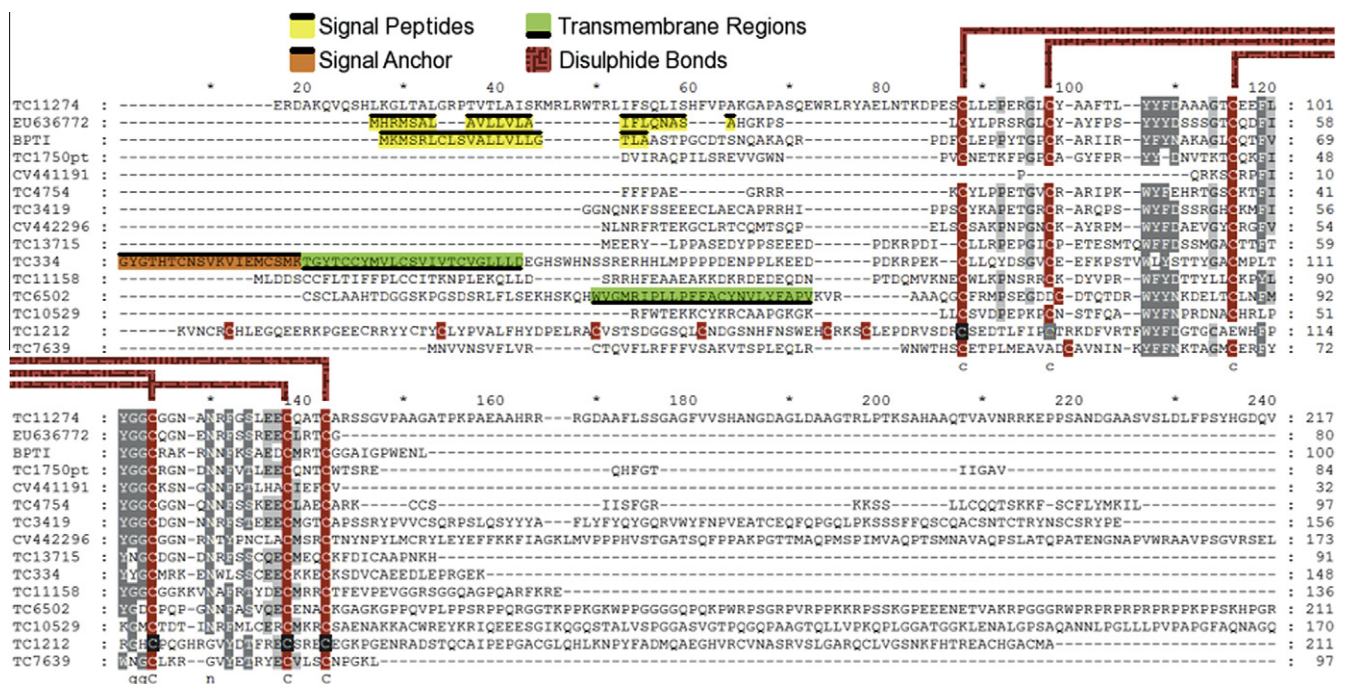
An outgroup of five soft tick sequences was used to study the relationships between the individual Kunitz domains from the homologues (Fig. 6). The alignment of the individual Kunitz domain nucleotide sequences is shown in Supplementary Fig. S2. Three taxon sets were defined and used to calibrate the phylogenetic tree and the recovered divergence times of the sets, together with the lower and upper 95% highest posterior density intervals (HPDs), are given in Table 2. Bayesian and ME analysis showed that the 93 Kunitz domains grouped into two major clades (Clades A and B) and one minor clade (Clade V), while the five soft tick sequences formed a basal outgroup to the *R. microplus* sequences. The major clades could each be subdivided into two minor clades (Clade A: Clades I and II; Clade B: Clades III and IV) resulting in a total of five minor clades. Fifty-one of the 96 nodes (53%) had posterior probability support, while 33 nodes had bootstrap support (34%).

#### 3.4. Evidence for domain duplications as a method of family expansion

We analysed the phylogenetic tree of the individual Kunitz domains for evidence of domain duplication. By focusing only on clades containing at least three individual Kunitz domains and a



**Fig. 1.** Distribution of Kunitz domain families in *Rhipicephalus microplus*. The Bilaris (Bi) and Monolaris (Mono) families are the predominant Kunitz domain families within the cattle tick, making up 41% and 33% of the total, respectively. The rest of the families (Tri-Kunitz to Nona-Kunitz) collectively make up less than one-third of the whole.



**Fig. 2.** Alignment of the Monolaris family in *Rhipicephalus microplus*. Fourteen of the 42 Kunitz domain-containing sequences retrieved during BLAST searches had a single predicted Kunitz domain, placing them in the Monolaris group. The query 'bovine pancreatic trypsin inhibitor' (BPTI, GenBank accession number 115114), which also has a single Kunitz domain, aligns with this group. The conserved cysteines involved in the disulphide bond formation and the bonds themselves are indicated. Bioinformatic prediction software showed that two of these have signal peptides, another has a signal anchor and two sequences have transmembrane regions. Black to grey shading represents conservation of amino acid similarity at each position with residues conserved between all sequences shaded in black. Sequence names are as obtained from the DFCI database, except for EU636772 which corresponds to a NCBI database entry.

maximum divergence time of 55 MYA, we identified nine sub-clades with evidence for both recent and ancient domain duplication (Fig. 6 and Table 3). All of the Kunitz domains that formed part of the nine sub-clades, except for CV450668K3 in Clade IIIb, were duplications of the other Kunitz domains within their respective sub-clade. The sub-clade sizes ranged from 15 Kunitz domains within Clade IIc to three sub-clades containing only three domains (Clade Ia, Clade IIIa, Clade IIIC).

The 15 Kunitz domains that formed Clade IIc belonged to only seven transcripts, and nine were Kunitz domains from transcripts TC1750 and TC4684. The Kunitz domains of these two homologues were spread over Clades I and II, but Clade I only contained a single domain from each transcript (TC1750K2 and TC4683K3) (Fig. 7). The Kunitz domains present in Clade II could be divided into three groups: TC1750K7 + TC4684K9, TC1750K1 + TC4684K2, with the third group containing the rest of the sequences. When the complete sequences of TC1750 and TC4684 were aligned their overall amino acid identity and similarity were 58% and 66%, respectively (Fig. 8).

#### 4. Discussion

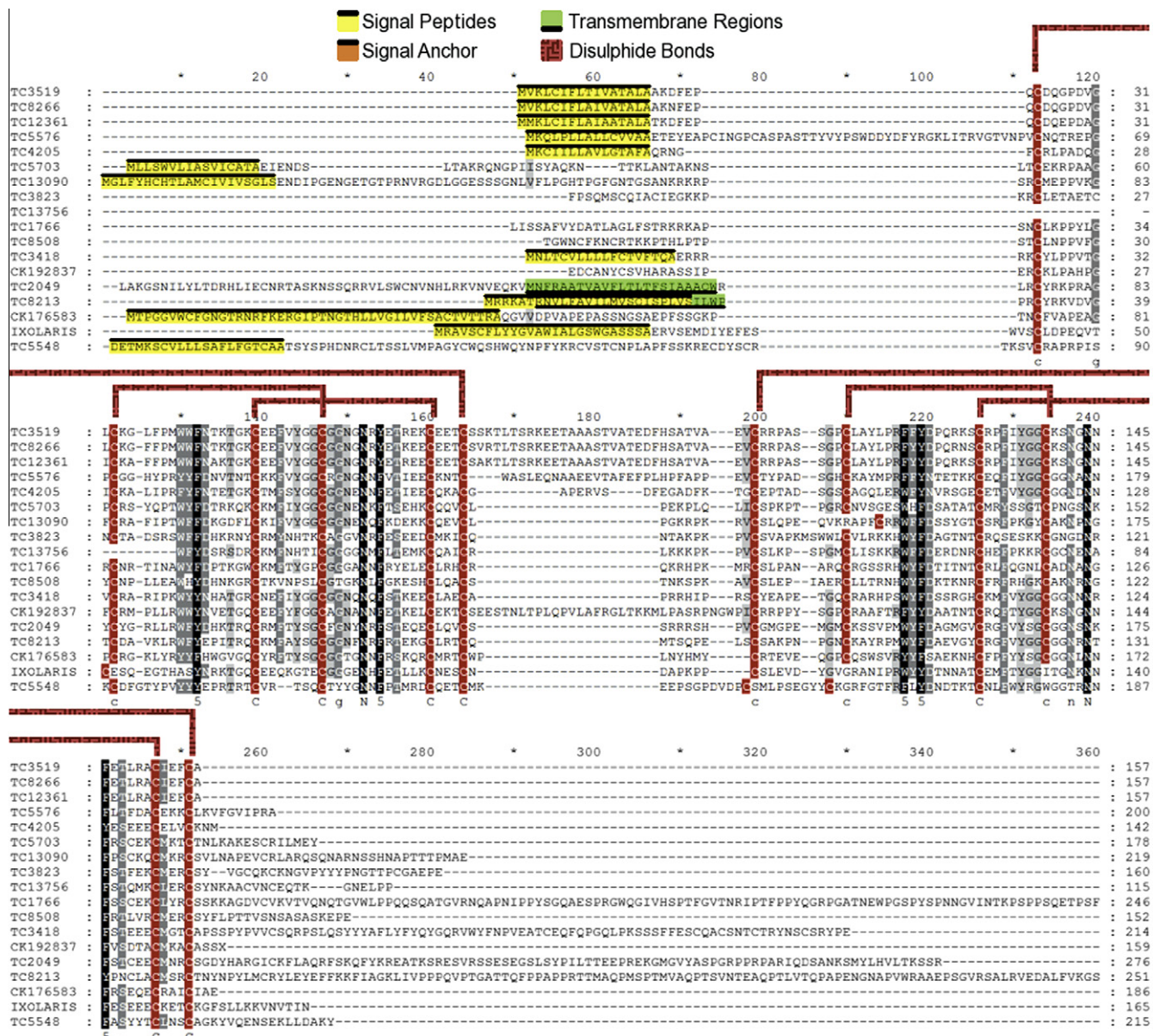
We set out to determine whether a family of TFPI-like Kunitz domain-containing proteins is present within *R. microplus*, and whether such a family had evolved through lineage-specific expansion of ancestral Kunitz domains. We successfully identified a TFPI-like Kunitz domain-containing protein family comprising 42 proteins that evolved through both ancient and recent domain- and gene duplication. This correlates with previous hard and soft tick studies in which similar expansions of secretory protein families, including Kunitz protein families, were observed. These lineage-specific expansions likely occurred during individual adaptations of hard and soft ticks to blood-feeding behaviour after their divergence from the ancestral tick (Mans et al., 2002b, 2008a;

Mans and Neitz, 2004). It is apparent that the Kunitz protein families are of integral importance to the haematophagous lifestyle of both hard and soft ticks.

Several previously identified tick Kunitz proteins have been shown to have anti-haemostatic functions, but despite their similarity to both TFPI and BPTI most have been shown to inhibit their targets through novel and non-canonical binding methods (Corral-Rodríguez et al., 2009). In some cases, loss of the C<sub>2</sub>–C<sub>4</sub> disulphide bond is thought to further facilitate unconventional binding manners (Demchenko, 2001). Furthermore, tick-derived Kunitz proteins that inhibit novel and unexpected targets have been described. In the family of TFPI-like Kunitz proteins we identified during this study, we noted both losses of the cysteines responsible for the C<sub>2</sub>–C<sub>4</sub> disulphide bridge as well as presentation of P<sub>1</sub> residues that were not expected to facilitate canonical inhibition of vital proteases of the blood coagulation cascade. This indicates that the TFPI-like Kunitz domain-containing proteins of *R. microplus* likely bind to their targets through novel ways and/or cause inhibition of unconventional targets. This family could therefore be a rich source of novel Kunitz-type inhibitors with diverse and unexpected targets.

One of the aims of this study was to identify TFPI homologues in *R. microplus* that may inhibit blood coagulation, similar to the anti-haemostatics, ixolaris and pentharalis, that were found in *I. scapularis* (Francischetti et al., 2002, 2004). We discerned a family of 42 of these proteins, which was comprised primarily of Kunitz domain-containing proteins belonging to the Monolaris and Bilaris groups, with the Bilaris group predominating. This distribution of Kunitz proteins is similar to that already reported for ticks where the Monolaris and Bilaris groups are the dominant groups. Which group predominates differs among species, for example in the hard tick *I. scapularis* the Monolaris group is approximately twice the size of the Bilaris group and this lineage-specific expansion has been shown to impart long-term blood feeding functionality (Dai





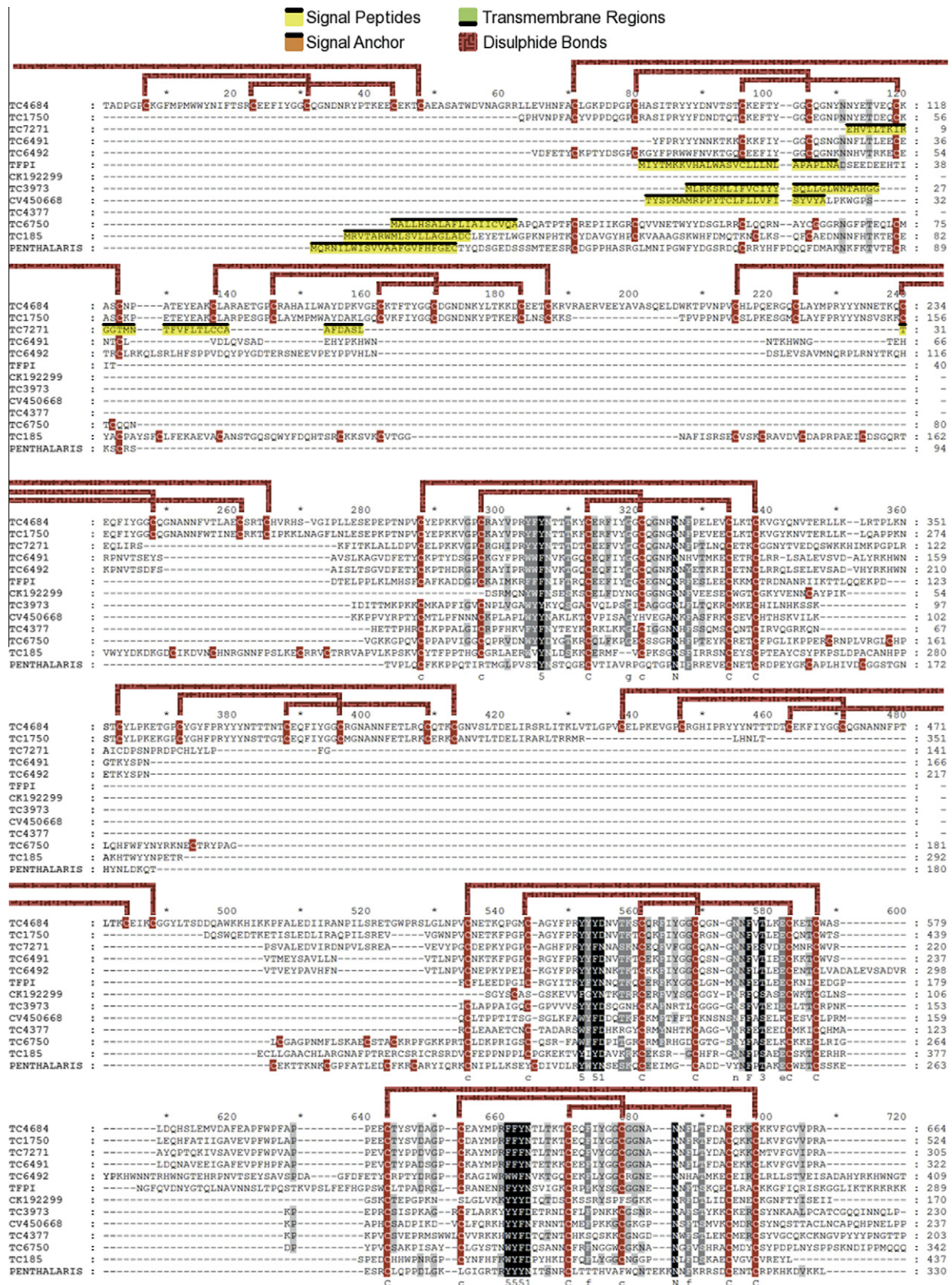
**Fig. 3.** Alignment of the Bilaris family in *Rhipicephalus microplus*. Seventeen sequences had two predicted Kunitz domains, placing them in the Bilaris group. 'Ixolaris' (GenBank accession number 15077002), from the *Ixodes scapularis* Bilaris family, is the query that aligns to this family. The conserved cysteines involved in the disulphide bond formation and the bonds themselves are indicated. Bioinformatic prediction software showed that 12 of these have signal peptides and two sequences have transmembrane regions. Black to grey shading represents conservation of amino acid similarity at each position with residues conserved between all sequences shaded in black. In the case of TC8213 the signal peptide region spans over the transmembrane region and therefore only the last four amino acids that are part of the transmembrane region are shown. Sequence names correspond to DFCI database entries.

et al., 2012). Proteins containing three or more Kunitz domains have to date only been observed in hard ticks. In fact, the observation of a Nonalaris transcript within ticks (TC4684) is the first to our knowledge. The Heptalaris transcript present in this study (TC1750) had previously been described by Rachinsky et al. (2007) (identified as TC9311) who suggested that due to variable splicing it could give rise to several shorter Kunitz domain-containing proteins. Whether this is also the case for TC4684, a close homologue of TC1750, should be confirmed through additional proteomic studies. Many of the transcripts were truncated, which prevented complete sequence annotation. Nonetheless, several transcripts possessed predicted signal peptides. This supports the expected function of these proteins, namely that anticoagulant Kunitz proteins would need to localise in the extracellular matrix so

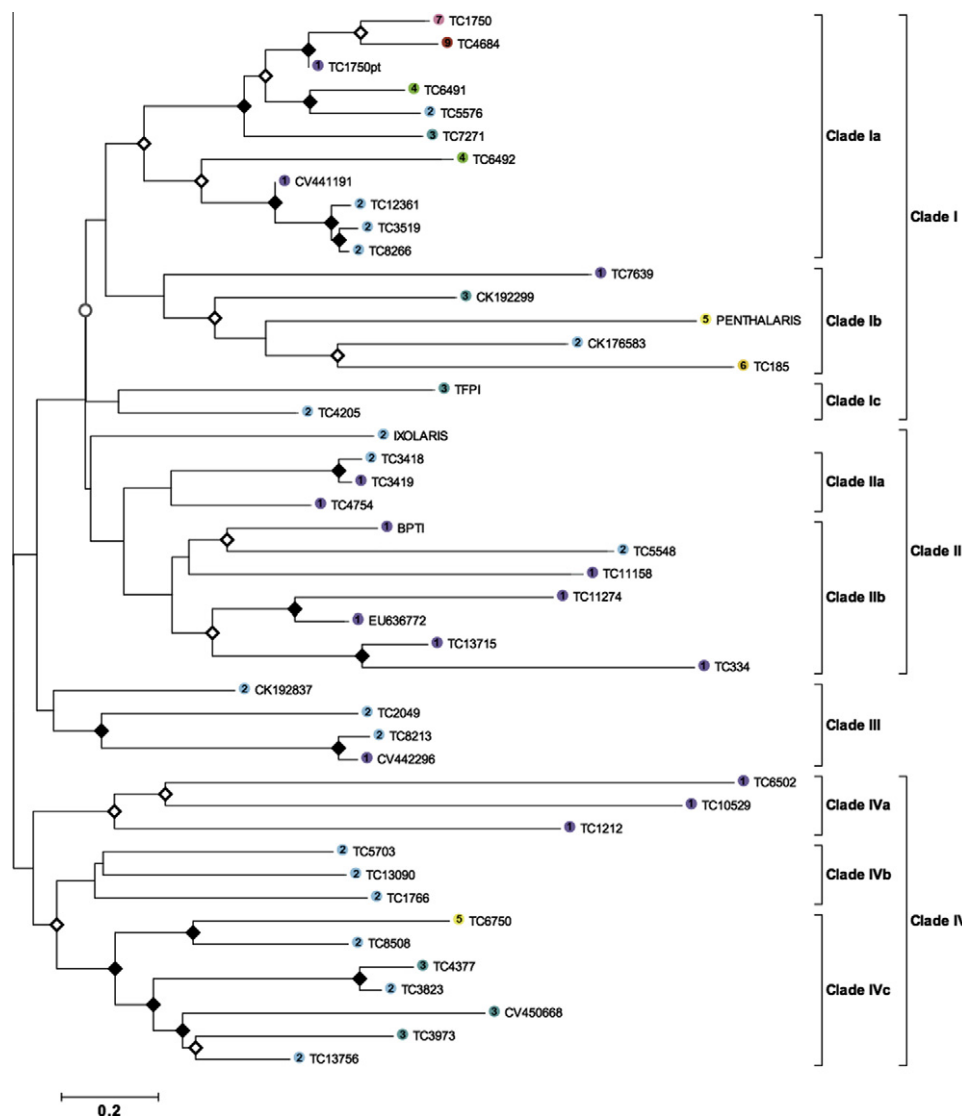
that they may be secreted into the feeding lesion or tick gut where maintenance of anti-haemostasis is required.

When the phylogeny of the TFPI-like Kunitz domain-containing proteins is considered, it is evident that the protein sequences did not group strongly according to the number of Kunitz domains they possess. Instead, all of the major clades are comprised of sequences belonging to at least two of the Kunitz families (Clades I–IV, Fig. 5) with Monalaris-type sequences generally in a basal position to the various clades. The phylogeny also showed some indications of gene duplication events, such as Monalaris sequences and sub-clades that were basal to multi-Kunitz sequences, as well as sequence groups in each major clade with strong support and short branch lengths, indicating little divergence from their most recent common ancestor (MRCA). Furthermore, the phylogeny of





**Fig. 4.** Alignment of the Trilaris and greater multilaris families in *Rhipicephalus microplus*. Eleven sequences had three or more predicted Kunitz domains, placing them in the groups ranging from the Trilaris to Nonalaris families. Seven of these have predicted signal peptides. Both 'tissue factor pathway inhibitor' (TFPI, GenBank accession number 125932) and 'penthalaris' (GenBank accession number 22164266) are queries that align well with these families. The conserved cysteines involved in the disulphide bond formation and the bonds themselves are indicated. Black to grey shading represents conservation of amino acid similarity at each position with residues conserved between all sequences shaded in black.



**Fig. 5.** Phylogenetic analysis of tissue factor pathway inhibitor-like Kunitz proteins in *Rhipicephalus microplus*. A Neighbor-Joining analysis with 10,000 bootstrap replicates was performed on the four query sequences as well as the 42 homologues that were retrieved during BLAST searches. Nodes with open diamonds indicate  $\geq 50\%$  confidence while filled diamonds are for  $\geq 75\%$  confidence. The scale is the number of amino acid substitutions per site. Coloured (shaded) circles show the number of Kunitz domains for each transcript and are as shown in Fig. 1. The open circle is included to show a node with a branch length that is less than 0.015 and difficult to discern. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

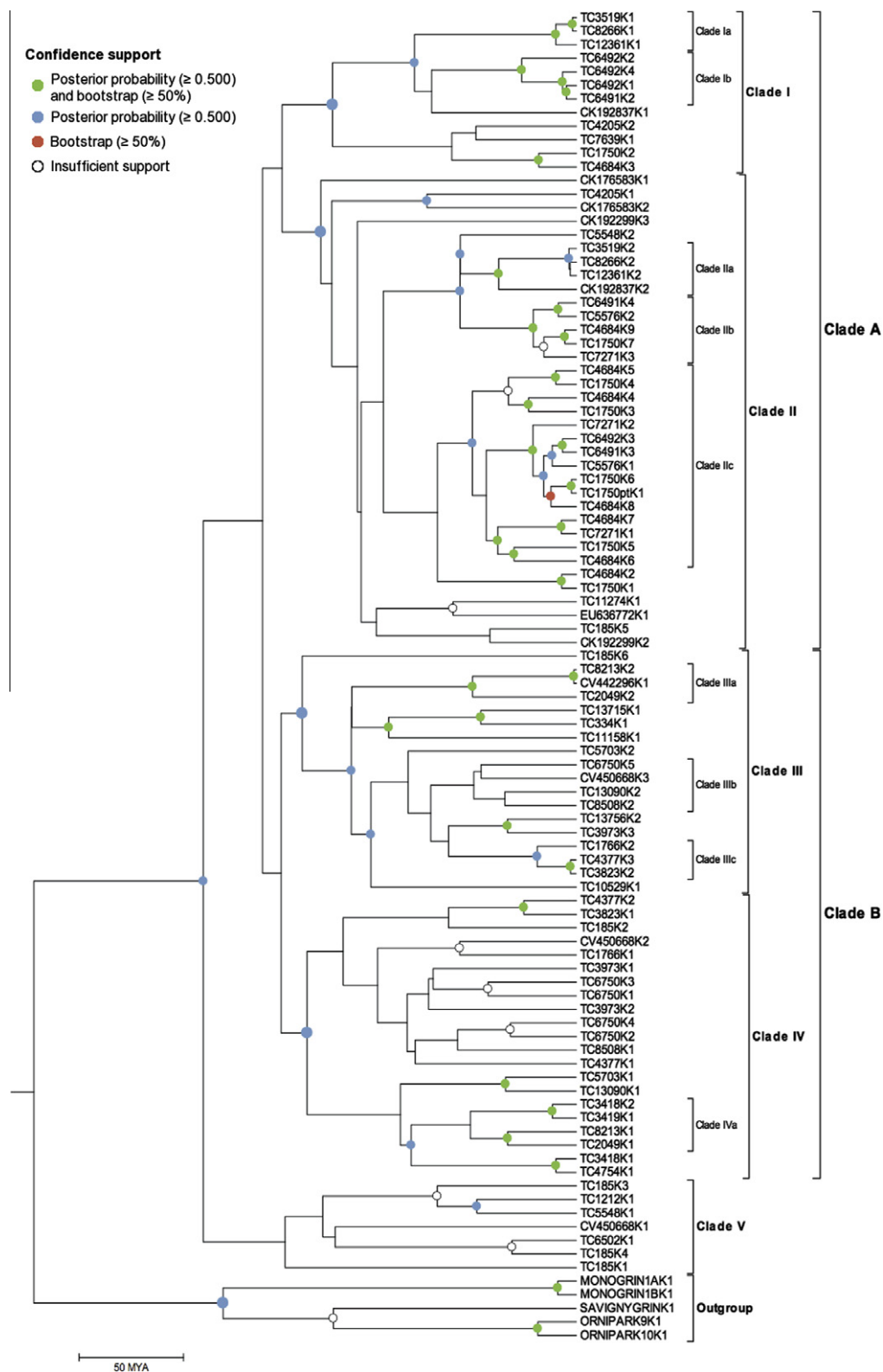
the individual Kunitz domains formed two major clades (Clades A and B, Fig. 6) that comprised almost all of the Kunitz domains from multi-domain transcripts. When the two phylogenies were compared, it became apparent that, except for seven transcripts (CK192837, TC185, TC2049, TC3418, TC5548, TC8213 and TC13756), all of the multi-Kunitz transcripts within Clade I have their Kunitz domains within Clade A. Similarly, multi-Kunitz transcripts residing in Clade IV exclusively grouped within Clade B. We therefore propose that the common ancestor of Clades A and B was likely a Monolaris-type transcript that underwent gene duplication and subsequent expansion to give rise to the two major families of Kunitz domain-containing proteins within *R. microplus*.

The estimated divergence times of the Ixodida, Argasidae and Metastriate lineages, as calculated in this paper by the phylogeny of the individual Kunitz domains, corresponded well with those presented by Jeyaprakash and Hoy (2009). In all cases there was an overlap between the 95% confidence intervals of the divergence times and in the case of the Argasidae the mean divergence time calculated in the present study (214 MYA) corresponded precisely with that previously estimated by Jeyaprakash and

Hoy (2009). The mean haploid genome size of *R. microplus* (~7.1 Gbp) has been reported as approximately double that of the rest of the Ixodidae (~2.67 Gbp) (Ullmann et al., 2005; Geraci et al., 2007). This difference in size could be indicative of a whole genome duplication event in this tick. The MRCA of the two major Kunitz domain clades that we observed was estimated to have occurred at 150 MYA and the subsequent divergence of these clades could correspond to such a genome duplication event. Therefore, we propose that the ancient whole genome duplication that is expected to have occurred within *R. microplus* took place at least 150 MYA.

Further analysis of the sub-clades within the phylogeny of the individual Kunitz domains revealed evidence for both ancient and recent gene and domain duplications. We observed several recent gene duplications and all of these had divergence times of less than 11.4 MYA, with the most recent divergence only 1.05 MYA. The most convincing evidence for duplication, manifesting both gene and domain duplications, was in the cases of TC1750 and TC4684. A potential evolutionary pathway for these transcripts is discussed below and shown in Fig. 9.





**Fig. 6.** Phylogenetic analysis of individual Kunitz domains from homologues. The Bayesian Evolutionary Analysis by Sampling Trees (BEAST) generated tree for the individual complete Kunitz domains from the various homologues, as well as five soft tick sequences that form an outgroup, is shown. The 93 *Rhipicephalus microplus* Kunitz domains group into five minor clades (Clades I–V). Confidence support from the BEAST and Minimum Evolution (ME) tree is indicated with coloured (shaded) circles: green (black) shows both posterior probability ( $\geq 0.500$ ) and bootstrap support ( $\geq 50\%$ ); blue (dark grey) shows only posterior probability support for the branch; red (light grey) shows the branch has bootstrap support; open circles indicate that the branch is predicted by both the BEAST and the ME trees but does not have sufficient support from either. The scale is 50 million years ago (MYA) with the age of the ancestral node recovered as  $255.3 \pm 0.9$  MYA (95% highest posterior density interval (HPD): 204.9, 310.2). Specific calibration points used as well as the divergence times recovered in this study are given in Table 2. The individual Kunitz domains were named by appending K and the Kunitz domain number within the homologue to the homologue's name, for example the third Kunitz domain of TC185 is named TC185K3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Taxon set calibration points used and recovered. Three taxon sets (Ixodida, Argasidae and Metastriate) were defined with divergence times as calculated by Jeyaprakash and Hoy (2009). The mean divergence times that were recovered for these taxa during phylogenetic analysis are shown.

| Taxon sets  | Calibration points<br>MYA ± SD (95% CI) | Recovered divergence times<br>MYA ± SD (95% HPD) |
|-------------|---|--|
| Ixodida     | 300 ± 27 (245, 351)                     | 255.3 ± 0.9 (204.9, 310.2)                       |
| Argasidae   | 214 ± 28 (160, 269)                     | 213.9 ± 1.5 (154.5, 272.7)                       |
| Metastriate | 134 ± 22 (95, 181)                      | 164.9 ± 0.4 (127.4, 204.2)                       |

MYA, million years ago; CI, confidence interval; HPD, highest posterior density interval.

The evolutionary path of the TC1750 and TC4684 homologues within *R. microplus* potentially started with the duplication of a

single ancestral Kunitz domain belonging to a Monolaris-type protein (Fig. 9). One of the domains could subsequently have become the ancestor of TC1750K2, TC4684K1 and TC4684K3 (we will call this the primary domain – Kunitz domains from Clade I, Fig. 6), while the other became the ancestor of all the other Kunitz domains within TC1750 and TC4684 (the secondary domain – Kunitz domains from Clade II, Fig. 6). A series of domain duplication events then probably followed, during which five additional copies of the secondary domain were produced. At this stage, the MRCA of TC1750 and TC4684 must have existed, namely a 7-Kunitz protein comprising of a single primary Kunitz domain and six secondary Kunitz domains.

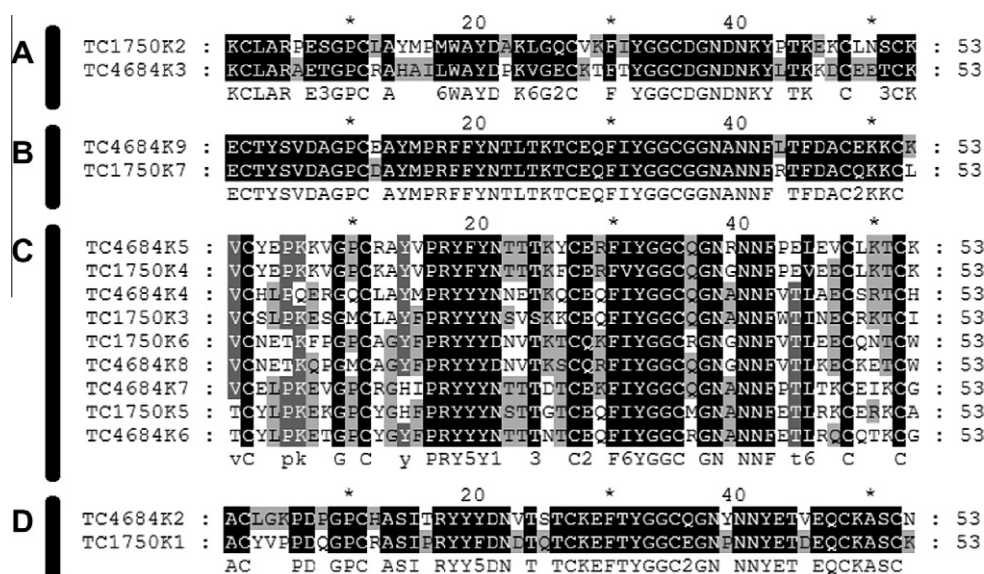
After the initial evolutionary processes involving domain duplications and expansions, a single gene duplication must have followed to allow divergence of TC1750 and TC4684. TC1750 probably continued to diverge without any further intra-transcript

**Table 3**

Recent Kunitz domain duplications. Of the nine sub-clades that were studied to find domain duplications (Fig. 6), seven sub-clades had sequences that were recently duplicated, indicated by an amino acid identity of more than 85% and a nucleotide identity of at least 90%. In this manner, eight groups of sequences that showed recent duplication were identified and the divergence times, amino acid identity and similarity as well as the nucleotide identity within these groups are shown.

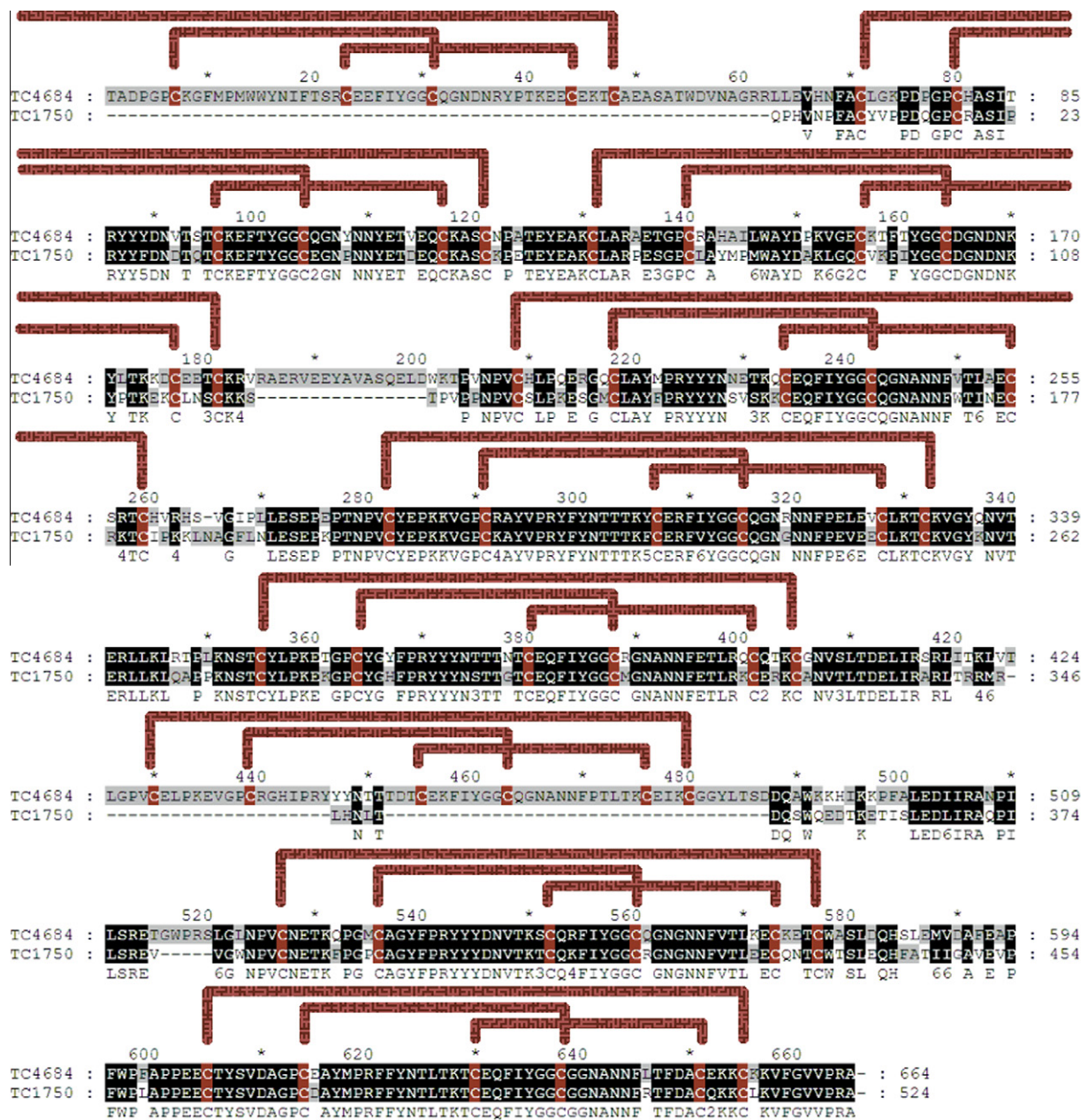
| Sub-clade  | Kunitz domain                                  | Divergence time (MYA) | Amino acid identity (%) | Amino acid similarity (%) | Nucleotide identity (%) |
|------------|--|-----------------------|-------------------------|---------------------------|-------------------------|
| Clade Ia   | TC3519K1<br>TC8266K1<br>TC12361K1              | 9.86                  | 86–94                   | 92–100                    | 91–94                   |
| Clade Ib   | TC6491K2<br>TC6492K1                           | 4.73                  | 96                      | 98                        | 94                      |
| Clade IIa  | TC3519K2<br>TC8266K2<br>TC12361K2              | 3.56                  | 96–98                   | 98–100                    | 96                      |
| Clade IIb  | TC5576K2<br>TC6491K4                           | 8.62                  | 96                      | 98                        | 92                      |
| Clade IIc  | TC1750K4<br>TC4684K5<br>TC1750K6<br>TC1750ptK1 | 9.82<br>2.27          | 88<br>96                | 96<br>96                  | 90<br>94                |
| Clade IIIa | CV442296K1<br>TC8213K2                         | 1.05                  | 100                     | 100                       | 96                      |
| Clade IVa  | TC3418K2<br>TC3419K1                           | 11.38                 | 90                      | 98                        | 93                      |

MYA, million years ago.



**Fig. 7.** Alignment of individual Kunitz domains from the closely homologous multimeric sequences TC1750 and TC4684. When the individual complete Kunitz domains of TC1750 and TC4684 are aligned they form four groups. Only group A falls within Clade I while groups B–D form part of Clade II. This suggests that an ancient domain duplication resulted in the ancestors of the Kunitz domains present in Clades I and II, and those in Clade II underwent further duplications to give the current transcripts. Black to grey shading represents conservation of amino acid similarity at each position with residues conserved between all sequences shaded in black.





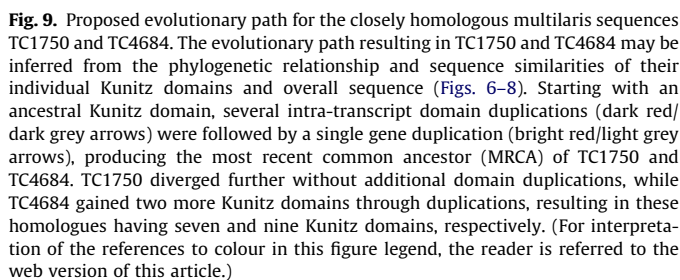
**Fig. 8.** Alignment of the closely homologous multimeric sequences TC1750 and TC4684. By aligning the full amino acid sequences of TC1750 and TC4684 it is clear that these two sequences evolved by gene duplication of a common ancestor. Furthermore TC4684 diverged from TC1750 by acquiring two more Kunitz domains through intra-domain duplication – these two domains were inserted before TC1750K1 and between TC1750K5 and TC1750K6. The overall amino acid identity between the two sequences is 58% with the amino acid similarity being 66%. The conserved cysteines involved in the disulphide bond formation and the bonds themselves are indicated. Black to grey shading represents conservation of amino acid similarity at each position with residues conserved between all sequences shaded in black.

domain duplications, resulting in seven Kunitz domains and placing it in the Heptalaris group of Kunitz proteins. It seems that one domain (TC1750K6) could have been duplicated from TC1750 to give rise to a new, separate Kunitz protein, namely TC1750ptK1, but a single nonsense mutation resulted in the appearance of a premature stop codon (Tyr → Stop).

TC4684 must have undergone a minimum of two further domain duplications to result in its current form consisting of at least nine Kunitz domains (Nonalaris group). Conceivably, one domain duplication produced TC4684K1 (primary domain type) and another produced TC4684K7 (secondary domain type). As TC4684 has an initial truncated Kunitz domain it is possible that it possesses additional Kunitz domains. Despite the inferred gene

duplication event allowing TC1750 and TC4684 to diverge from each other, they still retain a high degree of amino acid identity (58%) and similarity (66%).

When only a single gene copy is present, it is subject to purifying selection that prevents the accumulation of deleterious mutations (Ohno, 1970). However, as soon as gene duplication occurs, both homologues experience relaxed selection constraints that allow them to enter a period of accelerated evolution (Lynch and Katju, 2004). This could eventually result in novel functionality, whether through different temporal or spatial expression, the acquisition of additional targets or neofunctionalisation. Establishment of inhibitor variability is often associated with gene and domain duplication (Hill and Hastie, 1987; Christeller, 2005). Thus, it is common for small inhibitor families to contain individual



It is evident that the expansion of the TFPI-like Kunitz domain-containing protein family in *R. microplus* was facilitated through several gene and domain duplications, as well as a single genome duplication event. Such expansions are present in both hard and

During the course of this study we identified an extensive family of TFPI-like Kunitz domain-containing proteins in *R. microplus*. The distribution of these is similar to that previously found in other ticks such as *I. scapularis* and approximately half of these homologues have predicted signal peptides and present P<sub>1</sub>Lys or P<sub>1</sub>Arg, supporting their expected role as secreted anti-haemostatic proteins. Nonetheless, unexpected and novel functions have often been described for tick-derived Kunitz proteins (Corral-Rodríguez et al., 2009) and further studies will be needed to confirm the anti-haemostatic functions of these proteins. We have found considerable evidence for gene and domain duplications within this



family, including both ancient and recent duplications. We believe that this report represents the first comprehensive study of TFPI-like Kunitz proteins within *R. microplus* and expands our current knowledge of the evolution of Kunitz domain-containing proteins within this tick. However, the homologues we identified were only those present within the NCBI and *R. microplus* EST databases (BmiGI). A first draft of the *R. microplus* genome sequence has recently been released (Bellgard et al., 2012) and once full coverage of the genome is available, an exhaustive analysis of all TFPI-like Kunitz domain-containing proteins within *R. microplus* would be possible.

## Acknowledgements

This project was partially funded by the Wellcome Trust, United Kingdom, under the 'Animal Health in the Developing World' initiative through Project 0757990 entitled 'Adapting recombinant anti-tick vaccines to livestock in Africa'.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ijpara.2012.11.006>.

## References

- Aларcon-Chaidez, F.J., Sun, J., Wikel, S.K., 2007. Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). *Insect Biochem. Mol. Biol.* 37, 48–71.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Batista, I.F.C., Chudzinski-Tavassi, A.M., Faria, F., Simons, S.M., Barros-Batesti, D.M., Labruna, M.B., Leão, L.L., Ho, P.L., Junqueira-de-Azevedo, I.L.M., 2008. Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acar: Ixodidae). *Toxicon* 51, 823–834.
- Bellgard, M.I., Moolhuijzen, P.M., Guerrero, F.D., Schibeci, D., Rodriguez-Valle, M., Peterson, D.G., Dowd, S.E., Barrero, R., Hunter, A., Miller, R.J., Lew-Tabor, A.E., 2012. CattleTickBase: an integrated internet-based bioinformatics resource for *Rhipicephalus (Boophilus) microplus*. *Int. J. Parasitol.* 42, 161–169.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Burck, P.J., Hamill, R.L., Cerwinsky, E.W., Grinnan, E.L., 1967. Acidic bovine pancreatic trypsin inhibitor. II. Specificity of inhibition. *Biochemistry* 6, 3180–3184.
- Cerwinsky, E.W., Burck, P.J., Grinnan, E.L., 1967. Acidic bovine pancreatic trypsin inhibitor. I. Purification and physical characterization. *Biochemistry* 6, 3175–3179.
- Chmelaf, J., Anderson, J.M., Mu, J., Jochim, R.C., Valenzuela, J.G., Kopecký, J., 2008. Insight into the sialome of the castor bean tick, *Ixodes ricinus*. *BMC Genom.* 9, 233–253.
- Christeller, J.T., 2005. Evolutionary mechanisms acting on proteinase inhibitor variability. *FEBS J.* 272, 5710–5722.
- Chu, D., Bungiro Jr., R.D., Ibanez, M., Harrison, L.M., Campodonico, E., Jones, B.J., Mieszczynek, J., Kuzmic, P., Cappello, M., 2004. Molecular characterization of *Ancylostoma ceylanicum* Kunitz-type serine protease inhibitor: evidence for a role in hookworm-associated growth delay. *Infect. Immun.* 72, 2214–2221.
- Corral-Rodríguez, M.Á., Macedo-Ribeiro, S., Pereira, P.J.B., Fuentes-Prior, P., 2009. Tick-derived Kunitz-type inhibitors as antihemostatic factors. *Insect Biochem. Mol. Biol.* 39, 579–595.
- Costa, A.F.D.V., Gasser, R.B., Dias, S.R.C., Rabelo, É.M.L., 2009. Male-enriched transcription of genes encoding ASPs and Kunitz-type protease inhibitors in *Ancylostoma* species. *Mol. Cell. Probes* 23, 298–303.
- Dai, S.-X., Zhang, A.-D., Huang, J.-F., 2012. Evolution, expansion and expression of the Kunitz/BPTI gene family associated with long-term blood feeding in *Ixodes scapularis*. *BMC Evol. Biol.* 12, 4.
- de la Fuente, J., Almazán, C., Canales, M., de la Lastra, J.M.P., Kocan, K.M., Willadsen, P., 2007a. A ten-year review of commercial vaccine performance for control of tick infestations on cattle. *Anim. Health. Res. Rev.* 8, 23–28.
- de la Fuente, J., Blouin, E.F., Manzano-Roman, R., Naranjo, V., Almazán, C., de la Lastra, J.M.P., Zivkovic, Z., Jongejans, F., Kocan, K.M., 2007b. Functional genomic studies of tick cells in response to infection with the cattle pathogen, *Anaplasma marginale*. *Genomics* 90, 712–722.
- de Miranda Santos, I.K.F., Valenzuela, J.G., Ribeiro, J.M.C., de Castro, M., Costa, J.N., Costa, A.M., da Silva, E.R., Neto, O.B.R., Rocha, C., Daffre, S., Ferreira, B.R., da Silva, J.S., Szabó, M.P., Bechara, G.H., 2004. Gene discovery in *Boophilus microplus*, the cattle tick: the transcriptomes of ovaries, salivary glands and hemocytes. *Ann. NY Acad. Sci.* 1026, 242–246.
- Demchenko, A.P., 2001. Recognition between flexible protein molecules: Induced and assisted folding. *J. Mol. Recognit.* 14, 42–61.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Estrada-Peña, A., Sánchez Acedo, C., Quílez, J., Del Cacho, E., 2005. A retrospective study of climatic suitability for the tick *Rhipicephalus (Boophilus) microplus* in the Americas. *Global Ecol. Biogeogr.* 14, 565–573.
- Finn, R.D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunesekekan, P., Ceric, G., Forslund, K., Holm, L., Sonnhhammer, E.L., Eddy, S.R., Bateman, A., 2010. The Pfam protein families database. *Nucleic Acids Res.* D211–D222.
- Francischetti, I.M.B., Mans, B.J., Meng, Z., Gudderra, N., Veenstra, T.D., Pham, V.M., Ribeiro, J.M.C., 2008a. An insight into the sialome of the soft tick, *Ornithodoros parkeri*. *Insect Biochem. Mol. Biol.* 38, 1–21.
- Francischetti, I.M.B., Mather, T.N., Ribeiro, J.M.C., 2004. Penthalgarin, a novel recombinant five-Kunitz tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick vector of Lyme disease, *Ixodes scapularis*. *Thromb. Haemost.* 91, 886–898.
- Francischetti, I.M.B., Meng, Z., Mans, B.J., Gudderra, N., Hall, M., Veenstra, T.D., Pham, V.M., Kotsyfakis, M., Ribeiro, J.M.C., 2008b. An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. *J. Proteomics* 71, 493–512.
- Francischetti, I.M.B., Pham, V.M., Mans, B.J., Andersen, J.F., Mather, T.N., Lane, R.S., Ribeiro, J.M.C., 2005. The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). *Insect Biochem. Mol. Biol.* 35, 1142–1161.
- Francischetti, I.M.B., Valenzuela, J.G., Andersen, J.F., Mather, T.N., Ribeiro, J.M.C., 2002. Ixolarin, a novel recombinant tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick, *Ixodes scapularis*: identification of factor X and factor Xa as scaffolds for the inhibition of factor VIIa/tissue factor complex. *Blood* 99, 3602–3612.
- Fry, B.G., Roelants, K., Champagne, D.E., Scheib, H., Tyndall, J.D.A., King, G.F., Nevalainen, T.J., Norman, J.A., Lewis, R.J., Norton, R.S., Renjifo, C., de la Vega, R.C.R., 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* 10, 483–511.
- Geraci, N.S., Johnston, J.S., Robinson, J.P., Wikel, S.K., Hill, C.A., 2007. Variation in genome size of argasid and ixodid ticks. *Insect Biochem. Mol. Biol.* 37, 399–408.
- González, S., Fló, M., Margenat, M., Durán, R., González-Sapienza, G., Graña, M., Parkinson, J., Maizels, R.M., Salinas, G., Alvarez, B., Fernández, C., 2009. A family of diverse Kunitz inhibitors from *Echinococcus granulosus* potentially involved in host-parasite cross-talk. *PLoS ONE* 4, e7009.
- Guerrero, F.D., Moolhuijzen, P., Peterson, D.G., Bidwell, S., Caler, E., Bellgard, M.I., Nene, V.M., Djikeng, A., 2010. Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, *Rhipicephalus (Boophilus) microplus*. *BMC Genom.* 11, 374.
- Helland, R., Otlewski, J., Sundheim, O., Dadlez, M., Smalås, A.O., 1999. The crystal structures of the complexes between bovine  $\beta$ -trypsin and ten P1 variants of BPTI. *J. Mol. Biol.* 287, 923–942.
- Hill, C.A., Wikel, S.K., 2005. The *Ixodes scapularis* Genome Project: an opportunity for advancing tick research. *Trends Parasitol.* 21, 151–153.
- Hill, R.E., Hastie, N.D., 1987. Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature* 326, 96–99.
- Jeyaparakash, A., Hoy, M.A., 2009. First divergence time estimate of spiders, scorpions, mites and tick (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp. Appl. Acarol.* 47, 1–18.
- Joubert, A.M., Louw, A.I., Joubert, F., Neitz, A.W.H., 1998. Cloning, nucleotide sequence and expression of the gene encoding factor Xa inhibitor from the salivary glands of the tick, *Ornithodoros savignyi*. *Exp. Appl. Acarol.* 22, 603–619.
- Kahsay, R.Y., Gao, G., Liao, L., 2005. An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* 21, 1853–1858.
- Karczewski, J., Connolly, T.M., 1997. The interaction of disaggregin with platelet fibrinogen receptor, glycoprotein IIb–IIIa. *Biochem. Biophys. Res. Commun.* 241, 744–748.
- Karczewski, J., Endris, R., Connolly, T.M., 1994. Disaggregin is a fibrinogen receptor antagonist lacking the Arg–Gly–Asp sequence from the tick, *Ornithodoros moubata*. *J. Biol. Chem.* 269, 6702–6708.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentine, F., Wallace, I.M., Willm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Letunic, I., Doerks, T., Bork, P., 2008. SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232.
- Liao, M., Zhou, J., Gong, H., Bolbaatar, D., Shirafuji, R., Battur, B., Nishikawa, Y., Fujisaki, K., 2009. Hemalin, a thrombin inhibitor isolated from a midgut cDNA library from the hard tick *Haemaphysalis longicornis*. *J. Insect Physiol.* 55, 164–173.
- Lima, C.A., Torquato, R.J.S., Sasaki, S.D., Justo, G.Z., Tanaka, A.S., 2010. Biochemical characterization of a Kunitz type inhibitor similar to dendrotoxins produced by *Rhipicephalus (Boophilus) microplus* (Acari: Ixodidae) hemocytes. *Vet. Parasitol.* 167, 279–287.
- Lwaleed, B.A., Bass, P.S., 2006. Tissue factor pathway inhibitor: structure, biology and involvement in disease. *J. Pathol.* 208, 327–339.
- Lynch, M., Katju, V., 2004. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* 20, 544–549.

- Lynen, G., Zeman, P., Bakuname, C., Di Giulio, G., Mtui, P., Sanka, P., Jongejan, F., 2008. Shifts in the distributional ranges of *Boophilus* ticks in Tanzania: evidence that a parapatric boundary between *Boophilus microplus* and *B. decoloratus* follows climate gradients. *Exp. Appl. Acarol.* 44, 147–164.
- Macedo-Ribeiro, S., Almeida, C., Calisto, B.M., Friedrich, T., Mentel, R., Stürzebecher, J., Fuentes-Prior, P., Pereira, P.J.B., 2008. Isolation, cloning and structural characterisation of boophilin, a multifunctional Kunitz-type proteinase inhibitor from the cattle tick. *PLoS ONE* 3, e1624.
- Mans, B.J., Andersen, J.F., Francischetti, I.M.B., Valenzuela, J.G., Schwan, T.G., Pham, V.M., Garfield, M.K., Hammer, C.H., Ribeiro, J.M.C., 2008a. Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. *Insect Biochem. Mol. Biol.* 38, 42–58.
- Mans, B.J., Andersen, J.F., Schwan, T.G., Ribeiro, J.M.C., 2008b. Characterization of anti-hemostatic factors in the argasid, *Argas monolakensis*: implications for the evolution of blood-feeding in the soft tick family. *Insect Biochem. Mol. Biol.* 38, 22–41.
- Mans, B.J., Louw, A.I., Neitz, A.W.H., 2002a. Amino acid sequence and structure modeling of savignin, a thrombin inhibitor from the tick, *Ornithodoros savignyi*. *Insect Biochem. Mol. Biol.* 32, 821–828.
- Mans, B.J., Louw, A.I., Neitz, A.W.H., 2002b. Evolution of hematophagy in ticks: common origins for blood coagulation and platelet aggregation inhibitors from soft ticks of the genus *Ornithodoros*. *Mol. Biol. Evol.* 19, 1695–1705.
- Mans, B.J., Louw, A.I., Neitz, A.W.H., 2002c. Savignigrin, a platelet aggregation inhibitor from the soft tick *Ornithodoros savignyi*, presents the RGD integrin recognition motif on the Kunitz-BPTI fold. *J. Biol. Chem.* 277, 21371–21378.
- Mans, B.J., Neitz, A.W.H., 2004. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem. Mol. Biol.* 34, 1–17.
- Maritz-Olivier, C., Stutzer, C., Jongejan, F., Neitz, A.W.H., Gaspar, A.R.M., 2007. Tick anti-hemostatics: targets for future vaccines and therapeutics. *Trends Parasitol.* 23, 397–407.
- Monteiro, R.Q., Rezaie, A.R., Bae, J.S., Calvo, E., Andersen, J.F., Francischetti, I.M.B., 2008. Ixolarin binding to factor X reveals a precursor state of factor Xa heparin-binding exosite. *Protein Sci.* 17, 1–8.
- Monteiro, R.Q., Rezaie, A.R., Ribeiro, J.M.C., Francischetti, I.M.B., 2005. Ixolarin: a factor Xa heparin-binding exosite inhibitor. *Biochem. J.* 387, 871–877.
- Nazareth, R.A., Tomaz, L.S., Ortiz-Costa, S., Atella, G.C., Ribeiro, J.M.C., Francischetti, I.M.B., Monteiro, R.Q., 2006. Antithrombotic properties of Ixolarin, a potent inhibitor of the extrinsic pathway of the coagulation cascade. *J. Thromb. Haemost.* 96, 7–13.
- Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6.
- Ohno, S., 1970. *Evolution by Gene Duplication*. Springer-Verlag, Berlin, Germany.
- Ohta, T., 1994. On hypervariability at the reactive center of proteolytic enzymes and their inhibitors. *J. Mol. Evol.* 39, 614–619.
- Ohta, T., 2000. Evolution of gene families. *Gene* 259, 45–52.
- Peconick, A.P., Sossai, S., Girão, F.A., Rodrigues, M.Q.R.B., Souza e Silva, C.H., Guzman, F., Patarroyo, A.M., Vargas, M.I., Patarroyo, J.H., 2008. Synthetic vaccine (SBm7462) against the cattle tick *Rhipicephalus (Boophilus) microplus*: preservation of immunogenic determinants in different strains from South America. *Exp. Parasitol.* 119, 37–43.
- Rachinsky, A., Guerrero, F.D., Scoles, G.A., 2007. Differential protein expression in ovaries of uninfected and *Babesia*-infected southern cattle ticks, *Rhipicephalus (Boophilus) microplus*. *Insect Biochem. Mol. Biol.* 37, 1291–1308.
- Ribeiro, J.M.C., Alarcon-Chaidez, F.J., Francischetti, I.M.B., Mans, B.J., Mather, T.N., Valenzuela, J.G., Wikel, S.K., 2006. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mol. Biol.* 36, 111–129.
- Sasaki, S.D., Azzolini, S.S.A., Hirata, I.Y., Andreotti, R., Tanaka, A.S., 2004. *Boophilus microplus* tick larvae, a rich source of Kunitz type serine proteinase inhibitors. *Biochimie* 86, 643–649.
- Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–162.
- Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA* 95, 5857–5864.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Tanaka, A.S., Andreotti, R., Gomes, A., Torquato, R.J.S., Sampaio, M.U., Sampaio, C.A.M., 1999. A double headed serine proteinase inhibitor – human plasma kallikrein and elastase inhibitor – from *Boophilus microplus* larvae. *Immunopharmacology* 45, 171–177.
- Tsujiimoto, H., Kotsyfakis, M., Francischetti, I.M.B., Eum, J.H., Strand, M.R., Champagne, D.E., 2012. Simukunin from the salivary glands of the black fly *Simulium vittatum* inhibits enzymes that regulate clotting and inflammatory responses. *PLoS ONE* 7, e29964.
- Ullmann, A.J., Lima, C.M.R., Guerrero, F.D., Piesman, J., Black, W.C., 2005. Genome size and organization in the black-legged tick, *Ixodes scapularis*, and the Southern cattle tick, *Boophilus microplus*. *Insect Mol. Biol.* 14, 217–222.
- Valenzuela, J.G., Francischetti, I.M.B., Pham, V.M., Garfield, M.K., Mather, T.N., Ribeiro, J.M.C., 2002. Exploring the sialome of the tick *Ixodes scapularis*. *J. Exp. Biol.* 205, 2843–2864.
- van de Locht, A., Stubbs, M.T., Bode, W., Friedrich, T., Bollschweiler, C., Höffken, W., Huber, R., 1996. The ornithodorin–thrombin crystal structure, a key to the TAP enigma? *EMBO J.* 15, 6011–6017.
- van Zee, J.P., Geraci, N.S., Guerrero, F.D., Wikel, S.K., Stuart, J.J., Nene, V.M., Hill, C.A., 2007. Tick genomics: the *Ixodes* genome project and beyond. *Int. J. Parasitol.* 37, 1297–1305.
- Wei, A., Alexander, R.S., Duke, J., Ross, H., Rosenfeld, S.A., Chang, C.H., 1998. Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa. *J. Mol. Biol.* 283, 147–154.
- White, N., Sutherst, R.W., Hall, N., Whish-Wilson, P., 2003. The vulnerability of the Australian beef industry to impacts of the cattle tick (*Boophilus microplus*) under climate change. *Climatic Change* 61, 157–190.
- Willadsen, P., 2004. Anti-tick vaccines. *Parasitology* 129, S367–S387.