Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 133 (2004) 163–174
MOLECULAR & BIOCHEMICAL PARASITOLOGY

An unusual mosaic structure of the PIM gene of *Theileria parva* and its relationship to allelic diversity

Dirk Geysen [a,*], Thomas Bazarusanga [b], Jozef Brandt [a], Thomas T. Dolan [c]

[a] *Department of Animal Health, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerp, Belgium*
[b] *National Veterinary Laboratory, BP 804 Kigali, Rwanda*
[c] *Livestock Services, P.O. Box 24437, 00502 Karen, Nairobi, Kenya*

Received 24 February 2003; received in revised form 2 October 2003; accepted 9 October 2003

Abstract

Genetic diversity and structural organisation of the polymorphic immunodominant molecule (PIM) gene of the protozoan parasite *Theileria parva* was studied in isolates from sympatric and allopatric areas. The analyses revealed a mosaic structure consisting of highly conserved regions shared among some of the isolates from geographically different areas and homologous sequence runs shared among isolates from one area. The specific pattern of diversity in which large insertions and deletions were observed, giving a mosaic structure to the PIM locus, is quite exceptional for single-locus genes. The polymorphic middle region of the gene was characterised by large deletions or insertions in many isolates. There was no correlation between the copy number of the tetrapeptide repeats in this region and the total length of the sequence. The gene was highly polymorphic when compared with sequences from other known *T. parva* antigenic regions.

The findings support the concept that as yet unidentified mechanisms are generating extensive diversity and shaping the PIM locus. The relevance of this finding for diagnosis and the relationship between these mechanisms and the possible role of this protein in host immune responses is discussed.

Keywords: Theileria; Polymorphism; Immunodominant gene; Recombination

1. Introduction

East Coast fever (ECF) is a tick-transmitted protozoan disease caused by *Theileria parva*. It affects cattle in most of the eastern and southern countries of Africa, causing serious mortality and economic losses estimated to cost $169 million annually [1]. African Cape Buffalo (*Syncerus caffer*) are natural reservoirs of the parasite, harbouring antigenetic variants that are infectious to cattle and cause a special form of theileriosis called Corridor disease. The main differences between Corridor disease and classical cattle transmitted ECF are a high mortality and a relatively low number of intralymphocytic schizonts and intraerythrocytic piroplasms stages. The low number of piroplasms and rapid death are considered to be the main factors in limiting transmission of this form of the disease among cattle [2]. It is thought that the buffalo parasite adapts to the cattle host by a transformation process characterised by a gradual increase in the parasitosis and parasiteamia during passages [2,3]. It is generally accepted that this might happen occasionally under optimal transmission conditions and could undermine vaccination programmes [4,5]. Recovered animals have a solid cellular immunity against challenge with homologous isolates that may last for several years [6–8]. Schizont-infected lymphocytes can be established as long term cell lines following isolation from infected animals or by in vitro infection with mature sporozoites from ticks, provided the cells remain infected. Sexual reproduction of *T. parva* occurs in the tick vector [9,10] and this may give rise to recombination between different genotypes [11]. But the frequency of these events is not known. It is thought that one kinete infects one acinar cell in the tick salivary gland [12,13] and, in the field, only one acinar cell is usually found to be infected. These observations suggest that most cattle infections are derived from genetically identical sporozoites [11].

In vitro characterisation of *T. parva* isolates is based mainly on the polymorphic immunodominant molecule (PIM) that has been shown to be abundant and localised on the surface of the schizont stage [14]. The PIM molecule has been extensively characterised [15,16] and is utilised in recombinant form for diagnosis [17]. Serological char-

* Corresponding author. Tel.: +32-3-216-1431; fax: +32-3-247-6264.
E-mail address: dgeysen@itg.be (D. Geysen).

0166-6851/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2003.10.004

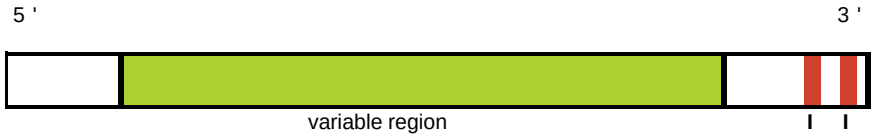

Fig. 1. General structure of the PIM gene with the central variable region and I (dark area) are the position of the introns.

acterisation is based on the use of a battery of mAbs raised against PIM that has revealed extensive diversity in buffalo-derived isolates and a more limited diversity in cattle-derived parasites [18–21].

The PIM molecule is encoded by a single copy gene and a limited sequence comparison has identified a central variable region, flanked by conserved 5′ and 3′ termini (Fig. 1), a common structure seen in some other important antigen genes of protozoa [22]. Sequence data has shown features typical of an intrinsic membrane protein [16].

There is strong evidence that the PIM protein is secreted by the microspheres [23] and microsphere-associated proteins of related parasites have been shown to be strongly immunogenic [24–27]. Sequence comparison of two other microsphere-associated antigens from several *T. parva* isolates revealed a dimorphism in cattle-derived isolates [23] that has been supported by RFLP data from field isolates [28]. This contrasts with the striking polymorphism of the PIM molecule that has been used to subdivide field populations of *T. parva* into groups correlating with the concept of 'strain' or immunogenic type [28].

The existence of such polymorphism in both sequence and length suggests that PIM is an important surface antigen of *T. parva*. Establishing the underlying mechanisms generating this polymorphism is important in understanding the host–parasite interactions and its influence on parasite population structure [29] that might have implications for the use of this antigen in diagnostic or immunogenic formulations. In this study of aspects of the diversity among *T. parva* isolates, several Zambian and Rwandian isolates were sequenced and compared with the published sequence data from one Zimbabwean Boleni (R. Bishop, ILRI), several Kenyan isolates, Muguga ([15], GenBank accession no. L06323), Marikebuni [22] and a buffalo-derived isolate Lawrencei 7014 [30].

2. Materials and methods

2.1. Parasite isolates

The isolates and stocks used in this study are summarised in Table 1. Parasite isolates were obtained from Rwanda and Zambia, countries with different *T. parva* epidemiologies. Sequences were derived from seven cell culture isolates of *T. parva* from the Eastern and Southern Provinces of Zambia. Three *T. parva* sequences were obtained from field bloodspot samples from a survey to characterise parasites in Rwanda [31]. The criterion for inclusion of parasites in this study was the difference in PCR-RFLP profiles of the

Table 1
Theileria parva isolates used in the study

Theileria parva	Origin	Time	Form	Identity and passages	
Zam3[a], Zam5	Zambia, Southern Province	May/August 1982–1983	Cell culture isolates, obtained from ILRI	2nd passage	Dr. F. Musisi (FAO Project, Lusaka)
Zam22, Zam23	Zambia, Southern Province	May/August 1985	Cell culture isolates, obtained from ILRI		Dr. F. Musisi (FAO Project, Lusaka)
Katete[b] and Genda	Zambia, Eastern Province	May/June 1983	Sporozoites stabilates	Katete and Gende (2nd passage) and Katete4 (5th passage)	Dr. D Geysen (Andico project, Chipata)
Muguga	Kenya		Published sequence	GenBank accession no. L06323	Baylis et al. (1993)
Marikebuni.	Kenya		Published sequence		Toye et al. (1995)
Boleni[c]	Zimbabwe		Sequence		R. Bischop, ILRI
Rwa R149	Rwanda, Kibungo	November 1998	Filter paper blood spot sample	Field sample	Dr. T. Bazarusanga
Rwa R76	Rwanda, Mutara	November 1998	Filter paper blood spot sample	Field sample	Dr. T. Bazarusanga
Rwa R225	Rwanda, Cyangugu	November 1998	Filter paper blood spot sample	Field sample	Dr. T. Bazarusanga
T. lawrencei	Kenya		Published sequence		Toye et al. (1995)

[a] *T. parva* Chitongo, used as vaccine in Southern Province.
[b] Used as vaccine in Eastern Province.
[c] Presently used as the vaccine stabilate in Zimbabwe.

various PIM amplification products, except in the case of the Zam22, Zam23, Zam3 group and the KateteB2-Genda group that had the same profiles. These were compared with the published *T. parva* sequences from Kenya and one unpublished sequence from Zimbabwe.

2.2. DNA extraction, amplification, cloning, sequencing and analysis

Genomic DNA was extracted from schizont-infected lymphocyte cultures as described by Conrad et al. [32] and from filter paper blood samples as described by de Almeida et al. [33]. PCR amplification was carried out in a total volume of 25 μl. Each reaction contained 50 mM KCl, 10 mM Tris–HCl (pH 8.3), 1.5 mM $MgCl_2$, 200 μM of each dNTP, 20 pmol of each primer, 0.5 U Taq polymerase enzyme and 5 μl of a DNA solution as template (at a concentration of 10 ng μl^{-1} in the case of cell culture extracts). Each mixture was overlaid with 50 μl of fine neutral mineral oil and placed on a heating block set at 90 °C for PCR amplification. The amplification programme was as follows: Step 1: 94 °C for 4 min; Step 2: 94 °C for 1 min; Step 3: annealing temperature of 62 °C for 2 min; Step 4: extension temperature of 72 °C for 2 min; Steps 2–4 were repeated for 39 times. Step 5 was a final extension phase at 72 °C for 8 min and standard detection with ethidium bromide was used after electrophoresis of the amplified samples. Primers used annealed in the conserved end regions: primers PIMF (ATTCCACTGGTTCTTCCGATGT) or PIMnF (AACACAAGTTGATACTGAAT) and as reverse PIMR (CAACCGTGGAATGGCGTATGTT). Cloning was done following purification of the PCR products using a commercial DNA clean up kit before ligation into a plasmid vector (usually pUC18 or TA pCR®2.1). Competent cells were transformed, cultured and glycerinated stabilates were sent for sequencing using the dideoxy chain termination method during a single run starting at both ends for about 500 bp.

Clones for sequencing of the Zambian isolates were selected using size difference as the criterion. Most isolates gave rise to identical-sized PIM amplification products. On more than one occasion, different sized clones were obtained from identical-sized amplicons. The sequences of Zam5 and the short Katete PIM alleles were obtained from two different transformed colonies after the TOPO cloning reaction. Three more clones from different Katete passages giving the short PIM form were obtained using the TA pCR®2.1 vector. Clones of the three Rwandian field samples were obtained through amplification of extracts from filter paper blood samples, collected in the field in Rwanda in 1998 and cloned into a TA pCR®2.1 vector and handled as described for the Zambian samples.

Seven sequences were obtained from Zambian isolates: Katete and Genda stocks from the Eastern Province, and Chitongo, Zam22, Zam23, and Zam5 isolates from the Southern Province, and three from Rwandian parasites. Sequences were aligned using the GeneJockeyII software corrected by inserting gaps so that identical amino acid motifs were in alignment and finalised by eye using SeqVu 1.0.1. software. Nucleotide and deduced amino acid sequences were analysed using GeneJockeyII software and diversity analysis was done using MEGA version 2.0 [34] and DnaSP version 3.50 program [35].

The two software programs were used to identify and quantify the polymorphism of the conserved 5′ and 3′ regions (5′CR and 3′CR) of the different sequences. The central variable region was too polymorphic in length for meaningful analysis with these programs due to the presence of large deletions in some sequences and the inclusion of numerous gaps for correct alignment. The MEGA2 program was used to quantify the diversity of the 5′CR and 3′CR regions within and between the Zambian isolates, the two Kenyan isolates and the buffalo-related isolates by using the complete deletion option for handling gaps and missing data.

The DnaSP program was used to estimate the synonymous against the non-synonymous replacements in the 5′CR and 3′CR regions, the minimum numbers of recombination events R_m and linkage disequilibrium within and between 5′CR and 3′CR regions.

3. Results

3.1. Description of diversity

3.1.1. Sequence analysis

Analysis of the overal gene structure of PIM was performed by comparing sequences of PIM products from various stocks, spanning almost the entire PIM gene. The open reading frame (ORF) of the PIM gene in *T. parva* Muguga has been reported to be 1445 bp long [15]. The products amplified using the nested PIM primers spanned a region between 159 and 1099 bp of the *T. parva* Muguga sequence, including the first intron of 55 bp near the 3′ end. The conserved areas including the signal sequence (57 bp) and membrane spanning region (117 bp) fell outside the amplified region. The size of the amplified product could vary from 616 to 1033 bp. Short and long sequences could coexist within a stock as in the Chitongo and Katete parent stocks but in all instances their presence was only revealed after cloning. The relationship between large and short sized sequences was determined by analysis of the sequence data. They were found to be identical, apart from the deletion of a large segment. Several clones and tick passages (up to four passages) of the Katete stock were sequenced and the deletion in the short PIM sequences was always at the same location but differed from the Chitongo short sequence (results not shown). Moreover, from the second tick passage onwards, only the short sequence was present in the Katete stock. This was also used as a vaccine and the short PIM sequence was found in all the *T. parva* positive field samples ($N = 245$, 6% positives) collected 10

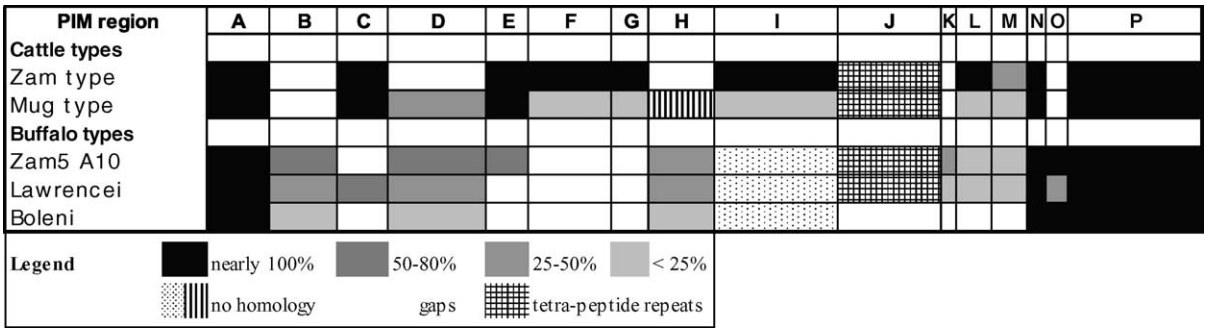

Fig. 2. Comparison of polymorphism in blocks of *T. parva*, *T. parva* buffalo-type and buffalo-derived PIM sequences. Darker regions are more conserved, gaps are indicated as blanks.

years after the vaccinations started [28]. The short sizes of the Marikebuni and Boleni PIM sequences were also due to a deletion of a large region. The exact locations where these deletions occurred differed but they affected the same zone in the central part of the gene, outside the repeated sequences. The predicted amino acid sequence alignment of the different PIM sequences is shown in Fig. 4.

Amino acid homology among the conserved 5′ and 3′ regions (totalling 435 sites) was 73% whereas the central variable region (totalling 1050 sites) showed a homology of only 15%. Glutamine, glycine and proline residues were the major amino acids of the central variable regions as calculated (results not shown). It is striking that the totality of their percentages per isolate are so well conserved (range: 64–76%) and this is most likely related to the high similarity in codons. Glutamine is coded by CAA/CAG; proline by CCA/CCG/CCC and glycine by GGA and GGG.

3.1.2. Mosaic pattern

The multiple sequence alignment identified a composite structure with regions of conserved (nearly 100% homology) and semi-conserved (50–80% homology) sequences alternating with polymorphic stretches (less than 25% homology). The sequences have been divided into 16 blocks to facilitate their description (Fig. 2 and for more details, see Fig. 3). Blocks A and P were extremely well conserved apart from a few point mutations in both blocks and one codon deletion in block A. Blocks B, K and O were specific for *T. parva* buffalo-type and buffalo-derived *T. parva* isolates, whereas all the other blocks showed large allelic polymorphisms, although blocks C, E, F, G, M and N were well conserved among all the bovine isolates. Block J contained tandemly repeated sequences present in the majority of stocks but absent in the short Katete4 sequence or only as a dual repeat in the Boleni isolate (Fig. 3).

Sequence relationships could be identified for most blocks of the different PIM genes to an extent that all blocks could be made up by sequences found in the PIM products that were analysed (Fig. 4). Using these relationships, it was possible to achieve correct alignments between gene sequences by inclusion of gaps. An obvious characteristic of this gene was the presence of duplications of large sequence regions in several parasites, with no geographical correlation. The following points summarise the main features of the multiple PIM sequence comparison:

1. Part or complete blocks C, E and G sequences (QDQPDHQQPTQGD **TSGQQ GQQPQDQPVQE-QDGQDSQGTP** EQTPDQSGQQPVQQPSGQQQ) are duplicated in Southern Zambian (block F), Rwandian (block F) and KateteB2 (block I) parasites, but not in the other Eastern Province parasites with the short sized PIM sequence. Block G sequence (DQSGQQPVQQPS-GQQQQ) is partly duplicated in Muguga (block I) and fully in Katete (block J). Block I of Rwanda R225 is identical to block I of Muguga.
2. A large region of sequence homology (GP(V/L)EPVD-QQQ(Q)PT) is shared between Muguga, Rwanda and the buffalo-related sequences in block D and as a duplication in Zam5 and *T. parva* buffalo-type parasites in block I.
3. Eastern Zambian isolates and Rwandian parasites share a large sequence (VQEQTPETPA(P/Q)Q EQPQTPDDQT-PEQKPDQP) with Marikebuni in the L–M blocks. Southern Zambian isolates show a large deletion of 10 AA at the start of the M block. This is the only difference found between the Marikebuni and the short Southern province sequence Zam3C1.
4. *Theileria parva* Muguga and the Rwandian parasites are the only cattle sequences that share a nona-peptide with *T. parva* buffalo-type and Zam5. This sequence PVD(Q/H)QQ(Q)P(V/T) is imperfectly repeated five times (*T. parva* buffalo-type), four times (Zam5) and one time (*T. parva* Boleni) in the buffalo-type sequences.
5. Some specific sequence blocks are shared among *T. parva* buffalo-type and *T. parva* Boleni and Zam5. These are unique for the three stocks and can be used to differentiate them from the remaining sequences as buffalo-type sequences.
6. Blocks A and P are remarkably conserved in cattle stocks whereas they contain variable but related sequences and point mutations in the buffalo-type sequences.

Genetic polymorphism was estimated using the computer programme MEGA2 with setting of the complete deletion option for handling gaps and missing data. The conserved

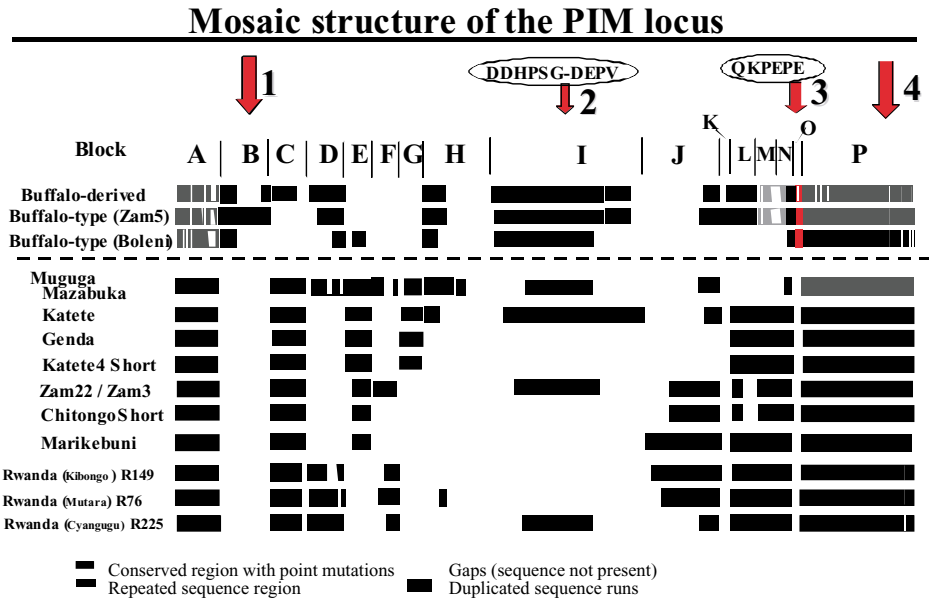

Fig. 3. Schematic representation of PIM alleles divided in blocks. Full coloured blocks indicate unique sequences, patterns indicate partly repeated sequences, gradation in darkness indicate percentage (black close to 100%) of homology. Squares indicate tandemly repeated sequence area. Small white lines in conserved regions indicate occurrence of non-synonymous mutations. The three top sequences are buffalo-derived while the rest are cattle-derived sequences. Numbers 1–4 indicate differences between *T. parva* and buffalo-derived or buffalo-type *T. parva* sequences.

areas and the variable central regions of the different PIM sequences were compared separately.

Due to the high variability and mosaic structure of the PIM locus, only the 5′ and 3′ conserved regions, a minor part of the sequence (blocks A, L, M, N, O and P), gave a workable alignment and could be used for polymorphism analysis. Alignment of the central polymorphic region sequences was not possible without the inclusion of large gaps, making it impossible for comparison due to the multiplicity of alignment possibilities. The high polymorphism in this region together with the multiplicity of gaps made quantification of the central polymorphic regions impossible as not enough sites (only 18%) could be used in the comparison.

3.1.3. Quantification of diversity

Alignment gaps were removed before analysis of the 5′ and 3′ conserved regions and nucleotide and synonymous versus non-synonymous methods were used to calculate the divergence as a proportion (p-distance) between the different sequences and geographical groups.

Table 2
Characterisation of sequence polymorphism

	3	8	1 3	2 0	2 1	2 7	2 8	2 9	3 3	4 9	5 0	5 8	5 9	6 0	6 1	7 2	7 6	7 7	8 4	9 7	1 3 0	1 3 5	1 4 2	1 4 4	1 5 1	1 5 2	1 6 0	1 6 1	1 6 9	1 7 3	1 7 6	2 4 7	2 7 6	2 8 7	2 9 5	3 0 0	3 0 2	3 0 8	3 1 6	3 1 7	3 3 5	3 3 6	3 4 4	4 3 5
	*						*	*					*		*	*			*												*		*			*	*	*		*	*	*	
KateteB2	A	T	A	C	T	G	A	C	A	T	C	G	G	A	A	T	G	A	A	C	A	A	A	A	A	A	G	C	G	T	C	A	T	G	A	C	G	A	G	A	T	A	G	T
GendaA8						T						T																												G				
Katete4 A15						T																																						
Zam3 B30						T																																					
Zam3 C1						T						G						T																									
Zam22 C22						T																																					
Katete4 C25						T																																					
Zam23 B26						T																																					
Marikebuni						T																			C	C			A														
Muguga						T																																					
Zam5 A10	G					T	G	A		G	G		C		C	A			T													C											
Lawrencei						T	G	A				C	C		C	A			T	A			T	T			C	T		A	T	C	A	A			T		C		G	G		
Boleni		C	T			T		G	G					C	C	A	A	T	T													C		A	T		T	G	C		C	G	A	
Rwa R146																																						G					A	
Rwa R76				G	G																																	G						
Rwa R225																																						G						A

Synonymous * Sites with parsimoneous information

Comparison and characterisation of the polymorphic sites in the 5′ and 3′ end conserved PIM regions. Base pair positions are presented (using MEGA2 program) where polymorphism is found between the *T. parva* sequences. These are subdivided in sites with non-synonymous AA substitutions and sites with parsimony information (sites that segregate for only two nucleotides that are present at least twice).

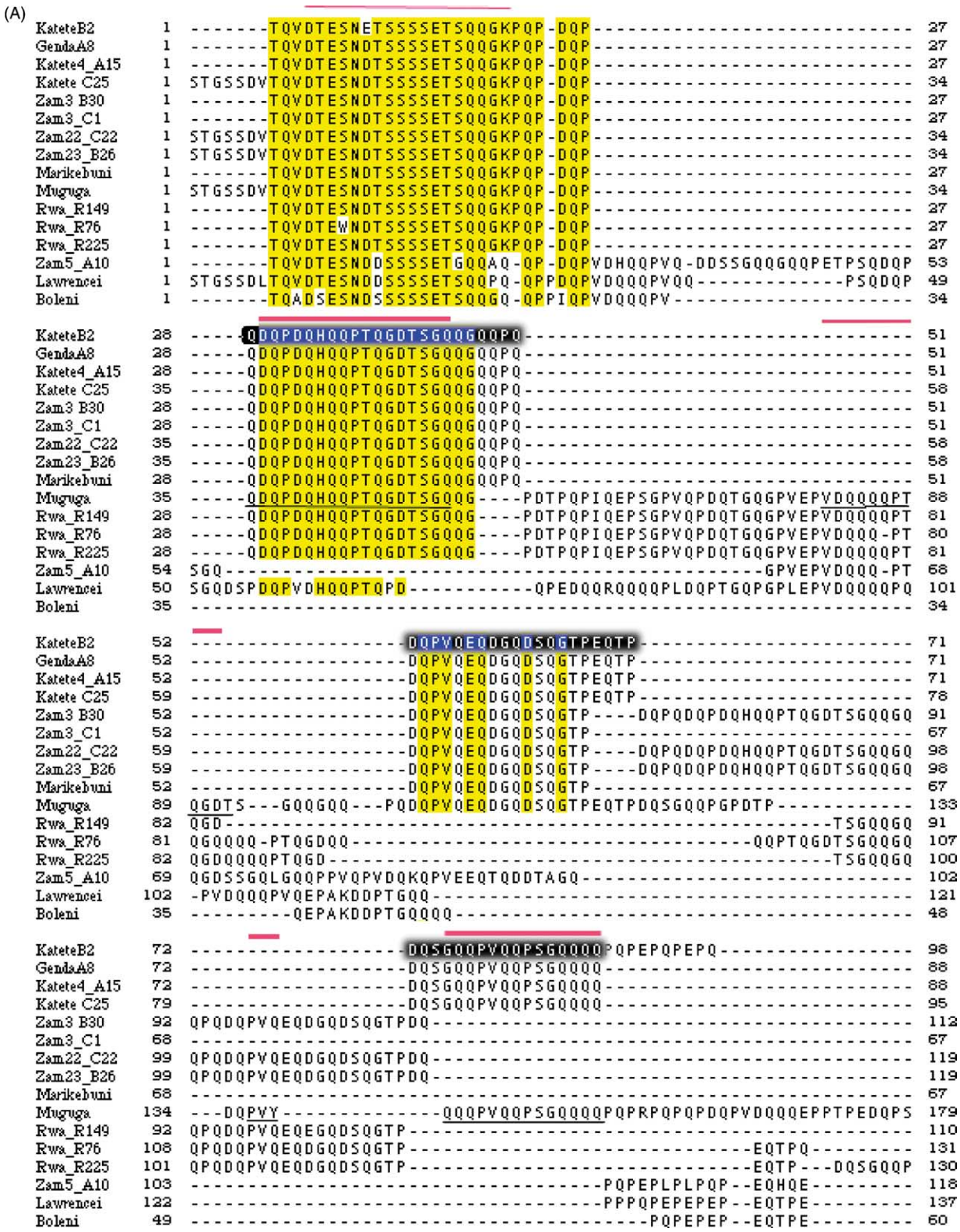

Fig. 4. Amino acid alignment maximising similarity of PIM sequences from various isolates. Shaded areas indicate homology among majority of sequences, dashes indicate missing amino acids, X is an undefined amino acid, due to unresolved sequencing data. Sequence runs which are affected in duplication events are shown in negative. mAbs epitopes are underlined.

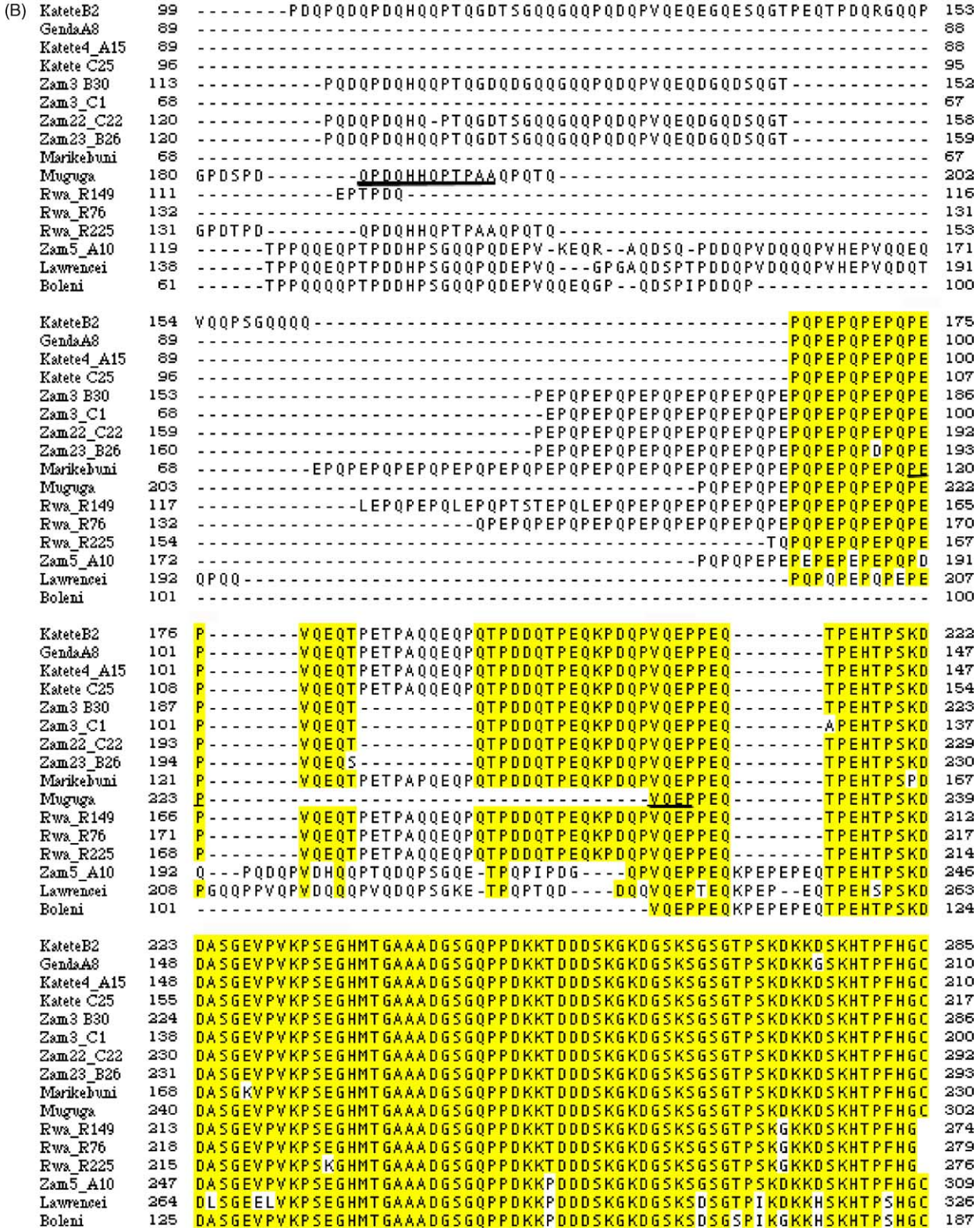

Fig. 4. (*Continued*).

The nucleotide based p-distance (pairwise distances) data calculated on the conserved regions using 314 sites (70% of the total sites) indicated that the Lawrencei isolate had the most outstanding sequence ($d = 0.07$) followed by Boleni ($d = 0.06$) and Zam5 ($d = 0.035$). The differences among the cattle isolates were in the range of $d = 0.01$ (results not shown).

3.1.4. Origin of diversity

The origin of the PIM diversity was further investigated by comparing the synonymous against non-synonymous substitutions (see Table 2). When sequences were grouped as cattle-derived and buffalo-derived, the synonymous (S) and non-synonymous (Ns) substitutions were 5 and 25 Ns changes against 2 and 9 S nucleotide changes giving a ratio Ns/S of 2.5 in cattle and 2.78 in buffalo-type PIM sequences for the conserved 5′ and 3′ end regions.

The degree of linkage disequilibrium (LD) among the 5′ and 3′ end regions of the 16 sequence data was tested by estimating D using the DnaSP program and plotting D against the nucleotide distance. Both the two-tailed Fisher's exact test and the Chi-square test were computed to determine the significance of the associations found [36]. This analysis gives an estimation of the non-random association between variants of different polymorphic sites. No decline in LD was found with increasing nucleotide distance between the 5′ and 3′ end regions of the sequences tested (results not shown).

4. Repeat sequence analysis

An area of PQ-rich, tandemly repeated sequences was present in most stocks towards the 3′ end. The repeats consisted of the tetrapeptide $(PQPE)_x$ in bovine stocks and $(PEPE)_x$ or $(P(Q/E/P)PQ)_x$ in buffalo stocks, and a mixture in buffalo-derived stocks (Table 3). The number of tetrapeptide repeats varied between 2 and 13 but there was no correlation between copy number and total length of the sequence. The 690 bp PIM sequence found in *T. parva* Marikebuni had the highest number of repeats whereas the number of tetrapeptide repeats of the 855 bp long sequence of *T. parva* Katete stock and the short 561 bp of Boleni was very low. Specific buffalo-related regions, highlighted in Fig. 3 as numbers 1–4, could be identified that were absent from the cattle sequences. These regions showed size and sequence polymorphism among the buffalo-related isolates.

The tandem repeats among *T. parva* strains are identical in sequence (PQPE) while the buffalo-derived *T. parva* has a mixture of two different repeats (PEPE and P(Q/E/P)P(E/Q/D)).

Other motifs were repeated one to five times throughout the PIM gene of all stocks (S((X)/G)QQ(G/P)) or a limited number of stocks (VDHQQ(P/Q)).

5. PIM epitopes

The PIM epitopes recognised by mAbs 2–7, IL-S32.3 and IL-S34.3 have been determined [15] and the sequence data correlated well with the mAbs results, and with the predicted presence of the epitopes (Table 4). Although the sequence of Zam5 differed to a large extent from the other Zambian isolates, its mAbs profile was identical. This could be explained by the presence of a very similar peptide, differing only in the middle amino acid (DQPVDHQ) position. The location of this difference would probably not restrict antibody interaction and therefore should not affect its reactivity with mAb7. The RFLP-PCR profiles show that Zam5 differs markedly from the other Zambian isolates [37].

6. Discussion

This study provides a detailed sequence analysis of the PIM gene of *T. parva*, including the entire variable regions, and a comparison of the sequence among a large number of isolates from different epidemiological environments. The conserved regions at both ends of the gene and the high proline (P) or glutamine (Q) content of the sequence described

Table 3
Amino acid repeats in PIM sequences

Stocks	Amino repeat	Base pair	No. of repeats
B2/Katete	PQPE	CTCCAACCTGAA	3//3*
A15/Katete4 S	PQPE	CTCCAACCTGAA	3
B30/Chitongo L	PQPE	CTCCAACCTGAA	*8*
C1/Chitongo S	PQPE	CTCCAACCTGAA	*8*
C22/Zam22	PQPE	CTCCAACCTGAA	8
B26/Zam23	PQPE	CTCCAACCTGAA	8
Marikebuni	PQPE	CTCCAACCTGAA	13
Muguga	PQPE	CTCCAACCTGAA	5
Lawrencei	PQPE–PEPE–P(E/P)PQ	TGAACCTGGA/CCTGAACCTGAA/CCA(CCA/CAA)CCTCAA	3–3–2
A10/Zam5	PQPE//PEPE + PQP(Q/D)	CCTCAACCTGAA//CCTGAACCTGAA//GAT	2//5 + 2*
Boleni	PQPE–PEPE	CCTGAACCTGAA	1 + 2

Main amino acid repeats and their length in PIM sequences from different *T. parva* stocks. S: short sequence; L: long sequence; The symbol (*) indicates 1 amino difference at start and/or end of repeats and // represents not adjoined.

Table 4
PIM sequences of mAbs epitopes

mAb origin	mAbs	Peptide epitopes	Peptide epitopes	Peptide epitopes	Present in		Comments
					Zams	Mug	
Muguga	2	HQPTPAA			0	3	
Marikebuni	7	DQPDQHQ	PVYQQQPVQQPS		1	3	All stocks except buffalo-type *T. parva*
Marikebuni	15				0	0	
Marikebuni	IL-S32.3	**ASGEVPVKPSEG**			0	0	All stocks except buffalo-type *T. parva*
	lL-S34.3	DQPDQHQ	PVYQQQP	QPDGHHQ	0	1	

PIM sequences of epitopes recognised by the different mAbs, some recognising more than one epitope. mAb origin indicates isolate used for the production of mAbs.

by Baylis et al. [16] and Toye et al. [22] were characteristic for all the sequences analysed. The specific pattern of diversity in which large insertions and deletions were observed, giving a mosaic structure to the PIM locus, is quite exceptional for single-locus genes and is reported here for the first time. Most mosaic genes described display small [38,39] to moderate size polymorphism [40] often related to variation in the number of repeated sequences.

Analysis of genetic polymorphism in the 5′ and 3′ end regions showed high homogeneity within the cattle isolates (less than 13% difference). The average difference (measured as pairwise distances) among buffalo-type sequences was 60% and between cattle and buffalo-derived isolates was in the order of 55%.

7. General structure

The structure of all the PIM genes analysed consisted of conserved termini flanking a central variable region and is in agreement with earlier observations [22].

There was a large area of short repeated sequences in the central regions of most bovine *T. parva* isolates that is absent from *T. parva* Boleni. A second and different type of polymorphism was found in the bovine *T. parva* sequences in which long sequence runs were duplicated or deleted and not showing any geographical correlation. This gave rise to a different RFLP profile. The identical size and location of deletion found in various independent PCR and cloning tests refute the possibility that these are due to PCR artefacts. Moreover, the profile corresponding with the shorter sequence was found as the dominant profile in bovine field samples collected in the Eastern Province of Zambia, 10 years after the start of mass vaccination using a tick derived sporozoite stabilate from a *T. parva* isolate showing the short PIM sequence [28]. The fact that three sequences revealed large deletions that could be related to the loss of a large continuous duplication, as in the case of the Zambian stocks, is remarkable and provides strong evidence of frequent sequence specific recombination events.

Size polymorphism in sequences coding for membrane proteins is usually related to antigenic variation [50,51]. Furthermore, there are indications that repeated sequences are implicated in recombination or the result of conversion events (see below). It would not be unusual for the sequences alongside such repeats to be regions of high variability that are correlated with epitope polymorphism.

8. Tetrapeptide repeats

The presence of variable numbers of repeats is often correlated with the presence of slipped strand mispairing during intra or interhelical exchange of DNA. The differences in repeat numbers in the PIM gene could be explained by such a mechanism as many sequences show large deletions adjacent to these repeats.

The results of the analysis of the *Theileria* repeats in the PIM gene are in support of the common ancestry hypothesis [41]. The presence of sequence homogeneities among bovine *T. parva* isolates from different geographic areas together with a history of possible buffalo-related origin are a strong indication of their common origin. Although more sequences from buffalo need to be compared, it is tempting to associate this with phylogenetic evidence for the long term divergence of *T. parva* buffalo and cattle-type *T. parva* parasites, and the intermediate status of isolates such as *T. parva* Boleni and Zam5.

9. Mosaic blocks

The analyses of PIM gene structures from different isolates revealed a general design characterised by multiple domains that are highly conserved and are shared among several isolates. Other blocks appear either to be common among stocks from the same geographical region or to be stock-specific. This gives rise to a mosaic appearance of the gene due to the alternation of shared blocks, deletions and specific sequences. The shared sequences might have resulted from common ancestry [42], convergent evolution [44], recombination or gene conversion [43].

The last two mechanisms could be incriminated in the generation of the mosaic appearance of the PIM gene as sufficient regions of homology are present to initiate intragenic recombination events. On the other hand, several X-like recombination motifs (Table 5) could be identified

Table 5
X-like recombination motifs

X-like motifs in *T. parva*				
GCTGGGG	GGGTGATTC	GTGGGG	CAGGGGGAT	CCCGTAG
GCTGGTG	GGGTG	GATGGGC	CAGGGTGAT	CCTGTAT
CCTGTGG	GTGGAC	CTCCGG	GCCTCGGCCTC	
GGTCAGG		GATGGTC	GACTCCTGATG	
X-motifs from literature				
GCTGGTGG[a]				
GGAGGTAGGCAGGCAG[b]				
GTGACTGGCCAGGAGG[c]				
GGGCAGGAXG[d]				
CXTCCTGCCC[e]				

X-like recombination motifs found in all PIM sequences.
[a] Motif in *Escherichia coli* (see [43]).
[b] Motif in mouse MHC gene (see [44]).
[c] Motif in hamster aprt locus (see [44]).
[d] Motif in human minisatellites (see [44]).
[e] Motif in ade8 gene of *Saccharomyces cerevisiae* (see [45]).

and might be responsible for initiation of gene conversion events [45,46].

Frequent inter or intra-allelic recombination would lead to a decline in linkage disequilibrium parameters with increasing gaps between polymorphic gene regions. In the case of the PIM sequences, no significant decline was found. This excludes recombination as the main mechanism although some polymorphism could have been generated through recombination events [52]. Gene conversion seems a more likely explanation for the polymorphism than intragenic recombination as the latter would imply an implausible frequency of recombination. Gene conversion and reciprocal intergenic exchanges have been identified as mechanisms responsible for the re-shuffling of important epitopes thereby creating novel alleles [47].

It is tempting to see the different PIM forms as variants originating from two basic allelic families and characterised by block D (Muguga type) and block F (Marikebuni/Zambian type sequence) sequences. This is in agreement with the dimorphism found in other cattle-type *T. parva* sequences [28].

10. Characterisation of difference

Results of the analysis of divergence using the conserved 5′ and 3′ ends (Table 2) are in agreement with the various reports showing large diversity in buffalo-type parasites compared to the cattle-derived parasites [19,48]. A second important finding is the relationship among buffalo-derived *T. parva* and *T. parva* buffalo-type sequences. *T. parva* Boleni and Zam5 display the highest sequence homology with *T. parva* buffalo-type as can be seen in Fig. 2.

The differential diversity characterising buffalo and non-buffalo-derived *T. parva* could be exploited for typing purposes. A probe could be designed using either the SNDTSS sequence in the 3′ block A for *T. parva* or the DKPDDD sequence of the O block for buffalo-derived *T. parva*. Another region would be the buffalo-derived specific PVD(H/Q)QQPV sequence of block B. These data support the notion that recombination between buffalo and cattle parasites occurs as data on the ITS region in buffalo parasites have shown [49]. Recombination among buffalo parasites and between buffalo and cattle parasites is occurring in those areas where cattle and buffalo share grazing. These are characterised by the occurrence of Corridor disease but it is unlikely that these recombinants become established in the cattle population. Zam5 was isolated from an area of known buffalo presence but a large scale characterisation survey conducted 10 years later did not reveal the presence of this or a related genotype [28]. Furthermore, no cattle infection was induced when between 10^9 and 10^{10} schizont-infected cells were inoculated on five different occasions (D. Geysen, personal communication) whereas infection could be established using the same technique with cattle-derived parasites.

11. Correlation of polymorphism with selection

The presence of a high Ns/S ratio in the PIM sequences is a possible indication of the influence of selective forces on the sequence and the evasion of host immune responses. Extensive polymorphism in surface proteins of parasites is a strong indication that immune selection pressure is being exerted. It has been shown that PIM is a predominant schizont antigen with reactivity in sera from infected animals as shown on immunoblots [22]. The mosaic nature of the polymorphism found in PIM excludes genetic drift or directional change. This study also confirms the highly conserved nature of the intron sequence for all PIM genes that have been analysed. This is in marked contrast to the polymorphism found in the coding part of the gene and is highly suggestive for selective pressure on the sequences in the coding regions, most likely imposed by the immune response of the host [30].

This study indicates that PIM may be an important antigen driving cellular immunity in ECF. A similar assumption has been reported by Schnittger et al. [39] who reported on the characterisation of a polymorphic *T. annulata* surface protein (TaSP) closely related to PIM. Further research should be directed to obtaining more sequence data from stocks for which good cross-immunity data are available to correlate sequence data with antigenic diversity.

Acknowledgements

This research was made possible through the financial support of the Belgian Administration for Development Co-operation and the collaboration of ILRI. We would like to thank especially Drs. R. Skilton, R. Bishop and S. Morzaria for their support and warm collaboration. We are grateful to R. Beudeker, F. Ceulemans and J. Dewitte at ITM, Antwerp for their technical inputs. We also wish to thank Drs. G. Van Der Auwera, P. Spooner, P. Toye, V. Nene and T. Musoke for their discussions and the anonymous reviewer for his helpful advice.

References

[1] Mukhebi AW, Perry BD, Kruska R. Estimated economics of Theileriosis control in Africa. Prev Vet Med 1992;12:73–85.
[2] Young AS, Branagan D, Brown CGD, Burridge MJ, Cunningham MP, Purnell RE. Preliminary observations on a *Theilerial* species pathogenic to cattle isolated from buffalo (*Syncerus caffer*) in Tanzania. Br Vet J 1973;129:382–9.
[3] Barnett SF, Brocklesby DW. The passage of "*Theileria Lawrencei* (Kenya)" through cattle. Br Vet J 1966;122:396–409.
[4] Lohr KF. A hypothesis on the role of *Theileria lawrencei* in the epizootiology of east coast fever and on its importance in attempting vaccination. In: Wilde JKH, editor. Tick-borne diseases and their vectors. Scotland: University of Edinburgh; 1978.
[5] Radley DE, Young AS, Grootenhuis JG, Cunningham MP, Dolan TT, Morzaria SP. Further studies on the immunization of cattle against *Theileria lawrencei* by infection and chemoprophylaxis. Vet Parasitol 1979;5:117–28.
[6] Morrison WI, Goddeeris BM, Teale AJ, Baldwin CL, Bensaid A, Ellis J. Cell mediated immune responses of cattle to *Theileria parva*. Immunol Today 1986;7:211–6.
[7] Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Karimi IM, Purnell RE, et al. East Coast fever: 1. Chemoprophylactic immunization of cattle against *Theileria parva* (Muguga) and five theilerial strains. Vet Parasitol 1975;1:35–42.
[8] Young AS, Brown CGD, Burridge MJ, Cunningham MP, Karimi IM, Irvin AD. Observations on the cross-immunity between *Theileria lawrencei* (Serengeti) and *Theileria parva* (Muguga) in cattle. Int J Parasitol 1973;3:723–8.
[9] Gonder R. The life cycle of *Theileria parva*: the cause of East Coast fever in cattle in South Africa. J Comp Pathol 1910;23:328–35.
[10] Mehlhorn H, Schein E. The piroplasms: life-cycle and sexual stages. Adv Parasitol 1984;23:37–103.
[11] Morzaria SP, Young JR, Spooner PR, Dolan TT, Young AS, Bishop RP. Evidence of a sexual cycle in *Theileria parva* and characterisation of the recombinants. In: Munderloch, Kurtti, editors. First international conference on tick-borne pathogens at the host–vector interface—an agenda for research: proceedings and abstracts. Saint Paul, MN: University of Minnesota; 1992.
[12] Young AS. The epidemiology of Theileriosis in East Africa. In: Irvin AD, Cunningham MP, Young AS, editors. Advances in the control of Theileriosis. The Hague: Martinus Nijhoff Publishers; 1981.
[13] Young AS, Leitch BL, Newson RL, Cunningham PM. Maintenance of *Theileria parva parva* infections in an endemic area of Kenya. Parasitology 1986;86:519–28.
[14] Shapiro SZ, Fujisaki K, Morzaria SP, Webster P, Fujinaga T, Spooner PR, et al. A life cycle stage specific antigen of *Theileria parva* recognised by anti-macroschizont monoclonal antibodies. Parasitology 1987;94:29–37.
[15] Toye P, Nyanjui J, Goddeeris B, Musoke AJ. Identification of neutralization and diagnostic epitopes on PIM, the polymorphic immunodominant molecule of *Theileria parva*. Infect Immun 1996;64(5):1832–8.
[16] Baylis HA, Allsopp BA, Hall R, Carrington M. Characterisation of a glutamine-rich and proline-rich protein (QP protein) from *Theileria parva*. Mol Biochem Parasitol 1993;61(2):171–8.
[17] Katende J, Morzaria S, Toye P, Skilton R, Nene V, Nkonge C, et al. An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule. Parasitol Res 1998;84(5):408–16.
[18] Conrad PA, Iams K, Brown WC, Sohanpal B, Ole-moiyoi O. DNA probes detect genomic diversity in *Theileria parva* stocks. Mol Biochem Parasitol 1987;25:213–26.
[19] Conrad PA, Ole-Moiyoi OK, Baldwin CL, Dolan TT, O'Callaghan CJ, Njamunggeh REG, et al. Characterisation of buffalo-derived theilerial parasites with monoclonal antibodies and DNA probes. Parasitology 1989;98:179–88.
[20] Minami T, Spooner PR, Irvin AD, Ocama JGR, Dobbelaere DAE, Fujinaga T. Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. Res Vet Sci 1983;35:334–40.
[21] Irvin AD, Dobbelaere DAE, Mwamachi DM, Minami T, Spooner PR, Ocama JGR. Immunisation against ECF: correlation between mAbs profiles of *T. parva* stocks and cross-immunity in vivo. Res Vet Sci 1983;35:341–6.
[22] Toye P, Gobright E, Nyanjui J, Nene V, Bishop R. Structure and sequence variation of the genes encoding the polymorphic immunodominant molecule (PIM), an antigen of *Theileria parva* recognized by inhibitory monoclonal antibodies. Mol Biochem Parasitol 1995;73(1/2):165–77.
[23] Skilton RA, Bishop RP, Wells CW, Spooner PR, Gobright E, Nkonge C, et al. Cloning and characterisation of a 150 kDa microsphere antigen of *Theileria parva* that is immunologically cross-reactive with the polymorphic immunodominant molecule (PIM). Parasitology 1998;117(4):321–30.
[24] Cesbron-Delauw MF. Dense granule organelles of *Toxoplasma gondii*: their role in the host–parasite relationship. Parasitol Today 1994;10:293–6.
[25] Goodger BV, Waltisbuhl DJ, Commins MA, Wright I. *Babesia bovis*: dextran sulphate as an adjuvant for and precipitant of protective immunogens. Int J Parasitol 1992;22:465–9.
[26] Doury JC, Bonnefoy S, Roger N, Dubremetz JF, Mercereau Puijalon O. Analysis of the high molecular weight rhoptry complex of *plasmodium falciparum* using monoclonal antibodies. Parasitology 1994;108(3):269–80.
[27] Ushe TC, Palmer GH, Sotomayor L, Figueroa JV, Buening GM, Perryman LE, et al. Antibody response to a babesia bigemina rhoptry-associated protein 1 surface-exposed and neutralization-sensitive epitope in immune cattle. Infect Immun 1994;62(12):5698–701.
[28] Geysen D. The application of molecular biology techniques to analyse diversity in *T. parva* populations in Zambia. Ph.D. thesis, Brunel University, Uxbridge; 2000.
[29] Gupta S, Day KP. A theoretical framework for the immuno epidemiology of *Plasmodium falciparum* malaria. Parasite Immunol 1994;16:361–70.

[30] Toye PG, Metzelaar MJ, Wijngaard PJL, Nene V, Iams K, Roose J, et al. Characterisation of the gene encoding the polymorphic immunodominant molecule, a neutralizing antigen of *Theileria parva*. J Immunol 1995;155:1370–81.
[31] Bazarusanga T. Molecular characterisation of *T. parva* parasites in the field of Rwanda. M.Sc. thesis, ITM, Antwerp; 1999.
[32] Conrad PA, Stagg DA, Grootenhuis JG, Irvin AD, Newson J, Njamunggeh REG, et al. Isolation of *Theileria parva* parasites from African buffalo (*Syncerus caffer*) and characterisation with antischizont monoclonal antibodies. Parasitology 1987;94:413–23.
[33] de Almeida PJLP, Ndao M, Van Meirvenne N, Geerts S. Diagnostic evaluation of PCR in goats experimentally infected with *Trypanosoma vivax*. Acta Trop 1997;66:45–50.
[34] Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 2001;17(12):1244–5.
[35] Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 1999;15:174–5.
[36] Sokal RR, Rohlf FJ. Biometry. 2nd ed. New York: W.H. Freeman and Company; 1981.
[37] Geysen D, Bishop R, Skilton R, Dolan TT, Morzaria S. Molecular epidemiology of *T. parva* in the field. TMIH 1999;4(9):A21–7.
[38] Gubbels MJ, Katzer F, Hide G, Jongejan F, Shiels BR. Generation of a mosaic pattern of diversity in the major merozoite-piroplasm surface antigen of *Theileria annulata*. Mol Biochem Parasitol 2000;110(1):23–32.
[39] Schnittger L, Katzer F, Biermann R, Shayan P, Boguslawski K, McKellar S, Beyer D, Shiels BR, Ahmed JS. Characterization of a polymorphic *Theileria annulata* surface protein (TaSP) closely related to PIM of *Theileria parva*: implications for use in diagnostic tests and subunit vaccines. Mol Biochem Parasitol 2002;120(2):247–56.
[40] Arnot DE, Barnwell JW, Stewart MJ. Does biased gene conversion influence polymorphism in the circumsporozoite protein-encoding gene of *Plasmodium vivax*? Proc Natl Acad Sci USA 1988;85:8102–6.
[41] Frontali C, Pizzi E. Conservation and divergence of repeated structures in *Plasmodium* genomes: the molecular drift. Acta Leid 1991;60(1):69–81.
[42] McDevitt H. Evolution of MHC class II allelic diversity. Immunol Rev 1995;143:113–22.
[43] Parham P, Adams EJ, Arnett KL. The origins of HLA-A, B, C polymorphism. Immunol Rev 1995;143:141–80.
[44] Klein J, O'hUigin C. Class II B MHC motifs in an evolutionary perspective. Immunol Rev 1995;143:89–111.
[45] Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL. Structure of Chi hotspots of generalized recombination. Cell 1981;24:429–36.
[46] Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. Nature 1985;316:76–9.
[47] Dormoy A, Reviron DV, Froelich N, Weiller PJ, Mercier PJ, Tongio MM. Birth of a new allele in a sibling: cis or trans gene conversion during meiosis? Immunogenetics 1997;46:520–3.
[48] Bishop RP, Spooner PR, Kanhai GK, Kiarie J, Latif AA, Hove T, et al. Molecular characterization of Theileria parasites: application to the epidemiology of theileriosis in Zimbabwe. Parasitology 1994;109(5):573–81.
[49] Collins NE, Allsopp BA. *Theileria parva* ribosomal internal transcribed spacer sequences exhibit extensive polymorphism and mosaic evolution: application to the characterization of parasites from cattle and buffalo. Parasitology 1999;118(6):541–51.
[50] Volkman SK, Hartl DL, Wirth DF, Nielsen KM, Choi M, Batalov S. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. Science 2002;298(5591):216–8.
[51] Palmer GH. The highest priority: what microbial genomes are telling us about immunity. Vet Immunol Immunopathol 2002;85(1/2):1–8.
[52] White JH, Di Martino JF, Anderson RW, Lusnak K, Hilbert D, Fogel S. A DNA sequence conferring high postmeiotic segregation frequency to heterozygous deletions in *Saccharomyces cerevisiae* is related to sequences associated with eukaryotic recombination hotspots. Mol Cell Biol 1988;8:1253–8.