MOLECULAR AND CELLULAR BIOLOGY, Mar. 1997, p. 1666–1673 Vol. 17, No. 3
0270-7306/97/$04.00+0

Concerted Evolution at a Multicopy Locus in the Protozoan Parasite *Theileria parva*: Extreme Divergence of Potential Protein-Coding Sequences†

RICHARD BISHOP,* ANTHONY MUSOKE, SUBHASH MORZARIA, BALJINDER SOHANPAL, AND ELKE GOBRIGHT

International Livestock Research Institute (ILRI), Nairobi, Kenya

Received 12 July 1996/Returned for modification 4 September 1996/Accepted 20 December 1996

Concerted evolution of multicopy gene families in vertebrates is recognized as an important force in the generation of biological novelty but has not been documented for the multicopy genes of protozoa. A multicopy locus, *Tpr*, which consists of tandemly arrayed open reading frames (ORFs) containing several repeated elements has been described for *Theileria parva*. Herein we show that probes derived from the 5′/N-terminal ends of ORFs in the genomic DNAs of *T. parva* Uganda (1,108 codons) and Boleni (699 codons) hybridized with multicopy sequences in homologous DNA but did not detect similar sequences in the DNA of 14 heterologous *T. parva* stocks and clones. The probe sequences were, however, protein coding according to predictive algorithms and codon usage. The 3′/C-terminal ends of the Uganda and Boleni ORFs exhibited 75% similarity and identity, respectively, to the previously identified *Tpr*1 and *Tpr*2 repetitive elements of *T. parva* Muguga. *Tpr*1-homologous sequences were detected in two additional species of *Theileria*. Eight different *Tpr*1-homologous transcripts were present in piroplasm mRNA from a single *T. parva* Muguga-infected animal. The *Tpr*1 and *Tpr*2 amino acid sequences contained six predicted membrane-associated segments. The ratio of synonymous to nonsynonymous substitutions indicates that *Tpr*1 evolves like protein-encoding DNA. The previously determined nucleotide sequence of the gene encoding the p67 antigen is completely identical in *T. parva* Muguga, Boleni, and Uganda, including the third base in codons. The data suggest that concerted evolution can lead to the radical divergence of coding sequences and that this can be a mechanism for the generation of novel genes.

Organisms contain two major categories of genes, ancient conserved housekeeping genes such as those encoding metabolic enzymes and rRNA and phylogenetically restricted genes with diverse functions specialized for the lifestyles of specific taxa (21). Among the parasitic protozoa, well-known examples of the latter class of genes are multicopy gene families, such as the variant surface glycoproteins of African trypanosomes (reviewed in references 11 and 42) and the recently described *var* genes of *Plasmodium falciparum* (4, 48, 49), which have evolved to allow the expression of peptide diversity in order to evade the immune responses of the vertebrate host. In the generation of the phylogenetically restricted repertoire of genes, the major mechanisms involved are believed to be gene duplication allowing the subsequent divergence of additional copies (41) and shuffling of preexisting exons (50). Other mechanisms include direct recruitment from existing genes as exemplified by the lens crystallins of the eye (53) and overprinting, the translation of an existing sequence in a different reading frame (21).

The sporozoan parasite *Theileria parva*, which causes East Coast fever in cattle, is transmitted by the tick *Rhipicephalus appendiculatus*, which introduces the mammal-infective sporozoite stage of the parasite into the blood of cattle. The parasite has two intracellular stages in the mammalian host, the multinucleate schizont, which transforms bovine lymphocytes, and the piroplasm in erythrocytes, which is infective for the vector. The *Tpr* locus, which is hyperpolymorphic between *T. parva* isolates (1, 5, 10), consists of an array of tandemly arranged open reading frames (ORFs), which are predicted to encode proteins, although many copies are partial in that they lack in-frame ATG codons near their 5′ ends (3). Herein we show that the *Tpr* ORFs are transcribed into heterogeneous mRNAs and contain predicted transmembrane domains characteristic of integral membrane proteins. The 3′/C-terminal ends of the ORFs evolve as if they code for protein(s) which are conserved between *Theileria* species. By contrast, multiple copies of the DNA sequences at the 5′ ends of the *Tpr* units are not detectably similar between different *T. parva* cloned isolates but are still conserved as long ORFs with the codon usage properties of protein-encoding DNA.

Concerted evolution is the process whereby the sequences of multicopy gene families diverge between populations but the individual gene copies within the family are homogenized. It is most studied in the context of the fixation of variant sequences within tandemly arrayed gene families of vertebrates, such as rRNA and histone genes (2, 14; reviewed in reference 13). The multicopy isolate-specific sequences at the 5′ ends of the *T. parva* ORFs constitute a dramatic example of concerted evolution. The data suggest that gene duplication, in conjunction with concerted evolution, can be responsible for the divergence of DNA sequences with protein-coding potential, to the extent that they exhibit no apparent similarity in nucleotide sequence or identity in deduced amino acid sequence.

MATERIALS AND METHODS

Parasite material and DNA preparation. Three *T. parva* stocks, Muguga (8), Boleni (26), and Uganda (30), were used, together with clones derived from the stocks (36). The two *Theileria taurotragi* stocks W575 and Z456 were schizont-

* Corresponding author. Mailing address: International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi, Kenya. Phone: 254-2 630 743. Fax: 254-2 631 399. E-mail: R.BISHOP@CGNET.COM.

† ILRI publication no. 063.

infected lymphocyte cell cultures, the former isolated from an eland and the latter isolated from cattle. The *Theileria annulata* Tova stock has been described previously (51). Purification of piroplasms and preparation of *T. parva* schizont-infected lymphocyte and piroplasm DNA were as described previously (10).

Isolation of genomic *Tpr* DNA sequences. Libraries of sheared *T. parva* Uganda and *T. parva* Boleni DNA fragments were prepared in λgt11 from purified piroplasm DNA by the methods of Young et al. (55). The libraries were screened with radiolabelled *T. parva* total genomic DNA as described elsewhere (10). The DNA inserts from *T. parva* Boleni (p*Tpr*Bol, 2,275 bp) and *T. parva* Uganda (p*Tpr*UgB, 2,370 bp) clones which hybridized strongly to homologous total DNA were subcloned into the *Eco*RI site of pUC19. A second *T. parva* Uganda *Tpr* clone (p*Tpr*UgA) was isolated by PCR amplification from genomic DNA with primers ILO 1099 and ILO 1322 (see Table 1 and Fig. 1). The 1,504-bp PCR product was cloned into the *Eco*RV site of pBluescript (Stratagene) by the T-vector procedure (27). Standard genetic manipulation procedures were as described in the work of Sambrook et al. (44).

***Tpr* DNA probes.** The *T. parva* Muguga probe located at the 3′ end of the *Tpr* ORF was a 623-bp genomic DNA sequence cloned in pUC8, the nucleotide sequence of which has been determined previously (1). The *T. parva* Uganda and Boleni 3′-end probes were 280-bp sequences which were PCR amplified from genomic DNA and cloned into *Sma*I-cut pUC19 (5). The *T. parva* Muguga probe containing sequences located at the 5′ end of the *Tpr* ORF was H1477, a PCR product corresponding to *Tpr*3 (3). Probes containing the 5′-end sequences of *T. parva* Boleni and Uganda *Tpr* ORFs were generated by taking advantage of unique *Nde*I sites in both the cloned *Tpr* sequences and pUC19 to delete the 3′ sequences. The extent of the probes with respect to the ORFs is indicated in Fig. 1.

RNA preparation and isolation of cDNA clones. For PCR amplification, piroplasm, schizont-infected lymphocyte, and sporozoite-infected tick salivary gland RNA was prepared by the acid phenol method (9). First-strand cDNA was prepared by using the Reverse Transcription System (Promega), and 1-μl aliquots were used directly in PCRs. For library construction, piroplasm RNA was prepared by the method of Han et al. (18) and poly(A)$^+$ RNA was purified as described previously (39). A Bethesda Research Laboratories cDNA synthesis kit was used, according to the manufacturer's instructions, to prepare cDNA which was cloned into λgt11 arms (Promega). The library was screened with the 623-bp *T. parva* Muguga *Tpr* fragment, and the inserts (1,310 and 1,338 bp) were subcloned into the *Eco*RI site of pUC19. Six additional cDNA clones were isolated by PCR amplification from first-strand cDNA primed with ILO 150, an oligo(dT) primer with a G/C-rich sequence at the 5′ end to allow the use of stringent priming conditions to amplify the 3′ ends of cDNA. The PCR used as primers ILO 194, which was located approximately 700 bp upstream of the 3′ end of the *Tpr*1 ORFs (see Fig. 1), and ILO 150 (Table 1). In control reactions, genomic DNA templates were not amplified with these primers under the conditions used. The PCR products (739 to 888 bp) were cloned into the *Eco*RV site of pBluescript KS.

PCR amplification of genomic DNA and cDNA. Amplifications were performed with Promega *Taq* polymerase in the buffer supplied by the manufacturer with the addition of 1.5 mM $MgCl_2$. The cycling conditions were denaturation at 94°C for 60 s, annealing at 55°C for 60 s, and extension at 72°C for 90 s for 30 cycles.

Nucleotide sequencing. Nucleotide sequencing strategies employed were either generation of nested deletions with exonuclease *Bal* 31 (Bethesda Research Laboratories) and subcloning of the deletions into *Sma*I/*Eco*RI-cut M13mp18 or synthesis of successive rounds of oligonucleotide primers. Sequencing reactions used either the Sequenase kit (United States Biochemicals) or the "fmol" DNA sequencing system (Promega). Sequences were determined on both strands.

Gel electrophoresis and Southern hybridization. Restriction endonuclease digests of piroplasm (1 μg) or schizont-infected lymphocyte DNA (20 μg) were carried out to completion according to the manufacturer's specifications (New England Biolabs). The DNA was size fractionated in 0.8% agarose gels and transferred onto nylon filters. Following hybridization, filters were washed in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at 60°C, unless otherwise indicated in the individual figure legends.

Pulsed-field gel electrophoresis. Preparation of *T. parva* high-molecular-weight DNA, digestion of the agarose-embedded DNA with *Sfi*I, and separation of restriction endonuclease-digested fragments by contour-clamped homogeneous electric field (CHEF) electrophoresis were performed according to the methods of Morzaria et al. (33). For CHEF electrophoresis, a sequential pulse frequency of 10 s for 16 h and 40 s for 2.5 h at 200 V was used. Size markers were concatemers of bacteriophage λ cI857 Sam 7 ladders and *Saccharomyces cerevisiae* chromosomes (Cambridge Bioscience).

Nucleotide sequence accession numbers. Nucleotide sequences were submitted to the GenBank database with accession numbers L48610 (Boleni genomic DNA, 2,276 bp), L48615 (Uganda genomic DNA, 1,504 bp), L48612 (Uganda genomic DNA, 2,370 bp), and L48611 (Muguga cDNA, 1,338 bp).

RESULTS

Comparative analysis of *Tpr* ORFs between *T. parva* isolates. Genomic clones containing DNA from *T. parva* Boleni and *T. parva* Uganda were isolated on the basis of their strong hybridization with radiolabelled homologous total DNA. The nucleotide sequences of two of the selected clones revealed the presence of long ORFs of 2,099 bp (699 codons) in a *T. parva* Boleni sequence and 1,872 bp (624 codons) in a *T. parva* Uganda sequence which were incomplete at the 5′ and 3′ ends, respectively. The nucleotide and deduced peptide sequences of the 3′/C-terminal end (233 codons) of the Boleni ORF exhibited 75% similarity and identity, respectively, with a section of *Tpr*1, a repetitive coding element (100 copies per genome) which is a constituent of tandemly arrayed ORFs within the *T. parva* Muguga *Tpr* locus (3). One of the copies of the *Tpr*1 element forms the 3′ end of a 2,415-bp (805-codon) ORF in *T. parva* Muguga DNA (3). In addition, a 360-bp sequence in the middle of the Boleni ORF was 75% identical to a second repeated element, *Tpr*2 (30 copies per genome), also present within the 2,415-bp ORF. The organization of the *T. parva* Muguga genomic ORFs and their constituent repeat units, which were determined in an earlier study, is shown in Fig. 1A. PCR amplification using primers ILO 1099 and 1322 (Fig. 1 and Table 1), derived from the 3′ end of the 1,872-bp Uganda ORF and the 3′ end of *Tpr*1, respectively, generated a PCR product containing a 1,504-bp (501-codon) ORF from Uganda genomic DNA, which overlapped the 1,872-bp Uganda ORF by 51 bp. The 1,504-bp Uganda ORF contained *Tpr*1- and *Tpr*2-homologous elements at the 3′ end, similar to those in the ORF in the cloned *T. parva* Boleni DNA. Amplification using primers ILO 1231, located near the 5′ end of the 1,872-bp Uganda ORF, and ILO 1322 (Fig. 1 and Table 1) generated a PCR product of approximately 3,300 bp from Uganda genomic DNA, which was consistent with the overlapping of the ORFs in the two cloned sequences as a 3,325-bp (1,108-codon) ORF in *T. parva* Uganda genomic DNA (Fig. 1).

There was no significant similarity (>50% over 100 bp) or identity (>25% over 100 amino acids) among the 5′/N-terminal ends of the Muguga (210 codons), Boleni (305 codons), and Uganda (665 codons) ORFs, when either the nucleotide or the deduced peptide sequences were compared. The regions of similarity and difference among the *T. parva* Muguga, Boleni, and Uganda ORFs are summarized diagrammatically in Fig. 1B.

All the *Tpr* ORFs were predicted to be protein-encoding DNA by the algorithms of Fickett (17) and Shepherd (46). This remained true even when the analysis was applied independently to the 5′ ends of the ORFs which were not conserved among the three different *T. parva* isolates. The nine codons which are rarely used in protein-encoding *Theileria* and *Babesia* genes (16) were represented rarely, or not at all, in the ORFs.

Blot analysis of *T. parva* genomic DNA using 5′ and 3′ *Tpr* ORF probes. Probes derived from the isolate-specific 5′ and conserved 3′ regions of the *Tpr* ORFs (Fig. 1) were hybridized to *Eco*RI-digested DNAs from stocks and clones of *T. parva*. The 5′ probes hybridized to many *Eco*RI fragments in homologous DNA derived from cloned isolates of Muguga, Boleni, and Uganda (Fig. 2A, B, and C), indicating the existence of multiple copies of these sequences in the *T. parva* genome. The 5′-end probes hybridized specifically to the DNA of the homologous parasite clone even at low washing stringencies (Fig. 2A, B, and C), except in the case of the *T. parva* Muguga probe, which hybridized weakly to *T. parva* Uganda DNA (Fig. 2C, lane 3). The 3′ probes hybridized to multiple polymorphic restriction fragments in the DNA of both homologous and heterologous isolates, and the restriction patterns were similar for the probes from all three stocks (Fig. 2D; a representative result with the *T. parva* Muguga probe is shown). The se-

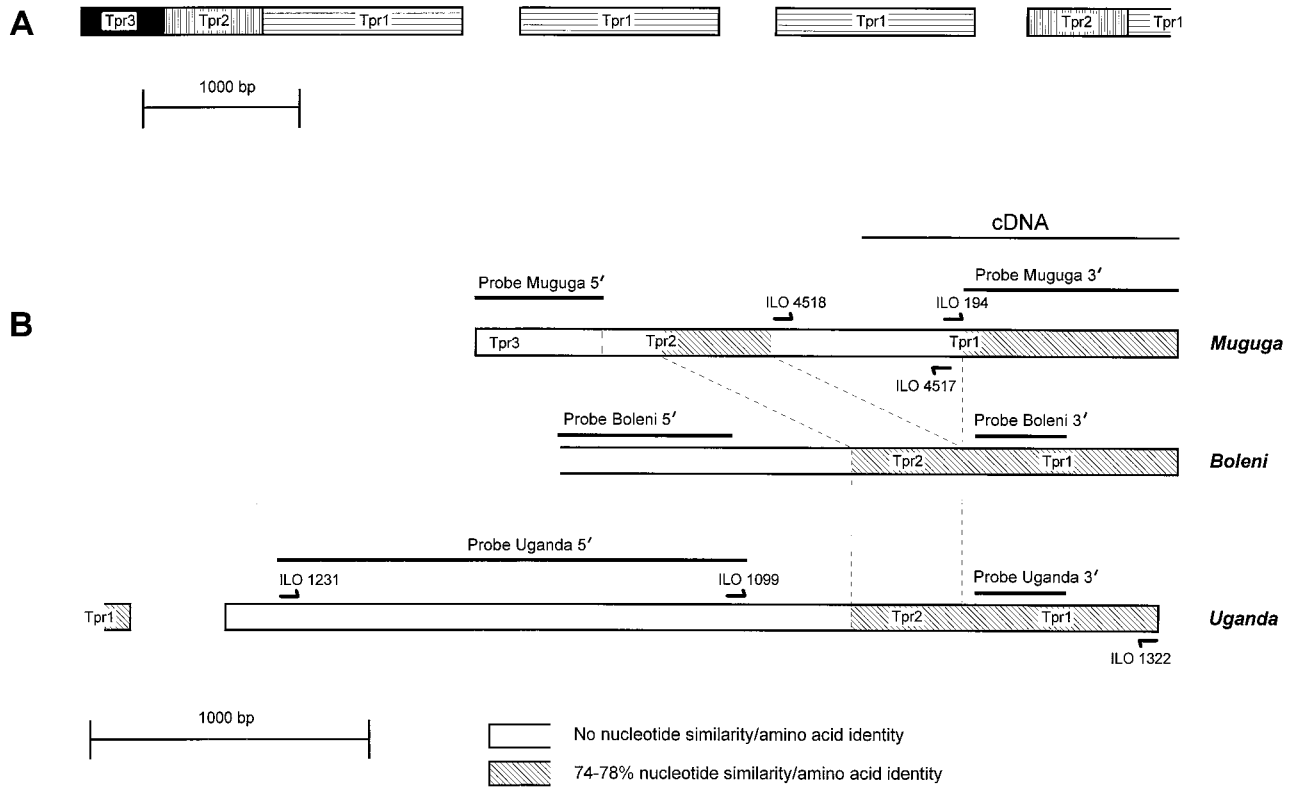

FIG. 1. Organization and similarity/identity of *Tpr* ORFs in genomic DNA fragments from *T. parva* Muguga, Boleni, and Uganda. (A) Arrangement of the repeat units *Tpr*1, *Tpr*2, and *Tpr*3 within ORFs located in an 8.1-kb fragment of *T. parva* Muguga DNA, which was sequenced in an earlier study (GenBank accession number X55385) (3). (B) Similarities and differences of ORFs in DNA sequences from *T. parva* Muguga (long first ORF from panel A; 805 codons), Boleni (699 codons), and Uganda (1,108 codons), which are shown with the 3′/C-terminal ends aligned. The regions exhibiting homology to the repeat units *Tpr*1 and *Tpr*2 are indicated for each ORF. The Uganda ORF is a composite derived from the sequence of two cloned DNAs. The Boleni ORF is incomplete at the 5′ end. Shaded regions within the ORFs exhibit significant nucleotide similarity and amino acid identity, and unshaded regions have no similarity/identity (<50% at the nucleotide level and <25% at the protein level). The position of the 3′ end of an additional partial copy of *Tpr*1 in the *T. parva* Uganda sequence, located 5′ to the long ORF, is also shown. The locations of DNA sequences used as probes and of oligonucleotide primers for PCR amplification of DNA are indicated. The extent of *T. parva* Muguga cDNA clones homologous to *Tpr*1 is also shown.

quences containing the 5′ ends of the *Tpr* ORFs were, therefore, absent from or highly divergent in the DNAs of other cloned isolates, although present in multiple copies in the genome of the homologous clone. The *T. parva* Uganda 5′ probe did not hybridize to the *Eco*RI-digested DNAs of 12 additional *T. parva* isolates from a wide geographical range (Fig. 2E). Similar results were obtained when the *T. parva* Boleni 5′ probe was hybridized to the same 12 isolates (data not shown).

The approximate genomic copy number of the isolate-specific 5′ section of the *T. parva* Uganda Tpr ORF was assessed by hybridization of the Uganda 5′ probe to a dilution series of known amounts of the probe sequence and genomic DNA from the cloned parasite. Quantitative comparison of the intensity of hybridization of the probe to the cloned sequence and the genomic DNA indicated a copy number of between 10 and 20.

Comparison of different copies of Tpr ORFs within *T. parva* isolates. The nucleotide sequences of two additional genomic clones which contained Tpr sequences, one from *T. parva* Uganda (2,392 bp) and one from *T. parva* Boleni (2,753 bp), were determined on one strand. The *T. parva* Uganda sequence contained a 656-codon ORF which was incomplete at the 3′ end. The ORF exhibited 79% nucleotide similarity and 60% amino acid identity over its whole length with the isolate-specific 5′ end of the first Uganda ORF. There were a total of

TABLE 1. PCR amplification primers

No.	Oligonucleotide sequence 5′ to 3′	Application
1099	GGTAGTACTAGTGCCACC	Amplification of *Tpr* ORFs from Uganda genomic DNA
1231	CCTGGGTAAAGAAGCTGC	Amplification of *Tpr* ORFs from Uganda genomic DNA
4697	GGTGGCACTAGTACTACC	Reverse complement of 1009; for use with 1231
1322	GTTCCTTGGTGAACAAGTC	For use with 1099 and 1231
4518	GGCCACTGAGGTAGTAAAA	For amplification of the 5′ section of *Tpr*1
4517	AGTACCTTGAGAGAGAGC	For use with 4518 to amplify the 5′ section of *Tpr*1
194	ATATATCCAGCCATAGCTCCTGGAATGATTGT	Amplification of *Tpr* from piroplasm cDNA
150	TAGGCGCGCC(T)^{20}	Oligo(dT) clamp primer for cDNA amplification

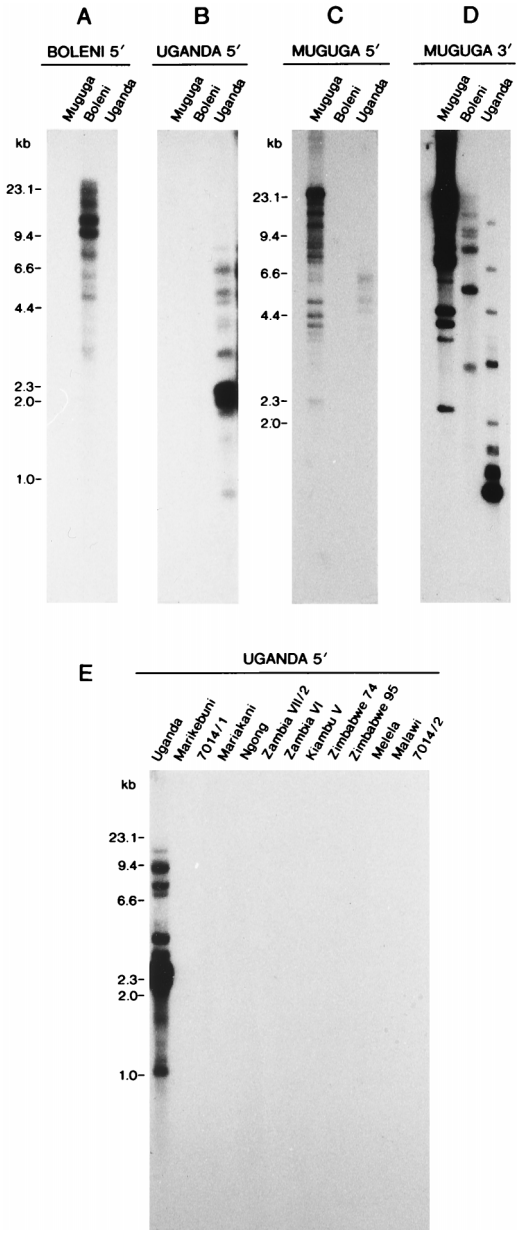

FIG. 2. Multiple copies of *T. parva* DNA sequences which are maintained as ORFs are isolate specific. Southern blots of *Eco*RI-digested piroplasm DNAs (1 μg) of cloned parasites derived from the *T. parva* Muguga, Boleni, and Uganda stocks are shown in panels A to D, and those of *Eco*RI-digested infected lymphocyte DNAs (20 μg) of *T. parva* field isolates are shown in panel E. The Southern blots were hybridized with probes containing the 5′ (A, B, C, and E) or 3′ (D) ends of *Tpr* ORFs as indicated. See Fig. 1 for location and extent of the probes with respect to the ORFs.

12 nucleotide insertions or deletions between the two sequences, but these were multiples of 3 (3 to 21 bp) such that the ORF was conserved. Part of a third *T. parva* Uganda Tpr ORF was isolated by PCR amplification from the genomic DNA of the cloned parasite by using as primers ILO 1231 and an oligonucleotide which was the reverse complement of ILO 1099 (Fig. 1 and Table 1). Nucleotide sequence analysis indicated that this partial copy was more closely related to the first ORF (an ORF which was identical in sequence to the first ORF was also amplified with these primers) in that it was 94% similar over the 325 bp of sequence which was determined. The additional *T. parva* Boleni sequence contained a 618-codon ORF which was incomplete at the 5′ end. The 3′/C-terminal 500 codons of this ORF exhibited approximately 90% similarity and identity with the corresponding region of the first sequenced Boleni ORF; the homologous section encompassed 160 codons within the isolate-specific section of the ORF. The N-terminal 270 and 120 codons of the first and second Boleni Tpr ORFs, respectively, were not similar at the DNA or protein levels. The isolate-specific 5′ sections of the Tpr sequences thus showed a high degree of similarity and identity between different copies within a genome and were maintained as ORFs in the examples studied. The ORFs were, however, not absolutely conserved in length and also contained domains which were divergent between the copies.

The *Tpr* ORFs are tandemly arrayed within a single region of the *T. parva* genome. Previous analysis has shown that in *T. parva* Muguga the *Tpr* ORFs are present as a tandem array (3) and that the *Tpr*1 sequences are located on two adjacent, chromosome-internal, *Sfi*I fragments (numbers 3 and 6) in the *T. parva* Muguga genome (34). In this study, a 115-bp sequence with 78% similarity to the extreme 3′ end of the *T. parva* Muguga *Tpr*1 was identified 300 bp upstream of the large 624- and 656-codon ORFs in Uganda DNA (Fig. 1), indicating that the ORFs are also tandemly arranged in *T. parva* Uganda. The 5′ and 3′ probes derived from *T. parva* Uganda *Tpr* ORFs (Fig. 1) both hybridized to a single *Sfi*I fragment of about 600 kb in size in CHEF gel separations of *T. parva* Uganda DNA (Fig. 3). Multiple copies of the conserved and nonconserved sections of the *Tpr* ORFs are thus located together on a single *Sfi*I fragment in *T. parva* Uganda DNA.

***Tpr*1-homologous sequences are present in other species of *Theileria*.** A Southern blot of *Eco*RI-digested DNAs of *T. parva*, *T. taurotragi*, and *T. annulata* was hybridized with the *T. parva* Muguga 3′ *Tpr*1 probe. Under low-stringency washing conditions, the probe hybridized to multiple *Eco*RI fragments in *T. taurotragi* and *T. annulata* DNA (Fig. 4, lanes 2 to 4) but not to trypanosome or uninfected bovine DNA (Fig. 4, lanes 5 and 6). The 3′ section of *Tpr*1 is therefore conserved among several *Theileria* species.

Polymorphic transcripts derived from *Tpr* sequences. It has previously been shown, by Northern blot analysis, that *Tpr*1 and *Tpr*3 are transcribed in the piroplasm, but not the schizont stage of *T. parva* Muguga, with hybridization to discrete RNA bands of 3.5 and 3.1 kb in size being observed (3). The nucleotide sequences of eight *Tpr*1-homologous piroplasm cDNAs (1,338 to 739 bp in length) which were derived from parasite RNA from a single animal experimentally infected with the uncloned *T. parva* Muguga stock were determined (data not shown; GenBank accession number L48611). The extent of the transcripts with respect to the 805-codon *T. parva* Muguga genomic ORF is indicated in Fig. 1. The transcripts were derived from a minimum of four loci, as assessed by the observation of four very distinct 3′ untranslated sequences among the eight clones, and all were polyadenylated. The 3′ untranslated region of one of the cDNAs was identical to the region 3′ of the 805-codon genomic ORF but differed within the coding region. None of the transcripts was identical in sequence within the potential protein-coding sequences. Within the 700 bp at the conserved 3′ end of *Tpr*1, which was also present in the Boleni and Uganda sequences (Fig. 1), the most similar pair of transcripts differed at 2 nucleotide positions and the most different differed at 18 nucleotide positions. A small proportion of the substitutions in the cDNAs generated by PCR amplification may be attributable to nucleotide misincorporation by *Taq* polymerase. However, about 50% of the substitu-

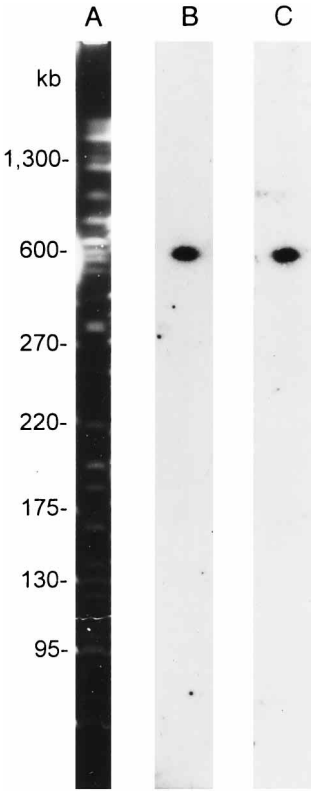

FIG. 3. The 5′ and 3′ ends of *T. parva* Uganda *Tpr* ORFs are located on a single genomic *Sfi*I fragment. (A) CHEF gel separation of *T. parva* Uganda agarose-embedded DNA digested with *Sfi*I and stained with ethidium bromide. (B and C) Duplicate Southern blots of the gel in panel A hybridized with *T. parva* Uganda *Tpr* probes derived from the 5′ and 3′ ends of the ORF, respectively. See Fig. 1 for the derivation of these probes.

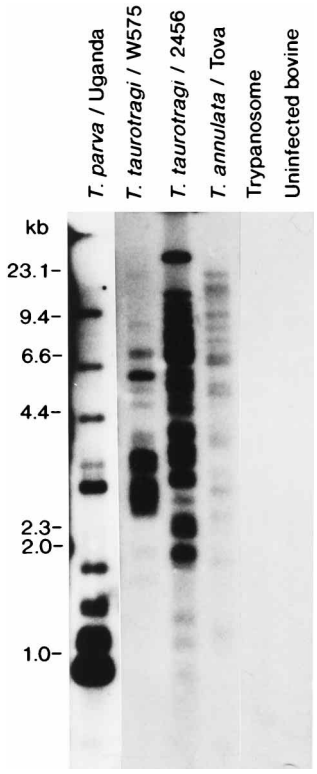

FIG. 4. DNA sequences homologous to the *Tpr*1 repeat element are present in the *T. taurotragi* and *T. annulata* genomes. A Southern blot of *Eco*RI-digested schizont-infected lymphocyte DNA (20 μg) of three *Theileria* species, together with uninfected bovine DNA (20 μg) and *Trypanosoma congolense* DNA (5 μg), is shown. The blot was hybridized at 55°C with the *T. parva* Muguga 3′ *Tpr*1 probe (Fig. 1) and washed in 2× SSC at the same temperature.

tions in the PCR-generated cDNAs were located at sites at which the two cDNAs isolated by library screening also differed from one another, strongly indicating that many of the sequence differences observed were real. The two longest cDNAs extended into the 5′ end of the *T. parva* Muguga *Tpr*1, for which no equivalent sequences were present in the *T. parva* Boleni and Uganda ORFs (Fig. 1). Many of the nucleotide sequence differences in the 5′ end of *Tpr*1 resulted in amino acid substitutions between the deduced peptides (14 amino acid substitutions from 32 nucleotide differences), whereas a greater proportion of the nucleotide substitutions within the 3′ end of *Tpr*1 were synonymous at the amino acid level (four amino acid substitutions from 13 nucleotide differences). The nonconserved 5′ end of *Tpr*1, including the region which was homologous to the 5′ end of the cDNAs, was amplified from genomic DNA by using primers ILO 4517 and ILO 4518 (Fig. 1 and Table 1). When this PCR product was used to probe blots of *T. parva* Muguga, Boleni, and Uganda DNA, it hybridized strongly to nine restriction fragments in Muguga DNA, less strongly to a single fragment in Uganda DNA, and not at all to Boleni DNA, after washing in 1× SSC–0.1% sodium dodecyl sulfate at 65°C (data not shown).

Predicted transmembrane segments at the C-terminal end of *Tpr* ORFs. The C-terminal regions of the *Tpr* ORFs exhibited high hydropathicity and were predicted to be integral membrane proteins (23, 24). A computer-assisted search of the deduced protein sequences encoded by the Muguga, Boleni, and Uganda ORFs using three different algorithms, those of Eisenberg et al. (15), Mohana-Rao and Argos (32), and Klein et al. (23), detected six predicted transmembrane helices within the conserved regions of *Tpr*2 and *Tpr*1. A CLUSTAL (19) alignment of the C-terminal region of the ORFs (Fig. 5) shows that the location and sequence of the membrane-associated segments are conserved in all three ORFs. The *Tpr*3 region at the 5′ end of the *T. parva* Muguga ORF contained three additional predicted transmembrane helices, without equivalents in the Boleni and Uganda ORFs. The isolate-specific 5′ ends of the Boleni and Uganda ORFs were relatively hydrophilic (23) and did not contain membrane-associated sequences.

The conserved region of *Tpr*1 evolves as if it is protein-encoding DNA. The evolution of the conserved 3′ end of *Tpr*1 (Fig. 1) was analyzed for the *T. parva* Muguga, Boleni, and Uganda sequences. The sequences are shown in a CLUSTAL (19) alignment in Fig. 6. Since there are multiple copies of *Tpr*1 in the *T. parva* genome, the sequences compared may not be exact homologs, but it is assumed that all the copies of *Tpr*1 are ultimately derived from a single ancestral copy by gene duplication. A comparison of the numbers of synonymous and nonsynonymous substitutions was performed by the methods of Nei and Gojobori (38). This analysis revealed that the ratio of synonymous substitutions per synonymous site to nonsynonymous substitutions per nonsynonymous site was approximately 4:1. The excess of synonymous substitutions suggests that the *Tpr*1 region of this repeat family is evolving under selective constraints with respect to mutations causing amino acid re-

```
MUGUGA                              VVYPSIAVQWLNFITYLILLLVYVPGD  220
BOLENI                              VIWTSVIVQWLNFITYVVLLFVYVPGD  333
UGANDA                              IIVPSVVVQWLNFVTYLILLFVYVPGD  148
                                       *  ****** **  ** ******
MUGUGA   QGHLTAFYFVIAVSGFVFGLSNVLTFAIDDNYLPIYIAGENSFPVLTSLI  270
BOLENI   QGHLTTFYWVIAVSGVVFGLNNVMVFAADWKYLPIYIAGENSFPALTSLI  383
UGANDA   HGHLTTFYWVIAVSGFVFGFNNVMVFAADLTYLPIYIAGENCFPTLTSLI  198
          **** ** ****** **   **  ** *  ********** ** *****
MUGUGA   HYLSSLMFGNRRKWNSDFLMVYIDLVVAILLSLLSAVMWTLGYVSKHPTE  320
BOLENI   HYVTTLSFGNRRKWNSDFIIVCIDIWVAIAISLAAAVVWTFGYLAKHPSD  433
UGANDA   HYVSMLTFGNRRKWNSDFLMVYIDLVVAIAISLAAAVVWTVAYLSQHPTD  248
         **     ***********  * **  ***  **  ** **  *   **
MUGUGA   DDALKSKATEVVKAIIQGLNGKVDKLATAVVAKVEQALKSAQAQQDPVNL  370
BOLENI   TTDTSSN-------------------------------------------  440
UGANDA   NKNR---------------------------------------------  252

MUGUGA   DLTETIQAVLNAANGLQGQVQSVKEALEKHLEITNIQGGVDKSAVQAAIQ  420

MUGUGA   NLQSKKQAAIAAVKGALGGPSANDTAKIEDKDLTAPATKALKTVKLEAQA  470

MUGUGA   GAVKTNLKSGGSGKTQINIGDSITQALTKLEQNKPGTTTGEKEPKPEAVQ  520

MUGUGA   DVQSALSQGTAKTQFTSLWHIKGTDGYSSLDSAHWYIMFPFLMVLIGMGL  570
BOLENI   ----------------RLWQILPPDSYGAIR-LTSSFIFPFLMVLVGMGL  473
UGANDA   ------------------YHHVHISTFTDIKELQWHIIFPILMVLVGMGL  284
                                          ** **** ****
MUGUGA   VYCIYPAIAPGMIVPFYLVDKIEMVLLIATIFPALYVAIARSGKLIPGFG  620
BOLENI   VYAIYPAIAPGMIVPFYLIDKIEMVLLILTIFPPVTVALLRK--Y--GKG  519
UGANDA   VFAIYPGIAPGMIVPFYLIDKIEMVLLILTIFPPFFVAAARGKWF--GYG  332
         *  *** ********** ********* ****   **  *      * *
MUGUGA   GFFNPVSPTCNWGAKNELPWIPGVPLGQGYHWHLTDLVIPTMIILAYLFI  670
BOLENI   GQS----PISHWDDTNQ-NFVTKAPLPSGFGWHVLDPLIPLMIILAIIFI  564
UGANDA   GFSSPWSPTSKWGPGNK-DWGG-TDLGNGAGWHFFDIVIPLMIILAAIFI  380
         *          *             *  *  **  *  ** ***** ***
MUGUGA   YSLHYRDSSVARSIINQPKMSTCLTILFYMCHEISLAVGFPGIFGGNGGG  720
BOLENI   YSLHYRESSLSRSIINQPKMSTFLTIMFYMCHEISLAVGFPGILGNSGFS  614
UGANDA   YSLHYRESSVARSIINQPKMSTCLTILFYMCHEVSLAVGFPGVFGGNGGG  430
         ****** **  ********** *** ****** ********  *  *
MUGUGA   SILALTAQLMGAFLMCLLAPYSEGYIIEYKRHDPSNWPTAGMTRWNALRY  770
BOLENI   SIMALLAQIIGAFLMVFLATYSEGYIIEYKRHTPEHWPTAGMTRWNAFCY  664
UGANDA   DTLALLCQLIGAFLMVFLALYSEGYIIEYKRHDSSHWPTTGMTKWNSLCY  480
            **  *  *****  ** ************    *** *** **   *
MUGUGA   WTKMASKNCNKNLAALFTKDL---RRDLLLCIERNELI              805
BOLENI   WAKMASRICNKNLRDLFTKDL---RRDLLLCIQTGDLI              699
UGANDA   WCKMASRITNKNLRDLFTKDR                               501
         * ****   ****  *****    ********    **
```

FIG. 5. Transmembrane segments located in the conserved C-terminal regions of *Tpr* ORFs. A CLUSTAL (19) alignment of *Tpr*1 and *Tpr*2 sequences at the C-terminal end of the *T. parva* Muguga, Boleni, and Uganda *Tpr* ORFs is shown. The six predicted transmembrane sequences are underlined and shown in boldface. The transmembrane segments, each a minimum of 17 amino acids in length, represent sequences identified by three different predictive algorithms (15, 24, 32). Amino acid residues which are conserved in all sequences are indicated by asterisks.

placements, a characteristic property of DNA sequences that encode proteins (22, 31).

DISCUSSION

The results presented herein demonstrate the presence of multiple copies of DNA sequences in the genome of individual *T. parva* cloned isolates, which are maintained as ORFs with protein-coding potential as assessed by patterns of codon usage, although similar sequences are not detectable by hybridization in other stocks and clones of the parasite. The isolate-specific regions of the *T. parva* Boleni (305 codons) and Uganda (665 codons) ORFs are considerably longer than the largest ORFs (200 codons) generated by chance in simulation studies of random DNA sequences (45). The data indicate strong selective pressure both for nucleotide sequence divergence and for maintenance of the ORFs. It is unclear whether maintenance of ORFs implies that they encode proteins, since evidence has recently been presented for the existence of ORFs in the genome of yeast which do not encode proteins (7, 54). If the N-terminal ends of the *Tpr* ORFs encode domains within proteins, their functions must be minimally dependent on amino acid sequence, so that like fibrinopeptides (12) they can tolerate almost any amino acid change. Despite their lack of detectable sequence identity, the proteins encoded by the *Tpr* ORFs could have common structures, since it is known from X-ray crystallographic data that protein tertiary structure is often more conserved than primary amino acid sequences (28). One example in protozoan parasites is the N-terminal variable domains of trypanosome variant surface glycoproteins (6).

The nucleotide sequence of the p67 sporozoite antigen gene (39), which is encoded on the same 780-kb *Sfi*I fragment as the *Tpr* locus (34), is identical in *T. parva* Muguga, Boleni, and Uganda (40). The unusual conservation of the p67 gene, which extends to the 29-bp intron and the third base position of all

```
    Muguga cattGgtacaTtATgTTTCCatTtCTtATGGTCcTgaTtGGTATGGGtCTaGTcTactgCATaTATCCaGccATaGCtCCTGGAATGATTGTtCCaTTtTACCTcgTgGAtAAGATTGAGATGGTtCTgtTGATagccACCATATTtCCA  150
    Boleni accaGttcatTcATaTTTCCatTcCTcATGGTCcTtgTgGGTATGGGaCTtGTcTatgcCATaTATCCaGccATtGCaCCTGGAATGATTGTgCCcTTcTACCTtaTtGAcAAGATTGAGATGGTaCTtcTGATccttACCATATTcCCA  150
    Uganda cagtGgcataTcATaTTTCCtaTtCTtATGGTCcTggTaGGTATGGGaCTgGTgTttgcCATtTATCCtGgaATaGCtCCTGGAATGATTGTgCCaTTtTACCTgaTtGAcAAGATTGAGATGGTaCTtcTGATccttACCATATTcCCA  150
Muguga pep h  w  y  i  m  F  P  f  L  M  V  L  i  G  M  G  L  V  y  c  I  Y  P  a  I  A  P  G  M  I  V  P  F  Y  L  v  D  K  I  E  M  V  L  L  I  a  T  I  F  P   50
Boleni pep t  s  s  f  i  F  P  f  L  M  V  L  v  G  M  G  L  V  y  a  I  Y  P  a  I  A  P  G  M  I  V  P  F  Y  L  i  D  K  I  E  M  V  L  L  I  l  T  I  F  P   50
Uganda pep q  w  h  i  i  F  P  i  L  M  V  L  v  G  M  G  L  V  f  a  I  Y  P  g  I  A  P  G  M  I  V  P  F  Y  L  i  D  K  I  E  M  V  L  L  I  l  T  I  F  P   50

    Muguga gCtcTgtatGTgGCaattgccAGAagtggtaaactGataccaGGtTtTGGTGGtTTCttcaaTcccgtGTCtCCtAcATgTaAtTGGGgagccaaaAAtgAActaccctggAtaccTggagtaccaTTaggccaaGGataccacTGGCAc  300
    Boleni cCagTcacaGTgGCtctttctgAGAaagtacgg----------------------CaaaggTggtcaGTCaCCaAtATcTcAcTGGGacgacactAAccAAaacttcgtcAc---TaaggctccgttgccttctggctttggtTGGCAt  273
    Uganda cCatTctttGTaGCtgcagccAGAggcaa------GtggtttggctatggtggTTCtctagTccatgGTCaCCtAcATcTaAgTGGGgaccaggtAAcaAgattgggg------TggaacagatTTgggcaatGGtgctggtTGGCAc  288
Muguga pep a  l  y  V  A  i  a  R  s  g  k  l  i  p  G  f  G  G  F  f  n  p  v  S  P  t  c  n  W  g  a  k  N  e  l  p  w  i  p  g  v  p  L  g  q  G  y  h  W  H  100
Boleni pep p  v  t  V  A  l  l  R  k  y  g  -  -  -  -  -  -  -  -  k  g  g  q  S  P  i  s  h  W  d  d  t  N  q  n  f  v  t  -  k  a  p  L  p  s  G  f  g  W  H   91
Uganda pep p  f  f  V  A  a  a  R  g  k  -  -  w  f  G  y  G  G  F  s  s  p  w  S  P  t  s  k  W  g  p  g  N  k  d  w  g  -  -  g  t  d  L  g  n  G  a  g  W  H   96

    Muguga cTaacTGAcctagTgATTCCcaccATGATcATttTAGCctattTATTtATTTACTCACTGCATTACAGAGAcTCatcagTtgCCAGATCaATcATCAAcCAACCCAAAATGTCTACATgTCTaACcATtcTaTTCTATATGTGTCATGAG  450
    Boleni gTtctTGAtccacTtATTCCactcATGATcATacTAGCtatcaTATTcATTTACTCACTGCATTACAGAGAaTCcagtcTatCCAGATCcATcATCAAtCAACCCAAAATGTCTACATtTCTgACcATcaTgTTCTATATGTGTCATGAG  423
    Uganda tTcttTGAtattgTaATTCCtcttATGATaATatTAGCtgctaTATTtATTTACTCACTGCATTACAGAGAaTCttcagTtgCCAGATCcATaATCAAtCAACCCAAAATGTCTACATgTCTaACtATtcTaTTCTATATGTGTCATGAG  438
Muguga pep l  t  D  l  v  I  P  t  M  I  I  L  A  y  l  F  I  Y  S  L  H  Y  R  d  S  S  v  a  R  S  I  I  N  Q  P  K  M  S  T  c  L  T  I  l  F  Y  M  C  H  E  150
Boleni pep v  l  D  p  l  I  P  l  M  I  I  L  A  i  i  F  I  Y  S  L  H  Y  R  e  S  S  l  s  R  S  I  I  N  Q  P  K  M  S  T  f  L  T  I  m  F  Y  M  C  H  E  141
Uganda pep f  f  D  i  v  I  P  l  M  I  I  L  A  a  i  F  I  Y  S  L  H  Y  R  e  S  S  v  a  R  S  I  I  N  Q  P  K  M  S  T  c  L  T  I  l  F  Y  M  C  H  E  146

    Muguga aTcTCatTgGCTGTAGGATTTCCAGGTaTCtTtGGTggtAaTGGaggtgGTagcAtCtTaGCccTaacGgcaCAgtTgATgGGtGCATTtCTGATGtgtcTgtTAGCtccaTATTCTGAAGGtTACATTATTGAGTATAAgaGACAtgaT  600
    Boleni aTcTCtcTtGCTGTAGGATTTCCAGGTaTCcTaGGTaacAgTGGgttcaGTagtAtCaTgGCccTgttGgctCAgaTaATaGGaGCATTcCTGATGgtatTtcTAGCtacaTATTCTGAAGGaTACATTATTGAGTATAAacGACAcacT  573
    Uganda gTaTCacTgGCTGTAGGATTTCCAGGTgTCtTtGGTggtAaTGGtggagGTgacAcCtTaGCatTactGtgtCAacTaATaGGtGCATTtCTGATGgtatTtcTAGCattgTATTCTGAAGGtTACATTATTGAGTATAAgaGACAtgaT  588
Muguga pep i  S  L  A  V  G  F  P  G  i  f  G  g  n  G  g  g  s  i  l  A  L  t  a  Q  l  m  G  A  F  L  M  c  l  L  A  p  Y  S  E  G  Y  I  I  E  Y  K  R  H  d  200
Boleni pep i  S  L  A  V  G  F  P  G  i  l  G  n  s  G  f  s  s  i  m  A  L  l  a  Q  i  i  G  A  F  L  M  v  f  L  A  t  Y  S  E  G  Y  I  I  E  Y  K  R  H  t  191
Uganda pep v  S  L  A  V  G  F  P  G  v  f  G  g  n  G  g  g  d  t  l  A  L  l  c  Q  l  i  G  A  F  L  M  v  f  L  A  l  Y  S  E  G  Y  I  I  E  Y  K  R  H  d  196

    Muguga cCTagcaATTGGCCTACagCaGGAATGACtAgGTGGAATgCccTtaGgTACTGGaccAAaATGGCcTCaAagAactgCAATAAGAACCTagcAGcaTTGTTcACCAAgGACCTgaGaAGAGACCTGcTACTaTGcATagAaAgAaaTGAg  750
    Boleni cCTgaacATTGGCCTACtgCtGGAATGACtAgGTGGAATgCctTctGtTACTGGgctAAgATGGCtTCtAggAtatgCAATAAGAACCTgagAGacTTGTTtACCAAaGACCTacGtAGAGACCTGtTACTcTGtATccAgAcAggTGAc  723
    Uganda tCTagtcATTGGCCTACtaCgGGAATGACaAaGTGGAATtCacTctGtTACTGGtgcAAgATGGCtTCcAgaAtaacCAATAAGAACCTaagAGacTTGTTcACCAAgGAC  699
Muguga pep p  s  n  W  P  T  a  G  M  T  r  W  N  a  l  r  Y  W  t  K  M  A  S  k  n  c  N  K  N  L  a  a  L  F  T  K  D  L  R  R  D  L  L  L  C  I  e  r  n  e  250
Boleni pep p  e  h  W  P  T  a  G  M  T  r  W  N  a  f  c  Y  W  a  K  M  A  S  r  i  c  N  K  N  L  r  d  L  F  T  K  D  L  R  R  D  L  L  L  C  I  q  t  g  d  241
Uganda pep s  s  h  W  P  T  t  G  M  T  k  W  N  s  l  c  Y  W  c  K  M  A  S  r  i  t  N  K  N  L  r  d  L  F  T  K  D                                          233


    Muguga CTAATC
    Boleni CTAATC
Muguga pep L  I
Boleni pep L  I
```

FIG. 6. Alignment of nucleotide and deduced peptide sequences encoding the 3′/C-terminal ends of *Tpr*1 from *T. parva* Muguga, Boleni, and Uganda. A CLUSTAL (19) alignment of genomic DNA sequences from the conserved 3′ end of *Tpr*1 is shown, with the corresponding deduced peptide sequences underneath. Mismatched bases and amino acids are shown in boldface and lowercase. Dashes are gaps inserted by the CLUSTAL program to optimize the alignment.

codons, is in marked contrast to the exceptional divergence of the *Tpr* sequences. These data provide evidence for the existence of domains in the genome which differ in sequence stability. Such domains have previously been proposed for other protozoan parasites (29; reviewed in reference 25). The chromosome-internal location of the hyperpolymorphic *Tpr* locus (34) is unusual, since a common theme in the genomic plasticity of parasitic protozoa, including *Plasmodium*, *Trypanosoma*, and *Giardia*, is the compartmentalization of chromosomes into conserved central domains and polymorphic ends (reviewed in reference 25).

The multiple copies of the isolate-specific 5′ ends of the *Tpr* sequences in *T. parva* Boleni and Uganda, which are absent from other stocks and clones, are a dramatic example of concerted evolution or molecular drive, whereby multicopy sequences diverge between, but are homogenized within, populations (2, 14; reviewed in reference 13). Concerted evolution was in addition apparent at the divergent 5′ ends of the *Tpr*1 repeat units which were demonstrated to be transcribed into mRNA by sequencing of cDNA clones. The relatively conserved regions of the *Tpr* locus also exhibit concerted evolution, as the *Tpr*1 and *Tpr*2 elements show 90 and 97% similarity, respectively, between copies within the *T. parva* Muguga genome (3) but only 75% similarity between genomes. Concerted evolution has been most studied in the context of the evolution of the rRNA and histone multicopy gene families (13), but the tandem arrangement of the *Tpr* genes would facilitate the rapid dissemination throughout the locus of mutations originating in individual copies of the gene family, through unequal crossing over (37), gene conversion (47), and other mechanisms of genetic turnover (13). We therefore suggest that in *T. parva* concerted evolution has been responsible for the generation of novel DNA sequences which are maintained as ORFs with coding potential and therefore represent raw material for the emergence of new proteins. Concerted evolution has primarily been observed through the comparison of sequences between different species of vertebrates and insects. An important aspect of the mechanism is that turnover of chromosomes within a reproductively isolated, sexually reproducing population is more rapid than sequence homogenization of a multicopy family within genomes, resulting in concerted evolution at the population level, which has been called molecular drive (13). Since the *T. parva* Muguga, Boleni, and Uganda isolates are from widely separated geographical locations (8, 26, 30), it is reasonable to assume that they originate from reproductively isolated parasite populations. However, a successful genetic cross between *T. parva* Muguga and *T. parva* Uganda has been performed in the laboratory (35).

The biological role of the *Tpr* locus is unknown, although the transcription of the *Tpr* sequences and the conservation of *Tpr*1 in different *Theileria* species indicate that they are functionally important. The presence of conserved transmembrane helices suggests that the C termini of the *Tpr* ORFs potentially encode integral membrane proteins. We have been unable to demonstrate the existence of peptides in *T. parva* corresponding to the *Tpr* genomic ORFs, by using antibodies generated against fusion proteins or synthetic peptides. A keyhole limpet hemocyanin-conjugated synthetic peptide, derived from a conserved sequence of *Tpr*1, which was a predicted antigenic determinant (20), was nonimmunogenic in rabbits and sheep (4a). Precedents exist for proteins and peptides of both *Trypanosoma* and *Plasmodium* which are refractory to the generation of experimental antisera (43, 52). The arrangement of the *Tpr* locus, which is reminiscent of that of vertebrate immunoglobulin genes (3), together with the observation that the eight *Tpr*1-homologous transcripts from a single *T. parva* Muguga-infected animal analyzed in this study all had different nucleotide and corresponding deduced peptide sequences, is consistent with the involvement of the *Tpr* locus in the generation of peptide diversity. In their erythrocytic expression and hypervariability, and in particular the presence of isolate-specific DNA sequences, the *Tpr* ORFs are most similar to the recently described *var* gene family of *P. falciparum* (4, 48, 49). The *var* genes are responsible for the erythrocyte cytoadherence and pathology associated with *P. falciparum* malaria. Further research will be required to determine the role of the *Tpr* locus in intraerythrocytic infection and transmission to the tick vector during *Theileria* infections of cattle.

ACKNOWLEDGMENTS

We appreciate the technical assistance of J. Kiarie. We thank H. Baylis for the H1477 probe, S. Williamson for the gift of *T. taurotragi* DNA, and M. Macklin for the use of the piroplasm cDNA library. We are grateful to N. Murphy and K. Wolfe for help with the analysis of the evolution of *Tpr*1 and to Basil Allsopp for the production of Fig. 6. We thank Noel Murphy, Onesmo ole-MoiYoi, and Per Hagblom for useful suggestions on the manuscript.

REFERENCES

1. **Allsopp, B., M. Carrington, H. Baylis, S. Sohal, T. Dolan, and K. Iams.** 1989. Improved characterisation of *Theileria parva* isolates using the polymerase chain reaction and oligonucleotide probes. Mol. Biochem. Parasitol. **35:**137–148.
2. **Arnheim, N., M. Krystal, R. Schmickel, G. Wilson, O. Ryder, and E. Zimmer.** 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and ape. Proc. Natl. Acad. Sci. USA **77:**7323–7327.
3. **Baylis, H. A., S. K. Sohal, M. Carrington, R. P. Bishop, and B. A. Allsopp.** 1991. An unusual repetitive gene family in *Theileria parva* which is stage-specifically transcribed. Mol. Biochem. Parasitol. **49:**133–142.
4. **Baruch, D. I., B. L. Pasloke, H. B. Singh, X. Bi, X. C. Ma, M. Feldman, T. F. Taraschi, and R. J. Howard.** 1995. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of human parasitised erythrocytes. Cell **82:**77–87.

4a. **Bishop, R., and A. Musoke.** Unpublished data.

5. **Bishop, R. P., B. K. Sohanpal, B. A. Allsopp, P. R. Spooner, T. T. Dolan, and S. P. Morzaria.** 1993. Detection of polymorphisms among *Theileria parva* stocks using repetitive, telomeric and ribosomal DNA probes and anti-schizont monoclonal antibodies. Parasitology **107:**19–31.
6. **Blum, M. L., J. A. Down, A. M. Gurnett, M. Carrington, M. J. Turner, and D. C. Wiley.** 1993. A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. Nature **362:**603–609.
7. **Boles, E., and F. K. Zimmerman.** 1994. Open reading frames in the antisense strands of genes coding for glycolytic enzymes in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **243:**363–368.
8. **Brocklesby, D. W., S. F. Barnett, and G. R. Scott.** 1961. Morbidity and mortality rates in East coast fever (*Theileria parva* infection) and their application to drug screening procedures. Br. Vet. J. **117:**529–531.
9. **Chomczynski, P., and N. Sacchi.** 1987. Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
10. **Conrad, P. A., K. Iams, W. C. Brown, B. Sohanpal, and O. K. ole-MoiYoi.** 1987. DNA probes detect genomic diversity in *Theileria parva* stocks. Mol. Biochem. Parasitol. **25:**213–226.
11. **Cross, G. A. M.** 1990. Cellular and genetic aspects of antigenic variation in trypanosomes. Annu. Rev. Immunol. **8:**83–110.
12. **Doolittle, R. F.** 1986. Of URFS and ORFS, p. 37–47. Oxford University Press, Oxford, United Kingdom.
13. **Dover, G.** 1982. Molecular drive: a cohesive mode of species evolution. Nature **299:**111–117.
14. **Dover, G., and E. Coen.** 1981. Spring cleaning ribosomal DNA: a model for multigene evolution? Nature **290:**731–732.
15. **Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179:**125–142.
16. **Ellis, J., H. Griffin, D. Morrison, and A. M. Johnson.** 1993. Analysis of dinucleotide frequency and codon usage in the phylum Apicomplexa. Gene **126:**163–170.
17. **Fickett, J. W.** 1982. Prediction of protein coding regions in DNA sequences. Nucleic Acids Res. **10:**5303–5318.
18. **Han, J. H., C. Stratowa, and W. J. Rutter.** 1987. Isolation of full-length putative rat lysophospholipase cDNA using improved methods for mRNA isolation and cDNA cloning. Biochemistry **26:**1617–1625.

19. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73:**237–244.
20. **Hopp, T. P., and K. R. Woods,** 1983. A computer program for predicting antigenic determinants. Mol. Immunol. **20:**483–489.
21. **Keese, P. K., and A. Gibbs.** 1992. Origins of genes: big bang or continuous creation? Proc. Natl. Acad. Sci. USA **89:**9489–9493.
22. **Kimura, M.** 1983. The neutral theory of molecular evolution. Harvard University Press, Cambridge, Mass.
23. **Klein, P., M. Kanehisa, and C. Delisi.** 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815:**468–476.
24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
25. **Lanzer, M., K. Fischer, and S. M. Le Blanq.** 1995. Parasitism and chromosome dynamics in parasitic protozoa: is there a connection? Mol. Biochem. Parasitol. **70:**1–8.
26. **Lawrence, J. A., and P. K. I. Mackenzie.** 1980. Isolation of a non-pathogenic *Theileria* of cattle transmitted by *Rhipicephalus appendiculatus*. Zimb. Vet. J. **11:**27–35.
27. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1990. Construction of T vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154.
28. **Mathews, B. W., and M. G. Rossmann.** 1985. Comparison of protein structures. Methods Enzymol. **115:**397–420.
29. **McCutchan, T. F., J. B. Dame, R. W. Gwadz, and K. D. Vernick.** 1988. The genome of *Plasmodium cynomolgi* is partitioned into separable domains which appear to differ in sequence stability. Nucleic Acids Res. **16:**4499–4510.
30. **Minami, T., P. R. Spooner, A. D. Irvin, J. G. R. Ocama, D. A. E. Dobbelaere, and T. Fujinaga.** 1983. Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. Res. Vet. Sci. **35:**334–340.
31. **Miyata, T., T. Yasunaga, and T. Nishida.** 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77:**7328–7332.
32. **Mohana-Rao, J. K., and P. Argos.** 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta **869:**197–214.
33. **Morzaria, S. P., P. R. Spooner, R. P. Bishop, A. J. Musoke, and J. R. Young.** 1990. *Sfi*I and *Not*I polymorphisms in *Theileria* stocks detected by pulsed field gel electrophoresis. Mol. Biochem. Parasitol. **40:**203–212.
34. **Morzaria, S. P., and J. R. Young.** 1992. Restriction mapping of the genome of the protozoan parasite *Theileria parva*. Proc. Natl. Acad. Sci. USA **89:** 5241–5245.
35. **Morzaria, S. P., J. R. Young, P. R. Spooner, T. T. Dolan, A. S. Young, and R. P. Bishop.** 1992. Evidence of a sexual cycle in *Theileria parva* and characterisation of the recombinants, p. 71–74. *In* U. G. Munderloch and T. J. Kurtti (ed.), First International Conference on Tick-Borne Pathogens—An Agenda for Research. Saint Paul, Minn.
36. **Morzaria, S. P., T. T. Dolan, R. A. I. Norval, R. P. Bishop, and P. R. Spooner.** 1995. Generation and characterisation of cloned *Theileria parva* parasites. Parasitology **111:**39–49.
37. **Nayglaki, T.** 1984. Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA **81:**3796–3800.
38. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
39. **Nene, V., K. P. Iams, E. Gobright, and A. J. Musoke.** 1992. Characterization of the gene encoding a candidate vaccine antigen of *Theileria parva* sporozoites. Mol. Biochem. Parasitol. **51:**17–28.
40. **Nene, V., A. Musoke, E. Gobright, and S. Morzaria.** 1996. Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. Infect. Immun. **64:**2056–2061.
41. **Ohno, S.** 1970. Evolution by gene duplication. Springer, Heidelberg, Germany.
42. **Pays, E., L. Vanhamme, and M. Bergerof.** 1994. Genetic controls for the expression of surface antigens in African trypanosomes. Annu. Rev. Microbiol. **48:**25–52.
43. **Roditi, I., and T. W. Pearson.** 1990. The procyclin coat of African trypanosomes. Parasitol. Today **6:**79–81.
44. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **Senapathy, P.** 1986. Origin of eukaryotic introns: a hypothesis, based on codon distribution statistics in genes, and its implications. Proc. Natl. Acad. Sci. USA **83:**2133–2137.
46. **Shepherd, J. C. W.** 1981. Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary significance. Proc. Natl. Acad. Sci. USA **78:**1596–1600.
47. **Smith, G. P.** 1973. Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. **38:**507–513.
48. **Smith, J. D., C. E. Chitnis, A. G. Craig, D. J. Roberts, D. E. Hudson-Taylor, D. S. Peterson, R. Pinches, C. I. Newbold, and L. H. Miller.** 1995. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell **82:**101–110.
49. **Su, X., V. M. Heatwole, S. P. Wertheimer, F. Guinet, J. A. Herrfeldt, D. S. Peterson, J. A. Ravetch, and T. E. Wellems.** 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell **82:**89–100.
50. **Sudhof, T. C., J. L. Goldstein, M. S. Brown, and D. W. Russell.** 1985. The LDL receptor gene: a mosaic of exons shared with different proteins. Science **228:**815–822.
51. **Tsur, I.** 1945. Multiplication *in vitro* of Koch bodies of *Theileria annulata*. Nature **156:**391.
52. **Wilson, R. J., and I. Ling.** 1979. Fractionation and characterization of *Plasmodium falciparum* antigens. Bull. W. H. O. **57:**123–133.
53. **Wistow, G.** 1993. Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem. Sci. **18:**301–306.
54. **Yomo, T., I. Urabe, and H. Okada.** 1992. No stop codons in the antisense strands of the genes for nylon oligomer degradation. Proc. Natl. Acad. Sci. USA **89:**3780–3784.
55. **Young, R. A., V. Mehra, D. Sweetser, T. Buchanan, J. Clark-Curtis, R. W. Davis, and B. R. Bloom.** 1985. Genes for the major protein antigens of the leprosy parasite *Mycobacterium leprae*. Nature **316:**450–452.