Characterisation of the telomeres at opposite ends of a 3 Mb *Theileria parva* chromosome

Baljinder K. Sohanpal, Subhash P. Morzaria, Elke I. Gobright and Richard P. Bishop*

International Livestock Research Institute, PO Box 30709, Nairobi, Kenya

Received February 13, 1995; Revised and Accepted April 20, 1995

GenBank accession nos L36963 and L36964

**ABSTRACT**

Bacterial phage \(\lambda\) clones containing *Theileria parva* genomic DNA derived from two different telomeres were isolated and the nucleotide sequences of the telomeric repeats and adjacent telomere-associated (TAS) DNA were determined. The *T.parva* telomeric repeat sequences, a tandem array of TTTAGGG or TTAGGG interspersed with a few variant copies, showed a high degree of sequence identity to those of the photosynthetic algae *Chlamydomonas reinhardtii* (97% identity) and *Chlorella vulgaris* (87.7% identity) and the angiosperm *Arabidopsis thaliana* (84.4% identity). Unlike most organisms which have been studied, no significant repetitive sequences were found in the nucleotide sequences of TAS DNA located centromere-proximal to the telomeric repeats. Restriction mapping and hybridisation analysis of \(\lambda\)EMBL3 clones containing 16 kilobases of TAS DNA derived from one telomere suggested that they did not contain long regions of repetitive DNA. The cloned TAS DNAs were mapped to *T.parva* Muguga genomic SM fragments 8 and 20, which are located at opposite ends of the largest *T.parva* chromosome. A 126 bp sequence located directly centromere-proximal to the telomeric repeats was 94% identical between the two cloned telomeres. The conserved 126 bp sequence was present on all *T.parva* Muguga telomeric SfiI fragments.

**INTRODUCTION**

Telomeres are the specialised structures at the ends of eukaryotic chromosomes which ensure their stability and allow for the complete replication of their ends (1). Recent experimental data from *Saccharomyces cerevisiae* demonstrates that telomeres are essential for chromosome survival (2). With the exception of some insects (3), chromosomal termini of eukaryotes thus far studied consist of arrays of 6–8 bp tandem repeats which conform to a consensus sequence \(5'T/[A/G]_4G[1-8]3'\) (4,5). The terminal repeats are flanked internally by telomere-associated sequences (TASs) which are usually moderately repetitive, often found at more than one telomere and frequently variable in sequence and distribution between individuals within a species (reviewed in 4,5). In the malaria parasite *Plasmodium falciparum* the TASs contain long arrays of repetitive sequence elements with a conserved organization (6,7). Homologous recombination and other rearrangements in *Plasmodium* TASs are responsible for extensive chromosome size polymorphisms (8,9) and novel restriction fragments are generated in the TAS DNA in association with meiosis in *P.falciparum* (10).

*Theileria parva* is an intracellular protozoan parasite which is classified together with *Plasmodium* in the class Sporozoa. *T.parva* reversibly transforms bovine lymphocytes (reviewed in 11) and causes a usually fatal disease of cattle in sub-Saharan Africa. The availability of an SfiI restriction map for *T.parva* DNA (12) constitutes a powerful tool for genome analysis in this organism.

In this report we describe the isolation of DNA from two *T.parva* telomeres which were located on the SfiI restriction map at opposite ends of the largest *T.parva* chromosome. The *T.parva* telomeric repeat sequences exhibited a very high degree of identity to those of photosynthetic algae. By contrast with *Plasmodium* and most other eukaryotes, the TAS DNA at the cloned telomeres was not internally repetitious, but a short conserved sequence was present centromere-proximal to the telomeric repeats.

**MATERIALS AND METHODS**

Parasite material and DNA preparation

The *T.parva* stocks used were Uganda (13) and Muguga (14). Purification of piroplasms and preparation of piroplasm DNA was as described (15).

Isolation of genomic clones containing *T.parva* telomeric DNA

A library of sheared *T.parva* Uganda DNA fragments was prepared in λgt 11 from purified piroplasm DNA (16). The library was screened with a 28mer oligonucleotide (CCCTGAAC-CCTAAA)\(^2\) corresponding to the telomeric repeats of a cloned *Plasmodium berghei* telomere (17). The 1.74 kb insert from a positive bacteriophage was subcloned into the EcoRI site of pBluescript KS (Stratagene). A construct containing only TAS DNA was generated by deleting the telomeric repeats using a unique PaeR7I site in the insert and the HindIII site in the pBluescript polylinker. A construct containing predominantly

* To whom correspondence should be addressed
telomeric repeat DNA was generated by deletion of the TAS DNA using an exonucleaseIII/mung bean nuclease kit (Stratagene). The constructs were verified by nucleotide sequencing. A library of partially Sau3AI-digested Tparva Muguga piroplasm DNA fragments was constructed in λEMBL3 (18). The λEMBL3 library was screened with the 1.74 kb telomeric insert and two positive clones were selected for analysis. SfiI fragments of 2.0, 3.2 and 8.5 kb were subcloned from λTpetl 7.11 into pBluescript KS.

Southern hybridisation

Restriction enzyme digestions were according to the manufacturer's instructions (New England Biolabs). Agarose gel electrophoresis, Southern blotting onto nylon (Hybond N, Amersham) filters and hybridisation of filters used standard procedures (19). DNA probes were labelled using a Prime-it kit (Stratagene). All filters were washed in 2 x SSC/0.1% SDS at 65°C unless indicated otherwise in the figure legends. Labelling, hybridisation and washing conditions for the fiberhegi telomeric oligonucleotide were as described previously (20).

BalB1 exonuclease treatment of genomic DNA

Tparva piroplasm DNA (14 μg) was digested with 10 U BalB1 exonuclease (Bethesda Research Laboratories), calibrated to digest 400 bp/min. Serial time points were taken at 0, 1, 2 and 3 min. For each time point 2 μg aliquots of BalB1 treated DNA were extracted with phenol–chloroform, ethanol precipitated, digested with EcoRI and size fractionated on 0.8% agarose gels for Southern blot analysis.

Nucleotide sequencing

For sequencing, nested deletions were generated using an exonuclease III/mung bean nuclease kit (Stratagene), in combination with the synthesis of oligonucleotides, based on acquired sequence, as primers for regions which were not covered by the deletions. DNA sequencing by the chain termination method (21) used either the Sequenase™ kit (US Biochemicals) or the ‘fmol™’ DNA sequencing system (Promega). For direct sequencing the ‘fmol’ system was used according to the manufacturer's instructions with 2 μg of uncloned genomic DNA as the template. Both strands of the DNA were sequenced. Sequences were deposited with the GenBank database under accession nos L36963 (1747 bp derived from a Tparva Uganda λgt11 clone) and L36964 (851 bp derived from Tparva Muguga λEMBL3 clone).

Pulsed-field gel electrophoresis (PFGE)

Preparation of high molecular weight Tparva piroplasm DNA embedded in low melting point agarose blocks and digestion of agarose-embedded DNA with the restriction enzyme SfiI was performed as described (20). The Tparva SfiI fragments were separated using contour clamped homogeneous electric field (CHEF) electrophoresis (22) and Tparva chromosome sized DNA molecules were separated using the pulspahor system with point electrodes (Pharmacia). For CHEF electrophoresis, 1.5% agarose gels prepared in 0.5 x TBE were used. The SfiI fragments were separated using a pulse frequency of 10 s for 16 h and 40 s for 2.5 h at 200 V. The Tparva chromosomes were separated in 1% agarose in 1 x TBE using pulses of 900 s for 24 h, 600 s for 24 h, 480 s for 24 h and 400 s for 24 h at 200 V. Size markers were concatemers of bacteriophage λ DNA CI 857 S7 and S.cerevisiae chromosomes (Cambridge Bioscience).

Polymerase chain reaction (PCR) amplification of DNA

Amplification reactions used Taq polymerase as specified by the manufacturer (Promega). Cycling conditions were 1 min at 94°C, 1 min at 55°C and 30 s at 72°C for 30 cycles. PCR products were cloned into the pCRtm11 vector (Invitrogen).

RESULTS

Isolation and validation of a Tparva telomeric DNA clone

A λgt11 clone containing a 1.74 kb Tparva genomic DNA fragment which hybridised strongly to an oligonucleotide derived from a fiberhegi telomeric repeat sequence (17) was isolated, and the insert subcloned into pBluescript KS. The plasmid clone was designated pTpUtel. To verify that pTpUtel contained telomere-specific sequences, the clone was hybridised to EcoRI-digested Tparva Muguga piroplasm DNA which had been treated with BalB1 exonuclease. The probe hybridised to eight EcoRI fragments in untreated Tparva genomic DNA (Fig. 1, lane 1) all of which were susceptible to BalB1 digestion (Fig. 1A, lanes 2–4) indicating that they were located at or near telomeres. The smallest EcoRI fragment recognised by the probe was diffuse (Fig. 1A, lane 1), a feature typical of telomeric restriction fragments due to length heterogeneity in the telomeric repeats (5). The blot was re-hybridised with a Tparva ribosomal RNA gene probe, pTmr5.7, which is located in a chromosome internal position (23). The rDNA gene probe recognised a 5.7 kb EcoRI fragment which did not reduce in size after BalB1 digestion (Fig. 1B). Two derivatives of pTpUtel were generated. The first, pTpUsubtel, contained only TAS DNA centromere-proximal to the single PaeR71 site at base pair 987, the second, pTpUtelrep, contained the telomeric repeat sequences and the centromere-proximal 10 bp.

Nucleotide sequences of telomeric DNAs

The nucleotide sequence of the 1747 bp insert in pTpUtel was determined (GenBank accession no. L36963). 490 bp at the 5' end of the cloned sequence consisted of interspersed tandem repeats of TTTTAGGGG (43 copies), TTTTAGGG (13 copies), TTTTAGG (3 copies) and TTTAGG (3 copies). There was no obvious order in the arrangement of the repeats. These sequences conform to the typical consensus for eukaryotic telomeric repeats (5). A search of the nucleotide sequence databases revealed very high levels of similarity to the telomeric repeat sequences of the photosynthetic algae, Chlamydomonas reinhardtii (97% identity in a 299 bp overlap; reference 24), Chlorella vulgaris (87.7% identity in a 261 bp overlap; GenBank accession no. D26374) and the angiosperm Arabidopsis thaliana (84.8% identity in a 299 bp overlap; reference 25). The telomeric repeats of Pfalciparum were less similar to Tparva (75% identity in a 411 bp overlap; reference 26). The 1250 bp of sequence centromere-proximal to the telomeric repeats was AT (70%) and T (44%) rich. The Ts in the sequence were often non-randomly arranged repetitions of two to six, as indicated by a high positive autocorrelation function for this base. A 22 bp sequence TAGATTAGAGTGGTTTAGAGT (two copies separated by
To isolate more extensive TAS DNA sequences, a XEMBL3 clone was isolated. The restriction maps of the TAS DNA from T. parva and A. tel. Two positively hybridising bacteriophages ATptel 4.12 and XTptel 7.11, containing insert sizes of -16.5 and 13.5 kb, were isolated. The restriction enzyme sites in the TAS DNA were shown in Figure 2, revealed frequent 6 bp recognition restriction enzyme sites. The similarity of the restriction enzyme sites in the TAS DNA contained in XTptel 7.11 and ATptel 7.11, containing insert sizes of -16.5 and 13.5 kb, was the only perfect repeat of 20 bp or more present in the TAS DNA. No perfect repeats were detected in 610 bp of TAS DNA. The diffuse nature of the genomic DNA also demonstrated that the 2.0, 3.2 and 8.5 kb fragments were adjacent in the T. parva genome. The 2.0, 3.2 and 8.5 kb SfiI restriction fragments from XTptel 7.11 (Fig. 2) predominantly recognised single SfiI fragments when hybridised to T. parva genomic DNA (Fig. 3). This result, together with the lack of cross-hybridisation between the 2.0, 3.2 and 8.5 kb fragments (data not shown), indicated the absence of extensive repeat sequences in the DNA. The diffuse nature of the genomic SfiI fragment recognised by the 2.0 kb fragment provided further confirmation of a telomeric location (Fig. 3, lane 1).

**Localisation of cloned telomeric DNA sequences on the SfiI restriction map**

The existence of an SfiI restriction map of the T. parva genome (12) allows cloned T. parva sequences to be mapped by hybridisation to blots of SfiI-digested DNA fractionated by PFGE. The plasmid construct containing only telomeric repeat sequences, pTpUtelrep, hybridised to a minimum of six distinct SfiI fragments in the DNA of the T. parva Muguga stock (Fig. 4A and B, lane 2), which had previously been identified as telomeric (12). Since T. parva Muguga has four chromosomes (12), and several of the telomeric SfiI fragments co-migrate on CHEF gels (20), the data was consistent with the presence of this repeat at all telomeres. The probe hybridised to at least six distinguishable SfiI fragments in the T. parva Uganda stock (Fig. 4A and B, lane 1). Several of the telomeric SfiI fragments were size polymorphic between the two T. parva stocks. The construct, pTpUtelrep, which contained only sequences located centromere-proximal to the telomeric repeats, hybridised strongly to two size-polymorphic variants of SfiI fragment 8 in T. parva Muguga DNA and to a single similar sized fragment in T. parva Uganda DNA. The

![Figure 1](image_url)

**Figure 1.** pTpUtel recognises EcoRI fragments in T. parva genomic DNA which are susceptible to digestion with exonuclease Bal31. T. parva Muguga genomic DNA was incubated for 1–3 min with Bal31 exonuclease, digested with EcoRI and size fractionated. Following Southern transfer the filter was hybridised with pTpUtel (A) and re-hybridised, after stripping, with pTpUtel 5.1, a ribosomal RNA gene probe (B). Lanes 1–4 show T. parva DNA treated with Bal31 for 0, 1, 2 and 3 min, respectively.

![Figure 2](image_url)

**Figure 2.** Physical maps of recombinant λ phage containing telomeric repeat and TAS DNA. A selective restriction map is shown of λEMBL3 recombinant clones λTptel 7.11 (13.5 kb) and λTptel 4.12 (16.5 kb). Restriction enzyme sites are designated as follows: BamHI = B, CiaI = C, EcoRI = E, HinfI = H, KpnI = K; PstI = P; SstI = S; XbaI = X. λTptel 4.12 was analysed only with enzymes EcoRI and SstI. The extent of the telomeric repeat sequences present in λTptel 7.11 is indicated. The location of 2.0, 3.2 and 8.5 kb SstI fragments in λTptel 7.11 which were subcloned and used as probes is also shown.
A conserved sequence located centromere-proximal to the telomeric repeats

To examine whether any sequence identity existed between different telomeres in the TAS DNA centromere-proximal to the telomeric repeats, 600 bp of sequence directly adjacent to the telomeric repeats was compared with the DNA cloned in pTpUtel and that cloned in pTpMte1.2.0. The DNA sequences were aligned using the CLUSTAL programme (28). The sequences in pTpUtel were derived from one telomere of T. parva chromosome 1 (Fig. 4), and the sequences in pTpMte1.2.0 were derived from the opposite telomere of T. parva chromosome 1 (Fig. 5). The only significant conserved region was a 126 bp sequence (shown in the lower panel of Fig. 6) directly centromere-proximal to the telomeric repeats which was 94% identical between the two telomeres. The conserved sequence contained a single copy of the minor repeat TTTAGG and one partial copy, TTAGGG, of the major repeat, but was otherwise unrelated to the telomeric repeats. To test whether the 126 bp sequence was present at other telomeres, it was amplified from pTpUtel using primers ILO 2299 (5'-ACT TCG GGA TTT TCA AAT TG-3), ILO 2359 (5'-CTT AAT GTC CTA AAG CT-3) and cloned. Nucleotide sequencing confirmed that the correct sequence had been amplified. The PCR product hybridised at moderate stringency to all eight T. parva Muguga telomeric SfiI fragments (Fig. 6B and D, upper panel). The apparent hybridisation to an additional non-telomeric SfiI fragment in Figure 6B is not evident in Figure 6D and is probably attributable to persistent partial digestion, which is often observed with SfiI digestion of T. parva DNA (12). The 126 bp PCR product hybridised only to the same eight Bul31-sensitive EcoRI fragments in T. parva Muguga DNA as the pTpUtel probe (data not shown). The results were consistent with the presence of a sequence with a high degree of identity to the conserved 126 bp sequence adjacent or close to the telomeric repeats at all telomeres.

DISCUSSION

Telomeric repeat sequences

The telomeric repeat sequences of T. parva conform to the consensus described for most eukaryotes (5). The TTAGGG repeat unit of T. parva and C. reinhardtii, is the least G+C rich
Telomeric repeat known (24), but the inverse correlation between genomic DNA and telomeric DNA G/C content observed in most organisms (24) does not hold for T.parva, in which the telomeric repeat G/C content is 37.5% but the genomic DNA G/C content is only 30% (29). The T.parva telomeric repeats are strikingly similar to those of two photosynthetic algae and an angiosperm. The sequence identity between the invariant TTATAGGG array of the alga C.reinhardtii and the T.parva telomeric repeats is 97%, the differences arising due to the occurrence of variants of the predominant repeat in T.parva. The similarity between the T.parva and plant telomeric repeat sequences may be an example of convergent evolution. However, evidence exists suggesting that sporozoan protozoa have evolved from photosynthetic algae (reviewed in 30). It is therefore tempting to speculate that a common ancestral telomeric repeat has been conserved in Thelellia and plants. A hypothesis which is supported by the similarity of the telomeric repeats of the algae C.reinhardtii and C.vulgaris and the angiosperm A.thaliana. Extensive sequence conservation over a wide evolutionary distance contrasts with the divergence observed in the telomeres of some other organisms such as relatively closely related species of budding yeasts (31). The extent of the telomeric repeats can vary from 20 bp to 20 kb in different organisms (5). In T.parva, as judged by the size of the smallest genomic restriction fragments recognised by the telomeric probes, the telomeric repeats are between 0.5 and 1 kb in length, which is similar to other lower eukaryotes (1,32).

Telomere-associated sequences

The organisation of the TASs in T.parva is dissimilar to that reported in most organisms. The TASs of yeast, protozoa, insects and vertebrates typically consist of middle repetitive sequences, which are often the location of genomic rearrangements and are polymorphic in their chromosomal distribution (reviewed in 5,33). These can be either short tandemly arranged 'minisatellites' (26,34—37) or moderately repetitive sequences ranging from 350 bp to several kb in length (38—41). Nucleotide sequence analysis did not reveal the presence of similar repeated sequences in T.parva TAS DNA at the two telomeres examined. The hybridisation of the TAS DNA contained in the X.4.12 recombinant primarily to a single Sfil restriction fragment (Fig. 5), the lack of cross-hybridisation between Sall restriction fragments from the λEMBL3 clones and the recognition of single genomic Sall fragments by the sequences subcloned from X.4.12, 7.11, was also consistent with the absence of extensive internal repetition in the DNA centromere-proximal to the telomeric repeat sequences.

In Plasmodium species chromosome size polymorphisms between clones can be detected by PFGE and the major source of this polymorphism is DNA rearrangements in TAS regions (reviewed in 9). In T.parva, chromosome size polymorphisms are difficult to detect due to the existence of only four chromosomes which are relatively large (2.1—3.0 Mb) and similar in size (12), but this study shows that some telomeric Sfil fragments, including fragment 20 from which the T.parva TAS DNA at the two telomeres examined. The hybridisation of the TAS DNA contained in the X.4.12 recombinant primarily to a single Sfil restriction fragment (Fig. 5), the lack of cross-hybridisation between Sall restriction fragments from the λEMBL3 clones and the recognition of single genomic Sall fragments by the sequences subcloned from X.4.12, 7.11, was also consistent with the absence of extensive internal repetition in the DNA centromere-proximal to the telomeric repeat sequences.

In Plasmodium species chromosome size polymorphisms between clones can be detected by PFGE and the major source of this polymorphism is DNA rearrangements in TAS regions (reviewed in 9). In T.parva, chromosome size polymorphisms are difficult to detect due to the existence of only four chromosomes which are relatively large (2.1—3.0 Mb) and similar in size (12), but this study shows that some telomeric Sfil fragments, including fragment 20 from which the T.parva λEMBL3 clones were derived, are size polymorphic between stocks (Fig. 4A and B). Blocks of tandemly repeated sequences located between 10 and 45 kb from most telomeres have been shown to be involved in the
generation of chromosome size polymorphisms in *P.falciparum* (8). Preliminary experiments in which *T.parva* Muguga telomeric SfiI fragments 20 (110 kb) and 9 (294 kb) were gel-purified and hybridised to blots of SfiI-digested *T.parva* DNA revealed significant hybridisation only to the homologous SfiI fragment (Baljinder Sohanpal and Richard Bishop, unpublished), suggesting the absence of major tracts of cross-hybridising repetitive sequences upstream of these telomeres.

It has been reported recently that repetitive DNA sequences with a conserved organisation extending over many kilobases are common to most *P.falciparum* telomeres (6,7). In *T.parva* Muguga we have shown that a conserved 126 bp sequence element is present centromere-proximal to the telomeric repeats. The hybridization of the pTpUsubtel probe (which does not contain the 126 bp sequence) with a subset of telomeric SfiI fragments indicates that additional TASs are shared between different telomeres. It is speculated that such homologies in the TASs may facilitate recombination between the telomeres of non-homologous chromosomes (5,6,33). However, the hybridisation of the XTptel 4.12 clone only to SfiI fragment 20 suggests that, by contrast with *P.falciparum*, such homologies are of limited extent in *T.parva*.

The scarcity of repeat sequences in the TASs suggests the possibility that genes may be located close to telomeres in *T.parva*, whose genome is only $10^7$ bp in size (20,29). The data also imply that the size polymorphism observed in *T.parva* telomeric SfiI fragments (Fig. 4; reference 20) is not caused only by rearrangements in repetitive DNA sequences. The cloned sequences described here will allow studies on the molecular basis of telomeric polymorphism in *T.parva* and investigation of whether protein coding genes are located in the TASs.

ACKNOWLEDGEMENTS

We thank Mr J. Kiarie and Mr T. Ndolo for technical assistance. We are indebted to Drs V. Nene and N. Murphy for their critical reading of and suggestions on this manuscript. We are grateful to Mr M. Kibe for the use of the AEMBL3 *T.parva* genomic DNA library. This is ILRAD publication no. 1384

REFERENCES