The EMBO Journal vol.13 no.21 pp.5086–5098, 1994

Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*

Julius Lukeš[1], Gert Jan Arts, Janny van den Burg, Annett de Haan, Fred Opperdoes[2], Paul Sloof and Rob Benne[3]

E.C.Slater Institute, University of Amsterdam, Academic Medical Centre, Meibergdreef 15, 1105 AZ, Amsterdam, The Netherlands and [2]International Institute of Cellular and Molecular Pathology, Avenue Hippocrate, 74-75, B-1200 Brussels, Belgium
[1]Present address: Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, Branišovská 31, 37005, Ceské Budéjovice, Czech Republic
[3]Corresponding author

Communicated by R.Benne

In mitochondria of Kinetoplastida belonging to the suborder Trypanosomatina, the nucleotide sequence of transcripts is post-transcriptionally edited via insertion and deletion of uridylate residues. In order to shed more light on the evolutionary history of this process we have searched for editing in mitochondrial RNAs of *Trypanoplasma borreli*, an organism belonging to the suborder Bodonina. We have cloned and sequenced a 5.3 kb fragment derived from a 37 kb mitochondrial DNA molecule which does not appear to be a part of a network structure and have found genes encoding cytochrome *c* oxidase (*cox*) subunit 1, *cox 2* and apocytochrome (*cyt*) *b*, and genes encoding the small and large subunit mitoribosomal RNAs. The order in which these genes occur is completely different from that of trypanosomatid maxicircle genes. The 5′ and 3′ termini of both the *cytb* and *cox1* gene are cryptic, the protein coding sequences being created by extensive insertion/deletion of Us in the corresponding mRNA sections. Phylogenetic analyses of the protein and ribosomal RNA sequences demonstrated that the separation between *T.borreli* and Trypanosomatina was an early event, implying that U-insertion/deletion processes are ancient. Different patterns of editing have persisted in different lineages, however, since editing of *cox1* RNA and of relatively small 3′-terminal RNA sections is not found in trypanosomatids. In contrast, *cox2* RNA which is edited in trypanosomatids by the insertion of four Us, is unedited in *T.borreli*.

Key words: evolution/kinetoplast/mitochondrion/RNA editing/trypanosomes

Introduction

RNA editing in trypanosomatid mitochondria is a post-transcriptional process which modifies the nucleotide sequence of transcripts via insertion and deletion of uridylate residues (for recent reviews, see Hajduk *et al.*, 1993; Simpson *et al.*, 1993; Stuart, 1993; Benne, 1994). Small guide (g)RNAs, which are partly complementary to the edited sequence if G:U basepairing is allowed (Blum *et al.*, 1990), provide the information for this remarkable form of RNA processing, which is essential for the production of functional mitochondrial (mt) mRNAs. It has been hypothesized that the 3′ oligo(U) extension of the gRNAs is involved in the U-sequence alteration of the mRNAs via one step transesterification reactions in analogy with splicing or, alternatively, two step 'cut and paste' processes mediated by (an) endonuclease(s) and RNA ligase (Blum *et al.*, 1991; Cech, 1991; Harris and Hajduk, 1992; Harris *et al.*, 1992; Koslowsky *et al.*, 1992; Simpson *et al.*, 1992; Sollner-Webb, 1992; Arts *et al.*, 1993). An efficient *in vitro* RNA editing assay system that could establish the mechanistic characteristics of the editing process and help to decide between these and other options is lacking however. The relation (if any) of trypanosome editing to other types of post-transcriptional mRNA sequence alteration for which the term RNA editing has been employed, such as the insertion of (mostly) Cs in mitochondrial RNAs of *Physarum polycephalum* (reviewed in Miller *et al.*, 1993), pyrimidine interconversions in plant organellar transcripts (reviewed in Gray and Covello, 1993) and the editing of mammalian apolipoprotein B and glutamate receptor RNAs (Sommer *et al.*, 1991; Hodges and Scott, 1992), is unknown.

It is also unknown whether RNA editing processes are ancient or recently acquired traits. So far, the U-insertion/deletion type of editing has been found in kinetoplastids belonging to the suborder Trypanosomatina (for trypanosome taxonomy, see Lumsden and Evans, 1976). Extensive editing over the entire length of the mRNAs ('panediting', Simpson and Shaw, 1989) is found in African trypanosomes such as *Trypanosoma brucei* (Feagin *et al.*, 1988, reviewed by Stuart, 1993) and *T.congolense* (Read *et al.*, 1993), in American trypanosomes such as *T.cruzi* (Maslov *et al.*, 1994) and in four monogenetic insect *Herpetomonas* species (Landweber and Gilbert, 1993; Maslov *et al.*, 1994). Smaller RNA sections are edited in *Leishmania, Crithidia* and *Blastocrithidia* species (Shaw *et al.*, 1988; Van Der Spek *et al.*, 1988, 1990; Maslov *et al.*, 1994), although a few transcripts in *L.tarentolae* and *C.fasciculata* are panedited (Maslov *et al.*, 1992, 1994). The exact evolutionary distance between different trypanosomatids is unknown but from phylogenetic analysis of the cytoplasmic small and large subunit rRNA sequences (Fernandes *et al.*, 1993; Landweber and Gilbert, 1994; Maslov *et al.*, 1994) it was concluded that the African trypanosomes represent the earliest trypanosomatid lineage, supporting the notion that extensive editing is the ancient primitive state. The possible mechanistic similarities between editing and splicing have been interpreted as a vestige of a common evolutionary origin of these processes (Blum *et al.*, 1991; Cech, 1991; reviewed in Benne, 1992,

1993). This could imply that RNA editing is very old indeed, perhaps dating back to prebiotic times in which it could have functioned as a primordial form of RNA synthesis (Benne, 1990). Alternatively, however, it has been proposed that RNA editing is a more recent acquisition which arose as a mechanism for the correction of genomic mutations (Covello and Gray, 1993). In this scheme, similarities in the mechanism between editing and splicing are the result of converging evolution by molecular determinism (see Weiner, 1993), rather than of a common evolutionary origin.

In order to explore further the evolutionary history of the trypanosome type of RNA editing, we have initiated the analysis of mitochondrial gene expression in a more distant kinetoplastid, *Trypanoplasma borreli*, which parasitizes fish species to which it is transmitted by a leech vector (Pecková and Lom, 1990). The organism belongs to the family Cryptobiidae of the suborder Bodonina characterized by the presence of two flagellae. The taxonomic status of *T.borreli* as a kinetoplastid has recently been confirmed by the existence of glycosomes (Opperdoes *et al.*, 1988), the presence of mini-exon genes in its nuclear DNA (Maslov *et al.*, 1993) and by kinetoplastid phylogeny analysis based on nuclear rRNA sequences (Maslov *et al.*, 1994). However, its different morphological characteristics and life-cycle, the relatively low degree of conservation of the mini-exon, nuclear rRNA and glycosomal protein gene sequences together with the observation that other Bodonina species do not possess the network of catenated maxi- and minicircles characteristic of the kDNA of Trypanosomatina (Hajduk *et al.*, 1986 and references therein) suggest a more distant relation to Trypanosomatina species.

In this report we describe the analysis of the organization and expression of a 5.3 kb mitochondrial DNA fragment from *T.borreli*. We find a novel gene order and diverged mitoribosomal RNA and protein gene sequences as strong evidence for an early separation of *T.borreli* from the kinetoplast lineage and, most importantly, two mRNAs edited in 5′ and 3′ terminal regions. Our results, therefore, support the hypothesis that RNA editing is an ancient process.

Results

Nucleotide sequence and gene content of a 5.3 kb fragment of T.borreli mt DNA

A 5293 nucleotide DNA fragment obtained by a partial *Sau*3A digestion of total DNA from *T.borreli* was cloned and sequenced as described in 'Materials and methods'. The complete sequence has been deposited in GenBank under accession number U11682. A check for the presence of sequences with similarity to those of mt genes in other organisms resulted in the identification of genes encoding subunits 1 and 2 of cytochrome *c* oxidase (cox) and apocytochrome (cyt) *b* with 65–68, 53–56 and 67–69% identity at the amino acid level, respectively, to the corresponding sequences from trypanosomatids (see Table I, Figure 1A). Furthermore, regions were found encoding abundant RNAs of ~600 and 1150 nucleotides which most likely correspond to the small and large subunit ribosomal (r)RNAs of 9S and 12S, respectively (Sloof *et al.*, 1985; for details see below). These results establish the mitochondrial origin of the DNA fragment analysed. As shown by Northern blot analysis (Figure 1B), all putative genes are transcribed into abundant transcripts.

The characterization via Field Inversion Gel Electrophoresis (FIGE, Carle *et al.*, 1986) and Southern blot analysis of the DNA from which the sequenced fragment was derived, is shown in Figure 2. Two DNA bands hybridized to the *cox1* probe used: a minor band of 50 kb and a major band of 37 kb (lane 1). The upper band disappeared upon digestion with *Eco*RI upon relatively short periods of incubation (lane 2), the lower one requires longer digestion times for complete disappearance (lane 3). Complete digestion by the enzyme gave rise to a 7 kb fragment, whereas products resulting from partial digestion could also be observed (lanes 2 and 3). From this result and the fact that under the same FIGE conditions linearized plasmid DNA migrates slightly ahead of the open circular form (results not shown) we conclude that the lower and upper bands of lane 1 correspond to linear and (non-catenated) circular versions, respectively, of a 37 kb *T.borreli* mt DNA molecule. Some of the hybridizing DNA remained in the wells of the gel, but extensive electron microscopical analysis of total and mitochondrial DNA of *T.borreli* isolated by a number of different procedures failed to provide evidence for the existence of a mt DNA network under conditions which routinely reveal the presence of networks of maxicircles and minicircles in *T.brucei, C.fasciculata* and *Phytomonas sp.* (mt) DNA (results not shown). Although a few linear and circular DNA molecules of a length approximately corresponding to 37 kb were observed, *T.borreli* mt DNA appeared to be surprisingly heterogeneous in length and large numbers of molecules of the same size that could represent the equivalent of trypanosomatid minicircles appeared to be absent. The identification of additional *T.borreli* kDNA molecules, therefore, must await further analysis.

T.borreli cox2 RNA is unedited

In all trypanosomatids analysed thus far a translatable *cox2* mRNA is generated by the insertion of four uridylate residues which remove a gene-encoded frameshift. The translation of the *T.borreli cox2* gene sequences as shown in Figure 3A, however, reveals that the protein is encoded in one continuous reading frame with complete conservation of the amino acids whose codons are created by editing in the frameshift region of trypanosomatid *cox2* mRNAs. This suggested that *T.borreli cox2* mRNA is unedited, which was confirmed by direct reverse transcriptase mediated sequence analysis of *T.borreli cox2* mRNA. All RNA sequences including those around the trypanosomatid editing region shown in Figure 3B are identical to those encoded by the gene.

The 5′ and 3′ regions of cox1 and cytb RNA are edited

As schematically indicated in Figure 1, the *T.borreli* genes for cox1 and cytb appear to lack information at both ends. When aligned with the corresponding trypanosomatid genes 48 (*cox1*) or 24 (*cytb*) amino acids at the N-terminus and 104 (cox1) or 98 (cytb) amino acids at the C-terminus of the *T.borreli* sequences do not seem to be encoded by the DNA, since the percentage of identity of these sections of the trypanosomatid sequences is around background

Table I. Sequence comparison of inferred protein sequences

	T.bor	*L.tar*	*Crit*[a]	*T.bru*	*P.tet*	*S.cer*
cox1/cox2						
T.borreli	100	54.3	53.1	56.2	29.6	24.4
L.tarentolae	65.4	100	90.0	81.0	26.0	22.0
Crithidia[a]	67.8	85.0	100	79.4	26.4	20.6
T.brucei	67.2	80.5	82.9	100	29.5	22.4
P.tetraurelia	33.2	34.7	36.6	34.4	100	21.9
S.cerevisiae	41.4	40.9	47.0	39.8	37.6	100
cytb						
T.borreli	100					
L.tarentolae	66.8	100				
C.oncopelti	68.5	85.0	100			
T.brucei	67.5	85.9	82.9	100		
P.tetraurelia	20.6	23.8	25.5	23.3	100	
S.cerevisiae	24.7	22.6	22.4	22.8	22.2	100

The numbers given represent % amino acid identity, according to the GCG Bestfit program. The *cox1* and *cytb* data are given below the diagonal, *cox2* data are given above the diagonal.
[a]For *cox1 C.oncopelti* has been used, for *cox2 C.fasciculata.*

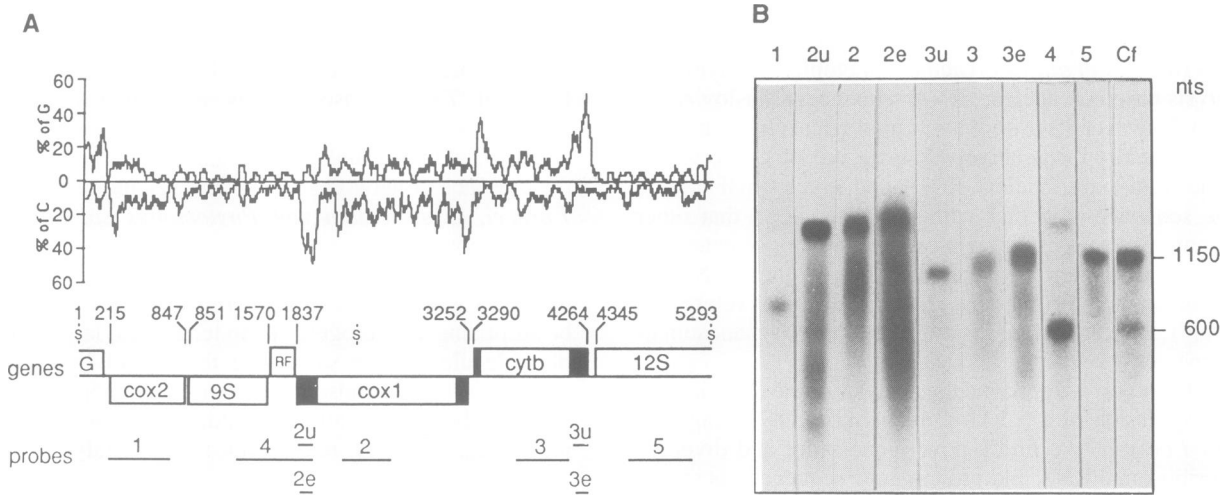

Fig. 1. Analysis of the 5.3 kb *T.borreli* mt DNA fragment. (**A**) Genomic map. The coordinates of the genes found and the position of the *Sau*3A sites (S) are indicated. The black areas of the *cox1* and *cytb* genes are cryptic, coding sequences being created by extensive editing of the corresponding RNA segments. The coordinates of the protein coding genes were derived from the amino acid sequences encoded by the 5.3 kb fragment and the cDNAs; for a given gene the A of the inferred translational initial codon is nucleotide #1, the last nucleotide of the stop codon (A or G) is the last nucleotide of the gene. The coordinates of the 9S and 12S rRNA genes have been inferred from the position of six universally conserved rRNA sequence motifs and the results of an RNase H experiment (see Figure 6). The direction of transcription is from left to right for the genes above the line and from right to left for the genes below. The upper part of the figure shows the G- and C-content of the left (5′) to right (3′) DNA strand. Abbreviations: cox, cytochrome *c* oxidase; cytb, apocytochrome *b*; 9S and 12S are the small and large subunit rRNAs, respectively; RF, a region containing a 68 codon open reading frame; G, a segment of a gene which most likely encodes an edited RNA, as judged by the high G-content of the coding strand. The approximate location of DNA probes used in the Northern blot analysis of Figure 1B is also indicated. The following probes were made by PCR amplification of cloned genomic DNA: 1, with oligonucleotides H23 and H25, for coordinates see Table II; 2, with C113 and C115; 2u (unedited), with H26 and H11; 3, with H28 and H18; 3u, with H29 and H31; 4, with H15 and H11, due to the location of the downstream primer (H11) this probe also hybridized to unedited *cox1* RNA, see lane 2u; 5, with H30 and H3. Probes 2e (edited) and 3e were gel purified inserts from cDNA clones of edited 3′ terminal *cox1* and *cytb* RNA sections, respectively. (**B**) Northern analysis. Northern blots with total *T.borreli* RNA were prepared as described in Materials and methods. Lanes are indicated by the probe used, see (A). In lane Cf., *C.fasciculata* RNA was used with a DNA fragment containing the 9S and 12S rRNA genes as a probe (Sloof *et al.*, 1985).

levels. The most likely explanation for these observations is that the 5′ and 3′ sections of *cox1* and *cytb* RNA are altered by editing. The relatively high GC-content of the DNA regions in question (see Figure 1) which in trypanosomatids is the hallmark of all cryptic DNAs (i.e. DNAs that are transcribed into RNAs that are edited, see Simpson and Shaw, 1989) is in support of this prediction.

This was further investigated with the aid of *cox1* and *cytb* RNA sequence analysis, the 5′ end sequence of these RNAs being determined via direct primer extension analysis, whereas the 3′ sections were sequenced with the aid of PCR-mediated cDNA cloning. The results show clearly that these RNAs are indeed edited by U-insertion/deletion at both ends. At the *cox1* RNA 5′ end (approximately) 73 Us are inserted and 12 Us deleted (Figure 4A) resulting in an RNA sequence encoding the N-terminus of the cox1 protein with a high degree of identity to the corresponding section of trypanosomatid cox1. The sequences obtained for the 3′ section of the 5′ editing region of *cox1* RNA are clear, comparable for example with those of *cox2* RNA (see Figure 3B), indicative of the fact that the majority of the *cox1* RNAs are edited in

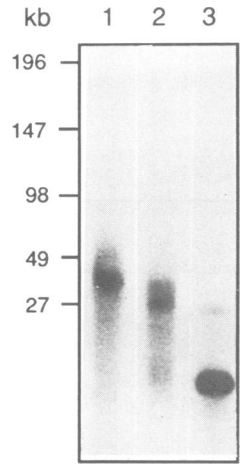

Fig. 2. Field inversion gel electrophoresis was carried out as described in Materials and methods. In lane 1, 5 μg of *T.borreli* DNA was applied; in lanes 2 and 3, 5 μg of *T.borreli* DNA digested for 1 and 18 h, respectively, with 25 U of *Eco*RI. Southern blot filters were probed with a 127 nucleotide *cox1* PCR fragment, (probe 2u, see legend to Figure 1). When other PCR fragments or the cloned 5.3 kb DNA fragment were used as a probe, the hybridization pattern with unrestricted DNA is essentially the same as that of lane 1. (Complete) digestion by *Eco*RI, however, resulted in (an) additional band(s) with the 5.3 kb probe, due to the presence of an *Eco*RI site in the *cytb* gene (results not shown).

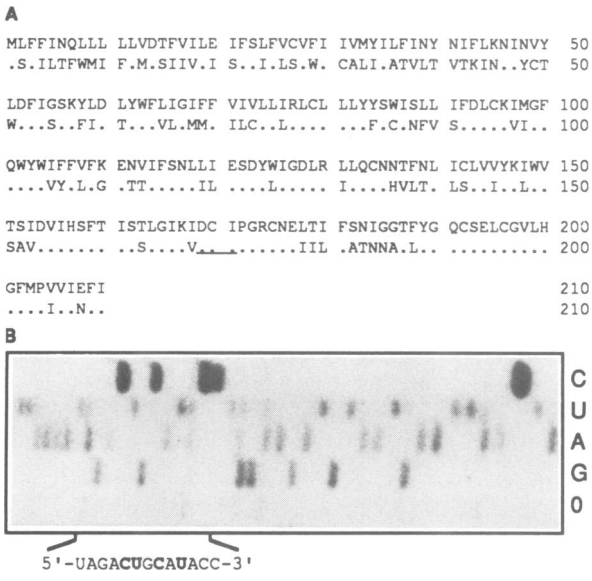

Fig. 3. Analysis of *T.borreli cox2* RNA. (**A**) The inferred amino acid sequence of the *T.borreli* cox2 protein (upper lines), as compared with that of *T.brucei* (lower lines). Identical amino acids are indicated by a dot. The three amino acids (DCI) that are derived from the edited segment of *T.brucei cox2* RNA are underlined. (**B**) Sequence of *T.borreli cox2* RNA. The sequence of *cox2* RNA was determined by primer extension with oligonucleotides H34 and H9 (see Table II), the section corresponding to the *T.brucei* frameshift region is given. The RNA sequence is completely colinear with the genomic sequence, four nucleotides that correspond to the inserted Us in trypansomatid *cox2* are in bold.

this section. The quality of the sequences deteriorates in the 3′ to 5′ direction, however, even if 'edited' oligonucleotide primers are used at close range, suggesting that fully edited RNAs become less abundant. There is a clear consensus sequence, nevertheless, all the way up to A_{12} (Figure 4A), but beyond this nucleotide the sequences are no longer readable with predominant signals in the G and U lanes and minor signals in the A lane. They are not aspecific, however, given the absence of signals in the C-lane and the control lane and they most likely indicate the presence of a large number of different, mostly incompletely, edited RNAs. As a result an AUG translational start codon at the position of the trypanosomatid *cox1* initiation codon is lacking, but at a number of other positions in frame AUG codons can be inferred to be present in a minor fraction of the RNAs (indicated in Figure 4A), all of which would produce slightly shorter versions of the protein. For the majority of the transcripts the AUG codon is out of phase or absent altogether, however (see also Discussion).

The consensus sequence of the edited section at the 3′ end of *cox1* RNA (177 insertions, 13 deletions), as derived from the analysis of 20 cDNA clones is shown in Figure 4B. Like the *cox1* RNA 5′ end, the 3′ end edited *cox1* RNA encodes a protein section with high similarity to the corresponding trypanosomatid sequence, giving an overall identity of *T.borreli cox1* to trypanosomatid *cox1* of 65–68% (Table I). cDNAs with partially edited sequences were found at a high frequency in the analysis. Only with an edited oligonucleotide as a downstream primer could we isolate some completely edited *cox1* cDNAs. Even in this case >80% of the clones contained incompletely 3′ edited cDNA, notwithstanding the fact that cDNA of the size corresponding to fully edited RNA was gel-purified before ligation into the vector. Also the editing of the 3′ region of *cox1* RNA appears, therefore, to be incomplete in most of the transcripts. The characteristics of partially edited *T.borreli cox1* RNAs are very similar to those of partially edited RNAs in trypanosomatids, the edited sequence being invariably found at the 3′ end and differently edited sequences, which differ from both the unedited and the consensus edited sequence, occurring at the junction of edited and unedited sections. Some of these 'junction' sequences are presented in Figure 4B. Last but not least, Figure 4C shows the alignment of the inferred *T.borreli* cox1 protein sequence, as assembled from genomic and consensus edited cDNA sequences, with that of *T.brucei*. The alignment shows that the sequences are colinear, although *T.borreli* cox1 is slightly shorter and heterogeneous at the N-terminus as discussed above. This strongly suggests that no further extensive editing occurs in the remainder of *T.borreli cox1* RNA, in line with the overall low G-content of the coding strand (Figure 1A). We have not determined the complete RNA sequence, however, so the presence of additional small edited regions cannot be formally excluded.

A similar analysis yielding similar results was carried out for *cytb* RNA (Figure 5). The edited RNA sequences encoding the 'missing' N- and C-terminal parts of the protein are created by the insertion/deletion of 42/2 and 144/40 Us respectively, producing a sequence with an overall identity to trypanosomatid *cytb* of 67–69%. As for *cox1* RNA, the sequence at the extreme 5′ terminus of the 5′ editing region of *cytb* RNA was heterogeneous, resulting in two possible locations of the translational start codon (indicated in Figure 5A), but also in this case an in-frame start codon appeared to be present in only a minor fraction of the transcripts. Figure 5B presents the consensus sequence of the 3′ editing region as determined

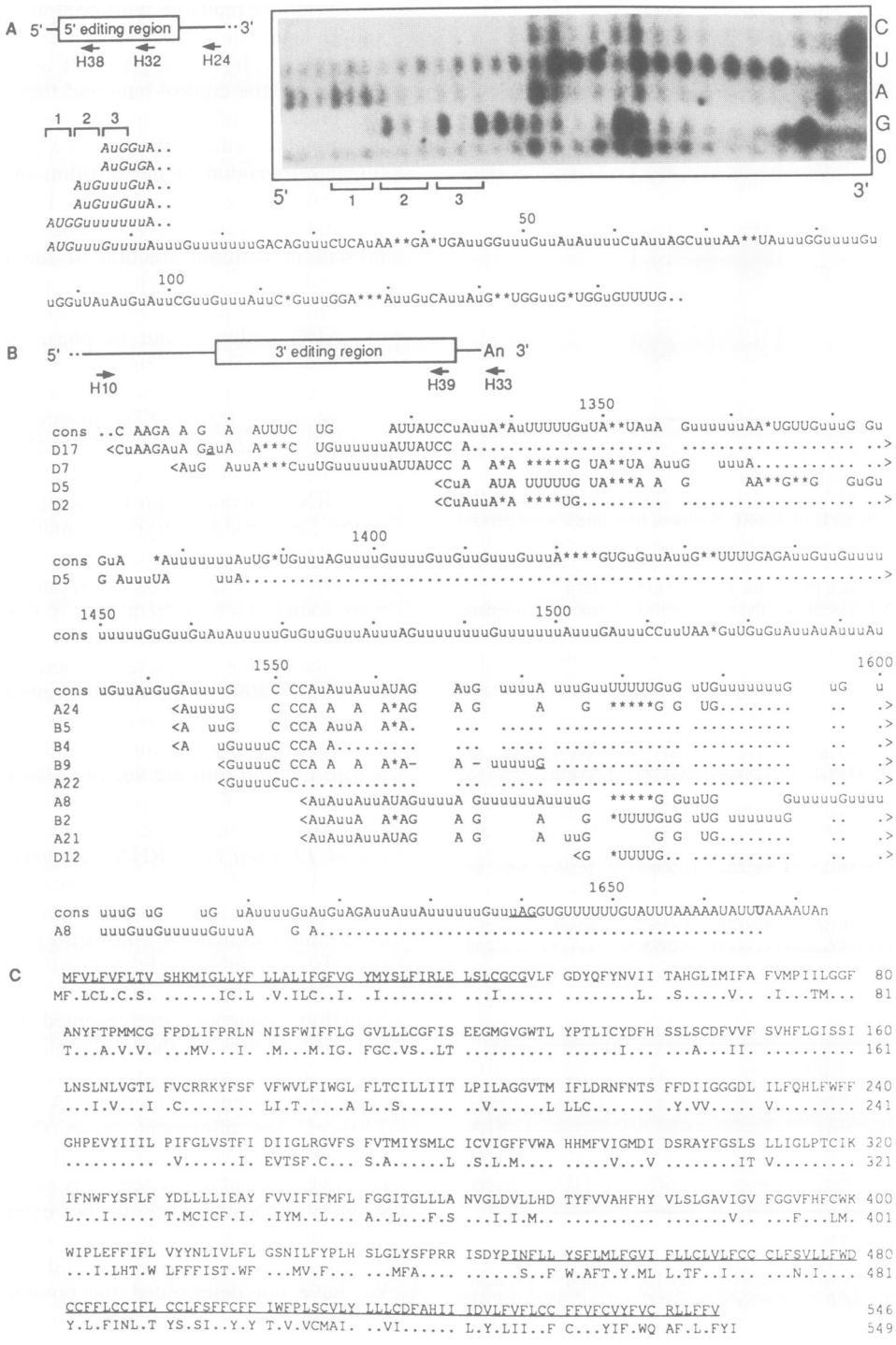

Fig. 4. Analysis of *T.borreli cox1* RNA. (**A**) Nucleotide sequence of the 5′ editing region as obtained from primer extension analyses with oligonucleotide H24, H32 and H38 (see Table II and cartoon). Genomically encoded nucleotides are in upper case lettering; deleted Us are indicated by asterisks, inserted Us are in lower case. At the 5′ side of nucleotide A12, some of the possible (but hypothetical) sequences containing in-frame translational initiation codons are presented in italics; they are inferred from primer extension analysis with oligonucleotide H38 (see insert). The A of AUG #1 is taken as the first nucleotide of the *cox1* cDNA sequence. (**B**) Consensus nucleotide sequence of the 3′ editing region. The sequence was assembled from that of 20 cDNA clones, 16 of which contained partially edited sequences. The cartoon gives the position of the oligonucleotide primers that were used in the RT-PCR mediated cDNA cloning procedures (see Materials and methods). The junction region between the edited and unedited sections of some cDNA clones is given underneath the consensus, 'cons'. Dots to the right of the junction section indicate identity to the consensus. A rightward arrowhead at the end of a line represents the 3′ remainder of a particular cDNA, containing the consensus edited sequence unless indicated otherwise. A leftward arrowhead 5′ of the junctions represents the 5′ unedited sections of the cDNAs in question. Underlined G and A residues and two missing nucleotides in clone B9 are most likely derived from a PCR-related artifact. An alternative polyadenylation site is indicated in bold face. The presumed UAG stop codon is underlined. (**C**) The inferred *T.borreli* cox1 protein sequence (upper lines) aligned with that of *T.brucei* (lower lines). Amino acids derived from edited RNA sections are underlined, only one of the possible N-terminal amino acid sequences is given. Identical amino acids are indicated by a dot.

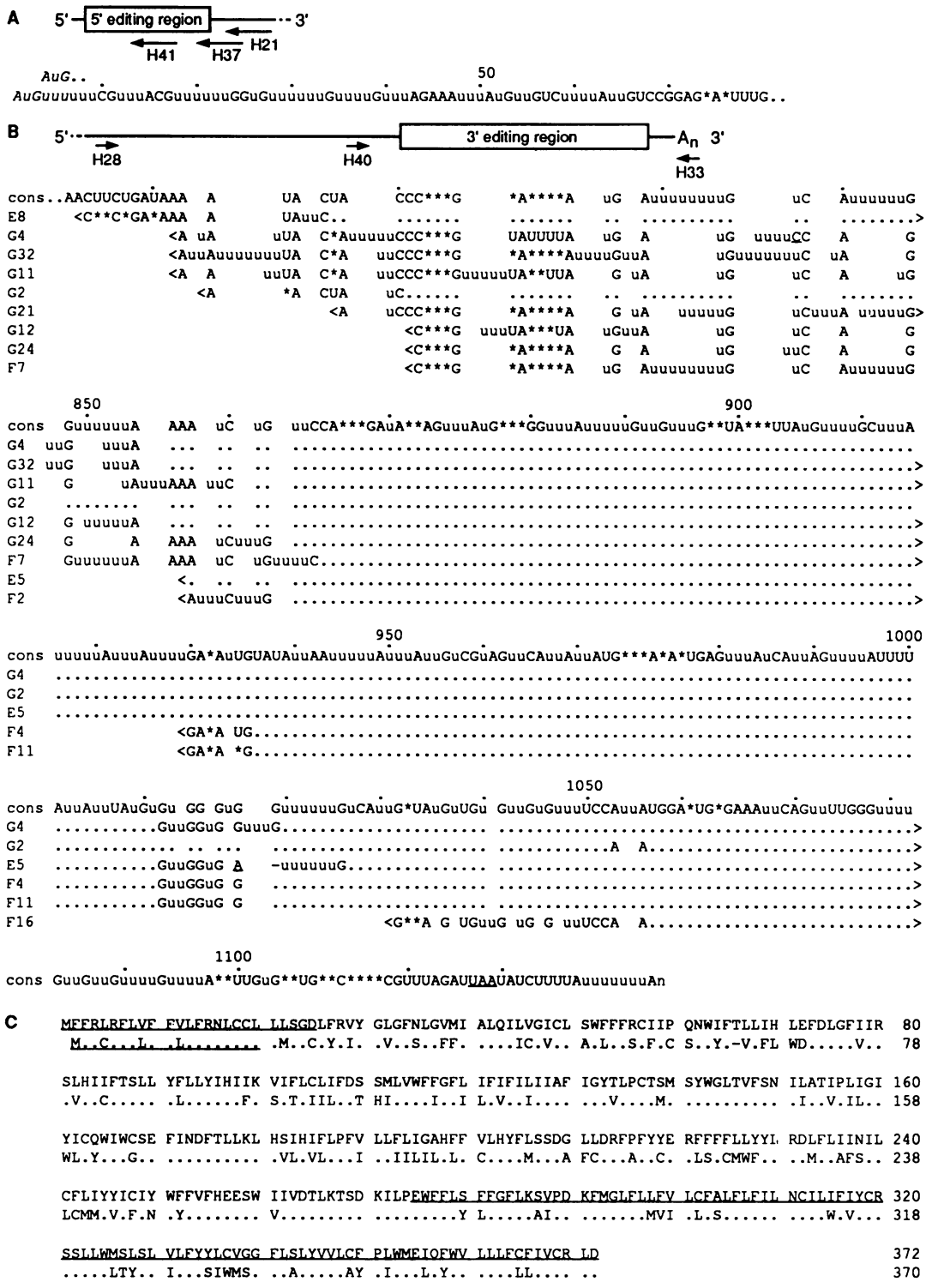

Fig. 5. Analysis of *T.borreli cytb* RNA. (**A**) Nucleotide sequence of the 5′ editing region as obtained from primer extension analyses with oligonucleotides H21, H37 and H41 (see Table II and cartoon). Two of the possible sequences at the extreme 5′ end that would result in in-frame translational initiation codons are given in italics. (**B**) Consensus nucleotide sequences of the 3′ editing region. The sequence was assembled from that of 15 clones, the cartoon gives the position of the oligonucleotide primers that were used in the RT-PCR mediated cDNA cloning procedures. Only one clone contained the complete 'consensus' sequence, in two others (E8 and G2) a small 5′ section was missing. The remainder of the consensus was assembled from three or more independent, different cDNAs. Five of the clones (G2, G4, E5, F4 and F11) display sequences deviating from the consensus in an otherwise fully edited region. For further details, see the text and the legend to Figure 4. (**C**) The inferred *T.borreli* cytb protein sequence (upper lines) aligned with that of *T.brucei* (lower lines). Identical amino acids are indicated by a dot, − indicates a gap. Underlined amino acids are derived from edited RNA sections.

from 15 cDNA clones obtained in the same experimental set up as the *cox1* cDNAs (i.e. cDNA of the size corresponding to fully edited RNA was gel-purified prior to cloning). Although the editing in most of the clones is virtually complete, only one clone contained a sequence encoding a protein that is fully colinear with trypanosomatid *cytb* (the 'consensus' sequence, Figure 5B, 5C). Two clones possessed a small differently edited section of 7–13 nts at the 5′ end of the edited sequence (clones E8, G2), whereas in other cDNAs the differently edited region was somewhat larger. Five cDNAs contain short differently edited stretches in an otherwise fully edited section at a considerable distance from the junction (clones G2, G4, E5, F4, F11, see Figure 5B). Figure 5C shows the alignment of the inferred *T.borreli* (consensus) cytb protein sequence and that of *T.brucei*, which suggests the absence of extensive editing in other sections of *T.borreli cytb* RNA.

We next analysed the frequency of editing of *cox1* and *cytb* RNA by Northern blot analysis (Figure 1B). With probes derived from the middle portion of the genes that recognize all RNA species present [3′ and 5′ (partially) edited and unedited], bands light up (lanes 2 and 3) migrating slightly more slowly than those hybridizing with a 3′ unedited probe (lanes 2u and 3u) but slightly faster than transcripts that hybridize to edited 3′ region probes (lanes 2e and 3e). These results are in complete agreement with those of the sequence analysis and again lead to the conclusion that the majority of the RNAs are edited to some extent in the 3′ part of the two editing regions, but that only a relatively small fraction of the transcripts is fully edited (see also Discussion).

The identification of 9S and 12S rRNA genes

Probes derived from the region flanked by the *cox2* and *cox1* genes and from that to the right of the *cytb* gene (Figure 1A) hybridize to abundant RNAs of ~600 and 1150 nucleotides, respectively (Figure 1B, lanes 4 and 5), which comigrate with small and large subunit mt rRNAs of 9S and 12S from *C.fasciculata* (Figure 1B, lane Cf) (a minor band hybridizing to probe 4 represents unedited *cox1* RNA, due to overlapping sequences in the probe, see the legend to Figure 1). *Trypanosoma borreli* mt DNA probes, however, do not cross-hybridize with the trypanosomatid rRNAs (or vice versa) in line with results from other groups who fail to observe cross-hybridization between trypanosomatid rRNA gene probes and *Bodo caudatus* mt DNA (Hajduk *et al.*, 1986; Fernandes *et al.*, 1993).

A homology search using the GCG Bestfit program revealed that the region downstream of the *cytb* gene possessed a 66–68% identity to the 5′ two-thirds of trypanosomatid large mitoribosomal 12S rRNA genes and showed the presence of three universally conserved large subunit rRNA sequence motifs [Figure 6A, motifs g, e and f (see also De la Cruz *et al.*, 1985b; Sloof *et al.*, 1985)]. Alignment with control sequences of similar size and AT-content such as *T.brucei* variable region segments consistently gave lower values (59–62%) of considerable lower quality, caused by the fact that many more gaps had to be introduced. Surprisingly, a similar search for sequence identity between the sequences derived from the region from which the 600 nt RNA is derived and trypanosomatid 9S rRNA gave only background values.

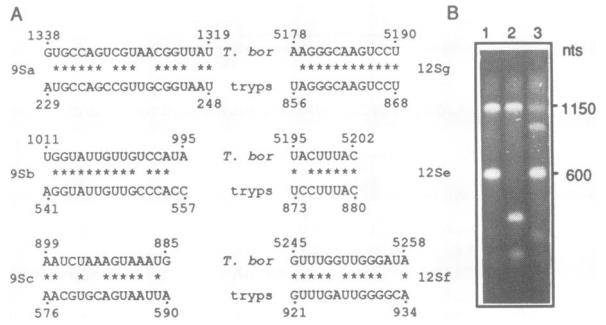

Fig. 6. Identification of 12S and 9S mt rRNAs in *T.borreli*. (**A**) The presence of six universal small and large subunit RNA motifs in *T.borreli* mt DNA sequences. The numbers above the sequences represent the coordinates of the *T.borreli* mt DNA sequence (see Figure 1A) at which these motifs are found, the numbers below the sequences give the approximate position of the motifs in the respective trypanosomatid rRNAs, for details see Sloof *et al.* (1985) and De la Cruz *et al.* (1985a,b). (**B**) RNase H treatment of *T.borreli* mt RNA. The experiment was carried out as described under Materials and methods. Lane 1, control without added oligonucleotide, in lanes 2 and 3 oligonucleotides H15 (9S) and H14 (12S) were added, respectively. The figure shows an ethidium bromide stained agarose gel.

In this region, however, we found universally conserved small subunit rRNA sequence motifs [Figure 6A, motifs a, b and c (see De la Cruz *et al.*, 1985a; Sloof *et al.*, 1985)]. The areas surrounding the conserved rRNA motifs present in the 600 and 1150 nucleotide RNAs could be folded into secondary structures highly characteristic for small and large subunit rRNA, respectively (results not shown, see Sloof *et al.*, 1985). Moreover, in an experiment in which mt RNA from *T.borreli* was included with RNase H and oligonucleotide H15 which is complementary to the putative 9S rRNA ~200 nucleotides downstream of the 5′ end, as based on the gene coordinates of Figure 1A, specific cleavage of an abundant ethidium bromide stainable RNA of 600 nucleotides was observed, generating two smaller fragments of 400 and 200 nucleotides (Figure 6B, lane 2). A similar experiment with oligonucleotide H14, which is complementary to the candidate 12S rRNA also at ~200 nucleotides of the expected 5′ end, resulted in the specific cleavage of an ethidium bromide stainable RNA of 1150 nucleotides generating 900 and 250 nucleotide fragments (Figure 6B, lane 3). Since in trypanosomatids abundant ethidium bromide stainable mt RNAs of ~600 and 1150 nucleotides are the small and large subunit rRNAs (De la Cruz *et al.*, 1985a,b; Sloof *et al.*, 1985), respectively, our data leave little doubt that we have identified the *T.borreli* mt rRNA genes. The high AT-content (Figure 1A) and partial primer extension analysis (results not shown) indicate that as in trypanosomatids (Van der Spek *et al.*, 1990) the rRNAs are not (extensively) edited.

We have also checked the remainder of the DNA fragment for the presence of genes. Downstream of the *cox2* gene, a region is found displaying the high G-content in the coding strand characteristic of panediting. Although we have not analysed this further our prediction is that this region encodes an extensively edited RNA. Analysis of the area sandwiched between the 9S rRNA and the *cox1* gene revealed the presence of a 68 codon open reading frame (including a putative AUG start codon), which encodes a highly hydrophobic protein without

identity to known protein sequences (RF, see Figure 1A). The low GC-content of the DNA sequence of this region suggests the absence of extensive RNA editing, but no transcript derived from this area could be detected (Figure 1B, lane 4). The mRNA encoding the protein may therefore be of low abundance or, alternatively, this region may have another function. Extensive computer analysis, however, revealed that conventional tRNA genes are absent, as are gRNA genes for the edited RNA sections that we have found.

Phylogenetic analysis

Multiple alignments of the inferred cox1, cox2 and cytb amino acid sequences from *T.borreli* with homologous sequences from trypanosomatids and other organisms allowed us to infer phylogenetic trees, typical examples of which are shown in Figure 7. Inclusion of homologous sequences from ciliates and from yeast, which are only distantly related, resulted in the unambiguous determination of the root of the kinetoplastid tree, the construction of which produced similar results for all three proteins irrespective of the specific algorithm that was used (neighbour-joining, least square, protein parsimony, see Materials and methods). It is evident from Figure 7 and Table I that the evolutionary distance between the ciliated and kinetoplastid protists is considerable. This is especially true for cox2 and cytb where the identities between the Kinetoplastida and the ciliates are 30% or less. More importantly, however, our analyses indicate that although Bodonidae and Trypanosomatidae are related with 53–69% identical protein sequences, their separation took place well before trypanosomatid diversification given the much higher degree of identity (79–90%) between the trypanosomatid sequences (see also Benne, 1985; Simpson *et al.*, 1987).

Phylogenetic analyses of the 145 3′ terminal nucleotides of the partial 12S rRNA sequence, which contain the three conserved elements, confirm the conclusions from the mt protein gene sequences (Figure 7). Also, in this case, essentially identical results were obtained using different methods (distance matrix, maximum likelihood, parsimony methods). Our results with the 12S phylogeny are similar to those of Landweber and Gilbert (1994) with the addition of being able to root the trypanosomatid 12S tree by using *T.borreli* as an outgroup. A similar analysis with the putative *T.borreli* 9S sequences was not performed, given the absence of primary sequence identity with trypanosomatid 9S genes.

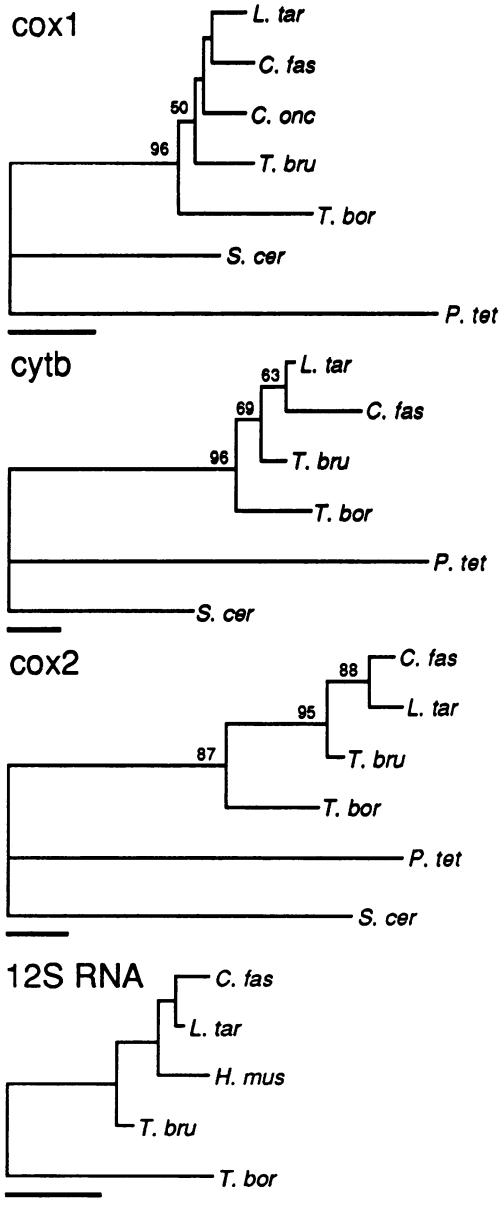

Fig. 7. Kinetoplastid phylogeny was inferred from the mt protein sequences and a segment of the 12S rRNA sequence, corresponding to the 145 3′ terminal nucleotides of the partial *T.borreli* 12S gene sequence (see Figure 1). Evolutionary distances are represented solely by the horizontal components of the tree. Bootstrap values ≥50% are indicated as percentage of 100 resamplings. Bars represent 20 accepted point mutation units for the protein trees and 10 changes per hundred nucleotide positions for the rRNA tree. Abbreviations: *L.tar, L.tarentolae; C.fas, C.fasciculata; C.onc, C.oncopelti; T.bru, T.brucei; T.bor, T.borreli; S.cer, Saccharomyces cerevisiae; P.tet, Paramecium tetraurelia; H.mus, Herpetomonas muscarum muscarum.*

Discussion

Features of T.borreli RNA editing

In this paper we report the cloning, sequencing and transcript analysis of a 5.3 kb mt DNA fragment from the cryptobiid *T.borreli*. The most important result is that two of the three protein genes that were identified are transcribed into RNAs, sizeable sections of which are edited by U-insertion/deletion creating the coding sequence for (approximately) 47 N-terminal and 102 C-terminal amino acids for *cox1* RNA (Figure 4) and ~25 N-terminal and 98 C-terminal amino acids for *cytb* RNA (Figure 5). This implies not only that editing is present in the kinetoplast suborder Bodonina but also that editing patterns differ with respect to Trypanosomatina species which without exception contain a completely unedited *cox1* RNA while editing of trypanosomatid *cytb* RNAs is limited to the 5′ end. In contrast *cox2* RNA, which is edited by the insertion of four Us in trypanosomatids, is unedited in *T.borreli* (Figure 3).

Our results indicate that the editing process in *T.borreli* has a number of features in common with that in trypanosomatids: (i) the close spacing of editing sites in large majority located 3′ to a purine, (ii) the numbers of U that are inserted and deleted per site, insertions greatly

outnumbering deletions and (iii) the existence of partially edited RNAs with the edited section invariably present at the 3′ end of the RNA, suggesting that an overall 3′ to 5′ direction is a conserved trait of the editing process (see Hajduk *et al.*, 1993; Simpson *et al.*, 1993; Stuart, 1993; Benne, 1994). Also other aspects are conserved, such as the occurrence of sequences at the junction of edited and unedited sections of *cox1* and *cytb* RNA with editing patterns different from those of the consensus fully edited mRNA sequence. Similar sequences in trypanosomatids have been called 'unexpectedly' edited to reflect the view that they are side products of the editing process produced via 'mis-editing' with the wrong (section of a) gRNA (Sturm and Simpson, 1990; Sturm *et al.*, 1992) or 'incompletely' edited, meaning to indicate that their formation is part of a normal editing process in which the gRNA is progressively realigned with the RNA, regions of low thermodynamic stability being edited first (Koslowsky *et al.*, 1991; Stuart, 1993). The distance between the boundaries of these sequences has been used to infer the length of the guiding section of the corresponding gRNAs (Landweber *et al.*, 1993), which makes it interesting to note that these distances in *T.borreli* (Figures 4B and 5B) are comparable with those observed in panedited RNAs in trypanosomatids. Given all the similarities between the edited RNAs in the two lineages it is to be expected that the general characteristics and mode of action of the gRNAs are similar. We have, however, not yet characterized the *T.borreli* gRNAs in any detail.

From the results of the Northern blots of Figure 1B and the cDNA sequence analyses of Figures 4 and 5 we concluded that a relatively low percentage of the *cox1* and *cytb* RNAs are fully and correctly edited. This could indicate that under our laboratory conditions *T.borreli* has a low requirement for respiratory chain proteins and that as in the different life-cycle phases of *T.brucei* (reviewed by Stuart, 1993) their production is regulated at the level of RNA editing. The frequency with which the RNA regions are edited declines rather steeply in a 3′ to 5′ direction, which could be caused by a low concentration of 5′ gRNAs. To date a clear correlation between the concentration of a particular gRNA and the frequency of editing of the corresponding RNA section is lacking however (reviewed in Benne, 1994). As an alternative it could be envisaged that the editing process in cultured *T.borreli* is unusually sloppy, in which case editing mistakes made by early acting gRNAs block the anchoring of the later acting ones (see Maslov *et al.*, 1992). The high incidence of differently edited sequences in our clones seems to support the latter possibility. The fact that many *cox1* and *cytb* RNAs have editing in the 3′ section of the 5′ editing region while few RNAs have a completely edited 3′ editing region together with the large distance that separates the two editing regions, indicates that they are independently edited and represent separate editing 'domains' (see Simpson *et al.*, 1993), similar to those found in *T.brucei* ND7 (which has two domains, Koslowsky *et al.*, 1990), *L.tarentolae* RPS12 (which has three, Maslov *et al.*, 1992) and the recently described *T.brucei* CR5 RNA which, like the two *T.borreli* RNAs, has a small independently edited 3′-terminal domain (Read *et al.*, 1994).

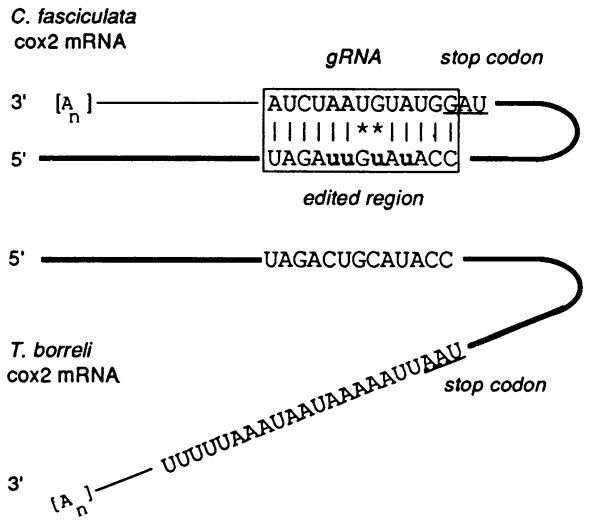

Fig. 8. *T.borreli cox2* mRNA is unedited. The figure shows an alignment of the editing region and a putative gRNA of *C.fasciculata cox2* mRNA with the corresponding sections of *T.borreli cox2* mRNA. Us that are created by editing in the *C.fasciculata* sequence are in bold lower case. *Trypanosoma borreli cox2* RNA is unedited and does not contain the gRNA sequence. Nucleotides that are completely conserved in different trypanosomatids (Kim *et al.*, 1994) are boxed. Protein coding parts of *cox2* RNA are indicated by a thick line.

An interesting aspect of trypanosomatid *cox2* RNA editing is that a putative guide (g)RNA sequence serving as a template for the insertion process is present *in cis* in the 3′ untranslated region of the RNA immediately downstream of the stop codon. Both the sequence and position of this gRNA are highly conserved in four trypanosomatid species (Kim *et al.*, 1994), providing strong evidence for its supposed role. A schematic representation of the relevant sections of *C.fasciculata cox2* RNA is given in Figure 8, conserved complementary mRNA/gRNA sequences being boxed. As expected in an organism which does not edit *cox2* RNA, *T.borreli cox2* RNA sequences downstream from the stop codon show no sign whatsoever of complementarity to the 'frameshift' area and no conservation is found with respect to the *C.fasciculata* gRNA sequence. The *T.borreli* counterpart of the *cox2* mRNA frameshift region, however, is highly conserved differing only by two (silent) U to C mutations at positions at which inserted Us are found in *C.fasciculata* (see also Figure 3B). These observations lend further support to the proposed role as intramolecular gRNA of the 3′ untranslated region of trypanosomatid *cox2* RNA, *T.borreli* serving as a negative control in which both *cox2* editing and the gRNA are absent.

Early divergence of T.borreli from the trypanosomatid lineage

Phylogenetic analyses utilizing nuclear rRNAs (Fernandes *et al.*, 1993; Landweber and Gilbert, 1994; Maslov *et al.*, 1994), mt rRNAs (Lake *et al.*, 1988; Landweber and Gilbert, 1994) and mini-exon and 5S rRNA sequences (Campbell, 1992) have been performed to establish the evolutionary relationship between species belonging to the Trypanosomatina and Bodonina suborder. Most of these studies resulted in trees in which the African trypanosomes branch off early from the trypanosomatid

lineage, the exception being a tree based on mitoribosomal RNA sequences obtained by Lake *et al.* (1988) which suggests a late divergence. The reason for this discrepancy is unclear, but it has been suggested that the results of Lake *et al.* can be explained by the absence of a proper outgroup and that there is in fact no conflict with the other trees (Fernandes *et al.*, 1993). We have included a partial *T.borreli* 12S mt rRNA sequence as an outgroup in the analysis of trypanosomatid mt rRNAs and our results confirm the early divergence of the African trypanosomes well before the other trypanosomatid species (Figure 7). The same tree topology was found when maximum likelihood, distance and parsimony methods were applied, making this tree highly reliable. The trees derived from the mt protein gene sequences fully support this scenario. Also these trees appear to be reliable as can be deduced from the highly significant confidence intervals for the branching order of the trees, as obtained by bootstrap analysis.

The phylogenetic analyses presented in Figure 7 also clearly demonstrated that the separation between *T.borreli* and Trypanosomatina was an early event taking place well before trypanosomatid diversification, in agreement with two other trees which included members of the suborder Bodonina (Fernandes *et al.*, 1993; Maslov *et al.*, 1994). Again, all genes gave very similar results irrespective of the method used. Other lines of evidence lead to a similar conclusion. First, although the size of the 37 kb *T.borreli* mt DNA molecule from which the cloned fragment was derived, is comparable with that of the maxicircle of various trypanosomatids (e.g. 36 kb in *C.fasciculata*, Hoeijmakers *et al.*, 1982), the gene order is completely different (see Hajduk *et al.*, 1993; Simpson *et al.*, 1993; Stuart, 1993; Benne, 1994). For example, in *T.borreli* the 9S and 12S genes are not juxtaposed as in trypanosomatids but instead encoded on opposite strands at a considerable distance from each other, precluding the possibility of coordinate expression of the two rRNAs via the production of a common precursor. Second, inspection of numerous EM pictures of *T.borreli* mt and total DNA failed to provide evidence for the existence of a DNA network, the typical hallmark of trypanosomatid kDNA, under conditions that routinely allow the visualization of such a network in various trypanosomatids. Although the existence of a particularly fragile network cannot be excluded,it is attractive to assume that, like another member of the Bodonina suborder (i.e. *Bodo caudatus*, Hajduk *et al.*, 1986), *T.borreli* does not possess a kDNA network. It has been hypothesized that the function of the kDNA network is to minimize the risk of losing minicircles harbouring essential gRNA genes during mt DNA replication (Borst, 1991). Without a network other mechanisms must have arisen in *T.borreli* to avoid this, for example via the clustering of gRNA genes in the 37 kb DNA itself or in another large molecule that cannot be easily lost. In this perspective the further characterization of *T.borreli* kDNA promises to be interesting.

The evolution of RNA editing

It remains difficult to assess accurately the date of divergence of the two kinetoplastid suborders, in view of the lack of fossil records and precise knowledge of the relative evolutionary rates. Based on small subunit rRNA sequence comparisons, Fernandes *et al.* suggested that the divergence between the Bodonid *Bodo caudatus* and the trypanosomatids approaches that observed between vertebrates and invertebrates or between plants, animals and fungi, species with estimated divergence times of 680 and 900 million years ago, respectively (Fernandes *et al.*, 1993 and references therein). Such data imply that the U-insertion/deletion type of RNA editing is a relatively ancient process, since our results indicate that it was already present in mitochondria of the common ancestor of the two kinetoplastid suborders. Exactly how ancient it is remains uncertain, however, without the demonstration of editing outside kinetoplastids, for example in euglenoids or α-Proteobacteria (Gray, 1994). The finding of editing in *T.borreli cox1* RNA shows that there is no special feature in the kinetoplastid *cox1* gene and RNA sequences that is incompatible *per se* with RNA editing. This finding, combined with the observation that also the editing of *T.borreli cytb* RNA is more extensive than that of trypanosomatids suggests that in ancestral kinetoplastids more (or all) RNAs were extensively edited (see Gray, 1994; Maslov *et al.*, 1994) and that different editing patterns have persisted in different lineages. It has been proposed that the disappearance of editing in certain transcripts has been the result of retrotransposition of edited cDNAs into the mt genome via homologous recombination (Fernandes *et al.*, 1993; Simpson *et al.*, 1993; Landweber and Gilbert, 1993, 1994; Maslov *et al.*, 1994), a view supported by the discovery of a reverse transcriptase activity in *C.fasciculata* (Gabriel and Boeke, 1991). In such a scheme the necessary homology between incoming and target DNA is provided by the untranslated leader and trailer regions of the cDNAs, the leader region not being edited and the trailer region possessing stretches of at least 15 unedited nucleotides at the 3′ terminus in all RNAs analysed. The question why in trypanosomatids a partially edited cDNA still requiring 5′ editing has often been incorporated can be answered by assuming that the frequency of recombination is greatly enhanced by the longer stretches of homology that arise from the presence of 5′ unedited regions. The large 3′ unedited sections present in two *T.borreli* RNAs could be the result of a similar requirement for homology at the other side of the incoming cDNA. It should be realized, however, that partially edited cDNAs in which both the 5′ and the 3′ terminal sections are unedited can only arise if these regions are edited independently of one another, given the fact that editing has an overall 3′ to 5′ direction. As argued above, this indeed appears to be the case for *T.borreli cox1* and *cytb* RNA.

However, it is difficult to explain why retrotransposition processes that have resulted in the incorporation of fully edited cDNAs for a number of RNAs that are no longer edited in present day kinetoplastids (e.g. trypanosomatid *cox1*, *ND4*, *ND5*, etc. and *T.borreli cox2* RNA), would produce a trypanosomatid *cox2* gene that still contains a small internal editing region. There are two possible explanations for this, which are not necessarily mutually exclusive: (i) the ability to regulate the expression of the mt *cox2* gene and/or to produce a truncated version of the cox2 protein at the level of RNA editing provided some evolutionary advantage in the trypanosomatid lineage or (ii), the retrotransposition process with a fully edited *cox2* cDNA was already complete before the divergence

between the trypanosomatids and cryptobiids, but *cox2* editing reappeared in the trypanosomatid lineage as a result of a frameshift mutation in the gene, combined with the creation of an intramolecular gRNA sequence along the lines of the hypothesis of Covello and Gray (1993, see Introduction). Although the latter possibility is not the most parsimonious one since it requires an extra step, it is certainly not unattractive since the other components of the editing machinery were present in the trypanosomatid lineage. On the other hand, as long as data on the production of the cox2 protein (and any mt protein for that matter) are lacking it is difficult to rule out the first explanation. More work is obviously required to shed light on these matters.

Materials and methods

Cell culture, cell fractionation and nucleic acid isolation

Trypanosoma borreli [strain Tt-JH, isolated in 1986 from the blood of infected tench (*Tinca tinca L.*) in South Bohemia, Czech Republic] was cultivated at 15°C in biphasic blood-agar medium as described by Peckoŕa and Lom (1990). Under these conditions the doubling time of the organisms was ~50 h. Mitochondrial vesicles from ~5 × 10^9 cells (10^7 cells/ml) were prepared by forcing the cells through a 26 gauge needle followed by fractionation on a Renographin density gradient (Braly *et al.*, 1974) using, in essence, the method described for the isolation of mt vesicles from *T.brucei* (Feagin *et al.*, 1987). Total cellular (or mt) DNA or RNA was isolated using the hot-phenol extraction method as described (Borst and Fase Fowler, 1979). For RNA isolation, DNA was removed by DNase I digestion, 5 μg/ml, 15 min 37°C, according to Tullis and Rubin (1980). The RNA obtained in virtually all preparations displayed a somewhat higher level of degradation than RNA preparations from *T.brucei* and *C.fasciculata* (e.g. compare Figure 1B, this paper with Figure 1, Van der Spek *et al.*, 1990). Plasmid DNA was prepared according to Birnboim and Doly (1979).

Electron microscopy

Trypanosoma borreli DNA was spread for electron microscopy using the formamide technique (Davis *et al.*, 1971). Spread DNA was stained in 10^{-5} M uranyl acetate for 10 s, rinsed in 90% ethanol and air-dried. After rotary shadowing with platinum/palladium at an angle of 7°, grids were viewed in a Philips EM 400. The magnification was determined using a grating replica.

Electrophoresis, blotting, hybridization and PCR

Standard agarose gel electrophoresis of DNA or RNA, Southern and Northern blotting and hybridization procedures were essentially as described (Sambrook *et al.*, 1989; Van der Spek *et al.*, 1990, 1991). Field inversion gel electrophoresis (FIGE) was performed as described by Carle *et al.* (1986). In a representative experiment, 5 μg of DNA was electrophoresed with an initial pulse length of 3 s increasing by 1 s/h to 20 s (forward 150 V, backward 50 V). Gels obtained from FIGE or normal gel electrophoresis were blotted following treatment of the gel with 0.5 N HCl for 20 min to facilitate the blotting of large DNA molecules.

For PCR amplification of DNA fragments the following protocol was used: 20 ng of recombinant DNA or 50 ng of total *T.borreli* DNA was incubated with 20 pmol of oligonucleotides for 5 min at 95°C after which it was amplified in a Hybaid temperature controller in 30 cycles of 1 min at 95°C, 1.5 min at 45–50°C, followed by 2 min at 74°C. The exact annealing temperature depended on the length and GC-content of the oligonucleotides used in a particular experiment: 4°C per G or C plus 2°C per A or T minus 5°C. For these reactions 1 U of Taq polymerase and buffers were used according to the manufacturer's (Promega) instructions.

For amplification of RNA sections (RT-PCR), the PCR protocol was preceded by cDNA synthesis: 1 μg of *T.borreli* RNA was denatured for 2 min at 70°C, then immediately put on ice and added to the RT mix containing 200 ng of the downstream primer, 9 U of reverse transcriptase, 10 U of RNasin (Promega) and Promega's RT buffer in a total volume of 100 μl. The mixture was incubated for 1 h at 42°C, followed by 5 min at 95°C; 5 μl of this reaction was used for PCR essentially as described above.

Cloning and sequencing of the 5293 nucleotide mtDNA fragment

Two fragments of ~410 nucleotides derived from an internal region of the *T.borreli cox1* gene were prepared by PCR, utilizing three degenerated oligonucleotides (C112, C113 and C115, see Table II) derived from universally highly conserved regions of the *cox1* gene. Following random priming, this fragment hybridized to 2.8 and 5.3 kb DNA fragments on a Southern blot obtained upon a partial *Sau*3A digestion of *T.borreli* DNA (results not shown). DNA present in the 2.5–3.0 kb and 5.0–5.5 kb regions of the gel was excised using the freeze squeeze or low melting agarose method (Sambrook *et al.*, 1989) and cloned into the *Bam*HI site of pUC19. Clones were screened with the *T.borreli cox1* probe and a number of clones containing 2.8 and the 5.3 kb fragments were obtained. The 5.3 kb fragment of one of the clones was sequenced in its entirety using the dideoxy sequencing procedure of Sanger *et al.* (1977). Both strands were sequenced obtaining overlapping sequences from the use of numerous oligonucleotide primers (see Table II) and *Alu*I, *Rsa*I and *Hin*dII subclones of the 5.3 kb fragment (subcloned in pUC19 cut with *Hin*dII). Control Southern blot experiments in which fragments derived from cloned DNA generated by restriction enzymes or PCR were compared with those derived from mt DNA from *T.borreli* showed that no rearrangements had occurred during the cloning procedures. The fragments generated from cloned DNA in all cases had a size identical to the corresponding mt DNA fragment. The sequence has been deposited in GenBank under accession number U 11682.

Sequence determination of RNA and cDNA

The sequence of *cox2*, 9S and 12S RNA sections, the complete 5′ edited region of *cox*1 and *cytb* RNA together with some unedited sections of these RNAs were determined utilizing primer extension sequencing protocols as described in Van der Spek *et al.* (1990), with oligonucleotides listed in Table II and in Figures 3–5. 3′ edited and unedited RNA sequences were obtained via RT-PCR mediated cloning of cDNA fragments, as outlined in Figures 4 and 5. The oligonucleotides were phosphorylated according to Sambrook *et al.* (1989) and PCR fragments were cloned in pUC19 which was restricted with *Hin*cII and dephosphorylated with calf intestine phosphatase. The complete *cox1* and *cytb* cDNA consensus sequences and the inferred amino acid sequences have been deposited in GenBank under accession numbers U11683 and U11684, respectively.

RNase H analysis

0.2 units of RNase H (BRL) was added to 15 μl samples containing 1 μg of mt *T.borreli* RNA in the presence or absence of 0.6 μg of oligonucleotide H15 (9S analysis) or H14 (12S), in 20 mM HEPES−KOH (pH 8.0), 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT And 0.1 mg/ml BSA. After a 20 min incubation at 37°C, the samples were analysed by gel electrophoresis.

Computer manipulations and phylogenetic analysis

Handling of the nucleotide and amino acid sequences for purposes other than phylogenetic analysis was mostly performed with Apple Macintosh computers, utilizing the MacVector 4.1.4 program. Most of the protein sequences were taken from the SwissProt database, only *C.oncopelti* cox1 being taken from the Protein Identification Resource (PIR) database. Multiple alignments of protein sequences were performed using the program Pileup of the GCG package available on the Belgian EMBNet Node (BEN). Percentage identity between pairwise aligned sequences was calculated for the sections common to both sequences in the alignment, not counting deletions and insertions. Multiple alignments of protein sequences were converted to the appropriate input format for the PHYLIP package (version 3.5 by Felsenstein, 1985) using the program ReadSeq by Gilbert (1993). The number of substitutions between amino acid sequences was measured using the program PROTDIST (PHYLIP package) correcting for multiple substitutions by using $d = -\ln(lk)L$ (Kimura, 1983), where K is the proportion of amino acid difference and L the homologous length (i.e. the number of residues that are not a deletion simultaneously in both sequences) of each pair of the sequences compared. For large distances, the Dayhof PAM250 method for the calculation of distances was used. Positions in which gaps are present in any of the aligned sequences were excluded from the analysis.

Based on the distance matrix, phylogenetic trees were inferred by the neighbour-joining method (Saitou and Nei, 1987), implemented in the program NEIGHBOR (PHYLIP package). To contrast the reliability of the tree, the bootstrap method (Efron, 1982; Felsenstein, 1985) was applied. From the original set of sequences 100 bootstrap replicas were obtained for the construction of the corresponding distance matrices and

Table II. Oligonucleotides used

Oligonucleotide (5′–3′)	Coordinates	
DNA and unedited RNA[a]		
C112	TTYTGRTTYTTYGGNCAYCCNGA	2605-2583
C113	TTYGGNCAYCCNGARGTNTAYAT	2596-2574
C115	GCNACNACRAARTANGTRTCRTG	2187-2209
H1	GCATACCTGGAAGATGTAAT	339-320
H2	CCCAAAATACAAAACTAAAATA	2764-2785
H3	GTGGAACAAGTTAGTAATTA	5100-5081
H6	AAAGACATAAAGTTTAAAAG	4872-4853
H7	CTAGGAAAGATTAAATCTGG	3021-3040
H9	TAACCTTAAATCTCCTATTC	455-474
H10	TATAGGAGTATTTGGGGGAG	2153-2134
H11	AGGGAAGGAGACCTAATGTG	1907-1888
H12	CAAGAATCACAAATGTATCC	790-809
H13	GGATTAGGATTTAATTTAGGAG	3340-3361
H14	GAAAGGAATTAAATTTTTAAGTG	4560-4538
H15	ATAACCGTTACGACTGGCAC	1319-1338
H16	CATTAATTAAATTTTTGTGTG	1541-1521
H18	GTTTTCCTGCTCTAAAATAC	4094-4075
H19	GCACCTATTAAAAATAATAG	3839-3820
H20	CCAAAAAATCAAACTAACATACTAG	3602-3578
H21	TAACTCCTAAATTAAATCCTAATCC	3364-3340
H22	CCAGATTTAATCTTTCCTAG	3040-3021
H23	CCATAAAATGTGCCTCCTATATTAG	279-303
H24	GCAAAAATCATTATTAGCCCATGAGCCG	3105-3132
H25	GGATACATTTGTGATTCTTG	809-790
H26	TATTTAATTTTTACATTTTTATTAC	1781-1805
H27	ATACTAATCCAAATATAGGC	2545-2564
H28	TTTATTGGTTATACTCTACC	3637-3656
H29	AATGTTTATATTGAGACAAG	4157-4176
H30	TAATTCCTTTCACATAGCAG	4550-4569
H31	ATTATAAAAGATATTAATCTAAACG	4277-4253
H34	CACACAATTCTGAACACTGCCCATA	259-283
H35	TTAAACTAAACATAAATTAG	1220-1239
H36	TTCTAAAGGTATTCATTTTC	2099-2118
H40	GGAATTCGTATTTCACGAAGAATCTTG	4009-4028
H42	GGCATAATATTATATGTATG	1206-1187
H43	CTTGCTAAATCCCAGTCCGTCTTAC	3183-3207
H44	AATACTCCGGACTGACCTTTTCTCC	3324-3300
Edited RNA[b]		
H32	GACAATTCCAAACGAATAAACAACG (cox1 edN)	125-101
H33	GGAATTCTTTTTTTTTTTTTTTTTTT (dT primer)	
H37	CTAATCCATATACACGAAACAAATC (cytb edN)	94-70
H38	TATAACAAACCAATCATCTTATGAG (cox1 edN)	56-32
H39	AATAATCTACATACAAAATACACAC (cox1 edC)	1628-1604
H41	GACAACATAAATTTCTAAACAAAAC (cytb edN)	55-31

[a]These oligonucleotides have been used in the sequence analysis of the 5.3 kb mt DNA fragment and as primers with unedited sections of RNAs; coordinates are according to Figure 1. C112, C113 and C115 are derived from two regions of the *cox1* gene which are almost completely conserved between trypanosomes, mammals and *S.cerevisiae* (Hensgens *et al.*, 1994), R = A,G; Y = T,C; N = any nucleotide.
[b]These oligonucleotides have been used in the sequence analysis of edited portions of the RNAs and cDNAs; coordinates according to the cDNA sequences of Figures 4 and 5.

phylogenetic trees. The program CONSENSE (PHYLIP package) was used to obtain a consensus tree as well as confidence levels for monophyletic groups. A parallel analysis was carried out in the phylogenetic tree reconstruction step using the least square method (Fitch and Margoliash, 1967) as implemented in the program FITCH (also from the PHYLIP package). Also, directly from the aligned protein sequences a maximal parsimony analysis, using the program PROTPARS (PHYLIP package) was carried out. A discussion of the reliability of the various tree construction methods can be found in Hasegawa and Fujiwara (1993).

Trypanosomatid 12S mt rRNA sequences were taken from GenBank: *T.brucei* 12S (M94286), *C.fasciculata* 12S (X02548) and *Hepetomonas muscarum muscarum* 12S (U01011). The *T.borreli* and trypanosomatid sequences were aligned through the introduction of gaps. The alignment was refined by taking into account phylogenetically conserved higher order structures (Sloof *et al.*, 1985). Sequence similarities in the 3′-terminal part of the 12S RNA for which a secondary structure was inferred by Sloof *et al.* (1985) and comprising the 145 3′ terminal nucleotides of the *T.borreli* sequence were used to infer phylogenetic relationships. The sequences were analysed using distance methods, the maximum likelihood method, as well as parsimony, using DNADIST, FITCH, DNAML and DNAPARS of the PHYLIP package.

All phylogenetic analyses were carried out on a Silicon Graphics R4000 Iris Indigo computer. The alignments of the rRNA and protein sequences are available from the authors by electronic mail (Opperdoes-@trop.ucl.ac.be.).

Acknowledgements

We thank Professor Dr Les Grivell, Dr Hans van der Spek and Dave Speijer for critical reading of the manuscript and valuable suggestions, Marlene van den Berg for performing the EM analysis, Dr Frank van den Berg for his help with the FIGE experiments, Dr Harry van Steeg for his generosity in providing numerous DNA oligonucleotides and Kamiel Zijderveld for his valuable contribution during the initial stages of the work. J.L. has been supported by a short term EMBO fellowship (EE 66) and by a grant from the E.C.Slater Institute; F.O. received financial support from the Belgian Prime Minister's Office Science

Policy Programme, grant no 88/93–122. The research is supported by the Netherlands Foundation for Chemical Research (SON), which is subsidized by the Netherlands Foundation for Scientific Research (NWO).

References

Arts,G.J., Van der Spek,H., Speijer,D., Van den Burg,J., Van Steeg,H., Sloof,P. and Benne,R. (1993) *EMBO J.*, **12**, 1523–1532.

Benne,R. (1985) *Trends Genet.*, **1**, 117–121.

Benne,R. (1990) *Trends Genet.*, **6**, 177–181.

Benne,R. (1992) *Curr. Biol.*, **2**, 425–427.

Benne,R. (1993) *Semin. Cell. Biol.*, **4**, 241–249.

Benne,R. (1994) *Eur. J. Biochem.*, **221**, 9–23.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.

Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.

Blum,B., Sturm,N.R., Simpson,A.M. and Simpson,L. (1991) *Cell*, **65**, 543–550.

Braly,P., Simpson,L. and Kretzer,F. (1974) *J. Protozool.*, **21**, 782–790.

Borst,P. and Fase-Fowler,F. (1979) *Biochim. Biophys. Acta*, **565**, 1–12.

Borst,P. (1991) *Trends Genet.*, **7**, 139–141.

Campbell,D.A. (1992) *Biochem. Biophys. Res. Commun.*, **182**, 1053–1058.

Carle,G.F., Frank,M. and Olson,M.V. (1986) *Science*, **232**, 65–68.

Cech,T.R. (1991) *Cell*, **64**, 667–669.

Covello,P.S. and Gray,M.W. (1993) *Trends Genet.*, **9**, 265–268.

Davis,R.W., Simon,M. and Davidson,N. (1971) *Methods Enzymol.*, **21**, 413–428.

De la Cruz,V.F., Lake,J.A., Simpson,A.M. and Simpson,L. (1985a) *Proc. Natl Acad. Sci. USA*, **82**, 1401–1405.

De la Cruz,V.F., Simpson,A.M. and Simpson,L. (1985b) *Nucleic Acids Res.*, **13**, 2337–2356.

Efron,B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics, Philadelphia.

Feagin,J.E., Abraham,J.M. and Stuart,K. (1988) *Cell*, **53**, 413–422.

Feagin,J.F., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.

Felsenstein,J. (1985) *Evolution*, **39**, 783–791.

Fernandes,A.P., Nelson,K. and Beverley,S.M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11608–11612.

Fitch,M.W. and Margoliash,E. (1967) *Science*, **155**, 279–284.

Gabriel,A. and Boeke,J. (1991) *Proc. Natl Acad. Sci. USA*, **90**, 9794–9798.

Gilbert,D.G. (1993) Readseq. documentation file. Internet gopher: ftp.bio.indiana.edu

Gray,M.W. and Covello,P.S. (1993) *FASEB J.*, **7**, 64–71.

Gray,M.W. (1994) *Nature*, **368**, 288.

Hajduk,S.L., Siqueira,A.M. and Vickerman,K. (1986) *Mol. Cell. Biol.*, **6**, 4372–4378.

Hajduk,S.L., Harris,M.E. and Pollard,V.W. (1993) *FASEB J.*, **7**, 54–63.

Harris,M.E. and Hajduk,S.L. (1992) *Cell*, **68**, 1091–1099.

Harris,M.E., Decker,C., Sollner-Webb,B. and Hajduk,S.L. (1992) *Mol. Cell. Biol.*, **12**, 2591–2598.

Hasegawa,M. and Fujiwara,M. (1993) *Mol. Phylogenet. Evol.*, **2**, 1–5.

Hensgens,L.A.M., Brakenhoff,J., De Vries,B.F., Sloof,P., Tromp,M.C., Van Boom,J.H. and Benne,R. (1984) *Nucleic Acids Res.*, **12**, 7327–7344.

Hodges,P.J. and Scott,J. (1992) *Trends Biochem. Sci.*, **17**, 77–81.

Hoeijmakers,J.H.J., Schoutsen,B. and Borst,P. (1982) *Plasmid*, **7**, 199–209.

Kim,K.S., Texeira,S.M.R., Kirchhoff,L.V. and Donelson,J.E. (1994) *J. Biol. Chem.*, **269**, 1206–1211.

Kimura,M. (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, UK.

Koslowsky,D.J., Bhat,G.J., Perrolaz,A.L., Feagin,J.E. and Stuart,K. (1990) *Cell*, **62**, 901–911.

Koslowsky,D.J., Bhat,G.J., Read,L.K. and Stuart,K. (1991) *Cell*, **67**, 537–546.

Koslowsky,D.J., Göringer,H.U., Morales,T.H. and Stuart,K. (1992) *Nature*, **356**, 807–809.

Lake,J.A., De la Cruz,V., Ferreira,P., Morel,C. and Simpson,L. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4779–4783.

Landweber,L.F. and Gilbert,W. (1993) *Nature*, **363**, 179–182.

Landweber,L.F. and Gilbert,W. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 918–921.

Landweber,L.F., Fiks,A.G. and Gilbert,W. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 9242–9246.

Lumsden,W.H.R. and Evans,D.A. (1976) *Biology of the Kinetoplastida.* Academic Press Inc., London, Vol. I and II.

Maslov,D.A., Sturm,N.R., Niner,B.M., Gruszinsky,E.S., Peris,M. and Simpson,L. (1992) *Mol. Cell. Biol.*, **12**, 56–67.

Maslov,D.A., Elgort,M.G., Wong,S., Peckova,H., Lom,J., Simpson,L. and Campbell,D. (1993) *Mol. Biochem. Parasitol.*, **61**, 127–136.

Maslov,D.A., Avila,H.A., Lake,J.A. and Simpson,L. (1994) *Nature*, **368**, 345–348.

Miller,D., Mahendran,R., Spottswood,M., Ling,M., Wang,S., Yang,N. and Constandy,H. (1993) In Benne,R. (ed.), *RNA Editing, the Alteration of Protein Coding Sequences of RNA.* Ellis Horwood, Chichester, UK, pp. 87–101.

Opperdoes,F.R., Nohynková,E., Van Schaftingen,E., Lambeir,A.-M., Veenhuis,M. and Van Roy,J. (1988) *Mol. Biochem. Parasitol.*, **30**, 171–173.

Pecková,H. and Lom,J. (1990) *Parsitol. Res.*, **76**, 553–558.

Read,L.K., Fish,W.R., Muthiani,A.M. and Stuart,K. (1993) *Nucleic Acids Res.*, **21**, 4073–4078.

Read,L.K., Wilson,K.D., Myler,P.J. and Stuart,K. (1994) *Nucleic Acids Res.*, **22**, 1489–1495.

Saitou,N.M. and Nei,M. (1987) *J. Mol. Evol.*, **4**, 406–426.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Shaw,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell*, **53**, 401–411.

Simpson,L. and Shaw,J.M. (1989) *Cell*, **57**, 355–366.

Simpson,L., Neckelmann,N., De la Cruz,V.F., Simpson,A.M., Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *J. Biol. Chem.*, **262**, 6182–6196.

Simpson,A.M., Bakalara,N. and Simpson,L. (1992) *J. Biol. Chem.*, **267**, 6782–6788.

Simpson,L., Maslov,D. and Blum,B. (1993) In Benne,R. (ed.), *RNA Editing, the Alteration of Protein Coding Sequences of RNA.* Ellis Horwood, Chichester, UK, pp. 53–85.

Sloof,P., Van den Burg,J., Voogd,A., Benne,R., Agostinelli,M., Borst,P., Gutell,R. and Noller,H. (1985) *Nucleic Acids Res.*, **11**, 4171–4190.

Sollner-Webb,B. (1992) *Nature*, **356**, 743–744.

Sommer,B., Kohler,M., Sprengel,R. and Seeburg,P.H. (1991) *Cell*, **67**, 11–19.

Stuart,K. (1993) In Benne,R. (ed.), *RNA Editing, the Alteration of Protein Coding Sequences of RNA.* Ellis Horwood, Chichester, UK, pp. 25–52.

Sturm,N.R. and Simpson,L. (1990) *Cell*, **61**, 871–878.

Sturm,N.R., Maslov,D.A., Blum,B. and Simpson,L. (1992) *Cell*, **70**, 469–476.

Tullis,R.H. and Rubin,H. (1980) *Anal. Biochem.*, **107**, 260–264.

Van der Spek,H., Van den Burg,J., Croiset,A., Van den Broek,M., Sloof,P. and Benne,R. (1988) *EMBO J.*, **7**, 2509–2514.

Van der Spek,H., Speijer,D., Arts,G.J., Van den Burg,J., Van Steeg,H., Sloof,P. and Benne,R. (1990) *EMBO J.*, **9**, 257–262.

Van der Spek,H., Arts,G.J., Zwaal,R.R., Van den Burg,J., Sloof,P. and Benne,R. (1991) *EMBO J.*, **10**, 1217–1224.

Weiner,A.M. (1993) *Cell*, **72**, 161.

Received on July 5, 1994; revised on August 16, 1994