Evolution of the U-Insertion/Deletion RNA Editing in Mitochondria of Kinetoplastid Protozoa

LARRY SIMPSON[a,b,d] AND DMITRI A. MASLOV[c]

[a]Howard Hughes Medical Institute and [b]Departments of Molecular, Cell and Developmental Biology and Medical Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California 90095-1662, USA

[c]Department of Biology, University of California, Riverside, California 92521, USA

R NA editing is a broad term used to describe a variety of unrelated RNA modification phenomena.[1] These phenomena can be divided into two basic types, nucleotide substitution editing and nucleotide insertion/deletion editing. The first type of editing to be discovered was the uridylate (U) insertion/deletion editing that occurs in the mitochondrion of trypanosomatid protozoa,[2] and this remains the most dramatic and unusual editing phenomenon in nature (FIG. 1).[3–6] In this paper we review this type of editing in an evolutionary perspective.

CELL BIOLOGY OF TRYPANOSOMATID PROTOZOA

To understand the mechanism and evolution of this type of editing, we must appreciate the cell biology of the trypanosomatid parasites. These cells contain a single tubular mitochondrion with a single giant network of mitochondrial or "kinetoplast" DNA situated in the matrix attached to the mitochondrial membrane in the region adjacent to the basal body of the flagellum.[7–9] The kDNA network consists of 5–10,000 catenated minicircle molecules and 20–50 catenated maxicircle molecules. The maxicircle DNA contains two rRNA genes, 6 protein-encoding genes with no required editing, 12 cryptogenes, the transcripts of which are modified by U-insertion/deletion editing usually within coding regions, and 15 guide RNA (gRNA) genes, the transcripts of which contain complementary sequences to edited RNAs (allowing G-U base pairs) (FIG. 2). The minicircles encode most of the gRNA genes.[10] This organization of the mitochondrial genome into two separate but interacting genomes is novel, and the evolutionary origin of this split is of great interest.

MECHANISM OF U-INSERTION/DELETION EDITING

A detailed knowledge of the molecular mechanism and biochemistry is essential to provide some constraints for evolutionary speculation on the origin and evolution of trypanosomatid RNA editing. The mechanism involves a cascade of protein enzyme-mediated

[d]Corresponding author: Larry Simpson, Howard Hughes Medical Institute - UCLA, 6780 MacDonald Research Laboratories, 675 Circle Dr. S., Los Angeles, CA 90095. Phone, 310/825-4215; fax, 310/206-8967; e-mail: simpson@hhmi.ucla.edu

```
  1  ---------+---------+---------+-- - -      -     -----+  - -----   46
       CGTTATATAATTTAATATCTATTAAGTATAA A G       A    TGTTTA  A GTTTC
       CGUUAUAUAAUUUAAUAUCUAUUAAGUAUAAuAuGuuuuuAuuuuUG   AuuAuGUUUC
  1  ---------+---------+---------+---------+-------    --+-------- 57
                                        M  F  L  F  L     I  M  F  R
                                        1 -  -  -  -  -  -  -  -  - + 10


 47  -  - - +   - - - - -       -   - - - +-- -------+--         72.
       G  G G G   G G  G G  A       G   G G   G AGA GGTTTTTTTG
       GuuGuGuGuuuGuGuuGuuGuuAuuuuuuuuGuuuGuGuuGuAGAuGG        Guuuuu
 58  --+---------+---------+---------+---------+----       -----+ 110
         C  V  F  V  L  L  L  F  F  C  L  C  C  R  W       V  F
 11  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -       -  -        27


 73  - -    -  --  - -+-------         -   -  + - -          -   93
       A G    G  GA  G AGTTTTTTG         A   G   G G G        A
       uuAuGuuuuGuuGAuuGuAG  UUUUGuuuuuuuuuAuuuGuuuGuGuGuuuuuuuuuuAu
111  ---------+---------+   ---------+---------+---------+-------- 168
       L  C  F  V  D  C  S    F  V  F  F  Y  L  F  V  C  F  F  L  F
 28  -  -  -  -  -  -  -    -  -  -  -  -  -  -  -  -  -  -  -  - 47


 94      ------ +      -        ---  -  --   --+---  -  -  -    --- 119
           GTTTGA G      A      TAA   A GA   TTTTTA  G A  G     AGA
         uuuuuG   UGAuGuuuuuAuuuuuUAAuuuAuGAuuuUUUUUAuuuGuAuuGuuuuAGAuuu
169  -+----    -----+---------+---------+---------+---------+------ 226
         F  V    M  F  L  F  F  N  L  W  F  F  L  L  Y  C  L  D  L
 48  -  -    -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -    66


120  +    - - ---      ----      +     ----------     +     ---- - ---- 149
       A    G A AGA     GCGG      G     GTTTTTTCG     A     GATA A GTTT
       AuuuuGuAuAGAuuuuuuGCGGuuuuuGuuuuG   UUCGuuuuAuuuuGAUAuAuG   U
227  ---+---------+---------+--------    -+---------+---------   + 280
       F  C  I  D  F  C  G  F  C  F  V     R  F  I  L  I  Y  V
 67  -  -  -  -  -  -  -  -  -  -       -  -  -  -  -  -  -    84


150  +---    -   -  -  - -     -+----------+    - -  --    -      175
       TTTG    G   A  A  A G    CGAGTTTTTAAG    G A  GA   G
       UUUGuuuuGuuuAuuAuuAuGuuuuuCGAG    UAAGuuuuGuAuuGAuuuGuuuuuuuu
281  ---------+---------+---------    +---------+---------+------ 336
       L  F  C  L  L  L  C  F  R  V     S  F  V  L  I  C  F  F  L
 85  -  -  -  -  -  -  -  -  -  -       -  -  -  -  -  -  -  -  - 103
```
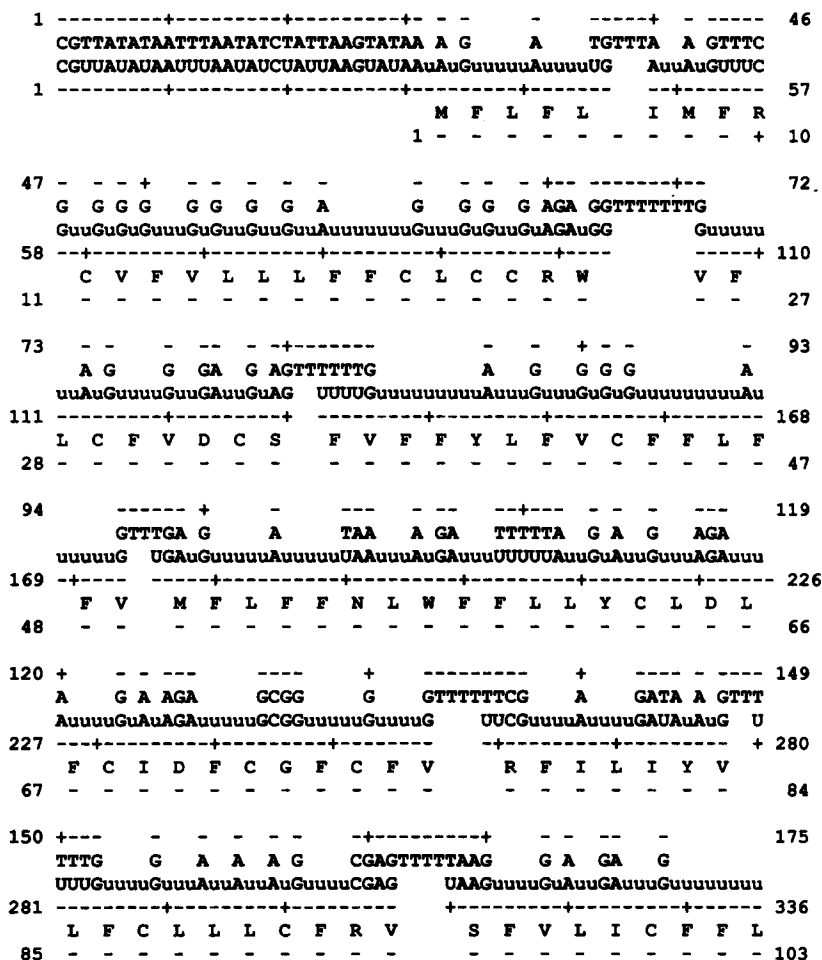
FIGURE 1. Pan-editing of ND9 mRNA from *L. tarentolae*. The DNA sequence is shown with gaps where Us are inserted. The edited RNA sequence and the translated amino acid sequence are also shown.

reactions[11,12] (FIG. 3) that are initiated with the hybridization of a specific gRNA molecule to the pre-edited mRNA just downstream of the first editing site, forming the "anchor duplex." An endonuclease then cleaves the mRNA at the first mismatch upstream of the anchor duplex. There is circumstantial evidence for the presence of at least three different nuclease activities in *Trypanosoma brucei* mitochondrial extracts, only one of which appears to be dependent on hybridization of a cognate gRNA.[13] Recent evidence also suggests that cleavage activity at U-deletion sites has different adenylate nucleotide requirements from that at U-insertion sites.[14] The only trypanosomatid mitochondrial nuclease to
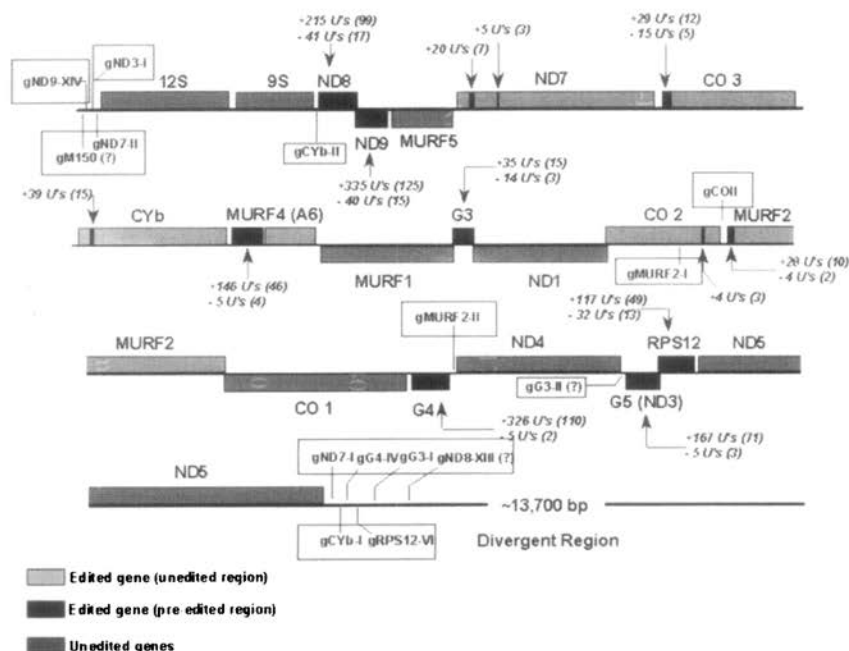
```
176  -        --    -+      --    -        - -- - --+   - - -     ------ 199
     G        GG    GG      AG    A        G AG  A GCG  A G A     TTGAAC
     uGuuuuuuGGuuuGGuuuuuAGuuuAuuuuuuuGuAGuuAuuGCGuuAuuGuAuuuUUGAAC
337  ---+---------+---------+---------+---------+---------+------ 396
       F  F  G  L  V  F  S  L  F  F  C  S  Y  A  L  C  I  F  E  R
104  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  - 124

200  + -------      --+---    --    -       - -- +- -- -    -----+- 231
     G GATTTAG      GATTTG    GG    G       G GG AA GA  G   ACATTCG
     GuGA    AGuuuuuGAUUUGuuuGGuuuuGuuuuuuGuGGuAAuGAuuGuuuACA UCGu
397  ---+      ---------+---------+---------+---------+-------- -+-- 452
       E       V  F  D  L  F  G  F  V  F  C  G  N  D  C  L  H   R
125  -        -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -   - 142

232      - -  ------      + --          - - ---- - +------- - 258
         A G  GATTGA      G  GG         G G AAAG A CCATTTTG  G
     uuuuAuGuuGAUUGAuuuuuuuGuuGGuuuuuuuuuuGuGuAAAGuuuuAuCCA  UUGuuuG
453  -------+---------+---------+---------+---------+--- ------+ 510
       F  Y  V  D  W  F  F  V  G  F  F  L  C  K  V  Y  P   L  F  G
143  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -   -  -  - 162

259  -    +    ------   ---+---     - - -  - - -+----- -   - - --  +-- 292
     G    G    ATTTTG   GTATTTA     G A G  A G GAAGATA  G   G GA GCC
     GuuuGuuuA    UGuuG A    AuuuuuuGuAuGuuAuGuGAAGAUAuuuGuuuGuGAuuGCC
511  ---------    +---- -    ----+---------+---------+---------+--- 563
       L  F  M    L  N    F  C  M  L  C  E  D  I  V  V  I  A
163  -  -  -    -  -    -  -  -  -  -  -  -  -  -  -  -  - 179

293  -- -----+ --   - - ---   --+-- ------- + --    -    ----- 328
     AC ATTTTG TG   G G GTG   TCCAA TTTGCAA A AG    A    TATTT
     ACuA    GuUGuuuuGuGuuGUGuuuuUCCAAuUUUGCAAuAuAGuuuuuAuuuUA  U
564  ----    --+---------+---------+---------+---------+------ - 617
       T  S      C  F  V  L  C  F  S  N  F  A  I  *
180  -  -      -  -  -  -  -  -  -  -  -  -  -       193

329  -  +---- --- --+---------+---------- 359
     A  TGTTA GAA CTGTCTATAAAATTAATTTATT
     AuuUGUUAuGAAuCUGUCUAUAAAAUUAAUUUAUU
618  --+---------+---------+---------+-- 652
```

┌─────────────────────────────────┐ ┌──────────────────────────────┐
│ **340 U-insertions in 125 sites** │ │ **40 U-deletions in 14 sites** │
└─────────────────────────────────┘ └──────────────────────────────┘

┌──────────────────────────┐
│ **196 codons created**   │
└──────────────────────────┘

FIGURE 1. *See legend on previous page.*

date isolated to homogeneity and the gene cloned is the MAR1 nuclease from *Leishmania tarentolae*.[14a] Experiments are in progress to determine if this nuclease is involved in the editing process.

It is presumed that the 5′ mRNA cleavage fragment is held together with the 3′ cleavage fragment by a combination of the base pairing between the nonencoded 3′ oligo(U)

FIGURE 2. Organization of maxicircle genome in *L. tarentolae*. The maxicircle was linearized at the single *EcoRI* site, and the unsequenced divergent region is indicated by a *broken line*. The 5' to 3' polarity of the genes is indicated by placement above or below the line. gRNA genes are *boxed*. The *shading* indicates whether the gene is unedited or is a cryptogene whose transcript is edited. The extent and localization of editing are indicated as shown.

tail of the gRNA and the purine-rich pre-edited region[12] and by interaction with yet undefined RNA-binding protein components of the RNP editing complexes. The next step of editing reaction is the addition of Us to the 3' end of the 5' cleavage fragment, apparently in an untemplated fashion by a mitochondrial 3' terminal uridylyl transferase (TUTase).[15] *In vitro* evidence[16-18] suggests that multiple Us are added to the 5' fragment, with the number of Us added being determined by the local UTP concentration. The added Us then form base pairs with guiding nucleotides in the gRNA, and the remaining single-stranded 3' overhang is trimmed by a 3'–5' U-specific exonuclease, which was recently shown to exist in an *L. tarentolae* mitochondrial extract (R. Aphasizhev and L. Simpson, unpublished results). This exonuclease is also thought to be responsible for the removal of unpaired Us at a U-deletion site. The two mRNA cleavage fragments, which are bridged by a cognate gRNA, are then ligated by a mitochondrial RNA ligase.[15,19] An RNA ligase has been localized to a 20S RNP complex and shown to consist of at least two adenylatable polypeptides of 45 and 50 kD.[20] This is consistent with recent evidence which indicates that a partially purified mitochondrial RNA ligase preparation from *L. tarentolae* is capable of ligating RNA-bridged fragments (V. Blanc and L. Simpson, unpublished results).
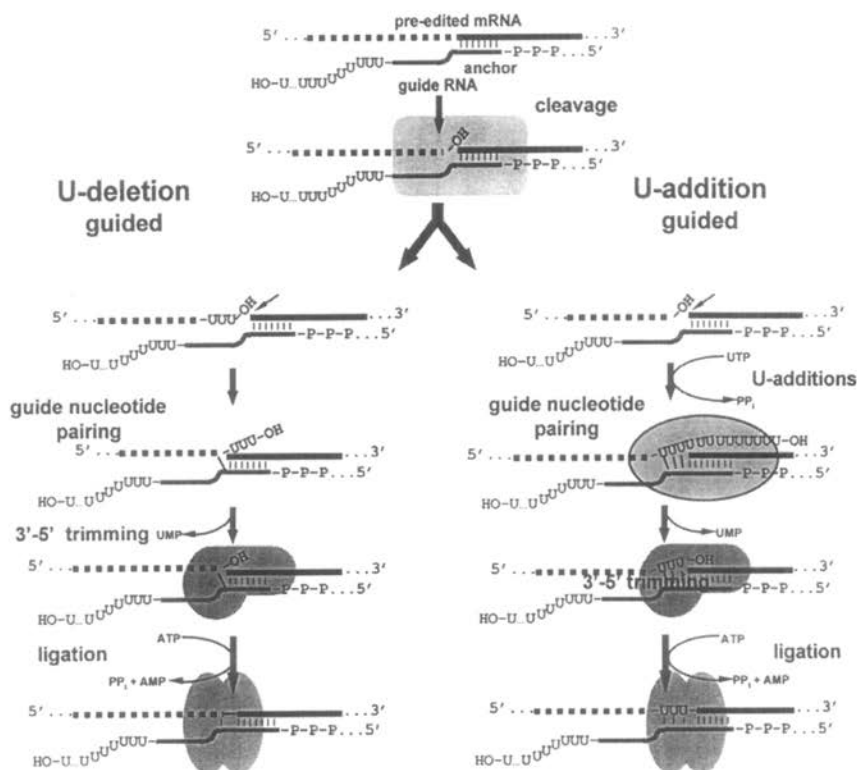
FIGURE 3. Modified enzyme cascade model of RNA editing. Both U-deletion and U-insertion pathways are shown.

Editing of some mRNAs is limited to a single block mediated by a single gRNA, and editing of others is more extensive and is mediated by two or more overlapping gRNAs. The known 3′ to 5′ polarity of editing in a multiple gRNA-mediated editing domain[21] is determined by the creation of upstream anchor sequences by downstream editing[22] (FIG. 4). Breathing of the upstream portion of the mRNA-gRNA duplex, which may be assisted by the observed asymmetric localization of G-U base pairs within the duplex, may allow the formation of the anchor duplex with the upstream gRNA. A mitochondrial RNA helicase may also be involved in unwinding the duplex, especially in regard to the 5′-most gRNA interaction.[23,24]

Evidence for this enzymatic cleavage-ligation model for editing involves the detection of the predicted enzymatic activities in mitochondrial extracts cosedimenting with *in vitro* editing activities[20,25] and the detection of the predicted 5′ and 3′ mRNA cleavage fragments during *in vitro* editing reactions.[16,17] The strongest evidence that the inserted Us are derived directly from UTP rather than from the 3′ oligo[U] tail of the gRNA is the lack of effect of blocking the 3′ end of the gRNA on a gRNA-directed *in vitro* U-insertion editing activity.[18]
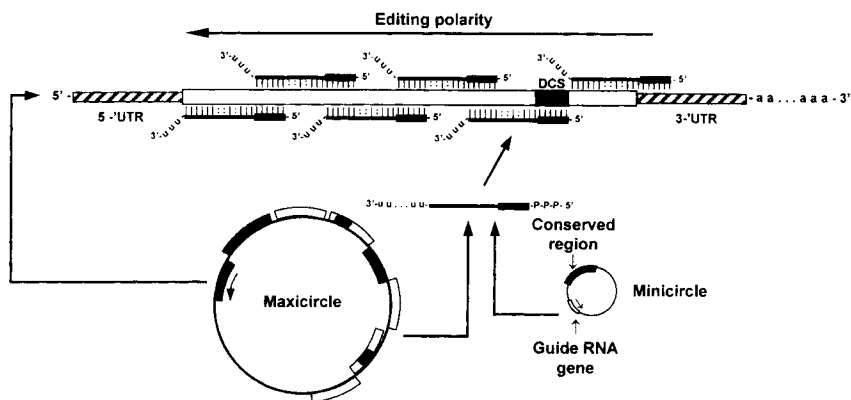
FIGURE 4. The 3' to 5' polarity of editing within an editing domain is determined by the creation of upstream anchor sequences by downstream editing. This is shown schematically for a two domain situation in which one gRNA mediates editing of the first domain and 5 overlapping gRNAs mediate editing of the second domain.

The editing process clearly involves protein enzyme-mediated reactions, and there is no evidence for the involvement of a ribozyme. Previous speculations that the mechanism of U-insertion/deletion editing is similar to the transesterification chemistry of RNA splicing[26,27] and that, therefore, this type of editing is as ancient as RNA splicing can probably be discarded.

EVOLUTION OF U-INSERTION/DELETION EDITING

It was noted shortly after the discovery of RNA editing that the number of Us inserted or deleted in a specific mRNA is not always the same in different species of trypanosomatids (FIG. 5). In extreme cases, an mRNA has to undergo extensive editing to become translatable in some species, but is translatable without any editing in other species. An example of this was the demonstration that the gene for subunit III of cytochrome c oxidase (COIII) is not missing from the genome of *T. brucei*, as was originally thought,[28] but is present in a cryptic, "pan-edited" form which makes it unrecognizable at the DNA level.[29] However, the fully edited mRNA derived from this cryptogene is highly similar to the COIII mRNA in *L. tarentolae*, which undergoes only limited editing at the 5' end.[30]

These and similar observations suggested that RNA editing represents a dynamic, evolving system, but it was not clear in which direction it evolves. Is editing being gradually replaced with unedited forms or is it continuing to spread in the genome? Is this process driven by the evolution of parasitic adaptations, such as acquisitions of new hosts or inventions of novel survival strategies, or does it evolve independently? And the most important questions about RNA editing are: how, why, and when did it appear for the first time?

FIGURE 5. Phylogenetic analysis of kinetoplastid RNA editing. A maximum likelihood ribosomal RNA phylogenetic tree representing only the main trypanosomatid and some bodonid lineages (J. Lukeš and D. Maslov, unpublished results) with the corresponding bootstrap values is shown on the left. A representation of the known cryptogenes is shown on the right: *open boxes* correspond to unedited sequences, *black boxes* to pre-edited regions or edited cryptogenes, and *cross-hatched boxes* indicate a lack of information.

One way to address these problems is by using a cladistic approach. Towards this end, the phylogeny of kinetoplastid protozoa can be used as a framework for an analysis of editing, in parallel with an analysis of the evolution of parasitism. In previous studies the clade of trypanosomes was paraphyletic, with *T. brucei* forming the earliest separating lineage.[31-34] The most parsimonious assumption was that pan-editing is primitive and 5'-editing is derived in trypanosomatids. In regard to the origin of parasitism, this topology suggested an early acquisition of a digenetic (two-host) life cycle followed by multiple and independent losses of a vertebrate host as a likely evolutionary scenario. However, more recent analyses indicate that the paraphyly is an artifact caused by unequal rates of sequence change, and the lineage of *T. brucei* is a member of a monophyletic clade that includes all investigated trypanosomatids.[35-40]

A recent phylogenetic tree constructed with small and partial large ribosomal RNA sequences (J. Lukeš and D.A. Maslov, unpublished results) is shown in FIGURE 5. The tree is rooted using the outgroup organisms, *Euglena gracilis* and *Diplonema papillatum*. The ingroup includes several representatives of the suborder Bodonina (*Bodo caudatus, Bodo saltans, Cryptobia helicis, Trypanoplasma borreli, Rhynchobodo* sp., and *Dimastigella trypaniformis*) and the suborder Trypanosomatina (*L. tarentolae, Herpetomonas muscarum, Phytomonas serpens, Blastocrithidia culicis,* and *T. brucei*). Bodonids appear on the tree as a paraphyletic group, and trypanosomatids as a late-diverging monophyletic clade. That analysis, together with the results of other studies, has demonstrated that there are at least five monophyletic clades of trypanosomatids: a clade that includes the genera *Leishmania, Crithidia,* and some other insect trypanosomatids, a clade composed of the genus *Herpetomonas,* a clade of plant parasites, *Phytomonas,*[41] a clade that unites endosymbiont-bearing members of several genera,[40] and a clade of trypanosomes.[35,37,40] Although not quite uniform in terms of their parasitic adaptations, members of the same clade seem to share editing patterns in corresponding mRNAs. For this reason, only a single representative of each major trypanosomatid clade is included in FIGURE 5.

With the new topology, the major divergence within *Trypanosomatidae* is represented by a dichotomic split between the clade of trypanosomes and all remaining clades. This topology strengthens the view that there were multiple and independent origins of digenetic life cycles. However, solving this problem by a cladistic approach requires that all major clades be identified and their relationships fully resolved, a task that has not yet been achieved. Nevertheless, given the striking differences in the mechanisms used by different parasites to avoid the defense systems of their vertebrate hosts, multiple origins of digeneity seem to be likely.

The new tree topology also affects the interpretation of the evolution of RNA editing in trypanosomatids. A monophyletic origin of all trypanosomatids implies that the dichotomy just mentioned also includes these two character-states, thus making a similar parsimony argument impossible to apply.

A solution to this problem lies in a more careful analysis of editing in bodonids, which represent the second major group of the kinetoplastid protozoa. However, it is evident from FIGURE 5 that editing has not yet been investigated in many known bodonid (and even some trypanosomatid) lineages. The problem is complicated even more by a growing realization that there may be many more bodonid lineages than have been found so far. All this means that the daunting task of conducting a strict cladistic analysis of editing must be left for the future.

Nevertheless, the available data already allow us to draw several conclusions. One conclusion is that for most of the investigated genes, the extent of editing in mRNA can vary dramatically. A striking example of this is the ND5 mRNA, which is unedited in all studied trypanosomatids and pan-edited in *B. saltans*.[42] This suggests a possible lack of correlation between editing and specific gene function. A second conclusion is that a pattern of changes observed for a certain gene (e.g., A6) is not paralleled by another gene (e.g., COIII), indicating that different genes change independently. Nevertheless, in some cases, similar editing patterns are indeed shared by the members of different clades. For example, the editing of the Cyb and COII mRNAs in *Leishmania* and *Trypanosoma* is very similar, as is the editing of the A6 mRNA in *Phytomonas* and *Blastocrithidia*. This indicates either homoplasy (parallel changes or convergence) or inheritance of a specific form from a common ancestor. Furthermore, because unique and shared forms are interspersed with each other, the changes, regardless of their nature, must have occurred at different times in evolutionary history. And finally, as shown by the discovery of editing in a free-living organism, *B. saltans*,[42] RNA editing is not a specific attribute of parasitism, although, as was discussed elsewhere,[43] it may play an important role.

On a larger scale, it seems likely that RNA editing, in any form, is a primitive feature in kinetoplastids. This conclusion, however, still has to be corroborated by a demonstration of editing in the most deeply diverging lineages, *C. helicis* and *B. caudatus*. U-insertion editing has not been found in the outgroup organisms, *D. papillatum* and *E. gracilis*[44,45] (D.A. Maslov, S. Yasuhira, and L. Simpson, Protist. In press), neither was it found in any other investigated group of protists.[46] This dates the origin of editing to the time after the split of the kinetoplastid and euglenoid lineages.

The separation of gRNA genes into a separate genome occurred within the bodonid lineage, since *T. borreli* has an 80-kb molecule containing the rRNA and structural genes and a 180-kb molecule containing gRNA genes.[47,48] At the same time, *B. saltans* has 1.4-kb minicircles, which probably encode gRNAs.[42] The origin of a catenated network of minicircles and maxicircles must have occurred in the progenitor of trypanosomatids, because catenation has not been found so far in bodonids.

Given these limitations of a cladistic analysis of the evolution of editing, an alternative approach to this problem can be taken. This approach includes consideration of the molecular mechanisms involved in the spread or reduction of editing and a search for their presence in the kinetoplast.

RETROPOSITION MODEL FOR THE LOSS OF PAN-EDITING IN EVOLUTION

As just discussed, editing proceeds 3′ to 5′ within an editing domain, and this polarity is due to the mediation of multiple overlapping gRNAs, each of which contains the information for editing a single sequence block (FIG. 4). In comparing the extent of editing of homologous genes in different kinetoplastid species (FIG. 5), it is striking that certain genes exhibit a progressive restriction of editing to the 5′ end of an editing domain.[32] This can be seen nicely in a comparison of the A6 cryptogenes. In *T. brucei* and *T. cruzi*, this mRNA is completely pan-edited by approximately 8–10 gRNAs. In other species, editing is restricted to differing extents of the 5′ portion of the mRNA and involves fewer numbers of gRNAs. For example, A6 editing requires six overlapping gRNAs in *L. tarentolae*, four

gRNAs in *C. fasciculata* and *H. muscarum*, and one gRNA in *B. culicis*. As another example, the NADH dehydrogenase 7 (ND7) mRNA is pan-edited in two domains in *T. brucei*, each with multiple gRNAs, but in *L. tarentolae* editing is limited to single blocks at the 5' ends of each domain, each of which was mediated by a single gRNA.

Because the 5' edited cryptogenes resemble the structures of partially edited mRNAs transcribed from pan-edited cryptogenes, it was attractive to speculate that pan-edited cryptogenes were substituted in evolution by retroposition of partially edited mRNAs (FIG. 6).[32,49,50] The recent demonstration of reverse transcriptase activity in whole cell extracts of modern trypanosomatids is consistent with this model.[51,52] Because most kinetoplastids are obligate aerobes or at least have such a stage in their life cycle and require functional respiratory and oxidative phosphorylation enzymes for viability, loss of the ability to fully edit an mRNA would be lethal. *Phytomonas* species resemble the bloodstream stage of *T. brucei* in that they can live glycolytically and do not have a functional respiratory chain, but recent evidence suggests that this is a derived character due to deletion of a portion of the maxicircle genome (D.A. Maslov, P. Nawathean, and J. Scheel. Mol. Biochem. Parasitol. In press). Free-living bodonid protozoa are facultative anaerobes and presumably would only be affected by loss of editing in the aerobic phase.

Because the copy number of different minicircle sequence classes ranges from a few to thousands of copies per network, and because minicircles are randomly segregated to daughter networks at division of the kinetoplast, it is possible that an entire sequence class could be lost by missegregation. For those cells that require aerobic respiration, substitution of a pan-edited cryptogene with a partially edited cryptogene in at least one of the 50 maxicircle molecules in the mitochondrion would allow the cells to survive the loss of an entire sequence class of minicircles encoding a specific gRNA involved in an editing cascade. Such a loss would apply a strong selective pressure for replacement of the specific pan-edited cryptogene by a 5' edited cryptogene.

A loss of over 30 different minicircle sequence classes and the encoded gRNAs for the editing of five mRNAs was proposed to occur during the long culture history of the UC laboratory strain of *L. tarentolae*.[53,54] The evidence is that a recently isolated strain of *L. tarentolae* can edit these five mRNAs and contains at least 32 additional minicircle-encoded gRNAs. It was presumed that the protein products of these edited mRNAs, at least three of which encode components of respiratory Complex I, were not required during the culture phase, and therefore there was no selective pressure to maintain the corresponding gRNA sets. The finding that a loss of apparently nonessential gRNA genes can occur during prolonged cultivation makes the model for loss of pan-editing in evolution by retroposition of partially edited RNAs more plausible.

An alternative explanation for the observed distribution of editing patterns in homologous genes from different kinetoplastid species is that editing is a derived character in each lineage. However, the *de novo* creation of a complex cascade of editing involving multiple gRNAs in each lineage is extremely unlikely. There is in fact only a single example in which a claim was made for the creation of editing and that is the frameshift editing of the COII mRNA.[42] This gene is unedited in bodonids *T. borreli* and *C. helicis*, but in *B. saltans* is edited in two of the three editing sites that are normally edited in all trypanosomatids. This is the sole editing that is mediated by a gRNA in *cis* rather than in *trans*, and the length of the base-paired region between the gRNA and mRNA is similar in the bodonid and trypanosomatid lineages. It was suggested that this is an example of a recent evolution of an editing domain by a gradual increase in the number of editing sites.[42] However,
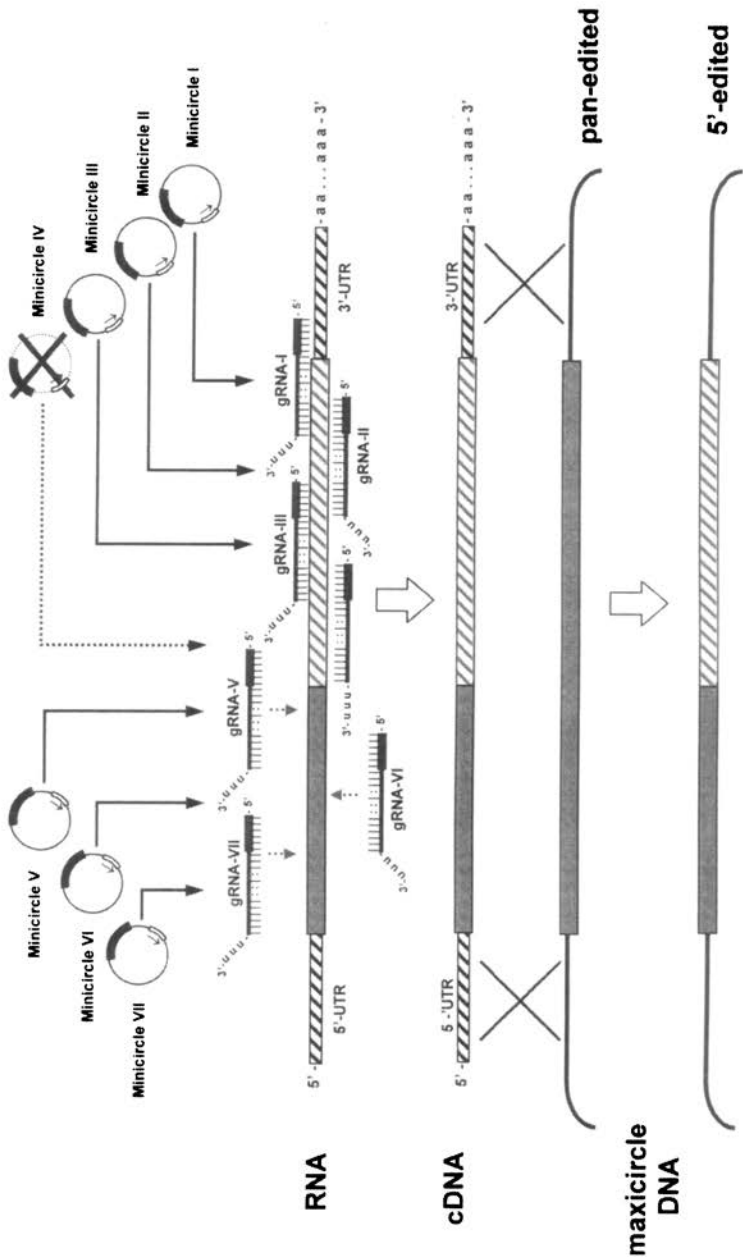
FIGURE 6. Retroposition model for evolution of U-insertion/deletion editing. A pan-edited domain originally mediated by 7 overlapping gRNAs is shown schematically. The loss of minicircle class IV is indicated by a diagonal cross (X).

although the form of COII with two Us inserted at two sites may be more ancient than the form with four Us inserted at three sites, it is easy to envision the loss of two encoded Ts by neutral drift, creating site 3 in the ancestor of trypanosomatids. This mutation is rescued at the RNA level by insertion of two Us mediated by the existing gRNA. Alternatively, the acquisition of two Ts in the *B. saltans* lineage could have resulted in elimination of site 3. A similar drift of T residues may have eliminated all the editing sites in COII in *T. borreli* and *C. helicis*, which was followed by random mutations in the gRNA sequence. Regardless of their direction, these events illustrate what could happen within a block of editing after establishment of a gRNA in regard to divergence of a pre-edited sequence and creation or elimination of new editing sites.

MODELS FOR THE ORIGIN OF U-INSERTION/DELETION EDITING

As just discussed, the origin of the mitochondrial U-insertion/deletion type of RNA editing probably occurred after the split of the euglenoid and kinetoplastid lineages, and a comparative analysis of editing patterns in extant species suggested that pan-editing was the most primitive form of editing.[32,33] It is likely that editing originated in a free-living bodonid protozoan, which are known today to be facultative anaerobes. Cavalier-Smith[55] speculated that there was a coevolution of the glycosome, a microbody-like organelle containing all the gycolytic enzymes which provided an extremely efficient source of energy production under anaerobic conditions and the editing machinery. He proposed that because a facultative anaerobe is temporarily shielded from selection against harmful mutations in mitochondrial DNA when growing anaerobically, RNA editing evolved under anaerobic conditions as a method of correcting genetic defects and that the presence of this process provided a strong selection under aerobic conditions.

Covello and Gray[56] proposed a general three-step model for the evolution of any type of editing phenomenon. These steps included the appearance of RNA editing activity, mutations at editable nucleotide positions which result in fixation of the editing information associated with the editing activity, and maintenance of RNA editing activity by natural selection. Because the mechanism of kinetoplastid editing is now known to involve several specific enzymatic activities, we speculate that these activities were already present in the mitochondrion of the free-living bodonid and were utilized for other pathways, such as mismatch repair or RNA turnover. The evolution of the gRNA could be imagined to occur in several ways. In one scenario, there is a partial gene duplication and inversion elsewhere in the mitochondrial genome. This gene would give rise to an antisense transcript, which would represent the proto-gRNA. The evolution of G-U base pairs might arise from the action of a cytidine deaminase, converting C residues to U residues in either the mRNA or the gRNA, and an asymmetric distribution of G-U base pairs would be selected for in the gRNA-mRNA duplex to allow breathing of the duplex. Another scenario would involve the evolution of gRNAs from random sequences in the mitochondrial genome. Statistical analysis of the distribution of gRNA-like sequences in the *L. tarentolae* maxicircle sequence indicates that these sequences are fairly random.[57] Again, an asymmetric distribution of G-U base pairs will then be selected for in these molecules to allow a 3′ to 5′ polarity of editing. The next step is to evolve the enzyme cascade machinery from preexisting activities. Possibly, the conversion of an endoribonuclease into an

enzyme that recognizes the initial mismatch upstream of the gRNA-mRNA anchor duplex was a critical event.

After the establishment of an RNA editing machinery to correct at the RNA level genetic drift at the loci with T residues, this process could have been utilized for gene regulation in the mitochondrion. This would provide a selective pressure to have maintained this process throughout more than 500 million years of evolutionary history in the face of strong pressure to convert edited sequences to unedited sequences by retroposition of partially or fully edited cDNAs. Regulation of editing does occur in the different life cycle stages of modern *T. brucei*,[58–62] and this regulation is clearly advantageous to regulate mitochondrial biosynthesis during the biphasic life cycle. Little is known about regulation in other modern trypanosomatids or bodonids, but it is possible that this does occur even in monogenetic (single-host) species.

If one assumes that gRNA genes originate within the maxicircle genome, the next step to segregate most of the gRNA genes within a separate genome, either as separate non-catenated self-replicating minicircles, as in *B. saltans*[42] and *C. helicis*,[63] or as multiple genes within a large circular chromosome, such as *T. borreli*.[48] The network of catenated minicircles must have originated in an ancestor of the entire trypanosomatid lineage, because all known trypanosomatid species contain a network. The selective advantage of the network may have been to act as a DNA-based mitotic apparatus to ensure random segregation of both the maxicircle molecules and a representative selection of minicircle molecules to daughter networks.[64]

CONCLUSIONS AND PERSPECTIVES

We conclude from this analysis that the U-insertion/deletion type of RNA editing probably originated in the mitochondrion of an ancestral free-living bodonid protozoan and rapidly evolved multiple overlapping gRNA pan-editing patterns for several if not all of the mitochondrial genes. The gRNA genome became partially separated from the cryptogene genome and evolved into self-replicating minicircle molecules, each containing 1–4 genes. In an ancestor of the trypanosomatids, minicircles and maxicircles became catenated into a single giant network of DNA. During the course of evolution, the primitive pan-editing patterns simplified, and editing became restricted to the 5' end of domains. Editing proved to have a selective advantage as it was utilized as a gene regulation method and allowed the cells to vary their mitochondrial metabolism to inhabit different environmental niches, including parasitic niches.

To confirm this speculative hypothesis and illuminate more clearly the evolutionary origin of U-insertion/deletion editing and the gRNA genome, a more extensive comparative analysis of the variety of genomic organizations and editing patterns in a larger number and variety of extant bodonid lineages is required. Substantiated negative evidence for an absence of editing in euglenoid mitochondria and in α-proteobacteria would also be very helpful.

REFERENCES

1. SIMPSON, L. & R.B. EMESON. 1996. RNA editing. Annu. Rev. Neurosci. **19:** 27–52.

2. BENNE, R., J. VAN DEN BURG, J. BRAKENHOFF, P. SLOOF, J. VAN BOOM & M. TROMP. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46: 819–826.

3. SIMPSON, L. & J. SHAW. 1989. RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. Cell 57: 355–366.

4. ARTS, G.J. & R. BENNE. 1996. Mechanism and evolution of RNA editing in kinetoplastida. Biochim. Biophys. Acta Gene Struct. Expression 1307: 39–54.

5. STUART, K., T.E. ALLEN, S. HEIDMANN & S.D. SEIWERT. 1997. RNA editing in kinetoplastid protozoa. Microbiol. Rev. 61: 105–120.

6. ADLER, B.K. & S.L. HAJDUK. 1994. Mechanisms and origins of RNA editing. Curr. Opin. Genet. Dev. 4: 316–322.

7. SIMPSON, L. 1987. The mitochondrial genome of kinetoplastid protozoa: Genomic organization, transcription, replication and evolution. Ann. Rev. Microbiol. 41: 363–382.

8. SIMPSON, L. 1986. Kinetoplast DNA in trypanosomid flagellates. Int. Rev. Cytol. 99: 119–179.

9. SIMPSON, L. 1972. The kinetoplast of the hemoflagellates. Int. Rev. Cytol. 32: 139–207.

10. STURM, N.R. & L. SIMPSON. 1990. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell 61: 879–884.

11. ALFONZO, J.D., O. THIEMANN & L. SIMPSON. 1997. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. Nucleic Acids Res. 25: 3751–3759.

12. BLUM, B., N. BAKALARA & L. SIMPSON. 1990. A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60: 189–198.

13. PILLER, K.J., L.N. RUSCHE, J. CRUZ-REYES & B. SOLLNER-WEBB. 1997. Resolution of the RNA editing gRNA-directed endonuclease from two other endonucleases of *Trypanosoma brucei* mitochondria. RNA 3: 279–290.

14. CRUZ-REYES, J., L. RUSCHE, K.J. PILLER & B. SOLLNER-WEBB. 1998. *T. brucei* RNA editing: Adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. Mol. Cell 1: 401–409.

14a. ALFONZO, J.D., O.H. THIEMANN & L. SIMPSON. 1998. Purification and characterization of MAR1: A mitochondrial-associated ribonuclease from *Leishmania tarentolae*. J. Biol. Chem. 273: 30003–30011.

15. BAKALARA, N., A.M. SIMPSON & L. SIMPSON. 1989. The *Leishmania* kinetoplast-mitochondrion contains terminal uridylyltransferase and RNA ligase activities. J. Biol. Chem. 264: 18679–18686.

16. KABLE, M.L., S.D. SEIWERT, S. HEIDMANN & K. STUART. 1996. RNA editing: A mechanism for gRNA-specified uridylate insertion into precursor mRNA [see comments]. Science 273: 1189–1195.

17. SEIWERT, S.D., S. HEIDMANN & K. STUART. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell 84: 831–841.

18. BYRNE, E.M., G.J. CONNELL & L. SIMPSON. 1996. Guide RNA-directed uridine insertion RNA editing in vitro. EMBO J. 15: 6758–6765.

19. SABATINI, R. & S.L. HAJDUK. 1995. RNA ligase and its involvement in guide RNA/mRNA chimera formation. J. Biol. Chem. 270: 7233–7240.

20. PERIS, M., G.C. FRECH, A.M. SIMPSON, F. BRINGAUD, E. BYRNE, A. BAKKER & L. SIMPSON. 1994. Characterization of two classes of ribonucleoprotein complexes possibly involved in RNA editing from *Leishmania tarentolae* mitochondria. EMBO J. 13: 1664–1672.

21. ABRAHAM, J., J. FEAGIN & K. STUART. 1988. Characterization of cytochrome c oxidase III transcripts that are edited only in the 3' region. Cell 55: 267–272.

22. MASLOV, D.A. & L. SIMPSON. 1992. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell 70: 459–467.

23. MISSEL, A., G. NORSKAU, H.H. SHU & H.U. GORINGER. 1995. A putative RNA helicase of the DEAD box family from *Trypanosoma brucei*. Mol. Biochem. Parasitol. 75: 123–126.

24. MISSEL, A., A.E. SOUZA, G. NORSKAU & H.U. GORINGER. 1997. Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. Mol. Cell Biol. 17: 4895–4903.

25. CORELL, R.A., L.K. READ, G.R. RILEY, J.K. NELLISSERY, T.E. ALLEN, M.L. KABLE, M.D. WACHAL, S.D. SEIWERT, P.J. MYLER & K.D. STUART. 1996. Complexes from *Trypanosoma brucei* that exhibit deletion editing and other editing-associated properties. Mol. Cell. Biol. **16:** 1410–1418.

26. BLUM, B., N.R. STURM, A.M. SIMPSON & L. SIMPSON. 1991. Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell **65:** 543–550.

27. CECH, T.R. 1991. RNA editing: World's smallest introns. Cell **64:** 667–669.

28. SIMPSON, L., N. NECKELMANN, V. DE LA CRUZ, A. SIMPSON, J. FEAGIN, D. JASMER & K. STUART. 1987. Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. J. Biol. Chem. **262:** 6182–6196.

29. FEAGIN, J.E., J. ABRAHAM & K. STUART. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell **53:** 413–422.

30. SHAW, J., J.E. FEAGIN, K. STUART & L. SIMPSON. 1988. Editing of mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell **53:** 401–411.

31. FERNANDES, A.P., K. NELSON & S.M. BEVERLEY. 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: Perspectives on the age and origins of parasitism. Proc. Natl. Acad. Sci. USA **90:** 11608–11612.

32. MASLOV, D.A., H.A. AVILA, J.A. LAKE & L. SIMPSON. 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature **365:** 345–348.

33. LANDWEBER, L.F. & W. GILBERT. 1994. Phylogenetic analysis of RNA editing: A primitive genetic phenomenon. Proc. Natl. Acad. Sci. USA **91:** 918–921.

34. MASLOV, D.A., J. LUKEŠ, M. JIRKU & L. SIMPSON. 1996. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. Mol. Biochem. Parasitol. **75:** 197–205.

35. LUKES, J., M. JIRKU, D. DOLEZEL, I. KRAL'OVÁ, L. HOLLAR & D.A. MASLOV. 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J. Mol. Evol. **44:** 521–527.

36. ALVAREZ, F., M.N. CORTINAS & H. MUSTO. 1996. The analysis of protein coding genes suggests monophyly of *Trypanosoma*. Mol. Phylogenet. Evol. **5:** 333–343.

37. HAAG, J., C. O'HUIGIN & P. OVERATH. 1998. The molecular phylogeny of trypanosomes: Evidence for an early divergence of the Salivaria. Mol. Biochem. Parasitol. **91:** 37–49.

38. BERCHTOLD, M., H. PHILIPPE, A. BREUNIG, G. BRUGEROLLE & H. KOENIG. 1995. The phylogenetic position of *Dimastigella trypaniformis* within the parasitic kinetoplastids. Parasitol. Res. **80:** 672–679.

39. MARCHE, S., C. ROTH, H. PHILIPPE, M. DOLLET & T. BALTZ. 1995. Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. Mol. Biochem. Parasitol. **71:** 15–26.

40. HOLLAR, L., J. LUKES & D.A. MASLOV. 1998. Monophyly of endosymbiont containing trypanosomatids: Phylogeny versus taxonomy. J. Eukaryot. Microbiol. **45:** 293–297.

41. HOLLAR, L. & D.A. MASLOV. 1997. A phylogenetic view on the genus *Phytomonas*. Mol. Biochem. Parasitol. **89:** 295–299.

42. BLOM, D., A. DE HAAN, M. VAN DEN BERG, P. SLOOF, M. JIRKU, J. LUKEŠ & R. BENNE. 1998. RNA editing in the free-living bodonid *Bodo saltans*. Nucleic Acids Res. **26:** 1205–1213.

43. MASLOV, D.A. & L. SIMPSON. 1995. Evolution of parasitism in kinetoplastid protozoa. Parasitol. Today **11:** 30–32.

44. YASUHIRA, S. & L. SIMPSON. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. J. Mol. Evol. **44:** 341–347.

45. TESSIER, L.H., H. VAN DER SPECK, J.M. GUALBERTO & J.M. GRIENENBERGER. 1997. The *cox1* gene from *Euglena gracilis*: A protist mitochondrial gene without introns and genetic code modifications. Curr. Genet. **31:** 208–213.

46. LANG, B.F., G. BURGER, C.J. O'KELLY, R. CEDERGREN, G.B. GOLDING, C. LEMIEUX, D. SANKOFF, M. TURMEL & M.W. GRAY. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387:** 493–497.

47. YASUHIRA, S. & L. SIMPSON. 1996. Guide RNAs and guide RNA genes in the cryptobiid kinetoplastid protzoan, *Trypanoplasma borreli*. RNA **2:** 1153–1160.

48. MASLOV, D.A. & L. SIMPSON. 1994. RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan, *Trypanoplasma borreli*. Mol. Cell. Biol. **14:** 8174–8182.
49. SIMPSON, L. & D.A. MASLOV. 1994. RNA editing and the evolution of parasites. Science **264:** 1870–1871.
50. LANDWEBER, L.F. 1992. The evolution of RNA editing in kinetoplastid protozoa. BioSystems **28:** 41–45.
51. GONZÁLEZ, C.I., M.C. THOMAS, F. MARTÍN, J. ALCAMI, C. ALONSO & M.C. LÓPEZ. 1997. Reverse transcriptase-like activity in *Trypanosoma cruzi*. Acta Tropica **63:** 117–126.
52. GABRIEL, A. & J.D. BOEKE. 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. Proc. Natl. Acad. Sci. USA **88:** 9794–9798.
53. THIEMANN, O.H., D.A. MASLOV & L. SIMPSON. 1994. Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes. EMBO J. **13:** 5689–5700.
54. MASLOV, D.A., O. THIEMANN & L. SIMPSON. 1994. Editing and misediting of transcripts of the kinetoplast maxicircle G5 (ND3) cryptogene in an old laboratory strain of *Leishmania tarentolae*. Mol. Biochem. Parasitol. **68:** 155–159.
55. CAVALIER-SMITH, T. 1997. Cell and genome coevolution: Facultative anaerobiosis, glycosomes and kinetoplastan RNA editing. Trends Genet. **13:** 6–9.
56. COVELLO, P.S. & M.W. GRAY. 1993. On the evolution of RNA editing. Trends Genet. **9:** 265–268.
57. VON HAESELER, A., B. BLUM, L. SIMPSON, N. STURM & M.S. WATERMAN. 1992. Computer methods for locating kinetoplastid cryptogenes. Nucleic Acids Res. **20:** 2717–2724.
58. FEAGIN, J., D. JASMER & K. STUART. 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell **49:** 337–345.
59. FEAGIN, J. & K. STUART. 1988. Developmental aspects of uridine addition within mitochondrial transcripts of *Trypanosoma brucei*. Mol. Cell Biol. **8:** 1259–1265.
60. KOSLOWSKY, D.J., G.J. BHAT, A.L. PERROLLAZ, J.E. FEAGIN & K. STUART. 1990. The MURF3 gene of T. brucei contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell **62:** 901–911.
61. SOUZA, A.E., P.J. MYLER& K. STUART. 1992. Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited, developmentally regulated, and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. Mol. Cell. Biol. **12:** 2100–2107.
62. READ, L.K., MYLER, P.J. & STUART, K. 1992. Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. J. Biol. Chem. **267:** 1123–1128.
63. LUKEŠ, J., M. JIRKU, N. AVLIYAKULOV & O. BENADA. 1998. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. EMBO J. **17:** 838–846.
64. BORST, P. 1991. Why kinetoplast DNA networks? Trends. Genet. **7:** 139–141.