

Evolution of RNA editing in trypanosome mitochondria

Larry Simpson^{*†‡}, Otavio H. Thiemann[§], Nicholas J. Savill^{||}, Juan D. Alfonzo^{*}, and D. A. Maslov^{**}

^{*}Howard Hughes Medical Institute and [†]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095; [‡]School of Biological Sciences, Manchester University, Manchester, United Kingdom M13 9PT; [§]Laboratory of Protein Crystallography and Structural Biology, Physics Institute of Sao Carlos, University of Sao Paulo, Av. Dr. Carlos Botelho 1465, PO Box 369, Sao Carlos, SP, Brazil 13560-970; and ^{**}Department of Biology, University of California, 3401 Watkins Drive, Riverside, CA 92521

Two different RNA editing systems have been described in the kinetoplast-mitochondrion of trypanosomatid protists. The first involves the precise insertion and deletion of U residues mostly within the coding regions of maxicircle-encoded mRNAs to produce open reading frames. This editing is mediated by short overlapping complementary guide RNAs encoded in both the maxicircle and the minicircle molecules and involves a series of enzymatic cleavage-ligation steps. The second editing system is a C₃₄ to U₃₄ modification in the anticodon of the imported tRNA^{Trp}, thereby permitting the decoding of the UGA stop codon as tryptophan. U-insertion editing probably originated in an ancestor of the kinetoplastid lineage and appears to have evolved in some cases by the replacement of the original pan-edited cryptogene with a partially edited cDNA. The driving force for the evolutionary fixation of these retroposition events was postulated to be the stochastic loss of entire minicircle sequence classes and their encoded guide RNAs upon segregation of the single kinetoplast DNA network into daughter cells at cell division. A large plasticity in the relative abundance of minicircle sequence classes has been observed during cell culture in the laboratory. Computer simulations provide theoretical evidence for this plasticity if a random distribution and segregation model of minicircles is assumed. The possible evolutionary relationship of the C to U and U-insertion editing systems is discussed.

The term RNA editing describes several types of posttranscriptional modifications of RNAs that involve either specific insertion/deletion or modifications of nucleotides (1). The uridine (U)-insertion/deletion type of editing has so far only been found to occur in the mitochondria of kinetoplastid protists (2, 3). We recently showed that C to U nucleotide modification editing also occurs in the mitochondria of these cells (4). The origin and evolution of these two genetic systems is the subject of this paper.

Kinetoplastid Protists Consist of Two Major Groups: the Trypanosomatids and the Bodonids

Kinetoplastid protists belonging to the Euglenozoa phylum, according to rRNA phylogenetic trees, represent one of the earliest mitochondrial-containing extant branches of the eukaryotic lineage (5). This view may change in the future, as protein-based phylogenies favor a later divergence of Euglenozoa (6–8). However, even in such a case, this phylum still demonstrates a long and independent evolutionary history and is well separated from other eukaryotic groups. Taxonomists previously have proposed the existence of two suborders in the Kinetoplastida, the Trypanosomatina and Bodonina. All of the pathogenic trypanosomatids belong to the suborder, Trypanosomatina, and to the single family, Trypanosomatidae. Phylogenetic reconstructions using nuclear SSU rRNA sequences have confirmed

the separation of the trypanosomatids as a derived late-emerging group. The trypanosomes, which initially were thought to be paraphyletic, with *Trypanosoma brucei* as an early-diverging branch (9–11), are now thought more likely to be monophyletic (12, 13) (Fig. 1). There are two major clades of trypanosomatids, the trypanosomes and the clade of *Leishmania*, *Crithidia*, *Leptomonas*, *Phytomonas*, *Herpetomonas*, and *Blastocrithidia*. An early divergence within the trypanosome lineage led to separate salivarian (e.g., *T. brucei*) and nonsalivarian trypanosomes (14). One study further splits the nonsalivarian trypanosomes into two clades, consisting of bird trypanosomes, such as *Trypanosoma avium*, and stercoarian trypanosomes such as *Trypanosoma cruzi* (14).

The bodonid group is poorly studied and is probably paraphyletic (15). The rRNA tree in Fig. 1 includes multiple species from this lineage. The deepest branches of the bodonid lineage include the poorly studied free-living organisms *Bodo designis*, *Rhynchobodo*, and *Dimastigella*. This is followed by a mixed clade of free-living *Bodo caudatus*, *Cryptobia helices*, and the parasitic *Cryptobia salmositica* and *Trypanoplasma borreli*. Another free-living organism, *Bodo saltans*, may represent the closest relative of trypanosomatids. *B. caudatus*, *T. borreli*, *C. helices*, and *B. saltans* are the only bodonids for which mitochondrial molecular data are available. *Diplonema papillatum* represents either a sister group to the kinetoplastids or a sister group to the euglenoids (16). It is of some interest that the UGA stop codon is used to encode tryptophan in the mitochondrial genome of all kinetoplastid species including *Diplonema*, but not in the Euglenoids (17).

Kinetoplastids Contain a Single Extended Tubular Mitochondrion with an Unusual Mitochondrial DNA

The trypanosomatids contain a single tubular mitochondrion (18, 19) that has a single mass of mitochondrial DNA situated within the matrix adjacent to the basal body of the flagellum (20). The trypanosomatid mitochondrial or kinetoplast DNA (kDNA) consists of a single highly structured disk-shaped network composed of thousands of catenated minicircles ranging in size from as small as 465 bp in *Trypanosoma vivax* (21) to as large as 10,000 bp in *T. avium* (22), and a smaller number of catenated maxicircles ranging in size from 23 to 36 kb. The maxicircles are the

This paper was presented at the National Academy of Sciences colloquium "Variation and Evolution in Plants and Microorganisms: Toward a New Synthesis 50 Years After Stebbins," held January 27–29, 2000, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviations: gRNA, guide RNA; kDNA, kinetoplast DNA.

[†]To whom reprint requests should be addressed at: Howard Hughes Medical Institute—University of California, 6708 MRL, 675 Charles Young Drive South, Los Angeles, CA 90095. E-mail: Simpson@hhmi.ucla.edu.

^{||}Present address: Department of Mathematics, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom.

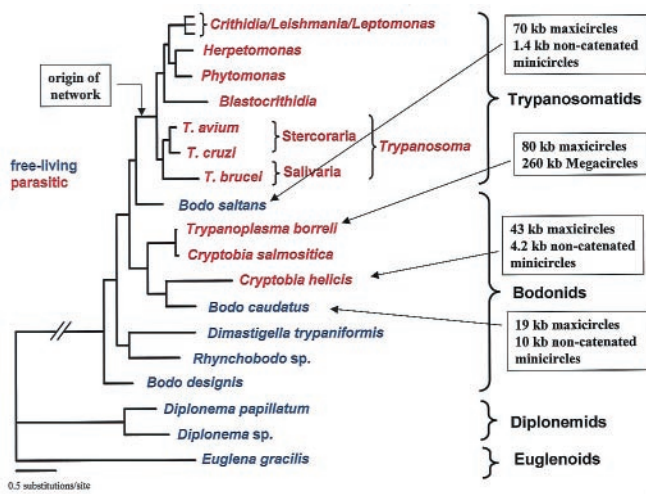


Fig. 1. Phylogenetic tree of Kinetoplastida based on SSU rRNA sequences. Only representative species for each trypanosomatid lineage are shown. The sequences of *B. designis*, *C. helicis*, and *B. saltans* are from unpublished data of D. Doležel, M. Jirků, D.A.M., and J. Lukeš. The tree was constructed by the method of maximum likelihood. The horizontal bar corresponds to 0.05 substitutions per site. This tree represents a tentative result based on a more extensive reconstruction using additional species (D. Doležel, M. Jirků, D.A.M., and J. Lukeš, unpublished results).

homologues of the informational mtDNA molecules found in other eukaryotes and contain 18 tightly clustered protein-coding genes and two rRNA genes; the gene order is conserved in all trypanosomatid species examined. No tRNA genes are encoded in the maxicircle. The mitochondrial tRNAs are encoded in the nucleus and are imported into the mitochondrion (23, 24).

The bodonids also have a single mitochondrion but the mtDNA is less structured and the molecules are not catenated. The DNA appears in thin sections as large oval fibrous structures. In *C. helicis* there are multiple nodules of DNA in the mitochondrion, an organization that has been termed pan-kinetoplastic (25, 26). The kDNA in *C. helicis* consists of 43-kb maxicircles and several thousand 4.2-kb noncatenated minicircles (26), in *B. saltans* 70-kb maxicircles and multiple noncatenated 1.4-kb minicircles [which encode two guide RNAs (gRNAs) each] (27), and in *B. caudatus* 19-kb maxicircles and 10- to 12-kb minicircles (28). The kDNA in *T. borreli*, however, contains 80-kb maxicircles and 180-kb minicircle homologues (megacircles) (29–31). Sequence information is available only for fragments of the maxicircle equivalents from *B. saltans* and *T. borreli* and for five gRNAs from *T. borreli* and 14 gRNAs from *B. saltans*. It is of interest that the *Cyb*, *COI*, and *COIII* gene order in the *T. borreli* maxicircle and also the *COII* and *ND5* gene order in *B. saltans* differ from that in the trypanosomatids, which is consistent with the large evolutionary distance separating bodonids and trypanosomatids.

U-Insertion/Deletion RNA Editing

Mechanism. The transcripts of 12 (the precise number varies with the species) of the maxicircle protein-coding genes are edited posttranscriptionally by the insertion and occasional deletion of uridine (U) residues mostly within coding regions, thereby correcting frameshifts and producing translatable mRNAs. The minicircles encode gRNAs, which are complementary to mature edited sequences if G:U as well as canonical base pairs are allowed (32). The gRNAs have 3' nonencoded oligo(U) tails that may be involved in stabilizing the initial interaction of the gRNA and the mRNA by either RNA–RNA or RNA–protein interactions (33, 34). Fifteen gRNAs are encoded in intergenic regions of the maxicircle DNA of *Leishmania tarentolae*. This division of

the mitochondrial genome into two physically separate genomes, with the RNA transcripts of one interacting with the incomplete mRNA transcripts of the other to produce translatable mRNAs is unprecedented and is suggestive of an unique evolutionary origin.

The mechanism of U-insertion/deletion editing involves a series of enzymatic cleavage-ligation steps, with the precise cleavages determined by base pairing with the cognate gRNAs (2). A single gRNA mediates the editing of a “block” of approximately 1–10 sites. Multiple overlapping gRNAs mediate the editing of a “domain” (35). The overall 3' to 5' polarity of editing site selection within a domain is determined by the creation of upstream mRNA “anchor” sequences by downstream editing (Fig. 2). Editing usually also proceeds 3' to 5' within a single block. A variable extent of “misediting” at the junction regions between fully edited and unedited sequences also has been observed, and this varies from gene to gene and from species to species (36–38). Misediting, which appears to be a consequence both of correct guiding by an incorrect gRNA and by stochastic errors in the editing process, is not deleterious, as misedited sequences appear to be re-edited correctly, in a 3' to 5' polarity (37–39).

Comparative Analysis of Editing in Different Kinetoplastid Species.

The U-insertion/deletion type of RNA editing has been detected in multiple trypanosomatid species. The extent of editing for several genes varies in different species. For example, the ND7 gene in the trypanosomes, *T. brucei* and *T. cruzi*, is pan-edited in two domains, whereas in the *Leishmania-Crithidia* clade this gene is edited only at the 5' end of each domain. The A6 gene in the trypanosomes is pan-edited, whereas in the *Leishmania-Crithidia* clade, the editing of the A6 gene shows a gradient of restriction to the 5' end of the single domain, from *L. tarentolae*, to *Herpetomonas muscarum*, to *Phytomonas serpens*, and to *Blastocrithidia culicis* (10).

To date, the deepest lineage in which U-insertion editing has been detected is the bodonid group. Because minicircles, which presumably encode gRNAs, are observed in *B. caudatus*, *B. saltans*, and *C. helicis*, this suggests that the free noncatenated state is a primitive feature. Catenation of minicircles to form the kDNA network probably arose in an ancestor of the trypanosomatids as a molecular mechanism designed to avoid minicircle losses by missegregation. Concatenation of minicircles in the 180-kb megacircle as observed in *T. borreli* might have independently arisen as another solution to the same problem. However, additional analyses of the kDNA structure in bodonids are required to shed more light on kDNA evolution.

The only mitochondrial gene isolated from the deeper branching *Diplonema* (16) and *Euglena gracilis* (17, 40) is the *COI* gene, which is unedited. In addition, no evidence was obtained for small gRNA-like molecules in *E. gracilis* mitochondria by 5' capping experiments (17). This preliminary evidence does not, of course, eliminate the possibility of editing in these cells, but the simplest hypothesis is that this type of editing evolved in the mitochondrion of an ancestral bodonid after the split from the euglenoid lineage.

Minicircle-Encoded gRNAs in Two Strains of *L. tarentolae*. The only species for which the entire complement of gRNAs is known (35) is the UC strain of *L. tarentolae*, which has been maintained as the promastigote form in culture in various laboratories for more than 60 years. There are 15 maxicircle-encoded gRNAs and 17 minicircle-encoded gRNAs in this strain. Five pan-edited genes (G1–G5) show a complete absence of productive editing in this strain, as evidenced by an inability to PCR-amplify mature edited transcripts by standard methods. These genes are productively edited in *T. brucei*. Two of the minicircle-encoded gRNAs in the *L. tarentolae* UC strain, gL19 (=gG4-III) and gB4 (=gND3-IX),

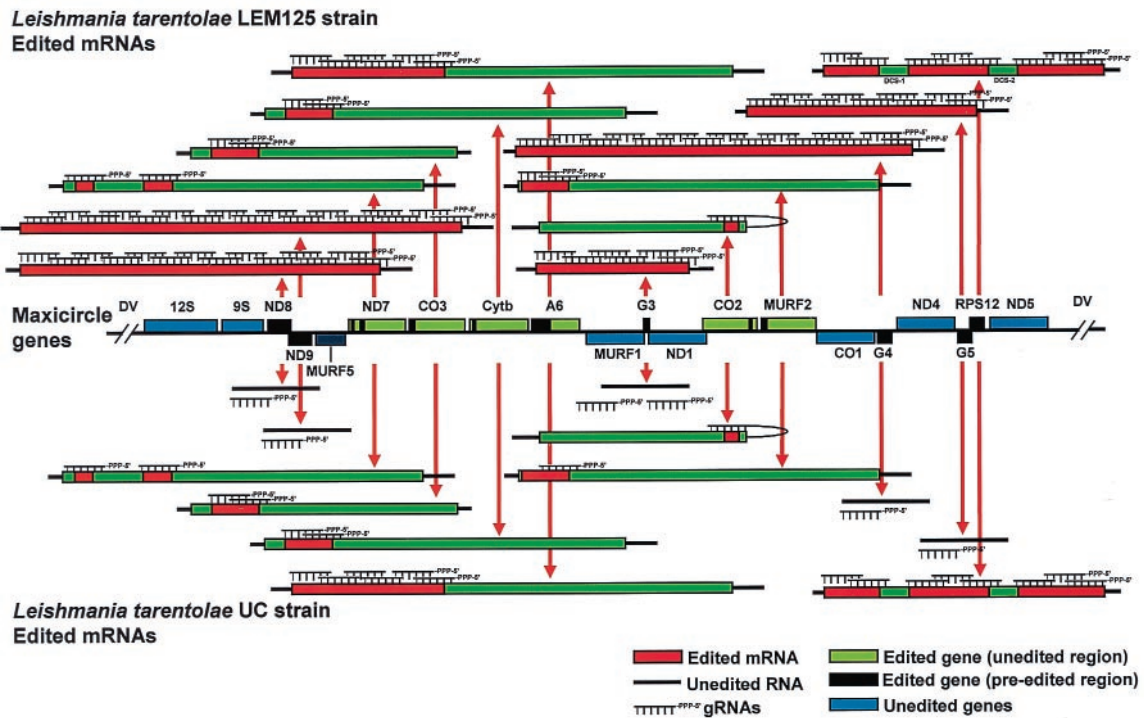


Fig. 2. Diagram of the extent of gRNA-mediated editing of maxicircle cryptogenes in the old laboratory UC strain of *L. tarentolae* and the recently isolated LEM125 strain. The overlapping gRNAs that give rise to the overall 3' to 5' editing within a domain are indicated. In the LEM125 strain, all of the approximately 80 predicted gRNAs are indicated although only 47 gRNAs have been identified.

represent nonessential gRNAs for these nonfunctional editing cascades. This was determined by analyzing the minicircle-encoded gRNA complement of LEM125, a recently isolated strain of *L. tarentolae* (41). LEM125 has the same 15 maxicircle gRNA genes but also has an estimated 80 total minicircle-encoded gRNAs, of which 30 have been cloned and sequenced and the remainder inferred to be present because of the existence of completely edited mRNA transcripts in this strain. These additional gRNAs mediate the editing of three components of complex I of the respiratory chain, *ND3*, *ND8*, and *ND9*, and also two unidentified genes, which were termed G-rich regions 3 and 4 (G3 and G4). It was proposed that multiple gRNA-encoding minicircle sequence classes had been lost from the UC strain probably because of a lack of a requirement for complex I activity in culture (41, 42). The presence of productively edited *ND8*, *ND9*, G3 (=CR3), G4 (=CR4), and *ND3* mRNAs in *T. brucei* and the presence of productively edited G3 mRNA in *P. serpens* (D.A.M., unpublished results) implies that the corresponding minicircle-encoded gRNAs also exist in these species, and this provides phylogenetic evidence for our hypothesis that the ancestral cell had a complete complement of minicircle classes. In addition, the presence of two minicircle-encoded gRNAs, gG4-III and gND3-IX, in the UC strain, which are remnants of the complete editing cascade of gRNAs for these two genes in LEM125, corroborates this evidence. To propose a loss of multiple minicircle classes from the UC strain is also more parsimonious than to propose a gain of multiple classes in the LEM125 strain. And finally, the existence of a 5' terminal block of misedited sequence in the LEM125 *ND3* mRNA (41) is indicative that this gene originally was completely edited and has lost the 5' terminal gRNA.

Minicircles from *L. tarentolae* and other members of the *Leishmania-Crithidia* clade contain a single gRNA gene situated at a constant distance from the origin of replication (43, 44). Minicircles from *T. brucei*, however, also have a single origin of replication but contain three gRNA genes situated between

18-mer inverted repeats (45), and minicircles from *T. cruzi* contain four gRNA genes situated within each of the four variable regions between four origins of replication (46). The total number of different minicircle sequence classes in *T. brucei* is estimated to be 200–300 (47), which would yield a total of 600–900 gRNAs. Although only 72 gRNAs have been identified so far in *T. brucei* (48), it is clear that there are extensive redundant gRNAs, which are gRNAs of different sequence but possessing the identical editing information because of the allowed G:U base pairing (49). In fact, 28 of the 72 identified gRNAs are redundant over the entire length of the gRNA. Only a single redundant gRNA pair has been observed in *L. tarentolae* (41). *T. brucei* also contains gRNAs with several mismatches in the anchor or guiding regions, which may be nonfunctional, but there is no evidence for or against this suggestion.

Retroposition Model for Loss of Editing in Evolution. Based on these observations and on the known 3'-5' polarity of editing, a retroposition model was proposed to explain both the gradual restriction of editing to the 5' end of domains and the complete loss of editing in some cases (42, 50). We proposed that partially edited mRNAs were being frequently converted to cDNAs by a postulated mitochondrial reverse transcriptase activity, and those cells that had replaced the original pan-edited cryptogene with a partially edited gene would survive a loss of an entire minicircle sequence class encoding a specific gRNA involved in that editing cascade. The retention of editing at the 5' end of a domain may allow regulation of translation by creation of a methionine initiation codon and a possible ribosome-binding site. This model is based on the assumption that minicircles are distributed randomly to daughter cells upon cell division.

Replication and Segregation of Minicircles. One possible mechanism involved in the random distribution of minicircles is the mode of replication and segregation. The mitochondrial S phase is fairly synchronous with the nuclear S phase, although the

kinetoplast network physically divides just before the nucleus (51). Closed minicircles are apparently randomly removed from the side of the network facing the basal body by a topoisomerase II activity and migrate by an unknown mechanism to one of two replisomes (52) that are located at the two antipodes of the kDNA nucleoid body (52–54). After replication, the daughter molecules remain nicked or gapped, which may be a mechanism to ensure replication of each minicircle. The daughter minicircles then are recatenated into the periphery of the network. There is microscopic evidence that the networks in *Leishmania* and *Crithidia* (and also *T. cruzi*) are actually rotating, and this movement produces a complete peripheral distribution of newly replicated minicircles (55–57). The networks in the middle of S phase consist of an expanding ring of nicked circles and a central core of closed circles, and at the end of S phase the networks consist entirely of nicked circles. The minicircles then become closed and then the network segregates into two daughter networks as the single mitochondrion divides (58). This mechanism of replication appears to introduce a certain amount of randomness into minicircle segregation. In other words, sister minicircles may not necessarily end up in different daughter cells. A pulse–chase experiment performed with *C. fasciculata* cells at the light microscope level previously showed that newly replicated minicircles are spread throughout the network after one cell cycle is completed (59).

In the case of *T. brucei*, the network apparently does not rotate and two dumbbell-shaped masses of nicked replicated minicircles accumulate at either end of the nucleoid body, which then divides in half into the daughter cells (56, 57). In this case there does not appear to be a mechanism for randomization throughout the network, other than the possible random selection and migration to the antipodal replisomes.

Plasticity of Minicircle Sequence Class Copy Number in *L. tarentolae* in Culture. The number of minicircles per network in *L. tarentolae* was assayed by counting 4',6-diamidino-2-phenylindole (DAPI)-stained networks in a cell counting chamber using a fluorescent microscope and measuring the DNA concentration spectrophotometrically. Quantitative dot blot hybridization using an oligonucleotide probe complementary to the conserved CSB-3 12-mer sequence yielded values of $12,600 \pm 300$ and $12,700 \pm 800$ for the UC and LEM125 strains, respectively. Similar dot blot hybridization analysis showed that the copy number of maxicircle DNA molecules was very similar in the UC and LEM125 strains (32 ± 2 and 25 ± 2 copies per network, respectively).

Quantitation of the copy numbers of 17 specific minicircle sequence classes in the UC strain was previously performed by Southern blot analysis using specific oligonucleotide probes for specific gRNAs. We have repeated these analyses with both UC strain kDNA and LEM125 strain kDNA, by dot blot hybridization of *MspI*-digested kDNA (all minicircles have at least one *MspI* site), and a known amount of specific cloned minicircles using primers specific to each gRNA. A primer to the conserved 12-mer sequence was used as a loading control. The results in Table 1 show that homologous minicircle sequence class frequencies are extremely variable, both between strains and between different kDNA isolates from the same strain taken after several years of culture. In general, the LEM125 strain kDNA exhibited lower copy numbers for the sequence classes in common between the strains, which is consistent with LEM125 possessing a more complex minicircle repertoire.

In the UC strain kDNA as mentioned above, two gRNAs, pLt19 (= G4-III) and pB4 (= gND3-IX), are nonfunctional, in that all of the other minicircle-encoded gRNAs in those editing cascades are missing from this strain. It is of interest that these nonfunctional minicircles showed the greatest plasticity in frequency.

As was found previously (35), there was no correlation of

Table 1. Minicircle copy numbers in kDNA networks from UC and LEM125 strains of *L. tarentolae*

Minicircle class	UC		LEM 125
	% Minicircles per network		% Minicircles per network
	1992*	1994	
RPS12-I	3.7	4.7	0.2
RPS12-II	1.4	0.2	0.8
RPS12-III	10.5	1.5	2.5
RPS12-IV	2.1	0.4	0.8
RPS12-V	5.0	0.2	0.4
RPS12-VII	0.9	15.0	2.8
RPS12-VIII	0.3	0.9	0.1
A6-I	2.2	0.4	6.0
A6-II	1.1	0.4	1.8
A6-III	3.8	1.0	0.5
A6-IV	3.3	0.1	0.1
A6-V	2.1	2.0	0.2
A6-VI	3.2	0.8	1.0
COIII-I	1.9	0.5	0.8
COIII-II	3.7	1.7	3.6
ND8-I	—	ND	4.8
ND8-II	—	ND	9.2
ND8-IV	—	ND	3.9
G4-III(Lt19)	25	66.8	1.7
ND3-II	—	ND	0.2
ND3-III	—	ND	6.4
ND3-V	—	ND	2.8
ND3-VI	—	ND	5.4
ND3-IX (B4)	29.8	3.4	3.1

ND, not detected above background.

*Data taken from ref. 35.

minicircle copy number and gRNA relative abundance (data not shown).

Computer Simulations of Minicircle Sequence Class Plasticity. Using a population dynamics model of minicircle segregation, Savill and Higgs (60) recently have shown that random segregation can indeed account for much of the above experimental observations on minicircle plasticity. The copy number of every minicircle class in every cell in a population is tracked over many generations. In every generation each cell replicates its minicircles, hence doubling the copy number of all its classes. Then the cells divide and the daughter cells receive a certain number of copies of each class. The actual number of copies is randomly chosen according to a binomial distribution that models a purely random segregation process. All daughter cells that receive the full complement of minicircle classes and have fewer than 12,000 minicircles in total are randomly chosen to populate the next generation up to a maximum population size. These two conditions model the reasonable assumptions that (i) if a cell does not receive any copies of a particular class it is therefore missing a gRNA and hence its mRNA cannot be correctly edited, which is assumed to be lethal, and (ii) the network is restricted in its maximum size because of physical constraints.

A typical simulation of a hypothetical species with 17 minicircle classes is shown in Fig. 3. It clearly demonstrates that random segregation causes fluctuations in the average minicircle class copy number from one generation to the next. Moreover, it also leads to the experimentally observed distribution of many

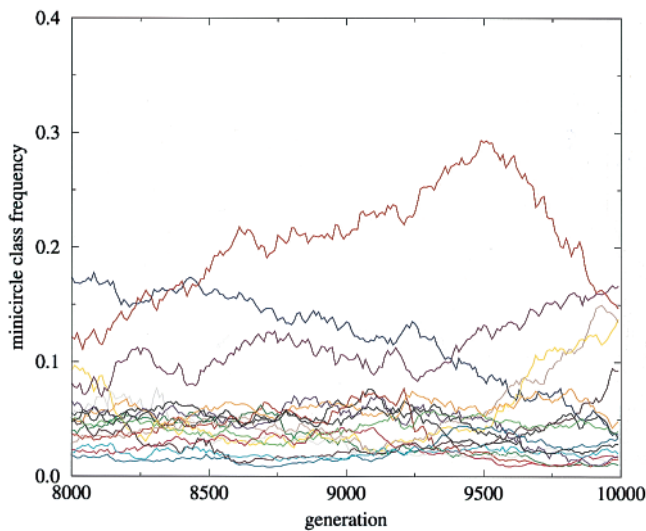


Fig. 3. The average frequencies of 17 minicircle classes undergoing random segregation over 2,000 generations. Random segregation causes fluctuations in the frequencies, giving rise to the experimentally observed distribution of many classes having low frequency and few classes having high frequency. Initially at generation zero, every class in every cell has 170 copies. There are 1,000 cells that have maximum network sizes of 12,000 minicircles.

classes having very low copy number and a few having very high copy number. No two runs are ever the same, thus explaining why homologous minicircle classes in different strains have different copy numbers.

The loss of minicircle classes during the long culture history of the UC strain also was modeled by starting with 70 classes, of which 15 are required and 55 are not. Fig. 4 shows the number of generations for each unnecessary class to be lost, from the time when the UC strain was first cultured. Many classes are lost fairly rapidly within the first few hundred generations, but it takes successively longer for the remaining classes to be lost, and the last few classes may take tens of thousands of generations to be lost. Moreover, by averaging over many simulations we found that the last remaining unnecessary class was also the most abundant class in 27% of cases. Therefore, random segregation can explain the observed long persistence time of unnecessary classes and their high abundance. However, as shown in Fig. 3, the highest frequency achieved by the most abundant class for a hypothetical species with 17 classes (similar to the UC strain) only reaches about 30% and never as large as the 67% observed in the 1,994 UC cells. This large abundance of one class is similar to the situation in the CFC1 strain of *C. fasciculata*, in which one minicircle sequence class shows over 90–95% abundance. It appears that random segregation alone cannot explain the large abundance of these classes, and therefore other selective forces must be present.

If the additional following assumptions are made, simulations can explain the experimental results: (i) The network has a minimum allowable size. If the network is too small, it may not abut the replisomes. (ii) The number of copies of each necessary minicircle class is regulated by an unknown mechanism. (iii) The number of copies of each unnecessary minicircle class is unregulated, i.e., once a minicircle becomes unnecessary—by loss of other gRNAs in a cascade, its copy number is not regulated and can vary freely. The model is modified so that if the total number of minicircles in a daughter cell falls below a predetermined threshold or if the copy number of each necessary class exceeds a predetermined threshold, the cell does not survive into the next generation. For simplicity, in the model this threshold is set to the same value for all necessary classes, but in reality it may vary between classes. The lower threshold for each necessary class is

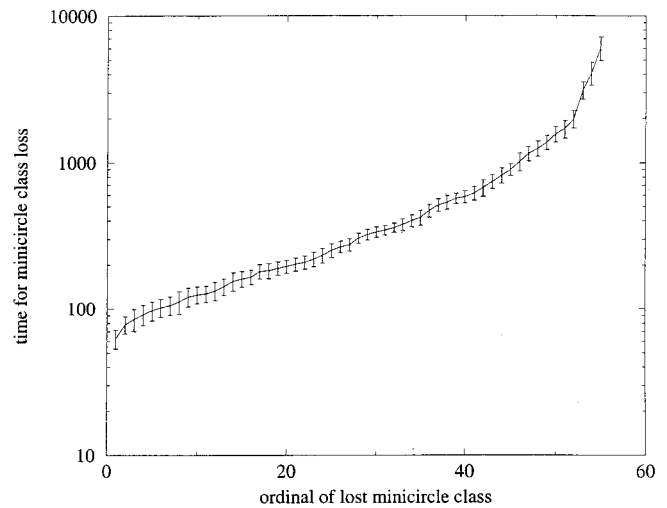


Fig. 4. The number of generations for consecutive unnecessary minicircle classes to be lost. The last few classes take many thousands of generations to be lost. The simulations are initially run for 2,000 generations with all classes being necessary. This is to lose the artificial initial conditions. Then, 55 classes become unnecessary, i.e., if their frequencies reach zero in a cell, the cell is still viable. The loss time was averaged over 10 simulations; error bars show ± 2 SEM. Previously published in ref. 60.

one copy, as in the original model. Again, in reality this may not be true. Fig. 5 shows a simulation where the minimum kinetoplast threshold size is set to 10,000 minicircles and the upper threshold for necessary classes is set to 200. Fifteen classes are required and 55 are not. Initially all classes have the same copy number of 170, giving a total of 11,900 minicircles per cell, which lies between the assumed lower and upper thresholds for the kinetoplast size (i.e., 10,000 and 12,000 minicircles, respectively). The figure shows the cumulative proportion of minicircles of all necessary classes, all unnecessary classes, and the proportion of the most abundant class. Initially the proportions are 21% ($170 \times 15/11,900$), 79% ($170 \times 55/11,900$), and 1.4% ($170/11,900$), respectively. Because of assumption ii, the proportion of minicircles of necessary classes cannot exceed 30% ($15 \times 200/10,000$). Therefore, the unnecessary classes must make up the difference for the kinetoplasts to maintain their minimum sizes. However, as unnecessary classes are lost because of random segregation over time, there are fewer classes that can make up this difference. Finally, there will be only one unnecessary class left to make up at least 70% of the minicircles in the kinetoplast. This class is now necessary only to maintain kinetoplast size, and the function of encoding gRNAs has now been replaced by a buffering function. By adjusting the parameters, it is even possible to obtain an unnecessary class with over 90% abundance, as in the CFC1 strain (58). This successful simulation of large frequencies for unnecessary minicircle classes actually provides support for assumption ii.

The model of random segregation also makes several interesting predictions. At every generation some daughter cells become unviable and do not survive into the next generation because they do not receive the full complement of minicircle classes. Hence some fraction of the total population of daughter cells is viable; we term this the daughter cell viability. We find that cell viability increases with increasing kinetoplast size and decreasing number of minicircle sequence classes (Fig. 6). If the cells have some mechanism that more evenly segregates sister minicircles between daughter cells, cell viability increases. This implies that there could be some selection pressure for trypanosomatids to segregate their minicircles more evenly, which may have led to the development of the rotating network in the *Leishmania-Crithidia* clade.

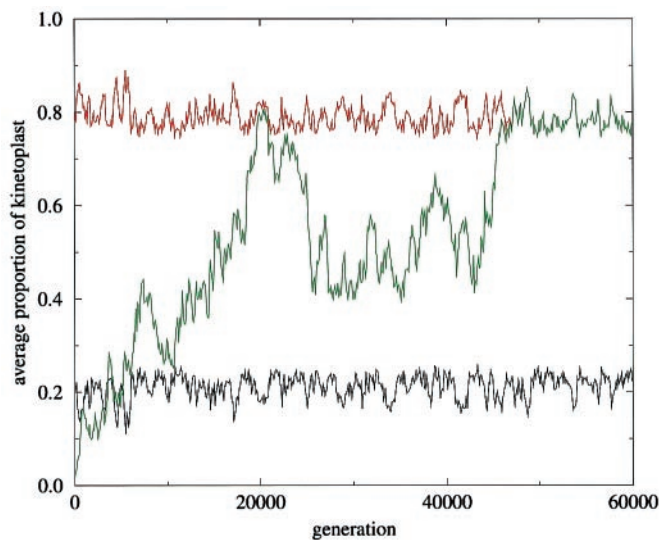


Fig. 5. As unnecessary classes are lost, one unnecessary class must become highly abundant to maintain the size of the kinetoplast, i.e., 10,000 minicircles. This is because the 15 necessary classes are restricted to a maximum of only 200 copies. Initially all classes have a copy number of 170, and 55 classes are unnecessary and 15 are necessary. The red line shows the cumulative frequency of all unnecessary classes, the black line the cumulative frequency of all necessary classes, and the green line the frequency of the most abundant class at any given time.

In the case of *T. brucei*, random segregation of the 250+ sequence classes would lead to a predicted cell viability in this model of less than 0.5, and hence population extinction. However, incorporating the information that each minicircle in this species encodes multiple gRNAs and that genetic exchange occurs, it has been shown that the model can produce the observed situation of evolutionary viability and multiple redundant and nonfunctional gRNAs (N.J.S. and P. G. Higgs, unpublished results). Mutation of the gRNA genes and drift in the minicircle copy numbers lead to an ever-increasing number of necessary classes encoding ever fewer functional gRNAs per minicircle.

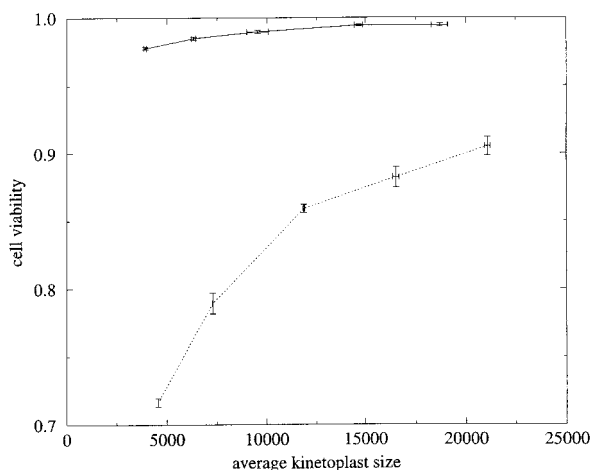


Fig. 6. Cell viability increases as the average kinetoplast size increases and as the number of minicircle classes decreases (dotted line 70 classes, solid line 17 classes). This is because classes have more copies and hence there is more chance that a daughter cell receives the full complement of classes. Averages were taken over 10 simulations; error bars show ± 2 SEM. Previously published in ref. 60.

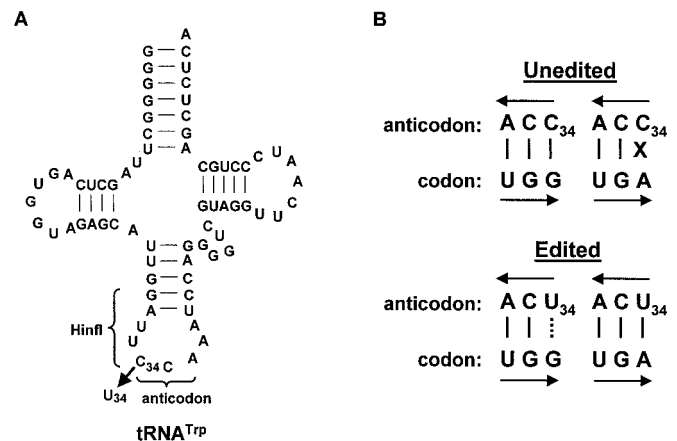


Fig. 7. C to U editing of the anticodon of the mitochondrial-imported tRNA^{Trp}. (A) tRNA^{Trp} showing the editing of C34. The *HinfI* site that is destroyed by the C to U editing event is also indicated. (B) The C34 to U34 editing allows the decoding of the UGA codon as tryptophan.

C to U Editing and the Origin of Uridine-Insertion Editing in Trypanosomes

UGA Codon Reassignment. Kinetoplastids use a nonuniversal genetic code in which the UGA stop codon is read as tryptophan (61). The codon capture hypothesis (62, 63) proposes that evolutionary reassignment of a stop codon involves first the disappearance of the stop codon and replacement with synonymous codons, then mutations in the peptide chain release factor so as not to recognize the stop codon, and finally duplication and mutation of a tRNA gene to allow decoding of the old codon with a new meaning. The occurrence of a nonuniversal genetic code in mitochondrial genomes is thought to be a derived character that arose independently in different organisms. In the Euglenozoa phylum, the use of a nonuniversal code is limited to the kinetoplastids (and diplomonads) (16, 17). However, the appearance of a new gene for a tRNA decoding UGA for tryptophan did not occur in these species, perhaps because of the early loss of all mitochondrial-encoded tRNA genes.

C to U Editing of tRNA^{Trp}. Alfonzo *et al.* (4) recently reported that, at least in the case of *L. tarentolae*, the problem of decoding UGA was solved by evolving an editing activity that changes the first position of the anticodon of the mitochondrial imported tRNA^{Trp} from C to U (CCA to UCA in the anticodon), thereby allowing the decoding of UGA codon as tryptophan (Fig. 7). The evidence for this editing involved the observation of a loss of a *HinfI* restriction site in a cDNA copy of the mitochondrial tRNA^{Trp}, which was confirmed by sequencing the reverse transcription-PCR-amplified product, and by direct analysis of the mitochondrial tRNA^{Trp} by poisoned primer extension experiments. More than 40% of the mitochondrial tRNA^{Trp} is edited at C34. A C to U editing event also has been described for the cytosolic 7SL RNA in *Leptomonas* (64). C to U editing is found in many phylogenetically diverse organisms, both in organellar and nuclear genomes, suggesting that this site-specific modification represents an ancient evolutionary activity (65–67).

The following hypothetical scenario could explain the origin of this tRNA editing and the tryptophan codon change in trypanosomes. We propose that tRNA importation into the mitochondrion was developed at a very early stage of evolution and that tRNA genes in the kDNA subsequently were lost because of redundancy. The original state included the encoding of tryptophan by UGG and the CCA anticodon in the tRNA. We also assume that a pre-existing activity performing some other function in the cell produced a promiscuous C to U modification in the anticodon of the imported tRNA^{Trp} at a low frequency. G

Table 2. Tryptophan and stop codons in *L. tarentolae* mitochondria

Gene	Trp			Created by editing			Stop	
	UGA	UGG	Total	UGA	UGG	Total	UAA	UAG
ND8	1	0	1	0	0	0		X*
ND9	2	1	3	1	0	2		X*
MURF5	1	1	2	0	0	0	X	
ND7	3	0	3	0	0	0	X	
COIII	8	0	8	0	0	0	X	
Cyb	15	1	16	0	0	0	X	
MURF4	2	0	2	2	0	2	X	
MURF1	4	1	5	0	0	0	X	
G3	0	1	1	0	1	0	X	
ND1	3	0	3	0	0	0	X	
COII	7	0	7	0	0	0		X
MURF2	3	0	3	0	0	0	X	
COI	13	1	14	0	0	0	X	
G4	1	2	3	1	2	3	X	
ND4	8	1	9	0	0	0		X
G5	2	2	4	2	2	4	X	
RPS12	1	0	1	1	0	1	X	
ND5	15	1	16	0	0	0	X	
Total	89	12	101	7	5	12	14	4 (2*)

*Created by editing.

to A transition mutations, perhaps driven by AT mutational pressure, led to the replacement of UGA with UAA stop codons and this was followed by mutations that affected the interaction of release factor with UGA. Similar mutational pressure led to the replacement of TGG tryptophan codons with TGA in essential mitochondrial genes, and this would have made the C to U tRNA editing indispensable for cell survival. This scenario combines the model of Covello and Gray (68) for the evolution of RNA editing systems in general and a modified codon capture hypothesis (62). The editing of an imported tRNA offers a new mechanism for codon capture that does not require a gene duplication event. The alternative hypothesis of duplication and mutation of an existing nuclear-encoded tRNA gene [tRNA^{Trp} (CCA)] did not occur perhaps because of the problems involved in maintaining a suppressor tRNA [tRNA^{Trp}(UCA)] in the cytosol.

The Relationship of C to U Editing and U-Insertion Editing. It is interesting that 7% of the UGA tryptophan codons are created by U-insertion editing (Table 2). This observation places some time constraints on hypotheses for the appearance of U-insertion editing. We previously have proposed a scenario for the origin of U-insertion editing (69) based on the models of Covello and Gray (68) and Cavalier-Smith (5), which involved the pre-existence of editing enzymatic activities that were used for other biochemical functions, genetic drift in a mitochondrial gene, appearance of complementary gRNAs by partial gene duplication and antisense transcription, and finally utilization of editing for gene regulation. Stoltzfus (70) has pointed out that DNA polymerases have a bias toward single nucleotide deletions and that, for this reason, an increase in the number of edited sites is more likely than a loss of an editing site. The same author also proposed that if gRNAs arose by duplication and antisense transcription, this must have occurred before the genetic drift that gave rise to the pre-edited sequence, because gRNAs are complementary to the edited sequence and not to the pre-edited sequence. If one accepts this proposal, then the observed guiding of U-insertions to produce UGA tryptophan codons suggests that this codon reassignment also occurred before the appearance of U-insertion editing.

It should be noted that the presence of guiding G residues in

the gRNAs that base pair with inserted Us presents a potential problem for the gene duplication scenario for the origin of gRNAs. One possible mechanism could have been deamination of the guiding A to produce inosine by an adenosine deaminase acting on RNA-like activity (71–74) and retroposition of the mutated gRNA back into the genome to replace the original gRNA gene. Another mechanism could have been a replication-associated deletion of a C in a mitochondrial gene that was corrected by U-specific insertional editing activity at the transcript level guided by the original G residue in the gRNA.

An alternative scenario for the origin of gRNAs can be derived from an analysis of computational algorithms for searching for possible gRNA genes that are complementary to candidate cryptogenes (75) in which it was shown that known gRNA sequences are in or very close to the statistical noise. Based on these results, one could speculate that the primordial gRNA was derived from some other mitochondrial RNA fragment that by chance base-paired with the mRNA downstream of the U-deletion site and contained a guiding A or G residue that could base-pair with an inserted U and thereby overcome this frameshift mutation and allow translation of the mRNA.

Conclusions

The mitochondrial genome of the kinetoplastid protists is a highly derived genome in which frameshift errors in the reading frames of 12 of the 18 genes are corrected at the RNA level by U-insertion/deletion editing, which probably arose in the early bodonid-kinetoplast lineage after divergence of the euglenoids. The sequence information for these corrections is partially located in a physically separate guide RNA genome. The most primitive type of organization of this genome may have been similar to that seen in the bodonids, *B. saltans*, *B. caudatus*, and *C. helicis*, in which the gRNA genes are present on multiple plasmid-like molecules. The next steps in evolution may have either been a concatenation of the plasmids into megacircles such as in *T. borreli* or a catenation of the plasmids into a network such as in the trypanosomatids.

The fact that that daughter cells must receive a complete complement of all of the minicircle sequence classes encoding the gRNAs required for editing has led to the evolution of mechanisms for the random distribution of minicircles within the single network. The highly structured organization of the catenated minicircles within the network must have placed additional constraints on the evolution of this system. Two different types of mechanisms evolved, both based on a decatenation of minicircles from the network and replication at two antipodal nodes before recatenation of the daughter molecules. In the *Leishmania-Crithidia* clade (and *T. cruzi*), random selection of closed molecules and rotation of the network during the recatenation process in S phase produced a high degree of randomization. We have shown that computer simulations provide evidence that random segregation of minicircles during replication can account for many of the phenomena observed in *L. tarentolae* and possibly for the observed restriction of editing that has occurred in evolution. However, to explain the high abundance of unnecessary minicircle classes in the UC strain of *L. tarentolae* and the CFC1 strain of *C. fasciculata*, further assumptions need to be made concerning the regulation of minicircle copy number.

In the *T. brucei* clade in which the network does not rotate, the randomization is mainly a function of random selection of closed molecules. However, the homogenizing effect of genetic exchange that occurs in the tsetse vector (76, 77), but which does not appear to occur at a detectable level in *Leishmania*, is another factor that may affect random distribution of minicircles to daughter cells. Relevant to this is the fact that two closely related trypanosome species, *T. equiperdum* and *T. evansi*, which have lost the sexual cycle in the fly and are transmitted by sexual intercourse

and by mechanism transmission by tabanid flies, respectively, have networks consisting of one of several single minicircle sequence classes and mutated or deleted maxicircle DNA (78–83).

Another derived feature of the kinetoplastid mitochondrial genome is the complete lack of tRNA genes and the importation of all mitochondrial tRNAs from the cytosol (23). To decode UGA as tryptophan, the imported tRNA^{Trp} is edited by a C to U modification within the anticodon (4).

RNA editing appears to have arisen in evolution multiple times in different organisms as a way to correct errors and modulate genetic sequences at the RNA level. In kinetoplastid protists, two types of editing that apparently arose early in the evolution of the kinetoplast-mitochondrion are intimately tied in with the unusual mitochondrial genome unique to these organisms. This provides yet another example of the evolutionary diversity of lower eukaryotes.

1. Smith, H. C., Gott, J. M. & Hanson, M. R. (1997) *RNA* **3**, 1105–1123.
2. Alfonzo, J. D., Thiemann, O. & Simpson, L. (1997) *Nucleic Acids Res.* **25**, 3751–3759.
3. Stuart, K., Kable, M. L., Allen, T. E. & Lawson, S. (1998) *Methods* **15**, 3–14.
4. Alfonzo, J. D., Blanc, V., Estevez, A. M., Rubio, M. A. & Simpson, L. (1999) *EMBO J.* **18**, 7056–7062.
5. Cavalier-Smith, T. (1997) *Trends Genet.* **13**, 6–9.
6. Philippe, H. & Forterre, P. (1999) *J. Mol. Evol.* **49**, 509–523.
7. Budin, K. & Philippe, H. (1998) *Mol. Biol. Evol.* **15**, 943–956.
8. Germot, A. & Philippe, H. (1999) *J. Eukaryotic Microbiol.* **46**, 116–124.
9. Fernandes, A. P., Nelson, K. & Beverley, S. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11608–11612.
10. Maslov, D. A., Avila, H. A., Lake, J. A. & Simpson, L. (1994) *Nature (London)* **365**, 345–348.
11. Landweber, L. F. & Gilbert, W. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 918–921.
12. Alvarez, F., Cortinas, M. N. & Musto, H. (1996) *Mol. Phylogenet. Evol.* **5**, 333–343.
13. Lukes, J., Jirku, M., Dolezel, D., Kral'ová, I., Hollar, L. & Maslov, D. A. (1997) *J. Mol. Evol.* **44**, 521–527.
14. Haag, J., O'Huigin, C. & Overath, P. (1998) *Mol. Biochem. Parasitol.* **91**, 37–49.
15. Wright, A. D., Li, S., Feng, S., Martin, D. S. & Lynn, D. H. (1999) *Mol. Biochem. Parasitol.* **99**, 69–76.
16. Maslov, D. A., Yasuhira, S. & Simpson, L. (1999) *Protist* **150**, 33–42.
17. Yasuhira, S. & Simpson, L. (1997) *J. Mol. Evol.* **44**, 341–347.
18. Paulin, J. J. (1975) *J. Cell Biol.* **66**, 404–413.
19. Simpson, L. & Kretzer, F. (1997) *Mol. Biochem. Parasitol.* **87**, 71–78.
20. Simpson, L. (1986) *Int. Rev. Cytol.* **99**, 119–179.
21. Borst, P., Fase-Fowler, F., Weijers, P., Barry, J., Tetley, L. & Vickerman, K. (1985) *Mol. Biochem. Parasitol.* **15**, 129–142.
22. Yurchenko, V., Hobza, R., Benada, O. & Lukes, J. (1999) *Exp. Parasitol.* **92**, 215–218.
23. Simpson, A. M., Suyama, Y., Dewes, H., Campbell, D. & Simpson, L. (1989) *Nucleic Acids Res.* **17**, 5427–5445.
24. Hancock, K. & Hajduk, S. L. (1990) *J. Biol. Chem.* **265**, 19208–19215.
25. Vickerman, K. (1977) *J. Protozool.* **24**, 221–233.
26. Lukes, J., Jirku, M., Avliyakov, N. & Benada, O. (1998) *EMBO J.* **17**, 838–846.
27. Blom, D., De Haan, A., Van den Berg, M., Sloof, P., Jirku, M., Lukes, J. & Benne, R. (1998) *Nucleic Acids Res.* **26**, 1205–1213.
28. Hajduk, S., Siqueira, A. & Vickerman, K. (1986) *Mol. Cell Biol.* **6**, 4372–4378.
29. Maslov, D. A. & Simpson, L. (1994) *Mol. Cell Biol.* **14**, 8174–8182.
30. Lukes, J., Arts, G. J., Van den Burg, J., De Haan, A., Opperdoes, F., Sloof, P. & Benne, R. (1994) *EMBO J.* **13**, 5086–5098.
31. Yasuhira, S. & Simpson, L. (1996) *RNA* **2**, 1153–1160.
32. Blum, B., Bakalara, N. & Simpson, L. (1990) *Cell* **60**, 189–198.
33. Blum, B. & Simpson, L. (1990) *Cell* **62**, 391–397.
34. Kapushoc, S. T. & Simpson, L. (1999) *RNA* **5**, 656–669.
35. Maslov, D. A. & Simpson, L. (1992) *Cell* **70**, 459–467.
36. Decker, C. J. & Sollner-Webb, B. (1990) *Cell* **61**, 1001–1011.
37. Sturm, N. R. & Simpson, L. (1990) *Cell* **61**, 871–878.
38. Sturm, N. R., Maslov, D. A., Blum, B. & Simpson, L. (1992) *Cell* **70**, 469–476.
39. Byrne, E. M., Connell, G. J. & Simpson, L. (1996) *EMBO J.* **15**, 6758–6765.
40. Tessier, L. H., Van der Speck, H., Gualberto, J. M. & Grienenberger, J. M. (1997) *Curr. Genet.* **31**, 208–213.
41. Thiemann, O. H., Maslov, D. A. & Simpson, L. (1994) *EMBO J.* **13**, 5689–5700.
42. Simpson, L. & Maslov, D. A. (1994) *Science* **264**, 1870–1871.
43. Sturm, N. R. & Simpson, L. (1990) *Cell* **61**, 879–884.
44. Yasuhira, S. & Simpson, L. (1995) *RNA* **1**, 634–643.
45. Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K. & Hajduk, S. L. (1990) *Cell* **63**, 783–790.
46. Avila, H. & Simpson, L. (1995) *RNA* **1**, 939–947.
47. Stuart, K. (1979) *Plasmid* **2**, 520–528.
48. Souza, A. E., Hermann, T. & Göringer, H. U. (1997) *Nucleic Acids Res.* **25**, 104–106.
49. Corell, R. A., Feagin, J. E., Riley, G. R., Strickland, T., Guderian, J. A., Myler, P. J. & Stuart, K. (1993) *Nucleic Acids Res.* **21**, 4313–4320.
50. Landweber, L. F. (1992) *BioSystems* **28**, 41–45.
51. Simpson, L. & Braly, P. (1970) *J. Protozool.* **17**, 511–517.
52. Ferguson, M., Torri, A. F., Ward, D. C. & Englund, P. T. (1992) *Cell* **70**, 621–629.
53. Simpson, A. & Simpson, L. (1976) *J. Protozool.* **23**, 583–587.
54. Shapiro, T. A. & Englund, P. T. (1995) *Annu. Rev. Microbiol.* **49**, 117–143.
55. Guilbride, D. L. & Englund, P. T. (1998) *J. Cell Sci.* **111**, 675–679.
56. Ferguson, M. L., Torri, A. F., Pérez-Morga, D., Ward, D. C. & Englund, P. T. (1994) *J. Cell Biol.* **126**, 631–639.
57. Robinson, D. R. & Gull, K. (1994) *J. Cell Biol.* **126**, 641–648.
58. Pérez-Morga, D. & Englund, P. T. (1993) *J. Cell Biol.* **123**, 1069–1079.
59. Simpson, L., Simpson, A. & Wesley, R. (1974) *Biochim. Biophys. Acta* **349**, 161–172.
60. Savill, N. J. & Higgs, P. G. (1999) *Proc. R. Soc. London Ser. B* **266**, 611–620.
61. de la Cruz, V., Neckelmann, N. & Simpson, L. (1984) *J. Biol. Chem.* **259**, 15136–15147.
62. Inagaki, Y., Ehara, M., Watanabe, K. I., Hasashi-Ishimaru, Y. & Ohama, T. (1998) *J. Mol. Evol.* **47**, 378–384.
63. Osawa, S., Jukes, T. H., Watanabe, K. & Muto, A. (1992) *Microbiol. Rev.* **56**, 229–264.
64. Ben Shlomo, H., Levitan, A., Shay, N. E., Goncharov, I. & Michaeli, S. (1999) *J. Biol. Chem.* **274**, 25642–25650.
65. Morl, M., Dorner, M. & Paabo, S. (1995) *Nucleic Acids Res.* **23**, 3380–3384.
66. Navaratnam, N., Bhattacharya, S., Fujino, T., Patel, D., Jarmuz, A. L. & Scott, J. (1995) *Cell* **81**, 187–195.
67. Covello, P. S. & Gray, M. W. (1989) *Nature (London)* **341**, 662–666.
68. Covello, P. S. & Gray, M. W. (1993) *Trends Genet.* **9**, 265–268.
69. Simpson, L. & Maslov, D. A. (1999) *Ann. N.Y. Acad. Sci.* **870**, 190–205.
70. Stoltzfus, A. (1999) *J. Mol. Evol.* **49**, 169–181.
71. Yang, J. H., Sklar, P., Axel, R. & Maniatis, T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 4354–4359.
72. Gerber, A., Grosjean, H., Melcher, T. & Keller, W. (1998) *EMBO J.* **17**, 4780–4789.
73. Polson, A. G., Bass, B. L. & Casey, J. L. (1996) *Nature (London)* **380**, 454–456.
74. Keller, W., Wolf, J. & Gerber, A. (1999) *FEBS Lett.* **452**, 71–76.
75. Von Haeseler, A., Blum, B., Simpson, L., Sturm, N. & Waterman, M. S. (1992) *Nucleic Acids Res.* **20**, 2717–2724.
76. Bogliolo, A. R., Lauria-Pires, L. & Gibson, W. C. (1996) *Acta Trop.* **61**, 31–40.
77. Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A. & Turner, C. M. (1999) *Mol. Biochem. Parasitol.* **104**, 1–9.
78. Shu, H.-H. & Stuart, K. (1994) *Nucleic Acids Res.* **22**, 1696–1700.
79. Barrois, M., Riou, G. & Galibert, F. (1982) *Proc. Natl. Acad. Sci. USA* **78**, 3323–3327.
80. Borst, P., Fase-Fowler, F. & Gibson, W. (1987) *Mol. Biochem. Parasitol.* **23**, 31–38.
81. Lun, Z.-R., Brun, R. & Gibson, W. (1992) *Mol. Biochem. Parasitol.* **50**, 189–196.
82. Songa, E. B., Paindavoine, P., Wittouck, E., Viseshakul, N., Muldermans, S., Steinert, M. & Hamers, R. (1990) *Mol. Biochem. Parasitol.* **43**, 167–180.
83. Frasch, A., Hajduk, S., Hoeijmakers, J., Borst, P., Brunel, F. & Davison, J. (1980) *Biochim. Biophys. Acta* **607**, 397–410.