Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular & Biochemical Parasitology 145 (2006) 158–170
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements

Frédéric Bringaud [a,*], Elodie Ghedin [b,c], Gaëlle Blandin [b], Daniella C. Bartholomeu [b], Elisabet Caler [b], Mariano J. Levin [d], Théo Baltz [a], Najib M. El-Sayed [b,c]

[a] *Laboratoire de Génomique Fonctionnelle Des Trypanosomatides, Université Victor Segalen Bordeaux 2, UMR-5162 CNRS, 146 Rue Léo Saignat, 33076 Bordeaux Cedex, France*
[b] *The Institute for Genomic Research (TIGR), 9712 Medical Center Drive, Rockville, MD 20850, USA*
[c] *Department of Microbiology and Tropical Medicine, George Washington University, Washington, DC 20052, USA*
[d] *Laboratorio de Biologia Molecular de la Enfermedad de Chagas, INGEBI-CONICET, DBMCF-School of Sciences, Centro de Genomica Aplicada (CeGA), University of Buenos Aires, Vuelta de Obligado 2490, 1428 Buenos Aires, Argentina*

Received 25 August 2005; received in revised form 25 September 2005; accepted 27 September 2005
Available online 10 October 2005

Abstract

The *ingi* and L1Tc non-LTR retrotransposons – which constitute the *ingi* clade – are abundant in the genome of the trypanosomatid species *Trypanosoma brucei* and *Trypanosoma cruzi*, respectively. The corresponding retroelements, however, are not present in the genome of a closely related trypanosomatid, *Leishmania major*. To study the evolution of non-LTR retrotransposons in trypanosomatids, we have analyzed all *ingi*/L1Tc elements and highly degenerate *ingi*/L1Tc-related sequences identified in the recently completed *T. brucei*, *T. cruzi* and *L. major* genomes. The coding sequences of 242 degenerate *ingi*/L1Tc-related elements (DIREs) in all three genomes were reconstituted by removing the numerous frame shifts. Three independent phylogenetic analyses conducted on the conserved domains encoded by these elements show that all DIREs, including the 52 *L. major* DIREs, form a monophyletic group belonging to the *ingi* clade. This indicates that the trypanosomatid ancestor contained active mobile elements that have been retained in the *Trypanosoma* species, but were lost from *L. major* genome, where only remnants (DIRE) are detectable. All 242 DIREs analyzed group together according to their species origin with the exception of 11 *T. cruzi* DIREs which are close to the *T. brucei ingi*/DIRE families. Considering the absence of known horizontal transfer between the African *T. brucei* and the South-American *T. cruzi*, this suggests that this group of elements evolved at a lower rate when compared to the other trypanosomatid elements. Interestingly, the only nucleotide sequence conserved between *ingi* and L1Tc (the first 79 residues) is also present at the 5′-extremity of all the full length DIREs and suggests a possible role for this conserved motif, as well as for DIREs.

Keywords: *Ingi*; L1Tc; Non-LTR retrotransposon; Retroposon; Degenerate retroelement; Evolution; *Trypanosoma brucei*; *Trypanosoma cruzi*; *Leishmania major*

1. Introduction

Transposable elements are widespread mobile genetic elements found in the genome of most organisms. They can be grouped into two main categories based on sequence organization and mode of transposition [1]. The first group consists of the cut-and-paste elements (DNA transposons), which move strictly through a DNA intermediate in both prokaryotic and eukaryotic genomes. The second group (retrotransposons) is transposed through an RNA intermediate and can be further divided into two lineages that utilize completely different mechanisms of integration. Those elements with long terminal repeats (LTR), called LTR retrotransposons, are similar both in structure and retrotransposition mechanism to retroviruses [2] and those elements that lack LTR, called non-LTR retrotransposons or retroposons, use a simpler mechanism of transposition. The current model for transposition of non-LTR retrotransposons was developed based on the analysis of the insect R2 element [3]. This model predicts that an element-encoded endonuclease (EN) performs a single-strand nick of the target DNA, generating an exposed 3′-hydroxyl that serves as a primer for reverse transcription of the element's RNA. The complementary strand of the new

* Corresponding author. Tel.: +33 557574632; fax: +33 557574803.
E-mail address: bringaud@u-bordeaux2.fr (F. Bringaud).

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.09.017

DNA copy of the element is, thus, directly synthesized onto the chromosome by the element-encoded reverse transcriptase (RT). The second single-strand nick is carried out on the other strand, a few base pairs downstream of the first nick, by the same element-encoded endonuclease, generating a primer for the second-strand synthesis of the retroelement. Consequently, the non-LTR retroelements are flanked by a direct repeat corresponding to the sequence between the two single-strand nicks performed by the element-encoded endonuclease, called target site duplication. They also have a variable length poly(A) or A-rich 3′-tail, due to the involvement of an RNA intermediate.

Since DNA transposons- and retrotransposons-like elements are present in prokaryotes, all mobile elements in eukaryotes are assumed to have descended from bacterial elements [4]. According to this model, ancestor(s) of eukaryotes contained both DNA transposons and retrotransposons suggesting that most, if not all, eukaryotes may contain mobile elements. Indeed, all the higher eukaryotes analyzed so far contain at least one family of mobile elements [5]. In contrast, 5 of the 15 unicellular eukaryotic genomes sequenced to date (http://genomesonline.org/), lack mobile elements, i.e. a Microsporidia intracellular parasite *Encephalitozoon cuniculi* [6] and 4 members of the Apicomplexa protozoan pathogens, *Plasmodium falciparum* [7], *P. yoelii yoelii* [8], *Cryptosporidium hominis* [9] and *C. parvum* [10]. This suggests that a significant fraction of unicellular eukaryotes may have lost active mobile elements. However, since none of these five genomes contain detectable vestiges of a mobile element, one cannot rule out the hypothesis that these genomes never contained mobile elements. To address this question, we have analyzed all the potentially active and highly degenerate non-LTR retrotransposons contained in the recently completed genome of three trypanosomatid protozoan parasites (*Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*) [11–13].

Trypanosomatids are protozoan parasites of major medical and veterinary significance. They cause serious disease in humans, such as sleeping sickness (*T. brucei*), Chagas disease (*T. cruzi*) and Leishmaniasis (*Leishmania* spp.). *T. brucei* and *T. cruzi* belong to the *Trypanosoma* genus and constitute a monophyletic group distantly related from the *Leishmania* spp. [14–16]. *L. major* is considered devoid of any mobile element, while both trypanosome species contain retrotransposons [17,18]. The genomes of *T. brucei* and *T. cruzi* contain similar retrotransposons, while no DNA transposons have been detected so far. VIPER is an LTR retrotransposon originally characterized in the *T. cruzi* genome [19] and recently identified in the *T. brucei* genome [11,12]. According to the current nomenclature, all trypanosomatid non-LTR retrotransposons analyzed so far are divided into the CRE and *ingi* clades [4]. The CRE clade is composed of the *T. brucei* SLACS, *T. cruzi* CZAR and *Crithidia fasciculata* CRE1/CRE2 elements, which are site-specific retroelements always inserted at the same relative position in the spliced leader (*SL*) *RNA* genes [20–23]. The *T. brucei ingi* and *T. cruzi* L1Tc elements, of the *ingi* clade, are dispersed in the host genome [24–26], although they show a relative site-specificity for insertion [27] (Bringaud, unpublished data). It is noteworthy that mobilization of trypanosomatid retroelements has not been observed so far, therefore, we consider as potentially functional

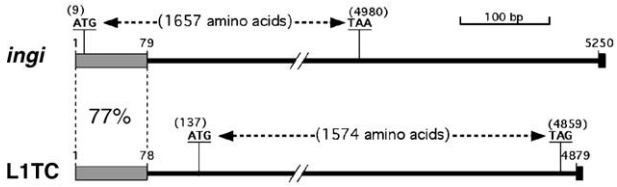

Fig. 1. Schematic representation and comparison of *ingi* and L1Tc. Schematic map of *ingi* (*T. brucei*) is based on the fifth (26P8i5) retroelement present in the fully sequenced *T. brucei* BAC RPCI93-26P8 (ACC: AC087701) [36,58] and the L1Tc map (*T. cruzi*) is derived from the retroelement present in BAC62 (ACC: AF208537) [36,59]. The potentially functional *ingi* retroelement contains a single long ORF (4971 bp), from position 9 (ATG codon) to position 4980 (TAA codon), which encodes a 1657 amino acid protein. The potentially functional L1Tc retroelement contains a single long ORF (4722 bp), from position 137 (ATG codon) to position 4859 (TAG codon), which encodes a 1574 amino acid protein. Black boxes at the end of both maps represent the poly(dA) terminal sequence. The first 79 bp of *ingi* are 77% identical to the corresponding region (first 78 bp) of L1Tc (grey boxes), which constitutes the only conserved nucleotide sequence between these two members of the *ingi* clade.

(or potentially active) *ingi* and L1Tc elements encoding a large single protein (1657 and 1574 amino acids, respectively) (Fig. 1) comprised of the central reverse transcriptase [28] and RNAse H (RH) [29] domains, C-terminal DNA-binding domains [30] and a N-terminal apurinic/apyrimidinic-like endonuclease domain [31]. We have previously identified a subset of highly degenerate group of non-LTR retroelements related to the *ingi* clade and named them DIREs for "degenerate *ingi*/L1Tc-related elements" [32]. In this paper, we report the identification and characterization of the full complement of DIREs in the *T. brucei*, *T. cruzi* and *L. major* genomes. Our analysis shows that *L. major* has eliminated all the active non-LTR retrotransposons present in its trypanosomatid ancestor, while trypanosome genomes still contain potentially active elements.

2. Materials and methods

2.1. Detection and reconstitution of the chimeric DIRE coding sequences

T. brucei, *T. cruzi* and *L. major* genome sequences are available at GeneDB (http://www.genedb.org/). *Ingi* and L1Tc peptide sequences were used to detect all DIREs in the *T. brucei* (Tb927.v3.0), *T. cruzi* (TcBr.v3.0) and *L. major* (LmjF.v4.0) genomes. An initial TBLASTN search was performed against all *T. brucei* and *L. major* chromosomes and *T. cruzi* contigs using the *ingi* and L1Tc peptides. Approximate coordinates of the DIREs were determined and putative gene models were created and translated. These peptides were then searched again against *ingi* and L1Tc peptides using the BLAST–extend–repraze (BER) algorithm developed at TIGR. This algorithm extends the boundaries of each ORF by 300 bp on both ends and a modified Smith–Waterman alignment is then performed between the proteins, including the translation of the extensions. The extensions allow the examination of all translation frames and past stop codons. Because of the degenerate nature of DIREs, this process allows to determine the precise coordinates of these ele-

ments. To tentatively reconstitute chimeric proteins from the analyzed DIREs, frame shifts were removed manually from the DNA sequences using the BER outputs to precisely determine the frame shift positions. This approach was used to generate a pseudogene for each DIRE encoding a single *ingi*/L1Tc-like sequence that contains numerous stop codons in most cases.

2.2. Other databases mining

The absence of detectable traces of retrotransposons, in the *Encephalitozoon*, *Cryptosporidium* and *Plasmodium* genomes was confirmed by performing TBLASTN searches with the reverse transcriptase domain of different LTR and non-LTR retrotransposons. The TBLASTN searches were performed on the *E. cuniculi* (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=6035), *Cryptosporidium parvum* and *hominis* (http://CryptoDB.org/) and *P. falciparum* (http://www.genedb.org/genedb/malaria/) genome web sites.

2.3. Phylogenetic analyses

The reverse transcriptase, apurinic/apyrimidinic-like endonuclease and RNase H amino acid domains were aligned using the multiple alignment option in CLUSTAL X [33], followed by minor manual adjustments using MacClade Version 4.06 (Sinauer Associates Inc.). The alignments of the RT, EN and RH domains have been deposited at EMBL with Accession numbers ALIGN_000836, ALIGN_000837 and ALIGN_000838, respectively. Phylogenetic trees were generated by the neighbor-joining and maximum parsimony heuristic methods as implemented in PAUP Version 4.0b10 (Sinauer Associates Inc.), using default parameters. Bootstrapping was also carried out using PAUP.

3. Results

3.1. Identification of degenerate Ingi/L1Tc-like sequences

In the course of the genome project analysis, we annotated *ingi* and L1Tc elements based on respective nucleotide sequence homology to the *T. brucei ingi* and *T. cruzi* L1Tc non-LTR retrotransposons. A total of 85 *ingi* and 296 L1Tc were identified, which corresponds to 115 and 320 expected retroelements per haploid genome, respectively [12]. The intra-species percentage of identity between the nucleotide sequence of these elements ranges between 49.9 and 99.8% for *T. brucei* (*ingi*) and 51.8 and 99.8% for *T. cruzi* (L1Tc) with a mean of 92.2% for *ingi* and 94% for L1Tc. TBLASTN searches of the three trypanosomatid genomes with the retroelements product revealed, in addition to the previously annotated elements, sequences presenting significant homology with the RT, RH and/or EN domains. These *ingi*/L1Tc-like sequences, which contain numerous frame shifts and stop codons, correspond to DIREs [32]. To analyze and compare the DIRE gene products, we used a BLAST-based tool (BER, see Section 2) to locate the frame shifts in the degenerate sequences. This allowed us to tentatively reconstitute chimeric proteins with matches to the *ingi* and L1Tc products for the purpose of phylogenetic analyses. Among the 53 and 52 DIREs identified in the *T. brucei* (TbDIRE) and *L. major* (LmDIRE) genomes, 47 and 31 were successfully reconstituted (Fig. 2). Intra-species comparison of the reconstituted DIRE proteins revealed the existence of nearly identical elements that were ordered into representative groups of related elements (22 for *T. brucei* and 21 for *L. major*). The elements in each group are depicted in Fig. 2. Due to the higher number of DIREs in the *T. cruzi* database (238 elements in the 1701 contigs >10 kb, which represent approximately 1.2× coverage of the haploid genome), 192 TcDIREs were first ordered by comparing their nucleotide sequence. We defined 28 groups of related sequences, including a large family composed of 104 elements. Chimeric protein sequences were then successfully reconstituted for a representative element of 22 groups. As expected, all the 65 reconstituted proteins from *T. brucei* (22), *T. cruzi* (22) and *L. major* (21), which represent a total 297 different DIREs, match with the trypanosome non-LTR retrotransposons. The mean percents of identity with the *ingi* product are 31.6 ± 6.2% for TbDIRE, 29.6 ± 6.6% for TcDIRE and 27.9 ± 4.3% for LmDIRE (Table 1). The values are lower when TbDIRE and LmDIRE are compared with the L1Tc product (15.7 ± 3.7 and 25.5 ± 4.1%, respectively). Sur-

Table 1
Percentage of identity between the proteins encoded by *ingi*, L1Tc and DIRE

Group	Subclade	Number[a]	% of identity with *ingi*[b]	% of identity with L1Tc[b]
ingi	*ingi*	1 (85)	–	23.8
L1Tc	L1Tc	1 (296)	23.8	–
TbDIRE1	*ingi*	3 (13)	44.7 (33.8–62.0)	18.3 (14.1–21.0)
TbDIRE2	*ingi*	6 (7)	29.5 (27.1–31.6)	14.1 (11.3–19.3)
TbDIRE3	*ingi*	9 (11)	27.5 (22.7–33.0)	14.4 (10.2–19.4)
TcDIRE1	*Ingi*	6 (11)	38.7 (33.8–42.4)	26.3 (18.9–32.6)
TcDIRE2	L1Tc	15 (169)	26.4 (21.0–33.3)	25.7 (18.9–33.8)
LmDIRE	LmDIRE	21 (31)	27.9 (18.2–33.7)	25.5 (14.8–31.8)

[a] Number of sequences compared with the *ingi* and/or L1Tc product. Number of sequences belonging to this family and annotated in the corresponding genome is indicated into brackets.
[b] Range of values is indicated into brackets.

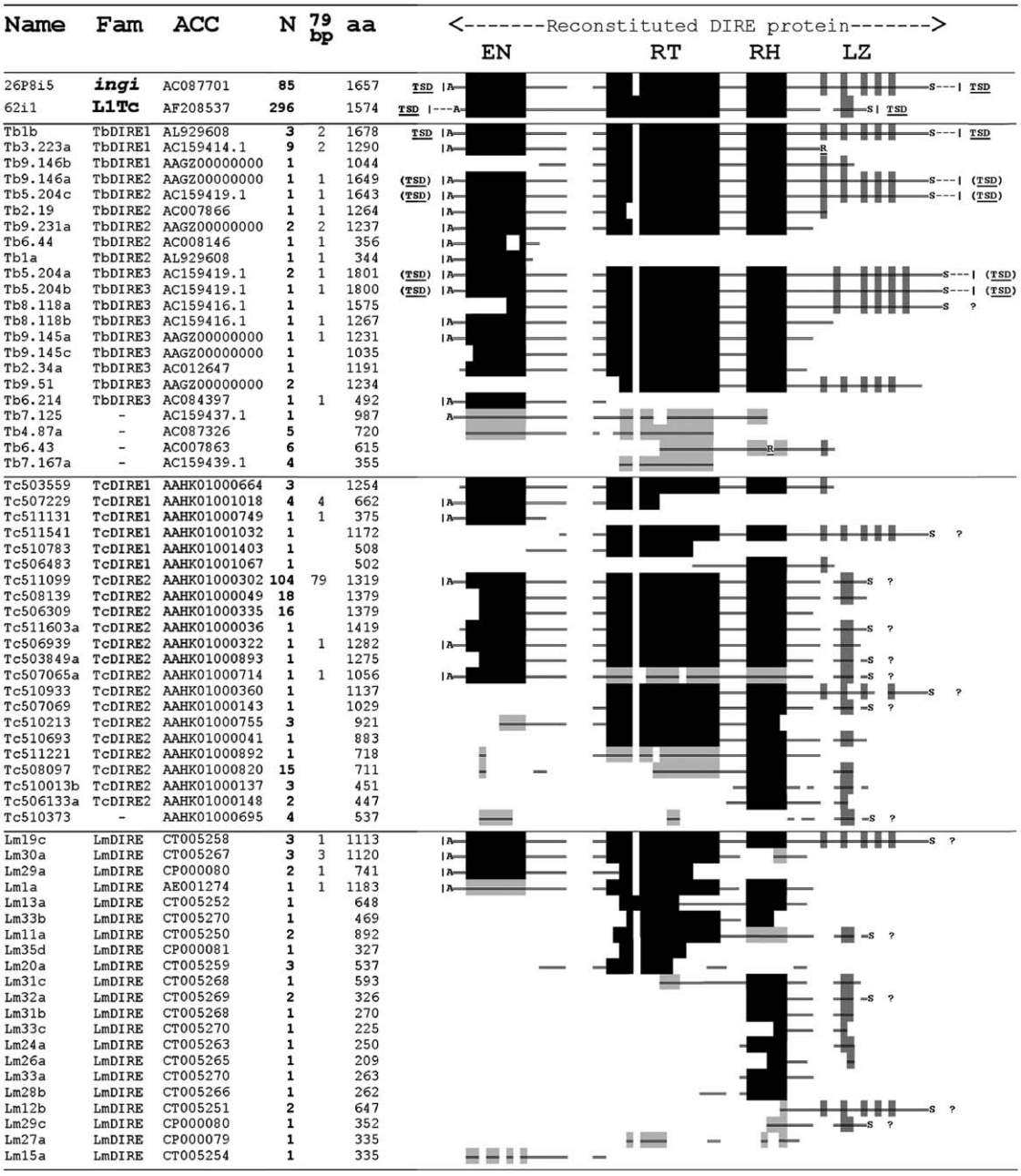

Fig. 2. Schematic map of the DIRE. A map of *ingi*, L1Tc and the selected representative of each group of nearly identical DIREs, is presented in the right panel. The five TbDIREs, eight TcDIREs and 21 LmDIREs for which the protein could not be successfully reconstituted, are not shown. In the left panel, are indicated the names of the selected DIREs ("name") along with their corresponding DIRE family ("Fam"), the accession number of the BAC, contig or chromosome containing the representative selected DIREs ("ACC"), number of elements in each group ("N"), number of elements containing the 79 bp signature in each group of DIREs ("79 bp") and size of the reconstituted protein in amino acids ("aa"). In the right panel, the EN, RT and RH domains are represented by black boxes or light grey boxes, corresponding to whether the domain was used or not for the phylogenetic analysis, respectively (Figs. 3–5). The leucine zipper motifs are indicated by dark grey boxes. Gaps introduced in the coding sequences (double lanes) reflect the alignment of the retroelement product. The initiation (ATG) and stop codons at the beginning and the end of the coding sequences are represented by "A" and "S", respectively. Dashes represent non-coding retroelement sequences and vertical bars flanking the maps indicate that the extremity of the element is identified. When present at the 5′-extremity, the bar indicates that the element contains the 79 bp trypanosomatid non-LTR retroelement signature. An interrogation mark indicates that the 3′-extremity is not known, while the stop codon was identified. Target site duplication (TSD) means that the retroelement is flanked by an identified duplicated sequence of TSD-like sequence, suggestive of a relative recent retrotransposition event. Two groups of TbDIREs, represented by the Tb3.223a and Tb6.43 elements, are interrupted by a RIME (non-autonomous non-LTR retrotransposon) and represented by an underlined "R". All the *T. brucei*, *T. cruzi* and *L. major* sequences are available on GeneDB (http://www.genedb.org).

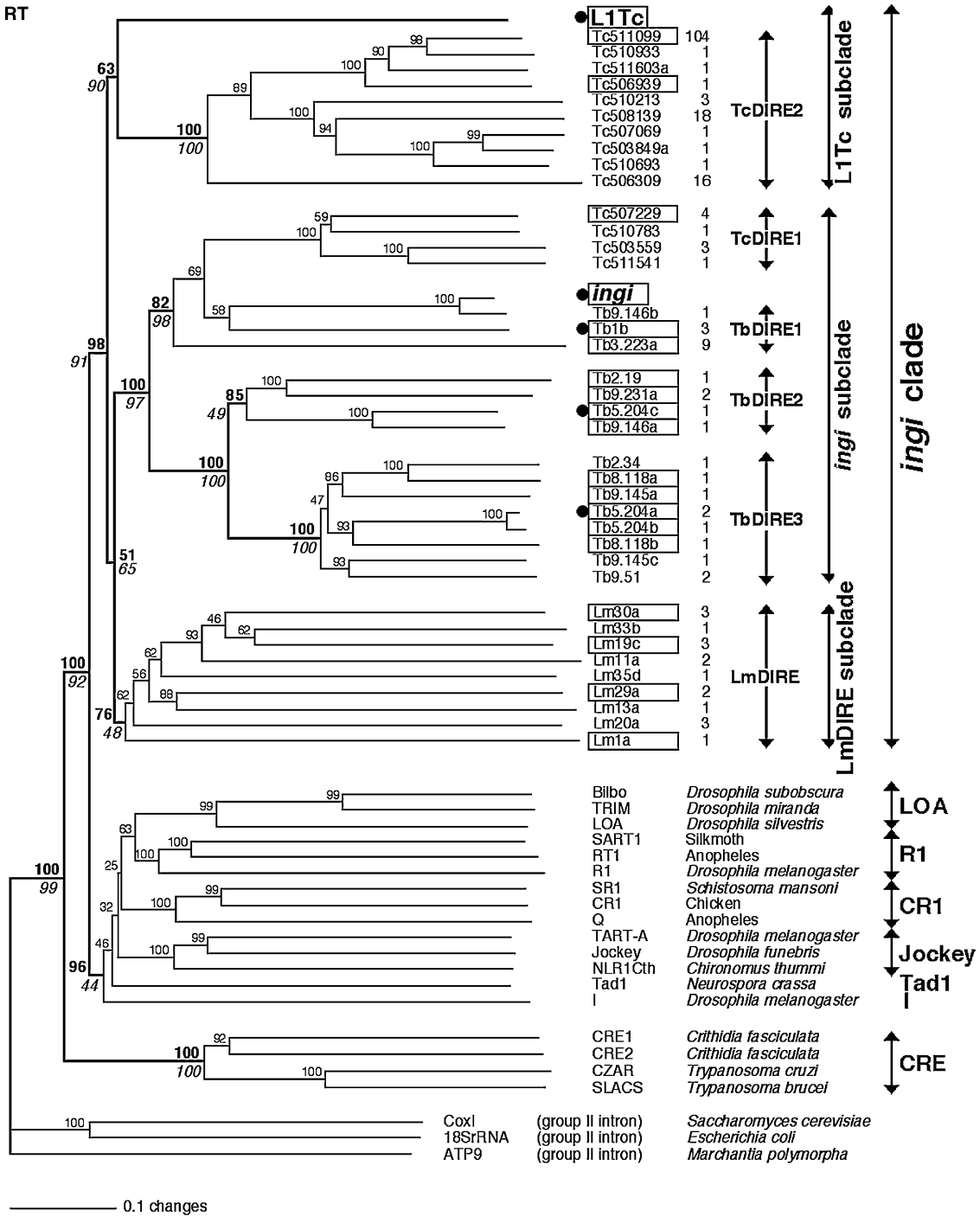

Fig. 3. Phylogenetic analysis of the RT domain. The phylogeny is based on approximately 450 aligned amino acid residues (alignment is available under the Accession number: ALIGN_000836), corresponding to the entire RT domain of non-trypanosomatid non-LTR elements, *ingi*, L1Tc and DIREs. Very few highly degenerate DIRE RT domains were removed from this analysis (grey boxes in Fig. 2). The 14 non-trypanosomatid non-LTR elements are representatives of the Jockey and I groups which are the closest relative of *ingi*/L1Tc among the five previously defined non-LTR retrotransposon groups [4]. The I group is comprised of the Tad1, R1, LOA, I and *ingi* clades. This consensus tree was generated with the neighbor-joining method and rooted on the RT sequences of group II introns. All numbers next to each node, except those in italic, indicate bootstrap values as percentage out of 100 replicates corresponding to the tree generated with the neighbor-joining method. A similar pattern was also obtained by the maximum parsimony method for the lower part of the tree, which define the DIRE families (italics numbers below nodes indicate bootstrap values as percentage out of 100 replicates). Names of each DIRE and number of elements constituting their group are shown in the right margin. The boxed DIREs contain the 79 bp trypanosomatid non-LTR retroelement signature and those with a dot are flanked by a TSD or a TSD-like sequence. For non-trypanosomatid non-LTR elements and group II introns, name and species of origin are given to the right. Arrows in the right margin indicate DIRE families (TbDIRE1/2/3, TcDIRE1/2 and LmDIRE), subclades (*ingi*, L1Tc and LmDIRE) and clades (*ingi*, LOA, R1, CR1, Jockey, Tad1, I and CRE).

prisingly, TcDIREs are apparently closer to *ingi* than to L1Tc (29.6 ± 6.6% versus 25.8 ± 4.5%) (Table 1).

3.2. Phylogenetic analysis of the trypanosomatid non-LTR retrotransposons

Using the reconstituted DIRE products, we attempted to reconstruct the phylogenetic relationships between non-LTR elements from different eukaryotes, including *ingi* (*T. brucei*), L1Tc (*T. cruzi*) and DIRE. Phylogenetic analyses of the non-LTR retrotransposons are commonly performed on the RT domain, which is the trademark of retrotransposons. RT phylogeny is statistically more robust than phylogenetic trees generated with the EN and RH domains of retroelements. Indeed, among the non-LTR retrotransposon domains, RT is the most conserved. It is also the only domain present in all elements and it is the longest one (∼450 aa as compared to ∼230 and ∼200 aa for the EN and RH domains, respectively). This increases considerably the number of evolutionary significant positions [34]. In addition, the non-LTR retrotransposon RT domain is related to the RT domain of group II introns, which can be used as a closely related outgroup to root the tree [35]. A neighbor-joining phylogram for the RT domain of 3 group II introns (outgroup), 4 trypanosomatid site-specific retroelements (CRE clade), 14 non-trypanosomatid non-LTR elements, *ingi*, L1Tc and 38 DIREs is shown in Fig. 3. We find that *ingi*, L1Tc and all the DIREs appear to form a monophyletic clade distinct from all the other non-LTR retroelements. This is supported by a high bootstrap value (98 and 91% for the neighbor-joining and maximum parsimony methods, respectively). As previously described, I factor element is the closest relative of the *ingi* clade [34]. The *ingi* clade can be subdivided into three subclades: the LmDIRE subclade contains all the LmDIRE, the L1Tc subclade is composed of L1Tc and most of the TcDIREs and the *ingi* subclade contains *ingi*, all the TbDIREs and a few TcDIREs (Fig. 3). Similar phylogenetic analyses were performed on the EN and RH domains, using cellular domains as outgroups (Figs. 4 and 5), displaying the same pattern as for RT. This includes the monophyly of the

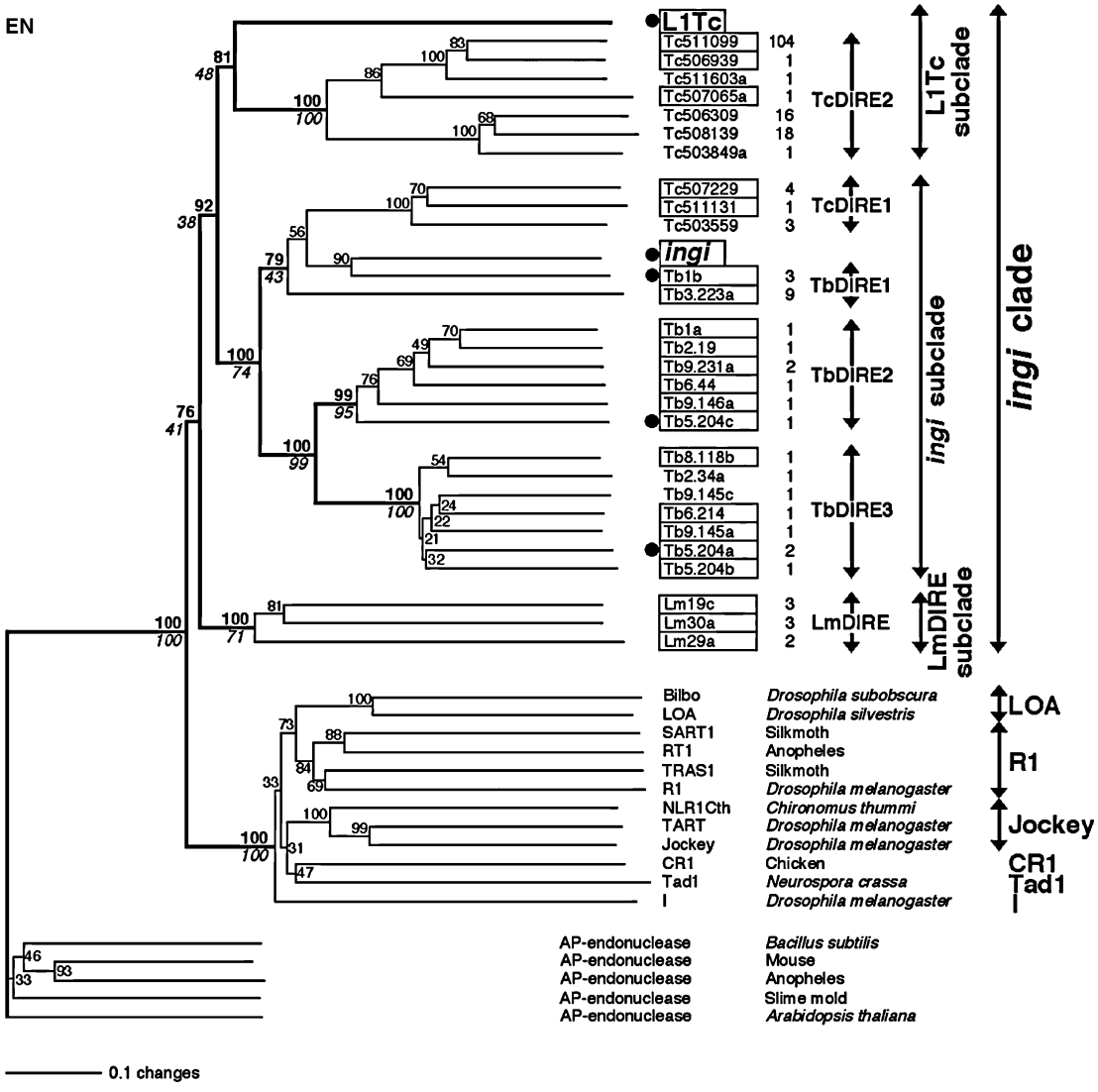

Fig. 4. Phylogenetic analysis of the EN domain. The phylogeny is based on approximately 230 aligned amino acid residues (alignment is available under the Accession number: ALIGN_000837), corresponding to the entire EN domain. This consensus tree was generated with the neighbor-joining method and rooted on the prokaryotic and eukaryotic housekeeping apurinic/apyrimidinic endonucleases (called "AP-endonuclease") (for more details, see Fig. 3).

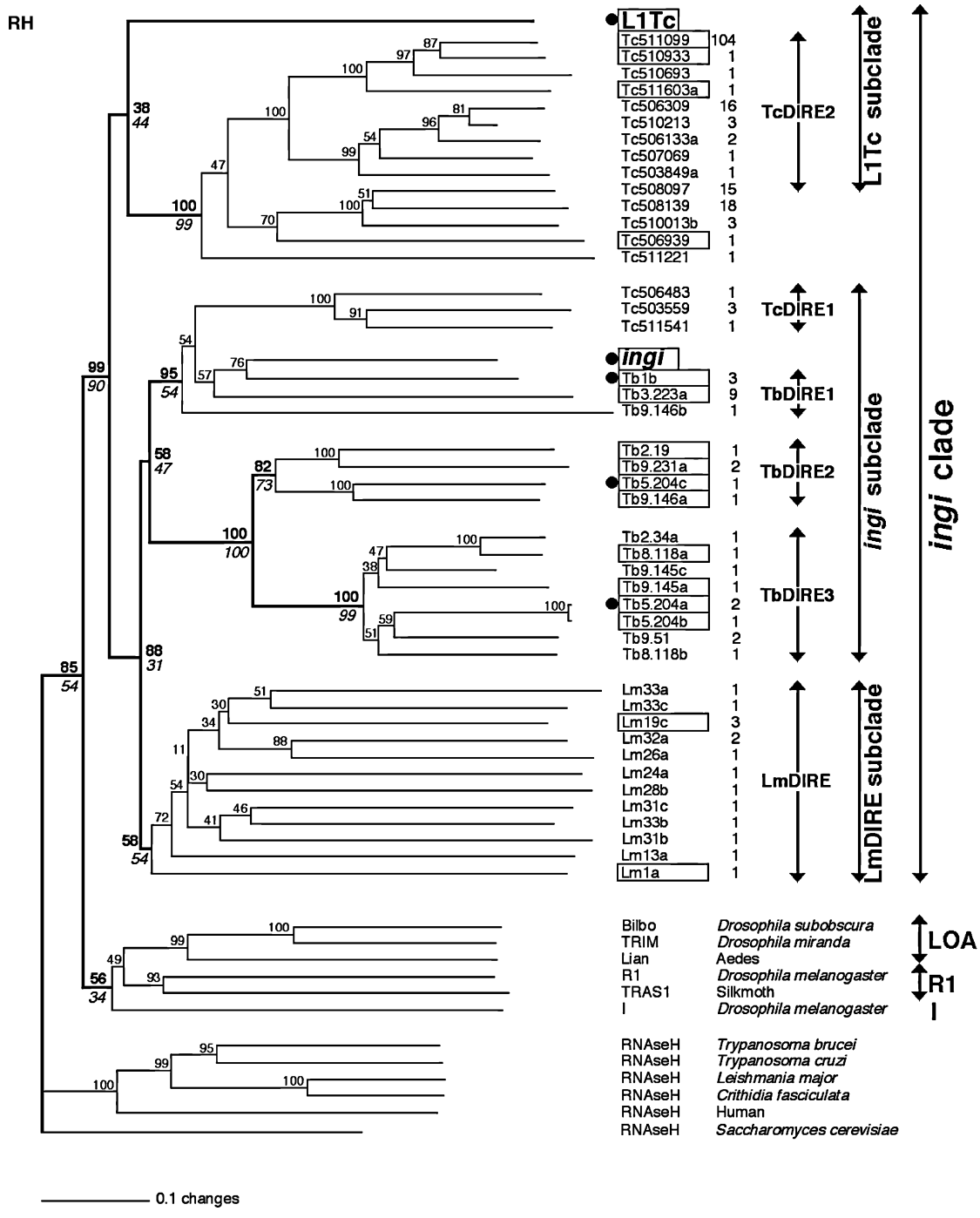

Fig. 5. Phylogenetic analysis of the RH domain. The phylogeny presented is based on approximately 200 aligned amino acid residues (alignment is accessible under the Accession number: ALIGN_000838), corresponding to the entire RH domain. This consensus tree was generated with the neighbor-joining method and rooted on the housekeeping eukaryotic RH (called "RNAse H"), including those for four trypanosomatid species (for more details, see Fig. 3).

ingi clade (*ingi*, L1Tc and DIRE) with the I factor element as the closest relative, the definition of three subclades corresponding to the three trypanosomatid species and the presence of a few TcDIREs in the *ingi* subclade.

On the basis of these three phylogenies, the *T. brucei* and *T. cruzi* DIREs form two and three families, respectively (Figs. 3–5). The 3 TbDIRE families belong to the *ingi* subclade, while for *T. cruzi*, the TcDIRE2 family (15 groups corresponding to 169 elements) belongs to the L1Tc subclade and the TcDIRE1 family (6 groups corresponding to 11 sequences) belongs to the *ingi* subclade (Fig. 2).

3.3. Most of the DIREs contain the trypanosomatid non-LTR retrotransposon signature

The protein-based phylogenetic analysis strongly supports the view that DIRE, *ingi* and L1Tc belong to the same clade of retroelements and have a common ancestor in the trypanoso-

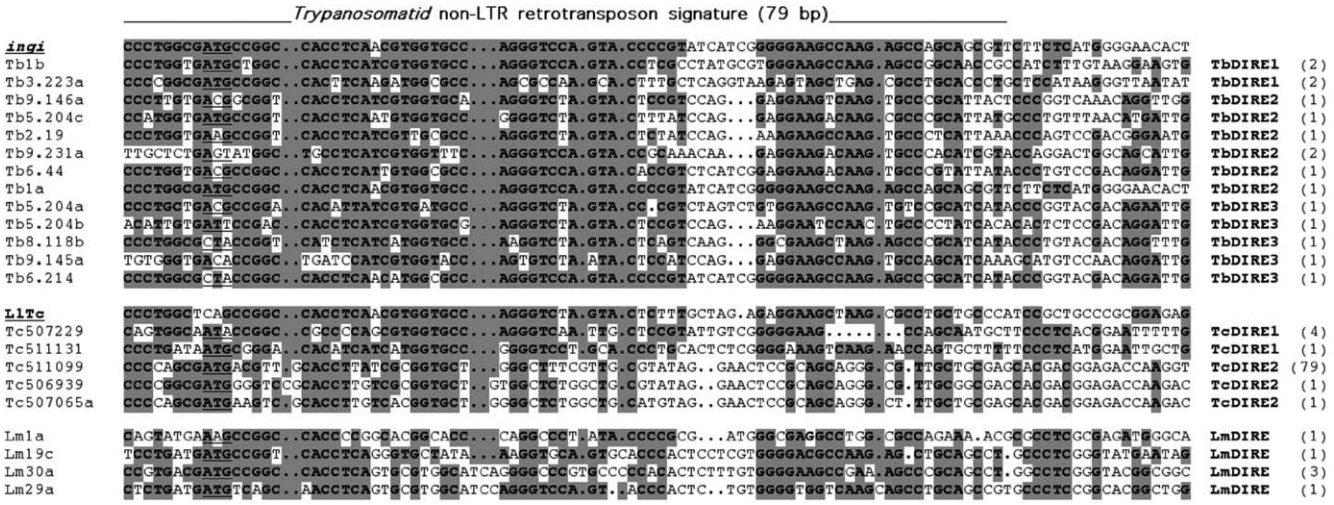

Fig. 6. Comparison of the DIRE 5′-extremity. The first 100 bp of *ingi* (26P8i5), L1Tc (Tc507835) and DIRE containing the trypanosomatid non-LTR retrotransposon signature were aligned with the introduction of gaps (.) to maximize the alignments. Identical residues are shaded in grey and the ATG or degenerate ATG initiation codons are underlined. The position of the trypanosomatid non-LTR retrotransposon signature (79 bp signature) is indicated above the alignment. Name and family of the aligned DIRE are indicated in the left and right margins, respectively. Number of identical or nearly identical elements containing the 79 bp signature is indicated into brackets in the right margin and only the representative of each group is shown in the alignment (see Fig. 2).

matid lineage. *Ingi* and L1Tc, which are 23.8% identical at the protein level, are not conserved at the nucleotide level; however, they share a conserved stretch of 79 bp [36]. This 79 bp sequence is also conserved in the short non-autonomous non-LTR retrotransposons identified in *T. brucei* (RIME) [24,25,37] and *T. cruzi* (NARTc) [36]. This suggests that this particular sequence, called the trypanosomatid non-LTR retrotransposon signature (or 79 bp signature), has a critical role for trypanosomatid retroelement function. Interestingly, the 79 bp signature is present at the 5′-extremity of one-third of the 297 DIREs analyzed. This represents 108 elements, i.e. 86 TcDIREs, 16 TbDIREs and 6 LmDIREs (Figs. 2 and 6), which belong to the 6 different DIRE families (boxed DIRE names in Figs. 3–5). The 5′-extremity of the other DIREs is truncated and consequently does not contain the 79 bp signature.

3.4. Few DIREs are flanked by a putative target site duplication

An essential question is to determine whether the DIREs are still active. Due to their significant degree of degeneration, we can exclude the possibility that DIREs code for their own retrotransposition. However, they could be mobilized using the retrotransposition machinery encoded by active elements (such as potentially active *ingi* in *T. brucei* and L1Tc in *T. cruzi*) as previously observed in other eukaryotes [38]. Recent mobilization of DIREs would imply the presence of a target site duplication flanking the element. However, the identification of TSD requires the precise definition of the retroelement extremities, which is straightforward for members of conserved retroelement families, such as *ingi* and L1Tc, but not obvious for highly degenerate elements.

Fortunately, we were able to precisely define the 5′-extremity for 16 TbDIREs containing the 79 bp signature. They include six full length elements which contain the entire coding sequence followed by a putative poly(A) sequence considered as the 3′-extremity of the elements (Fig. 2). The size of these full length TbDIREs is similar to *ingi* (5.25 kb), including Tb1b (5218 bp) and Tb9.231b (5258 bp), which are 88.9% identical at the nucleotide level. It is noteworthy that while Tb1b and Tb9.231b are similar, they are flanked by very different sequences, which suggests that these related DIREs are inserted

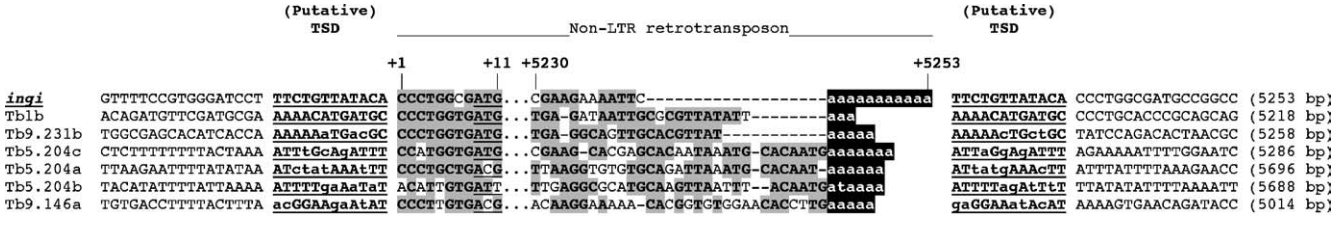

Fig. 7. Comparison of the 5′- and 3′-adjacent sequences flanking DIREs. The alignment of the selected *ingi* (26P8i5) and DIREs and their flanking regions is based on the retroelement sequence (headed "non-LTR retrotransposon") from which only the first 11 bp and the last 23–37 bp are shown. In the retroelement sequence, grey-shaded residues are conserved in the retroelement sequence, the poly(A) sequence is shaded in black and the ATG or degenerate ATG initiation codons are underlined. Names and lengths of each elements are indicated in the left and right margins, respectively. Tb9.231b belongs to the Tb1b group, which contains three differents DIREs. Numbers above the alignment show the nucleotide positions of the selected *ingi* element. Conserved residues in the target site duplication (TSD) or putative TSD are indicated by bold-faced and capital characters.

in different sequences. Furthermore, this comparison is helpful to determine precisely both extremities of Tb1b and Tb9.231b. Interestingly, Tb1b is flanked by a 12 bp identical sequence which fulfills the TSD definition (Fig. 7). The size of the Tb1b TSD is also consistent with the 12 bp long *ingi* TSD [27]. Furthermore, Tb1b (previously called TUBIS [39]) is inserted at the 3′-extremity of a β-tubulin pseudogene (901 bp), which did not accumulate point mutations, since its sequence is identical to the corresponding sequence of the other β-tubulin genes (1329 bp) found in the α-/β-tubulin cluster [40]. Altogether, this suggests that Tb1b has been recently retrotransposed into the α-/β-tubulin cluster. This hypothesis is strengthened by the absence of TUBIS-like sequence (Western blot analysis) in the *T. congolense* genome [41] and database mining of the ongoing *T. congolense* and *T. vivax* genome projects at the Wellcome Trust Sanger Institute (http://www.genedb.org/) did not reveal the presence of DIRE in the vicinity of the tubulin clusters. For the five other full length DIREs identified in the *T. brucei* genome, the situation is different. The 12 bp sequences flanking the elements are not identical, but 7 (Tb5.204b and Tb9.146a) to 9 residues (Tb9.231b and Tb5.204c) out of the 12 bp are conserved (Fig. 7), suggesting that these remnant TSDs, as well as the DIREs, have accumulated point mutations over time. This observation is suggestive of older retrotransposition events as compared to the Tb1b and *ingi* elements.

Most of the 86 *T. cruzi* DIREs containing the 79 bp signature belong to the Tc511099 group (79 out of the 104 elements composing this group). Interestingly, the analysis of the 5′-flanking sequence of these 79 DIREs revealed only 5 classes of nearly identical sequences, suggesting that these elements have relative site-specificity for insertion or they have been duplicated together with their flanking regions (data not shown). Among the 86 TcDIREs containing the 79 bp signature, 33 are full length, which corresponds to five groups of DIREs (Figs. 2 and 6). None of these full length elements are flanked by a TSD. However, three elements belonging to the Tc511099 group are inserted in the middle of a VIPER LTR retrotransposon that is present in 275 nearly identical copies in the haploid genome (data not shown), suggesting recent retrotransposition events.

For *L. major*, we were not able to determine the 3′-extremity of the six LmDIREs containing the 79 bp signature (Fig. 6), which enabled us to identify putative TSD for these elements.

4. Discussion

Three classes of potentially active mobile elements have been characterized so far in the nuclear genome of trypanosomatids: the LTR retrotransposons (VIPER), the site-specific non-LTR retrotransposons (SLACS, CZAR, CRE) and the non-site-specific non-LTR retrotransposons (*ingi* and L1Tc) [17]. TBLASTN analyses with the protein(s) encoded by these non-LTR retrotransposons revealed 456 highly degenerate sequences in the genomes of the three trypanosomatids sequenced so far (*T. brucei*, *T. cruzi* and *L. major*). All of these elements are related to the *ingi* and L1Tc products. Phylogenetic analyses of the reconstituted chimeric proteins encoded by 242 of these DIREs showed that: (i) all DIREs form a monophyletic group belonging to the *ingi* clade, composed of the *ingi* and L1Tc potentially active elements [4]; (ii) all 242 DIREs analyzed are grouped according to their species origin, except 11 *T. cruzi* DIREs (TcDIRE1 family), which are closely related to the *T. brucei ingi* and TbDIRE1 family (*ingi* subclade); (iii) the *L. major* haploid genome contains 52 DIREs, although this species was previously believed to lack mobile elements (Fig. 8A). This was widely accepted since none of the retrotransposons characterized so far in *T. brucei*, *T. cruzi* and *C. fasciculata* have been observed in the *Leishmania* spp. [17]. Indeed, in the course of this analysis, we did not detect any potentially active mobile elements in the completed *L. major* genome [12,13]. We have only detected highly degenerate sequences (LmDIREs) phylogenetically related to *ingi* and L1Tc. This clearly demonstrates that the *Leishmania* ancestor hosted active *ingi*/L1Tc-like retroelements, which have been lost over time to become vestigial retroelements in the genome of the present day *L. major*.

The fall and rise of retroelement subfamilies is well documented in multicellular eukaryotes [42,43]. However, the complete disappearance of all active mobile elements from eukaryotes has not yet been demonstrated. Indeed, the only previously sequenced eukaryotic genomes devoid of mobile elements (*E. cuniculi* [6], *Cryptosporidium* species [9,10] and *Plasmodium* species [7,8]) do not contain detectable traces of extinct elements. These unicellular eukaryotes probably lost mobile elements present in the genome of their ancestors, however, one cannot rule out the possibility that *Encephalitozoon*, *Cryptosporidium* and *Plasmodium* ancestors never contained mobile elements. The presence of solely degenerate dead non-LTR retroelements in *L. major* (LmDIREs), a close relative of *T. brucei* and *T. cruzi,* which contain related potentially active elements, clearly demonstrates that eukaryotic cells can eradicate active mobile elements from their genome. We can predict that in the absence of active elements, LmDIREs will continue to rapidly accumulate mutations over time. Ultimately, descendants of *Leishmania* will probably lose detectable traces of extinct retroelements, as observed for *Encephalitozoon*, *Cryptosporidium* and *Plasmodium* species.

Most of the 180 *T. cruzi* DIREs analyzed (TcDIRE2) form a monophyletic group with the *T. cruzi* potentially active non-LTR retrotransposons (L1Tc subclade) with the exception of 11 TcDIREs (TcDIRE1), which belong to the *ingi* subclade. The presence of these few TcDIREs in the *ingi* subclade could be explained by horizontal transfer between *T. brucei* and *T. cruzi*, as proposed in Fig. 8B. However, this hypothesis is unlikely, since *T. brucei brucei* subspecies (including the sequenced TREU927 strain) and *T. cruzi* are restricted to the African and South-American continents, respectively, due to the geographical localization of their insect vector. These two continents separated approximately 130 million years ago. In the absence of a fossil record, the time of divergence between Salivarian trypanosomes (group containing *T. brucei brucei* subspecies) and Stercorarian trypanosomes (group containing *T. cruzi*) is unknown, however, it preceded the separation of the continents [15,44]. In addition, no horizontal transfer has been described so far for non-LTR retrotransposons during the last 600 million

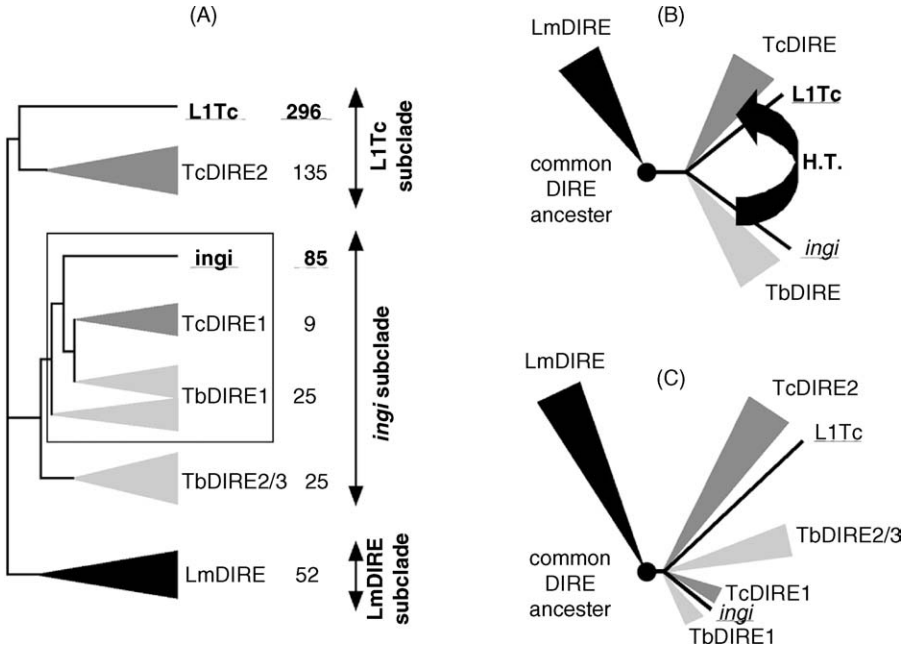

Fig. 8. Evolution of the retroelements composing the *ingi* clade. (Panel A) Compilation of the phylogenetic analyses performed in Figs. 3–5. The number of identified elements composing the seven families of the three subclades (L1Tc, *ingi* and LmDIRE) are indicated on the right. The *ingi* and L1Tc families, which contain potentially active retrotransposons are underlined. To tentatively explain the presence of *T. cruzi* DIREs (TcDIRE1) in the *ingi* subclade (as shown in panel A), two evolutionary scenarios are presented in panels B and C. Both hypotheses consider that all the elements composing the *ingi* clade derive from an ancestral *ingi*/L1Tc-like retroelement present in the common trypanosomatid ancestor. The first hypothesis is based on the unlikely horizontal transfer of element(s) (H.T.) from *T. brucei* to *T. cruzi* (panel B). The second hypothesis considers that the rate of mutation accumulation for some elements, boxed in panel A, is lower as compared to the other elements of the *ingi* clade (panel C).

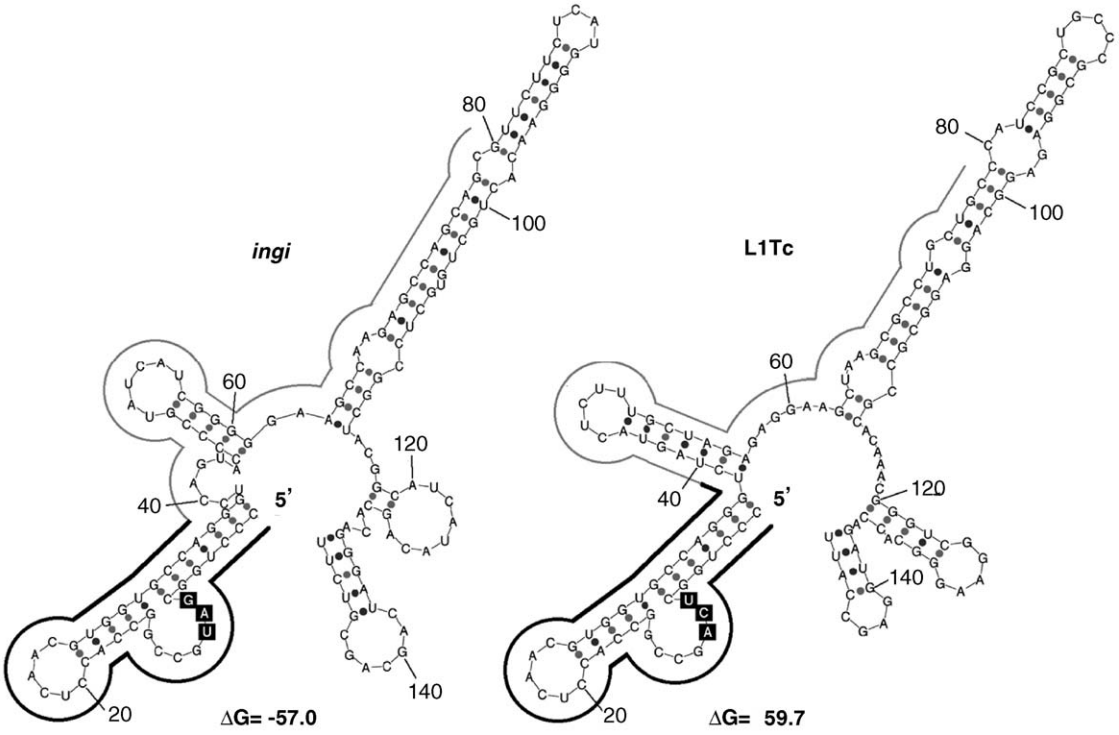

Fig. 9. Predicted secondary structure of the 5′-extremity of the *ingi* and L1Tc mRNA. This analysis was performed on the first 149 bp of *ingi* and L1Tc using the Mfold program, Version 3.1 (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1-2.3.cgi) [60]. The most stable structure, among the few predicted structures (five for *ingi* and three for L1Tc) are presented. The 5′-end conserved sequences (79 bp for *ingi* and 78 bp for L1Tc) are shown by a thick black lane for the first 39 bp (92% identity) and a thin grey lane the following 40 bp (57% identity). The three non-conserved residues in the first 39 bp conserved sequences are white characters on a black background.

years [34], as opposed to LTR retrotransposons [45]. The reason for this difference may reside in the retrotransposition mechanism of non-LTR retrotransposons (the target-primed reverse transcription reaction), in which the cDNA strand is reverse-transcribed from an RNA template directly onto a chromosomal target site [3], leading to the absence of free DNA element in the cytosol. However, the absence of evidence does not mean that non-LTR retrotransposons cannot be transferred horizontally. Considering that the horizontal transfer scenario is unlikely, we propose that certain elements constituting the *ingi* subclade, such as *ingi*, TbDIRE1 (13 elements) and TcDIRE1 (11 elements), evolved at a lower rate as compared to the other elements (Fig. 8C). This hypothesis suggests that selective pressure was imposed over time on these particular sequences forcing them to be conserved. In other words, these elements may have a particular unknown function useful to the cell, preventing them to accumulate point mutations.

For that matter, Obado et al. recently showed that the locus required for the mitotic stability of *T. cruzi* chromosome three centers on a 16 kb strand-switch region composed of degenerate retrotranspososons [46]. They proposed that this GC-rich region, which contains a TcDIRE1 of the Tc507229 group, acts as a centromere. Interestingly, the corresponding region of *T. brucei* chromosome 1, which is syntenic with *T. cruzi* chromosome 3, has been postulated to have centromeric properties on the basis of a low recombination frequency [40] and contains a TbDIRE1 belonging to the Tb3.223a group. The *T. brucei* genome contains nine TbDIRE1 of this Tb3.223a group (the larger TbDIRE group) located on five different chromosomes. The introduction in *T. brucei* of an episomal plasmid containing a degenerate non-LTR retrotransposon will be helpful to determine whether TbDIRE1 sequences play a role in mitotic stability.

The two presumed active families of retroelements of the *ingi* clade (*ingi* and L1Tc) are only 28.5% identical at the protein level. At the nucleotide level, the first 39 bp and the following 40 bp constitute the only motif conserved between *ingi* and L1Tc (92 and 57% identity, respectively). The conservation of this 79 bp motif in presumably active non-LTR retrotransposons suggests that this sequence plays an important role in retroelement function. It has been proposed that the 5′-extremity of *ingi* and L1Tc contains an internal promoter responsible for the transcription of the element [47], as observed for the numerous non-LTR retroelements, including L1, Doc, I factor and Jockey [48–51]. However, Garcia-Salcedo et al. recently demonstrated that the first 55 bp of *ingi* do not display any promoter activity in transient transfection assays [28]. Alternatively, this conserved 5′-terminal motif might play a post-transcriptional role, such as observed for the cap present at the 5′-extremity of all eukaryotic mRNAs, i.e. mRNA stabilization and translation [52]. In support of this hypothesis, none of the *ingi* and L1Tc mRNAs contain the spliced leader cap added to the 5′-extremity of all trypanosomatid mRNAs by *trans*-splicing [26,53]. Comparison of the predicted RNA secondary structure of the *ingi* and L1Tc 5′-extremities (first 149 bp) shows a similar organization pattern, although about half of the sequence is different (Fig. 9). Interestingly, the 39 bp signature (92% identical between *ingi* and L1Tc) form a stem-loop involving the 5′-extremity in G–pairs, suggesting that the 5′-extremity of the mRNA encoding *ingi* and L1Tc might be protected against exonucleases. We propose that in the absence of the SL cap, the conserved 5′-extremity plays a role in structuring the mRNA extremity for stabilization and translation of the *ingi* and L1Tc mRNAs. mRNAs of other non-LTR retrotransposons, as well as all eukaryotic mRNAs, are thought to be capped. However, these questions have not been directly studied in the past [54]. Interestingly, this 79 bp signature present at the 5′-extremity of *ingi* and L1Tc is the only nucleotide sequence conserved in DIREs and is found in all the full length DIREs analyzed to date (108 elements). The reason of this conservation is puzzling and may suggest that these DIREs, which presumably do not code for functional proteins, are expressed and may have an impact on the host genome.

Among the six full length *T. brucei* DIREs identified, only one (Tb1b) is clearly flanked by a 12 bp target site duplication. Interestingly, most of the TSD flanking the *ingi* elements are 12 bp long [27]. This suggests that the Tb1b DIRE recently moved into the tubulin gene cluster using the *ingi* machinery, by the so-called *trans* retrotransposition, which has been previously described for other non-LTR retrotransposons [38,55–57]. For instance, the human L1 element displays a *cis*-preference, however, L1 can also function in *trans* to retrotranspose mutant L1 RNAs and to generate processed pseudogenes, though at a much lower frequency [38,55,56].

In conclusion, degenerate non-LTR retrotransposons related to the trypanosome potentially active elements were identified in the *L. major* genome, indicating that active elements present in the trypanosomatid ancestors were lost in this species. Since DNA transposons and retrotransposon-like elements were present in prokaryotes, it was expected that primitive eukaryotes also contained both classes of mobile elements. However, among the 15 unicellular eukaryote genomes sequenced so far (http://genomesonline.org/), 5 genomes lack mobile elements. We propose that a significant fraction of unicellular eukaryotic species or lineages have lost active mobile elements. Since mobile elements are selfish in the host genome, their secondary loss could be the result of natural extinction. However, we cannot rule out the possibility that unicellular eukaryotes have developed a strategy to eradicate mobile elements from their genome.

Acknowledgments

FB and TB were supported by the CNRS, the Conseil Régional d'Aquitaine, the Ministère de l'Education Nationale de la Recherche et de la Technologie. NES and co-workers were supported by NIH Grants AI43062 and AI45038. MJL has received partial support from Howard Hughes Medical Institute (Chevy Chase, USA), FONCYT 1-14389 and University of Buenos Aires.

References

[1] Capy P, Bazin C, Higuet D, Langin T. Dynamics and evolution of transposable elements. Austin, Tex: Landes Bioscience; 1998.

[2] Whitcomb JM, Hughes SH. Retroviral reverse transcription and integration: progress and problems. Annu Rev Cell Biol 1992;8:275–306.
[3] Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993;72:595–605.
[4] Eickbush TH, Malik HS. In: Craig AG, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, DC: ASM Press; 2002. p. 1111–44.
[5] Kazazian Jr HH. Mobile elements: drivers of genome evolution. Science 2004;303:1626–32.
[6] Katinka MD, Duprat S, Cornillot E, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 2001;414:450–3.
[7] Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002;419:498–511.
[8] Carlton JM, Angiuoli SV, Suh BB, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature 2002;419:512–9.
[9] Xu P, Widmer G, Wang Y, et al. The genome of *Cryptosporidium hominis*. Nature 2004;431:1107–12.
[10] Abrahamsen MS, Templeton TJ, Enomoto S, et al. Complete genome sequence of the apicomplexan *Cryptosporidium parvum*. Science 2004;304:441–5.
[11] Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309:416–22.
[12] El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005;309:409–15.
[13] Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite *Leishmania major*. Science 2005;309:436–42.
[14] Lukes J, Jirku M, Dolezel D, Kral'ova I, Hollar L, Maslov DA. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol 1997;44:521–7.
[15] Haag J, O'Huigin C, Overath P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Mol Biochem Parasitol 1998;91:37–49.
[16] Stevens J, Rambaut A. Evolutionary rate differences in trypanosomes. Infect Genet Evol 2001;1:143–50.
[17] Bhattacharya S, Bakre A, Bhattacharya A. Mobile genetic elements in protozoan parasites. J Genet 2002;81:73–86.
[18] Wickstead B, Ersfeld K, Gull K. Repetitive elements in genomes of parasitic protozoa. Microbiol Mol Biol Rev 2003;67:360–75.
[19] Vazquez M, Ben-Dov C, Lorenzi H, Moore T, Schijman A, Levin MJ. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. Proc Natl Acad Sci USA 2000;97:2128–33.
[20] Aksoy S, Lalor TM, Martin J, Van der Ploeg LH, Richards FF. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome *Trypanosoma gambiense*. EMBO J 1987;6:3819–26.
[21] Gabriel A, Yen TJ, Schwartz DC, et al. A rapidly rearranging retrotransposon within the miniexon gene locus of *Crithidia fasciculata*. Mol Cell Biol 1990;10:615–24.
[22] Villanueva MS, Williams SP, Beard CB, Richards FF, Aksoy S. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. Mol Cell Biol 1991;11:6139–48.
[23] Teng SC, Wang SX, Gabriel A. A new non-LTR retrotransposon provides evidence for multiple distinct site-specific elements in *Crithidia fasciculata* miniexon arrays. Nucleic Acids Res 1995;23:2929–36.
[24] Kimmel BE, Ole-MoiYoi OK, Young JR. *Ingi*: a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol Cell Biol 1987;7:1465–75.
[25] Murphy NB, Pays A, Tebabi P, et al. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. J Mol Biol 1987;195:855–71.
[26] Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. J Mol Biol 1995;247:49–59.
[27] Bringaud F, Biteau N, Zuiderwijk E, et al. The *ingi* and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. Mol Biol Evol 2004;21:520–8.
[28] Garcia-Salcedo JA, Gijon P, Amiguet-Vercher A, Pays E. Searching for promoter activity in RIME/*Ingi* retrotransposons from *Trypanosoma brucei*: binding of a nuclear protein to their 5′ extremity. Exp Parasitol 2003;104:140–8.
[29] Olivares M, Garcia-Perez JL, Thomas MC, Heras SR, Lopez MC. The non-LTR (long terminal repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. J Biol Chem 2002;277:28025–30.
[30] Pays E, Murphy NB. DNA-binding fingers encoded by a trypanosome retroposon. J Mol Biol 1987;197:147–8.
[31] Olivares M, Alonso C, Lopez MC. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. J Biol Chem 1997;272:25224–8.
[32] Ghedin E, Bringaud F, Peterson J, et al. Gene synteny and evolution of genome architecture in trypanosomatids. Mol Biochem Parasitol 2004;134:183–91.
[33] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–82.
[34] Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 1999;16:793–805.
[35] Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 1990;9:3353–62.
[36] Bringaud F, García-Pérez JL, Heras SR, et al. Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. Mol Biochem Parasitol 2002;124:73–8.
[37] Hasan G, Turner MJ, Cordingley JS. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. Cell 1984;37:333–41.
[38] Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked *Alu* sequences. Nat Genet 2003;35:41–8.
[39] Affolter M, Rindisbacher L, Braun R. The tubulin gene cluster of *Trypanosoma brucei* starts with an intact beta-gene and ends with a truncated beta-gene interrupted by a retrotransposon-like sequence. Gene 1989;80:177–83.
[40] Hall N, Berriman M, Lennard NJ, et al. The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. Nucleic Acids Res 2003;31:4864–73.
[41] Braun R, Behrens K, Glauser A, Brun R. Evolution of the retrotransposons TRS/*ingi* and of the tubulin genes in trypanosomes. Acta trop 1992;52:175–87.
[42] Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999;9:657–63.
[43] International-Human-Genome-Sequencing-Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
[44] Maslov DA, Simpson L. Evolution of parasitism in kinetoplastid protozoa. Parasitol Today 1995;11:30–2.
[45] Doolittle RF, Feng DF, Johnson MS, McClure MA. Origins and evolutionary relationships of retroviruses. Q Rev Biol 1989;64:1–30.
[46] Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM. Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. Genome Res 2005;15:36–43.
[47] Olivares M, Lopez MC, Garcia-Perez JL, Briones P, Pulgar M, Thomas MC. The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. Biochim Biophys Acta 2003;1626:25–32.

[48] Mizrokhi LJ, Georgieva SG, Ilyin YV. *jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 1988;54:685–91.
[49] Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 1990;10:6718–29.
[50] McLean C, Bucheton A, Finnegan DJ. The 5′ untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. Mol Cell Biol 1993;13:1042–50.
[51] Contursi C, Minchiotti G, Di Nocera PP. Identification of sequences which regulate the expression of *Drosophila melanogaste*r Doc elements. J Biol Chem 1995;270:26570–6.
[52] Liang XH, Haritan A, Uliel S, Michaeli S. *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell 2003;2:830–40.
[53] Vassella E, Roditi I, Braun R. Heterogeneous transcripts of RIME/*ingi* retroposons in *Trypanosoma brucei* are unspliced. Mol Biochem Parasitol 1996;82:131–5.
[54] Boeke JD. The unusual phylogenetic distribution of retrotransposons: a hypothesis. Genome Res 2003;13:1975–83.
[55] Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 2000;24:363–7.
[56] Wei W, Gilbert N, Ooi SL, et al. Human L1 retrotransposition: *cis* preference versus *trans* complementation. Mol Cell Biol 2001;21:1429–39.
[57] Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 2002;111:433–44.
[58] LaCount DJ, El-Sayed NM, Kaul S, Wanless D, Turner CM, Donelson JE. Analysis of a donor gene region for a variant surface glycoprotein and its expression site in African trypanosomes. Nucleic Acids Res 2001;29:2012–9.
[59] Olivares M, Thomas MC, Lopez-Barajas A, et al. Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. Electrophoresis 2000;21:2973–82.
[60] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406–15.