Review

Cell PRESS

The biology and evolution of transposable elements in parasites

M. Carmen Thomas[1], Francisco Macias[1], Carlos Alonso[2] and Manuel C. López[1]

[1] Departamento de Biología Molecular, Instituto de Parasitología y Biomedicina 'López Neyra' – CSIC, Parque Tecnológico de Ciencias de la Salud, 18100 Granada, Spain
[2] Centro de Biología Molecular 'Severo Ochoa'- CSIC, Madrid, Spain

Transposable elements (TEs) are dynamic elements that can reshape host genomes by generating rearrangements with the potential to create or disrupt genes, to shuffle existing genes, and to modulate their patterns of expression. In the genomes of parasites that infect mammals several TEs have been identified that probably have been maintained throughout evolution due to their contribution to gene function and regulation of gene expression. This review addresses how TEs are organized, how they colonize the genomes of mammalian parasites, the functional role these elements play in parasite biology, and the interactions between these elements and the parasite genome.

The TEs of mammalian parasites

The first description of transposable elements (TEs) in genomic DNA was made in 1950 by Barbara McClintock. Since then several types of TEs have been found in the genomes of most prokaryotic and eukaryotic organisms. The transposable elements have been referred to as 'junk' DNA, selfish sequences or genomic parasites, because in the same cell they can move from chromosome to chromosome. However, there are increasingly solid indications that transposable elements play a role in the evolution of genes and genomes of a wide range of organisms.

To date, several TEs have been described in the genomes of parasites that infect mammals. The number of parasite genomes that have been sequenced continues to increase, and the number of identified TEs in these organisms has grown logarithmically. We believe that the study of TEs present in these parasites is of particular interest because of the close relationship between the parasites and the organisms they inhabit (both bearing a large quantity of TEs). According to their mechanism of mobilization, TEs can be subdivided into two major groups: (i) those that transpose via a DNA intermediate, termed DNA transposons or class II elements, and (ii) those that move by reverse transcription of an RNA intermediate, termed retrotransposons or class I elements (Figure 1).

Here we review the association and functional implications of TEs within parasite genomes as well as the role of the mobile elements in genome evolution. The organization, distribution, structure and mobilization mechanisms of TEs in the parasite genome will be analyzed. In addition, we examine our current understanding of what can be learned from the analysis of parasite TEs and how the dialog between TE elements and their hosts can improve the adaptability and survival of each. Questions such as how can we harness TEs to improve parasite

Glossary

DDX$_n$D or DDX$_n$E: a triad of highly conserved amino acids (Asp–Asp–Asp [DDD] or Asp–Asp–Glu [DDE]) that is found in one class of transposases and that is required for the coordination of metal ions that are necessary for catalysis.
DIRE (degenerate *ingi*-L1Tc-related retrotransposon): DIREs are devoid of coding sequences.
EhMLBP (*Entamoeba histolytica* methylated LINE-binding protein): an essential constituent of *E. histolytica* epigenetic machinery and a potential drug target.
***env* (envelope)**: the retroviral *env* gene encodes protein precursors for the virion envelope proteins. These proteins reside in the lipid layer and serve to form the viral envelope. In the HIV-1 genome, *env* genes encode 2 envelope proteins: *gp120* and *gp41*.
***gag* (group-specific antigen)**: the retroviral *gag* gene encodes the virus nucleoproteins, nucleocapsid proteins and matrix proteins. The encoded polypeptides products are typically synthesized as protein precursors or polyproteins that are then cleaved by viral proteases to yield the final products. The final products are generally associated with the nucleoprotein core of the virion, and the 'gag' protein itself is the protein of the nucleocapsid shell around the RNA of a retrovirus.
ICR (internal complementary repeats): related or identical sequences of deoxyribonucleic acid in inverted form occurring at opposite ends of some transposons.
ITR (inverted terminal repeats): repeat sequences in reverse orientation that serve as the viral origins of replication. ITRs are essential for virus packaging and rescue of the integrated viral genome.
LINE (long interspersed nuclear element): an autonomous retrotransposon whose mobility is dependent on target-primed reverse transcription.
L1Tc (LINE 1 *Trypanosoma cruzi*): a non-LTR LINE1 element from *T. cruzi*.
LTR (long terminal repeat): a long sequence (usually 200 to 600 bp in length) repeated at each end of a retroviral DNA or retrotransposon and that is necessary for reverse transcription, integration and transcription. The LTRs include promoter and termination sequences for RNA polymerase II.
Miniexon or spliced leader: short sequences that are joined to the 5′ ends of pre-mRNAs by trans-splicing. They are found primarily in primitive eukaryotes (protozoan and nematodes).
***pol* (polymerase) gene**: a retroviral gene that encodes reverse transcriptase, RNase H and integrase.
RNP (ribonucleoprotein): a protein–mRNA complex that mediates all the steps of retrotransposition, ensuring the fidelity of the reaction.
SIDER (short interspersed degenerate retroelement): SIDERs are short non-autonomous heterogeneous elements related to retrotransposons identified in trypanosomes (*ingi*/RIME, L1Tc/NARTc, and DIRE).
SINE (short interspersed nuclear element): a non-autonomous retrotransposon that requires the enzymatic machinery of LINE for its mobilization.
Telomere: a region of repetitive DNA located at the end of chromosomes. Telomeres are usually maintained by telomerase, a ribonuclear protein particle (RNP) that includes an RT enzyme and an RNA molecule that is reverse transcribed.
Tpase (transposase): the enzyme responsible for the catalysis of transposition.
Transposable element (TE): a DNA sequence that can move to different positions within the genome of a cell.
TSD (target site duplication): LINE flanking sequences corresponding to the sequence between the two single-strand nicks performed by a LINE-encoded endonuclease at the target site of the element insertion.
UTR (untranslated region): regions of mature mRNA that do not code for proteins, including the 5′- and 3′-UTRs. UTRs contain information for the regulation of translation and mRNA stability.

Corresponding author: López, M.C. (mclopez@ipb.csic.es).

1471-4922/$ – see front matter © 2010 Elsevier Ltd. All rights reserved. doi:10.1016/j.pt.2010.04.001 Available online xxxxxx

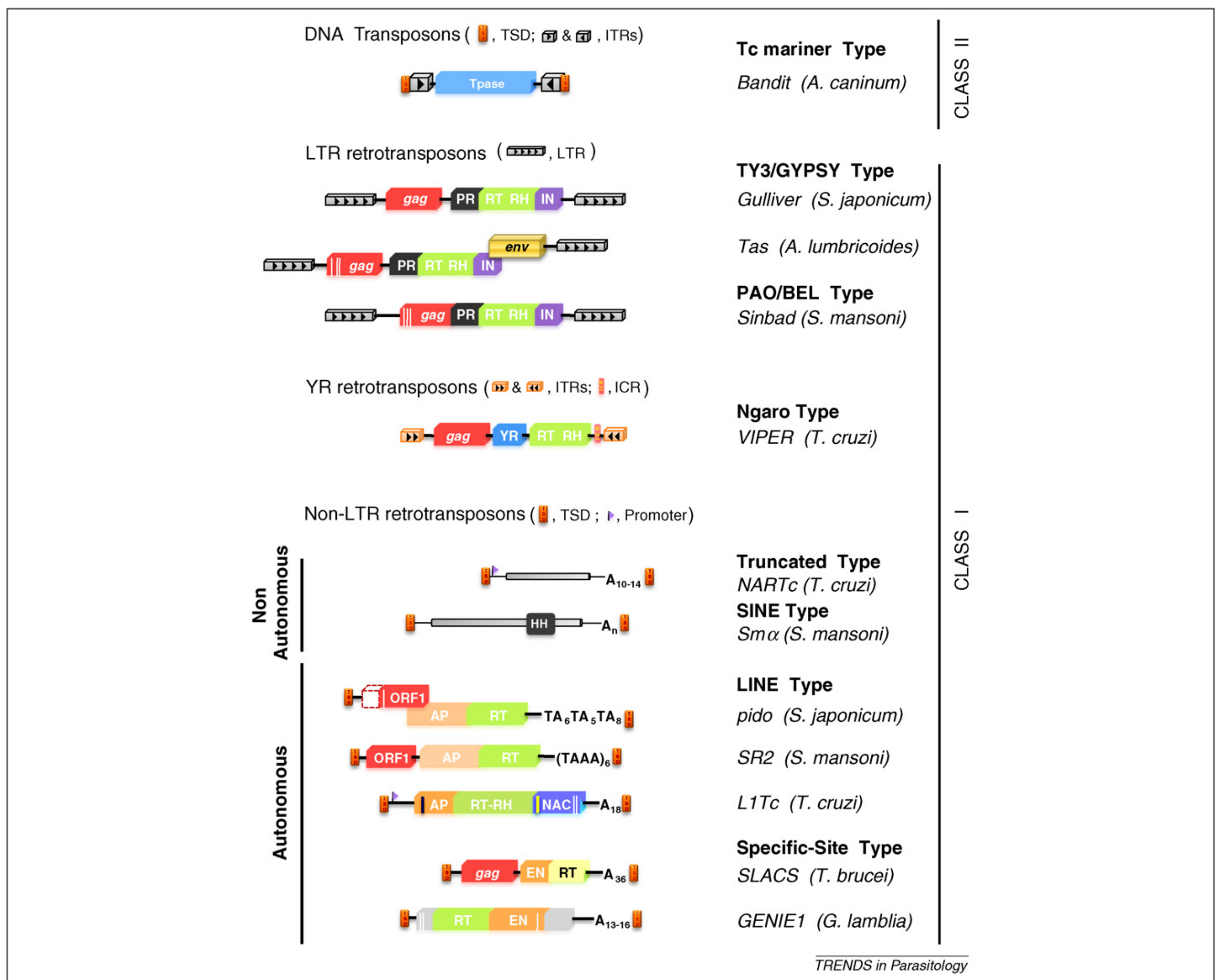

Figure 1. Schematic representation of different types of transposable elements from parasites infecting mammals. Most class II TEs or DNA transposons from parasites encode a Tpase, ITRs, and TSD. Class I elements are divided into three major groups: **(i)** LTR retrotransposons, **(ii)** YR retrotransposons, and **(iii)** non-LTR retrotransposons. The latter elements are differentiated into non-autonomous elements and autonomous elements. An example of each type of element is given in each case. The positions of sequences for *gag*, protease (PR), reverse transcriptase (RT), RNase H (RH), integrase (IN), tyrosine recombinase (YR), AP endonuclease (AP), endonuclease (EN), nucleic acids chaperone (NAC) and hammerhead (HH) domains are indicated. Nucleic acid binding domains in LTR and non-LTR retrotransposons are represented as vertical white lines. In the *L1Tc* element the blue and yellow vertical lines correspond to the 2A self-cleaving sequence and the nuclear localization signal (NLS), respectively. Abbreviations: Tpase, transposase protein; ITRs, inverted terminal repeats; TSD, target site duplication; UTRs, untranslated regions (shown as black horizontal lines); LTR, long terminal repeats; ICR, internal complementary repeats; the flags represent an internal promoter.

control and whether there are differences in TE content in the genomes of parasites compared to those of non-parasitic organisms will also be addressed.

Organization, structure, distribution and mobilization mechanisms of TEs in host genomes

DNA class II transposons

DNA class II transposons are the oldest group of TEs and are present in all kingdoms. The DNA class II transposons identified in eukaryotes belong to two subclasses. Subclass 1 predominantly comprises the so-called cut-and-paste DNA transposons, whereas subclass 2 comprises both the rolling-circle DNA transposons (*Helitrons*) and the self-synthesizing DNA transposons (*Maverick*, also known as *Polintons*) [1]. Most of the described DNA transposons (Table 1) in the genomes of parasites that infect mammals code for a transposase enzyme (Figure 1) that bears the catalytic domain DDX_nD or DDX_nE. These elements are flanked by target-site duplications (TSD) resulting from the cutting and ligation of the DNA element into the insertion site by a transposase. Frequently, these DNA transposons contain inverted terminal repeats (ITRs) that are recognized by the DNA-binding domains of the transposase and are important for transposon excision by the enzyme [2]. A list of DNA transposons from parasites that infect mammals and some of their structural and functional characteristics are presented in Table 1 [3–12], and a scheme for the mechanism of their mobilization is shown in Figure 2a.

Retrotransposons or class I elements

The class I of TEs includes LTR-retrotransposons, tyrosine recombinase retrotransposons and non-LTR retrotransposons. The sizes of parasite LTR-retrotransposons range

Table 1. Principal mobile genetic elements of parasites of mammals

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
ARTHROPODS								
Sola2-1_IS ***Sola2-2_IS***	DNA transposons; *Sola2 group*	*Ixodes scapularis*	5.1 5.5		Transposase (DDD motif); contains CCCC binding-motif	4 6	ITRs (712 and 900 bp, respectively) containing 5′-GRG and CYC-3′ termini	[3]
HELMINTHS								
bandit	DNA transposon; *Mariner-like* [A][a]	*Ancylostoma caninum*	1.3	1	Transposase (D,D34D motif)		Possible horizontal genetic transfer between host and parasite	[4]
mle-1	DNA transposon; *Mariner-like* [A]	*Trichostrongylus colubriformis*	0.9	1	Transposase	50	Transposase is closely related to *C.elegans* transposase element Inserted into the *tar-1* gene (X-chromosome)	[5]
Tas	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*Ascaris lumbricoides*	~7	3	ORF1: *gag-like* (leucine zipper, aspartic protease) ORF2: *pol*-like (RT, RNase H, integrase) ORF3: *env*-like	50	Incomplete *pol*-like (ORF2)	[13]
R4	Non-LTR; site-specific; R4-clade [A]	*A. lumbricoides*	4.7	1	EN; RT; nucleic acid binding motif (CCHC)		Target site of insertion: rRNA 26S gene Closely related to *Dong* element *(B. mori)* Homologs in *Paragonimus equorum* and *Hemonchus contortus*	[36]
dingo 1 & 2	Non-LTR; RTE-clade [A]	*Ancylostoma caninum*	~3	1	AP-endonuclease, RT	100–1000	Both closely related to RTE-1 from *C. elegans*, to BDDF from *Bos taurus* and to SR2 from *S. mansoni*	[37]
Merlin_Sm1	DNA transposons	*Schistosoma mansoni*	1.4	1	Transposase (DDE motif)	~500	First family of DNA transposons described in flatworms Related to the IS1016 group of bacterial insertion sequences	[6]
Merlin_Sj1		*S. japonicum*					Contains 24 bp ITRs	
SmTRC1	DNA transposon; *CACTA Superfamily*	*S. mansoni*	4.5	1	Transposase_21 domain (putative DDE motif) Contain CXXC motif (putative HHCC DNA-binding domain)	30–300	Contain the CACTA sequence in ITRs (54 bp), flanked by TSDs; transcriptionally active Generate multiple alternatively-spliced transcripts	[7]
Gulliver	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*S. japonicum*	4.8	2	ORF1: *gag*-like ORF2: *pol*-like (aspartic protease, RT, RNase H)	100–1000	The two ORFs are separated by one or several stop codons	[16]
Boudicca	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*S. mansoni*	5.9	3	ORF1: *gag*-like (CCHC motif) ORF2: *pol*-like (aspartic protease, RT, RNase H, integrase) OFR3: *env*-like	>1000	Putative *env*-like protein (ORF3) *A Boudicca-like* element is also present in *S. haematobium*	[14]
Sinbad	LTR retrotransposon; *Pao/BEL-like* [A]	*S. mansoni*	6.3	1	*gag*-like (3 CCHC motifs); *pol*-like (aspartic protease, RT, RNase H, integrase)	50	Contain a triple Cys-His RNA binding motif characteristic of *gag* from *Pao/BEL* elements A *Sinbad-like* element is also present in *S. haematobium*	[17]
Fugitive	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*S. mansoni*	4.8	1	*gag*-like (2 Cys/His motifs); *pol*-like (protease, RT, RNase H, integrase)	2000	Transcribed in at least six developmental stages of *S. mansoni*	[18]
Saci-1	LTR retrotransposon; *BEL-like* [A]	*S. mansoni*	6	1	*gag*-like (3 CCHC); *pol*-like (protease, RT, RNase H, integrase)	70–700		[19]
Saci-2 & 3	LTR retrotransposons; *Gypsy/Ty3-like* [A]	*S. mansoni*	4.9 5.2	1	*gag*-like (CCKCH)/(CHCC); *pol*-like (protease, RT, RNase H, integrase)	85–850 150–1500	*Saci-3* is closely related to *Boudicca (S. mansoni)* and *CsRn1*	[19]

Table 1 (*Continued*)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
CsRn1	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*Clonorchis sinensis*	5	1	*gag*-like (CHCC motif); *pol*-like (protease, RT, RNase H, integrase)	>100	Instead of a conventional *gag* motif (CCHC), *CsRn1* has a CHCC motif	[20]
PwRn1	LTR retrotransposon; *Gypsy/Ty3-like* [A]	*Paragonimus westermani*	3.6	1–2	*gag-like* (CHCC motif); *pol*-like (protease, RT, RNase H, integrase)	>1000	Several expression strategies: (i) ORF similar to *CsRn1* (ii) two ORFs within the same coding region (iii) ORFs overlapped by -1 frameshifting to encode *gag* and *pol* proteins	[21]
SR1, SR2 & SR3	Non-LTR; RTE-clade [A]	*S. mansoni*	2.3	1	AP-endonuclease, RT	2000–20000	*SR2* contains a nucleic acid binding domain (ORF1) The two ORFs are not overlapped	[38,39]
			3.9	2		2600–26000		
			~3.2	1		>1000		
SjR2	Non-LTR; RTE-clade [A]	*S. japonicum*	~3.9	1	AP-endonuclease, RT	>10000	Actively replicated	[16]
pido	Non-LTR; CR1-clade [A]	*S. japonicum*	3.6	2	ORF1: nucleic acid binding domain CCHC ORF2: AP-endonuclease, RT	~1000		[16]
Perere	Non-LTR; CR1-clade [A]	*S. mansoni*	4.9	1	Endonuclease, RT	250–2500	Closely related to *pido (S. japonicum*	[19]
Smα, Sjα Shα	Non-LTR; SINE-*like* elements [NA][b]	*S. mansoni* *S. japonicum* *S. haematobium*	<0.5		Hammerhead domain (ribozyme)	10000	Contain an RNA *pol*-III promoter element, a 3′poly(A) stretch; flanked by short TSDs sequences	[25]
SR2 DEL	Non-LTR; SINE-like element, RTE-clade [NA]	*S. mansoni*	0.1–0.5				Consists solely of the 5′ and 60–80 bp of 3′ UTRs of full-length SR2 elements, have both ORFs deleted	[38]
PROTISTS								
Tvmar1	DNA transposon; *Mariner Superfamily*	*Trichomonas vaginalis*	1.3	1	Transposase (D,D34D motif)	600–650	First *mariner* to be found in a protist Flanking ITRs (28 bp) and TSD (~9 bp) Presence of TA dinucleotide (the typical target site for *mariner* elements) near to ITRs Putative HTH and 2 NLS motifs were identified	[8]
TvMULE1	DNA transposon; *Mutator superfamily (MuDR clade)*	*T. vaginalis*	2.1	1	Transposase D34E motif (zinc finger CCHC-type domain)	~61	ITRs (31 bp) and TSD (~9 bp)	[9,10]
TvMULE 2	DNA transposons; *Mutator-like elements (TvCaMULE clade)*	*T. vaginalis*	2.5	1	Transposase (related to transposases of *C.albicans* elements)	~514	ITRs (20 bp) and TSD (10–14 bp)	[10]
TvMULE 3			2.9			~666	ITRs (30 bp) and TSD (11 bp)	
TvMULE 4			2.4			~1204	ITRs (38 bp) and TSD (12–13 bp)	
Mav-Tv1	Putative DNA transposon; *Mavericks group*	*T. vaginalis*	~15–20	various	Retroviral-like c-integrase; 5 genes, CMG 1–5: (1) DNA polymerase (2,3) coiled-coiled domain protein (4) ATPase (5) cysteine protease	~3000	Related to DNA viruses and encode 9–11 putative proteins Contain ITRs (400–700 bp) and are flanked by 5–6 bp putative TSDs Contain multiple independent transcription units; each gene possesses its own promoter	[6]
Polinton-1_TV	DNA transposon; *Polinton group* [A] or [NA]	*T. vaginalis*	20.7	10	Up to 10 proteins: DNA polymerase B, integrase, 2 ATPases (ATP and ATP1) and 6 additional proteins (PTV1–PTV6)		Constitute ~5% of *T. vaginalis* genome Share features with *Helitrons* elements (rolling-circle DNA transposons) but are self-synthesizing DNA transposons Contain 5′-AG and TC-3′ long ITRs (160 bp) and are flanked by 6 bp TSDs	[11]

Table 1 (*Continued*)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
Genie 1 (GilM)	Non-LTR; site-specific; (NC)[c] [A]	*Giardia lamblia*	5.5	1	2 CCHH motifs; RT; endonuclease (CCHC and PDXXXD motifs)	7–8	Inserted into 771 bp DNA repeats ORF of *Genie 1* contains an additional 150 aa segment at its C-terminus not found in *R2* or other non-LTR elements 3′-UTR of over 1.5 kb; 5′-UTR only 15–18 bp	[40]
Genie 1A (GilT)	Non-LTR; site-specific; (NC) [A]	*G. lamblia*	6	1	At least one CCHH motif; RT; endonuclease (CCHC and PDXXXD motifs)		Related to *Genie 1* element but divergent in sequence As in *Genie 1*, the Genie 1A ORF contains a 150 aa extension beyond the endonuclease domain	[40]
Genie 2 (GilD)	Non-LTR; site-specific; (NC) [A]	*G. lamblia*	3	1	RT; endonuclease (CCHC and PDXXXD motifs)	~10 (degenerate)	All copies contain inverted repeats up to 172 bp in length	[40]
EMULE-Eh	DNA transposon; *Mutator superfamily (EMULE clade)*	*Entamoeba histolytica*	3–5	1	Transposase (~700–900aa)	2	Relatively long ITRs (~100–200 bp) and TSD (~9 bp) Weak similarity with the IS256 group of prokaryotic Insertion sequences (IS)	[12]
EhLINE1, 2 & 3	Non-LTR; LINE; R4 clade [A]	*E. histolytica*	~4.8	1–2	ORF1 (coiled-coil domain) ORF2: RT, endonuclease (CCHC and PDXXXD motifs); the endonuclease (a restriction enzyme-like type) is closely related to *Dong*, *R4*, *SLACS*, *R2* endonuclease	140	Frequently located near protein-coding genes; predominently inserted into AT-rich sequences TSDs of most inserted elements are preceded by a short T-rich stretch Similar elements in *E. Dispar* genome	[41,42]
EhSINE1, 2 & 3	Non-LTR; SINE [NA]	*E. histolytica*	0.6				Frequently located near protein-coding genes Only one copy of *EhSINE3* has been described Homologous elements are present in *E. Dispar*	[41,42]
ingi	Non-LTR; LINE [A]	*Trypanosoma brucei*	~5.2	1	AP-endonuclease, RT; DNA binding domain (CCHH)	~500	Not randomly distributed in the genome Preceded by a highly conserved sequence with a consensus pattern (Pr77–79 hallmark) Contain a Poly(A) tail	[43,44]
L1Tc	Non-LTR; LINE [A]	*T. cruzi*	4.9	1	AP-endonuclease, RT, RNase H; nucleic acid binding domain (CCHH); nuclear localization signal (NLS)	320–1000	Present in all *T. Cruzi* chromosomes Contain an internal RNA polymerase II-dependent promoter *(Pr77)* Contain a functional virus-like self-cleaving 2A sequence and a poly(A) tail	[45]
RIME	Non-LTR; *ingi* truncated element [NA]	*T. brucei*	~0.5			~500	Inferred to be mobilized by *ingi*-encoded enzymatic machinery; Pr77–79 hallmark; poly(A) tail	[46]
NARTc	Non-LTR; *L1Tc* truncated e lement [A]	*T. cruzi*	~0.3			133	Contain an internal RNA *pol-II* dependent promoter *(Pr77)* Inferred to be mobilized by L1Tc enzymatic machinery; poly(A) tail	[27]
SIRE	*VIPER* truncated element [NA]	*T. cruzi*	0.4			1500–3000	Frequently located near telomeres and linked to protein-coding genes Highly conserved among *T.cruzi* strains	[24]
VIPER	Tyrosine recombinase	*T. cruzi*	4.5	3	ORF1: *gag*-like ORF2: tyrosine recombinase ORF3: RT-RNase H		Structurally related to the tyrosine recombinase retroelements *DIRS (D. discoideum)* and *Ngaro (zebrafish)*	[24]

Table 1 (Continued)

Name	Type / clade	Host organism	Size (kb)	ORFs	Domains	Copy number	Additional information	Refs.
LmSIDER1	Degenerate retrotransposon	*Leishmania major*	0.6			785	LmSIDERs could play a role in post-transcriptional downregulation of gene expression High GC content. Some bear Pr77–79-like hallmark and poly(A) tail	[47,48]
LmSIDER2 and LiSIDER2	Degenerate retrotransposons	*L. major* *L. infantum*	0.6		Pr77–79 promoter and poly(A) stretch hallmarks.	1073	Predominantly located in the 3′UTR of *L. major* mRNAs Most abundant TEs in trypanosomatid genomes High GC content Some bear Pr77–79-like hallmark and poly(A) tail	[47,48]

[a][A], autonomous element.
[b][NA], non-autonomous element.
[c](NC), not classified.

from 3.5 to 10 kb, a length similar to that of LTR-retrotransposons found in other organisms. These elements are flanked by long terminal repeats (LTRs) of variable length (~200 to 600 bp) in a direct orientation. Most of them contain two open reading frames (ORFs) (Figure 1). The protein encoded by the first ORF is similar to the retrovirus *gag* protein, although it exhibits high sequence variability. The second ORF, similar to *pol* genes of retroviruses, encodes a protease (PR), a reverse transcriptase (RT), an RNase H (RH) and an integrase (IN). The function of these genes is required for the synthesis of cDNA via reverse transcription for insertion into the genome (Figure 2b). Some LTR elements, such as *Tas* from *Ascaris lumbricoides* and *Boudicca* from *Schistosoma mansoni*, bear a third ORF (Figure 1) resembling the *env* gene from retroviruses [13,14]. As with retroviral env proteins, the third ORF could be involved in attachment of the parasite to the host cell to permit internalization (Figure 2b). In addition, the *Tas* envelope protein might have been obtained from *Herpesviridae* (double-stranded DNA viruses with no RNA stage), representing a case in which the *env* gene of a retrovirus has been traced back to its original source following the proposal that retroviruses evolved from LTR retrotransposons [15]. A list of the LTR-retroelements identified in mammalian parasites and some of their characteristics are presented in Table 1 [13,14,16–21].

The tyrosine recombinase retrotransposons (YR-retrotransposons) include three groups of elements (*DIRS*-like, *Ngaro* and *VIPER*) that harbor enzymatic domains similar to those present in LTR-retrotransposons. They contain, however, a tyrosine-recombinase (YR) domain instead of an integrase and protease domains [22,23]. Unlike LTRs, their terminal repeats are inverted in orientation (ITRs). A segment of the ITR sequences is repeated within the element, giving rise to internal complementary repeats (ICRs) (Figure 1). The transposition mechanism described for this type of element is similar to that described for retroviruses and LTR-retrotransposons [23]. However, biological assays to directly address the mechanism of reverse transcription and insertion of the YR-retrotransposon have not been reported [15]. To date, *VIPER* is the only YR-retrotransposon identified in a parasitic organism. This element was initially described in *Trypanosoma cruzi* associated with a short interspersed repetitive element (*SIRE*) [24]. Elements homologous to *VIPER* have been found in the genomes of *T. brucei* and *T. vivax* [24].

The non-LTR retrotransposons are not flanked either by inverted repeats or by LTR. However, these elements are flanked by direct duplications of 5 to 20 nucleotides in length and are actively transcribed as an intermediate mRNA (Figure 1). In this class of elements, two main groups can be discerned depending on whether they encode the proteins needed for their own retrotransposition or whether these coding sequences are absent. The latter group are termed non-autonomous transposable elements because they rely on full-length autonomous non-LTR retrotransposons or LINEs (long interspersed nuclear elements) to provide all the protein components required for mobility. They are short in length and have at their 3′ ends a putative recognition sequence for reverse transcriptase (RT) binding. The characteristics of both

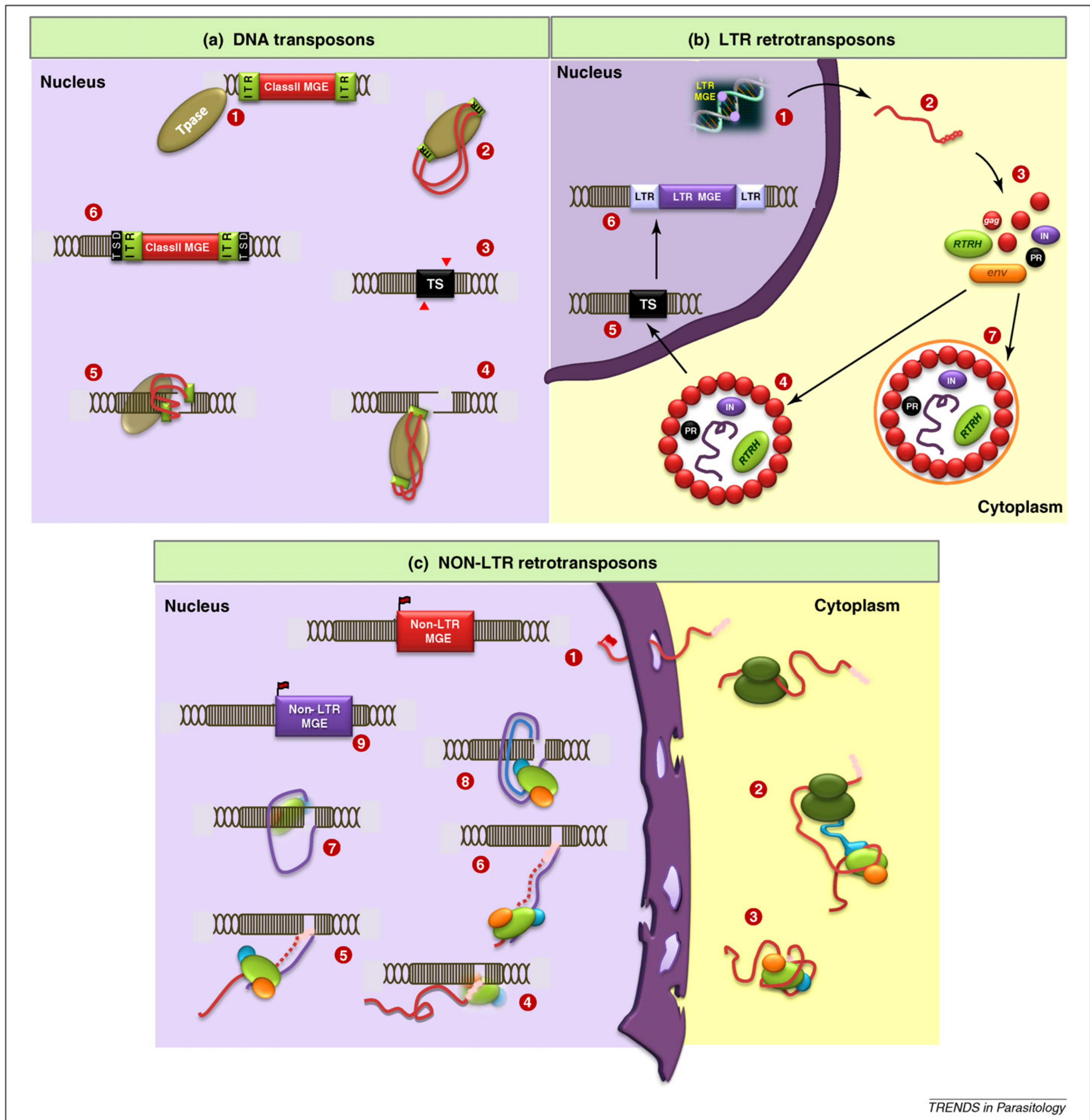

Figure 2. Mobilization mechanism of DNA transposons and retrotransposon elements. **(a)** DNA transposons: (1) the Tpase acts as an endonuclease and exscinds an integrated copy of a DNA transposon; (2) the Tpase recognizes the terminal inverted repeats and binds both ends of the exscinded sequence; (3) the complex Tpase-DNA transposon recognizes the target site (TS) for insertion into the genome; (4) the Tpase produces a staggered cut at the target site of the genome; (5,6) insertion of the exscinded DNA transposon and ligation. The inserted TE is then flanked by target site duplications (TSDs). **(b)** LTR-retrotransposons: (1,2) transcription of the element and mRNA export to the cytoplasm. The LTRs contain initiation and termination signals for transcription of the RNA intermediate; (3) translation of the element and generation of the enzymatic machinery for transposition; (4) reverse transcription of the LTR-retrotransposon by the reverse transcriptase (RT); a virus-like particle (VLP) is formed and dsDNA is generated by the RT and the nucleic acid chaperone. The function of the chaperone is to facilitate the rearrangements of the reverse-transcribed nucleic acid copies and promote the formation of properly folded and stabilized duplex conformations. The LTRs are involved in template jumps of the RT from one end of the transcript to the other resulting in the generation of the new copy of the element; (5) the new copy migrates to the nucleus and a potential site for insertion is detected; (6) the element is inserted into the genome through the integrase encoded by the element. (7) Elements that bear an *env* gene (such as *Tas* from *A. lumbicoides* and *Boudicca* from *S. mansoni*) are able to form an envelope that facilitates entry into the host cell. **(c)** non-LTR retrotransposons or LINEs: (1) Transcription of the LINE from an internal pol II promoter encoded in the 5′-end of the mRNA (red flag) takes place in the nucleus, and the full-length transcript is exported to the cytoplasm. (2) LINE is translated, generating the proteins that form the enzymatic machinery needed for transposition. These proteins bind to the LINE mRNA to generate a ribonucleoprotein particle (RNP); (3) the LINE RNP is transported into the nucleus; (4) DNA minus-strand cleavage takes place at the target insertion site, a process catalyzed by the endonuclease; (5,6) cDNA synthesis employing the LINE mRNA as template and hydrolysis of the RNA molecule of the RNA–cDNA duplex mediated by the RT and RNase H; (7) cleavage of the upper DNA strand, a process probably mediated by the element-encoded endonuclease; (8) second-strand cDNA synthesis mediated by the DNA polymerase activity of the RT, followed by repair and ligation; (9) the new copy of LINE is integrated into a new site of the genome.

non-autonomous and autonomous non-LTR retrotransposons of parasites are described in detail in Box 1 [25–35].

Most LINEs have two ORFs with a genomic organization similar to that of LTR-retrotransposons. The first ORF has homology with the *gag* genes and encodes a nucleic-acid-binding protein. The second ORF has similarity with *pol* genes. This ORF codes for at least a protein having RT and endonuclease activities (EN) [15]. A list of non-LTR retrotransposons from parasites of mammals and some of their characteristics are presented in Table 1 [16,19,25,27,36–48].

Given the crucial functional role that the LTRs play in the mobilization of LTR-retrotransposons, it is likely that non-LTR retrotransposons use a different retrotransposition–integration mechanism. In non-LTR retrotransposons the intermediate mRNA has to be transported into the nucleus for reverse transcription and integration via target-primed reverse transcription (TPRT) [15,49]. This is a process in which cleaved DNA targets are used to prime reverse transcription of the element's RNA transcript. This mechanism was initially predicted by analysis of the *Bombyx mori R2* element [49]. It appears to constitute a mechanism common to the mobilization of all LINEs [15]; a schematic representation of the mobilization mechanism of non-LTR retrotransposons is shown in Figure 2c. The best-characterized parasite retrotransposon in terms of the enzymology of mobilization is *L1Tc* from *T. cruzi* (Box 2) [50–58]. Despite the fact that this element encodes functional proteins required for TPRT, there is no direct evidence that *L1Tc* is functionally active.

Association and functional implications of TEs in parasite host genomes

In some protozoan parasite organisms the chromosomal locations of several TEs have been found to be clustered [12,30,42,59]. In *Entamoeba histolytica* and *E. dispar* genomes the highest TE densities are found in about 24% of their genomes [12]. This clustered organization of non-autonomous and autonomous retrotransposons has been also described in *T. cruzi* and *T. brucei* genomes [30,59]. The association could be a consequence of the mobilization

Box 1. Non-autonomous and autonomous non-LTR retrotransposons from mammalian parasites

SINEs (short interspersed nucleotide elements) are the most abundant non-autonomous elements and are shorter than 1 kb in length. The 5′ ends of eukaryotic SINEs bear an internal pol III promoter derived from host genes such as tRNAs, 7SL RNA, and 5S rRNA. In *S. mansoni* the pol III promoter from *Sm*α SINE element is derived from tRNA [25]. The central region of SINEs, termed core or body, is highly variable and is sometimes not present in these elements. The 3′ ends of SINEs contain a poly(A) tail. These characteristics can be considered as hallmarks of mature mRNAs and indicate that the origin of these elements could very well be a processed RNA that, after being copied, was integrated into a new position within the genome [26]. Furthermore, the *Sm*α*, Sh*α and *Sj*α elements from different species of *Schistosoma* encode an active ribozyme that bears a hammerhead domain; this domain was probably acquired by vertical transmission from a common *Schistosoma* ancestor [25]. Another type of non-autonomous element corresponds to truncated versions of the LINEs present in a host that has lost the enzymatic machinery for retrotransposition but has retained a 5′ polymerase II internal promoter and a 3′ poly(A) tail. That is the case for both *NARTc*, that bears the first 77 nt of *L1Tc* at its 5′-end, and *RIME*, that bears the conserved 79 nt fragment from the *ingi* 5′-end [27].

The autonomous non-LTR retrotransposons (LINEs) are long elements of about 5 kb in length that generally bear a 3′ poly(A) tail. These elements typically code for the enzymes implicated in their transposition including a reverse transcriptase (RT) that is present in all autonomous LINE elements. Depending on whether they are inserted in specific and non-specific sites in the genome, two types of LINEs may be described. Examples of site-specific non-LTR retrotransposons of parasites are the *SLACS* from *T. brucei* and *CZAR* from *T. cruzi* that insert between nucleotides 11 and 12 of the miniexon of these protozoan parasites. Transcription of these elements takes place from the spliced leader (SL) RNA promoter and the RNA polymerase II α-amanitin sensitive promoter [28]. An element, *TcTREZO*, was recently found to be inserted into hot-spots for homologous recombination in *MASP* genes in the *T. cruzi* genome [29]. An example of a parasite non-LTR retrotransposon that is not site-specific is the *L1Tc* element from *T. cruzi*. The element is present in most *T. cruzi* chromosomes and is frequently located in genomic regions that are rich in repetitive DNA sequences [30], often in association with the ABC genes [31] and with the subtelomeric regions of the parasite chromosomes [32,33]; however, as shown by *in silico* analysis, the *L1Tc* seems to display some site-specificity for insertion because a conserved pattern has been observed upstream of the site where this element is inserted [34]. The *ingi* element from *T. brucei*, as well as the *ingi*-related elements from *T. congolense* and *T. vivax*, maintain a similar distribution to that of *L1Tc* from *T. cruzi* [35].

Box 2. Enzymatic machinery from the *L1Tc* element

The *L1Tc* LINE from *T. cruzi* actively transcribes a polyadenylated mRNA [45]. Translation of *L1Tc* from *T. cruzi* gives rise to a polyprotein of 1574 amino acids. The 5′ end of the element codes for a protein endowed with AP endonuclease activity [50] as well as 3′-phosphatase and 3′-phosphodiesterase activities [51]. The endonuclease present in *L1Tc* is thought to cleave the minus strand of the genomic DNA at the target insertion site, generating a free 3′-hydroxyl end that is used by the RT as a primer to initiate reverse transcription. Similar to other LINE elements, the endonuclease encoded by *L1Tc* contains conserved domains involved in DNase I acid–base catalysis [52], suggesting that this enzyme can recognize sequence-dependent structural variations. A second single-stranded nick is produced on the plus strand, a few nucleotides downstream of the first nick, by the *L1Tc*-encoded endonuclease, generating a primer for second strand synthesis.

The central region of the *L1Tc* element codes for a protein with RT and DNA polymerase activity with the ability to use both homologous and heterologous templates [53]. The existence of RT and DNA polymerase activities associated with the polyprotein encoded by *L1Tc* facilitates the reverse transcription of the mRNA of the element and the generation of the complementary strand of the cDNA. The RT from *L1Tc* is also capable of synthesizing cDNA molecules by consecutive switching of the oligonucleotide molecule used as a template [53]. A similar activity capable of template switching has been proposed to be involved in the retrotransposition of non-LTR elements, and it is thought to be a property of non-LTR RTs [54]. Downstream and in-frame with the RT-coding domains, *L1Tc* codes for a protein with RNase H activity [55]. The *L1Tc* element is one of the few non-LTR retrotransposons that contains RNase H conserved domains [56].

The sequence located downstream of RNase H domain of *L1Tc* codes for a protein that has both nucleic acid chaperone activity (NAC) and high affinity for single- and double-stranded nucleic acid molecules [57]. This NAC activity could be involved in a TPRT process promoting both stabilization and destabilization of the nucleic acid helix, thus facilitating strand exchange between DNA strands [57,58]. The non-LTR retrotransposons are flanked by a direct repeat sequence derived from sequences located between the two single-stranded nicks generated by the element-encoded endonuclease, known as target-site duplications (TSDs).

of the non-autonomous retrotransposons by the protein machinery of autonomous retrotransposons. Because nascent autonomous retrotransposon-encoded proteins that mediate mobilization are bound to the intermediate mRNA during translation, the machinery of autonomous retrotransposons could generate retrotransposition of the non-autonomous retrotransposons *in trans* due to their proximity (Figure 3a). The association of TEs could also be a consequence of the existence of a strong negative selection against insertion of these elements into gene-coding regions in order to minimize deleterious effects (Figure 3b). Consistent with this hypothesis is the existence of a sequence-pattern preference for insertion of TEs in the parasite genomes. A multigene family called retrotransposon hot-spots (RHS) has been characterized that contains a hot-spot for insertion of the *ingi/RIME* retrotransposons in the *T. brucei* genome [60]. *T. cruzi* RHS-related sequences also contain a hot-spot for *L1Tc/NARTc* insertion [27]. In the genomes of *T. brucei* [44] and *T. cruzi* [34], the *ingi/RIME* and *L1Tc/NARTc* elements, respectively, are preceded by a sequence that maintains a highly-conserved consensus pattern and by a conserved TSD sequence that results from single-stranded nicks generated by the endonuclease. However, in the genomes of these parasites there are also elements that are not flanked by TSD. It is believed they were mobilized by homologous recombination processes (Figure 3c).

TEs can modify and manipulate the parasite genome in diverse ways. Consistent with the selfish DNA theory that considers TEs as 'molecular parasites', the mobility of transposable elements could produce a variety of detrimental effects on the parasite host genomes. All in all, harmful insertion of TEs will probably be minimized whereas those insertions that are beneficial are likely to be retained during genome evolution. In fact, several reports have shown that the TEs can substantially contribute to genome plasticity and to genome architecture. Thus they might play a fundamental role as drivers of genome evolution, shaping both genes and genomes with regard to function and structure [61].

The co-option by the host genome of TE sequences that generate new regulatory signals or functional genes useful to the host is a process referred to as 'molecular domestication' of transposable elements. It thus seems that many TEs have been domesticated and have evolved to fulfill essential functions in genome dynamics, for example by

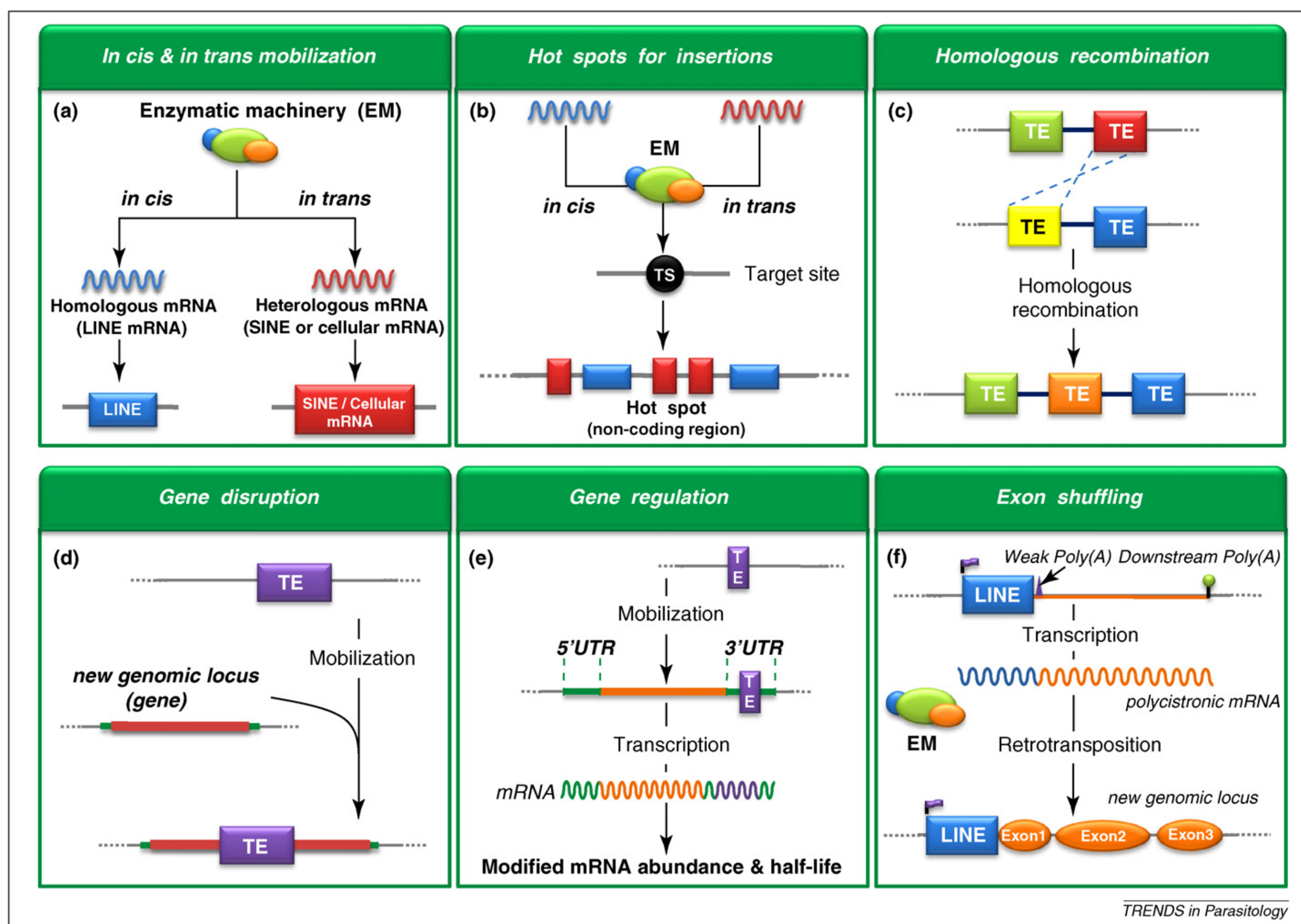

Figure 3. Impact of retrotransposons on parasite genomes. **(a)** Mobilization of the autonomous retrotransposons (in *cis*) and of the non-autonomous retrotransposons or cellular mRNAs (in *trans*) by the LINE protein machinery. **(b)** LINE enzymatic machinery mobilizes the messenger RNAs in *cis* and *trans* and inserts the reverse-transcribed copies into non-coding sequences. **(d)** TEs are substrates for recombination events and generate genomic rearrangements. **(d)** TEs are inserted inside a coding sequence at the gene 3′-untranslated regions, and **(e)** modify gene expression patterns. **(f)** Weak poly(A) signals of LINEs are bypassed during transcription, generating polycistronic transcripts that are mobilized together with the element.

acting as telomerases. A large body of data suggests that telomerases are likely to have originated from the retrotransposon reverse transcriptase [62].

TEs can also contribute to genome plasticity because they can produce large-scale chromosomal rearrangements (Figure 3c). Sequence alignments of different chromosomes of *T. brucei*, *T. cruzi* and *Leishmania major* parasites indicate that all the identified non-LTR retrotransposons or retrotransposon-like sequences tend to be located in areas of chromosome inversions, strand-switch regions and chromosome ends [59]. This was taken as an indication of the involvement of retrotransposon elements in chromosomal arrangements [59].

TEs can also play a regulatory role in gene expression patterns because they are frequently inserted inside a coding sequence or within the 3′-untranslated regions (UTR) of genes (Figure 3d,e). In three *Plasmodium* species (*P. falciparum*, *P. yoelii yoelii* and *P. vivax*) several host ORFs containing TE domains have been identified [63]. One of the ORFs found in *P. y. yoelii* with similarity to a phage integrase domain is a putative *yir*3 protein. Because the *P. y. yoelii yir*3 family is analogous to the *rif*, *stevor* and *var* superfamilies in *P. falciparum* [63], it could play an important role in antigen switching that generates the antigenic diversity of the parasite schizont stage, and that allows the parasite to evade the host immune response [64]. Thus, domestication of TEs could be advantageous to the malaria parasite in terms of being able to evade the immune system of the human host. Another example of a domesticated transposable element in *L. major* that regulates gene expression is the *SIDER2* element. This element is typically located in the 3′-UTR of mRNAs, and *LmSIDER2*-containing mRNAs are generally expressed at lower levels and have shorter half-lives than non-*SIDER2*-bearing transcripts [65]. Similarly, the insertion of the *SIRE* non-autonomous element [66], a truncated version of *VIPER*, at the 3′-UTR of some H2A gene units of several strains from *T. cruzi* [67] generates a longer mRNA that is less stable and has a shorter half-life than copies of H2A that lack the insertion [67,68]. In addition, TEs can generate variability in the composition of genomes because they have ‘weak’ poly(A) signals that can be bypassed during TE transcription (Figure 3f). In this case, polycistronic transcripts are generated that can be reverse transcribed and mobilized together with the TE. Together these data indicate that TEs can provide parasites with tools for the amplification of their genomes and produce a wide diversity of potential functional genes. It has been described that a large number of transposon-derived genes known to date have been recruited by the host to function as transcriptional regulators (61).

Role of mobile elements in the evolution of parasites' host genomes

It has been suggested that mobile elements and their hosts have coevolved in a way that balances the proliferation of TEs against the survival of the host parasite. How, why and when TE mobilization events take place are still open questions. Methylation of LINE promoters and the expression of specific small interfering RNAs are common processes involved in the expression control of many types of retrotransposons from a variety of organisms. Control of LINE mobilization based on DNA methylation has been described in *E. histolytica* [69]; also, an EhMLBP protein binds to the methylated form of a DNA encoding the RT of a LINE with higher affinity than to the non-methylated LINE RT gene [70]. Moreover, several cellular mechanisms inactivate retrotransposon mobility at different levels of the transposition process. The role of RNAi in silencing retrotransposon transcripts was suggested when 24- to 26-nucleotide fragments homologous to the ubiquitous retrotransposon *ingi* and the site-specific retrotransposon *SLACS* were detected in the *T. brucei* genome [28]. Because *SLACS* transcripts are more stable and more abundant in RNAi-deficient *T. brucei* parasites [28,71] it is likely that RNAi pathways regulate retrotransposon gene expression at transcriptional and/or post-transcriptional levels. These retrotransposon control mechanisms do not seem to play a similar role in other organisms because, for example, the *T. cruzi* parasite does not have RNAi machinery [33]. Many other parasites such as *Plasmodium* and most *Leishmania* species lack the enzymes required for RNAi-based ablation of gene expression [72]. In *Plasmodium* there might be no selective advantage in either retaining or gaining RNAi machinery because it apparently does not possess active retrotransposons or viral pathogens [72]. In *E. histolytica* there is a gene-silencing mechanism at the transcriptional level that depends on the presence of a truncated *EhSINE1* element. This type of gene silencing does not involve DNA methylation or RNAi. Remarkably, the silenced gene corresponds to the *amoebapore A* virulence gene [73].

TEs are essential drivers of genome evolution and can play an important role in speciation [74,75]. In addition, the analysis of TEs can be used as a tool to carry out phylogenetic analyses. Although the *in silico* analysis of parasite genomes has identified several TEs, sequence comparison analyses to known elements is frequently a difficult task due to a high degree of sequence divergence among TEs. In some parasites, elements classified as ‘degenerated’ based on their high degree of sequence divergence have been described. In *Leishmania* species, LINE-counterpart degenerate elements called *DIREs* have been identified [44], as well as their truncated versions known as *SIDERs* [47,48,65]. Identification of *SIDERs* and *DIREs* was possible because they bear at their 5′-end a highly conserved 79 bp sequence found in *ingi-L1Tc* related elements [65]. In the *T. cruzi L1Tc* element this sequence, called Pr77, is located at its 5′-end [76]. The Pr77 sequence is an RNA polymerase II promoter that generates abundant transcripts that are also efficiently translated [76]. In *L. major*, the *LmSIDER* sequences are located within the 3′ UTR of specific genes and downregulate mRNA steady-state levels [65].

The acquisition of gene fragments during evolution could be beneficial to the genome of the recipient organism, and lead to adaptive selection and maintenance. In fact, parasites often inhabit their hosts for long periods of time, sometimes intracellularly. The hookworm *Ancylostoma caninum,* a parasite of dogs that is frequently detected in the human small intestine, has a 1.3 kb element, *bandit*, that contains a single ORF encoding a transposase

(Figure 1) [4]. Interestingly, the closest phylogenetic relative of *bandit* is the human *Hsmar* DNA transposon, a fact that has been interpreted as suggesting that *bandit* had been transferred horizontally between the hookworm parasite and its mammalian host [4].

The existence of a viral sequence forming part of a non-LTR retrotransposon has also been described in some parasites that infect mammals. The viral 2A self-cleaving sequence, a 17 amino-acid motif present in some small viruses [77], is found at the N-terminus of the *L1Tc* LINE from *T. cruzi*. This sequence is functional *in vitro* and *in vivo* and, as in viruses, controls translation and influences the composition and abundance of the proteins that compose the mobilization machinery of the element [77]. The 2A consensus domain is not restricted to *T. cruzi* and is also present at the N-terminus of LINE-like elements from trypanosomatids such as *T. congolense*, *T. b. gambiense* and *T. vivax* [77]. The wide distribution of the 2A self-processing consensus sequence could indicate that its acquisition occurred in a common ancestor before species divergence, and thus could represent an example of a useful function that has been maintained during evolution.

Another example of LINE gain-of-function is the acquisition of an endonuclease (EN) with target specificity (Table 1). It has been postulated this is one of the factors leading to the successful amplification of the elements – moderate restriction of insertion targets provides a better chance for LINEs to be tolerated by the host and thus to proliferate within the host genome [78]. The ability of the *T. cruzi L1Tc* AP-endonuclease to repair *in vitro* and *in vivo* chromosomal DNA breaks, as well as its capacity to complement *Escherichia coli* null mutants deficient in both exonuclease III and endonuclease IV [50,51], suggests that the *L1Tc* retrotransposon could have a potential role in DNA repair [79].

It seems that there is no systematic difference in overall TE content between the genomes of parasites and those of corresponding non-parasitic organisms. Nevertheless, whereas the mammalian parasites *E. histolytica* and *E. dispar* are almost devoid of DNA transposons, the genomes of free-living *E. moshkovskii* and the reptilian parasite *E. invadens* contain an abundant number of these elements, although both *E. invadens* and *E. moshkovskii* contain very few LINE elements [6]. It has been shown, moreover, that the DNA transposons in the *E. moshkovskii* and *E. invadens* genomes are greatly diversified [6] and belong to different families of TEs. The presence of these elements in their genomes has been considered as a sign of recent transpositional activity. This seems to be the case for the *Tvmar1* element that is found in the genome of the protozoan parasite *Trichomonas vaginalis* [8]. The absence of *Tvmar1* in other trichomonads and its presence in a wide range of *T. vaginalis* isolates supports this suggestion. Recent studies regarding the content of *S. mansoni* families and *Schistosoma japonicum* LINEs indicate that there have been recent bursts of transposition in *S. mansoni* that were not paralleled in their *S. japonicum* counterparts [75]. These transposition bursts could be a consequence of the evolutionary pressure resulting from migration of *Schistosoma* from Asia to Africa and establishment in a new environment, helping both speciation and adaptation. Although it is possible that TEs might play a role in parasite pathogenicity and virulence, few data address this possibility. The fact that the content and structure of mobile elements in the non-pathogenic *E. dispar* is very similar to that in the pathogen *E. histolytica* [6,69,80] could indicate that TEs do not contribute to the pathogenicity of these parasites.

In summary, we suggest that because retroelements not only can cause non-selective deleterious effects but can also make a positive contribution to the evolution of the host genome, the renaming of 'molecular parasites' as 'potentially useful domesticated elements' could be justified.

Concluding remarks and future questions

As with the majority of organisms, the genomes of parasites that infect mammals are colonized by TEs of different types termed DNA transposons (class II) and retrotransposons (class I). The TE contents of the genomes of both parasitic and free-living organisms does not differ significantly. Most TEs present in parasite genomes have been identified *in silico* and, despite having high sequence variability, they conserve the functional domains that make them putative active mobile elements. Retrotransposons are often inserted into coding sequences or into the 3′-UTRs of genes where they can play a regulatory role in gene expression. Non-LTR retrotransposons (autonomous and non-autonomous) tend to be clustered in non-coding regions of the genomes of parasites. The distribution of these elements could be a consequence of the existence of a strong negative selection pressure against insertion in coding regions or a strategy for mobilizing non-autonomous elements.

It is believed that many TE-derived coding sequences have been 'domesticated' and have evolved to fulfill essential functions in genome dynamics. There is also evidence that some parasite TEs have acquired particular functions to increase their status as autonomous elements. Furthermore, some of the data reported to date suggest that recent TE transposition activity has occurred in parasites, probably as a consequence of an evolutionary pressure for environmental adaptation. Because TEs and their host genomes have coevolved, and continue to coevolve, it would appear that genomes should not be considered as static information systems.

Analysis of the content, structure and functional implications of TEs in parasite genomes will reveal both the recent and deep-rooted contributions of TEs to the complexity of genomes and will help in carrying out phylogenetic analyses. These studies will potentially elucidate the role TEs play in pathogenicity and virulence of parasites, with particular relevance in protozoa.

The most ambitious question to be resolved is to understand how, when and why TEs are mobilized. In this context, the characterization at the molecular level of the function and properties of the proteins that the elements encode will help to understand the mobilization mechanisms of each type of element. A key issue to be addressed is the need to develop *in vitro* retrotransposition systems to permit analysis of the molecular mechanisms of TE mobilization and to identify essential sequences and cofactor requirements. A proper understanding of the

mobilization mechanisms of parasite TEs will most likely facilitate their use as tools for gene transfer and/or gene silencing that could lead to new strategies for parasite control.

Acknowledgements
This work was supported by Grants BFU2007-64999 and BFU2007-65095) from the Plan Nacional I+D+Il of the Ministerio de Ciencia e Innovación (MICINN); P05-CVI-01227 from the Plan Andaluz de Investigación (PAI; Junta de Andalucía) and RD06/0021/0014 from the Instituto de Salud Carlos III – Redes Temáticas de Investigación Cooperativa (ISCIII–RETIC), Spain. F.M. was supported by a PAI Predoctoral Fellowship (Grant P05-CVI-01227; Junta de Andalucía).

References

1 Wicker, T. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982
2 Siefert, J.L. (2009) Defining the mobilome. In *Horizontal Gene Transfer: Genome in Flux* (Vol. 532), (Gogarten, M.N. *et al.*, eds) Chapter 2, pp 13–27, Humana Press
3 Bao, W. *et al.* (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. *Mol. Biol. Evol.* 26, 983–993
4 Laha, T. *et al.* (2007) The bandit, a new DNA transposon from a hookworm-possible horizontal genetic transfer between host and parasite. *PLoS Negl. Trop. Dis.* 1, e35
5 Wiley, L.J. *et al.* (1997) *mle-1*, a mariner-like transposable element in the nematode *Trichostrongylus colubriformis*. *Gene* 188, 235–237
6 Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368
7 DeMarco, R. *et al.* (2006) SmTRC1, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily. *BMC Evol. Biol.* 6, 89–102
8 Silva, J.C. *et al.* (2005) A potentially functional mariner transposable element in the protist *Trichomonas vaginalis*. *Mol. Biol. Evol.* 22, 126–134
9 Carlton, J.M. *et al.* (2007) Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science* 315, 207–212
10 Lopez F.R. *et al.* (2009) The protist *Trichomonas vaginalis* harbors multiple lineales of transcriptionally active *Mutator-like* elements. *BMC Genomics* 10:30, 1–16
11 Kapitonov, V.V. and Jurka, J. (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* 103, 4540–4545
12 Lorenzi, H. *et al.* (2008) Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. *BMC Genomics* 9, 595–610
13 Felder, H. *et al.* (1994) *Tas*, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. *Gene* 149, 219–225
14 Copeland, C.S. *et al.* (2003) *Boudicca*, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. *J. Virol.* 77, 6153–6166
15 Eickbush, T.H. and Jamburuthugoda, V.K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* 134, 221–234
16 Brindley, P.J. *et al.* (2003) Mobile genetic elements colonizing the genomes of metazoan parasites. *Trends Parasitol.* 19, 79–87
17 Copeland, C.S. *et al.* (2005) The *Sinbad* retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*, and the distribution of related *Pao*-like Elements. *BMC Evol. Biol.* 5, 20
18 Laha, T. *et al.* (2004) The *fugitive* LTR retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. *Int. J. Parasitol.* 34, 1365–1375
19 DeMarco, R. *et al.* (2004) *Saci-1, -2*, and *-3* and *Perere*, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. *J. Virol.* 78, 2967–2978
20 Bae, Y.A. *et al.* (2003) Evolutionary course of *CsRn1* long-terminal-repeat retrotransposon and its heterogeneous integrations into the genome of the liver fluke, *Clonorchis sinensis*. *Korean J. Parasitol.* 41, 209–219
21 Bae, Y.A. *et al.* (2008) *PwRn1*, a novel *Ty3/gypsy*-like retrotransposon of *Paragonimus westermani*: molecular characters and its differentially preserved mobile potential according to host chromosomal polyploidy. *BMC Genomics* 9, 482
22 Goodwin, T.J. and Poulter, R.T. (2004) A new group of tyrosine recombinase-encoding retrotransposons. *Mol. Biol. Evol.* 21, 746–759
23 Piednoël, M. and Bonnivard, E. (2009) DIRS1-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms. *BMC Evol. Biol.* 9, 86
24 Lorenzi, H.A. *et al.* (2006) The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Mol. Biochem. Parasitol.* 145, 184–194
25 Laha, T. *et al.* (2000) *Sjα* elements, short interspersed element-like retroposons bearing a hammerhead ribozyme motif from the genome of the Oriental blood fluke *Schistosoma japonicum*. *Biochim. Biophys. Acta* 1492, 477–482
26 Kramerov, D.A. and Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165–221
27 Bringaud, F. *et al.* (2002) Identification of non-autonomous non-LTR retrotransposon in the genome of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 124, 73–78
28 Patrick, K.L. *et al.* (2008) Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference deficient *Trypanosoma brucei*. *Mol. Microbiol.* 67, 435–447
29 Souza, R.T. *et al.* (2007) New *Trypanosoma cruzi* repeated element that shows site specificity for insertion. *Eukaryot. Cell* 6, 1228–1238
30 Olivares, M. *et al.* (2000) Genomic clustering of the *Trypanosoma cruzi* non-long terminal *L1Tc* retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis* 21, 2973–2982
31 Torres, C. *et al.* (1999) Characterization of a new ATP-binding cassette transporter in *Trypanosoma cruzi* associated to a *L1Tc* retrotransposon. *Biochim. Biophys. Acta* 1489, 428–432
32 Chiurillo, M.A. *et al.* (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 100, 173–183
33 El-Sayed, N.M. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415
34 Bringaud, F. *et al.* (2006) The *Trypanosoma cruzi L1Tc* and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Mol. Biol. Evol.* 23, 411–420
35 Bringaud, F. *et al.* (2009) Trypanosomatid genomes contain several subfamilies of ingi-related retroposons. *Eukaryotic Cell* 8, 1532–1542
36 Burke, W.D. *et al.* (1995) *R4*, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acids Res.* 23, 4628–4634
37 Laha, T. *et al.* (2006) The *dingo*, non-long terminal repeat retrotransposons from the genome of the hookworm, *Ancylostoma caninum*. *Exp. Parasitol.* 113, 142–153
38 Drew, A.C. *et al.* (1999) *SR2* elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni*. *Mol. Biol. Evol.* 16, 1256–1269
39 Laha, T. *et al.* (2005) Characterization of *SR3* reveals abundance of non-LTR retrotransposons of the RTE clade in the genome of the human blood fluke, *Schistosoma mansoni*. *BMC Genomics* 6, 154
40 Burke, W.D. *et al.* (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* 19, 619–630
41 Van Dellen, K. *et al.* (2002) LINEs and SINE-like elements of the protist *Entamoeba histolytica*. *Gene* 297, 229–239
42 Bakre, A.A. *et al.* (2005) The LINEs and SINEs of *Entamoeba histolytica*: Comparative analysis and genomic distribution. *Exp. Parasitol.* 110, 207–213
43 Kimmel, B.E. *et al.* (1987) 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.* 7, 1465–1475
44 Bringaud, F. *et al.* (2008) Role of transposable elements in trypanosomatids. *Microbes Infect.* 10, 575–581
45 Martin, F. *et al.* (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (*L1Tc*) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family DNA repair enzymes. *J. Mol. Biol.* 247, 49–59
46 Bringaud, F. *et al.* (2004) The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol. Biol. Evol.* 21, 520–528
47 Smith, M. *et al.* (2009) Organization and evolution of two SIDER retroposon subfamilies and their impact on the *Leishmania* genome. *BMC Genomics* 10, 240

48 Requena, J.M. *et al.* (2008) The SIDER2 elements, interspersed repeated sequences that populate the *Leishmania* genomes, constitute subfamilies showing chromosomal proximity relationship. *BMC Genomics* 9, 263
49 Luan, D.D. *et al.* (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605
50 Olivares, M. *et al.* (1997) The open reading frame 1 of the *L1Tc* retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *J. Biol. Chem.* 272, 25224–25228
51 Olivares, M. *et al.* (1999) The *L1Tc*, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3′-phosphatase and 3′-phosphodiesterase enzymatic activities. *J. Biol. Chem.* 274, 23883–23886
52 Martín, F. *et al.* (1996) Do non-long terminal repeat retrotransposons have nuclease activity? *Trends Biochem. Sci.* 21, 283–285
53 Garcia-Perez, J.L. *et al.* (2003) Characterization of reverse transcriptase activity of the *L1Tc* retroelement from *Trypanosoma cruzi*. *Cell. Mol. Life Sci.* 60, 2692–2701
54 Bibillo, A. and Eickbush, T.H. (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates *J. Mol. Biol.* 316, 459–473
55 Olivares, M. *et al.* (2002) The non-LTR (long terminal repeat) retrotransposon *L1Tc* from *Trypanosoma cruzi* codes for a protein with RNase H activity. *J. Biol. Chem.* 277, 28025–28030
56 Malik, H.S. (2005) Ribonuclease H evolution in retrotransposable elements. *Cytogenet. Genome Res.* 110, 392–401
57 Heras, S.R. *et al.* (2005) The *L1Tc* C-terminal domain from *Trypanosoma cruzi* non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. *Mol. Cell Biol.* 25, 9209–9220
58 Heras, S.R. *et al.* (2009) Nucleic-acid-binding properties of the C2-*L1Tc* nucleic acid chaperone encoded by *L1Tc* retrotransposon. *Biochem. J.* 424, 479–490
59 Ghedin, E. *et al.* (2004) Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* 134, 183–191
60 Bringaud, F. *et al.* (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot. Cell* 1, 137–151
61 Sinzelle, L. *et al.* (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell Mol. Life Sci.* 66, 1073–1093
62 Gladyshev, E.A. and Arkhipova, I.R. (2007) Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. USA* 104, 9352–9357
63 Durand, P.M. *et al.* (2006) An analysis of mobile genetic elements in three *Plasmodium* species and their potential impact on the nucleotide composition of the *P. falciparum* genome. *BMC Genomics* 7, 282
64 Niang, M. *et al.* (2009) The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLOS Pathogens* 5, e1000307
65 Bringaud, F. *et al.* (2007) Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog.* 3, 1291–1307
66 Vázquez, M.P. *et al.* (2000) 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* 97, 2128–2133
67 Thomas, M.C. *et al.* (2000) Plasticity of the histone H2A genes in a Brazilian and six Colombian strains of *Trypanosoma cruzi*. *Acta Trop.* 75, 203–210
68 Marañón, C. *et al.* (2000) The stability and maturation of the H2A histone mRNAs from *Trypanosoma cruzi* are implicated in their post-transcriptional regulation. *Biochim. Biophys. Acta* 1490, 1–10
69 Harony, H. *et al.* (2006) DNA methylation and targeting of LINE retrotransposons in *Entamoeba histolytica* and *Entamoeba invadens*. *Mol. Biochem. Parasitol.* 147, 55–63
70 Lavi, T. *et al.* (2006) Sensing DNA methylation in the protozoan parasite *Entamoeba histolytica*. *Mol. Microbiol.* 62, 1373–1386
71 Shi, H. *et al.* (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retrotransposon transcript abundance. *Mol. Cell Biol.* 24, 420–427
72 Baum, J. *et al.* (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res.* 37, 3788–3798
73 Huguenin, M. *et al.* (2010) Epigenetic transcriptional gene silencing in *Entamoeba histolytica*: insight into histone and chromatin modifications. *Parasitology* 37, 619–627
74 Böhne, A. *et al.* (2008) Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* 16, 203–215
75 Venancio, T.M. *et al.* (2010) Burst of transposition from non-long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. *Int. J. Parasitol* 40, 743–749
76 Heras, S.R. *et al.* (2007) The *L1Tc* non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res.* 35, 2199–2214
77 Heras, R.S. *et al.* (2006) *L1Tc* non-LTR retrotransposons contain a functional viral self-cleaving 2A sequence in frame with the active proteins they encode. *Cell. Mol. Life Sci.* 63, 1449–1460
78 Ichiyanagi, K. *et al.* (2007) Acquisition of endonuclease specificity during evolution of L1 retrotransposon. *Mol. Biol. Evol.* 24, 2009–2015
79 Olivares, M. *et al.* (2003) The endonuclease N*L1Tc* encoded by the LINE *L1Tc* from *Trypanosoma cruzi* shows repair activity *in vivo* on damaged DNA. *Biochim. Biophys. Acta* 1626, 25–32
80 Pritham, E.J. (2009) Transposable elements and factors influencing their success in eukaryotes. *J. Hered.* 100, 648–655