Molecular & Biochemical Parasitology 179 (2011) 30–36

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

TSIDER1, a short and non-autonomous Salivarian trypanosome-specific retroposon related to the ingi6 subclade

Frédéric Bringaud[a,*], Matthew Berriman[b], Christiane Hertz-Fowler[b,1]

[a] Centre de Résonance Magnétique des Systèmes Biologiques (RMSB), UMR 5536, Université Bordeaux Segalen, CNRS, 146 rue Léo Saignat, 33076 Bordeaux, France
[b] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

ARTICLE INFO

Article history:
Received 28 March 2011
Received in revised form 23 May 2011
Accepted 25 May 2011
Available online 1 June 2011

Keywords:
African trypanosomes
Ingi
Retroposon
Non-LTR retrotransposon
Non-autonomous
SIDER

ABSTRACT

Retroposons of the ingi clade are the most abundant transposable elements identified in the trypanosomatid genomes. Some are long autonomous elements (ingi, L1Tc) while others, such as RIME and NARTc, are short non-coding elements that parasitize the retrotransposition machinery of the active autonomous ones for their own mobilization. Here, we identified a new family of short non-autonomous retroposons of the ingi clade, called TSIDER1, which are present in the genome of Salivarian (African) trypanosomes, *Trypanosoma brucei*, *T. congolense* and *T. vivax*, but absent in the *T. cruzi* and *Leishmania* spp. genomes and, as such, TSIDER1 is the only retroposon subfamily conserved at the nucleotide level between African trypanosome species. We identified three TvSIDER1 families within the genome of *T. vivax* and the high level of sequence conservation within the TvSIDER1a and TvSIDER1b groups suggests that they are still active. We propose that TvSIDER1a/b elements are using the Tvingi retrotransposition machinery, as they are preceded by the same conserved pattern characteristic of the ingi6 subclade, which corresponds to the retroposon-encoded endonuclease binding site. In contrast, TcoSIDER1, TbSIDER1 and TvSIDER1c are too divergent to be considered as active retroposons. The relatively low number of SIDER elements identified in the *T. congolense* (70 copies), *T. vivax* (32 copies) and *T. brucei* (22 copies) genomes confirms that trypanosomes have not expanded short transposable elements, which is in contrast to *Leishmania* spp. (~2000 copies), where SIDER play a role in the regulation of gene expression.

1. Introduction

The trypanosomatid family includes some of the most important protist parasites of humans in the genera *Leishmania* and *Trypanosoma*, as well as other species parasitic in a wide variety of vertebrates, invertebrates, ciliates, and plants [1]. Two well-defined major groups have been identified within the genus *Trypanosoma*: (i) the section Stercoraria (also called South-American trypanosomes), including the causative agent of Chagas' disease (*T. cruzi*), and (ii) the section Salivaria (also called African trypanosomes), including the causative agents of African sleeping sickness (*T. brucei gambiense and T. b. rhodesiense*), along with other African parasites of mammals, such as *T. b. brucei* (*T. brucei*), *T. vivax* and *T. congolense*. The genome of six trypanosomatids has been sequenced and published to date, *i.e. T. brucei* [2], *T. b. gambiense* [3], *T. cruzi* [4], *L. major* [5], *L. infantum* [6] and *L. braziliansis* [6]. In addition, a number of ongoing trypanosomatid genome projects are available on line (http://tritrypdb.org/tritrypdb/showXmlDataContent.do?name=XmlQuestions.DataSources), such as two other African trypanosomes *T. congolense* and *T. vivax*. All these aforementioned genomes contain active and/or traces of inactive transposable elements (TE) (for reviews see: [7,8]).

Among the three main classes of eukaryotic transposable elements (TE: retroelements, DNA transposons and Miniature Inverted-repeat Transposable Elements – MITE), only retroelements have been described so far in the trypanosomatid genomes (2–5% of the nuclear genome). Retroelements transpose *via* reverse transcription of an RNA intermediate and are further divided into LTR retrotransposons with long terminal repeats (LTR) and the non-LTR retrotransposons, also called retroposons. Trypanosomatids contain LTR retrotransposons (VIPER), site-specific retroposons (SLACS/CZAR) and retroposons of the ingi clade (ingi/L1Tc) showing a relative site specificity for insertion [7,8]. Retroposons of the ingi clade, which will be further considered herein, contain two categories of active elements (Fig. 1 and Table S1): (i) the long (4736–5419 bp) and autonomous elements originally

Abbreviations: SIDER, Short Interspersed DEgenerate Retroposons; DIRE, Degenerate Ingi/L1Tc-Related Element.

* Corresponding author. Tel.: +33 5 57 57 46 32; fax: +33 5 57 57 45 56.
E-mail address: bringaud@rmsb.u-bordeaux2.fr (F. Bringaud).
URL: http://www.rmsb.u-bordeaux2.fr/iMET/Bringaud_Lab/Bringaud_Lab.html (F. Bringaud).
[1] Present address: Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

0166-6851/
doi:10.1016/j.molbiopara.2011.05.007

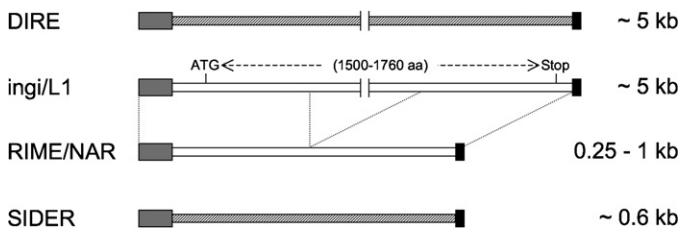

Fig. 1. Schematic representation of ingi-related retroposons identified in the trypanosomatid genomes. Potential active retroposons (ingi/L1 and RIME/NAR) and degenerate retroposons (DIRE and SIDER) are represented by white and hatched/grey bars, respectively. The conserved "76–79 bp signatures" and the poly(dA) tails are showed by dark grey and black boxes, respectively. The approximate size of these mobile elements is indicated in the right margin. The long and autonomous ingi/L1 elements are the only retroposons coding for a protein (from 1500 to 1760 aa long) responsible for retrotransposition of themselves or other ingi-related retroposons.

characterized in *T. brucei* (Tbingi) [9,10] and *T. cruzi* (L1Tc) [11], and subsequently described in *T. vivax* (Tvingi) [12] and *T. congolense* (Tcoingi and L1Tco) [12] and (ii) the short (260–1030 bp) and non-autonomous elements identified in *T. brucei* (TbRIME) [13], *T. cruzi* (NARTc) [14] and *T. vivax* (TvRIME) [12]. The short elements (TbRIME, TvRIME and NARTc) are truncated versions, which are mobilized by the retrotransposition machinery of the corresponding long elements (Tbingi, Tvingi and L1Tc, respectively) [12,15,16]. Consequently, the Tbingi/TbRIME, Tvingi/TvRIME and L1Tc/NARTc associations are considered as pairs of retroposons akin to the human LINE1/Alu, the eel UnaL2/UnaSINE1 and the plant LINE/S1 pairs [17–20].

In addition, the trypanosomatid genomes contain two kinds of degenerate vestigial ingi-related retroposons (Fig. 1 and Table S1): (i) the long elements (~5000 bp) detected in all the trypanosomatid genomes (DIRE, *D*egenerate *I*ngi/L1Tc-*R*elated *E*lement) [21] and (ii) the small elements (~600 bp), called SIDER (*S*hort *I*nterspersed *DE*generate *R*etroposon), in *T. brucei* (TbSIDER) and all the *Leishmania* spp. analyzed so far (LmSIDER, LbSIDER, etc.) [22,23]. Interestingly, SIDER retroposons were extensively expanded in the *Leishmania* genomes (~2000 copies per haploid genomes), where they have been co-opted into a role regulating gene expression [22,24–27]. In contrast, expansion of the SIDER elements has not occurred in the *T. brucei* genome, in which only ~20 copies have been detected [22]. Here we analyzed SIDER elements in the *T. congolense* and *T. vivax* genomes and report a Salivarian-specific group (TSIDER1) belonging to the ingi6 subclade.

2. Materials and methods

2.1. Detection of SIDER sequences

BLASTN was used to detect short ingi-related sequences in the *T. congolense* (strain IL3000, version 1 genome release) and *T. vivax* (strain Y486, version 2 genome release) datasets. The *T. congolense* genome assembly consisted of 3181 contigs, totaling 41.8 Mb, while the *T. vivax* assembly contained 10,250 contigs, totaling 47.4 Mb, including 8279 *T. vivax* contigs not assigned to chromosomes [12]. To identify all SIDER sequences, genome contigs were first probed with the 5′-terminal "76–79 bp signature" of Tbingi (79 bp), Tvingi (76 bp), L1Tc (78 bp), LmSIDER2a (79 bp) and LmSIDER2b (77 bp), and iteratively repeated with members of each of the SIDER families that were identified.

2.2. Bioinformatics analyses

Alignments were done using the multiple alignment software CLUSTAL X [28], followed by minor manual adjustments using MacClade version 4.06 (Sinauer Associates, Inc). Using online tools available from the European Bioinformatics Institute (EBI,

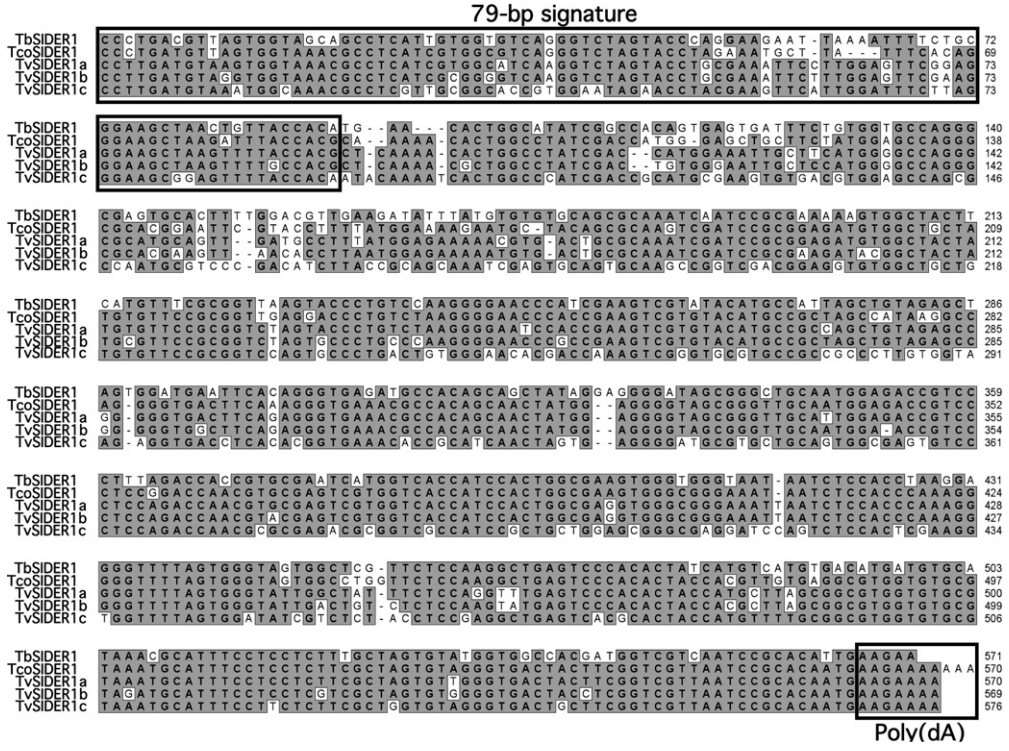

Fig. 2. Comparison of TSIDER1 consensus sequences identified in the African trypanosome genomes. The consensus sequences were generated by comparing 10 TbSIDER1 (*T. brucei*), 70 TcoSIDER1 (*T. congolense*), 7 TvSIDER1a, 4 TvSIDER1b and 21 TvSIDER1 (*T. vivax*) copies. Gaps (-) were introduced to maximize the alignments and the residues conserved among at least three sequences are shaded in grey. The "76–79 bp signature" conserved between ingi-related retroposons and the poly(dA) terminal stretch, which is a hallmark of retrotransposons, are boxed.