COMMENTARY

Genome wide occurrence and insertion preferences of INGI/RIME and SLACS CRE transposable elements in *Trypanosoma brucei*

Mohd. Faheem Khan^{a,b}, Kush Shrivastava^b, Rebeka Sinha^c, Virendra Kumar^a, and A. K. Jaitly^a

^aDepartment of Plant Sciences, M.J.P. Rohilkhand University, Bareilly, UP, India; ^bDivision of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly, UP, India; ^cDairy Cattle Breeding Division, National Dairy Research Institute, Karnal, Haryana, India

ABSTRACT

Retrotransposons play significant role in genome remodelling of *T. brucei* and about 5% of its genome consists of retrotransposons including INGI/RIME elements. INGI is one of the dispersed repetitive elements in *T. brucei* genome which is found distributed throughout all the chromosomes. SLACS (Spliced Leader Associated Conserved Sequence) however, is more conserved in its nature and lacks the typical poly-distributional pattern of LINE like transposons. We have found total 589 copies of these TEs with only 17.06 % (104 copies out of 589) copies with both ends intact thus showing a majority of truncated copies. Complete SLACS CRE were found only on chromosome 9, whereas, complete INGI/RIME were more ubiquitously distributed. The 50 – 70 bp upstream flanking sequence of these elements shows suitable nucleotide biophysical properties to favor transposition.

ARTICLE HISTORY

Received 14 June 2016
Revised 18 September 2016
Accepted 20 September 2016

KEYWORDS

distribution; INGI/RIME;
insertion preferences; SLACS
CRE; *T. brucei*; TEs

Introduction

Mobile genetic elements (MGE) play significant role in genome remodelling of parasitic protozoan like *Trypanosoma brucei* and other species. The genomes of these organisms consist of abundance of these elements and our current understanding about distribution of MGE in *T. brucei* genome emphasizes their need for maintenance of parasitaemia and in development of antigenic variation by the parasite.¹ The genome of *T. brucei* consists of 11 pairs of large chromosomes, 3–5 intermediate chromosome and around 100 mini-chromosomes.² Besides this about 5 % of its genome constitutes of retrotransposons *i.e.*, INGI/RIME elements.³ Retrotransposons are type of elements that transpose via an RNA intermediate⁴ and broadly, they can be grouped as LTR (long-terminal repeats) and non-LTRs (Non Long terminal repeats). INGI/RIME and SLACS CRE are mammalian LINEs (Long interspersed nuclear elements) like elements, falling in non-LTR category, found in *Trypanosoma* genome, having A-rich or oligo(dA) tails at their 3' end, flanked by short insertion loci duplications and encode for reverse transcriptases

which may be involved in their transposition.^{3–6}

Active transposition allow these elements to remodel the host genomes by generating various new arrangements and combinations, thus creating, destructing and shuffling the already interspersed genes as well as the elements in between which may further modulate the patterns of gene expression.^{3,5} INGI is one of the dispersed repetitive element in *T. brucei* genome which is characteristically a retrotransposon found distributed throughout in all the chromosomes,⁶ SLACS (Spliced Leader Associated Conserved Sequence) however, is more conserved in its nature and lacks the typical poly-distributional pattern of LINE like transposons.⁷ However, looking into their ubiquitous nature and abundance in the genome they are important candidates responsible for diverse set of genomic changes in the organism.⁸

Recent studies have shown that distribution of transposable elements (TEs), in particular INGI/RIME, in *T. brucei* genome is not entirely random and there are genomic hotspots that are preferred for transposition activity. Identification of these preferential insertion sites may help in prediction of

transposition events and their effect on evolution of genome.⁹⁻¹¹

Distribution and insertion preference of INGI/RIME and SLACS CRE in *T. brucei* genome

INGI/RIME and SLACS CRE are the most profoundly found TEs in *T. brucei* genome.⁷ INGI/RIME elements are found throughout the genome¹² however SLACS CRE are found only in spliced leader RNA genes.⁷ INGI elements are of 5.25 kb size and are autonomous LINE elements whereas RIMEs are non autonomous and shorter in length (0.5 kb). A complete 5.25 kb

INGI element consists of a 4.7 kb fragment bordered by 2 separate halves of RIME A and RIME B at 5' and 3' ends, respectively.^{10,13} Hence, INGI and SLACS CRE can be classified as complete or incomplete based on the completeness on their 5' or 3' end.¹³ We found total 589 copies of INGI and SLACS CRE retroelements in the genome of *T. brucei*.¹³ Most of the inserted copies of these transposable elements (TEs) belong to INGI group, however SLACS CRE TEs were observed in relatively low numbers. Distribution of these elements is affected by status of their corresponding 5' and 3' ends, whether both ends are truncated, only 3' or 5' end is truncated or both ends are

Figure 1. Genome wide status of INGI/RIME and SLACS CRE elements in *T. brucei* genome. (In figure: The outer circumference on the right represents different transposable elements, on the left it represents different chromosomes; the colored ribbons represents the number and percentage of each type of element and their distribution across different chromosomes; the more the number of elements the more thicker the ribbon is and is connected to its respective position on the designated chromosome).

intact. It was observed that within chromosomes a major portion was occupied by both ends truncated type of elements (70 % or more).

Out of total copies (104) of both TEs only 17.06 % were intact rest were either both end or one end truncated. Chromosome eleven had the maximum number of copies (174) of total TEs and complete SLACS CRE were exclusively found on chromosome 9, whereas complete INGI/RIME elements were found distributed throughout all the 11 chromosomes.¹³ Based on the

data of Khan et al.¹³ a complete map of INGI and SLACS CRE transposable elements of *T. brucei* genome (TREU927, strain927/4GUT at 10.1; last updated on 2014/08/07) was constructed using Circos, which depicted the genome wide distribution and status of their ends (either truncated or intact) (Fig. 1).

Beside the distribution of INGI & SLACS CRE in the *T. brucei* genome, it is not well known that how these elements target specific loci and what are their unique preferences to integrate within the genome. To

Figure 2. Thermodynamic and architectural signals of INGI/RIME and SLACS CRE TEs upstream flanking sequence. (In figure: X axis represents the base pair of upstream flanking sequence; Y axis represents the value of different plotted parameters).

integrate itself at specific position TEs require precise conformational mobility in terms of architectural and thermodynamic signals. The nucleotide composition, nucleosomal positioning, bending stiffness, stabilizing energy and propeller twist are the important parameters which are responsible for the transposition of TEs.^{11,14,15,16,17} During insertion process, the TEs causes DNA to unwind and nick at the target site, in addition it needs the supportive action of number of proteins such as reverse transcriptase and endonuclease for the breaking of bonds and insertion of transposable element.¹⁰ As SLACS CRE elements are site specific, more attention is paid for distribution pattern and insertion preference of INGI/RIME elements. A large multigene family, called RHS was characterized to be the insertion hotspot of INGI/RIME elements.⁹ We have analyzed 5' upstream 100 bp flanking region of these elements to find their possible insertion hotspots. The 50 – 70 bp region of this upstream flanking segment showed high A+T content, sudden multiple peaks of stabilizing energy and low bending stiffness and propeller twist (Fig. 2). Local alignment of this 100 bp upstream flanking region of SLACS CRE TEs showed a highly conserved region common to all copies, however a different pattern was observed in INGI elements. Using the MEME server (<http://meme-suite.org/>) we searched for different motifs, in upstream flanking sequence of INGI elements, into the whole genome of *T. brucei* genome. Total ten motifs were found at 57739 places for INGI/RIME elements with motif pattern 2 occurring in maximum copies (48 sites) in *T. brucei* genome.¹³ These motifs could form hotspots for future transposition and their further analysis may enable us to predict future transposition events.

Conclusion

Distribution and insertion preferences are 2 important aspects in studies related with mobile genetic elements. Previously, we studied rice genome for the distribution and insertion preferences of LINEs and SINEs with the help of nucleotide architectural signals in upstream flanking sequences.¹⁸ With the similar approach we studied INGI/RIME and SLACS CRE TEs in *T. brucei* genome. Previously, it was shown that upstream flanking sequences of TEs exhibit certain biophysical features that facilitate transposition.^{11,18} Similar approach was used for INGI/RIME

and SLACS CRE elements in *T. brucei* genome. The upstream flanking sequence of both the elements (between 50 to 70bp) exhibits nearly suitable nucleotide biophysical signals for successful transposition of these elements and searching for similar signals within the genome may help in identification of future transposition site(s). In addition to thermodynamic signals, sequence diversity was also studied, the upstream flanking sequence of SLACS CRE was conserved, however, INGI/RIME showed 10 different motif patterns throughout the genome which may be harbouring the appropriate biophysical properties and nucleotide content¹¹ that can assist in transposition.

Abbreviations

CRE	<i>Crithidia</i> retrotransposable element
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeats
MGE	Mobile Genetic Elements
SINE	Short Interspersed Nuclear Elements
TEs	Transposable Elements

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

References

- [1] Wickstead B, Ersfeld K, Gull K. Repetitive elements in genomes of parasitic protozoa. *Microbiol Mol Biol Rev* 2003; 67(3):360-75; PMID:12966140; <http://dx.doi.org/10.1128/MMBR.67.3.360-375.2003>
- [2] Daniels J, Gull K, Wickstead B. Cell biology of the trypanosome genome. *Microbiol Mol Biol R* 2010; 74(4):552-69; <http://dx.doi.org/10.1128/MMBR.00024-10>
- [3] Bhattacharya S, Bakre A, Bhattacharya A. Mobile genetic elements in protozoan parasites. *J Genet* 2002; 81:73-86; PMID:12532039; <http://dx.doi.org/10.1007/BF02715903>
- [4] Deininger P, Moran J, Batzer M, Kazazian H. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 2003; 13:651-8; PMID:14638329; <http://dx.doi.org/10.1016/j.gde.2003.10.013>
- [5] Kimmel B, ole-MoiYoi O, Young J. 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; PMID:3037321; <http://dx.doi.org/10.1128/MCB.7.4.1465>
- [6] Mellor J, Malim M, Gull K, Tuite M, McCready S, Dibbayawan T, Kingsman S, Kingsman A. Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature* 1985; 318:583-6; PMID:2415827; <http://dx.doi.org/10.1038/318583a0>

- [7] Aksoy S, Williams S, Chang S, Richards F. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucl Acids Res* 1990; 18:785-92; PMID:2156231; <http://dx.doi.org/10.1093/nar/18.4.785>
- [8] Simarro PP, Cecchi G, Franco JR, Paone M, Diarra A, Ruiz-Postigo JA, Fèvre EM, Mattioli RC, Jannin JG. Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis* 2012; 6(10): e1859; PMID:23145192; <http://dx.doi.org/10.1371/journal.pntd.0001859>
- [9] Bringaud F, Biteau N, Melville S, Hez S, El-Sayed N, Leech V, Berriman M, Hall N, Donelson J, Baltz T. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell* 2002; 1:137-51; PMID:12455980; <http://dx.doi.org/10.1128/EC.1.1.137-151.2002>
- [10] Bringaud F, Biteau N, Zuiderwijk E, Berriman M, El-Sayed NM, Ghedin E, Melville SE, Hall N, Baltz T. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. *Mol Biol Evol* 2003; 21(3):520-8; PMID:14694076; <http://dx.doi.org/10.1093/molbev/msh045>
- [11] Rawal K, Ramaswamy R. Genome-wide analysis of mobile genetic element insertion sites. *Nucleic Acids Res* 2011; 39:6864-78; PMID:21609951; <http://dx.doi.org/10.1093/nar/gkr337>
- [12] Murphy NB, Pays A, Tebabi P, Coquelet H, Guyaux M, Steinert M, Pays E. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *J Mol Biol* 1987; 195:855-71; PMID:3656436; [http://dx.doi.org/10.1016/0022-2836\(87\)90490-6](http://dx.doi.org/10.1016/0022-2836(87)90490-6)
- [13] Khan MF, Shrivastava K, Sinha R, Kumar V, Jaitly AK. Dataset for distribution of INGI/RIME and SLACS CRE transposable elements in *Trypanosoma brucei* genome. *Data in Brief* 2015; 5:818-21; <http://dx.doi.org/10.1016/j.dib.2015.10.040>
- [14] El Hassan MA, Calladine CR. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* 1996; 259:95-103; PMID:8648652; <http://dx.doi.org/10.1006/jmbi.1996.0304>
- [15] Brukner I, Sanchez R, Suck D, Pongor S. Sequence-dependent bending propensity of DNA as revealed by DNase I: Parameters for trinucleotides. *EMBO J* 1995; 18:1812-8
- [16] Ornstein RL, Rein R, Breen DL, MacElroy R. An Optimized potential function for calculation of nucleic-acid interaction energies I Base stacking. *Biopolymers* 1978; 17:2341-60; PMID:24624489; <http://dx.doi.org/10.1002/bip.1978.360171005>
- [17] Breslauer KJ, Frank R, Bolcker H, Marky LA. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci* 1986; 83:3746-50; PMID:3459152; <http://dx.doi.org/10.1073/pnas.83.11.3746>
- [18] Khan MF, Yadav BS, Ahmad K, Jaitly AK. Mapping and analysis of the LINE and SINE type of repetitive elements in rice. *Bioinformatics* 2011; 27(6):276-9; PMID:22355220; <http://dx.doi.org/10.1093/bioinformatics/btt276>