Proc Indian Natn Sci Acad **81** No. 2 March 2015 pp. 431-446

DOI: 10.16943/ptinsa/2015/v81i2/48097

Review Article

Circular Noncoding Ribosomal Transcripts and Recombination of Retrotransposons: Two Charming Secrets Revealed by a Less-Explored Human Parasite

SUDHA BHATTACHARYA*
School of Environmental Sciences, Jawaharlal Nehru University, New Delhi 110067

(Received on 19 May 2014; Revised on 16 October 2014; Accepted on 14 November 2014)

Parasites, through their diverse cellular metabolic pathways, help us to appreciate the different directions in which evolution works, while maintaining a unified theme. To illustrate this I describe our recent observations with two basic processes, namely regulation of ribosomal RNA transcription, and the biology of retrotransposition in *Entamoeba histolytica*, a highly prevalent protozoan parasite that causes amebiasis.

Ribosomal RNA synthesis is generally tightly regulated in response to growth rate such that cells subjected to growth stress shut down their rRNA transcription. In *E. histolytica* we observed that upon growth stress, rRNA synthesis did not shut down. Instead, unprocessed pre-rRNA accumulated to high levels along with a novel class of circular RNAs derived from the 5'-external transcribed spacer (etsRNA). The etsRNA can self circularize *in vitro*, a property not previously known in spacer RNAs. We hypothesize that circular etsRNAs would escape exonucleolytic decay and inhibit pre-rRNA processing, possibly by titrating away the processing factors which normally bind to them.

In the study on retrotransposition of non-long terminal repeat retrotransposons (EhLINEs/SINEs), we successfully mobilized EhSINE in a cell-line made retrotransposition-competent by transfection with multiple constructs to express the polypeptides required for retrotransposition, a first for any protozoan parasite. While tracking retrotransposition of a marked SINE copy we found that the newly retrotransposed SINEs had undergone high-frequency recombination, presumably due to the known ability of reverse transcriptase to perform template jumping. Such recombination has not been reported for retrotransposons, and may be important in generating sequence polymorphism.

Key Words: ***Entamoeba histolytica*; Circular Non Coding RNA; Ribosomal RNA Spacer; Ehsine; Non LTR Retrotransposon; SINE Recombination**

Introduction

Entamoeba histolytica, a protozoan parasite, is classically defined as the causative agent of the disease Amoebiasis. It resides in the human colon along with a number of nonpathogenic species of *Entamoeba*. The estimates available world-wide state that 500 million persons are infected with *Entamoeba* spp., with about 50 million cases of invasive amoebiasis per year; resulting in about 100,000 deaths (WHO/PAHO/UNESCO report, 1997). The parasite is directly transmitted through the faecal-oral route without the need of a vector system. It therefore flourishes where humans live in crowded conditions with poor sanitation. It continues to be a major public

*Author for Correspondence: E-mail: sbjnu110@gmail.com; Tel.: +91-11-26704308

health problem in India due to recurrent infections caused by the high parasite load in our environment, and lack of effective immunity in infected individuals. The morbidity caused by the intestinal form of the disease results in great economic loss, although mortality is reduced, thanks to the freely-available and over-used anti-amoebic drug, metronidazole (Upcroft and Upcroft, 2001). The parasite also invades the liver and other organs, causing abscesses which are fatal if not timely diagnosed (Choudhuri and Rangan, 2012). These patients need critical hospital support.

Entamoeba research in the past few decades has generated a number of valuable resources, including genome and transcriptome sequences **(**Loftus *et al*., 2005; Clark *et al*., 2007; Lorenzi *et al*., 2010); and technologies of gene manipulation, like stable transfection, inducible gene expression, and down regulation of genes. These, and other resources will undoubtedly help future researchers to address some of the key unanswered questions, foremost being-1) identification of parasite strains from clinical samples and correlation with severity and type of disease; 2) determination of host genetic and physiological factors that are correlated with invasive and extraintestinal disease; 3) identification of molecules and pathways involved in pathogenic processes; and 4) combating the possibility of drug resistance. In order to address these issues meaningfully, it is necessary to understand the basic biology of the parasite. Our lab has, therefore, focused its efforts on in depth analysis of selected biological processes in the parasite. Here I describe the significant observations we have made in two such processes, namely regulation of ribosomal RNA transcription, and the biology of retrotransposition. These two processes are not physiologically related, but are being reported together to inform the readers of the various ways in which *E. histolytica* uses unique mechanisms. Our data underscores the importance of studying the basic biology of parasites to help us understand and control their pathology; and also enriches our general understanding of how different life forms function.

Ribosomal RNA Transcription During Growth-Stress in *E. histolytica*

Ribosomal RNA genes are amongst the most highly repetitive and conserved genes in all cell types **(**Long and Dawid, 1980). Their transcripts- the rRNAs are components of ribosomes which drive all cellular protein synthesis. Each rDNA unit is transcribed as a long precursor (pre-rRNA) which, in addition to the sequences of the mature rRNAs (18S-, 5.8S- and 28S-rRNAs), contains intervening external and internal transcribed spacers (5'- and 3'-ETS; ITS1 and 2) (Fig. 1). The pre-rRNA gets processed to remove the mature rRNAs from the spacers by an intricate series of steps involving endonucleolytic cuts and exonucleolytic trimming. While the mature rRNAs get packaged into ribosomes, the spacers are degraded.

The transcription of rRNA genes is tightly regulated with respect to growth rate, and in the model systems studied most extensively, namely mammals and yeast, rRNA gene transcription stops in response to growth stress (McStay and Grummt, 2008; Moss, 2004; Warner, 1999). This is entirely logical as the cell would not waste resources on making ribosomal precursors at a time when protein synthesis has slowed down. Therefore it was unexpected when we found that *E. histolytica*, in fact does not shut down rRNA synthesis during growth stress. Instead unprocessed pre-rRNA accumulates in stressed cells, along with a novel class of circular RNAs derived from the 5'-ETS. We hypothesize that the accumulated circular 5'-ETS transcripts serve to titrate away the essential pre-rRNA processing factors, leading to inhibition of pre-rRNA processing during stress. A review of this work is presented below.

The rDNA Transcription Unit of *E. histolytica*

The rRNA Genes in E. histolytica are Extrachromosomal

The rRNA genes in *Entamoeba* are present exclusively extrachromosomally on circular DNAs without any chromosomal copy reported so far. The best-characterized circular rDNA in *E. histolytica*,

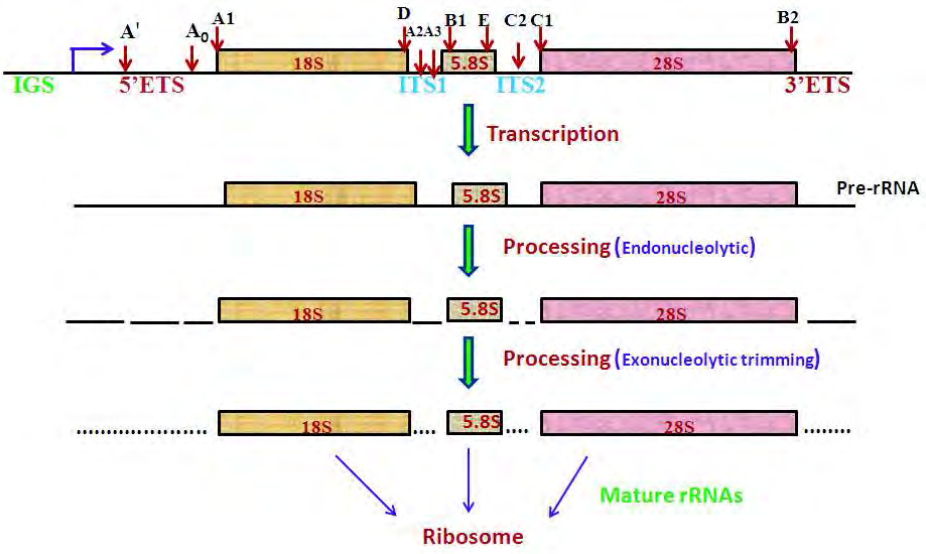
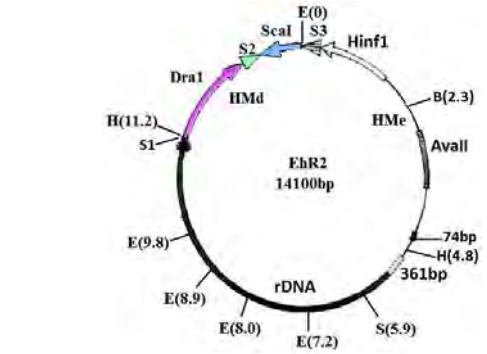


Fig. 1: General scheme of pre-rRNA synthesis and processing in eukaryotes. A single rDNA transcription unit is shown. Tandem repeats of rDNA units are separated by intergenic spacers (IGS). The transcription start point is depicted by bent arrow. The transcribed pre-rRNA consists of the coding regions (18S-, 5.8S-, and 28S rRNAs) separated by external- and internal transcribed spacers (ETS and ITS respectively). The pre-rRNA is processed endonucleolytically at the sites indicated by red arrows. Exonucleolytic trimming finally yields the mature rRNAs

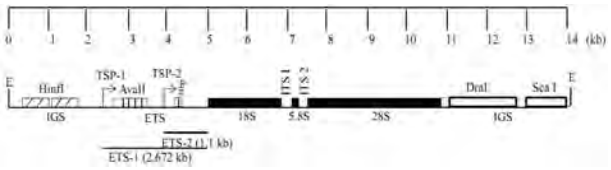
EhR1, is 24.5 kb and is present in about 200 copies per haploid genome equivalent **(**Bhattacharya *et al*., 1989; Huber *et al*., 1989; Sehgal *et al*., 1994; Bhattacharya *et al*., 1998). It contains two identical rRNA transcription units (rDNA I and rDNA II) arranged as inverted repeats (Fig. 2A). The rDNA circles (EhR2) with one rDNA transcription unit are also found in different strains of *E. histolytica* **(**Bhattacharya *et al*., 1998).

The Nucleolus in Entamoeba is at the Nuclear Periphery

The nucleolus is a well-defined compartment of the nucleus and is the site of rDNA transcription, pre-rRNA processing and modification, and pre-ribosome assembly (Venema and Tollervey, 1999). As viewed by electron microscopy the nucleolus in model organisms is nucleoplasmic and composed of three distinct regions (Koberna *et al*., 2002). Nucleolar organization in *Entamoeba* appears to be completely different as it appears to be confined to the nuclear periphery. Localization studies showed that the rDNA circles mapped to the inner membrane of the nucleus **(**Zurita *et al*., 1991). Using flouresence microscopy with antibodies against *E. histolytica* homologue of



A



B

Fig. 2: Organization of ribosomal RNA genes in *E. histolytica*. (A) The rDNA circle from *E. histolytica* (EhR2) is shown (Sehgal *et al*., 1994; Bhattacharya *et al*., 1998). The arrow marked rDNA gives the orientation of rRNA gene transcription. Positions of various short tandem repeats (*Hinf*I, *Ava*II, 74 bp, *Dra*I, and ScaI) are marked. E, *Eco*RI. (B) Schematic linear view of rDNA transcription unit and flanking sequences of EhR2 (Adapted from Gupta *et al*., 2012). Bent arrows represent the transcription start points

fibrillarin (a known nucleolar marker) and *E. histolytica* RNA Pol I subunit, it was shown that both antibodies co-localized to the nuclear periphery with very little labelling in the nucleoplasm (Jhingan *et al*., 2009).

The rRNA Genes in E. histolytica are Transcribed from Two Promoters

Studies with the rDNA transcription unit in *E. histolytica* show that it follows the general organization of regulatory sequences of rRNA genes found in most eukaryotes. For example, (i) the transcription start point (TSP) maps 1 to 2 kb upstream of the mature 18S rRNA (2.627 kb upstream of the mature 18S rRNA in rDNA I, and 1.224 kb upstream of the mature 18S rRNA in rDNA II in circles containing two rDNA units) (Michel *et al*., 1995; Panigrahi *et al*., 2009); (ii) the intergenic spacer of both rDNAs contains tandem repetitive elements, which may have regulatory roles; and (iii) the promoter is located within 100 nucleotides upstream of the TSP (Panigrahi *et al*., 2009).

Further analysis with a circle containing a single rDNA unit (equivalent of rDNA I) showed the presence of a second stronger promoter downstream to the previously mapped promoter **(**Gupta *et al*., 2012) (Fig. 2B). The TSP from this promoter (P2), mapped by primer extension, was located 1.1 kb upstream of the 5'-end of 18S rRNA. The relative role of these two promoters in rDNA transcription is not known. However it was observed that the two promoters responded differently to growth stress, as transcription from the weak promoter P1 was not visible during stress **(**Gupta *et al*., 2012).

Transcription Continues and Unprocessed pre-rRNA Accumulates During Growth Stress in *E. histolytica*

As mentioned earlier, transcription of rRNA genes is highly regulated in response to both general metabolism and growth stress. Conditions that are harmful to cellular metabolism, like nutrient starvation, environmental challenges, and specific inhibition of protein synthesis cause down regulation of rDNA transcription; and subsequent up regulation is observed upon reversal of adverse conditions **(**Zhao *et al*., 2003; Gokal *et al*., 1986). To study the mechanisms by which *E. histolytica* regulates rRNA synthesis the cells were starved of serum to arrest further growth, and the levels of pre-rRNA were measured by northern hybridization.

Contrary to expectations from the model systems, it was found that in *E. histolytica* the pre-rRNA accumulated 2.5-fold during serum starvation, and 1.5-fold when protein synthesis was inhibited by cycloheximide treatment **(**Gupta *et al*., 2012) **(**Fig. 3). Pre-rRNA accumulation was not due to increased transcription of rRNA genes. It appears that the accumulation may occur due to inhibition of processing rather than up regulation of transcription. Interestingly, in addition to the full-length pre-rRNA, we observed that serum-starved cells accumulated a heterogeneous population of RNA molecules (0.7 to 0.9 kb) corresponding to the 5'-ETS (Fig. 3). These RNAs and their possible link with pre-rRNA accumulation are described below.

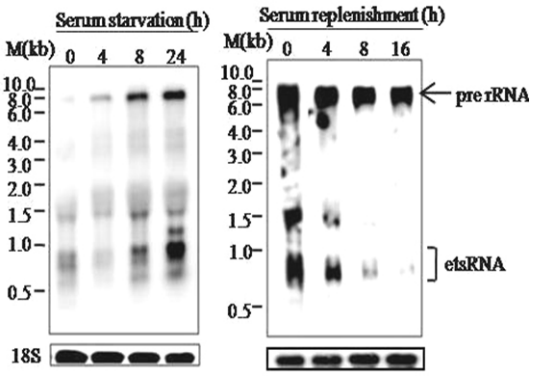


Fig. 3: Northern analysis showing the accumulation of pre-rRNA and etsRNA during serum starvation. *E. histolytica* cells growing in medium containing 15% adult bovine serum were resuspended in medium containing 0.5% serum to induce serum starvation. Serum was restored after 24 h of starvation. Cells were collected , total RNA was purified, electrophoresed, blotted and hybridized with ETS-2 probe (shown in Fig. 2B). The pre-rRNA and etsRNA bands are indicated on the right. 18S rRNA was used as loading control

The 5'-ETS RNA Accumulates as Circular Molecules Under Stress

The 5'-ETS is the first part of pre-rRNA to be transcribed, and is generally the longest spacer,

varying in length from ~ 4 kb in mouse (Bourbon *et al*., 1988) to 696 nt in yeast (Hughes and Ares, 1991). It binds to the U3 snoRNA, and serves a crucial role in initiating the correct processing of pre-rRNA and assembly of pre ribosomal particles (Borovjagin and Gerbi, 2000). After the pre-rRNA is processed, the free 5'-ETS fragment is rapidly degraded by the exosome (Houseley *et al*., 2006; Houseley and Tollervey, 2009).

We were therefore intrigued to observe that in *E. histolytica* the 1.1 kb 5'-ETS accumulated during stress as two bands of 0.7- and 0.9-kb (Gupta *et al.*, 2012). Further analysis of this RNA class, referred to as 'etsRNA' showed that it was not polyadenylated. It was transcribed in the same direction as rRNA, and it was nuclear-localized. When the endpoints were mapped, it was found that it consisted of circular RNAs. The circular nature of these RNAs was first detected by RT-PCR with out-facing primers. This was further confirmed by resistance to digestion with exonuclease T (which requires a free 3' terminus), and to nicking conditions (90ºC, with $NaHCO_3$) in which linear RNA species rapidly disappeared, but the etsRNA was more resistant (Fig. 4).Two major families of circular etsRNAs of size 766 nt and 912 nt were found. The 5'-junction of the circularization event was the same in both; at position +102G (with respect to the TSP). Although the 3'-junction was different, it was an A-residue in both cases (Fig. 5A). Circularization of etsRNAs did not seem to require protein factors, as linear transcripts obtained by *in vitro* transcription could spontaneously circularize **(**Gupta *et al*., 2012). Experiments are under way to show whether these molecules form a lariat structure.

This is the first report of autocatalytic RNA located in the 5'-ETS. Under normal growth conditions, the linear 5'-ETS is rapidly degraded. Under stress, by mechanisms which we are currently investigating, two internal fragments of this RNA undergo autocyclization. Once formed, the circles would be much more stable than the linear precursors and may serve regulatory functions, e.g. to inhibit pre-rRNA processing.

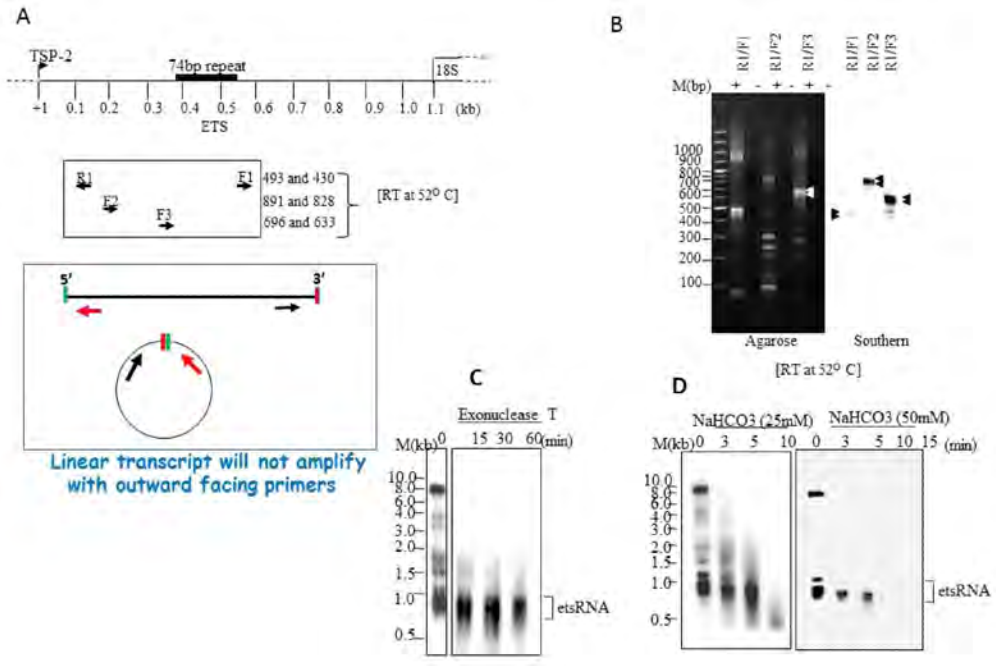


Fig. 4: The etsRNA is circular. (A) The circular nature of etsRNAs was shown by circular (c) RT-PCR using out-facing primers. The location of reverse (R) and forward (F) primers in ETS-2 is shown. The expected amplicon sizes (in bp) from each primer pair are indicated on the right. Each primer pair gave two amplicons, one of which contained a short internal deletion (see Fig. 5a). (B) The amplicons from cRT-PCR with total RNA were analyzed by Southern hybridization using ETS-2 probe. (C) The total RNA was treated with exonuclease T for the indicated times. Northern blot was hybridized with ETS-2 probe. (D) Same as (C) except that the RNA was treated with $NaHCO_3$

Circular Non Coding RNAs May Regulate Pre-rRNA Levels

Linear non coding RNAs with regulatory functions have been reported from the rDNA locus. Transcripts arising from the mouse rDNA intergenic spacer (IGS) are processed into 150–300 nt RNAs which are involved in epigenetic silencing of the rDNA locus **(**Mayer *et al*., 2008). In another study, an IGS transcript in lung cancer cells was implicated in regulating the levels of 45 S pre-rRNA (Shiao *et al*., 2009).

It is possible that the etsRNAs reported in *E. histolytica* also have regulatory roles. Their self-circularization ability would lead to their stabilization and consequent accumulation. Since the 5'-ETS region of pre-rRNA is the primary site for assembly of the pre-rRNA processing machinery, it is possible that accumulated ETS RNA in the form of circles may trap or store the processing factors in an inactive state under stress conditions when synthesis of new ribosomes is stopped (Fig. 5B). Accumulation of unprocessed pre-rRNA may permit the starving *E. histolytica* trophozoite to rapidly assemble new ribosomes from already available precursors when normal nutrient conditions are restored. It is interesting that the accumulated circular etsRNAs rapidly disappear, once serum is restored to serum-starved cells **(**Gupta *et al*., 2012), further suggesting a regulatory role for these RNAs during stress.

Why does *E. histolytica* resort to storage of unprocessed pre-rRNA rather than shutting down rRNA synthesis during stress? While further work is required to answer this question, a possible reason could be that the parasite is unable to obtain sufficiently high levels of nucleotide precursor pools to undertake rapid synthesis of rRNAs when stressed cells resume normal growth. The pools of unprocessed pre-rRNAs would therefore be required to fulfil the demand for new ribosomes in the absence of sufficiently high production of new rRNAs.

The universe of circular RNAs seems to be rapidly expanding as several important classes of these RNAs are being reported in the literature. Some of the well-known circular RNAs so far were Group I and II introns **(**Vicens and Cech, 2009; Murray *et al*., 2001), viroids and satellite RNAs **(**Hammann and Steger, 2012) and special cases of skipped exons **(**Zaphiropoulos, 1997; Burd *et al*., 2010; Capel *et al*., 1993). Due to their specialized nature they were not thought to be of general physiological relevance. However, the recently-discovered circular RNAs show that this class of RNAs could have important regulatory functions that are waiting to be discovered. One such class is the circ RNAs, which are circularized by back-splicing of exons, are cytoplasmically localized, and act as miRNA sponges (Memczak *et al*., 2013; Hansen *et al*., 2013). These RNAs are abundant and are transcribed from thousands of loci in human and mouse. It is estimated that in human ~1% of polyadenylated RNAs are circular (Jeck *et al*., 2013; Salzman *et al*., 2012). Another class of circular RNAs are ci RNAs. They are derived from circular introns, are nuclear localized, and positively regulate transcription of their parent genes (Zhang *et al*., 2013). We believe that our work has defined yet another class of circular regulatory RNA, the circular spacer (cs) RNAs which are derived from the rRNA spacer and are nuclear localized. We suggest that these RNAs may down-regulate pre-rRNA processing in *E. histolytica*.

The Biology of Retrotransposition in *E. histolytica*

Transposable elements are ubiquitously present in most genomes. They are of two major types- the DNA transposons and the retrotransposons. All retrotransposons can be divided into two major groups, i) LTR (which contain long terminal repeats), and ii) non-LTR. The LTR elements share similarity with retroviruses. Almost 35% of the human genome is occupied by non-LTR retrotransposons (Goodier and Kazazian, 2008). This is also the group of transposable elements predominantly found in *E. histolytica* **(**Sharma *et al*., 2001; Lorenzi *et al*., 2008). Hence further discussion will focus on these elements.

Although transposable elements were earlier relegated as 'junk' or 'selfish' DNA, they are now thought to be important players in genome evolution and may also influence the expression of genes involved both in differentiation and disease (Belancio

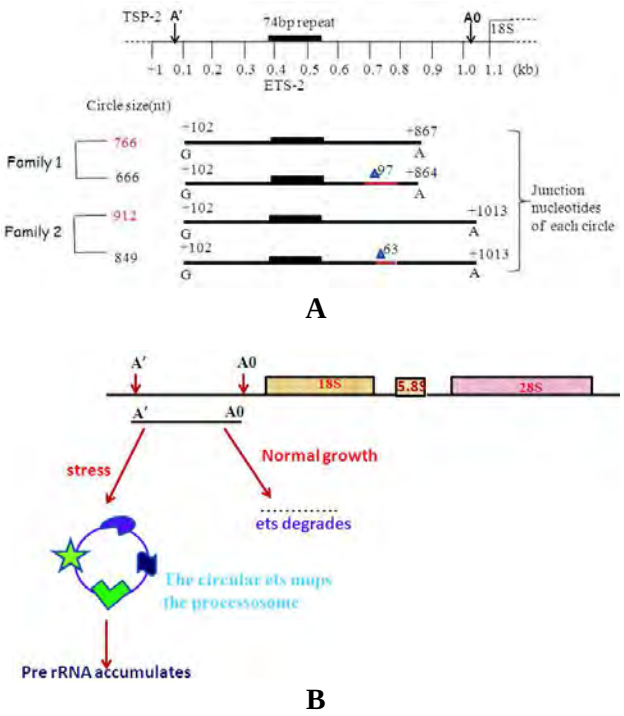


Fig. 5: Circles originate from the A'-A0 sub fragment of 5' ETS-2. (A) Top line gives schematic representation of ETS-2 with the processing sites A' and A0 indicated by arrows. The junction points of the two major families of circles found *in vivo* are shown. In each case the internal deletions leading to smaller circles are indicated. (B) Model showing the possible function of circular etsRNA. The A'-A0 subfragment contains the binding sites for assembly of the SSU processome. This linear fragment is rapidly degraded under normal growth conditions. Under stress its degradation is slow, allowing the fragment to circularize and further escape degradation. The accumulated fragment could titrate away the processome components, resulting in accumulation of unprocessed pre-rRNA

et al., 2008). The non-LTR retrotransposons include the well-studied Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs). Transposons which encode the functions required for their own transposition are called autonomous elements. Their transposition functions can sometimes be utilized by nonautonomous elements which lack the necessary coding capacity, but can transpose equally efficiently. Thus, amongst the non-LTR retrotransposons, LINEs are autonomous elements while partner SINEs retrotranspose using the LINE machinery. LINEs typically encode a reverse transriptase (RT), endonuclease (EN), and a nucleic-acid binding protein. The latter is thought to coat the RNA to be retrotransposed, converting it into a ribonucleoprotein.

Studies with non-LTR retrotransposons in model systems have revealed that these elements move by a mechanism called target primed reverse transcription in which the endonuclease encoded by the element nicks the new target site, generating a free 3'-OH which primes reverse transcription of element RNA using the element-encoded RT **(**Luan *et al*., 1993). Staggered nick in the second strand of target site, followed by copying of the first strand cDNA generates a DNA copy of the element RNA at the new site, which is accompanied by short target site duplication (TSD).

Working with these elements is challenging as most genomic copies have been inactivated during evolution by accumulation of multiple mutations and truncations. A model system has been developed to study non-LTR retrotransposon mobility in a human cell-line **(**Moran *et al*., 1996). Our current understanding of the biology of this process is very incomplete even in the human cell-line, and is almost non-existent in other systems. Given that these elements occupy a substantial fraction of the *E. histolytica* genome we undertook studies to understand the consequences of the act of retrotransposition. For this we made an *E. histolytica* cell line which expressed the element-encoded functions needed for retrotransposition. Interestingly, we found that mobilization of these elements was accompanied by high frequency of recombination between transcripts of the resident copies, resulting in sequence heterogeneity in the mobilized copies. Such recombination, which presumably results from the property of RT to undertake template switching during cDNA synthesis, has been reported for retroviruses but has never been suggested for non-LTR retrotransposons. A review of this work is presented below.

The LINEs and SINEs of *E. histolytica*

E. histolytica contains three classes of LINEs (EhLINE1, 2, and 3) and SINEs (EhSINE1, 2 and 3)

which together constitute ~11% of the genome (Van Dellen *et al*., 2002; Bakre *et al*., 2005; Lorenzi *et al*., 2008) (Fig. 6). They insert at AT-rich sites on all chromosomes, are not telomeric, and are close to protein-coding genes. By virtue of the sequence identity at the 3'-end between EhLINEs and EhSINEs (87% identity in a 73 nt stretch between EhLINE1 and EhSINE1, and 76% identity in a 84 nt stretch between EhLINE2 and EhSINE2), EhSINE1 and 2 may be considered as partner SINEs of EhLINE1 and 2 respectively **(**Cruz-Reyes *et al*., 1995; Bhattacharya *et al*., 2002; Willhoeft *et al*., 2002). EhLINE1 (4.8 kb) and EhSINE1 (550 bp) are the most abundant. EhLINEs typically encode two ORFs. The N-terminal one-third of EhLINE1 contains ORF1 which has nucleic acid-binding properties. The ORF2 contains RT domain and EN domain which resembles Type IIS restriction endonucleases.

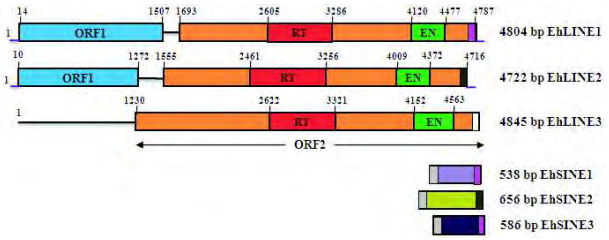


Fig. 6: Sequence organization of full-length EhLINEs and EhSINEs. Consensus sequence of each EhLINE family derived by comparative analysis of all database entries was used to mark the ORFs and other features, including the RT and EN domains in ORF2. The sizes of each consensus element are indicated. Numbers on top of each LINE family denote nucleotide positions. Regions identical between EhLINEs and EhSINEs at their 3'-ends, and between EhSINEs at their 5'-ends are shown by similar shading

Sequence Comparison of the Three LINE Families

Most copies of each element are truncated at the 5'- or 3'-end or at both ends. Of the full-length EhLINE copies none was found to contain a complete ORF, due to many point mutations (Bakre *et al*., 2005). EhLINE1 constitutes the largest family of TEs in *E. histolytica* with a total of 742 elements, including 88 complete copies and 46 putative complete elements, truncated due to their location at the end of assemblies (Lorenzi *et al*., 2008).

Although EhLINE2 and 3 are present in fewer copies than EhLINE1, their overall sequence organization is very similar, and the RT and EN functional domains are well conserved. However, the consensus sequence reconstructed for EhLINE3 does not have ORF1. This may be due to the accumulation of too many mutations in this part of the element (Bakre *et al*., 2005).

Sequence Organization of the Major *E. histolytica* Retrotransposon-EhLINE1

The consensus sequence of EhLINE1, with complete ORFs, was reconstructed manually by selecting the most common nucleotide at each position. Analysis of the consensus sequence showed that EhLINE1 had a length of 4804 bp (Fig. 6). The RT domain showed the closest match with RTs encoded by the R4 clade of non-LTR retrotransposons, most notably the R4 element of *Ascaris lumbricoides* and the Dong element of *Bombyx mori*. The EN domain had sequence features resembling Type IIS restriction endonucleases, and was very similar to the domains in R2, R4, and CRE clades of non-LTR elements (Bhattacharya *et al*., 2002; Van Dellen *et al*., 2002). The N-terminal one-third of the element encoded ORF1 which has matches with proteins containing coiled coil domains.

Properties of ORF1p and RT Domain of ORF2p

Amongst the non-LTR retrotransposons, ORF1 shows considerable sequence diversity in different organisms, but there appears to be functional conservation. ORF1p has nucleic acid-binding property (Martin, 1991), which in some elements (e.g. I and Jockey) is associated with three cysteine-histidine motifs (CCHC type) similar to that found in the gag protein of many LTR retrotransposons. Other elements contain conserved C2H2 zinc-finger and/or c-myb DNA binding motifs (Yang *et al*., 1999). The colocalization of human L1 ORF1p with L1 RNA in ribonucleoprotein particles (RNPs) indicates that the possible role of this protein is to associate with the RNA template after translation, and import the template back into the nucleus for reverse transcription.

The EhLINE1 ORF1 encodes a 498 amino acid long polypeptide of 60.5 kDa. Its sequence shows the presence of "coiled coil domain" at C-terminus while the most basic region is located at the N-terminus (pI= 10.67). Our preliminary analysis of recombinant EhLINE1 ORF1p shows that it binds ssRNA, ssDNA and dsDNA in a cooperative, non-sequence specific manner as previously reported for mouse L1 (Kolosha and Martin, 1997). The nucleic acid binding activity is located at the N-terminus, while the C-terminus appears to promote formation of multimeric complexes.

ORF2 encodes a multifunctional protein consisting of reverse transcriptase and endonuclease activities. Phylogenetic analysis demonstrated that RTs encoded by non-LTR retrotransposons represent a lineage that is distinct from the RTs encoded by LTR retrotransposons and retroviruses (Malik *et al*., 1999). The Non-LTR retrotransposon-encoded RT has both RNA- and DNA-dependent polymerase activities (Ivanov *et al*., 1991; Garcia-Perez *et al*., 2003). The latter would allow synthesis of both the complementary DNA and second strand DNA necessary to complete the element's integration starting from RNA. It is capable of adding non-template nucleotides (usually T residues), has high processivity (Luan and Eickbush, 1995), and low fidelity **(**Jamburuthugoda and Eickbush, 2011).

Our preliminary work showed that the ORF2 protein encoded by EhLINE1 (recombinant full-length ORF2p) has RT activity with various substrates, including poly (rA)-oligo(dT), and a 120 nt RNA template from the 3'-end of EhSINE1. EhLINE1-ORF2p seems to have considerably high processivity *in vitro* as full length cDNA was the major product with very few lower-sized products. A mutation in the two Asp residues in the conserved YXDD motif (YMDD to YMYY) abolished RT activity.

Properties of the EhLINE1-endonuclease

The non-LTR retrotransposons can be classified into two broad categories based on the nature of the endonucleases encoded by the elements. One class encodes the apurinic endonuclease (APE), while the other encodes a restriction enzyme-like endonuclease (EN). All the elements of the latter class (including EhLINEs) belong to the R2 group, which is considered to be of ancient origin (Malik *et al*., 1999). Many of them insert in a sequence-specific manner, presumably due to the nicking-specificity of the endonuclease.

To better understand the mode of transmission of EhLINE1 in the *E. histolytica* genome, the EN domain was cloned and expressed in *Escherichia coli* and its properties were studied with respect to target site specificity of nicking *in vitro* (Mandal *et al*., 2004). The purified protein could nick a completely unrelated substrate, supercoiled pBluescript DNA, to yield open circles and linear DNA smears on longer incubation, showing that the enzyme was, at best, loosely sequence specific. The conserved $PDX_{12–14}D$ motif was required for activity. To determine whether the enzyme exhibited any nicking hotspots in the *E. histolytica* genome, an empty target site was searched and one such site, known to be occupied by EhSINE1, was identified. When a 176-bp fragment containing this empty site was used as a substrate for EN, it was prominently nicked on the bottom strand at the precise point of insertion of EhSINE1, showing that the enzyme indeed preferentially nicked selected genomic sequences. This data also confirmed that EhSINE1 could use the EhLINE1-encoded endonuclease for its insertion. The sequence preference of the EN was determined *in vitro* with a variety of mutated substrates. It was possible to assign a consensus sequence, 5'-GCATT-3', which was efficiently nicked between A-T and T-T (Mandal *et al*., 2006).

Studies on the kinetics of EhLINE1 EN-catalyzed reaction, determined under steady-state and single turnover conditions, revealed a significant burst phase followed by a slower steady-state phase, indicating that release of product could be the slower step in this reaction (Yadav *et al*., 2009). Like restriction endonucleases, this enzyme displayed a low K_m, suggesting high affinity for DNA. It had a low turnover number that could be an evolutionary advantage to limit retrotransposition. The binding of the enzyme to DNA was accompanied by major

conformational change. These similarities with bacterial restriction endonucleases suggest that the endonuclease encoded by EhLINE1, and other related non-LTR retrotransposons could possibly be acquired from bacteria, through horizontal gene transfer.

Use of the Cell Culture System for Induction of *de novo* Retrotransposition in Cultured Cells

Although LINEs are typically present in high copy number in a large variety of genomes, it is widely seen that most LINE copies are, in fact, inactive due to the accumulation of multiple mutations. In the human genome most L1 copies (>99.8%) are inactive. About 80-100 L1s remain retrotransposition-competent (i.e. active) **(**Brouha *et al*., 2003) and a number of human diseases are caused by recent insertions of these L1s **(**Babushok and Kazazian, 2007; Belancio *et al*., 2008). In order to study *de novo* retrotransposition, episomal systems were developed in cultured mammalian cells in which a full length L1 element was expressed **(**Moran *et al*., 1996). Retrotransposition events were recovered and the flanking sequences demonstrated typical L1 structural hallmarks. This system could also retrotranspose SINEs, such as Alu and SVA **(**Dewannieux *et al*., 2003; Hancks *et al*., 2011).

The cell culture assay has established conclusively the requirement of L1-encoded proteins ORF1p, and ORF2p for retrotransposition (Feng *et al*., 1996; Martin *et al*., 2005; Kulpa and Moran, 2006; Doucet *et al*., 2010). It was found that co-transfection of L1 construct with L1-specific siRNA reduced retrotransposition **(**Soifer and Rossi, 2006). Thus the cell culture assay system has been extensively used to understand the regulation of retrotransposition, especially in the mammalian system.

Construction of a Retrotransposition-Competent Cell Line of *E. histolytica*

A cell culture assay would be similarly beneficial to study the retrotransposition of *Entamoeba* LINEs/ SINEs, and to address the role of DNA methylation and RNA silencing in the modulation of retrotransposition. It has been shown in *E. histolytica* that the methylated base 5-methylcytosine is localized predominantly in repetitive DNA elements, including LINEs (Harony *et al*., 2006), and the protein EhMLBP, which binds to methylated LINEs has been identified (Lavi *et al*., 2006). In addition, sequence analysis of 27-nt small RNAs of *E. histolytica* showed that a small population of these RNAs (0.5%) mapped to retrotransposons **(**Zhang *et al*., 2008). To analyze the functional role of DNA methylation and RNA interference in the control of retrotransposition in *E. histolytica*, a system for *de novo* induction of retrotransposition is required. The construction of such a system for *E. histolytica* is described below.

Although EhLINE1 ORF1p is abundantly expressed in cultured *E. histolytica* cells, full-length transcripts of EhLINE1, and ORF2p polypeptide are not detected; hence these cells are not expected to be retrotransposition-competent. A cell-line was obtained by transfecting *E. histolytica* cells with a plasmid construct containing the complete EhLINE1 ORF2 (reconstructed to remove the stop codons in resident EhLINE1 copies) (Yadav *et al*., 2012). This was cloned in a tetracycline (tet)-inducible expression vector (Fig. 7A), and it expressed the ORF2p (111 kDa) in a tet-inducible manner (Fig. 7B). This cell line also contained a constitutively expressed, marked EhSINE1 copy (with a 25-bp GC-rich tag), and a 176-bp fragment containing the target site sequence where EhSINE1 is known to insert in the *E. histolytica* genome (Mandal *et al*., 2004). Retrotransposition events occurring at this target site in the plasmid were directly scored by PCR amplification of total genomic DNA, followed by Southern hybridization with the marked SINE probe (Fig. 7C). Specific amplicons expected from the mobilization of the marked SINE to the insertion site were obtained only in the presence of tet, when ORF2 was expressed. Retrotransposition was accompanied by target site duplication, a hallmark of retrotransposition (Fig. 7D). It was concluded that the scored events were due to retrotransposition and not due to processes like DNA recombination (Yadav *et al*., 2012). This was a first demonstration of its type in an early branching eukaryote.

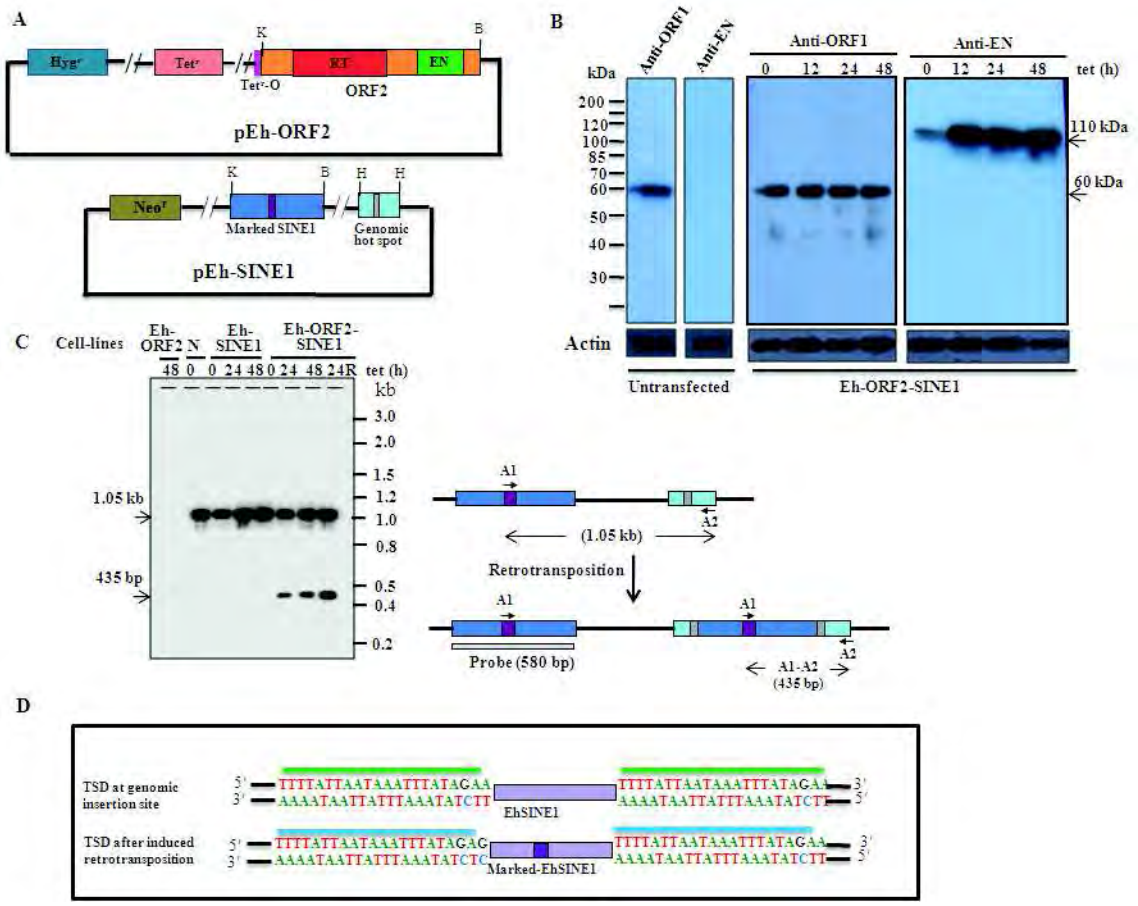


Fig. 7: Demonstration of retrotransposition in the doubly transfected retrotransposition-competent cell line. (A) Of the two plasmids in this cell line, pEh-ORF2 provides the RT and EN activities. pEh-SINE1 contains the marked EhSINE1 with a tag (Purple), and the insertion hot spot (grey box). K, *Kpn*I; B, *BamH*I; H, *Hind*III. (B) Western analyses show that ORF2p (110 kDa) is expressed only upon tet induction of cells transformed with pEh-ORF2. (C) Mobilization of the marked-EhSINE1 copy to the genomic hot spot was detected by 435 bp amplicon obtained with primer pair A1/B1 in the doubly-transfected cells. (D) Retrotransposition is accompanied by 22 bp Target Site Duplication (TSD)

Evidence of Recombination Between SINEs During Retrotransposition in *E. histolytica*

To get some insight into the process of retrotransposition, the question was asked whether the marked SINE copy suffered any changes consequent to retrotransposition (Yadav *et al.*, 2012). This was done by checking the sequences of the newly retrotransposed copies. Retrotransposition events at the target site were retrieved by PCR amplification with primers flanking the target site. The amplicons were cloned and 23 randomly selected clones were sequenced. The data showed that the sequences belonged to three different categories (Fig. 8A). Set I consisted of ten sequences matching completely with the marked SINE. Seven of these were 100% identical to the marked SINE, and three had one mismatch each. Set II contained eight sequences lacking the tag and showing 98-99% identity with various genomic SINE copies but not with the marked SINE copy. It is estimated that 142 SINE copies are transcribed in *E. histolytica* (Huntley *et al.*, 2010), some of which were mobilized upon tet induction. Set III consisted of five sequences containing the 25-bp tag at the expected location but, surprisingly, showed only 94-95% overall sequence identity (22-27 mismatches) with the marked SINE. They also showed at best 94-98% matches with the genomic SINE copies. However, when the sequence on either side of the tag was searched separately (5'-

half and 3'-half of each SINE separately), 98-100% matches were obtained, and each side matched with different genomic SINE sequences (Yadav *et al*., 2012). In these five instances of set III, the tag had associated itself with genomic SINE sequences to result in recombinants derived from at least three different SINE sequences, one of them being the marked SINE and two belonging to different genomic SINEs (Fig. 8B).

Control experiments confirmed that the acquisition of tag by the genomic SINE copies was not through a DNA-recombination event before retrotransposition, and that the recombinants did not exist before the induction of retrotransposition. This study therefore shows that recombinant SINEs are formed consequent to retrotransposition. The process is rapid, as these events were scored within 48 h of retrotransposition induction, and occurred at high frequency (> 20% of total events scored). Some of the events in set II might also be recombinants, as the number of mismatches reduced when the 5' and 3' halves of each sequence were searched separately with the database.

Chimeric molecules arising from reverse transcripts have been observed in yeast Ty elements, and were attributed to gene conversion (Derr and Strathern, 1993) whereas in retroviruses high-frequency recombination occured during reverse transcription of the two co-packaged RNAs in the virion as a result of template switching (Delviks-Frankenberry *et al*., 2011). In non-LTR elements, tripartite chimeric LINEs have earlier been reported

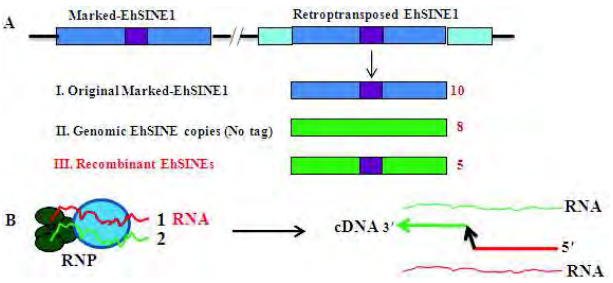


Fig. 8: Retrotransposition is accompanied by EhSINE1 recombination. (A) Sequencing of the newly retrotransposed EhSINE1 copies at the genomic hot spot revealed three categories. (B) Hypothetical model for generation of recombinant SINEs

in a fungal genome (Gogvadze *et al*., 2007). In mammalian genomes U6/L1 pseudogene chimeras have been experimentally demonstrated **(**Garcia-Perez *et al*., 2007). However, recombination between multiple copies of the same SINE family during retrotransposition as observed in *E. histolytica* is a novel observation.

These results are the first direct demonstration that SINE copies engage in active sequence exchange during retrotransposition, leading to the rapid spread of the sequence tag to the SINE population, and generation of diversity. The demonstrated properties of RT to displace the RNA template during cDNA synthesis, and to perform multiple template jumping (Bibillo and Eickbush, 2002; Bibillo and Eickbush, 2004) could lead to these recombinants. It is hypothesized that the RNP particle (formed when the EhLINE1-encoded polypeptides associate with EhSINE1 RNA) contains more than one RNA molecule per RNP. Reverse transcription of these closely associated RNAs could generate recombinant cDNAs by template jumping according to the model shown (Fig. 8B). Human Alu subfamilies show the existence of mosaic elements (Carroll *et al*., 2001), which could arise from such a mechanism. Messenger RNA transcripts in the cell are also templates of the same retrotransposition machinery during retropseudogene formation. If they too engage in similar recombination during reverse transcription, the sequence diversity thus generated may provide selective advantage to the host.

In conclusion, although *E. histolytica* is a challenging system to study due to fastidious growth properties, and poorly developed tools of analysis, this work shows that detailed analysis of *E. histolytica*, and other such organisms, could indeed be rewarding. More work with such systems will enrich our further understanding of biological processes by describing novel mechanisms not reported in model systems.

Future Perspectives

India remains endemic to *E. histolytica* due to poor hygienic conditions which permit continuous transmission of the parasite through the faecal-oral

route. There are large gaps in our understanding of this parasite's basic biology and its interaction with the human host. Apart from studying the parasite in lab culture and using animal models, there is need to address issues like the role of gut micro flora in influencing the severity of invasive disease in amoebiasis patients. Close interaction between clinicians and researchers must be forged to steer research into areas that ultimately benefit the patients.

Acknowledgements

I gratefully acknowledge the contributions of my former and present students who have carried out this work. The funding support was from DBT, DST, and ICMR, Govt. of India. I thank Dr. Vijay Pal Yadav for his help in editing the manuscript.

References

Babushok D V and Kazazian H H Jr (2007) Progress in understanding the biology of the human mutagen LINE-1 *Hum Mutat* **28** 527-539

Bakre A A, Rawal K, Ramaswamy R, Bhattacharya A and Bhattacharya S (2005) The LINEs and SINEs of *Entamoeba histolytica*: comparative analysis and genomic distribution *Exp Parasitol* **110** 207-213

Belancio V P, Hedges D J and Deininger P (2008) Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health *Genome Res* **18** 343-358

Bhattacharya S, Bakre A and Bhattacharya A (2002) Mobile genetic elements in protozoan parasites *J Genet* **81** 73-86

Bhattacharya S, Bhattacharya A, Diamond L S and Soldo A T (1989) Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA *J Protozool* **36** 455-458

Bhattacharya S, Som I and Bhattacharya A (1998) The ribosomal DNA plasmids of *Entamoeba Parasitol Today* **14** 181-185

Bibillo A and Eickbush T H (2004) End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon *J Biol Chem* **279** 14945–14953

Borovjagin A V and Gerbi S A (2000) The spacing between functional *cis*-elements of U3 snoRNA is critical for rRNA processing *J Mol Biol* **300** 57-74

Bourbon H, Michot B, Hassouna N, Feliu J and Bachellerie J P (1988) Sequence and secondary structure of the 5' external transcribed spacer of mouse pre-rRNA *DNA* **7** 181-191

Brouha B, Schustak J, Badge R M, Lutz-Prigge S, Farley A H, Moran J V and Kazazian H H Jr (2003) Hot L1s account for the bulk of retrotransposition in the human population *Proc Natl Acad Sci U S A*. **100** 5280-5285

Burd C E, Jeck W R, Liu Y, Sanoff H K, Wang Z and Sharpless N E (2010) Expression of linear and novel circular forms of an INK4/ARF-associated on-coding RNA correlates with atherosclerosis risk *PLoS Genet* **6**: e1001233

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P and Lovell-Badge R (1993) Circular transcripts of the testis-determining gene *Sry* in adult mouse testis *Cell* **73** 1019–1030

Carroll M L, Roy-Engel A M, Nguyen S V, Salem A H, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde L B, Deininger P L and Batzer M A (2001) Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity *J Mol Biol* **311** 17-40

Choudhuri G and Rangan M (2012) Amebic infection in humans *Indian J Gastroenterol* **311** 53-62

Clark C G, Alsmark U C, Tazreiter M, Saito-Nakano Y, Ali V, Marion S, Weber C, Mukherjee C, Bruchhaus I, Tannich E, Leippe M, Sicheritz-Ponten T, Foster P G, Samuelson J, Noël C J, Hirt R P, Embley T M, Gilchrist C A, Mann B J, Singh U, Ackers J P, Bhattacharya S, Bhattacharya A, Lohia A, Guillen N, Duchene M, Nozaki T and Hall N (2007) Structure and content of the *Entamoeba histolytica* genome *Adv Parasitol* **65** 51-190

Cruz-Reyes J, ur-Rehman T, Spice W M and Ackers J P (1995) A novel transcribed repeat element from *Entamoeba histolytica Gene* **166** 183-184

Delviks-Frankenberry K, Galli A, Nikolaitchik O, Mens H, Pathak V K and Hu W S (2011) Mechanisms and factors that influence high frequency retroviral recombination *Viruses* **3** 1650–1680

Derr L K and Strathern J N (1993) A role for reverse transcripts in gene conversion *Nature* **361** 170–173

Dewannieux M, Esnault C and Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences *Nat Genet* **35** 41-48

Doucet A J, Hulme A E, Sahinovic E, Kulpa D A, Moldovan J B, Kopera H C, Athanikar J N, Hasnaoui M, Bucheton A, Moran J V and Gilbert N (2010) Characterization of LINE-1 ribonucleoprotein particles *PLoS Genet* **6**: e1001150

Feng Q, Moran J V, Kazazian H H Jr and Boeke J D (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition *Cell* **87** 905-

916

Garcia-Perez J L, Doucet A J, Bucheton A, Moran J V and Gilbert N (2007) Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase *Genome Res* **17** 602–611

Garcia-Perez J L, Gonzalez C I, Thomas M C, Olivares M and Lopez M C (2003) Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi Cell Mol Life Sci* **60** 2692-2701

Gogvadze E, Barbisan C, Lebrun M H and Buzdin A (2007) Tripartite chimeric pseudogene from the genome of rice blast fungus *Magnaporthe grisea* suggests double template jumps during long interspersed nuclear element (LINE) reverse transcription *BMC Genomics* **8** 360

Gokal P K, Cavanaugh A H and Thompson E A Jr (1986) The effects of cycloheximide upon transcription of rRNA, 5 S RNA, and tRNA genes *J Biol Chem* **261** 2536-2541

Goodier J L and Kazazian H H Jr (2008) Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites *Cell* **135** 23-25

Gupta A K, Panigrahi S K, Bhattacharya A and Bhattacharya S (2012) Self-circularizing 5'-ETS RNAs accumulate along with unprocessed pre ribosomal RNAs in growth-stressed *Entamoeba histolytica Sci Rep* **2**: 303

Hammann C and Steger G (2012) Viroid-specific small RNA in plant disease *RNA Biol* **9** 809-819

Hancks D C, Goodier J L, Mandal P K, Cheung L E and Kazazian H H Jr (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells *Hum Mol Genet* **20** 3386-3400

Hansen T B, Jensen T I, Clausen B H, Bramsen J B, Finsen B, Damgaard C K and Kjems J (2013) Natural RNA circles function as efficient microRNA sponges *Nature* **495** 384-388

Harony H, Bernes S, Siman-Tov R and Ankri S (2006) DNA methylation and targeting of LINE retrotransposons in *Entamoeba histolytica* and *Entamoeba invadens Mol Biochem Parasitol* **147** 55-63

Houseley J, LaCava J and Tollervey D (2006) RNA-quality control by the exosome *Nat Rev Mol Cell Biol* **7** 529-539

Huber M, Koller B, Gitler C, Mirelman D, Revel M, Rozenblatt S and Garfinkel L (1989) *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules *Mol Biochem Parasitol* **32** 285-296

Hughes J M and Ares M Jr (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA *EMBO J* **10** 4231-4239

Huntley D M, Pandis I, Butcher S A and Ackers J P (2010) Bioinformatic analysis of *Entamoeba histolytica* SINE1 elements *BMC Genomics* **11**: 321

Ivanov V A, Melnikov A A, Siunov A V, Fodor II and Il'in IuV (1991) The jockey mobile genetic element codes a DNA polymerase similar to retroviral reverse transcriptase *Dokl Akad Nauk SSSR* **320** 473-476

Jamburuthugoda V K and Eickbush T H (2011) The reverse transcriptase encoded by the non-LTR retrotransposon R2 is as error-prone as that encoded by HIV-1 *J Mol Biol* **407** 661-672

Jeck W R, Sorrentino J A, Wang K, Slevin M K, Burd C E, Liu J, Marzluff W F and Sharpless N E (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats *RNA* **19** 141-157

Jhingan G D, Panigrahi S K, Bhattacharya A and Bhattacharya S (2009) The nucleolus in *Entamoeba histolytica* and *Entamoeba invadens* is located at the nuclear periphery *Mol Biochem Parasitol* **167** 72-80

Koberna K, Malinsky J, Pliss A, Masata M, Vecerova J, Fialova M, Bednar J and Raska I (2002) Ribosomal genes in focus: new transcripts label the dense fibrillar components and form clusters indicative of "Christmas trees" in situ *J Cell Biol* **157** 743-748

Kolosha V O and Martin S L (1997) In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition *Proc Natl Acad Sci USA* **94** 10155-10160

Kulpa D A and Moran J V (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles *Nat Struct Mol Biol* **13** 655-660

Lavi T, Isakov E, Harony H, Fisher O, Siman-Tov R and Ankri S (2006) Sensing DNA methylation in the protozoan parasite *Entamoeba histolytica Mol Microbiol* **62** 1373-86

Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP and Mann BJ *et al*. (2005) The genome of the protist parasite *Entamoeba histolytica Nature* **433** 865-868

Long E O and Dawid I B (1980) Repeated genes in eukaryotes *Ann Rev Biochem* **49** 727-764

Lorenzi H, Thiagarajan M, Haas B, Wortman J, Hall N and Caler E (2008) Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species *BMC Genomics* **9** 595

Lorenzi H A, Puiu D, Miller J R, Brinkac L M, Amedeo P, Hall N and Caler E V (2010) New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new

genomic features and protein content information *PLoS Negl Trop Dis* **4**: e716

Luan D D and Eickbush T H (1995) RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element *Mol Cell Biol* **15** 3882-3891

Luan D D, Korman M H, Jakubczak J L and Eickbush T H (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

Malik H S, Burke W D and Eickbush T H (1999) The age and evolution of non-LTR retrotransposable elements *Mol Biol Evol* **16** 793-805

Mandal P K, Bagchi A, Bhattacharya A and Bhattacharya S (2004) An *Entamoeba histolytica* LINE/SINE pair inserts at common target sites cleaved by the restriction enzyme-like LINE-encoded endonuclease *Eukaryot Cell* **3** 170-179

Mandal P K, Rawal K, Ramaswamy R, Bhattacharya A and Bhattacharya S (2006) Identification of insertion hot spots for non-LTR retrotransposons: computational and biochemical application to *Entamoeba histolytica Nucleic Acids Res* **34** 5752-5763

Martin S L (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells *Mol Cell Biol* **11** 4804-4807

Martin S L, Cruceanu M, Branciforte D, Wai-Lun Li P, Kwok S C, Hodges R S and Williams M C (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein *J Mol Biol* **348** 549-561

Mayer C, Neubert M and Grummt I (2008) The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus *EMBO Rep* **9** 774–780

McStay B and Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology *Annu Rev Cell Dev Biol*. **24** 131-157

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F and Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency *Nature* **495** 333-338

Michel B, Lizardi P M, Alagón A and Zurita M (1995) Identification and analysis of the start site of ribosomal RNA transcription of *Entamoeba histolytica Mol Biochem Parasitol* **73** 19-30

Moran J V, Holmes S E, Naas T P, DeBerardinis R J, Boeke J D and Kazazian H H Jr (1996) High frequency retrotransposition in cultured mammalian cells *Cell* **87** 917-927

Moss T (2004) At the crossroads of growth control; making ribosomal RNA *Curr Opin Genet Dev* **14** 210-217

Murray H L, Mikheeva S, Coljee V W, Turczyk B M, Donahue W F, Bar-Shalom A and Jarrell K A (2001) Excision of group II introns as circles *Mol Cell* **8** 201-211

Panigrahi S K, Jhingan G D, Som I, Bhattacharya A, Petri W A Jr and Bhattacharya S (2009) Promoter analysis of palindromic transcription units in the ribosomal DNA circle of *Entamoeba histolytica Eukaryot Cell* **8** 69-76

Salzman J, Gawad C, Wang P L, Lacayo N and Brown P O (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types *PLoS One* **7**: e30733

Sehgal D, Mittal V, Ramachandran S, Dhar S K, Bhattacharya, A and Bhattacharya S (1994) Nucleotide sequence organisation and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica Mol Biochem Parasitol* **67** 205-214

Sharma R, Bagchi A, Bhattacharya A and Bhattacharya S (2001) Characterization of a retrotransposon-like element from *Entamoeba histolytica Mol Biochem Parasitol* 116 45-53

Shiao YH, Lupascu ST, Gu YD, Kasprzak W, Hwang CJ, Fields JR, Leighty RM, Quiñones O, Shapiro BA, Alvord WG and Anderson LM (2009) An intergenic non-coding rRNA correlated with expression of the rRNA and frequency of an rRNA single nucleotide polymorphism in lung cancer cells *PLoS One* **4**: e7505

Soifer H S and Rossi J J (2006) Small interfering RNAs to the rescue: blocking L1 retrotransposition *Nat Struct Mol Biol* **13** 758-759

Upcroft P and Upcroft JA (2001) Drug targets and mechanisms of resistance in the anaerobic protozoa *Clin Microbiol Rev* **14** 150-164

Van Dellen K, Field J, Wang Z, Loftus B and Samuelson J (2002) LINEs and SINE-like elements of the protist *Entamoeba histolytica Gene* **297** 229-239

Venema J and Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae Annu Rev Genet* **33** 261-311

Vicens Q and Cech T R (2009) A natural ribozyme with 3',5' RNA ligase activity *Nat Chem Biol* **5** 97-99.

Warner J R (1999) The economics of ribosome biosynthesis in yeast *Trends Biochem Sci* **24** 437-440

WHO/PAHO/UNESCO report (1997) A consultation with experts on amoebiasis. Mexico City, Mexico. 28–29 January, 1997 Epidemiol Bull **18** 13-14

Willhoeft U, Buss H and Tannich E (2002) The abundant polyadenylated transcript 2 DNA sequence of the pathogenic protozoan parasite *Entamoeba histolytica* represents a nonautonomous non-long-terminal-repeat retrotransposon-like element which is absent in the closely related nonpathogenic species *Entamoeba dispar Infect Immun* **70** 6798-6804

Yadav V P, Mandal P K, Bhattacharya A and Bhattacharya S (2012) Recombinant SINEs are formed at high frequency during induced retrotransposition in vivo *Nat Commun* **3** 854

Yadav V P, Mandal P K, Rao DN and Bhattacharya S (2009) Characterization of the restriction enzyme-like endonuclease encoded by the *Entamoeba histolytica* non-long terminal repeat retrotransposon EhLINE1 *FEBS J* **276** 7070-7082

Yang J, Malik and Eickbush T H (1999) Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements *Proc Natl Acad Sci USA* **96** 7847-7852

Zaphiropoulos P G (1997) Exon skipping and circular RNA formation in transcripts of the human cytochrome P-4502C18 gene in epidermis and rat androgen binding protein gene in testis *Mol Cell Biol* **17** 2985-93

Zhang H, Ehrenkaufer G M, Pompey JM, Hackney J A and Singh U (2008) Small RNAs with 5'-polyphosphate termini associate with a Piwi-related protein and regulate gene expression in the single-celled eukaryote *Entamoeba histolytica PLoS Pathog* **4** e1000219

Zhang Y, Zhang X O, Chen T, Xiang J F, Yin Q F, Xing Y H, Zhu S, Yang L and Chen L L (2013) Circular intronic long noncoding RNAs *Mol Cell* **51** 792-806

Zhao J, Yuan X, Frodin M and Grummt I (2003) ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth *Mol Cell* **11** 405-413

Zurita M, Alagon, A, Vargas-Villarreal J and Lizardi P M (1991) The *Entamoeba histolytica* rDNA episome: nuclear localization, DNAase I sensitivity map, and specific DNA-protein interactions *Mol Microbiol* **5** 1843-1851.