Available online at www.sciencedirect.com
SCIENCE DIRECT
International Journal for Parasitology 34 (2004) 1365–1375
www.parasitology-online.com

The *fugitive* LTR retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*

Thewarach Laha[a], Alex Loukas[b], Danielle J. Smyth[b], Claudia S. Copeland[c], Paul J. Brindley[c,*]

[a]*Department of Parasitology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand*
[b]*Division of Infectious Diseases and Immunology, Queensland Institute of Medical Research, Brisbane, Qld 4029, Australia*
[c]*Department of Tropical Medicine, and Interdisciplinary Program in Molecular and Cellular Biology, School of Public Health and Tropical Medicine, SL-17, Tulane University, 1430 Tulane Avenue, New Orleans, LA 70112, USA*

Received 1 July 2004; received in revised form 19 August 2004; accepted 19 August 2004

Abstract

Inspection of the nucleotide sequence of bacterial artificial chromosome number 49_J_14 [Le Paslier, M.C., Pierce, R.J., Merlin, F., Hirai, H., Wu, W., Williams, D.L., Johnston, D., LoVerde, P.T., Le Paslier, D., 2000. Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. Genomics 65, 87–94] from chromosome 1 of the genome of *Schistosoma mansoni* (GenBank AC093105) revealed the likely presence of a proviral form of a novel schistosome retrotransposon. The novel element, which we named the *fugitive*, belonged to the *mag*-like family of the *gypsy-Ty3* clade of long terminal repeat retrotransposons. It was closely related to the *mag*-like retrotransposon *Gulliver* from *Schistosoma japonicum*, but was dissimilar to several other long terminal repeat retrotransposons known from *S. mansoni* including *Boudicca*, *Saci-1*, *Saci-2* and *Saci-3*. The full length *fugitive* element was 4811 bp constituted of a single read-through open reading frame of 4134 bp flanked at both ends by identical long terminal repeat sequences of 273 bp. The open reading frame encoded retroviral-like gag, with a distinctive double Cys–His motif, and pol polyprotein, with a pol domain order of protease, reverse transcriptase, RNaseH and integrase. Examination of schistosome transcriptome sequences in the public domain revealed that the *fugitive* was transcribed in at least six developmental stages of *S. mansoni*, while bioinformatics approaches and Southern hybridisation analysis indicated that as many as 2000 copies of the *fugitive* were interspersed throughout the schistosome genome.

Keywords: Schistosome; Mobile genetic element; Retrotransposon; Reverse transcriptase; Genome; *Gulliver*

1. Introduction

Mobile genetic elements including transposons, retrotransposons and retroviruses have colonised the genomes of most eukaryotes (Kazazian, 2004). These mobile elements have influenced genome evolution, often are active in the germ line, can be present in large numbers, and often constitute substantial proportions of the host genome. Uncontrolled proliferation of mobile genetic elements appears to be constrained by natural RNA interference processes including dicer enzyme activity (Sijen and Plasterk, 2003) although many examples of deleterious mutations resulting from retrotransposon mobilisations are known from humans and other species (e.g. Brouha et al., 2002).

Schistosomiasis is caused by infection with blood flukes of the trematode genus, *Schistosoma*, and is considered the most important of the human helminthoses in terms of morbidity and mortality (Engels et al., 2002). International efforts are currently underway to determine the entire genome sequence for *Schistosoma japonicum* and for *Schistosoma mansoni* (Hu et al., 2003; Verjovski-Almeida et al., 2003; El-Sayed et al., 2004). As much as half of the *S. mansoni* genome might be comprised of repetitive sequences, and many of these repetitive sequences can be expected to be mobile genetic elements (Brindley et al., 2003). Knowledge of these endogenous mobile elements is of value in order to facilitate the sequencing and annotation of the schistosome genome by shotgun sequencing

* Corresponding author. Tel.: +1 504 988 4645; fax: +1 504 988 6686.
E-mail address: paul.brindley@tulane.edu (P.J. Brindley).

0020-7519/$30.00 © 2004 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijpara.2004.08.007

approaches (El-Sayed et al., 2004), in investigation of comparative evolution of schistosome species (Stein et al., 2003), and because of the potential use of these kinds of elements in transgenesis studies (Ivics et al., 1997; Teysset et al., 1998). Several long terminal repeat (LTR) retrotransposons have been reported from schistosome genomes (Laha et al., 2001; Copeland et al., 2003; DeMarco et al., 2004) as have a number of other categories of mobile genetic elements (see Brindley et al., 2003).

Here we describe the presence, sequence, structure and evolutionary relationships of the *fugitive*, a novel LTR retrotransposon interspersed within the genome of *S. mansoni*. The *fugitive* is a *gypsy*-like element that appears to be transcriptionally active, is present in high copy number, and indeed may comprise as much as 4% of the *S. mansoni* genome.

2. Materials and methods

2.1. Bioinformatics approaches for detection of mobile sequences in the schistosome genome

The keyword phrase ⟨Reverse Transcriptase⟩ was used as the query to search the EST_others and GSS databases at GenBank for novel schistosome sequences associated with mobile genetic elements. Schistosome RT-like sequences that were retrieved were employed subsequently to search for matches in the GenBank non-redundant sequence database using Blastn, Blastx and/or tBlastn. Sequences of the previously characterised schistosome retrotransposons *Gulliver* (Laha et al., 2001) and *Boudicca* (Copeland et al., 2003) also were employed as queries. In addition, retrotransposon integration sites were investigated by interrogation of the *S. mansoni* genome survey sequences (GSS) at the Sanger Institute, Hinxton, UK (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_mansoni).

2.2. Sequence analysis, amino acid alignments, and phylogenetic trees

Retrotransposon-like sequences were conceptually translated into amino acid sequences, and functional domains and motifs were compared and contrasted with those of other retrotransposons. Alignments of amino acid sequences of functional domains were accomplished with ClustalW (Thompson et al., 1994) and edited with Bioedit v5.0.9 (Hall, 1999) (www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequences for phylogenetic analysis comparing the RT region of retrotransposons were trimmed from the large single polyprotein of each retrotransposon to retain only the conserved domains of RT (Xiong and Eickbush, 1990; Abe et al., 2001). Where available, amino acid sequences were taken directly from GenBank. For elements not included in the GenBank protein database, nucleotide sequences were examined for open reading frames (ORFs), translated, and trimmed to include only the seven domains of RT (assisted by MacVector, http://www.accelrys.com/). Alignments were accomplished using Clustal X (Thompson et al., 1997) after which bootstrapped (1000 repetitions) trees were assembled using the neighbor joining method and Njplot (Saitou and Nei, 1987). The accession numbers for sequences included in the phylogenetic analysis are as follows: *Gulliver* (AF243513), *Saci-1* (BK004068), *Saci-2* (BK004069) and *Saci-3* (BK004070) from species of *Schistosoma*, retrovirus-related polyprotein from *Caenorhabditis elegans* (AAB42240), *CsRn1* from *Clonorchis sinensis* (AAK07487), *412* from *Drosophila melanogaster* (CAA27750), *17.6* from *D. melanogaster* (CAA25702), *297* from *D. melanogaster* (X03431), *Tom* from *Drosophila ananassae* (CAA80824), *Gypsy*Dm from *D. melanogaster* (AAC82604), *Gypsy*Dv from *Drosophila virilis* (AAA28600), *Zam* from *D. melanogaster* (CAA04050), *Ted* from *Trichoplusia ni* (AAA92249), *Woot* from *Tribolium castaneum* (AAC47271), *Ulysses* from *D. virilis* (CAA39967), *mag* from *Bombyx mori* (S08405), *Osvaldo* from *Drosophila buzzatii* (AJ133521), *Micropia* from *D. melanogaster* (CAA32198), *Ty3* from *Saccharomyces cerevisiae* (S53577), *mdg3* from *D. melanogaster* (CAA65152), *Tf1* from *Schizosaccharomyces pombe* (AAA35339), *Del-1* from *Lilium henryi* (1510387A), *Maggy* from *Oryza sativa* (AAA33420), *Cer-1* from *C. elegans* (U15406), *Tnt1* from *Nicotiana tabacum* (CAA32025), *Cft1* from *Cladosporium fulvum* (CAA77891), *copia* from *D. melanogaster* (CAD27357), *Sushi* from *Takifugu rubripes* (AAC33526), *Kamikaze* from *B. mori* (AB042120). The sequence representing *Pao* was a reconstruction made by Abe et al. (2001), from accession numbers S33901, AB042118, and AB042119, and the sequence representing *Boudicca* was a composite of translated cDNA sequences introduced in Copeland et al. (2004), AY308018, AY308019, AY308021 and AY308022 (these composite sequences for *Pao* and *Boudicca* are available from the corresponding author).

2.3. Parasites and parasite DNA; Southern hybridisation analysis

The *S. mansoni* life cycle was maintained in experimentally infected mice and *Biomphalaria glabrata* snails at the Queensland Institute of Medical Research, Brisbane, Australia. Genomic DNA (gDNA) of adult mixed sexes of *S. mansoni* perfused from these mice was extracted as described (Drew and Brindley, 1997). For Southern blots, 30 μg of *S. mansoni* gDNA was digested with *Hin*dIII or *Bam*H1, size fractionated by electrophoresis through 0.8% agarose in Tris-acetate-EDTA, and transferred by capillarity to nylon membrane (Hybond-N+, Amersham Biosciences). Southern hybridisation analysis was performed using a horseradish peroxidase labelled (ECL Direct Nucleic Acid Labelling and Detection System, Amersham Biosciences) gene probe (below). The membrane was incubated in

hybridisation medium (provided with kit) supplemented with the labeled probe overnight at 42 °C, after which the membrane was washed in 0.4% SDS, 0.5× SSC (NaCl, Na Citrate buffer pH 7.0) at 42 °C (two washes, 20 min each wash) and subsequently in 2× SSC at room temperature (two washes, 5 min each). The gene probe was prepared from a PCR product amplified from the reverse transcriptase (RT) domain of the new schistosome retrotransposon. The RT encoding domain of the retrotransposon was amplified using specific forward, 5′-AGTGAAAGCTCATTTACCCC, and reverse primers, 5′-TTCTATATTGGCAGGGTCTGG, targeting nucleotides 1761–2448 of GenBank accession BK005225, using gDNA of *S. mansoni* as the template. PCR cycling conditions were 95 °C for 1 min, 55 °C for 1 min and 72 °C for 3 min, 35 cycles and final extension of 10 min at 72 °C. PCR products were sized by electrophoresis through 1% agarose, stained with ethidium bromide and photographed under UV illumination. PCR products that were isolated from the gel were cloned into plasmid pCR4-TOPO (Invitrogen), after which the identity of the cloned inserts was confirmed by automated DNA sequencing.

2.4. Copy number estimation

Estimates of the copy number of the *fugitive* retrotransposon were obtained by a comparative bioinformatics approach (Copeland et al., 2003; DeMarco et al., 2004) wherein Blast analysis of the bacterial artificial chromosome (BAC)-end database of *S. mansoni* genomic sequences targeted more well-characterised retrotransposable elements from *S. mansoni* for which copy numbers had been reported. These included the *Boudicca* LTR retrotransposon (Copeland et al., 2003), the non-LTR retrotransposons *SR1* and *SR2* (Drew and Brindley, 1997; Drew et al., 1999), and the SINE-like element, *SMα* (Spotila et al., 1989). The advanced search option was used for these Blastn searches, with the expect value for this analysis set at a more stringent value of 0.1 rather than at the default setting of 10. Furthermore, the searches were limited to include only *S. mansoni* sequences by employing the Entrez query, ⟨Schistosoma mansoni[organism]⟩. Only hits with a Blast score of ≥100 were counted.

2.5. Analysis of the developmental stage of expression of the new retrotransposon

To investigate developmental stage specificity of transcription of the new retrotransposon, the full length *fugitive* sequence was employed as the query in Blastn-based searches of the GenBank EST database, which includes more than 130,000 ESTs from six discrete developmental stages of *S. mansoni* (Verjovski-Almeida et al., 2003). Sequences with over 85% identity over at least 150 nucleotides (nt) were considered as representing *fugitive* transcripts. Subsequently, each of the positive EST matches was aligned to the consensus *fugitive* element to determine the identity of the transcribed domain of the *fugitive* retrotransposon.

2.6. Investigation of schistosome retrotransposon integration sites

Integration sites of the new retrotransposon within the *S. mansoni* genome were investigated by Blast analysis against the schistosome sequences in the non-redundant, GSS and EST databases at GenBank, and against *S. mansoni* genomic sequences retrieved from the databases at the Sanger Institute using, as the query sequences, about 2000 bp both upstream and downstream of the newly characterised *fugitive* retrotransposon within BAC number 49_J_14 (Le Paslier et al., 2000) (accession numbers AC093105 and BK005225). In addition, regions flanking the 5′- and 3′-LTRs of the *fugitive* were investigated to determine potential protein-encoding sequences and other motifs that might be targeted for insertion by the retrotransposon.

2.7. GenBank accession number

The genomic sequence of the new LTR retrotransposon, *fugitive*, has been assigned the Third Party Annotation accession number BK005225.

3. Results

3.1. Novel schistosome retrotransposons identified in bacterial artificial chromosome

Blastn searches indicated the presence of RT-encoding sequences in the bacterial artificial chromosome number 49_J_14 (Le Paslier et al., 2000), the entire sequence of which has been lodged in GenBank with accession number AC093105 by El Sayed and co-workers (El-Sayed et al., 2004). Inspection of the nucleotide sequence of BAC 49_J_14, of ∼123 kb, suggested the presence of a number of discrete retrotransposons, one of which appeared to encode a novel LTR retrotransposon that we have termed the *fugitive* (Fig. 1A), the characterisation of which forms the basis of this present report. The *fugitive* retrotransposon was located between nucleotide residues 80,490 and 75,680 of BAC 49_J_14, in the reverse orientation to the direction of the annotated sequence of AC093105. BAC 49_J_14 also includes at least three non-LTR retrotransposons, as illustrated in Fig. 1A as *SR2*, *SR3* and *SR4*, which exhibit identity to other non-LTR retrotransposons reported previously from the genome of *S. mansoni* (Drew and Brindley, 1997; Drew et al., 1999; DeMarco et al., 2004). Characterisation of these latter elements will be reported separately (Laha and others, manuscript in preparation). Annotation provided with GenBank AC093105 indicated

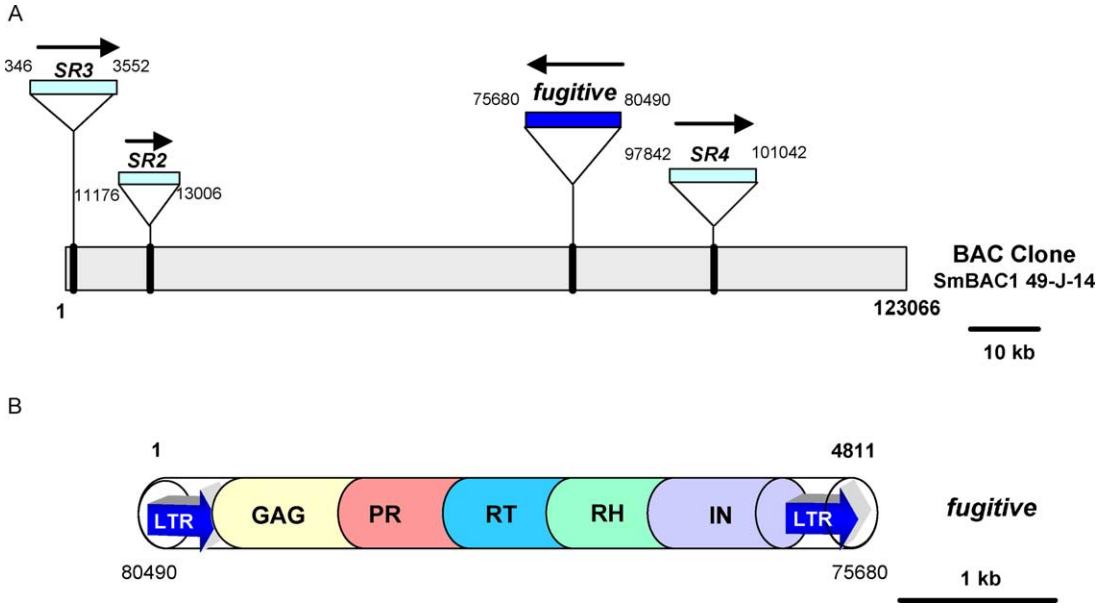

Fig. 1. Schematic diagram of the location, size and structure of the *fugitive* LTR retrotransposon in bacterial artificial chromosome number 49_J_14 from chromosome 1 of the genome of *Schistosoma mansoni*. Panel A: schematic diagram showing the location, size, and direction of the *fugitive* retrotransposon, and several non-LTR retrotransposons (SR2, SR3 and SR4) from *S. mansoni* in BAC 49_J_14. Panel B: schematic representation of the key structures and domains of the *fugitive* LTR retrotransposon. Gag, group specific antigen; PR, protease; RT, reverse transcriptase; RN, RNaseH; IN, integrase; LTR, long terminal repeat.

that the sequence included in BAC 49_J_14 is from chromosome 1 of the genome of *S. mansoni*.

3.2. *Fugitive, a novel LTR retrotransposon from the genome of* S. mansoni

The pro-viral form of the *fugitive* identified in BAC 49_J_14 was 4811 bp in length (residues 80,490 to 75,680 of BAC 49_J_14), and included a single read-through open reading frame (ORF) of 4134 bp flanked at both ends by identical LTR sequences of 273 bp (Fig. 1). Complete identity in sequence of both the 5′- and 3′-LTR indicated that the *fugitive* retrotransposon remained transpositionally active or had been transposed in the recent past. The LTRs of the *fugitive* began with TGA and ended with ACA, authentic versions of the common retrotransposon 5′- and 3′-terminal trinucleotide motifs, TGN and NCA, and they exhibited direct inverted repeats of TG and CA (Bowen and McDonald, 1999). Each of the *fugitive* LTRs contained a TATA box and a CAAT box (Supplementary Fig. 1), transcriptional signals for RNA polymerase II. LTRs are expected to include promoter as well as terminator signals for RNA polymerase II, although a polyadenylation signal similar to ATAAA was not apparent in the *fugitive* LTR. Blastn searches using the *fugitive* LTR as the query sequence revealed no significant matches in the non-redundant database at GenBank to the LTR of any other retrotransposon. However, comparisons of nucleotide sequences of the LTRs of the *fugitive* and the *Gulliver* retrotransposon of *S. japonicum* revealed 50% identity with gaps included and 27% identity when gaps were excluded (Supplementary Table 1).

Multiple sequence alignments of the Cys–His box region of gag and of the entire retroviral pol polyprotein of the *fugitive* and related retrotransposons are presented in Fig. 2. The strong sequence identity and similarity among these elements were evident in the alignment). At its NH_2 terminal-encoding side, the *fugitive* ORF encoded ~300 amino acid residues that shared identity with gag polyproteins from a number of retrotransposons including *Gulliver* of *S. japonicum* (35% identity/61% similarity). The gag (group specific antigen) polyprotein, which surrounds the RNA genome of retrotransposons and retroviruses, includes matrix, capsid and nucleocapsid domains, and is characterised by the presence of conserved domains termed the Major Homology Region (MHR) and the Cys–His box (Meric and Goff, 1989; Gonsky et al., 2001). The ORF was mutated by the presence of a single stop codon at amino acid position 95 (Supplementary Fig. 1) within the gag-encoding region. Further, the gag ORF included two Cys–His box-like motifs, $CX_2CX_3HX_4C$ and $CX_2CX_4HX_3HC$, as also reported for *mag* (Garel et al., 1994) (X denotes any amino acid). These motifs encode zinc finger binding domains of the retroviral-like nucleocapsid protein, which play a key role in RNA binding (Fig. 1B; Supplementary Fig. 1). The Cys–His box motif residues of the gag polyprotein of the *fugitive* are presented in multiple sequence alignment in Fig. 2, panel A, with orthologous residues from several closely related retrotransposons. The presence of two Cys–His motifs was notable because

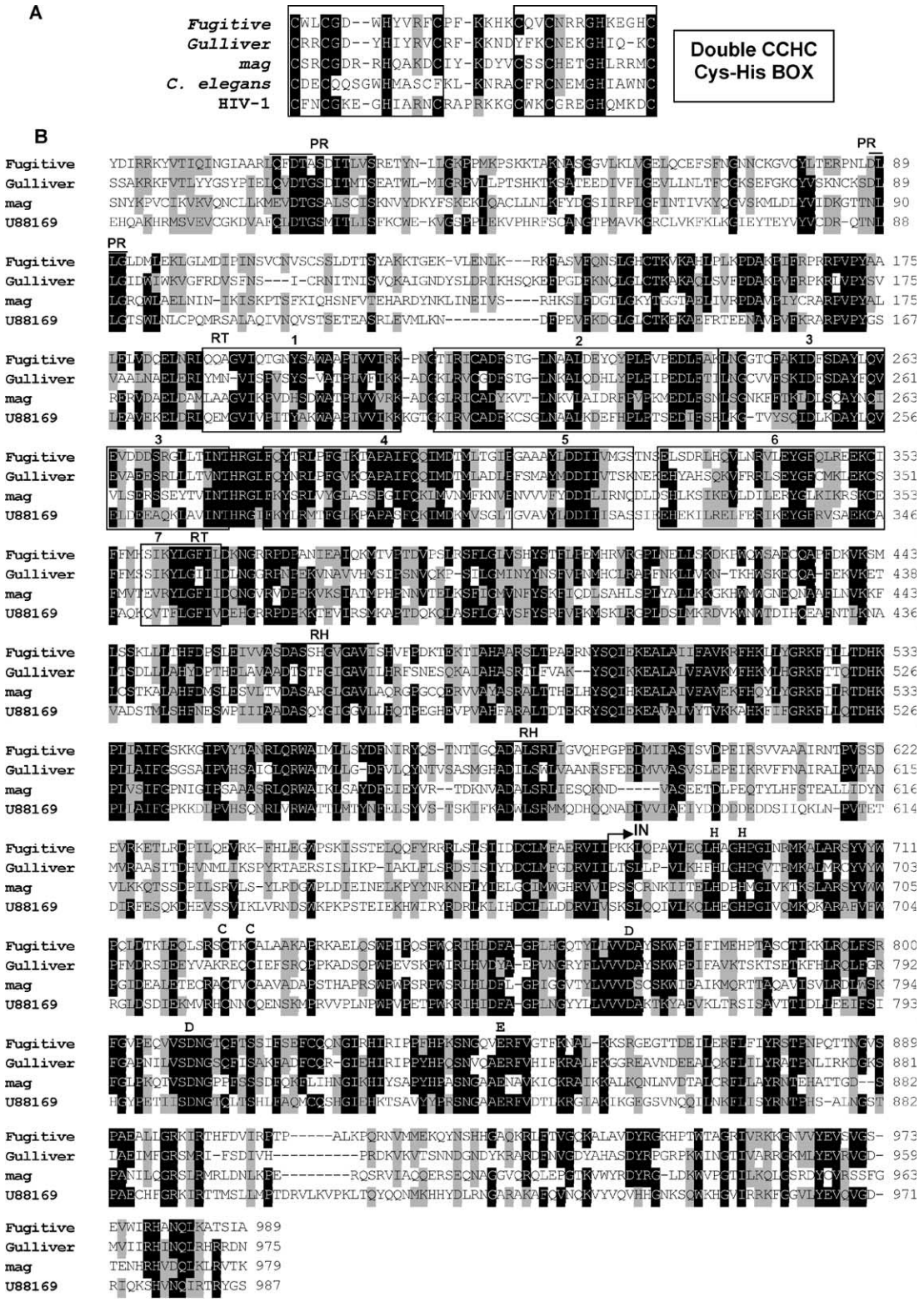

Fig. 2. Multiple sequence alignments of functional domains encoded by the ORF of the *fugitive*. Panel A: alignment of the Cys–His region of gag of the *fugitive* and other retrotransposable elements including *Gulliver* from *Schistosoma japonicum* (AF243513), *mag* from *Bombyx mori*, uncharacterised retrotransposon from *Caenorhabditis elegans* (U88169) and human immunodeficiency virus 1 (HIV1) (AF005496). Panel B: alignment of the predicted protease (PR), reverse transcriptase (RT), RNaseH (RH) and integrase (IN) domains of the *fugitive* from *S. mansoni* and related LTR retrotransposons including *Gulliver*, *mag* and the uncharacterised retrotransposon from *C. elegans* (U88169). Amino acids identical to at least half of the sequences are shown in black boxes while conservative substitutions are in light grey boxes. The sequences were adjusted manually to maximise the alignment of conserved residues; numbered boxes in the RT domain correspond to previously defined conserved blocks (Xiong and Eickbush, 1990).

retroviral-like gag polyproteins from many *gypsy/Ty3*-like retrotransposons include just a single Cys–His box, e.g. as in *Boudicca* (Copeland et al., 2003).

Downstream of *gag*, the ORF encoded an additional 1077 amino acid residues that shared strong identity with retroviral polyproteins (pol). More specifically, it encoded protease (PR), reverse transcriptase (RT), RNaseH (RH) and integrase (IN) domains, in that order, a domain order characteristic of the *gypsy/Ty3*-like clade of LTR retrotransposons (Boeke and Corces, 1989). Whereas the size of these domains is not clear at this time, based on the similarity of structure with other retroviral pol domains, the PR domain is expected to be ~100 amino acids in length. It included conserved active site DT/SG and LLG motifs, which characterise the active site cleft of other retroviral aspartic proteases (Pearl and Taylor, 1987; Bowen and McDonald, 1999). The region encoding the PR domain in BAC 49_J_14 had DTA at the active site, rather than the expected DTG. The positions of these active site residues of PR are indicated in Fig. 2. The RT domain was ~200 amino acids long, and included the seven conserved blocks of residues described by Xiong and Eickbush (1990) to constitute RT of retroviruses and LTR retrotransposons. The consensus active site residues, F/YXDD, were present in the RT of the *fugitive* as YLDD, in conserved block 5. The positions of these seven blocks are indicated in Fig. 2. The RH enzymatic domain was also ~200 amino acids long and included the conserved motifs, DASXXGXGAV and ADXLSRL (Springer and Britten, 1993) (Fig. 2). At its COOH-terminus, the *fugitive* pol included an IN domain of ~300 amino acids in length, and included the conserved D(38)D(35)E motif that characterises IN enzymes of other retrotransposons and transposons (Capy et al., 1996). Further, at its NH_2-terminus, the IN domain of the *fugitive* also included several conserved Cys and His residues that constitute the zinc-binding ligand of IN (Fig. 2). Finally, the *fugitive* showed high similarity to *Gulliver* from *S. japonicum* in both sequence composition (Supplementary Table 1) and structure (Laha et al., 2001). Inspection of the deduced amino acid sequences revealed that *Gulliver* and *fugitive* shared 42% identity and 66% similarity over the entire ORFs.

3.3. Fugitive is a member of the mag clade of gypsy/Ty3 LTR retrotransposons

The region of the ORF spanning the seven conserved blocks of the RT domain of the *fugitive* was aligned with orthologous RT regions of a number of other informative LTR retrotransposons and retroviruses, using ClustalW after which bootstrapped trees were assembled using the neighbor joining method and Njplot. The phylogenetic tree confirmed that the *fugitive* belonged to the *gypsy/Ty3* clade of LTR retrotransposons, and that its closest relatives were *Gulliver* from the Asian schistosome, *S. japonicum* (Laha et al., 2001), *mag* from *B. mori* (Garel et al., 1994), and an uncharacterised LTR retrotransposon of *C. elegans* (Fig. 3). The tree also confirmed the presence of a number of other discrete clades of elements that constitute the *gypsy/Ty3* assemblage of LTR retrotransposons, including the *mag*, *Ty3*, *gypsy*, *CsRn1/kabuki*, and *micropia* clades. Other *gypsy/Ty3*-like retrotransposons reported previously from schistosomes, *Boudicca* and *Saci-3* (Copeland et al., 2003; DeMarco et al., 2004) and *Saci-2* (DeMarco et al., 2004), fell into the *CsRn1/kabuki* and the *Ty3* groups, respectively. In addition, *Saci-3* from *S. mansoni* (DeMarco et al., 2004) is a *Pao/BEL*-like retrotransposon (Fig. 3). Phylogenetic analysis of the RNaseH domain using the same approach likewise revealed that the *fugitive* was closely related to *Gulliver* and belonged to the *mag*-like clade of the *gypsy/Ty3* type LTR retrotransposons (not shown).

3.4. Numerous copies of fugitive are interspersed throughout the S. mansoni genome

Numerous bands of hybridisation were evident when a Southern blot of gDNA of *S. mansoni* was probed with the labeled fragment encoding part of the RT domain of the *fugitive*, confirming the presence of the *fugitive* in the genome of NMRI strain of *S. mansoni*. Lanes of gDNA digested with *Hin*dIII and *Bam*H1 both revealed smeared, though distinct, patterns, but also revealed a strong band of hybridisation at ~1.5 kb in the *Hin*dIII digested gDNA. *Bam*HI cuts once whereas *Hin*dIII cuts twice (to release a fragment of ~1.5 kb as evident in the hybridisation signal) within the 4811 bp of the *fugitive* copy in BAC 49_J_14. The probe did not contain cleavage sites for these enzymes. The pattern of hybridisation indicated that many copies of the *fugitive* element were interspersed throughout the genome of *S. mansoni* rather than being localised at a discrete locus (Fig. 4).

Blastn searches of the *S. mansoni* BAC end sequences in the GenBank database (42,017 entries, dbGSS release 080604), using as the query the 4811 bp of the *fugitive* copy in BAC clone 49_J_14 (BK005225), returned 106 hits with a score >100. In comparison, queries with the non-LTR retrotransposons *SR1* (partial sequence, 2337 bp consensus, U66331) and *SR2* (3913 bp, AF025673) from *S. mansoni* returned 104 and 102 hits, respectively. Other comparisons with the *SMα* retroposon (331 bp, M27676), the 18S rRNA gene of *S. mansoni* (1739 bp, M62652), the SL RNA gene (595 bp, M62652) and the cDNA encoding the *S. mansoni* cathepsin D protease (1285 bp, U60995) returned 162, two, zero and zero hits, respectively (Table 1). Since gene copy numbers have been estimated for these other query sequences, e.g. SR1, 200–2000 copies, *SR2*, 1000–10,000 copies, *SMα*, 7000–10,000 copies, 18S rRNA gene, 100 copies, SL RNA gene, 200 copies, and cathepsin D, one copy (Table 1), and since the hits value for the *fugitive* was similar to those of *SR1* and *SR2*, yet *fugitive* was a longer query sequence (4811 bp), the findings suggested that there may be as many as 2000 copies of

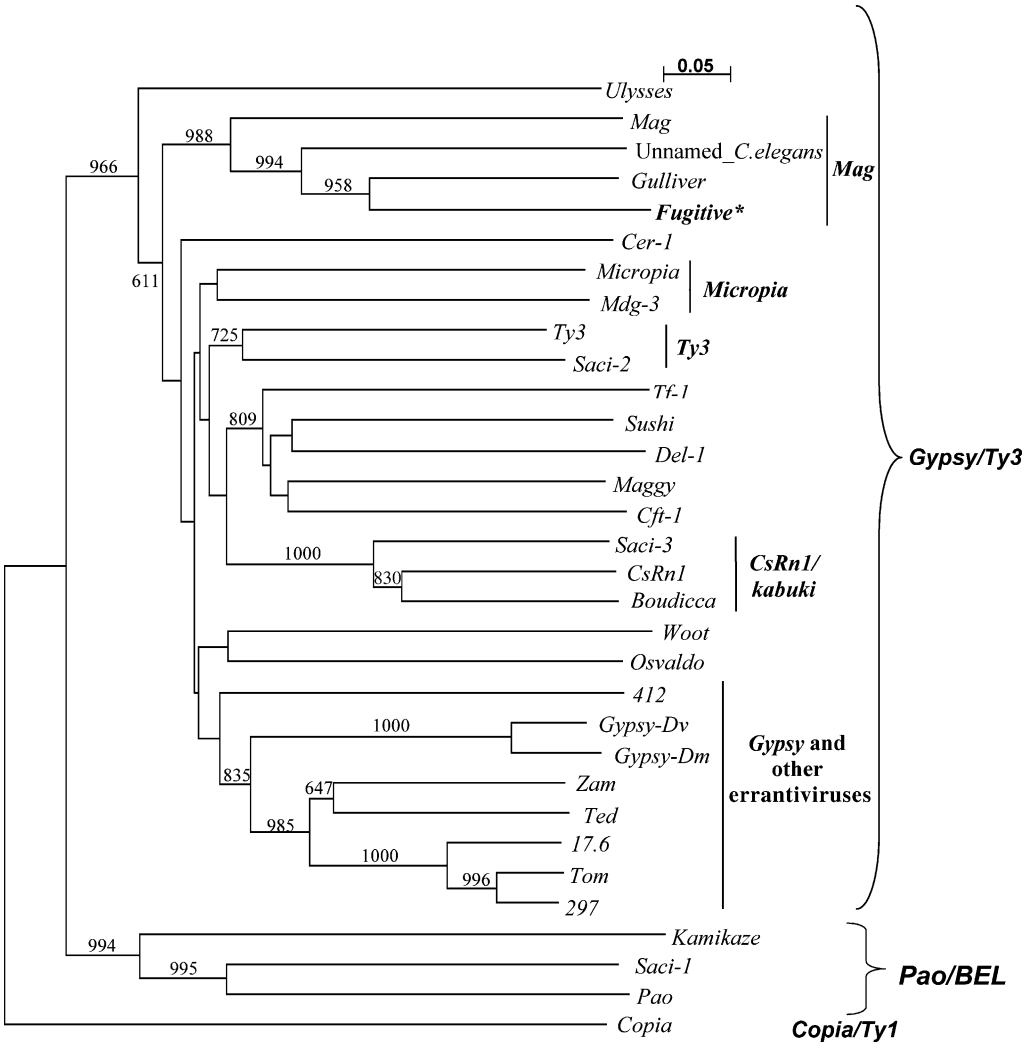

Fig. 3. Phylogram constructed using the neighbor joining method to compare the relationships among reverse transcriptases of the *fugitive* and of other LTR retrotransposons and retroviruses from a range of host genomes. Representatives of the three major clades of LTR retrotransposons—*gypsy/Ty3*, *Copia/Ty1*, and *Pao/BEL*—are included. Bootstrap values, where 500 or greater, are presented at the nodes.

this LTR retrotransposon interspersed throughout the genome of *S. mansoni*. Further, this analysis suggested that there may be >3000 copies of the LTR of the *fugitive* retrotransposon within the *S. mansoni* genome (Table 1). If there are as many as 2000 copies, and given that the full-length retrotransposon is 4811 bp in length and that the haploid genome size of *S. mansoni* is estimated at 270 Mbp (see Brindley et al., 2003), *fugitive* sequences may comprise as much as 4% of the genome of this schistosome.

3.5. Fugitive transcribed in all developmental stages of S. mansoni

Blastn analyses were undertaken using the full length of *fugitive* as the query sequence and the GenBank EST database of non-human, non-mouse sequences. The database includes more than 130,000 EST sequences from six developmental stages of *S. mansoni*—egg, miracidium, cercaria, germball (=sporocyst), schistosomulum, and mixed sex adults (Verjovski-Almeida et al., 2003). Significant hits were found to ESTs from all of these six developmental stages (not shown). Representative accession numbers of these positive matches are presented in Supplementary Table 2, along with some details of the region where the matches were located and on the percent identity. In brief, positive ESTs spanning most or all of the LTR, gag, PR, RT, RH and/or IN regions were located in most of these six developmental stages. Based on these results, it appeared that the *fugitive* was expressed in all developmental stages of *S. mansoni*.

3.5.1. Fugitive integration sites

In order to investigate the nature of integration sites or target sequences of the new retrotransposon within the schistosome genome, the *fugitive* LTR sequence was employed as the query to search the GSS of *S. mansoni*. There were 3431 significant hits among the 1,100,000 GSS of *S. mansoni* in the Sanger Institute database. Of these hits,

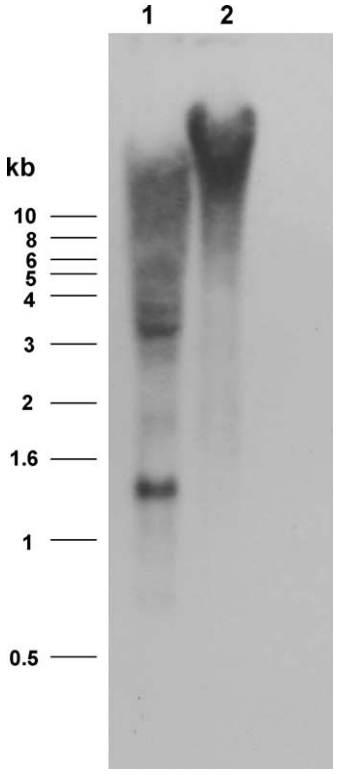

Fig. 4. Southern hybridisation analysis of genomic DNA of *Schistosoma mansoni* probed with a *fugitive*- specific gene fragment. Lanes 1 and 2, *Hin*dIII and *Bam*H1 digested genomic DNAs, respectively. Molecular size standards in kilobase are shown at the left.

only 19 exhibited 97% identity to the LTR of *fugitive* (not shown). These 19 sequences were aligned and compared with termini of the 5′- or the 3′-LTRs and flanking sequences of the BAC 49_J_14 copy of *fugitive* (Fig. 5). The alignment revealed that the flanking sequences of each of the six positive sequences flanking the 5′-LTR were almost identical over about 50 nt immediately upstream of the 5′-LTR and, moreover, all these six flanking sequences were AT rich. Whereas no similarly strong identity was seen among the sequences in the insertion region sequences flanking the 3′-LTR, these were all AT rich (60–70% AT) (Fig. 5). In addition, Blast searches employing the sequences flanking the *fugitive* in BAC 49_J_14 as the query against the GenBank non-redundant protein and nucleotide databases mostly revealed no significant matches to any *Schistosoma* species coding sequences.

4. Discussion

Although four discrete LTR retrotransposons have been characterised recently from the genome of *S. mansoni*, elements named *Boudicca*, *Saci-1*, *Saci-2*, and *Saci-3* (Copeland et al., 2003; DeMarco et al., 2004), along with some fragments of other elements (Foulk et al., 2002), the *fugitive* retrotransposon characterised in this present study is a novel retrotransposon discrete from these other elements. Sequence identity, structure, and phylogenetic relationships demonstrated that the *fugitive* belonged to the *mag* family of *gypsy/Ty3*-like retrotransposons. Whereas *mag*, the type element of this clade, resides in the genome of the silk moth, *B. mori* (Garel et al., 1994), the *fugitive* is more closely related in sequence and structure to the *Gulliver* LTR retrotransposon from the Asian schistosome, *S. japonicum* than to *mag*. *Gulliver* and the *fugitive* may have been transmitted vertically from a common ancestral retrotransposon that colonised the genome of a progenitor schistosome. The current findings obtained using the bioinformatics and Southern hybridisation approaches suggested that there might be as many as 2000 copies of the *fugitive* interspersed throughout the genome of

Table 1
Estimation of gene copy number of the *fugitive* LTR retrotransposon in the genome of *Schistosoma mansoni* by comparison of BLAST results of the GSS sequences in the GenBank database for a number of other *S. mansoni* genes for which copy numbers have been reported

Gene	GenBank accession	Length (bp)	Number of hits	*E* value	Estimated copy number	Key citation
SR1	U66331	2337	104 (score > 887)	0.0	200–2000	Drew and Brindley (1997)
SR2	AF025672	3913	102 (score > 1288)	0.0	1000–10000	Drew et al. (1999)
SM-α	M27676	331	162 (score > 264)	7×10^{-23}	7000–10000	Spotila et al. (1989)
Cathepsin D cDNA	U60995	1285	0	–	1	Morales et al. (2004)
Boudicca	AY662653	5858	100	0.0	2000–3000	Copeland et al. (2003)
Boudicca LTR	BK000439	324	116	9×10^{-53}	> 3000	Copeland et al. (2003)
SL RNA	M34074	595	0	–	100–200	Rajkovic et al. (1990)
18S rRNA	M62652	1739	2	1×10^{-26}	100	Simpson et al. (1984)
Perere	BK004067	4875	136 (> score 355)	1×10^{-96}	250–2500	DeMarco et al. (2004)
Saci-1	BK004068	5980	66	1×10^{-19}	70–700	DeMarco et al. (2004)
Saci-2	BK004069	4946	58	5×10^{-25}	85–850	DeMarco et al. (2004)
Saci-3	BK004070	5217	100 (> 186)	7×10^{-46}	150–1500	DeMarco et al. (2004)
fugitive	BK005225	4811	190	$\times10^{-121}$	> 2000	This study
fugitive LTR	BK005225	273	106 (> 127)	3×10^{-29}	> 3000	This study

The advanced search option for Blastn with the expect value changed from 10 (default) to 0.1 was employed in this analysis, and the search was limited by an Entrez query ⟨Schistosoma mansoni[organism]⟩. The number of Blast hits with a score 100 or higher is presented, along with the *E* value of the weakest matching hit.

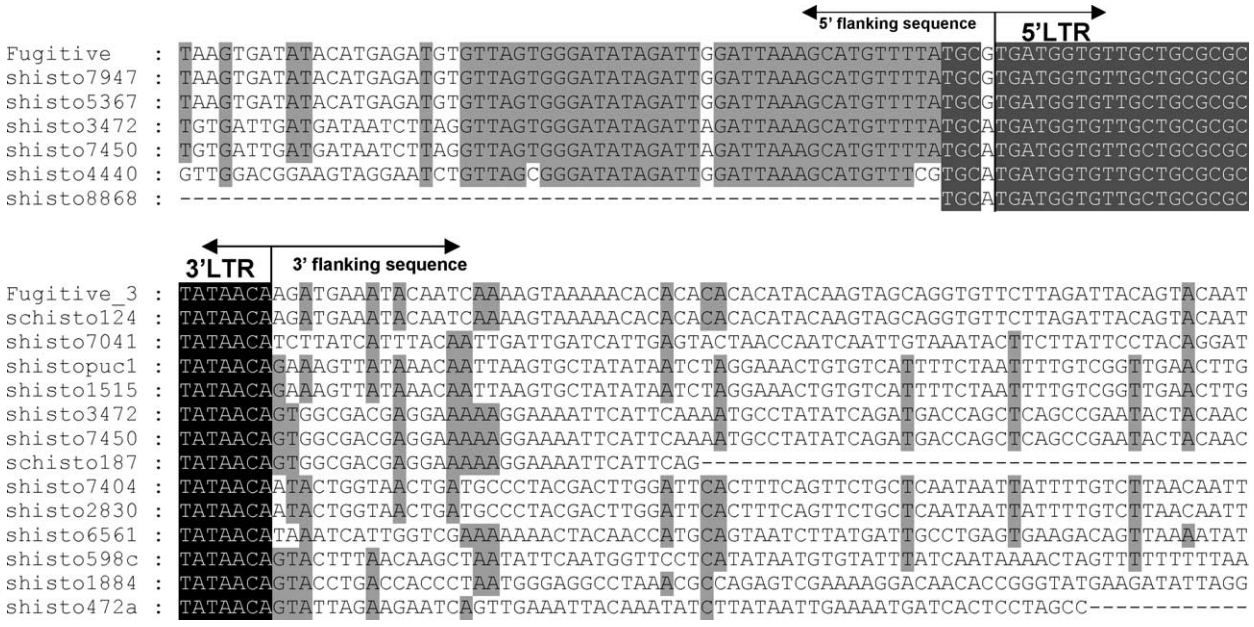

Fig. 5. Sequences flanking the 5′- and 3′- LTRs of the *fugitive* retrotransposon from BAC clone 49_J_14 aligned with a number of BAC-end sequences from randomly selected genomic clones from the *Schistosoma mansoni* genome. Conservation of residues is indicated by the shading of boxes.

S. mansoni. If so, given its length of 4811 bp and given that the size of the haploid genome of *S. mansoni* is estimated at 270 Mbp, the *fugitive* may comprise as much as 4% of the genome of *S. mansoni*.

The proviral form of the *fugitive* characterised from BAC 49_J_14 appears to be located within chromosome number 1, based on the annotation provided in GenBank accession number AC093105 (El-Sayed et al., 2004). Chromosome 1, an autosome, is one of eight chromosomes that constitute the haploid genome of *S. mansoni*. In addition to the *fugitive*, several non-LTR retrotransposons were located in BAC 49_J_14, in relatively close proximity to the *fugitive* (Fig. 1). Retrotransposons often target other retrotransposons as preferred insertion targets (SanMiguel et al., 1996). The smeared pattern of hybridisation seen in the Southern hybridisation of the *fugitive*-specific probe to genomic DNA of *S. mansoni* suggested that the copies of the retrotransposon were interspersed throughout the genome rather than being localised at a specific site. In addition to targeting sites rich in other retrotransposons, the *fugitive* appeared to prefer integration into AT-rich loci. The general absence of Blast hits to query sequences comprised of the *fugitive*, its LTRs or its flanking sequences in BAC 49_J_14 suggested that the *fugitive* may prefer intergenic regions or gene poor locations. Other schistosome retrotransposons are known to target introns (Foulk et al., 2002; Morales et al., 2004). Mobilisation of these kinds of mobile sequences to intergenic regions and/or to introns rather than to exons likely carries much less danger from deleterious mutations to the host genome. Host genomes, including the schistosome genome, have adapted to the presence of mobile genetic elements in proximity to and indeed within host genes and appear to have incorporated these sequences into regulatory elements and other structures (e.g. Ackerman et al., 2002). As more copies of the *fugitive* and its integration sites are characterised and mapped, it will be of interest to determine whether similar assimilations of the *fugitive* into schistosome regulatory elements have taken place. Moreover, their presence in high copy number can be expected to have exerted an influence on evolution of the schistosome genome through the effects on unequal crossing over (homologous recombination; Kazazian, 2004).

The LTRs of the *fugitive* at 273 bp in length are substantially longer than the short LTRs of *mag* (59 bp long) but are similar in length to the LTRs of *Gulliver*, which are 259 bp in length. Both the 5′-LTR and 3′-LTR of the *fugitive* copy located in BAC 49_J_14 were identical in sequence. Along with conservation of most residues contributing to the active sites of the enzyme domains (e.g. RT, IN) of pol, these structural characteristics suggested that the BAC 49_J_14 copy of the *fugitive* had been transpositionally active in the recent past (Bowen and McDonald, 1999). Several other features also indicated that the *fugitive* is transpositionally active. First, numerous ESTs (transcripts) spanning most of the enzymatic domains and LTRs of the *fugitive*, and from at least six developmental stages of *S. mansoni*, have been sequenced (Supplementary Table 2; Verjovski-Almeida et al., 2003). Second, the LTRs of the *fugitive* contain a putative promoter for initiation of transcription by RNA polymerase II, and the ORF between the terminal LTRs encodes gag and pol polyprotein domains essential for retrotransposition. Third, the *fugitive* exhibits structural and enzymatic domains required for retrotransposition. These include a dual Cys–His box motif as part of its nucleocapsid domain, a region of the gag polyprotein of retrotransposons and retroviruses that interacts with

the viral RNA genome through a zinc finger (e.g. Scarlata and Carter, 2003). Interestingly, whereas most *gypsy*/*Ty3*-like LTR retrotransposons include just a single Cys–His box in this domain, the *fugitive* exhibited two Cys–His boxes in like fashion to the closely related elements, *Gulliver* and *mag* (Garel et al., 1994; Laha et al., 2001). The RT domain of the *fugitive* included the highly conserved seven boxes of residues that characterise RT of retroviruses and LTR retrotransposons; the RT domain of non-LTR retrotransposons is more extensive (Xiong and Eickbush, 1990; Malik et al., 1999). The RNaseH domain of the *fugitive* included hallmark motifs of *mag*-like elements and the IN domain included the conserved CCHC and DD(35)E motifs characteristic of integrase enzymes (Haren et al., 1999). By contrast, the presence of several potentially inactivating mutations, including a stop codon within the gag-encoding domain and the presence of DTA rather than DT/SG at the protease active site, suggested that the copy of this element within BAC 49_J_14 may not be capable of autonomous retrotransposition.

There remains a pressing need to determine the role and importance of the new genes being sequenced at rapid pace from genome projects targeting the human schistosomes (Hu et al., 2003; Verjovski-Almeida et al., 2003). Many of the newly reported protein-encoding sequences may represent valuable intervention targets. Transgenesis techniques including insertional mutagenesis represent a rational approach to determine whether a specific gene plays an essential function in parasite development and longevity, as do approaches using double stranded RNAs for knock down of specific gene expression (Boyle and Yoshino, 2003; Heyers et al., 2003; Skelly et al., 2003). Harnessing endogenous mobile genetic elements for these kinds of applications should obviate difficulties with host specificity for the gene transfer vectors and thereby enhance prospects for successful introduction and integration of transgenes. Indeed, in like fashion to the situation in *S. cerevisiae* where the endogenous retrotransposon *Ty1* has been successfully employed in insertional mutagenesis (Garraway et al., 1997), it may be feasible to harness an active form of the *fugitive* or other schistosome retrotransposon for transgenesis and insertional mutagenesis of the genome of *S. mansoni*.

Acknowledgements

We thank Mary Duke for maintenance of the schistosome life cycle. This investigation received financial support from the UNICEP/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) (ID A20723) and an Infrastructure Grant from the Ellison Medical Foundation (award no. ID-IA-0037-02). PJB is a recipient of a Burroughs Wellcome Fund scholar award in Molecular Parasitology and AL is a recipient of an R. Douglas Wright Biomedical Career Development Award from the National Health and Medical Research Council of Australia.

Appendix. Supplementary Material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpara.2004.08.007

References

Abe, H., Ohbayashi, F., Sugasaki, T., Kanehara, M., Terada, T., Shimada, T., Kawai, S., Mita, K., Kanamori, Y., Yamamoto, M.T., Oshiki, T., 2001. Two novel *Pao*-like retrotransposons (*Kamikaze* and *Yamato*) from the silkworm species *Bombyx mori* and *B. mandarina*: common structural features of *Pao*-like elements. Mol. Genet. Genomics 265, 375–385.

Ackerman, H., Udalova, I., Hull, J., Kwiatkowski, D., 2002. Evolution of a polymorphic regulatory element in interferon-gamma through transposition and mutation. Mol. Biol. Evol. 19, 884–890.

Boeke, J.D., Corces, V.G., 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. 43, 403–434.

Bowen, N.J., McDonald, J.G., 1999. Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retrovirus-like elements. Genome Res. 9, 924–935.

Boyle, J.P., Yoshino, T.P., 2003. Gene manipulation in parasitic helminths. Int. J. Parasitol. 33, 1259–1268.

Brindley, P.J., Laha, T., McManus, D.P., Loukas, A., 2003. Mobile genetic elements colonizing the genomes of metazoan parasites. Trends Parasitol. 19, 79–87.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., Kazazian Jr.., H.H., 2002. Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am. J. Hum. Genet. 71, 327–336.

Capy, P., Vitalis, R., Langin, T., Higuet, D., Bazin, C., 1996. Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor?. J. Mol. Evol. 42, 359–368.

Copeland, C.S., Brindley, P.J., Heyers, O., Michael, S.F., Johnston, D.A., Williams, D.L., Ivens, A.C., Kalinna, B.H., 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. J. Virol. 77, 6153–6166.

Copeland, C.S., Heyers, O., Kalinna, B.H., Bachmair, A., Stadler, P.F., Hofacker, I.L., Brindley, P.J., 2004. Structural and evolutionary analysis of the transcribed sequence of Boudicca, a *Schistosoma mansoni* retrotransposon. Gene 329, 103–114.

DeMarco, R., Kowaltowski, A.T., Machado, A.A., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M., Wilson, R.A., Menck, C.F., Setubal, J.C., Dias-Neto, E., Leite, L.C., Verjovski-Almeida, S., 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.

Drew, A.C., Brindley, P.J., 1997. A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. Mol. Biol. Evol. 14, 602–610.

Drew, A.C., Minchella, D.J., King, L.T., Rollinson, D., Brindley, P.J., 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.

El-Sayed, N.M., Bartholomeu, D., Ivens, A., Johnston, D.A., LoVerde, P.T., 2004. Advances in schistosome genomics. Trends Parasitol. 20, 154–157.

Engels, D., Chitsulo, L., Montresor, A., Savioli, L., 2002. The global epidemiological situation of schistosomiasis and new approaches to control and research. Acta Trop. 82, 139–146.

Foulk, B.W., Pappas, G., Hirai, Y., Hirai, H., Williams, D.L., 2002. Adenylosuccinate lyase of *Schistosoma mansoni*: gene structure, mRNA expression, and analysis of the predicted peptide structure of a potential chemotherapeutic target. Int. J. Parasitol. 32, 1487–1495.

Garel, A., Nony, P., Prudhomme, J.C., 1994. Structural features of *mag*, a gypsy-like retrotransposon of *Bombyx mori*, with unusual short terminal repeats. Genetica 93, 125–137.

Garraway, L.A., Tosi, L.R., Wang, Y., Moore, J.B., Dobson, D.E., Beverley, S.M., 1997. Insertional mutagenesis by a modified in vitro Ty1 transposition system. Gene 198, 27–35.

Gonsky, J., Bacharach, E., Goff, S.P., 2001. Identification of residues of the Moloney murine leukemia virus nucleocapsid critical for viral DNA synthesis in vivo. J. Virol. 75, 2616–2626.

Hall, T., 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Haren, L., Ton-Hoang, B., Chandler, M., 1999. Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. 53, 245–281.

Heyers, O., Walduck, A.K., Brindley, P.J., Bleiss, W., Lucius, R., Dorbic, T., Wittig, B., Kalinna, B.H., 2003. *Schistosoma mansoni* miracidia transformed by particle bombardment infect *Biomphalaria glabrata* snails and develop into transgenic sporocysts. Exp. Parasitol. 105, 174–178.

Hu, W., Yan, Q., Shen, D.K., Liu, F., Zhu, Z.D., Song, H.D., Xu, X.R., Wang, Z.J., Rong, Y.P., Zeng, L.C., Wu, J., Zhang, X., Wang, J.J., Xu, X.N., Wang, S.Y., Fu, G., Zhang, X.L., Wang, Z.Q., Brindley, P.J., McManus, D.P., Xue, C.L., Feng, Z., Chen, Z., Han, Z.G., 2003. Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. Nat. Genet. 35, 139–147.

Ivics, Z., Hackett, P.B., Plasterk, R.H., Izsvak, Z., 1997. Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91, 501–510.

Kazazian Jr.., H.H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626–1632.

Laha, T., Loukas, A., Verity, C.K., McManus, D.P., Brindley, P.J., 2001. *Gulliver*, a long terminal repeat retrotransposon from the genome of the oriental blood fluke *Schistosoma japonicum*. Gene 264, 59–68.

Le Paslier, M.C., Pierce, R.J., Merlin, F., Hirai, H., Wu, W., Williams, D.L., Johnston, D., LoVerde, P.T., Le Paslier, D., 2000. Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. Genomics 65, 87–94.

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

Meric, C., Goff, S.P., 1989. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys–His box of the nucleocapsid protein. J. Virol. 63, 1558–1568.

Morales, M.E., Kalinna, B.H., Heyers, O., Schulmeister, A., Mann, V.H., Copeland, C.S., Loukas, A., Brindley, P.J., 2004. Genomic organization of the *Schistosoma mansoni* aspartic protease gene, a platyhelminth orthologue of mammalian lysosomal cathepsin D. Gene 338, 99–109.

Pearl, L.H., Taylor, W.R., 1987. A structural model for the retroviral proteases. Nature 329, 351–354.

Rajkovic, A., Davis, R.E., Simonsen, N., Rottman, F.M., 1990. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. Proc. Natl Acad. Sci. USA 87, 8879–8883.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274, 765–768.

Scarlata, S., Carter, C., 2003. Role of HIV-1 Gag domains in viral assembly. Biochim. Biophys. Acta 1614, 62–72.

Sijen, T., Plasterk, R.H., 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature 426, 310–314.

Simpson, A.J., Dame, J.B., Lewis, F.A., McCutchan, T.F., 1984. The arrangement of ribosomal RNA genes in *Schistosoma mansoni*. Identification of Polymorphic structural variants. Eur. J. Biochem. 139, 41–45.

Skelly, P.J., Da'dara, A., Harn, D.A., 2003. Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. Int. J. Parasitol. 33, 363–369.

Spotila, L.D., Hirai, H., Rekosh, D.M., LoVerde, P.T., 1989. A retroposon-like short repetitive DNA element in the genome of the human blood fluke, *Schistosoma mansoni*. Chromosoma 97, 421–428.

Springer, M.S., Britten, R.J., 1993. Phylogenetic relationships of reverse transcriptase and RNaseH sequences and aspects of genome structure in the gypsy group of retrotransposons. Mol. Biol. Evol. 10, 1370–1379.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., Waterston, R.H., 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1, E45.

Teysset, L., Burns, J.C., Shike, H., Sullivan, B.L., Bucheton, A., Terzian, C., 1998. A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral gypsy envelope can infect *Drosophila* cells. J. Virol. 72, 853–856.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Verjovski-Almeida, S., DeMarco, R., Martins, E.A., Guimaraes, P.E., Ojopi, E.P., Paquola, A.C., Piazza, J.P., Nishiyama Jr.., M.Y., Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F., Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L., Leite, R.A., Malaquias, L.C., Marques, R.C., Miyasato, P.A., Nascimento, A.L., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A., Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M., Wilson, R.A., Menck, C.F., Setubal, J.C., Leite, L.C., Dias-Neto, E., 2003. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. Nat. Genet. 35, 148–157.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.