Identification and characterization of endogenous viral elements for the three key schistosomes of humans

Na Li and Quhuan Li*
School of Bioscience and Engineering, South China University of Technology, Guangzhou Guangdong, China

Abstract: Endogenous viral elements (EVEs) are widely distributed throughout eukaryotic genomes, and their evolution and potential function have attracted a lot of interest. Draft genome sequences for *Schistosoma mansoni*, *Schistosoma japonicum* and *Schistosoma haematobium* are now available; however, information about EVEs in blood flukes of the genus schistosoma is scanty. Here, genome-wide survey into the putative EVE sequences of the three key schistosome genomes were present. Totally 4, 117 gene sequences were identified, including retrovirus-like *gypsy* elements, RNA viruses and dsDNA viruses. Compared with *S. japonicum* and *S. haematobium*, *S. mansoni* appeared to greatly outnumbered by *gypsy* members. Phylogenetic analysis revealed one novel endogenous retrovirus element in *S. mansoni*. This initial characterization of schistosomes showed that schistosomes harbour distinct EVEs that may have played an important evolutionary role. Studies of schistosomes' endogenous viruses helped us to glance at an earlier viral event in the class *Trematoda*, greatly broadening the field of palaeovirology.

Keywords: Endogenous viral elements; *Schistosoma mansoni*; *Schistosoma japonicum*; *Schistosoma haematobium*.

INTRODUCTION

Endogenous viral elements (EVEs) are copies or remnant of exogenous viruses that were once inserted into the host germ line cells and have become heritable as host alleles (Benveniste and Todaro, 1974, Holmes, 2011, Jaenisch, 1976). Over millions of years of evolution, most of EVEs have accumulated random mutations, eventually losing their ability to replicate, and are thought to have little effect on existing the hosts' gene expression. However, as long as exogenous viruses have become endogenous, they behave like any other genetic element, so they are subjected to selection, mutation, genetic drift and do co-evolve with the host genomes (Jern and Coffin, 2008). Hence, these genomic 'fossils' not only can offer new insights into the origin and evolutionary dynamics between viruses and their hosts, but also contribute to better understand host genome structure and function, including mutation, development and physiological effects (Feschotte and Gilbert, 2012, Jern and Coffin, 2008).

EVEs are widely distributed throughout the genomes of various eukaryotes, especially endogenous retroviral elements (the majority of EVEs that are recognized till today). More surprising were not just retroviruses but also non-retroviruses that can integrate and be endogenous (Crochu *et al.*, 2004, Holmes, 2011, Horie *et al.*, 2010, Maori *et al.*, 2007). Interestingly, many genomes contain mobile sequences such as LTR retrotransposons or viral retrotransposons, sharing similarities with integrated proviruses of retroviruses (Kim *et al.*, 1994). For instance, several lines of evidence have indicated that *gypsy* retrotransposons are the endogenous retrovirus of *Drosophila* species (Kim *et al.*, 1994, Llorens *et al.*, 2008, Mejlumian *et al.*, 2002, Pelisson *et al.*, 1994). Recently, these elements have been listed as *Metaviridae* family in the International Committee on Taxonomy of Viruses (ICTV) (Fauquet and Mayo, 2001). According to a hypothesis is that retroviruses might be evolved from *gypsy* family after possessing an envelope protein and adopting a viral lifestyle (Bucheton, 1995). Therefore, they are interesting for their ability to facilitate horizontal transmission and roles in evolutionary development. The term EVEs in this study will be used to refer to all endogenous virus-like elements, including DNA viruses, RNA viruses as well as retrovirus-like *gypsy* retrotransposons (including eight representative types *Boudicca*, *Scai-2*, *Scai-3*, *Scai-5*, *Scai-6*, *Scai-7*, *NoNaut-3* and *NoNaut-5*).

Schistosomes have come into the spotlight (organisms) for they can feed on blood and globulins and then cause a high-mortality disease, schistosomiasis (Gryseels *et al.*, 2006, Steinmann *et al.*, 2006). To date no vaccines are available and 'Praziquantel' is the only drug being used to treat this disease (Doenhoff *et al.*, 2008). Recent progress has shown that retroviral-mediated transduction offers a way to establish transgenic lines of schistosomes and horizontal gene transfer between host and parasite mediated by retroviruses (Kines *et al.*, 2006, Kines *et al.*, 2008). Furthermore, some researchers have proved that schistosomal liver disease is significantly associated with hepatitis C and HIV infection as well as cancer (Gad *et al.*, 2001, Ishii *et al.*, 1994, Ndeffo Mbah *et al.*, 2013). These results indicate that schistosomes have some interaction with viral infection. The characterization of endogenous viral elements in the schistosome genomes

**Corresponding author:* e-mail: liqh@scut.edu.cn

compared with molecular conservation can offer the source about host-parasite coevolution (Damian, 1997). However, to date only some retroviruses-like *gypsy* elements have been characterized in the schistosomes (Copeland *et al.*, 2006, DeMarco *et al.*, 2004).

Three principal disease-causing species include *S. mansoni*, *S. japonicum* and *S. haematobium*. Despite the availability of these three genome sequences since 2012 (Berriman *et al.*, 2009, Young *et al.*, 2012,Zhou *et al.*, 2009), the diversity of EVEs in bilharzial genomes has not been systematically explored. In this work we applied two search algorithms, LTR_STRUC (McCarthy and McDonald, 2003) and TBLASTN algorithm to detect EVEs in schistosome genome to provide an important source of information about EVEs in invertebrates. This combination strategy is benefited for comprehensive understanding of generating results than does any single program. The aims of this work included (i) to genome-wide detect and characterize EVEs that may be present in the schistosome genome, (ii) extending the likely host range of virus groups, (iii) to compare the detection results among the three key of schistosome species that can infect human beings.

MATERIALS AND METHODS

Detect strategy

The *S. mansoni* (v5.0) genome, ftp.sanger.ac.uk/pub4/pathogens/Schistosoma/mansoni/genome/Assembly-latest/; the *S. japonicum* genome, http://www.chgc.sh.cn/japonicum/Resources.html; the complete *S. haematobium* genome sequences were retrieved from BGI-Shenzhen. The viral genomes, ftp.ncbi.nlm.nih.gov/genomes/Viruses/.

LTR _STRUC and TBLASTN strategies were applied to detect the EVEs in schistosome genomes. The first strategy was used to detect LTR elements based on the LTR retrotransposons search algorithm. Briefly, LTR_STRUC identifies putative LTR-LTR elements and then looks for other defining features of LTR elements, such as primer binding site (PBS), polypurine tract (PPT), and target site duplications (TSDs). Lastly, the software assigns a final score to the successful match according to the presence or absence of these features. This program has proven effective in many genome sequences (for example chimpanzee (Polavarapu *et al.*, 2006). In the present work, default parameters were used. After running the program, the LTR-LTR full elements were obtained and further combined with conventional BLAST (e-values $<1\times10^{-8}$) analysis to get members that belong to *gypsy*-like family and ERVs.

In the second strategy was based on the similarity of sequences (for example horse (van der Kuyl, 2011)). All virus genomes were obtained that had been sequenced from NCBI and added several representative *gypsy* peptide sequences to construct an intact library of viral peptide sequences as search query sequences (Sup 1). Sequences that matched viral peptides in the query library with e-values 1×10^{-8} were extracted and were inferred through two protein databases, Nr and PFAM. For each genomic segment, putative EVE element was the best hit by removing the redundancy and getting a purified block. The sequences were assigned to taxonomic family or genus based on the most similar exogenous viral sequences.

Phylogenetic analysis

Putative EVE sequences were extracted and aligned with closely related viruses using the program ClustalW. MEGA version 4 for phylogenetic inference. Neighbor-joining (NJ) phylogenies were estimated using amino acid sequence alignments. Bootstrap analysis was carried out with 1,000 replicates.

Sequences and accession number

Human adenovirus B1, NC_011203.1; Simian adenovirus 3, NC_006144.1; Bovine adenovirus A, NC_006324.1; Canine adenovirus 1, AC_000003.1; Fowl adenovirus A , AC_000014.1; Frog adenovirus 1, NC_002501.1; Ovine adenovirus 7, U40839.3; Ovine adenovirus D, NC_004037.2; Simian retrovirus 4 , NC_014474.1; Human T-lymphotropic virus 1, NC_001436.1; Feline foamy virus, NC_001871.1; Saci-7, BN000785.1; Boudicca, BK004066.1; Saci-3, BK004070.1; Mason-Pfizer monkey virus, NC_001550.1; Rice tungro bacilliform virus, NC_001914.1; Dracaena mottle virus, NC_008034.1; Bougainvillea spectabilis chlorotic vein-banding virus, NC_011592.1; Sugarcane bacilliform virus, NC_013455.1.

RESULTS

Genome screening

A total of 4,117 nonredundant putative EVEs were detected in the three-schistosome genomes by two methods (table 1). Most of these were identified by just one program: 1,416 by LTR_STRUC and 2,805 by TBLASTN-based search. Only 104 EVEs were detected by both methods. These elements were almost *gypsy*-related sequences. One certain reason is that retroviruses are largely restricted to vertebrates, only some elements of the *gypsy* family are able to infect new individuals in invertebrates (Wicker *et al.*, 2007).

By LTR_STRUC program directly, putative 74 LTR-LTR elements were detected in *S. haematobium*, 77 were detected in *S. japonicum*. Surprisingly, up to 758 LTR_LTR elements were detected in *S. mansoni* that almost 10 fold compared to the other two schistosome genomes. Among all detected EVEs, the *S. mansoni* genome contained the most LTR-RT-LTR structural EVEs (197), followed by the *S. japonicum* (17) and *S. haematobium* (14) genomes. These elements reported by

LTR_STRUC were subjected to sequence analysis by TBLASTN to identify the *gypsy* family and ERVs. Up to 1,358 sequences were obtained in *S. mansoni*, while only 11 hits were identified in *S. haematobium* and 47 elements were found in *S. japonicum*.

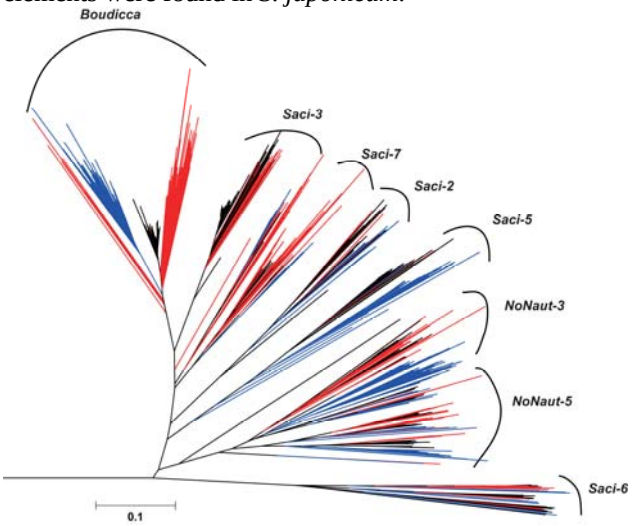

Fig. 1: Phylogenetic tree of gypsy family in the three schistosome genomes. The EVEs in *S. haematobium*, *S. japonicum* and *S. mansoni* are represented in red, blue and black respectively.

By TBLASTN-based method alone, there were 610 elements identified in *S. haematobium*, 726 in *S. japonicum* and 1,469 in *S. mansoni*. The *S. mansoni* genome also contained the most retrovirus-like *gypsy* EVEs. Although the genome sizes among them were similar, but the number of putative EVEs detected in *S. mansoni* were greatly outnumbered by the genomes of both *S. haematobium* and *S. japonicum*. It was therefore deduced that it could be due to the limitations of the detection method or stem from characteristics of the *S. mansoni* genome.

EVEs related to the retrovirus-like gypsy family

We identified numerous EVEs matched to some representative *gypsy* members (table 1). By LTR_STRUC program, less *gypsy*-like elements were identified compared with TBLASTN-based method, especially in *S. haematobium* and *S. japonicum*. The *gypsy* family identified by LTR_STRUC had 360 full-length *gypsy*-like elements (at least including *gag* and *pol*): 6 for *S. haematobium*, 23 for *S. japonicum* and 333 for *S. mansoni* (Sup 2). Among them, eight kinds of *gypsy* members all were presenting in *S. mansoni*; four kinds (*Saci-2*, *Saci-6*, *Saci-7* and *Nonaut-3*) were displayed in *S. haematobium*; three kinds (*Saci-5*, *Saci-6* and *Nonaut-5*) were existed in *S. japonicum*.

According to TBLASTN-based method, the results showed that *Boudicca* caused the highest gene copy number among the three genomes while *Saci-3* and *Saci-7* displayed intermediate copy numbers, other families displayed relatively low copy numbers. *Boudicca* may have a significant impact on gene and genome evolution in *Schistosome* species. A consensus constructed using the most conserved *pol* gene sequences (>80% coverage) that were identified in these three schistosome genomes by Neighbor-Joining phylogenies (fig. 1). In phylogeny, EVEs derived from *Boudicca* members grouped into four well-supported clades, strongly suggesting for the two distinct expansions in *S. haematobium* and one expansion in the other two schistosome genomes, respectively. EVEs derived from other members of the *gypsy* family, formed a robustly but not clearly supported cluster in phylogenies with the other two schistosome species, indicating that they were derived from a common ancestor.

All members of *gypsy* family in this study have been previously characterized in *S. mansoni* and members of *Boudicca* also exist in the genome of *S. haematobium* (Copeland *et al.*, 2003, Copeland *et al.*, 2006, DeMarco *et al.*, 2004). In this study, it was shown that all members existed in these three major species of schistosomes. What's more, some of them presented as full-length elements including LTR-RT-LTR. Taken together, this findings provided evidence for the first time that all eight kinds *gypsy* family are also be presented in the other two schistosome genomes.

EVEs related to viruses with RNA viruses

Several sequences that highly matched (e-values<1×10^{-8}) to RNA viruses in the genomes of schistosoma were identified, all derived from the *Retroviridae* family (Sup 3). An EVE representing ~83% of the *gap-pro-pol* gene of Mason-Pfizer monkey virus was identified in *S. mansoni* (fig. 2A). To further elucidate the relationship between the EVE and other viruses, phylogenetic analysis was made using the most conserved *pol* gene sequences. The phylogenetic tree showed that this EVE was chosen to represent well-supported clades with retroviruses but not *gypsy* family (fig. 2B). The EVE also displayed the LPQG and YMDD motifs, that generally were conserved in the representatives of Class II families of ERVs (Xiong and Eickbush, 1990). Therefore, this EVE was indeed an endogenous retrovirus for these decisive evidences. Due to lacking LTR sequences in both ends, the integration time based on sequence divergence could not be able to estimate; however, this finding might shed new light on the rooted history and diversity of retroviruses.

EVEs related to viruses with DNA viruses

Several numbers of EVEs that matched to several families of viruses with DNA genomes were identified in the three genomes, and all these belonged to dsDNA virus families (Sup 4). It is intriguing that matches to the genus *Mastadenovirus* of the *Adenoviridae* virus family with high similarity (>96% identity) were identified in *S. japonicum* genome (fig 2C). *Mastadenoviruses* can infect mammals only and have not been reported in other

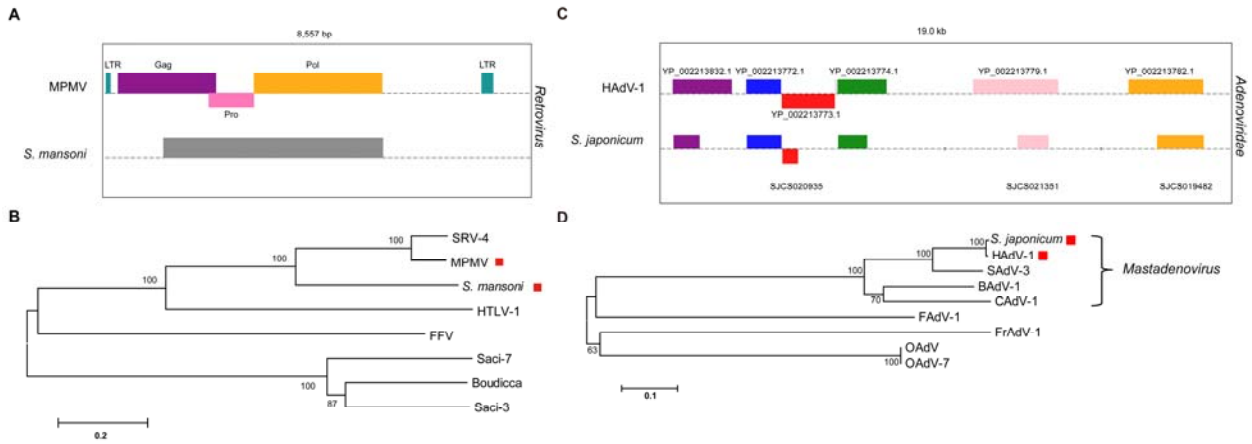

Fig. 2: Genetic structure and phylogeny tree of EVE. (A) Genetic structures of the most intact retrovirus EVE shown relative to the most closely species. (B) Phylogeny tree between the EVE, retrovirus and gypsy. (C) Summary genetic structures of Mast adenovirus EVE sets shown relative to the most similar viruses. Vertical lines between the EVEs in the host species indicate that the EVEs are not contiguous in the host genome. (D) Phylogenetic relationship of the Mast adenovirus EVE and the representative exogenous viruses in the Adenoviridae family. Taxa that are shown as genetic structures in (A and C) are indicated by red squares. Abbreviations: MPMV=Mason-Pfizer monkey virus; SRV-4=Simian retrovirus 4; HTLV-1=Human T-lymphotropic virus 1; FFV=Feline foamy virus. HAdV-1=Human adenovirus B 1; SAdV-3=Simian adenovirus 3; BAdV-1=Bovine adenovirus A; CAdV-1=Canine adenovirus 1; FAdV-1=Fowl adenovirus A; FrAdV-1=Frog adenovirus 1; OAdV=Ovine adenovirus D; OAdV-1=Ovine adenovirus 7.

vertebrates. In *S. japonicum*, several EVEs related to Human adenovirus B1 in the three scaffolds were identified: SJCS020935, SJCS019482 and SJCS021351. Because most of sequences of these three scaffolds were matched, thus the three scaffolds should be assembled for a bigger scaffold. These elements seemed to be functional, given that all the elements contain no frame shift mutations and no internal stop codons. Phylogenies constructed using 52k protein sequences that conserved in all genera of the *Adenoviridae* family, this insertion clustered consistently with the human adenovirus (fig. 2D). Consequently, the single EVE in *S. japonicum* was indeed an endogenous mast adenovirus element.

EVEs related to a total of five viral families were also identified including *Baculoviridae*, *Caulimoviridae*, *Iridoviridae*, *Nimaviridae*, and *Poxviridae* (fig. 3). The majority of these matches comprised fragments of genes. More fragments of the gene EVEs that encode reverse transcriptase, related to the genus *Tungrovirus* of the *Caulimoviridae* family were identified in the *S. japonicum* and *S. mansoni* genomes, but not in *S. haematobium* genome. The host ranges of tungroviruses are narrow, restricted to monocotyledonous plants(Fauquet and Mayo, 2001). However, it was found that a total of 7 distinct EVEs related to the reverse transcriptase sequence of the genus *Tungrovirus* on 7 distinct loci. These EVEs shared about 30 amino acids similarity to tungroviruses. In phylogenies, EVEs derived from the genes of the *Caulimoviridae* family grouped into two well- supported clades (fig. 3F), one clade of which included exogenous *Caulimovirus* and the other clade was based on EVEs derived from the *S. japonicum* and *S. mansoni*. Weak support for relationships was obtained from the tree, making it difficult to confidently place the EVEs with respect to the known viruses. EVEs derived from the *S. japonicum* formed a robustly supported cluster in phylogenies with EVEs in *S. mansoni*, suggesting that they are derived from the same exogenous lineage.

DISCUSSION

The objective of this work were to systematically screen and characterize sequences occur as endogenous viral elements in the three main schistosomes that infect human beings. For detection, two programs were applied, TBLASTN and LTR_STRUC, to generate comprehensive results about EVEs in schistosome genomes. The first approach was structure-based method that depended on detecting LTR-LTR structures, while the other approach was homology-based method that relied on query sequences. Many of the EVEs identified by these two methods were highly mutated or fragmented. For example, seven distinct endogenous viral fragment elements shared only 30% amino acids similarity with *Caulimoviridae* family.

As expected, these two methods generated different information about EVEs. Much less EVEs by LTR_STRUC program were identified compared with TBLASTN-based method, especially the *gypsy*-like elements in *S. haematobium* and *S. japonicum* genomes. Since LTR_STRUC detected specific models of LTR architecture, it was possible that most elements detected

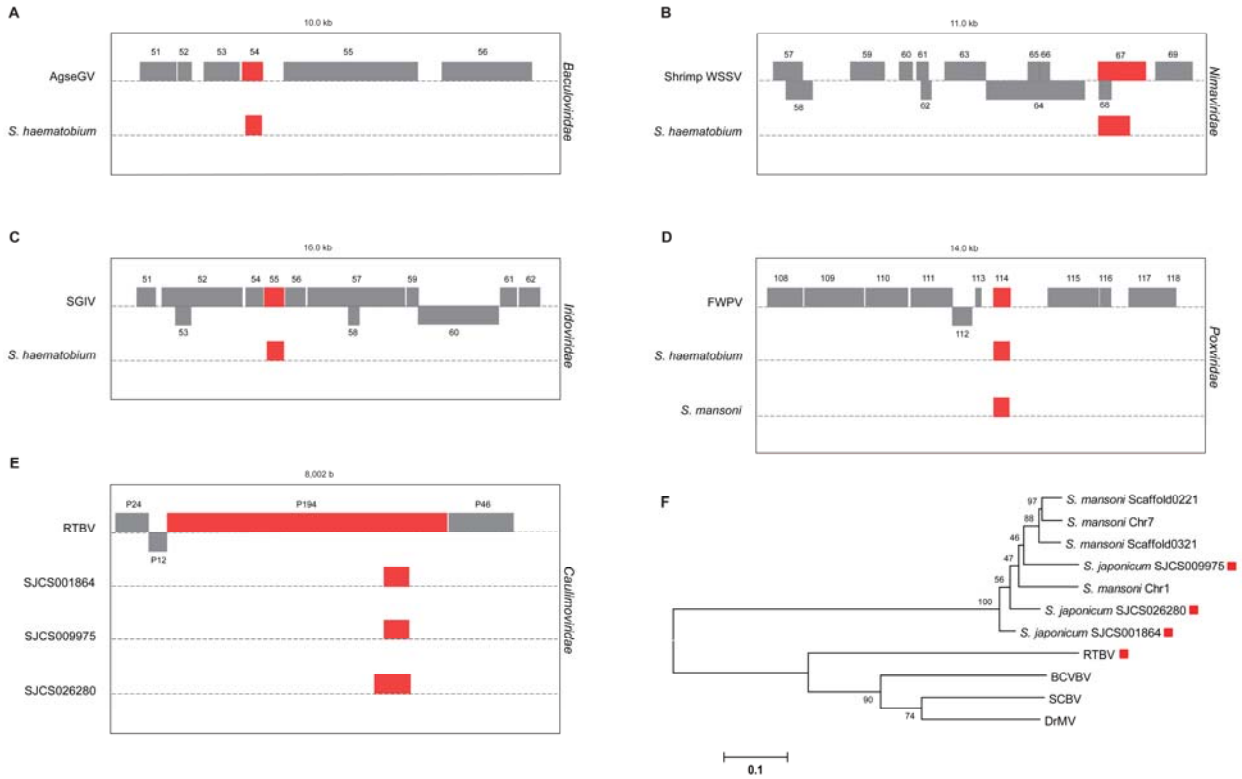

Fig. 3: Genetic structure of EVEs related to other dsDNA viruses and phylogenetic relationship of Caulimovirus EVEs. Summary genetic structures of EVEs derived from dsDNA viruses (A) Baculoviridae, (B) Nimaviridae, (C) Iridoviridae, (D) Poxviridae and (E) Caulimoviridae shown relative to the most similar viruses. Because of the large size, only parts of the genome sequence are shown. Neighbor-joining phylogenies of Caulimovirus EVEs and exogenous virus are shown (F). Red squares indicated taxa that are shown as genetic structures in (E). Abbreviations: RTBV=Rice tungro bacilliform virus; DrMV=Dracaena mottle virus; SCBV=Sugarcane bacilliform virus; BCVBV=Bougainvillea spectabilis chlorotic vein-banding virus; AgseGV=Agrotis segetum granulovirus; Shrimp WSSV=Shrimp white spot syndrome virus; SGIV=Singapore grouper iridovirus; FWPV=Fowlpox virus.

by TBLASTN method were solo LTRs or fragments. One hypothesis for less EVEs in *S. haematobium* and *S. japonicum* genomes is that more recombination events between LTRs occurred in these genomes, thereby causing a lower density of full *gypsy*-like elements (Garcia-Etxebarria and Jugo, 2012).

This study for the first time describes endogenous elements not only including retroviral elements but also non-retroviral elements such as DNA viruses and *gypsy* elements in the three schistosomes. For example, it was shown the first evidence that eight representative *gypsy* family members are present in the two other schistosome genomes apart from the *S. mansoni* genome. Furthermore, an EVE highly matched to RNA viruses in the genomes of *S. mansoni* (query=1,771 amino acid *gap-pro-pol*, best hit 83% identities/1449 amino acid fragment on Scaffold0908), and several dsDNA virus-like elements also existed in these genomes. Phylogenetic analysis of previous studies strongly supported that *S. mansoni* and *S. haematobium* form a clade relative to the *S. japonicum* species (Bowles *et al.*, 1995, Despres *et al.*, 1992, Rollinson *et al.*, 1997). Assuming that EVE occurs randomly, the presence of endogenous RNA-like element only in *S. mansoni* indicated that this insertion occurred after the divergence between *S. mansoni* and *S. haematobium*. The high level of identity (~83%) raises the possibility that this insertion was formed recently and may not increase in frequency and reached fixation. These EVEs identified in this study indicated that the wide distribution of viruses and extended the host range to the genus *Schistosoma*.

Of those EVEs discovered, it was most unexpected that several EVEs that derived from the genus *Mast adenovirus* of the *Adenoviridae* virus family were found, which generally infect a number of mammalian species. There are three possible explanations for the detection of the special EVEs. Firstly, horizontal transfer must be taken into consideration, since the ability to horizontal transfer is reflected in the genome. It was reported that not only vertical but also horizontal transfer of the host sequences occurred in schistosomes (Imase *et al.*, 2000, Imase *et al.*, 2001, Imase *et al.*, 2003). Thus, these

Table 1: Distribution and diversity of EVEs identified by both methods

Family	Host range	*S. Haematobium*	*S. Japonicum*	*S. Mansoni*	All loci
Methods	LTR_STRUC				
Boudicca	I	1	7	322	330
Scai-2	I	2	5	35	42
Saci-3	I	2	0	434	436
Saci-5	I	0	7	120	127
Saci-6	I	2	3	9	14
Saci-7	I	3	1	291	295
NoNaut-3	I	1	11	74	86
NoNaut-5	I	0	13	69	82
Retroviridae	V	0	0	4	4
SUM	1416				
Methods	TBLASTN				
Boudicca	I	198	266	438	902
Scai-2	I	9	34	62	105
Saci-3	I	168	64	461	693
Saci-5	I	11	57	67	135
Saci-6	I	21	24	46	91
Saci-7	I	116	115	234	465
NoNaut-3	I	33	55	55	143
NoNaut-5	I	49	98	100	247
Adenoviridae	V	0	6	0	6
Baculoviridae	I	1	0	0	1
Caulimoviridae	P	0	3	4	7
Iridoviridae	V,I	1	0	0	1
Nimaviridae	I	1	0	0	1
Poxviridae	V,I	1	0	1	2
Retroviridae	V	0	4	1	5
Unclassifiable	N/A	1	0	0	1
SUM	2805				

Footnote: Insertions were regarded as endogenous viral elements (EVEs) if the species-specific gene was identified with e-value 1×10^{-8} by TBLASTN with a length more than 100 amino acids. The most closely related exogenous viral sequences in searches of PFAM and Nr databases were not assigned to taxonomic groups according to the book of Virus Taxonomy 9thed Report of the International Committee on Taxonomy of Viruses was defined as unclassifiable. Abbreviations of the virus hosts: Invertebrates, I; Vertebrates, V.

identified EVEs may derive from the gene communication between the parasites *S. japonicum* and snails or vertebrate host recently. Another explanation is that the *S. japonicum* genome may have been exposed to *mastadenovirus* infection. Under this circumstance, the discovery of mast adenovirus EVEs establishes that a distinct lineage of exogenous *mastadenovirus* may exist in the species outside the mammalians. The high level of identify of the *mastadenovirus* EVEs reflect recent insertion rather than a long co-evolve history within the host genome. The one last explanation is that something were wrong during extract the genomic DNA.

The discovery and phylogenetic analyses in this report have indicated that EVEs might be more broadly distributed than previously thought and shed additional light on the diversity of endogenous viral elements in class *Trematoda*. However, a better understanding of the process of co-evolve between viruses and schistosomes will require more information about EVEs from other genomes. The discovery may be expected to contribute to future comparative evolution studies.

ACKNOWLEDGEMENTS

We would like to gratefully acknowledge the support and technical guidance of BGI institution and South China Normal University. This work was supported by National Natural Science Foundation of China Grants (31200705) and Specialized Research Fund for the Doctoral Program of Higher Education Grant (20100172120042) and Fundamental Research Funds for the Central Universities (2014ZZ0051).

REFERENCES

Benveniste RE and Todaro GJ (1974). Evolution of C-type viral genes: Inheritance of exogenously acquired

viral genes. *Nature,* **252**: 456-459.

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, DeMarco R, Djikeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, Lacerda D, Macedo CD, McVeigh P, Ning Z, Oliveira G, Overington JP, Parkhill J, Pertea M, Pierce RJ, Protasio AV, Quail MA, Rajandream MA, Rogers J, Sajid M, Salzberg SL, Stanke M, Tivey AR, White O, Williams DL, Wortman J, Wu W, Zamanian M, Zerlotini A, Fraser-Liggett CM, Barrell BG and El-Sayed NM (2009). The genome of the blood fluke Schistosoma mansoni. *Nature,* **460**: 352-358.

Bowles J, Blair D and McManus D (1995). A molecular phylogeny of the human schistosomes. *Molecular Phylogenetics and Evolution,* **4**: 103-109.

Bucheton A (1995). The relationship between the flamenco gene and gypsy in Drosophila: How to tame a retrovirus. *Trends in Genetics*, **11**: 349-353.

Copeland CS, Brindley PJ, Heyers O, Michael SF, Johnston DA, Williams DL, Ivens AC and Kalinna BH (2003). Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke schistosoma mansoni. *Journal of Virology,* **77**: 6153-6166.

Copeland CS, Lewis FA and Brindley PJ (2006). Identification of the boudicca and sinbad retrotransposons in the genome of the human blood fluke Schistosoma haematobium. *Memórias do Instituto Oswaldo Cruz,* **101**: 565-571.

Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson JJ, de Micco P and de Lamballerie X (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *J. Gen. Virol.,* **85**: 1971-1980.

Damian R (1997). Parasite immune evasion and exploitation: reflections and projections. *Parasitology,* **115**: 169-175.

DeMarco R, Kowaltowski AT, Machado AA, Soares MB, Gargioni C, Kawano T, Rodrigues V, Madeira AMBN, Wilson RA, Menck CFM, Setubal JC, Dias-Neto E, Leite LCC and Verjovski-Almeida S (2004). Saci-1,-2 and -3 and perere four novel retrotransposons with high transcriptional activities from the human parasite schistosoma mansoni. *Journal of Virology,* **78**: 2967-2978.

Despres L, Imbert-Establet D, Combes C and Bonhomme F (1992). Molecular evidence linking hominid evolution to recent radiation of schistosomes (Platyhelminthes: Trematoda). *Molecular Phylogenetics and Evolution,* **1**: 295-304.

Doenhoff MJ, Cioli D and Utzinger J (2008). Praziquantel: Mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr. Opin. Infect. Dis.,* **21**: 659-667.

Fauquet C and Mayo M (2001). The 7th ICTV Report. *Archives of virology,* **146**: 189-194.

Feschotte C and Gilbert C (2012). Endogenous viruses: Insights into viral evolution and impact on host biology. *Nat. Rev. Genet,* **13**: 283-296.

Gad A, Tanaka E, Orii K, Rokuhara A, Nooman Z, Serwah AH, Shoair M, Yoshizawa K and Kiyosawa K (2001). Relationship between hepatitis C virus infection and schistosomal liver disease: Not simply an additive effect. *Journal of Gastroenterology,* **36**: 753-758.

Garcia-Etxebarria K and Jugo BM (2012). Detection and characterization of endogenous retroviruses in the horse genome by in silico analysis. *Virology,* **434**: 59-67.

Gryseels B, Polman K, Clerinx J and Kestens L (2006). Human schistosomiasis. *The Lancet,* **368**: 1106-1118.

Holmes EC (2011). The evolution of endogenous viral elements. *Cell Host Microbe,* **10**: 368-377.

Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojobori T, Coffin JM and Tomonaga K (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature,* **463**: 84-87.

Imase A, Kobayashi K, Ohmae H, Irie Y and Iwamura Y (2000). Horizontal and vertical transfer of mouse endogenous retroviral DNA sequences in schistosomes. *Parasitology,* **121**: 155-162.

Imase A, Kobayashi K, Ohmae H, Matsuda H and Iwamura Y (2001). Horizontal and vertical transmission of mouse class I MHC sequence in Schistosoma mansoni. *Parasitology,* **123**: 163-168.

Imase A, Matsuda H, Irie Y and Iwamura Y (2003). Existence of host DNA sequences in schistosomes-horizontal and vertical transmission. *Parasitology International,* **52**: 369-373.

Ishii A, Matsuoka H, Aji T, Ohta N, Arimoto S, Wataya Y and Hayatsu H (1994). Parasite infection and cancer: with special emphasis on Schistosoma japonicum infections (Trematoda). A review. *Mutation Research Fundamental and Molecular Mechanisms of Mutagenesis,* **305**: 273-281.

Jaenisch R (1976). Germ line integration and mendelian transmission of the exogenous moloney leukemia virus. *Proceedings of the National Academy of Sciences,* **73**: 1260-1264.

Jern P and Coffin JM (2008). Effects of retroviruses on host genome function. *Annu. Rev. Genet,* **42**: 709-732.

Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N and Bucheton A (1994). Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. *Proceedings of the National Academy of Sciences,* **91**: 1285-1289.

Kines KJ, Mann VH, Morales ME, Shelby BD, Kalinna BH, Gobert GN, Chirgwin SR and Brindley PJ (2006).

Transduction of schistosoma mansoni by vesicular stomatitis virus glycoprotein-pseudotyped Moloney murine leukemia retrovirus. *Exp. Parasitol.,* **112**: 209-220.

Kines KJ, Morales ME, Mann VH, Gobert GN and Brindley PJ (2008). Integration of reporter transgenes into Schistosoma mansoni chromosomes mediated by pseudotyped murine leukemia virus. *FASEB J,* **22**: 2936-2948.

Llorens JV, Clark JB, Martinez-Garay I, Soriano S, de Frutos R and Martinez-Sebastian MJ (2008). Gypsy endogenous retrovirus maintains potential infectivity in several species of Drosophilids. *BMC Evol. Biol.,* **8**: 302.

Maori E, Tanne E and Sela I (2007). Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology,* **362**: 342-349.

McCarthy EM and McDonald JF (2003). LTR_STRUC: A novel search and identification program for LTR retrotransposons. *Bioinformatics,* **19**: 362-367.

Mejlumian L, Pélisson A, Bucheton A and Terzian C (2002). Comparative and functional studies of drosophila species invasion by the gypsy endogenous retrovirus. *Genetics,* **160**: 201-209.

Ndeffo Mbah ML, Kjetland EF, Atkins KE, Poolman EM, Orenstein EW, Meyers LA, Townsend JP and Galvani AP (2013). Cost-effectiveness of a community-based intervention for reducing the transmission of Schistosoma haematobium and HIV in Africa. *Proc. Natl. Acad. Sci. USA,* **110**: 7952-7957.

Pelisson A, Song S, Prud'homme N, Smith P, Bucheton A and Corces V (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *The EMBO Journal,* **13**: 4401.

Polavarapu N, Bowen NJ and McDonald JF (2006). Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biol.,* **7**: R51.

Rollinson D, Kaukas A, Johnston DA, Simpson AJ and Tanaka M (1997). Some molecular insights into schistosome evolution. *International Journal for Parasitology,* **21**: 11-28.

Steinmann P, Keiser J, Bos R, Tanner M and Utzinger J (2006). Schistosomiasis and water resources development: Systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases,* **6**: 411-425.

Van der Kuyl AC (2011). Characterization of a full-length endogenous beta-retrovirus, EqERV-beta1, in the genome of the horse (*Equus caballus*). *Viruses,* **3**: 620-628.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M and Panaud O (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics,* **8**: 973-982.

Xiong Y and Eickbush TH (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal,* **9**: 3353.

Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS, Xu X, Chen F, Wu X, Zerlotini A, Oliveira G, Hofmann A, Zhang G, Fang X, Kang Y, Campbell BE, Loukas A, Ranganathan S, Rollinson D, Rinaldi G, Brindley PJ, Yang H, Wang J and Gasser RB (2012). Whole-genome sequence of Schistosoma haematobium. *Nat Genet,* **44**: 221-225.

Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, Huang Z, Zhang B, Huang W and Jin K (2009). The Schistosoma japonicum genome reveals features of host-parasite interplay. *Nature,* **460**: 345-351.