Open Journal of Genetics, 2013, 3, 141-158 **OJGen**
http://dx.doi.org/10.4236/ojgen.2013.33017 Published Online September 2013 (http://www.scirp.org/journal/ojgen/)

Characterization of randomly amplified polymorphic DNA (RAPD) fragments revealing clonal variability in cercariae of avian schistosome *Trichobilharzia szidati* (Trematoda: Schistosomatidae)

Anna Korsunenko[1,2], Galina Chrisanfova[1], Alexander Arifov[1], Alexey Ryskov[1], Seraphima Semyenova[1]

[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia
[2]Center of Parasitology, A.N. Severtsov Institute of Problems of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia
Email: korsanna@mail.ru

Received 26 May 2013; revised 20 June 2013; accepted 8 July 2013

Copyright © 2013 Anna Korsunenko *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Recently we applied randomly amplified polymorphic DNA (RAPD) fingerprinting to detect clonal variability among individual cercariae within daughter sporocysts and rediae of 10 digenean trematodes (Platyhelminthes: Trematoda). The most variable RAPD patterns were obtained for Schistosomatidae representative—avian schistosome *Trichobilharzia szidati*. In this work, 50 polymorphic DNA fragments of approximately 300 - 1500 bp from RAPD patterns of individual *T. szidati* cercariae were cloned and sequenced. As a result genomic DNA sequences (total length of approximately 41,000 bp) revealing clonal variability in *T. szidati* cercariae were obtained and analyzed. The analysis indicated that these sequences contained tandem, inverted and dispersed repeats as well as regions homological to retroelements of two human parasites, *Schistosoma mansoni* and *S. japonicum*. Tandem and inverted repeats constituted 8.9% and 22.1% respectively, while the percentage of dispersed repeats was 21.0%. The average content of these components was 41.7% with the average AT content being 59.0%. About 40% of sequences included regions ranging in length from 96 to 1005 bp which displayed amino acid homology with open reading frame *pol* products of *S. mansoni* and *S. japonicum* retroelements: non-long terminal repeat retrotransposons (nLTRs, 76%), long terminal repeat retrotransposons (LTRs, 14%), and Penelope-like elements (PLEs, 10%). Most of these regions (86.4%) contained frameshifts, gaps, and stop-codons. The largest portion of them was homological to nLTRs of the RTE clade (67%). The number of sequences homologous to the members of CR1 lineage was 7 times smaller (9%). Homology with LTRs of Gypsy/Ty3 and BEL clades was revealed in 5% and 9% of cases respectively. We assume that the repetitive elements including retroelement-like sequences described in the current study may serve as the source of clonal variability detected previously in *T. szidati* and other digenean trematodes. Such genome regions rapidly accumulate mutations and thus may play an important functional role in the life history of the species.

Keywords: RAPD Variability; Cercariae Heterogeneity; *Trichobilharzia szidati* Repetitive DNA

1. INTRODUCTION

The vast majority of eukaryotic species reproduces bisexually, yet approximately one out of every 1000 multicellular eukaryotic taxa is unisexual (parthenogenetic) or asexual. Parthenogenetic reproduction occurs in many phyla, especially in plants and invertebrates [1]. Among invertebrates, the Digenea (Platyhelminthes: Trematoda) have by far the most complex life cycles which usually involve free-living and parasitic stages and always incorporates both parthenogenetic (within molluscan first intermediate host) and sexual (within vertebrate definitive host) reproduction. There are some stages in digenean life cycle (parthenitae) reproducing via diploid parthenogenesis—mother sporocyst and either daughter sporocysts or rediae [2,3]. The reproductive function in parthenitae is performed by actively functioning ger-

minal masses [4]. Numerous free-swimming larvae (cercariae) are formed in daughter sporocysts or rediae after undergoing parthenogenetic reproduction. Since parthenitae are the result of diploid parthenogenesis with only one parent involved, all cercariae forming within sporocyst or redia might have been expected to represent a group of genetically identical individuals—the clone. Functionally diploid parthenogenesis can be viewed as a simple cell division. Acquirement of such type of reproduction is an essential adaptation developed by trematodes which allowed them to be evolutionary successful [2,3].

In general, genetic variation was described previously for different invertebrate clonal systems using a variety of molecular approaches [5]. Among digeneans, the variable occurrence of W1 and W2 repetitive elements was detected within and among daughter sporocyst generations of *Schistosoma mansoni* cultured *in vitro* [6,7]. Since an unexpected heterogeneity was found even among clonal cercariae arising from monomiracidial snail infections the mitotic recombination events were suggested to occur during the parthenogenetic reproduction of schistosomes. The clonal variability was also detected in *S. mansoni* [8] and *S. japonicum* [9] using multi-locus microsatellite analysis. In addition, RAPD variability within daughter sporocysts was determined in *Microphallus pygmaeus* and *M. pseudopygmaeus* [10].

RAPD-PCR is known as a DNA polymorphism analysis based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence, usually 1 - 12 bp in length [11,12]. It detects DNA polymorphisms produced by point mutations, insertions or deletions in the genome. RAPD assay has a distinct advantage of not requiring any specific nucleotide sequence information for amplification and can be employed across species using universal primers. This method is targeted primarily to abundant sequences within the genome, and usually generates a population of amplification products that can be characteristic of a specific organism. Because RAPD technique can reveal considerable polymorphism even between closely related organisms it has been used successfully for identification and differentiation of various parasite species [13]. In addition, RAPD-derived sequences obtained via cloning of PCR products can be used to design more specific and sensitive primers to develop locus-specific, or SCAR (Sequence-Characterized Amplified Regions), markers for parasite diagnosis. For example, using RAPD-derived sequences we developed a specific primer pair to detect three European causative agents of cercarial dermatitis in humans (*T. franki*, *T. szidati*, and *T. regenti*) during the prepatent period and after cercariae shedding [14].

Since RAPD assay is a method suitable for DNA polymorphism detection in organisms containing even small amounts of sample material we recently used this technique to reveal clonal heterogeneity between individual cercariae within sporocysts and rediae of 10 digenean trematodes from Schistosomatidae, Strigeidae, Gorgoderidae, Bucephalidae, Diplostomatidae, Plagiorchiidae, Halipegidae, Notocotylidae, and Echinostomatidae families [15].

Here, we characterized RAPD fragments revealing clonal variability between individual cercariae within daughter sporocysts of the avian schistosome *T. szidati* since the most variable RAPD patterns were obtained previously for this Schistosomatidae representative [15]. The analysis of polymorphic DNA fragments included sequence homology search using available nucleotide and amino sequences of the avian (*Trichobilharzia* spp.) and mammalian (*S. mansoni* and *S. japonicum*) schistosomes, search for tandem, inverted and dispersed repeats as well as AT/GC ratio calculation. Such information may contribute in determining structural genome organization of the avian schistosomes of *Trichobilharzia* spp. which are the most frequent causative agents of cercarial dermatitis in humans. RAPD-derived sequences of *T. szidati* obtained in the current study can be also valuable in cytogenetic studies. Our data may help to further investigate the mechanisms underlying clonal variability detected in digenean trematodes.

2. MATERIALS AND METHODS

In total, 250 cercariae of *T. szidati* were isolated from 47 daughter sporocysts sampled from nine naturally infected snails of *Lymnaea stagnalis*. Four snails (Lsm1, Lsm2, Lsm7, Lsm20) were collected from Russian (Moscow) and five snails (NLst4, NLst9, NLst10, NLst11, Stag1) from Belarusian (Naroch Lake) freshwater ponds. The parasite species identification was resolved using intermediate host-specificity combined with ITS2 rDNA, the amplification reactions were performed by using the primers its3Trem and ITS4 Trem [16].

Pieces of individual sporocysts were isolated from snail hepatopancreas, and, after washing in distilled water, fixed in 70% ethanol. The parthenitae were dissected with a preparation needle to isolate three to six cercariae. DNA was extracted from individual cercariae as described previously [17].

Since DNA amount extracted from individual cercaria is limited, the most intensely stained and unambiguous RAPD profiles were obtained only when the DNA sample was used for no more than three primers. Among tested random primers, those which detected the largest number of polymorphic markers were selected for this study. RAPD-PCR was performed with each of the chosen 10-mer primers, P29 (5’-CCGGCCTTAC-3’), OPA09 (5’-ACCGGACACT-3’), and OPA10 (5’-GTGATCGCAG-3’), in 25 µl volume containing 10 ng of total DNA, 75 mMTris-HCl (pH 8.8), 20 mM

Copyright © 2013 SciRes.

$(NH_4)_2SO_4$, 0.01% Tween 20, 5 mM $MgCl_2$, 0.25 mM of each dNTPs, 1 μM of primer, and 0.6 - 0.7 u of Taq DNA polymerase (Fermentas, Vilnius, Lithuania). The following amplification profile was used: 2 min of denaturation at 95 C, 35 cycles of 1 min at 94 C, 1 min at 38 C, 15 sec at 45 C and 2 min at 72 C, followed by the final 10 min extension at 72 C. Reaction mixture containing no template DNA was used as negative control for PCR assays.

In total, 50 DNA fragments of approximately 300 - 1500 bp were cut and eluted from the agarose gel using GFX PCR DNA and Gel Band Purification Kit (Amersham, Piscataway, New Jersey) and cloned into the pGEM-T Easy Vector (Promega, Madison, Wisconsin) following the manufacturer's instructions. Plasmid isolation was made with GeneJET Plasmid Miniprep Kit (Fermentas, Vilnius, Lithuania). For each RAPD fragment one clone was selected for processing by the automatic sequencing system ABI PRISM 3100-Avant (Applied Biosystems, Foster City, California). The nucleotide sequences were deposited to GenBank: JX049928-JX049977 (**Supplementary Table 1**).

Repeats search within obtained *T. szidati* sequences was performed using Tandem Repeat Finder [18], Inverted Repeat Finder [19] and Spectral Repeat Finder (http://www.imtech.res.in/raghava/srf/). As a result of automatic repeats search, a specific nucleotide position may be included by analysis software into more than one repeat. So the calculation of an abundance of repeats based on a sum of lengths of all repeats found by the analysis software may produce an overrated result. To overcome this, we created an algorithm which uses the data from repeats search software and calculates the percentage of tandem, inverted and dispersed repeats as a ratio of the number of nucleotide positions included in at least one repeat to total length of analyzed sequence. AT/GC ratio was calculated by DNA/RNA Base Composition Calculator software (http://www.currentproto-cols.com/tools/dnarna-base-composition-calculator).

Using the Basic Local Alignment Search Tool (BLAST) sequences of *T. szidati* were subjected to search against the nucleotide (blastn) and protein (blastx) sequences databases of schistosomes (*S. mansoni*, *S. japonicum*, and *Trichobilharzia* spp.) available throughout the NCBI (http://www.ncbi.nlm.nih.gov). Sequence similarities identified by the BLAST algorithms were considered statistically significant with E value of $\leq 10^{-5}$.

3. RESULTS

To detect clonal variability among individual cercariae within sporocysts of *T. szidati* we selected three random primers (P29, OPA09, OPA10) which previously determined the largest number of polymorphic markers in cercariae RAPD profiles [15]. These primers yielded a total of 105 RAPD bands in the range of 220 - 2000 bp. A set of 50 bright, clear, intensively amplified polymorphic fragments of approximately 300 - 1500 bp was chosen from cercariae banding patterns to construct *T. szidati* DNA library (**Supplementary Table 1**). Among them, 20 fragments were obtained using the P29 primer, while 14 and 16 bands were obtained using the OPA09 and OPA10 primers respectively. Several cloned RAPD fragments obtained with the P29 (Ts54-27, Ts88-83), OPA09 (Ts52-50, Ts58-00), and OPA10 (Ts10-87, Ts72-64) primers are shown as examples in **Figure 1**.

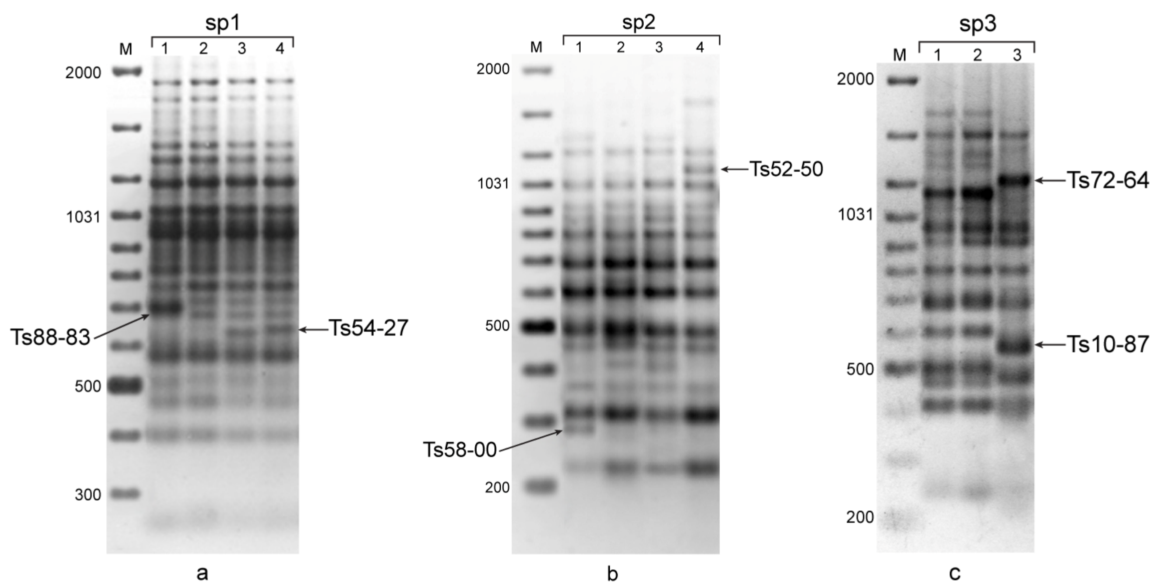

Figure 1. RAPD patterns of *Trichobilharzia szidati* cercariae obtained with the P29 (a), OPA09 (b), and OPA10 (c) primers. *Lane M*—molecular size marker (100-bp ladder), sp—sporocyst; the cloned bands are indicated by the arrow.

Copyright © 2013 SciRes.

For each RAPD fragment one clone containing DNA insert was randomly selected for sequencing. In total, 50 sequences ranging in size from 281 to 1505 bp were obtained. There was no similarity between the sequences. The analysis indicated the prevalence of AT-bases (59.0%). Repeats search revealed a number of tandem, inverted, and dispersed repetitive components within these sequences (**Supplementary Table 2**). For example, one of the sequences (Ts54-27) contained a microsatellite repeat $(CA)_{10}$, while a 309-bp-long minisatellite repeat was identified within another sequence (Ts59-00). This repeat consisted of nine 35-bp monomers. Interestingly, Ts49-66 included a 167-bp region similar to Sau3A minisatellite repeat of *T. ocellata* and *T. regenti*. **Supplementary Table 3(a)** gathers the information concerning sizes of homology regions, their identity/similarity (I/S) levels, and significance. In analyzed *T. szidati* sequences tandem and inverted repeats constituted 8.9% and 22.1% respectively, while the percentage of dispersed repeats was 21.0%. Average repeats content turned out to be 41.7% of total sequence length.

Additionally, BLAST algorithms (blastn and blastx) were used to find *T. szidati* sequence homology with nucleotide and amino acid sequences of the avian (*Trichobilharzia* spp.) and mammalian (*S. mansoni*, *S. japonicum*) schistosomes available throughout the NCBI. Based on nucleotide sequence similarity search (blastn) all the sequences revealed regions ranging from 35 to 643 bp homological to numerous parts of *S. mansoni* and *S. japonicum* genome. Similarity with available *Trichobilharzia* spp. sequences was either not found or had no significance ($E > 10^{-5}$) except for the case of Ts49-66 described above. Translated nucleotide sequence similarity search (blastx) detected 22 regions (fragments) within 19 *T. szidati* sequences which reveal homology with open reading frame *pol* products (reverse transcriptase, endonuclease, and integrase) of *S. mansoni* and *S. japonicum* retroelements from three major subclasses: long terminal repeat retrotransposons (LTRs), non-long terminal repeat retrotransposons (nLTRs), and Penelope-like elements (PLEs) (Supplementary **Table 3(b)**). All these regions had no similarity between each other at amino acid level, consequently, they represented non-overlapping regions homological to retroelements. Among them, three homology regions were found in Ts82-32, two regions were detected in Ts49-66, while the other sequences revealed one region each (**Supplementary Table 3(b)**). Nineteen of 22 detected fragments were found to be disrupted by gaps (sequences, containing these regions: Ts49-66, Ts51-24, Ts51-49, Ts53-51, Ts64-57, Ts72-64, Ts74-83, Ts76-85, Ts82-32), stop-codons (sequences, containing these regions: Ts48-65, Ts49-66, Ts51-49, Ts52-50, Ts64-57, Ts69-102, Ts72-64, Ts74-83, Ts76-85, Ts82-32), and frameshifts (sequences, containing these regions: Ts49-66, Ts51-24, Ts53-51, Ts71-00) (**Supplementary Table 3(b)**).

Our data indicated the predominance of regions (n = 13) homological to the following nLTRs of the RTE clade: SjCHGCS20 and SjR2 (one region in Ts51-82, Ts71-00, Ts82-32, Ts85-90 and two in Ts49-66) as well as SR3, Perere-3, and SjCHGCS19 (Ts52-50, Ts64-57, Ts72-64, Ts74-83, Ts76-85, Ts82-32, Ts88-83) (**Supplementary Table 3(b)**). Lengths of homology subregions varied from 45 to 546 bp with I/S level ranging from 26/50% to 77/86%. Moreover, a 55-bp fragment homological to untranslated region of Perere-3 was found in Ts47-78 (**Supplementary Table 3(a)**). In contrast, regions homological to the members of CR1 clade were found to be 2.5 times lower as compared to RTE. These regions demonstrating homology with Perere, Perere-4, -7, and Perere-2, -5, -6 were detected in Ts53-51 and Ts51-24 respectively (**Supplementary Table 3(b)**). Lengths of homology subregions were 177 - 405 bp, while values of I/S varied from 33/55% to 55/74%. Furthermore, a 132-bp-long region homological (I/S = 48/75%) to Perere-8 was revealed in Ts48-65 (**Supplementary Table 3(b)**). Previously, the truncated copy of Perere-8 was identified as CR1 member [20] whereas by examining a full-length copy of this element the apurinic/apyrimidinic endonuclease domain was detected which is characteristic of the more modern lineages of nLTRs [21].

The most lengthy homology region (1005 bp) was detected in Ts51-49 (**Supplementary Table 3(b)**). This region revealed homology (I/S = 54/68%) with SjCHGCS3-retrotransposon from Gypsy/Ty3 clade of LTRs. Lengths of subregions homological to the other members of this lineage (Saci-5, SjCHGCS1, and SjCHGCS4) ranged from 702 to 966 bp, while I/S varied from 27/42% to 57/71%. Two other *T. szidati* sequences (Ts48-00, Ts83-28) contained regions homologous to BEL representatives of LTRs (SjCHGCS16, Saci-1, etc.) (**Supplementary Table 3(b)**). Lengths of homology subregions were 96 - 252 bp with I/S varying in the range of 42/63% - 57/74%.

In total collection two regions demonstrated homology with PLEs: 273-bp region homological (I/S = 33/55%) to SjPenelope-2 was detected in Ts69-102, while Ts82-32 included 123-bp region similar to SjPenelope-2, -3, and Perere-10 (**Supplementary Table 3(b)**).

The most complex structure of homology regions was identified in Ts49-66 (**Figure 2**). This sequence contained two regions homological to RTE members (SjCHGCS20 and SjR2): located towards 5'-end subregion I and II formed the first region, while the second one (designated as VI) was located closer to 3'-end. These two regions were homological to different parts within amino acid sequences of SjCHGCS20 and SjR2: the first region had homology with 856 - 913 and 1029 -

Copyright © 2013 SciRes.

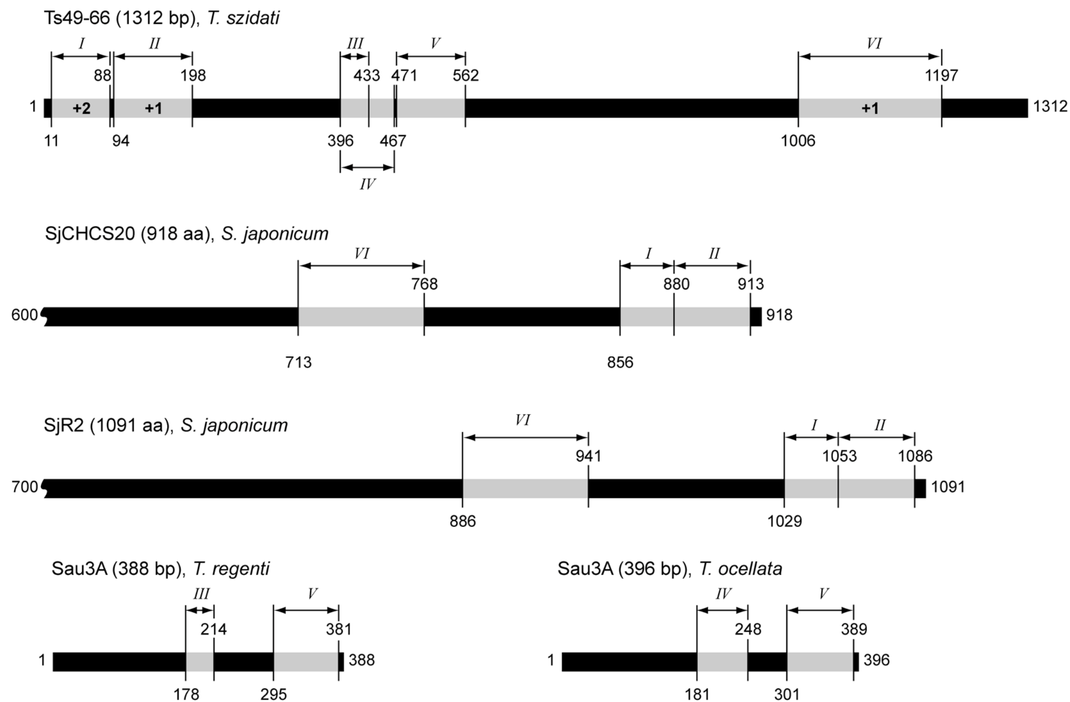

Figure 2. Location of homology regions in Ts49-66 nucleotide sequence of *Trichobilharzia szidati*. Subregion I and II formed the first region homological to SjCHGCS20 and SjR2 retrotransposons, the second region is designated as VI. Subregions III, IV, and V represented the region homological to Sau3A minisatellite repeat of *T. ocellata* and *T. regenti*. Regions without any homology are indicated in black. Reading frames are indicated as +1 and +2.

1086 aa respectively, while the second region had homology with 713 - 768 and 886 - 941 aa respectively. Therefore, Ts49-66 included regions homological to different copies of SjCHGCS20 and SjR2. Finally, subregions III, IV, and V (located in the central part of the sequence) represented the region homological to Sau3A minisatellite repeat of *T. ocellata* and *T. regenti*.

Besides that, computer-assisted sequence similarity search revealed two regions homological at amino acid level to *S. mansoni* regulatory elements. One of them (288 bp) detected in Ts67-61 was homologous to zinc finger transcription factor gli2 with I/S being 69/80%, while the other one (177 bp) detected in Ts04-84 was homological to transcription initiation factor tfiid with I/S being 92/98% (**Supplementary Table 3(c)**).

In summary, a set of RAPD-derived sequences (total length of approximately 41,000 bp) which reveal clonal variability in *T. szidati* cercariae were obtained and analyzed. The analysis indicated that these sequences contained tandem, inverted and dispersed repeats. The average content of these components was 41.7% with the average AT content being 59.0%. About 40% of sequences included regions ranging in length from 96 to 1005 bp which displayed amino acid homology with pol products of *S. mansoni* and *S. japonicum* retroelements: nLTRs (76%), LTRs (14%), and PLEs (10%) (**Figure 3**). Most of these regions (86.4%) contained frameshifts, gaps, and stop-codons. The largest portion of them was homological to nLTRs of the RTE clade (67%). The number of sequences homologous to CR1 members was 7 times smaller (9%). Homology with LTRs of Gypsy/Ty3 and BEL lineages was revealed in 5% and 9% of cases respectively.

4. DISCUSSION

The aim of this study was the molecular characterization of anonymous nuclear genome sequences produced by RAPD genotyping and revealing clonal variability between individual cercariae within sporocysts of *T. szidati*. The majority of RAPD-derived sequences belonged to repetitive elements. Among them there were tandem, inverted and dispersed repeats as well as regions homological to retroelements of two human parasites, *S. mansoni* and *S. japonicum*. The complete genome sequences of *S. mansoni*, *S. japonicum*, and *S. haematobium* were published recently [22-24]. Over a third part of their genomes consists of the repetitive DNA including mobile genetic elements (MGE), predominantly retrotransposons. The average tandem, inverted and dispersed repeats content (41.7%) in *T. szidati* sequences was comparable with total repetitive DNA constitution in *S. mansoni* (40.0%), *S. japonicum* (40.1%), and *S. haematobium* (43.0%) genomes [22-24]. The AT-content prevalence (59.0%) was

Copyright © 2013 SciRes.

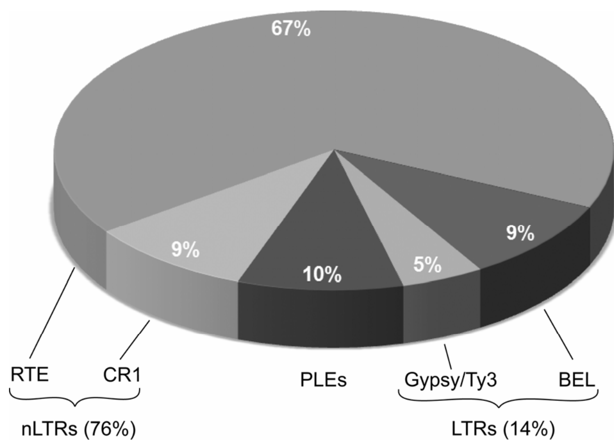

Figure 3. Distribution of homology regions detected in *Trichobilharzia szidati* nucleotide sequences by three retroelement subclasses. RTE, CR1—clades of non-long terminal repeat retrotransposons (nLTRs); Gypsy/Ty3, BEL—clades of long terminal repeat retrotransposons (LTRs); PLEs—Penelope-like elements.

detected in analyzed sequences of *T. szidati* as previously was reported for *S. mansoni*, *S. japonicum*, and *S. haematobium* (65.3%, 66.5%, and 65.7% respectively) [22-24]. Almost one-half of *T. szidati* sequences included fragments homological to retroelements of mammalian schistosomes (nLTRs, LTRs, PLEs). The largest portion of them (76%) demonstrated homology with nLTRs. This type of MGE was shown to be predominant in *S. mansoni* and *S. japonicum* genomes [22,24]. Both blood flukes are known to contain numerous families from the CR1 and RTE clades of nLTRs [21]. Likewise, most of homology regions detected in *T. szidati* sequences were similar to RTE and CR1 members. The elements from R2 lineage of nLTRs also occur in *S. mansoni* and *S. japonicum*, but do not comprise a significant fraction of their genomes [21]. Homologues of R2 members were not found among analyzed *T. szidati* DNA fragments. This may be explained by the low frequency or general absence of R2 representatives in genome of this avian schistosome. Additionally, we detected homologues of LTRs from two clades (Gypsy/Ty3 and BEL) common in mammalian schistosomes [22,24]. Among *T. szidati* sequences those homological to PLEs were revealed as well. PLEs are not very abundant in mammalian schistosomes, representing less than 2% of bases in the shortgun reads of their genomes [21].

Retrotransposons and other MGE colonize the genomes of nearly every eukaryote and they play an important role in their genome structural organization and evolution [25]. The genetic variability caused by MGE ranges from changes in single nucleotides to changes in the size and arrangement of whole genomes. Given that the majority of new MGE insertions tend to be deleterious to host, different mechanisms have been developed to mitigate the reduction in host fitness. MGE commonly integrates into non-coding regions, for example, introns to increase their probability of survival because of less visibility to natural selection. However, MGE activity can result in positive benefit to their hosts providing positive selection on elements [26,27].

The genomes of most organisms carry tens to thousands of retrotransposon copies. The majority of them is non-functional and contains disabling mutations. The coding sequences of such copies are highly degenerate, cluttered with stop codons, frameshifts, and large indels [28]. The mammalian schistosome genome also contains defective copies of transposable elements. For example, the coding sequences of different *S. mansoni* retrotransposons found to be truncated or contain insertions of related copies [20,29]. Furthermore, retrotransposons in *S. japonicum* were represented by one to 793 intact copies and hundreds to thousands of partial copies [24]. Previously, the degenerated inactive copies were found to be more abundant in heterochromatic regions than in the euchromatin [30,31].

Potential negative effects of MGE on the fitness of their hosts necessitate the development of strategies for transposon control. One of the most important mechanisms to regulate MGE expression rates and transposition frequencies is RNA interference (RNAi) pathway which recognizes intracellular double-stranded RNA (dsRNA). MGE may represent a source of dsRNA and thus represent the RNAi target [32]. RNAi as an important tool to elucidate gene function has been identified in *S. mansoni* and *S. japonicum* [33,34]. RNA silencing system is likely to be common in other flatworms, including avian schistosomes.

Recently, analysis of *S. japonicum* endogenous short interfering RNAs (siRNAs) detected that the majority of them are transposable-elements-derived [35,36]. It was discovered that such siRNAs in *Drosophila* were derived from heterochromatic genomic loci which have previously been identified as master regulators of transposon activity [37]. These loci contain numerous defective clustered MGE copies (fragments of MGE as well as nested copies). Likewise, described in this study *T. szidati* DNA fragments homological to retroelements may act as a source of siRNA to put down MGE expansion. The most part of these regions (86.4%) contains frameshifts, stop-codons and gaps. Nevertheless, these sequences possess a genetic memory of MGE bursts in the evolutionary history. Taking into account that transposable element activity facilitates emergence of new genes, modifies gene expression patterns and promotes chromosomal rearrangements, MGE bursts can contribute to the evolution of lineage-specific traits that increase adaptability of the species [38,39]. Recently, it was suggested that the higher nLTRs content in *S. mansoni* (15.4%) in relation

Copyright © 2013 SciRes.

to *S. japonicum* (8.3%) can be attributed to higher representation of two retrotransposon families (SR2 and Perere-3/SR3) of the RTE clade compared with the representation of their closest relative families in *S. japonicum* [21]. Considering the model of origin of African schistosomes from a migrating ancestral species dwelling in Asia it is proposed that bursts of SR2 and Perere-3/SR3 in *S. mansoni* would be a consequence of the selection of parasite populations in a new environment during the migration and speciation of blood flukes in Africa. Likewise, the bursts of MGE transpositions may have taken place in the evolutionary history of *T. szidati* studied in the present work.

As it was mentioned above, the germinal cells of parthenitae (daughter sporocysts and rediae) undergo mitotic division instead meiotic, which may have suggested their genetic identity. The main cause of genetic instability in this case is homologous recombination during mitosis. Mitotic recombination can lead to either reciprocal (mitotic crossing over) or non-reciprocal (gene conversion) transfer of genetic material.

The frequency of mitotic crossing over in somatic cells of the most of eukaryotes is quite low. For example, spontaneous homologous recombination occurs at a rate of 10^{-6} to 10^{-5} per cell cycle between repeated DNA sequences in mammalian cells [40]. In contrast, the frequency of homologous recombination can be very high in fertilized eggs (1/500) and embryonic cells (up to 10^{-1}) [41,42]. Likewise, the frequency of homologous recombination in proliferating germinal cells of digenean trematodes appears to be significantly high. The longevity of the germinal cell proliferation period was found to vary in different groups of trematodes [4]. Generative function is absent or weakly expressed in the mother sporocysts of the most archaic and primitive trematodes (Fasciolidae, Philophthalmidae, Cyclocoelidae, Notocotylidae, Halipegidae, and many Echinostomatidae). In contrast, the long proliferation period of germinal cells during the parthenitae development was found in more specialized groups (Sanguinicolidae, Schistosomatidae, Bucephalidae, Diplostomatidae, Strigeidae, Plagiorchiidae) [3]. These findings are well supported by our data obtained earlier [15] in that we found that species which have more germinal cell proliferation cycles tend to exhibit higher levels of clonal variability. Based on RAPD markers we revealed significant genetic heterogeneity (the percentage of polymorphic loci (*P*) ranged from 17.8% - 29.4%) between cercariae within daughter sporocysts of five specialized forms of digenean trematodes (Schistosomatidae, Gorgoderidae, Bucephalidae, Diplostomatidae, Plagiorchiidae). Additionally, the high level of clonal variability between cercariae within daughter sporocysts of Bucephalidae representative was confirmed using several snail-hosts in another study [43]. In contrast, the low level of clonal variability (*P* values ranged in 5.2% - 6.5%) was determined between cercariae within rediae of four archaic forms of digenean trematodes (Halipegidae, Notocotylidae, and Echinostomatidae). The average percentage of polymorphic loci for specialized forms was four times higher (25.4%) compared with archaic forms (6.2%). So, the numerous proliferation events of germinal cells in parthenitae of the studied specialized digeneans may lead to increased recombination frequency and consequent accumulation of different genome rearrangements, while the low level of clonal variability observed using RAPD markers in studied archaic digeneans reflects the limited multiplication of germinal cells and reduced frequency of recombination. The recombination frequency is also influenced by the content and chromosome localization of recombined sequences.

Taking into account the specificities of RAPD technique [11,12] and sequence analysis results obtained here we supposed that rearrangements occur in moderate and highly repetitive genome fraction of *T. szidati*. The clonal heterogeneity detected previously in mammalian schistosomes [6-9] was shown to be based on the repetitive DNA variability. To compare sequence analysis results of *T. szidati* (the specialized digenean representative) for one of the archaic digenean representative (*Echinoparyphium aconiatum*) two polymorphic RAPD fragments which reveal clonal variability in cercariae within rediae were cloned and sequenced by us previously (unpublished data). The obtained sequences of 441 bp and 667 bp contained regions homological at amino acid level to *pol* products of *S. mansoni* retrotransposons. One of these sequences (GenBank KC902786) revealed homology with Saci-3, the member of Gypsy/Ty3 clade of LTRs. The other one (GenBank KC902787) was homological to Perere-5 from CR1 clade of nLTRs. So, in comparison to *T. szidati* clonal variability in *E. aconiatum* cercariae also seemed to be associated with rearrangements in repetitive DNA including retroelement-like sequences. Since an important consequence of mitotic recombination is homozygosis of heterozygous markers, the increase of recombination frequency may be considered as the evolutionary adaptation of more specialized and evolutionary advanced digeneans to prevent the spatial distribution of MGE compared to the archaic forms. However, comparative analysis of MGE abundance in both specialized and archaic groups of digeneans is required to support this hypothesis.

5. ACKNOWLEDGEMENTS

This work received financial support from the Russian Foundation for Basic Research (12-04-01153-a, 12-04-90034-Bel_a), President RF Program of Leading Scientific Schools (S.S. -5233.2012.4), the Russian Academy of Sciences Programs (Molecular and Cell Biology) of the

Copyright © 2013 SciRes.

Presidium of the Russian Academy of Sciences and President Grants for Government Support of Young Russian Scientists (MK-1026. 2012.4).

REFERENCES

[1] Bell, G. (1982) The masterpiece of nature: The evolution and genetics of sexuality. University of California Press, Berkeley.

[2] Galaktionov, K. and Dobrovolskij, A. (1998) The origin and evolution of trematode life cycles. Nauka, Saint-Petersburg.

[3] Galaktionov, K. and Dobrovolskij, A. (2003) The biology and evolution of trematodes. Kluwer Academic Publishers, Boston, Dordrecht, London.

[4] Cort, W.W., Ameel, D.J. and Van Der Woude, A. (1954) Germinal development in the sporocysts and rediae of the digenetic trematodes. *Experimental Parasitology*, **3**, 185-225. doi:10.1016/0014-4894(54)90008-9

[5] Lushai, G. and Loxdale, H.D. (2002) The biological improbability of a clone. *Genetics Research*, **79**, 1-9. doi:10.1017/S0016672301009582

[6] Bayne, C.J. and Grevelding, C.G. (2003) Cloning of *Schistosoma mansoni* sporocysts *in vitro* and detection of genetic heterogeneity among individuals within clones. *Journal of Parasitology*, **89**, 1056-1060. doi:10.1645/GE-3186RN

[7] Grevelding, C.G. (1999) Genomic instability in Schistosoma mansoni. *Molecular and Biochemical Parasitology*, **101**, 207-216. doi:10.1016/S0166-6851(99)00078-X

[8] Gower, C.M., Shrivastava, J., Lamberton, P.H., *et al.* (2007) Development and application of an ethically and epidemiologically advantageous assay for the multi-locus microsatellite analysis of *Schistosoma mansoni*. *Parasitology*, **134**, 523-536. doi:10.1017/S0031182006001685

[9] Yin, M., Hu, W., Mo, X., *et al.* (2008) Multiple near-identical genotypes of *Schistosoma japonicum* can occur in snails and have implications for population-genetic analyses. *International Journal for Parasitology*, **38**, 1681-1691. doi:10.1016/j.ijpara.2008.05.015

[10] Khalturin, K.V., Mikhailova, N.A. and Granovich, A.I. (2000) Genetic heterogeneity in natural populations of *Microphallus piriformes* and *M. pygmaeus* parthenites (Trematoda: Microphallidae). *Parazitologiia*, **34**, 486-501.

[11] Welsh, J. and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, **18**, 7213-7218. doi:10.1093/nar/18.24.7213

[12] Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**, 6531-6535. doi:10.1093/nar/18.22.6531

[13] Zarlenga, D.S. and Higgins, J. (2001) PCR as a diagnostic and quantitative technique in veterinary parasitology. *Veterinary Parasitology*, **101**, 215-230. doi:10.1016/S0304-4017(01)00568-4

[14] Korsunenko, A.V., Chrisanfova, G.G., Ryskov, A.P., Movsessian, S.O., Vasilyev, V.A. and Semyenova, S.K. (2010) Detection of European Trichobilharzia schistosomes (*T. franki*, *T. szidati*, and *T. regenti*) based on novel genome sequences. *Journal of Parasitology*, **96**, 802-806. doi:10.1645/GE-2297.1

[15] Semyenova, S.K., Khrisanfova, G.G., Korsunenko, A.V., *et al.* (2007) Multilocus variation in cercariae, parthenogenetic progeny of different species of the class Trematoda. *Doklady Biological Sciences*, **414**, 235-238. doi:10.1134/S0012496607030192

[16] Dvorak, J., Vanacova, S., Hampl, V., Flegr, J. and Horak, P. (2002) Comparison of European Trichobilharzia species based on ITS1 and ITS2 sequences. *Parasitology*, **124**, 307-313.

[17] Korsunenko, A., Chrisanfova, G., Lopatkin, A., Beer, S. A., Voronin, M., Ryskov, A.P. and Semyenova, S.K. (2012) Genetic differentiation of cercariae infrapopulations of the avian schistosome Trichobilharzia szidati based on RAPD markers and mitochondrial cox1 gene. *Parasitology Research*, **110**, 833-841. doi:10.1007/s00436-011-2562-6

[18] Benson, G. (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573-580. doi:10.1093/nar/27.2.573

[19] Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y. and Benson, G. (2004) Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Research*, **14**, 1861-1869. doi:10.1101/gr.2542904

[20] DeMarco, R., Machado, A.A., Bisson-Filho, A.W. and Verjovski-Almeida, S. (2005) Identification of 18 new transcribed retrotransposons in *Schistosoma mansoni*. *Biochemical and Biophysical Research Communications*, **333**, 230-240. doi:10.1016/j.bbrc.2005.05.080

[21] Venancio, T.M., Wilson, R.A., Verjovski-Almeida, S. and DeMarco, R. (2010) Bursts of transposition from non-long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. *International Journal for Parasitology*, **40**, 743-749. doi:10.1016/j.ijpara.2009.11.013

[22] Berriman, M., Haas, B.J., LoVerde, P.T., *et al.* (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature*, **460**, 352-358. doi:10.1038/nature08160

[23] Young, N.D., Jex, A.R., Li, B., *et al.* (2012) Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics*, **44**, 221-225. doi:10.1038/ng.1065

[24] Zhou, Y., Zheng, H., Chen, Y., *et al.* (2009). The Schistosoma japonicum genome reveals features of host-parasite interplay. *Nature*, **460**, 345-351. doi:10.1038/nature08140

[25] Berg, D. and Howe, M. (1989). Mobile DNA. American Society of Microbiology, Washington DC.

[26] Kidwell, M.G. and Lisch, D. (1997) Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences of the United of America*, **94**, 7704-7711. doi:10.1073/pnas.94.15.7704

Copyright © 2013 SciRes.

[27] Thomas, M.C., Macias, F., Alonso, C. and Lopez, M.C. (2010) The biology and evolution of transposable elements in parasites. *Trends in Parasitology*, **26**, 350-362. doi:10.1016/j.pt.2010.04.001

[28] Laten, H.M., Mogil, L.S. and Wright, L.N. (2009) A shotgun approach to discovering and reconstructing consensus retrotransposons ex novo from dense contigs of short sequences derived from Genbank Genome Survey Sequence database records. *Gene*, **448**, 168-173. doi:10.1016/j.gene.2009.06.011

[29] Drew, A.C., Minchella, D.J., King, L.T., Rollinson, D. and Brindley, P.J. (1999) SR2 elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni*. *Molecular Biology and Evolution*, **16**, 1256-1269. doi:10.1093/oxfordjournals.molbev.a026216

[30] Dimitri, P., Junakovic, N. and Arca, B. (2003) Colonization of heterochromatic genes by transposable elements in Drosophila. *Molecular Biology and Evolution*, **20**, 503-512. doi:10.1093/molbev/msg048

[31] Mandrioli, M. (2000) Mariner-like transposable elements are interspersed within the rDNA-associated heterochromatin of the pufferfish Tetraodon fluviatilis (Osteichthyes). *Chromosome Research*, **8**, 177-179. doi:10.1023/A:1009254805686

[32] Sijen, T. and Plasterk, R.H. (2003) Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature*, **426**, 310-314. doi:10.1038/nature02107

[33] Boyle, J.P., Wu, X.J., Shoemaker, C.B. and Yoshino, T.P. (2003) Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Molecular and Biochemical Parasitology*, **128**, 205-215. doi:10.1016/S0166-6851(03)00078-1

[34] Zhao, Z.R., Lei, L., Liu, M., Zhu, S.C., Ren, C.P., Wang, X.N. and Shen, J.J. (2008) *Schistosoma japonicum*: Inhibition of Mago nashi gene expression by shRNA-mediated RNA interference. *Experimental Parasitology*, **119**, 379-384. doi:10.1016/j.exppara.2008.03.015

[35] Hao, L., Cai, P., Jiang, N., Wang, H. and Chen, Q. (2010) Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics*, **11**, 55. doi:10.1186/1471-2164-11-55

[36] Wang, Z., Xue, X., Sun, J., *et al.* (2010) An "in-depth" description of the small non-coding RNA population of *Schistosoma japonicum* schistosomulum. *PLoS Neglected Tropical Diseases*, **4**, e596. doi:10.1371/journal.pntd.0000596

[37] Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, **128**, 1089-1103. doi:10.1016/j.cell.2007.01.043

[38] Betran, E. and Long, M. (2002) Expansion of genome coding regions by acquisition of new genes. *Genetica*, **115**, 65-80. doi:10.1023/A:1016024131097

[39] Oliver, K.R. and Greene, W.K. (2009) Transposable elements: powerful facilitators of evolution. *BioEssays*, **31**, 703-714. doi:10.1002/bies.200800219

[40] Helleday, T. (2003) Pathways for mitotic homologous recombination in mammalian cells. *Mutation Research*, **532**, 103-115. doi:10.1016/j.mrfmmm.2003.08.013

[41] Brinster, R.L., Braun, R.E., Lo, D., Avarbock, M.R., Oram, F. and Palmiter, R.D. (1989) Targeted correction of a major histocompatibility class II E alpha gene by DNA microinjected into mouse eggs. *Proceedings of the National Academy of Sciences of the United of America*, **86**, 7087-7091. doi:10.1073/pnas.86.18.7087

[42] Wurst, W. and Joyner, A. (1993) Production of targeted embryonic stem cell clones. In A. Joyner, Ed., *Gene Targeting*: *A Practical Approach*. Oxford University Press, New York, 33-61.

[43] Korsunenko, A.V., Tiutin, A.V. and Semenova, S.K. (2009) Clonal and population RAPD variation of cercariae obtained from *Bucephalus polymorphus* sporocysts (Trematoda: Bucephalidae). *Genetika*, **45**, 73-80.

Copyright © 2013 SciRes. OPEN ACCESS

SUPPLEMENTARY MATERIALS

Supplementary Table 1. The list of cloned variable RAPD fragments from *Trichobilharzia szidati* cercariae banding patterns obtained with the P29, OPA09, and OPA10 primers.

No.	Fragment (clone) identification	GenBank Acc. Num.	Length, bp	Random primer	Snail-host identification	Significant ($E \leq 10^{-5}$) homology with protein database sequences
1	Ts04-84	JX049957	683	OPA09	Lsm2	+
2	Ts09-86	JX049976	535	OPA10	Lsm20	–
3	Ts10-87	JX049972	572	OPA10	Lsm1	–
4	Ts11-88	JX049966	792	OPA10	NLst10	–
5	Ts12-89	JX049973	549	OPA10	Lsm2	–
6	Ts27-79	JX049943	887	P29	Lsm7	–
7	Ts28-80	JX049945	510	P29	Lsm20	–
8	Ts31-81	JX049959	589	OPA09	Lsm7	–
9	Ts47-20	JX049933	1039	P29	NLst10	–
10	Ts47-26	JX049951	435	OPA09	NLst10	+
11	Ts47-78	JX049963	1243	OPA10	NLst4	–
12	Ts48-00	JX049930	428	P29	NLst9	+
13	Ts48-65	JX049944	1420	P29	Lsm7	+
14	Ts49-00	JX049950	503	OPA09	NLst9	–
15	Ts49-66	JX049937	1312	P29	Stag1	+
16	Ts50-67	JX049939	1235	P29	Lsm1	–
17	Ts51-24	JX049961	1320	OPA09	Lsm20	+
18	Ts51-49	JX049977	1137	OPA10	Lsm20	+
19	Ts51-82	JX049975	1018	OPA10	Lsm7	+
20	Ts52-50	JX049956	1082	OPA09	Lsm1	+
21	Ts53-51	JX049967	1383	OPA10	NLst10	+
22	Ts54-27	JX049942	640	P29	NLst11	–
23	Ts54-69	JX049954	739	OPA09	Stag1	–
24	Ts55-28	JX049947	1320	P29	Lsm20	–
25	Ts56-00	JX049952	522	OPA09	NLst10	–
26	Ts57-00	JX049953	288	OPA09	NLst11	–
27	Ts58-00	JX049955	281	OPA09	Lsm1	–
28	Ts59-00	JX049932	299	P29	NLst10	–
29	Ts59-60	JX049969	1190	OPA10	NLst11	–
30	Ts64-57	JX049940	1210	P29	Lsm2	+

Copyright © 2013 SciRes. OPEN ACCESS

Continued

31	Ts65-58	JX049968	1172	OPA10	NLst11	–
32	Ts65-97	JX049958	758	OPA09	Lsm2	–
33	Ts67-61	JX049965	1124	OPA10	NLst9	+
34	Ts69-102	JX049949	492	OPA09	NLst9	+
35	Ts71-00	JX049928	372	P29	NLst4	+
36	Ts72-64	JX049971	1215	OPA10	Lsm1	+
37	Ts74-83	JX049946	957	P29	Lsm20	+
38	Ts76-85	JX049962	794	OPA10	NLst4	+
39	Ts80-49	JX049974	857	OPA10	Lsm7	–
40	Ts81-50	JX049964	689	OPA10	NLst9	–
41	Ts82-32	JX049941	1505	P29	Lsm2	+
42	Ts83-28	JX049934	412	P29	NLst11	+
43	Ts84-00	JX049938	385	P29	Lsm1	–
44	Ts85-90	JX049931	1252	P29	NLst9	+
45	Ts88-83	JX049935	718	P29	NLst11	+
46	Ts91-86	JX049929	755	P29	NLst4	–
47	Ts92-87	JX049936	719	P29	Stag1	–
48	Ts94-00	JX049960	723	OPA09	Lsm7	–
49	Ts96-00	JX049948	356	OPA09	NLst4	–
50	Ts101-00	JX049970	317	OPA10	Stag1	–

Copyright © 2013 SciRes.

Supplementary Table 2. Tandem, inverted, and dispersed repeats content and AT/GC content in *Trichobilharzia szidati* sequences obtained using RAPD-PCR.

No.	Fragment (clone) identification	Length, bp	Inverted repeats length, bp	Inverted repeats content, %	Tandem repeats length, bp	Tandem repeats content, %	Disperse repeats length, bp	Disperse repeats content, %	Total repeats length, bp	Total repeats content, %	AT content, %	GC content, %
1	Ts04-84	863	134	15.5	151	17.5	119	13.8	272	31.5	58	42
2	Ts09-86	535	102	19.1	74	13.8	134	25.0	239	44.7	56	44
3	Ts10-87	572	111	19.4	0	0.0	145	25.3	187	32.7	61	39
4	Ts11-88	792	91	11.5	74	9.3	268	33.8	372	47.0	46	54
5	Ts12-89	549	135	24.6	94	17.1	89	16.2	243	44.3	63	37
6	Ts27-79	887	298	33.6	0	0.0	192	21.7	409	46.1	64	36
7	Ts28-80	510	108	21.2	0	0.0	134	26.3	210	41.2	63	37
8	Ts31-81	589	162	27.5	45	7.6	135	22.9	271	46.0	63	37
9	Ts47-20	1058	82	7.8	122	11.5	144	13.6	332	31.4	62	38
10	Ts47-26	435	94	21.6	24	5.5	81	18.6	166	38.2	54	46
11	Ts47-78	1243	419	33.7	50	4	236	19	563	45.3	63	37
12	Ts48-00	428	34	7.9	0	0	45	10.5	78	18.2	53	47
13	Ts48-65	1445	519	35.9	273	18.9	262	18.1	762	52.7	63	37
14	Ts49-00	503	29	5.8	52	10.3	154	30.6	221	43.9	62	38
15	Ts49-66	1330	364	27.4	173	13	198	14.9	588	44.2	57	43
16	Ts50-67	1255	467	37.2	23	1.8	442	35.2	891	71	61	39
17	Ts51-24	1337	373	27.9	111	8.3	297	22.2	653	48.8	60	40
18	Ts51-49	1159	211	18.2	52	4.5	166	14.3	398	34.3	56	44
19	Ts51-82	1018	198	19.5	55	5.4	243	23.9	385	37.8	62	38
20	Ts52-50	1110	290	26.1	44	4	245	22.1	475	42.8	64	36
21	Ts53-51	1404	175	12.5	118	8.4	445	31.7	621	44.2	58	42
22	Ts54-27	664	227	34.2	111	16.7	201	30.3	399	60.1	66	34
23	Ts54-69	739	289	39.1	132	17.9	102	13.8	357	48.3	56	44
24	Ts55-28	1346	258	19.2	34	2.5	561	41.7	562	41.8	52	48
25	Ts56-00	366	72	19.7	13	3.6	72	19.7	153	41.7	53	47
26	Ts57-00	311	55	17.7	0	0	86	27.7	129	41.4	62	38
27	Ts58-00	304	45	14.8	15	4.9	34	11.2	85	27.9	62	38

Copyright © 2013 SciRes.

Continued

28	Ts59-00	322	72	22.4	24	7.5	88	27.3	139	43.3	56	44
29	Ts59-60	1210	106	8.8	37	3.1	126	10.4	231	19.1	57	43
30	Ts64-57	1234	258	20.9	230	18.6	343	27.8	723	58.6	58	42
31	Ts65-58	1197	439	36.7	130	10.9	157	13.1	480	40.1	62	38
32	Ts65-97	781	160	20.5	95	12.2	208	26.6	391	50	58	42
33	Ts67-61	1146	221	19.3	222	19.4	238	20.8	513	44.8	63	37
34	Ts69-102	532	59	11.1	49	9.2	70	13.2	138	25.9	55	45
35	Ts71-00	372	24	6.5	0	0	75	20.2	98	26.3	56	44
36	Ts72-64	1241	209	16.8	143	11.5	111	8.9	396	31.9	54	46
37	Ts74-83	957	153	16	95	9.9	108	11.3	327	34.2	57	43
38	Ts76-85	809	219	27.1	0	0	66	8.2	258	31.9	62	38
39	Ts80-49	857	220	25.7	110	12.8	115	13.4	369	43.1	53	47
40	Ts81-50	689	121	17.6	86	12.5	144	20.9	263	38.2	49	51
41	Ts82-32	1504	426	28.3	86	5.7	359	23.9	686	45.6	58	42
42	Ts83-28	412	63	15.3	0	0	99	24	156	37.9	60	40
43	Ts84-00	385	26	6.8	67	17.4	42	10.9	112	29.1	61	39
44	Ts85-90	1252	238	19	163	13	269	21.5	511	40.8	61	39
45	Ts88-83	718	115	16.0	115	16.0	137	19.1	315	43.9	63	37
46	Ts91-86	755	64	8.5	90	11.9	143	18.9	253	33.5	60	40
47	Ts92-87	719	243	33.8	59	8.2	209	29.1	290	40.3	52	48
48	Ts94-00	746	120	16.1	12	1.6	139	18.6	194	25.9	57	44
49	Ts96-00	379	79	20.8	35	9.2	153	40.4	196	51.7	65	35
50	Ts101-00	340	136	40	0	0	66	19.4	160	47	65	36
	Average		182.3	22.1	73.8	8.9	173.9	21.0	344.4	41.7	59	41

Copyright © 2013 SciRes.

Supplementary Table 3. RAPD-derived sequences of *Trichobilharzia szidati* cercariae identified by matches to nucleotide (a) and protein (b, c) databases of *Trichobilharzia* spp., *Schistosoma mansoni*, and *S. japonicum*. nLTRs—non-long terminal repeat retrotransposons; LTRs—long terminal repeat retrotransposons; PLEs—Penelope-like elements. The unclear clade assignment of Perere-8 is designated by "?".

(a)

	T. szidati sequences							Nucleotide database sequences			
Fragment (clone) identification	Location of homology region (length), bp	Location of homology subregion, bp	Homology region length, bp	Identity (%)	*Score*	*E*	Gaps	GenBank Acc. Num.	Organism	Location of homology region, bp	Annotation
Ts47-78	755 - 808 (55)	755 - 808	55	83	55.4	9e-11	+	BN000794	*S. mansoni*	52 - 106	5'-untranslated regionofPerere-3nL TRretrotransposon
Ts49-66	396 - 562 (167)	396 - 467	72	79	62.6	8e-11	+	AF442689	*T. ocellata*	181 - 248	Sau3Arepeatedsequence
		396 - 433	38	76	51.8	1e-07	+	AF442688	*T. regenti*	178 - 214	Sau3A repeatedsequence
		471 - 562	92	74	59.0	9e-10	+	AF442689	*T. ocellata*	301 - 389	Sau3Arepeatedsequence
		471 - 562	92	92	60.8	3e-10	+	AF442688	*T. regenti*	295 - 381	Sau3Arepeatedsequence

Copyright © 2013 SciRes.

(b)

Fragment (clone) identification	*T. szidati* sequences									Protein database sequences					
	Location of homology region (length), bp	Frame	Location of homology subregion, bp	Homology subregion length, bp	Identity/ Similarity (%)	*score*	*E*	Stop-codons	Gaps	GenBank Acc. Num.	Organism	Putative retroelement homologue: Location of homology region, aa	ORF (element)	clade	sub-class
Ts47-26	101 - 394 (294)	+2	101 - 394	294	68/80	137.0	3e-40	+	–	AAK14815	*S. mansoni*	915 - 1012	pol (SjR2)	RTE	nLTRs
		+2	101 - 394	294	65/77	132.0	1e-38	+	–	CAX83711	*S. japonicum*	742 - 839	END-RT (SjCHGCS20)		
Ts48-00	1 - 258 (258)	+1	1 - 222	222	57/74	74.7	3e-19	–	–	CAX83707	*S. japonicum*	1599 - 1672	gag-pol (SjCHGCS16)	BEL	LTRs
		+1	1 - 222	222	57/68	72.4	2e-18	–	–	DAA04498	*S. mansoni*	1605 - 1678	pol (Saci-1)		
		+1	7 - 258	252	42/63	65.9	6e-12	–	–	CAX83698	*S. japonicum*	878 - 960	pol (SjCHGCS8)		
		+1	7 - 225	219	43/64	59.7	5e-10	–	–	CAJ00252	*S. mansoni*	1116 - 1188	gag-pol (Saci-6)		
Ts48-65	357 - 488 (132)	+3	357 - 488	132	48/75	47.0	2e-09	+	–	CAJ00245	*S. mansoni*	403 - 446	RT (Perere-8)	?	nLTRs
Ts49-66	11 - 198 (188) region 1	+2	11 - 88 (856 - 880)	78	73/85	43.9	8e-17	–	+	CAX83711	*S. japonicum*	856 - 880	END-RT (SjCHGCS20)	RTE	nLTRs
		+2	11 - 88 (1029 - 1053)	78	77/77	42.4	1e-15	–	+	AAK14815	*S. japonicum*	1029 - 1053	pol (SjR2),		
		+1	94 - 198 (880 - 913)	105	77/86	61.6	8e-17	–	+	CAX83711	*S. japonicum*	880 - 913	END-RT (SjCHGCS20)		
		+1	94 - 198 (1053 - 1086)	105	71/83	58.9	1e-15	–	+	AAK14815	*S. japonicum*	1053 - 1086	pol (SjR2)		
	1006 - 1197 (192) region 2	+1	1006 - 1197 (713 - 768)	192	47/59	38.5	8e-06	+	+	CAX83711	*S. japonicum*	713 - 768	END-RT (SjCHGCS20)		
		+1	1006 - 1197 (886 - 941)	192	50/61	44.7	1e-07	+	+	AAK14815	*S. japonicum*	886 - 941	pol (SjR2)		
Ts51-24	11 - 630 (620)	+2	11 - 250	240	49/60	78.6	9e-51	–	–	CAJ00235	*S. mansoni*	718 - 797	END-RT (Perere-2)	CR1	nLTRs
		+2	41 - 250	210	47/64	68.9	5e-49	–	–	CAJ00242	*S. mansoni*	733 - 802	END-RT (Perere-6)		
		+2	11 - 250	240	46/60	74.7	2e-44	–	–	CAJ00240	*S. mansoni*	688 - 767	END-RT (Perere-5)		
		+1	250 - 630	381	55/74	140.0	9e-51	–	+	CAJ00235	*S. mansoni*	808 - 935	END-RT (Perere-2)		
		+1	250 - 624	375	57/74	144.0	5e-49	–	+	CAJ00242	*S. mansoni*	813 - 938	END-RT (Perere-6)		

Copyright © 2013 SciRes.

Continued

Ts51-49	123 - 1127 (1005)	+1	253 - 624	372	52/66	124.0	2e-44	–	+	CAJ00240	*S. mansoni*	779 - 903	END-RT (Perere-5)	Gypsy/Ty 3	LTRs
		−2	123 - 1127	1005	54/68	337.0	5e-107	+	+	CAX83693	*S. japonicum*	918 - 1254	gag-pol (SjCHGCS3)		
		−2	426 - 1127	702	57/71	245.0	1e-74	+	+	CAJ00251	*S. mansoni*	918 - 1152	gag-pol(Saci-5)		
		−2	204 - 1112	909	27/42	84.7	6e-20	+	+	CAX83694	*S. japonicum*	976 - 1289	gag-pol(SjCHGCS4)		
		−2	150 - 1115	966	28/43	77.0	2e-17	+	+	CAX83691	*S. japonicum*	646 - 959	pol (SjCHGCS1)		
		−2	489 - 1112	624	28/45	76.3	3e-17	+	+	DAA04499	*S. mansoni*	996 - 1210	pol (Saci-2)		
Ts51-82	788 - 1018 (231)	+2	788 - 1018	231	77/86	83.6	2e-20	–	–	AAK14815	*S. japonicum*	879 - 955	pol (SjR2)	RTE	nLTRs
		+2	788 - 1018	231	77/86	83.2	2e-20	–	–	CAX83711	*S. japonicum*	706 - 782	END-RT (SjCHGCS20)		
Ts52-50	598 - 717 (120)	−3	598 - 717	120	35/65	38.5	1e-05	+	–	CAJ00236	*S. mansoni*	820 - 859	END-RT (Perere-3)	RTE	nLTRs
Ts53-51	11 - 612 (602)	−3	11 - 415	405	40/62	119.0	5e-42	–	+	CAJ00244	*S. mansoni*	191 - 326	END-RT (Perere-7)	CR1	nLTRs
		−3	26 - 415	390	43/65	119.0	1e-41	–	+	DAA04497	*S. mansoni*	346 - 476	END-RT (Perere)		
		−3	11 - 415	405	33/55	85.9	2e-28	–	+	CAJ00238	*S. mansoni*	63 - 199	END-RT (Perere-4)		
		−1	433 - 612	180	55/73	71.2	4e-42	–	–	CAJ00244	*S. mansoni*	125 - 184	END-RT (Perere-7)		
		−1	430 - 606	177	54/69	70.1	1e-41	–	–	DAA04497	*S. mansoni*	282 - 340	END-RT (Perere)		
		−1	415 - 600	186	47/63	58.5	2e-28	–	–	CAJ00238	*S. mansoni*	1 - 62	END-RT (Perere-4)		
Ts64-57	655 - 1200 (546)	−2	850 - 1200	351	44/58	80.1	8e-19	+	+	CAJ00236	*S. mansoni*	811 - 921	END-RT (Perere-3)	RTE	nLTRs
		−2	655 - 1200	546	30/48	70.9	8e-16	+	+	CAX83710	*S. japonicum*	824 - 1000	END-RT (SjCHGCS19)		
Ts69-102	11 - 283 (273)	−3	11 - 283	273	33/55	62.4	2e-15	+	–	CAX83714	*S. japonicum*	439 - 529	RT (SjPenelope-2)	–	PLEs
Ts71-00	1 - 367 (367)	+1	1 - 183	183	43/54	44.3	1e-16	+	–	AAK14815	*S. japonicum*	861 - 921	pol (SjR2)	RTE	nLTRs
		+1	1 - 183	183	48/64	51.2	3e-16	+	–	CAX83711	*S. japonicum*	688 - 748	END-RT (SjCHGCS20)		
		+2	230 - 367	138	59/74	57.0	1e-16	–	–	AAK14815	*S. japonicum*	937 - 982	pol (SjR2)		
		+2	230 - 364	135	51/64	48.9	3e-16	–	–	CAX83711	*S. japonicum*	764 - 808	END-RT (SjCHGCS20)		
Ts72-64	289 - 690 (402)	+1	289-690	402	31/47	52.0	6e-10	+	+	CAJ00236	*S. mansoni*	807 - 958	END-RT (Perere-3)	RTE	nLTRs

Copyright © 2013 SciRes.

Continued

Ts74-83	379-774 (396)	+1	379 - 774	396	29/49	52.4	2e-10	+	+	AAY56755	*S. mansoni*	96 - 229	SR3-left	RTE	nLTRs
		+1	379 - 774	396	28/49	49.7	1e-09	+	+	AAY56756	*S. mansoni*	96 - 229	SR3-right		
		+1	379 - 768	390	27/48	48.1	9e-09	+	+	CAJ00236	*S. mansoni*	627 - 758	END-RT (Perere-3)		
		+1	379 - 768	390	26/50	47.4	2e-08	+	+	CAX83710	*S. japonicum*	640 - 771	END-RT (SjCHGCS19)		
Ts76-85	11-316 (306)	+2	11 - 304	294	35/50	51.2	2e-10	+	+	CAJ00236	*S. mansoni*	823 - 921	END-RT (Perere-3)	RTE	nLTRs
		+2	11 - 316	306	29/49	41.6	3e-07	+	+	CAX83710	*S. mansoni*	836 - 938	END-RT (SjCHGCS19)		
Ts82-32	298-540 (243) region 1	+1	298 - 540	243	34/52	48.1	2e-08	+	+	CAJ00236	*S. mansoni*	838 - 921	END-RT (Perere-3)	RTE	nLTRs
	943-1065 (123) region 2	+1	943 - 1065	123	54/83	53.9	3e-10	+	–	CAX83714	*S. japonicum*	111 - 151	RT (SjPenelope-2)	–	PLEs
		+1	943 - 1065	123	44/80	43.1	6e-07	+	–	CAJ00247	*S. mansoni*	200 - 240	RT (Perere-10)		
		+1	943 - 1062	120	53/75	43.1	7e-07	+	–	CAX83715	*S. japonicum*	200 - 239	RT (SjPenelope-3)		
	1047-1442 (396) region 3	+3	1065 - 1439	375	73/83	192.0	8e-56	–	–	CAX83711	*S. japonicum*	666 - 790	END-RT (SjCHGCS20)	RTE	nLTRs
		+3	1047 - 1442	396	69/82	191.0	3e-55	+	–	AAK14815	*S. japonicum*	833 - 964	pol (SjR2)		
Ts83-28	293-388 (96)	+2	293 - 388	96	50/63	41.6	5e-08	–	–	CAX83707	*S. japonicum*	104 - 135	gag-pol (SjCHGCS16)	BEL	LTRs
		+2	293 - 388	96	47/63	40.0	2e-07	–	–	DAA04498	*S. mansoni*	97 - 128	pol (Saci-1)		
Ts85-90	960-1247 (288)	+3	960 - 1244	285	50/62	67.4	6e-15	+	+	CAX83711	*S. japonicum*	711 - 808	END-RT (SjCHGCS20)	RTE	nLTRs
		+3	960 - 1247	288	49/61	66.6	1e-14	+	+	AAK14815	*S. japonicum*	884 - 982	pol (SjR2)		
Ts88-83	56-262 (207)	+2	56 - 220	165	45/64	42.7	3e-08	+	+	CAJ00236	*S. mansoni*	600 - 652	END-RT (Perere-3)	RTE	nLTRs
		+2	59 - 262	204	38/57	40.8	1e-07	+	+	CAX83710	*S. japonicum*	614 - 676	END-RT (SjCHGCS19)		
		+2	56 - 220	165	41/64	39.3	3e-07	+	+	AAY56756	*S. mansoni*	69 - 121	SR3-right		
		+2	56 - 220	165	41/63	37.7	9e-07	+	+	AAY56755	*S. mansoni*	69 - 121	SR3-left		

Copyright © 2013 SciRes.

(c)

		T. szidati sequences							Protein database sequences			
Fragment (clone) identification	**Frame**	**Location of homology region, bp**	**Homology region length, bp**	**Identity/Similarity (%)**	***score***	***E***	**Stop- codons**	**Gaps**	**GenBank Acc. Num.**	**Organism**	**Location of homology region, aa**	**Product**
Ts04-84	−3	34 - 210	177	92/98	117.0	7e-33	–	–	XP_002577399	*S. mansoni*	279 - 337	transcription initiation factor tfiid
Ts67-61	+2	104 - 391	288	69/80	86.3	2e-21	–	+	XP_002571520	*S. mansoni*	885 - 972	zinc finger transcription factor gli2

Copyright © 2013 SciRes.