Parasitol Res
DOI 10.1007/s00436-010-1889-8

SHORT COMMUNICATION

The DL1 repeats in the genome of *Diphyllobothrium latum*

Nadezhda M. Usmanova · Vasiliy I. Kazakov

Received: 27 February 2010 / Accepted: 20 April 2010

Abstract *Diphyllobothrium latum* is a widespread intestinal parasite, which has a great clinical relevance, but there are no sequences of its nuclear genome. In this paper, a repetitive element in the *D. latum* genome is firstly described. The adult *D. latum* was obtained in the result of expulsion from intestinum of a patient suffering from diphyllobothriasis. Genomic DNA was isolated from several proglottids of this individual. *Pst*I restriction products of *D. latum* genomic DNA were sequenced. Polymerase chain reaction (PCR) amplification of these products using genomic DNA and selected primers was carried out. Thereby a cluster of a repetitive element, called DL1, was discovered. For precise identification of a beginning and an end of the repeat, a product of PCR amplification of *D. latum* genomic DNA with one specific primer was sequenced. In discussion, several evidences that DL1 repeat is a member of the SINE family of retroposons were adduced.

Introduction

The broad tapeworm *Diphyllobothrium latum* is an intestinal parasite, which is the most commonly diagnosed cause of diphyllobothriasis in humans. Approximately 20 million people are infected worldwide (Scholz et al. 2009). In spite of the great clinical relevance of this species, just its mitochondrial genome has been sequenced (Nakao et al. 2007). Nothing has been known about its nuclear genome. In this paper, DL1 repeats in the nuclear genome of *D. latum* is firstly described.[1]

Materials and methods

DNA isolation and restriction

The adult *D. latum* was obtained in the result of expulsion from intestinum of a Russian patient. A fragment of the strobile was washed well with physiological saline. Ten immature proglottids were reduced to powder in liquid nitrogen. Lysis buffer, containing 0.15 mg/ml proteinase K, 0.1 M NaCl, 20 mM EDTA, 1% sodium dodecylsulfate, and 50 mM Tris-HCl, pH 8.0, was added to the sample and incubated at 37°C during a night. Genomic DNA was extracted from the lysate using the phenol/chloroform method. Purified DNA was digested with restrictase *Pst*I (Invitrogen). The restriction products were analyzed by electrophoresis in a 1% agarose gel.

DNA cloning

The *Pst*I restriction product ~150 bp was extracted from the agarose gel using High Pure polymerase chain reaction (PCR) Product Purification Kit (Roche) and subcloned into *Pst*I restricted pUC18 plasmid.

PCR products were cleaned using the same kit and subcloned into pTZ57R/T vector using InsTAclone™ PCR Cloning Kit (Fermentas).

N. M. Usmanova (✉) · V. I. Kazakov
Russian Academy of Sciences, Institute of Cytology,
Tikchoretskii Avenue 4,
194064 St. Petersburg, Russia
e-mail: nmusmanova@gmail.com

[1] The nucleotide sequences data reported in this paper have been submitted to GenBank with the accession numbers GQ398237 and GU350458.

Published online: 04 May 2010

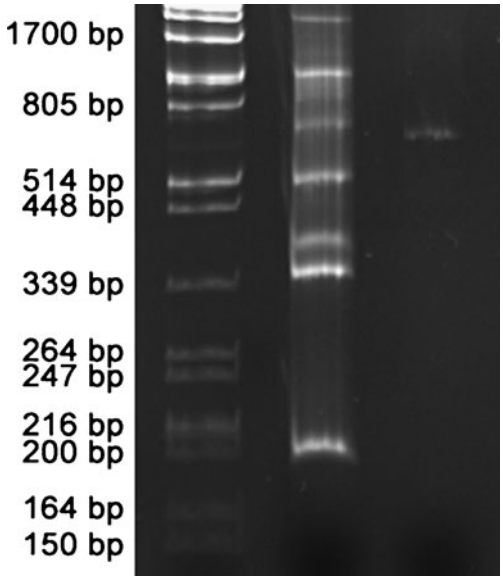

Fig. 1 Separation of PCR products in the 8% polyacrylamide gel. *Lane 1* bacteriophage lambda DNA digested with *Pst*I; *Lane 2* products of PCR amplification of *D. latum* genomic DNA with primers *DL1for* and *DL1rev*; *Lane 3* product of PCR amplification of *D. latum* genomic DNA with primer *DL1for*

DNA sequencing

DNA sequencing was carried out using BigDye® Terminator v3.1 Cycle Sequencing Kit on ABI PRISM® 3130*xl* Genetic Analyzer (Applied Biosystems). For the sequencing of clones, the standard M13/pUC forward and reverse sequencing primers were used.

PCR

PCR primers *DL1for* 5′-tgcaacccctcaacaactac-3′ and *DL1rev* 5′-taaagactcccgacagggta-3′ were selected using Vector NTI Advance™ Software (Invitrogen). They were used for amplification of the DL1 repeats. After amplification, the products were separated by electrophoresis in an 8% polyacrylamide gel.

Computer analysis

For an alignment of sequences we used the program CLUSTALW, available at http://www.ebi.ac.uk/Tools/clustalw2/index.html. Searching proteins using a translated nucleotide query was realized with the program BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi). To find A- and B-box in the sequence of the repeat, we used the program DNA-PATTERN, available at http://rsat.ulb.ac.be/rsat/. A- and B-box consensuses were: A-box, 5′-RRYNNRRYGG-3′, B-box, 5′-GWTCRANNC-3′ (Perez-Stable and Shen 1986). One mismatch with the consensus sequences was allowed.

Results

The restriction of genomic DNA of *D. latum* with *Pst*I led to an appearance of a blurred ~150 bp band on the agarose gel. This product was isolated from the gel, subcloned it into pUC18 plasmid, and three clones were sequenced. It

Fig. 2 Alignment of seven copies of DL1 repeat. Identical nucleotides in columns are marked by *asterisks*

```
DL1 seq 1 CAACAACTACACCATGTTCACCACCCTGTCGGGAGTCTTTAGGTACGGTTTTGGCTACCT 60
DL1 seq 2 CAACAACTACACCATGTTCACCACCCTGTCGGGAGTCTTTAGGTA-GGTTTTGGCTACCT 59
DL1 seq 3 CAACAACTACACCATGTTCACCACCCTGTCGAGAGTCTTTAGGTACGGTTTTGGCTACCT 60
DL1 seq 4 GAACAAATACACCATGTTCACCACCCTGTCGGGAGTCTTTAGGTACGGTTTTGGCTACCT 60
DL1 seq 5 GAACAACTACACCATGTTCACCACCCTGTCGGGAGTCTTCAGGTACGGTTTTGGCTAACT 60
DL1 seq 6 CAACAACTACACCATGTTCACCACCCTGTCGGGAGTCTTTATGTACGGTTTTGGCTACCA 60
DL1 seq 7 GAACAACTACACCATGTTCACCACCCTGTCGGGAGTTTTTAGGTACGGTTTTGGCTACCA 60
          ***** ********************** **** ** * *** ********** *

DL1 seq 1 TTGCTGCAGTGGGGCCTGGAGCTCGAACTGTGTGGTGTGACTGCTTGGTCGTCAGATCTG 120
DL1 seq 2 TTGCTGCAGTGGG-CCTGGAGCTCGAACTGTGTGGTGTGACTGCTTGGTCGTCAGATCTG 118
DL1 seq 3 TTGCTGCAGTGGG-CCTGGAGCTCGAACTGTGTGGTGTGACTGCTTGGTCGTCAGATCTG 119
DL1 seq 4 TTGCTGCAGTGGG-CCTGGAGCTCGAAATGTGTGGTGTGACCGCTTGGTCGTCAGAACTG 119
DL1 seq 5 TTGCTGCAGTGGG-CCTGGGGCTCGAACTGTGTGGTGTGCCTGCTTTGTCGTCAGAACTG 119
DL1 seq 6 TTGCTACAGTGTG-CCTGGAGCTC-AATTGTGTGGTGTGACTGCTTGGTCGTCAGAACTG 118
DL1 seq 7 TTGCTACAGTGTG-CCTGGAGCTA-AATTATGTGGTGTGACTGCTTGGTCGTCAGAACTG 118
          ***** ***** * ***** ***  ** * ********* * **** ********* ***

DL1 seq 1 TCTGAAGTTGCTTCTTGCAACAACT 145
DL1 seq 2 TCTGAAGTTGCTTCTTGCAACAACT 143
DL1 seq 3 TCTGAAGTTGCTTCTTGCAACAACT 144
DL1 seq 4 TCTGAAGTTGCTTCTTTGAACCCCT 144
DL1 seq 5 TCTGAAGTTGCTTCTTGCAACCCCT 144
DL1 seq 6 TCTGAAGTTGTTTGTTGCAACCCCT 143
DL1 seq 7 CCTGAAGTTGTTTGTTGCAACATCT 143
          ********* ** **  ***  **
```

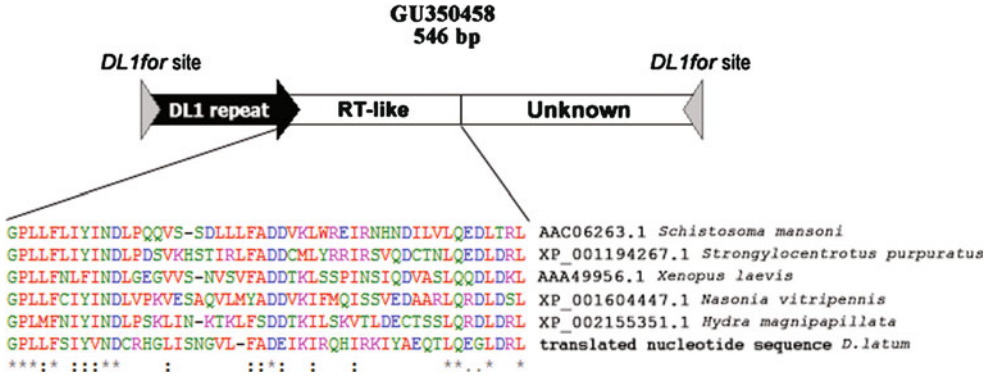

Fig. 3 Structure of the GU350458 sequence. Translated version of RT-like region (reverse transcriptase-like) of this sequence is similar to reverse transcriptases from several animals. *Coloured sequences* are alignment of fragments of these proteins. *Grey triangles* are sites for *DL1for* primer. Identical amino acid residues in columns are marked by *asterisks*. *Colons* mean that conserved substitutions are observed. *Dots* mean that semi-conserved substitutions are observed

turned out that they were very similar. Relying on an alignment of the inserts in these three clones, primers *DL1for* and *DL1rev* for PCR amplification of a fragment of these inserts were chosen. Rated length of the PCR product was approximately 52 bp. After PCR amplification of *D. latum*, genomic DNA with these primers the products were separated by electrophoresis in the polyacrylamide gel (Fig. 1, Lane 2). One can see that several products are generated in the course of the reaction. They look like a ladder on the gel. We suggested that there was some repeat in the *D. latum* genome, and several copies of this repeat could form a cluster. To check this suggestion, these products were subcloned into pTZ57R/T vector, and 11 clones were sequenced. Lengths of the shortest and longest inserts were 51 bp and 625 bp, respectively. Lengths of the others were 193-195 bp (five clones), 338-339 (three clones), and 483 (one clone). The sequence of the longest insert was deposited into GenBank (accession no. GQ398237). The sequences of the 193-625 bp inserts contain the sequence of the shortest insert two to five times, respectively. Spaces between these copies of the shortest insert are extremely similar to each other. Obviously, this is clusters of a repeat. This repeat was named DL1 by us.

To determine an accurate beginning and end of the DL1 repeats in the sequence GQ398237, the *D. latum* genome DNA with the only one primer *DL1for* was amplified. We expected to amplify convergent neighboring copies of the DL1 repeat. The result of the PCR is presented in Fig. 1 (Lane 3). The only major product was excised from the gel, subcloned it into pTZ57R/T vector, and the insert was sequenced. The sequence was deposited into GenBank (accession no. GU350458). It turned out that this sequence contained a fragment, which was extremely similar with repeat units in the GQ398237. Using program CLUSTALW, an alignment of the two sequences was carried out and 5′- and 3′-ends of the DL1 repeat were determined. Several copies of the DL1 repeat in our clones were aligned too (Fig. 2). It is noticeable that a divergence between different copies of DL1 repeat is insignificant.

Besides the DL1 repeat, the sequence GU350458 contains a region of unknown functions. Using the program BLASTX, protein sequences, which were the most similar to translated version of this region, were found (Fig. 3).

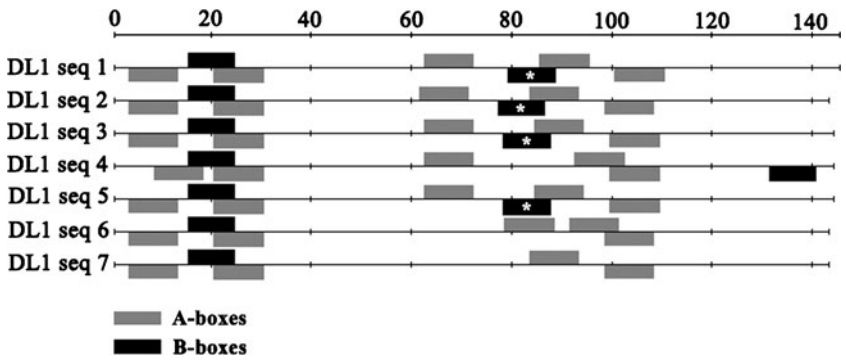

Fig. 4 Locations of A- and B-boxes in seven copies of DL1 repeat. One mismatch with the consensus sequences of these boxes was allowed. *Rectangles above the straight line* are A- and B-boxes in direct orientation; *rectangles under the straight line* are A- and B-boxes in reverse orientation. *Asterisks* mark B-boxes, which possibly initiate transcription of the DL1 repeat

These were proteins and predicted proteins with reverse transcriptase activity from *Xenopus laevis*, *Schistosoma mansoni*, *Strongylocentrotus purpuratus*, *Hydra magnipapillata*, and *Nasonia vitripennis*. Degree of the similarity between each from these proteins and translated nucleotide sequence of *D. latum* reached 40-46%.

Using the program DNA-PATTERN, several A- and B-boxes, which are elements of an internal promoter for RNA polymerase III, were discovered in the sequence of DL1 repeat (Fig. 4).

Discussion

In this work, a repetitive element in the genome of *D. latum* was firstly identified. This fact is not surprising because repeats have been found in genomes of many parasites. For example, genomes of *Schistosoma japonicum* and *S. mansoni* (members of phylum *Platyhelminthes*, as is *D. latum*) comprise 40% repetitive sequence (Berriman et al. 2009; Zhou et al. 2009). It is remarkable that DL1 repeat does not have any essential similarity to other nucleotide sequences deposited in GenBank.

Several signs indicate that DL1 repeat is similar to SINE family of repeats. SINEs are relatively short retroposons: 80–500 bp (typically 150–200 bp). In the majority of cases, SINEs are flanked by short (5–15 nucleotides) direct repeats of host DNA (Kramerov and Vassetzky 2005). One can see (Fig. 2) that the length of the DL1 repeats is 143-145 bp. The first eight nucleotides are similar to last eight ones. They may be those direct repeats of host DNA.

In the GU350458 sequence, DL1 repeat is close to the reverse transcriptase-like region (Fig. 3). It is possible that this sequence is a fragment of reverse transcriptase gene of *D. latum*, and DL1 repeat is located in an intron of this gene. It is known that a reverse transcriptase is essential for amplification of retroposons in genomes (Kramerov and Vassetzky 2005).

All known SINEs have internal promoters for RNA polymerase III, most use the type 2 promoter. This is a split promoter consisted from two regions, A- and B-boxes, separated by approximately 30-60 bp (Perez-Stable and Shen 1986). Computer analysis revealed several A- and B-boxes in DL1 repeat (Fig. 4). On this figure, it is rather difficult to select an appropriate pair of these elements, which could form a Pol III promoter. But only one mismatch with the consensus sequences was allowed. If two mismatches were allowed, the number of A- and B-boxes increased significantly. Moreover, it is known that deletion of the A-box sequence of the Alu repeat (human SINE retroposon) just reduces the efficiency of transcription by 10-20-fold, the remaining B-box is sufficient to direct initiation of transcription at approximately 70 bp upstream (Perez-Stable et al. 1984). It is possible that B-boxes, marked by asterisks on Fig. 4, can initiate transcription of DL1 repeat.

SINE elements have been found in genomes of many organisms, including members of phylum *Platyhelminthes*, e.g., Sjα element of *S. japonicum* (Laha et al. 2000). We suppose, it is quite probable that DL1 repeat is a retroposon of the SINE family of *D. latum*. In addition, the sequence GQ398237 is a cluster of such retroposons.

Acknowledgments Research described in this article was supported by the Program of Presidium of the Russian Academy of Sciences "Molecular and Cellular Biology".

References

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD et al (2009) The genome of the blood fluke *Schistosoma mansoni*. Nature 460:352–358. doi:10.1038/nature08160

Kramerov DA, Vassetzky NS (2005) Short retroposons in eukaryotic genomes. Int Rev Cytol 247:165–221. doi:10.1016/S0074-7696(05)47004-7

Laha T, McManus DP, Loukas A, Brindley PJ (2000) Sjalpha elements, short interspersed element–like retroposons bearing a hammerhead ribozyme motif from the genome of the oriental blood fluke *Schistosoma japonicum*. Biochim Biophys Acta 1492:477–482. doi:10.1016/S0167-4781(00)00112-3

Nakao M, Abmed D, Yamasaki H, Ito A (2007) Mitochondrial genomes of the human broad tapeworms *Diphyllobothrium latum* and *Diphyllobothrium nihonkaiense* (Cestoda: *Diphyllobothriidae*). Parasitol Res 101:233–236. doi:10.1007/s00436-006-0433-3

Perez-Stable C, Shen CK (1986) Competitive and cooperative functioning of the anterior and posterior promoter elements of an Alu family repeat. Mol Cell Biol 6:2041–2052

Perez-Stable C, Ayres TM, Shen CK (1984) Distinctive sequence organization and functional programming of an Alu repeat promoter. Proc Natl Acad Sci USA 81:5291–5295

Scholz T, Garcia HH, Kuchta R, Wicht B (2009) Update on the human broad tapeworm (genus *Diphyllobothrium*), including clinical relevance. Clin Microbiol Rev 22:146–160. doi:10.1128/CMR.00033-08

Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, Huang Z, Zhang B, Huang W, Jin K et al (2009) The *Schistosoma japonicum* genome reveals features of host-parasite interplay. Nature 460:345–351. doi:10.1038/nature08140