Genetic diversity of nuclear ITS1–5.8S–ITS2 rDNA sequence in Clonorchis sinensis Cobbold, 1875 (Trematoda: Opisthorchidae) from the Russian Far East

Yulia V. Tatonova, Galina N. Chelomina*, Vladimir V. Besprosvannykh

Institute of Biology and Soil Science, Far Eastern Branch, Russian Academy of Sciences, 100-letiya Street, 159, Vladivostok 690022, Russia

ARTICLE INFO

Article history:
Received 25 May 2012
Received in revised form 5 July 2012
Accepted 9 July 2012
Available online 21 July 2012

Keywords:
Trematode
Clonorchis sinensis
Genetic diversity
rDNA
ITS
Russia

ABSTRACT

The present study examined the molecular organisation and sequence variation in the nuclear ribosomal DNA (rDNA) region, including the two internal transcribed spacers (ITS1 and ITS2) and the 5.8S gene of the Clonorchis sinensis from the Russian Far East. The relevant sequences from other parts of this species’ area were downloaded from GenBank. The results showed 100% identity for all investigated 5.8S–ITS2 rDNA sequences. In contrast, two levels of intraspecific variations were revealed in the complete ITS1 sequences. The intra-genomic variation resulted from a C/T polymorphism in a single position. The inter-individual differences between the ITS1 sequences were both due to nucleotide and size polymorphisms resulting from a varying number of five-nucleotide repeats and followed by two ITS1 length variants. These variant frequencies correlate with the clonorchiosis level in some geographical localities. ITS1 differences, both in the mutation profile and mutation localisation, were revealed between northern and southern geographical samples. The presence of GC boxes that are identical to known regulatory motifs in eukaryotes was detected within the ITS1 sub-repeats. The predicted secondary structures for ITS1 consist of two large branches, one of which was invariable, while another depended on ITS1 length. The predicted secondary structure for ITS2 includes four helices around the core. The main differences between C. sinensis and other opisthorchids were localised on the tops of helices 2, 3, and 4. A phylogenetic MST reconstruction subdivided all ITS1 sequences into two well differentiated clusters, each with the major widespread ribotype, and showed that ribotype diversity in both Russia and Korea is much lower than in China. The results obtained demonstrate the feasibility of complete ITS1 sequences in C. sinensis population genetics and can be considered as a basis for further studies of the parasite infection because they may help to elucidate the molecular mechanisms of pathogen evolution and adaptation.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Clonorchis sinensis Cobbold, 1875 is the most common human liver fluke in East Asia and causes the important food-borne zoonosis, clonorchiasis. C. sinensis is the cause of a number of human diseases, such as cholecystitis, liver cirrhosis, and pancreatitis. Moreover, at present, this liver fluke is classified as a carcinogen because it increases the risk of cholangiocarcinoma, which is fatal for humans [1–4]. The impact of clonorchiasis on human health is reinforced due to the increase in human activity, including migrations as well as the import of raw, frozen or dried freshwater fish from endemic areas [5,6]. According to a recent report [7], 15–20 million people are infected, and more than 200 million people are at risk for this infection. Clonorchiasis is regularly registered in the south of the Russian Far East, although this region does not belong to areas that are stressed as sources of clonorchiasis [8]. Unlike neighbouring China and Korea, no population genetic studies of C. sinensis have yet been conducted in this region. To investigate the genetic variation of this species in the Russian Far East, we used nuclear rDNA markers that were earlier applied by the authors [9–14] to compare samples from China and Korea.

In eukaryotes, the ribosomal genes of the nuclear genome are known to usually be organised into tandemly repeated clusters. Each cluster contains the genes for the 18S, 5.8S, and 28S rRNA, two internal transcribed spacers (ITS1 and ITS2) that separate these genes, and a non-transcribed external (IGS) spacer separating the transcriptional units. The multiple copies of rDNA in the genome did not evolve independently, which were rather the subject of concerted evolution when the majority of mutations are quickly fixed in population or species while accumulating between species. Therefore, rDNA sequences appear to be nearly identical (“homogenised”) within a given organism and well differentiated between species [15,16]. The intraspecific rDNA variability of animals and helminths in particular is often limited to noncoding spacer sequences [17–24]. Notably, these regions can differ between closely related taxa in the extent of their variations [20,25]. Because the transcribed spacer regions are known to be very important for the correct processing of mature sequences.
rRNA and the biogenesis of active ribosomal subunits, the derivation of the secondary structures is a major step to understand their biological role in more detail [26–28].

The aim of this study was to broaden the molecular and genetic information on the parasitic worm *C. sinensis*. The complete sequences of ITS1–5.8S–ITS2 rDNA from *C. sinensis* samples from the Russian Far East were analysed, and their secondary structures were derived herein for the first time. In theory, such information can be useful for better understanding of the molecular mechanisms of species adaptation and evolution and the maintenance of the ability of the parasite to infect. Additionally, the intraspecific genetic variation was studied, and the phylogeny was inferred based on a combination of data from this study and GenBank to visualise different ribotypes and reveal the genetic structure of the species.

2. Materials and methods

Metacercariae of *C. sinensis* were harvested from the freshwater fish, *Pseudorasbora parva*, collected from two localities in the southern region of the Russian Far East. Metacercariae were orally fed to rats, which were killed approximately 1 month after infection, and the adult flukes were recovered. The specimens were identified using morphometric species–specific characteristics [8].

The liver flukes were washed several times in a physiological solution immediately after dissection and stored in 96% ethanol. Genomic DNA was extracted from 26 individual worms using the HotSHOT technique [29]. The complete sequences of the ITS1–5.8S–ITS2 rDNA were amplified by polymerase chain reaction (PCR) with the following universal primers: BD1 (5′-GTC-GTA-AGG-TTT-CCG-3′), BD2 (5′-TAT-GCT-TAA-G(A)-TCA-CCG-GGT-3′), 3S (5′-GGT-ACC-GTG-GGA-CCA-CTT-GTG-3′) [30], and a new reverse primer R15 (5′-CCA-TTC-TGA-CAG-AC-CC-GT-3′), designed with the Lasergene PrimerSelect program. The PCR was performed in a total volume of 20 μl containing 0.25 mM of each primer pair, 1 μl DNA in water, 1 × Taq buffer, 1.25 mM dNTP, 1.5 mM magnesium and 1.5 unit of Taq polymerase (“Medigen”, Russia). The amplification of the complete ITS1–5.8S–ITS2 region was performed in a GeneAmp 9700 (Applied Biosystems, USA) with a 1-min initial denaturation at 94 °C, 35 cycles of 15 s at 94 °C, 30 s at 55 °C, and 2 min at 72 °C, and a 5-min extension at 72 °C. Negative and positive controls, using both primers, were used.

The PCR products of 26 individuals of *C. sinensis* were sequenced directly using the ABI 3130 genetic analyser (at the Institute of Biology and Soil Sciences FEB RAS) and the ABI Big Dye Terminator v.3.1. Cycle Sequencing Kit (Applied Biosystems, USA). Chromatograms with double C + T peaks were found within 10 ITS1 sequences, suggesting that more than one allele variant of the ribosomal sequence could be present in the amplified ribosomal pool. Therefore, two such samples were amplified with forward primer BD1 and R15. Then, the purified PCR products of the ITS1 spacer region were cloned with theInstAclone PCR Product Cloning Kit (Fermentas, Lithuania) according to the manufacturer’s protocol. The clones were amplified and sequenced with the universal M13-F (−20) (5′-GTA-AAA-CGA-CGG-CCA-GT-3′) and M13-R (−24) (5′-AAC-AGT-TAT-GAC-CAT-G-3′) primers.

The 26 complete ITS1–5.8S–ITS2 sequences and 20 sequences of ITS1 clones that were obtained were deposited in the EMBL GenBank database under accession numbers JQ48576–JQ48601 and JQ48602–JQ48621.

| Table 1 |
| ITS1 sequences of *C. sinensis* investigated in this study. |

<table>
<thead>
<tr>
<th>rDNA region</th>
<th>Geographic origin (locality)</th>
<th>Aligned length, bp</th>
<th>n</th>
<th>GenBank accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS1</td>
<td>Russia (Primorye, Kronshtadtrka)</td>
<td>657</td>
<td>13</td>
<td>JQ48576–JQ48588</td>
</tr>
<tr>
<td>ITS1</td>
<td>Russia (Primorye, Kronshtadtrka)</td>
<td>662</td>
<td>13</td>
<td>JQ48599–JQ48601</td>
</tr>
<tr>
<td>ITS1</td>
<td>Northern China (Guangxi)</td>
<td>662</td>
<td>17</td>
<td>JQ48601–JQ48602</td>
</tr>
<tr>
<td>ITS1</td>
<td>Korea (Gimhae-si, Gurye-gun, Jingu-si)</td>
<td>456</td>
<td>15</td>
<td>JQ48603–JQ48604</td>
</tr>
<tr>
<td>ITS1</td>
<td>Southern China (Jangsu, Anhui, Zhejiang, Guangxi, Guandong)</td>
<td>456</td>
<td>37</td>
<td>JQ48605–JQ48606</td>
</tr>
<tr>
<td>ITS1</td>
<td>Central China (Henan, Hunan, Hubei, Jiangxi)</td>
<td>456</td>
<td>31</td>
<td>JQ48607–JQ48608</td>
</tr>
<tr>
<td>ITS1</td>
<td>Northern China (Jilin, Heilongjiang)</td>
<td>456</td>
<td>14</td>
<td>JQ48609–JQ48610</td>
</tr>
<tr>
<td>5.8S</td>
<td>Korea (Chilgok-gun)</td>
<td>456</td>
<td>2</td>
<td>JQ48511–JQ48512</td>
</tr>
<tr>
<td>5.8S</td>
<td>Russia (Primorye, Kronshtadtrka)</td>
<td>159</td>
<td>13</td>
<td>JQ48576–JQ48588</td>
</tr>
<tr>
<td>5.8S</td>
<td>Korea (Kimhae)</td>
<td>146</td>
<td>1</td>
<td>JQ48599–JQ48601</td>
</tr>
<tr>
<td>5.8S</td>
<td>Northern China (Shenyang)</td>
<td>146</td>
<td>1</td>
<td>JQ48601–JQ48602</td>
</tr>
<tr>
<td>ITS2</td>
<td>Russia (Primorye, Kronshtadtrka)</td>
<td>300</td>
<td>13</td>
<td>JQ48576–JQ48588</td>
</tr>
<tr>
<td>ITS2</td>
<td>Russia (Primorye, Kronshtadtrka)</td>
<td>300</td>
<td>13</td>
<td>JQ48599–JQ48601</td>
</tr>
<tr>
<td>ITS2</td>
<td>Russia (Khabarovskiy Kray)</td>
<td>272</td>
<td>1</td>
<td>JQ48603–JQ48604</td>
</tr>
<tr>
<td>ITS2</td>
<td>Korea (Kimhae)</td>
<td>300</td>
<td>1</td>
<td>JQ48601–JQ48602</td>
</tr>
<tr>
<td>ITS2</td>
<td>Northern China (Shenyang)</td>
<td>300</td>
<td>1</td>
<td>JQ48601–JQ48602</td>
</tr>
<tr>
<td>ITS2</td>
<td>Japan (Okayama)</td>
<td>272</td>
<td>1</td>
<td>JQ48603–JQ48604</td>
</tr>
</tbody>
</table>

n, number of sequences.
Fig. 1. Predicted secondary structure of *C. sinensis* ITS2 transcript created using the program Mfold [37]. Inter-species differences are shown in leaders and pointed with asterisks, squares, and daggers for *Opisthorchis felineus* (EF688142), *O. viverrini* (AY584735) and *Metorchis orientalis* (HM347226), respectively; $dG$, a free energy. The structure conserves the ring model known for ITS2.
respectively. The relevant sequences of this species from other parts of the region (in China, Korea, Vietnam, and Japan) were downloaded from GenBank. Overall, we analysed 29 complete and 143 partial sequences of the nuclear ITS1–5.8S–ITS2 region and 20 clones of the ITS1 spacer (Table 1). The ribosomal DNA sequences were assembled manually and aligned using the Clustal X program [32]. Nucleotide (Pi) diversity, gene fixation coefficient (Fs) and sequence differences (D) were estimated using DnaSP version 5.10 [33] and MEGA version 5.1 [34]. The distribution pattern of nucleotide diversity was analysed by means of “sliding-window” method [35] using DnaSP version 5.10 [33]. Phylogenetic relationships were analysed using neighbour-joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) methods implemented using MEGA version 5.1 [34]. The ITS1 sequences of Metorchis bilis (EU038154) and Metorchis orientalis (HM347226) were used as an outgroup to root the trees. Minimum spanning trees (MST) developed especially for intraspecific investigations were constructed using ARLEQUIN version 3.11 [36]. Both short tandem and long repeats were identified with the program UGENE version 1.10 (http://ugene.unipro.ru). The folding of the sequences into putative secondary structures was performed with Mfold version 3.0 [37] (http://mfold.rit.albany.edu) at a fixed temperature of 37 °C, and the structure chosen had the highest negative free energy.

3. Results

3.1. 5.8S–ITS2 rDNA analyses

All 26 complete sequences of the 5.8S–ITS2 rDNA regions obtained in this study were completely identical. The size and GC-content of 5.8S and ITS2 were estimated to be 159 bp and 54.4% and 300 bp and 51.2%, respectively. No differences were identified when both the complete and partial sequences of these regions belonging to samples from other Asian countries were added from GenBank. Small di-, tri- and tetra–nucleotide repeats were detected throughout ITS2. In total, 17 simple sequence repeats were identified, of which (TTG)2 was detected in three loci, and each of (CC)2, (GT)3, (GG)2, (TT)2 and (GTT)2 was detected in two loci.

The predicted secondary structure of the ITS2 region (Fig. 1) demonstrated a high similarity with that of other ophiuroids. The putative model includes four helices around the core structure, one of which was relatively long. The main inter-specific differences were located within three regions on the tops of helices 2, 3, and 4.

3.2. ITS1 rDNA analyses

The complete ITS1 region demonstrated two levels of intra-specific variations, which were inter-individual and intra-genomic. The GC content was 54.2%, and the average pairwise sequence difference (D) was estimated to be approximately 0.004. Table 2 demonstrates that the nucleotide diversity of the Russian samples is considerably lower compared with other Asian countries based on both the partial and complete ITS1 sequences. However, gene fixation between geographical regions was estimated to be quite low.

The intra-individual, i.e., intra-genomic, variation estimated in the ITS1 sequences of the Russian samples resulted from a nucleotide polymorphism (C ↔ T transition) in the 114 bp position. The frequencies of C, C/T and T were estimated after direct sequencing to be approximately 46.2%, 38.4% and 15.4%, respectively. After molecular cloning, a single C or T nucleotide was only detected in this position. The C frequency in the 114 bp position was 61% and 80% for individuals from China and Korea, respectively.

The inter-individual variation (Fig. 2) was due to both nucleotide polymorphisms and size variation. The main type of mutation was a C ↔ T transition (53%); 27% of the mutations were C → A, T ↔ A, G → T and G ↔ C transversions, and 13% and 7% were insertions and deletions, respectively. These mutations were found in 15 polymorphic sites, 5 of which were informative. The nucleotide diversity within ITS1 is distributed unequally, with two highly polymorphic sites (in the 114 and 507 bp positions) (Fig. 3).

Short (di-, tri-, tetra-, penta- and hexa-nucleotide) repeats or degenerate versions of these repeats were detected throughout ITS1 (Fig. 2, Table 3). In total, 11 simple sequence repeats were identified, of which 8 were invariant, and two – (TG)2n and (TG)n – were detected at more than one locus. In addition, three long, 47–50 bp direct repetitive elements with 75.5–88% identity (Fig. 4a–c) and one long, 30-bp inverted repeat with 60% homology (Fig. 4d) were detected within the ITS1 (Fig. 4). These sub-repeats were predicted to include three regulatory GC boxes with sequences of GGGCGG and CCGCCC (Fig. 4).

Size variation was due to a varying number of a five-nucleotide GCCCTG repeat that resulted in different lengths of the ITS1 variants, estimated to be 657 and 662 bp in size. With a few exceptions, this mutation is associated with a G/C transversion at position 507 bp, implying its compensatory characteristic. The frequency of this mutation, based on the combined complete and partial ITS1 sequence data (n = 143), varies widely among geographic localities, from 4% in Russia to 47% in Korea, and tends to increase towards southern (13%), central (32%) and northern (36%) China. The nucleotide diversity was estimated to be higher among 662-bp ITS1 sequences. The differentiation between the ITS1 sequences with different numbers of five-nucleotide repeats appeared to be significant (Table 2).

Based on both the mutation profile and mutation localisation, all the aligned variable sequences (40 of 143) were subdivided into two groups, conventionally named as the “northern” and “southern” groups (Fig. 5). The first group comprises mainly localities from Northern China and Korea and is characterised by only C ↔ T transitions upstream of the additional five-nucleotide repeat. The second group includes sequences only from the southern localities and is characterised by different types of transitions and transversions downstream of the additional 5-nucleotide repeat. There is a single exception, found in the sequences from the Heilongjiang (Northern China) sample, which is characterised by both types of variability. No such sequence variations were identified within the Russian samples. The nucleotide diversity of the southern group was almost two times higher than that of the northern group; however, no significant gene fixation was estimated between these groups (Table 2).

| Table 2 |
| Genetic diversity parameters estimated for ITS1 sequences of C. sinensis. |
| Samples | n | Pi (±S.D.) | Fst |
| Complete sequences | 59 | 0.00212 (±0.00021) | 0.30 |
| Russia | 26 | 0.00101 (±0.00014) | 0.30 |
| China | 18 | 0.00339 (±0.00041) | 0.30 |
| Korea | 15 | 0.00192 (±0.00039) | 0.30 |
| Russia/China, Korea | 26/33 | 0.00136 (±0.00020) | 0.30 |
| Repeat | 41 | 0.00169 (±0.00041) | 0.30 |
| Repeat | 18 | 0.00169 (±0.00041) | 0.30 |
| Repeat | 41/18 | 0.00169 (±0.00041) | 0.30 |
| Partial sequences | 143 | 0.00335 (±0.00038) | 0.87 |
| Russia | 26 | 0.00033 (±0.00020) | 0.87 |
| China | 100 | 0.00424 (±0.00047) | 0.87 |
| Korea | 17 | 0.00167 (±0.00028) | 0.87 |
| Russia/China, Korea | 26/117 | 0.00171 (±0.00036) | 0.87 |
| Repeat | 108 | 0.00390 (±0.00030) | 0.87 |
| Repeat | 35 | 0.00390 (±0.00030) | 0.87 |
| Repeat | 108/35 | 0.00390 (±0.00030) | 0.87 |
| Northern group | 11 | 0.00469 (±0.00047) | 0.87 |
| Southern group | 28 | 0.00800 (±0.00040) | 0.87 |
| Northern/southern groups | 11/28 | 0.00000 (±0.00000) | 0.87 |

Repeat +, repeat −, sequences with and those without additional five-nucleotide repeat; n, number of sequences; Pi, nucleotide diversity (± S.D.); Fst, coefficient of gene fixations. |
Fig. 2. Alignment of the complete *C. sinensis* ITS1 sequences. Short repeats are in bold on a grey background.
The predicted secondary structures for the complete ITS1 transcripts (Fig. 6) consist of two complex organised large branches separated by a relatively long helical structure. One of the branches was invariable, while the other depended on the length of ITS1. The variable branch involves 3′ and partial 5′ ends of ITS1 transcripts and includes a small open loop. Differences between putative models are estimated to be less pronounced for the longer ITS1 sequences.

3.3. Phylogenetic analyses

The topologies of the NJ, MP, and ML trees (data not shown) for both partial and complete sequences were unresolved, without significant phylogenetic branches or clusters of ribotypes corresponding to sampling localities. Therefore, to further investigate the shallow phylogeny of the species we used MST reconstructions, an approach explicitly developed for intraspecific data. The phylogenetic MST trees based on both complete (data not shown) and partial (Fig. 7) sequences of the ITS1 spacer divided all the sequences into two groups, those with and those without the additional five-nucleotide repeat, which are connected through the unique nucleotide sequence from Russia (with a substitution in position 507 bp but without the additional repeat). Both reconstructions demonstrated two major and widely spread ribotypes and a much higher ribotype diversity in China compared with those from both Russia and Korea. There are 13 and 18 ribotypes among the complete and
partial ITS1 sequences, respectively. Additionally, MST based on partial ITS1 sequences, i.e., sequences that are more numerous and covering a wider species’ area, revealed a few well-differentiated ribotypes among the samples from southern China.

4. Discussion

In trematodes, intraspecific variation in the ITS sequences is minimal [12,38], and the ITS2 has been shown to be a sensitive marker at the species level [25,31]. The study of genetic variation within C. sinensis in Korea and China demonstrated a near identity of the ITS rDNA sequences [9,10,12]. Moreover, no differences were estimated in ITS2, and only 15 nucleotide substitutions were identified in ITS1 when modern and ancient samples have been compared [11]. The low level of intraspecific variation was also observed in the ITS2 of the closely related species, Opisthorchis viverrini [12,39]. In the present study, the lack of intraspecific variability in the ITS2 sequences was confirmed, and new information about the intraspecific variation in the ITS1 sequences were described by studying the liver fluke individuals from new geographically isolated populations in the most northeastern marginal part of the C. sinensis range.

For many organisms, the ITS2 is organised into four helices around a preserved central core [40–42]. The derived structure demonstrated a high similarity with those of other opisthochid flukes: Metorchis orientalis (HM347226), O. viverrini (AY584735) and Opisthorchis felineus (EF688142). Such conservation should be maintained by natural selection, implying a high functional significance of the primary and secondary structures of this non-coding rDNA region. Although the biological role of the ITS2 spacer is not well understood, their importance for the production of the mature rRNA and for ribosomal assembly was shown by certain changes in the secondary structure leading to inter-individual ITS1 size variation were found in ITS1 when modern and ancient samples have been compared [11]. The low level of intraspecific variation was also observed in the ITS2 of the closely related species, Opisthorchis viverrini [12,39]. In the present study, the lack of intraspecific variability in the ITS2 sequences was confirmed, and new information about the intraspecific variation in the ITS1 sequences were described by studying the liver fluke individuals from new geographically isolated populations in the most northeastern marginal part of the C. sinensis range.

For many organisms, the ITS2 is organised into four helices around a preserved central core [40–42]. The derived structure demonstrated a high similarity with those of other opisthochid flukes: Metorchis orientalis (HM347226), O. viverrini (AY584735) and Opisthorchis felineus (EF688142). Such conservation should be maintained by natural selection, implying a high functional significance of the primary and secondary structures of this non-coding rDNA region. Although the biological role of the ITS2 spacer is not well understood, their importance for the production of the mature rRNA and for ribosomal assembly was shown by certain changes in the secondary structure leading to inter-individual ITS1 size variation were found in ITS1 when modern and ancient samples have been compared [11]. The low level of intraspecific variation was also observed in the ITS2 of the closely related species, Opisthorchis viverrini [12,39]. In the present study, the lack of intraspecific variability in the ITS2 sequences was confirmed, and new information about the intraspecific variation in the ITS1 sequences were described by studying the liver fluke individuals from new geographically isolated populations in the most northeastern marginal part of the C. sinensis range.

Table 3

Short nucleotide repeats detected in ITS1 variants of C. sinensis.

<table>
<thead>
<tr>
<th>n</th>
<th>AC-CA</th>
<th>GCCT</th>
<th>GCCTG</th>
<th>TGC</th>
<th>TA</th>
<th>TG</th>
<th>CG</th>
<th>TGG</th>
<th>TGGCC</th>
<th>TGT</th>
<th>CGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>ACACA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AAAA (GCCT)₁</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ACACA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ACACA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ACACA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AACCA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ACACA (GCCT)₂</td>
<td>GCCTG</td>
<td>(TGC)₂</td>
<td>(TA)₃</td>
<td>(TG)₁</td>
<td>(CG)₂</td>
<td>(TGG)₂</td>
<td>TGGCC</td>
<td>(TGT)₃</td>
<td>(CGG)₂</td>
<td></td>
</tr>
</tbody>
</table>

n, number of sequences; positions of repeat motifs are in brackets; and mutations within repeats are in italic.

be a valuable tool to distinguish new species [43,44]. The secondary structure models for C. sinensis ITS1 and ITS2 transcripts were derived herein for the first time.

Although there is little information about the ITS1 secondary structure, in most eukaryotes, these structures consist of an open multi-branched loop with several helices [45,46]. The sequence variation differs across the ITS1 spacer, and this region has been found to be conserved at the 3’ end, indicating the presence of functional constraints [25,47–49]. In fact, the 3’ end of the ITS1 in yeast rDNA has been revealed to contain recognition sites and conserved secondary structure motifs that are essential for RNA maturation [50].

In trematodes, ITS1, having less functional limitations, usually shows greater variability, both in its size and nucleotide sequence [51]. In the present study, two intra-genomic ITS1 variants with a single difference were found in the liver fluke C. sinensis, for the first time to our knowledge. The occurrence of a fixed variation in different clones and individuals among products of independent PCR excludes the possibility that they resulted from random PCR artefacts. The pronounced intragenomic variation of ITS1 is known in different organisms, including helminthes, e.g., Paragonimus westermani [25], Schistosoma japonicum [47], and Atractolyctesurus hurenosus [23]. The intragenomic rDNA polymorphism is usually associated with such genetic phenomena as polyploidy, introgressive hybridisation, parthenogenesis, and multiple rDNA loci [23,47,52]. However, the C. sinensis karyotype was recently reported to be characterised by the only nucleolus organiser region in one of the small chromosomes [53]. In the parthenogenetic reproduction of trematodes, the larval stages are associated with high replicative activity, which can also lead to heterogeneity in the ITS1 regions [23,54].

The internal repeats appear to be characteristic of the ITS1 evolution in different groups of organisms. Their evolution is subject to replication slippage, unequal crossing over, and biased gene conversion, which can generate variability or, in contrast, homogenise the sequences, i.e., lead to concerted evolution [55]. Both long and short repeats leading to inter-individual ITS1 size variation were found across a range of organisms, including helminthes, such as trematodes [25,47,56], nematodes [24], and cestodes [57]. However, both size and nucleotide variations in the C. sinensis ITS1 were estimated to be low enough that they can be evidence of the strong mechanisms of rDNA concerted evolution within its genome.

A number of sequence motifs are well known to be required to initiate and regulate ribosomal genes. There are three well-known regulatory motifs in eukaryotes that can act as promoters: the TATAAT, CCAAT and GC boxes. The GC boxes contain the sequences GCGG, CCCGCC, and CCGCCC. Regulatory motifs often occur in tandemly repeated sequences in the external transcribed spacer; however, they might also be occurring in other regions, such as ITS1, to increase their activity through the copy frequency [58]. Such multiple regulatory repeats, with a broad functional role suggested for them, were identified in ITS1 of a number of species belonging to the three trematode genera: Paragonimus, Schistosoma and Dolichosaccous [58].

Similar to our results, C. sinensis individuals from Heilongjiang (northern China) were found to be genetically closer, with fewer
polymorphisms than those collected from the Guangdong (southern China) province using the random amplified polymorphic DNA (RAPD)-PCR and mobile genetic elements (MGEs)–PCR techniques [59]. The authors suggested that the genetic variation of C. sinensis occurs in the subtropical region faster than that in colder regions due to the occurrence of more generations (life cycles).

A heavy endemic area is widely distributed in northern China (Heilongjiang Province) and Korea [7,60]. These territories were characterised in our study by certain peculiarities of the ITS1 sequence variation. Additionally, samples from Russia, where the situation with clonorchiasis is not as difficult, differed well from those of the adjacent regions of China and Korea with respect to the ITS1 variation pattern. Modelling the secondary structure implies that an additional repeat can stabilise the secondary structure and reduce the folding energy for ITS1; thus, such ribotypes could be more successful for the evolutionary adaptation of the liver fluke.

The available data suggest that different types of ITS1 polymorphisms may have some evolutionary and adaptive significance, including that of C. sinensis infectivity. Although the role of repetitive elements in rRNA gene sequences is still unknown, their contribution to parasite adaptive

Fig. 4. Variants of long sub-repeats within the ITS1 sequences of C. sinensis identified with the program UGENE (http://ugene.unipro.ru). a, b, and c, direct sub-repeats with 88%, 75.5%, and 76.6% identities, respectively; d, inverted sub-repeat with 60% identity. Arrows indicate the localisation of sub-repeats within the ITS1 region and their direction relative to each other. Substitutions between sub-repeats are in italic. GC boxes are in bold on a grey background.

Fig. 5. Variable sites for partial ITS1 sequences of C. sinensis. G/C transversion in the position 507 bp are in bold; transversions in this position, which are not accompanied by an increase in length, are in bold and italic; variable areas upstream and downstream the additional five-nucleotide repeat are in grey; and sequences marked with asterisks were found in samples of northern China.
evolution was recently discussed [24, 61]. The evidence that the ITS1 provides information about the virulence of the species was obtained for the Leishmania species [62]. The potential functional role for the ITS1 in the stage- or tissue-specific regulation of ribosomal gene transcription in helminths was suggested based on regulatory motif analysis [58]. In the genus Acanthamoeba, the ITS1 sequence variabilities correlate with the 18S rDNA sequence types [51]. In the malaria parasite, parasitic insects and vertebrates, two different genes of the small subunit ribosomal RNA are used at various stages of the life cycle. This report was the first and, thus far, only case of birth-and-death evolution in an rDNA gene family [63].

Thus, despite evolutionary conservatism, the selected markers provide new types of data to complement our understanding of the genetic diversity and molecular organisation of the sequences within

---

*Fig. 6.* Predicted secondary structures of the C. sinensis ITS1 transcripts created using the program Mfold [37]. Five-nucleotide repeats are in brackets. \( dG \), a free energy. Differences among the structures are only in the lower branches. The lowest free energy secondary structure (\( dG = -246.20 \) kcal/mol) is predicted for ITS1 sequences with additional five-nucleotide repeat.
rDNA cluster of C. sinensis and provided the basis to search for a relationship between the infectivity of the parasite and the structural organisation of the non-coding sequences of rDNA.

Acknowledgements

This work was partially funded by a grant from the Far Eastern Branch of RAS (projects nos. 11-III-D-06-014 and 12-I-06-022).

References