Available online at www.sciencedirect.com
ScienceDirect
Molecular & Biochemical Parasitology 154 (2007) 190–194
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Short communication

The complete mitochondrial DNA sequence of the monogenean *Gyrodactylus thymalli* (Platyhelminthes: Monogenea), a parasite of grayling (*Thymallus thymallus*)

Laetitia Plaisance [a,*], Tine Huyse [b], D.T.J. Littlewood [b], Tor A. Bakke [a], Lutz Bachmann [a]

[a] *The Natural History Museum, Department of Zoology, University of Oslo, P.O. Box 1172, Blindern, NO-0318 Oslo, Norway*
[b] *Parasitic Worms Group, Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*

Received 25 January 2007; received in revised form 18 April 2007; accepted 18 April 2007
Available online 25 April 2007

Abstract

We present the complete mitochondrial (mt) genome of *Gyrodactylus thymalli*, a monogenean ectoparasite on grayling (*Thymallus thymallus*). The circular genome is 14788 bp in size and includes all 35 genes recognized from other flatworm mt genomes. The overall A + T content of the mt genome is 62.8%. Twenty regions of non-coding DNA ranging from 1 to 111 bp in length were identified in addition to 2 highly conserved large non-coding regions 799 bp and 767 bp in size. Compared to the recently described mt DNA of the closely related *G. salaris* from Atlantic salmon from Signaldalselva, Norway, the mitochondrial genome of *G. thymalli* from Hnilec, Slovakia, differs on average by 2.2%.

Keywords: Fish parasites; Ectoparasites; Gyrodactylosis; Mitogenomics; Sequence annotation

Gyrodactylus thymalli Žitňan, 1960, is a relatively harmless monogenean ectoparasite infecting grayling (*Thymallus thymallus*). Morphologically and genetically, *G. thymalli* closely resembles *G. salaris* Malmberg, 1957, a gyrodactylid that is highly pathogenic to Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) stocks with devastating ecological and economical effects. More than 45 watercourses in Norway alone have been infected with *G. salaris* and many salmon stocks are threatened [1,2]. More than 15% of the Norwegian wild salmon catch in Norway has been lost since the introduction of the parasite and the costs from loss of fish, decline in fish tourism and the need to survey and eradicate infected stocks amount to over USD 50 million per year [3]. It is thus not surprising that *G. salaris* on salmon has been studied intensively, while *G. thymalli* on grayling has attracted only little attention, although comparative analyses may provide clues to a better understanding of gyrodactylosis of salmonid fish. An unambiguous discrimination of *G. thymalli* and *G. salaris* has proven difficult by use of pure morphometrical methodology [4], but molecular systematics has provided additional tools to identifying the parasites. The nucleotide sequences of the internal transcribed spacers (ITS) 1 and 2 of the nuclear ribosomal DNA (∼1200 bp) are identical for many strains of *G. salaris* and *G. thymalli* [5] and support the con-specificity of both species. Sequences of the intergenic spacer (IGS) of the nuclear ribosomal DNA, however, were initially considered supporting the species status of *G. salaris* and *G. thymalli* [6], but a more comprehensive study including 39 populations [7] challenges this interpretation. Recently, a 820 bp segment of the mitochondrial cytochrome oxidase I gene has been sequenced for many geographical isolates of *G. salaris* and *G. thymalli* in order to resolve the species' taxonomy and phylogeography [1,7,8]. Substantial sequence diversity has been detected and phylogenetic analyses are neither supporting monophyly of *G. thymalli* nor of *G. salaris*. Instead, there are several well-supported clades of haplotypes but the phylogenetic relationships of these clades could not be resolved. Currently there is, with the exception of host preference, apparently no reliable method for distinguishing *G. salaris* and *G. thymalli*. As mitochondrial (mt) DNA typically evolves faster than nuclear DNA, mitochondrial sequences may provide powerful markers [9] for

* Corresponding author. Present address: Scripps Institution of Oceanography, Marine Biology Research Division, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0202, USA. Tel.: +1 858 822 5388; fax: +1 858 822 1267.
E-mail address: lplaisance@ucsd.edu (L. Plaisance).

0166-6851/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2007.04.012

further elucidating the more recent evolutionary histories of *G. salaris* and *G. thymalli*. Recently, the mt genome of one isolate of *G. salaris* has been published [10]. In this study, we report the complete mtDNA sequence of *G. thymalli* from graylings. Along with the mtDNA sequence of *G. salaris* we provide an important source for the further development of molecular markers suitable for species or strain identification.

The complete mtDNA of *G. thymalli* from Hrable at River Hnilec, Slovakia, has been determined. From this type locality most specimens used for the species description of *G. thymalli* have been collected [11]. Total DNA was extracted from a single parasite specimen that had been stored in 95% ethanol using the DNeasy Tissue kit (Qiagen). The DNA was subsequently concentrated to a final volume of 15 μL using Microcon-100 columns (Millipore). Two long fragments amplified with the primer pairs Mit4F 5′-AGCTAGGAAAAGTCACAGTGCCAGC-3′–Mit6R 5′-GTACGCCTCTGAGTCCAATGCTAGG-3′ and Mit3R 5′-TGGCATCAATAGCCAAGCCCTTAAAGC-3′–Mit5F 5′-ATAGTCGGTGGGTTCGGTGTAACC-3′ both approximately 6500 bp long and a shorter fragment amplified with the primer pair ZMO4F 5′-GTGACAGGGATAGTGCTATCCTC-3′ [1]–Mit2R 5′-TAACCGCAGCTGCTGGCACTGTG-3′ represent the entire mitochondrial genome of *G. thymalli* in three overlapping fragments. The amplicons were purified with ExoSAP-IT™ (GE Healthcare) and both strands were sequenced by means of primer walking on an ABI3100 automated sequencer using BigDye chemistry (Applied Biosystems).

The mitochondrial genome of *G. thymalli* comprises of 14,788 bp (GenBank accession number: EF527269). Sequences were assembled and edited using Sequencher™ 4.5 (GeneCodes) and annotated with MacVector® 7.2.2 (Accelrys). The sequence analysis of the mtDNA of *G. thymalli* revealed the typical genes found in flatworm genomes (Fig. 1), i.e. 12 protein-coding, 22 tRNA, and the large and small subunit rRNA genes (Table 1). The *atp8* gene is missing as in other flatworm mitochondrial genomes [12], and all mtDNA genes are transcribed from the same strand [13]. The arrangement of mitochondrial genes (Fig. 1) matches those of most other neodermatan species, (i.e. all except the African and Asian species of *Schistosoma* in which gene rearrangements have taken place [12,14]) and the closely related *G. salaris*. A total of 20 short non-coding regions with a length ranging from 1 bp (between $tRNA^{Ile}$ and $tRNA^{Lys}$, $tRNA^{Lys}$ and rrnS, and $tRNA^{Leu2}$ and $tRNA^{Arg}$) to 111 bp (between *cox2* and $tRNA^{Glu}$ and *nad5* and $tRNA^{Gly}$, respectively) are found in the *G. thymalli* mtDNA making up a total of 197 nucleotides. Five overlapping regions are found in the *G. thymalli* genome ranging from 1 bp (*cox3* and $tRNA^{His}$, $tRNA^{Val}$ and $tRNA^{Ala}$, and $tRNA^{Thr}$ and *rrnL*) up to 28 bp (*nad4L* and *nad4*) (Table 1). This particular overlap between

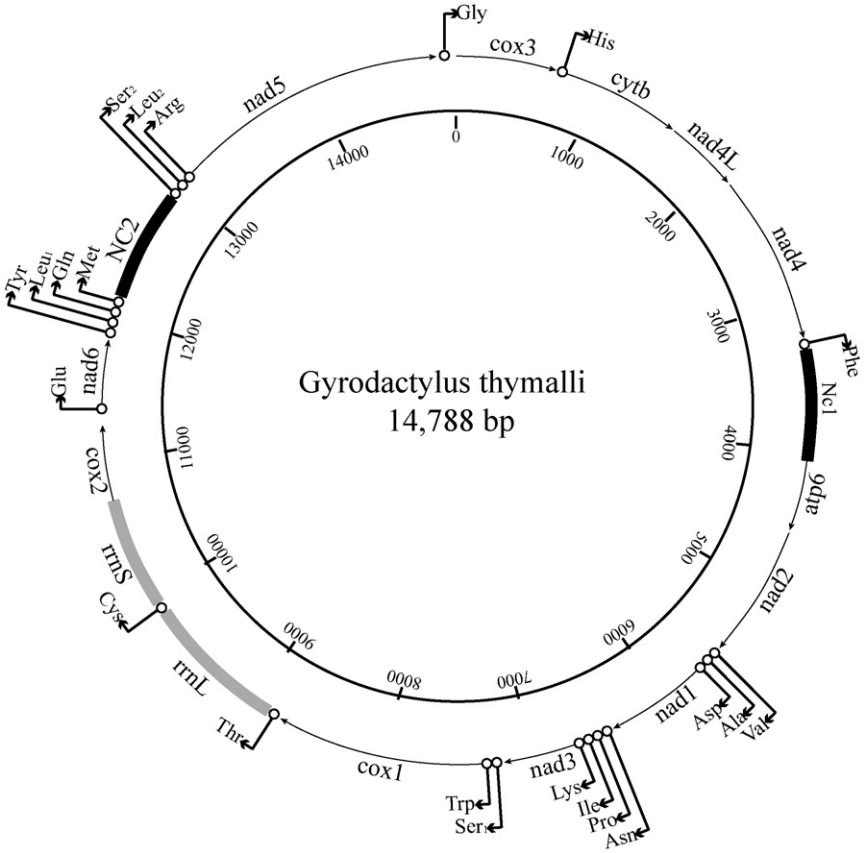

Fig. 1. Graphical representation of the mtDNA of *G. thymalli* (GenBank accession number: EF527269). The abbreviations for the genes are as follow: *atp6* refers to the ATP synthase; *cox1*, *2*, *3* refer to the cytochrome oxidase subunits; *cytb* refers to the cytochrome *b*; and *nad1–6* refers to the nicotinamide dehydrogenase subunits. tRNAs are denoted as three-letter symbol according to the IUPAC-IUB amino acid code. The two long non-coding regions are denoted as NC1 and NC2.

Table 1
Localization of genes and non-coding regions in the mitochondrial genome of *G. thymalli* and sequence similarity in comparison to the closely related *G. salaris* [10]

Annotation	Relative position	Length (bp)	Length (aa)	AT-content	*p* distance to DQ988931
cox3	1–639	639	213	60.9	0.036
$tRNA^{His}$	639–701	63		73.0	0.016
cytb	705–1778	1074	358	60.9	0.024
nad4L	1784–2032	249	83	67.9	0.032
nad4	2005–3213	1209	403	60.1	0.035
$tRNA^{Phe}$	3216–3281	66		74.2	0
NC1	3282–4080	799		63.9	0.004
atp6	4081–4593	513	171	61.0	0.012
nad2	4600–5456	857	286	63.0	0.032
$tRNA^{Val}$	5457–5521	65		75.8	0
$tRNA^{Ala}$	5520–5589	70		64.3	0
$tRNA^{Asp}$	5590–5656	67		74.6	0.015
nad1	5657–6544	888	296	54.8	0.020
$tRNA^{Asn}$	6554–6620	67		62.7	0.015
$tRNA^{Pro}$	6626–6691	66		78.8	0
$tRNA^{Ile}$	6685–6756	72		62.5	0
$tRNA^{Lys}$	6758–6823	66		59.1	0
nad3	6827–7177	351	117	59.8	0.009
$tRNA^{Ser1}$	7178–7237	60		66.7	0.017
$tRNA^{Trp}$	7242–7304	63		61.9	0
cox1	7309–8856	1548	516	58.9	0.032
$tRNA^{Thr}$	8866–8931	66		69.7	0.015
rrnL	8931–9888	958		70.7	0.010
$tRNA^{Cys}$	9889–9948	60		70.0	0
rrnS	9950–10659	710		64.7	0.021
cox2	10660–11241	582	194	63.1	0.029
$tRNA^{Glu}$	11353–11424	72		66.7	0
nad6	11428–11910	483	161	66.9	0.019
$tRNA^{Tyr}$	11914–11980	67		73.1	0.030
$tRNA^{Leu1}$	11981–12047	67		67.2	0.017
$tRNA^{Gln}$	12048–12110	63		63.5	0.016
$tRNA^{Met}$	12117–12182	66		75.8	0
NC2	12183–12949	767		64.5	0.012
$tRNA^{Ser2}$	12950–13012	63		68.3	0.016
$tRNA^{Leu2}$	13018–13085	68		63.2	0
$tRNA^{Arg}$	13087–13153	67		59.7	0
nad5	13154–14704	1551	517	59.2	0.030
$tRNA^{Gly}$	14718–14775	58		77.6	0.015

GenBank accession number DQ988931. The abbreviations for the genes are as follow: *atp6* refers to the ATP synthase; *cox1*, *2*, *3* refer to the cyotchrome oxidase subunits; *cytb* refers to the cytochrome *b*; and *nad1–6* refers to the nicotinamide dehydrogenase subunits. tRNAs are denoted as three-letter symbol according to the IUPAC-IUB amino acid code. The two long non-coding regions are denoted as NC1 and NC2.

nad4L and *nad4*, although variable in length, appears to be common among metazoan mtDNAs [13]. However, a 28 bp overlap has also been identified in some digenean species [14]. Two long non-coding regions (NC1 and NC2) could be detected. However, an AT-rich region as found in many other mitochondrial genomes [15] that is usually corresponding to the control region containing control elements for replication and transcription could not be found.

The mtDNA of *G. thymalli* is moderately AT-rich, i.e. 30.0% A, 32.8% T, 20.2% G, and 17.0% C. The AT content of 62.8% is thus as observed in other flatworm mitochondrial genomes. The most AT-rich sections correspond to the $tRNA^{Gly}$ and $tRNA^{Pro}$ two rRNA genes (77.6 and 78.8%).

In mitochondrial genomes, it is frequently the third codon positions in protein-coding genes that are particularly AT-rich [16] because it is assumed that these positions are under low purifying selection. In *G. thymalli* this is not the case and the third codon positions show a low AT content (55.5%) relative to the first and second codon positions (61.0 and 66.2%).

The amino acid sequences of the proteins of *G. thymalli* mtDNA were inferred using the genetic code for Platyhelminthes [17] that uses AAA to code for asparagine, AGA and AGG to code for serine, TGA to code for tryptophan, and TAA and TAG as stop codon [12]. The codons AUA (Ile, 6.5%), CUA (Leu, 5.2%), and UUC (Phe, 5.1%) are the most frequently used codons in the protein-coding sequences. Leucine and serine are the most frequently used amino acids with a proportion of 16.3 and 12.3%, respectively. The least used codons include CAG (Gln; 0.2%), CCG (Pro; 0.3%), and CGU, CGC, and CGG (0.3% each) and CGA (0.5%) that code for arginine. Not surprisingly, arginine (1.3%) and glutamine (0.8%) are the least frequently used amino acids.

The only codon used for initiation is ATG, and TAA and TAG are used for termination. The ND2 gene has an incomplete

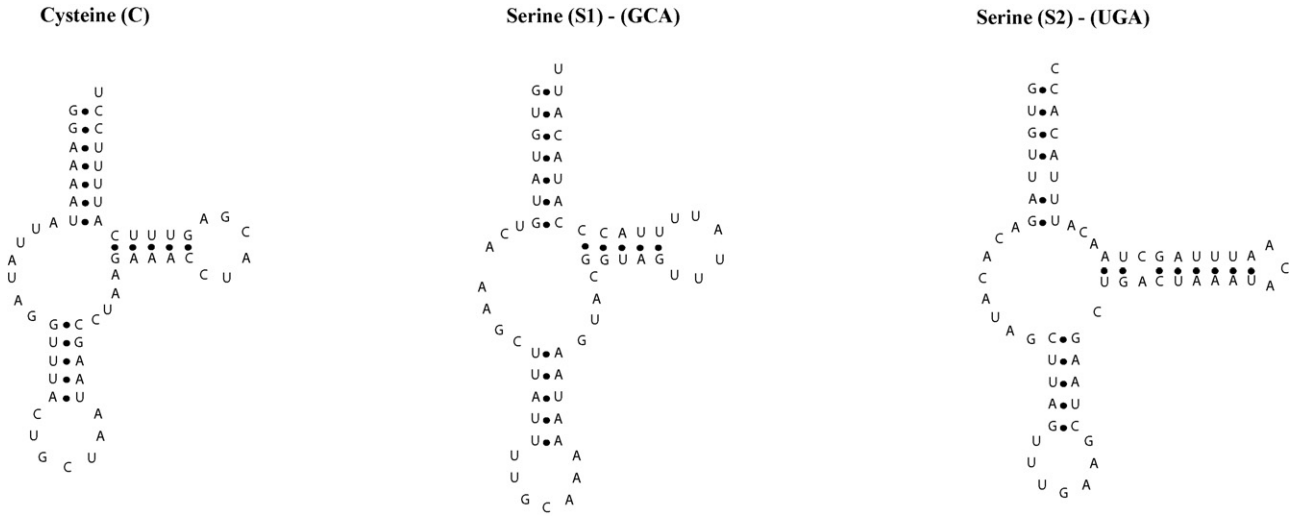

Fig. 2. Inferred secondary structure of $tRNA^{Cys}$, $tRNA^{Ser1}$ and $tRNA^{Ser2}$ of *G. thymalli* that have unpaired DHU-arms. The tRNAs are labeled with the abbreviations of their corresponding amino acids. Nucleotide sequences are presented from 5′ to 3′.

termination codon (TA) as found in many metazoan mitochondrial genomes and as has also been observed in cestodes [12]. A leucine zipper as has been described in the *nad4L* peptide of several flatworms [12,18] and other eukaryotes that consists of a leucine (L) residue repeated every seventh amino acid at least four times (leucine can be substituted by Met, Val or Ile), has not been detected in the *nad4L* peptide of *G. thymalli*. There are only three leucine residues whereas the fourth residue is substituted by a phenylalanine.

With the help of the program tRNAscan-SE 1.21 (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/) 17 tRNAs could easily be identified. The remaining 5 tRNAs that were not identified by tRNAscan-SE 1.21, $tRNA^{Ile}$, $tRNA^{Ser1}$, $tRNA^{Cys}$, $tRNA^{Ser2}$, and $tRNA^{Arg}$, were found by identifying the anticodon sequences, the conserved motif YUxxxR (xxx = the anti-codon), and the standard cloverleaf structure or cloverleaf structure lacking the DHU arm. A few mismatches occur in the structure of some tRNAs especially in the acceptor arm. Of a total of 426 potential bonds, 380 are canonical pairings, i.e. A–U (247) and G–C (133). Another 36 are non-Watson–Crick interactions, i.e. G–U, which are permitted in RNA secondary structures [19]. The remaining 10 bonds are considered mismatches. The mitochondrial genome of *G. thymalli* contains 22 tRNAs as is also common for other metazoans. Six tRNA genes are overlapping with other genes $tRNA^{His}$, $tRNA^{Val}$, $tRNA^{Ala}$, $tRNA^{Pro}$, $tRNA^{Ile}$, and $tRNA^{Thr}$. The lengths of these tRNA genes vary from 58 bp ($tRNA^{Gly}$) to 72 bp ($tRNA^{Ile}$ and $tRNA^{Glu}$) (Table 1). The inferred putative secondary structure of 19 tRNAs shows the typical cloverleaf shape. However, the secondary structures of $tRNA^{Cys}$, $tRNA^{Ser1}$ and $tRNA^{Ser2}$ have unpaired DHU-arms with a 7–8 bp long loop (Fig. 2). This feature has also been described for some other metazoan mtDNA [15].

The regions coding for the large and small ribosomal subunits genes are 958 and 710 bp long, respectively (Table 1). These values are small compared to most other metazoans but fall well into the length range in parasitic flatworms. The *rrnL* and *rrnS* are separated by the $tRNA^{Cys}$ which is usually the case in flatworms. The AT content of the rRNA encoding sequences is 64.7 and 70.7%, respectively.

In addition to 20 short non-coding regions that are all <112 bp, two long non-coding regions (NC) occur in the mitochondrial genome of *G. thymalli*. Both NCs are of similar length (799 and 767 bp, respectively) and are located between $tRNA^{Phe}$ and *atp6* and between $tRNA^{Met}$ and $tRNA^{Ser2}$. The AT content of the NCs (63.9 and 64.5%) is only slightly higher than the overall AT content of the mitochondrial genome (62.8%) and thus somewhat lower than what has been frequently observed in other metazoan (usually >70%). This low AT content may have facilitated to relatively easily amplify and sequence these non-coding regions. Furthermore, sequence analysis of the two NCs did not reveal clusters of repeated motifs in *G. thymalli* that are frequently found in other species. However, as has been reported for the closely related *G. salaris* [10] (GenBank accession number DQ988931) the sequences of the two NCs show a very high level of sequence similarity over 724 bp. In these regions both NCs differ only by 2.1% (10 transitions, 4 transversions and 1 indel). It is also noteworthy, that the two NCs have been amplified in two different amplicons and sequenced using primer walking; they can, therefore, not be explained as artifacts.

Compared to the recently described mtDNA of *G. salaris* [10] the mitochondrial genome of *G. thymalli* from Hnilec, Slovakia, differs on average by 2.2%. While some of the tRNA genes do not differ at all, the most differentiated sequence is the cytochrome oxidase subunit III gene (Table 1). Surprisingly, the long non-coding regions, in particular NC1, are highly conserved. Further studies need to be conducted in order to analyze why these regions that cannot be associated to any gene show such a low differentiation between the two species.

Acknowledgements

We are grateful to Vladka Hanzelova for providing *G. thymalli* samples from Slovakia. The project was supported by the

Norwegian Research Council Wild Salmon Programme (Project No. 145861/720) and the National Centre for Biosystematics (Project No. 146515/420), co-funded by the NRC and the NHM, University of Oslo, Norway. TH was funded by a Marie Curie Research Fellowship (Meif-ct-2004-501684).

References

[1] Hansen H, Bachmann L, Bakke TA. Mitochondrial DNA variation of *Gyrodactylus* spp. (Monogenea, Gyrodactylidae) populations infecting Atlantic salmon, grayling, and rainbow trout in Norway and Sweden. Int J Parasitol 2003;33:1471–8.

[2] Mo TA, Norheim K. The surveillance and control programme for *Gyrodactylus salaris* in Atlantic salmon and rainbow trout in Norway. In: Annual report 2004. National Veterinary Institute; 2005. p. 137–9.

[3] Mo TA, Norheim K, Hellesnes I. The surveillance and control programme for *Gyrodactylus salaris* in Atlantic salmon and rainbow trout in Norway. Norsk Veterinærtidsskrift 2004;3:157–63 (in Norwegian, English summary).

[4] Shinn AP, Hansen H, Olstad K, Bachmann L, Bakke TA. The use of morphometric characters to discriminate specimens of laboratory-reared and wild populations of *Gyrodactylus salaris* and *G. thymalli* (Monogenea). Folia Parasitol 2004;51:239–52.

[5] Ziętara MS, Huyse T, Lumme J, Volckaert FA. Deep divergence among subgenera of *Gyrodactylus* inferred from rDNA ITS region. Parasitology 2002;124:124:39–52.

[6] Sterud E, Mo TA, Collins CM, Cunningham CO. The use of host specificity, pathogenicity, and molecular markers to differentiate between *Gyrodactylus salaris* Malmberg, 1957 and *G. thymalli* Zitnan, 1960 (Monogenea: Gyrodactylidae). Parasitology 2002;124:203–13.

[7] Hansen H, Martinsen L, Bakke TA, Bachmann L. The incongruence of nuclear and mitochondrial DNA variation supports conspecificity of the monogenean parasites *Gyrodactylus salaris* and *G. thymalli*. Parasitology 2006;113:639–50.

[8] Meinila M, Kuusela J, Ziętara MS, Lumme J. Initial steps of speciation by geographic isolation and host switch in salmonid pathogen *Gyrodactylus salaris* (Monogenea: Gyrodactylidae). Int J Parasitol 2004;34: 515–26.

[9] Place AR, Feng XJ, Steven CR, Fourcade HM, Boore JL. Genetic markers in blue crabs (*Callinectes sapidus*). II: Complete mitochondrial genome sequence and characterization of genetic variation. J Exp Mar Biol Ecol 2005;319:15–27.

[10] Huyse T, Plaisance L, Webster BL, et al. The mitochondrial genome of Gyrodactylus salaris (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (Salmo salar). Parasitology 2007;134:739–47.

[11] Žitňan R. *Gyrodactylus thymalli* sp. nov. aus den Flossen der Äsche (*Thymallus thymallus*). Helminthologia 1960;2:266–9 (in German, Russian, English and French summary).

[12] Le TH, Blair D, McManus DP. Mitochondrial genomes of parasitic flatworms. Trends Parasitol 2002;19:206–13.

[13] von Nickisch-Rosenegk M, Brown WM, Boore JL. Sequence and structure of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: gene arrangement indicates that Platyhelminths are derived eutrochozoans. Mol Biol Evol 2001;18:721–30.

[14] Littlewood DTJ, Lockyer AE, Webster BL, Johnston DA, Le TH. The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. Mol Phylogenet Evol 2006;39:452–67.

[15] Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 1992;141:173–216.

[16] Jeon HK, Lee KH, Kim KH, Hwang UW, Eom KS. Complete sequence and structure of the mitochondrial genome of the human tapeworm, *Taenia asiatica* (Platyhelminthes; Cestoda). Parasitology 2005;130:717–26.

[17] Telford MJ, Herniou EA, Russell RB, Littlewood DTJ. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. Proc Natl Acad Sci USA 2000;97:11359–64.

[18] Le TH, Humair PF, Blair D, Agatsuma T, Littlewood DTJ, McManus DP. Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. Mol Biochem Parasitol 2001;117: 61–71.

[19] Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, Penny D. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol Biol Evol 1996;13:150–69.