ResearchOnline@JCU

JAMES COOK UNIVERSITY AUSTRALIA

This is the author-created version of the following work:

Le, Thanh Hoa, Nguyen, Khue Thi, Pham, Linh Thi Khanh, Doan, Huong Thi Thanh, Agatsuma, Takeshi, and Blair, David (2022) ***The complete mitogenome of the Asian lung fluke Paragonimus skrjabini miyazakii and its implications for the family Paragonimidae (Trematoda: Platyhelminthes)*****. Parasitology, . (In Press)**

Access to this file is available from:

https://researchonline.jcu.edu.au/76116/

© The Author(s), 2022. Published by Cambridge University Press

Please refer to the original source for the final version of this work:

https://doi.org/10.1017/S0031182022001184

The complete mitogenome of the Asian lung fluke ***Paragonimus skrjabini miyazakii*** and its implications for the family Paragonimidae (Trematoda: Platyhelminthes)

Thanh Hoa Le[1,2]*, Khue Thi Nguyen[1], Linh Thi Khanh Pham[1], Huong Thi Thanh Doan[1,2], Takeshi Agatsuma[3], David Blair[4*]

[1]Immunology Department, Institute of Biotechnology (IBT), Vietnam Academy of Science and Technology (VAST). 18. Hoang Quoc Viet Rd., Cau Giay, Hanoi, Vietnam.

[2]Graduate University of Science and Technology (GUST), Vietnam Academy of Science and Technology (VAST). 18. Hoang Quoc Viet Rd., Cau Giay, Hanoi, Vietnam.

[3]Department of Environmental Health Sciences, Kochi Medical School, Kohasu, Oko-cho 185-1, Nankoku, Kochi, 783-8505, Japan.

[4]College of Marine and Environmental Sciences, James Cook University, Townsville, Queensland 4811, Australia.

Corresponding authors: TH Le (email: imibtvn@gmail.com; Tel: 84-0912336855; Fax: 84-024-38363144); D. Blair (email: davidblair49@gmail.com)

Accepted Manuscript

This is an Accepted Manuscript for Parasitology. This version may be subject to change during the production process. DOI: 10.1017/S0031182022001184

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Abstract

The complete circular mitogenome of *Paragonimus skrjabini miyazakii* (Platyhelminthes: Paragonimidae) from Japan, obtained by PacBio long-read sequencing, was 17,591 bp and contained 12 protein-coding genes (PCGs), two mitoribosomal RNA and 22 transfer RNA genes. The *atp*8 gene was absent, and there was a 40 bp overlap between *nad*4L and *nad*4. The long non-coding region (NCR: 4.3 kb) included distinct types of long and short repeat units (L/SRUs). The pattern of base usage for PCGs and the mtDNA coding region overall in Asian and American *Paragonimus* species (*P. s. miyazakii*, *P. heterotremus*, *P. ohirai*, and *P. kellicotti*) and the Indian form of *P. westermani* was T > G > A > C. On the other hand, East-Asian *P. westermani* used T > G > C > A. Five Asian and American *Paragonimus* species and *P. westermani* had TTT/Phe, TTG/Leu, and GTT/Val as the most frequently used codons, whereas the least-used codons were different in each species and between regional forms of *P. westermani*. The phylogenetic tree reconstructed from a concatenated alignment of amino acids of 12 PCGs from 36 strains/26 species/5 families of trematodes confirmed that the Paragonimidae is monophyletic, with 100% nodal support. *Paragonimus skrjabini miyazakii* was resolved as a sister to *P. heterotremus.* The *P. westermani* clade was clearly separate from remaining congeners. The latter clade was comprised of two subclades, one of the East-Asian and the other of the Indian Type 1 samples. Additional mitogenomes in the Paragonimidae are needed for genomic characterization and are useful for diagnostics, identification, and genetic/ phylogenetic/ epidemiological/ evolutionary studies of the Paragonimidae.

Keywords: *Paragonimus skrjabini miyazakii*, Paragonimidae, Troglotremata, Mitogenome, Phylogenetic analysis, Repeat units, Skewness value.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Introduction

More than 50 species have been described in the genus *Paragonimus* (Trematoda: Paragonimidae) (Blair, 2008). As adults, *Paragonimus* species mature in the lungs of mammals. Freshwater snails and crustaceans act as first and second intermediate hosts, respectively. Eight species or subspecies of *Paragonimus* cause zoonotic diseases in humans, acquired by consumption of crustacean hosts containing metacercariae. These include *Paragonimus westermani*, *Paragonimus heterotremus*, *Paragonimus skrjabini miyazakii*, *Paragonimus skrjabini skrjabini*, *Paragonimus kellicotti*, *Paragonimus mexicanus*, *Paragonimus uterobilateralis* and *Paragonimus africanus* (see Blair, 2019; Chai and Jung, 2018).

Some of these names (in particular *P. westermani*, *P. heterotremus* and *P. skrjabini*) refer to species complexes with multiple genotypes and strains, often with a geographical pattern (Blair, 2008, 2019; Blair *et al*., 2016). The *P. westermani* complex has received the most attention. Its range extends from India and Sri Lanka eastwards to China, Japan and Korean Peninsula. Its northern limits are in the Russian Far East and its southern limits are in the Philippines, with a possible case in Papua New Guinea (Blair, 2019, Voronova *et al*., 2022). Human cases, which are mostly confined to parts of eastern Asia and the Philippines, are typically of pulmonary paragonimiasis: pairs of worms form capsules in the lungs from which their eggs pass into bronchioles and are voided from the body via sputum or feces (Blair, 2019). Molecular studies have revealed great diversity across this geographic range. There is also diversity within single regions: three distinct genotypes (Types 1–3) occur in northeastern India alone, two in Sri Lanka and two in Thailand (Blair, 2019). One case of pulmonary paragonimiasis has been reported from northeast India, likely caused by the Indian Type 2 form (Singh *et al*., 2013).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

The *P. skrjabini* complex includes two nominal subspecies, *P. skrjabini skrjabini* Chen, 1959 in China, and *P. skrjabini miyazakii* Kamo, Nishida, Hatsushika & Tomimura, 1961, in Japan and parts of eastern China (Blair *et al*., 2005). Additional named taxa may also belong within this complex (Blair *et al*., 2005). The complex extends from northeast India into eastern and southern China, and Japan. Its members rarely mature in human lungs. Instead, worms wander into extra-pulmonary sites causing ectopic paragonimiasis with a wide range of signs and symptoms (Blair, 2019).

For accuracy in species identification, taxonomic classification and inferring phylogenetic relationships of Paragonimidae, markers from the mitogenomes are among the best tools (Boore, 1999; Hardman and Hardman, 2006; Bernt *et al*., 2013*a*; Wey-Fabrizius *et al*., 2013; Solà *et al*., 2015). The mitogenome is a circular molecule with a length ranging from 13.5 kb to over 20 kb in trematodes (Le *et al*., 2002). As is typical for trematodes, the *Paragonimus* mitogenome consists of 12 protein-coding genes (PCGs) named *cox*1–3, *nad*1–6 and *nad4*L. *atp*6 and cytochrome b or *cob* or *cyt*B, 2 mitoribosomal RNA genes (MRGs) as *rrn*L (or 16S) and *rrn*S (or 12S), 22 transfer RNA genes (tRNAs), and a variable non-coding region that possesses multiple repeats (Biswal *et al*., 2014; Le *et al*., 2019, 2020; Oey *et al*., 2019*b*; Rosa *et al*., 2020).

Although many complete mtDNA genome sequences have been reported from trematodes (**Table S1** lists some examples, and see https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/Trematoda for a current list), relatively few are available for *Paragonimus* species (all are listed in **Table S1**). No mitogenome has been published for any member of the *P. skrjabini* complex. Next-generation sequencing (NGS) has become increasingly applied to generate mitogenomes for a range of species (Biswal *et al*., 2014; Oey *et al*., 2019*b*). Single-molecule real-time sequencing (SMRT) from Pacific

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Biosystems ('PacBio') offers very long read lengths with high accuracy, overcoming errors caused by problematic genomic regions (Rhoads and Au, 2015).

The aim of this paper is to present the complete mtDNA sequence, which was obtained using NGS, from the Asian lung fluke *P. s. miyazakii*, representing the first for the *P. skrjabini* complex, and to use the mitogenomic markers for taxonomic and phylogenetic studies of the Paragonimidae. We also provide a detailed comparative analysis of the mitogenomic sequences and a phylogenetic analysis, emphasizing their utility as molecular markers for estimation of pairwise genetic distance and molecular evolutionary studies. The data are useful for further studies of the family Paragonimidae.

Materials and methods

Parasite samples and DNA extraction

The adult *P. s. miyazakii* samples used in this study were collected from experimentally infected albino rats (*Rattus norvegicus*) fed with metacercariae collected from the crab *Geothelphusa dehaani* hosts near Okuyanai, Kochi Prefecture, Japan (34° 39' 42.01" N, 133°56' 6" E). The metacercariae from the crabs and the adults obtained from the infected rats were identified by morphology and molecular analysis. The *cox*1 and ITS-2 sequences from samples in this study were 99%–100% identical with these markers from all *P. s. miyazakii* samples collected previously in different geographical localities in Japan, including Kochi, Miyazaki, Tokushima, Shizuoka, and Yamaguchi (Kamo *et al*., 1961; Agatsuma *et al*., 1992; Blair *et al*., 1997, 2005; Ohari *et al*., 2019; Urabe and Marcaida, 2022).

Total genomic DNA was extracted from individual adult worms using the GeneJET™ Genomic DNA Purification Kit (Thermo Fisher Scientific Inc., MA, USA) according to the

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

manufacturer's instructions. The genomic DNA was eluted in 50 µL of the elution buffer and kept at –20 °C until use. The DNA content was quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer. A working concentration (50 ng/µL) was prepared and an amount of 2 µL was used in each LPCR in a 50 µL reaction volume.

Targeted enrichment of the mitogenome by long polymerase chain reaction

Trematode-PCR primers based on alignment of all mtDNA of trematodes available and specific primers binding to *nad*5, *cox*3, and cytochrome b (*cob*) were designed for amplifying the whole mtDNA of *P. s. miyazakii* in five overlapping fragments (**Tables S2** and **S3**). Among these, two amplicons (obtained by primer pairs PMI12F/PMI15R and GLYF/PMI20R, respectively) were used for comparative validation of the length of the NCR as well as these amplicons were used for the NGS (**Table S3**). Long-range PCRs (LPCR) were performed in a 50 µL volume reaction in a MJ PTC-100 Thermal Cycler. Each reaction contained 25 µL 2X LongAmp Master Mix (New England Biolabs, Ipswich, MA, USA), 2 µL each primer (10 pmol/µL), 2 µL of template DNA, and 19 µL DEPC-water). Long-PCRs were conducted with an initial denaturation at 94 °C for 1 min, followed by 30 cycles, each consisting of a denaturation step for 30 s at 94 °C, an annealing/extension step at 50 °C for 30 s, extension at 65 °C for 8 min, and a final extension at 65 °C for 10 min. The PCR products (10 µL of each) were examined on a 1% agarose gel, stained with ethidium bromide and visualized under UV light (Wealtec, Sparks, NV, USA)

All dsDNA products were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA) and the amplicon length was verified via 1.5% agarose gel electrophoresis. Three amplicons from the coding mtDNA and two from the NCR (**Table S3**) were pooled for NGS. The complete mitogenome of *P. s. miyazakii* was sequenced using the

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

PacBio SEQUEL system (https://www.pacb.com/) with a targeted long-read sequencing approach at the PacBio facility at the Institute of Biotechnology (Hanoi, Vietnam).

Library preparation

The dsDNA products of each sample from five overlapping amplicons were pooled into one tube and purified with AMPure XP beads (Pacific Biosciences, Menlo Park, CA, USA). Input dsDNA was quantified using the Qubit fluorometer 3.0 and Qubit dsDNA HS Assay reagents (Thermo Fisher Scientific, Waltham, MA, USA). SMRTbell Libraries were prepared using the Express Template Prep Kit 2.0 with multiplexing amplicons protocol with low DNA input (100 ng) (Pacific Biosciences, Menlo Park, CA, USA) for sequencing on the PacBio SEQUEL system according to the manufacturer's instructions. The SMRTbell templates were purified once with 1.2 volumes of AMPure PB beads and the size and amount of the library were checked again using the Bioanalyzer 2100 system (Agilent, CA, USA) and the Qubit fluorometer 3.0 with Qubit™ dsDNA HS Assay reagents, respectively. The libraries of all amplicons were then pooled before long-read sequencing.

Sequencing and de novo assembly

The pooled library was bound to polymerase using the Sequel Binding and the Internal Control Kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) and purified using Ampure PB beads. The DNA Control Complex 3.0 and the Internal Control Kit 3.0 from Sequel Binding and Internal Ctrl Kit 3.0 were used to control the sequencing procedure. The final library was loaded onto Sample Plate (Pacific Biosciences, Menlo Park, CA, USA). The run design was created by the Sample Setup software included in the SMRTLink portal v5.1 version 9.0 with an insert size of

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

1200 basepairs (bp). The sequencing signals were processed, evaluated, and converted into raw data by the Primary Analysis Computer server. All data was automatically transferred to the Secondary Analysis Server system via the intranet. High quality sequence data was proofread and generated by PacBio's circular consensus sequencing (CCS), then *de novo* assembled using Canu software v2.0 (Koren *et al.*, 2017), and the quality of the assembly was checked by using Quast software v5.0.2 (Gurevich *et al.*, 2013).

Mitogenome annotation and characterization

Twelve protein-coding genes were identified by comparative alignment with the available mitogenomes of the *Paragonimus* strains and species in GenBank or recently reported (Biswal *et al*., 2014; Qian *et al*., 2018; Wang *et al.*, 2018; Oey *et al*., 2019*b*; Le *et al*., 2019) (**Table S1**). ATG/GTG as start and TAA/TAG as stop codons were used to define individual PCG gene boundaries. The *echinoderm and flatworm mitochondrial genetic code* (Translation Table 9 in GenBank) was used for the translation of the PCGs. The nucleotide composition for PCGs, MRGs, and the mtDNA coding region (5'-*cox*3-*nad*5-3', designated as mtDNA*) and codon usage of the PCGs were analyzed using MEGA X (https://www.megasoftware.net/) (Kumar *et al*., 2018). Codon usage for all concatenated 12 PCGs from each strain/species was determined with the online program GENE INFINITY (Codon Usage: http://www.geneinfinity.org/sms/sms_codonusage.html). Genetic distance was estimated based on the pairwise amino-acid comparisons of concatenated 12 PCGs among 11 strains of five species (or species complexes) of the family Paragonimidae and was determined using GENEDOC 2.7 (https://softdeluxe.com/GeneDoc-180568/download/) for alignment and MEGA X for percentage estimation.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Transfer RNA genes were identified using the tRNAscan-SE 1.2.1 program (www.genetics.wustl.edu/eddy/tRNAscan-SE/) (Lowe and Chan, 2016), ARWEN (http://mbio-serv2.mbioekol.lu.se/ARWEN/) (Laslett and Canback, 2008), and the MITOS Alpha version (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt *et al*., 2013*b*). Final sequences and secondary structures were based on comparisons using all these programs. All transfer RNAs proposed by these programs were checked to confirm their typical "clover-leaf" structure, and any tRNAs not detected by these programs were found by inspection of the sequences based on the alignment with other trematode tRNA sequences known to form tRNA structures. Repeat sequences were detected in the NCR using the Tandem Repeats Finder v3.01 (Benson, 1999). The circular map and gene abbreviations on the map were generated using GenomeVx (Conant and Wolfe, 2008) with manual drawing of the tandem repeat units.

Skew values (ranging from −1 to +1) were determined by calculating the percentage of AT and GC nucleotide usage using Perna and Kocher's (1995) formula: AT skew = (A − T)/(A + T), and GC skew = (G − C)/(G + C), where the letters represent the absolute usage of the corresponding nucleotides in the sequences (see also Solà *et al*. (2015)).

Accepted Manuscript

Phylogenetic reconstruction

Concatenated aligned sequences of 12 PCGs from 36 strains/species of five families of trematodes listed in **Table S1** were used in a phylogenetic analysis. *Schistosoma haematobium* (Schistosomatidae) was used as an outgroup. In addition to *P. s. miyazakii* (OkuST1 strain, Japan), the ten available *Paragonimus* mtDNA sequences were included. Among these were two strains of *P. heterotremus* (LC, Vietnam and GX, China), one strain of *P. ohirai* (Kino strain, Japan), and six representatives of the *P. westermani* complex (two from South Korea, one from

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

China, and three from India. All of the Indian sequences appear to belong to Type 1 of Devi *et al*. (2013), based on comparisons of available *cox*1 sequences. One strain of *P. kellicotti* (Ozark, Missouri, the United States) was also included. The paper reporting this sequence has been retracted by the journal (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7748660/) but the sequence data appear to match expectations: the reader can chose whether to interpret these data with caution. The concatenated protein-coding nucleotide sequences were imported into GENEDOC 2.7 (available at: https://softdeluxe.com/GeneDoc-180568/download/) and translated using Translation Table 9 in GenBank. The inferred amino-acid sequences ranged in length from 3,354 residues (*Paragonimus* kellicotti, Pkel-Ozark-US; GenBank: MH322000; and *P. s. miyazakii*; GenBank: ON782295; Paragonimidae) to 3,403 residues (*Metagonimus yokogawai*, GenBank: KC330755; Heterophyidae). The alignment was performed by GENEDOC 2.7, confirmed by the MAFFT 7.122 program (available at https://www.ebi.ac.uk/Tools/msa/mafft/) (Katoh and Standley, 2013), and lightly trimmed by Gblocks 0.91b for the best-quality final sequence block to avoid excessive ambiguity (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) (Talavera and Castresana, 2007). The phylogenetic tree was reconstructed using the maximum-likelihood method in MEGA X with 1000 bootstrap resamplings with a final alignment of 3,146 amino-acid residues after removing gap sites (Kumar *et al*., 2018). The substitution model with the best score, according to the Bayesian information criterion, was the Jones, Taylor, and Thornton + F + G + I model, with residue frequencies estimated from the data (+F), rate variation along the length of the alignment (+G) and allowing for a proportion of invariant sites (+ I).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Results

Paragonimus skrjabini miyazakii *mitogenomic organization and features*

Paragonimus skrjabini miyazakii, strain OkuST1 from Kochi, Japan, possessed the complete mitogenome of 17,591 bp in length (GenBank accession no. ON782295) (**Table 1**; **Fig. 1**). The NCR in the mitogenome of *P. s. miyazakii*, which is about 4.3 kb, was successfully obtained using the long-reading PacBio system. The circular mtDNA molecule is comprised of 12 PCGs (*cox*1−3, *cob*, *nad*1−6, *nad*4L, *atp*6), two MRGs (16S or *rrn*L and 12S or *rrn*S), and 22 tRNAs or *trn*, and an NCR, rich in long and short tandem repeats (LRUs and SRUs). The length of the mitogenome was similar to that of members in the Echinostomatidae, e.g., *Artyfechinostomum malayanum* (17,175 bp, strain EMI3, Thailand) (Pham *et al*., 2022), *Echinostoma revolutum* (17,030 bp, strain MSD15, Thailand) (Le *et al*., 2020), shorter than *Paragonimus westermani* (20,273 bp, strain IND2009, India, GenBank: CM017921) (Oey *et al*., 2019*b*) and *Echinostoma paraensei* (20,298 bp, GenBank: KT008005) but longer than some other *Paragonimus* species (**Table S1**). The linearized map of the *P. s. miyazakii* mtDNA is 5'-*cox*3-H-*cob*-*nad*4L-*nad*4-QFM-*atp*6-*nad*2-VAD-*nad*1-NPIK-*nad*3-S_1W-*cox*1-T-*rrn*L-C-*rrn*S-*cox*2-*nad*6-$YL_1S_2L_2$R-*nad*5-G-NCR[LRU1(E)−8.6#]-[SRU1−6.8#]-3'. The NCR was located between $tRNA^{Glu}$ and *cox*3. The NCR of *P. s. miyazakii* mtDNA was bounded by $tRNA^{Glu}$ (*trn*E) and the *cox*3 gene, was relatively long (4,351 bp) and was divided into two subregions. The first subregion contained 8.6 LRUs (8 perfect LRUs of 384 bp/each and a partial one of 241 bp) and the second contained 6.8 SRUs (6 perfect SRUs of 118 bp/each and a partial one of 93 bp) (**Table 1; Figs. 1** and **S1**). The 8 LRUs were identical and each LRU (384 bp) contained a 66 bp complete sequence of $tRNA^{Glu}$ (*trn*E1–E8) in between each pair of LRUs (**Fig. S1; Table 1**).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Ribosomal and transfer RNA genes

The *16S* or *rrn*L gene is 976 bp and the *12S* or *rrn*S gene is 738 bp. These were located in the region between $tRNA^{Thr}$ (*trn*T) and *cox*2 and separated from one another by $tRNA^{Cys}$ (*trn*C) (**Table 1**). Twenty-two tRNAs ranged from 61 (*trn*S_1) to 71 nucleotides (*trn*N) in length. Except for two tRNAs that specify serine, $tRNA^{Ser1(AGN)}$ (*trn*S_1) and $tRNA^{Ser2(UCN)}$ (*trn*S_2), which possessed a special form, missing the DHU-arm (dihydrouridine-arm), the remaining twenty tRNAs had the common 'cloverleaf' secondary structures with the complete four arms (**Table 1**; **Fig. S2**).

Protein-coding genes and codon usage

ATG was used as the start codon in nine PCGs, and the other three genes used GTG. Three PCGs (*cox*1, *cox*2, and *nad*6) used TAA as the stop codon; the other nine genes used TAG. An overlap of 40 bp exists between *nad*4L and *nad*4. There is no *atp*8 gene (**Table 2**).

The codon usage values for 12 PCGs in the mitogenomes of *P. s. miyazakii* and the other 10 strains/species of the Paragonimidae are summarized in **Table S4**. The most frequently used codons are TTT (for phenylalanine), followed by TTG (for leucine) and GTT (for valine) in all paragonimids. In *P. s. miyazakii*, TTT/Phe (10.13%, n=341), TTG/Leu (8.02%, n=270), and GTT/Val (5.67%, n=191) were the most commonly used. The least frequently used codons were CGC/Arg (n=5) in *P. s. miyazakii* and *P. kellicotti*, and CCA/Pro (n=5) in *P. westermani* (two strains from Korea and one from China). Only one CGC/Arg codon was used in *P. ohirai*. The least-used codons are CGA/Arg in *P. heterotremus* (one strain from Vietnam and one from China) and CAA/Gln in *P. westermani* (2 strains from Korea and one from China) (**Table S4**).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Base composition and skew values

Table 2 presents the base composition of A, T, G, and C and skewness values of AT and GC content for PCGs, MRGs, and mtDNA* (the coding region from 5' *cox*3 to 3' *nad*5) of 11 strains/species of Paragonimidae. *Paragonimus skrjabini miyazakii* used 15.83% A and 44.76% T (A+T = 60.59%), 27.70% G and 11.71% C (G+C = 39.41%) for 12 PCGs; and 17.33% A and 42.90% T (A+T = 60.23%), 27.72% G and 12.05% C (G+C = 39.77%) for its coding mtDNA* region (13,240 bp). The AT-skew was highly negative (–0.477/PCGs and –0.425/mtDNA*), indicating that T was more frequently used than A. The GC-skew was highly positive (0.406/PCGs and 0.394/mtDNA*), indicating that greater numbers of C were used than G. The base composition and skewness values for PCGs and mtDNA* seemed to be different between *P. s. miyazakii* and other paragonimids, for example, *P. ohirai* (Le *et al*., 2019), and also slightly different between Korean/Chinese and Indian *P. westermani* strains. The AT skew values were the most negative (–0.491 to –0.492/PCGs and –0.441/mtDNA*) in *P. heterotremus* (Vietnamese and Chinese strains) and (–0.492/PCGs and –0.442/mtDNA*) in *P. kellicotti* (**Table 2**). The pattern of the base usage for PCGs and mtDNAs* is T > G > A > C in *P. s. miyazakii*, and is similar in *P. heterotremus*, *P. ohirai*, and *P. kellicotti*, whereas it is T > G > C > A in *P. westermani* (Korean and Chinese strains, the East-Asian *P. westermani* subgroup). However, in the Indian *P. westermani*, the T>G>A>C is for the mtDNAs*, but it is T > G > C > A for the PCGs.

For MRGs, *P. s. miyazakii* used 22.11% A and 37.11% T (A+T = 59.22%), and 27.77% G and 13.01% C (G+C = 40.78%) with an AT skew of –0.253 and a GC skew of 0.362. This pattern was similar to that of *P. heterotremus* and *P. kellicotti* but different from that of *P. westermani*. The East-Asian (Korean and Chinese) strains used A+T less frequently than the

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Indian *P. westermani* (51.58%–51.73% versus 54.85%–55.42%, respectively) (**Table 2**). For MRGs, the T > G > C > A pattern of nucleotide usage is recorded in all strains/species of Paragonimidae.

Taxonomic and phylogenetic relationships

Genetic distances between species of Paragonimidae

Pairwise genetic distances (p-distances) were estimated based on the analysis of concatenated amino acid sequences inferred from 12 PCGs each among 11 strains/five species in the family Paragonimidae (**Table 3**). *Paragonimus skrjabini miyazakii* had the lowest genetic distance (14.71%–13.99%) to *P. heterotremus* and, *P. kellicotti* (16.21%), and the highest (24.34%–24.72%) to Indian *P. westermani* (strains Pwes-AP-IN (KX943544) and Pwes-(typeI)-IN (KM280646)). There was a low intra-specific genetic distance within strains in the *P. heterotremus* complex (1.25%) and between the Korean and Chinese *P. westermani* (0.23%–1.21%), but interestingly, a relatively high distance between members of the *P. westermani* complex within India (3.71%–7.35%). The *P. westermani* complex seemed to have two evolutionary sublineages with different genetic distances among the geographical strains (**Table 3**).

Phylogenetic relationships

To resolve the phylogenetic relationships between *P. s. miyazakii* and the other congeners within Paragonimidae and between this family and other families in suborders Opisthorchiata, Troglotremata, and Echinostomata, a maximum likelihood phylogenetic analysis was performed on the alignment of the concatenated amino acid sequences of 12 PCGs for 36 strains/26

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

species, including *P. s. miyazakii* (listed in **Table S1**). *Schistosoma haematobium* (family Schistosomatidae) was used as an outgroup species. In the phylogenetic tree (**Fig. 2**), the family Paragonimidae representing the suborder Troglotremata was rendered as a monophyletic clade and was placed as a sister clade to Opisthorchiata, comprising two families, Heterophyidae and Opisthorchiidae. The topology clearly showed two subclades with absolute nodal support (100%). One contained all the sequences of the *P. westermani* complex (Korean, Chinese, and Indian strains), and the other was formed by sequences from the remaining *Paragonimus* species (*P. s. miyazakii*, *P. heterotremus*, *P. ohirai*, and *P. kellicotti*) (**Fig. 2**). The *P. westermani* clade itself is comprised of two subclades, one of the East-Asian and the other of the Indian Type 1 strains, indicating the complicated phylogenetic relationships within this complex.

Discussion

The complete mitogenome of *Paragonimus skrjabini miyazakii* was obtained by long-read NGS using the PacBio system and showed gene order and content similar to those of other trematodes (Biswal *et al*., 2014; Ma *et al*., 2016; Oey *et al*., 2019*b*; Li *et al*., 2019; Le *et al*., 2019; 2020; Rajapakse *et al*., 2020; Rosa *et al*., 2020; Pham *et al*., 2022). The NCRs of trematode mtDNAs typically contain long and short repeat units but obtaining complete sequences of these is difficult using conventional Sanger and/or second-generation sequencing methods (Kinkar *et al*., 2019, 2020; Rosa *et al*., 2020). However, the PacBio single-molecule real-time sequencing can yield the long, complicated, and repetitive region of the NCR, which was 4.3 kb for *P. s. miyazakii* in this study; 6.9 kb for *P. westermani* (Oey *et al*., 2019*b*); 4.4 kb for *Echinococcus granulosus* (Kinkar *et al*., 2019); and 4 kb for *Schistosoma bovis* (Oey *et al*., 2019*a*). The *P. s. miyazakii* mitogenome, at 17,591 bp in length, is one of the longest *Paragonimus* mtDNAs

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

obtained to-date (Oey *et al*., 2019*b*; Rosa *et al*., 2020).

A notable finding was the presence of multiple LRUs and SRUs in the NCR and the repeated 66 bp sequence of $tRNA^{Glu}$ (*trn*E1−E8) included between each pair of LRUs. Multiple long-repetitive sequences tandemly arranged in the NCRs are also found in other trematodes, as summarized in Le *et al*. (2020). The presence of long repeats in the mitogenome has been reported not only in *P. s. miyazakii* but also in the mtDNAs of *Paragonimus westermani*, *Clonorchis sinensis*, *Schistosoma bovis,* and the cestode *Echinococcus granulosus* G1 (Oey *et al*., 2019*a*, *b*; Kinkar *et al*., 2019, 2020). The multiple copies of *trn*E in the LRUs raise the question as to which copy might be functional, or whether all the tRNAs are active for transferring Glu. In the cestode *Echinococcus granulosus* G1, $tRNA^{Gly}$ was found in each of the ten repeat units of the NCR (Kinkar *et al*., 2019).

The mito-ribosomal gene order and location were conserved in paragonimids, as seen in *P. heterotremus, P. ohirai, P. westermani,* and *P. kellicotti* and many other trematodes reported to date (Biswal *et al*., 2014; Quan *et al*., 2018; Wang *et al*., 2018; Oey *et al*., 2019*b*; Le *et al*., 2019, 2020; Rosa *et al*., 2020). The absence of a DHU-arm, which was replaced by a DHU-replacement loop in both tRNAs specifying serine (*trn*S1 and *trn*S2) in the mitogenome of *P. s. miyazakii*, is commonly found in other *Paragonimus* species, i.e., *P. ohirai* (Le *et al*., 2019) and in many other trematodes in the Echinostomatidae, Fasciolidae, and Opisthorchiidae, i.e., *Clonorchis sinensis*, *Opisthorchis viverrini*, and *O. felineus* (Shekhovtsov *et al*., 2010; Cai *et al*., 2012), *Fasciola hepatica*, *F. gigantica*, *Fasciola jacksoni* and *Fascioloides magna* (Le *et al*., 2001, 2002; Liu *et al*., 2014; Ma *et al*., 2016; Rajapakse *et al*., 2020), *Echinostoma miyagawai*, *E. revolutum,* and *Echinostoma*/*Artyfechinostomum malayanum* (Li *et al*., 2019; Le *et al*., 2020; Pham *et al*., 2022).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

The base composition of A, T, G, and C, as well as skewness values of AT and GC content for PCGs, MRGs and mtDNA*, indicated the use of T was more frequent than A, and G more than C. These yield the highly negative values for AT-skew and highly positive GC-skew in the mtDNA of all *Paragonimus* species. However, the nucleotide usage was not similar among the Asian/American and the Indian paragonimids and was even different between the East-Asian and the Indian *P. westermani* mitogenomes. Whereas the T > G > C > A pattern for nucleotide use was found in the mtDNA* of the Asian/American *Paragonimus* spp. (*P. s. miyazakii*, *P. heterotremus*, *P. ohirai*, and *P. kellicotti*) and the East Asian *P. westermani* (Korean and Chinese strains), the Indian Type 1 *P. westermani* exhibited a different usage pattern (T > G > A > C).

The most frequently used codons (TTT for Phe, TTG for Leu, and GTT for Val) appeared similar in all the eleven mitogenomes of the five *Paragonimus* species, while the least frequently used codons were different in each species and between the East-Asian and Indian *P. westermani* subgroups. The least frequently used codons in *P. heterotremus* were (CGC/Arg, n=3), in East-Asian *P. westermani* (CAA/Gln, n=3). In *P. s. miyazakii* and *P. kellicotti*, five codons were equally least common and in *P. ohirai*, CGC/Arg was only used once.

We inferred a phylogenetic tree from an alignment of concatenated, translated PCGs. This revealed the Paragonimidae to be monophyletic, given the limited mitogenome sampling possible. The separation of *Paragonimus* species into two main groups (the *P. westermani* complex in one and all remaining species in the other) reflects the pattern seen using other sequence data (e.g. Blair, 2022). Similarly, the *P. westermani* group itself is subdivided, with the three Indian Type 1 mitogenomes being distinct from those from eastern Asia (and rather different from each other). There are too few complete mitogenomes from the complex for us to

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

be able to add to previous findings based on shorter sequence markers. But it is clear that relationships within the *P. westermani* complex defy easy explanation (Blair, 2019; Doanh *et al*., 2016). For example, within northeastern India, three distinct clusters have been recognized: Type 1, which the mitogenomes used here represent, Type 2, which is phylogenetically close to eastern Asian strains infecting humans and Type 3 which might represent *Paragonimus siamensis* (Devi *et al*., 2013). No mitogenomes are available as yet for either Type 2 or Type 3.

In our phylogenetic tree, *P. s miyazakii* appeared closest to *P. heterotremus*. This reflects previous findings based on molecular data, which place these two species in a clade with *P. macrorchis* and *P. vietnamensis* (e.g. see Blair, 2022). Unfortunately, there is no complete mtDNA sequence for *P. s. skrjabini* and for other related species, such as *P. hokuoensis*, *P. macrorchis* or *P. vietnamensis*, which could assist in resolution of the phylogenetic position of *P. s. miyazakii* and *P. s. skrjabini* in the Paragonimidae.

Conclusion

The present study provides the complete mitogenome of *Paragonimus skrjabini miyazakii* (from one individual worm, from Kochi, Japan), obtained by PacBio long-read sequencing of long-PCR amplicons. Annotations included a description of its genomic features, including base composition, genetic distance, and codon usage in comparison with those of other *Paragonimus* congeners. Mitogenomic organization and features were the same as in other *Paragonimus* species. The long non-coding region was completely sequenced for the *P. s. miyazakii* mtDNA and featured long repeat units. Two major groups were apparent in a phylogenetic tree: members of the *P. westermani* complex and the remaining members of the genus for which mitogenomes were available (*P. s. miyazakii*, *P. heterotremus*, *P. ohirai*, and *P. kellicotti*). Despite the

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

monophyly of family Paragonimidae being revealed by the phylogenetic analysis, the taxonomic and phylogenetic relationships of *P. s. miyazakii* and *P. s. skrjabini* need to be confirmed. Fully characterized mitogenomes of additional paragonimid species, particularly those in the major *Paragonimus* complexes, are necessary and will be useful for diagnostic, taxonomic, epidemiological, systematic, phylogenetic, and population genetic studies.

Supplementary material

The supplementary materials for this article can be found at https://www.cambridge.org/core/journals/parasitology

Author ORCIDs: https://orcid.org/0000-0003-3841-368X

Acknowledgements

We express our thanks to colleagues and technicians for their contribution to our laboratory work. We are grateful to Mr. Le Tung Lam and Ms. Le Thi Dzung for their technical help with NGS by PacBio at the Institute of Biotechnology, Hanoi, Vietnam.

Author contributions

Thanh Hoa Le conceived the study, analyzed the final data, prepared figures and tables and wrote the manuscript; Khue Thi Nguyen, Linh Thi Khanh Pham, Huong Thi Thanh Doan conducted laboratory works and sequence analyses. Takeshi Agatsuma collected, experimentally infected, molecularly identified and provided specimens. David Blair reviewed the drafts. Takeshi Agatsuma, David Blair and Thanh Hoa Le completed and approved the manuscript. All

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

authors read and approved the final manuscript.

Financial support

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 108.02–2020.07) (PI: Thanh Hoa Le).

Conflict of interest

None.

Ethical standards

None.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

References

Agatsuma, T, Ketudat, P, Thaithong, S, Shibahara, T, Sugiyama, H, Habe, S, Terasaki, K and Kawashima, K (1992) Electrophoretic analysis of a natural population of the Thai *Paragonimus heterotremus* and its genetic relationship to the three Japanese species *P. miyazakii, P. ohirai* and *P. westermani*. *Parasitology Research* **78(6)**, 463–468 https://doi.org/10.1007/BF00931564.

Benson, G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research **27**, 573–580. https://doi.org/10.1093/nar/27.2.573.

Bernt, M, Braband, A, Schierwater, B and Stadler, PF (2013*a*) Genetic aspects of mitochondrial genome evolution. *Molecular Phylogenetics and Evolution* **69(2)**, 328–338. https://10.1016/j.ympev.2012.10.020.

Bernt, M, Donath, A, Jühling, F, Externbrink, F, Florentz, C, Fritzsch, G, Pütz, J, Middendorf, M and Stadler, PF (2013*b*) MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Biswal, DK, Chatterjee, A, Bhattacharya, A and Tandon, V (2014) The mitochondrial genome of *Paragonimus westermani* (Kerbert, 1878), the Indian isolate of the lung fluke representative of the family Paragonimidae (Trematoda). *Peer J* **2**, e484. https://doi.org/10.7717/peerj.484

Blair, D (2008) Family Paragonimidae Dollfus, 1939. In Bray RA, Gibson DI, Jones A (editors). *Keys to the Trematoda*, *vol 3*. CAB International and Natural History Museum, London; 2008, pp. 271–273.

Blair, D (2019) Paragonimiasis. In Toledo R, Fried B (editors). *Digenetic Trematodes. Springer*,

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Switzerland: Springer Nature Switzerland AG; pp. 105–138.

Blair, D (2022) Lung flukes of the genus *Paragonimus*: ancient and re-emerging pathogens. *Parasitology* 1–10. https://doi.org/10.1017/S0031182022000300.

Blair, D, Agatsuma, T, Watanobe, T, Okamoto, M and Ito, A (1997) Geographical genetic structure within the human lung fluke, *Paragonimus westermani*, detected from DNA sequences. *Parasitology* **115(4)**, 411–417. https://doi.org/10.1017/s0031182097001534.

Blair, D, Chang, Z, Chen, M, Cui, A, Wu, B, Agatsuma, T, Iwagami, M, Corlis, D, Fu, C and Zhan, X (2005) *Paragonimus skrjabini* Chen, 1959 (Digenea: Paragonimidae) and related species in eastern Asia: a combined molecular and morphological approach to identification and taxonomy. *Systematic Parasitology* **60(1)**, 1–21. https://doi.org/10.1007/s11230-004-1378-5.

Blair, D, Nawa, Y, Mitreva, M and Doanh, PN (2016) Gene diversity and genetic variation in lung flukes (genus *Paragonimus*). *Transactions of The Royal Society of Tropical Medicine and Hygiene* **110(1)**, 6–12. https://doi.org/10.1093/trstmh/trv101.

Boore, JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research* **27**, 1767–1780. https://doi.org/10.1093/nar/27.8.1767.

Cai, XQ, Liu, GH, Song, HQ, Wu, CY, Zou, FC, Yan, H, Yuan, ZG, Lin, RQ and Zhu, XQ (2012) Sequences and gene organization of the mitochondrial genomes of the liver flukes Opisthorchis viverrini and Clonorchis sinensis (Trematoda). Parasitology Research **110**, 235–243. https://doi.org/10.1007/s00436-011-2477-2.

Chai, JY and Jung, BK (2018) Paragonimus spp. In Rose JB and Jiménez-Cisneros B (editors) *Global Water Pathogen Project*. http://www.waterpathogens.org (Robertson L (eds) Part 4 Helminths) http://www.waterpathogens.org/book/paragonimus Michigan State University,

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

E. Lansing, MI, UNESCO. https://doi.org/10.14321/waterpathogens.46

Conant, GC and Wolfe, KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24(6)**, 861–862. https://doi.org/10.1093/bioinformatics/btm598.

Devi, KR, Narain, K, Mahanta, J, Nirmolia, T, Blair, D, Saikia, SP and Agatsuma, T (2013) Presence of three distinct genotypes within the *Paragonimus westermani* complex in northeastern India. *Parasitology* **140(1)**, 76–86. https://doi.org/10.1017/S0031182012001229.

Doanh, NP, Tu, AL, Bui, TD, Loan, TH, Nonaka, N, Horii, Y, Blair, D and Nawa, Y (2016) Molecular and morphological variation of *Paragonimus westermani* in Vietnam with records of new second intermediate crab hosts and a new locality in a northern province. *Parasitology* **143(12)**, 1639–1646. https://doi.org/10.1017/S0031182016001219.

Gurevich, A, Saveliev, V, Vyahhi, N and Tesler, G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29(8)**, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hardman, M and Hardman, LM (2006) Comparison of the phylogenetic performance of neodermatan mitochondrial protein-coding genes. *Zoologica Scripta* **35**, 655–665. https://doi.org/10.1111/j.1463-6409.2006.00248.x

Kamo, H, Nishida, H, Hatsushika, R and Tomimura, T (1961) On the occurrence of a new lung fluke, *Paragonimus miyazakii* n.sp. in Japan. (Trematoda: Troglotrematidae). *Yonago Acta Medica* **5**, 43–52.

Katoh, K and Standley, DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

780. https://doi.org/10.1093/molbev/mst010.

Kinkar, L, Korhonen, PK, Cai, H, Gauci, CG, Lightowlers, MW, Saarma, U, Jenkins, DJ, Li, J, Li, J, Young, ND and Gasser, RB (2019) Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. *Parasites and Vectors* **12**, 238. https://doi.org/10.1186/s13071-019-3492-x

Kinkar, L, Young, ND, Sohn, WM, Stroehlein, AJ, Korhonen, PK and Gasser, RB (2020) First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLOS Neglected Tropical Diseases* **14(8)**, e0008552. https://doi.org/10.1371/journal.pntd.0008552.

Koren, S, Walenz, BP, Berlin, K, Miller, JR, Bergman, NH and Phillippy, AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27(5)**, 722–736. https://doi.org/10.1101/gr.215087.116.

Kumar, S, Stecher, G, Li, M, Knyaz, C and Tamura, K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Laslett, D and Canback, B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175. https://doi.org/10.1093/bioinformatics/btm573.

Le, TH, Blair, D and McManus, DP (2001) Complete DNA sequence and gene organization of the mitochondrial genome of the liverfluke, *Fasciola hepatica* L. (Platyhelminthes; Trematoda). *Parasitology* **123**, 609–621. https://doi.org/10.1017/s0031182001008733.

Le, TH, Blair, D and McManus, DP (2002) Mitochondrial genomes of parasitic flatworms. *Trends in Parasitology* **18**, 206–213. https://doi.org/10.1016/s1471-4922(02)02252-3.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Le, TH, Nguyen, KT, Nguyen, NTB, Doan, HTT, Agatsuma, T and Blair, D (2019) The complete mitochondrial genome of *Paragonimus ohirai* (Paragonimidae: Trematoda: Platyhelminthes) and its comparison with *P. westermani* congeners and other trematodes. *PeerJ* **7**, e7031. https://doi.org/10.7717/peerj.7031.

Le, TH, Pham, LTK, Doan, HTT, Le, XTK, Saijuntha, W, Rajapakse, RPVJ and Lawton, SP (2020) Comparative mitogenomics of the zoonotic parasite *Echinostoma revolutum* resolves taxonomic relationships within the '*E. revolutum*' species group and the Echinostomata (Platyhelminthes: Digenea). *Parasitology* **147**, 566–576. https://doi.org/10.1017/S0031182020000128.

Li, Y, Qiu, YY, Zeng, MH, Diao, PW, Chang, QC, Gao, Y, Zhang, Y and Wang, CR (2019) The complete mitochondrial genome of Echinostoma miyagawai: comparisons with closely related species and phylogenetic implications. *Infectinon Genetics and Evolution* **75**, 10396. https://doi.org/10.1016/j.meegid.2019.103961.

Liu, GH, Gasser, RB, Young, ND, Song, HQ, Ai, L and Zhu, XQ (2014) Complete mitochondrial genomes of the 'intermediate form' of *Fasciola* and *Fasciola gigantica*, and their comparison with *F. hepatica. Parasites and Vectors* **7**, 150. https://doi.org/10.1186/1756-3305-7-150.

Lowe, TM and Chan, PP (2016) tRNAscan on-line: Search and contextual analysis of transfer RNA genes. *Nucleic Acids Research* **44**, W54–57. https://doi.org/10.1093/nar/gkw413.

Ma, J, He, JJ, Liu, GH, Leontovyč, R, Kašný, M and Zhu, XQ (2016) Complete mitochondrial genome of the giant liver fluke *Fascioloides magna* (Digenea: Fasciolidae) and its comparison with selected trematodes. *Parasites and Vectors* **9**, 429. https://doi.org/10.1186/s13071-016-1699-7.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Oey, H, Zakrzewski, M, Gravermann, K, Young, ND, Korhonen, PK, Gobert, GN, Nawaratna, S, Hasan, S, Martínez, DM, You, H, Lavin, M, Jones, MK, Ragan, MA, Stoye, J, Oleaga, A, Emery, AM, Webster, BL, Rollinson, D, Gasser, RB, McManus, DP and Krause, L (2019*a*) Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*. *PLOS Pathogens* **15(1)**, e1007513. https://doi.org/10.1371/journal.ppat.1007513.

Oey, H, Zakrzewski, M, Narain, K, Devi, KR, Agatsuma, T, Nawaratna, S, Gobert, GN, Jones, MK, Ragan, MA, McManus, DP and Krause, L (2019*b*) Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *Gigascience* **8(1)**, giy146. https://doi.org/10.1093/gigascience/giy146.

Ohari, Y, Suzuki, Y, Shibahara, T and Itagaki, T (2019) First report of *Paragonimus skrjabini miyazakii* metacercariae in *Geothelphusa dehaani* (Sawagani) occurring in Iwate Prefecture, Japan. *Journal of Veterinary Internal Medicine* **81(8)**, 1109–1112. https://doi:10.1292/jvms.19-0164.

Perna, NT and Kocher, TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal* of *Molecular Evolution* **41**, 353–358. https://doi.org/10.1007/BF00186547.

Pham, KLT, Saijuntha, W, Lawton, SP and Le, TH (2022) Mitophylogenomics of the zoonotic fluke *Echinostoma malayanum* confirms it as a member of the genus *Artyfechinostomum* Lane, 1915 and illustrates the complexity of Echinostomatidae systematics. *Parasitology Research* **121**, 899–913. https://doi.org/10.1007/s00436-022-07449-z.

Qian, L, Zhou, P, Li, W, Wang, H, Miao, T and Hu, L (2018) Characterization of the

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

complete mitochondrial genome of the lung fluke, *Paragonimus heterotremus*. *Mitochondrial DNA Part B* **3(2)**, 560–561. https://doi.org/10.1080/23802359.2018.1462119.

Rajapakse, RPVJ, Pham, KLT, Karunathilake, KJK, Lawton, SP and Le, TH (2020) Characterization and phylogenetic properties of the complete mitochondrial genome of *Fascioloides jacksoni* (syn. Fasciola jacksoni) support the suggested intergeneric change from *Fasciola* to *Fascioloides* (Platyhelminthes: Trematoda: Plagiorchiida). *Infectinon Genetics and Evolution* **82**, 104281. https://doi.org/10.1016/j.meegid.2020.104281.

Rhoads, A and Au, KF (2015) PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13(5)**, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002.

Rosa, BA, Choi, YJ, McNulty, S, Jung, H, Martin, J, Agatsuma, T, Sugiyama, H, Le, TH, Doanh, PN, Maleewong, W, Blair, D, Brindley, PJ, Fischer, PU and Mitreva, M (2020) Comparative genomics and transcriptomics of four *Paragonimus* species provide insights into lung fluke parasitism and pathogenesis. *Gigascience* **9(7)**, giaa073. https://doi.org/10.1093/gigascience/giaa073

Shekhovtsov, SV, Katokhin, AV, Kolchanov, NA and Mordvinov, VA (2010) The complete mitochondrial genomes of the liver flukes Opisthorchis felineus and Clonorchis sinensis (Trematoda). Parasitology International **59**, 100–103. https://doi.org/10.1016/j.parint.2009.10.012.

Singh, T, Hiromu, S, Devi, K and Singh, W (2015) First case of *Paragonimus westermani* infection in a female patient in India. *Indian Journal of Medical Microbiology* **33(5)**, Suppl 156–159.

Solà, E, Álvarez-Presas, M, Frías-López, C, Littlewood, DTJ, Rozas, J and Riutort, M

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

(2015) Evolutionary Analysis of Mitogenomes from Parasitic and Free-Living Flatworms. *PLOS One* **10(3)**, e0120081. https://doi.org/10.1371/journal.pone.0120081

Talavera, G and Castresana, J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564–577. https://doi.org/10.1080/10635150701472164.

Urabe, M and Marcaida, AJB (2022) A record of juvenile *Paragonimus skrjabini miyazakii* from the urinary bladder of the Japanese toad, *Bufo japonicus formosus*. *Parasitology International* **86**, 102474. https://doi:10.1016/j.parint.2021.102474.

Voronova, AN, Vainutis, KS, Tabakaeva, TV, Sapotsky, MV, Karaeka, NN, Volkov, YG, Galkina, IV and Shchelkanov, MY (2022) Molecular identification of the trematode *P. ichunensis* stat. n. from lungs of siberian tigers justified reappraisal of *Paragonimus westermani* species complex. *Journal of Parasitic Diseases* https://doi.org/10.1007/s12639-022-01481-7

Wang, T, Wang, Y, Xu, F, Li, X, Qu, R, Song, L, Tang, Y and Lin, P (2018) Characterization of the complete mitochondrial genome of the lung fluke, *Paragonimus kellicotti*. *Mitochondrial DNA Part B* **3(2)**, 715–716. https://doi.org/10.1080/23802359.2018.1483763.

Wey-Fabrizius, AR, Podsiadlowski, L, Herlyn, H and Hankeln, T (2013) Platyzoan mitochondrial genomes. *Molecular Phylogenetics and Evolution* **69**, 365–375. https://doi.org/10.1016/j.ympev.2012.12.015.

Further reading references

Fu YT, Jin YC, Li F and Liu GH (2019) Characterization of the complete mitochondrial

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

genome of the echinostome Echinostoma miyagawai and phylogenetic implications. *Parasitology Research* **118(10)**, 3091–3097. https://doi.org/10.1007/s00436-019-06417-4.

Lee D, Choe S, Park H, Jeon HK, Chai JY, Sohn WM, Yong TS, Min DY, Rim HJ and Eom KS (2013) Complete mitochondrial genome of Haplorchis taichui and comparative analysis with other trematodes. *Korean Journal of Parasitology* **51(6)**, 719–726. https://doi.org/10.3347/kjp.2013.51.6.719.

Littlewood DT, Lockyer AE, Webster BL, Johnston DA and Le TH (2006) The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. *Molecular Phylogenetics and Evolution* **39**, 452–467. https://doi.org/10.1016/j.ympev.2005.12.012.

Ma J, He JJ, Zhou CY, Sun MM, Cevallos W, Sugiyama H, Zhu XQ and Calvopiña M (2019) Characterization of the mitochondrial genome sequences of the liver fluke Amphimerus sp. (Trematoda: Opisthorchiidae) from Ecuador and phylogenetic implications. *Acta Tropica* **195**, 90–96. https://doi.org/10.1016/j.actatropica.2019.04.025

Ma J, Sun MM, He JJ, Liu GH, Ai L, Chen MX and Zhu XQ (2017) *Fasciolopsis buski* (Digenea: Fasciolidae) from China and India may represent distinct taxa based on mitochondrial and nuclear ribosomal DNA sequences. *Parasites and Vectors* **10(1)**, 101. https://doi.org/10.1186/s13071-017-2039-2.

Na L, Gao JF, Liu GH, Fu X, Su X, Yue DM, Gao Y, Zhang Y and Wang CR (2016) The complete mitochondrial genome of Metorchis orientalis (Trematoda: Opisthorchiidae): comparison with other closely related species and phylogenetic implications. *Infection Genetics and Evolution* **39**, 45–50. https://doi.org/10.1016/j.meegid.2016.01.010

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Yang X, Gasser RB, Koehler AV, Wang L, Zhu K, Chen L, Feng H, Hu M and Fang R (2015) Mitochondrial genome of Hypoderaeum conoideum - comparison with selected trematodes. *Parasites and Vectors* **8**, 97. https://doi.org/10.1186/s13071-015-0720-x.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Table 1. Locations of genes and other features in the mitochondrial genome of *Paragonimus skrjabini miyazakii* (17,591 bp; GenBank: ON782295)

Gene	Position (5'>3')	Length		Codon		tRNA anti-codon	Int. seq. length (bp)
		bp	aa	start	stop		
*cox*3	1–645	645	214	ATG	TAG		+3
$tRNA^{His}$ (*trn*H)	649–712	63				GTG	+3
cob	716–1833	1119	372	ATG	TAG		+6
*nad*4L	1840–2103	264	87	ATG	TAG		–40
*nad*4	2064–3323	1260	419	GTG	TAG		+17
$tRNA^{Gln}$ (*trn*Q)	3341–3402	62				TTG	+13
$tRNA^{Phe}$ (*trn*F)	3416–3482	67				GAA	+15
$tRNA^{Met}$ (*trn*M)	3498–3564	67				CAT	0
*atp*6	3565–4080	516	171	ATG	TAG		0
*nad*2	4081–4947	867	288	ATG	TAG		+2
$tRNA^{Val}$ (*trn*V)	4950–5012	64				TAC	+8
$tRNA^{Ala}$ (*trn*A)	5022–5088	67				TGC	+9
$tRNA^{Asp}$ (*trn*D)	5098–5162	65				GTC	0
*nad*1	5163–6068	906	301	ATG	TAG		+10
$tRNA^{Asn}$ (*trn*N)	6069–6149	71				GTT	0
$tRNA^{Pro}$ (*trn*P)	6150–6217	68				TGG	+8
$ttRNA^{Ile}$ (*trn*I)	6226–6287	62				GAT	+6
$tRNA^{Lys}$ (*trn*K)	6294–6359	66				CTT	+1
*nad*3	6361–6717	357	118	ATG	TAG		+5
$tRNA^{Ser1(AGN)*}$ (*trn*S_1)	6723–6783	61				GCT	+9
$tRNA^{Trp}$ (*trn*W)	6793–6858	66				TCA	+3
*cox*1	6862–8397	1536	511	ATG	TAA		+1
$tRNA^{Thr}$ (*trn*T)	8399–8463	65				TGT	0
*rrn*L (16S)	8464–9439	976					0
$tRNA^{Cys}$ (*trn*C)	9440–9508	69				GCA	0

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

*rrn*S (12S)	9509–10246	738					0
*cox*2	10246–10843	597	198	ATG	TAA		+7
*nad*6	10851–11303	453	150	GTG	TAA		+3
$tRNA^{Tyr}$ (*trn*Y)	11307–11370	64				GTA	+12
$tRNA^{Leu1(CUN)}$ (*trn*L_1)	11382–11445	64				TAG	−3
$tRNA^{Ser2(UCN)*}$ (*trn*S_2)	11443–11505	63				TGA	+22
$tRNA^{Leu2(UUR)}$ (*trn*L_2)	11528–11592	65				TAA	+3
$tRNA^{Arg}$ (*trn*R)	11596–11662	67				TCG	0
*nad*5	11663–13240	1578	525	GTG	TAG		+10
$tRNA^{Gly}$ (*trn*G)	13251–13317	67				TCC	+33
LRU region	13451–16663						
LRU1	13451–13734	384					0
$tRNA^{Glu1}$ (*trn*E1)	13544–13599	66					
LRU2	13735–14118	384					0
$tRNA^{Glu2}$ (*trn*E2)	13928–13993	66					
LRU3	14119–14502	384					0
$tRNA^{Glu3}$ (*trn*E3)	14312–14377	66					
LRU4	14503–14886	384					0
$tRNA^{Glu4}$ (*trn*E4)	14696–14761	66					
LRU5	14887–15270	384					0
$tRNA^{Glu5}$ (*trn*E5)	15079–15145	66					
LRU6	15271–15654	384					0
$tRNA^{Glu6}$ (*trn*E6)	15464–15529	66					
LRU7	15655–16038	384					0
$tRNA^{Glu7}$ (*trn*E7)	15848–15913	66					
LRU8	16039–16422	384					0
$tRNA^{Glu8}$ (*trn*E8)	16232–16297	66					0
LRU8.6	16423–17026	241					+9
SRU region							
SRU1	16673–16790	118					0
SRU2	16791–16898	118					0

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

SRU3	16899−17026	118	0
SRU4	17027−17144	118	0
SRU5	17145−17262	118	0
SRU6	17263−17380	118	0
SRU6.8	17381−17473	93	0
Int. seq.	17474−17591	119	0

bp: basepair; aa: amino acid; Int. seq.: intergenic sequence (+, number of nucleotides before start of the following gene; −, number of nucleotides overlapping with the following gene); LRU: Long tandem repeat unit, which contain a $tRNA^{Glu}$; SRU: short tandem repeat unit; Sterisk $^{(*)}$: indicating tRNAs lacking DHU-arm.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Table 2. Base composition and skew/skewness value for AT and GC of the protein-coding genes (PCGs), ribosomal genes (MRGs), and the coding region overall (5′-*cox*3-*nad*5-3′, designated as mtDNA*) of the mitogenomes of *Paragonimus skrjabini miyazakii* and other members of Paragonimidae

	Species/Strains	Sequence	Length (nt)	A (%)	T (%)	G (%)	C (%)	A+T (%)	AT-skew	G+C (%)	GC-skew
1	***Paragonimus skrjabini miyazakii*** (Pmiy-OkuST1-JP) /ON782295	PCGs	10098	15.83	44.76	27.70	11.71	60.59	–0.477	39.41	0.406
		MRGs	1714	22.11	37.11	27.77	13.01	59.22	–0.253	40.78	0.362
		mtDNA*	13240	17.33	42.90	27.72	12.05	60.23	–0.425	39.77	0.394
2	***Paragonimus heterotremus*** (Phet-LC-VN) /KY952166	PCGs	10101	14.97	43.85	28.17	13.02	58.82	–0.491	41.19	0.368
		MRGs	1720	20.47	37.15	28.55	13.84	57.62	–0.289	42.39	0.347
		mtDNA*	13230	16.35	42.12	28.20	13.33	58.47	–0.441	41.53	0.358
3	***Paragonimus heterotremus*** (Phet-GX-CN) /MH059809	PCGs	10101	14.95	43.90	28.22	12.93	58.85	–0.492	41.15	0.372
		MRGs	1711	20.81	37.05	28.23	13.91	57.86	–0.281	42.14	0.340
		mtDNA*	13222	16.34	42.13	28.26	13..28	58.47	–0.441	41.53	0.361
4	***Paragonimus ohirai*** (Pohi-Kino-JP) /KX765277	PCGs	10104	17.00	46.00	26.17	10.83	63.00	–0.460	37.00	0.415
		MRGs	1710	22.92	37.19	26.90	12.98	60.01	–0.238	39.88	0.349
		mtDNA*	13202	18.38	43.86	26.32	11.45	62.24	–0.409	37.78	0.394
5	***Paragonimus westermani*** (Pwes(2n)-Haenam-KR) /AF540958	PCGs	10101	13.40	38.21	30.61	17.77	51.61	–0.481	48.39	0.265
		MRGs	1732	19.11	32.62	31.12	17.15	51.73	–0.261	48.27	0.289
		mtDNA*	13213	14.75	36.96	30.73	17.56	51.71	–0.430	48.29	0.273
6	***Paragonimus westermani*** (Pwes(3n)-Bogil-KR) /AF219379	PCGs	10101	13.33	38.25	30.61	17.81	51.58	–0.483	48.42	0.264
		MRGs	1732	18.94	32.79	31.29	16.97	51.73	–0.268	48.26	0.297
		mtDNA*	13213	14.65	37.02	30.76	17.56	51.67	–0.433	48.33	0.273

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

	Paragonimus westermani	PCGs	10104	13.41	38.33	30.56	17.70	51.74	–0.482	48.26	0.267
7	Pwes-(dog1)-CN	MRGs	1731	19.06	32.52	31.14	17.27	51.58	–0.261	48.42	0.287
	/MN412705	mtDNA*	13125	14.74	37.07	30.70	17.50	51.81	–0.431	48.19	0.274
	Paragonimus westermani	PCGs	10104	14.89	40.50	29.36	15.25	55.39	–0.462	44.61	0.316
	(Pwes-TypeI-IN)	MRGs	1729	20.24	35.11	29.84	14.81	55.35	–0.269	44.65	0.337
8	/KM280646	mtDNA*	13223	16.21	38.98	29.57	15.25	55.19	–0.413	44.81	0.320
	Paragonimus westermani	PCGs	10095	14.74	40.45	29.59	15.23	55.19	–0.466	44.82	0.320
	(Pwes-AP-IN)	MRGs	1721	20.22	34.63	30.16	14.99	54.85	–0.263	45.15	0.336
9	/KX943544	mtDNA*	13216	16.13	38.88	29.76	15.23	55.01	–0.414	44.99	0.323
	Paragonimus westermani	PCGs	10104	14.92	40.95	29.19	14.93	55.87	–0.466	44.12	0.323
1	(Pwes-IND2009-IN)	MRGs	1732	20.32	35.10	29.79	14.78	55.42	–0.267	44.57	0.337
0	/CM017921	mtDNA*	13230	16.25	39.34	29.42	14.99	55.59	–0.415	44.41	0.325
	Paragonimus kellicotti	PCGs	10098	15.22	44.74	27.78	12.26	59.96	–0.492	40.04	0.388
1	(Pkel-Ozark-US)	MRGs	1711	21.33	37.52	27.47	13.68	58.85	–0.275	41.15	0.335
1	/MH322000	mtDNA*	13196	16.57	42.85	27.89	12.69	59.42	–0.442	40.58	0.375

PCGs: protein-coding gene; MRGs: mitoribosomal gene; mtDNA*: mitochondrial coding nucleotide sequence (from 5' *cox*3 to 3' *nad*5).

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

Table 3. Estimation of pairwise genetic distances (%) among *Paragonimus skrjabini miyazakii* and strains/species of the family Paragonimidae inferred from the concatenated amino acid sequences of 12 mitochondrial protein-coding genes

	Strains	1	2	3	4	5	6	7	8	9	10
1	*Paragonimus skrjabini miyazakii* (Pmiy-OkuST1-JP, ON782295)										
2	*Paragonimus heterotremus* (Phet-LC-VN, KY952166)	**15.12**									
3	*Paragonimus heterotremus* (Phet-GX-CN, MH059809)	**14.71**	1.25								
4	*Paragonimus ohirai* (Pohi-Kino-JP, KX765277)	18.25	18.26	17.68							
5	*Paragonimus westermani* (Pwes(2n)-Haenam-KR, AF540958)	22.57	22.35	21.69	24.60						
6	*Paragonimus westermani* (Pwes(3n)-Bogil-KR, AF219379)	22.22	22.04	21.38	24.29	0.23					
7	*Paragonimus westermani* (Pwes-(dog1)-CN, MN412705)	21.88	21.65	21.00	23.80	1.21	0.98				
8	*Paragonimus westermani* (Pwes-AP-IN, KX943544)	**24.72**	24.77	24.18	27.24	13.74	13.44	13.58			
9	*Paragonimus westermani* (Pwes-(typeI)-IN, KM280646)	**24.34**	24.13	23.48	27.04	13.44	13.14	13.23	7.35		
10	*Paragonimus westermani* (Pwes-IND2009-IN, CM017921)	22.71	22.63	22.02	25.43	14.61	14.31	14.18	4.78	3.71	
11	*Paragonimus kellicotti* (Pkel-Ozark-US, MH322000)	**16.21**	16.09	15.57	18.54	21.52	21.18	20.90	24.26	23.83	22.47

Information for strain/ species is given in **Table S1**. Pairwise genetic distance (%) among *P. s. miyazakii* and species of Paragonimidae showing the lowest (numbers 2–3, and 11) and the highest (numbers 8–9) are bolded. The intra-specific distances between strains within the *P. heterotremus* complex and within the *P. westermani* complex (the East-Asian groups and the Indian groups) are boxed.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

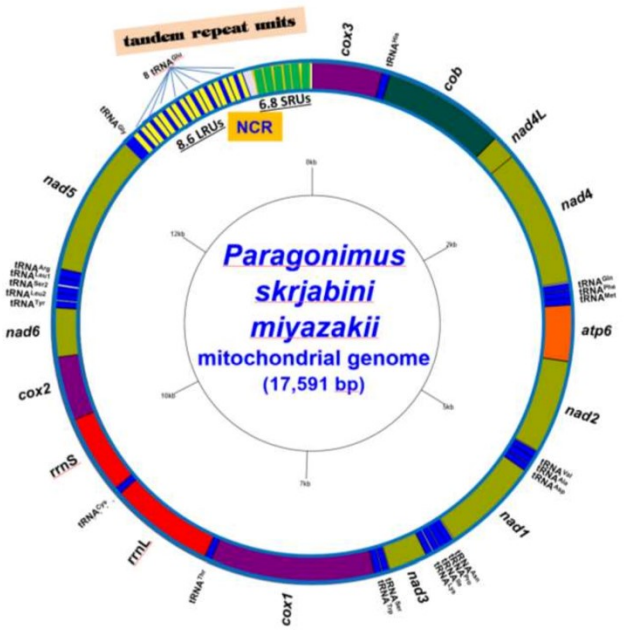

Fig. 1. A schematic drawing of a circular map of the mitogenome of *Paragonimus skrjabini miyazakii* (GenBank: ON782295). The protein-coding (PCGs) and mitoribosomal large and small subunit genes (MRGs) are abbreviated as in our previous publications (Le *et al*., 2019, 2020). The transfer RNA genes (tRNAs) are marked with three-letter amino acid abbreviations (see Table 1). The non-coding region (NCR) contains tandem repeat units, and is located between $tRNA^{Gly}$ and *cox*3 with 8.6 LRUs (long, LRU1–8.6) including $tRNA^{Glu}$ in each (8 $tRNA^{Glu}$) and 6.8 SRUs (short repeat units, SRU1–6.8).

Accepted Manu

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

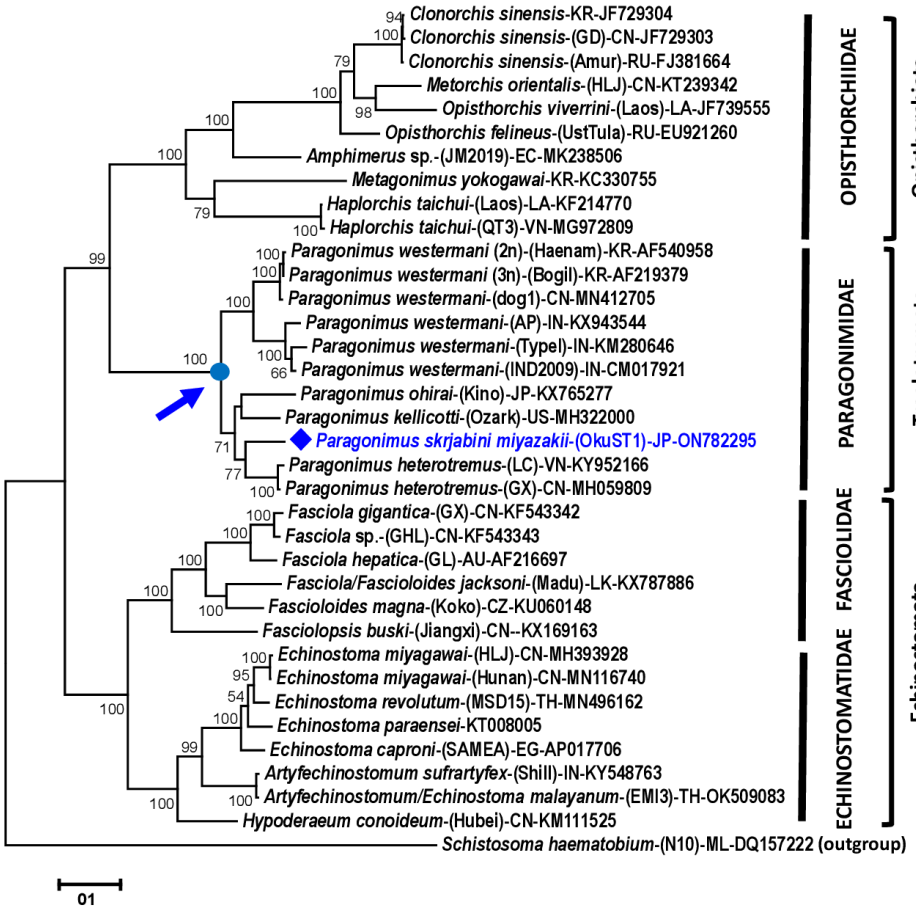

Fig. 2. A maximum likelihood phylogenetic tree showing the position of *Paragonimus skrjabini*

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press

miyazakii (diamond symbol) based on the analysis of the complete concatenated amino acids of 12 PCGs of 36 trematode species/strains. *Schistosoma haematobium* (Platyhelminthes: Schistosomatidae) is included as an outgroup. Nodal support values evaluated using 1000 bootstrap resamplings are shown on each branch. The basal node of the Paragonimidae is shown by an arrow. Accession numbers are given for each species/strain and country of origin (in brackets) (where available) at the end of each sequence. The scale bar represents the number of substitutions per site.

Accepted Manuscript

https://doi.org/10.1017/S0031182022001184 Published online by Cambridge University Press