# Marked mitochondrial genetic variation in individuals and populations of the carcinogenic liver fluke *Clonorchis sinensis*

**Liina Kinkar**[1], **Pasi K. Korhonen**[1], **Daxi Wang**[1,2], **Xing-Quan Zhu**[3], **Galina N. Chelomina**[4], **Tao Wang**[1], **Ross S. Hall**[1], **Anson V. Koehler**[1], **Ivon Harliwong**[5], **Bicheng Yang**[5], **J. Lynn Fink**[5], **Neil D. Young**[1]\*, **Robin B. Gasser**[1]\*

**1** Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia, **2** BGI International, Shenzhen, China, **3** State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, Gansu Province, China, **4** Department of Parasitology, Federal Scientific Center of the East Asia Terrestrial Biodiversity FEB RAS, Vladivostok, Russia, **5** BGI Australia, Herston, Queensland, Australia

\* nyoung@unimelb.edu.au (NDY); robinbg@unimelb.edu.au (RBG)

## Abstract

Clonorchiasis is a neglected tropical disease caused by the Chinese liver fluke, *Clonorchis sinensis*, and is often associated with a malignant form of bile duct cancer (cholangiocarcinoma). Although some aspects of the epidemiology of clonorchiasis are understood, little is known about the genetics of *C. sinensis* populations. Here, we conducted a comprehensive genetic exploration of *C. sinensis* from endemic geographic regions using complete mitochondrial protein gene sets. Genomic DNA samples from *C. sinensis* individuals (n = 183) collected from cats and dogs in China (provinces of Guangdong, Guangxi, Hunan, Heilongjiang and Jilin) as well as from rats infected with metacercariae from cyprinid fish from the Russian Far East (Primorsky Krai region) were deep sequenced using the BGISEQ-500 platform. Informatic analyses of mitochondrial protein gene data sets revealed marked genetic variation within *C. sinensis*; significant variation was identified within and among individual worms from distinct geographical locations. No clear affiliation with a particular location or host species was evident, suggesting a high rate of dispersal of the parasite across endemic regions. The present work provides a foundation for future biological, epidemiological and ecological studies using mitochondrial protein gene data sets, which could aid in elucidating associations between particular *C. sinensis* genotypes/haplotypes and the pathogenesis or severity of clonorchiasis and its complications (including cholangiocarcinoma) in humans.

## Author summary

Clonorchiasis is an important neglected tropical disease caused by the Chinese liver fluke, *Clonorchis sinensis*, which can induce malignant bile duct cancer (cholangiocarcinoma). Little precise information is available on the biology, epidemiology and population

genetics of *C. sinensis*. For this reason, we explored here the genetic composition of *C. sinensis* populations in distinct endemic areas in China and Russia. Using a deep sequencing-informatic approach, we revealed marked mitochondrial genetic variation within and between individuals and populations of *C. sinensis*, with no particular affiliation with geographic or host origin. These molecular findings and the methodology established should underpin future genetic studies of *C. sinensis* causing human clonorchiasis and associated complications (cancer) as well as transmission patterns in endemic regions.

## Introduction

Eleven pathogens are recognised to cause one fifth of all human cancers worldwide [1]. One of them is the Chinese liver fluke, *Clonorchis sinensis* (phylum Platyhelminthes; class Trematoda; family Opisthorchiidae). Chronic infection with this fluke can induce cholangiocarcinoma in humans–a malignant and fatal form of bile duct cancer [2].

The disease (clonorchiasis) caused by *C. sinensis* is of major public health concern in parts of Asia [3–5]; ~ 15 million people are infected, and the highest endemicity is in South Korea, Vietnam, China and the Russian Far East [6, 7]. Most of these people (~ 13 million) live within provinces in the southern (Guangdong, Guangxi, and Hunan) and north-eastern (Heilongjiang and Jilin) parts of China [4, 6–8]. In endemic areas, millions of tourists per annum are also at risk of infection, as they often consume local delicacies made from raw fish or shrimp [9]. Humans and other piscivorous mammals (e.g., canids and felids) are definitive hosts, and acquire infection by eating raw freshwater (cyprinid) fish or shrimp containing encysted larvae (= metacercariae) which ultimately establish as adult worms in the biliary system [7, 8]. Here, eggs produced by adult *C. sinensis* are released into aquatic environments to infect small snail intermediate hosts (order Mesogastropoda); larval stages (cercariae) released from the snails then usually infect fish (family Cyprinidae) and encyst as metacercariae.

An epidemiological factor contributing significantly to human clonorchiasis relates to human behaviour. In some endemic regions, such as in the Chinese provinces of Guangdong and Guangxi, it is commonplace in fish-producing areas for toilets to be built adjacent to fish ponds, so that human faeces containing *C. sinensis* eggs enter fish ponds; such eggs are thus accessible to snail hosts, and then infect cyprinid fish [8]. In such regions, the often high dependence on aquaculture for food; fish producers' unawareness of the life cycle of *C. sinensis*; the intimate involvement of canids and felids in the transmission; and environmental and climatic (temperature and humidity) conditions all contribute significantly to the endemicity of clonorchiasis [6–8, 10–12].

Although these aspects of the epidemiology of clonorchiasis (transmission and prevalence) are relatively well understood [4, 8], there is little comprehensive information on the genetics of *C. sinensis* and what impact genetic variation within and among populations has on the biology and transmission of *C. sinensis* infection and the severity of clonorchiasis and/or associated cancer. Published population genetic studies conducted using selected genetic markers, including the internal transcribed spacers (ITS-1 and/or ITS-2) of nuclear ribosomal DNA and the mitochondrial *cox*1 gene regions, have indicated variable levels of genetic variation and sub-structuring within *C. sinensis* populations [13–22]. Other studies have reported draft nuclear genomes of *C. sinensis* from Korea and China [23, 24] and four mitochondrial genomes of individual isolates from Korea, China and the Russian Far East [25–27].

Given the known utility of mitochondrial genomic data sets for investigating genetic variation within flatworm taxa [28, 29], encoded protein genes provide a sound basis for

comprehensive population genetic studies. For this reason, we undertook the first detailed population genetic exploration of *C. sinensis* from high-endemicity regions in China and the Russian Far East using complete mitochondrial protein-encoding gene data sets. This work provides a foundation for future biological, epidemiological and ecological studies of this parasite.

## Methods

### Ethics statement

The collection of *C. sinensis* in China was approved by the South China Agricultural University [30], following the Animal Ethics Guidelines of the People's Republic of China. The production of *C. sinensis* in rats was approved by the Ethics Committee for Animal Experimentation of the Institute of Biology and Soil Science, Russia (permit no. 3_02.06.2011; [21]).

### Parasite material

Adult *C. sinensis* worms (n = 183) were collected from China and the Russian Far East–two regions endemic for *C. sinensis* [4] (Table 1 and S1 Table). Specimens from China were from the livers of naturally infected cats or dogs originating from the provinces of Guangdong, Guangxi, Hunan, Heilongjiang and Jilin. Biliary ducts were flushed with physiological saline, intact *C. sinensis* worms collected, washed extensively in physiological saline and then fixed in 70% (v/v) ethanol. Specimens from the Russian Far East (Primorsky Krai region) were from rats experimentally infected with metacercariae collected from naturally infected freshwater fish (family Cyprinidae). All samples were fixed in ethanol and stored at -80°C until further use.

### DNA isolation, amplification and sequencing of mitochondrial protein genes

The most posterior tip (2–4 mm) of each of the 183 adult specimens of *C. sinensis* was excised under a dissecting microscope (10× magnification) using a sterile scalpel, in order to exclude eggs from the reproductive tract. Genomic DNA was extracted from each of these tips using the DNeasy Blood & Tissue Kit (Qiagen), following the manufacturer's protocol. DNA amounts were determined using a Qubit fluorometer dsDNA HS kit (Invitrogen).

**Table 1. Geographical origin and host species of the 183 specimens of *Clonorchis sinensis* studied herein.**

| Origin[a] | Dog | Cat | Cyprinid fish[b] | Totals |
|---|---|---|---|---|
| **China** | | | | |
| Guangdong | 70 | 61 | | 131 |
| Guangxi | | 13 | | 13 |
| Hunan | 5 | | | 5 |
| Heilongjiang | 9 | | | 9 |
| Jilin | 13 | | | 13 |
| **Russia** | | | | |
| Primorsky Krai | | | 12 | 12 |
| Totals: | 97 | 74 | 12 | 183 |

[a] Heilongjiang and Jilin are northern provinces, and Guangdong, Guangxi and Hunan are southern provinces in China.

[b] Specimens from rats experimentally infected with metacercariae from cyprinid fish.

https://doi.org/10.1371/journal.pntd.0008480.t001

Subsequently, individual DNA samples (representing individual *C. sinensis* adults) were whole genome-amplified using the REPLI-g Mini Kit (QIAGEN; cat. no. 150025), and individual genomic DNA libraries constructed using the DNA Library Prep Kit (MGI, item no. 1000002505), employing a well-established protocol [31], and then sequenced (100 bp paired-end reads) using the BGISEQ-500 platform (BGI Australia).

### Detection of unambiguous (i.e. fixed) nucleotide variations

For each of the 183 *C. sinensis* specimens, raw DNA sequence data in the FASTQ format [32] were filtered for quality (Phred quality score cut-off: 20) and trimmed using the program trimmomatic v.0.38 [33]. Filtered read-pairs were mapped to a reference mitochondrial genome of *C. sinensis* (GenBank accession no. KY564177–a sequence without polymorphism; [27]) using the Burrows-Wheeler Aligner (BWA) v.0.7.8 [34]. For each sample, single nucleotide variations (SNVs) at individual positions and insertion/deletion events (indels) in relation to this reference were recorded using a workflow in the Genome Analysis Toolkit (GATK v.4.0.8.1; [35]), and merged into one 'variant call format' (VCF) file that listed all variable sites for all samples. This file was used to produce a consensus FASTA file containing the sequence data for each individual *C. sinensis* specimen using BCFtools v.1.9 [36]. The open reading frame (ORF) of each protein gene was verified and conceptually translated using the program Geneious v.11.1.5 [37] employing the mitochondrial genetic code for echinoderms and flatworms (cf. [29]; https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi; translation_table 9). For each of the 183 *C. sinensis* specimens, the nucleotide sequences of all 12 protein genes and conceptually translated amino acid sequences were defined and deposited in the GenBank database (accession nos. MT292110–MT292292).

### Detection of nucleotide polymorphism within individual *C. sinensis* specimens

Duplicates were removed from the BWA-mapped sequence data using Picard v.2.0.1 (http://broadinstitute.github.io/picard). Subsequently, SAMtools v.1.9 [38] was used to compile an mpileup file; this file provided a summary of nucleotide coverage at each position following mapping to the reference sequence. For each specimen, any multiple-nucleotide variation (i.e. sequence polymorphism or microheterogeneity) at individual positions with respect to the reference was identified using the program VarScan v.2.4.3 [39]. The percentages of individual bases contributing to each polymorphic position (i.e. 0 to 100%) were recorded in a VCF file; the minimum base-contribution for a position to be recorded as polymorphic was $\geq$ 10%, with a minimum of 1400 nucleotides mapping to such a position. Each polymorphic position recorded was verified by eye using the integrative genomics viewer (IGV) v.2.5.2 [40].

### Population genetic analyses

All 12 protein gene sequences of all 183 *C. sinensis* individuals produced here (excluding polymorphic positions) and of four respective, published sequences (accession nos. KY564177 from Korea; JF729304 from Korea, JF729303 from China; FJ381664 from Russia; [25–27]) were aligned, as were their amino acid sequences.

The aligned nucleotide sequence data (Mendeley Data doi: 10.17632/dj4ybp77jw.2) were subjected to phylogenetic analyses, employing *Opisthorchis felineus* (family Opisthorchiidae) (accession no. EU921260; [25]) as an outgroup–as required. First, a Bayesian (BI) tree was built using Monte Carlo Markov Chain (MCMC) analysis [41, 42] in MrBayes v.3.2.6 [43]. The general time-reversible model, employing a gamma-distribution and a proportion of invariable sites (GTR+I+G; [44, 45]), was established to provide the best-fit model of sequence

evolution using the program PartitionFinder v.2.1.1 [46]. Posterior probability (pp) values were calculated by running 40 million generations, with 25% discarded as non-converged burn-in. Second, a maximum likelihood (ML) method was used to infer nodal support values for the BI tree using the program RAxML v.8.2.9 [47] employing 1000 bootstrap replicates. Third, a median-joining (MJ) network [48] was constructed using the program PopART [49], with epsilon set at zero. The aligned amino acid sequences were used to further assess phylogenetic clustering and support.

Subsequently, population genetic diversity and differentiation were assessed in relation to geographical origin or known host species. Population diversity indices (number of haplotypic sequences, haplotype diversity, nucleotide diversity, and mean number of nucleotide variations between haplotypic sequences) were estimated using DnaSP v.6.12.03 [50]. To estimate the degree of genetic differentiation between *C. sinensis* subpopulations (in relation to geographic origin or definitive host species with natural *C. sinensis* infection), the pairwise fixation index ($F_{ST}$) was calculated using Arlequin v.3.5.2.2 [51]. To assess patterns of nucleotide variation in all mitochondrial protein genes, a sliding window analysis of nucleotide diversity (steps of 10 bp over 300 bp-windows) was performed using the PopGenome package [52] in the R software environment [53]. For each data set, nucleotide diversity values were plotted against midpoint positions in each window using the R package ggplot2 [54].

## Results

### Sequence data sets and mapping results

A large amount of high-quality, trimmed short-read mitochondrial genomic data (mean: 665 Mb) for each of the 183 *C. sinensis* specimens was mapped to the reference mitochondrial genome sequence (accession no. KY564177), revealing substantial depth (median: 7268 nucleotides) at each position in each of the 12 protein genes. This mapping showed that most specimens (n = 164) had haplotypic sequences (without polymorphism within individuals), while a small subset of *C. sinensis* specimens (n = 19) had sequences with ≤ 3 polymorphic positions (Table 2). Such specimens originated mostly from Guangdong (n = 17), whereas single specimens were from Guangxi and Hunan.

### Marked mitochondrial genetic variation within *C. sinensis*

An analysis of haplotypic diversity indicated marked genetic variation among the 183 *C. sinensis* specimens (Hd = 0.998; Table 3), and within all sub-populations relating to geographical provenance (0.939 to 1.000) and host species (cat, dog or cyprinid fish) (0.939 to 1.000). The number of nucleotide differences among haplotypes/individuals (upon pairwise comparison) varied from 1 to 75, with a mean of 36 (Table 3). Nucleotide diversity values indicated that genetic variation within *C. sinensis* in the Chinese provinces of Jilin (π = 0.00223; Table 3) and Hunan (0.00167) was lower than in other geographical locations (ranging from 0.00300 to 0.00384). *C. sinensis* specimens originating from cats, dogs or cyprinid fish displayed similar levels of genetic diversity, with nucleotide diversity values ranging from 0.00344 to 0.00364. Having established marked variation among haplotypes, we proceeded to investigate the genetic relationships of haplotypes from different geographic provenances and definitive host species (cats and dogs).

### Affiliation of mitochondrial haplogroups with geographical provenance or host species

Most haplotypes defined herein clustered into 12 genetically closely-related groups (haplogroups A to L), with more than 3 haplotypes per haplogroup (Fig 1). Haplogroups were well-

**Table 2. List of polymorphic positions in the sequences of five mitochondrial protein genes derived from 19 of a total of 183 specimens of *Clonorchis sinensis*.**

| Gene | Position in the reference sequence[a] | Base in the reference sequence | Polymorphism recorded and base-contribution (in %) | Codon position | Specimen nos. |
|---|---|---|---|---|---|
| *cox*3 | 96 | T | T (71) and C (29) | 3 (synonymous) | 21[b] |
| *cox*3 | 191 | A | A (89) and G (11) | 2 (non-synonymous) | 58, 59, 60 and 61 |
| *cox*3 | 219 | A | A (87–90) and G (10–13) | 3 (synonymous) | 58, 60 and 61 |
| *cox*3 | 229 | T | T (86–90) and G (10–14) | 1 (non-synonymous) | 34, 35, 36, 37, 58, 60, 61 and 64 |
| *cox*3 | 330 | C | C (84) and T (16) | 3 (synonymous) | 12 |
| *nad*3 | 6718 | A | A (52) and G (48) | 1 (non-synonymous) | 189[b] |
| *cox*1 | 8152 | T | T (90) and A (10) | 3 (synonymous) | 160 |
| *cox*1 | 8414 | T | T (86–88) and G (12–14) | 1 (non-synonymous) | 75, 100 and 125 |
| *nad*6 | 11340 | T | T (52) and A (48) | 2 (non-synonymous) | 177[b] |
| *nad*5 | 12832 | A | A (75) and G (25) | 1 (non-synonymous) | 30[b] |
| *nad*5 | 13009 | A | A (82) and G (18) | 1 (non-synonymous) | 154 |

[a] Sequence with GenBank accession no. KY564177 [27].

[b] Base-contribution of the minor-frequency nucleotide: > 20%.

https://doi.org/10.1371/journal.pntd.0008480.t002

supported by BI and ML analyses, with near-absolute nodal support values (pp > 0.95 and bs ≥ 0.93), with the exception of haplogroup A, for which bs was 0.6 (S1 Fig). $F_{ST}$ values of 0.43692 to 0.80709 (S2 Table) indicated a clear genetic differentiation of individual

**Table 3. Diversity indices calculated for the 183 *Clonorchis sinensis* specimens used in the present study.**

| | No. of isolates | No. of haplotypes | Haplotype diversity (Hd) ± standard deviation | Nucleotide diversity (π) ± standard deviation | Average no. of nucleotide differences |
|---|---|---|---|---|---|
| Complete data set | 183 | 162 | 0.998 ± 0.001 | 0.00356 ± 0.00009 | 36 |
| **Regions** | | | | | |
| 'North' (northern China and Russia) | 34 | 29 | 0.989 ± 0.010 | 0.00325 ± 0.00027 | 33 |
| Northern China | 22 | 21 | 0.996 ± 0.015 | 0.00300 ± 0.00035 | 30 |
| Russia | 12 | 9 | 0.939 ± 0.058 | 0.00350 ± 0.00045 | 36 |
| 'South' (southern China) | 149 | 133 | 0.998 ± 0.001 | 0.00359 ± 0.00010 | 36 |
| **Provinces in China** | | | | | |
| Heilongjiang | 9 | 9 | 1.000 ± 0.052 | 0.00384 ± 0.00053 | 39 |
| Jilin | 13 | 13 | 1.000 ± 0.030 | 0.00223 ± 0.00031 | 23 |
| Guangdong | 131 | 119 | 0.998 ± 0.001 | 0.00346 ± 0.00009 | 35 |
| Guangxi | 13 | 13 | 1.000 ± 0.030 | 0.00351 ± 0.00034 | 36 |
| Hunan | 5 | 5 | 1.000 ± 0.126 | 0.00167 ± 0.00029 | 17 |
| **Host species** | | | | | |
| Cat | 74 | 73 | 1.000 ± 0.002 | 0.00364 ± 0.00012 | 37 |
| Dog | 97 | 88 | 0.998 ± 0.002 | 0.00344 ± 0.00014 | 35 |
| Cyprinid fish[a] | 12 | 9 | 0.939 ± 0.058 | 0.00350 ± 0.00045 | 36 |

[a] Specimens from rats experimentally infected with metacercariae from naturally infected cyprinid fish.

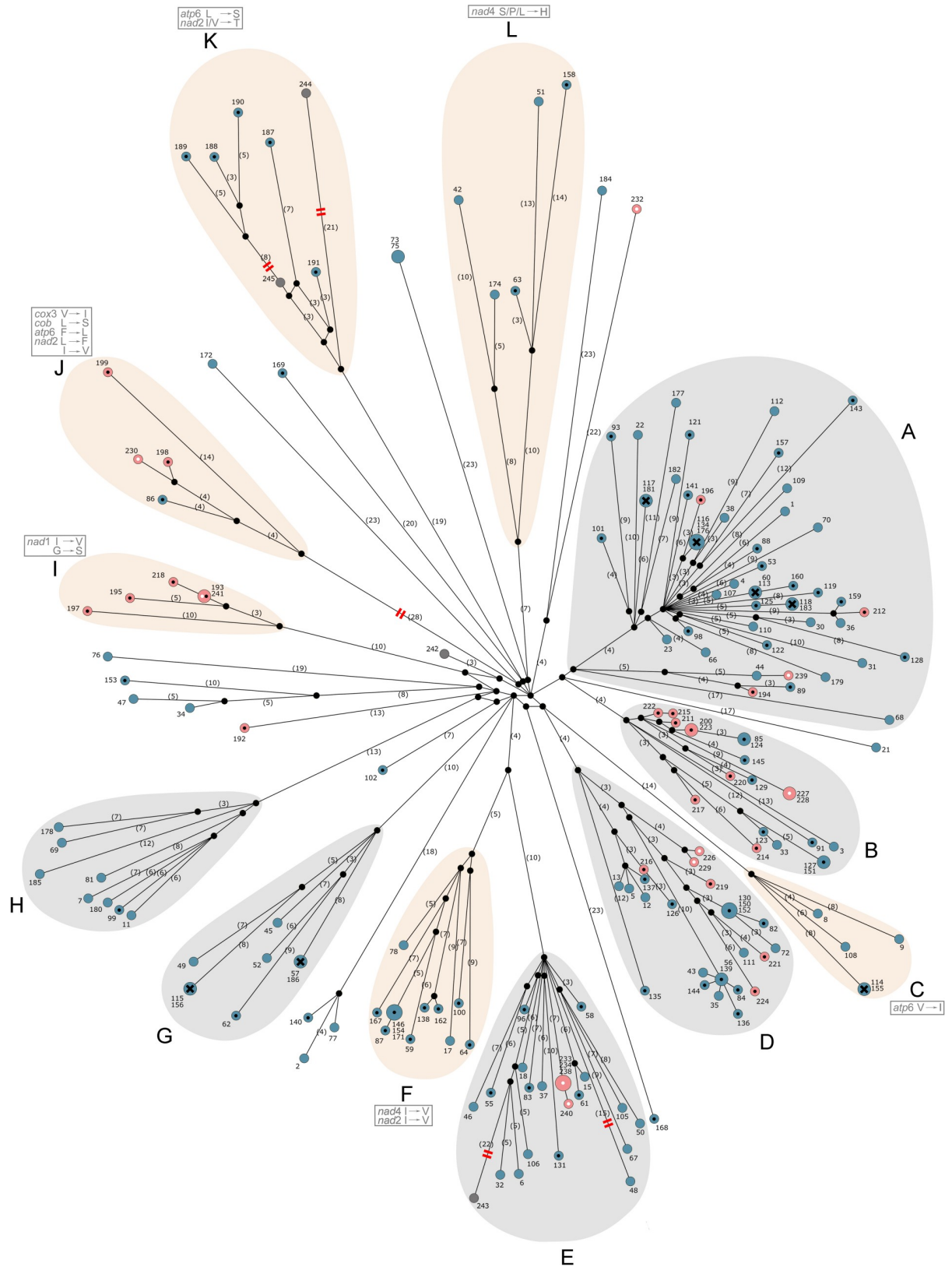https://doi.org/10.1371/journal.pntd.0008480.t003

**Fig 1. Median-joining haplotype network constructed using concatenated sequences for the 12 mitochondrial protein genes representing 183 *Clonorchis sinensis* specimens and four published mitochondrial genomes (GenBank accession nos. JF729303, FJ381664, JF729304**

**and KY564177).** Pink and blue circles represent haplotypic sequences produced in the present study; blue represents specimens from southern China (Guangdong, Hunan and Guangxi provinces), pink represents northern China (Heilongjiang and Jilin provinces) and the Russian Far East (Primorsky Krai region). Host species are indicated with a black dot (dog), solid colour (cat), black cross (cat and dog) or white dot (rats experimentally infected with metacercariae from cyprinid fish). Dark grey circles are haplotypic sequences from GenBank–haplotype 242 represents a *C. sinensis* specimen from a hamster experimentally infected with metacercariae obtained from cyprinid fish from Russia (FJ381664; [25]), haplotype 243 is from a cat from China (accession no. JF729303; [26]), haplotype 244 is from a cat from Korea (JF729304; [26]), and haplotype 245 is from a hamster experimentally infected with metacercariae obtained from cyprinid fish from South Korea (KY564177; [27]). Black circles at nodes represent median vectors (i.e. haplotypes not sampled, or extinct). Numbers of mutations ($\geq 3$) between haplotypes and/or median vectors in parentheses. Specimen numbers–not in parantheses (see S1 Table). Black lines indicate network edges, the lengths of which are proportional to the number of mutations (note that several mutations are shared by some edges in the network). Red lines on branches indicate reduced edge-lengths. Clusters A to L indicate individual haplogroups; a grey outline indicates a haplogroup supported by nucleotide sequence data; a beige outline indicates a haplogroup supported by both nucleotide and amino acid sequence data. Unique amino acid substitutions in protein sequences characteristic of all haplotypes of a respective cluster are indicated by grey boxes; amino acid abbreviations: V–valine, I–isoleucine, G–glycine, S–serine, L–leucine, F–phenylalanine, P–proline, H–histidine, T–threonine.

haplogroups from one other. Analyses of amino acid data sets further supported the clustering of six haplogroups (C, F, I, J, K and L; Fig 1).

Subsequently, we proceeded to investigate whether there was an affiliation between haplogroups and geographical provenance. First, we tested the hypothesis that there is a genetic distinction between the 'south' and the 'north'; however, no unequivocal genetic differentiation could be identified. Although haplogroup I comprised haplotypes exclusive to the 'north' (Primorsky Krai, Heilongijang and Jilin) and haplogroups C, F, G, H and L comprised haplotypes that were only in the 'south' (Guangdong and Guangxi), for several clusters (A, B, D, E and J) 'southern' and 'northern' haplotypes grouped together, indicating their close genetic relationship. This lack of differentiation was well-supported by low $F_{ST}$ values between 'north' and 'south' (0.02893, $p < 0.00001$). The sliding window analysis of nucleotide diversity also indicated a lack of genetic distinctiveness in *C. sinensis* between the two geographic regions. We detected similar patterns of genetic diversity between the two subpopulations, with minor differences in the range of values for particular genes, being often more pronounced for specimens originating from the 'north' (Fig 2). Subsequently, we subdivided northern specimens into northern China and the Russian Far East, to assess whether these subpopulations might be genetically differentiated. However, $F_{ST}$ values did not indicate a genetic differentiation in *C. sinensis* between these regions–low values were recorded between southern China and Russia (0.04095, $p < 0.01$), southern China and northern China (0.03542, $p < 0.001$), and the Russian Far East and northern China (0.04969, $p < 0.05$). Nonetheless, we did detect some evidence of geographical sub-structuring. Within haplogroup K, two samples from Korea represented by sequences with accession nos. KY564177 and JF729304 clustered together, as did all five specimens from Hunan. $F_{ST}$ values (0.42205 to 0.56811; S3 Table) revealed a clear genetic distinction between subpopulations from Hunan, other Chinese provinces and the Russian Far East (Primorsky Krai region), with significantly lower $F_{ST}$ values (0.02364 to 0.0999; S3 Table) between other provinces in China and Primorsky Krai in Russia–upon pairwise comparison.

Separate analyses did not find evidence for an affiliation between particular haplogroups and known definitive host species, as all genetic clusters, except haplogroups I and J, represented specimens from cats and dogs, and some specimens from cats and dog hosts were genetically identical (i.e. haplogroups A, C and G; see Fig 1). This finding was further supported by low values of $F_{ST}$ between the two host species (0.01023, $p < 0.01$).

## Discussion

Here, we discovered marked mitochondrial genetic variation within *C. sinensis*. This variation in protein genes was present within individual worms, between individual worms, and within
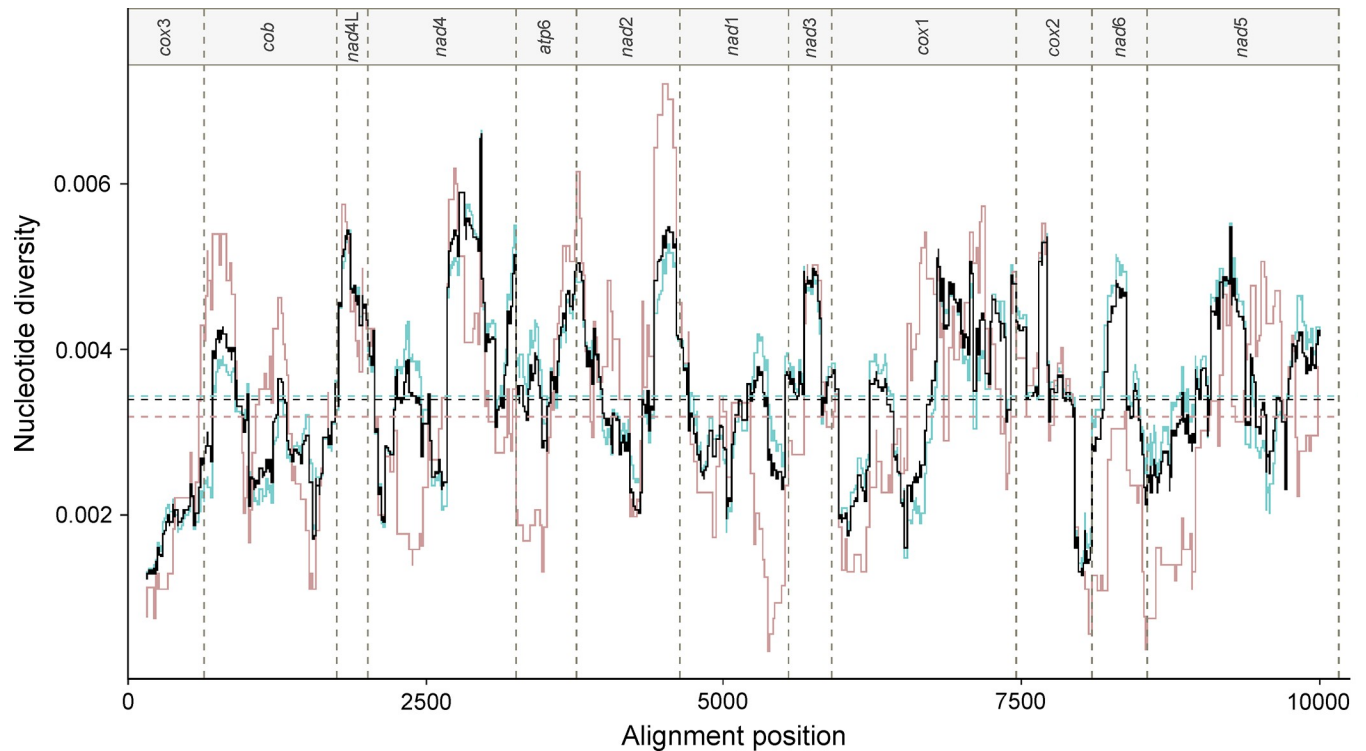
**Fig 2. Results from sliding window analysis of nucleotide diversity.** Nucleotide diversity values calculated based on all 12 concatenated mitochondrial protein gene sequences for: (i) all 183 *Clonorchis sinensis* specimens studied (black line); (ii) 149 specimens from the 'south' (Chinese provinces Guangdong, Guangxi and Hunan; blue line); and (iii) 34 specimens from the 'north' (Primorsky Krai region in the Russian Far East, and Chinese provinces Heilongjiang and Jilin; pink line).

populations from distinct geographical locations and different species of hosts. These results should have implications for future explorations of the biology, epidemiology and ecology of this parasite.

Nucleotide polymorphism (microheterogeneity) within some worms likely represents heteroplasmy as a result of multiple, distinct mitochondrial genomes being present within an individual. This interpretation is supported by the high sequencing depth achieved (the median value being 7268). However, it is possible that other factors, such as nuclear DNA of mitochondrial origin (called nuclear mitochondrial sequences, NUMTs; [55]) or sequencing error might cause 'background' issues. NUMTs can occur as a result of the transposition of mitochondrial sequence tracts to the nuclear genome; although their prevalence has not been investigated in *C. sinensis*, such DNA elements have been identified or characterised in a range of organisms, such as fungi, plants and metazoans including flatworms [55, 56]. In the sequencing procedure employed herein, such sequences or mutants thereof might have been co-amplified from homologous mitochondrial sequences, being reflected as polymorphism at a low frequency [57], given that mitochondrial sequences usually have a higher copy number than NUMTs in the nuclear genome [58]. Sequencing errors may also masquerade as polymorphism, and would occur randomly at low level frequency [59, 60]. To distinguish authentic polymorphism (i.e. heteroplasmy) from such 'background', a detection threshold is commonly applied (cf. [61]). Thresholds in the range of 10% to 20% have been shown to minimise the risk of 'false attribution' in the event that NUMTs or sequencing errors are present [61, 62]. Using detection thresholds of 10% and 20%, genuine polymorphism (heteroplasmy) was inferred here in 19 and 4 individual *C. sinensis* specimens, respectively (Table 2).

Heteroplasmy is relatively common in mitochondrial genomes of eukaryotes (e.g., [63–65]); however, the biological, evolutionary and molecular mechanisms leading to it have not been explored in-depth in helminths. On one hand, heteroplasmy is thought to occur as an intermediate state of a mutation before becoming fixed in the mitochondrial genome. On the other hand, multiple populations of mitochondrial haplotypes within an individual might be positively selected for, given that complex tissue-specific differences in mitochondria and their genomes are known to occur [66, 67]. Such genetic variation within individuals could be acquired *de novo*, inherited maternally or, in rare cases, inherited paternally [68–70]. Although mitochondrial genomes are believed to be maternally inherited, paternal leakage of mitochondria during fertilisation has been reported in metazoans [68, 69, 71, 72] and may occur in *C. sinensis*. Assuming that out-crossing occurs in *C. sinensis*, or that mitochondrial genomes in spermatozoa and oocytes of hermaphroditic individuals display some level of genetic variation, a small number of mitochondria of some sperm might populate the oocyte-cytoplasm, thus introducing variation into an individual. The transfer of sperm mitochondria into the cytoplasm of the oocyte is known to occur in some trematodes such as *Gonapodasmius* and *Paragonimus* [73, 74], and might, to some extent, be removed by autophagy triggered by fertilisation, as has been described for the hermaphroditic (free-living) nematode *Caenorhabditis elegans* [69, 75], which might, in turn, result in a 'recognition failure' and a paternal leakage of mitochondrial DNA [68]. Although challenging to address in a hermaphroditic parasite with an indirect life cycle, these aspects warrant investigation in *C. sinensis* to better understand its reproductive biology (fertilisation process) and the inheritance of mitochondrial DNA, which could have broader implications for other trematodes. It is possible that heteroplasmy within *C. sinensis* is more pronounced than currently recognised and that it may display intricate patterns within individuals (possibly allele- or tissue-specificity; [66, 67]). Future studies might utilise deep-sequencing and a double-strain validation approach [76] to elucidate levels of heteroplasmy within *C. sinensis*, and single-cell or single-mitochondrion sequencing [77] to elucidate intra-mitochondrial and tissue-specific patterns of mitochondrial genetic variation in this parasite.

In the present study, numerous fixed nucleotide variations were detected between some individuals of *C. sinensis*, which might reflect an 'adaptive capacity' in this species, particularly for factors linked to the establishment of infection. It is generally accepted that the adaptive capacity of a species is reflected in the levels of genetic diversity within populations, as a high level of variation provides a mechanism to respond to different selection pressures in the environment; such variation might be reflected in phenotypic variation. Phenotypic traits of particular interest in parasites, relevant to understanding parasite biology and potentially parasitic diseases, include differences in morphology, immune modulation or evasion, host preference or affiliation, sensitivity to drugs, virulence or pathogenicity, and/or adaptation to geographical regions with particular climatic conditions or host spectra. Investigating the impact that genetic variation within *C. sinensis* populations has on such traits relies on comprehensive genetic data and relevant biological, ecological, epidemiological or clinical information available upon sample collection. Our current data sets enabled us to investigate whether a genetic composition of adult *C. sinensis* populations correlated with geographical and/or known naturally infected definitive host species (cat or dog), as the worm isolates were obtained from disparate northern and southern zones in which *C. sinensis* is endemic [4].

We found no unequivocal mitochondrial genetic distinction between or among the geographically distinct sampling locations in 'northern' and 'southern' endemic regions, or between most other geographical sub-populations (Figs 1 and 2; S3 Table). The population from Hunan was genetically unique with respect to other provinces in China, and some haplogroups appeared to be specific to the 'south' (e.g., C, F and G) or 'north' (I) (Fig 1). However,

we cannot exclude an effect of sampling bias on the findings. The apparent lack of genetic differentiation between/among most geographical regions, particularly between the 'south' and 'north', suggests that the population genetic make-up of *C. sinensis* is similar across the distribution range of this species, which would suggest high levels of ongoing gene-flow between and among populations, probably due to the migration of humans and/or domestic or wild canids or felids, and possibly the transport of intermediate hosts or infected food products. This is concerning in the sense that, if new, evolutionarily advantaged phenotypes emerged (e.g., genetic variants less sensitive to anthelmintics), there would be a marked potential for such variants to be spread across the distribution range of *C. sinensis*; it might also facilitate and enhance parasite transmission or dissemination in endemic or newly established foci of infection. It is also interesting to note that the apparent endemic hot-spots for *C. sinensis* in the 'north' and 'south' might differ in their epidemiology (transmission and prevalence) linked to climatic conditions (temperature, rainfall and humidity), known to have a significant impact on the success of trematode transmission, presence, prevalence and biology of intermediate host species, such as higher temperatures enhancing snail growth, activity and survival [10, 78, 79]. Whether such ecological aspects relate to differing selection pressures on *C. sinensis* in these regions and what impact high human migration rates might have on these possible local adaptations (cf. [80]) remain to be assessed in the future.

The lack of a clear genetic differentiation between *C. sinensis* from cats and dogs appears to be indicative of a plasticity in adaptation to definitive host species. Not only were the samples obtained from the two host species genetically similar, remarkably, several of them were genetically identical in their mitochondrial protein-encoding sequences (Fig 1), which suggests that the same *C. sinensis* variant can infect both cat and dog, and that there is likely no specific affiliation to either host species. However, we have no precise information on how transmission takes place in nature, i.e. whether genetically similar or identical worms in cats and dogs transmit through the same intermediate host species. Thus, based on current knowledge, we do not know whether genetic plasticity relating to cat and dog hosts extends to the infection of other hosts, or whether genetic variants display particular levels of host specificity. The genetic determinants of transmission and host-specificity in *C. sinensis* could potentially relate to a complex network of associations, as this parasite is known to infect several other 'natural' definitive hosts (e.g., humans, badgers, weasels and mink), in addition to cats and dogs [81], and a wide range of snail and fish intermediate host species [7, 8], whereas differences in host genotypes could also play a role. Although challenging, it would be interesting to investigate genetic variants of *C. sinensis* in humans to establish whether there is a genetic differentiation between isolates from people and other host species, and whether particular genotypes/haplotypes have an association with the pathogenesis, the severity of disease and complications, such as cancer, in humans. Such an endeavour would be reliant on obtaining worms during surgery, which might be feasible in endemic regions, such as southern China (e.g., Guangdong), and the Russian Far East. Alternatively, eggs from human faecal samples could be collected during epidemiological studies and subjected to genetic analysis–preferably at an individual egg level. Given the knowledge gaps regarding the genetic composition of *C. sinensis* in human populations, finding unique variants of *C. sinensis* could have significant implications for transmission and control (cf. [82]). It seems reasonable to assume that the levels of genetic diversity would correlate negatively with transmission intensity, reflecting the success of control efforts. It would also be interesting to see whether repeated praziquantel treatment of people with *C. sinensis* infection induces drug-resistant phenotypes of the parasite with reference to untreated populations.

Clearly, the finding of marked genetic variability within *C. sinensis* in the present study and the establishment of an integrated deep-sequencing-informatic-phylogenetic approach

provide a solid foundation to tackle these research questions or areas. Clear proof-of-principle of this technological approach also opens the door to large-scale and comparative population genetic and molecular epidemiological analyses using whole mitochondrial and nuclear genomic data sets.

## Supporting information

**S1 Table. Information on the *Clonorchis sinensis* specimens (n = 183) used in this study, including their host and geographical origins as well as GenBank database accession nos. for the 12 mitochondrial protein gene sequences defined for individual specimens.** (DOCX)

**S2 Table. Pairwise fixation index ($F_{ST}$) values between haplogroups A to L inferred based on phylogenetic analyses.** (DOCX)

**S3 Table. Pairwise fixation ($F_{ST}$) indices for populations of *Clonorchis sinensis* from Russia (Primorsky Krai region), and China (Heilongjiang, Jilin, Guangdong, Guangxi and Hunan provinces).** (DOCX)

**S1 Fig. Phylogenetic tree.** This tree was constructed using concatentated sequences for the 12 mitochondrial protein genes representing 183 *Clonorchis sinensis* individuals studied here and four published mitochondrial genomes (GenBank accession nos. JF729303, FJ381664, JF729304 and KY564177) employing Bayesian (BI) and maximum likehood (ML) methods. Specimen numbers are indicated (see S1 Table); 's' stands for sample. Haplotype S242 represents a *C. sinensis* specimen from a hamster experimentally infected with metacercariae obtained from cyprinid fish from Russia (FJ381664; [25]), haplotype S243 is from a cat from China (accession no. JF729303; [26]), haplotype S244 is from a cat from Korea (JF729304; [26]), and haplotype S245 is from a hamster experimentally infected with metacercariae obtained from cyprinid fish from South Korea (KY564177; [27]). Clusters A to L correspond to haplogroups in Fig 1; a grey outline indicates a haplogroup supported by nucleotide sequence data; a beige outline indicates a haplogroup supported by both nucleotide and amino acid sequence data. Posterior probability values (pp) are indicated by black, grey, solid or dashed lines; bootstrap support (bs) values are indicated on nodes (coloured circles). Outgroup not included in the figure. (TIF)

## Acknowledgments

## Author Contributions

**Conceptualization:** Liina Kinkar, Pasi K. Korhonen, Neil D. Young, Robin B. Gasser.

**Formal analysis:** Liina Kinkar, Pasi K. Korhonen, Neil D. Young, Robin B. Gasser.

# References

1. Vandeven N, Nghiem P. Pathogen-driven cancers and emerging immune therapeutic strategies. Cancer Immunol Res. 2014; 2(1):9–14. https://doi.org/10.1158/2326-6066.CIR-13-0179 PMID: 24778160; PubMed Central PMCID: PMC4135058.

2. Kim T-S, Pak JH, Kim J-B, Bahk YY. *Clonorchis sinensis*, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: a brief review. BMB Rep. 2016; 49(11):590–7. https://doi.org/10.5483/bmbrep.2016.49.11.109 PMID: 27418285; PubMed Central PMCID: PMC5346318.

3. Lim MK, Ju Y-H, Franceschi S, Oh J-K, Kong H-J, Hwang S-S, et al. *Clonorchis sinensis* infection and increasing risk of cholangiocarcinoma in the Republic of Korea. Am J Trop Med Hyg. 2006; 75(1):93–6. PMID: 16837714.

4. Qian M-B, Chen Y-D, Liang S, Yang G-J, Zhou X-N. The global epidemiology of clonorchiasis and its relation with cholangiocarcinoma. Infect Dis Poverty. 2012; 1:4. https://doi.org/10.1186/2049-9957-1-4 PMID: 23849183; PubMed Central PMCID: PMC3710150.

5. Lai Y-S, Zhou X-N, Pan Z-H, Utzinger J, Vounatsou P. Risk mapping of clonorchiasis in the People's Republic of China: A systematic review and Bayesian geostatistical analysis. PLoS Negl Trop Dis. 2017; 11(3): e0005239. https://doi.org/10.1371/journal.pntd.0005239 PMID: 28253272; PubMed Central PMCID: PMC5416880.

6. Qian M-B, Utzinger J, Keiser J, Zhou X-N. Clonorchiasis. Lancet. 2016; 387(10020):800–10. https://doi.org/10.1016/S0140-6736(15)60313-0 PMID: 26299184.

7. Wang D, Young ND, Korhonen PK, Gasser RB. *Clonorchis sinensis* and clonorchiasis: the relevance of exploring genetic variation. Adv Parasitol. 2018; 100:155–208. https://doi.org/10.1016/bs.apar.2018.03.006 PMID: 29753338.

8. Lun Z-R, Gasser RB, Lai D-H, Li A-X, Zhu X-Q, Yu X-B, et al. Clonorchiasis: a key foodborne zoonosis in China. Lancet Infect Dis. 2005; 5(1):31–41. https://doi.org/10.1016/S1473-3099(04)01252-6 PMID: 15620559.

9. Deng Z-H, Fang Y-Y, Zhang Q-M, Mao Q, Pei F-Q, Liu M-R. The control of clonorchiasis in Guangdong province, southern China. Acta Trop. 2020; 202:105246. https://doi.org/10.1016/j.actatropica.2019.105246 PMID: 31672488.

10. Li T, Yang Z, Wang M. Correlation between clonorchiasis incidences and climatic factors in Guangzhou, China. Parasit Vectors. 2014; 7:29. https://doi.org/10.1186/1756-3305-7-29 PMID: 24428903; PubMed Central PMCID: PMC3896806.

11. Tang Z-L, Huang Y, Yu X-B. Current status and perspectives of *Clonorchis sinensis* and clonorchiasis: epidemiology, pathogenesis, omics, prevention and control. Infect Dis Poverty. 2016; 5(1):71. https://doi.org/10.1186/s40249-016-0166-1 PMID: 27384714; PubMed Central PMCID: PMC4933995.

12. Qian M-B. Clonorchiasis control: starting from awareness. Infect Dis Poverty. 2014; 3:33. https://doi.org/10.1186/2049-9957-3-33 PMID: 25243069; PubMed Central PMCID: PMC4168994.

13. Park GM, Yong TS. Geographical variation of the liver fluke, *Clonorchis sinensis*, from Korea and China based on the karyotypes, zymodeme and DNA sequences. Southeast Asian J Trop Med Public Health. 2001; 32(Suppl 2):12–6. PMID: 12041573.

14. Lee S-U, Huh S. Variation of nuclear and mitochondrial DNAs in Korean and Chinese isolates of *Clonorchis sinensis*. Korean J Parasitol. 2004; 42(3):145–8. https://doi.org/10.3347/kjp.2004.42.3.145 PMID: 15381864; PubMed Central PMCID: PMC2717366.

15. Liu GH, Li B, Li JY, Song HQ, Lin RQ, Cai XQ, et al. Genetic variation among *Clonorchis sinensis* isolates from different geographic regions in China revealed by sequence analyses of four mitochondrial

genes. J Helminthol. 2012; 86(4):479–84. https://doi.org/10.1017/S0022149X11000757 PMID: 22152452.

16. Tatonova YV, Chelomina GN, Besprosvannykh VV. Genetic diversity of nuclear ITS1-5.8S-ITS2 rDNA sequence in *Clonorchis sinensis* Cobbold, 1875 (Trematoda: Opisthorchidae) from the Russian Far East. Parasitol Int. 2012; 61(4):664–74. https://doi.org/10.1016/j.parint.2012.07.005 PMID: 22824337.

17. Tatonova YV, Chelomina GN, Besprozvannykh VV. Genetic diversity of *Clonorchis* sinensis (Trematoda: Opisthorchiidae) in the Russian southern Far East based on mtDNA cox1 sequence variation. Folia Parasitol. 2013; 60(2):155–62. https://doi.org/10.14411/fp.2013.017 PMID: 23724735.

18. Tatonova YV, Chelomina GN, Nguyen HM. Inter-individual and intragenomic variations in the ITS region of *Clonorchis sinensis* (Trematoda: Opisthorchiidae) from Russia and Vietnam. Infect Genet Evol. 2017; 55:350–7. https://doi.org/10.1016/j.meegid.2017.10.008 PMID: 28993292.

19. Sun J, Huang Y, Huang H, Liang P, Wang X, Mao Q, et al. Low divergence of *Clonorchis sinensis* in China based on multilocus analysis. PLoS One. 2013; 8(6):e67006. https://doi.org/10.1371/journal.pone.0067006 PMID: 23825605; PubMed Central PMCID: PMC3688995.

20. Xiao J-Y, Gao J-F, Cai L-S, Dai Y, Yang C-J, Luo L, et al. Genetic variation among *Clonorchis sinensis* isolates from different hosts and geographical locations revealed by sequence analysis of mitochondrial and ribosomal DNA regions. Mitochondr DNA. 2013; 24(5):559–64. https://doi.org/10.3109/19401736.2013.770490 PMID: 23463938.

21. Chelomina GN, Tatonova YV, Hung NM, Ngo HD. Genetic diversity of the Chinese liver fluke *Clonorchis sinensis* from Russia and Vietnam. Int J Parasitol. 2014; 44(11):795–810. https://doi.org/10.1016/j.ijpara.2014.06.009 PMID: 25123068.

22. Zhang X, Sun B, Tang Q, Chen R, Han S. Molecular identification and phylogenetic analysis of nuclear rDNA sequences of *Clonorchis sinensis* isolates from human fecal samples in Heilongjiang province, China. Front Microbiol. 2019; 10:26. https://doi.org/10.3389/fmicb.2019.00026 PMID: 30745896; PubMed Central PMCID: PMC6360181.

23. Wang X, Chen W, Huang Y, Sun J, Men J, Liu H, et al. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. Genome Biol. 2011; 12(10):R107. https://doi.org/10.1186/gb-2011-12-10-r107 PMID: 22023798; PubMed Central PMCID: PMC3333777.

24. Wang D, Korhonen PK, Gasser RB, Young ND. Improved genomic resources and new bioinformatic workflow for the carcinogenic parasite *Clonorchis sinensis*: Biotechnological implications. Biotechnol Adv. 2018; 36(4):894–904. https://doi.org/10.1016/j.biotechadv.2018.02.008 PMID: 29454982.

25. Shekhovtsov SV, Katokhin AV, Kolchanov NA, Mordvinov VA. The complete mitochondrial genomes of the liver flukes *Opisthorchis felineus* and *Clonorchis sinensis* (Trematoda). Parasitol Int. 2010; 59 (1):100–3. https://doi.org/10.1016/j.parint.2009.10.012 PMID: 19906359.

26. Cai XQ, Liu GH, Song HQ, Wu CY, Zou FC, Yan HK, et al. Sequences and gene organization of the mitochondrial genomes of the liver flukes Opisthorchis viverrini and *Clonorchis sinensis* (Trematoda). Parasitol Res. 2012; 110(1):235–43. https://doi.org/10.1007/s00436-011-2477-2 PMID: 21626421.

27. Wang D, Young ND, Koehler AV, Tan P, Sohn W-M, Korhonen PK, et al. Mitochondrial genomic comparison of *Clonorchis sinensis* from South Korea with other isolates of this species. Infect Genet Evol. 2017; 51:160–6. https://doi.org/10.1016/j.meegid.2017.02.015 PMID: 28254628.

28. Le TH, Blair D, McManus DP. Mitochondrial genomes of human helminths and their use as markers in population genetics and phylogeny. Acta Trop. 2000; 77(3):243–56. https://doi.org/10.1016/s0001-706x(00)00157-1 PMID: 11114386.

29. Le TH, Blair D, McManus DP. Mitochondrial genomes of parasitic flatworms. Trends Parasitol. 2002; 18(5):206–13. https://doi.org/10.1016/s1471-4922(02)02252-3 PMID: 11983601.

30. Lin R-Q, Tang J-D, Zhou D-H, Song H-Q, Huang S-Y, Chen J-X, et al. Prevalence of *Clonorchis sinensis* infection in dogs and cats in subtropical southern China. Parasit Vectors. 2011; 4:180. https://doi.org/10.1186/1756-3305-4-180 PMID: 21929783; PubMed Central PMCID: PMC3183008.

31. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, et al. A reference human genome dataset of the BGI-SEQ-500 sequencer. Gigascience. 2017; 6(5):1–9. https://doi.org/10.1093/gigascience/gix024 PMID: 28379488; PubMed Central PMCID: PMC5467036.

32. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010; 38(6):1767–71. https://doi.org/10.1093/nar/gkp1137 PMID: 20015970; PubMed Central PMCID: PMC2847217.

33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404; PubMed Central PMCID: PMC4103590.

34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168; PubMed Central PMCID: PMC2705234.

35. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199; PubMed Central PMCID: PMC2928508.

36. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 1; 27(21):2987–93. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627; PubMed Central PMCID: PMC3198575.

37. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12):1647–9. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367; PubMed Central PMCID: PMC3371832.

38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943; PubMed Central PMCID: PMC2723002.

39. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22(3):568–76. https://doi.org/10.1101/gr.129684.111 PMID: 22300766; PubMed Central PMCID: PMC3290792.

40. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14(2):178–92. https://doi.org/10.1093/bib/bbs017 PMID: 22517427; PubMed Central PMCID: PMC3603213.

41. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953; 21(6):1087–92. https://doi.org/10.1063/1.1699114.

42. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970; 57(1):97–109. https://doi.org/10.1093/biomet/57.1.97.

43. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572–4. https://doi.org/10.1093/bioinformatics/btg180 PMID: 12912839

44. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci. 1986; 17:57–86.

45. Gu X, Fu YX, Li WH. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol. 1995; 12(4):546–57. https://doi.org/10.1093/oxfordjournals.molbev.a040235 PMID: 7659011.

46. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 2017; 34(3):772–3. https://doi.org/10.1093/molbev/msw260 PMID: 28013191

47. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005; 21(4):456–63. https://doi.org/10.1093/bioinformatics/bti191 PMID: 15608047.

48. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16(1):37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250.

49. Leigh JW, Bryant D. PopART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015; 6(9):1110–6. https://doi.org/10.1111/2041-210X.12410.

50. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25(11):1451–2. https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325.

51. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2007; 1:47–50. PMID: 19325852; PubMed Central PMCID: PMC2658868.

52. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014; 31(7):1929–36. https://doi.org/10.1093/molbev/msu136 PMID: 24739305; PubMed Central PMCID: PMC4069620.

53. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2018, Vienna, Austria. Available from: https://www.R-project.org/

54. Wickham H. ggplot2: elegant graphics for data analysis. 2016, Springer-Verlag, New York. https://doi.org/10.1007/978-3-319-24277-4.

55. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 2010; 6(2):e1000834. https://doi.org/10.1371/journal.pgen.1000834 PMID: 20168995; PubMed Central PMCID: PMC2820518.

**56.** Brabec J, Scholz T, Králová-Hromadová I, Bazsalovicsová E, Olson PD. Substitution saturation and nuclear paralogs of commonly employed phylogenetic markers in the Caryophyllidea, an unusual group of non-segmented tapeworms (Platyhelminthes). Int J Parasitol. 2012; 42(3):259–67. https://doi.org/10.1016/j.ijpara.2012.01.005 PMID: 22366732.

**57.** Santibanez-Koref M, Griffin H, Turnbull DM, Chinnery PF, Herbert M, Hudson G. Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution. Mitochondrion. 2019; 46:302–6. https://doi.org/10.1016/j.mito.2018.08.003 PMID: 30098421; PubMed Central PMCID: PMC6509278.

**58.** Chinnery PF, Hudson G. Mitochondrial genetics. Br Med Bull. 2013 Jun; 106(1):135–59. https://doi.org/10.1093/bmb/ldt017 PMID: 23704099; PubMed Central PMCID: PMC3675899.

**59.** Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of next generation sequencing platforms. Next Gener Seq Appl. 2014; 1:1000106. https://doi.org/10.4172/jngsa.1000106 PMID: 25699289; PubMed Central PMCID: PMC4331009.

**60.** Mak SST, Gopalakrishnan S, Carøe C, Geng C, Liu S, Sinding M-HS, et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. Gigascience. 2017; 6(8):1–13. https://doi.org/10.1093/gigascience/gix049 PMID: 28854615; PubMed Central PMCID: PMC5570000.

**61.** Just RS, Irwin JA, Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. Forensic Sci Int Genet. 2015; 18:131–9. https://doi.org/10.1016/j.fsigen.2015.05.003 PMID: 26009256; PubMed Central PMCID: PMC4550493.

**62.** Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet. 2010; 87(2):237–49. https://doi.org/10.1016/j.ajhg.2010.07.014 PMID: 20696290; PubMed Central PMCID: PMC2917713.

**63.** Kmiec B, Woloszynska M, Janska H. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Curr Genet. 2006; 50(3):149–59. https://doi.org/10.1007/s00294-006-0082-1 PMID: 16763846.

**64.** Nunes MDS, Dolezal M, Schlötterer C. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. Mol Ecol. 2013; 22(8):2106–17. https://doi.org/10.1111/mec.12256 PMID: 23452233; PubMed Central PMCID: PMC3659417.

**65.** Robison GA, Balvin O, Schal C, Vargo EL, Booth W. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). J Med Entomol. 2015; 52(4):734–8. https://doi.org/10.1093/jme/tjv055 PMID: 26335484; PubMed Central PMCID: PMC4592348.

**66.** Samuels DC, Li C, Li B, Song Z, Torstenson E, Boyd Clay H, et al. Recurrent tissue-specific mtDNA mutations are common in humans. PLoS Genet. 2013; 9(11):e1003929. https://doi.org/10.1371/journal.pgen.1003929 PMID: 24244193; PubMed Central PMCID: PMC3820769.

**67.** Li M, Schröder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. PNAS. 2015; 112(8):2491–6. https://doi.org/10.1073/pnas.1419651112 PMID: 25675502; PubMed Central PMCID: PMC4345623.

**68.** Rokas A, Ladoukakis E, Zouros E. Animal mitochondrial DNA recombination revisited. Trends Ecol Evol. 2003; 18(8):411–7. https://doi.org/10.1016/S0169-5347(03)00125-3.

**69.** Sato K, Sato M. Multiple ways to prevent transmission of paternal mitochondrial DNA for maternal inheritance in animals. J Biochem. 2017; 162(4):247–53. https://doi.org/10.1093/jb/mvx052 PMID: 28981751.

**70.** Burr SP, Pezet M, Chinnery PF. Mitochondrial DNA heteroplasmy and purifying selection in the mammalian female germ line. Dev Growth Differ. 2018; 60(1):21–32. https://doi.org/10.1111/dgd.12420 PMID: 29363102.

**71.** Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahata N, Chigusa SI. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. Genetics. 1990; 126(3):657–63. PMID: 2249764; PubMed Central PMCID: PMC1204221.

**72.** Gyllensten U, Wharton D, Josefsson A, Wilson AC. Paternal inheritance of mitochondrial DNA in mice. Nature. 1991; 352(6332):255–7. https://doi.org/10.1038/352255a0 PMID: 1857422.

**73.** Justine J-L, Mattei X. Ultrastructural observations on the spermatozoon, oocyte and fertilization process in *Gonapodasmius*, a gonochoristic trematode (Trematoda: Digenea: Didymozoidae). Acta Zoologica. 1984; 65(3):171–7. https://doi.org/10.1111/j.1463-6395.1984.tb00822.x.

**74.** Orido Y. Fertilization and oogenesis of the lung fluke, *Paragonimus ohirai* (Trematoda:Troglotrematidae). Int J Parasitol. 1988; 18(7):973–80. https://doi.org/10.1016/0020-7519(88)90180-4 PMID: 3225118.

**75.** Sato M, Sato K. Degradation of paternal mitochondria by fertilization-triggered autophagy in *C. elegans* embryos. Science. 2011; 334(6059):1141–4. https://doi.org/10.1126/science.1210333 PMID: 21998252.

**76.** Duan M, Tu J, Lu Z. Recent advances in detecting mitochondrial DNA heteroplasmic variations. Molecules. 2018; 23(2):323. https://doi.org/10.3390/molecules23020323 PMID: 29401641; PubMed Central PMCID: PMC6017848.

**77.** Morris J, Na Y-J, Zhu H, Lee J-H, Giang H, Ulyanova AV, et al. Pervasive within-mitochondrion single-nucleotide variant heteroplasmy as revealed by single-mitochondrion sequencing. Cell Rep. 2017; 21 (10):2706–13. https://doi.org/10.1016/j.celrep.2017.11.031 PMID: 29212019; PubMed Central PMCID: PMC5771502.

**78.** Paull SH, Johnson PTJ. High temperature enhances host pathology in a snail-trematode system: possible consequences of climate change for the emergence of disease. Freshw Biol. 2011; 56(4):767–78. https://doi.org/10.1111/j.1365-2427.2010.02547.x.

**79.** Prasopdee S, Kulsantiwong J, Piratae S, Khampoosa P, Thammasiri C, Suwannatrai A, et al. Temperature dependence of *Opisthorchis viverrini* infection in first intermediate host snail, *Bithynia siamensis goniomphalos*. Acta Trop. 2015; 141(Pt A):112–7. https://doi.org/10.1016/j.actatropica.2013.10.011 PMID: 24161535.

**80.** Morgan AD, Gandon S, Buckling A. The effect of migration on local adaptation in a coevolving host-parasite system. Nature. 2005; 437(7056):253–6. https://doi.org/10.1038/nature03913 PMID: 16148933.

**81.** Choi BI, Han JK, Hong ST, Lee KH. Clonorchiasis and cholangiocarcinoma: etiologic relationship and imaging diagnosis. Clin Microbiol Rev. 2004; 17(3):540–52. https://doi.org/10.1128/CMR.17.3.540-552.2004 PMID: 15258092; PubMed Central PMCID: PMC452546.

**82.** Webster JP, Gower CM, Norton AJ. Evolutionary concepts in predicting and evaluating the impact of mass chemotherapy schistosomiasis control programmes on parasites and their hosts. Evol Appl. 2008; 1(1):66–83. https://doi.org/10.1111/j.1752-4571.2007.00012.x PMID: 25567492; PubMed Central PMCID: PMC3352399.