

RESEARCH ARTICLE

First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach

Liina Kinkar¹, Neil D. Young^{1*}, Woon-Mok Sohn², Andreas J. Stroehlein¹, Pasi K. Korhonen¹, 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** Department of Parasitology and Tropical Medicine, and Institute of Health Sciences, Gyeongsang National University College of Medicine, Jinju, Korea

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



OPEN ACCESS

Citation: Kinkar L, Young ND, Sohn W-M, Stroehlein AJ, Korhonen PK, Gasser RB (2020) First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLoS Negl Trop Dis* 14(8): e0008552. <https://doi.org/10.1371/journal.pntd.0008552>

Editor: Stephen W. Attwood, University of Oxford, UNITED KINGDOM

Received: May 18, 2020

Accepted: July 1, 2020

Published: August 26, 2020

Copyright: © 2020 Kinkar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The complete mitochondrial genome sequence has been deposited in the GenBank database under the accession no. MT607652; raw data are available in the Sequence Read Archive (SRA) under the accession no. PRJNA386618.

Funding: Project support was through Australian Research Council (ARC) grants LP180101334 (N. D.Y. and P.K.K.) and LP180101085 (R.B.G.) and Yougene Health Singapore. N.D.Y. and P.K.K.

Abstract

Background

Mitochondrial genomes provide useful genetic markers for systematic and population genetic studies of parasitic helminths. Although many such genome sequences have been published and deposited in public databases, there is evidence that some of them are incomplete relating to an inability of conventional techniques to reliably sequence non-coding (repetitive) regions. In the present study, we characterise the complete mitochondrial genome—including the long, non-coding region—of the carcinogenic Chinese liver fluke, *Clonorchis sinensis*, using long-read sequencing.

Methods

The mitochondrial genome was sequenced from total high molecular-weight genomic DNA isolated from a pool of 100 adult worms of *C. sinensis* using the MinION sequencing platform (Oxford Nanopore Technologies), and assembled and annotated using an informatic approach.

Results

From > 93,500 long-reads, we assembled a 18,304 bp-mitochondrial genome for *C. sinensis*. Within this genome we identified a novel non-coding region of 4,549 bp containing six tandem-repetitive units of 719–809 bp each. Given that genomic DNA from pooled worms was used for sequencing, some variability in length/sequence in this tandem-repetitive region was detectable, reflecting population variation.

Conclusions

For *C. sinensis*, we report the complete mitochondrial genome, which includes a long (> 4.5 kb) tandem-repetitive region. The discovery of this non-coding region using a nanopore-

were recipients of Career Development and Early Career Research Fellowships, respectively, from the National Health and Medical Research Council (NHMRC) of Australia. The LIEF HPC-GPGPU Facility is supported by ARC LIEF Grant LE170100200. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

sequencing/informatic approach now paves the way to investigating the nature and extent of length/sequence variation in this region within and among individual worms, both within and among *C. sinensis* populations, and to exploring whether this region has a functional role in the regulation of replication and transcription, akin to the mitochondrial control region in mammals. Although applied to *C. sinensis*, the technological approach established here should be broadly applicable to characterise complex tandem-repetitive or homo-polymeric regions in the mitochondrial genomes of a wide range of taxa.

Author summary

In the present study, we characterised the complete mitochondrial genome of *Clonorchis sinensis*—a carcinogenic liver fluke. To do this, we sequenced from total genomic DNA from multiple adult worms using a new method (Oxford Nanopore technology) to obtain data for long stretches of DNA, and then assembled these data to construct a mitochondrial genome of 18,304 bp, containing a > 4.5 kb-long tandem-repetitive region—not previously detected in this species. The results demonstrate that this method is effective at sequencing long and complex non-coding elements—not achievable using conventional techniques. The discovery of this long tandem-repetitive region in *C. sinensis* provides an opportunity to now explore its origin(s) and length/sequence diversity in populations of this species, and also to characterise its function(s). The technological approach employed here should have broad applicability to characterise previously-elusive non-coding mitochondrial genomic regions in a wide range of taxa.

Introduction

Substantial progress in nuclear and mitochondrial genomics has been made over the last two decades through the use of DNA sequencing methods [1]. This progress is starting to have a major positive impact in many areas of parasitology, both fundamental and applied. For instance, exploring the mitochondrial genomes has enabled systematic (taxonomic and phylogenetic) and population genetic investigations of helminths (flatworms and roundworms) [2–6]. Such genomes provide a rich source of markers for such investigations and are particularly applicable to systematic investigations of species of flatworms (platyhelminths) [7], because the mitochondrial genes are usually considerably less variable in sequence than for many roundworm (nematode) species [8–11]. Thus, there have been numerous studies of members of the classes Trematoda and Cestoda [7, 12–15].

Seminal work on mitochondrial genomes was conducted using PCR-based cloning combined with conventional (Sanger) sequencing (e.g., [7, 16]). Subsequently, high throughput sequencing (e.g., 454 and Illumina) became the approach of choice, allowing sequencing from small amounts of genomic DNA at reduced cost and time [1]. With the advent of ‘short-read’ sequencing (e.g., Illumina) came the confidence that sequencing at high coverage in a high throughput manner would readily allow the sequencing and assembly of complete mitochondrial genomes, because of their relatively small size (~ 14 kb ± 1 kb in flatworms; [14]). However, there have been challenges with sequencing through tandem-repetitive elements and regions with a biased nucleotide composition using Sanger and short-read technologies [16–18], and little attention has been paid to the impact of these issues.

Indeed, recently, when we explored mitochondrial genomes of parasitic flatworms of the genus *Echinococcus*, we noticed a gap of > 1 kb between the 3'-end of the *nad5* gene and the 5'-end of the *cox3* gene in *E. granulosus* genotype G1 [19]. Despite our efforts using a PCR-based sequencing strategy, we were not able to sequence this gap. However, employing a single molecule, real-time (SMRT) sequencing technology, we obtained long sequence reads that bridged the entire gap, allowing us to characterise a 4,417 bp-long tandem-repetitive region consisting of ten near-identical repeat units (441–445 bp), each harbouring a 184 bp non-coding region and flanking regions [17]. Although three mitochondrial genomes for *E. granulosus* genotype G1 had been published and/or deposited in public gene databases (including GenBank), closing this gap allowed us to define (what we considered to be) the first complete mt genome (17,675 bp) for this genotype, being > 4 kb larger than any previously reported genome for this taxon.

This work stimulated us to scrutinise published mitochondrial genomic data sets of other flatworms, including the carcinogenic liver flukes *Clonorchis sinensis* (Chinese liver fluke), *Opisthorchis viverrini* (Southeast Asian liver fluke) and *Opisthorchis felinus* (cat liver fluke) [20–22]. There were indications of sequence complexity in mitochondrial non-coding regions and the potential for gaps in the published genomes. In the present study, our goal was to critically investigate the completeness of the mitochondrial genome of *C. sinensis* using Oxford Nanopore long-read sequencing technology (<https://nanoporetech.com>). We show the effectiveness of this technology to rapidly sequence the complete mitochondrial genome, irrespective of its length, nature or the structure of intergenic spacer region(s), and to enable the characterisation of large tandem-repeat regions within the mitochondrial genome of *C. sinensis*.

Methods

Parasite material

Adult worms of *C. sinensis* ($n = 100$) were collected in 2009 from Syrian golden hamsters (*Mesocricetus auratus*) experimentally infected with metacercariae isolated from naturally infected cyprinid fish (*Pseudorasbora parva*) originating from Jinju-si, Gyeongsangnam-do, the Republic of Korea, as described previously [23]. This work was conducted by one of the authors (W.-M.S.), in accordance with protocols approved by the animal ethics committee at Gyeongsang National University.

Isolation of high molecular weight genomic DNA, library construction and sequencing

High quality DNA was isolated from the pool of 100 adults of *C. sinensis* using the Circulomics Tissue Kit (Circulomics, Baltimore, MD, USA). Subsequently, low molecular weight DNA was removed using the 5 kb- or 20 kb-Short Read Eliminator (SRE) kit (Circulomics, Baltimore, MD, USA). High molecular weight *C. sinensis* genomic DNA was used to construct rapid-sequencing (SQK-RAD004; Oxford Nanopore Technologies; 5 kb SRE) and ligation-sequencing genomic DNA libraries (SQK-LSK109; Oxford Nanopore Technologies; 5 and 20 kb SRE), according to the manufacturer's instructions. The SQK-RAD004 (5 kb SRE) and SQK-LSK109 (5 kb SRE) libraries were sequenced using separate flow cells (R9.4.1; Oxford Nanopore Technologies). The flow cell used to sequence the SQK-LSK109 (5 kb SRE) library was washed using a Flow Cell Wash Kit (EXP-WSH003; Oxford Nanopore Technologies) and re-used to sequence the SQK-LSK109 (20 kb SRE) library. All genomic DNA libraries were sequenced (48 h) on the MinION sequencer (Oxford Nanopore Technologies). Following sequencing,

bases were 'called' from raw FAST5 reads using the program Guppy v.3.1.5 (Oxford Nanopore Technologies) and stored in the FASTQ format [24].

Assembly of the mitochondrial genome

The reads were mapped to the reference mitochondrial genome of a Korean isolate of *C. sinensis* (GenBank accession no. KY564177; [22]) using Minimap2 v.2.17-r941 [25]; mapped reads and their alignment positions were stored in the BAM format [26]. The mapped reads were extracted from the BAM file using SAMtools v.1.9 [26] and initially assembled using the program Canu v.2.0 [27]. Repeat sequences in the assembled mitochondrial genome were identified using the program repeat-match in the MUMmer package v.3.23 [28]. A library of identified repeat sequences and published mitochondrial protein genes of *C. sinensis* (GenBank accession no. KY564177; [22]) was used to assess the number of repeat units and completeness of the repeat region using the program RepeatMasker v.4.0.5 (<http://www.repeatmasker.org>). The final representative mitochondrial genome was assembled using reads that spanned the entire repetitive region encoding the commonest tandem-repeat unit frequency (± 1 repeat unit) and the program Canu. The non-repetitive region of the assembled genome was then polished with Pilon v.1.23 [29] using available Illumina short-read data [22]. Finally, all long-read data produced were mapped to the assembled mitochondrial genome using Minimap2, and coverage of the genome was determined using mpileup in the SAMtools package [26].

Annotation of the mitochondrial genome and characterisation of the repeat region

The new assembly was compared with those of published mitochondrial genomes of *C. sinensis* (GenBank accession nos. KY564177, JF729304, JF729303 and FJ381664; [20–22]); subsequently, tRNA, rRNA and protein-encoding gene annotations were transferred to the assembled genome. The open reading frame (ORF) of each protein gene was verified using the program Geneious v.11.1.5 [30], employing the mitochondrial genetic code for echinoderms and flatworms ([7]; <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG9>). Secondary structures were predicted using the Vienna RNA Websuite ([31]; <http://rna.tbi.univie.ac.at>) and drawn using the tool Forna [32]. The complete mitochondrial genome sequence was deposited in the GenBank database under the accession no. MT607652; raw data are also available in the Sequence Read Archive (SRA) under the accession no. PRJNA386618.

Results and discussion

The mitochondrial genome of *C. sinensis* contains a tandem-repetitive region of > 4.5 kb

From a total of 93,729 long-reads (equating to 310 Mb), we *de novo*-assembled a 18,304 bp mitochondrial genome for *C. sinensis* at high coverage (average: 2,381; median: 1,615; Fig 1), including a tandem-repetitive region (Fig 2). The initial assembly indicated variation in the number of repeats spanning this region, which likely related to sequence-length variation among individual worms used for the preparation of genomic DNA. In the first instance, we selected six repeats to represent this region. However, it was somewhat challenging to unequivocally assemble all sequences across this tandem-repeat region and to define its precise length. In order to establish the nature and extent of variation in the number and length of repeat sequences, we mapped all long-read data to the mitochondrial genome containing six tandem-repeats and showed a substantial increase in coverage (mean of 1,530 to 5,018; peak at 7,627)

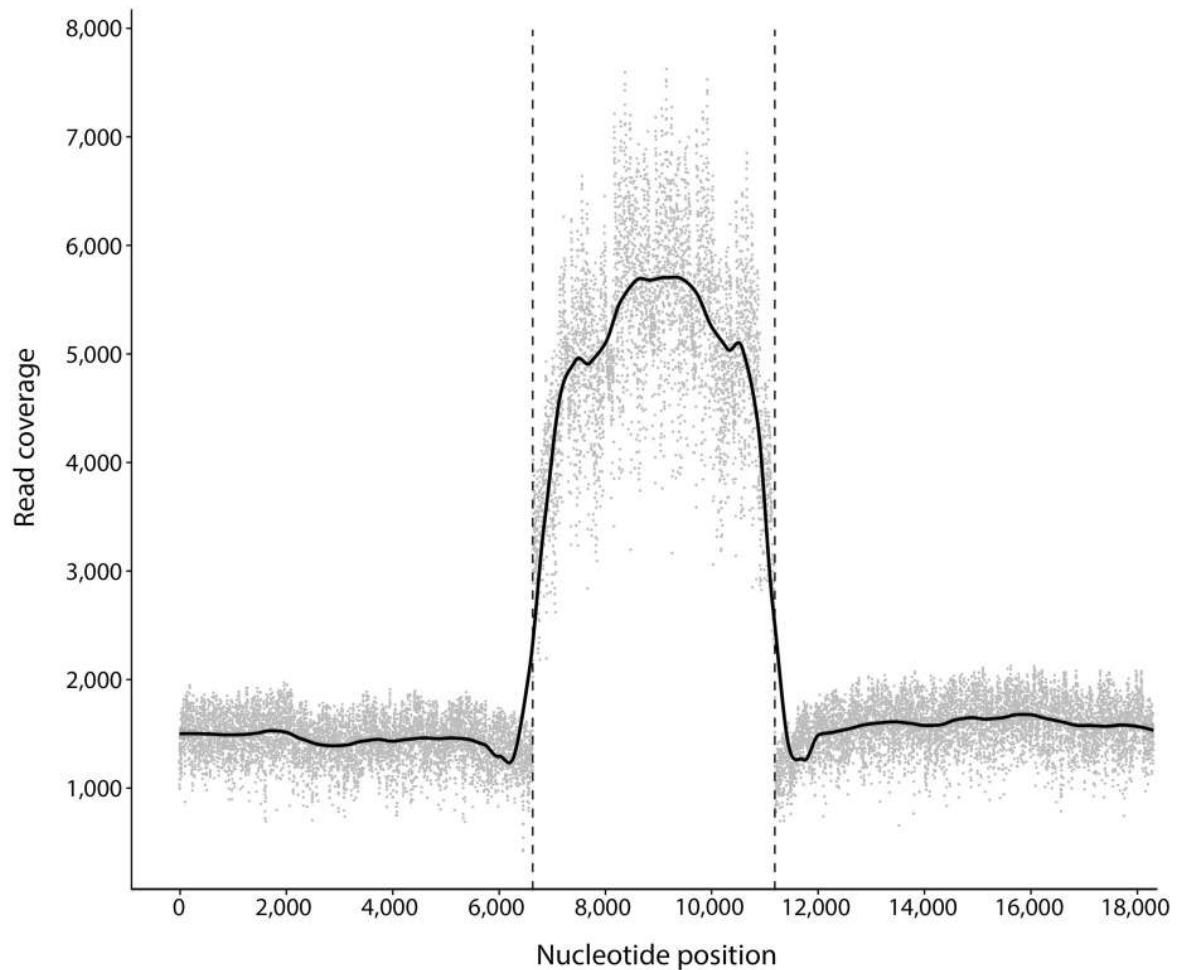


Fig 1. Coverage of long-reads (produced by Oxford Nanopore sequencing) across the mitochondrial genome (18,304 bp) of *Clonorchis sinensis*. The graph shows the depth of nucleotides at each position (grey dots) and the smoothed average of depth across the genome (solid black line). Dashed lines demarcate the start (position 6,640) and the end (11,188) of the long tandem-repeat region of 4,549 bp (Fig 2).

<https://doi.org/10.1371/journal.pntd.0008552.g001>

across this region (positions 6,640 to 11,188; Fig 1). Although mapping results identified reads containing more ($n > 1,200$) or less ($n > 18,900$) than six tandem-repeats, scrutiny of the data revealed 40 sequences (with 3 to 41 repeat units) that bridged the entirety of the tandem-repeat region and were flanked at each terminus by sequences that matched perfectly the expected genes (tRNA-Glu and *nad5* at the 5'-end, and tRNA-Gly and *cox3* at the 3'-end). Irrespective of this variation, reads with six tandem-repeats predominated. Hence, this number of repeats was selected to represent the mitochondrial genome of *C. sinensis* without considering the variation that exists among (or within) individual worms. In this representative mitochondrial genome, repeat units R1 to R6 (Fig 2) were 719–809 bp in length and had 91% identity upon pairwise comparison. Most differences related to length variation in TA- (69 to 138 bp) and GA-rich (26 to 35 bp) sequence tracts, although a 58 bp deletion occurred in a non-repetitive DNA segment (Fig 2). Parts of the repeat units were predicted to fold into secondary structures; some of these predicted structures were complex, with internal loops (≤ 10 bp) and multiple hairpins (stems: ≤ 39 bp; Fig 2).

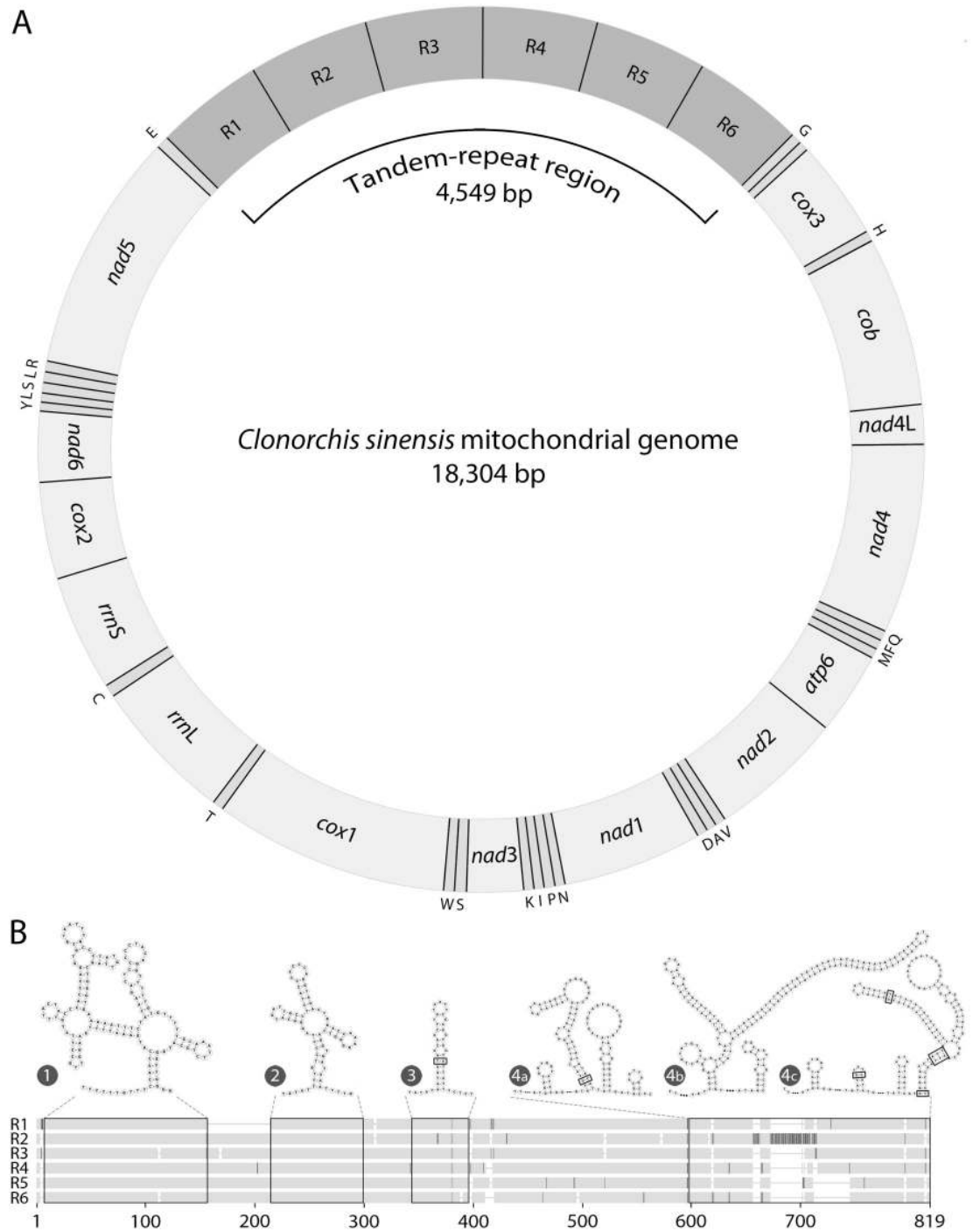


Fig 2. Complete, annotated mitochondrial genome of *Clonorchis sinensis*. (A) Schematic representation of the mitochondrial genome, including the newly-identified tandem-repeat region (4,549 bp); the 12 protein-encoding genes, 2 rRNAs and 22 tRNAs (designated by their one-letter amino acid abbreviations) are in accord with a published reference mitochondrial genome available in GenBank (accession no. KY564177; [22]). The short non-coding region is between tRNA-Gly (G) and *cox3*. (B) A schematic alignment of the tandem-repeat units (R1 to R6; bottom), showing nucleotide identities (light grey) or differences (dark grey) and regions predicted to assume structures (boxed in black). Secondary structures predicted for individual repeat units are indicated above the alignment; mis-matches in stems are indicated (boxed). Positions 598 to 819 include a variable TA-rich region and was predicted to fold into three distinct structures (4a to 4c).

<https://doi.org/10.1371/journal.pntd.0008552.g002>

Variation in the tandem-repetitive region

Evidence of variation in the number of repeats spanning this long non-coding region raised a question about possible technical artefacts. However, because long, intact single-molecule DNA strands were sequenced here using Nanopore technology, such artefacts can be excluded (cf. [33]). Using this technology, we obtained long sequence reads for the entire long tandem-repetitive region, without the need for any read assembly. The use of direct library construction methods excludes artefacts, such as chimeric sequences, resulting from amplification [34–36]. Thus, reads that bridged the entire repeat region and had termini that matched respective flanking regions in the reference mitochondrial genome represented the tandem-repetitive region in *C. sinensis*.

Given that sequence/length variation in mitochondrial non-coding (e.g., control or intergenic) regions is commonly recorded among individuals of an animal species [37], we expected to find such variation in the tandem-repetitive region of *C. sinensis*, because we used a pool of *C. sinensis* adults to prepare genomic DNA for sequencing. Indeed, the mapping results revealed marked variation in sequence, length and repeat numbers as well as sequence coverage. This variation could be among individual worms, because DNA was isolated from 100 worms, but intraindividual or tissue-specific variability (i.e. heteroplasmy) cannot be excluded. Length variation in mitochondrial repeat regions, established using PacBio long-read sequence data, have been reported recently in other trematodes, such as *Paragonimus westermani* and *Schistosoma bovis* [18, 38], but the frequencies and patterns of occurrence within worm populations are unexplored. We believe that further sequencing is warranted to obtain complete (long) read data from individual worms of *C. sinensis* (preferably from disparate geographical areas) to gain an appreciation of the diversity in number and sequence of repeat elements within this non-coding region in *C. sinensis*. Although the origin(s) of such variation in flatworms is presently unknown, it might be the result of double-strand break repair or slipped-strand mispairing during replication [39, 40].

The identification in the sequence data set of long-reads containing > 6 repeat units that did not span the non-coding region (4.5 kb) suggested partial degradation of mitochondrial DNA in the total DNA sample—extracted from *C. sinensis* worms collected in 2009—used for nanopore-sequencing. Some degradation or nicking of repetitive DNA would be expected to occur in a sample stored frozen for such an extended period (11 years). However, it is also possible that secondary structural arrangements in repetitive elements (Fig 2) might have led to some nicking during sequencing, resulting in a proportion of incomplete sequences, which is plausible for long DNA strands.

Overcoming the challenges of sequencing the tandem-repetitive region

The mitochondrial genomes of a range of flatworms (cestodes and trematodes) are known to harbour non-coding regions containing repetitive elements [2, 7]. Short and long non-coding regions appear to be characteristic of trematodes, although often partially sequenced using Sanger- or short-read sequencing methods [3, 7, 21]. The comparison of the present mitochondrial genome assembly with published mitochondrial genomes of *C. sinensis* revealed that the newly-characterised tandem-repeat region occurs between tRNA-Glu and tRNA-Gly, formerly estimated at 153–154 bp in size [20–22]. A short non-coding region between tRNA-Gly and *cox3* equated to 67 bp, as reported previously (67 or 68 bp). All 12 protein-encoding genes, 22 tRNAs and two rRNAs had high sequence similarities (> 99.2%) to those in published mitochondrial genomes and occurred in the same order. However, there is clear evidence [17, 18, 38] that conventional sequencing methods are not suited to the sequencing of long non-coding regions in mitochondrial genomes. This obstacle has been overcome through

the use of nanopore-sequencing, which bodes well for future mitochondrial genome investigations.

Speculating about the role(s) of non-coding elements in the mitochondrial genome

Although the functions of long non-coding elements in the mitochondrial genome of parasitic flatworms are unexplored, they are hypothesised to be ‘control’ regions, which initiate replication and transcription [7, 41–44]. In bilaterian animals, the control region is typically ~ 1 kb in size [45–49] and often contains short repeat elements, predicted to fold into secondary structures [37]. Although significant deviations from a ‘typical’ animal mitochondrial genome exist [50] and duplications of control regions are known to occur [51–55], expansive repetitive non-coding regions with substantial size variation within a species seem to be unusual. For parasitic flatworms, we propose that each tandemly-repeated unit represents a distinct control region possibly enhancing replication and transcription efficiency [17]. Multiple control regions within the mitochondrial genome might provide an advantage in terms of being able to adapt cellular energy production and metabolism during particular life-cycle phases while under strong selective pressure in different environments, both outside of or within a host animal (e.g., O₂, pH, salinity, temperature, light, osmotic pressure and/or nutrient accessibility).

Efficient replication might also limit the detrimental effect of extreme environments on mitochondrial DNA integrity. A plethora of internal and external agents (e.g., reactive oxygen species, metabolites, radiation, environmental chemicals and toxins) are known to cause DNA damage such as mutations and lesions, of which double-strand breaks (DSBs) are particularly harmful [56–58]. Although animal DNA is constantly exposed to such stressors, it could be proposed that many organisms, such as parasitic helminths, inhabit particularly inhospitable environments that cause chronic damage to mitochondrial DNA and that unique strategies might have evolved to achieve efficient genome maintenance and ensure cellular viability. Conditions potentially disrupting the mitochondrial DNA integrity of *C. sinensis* could include exposure to toxic bile salts and acids and/or desiccation, which have been shown to cause DNA anomalies such as DSBs in some microbe and metazoan species [59–66]. In response to this stress, replication of the mitochondrial genome might need to be highly efficient, in order to have a high number of genomes in the cell at any one time. This might avoid harmful mutations in the mitochondrial genome by increasing the number of template molecules in each cell, required to repair DNA in the least error-prone way [58, 67, 68]. A large number of genomes might act also as a ‘buffer’ in the cell—even if some get damaged, many functionally intact genomes will be present, ensuring that replication and transcription of mitochondrial genes are not disrupted within the cell. Whether selection acts upon the size of the repeat region in the mitochondrial genome of *C. sinensis*, or whether repeat expansions and contractions represent stochastic events, such as errors during DNA repair (e.g., [40]), warrants investigation. Future work might explore whether the repetitive region might function as an ‘origin of replication’ using a combination of two-dimensional neutral agarose gel electrophoresis and electron microscopy techniques [69].

Concluding remarks

The first characterisation of a novel tandem-repetitive region (> 4.5 kb) in *C. sinensis* and variation in the sequence and number of repeat elements within this region raise questions about (i) the functional role(s) of this region within cells and mitochondria; (ii) the origin of such variation and whether it occurs within cells or tissues within individual worms, or among worms; and (iii) what impact such variation has on mitochondrial, nuclear and/or cellular

functions. In our opinion, these research questions would be interesting to pursue in the near future.

Acknowledgments

We acknowledge the use of HPC-GPGPU Facility computer resources hosted at the University of Melbourne.

Author Contributions

Conceptualization: Liina Kinkar, Neil D. Young, Andreas J. Stroehlein, Pasi K. Korhonen, Robin B. Gasser.

Formal analysis: Liina Kinkar, Neil D. Young, Robin B. Gasser.

Funding acquisition: Neil D. Young, Pasi K. Korhonen, Robin B. Gasser.

Methodology: Liina Kinkar, Neil D. Young, Andreas J. Stroehlein, Pasi K. Korhonen, Robin B. Gasser.

Project administration: Robin B. Gasser.

Resources: Woon-Mok Sohn, Robin B. Gasser.

Writing – original draft: Liina Kinkar, Neil D. Young, Robin B. Gasser.

Writing – review & editing: Andreas J. Stroehlein, Pasi K. Korhonen.

References

1. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. *Nature*. 2017; 550(7676):345–53. <https://doi.org/10.1038/nature24286> PMID: [29019985](https://pubmed.ncbi.nlm.nih.gov/29019985/).
2. 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](https://doi.org/10.1016/S0001-706X(00)00157-1) PMID: [11114386](https://pubmed.ncbi.nlm.nih.gov/11114386/).
3. Littlewood DTJ, Lockyer AE, Webster BL, Johnston DA, Le TH. The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. *Mol Phylogenet Evol*. 2006; 39(2):452–67. <https://doi.org/10.1016/j.ympev.2005.12.012> PMID: [16464618](https://pubmed.ncbi.nlm.nih.gov/16464618/).
4. Zarowiecki MZ, Huyse T, Littlewood DTJ. Making the most of mitochondrial genomes—markers for phylogeny, molecular ecology and barcodes in *Schistosoma* (Platyhelminthes: Digenea). *Int J Parasitol*. 2007; 37(12):1401–18. <https://doi.org/10.1016/j.ijpara.2007.04.014> PMID: [17570370](https://pubmed.ncbi.nlm.nih.gov/17570370/).
5. Mohandas N, Pozio E, La Rosa G, Korhonen PK, Young ND, Koehler AV, et al. Mitochondrial genomes of *Trichinella* species and genotypes—a basis for diagnosis, and systematic and epidemiological explorations. *Int J Parasitol*. 2014; 44(14):1073–80. <https://doi.org/10.1016/j.ijpara.2014.08.010> PMID: [25245252](https://pubmed.ncbi.nlm.nih.gov/25245252/).
6. Le TH, Nguyen KT, Nguyen NTB, Doan HTT, Agatsuma T, Blair D. The complete mitochondrial genome of *Paragonimus ohirai* (Paragonimidae: Trematoda: Platyhelminthes) and its comparison with *P. westermani* congeners and other trematodes. *PeerJ*. 2019; 7:e7031. <https://doi.org/10.7717/peerj.7031> PMID: [31259095](https://pubmed.ncbi.nlm.nih.gov/31259095/).
7. 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](https://doi.org/10.1016/s1471-4922(02)02252-3) PMID: [11983601](https://pubmed.ncbi.nlm.nih.gov/11983601/).
8. Hu M, Chilton NB, Gasser RB. The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). *Int J Parasitol*. 2002; 32(2):145–58. [https://doi.org/10.1016/S0020-7519\(01\)00316-2](https://doi.org/10.1016/S0020-7519(01)00316-2) PMID: [11812491](https://pubmed.ncbi.nlm.nih.gov/11812491/).
9. Hu M, Gasser RB, El-Osta YGA, Chilton NB. Structure and organization of the mitochondrial genome of the canine heartworm, *Dirofilaria immitis*. *Parasitology*. 2003; 127(1):37–51. <https://doi.org/10.1017/S0031182003003275> PMID: [12885187](https://pubmed.ncbi.nlm.nih.gov/12885187/).

10. Hyman BC, Lewis SC, Tang S, Wu Z. Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. *Genetica*. 2011; 139(5):611–5. <https://doi.org/10.1007/s10709-010-9531-3> PMID: 21136141.
11. Zou H, Jakovlić I, Chen R, Zhang D, Zhang J, Li W-X, et al. The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. *BMC Genomics*. 2017; 18(1):840. <https://doi.org/10.1186/s12864-017-4237-x> PMID: 29096600.
12. Hu M, Gasser RB. Mitochondrial genomes of parasitic nematodes—progress and perspectives. *Trends Parasitol*. 2006; 22(2):78–84. <https://doi.org/10.1016/j.pt.2005.12.003> PMID: 16377245.
13. Littlewood DTJ. Platyhelminth systematics and the emergence of new characters. *Parasite*. 2008; 15(3):333–41. <https://doi.org/10.1051/parasite/2008153333> PMID: 18814704.
14. Wey-Fabrizius AR, Podsiadlowski L, Herlyn H, Hankeln T. Platyzoan mitochondrial genomes. *Mol Phylogenet Evol*. 2013; 69(2):365–75. <https://doi.org/10.1016/j.ympev.2012.12.015> PMID: 23274056.
15. Solà E, Álvarez-Presas M, Frías-López C, Littlewood DTJ, Rozas J, Riutort M. Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. *PLoS One*. 2015; 10(3):e0120081. <https://doi.org/10.1371/journal.pone.0120081> PMID: 25793530.
16. Le TH, Humair P-F, Blair D, Agatsuma T, Littlewood DTJ, McManus DP. Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. *Mol Biochem Parasitol*. 2001; 117(1):61–71. <https://doi.org/10.1016/S0166-6851> PMID: 11551632.
17. Kinkar L, Korhonen PK, Cai H, Gauci CG, Lightowlers MW, Saarma U, et al. Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. *Parasit Vectors*. 2019; 12(1):238. <https://doi.org/10.1186/s13071-019-3492-x> PMID: 31097022.
18. Oey H, Zakrzewski M, Narain K, Devi KR, Agatsuma T, Nawaratna S, et al. Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *Gigascience*. 2019; 8(1):gij146. <https://doi.org/10.1093/gigascience/giy146> PMID: 30520948.
19. Kinkar L, Laurimäe T, Acosta-Jamett G, Andresiuk V, Balkaya I, Casulli A, et al. Global phylogeography and genetic diversity of the zoonotic tapeworm *Echinococcus granulosus sensu stricto* genotype G1. *Int J Parasitol*. 2018; 48(9–10):729–42. <https://doi.org/10.1016/j.ijpara.2018.03.006> PMID: 29782829.
20. Shekhovtsov SV, Katokhin AV, Kolchanov NA, Mordvinov VA. The complete mitochondrial genomes of the liver flukes *Opisthorchis felinus* and *Clonorchis sinensis* (Trematoda). *Parasitol Int*. 2010; 59(1):100–3. <https://doi.org/10.1016/j.parint.2009.10.012> PMID: 19906359.
21. 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.
22. 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.
23. Sohn W-M, Zhang H, Choi M-H, Hong S-T. Susceptibility of experimental animals to reinfection with *Clonorchis sinensis*. *Korean J Parasitol*. 2006; 44(2):163–6. <https://doi.org/10.3347/kjp.2006.44.2.163> PMID: 16809966.
24. 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.
25. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; 34(18):3094–100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242.
26. 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.
27. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013; 10(6):563–9. <https://doi.org/10.1038/nmeth.2474> PMID: 23644548.
28. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004; 5(2):R12. <https://doi.org/10.1186/gb-2004-5-2-r12> PMID: 14759262.
29. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014; 9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509.

30. 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.
31. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res*. 2008; 36:W70–74. <https://doi.org/10.1093/nar/gkn188> PMID: 18424795.
32. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*. 2015; 31(20):3377–9. <https://doi.org/10.1093/bioinformatics/btv372> PMID: 26099263.
33. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol*. 2008; 26(10):1146–53. <https://doi.org/10.1038/nbt.1495> PMID: 18846088.
34. Gonzalez JM, Zimmermann J, Saiz-Jimenez C. Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics*. 2005; 21(3):333–7. <https://doi.org/10.1093/bioinformatics/bti008> PMID: 15347575.
35. Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, et al. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene*. 2010; 469(1):45–51. <https://doi.org/10.1016/j.gene.2010.08.009> PMID: 20833233.
36. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res*. 2011; 21(3):494–504. <https://doi.org/10.1101/gr.112730.110> PMID: 21212162.
37. Lunt DH, Whipple LE, Hyman BC. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol*. 1998; 7(11):1441–55. <https://doi.org/10.1046/j.1365-294x.1998.00495.x> PMID: 9819900.
38. Oey H, Zakrzewski M, Gravermann K, Young ND, Korhonen PK, Gobert GN, et al. Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*. *PLoS Pathog*. 2019; 15(1):e1007513. <https://doi.org/10.1371/journal.ppat.1007513> PMID: 30673782.
39. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. 1987; 4(3):203–21. <https://doi.org/10.1093/oxfordjournals.molbev.a040442> PMID: 3328815.
40. Pâques F, Leung W-Y, Haber JE. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol Cell Biol*. 1998; 18(4):2045–54. <https://doi.org/10.1128/MCB.18.4.2045> PMID: 9528777.
41. Nakao M, Yokoyama N, Sako Y, Fukunaga M, Ito A. The complete mitochondrial DNA sequence of the cestode *Echinococcus multilocularis* (Cyclophyllidae: Taeniidae). *Mitochondrion*. 2002; 1(6):497–509. [https://doi.org/10.1016/s1567-7249\(02\)00040-5](https://doi.org/10.1016/s1567-7249(02)00040-5) PMID: 16120302.
42. Huyse T, Buchmann K, Littlewood DTJ. The mitochondrial genome of *Gyrodactylus derjavinoidea* (Platyhelminthes: Monogenea)—A mitogenomic approach for *Gyrodactylus* species and strain identification. *Gene*. 2008; 417(1):27–34. <https://doi.org/10.1016/j.gene.2008.03.008> PMID: 18448274.
43. Sakai M, Sakaizumi M. The complete mitochondrial genome of *Dugesia japonica* (Platyhelminthes; Order Tricladida). *Zool Sci*. 2012; 29(10):672–80. <https://doi.org/10.2108/zsj.29.672> PMID: 23030340.
44. Egger B, Bachmann L, Fromm B. Atp8 is in the ground pattern of flatworm mitochondrial genomes. *BMC Genomics*. 2017; 18(1):414. <https://doi.org/10.1186/s12864-017-3807-2> PMID: 28549457.
45. Taanman J-W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta Bioenerg*. 1999; 1410(2):103–23. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3) PMID: 10076021.
46. Duarte GT, De Azeredo-Espin AML, Junqueira ACM. The mitochondrial control region of blowflies (Diptera: Calliphoridae): a hot spot for mitochondrial genome rearrangements. *J Med Entomol*. 2008; 45(4):667–76. <https://doi.org/10.1093/jmedent/45.4.667> PMID: 18714866.
47. Terencio ML, Schneider CH, Gross MC, Feldberg E, Porto JIR. Structure and organization of the mitochondrial DNA control region with tandemly repeated sequence in the Amazon ornamental fish. *Mitochondrial DNA*. 2013; 24(1):74–82. <https://doi.org/10.3109/19401736.2012.717934> PMID: 22954310.
48. Gonçalves R, Freitas AI, Jesus J, De la Rúa P, Brehm A. Structure and genetic variation of the mitochondrial control region in the honey bee *Apis mellifera*. *Apidologie*. 2015; 46(4):515–26. <https://doi.org/10.1007/s13592-014-0341-y>.
49. Rahman MM, Yoon KB, Park YC. Structural characteristics of a mitochondrial control region from *Myotis bat* (Vespertilionidae) mitogenomes based on sequence datasets. *Data Brief*. 2019; 24:103830. <https://doi.org/10.1016/j.dib.2019.103830> PMID: 31032389.

50. Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol Evol.* 2016; 8(9):2896–913. <https://doi.org/10.1093/gbe/evw195> PMID: 27557826.
51. Eberhard JR, Wright TF, Bermingham E. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. *Mol Biol Evol.* 2001; 18(7):1330–42. <https://doi.org/10.1093/oxfordjournals.molbev.a003917> PMID: 11420371.
52. Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks. *Mol Biol Evol.* 2005; 22(3):620–9. <https://doi.org/10.1093/molbev/msi047> PMID: 15537802.
53. Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, et al. Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol Phylogenet Evol.* 2012; 64(2):342–56. <https://doi.org/10.1016/j.ympev.2012.04.009> PMID: 22543055.
54. Zheng C, Nie L, Wang J, Zhou H, Hou H, Wang H, et al. Recombination and evolution of duplicate control regions in the mitochondrial genome of the Asian big-headed turtle, *Platysternon megacephalum*. *PLoS One.* 2013; 8(12):e82854. <https://doi.org/10.1371/journal.pone.0082854> PMID: 24367563.
55. Akiyama T, Nishida C, Momose K, Onuma M, Takami K, Masuda R. Gene duplication and concerted evolution of mitochondrial DNA in crane species. *Mol Phylogenet Evol.* 2017; 106:158–63. <https://doi.org/10.1016/j.ympev.2016.09.026> PMID: 27693570.
56. Jeggo PA, Löbrich M. DNA double-strand breaks: their cellular and clinical impact? *Oncogene.* 2007; 26(56):7717–9. <https://doi.org/10.1038/sj.onc.1210868> PMID: 18066083.
57. Alexeyev M, Shokolenko I, Wilson G, LeDoux S. The maintenance of mitochondrial DNA integrity—critical analysis and update. *Cold Spring Harb Perspect Biol.* 2013; 5(5):a012641. <https://doi.org/10.1101/cshperspect.a012641> PMID: 23637283.
58. García-Lepe UO, Bermúdez-Cruz RM. Mitochondrial genome maintenance: damage and repair pathways. In: Mognato M, editor. *DNA Repair—An Update*. London, UK: IntechOpen; 2019. <https://doi.org/10.5772/intechopen.84627>.
59. Kandell RL, Bernstein C. Bile salt/acid induction of DNA damage in bacterial and mammalian cells: implications for colon cancer. *Nutr Cancer.* 1991; 16(3–4):227–38. <https://doi.org/10.1080/01635589109514161> PMID: 1775385.
60. Prieto AI, Ramos-Morales F, Casadesús J. Repair of DNA damage induced by bile salts in *Salmonella enterica*. *Genetics.* 2006; 174(2):575–84. <https://doi.org/10.1534/genetics.106.060889> PMID: 16888329.
61. Payne CM, Bernstein C, Dvorak K, Bernstein H. Hydrophobic bile acids, genomic instability, Darwinian selection, and colon carcinogenesis. *Clin Exp Gastroenterol.* 2008; 1:19–47. <https://doi.org/10.2147/ceg.s4343> PMID: 21677822.
62. Merritt ME, Donaldson JR. Effect of bile salts on the DNA and membrane integrity of enteric bacteria. *J Med Microbiol.* 2009; 58(Pt 12):1533–41. <https://doi.org/10.1099/jmm.0.014092-0> PMID: 19762477.
63. Hardie LJ. The genotoxicity of bile acids. In: Jankins G, Hardie LJ, editors. *Bile Acids: Toxicology and Bioactivity*. Royal Society of Chemistry, Cambridge; 2008. pp. 72–83.
64. Gusev O, Nakahara Y, Vanyagina V, Malutina L, Cornette R, Sakashita T, et al. Anhydrobiosis-associated nuclear DNA damage and repair in the sleeping chironomid: linkage with radioresistance. *PLoS One.* 2010; 5(11):e14008. <https://doi.org/10.1371/journal.pone.0014008> PMID: 21103355.
65. Hespeels B, Knapen M, Hanot-Mambres D, Heuskin A-C, Pineux F, Lucas S, et al. Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. *J Evol Biol.* 2014; 27(7):1334–45. <https://doi.org/10.1111/jeb.12326> PMID: 25105197.
66. Negretti NM, Gourley CR, Clair G, Adkins JN, Konkel ME. The food-borne pathogen *Campylobacter jejuni* responds to the bile salt deoxycholate with countermeasures to reactive oxygen species. *Sci Rep.* 2017; 7(1):15455. <https://doi.org/10.1038/s41598-017-15379-5> PMID: 29133896.
67. Li X, Heyer W-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* 2008; 18(1):99–113. <https://doi.org/10.1038/cr.2008.1> PMID: 18166982.
68. Chen XJ. Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiol Mol Biol Rev.* 2013; 77(3):476–96. <https://doi.org/10.1128/MMBR.00007-13> PMID: 24006472.
69. Lewis SC, Joers P, Willcox S, Griffith JD, Jacobs HT, Hyman BC. A rolling circle replication mechanism produces multimeric lariats of mitochondrial DNA in *Caenorhabditis elegans*. *PLoS Genet.* 2015; 11:e1004985. <https://doi.org/10.1371/journal.pgen.1004985> PMID: 25693201.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Kinkar, L;Young, ND;Sohn, W-M;Stroehlein, AJ;Korhonen, PK;Gasser, RB

Title:

First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach

Date:

2020-08-01

Citation:

Kinkar, L., Young, N. D., Sohn, W. -M., Stroehlein, A. J., Korhonen, P. K. & Gasser, R. B. (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), <https://doi.org/10.1371/journal.pntd.0008552>.

Persistent Link:

<http://hdl.handle.net/11343/244613>

License:

CC BY