

RESEARCH

Open Access



# Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1

Liina Kinkar<sup>1</sup>, Pasi K. Korhonen<sup>1</sup>, Huimin Cai<sup>2</sup>, Charles G. Gauci<sup>1</sup>, Marshall W. Lightowlers<sup>1</sup>, Urmas Saarma<sup>3</sup>, David J. Jenkins<sup>4</sup>, Jiandong Li<sup>2</sup>, Junhua Li<sup>2</sup>, Neil D. Young<sup>1</sup> and Robin B. Gasser<sup>1\*</sup>

## Abstract

**Background:** *Echinococcus* tapeworms cause a severe helminthic zoonosis called echinococcosis. The genus comprises various species and genotypes, of which *E. granulosus* (*sensu stricto*) represents a significant global public health and socioeconomic burden. Mitochondrial (mt) genomes have provided useful genetic markers to explore the nature and extent of genetic diversity within *Echinococcus* and have underpinned phylogenetic and population structure analyses of this genus. Our recent work indicated a sequence gap (> 1 kb) in the mt genomes of *E. granulosus* genotype G1, which could not be determined by PCR-based Sanger sequencing. The aim of the present study was to define the complete mt genome, irrespective of structural complexities, using a long-read sequencing method.

**Methods:** We extracted high molecular weight genomic DNA from protoscoleces from a single cyst of *E. granulosus* genotype G1 from a sheep from Australia using a conventional method and sequenced it using PacBio Sequel (long-read) technology, complemented by BGISEQ-500 short-read sequencing. Sequence data obtained were assembled using a recently-developed workflow.

**Results:** We assembled a complete mt genome sequence of 17,675 bp, which is > 4 kb larger than the complete mt genomes known for *E. granulosus* genotype G1. This assembly includes a previously-elusive tandem repeat region, which is 4417 bp long and consists of ten near-identical 441–445 bp repeat units, each harbouring a 184 bp non-coding region and adjacent regions. We also identified a short non-coding region of 183 bp, which includes an inverted repeat.

**Conclusions:** We report what we consider to be the first complete mt genome of *E. granulosus* genotype G1 and characterise all repeat regions in this genome. The numbers, sizes, sequences and functions of tandem repeat regions remain to be studied in different isolates of genotype G1 and in other genotypes and species. The discovery of such 'new' repeat elements in the mt genome of genotype G1 by PacBio sequencing raises a question about the completeness of some published genomes of taeniid cestodes assembled from conventional or short-read sequence datasets. This study shows that long-read sequencing readily overcomes the challenges of assembling repeat elements to achieve improved genomes.

**Keywords:** *Echinococcus*, Genotype G1, Complete mitochondrial (mt) genome, Repetitive DNA, PacBio sequencing

\*Correspondence: robinbg@unimelb.edu.au

<sup>1</sup> Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia

Full list of author information is available at the end of the article



## Background

Cestodes of the genus *Echinococcus* cause a disease called echinococcosis, which affects humans, and various domestic and wild mammals [1]. *Echinococcus* spp. are distributed worldwide and represent a substantial global public health and socioeconomic burden [2]. Echinococcosis is recognised by the World Health Organization (WHO) as a neglected tropical disease (NTD), requiring a prioritisation of global research and control efforts [3].

Genetically, *Echinococcus* is a diverse cestode group, currently consisting of 10 species [4–7]: *E. multilocularis*, *E. oligarthra*, *E. vogeli*, *E. shiquicus*, *E. granulosus* (*sensu stricto*; genotypes G1 and G3), *E. equinus* (genotype G4), *E. ortleppi* (genotype G5), *E. intermedius* (species name is being debated [6, 8–10]; comprising genotypes G6 and G7), *E. canadensis* (genotypes G8 and G10) and *E. felidis*. These species are distinctly different from one another in their ecology (e.g. infectivity to humans, prevalence, distribution and host ranges) [1]; thus, exploring the extent of genetic variation within the genus *Echinococcus* is central to understanding disease transmission patterns. *Echinococcus granulosus* genotype G1 is recognised as the most wide-spread of all *Echinococcus* taxa, and is, thus, of particular importance [2, 11].

Mitochondrial (mt) genomes have provided useful genetic markers to discover the nature and extent of genetic diversity within *Echinococcus* [7, 12–15], and have underpinned extensive phylogenetic and population structure analyses of this genus over the years (e.g. [16–20]). Published reports show that mt genomes of genotype G1 sequenced using Sanger-, 454- or Illumina-methods are ~13,600 bp in length [16, 21, 22]; in addition to 12 protein-encoding genes, 22 tRNAs and 2 rRNAs, the mt genome contains two non-coding regions (NCRs), one of which is between tRNA-Tyr and tRNA-Leu genes (NR1; estimated at 87 bp [16] or 66 bp [21]) and the other between *nad5* and tRNA-Gly (NR2, estimated at 184 bp; [16, 21]). However, our recent work [23], exploring the “global” genetic structure of *E. granulosus* genotype G1 using near-complete mitogenome sequences of 211 individual samples of this genotype revealed a gap (estimated at >1 kb) between the 3'-end of the *nad5* gene and the 5'-end of the *cox3* gene. In spite of many attempts to PCR-amplify (using a range of different oligonucleotide primer sets designed specifically to the *nad5* and *cox3* genes flanking this enigmatic region), we were not able to define this region for any of the 211 genotype G1 isolates investigated using a Sanger-based sequencing approach (cf. [23]). This finding was suggestive of a repetitive, non-coding region of complex structure(s) that is “resistant” to amplification by conventional PCR. This challenge needed to be circumvented using a different approach.

The availability of PacBio sequencing technology [24] and the advent of an automated pipeline [25] for the assembly of long-read sequence data provided an opportunity to overcome this obstacle and to, for the first time, directly define a complete mt genome of *Echinococcus*, in one sweep, irrespective of the nature or structural complexities in intergenic regions. Here, we report what we consider to be the first complete mt genome of *E. granulosus* genotype G1 and characterise all repeat regions in this genome.

## Methods

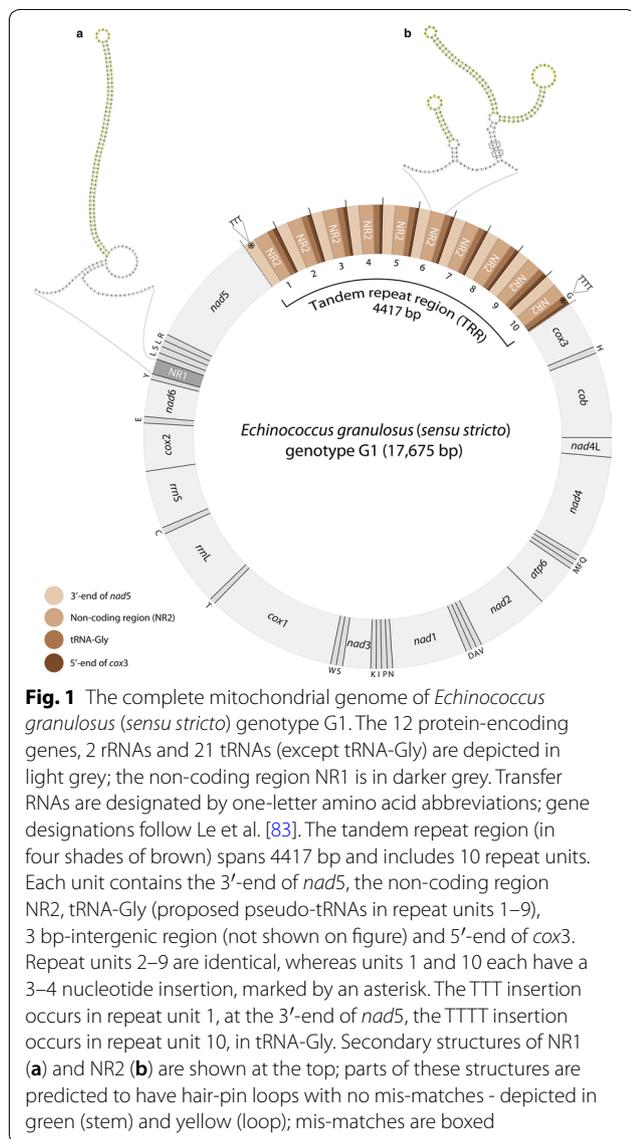
High molecular weight genomic DNA, extracted from protoscoleces from a single cyst of *E. granulosus* genotype G1 from a liver of a sheep from New South Wales in Australia using a conventional phenol:chloroform method [26], was sequenced using the PacBio Sequel [27] and BGISEQ-500 short-read sequencing [28] platforms employing established protocols. Sequence data obtained were assembled using a recently established common workflow language (CWL)-based pipeline [25]. To check the contiguity and sequencing depth of the assembly, the program Circlator v1.5.5 [29] was used to map corrected PacBio reads to the assembly; subsequently, short-read data were mapped using the program Bowtie2 v2.1.0 [30] and sorted using the program SAMtools v1.3.1 [31]. Aligned reads were inspected for any nucleotide inconsistencies using the program IGV v2.3.97 [32].

The mt genome was compared with a representative, published mt genome of *E. granulosus* genotype G1 (GenBank: AF297617; [21]), and its sequence deposited in the GenBank database under accession no. MK774655. Protein-encoding genes and rRNAs were annotated using an established bioinformatics pipeline [33]; tRNAs were identified using the same bioinformatic pipeline [33] and/or by a BLAST search [34] against an *E. granulosus* genotype G1 mt genome sequence available in GenBank (accession no. AB786664; [16]). The repeat region was characterised using the tandem repeat finder Dot2dot [35], and the secondary structures of non-coding regions were predicted using the RNAfold web server [36]. All annotations were curated manually.

## Results and discussion

From totals of 2757 PacBio long-reads equating to 41 Mb, we assembled a complete mt genome for *E. granulosus* genotype G1 at an average sequencing depth of 2268, resulting in a contig of 17,675 bp, which is >4 kb larger than all published mt genomes representing genotype G1 (~13,600 bp) [16, 21, 22], but with the same order of protein-coding genes (Fig. 1).

We succeeded in assembling a tandem repeat region (TRR) of 4417 bp; 1895 reads spanned this and flanking



regions ( $\geq 1$  kb, both 5' and 3'). No length or sequence variation was detected among these reads, indicating a lack of polymorphism within or among protoscolecemes within the sample from one cyst. The annotation of TRR revealed that it contains 10 tandemly repeated, near-identical repeat units: the first unit is 444 bp in length; units two to nine are identical in sequence, being 441 bp; and the tenth unit is 445 bp (Fig. 1). Each repeat unit within TRR contains 144–147 bp of the 3'-end of *nad5*, 184 bp of the non-coding region NR2, 63–67 bp of tRNA-Gly, a 3 bp-intergenic region and 47 bp of the start of the *cox3* gene (Fig. 1). The tRNA in the tenth unit appears to be the only functional tRNA-Gly in the genome, but this suggestion requires experimental verification. In addition to characterising the 4417 bp intergenic region (TRR),

we also succeeded in defining the complete sequence of the non-coding region NR1, which equated to a total of 183 bp (Fig. 1), the same length as estimated previously for *E. multilocularis* [37] but longer than the 87 bp [16] or 66 bp [21] recorded for *E. granulosus* genotype G1. Parts of NR1 and NR2 were predicted to fold into secondary stem-loop structures with possible roles in the regulation of replication and/or transcription of the mt genome (Fig. 1).

Sanger-sequencing and second-generation short-read sequencing, used previously to sequence mt genomes of *Echinococcus* [7, 16, 21–23, 37–39], are not suited to defining complex genomic regions, such as repeat elements. However, PacBio single-molecule real-time sequencing offers the long-read lengths to identify and characterise long, complicated repetitive regions [24], as achieved here. Other recent examples of success with PacBio sequencing include the resolution of unique, complex sequence tracts of ~4 kb and ~6.9 kb in the mt genomes of *Schistosoma bovis* [40] and *Paragonimus westermani* [41], respectively.

Published mt genomes of parasitic and free-living flatworms (e.g. [40–46]) are known to harbour non-coding regions of varying sizes, which may contain repeat elements. Two NCRs (66–875 bp) [47] appear to be characteristic of cestodes, in which repetitive elements are relatively common [47–55]. In *E. granulosus* genotype G1, the two NCRs are located between two tRNAs upstream of the *nad5* gene (NR1) and between *nad5* and tRNA-Gly (NR2), and have both been estimated at <200 bp [16, 21]. Our results are consistent with the previous observations, in terms of the location of NR1 and NR2, and the length and sequence of NR2; however, the tandem replication of NR2 and its adjacent sequences are unique features not previously reported for mt genomes of cestodes.

In the future, the nature and extent of polymorphism in the tandem repeat region should be assessed by sequencing a large number of genotype G1 samples from individual cysts from distinct hosts and geographical locations using the PacBio approach, as the consistency of occurrence, size and sequence in this repeat region are presently unknown. It is possible that not all G1 isolates harbour tandem repeats, as they had not been observed in the published mitogenomes of G1 [16, 21, 22]. However, as the sequencing approaches used previously might not have been able to resolve complex regions, there is a question regarding the completeness of previously published mt genome sequences. In addition to the intra-genotypic variation, length and/or structural differences in TRR among different *Echinococcus* taxa should also be explored. The mt genome of genotype G3 is likely to harbour the tandem repeat region as well, as we detected but

could not resolve the enigmatic region for 39 G3 samples using Sanger-based sequencing [56]. However, this might not be the case for genotypes G6 and G7, as Sanger-based sequencing defined, without complication, complete mt genomes ( $n=94$ ) for these genotypes in a recent study [7]. Taken together, these findings suggest that there is significant scope for studies of the nature and extent of variation in repeat regions within and among different *Echinococcus* species and genotypes, and their evolution.

We hypothesise that tandem repeats within genotype G1 might provide an evolutionary advantage over mt genomes with no such replications. Most mt genomes of animals are relatively small (typically 15–20 kb in size; [57, 58]), lack introns and have short intergenic regions (usually only a few bp; [59]) and are, thus, thought to be under selection for compactness (cf. [60]). Non-functional replications could be rare and would be expected to be eliminated relatively quickly due to the rapid rate of replication of compact mt genomes [60, 61]. It could be speculated that the existence of the tandem repeat region (TRR) within the mt genome of G1 overrides the selection for a small genome size and might provide an evolutionary advantage. A key element of this proposal could be the existence of replicated control regions (CRs) within TRR.

It is well established that mt genomes of animals contain a control region that initiates replication and transcription [62, 63]. Interestingly, there have been several reports of duplications of the control region in the mt genomes of various species of animals [64–68], which are thought to be advantageous, in terms of more efficient transcription and/or replication of mt genes [64, 65, 69]. As a working hypothesis, we propose that the 184 bp non-coding sequence (NR2) within each repeat unit of TRR is a putative control region of genotype G1 and, thus, the mt genome contains 10 identical copies of CR which might be beneficial, in terms of more efficient replication and/or transcription. Parts of this region appear to be capable of folding into secondary stem-loop structures (see Fig. 1), which, as suggested previously [37, 42, 44, 55], could be associated with mt genome replication in cestodes. This hypothesis warrants testing.

If the mt genome of *E. granulosus* genotype G1 did consistently contain 10 identical CRs, this might provide an advantage, in terms of cellular energy production, especially during life-cycle phases that require short-term bursts of energy in a micro-aerobic habitat. As an adaptation to this environment, it has been hypothesised that the parasite uses fermentative pathways to generate cellular energy, specifically lactic fermentation and malate dismutation [70–72]. While lactate is produced in the cytosol and excreted, mt fermentation of malate is known to occur in helminths [73] and is encoded in

*Echinococcus* [74]. More effective mt replication and/or transcription mechanisms might compensate for the lower energy yield of fermentation [73] compared with aerobic respiration [75] and be under strong selective pressure. Efficient energy production would be particularly important during the phase in which eggs hatch, oncospheres activate and are then required to rapidly penetrate the intestinal wall of the intermediate host animal [76, 77]. The successful development of an *Echinococcus* cyst in an intermediate host is highly dependent on a rapid penetration of the oncosphere and immediate post-oncospherical establishment [77, 78]. Interestingly, genotype G1 has the broadest host range of all *Echinococcus* taxa [2]. Thus, it could be proposed that efficient energy production at the oncosphere stage might be one of the factors contributing to this genotype's success at infecting a diverse range of host species. Another crucial phase requiring rapid energy production is during the development of protoscoleces into adult worms in the small intestine of the definitive host [77].

We suggest that the other non-coding region, NR1, might have an exclusive functional role in the replication of the mt genome. Several mechanisms, including rolling circle, strand-displacement, and strand-coupled replication, have been proposed for mt DNA in vertebrates and invertebrates [62]. Although the replication mechanisms in cestodes are not understood, the secondary structure of NR1 (see Fig. 1) seems to lend support to the strand-displacement mechanism being utilised (cf. [63]). According to this model, there are two distinct origins of replication, a CR containing the origin of leading-strand replication and another origin initiating lagging-strand replication, which is characterised by a stem-loop structure [63]. The 183 bp non-coding region assembled here, for the first time, for genotype G1 appears to assume a long stem-loop (Fig. 1), suggesting that, if the strand-displacement mechanism is utilised by the *Echinococcus* tapeworms, NR1 could be the initiation site for lagging-strand replication. The NR1 (183 bp) identified in *E. multilocularis* [37] is also predicted to fold into a long stem-loop of a similar size [44], suggesting that it has structural and functional significance in the mt genome. Future work might focus on exploring the roles of both NR1 and NR2 using 2D neutral agarose gel electrophoresis, Southern blot-hybridisation and electron microscopy techniques [79–82].

## Conclusions

Here, we report what we consider to be the first complete mt genome of *E. granulosus* genotype G1. We succeeded in defining an elusive tandem repeat region (4.4 kb), which consists of ten repeat units, each harbouring a

184 bp non-coding region and adjacent regions, a unique feature, not previously observed in mt genomes of cestodes. We also characterised a short non-coding region (183 bp; containing a long, inverted repeat) for the first time for genotype G1. The presence, size, sequence and function of tandem repeat regions in different isolates of genotype G1, and in other genotypes and species, remain to be studied. The discovery here of “new” repeat elements in the mt genome of G1 raises a question about the completeness of some published genomes of taeniidae assembled previously from conventional or short-read sequence data sets. The present study shows that PacBio sequencing overcomes the challenges associated with the assembly of repeat elements in genomes and indicates its benefits for investigating the genomes of cestodes and other parasites.

#### Abbreviations

WHO: World Health Organization; NTD: neglected tropical disease; mt: mitochondrial; bp: base pair or base pairs; kb: kilobase or kilobases; Mb: megabase or megabases; PCR: polymerase chain reaction; tRNA: transfer RNA; rRNA: ribosomal RNA; Leu: Leucine; Tyr: Tyrosine; Gly: Glycine; *cox3*: cytochrome c oxidase subunit 3; *nad5*: NADH dehydrogenase subunit 5; TRR: tandem repeat region; NCR: non-coding region; NR1: non-coding region 1; NR2: non-coding region 2; CR: control region; CWL: common workflow language; 2D: two-dimensional.

#### Acknowledgements

Thanks to Ross S. Hall for technical support and to Professor Ian Beveridge for advice on the taxonomy of cestodes.

#### Authors' contributions

CGG, MWL and DJJ sourced the isolate of *E. granulosus* genotype G1 and prepared genomic DNA for sequencing. HC, JiL and JuL performed the sequencing. LK, PKK, US, NDY and RBG analysed and interpreted the sequence data. LK and RBG wrote the manuscript with inputs from other authors. All authors read and approved the final manuscript.

#### Funding

Research funding from the Australian Research Council (RBG et al.) and the Australian National Health and Medical Research Council (grant GTN1105448; MWL) is gratefully acknowledged.

#### Availability of data and materials

The data generated and analysed during the present study are available in the GenBank database under the accession no. MK774655. Raw data are available in the CNSA database under the accession no. CNP0000438.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup> Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia. <sup>2</sup> BGI Research, Shenzhen, Guangdong, China. <sup>3</sup> Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia. <sup>4</sup> School of Animal and Veterinary Sciences, Charles Sturt University, Wagga, Wagga, NSW, Australia.

Received: 28 March 2019 Accepted: 6 May 2019

Published online: 16 May 2019

#### References

- Romig T, Deplazes P, Jenkins D, Giraudoux P, Massolo A, Craig PS, et al. Ecology and life cycle patterns of *Echinococcus* species. *Adv Parasitol.* 2017;95:213–314.
- Deplazes P, Rinaldi L, Alvarez Rojas CA, Torgerson PR, Harandi MF, Romig T, et al. Global distribution of alveolar and cystic echinococcosis. *Adv Parasitol.* 2017;95:315–493.
- World Health Organization. Fourth WHO report on neglected tropical diseases. Integrating neglected tropical diseases into global health and development. Geneva: World Health Organization; 2017.
- Lymbery AJ. Phylogenetic pattern, evolutionary processes and species delimitation in the genus *Echinococcus*. *Adv Parasitol.* 2017;95:111–45.
- Kinkar L, Laurimäe T, Sharbatkhori M, Mirhendi H, Kia EB, Ponce-Gordo F, et al. New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm *Echinococcus granulosus sensu stricto*. *Infect Genet Evol.* 2017;52:52–8.
- Laurimäe T, Kinkar L, Moks E, Romig T, Omer RA, Casulli A, et al. Molecular phylogeny based on six nuclear genes suggests that *Echinococcus granulosus sensu lato* genotypes G6/G7 and G8/G10 can be regarded as two distinct species. *Parasitology.* 2018;145:1929–37.
- Laurimäe T, Kinkar L, Romig T, Omer RA, Casulli A, Umhang G, et al. The benefits of analysing complete mitochondrial genomes: deep insights into the phylogeny and population structure of *Echinococcus granulosus sensu lato* genotypes G6 and G7. *Infect Genet Evol.* 2018;64:85–94.
- Lymbery AJ, Jenkins EJ, Schurer JM, Thompson RCA. *Echinococcus canadensis*, *E. borealis*, and *E. intermedicus*. What's in a name? *Trends Parasitol.* 2015;31:23–9.
- Lymbery AJ, Jenkins EJ, Schurer JM, Thompson RCA. Response to Nakao et al. —Is *Echinococcus intermedicus* a valid species? *Trends Parasitol.* 2015;31:343–4.
- Nakao M, Lavikainen A, Hoberg E. Is *Echinococcus intermedicus* a valid species? *Trends Parasitol.* 2015;31:342–3.
- Alvarez Rojas CA, Romig T, Lightowlers MW. *Echinococcus granulosus sensu lato* genotypes infecting humans—review of current knowledge. *Int J Parasitol.* 2014;44:9–18.
- Bowles J, Blair D, McManus D. Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Mol Biochem Parasitol.* 1992;54:165–74.
- Bowles J, Blair D, McManus DP. Molecular genetic characterization of the cervid strain ('northern form') of *Echinococcus granulosus*. *Parasitology.* 1994;109:215–21.
- Lavikainen A, Lehtinen MJ, Meri T, Hirvelä-Koski V, Meri S. Molecular genetic characterization of the Fennoscandian cervid strain, a new genotypic group (G10) of *Echinococcus granulosus*. *Parasitology.* 2003;127:207–15.
- Wassermann M, Woldeyes D, Gerbi BM, Ebi D, Zeyhle E, Mackenstedt U, et al. A novel zoonotic genotype related to *Echinococcus granulosus sensu stricto* from southern Ethiopia. *Int J Parasitol.* 2016;46:663–8.
- Nakao M, Yanagida T, Konyaev S, Lavikainen A, Odnokurtsev VA, Zaikov VA, et al. Mitochondrial phylogeny of the genus *Echinococcus* (Cestoda: Taeniidae) with emphasis on relationships among *Echinococcus canadensis* genotypes. *Parasitology.* 2013;140:1625–36.
- Alvarez Rojas CA, Ebi D, Gauci CG, Scheerlinck JP, Wassermann M, Jenkins DJ, et al. Microdiversity of *Echinococcus granulosus sensu stricto* in Australia. *Parasitology.* 2016;143:1026–33.
- Moks E, Jögisalu I, Valdmann H, Saarma U. First report of *Echinococcus granulosus* G8 in Eurasia and a reappraisal of the phylogenetic relationships of 'genotypes' G5–G10. *Parasitology.* 2008;135:647–54.
- Casulli A, Interisano M, Sreter T, Chitima L, Kirkova Z, La Rosa G, et al. Genetic variability of *Echinococcus granulosus sensu stricto* in Europe inferred by mitochondrial DNA sequences. *Infect Genet Evol.* 2012;12:377–83.
- Yanagida T, Mohammadzadeh T, Kamhawi S, Nakao M, Sadjadi SM, Hijawi N, et al. Genetic polymorphisms of *Echinococcus granulosus sensu stricto* in the Middle East. *Parasitol Int.* 2012;61:599–603.

21. Le TH, Pearson MS, Blair D, Dai N, Zhang LH, McManus DP. Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*. *Parasitology*. 2002;124:97–112.
22. Tsai IJ, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496:57–63.
23. 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;48:729–42.
24. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13:278–89.
25. Korhonen PK, Hall RS, Young ND, Gasser BG. Common Workflow Language (CWL)-based software pipeline for *de novo* genome assembly from long- and short-read data. *GigaScience*. 2019;8:giz014.
26. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: A Laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 1989.
27. PacBio sequencing. <https://www.pacb.com>. Accessed 20 March 2019.
28. BGI Australia. <https://bgi-australia.com.au>. Accessed 20 March 2019.
29. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*. 2015;16:294.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
32. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.
33. Jex AR, Hall RS, Littlewood DTJ, Gasser RB. An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Res*. 2010;38:522–33.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
35. Genovese LM, Mosca MM, Pellegrini M, Geraci F. Dot2dot: accurate whole-genome tandem repeats discovery. *Bioinformatics*. 2019;35:914–22.
36. RNAfold web server. <http://rna.tbi.univie.ac.at>. Accessed 20 March 2019.
37. Nakao M, Yokoyama N, Sako Y, Fukunaga M, Ito A. The complete mitochondrial DNA sequence of the cestode *Echinococcus multilocularis* (Cyclophyllidae: Taeniidae). *Mitochondrion*. 2002;1:497–509.
38. Nakao M, McManus DP, Schantz PM, Craig PS, Ito A. A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. *Parasitology*. 2007;134:713–22.
39. Wang N, Xie Y, Liu T, Zhong X, Wang J, Hu D, et al. The complete mitochondrial genome of G3 genotype of *Echinococcus granulosus* (Cestoda: Taeniidae). *Mitochondrial DNA Part A*. 2016;27:1701–2.
40. 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:e1007513.
41. 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:giy146.
42. Le TH, Blair D, McManus DP. Mitochondrial genomes of parasitic flatworms. *Trends Parasitol*. 2002;18:206–13.
43. Le TH, Blair D, McManus DP. Complete DNA sequence and gene organization of the mitochondrial genome of the liverfluke, *Fasciola hepatica* L. (Platyhelminthes: Trematoda). *Parasitology*. 2001;123:609–21.
44. von Nickisch-Rosenegk M, Brown WM, Boore JL. Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: gene arrangements indicate that Platyhelminths are Eutrochozoans. *Mol Biol Evol*. 2001;18:721–30.
45. Huyse T, Buchmann K, Littlewood DTJ. The mitochondrial genome of *Gyrodactylus derjavinoideus* (Platyhelminthes: Monogenea)—a mitogenomic approach for *Gyrodactylus* species and strain identification. *Gene*. 2008;417:27–34.
46. 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:e0120081.
47. Guo A. The complete mitochondrial genome of *Anoplocephala perfoliata*, the first representative for the family Anoplocephalidae. *Parasit Vectors*. 2015;8:549.
48. Li WX, Zhang D, Boyce K, Xi BW, Zou H, Wu SG, et al. The complete mitochondrial DNA of three monozoic tapeworms in the Caryophyllidea: a mitogenomic perspective on the phylogeny of eucestodes. *Parasit Vectors*. 2017;10:314.
49. Xi B-W, Zhang D, Li W-X, Yang B-J, Xie J. Characterization of the complete mitochondrial genome of *Parabreviscolex niepini* (Cestoda, Caryophyllidea). *ZooKeys*. 2018;783:97–112.
50. Kim K-H, Jeon H-K, Kang S, Sultana T, Kim GJ, Eom K, et al. Characterization of the complete mitochondrial genome of *Diphyllobothrium nihonkaiense* (Diphyllobothriidae: Cestoda), and development of molecular markers for differentiating fish tapeworms. *Mol Cells*. 2007;23:379–90.
51. Yamasaki H, Ohmae H, Kuramochi T. Complete mitochondrial genomes of *Diplogonoporus balaenopterae* and *Diplogonoporus grandis* (Cestoda: Diphyllobothriidae) and clarification of their taxonomic relationships. *Parasitol Int*. 2012;61:260–6.
52. Yamasaki H, Izumiya S, Nozaki T. Complete sequence and characterization of the mitochondrial genome of *Diphyllobothrium stemmacephalum*, the type species of genus *Diphyllobothrium* (Cestoda: Diphyllobothriidae), using next generation sequencing. *Parasitol Int*. 2017;66:573–8.
53. Li WX, Fu PP, Zhang D, Boyce K, Xi BW, Zou H, et al. Comparative mitogenomics supports synonymy of the genera *Ligula* and *Digramma* (Cestoda: Diphyllobothriidae). *Parasit Vectors*. 2018;11:324.
54. Liu G-H, Lin R-Q, Li M-W, Liu W, Liu Y, Yuan Z-G, et al. The complete mitochondrial genomes of three cestode species of *Taenia* infecting animals and humans. *Mol Biol Rep*. 2011;38:2249–56.
55. Guo A. Characterization of the complete mitochondrial genome of the cloacal tapeworm *Cloacotaenia megalops* (Cestoda: Hymenolepididae). *Parasit Vectors*. 2016;9:490.
56. Kinkar L, Laurimäe T, Balkaya I, Casulli A, Zait H, Irshadullah M, et al. Genetic diversity and phylogeography of the elusive, but epidemiologically important *Echinococcus granulosus* sensu stricto genotype G3. *Parasitology*. 2018;145:1613–22.
57. Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*. 2008;101:301–20.
58. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res*. 1999;27:1767–80.
59. Taanman J-W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta Bioenerg*. 1999;1410:103–23.
60. 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;64:342–56.
61. Selosse M-A, Albert B, Godelle B. Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evol*. 2001;16:135–41.
62. Ciesielski GL, Oliveira MT, Kaguni LS. Animal mitochondrial DNA replication. *Enzymes*. 2016;39:255–92.
63. Falkenberg M. Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem*. 2018;62:287–96.
64. 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;22:620–9.
65. 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.
66. Eberhard JR, Wright TF, Bermingham E. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. *Mol Biol Evol*. 2001;18:1330–42.
67. Morris-Pocock JA, Taylor SA, Birt TP, Friesen VL. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Evol Biol*. 2010;10:14.
68. 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:e82854.
69. Kumazawa Y, Ota H, Nishida M, Ozawa T. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of

- control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Mol Biol Evol.* 1996;13:1242–54.
70. McManus DP, Smyth JD. Differences in the chemical composition and carbohydrate metabolism of *Echinococcus granulosus* (horse and sheep strains) and *E. multilocularis*. *Parasitology.* 1978;77:103–9.
  71. McManus DP, Smyth JD. Intermediary carbohydrate metabolism in protoscolecids of *Echinococcus granulosus* (horse and sheep strains) and *E. multilocularis*. *Parasitology.* 1982;84:351–66.
  72. Brehm K, Koziol U. *Echinococcus*-host interactions at cellular and molecular levels. *Adv Parasitol.* 2017;95:147–212.
  73. Mehlhorn H. *Encyclopedia of parasitology.* 3rd ed. Berlin, Heidelberg: Springer-Verlag; 2016. p. 478.
  74. Parkinson J, Wasmuth JD, Salinas G, Bizarro CV, Sanford C, Berriman M, et al. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl Trop Dis.* 2012;6:e1897.
  75. Berg JM, Tymoczko JL, Stryer L, Berg JM, Tymoczko JL, Stryer L. *Biochemistry.* 5th ed. New York: WH Freeman; 2002.
  76. Swiderski Z. *Echinococcus granulosus*: hook-muscle systems and cellular organisation of infective oncospheres. *Int J Parasitol.* 1983;13:289–99.
  77. Thompson RCA. Biology and systematics of *Echinococcus*. *Adv Parasitol.* 2017;95:65–109.
  78. Thompson RCA, Lymbery AJ. Biology and systematics of *Echinococcus*. *Echinococcus and hydatid disease.* Wallingford: CAB International; 1995.
  79. 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.
  80. Kuzminov A, Schabtach E, Stahl FW. Study of plasmid replication in *Escherichia coli* with a combination of 2D gel electrophoresis and electron microscopy. *J Mol Biol.* 1997;268:1–7.
  81. Dandjinou AT, Larrivée M, Wellinger RE, Wellinger RJ. Two-dimensional agarose gel analysis of DNA replication intermediates. *Methods Mol Biol.* 2006;313:193–208.
  82. Jöers P, Lewis SC, Fukuoh A, Parhiala M, Ellilä S, Holt IJ, et al. Mitochondrial transcription terminator family members mTTF and mTerF5 have opposing roles in coordination of mtDNA synthesis. *PLoS Genet.* 2013;9:e1003800.
  83. Le TH, Blair D, McManus DP. Mitochondrial genomes of human helminths and their use as markers in population genetics and phylogeny. *Acta Trop.* 2000;77:243–56.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

