Mitochondrial genome of *Bulinus truncatus* (Gastropoda: Lymnaeoida): Implications for snail systematics and schistosome epidemiology

Neil D. Young a,*, Liina Kinkar a, Andreas J. Stroehlein a, Pasi K. Korhonen a, J. Russell Stothard b, David Rollinson c,d, Robin B. Gasser a

a Melbourne Veterinary School, The University of Melbourne, Parkville, Victoria, Australia
b Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK
c, d Department of Life Sciences, Natural History Museum, London, UK

London Centre for Neglected Tropical Disease Research, London, UK

1. Introduction

Human schistosomiasis represents a group of neglected tropical diseases (NTDs) caused by blood flukes of the genus *Schistosoma* (schistosomes). More than 200 million people worldwide are affected by the hepato-intestinal or urogenital form of schistosomiasis. Notable species within the *S. haematobium*-group include *Schistosoma bovis, Schistosoma curassoni, Schistosoma mattheei, Schistosoma leiperi* and *Schistosoma margrebowiei* (pathogens infecting domestic and wild ruminants). No vaccines are available to prevent schistosome infections or disease, and the reliance on praziquantel – an anthelmintic drug which is not highly effective against all developmental stages of schistosomes (Gryseels et al., 2006) – for the mass treatment of humans against schistosomiasis, carries a significant risk of treatment failures, and the excessive treatment over time has a risk of resistance against this drug emerging in schistosomes (Greenberg, 2013). For these reasons, the effective prevention and control should rely on a sound understanding of the life-cycle and transmission of schistosomiasis for the implementation of integrated prevention/control, focused on the disruption of transmission to humans (World Health Organization, 2020).

As some species of freshwater snails of the genus *Bulinus* act as intermediate hosts in the life-cycles of schistosomes in Africa and adjacent regions. Currently, 37 species of *Bulinus* representing four groups are recognised. The mitochondrial cytochrome c oxidase subunit 1 (cox1) gene has shown utility for identifying and differentiating *Bulinus* species and groups, but taxonomic relationships based on genetic data are not entirely consistent with those inferred using morphological and biological features. To underpin future studies of members of the genus, we characterised here the mitochondrial genome of *Bulinus truncatus* (from a defined laboratory strain) using a combined second- and third-generation sequencing and informatics approach, enabling taxonomic comparisons with other planorbid snails for which mitochondrial (mt) genomes were available. Analyses showed consistency in gene order and length among mitochondrial genomes of representative planorbid snails, with the lowest and highest nucleotide diversities being in the cytochrome c oxidase and nicotinamide dehydrogenase subunit genes, respectively. This first mt genome for a representative of the genus *Bulinus* should provide a useful resource for future investigations of the systematics, population genetics, epidemiology and/or ecology of *Bulinus* and related snails. The sequencing and informatic workflow employed here should find broad applicability to a range of other snail intermediate hosts of parasitic trematodes.

**ARTICLE INFO**

**Keywords:** Mitochondrial genome *Bulinus truncatus* Snail intermediate host *Schistosoma*

**ABSTRACT**

Many freshwater snails of the genus *Bulinus* act as intermediate hosts in the life-cycles of schistosomes in Africa and adjacent regions. Currently, 37 species of *Bulinus* representing four groups are recognised. The mitochondrial cytochrome c oxidase subunit 1 (cox1) gene has shown utility for identifying and differentiating *Bulinus* species and groups, but taxonomic relationships based on genetic data are not entirely consistent with those inferred using morphological and biological features. To underpin future studies of members of the genus, we characterised here the mitochondrial genome of *Bulinus truncatus* (from a defined laboratory strain) using a combined second- and third-generation sequencing and informatics approach, enabling taxonomic comparisons with other planorbid snails for which mitochondrial (mt) genomes were available. Analyses showed consistency in gene order and length among mitochondrial genomes of representative planorbid snails, with the lowest and highest nucleotide diversities being in the cytochrome c oxidase and nicotinamide dehydrogenase subunit genes, respectively. This first mt genome for a representative of the genus *Bulinus* should provide a useful resource for future investigations of the systematics, population genetics, epidemiology and/or ecology of *Bulinus* and related snails. The sequencing and informatic workflow employed here should find broad applicability to a range of other snail intermediate hosts of parasitic trematodes.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2.1. Isolation and procurement of the snails

2. Materials and methods

As intermediate hosts of members of the

mitogenomic and genetic explorations of the snails that act

nected
coccus granulosus

accelerate future

(short-read) sequencing and advanced informatics, to de

2020), we here used this technology, combined with second-generation

long-range) sequencing technologies to de

improve the phylogenetic reconstruction of systematic relationships. A

the taxonomic classi

number of phylogenetically-informative genetic markers should enhance

the taxonomic classification of Bulinus species or subspecies, and could improve the phylogenetic reconstruction of systematic relationships. A logical extension to this published work is to sequence and annotate a complete mt genome to represent the genus Bulinus, in order to underpin future taxonomic/phylogenetic studies of members of the genus.

Given our recent success with the use of third-generation (long-read or long-range) sequencing technologies to define the complete mt genomes of a number of invertebrate species (parasitic flatworms), including Echinococcus granulosus (genotype G1) and Clonorchis sinensis (Kinker et al., 2019, 2020), we here used this technology, combined with second-generation (short-read) sequencing and advanced informatics, to define the mt genome of a first, key representative of the genus – Bu. truncatus – to accelerate future mitogenomic and genetic explorations of the snails that act as intermediate hosts of members of the S. haematobium group.

2. Materials and methods

2.1. Isolation and procurement of the snails

Samples of Bu. truncatus originated from a laboratory strain (designated ‘BRI’) which is maintained in the Biomedical Research Institute (BRI), Rockville, Maryland (Lewis et al., 2008); this strain was originally sourced from Egypt (Dr Margaret Mentink-Kane, personal communication, 10 October 2020). Individual snails were washed extensively in phosphate-buffered saline (PBS, pH 7.0) and frozen at –80°C.

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

High quality genomic DNA was isolated from two adult Bu. truncatus snails using the Circulomics Tissue Kit (Circulomics, Baltimore, MD, USA). The integrity of the DNA was assessed using an Agilent 4200 TapeStation system (ThermoFisher) and using Genomic DNA ScreenTape (ThermoFisher). Low molecular weight DNA was removed using a 10 kb short-read eliminator kit (Circulomics, Baltimore, MD, USA). The high molecular weight DNA was used to construct Nanopore Rapid Sequencing (SQK-RAD004; Oxford Nanopore Technologies) and Ligation Sequencing (SQK-LSK109; Oxford Nanopore Technologies) genomic DNA libraries, according to the manufacturer’s instructions. Each flow cell used to sequence the first libraries was washed using a Flow Cell Wash Kit (EXP-WSH003; Oxford Nanopore Technologies) and re-used to sequence additional SQK-LSK109 libraries. All libraries were sequenced using 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 (Cock et al., 2010). A short-insert (500 bp) genomic DNA library was also constructed and paired-end sequenced (150 bp reads) using TruSeq sequencing chemistry and the NovaSeq sequencing platform (Illumina).

2.3. Assembly and annotation of the mt genome

Long reads were assembled using FLYE v.2.6 (Kolmogorov et al., 2019) with the --nano-raw option and setting a genome size estimate of 900 megabases (Gregory, 2003). Errors in long-read sequence data were initially corrected using medaka consensus in the Medaka package v.0.10.0 (https://github.com/nanoporetech/medaka). The assembled genome was then polished with Pilon v.1.23 (Walker et al., 2014) using the short-read Illumina data. Finally, long- and short-reads were mapped to the assembled mt genome using Minimap2 v.2.0 (Li, 2018), and coverage of the genome was determined using mpileup in the SAMtools package v.1.9 (Li et al., 2009).

Annotation of tRNA, rRNA and protein-encoding gene regions was initially undertaken using MITOS webserver (Bernt et al., 2013) and employing the mt genetic code for invertebrates (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi; translation_table 5). The open reading frame (ORF) of each protein gene was then further curated in the program Geneious v.11.1.5 (Kearse et al., 2012) and using gene regions of published mt genomes of planorbid snails as a guide (Table 1). The complete mt genome sequence was deposited in the GenBank database under the accession no. MT947902; raw data are also available from the Sequence Read Archive (SRA) under the accession no. SAMN17050146 with the NCBI BioProject accession number PRJNA680620.

2.4. Single loci and whole genome comparative analyses

Available cox1 sequence data for Bulinus spp., selected taxa of the family Planorbidae and Radix auricularia (family Lymnaeidae; outgroup) were downloaded from NCBI (Supplementary Table S1) and aligned as nucleotide sequence using MUSCLE v.3.7 (Edgar, 2004). Sequence alignments were trimmed to the coding region of cox1 (164–1,098 of the

Table 1

Mitochondrial genomic sequences for snail species/subspecies used in the present study, with accession numbers and references listed

| Accession number | Species “strain”            | Length (bp) | G + C content (%) | Reference                  |
|------------------|----------------------------|-------------|-------------------|----------------------------|
| MT947902         | Bulinus truncatus          | 13,767      | 24.3              | Present study              |
| NC_026538        | Radix auricularis – ZUE strain | 13,745      | 28.6              | Feldmeyer et al. (2015)    |
| MT628577.1       | Gyraulus laevis            | 13,685      | 25.2              | Unpublished                |
| KY134384.1       | Planorbella duryi          | 14,217      | 27.3              | Schultz et al. (2018)      |
| MT628573.1       | Bathymphalus contortus     | 13,679      | 24.9              | Unpublished                |
| EP13576          | Biomphalaria tenagophila – Taim-RS strain | 13,722      | 24.2              | Unpublished                |
| MF480756         | Biomphalaria straminea voucher Xiaoping Wuxi-BS-2017 | 13,650      | 24.7              | Zhou et al. (2019)         |
| MG431962         | Biomphalaria pfeiffer      | 13,624      | 23.4              | Zhang et al. (2018)        |
| MG431964         | Biomphalaria chuanchinapha | 13,672      | 23.4              | Zhang et al. (2018)        |
| MG431963         | Biomphalaria sudanica      | 13,671      | 23.4              | Zhang et al. (2018)        |
| MG431966         | Biomphalaria glabrata – G16 BS90 strain | 13,676      | 24.8              | Zhang et al. (2018)        |
| MG431965         | Bi. glabrata – G72 M strain | 13,667      | 25.3              | Zhang et al. (2018)        |
| AY380531         | Bi. glabrata – 1742 strain | 13,670      | 25.4              | DeJong et al. (2004)       |
| AY380567         | Bi. glabrata – M strain    | 13,670      | 25.4              | DeJong et al. (2004)       |

* Member of the family Lymnaeidae – used as an outgroup in phylogenetic analyses.
1,542 bp *Bu. truncatus* cox1 gene) in order to minimise missing data from the 5' and 3' regions of previously sequenced amplicons (Supplementary Table S1). Identical nucleotide sequences were removed using CD-HIT-EST v.4.6 (Fu et al., 2012). The optimal nucleotide substitution model for aligned sequences was then assessed using the program ModelTest-NG v.0.1.6 (Darriba et al., 2020). The aligned sequences were then subjected to phylogenetic analysis using Bayesian inference (BI) employing Monte Carlo Markov Chain analysis in the program MrBayes v.3.2.2 (Ronquist et al., 2012). Posterior probabilities (pp) were calculated using the GTR + I + G model, generating 1,000,000 trees and sampling every 200th tree until potential scale reduction factors for each parameter approached one. The initial 25% of trees were discarded as "burn-in", and the others were used to construct a majority rule tree. Phylogenetic trees were rendered and annotated using ggtree v.1.10.5 (Yu et al., 2017) in R v.3.4.3 (http://www.R-project.org/).

The complete mt genome of *Bu. truncatus* was compared to the available reference mt genomes of other planorbid snails (Table 1) and *R. auricularia* (family Lymnaeidae; outgroup) using progressiveMaue v.2.4.0 (Darling et al., 2004), and using the following settings: --hmm-identity = 0.95 --island-gap-size = 10. Mitochondrial protein-coding genes were then extracted and aligned as individual nucleotide coding or inferred amino acid sequences using MUSCLE. The optimal substitution model for each aligned sequence was then assessed using the program ModelTest-NG v.0.1.6 (Darriba et al., 2020). The aligned sequences were then subjected to phylogenetic analysis using BI or maximum likelihood (ML) methods employing Monte Carlo Markov Chain analysis in the program MrBayes and RaxML v.8.2.12 (Stamatakis, 2014). For BI, posterior probabilities (pp) were calculated using the selected substitution model (Nucleotide sequence: GTR + I + G; Protein sequence: MtMam for ATP6, COX1-2, CYTB, NAD1-2 and MtRev for COX3 and NAD3-6), generating 1,000,000 trees and sampling every 200th tree until potential scale reduction factors for each parameter approached one. The initial 25% of trees were discarded as "burn-in", and the others were used to construct a majority rule tree. Phylogenetic trees were rendered and annotated using ggtree v.1.10.5 (Yu et al., 2017) in R v.3.4.3 (http://www.R-project.org/).

Patterns of nucleotide diversity were compared between the aligned mt protein-coding regions of *Bu. truncatus* and protein-coding regions of available reference mt genomes of planorbid snails and *R. auricularia* (family Lymnaeidae) as an outgroup (Table 1). A sliding window analysis of nucleotide diversity (steps of 10 bp over 200-bp windows) was performed for each pairwise-alignment of concatenated genes using the PopGenome package (Peifer et al., 2014) in R. For each comparison, nucleotide diversity values were plotted using the R package ggplot2 (Wickham, 2009).

3. Results and discussion

3.1. Characteristics of the mitochondrial genome of *Bulinus truncatus*

From a total of 1,327 long- and 917,758 short-reads (equating to ~139 Mb), a complete, 13,767 bp circular mt genome was assembled (GenBank: MT947902). The mean coverage of long-reads mapped to the genome was 1.36 (range: 0.53–199; Fig. 1). The mean coverage of mapped short-reads was 8,902 (range: 3,892–9,730), with the depth being ~4,000 reads at the introduced start/stop positions (Fig. 1). We identified 37 genes, including 13 protein-coding (adenosine triphosphatase

### Table 2

| Gene designations | Location (start/stop) | GC (%) | Length (bp) | Transcription direction | Start/stop codons |
|-------------------|-----------------------|--------|-------------|-------------------------|-------------------|
| Protein-coding genes |                       |        |             |                         |                   |
| apr6              | 5284/5940             | 22.4   | 789         | Reverse                 | AAT/TAA          |
| apr8              | 5128/5235             | 18.3   | 108         | Reverse                 | ATT/TTA          |
| cox1              | 10517/12058           | 29.6   | 1542        | Forward                 | ATT/TTA          |
| cox2              | 4114/4785             | 27.7   | 729         | Forward                 | ATT/TTA          |
| cox3              | 8666/9463             | 28.8   | 882         | Reverse                 | ATG/TTA          |
| cytB              | 2838/3292             | 27.3   | 1107        | Forward                 | TGG/TTA          |
| nad1              | 1647/2540             | 26.2   | 894         | Forward                 | ATT/TTA          |
| nad2              | 9570/10514            | 23.0   | 1044        | Forward                 | ATT/TTA          |
| nad3              | 6819/7184             | 18.6   | 378         | Reverse                 | ATT/TTA          |
| nad4              | 7308/8612             | 22.3   | 1404        | Forward                 | ATT/TTA          |
| nadM              | 2566/2862             | 21.2   | 327         | Forward                 | ATG/TTA          |
| nadS              | 61312/13767           | 18.4   | 456         | Forward                 | ATT/TTA          |
| Ribosomal RNA genes |                     |        |             |                         |                   |
| rrnS              | 6064/6776             | 23.6   | 713         | Reverse                 | na                |
| rrnL              | 12104/13153           | 21.4   | 1093        | Forward                 | na                |
| Transfer RNA genes |                     |        |             |                         |                   |
| tRNA-G(ggc)       | 3925/3991             | 67     | Forward     | na                      |                   |
| tRNA-G(gca)       | 3987/4048             | 62     | Forward     | na                      |                   |
| tRNA-G(gau)       | 4051/4112             | 62     | Forward     | na                      |                   |
| tRNA-Y(gca)       | 4768/4826             | 59     | Forward     | na                      |                   |
| tRNA-W(cca)       | 4823/4885             | 63     | Forward     | na                      |                   |
| tRNA-G(cca)       | 4885/4950             | 66     | Forward     | na                      |                   |
| tRNA-H(gca)       | 4942/5003             | 62     | Forward     | na                      |                   |
| tRNA-Q(taa)       | 5006/5066             | 61     | Reverse     | na                      |                   |
| tRNA-L2(taa)      | 5056/5125             | 70     | Reverse     | na                      |                   |
| tRNA-L5(tac)      | 5236/5303             | 68     | Reverse     | na                      |                   |
| tRNA-R(tatc)      | 5944/6005             | 62     | Reverse     | na                      |                   |
| tRNA-E(tta)       | 6001/6064             | 64     | Reverse     | na                      |                   |
| tRNA-M(cat)       | 6775/6838             | 64     | Reverse     | na                      |                   |
| tRNA-S2(tgca)     | 7185/7240             | 56     | Reverse     | na                      |                   |
| tRNA-S1(tgc)      | 7242/7299             | 58     | Forward     | na                      |                   |
| tRNA-T5(tgg)      | 6821/6855             | 65     | Reverse     | na                      |                   |
| tRNA-L2(tac)      | 9504/9570             | 67     | Reverse     | na                      |                   |
| tRNA-V(tac)       | 10474/10537           | 64     | Forward     | na                      |                   |
| tRNA-L1(tag)      | 12062/12125           | 64     | Forward     | na                      |                   |
| tRNA-A(tag)       | 13191/13254           | 64     | Forward     | na                      |                   |
| tRNA-N(tgg)       | 13255/13315           | 61     | Forward     | na                      |                   |

Abbreviation: na, not applicable.

* The locations of start and stop codons were identified based on the alignment of protein-coding regions, inferred using published gene models for planorbid snails (Feldmeyer et al., 2015; Schultz et al., 2018; Zhang et al., 2018).
subunit 6 (atp6), adenosine triphosphatase subunit 8 (atp8), the cytochrome c oxidase subunits 1, 2 and 3 (cox1–cox3), cytochrome b (cytb) and the nicotinamide dehydrogenase subunit 1–6 (nad1–nad6 and nad4l), 22 transfer RNA (two coding for leucine and two coding for serine)) and the small (rrnS) and large (rrnL) subunits of ribosomal RNA genes (Fig. 1 and Table 2). In most instances, start and stop codons were consistent with those of the mt genomes of other invertebrates characterised to date (Lavrov, 2014).

The mt genomic sequence of *Bu. truncatus* is biased toward A + T (75.7%), with T (41.7%) being the most frequent nucleotide, and C (10.9%) the least frequent one, in accord with mt genomes of some other planorbid snails including *Bathyomphalus contortus*, *Biomphalaria* spp., *Gyrulus laevis* and *Planorbello duryi* (Table 1). A + T content was highest in the genes *atp6*, *nad3* and *nad6* (81.4–81.7%) and lowest in the genes *cox1*, *cox2* and *cox3* (70.4–72.3%) (Table 2). The tRNAs (Table 2) were inferred to have a canonical structure (Supplementary Figure S1). Two copies of a serine and a leucine tRNA were encoded, and all tRNAs were predicted to have DHU and TψC arms, except for tRNA-G(tcc) (without a TψC arm), tRNA-S1(gct) (without a DHU arm) and tRNA-S2(tga) (without a DHU arm).

### 3.2. Verifying the taxonomic status of *Bu. truncatus*

As the molecular taxonomy of snails relies on the use of *cox1* as the mt genetic marker (Stothard et al., 2002; Pernance et al., 2020), we used extensive, publicly-available (partial) *cox1* sequence data sets in a phylogenetic analysis to verify the position of the BRI strain of *Bu. truncatus* in relation to previously characterised *Bulinus* groups and taxa (Fig. 2). There was strong nodal support for the present strain grouping with a large number of other *Bu. truncatus* specimens of the *Bu. truncatus/tropicus* complex, to the exclusion of members of the three other groups (*Bu. reticulatus*, *Bu. forskalli* and *Bu. africana*), although there was limited support for grouping with other *Bu. truncatus* specimens from Egypt (Fig. 2). The BRI strain of *Bu. truncatus* clustered most closely with isolates KM272998, MG407308, MG407310, MG407312, MG407326 and MG759391, all of which originated from Egypt (Supplementary Table S1; Fig. 2B). The phylogenetic tree constructed using partial *cox1* sequence data also revealed limited nodal support for some nodes within some recognised *Bulinus* groups.

### 3.3. Mt genomic comparisons

Having established the characteristics of the mt genome of *Bu. truncatus* and verified the taxonomic status of our BRI strain, we established levels of nucleotide variability along the whole mt genome between *Bu. truncatus*, other planorbid snails (*Bu. contortus*, *Biomphalaria* spp., *G. laevis* and *P. duryi*) using *R. auricularia* as a distant taxon (outgroup) (Fig. 3). The comparison of the aligned genome sequences revealed 38 conserved, co-linear blocks with ≥95% nucleotide identity (comprising 8,562 of the 13,506 aligned nucleotide positions). There was no observed difference in gene order or direction of transcription between *Bu. truncatus* and other planorbid snails studied here; *cox1* (1,563 bp), *cox3* (1,106 bp) and *cytb* (860 bp) were the longest protein-encoding genes (Table 2). A guide-tree, inferred from a whole genome alignment (Fig. 3, left), indicated that *Bu. truncatus* grouped with *G. laevis* to the exclusion of *Biomphalaria* species and other planorbid (*Bu. contortus* and *P. duryi*). An assessment of nucleotide diversity in protein-coding genes between *Bu. truncatus* and other snail species upon pairwise comparison (Fig. 4) revealed low (0.153–0.244) and high (0.215–0.391) mean diversities in the *cox* and *nad* genes, respectively. Across all protein-coding genes of *Bu. truncatus*, most diversity was seen with the lymnaeid *R. auricularia* (mean nucleotide diversity = 0.304; Fig. 4).

### 3.4. Relationship of *Bu. truncatus* with other planorbid snails

The phylogenetic/systematic relationship of *Bu. truncatus* with other planorbid species was constructed using complete data sets publicly available for mt protein-coding genes (Table 1; Fig. 5). Trees constructed
using nucleotide or amino acid sequence data provided strong support (posterior probability = 1.00; bootstrap support = 100%) for most nodes, with *Bu. truncatus* being basal. Using nucleotide sequence data, support was limited (pp = 0.82; bss = 70%) only at nodes grouping *Bi. glabrata* 1742 with *Bi. glabrata* M-line as well as *Bi. straminea* with some other species, such as *Bi. tenagophila* (pp = 0.89; bss < 50%). Using amino acid sequence data, *Bu. truncatus* grouped with *Ba. contortus* and *G. laevis* (pp = 0.99; bss = 73%), with *P. duryi* being basal. The basal position of *Bu. truncatus* in the tree (Fig. 5) agrees with a proposed ancestral position of *Bulinus* and *Indoplanorbis* (Albrecht et al., 2007) and supports the previous proposals that *Bulinus* belong to the family Bulinidae (Albrecht et al., 2007; Bouchet et al., 2017). The characterisation of the mt genomes of other species of *Bulinus* and *Indoplanorbis* should help resolve the taxonomic ambiguities within the superfamily Lymnaeoidea.

### 3.5. Implications and future work

Here, we characterised the first complete mt genome representing the genus *Bulinus* – a complex of at least 37 species presently divided into four main ‘groups’ (Brown, 1994). This mt genome for the BRI strain of *Bu. truncatus* was assembled using a combination of second- (short-read) and third-generation (long-read) sequence data, enabling taxonomic verification (using cox1 sequence data) as well as comparative analyses with other planorbid snails for which mt genomes were known.

The taxonomic investigation using cox1 sequence data showed that the BRI strain of *Bu. truncatus* clustered with six samples from Egypt, including those from locations in Tanta (KM272998), Mansouria (MG407308, MG407310 and MG407312), Behera (MG407312) and Giza (MG795391) (Zein-Eddine et al., 2014; Abe et al., 2018). While an exact phyleogeographic location was not determined, the clustering of the BRI strain with *Bulinus* *cox* partial evidence provided in previous studies (Zein-Eddine et al., 2014; Abe et al., 2018; Tumwebaze et al., 2019) – although identification and differentiation should be achieved if cox1 sequences of at least 1 kb are employed (Kane et al., 2008). The molecular characterisation of *Bulinus* species has usually relied on the use of single mt and/or nuclear (e.g. ITS-2) genetic markers (cf. Kane et al., 2008; Chibwana et al., 2020; Pennance et al., 2020). In most cases, identification or differentiation of members of the four major *Bulinus* groups was achieved (Kane et al., 2008; Zein-Eddine et al., 2014; Pennance et al., 2020), but, in some instances, incongruences were observed due to limited signal (saturated mutational variation) or nodal support (Stothard et al., 2001; Jorgensen et al., 2013; Zein-Eddine et al., 2014; Abe et al., 2018; Tumwebaze et al., 2019).

Defining the first complete reference mt genome for *Bu. truncatus* (BRI strain) using short- and long-read data sets paves the way to characterising such genomes for the 37 known members of the genus *Bulinus* and then, using complete protein-coding gene data sets, to test the validity of the four currently-recognised/-proposed taxonomic groups (*“Bu. truncatus/tropicus”* complex, “*Bu. reticulatus*”, “*Bu. forskalii*” and “*Bu. africanaus*”). In the future, it will be important to conduct mt genome studies of type (museum) specimens of *Bulinus* species using refined methods for the isolation and sequencing of ancient DNA (e.g. Spyrou et al., 2019). Using all of these mt genomic data sets, it should be possible to define genetic markers for distinct OTUs as a basis to conduct molecular epidemiological investigations of the prevalence and geographical distribution of particular *Bulinus* taxa involved in the transmission of urogenital schistosomiasis, explorations of OTU-schistosome affiliations as well as large-scale population genetic studies of the snail vectors of the S. haematobium-group (cf. Rollinson et al., 2001). These efforts would assist in providing an informed position regarding the control of urogenital schistosomiasis and the feasibility of eliminating urogenital schistosomiasis in endemic regions in the context of mass drug administration regimens (Knopp et al., 2012, 2019; Allan et al., 2020).
The relatively conserved gene order in the mt genomes and the phylogenetic relationships of the snails studied here (Figs. 3 and 5) show the feasibility of rapidly characterising the mt genomes of a range of members of the family Planorbidae – some of which are vectors of socioeconomically important parasitic trematodes other than schistosomes – using the technological approach established here. Such an effort could assist significantly in closing some of the knowledge gaps that exist in the understanding of systematics of these groups, and would allow some insights into molecular evolution and codon usage (cf. Clary & Wolstenholme, 1985; Batuecas et al., 1988; Hoffmann et al., 1992; Boore & Brown, 1994). Complementing such an effort, it would be imperative to start to sequence the nuclear genomes of respective snail vectors of key parasitic trematodes.

Given the prime importance of *Bi. glabrata* as a vector of *Schistosoma mansoni*, draft nuclear genomes have been published for this species (Adema et al., 2017; Lu et al., 2020), although there have been challenges associated with karyotyping (Goldman et al., 1984) and ploidy determination (Wheeler et al., 2018) for this snail. Nonetheless, through major advances in sequencing and informatics, a recent study (Kenny et al., 2020) has succeeded in yielding a chromosome-level assembly of a large and complex genome of the mollusc *Pecten maximus* – great scallop (family Pectinidae) – which bodes well for the application of a similar workflow to sequence the genomes of molluscan vectors of trematodes. As a key intermediate host for *S. haematobium*, *Bu. truncatus* would be an ideal candidate for future nuclear genome characterisation and comparative genomic explorations. We anticipate that the ploidy/karyotype(s) of *Bulinus* (Wu, 1972) will present unique challenges for genome assembly efforts, but, if successful, the availability of both mt and nuclear genomes for this snail species would represent major resources for studies of its ecology, population genetics, origin and evolution as well as the epidemiology of urogenital schistosomiasis and snail-host-parasite affiliation(s).

In conclusion, the present study emphasises the relevance of genomic work on *Bu. truncatus* and other key snail vectors of parasitic trematodes, as a basis for future systematic, population genetic, epidemiological, ecological and biological investigations. The sequencing and bioinformatic platform established here provides exciting prospects for large-scale comparative and population genetic studies of such snail vectors.
vectors. Fundamental insights into snail genetics and biology could assist in better understanding the transmission patterns of trematodiases and in finding ways of blocking disease transmission to humans and other vertebrates.

Funding

This research project was supported by grants from the National Health and Medical Research Council (NHMRC) and the Australian Research Council (ARC) to RBG, NDY and PKK.

CRediT author statement

Neil D. Young: Methodology, Writing - Original Draft, Conceptualisation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. Liina Kinkar: Methodology, Writing - Review & Editing. Andreas J. Stroehlein: Methodology, Writing - Review & Editing. Pasi K. Korhonen: Writing - Review & Editing, Funding acquisition. David Rollinson: Writing - Review & Editing. J. Russell Stothard: Writing - Review & Editing. Robin B. Gasser: Conceptualisation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. All authors read and approved the final manuscript.

Data availability

The nucleotide sequence of the mitochondrial (mt) genome reported in this article is publicly available in the GenBank database under accession no. MT947902. Raw data are also available from the Sequence Read Archive (SRA) under the accession no. SAMN17050146 with the NCBI BioProject accession number PRJNA680620. All other data analysed in this study are referred to in this article and its supplementary files.

Declaration of competing interests

The authors declare that they have no competing interests.

Fig. 5 Phylogenetic relationship of *Bulinus truncatus* with selected species of planorbid snails and *Radix auricularia* (family Lymnaeidae) as an outgroup (cf. Table 1). A Phylogeny inferred from concatenated nucleotide sequences derived from 12 mitochondrial protein-encoding genes using Bayesian inference (BI) or maximum likelihood (ML). B Phylogeny inferred from concatenated amino acid sequences derived from 12 mitochondrial protein-encoding genes using BI or ML. Posterior probability (pp; BI) or percentage bootstrap (bs; ML) support values are indicated at each node of the tree (BI/ML). The scale-bar indicates phylogenetic distance (in substitutions per site).

Acknowledgements

*Bulinus truncatus* snails were kindly provided by Dr Margaret Mentink-Kane of the NIH-NIAID Schistosomiasis Resource Center, Biomedical Resource Institute, Rockville, MD 20850, USA.

Abbreviations

| Abbreviation | Description                        |
|--------------|------------------------------------|
| *atp*        | adenosine triphosphatase subunit   |
| *bp*         | base pair                          |
| *BI*         | Bayesian inference                 |
| *BRI*        | Biomedical Research Institute      |
| *bss*        | bootstrap support                  |
| *cox*        | cytochrome c oxidase subunit       |
| *cyb*        | cytochrome b                        |
| *ITS*        | internal transcribed spacer         |
| *LR*         | long-read                          |
| *MEE*        | multilocus enzyme electrophoresis  |
| *ML*         | maximum likelihood                 |
| *mt*         | mitochondrial                       |
| *pp*         | posterior probability               |
| *nad*        | nicotinamide dehydrogenase subunit |
| *nt*         | nucleotides                        |
| *NTD*        | neglected tropical disease         |
| *ORF*        | open reading frame                 |
| *OTU*        | operation taxonomic units          |
| *PBS*        | phosphate-buffered saline          |
| *rnl*        | small subunit of ribosomal RNA gene|
| *rns*        | small subunit of ribosomal RNA gene|
| *SR*         | short-read                         |
| *tRNA*       | transfer RNA                        |

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crpvbd.2021.100017.
References

Abe, E.M., Guo, Y.H., Shen, H., Muttsaka-Makuvaza, M.J., Habib, M.R., Xue, J.B., et al., 2018. Phylogeny of the genus Biomphalaria (Gastropoda: Planorbidae) in selected African countries. Trop. Med. Infect. Dis. 3, 1–11.

Adema, C.M., Hillier, L.W., Jones, C.S., Leker, E.S., Knight, M., Minx, P., et al., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat. Genet. 49, 1545–1551.

Albrecht, C., Kuhn, K., Streit, B., 2007. A molecular phylogeny of Planorboidae (Gastropoda, Pulmonata): Insights from enhanced taxon sampling. Zool. Scr. 36, 27–39.

Allan, F., Ame, S.M., Tan-Bi, Y.T., Hofkin, B.V., Webster, B.L., Diakite, N.R., et al., 2020. Small related contributions from the schistosomiasis consortium for operational research and evaluation program including xenomonitoring, focal mollusciciding, biological control, and monitoring. Am. J. Trop. Med. Hyg. 103, 66–79.

Batuecas, B., Garesse, R., Calleja, M., Valverde, J.R., Marco, R., 1988. Genome organization of Actinocyclus (Mollusca: Pulmonata) nuclear DNA. Nucleic Acids Res. 16, 6515–6529.

Bernet, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

Boore, J.L., Brown, W.M., 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata. Genetics 138, 423–443.

Bouchet, P., Rocroi, J.-C., Hausdorf, B., Kaim, A., Kano, Y., Nübel, A., et al., 2017. Revised classification, nomenclator and typification of gastropod and monoplacophoran families. Malacologia 51, 1–526.

Brown, D.S., 1994. Freshwater snails of Africa and their medical importance. CRC Press, London.

Chibwana, F.D., Tumwebaze, I., Mahulu, A., Sands, A.F., Albrecht, C., 2020. Assessing the black chiton, Mitra ficata, nomenclator and typifiation of the genus Mitra (Gastropoda: Mitridae) in selected African countries. Trop. Med. Infect. Dis. 3, 1.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: Multiple alignment of genome sequences. Genome Res. 14, 1394–1403.

DeJong, R.J., Emery, A.M., Adema, C.M., 2004. The mitochondrial genome of Schistosoma mansoni. J. Exp. Biol. 207, 2541–2550.

DeJong, R.J., Emery, A.M., Adema, C.M., 2004. The mitochondrial genome of Biomphalaria glabrata (Gastropoda: Basommatophora), intermediate host of the schistosome Schistosoma mansoni. Mol. Biochem. Parasitol. 90, 991–997.

Edgar, R.C., 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Feldmeyer, B., Grehake, B., Punke, E., Ebersberger, I., Flenner, M., et al., 2015. Positive selection in development and growth rate regulation genes involved in species divergence of the genus Radix. Biodivers. Conserv. 24, 1551–1566.

Goto, N., Heuer, M.L., Rice, P.M., 2010. The Sanger FASTQ format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38, 1767–1771.

Greenberg, R.M., 2013. New approaches for understanding mechanisms of drug resistance in parasites. PLoS Pathog. 9, e1003673.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al., 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kane, R.A., Stothard, J.R., Emery, A.M., Rollinson, D., 2008. Molecular characterization of schistosome-hosts of the schistosoma haematobium group. Parasitology 135, S235–S247.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al., 2012. Revisiting classification, nomenclator and typification of gastropod and monoplacophoran families. Malacologia 51, 1–526.

Liu, L., Loker, E.S., Zhang, S.M., Buddenbrock, S.K., Bo, L., 2020. Genome-wide discovery, and computational and transcriptional characterization of an AIG gene family in the freshwater snail Biomphalaria glabrata, a vector for Schistosoma mansoni. BMC Genomics 21, 190.

Penncence, T., Allan, F., Emery, A., Rabone, M., Cable, J., Garba, A.D., et al., 2020. Interactions between Schistosoma haematobium group species and their Bulinus spp. intermediate hosts along the Niger River basin. Parasit. Vectors 13, 268.

Pfeifer, U., Wittelsburger, U., Ramsen-Ostvin, S.E., Lercher, M.J., 2014. PopGenome: An efficient Swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31, 1929–1936.

Raahauge, P., Kristensen, T.K., 2000. A comparison of Bulinus africanus group species (Planorbidae; Gastropoda) by use of the internal transcribed spacer 1 region combined by morphological and anatomical characters. Acta Trop. 85, 75–84.

Rollinson, D., Southgate, V.R., 1979. Enzyme analyses of Bulinus africanus group snails (Mollusca: Planorbidae) from Tanzania. Trans. R. Soc. Trop. Med. Hyg. 73, 667–672.

Rollinson, D., Stothard, J.R., Southgate, V.R., 2001. Interactions between intermediate snail hosts of the genus Bulinus and schistosomiasis of the Schistosoma haematobium group. Parasitology 123 (Suppl.), S245–S260.

Rongquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Holma, S., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice for large models. Syst. Biol. 61, 539–542.

Schultz, J.H., Bambah, L.M., Brem, J.A., Sellin, B., Sellin, E., Rollinson, D., 2001. Bulinus species on Madagascar: ecological evolution, genetic markers and compatibility with Schistosoma haematobium. Parasitology 123 (Suppl.), S261–S275.

Stothard, J.R., Loker, E.S., Southgate, V.R., 2001. Identification of snails within the Bulinus africanus group from East Africa by multiplex SNPShot trade mark analysis of single nucleotide polymorphisms in the cytochrome oxidase subunit I gene. Mem. Inst. Oswaldo Cruz 97 (Suppl. 1), 31–36.

Thompson, J.J., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963.

Wright, L.G., Dittrich, M., Thomas, J., Manduley, L., Fregert, S., 2000. Genetic and ecological aspects of the schistosome-hosts of the schistosome haematobium group parasites. Parasit. Vectors 12, 565.

Wright, M.R., Allen, R., 2001. PhyloPLOT2 elegant graphics for data analysis. Use R! Springer.

Wright, M.R., 2001. Odyssey - New york online resource (viii, 212 p.).

World Health Organization, 2020. Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030. World Health Organization, Geneva. www.who.int/neglected_diseases/resources/who-ucn-mid-2020-01/en/.

Yu, S.K., 1972. Comparative studies on a polyploid series of the African snail genus Bulinus (Bassommatophora: Planorbidae). Malacol. Rev. 5, 95–164.

Zhang, S.M., Bu, L., Laidemitt, M.R., Lu, L., Mkoji, G.M., et al., 2018. Genome-wide discovery, and computational and transcriptional characterization of an AIG gene family in the freshwater snail Biomphalaria glabrata, a vector for Schistosoma mansoni. BMC Genomics 19, 1.

Zhou, Y., Liu, X., Wu, R., Zhou, C., Ouyang, S., Wu, X., et al., 2019. The complete mitochondrial genome of invasive species Biomphalaria straminea (Planorbidae: Gastropoda) and phylogenetic analysis. Mitochondrial DNA Part B 4, 937–938.

Zhou, Y., Liu, X., Wu, R., Zhou, C., Ouyang, S., Wu, X., et al., 2019. The complete mitochondrial genome of invasive species Biomphalaria straminea (Planorbidae: Gastropoda) and phylogenetic analysis. Mitochondrial DNA Part B 4, 937–938. 10.1080/23802359.2019.1579060.