Journal Pre-proof

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

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

PII: S2667-114X(21)00011-X

DOI: https://doi.org/10.1016/j.crpvbd.2021.100017

Reference: CRPVBD 100017

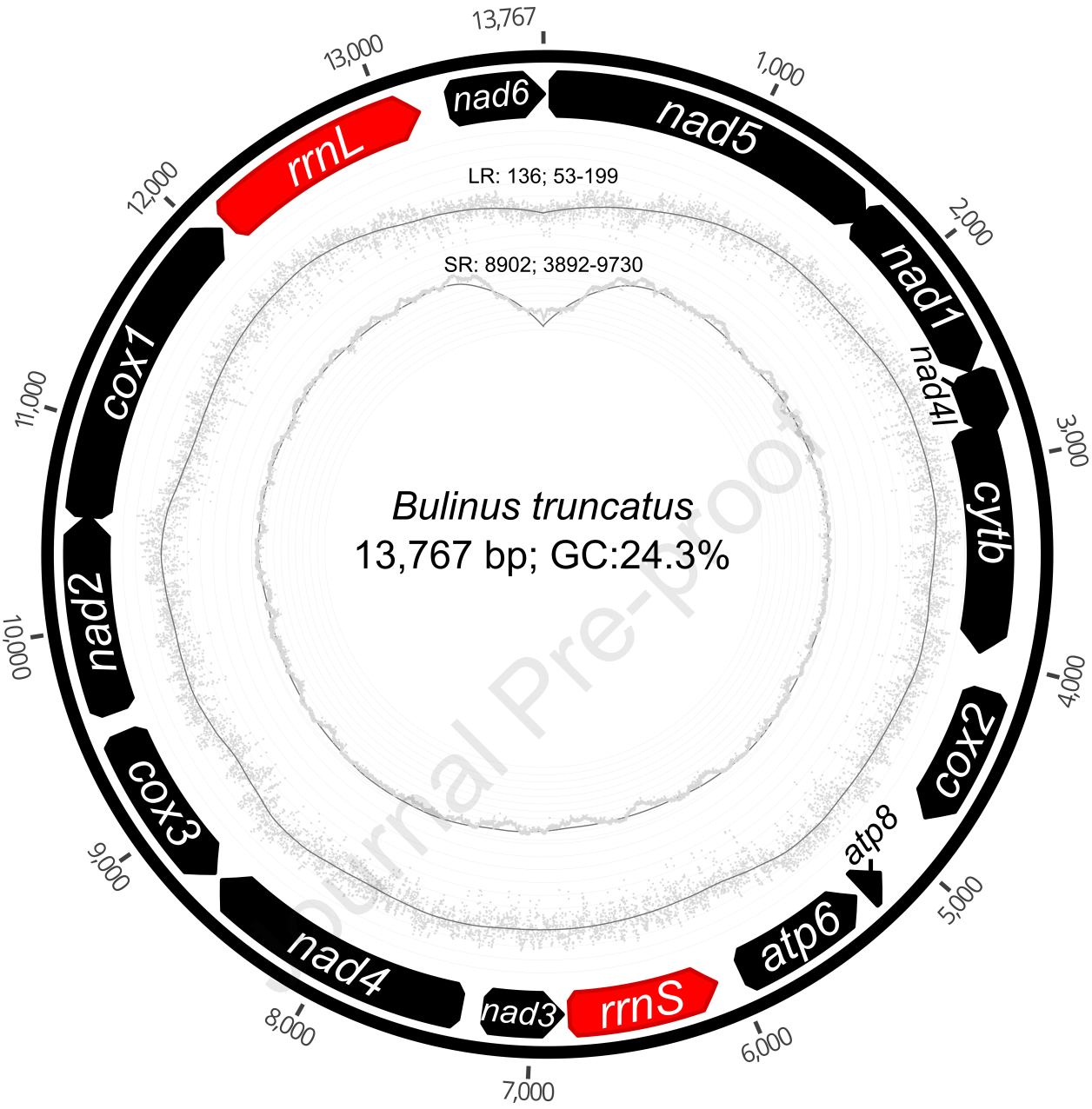
To appear in: *Current Research in Parasitology and Vector-Borne Diseases*

Received Date: 21 January 2021

Revised Date: 10 February 2021

Accepted Date: 11 February 2021

Please cite this article as: Young ND, Kinkar L, Stroehlein AJ, Korhonen PK, Stothard JR, Rollinson D, Gasser RB, Mitochondrial genome of *Bulinus truncatus* (Gastropoda: Lymnaeoidea): implications for snail systematics and schistosome epidemiology, *CORTEX*, https://doi.org/10.1016/j.crpvbd.2021.100017.



Mitochondrial genome of *Bulinus truncatus* (Gastropoda: Lymnaeoidea): 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] *Department of Life Sciences, Natural History Museum, London, UK*

[d] *London Centre for Neglected Tropical Disease Research, London, UK*

*Corresponding author at: Melbourne Veterinary School, The University of Melbourne, Parkville, Victoria, Australia.

E-mail address: nyoung@unimelb.edu.au (N.D. Young).

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 (*cox*1) 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 systematic 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 vectors of parasitic trematodes.

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

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 *Schistosoma haematobium*-group include *S. haematobium*, *S. intercalatum* and *S. guineensis* (human pathogens), and *S. bovis*, *S. curassoni*, *S. mattheei*, *S. leiperi* and *S. 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, Madagascar and adjacent regions, knowing which species of *Bulinus* transmits which species of *Schistosoma* is crucial. Therefore, the accurate identification of species of *Bulinus* is central to elucidating and understanding the epidemiology of schistosomiasis. Currently, 37 species of *Bulinus* represent four "groups": the "*Bu. truncatus*/*tropicus*" complex, "*Bu. reticulatus*", "*Bu. forskalii*" and "*Bu. africanus*" (Brown, 1994; Rollinson et al., 2001). Traditionally, members of these groups (or operational taxonomic units, OTUs) have been identified using morphological characters (Brown, 1994; Southgate et al., 2000), but specific identification can be unreliable due to a limited number of informative phenotypic characters (e.g. shell morphology) (Stothard et al., 1997). Nonetheless, karyotyping, multilocus enzyme electrophoresis (MEE) and DNA-based methods, particularly those based on PCR, have assisted in discerning species or taxa (Rollinson & Southgate, 1979; Raahauge & Kristensen, 2000; Zein-Eddine et al., 2014). For both fundamental and applied studies, there is a clear need for molecular tools for the accurate assignment of snails to the *Bulinus* groups, but, more importantly, for the unequivocal identification and differentiation of species within each of the four recognised groups.

Considerable work (Stothard et al., 2002) has been done in this direction, with a particular focus on defining, assessing and/or using genetic markers in nuclear ribosomal and in mitochondrial (mt) DNAs for specific identification, particularly the second internal transcribed spacer (ITS-2) and cytochrome *c* oxidase subunit 1 gene (*cox*1), respectively (Stothard et al., 2002; Jorgensen et al., 2013). Indeed, mt *cox*1 sequences have shown major utility for the identification of species and groups, and to establish the genetic/systematic relationships of species/taxa within the genus *Bulinus* (Tumwebaze et al., 2019; Chibwana et al., 2020), although, in some instances, there has been limited statistical support for some relationships (Kane et al., 2008). The latter finding indicates that the use of a greater 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* (Kinkar 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 ºC.

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 *cox*1 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 *cox*1 (164 to 1,098 of the 1,542 bp *Bu. truncatus cox*1 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 progressiveMauve 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. For ML the same sequence set was used to construct a partitioned ML tree using the selected models for

each partition and performing 20 iterations. Nodal support values for the ML tree were inferred from 100 bootstrap replicates. Phylogenetic trees were rendered and annotated using *ggtree* in R.

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 (Pfeifer 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 Bulunus 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 136 (range: 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 subunit 6 (*atp*6), adenosine triphosphatase subunit 8 (*atp*8), the cytochrome *c* oxidase subunits 1, 2 and 3 (*cox*1–*cox*3), cytochrome b (*cyt*b) and the nicotinamide dehydrogenase subunit 1–6 (*nad*1–*nad*6 and *nad*4*l*), 22 transfer RNA (two coding for leucine and two coding for serine) and the small (*rrn*S) and large (*rrn*L) 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., *Gyraulus laevis* and *Planorbella duryi* (Table 1). A+T content was highest in the genes *atp*8, *nad*3 and *nad*6 (81.4–81.7%) and lowest in the genes *cox*1, *cox*2 and *cox*3 (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 Bulinus truncatus

As the molecular taxonomy of snails relies on the use of *cox*1 as the mt genetic marker (Stothard et al., 2002; Pennance et al., 2020), we used extensive, publicly-available (partial) *cox*1 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. forskalii* and *Bu. africanus*), 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 *cox*1 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 (*Ba. 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 8,582 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; *cox*1 (1,563 bp), *cox*3 (1,106 bp) and *cyt*b (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 planorbids (*Ba. 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 Bulinus truncatus *with other planorbids*

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 *cox*1 sequence data) as well as comparative analyses with other planorbid snails for which mt genomes were known.

The taxonomic investigation using *cox*1 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 (MG759391) (Zein-Eddine et al., 2014; Abe et al., 2018). While an exact phylogeographic location was not determined, the clustering of the BRI strain with these *Bu. truncatus* isolates is consistent with information that this laboratory strain was originally collected from Egypt (Dr Margaret Mentink-Kane, BRI, personal communication, 10 October 2020). The partial *cox*1 sequence data sets publicly available were not sufficiently phylogenetically informative to yield statistically well-supported trees/phylogeographic relationships (Fig. 2) – a finding that is consistent with 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 *cox*1 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 (sequence 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. africanus*"). 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 members 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 *S. 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 *B. 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. 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.

Abbreviations

atp: adenosine triphosphatase subunit; bp: base pair; BI: Bayesian inference; BRI: Biomedical Research Institute; bss: bootstrap support; *cox*: cytochrome *c* oxidase subunit; *cyt*b: 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: nucelotides; NTD: neglected tropical disease; ORF: open reading frame; OTU: operation taxonomic units; PBS: phosphate-buffered saline; *rrn*L: small subunit of ribosomal RNA gene; *rrn*S: small subunit of ribosomal RNA gene; SR: short-read; tRNA: transfer RNA.

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.

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.

References

Abe, E.M., Guo, Y.H., Shen, H., Mutsaka-Makuvaza, M.J., Habib, M.R., Xue, J.B., et al., 2018.

Phylogeography of *Bulinus truncatus* (Audouin, 1827) (Gastropoda: Planorbidae) in selected African countries. Trop. Med. Infect. Dis. 3, 1-11.

Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., Minx, P., et al., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat. Commun. 8, 15451.

Akinwale, O.P., Kane, R.A., Rollinson, D., Stothard, J.R., Ajayi, M.B., Akande, D.O., et al., 2011. Molecular approaches to the identification of *Bulinus* species in south-west Nigeria and observations on natural snail infections with schistosomes. J. Helminthol. 85, 283-293.

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

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

Batuecas, B., Garesse, R., Calleja, M., Valverde, J.R., Marco, R., 1988. Genome organization of *Artemia* mitochondrial DNA. Nucleic Acids Res. 16, 6515-6529.

Bernt, 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.-P., Hausdorf, B., Kaim, A., Kano, Y., Nützel, A., et al., 2017. Revised classification, nomenclator and typification of gastropod and monoplacophoran families. Malacologia 61, 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 diversity and distribution of potential intermediate hosts snails for urogenital schistosomiasis: *Bulinus* spp. (Gastropoda: Planorbidae) of Lake Victoria. Parasit. Vectors 13, 418.

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252-271.

Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M., 2010. The Sanger FASTQ file format

for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38, 1767-1771.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394-1403.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 37, 291-294.

DeJong, R.J., Emery, A.M., Adema, C.M., 2004. The mitochondrial genome of *Biomphalaria glabrata* (Gastropoda: Basommatophora), intermediate host of *Schistosoma mansoni*. J. 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., Greshake, B., Funke, E., Ebersberger, I., Pfenninger, M., 2015. Positive selection in development and growth rate regulation genes involved in species divergence of the genus *Radix*. BMC Evol. Biol. 15, 164.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150-3152.

Goldman, M.A., Loverde, P.T., Chrisman, C.L., Franklin, D.A., 1984. Chromosomal evolution in planorbid snails of the genera *Bulinus* and *Biomphalaria*. Malacologia 25, 427-446.

Grabner, D.S., Mohamed, F.A., Nachev, M., Meabed, E.M., Sabry, A.H., Sures, B., 2014. Invasion biology meets parasitology: a case study of parasite spill-back with Egyptian *Fasciola gigantica* in the invasive snail *Pseudosuccinea columella*. PLoS ONE 9, e88537.

Greenberg, R.M., 2013. New approaches for understanding mechanisms of drug resistance in schistosomes. Parasitology 140, 1534-1546.

Gregory, T.R., 2003. Genome size estimates for two important freshwater molluscs, the zebra mussel (*Dreissena polymorpha*) and the schistosomiasis vector snail (*Biomphalaria glabrata*). Genome 46, 841-844.

Gryseels, B., Polman, K., Clerinx, J., Kestens, L., 2006. Human schistosomiasis. Lancet 368, 1106-1118.

Hoffmann, R.J., Boore, J.L., Brown, W.M., 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics 131, 397-412.

Jorgensen, A., Stothard, J.R., Madsen, H., Nalugwa, A., Nyakaana, S., Rollinson, D., 2013. The ITS2 of the genus *Bulinus*: novel secondary structure among freshwater snails and potential new taxonomic markers. Acta Trop. 128, 218-225.

Kane, R.A., Stothard, J.R., Emery, A.M., Rollinson, D., 2008. Molecular characterization of freshwater snails in the genus *Bulinus*: a role for barcodes? Parasit. Vectors 1, 15.

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.

Kenny, N.J., McCarthy, S.A., Dudchenko, O., James, K., Betteridge, E., Corton, C., et al., 2020. The gene-rich genome of the scallop *Pecten maximus*. Gigascience 9.

Kinkar, L., Korhonen, P.K., Cai, H., Gauci, C.G., Lightowlers, M.W., Saarma, U., et al., 2019. Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. Parasit. Vectors 12, 238.

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 Negl. Trop. Dis. 14, e0008552.

Knopp, S., Mohammed, K.A., Ali, S.M., Khamis, I.S., Ame, S.M., Albonico, M., et al., 2012. Study and implementation of urogenital schistosomiasis elimination in Zanzibar (Unguja and Pemba islands) using an integrated multidisciplinary approach. BMC Public Health 12, 930.

Knopp, S., Person, B., Ame, S.M., Ali, S.M., Hattendorf, J., Juma, S., et al., 2019. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. Lancet Glob. Health 7, e1118-e1129.

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540-546.

Lavrov, D.V., 2014. Mitochondrial Genomes in Invertebrate Animals. In: Bell, E (Ed.), Molecular Life Sciences: An Encyclopedic Reference. Springer New York, New York, NY, pp. 1-8.

Lewis, F.A., Liang, Y.S., Raghavan, N., Knight, M., 2008. The NIH-NIAID schistosomiasis resource center. PLoS Negl. Trop. Dis. 2, e267.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100.

Lu, L., Loker, E.S., Zhang, S.M., Buddenborg, S.K., Bu, 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.

Martínez-Ortí, A., Bargues, M., Mas-Coma, S., 2015. Dos nuevas localizaciones para España de *Bulinus truncatus* (Audouin, 1827) (Gastropoda, Planorbidae), huésped intermediario de la esquistosomiasis urinaria. Arx. Misc. Zool. 13, 25-31.

Mutsaka-Makuvaza, M.J., Zhou, X.N., Tshuma, C., Abe, E., Manasa, J., Manyangadze, T., et al., 2020. Molecular diversity of *Bulinus* species in Madziwa area, Shamva district in Zimbabwe: implications for urogenital schistosomiasis transmission. Parasit. Vectors 13, 14.

Nalugwa, A., Jorgensen, A., Nyakaana, S., Kristensen, T.K., 2010a. Molecular phylogeny of *Bulinus* (Gastropoda: Planorbidae) reveals the presence of three species complexes in the Albertine Rift freshwater bodies. Int. J. Genet. Mol. Biol. 2, 130-139.

Nalugwa, A., Kristensen, T.K., Nyakaana, S., Jorgensen, A., 2010b. Mitochondrial DNA variations in sibling species of the *Bulinus truncatus*/*tropicus* complex in Lake Albert, Western Uganda. Zool. Stud. 49, 515-522.

Pennance, 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 Valley. Parasit. Vectors 13, 268.

Pfeifer, B., Wittelsburger, U., Ramos-Onsins, 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. 75, 85-94.

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 schistosomes of the *Schistosoma haematobium* group. Parasitology 123 Suppl, S245-260.

Schultz, J.H., Bansbach, L.M., Bremmer, J.A., Dimmler, K.E., Forde, Q.A., Gagliano, E.M., et al., 2018. The mitochondrial genome of the planorbid snail *Planorbella duryi*. Mitochondrial DNA B 3, 972-973.

Southgate, V.R., de Clercq, D., Sene, M., Rollinson, D., Ly A., Vercruysse, J., 2000. Observations on the compatibility between *Bulinus* spp. and *Schistosoma haematobium* in the Senegal River basin. Ann. Trop. Med. Parasitol. 94, 157-164.

Spyrou, M.A., Bos, K.I., Herbig, A., Krause, J., 2019. Ancient pathogen genomics as an

emerging tool for infectious disease research. Nat. Rev. Genet. 20, 323-340.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stothard, J.R., Mgeni, A.F., Alawi, K.S., Savioli, L., Rollinson, D., 1997. Observations on shell morphology, enzymes and random amplified polymorphic DNA (RAPD) in *Bulinus africanus* group snails (Gastropoda: Planorbidae) in Zanzibar. J. Molluscan Studies 63, 489-503.

Stothard, J.R., Bremond, P., Andriamaro, L., Sellin, B., Sellin, E., Rollinson, D., 2001. *Bulinus* species on Madagascar: molecular evolution, genetic markers and compatibility with *Schistosoma haematobium*. Parasitology 123 Suppl, S261-275.

Stothard, J.R., Llewellyn-Hughes, J., Griffin, C.E., Hubbard, S.J., Kristensen, T.K., Rollinson, D., 2002. Identification of snails within the *Bulinus africanus* group from East Africa by multiplex SNaPshot trade mark analysis of single nucleotide polymorphisms within the cytochrome oxidase subunit I. Mem. Inst. Oswaldo Cruz 97 Suppl 1, 31-36.

Tumwebaze, I., Clewing, C., Dusabe, M.C., Tumusiime, J., Kagoro-Rugunda, G., Hammoud, C., et al., 2019. Molecular identification of *Bulinus* spp. intermediate host snails of *Schistosoma* spp. in crater lakes of western Uganda with implications for the transmission of the *Schistosoma haematobium* group parasites. Parasit. Vectors 12, 565.

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.

Wheeler, N.J., Dinguirard, N., Marquez, J., Gonzalez, A., Zamanian, M., Yoshino, T.P., et al., 2018. Sequence and structural variation in the genome of the *Biomphalaria glabrata* embryonic (Bge) cell line. Parasit. Vectors 11, 496.

Wickham, H., 2009. GGPLOT2 elegant graphics for data analysis, Use R! Springer, Dordrecht; New York, pp. 1 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. www.who.int/neglected_diseases/resources/who-ucn-ntd-2020.01/en/. World Health Organization, Geneva.

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

Yu, G.C., Smith, D.K., Zhu, H.C., Guan, Y., Lam, T.T.Y., 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28-36.

Zein-Eddine, R., Djuikwo-Teukeng, F.F., Al-Jawhari, M., Senghor, B., Huyse, T., Dreyfuss, G., 2014. Phylogeny of seven *Bulinus* species originating from endemic areas in three African countries, in relation to the human blood fluke *Schistosoma haematobium*. BMC Evol. Biol. 14, 271.

Zhang, S.M., Bu, L., Laidemitt, M.R., Lu, L., Mutuku, M.W., Mkoji, G.M., et al., 2018. Complete mitochondrial and rDNA complex sequences of important vector species of *Biomphalaria*, obligatory hosts of the human-infecting blood fluke, *Schistosoma mansoni*. Sci. Rep. 8, 7341.

FIGURE LEGENDS

Fig. 1 Reference mitochondrial genome of *Bulinus truncatus*. Direction of gene transcription is shown with an arrow. Long (*rrn*L) and short (*rrn*S) ribosomal RNA subunits are shown in red and protein-encoding genes are shown in black. The graph shows the depth of individual nucleotides of individual long- (LR) and short-read (SR) reads at individual positions in the mt genome (grey dots); indicated is also the mean nucleotide depth across this genome (solid, dark-grey line)

Fig. 2 A. Phylogenetic relationships of *Bulinus truncatus* with other representatives of this species and other (selected) species of planorbid snails (Supplementary Table S1) inferred based on an analysis of the aligned partial *cox*1 gene sequences by Bayesian inference using *Radix auricularia* (family Lymnaeidae) as an outgroup. The outermost black line indicates recognised groups of *Bu. truncatus* (see Brown, 1994). **B.** Part of the phylogenetic tree includes a group comprising *Bu. truncatus* from the Biomedical Research Institute [BRI] (Lewis et al., 2008) – sequenced herein – and other representatives from Egypt (EGY), Sudan (SUD) and Zimbabwe (ZWE). Posterior probability (pp) support is given at each node of the tree and relates to a heatmap (red: pp = 1; black: pp = 0.5). The partial *cox*1 gene sequence of the BRI strain of *Bu. truncatus* sequenced here is denoted in bold-type

Fig. 3 Alignment of the *Bulinus truncatus* mt genome with other (selected) species of planorbid snails and an outgroup – *Radix auricularia* (family Lymnaeidae) (cf. Table 1). Co-linear blocks of nucleotides $\geq$ 95% identity (maximum of 10 nucleotide gaps allowed within block) are shown. The guide-tree represents the genetic distance between genomes based on shared gene content among aligned genome pairs. The scale-bar indicates estimated evolutionary distance

Fig. 4 Nucleotide diversity in coding genes between *Bulinus truncatus* and other available planorbid snails as well as *Radix auricularia* (outgroup) upon pairwise comparison (cf. Table 1). Each solid line represents the average nucleotide diversity across 200, 10-base overlapping nucleotide windows. All species of *Biomphalaria* are represented in one plot, with each species/strain represented by a distinct grey line. Average nucleotide diversity across all coding regions is indicated by a horizontal, dashed line. Average nucleotide diversities for each gene region are indicated

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)

Table 1

Mitochondrial genomic sequences for snail species/strains 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[a]	*Radix auricularia* – ZUE strain	13,745	28.6	Feldmeyer et al. (2015)
MT628577.1	*Gyraulus laevis*	13,685	25.2	Unpublished
KY514384.1	*Planorbella duryi*	14,217	27.3	Schultz et al. (2018)
MT628573.1	*Bathyomphalus contortus*	13,679	24.9	Unpublished
EF433576	*Biomphalaria tenagophila* – Taim-RS strain	13,722	24.2	Unpublished
MF480756	*Biomphalaria straminea* voucher Xiaoping Wu:04-BS-2017	13,650	24.7	Zhang et al. (2018)
NC_038059	*Biomphalaria pfeifferi*	13,624	18.5	Zhang et al. (2018)
NC_038061	*Biomphalaria choanomphala*	13,672	23.4	Zhang et al. (2018)
NC_038060	*Biomphalaria sudanica*	13,671	18.0	Zhang et al. (2018)
MG431966	*Biomphalaria glabrata* – G16 BS90 strain	13,676	17.9	Zhang et al. (2018)
MG431965	*Bi. glabrata* – G72 M strain	13,667	19.3	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)

[a] Member of the family Lymnaeidae – used as an outgroup in phylogenetic analyses.

Table 2

Mitochondrial genes of *Bulinus truncatus* and their locations, GC contents, lengths, direction of transcription and start/stop codons

Gene designations	Location (start/end)	GC (%)	Length (bp)	Transcription direction	Start/stop codons
Protein-coding genes					
*atp*6	5284/5940	22.4	789	Reverse	AAT/TAA[a]
*atp*8	5128/5235	18.3	108	Reverse	ATT/TAA
*cox*1	10517/12058	29.6	1542	Forward	ATT/TAA
*cox*2	4114/4785	27.7	729	Forward	ATT/TAG
*cox*3	8666/9463	28.8	882	Reverse	ATG/TAG
cytb	2828/3924	27.3	1107	Forward	TTG[a] /TTA[a]
*nad*1	1647/2540	26.2	894	Forward	ATG/TAA
*nad*2	9570/10514	23.0	1044	Forward	ATG/TAA
*nad*3	6819/7184	18.6	378	Reverse	ATA/TAA
*nad*4	7308/8612	22.3	1404	Forward	ATA/TAA
*nad*4*l*	2566/2862	21.2	327	Forward	ATA/TAG
*nad*5	14/1663	22.8	1650	Forward	ATA/TAG
*nad*6	13312/13767	18.4	456	Forward	TTG/TAA
Ribosomal RNA genes					
rrnS	6064/6776	23.6	713	Reverse	na
rrnL	12104/13153	21.4	1093	Forward	na
Transfer RNA genes					
tRNA-D(gtc)	3925/3991		67	Forward	na
tRNA-C(gca)	3987/4048		62	Forward	na
tRNA-F(gaa)	4051/4112		62	Forward	na
tRNA-Y(gta)	4768/4826		59	Forward	na
tRNA-W(tca)	4823/4885		63	Forward	na
tRNA-G(tcc)	4885/4950		66	Forward	na
tRNA-H(gtg)	4942/5003		62	Forward	na
tRNA-Q(ttg)	5006/5066		61	Reverse	na
tRNA-L2(taa)	5056/5125		70	Reverse	na
tRNA-N(gtt)	5236/5303		68	Reverse	na
tRNA-R(tcg)	5944/6005		62	Reverse	na
tRNA-E(ttc)	6001/6064		64	Reverse	na
tRNA-M(cat)	6775/6838		64	Reverse	na
tRNA-S2(tga)	7185/7240		56	Reverse	na
tRNA-S1(gct)	7242/7299		58	Forward	na
tRNA-T(tgt)	8621/8685		65	Reverse	na
tRNA-I(gat)	9504/9570		67	Forward	na
tRNA-K(ttt)	10474/10537		64	Forward	na
tRNA-V(tac)	12062/12125		64	Forward	na
tRNA-L1(tag)	13130/13194		65	Forward	na
tRNA-A(tgc)	13191/13254		64	Forward	na
tRNA-P(tgg)	13255/13315		61	Forward	na

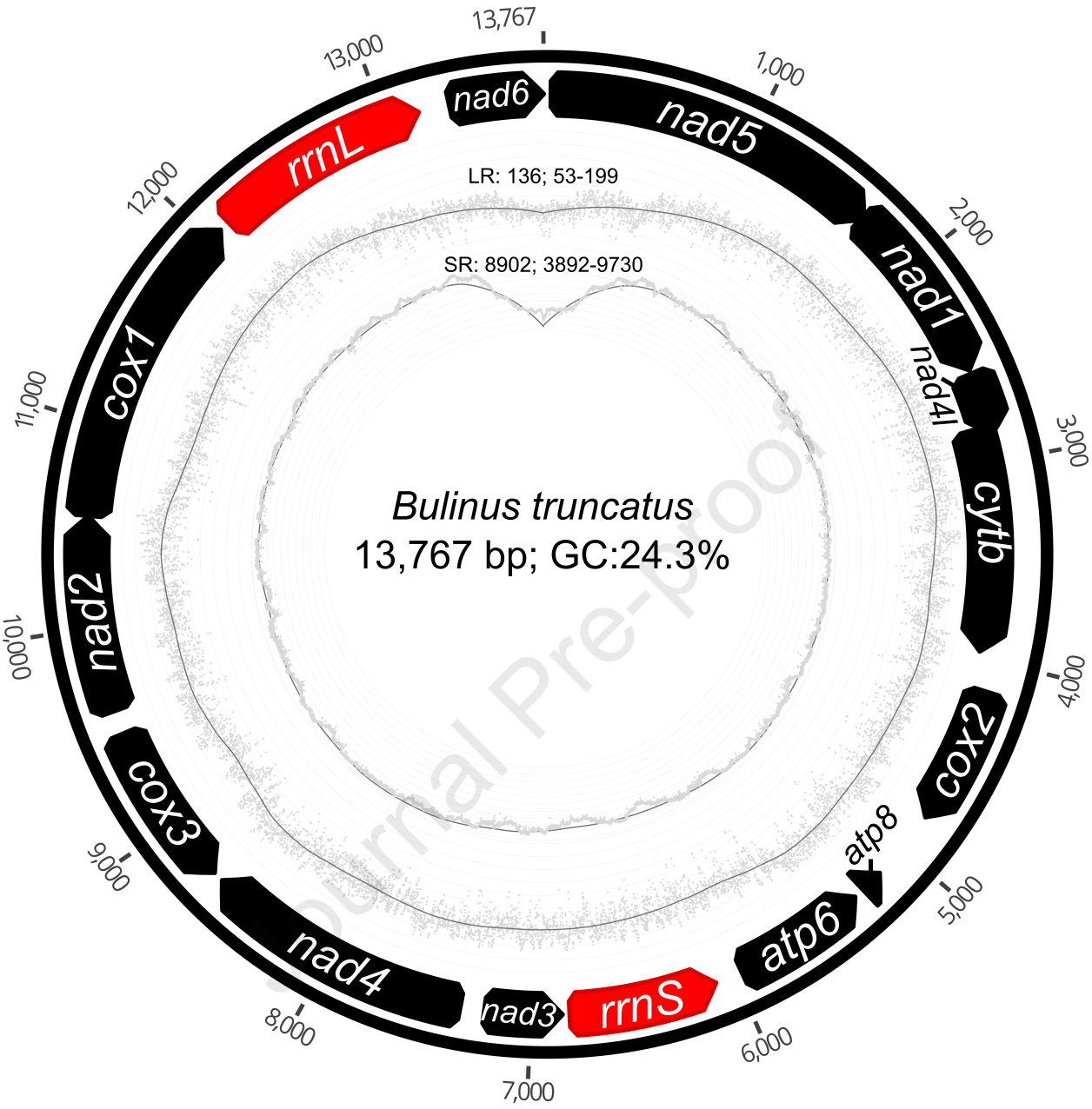
[a] The locations of start and stop codons were identified based on the alignment of protein-coding regions, inferred using published gene models for planorbid snails (Feldmeyer et al., 2015; Schultz et al., 2018; Zhang et al., 2018).

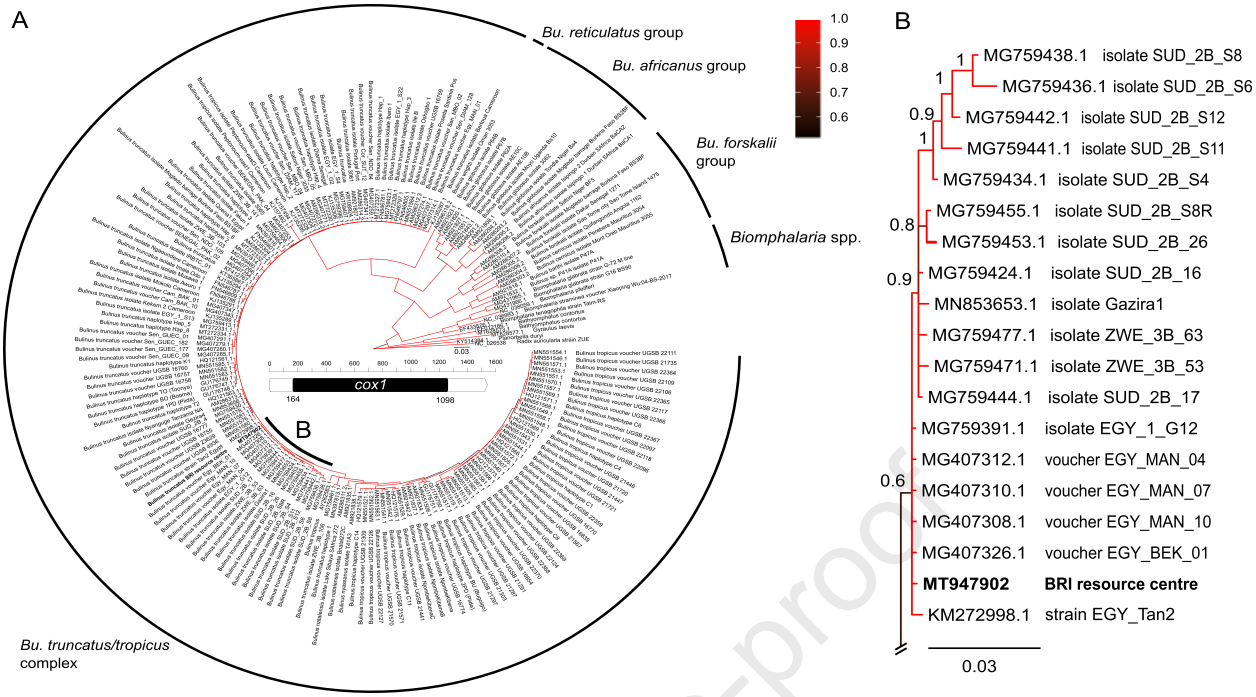
Abbreviation: na, not applicable.

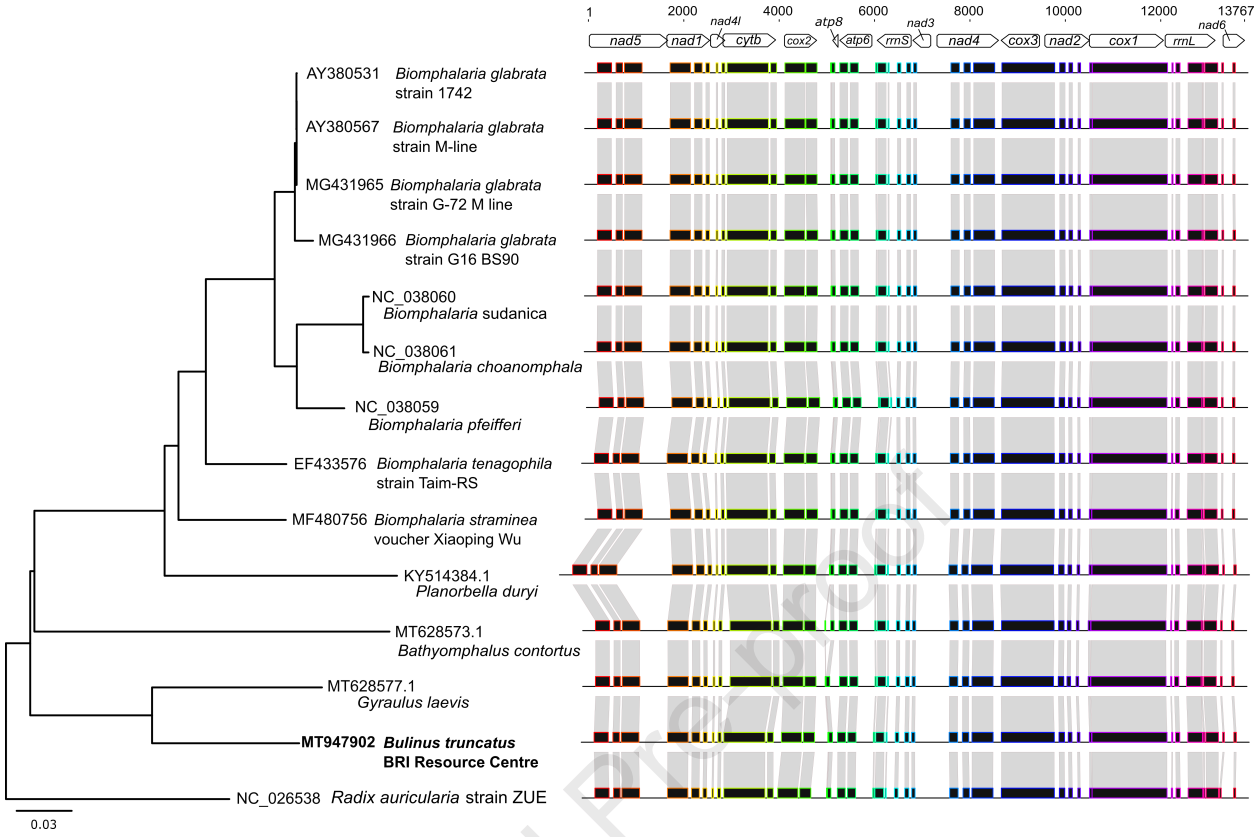
Supplementary material

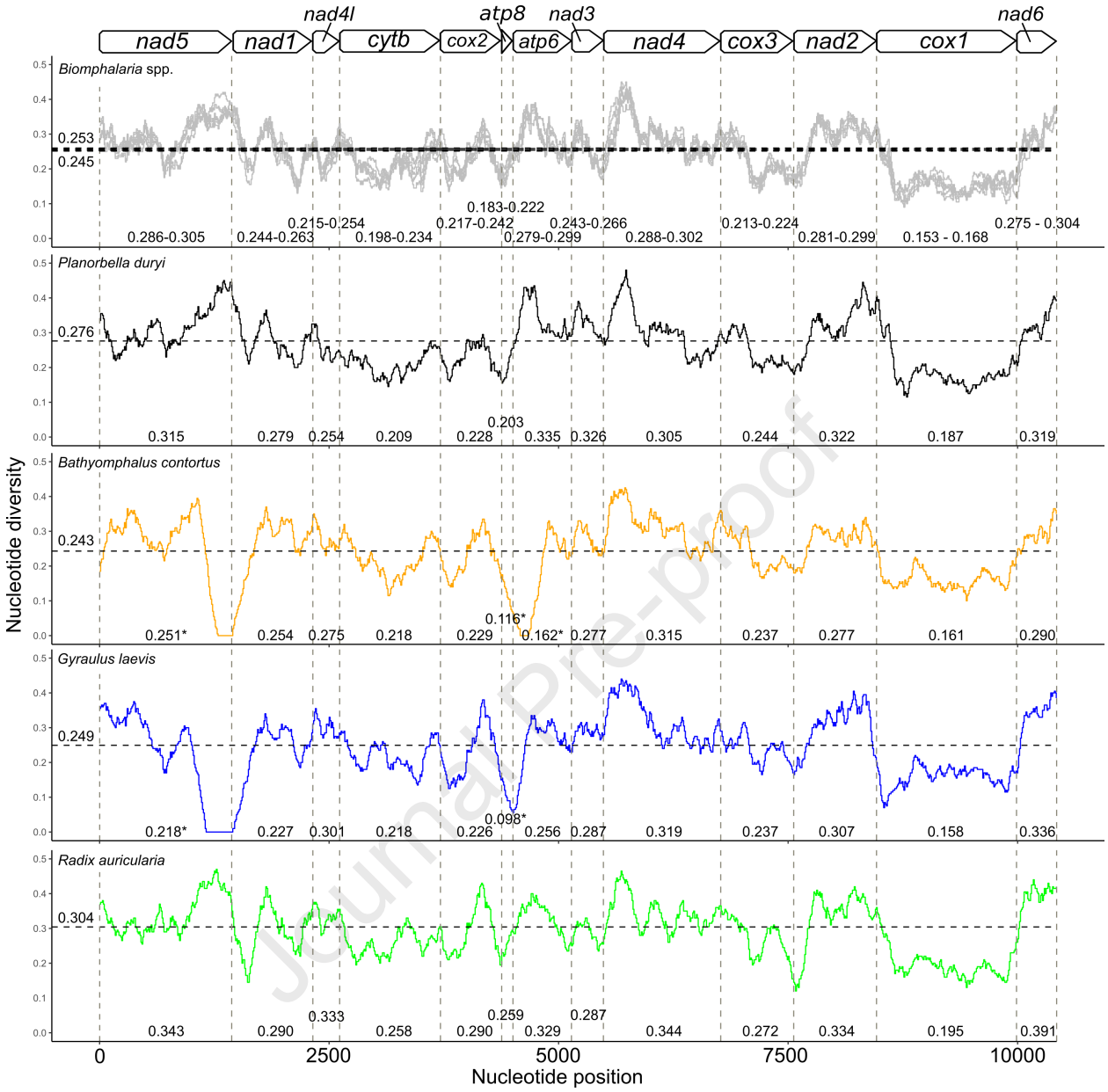
Supplementary Figure S1. Secondary structures predicted for the 22 tRNA genes encoded in the mitochondrial genome of *Bulinus truncatus*. The locations of individual tRNA genes are indicated.

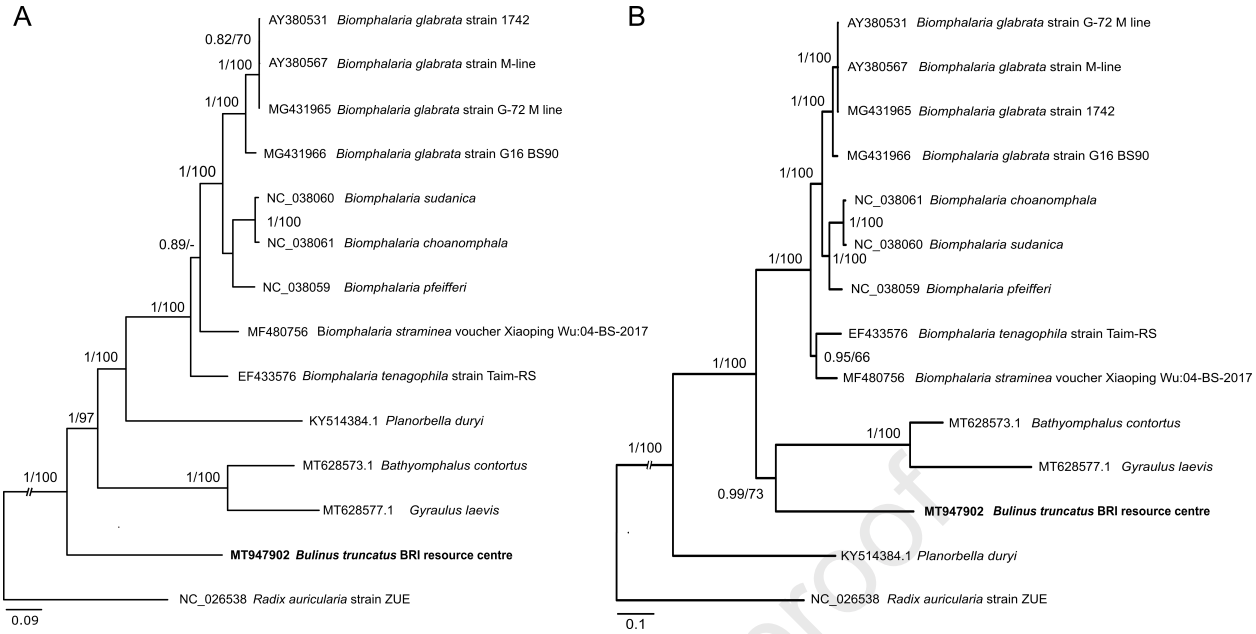
Supplementary Table S1. Partial *cox*1 gene sequences used in the present study, with accession numbers and references listed.











Highlights

- Combined use of second- and third-generation DNA sequencing and advanced informatics
- Characterisation of the first mitochondrial genome for *Bulinus truncatus* – intermediate host of *Schistosoma haematobium*
- Detailed comparative genetic analyses of *Bulinus* and other snails

AUTHOR DECLARATION

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from (nyoung@unimelb.edu.au).

Signed by all authors as follows:

Liina Kinkar 21.1.2021

Andreas J. Stroehlein 21.1.2021

Pasi K. Korhonen 21.1.2021

David Rollinson 21.1.2021

J. Russell Stothard 21.1.2021

Robin B. Gasser 21.1.2021

Neil D. Young 21.1.2021