# Simultaneous genotyping of snails and infecting trematode parasites using high- throughput amplicon sequencing

Hammoud Cyril [1, 2, *], Mulero Stephen [3], Van Bocxlaer Bert [1, 4], Boissier Jérôme [3], Verschuren Dirk [1], Albrecht Christian [5], Huyse Tine [2, 6]

[1] Ghent University, Limnology Unit Department of Biology 9000 Gent, Belgium
[2] Royal Museum for Central Africa, Department of Biology 3080 Tervuren, Belgium
[3] IHPE, Univ. Montpellier, CNRS, IFREMER, Univ. Perpignan Via Domitia Perpignan ,France
[4] CNRS, Univ. Lille, UMR 8198 Evo-Eco-Paleo F-59000 Lille, France
[5] Justus Liebig University ,Systematics & Biodiversity Lab Department of Animal Ecology & Systematics D-35392 Giessen, Germany
[6] University of Leuven ,Laboratory of Biodiversity and Evolutionary Genomics B-3000 Leuven, Belgium

* Corresponding author : Cyril Hammoud, email address : cyril.hammoud@ugent.be

**Abstract :**

Several methodological issues currently hamper the study of entire trematode communities within populations of their intermediate snail hosts. Here we develop a new workflow using high-throughput amplicon sequencing to simultaneously genotype snail hosts and their infecting trematode parasites. We designed primers to amplify 4 snail and 5 trematode markers in a single multiplex PCR. While also applicable to other genera, we focused on medically and economically important snail genera within the Superorder Hygrophila and targeted a broad taxonomic range of parasites within the Class Trematoda. We tested the workflow using 417 Biomphalaria glabrata specimens experimentally infected with Schistosoma rodhaini, two strains of Schistosoma mansoni, and combinations thereof. We evaluated the reliability of infection diagnostics, the robustness of the workflow, its specificity related to host and parasite identification, and the sensitivity to detect co-infections, immature infections, and changes of parasite biomass during the infection process. Finally, we investigated its applicability in wild-caught snails of other genera naturally infected with a diverse range of trematodes. After stringent quality control the workflow allows the identification of snails to species level, and of trematodes to taxonomic levels ranging from family to strain. It is sensitive to detect immature infections and changes in parasite biomass described in previous experimental studies. Co-infections were successfully identified, opening the possibility to examine parasite-parasite interactions such as interspecific competition. Altogether, these results demonstrate that our workflow provides a powerful tool to analyze the processes shaping trematode communities within natural snail populations.

**Keywords** : co-infections, community ecology, diagnostics, Gastropoda, high-throughput amplicon sequencing, schistosomiasis

## 1. Introduction

Trematodes are a highly diversified group of parasitic flatworms, with ~18,000 extant species (Toledo 2014). They are best known for the various diseases they cause in humans and livestock (Hotez et al. 2014), such as schistosomiasis, which affects almost 220 million people worldwide (92% in Africa, WHO 2017). More information on neglected tropical diseases and their control can be found in the new road map (WHO 2020). Most trematodes start their complex life cycle within an obligate intermediate snail host (Esch et al. 2001) towards which they are highly specific (Lockyer et al. 2004). The rest of the life cycle varies greatly among species in terms of subsequent hosts and modes of transmission (Fig. 1). In Africa, most trematodes of high medical or veterinary importance have as intermediate host freshwater snail species from the families Planorbidae and Lymnaeidae, which belong to the superorder Hygrophila (Brown 1994). Importantly, several trematode species may co-occur in the same planorbid or lymnaeid snail species. Therefore, individual snails can be exposed to, and simultaneously infected by, multiple parasite species during their lifetime (e.g., Loker et al. 1981; Soldánová et al. 2011), or by multiple strains of the same species (Minchella et al. 1995; Steinauer et al. 2009). The outcome of co-exposure to multiple parasites depends on 1) the capacity of the snail to be infected by the species and to transmit it (competency, Hopkins et al. 2016), 2) the chronology of infections (immune priming, Portela et al. 2013) and 3) parasite-parasite interactions, which may be synergistic or antagonistic (Esch et al. 2002; Soldánová et al. 2011). Recent studies suggest that parasite-parasite interactions within snails can strongly impact the dynamics of trematode transmission to their subsequent hosts (e.g., Laidemitt et al. 2019). For example, at the interspecific level, Laidemitt et al. (2019) showed that within the snail *Biomphalaria pfeifferi*, the cattle parasite *Calicophoron sukari* antagonizes the human parasite *Schistosoma mansoni*, decreasing prevalence of the latter in snail populations and limiting its transmission to humans. Thus, to better understand disease dynamics, we need to uncover the processes that shape natural parasite communities within snail host populations.

The fragmentary nature of current knowledge on parasite communities within lymnaeid or planorbid snail hosts (but see Soldánová et al. 2012 or Laidemitt et al. 2019) can to a large extent be ascribed to technical research limitations. First, the traditional parasite shedding method to detect infections is highly time-consuming (Born-Torrijos et al. 2014; Kamel et al. 2021). Because typically only a limited fraction of the snails in a population are infected (overdispersion, Curtis et al. 2002), large numbers of snail specimens must be screened to adequately characterize the trematode community (Loker et al.

1981; Eppert et al. 2002), limiting the scalability of these methods. Also, shedding patterns are parasite-specific, usually depending on the activity pattern of the subsequent host (Pages & Théron 1990). For example, *S. mansoni* primarily sheds during the day, and *Schistosoma rodhaini* at night (Pitchford et al. 1969). Furthermore, shedding does not allow detection of immature infections (Born-Torrijos et al. 2014), and due to paucity of diagnostic morphological characters the identification of larval trematodes to species level is often impossible (Nolan & Cribb 2005). Cercarial genotyping resolves this issue but is labor-intensive as multiple cercariae need to be sequenced from each snail to detect all co-infecting strains and species (Eppert et al. 2002; Van den Broeck et al. 2011; Thiele & Minchella 2013; Bakuza et al. 2017), which limits scalability in co-infection studies. Recent methodological advances include the development of molecular xenomonitoring tools such as PCR assays that identify parasite infections in snails to family, genus, or species level (see Kamel et al. 2021 for a review). Compared to shedding, these tend to be more sensitive for detecting infections (an additional 10% to 60%; Born-Torrijos et al. 2014) and they allow cost-efficient screening of large snail populations when combined with a pooling design (Carolus et al., 2019). However, they typically target a single species or limited set of species, usually those of medical or veterinary importance (e.g., Hamburger et al. 2004; Jannotti-Passos et al. 2006; Caron et al. 2011; Schols et al. 2019, Mesquita et al. 2020, Pennance et al. 2020). The ability of xenomonitoring PCRs to detect co-infections is therefore limited to a very restricted set of target species (e.g., Mesquita et al. 2020), making these methods inappropriate for trematode community studies.

Given the above, what is missing is a tool that reliably detects and identifies all trematode infections within individual snails in a cost- and time-efficient manner. At the same time, molecular identification of the snail host is also needed as the morphological identification of certain taxa of lymnaeids and planorbids remains complicated (Brown et al. 1994). In this study, we make use of the increasingly accessible and affordable Next Generation Sequencing (NGS) technology of high-throughput amplicon sequencing (HTAS) to develop an efficient and comprehensive method for simultaneous identification of parasitic trematodes and their snail hosts, thereby preserving the host-parasite relationships. HTAS, which involves multiplexing and pooling, is commonly used in meta-barcoding studies (Herbold et al. 2015; Cruaud et al. 2017; Braukmann et al. 2019) and has been successfully applied to identify parasite infections from DNA extracts of their hosts (Cooper et al. 2018, Chaudhry et al. 2019). By multiplexing various markers in one single PCR and pooling numerous samples per sequencing run (Bybee et al. 2011, Herbold et al. 2015, Cruaud et al. 2017), the study of snail-trematode communities becomes scalable. Beyond its high sensitivity and specificity, the method may also allow reliable detection of co-infections.

Our goal is to develop a versatile HTAS workflow to study taxonomically diverse trematode communities hosted by planorbids and lymnaeids as well as other Hygrophila snails. We performed validation experiments on the snail genus *Biomphalaria*, because of its central role in the transmission of medically and economically important trematodes (Brown 1994). However, the workflow is designed to be easily adaptable to other snails within Hygrophila. We coupled workflow development to various infection experiments involving laboratory-reared snails and well-characterized strains of schistosome species. These experiments provide a controlled set-up to evaluate whether the method 1) provides correct diagnostics (reliability); 2) is robust in technical replicates (robustness); 3) enables the identification of host and parasite to species and even strain level (specificity); 4) discriminates among early/aborted and mature infections (sensitivity); and 5) can detect co-infections (reliability/sensitivity). Finally, we investigated the versatility of our HTAS workflow *via* analysis of wild-caught *Bulinus* and *Radix* snails with documented infection status.

## 2. Materials and methods

### 2.1 Set-up of snail infection experiments

We conducted three infection experiments, one time-series experiment and two co-infection experiments, under controlled conditions at the Host-Pathogen-Environment Interactions laboratory of the University of Perpignan (UPVD, France) with laboratory-reared snails of *Biomphalaria glabrata* strain BgGUA2 from Guadeloupe. These were exposed to miracidia of *Schistosoma mansoni* strains SmBre and SmLE from Brazil and one strain of *Schistosoma rodhaini* from Burundi. We selected these species to maximize the experimental infection rate, as BgGUA2 is a highly permissive snail strain and the selected parasite strains are highly infective (Théron et al. 2014; Portet et al. 2019). The snails were infected by exposure to miracidia obtained by hatching eggs filtered from the liver of white mice strain Swiss OFI (Fig. 2). In the time-series experiment, 144 snails were individually exposed to one miracidium of SmBRE each and subsets of snails were sacrificed at specific moments between 2- and 40-days post-infection (i.e., the approximate time of cercarial development, Schwanbek et al. 1986) to assess whether the method allows to distinguish between early (immature) and patent (mature) infections (Fig. 2). In the first co-infection experiment we exposed 96 snails to 10 miracidia of both *S. mansoni* strains and in the second we exposed 96 snails to 10 miracidia each of SmLE and *S. rodhaini*, thus 20 parasites per snail (Fig. 2). As positive controls we quantified the baseline infection rate of each

of the three parasite strains by exposing 30 snails to 10 miracidia each. To test for contamination, negative controls for all experiments consisted of 1) subjecting 5 non-exposed snails to all subsequent procedures, and 2) adding a single miracidium to ethanol-preserved samples of the non-exposed snails to simulate environmental contamination with parasite DNA. In each experiment we exposed snails to miracidia for 24h. After the exposure phase, the snails were transferred to different aquaria (10L) depending on their treatment, maintained at 25 °C under a 12h:12h light-dark regime and fed *ad libitum* before being sacrificed by heat shock (70 °C for 1 minute). A total of 20 snails (4.6%) died prematurely and were excluded from analysis. The remaining 417 snails were preserved in 80% ethanol and subjected to the HTAS workflow at the Royal Museum for Central Africa (Tervuren, Belgium).

### *2.2 HTAS workflow*

#### 2.2.1 DNA Extraction

We separated the whole animal from its shell, dried it for 10 seconds on disposable absorbent paper to remove the remaining ethanol, and subsequently homogenized it with a sterile scalpel. We changed the absorbent paper between specimens and sterilized the scalpel blade and the pliers in a flame for 20 seconds. DNA was extracted using the E.Z.N.A. Mollusc DNA Kit (OMEGA Bio-tek, Norcross, GA, USA), according to manufacturer's instructions. The volume of final DNA extracts was 150μL, and the DNA concentration was measured using Qubit (Qubit Fluorometer 2 with Qubit™ 1X dsDNA HS Assay, Invitrogen, ThermoFisher, USA). DNA aliquots were diluted to 10ng/μL in Milli-Q water and stored at −20 °C until subsequent analysis.

#### 2.2.2 Primer design

Nine primer pairs (Table 1) were selected to simultaneously amplify one nuclear and three mitochondrial markers of the snails, and one nuclear and four mitochondrial markers of the trematodes. These included two fragments of the snails' cytochrome c oxidase subunit 1 (COI1_snail and COI2_snail, Fig. S1A), a fragment of NADH dehydrogenase subunit 1 (NAD1_snail), a fragment of the internal transcribed spacer 1 of the rRNA cluster (ITS1_snail) and for the trematodes two fragments of COI (COI1_trematode and COI2_trematode, Fig. S1B), one fragment of NAD1 (NAD1_trematode), one fragment of cytochrome b (cytb_trematode) and one fragment of internal transcribed spacer 2 (ITS2_trematode). The detailed workflow for primer design can be found in supplementary information (Text S1, Table S1, Table S2).

### 2.2.3 PCR conditions

Simultaneous amplification of the nine amplicons was performed in a single multiplex PCR using the ThermoFisher Platinum™ Multiplex PCR Master Mix according to the following program: 2 minutes of initial denaturation at 95 °C, followed by 40 cycles each of 30 seconds at 95 °C, 90 seconds at 57.5 °C, 40 seconds at 72 °C, and then a final elongation step of 10 minutes at 72 °C, using a Professional Thermal Cycler (Biometra, Göttingen, Germany). PCR products were visually checked by gel electrophoresis for 1h45 min at 120V in a 3% agarose gel with Midori Green direct staining and ultraviolet light exposure. The multiplex PCR mix was prepared using 12.5 µL of master mix, 2.5 µL of multiplex primer mix (prepared with individual primer pair concentrations as shown in Table 1), 5µL of water and 5 µL of template DNA at 10 ng/µL, for a total of 25 µL of PCR mix per sample. A blank consisting of Milli-Q water (Merck, Darmstadt, Germany) was used as negative control.

### 2.2.4 Illumina MiSeq sequencing

The multiplexed amplicons were sent to the Genomics Core of the University Hospital Leuven (Belgium) for library preparation with the Illumina MiSeq v3 kit (indexing PCR, balancing amplicon concentrations, EMPURE Bead purification) and sequencing. Specifically, this included the PCR products of the 417 *Biomphalaria* specimens from the controlled infection experiments, 7 wild-caught *Bulinus* specimens, 7 wild-caught *Radix natalensis* and the blank. In addition, to examine the robustness of the HTAS procedure, we aliquoted the PCR products of 6 *B. glabrata* specimens in triplicates. The resulting 444 libraries were sequenced in a single sequencing run on an Illumina MiSeq. A first sequencing run was performed with 10% of Phi X spiking but generated an insufficient quantity of reads ($\sim 3 \times 10^6$) with overall low Phred Quality Score. As run failure can result from low library diversity, we increased Phi X spiking to 25% in a second run, which successfully increased the quality and quantity of reads obtained ($>20 \times 10^6$).

### 2.2.5 Bioinformatics processing

Raw reads were demultiplexed using the sample-specific unique dual indexes (no mismatch) before further processing. We removed primer sequences from the reads using cutadapt v.2.7 (Martin 2011) in bash v.4.4.19 (GNU 2007) with a minimum overlap of 15bp and 15% maximum error rate. Subsequently, the reads were processed in the dada2 package v.1.12.1 (Callahan et al. 2016) of the statistical environment R v.3.6.1 (R Core Team, 2019) to infer Amplicon Sequence Variants (ASVs) without

activating the 'pool' parameter. The reads were first trimmed and filtered (maximum expected error = 2, length after trimming > 150 bp) in paired-end mode (i.e., both forward and reverse reads must pass to be output), then de-replicated. The base-specific error rates were estimated from $5 \times 10^8$ bases in randomly selected reads, both for the subset of forward and reverse reads. Quality-controlled reads were then grouped and subsequently merged into ASVs (minimum overlap = 20 bp, maximum 1 mismatch) before removing chimerae using the 'consensus' method (Callahan et al. 2016). To organize ASVs per taxon and per target region we constructed a local reference database from BLASTing ASVs against the NCBI nucleotide database. Subsequently, ASVs were compared to this database using BLAST+ v. 2.9.0 (Camacho et al. 2009) in bash (word_size = 10, penalty = -1, gapopen = 2 and gapextend = 1). The results were compiled in a table summarizing the ASVs identified per sample, with associated read count and proportion of reads supporting the ASV (count divided by total number of filtered reads attributed to the sample). Based on the BLAST results, we distinguished ASVs of target species from contaminants (i.e., ASVs belonging to organisms other than the studied model taxa). Additionally, ASVs belonging to one of the model taxa but not to the taxon of that specific sample (as based on the experimental set-up), were classified as cross-contaminants (e.g., a *S. rodhaini* COI1 sequence found in a snail exposed to *S. mansoni* only). Finally, for each non-contaminant sequence we distinguished between primary and secondary ASV. The primary ASV of a specific sample and marker corresponds to the on-target ASV with the highest proportion of reads supporting it, whereas secondary ASVs include all other ASVs for the same sample and marker. The presence of secondary ASVs may indicate polymorphism in targeted markers, or depending on the specific experiment, co-infection. Contaminant ASVs can either result from non-specific primer annealing, wet laboratory contamination during DNA extraction or PCR (Ballenghien et al. 2017), or cross-talk (index hopping) during MiSeq sequencing (MacConaill et al. 2018). In any of these cases, we would expect the number of reads associated with contamination to be much lower than those related to targeted ASVs: the concentration of contaminating template DNA ought to be minimal compared to that of the original sample (Lusk 2014), and cross-talk occurs at an extremely low rate in sequencing with dual indexes (MacConaill et al. 2018). To remove contaminating and cross-contaminating sequences, we applied a stringent, marker-specific filtering of ASVs based on read support. First, based on the ASV classification described above, we determined the highest proportion of reads supporting cross-contaminating ASVs for each marker. Then, we used these maximum values (rounded up with 0.005%) as threshold to filter all ASVs with inferior read support, for each marker. We adopted a marker-specific threshold because we observed that the patterns of cross-contamination

varied among markers; markers with ASVs supported by a higher proportion of reads (e.g., COI2_snail) also showed higher read-support for cross-contaminating ASVs. Therefore, applying a uniform threshold for filtering would either result in removing most sequences for markers with low read-support or retaining cross-contaminating ASVs in markers with high read support. This threshold-based filtering was aimed at increasing the reliability of subsequent diagnostics and genotyping. A snail was diagnosed as being infected by a specific trematode strain if at least one ASV of that parasite strain was retained after filtering. All relevant datasets, FASTA files as well as bash and R scripts are available on the Dryad Data Repository (Hammoud et al. 2020a). Raw reads are available in NCBI SRA (Hammoud et al. 2020b) and the filtered ASVs are available in GenBank (Hammoud et al. 2020c).

## 2.3 Performance evaluation under controlled experimental conditions

We determined the performance of our HTAS workflow by investigating whether it 1) provides correct diagnostics (reliability); 2) is robust in technical replicates (robustness); 3) enables the identification of host and parasite species to species and strain level (specificity); 4) discriminates among early/aborted and mature infections (sensitivity); and 5) can detect co-infections (reliability/sensitivity).

A first examination of reliability, already mentioned in section 2.5, involved the characterization of contamination. Subsequently, we compared the proportion of reads from infected and non-infected snails across the various markers. We statistically tested for differences in read support for each category of ASVs (primary, secondary and contamination) per marker in infected *versus* uninfected snails using pairwise Wilcoxon tests with Bonferroni correction for multiple comparisons in R. In addition, we evaluated the reliability of the diagnostics inferred with the HTAS workflow by comparing it with the results of an infection rapid-diagnostic PCR (infection RD-PCR, Schols et al. 2019). We opted for this method as it allows to simultaneously detect the presence of any trematode within the snail and verify whether some of these trematodes belong to the genus *Schistosoma*, which is not possible with other existing qPCR or ddPCR assays (Kamel et al. 2021). The method consists of multiplexing 3 primer pairs in a PCR reaction and evaluating the infection diagnostic based on amplicon size discrimination using gel electrophoresis. The PCR amplifies a fragment of 18S rRNA of the snail, used to validate success of the PCR reaction, a fragment of 18S rRNA of trematodes aimed at general parasite detection and a fragment of COI designed specifically to detect *Schistosoma* spp. Infection RD-PCRs were performed

according to the protocol described in Schols et al. (2019), except for the number of PCR cycles, which was reduced from 40 to 25 to calibrate the assay to avoid detecting early infections (< 6 days post-exposure). Additionally, we used published data on infection dynamics to verify whether the infection rates observed in our experiments were consistent with those reported for the specific snail and parasite strains used. The infection dynamics of *S. mansoni* strains SmBre and SmLE with *B. glabrata* strain BgGUA (i.e., a strain sampled from the same wild population as BgGUA2) was extensively described in Théron et al. (2014). For *S. rodhaini*, we compared our data with those of Portet et al. (2019) who reported the infection rate of *S. rodhaini* from Burundi with another strain of *B. glabrata* (BRE), as no published data exist on strain BgGUA2.

The robustness of post-PCR procedures in our HTAS workflow was examined from technical replicates on 6 *B. glabrata* specimens; 3 were co-infected with SmLE and SmBre and the other 3 with SmLE and *S. rodhaini*. We compared the pools of ASVs retained after filtering for all replicates, both for snails and their infecting trematodes.

To identify hosts and parasites to species or strain level, the ASVs retained after contaminant removal were first BLASTed to our local reference database, then aligned per taxon (snails or trematodes) in Geneious, using the MUSCLE algorithm to 1) calculate pairwise genetic distances (uncorrected p-distance), 2) check for the presence of STOP codons in the coding regions (to identify nuclear mitochondrial DNA or NUMT; Lopez et al. 1994), and 3) identify diagnostic single nucleotide polymorphisms (SNPs) specific to the species or strains used in our experiments.

The sensitivity of the HTAS workflow was studied using the samples of the time-series experiment (Fig. 2). Specifically, we hypothesized that the proportion of reads attributed to trematode ASVs should reflect the volume of parasite tissue within the host and, therewith, the development of the infection over time. The progressive changes in the volume of *S. mansoni* tissues within the digestive gland of *B. glabrata* as described by Théron et al. (1992) provided a reference frame for the proportion of reads attributed to *S. mansoni* ASVs (all markers pooled).

We investigated the reliability and sensitivity of the HTAS workflow for detection of co-infections by using the samples co-exposed to multiple Schistosoma strains/species (Fig. 2). We compared the rates of successful infections under co-exposure with those of exposure to single strains/species (i.e., positive

controls for the rates of infections). Information on the statistical analyses for co-infection experiments can be found in the supplementary information (Text S2).

### 2.4 Validation in other Hygrophila

We assessed the applicability of our HTAS workflow to other snail taxa, i.e., to 7 wild-caught *Bulinus* specimens (1 *B. globosus*, 3 *B. truncatus,* 2 *B. tropicus* and 1 *B. forskalii*) and 7 wild-caught *R. natalensis* of which the infection status had already been documented (Table S3). These specimens were processed identically to the *B. glabrata* specimens, were sequenced in the same MiSeq run and subjected to the same bioinformatics processing up to chimera removal. After chimera removal some of the samples showed substantial contamination by *B. glabrata* and *S. mansoni*, which was removed before filtering based on read support. This filtering was performed with the same marker-specific threshold approach used in the controlled laboratory experiments, but thresholds were recalculated per genus because marker amplification varied substantially among taxa. Retained snail and parasite ASVs were BLASTed against Genbank for identification to the lowest-possible taxonomic level.

## 3. Results

### 3.1 DNA extraction, MiSeq library preparation and sequencing

All 417 DNA extractions performed for the controlled experiments yielded DNA concentrations >10 ng/μL (on average 103 ± 100 ng/μL), except for 4 (0.3, 4.3, 6.9 and 9.7 ng/μL) which were nevertheless included in library preparation. Library preparation was successful for all these samples as well as the triplicates and the *Bulinus* and *Radix* samples (total of 444 libraries). After removal of phage X sequences and de-multiplexing by sample, MiSeq sequencing resulted in a total of $13.4 \times 10^6$ reads. Trimming, filtering and merging resulted in a loss of 48% of reads per sample, on average. Chimera detection prompted removal of 440 out of the 917 constructed ASVs, but the number of reads lost was modest (<1% of the filtered reads). The number of ASVs per marker and per sample varied between 0 and 15. These data were tabulated for subsequent analysis, which started with the detection of contaminants. Out of the 477 filtered ASVs, 13 belonged to taxa not included in the experiment (bacteria, teleost and plants), and 5 ASVs were off-target markers from the studied snails (Table S4). These contaminants were

found in 6 samples only, were supported by a low proportion of reads (0.1 ± 0.2% on average) and were easy to identify. However, cross-contaminant ASVs were observed in 273 samples (~63%) and represented on average 0.8 ± 1.1% of reads/ASV/sample (cumulatively 2.6 ± 3.6% of reads/sample). Of the 1.1 x 10⁶ reads attributed to the PCR1 blank 99.7% were removed after trimming. ASVs resulting from the remaining 0.3 % of reads belonged to various organisms of the study (*B. glabrata*, *R. natalensis*, etc.), testifying of a low background level of cross-contamination. Rigorous filtering of contaminants was thus required, which was achieved via marker-specific thresholds on the frequency of reads supporting each ASV.

### 3.2 HTAS performance evaluation with controlled laboratory experiments

3.2.1 Read support, data reliability and diagnosis of infection status.

The number of reads per sample in the controlled laboratory experiments ranged from 6,080 to 81,482 with an average of 25,659 ± 15,053. ASVs merged from these reads were in average 391 ± 67 bp long. The representation of these reads over the nine markers is provided in Table 1 (see Table S5 for details on read distribution among markers and the infection status, as well as threshold levels for filtering). The relative proportion of reads supporting the various ASVs obtained per marker for all samples in our laboratory experiments is shown in Fig. 3. Noteworthy aspects are that the proportion of contamination is higher for trematode ASVs than for those of snails, because contamination is inflated in trematode ASVs of non-infected snails. Second, NAD1_trematode and cytb_trematode are overall substantially less amplified than other markers. Nevertheless, removal of contaminants based on threshold levels resulted in the retention of the primary ASVs for all nine markers and, except for COI1_trematode, NAD1_trematode and cytb_trematode, also secondary ASVs (Fig. S2).

Upon examining the relative proportion of reads per sample prior to contaminant filtering, that is when all reads for a sample together represent 100%, strong differences existed in read support for the various markers and within markers for ASVs depending on infection status (Table 2, Fig. S3). Specifically, we observed that COI2_snail amplified more than other markers, and its representation differed strongly between infected and non-infected snails (~23% *vs* ~54%, respectively). This ~30% difference overall relates strongly to the total representation of trematode-targeting ASVs in infected snails. Pairwise Wilcoxon tests indicated that for all markers except COI1_snail, the proportions of

primary ASVs differed significantly ($\alpha$ = 0.05) depending on infection status. The three other snail markers have thus more reads in uninfected than in infected snails, whereas all five trematode markers have significantly higher read support in infected snails. Significantly more secondary ASVs were amplified for COI1_snail, COI2_snail and NAD1_snail in infected snails than in uninfected specimens, and *vice versa* for ITS1_snail. Secondary ASVs for trematode markers COI2_trematode and ITS2_trematode had higher read support in infected snails, whereas no significant differences between infected and uninfected snails were detected for the other markers. No significant pairwise differences were observed in the read support for contaminant ASVs according to infection status. In both exposed (n = 5) and unexposed (n = 5) negative controls, ASVs of *S. mansoni* were inferred but removed during filtering.

Inferring infection status with HTAS and infection RD-PCR yielded very similar results for the 417 *B. glabrata* snails included in our controlled experiments. For 223 specimens both methods indicated infection, and 186 specimens were diagnosed as uninfected by both methods (overall congruence >98%). Six individuals were diagnosed as infected by the HTAS workflow but not by infection RD-PCR, and *vice versa* for one specimen. One specimen was excluded from this comparison because no band (internal control or infection) was observed with the infection RD-PCR.

Comparison of our results with published data indicates that infection rates inferred with the HTAS workflow fall within the expected range for each specific host-parasite association under study. For the association of BgGUA2 with SmBRE, 26% of the snails (n = 142) were infected in the time-series experiment (30% when excluding the 20 snails sacrificed after 2 days for which no infections were detected), which closely approximates the value of 28 ± 11% described by Théron et al. (2014). Note that these rates are specific for infections with a single miracidium. The infection rate increased to 96% when individual snails were exposed to 10 miracidia (positive control for the co-infection experiment, n = 27), somewhat higher than the value of 80 ± 10% reported by Théron et al. (2014). For the other two associations, we examined infection dynamics only with data from those positive controls. The association of BgGUA2 with SmLE resulted in 92% infection rate (n = 26), identical to the value of 92 ± 8% published by Théron et al. (2014) and the association of BgGUA2 with *S. rodhaini* resulted in 82% infection which is again highly similar to the 80% reported by Portet et al. (2019).

### 3.2.2 Robustness

Genotyping of technical replicates of the same samples used for the verification of pre-PCR procedures yielded identical diagnostics and similar ASV profiles, indicating high robustness of the new HTAS methodology. Minor differences were only observed in the ASVs of snail markers, and except for one case all related to secondary ASVs (Table S6).

### 3.2.3 Specificity of host and parasite identifications

Among the 417 *B. glabrata* specimens involved in our laboratory experiments, we identified 3 distinct haplotypes of the marker COI1_snail (differing from one another by 1 'silent' SNP), 2 of COI2_snail (differing from one another by 2 'silent' SNPs), 24 of ITS1_snail (differing by a combination of multiple SNPs and INDELs), and 3 of NAD1_snail (2 differing from one another by 1 synonymous substitution, the third having a non-synonymous mutation and a 'silent' SNP). These data indicate that substantial genetic diversity exists in the BgGUA2 laboratory strain. The genetic diversity in the parasite strains of our experiments was substantially lower, but nevertheless allowed to recognize both *Schistosoma* species and strains. We recovered 3 haplotypes of COI2_trematode (1 in each *Schistosoma* species/strain included), 2 of COI1_trematode (1 in *S. rodhaini* and the other in both *S. mansoni* strains), 3 of ITS2_trematode (1 in *S. rodhaini* and 2 common to both *S. mansoni* strains that differ from one another by 1 SNP). Both strains of *S. mansoni* have the same haplotype of NAD1_trematode and cytb_trematode, but as verified with the control specimens, these markers were not amplified for *S. rodhaini*. The pairwise genetic distances among *B. glabrata* specimens and among the different parasite species/strains are summarized per marker in Tables S7.A to S7.I. Here we focus on reporting genetic distances between parasite species/strains. The genetic divergence between *S. mansoni* and *S. rodhaini* for COI1_trematode, COI2_trematode and ITS2_trematode was ~13.6%, 12.3%, 1.0%, respectively. SmBre and SmLE strains were divergent only in COI2 (0.2%).

### 3.2.4 Sensitivity to detect infection stages

In the course of the time-series experiment, the proportion of infected snails increased from day 2 to 16 after which it remained stable until day 30 and subsequently dropped dramatically by day 40 (Table 3). The proportion of reads attributed to *S. mansoni* ASVs in these specimens shows a similar trend (Fig. 4). Interestingly, the increase in read support for parasite ASVs occurred earlier than the increase in volume of parasite tissue within the snails' digestive gland as documented by Théron et al. (1992). This volume started to increase after 2 weeks with its steepest growth between day 14 and 30.

### 3.2.5 Reliability and sensitivity to detect co-infections

The null models of expected rates of co-infection based on the rates of infection observed in the positive control experiments are indicated in Table 4 together with observed levels of co-infection. For all time period later than 2 days after infection, our results indicate statistically significant differences between the expected levels of co-infection assuming random/independent interactions between parasite species/strains *versus* infection levels observed in our HTAS diagnostics. Co-exposure to both *S. mansoni* strains led to an infection dynamic that differed significantly from expectation under the abovementioned assumptions on day 10, but even more so by day 40 (Table 4). Overall the level of co-infection was lower than expected, and the proportion of snails infected with SmBre (either co-infected or infected by SmBre alone) was much lower than expected. A higher proportion of snails was infected with SmLE (co-infected or not) than with SmBre, but not as high as in positive controls. Even larger differences were observed in the co-infection experiment with *S. mansoni* and *S. rodhaini*: co-exposure resulted in very low levels of co-infection, and of infection by *S. rodhaini* on day 10 and day 40 when compared to expectations based on literature (Table 4).

Furthermore, co-infection resulted in lower levels of successful amplification/sequencing of the trematode markers, and if markers were successfully obtained they had lower read support. Regression analysis indicated that the proportion of reads supporting the COI2_trematode ASVs of both strains of *S. mansoni* (the only marker enabling distinction between strains) was significantly lower in co-infections than in single infections and remained stable over time, whereas in single infections it increased over time (Fig. 5). In co-infections with *S. mansoni* and *S. rodhaini,* an ASV of ITS2_trematode was obtained from all six snails that indicated co-infection, whereas COI1_trematode was obtained from one specimen only, and COI2_trematode from none. As for co-infection with both *S. mansoni* strains, regression analysis for co-infection with *S. mansoni* and *S. rodhaini* showed a significantly lower proportion of reads attributed to the ITS2_trematode ASVs under co-infections compared to single infections (Fig. 6).

### 3.3  Validation in other Hygrophila

On average the number of de-multiplexed reads from *Bulinus* specimens (4 different species, see above) (36,708 ± 35,015) was higher but of the same order of magnitude as those from *B. glabrata* in the controlled experiment (25,659 ± 15,053), whereas those from *R. natalensis* were an order of magnitude higher (210,019 ± 79,796). The distribution of raw reads after removal of unambiguous contaminants,

but before threshold-based filtering, was strongly taxon-specific (Table 2). Whereas the proportions of *Bulinus* reads were comparable to those of infected *B. glabrata* we observed a strong overrepresentation of COI2_snail in *R. natalensis* (~74% of the total reads *vs* ~20% for *Bulinus* and ~23% for infected *B. glabrata*). The distribution among markers of trematode reads in the wild-caught snails strongly resembled those in the laboratory experiments (namely, lower read support for NAD1_trematode and cytb_trematode than for COI1_trematode, COI2_trematode and ITS2_trematode).

The success of amplification and sequencing for each marker and the level of intragenomic variation detected at the end of the HTAS workflow is summarized in Table 5 (snails) and 6 (trematodes). These results indicate that the success rate of amplification is high overall, but higher for snails than for trematodes and with substantial variation among markers. Specifically, ITS1_snail was consistently obtained from all snails, whereas the success rate for COI1_snail varied between *Bulinus* and *Radix*, for COI2_snail among *Bulinus* species, and NAD1_snail was more variable across all snails. Overall, polymorphism is low in these markers, apart from ITS1_snail, and to lesser extent NAD1_snail. However, there was substantial variation among snail specimens in the success of amplifying and sequencing parasite amplicons. ITS2_trematode was most consistently obtained (in 10 out of 12 parasites), followed by COI2_trematode (9/12), cytb_trematode (8/12), COI1_trematode (7/12) and NAD1_trematode (5/12). In some specimens, only 1 of these 5 markers was amplified. Intragenomic variation was observed in four specimens for ITS2_trematode and one specimen for COI2_trematode.

Screening of wild-caught samples with our HTAS workflow confirmed infection in all 7 *Bulinus* and 7 *R. natalensis* specimens analyzed, and detected in total 12 different parasites belonging to two orders and 8 families of the Trematoda (Table 6, Table S3). Eight of these were specific to *Bulinus* spp., 3 to *R. natalensis* and 1 was found in both *Bulinus tropicus* and *R. natalensis*. For *Bulinus* spp., the identification of parasite infections was highly consistent with previous diagnostic efforts. All previous infections were recovered, except for 2 *B. tropicus* specimens where only one infection was identified with HTAS workflow compared to 2 infections using a combination of RD-PCR and genotyping (Table S3). For *R. natalensis* the HTAS diagnostics differed more substantially from previous efforts, perhaps due to overrepresentation of the marker COI2_snail. Specifically, *Fasciola* sp. infections in 5 snails that had been detected using RD-PCR were not recovered. However, the HTAS workflow detected 3 additional infections in these same snails (Table S3). Furthermore, in 1 *R. natalensis* specimen it detected the

presence of 9 distinct but highly similar haplotypes of the COI2_trematode amplicon, possibly indicating intraspecific co-infection by *Cotylurus* sp. (Table S3).

4. **Discussion**

In this study, we aimed to design a new workflow based on HTAS to analyze entire communities of trematodes within populations of their snail hosts. By testing this workflow on artificially and naturally infected snails, we showed that it yields reliable and robust diagnostics, and it enables the simultaneous molecular characterization of both hosts and parasites. Below, we discuss our evaluation of the HTAS workflow, assess its limitations and provide further recommendations for its application to field surveys.

*4.1 Specificity of host and parasite identification and intragenomic variations*

Our workflow enabled the molecular characterization of snails belonging to 6 different species from 3 genera, and of parasites belonging to 16 different species from 8 families, demonstrating its broad taxonomic versatility. While all snails were identified to species level, the taxonomic resolution of the identification of their trematode was lower, sometimes only up to family level. This lack of specificity is mostly linked to a lack of reference sequences for African trematodes in the NCBI databases (see Schols et al. 2020). In addition, the high molecular diversity within this group complicates the design of general primers that would amplify all trematode species (Moszczyńska et al. 2009; Vanhove et al. 2013). More field studies that combine molecular and morphological methods (e.g., Laidemitt et al., 2017) and use general primers like our HTAS workflow are needed to enrich these databases and overcome such limitations.

The 4 snail markers were consistently retrieved, except for low amplification/sequencing success of NAD1_snail in *Bulinus* and *Radix* spp. The ITS1_snail and both COI markers were particularly informative for species identification because of the high representation of these makers for Hygrophila in the GenBank. Sequencing success of the parasites was high in our controlled experiment. Notably, we could discriminate between the 2 *S. mansoni* laboratory strains based on a single SNP in COI2_trematode, highlighting the specificity of our workflow. Parasite identity was mainly determined using ITS2 or COI sequences, which are well-represented in molecular databases due to their broad use in studies of trematode diversity (Blasco-Costa et al. 2016).

The HTAS workflow also enabled the detection of intragenomic variation, foremost in the nuclear markers ITS1_snail and ITS2_trematode and to a lesser extent in the mitochondrial markers NAD1_snail, COI1_snail and COI2_snail. As ITS is part of the ribosomal tandem repeat, intragenomic variation has already been documented in a variety of animal taxa such as trematodes (Steinauer et al. 2008), molluscs (Stothard et al. 2000; Vierna et al. 2010), insects (Parkin & Butlin 2004; Sword et al. 2007), anthozoans (Rodriguez-Lanetty & Hoegh-Guldberg 2002), crustaceans (Harris & Crandall 2000) and teleost fishes (Xu et al. 2009). Some authors suggested that such variation could be more widespread than hitherto assumed (Harris & Crandall 2000; Sword et al. 2007) and may disturb phylogenetic reconstruction by violating orthology assumptions (Harris and Crandall 2000; Alvarez & Wendel 2003). However, multiple ITS copies can also provide information on meaningful biological processes such as hybridization in schistosomes (Steinauer et al. 2008) sequenced ITS copies of 2 distinct species from individual hybrid schistosomes. Nevertheless, ITS sequences should be interpreted with caution.

Intragenomic variability in mitochondrial markers can occur either as natural functional variants from different populations of mitochondria found in a single organism (heteroplasmy) or as non-functional, often shorter versions of the original gene that are integrated in the nuclear genome (NUMT; Lopez et al. 1994). In molluscs, heteroplasmy has been described for bivalves, where it is caused by double uniparental inheritance of mitochondria (Gusman et al. 2017). To our knowledge, heteroplasmy has not previously been observed in snails, although the issue was investigated by various authors (Thomaz et al. 1996; Davison 2000; Parakatselaki et al. 2016; Gusman et al. 2017). Our data favor heteroplasmy above NUMT as we did not find STOP codons in the variants found in *B. glabrata* and *B. globosus*. The number of variants differed across markers (for *B. glabrata*: 2.4% in COI2_snail, 8.2% in COI1_snail and 52% in NAD1_snail), which may be due to methodological aspects such as filtering. Applying clustering methods onto the dada2 output (see Palmer et al. 2018; Jusino et al. 2019) may allow further filtering and provide additional insight into this issue. Furthermore, including a 'host-parasite mock community' (an artificial sample composed of single-copy target template DNA) in the experiment could provide an additional way to control the sequencing outputs, though generating such a sample might be technically demanding (see the development of the 'arthropod mock community' in Jusino et al. 2019).

### 4.2 Infection diagnostics, cross-contamination and read support

Infection diagnostics inferred from HTAS results were highly consistent with those of the infection RD-PCR and with the infection rates obtained in previous infection experiments using the same host and

parasite strains (Théron et al., 2014), emphasizing the reliability of the workflow. Interestingly, more samples were classified as HTAS positives and RD-PCR-negatives (6 occurrences) than *vice versa* (1 occurrence), suggesting that the HTAS workflow is more sensitive than the infection RD-PCR.

In addition, our results indicate that the different molecular parts of the workflow (DNA extraction, amplification and sequencing) were successful for all specimens included in the controlled laboratory experiment, confirming its effectiveness. The level of cross-contamination was relatively high, which is a common observation in NGS applications, especially those involving PCR steps (Lusk 2014; Ballenghien et al. 2017; Karstens et al. 2019). This was successfully resolved by applying stringent filtering of ASVs based on thresholds of read support. Simple precautions in the laboratory like leaving empty wells between samples during PCR1 also significantly lowered the level of cross-contamination (personal observation). Additionally, including additional negative controls such as an extraction blank and a sequencing blank can also help identify the sources of cross-contamination and control its impact on the ASVs profiles. Although most markers were effectively sequenced, the distribution of reads among them was generally unequal and varied among taxa. This uneven distribution is a common feature in HTAS workflows (e.g., Lerch et al. 2017) and does not impair the outcome. Furthermore, amplicon balance can be improved by careful optimization of primer concentrations in the multiplex PCRs. Read distributions also depended strongly on infection status: snail markers in infected specimens had on average significantly fewer reads than in non-infected specimens because the relative proportion of trematode reads increased.

### *4.3 Sensitivity to discriminate between immature and mature infections*

We optimized our workflow to detect infections after 6 days and while being insensitive to earlier infections (after 2 days). In the first few hours or days after miracidial penetration, schistosome infections are still being suppressed by the snail's immunological system (Basch & DiConza 1974; Pinaud et al. 2016). Therefore, the probability of abortion of the infection in case of an incompatible host-parasite combination is high compared to that in later stages. Consequently, the fact that the HTAS workflow is only sensitive to 'successful' infections (i.e., it reliably detects infections from 6 days onward) helps avoiding false conclusions on host-parasite associations. However, additional research would be necessary to fully characterize the sensitivity of the workflow within a broader range of host and parasite species.

The proportion of parasite reads obtained with the HTAS workflow in our time-series experiment reflected, at least semi-quantitatively, the changes in parasite biomass within the snail throughout the infection process. Notably, the increase in parasite reads precedes the surge of parasite biomass caused by the rapid proliferation of daughter sporocysts (Théron et al. 1992), rather than following it. The proportion of parasite reads started to increase 6 days after infection, which coincides with a phase of intensive cellular division in the mother sporocyst, characterized by high mitotic activity in layers of germinal cells (Buddenborg et al. 2019). From 16 days onwards, the proportion of reads leveled off, whereas Théron et al. (1992) observed that the parasite biomass continues to increase. Thus, parasite DNA concentration may have passed a threshold after which PCR reactions saturates and does not yield a higher quantity of amplicons. By day 40, the proportion of parasite reads decreased slightly. Note that the number of infected snails (i.e., the infection rate) in our time-series experiment decreased at 40 days compared to previous time points, which was unexpected. An explanation could be that, because snails were sacrificed at successive points in time, individuals that appeared unhealthy or moribund were preferentially sacrificed to limit background mortality between the sampling points. As a consequence of this sampling strategy, the proportion of uninfected snails may have artificially increased towards the end of the experiment. Also, the decrease in the proportion of parasite reads could also reflect the start of parasite shedding, which reduces parasite biomass within the snail, but the number of infected snails at this time point (n = 2; Table 3) is too low to draw any conclusion.

### 4.4 Capacity to detect co-infections and parasite-parasite interactions

Our workflow successfully distinguished co-infections by trematodes of different species, and even co-infections by different strains. This feature represents a major improvement over traditional shedding, genotyping using Sanger sequencing or RD-PCR (Eppert et al. 2002; Thiele & Minchella 2013; Bakuza et al. 2017; Schols et al. 2019) and will greatly facilitate the study of inter- and intraspecific parasite interactions within populations of their snail hosts.

We generally observed lower read support for parasite ASVs in experimental co-infections compared to single infections. Given that the proportion of parasite reads is linked to changes in parasite biomass, this result likely indicates a relatively lower quantity of parasite template DNA within co-infected snails. In addition, the rates of co-infection in the controlled experiments were significantly lower than expected under random co-infection. This pattern implies either that our method cannot detect all co-infections or that the observed rates of co-infection reflect antagonistic parasite-parasite interactions.

We cannot fully exclude the former hypothesis since we did not dissect the snails, but Norton et al. (2008) and Steinauer et al. (2009) reported laboratory and field observations corroborating intra-molluscal competition between *S. mansoni* and *S. rodhaini*. This competition appeared mainly detrimental to *S. rodhaini*, which is consistent with the extremely low rate of infection by this parasite in our co-infection experiments. A potential biological explanation might be that developing *S. mansoni* larvae migrate earlier to the snail digestive gland than those of *S. rodhaini* (Théron & Touassem 1989), potentially pre-empting the niche where cercariogenesis happens. Similar mechanisms may have caused the apparent advantage of SmLE over SmBRE in our experiments. Indeed, Le Clec'h et al. (2019) showed that SmLE sporocysts grow faster and larger than those of SmBRE, and that the former is more virulent and deadly to the snail than the latter. The higher proportion of reads attributed to SmLE over both SmBRE and *S. rodhaini* corroborates the hypothesis that the former is the superior competitor in co-infections.

### 4.5 *Applicability to other snail and parasite species*

Applying HTAS on wild caught *Bulinus* and *Radix* snails with previously established infection status indicated that our method is applicable to a wide diversity of Hygrophila and Trematoda. HTAS allowed the identification of most trematode infections detected previously and detected a few new ones (Table S3), thus showing a slightly higher sensitivity compared to the combination of cercarial genotyping and RD-PCR. The taxonomic resolution was also generally higher, as was the sensitivity towards detecting co-infection by strains belonging to the same species. However, some co-infections (primarily of *Fasciola* sp.) were consistently missed, suggesting either an issue with the primers or with the previously performed diagnostic. This will be investigated in future research. Given that the read support of individual ASVs might decrease under co-infection, the reliability to detect co-infections with HTAS may be optimized by increasing sequencing depth.

### 4.6 Advantages and drawbacks of the workflow

Our HTAS workflow produces a wealth of molecular data on both hosts and parasites, enabling the analysis of trematode communities within snails with an unprecedented precision. The approach is cost-efficient compared to Sanger sequencing. Assuming a cost of 0.5€ for PCR amplification of one marker/sample, and 0.5€ for Sanger sequencing of one PCR product, it would cost 9€/sample to sequence all 9 markers. The cost would increase to 12.7€/sample if we include DNA extraction and

detection PCR, which is ~25% more expensive than HTAS (see Table S8). Furthermore, in case of co-infections, a Sanger sequencing approach would yield much less information as it would provide only one sequence per marker, therefore reducing the sensitivity to detect co-infections. The long time required to perform the HTAS workflow (~8 weeks starting with PCR1) is compensated by the efficiency of multiplexing the PCR reactions, resulting in similar time investment for HTAS and Sanger sequencing. Nevertheless, the substantial cost and time required to implement HTAS limit its use for routine diagnostics or monitoring. For this, shedding experiments, xenomonitoring PCR, qPCR or environmental DNA tools that detect cell-free DNA of larval parasite stages (Alzaylaee et al. 2020) or intermediate snail host (Mulero et al. 2020) in water samples are much more appropriate. Also, in the broader perspective of transmission ecology, it is known that not all infected snails contribute equally to transmission to further hosts. Therefore, surveys focusing on the final hosts are required to investigate actual cercarial transmission as xenomonitoring of intermediate host populations by itself does note address these aspects.

Finally, both the concept and protocol of the workflow are highly versatile for studying other snail-borne trematode communities. Depending on the specific target, primer concentrations in the multiplex PCR may be optimized to balance amplicon representation. Alternatively, depending on research interests, one may choose to remove snail or trematode markers to simplify the workflow and increase the coverage for the remaining markers. Altogether, our HTAS workflow opens new and exciting avenues to efficiently and accurately study trematode communities and parasite-parasite interactions within snail hosts.

## References

Alvarez, I., Wendel, J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution,* 29: 417–434. doi: 10.1016/s1055-7903(03)00208-2

Alzaylaee, H., Collins, R. A., Shechonge, A., Ngatunga, B. P., Morgan, E. R., & Genner, M. J. (2020). Environmental DNA-based xenomonitoring for determining Schistosoma presence in tropical freshwaters. *Parasites & vectors*, 13(1), 63. https://doi.org/10.1186/s13071-020-3941-6

Bakuza, J.S., Denwood, M.J., Nkwengulila, G., Mable, B.K. (2017). Estimating the prevalence and intensity of *Schistosoma mansoni* infection among rural communities in Western Tanzania: The influence of sampling strategy and statistical approach. *PLOS Neglected Tropical Diseases* 11(9): e0005937. https://doi.org/10.1371/journal.pntd.0005937

Ballenghien, M., Faivre, N. Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol.,* 15, 25 https://doi.org/10.1186/s12915-017-0366-6

Basch, P., DiConza, J. (1974). The Miracidium-Sporocyst Transition in *Schistosoma mansoni*: Surface Changes In vitro with Ultrastructural Correlation. *The Journal of Parasitology*, 60(6), 935-941. doi:10.2307/3278518

Blasco-Costa, I., Cutmore, S. C., Miller, T. L., Nolan, M. J. (2016). Molecular approaches to trematode systematics: 'best practice' and implications for future study. *Systematic parasitology*, 93(3), 295–306. https://doi.org/10.1007/s11230-016-9631-2

Born-Torrijos, A., Poulin, R., Raga, J. A., Holzer, A. S. (2014). Estimating trematode prevalence in snail hosts using a single-step duplex PCR: How badly does cercarial shedding underestimate infection rates? *Parasites and Vectors*, 7(1), 243. https://doi.org/10.1186/1756-3305-7-243 LK

Brown, D. S. (1994) Freshwater snails of Africa and their medical importance. 2nd edn, *Transactions of the Royal Society of Tropical Medicine and Hygiene*. UK Taylor & Francis Ltd. doi: 10.1016/0035-9203(81)90097-3.

Buddenborg, S.K., Kamel, B., Hanelt, B., Bu, L., Zhang, S.M., Mkoji, G., Loker, E. (2019) The in vivo transcriptome of *Schistosoma mansoni* in the prominent vector species *Biomphalaria pfeifferi* with supporting observations from *Biomphalaria glabrata*. *PLOS Neglected Tropical Diseases* 13(9): e0007013. https://doi.org/10.1371/journal.pntd.0007013

Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., Crandall, K. A. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution*, 3, 1312–1323. https://doi.org/10.1093/gbe/evr106

Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Elbrecht V., Steinke D., Ratnasingham S., de Waard J., Sones J., Zakharov E., Hebert, P. (2019) Metabarcoding a diverse arthropod mock community. *Mol Ecol Resour.*, 19: 711– 727. https://doi.org/10.1111/1755-0998.13008

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Carolus, H., Muzarabani, K. C., Hammoud, C., Schols, R., Volckaert, F., Barson, M., Huyse, T. (2019). A cascade of biological invasions and parasite spillback in man-made Lake Kariba. *The Science of the total environment*, 659, 1283–1292. https://doi.org/10.1016/j.scitotenv.2018.12.307

Caron, Y., Righi, S., Lempereur, L., Saegerman, C., Losson, B. (2011). An optimized DNA extraction and multiplex PCR for the detection of *Fasciola* sp. in lymnaeid snails. *Veterinary Parasitology*, 178(1), 93–99. https://doi.org/https://doi.org/10.1016/j.vetpar.2010.12.020

Chaudhry, U.N., Ali, Q., Rashid, I., Shabbir, M.Z., Abbas, M.A., Numan, M., Evans, M., Ashraf, K., Morrison, I., Morrison, L.J., Sargison, N. (2019). Development of a deep amplicon sequencing method to

determine the proportional species composition of piroplasm haemoprotozoa as an aid in their control. *Ticks and Tick-Borne diseases*, 10(6). https://doi.org/10.1016/j.ttbdis.2019.101276

Cooper, C., Keatley, S., Northover, A., Gofton, A. W., Brigg, F., Lymbery, A. J., Thompson, R. C. A. (2018). Next generation sequencing reveals widespread trypanosome diversity and polyparasitism in marsupials from Western Australia. *International Journal For Parasitology*: *Parasites And Wildlife*, 7(1), 58-67. https://doi.org/10.1016/j.ijppaw.2018.01.005

Cruaud, P., Rasplus, J., Rodriguez, L.J., Cruaud, A. (2016). High throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific reports*, 41948. https://doi.org/10.1101/073304

Curtis, J., Sorensen, R. E., Minchella, D. J. (2002). Schistosome genetic diversity: the implications of population structure as detected with microsatellite markers. *Parasitology*, 125 Suppl., S51–S59. https://doi.org/10.1017/s0031182002002020

Davison, A. (2000). The inheritance of divergent mitochondria in the land snail, *Cepaea nemoralis*. *Journal of Molluscan Studies*, 66(2), 143–147. https://doi.org/10.1093/mollus/66.2.143

Eppert, A., Lewis, F. A., Grzywacz, C., Coura-Filho, P., Caldas, I., Minchella, D. J. (2002). Distribution of schistosome infections in molluscan hosts at different levels of parasite prevalence. *The Journal of Parasitology*, 88(2), 232–236. https://doi.org/10.1645/0022-3395(2002)088[0232: DOSIIM]2.0.CO;2

Esch, G. W., Curtis, L., Barger, M. (2001). A perspective on the ecology of trematode communities in snails. *Parasitology*, 123(7), 57-75. doi:10.1017/S0031182001007697

Esch, G. W., Barger, M. A., Fellis, K. J. (2002). The Transmission of Digenetic Trematodes: Style, Elegance, Complexity. *Integrative and Comparative Biology*, 42(2), 304–312. https://doi.org/

GNU, P. (2007). Free Software Foundation. Bash (3.2. 48)[Unix shell program].

Gusman, A., Azuelos, C., Breton, S. (2017). No evidence of sex-linked heteroplasmy or doubly-uniparental inheritance of mtDNA in five snail species. *Journal of Molluscan Studies*, 83(1), 119–122. https://doi.org/10.1093/mollus/eyw034

Hamburger, J., Hoffman, O., Kariuki, H.C., Muchiri, E.M., Ouma, J.H., Koech, D.K., Sturrock, R.F., King, C.H. (2004). Large-scale, polymerase chain reaction-based surveillance of *Schistosoma haematobium* DNA in snails from transmission sites in coastal Kenya: a new tool for studying the dynamics of snail infection. *The American journal of tropical medicine and hygiene*, 71 6, 765-73 . https://doi.org/10.4269/ajtmh.2004.71.765

[dataset] Hammoud, C., Mulero, S., Van Bocxlaer, B., Boissier, J., Verschuren, D., Albrecht, C., Huyse, T. (2020). Datasets, FASTA files, R and bash scripts for " Simultaneous genotyping of snails and infecting trematode parasites using high-throughput amplicon sequencing. ", *Dryad data*, doi:10.5061/dryad.qnk98sfg1.

[dataset] Hammoud, C., Mulero S., Van Bocxlaer, B., Boissier, J., Verschuren, D., Albrecht, C., Huyse, T. (2020). Raw sequencing reads from " Simultaneous genotyping of snails and infecting trematode parasites using high-throughput amplicon sequencing. ", *NCBI SRAI*: PRJNA743924.

[dataset] Hammoud, C., Mulero S., Van Bocxlaer, B., Boissier, J., Verschuren, D., Albrecht, C., Huyse, T. (2020). Processed Amplicon Sequence Variants from " Simultaneous genotyping of snails and infecting trematode parasites using high-throughput amplicon sequencing. ", *NCBI GenBank*, Genbank accessions: MZ546824 - MZ546834 (COI1_snail); MZ540237 - MZ540247 (COI2_snail); MZ600069 - MZ600118 (ITS1_snail); MZ546798 - MZ546803 (NAD1_snail); MZ567152 - MZ567158 (COI1_trematode); MZ558220 - MZ558244 (COI2_trematode); MZ600119 - MZ600142 (ITS2_trematode); MZ546812 - MZ546823 (cytb_trematode); MZ546804 - MZ546811 (NAD1_trematode).

Harris, D. J., Crandall, K. A. (2000). Intragenomic Variation Within ITS1 and ITS2 of Freshwater Crayfishes (Decapoda: Cambaridae): Implications for Phylogenetic and Microsatellite Studies. *Mol. Biol. and Evol.*, 17(2), 284–291. https://doi.org/10.1093/oxfordjournals.molbev.a026308

Herbold, C. W., Pelikan, C., Kuzyk, O., Hausmann, B., Angel, R., Berry, D., Loy, A. (2015). A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. *Frontiers in microbiology*, 6, 731. https://doi.org/10.3389/fmicb.2015.00731

Hopkins, S. R., Ocampo, J. M., Wojdak, J. M., Belden, L. K. (2016). Host community composition and defensive symbionts determine trematode parasite abundance in host communities. *Ecosphere* 7(3):e01278. 10.1002/ecs2.1278

Hotez, P.J., Alvarado, M., Basáñez, M.G., Bolliger, I., Bourne, R., Boussinesq, M., Brooker, S.J., Brown, A. S., Buckle, G., Budke, C.M., Carabin, H., Coffeng, L.E., Fèvre, E.M., Fürst, T., Halasa, Y. A., Jasrasaria, R., Johns, N. E., Keiser, J., King, C. H., Lozano, R., Murdoch, M.E., O'Hanlon, S., Pion, S.D., Pullan, R.L., Ramaiah, K.D., Roberts, T., Shepard, D.S., Smith, J.L., Stolk, W.A., Undurraga, E. A., Utzinger, J., Wang, M., Murray, C. J., Naghavi, M. (2015). The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. *PLoS Negl Trop Dis.*, 8(7) doi:10.1371/journal.pntd.0002865

Jannotti-Passos, L., Magalhães, K., Carvalho, O., & Vidigal, T. (2006). Multiplex PCR for Both Identification of Brazilian *Biomphalaria* Species (Snaila: Planorbidae) and Diagnosis of Infection by *Schistosoma mansoni* (Trematoda: Schistosomatidae). *The Journal of Parasitology*, 92(2), 401-403. https://doi.org/10.1645/GE-593R1.1

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular ecology resources*, 19(1), 176–190. https://doi.org/10.1111/1755-0998.12951

Kamel, B., Laidemitt, M. R., Lu, L., Babbitt, C., Weinbaum, O. L., Mkoji, G. M., Loker, E. S. (2021). Detecting and identifying Schistosoma infections in snails and aquatic habitats: A systematic review. *PLoS neglected tropical diseases*, 15(3), e0009175. https://doi.org/10.1371/journal.pntd.0009175

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W.T., Wolfe, A.J., Braun, J., McWeeney, S. (2019). Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* 4:e00290-19. https://doi.org/10.1128/mSystems.00290-19.

Laidemitt, M.R., Anderson, L.C., Wearing, H.J., Mutuku, M.W., Mkoji, G.M., Loker, E.S. (2019). Antagonism between parasites within snail hosts impacts the transmission of human schistosomiasis. *eLife*, 8. https://doi.org 10.7554/eLife.50095

Le Clec'h, W., Diaz, R., Chevalier, F. D., McDew-White, M., Anderson, T. (2019). Striking differences in virulence, transmission and sporocyst growth dynamics between two schistosome populations. *Parasites & vectors*, 12(1), 485. https://doi.org/10.1186/s13071-019-3741-z

Lerch, A., Koepfli, C., Hofmann, N.E., Messerli C., Wilcox, S., Kattenberg, J., Betuela, I., O'Connor, L.,Mueller, I., Felger, I. (2017). Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genomics*, 18(1), 864. https://doi.org/DOI:10.1186/s12864-017-4260-y

Lockyer, A. E., Jones, C. S., Noble, L. R., Rollinson, D. (2004). Trematodes and snails: an intimate association. *Canadian Journal Of Zoology/Revue Canadien De Zoologie*, 82(2), 251-269. https://doi.org/10.1139/z03-215

Loker, E. S., Moyo, H. G., Gardner, S. L. (1981). Trematode–snail associations in nine non-lacustrine habitats in the Mwanza region of Tanzania. *Parasitology*, 83(2), 381–399. https://doi.org/DOI: 10.1017/S0031182000085383

Lopez, J. V, Yuhki, N., Masuda, R., Modi, W., O'Brien, S. J. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution,* 39(2), 174–190. https://doi.org/10.1007/bf00163806

Lusk, R.W. (2014) Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLOS ONE* 9(10): e110808. https://doi.org/10.1371/journal.pone.0110808

MacConaill, L. E., Burns, R. T., Nag, A., Coleman, H. A., Slevin, M. K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M. S., Ducar, M. D., Meyerson, M., Thorner, A. R. (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of

massively parallel sequencing. *BMC genomics*, 19(1), 30. https://doi.org/10.1186/s12864-017-4428-5

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10:12. https://doi.org/https://doi.org/10.14806/ej.17.1.200

Mesquita, S. G., Rodrigues-Luiz, G. F., Reis-Cunha, J. L., Cardoso, M. S., De Mendonça, C., Bueno, L. L., Fujiwara, R. T., Pinto, H. A., Caldeira, R. L., Bartholomeu, D. C. (2020). A multiplex PCR protocol for rapid differential identification of four families of trematodes with medical and veterinary importance transmitted by Biomphalaria Preston, 1910 snails. *Acta tropica*, 211, 105655. https://doi.org/10.1016/j.actatropica.2020.105655

Minchella, D., Sollenberger, K., Pereira De Souza, C. (1995). Distribution of schistosome genetic diversity within molluscan intermediate hosts. *Parasitology*, 111(2), 217-220. doi:10.1017/S0031182000064970

Moszczyńska, A., Locke, S.A., McLaughlin, J.D., Marcogliese, D.J., Crease, T.J. (2009). Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Molecular ecology resource*s, 9 Suppl s1, 75-82 . DOI: 10.1111/j.1755-0998.2009.02634.x.

Mulero, S., Boissier, J., Allienne, J.-F., Quilichini, Y., Foata, J., Pointier, J.-P., Rey, O. (2020). Environmental DNA for detecting *Bulinus truncatus*: A new environmental surveillance tool for schistosomiasis emergence risk assessment. *Environmental DNA*. 2: 161– 174. https://doi.org/10.1002/edn3.53

Nolan, M. J., Cribb, T. H. (2005). The use and implications of ribosomal DNA sequencing for the discrimination of trematode species. *Advances in Parasitology*, 60, 101–163. https://doi.org/10.1016/S0065-308X(05)60002-4

Norton, A., Rollinson, D., Richards, L., Webster, J. (2008). Simultaneous infection of *Schistosoma mansoni* and *S. rodhaini* in *Biomphalaria glabrata*: Impact on chronobiology and cercarial behaviour. *Parasites and Vectors*, 1(1), 43. https://doi.org/10.1186/1756-3305-1-43

Pages, J.-R., Théron, A. (1990). Analysis and comparison of cercarial emergence rhythms of *Schistosoma haematobium*, *S. intercalatum, S. bovis*, and their hybrid progeny. *Int. Journal for Parasitology*, 20(2), 193-197. https://doi.org/10.1016/0020-7519(90)90100-2

Palmer, J. M., Jusino, M. A., Banik, M. T., Lindner, D. L. (2018). Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, 6, e4925. https://doi.org/10.7717/peerj.4925

Parakatselaki, M. E., Saavedra, C., Ladoukakis, E. D. (2016). Searching for doubly uniparental inheritance of mtDNA in the apple snail *Pomacea diffusa*. *Mitochondrial DNA*, 27(6), 4000–4002. https://doi.org/10.3109/19401736.2014.989521

Parkin, E. J., Butlin, R. K. (2004). Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Molecular Biology and Evolution*, 21(8), 1595–1601. https://doi.org/10.1093/molbev/msh163

Pennance, T., Archer, J., Lugli, E. B., Rostron, P., Llanwarne, F., Ali, S. M., Amour, A. K., Suleiman, K. R., Li, S., Rollinson, D., Cable, J., Knopp, S., Allan, F., Ame, S. M., Webster, B. L. (2020). Development of a Molecular Snail Xenomonitoring Assay to Detect *Schistosoma haematobium* and *Schistosoma bovis* Infections in their *Bulinus* Snail Hosts. *Molecules*, 25(17), 4011. https://doi.org/10.3390/molecules25174011

Pinaud, S., Portela, J., Duval, D., Nowacki, F. C., Olive, M. A., Allienne, J. F., Galinier, R., Dheilly, N. M., Kieffer-Jaquinod, S., Mitta, G., Théron, A., Gourbal, B. (2016). A Shift from Cellular to Humoral Responses Contributes to Innate Immune Memory in the Vector Snail *Biomphalaria glabrata*. *PLoS pathogens*, 12(1), e1005361. https://doi.org/10.1371/journal.ppat.1005361

Pitchford, R. J., Meyling, A. H., Meyling, J., Du Toit, J. F. (1969). Cercarial shedding patterns of various schistosome species under outdoor conditions in the Transvaal, *Annals of Tropical Medicine & Parasitology,* 63:3, 359-371, DOI: 10.1080/00034983.1969.11686637

Portela, J., Duval, D., Rognon, A., Galinier, R., Boissier, J., Coustau, C., Mitta, G., Theron, A., Gourbal, B., (2013). Evidence for specific genotype-dependent immune priming in the lophotrochozoan *Biomphalaria glabrata* snail. *Journal of Innate Immunity*. 5(3):261-276. DOI: 10.1159/000345909.

Portet, A., Pinaud, S., Chaparro, C., Galinier, R., Dheilly, N.M., Portela, J., Charriere, G., Allienne, J.F., Duval, D., Gourbal, B. (2019) Sympatric versus allopatric evolutionary contexts shape differential immune response in *Biomphalaria / Schistosoma* interaction. *PLOS Pathogens* 15(3): e1007647. https://doi.org/10.1371/journal.ppat.1007647

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.  URL http://www.R-project.org/.

Rodriguez-Lanetty, M., Hoegh-Guldberg, O. (2002). The phylogeography and connectivity of the latitudinally widespread scleractinian coral *Plesiastrea versipora* in the Western Pacific. *Molecular Ecology*, 11(7), 1177–1189. https://doi.org/10.1046/j.1365-294X.2002.01511.x

Schols, R., Carolus, H., Hammoud, C., Mulero, S., Mudavanhu, A., Huyse, T. (2019). A rapid diagnostic multiplex PCR approach for xenomonitoring of human and animal schistosomiasis in a 'One Health' context. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 113(11), 722–729. https://doi.org/10.1093/trstmh/trz067

Schols, R., Mudavanhu, A., Carolus, H., Hammoud, C., Muzarabani, K.C., Barson, M., Huyse, T. (2020) Exposing the Barcoding Void: An Integrative Approach to Study Snail-Borne Parasites in a One Health Context. *Frontiers in Veterinary Science* 7:605280. doi: 10.3389/fvets.2020.605280

Schwanbek, A., Becker, W., Rupprecht, H. (1986). Quantification of parasite development in the host-parasite system *Biomphalaria glabrata* and *Schistosoma mansoni*. *Z. Parasitenkd*., 72(3), 365–373. https://doi.org/https://doi.org/10.1007/BF00928747

Signorell A., et mult. al. (2020). DescTools: Tools for Descriptive Statistics. R Package Version 0.99.32. https://cran.r-project.org/package=DescTools

Soldánová, M., Faltýnková, A., Scholz, T., Kostadinova, A. (2011). Parasites in a man-made landscape: Contrasting patterns of trematode flow in a fishpond area in Central Europe. *Parasitology*, 138(6), 789-807. doi:10.1017/S0031182011000291

Soldánová, M., Kuris, A. M., Scholz, T., Lafferty, K. D. (2012). The role of spatial and temporal heterogeneity and competition in structuring trematode communities in the great pond snail, *Lymnaea stagnalis* (L.). *The Journal of Parasitology*, 98(3), 460–471. https://doi.org/10.1645/GE-2964.1

Steinauer, M. L., Hanelt, B., Mwangi, I. N., Maina, G. M., Agola, L. E., Kinuthia, J. M., Mutuku, M. W., Mungai, B. N., Wilson, W. D., Mkoji, G. M., & Loker, E. S. (2008). Introgressive hybridization of human and rodent schistosome parasites in western Kenya. *Molecular ecology*, 17(23), 5062–5074. https://doi.org/10.1111/j.1365-294X.2008.03957.x

Steinauer, M. L., Hanelt, B., Agola, L. E., Mkoji, G. M., Loker, E. S. (2009). Genetic structure of *Schistosoma mansoni* in western Kenya: the effects of geography and host sharing. *International journal for parasitology*, 39(12), 1353–1362. https://doi.org/10.1016/j.ijpara.2009.04.010

Stothard, J.R., Brémond, P., Andriamaro, L., Loxton, N.J., Sellin, B., Sellin, E., Rollinson, D. (2000), Molecular characterization of the freshwater snail *Lymnaea natalensis* (Gastropoda: Lymnaeidae) on Madagascar with an observation of an unusual polymorphism in ribosomal small subunit genes. *Journal of Zoology*, 252: 303-315. https://doi.org/10.1111/j.1469-7998.2000.tb00625.x

Sword, G. A., Senior, L. B., Gaskin, J. F., Joern, A. (2007). Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera: Acrididae). *Systematic Entomology*, 32(3), 420–428. https://doi.org/10.1111/j.1365-3113.2007.00385.x

Théron, A., Touassem, R. (1989). *Schistosoma rodhaini*: intramolluscan larval development, migration and replication processes of daughter sporocysts. *Acta Tropica*, 46(1), 39–45. https://doi.org/https://doi.org/10.1016/0001-706X(89)90014-4

Théron, A., Gérard, C., Moné, H. (1992). Early enhanced growth of the digestive gland of *Biomphalaria glabrata* infected with *Schistosoma mansoni*: side effect or parasite manipulation? *Parasitology research*, 78(5), 445–450. https://doi.org/10.1007/BF00931703

Théron, A., Rognon, A., Gourbal, B., Mitta, G. (2014). Multi-parasite host susceptibility and multi-host parasite infectivity: a new approach of the *Biomphalaria glabrata/Schistosoma mansoni* compatibility polymorphism. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 26, 80–88. https://doi.org/10.1016/j.meegid.2014.04.025

Thiele, E.A., Minchella, D.J. (2013). Molecular assessment of trematode co-infection and intraspecific competition in molluscan intermediate hosts. *Molecular and biochemical parasitology*, 187 1, 52-9 . DOI: 10.1016/j.molbiopara.2012.12.003

Thomaz, D., Guiller, A., Clarke, B. C. (1996). Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society of London*. Series B: Biological Sciences, 263(1368), 363–368. https://doi.org/10.1098/rspb.1996.0056

Toledo, R., Fried, B. (2014) Digenetic Trematodes, *Advances in Experimental Medicine and Biology*. Springer. doi: 10.1007/978-1-4939-0915-5.

Van den Broeck, F., Geldof, S., Polman, K., Volckaert, F. A. M., Huyse, T. (2011). Optimal sample storage and extraction procotols for reliable multilocus genotyping of the human parasite *Schistosoma mansoni*. Infection, Genetics and Evolution. *Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 11(6), 1413–1418. https://doi.org/10.1016/j.meegid.2011.05.006

Vanhove, M. P., Tessens, B., Schoelinck, C., Jondelius, U., Littlewood, D. T., Artois, T., Huyse, T. (2013). Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes). *ZooKeys*, (365), 355–379. https://doi.org/10.3897/zookeys.365.5776

Vierna, J., Martínez-Lage, A., González-Tizón, A. M. (2009). Analysis of ITS1 and ITS2 sequences in *Ensis razor* shells: suitability as molecular markers at the population and species levels, and evolution of these ribosomal DNA spacers. *Genome*, 53(1), 23–34. https://doi.org/10.1139/G09-080

World Health Organization (2017) Integrating neglected tropical diseases into global health and development: fourth WHO report on neglected tropical diseases. Geneva. Licence: CC BY-NC-SA 3.0 IGO

World Health Organization. (2020). Ending the neglect to attain the Sustainable Development Goals – A road map for neglected tropical diseases 2021–2030. Geneva: Licence: CC BY-NC-SA 3.0 IGO.

Xu, J., Zhang, Q., Xu, X., Wang, Z., Qi, J. (2009). Intragenomic variability and pseudogenes of ribosomal DNA in stone flounder *Kareius bicoloratus. Molecular Phylogenetics and Evolution*, 52(1), 157–166. https://doi.org/10.1016/j.ympev.2009.03.031

**Data accessibility**

- DNA sequences: NCBI SRA: PRJNA743924. Genbank accessions: MZ546824 - MZ546834 (COI1_snail); MZ540237 - MZ540247 (COI2_snail); MZ600069 - MZ600118 (ITS1_snail); MZ546798 - MZ546803 (NAD1_snail); MZ567152 - MZ567158 (COI1_trematode); MZ558220 - MZ558244 (COI2_trematode); MZ600119 - MZ600142 (ITS2_trematode); MZ546812 - MZ546823 (cytb_trematode); MZ546804 - MZ546811 (NAD1_trematode).

- bash and R scripts: Dryad https://doi.org/10.5061/dryad.qnk98sfg1

- Infection data, alignments for primers design and input files for analyses: Dryad https://doi:10.5061/dryad.qnk98sfg1

**Ethics statements**

Housing, feeding and animal care followed the French ethical standards established in the writ of February 1st, 2013 (NOR: AGRG1238753A). Animal experiments were authorized following permit A66040 of the Host-Pathogen-Environment Interactions lab (UPVD) and SM, CH and JB have the required certificates to act as experimenters (authorization 007083, decree 87–848).

**Author's contributions**

Authors' contributions: TH, BVB, and CH conceived the study. CH, SM, JB and TH designed the study protocols. SM, CH and JB conducted the experimental infections. CH carried out the HTAS sequencing assay and analyzed and interpreted the data. CH and BVB drafted the manuscript with the help of TH. CH, SM, TH, BVB, DV, CA, CH and JB critically revised the manuscript. All authors read and approved the final manuscript.

**Tables**

Table 1: Final primer pairs designed for the HTAS workflow, with sequence, marker name, expected amplicon size (which includes primers), and the final concentration of each primer in the multiplex primer mix. The last column indicates the average number of reads per sample attributed to the marker (after trimming), with standard deviation. For snail markers, all specimens (n = 417) were included in the calculation of read support whereas only infected specimens were considered for the read support of

| Primer name | Primer sequence (5′–3′) | Marker | Concentration (nM) | Expected length (bases) | Read support |
|---|---|---|---|---|---|
| COI1_snailF | TAATTWATTGTTACDGCWCATGC | COI (1) | 100 | 514 | 1071 |
| COI1_snailR | CWCCHCCTGCWGGRTCAAA | | 100 | | ± 641 |
| COI2_snailF | CARCATTTRTTTTGRTTTTTTGG | COI (2) | 100 | 380 | 6898 |
| COI2_snailR | GATARWACAATHCCWGTWARHCCHCC | | 100 | | ± 11090 |
| ITS1_snailF | GCTCGCCGGGTCGCVA | ITS1 | 25 | 333 - 549 | 1795 |
| ITS1_snailR | GACCCTCAGACGGRCGT | | 25 | | ± 1851 |
| NAD1_snailF | GGNTATATTCARATTCGWAARGG | NAD1 | 170 | 533 - 536 | 1760 |
| NAD1_snailR | CCWGANACTAATTCNGAYTCNCC | | 170 | | ± 978 |
| COI1_tremF | CNATGATNTTNTTYTTYTTDATGCC | COI (1) | 180 | 439 - 453 | 1425 |
| COI1_tremR | GMDSWACHAAAWTTHCGATCAAA | | 180 | | ± 1020 |
| COI2_tremF | GDGGBGATCCNDTDTTDTTHCA | COI (2) | 200 | 469 - 472 | 1152 |
| COI2_tremR | ACATAATGAAARTGAGCNAYMACRA | | 200 | | ± 494 |
| ITS2_tremF | CAAHAAGTCGTGGMTTGG | ITS2 | 25 | 364 - 879 | 713 |
| ITS2_tremR | CCTAAACACCACATTGCCT | | 25 | | ± 495 |
| NAD1_tremF | TTCGTAAGGGDCCBAAHAAGGT | NAD1 | 100 | 505 - 511 | 121 |
| NAD1_tremR | CTCTCMGMCTCMSMRTAATC | | 100 | | ± 66 |
| cytb_tremF | GRCATCARATGTCTTATTGRGCD | cytb | 100 | 412 - 427 | 119 |
| cytb_tremR | ARAAARTAYCACTCHGGCTTWAT | | 100 | | ± 60 |
| custom_tailF | CTTTCCCTACACGACGCTCTTCCGATCT | - | - | - | |
| custom_tailR | CTGGAGTTCAGACGTGTGCTCTTCCGATCT | - | - | - | |

trematode markers (n = 229).

Table 2: Average proportion of reads (mean ± standard deviation for the Amplicon Sequence Variant (ASV) class / total reads per sample, in %) that were amplified for the nine primer markers per ASV category. Values in bold indicate statistically significant differences between the infected and uninfected specimens. Marker-specific thresholds for ASV filtering are indicated in the last column.

| Taxon | Marker | Primary ASV | | Secondary ASVs | | Contamination | | Threshold |
|---|---|---|---|---|---|---|---|---|
| | | Infected | Uninfected | Infected | Uninfected | Infected | Uninfected | |
| *Biompalaria glabrata* n = 417 | COI1_snail | 9.0 ± 2.4 | 8.4 ± 1.8 | **1.7 ± 0.7** | **0.9 ± 1.4** | 0.5 ± 0.3 | 0.4 ± 0.4 | 1.5 |
| | COI2_snail | **23.4 ± 9.1** | **54.4 ± 6.2** | **2.2 ± 3.1** | **1.3 ± 2.6** | 1.5 ± 1.3 | 1.4 ± 1.5 | 6.5 |
| | ITS1_snail | **4.5 ± 2.7** | **11.6 ± 4** | **2.2 ± 2.3** | **5.8 ± 4.9** | 0.2 ± 0.1 | 0.1 ± 0.1 | 1.0 |
| | NAD1_snail | **8.7 ± 2.9** | **10.9 ± 2** | **3.0 ± 1.6** | **1.9 ± 1.4** | 0.3 ± 0.2 | 0.3 ± 0.2 | 1.0 |
| | COI1_trem | **14.2 ± 4.3** | **2.1 ± 1.6** | 1.0 ± 1.1 | 1.2 ± 1.2 | 1.1 ± 1.4 | 0.9 ± 1.2 | 9.0 |
| | COI2_trem | **18.6 ± 8.1** | **1.5 ± 1** | **3.9 ± 3.2** | **1.0 ± 0.8** | 1.1 ± 1.4 | 0.8 ± 1.2 | 7.0 |
| | ITS2_trem | **7.6 ± 3.3** | **0.6 ± 0.4** | **1.4 ± 0.7** | **0.5 ± 0.3** | 0.5 ± 0.4 | 0.3 ± 0.3 | 2.0 |
| | NAD1_trem | **1.8 ± 0.8** | **0.3 ± 0.2** | 0.0 | 0.0 | 0.4 ± 0.4 | 0.2 ± 0.2 | 1.5 |
| | Cytb_trem | **1.8 ± 0.7** | **0.4 ± 0.3** | 0.0 | 0.0 | 0.5 ± 0.4 | 0.3 ± 0.2 | 1.5 |
| *Bulinus* spp. n = 7 | COI1_snail | 2.7 ± 1.2 | - | 0.5 ± 0.04 | - | 0.03 ± 0.008 | - | 0.04 |
| | COI2_snail | 19.7 ± 21.0 | - | 0.4 ± 0.4 | - | 1.2 ± 1.4 | - | 3.52 |
| | ITS1_snail | 8.7 ± 2.0 | - | 9.0 ± 9.3 | - | 0.06 ± 0 | - | 0.06 |
| | NAD1_snail | 0.8 ± 0.5 | - | 0 | - | 0 | - | 0 |
| | COI1_trem | 6.8 ± 4.4 | - | 2.0 ± 2.9 | - | 0.2 ± 0.1 | - | 0.4 |
| | COI2_trem | 26.8 ± 7.0 | - | 10.0 ± 7.7 | - | 0.2 ± 0.3 | - | 0.83 |
| | ITS2_trem | 19.0 ± 10.2 | - | 9.6 ± 7.3 | - | 0.3 ± 0.3 | - | 0.53 |
| | NAD1_trem | 2.6 ± 2.5 | - | 1.5 ± 0 | - | 0.8 ± 0 | - | 0.8 |
| | Cytb_trem | 4.5 ± 2.5 | - | 2.3 ± 1.2 | - | 0.1 ± 0.2 | - | 0.32 |
| *Radix natalensis* n = 7 | COI1_snail | 0.08 ± 0.06 | - | 0 | - | 0.02 ± 0.01 | - | 0.04 |
| | COI2_snail | 73.7 ± 25.9 | - | 0.02 ± 0.01 | - | 0.05 ± 0.04 | - | 0.07 |
| | ITS1_snail | 2.9 ± 2.1 | - | 1.3 ± 0.6 | - | 0 | - | 0 |
| | NAD1_snail | 0.3 ± 0.3 | - | 0.2 ± 0 | - | 0 | - | 0 |
| | COI1_trem | 3.0 ± 0 | - | 0.04 ± 0.06 | - | 0.009 ± 0.003 | - | 0.02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| COI2_trem | 6.4 ± 6.0 | - | 2.3 ± 2.6 | - | 0.05 ± 0.05 | - | 0.06 |
| ITS2_trem | 10.2 ± 15.0 | - | 2.8 ± 4.6 | - | 0.02 ± 0.02 | - | 0.03 |
| NAD1_trem | 0.04 ± 0 | - | 0.003 ± 0 | - | 0.003 ± 0.001 | - | 0.01 |
| Cytb_trem | 0.03 ± 0.03 | - | 0.02 ± 0.01 | - | 0.004 ± 0.002 | - | 0.01 |

Table 3: Rates of parasite infection (in % of n exposed snail specimens) in function of time (in days) for the time-series experiment and controls. SmBRE refers to *Schistosoma mansoni* strain BRE and SmLE to *S. mansoni* strain LE.

| | Time series | | | | | | |
|---|---|---|---|---|---|---|---|
| Strain | SmBRE | | | | | | |
| Time (days) | 2 (n = 20) | 6 (n = 20) | 10 (n = 20) | 16 (n = 20) | 20 (n = 20) | 30 (n = 20) | 40 (n = 22) |
| Rate of infection | 0.0 | 20.0 | 25.0 | 45.0 | 40.0 | 45.0 | 9.1 |
| | **Controls** | | | | | | |
| Strain | SmBRE | | SmLE | | *Schistosoma rodhaini* | Not exposed | Uninfected exposed (SmBRE) |
| Time (days) | 40 (n = 27) | | 40 (n = 26) | | 40 (n = 28) | 40 (n = 5) | 0 (n = 5) |
| Rate of infection | 96.3 | | 92.3 | | 82.1 | 0.0 | 0.0 |

Table 4: Expected (see Théron et al. 2014) and observed proportions of infections by the different strains/species in the co-infection experiment, based respectively on the proportions observed in the controls and on the genotyping results for the co-exposed snails. P-values are provided for the G-test with null hypothesis of random co-infection, i.e. independent infection histories for both species/strains included. SmBRE refers to *Schistosoma mansoni* strain BRE and SmLE to *S. mansoni* strain LE.

| Exposure | *Schistosoma mansoni* BRE + *Schistosoma mansoni* LE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days | 2 (n= 32) | | | | 10 (n= 32) | | | | 40 (n= 27) | | | |
| Strain present | None | Only SmBRE | Only SmLE | Both | None | Only SmBRE | Only SmLE | Both | None | Only SmBRE | Only SmLE | Both |
| Expected | 32.0 | 0.0 | 0.0 | 0.0 | 0.1 | 2.5 | 1.2 | 28.3 | 0.1 | 2.1 | 1.0 | 23.8 |
| Observed | 32 | 0 | 0 | 0 | 2 | 5 | 6 | 19 | 1 | 4 | 16 | 6 |
| p-value | 1 | | | | 0.04 | | | | $10^{-6}$ | | | |
| Exposure | *Schistosoma mansoni* LE + *Schistosoma rodhaini* | | | | | | | | | | | |
| Days | 2 (n= 32) | | | | 10 (n= 31) | | | | 40 (n= 30) | | | |
| Strain present | None | Only SmLE | Only *S. rodhaini* | Both | None | Only SmLE | Only *S. rodhaini* | Both | None | Only SmLE | Only *S. rodhaini* | Both |
| Expected | 32.0 | 0.0 | 0.0 | 0.0 | 0.4 | 5.1 | 2.0 | 23.4 | 0.4 | 5.0 | 2.0 | 22.6 |
| Observed | 29 | 0 | 3 | 0 | 0 | 28 | 0 | 3 | 1 | 26 | 0 | 3 |
| p-value | 0.23 | | | | $10^{-8}$ | | | | $10^{-8}$ | | | |

Table 5: Proportion of specimens per snail taxon for which each of the four snail markers was successfully sequenced (in %). The proportion of samples that contained secondary Amplicon Sequence Variants, which indicates marker polymorphism, is indicated in brackets (in %, absence of value = 0).

| Taxon | Snail marker | | | |
|---|---|---|---|---|
| | COI1 | COI2 | ITS1 | NAD1 |
| *Biomphalaria glabrata* (n = 417) | 100.0 (8.2) | 100.0 (2.4) | 99.5 (44.1) | 100.0 (52) |
| *Bulinus globosus* (n = 1) | 100.0 (100.0) | 100.0 | 100.0 (100.0) | 0.0 |
| *Bulinus truncatus* (n = 3) | 100.0 | 66.7 | 100.0 (100.0) | 66.7 |
| *Bulinus tropicus* (n = 2) | 100.0 | 100.0 | 100.0 (100.0) | 100.0 |
| *Bulinus forskalii* (n = 1) | 100.0 | 0.0 | 100.0 (100.0) | 0.0 |
| *Radix natalensis* (n = 7) | 42.9 | 100.0 | 100.0 (100.0) | 42.9 (14.3) |

Table 6: Proportion of parasite infections per snail and trematode taxon for which each of the five trematode markers were successfully sequenced (in %). The proportion of samples that contained secondary Amplicon Sequence Variants, which indicates intragenomic variation, is indicated in brackets (in %, absence of variation = 0).

| Snail host | Parasite | Infection type | Trematode marker | | | | |
|---|---|---|---|---|---|---|---|
| | | | COI1 | COI2 | ITS2 | NAD1 | cytb |
| *Biomphalaria glabrata* | *S. mansoni* LE | Single infection (n = 100) | 97.0 | 99.0 | 100.0 (22.0) | 83.0 | 84.0 |
| | | Co-infection (n = 31) | 96.8 | 93.5 | 96.8 | 58.1 | 64.5 |
| | *S. mansoni* Bre | Single infection (n = 72) | 80.6 | 97.2 | 100.0 (19.4) | 46.8 | 46.8 |
| | | Co-infection (n = 25) | 100.0 | 100.0 | 100.0 | 56.0 | 60.0 |
| | *S. rodhaini* | Single infection (n = 26) | 57.7 | 46.2 | 100.0 | 0.0 | 0.0 |
| | | Co-infection (n = 6) | 16.7 | 0.0 | 100.0 | 0.0 | 0.0 |
| *Bulinus* spp. | *S. haematobium* | Co-infection (n = 1) | 100.0 | 0.0 | 0.0 | 0.0 | 100.0 |
| | Paramphistomidae sp. | Co-infection (n = 1) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | *Uvulifer* sp. | Co-infection (n = 1) | 100.0 | 100.0 | 100.0 | 0.0 | 0.0 |
| | Plagiorchioidea sp. | Co-infection (n = 1) | 0.0 | 0.0 | 100.0 | 0.0 | 100.0 |
| | *Orientocreadium* sp. | Single infection (n = 2) | 0.0 | 0.0 | 100.0 (100.0) | 0.0 | 0.0 |
| | Telorchiidae *sp.* | Single infection (n = 1) | 0.0 | 0.0 | 100.0 (100.0) | 0.0 | 0.0 |
| | Plagiorchioidea sp. II | Single infection (n = 1) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | *Haematoloechus* sp. | Co-infection (n = 1) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | *Calicophoron* sp. | Co-infection (n = 1) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| *Radix natalensis* | *Fasciola gigantica* | Co-infection (n = 1) | 0.0 | 100.0 | 100.0 | 100.0 | 0.0 |
| | *Cotylurus sp.* | Single infection (n = 3) | 0.0 | 67.0 | 100.0 | 0.0 | 100.0 |
| | | Co-infection (n = 4) | 0.0 | 75.0 (25.0) | 75.0 | 0.0 | 75.0 |
| | *Orientocreadium* sp. | Co-infection (n = 3) | 0.0 | 67.0 | 100.0 | 0.0 | 0.0 |

| | | | | (100.0) | | |
|---|---|---|---|---|---|---|
| Plagiorchiida sp. | Co-infection (n = 1) | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 |

**Figure captions**

Figure 1: Schematic view of possible trematode life cycles in the freshwater environment. Trematodes parasitize their first intermediate snail host as a sporocyst or (if having a mouth) a redia. After reproducing asexually within the host, they develop into cercariae (motile larvae) that are released in the water. Cercariae may either infect a second intermediate host or a final host. Within the environment or the second and possibly third intermediate hosts, trematodes may survive encysted as metacercariae, which infect the subsequent host through ingestion. Within the final host, the larvae transform into mature adult worms and start producing eggs. Once the eggs are released into the water the eggs may either hatch into motile larvae called miracidia or they may be ingested by a snail. Parasite transmission between hosts can either be active, if the cercariae or miracidia actively seek their host, or passive through trophic transmission. Trematode larvae may also infect snail species in which they are unable to develop (incompetent host), which results in a dead-end. Note that host species and cercarial morphotypes are non-exhaustive.

Figure 2: A) Workflow of our controlled infection experiments, up to parasite exposure. From top to bottom, the represented parasite life forms are cercariae, eggs and miracidia, which are then used to infect snails (see main text). B) Graphic representation of the infection experiments, showing the number of snails exposed to the various parasite strains (N) and the number of individuals sampled (n) at different times (T, in days).

Figure 3: Marker-specific proportion of reads supporting primary and secondary Amplicon Sequence Variants (ASVs) of the four snail and five trematode markers, as obtained from the 417 laboratory-infected *Biomphalaria glabrata* specimens.

Figure 4: Comparison between the proportion of reads attributed to *Schistosoma mansoni* Amplicon Sequence Variants in infected *Biomphalaria glabrata* snails throughout the time-series experiment (solid

line; error bars reflect standard deviation on the mean) and the volume of parasite biomass within the snail's digestive gland as reported by Théron et al. (1992; dashed line).

Figure 5: Comparison of the proportion of sequence reads supporting the COI2_trematode Amplicon Sequence Variants of *Schistosoma mansoni* strain BRE (SmBRE) and *S. mansoni* strain LE (SmLE) in snails infected by either of the parasites alone or co-infected (controls and co-infection experiments included), on day 10 and day 40 after exposure. COI2_trematode was used because it is the only marker that allows distinction between both strains.

Figure 6: Comparison of the proportion of sequence reads supporting the ITS2_trematode Amplicon Sequence Variants of *Schistosoma mansoni* strain LE (SmLE) and *Schistosoma rodhaini* in snails infected by either of the parasites alone or co-infected (controls and co-infection experiments included), on day 10 and day 40 after exposure. ITS2_trematode was used because it was the most consistently represented amplicon.
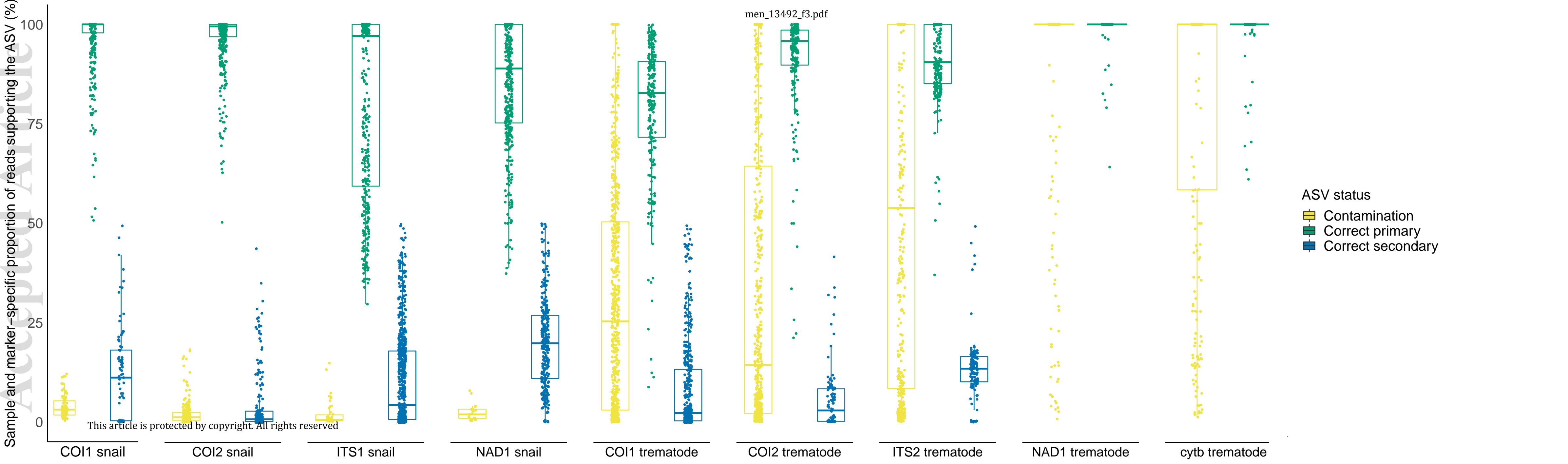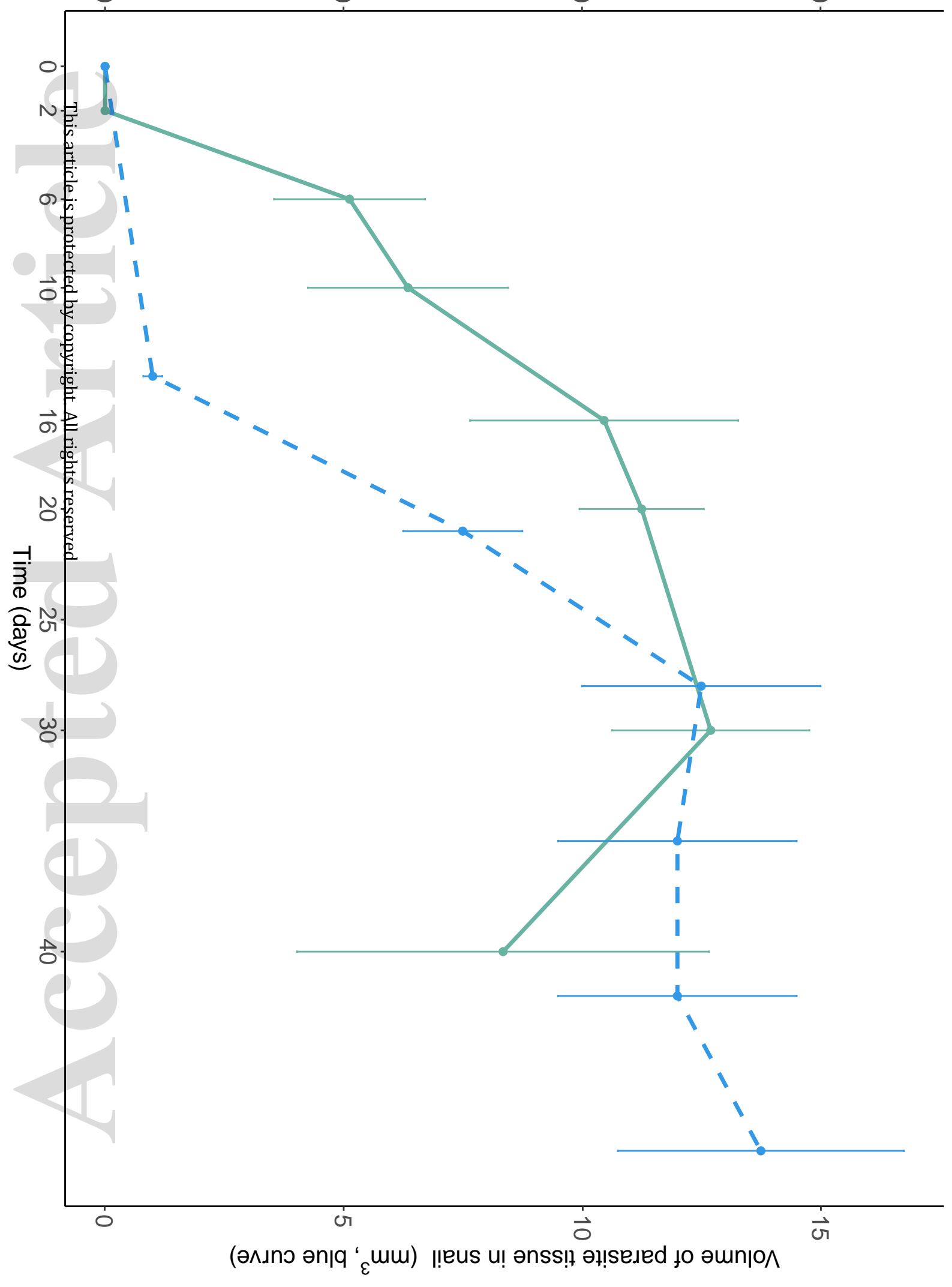
A

*S. mansoni* Bre   *S. mansoni* LE   *S. rodhaini*

Egg collection

Egg hatching

Snail exposure
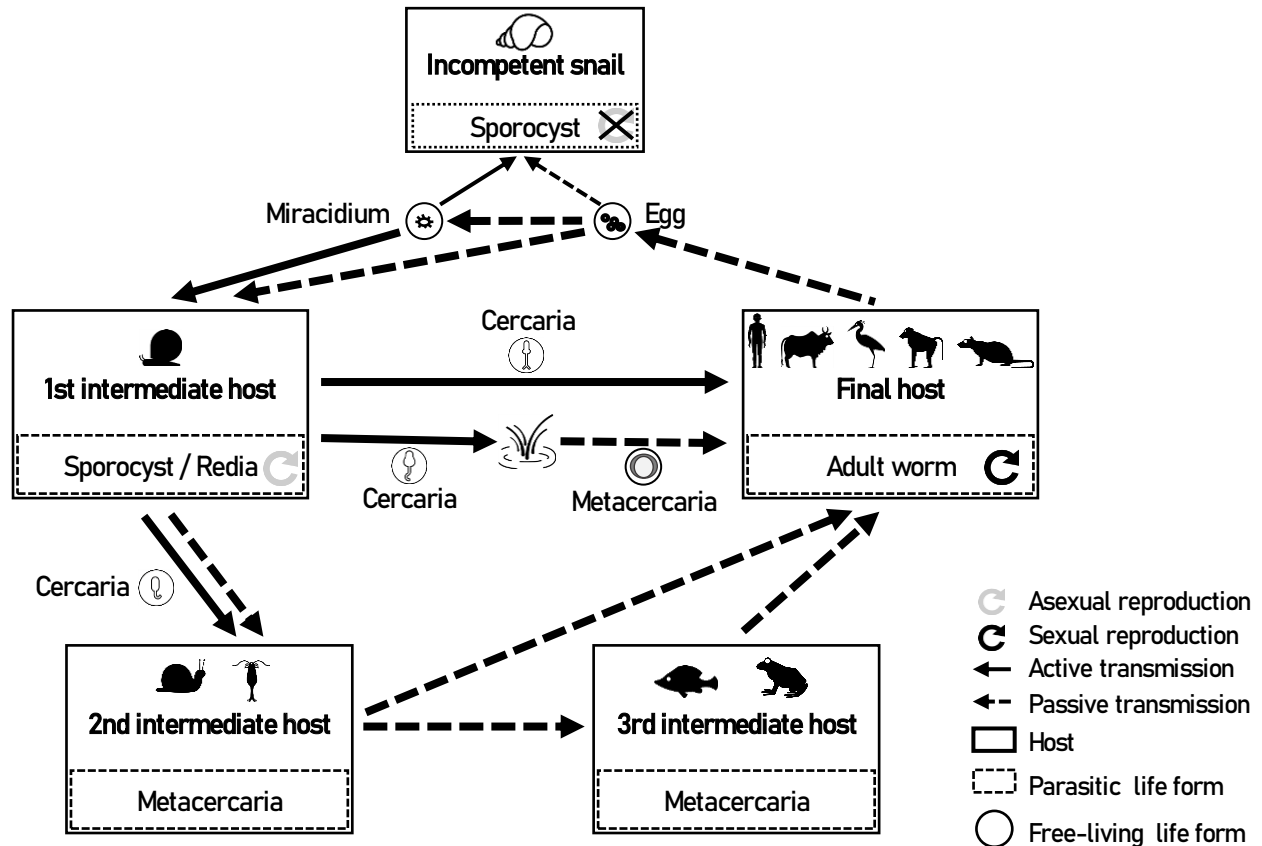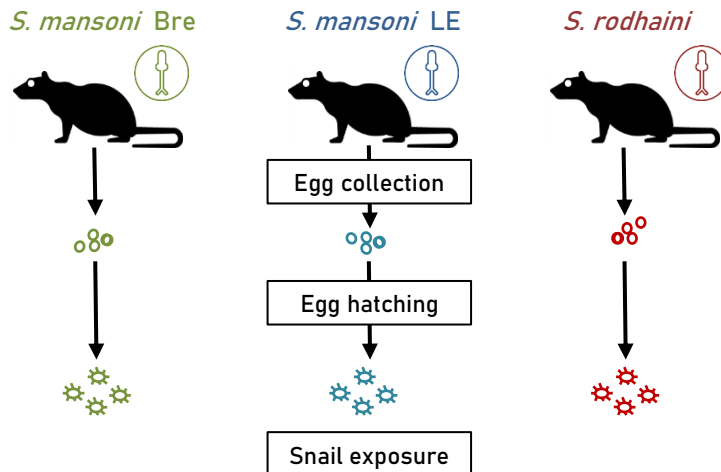
B

| | Time series | | | | | | | Controls | | | | | Co-infections | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☼ x 1 | | | | | | | | ☼ x 10  ☼ x 10  ☼ x 10 | | | | | ☼ x 10  ☼ x 10 | | | ☼ x 10  ☼ x 10 | | |
| T | 2 | 6 | 10 | 16 | 20 | 30 | 40 | 40 | 40 | 40 | 40 | 0 | 2 | 10 | 40 | 2 | 10 | 40 |
| N | | | 144 | | | | | 30 | 30 | 30 | 5 | 5 | | 96 | | | 96 | |
| n | 20 | 20 | 20 | 20 | 20 | 20 | 22 | 26 | 27 | 28 | 5 | 5 | 32 | 31 | 30 | 32 | 32 | 27 |

men_13492_f3.pdf

10 days after exposure

40 days after exposure

10 days after exposure

40 days after exposure

Propotion of reads from parasite ASV's

Species

SmLE

S. rodhaini

SmLE only    S. rodhaini only    Both    SmLE only    S. rodhaini only    Both

Infection status