# Improved Molecular Diagnostics for Soil-Transmitted Helminths

# IMPROVED MOLECULAR DIAGNOSTICS FOR SOIL-TRANSMITTED HELMINTHS

A Dissertation Presented

by

NILS PILOTTE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2020

Molecular and Cellular Biology

**IMPROVED MOLECULAR DIAGNOSTICS FOR SOIL-TRANSMITTED HELMINTHS**

A Dissertation Presented

by

NILS PILOTTE

Approved as to style and content by:

_____
Steven A. Williams, Chair

_____
Michele M. Klingbeil, Member

_____
M. Sloan Siegrist, Member

_____
Patrick Flaherty, Member

_____
Scott C. Garmin,
MCB Graduate Program Director

## DEDICATION

To my parents Claude and LeaAnn Pilotte who taught me from the youngest of ages that the only way to approach anything is with hard work and determination. Since that day in high school that I woke up to the newspaper clipping of this quote, it has been my unwavering inspiration:

> *It is not the critic who counts; not the man who points out how the strong man stumbles, or where the doer of deeds could have done them better. The credit belongs to the man who is actually in the arena, whose face is marred by dust and sweat and blood; who strives valiantly; who errs, who comes short again and again, because there is no effort without error and shortcoming; but who does actually strive to do the deeds; who knows great enthusiasms, the great devotions; who spends himself in a worthy cause; who at the best knows in the end the triumph of high achievement, and who at the worst, if he fails, at least fails while daring greatly, so that his place shall never be with those cold and timid souls who neither know victory nor defeat.*
>
> - *Theodore Roosevelt*

And to Mémère Pilotte. It's certainly taken me long enough, but we have a Doctor (of some sort) in the family...

## ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor, mentor, colleague, and friend Dr. Steven Williams. I remember walking up the stairs to the fourth floor of Sabin-Reed Hall all those years ago to meet with you and interview about a possible Master's Degree. Nervous and unsure of my abilities, you started me on a path that has led me here and has provided me with the sturdiest of foundations for what is to come next. It's been a long and windy road, but the journey has been a blast.

To my dissertation committee members Dr. Michele Klingbeil, Dr. Sloan Siegrist and Dr. Patrick Flaherty. Your encouragement, support, motivation, and willingness to listen has been invaluable and is most sincerely appreciated. Never once have you made me feel like an obligation or one of your countless responsibilities.

To my undergraduate advisor Dr. Janet Williams. At a time when I was unsure of my direction, you encouraged me to pursue a Master's Degree with your husband Steve. I almost said "no" and find myself wondering how different things might have been. I owe two jobs and three degrees to Steve and yourself.

To the late, Sr. Mary Lou Wright (Ph.D.). The first, and most demanding PI that I ever had. You were a scientist in the truest sense of the word. Dedicated, passionate, methodical, and ruthless. Working for you taught me that while fun and more rewarding than could possibly be imagined, science is not for the faint of heart.

To my "lab moms" Dr. Sandra Laney and Dr. Lori Saunders. You've encouraged and supported me in more ways than I can count and have taught me some pretty great science along the way. And Sandra, I'm sorry if my music is too loud.

To Gaby, my first SAW lab mate. You are the best of friends and have always been there for me when I've needed you. I'm not sure I would have made it through the Master's Degree program without you and I find it absolutely amazing that I was able to negotiate the paperwork of a Ph.D. program without your reminders and guidance!

To Sue Haynes. We infuriate each other, but I'm not sure I'd have it any other way. Ok, maybe…

To Dr. Eric Ottesen and Dr. Patrick Lammie. Despite your accomplishments, titles, and countless obligations, you have always treated me as a colleague and an equal. In a field where ego can all too often be a driving force, you remind me on a daily basis why the Neglected Tropical Disease community is such a special place to be. The NTD crowd is an exceptional one and I could not be more proud to be a part of it.

To my wife and fellow MCB student Allison. We WILL find a way to publish at least one paper together ☺

## ABSTRACT

### IMPROVED MOLECULAR DIAGNOSTICS FOR SOIL-TRANSMITTED HELMINTHS

MAY 2020

NILS PILOTTE

B.A., ELMS COLLEGE

M.S., SMITH COLLEGE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Steven A. Williams

Current World Health Organization recommendations for the diagnosis of soil-transmitted helminths (STH) rely on antiquated microscopy-based techniques that lack both diagnostic sensitivity and specificity. While sufficient for providing rough estimates of infection frequency and intensity within high prevalence settings, these techniques lack the capacity to effectively estimate infection levels following successful intervention efforts, as worm burdens decline and community prevalences decrease. Furthermore, an expanding body of evidence is suggesting that microscopy-based misdiagnosis of infection is likely a larger concern then previously believed. As such, with an increase in programmatic support for transmission interruption and an escalating belief in the possibility of regional eradication, recognition of the need for improved diagnostics is expanding. Such diagnostics are critically important for accurately measuring intervention successes, and for making determinations about where and when interventions can be

discontinued allowing for the re-prioritization of resources. Given the stakes and acknowledging this need, recent years have witnessed the growth of molecular diagnostic development efforts, centering primarily upon the creation of real-time PCR-based assays. However, while a variety of assays have been developed, these assays have all utilized suboptimal DNA targets, typically exploiting ribosomal and/or mitochondrial sequences for parasite detection.

By coupling next-generation sequencing with bioinformatics-based approaches to the analysis of raw sequencing read data, the selection of optimal DNA-based molecular targets becomes possible. Software, such as RepeatExplorer can be utilized to identify high copy-number, species-specific, repetitive DNA elements within a pathogen. These repeat sequences can then be utilized to design sensitive and specific real-time PCR assays. Utilizing such an approach, we have designed, developed, and extensively validated a panel of assays facilitating the improved detection of the human infecting STH. These assays are proving useful in a number of settings, gaining widespread traction within the operational research community and helping to shape and define future intervention strategies. Furthermore, our assays continue to highlight the inadequacies of alternative diagnostic methods, illustrating the potential challenges and risks for misdiagnosis associated with the use of both microscopy-based diagnostics, and assays targeting less sensitive, less specific genomic regions.

## TABLE OF CONTENTS

Page

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 The soil-transmitted helminths

Spanning 120 countries and placing an estimated one quarter of the world's total population at risk (Figure 1.1), the soil-transmitted helminths (STH) are a leading cause of morbidity and of lost economic productivity [1,2]. Categorized and named based upon the important role played by soil in their transmission cycles [3] (Figure 1.2), the human-infecting STH (roundworm - *Ascaris lumbricoides*; hookworm - *Necator americanus*, *Ancylostoma duodenale*, and *Ancylostoma ceylanicum*; whipworm - *Trichuris trichiura*; and threadworm - *Strongyloides stercoralis*) are estimated to result in the annual loss of over 2 million healthy-life-years lived (as measured by disability adjusted life years [DALYs]) [4], with some experts suggesting the burden is significantly greater [5]. While infrequently fatal, morbidity associated with STH infections is substantial, and is most pronounced in school-aged children, resulting in lethargy, wasting, growth stunting, chronic anemia, poor school attendance, and underperformance vs. peers in the educational and cognitive domains of learning, and memory [6]. Physical performance is also impacted, with infected children demonstrating reduced capacities on a variety of fitness-related measures [7, 8]. As the prevalence of STH infection is highest in economically-disadvantaged locations [9–11], these wide-ranging health and wellness impacts perpetuate the cycle of poverty, reinforcing existing disparities and curtailing the potential of young children living in disadvantaged locations [7].

In addition to its significant impacts on childhood health, development, and achievement, STH infection, primarily with hookworm and *T. trichiura*, is correlated with increased risk for anemia during pregnancy. This increased risk results in higher rates of pregnancy-related and child-delivery-related mortality [12, 13]. Growing evidence also suggests that STH co-infection may increase susceptibility to other diseases such as HIV and malaria [14, 15], while simultaneously exacerbating morbidity and increasing rates of mortality upon the establishment of co-infection [10]. Furthermore, the efficacy of vaccines providing protection against a variety of dangerous pathogens may also be reduced as a result of STH infection [10, 16]. Taken together, these deleterious impacts greatly reduce the capacity for work among populations already facing significant disadvantages. This results in extreme consequences for economic productivity, with annual losses estimated to reach US$3.5 billion in sub-Saharan Africa alone [17].

Given the broad, and multi-faceted impact which STH disease has on global health and economic productivity, international recognition of the need for organized and standardized intervention efforts is growing [18–21]. Current efforts typically involve the mass drug administration (MDA)-based treatment of children with benzimidazoles, well-tolerated anthelminthic chemotherapeutics which are given to a target population irrespective of the individual infection statuses among the constituents [22–24]. When intervention coverage and compliance are sufficient, these efforts are effective for de-worming infected children under treatment. However, as the majority of such interventions do not target large segments of the adult population for whom infection is generally considered to be less detrimental to overall health, human reservoirs of infection remain within the treated communities and re-infection of children occurs rapidly [25–28]. Odds

for reinfection are further bolstered by environmental reservoirs, as shed *A. lumbricoides* and *T. trichiura* eggs may remain viable for extended periods of time outside of the host [29–33]. This results in the persistence of infectious eggs for a period of years even if community-wide treatment and intervention efforts are able to effectively clear infection within an afflicted region. Recognizing these shortcomings, operational research-based studies are now being conducted to evaluate whether transmission interruption, and regional elimination of STH disease may be possible through the expansion and revision of MDA-based treatment strategies, in combination with additional interventions such as improvements to water, sanitation, and hygiene (WASH) [34–39].

### 1.2 Soil-transmitted helminth diagnostics

Historically, diagnosis of STH infection has relied on the microscopic identification of eggs or larvae within the stool shed by infected individuals [40]. The Kato-Katz technique, originally developed to facilitate detection of *Schistosoma mansoni* in stool [41], remains the most commonly employed diagnostic methodology for use in human surveys [42–45]. This technique involves the direct microscopic observation of shed STH eggs within a recently-collected stool sample. Other less frequently utilized copromicroscopic techniques such as the McMaster egg counting technique [46, 47], FLOTAC [48, 49], mini-FLOTAC [50, 51] and FECPAC [52, 53] couple an egg-flotation step with microscopic observation. However, despite widespread use, microscopy-based techniques have long been known and acknowledged to lack both sensitivity and specificity of detection, rendering them suboptimal [43, 54–56]. While such methods remain largely sufficient for disease mapping in areas suspected of having high-prevalence/high-intensity infections

[57, 58], the expansion of intervention strategies with the lofty goal of interrupting transmission and achieving local/regional elimination has resulted in a recognized need for improved diagnostics. Programmatic acknowledgement of the need for a diagnostic technique capable of accurately assessing infection rates under conditions of declining prevalence is expanding [58] as such measures are critical to gauging intervention effectiveness. As a result, recent years have witnessed the championing of polymerase chain reaction (PCR)-based approaches as a possible solution to this diagnostic challenge [59]. In response, a number of multiplexed and multi-parallel quantitative real-time PCR (qPCR) assays have been developed [60–64] and a growing body of evidence is demonstrating that these assays outperform the historically-employed microscopy-based techniques [59, 62–66]. Yet despite these improvements, assay development efforts have exclusively targeted ribosomal and mitochondrial sequences. These targets, historically selected for their ease of discovery and multi-copy nature, remain the most commonly selected targets for PCR-based diagnostics spanning the breadth of eukaryotic pathogens [67–69]. However, despite their frequency of use, such targets are oftentimes suboptimal due to their semi-conserved nature and moderate genomic copy-numbers.

### 1.3 Satellite sequence-based PCR diagnostics

Satellite sequences are repetitive DNA elements, most commonly between 150 to 400 base pairs (bp) in length, which form large tandemly repeated arrays within the genomes of plants and animals [70]. These elements, whose biological functions are just now being elucidated, play important roles in a variety of processes including chromatin organization, kinetochore formation, gene regulation, and speciation [71–74]. Yet despite

these critical functions, satellite DNA sequences are known to diverge rapidly between species, while undergoing periodic homogenization within a species [75]. These processes result in high copy-number sequences that are maintained within a species, but bear little resemblance to satellite regions in even the most closely related of species.

With single satellite families comprising as much as 35% of the eukaryotic genome [76], the diagnostic utility of satellite sequences is unmatched by other genomic elements. Recognizing that an ideal DNA-based diagnostic target should be highly repetitive and species-specific, efforts to utilize satellite sequences as targets for DNA-based molecular diagnostics predate the existence of PCR [77–79]. However, prior to the advent of next-generation sequencing (NGS), attempts to discover optimal satellite sequences for diagnostic exploitation were dependent upon "guess and check"-type strategies. These tactics relyed upon the coupling of restriction enzyme digestion with agarose gel electrophoresis to identify repetitive genomic elements which would appear as high copy-number bands following size discrimination [80] (Figure 1.3). While such labors have led to the identification of diagnostic target sequences for a variety of eukaryotic pathogens [81–83], verification of such targets as optimal genomic elements was not possible utilizing pre-NGS approaches. However, with the advent of NGS, the capacity to efficiently and effectively estimate the genomic abundances of various DNA elements has emerged, fundamentally changing the methods underlying optimal target discovery.

### 1.4 RepeatExplorer

With the advent of NGS technologies, the ability to gather high-throughput sequence data for virtually any organism of interest has become straightforward and

inexpensive. Such sequencing efforts result in vast quantities of raw sequence reads, which are most frequently used to assemble draft/reference genomes. During such assemblies, repetitive sequence elements are typically "masked" because of the challenges they pose for the assembly process [84–86]. However, with the development of specialized informatics tools, these raw sequence reads can now be repurposed to answer specific questions and allow specific investigations.

RepeatExplorer is a collection of Galaxy-based web tools capable of identifying repetitive elements from raw NGS sequence reads [87]. Following initial trimming, grooming, and quality control checks, a randomly sampled subset of reads are uploaded for analysis. Input sequences then undergo an all-to-all sequence comparison, akin to a BLAST analysis, and sequences are binned based upon their similarity. Reads attaining a threshold similarity of $\geq 90\%$ sequence identity over 55% or more of their read lengths are then "clustered", resulting in graph-based output. Within cluster graphs individual reads are represented by vertices/nodes, which are connected by edges when the minimum threshold similarity is met or exceeded. In such clusters, edge length identifies degree of similarity, with shorter edges representing sequences with greater degrees of sequence identity. Based upon the organization and length of the underlying repeat within the genome, clusters then take on characteristic appearances (Figure 1.4). Combining graphical interpretation with statistical output such as cluster density (a measure of the percentage of edges within a cluster relative to the maximum number of possible edges that would exist if each node successfully paired with every other node within the cluster), optimal targets can be selected. Further examination of the selected cluster then allows for the identification of genomic repeat regions that show the greatest element of conservation

and representation. Through the assignment of representation scores to each nucleotide within a given repeat's sequence, target regions which lack diagnostically problematic elements such as insertions/deletions, single nucleotide polymorphisms, and truncations can be targeted for assay design (Figure 1.5).

By designing assays that target the most represented regions within clusters built from the most numerous sequence elements, it becomes possible to develop optimal PCR-based diagnostic tools. While repetitive sequence composition can vary considerably from species to species [88], the methodical analysis of underlying genome content ensures that each assay target is optimal, reducing reliance on targets selected because of factors such as prior availability or ease of discovery.

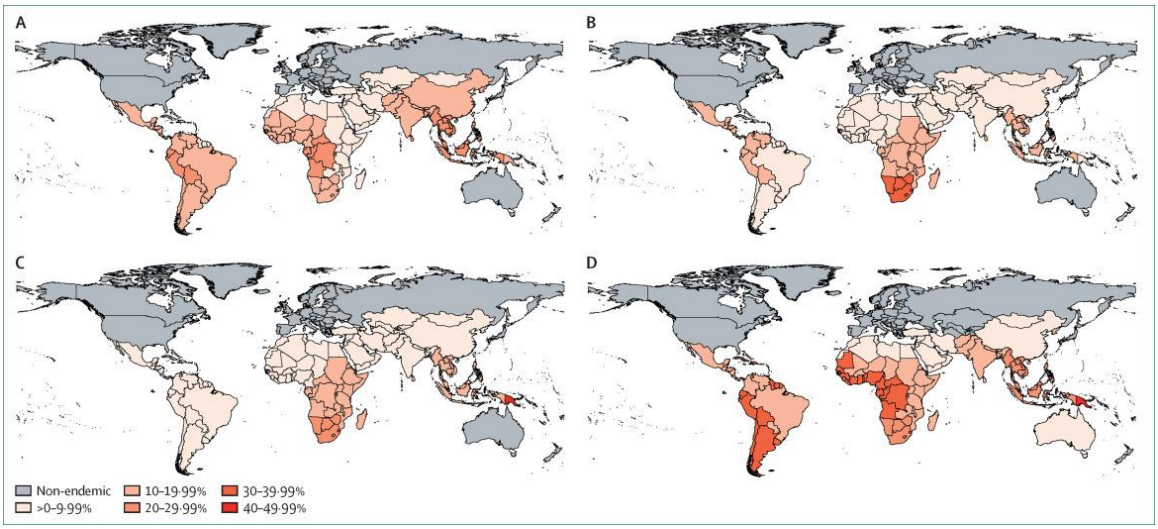


**Figure 1.1 Global distributions of soil-transmitted helminth infections.** Estimated prevalence rates as presented by Jourdan and colleagues, 2017. Prevalences are striated by color and country for (A) *Ascaris lumbricoides*, (B) *Trichuris trichiura*, (C) Human infecting hookworm, and (D) *Strongyloides stercoralis*.

This image adapted from: [2]

Jourdan P, Lamberton P, Fenwick A, Addiss D. Soil-transmitted helminth infections. Lancet. 2018; 391: 2322–2323.

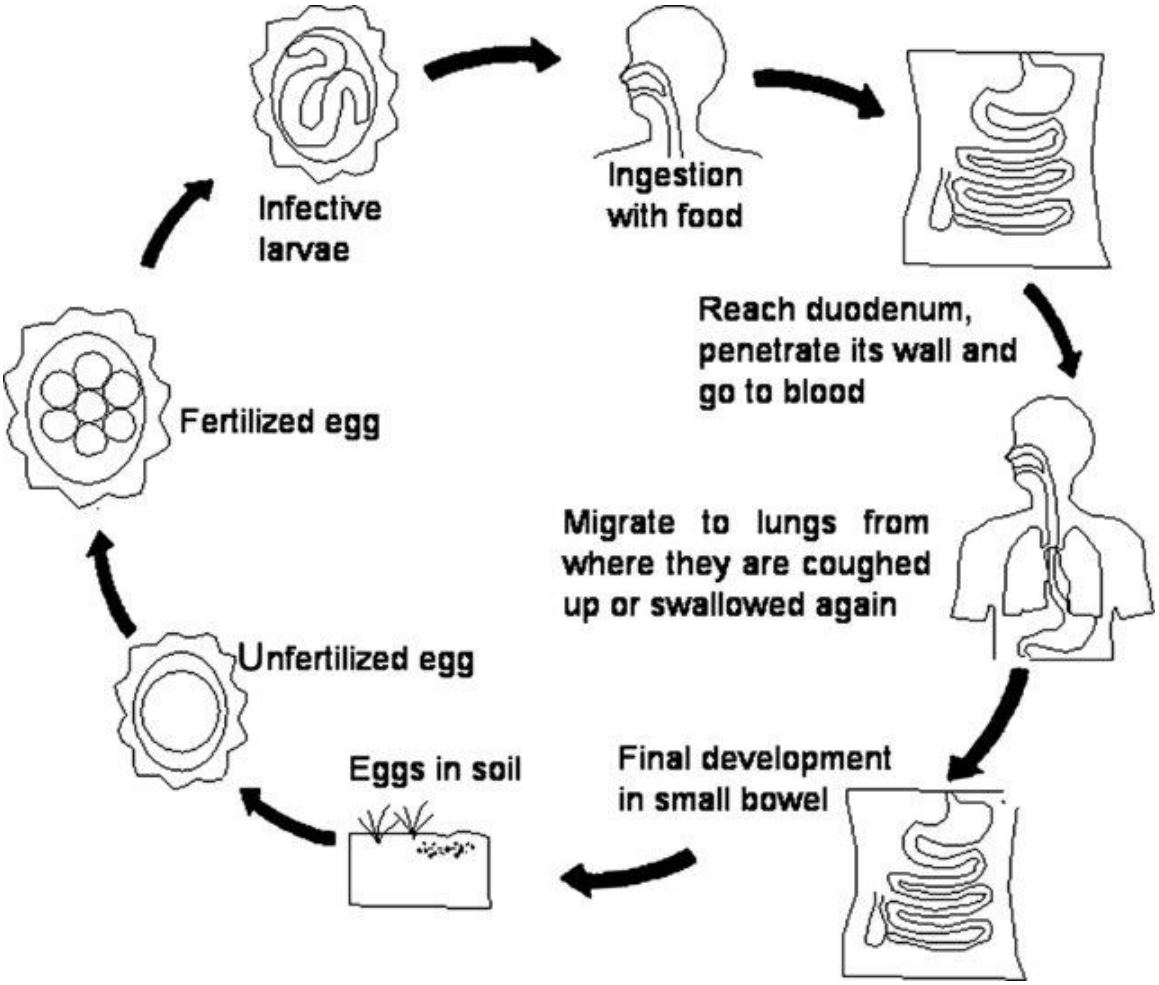


**Figure 1.2 A simplified overview of the soil-transmitted helminth lifecycle.** Infective larvae are ingested with food, water, or through oral contact with other contaminated substance. Hookworm species may also enter the body by directly penetrating the skin (not pictured). Once internalized, larvae travel to the bloodstream and migrate to the lungs. After reaching the lungs, larvae are cleared via the process of coughing, after which they are swallowed and travel back to the intestinal track. After reaching their species-specific intestinal regions of preference, colonization occurs and infection is established. The mating of male and female worms within the intestinal lumen then becomes possible, resulting in the daily shedding of hundreds, to tens of thousands of infective eggs or larvae. Depending upon species, shed pathogen may remain viable in the external environment for months or years. Such extended viability greatly complicated intervention efforts, as it is possible for environmental reservoirs to remain even following the clearance of pathogen from a population.

This image adapted from: [3]

Das CJ, Kumar J, Debnath J, Chaudhry A. Imaging of ascariasis. Australas Radiol. 2007; 51: 500-506.

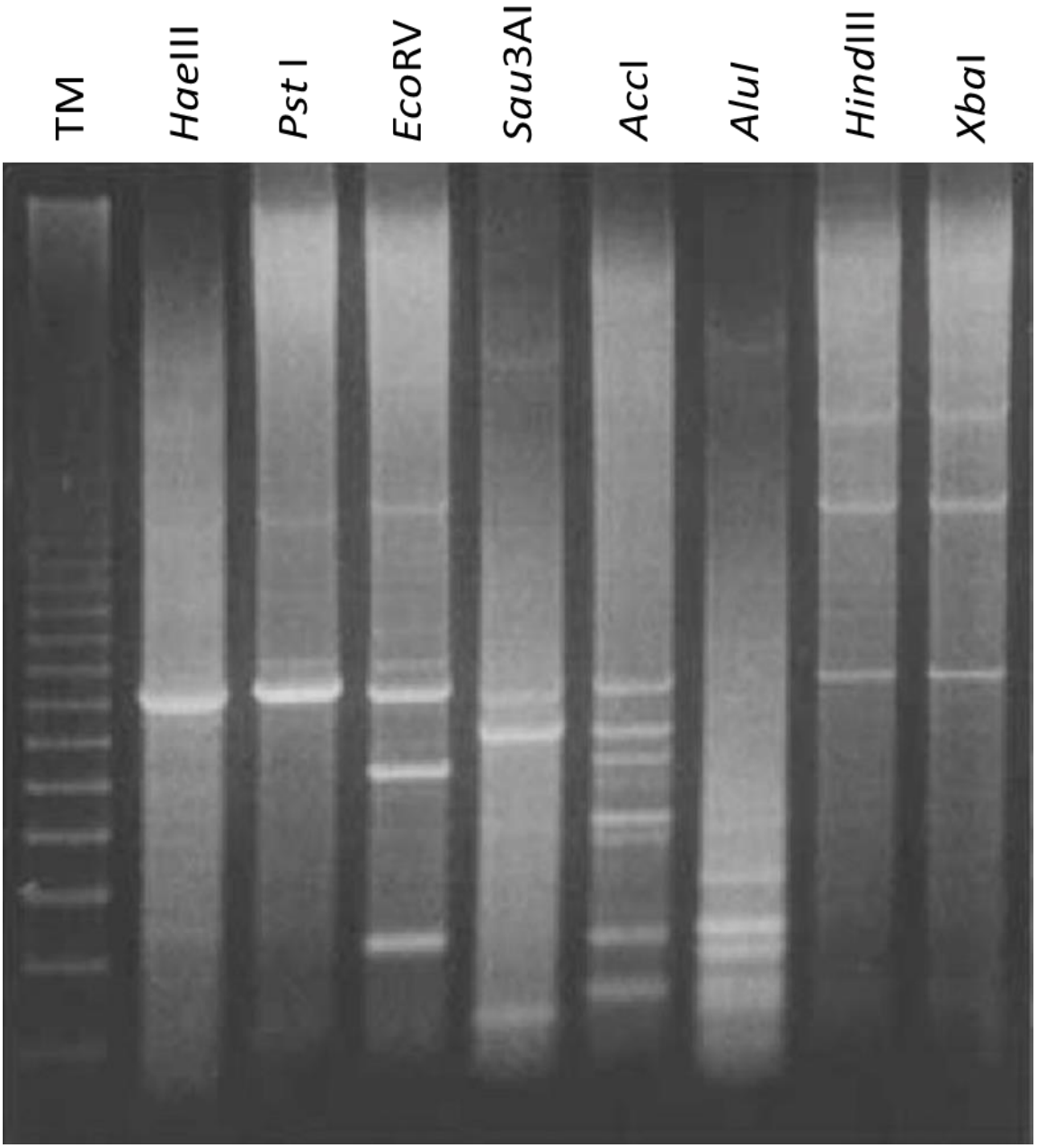


**Figure 1.3 An example of satellite DNA detection using restriction enzyme digestion and agarose gel electrophoresis.** In this example from the literature, genomic DNA from the copepod *Tigriopus brevicornis* was digested with a variety of restriction enzymes. Digested products were then visualized using agarose gel electrophoresis, allowing for the identification of highly represented gDNA elements. These repetitive sequences appear as bands due to the presence of restriction sites in the gDNA regions of each repeat monomer. Due to the presence of these restircition sites, digestion results in the production of many copies of predictably digested monomer, resulting in digestion products of identical, or near identical size. Due to their reptetive nature and resulting high copy numbers, these elements then appear as distinct DNA bands when separated by size on an agarose gel.

Such bands can then be excised, recovered, and sequenced to determine repeat identity. TM; partially digested *Tenebrio molitor* satellite DNA utilized as a standard.

This image adapted from: [80]

Mravinac B, Ugarković E, Franjević D, Plohl M. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. J Mol Evol. 2005; 60: 513-525.

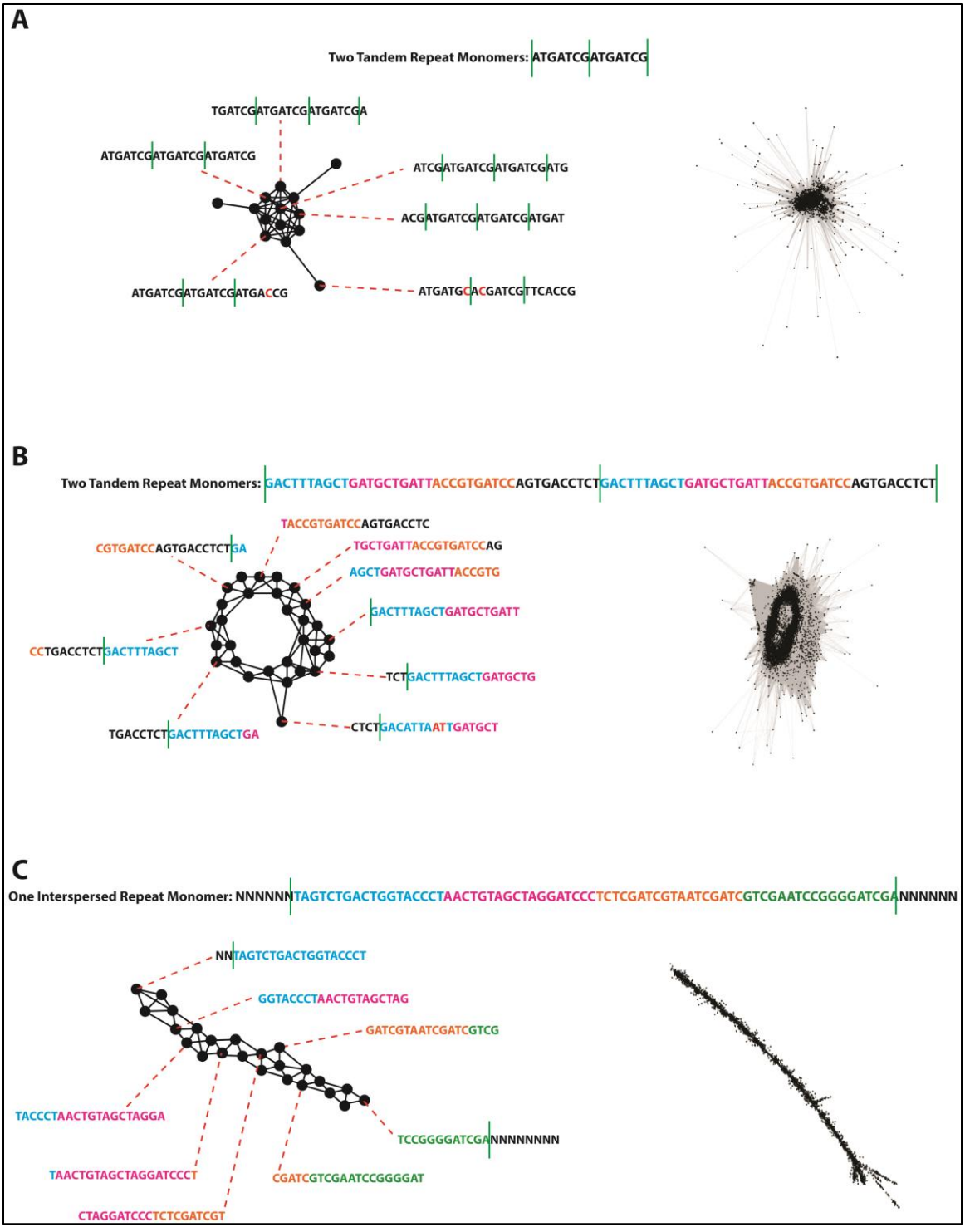


**Figure 1.4 Basic principles underlying RepeatExplorer cluster formation.** By performing an all-to-all BLAST analysis of raw sequence reads, RepeatExplorer joins reads which share 90% or greater sequence identity over 55% or more of the read lengths. Reads (represented by nodes) meeting or exceeding this criteria for identity are joined by edges, forming clusters. (A) A characteristic cluster for a short tandem repeat. Because

read lengths are closer in length to the length of individual repeat monomers, reads exceed RepeatExplorer's similarity threshold when compared with many other reads within the cluster. For this reason, a compact cluster is formed comprised of short verticies connecting each individual node to many other nodes. Reads that are still able to meet the required identity threshold but show greater variability due to the presence of single nucleotide polymorphisms, insertions/deletions, portions of a flanking sequences, or other dissimilarities are represented by the more distant nodes, giving the cluster a "star burst"-like appearance. (B) A characteristic cluster for a long tandem repeat. Because individual repeat monomers are considerably longer than the length of an individual read, reads within a cluster share threshold similarity only with reads mapping to adjacent regions of the repeat. Two reads mapping to separated regions of the monomer will not share threshold similarity, and therefore will not be joined by an edge. However, because the repeat is tandemly arranged, a read spanning the junction between two monomers will share similarity with both the 5' and 3' regions of the adjacent monomers at the junction. For this reason, the resulting cluser forms a closed circle. (C) A characteristic cluster for a long interspersed repeat. Much like a long tandem repeat, individual nodes within a long interspersed repeat meet the similarity threshold only with reads mapping to neighboring regions of the monomer. However, because the repat monomer is not tandemly arranged, it is boardered by variable flanking sequences. As such, reads mapping to the 5' end of the repeat do not share threshold similarity with reads mapping to the repeat's 3' end. This results in the formation of a "line"-like cluster. Because short, tandemly arranged repeats contain the greatest number of repeat monomers per unit length of DNA, an assay targeting such a monomer typically provides the ideal diagnostic target, comprising the most prevalent genomic element of sufficient length to accommodate assay design. As such these clusters represent the "first choice" for target selection.

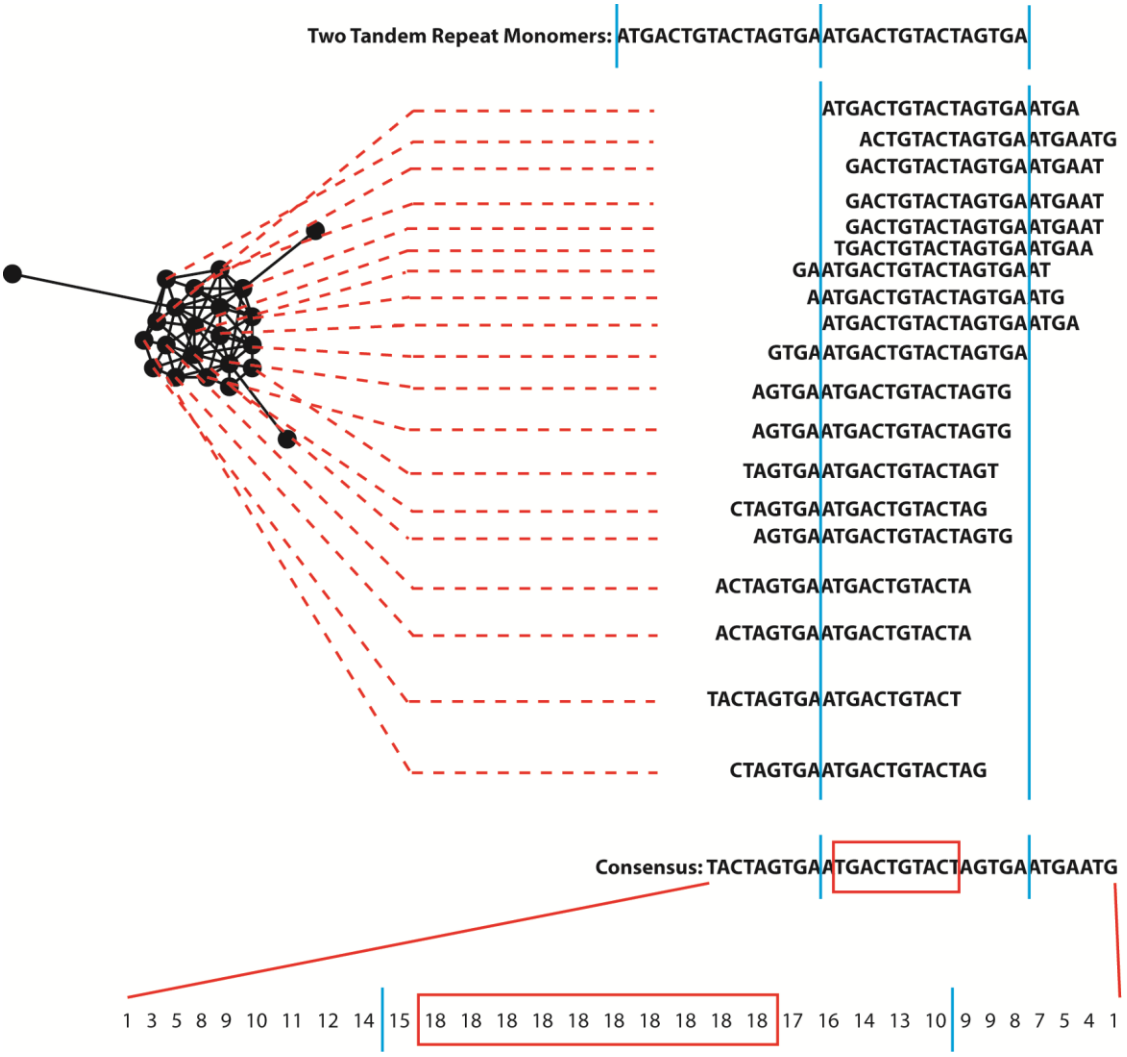


**Figure 1.5 Mechanism for identifying optimal primer/probe regions within a candidate cluster.** Individual reads within a cluster are aligned by RepeatExplorer and this alignment is used to generate a representation score for each nucleotide position within the cluster. This score is equal to the total number of clustered reads that a given nucleotide is found within. In this figure, these scores are represented by the string of numbers, with each number corresponding to an individual nucleotide within the consensus sequence. Utilizing these scores, the region of the cluster containing nucleotides with the greatest representation can be selected as the preferred region for primer/probe design. This region putatively represents the repeat's "core" region, allowing for the avoidance of flanking regions, regions lost within truncated repeats, or highly variable regions within the repeat.

## 1.5 References


1. Ásbjörnsdóttir KH, Means AR, Werkman M, Walson JL. Prospects for elimination of soil-transmitted helminths. Curr Opin Infect Dis. 2017; 30: 482–488.

2. Jourdan P, Lamberton P, Fenwick A, Addiss D. Soil-transmitted helminth infections. Lancet. 2018; 391: 2322–2323.

3. Das CJ, Kumar J, Debnath J, Chaudhry A. Imaging of ascariasis. Australas Radiol. 2007; 51: 500-506.

4. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392: 1859–1922.

5. King CH, Bertino A-M. Asymmetries of Poverty: Why Global Burden of Disease Valuations Underestimate the Burden of Neglected Tropical Diseases. PLoS Negl Trop Dis. 2008; 2: e209.

6. Pabalan N, Singian E, Tabangay L, Jarjanazi H, Boivin MJ, Ezeamama AE. Soil-transmitted helminth infection, loss of education and cognitive impairment in school-aged children: A systematic review and meta-analysis. PLoS Negl Trop Dis. 2018;12: e0005523.

7. Gall S, Müller I, Walter C, Seelig H, Steenkamp L, Pühse U, et al. Associations between selective attention and soil-transmitted helminth infections, socioeconomic status, and physical fitness in disadvantaged children in Port Elizabeth, South Africa: An observational study. PLoS Negl Trop Dis. 2017; 11: e0005573.

8. Müller I, Yap P, Steinmann P, Damons BP, Schindler C, Seelig H, et al. Intestinal parasites, growth and physical fitness of schoolchildren in poor neighbourhoods of Port Elizabeth, South Africa: a cross-sectional survey. Parasit Vectors. 2016; 9: 488.

9. Beknazarova M, Whiley H, Ross K. Strongyloidiasis: A Disease of Socioeconomic Disadvantage. Int J Environ Res Public Health. 2016; 13: 517.

10. Lustigman S, Prichard RK, Gazzinelli A, Grant WN, Boatin BA, McCarthy JS, et al. A Research Agenda for Helminth Diseases of Humans: The Problem of Helminthiases. PLoS Negl Trop Dis. 2012; 6: e1582.

11. Parija SC, Chidambaram M, Mandal J. Epidemiology and clinical features of soil-transmitted helminths. Trop Parasitol. 2017; 7: 81–85.

12. Brooker S, Hotez PJ, Bundy DAP. Hookworm-Related Anaemia among Pregnant Women: A Systematic Review. PLoS Negl Trop Dis. 2008; 2: e291.

13. Gyorkos TW, Maheu-Giroux M, Blouin B, Casapia M. Impact of Health Education on Soil-Transmitted Helminth Infections in Schoolchildren of the Peruvian Amazon: A Cluster-Randomized Controlled Trial. PLoS Negl Trop Dis. 2013; 7: e2397.

14. Craig JM, Scott AL. Antecedent *Nippostrongylus* infection alters the lung immune response to *Plasmodium berghei*. Parasite Immunol. 2017; 39: e12441.

15. Mbabazi PS, Andan O, Fitzgerald DW, Chitsulo L, Engels D, Downs JA. Examining the Relationship between Urogenital Schistosomiasis and HIV Infection. PLoS Negl Trop Dis. 2011; 5: e1396.

16. Mabbott NA. The Influence of Parasite Infections on Host Immunity to Co-infection With Other Pathogens. Front Immunol. 2018; 9: 2579.

17. Lo NC, Addiss DG, Hotez PJ, King CH, Stothard JR, Evans DS, et al. A call to strengthen the global strategy against schistosomiasis and soil-transmitted helminthiasis: the time is now. Lancet Infect Dis. 2017; 17: e64–e69.

18. Gyorkos TW, Gilbert NL, Larocque R, Casapía M. *Trichuris* and hookworm infections associated with anaemia during pregnancy: *Trichuris*, hookworm and anaemia in pregnancy. Trop Med Int Health. 2011; 16: 531–537.

19. Albonico M, Montresor A, Crompton DWT, Savioli L. Intervention for the Control of Soil-Transmitted Helminthiasis in the Community. Advances in Parasitology. Elsevier; 2006. pp. 311–348.

20. Mascarini-Serra L. Prevention of soil-transmitted helminth infection. J Glob Infect Dis. 2011; 3: 175.

21. Freeman MC, Akogun O, Belizario V, Brooker SJ, Gyorkos TW, Imtiaz R, et al. Challenges and opportunities for control and elimination of soil-transmitted helminth infection beyond 2020. PLoS Negl Trop Dis. 2019; 13: e0007201.

22. Okoyo C, Nikolay B, Kihara J, Simiyu E, Garn JV, Freeman MC, et al. Monitoring the impact of a national school based deworming programme on soil-transmitted helminths in Kenya: the first three years, 2012 – 2014. Parasit Vectors. 2016; 9: 408.

23. Halliday KE, Oswald WE, Mcharo C, Beaumont E, Gichuki PM, Kepha S, et al. Community-level epidemiology of soil-transmitted helminths in the context of school-based deworming: Baseline results of a cluster randomised trial on the coast of Kenya. PLoS Negl Trop Dis. 2019; 13: e0007427.

24. Sungkar S, Ridwan AS, Kusumowidagdo G. The Effect of Deworming Using Triple-Dose Albendazole on Nutritional Status of Children in Perobatang Village, Southwest Sumba, Indonesia. J Parasitol Res. 2017; 2017: 1–5.

25. Dunn JC, Bettis AA, Wyine NY, Lwin AMM, Tun A, Maung NS, et al. Soil-transmitted helminth reinfection four and six months after mass drug administration: results from the delta region of Myanmar. PLoS Negl Trop Dis. 2019; 13: e0006591.

26. Jia T-W, Melville S, Utzinger J, King CH, Zhou X-N. Soil-Transmitted Helminth Reinfection after Drug Treatment: A Systematic Review and Meta-Analysis. PLoS Negl Trop Dis. 2012; 6: e1621.

27. Zerdo Z, Yohanes T, Tariku B. Soil-Transmitted Helminth Reinfection and Associated Risk Factors among School-Age Children in Chencha District, Southern Ethiopia: A Cross-Sectional Study. J Parasitol Res. 2016; 2016: 1–7.

28. Halpenny CM, Paller C, Koski KG, Valdés VE, Scott ME. Regional, Household and Individual Factors that Influence Soil Transmitted Helminth Reinfection Dynamics in Preschool Children from Rural Indigenous Panamá. PLoS Negl Trop Dis. 2013; 7: e2070.

29. Brown HW. Further studies on the longevity of the eggs of *Ascaris lumbricoides* and *A. suum*. J Parasitol. 1928; 15: 14–22.

30. Dąbrowska J, Zdybel J, Karamon J, Kochanowski M, Stojecki K, Cencek T, et al. Assessment of viability of the nematode eggs (*Ascaris*, *Toxocara*, *Trichuris*) in sewage sludge with the use of LIVE/DEAD Bacterial Viability Kit. Ann Agric Environ Med AAEM. 2014; 21: 35–41.

31. Pham-Duc P, Nguyen-Viet H, Hattendorf J, Zinsstag J, Phung-Dac C, Zurbrügg C, et al. *Ascaris lumbricoides* and *Trichuris trichiura* infections associated with wastewater and human excreta use in agriculture in Vietnam. Parasitol Int. 2013; 62: 172–180.

32. Amoah ID, Adegoke AA, Stenström TA. Soil-transmitted helminth infections associated with wastewater and sludge reuse: a review of current evidence. Trop Med Int Health. 2018; 23: 692–703.

33. Zdybel J, Karamon J, Dąbrowska J, Różycki M, Bilska-Zając E, Kłapeć T, et al. Parasitological contamination with eggs *Ascaris* spp., *Trichuris* spp. and *Toxocara* spp. of dehydrated municipal sewage sludge in Poland. Environ Pollut. 2019; 248: 621–626.

34. Vaz Nery S, Traub RJ, McCarthy JS, Clarke NE, Amaral S, Llewellyn S, et al. WASH for WORMS: A Cluster-Randomized Controlled Trial of the Impact of a Community Integrated Water, Sanitation, and Hygiene and Deworming Intervention on Soil-Transmitted Helminth Infections. Am J Trop Med Hyg. 2019; 100: 750–761.

35. Chard AN, Garn JV, Chang HH, Clasen T, Freeman MC. Impact of a school-based water, sanitation, and hygiene intervention on school absence, diarrhea, respiratory infection, and soil-transmitted helminths: results from the WASH HELPS cluster-randomized trial. J Glob Health. 2019; 9: 020402.

36. Vaz Nery S, Pickering AJ, Abate E, Asmare A, Barrett L, Benjamin-Chung J, et al. The role of water, sanitation and hygiene interventions in reducing soil-transmitted helminths: interpreting the evidence and identifying next steps. Parasit Vectors. 2019; 12: 273.

37. Ercumen A, Benjamin-Chung J, Arnold BF, Lin A, Hubbard AE, Stewart C, et al. Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: A cluster-randomized controlled trial in rural Bangladesh. PLoS Negl Trop Dis. 2019; 13: e0007323.

38. Coffeng LE, Vaz Nery S, Gray DJ, Bakker R, de Vlas SJ, Clements ACA. Predicted short and long-term impact of deworming and water, hygiene, and sanitation on transmission of soil-transmitted helminths. PLoS Negl Trop Dis. 2018; 12: e0006758.

39. Pickering AJ, Njenga SM, Steinbaum L, Swarthout J, Lin A, Arnold BF, et al. Effects of single and integrated water, sanitation, handwashing, and nutrition interventions on child soil-transmitted helminth and Giardia infections: A cluster-randomized controlled trial in rural Kenya. PLOS Med. 2019; 16: e1002841.

40. O'Connell EM, Nutman TB. Molecular Diagnostics for Soil-Transmitted Helminths. Am J Trop Med Hyg. 2016; 95: 508–513.

41. Sleigh A, Hoff R, Mott K, Barreto M, de Paiva TM, de Souza Pedrosa J, et al. Comparison of filtration staining (Bell) and thick smear (Kato) for the detection and quantitation of *Schistosoma mansoni* eggs in faeces. Trans R Soc Trop Med Hyg. 1982; 76: 403–406.

42. Speich B, Ali SM, Ame SM, Albonico M, Utzinger J, Keiser J. Quality control in the diagnosis of *Trichuris trichiura* and *Ascaris lumbricoides* using the Kato-Katz technique: experience from three randomised controlled trials. Parasit Vectors. 2015; 8: 82.

43. Turner HC, Bettis AA, Dunn JC, Whitton JM, Hollingsworth TD, Fleming FM, et al. Economic Considerations for Moving beyond the Kato-Katz Technique for Diagnosing Intestinal Parasites As We Move Towards Elimination. Trends Parasitol. 2017; 33: 435–443.

44. Melrose W, Menzies H, Boer M, Joseph H, Reeve D, Speare R. Short Communication: A Simple Method for Performing Worm-Egg Counts on Sodium Acetate Formaldehyde-Preserved Samples. J Parasitol Res. 2012; 2012: 1–3.

45. Bärenbold O, Raso G, Coulibaly JT, N'Goran EK, Utzinger J, Vounatsou P. Estimating sensitivity of the Kato-Katz technique for the diagnosis of *Schistosoma mansoni* and hookworm in relation to infection intensity. PLoS Negl Trop Dis. 2017; 11: e0005953.

46. Dunn A, Keymer A. Factors affecting the reliability of the McMaster technique. J Helminthol. 1986;60: 260–262.

47. Egwang TG, Slocombe JO. Efficiency and sensitivity of techniques for recovering nematode eggs from bovine feces. Can J Comp Med. 1981; 45: 243–248.

48. Kochanowski M, Karamon J, Dąbrowska J, Cencek T. Experimental Estimation of the Efficacy of the Flotac Basic Technique. J Parasitol. 2014; 100: 633–639.

49. Habtamu K, Degarege A, Ye-Ebiyo Y, Erko B. Comparison of the Kato-Katz and FLOTAC techniques for the diagnosis of soil-transmitted helminth infections. Parasitol Int. 2011; 60: 398–402.

50. Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruysse J, Utzinger J, et al. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. Nat Protoc. 2017; 12: 1723–1732.

51. Maurelli MP, Rinaldi L, Alfano S, Pepe P, Coles GC, Cringoli G, et al. Mini-FLOTAC, a new tool for copromicroscopic diagnosis of common intestinal nematodes in dogs. Parasit Vectors. 2014; 7: 356.

52. Godber OF, Phythian CJ, Bosco A, Ianniello D, Coles G, Rinaldi L, et al. A comparison of the FECPAK and Mini-FLOTAC faecal egg counting techniques. Vet Parasitol. 2015; 207: 342–345.

53. Bosco A, Rinaldi L, Maurelli M, Musella V, Coles G, Cringoli G. The comparison of FLOTAC, FECPAK and McMaster techniques for nematode egg counts in cattle. Acta Parasitol. 2014; 59.

54. Mationg MLS, Gordon CA, Tallo VL, Olveda RM, Alday PP, Reñosa MDC, et al. Status of soil-transmitted helminth infections in schoolchildren in Laguna Province, the Philippines: Determined by parasitological and molecular diagnostic techniques. PLoS Negl Trop Dis. 2017; 11: e0006022.

55. Knopp S, Rinaldi L, Khamis IS, Stothard JR, Rollinson D, Maurelli MP, et al. A single FLOTAC is more sensitive than triplicate Kato–Katz for the diagnosis of low-intensity soil-transmitted helminth infections. Trans R Soc Trop Med Hyg. 2009; 103: 347–354.

56. Booth M, Vounatsou P, N'Goran EK, Tanner M, Utzinger J. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. Parasitology. 2003; 127: 525–531.

57. Lin D-D, Liu J-X, Liu Y-M, Hu F, Zhang Y-Y, Xu J-M, et al. Routine Kato–Katz technique underestimates the prevalence of *Schistosoma japonicum*: A case study in an endemic area of the People's Republic of China. Parasitol Int. 2008; 57: 281–286.

58. Lim MD, Brooker SJ, Belizario VY Jr, Gay-Andrieu F, Gilleard J, Levecke B, et al. Diagnostic tools for soil-transmitted helminths control and elimination programs: A pathway for diagnostic product development. PLoS Negl Trop Dis. 2018; 12: e0006213.

59. Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

60. Verweij JJ, Brienen EAT, Ziem J, Yelifari L, Polderman AM, Van Lieshout L. Simultaneous Detection and Quantification of *Ancylostoma duodenale, Necator americanus,* and *Oesophagostomum bifurcum* in Fecal Samples Using Multiplex Real-Time PCR. Am J Trop Med Hyg. 2007; 77: 685–690.

61. Phuphisut O, Yoonuan T, Sanguankiat S, Chaisiri K, Maipanich W, Pubampen S, et al. Triplex polymerase chain reaction assay for detection of major soil-transmitted helminths, *Ascaris lumbricoides, Trichuris trichiura, Necator americanus,* in fecal samples. Southeast Asian J Trop Med Public Health. 2014; 45: 267–275.

62. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a Multiplex Quantitative PCR to Assess Prevalence and Intensity Of Intestinal Parasite Infections in a Controlled Clinical Trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

63. Mejia R, Vicuña Y, Broncano N, Sandoval C, Vaca M, Chico M, et al. A Novel, Multi-Parallel, Real-Time Polymerase Chain Reaction Approach for Eight Gastrointestinal Parasites Provides Improved Diagnostic Capabilities to Resource-Limited At-Risk Populations. Am J Trop Med Hyg. 2013; 88: 1041–1047.

64. Basuni M, Muhi J, Othman N, Verweij JJ, Ahmad M, Miswan N, et al. A pentaplex real-time polymerase chain reaction assay for detection of four species of soil-transmitted helminths. Am J Trop Med Hyg. 2011; 84: 338–343.

65. Cools P, Vlaminck J, Albonico M, Ame S, Ayana M, Antonio BPJ, et al. Diagnostic performance of a single and duplicate Kato-Katz, Mini-FLOTAC, FECPAKG2 and qPCR for the detection and quantification of soil-transmitted helminths in three endemic countries. PLoS Negl Trop Dis. 2019; 13: e0007446.

66. Knopp S, Salim N, Schindler T, Voules DAK, Rothen J, Lweno O, et al. Diagnostic Accuracy of Kato–Katz, FLOTAC, Baermann, and PCR Methods for the Detection of Light-Intensity Hookworm and *Strongyloides stercoralis* Infections in Tanzania. Am J Trop Med Hyg. 2014; 90: 535–545.

67. Amaral LC, Robortella DR, Guimarães LFF, Limongi JE, Fontes CJF, Pereira DB, et al. Ribosomal and non-ribosomal PCR targets for the detection of low-density and mixed malaria infections. Malar J. 2019; 18: 154.

68. Sanprasert V, Kerdkaew R, Srirungruang S, Charuchaibovorn S, Phadungsaksawasdi K, Nuchprayoon S. Development of Conventional Multiplex PCR: A Rapid Technique for Simultaneous Detection of Soil-Transmitted Helminths. Pathogens. 2019; 8: 152.

69. Iqbal A, Mahdy MAK, Dixon BR, Surin J, Lim YAL. Cryptosporidisis in HIV-Infected Individuals: A Global Perspective. J AIDS Clin Res. 2012; 01.

70. Shen C-H. Diagnostic molecular biology. 1st edition. San Diego, CA: Elsevier; 2018.

71. Pezer Ž., Brajković J, Feliciello I, Ugarkovc Đ. Satellite DNA-Mediated Effects on Genome Regulation. Genome Dyn. 2012; 7: 153-169.

72. Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M. Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. 2015; 23: 463–477.

73. Rodríguez FR, de la Herrán R, Navajas-Pérez R, Cano-Roldán B, Sola-Campoy PJ, García-Zea JA, et al. Centromeric Satellite DNA in Flatfish (Order Pleuronectiformes) and Its Relation to Speciation Processes. J Hered. 2016; esw076.

74. Grenfell AW, Strzelecka M, Heald R. Transcription brings the complex(ity) to the centromere. Cell Cycle. 2017; 16: 235–236.

75. Roizes G. Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Res. 2006; 34: 1912–1924.

76. Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblížková A, et al. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. Ann Bot. 2011; 107: 255–268.

77. Poole CB, Williams SA. A rapid DNA assay for the species-specific detection and quantification of *Brugia* in blood samples. Mol Biochem Parasitol. 1990; 40: 129–136.

78. McReynolds LA, Poole CB, Williams SA, Partono F. Diagnosis of Filariasis with DNA Probes. In: Rapid Methods and Automation in Microbiology and Immunology. Berlin, Heidelberg: Springer Berlin Heidelberg; 1991. pp. 402–409.

79. Weiss JB. DNA probes and PCR for diagnosis of parasitic infections. Clin Microbiol Rev. 1995; 8: 113–130.

80. Mravinac B, Ugarković E, Franjević D, Plohl M. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. J Mol Evol. 2005; 60: 513-525.

81. McReynolds LA, DeSimone SM, Williams SA. Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi*. Proc Natl Acad Sci USA. 1986; 83: 797–801.

82. Zhong M, McCarthy J, Bierwert L, Lizotte-Waniewski M, Chanteau S, Nutman TB, et al. A polymerase chain reaction assay for detection of the parasite *Wuchereria bancrofti* in human blood samples. Am J Trop Med Hyg. 1996; 54: 357–363.

83. Vargas N, Souto RP, Carranza JC, Vallejo GA, Zingales B. Amplification of a specific repetitive DNA sequence for *Trypanosoma rangeli* identification and its potential application in epidemiological investigations. Exp Parasitol. 2000; 96: 147–159.

84. Wang J. RePS: A Sequence Assembler That Masks Exact Repeats Identified from the Shotgun Data. Genome Res. 2002; 12: 824–831.

85. Bedell JA, Korf I, Gish W. MaskerAid: a performance enhancement to RepeatMasker. Bioinforma Oxf Engl. 2000; 16: 1040–1041.

86. Kong J, Huh S, Won J-I, Yoon J, Kim B, Kim K. GAAP: A Genome Assembly + Annotation Pipeline. BioMed Res Int. 2019; 2019: 1–12.

87. Novak P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792–793.

88. Subirana JA, Messeguer X. A Satellite Explosion in the Genome of Holocentric Nematodes. PLoS ONE. 2013; 8: e62221.

# CHAPTER 2

# IMPROVED PCR-BASED DETECTION OF SOIL-TRANSMITTED HELMINTH INFECTIONS USING A NEXT-GENERATION SEQUENCEING APPROACH TO ASSAY DESIGN

*as written by Pilotte N., Papaiakovou M., Grant J.R., Bierwert L.A., Llewellyn S., McCarthy J.S., Williams S.A.*
*PLoS Neglected Tropical Diseases, 2016*

## 2.1 General Chapter Overview

This chapter describes a new pipeline for the discovery of highly repetitive DNA sequences within eukaryotic organisms, facilitating the development of real-time PCR (qPCR) assays targeting optimal sequences. Following the generation of raw NGS reads, a Galaxy-based analysis tool known as RepeatExplorer is utilized to identify the most common genomic elements within the human infecting soil-transmitted helminth species *Necator americanus*, *Ancylostoma duodenale*, *Trichuris trichiura*, *Strongyloides stercoralis,* and *Ascaris lumbricoides*. Following the identification of optimal targets, novel small-volume, multi-parallel assays are developed, optimized, and validated through comparative testing with published assays. Discordant results, coupled with confirmatory testing and sequencing efforts demonstrate that the novel index assays provide improved sensitivity and specificity over their reference comparators, in certain instances correcting pathogen misclassifications which occur when using the less specific reference assays targeting ribosomal sequences. Of note, the analysis of *A. lumbricoides* repeats, performed using DNA extracted from adult *A. lumbricoides* tissues, identifies the most repetitive

sequence elements as mapping to the ribosome-encoding and mitochondrial genes, rather than to a "non-coding" repetitive sequence. This finding is likely the result of *A. lumbricoides* undergoing chromosome diminution during embryonic developement. During this process, many of the genomic repeats are predictably and repeatibly eliminated from somatic tissues, rendering the moderately-repeated ribosomal targets the most prevalent remaining.

My contributions to this work include isolating all gDNA specimens, performing all NGS library preparations and sequencing reactions, and analyzing all output. I am also responsible for the re-purposing of RepeatExplorer software, a software designed and intended for use with plants, by developing a pipeline for its use as a repetitive element discovery tool for soil-transmitted helminths and other eukaryotic pathogens. I performed all analyses, designed all primers and probes, directly performed, or oversaw the performance of all optimization and validation testing, and wrote the manuscript.

## 2.2 Abstract


### 2.2.1 Background

The soil transmitted helminths are a group of parasitic worms responsible for extensive morbidity in many of the world's most economically depressed locations. With growing emphasis on disease mapping and eradication, the availability of accurate and cost-effective diagnostic measures is of paramount importance to global control and elimination efforts. While real-time PCR-based molecular detection assays have shown great promise, to date, these assays have utilized sub-optimal targets. By performing next-generation sequencing-based repeat analyses, we have identified high copy-number, non-coding DNA sequences from a series of soil transmitted pathogens. We have used these repetitive DNA elements as targets in the development of novel, multi-parallel, PCR-based diagnostic assays.

### 2.2.2 Methodology/principal findings

Utilizing next-generation sequencing and the Galaxy-based RepeatExplorer web server, we performed repeat DNA analysis on five species of soil transmitted helminths (*Necator americanus*, *Ancylostoma duodenale*, *Trichuris trichiura*, *Ascaris lumbricoides*, and *Strongyloides stercoralis*). Employing high copy-number, non-coding repeat DNA sequences as targets, novel real-time PCR assays were designed, and assays were tested against established molecular detection methods. Each assay provided consistent detection of genomic DNA at quantities of 2 fg or less, demonstrated species-specificity, and showed an improved limit of detection over the existing, proven PCR-based assay.

### 2.2.3 Conclusions/significance

The utilization of next-generation sequencing-based repeat DNA analysis methodologies for the identification of molecular diagnostic targets has the ability to improve assay species-specificity and limits of detection. By exploiting such high copy-number repeat sequences, the assays described here will facilitate soil transmitted helminth diagnostic efforts. We recommend similar analyses when designing PCR-based diagnostic tests for the detection of other eukaryotic pathogens.

## 2.3 Author Summary

With a growing emphasis on the mapping and elimination of soil transmitted helminth (STH) infections, the need for optimal and specific diagnostic methods is increasing. While PCR-based diagnostic methods for the detection of these parasitic organisms exist, these assays make use of sub-optimal target sequences. By designing assays that target non-coding, high copy-number repetitive sequences, both the limit of detection and species-specificity of detection improve. Using next-generation sequencing technology, we have identified high copy-number repeats for a series of STH species responsible for the greatest burden of disease. Using these repetitive sequences as targets in the design of novel real-time PCR assays, we have improved both the limits of detection and species-specificity of detection, and we have demonstrated this improved detection by testing these assays against an established PCR-based diagnostic methodology. Accordingly, these assays should facilitate mapping and monitoring efforts, and the generalized application of this approach to assay design should improve detection efforts for other eukaryotic pathogens.

## 2.4 Introduction

Estimated to infect more than one quarter of the world's total population, the soil transmitted helminths (STH) are responsible for profound morbidity and nutritional insufficiency [1]. Concentrated in the world's most impoverished locations, the results of widespread infection on economic capacity are equally burdensome. Yet despite the scope of such disease, and continuing efforts to improve treatment programs and integration strategies, reliable and accurate diagnosis of STH infections remains difficult, and resulting prevalence estimates remain imprecise [1-2].

In recent years, the interest in molecular diagnostic methods for the detection of gastrointestinal helminths has grown substantially. Largely, this escalation in interest has occurred in parallel with the belief that standard microscopy-based methodologies for the examination of stool samples are sub-optimal, leading to underrepresentation of infection [3-5]. Further complicating matters, rates of STH egg/larval excretion have been shown to vary considerably within sequentially collected stool samples originating from a single infected individual [6-7]. This variability in egg/larval count can result in false negative samples, particularly when non-amplification-based diagnostic methodologies are utilized [7]. Such underrepresentation of disease complicates programmatic efforts, making the accurate assessment of the effects of intervention difficult, and frequently leaving low-level infections undiagnosed [5, 8-9]. Additionally, microscopy-based diagnostic methods have been linked with pathogen misidentification due to the morphological similarities that exist between species [5, 10]. Because of such concerns, a number of conventional and real-time PCR-based assays have been developed with the objective of improving both species-specificity and limits of detection [4, 11-17]. These assays have proven valuable,

and as global efforts to estimate the burden of disease caused by the soil transmitted helminths (STHs) continue to increase, the number of studies incorporating such assays has risen in response [3, 5, 9, 18-21]. To date, the target sequences for such assays have been ribosomal internal transcribed spacer (ITS) sequences, 18S or ribosomal subunit sequences, or mitochondrial genes such as cytochrome oxidase I (COI) [4, 11-14]. Ribosomal sequences have been selected as diagnostic targets because they are typically found as easily identified moderate copy number tandem repeats in nucleated organisms [22-25]. Similarly, multiple copies of mitochondrial targets are found in the vast majority of eukaryotic cells [26], making them attractive target choices. However, while effective, such diagnostic targets are often sub-optimal. This is particularly true in the case of nematodes and other multi-cellular organisms where species-specific, highly repetitive DNA elements frequently make up a substantial portion of the genome, and are often present at copy-numbers exceeding 1,000 per haploid genome [27-29]. Due to such overrepresentation, non-coding repetitive sequence elements have become the targets of choice for many PCR-based diagnostic assays for the detection of various helminth species [30-31]. However, the identification of such repeats has historically been complicated and labor intensive. This identification has relied on techniques such as restriction endonuclease digestion of genomic DNA, followed by gel electrophoresis and Sanger DNA sequencing or polyacrylamide slab gel sequencing [32-34]. However, the advent of next-generation sequencing (NGS) technologies and associated informatics tools has expedited the search for highly repetitive sequence elements [35-39], and greater confidence can be placed in the accuracy of the results of such searches. Furthermore, as ribosomal and mitochondrial sequences tend to demonstrate high degrees of conservation

between species, species-specificity of detection is also improved through the targeting of unique, highly-divergent, non-coding repeat DNA elements.

Here we describe the development of multi-parallel real-time PCR assays for the detection of five species of soil transmitted helminths (*Necator americanus*, *Ancylostoma duodenale*, *Trichuris trichiura*, *Strongyloides stercoralis*, and *Ascaris lumbricoides*). Using NGS-generated sequence data and the Galaxy-based RepeatExplorer computational pipeline [38-39], we have searched the genomes of each organism for highly repetitive, non-coding DNA elements in order to identify diagnostic targets capable of providing optimal limits of detection and species-specificity of detection. Using these targets to design small-volume, multi-parallel tests [4], we have created a platform that provides cost-minimizing implementation of only those assays appropriate for a specific geographic region based upon the infections present. While performing multiplex assays may provide labor and time savings in locations where many parasites are co-endemic, such assays result in considerable waste when used in areas harboring only one or a few of the target species. In such settings, the "pick-and-choose" nature of multi-parallel assays minimizes reagent waste, and by improving upon limits of detection, the species-specific platform we describe here should facilitate improved STH monitoring and mapping efforts. Since NGS-based repeat analyses allow for the selection of the most efficacious target sequences, this approach to assay design should be applied to the development of additional diagnostics tests for other eukaryotic pathogens.

## 2.5 Materials and Methods

### 2.5.1 Isolation of parasite genomic DNA

For isolation of genomic DNA from *N. americanus*, *A. duodenale*, and *T. trichiura*, extractions were performed on cryopreserved adult worms in accordance with the "SWDNA1" protocol available on the Filarial Research Reagent Resource Center website (http://www.filariasiscenter.org/parasite-resources/Protocols/materials-1/). For *N. americanus* and *A. duodenale*, DNA extractions were conducted using a pool of approximately 10 adult worms. Both hookworm species belonged to strains originating in China. In the case of *T. trichiura*, extraction was performed using a single adult female worm of Ugandan origin. For *S. stercoralis* and *A. lumbricoides*, previously extracted genomic DNA was received from collaborators. *S. stercoralis* DNA was obtained from laboratory-reared worms originating from Pennsylvania, USA, and *A. lumbricoides* DNA was isolated from worms obtained from Ecuador.

### 2.5.2 Next-generation sequencing of genomic DNA

#### 2.5.2.1 Library preparation

50 ng of genomic DNA, at a concentration of 2.5 ng/µl, was utilized for the NGS library preparation of all organisms except *S. stercoralis*. For *S. stercoralis*, sequencing was not performed, as publically available sequence reads were used for the bioinformatics analyses (Sequence Read Archive ID: ERX044031). For all remaining parasites, libraries were prepared using the Nextera DNA Sample Preparation Kit (Illumina, San Diego, CA),

the Nextera DNA Sample Preparation Index Kit (Illumina), and the ZR-96 DNA Clean & Concentrator-5 Kit (Zymo Research Corporation, Irvine, CA) in accordance with the manufacturer's protocols and previous description [40]. Following library preparation, the concentration of each library was determined using the Qubit 1.0 Fluorometer (Life Technologies, Carlsbad, CA) and the Qubit dsDNA Broad Range Assay Kit (Life Technologies). Additionally, the size distribution of each library was analyzed using the Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA) and the Agilent High Sensitivity DNA Kit (Agilent Technologies).

#### 2.5.2.2 Next-generation sequencing

Based upon the DNA concentrations and size distributions for each prepared library, aliquots containing approximately 12 pmol of library were created for all parasites. Library aliquots were then sequenced individually on the MiSeq platform (Illumina) using the MiSeq Reagent Kit v3 (150 cycles) (Illumina) and the single-ended read approach.

### 2.5.3 Repeat analysis

For each parasite analyzed, raw sequencing reads were uploaded to the Galaxy-based RepeatExplorer web server [39]. Reads were processed according to the workflow in Figure 2.1, enabling the identification of high copy-number repeat DNA sequences for each organism. Promising repeat families were further analyzed using the Nucleotide BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) available from the National Center for Biotechnology Information (NCBI). Results from each organism were screened for repetitive DNA elements found to have high degrees of homology with elements of the

human genome, common bacteria of the human microbiome, or other parasitic organisms likely to be found within the human gut. Had such sequences been identified as among the most repetitive, they would have been eliminated from further consideration as they would be expected to cause species-specificity challenges during downstream PCR assay development. However, no such conserved highly repetitive elements were identified. Following screening, sequences from each organism, putatively determined to be among the most highly repetitive, were utilized for further assay development (Figure 2.2).

### 2.5.4 Primer and probe design

Candidate primer and probe pairings for each organism, excluding *A. lumbricoides*, were designed using the PrimerQuest online tool (Integrated DNA Technologies, Coralville, IA), utilizing the default parameters for probe-based qPCR. The putative species-specificity of each primer pair was further examined using Primer-BLAST software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi). In the case of *S. stercoralis* the highest copy-number repeat (as determined by RepeatExplorer) was not selected as a target sequence, due to design difficulties associated with the extreme A-T richness of the repeat (A-T % = 80.25). As a result, a second repeat analysis was performed, selecting only for sequence reads with > 30 % G-C content, and a second candidate sequence was selected based on these results. In the case of *A. lumbricoides*, RepeatExplorer analyses of two different sequencing runs performed from two distinct libraries both resulted in the identification of ribosomal and mitochondrial sequences as the most highly repetitive. For this reason, sequences from an existing, proven, primer and probe set targeting the ITS1 region were selected for further analysis [14, 16]. With the

exception of the previously published *A. lumbricoides* probe, all probes were labeled with a 6FAM fluorophore at the 5' end, and were double quenched using the internal quencher ZEN and 3IABkFQ (IOWA BLACK) at the 3' end (Integrated DNA Technologies). This fluorophore-quencher combination was chosen as comparative testing of each probe revealed improved Ct values and greater ΔRn values using this chemistry when compared to typical TAMRA quenching (Figure 2.3). Primer and probe sets for each organism can be found in Table 2.1.

### 2.5.5 Primer and probe validation

#### 2.5.5.1 Primer optimization reactions

In order to determine the optimal primer concentrations for each assay, a concentration matrix was created. For all primers, testing at 62.5 nM, 125 nM, 250 nM, 500 nM and 1000 nM was performed, and all forward primer concentrations were tested in combination with all reverse primer concentrations. Optimization assays were conducted in 7 µl volumes, containing 3.5 µl of 2X TaqMan Fast Universal PCR Master Mix (Life Technologies), 125 nmol of each assay's respective probe, and 2 µl of template DNA at a concentration of 1 ng/µl. Cycling conditions consisted of an initial 2 min incubation step at 50 °C, followed by a 10 min incubation at 95 °C. These incubations were followed by 40 cycles of 95 °C for 15 sec for denaturation, followed by 1 min at 59 °C for annealing and extension. All reactions were conducted using the StepOnePlus Real-Time PCR System (Life Technologies).

#### 2.5.5.2 Determination of assay detection limits

In order to determine the limits of detection for each assay, genomic template DNA stocks were titrated for each parasite. DNA stock concentrations of 1 ng/µl, 100 pg/µl, 10 pg/µl, 1 pg/µl, 100 fg/µl, 10 fg/µl, 1 fg/µl, 100 ag/µl, 10 ag/µl and 1 ag/µl were tested with each assay using the optimized primer concentrations and assays were again conducted in 7 µl total volumes. Reagent concentrations and cycling conditions were identical to those used for primer optimization reactions.

#### 2.5.5.3 Assay species-specificity testing

In order to ensure the species-specificity of each assay, the primer-probe set for each parasite was tested using template DNA from each of the other STH species. Furthermore, each primer-probe combination was tested against human genomic DNA and the DNA of the common gastrointestinal tract commensal *Escherichia coli* (K-12 strain). All template stocks were at a concentration of 1 ng/µl, and all assays were performed using the same reagent volumes and concentrations as used for the primer optimization reactions and for the determination of assay detection limits.

### 2.5.6 Comparative testing of field-collected samples

#### 2.5.6.1 Collection of samples

For comparative assay testing, a panel of 79 samples was employed. All samples had been previously collected as part of the "Wash for Worms" interventional trial in

Timor-Leste (Trial registration: ACTRN12614000680662). The specific procedures used during the collection and storage of these samples have been previously described [41].

#### 2.5.6.2 DNA extraction

All DNA extractions were performed at QIMR Berghofer using the PowerSoil DNA isolation Kit (Mo Bio, Carlsbad, CA, USA) in accordance with the previously described, modified version of the manufacturer's protocol [42]. Following extraction, an aliquot of each sample was retained at QIMR Berghofer and another was shipped to Smith College (Northampton, MA, USA).

#### 2.5.6.3 Real-time PCR testing

DNA extracts from all samples were assayed at both QIMR Berghofer and Smith College. Testing at QIMR Berghofer was conducted using previously described real-time PCR primer/probe sets [4, 13-15], optimized for use in pentaplex assays [42] to test for the presence of *N. americanus*, *T. trichiura*, *Ascaris* ssp., *Ancylostoma* ssp., and *S. stercoralis*. For ease of reporting, hereafter, these assays will be referred to as the "QIMR assays". Testing which occurred at Smith College made use of the optimized, previously undescribed multi-parallel assays for the detection of *N. americanus*, *T. trichiura*, *A. lumbricoides*, *A. duodenale*, and *S. stercoralis* introduced in this manuscript (hereafter referred to as the "Smith assays"). All sample aliquots tested at Smith College were coded blind by QIMR and assay results were not shared between institutions until all testing had been completed.

#### 2.5.6.4 Statistical analysis

Positive, negative and overall agreements were calculated to assess concordance between the QIMR and Smith assays using the equations found in "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests" [43]. Kappa statistics, which account for the possibility that concordance may occur by chance, were calculated for each taxon using R version 3.1.3 and the R package irr [44-45].

#### 2.5.6.5 *Trichuris* speciation of Smith-negative, QIMR-positive samples

Samples testing positive for *Trichuris* ssp. using the QIMR assay, but negative for *Trichuris trichiura* using the Smith assay were further analyzed to determine the identity of the infecting species. Using a previously described primer-probe set targeting the coding sequence for the 18S ribosomal subunit [46], these samples were PCR amplified in 25 µl reactions using the Phusion Hot Start Flex DNA Polymerase Kit (New England Biolabs, Ipswich, MA). PCR reaction conditions were as follows: 16.5 µl PCR-grade water, 400 nM forward primer, 400 nM reverse primer, 0.5 µl dNTPs, 0.75 µl DMSO, 5.0 µl Phusion HF buffer, 0.25 µl Phusion Polymerase, and 1 µl template DNA. Cycling conditions consisted of an initial denaturing step at 98 °C for 15 min, followed by 35 cycles of 98 °C for 10 sec, 56 °C for 15 sec, and 72 °C for 15 sec. Following 35 cycles, a final extension step of 72 °C for 7 min was performed. PCR products were then sequenced using standard Sanger methodology and resulting sequence data were analyzed using NCBI's Nucleotide BLAST tool.

#### 2.5.6.6 Differentiation of *A. duodenale* and *A. ceylanicum* infections

Samples testing positive for *Ancylostoma* ssp. using the QIMR assay, but negative for *A. duodenale* using the Smith assay, underwent further testing to discriminate between infection with *A. duodenale* and *A. ceylanicum*. For differential detection, a semi-nested PCR-Restriction Fragment Length Polymorphism (RFLP) assay was employed, and samples were tested in accordance with the published protocol [47]. Digestion of *Ancylostoma* ssp. PCR product using the MvaI enzyme (Life Technologies) was indicative of the presence of *A. ceylanicum*, while digestion with Psp1406I (Life Technologies) was indicative of *A. duodenale*.

## 2.6 Results

### 2.6.1 Primer and probe validation

#### 2.6.1.1 Primer optimization

The use of a primer matrix resulted in the determination of optimal concentrations for each assay. Optimal conditions were determined to be those at which Ct values were lowest when testing 2 µl of the appropriate template stock at a concentration of 1 ng/µl. For *N. americanus*, *A. duodenale*, and *S. stercoralis*, the optimal concentrations were determined to be 250 nM for both forward and reverse primers. For *A. lumbricoides* the optimal concentrations were determined to be 62.5 nM for both forward and reverse primers, and for *T. trichiura*, the optimal concentrations were determined to be 62.5 nM for the forward primer and 250 nM for the reverse primer.

#### 2.6.1.2 Assay sensitivities

In order to demonstrate the detection limits of each multi-parallel PCR assay, genomic DNA stocks for each parasite were serially diluted, and optimal primer concentrations for each assay were used to test 2 µl of the appropriate template at concentrations of 1 ng/µl, 100 pg/µl, 10 pg/µl, 1 pg/µl, 100 fg/µl, 10 fg/µl, 1 fg/µl, 100 ag/µl, 10 ag/µl, and 1 ag/µl. For all species, consistent detection of parasite DNA was possible at all concentrations at or above 1 fg/µl. For the detection of *A. duodenale*, consistent detection occurred at all concentrations at or above 10 ag/µl, and for the detection of *N. americanus*, sporadic detection proved possible at the 100 ag/µl and 10 ag/µl concentrations.

#### 2.6.1.3 Assay specificities

To verify that the primer-probe combinations for the detection of each parasite were species-specific, each optimized assay was tested against genomic template DNA from each of the other parasite species included within this multi-parallel platform. They were also tested against human genomic DNA and *E. coli* genomic DNA. In no instance did species-specificity testing result in the amplification of any non-target DNA template, indicating that each assay demonstrated excellent species specificity.

### 2.6.2 Comparative Testing of Field Collected Samples

A panel of 79 blindly-coded patient samples, obtained in Timor-Leste as part of a previously described study [42], was tested using the newly described multi-parallel Smith assays, as well as the previously described, multiplex real-time PCR detection

methodology (QIMR assay) (Table 2.2, Table 2.3). As samples were patient-obtained and no true "gold standard" exists for the detection of the various STH infections examined here, it is difficult to definitively determine whether increased sample positivity is a result of improved assay detection limits or non-specific, off-target amplification. For this reason, the comparative performances of each assay were assessed through calculations of positive, negative, and overall agreement [43]. For the detection of *N. americanus*, a positive agreement (PA) of 100 % and a negative agreement (NA) of 61 % were calculated. This resulted in an overall agreement (PO) of 85 % (Kappa 0.658). Use of the Smith assay resulted in the detection of 60 positive samples, while the QIMR assay resulted in the detection of 48 positives. All 48 QIMR-positive samples were among the 60 positive samples detected using the Smith methodology.

For the detection of *A. lumbricoides*, a PA of 100 %, an NA of 82 %, and a PO of 91 % (Kappa 0.822) were seen. The Smith assay for *A. lumbricoides* detection resulted in the identification of 47 positive samples, while the corresponding QIMR assay resulted in 40 positives. Again, all 40 QIMR-positive samples were among the 47 Smith-positive samples which were identified.

Detection of *Trichuris* gave a PA of 71 %, an NA of 88 % and a PO of 85 % (Kappa 0.580). Sample examination using the Smith assay identified 18 positive extracts, while examination with the QIMR assay identified 14 positives. However, only 10 positives were common to both assays, with 8 samples identified as positive only by the Smith assay, and 4 samples demonstrating the presence of parasite DNA using only the QIMR methodology. Amplification in control reactions demonstrated that the QIMR assay, but not the Smith assay, would provide for the detection of the closely related parasite *Trichuris*

*vulpis*, a whipworm species common to canines, but also known to cause zoonotic infection [48-49]. As *Trichuris* spp. including *T. vulpis*, *Trichuris suis*, and *Trichuris ovis* have a wide geographic distribution with increased prevalence in tropical and sub-tropical locations [50-51], the four QIMR-positive, Smith-negative samples were sequenced to determine the identity of the *Trichuris* species present within these samples. BLAST analysis indicated that two of the samples contained DNA from the ruminant parasite *T. ovis* (E values = 0.0). Unfortunately, two independent trials failed to produce usable sequence for the remaining two samples, after which both sample stocks had been exhausted, making further examination impossible.

Examination of all 79 samples for the presence of *S. stercoralis* resulted in the detection of only a single positive sample. This single sample was identified using both the Smith and QIMR assays. Sample examination for the presence of *Ancylostoma* resulted in the identification of 22 *Ancylostoma* spp. positive samples using the QIMR methodology. However, not a single *A. duodenale*-positive sample was identified using the Smith assay. As the zoonotic parasite *Ancylostoma ceylanicum* has been suspected of causing human infection in Timor-Leste [52], a previously described, semi-nested PCR-RFLP assay was employed to discriminate infection with *A. duodenale* from infection with *A. ceylanicum* [47]. In this assay, an MvaI restriction digest of PCR product is indicative of the presence of *A. ceylanicum*, while digestion with Psp1406I is indicative of *A. duodenale*. 21 of the 22 *Ancylostoma* spp. positive samples were digested by MvaI, identifying the infections as *A. ceylanicum* in origin. Two independent PCR trials (four replicates) failed to amplify the remaining *Ancylostoma* spp.-positive sample, preventing a definitive determination of the identity of the parasite in that sample.

Because a sizeable panel of field-collected samples was analyzed using the two different real-time PCR methodologies discussed here, a comparison of Ct values was conducted for all samples testing positive for a given parasite by both the Smith and QIMR methods (Table 2.3). All 10 samples demonstrating positive results for *T. trichiura* when tested by both assays showed lower Ct values using the Smith methodology (mean difference in Ct value = 7.86 +/- 2.46). Examination for *N. americanus* resulted in a similar pattern, with all 48 samples testing positive by both methodologies possessing lower Ct values when tested using the Smith assay (mean difference in Ct value = 4.94 +/- 1.22). In the case of *A. lumbricoides*, Ct values were lower using the QIMR methodology for 38 of 40 samples demonstrating positive results for both assays. However, at 0.896 +/- 0.767, the mean difference in Ct values was low. For *S. stercoralis* testing, only a single positive sample was identified. This sample possessed a lower Ct value when tested using the Smith assay. As no samples tested positive for *Ancylostoma* using the Smith assay (QIMR-positive samples were demonstrated to be *A. ceylanicum*), a Ct comparison could not be made.

## 2.7 Discussion

In light of their impact on global health, the importance of optimal and species-specific diagnostic methods for the detection of soil transmitted helminths cannot be overestimated. While current molecular assays making use of ribosomal and mitochondrial targets have vastly improved the diagnosis of STH infection, these targets are frequently sub-optimal, potentially leaving low-level infections undiagnosed. Furthermore, such sequences may lack the species-specificity required to discriminate between different

species of the same genus. In contrast, assays targeting high copy-number repetitive sequences improve upon assay detection limits, as many eukaryotic pathogens contain large numbers of such non-coding repeat DNA elements. Accordingly, by coupling the high throughput nature of NGS with the Galaxy-based RepeatExplorer computational pipeline, a cost effective, accurate, and expedited methodology for the identification of high copy-number repeat DNA elements was developed. Through the design of real-time PCR primer/probe pairings that uniquely target such repetitive sequences in a species-specific manner, diagnostic accuracy and limits of detection are improved dramatically when compared with microscopy-based diagnostic techniques and PCR-based diagnostics targeting mitochondrial or ribosomal sequences. Utilizing this strategy, we have successfully identified novel target sequences for the detection of *N. americanus*, *A. duodenale*, *T. trichiura*, and *S. stercoralis*. Furthermore, we have demonstrated the consistent detection of genomic DNA from each target organism at quantities of 2 fg or less, and have presented evidence to suggest improved limits of detection and species–specificity relative to an established and validated PCR diagnostic methodology [Llewellyn, 2016]. Although further testing utilizing "spiked" samples containing known quantities of eggs/larvae is currently underway, 2 fg of DNA is far less than the quantity present within a single fertilized egg or L1 larvae of each species [53-55] (Table 2.4). In principle, we have therefore demonstrated the potential of these assays to detect a single egg within a tested patient stool sample.

While the high copy-number nature of non-coding repetitive sequence elements makes them attractive diagnostic targets, such elements also frequently demonstrate rapid evolutionary divergence [56-57]. This divergence increases the diagnostic appeal of these

sequences, as divergence reduces the risk for non-specific, off-target amplification, a characteristic essential for the development of species-specific PCR assays capable of discriminating between closely related organisms. Accordingly, while additional testing against genomic DNA from a growing panel of closely related parasites will continue to be used to evaluate the species-specificity of each selected primer/probe set, we have successfully demonstrated that each Smith assay does not amplify off-target templates from any other parasite species included within this multi-parallel platform. Furthermore, by employing a semi-nested PCR-RFLP tool, we were able to successfully demonstrate that our assay for the detection of *A. duodenale* does not amplify the closely related parasite *A. ceylanicum*. In contrast, the previously published primer/probe set employed for comparative testing was unable to distinguish between these two species, resulting in consistent off-target amplification of *A. ceylanicum* DNA. Similarly, while our *T. trichiura* assay failed to amplify four samples containing genetic material from other *Trichuris* spp., the comparative QIMR assay again demonstrated non-specific, off-target amplification for at least two of these samples, as sequence analysis demonstrated the presence of DNA from the ruminant parasite *T. ovis*. Taken together, these findings support the notion that improved assay species-specificity results from non-coding, repeat-based PCR assay design. Of note, to our knowledge, this is the first example of *T. ovis* potentially serving as a causative agent of zoonotic infection. However, as sheep are considered a major agricultural commodity of Timor-Leste [58], the possibility exists that individuals testing positive for *T. ovis* may have ingested intestinal material from an animal harboring infection, making it conceivable that the *T. ovis* DNA present was not the result of zoonotic

infection. Given that *T. ovis* is not known to cause human infection, further exploration of this possible zoonosis is warranted.

Attempting to design a non-coding, repetitive DNA sequence-based assay for the species-specific detection of *A. lumbricoides* presented a unique set of challenges. *A. lumbricoides*, like many species of Ascaridae, discards large portions of its highly repetitive, non-coding genomic DNA during embryonic development. This process, known as chromosome diminution, eliminates the presence of such DNA elements from post-embryonic somatic cells [59-61]. Presumably for this reason, two separate repeat analyses, performed on two distinct library preparations, failed to identify any repetitive sequences with copy numbers greater than ribosomal and mitochondrial targets. Accordingly, a previously described primer/probe set targeting the ITS1 ribosomal region was chosen for inclusion in our multi-parallel platform [14, 16]. In order to improve diagnostics for this parasite, further analysis of *A. lumbricoides* using DNA extracted from eggs alone (before chromosome diminution) will be undertaken.

In addition to the potential detection limit improvements and species-specificity gains realized when diagnostically targeting non-coding repetitive DNA sequences, designing multi-parallel assays provides another unique set of advantages over previous design strategies [4]. By reducing the number of tests required, multiplex assays can provide labor and reagent savings over alternative diagnostic measures when used in environments that harbor the full complement of organisms targeted by the assay [62-63]. However, as the geographic distribution of STH species is not uniform, the use of multi-parallel assays makes it possible to select only the assays appropriate for a given location, reducing primer/probe costs associated with testing for unnecessary targets [4]. By running

these assays as “small-volume” 7 µl reactions, reagent use is minimized, resulting in cost savings. Furthermore, as multi-parallel reactions are run independently, this enables the development of new assays for new pathogens and their subsequent addition to the testing platform without the complex re-optimization of assay conditions required for multiplex PCR assays.

While reagent costs associated with performing molecular diagnostic testing are higher than costs associated with conducting traditional microscopy-based diagnostics, expenses associated with molecular techniques are declining as improved reagents and enzymes have allowed reaction volumes to decrease, minimizing reagent needs [4, 64]. Furthermore, reagent improvements have increased the practicality of sample pooling, a practice already adopted by many tropical disease surveillance and diagnostic efforts [65-69]. Such pooling allows for cost-reducing high-throughput screening of stool samples [70-71]. Thus, while the total cost associated with performing a duplicate Kato-Katz thick smear under field conditions has been estimated at $2.06 [72] and we estimate the total cost associated with the duplicate testing a single stool sample using all five multi-parallel assays to be approximately $10, the pooling of as few as five samples would render small volume, multi-parallel PCR testing more cost effective than Kato-Katz testing. Furthermore, molecular diagnostic accuracy and reliability provide increased clarity of results [64], allowing for the implementation of more informed and effective treatment and control strategies. Such improvements in efficiency result in greater programmatic gains, drastically reducing long-term costs and expenses of control or elimination programs.

One profound shortcoming which hampers STH diagnostic development is the lack of a reliable gold standard for detection [8]. While still used in many clinical, mapping,

and research efforts, microscopy-based methodologies are known to lack both adequate limits of detection and species-specificity of detection [3-5, 10, 64]. Similarly, while currently available molecular methods have greatly improved upon many of the shortcomings inherent to microscopy, the use of sub-optimal ribosomal or mitochondrial targets possessing relatively high degrees of conservation can result in both false-negative, and off-target, false-positive results. Thus, a gold standard of detection is sorely needed. Unfortunately, without a definitive method for assigning positive/negative status to an unknown sample, distinguishing improved limits of detection from false-positive amplification can be difficult. Nonetheless, comparative assay testing remains an important aspect of designing any diagnostic test. As such, we believe the evaluation of Timor-Leste patient samples presented in this paper provides strong evidence for improved limits of detection when utilizing the newly described Smith assays. While strain-specific genetic differences arising within divergent geographic isolates could present detection challenges, testing on a limited number of patient-derived samples from Argentina and Ethiopia aimed at providing evidence for the global applicability of these multi-parallel assays is currently underway. Additional studies to further validate these assays on a variety of geographic isolates will continue.

In all instances, and for all parasites excluding *Ancylostoma* and *Trichuris* (where off-target amplification of *A. ceylanicum* and *T. ovis* by the QIMR assay was demonstrated), each Timor-Leste patient sample that provided a positive QIMR assay result also demonstrated positivity with the corresponding Smith assay. Furthermore, all *N. americanus*, *T. trichiura*, and *S. stercoralis* samples that were positive by both assays exhibited lower Ct values for the Smith assay results. These findings strongly suggest

improved limits of detection for the Smith assays, and support our contention that samples returning Smith assay positive results, but QIMR assay negative results, are likely low-level positives escaping detection by the sub-optimal PCR platform. This conclusion is further supported by the finding that the Smith assays do not show off-target amplification of any other STH parasites, human DNA or *E. coli* DNA.

As both the QIMR and Smith assays for the detection of *A. lumbricoides* make use of the same previously published primer/probe combination [14, 16], comparative assay testing for this parasite provided results which were more difficult to interpret. As increased reaction volumes are known to frequently improve detection limits for an assay, likely due to the large volume nature of the QIMR assay (25 µl vs. 7 µl for Smith), 38 of 40 samples returning positive results for both testing platforms demonstrated lower Ct values when examined using the QIMR method. Interestingly, despite this tendency for QIMR testing to result in lower Ct values, seven samples identified as positive using the Smith assay were found to be QIMR-negative. In contrast, not a single sample was found to be QIMR-positive and Smith-negative. As the QIMR assays are multiplexed, one explanation for this apparent contradiction is that the multiplex methodology failed to detect *A. lumbricoides* in a subset of samples that were positive for multiple STH parasites (Table 2.3). Such failures are known to occur in multiplex reactions, particularly when primer concentrations are suboptimal, as reagents are utilized for the amplification of a more prevalent target, preventing the amplification of the lower copy-number target sequences within the sample [73]. Alternatively, while the results of our assay specificity testing present compelling evidence to the contrary, the possibility of false positive amplification cannot be definitively ruled out.

Non-coding repetitive DNA elements are found in nearly all eukaryotic organisms. Such sequences are typically highly divergent, and frequently exist in high copy-number. These characteristics make them ideal molecular diagnostic targets, particularly for the detection of pathogens such as the STHs, which remain an underdiagnosed, poorly mapped global health concern. By applying next-generation sequencing technology to the challenge of repeat DNA discovery, we have designed highly specific multi-parallel PCR assays with improved limits of detection over existing diagnostic platforms. We believe that these assays will greatly aid in the global efforts to map STH infection, facilitating accurate disease prevalence estimates. Furthermore, we intend to apply this approach to molecular target discovery of other parasitic organisms and NTDs, as optimal limits of detection and species-specificity of detection are vital to all diagnostic efforts. This is particularly true when implementing diagnostics in climates of decreasing disease prevalence. Accordingly, as NTD elimination efforts continue to progress, optimized assays will play an increasingly critical role in the detection of sporadic and focal infections and the monitoring for disease recrudescence.

## 2.8 Acknowledgements

The authors thank Dr. Peter Nejsum (University of Copenhagen, Denmark) for graciously providing us with *T. trichiura* worms. Additionally, we thank Dr. Peter Hotez and Dr. Bin Zhan (Baylor College of Medicine, Houston, TX, USA) for supplying both *N. americanus* and *A. duodenale* parasites, Dr. James Lok (University of Pennsylvania, Philadelphia, PA, USA) for providing *S. stercoralis* genomic DNA, and Dr. Roger Prichard (McGill University, Montreal, Québec, Canada), Dr. Matthew Berryman (Wellcome Trust

Sanger Institute, Hinxton, Cambridge, UK) and Nancy Holroyd (Wellcome Trust Sanger Institute) for their gifts of *A. lumbricoides* DNA. We would also like to extend our sincere appreciation to the WASH for Worms Investigators, especially Dr. Susana Vaz Nery and the field team in Timor-Leste. Furthermore, we thank Dr. Eric Ottesen and Dr. Patrick Lammie (Task Force for Global Health, Decatur, GA, USA) for their constant support and advice.

**Table 2.1 Selected primer and probe sequences for each multi-parallel assay.**

| Parasite | Forward Primer | Reverse Primer | Probe |
|---|---|---|---|
| ***Necator americanus*** | 5'- CCAGAATCGCCACAAATTGTAT -3' | 5'- GGGTTTGAGGCTTATCATAAAGAA -3' | 5'-/56 FAM/CCCG ATTTG/ZEN/AGCT GAATTGTCAAA/ 3IABkFQ/-3' |
| ***Ancylostoma duodenale*** | 5'- GTATTTCACTCATATGATCGAGTGTTC -3' | 5'- GTTTGAATTTGAGGTATTTCGACCA -3' | 5'-/56-FAM/TGAC AGTGT/ZEN/GTC ATACTGTGGAAA/ 3IABkFQ/-3' |
| ***Trichuris trichiura*** | 5'- GGCGTAGAGGAGCGATTT -3' | 5'- TACTACCCATCACACATTAGCC -3' | 5'-/56-FAM/TTTG CGGGC/ZEN/GAG AACGGAAATATT/ 3IABkFQ/-3' |
| ***Strongyloides stercoralis*** | 5'- CGCTCCAGAATTAGTTCCAGTT -3' | 5'- GCAGCTTAGTCGAAAGCATAGA -3' | 5'-/56-FAM/ACA GTCTCC/ZEN/AGT TCACTCCAGAAGA GT/3IABkFQ/-3' |
| ***Ascaris lumbricoides*** | 5'- GTAATAGCAGTCGGCGGTTTCTT -3' | 5'- GCCCAACATGCCACCTATTC -3' | 5'-/56-FAM/ TTG GCGGACAATTGC ATGCGAT/MBG/- 3' |

**Table 2.2 Comparative assay results for each species of parasite.**

| ***Necator americanus*** | *Smith-positive* | *Smith-negative* |
|---|---|---|
| *QIMR-positive* | 48 | 0 |
| *QIMR-negative* | 12 | 19 |
| ***Ascaris lumbricoides*** | *Smith-positive* | *Smith-negative* |
| *QIMR-positive* | 40 | 0 |
| *QIMR-negative* | 7 | 32 |
| ***Trichuris trichiura*** | *Smith-positive* | *Smith-negative* |
| *QIMR-positive* | 10 | 4* |
| *QIMR-negative* | 8 | 57 |
| ***Strongyloides stercoralis*** | *Smith-positive* | *Smith-negative* |
| *QIMR-positive* | 1 | 0 |
| *QIMR-negative* | 0 | 78 |
| ***Ancylostoma duodenale*** | *Smith-positive* | *Smith-negative* |
| *QIMR-positive* | 0 | 22 ** |
| *QIMR-negative* | 0 | 57 |

***** Two of these four samples contained *T. ovis*. The identity of the pathogen in the remaining 2 samples could not be determined due to a lack of material.

** 21 of these 22 samples contained *A. ceylanicum*. The identity of the pathogen in the 22nd sample could not be determined due to a lack of material.

## Table 2.3 Comparative assay results for patient-obtained samples from Timor-Leste.

| | *Trichuris* | | *Ascaris* | | *Necator* | | *Strongyloides* | | *Ancylostoma* | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sample #** | **QIMR Assay* (Ct value)** | **Smith Assay* (Ct value)** | **QIMR Assay* (Ct value)** | **Smith Assay* (Ct value)** | **QIMR Assay* (Ct value)** | **Smith Assay* (Ct value)** | **QIMR Assay* (Ct value)** | **Smith Assay* (Ct value)** | **QIMR Assay* (Ct value)** | **Smith Assay* (Ct value)** |
| **1** | | | 20.5 | 21.0 | 18.3 | 14.2 | | | | |
| **2** | 29.4 | 21.5 | 12.7 | 12.9 | 20.6 | 16.3 | | | | |
| **3** | 31.2 | 22.3 | | | | | | | | |
| **4** | 34.2 | 24.5 | 27.4 | 27.0 | | 25.8 | | | | |
| **5** | 35.5 | 23.2 | | | | 33.6 | | | | |
| **6** | 26.5 | 21.9 | 11.5 | 11.9 | 15.3 | 10.8 | | | 20.7 | |
| **7** | | | 22.4 | 22.5 | 22.4 | 17.3 | | | | |
| **8** | 34.2 | 24.7 | 13.7 | 13.8 | 22.0 | 18.0 | | | | |
| **9** | 18.5 | 28.7 | | | | | | | | |
| **10** | | | 18.1 | 18.3 | 18.2 | 13.0 | 25.2 | 24.2 | | |
| **11** | | | 13.3 | 13.8 | 18.2 | 14.0 | | | 21.93 | |
| **12** | | | | | | | | | | |
| **13** | | | 14.4 | 14.9 | 22.0 | 18.5 | | | | |
| **14** | | | 15.6 | 16.2 | 25.4 | 21.1 | | | | |
| **15** | | | | | 22.5 | 17.2 | | | | |
| **16** | | | 30.7 | 32.7 | 30.9 | 26.4 | | | 35.3 | |
| **17** | | 22.6 | 18.5 | 18.5 | 25.4 | 18.8 | | | | |
| **18** | | | 11.6 | 13.2 | 19.3 | 14.7 | | | 22.07 | |
| **19** | | | | 33.8 | 18.1 | 13.3 | | | | |
| **20** | | | 13.9 | 14.4 | 25.9 | 17.0 | | | | |
| **21** | | | 18.5 | 17.1 | 23.9 | 15.7 | | | | |
| **22** | | | | 31.5 | | 36.5 | | | | |
| **23** | | | 14.1 | 14.3 | 22.0 | 14.0 | | | 23.02 | |
| **24** | | | | | | | | | | |
| **25** | | | 14.8 | 16.7 | | 28.7 | | | | |
| **26** | | | 24.6 | 25.3 | | 34.4 | | | | |
| **27** | | | 14.1 | 15.5 | 18.1 | 13.2 | | | 22.25 | |
| **28** | | | | 32.3 | | 27.7 | | | 24.87 | |
| **29** | | | | | | | | | | |
| **30** | | | 14.8 | 15.6 | 20.2 | 15.3 | | | | |
| **31** | | | | | 18.0 | 14.4 | | | 21.545 | |
| **32** | | 28.8 | | | 25.8 | 20.5 | | | | |
| **33** | | | | | | | | | | |
| **34** | 22.8 | 16.8 | | | 29.0 | 23.7 | | | | |
| **35** | | | 13.3 | 14.0 | 19.0 | 14.3 | | | | |
| **36** | | | | | | | | | | |
| **37** | | | 26.4 | 27.7 | | | | | 23 | |
| **38** | | | 14.0 | 14.9 | 16.0 | 11.5 | | | 21.415 | |
| **39** | | | | 32.1 | 21.9 | 17.8 | | | | |
| **40** | | | 9.8 | 10.8 | | 35.0 | | | | |
| **41** | | | 21.0 | 21.6 | 20.8 | 16.6 | | | 24.965 | |
| **42** | | | | 35.0 | | 31.7 | | | | |
| **43** | | | 14.5 | 15.4 | | | | | | |
| **44** | | 33.5 | | | 15.5 | 10.7 | | | 20.15 | |
| **45** | 23.3 | 18.0 | | | | 28.9 | | | | |
| **46** | | | 27.4 | 28.7 | 20.3 | 14.5 | | | | |
| **47** | | | | | | | | | | |
| **48** | | | 28.4 | 29.5 | 15.3 | 10.2 | | | | |
| **49** | | | 16.5 | 17.6 | | 30.3 | | | | |
| **50** | 31.0 | | | | | | | | | |
| **51** | 22.0 | | | | | | | | | |
| **52** | 22.7 | | | | | | | | | |
| **53** | 28.1 | | | | | | | | | |
| **54** | | | | 35.0 | 21.1 | 15.6 | | | 23.235 | |
| **55** | | | | | 24.3 | 19.2 | | | 25.725 | |
| **56** | | | 17.3 | 18.6 | | 29.9 | | | | |
| **57** | | | | | | | | | | |
| **58** | 32.8 | 27.4 | | | 24.4 | 20.2 | | | | |
| **59** | | | | | | 35.1 | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **60** | | | | | 20.9 | 16.4 | | | 25.735 | |
| **61** | | | | | 26.9 | 19.4 | | | | |
| **62** | | | 30.4 | 31.6 | 21.2 | 16.7 | | | | |
| **63** | | | 13.8 | 15.1 | 16.6 | 11.6 | | | | |
| **64** | | 30.8 | | | 17.2 | 12.1 | | | 22.36 | |
| **65** | | | | | | | | | | |
| **66** | | | 13.9 | 15.7 | 19.4 | 15.2 | | | 20.93 | |
| **67** | | | 29.5 | 29.9 | 18.5 | 13.7 | | | | |
| **68** | | | 9.4 | 12.1 | | | | | | |
| **69** | | | | | 15.0 | 10.5 | | | 20.935 | |
| **70** | | | 13.3 | 14.2 | 19.2 | 14.1 | | | | |
| **71** | | 21.7 | 10.5 | 12.3 | 17.6 | 13.2 | | | 19.66 | |
| **72** | 31.1 | 22.2 | | | | | | | | |
| **73** | | | | | | | | | | |
| **74** | | | | | 24.3 | 18.8 | | | 26.995 | |
| **75** | | | | | 21.2 | 17.0 | | | 20.705 | |
| **76** | | | 28.1 | 29.1 | 20.4 | 17.0 | | | 23.47 | |
| **77** | | 28.6 | 16.4 | 18.7 | 22.7 | 20.7 | | | | |
| **78** | | 30.9 | 17.6 | 18.4 | 22.9 | 17.7 | | | | |
| **79** | | | | 33.4 | 20.2 | 15.3 | | | | |

*Results are reported as mean Ct values.

**Table 2.4 Estimated minimum quantities of DNA in the eggs of each species of STH.**

| Species | Haploid Genome Size (Genbank Assembly Accession) | Estimated Minimum Quantity of DNA/Diploid Cell | Estimated Number of Diploid Cells/Egg [55] | Estimated Total Quantity of DNA/Egg |
|---|---|---|---|---|
| *Necator americanus* | 244 Mb [54] (GCA_000507365.1) | 0.54 pg | 4-8 | 2.16 pg – 4.32 pg |
| *Ascaris lumbricoides* | 317 Mb (GCA_000951055.1) | 0.70 pg | As few as 1 | 0.70 pg |
| *Trichuris trichiura* | 75 Mb (GCA_000613005.1) | 0.17 pg | 1 | 0.17 pg |
| *Ancylostoma duodenale* | 332 Mb (GCA_000816745.1) | 0.73 pg | 4-8 | 2.92 pg – 5.84 pg |
| *Strongyloides stercoralis* | 42 – 60 Mb [53] (GCA_000947215.1) | 0.09 pg – 0.13 pg | 2-8* | 0.18 pg – 1.04 pg* |

* As *S. stercoralis* eggs typically hatch in the intestinal lumen, this parasite is released into the stool as rhabditiform larvae which possess an even greater number of cells. Therefore, a patient stool sample harboring a single larval worm will contain a much higher quantity of parasite DNA than the quantity listed here.

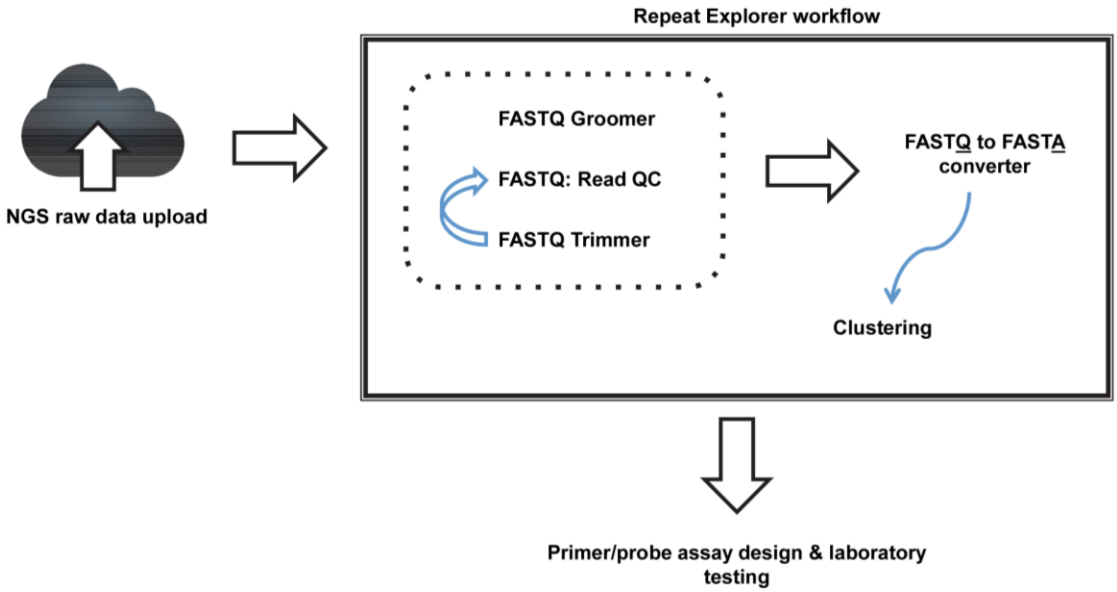


**Figure 2.1 Workflow for repeat analysis**. Output data from a next-generation sequencing run are uploaded to the RepeatExplorer Galaxy-based platform. During the QC and manipulation phase, the *FASTQ Groomer* tool is used to convert sequence reads into Sanger format. The *FASTQ: READ QC* tool is then used to verify the quality of the reads before removing unnecessary sequence (i.e. adapter sequences, etc.) from the ends of each read using the *FASTQ Trimmer* tool. The QC analysis is then repeated, and the *FASTQ to FASTA converter* tool is used to convert each read into FASTA format. Using these DNA sequence reads as input, sequences undergo clustering, during which an "all-to-all" sequence comparison is performed, and similar sequences are grouped together into clusters. Clusters containing the most highly repetitive sequences are then selected as putative diagnostic targets to be used for primer and probe-based real-time PCR assay design.

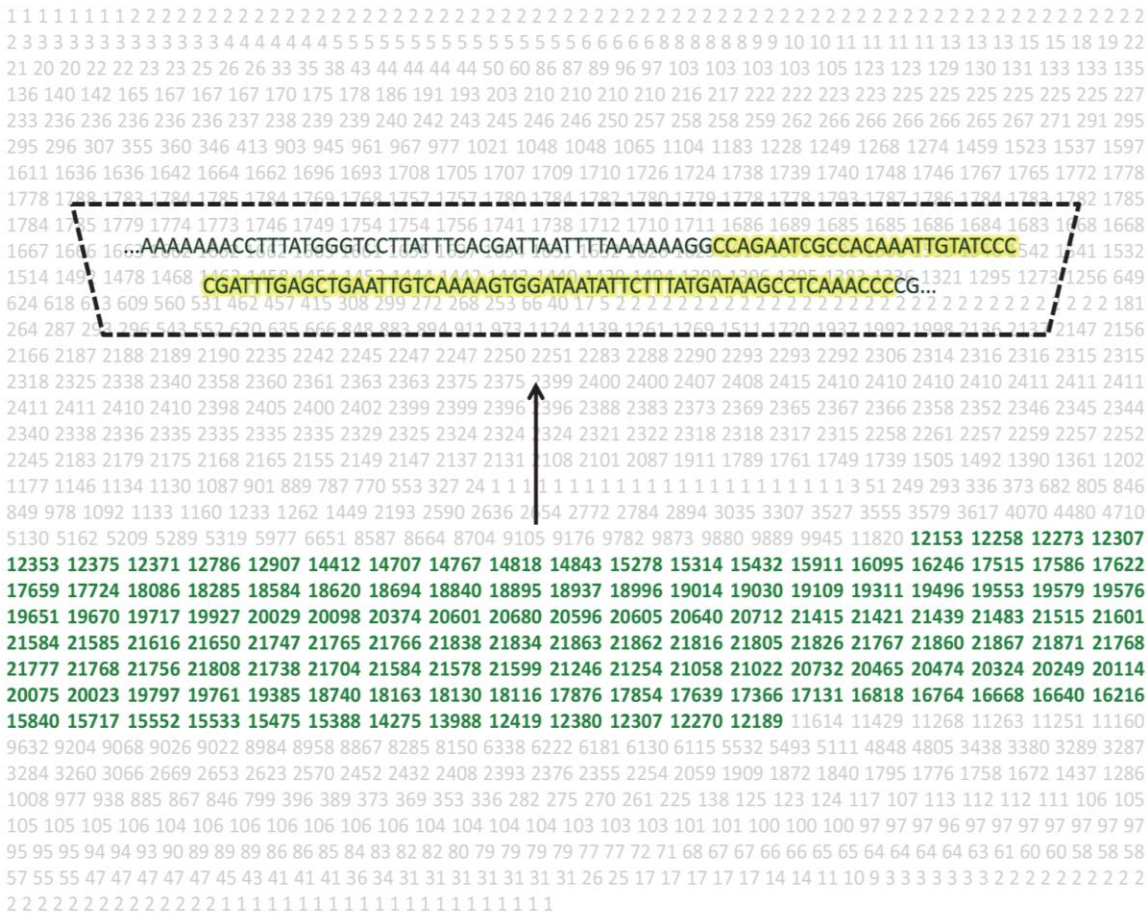


**Figure 2.2 Illustrative output from RepeatExplorer analysis of *Necator americanus*.** During "clustering" each nucleotide within a cluster is assigned a number. That number corresponds to how many individual next-generation sequencing reads that particular nucleotide appeared in. Using this output, a stretch of the most abundant nucleotides (depicted in green within the larger cluster's sequence) is selected, and the corresponding nucleotides (highlighted in yellow) are selected as the candidate sequence from which the primers and probe are designed.

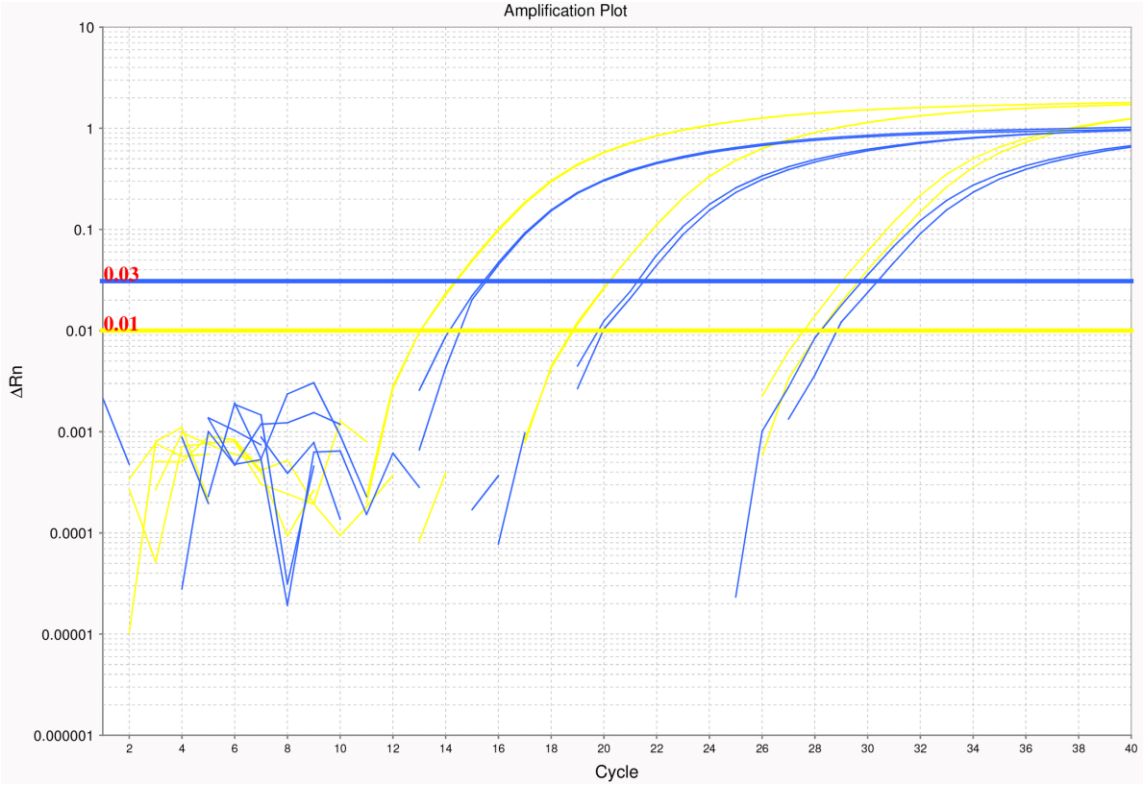


**Figure 2.3 Comparative probe testing.** For each novel probe design, FAM-TAMRA and double quenched FAM-ZEN-IOWA BLACK probes were synthesized. Comparative testing revealed that double quenched probes outperformed traditional probes, as evidenced by lower Ct values and greater ΔRn values. The plot above demonstrates these findings with the amplification of three concentrations of *N. americanus* template DNA using both double quenched (yellow) and traditional (blue) probe designs.

## 2.9 References


1. Hotez PJ, Bundy DAP, Beegle K, Brooker S, Drake L, de Silva N, et al. Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis. In: Disease Control Priorities in Developing Countries. Washington: World Bank; 2006. pp. 467-482.

2. Buonfrate D, Formenti F, Perandin F, Bisoffi Z. Novel approaches to the diagnosis of *Strongyloides stercoralis* infection. Clin Microbiol Infect. 2015; 21: 543-552.

3. Schär F, Odermatt P, Khieu V, Panning M, Duong S, Muth S, et al. Evaluation of real-time PCR for *Strongyloides stercoralis* and hookworm as diagnostic tool in asymptomatic schoolchildren in Cambodia. Acta Trop. 2013; 126: 89-92.

4. Mejia R, Vicuña Y, Broncano N, Sandoval C, Vaca M, Chico M, et al. A novel, multi-parallel, real-time polymerase chain reaction approach for eight gastrointestinal parasites provides improved diagnostic capabilities to resource-limited at-risk populations. Am J Trop Med Hyg. 2013; 88: 1041-1047.

5. Gordon CA, McManus DP, Acosta LP, Olveda RM, Williams GM, Ross AG, et al. Multiplex real-time PCR monitoring of intestinal helminthes in humans reveals widespread polyparasitism in Northern Samar, the Philippines. Int J Parasitol. 2015; 45: 477-483.

6. Anderson RM, Schad GA. Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. Trans R Soc Trop Med Hyg. 1985; 79: 812-825.

7. Booth M, Vounatsou P, N'Goran EK, Tanner M, Utzinger J. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. Parasitology. 2003; 127: 525-531.

8. Nikolay B, Brooker SJ, Pullan RL. Sensitivity of diagnostic tests for human soil-transmitted helminths infections: a meta-analysis in the absence of a true gold standard. Int J Parasitol. 2014; 44: 765-774.

9. Knopp S, Salim N, Schindler T, Voules DAK, Rothen J, Lweno O, et al. Diagnostic accuracy of Kato-Katz, FLOTAC, Baermann, and PCR Methods for the Detection of light-intensity hookworm and *Strongyloides stercoralis* Infections in Tanzania. Am J Trop Med Hyg. 2014; 90: 535-545.

10. Haque R, Mondal D, Karim A, Molla IH, Rahim A, Faruque ASG, et al. Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. Clin Infect Dis. 2009; 48: 1191-1197.

11. Phuphisut O, Yoonuan T, Sanguankiat S, Chaisiri K, Maipanich W, Pubampen S, et al. Triplex Polymerase Chain Reaction Assay for Detection of Major Soil-Transmitted Helminths, *Ascaris lumbricoides, Trichuris trichiura, Necator americanus*, in Fecal Samples. Southeast Asian J Trop Med Public Health. 2014; 45: 267-275.

12. Basuni M, Muhi J, Othman N, Verweij JJ, Ahmad M, Miswan N, et al. A pentaplex real-time polymerase chain reaction assay for detection of four species of soil-transmitted helminths. Am J Trop Med Hyg. 2011; 84: 338-343.

13. Verweij JJ, Brienen EAT, Ziem J, Yelifari L, Polderman AM, Van Lieshout L. Simultaneous Detection and Quantification of *Ancylostoma duodenale, Necator americanus*, and *Oesophagostomum bifurcum* in Fecal Samples Using Multiplex Real-Time PCR. Am J Trop Med Hyg. 2007; 77: 685-690.

14. Wiria AE, Prasetyani MA, Hamid F, Wammes LJ, Lell B, Ariawa I, et al. Does treatment of intestinal helminth infections influence malaria? Background and methodology of a longitudinal study of clinical, parasitological and immunological parameters in Nangapanda, Flores, Indonesia (ImmunoSPIN Study). BMC Infect Dis. 2010; 10: 77.

15. Verweij JJ, Canales M, Polman K, Ziem J, Brienen EAT, Polderman AM, et al. Molecular diagnosis of *Strongyloides stercoralis* in fecal samples using real-time PCR. Trans R Soc Trop Med Hyg. 2009; 103: 342-346.

16. Liu J, Gratz J, Amour C, Kibiki G, Becker S, Janaki L, et al. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. J Clin Microbiol. 2013; 51: 472-480.

17. Janwan P, Intapan PM, Thanchomnang T, Lulitanond V, Anamnart W, Maleewong W. Rapid detection of *Opisthorchis viverrini* and *Strongyloides stercoralis* in human fecal samples using a duplex real-time PCR and melting curve analysis. Parasitol Res. 2011; 109: 1593-1601.

18. Fabian S, Odermatt P, Khieu V, Panning M, Duong S, Muth S, et al. Evaluation of real-time PCR for *Strongyloides stercoralis* and hookworm as diagnostic tool in asymptomatic schoolchildren in Cambodia. Acta Trop. 2013; 126: 89-92.

19. Cimino RO, Jeun R, Juarez M, Cajal PS, Vargas P, Echazú A, et al. Identification of human intestinal parasites affecting an asymptomatic peri-urban Argentinian population using multi-parallel quantitative real-time polymerase chain reaction. Parasit Vectors. 2015; 8: 380.

20. Tahapary DL, de Ruiter K, Martin I, van Lieshout L, Guigas B, Soewondo P, et al. Helminth infections and type 2 diabetes: a cluster-randomized placebo controlled SUGARSPIN trial in Nangapanda, Flores, Indonesia. BMC Infect Dis. 2015; 15: 133.

21. Staudacher O, Heimer J, Steiner F, Kayonga Y, Havugimana JM, Ignatius R, et al. Soil-transmitted helminths in southern highland Rwanda: associated factors and effectiveness of school-based preventive chemotherapy. Trop Med Int Health. 2014; 19: 812-824.

22. Gatehouse HS, Malone LA. The Ribosomal RNA Gene Region of *Nosema apis* (Microspora): DNA Sequence for Small and Large Subunit rRNA Genes and Evidence of a Large Tandem Repeat Unit Size. J Invertebr Pathol. 1998; 71: 97-105.

23. Hlinka O, Murrell A, Barker SC. Evolution of the secondary structure of the rRNA internal transcriber spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). Heredity. 2002; 88: 275-279.

24. Gerbi SA. Evolution of ribosomal DNA. In: Molecular Evolutionary Genetics. New York/London: Plenum Press; 1985. pp. 419-517.

25. Fritz GN, Conn J, Cockburn A, Seawright J. Sequence Analysis of the Ribosomal DNA Internal Transcribed Spacer 2 from Populations of *Anopheles nuneztovari* (Diptera: Culicidae). Mol Biol Evol. 1994; 11: 406-416.

26. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 5th ed. New York: Garland Science; 2008.

27. Cooper GM. The Cell: A Molecular Approach. 2nd Edition. Sunderland (MA): Sinauer Associates; 2000.

28. The *C. elegans* Sequencing Consortium. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. Science. 1998; 282: 2012-2018.

29. Thanchomnang T, Intapan PM, Lulitanond V, Choochote W, Manjai A, Prasongdee TK, et al. Rapid detection of *Brugia malayi* in mosquito vectors using a real-time fluorescence resonance energy transfer PCR and melting curve analysis. Am J Trop Med Hyg. 2008; 78: 509-513.

30. Alhassan A, Li Z, Poole CB, Carlow CK. Expanding the MDx toolbox for filarial diagnosis and surveillance. Trends Parasitol. 2015; 31: 391-400.

31. Ricciardi A, Ndao M. Diagnosis of Parasitic Infections: What's Going On? J Biomol Screen. 2015; 20: 6-21.

32. McReynolds LA, DeSimone SM, Williams SA. Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi*. Proc Natl Acad Sci USA. 1986; 83: 797-801.

33. Zhong M, McCarthy J, Bierwert L, Lizotte-Waniewski M, Chanteau S, Nutman TB, et al. A polymerase chain reaction assay for detection of the parasite *Wuchereria bancrofti* in human blood samples. Am J Trop Med Hyg. 1996; 54: 357-363.

34. Vargas N, Souto RP, Carranza JC, Vallejo GA, Zingales B. Amplification of a specific repetitive DNA sequence for *Trypanosoma rangeli* identification and its potential application in epidemiological investigations. Exp Parasitol. 2000; 96: 147-159.

35. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011; 13: 36-46.

36. Subirana JA, Messeguer X. A Satellite Explosion in the Genome of Holocentric Nematodes. PLoS One. 2013; 8: e62221.

37. Roset R, Subirana JA, Messeguer X. MREPATT: detection and analysis of exact consecutive repeats in genomic sequences. Bioinformatics. 2003; 19: 2475-2476.

38. Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010; 11: 378.

39. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792-793.

40. Caruccio N. Preparation of next-generation sequencing libraries using Nextera technology: simultaneous DNA fragmentation and adapter tagging by in vitro transposition. Methods Mol Biol. 2011; 733: 241-55.

41. Nery SV, McCarthy JS, Traub R, Andrews RM, Black J, Gray D, et al. A cluster-randomized controlled trial integrating a community-based water, sanitation and hygiene programme with mass distribution of albendazole to reduce intestinal parasites in Timor-Leste: the WASH for WORMS research protocol. BMJ Open. 2015; 5: e009293.

42. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a multiplex quantitative PCR to assess prevalence and intensity of intestinal parasite infections in a controlled clinical trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

43. U.S. Department of Health and Human Services Food and Drug Administration. Statistical guidance on reporting results from studies evaluating diagnostic tests. 2007. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.

44. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia. 2015. Available: https://www.r-project.org.

45. Gamer M, Lemon J, Fellows I, Singh P. Package 'irr': Various Coefficients of Interrater Reliability and Agreement. 2012. Available: https://cran.r-project.org/web/packages/irr/irr.pdf.

46. Meekums H, Hawash MB, Sparks AM, Oviedo Y, Sandoval C, Chico ME, et al. A genetic analysis of *Trichuris trichiura* and *Trichuris suis* from Ecuador. Parasit Vectors. 2015; 8: 168.

47. George S, Kaliappan SP, Kattula D, Roy S, Geldhof P, Kang G, et al. Identification of *Ancylostoma ceylanicum* in Children from a tribal community in Tamil Nadu, India using a semi-nested PCR-RFLP tool. Trans R Soc Trop Med Hyg. 2015; 109: 283-285.

48. Masuda Y, Kishimoto T, Ito H, Tsuji M. Visceral larva migrans caused by *Trichuris vulpis* presenting as a pulmonary mass. Thorax. 1987; 42: 990-991.

49. Sakano T, Hamamoto K, Kobayashi Y, Sakata Y, Tsuji M, Usui T. Visceral larva migrans caused by *Trichuris vulpis*. Arch Dis Child. 1980; 55: 631-633.

50. Areekul P, Putaporntip C, Pattanawong U, Sitthicharoenchai P, Jongwutiwes S. *Trichuris vulpis* and *T. trichiura* infections among schoolchildren of a rural community in northwestern Thailand: the possible role of dogs in disease transmission. Asian Biomed. 2010; 4: 49-60.

51. Hendrix CM, Robinson E. Diagnostic Parasitology for Veterinary Technicians. 4th ed. Philadelphia: Elsevier Health Sciences; 2014.

52. Berger S. Infectious Diseases of East Timor. Los Angeles: GIDEON Informatics, Inc; 2015.

53. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. Nat Genet. 2016; 48: 299-307.

54. Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, et al. Genome of the human hookworm *Necator americanus*. Nat Genet. 2014; 46: 261-269.

55. Cuomo MJ, Noel LB, White DB. Diagnosing Medical Parasites: A Public Health Officers Guide to Assisting Laboratory and Medical Officers. USAF Air Education and Training Command, Randolph, TX: http://www.phsource.us/PH/PARA/; 2012.

56. Meštrović N, Pavlek M, Car A, Castagnone-Sereno P, Abad P, Plohl M. Conserved DNA Motifs, Including the CENP-B Box-like, are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. PLoS One. 2013; 8: e67328.

57. Plohl M, Luchetti A, Mestrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. 2008; 409: 72-82.

58. da Cruz CJ. Livestock development in East Timor. In: Agriculture: New Directions for a New Nation – East Timor (Timor Leste). ACIAR Proceedings No. 113; 2003. pp. 11-16.

59. Tobler H, Etter A, Müller F. Chromatin diminution in nematode development. Trends Genet. 1992; 8: 427-432.

60. Goday C, Pimpinelli S. The occurrence, role and evolution of chromatin diminution in nematodes. Parasitol Today. 1993; 9: 319-322.

61. Etter A, Bernard V, Kenzelmann M, Tobler H, Müller F. Ribosomal heterogeneity from chromatin diminution in *Ascaris lumbricoides*. Science. 1994; 265: 954-956.

62. Pilotte N, Torres M, Tomaino FR, Laney SJ, Williams SA. A TaqMan-based multiplex real-time PCR assay for the simultaneous detection of *Wuchereria bancrofti* and *Brugia malayi*. Mol Biochem Parasitol. 2013; 189: 33-37.

63. Tran AC, Halse TA, Escuyer VE, Musser KA. Detection of *Mycobacterium avium* complex DNA directly in clinical respiratory specimens: opportunities for improved turn-around time and cost savings. Diagn Microbiol Infect Dis. 2014; 79: 43-48.

64. Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenja SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

65. Schmaedick MA, Koppel AL, Pilotte N, Torres M, Williams SA, Dobson SL, et al. Molecular xenomonitoring using mosquitoes to map lymphatic filariasis after mass drug administration in American Samoa. PLoS Negl Trop Dis. 2014; 8: e3087.

66. Rao RU, Nagodavithana KC, Samarasekera SD, Wijegunawardana AD, Premakumara WD, Perera SN, et al. A comprehensive assessment of lymphatic filariasis in Sri Lanka six years after cessation of mass drug administration. PLoS Negl Trop Dis. 2014; 8: e3281.

67. Lamberton PH, Cheke RA, Winskill P, Tirados I, Walker M, Osei-Atweneboana MY, et al. Onchocerciasis transmission in Ghana: persistence under different control strategies and the roll of the simulidd vectors. PLoS Negl Trop Dis. 2015; 9: e0003688.

68. Rodríguez-Pérez MA, Lilley BG, Domínguez-Vázquez A, Segura-Arenas R, Lizarazo-Ortega C, Mendoza-Herrera A, et al. Polymerase chain reaction monitoring of transmission of *Onchocerca volvulus* of two endemic states in Mexico. Am J Trop Med Hyg. 2004; 70: 38-45.

69. Rodríguez-Pérez MA, Gopal H, Adeleke MA, De Luna-Santillana EJ, Gurrola-Reyes JN, Guo X. Detection of *Onchocerca volvulus* in Latin American black flies for pool screening PCR using high-throughput automated DNA isolation for transmission surveillance. Parasitol Res. 2013; 112: 3925-3931.

70. Fung MS, Xiao N, Wang S, Carlton EJ. Field evaluation of a PCR test for *Schistosoma japonicum* egg detection in low-prevalence regions of China. Am J Trop Med Hyg. 2012; 87: 1053-1058.

71. Kure A, Mekonnen Z, Dana D, Bajiro M, Ayana M, Vercruysse J, et al. Comparison of individual and pooled stool samples for the assessment of intensity of *Schistosoma mansoni* and soil-transmitted helminth infections using the Kato-Katz technique. Parasit Vectors. 2015; 8: 489.

72. Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, et al. Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted helminth diagnosis in epidemiological surveys. Parasit Vectors. 2010; 3: 71.

73. Markoulatos P, Siafakas N, Moncany M. Multiplex Polymerase Chain Reaction: A Practical Approach. J Clin Lab Anal. 2002; 16: 47-51.

# CHAPTER 3

# A NOVEL, SPECIES-SPECIFIC, REAL-TIME PCR ASSAY FOR THE DETECTION OF THE EMERGING ZOONOTIC PARASITE *ANCYLOSTOMA CEYLANICUM* IN HUMAN STOOL

*as written by Papaiakovou M.[#], Pilotte N.[#], Grant J.R., Traub R.J., Llewellyn S, McCarthy J.S., Krolewiecki A.J., Cimino R., Mejia R., Williams S.A.*
*[#] Equal contribution*
*PLoS Neglected Tropical Diseases, 2017*

## 3.1 General Chapter Overview

This chapter describes the utilization of our previously described RepeatExplorer-based target-discovery pipeline for the development of a novel qPCR assay enabling the species-specific detection of the zoonotic intestinal helminth *Ancylostoma ceylanicum*. This work represents the first description of a qPCR assay capable of discriminating between *A. ceylanicum* and *Ancylostoma duodenale*, the related, obligate human pathogen. Following development and optimization, this index assay is validated on samples originating from both Timor Leste and Argentina through comparisons with both a ribosomal sequence-targeting, genus-specific *Ancylostoma* spp. assay and a PCR-Restriction Fragment Length Polymorphism (RFLP) assay. These comparisons demonstrate the ability of the index assay to discriminate to the species level, highlighting the misidentification dangers associated with the use of less specific ribosomal sequence-targeting assays and providing a discussion of the reasons why species-level diagnosis is important.

My contributions to this work include performing all bioinformatics analyses, designing the assay and either directly performing or overseeing the performance of all optimization and validation assays, as well as all comparative testing. I also co-wrote the manuscript.

## 3.2 Abstract


### 3.2.1 Background

Molecular-based surveys have indicated that *Ancylostoma ceylanicum*, a zoonotic hookworm, is likely the second most prevalent hookworm species infecting humans in Asia. Most current PCR-based diagnostic options for the detection of *Ancylostoma* species target the Internal Transcribed Spacer (ITS) regions of the ribosomal gene cluster. These regions possess a considerable degree of conservation among the species of this genus and this conservation can lead to the misidentification of infecting species or require additional labor for accurate species-level determination. We have developed a novel, real-time PCR-based assay for the sensitive and species-specific detection of *A. ceylanicum* that targets a non-coding, highly repetitive genomic DNA element. Comparative testing of this PCR assay with an assay that targets ITS sequences was conducted on field-collected samples from Argentina and Timor-Leste to provide further evidence of the sensitivity and species-specificity of this assay.

### 3.2.2 Methods/principal findings

A previously described platform for the design of primers/probe targeting non-coding highly repetitive regions was used for the development of this novel assay. The assay's limits of detection (sensitivity) and cross-reactivity with other soil-transmitted helminth species (specificity) were assessed with real-time PCR experiments. The assay was successfully used to identify infections caused by *A. ceylanicum* that were previously only identified to the genus level as *Ancylostoma* spp. when analyzed using other published

primer-probe pairings. Further proof of sensitive, species-specific detection was provided using a published, semi-nested restriction fragment length polymorphism-PCR assay that differentiates between *Ancylostoma* species.

### 3.2.3 Conclusions/significance

Due to the close proximity of people and domestic/wild animals in many regions of the world, the potential for zoonotic infections is substantial. Sensitive tools enabling the screening for different soil-transmitted helminth infections are essential to the success of mass deworming efforts and facilitate the appropriate interpretation of data. This study describes a novel, species-specific, real-time PCR-based assay for the detection of *A. ceylanicum* that will help to address the need for such tools in integrated STH deworming programs.

## 3.3. Author Summary

Historically, *Ancylostoma ceylanicum* has been viewed as an uncommon cause of human hookworm infection, with minimal public health importance. However, recent reports have indicated that this zoonotic hookworm causes a much greater incidence of infection within certain human populations than was previously believed. Current methods for the species-level detection of *A. ceylanicum* rely on techniques that involve conventional PCR accompanied by restriction enzyme digestions. These PCR-based assays are not only laborious but they lack sensitivity as they target suboptimal regions on the DNA. As efforts aimed at the eradication of hookworm disease have grown substantially over the last decade, the need for sensitive and specific tools to monitor and evaluate

programmatic successes has correspondingly escalated. Since a growing body of evidence suggests that patient responses to drug treatment can vary based upon the species of hookworm that is causing infection, accurate species-level diagnostics are advantageous. Accordingly, the novel real-time PCR-based assay described here provides a sensitive, species-specific diagnostic tool that will facilitate the accurate mapping of disease endemicity and will aid in the evaluation of progress of programmatic deworming efforts.

### 3.4 Introduction

*Ancylostoma ceylanicum* is the only predominantly animal-infecting hookworm species known to successfully develop into adults within the human intestine [1]. This zoonotic species has been reported to infect animals and humans in Southeast Asia, South America, South Africa, Melanesia and Australia [2–13], and prevalence studies have demonstrated *A. ceylanicum* to be the second most common hookworm species causing human infection in many parts of Asia [5, 14-16]. Equally troubling, a recent study investigating *A. ceylanicum* transmission conducted in Cambodia demonstrated the ability of this worm to result in burdens of infection comparable to levels caused by *Necator americanus* [6]. Furthermore, it has been demonstrated that *Ancylostoma duodenale* can ingest eight times the volume of blood consumed by *N. americanus*, resulting in increased iron depletion [17]. Accordingly, frequent cases of *Ancylostoma*-associated anemia have been reported, even among properly nourished individuals [12, 18]. Although only speculative, similar pathology might exist for human infections with *A. ceylanicum*, underscoring the potential importance of this zoonotic parasite for both child and maternal health [19]. Despite these findings, methodologies for the reliable, species-specific, and

sensitive molecular detection of *A. ceylanicum* are lacking, resulting in a need for improved diagnostic tools.

Historically, diagnosis and detection of all hookworm species has relied heavily on the use of coprological methods [5]. Widespread use of these methods stems from their low cost, simplicity, and minimal infrastructure requirements. By default, microscopy has served as the standard method for gastrointestinal helminth detection [5]. However, at the egg level, and in some cases even at the larval level, certain hookworm species are morphologically indistinguishable [20] and further culturing of worms is required for differentiation. Conducting such coprological techniques is time-consuming and requires highly skilled microscopists for accurate assessment of results [21]. Accordingly, PCR-based molecular diagnostic tests for hookworm provide an attractive alternative to microscopy with more modest training requirements. However, such assays frequently target the internal transcribed spacer regions (ITS) of the rDNA gene clusters, which demonstrate considerable homology between closely related *Ancylostoma* species [5]. This similarity can lead to species misidentification in hookworm positive samples. Several PCR variations such as nested PCR and PCR coupled with restriction fragment length polymorphism (PCR-RFLP) have been developed to differentiate between the hookworm species. Unfortunately, these methods tend to be laborious, requiring many steps and/or restriction enzyme digestions coupled with agarose gel electrophoresis [4, 21, 23, 24]. In addition, nested PCR is far more likely to result in false positivity due to sample contamination and increased numbers of PCR cycles [22]. Thus, simplified PCR assays, not reliant on nested PCR or restriction enzyme digestions, would be of great benefit to programmatic efforts. Furthermore, while drug intervention and hygiene programs have

made substantial gains and continue to show great promise, they frequently fail to account for the control of zoonotic infections [15]. Such concerns are particularly important when proper sanitation is lacking as the potential for zoonotic transmission from wild or domestic animals to humans is dramatically increased [5]. Accordingly, species-level identification provides insight into possible infection reservoirs, which might not be considered utilizing genus-level diagnostics.

Herein we report the development and validation of a novel, real-time PCR-based diagnostic assay for the species-specific detection of *A. ceylanicum*. The development of this test is based on the identification of a highly repetitive, non-coding DNA element using advanced bioinformatics to analyze whole genome sequence data for target discovery and assay design [25]. We posit that the incorporation of this new assay into our previously established diagnostic platform [25] will aid programmatic efforts, as accurate species-level detection and differentiation of *Ancylostoma* ssp. is now possible.

## 3.5 Materials and Methods

### 3.5.1 Assay design

Genome sequences for *A. ceylanicum* were obtained from NCBI (Sequence Read Archive ID: SRR2037027 and SRR2037046). Analysis of repeat DNA content was performed using the publically available platform, RepeatExplorer [26], following our previously established bioinformatics workflow [25], and a primer-probe set was designed using the PrimerQuest online tool (Integrated DNA Technologies, Coralville, IA) (Table 3.1). This candidate primer-probe pairing then underwent further bioinformatics analysis

using the Primer-BLAST tool available from the National Center for Biotechnology Information (NCBI) website. Utilizing default parameters, the candidate primer pair did not return any matches to off target templates within NCBI's Nucleotide Collection database, indicating target specificity and ensuring that the occurrence of off target amplification would be extremely unlikely. Default parameters for probe-based PCR were utilized for the assay design and the probe was labeled with 6FAM fluorophore at the 5' end and double-quenched with ZEN (internal) and 3IABkFQ at the 3' end.

### 3.5.2 Assay optimization

A concentration matrix was created to determine the optimal forward and reverse primer concentration as previously described [25]. All reactions were performed in 7 µl total volumes utilizing 2 µl of template, 3.5 µl of TaqMan Fast Universal PCR Master Mix (ThermoFisher Scientific, Waltham, MA), and the appropriate concentrations of primers and probe. Cycling conditions consisted of an initial 2-minute incubation step at 50 °C, followed by a 10-minute incubation at 95 °C. These incubations were followed by 40 cycles of 95 °C for 15 seconds for denaturation and 1 minute at 59 °C for primer annealing and polymerase extension. All reactions were conducted using the StepOnePlus Real-Time PCR System (Life Technologies, Carlsbad, CA).

### 3.5.3 Sensitivity and specificity testing

Following the completion of primer optimization reactions, the sensitivity and specificity of our assay was evaluated. For sensitivity testing, the optimized assay was performed utilizing a titration of genomic DNA concentrations as template. Reactions were

conducted using 2 µl of *A. ceylanicum* genomic template at concentrations of 1 ng/µl, 100 pg/µl, 10 pg/µl, 1 pg/µl, 100 fg/µl, 10 fg/µl, 1 fg/µl, 100 ag/µl 10 ag/µl, and 1 ag/µl. In order to verify assay specificity, optimized conditions were employed to evaluate the potential for non-specific amplification of control genomic DNA from six other species of soil-transmitted helminths (*A. duodenale*, *N. americanus*, *Trichuris trichiura*, *Strongyloides stercoralis*, *Ascaris lumbricoides and Ancylostoma caninum*), along with human genomic DNA and DNA of the common gastrointestinal tract commensal bacterium *Escherichia coli* (strain K-12). All of these control samples were tested utilizing 2 µl of DNA template at concentrations of 1 ng/µl.

### 3.5.4 Mixed infection testing

A panel of six 10-fold serial dilutions of an initial 1 ng/µl stock of *A. ceylanicum* genomic DNA was created. To each dilution in this panel was added genomic DNA from both *A. caninum* and *A. duodenale* such that concentrations of DNA from both of these other species was at a final concentration of 1 ng/µl in each panel dilution. The aforementioned panel was tested with the optimized real-time PCR based assay for *A. ceylanicum* to prove that the presence of other *Ancylostoma* spp., simulating mixed infections, would not affect our assay's specificity. Furthermore, testing of this panel enabled us to determine the assay's limits of detection when increasingly limited concentrations of target DNA (*A. ceylanicum*) were intermixed with genomic DNA from the two other *Ancylostoma* spp. (Table 3.2).

### 3.5.5 Selection of field-samples for inclusion in this study

Sixty-one human stool samples previously collected as part of the "Wash for Worms" intervention trial in Timor-Leste (Trial registration: ACTRN12614000680662) were selected for inclusion in this study. DNA extractions for these samples were previously performed at QIMR Berghofer, and real-time PCR analysis demonstrated the presence of *Ancylostoma* spp. DNA in 22 of these samples [27]. Eight additional samples, collected in Orán, Argentina, as part of a larger collection effort (IRB00008019), were also selected for inclusion in this study. For Argentinian samples, DNA was previously extracted at Baylor College of Medicine utilizing a published methodology [28]. Prior real-time PCR analysis of these eight samples had demonstrated the presence of *Ancylostoma* spp. DNA [28].

### 3.5.6 PCR-RFLP analysis

A previously published semi-nested conventional PCR assay coupled with restriction fragment length polymorphism analysis [23] was employed to distinguish between the different hookworm species in the samples obtained from both Timor-Leste and Orán, Argentina. All amplification reactions and digestions were conducted in accordance with the previously described methodologies [23].

### 3.5.7 Real-time PCR analysis

DNA aliquots of all samples from both Timor-Leste and Orán, Argentina were sent to Smith College for further analysis. Samples were blindly coded and assayed for the presence of both *A. duodenale* and *A. ceylanicum.* Testing for the presence of *A. duodenale*

occurred in accordance with the previously described protocol [25], while *A. ceylanicum* testing was conducted using the novel assay described here.

### 3.6 Results

Utilization of the previously published workflow for repeat analysis led to the identification of a novel target for the sensitive and species-specific real-time PCR-based detection of *A. ceylanicum* [25]. Employing the PrimerQuest online tool, a primer/probe pairing was identified and oligonucleotides were synthesized. Through the analysis of a titration matrix, primer optimization reactions were performed, and the optimal forward primer concentration was determined to be 125 nM, while the reverse primer was demonstrated to have an optimal concentration of 1000 nM. This combination of concentrations resulted in the lowest Ct values when amplifying 2 µl of *A. ceylanicum* genomic DNA at a concentration of 1 ng/µl.

Assay specificity testing failed to amplify purified genomic DNA from the STH parasites *A. duodenale*, *N. americanus*, *T. trichiura*, *S. stercoralis*, *A. lumbricoides* and *A. caninum*. Testing also failed to amplify human DNA or DNA from the common gastrointestinal bacteria, *E. coli* (strain K-12), thus demonstrating the assay's species-specific detection properties. Furthermore, assay sensitivity testing demonstrated consistent detection of purified *A. ceylanicum* genomic DNA template at all quantities above 200 ag. However, when the *A. ceylanicum* assay was utilized to analyze samples containing simulated mixed infections of both *A. ceylanicum* and other *Ancylostoma* species, the limit of detection was determined to be 13.3 fg of *A. ceylanicum* DNA (Table 3.2). Importantly, based upon the genome size of *A. ceylanicum* [29], and assuming that a

single egg contains approximately 8 cells, this quantity of template DNA is less than the quantity expected to be found within a single *A. ceylanicum* egg (5520 fg of DNA theoretically based on *A. ceylanicum*'s genome size).

To further validate this novel assay, 61 DNA extracts from stool samples collected in Timor-Leste were analyzed. Utilizing a genus-specific real-time PCR assay, previously performed testing conducted at QIMR Berghofer had demonstrated the presence of *Ancylostoma* spp. DNA in 22 of these samples [27]. However, follow-up testing of these samples utilizing our previously described *A. duodenale*-specific primer/probe set excluded the presence of *A. duodenale* [25], and semi-nested PCR-RFLP analysis demonstrated the presence of *A. ceylanicum* in 21 of the 22 samples analyzed (Figure 3.1) [25].The identity of the 22$^{nd}$ sample could not be determined as this sample failed to allow for amplification using either our *A. duodenale*-specific assay or the semi-nested PRC-RFLP analysis. Two independent sequencing trials were also performed, but meaningful results could not be obtained after which the sample stock had been exhausted. This subset of *Ancylostoma* spp.-positive samples was then employed to demonstrate the comparative sensitivity and species-specificity of our newly designed *A. ceylanicum* assay. Utilizing this assay, the same 21 samples which were determined to be *A. ceylanicum*-positive by RFLP-PCR again tested positive for the presence of *A. ceylanicum* DNA, while the 22$^{nd}$ sample (PCR-RFLP-negative for *A. ceylanicum*) was also negative by real-time PCR, likely indicating the amplification of a non-*duodenale*, non-*ceylanicum* target utilizing the *Ancylostoma* genus-specific assay (Table 3.3). Equally important, the remaining 39 Timor-Leste-derived samples that tested negative for *Ancylostoma* spp. using the previously described genus-specific PCR assay [27] also produced a negative result when tested using

the newly described *A. ceylanicum*-specific diagnostic (Table 3.3). Thus, results agreed across all samples, in all replicates, with all utilized assays.

While testing of field-collected samples from Timor-Leste demonstrated the ability of our new assay to detect DNA from *A. ceylanicum*, we next sought to demonstrate that this *A. ceylanicum* real-time PCR assay would not amplify DNA from *A. duodenale*-containing field samples. For this purpose, eight additional DNA extracts from stool samples, obtained from an *A. duodenale* endemic region near Orán, Argentina, were tested in duplicate. Utilizing a combination of the previously described *A. duodenale*-specific PCR [25] and the semi-nested PCR-RFLP assays [23], we verified the presence of *A. duodenale* and the absence of *A. ceylanicum* within each sample (Figure 3.2, Table 3.4). Individual samples were then tested using the newly described *A. ceylanicum*-specific PCR assay, and for all samples, in all replicates, amplification failed to occur. Thus, by selectively amplifying *A. ceylanicum*-containing samples from Timor-Leste, but failing to amplify Argentinian samples positive for *A. duodenale*, we successfully demonstrated the species-specific nature of the *A. ceylanicum* PCR assay on human samples collected in the field. Table 3.5 summarizes the results of the various molecular approaches used to determine the species of *Ancylostoma* within samples from both field studies.

### 3.7 Discussion

Highlighting the under-recognized importance of *A. ceylanicum* as a public health concern, a study in northern Cambodia recently implicated *A. ceylanicum* as the causative agent responsible for as many as half of all human hookworm infections among the young adult population [6]. However, despite such prevalence, zoonotic parasites are typically

given little consideration during the implementation of mass deworming efforts. Accordingly, increased awareness of the potential impact of *A. ceylanicum* infection on the human population is required, and more sensitive tools are needed for monitoring this important zoonotic STH. Such tools will provide a more complete picture of the role such zoonotic parasites play in global health.

While current treatments for human hookworm infection are not species-dependent, it has been demonstrated that responses to drug treatments can vary in a species-dependent manner [31]. Because species-specific diagnostics are lacking, misidentification of the infecting species may lead to a lack of clarity in understanding drug-responses in infected populations. While benzimidazole-based interventions have not yet resulted in the reporting of any such species-specific responses, such responses may develop, or they may simply be underreported. As *A. duodenale* and *A. ceylanicum* are both capable of causing significant human infection, and *A. caninum* has the capacity to infect humans in an "unsuccessful" manner [23, 31], species-specific diagnostic assays are needed to fully evaluate the progress of drug intervention programs. Furthermore, given the increasing concerns surrounding the possible development of drug resistance [32-34], species-level knowledge of infection is critical, as it is likely that different hookworm species may evolve resistance *via* different mechanisms and at different rates. The importance of species-level identification is further supported by the phenomenon of *refugia*, as the presence of an untreated animal reservoir may slow the evolutionary pressures on a pathogen, in turn slowing the development of drug resistance mechanisms [32, 34].

While the *A. ceylanicum*-specific assay described here has potential usefulness as a tool for monitoring veterinary infections, its specificity and performance in such settings

has not yet been evaluated. In order to evaluate such applications, additional testing and validation of this assay against common animal hookworm species such as *Ancylostoma braziliense*, *Ancylostoma tubaeforme*, and *Uncinaria stenocephala*, as well as other common veterinary intestinal helminths would be required. Similarly, employment of this assay for the purposes of environmental monitoring would require, at a minimum, testing against common free-living nematode species.

Given the importance of the accurate, species-level identification of STH infections, we believe that the real-time PCR-based assay for the detection of *A. ceylanicum* described here, in combination with our previously described real-time PCR assays, will aid future hookworm monitoring efforts. With the demonstrated capacity to detect DNA isolated from a single egg, this assay provides a sensitive and species-specific diagnostic tool capable of more fully informing program managers, enabling more appropriate decision-making and allowing for improved programmatic outcomes. Finally, it should be noted that this *A. ceylanicum* PCR assay is easily integrated into our current multi-parallel PCR system [25] to provide a more complete picture of the STH infection status in studied populations.

### 3.8 Acknowledgements

The authors would like to thank all of the individuals who collected samples in the field sites (Nicolás Caro, Paola Vargas, Silvana Pamela Cajal for the study in Orán, Argentina; and the WASH for Worms investigators including Dr. Susana Vaz Nery and her field team in Timor-Leste) as well as all those individuals who graciously donated control materials for testing the specificity of the newly developed assay: Dr. Peter Nejsum

(University of Copenhagen, Denmark) for the *T. trichiura* worms; Dr. James Lok (University of Pennsylvania, Philadelphia, PA, USA) for the *Strongyloides stercoralis* DNA; Dr. Bin Zhan and Dr. Peter Hotez (Baylor College of Medicine, Houston, TX, USA) for providing *N. americanus* and *A. duodenale*; Dr. Matthew Berriman and Nancy Holroyd (Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK) for their generous contribution of *Ascaris lumbricoides* DNA. The authors would also like to express their gratitude to Dr. Eric Ottesen and Dr. Patrick Lammie (Task Force for Global Health) for their encouragement and support.

**Table 3.1 Primer and probe sequences for *A. ceylanicum*.**

| Parasite | Forward | Reverse | Probe |
|---|---|---|---|
| *A. ceylanicum* | 5'-CAAATATTACTGTGCGCATTTAGC-3' | 5'-GCGAATATTTAGTGGGTTTACTGG-3' | 5'-/56-FAM CGGTGAAAG/ZEN/CTTTGCG TTATTGCGA/3IABkFQ/ -3' |

**Table 3.2 Results of real-time PCR testing for *A. ceylanicum* of samples spiked with DNA from mixed *Ancylostoma* spp. to replicate mixed infections.**

| Sample Identifier | Ancylostoma ceylanicum | Ancylostoma duodenale | Ancylostoma caninum | Final Volume | Real-Time PCR Assay * |
|---|---|---|---|---|---|
| A | 4 ng | 4 ng | 4 ng | 12 ul | 14 |
| B | 400 pg | 4 ng | 4 ng | 12 ul | 17 |
| C | 40 pg | 4 ng | 4 ng | 12 ul | 20 |
| D | 4 pg | 4 ng | 4 ng | 12 ul | 24 |
| E | 400 fg | 4 ng | 4 ng | 12 ul | 27 |
| F | 40 fg | 4 ng | 4 ng | 12 ul | 31 |
| G | 4 fg | 4 ng | 4 ng | 12 ul | not detected |
| H | - | 4 ng | 4 ng | 8 ul | not detected |
| I (+) | 2 ng | - | - | 7 ul | 12 |
| J (NTC) | - | - | - | 7 ul | non detected |

* Results are reported as mean Ct values.
# "-" Indicates an absence of DNA in the sample.

**Table 3.3 Results of real-time PCR testing of samples from Timor-Leste for *A. duodenale* and *A. ceylanicum*.**

| Sample Number | Real-Time PCR Assay | | |
|---|---|---|---|
| | QIMR Assay | Smith Assays | |
| | *Ancylostoma* spp. *,# | *Ancylostoma duodenale* *,# | *Ancylostoma ceylanicum* *,# |
| **3** | - | - | - |
| **4** | - | - | - |
| **5** | - | - | - |
| **6** | - | - | - |
| **7** | - | - | - |
| **8** | 21 | - | 18 |
| **9** | - | - | - |
| **10** | - | - | - |
| **11** | - | - | - |
| **12** | - | - | - |
| **13** | 22 | - | 19 |
| **14** | - | - | - |
| **15** | - | - | - |
| **16** | - | - | - |
| **17** | - | - | - |
| **18** | 35 | - | - |
| **19** | - | - | - |
| **20** | 22 | - | 18 |
| **25** | 23 | - | 22 |
| **26** | - | - | - |
| **27** | - | - | - |
| **28** | - | - | - |
| **29** | 22 | - | 20 |
| **30** | 25 | - | 24 |
| **31** | - | - | - |
| **32** | - | - | - |
| **33** | 22 | - | 19 |
| **34** | - | - | - |
| **35** | - | - | - |
| **36** | - | - | - |
| **37** | - | - | - |
| **38** | - | - | - |
| **39** | 23 | - | 20 |
| **40** | 21 | - | 18 |
| **41** | - | - | - |
| **42** | - | - | - |
| **43** | 25 | - | 22 |
| **44** | - | - | - |
| **45** | - | - | - |
| **46** | 20 | - | 18 |
| **47** | - | - | - |
| **48** | - | - | - |
| **49** | - | - | - |
| **50** | - | - | - |
| **51** | - | - | - |
| **52** | - | - | - |
| **53** | - | - | - |
| **54** | - | - | - |
| **55** | - | - | - |
| **56** | 23 | - | 19 |
| **57** | 26 | - | 22 |
| **58** | - | - | - |

| | | | |
|---|---|---|---|
| **59** | - | - | - |
| **60** | - | - | - |
| **61** | - | - | - |
| **62** | 26 | - | 23 |
| **63** | - | - | - |
| **64** | - | - | - |
| **65** | - | - | - |
| **66** | 22 | - | 19 |
| **67** | - | - | - |
| **68** | 21 | - | 19 |
| **70** | - | - | - |
| **71** | - | - | - |
| **72** | 21 | - | 19 |
| **73** | - | - | - |
| **74** | 20 | - | 15 |
| **75** | - | - | - |
| **76** | - | - | - |
| **77** | 27 | - | 23 |
| **79** | 21 | - | 20 |
| **80** | 23 | - | 22 |
| **81** | - | - | - |
| **82** | - | - | - |
| **83** | - | - | - |
| **3** | - | - | - |

[*] Results are reported as mean Ct values.
[#] "-" Indicates a negative qPCR result.

**Table 3.4 Results of real-time PCR testing of samples from Orán, Argentina for *A. duodenale* and *A. ceylanicum*.**

| Sample Number | Real-time PCR Assay | | |
|---|---|---|---|
| | BCM Assay | Smith Assays | |
| | *Ancylostoma* spp. [*, #] | *Ancylostoma duodenale* [*, #] | *Ancylostoma ceylanicum* [*, #] |
| **6** | 32 | 34 | - |
| **7** | 29 | 33 | - |
| **10** | 30 | 37 | - |
| **11** | 29 | 33 | - |
| **13** | - | 34 | - |
| **36** | 32 | 32 | - |
| **37** | 27 | 25 | - |
| **38** | 34 | 33 | - |

[*] Results are reported as mean Ct values.
[#] "-" Indicates a negative qPCR result.

**Table 3.5 Results of three separate molecular assays for *Ancylostoma* spp. differentiation.**

| Origin of Samples | Real-Time PCR for *A. duodenale* [25] | Real-Time PCR for *A. ceylanicum* | Semi-Nested PCR [23] |
|---|---|---|---|
| **Timor-Leste** | (-) | (+)* | Figure 3.1 |
| **Orán, Argentina** | (+) | (-) | Figure 3.2 |

*Twenty-one out of twenty-two samples were positive for *A. ceylanicum*. The identity of the species in the 22nd sample could not be determined utilizing any of the employed assays or Sanger sequencing.

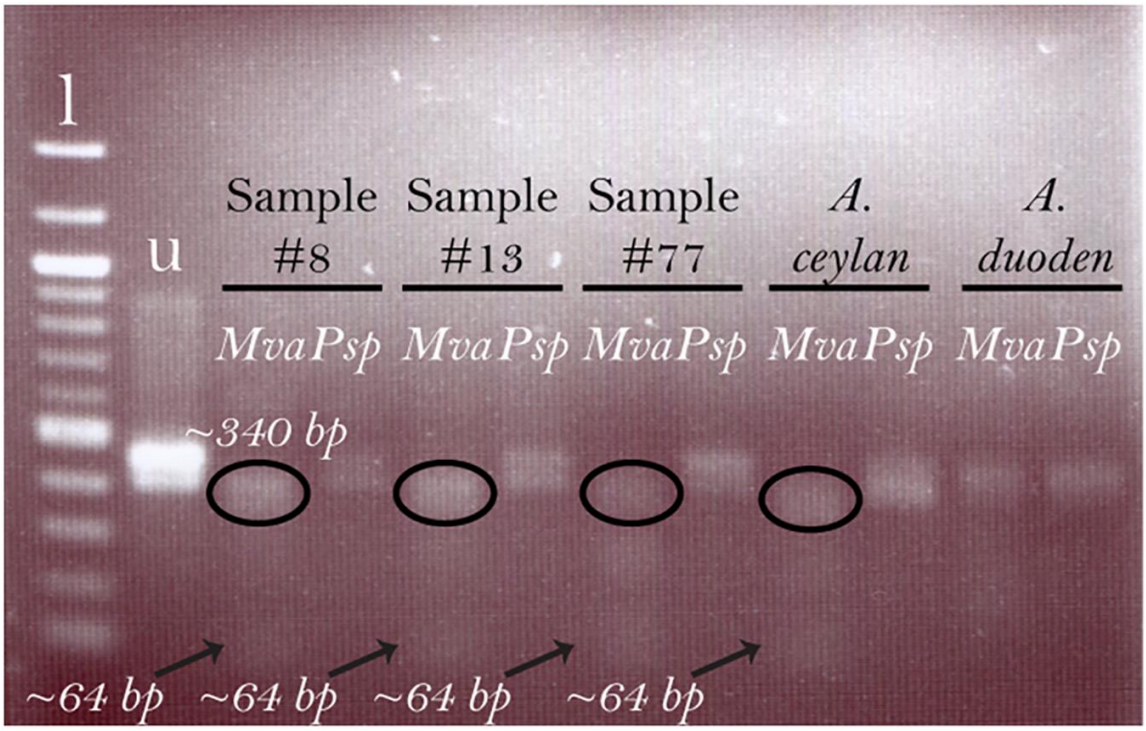


**Figure 3.1 *Ancylostoma* spp. determination using semi-nested PCR-RFLP analysis**. Following two rounds of conventional PCR targeting the ITS region, the product (~ 400 bp) was subjected to two separate restriction enzyme digestions (MvaI/BstN1 [#ER0551] and Psp1406I/AcII [#ER0942], ThermoFisher Scientific) following the manufacturer's suggested protocol. The MvaI enzyme digests PCR amplicons of *A. ceylanicum* into two products (bands at 340 bp and 64 bp) but does not digest amplicons of *A. canin*um or *A. duodenale*. Psp1406I digests *A. duodenale* amplicons into two products (bands at 255 bp and 149 bp) but does not digest amplicons of *A. caninum* or *A. ceylanicum*. In Figure 3.1, the uncut product (2nd round PCR product), the Mva product (2nd round PCR product digested with MvaI) and the Psp product (2nd round PCR product digested with Psp1 enzyme) for the positive control (*A. ceylanicum*) and for samples from Timor-Leste are shown. The banding pattern demonstrates the presence of *A. ceylanicum* in these samples and validates the positive results from the newly described real-time PCR assay for *A. ceylanicum* (see Additional File 1; l = 100 base pair ladder; u = undigested control).

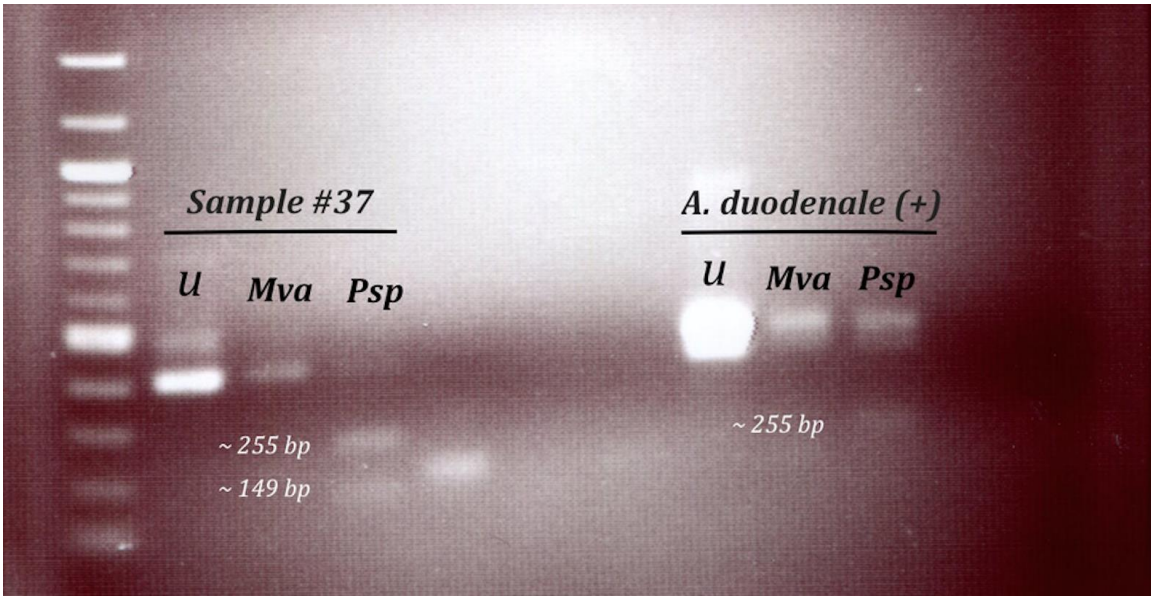


**Figure 3.2 *Ancylostoma* spp. determination on field samples from Orán, Argentina using semi-nested PCR-RFLP analysis**. Testing was performed in the same manner as described in Figure 3.1. The banding pattern from a representative sample demonstrates the presence of *A. duodenale* in the pictured sample and validates the negative results from the newly described real-time PCR assay for *A. ceylanicum*.

### 3.9 References


1. Ngui R, Lim YAL, Ismail WHW, Lim KN, Mahmud R. Zoonotic *Ancylostoma ceylanicum* infection detected by endoscopy. Am J Trop Med Hyg. 2014; 91: 86-88.

2. Bradbury RS, Hii SF, Harrington H, Speare R, Traub R. *Ancylostoma ceylanicum* hookworm in the Solomon Islands. Emerg Infect Dis. 2017; 23: 252-257.

3. Hu W, Wu S, Yu X, Abullahi AY, Song M, Tan L, et al. A multiplex PCR for simultaneous detection of three zoonotic parasites *Ancylostoma ceylanicum*, *Ancylostoma caninum*, and *Giardia lamblia* assemblage A. BioMed Res Int. 2015; 2015: 406168.

4. Traub RJ, Inpankaew T, Sutthikornchai C, Sukthana Y, Thompson RCA. PCR-based coprodiagnostic tools reveal dogs as reservoirs of zoonotic ancylostomiasis caused by *Ancylostoma ceylanicum* in temple communities in Bangkok. Vet Parasitol. 2008; 155: 67–73.

5. Ngui R, Lim YAL, Traub RJ, Mahmud R, Mistam MS. Epidemiological and genetic data supporting the transmission of *Ancylostoma ceylanicum* among human and domestic animals. PLoS Negl Trop Dis. 2012; 6: e1522.

6. Inpankaew T, Schär F, Dalsgaard A, Khieu V, Chimnoi W, Chhoun C, et al. High prevalence of *Ancylostoma ceylanicum* hookworm infections in humans, Cambodia, 2012. Emerg Infect Dis. 2014; 20: 976–982.

7. Conlan JV, Khamlome B, Vongxay K, Elliot A, Pallant L, Sripa B, et al. Soil-transmitted helminthiasis in Laos: a community-wide cross-sectional study of humans and dogs in a mass drug administration environment. Am J Trop Med Hyg. 2012; 86: 624–634.

8. Jiraanankul V, Aphijirawat W, Mungthin M, Khositnithikul R, Rangsin R, Traub RJ, et al. Incidence and risk factors of hookworm infection in a rural community of central Thailand. Am J Trop Med Hyg. 2011; 84: 594–598.

9. Baker MK, Lange L, Verster A, van der Plaat S. A survey of helminths in domestic cats in the Pretoria area of Transvaal, Republic of South Africa. Part 1: The prevalence and comparison of burdens of helminths in adult and juvenile cats. J S Afr Vet Assoc. 1989; 60: 139–142.

10. Rep BH, Heinemann DW. Changes in hookworm distribution in Surinam. Trop Geogr Med. 1976; 28: 104–110.

11. Schuster RK, Thomas K, Sivakumar S, O'Donovan D. The parasite fauna of stray domestic cats (*Felis catus*) in Dubai, United Arab Emirates. Parasitol Res. 2009; 105: 125–134.

12. Anten JF, Zuidema PJ. Hookworm infection in Dutch servicemen returning from West New Guinea. Trop Geogr Med. 1964; 16: 216–224.

13. Palmer CS, Traub RJ, Robertson ID, Hobbs RP, Elliot A, While L, et al. The veterinary and public health significance of hookworm in dogs and cats in Australia and the status of *A. ceylanicum*. Vet Parasitol. 2007; 145: 304–313.

14. Scholz T, Uhlírová M, Ditrich O. Helminth parasites of cats from the Vientiane province, Laos, as indicators of the occurrence of causative agents of human parasitoses. Parasite. 2003; 10: 343–350.

15. Traub RJ. *Ancylostoma ceylanicum*, a re-emerging but neglected parasitic zoonosis. Int J Parasitol. 2013; 43: 1009-1015.

16. Setasuban P, Vajrasthira S, Muennoo C. Prevalence and zoonotic potential of *Ancylostoma ceylanicum* in cats in Thailand. Southeast Asian J Trop Med Public Health. 1976; 7: 534–539.

17. Albonico M, Stoltzfus RJ, Savioli L, Tielsch JM, Chwaya HM. Epidemiological evidence for a different effect of hookworm species, *Ancylostoma duodenale* or *Necator americanus*, on iron status of children. Int J Epidemiol. 1998; 27: 530-537.

18. Jonker FAM, Calis JCJ, Phiri K, Brienen EAT, Khoffi H, Brabin BJ, et al. Real-time PCR demonstrates *Ancylostoma duodenale* is a key factor in the etiology of severe anemia and iron deficiency in Malawian pre-school children. PLoS Negl Trop Dis. 2012; 6: e1555.

19. Crompton DW. The public health importance of hookworm disease. Parasitology. 2000; 121 Suppl: S39-S50.

20. Lucio-Forster A, Liotta JL, Yaros JP, Briggs KR, Mohammed HO, Bowman DD. Morphological differentiation of eggs of *Ancylostoma caninum, Ancylostoma tubaeforme*, and *Ancylostoma braziliense* from dogs and cats in the United States. J Parasitol. 2012; 98: 1041–1044.

21. de Gruijter JM, van Lieshout L, Gasser RB, Verweij JJ, Brienen EAT, Ziem JB, et al. Polymerase chain reaction-based differential diagnosis of *Ancylostoma duodenale* and *Necator americanus* infections in humans in northern Ghana. Trop Med Int Health. 2005;10: 574–580.

22. Yamamoto Y. PCR in diagnosis of Infection: detection of bacteria in cerebrospinal fluids. Clin Diagn Lab Immunol. 2002; 9: 508–514.

23. George S, Kaliappan SP, Kattula D, Roy S, Geldhof P, Kang G, et al. Identification of *Ancylostoma ceylanicum* in children from a tribal community in Tamil Nadu, India using a semi-nested PCR-RFLP tool. Trans R Soc Trop Med Hyg. 2015; 109: 283–285.

24. Traub RJ, Robertson ID, Irwin P, Mencke N, Thompson RCA. Application of a species-specific PCR-RFLP to identify *Ancylostoma* eggs directly from canine faeces. Vet Parasitol. 2004; 123: 245–255.

25. Pilotte N, Papaiakovou M, Grant JR, Bierwert LA, Llewellyn S, McCarthy JS, et al. Improved PCR-based detection of soil transmitted helminth infections using a next-generation sequencing approach to assay design. PLoS Negl Trop Dis. 2016; 10: e0004578.

26. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792–793.

27. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a multiplex quantitative PCR to assess prevalence and intensity of intestinal parasite infections in a controlled clinical trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

28. Cimino RO, Jeun R, Juarez M, Cajal PS, Vargas P, Echazú A, et al. Identification of human intestinal parasites affecting an asymptomatic peri-urban Argentinian population using multi-parallel quantitative real-time polymerase chain reaction. Parasit Vectors. 2015; 8: 380.

29. Schwarz EM, Hu Y, Antoshechkin I, Miller MM, Sternberg PW, Aroian RV. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. Nat genet. 2015; 47: 416-422.

30. Reynoldson JA, Behnke, JM, Pallant LJ, Macnish MG, Gilbert F, Giles S, et al. Failure of pyrantel in treatment of human hookworm infections (*Ancylostoma duodenale*) in the Kimberley region of north west Australia. Acta Trop. 1997; 68: 301-312.

31. Landmann JK, Prociv P. Experimental human infection with the dog hookworm, *Ancylostoma caninum*. Med J Aust. 2003; 178: 69–71.

32. Vercruysse J, Albonico M, Behnke JM, Kotze AC, Prichard RK, McCarthy JS, et al. Is anthelminthic resistance a concern for the control of human soil-transmitted helminths? Int J Parasitol Drugs Drug Resist. 2011; 1: 14-27.

33. Vercruysse J, Levecke B, Prichard R. Human soil-transmitted helminths: implications of mass drug administration. Curr Opin Infect Dis. 2012; 25: 703-708.

34. Shalaby HA. Anthelmintics Resistance; How to Overcome it? Iran J Parasitol. 2013; 8: 18-32.

# CHAPTER 4

# TARGETING A HIGHLY REPEATED GERMLINE DNA SEQUENCE FOR IMPOROVED REAL-TIME PCR-BASED DETECTION OF *ASCARIS* INFECTION IN HUMAN STOOL

*as written by Pilotte N., Maasch J.R.M.A., Easton A.V., Dahlstrom E., Nutman T.B., Williams S.A.*
*PLoS Neglected Tropical Diseases, 2019*

## 4.1 General Chapter Overview

This chapter describes the development of an assay for the detection of human-infecting *Ascaris* spp. This assay is intended to improve upon the ribosomal sequence-targeting assay described in our original paper (Pilotte, et al., 2016). Improvement is made possible by analysis of genomic DNA isolated from germline, rather than adult, *Ascaris* tissue. Because members of the order Ascaridida are capable of a phenomenon known as chromosome diminution, whereby large sections of their repetitive gDNA are predictably and repeatibly eliminated from somatic cells during embryonic development, prior analysis of gDNA isolated from adult worms resulted in the identification of rDNA as the most prevalent genomic element. This occurred because the post-diminution adult tissues no longer contained the majority of these repeat sequences. However, as diagnosis of soil-transmitted helminths relies on identification of material within shed stool, identification of pathogen is dependent primarily upon the presence of pre-diminution eggs within a sample. For this reason, an index assay targeting a satellite repeat that is largely eliminated during diminution is described, and this assay's performance is comparatively evaluated

against the previously used ribosomal target. Analysis demonstrates the vastly improved analytical and clinical sensitivities of the new assay, and highlight the importance of methodical and systematic selection of assay targets.

My contributions to this chapter include performing all informatics analyses, designing primers and probe, and either directly performing, or overseeing the performance of all optimization and validation testing. I also wrote the manuscript.

## 4.2 Abstract

### 4.2.1 Background

With the expansion of soil transmitted helminth (STH) intervention efforts and the corresponding decline in infection prevalence, there is an increased need for sensitive and specific STH diagnostic assays. Previously, through next generation sequencing (NGS)-based identification and targeting of non-coding, high copy-number repetitive DNA sequences, we described the development of a panel of improved quantitative real-time PCR (qPCR)-based assays for the detection of *Necator americanus*, *Ancylostoma duodenale*, *Ancylostoma ceylanicum*, *Trichuris trichiura*, and *Strongyloides stercoralis*. However, due to the phenomenon of chromosome diminution, a similar assay based on high copy-number repetitive DNA was not developed for the detection of *Ascaris lumbricoides*. Recently, the publication of a reference-level germline genome sequence for *A. lumbricoides* has facilitated our development of an improved assay for this human pathogen of vast global importance.

### 4.2.2 Methodology/principal findings

Repurposing raw DNA sequence reads from a previously published Illumina-generated, NGS-based *A. lumbricoides* germline genome sequencing project, we performed a cluster-based repeat analysis utilizing RepeatExplorer2 software. This analysis identified the most prevalent repetitive DNA element of the *A. lumbricoides* germline genome (AGR, *Ascaris* germline repeat), which was then used to develop an improved qPCR assay. During experimental validation, this assay demonstrated a fold

increase in sensitivity of ~3,100, as determined by relative Cq values, when compared with an assay utilizing a previously published, frequently employed, ribosomal ITS DNA target. A comparative analysis of 2,784 field-collected samples was then performed, successfully verifying this improved sensitivity.

### 4.2.3 Conclusions/significance

Through analysis of the germline genome sequence of *A. lumbricoides*, a vastly improved qPCR assay has been developed. This assay, utilizing a high copy number repeat target found in eggs and embryos (the AGR repeat), will improve prevalence estimates that are fundamental to the programmatic decision-making process, while simultaneously strengthening mathematical models used to examine STH infection rates. Furthermore, through the identification of an optimal target for PCR, future assay development efforts will also benefit, as the identity of the optimized repeat DNA target is likely to remain unchanged despite continued improvement in PCR-based diagnostic technologies.

## 4.3 Author Summary

With an at-risk population in the billions, *Ascaris lumbricoides* is a pathogen of great global importance. In recent years, efforts to control the spread of this parasitic helminth have expanded, resulting in declining infection rates and worm burdens in some regions. While immeasurably important for global health, these declines have also served to expose the shortcomings of traditional diagnostic methods, as low-levels of pathogen generate a need for more sensitive tools, and microscopy-based techniques are proving ill-suited to the task at hand. Thankfully, improved sensitivity can be achieved through the

careful selection of optimal repetitive DNA targets for PCR. However, previous attempts to identify such targets in *A. lumbricoides* were unsuccessful, largely due to chromosome diminution, an unusual phenomenon occurring in the Ascaridida, whereby large portions of the germline genome are reproducibly eliminated during early development, resulting in their absence in larvae or adult worms. As the stool-based molecular diagnosis of *A. lumbricoides* infection is primarily dependent upon the identification of egg-derived DNA, utilizing genomic DNA from adult worms for molecular target selection eliminates germline candidates and results in suboptimal target sequence choices. Recently, the publication of a pre-diminution germline genome of *A. lumbricoides* has provided us with an opportunity to re-evaluate target selection, facilitating the development of a novel quantitative real-time PCR assay with greatly improved sensitivity (~3100-fold as determined by relative Cq value) over previously developed assays that were based on ribosomal repeat DNA sequences with lower copy numbers.

### 4.4 Introduction

Believed responsible for more than 800 million global infections, *Ascaris lumbricoides* is the most prevalent of the human-infecting soil transmitted helminths (STH) [1-2]. As recently as 2017, infections with this parasite were believed to result in approximaely 861,000 disability adjusted life years [3], generating nearly 45% of the global years lived with disability attributable to the overall burden of STH infections [3]. Due to an improved understanding of the scope of this disease burden, there is now an increased recognition of the global health impact of *A. lumbricoides* and the other STH infections. Such awareness has resulted in the expansion of infection and risk mapping

efforts [4-9] and operational research studies intended to improve, expand, and more fully understand the impacts associated with interventions [10-15]. Similarly, due to exponential improvements in approaches to mathematical modelling, the roles played by these valuable tools for shaping and informing the STH programmatic decision making process continue to increase [5-6, 16-18]. Fundamentally, such operational research efforts and modelling strategies rely heavily upon the availability of accurate data. Such reliance is particularly critical following interventions that have resulted in declining prevalence, drawing greater attention to the ramifications of employing insensitive diagnostic methods such as Kato-Katz [19]. Therefore, sensitive and specific diagnostic tools facilitating the collection of accurate data are increasingly critical for the proper interpretation of findings and the veracity of resulting conclusions.

Previously, we described a pipeline for the identification of high-copy number repetitive DNA elements for use as semi-quantitative real-time PCR (qPCR) targets for the detection of various STH species [20-21]. These targets, identified utilizing next-generation sequencing (NGS)-based analysis tools, have facilitated improved sensitivity and specificity of detection, leading to their adoption in various diagnostic efforts and operational research (OR) studies such as the DeWorm3 cluster randomized trials [11, 14, 22]. Despite the availability of such tools, to date, the qPCR-based detection of *A. lumbricoides* has depended upon less optimal targets, such as ribosomal internal transcribed spacer (ITS) sequences [20, 23-24]. This shortcoming is rooted in the unique process of chromosome diminution, whereby some species, including certain members of the order Ascaridida, undergo programmed elimination of select and reproducible regions of their gDNA during development [25-26]. In the case of *A. lumbricoides*, diminution

occurs between the third and seventh embryonic divisions [27], and an estimated 13% of the haploid germline genome is eliminated by this process [25], including the most abundant of the genome's tandemly repeated sequences [26]. Such elimination of highly repetitive, non-coding sequences during embryonic development renders ribosomal repeats the highest copy number gDNA sequences remaining in the genomes of larval and adult *Ascaris* worms. As pure gDNA is more easily obtained from adult worms, initial analyses using our pipeline utilized adult DNA extracts, and therefore failed to identify repeats present at higher copy number than the ribosomal ITS sequences [20]. However, STH diagnosis is dependent upon the detection of DNA from eggs/early embryos extracted from the stool of infected individuals. Thus, identifying an optimal qPCR target requires examination of egg-derived DNA, possessing pre-diminution gDNA sequences.

Acknowledging this shortcoming in the currently available PCR diagnostic toolkit, we now describe the development of an *Ascaris* germline assay utilizing a highly repetitive DNA element whose copy number is reduced by an estimated 99% in the post-diminution genome of larval and adult worms [26]. This 120 bp target, hereafter referred to as the *Ascaris* germline repeat (AGR), was previously estimated to constitute approximately 8.9% of the *Ascaris* germline genome [26], and further analysis of germline sequence reads utilizing RepeatExplorer2, a Galaxy-based computational tool [28] supports the prediction that this tandem repeat represents the most abundant germline gDNA sequence. The incorporation of a new PCR-based assay utilizing this improved target into our previously described STH diagnostic pipeline [20-21], represents a significant diagnostic improvement with the capacity to aid future programmatic efforts.

## 4.5 Materials and Methods

### 4.5.1 Ethics statement

The use of human samples in this study was approved by the reviewing body at the International Centre for Diarrhoeal Disease Research, Bangladesh (protocol # PR-14105) and by the University of California at Berkeley Committee for Protection of Human Subjects (protocol # 2014-08-6658).

### 4.5.2 Repetitive DNA sequence analysis

Three independently prepared, paired-end DNA libraries of raw Illumina sequencing reads, previously utilized for the assembly of a reference-quality *A. lumbricoides* germline genome [26], were repurposed for use in this study (Sequence Read Archive [SRA] BioProject number PRJNA511996). Prior to SRA upload, all reads were trimmed to a uniform length of 93 bp and a full description of sequencing and filtering methodologies has been described elsewhere [26]. Utilizing a randomly selected subset of 500,000 reads, each paired-end library was analyzed using RepeatExplorer2, a Galaxy-based analysis tool for the identification of repetitive DNA elements [28]. These analyses were used to identify the highest copy number genomic DNA sequences, which were selected as PCR targets for further analysis (Figure 4.1). All RepeatExplorer analyses were performed using default settings, without advanced options, and with the "Select Queue" set to "Basic and Fast".

### 4.5.3 Assay design

Utilizing default parameters for PrimerQuest Tool software (Integrated DNA Technologies, Coralville, IA), a candidate primer-probe pairing was designed that targeted our identified DNA repeat sequence. Primer-BLAST, available from the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov), was employed to determine whether or not our candidate primers matched off-target template sequences found within the RefSeq Representative Genome Database, and NCBI's nucleotide collection database. Following analysis, primers and probe were synthesized by Integrated DNA Technologies. Probe chemistry included labeling with a 6-FAM fluorophore at the 5' end, and double quenching with ZEN (internal) and 3IABkFQ (3' end) chemistries (Table 4.1).

### 4.5.4 Assay validation and optimization

Assay validation and optimization experiments were performed as previously described [20-21]. Briefly, utilizing 200 pg of pure *A. lumbricoides* gDNA isolated from an adult female worm as template, optimal primer concentrations were determined by titrating forward and reverse primers in independent 7 µL reactions containing 3.5 µL of TaqPath ProAmp Master Mix (ThermoFisher Scientific, Waltham, MA). Employing doubling dilutions, primers were tested at concentrations ranging from 1000 nM to 62.5 nM, with forward and reverse primer concentrations tested in all possible dilution combinations. Optimal AGR primer concentrations were then utilized in reactions intended to verify assay specificity, whereby 2 ng of purified genomic DNA isolated from adult *Necator americanus*, adult *Ancylostoma duodenale*, adult *Ancylostoma ceylanicum*,

adult *Trichuris trichiura*, *Strongyloides stercoralis* L1 larvae, adult *Schistosoma mansoni*, adult *Anisakis typica*, adult *Baylisascaris procyonis*, and adult *Parascaris univalens*, were used as template in separate reactions. Additionally, testing against human DNA, gDNA from *Candida albicans* (strain L26) (BEI Resources, Manassas, VA), DNA from the common gut bacteria *Escherichia coli*, and gDNA from a "mock" microbial community (v5.2H) (BEI Resources) also occurred. As a final validation, a panel of 20 infection-naïve, commercially available human stool samples were obtained for testing (BioIVT, Westbury, NY). DNA extraction was performed as previously described [29] and each extract was then tested for the presence of *Ascaris* signal.

### 4.5.5 Generation of a plasmid control containing the assay target sequence

Utilizing our AGR qPCR assay primers, pure *A. lumbricoides* gDNA was amplified by conventional PCR. Reactions in 25 µL volumes, containing 12.5 µL of Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) and 500 nM concentrations of each primer were amplified with an initial 30 second incubation at 98 °C; followed by 35 cycles of 98 °C for 10 seconds, 63 °C for 30 seconds, and 72 °C for 30 seconds; and a final 2 minute extension step at 72 °C. Following cycling, PCR products were cloned into the pCR-Blunt II-TOPO vector (ThermoFisher Scientific) in accordance with the manufacturer's suggested protocol, and NEB Express Competent *E. coli* (New England Biolabs) were transformed with 3 µL of the ligated plasmid. Transformed competent cells were then plated on LB-kanamycin plates, and grown at 37 °C overnight. Colonies were picked, and colony PCR was performed in 25 µL reactions containing 12.5 µL of One Taq 2x Master Mix (New England Biolabs), with 500 nM M13 forward and

reverse primers. Cycling began with an initial 30 second denaturation at 94 °C, followed by 35 reaction cycles of 94 °C for 15 seconds, 44 °C for 30 seconds, and 68 °C for 90 seconds; and a final extension for 5 minutes at 68 °C. Reaction products were sequenced, and a plasmid clone containing a single copy of the correct AGR repeat element was selected for use as a positive control in all future experiments.

### 4.5.6 Determination of assay efficiency

In order to determine assay efficiency, a panel of 10-fold plasmid serial dilutions was generated. Dilutions ranged from 100 pg/µL to 100 ag/µL. Because the control plasmid is 3,595 bp in size, and the average mass of a single nucleotide base pair is estimated to be 650 Da, 100 ag of plasmid was estimated to correspond to approximately 50 copies of the plasmid. Utilizing this information, approximate copy numbers were calculated for each concentration within the serial dilution series. Optimized reaction conditions were then employed to perform 11 or 12 reaction replicates for each dilution. Mean Cq values were calculated for reactions performed on each concentration of template, and a reaction efficiency was calculated.

### 4.5.7 Determination of assay detection limits

To determine assay detection limits, a panel of banked DNA extracts, previously isolated (as described elsewhere [29]) from 50 mg naïve stool samples that had been spiked with known numbers of *A. lumbricoides* eggs, was tested for the presence of detectable levels of *A. lumbricoides* target DNA. These samples were prepared and extracted as part of an ongoing, unrelated study in which the identification and isolation of eggs utilized for

spiking was performed using the McMaster egg counting technique as previously described [30]. Eggs were carefully removed from their parent samples under the microscope, briefly rinsed in nuclease-free water, and then added to the naïve stool. Following the addition of eggs, DNA was extracted from spiked aliquots as previously described [29]. All testing occurred in duplicate, and was performed using the experimentally-determined optimal AGR assay conditions. To facilitate inter-assay comparison, samples were similarly tested using a previously described assay that targets a ribosomal ITS2 sequence [20]. In total, 19 samples were tested. Four samples were spiked with 40 eggs, four with 10 eggs, four with 5 eggs, and four with 2 eggs. An additional three samples containing DNA from a single egg completed the panel.

### 4.5.8 Assay validation utilizing field-collected samples

#### 4.5.8.1 Collection of samples and assessment of quality

A panel of 2,784 previously isolated (as described elsewhere [29]) human stool DNA extracts were utilized in this study. All samples were collected in Bangladesh as part of the WASH Benefits Bangladesh trial [15]. At the time of processing, extraction quality was assessed, as previously described [29], through the procedural inclusion of an internal amplification control (IAC) plasmid. Following the extraction of all DNA samples, the recovery of IAC plasmid, spiked into each sample during the extraction procedure, was assessed using qPCR, and a mean Cq value and standard deviation (SD) was calculated for all samples. To ensure extraction quality and data comparability, all samples which produced a Cq value > 3SD from the mean underwent re-extraction and re-testing.

#### 4.5.8.2 qPCR testing

All samples and standards were tested in duplicate, utilizing the optimized version of the newly described AGR assay. After assaying all samples, results for each sample were compared with results obtained using the previously described ITS2-targeting assay referenced above [20]. While testing with both assays occurred sequentially, rather than simultaneously, samples were properly stored at -20 °C until all testing was completed. On all experimental plates, plasmid controls were run at concentrations of 10 pg/µL, 100 fg/µL, and 1 fg/µL. Following the completion of all testing, standard deviations were calculated for the mean Cq values for each control concentration. For a given experimental plate, in the event that the mean Cq value for any control dilution was > 2SD from the mean, the entire plate was retested. In the event that a given sample produced one positive and one negative result, the sample was re-tested in duplicate. For such samples, a positive result was ascribed to the sample if it again produced one or more positive results upon re-testing. Samples which produced two negative results upon re-testing were considered to be negative. All criteria for test validity and sample positivity/negativity determination were identical to those used for testing with the ITS2-targeting assay.

#### 4.5.8.3 Statistical analysis

An agreement table was generated to assess assay concordance/discordance. Samples were scored as "true positives" when they were found to be positive by either experimental assay. Assay sensitivities were calculated by dividing the number of positives as determined by a given assay, by the total number of "true positives".

## 4.6 Results

### 4.6.1 Target sequence identification and assay design

RepeatExplorer2 analysis software was employed to identify genomic DNA elements of putatively greatest copy number from sequence data derived from three, previously prepared, paired-end libraries of egg-derived *A. lumbricoides* DNA [26]. For each library, the RepeatExplorer-generated cluster containing the largest number of DNA sequence reads contained a 120 bp satellite element, previously identified as the most numerous within the *A. lumbricoides* germline genome [26]. Similarly, for each analyzed library, a supercluster comprised of multiple clusters each mapping to this repeat was predicted to represent between 7.9% and 16.3% of the germline genome. While it is important to note that such superclusters contain additional sequence fragments (flanking regions, etc.), when considered as rough representations of a sequence's genome percentage, these estimates are consistent with the 8.9% prediction made by Wang, et al [26]. These results strongly suggest that this repetitive sequence is the most prevalent repeat DNA element within the germline genome of *A. lumbricoides*. Utilizing this 120 bp AGR sequence, PrimerQuest Tool software was employed to design a candidate primer-probe set and Primer-BLAST analysis of this set returned only *Ascaris*-derived product predictions, minimizing the likelihood of experimental off-target PCR amplification.

### 4.6.2 Validation and optimization

As previously described, a titration of doubling dilutions of primer candidates was employed to determine optimal primer concentrations [20]. As determined by mean Cq

value, optimal concentrations were determined to be 125 nM for the forward primer and 500 nM for the reverse primer. Utilizing these primer concentrations, assay specificity was verified: 2 ng of template failed to produce off-target amplification for any of the species or samples tested. Similarly, testing of all DNA extracts from the infection-naïve stool panel failed to produce *Ascaris* signal, indicating that cross-reactivity with common elements of the gut flora is unlikely to occur.

### 4.6.3 Determination of assay efficiency

By testing a titration of our generated control plasmid (Figure 4.2), assay efficiency was determined. Utilizing plasmid size to determine target copy number per titration, a standard curve was generated by plotting Log(target copy #) vs. mean Cq value (Figure 4.3). The slope of this curve was determined to be -3.3216, with a reaction efficiency of 100.1% and an amplification factor of 2.00.

### 4.6.4 Determination of Assay Detection Limits

Utilizing DNA extracts obtained from naïve stool samples spiked with known numbers of *A. lumbricoides* eggs, assay detection limits were determined. Results using the new AGR assay indicated that target detection was possible from all stool samples spiked with all tested concentrations of eggs ranging from 40 eggs to a single egg (Table 4.2). In contrast, results obtained when testing with the ITS-targeting assay failed to allow for consistent detection at both 1 and 2 egg concentrations (Table 4.2).

### 4.6.5 Assay validation utilizing field-collected samples

Comparing results obtained using the newly described qPCR AGR assay with those generated through testing with a previously described, ribosomal ITS-targeting qPCR assay [20], an analysis of 2,784 human stool DNA extracts was performed. 349 samples were determined to be positive using the ITS-targeting assay, while 643 samples were determined to be positive utilizing the newly described qPCR AGR assay. Of the 349 ITS-assay positives, only two were negative when tested by the new AGR assay. In contrast, of the 643 samples determined to be positive by the AGR assay, 296 were negative when tested by the ribosomal ITS assay (Table 4.3). This led to a sensitivity of 99.69% for the AGR assay, and an ITS-targeting assay sensitivity of 54.11%. Minimum, maximum, median, and quartile values for the ITS-assay-positive sample population, the AGR-assay-positive sample population, and the AGR-assay-positive, ITS-assay-negative sample population are shown in Figure 4.4.

In an attempt to quantify the improvement in reaction sensitivity offered by the new AGR assay, an average reduction in mean Cq value was calculated for all samples which tested positive by both experimental assays, excluding a single sample which produced a lower Cq value when tested using the ribosomal ITS assay (n=346). To calculate this average reduction in mean Cq, the difference in mean Cq values for each co-positive sample was determined by subtracting the mean Cq value for the ribosomal ITS-targeting assay from the mean Cq value for the AGR assay. The average of these differences was then determined to be 11.51 cycles (range of 0.55 – 14.99) (Figure 4.5). This average change in Cq value corresponds to a fold increase in target number between the two qPCR

assays of ~3,100, which resulted in the detection of *Ascaris* DNA in nearly twice as many stool samples.

### 4.7 Discussion

With the expansion of treatment efforts and the resulting declines in infection prevalence and intensity, the sensitivity and specificity of STH diagnostic methods are increasingly important. Post-treatment surveys and population surveillance efforts are only as precise as the tools used to perform them and inconsistent tools may result in mismeasurement or misinterpretation of intervention impact. [31]. As such, diagnostic accuracy is critical for making assessments, and a given study's programmatic value is inherently tied to diagnostic capability. As OR efforts increasingly work towards the definition of transmission breakpoints, important decisions will be made based upon diagnostically determined prevalence levels under settings of declining parasite burden and decreasing infection intensities. The importance of diagnostic accuracy is embodied by the criteria governing the DeWorm3 cluster randomized trials, which state that "transmission interruption in a cluster will be defined as achieving a prevalence of each STH species of ≤2%...by qPCR 24 months after the final round of MDA" [11]. However, the attainment of a 2% prevalence rate is inherently linked to the test used for prevalence determination. Therefore, understanding diagnostic performance is critical for proper decision making, and maximizing diagnostic sensitivity increases confidence when breakpoint thresholds are attained, minimizing the odds of future recrudescence.

Previously, Easton, et al., described the theoretical limits of detection for both Kato-Katz and PCR as a function of the sample volume used for diagnosis [31]. As such, a 50

mg stool sample, analyzed by PCR has a theoretical limit of detection of 20 eggs per gram, should a single egg be present within the analyzed aliquot. However, Easton and colleagues also point out that with sufficient sensitivity, shed DNA, or DNA resulting from egg degradation could also be detected, allowing for further improvement over microscopy-based techniques that are dependent upon the presence of intact eggs within the sample aliquot tested [31]. While such levels of sensitivity may appear to have reduced importance when one considers that a single adult female *Ascaris* worm has been estimated to shed as many as 200,000 eggs per day [32], egg shedding varies considerably from person to person, and factors such as individual host immunity, geography, age of worm, worm burden, and intervention history can drastically alter patterns of egg production [33]. By selecting a molecular target with dramatically improved copy number, the capacity to detect pathogen signal is greatly improved, theoretically pushing limits of detection to previously impossible levels (Table 4.4).

Recognizing the need for optimal sensitivity in molecular diagnosis, we previously described the identification of improved qPCR targets for the detection of a number of human-infecting soil transmitted helminths [20-21]. However, due to the unusual phenomenon of chromosome diminution, whereby repeat-enriched portions of the genomic DNA are eliminated between the third and seventh cellular divisions, we were unable to identify an appropriate, novel, high copy-number repeat DNA element within the adult genome of *A. lumbricoides*. Recently, due to the publication of a reference-quality germline genome sequence for *A. lumbricoides* [26], we have been able to overcome this challenge with the selection of a highly repetitive DNA target that yields vastly superior sensitivity over previously utilized target DNA sequences. Present in both the *A.*

*lumbricoides* and *A. suum* germline genomes, this target facilitates improved diagnostic detection of all human *Ascaris* infections.

Representing an estimated 8.9% of the germline genome, yet only 120 bp in length, it is not surprising that the DNA target utilized by our new AGR assay facilitated a dramatic decrease in Cq values when compared to qPCR tests based on ribosomal DNA targets. With an estimated genome size of 334 Mb [26], nearly 2.5 x $10^5$ copies of this AGR element are believed to exist per haploid *A. lumbricoides* genome. This is in sharp contrast to the estimated 42 copies of ribosomal DNA present in *Ascaris* [34]. Interestingly, assuming similar reaction efficiencies, these copy numbers would suggest a Cq difference of just over 12, in near agreement with the 11.51 mean cycle difference which was determined during the experimental testing of field samples described here. Such drastic improvement in sensitivity should facilitate detection of *Ascaris* DNA at levels well below the quantity which is recoverable from a single egg, a hypothesis further supported by our spiking experiment results (Table 4.2). The validity of this sensitivity increase was reinforced by the results of the extensive specificity testing which we performed, providing strong evidence that the increased rates of positivity do not result from non-specific, off-target amplification.

It should be noted that a shortcoming of the performed spiking experiment was a failure to utilize an IAC during the DNA extraction procedure. However, as results for spiked sample testing existed for both the AGR assay and the ITS-targeting assay, an assessment of comparative sensitivity remained possible. While it is unfortunate that this failure to include an IAC prevented the drawing of meaningful correlations between Cq values and EPG levels, it is worth mentioning that large OR efforts, such as the DeWorm3

cluster randomized trials, aim to assess transmission break points based solely upon infection prevalence, irrespective of infection intensity [11].

In addition to their direct OR and surveillance functions, sensitive and specific diagnostic tools allow the research community to amass large bodies of accurate data, essential for the expansion and development of novel ideas and methodologies. The increased reliance on such data is seen in the incremental advancement of modeling efforts, a group of tools playing an ever-increasing role in both research and programmatic communities. Similarly, innovative ideas, such as the possibility of utilizing environmental sampling for STH surveillance [35] have historically been hampered by insufficient diagnostic options. However, with a resurging interest in these alternative methodologies [36-37], the availability of more sensitive tools will be critical, as such samples will likely rely upon larger sample masses, resulting in the dilution of molecular signal. Furthermore, while it is likely that future technologies will eventually render the current methods of qPCR-based diagnostics obsolete, prevalent targets will remain prevalent and may prove useful as new technologies come online. As such, the discovery of optimal targets should have a lasting impact on the field of infectious disease diagnostics.

While detection of parasite DNA target at sub-single egg concentrations greatly improves the sensitivity of detection, expanded sensitivity can also result in a potential complication. The issue is that higher copy-number DNA targets, coupled with excellent qPCR efficiencies, render assays increasingly susceptible to the possibility of sample-to-sample contamination. Such concerns are especially valid when the transfer of the technology to endemic countries is a priority. Deployment of such assays to varied laboratory environments can lead to an increased risk of false positive and false negative

results. Accordingly, highly sensitive qPCR assays require added attention to detail, and highlight the need for a renewed programmatic focus on proper training and project oversight. Equally important, appropriate quality assurance and quality control practices must be implemented, as must the use of consistent and standardized procedures and controls. Recognizing this need, options for external laboratory quality assessment are growing, and participation in assessment programs such as the Helminth External Molecular Quality Assessment Scheme offered by the Dutch Foundation for Quality Assessment in Medical Laboratories (SKML) should be considered whenever possible. Submitting to such external evaluations will help to ensure the accuracy of results and the inter-lab comparability of data.

Although infrequently voiced, an additional concern stems from the sometimes stated belief that optimization of a diagnostic assay can theoretically lead to the development of a test that is "too sensitive". The argument has been made that the detection of sub-cellular levels of DNA from cellular debris may result in the false attribution of a "positive" status to individuals who are not actually harboring active infection [38]. Similarly, for certain pathogens, detection of an individual microorganism may lead to diagnostic "positivity" under non-pathogenic concentrations [39-40]. However, such concerns are more relevant in the context of the clinical diagnosis of an individual patient. It is certainly true that sub-infectious levels of a pathogen may not pose a significant risk to the individual patient. Yet when used in a surveillance capacity, even sub-clinical levels of pathogen, or pathogen-derived material, are indicative of pathogen presence within the population. Oftentimes, such sub-clinical levels of pathogen may still pose a transmission risk within the community, facilitating persistence or providing an

early indication of possible infection recrudescence [41-42]. As such, when used for surveillance purposes, maximizing sensitivity should always be the diagnostic goal. However, it is equally important to remember that presence of pathogen signal is not necessarily an indicator of the potential for transmission. Factors such as single sex infections and expulsion of pathogen material can result in sample positivity despite failing to pose a transmission risk. For this reason it is critical that assay results be interpreted in the context of the study environment. Should the aims of a particular study dictate that only more heavily infected samples be of interest, a Cq value cutoff could be imposed, allowing the investigators to effectively filter out "light", potentially sub single-egg positive results without requiring changes to the testing procedure.

By targeting a highly repetitive element of the germline genome, the AGR qPCR assay described here has the capacity to greatly improve the sensitivity of detection of human *Ascaris* infections. This improvement should aid both operational research and programmatic efforts, increasing the accuracy of diagnostic results and facilitating better-informed decision making processes. Given the vast global prevalence of human *Ascaris* infection, the addition of this novel assay to the list of available molecular tools is of considerable significance.

### 4.8 Acknowledgements

The authors are indebted to Dr. Jack Colford, Dr. Jade Benjamin-Chung, Dr. Benjamin Arnold, and Dr. Ayse Ercumen (University of California, Berkeley) for providing the field-collected samples used for assay validation. The authors also thank Dr. Rojelio Mejia (Baylor College of Medicine) and Dr. Alejandro Krolewiecki (National

University of Salta, Argentina) for contributing “spiked” samples used for the determination of assay detection limits. We would also like to express our sincere gratitude to Dr. Ray Kaplan (University of Georgia) for providing *Anisakis* DNA, Dr. Martin Nielsen (University of Kentucky) for the pr*ovis*ion of *Parascaris* worms, and Dr. Michael Yabsley (University of Georgia) for graciously donating *Baylisascaris*. Finally, we thank Brian Abrams, Ashanta Ester, and Andrew Gonzalez Samara Loewenstein, and Marina Papaiakovou for their valuable assistance with laboratory procedures.

**Table 4.1 Primer and probe sequences for the AGR assay.**

| | |
|---|---|
| Forward Primer | 5’ – CTTGTACCACGATAAAGGGCAT - 3’ |
| Reverse Primer | 5’- TCCCTTCCAATTGATCATCGAATAA - 3’ |
| Probe | 5’ - /56-FAM/TCTGTGCAT/ZEN/TATTGCTGCAATTGGGA/3IABkFQ/ - 3’ |

**Table 4.2 Evaluation of limits of detection utilizing DNA extracts from spiked stool samples.**

| Number of Samples | Number of Eggs per Sample (EPG) | Mean Cq Value [Range][a] AGR Assay | Mean Cq Value [Range][a] ITS Assay |
|---|---|---|---|
| 4 | 40 (800) | 15.61 [14.69-16.77] | 25.40 [24.69-26.65] |
| 4 | 10 (200) | 15.99 [14.43-17.36] | 25.60 [24.11-26.78] |
| 4 | 5 (100) | 20.16 [18.52-21.85] | 30.56 [28.36-32.81] |
| 4 | 2 (40) | 25.26 [20.85-29.65] | 32.76 [31.33-33.85][b] |
| 3 | 1 (20) | 24.21 [19.12-32.44][d] | 30.21 [28.46-32.87][c] |
| 2 | No Template Control | 0.00 [0.00-0.00] | 0.00 [0.00-0.00] |

**[a]** Each spiked sample was run in duplicate, resulting in a mean for each sample. Mean Cq values for a given egg concentration were calculated as the mean value of all component sample means. The reported "Range" is the lowest individual Cq value and the highest individual Cq value for a given egg concentration. EPG, eggs per gram.

[b] Only two samples resulted in amplification of *Ascaris* target, with one of these samples amplifying in only 1 of 2 replicate reactions.

[c] Only two samples resulted in amplification of *Ascaris* target.

[d] One sample amplified in only 1 of 2 replicate reactions.

**Table 4.3 Agreement of assay results upon comparative testing of field-collected stool extracts.**

| | | Ribosomal ITS Assay | | |
|---|---|---|---|---|
| **AGR qPCR Assay** | | **Negative** | **Positive** | **Total** |
| | **Negative** | 2,139 (76.83%) | 2 (0.07%) | 2,141 (76.90%) |
| | **Positive** | 296 (10.63%) | 347 (12.47%) | 643 (23.10%) |
| | **Total** | 2,435 (87.46%) | 349 (12.54%) | 2,784 (100.00%) |

**Table 4.4 Causes of PCR positivity in the absence of microscopic identification of pathogen.**

| Circumstance Underlying Microscopic Detection Failure | Mechanism Facilitating qPCR Detection in Absence of Microscopic Detection |
|---|---|
| • Improper sample storage resulting in degradation of eggs | • Detection of short, tandem DNA repeats does not require intact eggs and repeats remain amplifiable even after significant fragmentation due to tandem arrangement |
| • Eggs are absent, but gDNA is present | • qPCR will detect target DNA, regardless of the DNA's source material (eggs, shed tissues, etc.) |
| • Light infection is missed due to human error | • Molecular detection is not subject to human error or fatigue |

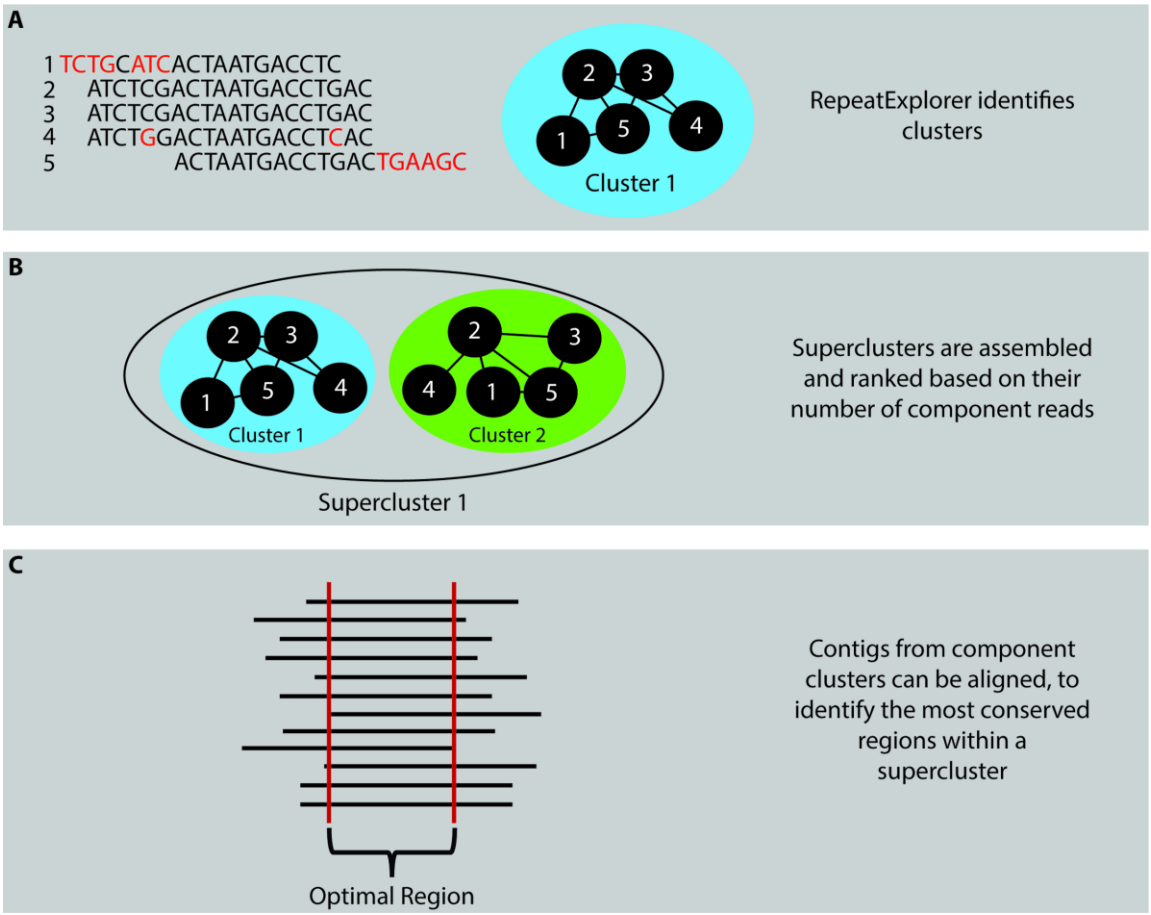


**Figure 4.1 Identification of an optimal PCR target using RepeatExplorer2.** (A) By comparing the individual sequence of each read to the sequence of every other read within the dataset of interest, RepeatExplorer builds "clusters" from reads meeting the cut-off criteria of having 90% or greater sequence identity over 55% or more of the read lengths. (B) Further comparison then identifies superclusters, comprised of clusters reaching a threshold level for paired-end read mates shared between clusters. (C) By aligning contigs/reads comprising component clusters within a supercluster, it becomes possible to identify regions of DNA sequence within each supercluster that have the greatest coverage. These highly repeated DNA regions are selected as targets for qPCR reaction primer and probe design.

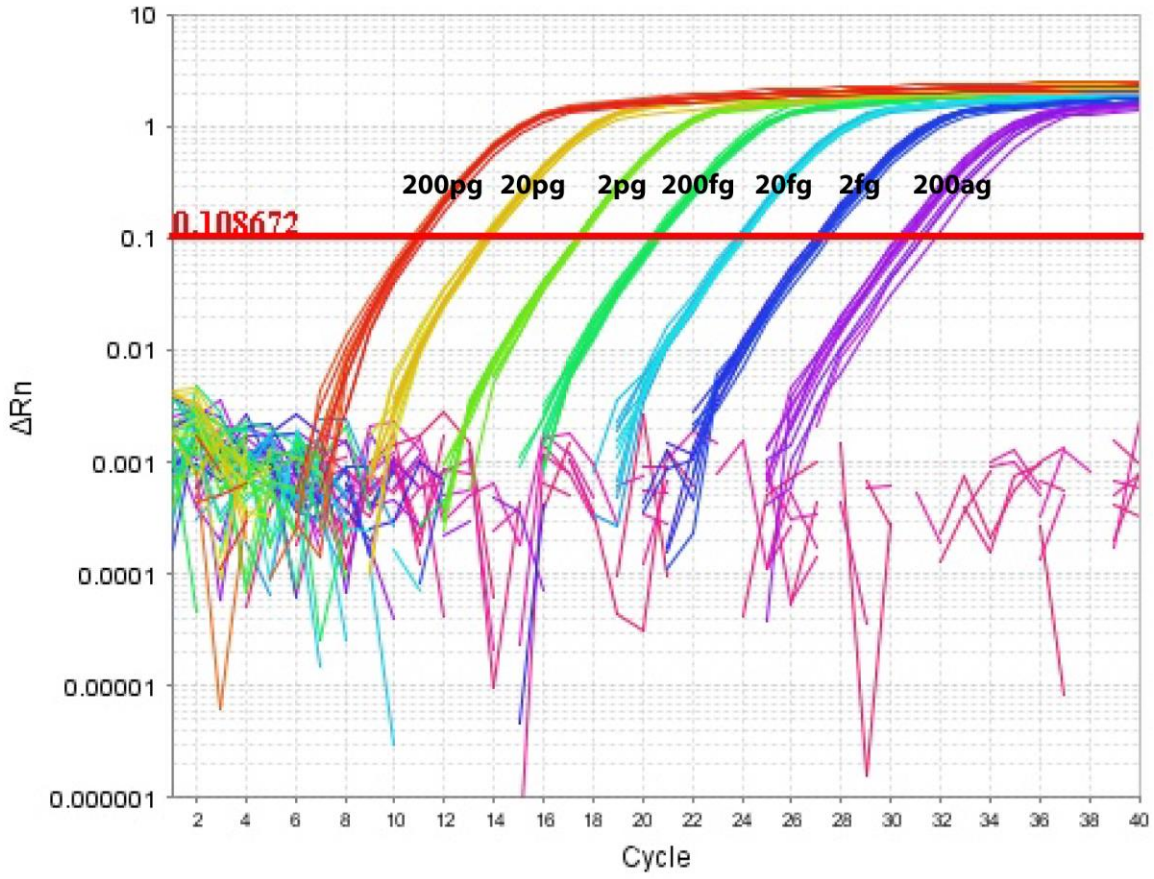


**Figure 4.2 Amplification plot for qPCR reactions used to calculate reaction efficiency.** Through replicate testing of titrated plasmid DNA containing a single copy of the reaction target sequence, amplification curves were used to determine mean Cq values for reactions occurring with each concentration of template.

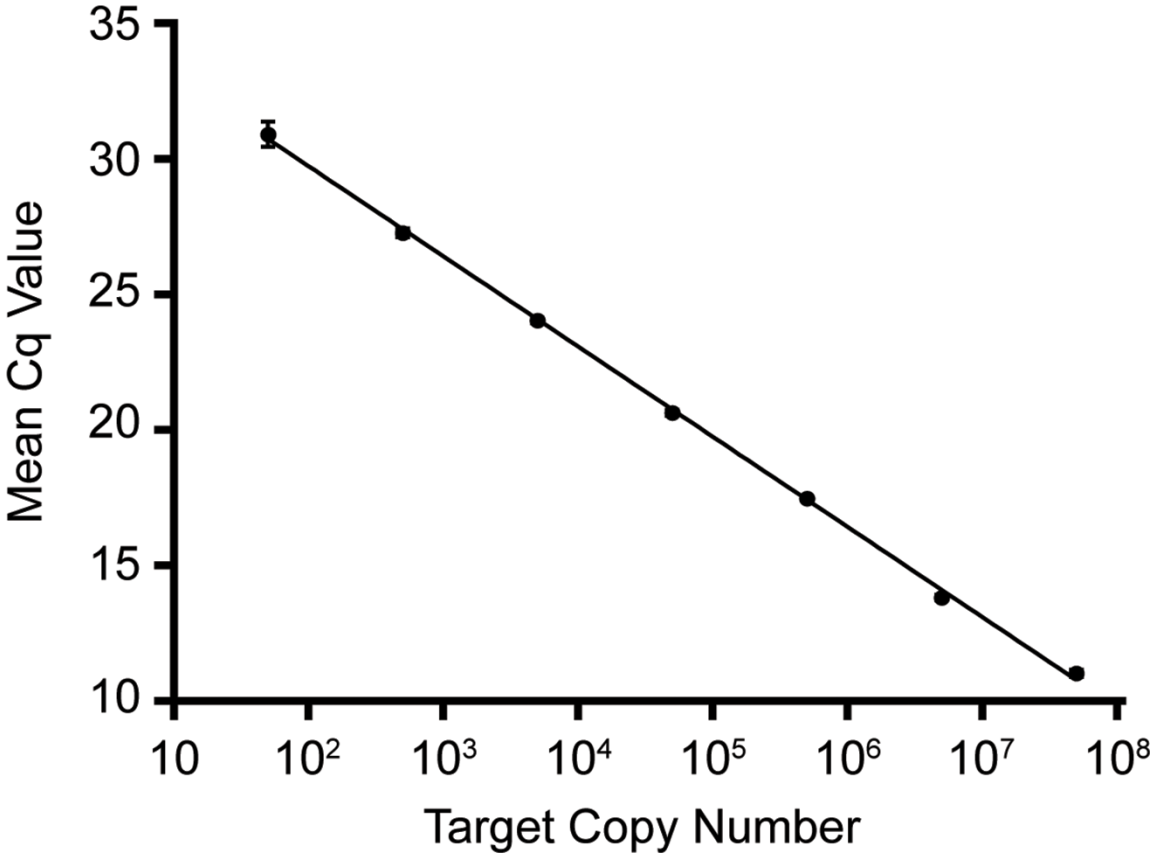


**Figure 4.3 Calculation of assay efficiency.** To determine assay efficiency, 10-fold serial dilutions of control plasmid were prepared. All dilutions, ranging in concentration from 100 pg/µL to 100 ag/µL, were analyzed in either 11 or 12 replicate reactions. Mean Cq values and standard deviations were then calculated for each concentration of plasmid template, a slope was plotted, and reaction efficiency and amplification factor were determined.

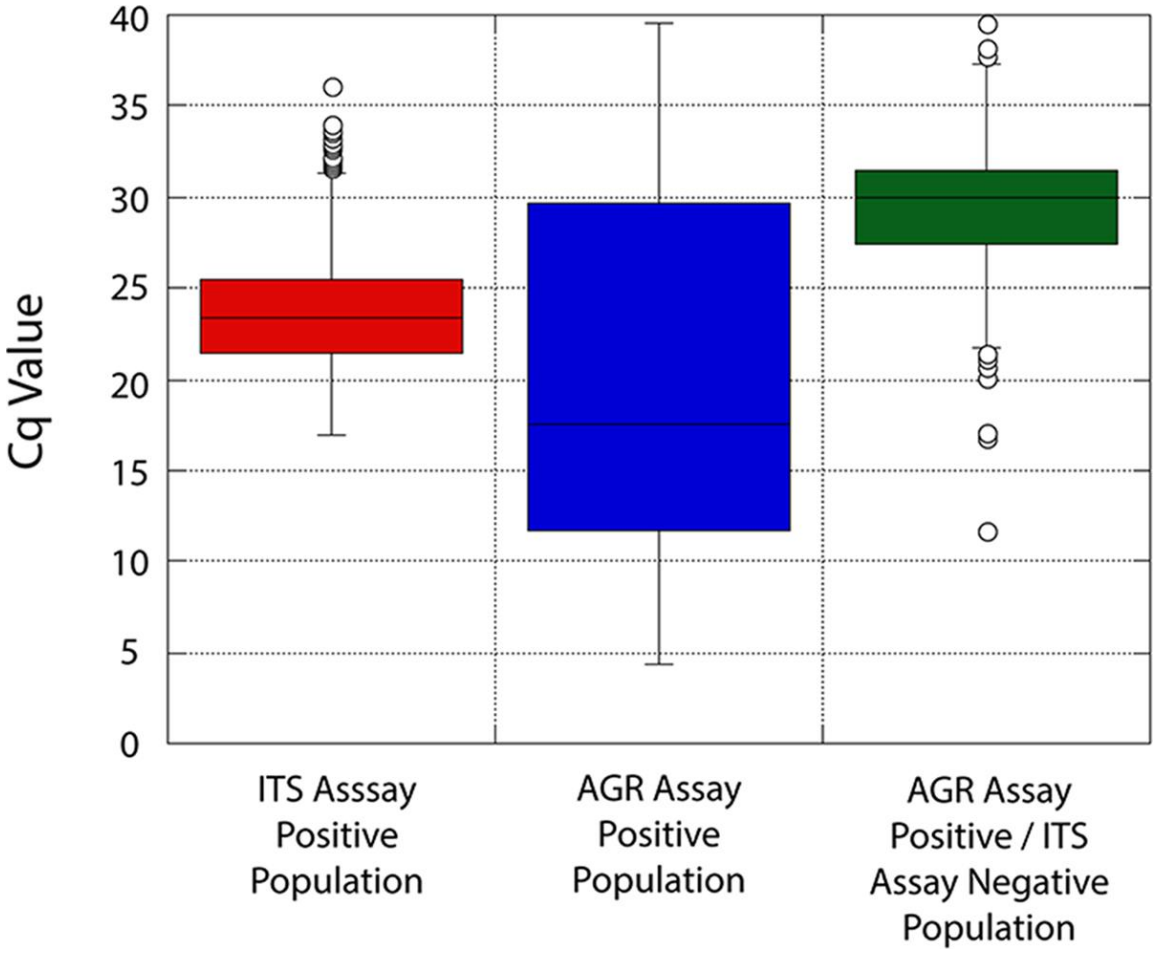


**Figure 4.4 Boxplots of positive sample populations from field-collected samples.** Plots represent the total population of ITS-assay-positive samples (n = 349), the total population of AGR-assay-positive samples (n = 643), and the population of ITS-assay-negative, AGR-assay-positive samples (n = 296). Medians are depicted by the horizontal lines, while the box for each plot represents the interquartile range (IQR), and whiskers represent Q3 + (1.5)(IQR) and Q1 – (1.5)(IQR).

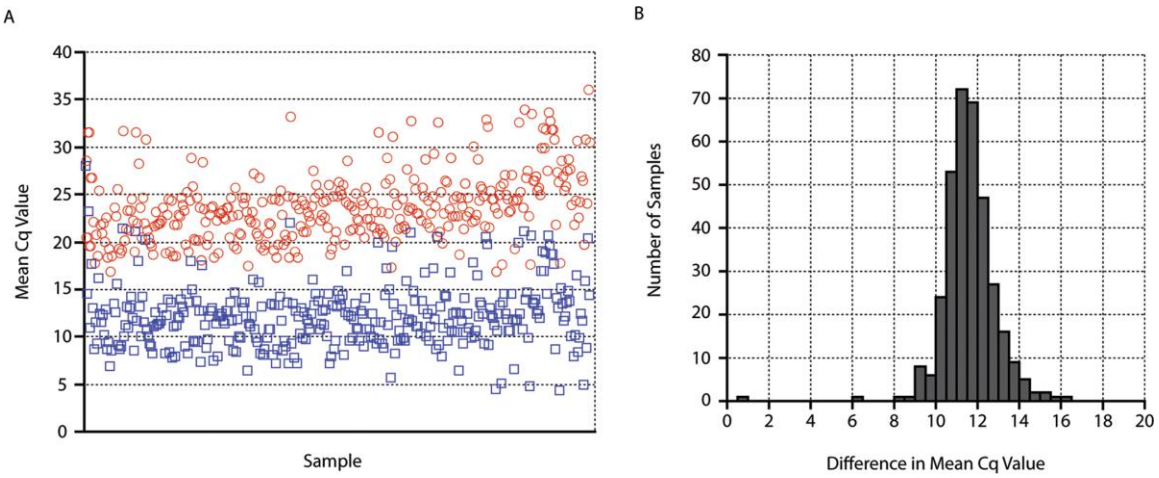


**Figure 4.5 Differences between mean Cq values for all samples co-positive using both the ribosomal ITS-targeting assay and AGR qPCR assay.** (A) For each co-positive sample, mean Cq values were plotted for both the ribosomal ITS-targeting assay (red circles) and the AGR assay (blue squares). (B) For each co-positive sample, a difference in mean Cq values was calculated by subtracting the mean value for the ribosomal ITS qPCR assay from the mean value for the AGR qPCR assay. Results were binned by difference and plotted. The average difference across all plotted samples was determined to be 11.51 cycles.

## 4.9 References


1. Brooker SJ, Pullan RL. *Ascaris lumbricoides* and Ascariasis: Estimating Numbers Infected and Burden of Disease. In: *Ascaris:* The Neglected Parasite. London: Academic Press; 2013. pp. 343-360.

2. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit Vectors. 2014; 7: 37.

3. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392: 1859-1922.

4. Scholte RG, Schur N, Bavia ME, Carvahlo EM, Chammartin F, Utzinger J, et al. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. Geospat Health. 2013; 8: 97-110.

5. Chammartin F, Scholte RG, Malone JB, Bavia ME, Nieto P, Utzinger J, et al. Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. Parasit Vectors. 2013; 6: 152.

6. Lai Y-S, Zhou X-N, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. Paraist Vectors. 2013; 6: 359.

7. Wardell R, Clements ACA, Lal A, Summers D, Llewellyn S, Campbell SJ, et al. An environmental assessment and risk map of *Ascaris lumbricoides* and *Necator americanus* distributions in Manufahi District, Timor-Leste. PLoS Negl Trop Dis. 2017; 11: e0005565.

8. Müller I, Gall S, Beyleveld L, Gerber M, Pühse U, Du Randt R, et al. Shrinking risk profiles after deworming of children in Port Elizabeth, South Africa, with special reference to *Ascaris lumbricoides* and *Trichuris trichiura*. Geospat Health. 2017; 12: 601.

9. Assoum M, Ortu G, Basáñez M-G, Lau C, Clements ACA, Halton K, et al. Spatiotemporal distribution and population at risk of soil-transmitted helminth infections following an eight-year school-based deworming programme in Burundi, 2007-2014. Parasit Vectors. 2017; 10: 583.

10. Means AR, Ajjampur SSR, Bailey R, Galactionova K, Gwayi-Chore MC, Halliday K, et al. Evaluating the sustainability, scalability, and replicability of an STH transmission interruption intervention: The DeWorm3 implementation science protocol. PLoS Negl Trop Dis. 2018; 12: e0005988.

11. Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, Bailey R, Gardiner I, Halliday KE, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3 cluster randomized trial protocol. PLoS Negl Trop Dis. 2018; 12: e0006166.

12. Arnold BF, Null C, Luby SP, Unicomb L, Stewart CP, Dewey KG, et al. Cluster-randomized controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. BMJ Open. 2013; 3: e003476.

13. Christensen G, Dentz HN, Pickering AJ, Bourdier T, Arnold BF, Colford Jr JM, et al. Pilot Cluster Randomized Controlled Trials to Evaluate Adoption of Water, Sanitation, and Hygiene Interventions and Their Combination in Rural Western Kenya. Am J Trop Med Hyg. 2015; 92: 437-447.

14. Pickering AJ, Njenga SM, Steinbaum L, Swarthout J, Lin A, Arnold BF, et al. Effects of Single and Integrated Water, Sanitation, Handwashing, and Nutrition Interventions on Child Soil-Transmitted Helminth and Giardia Infections: A Cluster-Randomized Controlled Trial in Rural Kenya. 2019; PLoS Med. 16; e1002841.

15. Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, et al. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. Lancet Glob Health. 2018; 6: e302-e315.

16. Farrell SH, Truscott JE, Anderson RM. The importance of patient compliance in repeated rounds of mass drug administration (MDA) for the elimination of intestinal helminth transmission. Parasit Vectors. 2017; 10: 291.

17. Coffeng LE, Truscott JE, Farrell SH, Turner HC, Sarkar R, Kang G, et al. Comparison and validation of two mathematical models for the impact of mass drug administration on *Ascaris lumbricoides* and hookworm infection. Epidemics. 2017; 18: 38-47.

18. Midzi N, Kavhu B, Manangazira P, Phiri I, Mutambu SL, Tshuma C, et al. Inclusion of edaphic predictors for enhancement of models to determine distribution of soil-transmitted helminths: the case of Zimbabwe. Parasit Vectors. 2018; 11: 47.

19. Werkman M, Wright JE, Truscott JE, Easton AV, Oliveira RG, Toor J, et al. Testing for soil-transmitted helminth transmission elimination: Analysing the impact of the sensitivity of different diagnostic tools. PLoS Negl Trop Dis. 2018; 12: e0006114.

20. Pilotte N, Papaiakovou M, Grant JR, Bierwert L, Llewellyn S, McCarthy JS, Williams SA. Improved PCR-Based Detection of Soil Transmitted Helminth Infections Using a Next-Generation Sequencing Approach to Assay Design. PLoS Negl Trop Dis. 2016; 10: e0004578.

21. Papaiakovou M, Pilotte N, Grant JR, Traub RJ, Llewellyn S, McCarthy JS, et al. A novel, species-specific, real-time PCR assay for the detection of the emerging zoonotic parasite *Ancylostoma ceylanicum* in human stool. PLoS Negl Trop Dis. 2017; 11: e0005734.

22. O'Connell EM, Mitchell T, Papaiakovou M, Pilotte N, Lee D, Weinberg M, et al. *Ancylostoma ceylanicum* Hookworm in Myanmar Refugees, Thailand, 2012-2015. Emerg Infect Dis. 2018; 24: 1472-1481.

23. Mejia R, Vicuña Y, Broncano N, Sandoval C, Vaca M, Chico M, et al. A Novel, Multi-Parallel, Real-Time Polymerase Chain Reaction Approach for Eight Gastrointestinal Parasites Provides Improved Diagnostic Capabilities to Resource-Limited At-Risk Populations. Am J Trop Med Hyg. 2013; 88: 1041-1047.

24. Liu J, Gratz J, Amour C, Kibiki G, Becker S, Janaki L, et al. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. J Clin Microbiol. 2013; 51: 472-480.

25. Wang J, Mitreva M, Berriman M, Thorne A, Magrini V, Koutsovoulos G, et al. Silencing of germline-expressed genes by DNA elimination in somatic cells. Dev Cell. 2012; 23: 1072-1080.

26. Wang J, Gao S, Mostovoy Y, Kang Y, Zagoskin M, Sun Y, et al. Comparative genome analysis of programmed DNA elimination in nematodes. Genome Res. 2017; 27: 2001-2014.

27. Goldstein P. Chromatin diminution in early embryogenesis of *Ascaris lumbricoides* L. var. *suum*. J Morphol. 1977; 152: 141-151.

28. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792-793.

29. Papaiakovou M, Pilotte N, Baumer B, Grant J, Asbjornsdottir K, Schaer F, et al. A comparative analysis of preservation techniques for the optimal molecular detection of hookworm DNA in a human fecal specimen. PLoS Negl Trop Dis. 2018; 12: e0006130.

30. Levecke B, Behnke JM, Ajjampur SSR, Albonico M, Ame SM, Charlier J, et al. A comparison of the sensitivity and fecal egg counts of the McMaster egg counting and Kato-Katz thick smear methods for soil-transmitted helminths. PLoS Negl Trop Dis. 2011; 5: e1201.

31. Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

32. Sinniah B. Daily egg production of *Ascaris lumbricoides*: the distribution of eggs in the faeces and the variability of egg counts. Parasitology. 1982; 84: 167-175.

33. Scott ME. *Ascaris lumbricoides*: A Review of Its Epidemiology and Relationship to Other Infections. Ann Nestlé. 2008; 66: 7-22.

34. Pecson BM, Antonio Barrios J, Johnson DR, Nelson KL. A Real-Time PCR Method for Quantifying Viable *Ascaris* Eggs Using the First Internally Transcribed Spacer Region of Ribosomal DNA. Appl Environ Microbiol. 2006; 72: 7864-7872.

35. Ulukanligil M, Seyrek A, Aslan G, Ozbilge H, Atay S. Environmental pollution with soil-transmitted helminths in Sanliurfa, Turkey. Mem Inst Oswaldo Cruz. 2001; 96: 903-909.

36. Steinbaum L, Njenga SM, Kihara J, Boehm AB, Davis J, Null C, et al. Soil-Transmitted Helminth Eggs Are Present in Soil at Multiple Locations within Households in Rural Kenya. PLoS One. 2016; 11: e0157780.

37. Steinbaum L, Kwong LH, Ercumen A, Negash MS, Lovely AJ, Njenga SM, et al. Detecting and enumerating soil-transmitted helminth eggs in soil: New method development and results from field testing in Kenya and Bangladesh. PLoS Negl Trop Dis. 2017; 11: e0005522.

38. MacGregor RR, Dreyer K, Herman S, Hocknell PK, Nghiem L, Tevere VJ, et al. Use of PCR in Detection of *Mycobacterium avium* Complex (MAC) Bacteremia: Sensitivity of the Assay and Effect of Treatment for MAC Infection on Concentrations of Human Immunodeficiency Virus in Plasma. J Clin Microbiol. 1999; 37: 90-94.

39. Machida U, Kami M, Fukui T, Kazuyama Y, Kinoshita M, Tanaka Y, et al. Real-Time Automated PCR for Early Diagnosis and Monitoring of Cytomegalovirus Infection after Bone Marrow Transplantation. J Clin Microbiol. 2000; 38: 2536-2542.

40. Mackay IM. Real-time PCR in the microbiology laboratory. Clin Microbiol Infect. 2004; 10: 190-212.

41. Martinez-Bakker M, King AA, Rohani P. Unraveling the Transmission Ecology of Polio. PLoS Biol. 2015; 13: e1002172.

42. Parker DM, Tripura R, Peto TJ, Maude RJ, Nguon C, Chalk J, et al. A multi-level spatial analysis of clinical malaria and subclinical *Plasmodium* infections in Pailin Province, Cambodia. Heliyon. 2017; 3: e00447.

# CHAPTER 5

# COMPARISON OF MULTI-PARALLEL qPCR AND DOUBLE-SLIDE KATO-KATZ FOR DETECTION OF SOIL-TRANSMITTED HELMINTH INFECTION AMONG CHILDREN IN RURAL BANGLADESH

*as written by Benjamin-Chung J.[#], Pilotte N.[#], Ercumen A., Grant J.R., Maasch J.R.M.A, Gonzalez A.M., Abrams B.P., Ester A.C., Arnold B.F., Rahman M., Haque R., Hubbard A.E., Luby S.P., Williams S.A., Colford, Jr. J.M.*
*[#] Equal contribution*
*bioRxiv, 2019*
*PLoS Neglected Tropical Diseases, In Press*

## 5.1 Chapter Overview

This chapter illustrates the advantages of using species-specific real-time PCR assays. Through comparison with Kato-Katz microscopy, the World Health Organizaton-recommended diagnostic technique, the increased clinical sensitivities of our molecular methodologies are demonstrated. The importance of assay specificity, made possible by the careful selection of molecular targets is also demonstrated, as Kato-Katz microscopy results in the misidentification of over 700 samples, wrongly classified as *Ascaris*-positive. A combination of techniques including amplicon sequencing and qPCR-based testing with multiple assays each targeting a unique region of the gDNA are used to show that samples that are positive for *Ascaris* by Kato-Katz, but negative for *Ascaris* by qPCR, are in fact *Ascaris* negative. This chapter tells a cautionary tale, illustrating the importance of employing sensitive and specific assays for programmatic research efforts.

A secondary objective of the analyses described in this chapter was to correlate qPCR-derived results with infection intensities as determined by Kato-Katz microscopy.

However, due to the wide-spread microscopy-based misidentifications, meaningful conclusions are not drawn.

My contributions to this chapter included directly performing, or overseeing the performance of all real-time PCR testing. I also prepared all libraries, performed all sequencing reactions and analyzed all sequencing results. I further contributed to data analysis, table and figure generation, and I co-wrote the manuscript.

## 5.2 Abstract


### 5.2.1 Background

There is growing interest in local elimination of soil-transmitted helminth (STH) infection in endemic settings. In such settings, highly sensitive diagnostics are needed to detect STH infection.

### 5.2.2 Methodology

We compared double-slide Kato-Katz, the most commonly used copromicroscopic detection method, to multi-parallel quantitative polymerase chain reaction (qPCR) in 2,799 stool samples from children aged 2-12 years in a setting in rural Bangladesh with predominantly low STH infection intensity. We estimated the sensitivity and specificity of each diagnostic using Bayesian latent class analysis.

### 5.2.3 Principal findings

Compared to double-slide Kato-Katz, STH prevalence using qPCR was almost 3-fold higher for hookworm species and nearly 2-fold higher for *Trichuris trichiura*. *Ascaris lumbricoides* prevalence was lower using qPCR, and 26% of samples classified as *A. lumbricoides* positive by Kato-Katz were negative by qPCR. Amplicon sequencing of the 18S rDNA from 10 samples confirmed that *A. lumbricoides* was absent in samples classified as positive by Kato-Katz and negative by qPCR. The sensitivity of Kato-Katz was 49% for *A. lumbricoides*, 32% for hookworm, and 52% for *T. trichiura*; the sensitivity of qPCR was 79% for *A. lumbricoides*, 93% for hookworm, and 90% for *T. trichiura*.

Specificity was ≥ 97% for both tests for all STH except for Kato-Katz for *A. lumbricoides* (specificity = 68%). There were moderate negative, monotonic correlations between qPCR cycle quantification values and eggs per gram quantified by Kato-Katz.

### 5.2.4 Conclusions

While it is widely assumed that double-slide Kato-Katz has few false positives, our results suggest otherwise and highlight inherent limitations of the Kato-Katz technique. qPCR had higher sensitivity than Kato-Katz in this low intensity infection setting.

## 5.3 Author Summary

Soil-transmitted helminth infections (STH) (e.g., *Ascaris*, hookworm, *Trichuris*) contribute to a large burden of disease among children in low- and middle-income countries. There is increasing interest in implementing large-scale deworming programs to eliminate STH in certain settings. Efforts to monitor whether local elimination has occurred require sensitive diagnostic tests that will not miss positive cases. Kato-Katz, a microscopy-based diagnostic test, has commonly been used to identify STH eggs in stool, but in settings where infection intensity is low, this method frequently misses positive samples because it requires visual identification of small numbers of eggs, and hookworm eggs may degrade prior to visualization. Quantitative polymerase chain reaction (qPCR) is a molecular diagnostic method for detecting STH. It may detect more low intensity infections than Kato-Katz because it identifies STH DNA in stool; DNA can be detected in very small quantities and is less likely to degrade than STH ova. Thus, qPCR is likely to be more accurate than Kato-Katz. This study compared the performance of double-slide

Kato-Katz and qPCR using 2,799 stool samples from children aged 2-12 years in a setting in rural Bangladesh with predominantly low STH infection intensity. qPCR was more sensitive than Kato-Katz for hookworm and *Trichuris* infections. 26% of samples were classified as *Ascaris* positive by Kato-Katz and negative by qPCR. DNA sequencing of 10 samples confirmed that *Ascaris* was absent in samples classified as positive by Kato-Katz and negative by qPCR. We conclude that Kato-Katz likely produced false positive results for *Ascaris* and that qPCR had a higher sensitivity than double-slide Kato-Katz in this low infection intensity setting.

## 5.4 Introduction

Soil-transmitted helminths (STH) infect an estimated 1.5 billion individuals, almost one fifth of the global population [1]. These infections contribute to a substantial burden of disability and disease, particularly for children [2]. To control the burden of STH infection, mass drug administration programs deliver preventive chemotherapy to populations at risk, such as school children. In recent years, approximately 33% of children in endemic settings received deworming medication for STH in the past six months according to caregiver report [3]. Such programs have expanded dramatically following the London Declaration on Neglected Tropical Diseases in 2012. Mass drug administration (MDA) programs have yielded large reductions in STH infection prevalence and intensity to date [4]. While the World Health Organization has set goals for STH focused on morbidity control [5], there is an active body of research focused on whether local elimination of STH is possible in endemic settings through community-wide MDA for deworming [6–9]. As STH infection intensity and prevalence wane, increasingly sensitive

diagnostic methods are required to determine whether STH transmission has been interrupted and when interventions can be scaled back or discontinued [10].

Diagnostic methods for STH include copromicroscopic methods, such as the Kato-Katz method [11, 12], formol-ether concentration [13], FLOTAC [14], mini-FLOTAC [15–18], and FECPAKG2 [18, 19]. Of these, Kato-Katz is the most commonly used diagnostic method for STH surveillance because it is inexpensive and relatively easy to perform in low-resource settings [20]. A significant limitation of this method is that samples must be evaluated within an hour of preparation in order to detect fragile hookworm ova. In addition, *Necator americanus* and *Ancylostoma duodenale* cannot be distinguished using Kato-Katz because of their morphological similarities. Because egg excretion is highly variable over time [21], there can be substantial variation in the count of eggs per gram (epg) of stool collected on different days. Kato-Katz has improved sensitivity when performed on multiple samples from defecation events on different days, but such samples are often logistically prohibitive to collect. Double-slide Kato-Katz examines stool using two separate aliquots from a single defecation event prepared as separate slides in order to help reduce the variability and increase sensitivity. Furthermore, a meta-analysis estimated the sensitivity of double-slide Kato-Katz to be 64.6% for *Ascaris lumbricoides*, 84.8% for *Trichuris trichiura*, and 63.0% for hookworm [22]. Another limitation of Kato-Katz is that its sensitivity is lower in low infection intensity settings (55.2% for *A. lumbricoides*, 79.8% for *T. trichiura*, and 52.6% for hookworm) [22]. Alternative copromicroscopic STH diagnostic methods, such as formol-ether concentration, FLOTAC, mini-FLOTAC, and the McMaster method also have moderate to low sensitivity for different types of STH in low-intensity settings [22].

Multiplex [23, 24] and multi-parallel [25, 26] quantitative polymerase chain reaction (qPCR) assays have been developed to detect STH DNA in stool. Multiplex and multi-parallel assays allow for detection of multiple helminths from a single stool sample. Growing evidence suggests that these methods may be used to estimate infection intensity, although qPCR-based cutoffs for light, moderate, and heavy intensity infections have not yet been defined [24, 27, 28]. To our knowledge, six studies have compared diagnostic performance of copromicroscopic methods and qPCR; these studies reported that analyses classifying STH status using qPCR as a diagnostic had higher prevalence [25], sensitivity, and specificity compared to those using single- and/or double-slide Kato-Katz as a diagnostic [18, 23, 27-29].

The greater sensitivity, specificity, and precision of qPCR relative to Kato-Katz makes it an attractive diagnostic for monitoring the successes of STH control programs deploying mass deworming since infection intensity tends to decline after several years of intervention [30]. As a population reaches an STH transmission breakpoint, worm burdens decrease, and the comparatively low sensitivity of Kato-Katz may preclude its use in monitoring STH prevalence. In addition, a growing body of work suggests that qPCR may also be used to estimate infection intensity; one study found that qPCR is approximately 3.6 times more precise than Kato-Katz in estimating *A. lumbricoides* egg intensity [31]. Such additional precision could increase the statistical power available to detect whether the STH prevalence meets the threshold for elimination.

The primary objectives of this study were to estimate the prevalence of STH infection and infection intensity using double-slide Kato-Katz and multi-parallel qPCR and

to estimate the sensitivity and specificity of each method using Bayesian latent class models and archived samples collected from children aged 2-12 years in rural Bangladesh.

## 5.5 Materials and Methods

### 5.5.1 Study population

This study analyzed archived stool samples that were collected from participants in the WASH Benefits Bangladesh trial between May 2015 and May 2016 (Clinicaltrials.gov NCT01590095) [32, 33]. The trial was implemented in Gazipur, Mymensingh, Tangail and Kishoreganj districts of Bangladesh. Biannual school-based mass drug administration with mebendazole had been implemented nationally in Bangladesh since 2008 and was ongoing in the study area during the trial. The trial delivered interventions to improve water, sanitation, handwashing, and nutrition and included single and combined intervention arms. Full details of interventions and user adherence have been published elsewhere [33, 34]. This study randomly selected 2,800 stool specimens collected from children aged 22 months to 12 years of age (mean = 57 months) enrolled in the individual improved water, sanitation, handwashing, and combined water+sanitation+handwashing (WSH) arms. This sample size was chosen in order to estimate sensitivity with precision of +/- 5% and 80% statistical power; we assumed STH prevalence estimates from a prior study in rural Bangladesh [35] and double-slide Kato-Katz sensitivity and specificity estimates from a prior meta-analysis of STH diagnostic accuracy [22].

### 5.5.2 Stool collection and storage

Field staff provided study participants' primary caregivers with sterile containers and requested that they collect stool from the child's defecation event the next morning. Field staff returned to the house at least three times to attempt to collect stool containers. As part of the WASH Benefits trial, all study participants received a single dose of albendazole following stool collection. Field staff transported stool specimens on ice to the field laboratory, where 1 g of stool was archived in 1 ml of 100% ethanol. Specimens were stored at -20 °C prior to relocation to a -80 °C freezer. In the WASH Benefits parasites sub-study, the median time that elapsed between the reported time of defecation and when the stool was first placed on ice was 2.4 hours (range: 0.03 to 18 hours), and the median time that elapsed between the date of specimen collection and the date when specimens were stored in a -80 °C freezer was 11 days (range: 1 to 330 days).

### 5.5.3 Kato-Katz procedures

Technicians were trained in double-slide Kato-Katz using the Vestergaard Frandsen protocol at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) parasitology laboratory. On the day of stool collection, trained technicians performed double-slide Kato-Katz [11, 12] on fresh stool within 30 minutes of preparing each sample. For quality assurance, 10% of slides were evaluated independently by two technicians, and 5% were evaluated by the same experienced parasitologist who conducted Kato-Katz training. The quality assurance results have previously been published elsewhere [34]. In brief, agreement between laboratory technicians was high (Kappa statistic > 0.99 for each STH). The Kappa statistic for agreement between laboratory

technicians and experienced parasitologists was 0.92 for *A. lumbricoides*, 0.20 for hookworm, and 0.86 for *T. trichiura*. The agreement is likely lower for hookworm because experienced parasitologists examined slides up to a few days after laboratory technicians, and hookworm ova may have begun to disintegrate by that time. For the subset of samples for which quality control was performed on the same day, the agreement for hookworm was high. Agreement decreased as the time between the original assessment and the senior parasitologist's assessment increased (Table 5.1). Samples in which at least one STH egg was visualized during Kato-Katz were classified as positive. Kato-Katz technicians did not have access to any clinical information about study participants.

### 5.5.4 DNA isolation and qPCR procedures

Approximately 1-2 years after stool collection, preserved stool specimens were shipped to Smith College in Northampton, MA, United States for qPCR analyses. Prior to shipment, ethanol was evaporated from all samples for compliance with shipping regulations, and samples were shipped on dry ice. Upon receipt at Smith College, all samples were stored at -20 °C until DNA extraction. DNA was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals, Solon, OH) utilizing a previously published modified version of the manufacturer's methodology [25]. Following DNA extraction and prior to qPCR analyses, extractions were stored at -20 °C. We assigned each sample a unique number following extraction, and this number was electronically linked to the initial barcode on the stool sample tube at the time of extraction. Results were then linked back to the initial sample using both the Smith-assigned number and the barcode, and samples underwent qPCR processing in the same batches that they were extracted in. An internal

amplification control (IAC) plasmid was employed during the extraction of each sample to ensure successful isolation, and adequate recovery of DNA [36]. Utilizing previously described reaction methods [37], any sample which failed to produce a positive qPCR result for one or more of the IAC replicates underwent re-extraction. Similarly, following the completion of all extractions, the mean quantitation cycle (Cq) value for the IAC results from all samples was calculated, as was the standard deviation of the mean. Any sample with an individual IAC result determined to be 3 or more standard deviations greater than the mean underwent re-extraction, as such a recovery was deemed suboptimal. Following re-extraction, only the extracted DNA sample producing a positive IAC result within the defined Cq range (not the previously conducted invalid extract) was used for experimental testing.

Experimental qPCR reactions were performed using previously published multi-parallel assays targeting non-coding repetitive sequences to detect *Ascaris lumbricoides* [38], *Trichuris trichiura*, *Strongyloides stercoralis*, *Necator americanus*, *Ancylostoma duodenale* [26] and *Ancylostoma ceylanicum* [39] (Table 5.2). Laboratory staff were blinded to Kato-Katz results for each sample during the initial qPCR analyses. However, upon completion of initial testing, all samples determined to be Kato-Katz positive but qPCR negative for *A. lumbricoides* were tested using a second qPCR assay targeting an unrelated, ribosomal DNA sequence in order to validate the negative qPCR result and confirm the absence of *A. lumbricoides* DNA [26, 40].

For each assay, all samples were tested in replicate reactions and samples were classified as positive when amplification occurred in both reactions with a Cq value of $< 40$, consistent with prior studies [41]. Samples which produced Cq values $\geq 40$ in both

replicate reactions or failed to amplify in both replicate reactions were classified as negative. Consistent with common practice [42], samples that were positive in one of two replicate reactions were re-tested in duplicate and classified as positive if the Cq value was < 40 in at least one of two re-test reactions. In all cases of re-testing, the Cq value reported as the sample's final result was the re-test Cq value.

In addition to the IAC control, all experimental qPCR reaction plates were accompanied by the testing of a titration of the appropriate assay's control plasmid. Plasmid controls for each target sequence were prepared as previously described [38]. Each plasmid contained a single copy of the corresponding assay's target sequence and 20 pg, 200 fg and 2 fg of plasmid were tested in duplicate reactions. Following the completion of all testing, the mean Cq value, across all experimental plates, for each plasmid concentration was determined. Any plate which produced a Cq value for any concentration of plasmid that was 3 or more standard deviations greater than the mean calculated for all plates was retested in its entirety, and all results from that plate were considered to be invalid. Only results from valid plates were reported (Table 5.3).

### 5.5.5 Sequencing

To verify that Kato-Katz positive/qPCR negative samples were truly qPCR negative for *A. lumbricoides*, we performed sequencing on a subset of samples. Because < 1% of samples tested negative for hookworm or *T. trichiura* using qPCR but positive using Kato-Katz, we only performed sequencing for samples with discordant *A. lumbricoides* classification. A subset of 10 samples was selected to undergo amplicon sequencing-based analysis. To facilitate the selection of samples, all samples that were double-slide Kato-

Katz positive for the presence of *A. lumbricoides* but qPCR negative for *A. lumbricoides* were identified. From this list, a convenience sample (n=7) was then selected, such that two of these samples would contain epg counts characteristic of moderate-intensity infections as determined by World Health Organization guidelines, while five samples would contain egg counts that were characteristic of light intensity infections [43]. All seven of the Kato-Katz positive samples were negative by qPCR using both *A. lumbricoides* assays. Additionally, three Kato-Katz negative samples were chosen for inclusion in this panel. Two of these samples were selected due to their status as *A. lumbricoides*-positive as determined by qPCR analysis, while a single sample was selected that was both Kato-Katz negative and qPCR negative to serve as an uninfected control.

Samples were prepared for sequencing using a modified version of the Earth Microbiome Project's 18S Illumina Amplicon Protocol available at: http://www.earthmicrobiome.org/protocols-and-standards/18s/. This protocol utilizes primers targeting the variable region 9 (V9) of the eukaryotic small subunit (SSU) rDNA [44]. This protocol has been used in the analysis of helminth biodiversity in the gut of rats [45] and validated as a method for assessing global parasite diversity [46]. Briefly, the targets within each sample were amplified in triplicate 25 µl reactions using a uniquely barcoded reverse primer coupled with a conserved forward primer (Table 5.4). A "mammalian blocking" primer, based on a previously described strategy, was also included in all reactions to minimize the prevalence of mammal-derived amplicons [47]. All reactions contained 10 µl of Platinum Hot Start PCR Master Mix (2X) (ThermoFisher, Waltham, MA), 0.2 µM forward and reverse primers, and 1.6 µM "mammalian blocking" primer (Integrated DNA Technologies, Coralville, IA). Cycling consisted of an initial

denaturation step of 94 °C for 3 minutes, followed by 35 cycles of 94 °C for 45 seconds, 65 °C for 15 seconds, 57 °C for 30 seconds, and 72 °C for 90 seconds. Samples then underwent a final extension step at 72 °C for 10 minutes. Following amplification, triplicate reactions were combined, and pooled products were run on a 1.5% agarose gel to confirm the presence of a band of the expected size (~260 bp). 240 ng of each pooled product was then combined in preparation for sequencing, and this library was purified using the ZR-96 Clean & Concentrator purification kit (Zymo Research, Irvine, CA). The purified library was diluted to a 9 pM concentration, and 30% PhiX was added to improve diversity. Sequencing then occurred on the Illumina MiSeq platform, using a MiSeq Reagent Kit v2 (300-cycles) (Illumina, San Diego, CA).

Sequencing results were analyzed using the Quantitative Insights into Microbial Ecology 2 (QIIME2) pipeline [48]. Briefly, all interlaced read pairs which did not contain a valid barcode were filtered out and all read pairs whose best match was to bacterial 16S ribosomal sequence were also removed. This resulted in a list of reads exclusively matching eukaryotic 18S sequence. Reads with 99% or greater identity were then classified into operational taxonomic units (OTUs) and consensus sequences were built. Taxonomy was then assigned to all OTUs that had ≥ 95% identity to sequences within the Silva release 132 QIIME-compatible database.

### 5.5.6 Statistical methods

All statistical analyses were pooled across intervention arms to ensure sufficient statistical power. Unless otherwise specified, analyses were conducted in R version 3.4.3.

Data and replication scripts are available through the Open Science Framework (https://osf.io/agk6w/).

### 5.5.7 Assessment of quality of Kato-Katz slide readings

For the 2,800 samples included in this analysis, we estimated concordance between individual Kato-Katz technicians and expert technicians by comparing the classification of positive/negative status for each STH on each slide as well as the mean difference in eggs counted per slide.

### 5.5.8 Prevalence and correlation between eggs per gram and Cq values

We estimated the prevalence of STH infection using each diagnostic method with robust sandwich standard errors to account for clustering at the village level [49]. We calculated the agreement between prevalence estimates from double-slide Kato-Katz and qPCR using a kappa statistic. We tested whether prevalence differed between the two diagnostic methods using the Wilcoxon matched-pairs signed rank test that conditioned on randomized block in the original trial [32, 50]. To classify infection intensity, we used the World Health Organization cutoffs (light intensity infections were defined as < 5,000 epg for *A. lumbricoides*, < 1,000 epg for *T. trichiura*, and < 2,000 epg for hookworm; moderate intensity infections were defined as 5,000 ≤ epg < 50,000 for *A. lumbricoides*, 1,000 ≤ epg < 10,000 for *T. trichiura*, and 2,000 ≤ epg < 4,000 for hookworm; heavy intensity infections were defined as ≥ 50,000 epg for *A. lumbricoides*, ≥ 10,000 epg for *T. trichiura*, and ≥ 4,000 epg for hookworm) [43]. To assess the correlation between epg estimated by double-slide Kato-Katz and Cq value estimated by qPCR, we estimated Kendall's tau and

calculated two-sided p-values using a bootstrap with 1,000 replicates that resampled blocks to account for the study's cluster-randomized design using Stata version 14.2. We compared mean Cq values within levels of infection intensity using a paired t-test paired within randomized blocks.

### 5.5.9 Estimation of sensitivity and specificity

A challenge in studies assessing accuracy of STH diagnostics is the lack of a gold standard diagnostic method [51]. We thus used two different approaches to estimate sensitivity and specificity of each method, consistent with prior studies. We created a reference using both detection methods and defined positivity as qPCR Cq values < 40 or at least one egg detected by double-slide Kato-Katz. We also used Bayesian latent class analysis, which defines the true prevalence of STH infection, sensitivity, and specificity as latent variables that are estimated simultaneously from the data and assumes no gold standard [22, 28, 52–54]. We hypothesized that the number of STH worms in the intestines influences STH DNA concentration and epg in stool, each of which determines observed STH infection measured by qPCR and Kato-Katz. Because observed STH status measured by qPCR and Kato-Katz each depends upon true worm burden, we defined a model that allowed for conditional dependence between each test by including terms for covariance between diagnostic methods since both methods are dependent upon the same underlying biological process [52].

Consistent with prior studies, our model did not include separate latent variables for worm burden, DNA concentration and epg, which were unobserved, because the model would have been non-identifiable. Latent class models for assessing diagnostic accuracy

of two diagnostic tests in the absence of a gold standard are not identifiable because the number of unknown parameters exceeds the degrees of freedom [55]; in this case, multiple unique parameter estimates can be produced from the same model, and a single set of estimates cannot be identified. Within a Bayesian framework, when informative prior distributions are defined for at least two parameters, the model is identifiable [56]. Consistent with prior studies, we used a highly informative prior for specificity of each diagnostic test to allow for identifiability [22, 28, 57, 58] since prior studies have reported high specificity values for double-slide Kato-Katz [59, 60] and qPCR [25, 61] (Table 5.5). For *A. lumbricoides*, we chose to specify a non-informative prior for double-slide Kato-Katz sensitivity and specificity and a less informative prior for qPCR sensitivity because of discordant test results (summarized below). Since the model's parameter estimates are sensitive to the assumed prior distributions [62], we conducted a sensitivity analysis using alternative, more informative priors for *A. lumbricoides* (Table 5.6). To estimate parameters, we used Monte Carlo simulation with 4 chains and 100,000 iterations, recording every 10th result. We computed the mean and 2.5 and 97.5 percentiles of the posterior distribution using Gibbs sampling in WinBUGS version 14 [63].

### 5.5.10 Ethical statement

The study protocol was approved by the Ethical Review Committee at icddr,b (PR-14105), the Committee for the Protection of Human Subjects at the University of California, Berkeley (2014-08-6658), and the institutional review board at Stanford University (27864). Prior to enrollment, all adult subjects provided written informed

consent. Parents or guardians of children provided written informed consent on behalf of children.

### 5.6 Results

We analyzed stool samples from 2,799 children using both double-slide Kato-Katz and qPCR. 51% of children were female, and 66% of children's caregivers reported that they had received deworming in the past 6 months (Figure 5.1). Concordance between individual Kato-Katz technicians and expert parasitologists in the classification of STH positive/negative status was close to 100% for all technicians and STH (Figure 5.2). For *A. lumbricoides* and *T. trichiura*, kappa statistics comparing individual technicians and expert parasitologists were high, and the kappa statistic for hookworm was high when the expert parasitologist performed Kato-Katz on the same day as the original count (Table 5.7). The difference in mean eggs per slide between individual Kato-Katz technicians and experts was < 60 for *A. lumbricoides*, < 15 for hookworm, and < 20 for *T. trichiura* (Figure 5.3). We excluded the qPCR results of one sample from statistical analyses for all species because the internal amplification control was > 3 standard deviations from the mean upon initial testing and following re-extraction.

The observed prevalence was higher for qPCR than double-slide Kato-Katz for hookworm and *T. trichiura* but lower for *A. lumbricoides* (Table 5.8, Figure 5.4). The observed prevalence of *A. lumbricoides* was 37.0% (95% CI 34.7%, 39.3%) using Kato-Katz and 23.3% (95% CI 20.9%, 25.7%) using qPCR. The observed prevalence of hookworm was 7.5% (95% CI 6.3%, 8.7%) using Kato-Katz. Using qPCR the observed prevalence of *Necator americanus* was 18.4% (95% CI 16.5%, 20.3%), and the observed

prevalence of *Ancylostoma ceylanicum* was 3.8% (95% CI 2.9%, 4.6%); only one out of 2,799 samples tested positive for *Ancylostoma duodenale*. The observed prevalence of any hookworm species using qPCR was 21.4% (95% CI 19.4%, 23.3%). For *T. trichiura*, the observed prevalence was 7.0% (95% 5.6%, 8.3%) by Kato-Katz and 12.3% (95% CI 10.3%, 14.2%) by qPCR. Using Kato-Katz, 7% of children were infected with more than one STH, and using qPCR it was 13% (Figure 5.5). The observed prevalence of moderate intensity infection using Kato-Katz was 11% for *A. lumbricoides*, less than 1% for hookworm, and 5% for *T. trichiura*. The observed prevalence of heavy intensity infection using Kato-Katz was < 1% for *A. lumbricoides*, 1% for hookworm, and 0% for *T. trichiura*. Using Kato-Katz, the geometric mean of epg was 5.14 for *A. lumbricoides* (134 among positives), 0.43 for hookworm (125 among positives), and 0.40 for *T. trichiura* (124 among positives).

Concordance between double-slide Kato-Katz and qPCR was higher for hookworm and *T. trichiura* than for *A. lumbricoides*. The p-value for permutation tests assessing whether the observed prevalence estimated by the two detection methods differed was <0.001 for each STH (Table 5.9). 6-14% of samples classified as negative for the different species by Kato-Katz were positive using qPCR. 26% of samples classified as positive for *A. lumbricoides* by Kato-Katz were negative using qPCR. All samples determined to be Kato-Katz positive for *A. lumbricoides* but negative by qPCR analysis targeting the non-coding repeat also failed to produce a positive result when tested with the ITS-targeting assay. The majority of epg values among *A. lumbricoides* Kato-Katz-positive, qPCR-negative samples were < 100, although a small number of samples had epg between 1,000 and 50,000 (Figure 5.6). The probability that an individual Kato-Katz technician classified

a stool sample as *A. lumbricoides* positive among those classified as negative by qPCR ranged from 20% to 31%, and four out of six technicians were significantly more likely to misclassify a sample than the technician with the lowest misclassification rate (Figure 5.7). There was no evidence that the proportion of samples classified as positive for *A. lumbricoides* by Kato-Katz and negative by qPCR (false positives assuming qPCR was the gold standard) followed a trend by date of Kato-Katz or date of DNA extraction (Figure 5.8, Figure 5.9); patterns by date were similar for false negatives (Figure 5.10, Figure 5.11).

Using Bayesian latent class models, the estimated true prevalence was 27% (95% Bayesian Credible Interval (BCI) 20%, 37%) for *A. lumbricoides*, 20% (95% BCI 17%, 24%) for hookworm, and 11% (95% BCI 8%, 14%) for *T. trichiura*. For double-slide Kato-Katz, the sensitivity was 49% (95% BCI 34%, 64%) for *A. lumbricoides*, 32% (95% BCI 22%, 41%) for hookworm, and 52% (95% BCI 33%, 71%) for *T. trichiura*, and the specificity was 68% (95% BCI 61%, 77%) for *A. lumbricoides*, 99% (95% BCI 96%, 100%) for hookworm, and 98% (95% BCI 96%, 100%) for *T. trichiura* (Table 5.10). For qPCR, the sensitivity was 79% (95% BCI 61%, 99%) for *A. lumbricoides*, 93% (95% BCI 82%, 100%) for hookworm, and 90% (95% BCI 81%, 100%) for *T. trichiura*, and the specificity was 97% (95% BCI 95%, 100%) for all three STH. The sensitivity analysis for *A. lumbricoides* using alternative more informative priors produced similar estimates for Kato-Katz and higher sensitivity for qPCR; when more informative priors were used for both Kato-Katz and qPCR, the estimated sensitivity of qPCR was 90 (95% BCI 80, 99) (Table 5.11). Pooling both methods as the gold standard, the sensitivity and specificity of Kato-Katz was similar to those estimated from the Bayesian model, and for qPCR the

sensitivity of hookworm and *T. trichiura* was higher but the specificity was lower than estimated by the Bayesian model.

The distribution of Cq values for *A. lumbricoides* was bimodal, and to a lesser extent the same pattern was present for *N. americanus* and *T. trichiura* (Figure 5.12). For each STH, there was a negative, monotonic relationship between qPCR Cq values and epg estimated using double-slide Kato-Katz (Figure 5.13). The Kendall's tau comparing these two quantities was -0.442 for *A. lumbricoides*, -0.346 for *N. americanus*, -0.266 for *A. ceylanicum*, and -0.248 for *T. trichiura* (for each, the p-value was <0.001); we compared *N. americanus* and *A. ceylanicum* to any hookworm species detected by Kato-Katz. We also examined the distribution of Cq values by Kato-Katz infection intensity status for *A. lumbricoides* (very few children had moderate-to-heavy intensity hookworm or *T. trichiura* infections using Kato-Katz) (Figure 5.14). The distribution of Cq values was higher for children who were Kato-Katz negative and children with light intensity infections compared to those with moderate-to-heavy infections. The median Cq value was 9 (range: 4-34) among children with moderate to heavy *A. lumbricoides* infection, 14 (range: 8-38) among those with light intensity infection, and 27 (range: 9-40) among those who were Kato-Katz negative (t-test p-value < 0.001 for all comparisons).

In the sequencing analyses of 10 samples, we detected *Enterobius vermicularis* in one sample that was double-slide Kato-Katz negative and qPCR positive, one sample that was Kato-Katz positive and qPCR negative, and one sample that was negative using both methods (Table 5.12). In addition, we identified *Giardia intestinalis* in 3 samples that were Kato-Katz positive and qPCR negative and one sample that was negative using both methods; we detected *Dientamoeba fragilis* in 3 samples that were Kato-Katz positive and

qPCR negative, one sample that was Kato-Katz negative and qPCR positive, and one sample that was negative using both methods. Analysis using the QIIME2 pipeline failed to identify any reads mapping to *A. lumbricoides* in all seven Kato-Katz positive, qPCR negative samples. *A. lumbricoides* was detected in one of the two samples that were Kato-Katz negative and qPCR positive. However, this analysis failed to identify *A. lumbricoides* in the second sample. Because of this discrepancy, the *A. lumbricoides*-derived consensus sequence of the reads from the *A. lumbricoides*-positive sample was compared, using BLAST, to all other reads from the entire sample set. This resulted in the identification of a small number of *A. lumbricoides*-derived reads ($n = 5$, identity $\geq 97\%$) within the second Kato-Katz negative, qPCR positive sample, while no *A. lumbricoides* reads were found within the pool of reads mapping to any of the other samples. As only a single read within this Kato-Katz negative, qPCR positive sample had 100% identity with the *A. lumbricoides*-derived consensus sequence used as the query sequence in the BLAST analysis, QIIME2's failure to generate an OTU from these reads follows logically, as it is likely that the 99% similarity threshold was not met for any two reads mapping to *A. lumbricoides* from this sample. However, the presence of *A. lumbricoides*-derived reads provides further support for the existence of *A. lumbricoides* DNA within the sample and speaks to the sensitivity of the qPCR assays used during the analysis.

### 5.7 Discussion

This study compared the performance of double-slide Kato-Katz and qPCR for detecting STH infections in a large sample of 2,800 children in a setting in rural Bangladesh with biannual mass deworming administration. Consistent with prior studies in low STH

infection intensity settings, using qPCR led to substantial increases in hookworm detection and moderate increases in *T. trichiura* detection compared to Kato-Katz [23, 25, 27, 28, 51]. Poorer performance of Kato-Katz, particularly for hookworm, is likely due to rapid disintegration of hookworm ova following stool collection [51, 64]. However, unexpectedly, we found that one third of samples classified as *A. lumbricoides* negative by qPCR were classified as positive by Kato-Katz, and the specificity of Kato-Katz was lower for *A. lumbricoides* than for hookworm or *T. trichiura*. This discordance in classification occurred primarily among samples with epg between 10 and 100, but surprisingly, a small number of samples with *A. lumbricoides* epg between 1,000 and 50,000 classified as negative by qPCR (Figure 5.6). Experimental testing with two qPCR assays utilizing independent targets as well as sequencing analysis of a subset of samples confirmed that samples classified as *A. lumbricoides* negative by qPCR contained no genomic targets for *A. lumbricoides*. These unexpected findings call into question the common assumption that Kato-Katz yields few false positives [22, 28, 57, 58].

There are several potential explanations for the unexpected discrepancy between Kato-Katz and qPCR for *A. lumbricoides*. First, poor quality control of Kato-Katz could have caused the discrepancy. The quality control assessment found minimal differences between classification of *A. lumbricoides* between individual laboratory technicians and expert parasitologists (Figure 5.2, Figure 5.3, Table 5.7), suggesting that Kato-Katz was performed with typical quality. However, the expert parasitologist that performed quality assurance also conducted Kato-Katz training; thus, if the expert mistook regularly misclassified substances in stool as *A. lumbricoides*, misclassification may have been passed on to technicians during training and would not have been caught during quality

assurance. Second, the discrepancy could have resulted from mislabeling of stool samples. While we cannot rule out this possibility, we consider this highly unlikely; if stool samples were mislabeled, we would expect to see discrepancies for hookworm and *T. trichiura* as well. There was no association between the false positive rate and false negative rate (using qPCR as the gold standard) by date of Kato-Katz or date of DNA extraction, which suggests that mislabeling or other laboratory protocol violations were unlikely since both may have caused an unexpected increase in false positives/negatives at a specific point in time (Figure 5.8 – 5.11). Third, it is possible that improper sample storage could have resulted in STH DNA degradation that led to our discrepant findings for *A. lumbricoides*, however, if that was the case we would expect to also see discrepancies for other STH and in particular for hookworm, which is more fragile and is known to degrade more rapidly than *A. lumbricoides*. Finally, it is possible that the discrepancy was due to insufficient DNA extraction, but if this was the cause we would expect to see similarly poor performance for all qPCR assays relative to Kato-Katz results. However, qPCR had higher sensitivity than Kato-Katz for hookworm and *T. trichiura*. Taken together, these findings suggest that there is an alternative explanation for the *A. lumbricoides* discrepancy.

Misclassification of STH ova has been previously reported: *Capillaria philippinensis* or *Capillaria hepatica* ova can be mistaken for *A. lumbricoides* or *T. trichiura* ova [65, 66] and *Trichostrongylus* spp. or *Schistosoma* ova can resemble hookworm ova [67, 68]. Though we did identify other enteropathogens (*Giardia intestinalis* and *Dientamoeba fragilis*) when sequencing a subset of samples, in stool, these parasites do not share morphological features with most stages of *A. lumbricoides* [69]. We hypothesize that the substance mistaken for *A. lumbricoides* in Kato-Katz was most likely

plant material, pollen grains, or fungal spores, all of which can resemble certain stages of *A. lumbricoides* [70]. Some of the sequenced stool samples contained DNA of flowering plants of the division *Magnoliophyta*, which produce pollen, and fungi of the genera *Saccharomyces* and *Candida*, which are commonly found in healthy intestinal microbiota [71] (Table 5.12). We cannot definitively identify whether a different organism or substance was mistaken for *A. lumbricoides* because the original slides used in Kato-Katz were not preserved. Nevertheless, our findings raise concerns about the quality of the Kato-Katz procedures and underscore the advantage of using molecular methods over copromicroscopy since molecular methods do not rely on visualization of stool contents.

Overall, we found that sensitivity was lower for both double-slide Kato-Katz and qPCR than in prior studies [22, 28]. A challenge in quantifying the accuracy of STH diagnostics is the lack of a gold standard measure of infection [10]. Though we used Bayesian latent class analysis, which assumes no gold standard, our estimates of sensitivity and specificity from this method should be interpreted with caution. In studies evaluating fewer than four diagnostic tests, parameters are not identifiable, and as a result parameter estimates are highly dependent upon the assumed prior distributions [52]. Our latent class models used informative priors for the specificity of each diagnostic, which we assumed was between 0.95 and 1.00 for hookworm and *T. trichiura*; for these STH there was agreement between results from the Bayesian latent class analysis and an analysis pooling both methods as the reference, which suggests that our assumption was appropriate. However, because of the apparent false positives for *A. lumbricoides* using Kato-Katz, our prior distribution for the specificity of Kato-Katz ranged from 0 to 1. Using this model with less informative priors, our estimate of sensitivity of qPCR for detecting *A. lumbricoides*

using Bayesian latent class analysis was lower than a previous estimate for multi-parallel qPCR using the same method [28]. Using less informative priors for *A. lumbricoides* may have reduced the identifiability of our model and contributed to differing sensitivity and specificity estimates in this study. However, our sensitivity analysis using more informative priors for *A. lumbricoides* produced estimates that were similar to our primary analysis and/or closer to previously published estimates (Table 5.6, Table 5.11). Our model also assumed that the sensitivity and specificity of each diagnostic was constant for each study subject [72]. It is possible that sensitivity and specificity varied due to variation in STH prevalence and infection intensity, differing laboratory technicians, time from stool collection to Kato-Katz or stool archiving, or other factors; however, we expect such heterogeneity to be relatively low because our study used standardized procedures and was in a relatively small geographic area.

We observed a bimodal distribution of Cq values for *A. lumbricoides* and to a lesser extent for *N. americanus* and *T. trichiura*. While we cannot definitively identify the cause of this pattern, we propose two potential hypotheses. First, it is possible that the lower mode of Cq values (near Cq=12) reflects target sequences from intact eggs, while the higher mode (near Cq=32) reflects the presence of target sequences from samples with sub-egg quantities (i.e. sub-genome quantities) of DNA (Figure 5.12). While the underlying cause was not addressed, similar distributions have been reported in prior studies [24]. Furthermore, an egg spiking study utilizing the same repeat-targeting assay employed in this study and employing the same controls and instrumentation used in this study, determined that on average, DNA extraction from a sample containing a single *A. lumbricoides* egg gives a Cq value of approximately 24, with the most efficient extractions

commonly producing Cq values in the range of 19-22 [38]. This implies that efficiently extracted sub-egg quantities of target DNA should result in Cq values greater than 22. At nearly 250,000 copies per haploid genome, the genomic target of the employed *Ascaris* assay is more numerous than the repetitive sequences utilized as targets in any of the other assays. As such, its ability to detect trace amounts of genomic DNA is likely greater, providing an explanation for the more pronounced secondary distribution with higher Cq values. Additional research is needed to determine whether individuals with sub-egg levels of STH DNA in their stool experience clinical symptoms of infection and/or contribute to STH transmission.

A second possible explanation for the bimodal Cq distribution is that the lower mode represents fertilized eggs with diploid genomes, and the higher mode represents unfertilized eggs shed from female-only *A. lumbricoides* infections with haploid genomes. Unfertilized adult female *A. lumbricoides* worms may still shed large numbers of single-genome eggs. However, fertilized eggs undergo multiple cellular divisions during the process of embryonation, meaning that a single egg may contain many copies of the genome. Typically, under permissive environmental conditions, embryonation occurs over the course of approximately 2-3 weeks following fecal shedding, generally requiring approximately 5 days for eggs to reach the 8-cell stage [73]. When stored at colder temperatures approaching 5 °C, such development is arrested [73]. Given that all stool samples collected as part of this study were placed on ice within 18 hours of sample production, it is extremely unlikely that significant embryonic development could have occurred prior to DNA extraction and qPCR testing. Therefore, while fertilized eggs would contain diploid genomes it is unlikely that significant cell division would have occurred.

As such, in this study, any effect on Cq value resulting from fertilization status is likely to have been minimal, making this unlikely to account for the presence of the observed bimodal distribution.

Whether Cq values can be used to approximate intensity of STH infection is an open question. The strength of the correlation between epg and Cq values informs whether Cq value thresholds can be defined for STH infection intensity as is done using Kato-Katz [74]. Prior studies have reported moderate-to-high correlations between epg and DNA concentration for all three STH [23–25, 28, 75], and one study reported that Cq values correlated well with expelled adult worms following anthelminthic treatment [28]. We found moderate correlations between epg and Cq values that were lower than the correlations reported in prior studies [24, 25, 28]. *A. lumbricoides* was the only STH in this study with a sufficiently large number of samples classified as moderate-to-heavy intensity using Kato-Katz. The majority of moderate-to-heavy intensity infections had Cq values of 10 or lower (Figure 5.14). However, because of the misclassification of *A. lumbricoides*, we interpret this pattern with caution because epg counts may have been overestimated in some samples and underestimated in others. In general, a limitation of using Cq values to approximate infection intensity is that the quantity of cells per egg (and thus the copy number of a target) varies depending on fertilization status, cell number, diminution status, and the STH species [76], complicating inferences about epg using Cq values. When comparing Cq values to epg estimated from Kato-Katz, another limitation is that there is variability in the number of STH eggs secreted in stool, the number of haploid genomes present per egg, and eggs are not distributed evenly within stool or in stool samples [77, 78]. Despite these limitations, Cq values represent the only quantitative output from a

qPCR reaction, and as such, they may represent the best (albeit imperfect) measure for approximating infection intensity.

In settings with predominantly low STH infection intensity, highly sensitive diagnostics are needed to detect low intensity infections prior to elimination and to detect resurgent infections. Though a common critique of qPCR is its higher cost relative to Kato-Katz, comparisons of the materials required for each method suggest that costs for multi-parallel qPCR can be lower than those for Kato-Katz [25]. As pointed out by Turner et al., even if qPCR is more costly than Kato-Katz, in an elimination setting, the continued use of low sensitivity diagnostics may impede efforts to determine when STH transmission has been interrupted which may unnecessarily prolong mass deworming administration. The cost of prolonged MDA is likely to exceed that of a more costly diagnostic [79]. Our results support the use of qPCR in elimination settings due to its higher sensitivity and specificity. In addition, importantly, using qPCR instead of Kato-Katz in intervention trials can also reduce misclassification and bias of intervention effect estimates [80].

### 5.8 Conclusion

Kato-Katz remains a highly cost-effective, feasible diagnostic method appropriate for certain uses cases, such as identifying where to perform MDA, determining the frequency of MDA, and assessing progress towards program goals [30]. Prior to this study, the greatest disadvantage of Kato-Katz was typically considered to be its low sensitivity; here we show that not only was sensitivity lower for Kato-Katz than for qPCR, but also in our setting, Kato-Katz had low specificity for *A. lumbricoides*. Though we were not able to definitively determine the cause of the apparent *A. lumbricoides* misclassification using

Kato-Katz, regardless of the cause, our results highlight an inherent limitation of Kato-Katz. In addition, our results echo conclusions of other recent studies that qPCR may be more appropriate for use cases such as identifying whether STH transmission has been interrupted and confirming sustained transmission interruption due to its higher sensitivity in low STH intensity settings [30].

**Table 5.1 Percent agreement between results as determined by the original Kato-Katz technician and the senior parasitologist by days elapsed between each assessment.**

| Days | *Ascaris lumbricoides* | Hookworm | *Trichuris trichiura* |
|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 |
| 1 | 0.91 | 0.33 | 0.81 |
| 2-4 | 0.94 | 0.11 | 0.91 |
| >4 | 0.88 | 0.00 | 0.83 |
| Overall | 0.92 | 0.20 | 0.86 |

**Table 5.2 Target information, sequences, manufacturers, and relevant references for all qPCR primers and probes used during this study.**

| Target | Forward Primer | | Reverse Primer | | Probe | | Relevant Reference |
|---|---|---|---|---|---|---|---|
| | Sequence (5' → 3') | Manufacturer | Sequence (5' → 3') | Manufacturer | Sequence (5' → 3') | Manufacturer | |
| ***Ancylostoma ceylanicum* Repetitive Sequence** | CAAATATTAC TGTGCGCATT TAGC | IDT DNA | GCGAATATTT AGTGGGTTTA CTGG | IDT DNA | /56-FAM/CGGTGA AAG/ZEN/CTT TGCGTTATTG CGA/3IABkFQ/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Ancylostoma duodenale* Repetitive Sequence** | GTATTTCACT CATATGATCG AGTGTTC | IDT DNA | GTTTGAATTT GAGGTATTTC GACCA | IDT DNA | /56-FAM/TGACAG TGT/ZEN/GTC ATACTGTGGA AA/3IABkFQ/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Necator americanus* Repetitive Sequence** | CCAGAATCG CCACAAATTG TAT | IDT DNA | GGGTTTGAG GCTTATCATA AAGAA | IDT DNA | /56-FAM/CCCGAT TTG/ZEN/AGC TGAATTGTCA AA/3IABkFQ/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Strongyloides stercoralis* Repetitive Sequence** | CGCTCCAGA ATTAGTTCCA GTT | IDT DNA | GCAGCTTAGT CGAAAGCAT AGA | IDT DNA | /56-FAM/ACAGTC TCC/ZEN/AGT TCACTCCAGA AGAGT/3IABk FQ/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Trichuris trichiura* Repetitive Sequence** | GGCGTAGAG GAGCGATTT | IDT DNA | TACTACCCAT CACACATTAG CC | IDT DNA | /56-FAM/TTTGCG GGC/ZEN/GAG AACGGAAAT ATT/3IABkFQ/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Ascaris lumbricoides* Repetitive Sequence** | CTTGTACCAC GATAAAGGG CAT | IDT DNA | TCCCTTCCAA TTGATCATCG AATAA | IDT DNA | /56-FAM/TCTGTG CAT/ZEN/TAT TGCTGCAATT GGGA/3IABkF Q/ | IDT DNA | Pilotte, et al. PLoS Negl Trop Dis. (2016) |
| ***Ascaris lumbricoides* ITS Sequence** | GTAATAGCA GTCGGCGGTT TCTT | IDT DNA | GCCCAACAT GCCACCTATT C | IDT DNA | /56-FAM/TTGGCG GACAATTGC ATGCGAT/MB G/ | ThermoFisher | Pilotte, et al. PLoS Negl Trop Dis. (2018) |
| **Internal amplification control** | CTAACCTTCG TGATGAGCA ATCG | IDT DNA | GATCAGCTAC GTGAGGTCCT AC | IDT DNA | /56-FAM/AGCTAG TCG/ZEN/ATG CACTCCAGTC CTCCT/3IABkF Q/ | IDT DNA | Deer, et al. Lett Appl Microbiol. (2010) |

**Table 5.3 Comparison of multi-parallel qPCR and double-slide Kato-Katz for detection of soil-transmitted helminth infection among children in rural Bangladesh.**

| Species | Control Conc. | Mean | Standard Deviation | +/- 3SD | Total Plates Tested | Total Plates Requiring Retests |
|---|---|---|---|---|---|---|
| *Ancylostoma ceylanicum* | 20 pg | 10.57428624 | 0.629269472 | 9.315747 - 11.83283 | 66 | 6 |
| | 200 fg | 16.76210483 | 0.89796325 | 14.96618 - 18.55803 | | |
| | 2 fg | 23.58155042 | 1.429190388 | 20.72317 - 26.43993 | | |
| *Necator americanus* | 20 pg | 14.90265768 | 1.546944972 | 11.80877 - 17.99655 | 66 | 6 |
| | 200 fg | 21.31265956 | 1.794560709 | 17.72354 - 24.90178 | | |
| | 2 fg | 27.7174806 | 1.0135431 | 25.69039 - 29.74457 | | |
| *Ancylostoma duodenale* | 20 pg | 11.58566395 | 2.354262378 | 6.877139 - 16.29419 | 66 | 7 |
| | 200 fg | 17.9120955 | 2.416993482 | 13.07811 - 22.74608 | | |
| | 2 fg | 24.6325359 | 2.150720837 | 20.33109 - 28.93398 | | |
| *Trichuris trichiura* | 20 pg | 14.54309587 | 1.10072496 | 12.34165 - 16.74455 | 66 | 5 |
| | 200 fg | 21.70388865 | 2.280166759 | 17.14356 - 26.26422 | | |
| | 2 fg | 28.74177976 | 1.995048103 | 24.75168 - 32.73188 | | |
| *Ascaris lumbricoides* | 20 pg | 11.35916323 | 1.225636444 | 8.90789 - 13.81004 | 66 | 5 |
| | 200 fg | 17.79485853 | 1.452208619 | 14.89044 - 20.69928 | | |
| | 2 fg | 24.59787477 | 1.855364044 | 20.88715 - 28.3086 | | |
| *Strongyloides stercoralis* | 20 pg | 15.34083837 | 0.84766427 | 13.64551 - 17.03617 | 66 | 3 |
| | 200 fg | 21.70180233 | 0.876749866 | 19.9483 - 23.4553 | | |
| | 2 fg | 28.44642298 | 1.353544442 | 25.73933 - 31.15351 | | |

**Table 5.4 Primer sequences used for 18S SSU sequencing.**

| Primer Name | Primer Sequence (5′ → 3′) [a] |
|---|---|
| Read 1 Sequencing Primer | TATCGCCGTTCGGTACACACCGCCCGTC |
| Read 2 Sequencing Primer | AGTCAGCCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Index Sequence Primer | GTAGGTGAACCTGCAGAAGGATCATGCTGACTGACT |
| | |
| Forward Primer for PCR | AATGATACGGCGACCACCGAGATCTACACTATCGCCGTTCGGTACACACCGCCCGTC |
| | |
| Reverse Primer for PCR #1 | CAAGCAGAAGACGGCATACGAGATACGAGACTGATTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #2 | CAAGCAGAAGACGGCATACGAGATGCTGTACGGATTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #3 | CAAGCAGAAGACGGCATACGAGATATCACCAGGTGTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #4 | CAAGCAGAAGACGGCATACGAGATTGGTCAACGATAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #5 | CAAGCAGAAGACGGCATACGAGATATCGCACAGTAAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #6 | CAAGCAGAAGACGGCATACGAGATGTCGTGTAGCCTAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #7 | CAAGCAGAAGACGGCATACGAGATAGCGGAGGTTAGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #8 | CAAGCAGAAGACGGCATACGAGATTACAGCGCATACAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #9 | CAAGCAGAAGACGGCATACGAGATAATTGTGTCGGAAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| Reverse Primer for PCR #10 | CAAGCAGAAGACGGCATACGAGATTGCATACACTGGAGTCAGTCAGCATGATCCTTCTGCAGGTTCACCTAC |
| | |
| Mammal Blocking Primer | GCCCGTCGCTACTACCGATTGG/ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/TTAGTGAGGCCCT/3SpC3/ |

[a] The barcode sequence in each reverse primer is underlined.

**Table 5.5 Prior distributions used in Bayesian latent class analysis models.**

| | ***A. lumbricoides*** | **Hookworm** | ***T. trichiura*** |
|---|---|---|---|
| Prevalence | Beta distribution with shape parameters $\alpha$=1, $\beta$=1 | Beta distribution with shape parameters $\alpha$=1, $\beta$=1 | Beta distribution with shape parameters $\alpha$=1, $\beta$=1 |
| Kato-Katz sensitivity | Beta distribution with shape parameters $\alpha$=1, $\beta$=1 | Beta distribution with shape parameters $\alpha$=3, $\beta$=3 | Beta distribution with shape parameters $\alpha$=3, $\beta$=3 |
| Kato-Katz specificity | Beta distribution with shape parameters $\alpha$=1, $\beta$=1 | Uniform distribution with minimum=0.95, maximum=1 | Uniform distribution with minimum=0.95, maximum=1 |
| qPCR sensitivity | Uniform distribution with minimum=0.60, maximum=1 | Uniform distribution with minimum=0.80, maximum=1 | Uniform distribution with minimum=0.80, maximum=1 |
| qPCR specificity | Uniform distribution with minimum=0.95, maximum=1 | Uniform distribution with minimum=0.95, maximum=1 | Uniform distribution with minimum=0.95, maximum=1 |

**Table 5.6 Sensitivity analysis using alternative prior distributions in Bayesian latent class analysis models for *A. lumbricoides*.**

| | | Kato-Katz | | qPCR | |
|---|---|---|---|---|---|
| **Analysis** | **Description of priors** | **Sensitivity (95% BCI)** | **Specificity (95% BCI)** | **Sensitivity (95% BCI)** | **Specificity (95% BCI)** |
| Primary analysis result | Non-informative priors | 49 (34, 64) | 68 (61, 77) | 79 (61, 99) | 97 (95, 100) |
| Sensitivity analysis 1 | More informative prior for Kato-Katz sensitivity and specificity | 49 (34, 64) | 68 (62, 76) | 80 (61, 99) | 97 (95, 100) |
| Sensitivity analysis 2 | More informative prior for qPCR sensitivity | 49 (38, 59) | 67 (63, 71) | 90 (80, 99) | 97 (95, 100) |
| Sensitivity analysis 3 | More informative prior for sensitivity and specificity of both Kato-Katz and qPCR | 49 (39, 59) | 67 (63, 71) | 90 (80, 99) | 97 (95, 100) |

**Table 5.7 Kappa statistics comparing the original classification of a single slide as having any STH ova between individual Kato-Katz technicians and expert parasitologists.**

| | ***Ascaris*** | | **Hookworm** | | ***Trichuris*** | |
|---|---|---|---|---|---|---|
| Technician | Kappa statistic | p-value | Kappa statistic[#] | p-value[#] | Kappa statistic[#] | p-value[#] |
| 1 | 0.83 | <0.001 | -- | -- | -- | -- |
| 2 | 0.90 | <0.001 | 0.18 | 0.004 | 0.79 | <0.001 |
| 3 | 1.00 | <0.001 | -- | -- | 0.66 | <0.001 |
| 4 | 0.91 | <0.001 | 0 | -- | 1.00 | <0.001 |

[#] Some values were not estimated (as indicated by --) because neither the technician nor the expert parasitologist identified any ova. The agreement is likely lower for hookworm because experienced parasitologists examined slides up to a few days after laboratory technicians, and hookworm ova may have begun to disintegrate by that time. In the primary study, in the subset of samples for which quality control was performed on the same day, the agreement for hookworm was high (Ercumen et al., 2019. https://doi.org/10.1371/journal.pntd.0007323).

**Table 5.8 Observed soil-transmitted helminth prevalence, double-slide Kato-Katz eggs per gram, and qPCR Cq values (N=2,799).**

| | double-slide Kato-Katz | | | qPCR | | |
|---|---|---|---|---|---|---|
| | **Number of positive samples** | **Observed prevalence (95% CI)** | **Median EPG in positive stool samples (range)** | **Number of positive samples** | **Observed prevalence (95% CI)** | **Median Cq value in positive stool samples (range)** |
| ***Ascaris lumbricoides*** | 1,035 | 37.0 (34.7, 39.3) | 72 (12, 100,824) | 652 | 23.3 (20.9, 25.7) | 17.4 (4.4, 39.6) |
| **Hookworm** | 209 | 7.5 (6.3, 8.7) | 96 (24, 6,732) | 598 | 21.4 (19.4, 23.3) | -- |
| ***Necator americanus*** | -- | -- | -- | 515 | 18.4 (16.5, 20.3) | 20.8 (13.9, 35.5) |
| ***Ancylostoma ceylanicum*** | -- | -- | -- | 106 | 3.8 (2.9, 4.6) | 23.6 (14.4, 37.8) |
| ***Ancylostoma duodenale*** | -- | -- | -- | 1 | <0.001 (0.0, 0.10) | 27.2 (27.2, 27.2) |
| ***Trichuris trichiura*** | 195 | 7.0 (5.6, 8.3) | 96 (24, 5,040) | 343 | 12.3 (10.3, 14.2) | 27.7 (21.4, 40.0) |
| ***Strongyloides stercoralis*** | -- | -- | -- | 17 | 0.6 (0.3, 0.9) | 24.6 (19.1, 31.8) |

**Table 5.9 Classification of qPCR and double-slide Kato-Katz for each type of STH (N=2,799).**

| ***Ascaris lumbricoides*** | **qPCR +** | **qPCR -** | **Kappa statistic (p-value)** |
|---|---|---|---|
| Kato-Katz + | 319 (11%) | 716 (26%) | 0.13 (<0.001) |
| Kato-Katz - | 333 (12%) | 1431 (51%) | |
| **Hookworm** | **qPCR +** | **qPCR -** | |
| Kato-Katz + | 195 (7%) | 14 (1%) | 0.42 (<0.001) |
| Kato-Katz - | 403 (14%) | 2187 (78%) | |
| ***Trichuris trichiura*** | **qPCR +** | **qPCR -** | |
| Kato-Katz + | 167 (6%) | 28 (1%) | 0.58 (<0.001) |
| Kato-Katz - | 176 (6%) | 2428 (87%) | |

**Table 5.10 Estimated sensitivity, and specificity of each diagnostic method using Bayesian latent class models.**

| | **double-slide Kato-Katz** | | **qPCR** | |
|---|---|---|---|---|
| | Sensitivity (%) (95% BCI) | Specificity (%) (95% BCI) | Sensitivity (%) (95% BCI) | Specificity (%) (95% BCI) |
| **Bayesian latent class model** | | | | |
| *Ascaris lumbricoides* | 49 (34, 64) | 68 (61, 77) | 79 (61, 99) | 97 (95, 100) |
| Hookworm[a] | 32 (22, 41) | 99 (96, 100) | 93 (82, 100) | 97 (95, 100) |
| *Trichuris trichiura* | 52 (33, 71) | 98 (96, 100) | 90 (81, 100) | 97 (95, 100) |
| **Any positive test (sensitivity) or any negative test (specificity) using double-slide Kato-Katz or qPCR** | | | | |
| | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) |
| *Ascaris lumbricoides*[b] | 49 (45, 53) | 67 (64, 69) | -- | -- |
| Hookworm[a] | 34 (30, 38) | 99 (99, 100) | 98 (97, 99) | 85 (83, 86) |
| *Trichuris trichiura* | 53 (47, 58) | 99 (99, 99) | 92 (90, 95) | 93 (92, 94) |

95% BCI: 95% Bayesian credible interval

95% CI: 95% confidence interval

[a] *Necator americanus*, *Ancylostoma duodenale*, and *Ancylostoma ceylanicum* combined

[b] Due to the high discordance between qPCR and double-slide Kato-Katz for *A. lumbricoides*, we estimated sensitivity and specificity using qPCR alone as the gold standard

**Table 5.11 Sensitivity analysis using alternative prior distributions in Bayesian latent class analysis models for *A. lumbricoides*.**

| | | Kato-Katz | | qPCR | |
|---|---|---|---|---|---|
| **Analysis** | **Description of priors** | **Sensitivity (95% BCI)** | **Specificity (95% BCI)** | **Sensitivity (95% BCI)** | **Specificity (95% BCI)** |
| **Primary analysis result** | Non-informative priors | 49 (34, 64) | 68 (61, 77) | 79 (61, 99) | 97 (95, 100) |
| **Sensitivity analysis 1** | More informative prior for Kato-Katz sensitivity and specificity | 49 (34, 64) | 68 (62, 76) | 80 (61, 99) | 97 (95, 100) |
| **Sensitivity analysis 2** | More informative prior for qPCR sensitivity | 49 (38, 59) | 67 (63, 71) | 90 (80, 99) | 97 (95, 100) |
| **Sensitivity analysis 3** | More informative prior for sensitivity and specificity of both Kato-Katz and qPCR | 49 (39, 59) | 67 (63, 71) | 90 (80, 99) | 97 (95, 100) |

**Table 5.12 Taxonomic assignments for OTUs resulting from sequencing analysis with the Qiime2 pipeline.**

| | Eukaryota;Amoebozoa;Discosea;Longamoebia;Centramoebida;Acanthamoeba;Acanthamoeba stevensoni;__;__;__;__;__;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Apiales;Cnidium;Cnidium monnieri;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Brassicales;Brassica;Brassica napus (rape);__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Caryophyllales;Amaranthus;Amaranthus tuberculatus;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Caryophyllales;Rheum;Rheum palmatum;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Cucurbitales;Cucumis;Cucumis sativus (cucumber);__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Fabales;Glycine;Glycine max (soybean);__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Fabales;Vigna;Vigna unguiculata subsp. unguiculata (cowpea);__ |
|---|---|---|---|---|---|---|---|---|
| K-K +/PCR - #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #2 | 0 | 233 | 0 | 0 | 0 | 195 | 0 | 93 |
| K-K +/PCR - #3 | 0 | 2824 | 0 | 0 | 958 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 538 | 13 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 0 | 45 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 13 | 1067 | 0 | 0 | 207 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 16 | 0 | 79 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR - | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Alismatales;Zantedeschia;Zantedeschia aethiopica | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Arecaceae;Elaeis;Elaeis guineensis (African oil palm) | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Poales;Poaceae;Oryza;Zea mays subsp. mays (maize) | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Zingiberales;Curcuma;Curcuma elata | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Zingiberales;Curcuma;Zingiber mioga | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Liliopsida;Zingiberales;Musa;Musa itinerans | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Malvales;Gossypium;Gossypium barbadense (sea-island cotton);__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Medicago truncatula (barrel medic);__;__;__ |
| K-K +/PCR - #1 | 0 | 144 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #2 | 0 | 0 | 277 | 0 | 0 | 190 | 0 | 173 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 88 | 108 | 8308 | 0 | 27 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 110 |
| K-K +/PCR - #5 | 0 | 0 | 556 | 0 | 0 | 0 | 0 | 640 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 259 |
| K-K +/PCR - #7 | 120 | 0 | 62 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 373 | 0 | 0 | 0 | 0 | 230 | 0 |
| K-K -/PCR - | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Myrtales;Eucalyptus;Eucalyptus grandis;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Rosales;Broussonetia;Broussonetia papyrifera;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Rosales;Ziziphus;Ziziphus jujuba;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Solanales;Nicotiana;Nicotiana attenuata;__ | Eukaryota;Archaeplastida;Chloroplastida;Charophyta;Phragmoplastophyta;Streptophyta;Embryophyta;Tracheophyta;Spermatophyta;Magnoliophyta;Solanales;Solanum;Solanum tuberosum (potato);__ | Eukaryota;Archaeplastida;Chloroplastida;Chlorophyta;Chlorophyceae;Pectinodesmus pectinatus;__;__;__;__;__;__ | Eukaryota;Archaeplastida;Chloroplastida;Chlorophyta;Chlorophyceae;Sphaeroplea les;Scenedesmus;Desmodesmus communis;__;__;__;__;__;__ | Eukaryota;Archaeplastida;Chloroplastida;Chlorophyta;Trebouxiophyceae;Apatococcus lobatus;__;__;__;__;__;__ |
| K-K +/PCR - #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #2 | 0 | 0 | 0 | 127 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 14274 | 87 | 43 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 140 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 2200 | 0 | 16211 | 0 | 0 | 35 | 0 | 11 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 132 | 0 | 0 | 39 | 0 |
| K-K +/PCR - #7 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR - | 0 | 0 | 0 | 142 | 0 | 0 | 0 | 0 |

| | Eukaryota;Archaeplastida;Chloroplastida;Chlorophyta;Trebouxiophyceae;Prototheca zopfii var. hydrocarbonea;_;_;_;_;_;_;_;_ | Eukaryota;Archaeplastida;Rhodophyceae;Florideophycidae;Nemaliophycidae;Petrohuabernabei;_;_;_;_;_;_;_ | Eukaryota;Excavata;Discoba;Discicristata;Euglenozoa;Kinetoplastea;Metakinetoplastina;Trypanosomatida;Zelonia;Zelonia cf. costaricensis;_;_;_;_ | Eukaryota;Excavata;Metamonada;Fornicata;Diplomonadida;Giardiinae;Giardia;Giardia intestinalis;_;_;_;_;_;_ | Eukaryota;Excavata;Metamonada;Parabasalia;Trichomonadea;Dientamoeba;Dientamoeba fragilis;_;_;_;_;_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Arthropoda;Chelicerata;Arachnida;Acari;Puccinia striiformis f. sp. tritici 87/7;_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Arthropoda;Crustacea;Maxillopoda;Copepoda;Calanoida;unculturedeukaryote;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Chordata;Vertebrata;Gnathostomata;Euteleostomi;Tetrapoda;Mammalia;Homo sapiens (human);_ |
|---|---|---|---|---|---|---|---|---|
| K-K +/PCR - #1 | 0 | 0 | 0 | 3421 | 127305 | 81 | 0 | 0 |
| K-K +/PCR - #2 | 0 | 0 | 35 | 0 | 0 | 0 | 5 | 125 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 179 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 311 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 103 | 0 | 0 | 0 | 13658 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 94 | 0 | 0 | 96 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 117 | 60 | 0 | 0 | 5513 |
| K-K -/PCR + #1 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 1043 |
| K-K -/PCR + #2 | 58 | 12 | 0 | 0 | 333 | 0 | 356 | 1014 |
| K-K -/PCR - | 0 | 0 | 0 | 7910 | 10752 | 0 | 0 | 79 |
| | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Chordata;Vertebrata;Gnathostomata;Euteleostomi;Tetrapoda;Mammalia;Tursiops truncatus (bottlenose dolphin);_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Mollusca;Bivalvia;Palaeoheterodonta;Unionoida;Elliptio complanata (eastern elliptio);_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Nematoda;Chromadorea;Oxyurida;Enterobius vermicularis (human pinworm);_;_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Nematoda;Chromadorea;Spirurida;*Ascaris lumbricoides* (common roundworm);_;_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Bilateria;Platyhelminthes;Rhabditophora;Seriata;Tricladida;Continenticola;Schmidtea mediterranea;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Cnidaria;Hydrozoa;Hydroidolina;Siphonophorae;unculturedeukaryote;_;_;_;_ | Eukaryota;Opisthokonta;Holozoa;Metazoa (Animalia);Eumetazoa;Cnidaria;Myxozoa;Myxosporea;Bivalvulida;Thelohanellus kitauei;_;_;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus candidus;_;_ |
| K-K +/PCR - #1 | 0 | 8 | 252725 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #2 | 0 | 0 | 0 | 0 | 0 | 0 | 129 | 0 |
| K-K +/PCR - #3 | 63 | 0 | 0 | 0 | 0 | 7 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 655 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 191 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 115 | 0 | 0 | 0 | 0 | 0 | 0 | 161 |
| K-K -/PCR + #2 | 74 | 0 | 110829 | 383 | 0 | 0 | 0 | 0 |
| K-K -/PCR - | 0 | 0 | 85879 | 0 | 21 | 0 | 0 | 189 |
| | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus cristatus;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus lentulus;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus niger ATCC 1015;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus penicillioides;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Aspergillus;Aspergillus ustus;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Aspergillaceae;Penicilliopsis;Penicilliopsis clavariiformis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Eurotiomycetes;Eurotiales;Trichocomaceae;Talaromyces;Talaromyces purpureogenus;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Pezizomycotina;Pezizales;Rhizinaceae;Phymatotrichopsis;uncultured eukaryote;_;_ |
| K-K +/PCR - #1 | 0 | 0 | 437 | 0 | 0 | 0 | 36 | 0 |
| K-K +/PCR - #2 | 0 | 0 | 0 | 0 | 0 | 53 | 0 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 274 | 68 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 309 | 0 | 0 | 140 | 0 |
| K-K +/PCR - #6 | 211 | 0 | 0 | 0 | 0 | 0 | 0 | 37 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 252 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 1184 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 120 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR - | 0 | 0 | 0 | 40 | 0 | 0 | 0 | 0 |
| | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Debaryomycetaceae;Candida-Lodderomyces clade;Candida dubliniensis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Debaryomycetaceae;Candida-Lodderomyces clade;Candida parapsilosis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Debaryomycetaceae;Yamadazyma-Candida clade;[Candida] membranifaciens;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Candida;Candida ghanaensis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Candida;Candida versatilis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Starmerella-Candida clade;Starmerella bacillaris;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Starmerella-Candida clade;[Candida] apicola;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Starmerella-Candida clade;[Candida] etchellsii;_;_ |
| K-K +/PCR - #2 | 0 | 0 | 0 | 0 | 0 | 3069 | 0 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 0 | 41 | 0 | 0 | 1428 |
| K-K -/PCR + #2 | 109 | 0 | 0 | 0 | 0 | 0 | 0 | 477 |
| K-K -/PCR - | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 15 |
| | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Incertae Sedis;Starmerella-Candida clade;[Candida] stigmatis;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Pichiaceae;Pichia;Pichia kudriavzevii;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Pichiaceae;Pichia;Pichia sp.;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Saccharomycetaceae;Saccharomyces;Saccharomyces cerevisiae (baker's yeast);_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Saccharomycopsidaceae;Saccharomycopsis;Saccharomycopsis malanga;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Trichomonascaceae;Zygoascus-Candida clade;Zygoascus hellenicus;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Ascomycota;Saccharomycotina;Saccharomycetes;Saccharomycetales;Trichomonascaceae;Zygoascus-Candida clade;Zygoascus meyerae;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Agaricales;Pterulaceae;Pterula;Pterula echo;_;_ |
| K-K +/PCR - #2 | 6 | 624 | 115 | 0 | 0 | 14010 | 30 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 3812 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Polyporales;Lentinaceae;Lentinus;Lentinus polychrous;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Tremellomycetes;Tremellales;Rhynchogastremataceae;Papiliotrema;Papiliotrema bandonii;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Tremellomycetes;Trichosporonales;Trichosporonaceae;Trichosporon;Trichosporon asahii var. asahii CBS 2479;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Wallemiomycetes;Wallemiales;Incertae Sedis;Wallemia;Wallemia muriae;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Pucciniomycotina;Pucciniomycetes;Pucciniales;Pucciniaceae;Gymnosporangium;Gymnosporangium cupressi;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Dikarya;Basidiomycota;Ustilaginomycotina;Malasseziomycetes;Malasseziales;Malasseziaceae;Malassezia;Malassezia japonica;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Mucoromycota;Mucoromycotina;Incertae Sedis;Mucorales;Choanephoraceae;Blakeslea;Blakeslea trispora;_;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Mucoromycota;Mucoromycotina;Incertae Sedis;Mucorales;Lichtheimiaceae;Rhizomucor;Rhizomucor miehei;_;_;_ |
| K-K +/PCR - #2 | 0 | 0 | 68 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 75 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 52 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 11 | 0 | 0 | 35 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 5 | 0 | 187 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 0 | 0 | 0 | 0 | 75 | 0 | 56 |
| | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Mucoromycota;Mucoromycotina;Incertae Sedis;Mucorales;Mucoraceae;Mucor;Mucor rudolphii;_;_;_ | Eukaryota;Opisthokonta;Nucletmycea;Fungi;Mucoromycota;Mucoromycotina;Incertae Sedis;Mucorales;Rhizopodaceae;Rhizopus;Rhizopus delemar RA 99-880;_;_;_ | Eukaryota;SAR;Alveolata;Apicomplexa;Aconoidasida;Haemosporoidia;Plasmodium;Plasmodium ovale wallikeri;_;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Apicomplexa;Conoidasida;Cryptosporida;Cryptosporidium;Colpodella sp. ATCC50594;_;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Apicomplexa;Conoidasida;Gregarinasina;Eugregarinorida;Gregarina sp. Phaedon/Daegwallyeong/2010;_;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Apicomplexa;Conoidasida;Gregarinasina;Eugregarinorida;Gregarina;Gregarina chortiocetes;_;_;_;_ | Eukaryota;SAR;Alveolata;Apicomplexa;Conoidasida;Gregarinasina;Eugregarinorida;uncultured;Gregarinidae sp. from Paranemertes peregrina SR-2009a;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Conthreep;Colpodea;Colpodida;Bromeliothrix;Colpoda sp. PRA-118;_;_;_;_ |
| K-K +/PCR - #2 | 0 | 0 | 470 | 0 | 0 | 31 | 5116 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 97 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 460 | 22 | 132 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 114 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 0 | 2489 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 33 | 29 | 0 | 0 | 0 | 0 | 22 |
| | Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Conthreep;Colpodea;Colpodida;Colpoda;metagenome;_;_;_;_ | Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Litostomatea;Haptoria;Arcuospathidium;Arcuospathidium namibiense tristicha;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Litostomatea;Mesodiniidae;uncultured;uncultured marine eukaryote;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Spirotrichea;Hypotrichia;Anteholosticha;Anteholosticha monilata;_;_;_;_;_ | Eukaryota;SAR;Alveolata;Protalveolata;Syndiniales;Syndiniales Group I;Karlodinium veneficum;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Cercomonadidae;Paracercomonas;Uncultured Cercozoa;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Imbricatea;Nudifila;metagenome;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Imbricatea;Silicofilosea;Euglyphida;Trinematidae;Trinema;metagenome;_;_;_;_ |
| K-K +/PCR - #2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 0 | 0 | 0 | 815 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 0 | 0 | 209 | 25 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| K-K -/PCR + #1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #2 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 |
| | Eukaryota;SAR;Rhizaria;Cercozoa;Incertae Sedis;Gymnophrys;uncultured eukaryote;_;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Incertae Sedis;metagenome;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Incertae Sedis;uncultured eukaryote;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Cercozoa;Novel Clade Gran-5;uncultured eukaryote;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Rhizaria;Retaria;RAD B;uncultured eukaryote;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Stramenopiles;Incertae Sedis;Blastocystis;Blastocystis sp. ATCC 50177/Nand II;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Stramenopiles;Ochrophyta;Chrysophyceae;uncultured;uncultured marine eukaryote;_;_;_;_;_;_;_ | Eukaryota;SAR;Stramenopiles;Ochrophyta;Eustigmatophyceae;Eustigmatales;Goniochloris;Goniochloris sculpta;_;_;_;_;_;_;_ |
| K-K +/PCR - #2 | 0 | 0 | 0 | 0 | 0 | 0 | 11721 | 0 |
| K-K +/PCR - #3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #4 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #5 | 0 | 0 | 0 | 0 | 0 | 30498 | 0 | 0 |
| K-K +/PCR - #6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K-K +/PCR - #7 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 |
| K-K -/PCR + #1 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 28 |
| K-K -/PCR + #2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Eukaryota;SAR;Stramenopiles;Ochrophyta;Eustigmatophyceae;Eustigmatales;uncultured phytoplankton;_;_;_;_;_;_;_;_ | Eukaryota;SAR;Stramenopiles;Peronosporomycetes;Phytophthora;Plasmopara viticola;_;_;_;_;_;_;_;_ | Unassigned;_;_;_;_;_;_;_;_;_;_;_;_;_;_;_ | | | | | |
| K-K +/PCR - #2 | 0 | 0 | 722 | | | | | |
| K-K +/PCR - #3 | 0 | 0 | 14 | | | | | |
| K-K +/PCR - #4 | 0 | 235 | 34 | | | | | |
| K-K +/PCR - #5 | 0 | 0 | 4744 | | | | | |
| K-K +/PCR - #6 | 0 | 0 | 3846 | | | | | |
| K-K +/PCR - #7 | 0 | 0 | 67 | | | | | |
| K-K -/PCR + #1 | 10 | 0 | 32 | | | | | |
| K-K -/PCR + #2 | 0 | 9 | 1031 | | | | | |

Following de-multiplexing by barcode, OTUs underwent taxonomic assignment. For each sample, the number of interlaced read pairs mapping to each taxonomic assignment is provided.

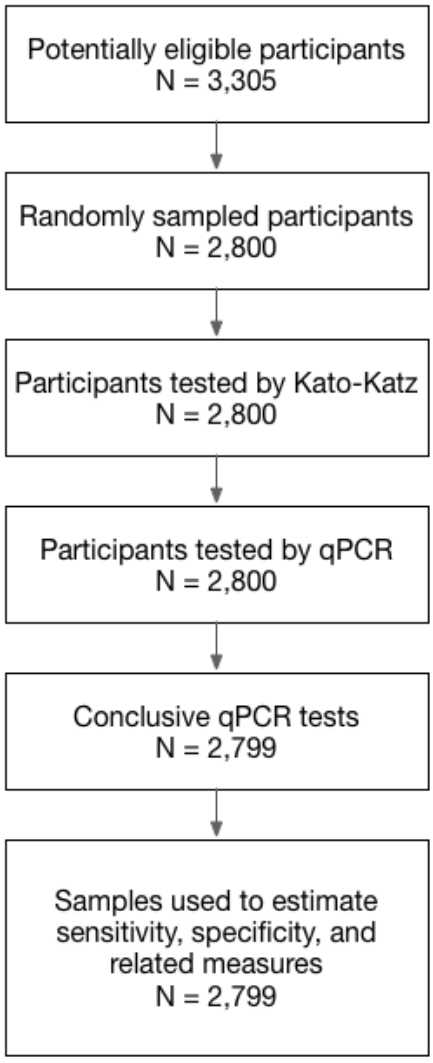


**Figure 5.1 Participant flow diagram.** An overview of participant selection and testing is provided.

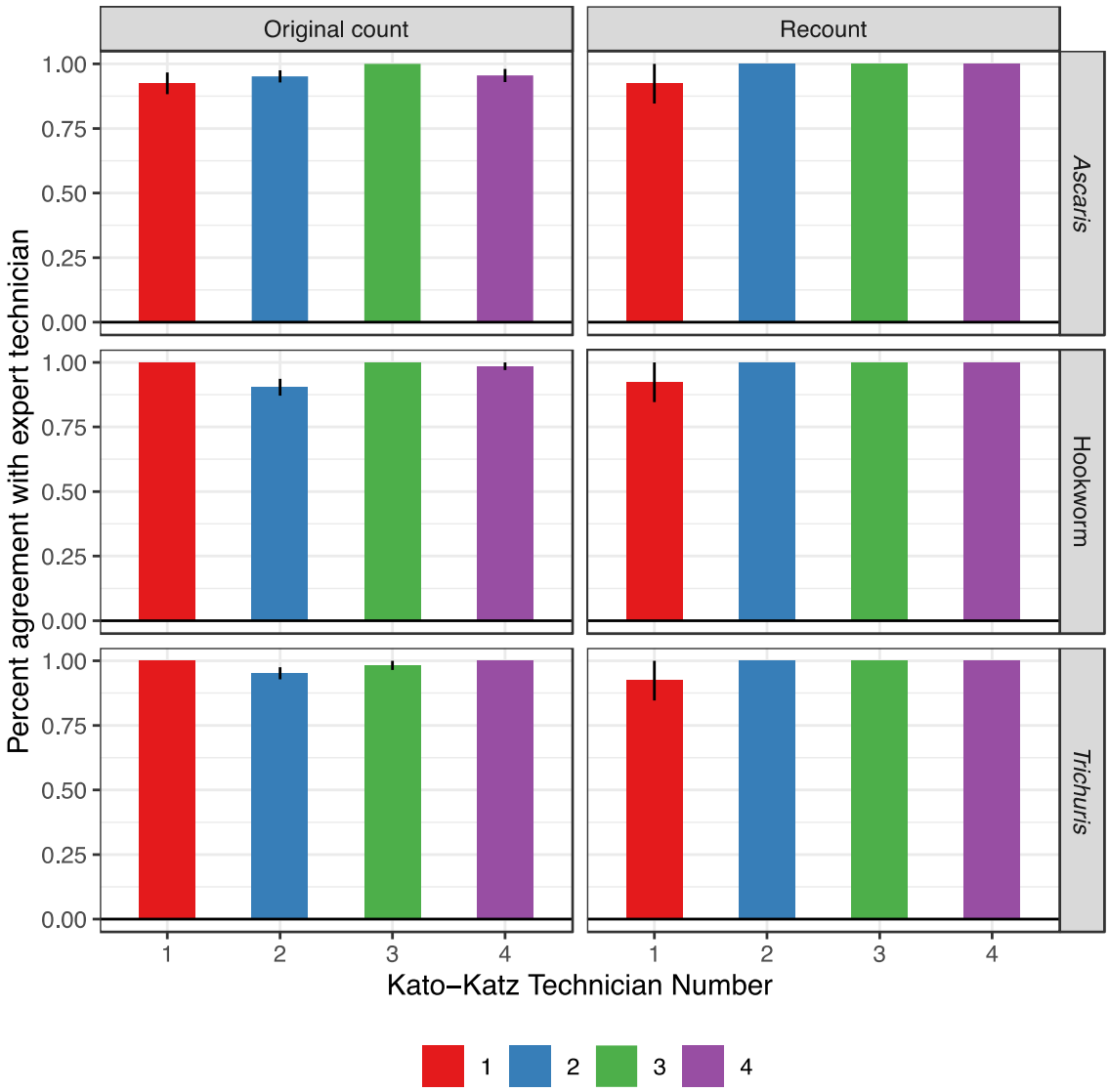


**Figure 5.2 Percent agreement between laboratory technicians and expert technicians in STH positive/negative status for individual slides.** Each bar indicates the percent agreement in STH positive/negative status between an individual technician and an expert technician on a single slide reading. Black vertical lines indicate the 95% confidence intervals.

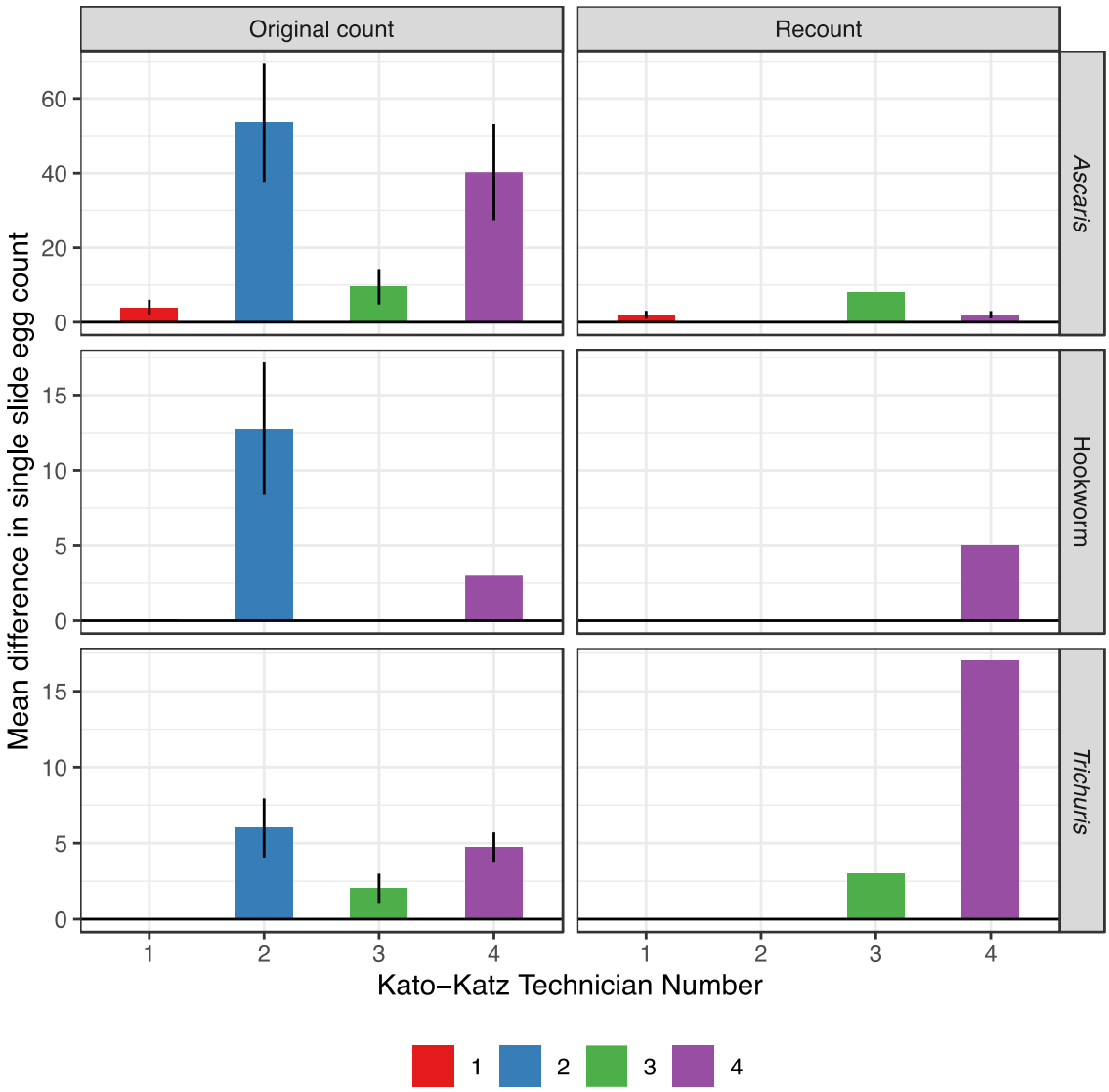


**Figure 5.3 Mean difference in Kato-Katz single slide egg count between laboratory technician and expert technician among samples that were positive by either the laboratory technician or the expert.** Each bar indicates the mean difference in single slide egg counts between an individual technician and an expert technician. Black vertical lines indicate the 95% confidence intervals.

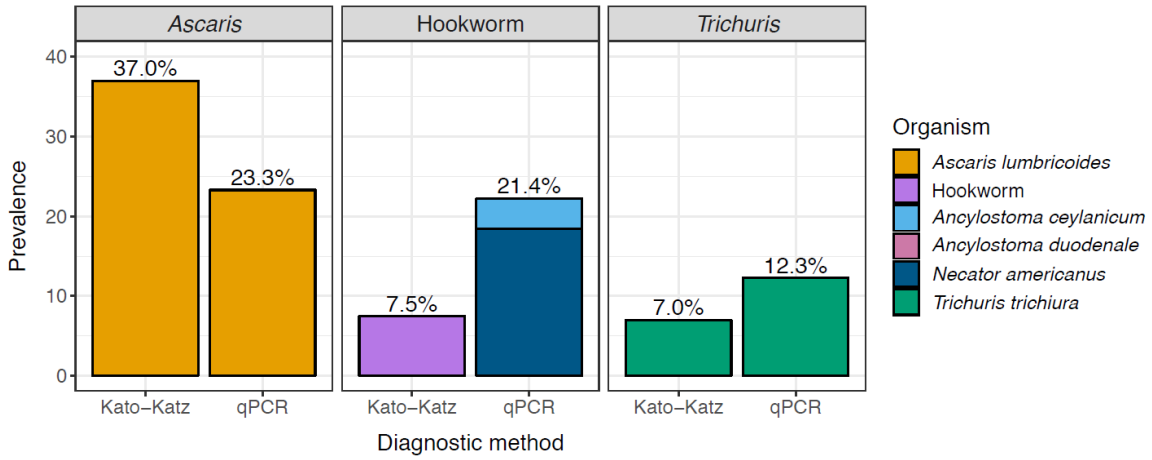


**Figure 5.4 Observed soil-transmitted helminth prevalence by double slide Kato-Katz and qPCR.** Prevalence was estimated from stool samples collected from children aged 2-12 years in rural Bangladesh (N=2,799).

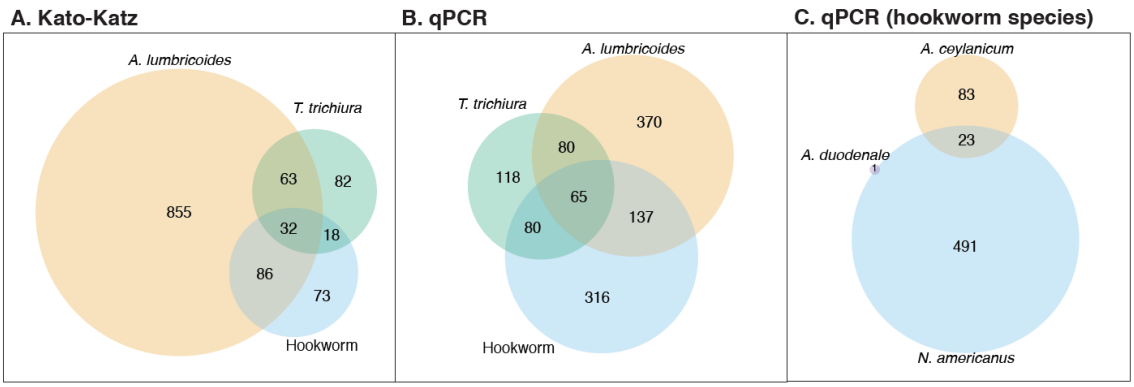


**Figure 5.5 Venn diagram of co-infections detected by double slide Kato-Katz and qPCR.** Soil-transmitted helminth ova or DNA were detected in stool samples collected from children aged 2-12 years in rural Bangladesh using Kato-Katz or multi-parallel qPCR (N=2,799).

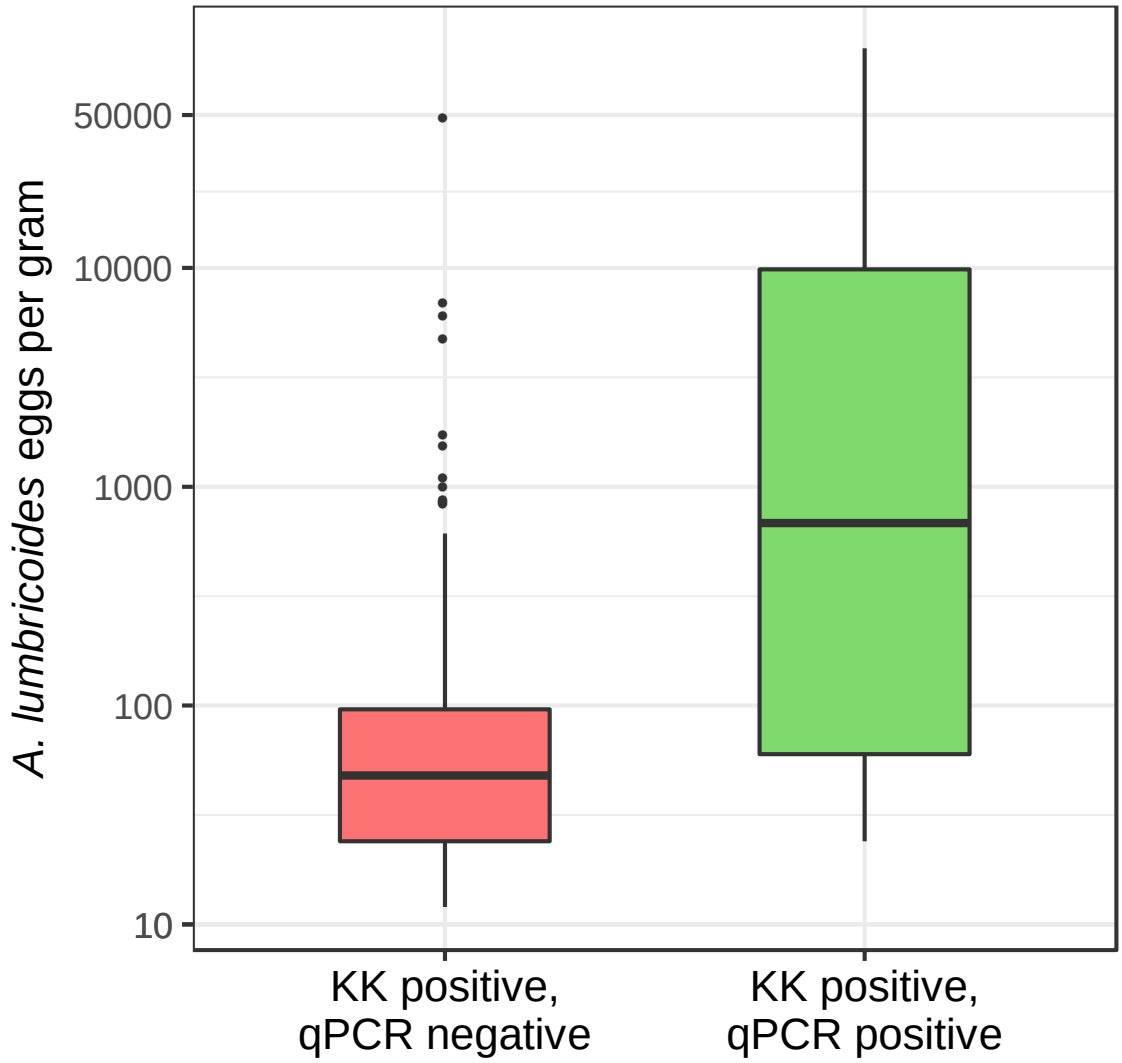


**Figure 5.6 Distribution of *A. lumbricoides* eggs per gram classified by concordance status between Kato-Katz and qPCR.** For this comparison, qPCR is assumed to be the gold standard of detection.

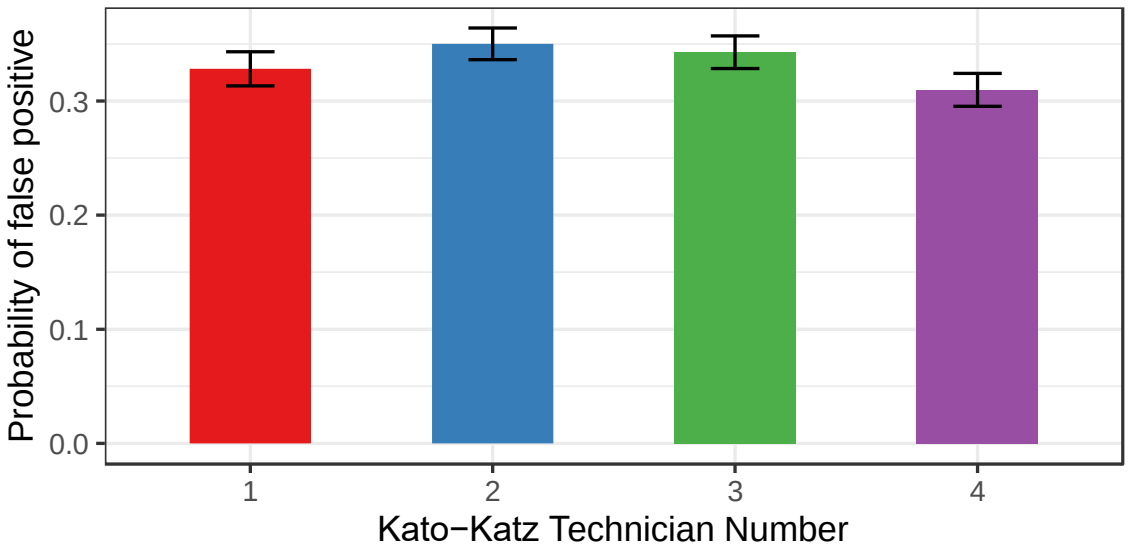


**Figure 5.7 Probability that a stool sample was classified as positive for *A. lumbricoides* using Kato-Katz among those classified as negative by qPCR by individual Kato-Katz technician.** Each bar indicates the probability that a sample was classified as positive by Kato-Katz among those classified as negative by qPCR for an individual technician. Black vertical lines indicate the 95% confidence intervals.

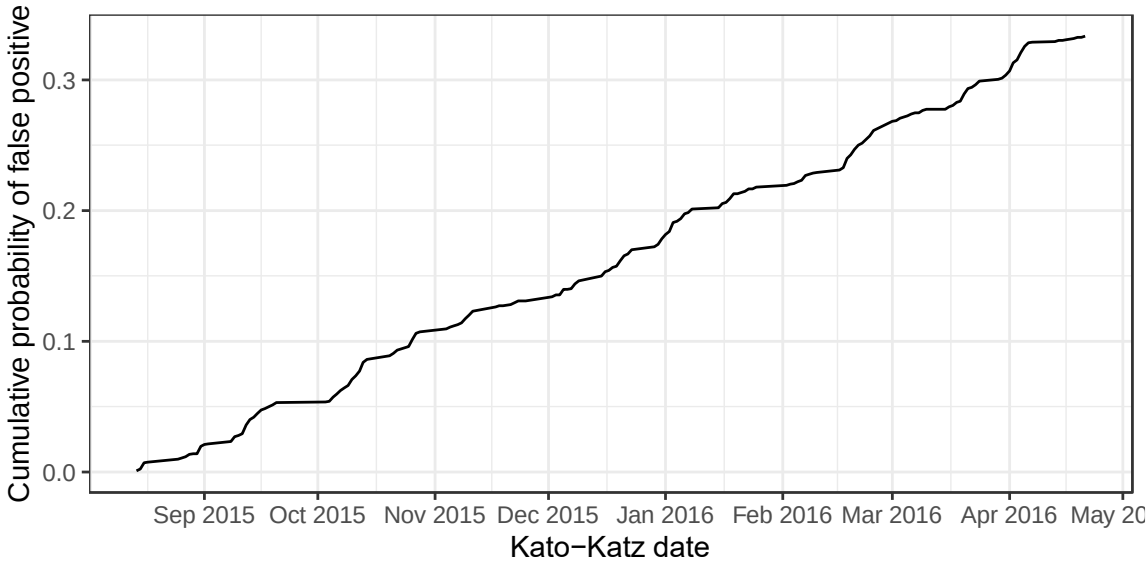


**Figure 5.8 Cumulative probability that a stool sample was classified as positive for *A. lumbricoides* using Kato-Katz among those classified as negative by qPCR by date of Kato-Katz.** Data were not included for hookworm and *Trichuris* since only 1% of samples were classified as positive by Kato-Katz among those classified as negative by qPCR for those STH.

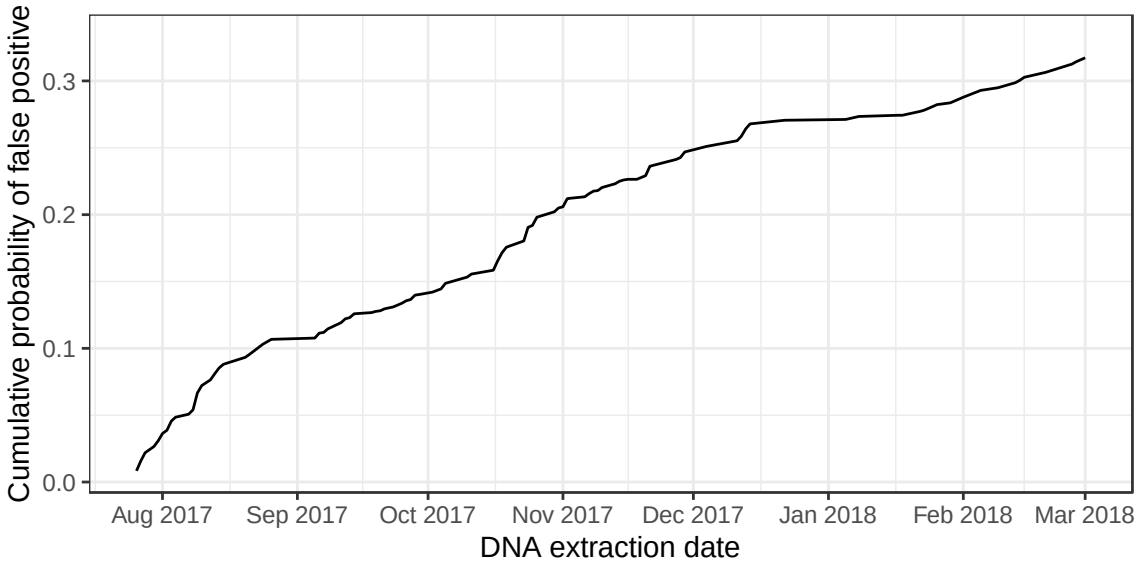


**Figure 5.9 Cumulative probability that a stool sample was classified as positive for *A. lumbricoides* using Kato-Katz among those classified as negative by qPCR by date of DNA extraction.** The date of DNA extraction was not recorded for the last 242 samples; we imputed the date for those samples as 1 day later than the last recorded date. Data were not included for hookworm and *Trichuris* since only 1% of samples were classified as positive by Kato-Katz among those classified as negative by qPCR for those STH.

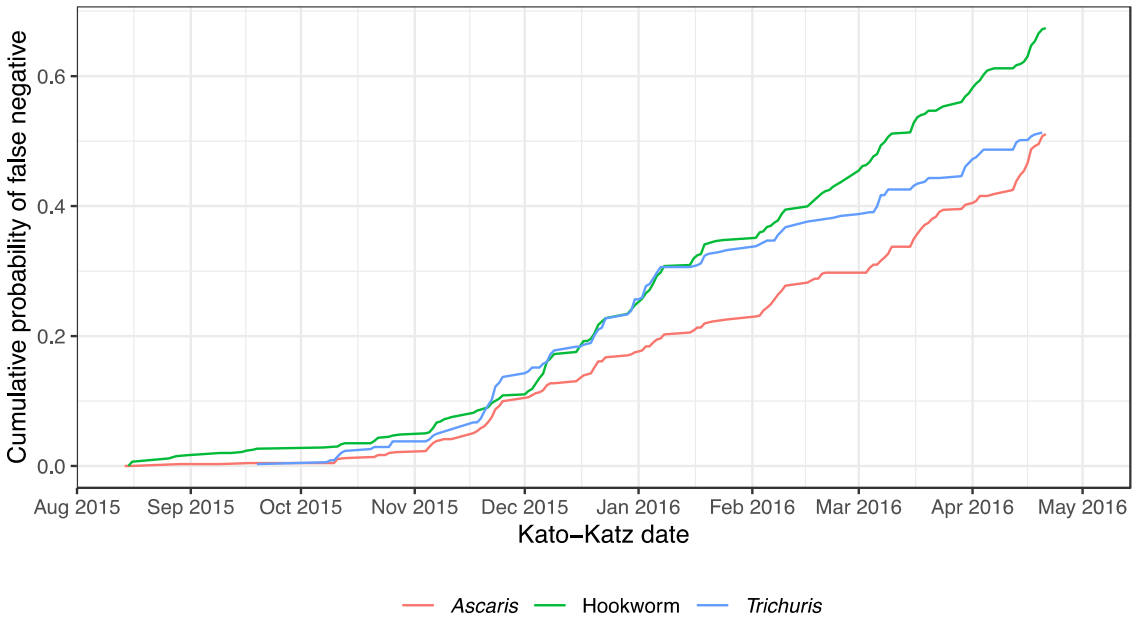


**Figure 5.10 Cumulative probability that a stool sample was classified as negative using Kato-Katz among those classified as positive by qPCR by date of Kato-Katz.** All data were included for all species of STH.

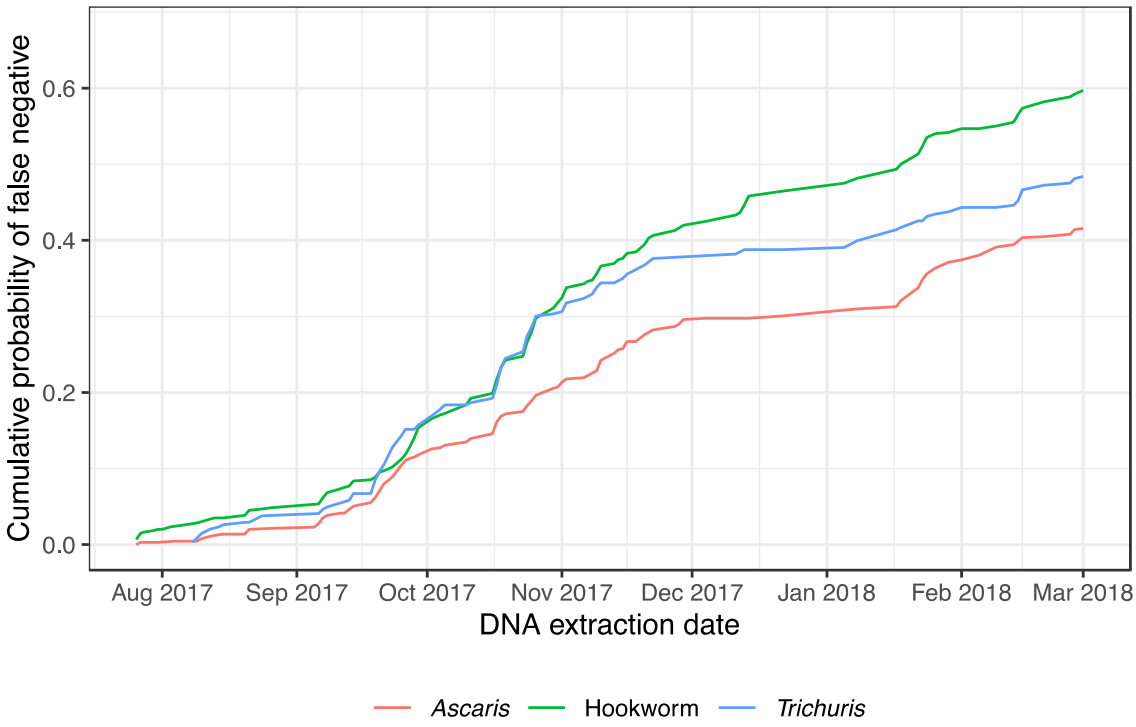


**Figure 5.11 Cumulative probability that a stool sample was classified as negative using Kato-Katz among those classified as positive by qPCR by date of DNA extraction.** The date of DNA extraction was not recorded for the last 242 samples; we imputed the date for those samples as 1 day later than the last recorded date.

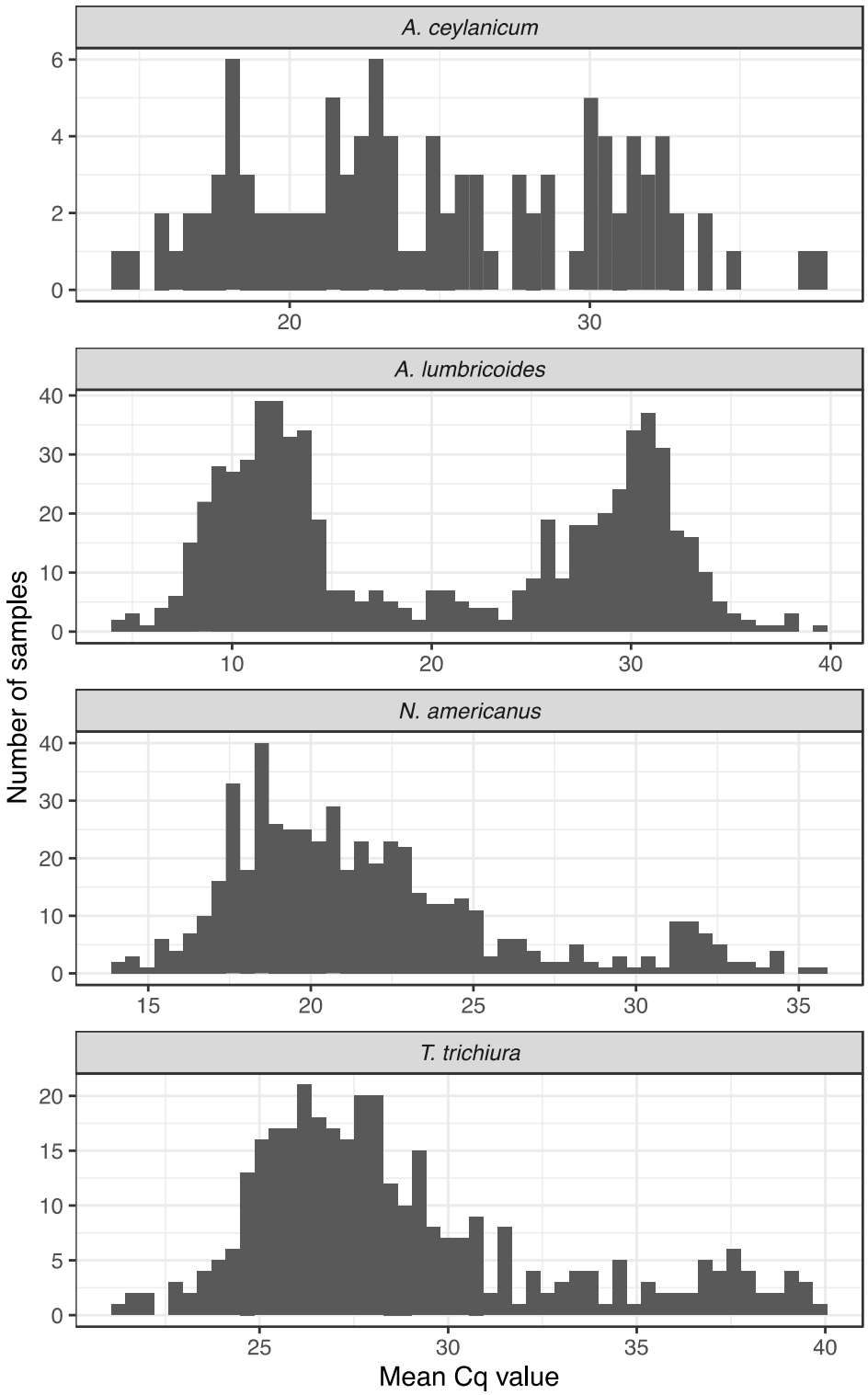


**Figure 5.12 Distribution of Cq values for each soil-transmitted helminth.** A panel was not included for *A. duodenale* as DNA was detected from only a single sample for that species.

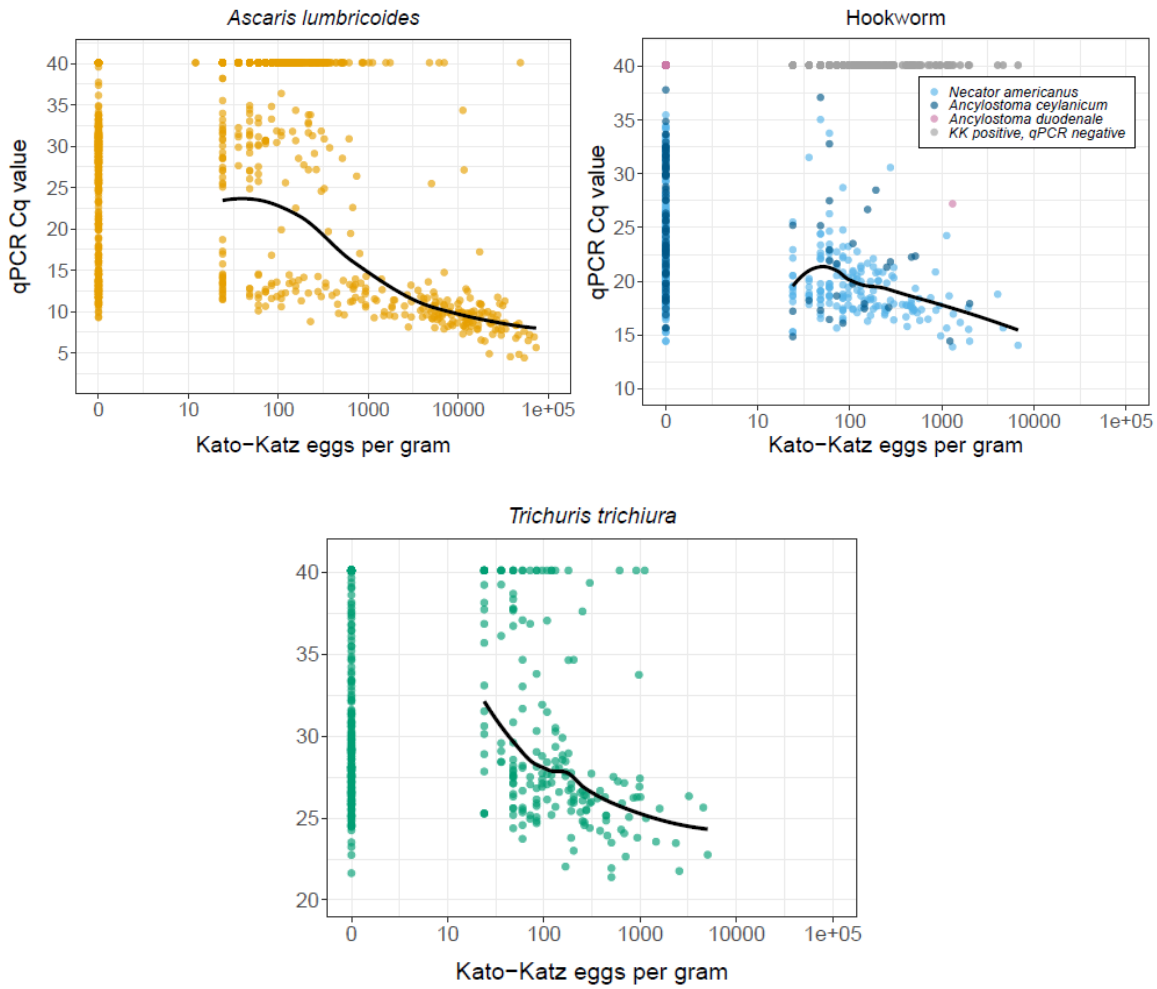


**Figure 5.13 Relationship between Cq values measured by qPCR and eggs per gram estimated using double slide lide Kato-Katz.** Soil-transmitted helminth ova or DNA were detected by Kato-Katz and qPCR in stool samples collected from children aged 2-12 years in rural Bangladesh (N=2,799). The black solid line is the LOESS smoother for values that were classified as positive using both tests. Gray points indicate results that were negative by qPCR but positive by Kato-Katz; they are gray because Kato-Katz cannot differentiate between hookworm species. The Kendall's tau comparing eggs per gram and Cq values was -0.442 for *A. lumbricoides*, -0.346 for *N. americanus*, -0.266 for *A. ceylanicum*, and -0.248 for *T. trichiura* (for each, the p-value was <0.001).

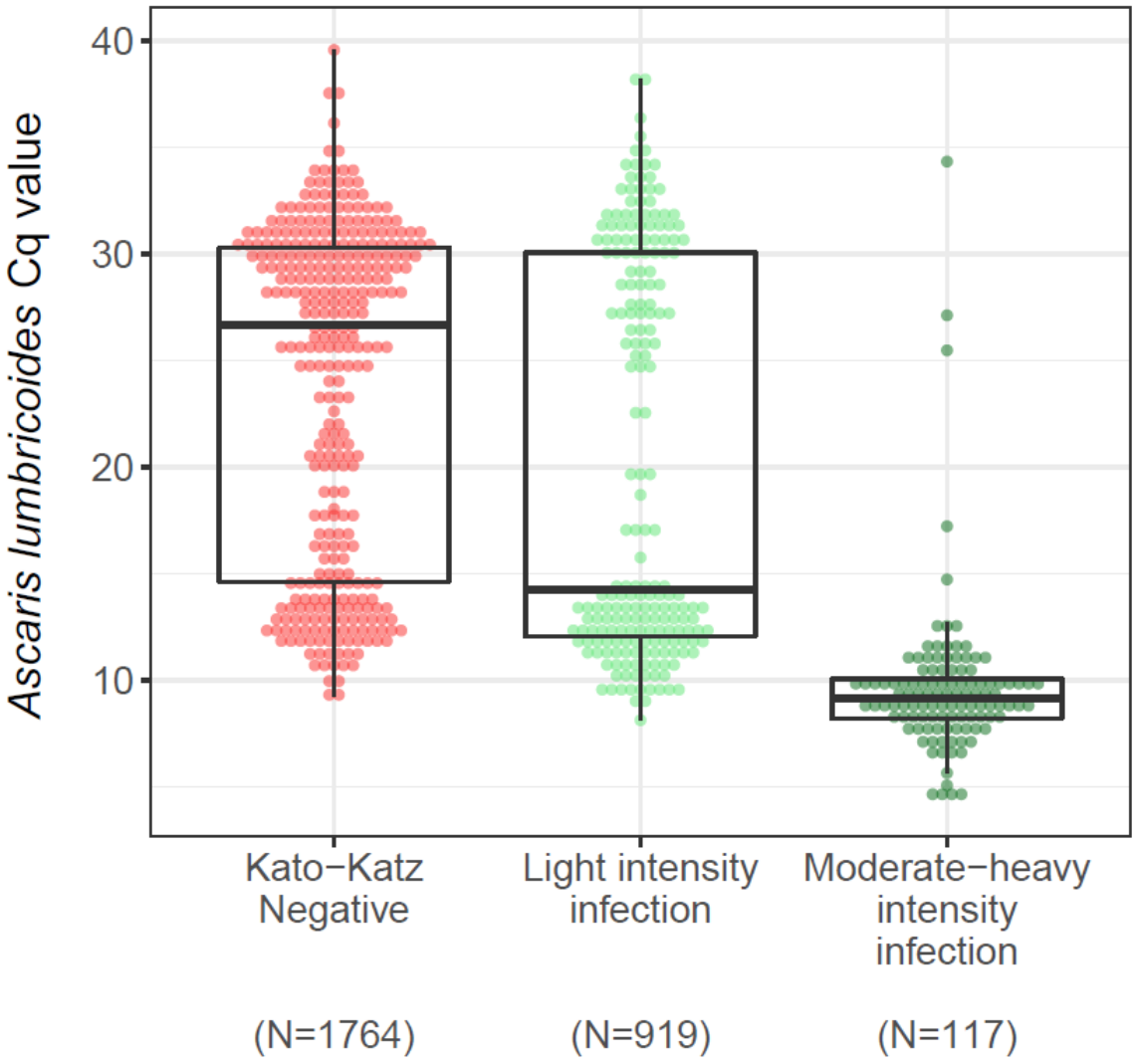


**Figure 5.14 Distribution of *A. lumbricoides* Cq values within infection intensity categories as determined by double slide Kato-Katz.** Light intensity infections were defined as < 5,000 epg for *A. lumbricoides*, < 1,000 epg for *T. trichiura*, and < 2,000 epg for hookworm; moderate intensity infections were defined as 5,000 ≤ epg < 50,000 for *A. lumbricoides*, 1,000 ≤ epg < 10,000 for *T. trichiura*, and 2,000 ≤ epg < 4,000 for hookworm; heavy intensity infections were defined as ≥ 50,000 epg for *A. lumbricoides*, ≥ 10,000 epg for *T. trichiura*, and ≥ 4,000 epg for hookworm per the World Health Organization definition.

## 5.9 References


1. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390: 1260–1344.

2. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet. 2006; 367: 1521–1532.

3. Lo NC, Heft-Neal S, Coulibaly JT, Leonard L, Bendavid E, Addiss DG. State of deworming coverage and equity in low-income and middle-income countries using household health surveys: a spatiotemporal cross-sectional study. Lancet Glob Health. 2019; 7: e1511–e1520.

4. Bundy DAP, Silva N de, Horton S, Jamison DT. Disease Control Priorities, Third Edition (Volume 8): Child and Adolescent Health and Development. World Bank Publications; 2017.

5. World Health Organization. Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation. 2012. Available: http://www.who.int/neglected_diseases/NTD_RoadMap_2012_Fullversion.pdf

6. Anderson R, Hollingsworth TD, Truscott J, Brooker S. Optimisation of mass chemotherapy to control soil-transmitted helminth infection. Lancet. 2012; 379: 289–290.

7. Anderson R, Truscott J, Hollingsworth TD. The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths. Philos Trans R Soc Long B Biol Sci. 2014; 369: 20130435.

8. Ásbjörnsdóttir KH, Means AR, Werkman M, Walson JL. Prospects for elimination of soil-transmitted helminths. Curr Opin Infect Dis. 2017; 30: 482-488.

9. Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, Bailey R, Gardiner I, Halliday KE, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3 cluster randomized trial protocol. PLoS Negl Trop Dis. 2018; 12: e0006166.

10. Medley GF, Turner HC, Baggaley RF, Holland C, Hollingsworth TD. The Role of More Sensitive Helminth Diagnostics in Mass Drug Administration Campaigns: Elimination and Health Impacts. Adv Parasitol. 2016; 94: 343–392.

11. Kato K, Miura M. Comparative examinations. Jpn J Parasitol. 1954; 3: 35.

12. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in *Schistosomiasis mansoni*. Rev Inst Med Trop Sao Paulo. 1972; 14: 397–400.

13. Allen AV, Ridley DS. Further observations on the formol-ether concentration technique for faecal parasites. J Clin Pathol. 1970; 23: 545–546.

14. Cringoli G, Rinaldi L, Maurelli MP, Utzinger J. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat Protoc. 2010; 5: 503–515.

15. Barda B, Cajal P, Villagran E, Cimino R, Juarez M, Krolewiecki A, et al. Mini-FLOTAC, Kato-Katz and McMaster: three methods, one goal; highlights from north Argentina. Parasit Vectors. 2014; 7: 271.

16. Barda B, Albonico M, Ianniello D, Ame SM, Keiser J, Speich B, et al. How Long Can Stool Samples Be Fixed for an Accurate Diagnosis of Soil-Transmitted Helminth Infection Using Mini-FLOTAC? PLoS Negl Trop Dis. 2015; 9: e0003698.

17. Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruysse J, Utzinger J, et al. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. Nat Protoc. 2017; 12: 1723–1732.

18. Cools P, Vlaminck J, Albonico M, Ame S, Ayana M, Antonio BPJ, et al. Diagnostic performance of a single and duplicate Kato-Katz, Mini-FLOTAC, FECPAKG2 and qPCR for the detection and quantification of soil-transmitted helminths in three endemic countries. PLoS Negl Trop Dis. 2019; 13: e0007446.

19. Ayana M, Vlaminck J, Cools P, Ame S, Albonico M, Dana D, et al. Modification and optimization of the FECPAKG2 protocol for the detection and quantification of soil-transmitted helminth eggs in human stool. PLoS Negl Trop Dis. 2018; 12: e0006655.

20. WHO. Assessing the Epidemiology of Soil-Transmitted Helminths During a Transmission Assessment Survey in the Global Programme for the Elimination of Lymphatic Filariasis. Geneva, Switzerland: WHO Press; 2015.

21. Anderson RM, Schad GA. Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. Trans R Soc Trop Med Hyg. 1985; 79: 812–825.

22. Nikolay B, Brooker SJ, Pullan RL. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. Int J Parasitol. 2014; 44: 765–774.

23. Verweij JJ, Brienen EAT, Ziem J, Yelifari L, Polderman AM, Van Lieshout L. Simultaneous Detection and Quantification of *Ancylostoma duodenale*, *Necator americanus*, and *Oesophagostomum bifurcum* in Fecal Samples Using Multiplex Real-Time PCR. Am J Trop Med Hyg. 2007; 77: 685–690.

24. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a Multiplex Quantitative PCR to Assess Prevalence and Intensity Of Intestinal Parasite Infections in a Controlled Clinical Trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

25. Mejia R, Vicuña Y, Broncano N, Sandoval C, Vaca M, Chico M, et al. A Novel, Multi-Parallel, Real-Time Polymerase Chain Reaction Approach for Eight Gastrointestinal Parasites Provides Improved Diagnostic Capabilities to Resource-Limited At-Risk Populations. Am J Trop Med Hyg. 2013; 88: 1041–1047.

26. Pilotte N, Papaiakovou M, Grant JR, Bierwert LA, Llewellyn S, McCarthy JS, et al. Improved PCR-Based Detection of Soil Transmitted Helminth Infections Using a Next-Generation Sequencing Approach to Assay Design. PLoS Negl Trop Dis. 2016; 10: e0004578.

27. Cimino RO, Jeun R, Juarez M, Cajal PS, Vargas P, Echazú A, et al. Identification of human intestinal parasites affecting an asymptomatic peri-urban Argentinian population using multi-parallel quantitative real-time polymerase chain reaction. Parasit Vectors. 2015; 8: 380.

28. Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

29. van Mens SP, Aryeetey Y, Yazdanbakhsh M, van Lieshout L, Boakye D, Verweij JJ. Comparison of real-time PCR and Kato smear microscopy for the detection of hookworm infections in three consecutive faecal samples from schoolchildren in Ghana. Trans R Soc Trop Med Hyg. 2013; 107: 269–271.

30. Lim MD, Brooker SJ, Jr BelizarioVY Jr, Gay-Andrieu F, Gilleard J, Levecke B, et al. Diagnostic tools for soil-transmitted helminths control and elimination programs: A pathway for diagnostic product development. PLoS Negl Trop Dis. 2018; 12: e0006213.

31. Easton AV, Oliveira RG, Walker M, O'Connell EM, Njenga SM, Mwandawiro CS, et al. Sources of variability in the measurement of *Ascaris lumbricoides* infection intensity by Kato-Katz and qPCR. Parasit Vectors. 2017; 10: 256.

32. Arnold BF, Null C, Luby SP, Unicomb L, Stewart CP, Dewey KG, et al. Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. BMJ Open. 2013; 3: e003476.

33. Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, et al. Effects of water quality, sanitation, handwashing and nutritional interventions on diarrhoea and child growth in rural Bangladesh: A cluster randomised trial. Lancet Glob Health. 2018; 6: e302–e315.

34. Ercumen A, Benjamin-Chung J, Arnold BF, Lin A, Hubbard AE, Stewart C, et al. Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: a cluster-randomized controlled trial in rural Bangladesh. PLoS Negl Trop Dis. 2019; 13: e0007323.

35. Benjamin-Chung J, Nazneen A, Halder AK, Haque R, Siddique A, Uddin MS, et al. The Interaction of Deworming, Improved Sanitation, and Household Flooring with Soil-Transmitted Helminth Infection in Rural Bangladesh. PLoS Negl Trop Dis. 2015; 9: e0004256.

36. Deer DM, Lampel KA, González-Escalona N. A versatile internal control for use as DNA in real-time PCR and as RNA in real-time reverse transcription PCR assays. Lett Appl Microbiol. 2010; 50: 366–372.

37. Papaiakovou M, Pilotte N, Baumer B, Grant J, Asbjornsdottir K, Schaer F, et al. A comparative analysis of preservation techniques for the optimal molecular detection of hookworm DNA in a human fecal specimen. PLoS Negl Trop Dis. 2018; 12: e0006130.

38. Pilotte N, Maasch JRMA, Easton AV, Dahlstrom E, Nutman TB, Williams SA. Targeting a highly repeated germline DNA sequence for improved real-time PCR-based detection of *Ascaris* infection in human stool. PLoS Negl Trop Dis. 2019; 13: e0007593.

39. Papaiakovou M, Pilotte N, Grant JR, Traub RJ, Llewellyn S, McCarthy JS, et al. A novel, species-specific, real-time PCR assay for the detection of the emerging zoonotic parasite *Ancylostoma ceylanicum* in human stool. PLoS Negl Trop Dis. 2017; 11: e0005734.

40. Liu J, Gratz J, Amour C, Kibiki G, Becker S, Janaki L, et al. A Laboratory-Developed TaqMan Array Card for Simultaneous Detection of 19 Enteropathogens. J Clin Microbiol. 2013; 51: 472–480.

41. Caraguel CGB, Stryhn H, Gagné N, Dohoo IR, Hammell KL. Selection of a Cutoff Value for Real-Time Polymerase Chain Reaction Results to Fit a Diagnostic Purpose: Analytical and Epidemiologic Approaches. J Vet Diag Invest. 2011; 23: 2-15.

42. Schmaedick MA, Koppel AL, Pilotte N, Torres M, Williams SA, Dobson SL, et al. Molecular Xenomonitoring Using Mosquitoes to Map Lymphatic Filariasis after Mass Drug Administration in American Samoa. PLoS Negl Trop Dis. 2014; 8: e3087.

43. WHO Expert Committee. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. World Health Organ Tech Rep Ser. 2002; 912: i–vi, 1–57.

44. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS ONE. 2009; 4: e6372.

45. Tanaka R, Hino A, Tsai IJ, Palomares-Rius JE, Yoshida A, Ogura Y, et al. Assessment of Helminth Biodiversity in Wild Rats Using 18S rDNA Based Metagenomics. PLoS ONE. 2014; 9: e110769.

46. Hino A, Maruyama H, Kikuchi T. A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. Parasitol Int. 2016; 65: 572–575.

47. Vestheim H, Jarman SN. Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. Front Zool. 2008; 5: 12.

48. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. Reproducible, Interactive, Scalable, and Extensible Microbiome Data Science Using QIIME 2. Nat Biotechnol 2019; 37: 852-857.

49. Freedman DA. On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors." The American Statistician. 2006; 60: 299–302.

50. Imbens GW, Rubin DB. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. New York, NY, USA: Cambridge University Press; 2015.

51. O'Connell EM, Nutman TB. Molecular Diagnostics for Soil-Transmitted Helminths. Am J Trop Med Hyg. 2016; 95: 508–513.

52. Dendukuri N, Joseph L. Bayesian Approaches to Modeling the Conditional Dependence Between Multiple Diagnostic Tests. Biometrics. 2001; 57: 158–167.

53. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev Vet Med. 2005; 68: 145–163.

54. Assefa LM, Crellen T, Kepha S, Kihara JH, Njenga SM, Pullan RL, et al. Diagnostic Accuracy and Cost-Effectiveness of Alternative Methods for Detection of Soil-Transmitted Helminths in a Post-Treatment Setting in Western Kenya. PLoS Negl Trop Dis. 2014; 8: e2843.

55. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988; 41: 923–937.

56. Joseph L, Gyorkos TW, Coupal L. Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard. Am J Epidemiol. 1995; 141: 263–272.

57. Knopp S, Speich B, Hattendorf J, Rinaldi L, Mohammed KA, Khamis IS, et al. Diagnostic Accuracy of Kato-Katz and FLOTAC for Assessing Anthelmintic Drug Efficacy. PLoS Negl Trop Dis. 2011; 5: e1036.

58. Levecke B, Behnke JM, Ajjampur SSR, Albonico M, Ame SM, Charlier J, et al. A Comparison of the Sensitivity and Fecal Egg Counts of the McMaster Egg Counting and Kato-Katz Thick Smear Methods for Soil-Transmitted Helminths. PLoS Negl Trop Dis. 2011; 5: e1201.

59. Knopp S, Mgeni AF, Khamis IS, Steinmann P, Stothard JR, Rollinson D, et al. Diagnosis of Soil-Transmitted Helminths in the Era of Preventive Chemotherapy: Effect of Multiple Stool Sampling and Use of Different Diagnostic Techniques. PLoS Negl Trop Dis. 2008; 2: e331.

60. Booth M, Vounatsou P, N'Goran EK, Tanner M, Utzinger J. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. Parasitology. 2003; 127: 525–531.

61. Verweij JJ. Application of PCR-based methods for diagnosis of intestinal parasitic infections in the clinical laboratory. Parasitol. 2014; 141: 1863–1872.

62. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent Class Models in Diagnostic Studies When There is No Reference Standard—A Systematic Review. Am J Epidemiol. 2014; 179: 423–431.

63. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10: 325–337.

64. Dacombe RJ, Crampin AC, Floyd S, Randall A, Ndhlovu R, Bickle Q, et al. Time delays between patient and laboratory selectively affect accuracy of helminth diagnosis. Trans R Soc Trop Med Hyg. 2007; 101: 140–145.

65. Walden J. Parasitic diseases. Other roundworms. *Trichuris*, hookworm, and *Strongyloides*. Prim Care. 1991; 18: 53–74.

66. Fischer K, Gankpala A, Gankpala L, Bolay FK, Curtis KC, Weil GJ, et al. *Capillaria* Ova and Diagnosis of *Trichuris trichiura* Infection in Humans by Kato-Katz Smear, Liberia. Emerg Infect Dis. 2018; 24: 1551–1554.

67. Ralph A, O'Sullivan MVN, Sangster NC, Walker JC. Abdominal pain and eosinophilia in suburban goat keepers - trichostrongylosis. Med J Aust. 2006; 184: 467–469.

68. Tarafder MR, Carabin H, Joseph L, Balolong E, Olveda R, McGarvey ST. Estimating the sensitivity and specificity of Kato-Katz stool examination technique for detection of hookworms, *Ascaris lumbricoides* and *Trichuris trichiura* infections in humans in the absence of a 'gold standard.' Int J Parasitol. 2010; 40: 399–404.

69. Garcia LS. Diagnostic Medical Parasitology. American Society for Microbiology Press; 2006.

70. Colmer-Hamood JA. Fecal Microscopy Artifacts Mimicking Ova and Parasites. Lab Med. 2001; 32: 80–84.

71. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016; 8.

72. Bermingham ML, Handel IG, Glass EJ, Woolliams JA, Bronsvoort BM de C, McBride SH, et al. Hui and Walter's latent-class model extended to estimate diagnostic test properties from surveillance data: a latent model for latent data. Scientific Reports. 2015; 5: 11861.

73. Kim M-K, Pyo K-H, Hwang Y-S, Park KH, Hwang IG, Chai J-Y, et al. Effect of Temperature on Embryonation of *Ascaris suum* Eggs in an Environmental Chamber. Korean J Parasitol. 2012; 50: 239–242.

74. Anderson RM, May RM. Helminth Infections of Humans: Mathematical Models, Population Dynamics, and Control. Adv Parasitol. 1985; 24: 1-101.

75. Knopp S, Salim N, Schindler T, Voules DAK, Rothen J, Lweno O, et al. Diagnostic Accuracy of Kato–Katz, FLOTAC, Baermann, and PCR Methods for the Detection of Light-Intensity Hookworm and *Strongyloides stercoralis* Infections in Tanzania. Am J Trop Med Hyg. 2014; 90: 535–545.

76. Papaiakovou M, Gasser RB, Littlewood DTJ. Quantitative PCR-Based Diagnosis of Soil-Transmitted Helminth Infections: Faecal or Fickle? Trends Parasitol. 2019; 35: 491–500.

77. Speich B, Ali SM, Ame SM, Albonico M, Utzinger J, Keiser J. Quality control in the diagnosis of *Trichuris trichiura* and *Ascaris lumbricoides* using the Kato-Katz technique: experience from three randomised controlled trials. Parasit Vectors. 2015; 8: 82.

78. Sinniah B. Daily egg production of *Ascaris lumbricoides*: the distribution of eggs in the faeces and the variability of egg counts. Parasitology. 1982; 84: 167–175.

79. Turner HC, Bettis AA, Dunn JC, Whitton JM, Hollingsworth TD, Fleming FM, et al. Economic Considerations for Moving beyond the Kato-Katz Technique for Diagnosing Intestinal Parasites As We Move Towards Elimination. Trends Parasitol. 2017; 33: 435–443.

80. Ercumen A, Benjamin-Chung J, Arnold BF, Lin A, Hubbard AE, Stewart C, et al. Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: a cluster-randomized controlled trial in rural Bangladesh. PLoS Negl Trop Dis. 2019; 13: e0007323.

# CHAPTER 6

# A CASE FOR USING GENOMICS AND A BIOINFORMATICS PIPELINE TO DEVELOP SENSITIVE AND SPECIES-SPECIFIC PCR-BASED DIAGNOSTICS FOR SOIL-TRANSMITTED HELMINTHS

*as written by Grant J.R., Pilotte N., Williams S.A.*
*Frontiers in Genetics, 2019*

## 6.1 General Chapter Overview

This chapter provides an overview of the merits of utilizing methodically and intelligently designed assay for the real-time PCR-based detection of soil-transmitted helminths. Providing a brief summary of the growing list of studies which have employed the assays that we have developed, examples of the advantages of utilizing sensitive and species-specific assays are detailed. Through a recounting of these individual examples, instances in which our assays have been instrumental in disentangling and/or elucidating otherwise confounded data are provided. Furthermore, a description of the methods used to develop these assays is given, as is a brief explanation of the software utilized for assay-target identification.

My contributions to this chapter include compiling details on all studies which have employed our assays. I have been personally involved in nearly all of these studies, either as an advisor, or directly as an investigator. I was also responsible for generating the figure and for writing the portions of the paper that detailed the RepeatExplorer and Tandem Repeat Analyzer (TAREAN) software platforms.

## 6.2 Abstract

The balance of expense and ease of use vs. specificity and sensitivity in diagnostic assays for helminth disease is an important consideration, with expense and ease often winning out in endemic areas where funds and sophisticated equipment may be scarce. In this review we argue that molecular diagnostics, specifically new assays that have been developed with the aid of next generation sequence data and robust bioinformatic tools, more than make up for their expense with the benefit of a clear and precise assessment of the situation on the ground. Elimination efforts associated with the London Declaration and the World Health Organization (WHO) 2020 Roadmap have resulted in areas of low disease incidence and reduced infection burdens. An accurate assessment of infection levels is critical for determining where and when the programs can be successfully ended. Thus, more sensitive assays are needed in locations where elimination efforts are approaching a successful conclusion. While microscopy or more general PCR targets have a role to play, they can mislead and cause study results to be confounded. Hyper-specific qPCR assays enable a more definitive assessment of the situation in the field, as well as of shifting dynamics and emerging diseases.

## 6.3 Introduction

Parasitic worms impact the health and economic well-being of billions of people worldwide. Soil-transmitted helminths (STH) are a burden in the tropics and subtropics, and contribute to an estimated 1.9 - 2.1 million Disability-Adjusted Life Years (DALYs) and $7.5 billion to $138.9 billion in loss of productivity [1, 2]. Efforts are underway to eliminate STH, with the ambitious goal of controlling morbidity by the year 2020 [3, 4].

Mass drug administrations (MDA) and water, sanitation and hygiene (WASH) programs across endemic countries are making headway [5-8], but with 2020 fast approaching, there are still many challenges to reaching this goal. An important concern is where to enact and when to cease MDA. This depends on accurately mapping the current burden in communities [9]. Sensitive, species-specific diagnostics are critical to properly evaluating the success of these programs, as well as addressing where to focus efforts and when interventions can be ended [6].

Diagnostic techniques need to be inexpensive, practical, and give consistent results across technicians and labs. Importantly, they must be accurate, sensitive, and easily interpreted. Microscopy has long been relied upon as the standard for diagnosis of intestinal parasites including soil transmitted helminths [10]. Several copromicroscopic methods are in use, including FLOTAC [11], MINI-FLOTAC [12], several modifications of the McMaster technique [13], and Kato-Katz [14] (KK). Of these, KK is the most commonly used for STH diagnosis because it is relatively easy to perform in the field and is generally more sensitive than other microscopic methods [15]. With any of these tests, even highly-trained microscopists can misidentify species or give inconsistent results [16], and they are notoriously insensitive in regions with low infection rates [17-20].

Molecular diagnostics have been garnering more interest in the last few years, as their superior sensitivity has been proven and their acceptance by the research community has increased [21-23]. However, as with any advance, there are technical problems to overcome. For example, DNA extraction efficiency and preservation of samples prior to testing will affect diagnostic reliability [24-27]. Notably, *Trichuris trichiura* eggs are notoriously difficult to break open and this impacts the sensitivity of molecular assays, but

techniques are being developed and improved to the point where consistently good results are achievable [28-30]. While molecular diagnostics are not inexpensive, microscopy techniques are also expensive and can be difficult to scale up, while the costs of qPCR have the potential to decrease, as studies show that multiple technical replicates may not be crucial and other cost-cutting measures such as cheaper, more effective sample preservation and pooling are explored [8, 24, 31]. Until recently, most PCR-based assays have targeted well-characterized and conserved regions such as ITS and 18S [32, 33] but increased availability of whole genome sequence data is facilitating the discovery of more sensitive and species-specific genomic targets [34, 35].

Repetitive elements are essential parts of eukaryotic genomes that have structural and regulatory functions [36, 37] and different types of repetitive DNA elements have been studied and classified [37-40]. Ribosomal DNA is found in repeat arrays [41, 42]. These have traditionally been used for primer design, and can give sensitive results depending on the size of the array [33]. However, the repeat is conserved between species and even genera, and rDNA-based assays are often less specific than those designed from other repeat types [34, 43].

Tandemly repeated DNA is classified by the size of the repeated monomer, resulting in microsatellites (< 9 bp), minisatellites (< 15 bp in arrays of 0.5-30 kb), and satellites (satDNA, up to ~200 bp per monomer, in megabase-sized arrays) that are generally enriched within the centromeric, pericentromeric and subtelomeric regions of the chromosome [37]. Copy number can be quite variable in mini- and micro-satellites but larger satDNA monomers are more consistent within species [37]. Microsatellites and minisatellites are not useful for assay design, as the repeats tend to be too short to allow for

primer/probe design. Larger satDNA monomers, on the other hand, offer the best options for assay design, in that the repeat monomers are an optimal size for qPCR, they are extremely abundant, and the copy number is relatively stable within species [37].

The other major types of repeat element include transposons and retrotransposons which are dispersed throughout the genome. These include Short Interspersed Nuclear Elements (SINEs) which are 100-500 bases long, and Long Interspersed Nuclear Elements (LINEs) which are larger – 6000 or 7000 bases long. These may also be useful for repeat-based assay design [44].

The amount of repetitive DNA in any eukaryotic species is variable and can make up quite a large percentage of the genome. A recent study of parasitic worms revealed that both genome size and repeat content of the genomes range widely, with repeat elements forming up to 37% of the genomes in STH of interest (Table 6.1). Repetitive elements make up even greater percentages in other eukaryotes [36, 45-47], up to an astonishing 97% in some plants [48, 49]. These repetitive elements, because of their abundance in the genome, provide targets for molecular assays of exquisite sensitivity. In addition, since many appear unbound by selective pressure, they can be highly species-specific. This combination of improved sensitivity and specificity makes repetitive elements a prime target for molecular diagnostics. However, as mentioned above, some forms of repetitive DNA, such as simple short nucleotide repeats, will be unsuitable for assay design. Another potential stumbling block is sequence variation in the repeat itself as polymorphism in the primer and probe sites will decrease the sensitivity of the assay. However, while the repeats we have targeted have not been specifically investigated, maintenance of homogeneity in repetitive elements by concerted evolution has been discussed in relation to other repeats

in other species [50, 51]. Concerted evolution of repeats conserves the sequence within a species, while allowing significant heterogeneity between species. There can, of course, be some variation within repeats [52, 53] which could lessen the sensitivity of repeat-based molecular assays. Copy number variation of repeats between individuals may also impact the sensitivity of these assays in some populations. Studies have found population-level copy number variation in ribosomal repeats of different species [54-57] but there is a suggestion that there are both copy-number variable-type repeats and constant-type repeats, whose copy number is consistent within species [58]. An understanding of copy number variation and its impact on assay sensitivity will likely need to be studied on an individual basis.

The value of repeats as diagnostic tools has been understood for some time but these sequences were more difficult to find in the pre-genomics era [59-66]. Now, abundant genomic data and robust bioinformatics tools are available to make these targets easier to identify and use in developing PCR-based assays. The pipeline leading from low coverage NGS data to hyper-sensitive and specific qPCR assay is not overly complicated or time-consuming, and the ability to repurpose low coverage NGS data from other studies makes this an attractive option for diagnostic development for helminthology and many other fields.

### 6.4 Findings

Diagnostics give an estimate of the true prevalence of a disease, with the probability of correctly estimating the truth given by the sensitivity of the diagnostic. WHO guidelines for when to treat a community are informed by the prevalence of disease in that community,

and are thus influenced by the sensitivity of the diagnostic used. A study modeling the probability of making the correct treatment decisions given WHO guidelines for treatment and varying the true prevalence and diagnostic sensitivity, shows that there is a significant difference in outcome when more sensitive diagnostics are used [67]. Medley et al (2016) found it especially true in areas of intermediate true prevalence (between 30 and 50 %). More sensitive tests allowed the correct treatment decision to be made more often in intermediate cases.

The aforementioned study measured outcome by looking at DALYs and found that these were not as influenced by diagnostic sensitivity in low or high prevalence areas. However, there are other reasons to prefer more sensitive tests in low prevalence areas as well. As the goals of the WHO elimination programs are reached, there will be pressure to reallocate the funds spent on MDA to other programs. A well-defined threshold under which recrudescence will not occur is critical to preventing the reoccurrence of disease after the completion of MDA. Restarting such programs would be extremely difficult and expensive once they have ended. Modeling has shown that the threshold must be based on true prevalence [68, 69] which can only be accurately estimated with highly sensitive diagnostics. Improved diagnostics are crucial to meet this need [26, 27, 71, 72]. Post-MDA surveillance is also necessary. With highly sensitive diagnostics, the reappearance of disease can be recognized and addressed before infection levels rise, increasing the probability of controlling the recrudescence [73]. In addition to testing human populations for infection, vectors that transmit helminths (or other parasites) or intermediate hosts can be screened to track disease prevalence in the community without the need for taking human samples [74-77]. Pooling of samples is a common way to decrease cost so

diagnostic tests used for the screening of pooled samples need to be highly sensitive [78-80], especially when the infection level is low, so as not to miss positive results in dilute samples.

In recent years, efforts to sequence nematode genomes by groups such as the International Helminth Genomes Consortium have made great strides in increasing the availability of helminth genomic sequence data [81]. While many of the genomes are in draft form, this is sufficient for probing the genome for species-specific repeats. Our method for recovering highly repetitive sequence from low depth, raw, short-read genome sequence data uses the Galaxy-based tool RepeatExplorer [34, 82, 83]. Originally used to investigate repeat sequences in plant genomes, RepeatExplorer takes as input short-read next generation sequence data, and creates graph-based clusters based on the similarity of the sequences. In these graphs (see Figure 6.1), each read is represented by a node and each sequence pair (by default defined as $\geq$ 90% identity over at least 55% of the read lengths) is represented by an edge. The density of the graph represents the number and similarity of reads in the cluster. In low-depth sequence data sets, low copy number sequences will not be well-represented and will therefore graph as individual nodes or small clusters, while high copy number repetitive sequences will be found in dense clusters. The number of reads in a given cluster, combined with the structure and density of that cluster can be used as a proxy for the representation of the number of repeats of that sequence in the genome. The depth of read coverage will impact the sensitivity of the assay, since each copy in the genome will be an additional target for the assay. In addition, RepeatExplorer and its sister software TAREAN [84], provide information on the count of individual nucleotides in the repeat contigs. These counts can be used to design assays to

the most conserved regions, limiting the problems associated with intra-genomic variation. Primer/probe qPCR assays targeting sequences discovered using this technique have been shown to amplify as little as 20 ag of genomic DNA, or less than the amount of DNA found in a single egg [34]. Care is needed in choosing which repeat to select for use as a diagnostic target, as some may be found in closely related species [85, 86], however, similar to what has been reported in the literature [87, 88], we have found that many repetitive sequences are species-specific. There may be times when a more general assay – one that will amplify several species of the same genus, for example – may be desired [86]. A careful search of RepeatExplorer output can often reveal both species- and genera-specific targets.

Diagnostics based on targets discovered using this technique have proven useful in both past, and ongoing field tests in Kenya [89], Bangladesh [93], Ethiopia, Uganda, Timor Leste [34, 35], Thailand [43], Liberia [90], Japan, Benin, Malawi, India, and the Southern US, and have been adopted for use by large operational research efforts such as the DeWorm3 cluster randomized trials [91]. However, testing biological samples, whether for diagnosing individuals or getting an overview of the epidemiological environment of a region, involves myriad factors such as unexpected or emerging parasites, zoonotic infections, and misleading material in the samples. While most of the criticism of the KK technique has been on the lack of sensitivity and potential for missed infection, there is also the risk of false positives, for example, mistaking other material in stool as eggs [18]. Some fecal elements may resemble parasite ova, depending on environmental or dietary factors. Confounding elements may include pollen grains, fungal spores, diatoms, or any number of items. An entire chapter in Ash and Orihel's "Atlas of Human Parasitology" is

dedicated to artifacts in fecal samples that can mislead copromicroscopic diagnostics [92]. Thus, this problem is more frequent than many researchers realize or acknowledge; what follows are some examples demonstrating the importance of this issue.

A field study comparing KK with repeat-based qPCR in Bangladesh [93] found that hookworm species and *Trichuris trichiura* prevalence, as measured by qPCR, was significantly higher than was measured by KK. This was expected, given the greater sensitivity of the qPCR assays and the results of many previous studies comparing KK and PCR [31, 34, 94-97]. For *Ascaris*, however, prevalence as measured by KK was significantly higher than by qPCR. This surprising result was investigated further, both by qPCR targeting a different part of the *Ascaris* genome, and also by amplicon sequencing targeting the 18S gene of all eukaryotes in several of the KK positive/qPCR negative samples. All of the samples that were positive by KK but negative by the initial qPCR assay were also negative using the second qPCR target. Additionally, the 18S amplicon sequencing revealed no *Ascaris* in these samples, but did find it in the control samples that were positive by both KK and qPCR. There was no one organism found in the amplicon sequencing that could explain all of the false-positive results; what had confounded the microscopists is still unknown, but there was no evidence by any of the molecular assays that *Ascaris* was present in the samples. Had the study relied upon copromicroscopic results alone, the conclusion would have been that MDA or WASH interventions were less effective as an *Ascaris* intervention than they likely were, since the true prevalence of the parasite was in fact much lower than was measured by KK.

In a similar case, higher than expected rates of hookworm were noted in a survey of children in rural Rwanda. Further investigation suggested these results may have been

confounded by *Caenorhabditis elegans* eggs [98]. Additional examples of misidentification of hookworm ova as other eggs [99-101] show that such confusion may be a more common problem than previously thought. Thus, relying solely on microscopy may be misleading in some instances.

Discrepancies can occur between molecular assays as well, since some PCR targets are less species-specific than others. In developing our pipeline for repeat-based primer discovery, a previously published qPCR assay targeting the internal transcribed spacer region was compared against our newly developed assay targeting an *Ancylostoma duodenale* species-specific repeat [34, 102]. Surprisingly, the repeat-based assay failed to detect any of the samples that were determined to be positive for *A. duodenale* by the ITS-based assay. A previously published semi-nested PCR assay [103, 104] and Sanger sequencing later determined that the discordant results were due to all of the infections being the zoonotic species *Ancylostoma ceylanicum*. The ITS of these two species is highly conserved in the region targeted by the original qPCR assay and so the ITS-based assay did not distinguish between these closely related species. The repeat-based assay, on the other hand, only detected *A. duodenale*. We have since used our pipeline to develop a species-specific qPCR assay for *A. ceylanicum* [35], which is more specific than the ITS-based assay and easier to use than semi-nested PCR.

In a similar case, the same ITS-based primer set [102] was used to investigate *Ancylostoma duodenale* in a field study of a refugee population in Thailand [43]. Since the most common human *Ancylostoma* parasite is *A. duodenale*, the results were initially believed to indicate *A. duodenale* infection. Again, however, a corroboratory qPCR targeting the highly specific *A. duodenale* repeat failed to detect any *A. duodenale*. Use of

the *A. ceylanicum*-specific assay [35], as well as confirmation with semi-nested PCR and Sanger sequencing, revealed that all of these infections were, in fact, caused by *A. ceylanicum* and not *A. duodenale*. In this case, while the more general ITS-based assay misdiagnosed the species causing the infection, the more specific assay for the expected parasite (*A. duodenale*) would have missed the infection. This highlights a risk of using extremely specific qPCR assays in the field if the precise parasite community is unknown.

The specificity of *Trichuris trichiura* (whipworm) detection by microscopy is assumed to be high, given the relatively distinct morphology of *Trichuris* eggs. However, there are several species of *Trichuris*, including some that infect companion or farm animals. Distinction between species of *Trichuris* relies on size differentiation, but there are overlaps in some species making misdiagnosis by microscopy possible. The most common species in humans is *T. trichiura*, but *T. suis*, which commonly infects pigs, and *T. vulpis*, usually found in dogs, have also been found infecting humans [105, 106]. Microscopy and genera-specific qPCR assays may be confounded by these zoonotic species. Discordance between the ITS-based *Trichuris* qPCR assay and the repeat-based qPCR assay has been noted in one study where all but one of the discordant results were later shown to be *Trichuris ovis*, a species found in sheep and goats that has not been known to infect humans. Whether this is evidence of human infection or merely false positives due to close contact with infected animals or ingesting food contaminated with animal feces is an open question. In these cases, qPCR targeting highly species-specific repetitive targets alone could easily miss the prevalence of zoonotic infection, leading to a misunderstanding of the health of the population.

Despite the distinct morphology of *Trichuris*, the potential of misidentifying unexpected infection with zoonotic species is not the only risk when relying solely on microscopy for investigating whipworm. A study of STH in two regions in Liberia that used both microscopy and qPCR found discordance between the tests for *T. trichiura* in one of the regions [90], with 25 of 27 putatively positive samples for *T. trichiura*, as determined by KK, being negative by qPCR. In the second region, the agreement between the two tests was high. In this case, the discrepancies were investigated further and the eggs were determined to be a species of *Capillaria*, a human parasite that is associated with eating raw fish and not expected in this region. Reinvestigation by microscopy in this case, elucidated subtle differences in the eggs found in the KK positive, qPCR negative samples. A microscopist highly trained and expecting to have to differentiate between two extremely similar egg morphologies could have noticed the difference and provided a correct result. But, since *Capillaria* had not previously been reported in Liberia, the eggs that looked like *Trichuris* were reported as such. Without the complementary qPCR, this mistake would not have been discovered. This study also found discrepancies between the microscopy and qPCR measuring the prevalence of *Ascaris lumbricoides* in the same region where *Capillaria* was discovered. Surprisingly, these samples were also determined to be confounded by the presence of *Capillaria*, which can appear rounder and have more subtle polar plugs, leading to its misidentification as the eggs of *A. lumbricoides*. These examples highlight the tendency for microscopists to sometimes see what they are looking for. In this case, the unexpected presence of *Capillaria* eggs misled the microscopists and would have resulted in significant misinterpretation of the distribution of STH species in the study region.

## 6.5 Conclusions

The elimination of STH is a worthy goal, given the distress and disability they cause to a large portion of the global population. The goal is also attainable, but will not be easily reached. Monitoring and evaluation of progress is critical and depends on highly accurate reporting. Species-specific repeat-based target discovery and qPCR assays deliver this accuracy. Microscopy and genera-specific molecular assays have a place in this effort, especially in surveys where full mapping of parasite diversity has not occurred. These tools, used in conjunction with highly sensitive and specific molecular assays targeting repetitive elements, can give a clear and accurate assessment where one tool alone could yield misleading results.

New target development is fairly easy, since the web-based bioinformatics tool RepeatExplorer provides output in a manner that makes finding repeat-based qPCR targets straightforward. Only a skim of the genome is necessary, so the data needed to develop new assays is already available for many species, and is fairly inexpensive to produce for a species whose genome has not yet been sequenced. We have used this technique to explore new diagnostics for other nematode and protist parasites, and feel the pipeline for assay development has potential for improving diagnostic sensitivity in many other disciplines.

Correct identification of species is of interest if one wants to understand evolution, biogeography and emerging disease. The treatment for infection with one species of helminth is often the same as for another, however, we would argue that lumping all species of a genus together is careless. It is not clear that different species respond the same way under the same drug treatment and misidentification of species in the past might be

confounding the results of studies of resistance or drug response based on microscopy alone. Different species within a genus may also vary in their capacity for animal infection, resulting in some parasites having animal reservoirs while others remain obligate human pathogens. A more detailed understanding of the underlying community structure will offer crucial insight into subjects such as anthelminthic resistance, emerging or zoonotic diseases, and optimal threshold levels for elimination of disease.

### 6.6 Acknowledgements

This work was supported in part by the Bill and Melinda Gates Foundation, the DeWorm3 project through a grant to the Natural History Museum, London from the Bill and Melinda Gates Foundation, and by the Task Force for Global Health. The authors would like to thank Marina Papaiakovou for helpful conversations, as well as Dr. Judd Walson and Dr. D. Timothy J. Littlewood (DeWorm3 [91]). We also thank Dr. Eric Ottesen and Dr. Patrick Lammie (Task Force for Global Health, Decatur, GA, USA) for their ongoing support and advice.

### 6.7 Bioinformatics resources

RepeatExplorer [82, 83] is made freely available here:

http://repeatexplorer.org/

**Table 6.1 Genome size of representative helminth species and the amount/percent of genome masked as repetitive.**

| Species | Assembly size (Mb) | Repeat-masked (Mb) | Repeat % | % of assembly that is low complexity/simple repeats[#] |
|---|---|---|---|---|
| *Ancylostoma ceylanicum* [32] | 349 | 128.6 | 36.8 | 0.5 |
| *Ancylostoma duodenale* [32] | 332.9 | 116.1 | 34.9 | 0.7 |
| *Ascaris lumbricoides* germline [79] | 334 | n/a | 16.8 | n/a |
| *Ascaris lumbricoides* soma [79] | 291 | n/a | 7 | 1.5 |
| *Necator americanus* [32, 78] | 244.1 | 67.1 | 27.5 | 1.2 |
| *Strongyloides stercoralis* [32, 80] | 42.7 | 4.4 | 10.3 | 4.2 |
| *Trichuris trichiura* [32] | 75.5 | 18.4 | 24.4 | 0.4 |

[#]Low complexity repeats and simple (for example di- or tri-nucleotide repeats) are not targets for molecular assays.

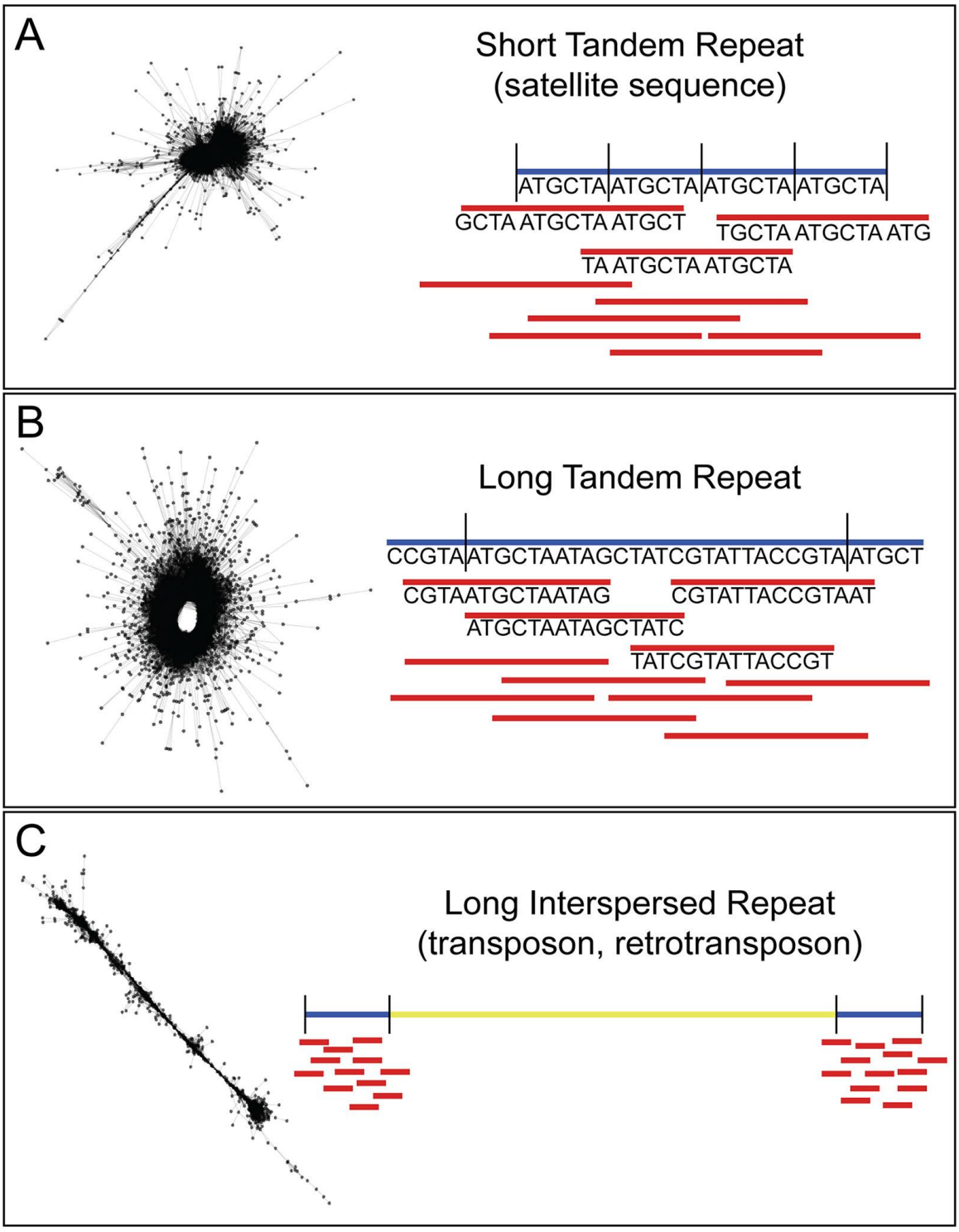


**Figure 6.1 Description of RepeatExplorer cluster output**. (A) Short tandem repeats, including many satellite sequences, form characteristic "star burst"-type clusters. Because they are tandem, and such repeats are of similar size or shorter than the length of an individual sequence read, a very high percentage of reads within the cluster meet the RepeatExplorer-defined criteria for pair formation. This results in each read successfully

pairing with a very high percentage of the other reads assigned to the cluster. Because nearly all of the reads within the cluster are paired with nearly all of the other reads within this same cluster, a compact network of very short edges forms between reads. This in turn generates a very dense cluster with a core of paired reads possessing nearly identical sequences. If of sufficient length for assay design, the consensus sequences for these clusters make ideal diagnostic targets, as they contain the greatest number of repeats per read. (B) Long tandem repeats characteristically result in "doughnut"-like clusters. In such clusters, neighboring reads within the underlying scaffold meet the criteria for pair formation. However, because the length of the repeat monomers generating these clusters is longer, reads may be significantly shorter than the monomer itself. This results in many reads within the cluster that do not meet the criteria for pair formation as they map to different regions of the same monomer unit. Yet because they are tandemly arranged, reads spanning a repeat-repeat junction will meet the pair formation criteria, closing the sequence "loop" and resulting in a "doughnut"-shaped cluster. (C) Long interspersed repeats, such as transposable elements form characteristic "line"-type clusters. While reads neighboring each other in the underlying scaffold meet the criteria for pair formation, the extended length of a repeat monomer means that distant reads within a single monomer will not meet this threshold. This results in similar pairings to those seen in clusters generated from long tandem repeats. However, because these elements are interspersed, reads do not span repeat-repeat junctions, so a "loop" is not formed, and clusters attain a linear appearance.

## 6.8 References


1. Bartsch SM, Hotez PJ, Asti L, Zapf KM, Bottazzi ME, Diemert DJ, et al. The Global Economic and Health Burden of Human Hookworm Infection. PLoS Negl Trop Dis. 2016; 10, e0004922.

2. Kyu, HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392; 1859–1922.

3. Becker SL, Liwanag HJ, Snyder JS, Akogun O, Belizario Jr V, Freeman MC, et al. Toward the 2020 goal of soil-transmitted helminthiasis control and elimination. PLoS Negl Trop Dis. 2018; 12: e0006606.

4. London Declaration on Neglected Tropical Diseases. Uniting to Combat NTDs. Available at: /london-declaration-neglected-tropical-diseases/. (Accessed: 9th April 2019)

5. Hicks JH, Kremer M, Miguel E. The Case for Mass Treatment of Intestinal Helminths in Endemic Areas. PLoS Negl Trop Dis. 2015; 9: e0004214.

6. Weatherhead JE, Hotez PJ, Mejia R. The Global State of Helminth Control and Elimination in Children. Pediatr Clin. 2017; 64: 867–877.

7. Truscott JE, Turner HC, Farrell SH, Anderson RM. Soil-Transmitted Helminths: Mathematical Models of Transmission, the Impact of Mass Drug Administration and Transmission Elimination Criteria. Adv. Parasitol. 2016; 94: 133–198.

8. Truscott JE, Dunn JC, Papaiakovou M, Schaer F, Werkman M, Littlewood DTJ, et al. Calculating the prevalence of soil-transmitted helminth infection through pooling of stool samples: Choosing and optimizing the pooling strategy. PLoS Negl Trop Dis 2019; 13: e0007196.

9. Children Without Worms. (2018). STH Coalition Action Group Meeting 2018 Report. Retrieved April 2019 from http://www.childrenwithoutworms.org/sites/default/files/ 2018%20Action%20Meeting%20Report_FINAL%2007.16.18.pdf

10. Beaver PC, Martin LK. Evaluation of Kato Thick-Smear Technique for Quantitative Diagnosis of Helminth Infections. Am J Trop Med Hyg. 1968; 17: 382–391.

11. Cringoli G, Rinaldi L, Maurelli MP, Utzinger J. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat Protoc. 2010; 5: 503–515.

12. Maurelli MP, Rinaldi L, Alfano S, Pepe P, Coles GC, Cringoli G, et al. Mini-FLOTAC, a new tool for copromicroscopic diagnosis of common intestinal nematodes in dogs. Parasit Vectors. 2014; 7: 356.

13. Mines J. Modifications of Mcmaster Worm Egg Counting Method. Aust Vet J. 1977; 53: 342–343.

14. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in *Schistosomiasis mansoni*. Rev Inst Med Trop Sao Paulo. 1972; 14: 397–400.

15. Moser W, Bärenbold O, Mirams GJ, Cools P, Vlaminck J, Ali SM, et al. Diagnostic comparison between FECPAKG2 and the Kato-Katz method for analyzing soil-transmitted helminth eggs in stool. PLoS Negl Trop Dis. 2018; 12: e0006562.

16. Krauth SJ, Coulibaly JT, Knopp S, Traoré M, N'Goran EK, Utzinger J. An In-Depth Analysis of a Piece of Shit: Distribution of *Schistosoma mansoni* and Hookworm Eggs in Human Stool. PLoS Negl Trop Dis. 2012; 6: e1969.

17. Soto LA, Santísma-Trinidad AB, Bornay-Llinares FJ, González MM, Valero JAP, Muñoz MR. Quantitative PCR and Digital PCR for Detection of *Ascaris lumbricoides* Eggs in Reclaimed Water. BioMed Res Int. 2017; 2017: 1–9.

18. Speich B, Ali SM, Ame SM, Albonico M, Utzinger J, Keiser J. Quality control in the diagnosis of *Trichuris trichiura* and *Ascaris lumbricoides* using the Kato-Katz technique: experience from three randomised controlled trials. Parasit Vectors. 2015; 8: 82.

19. Nikolay B, Brooker SJ, Pullan RL. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. Int J Parasitol. 2014; 44: 765–774.

20. Buonfrate D, Formenti F, Perandin F, Bisoffi Z. Novel approaches to the diagnosis of *Strongyloides stercoralis* infection. Clin Microbiol Infect. 2015; 21: 543–552.

21. Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

22. Holt DC, Shield J, Harris TM, Mounsey KE, Aland K, McCarthy JS, et al. Soil-Transmitted Helminths in Children in a Remote Aboriginal Community in the Northern Territory: Hookworm is Rare but *Strongyloides stercoralis* and *Trichuris trichiura* Persist. Trop Med Infect Dis. 2017; 2: 51.

23. Sow D, Parola P, Sylla K, Ndiaye M, Delaunay P, Halfon P, et al. Performance of Real-Time Polymerase Chain Reaction Assays for the Detection of 20 Gastrointestinal Parasites in Clinical Samples from Senegal. Am J Trop Med Hyg. 2017; 97: 173–182.

24. Papaiakovou M, Pilotte N, Baumer B, Grant J, Asbjornsdottir K, Schaer F, et al. A comparative analysis of preservation techniques for the optimal molecular detection of hookworm DNA in a human fecal specimen. PLoS Negl Trop Dis. 2018; 12: e0006130.

25. Sarhan RM, Kamel HH, Saad GA, Ahmed OA. Evaluation of three extraction methods for molecular detection of *Schistosoma mansoni* infection in human urine and serum samples. J Parasit Dis. 2015; 39: 499–507.

26. Andersen LO, Röser D, Nejsum P, Nielsen HV, Stensvold CR. Is Supplementary Bead Beating for DNA Extraction from Nematode Eggs by Use of the NucliSENS easyMag Protocol Necessary? J Clin Microbiol. 2013; 51: 1345–1347.

27. Hidalgo A, Melo A, Romero F, Hidalgo V, Villanueva J, Fonseca-Salamanca F. DNA extraction in *Echinococcus granulosus* and *Taenia* spp. eggs in dogs stool samples applying thermal shock. Exp Parasitol. 2018; 186: 10–16.

28. Kaisar MMM, Brienen EAT, Djuardi Y, Sartono E, Yazdanbakhsh M, Verweij JJ, et al. Improved diagnosis of *Trichuris trichiura* by using a bead-beating procedure on ethanol preserved stool samples prior to DNA isolation and the performance of multiplex real-time PCR for intestinal parasites. Parasitology. 2017; 144: 965–974.

29. Harmon AF, Zarlenga DS, Hildreth MB. Improved methods for isolating DNA from *Ostertagia ostertagi* eggs in cattle feces. Vet Parasitol. 2006; 135: 297–302.

30. Nunes CM, Lima LGF, Manoel CS, Pereira RN, Nakano MM, Garcia JF. Fecal Specimens Preparation Methods for PCR Diagnosis of Human Taeniosis. Rev Inst Med trop São Paulo. 2006; 48, 45–47.

31. Easton AV, Oliveira RG, Walker M, O'Connell EM, Njenga SM, Mwandawiro CS, et al. Sources of variability in the measurement of *Ascaris lumbricoides* infection intensity by Kato-Katz and qPCR. Parasit Vectors. 2017; 10: 256.

32. Hii SF, Senevirathna D, Llewellyn S, Inpankaew T, Odermatt P, Khieu V, et al. Development and Evaluation of a Multiplex Quantitative Real-Time Polymerase Chain Reaction for Hookworm Species in Human Stool. Am J Trop Med Hyg. 2018; 99: 1186–1193.

33. Verweij JJ, Stensvold, C. R. Molecular testing for clinical diagnosis and epidemiological investigations of intestinal parasitic infections. 2014; Clin Microbiol Rev. 2014; 27: 371–418.

34. Pilotte N, Papaiakovou M, Grant JR, Bierwert LA, Llewellyn S, McCarthy JS, et al. Improved PCR-Based Detection of Soil Transmitted Helminth Infections Using a Next-Generation Sequencing Approach to Assay Design. PLoS Negl Trop Dis. 2016; 10: e0004578.

35. Papaiakovou M, Pilotte N, Grant JR, Traub RJ, Llewellyn S, McCarthy JS, et al. A novel, species-specific, real-time PCR assay for the detection of the emerging zoonotic parasite *Ancylostoma ceylanicum* in human stool. PLoS Negl Trop Dis. 2017; 11: e0005734.

36. Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. Biol Rev. 2005; 80: 227–250.

37. López-Flores I, Garrido-Ramos MA. The Repetitive DNA Content of Eukaryotic Genomes. In: Genome Dynamics: Karger; 2012.

38. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. Chromosome Res. 2015; 23: 415–420.

39. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 1994; 371: 215.

40. Plohl M, Luchetti A, Mestrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. 2008; 409: 72–82.

41. Lafontaine DLJ, Tollervey D. The function and synthesis of ribosomes. Nat Rev Mol Cell Biol. 2001; 2: 514.

42. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007; 35: 7188–7196.

43. O'Connell EM, Mitchell T, Papaiakovou M, Pilotte N, Lee D, Weinberg M, et al. *Ancylostoma ceylanicum* Hookworm in Myanmar Refugees, Thailand, 2012-2015. Emerg Infect Dis. 2018; 24: 1472-1481.

44. Funakoshi K, Bagheri M, Zhou M, Suzuki R, Abe H, Akashi H. Highly sensitive and specific Alu-based quantification of human cells among rodent cells. Sci Rep. 2017; 7: 13202.

45. International Helminth Genomes Consortium, Coghlan A, Mitreva M, Berriman M. Comparative genomics of the major parasitic worms. Nat Genet. 2019; 51: 163-174.

46. Brindley PJ, Laha T, McManus DP, Loukas A. Mobile genetic elements colonizing the genomes of metazoan parasites. Trends Parasitol. 2003; 19: 79–87.

47. Wickstead B, Ersfeld K, Gull K. Repetitive Elements in Genomes of Parasitic Protozoa. Microbiol Mol Biol Rev. 2003; 67: 360–375.

48. Flavell RB, Bennett MD, Smith JB, Smith DB. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet. 1974; 12: 257–269.

49. Piednoël M, Aberer AJ, Schneeweiss GM, Makas J, Novak P, Gundlach H, et al. Next-Generation Sequencing Reveals the Impact of Repetitive DNA Across Phylogenetically Closely Related Genomes of *Orobanchaceae*. Mol Biol Evol. 2012; 29: 3601–3611.

50. Teruel M, Ruíz-Ruano FJ, Marchal JA, Sánchez A, Cabrero J, Camacho JP, et al. Disparate molecular evolution of two types of repetitive DNAs in the genome of the grasshopper *Eyprepocnemis plorans*. Heredity. 2014; 112: 531–542.

51. Ganley ARD, Kobayashi T. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. 2007; 17: 184–191.

52. Lindner, DL, Banik MT. Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. Mycologia. 2011; 103: 731–740.

53. Onyabe DY, Conn JE. Intragenomic heterogeneity of a ribosomal DNA spacer (ITS2) varies regionally in the neotropical malaria vector *Anopheles nuneztovari* (Diptera: Culicidae). Insect Mol Biol. 1999; 8: 435–442.

54. Bik HM, Fournier D, Sung W, Bergeron RD, Thomas WK. Intra-Genomic Variation in the Ribosomal Repeats of Nematodes. PLoS ONE. 2013; 8: e78230.

55. Schaap M, Lemmers RJ, Maassen R, van der Vliet PJ, Hoogerheide LF, van Dijk HK, et al. Genome-wide analysis of macrosatellite repeat copy number variation in worldwide populations: evidence for differences and commonalities in size distributions and size restrictions. BMC Genomics. 2013; 14: 143.

56. Zhao S, Gibbons JGA. Population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*. PLoS ONE. 2018; 13: e0201611.

57. Mascagni F, Vangelisti A, Giordani T, Cavallini A, Natali L. Specific LTR-Retrotransposons Show Copy Number Variations between Wild and Cultivated Sunflowers. Genes. 2018; 9.

58. Umemori J, Mori A, Ichiyanagi K, Uno T, Koide T. Identification of both copy number variation-type and constant-type core elements in a large segmental duplication region of the mouse genome. BMC Genomics. 2013; 14: 455.

59. McReynolds LA, DeSimone SM, Williams SA. Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi*. Proc Natl Acad Sci USA. 1986; 83: 797–801.

60. Meredith SE. Unnasch TR, Karam M, Piessens WF, Wirth DF. Cloning and characterization of an *Onchocerca volvulus* specific DNA sequence. Mol Biochem Parasitol. 1989; 36: 1–10.

61. Chanteau S, Luquiaud P, Failloux AB, Williams SA. Detection of *Wuchereria bancrofti* larvae in pools of mosquitoes by the polymerase chain reaction. Trans R Soc Trop Med Hyg. 1994; 88: 665–666.

62. Hamburger J, He-NA, Abbasi I, Ramzy RM, Jourdane J. Polymerase chain reaction assay based on a highly repeated sequence of *Schistosoma haematobium*: a potential tool for monitoring schistosome-infested water. Am J Trop Med Hyg. 2001; 65: 907–911.

63. Rao RU, Williams SA, Klei TR. Quantification of PCR amplification products of *Brugia* HhaI repeat DNA using a semiautomated Q-PCR system. Mol Cell Probes. 2002; 16: 13–23.

64. Nekrutenko A, Makova KD, Baker RJ. Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene. 2000; 249: 47–51.

65. Lodh N, Caro R, Sofer S, Scott A, Krolewiecki A, Shiff C. Diagnosis of *Strongyloides stercoralis* : Detection of parasite-derived DNA in urine. Acta Trop. 2016; 163: 9–13.

66. Demas A, Oberstaller J, DeBarry J, Lucchi NW, Srinivasamoorthy G, Sumari D, et al. Applied Genomics: Data Mining Reveals Species-Specific Malaria Diagnostic Targets More Sensitive than 18S rRNA. J Clin Microbiol. 2011; 4: 2411–2418.

67. Medley GF, Turner HC, Baggaley RF, Holland C, Hollingsworth TD. Chapter Six - The Role of More Sensitive Helminth Diagnostics in Mass Drug Administration Campaigns: Elimination and Health Impacts. In: Adv Parasitol. 2016; 94: 343-392.

68. Truscott JE, Werkman M, Wright JE, Farrell SH, Sarkar R, Ásbjörnsdóttir K, et al. Identifying optimal threshold statistics for elimination of hookworm using a stochastic simulation model. Parasit Vectors. 2017; 10: 321.

69. Ásbjörnsdóttir KH, Means AR, Werkman M, Walson JL. Prospects for elimination of soil-transmitted helminths. Curr Opin Infect Dis. 2017; 30: 482-488.

70. Lim MD, Brooker SJ, Belizario VY Jr, Gay-Andrieu F, Gilleard J, Levecke B, et al. Diagnostic tools for soil-transmitted helminths control and elimination programs: A pathway for diagnostic product development. PLoS Negl Trop Dis. 2018; 12: e0006213.

71. Kongs A, Marks G, Verlé P, der Stuyft PV. The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. Trop Med Int Health. 2001; 6: 163–169.

72. Clarke NE, Llewellyn S, Traub RJ, McCarthy J, Richardson A, Nery SV. Quantitative Polymerase Chain Reaction for Diagnosis of Soil-Transmitted Helminth Infections: A Comparison with a Flotation-Based Technique and an Investigation of Variability in DNA Detection. Am J Trop Med Hyg. 2018; 99: 1033–1040.

73. Farrell SH, Anderson RM. Helminth lifespan interacts with non-compliance in reducing the effectiveness of anthelmintic treatment. Parasit Vectors. 2018; 11: 66.

74. Zaky WI, Tomaino FR, Pilotte N, Laney SJ, Williams SA. Backpack PCR: A point-of-collection diagnostic platform for the rapid detection of *Brugia* parasites in mosquitoes. PLoS Negl Trop Dis. 2018; 12: e0006962.

75. Sanpool O, Intapan PM, Thanchomnang T, Sri-Aroon P, Lulitanond V, Sadaow L, et al. Development of a real-time PCR assay with fluorophore-labelled hybridization probes for detection of *Schistosoma mekongi* in infected snails and rat feces. Parasitology. 2012; 139: 1266–1272.

76. Pilotte N, Zaky WI, Abrams BP, Chadee DD, Williams SA. A Novel Xenomonitoring Technique Using Mosquito Excreta/Feces for the Detection of Filarial Parasites and Malaria. PLoS Negl Trop Dis. 2016; 10, e0004641.

77. Ramírez AL, Hall-Mendelin S, Doggett SL, Hewitson GR, McMahon JL, Ritchie SA. Mosquito excreta: A sample type with many potential applications for the investigation of Ross River virus and West Nile virus ecology. 2018; PLoS Negl Trop Dis. 12: e0006771.

78. Masny A, Sałamatin R, Rozej-Bielicka W, Golab E. Is molecular xenomonitoring of mosquitoes for *Dirofilaria repens* suitable for dirofilariosis surveillance in endemic regions? Parasitol Res. 2016; 115: 511–525.

79. Okorie PN, de Souza DK. Prospects, drawbacks and future needs of xenomonitoring for the endpoint evaluation of lymphatic filariasis elimination programs in Africa. Trans R Soc Trop Med Hyg. 2016; 110: 90–97.

80. Pilotte N, Unnasch TR, Williams SA. The Current Status of Molecular Xenomonitoring for Lymphatic Filariasis and Onchocerciasis. Trends Parasitol. 2017; 33: 788–798.

81. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016; 44: D774–D780.

82. Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010; 11: 378.

83. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792–793.

84. Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, et al. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 2017; 45: e111–e111.

85. Williams SA, Lizotte-Waniewski MR, Foster J, Guiliano D, Daub J, Scott AL, et al. The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. Int J Parasitol. 2000; 30: 411–419.

86. Rao RU, Weil GJ, Fischer K, Supali T, Fischer P. Detection of *Brugia* Parasite DNA in Human Blood by Real-Time PCR. J Clin Microbiol. 2006; 44: 3887–3893.

87. Subirana JA, Messeguer XA. Satellite explosion in the genome of holocentric nematodes. PloS One. 2013; 8: e62221.

88. Subirana JA, Albà MM, Messeguer X. High evolutionary turnover of satellite families in *Caenorhabditis*. BMC Evol Biol. 2015; 15: 218.

89. Pickering AJ, Njenga SM, Steinbaum L, Swarthout J, Lin A, Arnold BF, et al. Effects of single and integrated water, sanitation, handwashing, and nutrition interventions on child soil-transmitted helminth and Giardia infections: A cluster-randomized controlled trial in rural Kenya. PLOS Med. 2019; 16: e1002841.

90. Fischer K, Gankpala A, Gankpala L, Bolay FK, Curtis KC, Weil GJ, et al. *Capillaria* Ova and Diagnosis of *Trichuris trichiura* Infection in Humans by Kato-Katz Smear, Liberia. Emerg Infect Dis. 2018; 24: 1551–1554.

91. DeWorm3 | Natural History Museum. Available at: http://www.nhm.ac.uk/our-science/our-work/sustainability/deworm3.html. (Accessed: 23rd April 2019)

92. Ash LR, Orihel TC. 1992. Atlas of Human Parasitology. J Protozool. 1992; 39: 741–742.

93. Benjamin-Chung J, Pilotte N, Ercumen A, Grant JR, Maasch JRMA, Gonzalez AM, et al. Comparison of multi-parallel qPCR and Kato-Katz for detection of soil-transmitted helminth infection among children in rural Bangladesh. bioRxiv. 2019; 629501.

94. Ng-Nguyen D, Stevenson MA, Dorny P, Gabriël S, Vo TV, Nguyen VT, et al. Comparison of a new multiplex real-time PCR with the Kato Katz thick smear and copro-antigen ELISA for the detection and differentiation of *Taenia* spp. in human stools. PLoS Negl Trop Dis. 2017; 11: e0005743.

95. Stensvold CR, Saijuntha W, Sithithaworn P, Wongratanacheewin S, Strandgaard H, Ornbjerg N, et al. Evaluation of PCR based coprodiagnosis of human opisthorchiasis. Acta Trop. 2006; 97: 26–30.

96. Knopp S, Salim N, Schindler T, Voules DAK, Rothen J, Lweno O, et al. Diagnostic Accuracy of Kato–Katz, FLOTAC, Baermann, and PCR Methods for the Detection of Light-Intensity Hookworm and *Strongyloides stercoralis* Infections in Tanzania. Am J Trop Med Hyg. 2014; 90: 535–545.

97. Pontes LA, Oliveira MC, Katz N, Dias-Neto E, Rabello A. Comparison of a polymerase chain reaction and the Kato-Katz technique for diagnosing infection with *Schistosoma mansoni*. Am J Trop Med Hyg. 2003; 68: 652–656.

98. Irisarri-Gutiérrez MJ, Muñoz-Antolí C, Acosta L, Parker LA, Toledo R, Bornay-Llinares FJ, et al. Hookworm-like eggs in children's faecal samples from a rural area of Rwanda. Afr Health Sci. 2016; 16: 83–88.

99. Yong TS, Lee JH, Sim S, Lee J, Min DY, Chai JY, et al. Differential diagnosis of *Trichostrongylus* and hookworm eggs via PCR using ITS-1 sequence. Korean J Parasitol. 2007; 45: 69–74.

100. Ralph A, O'Sullivan MVN, Sangster NC, Walker JC. Abdominal pain and eosinophilia in suburban goat keepers — trichostrongylosis. Med J Aust. 2006; 184: 467–469.

101. Werneck JS, Carniato T, Gabriel A, Tufik S, Andrade SS. Mites in clinical stool specimens: potential misidentification as helminth eggs. Trans R Soc Trop Med Hyg. 2007; 101: 1154–1156.

102. Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a Multiplex Quantitative PCR to Assess Prevalence and Intensity Of Intestinal Parasite Infections in a Controlled Clinical Trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

103. George S, Kaliappan SP, Kattula D, Roy S, Geldhof P, Kang G, et al. Identification of *Ancylostoma ceylanicum* in children from a tribal community in Tamil Nadu, India using a semi-nested PCR-RFLP tool. Trans R Soc Trop Med Hyg. 2015; 109: 283–285.

104. Chidambaram M, Parija SC, Toi PC, Mandal J, Sankaramoorthy D, George S, et al. Evaluation of the utility of conventional polymerase chain reaction for detection and species differentiation in human hookworm infections. Trop Parasitol. 2017; 7: 111.

105. Areekul P, Putaporntip C, Pattanawong U, Sitthicharoenchai P, Jongwutiwes S. *Trichuris vulpis* and *T. trichiura* infections among schoolchildren of a rural community in northwestern Thailand: The possible role of dogs in disease transmission. Asian Biomed. 2010; 4: 49–60.

106. Mohd-Shaharuddin N, Lim YAL, Hassan N-A, Nathan S, Ngui R. Molecular characterization of *Trichuris* species isolated from humans, dogs and cats in a rural community in Peninsular Malaysia. Acta Trop. 2019; 190: 269–272.

# CHAPTER 7

# FUTURE DIRECTIONS AND OPPORTUNITIES

## 7.1 The Future of Molecular Diagnostics for Soil-Transmitted Helminths

Implementing methodical target selection strategies has long term advantages for the soil-transmitted helminth (STH) research and monitoring communities. While nucleic acid-based assay technologies continue to advance and evolve, parasite genomes evolve on an evolutionary scale. Therefore, even when targeting rapidly-diverging repetitive DNA regions, methodical selection strategies ensures that next-generation assays will have the ability to exploit optimal targets during assay design. In this manner, the identification of optimal targets constitutes an important and lasting research output with the capacity to significantly outlive the utility of the real-time PCR (qPCR) assays described in this dissertation.

Understanding that multi-parallel qPCR assays have limitations that present challenges for the STH research and monitoring communities, improving their utility is critical and attempts to build upon current assay technologies should be pursued. Foremost, opportunities to improve upon throughput, thereby reducing both resource/reagent costs and labor costs, should be prioritized as a short term goal. As such, with support from the Bill and Melinda Gates Foundation, we are currently collaborating with an industry partner (Quantigen Genomics, Fishers, IN) towards the adaptation of our assays into a multiplexed format, coupled with lyophilized, pre-plated reagents. A successful outcome will eliminate the need for cold chain, greatly facilitating assay exportability. While multi-parallel assays provide advantages over multiplexed assays in many settings, the increased throughput

provided by multiplexing can result in significant economic savings during large-scale studies. As such, through the multiplexing of targets, sample testing at scale becomes incresingly feasible. Accordingly, our assays are currently slated for use as the molecular discriminators of infection for the DeWorm3 Cluster Randomized Trials [1]. These studies, aiming to gauge the ability for chemotherapeutic intervention alone to allow for regional STH elimination will include the testing of approximately 240,000 individuals spanning three endemic countries (India, Malawi, and Benin). The trials constitute the largest STH operational research study ever performed, rendering throughput critical to feasibility and potential for success.

While lyophylization and multiplexing provide readily attainable benefits to research efforts in the short-term, future diagnostic platforms should aim to facilitate field-deployability, minimizing the need for laboratory infrastructure. Lateral flow-based detection systems provide an attractive option for next-generaion nucleic-acid based diganostics, as assay "read out" occurs rapidly, and in a presence/absence fashion much like an at-home pregnancy test. Recognizing the benefits of such a detection method, we have previoiusly adapted lateral flow technology to the detection of the filarial nematode parasite *Brugia malayi* [2]. However, this assay remains reliant upon the use of PCR, providing logistical challenges for programmatic implementation. Assays based upon technologies such as loop-mediated isothermal amplification (LAMP) [3], helicase-dependent amplification (HDA) [4], and recombinase polymerase amplification (RPA) [5] may help to overcome this reliance on PCR as all such amplification techniques are isothermal, requiring only a single temperature for product generation.

An alternative solution to the formidable challenges associated with a reliance on amplification may be to employ signal amplification-based strategies in place of target amplification-based approaches. Initially stemming from PCR-product-based contamination concerns, recent years have witnessed the growth of signal amplication strategies such as the use of branched DNA (bDNA) probes [6]. Fundamentally, these technologies remain reliant upon the complementary base pairing of a target sequence to a capture probe, typically bound to a solid-state surface. However, through the sequential addition of "amplifier" molecules that are capable of binding only in the presence of a captured DNA target, signal is increased, allowing for visual detection of product [7] (Figure 7.1). Due to their reliance on the flow of target molecules and assay components across a membrane-bound capture probe, bDNA-based assays are inherently adaptable to lateral flow designs, making the future exploitation of these technologies extremely attractive to the STH community, and to the neglected tropical disease (NTD) community as a whole.

### 7.2 Novel Approaches to Monitoring "Gut" Helminths

Historically, NTD monitoring and surveillance efforts have primarily gravitated towards antigen/antibody-based lateral flow assays for intervention monitoring and disease surveillance efforts. For the reasons discussed previously, the decision to rely on such immunoassays stems largely from their capacity for use as point-of-collection-based tests, requiring minimal infrastructure and producing a negative/positive readout in a matter of seconds or minutes. However, in recent years, the utility of these tests has been increasingly scrutinized, as a growing body of evidence is suggesting that many such

assays lack sensitivity and demonstrate considerable cross-reactivity leading, respectively, to both false negative and false positive outcomes [8, 9]. In response to such concerns, programmatic decision makers for diseases such as lymphatic filariasis (LF) are now re-examining the utility of alternative surveillance strategies such as molecular xenomonitoring (MX). MX, the testing of hematophagic insects for the presence of pathogen DNA [10] is attractive due to its non-invasive nature, eliminating the need for direct human sampling. As LF is a mosquito transmitted disease caused by infection with a lymph-dwelling nematode, vector insects can be monitored for the presence/absence of pathogen DNA, allowing for their use as a proxy measure of human infection [11]. However, while the potential utility of MX has long been recognized, throughput challenges, along with the associated costs and labor requirements have historically prevented the widespread programmatic use of this technique. Yet despite these limitations, the appeal of MX remains strong, and efforts to reimagine its use are increasing [12].

While taking a bloodmeal from an infected human host, a mosquito may ingest filarial nematodes that serve as the causative agents of LF. Following uptake, the filarial pathogen first travels to the mosquito's gut, where infection is established. From here, the mosquito's innate defenses attempt to eliminate the pathogen, and if successful, the pathogen is excreted within the feces produced by the mosquito. Reasoning that such a process is inherently similar to the processes underlying human infection with an intestinal helminth, we have developed a procedure facilitating the detection of pathogen signal from the excreta/feces (E/F) of a mosquito [13]. As the biological mass of mosquito E/F is significantly less than the corresponding mass of the E/F-producing mosquitoes, E/F

collected from many hundreds of mosquitoes can be tested as a single sample pool [13 – 15]. In contrast, testing of intact mosquitoes is limited to pool sizes of approximately 25 insects, as larger pools and the resulting dilution of DNA target may result in failed detection when the parasite burdens within the pools are low [16]. As such, E/F-based testing has the capacity to greatly improve the throughput of MX, overcoming one of the primary obstacles associated with traditional approaches. Coupling this technique with the detection of methodically selected repeat-based PCR targets and a novel methodology for the physical collecting of E/F [14] (Figure 7.2), we have developed a platform capable of serving as a high-throughput disease surveillance system. Currently, field-testing of this platform is underway, with studies in Ghana [17] and Cameroon having been completed. While further testing will be required, this platform has significant potential to augment traditional human and vector sampling strategies. As such, by re-imagining what it means to monitor for "gut"-infecting helminths, we have provided an important tool to the operational research, disease monitoring, and surveillance communities for vector borne diseases.

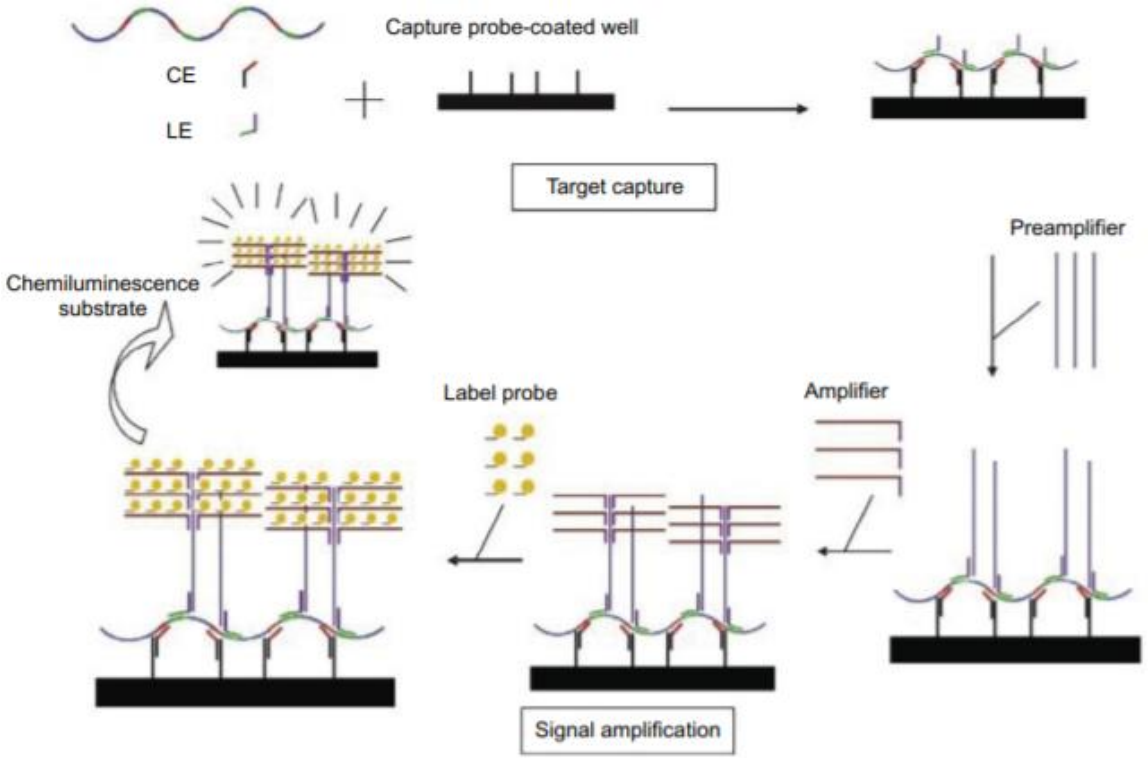


**Figure 7.1 Schematic representation of an example branched DNA assay.** Utilizing capture extenders (CE) designed to have complementarity to both the target DNA molecule and the universal capture probe, diagnostic targets are captured on a solid-state medium. Label extenders (LE) with complimentarity to the target are then bound to captured target molecules. Next, preamplifiers hybridize to complementary regions on each LE, and amplifiers anneal with preamplifiers. Label probe is then exposed to this assembled scaffold, allowing for either fluorescent or chemiluminescent signal amplification.

This image adapted from: [7]

Makowski GS. Advances in clinical chemistry. Volume 80. 2017. Available: http://www.sciencedirect.com/science/bookseries/00652423/80

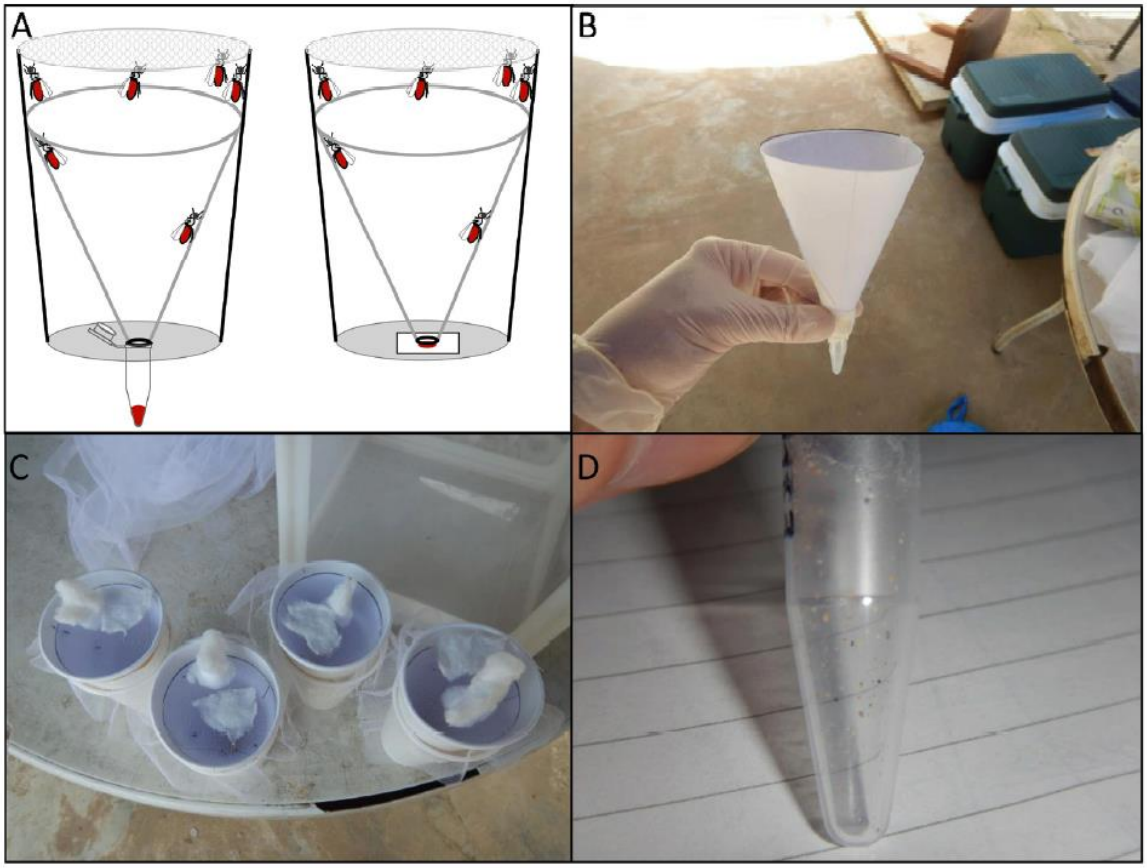


**Figure 7.2 Schematic of the superhydrophobic cone for E/F collection and images from field collections.** (A, B) By coating A4 printer paper with a superhydrophobic substance (NeverWet, Rust-Oleum, Durham, UK), cone-shaped collection funnels can be fabricated and placed above a collection vessel such as a microcentrifuge tube or a nucleic acid-preserving membrane. Funnels can then be fit into cups, and captured mosquitoes are transferred to the funnel. (C) Mosquitoes are housed above funnels for a period of days, allowing for the production of E/F which travels down the superhydrophobic surface and collects below. (D) Mosquito excreta feces following collection in a 1.7 mL microcentrifuge tube.

This image adapted from: [14]

Cook DAN, Pilotte N, Minetti C, Williams SA, Reimer LJ. A superhydrophobic cone to facilitate the xenomonitoring of filarial parasites, malaria, and trypanosomes using mosquito excreta/feces. Gates Open Res. 2018; 1: 7.

## 7.3 References


1. Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, Bailey R, Gardiner I, Halliday KE, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3 cluster randomized trial protocol. PLoS Negl Trop Dis. 2018; 12: e0006166.

2. Zaky WI, Tomaino FR, Pilotte N, Laney SJ, Williams SA. Backpack PCR: A point-of-collection diagnostic platform for the rapid detection of *Brugia* parasites in mosquitoes. PLoS Negl Trop Dis. 2018; 12: e0006962.

3. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. 2000; 28: E63.

4. Vincent M, Xu Y, Kong H. Helicase-dependent isothermal DNA amplification. EMBO Rep. 2004; 5: 795–800.

5. Piepenburg O, Williams CH, Stemple DL, Armes NA. DNA detection using recombination proteins. PLoS Biol. 2006; 4: e204.

6. Collins M. A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. Nucleic Acids Res. 1997; 25: 2979–2984.

7. Makowski GS. Advances in clinical chemistry. Volume 80. 2017. Available: http://www.sciencedirect.com/science/bookseries/00652423/80

8. Wanji S, Amvongo-Adjia N, Njouendou AJ, Kengne-Ouafo JA, Ndongmo WPC, Fombad FF, et al. Further evidence of the cross-reactivity of the Binax NOW® Filariasis ICT cards to non-*Wuchereria bancrofti* filariae: experimental studies with *Loa loa* and *Onchocerca ochengi*. Parasites Vectors. 2016; 9: 267.

9. Luz SLB, Crainey JL, Shelley AJ, Rubio M. Outstanding insecurities concerning the use of an Ov16-based ELISA in the Amazonia onchocerciasis focus. Mem Inst Oswaldo Cruz. 2014; 109: 506–508.

10. Opoku M, Minetti C, Kartey-Attipoe WD, Otoo S, Otchere J, Gomes B, et al. An assessment of mosquito collection techniques for xenomonitoring of anopheline-transmitted Lymphatic Filariasis in Ghana. Parasitology. 2018; 145: 1783–1791.

11. Ramesh A, Cameron M, Spence K, Hoek Spaans R, Melo-Santos MAV, Paiva MHS, et al. Development of an urban molecular xenomonitoring system for lymphatic filariasis in the Recife Metropolitan Region, Brazil. PLoS Negl Trop Dis. 2018; 12: e0006816.

12. Rao RU, Samarasekera SD, Nagodavithana KC, Punchihewa MW, Dassanayaka TDM, Gamini PDK, et al. Programmatic Use of Molecular Xenomonitoring at the Level of Evaluation Units to Assess Persistance of Lymphatic Filariasis in Sri Lanka. PLoS Negl Trop Dis. 2016; 10: e0004722.

13. Pilotte N, Zaky WI, Abrams BP, Chadee DD, Williams SA. A Novel Xenomonitoring Technique Using Mosquito Excreta/Feces for the Detection of Filarial Parasites and Malaria. PLoS Negl Trop Dis. 2016; 10: e0004641.

14. Cook DAN, Pilotte N, Minetti C, Williams SA, Reimer LJ. A superhydrophobic cone to facilitate the xenomonitoring of filarial parasites, malaria, and trypanosomes using mosquito excreta/feces. Gates Open Res. 2018; 1: 7.

15. Pilotte N, Cook DAN, Pryce J, Zulch MF, Minetti C, Reimer LJ, et al. Laboratory evaluation of molecular xenomonitoring using mosquito excreta/feces to amplify *Plasmodium*, *Brugia*, and *Trypanosoma* DNA. Gates Open Res. 2019; In press.

16. Rao RU, Atkinson LJ, Ramzy RMR, Helmy H, Farid HA, Bockarie MJ, et al. A Real-Time PCR Assay for Detection of *Wuchereria bancrofti* DNA in Blood and Mosquitoes. Am J Trop Med Hyg. 2006; 74: 826-832.

17. Minetti C, Pilotte N, Zulch MF, Canelas T, Tettevi EJ, Balunnaa F, et al. Amplification of human bloodborne eukaryotic pathogens using mosquito feces: A field evaluation of filariasis, malaria and mansonellosis. PLoS Negl Trop Dis. 2019; Under Review.

# BIBLIOGRAPHY


Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 5th ed. New York: Garland Science; 2008.

Albonico M, Montresor A, Crompton DWT, Savioli L. Intervention for the Control of Soil-Transmitted Helminthiasis in the Community. Advances in Parasitology. Elsevier; 2006. pp. 311–348.

Albonico M, Stoltzfus RJ, Savioli L, Tielsch JM, Chwaya HM. Epidemiological evidence for a different effect of hookworm species, *Ancylostoma duodenale* or *Necator americanus*, on iron status of children. Int J Epidemiol. 1998; 27: 530-537.

Alhassan A, Li Z, Poole CB, Carlow CK. Expanding the MDx toolbox for filarial diagnosis and surveillance. Trends Parasitol. 2015; 31: 391-400.

Allen AV, Ridley DS. Further observations on the formol-ether concentration technique for faecal parasites. J Clin Pathol. 1970; 23: 545–546.

Amaral LC, Robortella DR, Guimarães LFF, Limongi JE, Fontes CJF, Pereira DB, et al. Ribosomal and non-ribosomal PCR targets for the detection of low-density and mixed malaria infections. Malar J. 2019; 18: 154.

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLoS ONE. 2009; 4: e6372.

Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblížková A, et al. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. Ann Bot. 2011; 107: 255–268.

Amoah ID, Adegoke AA, Stenström TA. Soil-transmitted helminth infections associated with wastewater and sludge reuse: a review of current evidence. Trop Med Int Health. 2018; 23: 692–703.

Andersen LO, Röser D, Nejsum P, Nielsen HV, Stensvold CR. Is Supplementary Bead Beating for DNA Extraction from Nematode Eggs by Use of the NucliSENS easyMag Protocol Necessary? J Clin Microbiol. 2013; 51: 1345–1347.

Anderson R, Hollingsworth TD, Truscott J, Brooker S. Optimisation of mass chemotherapy to control soil-transmitted helminth infection. Lancet. 2012; 379: 289–290.

Anderson RM, May RM. Helminth Infections of Humans: Mathematical Models, Population Dynamics, and Control. Adv Parasitol. 1985; 24: 1-101.

Anderson RM, Schad GA. Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. Trans R Soc Trop Med Hyg. 1985; 79: 812–825.

Anderson R, Truscott J, Hollingsworth TD. The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths. Philos Trans R Soc Long B Biol Sci. 2014; 369: 20130435.

Anten JF, Zuidema PJ. Hookworm infection in Dutch servicemen returning from West New Guinea. Trop Geogr Med. 1964; 16: 216–224.

Areekul P, Putaporntip C, Pattanawong U, Sitthicharoenchai P, Jongwutiwes S. *Trichuris vulpis* and *T. trichiura* infections among schoolchildren of a rural community in northwestern Thailand: the possible role of dogs in disease transmission. Asian Biomed. 2010; 4: 49-60.

Arnold BF, Null C, Luby SP, Unicomb L, Stewart CP, Dewey KG, et al. Cluster-randomized controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. BMJ Open. 2013; 3: e003476.

Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, Bailey R, Gardiner I, Halliday KE, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3 cluster randomized trial protocol. PLoS Negl Trop Dis. 2018; 12: e0006166.

Ásbjörnsdóttir KH, Means AR, Werkman M, Walson JL. Prospects for elimination of soil-transmitted helminths. Curr Opin Infect Dis. 2017; 30: 482-488.

Ash LR, Orihel TC. 1992. Atlas of Human Parasitology. J Protozool. 1992; 39: 741–742.

Assefa LM, Crellen T, Kepha S, Kihara JH, Njenga SM, Pullan RL, et al. Diagnostic Accuracy and Cost-Effectiveness of Alternative Methods for Detection of Soil-Transmitted Helminths in a Post-Treatment Setting in Western Kenya. PLoS Negl Trop Dis. 2014; 8: e2843.

Assoum M, Ortu G, Basáñez M-G, Lau C, Clements ACA, Halton K, et al. Spatiotemporal distribution and population at risk of soil-transmitted helminth infections following an eight-year school-based deworming programme in Burundi, 2007-2014. Parasit Vectors. 2017; 10: 583.

Ayana M, Vlaminck J, Cools P, Ame S, Albonico M, Dana D, et al. Modification and optimization of the FECPAKG2 protocol for the detection and quantification of soil-transmitted helminth eggs in human stool. PLoS Negl Trop Dis. 2018; 12: e0006655.

Baker MK, Lange L, Verster A, van der Plaat S. A survey of helminths in domestic cats in the Pretoria area of Transvaal, Republic of South Africa. Part 1: The prevalence and comparison of burdens of helminths in adult and juvenile cats. J S Afr Vet Assoc. 1989; 60: 139–142.

Barda B, Albonico M, Ianniello D, Ame SM, Keiser J, Speich B, et al. How Long Can Stool Samples Be Fixed for an Accurate Diagnosis of Soil-Transmitted Helminth Infection Using Mini-FLOTAC? PLoS Negl Trop Dis. 2015; 9: e0003698.

Barda B, Cajal P, Villagran E, Cimino R, Juarez M, Krolewiecki A, et al. Mini-FLOTAC, Kato-Katz and McMaster: three methods, one goal; highlights from north Argentina. Parasit Vectors. 2014; 7: 271.

Bärenbold O, Raso G, Coulibaly JT, N'Goran EK, Utzinger J, Vounatsou P. Estimating sensitivity of the Kato-Katz technique for the diagnosis of *Schistosoma mansoni* and hookworm in relation to infection intensity. PLoS Negl Trop Dis. 2017; 11: e0005953.

Bartsch SM, Hotez PJ, Asti L, Zapf KM, Bottazzi ME, Diemert DJ, et al. The Global Economic and Health Burden of Human Hookworm Infection. PLoS Negl Trop Dis. 2016; 10, e0004922.

Basuni M, Muhi J, Othman N, Verweij JJ, Ahmad M, Miswan N, et al. A pentaplex real-time polymerase chain reaction assay for detection of four species of soil-transmitted helminths. Am J Trop Med Hyg. 2011; 84: 338–343.

Beaver PC, Martin LK. Evaluation of Kato Thick-Smear Technique for Quantitative Diagnosis of Helminth Infections. Am J Trop Med Hyg. 1968; 17: 382–391.

Becker SL, Liwanag HJ, Snyder JS, Akogun O, Belizario Jr V, Freeman MC, et al. Toward the 2020 goal of soil-transmitted helminthiasis control and elimination. PLoS Negl Trop Dis. 2018; 12: e0006606.

Bedell JA, Korf I, Gish W. MaskerAid: a performance enhancement to RepeatMasker. Bioinforma Oxf Engl. 2000; 16: 1040–1041.

Beknazarova M, Whiley H, Ross K. Strongyloidiasis: A Disease of Socioeconomic Disadvantage. Int J Environ Res Public Health. 2016; 13: 517.

Benjamin-Chung J, Nazneen A, Halder AK, Haque R, Siddique A, Uddin MS, et al. The Interaction of Deworming, Improved Sanitation, and Household Flooring with Soil-Transmitted Helminth Infection in Rural Bangladesh. PLoS Negl Trop Dis. 2015; 9: e0004256.

Benjamin-Chung J, Pilotte N, Ercumen A, Grant JR, Maasch JRMA, Gonzalez AM, et al. Comparison of multi-parallel qPCR and Kato-Katz for detection of soil-transmitted helminth infection among children in rural Bangladesh. bioRxiv. 2019; 629501.

Berger S. Infectious Diseases of East Timor. Los Angeles: GIDEON Informatics, Inc; 2015.

Bermingham ML, Handel IG, Glass EJ, Woolliams JA, Bronsvoort BM de C, McBride SH, et al. Hui and Walter's latent-class model extended to estimate diagnostic test properties from surveillance data: a latent model for latent data. Scientific Reports. 2015; 5: 11861.

Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet. 2006; 367: 1521–1532.

Bik HM, Fournier D, Sung W, Bergeron RD, Thomas WK. Intra-Genomic Variation in the Ribosomal Repeats of Nematodes. PLoS ONE. 2013; 8: e78230.

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M. Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. 2015; 23: 463–477.

Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. Chromosome Res. 2015; 23: 415–420.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. Reproducible, Interactive, Scalable, and Extensible Microbiome Data Science Using QIIME 2. Nat Biotechnol 2019; 37: 852-857.

Booth M, Vounatsou P, N'Goran EK, Tanner M, Utzinger J. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. Parasitology. 2003; 127: 525–531.

Bosco A, Rinaldi L, Maurelli M, Musella V, Coles G, Cringoli G. The comparison of FLOTAC, FECPAK and McMaster techniques for nematode egg counts in cattle. Acta Parasitol. 2014; 59.

Bradbury RS, Hii SF, Harrington H, Speare R, Traub R. *Ancylostoma ceylanicum* hookworm in the Solomon Islands. Emerg Infect Dis. 2017; 23: 252-257.

Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev Vet Med. 2005; 68: 145–163.

Brindley PJ, Laha T, McManus DP, Loukas A. Mobile genetic elements colonizing the genomes of metazoan parasites. Trends Parasitol. 2003; 19: 79–87.

Brooker S, Hotez PJ, Bundy DAP. Hookworm-Related Anaemia among Pregnant Women: A Systematic Review. PLoS Negl Trop Dis. 2008; 2: e291.

Brooker SJ, Pullan RL. *Ascaris lumbricoides* and Ascariasis: Estimating Numbers Infected and Burden of Disease. In: *Ascaris*: The Neglected Parasite. London: Academic Press; 2013. pp. 343-360.

Brown HW. Further studies on the longevity of the eggs of *Ascaris lumbricoides* and A. *suum*. J Parasitol. 1928; 15: 14–22.

Bundy DAP, Silva N de, Horton S, Jamison DT. Disease Control Priorities, Third Edition (Volume 8): Child and Adolescent Health and Development. World Bank Publications; 2017.

Buonfrate D, Formenti F, Perandin F, Bisoffi Z. Novel approaches to the diagnosis of *Strongyloides stercoralis* infection. Clin Microbiol Infect. 2015; 21: 543–552.

Caraguel CGB, Stryhn H, Gagné N, Dohoo IR, Hammell KL. Selection of a Cutoff Value for Real-Time Polymerase Chain Reaction Results to Fit a Diagnostic Purpose: Analytical and Epidemiologic Approaches. J Vet Diag Invest. 2011; 23: 2-15.

Caruccio N. Preparation of next-generation sequencing libraries using Nextera technology: simultaneous DNA fragmentation and adapter tagging by in vitro transposition. Methods Mol Biol. 2011; 733: 241-55.

Chammartin F, Scholte RG, Malone JB, Bavia ME, Nieto P, Utzinger J, et al. Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. Parasit Vectors. 2013; 6: 152.

Chanteau S, Luquiaud P, Failloux AB, Williams SA. Detection of *Wuchereria bancrofti* larvae in pools of mosquitoes by the polymerase chain reaction. Trans R Soc Trop Med Hyg. 1994; 88: 665–666.

Chard AN, Garn JV, Chang HH, Clasen T, Freeman MC. Impact of a school-based water, sanitation, and hygiene intervention on school absence, diarrhea, respiratory infection, and soil-transmitted helminths: results from the WASH HELPS cluster-randomized trial. J Glob Health. 2019; 9: 020402.

Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 1994; 371: 215.

Chidambaram M, Parija SC, Toi PC, Mandal J, Sankaramoorthy D, George S, et al. Evaluation of the utility of conventional polymerase chain reaction for detection and species differentiation in human hookworm infections. Trop Parasitol. 2017; 7: 111.

Children Without Worms. (2018). STH Coalition Action Group Meeting 2018 Report. Retrieved April 2019 from http://www.childrenwithoutworms.org/sites/ default/files/ 2018%20Action%20Meeting%20Report_FINAL%2007.16.18.pdf.

Christensen G, Dentz HN, Pickering AJ, Bourdier T, Arnold BF, Colford Jr JM, et al. Pilot Cluster Randomized Controlled Trials to Evaluate Adoption of Water, Sanitation, and Hygiene Interventions and Their Combination in Rural Western Kenya. Am J Trop Med Hyg. 2015; 92: 437-447.

Cimino RO, Jeun R, Juarez M, Cajal PS, Vargas P, Echazú A, et al. Identification of human intestinal parasites affecting an asymptomatic peri-urban Argentinian population using multi-parallel quantitative real-time polymerase chain reaction. Parasit Vectors. 2015; 8: 380.

Clarke NE, Llewellyn S, Traub RJ, McCarthy J, Richardson A, Nery SV. Quantitative Polymerase Chain Reaction for Diagnosis of Soil-Transmitted Helminth Infections: A Comparison with a Flotation-Based Technique and an Investigation of Variability in DNA Detection. Am J Trop Med Hyg. 2018; 99: 1033–1040.

Coffeng LE, Truscott JE, Farrell SH, Turner HC, Sarkar R, Kang G, et al. Comparison and validation of two mathematical models for the impact of mass drug administration on *Ascaris lumbricoides* and hookworm infection. Epidemics. 2017; 18: 38-47.

Coffeng LE, Vaz Nery S, Gray DJ, Bakker R, de Vlas SJ, Clements ACA. Predicted short and long-term impact of deworming and water, hygiene, and sanitation on transmission of soil-transmitted helminths. PLoS Negl Trop Dis. 2018; 12: e0006758.

Collins M. A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. Nucleic Acids Res. 1997; 25: 2979–2984.

Colmer-Hamood JA. Fecal Microscopy Artifacts Mimicking Ova and Parasites. Lab Med. 2001; 32: 80–84.

Conlan JV, Khamlome B, Vongxay K, Elliot A, Pallant L, Sripa B, et al. Soil-transmitted helminthiasis in Laos: a community-wide cross-sectional study of humans and dogs in a mass drug administration environment. Am J Trop Med Hyg. 2012; 86: 624–634.

Cook DAN, Pilotte N, Minetti C, Williams SA, Reimer LJ. A superhydrophobic cone to facilitate the xenomonitoring of filarial parasites, malaria, and trypanosomes using mosquito excreta/feces. Gates Open Res. 2018; 1: 7.

Cools P, Vlaminck J, Albonico M, Ame S, Ayana M, Antonio BPJ, et al. Diagnostic performance of a single and duplicate Kato-Katz, Mini-FLOTAC, FECPAKG2 and qPCR for the detection and quantification of soil-transmitted helminths in three endemic countries. PLoS Negl Trop Dis. 2019; 13: e0007446.

Cooper GM. The Cell: A Molecular Approach. 2nd Edition. Sunderland (MA): Sinauer Associates; 2000.

Craig JM, Scott AL. Antecedent *Nippostrongylus* infection alters the lung immune response to *Plasmodium berghei*. Parasite Immunol. 2017; 39: e12441.

Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruysse J, Utzinger J, et al. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. Nat Protoc. 2017; 12: 1723–1732.

Cringoli G, Rinaldi L, Maurelli MP, Utzinger J. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat Protoc. 2010; 5: 503–515.

Crompton DW. The public health importance of hookworm disease. Parasitology. 2000; 121 Suppl: S39-S50.

Cuomo MJ, Noel LB, White DB. Diagnosing Medical Parasites: A Public Health Officers Guide to Assisting Laboratory and Medical Officers. USAF Air Education and Training Command, Randolph, TX: http://www.phsource.us /PH/PARA/; 2012.

da Cruz CJ. Livestock development in East Timor. In: Agriculture: New Directions for a New Nation – East Timor (Timor Leste). ACIAR Proceedings No. 113; 2003. pp. 11-16.

Dąbrowska J, Zdybel J, Karamon J, Kochanowski M, Stojecki K, Cencek T, et al. Assessment of viability of the nematode eggs (*Ascaris*, *Toxocara*, *Trichuris*) in sewage sludge with the use of LIVE/DEAD Bacterial Viability Kit. Ann Agric Environ Med AAEM. 2014; 21: 35–41.

Dacombe RJ, Crampin AC, Floyd S, Randall A, Ndhlovu R, Bickle Q, et al. Time delays between patient and laboratory selectively affect accuracy of helminth diagnosis. Trans R Soc Trop Med Hyg. 2007; 101: 140–145.

Das CJ, Kumar J, Debnath J, Chaudhry A. Imaging of ascariasis. Australas Radiol. 2007; 51: 500-506.

de Gruijter JM, van Lieshout L, Gasser RB, Verweij JJ, Brienen EAT, Ziem JB, et al. Polymerase chain reaction-based differential diagnosis of *Ancylostoma duodenale* and *Necator americanus* infections in humans in northern Ghana. Trop Med Int Health. 2005;10: 574–580.

Deer DM, Lampel KA, González-Escalona N. A versatile internal control for use as DNA in real-time PCR and as RNA in real-time reverse transcription PCR assays. Lett Appl Microbiol. 2010; 50: 366–372.

Demas A, Oberstaller J, DeBarry J, Lucchi NW, Srinivasamoorthy G, Sumari D, et al. Applied Genomics: Data Mining Reveals Species-Specific Malaria Diagnostic Targets More Sensitive than 18S rRNA. J Clin Microbiol. 2011; 4: 2411–2418.

Dendukuri N, Joseph L. Bayesian Approaches to Modeling the Conditional Dependence Between Multiple Diagnostic Tests. Biometrics. 2001; 57: 158–167.

DeWorm3 | Natural History Museum. Available at: http://www.nhm.ac.uk/our-science/our-work/sustainability/deworm3.html. (Accessed: 23rd April 2019)

Dunn A, Keymer A. Factors affecting the reliability of the McMaster technique. J Helminthol. 1986;60: 260–262.

Dunn JC, Bettis AA, Wyine NY, Lwin AMM, Tun A, Maung NS, et al. Soil-transmitted helminth reinfection four and six months after mass drug administration: results from the delta region of Myanmar. PLoS Negl Trop Dis. 2019; 13: e0006591.

Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, et al. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. Parasit Vectors. 2016; 9: 38.

Easton AV, Oliveira RG, Walker M, O'Connell EM, Njenga SM, Mwandawiro CS, et al. Sources of variability in the measurement of *Ascaris lumbricoides* infection intensity by Kato-Katz and qPCR. Parasit Vectors. 2017; 10: 256.

Egwang TG, Slocombe JO. Efficiency and sensitivity of techniques for recovering nematode eggs from bovine feces. Can J Comp Med. 1981; 45: 243–248.

Ercumen A, Benjamin-Chung J, Arnold BF, Lin A, Hubbard AE, Stewart C, et al. Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: A cluster-randomized controlled trial in rural Bangladesh. PLoS Negl Trop Dis. 2019; 13: e0007323.

Etter A, Bernard V, Kenzelmann M, Tobler H, Müller F. Ribosomal heterogeneity from chromatin diminution in *Ascaris lumbricoides*. Science. 1994; 265: 954-956.

Fabian S, Odermatt P, Khieu V, Panning M, Duong S, Muth S, et al. Evaluation of real-time PCR for *Strongyloides stercoralis* and hookworm as diagnostic tool in asymptomatic schoolchildren in Cambodia. Acta Trop. 2013; 126: 89-92.

Farrell SH, Anderson RM. Helminth lifespan interacts with non-compliance in reducing the effectiveness of anthelmintic treatment. Parasit Vectors. 2018; 11: 66.

Farrell SH, Truscott JE, Anderson RM. The importance of patient compliance in repeated rounds of mass drug administration (MDA) for the elimination of intestinal helminth transmission. Parasit Vectors. 2017; 10: 291.

Fischer K, Gankpala A, Gankpala L, Bolay FK, Curtis KC, Weil GJ, et al. *Capillaria* Ova and Diagnosis of *Trichuris trichiura* Infection in Humans by Kato-Katz Smear, Liberia. Emerg Infect Dis. 2018; 24: 1551–1554.

Flavell RB, Bennett MD, Smith JB, Smith DB. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet. 1974; 12: 257–269.

Freedman DA. On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors." The American Statistician. 2006; 60: 299–302.

Freeman MC, Akogun O, Belizario V, Brooker SJ, Gyorkos TW, Imtiaz R, et al. Challenges and opportunities for control and elimination of soil-transmitted helminth infection beyond 2020. PLoS Negl Trop Dis. 2019; 13: e0007201.

Fritz GN, Conn J, Cockburn A, Seawright J. Sequence Analysis of the Ribosomal DNA Internal Transcribed Spacer 2 from Populations of *Anopheles nuneztovari* (Diptera: Culicidae). Mol Biol Evol. 1994; 11: 406-416.

Funakoshi K, Bagheri M, Zhou M, Suzuki R, Abe H, Akashi H. Highly sensitive and specific Alu-based quantification of human cells among rodent cells. Sci Rep. 2017; 7: 13202.

Fung MS, Xiao N, Wang S, Carlton EJ. Field evaluation of a PCR test for *Schistosoma japonicum* egg detection in low-prevalence regions of China. Am J Trop Med Hyg. 2012; 87: 1053-1058.

Gall S, Müller I, Walter C, Seelig H, Steenkamp L, Pühse U, et al. Associations between selective attention and soil-transmitted helminth infections, socioeconomic status, and physical fitness in disadvantaged children in Port Elizabeth, South Africa: An observational study. PLoS Negl Trop Dis. 2017; 11: e0005573.

Gamer M, Lemon J, Fellows I, Singh P. Package 'irr': Various Coefficients of Interrater Reliability and Agreement. 2012. Available: https://cran.r-project.org/web/packages/irr/irr.pdf.

Ganley ARD, Kobayashi T. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. 2007; 17: 184–191.

Garcia LS. Diagnostic Medical Parasitology. American Society for Microbiology Press; 2006.

Gatehouse HS, Malone LA. The Ribosomal RNA Gene Region of *Nosema apis* (Microspora): DNA Sequence for Small and Large Subunit rRNA Genes and Evidence of a Large Tandem Repeat Unit Size. J Invertebr Pathol. 1998; 71: 97-105.

GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017; 390: 1260–1344.

GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392: 1859-1922.

George S, Kaliappan SP, Kattula D, Roy S, Geldhof P, Kang G, et al. Identification of *Ancylostoma ceylanicum* in children from a tribal community in Tamil Nadu, India using a semi-nested PCR-RFLP tool. Trans R Soc Trop Med Hyg. 2015; 109: 283–285.

Gerbi SA. Evolution of ribosomal DNA. In: Molecular Evolutionary Genetics. New York/London: Plenum Press; 1985. pp. 419-517.

Goday C, Pimpinelli S. The occurrence, role and evolution of chromatin diminution in nematodes. Parasitol Today. 1993; 9: 319-322.

Godber OF, Phythian CJ, Bosco A, Ianniello D, Coles G, Rinaldi L, et al. A comparison of the FECPAK and Mini-FLOTAC faecal egg counting techniques. Vet Parasitol. 2015; 207: 342–345.

Goldstein P. Chromatin diminution in early embryogenesis of *Ascaris lumbricoides* L. var. *suum*. J Morphol. 1977; 152: 141-151.

Gordon CA, McManus DP, Acosta LP, Olveda RM, Williams GM, Ross AG, et al. Multiplex real-time PCR monitoring of intestinal helminthes in humans reveals widespread polyparasitism in Northern Samar, the Philippines. Int J Parasitol. 2015; 45: 477-483.

Grenfell AW, Strzelecka M, Heald R. Transcription brings the complex(ity) to the centromere. Cell Cycle. 2017; 16: 235–236.

Gyorkos TW, Gilbert NL, Larocque R, Casapía M. *Trichuris* and hookworm infections associated with anaemia during pregnancy: *Trichuris*, hookworm and anaemia in pregnancy. Trop Med Int Health. 2011; 16: 531–537.

Gyorkos TW, Maheu-Giroux M, Blouin B, Casapia M. Impact of Health Education on Soil-Transmitted Helminth Infections in Schoolchildren of the Peruvian Amazon: A Cluster-Randomized Controlled Trial. PLoS Negl Trop Dis. 2013; 7: e2397.

Habtamu K, Degarege A, Ye-Ebiyo Y, Erko B. Comparison of the Kato-Katz and FLOTAC techniques for the diagnosis of soil-transmitted helminth infections. Parasitol Int. 2011; 60: 398–402.

Halliday KE, Oswald WE, Mcharo C, Beaumont E, Gichuki PM, Kepha S, et al. Community-level epidemiology of soil-transmitted helminths in the context of school-based deworming: Baseline results of a cluster randomised trial on the coast of Kenya. PLoS Negl Trop Dis. 2019; 13: e0007427.

Halpenny CM, Paller C, Koski KG, Valdés VE, Scott ME. Regional, Household and Individual Factors that Influence Soil Transmitted Helminth Reinfection Dynamics in Preschool Children from Rural Indigenous Panamá. PLoS Negl Trop Dis. 2013; 7: e2070.

Hamburger J, He-NA, Abbasi I, Ramzy RM, Jourdane J. Polymerase chain reaction assay based on a highly repeated sequence of *Schistosoma haematobium*: a potential tool for monitoring schistosome-infested water. Am J Trop Med Hyg. 2001; 65: 907–911.

Haque R, Mondal D, Karim A, Molla IH, Rahim A, Faruque ASG, et al. Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. Clin Infect Dis. 2009; 48: 1191-1197.

Harmon AF, Zarlenga DS, Hildreth MB. Improved methods for isolating DNA from *Ostertagia ostertagi* eggs in cattle feces. Vet Parasitol. 2006; 135: 297–302.

Hendrix CM, Robinson E. Diagnostic Parasitology for Veterinary Technicians. 4th ed. Philadelphia: Elsevier Health Sciences; 2014.

Hicks JH, Kremer M, Miguel E. The Case for Mass Treatment of Intestinal Helminths in Endemic Areas. PLoS Negl Trop Dis. 2015; 9: e0004214.

Hidalgo A, Melo A, Romero F, Hidalgo V, Villanueva J, Fonseca-Salamanca F. DNA extraction in *Echinococcus granulosus* and *Taenia* spp. eggs in dogs stool samples applying thermal shock. Exp Parasitol. 2018; 186: 10–16.

Hii SF, Senevirathna D, Llewellyn S, Inpankaew T, Odermatt P, Khieu V, et al. Development and Evaluation of a Multiplex Quantitative Real-Time Polymerase Chain Reaction for Hookworm Species in Human Stool. Am J Trop Med Hyg. 2018; 99: 1186–1193.

Hino A, Maruyama H, Kikuchi T. A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. Parasitol Int. 2016; 65: 572–575.

Hlinka O, Murrell A, Barker SC. Evolution of the secondary structure of the rRNA internal transcriber spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). Heredity. 2002; 88: 275-279.

Holt DC, Shield J, Harris TM, Mounsey KE, Aland K, McCarthy JS, et al. Soil-Transmitted Helminths in Children in a Remote Aboriginal Community in the Northern Territory: Hookworm is Rare but *Strongyloides stercoralis* and *Trichuris trichiura* Persist. Trop Med Infect Dis. 2017; 2: 51.

Hotez PJ, Bundy DAP, Beegle K, Brooker S, Drake L, de Silva N, et al. Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis. In: Disease Control Priorities in Developing Countries. Washington: World Bank; 2006. pp. 467-482.

Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016; 44: D774–D780.

Hu W, Wu S, Yu X, Abullahi AY, Song M, Tan L, et al. A multiplex PCR for simultaneous detection of three zoonotic parasites *Ancylostoma ceylanicum*, *Ancylostoma* caninum, and *Giardia lamblia* assemblage A. BioMed Res Int. 2015; 2015: 406168.

Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. Nat Genet. 2016; 48: 299-307.

Imbens GW, Rubin DB. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. New York, NY, USA: Cambridge University Press; 2015.

Inpankaew T, Schär F, Dalsgaard A, Khieu V, Chimnoi W, Chhoun C, et al. High prevalence of *Ancylostoma ceylanicum* hookworm infections in humans, Cambodia, 2012. Emerg Infect Dis. 2014; 20: 976–982.

International Helminth Genomes Consortium, Coghlan A, Mitreva M, Berriman M. Comparative genomics of the major parasitic worms. Nat Genet. 2019; 51: 163-174.

Iqbal A, Mahdy MAK, Dixon BR, Surin J, Lim YAL. Cryptosporidisis in HIV-Infected Individuals: A Global Perspective. J AIDS Clin Res. 2012; 01.

Irisarri-Gutiérrez MJ, Muñoz-Antolí C, Acosta L, Parker LA, Toledo R, Bornay-Llinares FJ, et al. Hookworm-like eggs in children's faecal samples from a rural area of Rwanda. Afr Health Sci. 2016; 16: 83–88.

Janwan P, Intapan PM, Thanchomnang T, Lulitanond V, Anamnart W, Maleewong W. Rapid detection of *Opisthorchis viverrini* and *Strongyloides stercoralis* in human fecal samples using a duplex real-time PCR and melting curve analysis. Parasitol Res. 2011; 109: 1593-1601.

Jia T-W, Melville S, Utzinger J, King CH, Zhou X-N. Soil-Transmitted Helminth Reinfection after Drug Treatment: A Systematic Review and Meta-Analysis. PLoS Negl Trop Dis. 2012; 6: e1621.

Jiraanankul V, Aphijirawat W, Mungthin M, Khositnithikul R, Rangsin R, Traub RJ, et al. Incidence and risk factors of hookworm infection in a rural community of central Thailand. Am J Trop Med Hyg. 2011; 84: 594–598.

Jonker FAM, Calis JCJ, Phiri K, Brienen EAT, Khoffi H, Brabin BJ, et al. Real-time PCR demonstrates *Ancylostoma duodenale* is a key factor in the etiology of severe anemia and iron deficiency in Malawian pre-school children. PLoS Negl Trop Dis. 2012; 6: e1555.

Joseph L, Gyorkos TW, Coupal L. Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard. Am J Epidemiol. 1995; 141: 263–272.

Jourdan P, Lamberton P, Fenwick A, Addiss D. Soil-transmitted helminth infections. Lancet. 2018; 391: 2322–2323.

Kaisar MMM, Brienen EAT, Djuardi Y, Sartono E, Yazdanbakhsh M, Verweij JJ, et al. Improved diagnosis of *Trichuris trichiura* by using a bead-beating procedure on ethanol preserved stool samples prior to DNA isolation and the performance of multiplex real-time PCR for intestinal parasites. Parasitology. 2017; 144: 965–974.

Kato K, Miura M. Comparative examinations. Jpn J Parasitol. 1954; 3: 35.

Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in *Schistosomia mansoni*. Rev Inst Med Trop Sao Paulo. 1972; 14: 397–400.

Kim M-K, Pyo K-H, Hwang Y-S, Park KH, Hwang IG, Chai J-Y, et al. Effect of Temperature on Embryonation of *Ascaris suum* Eggs in an Environmental Chamber. Korean J Parasitol. 2012; 50: 239–242.

King CH, Bertino A-M. Asymmetries of Poverty: Why Global Burden of Disease Valuations Underestimate the Burden of Neglected Tropical Diseases. PLoS Negl Trop Dis. 2008; 2: e209.

Knopp S, Mgeni AF, Khamis IS, Steinmann P, Stothard JR, Rollinson D, et al. Diagnosis of Soil-Transmitted Helminths in the Era of Preventive Chemotherapy: Effect of Multiple Stool Sampling and Use of Different Diagnostic Techniques. PLoS Negl Trop Dis. 2008; 2: e331.

Knopp S, Rinaldi L, Khamis IS, Stothard JR, Rollinson D, Maurelli MP, et al. A single FLOTAC is more sensitive than triplicate Kato–Katz for the diagnosis of low-intensity soil-transmitted helminth infections. Trans R Soc Trop Med Hyg. 2009; 103: 347–354.

Knopp S, Salim N, Schindler T, Voules DAK, Rothen J, Lweno O, et al. Diagnostic Accuracy of Kato–Katz, FLOTAC, Baermann, and PCR Methods for the Detection of Light-Intensity Hookworm and *Strongyloides stercoralis* Infections in Tanzania. Am J Trop Med Hyg. 2014; 90: 535–545.

Knopp S, Speich B, Hattendorf J, Rinaldi L, Mohammed KA, Khamis IS, et al. Diagnostic Accuracy of Kato-Katz and FLOTAC for Assessing Anthelmintic Drug Efficacy. PLoS Negl Trop Dis. 2011; 5: e1036.

Kochanowski M, Karamon J, Dąbrowska J, Cencek T. Experimental Estimation of the Efficacy of the Flotac Basic Technique. J Parasitol. 2014; 100: 633–639.

Kong J, Huh S, Won J-I, Yoon J, Kim B, Kim K. GAAP: A Genome Assembly + Annotation Pipeline. BioMed Res Int. 2019; 2019: 1–12.

Kongs A, Marks G, Verlé P, der Stuyft PV. The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. Trop Med Int Health. 2001; 6: 163–169.

Krauth SJ, Coulibaly JT, Knopp S, Traoré M, N'Goran EK, Utzinger J. An In-Depth Analysis of a Piece of Shit: Distribution of *Schistosoma mansoni* and Hookworm Eggs in Human Stool. PLoS Negl Trop Dis. 2012; 6: e1969.

Kure A, Mekonnen Z, Dana D, Bajiro M, Ayana M, Vercruysse J, et al. Comparison of individual and pooled stool samples for the assessment of intensity of *Schistosoma mansoni* and soil-transmitted helminth infections using the Kato-Katz technique. Parasit Vectors. 2015; 8: 489.

Kyu, HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392; 1859–1922.

Lafontaine DLJ, Tollervey D. The function and synthesis of ribosomes. Nat Rev Mol Cell Biol. 2001; 2: 514.

Lai Y-S, Zhou X-N, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. Paraist Vectors. 2013; 6: 359.

Lamberton PH, Cheke RA, Winskill P, Tirados I, Walker M, Osei-Atweneboana MY, et al. Onchocerciasis transmission in Ghana: persistence under different control strategies and the roll of the simulidd vectors. PLoS Negl Trop Dis. 2015; 9: e0003688.

Landmann JK, Prociv P. Experimental human infection with the dog hookworm, *Ancylostoma caninum*. Med J Aust. 2003; 178: 69–71.

Levecke B, Behnke JM, Ajjampur SSR, Albonico M, Ame SM, Charlier J, et al. A Comparison of the Sensitivity and Fecal Egg Counts of the McMaster Egg Counting and Kato-Katz Thick Smear Methods for Soil-Transmitted Helminths. PLoS Negl Trop Dis. 2011; 5: e1201.

Lim MD, Brooker SJ, Belizario VY Jr, Gay-Andrieu F, Gilleard J, Levecke B, et al. Diagnostic tools for soil-transmitted helminths control and elimination programs: A pathway for diagnostic product development. PLoS Negl Trop Dis. 2018; 12: e0006213.

Lin D-D, Liu J-X, Liu Y-M, Hu F, Zhang Y-Y, Xu J-M, et al. Routine Kato–Katz technique underestimates the prevalence of *Schistosoma japonicum*: A case study in an endemic area of the People's Republic of China. Parasitol Int. 2008; 57: 281–286.

Lindner, DL, Banik MT. Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. Mycologia. 2011; 103: 731–740.

Liu J, Gratz J, Amour C, Kibiki G, Becker S, Janaki L, et al. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. J Clin Microbiol. 2013; 51: 472-480.

Llewellyn S, Inpankaew T, Nery SV, Gray DJ, Verweij JJ, Clements ACA, et al. Application of a multiplex quantitative PCR to assess prevalence and intensity of intestinal parasite infections in a controlled clinical trial. PLoS Negl Trop Dis. 2016; 10: e0004380.

Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016; 8.

Lo NC, Addiss DG, Hotez PJ, King CH, Stothard JR, Evans DS, et al. A call to strengthen the global strategy against schistosomiasis and soil-transmitted helminthiasis: the time is now. Lancet Infect Dis. 2017; 17: e64–e69.

Lo NC, Heft-Neal S, Coulibaly JT, Leonard L, Bendavid E, Addiss DG. State of deworming coverage and equity in low-income and middle-income countries using household health surveys: a spatiotemporal cross-sectional study. Lancet Glob Health. 2019; 7: e1511–e1520.

Lodh N, Caro R, Sofer S, Scott A, Krolewiecki A, Shiff C. Diagnosis of *Strongyloides stercoralis* : Detection of parasite-derived DNA in urine. Acta Trop. 2016; 163: 9–13.

London Declaration on Neglected Tropical Diseases. Uniting to Combat NTDs. Available at: /london-declaration-neglected-tropical-diseases/. (Accessed: 9th April 2019)

López-Flores I, Garrido-Ramos MA. The Repetitive DNA Content of Eukaryotic Genomes. In: Genome Dynamics: Karger; 2012.

Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, et al. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. Lancet Glob Health. 2018; 6: e302-e315.

Lucio-Forster A, Liotta JL, Yaros JP, Briggs KR, Mohammed HO, Bowman DD. Morphological differentiation of eggs of *Ancylostoma caninum, Ancylostoma tubaeforme*, and *Ancylostoma braziliense* from dogs and cats in the United States. J Parasitol. 2012; 98: 1041–1044.

Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10: 325–337.

Lustigman S, Prichard RK, Gazzinelli A, Grant WN, Boatin BA, McCarthy JS, et al. A Research Agenda for Helminth Diseases of Humans: The Problem of Helminthiases. PLoS Negl Trop Dis. 2012; 6: e1582.

Luz SLB, Crainey JL, Shelley AJ, Rubio M. Outstanding insecurities concerning the use of an Ov16-based ELISA in the Amazonia onchocerciasis focus. Mem Inst Oswaldo Cruz. 2014; 109: 506–508.

Mabbott NA. The Influence of Parasite Infections on Host Immunity to Co-infection With Other Pathogens. Front Immunol. 2018; 9: 2579.

MacGregor RR, Dreyer K, Herman S, Hocknell PK, Nghiem L, Tevere VJ, et al. Use of PCR in Detection of *Mycobacterium avium* Complex (MAC) Bacteremia: Sensitivity of the Assay and Effect of Treatment for MAC Infection on Concentrations of Human Immunodeficiency Virus in Plasma. J Clin Microbiol. 1999; 37: 90-94.

Machida U, Kami M, Fukui T, Kazuyama Y, Kinoshita M, Tanaka Y, et al. Real-Time Automated PCR for Early Diagnosis and Monitoring of Cytomegalovirus Infection after Bone Marrow Transplantation. J Clin Microbiol. 2000; 38: 2536-2542.

Mackay IM. Real-time PCR in the microbiology laboratory. Clin Microbiol Infect. 2004; 10: 190-212.

Makowski GS. Advances in clinical chemistry. Volume 80. 2017. Available: http://www.sciencedirect.com/science/bookseries/00652423/80.

Markoulatos P, Siafakas N, Moncany M. Multiplex Polymerase Chain Reaction: A Practical Approach. J Clin Lab Anal. 2002; 16: 47-51.

Martinez-Bakker M, King AA, Rohani P. Unraveling the Transmission Ecology of Polio. PLoS Biol. 2015; 13: e1002172.

Mascagni F, Vangelisti A, Giordani T, Cavallini A, Natali L. Specific LTR-Retrotransposons Show Copy Number Variations between Wild and Cultivated Sunflowers. Genes. 2018; 9.

Mascarini-Serra L. Prevention of soil-transmitted helminth infection. J Glob Infect Dis. 2011; 3: 175.

Masny A, Sałamatin R, Rozej-Bielicka W, Golab E. Is molecular xenomonitoring of mosquitoes for *Dirofilaria repens* suitable for dirofilariosis surveillance in endemic regions? Parasitol Res. 2016; 115: 511–525.

Masuda Y, Kishimoto T, Ito H, Tsuji M. Visceral larva migrans caused by *Trichuris vulpis* presenting as a pulmonary mass. Thorax. 1987; 42: 990-991.

Mationg MLS, Gordon CA, Tallo VL, Olveda RM, Alday PP, Reñosa MDC, et al. Status of soil-transmitted helminth infections in schoolchildren in Laguna Province, the Philippines: Determined by parasitological and molecular diagnostic techniques. PLoS Negl Trop Dis. 2017; 11: e0006022.

Maurelli MP, Rinaldi L, Alfano S, Pepe P, Coles GC, Cringoli G, et al. Mini-FLOTAC, a new tool for copromicroscopic diagnosis of common intestinal nematodes in dogs. Parasit Vectors. 2014; 7: 356.

Mbabazi PS, Andan O, Fitzgerald DW, Chitsulo L, Engels D, Downs JA. Examining the Relationship between Urogenital Schistosomiasis and HIV Infection. PLoS Negl Trop Dis. 2011; 5: e1396.

McReynolds LA, DeSimone SM, Williams SA. Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi*. Proc Natl Acad Sci USA. 1986; 83: 797–801.

McReynolds LA, Poole CB, Williams SA, Partono F. Diagnosis of Filariasis with DNA Probes. In: Rapid Methods and Automation in Microbiology and Immunology. Berlin, Heidelberg: Springer Berlin Heidelberg; 1991. pp. 402–409.

Means AR, Ajjampur SSR, Bailey R, Galactionova K, Gwayi-Chore MC, Halliday K, et al. Evaluating the sustainability, scalability, and replicability of an STH transmission interruption intervention: The DeWorm3 implementation science protocol. PLoS Negl Trop Dis. 2018; 12: e0005988.

Medley GF, Turner HC, Baggaley RF, Holland C, Hollingsworth TD. The Role of More Sensitive Helminth Diagnostics in Mass Drug Administration Campaigns: Elimination and Health Impacts. Adv Parasitol. 2016; 94: 343–392.

Meekums H, Hawash MB, Sparks AM, Oviedo Y, Sandoval C, Chico ME, et al. A genetic analysis of *Trichuris trichiura* and *Trichuris suis* from Ecuador. Parasit Vectors. 2015; 8: 168.

Mejia R, Vicuña Y, Broncano N, Sandoval C, Vaca M, Chico M, et al. A novel, multi-parallel, real-time polymerase chain reaction approach for eight gastrointestinal parasites provides improved diagnostic capabilities to resource-limited at-risk populations. Am J Trop Med Hyg. 2013; 88: 1041-1047.

Melrose W, Menzies H, Boer M, Joseph H, Reeve D, Speare R. Short Communication: A Simple Method for Performing Worm-Egg Counts on Sodium Acetate Formaldehyde-Preserved Samples. J Parasitol Res. 2012; 2012: 1–3.

Meredith SE. Unnasch TR, Karam M, Piessens WF, Wirth DF. Cloning and characterization of an *Onchocerca volvulus* specific DNA sequence. Mol Biochem Parasitol. 1989; 36: 1–10.

Meštrović N, Pavlek M, Car A, Castagnone-Sereno P, Abad P, Plohl M. Conserved DNA Motifs, Including the CENP-B Box-like, are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. PLoS One. 2013; 8: e67328.

Midzi N, Kavhu B, Manangazira P, Phiri I, Mutambu SL, Tshuma C, et al. Inclusion of edaphic predictors for enhancement of models to determine distribution of soil-transmitted helminths: the case of Zimbabwe. Parasit Vectors. 2018; 11: 47.

Mines J. Modifications of Mcmaster Worm Egg Counting Method. Aust Vet J. 1977; 53: 342–343.

Minetti C, Pilotte N, Zulch MF, Canelas T, Tettevi EJ, Balunnaa F, et al. Amplification of human bloodborne eukaryotic pathogens using mosquito feces: A field evaluation of filariasis, malaria and mansonellosis. PLoS Negl Trop Dis. 2019; Under Review.

Mohd-Shaharuddin N, Lim YAL, Hassan N-A, Nathan S, Ngui R. Molecular characterization of *Trichuris* species isolated from humans, dogs and cats in a rural community in Peninsular Malaysia. Acta Trop. 2019; 190: 269–272.

Moser W, Bärenbold O, Mirams GJ, Cools P, Vlaminck J, Ali SM, et al. Diagnostic comparison between FECPAKG2 and the Kato-Katz method for analyzing soil-transmitted helminth eggs in stool. PLoS Negl Trop Dis. 2018; 12: e0006562.

Mravinac B, Ugarković E, Franjević D, Plohl M. Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. J Mol Evol. 2005; 60: 513-525.

Müller I, Gall S, Beyleveld L, Gerber M, Pühse U, Du Randt R, et al. Shrinking risk profiles after deworming of children in Port Elizabeth, South Africa, with special reference to *Ascaris lumbricoides* and *Trichuris trichiura*. Geospat Health. 2017; 12: 601.

Müller I, Yap P, Steinmann P, Damons BP, Schindler C, Seelig H, et al. Intestinal parasites, growth and physical fitness of schoolchildren in poor neighbourhoods of Port Elizabeth, South Africa: a cross-sectional survey. Parasit Vectors. 2016; 9: 488.

Nekrutenko A, Makova KD, Baker RJ. Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene. 2000; 249: 47–51.

Nery SV, McCarthy JS, Traub R, Andrews RM, Black J, Gray D, et al. A cluster-randomized controlled trial integrating a community-based water, sanitation and hygiene programme with mass distribution of albendazole to reduce intestinal parasites in Timor-Leste: the WASH for WORMS research protocol. BMJ Open. 2015; 5: e009293.

Ng-Nguyen D, Stevenson MA, Dorny P, Gabriël S, Vo TV, Nguyen VT, et al. Comparison of a new multiplex real-time PCR with the Kato Katz thick smear and copro-antigen ELISA for the detection and differentiation of *Taenia* spp. in human stools. PLoS Negl Trop Dis. 2017; 11: e0005743.

Ngui R, Lim YAL, Traub RJ, Mahmud R, Mistam MS. Epidemiological and genetic data supporting the transmission of *Ancylostoma ceylanicum* among human and domestic animals. PLoS Negl Trop Dis. 2012; 6: e1522.

Nikolay B, Brooker SJ, Pullan RL. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. Int J Parasitol. 2014; 44: 765–774.

Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. 2000; 28: E63.

Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, et al. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 2017; 45: e111–e111.

Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010; 11: 378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792-793.

Nunes CM, Lima LGF, Manoel CS, Pereira RN, Nakano MM, Garcia JF. Fecal Specimens Preparation Methods for PCR Diagnosis of Human Taeniosis. Rev Inst Med trop São Paulo. 2006; 48, 45–47.

O'Connell EM, Mitchell T, Papaiakovou M, Pilotte N, Lee D, Weinberg M, et al. *Ancylostoma ceylanicum* Hookworm in Myanmar Refugees, Thailand, 2012-2015. Emerg Infect Dis. 2018; 24: 1472-1481.

O'Connell EM, Nutman TB. Molecular Diagnostics for Soil-Transmitted Helminths. Am J Trop Med Hyg. 2016; 95: 508–513.

Okorie PN, de Souza DK. Prospects, drawbacks and future needs of xenomonitoring for the endpoint evaluation of lymphatic filariasis elimination programs in Africa. Trans R Soc Trop Med Hyg. 2016; 110: 90–97.

Okoyo C, Nikolay B, Kihara J, Simiyu E, Garn JV, Freeman MC, et al. Monitoring the impact of a national school based deworming programme on soil-transmitted helminths in Kenya: the first three years, 2012 – 2014. Parasit Vectors. 2016; 9: 408.

Onyabe DY, Conn JE. Intragenomic heterogeneity of a ribosomal DNA spacer (ITS2) varies regionally in the neotropical malaria vector *Anopheles nuneztovari* (Diptera: Culicidae). Insect Mol Biol. 1999; 8: 435–442.

Opoku M, Minetti C, Kartey-Attipoe WD, Otoo S, Otchere J, Gomes B, et al. An assessment of mosquito collection techniques for xenomonitoring of anopheline-transmitted Lymphatic Filariasis in Ghana. Parasitology. 2018; 145: 1783–1791.

Pabalan N, Singian E, Tabangay L, Jarjanazi H, Boivin MJ, Ezeamama AE. Soil-transmitted helminth infection, loss of education and cognitive impairment in school-aged children: A systematic review and meta-analysis. PLoS Negl Trop Dis. 2018;12: e0005523.

Palmer CS, Traub RJ, Robertson ID, Hobbs RP, Elliot A, While L, et al. The veterinary and public health significance of hookworm in dogs and cats in Australia and the status of *A. ceylanicum*. Vet Parasitol. 2007; 145: 304–313.

Papaiakovou M, Gasser RB, Littlewood DTJ. Quantitative PCR-Based Diagnosis of Soil-Transmitted Helminth Infections: Faecal or Fickle? Trends Parasitol. 2019; 35: 491–500.

Papaiakovou M, Pilotte N, Baumer B, Grant J, Asbjornsdottir K, Schaer F, et al. A comparative analysis of preservation techniques for the optimal molecular detection of hookworm DNA in a human fecal specimen. PLoS Negl Trop Dis. 2018; 12: e0006130.

Papaiakovou M, Pilotte N, Grant JR, Traub RJ, Llewellyn S, McCarthy JS, et al. A novel, species-specific, real-time PCR assay for the detection of the emerging zoonotic parasite *Ancylostoma ceylanicum* in human stool. PLoS Negl Trop Dis. 2017; 11: e0005734.

Parija SC, Chidambaram M, Mandal J. Epidemiology and clinical features of soil-transmitted helminths. Trop Parasitol. 2017; 7: 81–85.

Parker DM, Tripura R, Peto TJ, Maude RJ, Nguon C, Chalk J, et al. A multi-level spatial analysis of clinical malaria and subclinical Plasmodium infections in Pailin Province, Cambodia. Heliyon. 2017; 3: e00447.

Pecson BM, Antonio Barrios J, Johnson DR, Nelson KL. A Real-Time PCR Method for Quantifying Viable *Ascaris* Eggs Using the First Internally Transcribed Spacer Region of Ribosomal DNA. Appl Environ Microbiol. 2006; 72: 7864-7872.

Pezer Ž., Brajković J, Feliciello I, Ugarkovc Đ. Satellite DNA-Mediated Effects on Genome Regulation. Genome Dyn. 2012; 7: 153-169.

Pham-Duc P, Nguyen-Viet H, Hattendorf J, Zinsstag J, Phung-Dac C, Zurbrügg C, et al. *Ascaris lumbricoides* and *Trichuris trichiura* infections associated with wastewater and human excreta use in agriculture in Vietnam. Parasitol Int. 2013; 62: 172–180.

Phuphisut O, Yoonuan T, Sanguankiat S, Chaisiri K, Maipanich W, Pubampen S, et al. Triplex Polymerase Chain Reaction Assay for Detection of Major Soil-Transmitted Helminths, *Ascaris lumbricoides*, *Trichuris trichiura*, *Necator americanus*, in Fecal Samples. Southeast Asian J Trop Med Public Health. 2014; 45: 267-275.

Pickering AJ, Njenga SM, Steinbaum L, Swarthout J, Lin A, Arnold BF, et al. Effects of Single and Integrated Water, Sanitation, Handwashing, and Nutrition Interventions on Child Soil-Transmitted Helminth and Giardia Infections: A Cluster-Randomized Controlled Trial in Rural Kenya. 2019; PLoS Med. 16; e1002841.

Piednoël M, Aberer AJ, Schneeweiss GM, Makas J, Novak P, Gundlach H, et al. Next-Generation Sequencing Reveals the Impact of Repetitive DNA Across Phylogenetically Closely Related Genomes of *Orobanchaceae*. Mol Biol Evol. 2012; 29: 3601–3611.

Piepenburg O, Williams CH, Stemple DL, Armes NA. DNA detection using recombination proteins. PLoS Biol. 2006; 4: e204.

Pilotte N, Cook DAN, Pryce J, Zulch MF, Minetti C, Reimer LJ, et al. Laboratory evaluation of molecular xenomonitoring using mosquito excreta/feces to amplify *Plasmodium*, *Brugia*, and *Trypanosoma* DNA. Gates Open Res. 2019; In press.

Pilotte N, Maasch JRMA, Easton AV, Dahlstrom E, Nutman TB, Williams SA. Targeting a highly repeated germline DNA sequence for improved real-time PCR-based detection of *Ascaris* infection in human stool. PLoS Negl Trop Dis. 2019; 13: e0007593.

Pilotte N, Papaiakovou M, Grant JR, Bierwert L, Llewellyn S, McCarthy JS, Williams SA. Improved PCR-Based Detection of Soil Transmitted Helminth Infections Using a Next-Generation Sequencing Approach to Assay Design. PLoS Negl Trop Dis. 2016; 10: e0004578.

Pilotte N, Torres M, Tomaino FR, Laney SJ, Williams SA. A TaqMan-based multiplex real-time PCR assay for the simultaneous detection of *Wuchereria bancrofti* and *Brugia malayi*. Mol Biochem Parasitol. 2013; 189: 33-37.

Pilotte N, Unnasch TR, Williams SA. The Current Status of Molecular Xenomonitoring for Lymphatic Filariasis and Onchocerciasis. Trends Parasitol. 2017; 33: 788–798.

Pilotte N, Zaky WI, Abrams BP, Chadee DD, Williams SA. A Novel Xenomonitoring Technique Using Mosquito Excreta/Feces for the Detection of Filarial Parasites and Malaria. PLoS Negl Trop Dis. 2016; 10, e0004641.

Plohl M, Luchetti A, Mestrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. 2008; 409: 72–82.

Pontes LA, Oliveira MC, Katz N, Dias-Neto E, Rabello A. Comparison of a polymerase chain reaction and the Kato-Katz technique for diagnosing infection with *Schistosoma mansoni*. Am J Trop Med Hyg. 2003; 68: 652–656.

Poole CB, Williams SA. A rapid DNA assay for the species-specific detection and quantification of *Brugia* in blood samples. Mol Biochem Parasitol. 1990; 40: 129–136.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007; 35: 7188–7196.

Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit Vectors. 2014; 7: 37.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia. 2015. Available: https://www.r-project.org.

Ralph A, O'Sullivan MVN, Sangster NC, Walker JC. Abdominal pain and eosinophilia in suburban goat keepers - trichostrongylosis. Med J Aust. 2006; 184: 467–469.

Ramesh A, Cameron M, Spence K, Hoek Spaans R, Melo-Santos MAV, Paiva MHS, et al. Development of an urban molecular xenomonitoring system for lymphatic filariasis in the Recife Metropolitan Region, Brazil. PLoS Negl Trop Dis. 2018; 12: e0006816.

Ramírez AL, Hall-Mendelin S, Doggett SL, Hewitson GR, McMahon JL, Ritchie SA. Mosquito excreta: A sample type with many potential applications for the investigation of Ross River virus and West Nile virus ecology. 2018; PLoS Negl Trop Dis. 12: e0006771.

Rao RU, Atkinson LJ, Ramzy RMR, Helmy H, Farid HA, Bockarie MJ, et al. A Real-Time PCR Assay for Detection of *Wuchereria bancrofti* DNA in Blood and Mosquitoes. Am J Trop Med Hyg. 2006; 74: 826-832.

Rao RU, Nagodavithana KC, Samarasekera SD, Wijegunawardana AD, Premakumara WD, Perera SN, et al. A comprehensive assessment of lymphatic filariasis in Sri Lanka six years after cessation of mass drug administration. PLoS Negl Trop Dis. 2014; 8: e3281.

Rao RU, Samarasekera SD, Nagodavithana KC, Punchihewa MW, Dassanayaka TDM, Gamini PDK, et al. Programmatic Use of Molecular Xenomonitoring at the Level of Evaluation Units to Assess Persistance of Lymphatic Filariasis in Sri Lanka. PLoS Negl Trop Dis. 2016; 10: e0004722.

Rao RU, Weil GJ, Fischer K, Supali T, Fischer P. Detection of Brugia Parasite DNA in Human Blood by Real-Time PCR. J Clin Microbiol. 2006; 44: 3887–3893.

Rao RU, Williams SA, Klei TR. Quantification of PCR amplification products of *Brugia* HhaI repeat DNA using a semiautomated Q-PCR system. Mol Cell Probes. 2002; 16: 13–23.

Rep BH, Heinemann DW. Changes in hookworm distribution in Surinam. Trop Geogr Med. 1976; 28: 104–110.

Reynoldson JA, Behnke, JM, Pallant LJ, Macnish MG, Gilbert F, Giles S, et al. Failure of pyrantel in treatment of human hookworm infections (*Ancylostoma duodenale*) in the Kimberley region of north west Australia. Acta Trop. 1997; 68: 301-312.

Ricciardi A, Ndao M. Diagnosis of Parasitic Infections: What's Going On? J Biomol Screen. 2015; 20: 6-21.

Rodríguez FR, de la Herrán R, Navajas-Pérez R, Cano-Roldán B, Sola-Campoy PJ, García-Zea JA, et al. Centromeric Satellite DNA in Flatfish (Order Pleuronectiformes) and Its Relation to Speciation Processes. J Hered. 2016; esw076.

Rodríguez-Pérez MA, Gopal H, Adeleke MA, De Luna-Santillana EJ, Gurrola-Reyes JN, Guo X. Detection of *Onchocerca volvulus* in Latin American black flies for pool screening PCR using high-throughput automated DNA isolation for transmission surveillance. Parasitol Res. 2013; 112: 3925-3931.

Rodríguez-Pérez MA, Lilley BG, Domínguez-Vázquez A, Segura-Arenas R, Lizarazo-Ortega C, Mendoza-Herrera A, et al. Polymerase chain reaction monitoring of transmission of *Onchocerca volvulus* of two endemic states in Mexico. Am J Trop Med Hyg. 2004; 70: 38-45.

Roizes G. Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. Nucleic Acids Res. 2006; 34: 1912–1924.

Roset R, Subirana JA, Messeguer X. MREPATT: detection and analysis of exact consecutive repeats in genomic sequences. Bioinformatics. 2003; 19: 2475-2476.

Sakano T, Hamamoto K, Kobayashi Y, Sakata Y, Tsuji M, Usui T. Visceral larva migrans caused by *Trichuris vulpis*. Arch Dis Child. 1980; 55: 631-633.

Sanpool O, Intapan PM, Thanchomnang T, Sri-Aroon P, Lulitanond V, Sadaow L, et al. Development of a real-time PCR assay with fluorophore-labelled hybridization probes for detection of *Schistosoma mekongi* in infected snails and rat feces. Parasitology. 2012; 139: 1266–1272.

Sanprasert V, Kerdkaew R, Srirungruang S, Charuchaibovorn S, Phadungsaksawasdi K, Nuchprayoon S. Development of Conventional Multiplex PCR: A Rapid Technique for Simultaneous Detection of Soil-Transmitted Helminths. Pathogens. 2019; 8: 152.

Sarhan RM, Kamel HH, Saad GA, Ahmed OA. Evaluation of three extraction methods for molecular detection of *Schistosoma mansoni* infection in human urine and serum samples. J Parasit Dis. 2015; 39: 499–507.

Schaap M, Lemmers RJ, Maassen R, van der Vliet PJ, Hoogerheide LF, van Dijk HK, et al. Genome-wide analysis of macrosatellite repeat copy number variation in worldwide populations: evidence for differences and commonalities in size distributions and size restrictions. BMC Genomics. 2013; 14: 143.

Schär F, Odermatt P, Khieu V, Panning M, Duong S, Muth S, et al. Evaluation of real-time PCR for *Strongyloides stercoralis* and hookworm as diagnostic tool in asymptomatic schoolchildren in Cambodia. Acta Trop. 2013; 126: 89-92.

Schmaedick MA, Koppel AL, Pilotte N, Torres M, Williams SA, Dobson SL, et al. Molecular xenomonitoring using mosquitoes to map lymphatic filariasis after mass drug administration in American Samoa. PLoS Negl Trop Dis. 2014; 8: e3087.

Scholte RG, Schur N, Bavia ME, Carvahlo EM, Chammartin F, Utzinger J, et al. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. Geospat Health. 2013; 8: 97-110.

Scholz T, Uhlírová M, Ditrich O. Helminth parasites of cats from the Vientiane province, Laos, as indicators of the occurrence of causative agents of human parasitoses. Parasite. 2003; 10: 343–350.

Schuster RK, Thomas K, Sivakumar S, O'Donovan D. The parasite fauna of stray domestic cats (Felis catus) in Dubai, United Arab Emirates. Parasitol Res. 2009; 105: 125–134.

Schwarz EM, Hu Y, Antoshechkin I, Miller MM, Sternberg PW, Aroian RV. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. Nat genet. 2015; 47: 416-422.

Scott ME. *Ascaris lumbricoides*: A Review of Its Epidemiology and Relationship to Other Infections. Ann Nestlé. 2008; 66: 7-22.

Setasuban P, Vajrasthira S, Muennoo C. Prevalence and zoonotic potential of *Ancylostoma ceylanicum* in cats in Thailand. Southeast Asian J Trop Med Public Health. 1976; 7: 534–539.

Shalaby HA. Anthelmintics Resistance; How to Overcome it? Iran J Parasitol. 2013; 8: 18-32.

Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. Biol Rev. 2005; 80: 227–250.

Shen C-H. Diagnostic molecular biology. 1st edition. San Diego, CA: Elsevier; 2018.

Sinniah B. Daily egg production of *Ascaris lumbricoides*: the distribution of eggs in the faeces and the variability of egg counts. Parasitology. 1982; 84: 167-175.

Sleigh A, Hoff R, Mott K, Barreto M, de Paiva TM, de Souza Pedrosa J, et al. Comparison of filtration staining (Bell) and thick smear (Kato) for the detection and quantitation of *Schistosoma mansoni* eggs in faeces. Trans R Soc Trop Med Hyg. 1982; 76: 403–406.

Soto LA, Santísma-Trinidad AB, Bornay-Llinares FJ, González MM, Valero JAP, Muñoz MR. Quantitative PCR and Digital PCR for Detection of *Ascaris lumbricoides* Eggs in Reclaimed Water. BioMed Res Int. 2017; 2017: 1–9.

Sow D, Parola P, Sylla K, Ndiaye M, Delaunay P, Halfon P, et al. Performance of Real-Time Polymerase Chain Reaction Assays for the Detection of 20 Gastrointestinal Parasites in Clinical Samples from Senegal. Am J Trop Med Hyg. 2017; 97: 173–182.

Speich B, Ali SM, Ame SM, Albonico M, Utzinger J, Keiser J. Quality control in the diagnosis of *Trichuris trichiura* and *Ascaris lumbricoides* using the Kato-Katz technique: experience from three randomised controlled trials. Parasit Vectors. 2015; 8: 82.

Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, et al. Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted helminth diagnosis in epidemiological surveys. Parasit Vectors. 2010; 3: 71.

Staudacher O, Heimer J, Steiner F, Kayonga Y, Havugimana JM, Ignatius R, et al. Soil-transmitted helminths in southern highland Rwanda: associated factors and effectiveness of school-based preventive chemotherapy. Trop Med Int Health. 2014; 19: 812-824.

Steinbaum L, Kwong LH, Ercumen A, Negash MS, Lovely AJ, Njenga SM, et al. Detecting and enumerating soil-transmitted helminth eggs in soil: New method development and results from field testing in Kenya and Bangladesh. PLoS Negl Trop Dis. 2017; 11: e0005522.

Steinbaum L, Njenga SM, Kihara J, Boehm AB, Davis J, Null C, et al. Soil-Transmitted Helminth Eggs Are Present in Soil at Multiple Locations within Households in Rural Kenya. PLoS One. 2016; 11: e0157780.

Stensvold CR, Saijuntha W, Sithithaworn P, Wongratanacheewin S, Strandgaard H, Ornbjerg N, et al. Evaluation of PCR based coprodiagnosis of human opisthorchiasis. Acta Trop. 2006; 97: 26–30.

Subirana JA, Albà MM, Messeguer X. High evolutionary turnover of satellite families in Caenorhabditis. BMC Evol Biol. 2015; 15: 218.

Subirana JA, Messeguer X. A Satellite Explosion in the Genome of Holocentric Nematodes. PLoS One. 2013; 8: e62221.

Subirana JA, Messeguer XA. Satellite explosion in the genome of holocentric nematodes. PloS One. 2013; 8: e62221.

Sungkar S, Ridwan AS, Kusumowidagdo G. The Effect of Deworming Using Triple-Dose Albendazole on Nutritional Status of Children in Perobatang Village, Southwest Sumba, Indonesia. J Parasitol Res. 2017; 2017: 1–5.

Tahapary DL, de Ruiter K, Martin I, van Lieshout L, Guigas B, Soewondo P, et al. Helminth infections and type 2 diabetes: a cluster-randomized placebo controlled SUGARSPIN trial in Nangapanda, Flores, Indonesia. BMC Infect Dis. 2015; 15: 133.

Tanaka R, Hino A, Tsai IJ, Palomares-Rius JE, Yoshida A, Ogura Y, et al. Assessment of Helminth Biodiversity in Wild Rats Using 18S rDNA Based Metagenomics. PLoS ONE. 2014; 9: e110769.

Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, et al. Genome of the human hookworm *Necator americanus*. Nat Genet. 2014; 46: 261-269.

Tarafder MR, Carabin H, Joseph L, Balolong E, Olveda R, McGarvey ST. Estimating the sensitivity and specificity of Kato-Katz stool examination technique for detection of hookworms, *Ascaris lumbricoides* and *Trichuris trichiura* infections in humans in the absence of a 'gold standard.' Int J Parasitol. 2010; 40: 399–404.

Teruel M, Ruíz-Ruano FJ, Marchal JA, Sánchez A, Cabrero J, Camacho JP, et al. Disparate molecular evolution of two types of repetitive DNAs in the genome of the grasshopper *Eyprepocnemis plorans*. Heredity. 2014; 112: 531–542.

Thanchomnang T, Intapan PM, Lulitanond V, Choochote W, Manjai A, Prasongdee TK, et al. Rapid detection of *Brugia malayi* in mosquito vectors using a real-time fluorescence resonance energy transfer PCR and melting curve analysis. Am J Trop Med Hyg. 2008; 78: 509-513.

The C. elegans Sequencing Consortium. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. Science. 1998; 282: 2012-2018.

Tobler H, Etter A, Müller F. Chromatin diminution in nematode development. Trends Genet. 1992; 8: 427-432.

Tran AC, Halse TA, Escuyer VE, Musser KA. Detection of *Mycobacterium avium* complex DNA directly in clinical respiratory specimens: opportunities for improved turn-around time and cost savings. Diagn Microbiol Infect Dis. 2014; 79: 43-48.

Traub RJ, Inpankaew T, Sutthikornchai C, Sukthana Y, Thompson RCA. PCR-based coprodiagnostic tools reveal dogs as reservoirs of zoonotic ancylostomiasis caused by *Ancylostoma ceylanicum* in temple communities in Bangkok. Vet Parasitol. 2008; 155: 67–73.

Traub RJ, Robertson ID, Irwin P, Mencke N, Thompson RCA. Application of a species-specific PCR-RFLP to identify *Ancylostoma* eggs directly from canine faeces. Vet Parasitol. 2004; 123: 245–255.

Traub RJ. *Ancylostoma ceylanicum*, a re-emerging but neglected parasitic zoonosis. Int J Parasitol. 2013; 43: 1009-1015.

Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011; 13: 36-46.

Truscott JE, Dunn JC, Papaiakovou M, Schaer F, Werkman M, Littlewood DTJ, et al. Calculating the prevalence of soil-transmitted helminth infection through pooling of stool samples: Choosing and optimizing the pooling strategy. PLoS Negl Trop Dis 2019; 13: e0007196.

Truscott JE, Turner HC, Farrell SH, Anderson RM. Soil-Transmitted Helminths: Mathematical Models of Transmission, the Impact of Mass Drug Administration and Transmission Elimination Criteria. Adv. Parasitol. 2016; 94: 133–198.

Truscott JE, Werkman M, Wright JE, Farrell SH, Sarkar R, Ásbjörnsdóttir K, et al. Identifying optimal threshold statistics for elimination of hookworm using a stochastic simulation model. Parasit Vectors. 2017; 10: 321.

Turner HC, Bettis AA, Dunn JC, Whitton JM, Hollingsworth TD, Fleming FM, et al. Economic Considerations for Moving beyond the Kato-Katz Technique for Diagnosing Intestinal Parasites As We Move Towards Elimination. Trends Parasitol. 2017; 33: 435–443.

U.S. Department of Health and Human Services Food and Drug Administration. Statistical guidance on reporting results from studies evaluating diagnostic tests. 2007. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.

Ulukanligil M, Seyrek A, Aslan G, Ozbilge H, Atay S. Environmental pollution with soil-transmitted helminths in Sanliurfa, Turkey. Mem Inst Oswaldo Cruz. 2001; 96: 903-909.

Umemori J, Mori A, Ichiyanagi K, Uno T, Koide T. Identification of both copy number variation-type and constant-type core elements in a large segmental duplication region of the mouse genome. BMC Genomics. 2013; 14: 455.

van Mens SP, Aryeetey Y, Yazdanbakhsh M, van Lieshout L, Boakye D, Verweij JJ. Comparison of real-time PCR and Kato smear microscopy for the detection of hookworm infections in three consecutive faecal samples from schoolchildren in Ghana. Trans R Soc Trop Med Hyg. 2013; 107: 269–271.

van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent Class Models in Diagnostic Studies When There is No Reference Standard—A Systematic Review. Am J Epidemiol. 2014; 179: 423–431.

Vargas N, Souto RP, Carranza JC, Vallejo GA, Zingales B. Amplification of a specific repetitive DNA sequence for *Trypanosoma rangeli* identification and its potential application in epidemiological investigations. Exp Parasitol. 2000; 96: 147–159.

Vaz Nery S, Pickering AJ, Abate E, Asmare A, Barrett L, Benjamin-Chung J, et al. The role of water, sanitation and hygiene interventions in reducing soil-transmitted helminths: interpreting the evidence and identifying next steps. Parasit Vectors. 2019; 12: 273.

Vaz Nery S, Traub RJ, McCarthy JS, Clarke NE, Amaral S, Llewellyn S, et al. WASH for WORMS: A Cluster-Randomized Controlled Trial of the Impact of a Community Integrated Water, Sanitation, and Hygiene and Deworming Intervention on Soil-Transmitted Helminth Infections. Am J Trop Med Hyg. 2019; 100: 750–761.

Vercruysse J, Albonico M, Behnke JM, Kotze AC, Prichard RK, McCarthy JS, et al. Is anthelminthic resistance a concern for the control of human soil-transmitted helminths? Int J Parasitol Drugs Drug Resist. 2011; 1: 14-27.

Vercruysse J, Levecke B, Prichard R. Human soil-transmitted helminths: implications of mass drug administration. Curr Opin Infect Dis. 2012; 25: 703-708.

Verweij JJ, Brienen EAT, Ziem J, Yelifari L, Polderman AM, Van Lieshout L. Simultaneous Detection and Quantification of *Ancylostoma duodenale*, *Necator americanus*, and *Oesophagostomum bifurcum* in Fecal Samples Using Multiplex Real-Time PCR. Am J Trop Med Hyg. 2007; 77: 685-690.

Verweij JJ, Canales M, Polman K, Ziem J, Brienen EAT, Polderman AM, et al. Molecular diagnosis of *Strongyloides stercoralis* in fecal samples using real-time PCR. Trans R Soc Trop Med Hyg. 2009; 103: 342-346.

Verweij JJ, Stensvold, C. R. Molecular testing for clinical diagnosis and epidemiological investigations of intestinal parasitic infections. 2014; Clin Microbiol Rev. 2014; 27: 371–418.

Verweij JJ. Application of PCR-based methods for diagnosis of intestinal parasitic infections in the clinical laboratory. Parasitol. 2014; 141: 1863–1872.

Vestheim H, Jarman SN. Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. Front Zool. 2008; 5: 12.

Vincent M, Xu Y, Kong H. Helicase-dependent isothermal DNA amplification. EMBO Rep. 2004; 5: 795–800.

Walden J. Parasitic diseases. Other roundworms. *Trichuris*, hookworm, and *Strongyloides*. Prim Care. 1991; 18: 53–74.

Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988; 41: 923–937.

Wang J, Gao S, Mostovoy Y, Kang Y, Zagoskin M, Sun Y, et al. Comparative genome analysis of programmed DNA elimination in nematodes. Genome Res. 2017; 27: 2001-2014.

Wang J, Mitreva M, Berriman M, Thorne A, Magrini V, Koutsovoulos G, et al. Silencing of germline-expressed genes by DNA elimination in somatic cells. Dev Cell. 2012; 23: 1072-1080.

Wang J. RePS: A Sequence Assembler That Masks Exact Repeats Identified from the Shotgun Data. Genome Res. 2002; 12: 824–831.

Wanji S, Amvongo-Adjia N, Njouendou AJ, Kengne-Ouafo JA, Ndongmo WPC, Fombad FF, et al. Further evidence of the cross-reactivity of the Binax NOW® Filariasis ICT cards to non-*Wuchereria bancrofti* filariae: experimental studies with *Loa loa* and *Onchocerca ochengi*. Parasites Vectors. 2016; 9: 267.

Wardell R, Clements ACA, Lal A, Summers D, Llewellyn S, Campbell SJ, et al. An environmental assessment and risk map of *Ascaris lumbricoides* and *Necator americanus* distributions in Manufahi District, Timor-Leste. PLoS Negl Trop Dis. 2017; 11: e0005565.

Weatherhead JE, Hotez PJ, Mejia R. The Global State of Helminth Control and Elimination in Children. Pediatr Clin. 2017; 64: 867–877.

Weiss JB. DNA probes and PCR for diagnosis of parasitic infections. Clin Microbiol Rev. 1995; 8: 113–130.

Werkman M, Wright JE, Truscott JE, Easton AV, Oliveira RG, Toor J, et al. Testing for soil-transmitted helminth transmission elimination: Analysing the impact of the sensitivity of different diagnostic tools. PLoS Negl Trop Dis. 2018; 12: e0006114.

Werneck JS, Carniato T, Gabriel A, Tufik S, Andrade SS. Mites in clinical stool specimens: potential misidentification as helminth eggs. Trans R Soc Trop Med Hyg. 2007; 101: 1154–1156.

WHO Expert Committee. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. World Health Organ Tech Rep Ser. 2002; 912: i–vi, 1–57.

WHO. Assessing the Epidemiology of Soil-Transmitted Helminths During a Transmission Assessment Survey in the Global Programme for the Elimination of Lymphatic Filariasis. Geneva, Switzerland: WHO Press; 2015.

Wickstead B, Ersfeld K, Gull K. Repetitive Elements in Genomes of Parasitic Protozoa. Microbiol Mol Biol Rev. 2003; 67: 360–375.

Williams SA, Lizotte-Waniewski MR, Foster J, Guiliano D, Daub J, Scott AL, et al. The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. Int J Parasitol. 2000; 30: 411–419.

Wiria AE, Prasetyani MA, Hamid F, Wammes LJ, Lell B, Ariawa I, et al. Does treatment of intestinal helminth infections influence malaria? Background and methodology of a longitudinal study of clinical, parasitological and immunological parameters in Nangapanda, Flores, Indonesia (ImmunoSPIN Study). BMC Infect Dis. 2010; 10: 77.

World Health Organization. Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation. 2012. Available: http://www.who.int/neglected_diseases/NTD_RoadMap_2012_Fullversion.pdf

Yamamoto Y. PCR in diagnosis of Infection: detection of bacteria in cerebrospinal fluids. Clin Diagn Lab Immunol. 2002; 9: 508–514.

Yong TS, Lee JH, Sim S, Lee J, Min DY, Chai JY, et al. Differential diagnosis of *Trichostrongylus* and hookworm eggs via PCR using ITS-1 sequence. Korean J Parasitol. 2007; 45: 69–74.

Zaky WI, Tomaino FR, Pilotte N, Laney SJ, Williams SA. Backpack PCR: A point-of-collection diagnostic platform for the rapid detection of Brugia parasites in mosquitoes. PLoS Negl Trop Dis. 2018; 12: e0006962.

Zdybel J, Karamon J, Dąbrowska J, Różycki M, Bilska-Zając E, Kłapeć T, et al. Parasitological contamination with eggs *Ascaris* spp., *Trichuris* spp. and *Toxocara* spp. of dehydrated municipal sewage sludge in Poland. Environ Pollut. 2019; 248: 621–626.

Zerdo Z, Yohanes T, Tariku B. Soil-Transmitted Helminth Reinfection and Associated Risk Factors among School-Age Children in Chencha District, Southern Ethiopia: A Cross-Sectional Study. J Parasitol Res. 2016; 2016: 1–7.

Zhao S, Gibbons JGA. Population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*. PLoS ONE. 2018; 13: e0201611.

Zhong M, McCarthy J, Bierwert L, Lizotte-Waniewski M, Chanteau S, Nutman TB, et al. A polymerase chain reaction assay for detection of the parasite *Wuchereria bancrofti* in human blood samples. Am J Trop Med Hyg. 1996; 54: 357–363.